Final: May 17 (W) 6:30 pm In WH 100E Notice the time changes !

Introduction to Statistics (Math 502)

WH 100E MWF 8:30am-9:30am
Office: WH 132
Office hours: M, T 3-4pm
Textbook: Statistical Inference (2nd ed.)
by George Casella and Roger L. Berger
Chapter 6 - Chapter 10.
Homework due: W in class.
Quiz: 8:20am Every Friday,
Midterm: March 20 (M)

Final: May 17 (W) 6:30 pm In WH 100E Notice the time changes !

Each is allowed to bring a piece of paper with anything on it.

Homework assigned during last week is due each Wednesday.
It is on my website: http://www.math.binghamton.edu/qyu
Remind me if you do not see it by Saturday morning !
Homework due this Friday is on my website !!!
The solution is on my website. Grade yourself carefully and hand in.

Grading Policy: 50% hw and quizzes +50% exams
B = 70 ±

Chapter 0. Introduction

Question: What is Statistics ?
One can use the following example to explain in short.

Example (capture-recapture problem).
In a pond, there are N fish.
Catch m, say $m = 10$,
tag them and put them back.
Re-catch n fish, say $n = 10$,
X of them are tagged, say $X = 3$.

Question: $\begin{cases} P(X=x)=? & \text{probability problem} \\ N = ? & \text{statistics problem.} \end{cases}$

Answer:

1. $f(x;N) = P(X = x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$, $x \in \{0, 1, ..., n \wedge m\}$, $n \vee m < N$.
2. Many estimates of N: MLE, MME, LSE, etc. *e.g.* MLE: $\check{N} = \text{argmax}_N f(3; N) => \check{N} = 40$. MME: Solve $\overline{X} = E(X) = nm/N => \hat{N} = nm/X = 33\frac{1}{3}$.
   Properties of these estimators ?
   What is the best estimator ?

Typically, statistics deals with such problems:
Given a random sample, say X_1, ..., X_n, i.i.d. from X,
assuming they are from a model with cdf $F(x;\theta)$, where θ unknown in Θ
find out: θ = ? or $P(X \le x)$ = ? (This is called *point estimation*).
What is θ in the capture-recapture problem ?

We shall study

1. how to summarize X_1, ..., X_n,
2. how to find a formula to guess θ based on the summary,
3. whether the guess with such a formula is good.

Chapter 6. Principles of Data Reduction

Denote $\mathbf{X} = (X_1, ..., X_n)$, where X_1, ..., X_n, i.i.d. from $X \sim F(x;\theta)$.
We call $\mathbf{X}$ a data set or observations from X.
One can use R to generate data set in simulation:

```
> x=rnorm(3,0,1)
```

[1] 0.3163466 0.4865695 -0.2163855
> x=rexp(30,3) # 3=E(X) or 1/E(X) ? ($f(x) \propto e^{-x/\mu} = e^{-\rho x}$, $x > 0$).
> mean(x)
[1] 0.3559676

Definition. Given data $\mathbf{X}$, a statistic $T(\mathbf{X})$ is a function of $\mathbf{X}$, where $T(\cdot)$ does not depend on θ.
A data set is often quite large,
for estimation purpose, it is desirable to simplify it to a statistic.
However, we do not want to lose information during the simplification.
This is called data reduction.

Several principles for data reduction:
(1) sufficiency principle,
(2) likelihood principle (maybe ignored in the lecture),
(3) invariancy principle (maybe ignored in the lecture).

§1. **Sufficiency**

Let $\mathbf{X}$ be a random vector (continuous or discrete),
with the density function (d.f.) $f_{\mathbf{X}}(\mathbf{x};\theta)$.

Definition. If $T(\mathbf{X})$ is a statistic and the conditional distribution of $\mathbf{X}$ given T, say $(\mathbf{X}|T)$, is independent of θ, then T is a sufficient statistic for θ
(or we say that T is sufficient for θ).

Sufficiency principle: Reduce the data to a sufficient statistic.

Theorem 1. (Factorization theorem).
Let f be the d.f. of $\mathbf{X}$, and $T(\mathbf{X})$ a statistic.
T is sufficient for θ iff

$$f(\mathbf{x};\theta) = g(T(\mathbf{x});\theta)h(\mathbf{x}), \ \forall \ (\mathbf{x},\theta) \text{ where } h \text{ does not depend on } \theta. \tag{1}$$

Recall that a family of distributions, say $\{f(\cdot;\theta): \ \theta \in \Theta\}$,
is said to belong to an exponential family if

$$f(x;\theta) = h(x)c(\theta)exp(\sum_{j=1}^{k} w_j(\theta)t_j(x)),$$

where h and t_i's are independent of θ and c and w_i's are independent of $\mathbf{x}$.

Theorem 2. If X_1, ..., X_n are i.i.d. from an exponential family, and if $T(\mathbf{X}) = \sum_{i=1}^n (t_1(X_i), ..., t_k(X_i))$, then T is sufficient for θ.

Remark. 3 methods for determining a sufficient statistic:
(1) Definition. $f_{\mathbf{X}|T(\mathbf{X})}$ is independent of θ.
(2) Factorization Th. $f(\mathbf{x};\theta) = g(T(\mathbf{x});\theta)h(\mathbf{x})$.
(3) Exponential family. $T(\mathbf{X}) = \sum_{i=1}^n \mathbf{t}(X_i)$, $\mathbf{t} = (t_1, ..., t_k)$.

Method (3) is most convenient, but not always work. **Why ?**
Method (1) is not convenient, but always works.
Method (2) is convenient most of the time and always works.

Example 1. Suppose that X_1, ..., X_n ($n \geq 3$) is a random sample from $bin(1,\theta)$. Are $T(\mathbf{X})$ sufficient for θ in the following cases ?
(a) $T = \mathbf{X}$,
(b) $T = \sum_{i=1}^n X_i + 1$,
(c) $T = \overline{X}$,
(d) $T = X_1 + X_2$,
(e) $T = \overline{X} + \theta$.

Sol. (a) $\vdash$: $T(\mathbf{X}) = \mathbf{X}$ is sufficient for θ.
By (1). $P(\mathbf{X} = \mathbf{x}|T(\mathbf{X}) = \mathbf{y}) = \mathbf{1}(\mathbf{x} = \mathbf{y})$ is independent of θ.
By (2). $f_{\mathbf{X}}(\mathbf{x};\theta) = \prod_{i=1}^n f_X(x_i;\theta) = \underbrace{\prod_{i=1}^{n} f_X(x_i;\theta)}_{g(T(\mathbf{X});\theta)} \times \underbrace{1}_{h(\mathbf{X})}$.

Way (3) is not applicable though $bin(1,\theta)$ belongs to an exponential family, as $\sum_{i=1}^n \mathbf{t}(X_i)$ $(= \sum_i X_i)$ is 1 dimensional, but $T = \mathbf{X}$ is $n-$dimensional.

(b) $\vdash$: $T(\mathbf{X}) = \sum_i X_i + 1$ is sufficient for θ.
By (1). $\vdash$: $f_{\mathbf{X}|T}$ is independent of θ.
$P(\mathbf{X} = \mathbf{x}|T = t) = P(A|B) = P(AB)/P(B)$.
$\sum_i X_i \sim bin(n,\theta)$.
$P(B) = P(\sum_i X_i + 1 = t) = P(\sum_i X_i = t-1)$.
$P(AB) = P(\mathbf{X} = \mathbf{x}, \sum_i X_i + 1 = t)$
$= \mathbf{1}(\sum_i x_i = t-1)P(\mathbf{X} = \mathbf{x})$
$= \mathbf{1}(\sum_i x_i = t-1)\theta^{t-1}(1-\theta)^{n-t+1}$ **why ?**
Thus $P(\mathbf{X} = \mathbf{x}|T = t) = \frac{\mathbf{1}(\sum_i x_i = t-1)}{\binom{n}{t-1}}$ is independent of θ.
By (2). $f_{\mathbf{X}}(\mathbf{x};\theta) = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i} = \underbrace{\theta^{T(\mathbf{X})-1}(1-\theta)^{n-T(\mathbf{X})+1}}_{g(T(\mathbf{X});\theta)} \times \underbrace{1}_{h(\mathbf{X})}$.
By (3). $\vdash$: $T = \sum_{i=1}^n t(X_i) = \sum_i X_i + 1$, where $f_X(x) = h(x)c(\theta)\exp(w(\theta)t(x))$.
$f_X(x;\theta) = \theta^x(1-\theta)^{1-x} = (\frac{\theta}{1-\theta})^x(1-\theta) = (\frac{\theta}{1-\theta})^{x+1/n}(\frac{\theta}{1-\theta})^{-1/n}(1-\theta)$
$= (\frac{\theta}{1-\theta})^{-1/n}(1-\theta)exp((x+1/n)\ln\frac{\theta}{1-\theta})$.
Thus $k = 1$, $w_1(\theta) = \ln\frac{\theta}{1-\theta}$ and $t_1(x) = x + 1/n$.
It yields $T = \sum_{i=1}^n t_1(X_i) = \sum_i X_i + 1$.

(c) $\vdash$: $\overline{X}$ is sufficient for θ.
$\overline{X} = \frac{1}{n}\sum_{=1}^n X_i$ $(= \frac{1}{n}((\sum_i X_i + 1) - 1/n))$,
a linear combination of T in (b)).
The proof is similar to that in (b).

(d) $\vdash$: $X_1 + X_2$ is not sufficient for θ.
Choose a counterexample:
$P(\mathbf{X} = (0,\cdots,0,1)|T = 0) = \frac{P(X_1=\cdots=X_{n-1}=0,X_n=1)}{P(X_1=X_2=0)}$
$= P(X_3 = \cdots = X_{n-1} = 0, X_n = 1) = \theta(1-\theta)^{(n-2)-1} = \begin{cases} 0 & \text{if } \theta = 0 \\ 0.5^{n-2} & \text{if } \theta = 0.5 \end{cases}$.
It depends on θ.
Q: Can we use methods (2) and (3) ? Why ?

(e) $\vdash$: $\overline{X} + \theta$ is not a sufficient statistic for θ.
Reason: T depends on θ, **Proof ?** Let $n = 3$ and $X_i = 1$, then $T(\mathbf{X}) = \begin{cases} 2 & \text{if } \theta = 1 \\ 1.5 & \text{if } \theta = 0.5 \end{cases}$
thus, it is not even a statistic,
let alone a sufficient statistic.

Remark. Sufficient statistics are not unique or equivalent.
$\mathbf{X}$, $\overline{X}$, $\sum_{i=1}^n X_i + 1$ are all sufficient for θ. Which is preferred ?
Remark. 3 methods for determining a sufficient statistic:
(1) Definition. (2) Factorization Th. (3) Exponential family.
Example 2. Does the family of distributions belong to an exponential family in the following cases ?
(a) $N(\mu,\sigma^2)$,
(b) $bin(m,p)$,
(c) Poisson with mean μ (P(μ)),
(d) $Exp(\theta)$ with mean $1/\theta$,
(e) Double exponential distribution $f = \frac{1}{2\lambda}\exp(-\frac{|x-\mu|}{\lambda})$,
(f) U(a,b).
Sol. Yes for (a) through (d), explained in 501. No for (e) and (f).
Reason for (e) is explained in 501.

Reason for $U(a, b)$: Note that $\theta = (a, b)$ is the parameter.
If it belongs to an exponential family,
say

$$f_X(x;\theta) = h(x)c(\theta)exp\{\sum_{j=1}^{k} w_j(\theta)t_j(x)\} = \frac{\mathbf{1}(x \in (a,b))}{b-a} \quad (1)$$

then it is impossible that $f_X(x;\theta) = \frac{\mathbf{1}_{(x\in(a,b))}}{b-a} = 0$ for all $x \notin [a, b]$,
as h and t_i's are independent of (a, b). **Done ??**

Give a counterexample: If Eq.(1) holds, for $(a, b) = (0, 2)$, $f_X(x;\theta) = \frac{\mathbf{1}_{(x\in(a,b))}}{b-a} = 1/2$ for $x \in (0, 2)$yields $h(x) > 0$ for $x \in (0, 2)$ **Why ??**
but for $(a, b) = (0, 1)$ $f_X(x;\theta) = \frac{\mathbf{1}_{(x\in(a,b))}}{b-a} = 0$ for $x > 1$yields $h(x) = 0$ for $x \in (1, 2)$ **Why ??**
A contradiction. Thus it does not belong to an exponential family.

Example 3. Let X_1, ..., X_n be i.i.d. from $N(\mu, \sigma^2)$. Find a sufficient statistic for θ in the following cases.
(a) μ known, (b) σ known,(c) both unknown.

Sol. $f(x;\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2}\frac{x^2}{\sigma^2} + \frac{x\mu}{\sigma^2} - \frac{1}{2}\frac{\mu^2}{\sigma^2})$

$$f_X(x;\theta) = h(x)c(\theta)exp\{\sum_{j=1}^{k} w_j(\theta)t_j(x)\}$$

(a) $\theta = \sigma^2$, $T = \sum_i(X_i - \mu)^2$ or $-\frac{1}{2}\overline{(X-\mu)^2}$, etc. **Why ?**
(b) $\theta = \mu$, $T = \sum_i X_i$ or $\overline{X}$.
(c) $\theta = (\mu, \sigma^2)$, $T = (\sum_i X_i, \sum_i X_i^2)$, etc.

Example 4. Find a (non-trivial) sufficient statistic for $U(0, b)$ if X_1, ..., X_n is a random sample from $U(0, b)$.

Sol. Question: What is a trivial sufficient statistics fo $U(0, b)$?
Question: Can we use Method (3) ?
No! $U(0, b)$ does not belong to an Exponential family.

Question: Can we use Method (1) ?
Not convenient ! as we have no idea on what is T.

Method (2) is a good tool for finding a suitable sufficient statistic for non-exponential family.

$$f_{\mathbf{X}}(\mathbf{x};\theta) = \prod_{i=1}^{n} f_X(x_i;\theta) = \prod_{i=1}^{n} \frac{1}{b}, \quad x_1, ..., x_n \in (0, b)\textbf{??}$$

Correct approach:

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x};\theta) &= \prod_{i=1}^{n} \frac{\mathbf{1}(x_i \in (0,b))}{b} \\ &= \underbrace{\frac{1}{b^n}\mathbf{1}(x_1, ..., x_n \in (0,b))}_{g(T(\mathbf{X});b)} \times \underbrace{1}_{h(\mathbf{X})} \\ &= \underbrace{\frac{1}{b^n}\mathbf{1}(x_{(1)}, x_{(n)} \in (0,b))}_{g(T(\mathbf{X});b)} \times \underbrace{1}_{h(\mathbf{X})} \qquad (x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)} - \text{order statistics}) \\ &= \underbrace{\frac{1}{b^n}\mathbf{1}(x_{(n)} < b)}_{g(T(\mathbf{X});b)} \times \underbrace{1_{(x_{(1)}>0)}}_{h(\mathbf{X})} \end{aligned}$$

Sufficient statistics:
(a) $T = \mathbf{X}$ (trivial one),
(b) $T_2 = (X_{(1)}, X_{(n)})$,

(c) $T_3 = X_{(n)}$.
Which one you prefer ?

To find a sufficient statistic, it is not convenient to use the definition.
Seen from the examples, the dimension of a sufficient statistic can be n or smaller.
It is desirable to find a sufficient statistic that has the smallest dimension.
Definition. A sufficient statistic T is called a minimal sufficient statistic (MSS), if for any other sufficient statistic T^*, T is a function of T^*.

A MSS = a sufficient statistic with the least dimension ?

Consider the case that $n = 2$, X_1 and X_2 are i.i.d. from $U(\theta, \theta+1)$, θ is unkown. $T_1 = (X_1, X_2)$ and $T_2 = (X_{(1)}, X_{(2)})$.
Which is likely an MSS ?
Theorem 3. *Suppose that*
(1) $f(\mathbf{x};\theta)$ is the density function of $\mathbf{X}$;
(2) $T(\mathbf{X})$ is a statistic;
(3) $f(\mathbf{x};\theta)/f(\mathbf{y};\theta)$ is independent of θ iff $T(\mathbf{x}) = T(\mathbf{y})$ $\forall$ $(\mathbf{x}, \mathbf{y})$.
Then T is MSS.
Question. How to get the density function $f_{\mathbf{X}}$ for a random sample from f_X ?
Remark. Two ways to determine a MSS.
1. Definition.
2. Theorem 3.

Example 5. Let X_1, ..., X_n be a random sample from $N(\mu, \sigma^2)$, where $\theta = (\mu, \sigma^2)$. Find a MSS for θ.
Sol. A sufficient statistic is $T(\mathbf{X}) = (\overline{X}, \overline{X^2})$.
To show that it is MSS,
the definition is not convenient, we use Theorem 3.
Since $f(\mathbf{x};\theta) \propto \exp(-\frac{n}{2}\frac{\overline{x^2}}{\sigma^2} + \frac{n\overline{x}\mu}{\sigma^2} - \frac{n\mu^2}{2\sigma^2})$
$f(\mathbf{x};\theta)/f(\mathbf{y};\theta) = \exp(-\frac{n}{2}\frac{\overline{x^2}-\overline{y^2}}{\sigma^2} + \frac{n(\overline{x}-\overline{y})\mu}{\sigma^2}) = 1$ $\forall$ θ iff $T(\mathbf{x}) = T(\mathbf{y})$, $\forall$ $(\mathbf{x}, \mathbf{y})$.
Thus T is MSS.
Example 6. Suppose thatX_1, ..., X_n are i.i.d. $\sim U(\theta, \theta+1)$.
Find a MSS for $U(\theta, \theta+1)$.
Sol. (1) Find a suitable sufficient statistic; (2) Show that it is MSS.

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x};\theta) &= \prod_{i=1}^{n} \mathbf{1}(x_i \in (\theta, \theta+1)) \\ &= \underbrace{\mathbf{1}(\theta < x_{(1)}, x_{(n)} < \theta+1)}_{g(x_{(1)}, x_{(n)};\theta)} \times \underbrace{1}_{h(\mathbf{x})}, \quad \theta \in (-\infty, \infty) \end{aligned}$$

$T = (X_{(1)}, X_{(n)})$ is sufficient for θ by the Factorization theorem.
$\vdash$: T is MSS. That is,

$$\frac{f_{\mathbf{X}}(\mathbf{x};\theta)}{f_{\mathbf{X}}(\mathbf{y};\theta)} = \frac{\mathbf{1}(\theta < x_{(1)}, x_{(n)} < \theta+1)}{\mathbf{1}(\theta < y_{(1)}, y_{(n)} < \theta+1)} = \begin{cases} 1 & \text{if } T(\mathbf{x}) = T(\mathbf{y}) \\ \textit{depends on } \theta & \text{if } T(\mathbf{x}) \neq T(\mathbf{y}) \end{cases}$$

where $\frac{0}{0} \stackrel{def}{=} 1$. It suffices to show to show $\frac{f_{\mathbf{X}}(\mathbf{x};\theta)}{f_{\mathbf{X}}(\mathbf{y};\theta)}$ is not constant in θ, if $T(\mathbf{x}) \neq T(\mathbf{y})$.

$T(\mathbf{x}) \neq T(\mathbf{y})$ implies either (1) $x_{(1)} < y_{(1)}$ (or $x_{(1)} > y_{(1)}$), or (2) $x_{(n)} < y_{(n)}$ (or $x_{(n)} > y_{(n)}$).
By symmetry between $\mathbf{x}$ and $\mathbf{y}$, we just need to consider either (1) $x_{(1)} < y_{(1)}$ or (2) $x_{(n)} < y_{(n)}$.
By symmetry between $-(\mathbf{x}, \mathbf{y})$ and $(\mathbf{x}, \mathbf{y})$,
we just need to prove case (1) $x_{(1)} < y_{(1)}$.
$\frac{f_{\mathbf{X}}(\mathbf{x};\theta)}{f_{\mathbf{X}}(\mathbf{y};\theta)} = \begin{cases} \frac{0}{0} & \text{if } y_{(1)} < \theta \\ \frac{0}{1} & \text{if } x_{(1)} < \theta < y_{(1)} \text{ and } y_{(n)} < \theta+1 \end{cases}$ **done ?**
Need to give 2 θ's

Since $f(x;\theta) = \mathbf{1}(x \in (\theta, \theta+1))$, $x_{(n)} - x_{(1)} \in (0,1)$ and $y_{(n)} - y_{(1)} \in (0,1)$, $(x_{(n)} \vee y_{(n)}) - (x_{(1)} \wedge y_{(1)}) \in (0,1)$ **why ?**

$$\frac{f_{\mathbf{X}}(\mathbf{x};\theta)}{f_{\mathbf{X}}(\mathbf{y};\theta)} = \begin{cases} \frac{0}{0} & \text{if } \theta = y_{(1)} + 2 \textbf{ Why 2 ?} \\ \frac{0}{1} & \text{if } \theta = \frac{x_{(1)}+y_{(1)}}{2} \\ & (\text{as } x_{(1)} < \theta < y_{(1)} \text{ and } y_{(n)} < \theta + 1) \\ & \text{due to } 0 < (y_{(n)} \vee x_{(n)}) - (y_{(1)} \wedge x_{(1)}) < 1 \\ & => y_{(1)} \wedge x_{(1)} < \theta < y_{(n)} \vee x_{(n)} < \theta + 1 \end{cases} \tag{2}$$

=> It depends on θ if $x_{(1)} < y_{(1)}$.

Thus $\frac{f_{\mathbf{X}}(\mathbf{x};\theta)}{f_{\mathbf{X}}(\mathbf{y};\theta)}$ is independent of θ iff $T(\mathbf{x}) = T(\mathbf{y})$ $\forall$ $(\mathbf{x},\mathbf{y})$. □

Definition. Suppose that $\{f(x;\theta) : \theta \in \Theta\}$ is a family of density functions, X_1, ..., X_n are i.i.d. from f. $T = T(\mathbf{X})$ is a statistic.
T is said to be *ancillary* if f_T does not depend on θ.
T is said to be a *complete* statistic or complete for θ (or for the distribution family), if $\forall$ function g such that $g(T)$ is a statistic, we have

$$E(g(T)) = 0 \ \forall \ \theta \Rightarrow P(g(T) = 0) = 1 \ \forall \ \theta.$$

Theorem 4. If an MSS exists, then each complete statistics is MSS.

Theorem 5. Suppose (1) X_1, ..., X_n are i.i.d. from $f(x;\theta)$, $\theta \in \Theta \subset R^k$. (2) $f = h(x)c(\theta)exp(\sum_{j=1}^k w_j(\theta)t_j(x))$; write $\mathbf{w} = (w_1, ..., w_k)$; (3) $\{\mathbf{w}(\theta) : \theta \in \Theta\}$ contains a non-empty open set of R^k; then $T = \sum_{i=1}^n \mathbf{t}(X_i)$ is complete; where $\mathbf{t} = (t_1, ..., t_k)$.

Q: Are w and t uniquely determined ?

Remark. Two ways to determine whether T is complete:

1. Definition;
2. Exponential family by Theorem 5.

Example 7. Let X_1, ..., X_n be i.i.d. from X. Is $T(\mathbf{X})$ complete for θ ?
(a) $T = (\overline{X}, \overline{X^2})$, where $X \sim N(\mu, \sigma^2)$ and $\theta = (\mu, \sigma^2)$.
(b) $T = (\overline{X}, \overline{X^2})$, where $X \sim N(\theta, \theta^2)$.
(c) $T = X_{(n)}$, where $X \sim U(0, \theta)$.

Sol.(a) Exponential family. $\{\mathbf{w}(\theta) : \theta \in \Theta\}$=?
Notice $f \propto exp(-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2}) \propto exp(-\frac{n}{2\sigma^2}x^2/n + \frac{n\mu}{\sigma^2}x/n)$
$(w_1, w_2) = (-\frac{n}{2\sigma^2}, \frac{n\mu}{\sigma^2}) = (-\frac{n}{2\sigma^2}, -2\mu\frac{-n}{2\sigma^2})$, $\mu \in (-\infty, \infty)$, $\sigma \in (0, \infty)$,
Why factor n **?** (check Th 5).
$\{\mathbf{w}(\theta) : \theta \in \Theta\} = (-\infty, 0) \times (-\infty, \infty)$. **Question:** Why ?
It follows that $\{\mathbf{w}(\theta) : \theta \in \Theta\}$ contains a non-empty open set in R^2.
Thus by Theorem 5, T is complete.

Remark. Notice that T is also MSS by Example 5.

(b) Q: $\{\mathbf{w}(\theta) : \theta \in \Theta\} = (-\infty, 0) \times (-\infty, \infty)$ **?**
$(w_1, w_2) = (-\frac{n}{2\theta^2}, \frac{n}{\theta})$, $\theta > 0$.

$\{\mathbf{w}(\theta) : \theta \in \Theta\} = \{(w_1, w_2) : \ 2w_1/n = (w_2/n)^2, w_2 > 0\}$ is a curve in $\mathcal{R}^2$.
It does not contain an open set in $\mathcal{R}^2$.
Condition (3) in Theorem 5 does not hold.
Cannot use Theorem 5, as it only gives sufficient condition for completeness.
$\vdash$: $(\overline{X}, \overline{X^2})$ is not complete for θ.
Use the definition. Need to construct a g such that $E(g(T)) = 0$ but $P(g(T) = 0) < 1$.

How ? Notice that
(1) $E(\overline{X}) = \theta$,
(2) $E((\overline{X})^2) = \mu_{\overline{X}}^2 + \sigma_{\overline{X}}^2 = \theta^2 + \theta^2/n = (1 + \frac{1}{n})\theta^2$,
(3) $E(\overline{X^2}) = E(X^2) = \sigma^2 + \theta^2 = 2\theta^2$.

Now from (2) and (3), setting $g(T) = \frac{(\overline{X})^2}{1+1/n} - \frac{\overline{X^2}}{2}$. Verify $E(g(T)) = 0$, but $P(g(T) = 0) = 0 < 1$ as $g(T)$ is continuous.
Thus T is not complete.

Question: Is T MSS ?

Yes, by Example 5.

Remark. This is an example that T is MSS but it is not complete.

(c) Claim: T is complete.

$U(0, \theta)$ does not belong to an exponential family, thus use the definition.

Need to compute $E(g(T)) = \int g(t) f_T(t) dt$.

f_T= ?

Formula: $f_{X_{(n)}}(t) = \frac{n!}{(n-1)!1!}(F_X(t))^{n-1}(f_X(t))^1$.

$f_T(t) = nt^{n-1}/\theta^n$, $t \in (0, \theta)$.

(**Or derive it directly as follows.**

$f_T(t) = F'_T(t)$.

$F_T(t) = P(T \leq t) = P(X_{(n)} \leq t) = P(X_1 \leq t, ..., X_n \leq t)$

$= P(X_1 \leq t)) \cdots P(X_n \leq t) = (F(t))^n = t^n/\theta^n$, $t \in (0, \theta)$.

$f_T(t) = nt^{n-1}/\theta^n$, $t \in (0, \theta)$.)

$E(g(T)) = \int_0^\theta g(t) \frac{nt^{n-1}}{\theta^n} dt = 0 \; \forall \; \theta > 0$,

Does it imply $P(g(T) = 0) = 1$?

Answer. Yes, as $h(t) = g(t) nt^{n-1}/\theta^n = 0$ a.e. (or $\int |h(t)| dt = 0$), *i.e.*, $g(t) = 0$ a.e. by the lemma as follows.

Lemma 1. If $\int_0^y h(t) dt = 0 \; \forall \; y > 0$, then $h(t) = 0$ a.e..

Note $\int \mathbf{1}(t \in \{-1, 1\}) dt = 0$, but it is not true that $\mathbf{1}(t \in \{-1, 1\}) \equiv 0$).

The proof of Lemma 1 is an exercise in Real Analysis and is quite long. We consider one that is easy to prove (though not quite precise).

Lemma 2. If h is continuous and $\int_0^x h(t) dt = 0 \; \forall \; x > 0$, then $h(t) = 0$.

Proof. $(\int_0^x h(t) dt)' = h(x) = 0 \; \forall \; x > 0$. □

Note that $g(t) t^{n-1}$ may not be continuous. *e.g.*, $g(t) = \mathbf{1}(t \in \{-1, 1\})$.

Recall that

if f_T does not depend on θ, the statistic T is called *ancillary.*

Basu's Theorem. *If $T(\mathbf{X})$ is a complete and MSS statistic, then $T(\mathbf{X}) \perp U(X)$, $\forall$ ancillary statistic $U(\mathbf{X})$.*

Example 8. Suppose that X_1, ..., X_n is a random sample from $U(\theta, \theta + 1)$, $T = X_{(n)} - X_{(1)}$. Show that T is ancillary.

Sol. Two ways to check (1) Direct. (2) Pivotal method.

Direct Way (1): Derive $f_T(t)$ by cdf or Jacobian method.

(1a) By cdf:

$$\begin{aligned}
& P(X_{(n)} - X_{(1)} \leq t) \\
= & \int\int_{y-x \leq t} f_{X_{(1)}, X_{(n)}}(x, y) dx dy \\
= & \int\int_{y-x \leq t} \mathbf{1}(x < y) \underbrace{\frac{n!}{1! \times 1! \times (n-2)!} f(x) f(y) (F(y) - F(x))^{n-2}}_{\textit{similar to trinomial dist}} dx dy \qquad f(x) = \mathbf{1}(x \in (\theta, \theta + 1)) \\
= & \int\int_{y-x \leq t} n(n-1) \mathbf{1}(\theta < x < y < \theta + 1)(y - x)^{n-2} dx dy \qquad \textbf{why}? \\
= & \int\int_{v-u \leq t} n(n-1) \mathbf{1}(0 < u < v < 1)(v - u)^{n-2} du dv \quad (u = x - \theta \;\; v = y - \theta) \\
= & \int\int_{v-u \leq t, 0 < u < v < 1} n(n-1)(v - u)^{n-2} du dv \\
= & \begin{cases} 0 & \text{if } t \leq 0 \\ \int_t^1 \int_0^{v-t} n(n-1)(v - u)^{n-2} du dv & \text{if } t \in (0, 1] \\ 1 & \text{if } t > 1 \end{cases}
\end{aligned}$$

Note that θ disappears, thus $f_T(t) = F_T'(t)$ is independent of θ.
T is ancillary.

(1b) By Jacobian. $(T,W) = (X_{(n)} - X_{(1)}, X_{(n)})$, $f_{T,W}(t,w) = f_{X_{(1)},X_{(n)}}(w-t,w)|J|$. $|J| = ?$ $f_T(t) = \int f_{T,W}(t,w)dw =$ ......
$f_T(t) = \int f_{X_{(1)},X_{(n)}}(w-t,w)dw$..., where

$$f_{X_{(1)},X_{(n)}}(x,y) = \frac{\mathbf{1}(x<y)n!}{1! \times 1! \times (n-2)!} f(x)f(y)(F(y)-F(x))^{n-2} \qquad f(x) = \mathbf{1}(x \in (\theta, \theta+1))$$
$$= n(n-1)\mathbf{1}(\theta < x < y < \theta+1)(y-x)^{n-2} = ...$$

(2) Pivotal method: That is, given $f_X(\cdot;\theta)$, find a pivotal $Z = g(X,\theta)$ such that the density f_Z is independent of θ.
Typical pivotals are related to the location-scale family:
$$Z = \begin{cases} X-\theta & \text{if } f_X(x;\theta) = f(x-\theta) \\ X/\theta & \text{if } f_X(x;\theta) = f(x/\theta)/\theta \\ \frac{X-\mu}{\lambda} & \text{if } f_X(x;\theta) = f(\frac{x-\mu}{\lambda})/\lambda. \end{cases}$$
Then $f_Z(t) = f(t)$.

$$f_X(x) = \mathbf{1}(x \in (\theta, \theta+1)) = \mathbf{1}(\underbrace{x-\theta}_{pivatol} \in (0,1)) = f_Z(x-\theta). \mathbf{= ??}$$
where $Z = X - \theta$ is called a pivotal, and $f_Z(t) = \mathbf{1}(t \in (0,1))$.
To prove T is ancillary, need to show $\vdash$: (i) f_Z is indenpendent of θ $\vdash$: (ii) $T = X_{(n)} - X_{(1)} = Z_{(n)} - Z_{(1)}$.
Then $F_T(t) = P(X_{(n)} - X_{(1)} \le t) = P(Z_{(n)} - Z_{(1)} \le t) = \int \cdots \int_A f_{Z_{(1)},Z_{(n}}(x,y)dxdy$
where $A = \{(x,y):\ y - x \le t\}$ and

$$f_{Z_{(1)},Z_{(n}}(x,y) = \frac{n!}{1!n-2)!1!}(f_Z(x))^1(F(y)-F(x))^{n-2}(f_Z(y))^1 dxdy$$

$\vdash$: (i) f_Z is indenpendent of θ There are two approaches to prove Z is a pivotal as well:
(a) cdf and (b) df.
Approach (a).
Since $F_Z(t) = P(Z \le t) = P(X - \theta \le t)$
$$= P(X \le t+\theta) = \begin{cases} 0 & \text{it } t+\theta < \theta \\ (t+\theta)-\theta & \text{it } \theta \le t+\theta < \theta+1 \\ 1 & \text{it } t+\theta \ge \theta+1 \end{cases}$$
$$= \begin{cases} 0 & \text{if } t < 0 \\ t & \text{if } t \in [0,1) \\ 1 & \text{if } t \ge 1 \end{cases}.$$
$f_{\mathbf{Z}}(\mathbf{z}) = \prod_{i=1}^n \mathbf{1}(z_i \in (0,1))$ is independent of θ, and T is ancillary.
Approach (b). $f_Z(z) = f_X(g^{-1}(z))\left|\frac{\partial g^{-1}}{\partial z}\right|$ where $z = g(x) = x - \theta$.
$g^{-1}(z) = z + \theta$ and $\left|\frac{\partial g^{-1}}{\partial z}\right| = 1$.
Thus $f_Z(z) = f_X(g^{-1}(z)) = \mathbf{1}(z + \theta \in (\theta, \theta+1)) = \mathbf{1}(z \in (0,1))$.
$f_{\mathbf{Z}}(\mathbf{z}) = \prod_{i=1}^n \mathbf{1}(z_i \in (0,1))$ is independent of θ, T is ancillary.
Example 9. Let X_1, ..., X_n be a random sample from $X \sim f(x;\theta)$, where $f = \frac{1}{\theta}f_o(\frac{x}{\theta})$, $\theta > 0$ and f_o is a density function, i.e., $\int f_o(x)dx = 1$, and $f_o \ge 0$. Show that T is ancillary in the two cases: (1) $T = \frac{\overline{X}}{S}$, where $S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$; (2) $T = \frac{X_n}{\overline{X}}$.
Sol. (1) Two ways as in Ex 8. Use the simpler way. Since $f(x) \propto f_o(\frac{x}{\theta})$, $Z = \frac{X}{\theta}$ is a pivotal.
$f_Z(z) = f_X(g^{-1}(z))\left|\frac{\partial g^{-1}}{\partial z}\right|$ $(g = ?)$
$= \frac{1}{\theta}f_o(z)\theta = f_o(z)$.
Let $Z_i = \frac{X_i}{\theta}$. Then $T = \frac{\overline{X}}{S} = \frac{\overline{Z}}{S_Z}$ where $S_Z^2 = \frac{1}{n}\sum_{i=1}^n (Z_i - \overline{Z})^2$.
$F_T(t) = P(\frac{\overline{Z}}{S_Z} \le t) = \int \cdots \int_A f_{\mathbf{Z}}(\mathbf{z})d\mathbf{z}$, where $A = \{\mathbf{z} : \frac{\bar{z}}{S_z} \le t\}$.

$f_{\mathbf{Z}}(\mathbf{z}) = \prod_{i=1}^n f_o(z_i)$ is independent of θ.
Thus T is ancillary.

(2) $T = \frac{X_n}{\overline{X}} = \frac{Z_n}{\overline{Z}}$. Thus T is ancillary too.

Example 10. Let X_1, ..., X_n be i.i.d. $Exp(\theta)$, where $E(X_i) = \theta$. Let $U(\mathbf{X}) = (X_n/\overline{X})^2$. $E(U) = ?$

Sol. Usual way

$$E(U) = \int x f_U(x)dx = \int\int (x/y)^2 f_{X_n,\overline{X}}(x,y)dxdy......$$

Another way: Make use of **Basu's Theorem.**
If $T(\mathbf{X})$ is a complete and MSS statistic, then $T(\mathbf{X}) \perp U(X)$, $\forall$ ancillary statistic $U(\mathbf{X})$.

T is said to be *ancillary* if f_T does not depend on θ.

Recall **Example 9.** Let X_1, ..., X_n be a random sample from $X \sim f(x;\theta)$, where $f = \frac{1}{\theta}f_o(\frac{x}{\theta})$, $\theta > 0$ and f_o is a density function, We show that

$T = X_n/\overline{X}$ is ancillary.

Note that $f_{X_1}(x) \propto e^{-x/\theta}$.

Thus $U(\mathbf{X}) = (X_n/T(\mathbf{X}))^2$ is an ancillary statistic. **How to show it ?**

Let $T(\mathbf{X}) = \overline{X}$. Then T is a complete and MSS statistic. **Why ?**

$E(X_n^2) = E(UT^2) = E(U)E(T^2)$ (by Basu Theorem) $=> E(U) = E(X_n^2)/E(T^2)$.

$\theta^2 + \theta^2 = E(U)(\frac{1}{n^2}(n\theta^2 + (n\theta)^2)$.

$$E(U) = \frac{2\theta^2}{\frac{1}{n^2}(n\theta^2 + (n\theta)^2)}$$

Remark. If $X \sim U(0,1)$
$E(U) = E(X_n^2/(\overline{X})^2) = E(X_n^2)/E((\overline{X})^2)$ **???**

Chapter 7. Point estimation.

Definition. A point estimator is a statistic. Its values are called estimates.

We shall discuss methods of estimation and their optimal properties.

§7.2. Methods of estimation

§7.2.1. Methods of moments estimator (MME)

Suppose that X_1, ..., X_n are i.i.d. from $X \sim f(x;\theta)$, $\theta = (\theta_1, ..., \theta_k) \in \Theta$.
A MME of θ is a solution of θ to equations

$$\begin{cases} \overline{X^{i_1}} = E(X^{i_1}) \\ ...... \\ \overline{X^{i_k}} = E(X^{i_k}), \quad \text{where } i_1, ..., i_k \text{ are distinct integers} \end{cases}$$

Question: Where is θ in these equations ?
In particular, a MME is a solution to

$\overline{X^i} = E(X^i)$, $i = 1$, ..., k.

Remark. The solution to the MME is not unique.

Example 1. Suppose that $X_1 \sim bin(n,p)$, $\theta = p$. MME of θ ?

Sol. We present two solutions, denoted by $\hat{p}$ and $\tilde{p}$.

(1) $\overline{X} = \mu_X$ with $k = 1$.

$X_1 = np$ **Why ?**

$\Rightarrow \hat{p} = X_1/n$.

Question: Why do not say MME is $p = X_1/n$?

(2) $\overline{X^2} = E(X^2)$ with $k = 1$.

$X_1^2 = \sigma^2 + \mu^2 = np(1-p) + (np)^2 = np + (n^2 - n)p^2$

that is, $-X_1^2 + np + (n^2 - n)p^2 = 0$

$\Rightarrow p = \frac{-n \pm \sqrt{n^2 + 4(n^2-n)X_1^2}}{2(n^2-n)}$

Question: Two solutions. Are they both MME ?

Answer: $\tilde{p} = \frac{-n + \sqrt{n^2 + 4(n^2-n)X_1^2}}{2(n^2-n)}$

Example 2. Suppose that $X_1, ..., X_n$ are iid. from $bin(1,p)$, $\theta = p$. MME of θ ?

Sol. Two approaches: (1) Standard, (2) MSS. $T = \sum_{i=1}^n X_i$.

(1) $\overline{X} = p => \hat{p} = \overline{X}$.

(2) $\overline{T} = E(T) => T = np => \hat{p} = \overline{X}$.

Example 3. Suppose that X_1, ..., X_n are i.i.d. from $N(\mu, \sigma^2)$, $\theta = (\mu, \sigma)$. MME of θ ?

Sol. $\begin{cases} \overline{X} = \mu \\ \overline{X^2} = \mu^2 + \sigma^2 \end{cases}$

$\Rightarrow \begin{cases} \hat{\mu} = \overline{X} \\ \hat{\sigma} = \sqrt{\overline{X^2} - (\overline{X})^2} \end{cases}$

§7.2.2. Maximum likelihood estimator (MLE). Assume that $f_{\mathbf{X}}(\mathbf{x};\theta)$ is the density function of $\mathbf{X}$, where $\theta \in \Theta$. Write $\mathcal{L}(\theta) = f_{\mathbf{X}}(\mathbf{x};\theta)$ and call it the likelihood function of θ. The value of θ that maximizes $\mathcal{L}(\cdot)$ over all possible θ in Θ is call the MLE of θ.

$$\hat{\theta} = \text{argmax}_{\theta \in \Theta} \mathcal{L}(\theta)$$

Interpretation: Given $\mathbf{x}$, the MLE chooses θ such that the probability that $\mathbf{X} \approx \mathbf{x}$ is the largest

$$f_{\mathbf{X}}(\mathbf{x}) \begin{cases} = P(\mathbf{X} = \mathbf{x}) & \text{if } \mathbf{X} \text{ is discrete} \\ \approx \frac{P(|\mathbf{X} - \mathbf{x}| < \epsilon)}{(2\epsilon)^n} & \text{if } \mathbf{X} \text{ is continuous} \end{cases}$$

Typical steps for the MLE with differentiable $\mathcal{L}$:

Step 1. Solve for critical points of $\ln\mathcal{L}$

(*i.e.*, all t's such that $(\ln\mathcal{L})'(t) = 0$ or $\mathcal{L}'(t)$ does not exist, or the boundary).

Step 2. Check whether t is the maximum point by

either the second derivative test if $\mathcal{L}'$ exists everywhere,

or comparing the value $\mathcal{L}(t)$ over all t obtained in step 1.

Example 1. Suppose that X_1, ..., X_n are i.i.d. from $N(\theta, 1)$. Find the MLE of θ in the following cases: (a) $\Theta = (-\infty, \infty)$, (b) $\Theta = [0, \infty)$,(c) $\Theta = [-1, 1]$.

Sol. Denote $\ln\mathcal{L}(\theta) = \ln \prod_{i=1}^n f_X(X_i;\theta)$

$= \ln\{\frac{1}{(2\pi)^{n/2}} \exp(-\sum_{i=1}^n \frac{(X_i-\theta)^2}{2})\}$

$= \ln\frac{1}{(2\pi)^{n/2}} - \sum_{i=1}^n \frac{(X_i-\theta)^2}{2}$

Remark. It is much clearer by drawing the graph of $y = \ln\mathcal{L}(x)$. A parabola concaving down.

(a) $\Theta = \mathcal{R}^1$.

$\ln\mathcal{L}(\theta)' = \sum_i (X_i - \theta) = 0 \Rightarrow \theta = \overline{X}$.

Check: $(\ln\mathcal{L})'$ exists on $\mathcal{R}^1$, and $(\ln\mathcal{L})'' < 0$. Thus $\hat{\theta} = \overline{X}$ is the MLE.

(b) $\Theta = [0, \infty)$.

Possible critical points: $\theta = \overline{X}$, 0, ∞.

Check: Two cases: (1) $\overline{X} > 0$, (2) $\overline{X} \leq 0$.

$$(1) \quad \begin{array}{cccc} \textit{critical pts}: & 0 & \overline{X} & \infty \\ \ln\mathcal{L}(\cdot) & c - \frac{n}{2}\overline{X^2} & c - \frac{n}{2}(\overline{X^2} - (\overline{X})^2) & -\infty \\ (\ln\mathcal{L}(\cdot))' & + & 0 & - \\ & & MLE & \end{array} \quad \textbf{Do we need both?}$$

$$(2) \quad \begin{array}{ccc} \textit{critical pts}: & 0 & \infty \\ \ln\mathcal{L}(\cdot) & \textit{finite} & -\infty \\ & MLE & \end{array}$$

Thus the MLE $\hat{\theta} = \max\{0, \overline{X}\}$.

(c) $\Theta = [-1, 1]$.

Possible critical points: $\theta = \overline{X}$, -1, 1.

Check: 3 cases: (1) $\overline{X} \in (-1, 1)$, (2) $\overline{X} \leq -1$, (3) $\overline{X} \geq 1$.

$$(1) \quad \begin{array}{cccc} \textit{critical points}: & -1 & \overline{X} & 1 \\ \ln\mathcal{L}(\cdot) & ? & ? & ? \\ (\ln\mathcal{L}(\cdot))' & + & 0 & - \\ & & MLE & \end{array} \quad \text{simple if } \mathbf{X} \text{ is given}$$

$$(2) \quad \begin{array}{ccc} \textit{critical points}: & -1 & 1 \\ \ln\mathcal{L}(\cdot) & ? & ? \\ (\ln\mathcal{L}(\cdot))' & - & - \\ & MLE & \end{array}$$

$$
(3) \quad \begin{array}{ccc} \textit{critical points}: & -1 & 1 \\ \ln\mathcal{L}(\cdot) & ? & ? \\ (\ln\mathcal{L}(\cdot))' & + & + \\ & & MLE \end{array}
$$

Thus the MLE $\hat{\theta} = \begin{cases} \overline{X} & \text{if } \overline{X} \in [-1, 1] \\ -1 & \text{if } \overline{X} < -1 \\ 1 & \text{if } \overline{X} > 1. \end{cases}$

Example 2. Suppose that X_1, ..., X_n are i.i.d. from $bin(k, p)$ where p is known, $p \in (0, 1)$, and k is unknown. MLE of k ?

Solution. **Question:** What is Θ ?

$$
\mathcal{L} = \prod_{i=1}^{n} \binom{k}{X_i} p^{X_i} q^{k-X_i} = (\prod_{i=1}^{n} \frac{1}{X_i!})(\frac{p}{q})^{\sum_i X_i} (\prod_{i=1}^{n} \frac{1}{(k-X_i)!})(k!)^n q^{nk}
$$

Remark. If $X_{(n)} = 0$, $\mathcal{L} = q^{nk}$ is maximized by $k = 1$. Thus $\hat{k} = 1$ if $X_{(n)} = 0$. WLOG, assume $X_{(n)} \geq 1$.

Question: Should we use the typical method ? *i.e.*, $\frac{\partial \ln\mathcal{L}(k)}{\partial k} = 0$?

(1) $\frac{\partial}{\partial k}(k!) = ?$ (2) $\theta = k$ is discrete, the root of $\frac{\partial \ln\mathcal{L}(k)}{\partial k}$ may not be an integer.

Notice that X_1, ..., $X_n \leq k$. Thus the MLE

$$
\hat{k} \geq \max\{X_{(n)}, 1\}.
$$

One method: Guess and try.

The MLE $\hat{k} = \text{argmax}_{k \geq X_{(n)} \vee 1} \mathcal{L}(k)$.

An R program in the special case of $(n, p, X_1) = (1, 0.8, 5)$:

```
X=5
p=0.8
N=20
K=X:N
f=choose(K,X)*p**X*(1-p)**(K-X)
F=max(f)
round(F,3)
round(f,3)
[1] 0.393
[1] 0.328 0.393 0.275 0.147 0.066 0.026 0.010 0.003 0.001 0.000 0.000 0.000
[13] 0.000 0.000 0.000 0.000
```

$(= \binom{k}{X_1} p^{X_1} q^{k-X_1})$ **($\hat{k}$ = ??)**

Then the MLE is $\hat{k} = 6$, **Why ?** according to $f(5), ..., f(20)$.

```
K[f==F]
[1] 6
```

Remark. Drawback of this approach: It is not clear that $\hat{k}$ is the MLE, as we only list $k \in \{5, 6, ..., 20\}$.

Second approach: Consider $g(k) = \frac{\mathcal{L}(k)}{\mathcal{L}(k-1)}$. *e.g.* let $n = 1$, then

$g(k) = \frac{k}{k-X_1} q = \frac{1}{1-X_1/k} q$ decreases from ∞ to q (< 1) for $k \in [X_{(n)} \vee 1, \infty)$; $X_{(n)}$ **= ??**

=> (1) $g(\hat{k}) \geq 1$ and (2) $g(\hat{k}+1) \leq 1$. **Why ?**

(1) $\quad \frac{q}{1-X_1/k} \geq 1 \Rightarrow q \geq 1 - X_1/k \Rightarrow X_1/k \geq p \ (= 1-q) \Rightarrow X_1/p \geq k.$

(2) $\quad \frac{q}{1-X_1/(k+1)} \leq 1 \Rightarrow q \leq 1 - X_1/(k+1) \Rightarrow X_1/(k+1) \leq p \Rightarrow X_1/p \leq k+1.$

Thus $\frac{X_1}{p} - 1 \leq k \leq \frac{X_1}{p}$. If $(n, p, X_1) = (1, 0.8, 5)$, then $5.25 \leq \hat{k} \leq 6.25 \Rightarrow \hat{k} = 6$. **Why ?**

Now in general,

$$
g(k) = \frac{(\prod_{i=1}^{n} \frac{k!}{X_i!(k-X_i)!}) p^{\sum_i X_i} q^{nk - \sum_i X_i}}{(\prod_{i=1}^{n} \frac{(k-1)!}{X_i!((k-1)-X_i)!}) p^{\sum_i X_i} q^{n(k-1) - \sum_i X_i}} = (\prod_{i=1}^{n} \frac{k}{k-X_i}) q^n
$$

$$
g(k) = \frac{\mathcal{L}(k)}{\mathcal{L}(k-1)} = (\prod_{i=1}^{n} \frac{1}{1-X_i/k}) q^n, \ k \geq (X_{(n)} \vee 1), \text{ where } \tfrac{1}{0} = \infty. \tag{1}
$$

WLOG, assume $X_{(n)} \geq 1$. $\frac{1}{1-X_i/k} \downarrow$ in $k \in [X_{(n)}, \infty)$, $\forall\, i$.

$$
\text{Then } g(k) \text{ decreases from } \infty \text{ to } q^n \ (< 1) \text{ on } [X_{(n)}, \infty). \tag{2}
$$

By Eq. (1), at the MLE $\hat{k}$, $\begin{cases} \mathcal{L}(\hat{k}-1) \leq \mathcal{L}(\hat{k}) \\ \mathcal{L}(\hat{k}+1) \leq \mathcal{L}(\hat{k}). \end{cases}$ *i.e.*, $\begin{cases} g(\hat{k}) \geq 1 \\ g(\hat{k}+1) \leq 1. \end{cases}$

Statement (2) says $y = g(x)$ is a decreasing curve that crosses $y = 1$.
We should look for $\hat{k} \geq X_{(n)}$ such that

$g(\hat{k}) \geq 1$;
$g(\hat{k}+1) \leq 1$.

Q: $g(\hat{k}) = 1$ **??**
The MLE can be written as

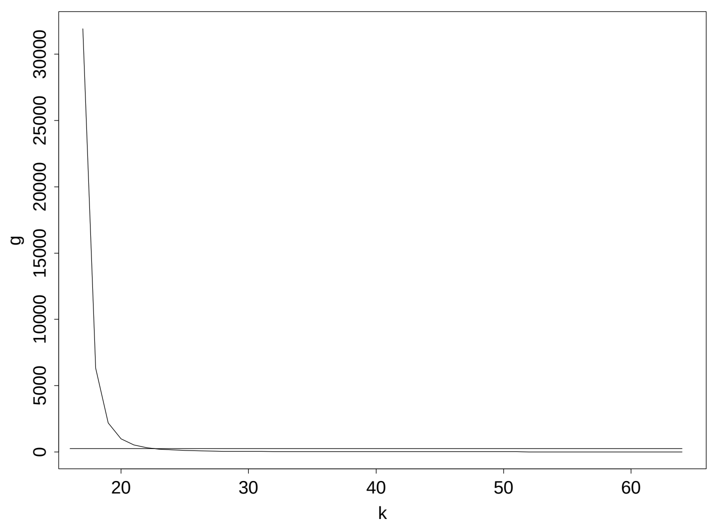
$\hat{k} = \max\{k : \ g(k) \geq 1, k \geq X_{(n)}\}$ **?**
$\hat{k} = \min\{k : \ g(k) \leq 1, k \geq X_{(n)}\}$ **?**
$\hat{k} = \min\{k : \ g(k) \leq 1, k \geq X_{(n)}\} - 1$ **?**

Given a data set, we can solve it easily:

Solve $y = g(x)$ and $y = 1$, $x \in \{X_{(n)}, X_{(n)} + 1, ...\}$;

or solve $y = \prod_{i=1}^n (1 - X_i/k)$ and $y = q^n$, as $g(k) = (\prod_{i=1}^n \frac{1}{1-X_i/k})q^n$.

(I) draw graph $y = \prod_{i=1}^n (1 - X_i/k)$ and $y = q^n$, $x \in \{X_{(n)}, X_{(n)} + 1, ...\}$.
(II) find their solution $\hat{x}$ and $\hat{k} = \max\{k : k \leq \hat{x}\}$



The R program:

```
p=0.6
n=6
x=rbinom(n,20,p) # simulation to get data x
m=max(x)
if (m==0)
    h=1
if (m>0) {
    j=4*m
    k=m:j
    g=rep(0,(j-m+1))
    q=(1-p)**n # q**n
    for(i in m:j)
            g[i-m+1]=q/prod(1-(x/i)) #g
    h=min(k[g<=1])-1
# or use
    H=max(k[g>=1])
    }
```

```
h
H
```

I ran the program 3 times and got 16, 22, 19. **Why 3 values ? True k ?**

The revised R program:

```
p=0.6
n=6
x=rbinom(n,20,p)
m=max(x)
if (m==0)
    h=1
if (m>0) {
    j=4*m
    k=m:j
    g=rep(0,(j-m+1))
    for(i in m:j)
            g[i-m+1]=prod(1-(x/i)) #1/g
    q=(1-p)**n # q**n
    plot(k,g,type="l") # not necessary
    lines(c(m,j),c(q,q)) # not necessary
    h=min(k[g>=q])-1
# or use
    H=max(k[g<=q])
    }
h
H
```

Theorem 1. (Invariance property of the MLE). If $\hat{\theta}$ is the MLE of θ and $\tau = g(\theta)$ is a function of θ, then the MLE of τ is $\hat{\tau} = g(\hat{\theta})$.

Example 3. Let X_1, ..., X_n be a random sample from $N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$, $\theta \in \Theta = [0, \infty) \times (0, \infty)$. Find the MLE of μ, σ, σ^2 and $E(X^2)$.

Sol. Let $\tau = \sigma^2$ and $\gamma = E(X^2)$. MLE of μ, τ, σ and γ ? (σ, γ) are functions of θ. First get the MLE $\hat{\theta}$, then $(\hat{\sigma}, \hat{\gamma})$ can be obained by the **invariance property of the MLE**.

$$\mathcal{L} = \prod_{i=1}^{n} f(X_i; \theta) = (2\pi\sigma^2)^{-n/2} \exp(-\frac{1}{2\sigma^2} \sum_i (X_i - \mu)^2),$$

$$\mathcal{L} = (2\pi\tau)^{-n/2} \exp(-\frac{1}{2\tau} \sum_i (X_i - \mu)^2), \qquad \textbf{why ??}$$

$$\ln\mathcal{L} = c - \frac{n}{2}\ln\tau - \frac{1}{2\tau} \sum_i (X_i - \mu)^2, \tag{1}$$

$$\frac{\partial \ln\mathcal{L}}{\partial \mu} = 2 \times \frac{1}{2\tau} \sum_i (X_i - \mu) = 0 \Rightarrow \mu = \overline{X},$$

$$\frac{\partial \ln\mathcal{L}}{\partial \tau} = -\frac{n}{2}\frac{1}{\tau} + \frac{1}{2\tau^2} \sum_i (X_i - \mu)^2 = 0 \Rightarrow \tau = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 \quad \textbf{Done ?} \tag{2}$$

Check: Two ways: (A) one-by-one, (B) Two dimensions.

(A). Fix τ, maximize $\ln\mathcal{L}(\mu, \tau)$ w.r.t. μ, say $\mu = g(\tau)$.
    Then maximize $\ln\mathcal{L}(g(\tau), \tau)$ w.r.t. τ.

The MLE is $(g(\hat{\tau}), \hat{\tau})$.

Now $\ln\mathcal{L}$ is maximized by $\hat{\mu} = 0 \vee \overline{X}$, (see Example 1 in §72.2.) regardless τ. That is $g(\tau) = \hat{\mu}$.

Replacing μ by $\hat{\mu}$ in $\mathcal{L}(\mu, \tau)$ and Eq. (2),

the critial points for τ:	0	$\hat{\tau}$	∞
$\log \mathcal{L}(\hat{\mu}, \tau)$	$-\infty$	$finite$	$-\infty$
$(see\ Eq.(1))$		MLE	

Thus the MLE of (μ,τ,σ,γ) is $\hat\mu = 0\vee\overline{X}$, $\hat\tau = \frac{1}{n}\sum_{i=1}^n (X_i-\hat\mu)^2$, $\hat\sigma=\sqrt{\hat\tau}$ as $\tau=\sigma^2$;
$\hat\gamma = (\hat\mu)^2+\hat\sigma^2$ as $E(X^2)=\mu^2+\sigma^2$.

$$\begin{aligned}\hat\gamma &= (\hat\mu)^2+\hat\sigma^2\\ &= (\hat\mu)^2+\frac{1}{n}\sum_i (X_i-\hat\mu)^2\\ &= (\hat\mu)^2+\frac{1}{n}\sum_i (X_i^2-2X_i\hat\mu+(\hat\mu)^2)\\ &= (\hat\mu)^2+\overline{X^2}-2\overline{X}\cdot\hat\mu+(\hat\mu)^2\\ &= \overline{X^2}\ \textbf{why ?}\end{aligned}$$

(B). (1) Critical points of $\mathcal{L}(\mu,\tau)$:
for μ: $\overline{X}$ (if $\overline{X}>0$), 0, ∞.
for τ: $\frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2$, 0, ∞,

(2.a) Compare $\mathcal{L}(\mu,\tau)$ over critical points if $\overline{X}>0$. $(\mu,\tau)\in[0,\infty)\times(0,\infty)$.
$\frac{\partial \ln\mathcal{L}}{\partial\theta}=\vec{0}$: $\mu=\overline{X}$ and $\tau=\frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2=\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})^2$.
4 boundary lines and a point.
$\mu=0$, $\mu=\infty$, $\tau=0$, $\tau=\infty$ and $(\overline{X},\overline{X^2}-(\overline{X})^2)$.
$\mu=0$ reduces to $(0,0)$, $(0,\infty)$ and $(0,\overline{X^2}-(0)^2)$, or **only the latter ??**

(μ,τ)	$\underbrace{\mu=0}_{(0,\overline{X^2}-(0)^2)}$	$(\overline{X},\overline{X^2}-(\overline{X})^2)$	$\mu=\infty$	$\tau=0$	$\tau=\infty$
$\ln\mathcal{L}=c-\frac{n}{2}\ln\tau-\frac{1}{2\tau}\sum_i (X_i-\mu)^2$	$finite$	$finite$	$-\infty$	$-\infty$	$-\infty$
$\frac{\partial\ln\mathcal{L}}{\partial\mu}(\mu,\overline{X^2}-(\overline{X})^2)$	$+$	0	$-$		
		MLE			

(2.b) compare $\mathcal{L}(\mu,\tau)$ over critical points (μ,τ) if $\overline{X}\le 0$(only 4 boundary lines):

(μ,τ)	$\mu=\infty$	$\tau=0$	$\tau=\infty$	$(0,\overline{X^2})$	$<=\mu=0$
$\ln\mathcal{L}$	$-\infty$	$-\infty$	$-\infty$	$finite$	
				MLE	

§7.2.3. Bayes estimator.

We have learned two estimators: MME and MLE under the assumption that X_1, ..., X_n are i.i.d. from $f(x;\theta)$, $\theta\in\Theta$.
θ is a constant (not random), unknown.
In this section, we consider Bayesian approach:
Conditional on θ, X_1, ..., X_n are i.i.d. from $f(x|\theta)$,
θ is a random variable with df $\pi(\theta)$,
$f(x|\theta)$ is a conditional df of $X|\theta$.
Bayes estimator of θ is $\hat\theta=E(\theta|\mathbf{X})$.
Recall the formula

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}. \tag{1}$$

Now
$f(\mathbf{x},\theta)$ is the joint df of $(\mathbf{X},\theta)$,
$f_{\mathbf{X}}(\mathbf{x})$ is the marginal df of $\mathbf{X}$,
$\pi(\theta)$ is the marginal df of θ, called **prior df** now,
$f(\mathbf{x}|\theta)$ is the conditional df of $\mathbf{X}|\theta$,
$\pi(\theta|\mathbf{x})$ is the conditional df of $\theta|\mathbf{X}$, called the **posterior df** now,
$f_{\mathbf{X}}(\mathbf{x})=\begin{cases}\int f(\mathbf{x},\theta)d\theta & \text{if } \theta \text{ is continuous}\\ \sum_\theta f(\mathbf{x},\theta) & \text{if } \theta \text{ is discrete.}\end{cases}$

$\pi(\theta) = \begin{cases} \int f(\mathbf{x},\theta)d\mathbf{x} & \text{if } X \text{ is continuous} \\ \sum_{\mathbf{x}} f(\mathbf{x},\theta) & \text{if } X \text{ is discrete.} \end{cases}$

$f(\mathbf{x}|\theta) = \frac{f(\mathbf{X},\theta)}{\pi(\theta)}$ by Eq. (1),

$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{X},\theta)}{f_{\mathbf{X}}(\mathbf{X})}$ by Eq. (1),

$E(\theta|\mathbf{X}=\mathbf{x}) = \int_\theta \theta \frac{dF(\mathbf{X},\theta)}{f_{\mathbf{X}}(\mathbf{X})} = \begin{cases} \int \theta\pi(\theta|\mathbf{x})d\theta & \text{if } \theta \text{ is continuous} \\ \sum \theta\pi(\theta|\mathbf{x}) & \text{if } \theta \text{ is discrete.} \end{cases}$

Recall the Bayes set-up: **conditional on** θ, X_1, ..., X_n are i.i.d. from $f(x|\theta)$,

Are X_i's i.i.d. ?

Homework. Answer it through the assumption as follows. Let X_1, ..., X_n be i.i.d. $\sim bin(1,p)$, and $p \sim U(0,1)$.
Ans: No !

Remark. Two ways to compute the Bayes estimator:

1. $E(\theta|\mathbf{X})$,
2. $E(\theta|T(\mathbf{X}))$ where T is a MSS.

They lead to the same estimator.
The second method is often simpler in derivation.

Example 1. Let X_1, ..., X_n be a random sample from $bin(k,\theta)$, $\theta \sim beta(\alpha,\beta)$ with $\pi(t) = \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha,\beta)}$, $t \in [0,1]$, where $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, $\alpha,\beta > 0$ and (k,α,β) is known. Bayes estimator of θ ?

Sol. Recall $T(\mathbf{X}) = \sum_{i=1}^n X_i$ is MSS if θ is a parameter.
(1) $E(\theta|\mathbf{X})$= ? (2) $E(\theta|T(\mathbf{X}))$= ?

Method 1. Based on $\mathbf{X}$.
$f(\mathbf{x}|\theta) = \prod_{i=1}^n \binom{k}{x_i}\theta^{x_i}(1-\theta)^{k-x_i} = (\prod_{i=1}^n \binom{k}{x_i})\theta^{\sum_i x_i}(1-\theta)^{nk-\sum_i x_i}$.

$$\begin{aligned} \pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x},\theta)}{f_{\mathbf{X}}(\mathbf{x})} \\ &= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{f_{\mathbf{X}}(\mathbf{x})} \\ &\propto \theta^{\sum_i x_i}(1-\theta)^{nk-\sum_i x_i}\theta^{\alpha-1}(1-\theta)^{\beta-1} \textbf{(main trick!!)} \\ &= \theta^{\sum_i x_i+\alpha-1}(1-\theta)^{kn-\sum_i x_i+\beta-1} \end{aligned} \tag{1}$$

Thus $\theta|(\mathbf{X}=\mathbf{x}) \sim beta(\sum_i x_i + \alpha, nk - \sum_i x_i + \beta)$ $(= beta(a,b))$,

The Bayes estimator is

$$\begin{aligned} \hat\theta = E(\theta|\mathbf{X}) = \frac{a}{a+b} &= \frac{\sum_i X_i + \alpha}{nk+\alpha+\beta} \\ &= \frac{nk}{nk+\alpha+\beta}\frac{\sum_i X_i}{nk} + \frac{\alpha+\beta}{nk+\alpha+\beta}\frac{\alpha}{\alpha+\beta} \\ &= r\frac{\sum_{i=1}^n X_i}{nk} + (1-r)\frac{\alpha}{\alpha+\beta} \qquad \approx \begin{cases} MLE & \text{if } r \approx 1 \\ E(\theta) & \text{if } r \approx 0 \end{cases} \end{aligned}$$

a weighted average of the MLE $\frac{\sum_{i=1}^n X_i}{nk}$ and the prior mean $\frac{\alpha}{\alpha+\beta}$.

Method 2. Based on MSS $T = \sum_i X_i$. $T|\theta \sim bin(nk,\theta)$? or $T \sim bin(nk,\theta)$? $f_{T|\theta}(t|\theta) = \binom{nk}{t}\theta^t(1-\theta)^{nk-t}$,

$$\begin{aligned} \pi(\theta|t) &= \frac{\binom{nk}{t}\theta^t(1-\theta)^{nk-t}\theta^{\alpha-1}(1-\theta)^{\beta-1}/B(\alpha,\beta)}{f_T(t)} \\ &\propto \theta^{t+\alpha-1}(1-\theta)^{kn-t+\beta-1} \qquad \textbf{same as (1), why ?} \end{aligned}$$

...

Example 2. Suppose that X_1, ..., X_n is a random sample from $N(\theta,\sigma^2)$, $\theta \sim N(\mu,\tau^2)$, where (σ,μ,τ) is known. Bayes estimator of θ ?

Sol. $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} exp(-(\frac{x-\mu}{\sigma})^2/2) \propto e^{-ax^2+bx}$ (kernel of f).
Two ways: (1) $E(\theta|\mathbf{X})$ and (2) $E(\theta|T(\mathbf{X}))$. **Which to choose ?**
MSS of θ is $T = \overline{X}$. $T|\theta \sim N(\theta, \sigma^2/n)$.

$$E(\theta|T(\mathbf{X}) = t) = \int \theta \underbrace{\pi(\theta|t)} d\theta.$$

$$\begin{aligned}
\pi(\theta|t) &= \frac{f(t|\theta)\pi(\theta)}{f_T(t)} = \mathbf{??} \\
&\propto f(t|\theta)\pi(\theta) && \textbf{(main trick)} \\
&\propto \exp\big(-\frac{1}{2}\frac{(t-\theta)^2}{\sigma^2/n} - \frac{1}{2}\frac{(\theta-\mu)^2}{\tau^2}\big) \\
&\propto \exp\big(-\frac{1}{2}\frac{-2t\theta+\theta^2}{\sigma^2/n} - \frac{1}{2}\frac{\theta^2-2\theta\mu}{\tau^2}\big) \\
&= \exp\big(-\frac{1}{2}\frac{\theta^2}{\sigma^2/n} - \frac{1}{2}\frac{\theta^2}{\tau^2} + \frac{1}{2}\frac{2t\theta}{\sigma^2/n} + \frac{1}{2}\frac{2\theta\mu}{\tau^2}\big) && = e^{-a\theta^2+b\theta} \\
&= \exp\big(-\frac{1}{2}\{\theta^2[\frac{1}{\sigma^2/n} + \frac{1}{\tau^2}] + (-2\theta)[\frac{t}{\sigma^2/n} + \frac{\mu}{\tau^2}]\}\big) \propto e^{-(\frac{\theta-\mu_*}{\sigma_*})^2/2}
\end{aligned}$$

$$\exp(-\frac{1}{2}\frac{(\theta-\mu_*)^2}{\sigma_*^2}) = \exp(-\frac{1}{2}[\theta^2\frac{1}{\sigma_*^2} - 2\theta\frac{\mu_*}{\sigma_*^2} + \frac{\mu_*^2}{\sigma_*^2}])$$

$$\frac{1}{\sigma_*^2} = [\frac{1}{\sigma^2/n} + \frac{1}{\tau^2}] \text{ and } \frac{\mu_*}{\sigma_*^2} = [\frac{t}{\sigma^2/n} + \frac{\mu}{\tau^2}]$$

$$\sigma_*^2 = \frac{1}{[\frac{1}{\sigma^2/n} + \frac{1}{\tau^2}]} \text{ and } \mu_* = \frac{[\frac{t}{\sigma^2/n} + \frac{\mu}{\tau^2}]}{[\frac{1}{\sigma^2/n} + \frac{1}{\tau^2}]}$$

Thus $\theta|(T = t) \sim N(\mu_*, \sigma_*^2)$ and the Bayes estimator

$$\hat{\theta} = E(\theta|T) = \mu_* = \frac{\frac{\overline{X}}{\sigma^2/n} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma^2/n} + \frac{1}{\tau^2}}.$$

Remark. It is interesting to notice the following fact again.

In Example 2, the Bayes estimator is

$$\begin{aligned}
\hat{\theta} &= \frac{\frac{\overline{X}}{\sigma^2/n} + \frac{\mu}{\tau^2}}{\frac{1}{\sigma^2/n} + \frac{1}{\tau^2}} \\
&= \frac{\frac{1}{\sigma^2/n}}{\frac{1}{\sigma^2/n} + \frac{1}{\tau^2}}\overline{X} + \frac{\frac{1}{\tau^2}}{\frac{1}{\sigma^2/n} + \frac{1}{\tau^2}}\mu \\
&= r\overline{X} + (1-r)\mu \\
&\approx \begin{cases} \overline{X} & \text{if } n \text{ is large or } r \approx 1 \\ E(\theta) & \text{if } r \approx 0 \end{cases}
\end{aligned}$$

a weighted average of the MLE $\overline{X}$ and the prior mean μ.

§7.3. Methods of evaluating estimators.

Notice that the MME, MLE and Bayes estimators may not be the same.

Question: How to compare estimators ?

$\hat{\theta} - \theta$ — error, **Not good, Why ?**

$|\hat{\theta} - \theta|$ — absolute error, **Not good, Why ?**

$E(\hat{\theta}) - \theta$ — bias, denoted by bias($\hat{\theta}$) or $B(\hat{\theta})$;

$E(|\hat{\theta} - \theta|)$ — mean absolute error, **Not ideal, Why ?**

$E((\hat{\theta}-\theta)^2)$ — mean-squared error of $\hat{\theta}$;

A naive approaches:

Select $\hat{\theta}$ that has smaller MSE($\hat{\theta}$).

Formula:

$$E(\hat{\theta}-\theta)^2 = Var(\hat{\theta}) + (bias(\hat{\theta}))^2$$

$$\begin{aligned}\text{Reason: } E((\hat{\theta}-\theta)^2) =& E[(\hat{\theta}-E(\hat{\theta})+E(\hat{\theta})-\theta)^2] \\ =& E((\hat{\theta}-E(\hat{\theta}))^2) + E((E(\hat{\theta})-\theta)^2) + 2E[(\hat{\theta}-E(\hat{\theta}))(E(\hat{\theta})-\theta)] \\ =& E((\hat{\theta}-E(\hat{\theta}))^2) + (E(\hat{\theta})-\theta)^2 + 2(E(\hat{\theta})-E(\hat{\theta}))(E(\hat{\theta})-\theta) \\ =& Var(\hat{\theta}) + (bias(\hat{\theta}))^2\end{aligned}$$

Definition. If $bias(\hat{\theta}) = 0$, $\hat{\theta}$ is called an unbiased estimator of θ.

Example 1. Suppose X_1, ..., X_n are i.i.d. with mean μ and variance σ^2. A common estimator of μ is $\hat{\mu} = \overline{X}$, and two common estimators of σ^2 are $\hat{\sigma}^2 = \frac{1}{n}\sum_i(X_i-\overline{X})^2$ and $S^2 = \frac{1}{n-1}\sum_i(X_i-\overline{X})^2$. (a) Are they unbiased ? (b) Compute the MSE of $\overline{X}$, S^2 and $\hat{\sigma}^2$ under $N(\mu,\sigma^2)$; (c) Compare $\hat{\sigma}^2$ to S^2 under $N(\mu,\sigma^2)$.

Sol. (a) Recall: $E(\overline{X}) = \mu_X$, unbiased estimator of μ_X. $Var(X) = E((X-\mu)^2) = E(X^2)-\mu^2$, $\hat{\sigma}^2 = \frac{1}{n}\sum_i(X_i-\overline{X})^2 = \overline{X^2}-(\overline{X})^2$ and $S^2 = \frac{n}{n-1}\hat{\sigma}^2$.

$$\begin{aligned}E(\hat{\sigma}^2) =& E(\overline{X^2}) - E((\overline{X})^2) \\ =& E(X^2) - ((E(\overline{X}))^2 + \sigma^2_{\overline{X}}) && \text{(Why ?)} \\ =& \sigma^2 - \sigma^2/n && \textbf{Why ?} \\ E(S^2) =& \frac{n}{n-1}E(\hat{\sigma}^2) = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2.\end{aligned}$$

Thus S^2 is unbiased but not $\hat{\sigma}^2$.

(b) $MSE(\hat{\mu}) = Var(\overline{X}) + (bias(\hat{\mu}))^2 = \sigma^2/n + 0$. $MSE(S^2) = Var(S^2) + (bias(S^2))^2 = Var(S^2)$ $MSE(\hat{\sigma}^2) = (\frac{n-1}{n})^2 Var(S^2) + (\sigma^2/n)^2$

(1) Recall a theorem: *Under i.i.d. normal assumption,*

1. $\overline{X} \sim N(\mu,\sigma^2/n)$*;*

2. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$*, that is,* $S^2 \sim \frac{\sigma^2}{n-1}\chi^2(n-1)$*;*

3. $\overline{X} \perp S^2$.

(2) Moreover, recall $E(\chi^2(m)) = m$ and $Var(\chi^2(m)) = 2m$.

$MSE(S^2) = (\frac{\sigma^2}{n-1})^2 \times 2(n-1) = 2(\frac{\sigma^4}{n-1})$.

$MSE(\hat{\sigma}^2) = (\frac{n-1}{n})^2 Var(S^2) + (\sigma^2/n)^2 = 2\frac{n-1/2}{n^2}\sigma^4$.

(c) $MSE(\hat{\sigma}^2)/MSE(S^2) = (\frac{n-1}{n^2})(n-\frac{1}{2}) < 1$.

Thus $\hat{\sigma}^2$ is better in terms of the MSE, (though S^2 is better than $\hat{\sigma}^2$) in terms of unbiasedness.

Question. Is $\hat{\sigma}^2$ is the best in terms of the MSE ?

$MSE(\hat{\sigma}^2) = 2\frac{n-1/2}{n^2}\sigma^4$.

Let $\tilde{\sigma}^2 = 1$, then $MSE(\tilde{\sigma}^2) = (1-\sigma^2)^2$.

$$\begin{cases} MSE(\hat{\sigma}^2) > 0 = MSE(\tilde{\sigma}^2) & \text{if } \sigma = 1 \\ MSE(\hat{\sigma}^2) = 2\frac{2^8-0.5}{2^{16}}2^2 = \frac{2^8-0.5}{2^{13}} < 1 = MSE(\tilde{\sigma}^2) & \text{if } \sigma^2 = 2 \text{ and } n = 2^8 \end{cases}$$

Question: How to compare estimators ?

1. Select $\hat{\theta}$ with smaller $MSE(\hat{\theta})$,
2. Select $\hat{\theta}$ with the smallest $MSE(\hat{\theta})$ $(= E((\hat{\theta}-\theta)^2))$ **(impossible) !**.
3. Select $\hat{\theta}$ with smaller bias.
4. Select unbiased $\hat{\theta}$ with the smallest $Var(\hat{\theta})$.

Definition. An estimator $\hat{\tau}$ is called the best unbiased estimator or uniformly minimum variance unbiased estimator (UMVUE) of $\tau(\theta)$ if

(a) $E(\hat{\tau}) = \tau(\theta)\ \forall\ \theta \in \Theta$;

(b) $Var(\hat{\tau}) \leq Var(\tilde{\tau})\ \forall\ \theta \in \Theta$ and $\forall$ unbiased $\tilde{\tau}$.

In many situations, the UMVUE exists.

Question: How can we determine that $\hat{\tau}$ is UMVUE ?

To answer the question, we need several theorems.

Theorem 1 (Cramér-Rao Inequality (CR- Ineq.)) Let X_1, ..., X_n be i.i.d. from $X \sim f(x;\theta)$ and let $W(\mathbf{X})$ be a statistic. Suppose that

(1) $\frac{d}{d\theta}E(W) = \begin{cases} \int \frac{\partial}{\partial\theta} W(\mathbf{x}) f(\mathbf{x};\theta) d\mathbf{x} & \text{if } \mathbf{X} \text{ is continuous} \\ \sum_{\mathbf{x}} \frac{\partial}{\partial\theta} W(\mathbf{x}) f(\mathbf{x};\theta) & \text{if } \mathbf{X} \text{ is discrete;} \end{cases}$

(2) $Var(W) < \infty$.

Let $\tau = E(W)$. Then

$Var(W) \geq \frac{(\frac{d}{d\theta}E(W))^2}{E((\frac{\partial}{\partial\theta}\ln f(\mathbf{X};\theta))^2)}$ $(= \frac{(\frac{d}{d\theta}E(W))^2}{nE((\frac{\partial}{\partial\theta}\ln f(X;\theta))^2)})$. **Why = ?**

The latter is called the Cramér-Rao Lower Bound (CRLB) of $\hat{\tau}(\theta)$.

Remark. A CR-ineq gives a tool for determining an UMVUE. If

(1) the assumptions in CR-inequality hold,

(2) $E(W) = \tau(\theta)$ and

(3) $Var(W) = \frac{(\frac{d}{d\theta}\tau(\theta))^2}{E(\frac{\partial}{\partial\theta}\ln f(\mathbf{X};\theta))^2}$,

then W is an UMVUE of $\tau(\theta)$.

Results: The assumptions in CR-inequality

1. **hold** if $f(x;\theta)$ belongs to an exponential family;
2. **often fail** if the domain of the f depends on θ such as $U(0,\theta)$.

Example 1. Let X_1, ..., X_n be i.i.d. from $N(\mu, 4)$, UMVUE of μ ?

Sol. $N(\mu, \sigma^2)$ belongs to the exponential family.

Thus Condition (1) in CR-inequality holds. $\theta =$ **??** $\tau(\theta)=$ **??**

Candidate of an UMVUE of μ: $W = \overline{X}$;

$E(W) = \mu = \theta$ (Condition (2) in Remark);

$Var(W) = Var(\overline{X}) = \sigma^2/n < \infty$ (Condition (2) in CR-In.);

$CRLB = \frac{(\frac{d}{d\theta}\tau(\theta))^2}{nE((\frac{\partial}{\partial\theta}\ln f(X;\theta))^2)}$;

$(\frac{d}{d\theta}\tau(\theta))^2 = 1$;

$\frac{\partial}{\partial\theta}\ln f(X;\theta) = \frac{\partial}{\partial\theta}[\ln c - \frac{1}{2}(X-\theta)^2/\sigma^2] = \frac{X-\theta}{\sigma^2}$,

$E((\frac{\partial}{\partial\theta}\ln f(X;\theta))^2) = E((\frac{X-\theta}{\sigma^2})^2) = \frac{\sigma^2}{\sigma^4} = 1/\sigma^2$.

$CRLB = \frac{(\frac{d}{d\theta}E(W))^2}{nE(\frac{\partial}{\partial\theta}\ln f(X;\theta))^2} = \frac{1}{n \times \frac{1}{\sigma^2}} = \sigma^2/n = Var(\overline{X})$

Thus $\overline{X}$ is an UMVUE of μ.

One of 6.8 or 6.9 will be in the midterm.

Definition. An estimator $\hat{\tau}$ is the best unbiased estimator or UMVUE of $\tau(\theta)$ if

(a) $E(\hat{\tau}) = \tau(\theta)\ \forall\ \theta \in \Theta$;

(b) $Var(\hat{\tau}) \leq Var(\tilde{\tau})\ \forall\ \theta \in \Theta$ and $\forall$ unbiased $\tilde{\tau}$.

The Cramér-Rao Lower Bound (CRLB) gives a tool for determining an UMVUE.

$\hat{\theta}$ is an UMVUE of $\tau(\theta)$ if

(1) $\frac{d}{d\theta}E(\hat{\theta}) = \begin{cases} \int \hat{\theta}(\mathbf{x}) \frac{\partial}{\partial\theta} f(\mathbf{x};\theta) d\mathbf{x} & \text{if } \mathbf{X} \text{ is continuous} \\ \sum_{\mathbf{x}} \hat{\theta}(\mathbf{x}) \frac{\partial}{\partial\theta} f(\mathbf{x};\theta) & \text{if } \mathbf{X} \text{ is discrete;} \end{cases}$

(2) $E(\hat{\theta}) = \tau(\theta)$,

(3) $Var(\hat{\theta}) = \frac{(\frac{d}{d\theta}\tau(\theta))^2}{E(\frac{\partial}{\partial\theta}\ln f(\mathbf{x};\theta))^2}$.

Results:

Assumptions in CR Th $\begin{cases} \textbf{hold} & \text{if } f(x;\theta) \text{ belongs to an exponential family;} \\ \textbf{often fail} & \text{if the domain of the } f \text{ depends on } \theta \text{ } e.g. \text{ } U(0,\theta). \end{cases}$

Example 2. Let X_1, ..., X_n be i.i.d. from $X \sim f(x;\theta) = \frac{1}{\theta}\mathbf{1}(x \in (0,\theta))$.

a. MLE $\hat{\theta}$ of θ ?

b. Find an unbiased estimator of θ based on $\hat{\theta}$.

c. Show that the CR-inequality fails.
d. Why does it fail ?

Sol. 1. Solve for MLE:

$$\mathcal{L}(\theta)=\prod_{i=1}^{n} f(X_i;\theta)=\prod_i \frac{\mathbf{1}(X_i\in(0,\theta))}{\theta}=\theta^{-n}\mathbf{1}(X_{(n)}\in(0,\theta))$$

Typical way: $\frac{\partial}{\partial\theta}\ln\mathcal{L}(\theta)=-n/\theta=0$ **???** if $X_{(n)}<\theta$.

Notice that $\frac{\partial}{\partial\theta}\ln\mathcal{L}(\theta)\begin{cases}<0 & \text{if } \theta>X_{(n)}\\ ? & \text{if } \theta=X_{(n)}\\ =0 & \text{if } \theta<X_{(n)}\end{cases}$,

Check: $\begin{pmatrix}\textbf{Critical points:} & 0 & X_{(n)} & \infty\\ \mathcal{L}(\theta): & 0 & 0 & 0\\ \mathcal{L}(\theta) & \to & - & \searrow\end{pmatrix}$ thus the **maximum value does not exist, based on the likelihood !**

However, the density function of $U(0,\theta)$ is unique in the sense that

$E(|f(X;\theta)-f_2(X;\theta)|)=0$ if f and f_2 are two density functions of $U(0,\theta)$.

Here $f(x;\theta)=\mathbf{1}(x\in(0,\theta))\frac{1}{\theta}$ *and* $f_2(x;\theta)=\mathbf{1}(x\in[0,\theta])\frac{1}{\theta}$.

The latter leads to the likelihood

$\mathcal{L}_2(\theta)=\theta^{-n}\mathbf{1}(X_{(n)}\in[0,\theta])\mathbf{1}(X_{(1)}\geq 0)$.

Then **the maximum value does exist !!**

The MLE is $\hat\theta=X_{(n)}$.

2. To find an unbiased estimator, consider $E(\tilde\theta)=E(c\hat\theta)=\theta$.

$E(\hat\theta)=\int tf_{X_{(n)}}(t)dt$

$=\int_0^\theta tn(\frac{t}{\theta})^{n-1}\frac{1}{\theta}dt$ (as $f_{X_{(n)}}(t)=\frac{n!}{(n-1)!1!}(F(t))^{n-1}f(t)$)

$=\int_0^\theta n\frac{t^n}{\theta^n}dt$

$=\frac{n}{n+1}\theta.$

An unbiased estimator related to the MLE is $\tilde\theta=\frac{n+1}{n}\hat\theta$.

3. To show the CR-inequality fails, one needs to show $Var(\tilde\theta)<CRLB$.

Now $Var(\tilde\theta)=E((\tilde\theta)^2)-\theta^2$.

$$\begin{aligned}E(\tilde\theta^2)=&\int(\frac{n+1}{n})^2t^2f_{X_{(n)}}(t)dt\quad (as\ E(g(Y))=\int tf_{g(Y)}(t)dt=\int g(x)f_Y(x)dx)\\
=&\int_0^\theta(\frac{n+1}{n})^2t^2n(\frac{t}{\theta})^{n-1}\frac{1}{\theta}dt \qquad as\ f_{X_{(n)}}(t)=n(F(t))^{n-1}f(t)\\
=&n(\frac{n+1}{n})^2\frac{1}{\theta^n}\int_0^\theta t^{n+1}dt\\
=&n(\frac{n+1}{n})^2\frac{1}{n+2}\theta^2\\
=&\frac{(n+1)^2}{n(n+2)}\theta^2.\end{aligned}$$

(1) $Var(\tilde\theta)=\frac{(n+1)^2}{n(n+2)}\theta^2-\theta^2=\frac{1}{n(n+2)}\theta^2$.

$$CRLB=\frac{(\frac{d}{d\theta}\theta)^2}{nE(\frac{\partial}{\partial\theta}\ln f(X;\theta))^2}.$$

$\ln f(x;\theta)=-\ln\theta$, $x\in(0,\theta)$.

$(\ln f(x;\theta))'=\frac{-1}{\theta}$, $x\in(0,\theta)$.

$E((\frac{\partial \ln f(X;\theta)}{\partial\theta})^2) = \begin{cases} E(\frac{\mathbf{1}(X\in(0,\theta))}{\theta^2}) & ? \\ E(\frac{1}{\theta^2}) & ? \end{cases}$ **Which is correct ??**
$= \frac{P(X\in(0,\theta))}{\theta^2} = \frac{1}{\theta^2}$ **by accident!** *e.g.*,

$$X \sim U(0,\theta) \;\; => \;\; E(\frac{1(X\in(\frac{\theta}{2},\theta))}{\theta^2}) = E(\frac{1}{\theta^2}) \;\; \mathbf{??}$$

$CRLB = \frac{1}{n\frac{1}{\theta^2}} = \frac{\theta^2}{n} > \frac{\theta^2}{n(n+2)} = Var(\tilde\theta)$

Thus the CR-inequality fails. **In fact, we shall show $\tilde\theta$ is UMVUE of θ.**
d. **Reason that the CRLB fails**: (condition (1) in theorem fails).
$\frac{\partial}{\partial\theta}E(W) \neq \int \frac{\partial}{\partial\theta}W(\mathbf{x})f(\mathbf{x};\theta)d\mathbf{x}$ where $W = \tilde\theta = X_{(n)}\frac{n+1}{n}$,

$$\frac{\partial}{\partial\theta}E(W) = \int \frac{\partial}{\partial\theta}W(\mathbf{x})f(\mathbf{x};\theta)d\mathbf{x}$$
$$(= \int \frac{\partial}{\partial\theta} y f_W(y;\theta)dy\mathbf{??})$$

$E(W) = \theta$,
LHS$=\frac{\partial}{\partial\theta}E(W) = 1$. But RHS$= -n$, as

$$\begin{aligned}
RHS =& \int \frac{\partial}{\partial\theta}\frac{n+1}{n} w f_{X_{(n)}}(w)dw \\
=& \int_0^\theta \frac{\partial}{\partial\theta}\frac{n+1}{n} wn(\frac{w}{\theta})^{n-1}\frac{1}{\theta}dw \\
=& (n+1)\int_0^\theta \frac{\partial}{\partial\theta}(\frac{w}{\theta})^n dw \\
=& (n+1)\int_0^\theta w^n(-n)\theta^{-n-1}dw \\
=& \theta^{n+1}(-n)\theta^{-n-1} \\
=& (-n) \\
or\ RHS =& \underbrace{\int\cdots\int}_{how\ many?} \frac{\partial}{\partial\theta}W(\mathbf{x})f(\mathbf{x};\theta)d\mathbf{x} \\
=& \int\cdots\int \frac{\partial}{\partial\theta}\frac{n+1}{n}x_{(n)}\theta^{-n}\mathbf{1}(x_{(1)},x_{(n)}\in(0,\theta))d\mathbf{x} \\
=& n!\int_0^\theta \underbrace{\int_0^{x_n}\cdots\int_0^{x_2} x_n dx_1\cdots dx_{n-1}}\, dx_n\{\frac{\partial}{\partial\theta}\theta^{-n}\frac{n+1}{n}\} \;\; \mathbf{Why}?? \\
& \textbf{(by induction on } n\textbf{)} \\
=& n!\int_0^\theta \overbrace{x_n\frac{(x_n)^{n-1}}{(n-1)!}}\, dx_n\{(-n)\theta^{-n-1}\frac{n+1}{n}\} \\
=& n!\int_0^\theta \frac{(x_n)^n}{(n-1)!}dx_n(-n)\theta^{-n-1}\frac{n+1}{n} \\
=& n!\theta^{n+1}\frac{1}{(n+1)(n-1)!}(-n)\theta^{-n-1}\frac{n+1}{n} = -n
\end{aligned}$$

$\frac{\partial}{\partial\theta}E(W) \neq \int \frac{\partial}{\partial\theta}W(\mathbf{x})f(\mathbf{x};\theta)d\mathbf{x}$ (condition (1) in theorem fails).

Theorem 2. *If (1) T is a sufficient and complete statistic for θ; (2) $\phi(T)$ is a statistic that only depends on T, Then $\phi(T)$ is the unique UMVUE of $E(\phi(T))$.*

Corollary. $\tilde{\theta} = \frac{n+1}{n} X_{(n)}$ is UMVUE of θ if X_i's are i.i.d. $\sim U(0,\theta)$. **Why ?**

Remark. 2 more ways for finding a UMVUE of $\tau(\theta)$ based on Theorem 2:

2. Find a sufficient and complete statistic T and
   a $\phi(T)$ that is unbiased of $\tau(\theta)$,
   then $\phi(T)$ is the UMVUE of $\tau(\theta)$.

3. Find a sufficient and complete statistic T and
   an unbiased estimator W of $\tau(\theta)$,
   then $\hat{\tau} = E(W|T)$ is the UMVUE of $\tau(\theta)$.

Example 1. Let X_i's be i.i.d. from $N(\mu, \sigma^2)$. UMVUE of μ^2 and σ^2 ?

Sol. Use Method 2.
$T = (\overline{X}, \overline{X^2})$ is sufficient and complete (known due to exponential family).
A function $\phi(T)$ such that $E(\phi(T)) = \theta$?
$E(S^2) = \sigma^2$ and $S^2 = \frac{n}{n-1}(\overline{X^2} - (\overline{X})^2)$, a function of T.
$E(\overline{X^2}) = E(X^2) = \mu^2 + \sigma^2$ and $\overline{X^2}$ is a function of T;
$E(\overline{X^2} - S^2) = \mu^2 + \sigma^2 - \sigma^2 = \mu^2$ and $\overline{X^2} - S^2$ is a function of T;
Thus $\overline{X^2} - S^2$ and S^2 are the UMVUEs of μ^2 and σ^2, respectively.

Example 2. Let X_1, ..., X_n be i.i.d. from Poisson(λ). UMVUE of λ ?

Sol. Recall $E(X_1) = \lambda = Var(X_1)$ for Poisson(λ). $T = \sum_{i=1}^n X_i$ is sufficient and complete.
Two unbiased estimators: $\hat{\lambda} = \overline{X}$, $\check{\lambda} = S^2$.
Method 1. Check: Cramer-Rao Lower Bound $= V(\hat{\lambda})$ or $V(\check{\lambda})$?
Method 2. $\hat{\lambda} = \overline{X}$, as $E(\overline{X}) = \mu = \lambda$.
Method 3. $\tilde{\lambda} = E(W|T)$, where $W = S^2$ or $\overline{X}$.

Question:
(1) Which method is better here ?
(2) $E(\overline{X}|T) = \hat{\lambda}$?
(3) $E(S^2|T) = \hat{\lambda}$?
Consider the case $n = 2$.

Let $T = \sum_i X_i$.

$E(S^2|\overline{X}) = E(S^2|T/n) = E(S^2|T) = 2[E(\overline{X^2}|\overline{X}) - (\overline{X})^2]$.

$f_{X_1|X_1+X_2}(x|t) = P(X_1 = x, X_2 = t-x)/P(T = t) = \binom{t}{x} 0.5^x 0.5^{t-x} \quad (bin(t, 0.5))$.

$E(\overline{X^2}|T) = E(X^2|T) = (Tpq + (Tp)^2) = (T/4 - T^2/4) = (\overline{X}/2) - (\overline{X})^2$.

$E(S^2|\overline{X}) = E(S^2|T) = 2[E(\overline{X^2}|\overline{X}) - (\overline{X})^2] = \overline{X}$.
In general, $n \geq 2$. $X_i | \sum_i X_i \sim bin(t, 1/n)$, $i = 1$, ..., n.

Example 3. Let X_1, ..., X_n be a random sample from $X \sim bin(5, \theta)$.
$\tau = P(X \le 1)$. UMVUE of τ ?
Sol. $\tau = ?$ $(= P(X \le 1))$.

$\tau = (1-\theta)^5 + 5\theta(1-\theta)^4$.

3 methods for UNVUE:

1. Find an unbiased $\hat\tau$, compare $\sigma^2_{\hat\tau}$ to CRLB.
2. Find a complete sufficient T and $g(T)$ so that $E(g(T)) = \tau$.
3. Find a complete sufficient T and an unbiased $\hat\tau$, compute $E(\hat\tau|T)$.

Method 3. $\hat\tau = E(W|T)$.

W=? T=? E(W|T)= ?

$W = \mathbf{1}(X_1 \le 1)$. Then $E(W) = P(X_1 \le 1) = P(X \le 1)$.

Why not $W = \mathbf{1}(X \le 1)$?

$T = \sum_{i=1}^n X_i$ is sufficient and complete (due to the exponential family).

Why T, not $\overline{X}$? Either is fine, but $T \sim bin(5n, \theta)$, $T = n\overline{X}$, $f_{\overline{X}}(y) =$?

$E(W|T) =$ **?**

Ans.: (1) $g(t) = E(W|T = t)$, $t = 0, ..., 5n$. (2) $E(W|T) = g(T)$.

$E(W|T = t) = \int w dF_{W|T}(w|t)$ **meaning ?**

$= 0 \cdot f_{W|T}(0|t) + 1 \cdot f_{W|T}(1|t)$

$E(W|T = t) = f_{W|T}(1|t)$.

$f_{W|T}(1|t) = \frac{P(W=1,T=t)}{P(T=t)}$, $t \in \{0, 1, ..., 5n\}$.

$P(T = t) = ?$

$P(W = 1, T = t) = ?$

If $t = 0$, then $P(W = 1, T = t) = P(X_1 \in \{0, 1\}, T = 0) = \begin{cases} P(X_1 = 0) & ? \\ P(T = 0) & ? \end{cases}$

$\{T = 0\} = \{\sum_{i=1}^n X_i = 0\} = \{X_1 = \cdots = X_n = 0\}$.

$E(W|T = 0) = f_{W|T}(1|0) = \frac{P(T=0)}{P(T=0)}$.

If $t \ge 1$,

$$\begin{aligned}
P(W = 1, T = t) =& P(X_1 \in \{0, 1\}, T = t) \textbf{ how to proceed ?} \\
& (=P(X_1 \in \{0, 1\})P(T = t)?) \\
\\
=& P(X_1 \in \{0, 1\}, \sum_{i=1}^n X_i = t) \\
=& P(X_1 = 0, \sum_{i=1}^n X_i = t) + P(X_1 = 1, \sum_{i=1}^n X_i = t) \\
=& P(X_1 = 0, \sum_{i=2}^n X_i = t) + P(X_1 = 1, \sum_{i=2}^n X_i = t-1) \\
=& P(X_1 = 0)P(\sum_{i=2}^n X_i = t) + P(X_1 = 1)P(\sum_{i=2}^n X_i = t-1) \\
=& (1-\theta)^5 \binom{5(n-1)}{t} \theta^t (1-\theta)^{5(n-1)-t} \\
& + 5(1-\theta)^4\theta \binom{5(n-1)}{t-1} \theta^{t-1}(1-\theta)^{5(n-1)-t+1} \\
=& [\binom{5(n-1)}{t} + 5\binom{5(n-1)}{t-1}]\theta^t(1-\theta)^{5n-t}
\end{aligned}$$

Since $P(T = t) = \binom{5n}{t}\theta^t(1-\theta)^{5n-t}$.

$$\hat\tau = \begin{cases} 1 & \text{if } T = 0 \\ [\binom{5(n-1)}{T} + 5\binom{5(n-1)}{T-1}]/\binom{5n}{T} & \text{if } T \ge 1, \end{cases} \quad where\ T = \sum_{i=1}^n X_i.$$

Theorem 2. *If (1) T is a sufficient and complete statistic for θ; (2) $\phi(T)$ is a statistic that only depends on T, Then $\phi(T)$ is the unique UMVUE of $E(\phi(T))$.*
Theorem 3 (Rao-Blackwell). *Suppose that*
(1) W is an unbiased estimator of $\tau(\theta)$,
(2) T is sufficient for θ and
(3) $\hat{\tau} = E(W|T)$.
Then $Var(\hat{\tau}) \leq Var(W)$ and $E(\hat{\tau}) = \tau(\theta)$.

What is the diffenrece between Th 2 and 3 ?

Remark. The R-B Theorem does not say that $\hat{\tau}$ is the UMVUE.

Proof of R-B Th. $E(\hat{\tau}) = E(E(W|T)) = E(W) = \tau(\theta)$.
$Var(W) = Var(E(W|T)) + E(Var(W|T))$
Thus $Var(W) \geq Var(E(W|T)) = Var(\hat{\tau})$. □

Proof of Theorem 2.
Step (1) Claim: $\phi(T)$ is a UMVUE of $\tau(\theta) = E(\phi(T))$.
If $\phi(T)$ is not a UMVUE of $\tau(\theta)$,
then there exists an unbiased estimator W such that

$$Var(W) < Var(\phi(T)) \text{ for a } \theta = \theta_o \textbf{ (or for all } \theta \textbf{ ?)}.$$

We shall show that it leads to a contradiction.
Now $\hat{\tau} = E(W|T)$ is an unbiased estimator and

$$Var(\hat{\tau}) \leq Var(W) < Var(\phi(T)) \text{ for } \theta = \theta_o \text{ by R-B theorem.} \tag{1}$$

Let $g(T) = E(W|T) - \phi(T)$,
then $E(g(T)) = \tau(\theta) - \tau(\theta) = 0 \ \forall\ \theta$.
It follows that $P(g(T) = 0) = 1 \ \forall\ \theta$, **Why ?** that is,
$\phi(T) = E(W|T)$ w.p.1, a contradiction to Inequality (1) **Why ??**
The contradiction implies that $\phi(T)$ is an UMVUE of τ.

Step (2) $\vdash$: $Cov(W, \phi(T)) = \sigma_W \sigma_{\phi(T)}$ where $\underline{W \text{ is an arbitrary UMVUE of } \tau(\theta)}$.

$W^* = (W + \phi(T))/2$ is also unbiased, and

$$\begin{aligned}
Var(W) \leq& Var(W^*) && \text{(as } W \text{ is an UMVUE)}\\
=& \frac{1}{2^2}Var(\phi(T)) + \frac{1}{2^2}Var(W) + \frac{1}{2}Cov(\phi(T), W) &&\\
\leq& \frac{1}{2^2}Var(W) + \frac{1}{2^2}Var(W) + \frac{1}{2}Var(W) && \textbf{Why ??}\\
=& Var(W) && (Cov(X,Y) \leq \sigma_X\sigma_Y)
\end{aligned}$$

$\Rightarrow Cov(\phi(T), W) = Var(W) = \sqrt{Var(\phi(T))Var(W)}$ **Why ?**
Step (3) Claim: $\phi(T)$ is the unique (w.p.1) UMVUE of $\tau(\theta)$.
Recall that $Cov(X, Y) \leq \sigma_X\sigma_Y$
with equality iff $P(Y = a + bX) = 1$ for some constants a and b.
Let W be an arbitrary UMVUE of $\tau(\theta)$.
Thus Step (2) => $P(\phi(T) = a + bW) = 1$ for some constants a and b.
Then $E(\phi(T)) = a + bE(W)$ and thus $\tau(\theta) = a + b\tau(\theta) \ \forall\ \theta$.
It follows that $a = 0$ and $b = 1$, and thus $P(W = \phi(T)) = 1$. □

Theorem 1 (Cramér-Rao Inequality (CR- Ineq.)) Let X_1, ..., X_n be i.i.d. from $X \sim f(x;\theta)$ and let $W(\mathbf{X})$ be a statistic. Suppose that

(1) $\frac{d}{d\theta}E(W) = \begin{cases} \int \frac{\partial}{\partial\theta}W(\mathbf{x})f(\mathbf{x};\theta)d\mathbf{x} & \text{if } \mathbf{X} \text{ is continuous} \\ \sum_{\mathbf{x}} \frac{\partial}{\partial\theta}W(\mathbf{x})f(\mathbf{x};\theta) & \text{if } \mathbf{X} \text{ is discrete;} \end{cases}$

(2) $Var(W) < \infty$.
Let $\tau = E(W)$. Then
$Var(W) \geq \frac{(\frac{d}{d\theta}E(W))^2}{E((\frac{\partial}{\partial\theta}\ln f(\mathbf{X};\theta))^2)}$ $(= \frac{(\frac{d}{d\theta}E(W))^2}{nE((\frac{\partial}{\partial\theta}\ln f(X;\theta))^2)})$. **Why = ?**

Remark. In general, the CRLB $= \frac{(\tau'(\theta))^2}{I_n(\theta)}$, where $I_n(\theta) = E((\frac{\partial}{\partial\theta}\ln f(\mathbf{X};\theta))^2)$, $I_n(\theta)$ is called the Fisher information number.
Here $\mathbf{X} = (X_1, ..., X_n)$ and X_1, ..., X_n do **not** need to be i.i.d..
If they are, then $I_n(\theta) = nI_1(\theta)$, where $I_1(\theta) = E((\frac{\partial}{\partial\theta}\ln f(X_i;\theta))^2)$; Moreover,
if $\frac{\partial}{\partial\theta}E(\frac{\partial}{\partial\theta}\ln f(X_1;\theta)) = \begin{cases} \int \frac{\partial}{\partial\theta}(\frac{\partial}{\partial\theta}\ln f(x;\theta)f(x;\theta))dx & \text{if } X_1 \text{ is continuous} \\ \sum \frac{\partial}{\partial\theta}(\frac{\partial}{\partial\theta}\ln f(x;\theta)f(x;\theta)) & \text{if } X_1 \text{ is discrete,} \end{cases}$

then

$$I_n(\theta) = -E(\frac{\partial}{\partial\theta}(\frac{\partial}{\partial\theta}\ln f(\mathbf{X};\theta))) = -nE(\frac{\partial^2}{\partial\theta^2}\ln f(X_1;\theta)) \tag{2}$$

Proof of (2) under the assumption that X is continuous.
Let $Y_i = \frac{\partial}{\partial\theta}\ln f(X_i;\theta)$, then

$$\begin{aligned} E(Y_i) =& E(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)) \\ =& E(\frac{\frac{\partial}{\partial\theta}f(X_i;\theta)}{f(X_i;\theta)}) \\ =& \int \frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)}f(x;\theta)dx \\ =& \int \frac{\partial}{\partial\theta}f(x;\theta)dx \\ =& \frac{\partial}{\partial\theta}\int f(x;\theta)dx && \text{(by (1) in the theorem)} \\ =& 0. \end{aligned}$$

$E(Y_i^2) = Var(Y_i) = V(Y_i)$. $E(\sum_i Y_i) = 0$.

$$I_n(\theta) = E((\frac{\partial}{\partial\theta}\sum_i \ln f(X_i;\theta))^2) = E[(\sum_i Y_i)^2] = V(\sum_i Y_i) = \sum_i V(Y_i) = nI_1(\theta)$$

$$\begin{aligned} 0 =& E(Y_i) = \int(\frac{\partial}{\partial\theta}\ln f(x;\theta))f(x;\theta)dx \quad => \\ \frac{\partial}{\partial\theta}0 =& \frac{\partial}{\partial\theta}\int(\frac{\partial}{\partial\theta}\ln f(x;\theta))f(x;\theta)dx \\ =& \int \frac{\partial}{\partial\theta}\big[(\frac{\partial}{\partial\theta}\ln f(x;\theta))f(x;\theta)\big]dx && \text{(by assumption)} \\ =& \int [\frac{\partial}{\partial\theta}(\frac{\partial}{\partial\theta}\ln f(x;\theta))]f(x;\theta) + (\frac{\partial}{\partial\theta}\ln f(x;\theta))\frac{\partial}{\partial\theta}f(x;\theta)dx \\ =& \int [\frac{\partial^2}{\partial\theta^2}\ln f(x;\theta))]f(x;\theta)dx + \int(\frac{\partial}{\partial\theta}\ln f(x;\theta))\frac{\partial}{\partial\theta}f(x;\theta)dx \\ =& \int [\frac{\partial^2}{\partial\theta^2}\ln f(x;\theta))]f(x;\theta)dx + \int(\frac{\partial}{\partial\theta}\ln f(x;\theta))(\frac{\partial}{\partial\theta}\ln f(x;\theta))f(x;\theta)dx \\ =& E(\frac{\partial^2}{\partial\theta^2}\ln f(X_i;\theta)) + E((\frac{\partial}{\partial\theta}\ln f(X_i;\theta))^2) \end{aligned}$$

Thus $E(\frac{\partial^2}{\partial\theta^2}\ln f(X_i;\theta)) = -E((\frac{\partial}{\partial\theta}\ln f(X_i;\theta))^2)$.

§7.3.4. More about the Bayes estimator.
Interpretation of various estimation methods: MLE $\hat\theta$ maximizes $\mathcal{L}(\theta) = f_{\mathbf{X}}(\mathbf{x};\theta)$, maximizing the chance for given

$\mathbf{X} = \mathbf{x}$. MME $\tilde{\theta}$ solves θ through $E_\theta(\mathbf{X}^k) = \overline{\mathbf{X}^k}$. Unbiased estimator $\check{\theta}$ set $E(\check{\theta})) = \theta$., UMVUE $\check{\theta}$ is the best unbiased estimator in terms of variance. Why Bayes estimator $E(\theta|\mathbf{X})$?

Definitions: A decision problem consists of

$\mathcal{X}$ – sample space,

$\mathcal{A}$ – action space,

Θ – parameter space,

$L(\theta, a)$ – loss function, that is, L: $\Theta \times \mathcal{A} \to \mathcal{R}$.

A decision rule δ is a (measurable) function from $\mathcal{X}$ to $\mathcal{A}$, that is,

δ: $\mathcal{X} \to \mathcal{A}$.

$R(\theta, \delta) = E(L(\theta, \delta(X)))$ – risk function of δ, or more precisely, $R(\theta, \delta) = E(L(\theta, \delta(X))|\theta)$ **(function of** (θ, δ)**, not** X**))**.

$r(\pi, \delta) = E_\pi(R(\theta, \delta))$ – Bayes risk of δ. It is **not a function of** $(\mathbf{X}, \theta)$ **!!**

$\delta_B = \arg\inf_\delta r(\pi, \delta)$ is called the Bayes rule of θ w.r.t. prior π and loss L.

Remark. If $L = (\theta - a)^2$ (called the quadratic loss function or the squared error loss), then $E_\pi(\theta|\mathbf{X})$ is the Bayes rule w.r.t. π and L **(or Bayes estimator)**.

The Bayes estimator is the best in term of $E(E(\hat{\theta}(\mathbf{X}) - \theta)^2|\theta)$, **average error over** $(\mathbf{X}, \theta)$.

Example 1. Let $X \sim bin(n, \theta)$, $\pi(\theta) \sim beta(\alpha, \beta)$ where $\alpha = \beta = \sqrt{n}/2$. Then the MLE is $\hat{\theta} = X/n$, and the Bayes estimator under the square error loss is $\tilde{\theta} = E(\theta|X) = \frac{X+\alpha}{n+\alpha+\beta} = \frac{X+\sqrt{n}/2}{n+\sqrt{n}}$ **why ?**

Can we write $\tilde{\theta} = E(\theta|x) = \frac{x+\sqrt{n}/2}{n+\sqrt{n}}$?

Can we write $\tilde{\theta} = E(\theta|X) = \frac{x+\sqrt{n}/2}{n+\sqrt{n}}$?

This is an estimation problem and is also called a decision problem. In this decision problem,

$\mathcal{X} = \{0, 1, ..., n\}$ (set of possible observations)

$\Theta = [0, 1] = \mathcal{A}$ (set of possible estimates)

$L = (a - \theta)^2$ (error).

A decision rule δ is an estimator.

$\hat{\theta}$ and $\tilde{\theta}$ are both decision rules. Then

$R(\theta, \hat{\theta}) = E((\hat{\theta} - \theta)^2) = E((\frac{X}{n} - \theta)^2) = \sigma^2_{X/n} = V(X/n) = \frac{\theta(1-\theta)}{n}$.

$r(\pi, \hat{\theta}) = E(\frac{\theta(1-\theta)}{n}) = \frac{B(\alpha+1,\beta+1)}{B(\alpha,\beta)n}$ **Why ??**

Recall $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ and $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

$r(\pi, \hat{\theta}) = \frac{\alpha}{\alpha+\beta+1} \cdot \frac{\beta}{\alpha+\beta} \cdot \frac{1}{n} = \frac{1}{4\sqrt{n}(\sqrt{n}+1)}$

$$R(\theta, \tilde{\theta}) = E((\tilde{\theta} - \theta)^2) = MSE(\tilde{\theta}) = Var(\underbrace{\frac{X + \sqrt{n}/2}{n + \sqrt{n}}}_{\tilde{\theta}}) + (bias(\tilde{\theta}))^2$$

$R(\theta, \tilde{\theta}) = \frac{n\theta(1-\theta)}{(n+\sqrt{n})^2} + (\frac{n\theta+\sqrt{n}/2}{n+\sqrt{n}} - \theta)^2 = \frac{n}{4(n+\sqrt{n})^2}$

$r(\pi, \tilde{\theta}) = E(R(\theta, \tilde{\theta})) = \frac{n}{4(n+\sqrt{n})^2} = \frac{1}{4(\sqrt{n}+1)^2} < r(\pi, \hat{\theta})$.

It can be checked that $r(\pi, \tilde{\theta}) = \inf_\delta r(\pi, \delta)$ (see Remark later).

Thus $\tilde{\theta}$ minimizes the average error. w.r.t. L and π.

Example 2. Other loss functions:

$L(\theta, a) = |a - \theta|$,

$L(\theta, a) = \frac{(a-\theta)^2}{\theta(1-\theta)}$, where $\Theta = [0, 1]$, and $\frac{1}{0} \stackrel{def}{=} \infty$

Is $E(\theta|X)$ **still the Bayes rule w.r.t.** L **and** π **?**

Example 3. Suppose that $X \sim bin(n, p)$, $\pi(p) \sim beta(\alpha, \beta)$, with $\alpha = \beta = \sqrt{n}/2$, and $L(p, a) = \frac{(a-p)^2}{p(1-p)}$. Let $\hat{p}_1 = X/n$, and $\hat{p}_2 = \frac{X+\alpha}{n+\alpha+\beta}$. $R(p, \hat{p}_i) =$? $r(\pi, \hat{p}_i) =$?

Sol. $R(p, \hat{p}_1) = E(\frac{(X/n-p)^2}{p(1-p)}) = \frac{p(1-p)/n}{p(1-p)} = 1/n$.

$r(\pi, \hat{p}_1) = E(1/n) = 1/n$.

$$R(p, \hat{p}_2) = E(\frac{(\frac{X+\alpha}{n+\alpha+\beta} - p)^2}{p(1-p)}) = E((\frac{X + \alpha}{n + \alpha + \beta} - p)^2)\frac{1}{p(1 - p)} = \frac{n}{4(n + \sqrt{n})^2} \cdot \frac{1}{p(1 - p)}$$

(by Ex. 1).

$$r(\pi,\hat p_2)=cE(\frac{1}{p(1-p)})=c\frac{B(\alpha-1,\beta-1)}{B(\alpha,\beta)} \textbf{ Why ?}$$
$$=c\frac{(\alpha+\beta-1)(\alpha+\beta-2)}{(\alpha-1)(\beta-1)}=\frac{n}{4(n+\sqrt{n})^2}\cdot\frac{(\sqrt{n}-1)(\sqrt{n}-2)}{(\frac{\sqrt{n}}{2}-1)^2}$$
$$=\frac{(\sqrt{n}-1)}{(\sqrt{n}+1)^2(\sqrt{n}-2)}\begin{cases}>1/n=r(\pi,\hat p_1) & \text{if } n=4 \text{ or } 9,\\ <1/n=r(\pi,\hat p_1) & \text{if } n=100\end{cases}$$

Can we tell whether $\hat p_1$ or $\hat p_2$ is Bayes rule (w.r..t. π and L) ?
Remark. Under certain regularity conditions (in the Fubini Theorem),
(1) If $E(L(\theta,\delta)|\mathbf{X})$ is finite, then the Bayes rule is
$\delta_B(\mathbf{x})=\arg\min_a E(L(\theta,a)|\mathbf{X}=\mathbf{x})$.
Or, if T is sufficient and $E(L(\theta,a)|T)$ is finite, then the Bayes rule is
$\delta_B(t)=\arg\min_a E(L(\theta,a)|T=t)$. (2) If $L=(a-\theta)^2$, then $\delta_B=E(\theta|\mathbf{X})$.
Proof: Note that both $\mathbf{X}$ and θ are random.

$$\begin{aligned} r(\pi,\delta)=&E(E(L(\theta,\delta(\mathbf{X}))|\theta))\\ =&E(E(L(\theta,\delta(\mathbf{X}))|\mathbf{X})) \qquad \text{(by Fubini Theorem)}\end{aligned}$$

is minimized by minimizing $E(L(\theta,\delta(\mathbf{X}))|\mathbf{X}=\mathbf{x})$ for each $\mathbf{x}$
or minimizing $E(L(\theta,a)|\mathbf{X}=\mathbf{x})$ over all $a\in\mathcal{A}$ for each $\mathbf{x}$. **Why ?**
(2) If $L=(a-\theta)^2$, then
$$E(L(\theta,a)|\mathbf{X}=\mathbf{x})=E((a-\theta)^2|\mathbf{X}=\mathbf{x})$$
If $E((a-\theta)^2|\mathbf{X}=\mathbf{x})$ is finite, then
$\frac{\partial}{\partial a}E(L(\theta,a)|\mathbf{X}=\mathbf{x})=2E((a-\theta)|\mathbf{X}=\mathbf{x})$ (**Why ?? Is it right ?**)
$=\frac{\partial}{\partial a}[a^2-2aE(\theta|\mathbf{X}=\mathbf{x})+E(\theta^2|\mathbf{X}=\mathbf{x})]=2a-2E(\theta|\mathbf{X}=\mathbf{x})$.
$\frac{\partial^2}{\partial a^2}E(L(\theta,a)|\mathbf{X}=\mathbf{x})=2>0$.
Thus $a=E(\theta|\mathbf{X}=\mathbf{x})$ is the minimum point.
That is, $\delta_B=E(\theta|\mathbf{X})$ is the Bayes estimator w.r.t. L and π.

Remark. Hereafter, if we do not mention L in the problem, the Bayes estimator is $E(\theta|\mathbf{X})$, otherwise, the Bayes estimator is the Bayes rule w.r.t. the loss L and the prior π.
Remark. Under certain regularity conditions (in the Fubini Theorem),
If $E(L(\theta,\delta)|\mathbf{X})$ is finite, then the Bayes rule is
$\delta_B(\mathbf{x})=\arg\min_a E(L(\theta,a)|\mathbf{X}=\mathbf{x})$.
Or, if T is sufficient and $E(L(\theta,a)|T)$ is finite, then the Bayes rule is
$\delta_B(t)=\arg\min_a E(L(\theta,a)|T=t)$.
Example 4. If one observes X, where $X\sim bin(n,p)$, $L=\frac{(a-p)^2}{p(1-p)}$, $\pi(p)\sim U(0,1)$,
then Bayes estimator $\hat p$= ?
Sol. $\hat p=\delta_B(x)=\arg\min_a \underbrace{E(L(p,a)|X=x)}_{=?}$.

$\int L(p,a)\underbrace{\pi(p|x)}_{=?}dp$

$\frac{f(x,p)}{f_X(x)}=?$
The joint distribution of (X,p) is

$$\begin{aligned} f(x,p)=&f(x|p)\pi(p)\\ =&\binom{n}{x}p^x(1-p)^{n-x}\mathbf{1}(p\in[0,1])\\ \propto&p^x(1-p)^{n-x}\mathbf{1}(p\in(0,1)).\end{aligned}$$
$\pi(p|x)\sim$ **?**

Thus $\pi(p|x) \sim beta(x+1, n-x+1)$.

$$g(a) = E(L(p,a)|X=x) = \int_0^1 \frac{(a-p)^2}{p(1-p)} cp^x(1-p)^{n-x}dp$$

$$g(a) = c\int_0^1 (a-p)^2 p^{x-1}(1-p)^{n-x-1}dp$$

Step (1) $g'(a) = c\int_0^1 2(a-p)p^{x-1}(1-p)^{n-x-1}dp$. $g'(a) = 0 \Rightarrow$

$$\begin{aligned} a &= \frac{\int_0^1 pp^{x-1}(1-p)^{n-x-1}dp}{\int_0^1 p^{x-1}(1-p)^{n-x-1}dp} \\ &= \frac{\int_0^1 \frac{pp^{x-1}(1-p)^{n-x-1}}{B(x,n-x)}dp}{\int_0^1 \frac{p^{x-1}(1-p)^{n-x-1}}{B(x,n-x)}dp} \\ &= \text{mean of a beta distribution} = \frac{x}{x+(n-x)} \end{aligned}$$

Step (2) $g''(a) = c\int_0^1 2p^{x-1}(1-p)^{n-x-1}dp > 0$.
=> $a = \hat{p}(x) = x/n$ is the Bayes estimator of p.
Are we done ???

$$\begin{aligned} g(a) =& E(L(p,a)|X=x) \\ =& c\int_0^1 (a^2 - 2ap + p^2)p^{x-1}(1-p)^{n-x-1}dp \\ \propto& a^2B(x,n-x) - 2aB(x+1,n-x) + B(x+2,n-x). \end{aligned}$$

$$B(\alpha,\beta) < \infty \text{ iff } \alpha > 0 \text{ and } \beta > 0. \tag{1}$$

1. **Notice that** if $x \neq 0$ or n, $g(a)$ is finite for all $a \in [0,1]$.
   $g'(a) = 2c[aB(x,n-x) - B(x+1,n-x)] = 0$
   $g''(a) = 2cB(x,n-x) > 0$.
   Thus $g(a)$ is minimized by
   $a = \delta_B(x) = \frac{\int_0^1 p^x(1-p)^{n-x-1}dp}{\int_0^1 p^{x-1}(1-p)^{n-x-1}dp} = \frac{B(x+1,n-x)}{B(x,n-x)} = \frac{\Gamma(x+1)\Gamma(n-x)\Gamma(n)}{\Gamma(x)\Gamma(n-x)\Gamma(n+1)} = x/n$.
2. **Notice that** if $x = 0$, $g(a)$ is finite only when $a = 0$, as

$$g(0) = c\int_0^1 p^{x+1}(1-p)^{n-x-1}dp = cB(2,n)$$

   Otherwise, **(if unaware of (1))** $g(a) = c\{\int_0^1 a^2p^{-1}(1-p)^{n-1}dp - 2aB(1,n) + B(2,n)\}$
   $\geq c\int_0^{1/2} a^2p^{-1}(0.5)^{n-1}dp + c\{-2aB(1,n) + B(2,n)\} = \lim_{y\downarrow 0} ca^2(0.5)^{n-1}(\ln p\big|_y^{0.5}) = \infty$.
   Thus $g(a)$ is minimized by $a = \delta_B(0) = 0 = 0/n$.
   Can we say that $aB(x,n-x) = 0$ **if** $x = 0 = a$ **?**
3. **Notice that** if $x = n$, $g(a)$ is finite only when $a = 1$, by symmetry.
   Thus $g(a)$ is minimized by $a = \delta_B(n) = 1 = n/n$.

Answer: The Bayes estimator w.r.t. π and L is $\delta_B(X) = X/n$.
Question about

$$\frac{\partial}{\partial y}\int_{a(y)}^{b(y)} g(x,y)dx = g(b(y),y)b'(y) - g(a(y),y)a'(y) + \int_{a(y)}^{b(y)} \frac{\partial}{\partial y}g(x,y)dx$$

$\frac{\partial}{\partial y}\int_0^1 \sin(xy)\mathbf{1}(x<y)dx$ **= ?**

Chapter 8. Hypothesis Testing

§8.1. Two types of inferences: $\begin{cases} \text{1. Estimation problem: } \theta = ? \\ \text{2. Testing Problem: } \theta = \theta_o \text{ ? Here } \theta_o \text{ is given.} \end{cases}$

Example 1. A slot machine is claimed to have winning rate 40%. To test the claim, 5 runs are made. Observe X times of winning. Let p be the winning rate of the machine.

Possible Questions:

H_0 : null hypothesis	H_1 : alternative hypothesis	made by
$p = 40\%$?	$p \neq 40\%$?	manufacturer
	$p > 40\%$?	casino owner
	$p < 40\%$?	player

If H_0 is correct, then $X \sim bin(5, 2/5)$ and one expects 2 winnings.

The maker rejects H_0	if	$X = 0, 4, 5$ but nor 1,2,3.
The owner rejects H_0	if	$X = 4, 5$ but not 0, 1,2,3
A player rejects H_0	if	$X = 0$, but not $1, 2, 3, 4, 5$.
		rejection region (RR)

A test statistic or test function is $\phi = \mathbf{1}(X \in RR)$,

which has two interpretations:

1. If $X \in RR$, then $\phi = 1$ or H_1 is accepted (often say **rejecting** H_0);
   if $X \notin RR$, then $\phi = 0$ or H_0 is accepted (often say **not rejecting** H_0).
2. The probability of rejecting H_0 is $\begin{cases} 1 & \text{if } X \in RR \\ 0 & \text{otherwise.} \end{cases}$

A testing hypothesis for $\theta \in \Theta$ consists of 5 elements:

1. H_0: $\theta \in \Theta_o$ ($\Theta_o = \{0.4\}$ in Example 1).
2. H_1: $\theta \in \Theta_o^c = \Theta \setminus \Theta_o$

	H_1	Θ	Θ_o^c
(in Example 1)	$\theta \neq 0.4$	$[0, 1]$	$[0, 0.4) \cup (0.4, 1]$
	$\theta > 0.4$	$[0.4, 1]$	$(0.4, 1]$
	$\theta < 0.4$	$[0, 0.4]$	$[0, 0.4)$

3. Test statistic ϕ ($= \mathbf{1}(X \in RR)$ in Example 1).
4. α – size of the test defined by $\alpha = \sup_{\theta \in \Theta_o} E_\theta(\phi)$.
5. Conclusion: Reject or do not reject H_0.

Two types of errors:

1. Type I error: reject correct H_0, denoted by $H_1|H_0$.
2. Type II error: do not reject wrong H_0, denoted by $H_0|H_1$.

Definition. $\beta(\theta)$ – power function of the test defined by $\beta(\theta) = E_\theta(\phi)$.

For $\theta \in \Theta_o^c$, $\beta(\theta)$ is called the power (at θ) of the test.

If $\theta \in \Theta_o$ then $\beta(\theta) = P(H_1|H_0)$, the probability of type I error;
If $\Theta_o = \{\theta_o\}$ then $\beta(\theta_o) = \alpha$, the size of the test;
If $\theta \in \Theta_o^c$, then $\beta(\theta) = 1 - P_\theta(H_0|H_1)$,
where $P_\theta(H_0|H_1)$ is the probability of type II error.

Example 1 (continued). Compute $\beta(p)$ and α.

$\beta(p) = E(\mathbf{1}(X \in RR)) = \sum_{x \in RR} \binom{5}{x} p^x (1-p)^{5-x}$.

$\alpha = P_p(X \in RR)$ when $p = 0.4$.

```
R
x=0:5
round(dbinom(x,5,0.4),3)
[1] 0.078 0.259 0.346 0.230 0.077 0.010
```

1. H_1: $p < 0.4$. $\alpha = (P(X \in \{0\})) = 0.078$.
2. H_1: $p > 0.4$. $\alpha = (P(X \in \{4, 5\})) = 0.077 + 0.010 = 0.087$.
3. H_1: $p \neq 0.4$. $\alpha = (P(X \in \{0, 4, 5\})) = 0.078 + 0.077 + 0.010 = 0.165$.

§**8.2. Question:** How to construct a test ?
Ans. Method 1. Likelihood ratio test (LRT):

Let X_1, ..., X_n be a random sample from $f(x;\theta)$.
For testing H_0: $\theta \in \Theta_o$ v.s. H_1: $\theta \in \Theta_o^c$,

LRT $\phi = \mathbf{1}(\lambda \le c)$, where

$$\lambda = \frac{\sup_{\theta\in\Theta_o} L(\theta|\mathbf{x})}{\sup_{\theta\in\Theta} L(\theta|\mathbf{x})} = \frac{L(\hat\theta_o|\mathbf{x})}{L(\hat\theta|\mathbf{x})}$$

$L(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i;\theta)$, $\hat\theta$ is the MLE of θ under Θ, $\hat\theta_o$ is the MLE of θ under Θ_o, c is determined by $\alpha = \sup_{\theta\in\Theta_o} P(\lambda \le c)$, or otherwise, $c = \sup\{t:\ \alpha \ge \sup_{\theta\in\Theta_o} P(\lambda \le t)\}$.
Q: How to understand λ ?
Two extremes ?

Is $\lambda = 1$ (or $\lambda >> c$) likely under H_0 or H_1 ?
Is $\lambda = 0$ (or $\lambda << c$) likely under H_0 or H_1 ?

Example 1. A random sample from $N(\mu, 1)$ results in $\overline{X} = 1.1$, where $n = 100$. Do you believe $\mu = 1$?
Sol. Use LRT.
H_0: $\mu = 1$ v.s. H_1: $\mu \ne 1$.

$\alpha = 0.05$.
$L(\mu|\mathbf{x}) = c\exp(-\frac{1}{2}\sum_{i=1}^n (X_i - \mu)^2)$,
$\Theta_o = \{1\}$: MLE $\hat\mu_0 = 1$ $(= \mu_0)$;
$\Theta = (-\infty, \infty)$: MLE $\hat\mu = \overline{X}$;

$$\begin{aligned}\lambda =& \frac{L(\hat\mu_0|\mathbf{x})}{L(\hat\mu|\mathbf{x})}\\
=& \frac{c\exp(-\frac{1}{2}\sum_{i=1}^n (X_i - \mu_0)^2)}{c\exp(-\frac{1}{2}\sum_{i=1}^n (X_i - \overline{X})^2)}\\
=& \exp(-\frac{1}{2}\sum_{i=1}^n [(X_i-\mu_0)^2 - (X_i - \overline{X})^2])\\
=& \exp(-\frac{1}{2}\sum_{i=1}^n [(2X_i - \overline{X} - \mu_0)(\overline{X} - \mu_0)])\\
=& \exp(-\frac{1}{2}[(2n\overline{X} - n\overline{X} - n\mu_0)(\overline{X} - \mu_0)])\\
=& \exp(-\frac{n}{2}(\overline{X} - \mu_0)(\overline{X} - \mu_0)])\\
=& \exp(-\frac{n}{2}(\overline{X} - \mu_0)^2).\end{aligned}$$

$\phi = \mathbf{1}(\lambda \le c) = \mathbf{1}(|\overline{X} - \mu_0| \ge c_1)$.

Since $\alpha = E_{\mu_0}(\phi) = 0.05$,
$\overline{X} \sim N(\mu_0, 1/n)$,
$\frac{\overline{X}-1}{1/\sqrt{n}} \sim N(0,1)$,
$P(|\frac{\overline{X}-1}{1/\sqrt{n}}| > 1.96) \approx 0.05$,
$c_1 = 1.96/\sqrt{n}$. Or $c = \exp(-\frac{n}{2}c_1^2)$. **(It is important to find c_1 and c).**

That is $\phi = \mathbf{1}(|\overline{X} - 1| \ge 1.96/\sqrt{n})$.

Thus do not reject H_o.
It is likely that $\mu = 1$.

Where are the 5 elements of a test ? A testing hypothesis for $\theta \in \Theta$ consists of 5 elements:

1. H_0
2. H_1
3. Test statistic ϕ $(= \mathbf{1}(\lambda \le c)$ for LRT)

4. α – size of the test defined by $\alpha = \sup_{\theta\in\Theta_o} E_\theta(\phi)$.
5. Conclusion: Reject or do not reject H_0 and answer to the related question.

Example 3. Suppose that $X_1, ..., X_4 \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$ are unknown, $\overline{X} = 3$ and $S^2 = 4$. H_0: $\mu \le 0$ $(= \mu_0)$ v.s. H_1: $\mu > \mu_0$. LRT ?

Sol.

Remark. A natural estimator of μ = ?

If $\overline{X} = 100$, H_0 or H_1 ?

If $\overline{X} = -0.001$, H_0 or H_1 ?

If $\overline{X} = 3$, H_0 or H_1 ? **need to find out now).**

A natural test is $\phi_+ = \mathbf{1}(\hat\mu > b)$ **Why ?**

5 elements of a test: (1) ? (2) ?

(3) Choose size $\alpha = 0.05$.

(4) Test statistics: $\mathbf{1}(\lambda \le c)$ = ? $\lambda = \frac{L(\hat\theta_o|\mathbf{X})}{L(\hat\theta|\mathbf{X})}$ = ? c= ? **The main work!!**

(5) Conclusion. **Don't forget !**

$L = (\frac{1}{\sqrt{2\pi\sigma^2}})^n \exp(-\frac{1}{2}\sum_i (X_i - \mu)^2/\sigma^2)$

$\Theta_o = \{(\mu, \sigma^2): \ \mu \le \mu_0, \sigma > 0\}$.

$\Theta = \{(\mu, \sigma^2): \ \mu \in \mathcal{R}^1, \sigma > 0\}$.

MLE under Θ: $\hat\theta = (\overline{X}, \hat\sigma^2)$, where $\hat\sigma^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$.

MLE under Θ_o: $\hat\theta_o = (\hat\mu_0, \hat\sigma_0^2) = (\overline{X}\wedge\mu_0, \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}\wedge\mu_0)^2)$

(see Example 3 in MLE section), or the derivation as follows.

If $\overline{X} \le \mu_0$, then $\hat\theta \in \Theta_o$ and thus it is the maximum point of the likelihood $L(\theta|\mathbf{X})$.

If $\overline{X} > \mu_0$, then,

since $\hat\theta$ is the unique stationary point in Θ,

the maximum point of $L(\theta|\mathbf{X})$ must be on the boundary:

$$\begin{array}{lcccc} boundaries: & \mu = -\infty & \mu = \mu_0 & \sigma = 0 & \sigma = \infty \\ L(\theta|\mathbf{X}): & 0 & finite & 0 & 0 \end{array}$$

It is easy to show that on the boundary $\mu = \mu_0$, the maximum point of the likelihood is achieved at $\hat\theta_o = (\hat\mu_0, \hat\sigma_0^2) = (\mu_0, \frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2)$.

$L(\hat\theta) = (\frac{1}{\sqrt{2\pi\hat\sigma^2}})^n \exp(-\frac{1}{2}\sum_i (X_i - \hat\mu)^2/\hat\sigma^2) = (\frac{1}{\sqrt{2\pi\hat\sigma^2}})^n \exp(-\frac{n}{2})$.

$$L(\hat\theta_o) = (\frac{1}{\sqrt{2\pi\hat\sigma_0^2}})^n \exp(-\frac{1}{2}\sum_i (X_i - \hat\mu_0)^2/\hat\sigma_0^2)$$
$$= \begin{cases} (\frac{1}{\sqrt{2\pi\hat\sigma^2}})^n \exp(-\frac{n}{2}) & \text{if } \overline{X} \le \mu_0 \\ (\frac{1}{\sqrt{2\pi\hat\sigma_0^2}})^n \exp(-\frac{n}{2}) & \text{if } \overline{X} > \mu_0. \end{cases}$$

$$\lambda = \begin{cases} 1 & \text{if } \overline{X} \le \mu_0 \\ (\frac{\sum_i (X_i - \overline{X})^2}{\sum_i (X_i - \mu_o)^2})^{n/2} & \text{if } \overline{X} > \mu_0. \end{cases}$$

$$\begin{aligned} \phi =& \mathbf{1}(\lambda \le c) && c \approx 0? \ \ or \ \ c \approx 1? \\ =& \mathbf{1}(\frac{\sum_i (X_i - \overline{X})^2}{\sum_i (X_i - \mu_o)^2} \le c^{2/n}) && \textbf{Why ? Is it correct ?} \end{aligned}$$

$$\phi_+ = \mathbf{1}(\frac{\sum_i (X_i - \overline{X})^2}{\sum_i (X_i - \mu_o)^2} \le c^{2/n})\mathbf{1}(\overline{X} > \mu_o) \qquad \textbf{Is it correct ?}$$

$$\begin{aligned} \phi =& \mathbf{1}(\frac{\sum_i (X_i - \overline{X})^2}{\sum_i (X_i - \overline{X})^2 + \sum_i (\overline{X} - \mu_o)^2} \le c^{2/n}) \ \ (as \ \ \sum_i (X_i - \mu_o)^2 = \sum_i (X_i - \overline{X} + \overline{X} - \mu_o)^2 \\ & \qquad = \textstyle\sum_i (X_i - \overline{X})^2 + \sum_i (\overline{X} - \mu_o)^2) \\ =& \mathbf{1}(\frac{\sum_i (X_i - \overline{X})^2 + \sum_i (\overline{X} - \mu_o)^2}{\sum_i (X_i - \overline{X})^2} \ge 1/c^{2/n}) \end{aligned}$$

$$=\mathbf{1}(\frac{\sum_i(\overline{X}-\mu_o)^2}{\sum_i(X_i-\overline{X})^2} \geq 1/c^{2/n}-1)$$

$$=\mathbf{1}(\frac{n(\overline{X}-\mu_o)^2}{\sum_i(X_i-\overline{X})^2} \geq c_2) \qquad c_2 = ?$$

$$=\mathbf{1}(\frac{|\overline{X}-\mu_o|}{\sqrt{S^2/n}} \geq c_3) \qquad c_3 = ?$$

Recall that $\frac{\overline{X}-\mu}{\sqrt{S^2/n}} \sim t_{n-1}$ if X_i's i.i.d. $\sim N(\mu, \sigma^2)$.

$\alpha = \sup_{\mu\leq\mu_0} E(\phi) = \sup_{\mu\leq\mu_0} P(\frac{|\overline{X}-\mu_o|}{\sqrt{S^2/n}} \geq c_3)$

$= \sup_{\mu\leq\mu_0} P(\frac{|\overline{X}-\mu+(\mu-\mu_o)|}{\sqrt{S^2/n}} \geq c_3)$.

Since t_{n-1} density function is bell-shaped and symmetric about 0,

$$\alpha = \sup_{\mu\leq\mu_0} E(\phi) = \sup_{\mu\leq\mu_0} P(\frac{|\overline{X}-\mu+(\mu-\mu_o)|}{\sqrt{S^2/n}} \geq c_3) = P(\frac{|\overline{X}-\mu_o|}{\sqrt{S^2/n}} \geq c_3)$$

with $\mu_0 = 0$ here. That is,

$$\phi = \mathbf{1}(\frac{|\overline{X}-\mu_o|}{\sqrt{S^2/n}} \geq t_{0.025,n-1}).$$

Q: Two tests: $\begin{cases} \phi_+ = \mathbf{1}(\frac{\overline{X}-\mu_o}{\sqrt{S^2/n}} \geq t_{0.05,n-1}) => 1_{(3>2.353)} \text{ (reject } H_0), \\ \phi = \mathbf{1}(\frac{|\overline{X}-\mu_o|}{\sqrt{S^2/n}} \geq t_{0.025,n-1}) => 1_{(|3|>3.182)} \text{ (don't reject } H_0) \end{cases}$

Which makes more sense ?

Recall H_0: $\mu \leq 0$ $(= \mu_0)$ v.s. H_1: $\mu > \mu_0$.

Question: Something goes wrong ?

$$\mathbf{1}(\lambda \leq c) = \begin{cases} \mathbf{1}(1 \leq c) & \text{if } \overline{X} < 0 \\ \mathbf{1}((\frac{\sum_i(X_i-\overline{X})^2}{\sum_i(X_i-\mu_o)^2})^{n/2} \leq c) & \text{if } \overline{X} > \mu_0 \end{cases}$$

$$=\mathbf{1}(\frac{|\overline{X}-\mu_o|}{\sqrt{S^2/n}} \geq c_3)\mathbf{1}(\overline{X} > \mu_0)$$

$$=\mathbf{1}(\frac{\overline{X}-\mu_o}{\sqrt{S^2/n}} \geq c_3) = \phi_+$$

(4) Test statistic is ϕ_+.

(5) Reject H_0. The data does not support the claim that $\mu \leq 0$. □

Example 2. A random sample X_1, ..., X_4 from $f = \exp(-(x-\theta))$, $x \geq \theta$.

H_0: $\theta \leq 1$ $(= \theta_o)$ v.s. H_1: $\theta > 1$. LRT of size $\alpha = 0.01$ if $X_{(1)} = 1.1$?

Sol. $\mathbf{1}(\lambda \leq c) = ?$ $\lambda = \frac{L(\hat{\theta}_o|\mathbf{X})}{L(\hat{\theta}|\mathbf{X})} = ?$ $c= ?$ **The main task !!**

What will you do if $\hat{\theta} = 0.1$? $\hat{\theta} = 100$?

Remark: A natural test is $\mathbf{1}(\hat{\theta} > b)$ **Why ?**

Step (1) MLE under $\Theta = (-\infty, \infty)$:

$$L = \prod_{i=1}^n \exp(-(X_i-\theta)) = \exp(-\sum_{i=1}^n X_i + n\theta) \uparrow \ in\ \theta......$$

MLE = ?

$$L = \prod_{i=1}^n [\exp(-(X_i-\theta))\mathbf{1}(X_i \geq \theta)]$$

$$= \exp(-\sum_{i=1}^{n} X_i + n\theta)\mathbf{1}(X_{(1)} \geq \theta)$$

$$= \exp(-n\overline{X} + n\theta)\mathbf{1}(X_{(1)} \geq \theta) \begin{cases} \uparrow \ in\ \theta & \text{if } \theta \leq X_{(1)} \\ = 0 & \text{if } \theta > X_{(1)}. \end{cases} \quad => \quad \hat{\theta} = X_{(1)}.$$

Step (2) MLE under $\Theta_o = (-\infty, 1]$:

$$L = \prod_{i=1}^{n} [\exp(-(X_i - \theta))\mathbf{1}(X_i \geq \theta)\mathbf{1}(1 \geq \theta)]$$

$$= \exp(-\sum_{i=1}^{n} X_i + n\theta)\mathbf{1}(X_{(1)} \geq \theta)\mathbf{1}(1 \geq \theta)$$

$$= \exp(-n\overline{X} + n\theta)\mathbf{1}(\theta \leq X_{(1)} \wedge 1) \begin{cases} \uparrow \ in\ \theta & \text{if } \theta \leq X_{(1)} \wedge 1 \\ = 0 & \text{if } \theta > X_{(1)} \wedge 1. \end{cases} \quad => \ \hat{\theta}_o = X_{(1)} \wedge 1.$$

Step (3) $\lambda = \frac{L(\hat{\theta}_o|\mathbf{X})}{L(\hat{\theta}|\mathbf{X})} = \begin{cases} 1 & \text{if } X_{(1)} \leq 1 \\ \frac{\exp(-n\overline{X}+n\cdot 1)\mathbf{1}(\hat{\theta}_o \leq X_{(1)} \wedge 1)}{\exp(-n\overline{X}+nX_{(1)})\mathbf{1}(\hat{\theta} \leq X_{(1)})} & \text{if } X_{(1)} > 1. \end{cases}$ $\lambda = 1(X_{(1)} \leq 1) + \frac{\exp(-n\overline{X}+n)}{\exp(-n\overline{X}+nX_{(1)})}1(X_{(1)} > 1)$

$\lambda = \begin{cases} 1 & \text{if } X_{(1)} \leq 1 \\ \exp(n(1 - X_{(1)})) & \text{if } X_{(1)} > 1 \end{cases} = [\exp(n(1 - X_{(1)}))]^{1(X_{(1)} > 1)}.$

$\phi = \mathbf{1}(\lambda \leq c) = \mathbf{1}(\exp(n(1 - X_{(1)})) \leq c)\mathbf{1}(X_{(1)} > 1)$ **Why**??

$\phi = \mathbf{1}(X_{(1)} \geq c_1)\mathbf{1}(X_{(1)} > 1)$

= the natural test $\mathbf{1}(\hat{\theta} > b)$, where $\hat{\theta} = X_{(1)}$ and $b > 1$.

Step (4) c= ? or c_1= ?

Use $\alpha = \sup_{\theta \leq 1} E(\phi)$.

$$E(\phi) = P(X_{(1)} \geq c_1) = \begin{cases} P(X_1 \geq c_1, ..., X_n \geq c_1) \\ \int_{c_1}^{\infty} \frac{n!}{1!(n-1)!}(f_X(x))^1(1 - F(x))^{n-1}dx \end{cases} = (P(X_1 \geq c_1))^n$$

Note that $f(x;\theta) = e^{-(x-\theta)} = P(X > x)$, $x > \theta$ **in this case !!**.

$E(\phi) = (e^{-(c_1-\theta)})^n = e^{-nc_1+n\theta}$,

$\alpha = \sup_{\theta \leq 1} e^{-nc_1+n\theta} = e^{-nc_1+n} = 0.01$. **Why ??**

$c_1 = (-\ln 0.01)/n + 1 \approx 2.15$,

(or $c_1 = \frac{(-\ln 0.01)}{n} + \theta_o$ if H_o: $\theta \leq \theta_o$).

Thus the test is $\phi = \mathbf{1}(X_{(1)} \geq \frac{\ln 100}{n} + 1 \approx 2.15)$.

Step (5) Do not reject H_0. θ is likely ≤ 1.

Midterm on March 20.

§8.3.

Two types of errors:

1. Type I error: reject correct H_0, denoted by $H_1|H_0$.
2. Type II error: do not reject wrong H_0, denoted by $H_0|H_1$.

Definition.

$\beta(\theta)$ – power function of the test defined by $\beta(\theta) = E_\theta(\phi)$.

If $\theta \in \Theta_o$ then $\beta(\theta) = P(H_1|H_0)$ is the probability of type I error;

If $\theta \in \Theta_o^c$, then $1 - \beta(\theta) = P_\theta(H_0|H_1)$ is the probability of type II error.

Question 1: How to find a good test ?

Ideally, $P(H_1|H_0) = 0$ and $P(H_0|H_1) = 0$, that is

$$\beta(\theta) = \begin{cases} 0 & \text{if } \theta \in \Theta_o \text{ (size } \alpha = 0) \\ 1 & \text{if } \theta \in \Theta_o^c \text{ (power } = 1). \end{cases} \tag{1}$$

Question 2: Is it possible ?

Example 1. Suppose $X \sim bin(5, \theta)$, H_0: $\theta = 0.5$ v.s. H_1: $\theta = 0.4$.
$\phi = \mathbf{1}(X \in RR)$.

$$\beta(\theta) = E(\mathbf{1}(X \in RR)) = P(X \in RR) = \sum_{x \in RR} \binom{5}{x} \theta^x (1-\theta)^{5-x}$$

$\beta(0.5) = 0 \Rightarrow RR = \emptyset$. **(Why ?)**
$\beta(\theta) = 1$ with $\theta \in \{0.4\} \Rightarrow RR = \{0, 1, 2, 3, 4, 5\} = \mathcal{X}$, the sample space.
Answer to Q2: It is impossible that (1) holds.
Recall $\beta_\phi(\theta) = E_\theta(\phi)$, size of $\phi = \sup_{\theta \in \Theta_o} \beta_\phi(\theta)$.
Definition. A test is a level α test if its size $\leq \alpha$.
A test ϕ is unbiased if $\beta_\phi(\theta_1) \geq \beta_\phi(\theta_o)$ $\forall$ $(\theta_1, \theta_o) \in \Theta_o^c \times \Theta_o$.
A test ϕ is uniformly most powerful (UMP) within a class $\mathcal{C}$ if

$$\phi \in \mathcal{C} \text{ and } E(\phi) \geq E(\phi_*) \ \forall \ \theta \in \Theta_o^c \text{ and } \forall \ \phi_* \in \mathcal{C}.$$

Alternatives for optimal tests:
(1) the UMP level α test,
(2) the UMP unbiased test.
Why not the UMP size α test ?
Example 1 (continued). $X_1 \sim bin(5, \theta)$. H_0: $\theta = 0.5$, H_1: $\theta = 0.4$. Ideally, size of ϕ is α for the optimal test ϕ. Let $\alpha = 0.05$, is there a size α LRT ?
Sol. The LRT is of the form $\phi = \mathbf{1}(\lambda \leq a)$ or $\mathbf{1}(\hat{\theta} \leq c)$ **Why ?**
or $\phi = \mathbf{1}(X_1 \leq c)$. **Why?**
(1) Intuition based on $\hat{\theta}$;
(2) Direct derivation from $\mathbf{1}(\lambda \leq c)$.
$\lambda = f(X_1; \hat{\theta}_o)/f(X_1; \hat{\theta}) = f(X_1; 0.5)/f(X_1; \hat{\theta})$ **Why ?**
Θ_0 = ? =0.5 ? ={0.5} ? Θ= ?

```
x=dbinom(0:5,5,0.5)
[1] 0.031 0.156 0.312 0.312 0.156 0.031
y=dbinom(0:5,5,0.4)
[1] 0.078 0.259 0.346 0.230 0.077 0.010
```

$$\begin{pmatrix} X: & 0 & 1 & 2 & 3 & 4 & 5 \\ \hat{\theta}: & 0.4 & 0.4 & 0.4 & 0.5 & 0.5 & 0.5 \\ \lambda: & ? & ? & \frac{312}{346} & 1 & 1 & 1 \end{pmatrix}$$

$\alpha = E(\phi) = P(X_1 \leq c)$.

$c =$	-0.1	0	1	2	3	4	5	6
$\alpha =$	0	$\frac{1}{32}$	$\frac{6}{32}$	*larger*				
$\approx$	0	0.031	0.187	0.500	0.812	0.969	1	1

Answer: No size 0.05 LRT.
Definition. A test of form $\phi = \mathbf{1}(X \in RR)$ is called a non-randomized test.
A randomized test for testing H_o v.s. H_1 is
a function ϕ from the sample space $\mathcal{X}$ to $[0, 1]$.
$\phi = \phi(\mathbf{X})$ is the probability of rejecting H_0 for observing $\mathbf{X}$.
$\phi(x) = 0$, reject H_0 w.p.0 if x is observed.
$\phi(x) = 1$, reject H_0 w.p.1 if x is observed.
$\phi(x) = 1/2$, reject H_0 w.p.1/2, if x is observed; *e.g.*,
flip a coin, reject H_0 if the head faces up.
Theorem 1. (Neymann-Pearson Lemma). *Consider testing H_0: $\theta = \theta_o$ v.s. H_1: $\theta = \theta_1$, based on $\mathbf{X} \sim f(\mathbf{x}; \theta)$, $\theta \in \Theta = \{\theta_o, \theta_1\}$.*
Let ϕ be a test such that for some $k \geq 0$,

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } f(\mathbf{x}; \theta_1) > kf(\mathbf{x}; \theta_o) \\ 0 & \text{if } f(\mathbf{x}; \theta_1) < kf(\mathbf{x}; \theta_o) \end{cases} \quad \textbf{(Why not } \tfrac{f(\mathbf{X}:\theta_1)}{f(\mathbf{X}:\theta_o)} \textbf{ ?)} \tag{2}$$

and

$$E_{\theta_o}(\phi(\mathbf{X})) = \alpha. \tag{3}$$

Then

*a. (**Sufficiency**) Each level α test ϕ^* satisfying Eq. (2) and (3) is also a UMP level α test.*

*b. (**Necessity**) If $\exists$ ϕ satisfying Eq. (2) and (3) with $k > 0$, then each UMP level α test ϕ^* satisfying Eq. (3) and also satisfies Eq. (2) except on a set A satisfying $P_\theta(\mathbf{X} \in A) = 0$ $\forall$ θ ($P(\phi^*(\mathbf{X}) = \phi(\mathbf{X})) = 1$ $\forall$ θ).*

LRT $\phi = \mathbf{1}(\frac{f(\mathbf{X};\hat\theta_0)}{f(\mathbf{X};\hat\theta)} \le c)$.

A test of form $\phi = \mathbf{1}(X \in RR)$ is called a non-randomized test. $\phi = 1 => H_1$, $\phi = 0 => H_0$.

A randomized test $\phi = \phi(\mathbf{X})$ is the probability of rejecting H_0 for observing $\mathbf{X}$.

$\phi(x) = 0$, reject H_0 w.p.0 if x is observed.

$\phi(x) = 1$, reject H_0 w.p.1 if x is observed.

$\phi(x) = 1/2$, reject H_0 w.p.1/2, if x is observed; *e.g.*,

flip a coin, reject H_0 if the head faces up.

Neymann-Pearson Lemma. H_0: $\theta = \theta_o$ v.s. H_1: $\theta = \theta_1$,

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } f(\mathbf{x};\theta_1) > kf(\mathbf{x};\theta_o) \\ 0 & \text{if } f(\mathbf{x};\theta_1) < kf(\mathbf{x};\theta_o) \end{cases} \quad \textbf{(Why not } \frac{f(\mathbf{X}:\theta_1)}{f(\mathbf{X}:\theta_o)} \textbf{ ?)} \tag{2}$$

and

$$E_{\theta_o}(\phi(\mathbf{X})) = \alpha. \tag{3}$$

"iff" ϕ is a UMP level α test.

Example 1 (continued). $X_1 \sim bin(5,\theta)$. H_0: $\theta = 0.5$, H_1: $\theta = 0.4$. Ideally, size of ϕ is α for the optimal test ϕ. Let $\alpha = 0.05$, is there a size α LRT ?

$\alpha = E(\phi) = P(X_1 \le c)$.

$c =$	-0.1	0	1	2	3	4	5	6
$\alpha =$	0	$\frac{1}{32}$	$\frac{6}{32}$	*larger*				
$\approx$	0	0.031	0.187	0.500	0.812	0.969	1	1

Answer: No size 0.05 LRT.

Alternative ?

$$\phi_o = \begin{cases} 1 & \text{if } X < 1 \\ \frac{\alpha - \frac{1}{32}}{5/32} & \text{if } X = 1 \\ 0 & \text{if } X > 1. \end{cases}$$

if $\theta = 0.5$ (under Θ_o), $E(\phi_o) = P(X = 0) + \frac{\alpha - \frac{1}{32}}{5/32} P(X = 1) = \alpha$.

Notice that $\phi = \mathbf{1}(\lambda < a)$ is a LRT, where $a = \frac{f(1;\hat\theta_o)}{f(1;\hat\theta)}$. We often write LRT $\phi = \mathbf{1}(\lambda \le c)$, $c = ?$

Is ϕ_o a LRT ?

$\phi_o = \mathbf{1}(\lambda < a) + \frac{\alpha - \frac{1}{32}}{5/32} \mathbf{1}(\lambda = a)$.

Example 1 (continued). Show that $\phi = \begin{cases} 1 & \text{if } X < 1 \\ \frac{\alpha - \frac{1}{32}}{5/32} & \text{if } X = 1 \\ 0 & \text{if } X > 1 \end{cases}$ is a UMP level α test for testing H_0: $p = 0.5$ against H_1: $p < 0.5$.

Sol. Let $\theta_o = 0.5$, $\theta_1 = p$, $q = 1 - p$ and

$$r = f(\mathbf{x};\theta_1)/f(\mathbf{x};\theta_o) = \frac{\binom{5}{x} p^x q^{5-x}}{\binom{5}{x} (\frac{1}{2})^5} = (2p)^x (2q)^{5-x} = (p/q)^x (2q)^5$$

$$r = (\frac{p}{q})^x (2q)^5 \textbf{ for all } x \textbf{ ?}$$

$\forall$ $p < 1/2$, we have $p/q < 1$, $(p/q)^x (2q)^5 \downarrow$ in x, then

$r > k$ iff $x < c$ for some c;

$r < k$ iff $x > c$.

That is, ϕ satisfies Eq. (2) and (3). Thus ϕ is the UMP level α test for testing

H_0: $p = 0.5$ v.s. H_1: $p = p_1$ ($< 1/2$).

Since this is valid for all p_1, provided $p_1 < 0.5$, and $E(\phi) = \alpha$ at $\theta = 0.5$, ϕ is the UMP level α test for testing H_0: $p = 0.5$ v.s. H_1: $p < 1/2$.

Proof of NPL: Note that $\int_{\mathbf{x}\in A} g(\mathbf{x})dF(\mathbf{x}) = \int_{\mathbf{x}\in A} g(\mathbf{x})f(\mathbf{x})d\mu(\mathbf{x})$

$$= \begin{cases} \sum_{\mathbf{x}\in A} g(\mathbf{x})f(\mathbf{x}) & \text{in discrete case} \\ \int_{\mathbf{x}\in A} g(\mathbf{x})f(\mathbf{x})d\mathbf{x} & \text{in continuous case} \\ \sum_{\mathbf{x}\in A\cap D} g(\mathbf{x})f(\mathbf{x}) + \int_{\mathbf{x}\in A\cap D^c} g(\mathbf{x})f(\mathbf{x})d\mathbf{x} & \text{in mixed case} \end{cases}$$

if the d.f. f exists, where D is the set of discrete points of $\mathbf{X}$.

WLOG, we can assume that $\mathbf{X}$ is continuous.

Sufficiency. Suppose that ϕ satisfies (2) and (3), and ϕ^* is a level α test. Let

$A_+ = \{\mathbf{x} : f(\mathbf{x};\theta_1) > kf(\mathbf{x};\theta_o)\}$; On A_+, $\phi(\mathbf{x}) = 1 \ge \phi^*(\mathbf{x})$;
$A_- = \{\mathbf{x} : f(\mathbf{x};\theta_1) < kf(\mathbf{x};\theta_o)\}$; On A_-, $\phi(\mathbf{x}) = 0 \le \phi^*(\mathbf{x})$;
$A_o = \{\mathbf{x} : f(\mathbf{x};\theta_1) = kf(\mathbf{x};\theta_o)\}$; On A_o, $f(\mathbf{x};\theta_1) - kf(\mathbf{x};\theta_o) = 0$.

$$\begin{aligned}
&\int (\phi(\mathbf{x}) - \phi^*(\mathbf{x}))(f(\mathbf{x};\theta_1) - kf(\mathbf{x};\theta_o))d\mathbf{x} &(4)\\
=&(\int_{A_+} + \int_{A_-} + \int_{A_o})(\phi(\mathbf{x}) - \phi^*(\mathbf{x}))(f(\mathbf{x};\theta_1) - kf(\mathbf{x};\theta_o))d\mathbf{x}\\
=&(\int_{A_+^*} + \int_{A_-^*} + \int_{A_o})(\phi(\mathbf{x}) - \phi^*(\mathbf{x}))(f(\mathbf{x};\theta_1) - kf(\mathbf{x};\theta_o))d\mathbf{x} \textbf{ Why ??}\\
&\qquad (A_+^* = \{\mathbf{x} : f(\mathbf{x};\theta_1) > kf(\mathbf{x};\theta_o),\ \phi(\mathbf{x}) = 1 > \phi^*(\mathbf{x})\}\\
&\qquad \text{and } A_-^* = \{\mathbf{x} : f(\mathbf{x};\theta_1) < kf(\mathbf{x};\theta_o),\ \ \phi(\mathbf{x}) = 0 < \phi^*(\mathbf{x})\})\\
\ge&\int_{A_+^*} (+)(+)d\mathbf{x} + \int_{A_-^*} (-)(-)d\mathbf{x} + \int_{A_o} \underbrace{(-1)}_{\textbf{Why}?}(0)d\mathbf{x} &(5)\\
\ge&0,
\end{aligned}$$

Inequality (5) yields $\beta_\phi(\theta_1) - \beta_{\phi^*}(\theta_1) - k(\beta_\phi(\theta_o) - \beta_{\phi^*}(\theta_o)) \ge 0$. **Why ?**

$$\Rightarrow \quad \beta_\phi(\theta_1) - \beta_{\phi^*}(\theta_1) \ge k(\underbrace{\beta_\phi(\theta_o)}_{=\alpha\ ?} - \underbrace{\beta_{\phi^*}(\theta_o)}_{\le\alpha\ ?}) \ge 0, \tag{6}$$

as ϕ^* is an arbitrary level α test, ϕ satisfies Eq. (3).

$\Rightarrow \quad \beta_\phi(\theta_1) \ge \beta_{\phi^*}(\theta_1)$.

Necessary. If $\exists\ \phi$ satisfying (2) and (3) with $k > 0$, and ϕ^* is a UMP level α test, then Expression (4) and inequality (6) yield

$$\underbrace{0 = \beta_\phi(\theta_1) - \beta_{\phi^*}(\theta_1)}_{\text{as both are UMP}} \ge k(\underbrace{\beta_\phi(\theta_o)}_{=\alpha} - \underbrace{\beta_{\phi^*}(\theta_o)}_{\le\alpha}) \ge 0.$$

$\Rightarrow 0 = \beta_\phi(\theta_1) - \beta_{\phi^*}(\theta_1) = k(\beta_\phi(\theta_o) - \beta_{\phi^*}(\theta_o)) = 0$.
$\Rightarrow \beta_\phi(\theta_o) - \beta_{\phi^*}(\theta_o) = 0$, as $k > 0$.
$\Rightarrow \beta_{\phi^*}(\theta_o) = \beta_\phi(\theta_o) = \alpha$.
Thus ϕ^* satisfies Eq. (3).
Moreover, Inequality (5) yields $0 \ge 0$, with "$>$" iff
either (1) $\int_{A_+^*\cup A_-^*} d\mathbf{x} > 0$ and $k > 0$ or (2) $\int_{A_+^*} d\mathbf{x} > 0$ and $k = 0$.
Since $k > 0$, in order to avoid the contradiction $0 > 0$,
it must be the case that $\int_{A_+^*\cup A_-^*} d\mathbf{x} = 0$.
Thus ϕ^* satisfies (2), except on $A = A_+^* \cup A_-^*$ if $k > 0$. □

Remark. If one deletes the condition "with $k > 0$" in the necessary condition of the NPL, then a UMP level α test ϕ^* must satisfy Eq. (2) w.p.1, but may not satisfies Eq. (3). Then its size $\begin{cases} \le \alpha & ? \\ > \alpha & ? \end{cases}$

Remark. Given α, the LRT $\phi = 1(\frac{L(\mathbf{X};\hat{\theta}_o)}{L(\mathbf{X};\hat{\theta})} \le c)$ (**a non-randomized test**) may not attain the size α, while NP Lemma yields a size α test (Eq. (3)) $\phi(\mathbf{X}) = \begin{cases} 1 & \text{if } f(\mathbf{X};\theta_1)/f(\mathbf{X};\theta_o) > k \\ ? & \text{if } f(\mathbf{X};\theta_1)/f(\mathbf{X};\theta_o) = k \\ 0 & \text{if } f(\mathbf{X};\theta_1)/f(\mathbf{X};\theta_o) < k \end{cases}$ (a **randomized test** Eq. (2)).

In Example 1, we extend the UMP level α test for simple hypotheses to composite hypotheses, that is, Θ contains more than 2 elements. But it is not convenient to check each time. We need a general tool.

Definition. A family of d.f. $\{g(t;\theta): \ \theta \in \Theta\}$ for a univariate random variable T with $\Theta \subset \mathcal{R}^1$ has a monotone likelihood ratio (MLR) if either of the following statements is true (defining $\frac{1}{0} \overset{def}{=} \infty$):

(1) $\forall\ \theta_2 > \theta_1$, $\frac{g(t;\theta_2)}{g(t;\theta_1)}$ is $\uparrow$ in t on $\{t: \ g(t;\theta_1) > 0 \text{ or } g(t;\theta_2) > 0\}$.

(2) $\forall\ \theta_2 > \theta_1$, $\frac{g(t;\theta_2)}{g(t;\theta_1)}$ is $\downarrow$ in t on $\{t: \ g(t;\theta_1) > 0 \text{ or } g(t;\theta_2) > 0\}$.

Theorem 2. (Karlin-Rubin). *Consider testing H_0: $\theta \le \theta_o$ v.s. H_1: $\theta > \theta_o$.*
Suppose that (1) T is a sufficient statistic for θ and
(2) the family of d.f. of T has $\uparrow$ MLR.
Then for each c, $\phi = \mathbf{1}(T > c)$ is a UMP level α test with $\alpha = P_{\theta_o}(T > c)$.

Corollary. *Consider testing H_0: $\theta \ge \theta_o$ v.s. H_1: $\theta < \theta_o$.*
Suppose that (1) T is a sufficient statistic for θ and
(2) the family of d.f. of T has $\uparrow$ MLR.
Then for each c, $\phi = \mathbf{1}(T < c)$ is a UMP level α test with $\alpha = P_{\theta_o}(T < c)$.

Proof of Corollary. Let $W = -T$ and $\gamma = -\theta$. Then it becomes
H_0^*: $\gamma \le \gamma_0$ v.s. H_1^*: $\gamma > \gamma_0$.

The family of the df of W has $\uparrow$ MLR in γ.

Thus $\forall\ w$, $\phi = \mathbf{1}(W > w)$ is a UMP level α test with $\alpha = P_{\gamma_0}(W > w)$,
i.e., $\forall\ c = -w$, $\phi = \mathbf{1}(T < c)$ is a UMP level α test with $\alpha = P_{\theta_o}(T < c)$.

Example 2. Let X_1, ..., X_n be a random sample from $N(\mu, \sigma^2)$, where σ is known. H_0: $\mu \le \mu_0$ v.s. H_1: $\mu > \mu_0$. A UMP level 0.05 test ?

Sol. $\overline{X}$ is a sufficient statistic for μ $(= \theta)$ and
$T = \overline{X} \sim N(\mu, \sigma^2/n)$ with d.f. g.
$\frac{g(t;\theta_2)}{g(t;\theta_1)} = \exp(-\frac{1}{2}\frac{(t-\mu_2)^2}{\sigma^2/n} + \frac{1}{2}\frac{(t-\mu_1)^2}{\sigma^2/n}) = \exp(\frac{1}{2}(\mu_2 - \mu_1)\frac{(2t-(\mu_2+\mu_1))}{\sigma^2/n})$ $\uparrow$ in t.
Thus the family of d.f. of T has $\uparrow$ MLR.

$\phi = \mathbf{1}(T > c)$ is a UMP level α test,
where $\alpha = P_{\theta_o}(T > c) = 1 - \Phi(\frac{c-\mu_0}{\sigma/\sqrt{n}}) = 1 - \Phi(1.645)$,
that is, $c = \mu_0 + 1.645\sigma/\sqrt{n}$.

Remark. Let $X \sim N(\mu, 0.01)$, H_a: $\mu = 1$, H_b: $\mu = -1$.
The UMP level 0.05 test for testing
$$\begin{cases} H_0 = H_a \text{ against } H_b \text{ is } \mathbf{1}(X < 1 - 0.1 \cdot 1.65)\ (= \mathbf{1}(X < 0.835)). \\ H_0 = H_b \text{ against } H_a \text{ is } \mathbf{1}(X > -1 + 0.1 \cdot 1.65)\ (= \mathbf{1}(X > -0.835)). \end{cases}$$

Remark. In general, given a parameter θ, let $\hat{\theta}$ be its MLE, then for testing H_0: $\theta = \theta_o$ v.s. H_1: $\theta > \theta_o$, $\phi = \mathbf{1}(\hat{\theta} > c)$ is a reasonable test for c given.

Question: Are they UMP level α test ?

Answer: Not necessary !

Results: Let μ be the mean of a distribution.
For testing H_0: $\mu = \mu_o$, v.s. H_1: $\mu > \mu_o$,
$\phi = \mathbf{1}(X > c)$ is UMP level α test,
 if $X \sim N(\mu, 1)$;
 if $X \sim Poisson(\mu)$;
 if $nX \sim bin(n, \mu)$;
 if $X \sim f(x) = \frac{1}{\mu}e^{-x/\mu}$, $x > 0$;
etc, as they belong to the families that have MLR.
For testing H_0: $\mu = \mu_o$, v.s. H_1: $\mu < \mu_o$,
 $\phi = \mathbf{1}(X < c)$ is UMP level α test if ... **?**

Remark. Let $X_1, ..., X_n$ be i.i.d from $N(\mu, \sigma^2)$. for test H_0: $\mu = \mu_o$, v.s. H_1: $\mu > \mu_o$.
 (1) Is $\phi = \mathbf{1}(\overline{X} > c)$ a UMP level α test if σ is known ?

(2) Is $\phi = \mathbf{1}(\overline{X} > c)$ a UMP level α test if σ is unknown ?

(3) Is $\phi^* = \mathbf{1}(T > c)$ a UMP level α test if σ is unknown and $T = \frac{\overline{X}-\mu_o}{S/\sqrt{n}}$?

Method: (a) check the size of ϕ, (b) try KR Th or NP Lemma.

Ans to (1): $\phi = \mathbf{1}(\overline{X} > c)$ is UMP level α test if σ is known, and if $c = \mu_0 + z_\alpha \sigma/\sqrt{n}$.

Reason: $f_{\overline{X}}(t; \mu)$ has MLR and $\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$.

Ans to (2): $\phi = \mathbf{1}(\overline{X} > c)$ is **not** UMP level α test if σ is unknown.

Reason: (a) Given finite c, size of ϕ ?

$size = E_{\mu_o}(\phi)$?

or $= \sup_{\theta \in \Theta_o} E(\phi)$? $\Theta_o = ?$

size $= \sup_{\theta \in \Theta_o} P(\frac{\overline{X}-\mu_o}{\sigma/\sqrt{n}} > \frac{c-\mu_o}{\sigma/\sqrt{n}}) =$? Why ϕ is not UMP level α test ?

(3) Question (continued) Let $T = \frac{\overline{X}-\mu_o}{S/\sqrt{n}}$ and $c = t_{n-1,\alpha}$.

Is $\phi^* = \mathbf{1}(T > c)$ is a UMP level α test ?

Answer: No.

Reason : (a) Given finite $c = t_{n-1,\alpha}$, size of ϕ^* ?

(a) size of $\phi^* = E_{\mu_o}(\phi^*) = \alpha$? or size of $\phi^* = sup_{\sigma>0} E_{\mu_o,\sigma}(\phi^*) = \alpha$?

$\alpha = \sup_{\theta \in \Theta_o} E(\phi^*)$, (where $\theta = (\mu, \sigma)$, $\Theta_o = \{(\mu_o, \sigma) : \sigma > 0\}$) **So the reason is not due to the size.**

(b) Why is it not UMP ? Try KR theorem: The sufficient statistic is $(\overline{X}, S^2)$ if $\theta = (\mu, \sigma^2)$ unknown.

$T = \frac{\overline{X}-\mu_o}{S/\sqrt{n}}$ is a statistic, but it is not sufficient.

$T \not\sim t_{n-1}$ if $\mu \neq \mu_o$

$Y = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$, but it is not a statistic if $\mu \neq \mu_o$.

So Karlin-Rubin Theorem does not work. Try NPL.

By NP Lemma, if H_0: $\theta = \theta_0$ v.s. H_1: $\theta = \theta_1$,

where $\theta_0 = (\mu_0, \sigma_0)$ and $\theta_1 = (\mu_1, \sigma_1)$, then a UMP test satisfies

$E_{\theta_o}(\phi) = \alpha$ and $\phi = \mathbf{1}(f_{\mathbf{X}}(\mathbf{x}; \theta_1) > k f_{\mathbf{X}}(\mathbf{x}; \theta_0)) + \mathbf{1}(f_{\mathbf{X}}(\mathbf{x}; \theta_1) = k f_{\mathbf{X}}(\mathbf{x}; \theta_0)\phi$ **???**

$= \mathbf{1}(f_{\overline{X},S^2}(y, t; \theta_1) > k f_{\overline{X},S^2}(y, t; \theta_0))$ **???**

$= \mathbf{1}(\frac{f_{\overline{X},S^2}(u,v;\theta_1)}{f_{\overline{X},S^2}(u,v;\theta_0)} > k)$

If $\phi = 1(T > c)$ is the UMP, it would lead to a contradiction, by selecting different θ.

Remark. $\overline{X} \perp S^2$ and $f_{\overline{X},S^2} = f_{\overline{X}} f_{S^2}$.

$\overline{X} \sim$? $(n-1)S^2/\sigma^2 \sim$?

$S^2 = \frac{\sigma^2}{n-1}\chi^2_{n-1}$ with

$$f_{S^2}(t; \sigma) \propto \frac{1}{\sigma^2/(n-1)} (\frac{t}{\sigma^2/(n-1)})^{\frac{n-1}{2}-1} \exp(-(\frac{t}{\sigma^2/(n-1)})/2), \ t > 0.$$

$$f_{\overline{X}}(y; \mu, \sigma) \propto \frac{1}{\sqrt{2\pi\sigma^2/n}} \exp(-\frac{n}{2}(\frac{y-\mu}{\sigma})^2)$$

$$\begin{aligned} r =& \frac{f_{\overline{X},S^2}(y, t; \mu_1, \sigma_1)}{f_{\overline{X},S^2}(y, t; \mu_0, \sigma_0)} \\ =& \frac{f_{\overline{X}}(y; \mu_1, \sigma)}{f_{\overline{X}}(y; \mu_0, \sigma)} \text{ if } \sigma_1 = \sigma_0 = \sigma. \\ =& \frac{\exp(-\frac{n}{2}(\frac{y-\mu_1}{\sigma})^2)}{\exp(-\frac{n}{2}(\frac{y-\mu_0}{\sigma})^2)} \\ =& \exp(\frac{n}{2\sigma^2}(2y - (\mu_1 - \mu_0))(\mu_1 - \mu_0)) \\ =& \exp(\frac{(y - \frac{\mu_1-\mu_0}{2})(\mu_1 - \mu_0)}{\sigma/\sqrt{n}})) \end{aligned}$$

If $\sigma = 1$, then $\phi = \mathbf{1}(r > k) = \mathbf{1}(\overline{X} > c) = \mathbf{1}(\overline{X} > \mu_0 + z_\alpha * 1/\sqrt{n})$ is a UMP level α test.

If $\sigma = 2$, then $\phi = \mathbf{1}(r > k) = \mathbf{1}(\overline{X} > c) = \mathbf{1}(\overline{X} > \mu_0 + z_\alpha * 2/\sqrt{n})$ is a UMP level α test.

By letting $\sigma = 1$ and $\sigma = 2$, the NP lemma leads to two different UMP level α tests, with RR_1 and RR_2 satisfying $P((RR_1 \setminus RR_2) \cup (RR_2 \setminus RR_1)) > 0$. Thus there exists no UMP level α test for H_0: $\mu = \mu_0$, v.s. H_1: $\mu > \mu_0$ if σ is unknown. ϕ^* is not UMP level α test.
Example 3. If $X \sim N(\mu, 1)$, then $\nexists$ the UMP level α test for testing
H_0: $\mu = 0$, v.s. H_1: $\mu \neq 0$.

What is the difference between these two set-ups ?
Proof. Suppose that such UMP level α test exists and is ϕ.
Then for testing H_0^*: $\mu = 0$ v.s. H_1^*: $\mu = 1$.
$\phi = \mathbf{1}(X > z_\alpha)$ w.p.1 by NPL or Example 2.
For testing H_0^*: $\mu = 0$ v.s. H_1^*: $\mu = -1$
$\phi = \mathbf{1}(X < -z_\alpha)$ w.p.1 by NPL or Example 2.
Since $\{X < -z_\alpha\} \cap \{X > z_\alpha\} = \emptyset$,
$0 = P(\mathbf{1}(X > z_\alpha) = \mathbf{1}(X < -z_\alpha)) = 1$ by NPL.
The contradiction indicates that the UMP level α does not exist. $\square$
Example 3 (continued). If $X \sim N(\mu, 1)$, H_0: $\mu = 0$, v.s. H_1: $\mu \neq 0$. Show that the LRT is an unbiased test.
Q: What should the LRT $\mathbf{1}(\lambda \leq c)$ look like ?
$\mathbf{1}(|\hat{\mu}| \geq a)$ (see Example 1 in §8.2).
Proof. $\vdash$: the LRT test $\phi = \mathbf{1}(|X| > z_{\alpha/2})$, where $\Phi(z_{\alpha/2}) = 1 - \alpha/2$.
Reason: $\hat{\mu} = X$, $\hat{\mu}_o = 0$,

$$\lambda = \exp(-\frac{1}{2}(X-0)^2)/\exp(-\frac{1}{2}(X-X)^2) = \exp(-\frac{1}{2}(X-0)^2) \leq c$$

$$<=> \ |X| \geq z_{\alpha/2}$$

Recall that ϕ is unbiased if $\beta_\phi(\theta_1) \geq \beta_\phi(\theta_o)$ $\forall$ $\theta_1 \in \Theta_o^c$ and $\theta_o \in \Theta_o$. $\alpha = \sup_{\theta \in \Theta_o} E_\theta(\phi)$.
$\vdash$: ϕ is an unbiased test.
Reason: It suffices to show that

$$\begin{cases} \frac{\partial}{\partial \mu}\beta_\phi(\mu) > 0 & \text{if } \mu > 0 \\ \frac{\partial}{\partial \mu}\beta_\phi(\mu) < 0 & \text{if } \mu < 0. \end{cases} \textbf{Why??} \qquad (1)$$

$$\begin{aligned} \beta_\phi(\mu) =& P_\mu(|X| > z_{\alpha/2}) \\ =& 1 - P_\mu(-z_{\alpha/2} \leq X \leq z_{\alpha/2}) \\ =& 1 - P_\mu(-z_{\alpha/2} - \mu \leq X - \mu \leq z_{\alpha/2} - \mu) \\ =& 1 - \int_{-z_{\alpha/2}-\mu}^{z_{\alpha/2}-\mu} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \\ \frac{\partial}{\partial \mu}\beta_\phi(\mu) =& \frac{1}{\sqrt{2\pi}}[\exp(-\frac{1}{2}(z_{\alpha/2}-\mu)^2) - \exp(-\frac{1}{2}(-z_{\alpha/2}-\mu)^2)] \end{aligned}$$

Assume $\mu > 0$. Then
Eq. (1) $<=>$ $\frac{1}{\sqrt{2\pi}}[\exp(-\frac{1}{2}(z_{\alpha/2}-\mu)^2) - \exp(-\frac{1}{2}(-z_{\alpha/2}-\mu)^2)] > 0$; $<=>$ $\exp(-\frac{1}{2}(z_{\alpha/2}-\mu)^2) > \exp(-\frac{1}{2}(-z_{\alpha/2}-\mu)^2)$;
$<=>$ $-\frac{1}{2}(z_{\alpha/2}-\mu)^2 > -\frac{1}{2}(-z_{\alpha/2}-\mu)^2$; $<=>$ $(z_{\alpha/2}-\mu)^2 < (-z_{\alpha/2}-\mu)^2$;
$<=>$ $(z_{\alpha/2}-\mu)^2 - (-z_{\alpha/2}-\mu)^2 < 0$;
$<=>$ $-2\mu(2z_{\alpha/2}) < 0$ (which always holds).
Thus if $\mu > 0$, then $\frac{\partial}{\partial \mu}\beta_\phi(\mu) > 0$.
The proof of Eq.(1) for $\mu < 0$ is similar and is skipped.
As a consequence, ϕ is unbiased. $\square$
Remark. The LRT test in Example 3 is actually a UMP unbiased test. The proof is given in Lehmann's textbook "Testing Statistical Hypotheses".
Theorem 2. (Karlin-Rubin). *Consider testing H_0: $\theta \leq \theta_o$ v.s. H_1: $\theta > \theta_o$.*
Suppose that (1) T is a sufficient statistic for θ and

(2) the family of d.f. of T has $\uparrow$ MLR.
Then for each c, $\phi = \mathbf{1}(T > c)$ is a UMP level α test with $\alpha = P_{\theta_o}(T > c)$.
Proof of KR theorem. Since T is sufficient for θ, $f_{\mathbf{X}}(\mathbf{x};\theta) = g(T(\mathbf{x});\theta)h(\mathbf{x})$. Let H_0^*: $\theta = \theta_1$ v.s. H_1^*: $\theta = \theta_2$, where $\theta_1 \leq \theta_o < \theta_2$. By the NPL, the UMP level α^* test satisfies

$$\phi = \begin{cases} 1 & \text{if } \frac{f_{\mathbf{X}}(\mathbf{X};\theta_2)}{f_{\mathbf{X}}(\mathbf{X};\theta_1)} > k \\ 0 & \text{if } \frac{f_{\mathbf{X}}(\mathbf{X};\theta_2)}{f_{\mathbf{X}}(\mathbf{X};\theta_1)} < k \end{cases} \text{ with } \alpha^* = E_{\theta_1}(\phi).$$

Claim: $f_T(t;\theta) = g(t;\theta)c(t)$, where c is a function of t.
Reason: 3 cases: (1) Discrete. (2) Continuous. (3) mixed distribution.

$$\begin{aligned} f_T(t;\theta) = & \int_{\mathbf{X}:\ T(\mathbf{X})=t} f_{\mathbf{X}}(\mathbf{x};\theta)d\mu(\mathbf{x}) \\ & (= \underbrace{\int \cdots \int_{T(\mathbf{X})=t}}_{n-1} f_{\mathbf{X}}(\mathbf{x};\theta)|J|dx_{i_2}\cdots dx_{i_n} + \sum_{\mathbf{X}\in D:\ T(\mathbf{X})=t} f_{\mathbf{X}}(\mathbf{x};\theta) \\ & \quad t_1 = T(\mathbf{x}), t_2 = x_{i_2}, ..., t_n = x_{i_n},\ J \text{ is Jacobian}) \\ = & \int_{\mathbf{X}:\ T(\mathbf{X})=t} g(T(\mathbf{x});\theta)h(\mathbf{x})d\mu(\mathbf{x}) \\ = & g(t;\theta)\int_{\mathbf{X}:\ T(\mathbf{X})=t} h(\mathbf{x})d\mu(\mathbf{x}) \\ = & g(t;\theta)c(t) \end{aligned}$$

$$\frac{f_{\mathbf{X}}(\mathbf{x};\theta_2)}{f_{\mathbf{X}}(\mathbf{x};\theta_1)} = \frac{g(T(\mathbf{x});\theta_2)h(\mathbf{x})}{g(T(\mathbf{x});\theta_1)h(\mathbf{x})} = \frac{g(t;\theta_2)c(t)}{g(t;\theta_1)c(t)} = \frac{f_T(t;\theta_2)}{f_T(t;\theta_1)}, \quad t = T(\mathbf{x}),$$

$$\phi = \begin{cases} 1 & \text{if } \frac{f_{\mathbf{X}}(\mathbf{X};\theta_2)}{f_{\mathbf{X}}(\mathbf{X};\theta_1)} > k \\ 0 & \text{if } \frac{f_{\mathbf{X}}(\mathbf{X};\theta_2)}{f_{\mathbf{X}}(\mathbf{X};\theta_1)} < k \end{cases} = \begin{cases} 1 & \text{if } \frac{f_T(t;\theta_2)}{f_T(t;\theta_1)} > k \\ 0 & \text{if } \frac{f_T(t;\theta_2)}{f_T(t;\theta_1)} < k \end{cases} = \begin{cases} 1 & \text{if } t > c \\ 0 & \text{if } t < c \end{cases}$$

as $\frac{f_T(t;\theta_2)}{f_T(t;\theta_1)} \uparrow$ in t. =>
$\phi = \mathbf{1}(T > c)$ is a level α^* UMP test for testing H_0^* vs H_1^*, where $\alpha^* = E_{\theta_1}(\phi)$.
In fact, it is true for each H_0^{**}: $\theta = \theta_1$ v.s. H_1^{**}, $\theta > \theta_1$.
$\vdash$: $\beta_\phi(\theta) \geq \beta_\phi(\theta_1)\ \forall\ \theta > \theta_1$.
Reason: Verify that (1) $\phi^* = \alpha^*$ is a level α^* test; (2) $E_\theta(\phi^*) = \alpha^*\ \forall\ \theta$;
(3) $\beta_\phi(\theta) \geq \beta_{\phi^*}(\theta) = \alpha^* = \beta_\phi(\theta_1)\ \ \forall\ \ \theta > \theta_1$ **Why ??**
(NPL). Thus $\beta_\phi(\theta) \uparrow$ in θ. It follows that $\beta_\phi(\theta_1) \leq \beta_\phi(\theta_o)$ as $\theta_1 \leq \theta_o < \theta_2$, and the size of ϕ is $\alpha = \sup_{\theta \leq \theta_o} \beta_\phi(\theta) = \beta_\phi(\theta_o)$. □

Chapter 9. Interval Estimation

There are 3 statistical inferences:
(1) point estimation $\theta = ?$ MLE, MME, Bayes estimator.
(2) test $\theta = \theta_o$? LRT, NPL.
(3) interval estimation: a likely interval $[L, U]$ for θ ?

Undergradaute statistics: 95% Confidence interval for μ is
$\overline{X} \pm 1.96\sigma/\sqrt{n}$ or
$\overline{X} \pm t_{n-1,0.025}S/\sqrt{n}$
Definition. Let $\mathbf{X} \sim f(\mathbf{x};\theta)$, $L(\mathbf{X})$ and $U(\mathbf{X})$ be two statistics such that $L(\mathbf{X}) \leq U(\mathbf{X})$. Then the interval $[L(\mathbf{X}), U(\mathbf{X})]$ or $[L, U]$ is called an interval estimator (IE) of θ;
$P_\theta(\theta \in [L, U])$ is called the coverage probability of $[L, U]$;
$1 - \alpha \stackrel{def}{=} \inf_\theta P_\theta(\theta \in [L, U])$ is called the confidence coefficient of $[L, U]$;

$[L, U]$ is also called a $1-\alpha$ confidence interval (CI) of θ.

Example 1. If $X_1, ..., X_n$ are i.i.d.$\sim N(\mu, \sigma^2)$ with σ known, a CI of μ is $[\overline{X} - 1.96\sigma/\sqrt{n}, \overline{X} + 1.96\sigma/\sqrt{n}]$ or $\overline{X} \pm 1.96\sigma/\sqrt{n}$. Its coverage probability ? Its confidence coefficient ?

Sol. Some particular quantiles values z_α of $N(0,1)$.

```
> a=c(0.05,0.025,0.01,0.005)
> round(qnorm(1-a),2)
[1] 1.64 1.96 2.33 2.58
> round(qnorm(1-a),3)
[1] 1.645 1.960 2.326 2.576
```

Coverage probability $= P_\mu(\mu \in [\overline{X} - 1.96\sigma/\sqrt{n}, \overline{X} + 1.96\sigma/\sqrt{n}\ ]) = P_\mu(\overline{X} - 1.96\sigma/\sqrt{n} \leq \mu \leq \overline{X} + 1.96\sigma/\sqrt{n}) = P_\mu(-1.96\sigma/\sqrt{n} \leq \mu - \overline{X} \leq 1.96\sigma/\sqrt{n}) = P_\mu(-1.96\sigma/\sqrt{n} \leq \overline{X} - \mu \leq 1.96\sigma/\sqrt{n},)$

$= P_\mu(-1.96 \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq 1.96) \approx 0.95$

Confidence coefficient $= \inf_\mu P_\mu(\mu \in [\overline{X} - 1.96\sigma/\sqrt{n}, \overline{X} + 1.96\sigma/\sqrt{n}\ ]) \approx \inf_\mu 0.95 = 0.95$.

Example 2. If $X_1, ..., X_n$ are i.i.d.$\sim N(\mu, \sigma^2)$, where (μ, σ) is unknown. A 95% CI of μ is $[\overline{X} - t_{n-1,0.025}S/\sqrt{n}, \overline{X} + t_{n-1,0.025}S/\sqrt{n}]$ or $\overline{X} \pm t_{n-1,0.025}S/\sqrt{n}$.

Coverage probability and Confidence coefficient ?

Solution: In class exercise. **Hint:**

$\overline{X} \pm t_{n-1,0.025}\frac{S}{\sqrt{n}}$ now. $\overline{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$ in Example 1.

A simulation example in R:

```
> x=rnorm(20)
> round(x,2)
[1] -1.39 -1.55 1.00 0.29 -0.17 -1.50 -1.65 0.33 -0.07 2.24 -0.62 0.32
[13] -0.54 -0.11 -0.17 -0.23 -0.21 1.47 0.77 -0.26
> t.test(x)
t = -0.4541, df = 19, p-value = 0.6549
95 percent confidence interval:
-0.5708114 0.3672875
mean of x -0.1017619
```

Example 3. Let X_1, ..., X_n be i.i.d. from $U(0, \theta)$. Then 2 IEs of θ are (a) $[X_{(n)}, 2X_{(n)}]$, and (b) $[X_{(n)}, X_{(n)} + 1/n]$. What are their coverage probabilities and confidence coefficients ?

Sol. Recall that the cdf of $X_{(n)}$ is $F_{X_{(n)}}(t) = P(X_{(n)} \leq t) = (F_X(t))^n$ **= ?**

(a) $P_\theta(\theta \in [X_{(n)}, 2X_{(n)}])$ (= coverage prob.)

$= P_\theta(X_{(n)} \leq \theta \leq 2X_{(n)})$

$= P_\theta(\theta/2 \leq X_{(n)} \leq \theta)$

$= P_\theta(X_{(n)} \leq \theta) - P_\theta(X_{(n)} < \theta/2) \qquad (F(b) - F(a-))$

$= (\theta/\theta)^n - (\frac{\theta/2}{\theta})^n = 1 - (1/2)^n$.

Coverage probability $= P_\theta(\theta \in [X_{(n)}, 2X_{(n)}]) = 1 - (1/2)^n$ (**independent of** θ).

Confidence coefficient $= \inf_\theta\{1 - (1/2)^n\} = 1 - (1/2)^n$.

(b) $P_\theta(\theta \in [X_{(n)}, X_{(n)} + 1/n])$ (= coverage prob.)

$= P_\theta(X_{(n)} \leq \theta \leq X_{(n)} + 1/n)$

$= P_\theta(\theta - 1/n \leq X_{(n)} \leq \theta)$

$= \begin{cases} (\theta/\theta)^n - (\frac{\theta-1/n}{\theta})^n & \text{if } \theta > 1/n \\ (\theta/\theta)^n - 0 & \text{otherwise} \end{cases}$ **why ?**

$= \begin{cases} 1 - (1 - \frac{1}{n\theta})^n & \text{if } \theta > 1/n \\ 1 & \text{otherwise} \end{cases}$

=Coverage probability **depending on** θ)

Confidence coefficient $= \inf_{\theta \geq 1/n}\{1 - (1 - \frac{1}{n\theta})^n\} = 0$

Question: How to construct a CI ?

Answer: Two methods:

(1) Acceptance interval of LRT,

(2) Pivotal Method.

1. **Acceptance interval of LRT.**

Let $RR(\theta_o)$ be the rejection region for testing
H_0: $\theta = \theta_o$ vs. H_1: $\theta \neq \theta_o$.

$$\text{Then a CI of } \theta \text{ is } \Theta \setminus RR(\theta). \quad (S1)$$

Example 2. Suppose that $X_1, ..., X_{100}$ are i.i.d.$\sim N(\mu, \sigma^2)$ with σ unknown. The sample results in $\overline{X} = 2$ and $S^2 = 1$. Construct a 95% CI for μ.
Sol. Set $\theta = \mu$. Then H_0: $\theta = \theta_o$ vs. H_1: $\theta \neq \theta_o$.
The LRT is $\phi = \mathbf{1}(|\frac{\overline{X}-\mu_0}{S/\sqrt{n}}| \geq t_{n-1,\alpha/2})$.
The RR is $\{(X_1, ..., X_n) : |\frac{\overline{X}-\mu_0}{S/\sqrt{n}}| \geq t_{n-1,\alpha/2}\}$.
The acceptance region is $\{(X_1, ..., X_n) : |\frac{\overline{X}-\mu_0}{S/\sqrt{n}}| < t_{n-1,\alpha/2}\}$.
Replacing μ_0 by μ results in the $(1-\alpha)$ CI of μ:

$$\{\mu : |\frac{\mu - \overline{X}}{S/\sqrt{n}}| \leq t_{n-1,\alpha/2}\} \ \ (= R^1 \setminus \{\mu : |\frac{\overline{X} - \mu}{S/\sqrt{n}}| > t_{n-1,\alpha/2}\}) \text{ (see (S1))}\textbf{Any problem ?}$$

Simplify it as in Ex 2: $\overline{X} - t_{n-1,\alpha/2}S/\sqrt{n} \leq \mu \leq \overline{X} + t_{n-1,\alpha/2}S/\sqrt{n}$.
$[2 - 0.196, 2 + 0.196]$ or 2 ± 0.196.
2. **Pivotal method.** Let $\mathbf{X} \sim f(\mathbf{x}; \theta)$ and $T = h(\mathbf{X}, \theta)$ be a pivotal rv,
i.e, its density f_T or cdf F_T does not depend on θ.
Derive a CI from $P(a \leq h(\mathbf{X}, \theta) \leq b) = 1 - \alpha$.
Example 2 (continued). Since $X \sim N(\mu, \sigma^2)$, Derive the CI by the pivotal method.
Sol. Notice $T = \frac{\overline{X}-\mu}{S/\sqrt{n}} = h(X_1, ..., X_n, \mu, \sigma)$ is a pivotal r.v., with t_{n-1} distribution and f_T does not depend on (μ, σ).
$P(a \leq T \leq b) = 1 - \alpha = P(-t_{n-1,\alpha/2} \leq T \leq t_{n-1,\alpha/2})$.
$-t_{n-1,\alpha/2} \leq \frac{\overline{X}-\mu}{S/\sqrt{n}} \leq t_{n-1,\alpha/2}$.
$\overline{X} - t_{n-1,\alpha/2}S/\sqrt{n} \leq \mu \leq \overline{X} + t_{n-1,\alpha/2}S/\sqrt{n}$.
Example 3. Let X_1, ..., X_9 be i.i.d. $\sim U(0, \theta)$ and $X_{(9)} = 3$. Construct a 95% CI for θ.
Sol. Recall that $X_{(n)} \sim F_{X_{(n)}}(x) = (x/\theta)^n \mathbf{1}(x \in (0, \theta)) + \mathbf{1}(x \geq \theta)$. Let $T = X_{(n)}/\theta$. Then $F_T(t) = t^n \mathbf{1}(t \in (0, 1)) + \mathbf{1}(t \geq 1)$, T **is pivotal**.
$P(a \leq T \leq b) = 0.95$ yields $a \leq \frac{X_{(n)}}{\theta} \leq b$ or

$$\frac{X_{(n)}}{b} \leq \theta \leq \frac{X_{(n)}}{a}.$$

Q: $(a, b) =$ **??**
$F_T(b) - F_T(a) = b^n - a^n = 0.975 - 0.025$.
$a^n = 0.025$ yields $a = (0.025)^{1/n}$, and
$b^n = 0.975$ yields $b = (0.975)^{1/n}$.
The 95% CI for θ:
$\frac{X_{(n)}}{(0.975)^{1/n}} \leq \theta \leq \frac{X_{(n)}}{(0.025)^{1/n}}$
or $[\frac{3}{(0.975)^{1/9}}, \frac{3}{(0.025)^{1/9}}]$ $(= [3.01, 4.52])$.
Question: Why choose $F(b) - F(a) = b^n - a^n = 0.975 - 0.025$?
Answer: Symmetry,
but it is even better to choose $F(b) - F(a) = b^n - a^n = 1 - 0.05$.
It results $[3, 3/0.05^{1/9}]$ or $[3, 4.18]$, which is the shortest.
This is due to the following results: $[a, b]$ is the shortest $1 - \alpha$ CI if (1) the density $f_T(t)$ $(= nt^{n-1}\mathbf{1}(t \in [0, 1])$ is unimodal, (2) $f_T(a) = f_T(b)$ and (3) $\int_a^b f_T(t)dt = 1 - \alpha$.
Here f_T can be defined arbitrary at $t = 0$ or 1, since it is on the boundary.
Remark.
1. In the statement $P_\theta(\theta \in [L, U])$, θ is not a random variable,
but L and U are.
$P_\theta(\theta \in [L, U]) = P_\theta(L \leq \theta \text{ and } U \geq \theta)$.

2. Under the same model such as in Example 1 with σ unknown,
   the CI $\overline{X} \pm t_{n-1,\alpha/2}S/\sqrt{n}$ changes each time and
   even its length $U - L$ $(= 2t_{n-1,\alpha/2}S/\sqrt{n})$ may change each time **Why ?**
3. For each parameter θ,
   $(-\infty, \infty)$ is always a 100% CI for θ, but it is useless.
   $[\hat{\theta}, \hat{\theta}]$ is often a 0% CI for θ,
   an interval that it is unlikely that θ is covered inside.
4. For the same confidence level and the same sample,
   we prefer a CI that is shorter,
   as it provides more accurate information about θ.
5. If $g(\theta)$ is a monotone function of θ *e.g.*, $F(t;\theta)$, then the confidence interval of $g(\theta)$ can be derived directly or by the CI $[L, U]$ of θ, say $[g(L), g(U)]$ or $[g(U), g(L)]$ **why ?** whenever it is appropriate.

Example 4: Suppose that $S(x) = 1 - F_X(x) = e^{-x/\theta}$, $x > 0$. $X = 2$ is observed. Derive a 95% CI for θ and $S(1)$.
Sol. Pivotal method: Find $T = h(\mathbf{X}, \theta)$ such that F_T does not depend on θ. Then derive a CI from $P(a \leq T \leq b)$.
Let $T = X/\theta$, **why ?** then $F_T(t) = P(X/\theta \leq t) = P(X \leq \theta t)$,
$F_T(t) = 1 - e^{-t}$ if $t > 0$.
which is a pivotal.
$P(a \leq T \leq b) = 0.95 = e^{-a} - e^{-b}$ yields $a \leq \frac{X}{\theta} \leq b$, or

$$\frac{X}{b} \leq \theta \leq \frac{X}{a}.$$

Take $e^{-a} - e^{-b} = 1 - 0.05 = 0.95$, it yields $(a, b) = -(\ln 1, \ln 0.05)$.
CI for θ: $\theta \geq X/\ln 20$ or $[2/\ln 20, \infty)$.
Take $0.975 - 0.025 = 0.95$, it yields $(a, b) = -(\ln 0.975, \ln 0.025)$.
CI for θ: $[0.27X, 39.5X]$ or $[0.54, 79]$.
Take $0.95 - 0 = 0.95$, it yields $(a, b) = (-\ln 0.95, \infty)$
CI for θ: $[0, 19.5X]$ or $[0, 39]$.
LRT Method. Acceptance interval of LRT.
Let $RR(\theta_o)$ be the rejection region for testing
H_0: $\theta = \theta_o$ vs. H_1: $\theta \neq \theta_o$.

$$\text{Then a CI of } \theta \text{ is } \Theta \setminus RR(\theta). \tag{S1}$$

Consider testing
H_0: $\theta = \theta_o$ v.s. H_1: $\theta \neq \theta_o$. $\Theta = (0, \infty)$.
Under H_0, the MLE $\hat{\theta}_o = \theta_o$;
Under Θ, the MLE $\hat{\theta} = X$.

$$\lambda = \frac{f_{\hat{\theta}}(X;\theta_o)}{f_{\hat{\theta}}(X;\hat{\theta})} = \begin{cases} 1 & \text{if } X = \theta_o \\ \frac{\frac{1}{\theta_o}e^{-X/\theta_o}}{\frac{1}{\hat{\theta}}e^{-X/\hat{\theta}}} & \text{if } X \neq \theta_o \end{cases} = \begin{cases} 1 & \text{if } X = \theta_o \\ \frac{X}{\theta_o}e^{1-\frac{X}{\theta_o}} & \text{if } X \neq \theta_o \end{cases}$$

$\ln\lambda = \ln X - \ln\theta_o + 1 - X/\theta_o$, $(\ln\lambda)'_x = \frac{1}{x} - \frac{1}{\theta_o}$. $(\ln\lambda)''_x = \frac{-1}{x^2}$.
$\ln\lambda$ is concave down with maximum at $X = \theta_o$.
LRT $\phi = \mathbf{1}(\lambda \leq c) = \mathbf{1}(\lambda(X) \leq c) = \mathbf{1}(X \leq k_1 \text{ or } X \geq k_2)$, where
(a) $\lambda(k_1) = \lambda(k_2)$ and (b) $E_{\theta_o}(\phi) = \alpha$.
RR: $X \notin [k_1, k_2]$.
Acceptance region: $k_1 < X < k_2$, where
(a) $\frac{k_1}{\theta_o}e^{1-\frac{k_1}{\theta_o}} = \frac{k_2}{\theta_o}e^{1-\frac{k_2}{\theta_o}}$ and (b) $e^{-k_1/\theta_o} - e^{-k_2/\theta_o} = 1 - \alpha = 0.95$ **Why ?**
<=> $t_1e^{1-t_1} = t_2e^{1-t_2}$ and $e^{-t_1} - e^{-t_2} = 0.95$ (where $t_i = k_i/\theta_o$).

$$<=> \quad G(t_1) = t_1e^{1-t_1} - t_2e^{1-t_2} = 0 \text{ where } t_2 = -\ln[e^{-t_1} - 0.95] \tag{1}$$

$$<=> \quad g(t_1) = t_1e^{-t_1} - t_2e^{-t_2} = 0 \text{ where } t_2 = -\ln[e^{-t_1} - 0.95]. \tag{2}$$

Then the **acceptance region** $k_1 < X < k_2$ yields $t_1\theta_o < X < t_2\theta_o$.

Replace θ_o by θ: $t_1\theta < X < t_2\theta$.

It yields a 95% CI of θ: $\frac{X}{t_2} < \theta < \frac{X}{t_1}$ **Are we done ?**

There is no closed form solution for Eq. (1) or (2). Solve the equation by R:

x=(1:49)/1000 # probabilities in (0,0.05)

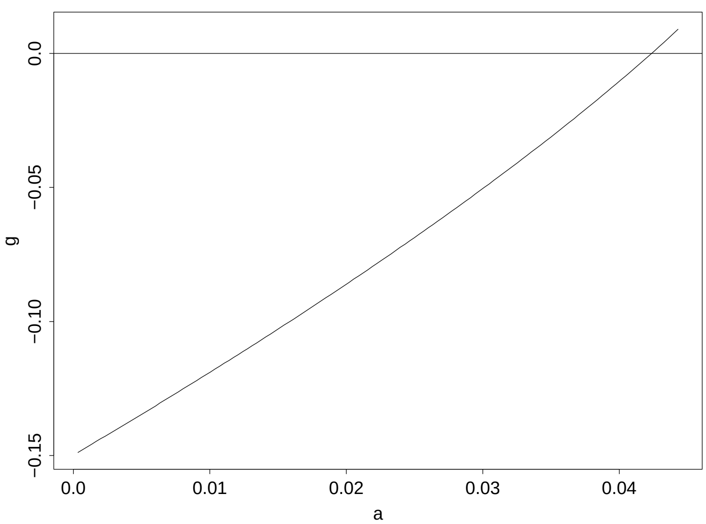
a=qexp(x) # t_1, quantile of Exponential at x

b=-log(exp(-a)-0.95) # t_2,

g=a*exp(-a)-b*exp(-b)

plot(a,g,type="l")

abline(h=0)



max(a[g<=0])

max(b[g<=0])

[1] 0.0418642

[1] 4.710531 It yields a 95% CI of θ: $(2/4.75, 2/0.042)$, or $(0.42, 47.62)$.

¿ ¿ max(b[g¡=0])

4 95% **CI's in this example:** $[0.67, \infty)$, $[0.54, 79]$, $[0, 39]$ or $[0.42, 47.62]$.

Which is better ?

Notice that the density is single-moded.

CI for $S(1) = e^{-1/\theta}$?

CI for $S(1) = e^{-1/\theta}$: $[e^{-1/0}, e^{-1/39}]$ $(= [0, e^{-1/39}])$.

Announcement: The class on April 7, Friday is a seminar about Intership application. 8-9:30am

Example 4. Suppose that $X \sim bin(3,p)$. Compute the confidence coefficient of $I = [\frac{X-1}{3} \vee 0, 1]$.

Sol. Formula: The confidence coefficient $= \inf_p P_p(p \in I)$. $I =$?, $\mathrm{P}_p(p \in I)=$? $\inf_p P_p(p \in I)=$?

$$I = \begin{cases} [2/3, 1] & \text{if } X = 3 \\ [1/3, 1] & \text{if } X = 2 \\ [0, 1] & \text{if } X \le 1 \end{cases}.$$

The coverage probability $P_p(p \in I)$ is a function of p, try $p = 0$, 1, 0.5.

$P_0(0 \in I) = P_0(X \le 1) = \sum_{i \le 1} \binom{3}{i} 0^i (1-0)^{3-i} = 1$. $P_1(1 \in I) = P_1(X \in \{0, 1, 2, 3\})=$? $P_{0.5}(0.5 \in I) = P_{0.5}(X \ne 3) = 1 - (0.5)^3$.

$$I = \begin{cases} [2/3, 1] & \text{if } X = 3 \\ [1/3, 1] & \text{if } X = 2 \\ [0, 1] & \text{if } X \le 1 \end{cases} = \begin{cases} [\frac{2}{3}, 1] & \text{if } X = 3 \\ [\frac{2}{3}, 1] \cup [\frac{1}{3}, \frac{2}{3}) & \text{if } X = 2 \\ [\frac{2}{3}, 1] \cup [\frac{1}{3}, \frac{2}{3}) \cup [0, \frac{1}{3}) & \text{if } X \le 1. \end{cases}$$

$$P_p(p \in I) = \begin{cases} (1-p)^3 + 3p(1-p)^2 & \text{if } p \in [0, 1/3) \ \textbf{Why ??} \\ 1 - p^3 & \text{if } p \in [1/3, 2/3) \\ 1 & \text{if } p \in [2/3, 1] \end{cases}$$

The confidence coefficient is
$\inf_p P_p(p \in I)$.

$$\frac{dP_p(p \in I)}{dp} = \begin{cases} -3(1-p)^2 + 3(1-p)^2 - 6p(1-p) = -6p(1-p) \le 0 & \text{if } p \in (0, 1/3) \\ -3p^2 & \text{if } p \in (1/3, 2/3) \\ 0 & \text{if } p \in (2/3, 1] \end{cases}$$

The confidence coefficient is
$\inf_p P_p(p \in I) = \min\{1, 1 - (2/3)^3, (2/3)^3 + (2/3)^2\} \approx \min\{1, 0.70, 0.74\} = 0.70$

Note. The CI in Example 4 is also called confidence bound, as it is one-sided.

Example 5. Suppose that $X \sim bin(5, p)$ and observed $X = 3$. Construct a 95% CI for p.

Sol. Two methods: (1) Inverting RR of LRT, (2) Pivotal (does not work here).

For simplicity, one may consider a $(1-\alpha)$ CI of form $[L, 1]$. It can be obtained by inverting the acceptance region of the LRT for testing

H_0: $p = p_o$ v.s. H_1: $p > p_o$ **Why not** $p \neq p_o$ **?** Then a non-randomized level α test is $\phi = \mathbf{1}(\underbrace{X > k}_{RR})$, where k satisfies

$$P_{p_o}(X \ge k) > \alpha \text{ and } P_{p_o}(X > k) \le \alpha \ \textbf{Why ?}$$

$$P_{p_o}(X < k) < 1 - \alpha \text{ and } P_{p_o}(\underbrace{X \le k}_{\textbf{what region}?}) \ge 1 - \alpha \ \textbf{Why ?}$$

$$<=> \sum_{i=0}^{k-1} \binom{5}{i} p_o^i (1-p_o)^{5-i} < 1 - \alpha \text{ and } \sum_{i=0}^{k} \binom{5}{i} p_o^i (1-p_o)^{5-i} \ge 1 - \alpha.$$

Acceptance region: $\{x : \ x \le k(p_o)\}$

Replacing p_o by p yields

$$C(x) = \{p : x \le k(p)\} = \{p : p > k^{-1}(x)\}$$

(for proof, see Page 426 in the textbook), where

$$k^{-1}(x) = \sup\{p : \sum_{i=0}^{x-1} \binom{5}{i} p^i (1-p)^{5-i} \ge 1 - \alpha\}.$$

In particular, $k^{-1}(3) = \sup\{p : \ \sum_{i=0}^{3-1} \binom{5}{i} p^i (1-p)^{5-i} \ge 0.95\}$.

Solve by R:

p=(0:1000)/1000
x=3
y=pbinom(x-1,5,p)
max(p[y>=0.95]) # **Why ?**
Reason: The graph of $y \downarrow$ from 1 to 0. **How can tell ?**
[1] 0.189

Thus if $X = 3$ then a 95% CI for p is $(0.189, 1]$ (not exact solution !!)

For a given X, a 95% CI for p is (approximately)

$$I = \begin{cases} [0, 1] & \text{if } X = 0 \\ (0.01, 1] & \text{if } X = 1 \\ (0.076, 1] & \text{if } X = 2 \\ (0.189, 1] & \text{if } X = 3 \\ (0.342, 1] & \text{if } X = 4 \\ (0.549, 1] & \text{if } X = 5 \end{cases}$$

The coverage probability varies in p, for instance, $P_p(p \in I) = (1-p)^5 \in [0.95, 1]$ for $p \in [0, 0.01]$, but the confidence coefficient of the IE is 0.95

A second approach: Observe a value of X, say x $(=3)$. If the p-value $P_{p_o}(X \geq x) > \alpha$, then we do not reject H_o. Thus if $P_p(X \geq x) > \alpha$, then p belongs to the $(1-\alpha)$ CI.
Then the left end $L = \inf\{p: \ P_p(X \geq x) > \alpha\}$, where $P_p(X \geq x) = 1 - P(X \leq x-1)$.
Solve by R

```
x = 3
p=(0:1000)/1000
y=1-pbinom(x-1,5,p)
min(p[y>0.05])
```

Chapter 10. Asymptotic Evaluations

10.1. Point estimation.

Suppose $\mathbf{X}_n = (X_1, ..., X_n) \sim f_{\mathbf{X}}(\mathbf{x}; \theta)$, $\theta \in \Theta$.
$\hat{\theta}$ is an estimator of θ.

Given n, desirable (or "optimal") properties of $\hat{\theta}$ are

1. $E(\hat{\theta}) = \theta$, $\forall\ \theta \in \Theta$ (unbiasedness),
2. UMVUE,
3. Bayes estimator $E(E(L(\theta, \hat{\theta}(\mathbf{X}))|\mathbf{X})) = \inf_\delta E(E(L(\theta, \delta(\mathbf{X}))|\mathbf{X}))$,

These are finite sample properties. We shall consider large sample properties.
Recall in general, for random variables Y_n and X on the sample space Ω.

$Y_n \overset{\cdot}{\to} X \iff \{\omega \in \Omega: \ Y_n(\omega) \to X(\omega)\} = \Omega$,
$Y_n \overset{a.s.}{\to} X \iff P(Y_n \to X) = 1$.
$Y_n \overset{P}{\to} X \iff P(|Y_n - X| < \epsilon) \to 1\ \forall\ \epsilon > 0$
$Y_n \overset{D}{\to} X \iff F_{Y_n}(x) \to F_X(x)$ at each continuity point x of F_X.

Definition. Let $\hat{\theta} = \hat{\theta}_n = W_n(X_1, ..., X_n)$, $n = 1, 2, 3, ...$
$\mathbf{X}_n = (X_1, ..., X_n) \sim f_{\mathbf{X}_n}(\mathbf{x}_n; \theta)$, $\theta \in \Theta$.

If $\hat{\theta}_n \overset{P}{\to} \theta\ \forall\ \theta \in \Theta$, then we say $\hat{\theta}$ is a consistent estimator of θ or $\hat{\theta}$ is consistent.
If $\hat{\theta}_n \overset{a.s.}{\to} \theta\ \forall\ \theta \in \Theta$, then we say $\hat{\theta}$ is a strongly consistent estimator of θ or $\hat{\theta}$ is strongly consistent.
$\hat{\theta}_n \overset{P}{\to} \theta \iff P(\omega \in \Omega: \ |\hat{\theta}_n(\omega) - \theta| \geq \epsilon) \to 0\ \forall\ \epsilon > 0 \iff P(|\hat{\theta}_n - \theta| > \epsilon) \to 0\ \forall\ \epsilon > 0 \iff P(|\hat{\theta}_n - \theta| < \epsilon) \to 1\ \forall\ \epsilon > 0$
$\hat{\theta}_n \overset{a.s.}{\to} \theta \iff P(\omega \in \Omega: \ \hat{\theta}_n(\omega) \to \theta) = 1 \iff \hat{\theta}_n \overset{w.p.1}{\to} \theta$

Remark. Consistency is the most important property of an estimator.
Reason:

Most of the time $\hat{\theta}_n \neq \theta$. $P(\hat{\theta} = \theta) =$ **??**
One can only hope that it is getting close to θ as $n \to \infty$.
Consistency says that this is so if the sample size is large enough.

Example 1. Let X_1, ..., X_n be i.i.d. from $N(\mu, \sigma^2)$. Is $\hat{\mu} = \overline{X}$ consistent ? Does $\hat{\mu}$ converges in distribution to μ ?
Sol. $\hat{\mu} = \overline{X} \overset{a.s.}{\to} E(X) = \mu$ by the strong law of large numbers (SLLN). $\hat{\mu}$ is strongly consistent.

$Y_n \to X \Rightarrow Y_n \overset{a.s.}{\to} X \Rightarrow Y_n \overset{P}{\to} X \Rightarrow Y_n \overset{D}{\to} X$.

Thus $\overline{X}$ is consistent and converges in distribution to μ. $F_\mu(t) = ?$
Another direct proof of consistency:

$$\begin{aligned} P(|\overline{X} - \mu| < \epsilon) =& P(\left|\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right| < \frac{\epsilon}{\sigma/\sqrt{n}}) \\ =& \Phi(\frac{\epsilon}{\sigma/\sqrt{n}}) - \Phi(-\frac{\epsilon}{\sigma/\sqrt{n}}) \\ \to& 1 - 0 = 1\ \forall\ \epsilon > 0. \end{aligned}$$

Does it prove strong consistency here ? Question:

$P(|\overline{X} - \mu| < \epsilon) \overset{a.s.}{\to} 1$?
$P(|\overline{X} - \mu| < \epsilon) \overset{P}{\to} 1$?

$P(|\overline{X}-\mu|<\epsilon)\xrightarrow{D}1$?
$\overline{X}\to\mu$?

Review of probability theory:

Suppose X_i, X, Y_n are random variables and a, b, c are constant.
$\mathbf{Y}_n$ and $\mathbf{Y}$ are $p-$dimensional random vectors and $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ are constant in $\mathcal{R}^p$.

1. SLLN: If X_1, ..., X_n are i.i.d. from X, $E(|X|)<\infty$, then $\overline{X}\xrightarrow{a.s.}E(X)$.

2. $Y_n\xrightarrow{a.s.}X \Rightarrow Y_n\xrightarrow{P}X \Rightarrow Y_n\xrightarrow{D}X$.

3. If $Y_n\xrightarrow{a.s.}a$ and g is a continuous function then $g(Y_n)\xrightarrow{a.s.}g(a)$.

4. If $Y_n\xrightarrow{P}a$ and g is a continuous function then $g(Y_n)\xrightarrow{P}g(a)$.

5. Continuous mapping theorem. If $g(\cdot)$ is a continuous function, then
$\mathbf{Y}_n\xrightarrow{a.s.}\mathbf{Y}$ (**?**) $\Rightarrow g(\mathbf{Y}_n)\xrightarrow{a.s.}g(\mathbf{Y})$.
$\mathbf{Y}_n\xrightarrow{P}\mathbf{Y} \Rightarrow g(\mathbf{Y}_n)\xrightarrow{P}g(\mathbf{Y})$.
$\mathbf{Y}_n\xrightarrow{D}\mathbf{a} \Rightarrow g(\mathbf{Y}_n)\xrightarrow{D}g(\mathbf{a})$.

6. If a_n is constant and $a_n\to a$, then $a_n\xrightarrow{a.s.}a$, $a_n\xrightarrow{P}a$ and $a_n\xrightarrow{D}a$. **Why ??**

7. If $Var(\hat\theta_n)\to 0$ and $Bias(\hat\theta_n)\to 0\ \forall\ \theta\in\Theta$,
then $\hat\theta_n$ is a consistent estimator of θ.

Proof. By Chebychev's inequality, $P(|X|\ge\epsilon)\le E(X^2)/\epsilon^2$

$$\begin{aligned}P(|\hat\theta_n-\theta|\ge\epsilon)\le& E((\hat\theta_n-\theta)^2)/\epsilon^2\\ =&\frac{Var(\hat\theta_n)+(Bais(\hat\theta_n))^2}{\epsilon^2}\to 0,\quad\forall\ \epsilon>0.\end{aligned}$$

Thus $\hat\theta_n\xrightarrow{P}\theta$. □

8. If $\hat\theta_n$ is consistent and a_n and b_n are constant satisfying $a_n\to 1$ and $b_n\to 0$, then $a_n\hat\theta_n+b_n$ are consistent.

Proof. Make use of Result 6.Let $\mathbf{Y}_n=(\hat\theta_n,a_n,b_n)$ and $g(y,a,b)=ay+b$, then $\mathbf{Y}_n\xrightarrow{P}(\theta,1,0)$, $g(\mathbf{Y}_n)=a_n\hat\theta_n+b_n\xrightarrow{P}g(\theta,1,0)=1\cdot\theta+0=\theta$. □

Theorem 1 (consistency of the MLE). *Assume that the following conditions:*
(A1) X_1, ..., X_n *are i.i.d. with* $f(\cdot;\theta_o)$, $\theta_o\in\Theta$;
(A2) $f(\cdot;\theta)\ne f(\cdot;\theta^*)\ \forall\ \theta\ne\theta^*$ *and* $\theta,\theta^*\in\Theta$ *(identifiability);*
(A3) $\{x:\ f(x;\theta)>0\}$ *does not depend on* θ *and* $\frac{\partial}{\partial\theta}f(x;\theta)$ *exists;*
(A4) Θ *contains an open set* O *and* $\theta_o\in O$;
(A5) $\tau=\tau(\theta)$ *is a continuous function of* θ.
Then the MLE $\hat\tau$ *of* $\tau(\theta)$ *is consistent i.e.,* $\hat\tau\xrightarrow{P}\tau(\theta_o)$.

Theorem 1 explains why we like the MLE. Theorem 1 only gives a sufficient condition.

Q: (1) Examples that (A3) fails ? (2) Let $X=rnorm(100)$. $\theta=$? What is the difference between θ_o and θ ? (3) Let $X_{ij}\sim N(\underbrace{\gamma_i+\alpha_j}_{\mu_{ij}},\sigma^2)$ with unknown $(\gamma_i,\alpha_j,\sigma)$, $i,j\in\{1,2\}$. Does (A2) hold ?

$$\begin{pmatrix}\mu_{11}\\ \mu_{12}\\ \mu_{21}\\ \mu_{22}\end{pmatrix}=\begin{pmatrix}1&0&1&0\\1&0&0&1\\0&1&1&0\\0&1&0&1\end{pmatrix}\begin{pmatrix}\gamma_1\\ \gamma_2\\ \alpha_1\\ \alpha_2\end{pmatrix}=\begin{pmatrix}0\\0\\0\\0\end{pmatrix}$$

$$\begin{pmatrix}1&0&1&0\\0&1&1&0\\1&0&0&1\\0&1&0&1\end{pmatrix}\begin{pmatrix}\gamma_1\\ \alpha_1\\ \gamma_2\\ \alpha_2\end{pmatrix}=\begin{pmatrix}0\\0\\0\\0\end{pmatrix}$$

$$\begin{pmatrix}1&0&1&0\\0&1&1&0\\0&0&-1&1\\0&0&-1&1\end{pmatrix}\begin{pmatrix}\gamma_1\\ \alpha_1\\ \gamma_2-\gamma_1\\ \alpha_2-\alpha_1\end{pmatrix}=\begin{pmatrix}0\\0\\0\\0\end{pmatrix}$$

Counterexample. ......

Example 2. Suppose that X_1, ..., X_n are i.i.d. from X, with $Var(X) = \sigma^2 < \infty$.
$\hat{\sigma}^2 = \frac{1}{n}\sum_i (X_i - \overline{X})^2$ and $S^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X})^2$.
Are they consistent estimators of σ^2 ?
Sol. Q: Can we apply Theorem 1 ?
1. SLLN: If X_1, ..., X_n are i.i.d. from X, $E(|X|) < \infty$, then $\overline{X} \stackrel{a.s.}{\to} E(X)$.
Notice that $\hat{\sigma}^2 = \overline{X^2} - (\overline{X})^2$.
$\overline{X} \stackrel{a.s.}{\to} E(X) = \mu$ by the SLLN;
$\overline{X^2} \stackrel{a.s.}{\to} E(X^2) = \sigma^2 + \mu^2$ by the SLLN.
$g(x, y) = x - y^2$ is continuous.
By Results 5, $g(\overline{X^2}, \overline{X}) \stackrel{a.s.}{\to} g(\sigma^2 + \mu^2, \mu) = \sigma^2 + \mu^2 - (\mu)^2 = \sigma^2$.
That is, $\hat{\sigma}^2$ is strongly consistent and thus is consistent.
Notice that $S^2 = \frac{n}{n-1}\hat{\sigma}^2$. Let $a_n = \frac{n}{n-1}$ and $b_n = 0$, then $a_n \to 1$ and $b_n \to 0$. By Result 9, S^2 is consistent estimator of σ^2.
Example 3. Supposet that $X \sim bin(n, p)$. $\tau = p(1 - p)$. Are the MLE of (p, τ) consistent ?
Sol. MLE $\hat{p} = X/n$ and $\hat{\tau} = \hat{p}(1 - \hat{p})$. **Can we use Theorem 1 ?**
Notice that $X = Y_1 + \cdots + Y_n$, where Y_i's are i.i.d. from $bin(1, p)$. Thus $X/n = \overline{Y}$. It can verified that the conditions in Theorem 1 are satisfied. Thus $\hat{p}$ and $\hat{\tau}$ are consistent.
In particular, (A3) holds *i.e.* $f(x; \theta) = f_Y(y; \theta) = \theta^y(1 - \theta)^{1-y}$.
$f(x; \theta) = \theta^y(1 - \theta)^{1-y} = f(x; \theta_1) = \theta_1^y(1 - \theta_1)^{1-y}\ \forall\ y \in \{0, 1\}$.
$=> \theta = \theta_1$.
Another way: By the SLLN, $\hat{p} \stackrel{a.s.}{\to} p$.
Let $g(p) = p(1 - p)$. g is continuous, thus $\hat{\tau}$ is strongly consistent.

10.2. Efficiency.

Definition. An estimator $\hat{\tau}$ is asymptotically efficient for $\tau(\theta)$ if

$$\sqrt{n}(\hat{\tau} - \tau(\theta)) \stackrel{D}{\to} N(0, v(\theta)) \text{ and } v(\theta) = \frac{(\tau'(\theta))^2}{E((\frac{\partial}{\partial\theta}\ln f(X; \theta))^2)}. \tag{1}$$

Recall $Y_n \stackrel{D}{\to} X <=> F_{Y_n}(x) \to F_X(x)$ for each continuity point x of F_X.
$Y_n \to X \quad => \quad Y_n \stackrel{a.s.}{\to} X \quad => \quad Y_n \stackrel{P}{\to} X \quad => \quad Y_n \stackrel{D}{\to} X.$
Results:

1. **The central limit theorem (CLT).** If X_1, ..., X_n are i.i.d. from X, $\mu = E(X)$ and $\sigma^2 = Var(X) < \infty$, then
   $\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0, 1)$.
   $\overline{X} - \mu \xrightarrow{D} N(0, \sigma^2/n)$ **??**
   $(\overline{X} - \mu)\sqrt{n} \xrightarrow{D} N(0, \sigma^2)$ **??**
2. **Slutsky's Theorem**. If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} a$, then
   $X_n + Y_n \xrightarrow{D} X + a$, **How about** $X_n - Y_n \xrightarrow{D} X - a$ **?**
   $Y_n X_n \xrightarrow{D} aX$. **How about** $X_n/Y_n \xrightarrow{D} X/a$ **?**
   How about $Y_n/X_n \xrightarrow{D} a/X$ **?**
3. **Delta method.** Suppose that
   (1) $Y_n(\omega)$ and θ are $p \times 1$ vectors;
   (2) $\sqrt{n}(Y_n - \theta) \xrightarrow{D} N(0, \Sigma)$, or $(\hat{\Sigma}_{Y_n})^{-1/2}(Y_n - \theta) \xrightarrow{D} N(0, I_{p\times p})$;
   (3) $g(\cdot)$ is a function, $\bigtriangledown g$ is continuous and $\bigtriangledown g(\theta) \neq 0$. Then

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{D} N(0, v(\theta)), \text{ where } v(\theta) = (\bigtriangledown g(\theta))^t \Sigma \bigtriangledown g(\theta); \tag{2}$$

$$\frac{g(Y_n) - g(\theta)}{\sqrt{\hat{v}/n}} \xrightarrow{D} N(0, 1), \text{ where } \hat{v}/n = (\bigtriangledown g(\hat{\theta}))^t \hat{\Sigma}_{Y_n} \bigtriangledown g(\hat{\theta}), \tag{3}$$

and $\hat{\Sigma}$ $(= n\hat{\Sigma}_{Y_n})$ is a consistent estimator of Σ.

Note: $A^{-\frac{1}{2}} = Q'D^{-\frac{1}{2}}Q$ if $A = Q'DQ$. $\quad Q = ?\ D = ?\ D^{-\frac{1}{2}} = ?$

Remark. Roughly speaking, Eq. (2) means that

(1) $E(g(Y_n)) \approx g(\theta)$ if n is large enough;

(2) $V(g(Y_n)) \approx (\bigtriangledown g(\theta))^t \Sigma \bigtriangledown g(\theta)/n$ if n is large enough

(asymptotic variance of $g(Y_n)$).

4. **Cramér-Rao Inequality (CR- Ineq.)** Let X_1, ..., X_n be i.i.d. from $X \sim f(x;\theta)$ and let $W(\mathbf{X})$ be a statistic. Suppose that

(1) $\frac{d}{d\theta}E(W) = \begin{cases} \int \frac{\partial}{\partial\theta}W(\mathbf{x})f(\mathbf{x};\theta)d\mathbf{x} & \text{if } \mathbf{X} \text{ is continuous} \\ \sum_{\mathbf{x}} \frac{\partial}{\partial\theta}W(\mathbf{x})f(\mathbf{x};\theta) & \text{if } \mathbf{X} \text{ is discrete;} \end{cases}$

(2) $Var(W) < \infty$.

$$Then\ \ Var(W) \geq \frac{(\frac{d}{d\theta}E(W))^2}{E((\frac{\partial}{\partial\theta}\ln f(\mathbf{X};\theta))^2)} = \frac{(\frac{d}{d\theta}E(W))^2}{nE((\frac{\partial}{\partial\theta}\ln f(X;\theta))^2)}.\ \ (CRLB).$$

(Note: If σ_W^2 =CRLB then W is the UMVUE of $\tau(\theta)$ = ??. Thus $\hat{\tau}$ is efficient = $\hat{\tau}$ is approximately UMVUE.)

To prove efficiency,

use CLT or Delta method to find $Var(\hat{\tau})$ and show $Var(\hat{\tau}) \approx$ CRLB.

Theorem 1. Suppose that assumptions (A1) – (A5) in Theorem 1 of §10.1 hold; $\hat{\theta}$ is the MLE of θ; $\bigtriangledown\tau(\theta)$ is a continuous function of θ;

(A6) For each $x \in \mathcal{X}$ (the sample space), $f'''(x;\theta)$ is continuous, and $\frac{\partial^3}{\partial\theta^3}\int f(x;\theta)dx = \int \frac{\partial^3}{\partial\theta^3}f(x;\theta)dx$;

(A7) $\forall\ \theta_o \in \Theta$, $\exists\ c > 0$ and a function $M(x)$ such that $E(M(X)) < \infty$ and $|\frac{\partial^3}{\partial\theta^3}\ln f(x;\theta)| \leq M(x)\ \forall\ x \in \mathcal{X}$ and $|\theta - \theta_o| < c$.

Then $\sqrt{n}(\tau(\hat{\theta}) - \tau(\theta_o)) \xrightarrow{D} N(0, v(\theta_o))$ (see Eq. (1)).

Remark.

1. $\hat{\tau}$ is efficient if n is large enough and A1-A7 hold;
2. Roughly speaking, $E(\tau(\hat{\theta})) \approx \tau(\theta_o)$ if n is large enough and A1-A7 hold;
4. Roughly speaking, $V(\tau(\hat{\theta})) \approx v(\theta_o)/n$ if n is large enough and A1-A7 hold;
4. If $f(\cdot;\theta)$ belongs to an exponential family, then A1–A7 hold.

Example 1. Suppose that $X \sim bin(n,p)$. The odd ratio $\tau = p/(1-p)$.

(1) Is the MLE $\hat{\tau}$ of τ consistent and efficient ?

(2) $V(\hat{\tau}) \approx$?

Sol The MLE of p is $\hat{p} = X/n$ and can be viewed as $\hat{p} = \overline{Y}$, $Y \sim bin(1,p)$.

Reason: $X = Y_1 + \cdots + Y_n$, where Y_1, ..., Y_n are i.i.d. from $bin(1,p)$. *i.e.*, $\hat{p} = \overline{Y}$.

(1) By the SLLN, $\hat{p} \xrightarrow{a.s.} p$. Thus it is strongly consistent.

MLE of τ is $\hat{\tau} = \tau(\hat{p}) = \hat{p}/(1-\hat{p})$ (**Why ??**).

Assume $p \neq 1$. Then τ is continuous. Thus $\hat{\tau}$ is strongly consistent.

$bin(1,p)$ belongs to an exponential family, thus A1-A7 hold.

Thus $\hat{\tau}$ is asymptotically efficient.

(2) Two ways to <u>*approximate*</u> $V(\hat{\tau})$:

(I) CRLB or Fisher information formula

$V(\hat{\tau}) \approx (\tau'(\theta))^2/I_n(\theta)$, where $(I_n(\theta) = E((\frac{\partial}{\partial\theta}\ln f(\mathbf{X};\theta))^2))$,

(II) Delta method.

$\sqrt{n}(\hat{p} - p) \xrightarrow{D} N(0,\sigma^2)$ where $\sigma^2 = p(1-p)$ **Why ?**

$\tau' = (-1 + \frac{1}{1-p})' = \frac{1}{(1-p)^2} \neq 0$ if $p \neq 1$.

$\sqrt{n}(\tau(\hat{p}) - \tau(p)) \xrightarrow{D} N(0, v(p))$ (**Why ?**)

where $v(p) = \tau'(p)\sigma^2\tau'(p) = \frac{p(1-p)}{(1-p)^4} = p/(1-p)^3$

Question: Note $P(\hat{\tau} \leq x) \approx \Phi(\frac{x-\tau(p)}{\sqrt{p/(1-p)^3 n}})$. Can we say $\hat{\tau} \xrightarrow{D} N(\tau(p), \frac{p}{n(1-p)^3})$?

Question: $\hat{v} = ?$
Question: $V(\hat{\tau}) \approx ?$
Question: Can we say $\sqrt{n}(\tau(\hat{p}) - \tau(p)) \xrightarrow{D} N(0, \hat{v})$ (**?**)
Question: Can we say $\sqrt{n}\frac{\tau(\hat{p})-\tau(p)}{\sqrt{\hat{v}}} \xrightarrow{D} N(0,1)$?
Question: What happens to $\sqrt{n}(\tau(\hat{p}) - \tau(p))$ if $p = 1$?
If $p = 1$, $\hat{p}$= ? $\tau(p)$ = ? $\tau(\hat{p})$ = ?
Ans: If $p = 1$ then $P(\hat{p} = 1) = P(\tau(\hat{p}) = \tau(p)) = 1$ and $\sqrt{n}(\tau(\hat{p}) - \tau(p)) \xrightarrow{D} 0$.
Example 2. Suppose that $X \sim bin(n, p)$, $\hat{p} = X/n$ and $\hat{\sigma}^2 = \hat{p}(1 - \hat{p})$.
(1) Is $\hat{\sigma}^2$ efficient ? (2) $Var(\hat{\sigma}^2)$ = ?
Sol. (1) **Question.** Can we use Theorem 1 (for MLE) here ?
(2) There are two ways. **First way:** We first use the Delta method.
Let $\tau(p) = \sigma^2 = p(1 - p)$. Recall $\hat{\sigma}^2 = \hat{\tau}$ is strongly consistent.
$\tau' = 1 - 2p \neq 0$ unless $p = 1/2$.
$\sqrt{n}(\tau(\hat{p}) - \tau(p)) \xrightarrow{D} N(0, v(p))$,
where $v(p) = (1 - 2p)^2 p(1 - p)$ by the Delta method. **Are we done ?**
$V(\hat{\tau}) = (1 - 2p)^2 p(1 - p)/n$?
$V(\hat{\tau}) \approx (1 - 2p)^2 p(1 - p)/n$?

2nd way: Compute $V(\hat{\tau})$. $\hat{\tau} = X/n - (X/n)^2$.
$V(\hat{\tau}) = E(X^2/n^2 - 2X^3/n^3 + X^4/n^4) - (E(X/n - X^2/n^2))^2$,
where $E(X^k)$ can be obtained by the mgf:
$M_X(t) = E(e^{Xt}) = (E(e^{Y_1 t}))^n = (q + pe^t)^n$.
$M'(t) = n(q + pe^t)^{n-1} pe^t$, => $E(X) = np$,
$M''(t) = n(n - 1)(q + pe^t)^{n-2}(pe^t)^2 + n(q + pe^t)^{n-1} pe^t$, => $E(X^2) = npq + (np)^2$,
$M'''(t) = n(n - 1)(n - 2)(q + pe^t)^{n-3}(pe^t)^3 + 3n(n - 1)(q + pe^t)^{n-2}(pe^t)^2 + n(q + pe^t)^{n-1} pe^t$ => $E(X^3) = n(n - 1)(n - 2)p^3 + 3n(n - 1)p^2 + np$,
$M^{(4)}t) = n(n-1)(n-2)(n-3)(q+pe^t)^{n-4}(pe^t)^4 + 6n(n-1)(n-2)(q+pe^t)^{n-3}(pe^t)^3 + 7n(n-1)(q+pe^t)^{n-2}(pe^t)^2 + n(q + pe^t)^{n-1} pe^t$
=> $E(X^4) = n(n - 1)(n - 2)(n - 3)p^4 + 6(n(n - 1)(n - 2)p^3 + 7(n(n - 1)p^2 + np$.
Question. What can be said if $p = 1/2$?
Ans: $Y_n = n\big(\frac{\hat{p}-1/2}{\frac{1}{2}}\big)^2 \xrightarrow{D} \chi^2(1)$.
This is proved in homework of 501. It is also included as follows.
$\tau = p - p^2$, $\tau'(1/2) = 0$, $\tau''(p) = -2$, $\tau^{(k)}(p) = 0$, $k \geq 3$.
By the Taylor expansion,
$\tau(\hat{p}) - \tau(1/2) = \tau'(1/2)(\hat{p} - 1/2)/1! + \tau''(o.5)(\hat{p} - 1/2)^2/2! + \sum_{k=3}^{\infty} \tau^{(k)}(0.5)(\hat{p} - 0.5)^k/k!$.
$\tau(\hat{p}) - \tau(1/2)) = \tau''(0.5)(\hat{p} - 1/2)^2/2 = -(\hat{p} - 1/2)^2$.
Let $Y_n = -4n(\tau(\hat{p}) - \tau(1/2))$, *then* $Y_n = n(\frac{\hat{p}-1/2}{1/2})^2$.
$\vdash$: $Y_n = n\big(\frac{\hat{p}-1/2}{\frac{1}{2}}\big)^2 \xrightarrow{D} \chi^2(1)$.
Reason: *Let* $Z_n = \sqrt{n}\frac{\hat{p}-1/2}{\frac{1}{2}} \xrightarrow{D} N(0, 1)$. *Letting* $Z \sim N(0, 1)$, *then for* $t \geq 0$,
$F_{Y_n}(t) = P(Y_n \leq t) = P(Z_n^2 \leq t)$
$= P(-\sqrt{t} \leq Z_n \leq \sqrt{t})$
$\rightarrow P(-\sqrt{t} \leq Z \leq \sqrt{t})$ **??**
$= P(Z^2 \leq t)$.

Thus if $p = 1/2$, then
(a) $-4n(\tau(\hat{p}) - \tau(1/2)) \xrightarrow{D} \chi^2(1)$.
Remark. If $p = 1/2$, $v(p) = (1 - 2p)^2 p(1 - p) = 0$ by Th. 1, and $\sqrt{n}(\tau(\hat{p}) - \tau(p)) \xrightarrow{D} N(0, v(p)) = N(0, 0)$ **what does it mean ?**
(b) $\sqrt{n}(\tau(\hat{p}) - \tau(1/2)) \xrightarrow{D} 0$
But it is not very useful, as it is the same as

(c) $\sqrt{n}(\tau(\hat{p}) - \tau(1/2)) \xrightarrow{P} 0$ or

(d) $(\tau(\hat{p}) - \tau(1/2)) \xrightarrow{P} 0$.

In particular, we cannot approximate $F_{\hat{\tau}}(x)$ based on either (b) or (c), but we can approximate $F_{\hat{\tau}}(x)$ based on (a). *e.g.* for $x \in (0, 1/4)$,

$F_{\hat{\tau}}(x) = P(\hat{\tau} \le x) = P(-4n(\hat{\tau} - 1/4) \ge -4n(x - 1/4)) \approx P(\chi^2(1) \ge -4n(x - 1/4))$, $F_{\hat{\tau}}(x) = P(\hat{\tau} \le x) = P(\sqrt{n}(\tau(\hat{p}) - \tau(1/2)) \le \sqrt{n}(x - (1/4)) \approx \mathbf{1}(\sqrt{n}(x - (1/4) \ge 0) \approx \mathbf{1}(x \ge 0.25)$ **Any contradiction ??**

Theorem 1 in §10.2. Suppose that A1-A4 hold, $\hat{\theta}$ is the MLE of θ;

Moreover, $\tau(\hat{\theta})$ is efficient if $\bigtriangledown\tau(\theta)$ is a continuous function of θ;

(A1) $X_1, ..., X_n$ are i.i.d. with $f(x;\theta)$, $\theta \in \Theta$;

(A2) $f(\cdot;\theta) \ne f(\cdot;\theta^*)\ \forall\ \theta \ne \theta^*$;

(A3) $\{x:\ f(x;\theta) > 0\}$ does not depend on θ and $\frac{\partial}{\partial\theta} f(x;\theta)$ exists;

(A4) Θ contains an open set O and the true parameter $\theta_o \in O$.

(A5) $\tau = \tau(\theta)$ is a continuous function of θ.

(A6) For each $x \in \mathcal{X}$ (the sample space), $f'''(x;\theta)$ is continuous, and $\frac{\partial^3}{\partial\theta^3} \int f(x;\theta)dx = \int \frac{\partial^3}{\partial\theta^3} f(x;\theta)dx$ (or $\sum <-> \int$);

(A7) For each $\theta_o \in \Omega$, $\exists$ a $c > 0$ and a function $M(x)$ such that $E(M(X)) < \infty$.

Counterexample if the assumptions in Theorem 1 are not valid.

Example 2 in §7.3. Let $X_1, ..., X_n$ be i.i.d. from $X \sim U(0,\theta)$.

The MLE is $\hat{\theta} = X_{(n)}$. An unbiased estimator is $\tilde{\theta} = \frac{n+1}{n} X_{(n)}$.

$V(\tilde{\theta}) = \frac{1}{n(n+2)}\theta^2$

$Var(\hat{\theta}) = \frac{n}{(n+1)^2(n+2)}\theta^2$.

$CRLB = v(\theta)/n = \frac{(\frac{d}{d\theta}\theta)^2}{nE(\frac{\partial}{\partial\theta}\ln f(X;\theta))^2} = \frac{\theta^2}{n} >> \frac{n}{(n+1)^2(n+2)}\theta^2 = Var(\hat{\theta})$. $\sqrt{n}(\hat{\theta} - \theta) \overset{D}{\not\longrightarrow} N(0, v(\theta))$.

Reason that the CRLB fails:

$\{x:\ f(x;\theta) > 0\} = (0,\theta)$ depends on θ. Thus A3 fails.

Remark. If $n < 20$, $Var(\hat{\tau}) \approx CRLB$ is not valid !!

Counterexample. Let $X \sim bin(4,p)$, the MLE of p is again $\hat{p} = X/4$. Let $p = .9$, $Y = \sqrt{4}(\frac{X}{4} - p)$ and $Z \sim N(0, p(1-p))$. $n < 20$?

$\vdash$: $F_Y \not\approx F_Z$.

Question: How should we prove it ?

Find a t such that

$F_Y(t) \not\approx F_Z(t)$.

x :	0	1	2	3	4
$Y(x) = 2(\frac{x}{4} - 0.9)$:	-1.8	-1.3	-0.8	-0.3	0.2

Thus $F_Z(0.2) = \Phi(\frac{0.2}{\sqrt{0.9(1-0.9)}}) = \Phi(0.67) \approx 0.75 << 1 = F_Y(0.2)$.

Remark. If $X \sim bin(n,p)$ with $n = 100$, $p = 0.9$ and $Y = \sqrt{n}(X/n - p)$, then $\sigma_Y = 0.3$, $F_Y(0.3) = 0.883$ and $F_Z(0.3) = 0.841$, where $Z \sim N(0, 0.09)$.

Q: Why efficiency of $\hat{\theta}$?

1. To find an estimator that has the smallest asymptotic variance (asymptotic UMVUE).
2. To approximate $F_{\hat{\theta}}$.

e.g. in Example 2, for $x \ge 0$ (**why ?**)

$$F_{\hat{\sigma}^2}(x) \approx \begin{cases} \Phi(\frac{x - p(1-p)}{\sqrt{(1-2p)^2 p(1-p)/n}}) & \text{if } p \ne 1/2 \\ 1 - F_{\chi^2(1)}(-n(x - 1/4)) & \text{if } p = 1/2 \text{ and } 0 < x < 1/4, \\ 1 & \text{if } p = 1/2 \text{ and } x \ge 1/4. \end{cases}$$

§10.3. Hypothesis testing. For large samples, say $X_1, ..., X_n$ ($n \ge 20$), we have three approximate large sample testing procedures:

	$H_0:\ \theta = \theta_o$ v.s. $H_1:$	$\theta \ne \theta_o$	$\theta > \theta_o$	$\theta < \theta_o$
A. Z-test:	$\phi =$	$\mathbf{1}(\lvert Z\rvert > z_{\alpha/2})$	$\mathbf{1}(Z > z_\alpha)$	$\mathbf{1}(Z < -z_\alpha)$
	$P - value \approx$	$2(1 - \Phi(\lvert Z\rvert))$	$(1 - \Phi(Z))$	$\Phi(Z)$

where $Z = \frac{\hat{\theta}-\theta_o}{\hat{\sigma}_{\hat{\theta}}}$ and $\hat{\theta}$ is an estimae. Reject H_0 if $\phi = 1$ or P-value$< \alpha$.

Remark 1. $\hat{\sigma}^2_{\hat{\theta}}$ can be obtained as follows.

1. If $\hat{\theta} = \overline{X}$, use CLT with $Var(\overline{X})$ estimated by S^2/n.
2. If $\hat{\theta} = g(\overline{X})$, then use the delta method.
3. If $\hat{\theta}$ is an MLE and
   assumptions (A1)–(A7) in Theorem 1 of §10.2 hold, then
   $\hat{\sigma}^2_{\hat{\theta}} = \begin{cases} (nI_1(\theta_o))^{-1} & \text{(CRLB), if doable} \\ (n\hat{I}_1(\theta_o))^{-1} & \text{(empirical CRLB), otherwise} \end{cases}$ where
   $I_1(\theta) = E((\frac{\partial}{\partial\theta}\ln f(X;\theta))^2) = -E(\frac{\partial^2}{\partial\theta^2}\ln f(X;\theta))$,
   $\hat{I}_1(\theta) = \frac{1}{n}\sum_{i=1}^n(\frac{\partial}{\partial\theta}\ln f(X_i;\theta))^2 = -\frac{1}{n}\sum_{i=1}^n(\frac{\partial^2}{\partial\theta^2}\ln f(X_i;\theta))$. **Why ?**

Question: How about $\hat{\sigma}^2_{\hat{\theta}} = (n\hat{I}_1(\hat{\theta}))^{-1}$?

Remark 2. The Z-test makes use of the statement $\frac{\hat{\theta}-\theta}{\hat{\sigma}_{\hat{\theta}}} \xrightarrow{D} N(0,1)$. For instance, in case 1 of Remark 1,

$$\frac{\overline{X}-\mu}{\hat{\sigma}_{\overline{X}}} = \frac{\overline{X}-\mu}{S/\sqrt{n}} \xrightarrow{D} N(0,1), \tag{1}$$

v.s. the CLT

$$\sqrt{n}\frac{\overline{X}-\mu}{\sigma_X} \xrightarrow{D} N(0,1).$$

Notice that

$$\frac{\overline{X}-\mu}{\hat{\sigma}_{\overline{X}}} = \frac{\overline{X}-\mu}{S/\sqrt{n}} = \underbrace{\frac{\overline{X}-\mu}{\sigma_X/\sqrt{n}}}_{X_n \xrightarrow{D} N(0,1)} \underbrace{\frac{\sigma_X}{S}}_{Y_n \xrightarrow{P} 1} = X_n Y_n.$$

Thus Eq. (1) follows from the CLT and Slutsky's theorem.

Questions: Are these expressions accurate ?

1. $\sigma^2_{\overline{X}} \approx \sigma^2_X/n$,
2. $\hat{\sigma}^2_{\overline{X}} \approx S^2/n$.
3. $\sigma^2_{\overline{X}} = S^2/n$.

B. Score test (assuming A1-A7). For testing H_0: $\theta = \theta_o$ v.s. H_1: $\theta \neq \theta_o$,

$$\phi = \mathbf{1}(|\frac{S(\theta_o)}{\sqrt{nI_1(\theta_o)}}| > z_{\alpha/2}) \text{ or } \phi = \mathbf{1}(|\frac{S(\theta_o)}{\sqrt{n\hat{I}_1(\theta_o)}}| > z_{\alpha/2})$$

where $S(\theta) = \frac{\partial}{\partial\theta}\sum_{i=1}^n \ln f(X_i;\theta)$ (score function), and
$I_1(\theta) = E((\frac{\partial}{\partial\theta}\ln f(X_i;\theta))^2)$ is the Fisher information number.

C. LRT. For testing H_0: $\theta \in \Theta_o$, v.s. H_1: $\theta \notin \Theta_o$, if the assumptions A1-A7 hold, then the LRT $\mathbf{1}(\lambda \leq c)$ can be approximated by

$$\phi = \mathbf{1}(-2\ln\lambda \geq \chi^2_{d,\alpha})$$

where d = degree of freedom in Θ − degree of freedom in Θ_o, λ is the likelihood ratio statistic and $1 - F_{\chi^2_d}(\chi^2_{d,\alpha}) = \alpha$.

Example 1. Let X_1, ..., X_n be i.i.d. from $f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$, $x, \theta > 0$. Test H_0: $\theta = 1$ v.s. H_1: $\theta \neq 1$ at level $\alpha = 0.05$ in the 2 cases: Case 1. $n = 25$ and $\overline{X} = 1.44$. **Why just record $\overline{X}$?** Case 2. $n = 1$ and $X_1 = 0.05$,

Sol. Case 1. We can use either of the three large sample tests.

A. Z-test or Wald test. $\phi = \mathbf{1}(|\frac{\hat{\theta}-\theta_o}{\hat{\sigma}_{\hat{\theta}}}| > z_{\alpha/2})$, $z_{0.025} \approx 1.96$.

$\hat{\theta} = \overline{X}$, $\sigma^2_{\hat{\theta}} = \sigma^2_X/n = \theta^2/n$.

$$\hat{\sigma}_{\hat{\theta}} = \begin{cases} \sqrt{1.44^2/25} = 1.44/5 = 0.288 & \text{if use } \hat{\sigma}^2_{\hat{\theta}} = \hat{\theta}^2/n, \\ \sqrt{1/25} = 0.2 & \text{if use } \hat{\sigma}^2_{\hat{\theta}} = \theta_o^2/n. \end{cases}$$

$$Z = \begin{cases} \frac{|\hat{\theta}-1|}{0.288} = 1.57 & \text{if use } \hat{\sigma}^2_{\hat{\theta}} = \hat{\theta}^2/n \\ \frac{|\hat{\theta}-1|}{0.2} = 2.2 & \text{if use } \hat{\sigma}^2_{\hat{\theta}} = \theta_o^2/n, \\ ? & \text{if use } \hat{\sigma}^2_{\overline{X}} = S^2/n \end{cases} \Rightarrow \begin{cases} \text{do not reject } H_0 \\ \text{reject } H_0. \\ ? \end{cases}$$

Q: Which decision is more "reliable" ?
B. Score test.

$$\phi = \mathbf{1}(|\frac{S(\theta_o)}{\sqrt{nI_1(\theta_o)}}| > z_{\alpha/2}) \text{ or } \mathbf{1}(|\frac{S(\theta_o)}{\sqrt{n\hat{I}_1(\theta_o)}}| > z_{\alpha/2}) \textit{ where}$$

$S(\theta) = \frac{\partial}{\partial\theta}\sum_{i=1}^n \ln f(X_i;\theta)$ and $nI_1(\theta) = E((S(\theta))^2) = -nE(\frac{\partial^2 \ln f(X;\theta)}{\partial\theta^2})$.
$S(\theta) = \sum_i(-\ln\theta - X_i/\theta)'_\theta = \sum_i(-\frac{1}{\theta} + \frac{X_i}{\theta^2}) = \frac{n(\overline{X}-\theta)}{\theta^2}$.
$I_1(\theta) = Var(-\frac{1}{\theta} + \frac{X_i}{\theta^2}) = Var(X_i)/\theta^4 = \frac{1}{\theta^2}$.

$|\frac{S(\theta_o)}{\sqrt{nI_1(\theta_o)}}| = |\frac{\overline{X}-\theta_o}{\theta_o/\sqrt{n}}| = |Z|$

which is the same as the Z-test.
Q: $\hat{I}_1(\theta_o) = ?$ $\hat{I}_1(\theta) = \frac{1}{n}\sum_{i=1}^n(\frac{\partial}{\partial\theta}\ln f(X_i;\theta))^2 = \frac{\overline{(X-\theta)^2}}{\theta^2}$

$= \frac{-1}{n}\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2}\ln f(X_i;\theta) = -\frac{\partial}{\partial\theta}\frac{\overline{X}-\theta}{\theta^2} = \frac{2\overline{X}}{\theta^3} - \frac{1}{\theta^2}$. $\hat{I}_1(\theta_o) = 2\overline{X} - 1$ or $\overline{(X-1)^2}$

Should we use $|Z| = |\frac{S(\theta_o)}{\sqrt{n\hat{I}_1(\theta_o)}}|$ **or** $|Z| = |\frac{S(\theta_o)}{\sqrt{nI_1(\theta_o)}}|$ **?**

Example 1 (continued). Let X_1, ..., X_n be i.i.d. from $f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$, $x, \theta > 0$. Test H_0: $\theta = 1$ v.s. H_1: $\theta \neq 1$ at level $\alpha = 0.05$ in the 2 cases:
Case 1. $n = 25$ and $\overline{X} = 1.44$.
Case 2. $n = 1$ and $X_1 = 0.05$,
Sol. Case 1. We can use either of the three large sample tests.
A. Z-test or Wald test. $\phi = \mathbf{1}(|Z| > 1.96)$, where $Z = \frac{\hat{\theta}-\theta_o}{\hat{\sigma}_{\hat{\theta}}}$.
B. Score test.

$$\phi = \mathbf{1}(|\frac{S(\theta_o)}{\sqrt{nI_1(\theta_o)}}| > z_{\alpha/2}) \text{ or } \mathbf{1}(|\frac{S(\theta_o)}{\sqrt{n\hat{I}_1(\theta_o)}}| > z_{\alpha/2}) \textit{ where}$$

$S(\theta) = \frac{\partial}{\partial\theta}\sum_{i=1}^n \ln f(X_i;\theta)$ and $nI_1(\theta) = E((S(\theta))^2) = -nE(\frac{\partial^2 \ln f(X;\theta)}{\partial\theta^2})$.
Here it is the same as Z test.
C. LRT. $\phi = \mathbf{1}(-2\ln\lambda \geq \chi^2_{d,\alpha})$.

$L = \theta^{-n}\exp(-n\overline{X}/\theta)$. **How to get it ?**

$$\lambda = \frac{\exp(-n\overline{X})}{(\overline{X})^{-n}\exp(-n)} = (\overline{X})^n \exp(n(1-\overline{X})). \tag{2}$$

$$d = \underbrace{1}_{d.f.\ under\ \Theta} - \underbrace{0}_{d.f.\ under\ H_o},$$

RR: $-2\ln\lambda \geq \chi^2_{1,0.05} = 3.84$.
Since $-2\ln\lambda = 3.767844$,
Do not reject H_0.

Is it consistent with Z-test or Score test ? Why ?
Recall in Z-test, $z_{\alpha/2} = 1.96 = \sqrt{3.84}$, and for comparison,

$$\begin{cases} \sqrt{-2\ln\lambda} = 1.94 & \text{do not reject } H_o, \\ Z = \frac{\hat{\theta}-\theta_o}{\hat{\sigma}_{\hat{\theta}}(\theta_o)} = 2.2, & \text{reject } H_o. \\ Z = \frac{\hat{\theta}-\theta_o}{\hat{\sigma}_{\hat{\theta}}(\hat{\theta})} = 1.57, & \text{do not reject } H_o. \end{cases} \tag{1}$$

Case 2. $n = 1$. **Can we apply Z-test or Score test ?**
LRT Method.

$\ln\lambda = \ln X_1 + 1 - X_1$ by Eq. (2), $(\ln\lambda(x))'_x = \frac{1}{x} - 1$. $(\ln\lambda(x))''_x = \frac{-1}{x^2}$. $\ln\lambda(x)$ is concave down with maximum at $x = 1$.
LRT $\phi = \lambda(X_1) = \mathbf{1}(\lambda \leq c) = \mathbf{1}(X_1 \leq k_1 \textit{ or } X_1 \geq k_2)$, $k_1 < k_2$,
where $E_{\theta_o}(\lambda) = \alpha$ and $\lambda(k_1) = \lambda(k_2)$. These two equations are equivalent to

$$k_1 e^{1-k_1} = k_2 e^{1-k_2} \text{ and } e^{-k_1} - e^{-k_2} = 1 - \alpha = 0.95.$$

$=> k_1 e^{1-k_1} - k_2(k_1) e^{1-k_2(k_1)} = 0.$ $\quad k_2(k_1)$**= ??** Solving the equations yields $k_1 = 0.04221185$ and $k_2 = 4.748271$

Since $X_1 = 0.05 \in (0.042, 4.75)$, we do not reject H_0.

Question: Can we use RR: $-2\ln\lambda > \chi^2_{1,0.05}$ **?**

Remark. If we use RR: $-2\ln\lambda > \chi^2_{1,0.05}$, then $-2\ln\lambda = 4.09 > 3.84$ and we reject H_0.

Reasoning of the Score test (assuming A1-A7).

$\vdash$: $\frac{S(\theta_o)}{\sqrt{n\hat{I}_1(\theta_o)}} \xrightarrow{D} N(0,1)$ assuming X_i's are discrete.

Let $Y_i = \frac{\partial}{\partial\theta}\ln f(X_i;\theta)$, then

$S(\theta) = \sum_{i=1}^n Y_i$; by assumption A1-A7,

$$E_\theta(Y_i) = E_\theta(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)) = \sum_x \frac{\frac{\partial}{\partial\theta} f(x;\theta)}{f(x;\theta)} f(x;\theta) = (\sum_x f(x;\theta))'_\theta = 0$$

$I_1(\theta) = E(Y_i^2) = Var_\theta(Y_i) = \sigma_Y^2$,

$$\frac{\overline{Y} - E(Y_i)}{\sigma_Y/\sqrt{n}} \; (= \frac{n\overline{Y} - nE(Y_i)}{n\sigma_Y/\sqrt{n}} = \frac{S(\theta)}{\sqrt{nI_1(\theta)}}) \xrightarrow{D} N(0,1);$$

By the SLLN, $(\overline{X} \xrightarrow{a.s.} E(X))$,

$$\hat{I}_1(\theta) = -\frac{1}{n}\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2}\ln f(X_i;\theta) \xrightarrow{a.s.} E(-\frac{\partial^2}{\partial\theta^2}\ln f(X_i;\theta)) = I_1(\theta)$$

$$\frac{S(\theta_o)}{\sqrt{n\hat{I}_1(\theta_o)}} = \frac{S(\theta_o)}{\sqrt{nI_1(\theta_o)}} \frac{\sqrt{I_1(\theta_o)}}{\sqrt{\hat{I}_1(\theta_o)}} \xrightarrow{D} N(0,1)$$

Remark. The large sample LRT test $\phi = \mathbf{1}(-2\ln\lambda > \chi^2_{d,\alpha})$ makes use of the result

$$-2\ln\lambda \xrightarrow{D} \chi^2_d \text{ or } F_{-2\ln\lambda} \approx F_{\chi^2_d}.$$

There are empirical method and rigorous method to check whether

$$F_{-2\ln\lambda} \approx F_{\chi^2_1}.$$

Rigorous method: Compare $f_{\chi^2_1}$ and $f_{-2\ln\lambda}$ by

1. derive the density function of χ^2_1 $(= G(\frac{1}{2}, 2))$,
2. derive the density function of $-2\ln\lambda$ $(= f_X(g^{-1}(y))|\frac{dg^{-1}(y)}{dy}|$, where $y = g(x) = -2\ln\lambda(x))$ (no explicit solution).

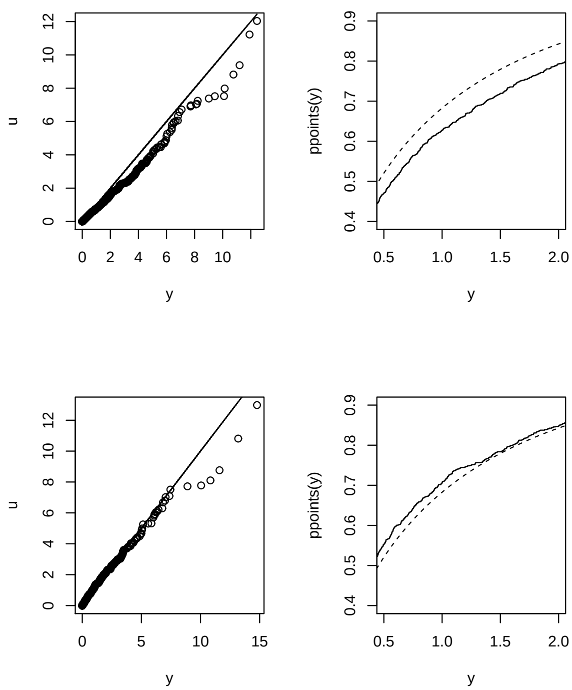
Empirical method. Two ways:

1. The empirical distribution function (edf) $\hat{F}(t) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}(X_i \le t)$ $(= \overline{W}) \xrightarrow{a.s.} F(t)$ if X_1, ..., X_n are i.i.d. from $F(t)$.
2. qqplot (quantile-quantile plot). We expect a straight line if the two samples are from the same distributions.

Definition. If F is the cdf, $F^{-1}(p) = \inf\{x : F(x) \ge p\}$ is the quantile function.

In Example 1, if $n = 100$, we can use the empirical method to see that

$$F_{-2\ln\lambda} \approx F_{\chi^2_1}$$



The figure is created by R program as follows.
myfun=function(n){
m=10000
x=rgamma(m,n,1)/n # m $\overline{X}$'s
y=-2*(n*log(x)+n*(1-x)) # m λ's
u=rchisq(m,1)
qqplot(y,u)
lines(y,y)
y=sort(y)
plot(y,ppoints(y), xlim=c(0.5,2), ylim=c(0.4,0.9), type="S",lty=1)
lines(y,pchisq(y,1), xlim=c(0.5,2), ylim=c(0.4,0.9), type="l",lty=2)
}
makepsfile = function(a,b) {
ps.options(horizontal = FALSE)
ps.options(height=9.0, width=6.5)
postscript("fig10.ps")
par(mfrow=c(2,2))
n=1
myfun(n)
n=100
myfun(n)
dev.off()
}
makepsfile()
You can use "gv fig10.ps" to view the graph and "lpr fig10.ps" to print the graph in the Linex system.

Example 2. Suppose that two independent random samples: X_1, ..., X_n i.i.d. $\sim N(0,\sigma_X^2)$, Y_1, ..., Y_m i.i.d. $\sim N(0,\sigma_Y^2)$, $\gamma=\sigma_Y^2/\sigma_X^2$, Observing $\overline{X^2}=1.2$ and $\overline{Y^2}/\overline{X^2}=2$ with $n=m=25$, do the data support $\gamma=1$ at level 0.05 ?

Sol. Three ways: LRT, Z-tes and Score test.

LRT. $\theta=(\sigma_X^2,\sigma_Y^2)\in\Theta$ and $\gamma=\sigma_Y^2/\sigma_X^2$. **Degree of freedom of Θ ?** $\Theta=\{(x,y):\ x,y>0\}$, $(\sigma_X^2,\sigma_Y^2)\in\Theta$,

$\Theta_o = \{(x,y): \ x=y>0\}$ and $\gamma=\gamma_o=1$.
H_0: $\gamma=\gamma_o=1$, H_1: $\gamma\neq 1$. $\phi=\mathbf{1}(-2\ln\lambda \geq \chi^2_{d,\alpha})$. d= ? α= ? λ= ?
Degree of freedom under Θ is 2,
Degree of freedom under H_0 is 1. $d=2-1$.
$\lambda = L(\hat\theta_o)/L(\hat\theta)$.

$$L=(2\pi\sigma_X^2)^{-n/2}\exp(-\frac{1}{2}\sum_i X_i^2/\sigma_X^2)\cdot(2\pi\sigma_Y^2)^{-m/2}\exp(-\frac{1}{2}\sum_i Y_i^2/\sigma_Y^2).$$

MLE under Θ: $\hat\sigma_X^2=\sum_i X_i^2/n$, $\hat\sigma_Y^2=\sum_i Y_i^2/m$, $\hat\gamma=\hat\sigma_Y^2/\hat\sigma_X^2$.
MLE under H_0: $\tilde\sigma_X^2=\frac{\sum_i X_i^2+\sum_j Y_j^2}{n+m}$, $\tilde\sigma_Y^2=\tilde\sigma_X^2$, $\tilde\gamma=\gamma_o$.
$\lambda = \frac{(\tilde\sigma_X^2)^{-n/2}(\tilde\sigma_Y^2)^{-m/2}\exp(-\frac{m+n}{2})}{(\hat\sigma_X^2)^{-n/2}(\hat\sigma_Y^2)^{-m/2}\exp(-\frac{n}{2}-\frac{m}{2})}$
$=(\frac{n}{n+m}\frac{\sum_i X_i^2+\sum_j Y_j^2}{\sum_i X_i^2})^{-n/2}(\frac{m}{n+m}\frac{\sum_i X_i^2+\sum_j Y_j^2}{\sum_i Y_i^2})^{-m/2}$
$=[(\frac{n}{m+n})(1+\frac{\sum_j Y_j^2}{\sum_i X_i^2})]^{-n/2}[(\frac{m}{n+m})(\frac{\sum_i X_i^2}{\sum_j Y_j^2}+1/\gamma_o)]^{-m/2}$,
$\lambda=((1/2)(1+2))^{-25/2}((1/2)(\frac{1}{2}+1))^{-25/2}=(\frac{9}{8})^{-25/2}$
Test: $\phi=\mathbf{1}(-2\ln\lambda\geq\chi^2_{d,\alpha})$.
$-2\ln\lambda=25\ln\frac{9}{8}=1.28\geq\chi^2_{1,0.05}=3.84$? Thus the data support the claim that $\gamma=1$.
Z-test. $\phi=\mathbf{1}(|Z|>1.96)$, where $Z=\frac{\hat\gamma-1}{\hat\sigma_{\hat\gamma}}\xrightarrow{D}N(0,1)$.
Notice

$$\gamma=\sigma_Y^2/\sigma_X^2\overset{def}{=}g(\sigma_Y^2,\sigma_X^2).$$

Delta method. Suppose that
(1) $W_n(\omega)$ and θ are $p\times 1$ vectors;
(2) $\sqrt{n}(W_n-\theta)\xrightarrow{D}N(0,\Sigma)$, or $(\hat\Sigma_{W_n})^{-1/2}(W_n-\theta)\xrightarrow{D}N(0,I_{p\times p})$;
(3) $\bigtriangledown g$ is continuous and $\bigtriangledown g(\theta)\neq 0$. Then

$$\sqrt{n}(g(W_n)-g(\theta))\xrightarrow{D}N(0,v(\theta)),\ \text{where}\ v(\theta)=(\bigtriangledown g(\theta))^t\Sigma\bigtriangledown g(\theta); \tag{2}$$

$$\frac{g(W_n)-g(\theta)}{\sqrt{\hat v}}\xrightarrow{D}N(0,1),\ \text{where}\ \hat v=(\bigtriangledown g(\hat\theta))^t\hat\Sigma_{W_n}\bigtriangledown g(\hat\theta),\ (if\ v\neq 0), \tag{3}$$

and $\hat\Sigma$ is a consistent estimator of Σ. θ= ? $g(\theta)$ = ? $g(y,x)$= ? p = ? W_n = ?
$\hat\gamma=\overline{Y^2}/\overline{X^2}=2$, $(=g(\overline{Y^2},\overline{X^2}))$.
$Y/\sigma_Y\sim N(0,1)$,
$Y^2/\sigma_Y^2\sim\chi^2(1)$ with $E(\chi^2(1))=1$ and $V(\chi^2(1))=2$.
$\sigma^2_{\hat\gamma}\approx(1/x,-y/x^2)\hat{Cov}(\begin{pmatrix}\overline{Y^2}\\ \overline{X^2}\end{pmatrix})\begin{pmatrix}1/x\\ -y/x^2\end{pmatrix}\Big|_{(y,x)=E(Y^2,X^2)}$
$=(1/x,-y/x^2)\begin{pmatrix}2\sigma_Y^4/m & 0\\ 0 & 2\sigma_X^4/n\end{pmatrix}\begin{pmatrix}1/x\\ -y/x^2\end{pmatrix}$ $(Y^2/\sigma_Y^2\sim\chi^2(1))$
$=[2\frac{\sigma_Y^4}{\sigma_X^4}+2\sigma_X^4\sigma_Y^4/\sigma_X^8]/n=4\gamma_o^2/25=4/25$.
$\hat\sigma_{\hat\gamma}$ = ?
$Z=\frac{\hat\gamma-\gamma_o}{\hat\sigma_{\hat\gamma}}=2.5>1.96$,
The data do not suggest that $\gamma=1$.
Question: Why is there a discrepancy in conclusion ?
Ans. 1. They are different tests and there is randomness. 2. Z-test is more acurate here as it has less approximation than LRT. ($\sigma^2_{\hat\gamma}\approx 4\gamma_o/25=4/25$).
Score test ? Recall $\phi=\mathbf{1}(|\frac{S(\theta_o)}{\sqrt{nI_1(\theta_o)}}|>z_{\alpha/2})$ or $\phi=\mathbf{1}(|\frac{S(\theta_o)}{\sqrt{n\hat I_1(\theta_o)}}|>z_{\alpha/2})$
For vector θ,

$$\phi=\mathbf{1}(||I_{n+m}^{\frac{-1}{2}}(\theta_o)S(\theta_o)||>z_{a/2})$$

$$=\mathbf{1}(||I_{n+m}^{\frac{-1}{2}}(\theta_o)S(\theta_o)||^2 > \chi^2_{1,\alpha})$$
$$=\mathbf{1}(S(\theta_o)^t I_{m+n}^{-1}(\theta_o)S(\theta_o) > \chi^2_{1,\alpha})$$

or $\phi = \mathbf{1}(||\hat{I}_{m+n}^{\frac{-1}{2}}(\theta_o)S(\theta_o)||^2 > \chi^2_{1,\alpha})$.

$\boldsymbol{\theta = ?}$

Vector approach: $S(\theta) = \frac{\partial \ln \prod_{j=1}^m f(Y_j;\theta_1)\prod_{i=1}^n f(X_i;\theta_2)}{\partial \theta^t}$,

$$I_{m+n}(\theta) = E(S(\theta)S(\theta)^t) = -E(\frac{\partial^2 \ln \prod_{j=1}^m f(Y_j;\theta_1)\prod_{i=1}^n f(X_i;\theta_2)}{\partial \theta^t \partial \theta}. \quad (4)$$

$$S(\theta) = \frac{\partial}{\partial \theta}[(-n/2)\log(\sigma_X^2) - \frac{1}{2}\sum_i X_i^2/\sigma_X^2 + (-n/2)\log(\sigma_Y^2) - \frac{1}{2}\sum_i Y_i^2/\sigma_Y^2] \quad (\theta = (theta_1, \theta_2))$$
$$= \begin{pmatrix} \frac{-n/2}{\sigma_Y^2} + \frac{1}{2}\frac{\sum_i Y_i^2}{\sigma_Y^4} \\ \frac{-n/2}{\sigma_X^2} + \frac{1}{2}\frac{\sum_i X_i^2}{\sigma_X^4} \end{pmatrix}$$
$$= \frac{n}{2}\begin{pmatrix} \frac{-1}{\sigma_Y^2} + \frac{\overline{Y^2}}{\sigma_Y^4} \\ \frac{-1}{\sigma_X^2} + \frac{\overline{X^2}}{\sigma_X^4} \end{pmatrix}$$

$$(4) => I_{m+n}(\theta) = -E\begin{pmatrix} \frac{n/2}{\sigma_Y^4} - \frac{2}{2}\frac{\sum_i Y_i^2}{\sigma_Y^6} & 0 \\ 0 & \frac{n/2}{\sigma_X^4} - \frac{2}{2}\frac{\sum_i X_i^2}{\sigma_X^6} \end{pmatrix} = \frac{n}{2}\begin{pmatrix} 1/\sigma_Y^4 & 0 \\ 0 & 1/\sigma_X^4 \end{pmatrix}$$

$S(\theta)' I_{n+m}^{-1}(\theta)S(\theta) = \frac{n}{2}(\frac{(\overline{Y^2}-\sigma_Y^2)^2}{\sigma_Y^4} + \frac{(\overline{X^2}-\sigma_X^2)^2}{\sigma_X^4})$

$\phi = \mathbf{1}(\frac{n}{2}(\frac{(\overline{Y^2}-\sigma_Y^2)^2}{\sigma_Y^4} + \frac{(\overline{X^2}-\sigma_X^2)^2)}{\sigma_X^4} > 3.84)$

$\sigma_X = \sigma_Y$, $\overline{Y^2} = 2\overline{X^2}$, and $\overline{X^2} = 1.2 => \hat{\sigma}_X^2 = \hat{\sigma}_Y^2 = 3 \times 1.2/2$.

$\phi = \mathbf{1}(\frac{n}{2}(\frac{(\overline{Y^2}-\hat{\sigma}_Y^2)^2}{\hat{\sigma}_Y^4} + \frac{(\overline{X^2}-\hat{\sigma}_X^2)^2)}{\hat{\sigma}_X^4} > 3.84) = \mathbf{1}(12.5[(0.5 \times 1.2/1.5)^2 + (1.5 \times 1.2/1.5)^2] > 3.84) = \mathbf{1}(4 > 3)$. Reject H_0.

The next approach does not work. $\theta = \sigma_X^2 = \sigma_Y^2$.

$S(\theta) = \frac{\partial \ln \prod_{j=1}^m f(Y_j;\theta)\prod_{i=1}^n f(X_i;\theta)}{\partial \theta}$,

$I_{m+n}(\theta) = E((S(\theta))^2) = -E(\frac{\partial^2 \ln \prod_{j=1}^m f(Y_j;\theta)\prod_{i=1}^n f(X_i;\theta)}{\partial \theta \partial \theta}$.

$S(\theta) = \frac{\partial}{\partial \theta}[(-n/2)\log(\sigma_X^2) - \frac{1}{2}\sum_i X_i^2/\sigma_X^2 + (-n/2)\log(\sigma_Y^2) - \frac{1}{2}\sum_i Y_i^2/\sigma_Y^2]$

$= \frac{-n/2}{\sigma_X^2} + \frac{1}{2}\frac{\sum_i Y_i^2}{\sigma_X^4} + \frac{-n/2}{\sigma_X^2} + \frac{1}{2}\frac{\sum_i X_i^2}{\sigma_X^4}$

$= \frac{n}{2}(\frac{-1}{\sigma_X^2} + \frac{\overline{Y^2}}{\sigma_X^4} + \frac{-1}{\sigma_X^2} + \frac{\overline{X^2}}{\sigma_X^4})$

$= \frac{n}{2\sigma_X^4}(-\sigma_X^2 + \overline{Y^2} - \sigma_X^2 + \overline{X^2})$

$= \frac{n}{2\theta^2}(-\theta + \overline{Y^2} - \theta + \overline{X^2})$

$= \frac{n}{2}(-2\theta^{-1} + (\overline{Y^2} + \overline{X^2})\theta^{-2})$

$S(\hat{\theta}) = ??$

Under H_o, $S(\theta_o)$ depends on θ_o which is unknown, and $S(\hat{\theta}_o) = 0$, as $\hat{\theta}_o = \frac{n\overline{X^2}+n\overline{Y^2}}{2n} = \frac{\overline{X^2}+\overline{Y^2}}{2}$. Thus the score test approach does not work. **Ignore the rest arguements.**

$\hat{\sigma}_X^2 = \overline{X^2}$, $\overline{X^2} = 1.2$ and $\overline{Y^2} = 2.4$.

$S(\hat{\theta}) = \frac{n}{2(\overline{X^2})^2}(\overline{X^2}) = \frac{25}{2(1.2)^2}(1.2) \approx 10.4$

$I_{m+n}(\theta) = -E(\frac{n}{2}(\theta^{-2} - 2\overline{Y^2}\theta^{-3} + \theta^{-2} - 2\overline{X^2}\theta^{-3})) = n/\sigma_X^4$, $I_{m+n}(\hat{\theta}) = ??$

$\phi = \mathbf{1}(|\frac{S(\hat{\theta}_o)}{\sqrt{n\hat{I}_1(\hat{\theta}_o)}}| > z_{\alpha/2}) = \mathbf{1}(\frac{\frac{n}{2(\overline{X^2})^2}(\overline{X^2})}{\sqrt{n}/\overline{X^2}} > 1.96) = \mathbf{1}(2.5 > 1.96)$

Reject H_0.

§10.4. Approximate CI.

For finite samples, construction of CI is based on
inverting LRT and
pivotal method.

Def. $\frac{\hat\theta-\theta}{\hat\sigma_{\hat\theta}}$ is called a Wald-type statistic if θ is known and if

$$P(|\frac{\hat\theta-\theta}{\hat\sigma_{\hat\theta}}| \le z_{\alpha/2}) \approx 1-\alpha \tag{1}$$

A. CI based on Wald-type statistic.
CI for θ: $\hat\theta \pm z_{\alpha/2}\hat\sigma_{\hat\theta}$ due to $P(\{\theta: |\frac{\hat\theta-\theta}{\hat\sigma_{\hat\theta}}| \le z_{\alpha/2}\}) \approx 1-\alpha$.

Remark. (1) A smooth $g(\theta)$ can also be viewed as a parameter, thus
CI for $g(\theta)$: $g(\hat\theta) \pm z_{\alpha/2}\hat\sigma_{g(\hat\theta)}$,
where $\sigma^2_{g(\hat\theta)} \approx (g'(\theta))^2\sigma^2_{\hat\theta}$ or $(\bigtriangledown g(\theta))^t\Sigma_{\hat\theta} \bigtriangledown g(\theta)$. $\hat\sigma^2_{g(\hat\theta)} = ?$
(2) A monotone function $g(\theta)$: $g(\hat\theta) \pm g(\hat\sigma_{\hat\theta})$.

B. CI based on MLE. $\{\theta: |\frac{\hat\theta-\theta}{\sigma_{\hat\theta}}| \le z_{\alpha/2}\}$.

C. CI based on score function.
$\{\theta: |\frac{\sum_{i=1}^n \frac{\partial}{\partial\theta}\ln f(X_i;\theta)}{\sqrt{nI_1(\theta)}}| \le z_{\alpha/2}\}$.

D. CI based on LRT.
$\{\theta: -2\ln\lambda(\mathbf{x};\theta) \le \chi^2_{d,\alpha}\}$

Example 1. Let $X \sim bin(n,p)$, $n=100$. Observe $X=30$. Approximiate 95% CI for p ?

Sol. A 95% CI for p :

(1) Wald-type. $|\frac{\hat p-p}{\sqrt{\hat p(1-\hat p)/n}}| \le 1.96$, or $\hat p \pm 1.96\sqrt{\hat p(1-\hat p)/n}$ or
$0.3 \pm 1.96 \times \sqrt{0.21}/10$. Thus a 95% CI for p is $(0.210, 0.390)$.

(2) MLE method:

$$|\frac{\hat p-p}{\sqrt{p(1-p)/100}}| \le 1.96 \text{ (due to Eq. (1)).} \tag{2}$$

$\hat p^2 - 2p\hat p + p^2 \le (1.96^2/100)(p-p^2)$.
$\hat p^2 - (2\hat p + 0.01*1.96^2)p + (1+0.01*1.96^2)p^2 = c + bp + ap^2 \le 0$.
$\frac{-b-\sqrt{b^2-4ac}}{2a} \le p \le \frac{-b+\sqrt{b^2-4ac}}{2a}$. A 95% CI for p is $(0.219, 0.396)$

(3) Score method: $\ln f(X;p) \propto X\ln p + (n-X)\ln(1-p)$.
Score function: $S(p) = \frac{X}{p} - \frac{n-X}{1-p}$.
$-S'(p) = \frac{X}{p^2} + \frac{n-X}{(1-p)^2}$.
$I_n(p) = -E(S'(p)) = \frac{np}{p^2} + \frac{n-np}{(1-p)^2} = \frac{n}{p(1-p)}$.
$|\frac{S(p)}{\sqrt{I_n(p)}}| = |\frac{\frac{X}{p}-\frac{n-X}{1-p}}{\sqrt{\frac{n}{p(1-p)}}}| \le 1.96$.
$\frac{\frac{X}{p}-\frac{n-X}{1-p}}{\sqrt{\frac{n}{p(1-p)}}} = \frac{X-np}{\sqrt{np(1-p)}} = \frac{X/n-p}{\sqrt{p(1-p)/n}}$. (see Eq. (2)). A 95% CI for p is $(0.219, 0.396)$

(4) LRT. $\lambda = \frac{p^X(1-p)^{n-X}}{(X/n)^X(1-X/n)^{n-X}}$.
$\ln\lambda = X\ln(p*n/X) + (n-X)\ln((1-p)/(1-X/n))$,
$\{p: -2\ln\lambda \le \chi^2_{1,\alpha} = 3.84\}$.

A 95% CI for p is $(0.216, 0.395)$ (solving by R).

Question: Which is more convenient ? Which is better ?

R-program:

```
myfun=function(p){
x=30
n=100
y=-2*(x*log(p*n/x)+(n-x)*log((1-p)/(1-x/n))) # −2lnλ
```

```
u=min(y[y>=3.84])
p[y==u]
}
p=(200:300)/1000
myfun(p)
[1] 0.216
p=(300:400)/1000
myfun(p)
[1] 0.395
```

Example 1 (continued) Suppose that $X \sim bin(n,p)$, $n = 100$. Observe $X = 30$. A 95% CI for the odd ratio $g(p) = p/(1-p)$?

Sol. A 95% CI for $g(\theta)$: $g(\hat\theta) \pm z_{\alpha/2}\hat\sigma_{g(\hat\theta)}$,
where $\hat\sigma^2_{g(\hat\theta)} = (g'(\hat\theta))^2\hat\sigma^2_{\hat\theta}$,

$g'(p) = (-1 + \frac{1}{1-p})' = \frac{1}{(1-p)^2}$.

95% CI for $g(p)$ is

$\frac{\hat p}{1-\hat p} \pm 1.96\sqrt{\frac{\hat p(1-\hat p)}{(1-\hat p)^4 n}}$ or $\frac{0.3}{0.7} \pm 1.96 \times \sqrt{\frac{0.3}{0.7^3 \times 100}}$.

A 95% CI for $g(p)$ is $(0.245, 0.612)$

How about the other approaches ?

If the CI is $\tilde p \pm a$ and g is monotone, then $g(\tilde p) \pm g(a)$. **Pay attention to ↑ or ↓.**

Example 2. Suppose that $X \sim bin(n,p)$, $n = 3$. Observe $X = 1$. $\hat p \pm 1.96\sqrt{\hat p(1-\hat p)/n}$ yields $\frac{1}{3} \pm 1.96\sqrt{\frac{2}{27}}$. (a) **Is it a 95% CI for p ?** (b) How about if $X = 1$ and $n = 25$?

Sol. Confidence coefficient of CI $= \inf_p P_p(p \in I)$. **???**
The CI is of the form

$$I = \begin{cases} 0 \pm 0 & \text{if } X = 0 \\ \frac{1}{3} \pm 1.96\sqrt{2/27} & \text{if } X = 1 \\ \frac{2}{3} \pm 1.96\sqrt{2/27} & \text{if } X = 2 \\ 1 \pm 0 & \text{if } X = 3 \end{cases} = \begin{cases} [0,0] & \text{if } X = 0 \\ [0, \frac{1}{3} + a] & \text{if } X = 1 \\ [\frac{2}{3} - a, 1] & \text{if } X = 2 \\ [1,1] & \text{if } X = 3 \end{cases} \quad (a \approx 0.53).$$

$$\tfrac{1}{3} - a < 0 < \tfrac{2}{3} - a < \tfrac{1}{3} + a < 1 < \tfrac{2}{3} + a.$$

The coverage probability is $P_p(p \in I)$

$$= \begin{cases} P_p(X = 0 \text{ or } 1) & \text{if } p = 0 \\ P_p(X = 1) & \text{if } p \in (0, \frac{2}{3} - a) \\ P_p(X = 1 \text{ or } 2) & \text{if } p \in [\frac{2}{3} - a, \frac{1}{3} + a] \\ P_p(X = 2) & \text{if } p \in (\frac{1}{3} + a, 1) \\ P_p(X = 2 \text{ or } 3) & \text{if } p = 1 \end{cases} = \begin{cases} (1-p)^3 & \text{if } p = 0 \\ 3p(1-p)^2 & \text{if } p \in (0, \frac{2}{3} - a) \\ 3p(1-p) & \text{if } p \in [\frac{2}{3} - a, \frac{1}{3} + a] \\ 3p^2(1-p) & \text{if } p \in (\frac{1}{3} + a, 1) \\ p^3 & \text{if } p = 1 \end{cases}$$

Confidence coefficient of CI $= \inf_p P_p(p \subset I) = \min\{1, 0, 3/4, 0, 1\} = 0$

Answer: No, it is a 0% CI for p.

(b). What is the question ?

The approximate 95% CI is $0.04 \pm 1.96\sqrt{24/25^3}$ if $X = 1$ and $n = 25$.

```
p=(0:10)/25
round(p-1.96*sqrt(p*(1-p)/25),3)
round(p+1.96*sqrt(p*(1-p)/25),3)
```

X	0	1	2	3	4	5	6	7	8	9	10
L	0.000	−0.037	−0.026	−0.007	0.016	0.043	0.073	0.104	0.137	0.172	0.208
R	0.000	0.117	0.186	0.247	0.304	0.357	0.407	0.456	0.503	0.548	0.592

Coverage probability $P_p(p \in I) = \begin{cases} P_p(X \in \{0,1,2,3\}) & \text{if } p \in [0,0], \\ P_p(X \in \{1,2,3\}) & \text{if } p \in (0,0.016), \\ P_p(X \in \{1,2,3,4\}) & \text{if } p \in [0.016,0.043), \\ P_p(X \in \{1,2,3,4,5\}) & \text{if } p \in [0.043,0.073), \\ P_p(X \in \{1,2,3,4,5,6\}) & \text{if } p \in [0.073,0.104], \\ P_p(X \in \{1,2,3,4,5,6,7\}) & \text{if } p \in (0.104,0.117], \\ P_p(X \in \{2,3,4,5,6,7\}) & \text{if } p \in (0.117,186] \\ \cdots & \cdots \end{cases} = \begin{cases} p^0(1-p)^{25} + \binom{25}{1}p^1(1-p)^{24} + \binom{25}{2} \\ p(1-p)[\binom{25}{1}p^0(1-p)^{23} + \binom{25}{2}p^1(1 \\ \cdots \end{cases}$

Coeffidence coefficient of $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n}$ is

$\inf_p P_p(p \in I) \le \inf_{p\in(0,0.016]} P_p(X \in \{1,2,3\}) = 0.$

Any contradiction ? Approximated 95% **CI:** If n is large,

$P(\hat{p} - 1.96\sqrt{\hat{p}(1-\hat{p})/n} \le p \le \hat{p} + 1.96\sqrt{\hat{p}(1-\hat{p})/n}) = P(\frac{|\hat{p}-p|}{\sqrt{\hat{p}(1-\hat{p})/n}} < 1.96) \approx 0.95$

95% **CI:** $\inf_p P_p(p \in [L,R]) = 0.95$.

Remark. Recall in Example 4 of Chapter 9. If $X \sim bin(3,p)$ and $I = [\frac{X-1}{3} \vee 0, 1]$, the confidence coefficient of I is 0.59.

Review of testing and CI.

Assume that X_1, ..., X_n are i.i.d. from $N(\mu, \sigma^2)$.

1. If H_0: $\mu = \mu_0$ and σ is known,

LRT test statistic λ yields $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ and observe z.

H_1	ϕ	$P-value$	CI
$\mu \ne \mu_0$	$\mathbf{1}(\lvert Z\rvert > z_{\alpha/2})$	$2P(Z > \lvert z\rvert)$	$(\overline{X} - z_{\alpha/2}\sigma/\sqrt{n}, \overline{X} + z_{\alpha/2}\sigma/\sqrt{n})$
$\mu > \mu_0$	$\mathbf{1}(Z > z_\alpha)$	$P(Z > z)$	$(\overline{X} - z_\alpha\sigma/\sqrt{n}, \infty)$
$\mu < \mu_0$	$\mathbf{1}(Z < -z_\alpha)$	$P(Z < z)$	$(-\infty, \overline{X} + z_\alpha\sigma/\sqrt{n})$

2. If H_0: $\mu = \mu_0$ and σ is unknown,

LRT test statistic λ yields $T = \frac{\overline{X}-\mu_0}{S/\sqrt{n}}$ and observe t.

H_1	ϕ	$P-value$	CI
$\mu \ne \mu_0$	$\mathbf{1}(\lvert T\rvert > t_{n-1,\alpha/2})$	$2P(T > \lvert t\rvert)$	$(\overline{X} - \frac{t_{n-1,\alpha/2}S}{\sqrt{n}}, \overline{X} + \frac{t_{n-1,\alpha/2}S}{\sqrt{n}})$
$\mu > \mu_0$	$\mathbf{1}(T > t_{n-1,\alpha})$	$P(T > t)$	$(\overline{X} - \frac{t_{n-1,\alpha}S}{\sqrt{n}}, \infty)$
$\mu < \mu_0$	$\mathbf{1}(T < -t_{n-1,\alpha})$	$P(T < t)$	$(-\infty, \overline{X} + \frac{t_{n-1,\alpha}S}{\sqrt{n}})$

3. If H_0: $\sigma = \sigma_0$ and μ is unknown, LRT test statistic λ yields $Y = (n-1)S^2/\sigma_0^2$ and observe y.

H_1	ϕ	$P-value$	$CI\ of\ \sigma$
$\sigma \ne \sigma_0$	$\mathbf{1}(Y \notin (a,b))$	$\begin{cases} 2P(Y < y) & \text{if } y < n-1 \\ 2P(Y > y) & \text{if } y > n-1 \end{cases}$	$(\sqrt{Y/b}, \sqrt{Y/a})$
$\sigma > \sigma_0$	$\mathbf{1}(Y > \chi^2_{n-1,\alpha})$	$P(Y > y)$	$(\sqrt{\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha}}}, \infty)$
$\sigma < \sigma_0$	$\mathbf{1}(Y < \chi^2_{n-1,1-\alpha})$	$P(Y < y)$	$(0, \sqrt{\frac{(n-1)S^2}{\chi^2_{n-1,\alpha}}})$

$\lambda(a) = \lambda(b)$ and $E(\phi) = \alpha$.

Remark. If $N(\mu, \sigma^2)$ is not valid, the derivation cannot be unified.

Chapter 11. Introduction to non-parametric analysis

Common interests of estimation are

1. Mean μ,
2. SD σ,
3. cdf $F(t)$.

In this course, we make the assumption that

X_1, ..., X_n are i.i.d. with cdf $F_o(t;\theta)$, where F_o is known, but not θ ($\in \Theta$),

Then $\mu = \mu(\theta)$ and $\sigma = \sigma(\theta)$.
We derive either the MLEs or other types of estimators:

$$\begin{pmatrix} \hat\theta \\ \hat\mu = \mu(\hat\theta) \\ \hat\sigma = \sigma(\hat\theta) \\ \hat F(t) = F_o(t;\hat\theta) \end{pmatrix} \quad e.g.\ X \sim bin(n,p), \text{ then } \begin{pmatrix} \hat p = X/n \\ \hat\mu_X = n\hat p \\ \hat\sigma = \sqrt{n\hat p(1-\hat p)} \\ \hat F(t) = \sum_{i\le t}\binom{n}{i}(\hat p)^i(1-\hat p)^{n-i} \end{pmatrix}$$

This is called point estimation of the parametric analysis. The CI and testing hypotheses discussed so far are also under the parametric analysis frame work.

Question:

How do we know that the assumption $F(t) = F_o(t;\theta)$ is correct ?

Answer: One approach is to compare the parametric estimator $F_o(t;\hat\theta)$ to the empirical distribution function (edf)

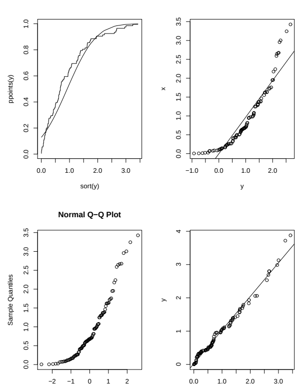
$$\hat F(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \le t).$$

Example 1. Here is a simulation study for checking normality assumption. Given a data set, assume that it is normal, then compute the MLE $\hat\mu$ and $\hat\sigma$, and estimate the cdf by $F_o(t;\hat\mu,\hat\sigma)$. Then compare it to its edf $\hat F$.

```
    par(mfrow=c(2,2))
    x=rexp(100)
# Now pretend we only have data x without knowing it is from Exp(1)
    u=mean(x)
    s=sd(x)
    x=sort(x)
    plot(x,ppoints(x),type="S") # edf
    lines(x,pnorm(x,u,s),type="l") # F̂(t; μ̂, σ̂)
    y=rnorm(100,u,s)
    qqplot(y,x) # check N(·,·)
    abline(lm(x~ y))
```

Check for linearity.

```
    qqnorm(x)
    y=rexp(100,1/u)
    qqplot(x,y)
    abline(lm(y~ x))
```



Properties of the edf: $\hat F(t) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(X_i \le t)$.

1. $\hat F$ is a cdf and for each t, $\hat F(t)$ can be viewed as $\overline{Y}$, where $Y_1, ..., Y_n$ are i.i.d. $\sim bin(1,p)$ and $p = F(t)$.
2. $E(\hat F(t)) = E(\overline{Y}) = E(Y) = p = F(t)$.
3. $Var(\hat F(t)) = V(\overline{Y}) = pq/n = F(t)(1-F(t))/n$.
4. $Cov(\hat F(t), \hat F(s)) = \frac{1}{n}Cov(\mathbf{1}(X_1 \le t), \mathbf{1}(X_1 \le s)) \quad (\Sigma_{\overline{\mathbf{Z}}} = \Sigma_{\mathbf{Z}}/n)$
   $= \frac{1}{n}(E(\mathbf{1}(X_1 \le s)\mathbf{1}(X_1 \le t)) - E(\mathbf{1}(X_1 \le s))E(\mathbf{1}(X_1 \le t))) = \frac{F(t\wedge s) - F(t)F(s)}{n}$. **Why ?**

5. $\overline{Y} \xrightarrow{a.s} p$ by the SLLN. $\hat{F}(t) \xrightarrow{a.s} F(t)$ by the SLLN.
6. $\sqrt{n}(\hat{F}(t) - F(t)) \xrightarrow{D} N(0, F(t)(1 - F(t)))$ **Why ??**

Remark. Since $E(\hat{F}(t)) = F(t) < \infty$ and $nV(\hat{F}(t)) = F(t)(1 - F(t)) < \infty$, thus we can apply the SLLN and CLT.

the density $\hat{f}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i = t)$,
$\int t d\hat{F}(t) = \sum_{i=1}^{n} t\hat{f}(t) = ?$
$\int t^2 d\hat{F}(t) = \sum_{i=1}^{n} t^2\hat{f}(t) = ?$
$\int (t - \overline{X})^2 d\hat{F}(t) = ?$

Homework: Answer the following questions: If X_1, ..., X_n are i.i.d. from Cauchy, where

$$f(x) = (\pi(1 + (x)^2))^{-1},$$

then $\int_{-\infty}^{\infty} xf(x)dx = 0$ as $xf(x)$ is odd, $\int_{-\infty}^{\infty} |x|f(x)dx = \frac{2}{\pi}\ln(1 + x^2)|_0^{\infty} = \infty$.
$E(X) = ?$
$\overline{X} \xrightarrow{a.s} \mu_X$? (Yes, No, Not sure, explain).
$\sqrt{n}(\overline{X} - \mu_X) \xrightarrow{D} N(0, \tau^2)$? (Yes, No, Not sure, explain).
$\hat{F}(t) \xrightarrow{a.s} F(t)$?
$\sqrt{n}(\hat{F}(t) - F(t)) \xrightarrow{D} N(0, F(t)(1 - F(t)))$?
$\sigma^2_{\hat{F}(t)} = ?$
$\hat{\sigma}^2_{\hat{F}(t)} = ?$
$\hat{\sigma}^2_{\hat{F}(t)} \to ?$
$n\hat{\sigma}^2_{\hat{F}(t)} \xrightarrow{a.s} ?$
$\sqrt{n}(\hat{\sigma}^2_{\hat{F}(t)} - \sigma^2_{\hat{F}(t)}) \xrightarrow{D} ?$
$E(\hat{\sigma}^2_{\hat{F}(t)}) = ?$
$V(\hat{\sigma}^2_{\hat{F}(t)}) = ?$

7. $\hat{F}(t)$ is admissible w.r.t. the squared error loss and the weighted squared error loss

$$L(F(t), \tilde{F}(t)) = \frac{(\tilde{F}(t) - F(t))^2}{F(t)(1 - F(t))}. \tag{1}$$

   That is, if
   (a) t is fixed,
   (b) $\tilde{F} \in \mathcal{A}$, the collection of all estimators of the cdf F,
   (c) $R(F(t), \tilde{F}(t)) = E(L(F(t), \tilde{F}(t)))$,
   (d) Θ_o is the collection of all cdf's,
   (e) $R(F(t), \tilde{F}(t)) \leq R(F(t), \hat{F}(t))\ \forall\ F \in \Theta_o$,
   then $R(F(t), \tilde{F}(t)) = R(F(t), \hat{F}(t))\ \forall\ F \in \Theta_o$.
   No estimator can dominate $\hat{F}$.
8. $\hat{F}(t)$ is minimax w.r.t. the weighted squared error loss in Eq. (1).
   That is, if (a), (b), (c) and (d) hold, then
   $\sup_F R(F(t), \hat{F}(t)) = \inf_{\tilde{F}\in A} \sup_F R(F(t), \tilde{F}(t))$. $\hat{F}$ is the best in the worst scenario.

Remark. Since $\hat{F}$ is a functional, the functional properties are as follows.

9. $\sup_t |\hat{F}(t) - F(t)| \xrightarrow{a.s} 0$ (uniform strong consistency) (Glivenko-Cantelli Theorem).
10. $\sqrt{n}(\hat{F} - F) \xrightarrow{D} W$ where W is a Gaussian process with the covariance specified in Part 4.
11. $\hat{F}$ is inadmissible w.r.t the loss function

$$L(F, a) = \int (F(t) - a(t))^2 dF(t)$$

   and the parameter space being the collection of all continuous cdfs (Aggarwal (1955)).

12. $\hat{F}$ is not minimax w.r.t the loss function

$$L(F,a) = \int (F(t) - a(t))^2 dW(t)$$

but is minimax w.r.t the loss function

$$L(F,a) = \int \frac{(F(t) - a(t))^2}{F(t)(1-F(t))} dW(t)$$

where W is a finite measure, and the parameter space being the collection of all cdfs (Phadia (1973)).

13. $\hat{F}$ is admissible w.r.t the loss function

$$L(F,a) = \int (F(t) - a(t))^2 dW(t)$$

where W is a finite measure, and the parameter space being the collection of all cdfs (Cohen and Kuo (1985)).

14. $\hat{F}$ is admissible w.r.t the loss function

$$L(F,a) = \int (F(t) - a(t))^2 dF(t)$$

and the parameter space being the collection of all cdfs (Brown (1985)).

15. Whether $\hat{F}$ is admissible w.r.t the loss function

$$L(F,a) = \int \frac{(F(t) - a(t))^2}{F(t)(1-F(t))} dF(t)$$

and the parameter space being the collection of all continuous cdfs was an open question between 1950's and 1980's. Yu (1989) shows that it is admissible if $n = 1$ or 2, and is inadmissible if $n \geq 3$.

16. $\hat{F}$ is minimax w.r.t the loss function

$$L(F,a) = \int \frac{(F(t) - a(t))^2}{F(t)(1-F(t))} dF(t)$$

and the parameter space being the collection of all continuous cdfs was an longstanding conjecture between 1950's and 1980's. Yu and Chow (1991) shows that it is indeed minimax.

Chapter 12. Decision Theory

§12.1. Introduction.

There are several definitions of the optimality for an estimator.

Large samples:

1. consistency,
2. efficiency.

Small samples:

3. unbiasedness,
4. UMVUE,
5. Bayesian,
6. admissibility,
7. Minimaxity.

The last two has just been briefly mentioned and will be studied here. The last three all belong to the decision theory frame work.

Recall in §7.3.4 that a decision problem consists of

data $\mathbf{X}$ from the sample space $\mathcal{X}$ ($\mathbf{X} = \mathbf{x} \in \mathcal{X}$), with density function $f_{\mathbf{X}}(\mathbf{x};\theta)$,
the parameter θ from the parameter space Θ,
an action a from the action space $\mathcal{A}$,
a loss function $L(\theta, a)$ on $\Theta \times \mathcal{A}$, *e.g.* $|a-\theta|$, $(a-\theta)^2$, $\frac{(a-\theta)^2}{\theta(1-\theta)}$,
a (nonrandomized) decision rule d: $\mathcal{X} \mapsto \mathcal{A}$. the risk function of d: $R(\theta, d) = E(L(\theta, d(\mathbf{X}))$.

Example 1 Suppose that X_1, ..., X_n are i.i.d. from $N(\mu, \sigma^2)$. Consider estimation of σ^2. This can be viewed as a decision problem.

$\mathbf{X}$ =(X_1, ..., X_n), from $\mathcal{X} = \mathcal{R}^n$,

$\theta = (\mu, \sigma^2)$, in $\Theta = (-\infty, \infty) \times (0, \infty)$,

$\mathcal{A} = [0, \infty)$,

an action is an estimate of σ^2, say $a \in \mathcal{A}$,

a decision rule is an estimator of σ^2, say d: $\mathcal{R}^n \mapsto [0, \infty)$,

a loss function L is the squared error $L(\theta, a) = (a - \sigma^2)^2$.

Recall that under this set-up:

$R(\theta, d) = \text{MSE}(d) = E(L(\theta, d(\mathbf{X})) = E((d(\mathbf{X}) - \sigma^2)^2)$.

Two decision rules (estimators) S^2 and $\hat{\sigma}^2$,

where $S^2 = \frac{1}{n-1}\sum_i (X_i - \overline{X})^2$ and $\hat{\sigma}^2 = \frac{n-1}{n} S^2 \qquad (= \overline{X^2} - (\overline{X})^2)$.

S^2 is the UMVUE of σ^2.

$\hat{\sigma}^2$ is biased.

$R(\theta, S^2) = Var(S^2) = \frac{\sigma^4}{(n-1)^2} 2(n-1) = 2\frac{\sigma^4}{(n-1)}$,

as $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$.

$R(\theta, \hat{\sigma}^2) = Var(\hat{\sigma}^2) + (bias(\hat{\sigma}^2))^2 = (\frac{n-1}{n})^2 Var(S^2) + (-\sigma^2/n)^2$

$= \frac{\sigma^4}{n}[2 - 1/n] < \frac{\sigma^4}{n-1}2$.

Thus $R(\theta, \hat{\sigma}^2) < R(\theta, S^2)$.

Hence the MLE $\hat{\sigma}^2$ is preferable over S^2 in terms of the MSE.

Question: Can we find an estimator that is the best w.r.t. the MSE ?

Answer: No ! In Example 1, idealy, $R(\theta, d) = 0 \; \forall \; \theta$. However,

$0 = R(\theta, d) = E((d(\mathbf{X}) - \sigma^2)^2)$ for a given $\sigma^2 => P(d(\mathbf{X}) = \sigma^2) = 1$ **Why ??**

Thus, for each estimator d, $\exists$ a θ_o such that $R(\theta_o, d) > 0$.

Set $\tilde{\sigma}^2(\mathbf{x}) = \sigma_o^2 \; \forall \; \mathbf{x}$, then $R(\theta_o, \tilde{\sigma}^2) = 0 < R(\theta_o, d)$.

Thus the usual global optimality ($R(\theta, \delta) \le R(\theta, d) \; \forall \; (\theta, d)$) is not applicable. In decision theory, two other types of optimalitiy are considered: admissibility and minimaxity.

Remark. Decision theory can be applied to point estimation, as well as to hypothesis testing and confidence interval.

Two classical textbooks in decision theory:

Mathematical Statistics, a Decision Theory approach, by Thomas Ferguson.

Statistical Decision Theory and Bayesian Analysis, by James Berger.

§**12.2. Admissibility.**

Definition. Let δ and d be two decision rules.

δ is as good as d if $R(\theta, d) \ge R(\theta, \delta) \; \forall \; \theta$.

δ is better than d if $\begin{cases} R(\theta, d) \ge R(\theta, \delta) & \forall \; \theta \\ R(\theta, d) > R(\theta, \delta) & \text{for at least one } \theta. \end{cases}$

In the latter case, the decision rule d is said to be inadmissible.

If a decision rule is not inadmissible, we say that it is admissible.

Example 1 (continued).

S^2 is inadmissible, even though it is UMVUE.

$\hat{\sigma}^2$ is biased, but it is better than S^2 in terms of the MSE.

$\tilde{\sigma}^2(\mathbf{X}) \equiv c \; (> 0)$ is admissible.

The example of $\tilde{\sigma}^2$ suggests that admissibility may not be an appealing property, but it is clear that inadmissible estimators are definitely not desirable, as far as the risk is concerned.

Question: How to determine that an estimator is admissible ?

Answer:

(1) By definition as in Example 1,

(2) by the following theorem.

Theorem 1. *Suppose that the following conditions hold:*

1. *Θ is a subset of the real line;*
2. *$R(\theta, d)$ is continuous in θ for each decision rule d;*
3. *π is a prior density on Θ such that $\int_{\theta_o - \epsilon}^{\theta_o + \epsilon} \pi(\theta) d\theta > 0 \; \forall \; \epsilon > 0, \; \forall \; \theta_o \in \Theta$;*
4. *δ^π is the Bayes estimator w.r.t. π and has a finite Bayes risk $r(\pi, \delta^\pi)$.*

Then δ^π is admissible.
Example 2. Suppose that $X \sim bin(n,p)$ and $L(p,a) = \frac{(a-p)^2}{p(1-p)}$. We have shown that $\hat{p} = X/n$ is Bayesian w.r.t. L and $\pi \sim U(0,1)$ (§7.3.4). **Do the 4 conditions in Theorem 1 hold ?** $\hat{p}$ is admissible.
Remark. If $\pi(\theta)$ is a non-negative function of θ and $\int \pi(\theta)d\theta = \infty$, it is called an improper prior density of θ. Theorem 1 is still applicable if π is an improper prior density.
Example 2 (continued). Suppose that $X \sim bin(n,p)$ and $L(p,a) = (a-p)^2$. Show that $\hat{p} = \frac{X}{n}$ is admissible.
Sol. Two ways:

(1) Definition $\vdash$: $R(\hat{p},p) \le R(\tilde{p},p) => R(\hat{p},p) = R(\tilde{p},p)\ \forall\ p \in \Theta$;

(2) Theorem 1, main condition: $\hat{p}$ is a Bayes estimator w.r.t. L and a prior $\pi(\cdot)$.

Notice that if $\pi(p) = \frac{1}{p(1-p)}$, $p \in (0,1)$, π is not a proper prior.
However, the "Bayes estimator " w.r.t. π exists and Theorem 1 is also applicable to non-proper prior π.
To obtain the Bayes estimator, it suffices to solve

$$d(x) = \text{argmin}_a E(L(p,a)|X=x).$$

$$E(g(Y)|X=x) = \begin{cases} \int g(y) f_{Y|X}(y|x)dy & \text{if cts} \\ \sum_y g(y) f_{Y|X}(y|x) & \text{if dis} \end{cases} \qquad \mathbf{g(y)=\ ?}$$

$f_{Y|X}(y|x) = f(x,y)/f_X(x)$ and $f(x,y) = f_{X|Y}(x|y)f_Y(y) = ??$ $f_Y(\cdot) = \pi(\cdot)$. $f_{X|Y}(x|y) = ?$
$f_{X|p}(x|p)\pi(p) \propto p^{x-1}(1-p)^{n-x-1}$.
$\pi(p|x) \propto p^{x-1}(1-p)^{n-x-1}$.

If $x \in \{1,...,n-1\}$, $\pi(p|x)$ can be viewed as a $beta(x, n-x)$ density. $beta(\alpha,\beta)$, $\alpha,\beta > 0$.
$E((L(p,a)|X=x) = \int_0^1 (a-p)^2 cp^{x-1}(1-p)^{n-x-1}dp$ is finite if $x \notin \{0,n\}$.$(= a^2 - 2aE(p|X=x) + E(p^2|X=x))$
It is minimized by $a = E(p|X=x) = \frac{\alpha}{\alpha+\beta} = \frac{x}{x+n-x} = x/n$. if $x \notin \{0,n\}$.
If $x = 0$ then $E(L(p,a)|X=x)$ is finite iff $a = 0 = x/n$.
If $x = n$ then $E(L(p,a)|X=x)$ is finite iff $a = 1 = x/n$.
Thus $\hat{p} = x/n$ is the Bayes estimator w.r.t. to the improper prior π.
Do the other 3 conditions in Theorem 1 hold ?
Consequently, it is admissible.
Remark. Thus $\hat{p}$ is admissible under the weighted squared error loss, admissible under the squared error loss, UMVUE, consistent, efficient.
Remark. Since the edf $\hat{F}(t) = \overline{Y}$, where $Y = \mathbf{1}(X \le t) \sim bin(1, F(t))$, $\hat{F}(t)$ is admissible w.r.t. $L(F(t), \tilde{F}(t)) = \frac{(\tilde{F}(t)-F(t))^2}{F(t)(1-F(t))}$ and $\hat{F}(t)$ is admissible w.r.t. $L(F(t), \tilde{F}(t)) = (\tilde{F}(t) - F(t))^2$.
Example 3. Suppose that $X \sim bin(n,p)$, $L(p,a) = |a-p|$. $\delta(x) = 1/3$. Show that δ is admissible.
Proof. Two possible approaches: (1) Bayes estimator, (2) definition. Since

$$R(p,d) = E|d(X)-p| = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} |d(x)-p|. \tag{1}$$

Bayesian approach: Set $\pi(p) = 1(p = 1/3)$. Bayes estimator $= \inf_d E(R(p,d)) = ??$ Then verify that δ is the Bayes estimator w.r.t. L and π.

Can we apply Theorem 1 ?

The second apporoach: $\vdash$: $R(p,d) \le R(p,\delta)\ \forall\ p \in \Theta$; $=> R(p,\delta) = R(p,d)\ \forall\ p \in \Theta$;
If d is as good as δ, then $R(p,d) \le R(p,\delta)$ for all $p \in [0,1]$.
$R(1/3,\delta) = 0$. $=> 0 \le R(1/3,d) = E|d(X) - 1/3| \le R(1/3,\delta)$. $=> P(|d(X) - 1/3| = 0) = 1$.
Thus $d(x) = \delta(x)$ for all possible x.
Thus δ is admissible. □

Notice that if X_1, ..., X_n are i.i.d. from $N(\theta,\sigma^2)$, then $\overline{X} \sim N(\theta, \sigma^2/n)$.
$\overline{X}$ is the MLE, UMVUE, consistent and efficient.

Is it admissible under the squared error loss ?

It suffices to ask whether X is admissible if $X \sim N(\theta,\sigma^2)$, by setting $n = 1$.
Example 4. Suppose that $X \sim N(\theta,\sigma^2)$ and $L = (a-\theta)^2$. Show that

1. If σ is known, $\hat{\theta} = X$ is admissible.
2. If σ is unknown, $\hat{\theta} = X$ is admissible.

Proof. Case 1. Two ways to prove admissibility: (1) Bayesian, (2) Definition.
Recall that in the Baysian approach, a candidate of the prior is

$$\pi(\theta) \sim N(\mu, \tau^2).$$

$\pi(\theta|x)$ is $N(\mu_*, \sigma_*^2)$, where

$\mu_* = \frac{\tau^2}{\tau^2+\sigma^2}x + \frac{\sigma^2}{\tau^2+\sigma^2}\mu \stackrel{def}{=} (1-\eta)x + \eta\mu$ and $\sigma_*^2 = \tau^2\eta$.

The Bayes estimator of θ is $\delta^\pi = E(\theta|X)$, that is,

$$\delta^\pi(x) = \mu_* = (1-\eta)x + \eta\mu.$$

The second approach needs **MLE=Bayes estimator.**
Q: Can we set $\delta^\pi(x) = x$?

The second approach: $\vdash$: $R(\theta, d) \leq R(\theta, \hat{\theta})\ \forall\ \theta \in \Theta$; $=>$ $R(\theta, d) = R(\theta, \hat{\theta})\ \forall\ \theta \in \Theta$.

Suppose that σ^2 is known and thus the parameter is θ. We shall assume that $\hat{\theta}$ is inadmissible and show that it leads to a contradiction.

If $\hat{\theta}$ is inadmissible, then there is a d such that

$$R(\theta, \hat{\theta}) - R(\theta, d) \begin{cases} \geq 0 & \forall\ \theta \\ = 2c > 0 & \text{for } \theta = \theta_o \end{cases} \tag{1}$$

$R(\theta, d) = \int (\theta - d(x))^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\theta)^2}{2\sigma^2}} dx$ is continuous in θ $\forall$ estimator d,
thus $R(\theta, d) - R(\theta, \hat{\theta})$ is continuous in θ too.
Then by Eq. (1), there is a $b > 0$ such that

$R(\theta, \hat{\theta}) - R(\theta, d) > c$ if $|\theta - \theta_o| < b$.

$$\begin{aligned} r(\pi, \hat{\theta}) - r(\pi, d) =& \int (R(\theta, \hat{\theta}) - R(\theta, d))\pi(\theta) d\theta \quad (\text{let } \pi(\theta) \sim N(\mu, \tau^2)) \\ =& (\int_{-\infty}^{\theta_o - b} + \int_{\theta_o - b}^{\theta_o + b} + \int_{\theta_o + b}^{\infty})(R(\theta, \hat{\theta}) - R(\theta, d))\pi(\theta) d\theta \\ \geq& \int_{\theta_o - b}^{\theta_o + b} c\pi(\theta) d\theta \\ =& \int_{\theta_o - b}^{\theta_o + b} \frac{c}{\sqrt{2\pi\tau^2}} e^{-\frac{\theta^2}{2\tau^2}} d\theta \end{aligned} \tag{2}$$

$$(r(\pi, \hat{\theta}) - r(\pi, d))\tau \geq \int_{\theta_o - b}^{\theta_o + b} \frac{c}{\sqrt{2\pi}} e^{-\frac{\theta^2}{2\tau^2}} d\theta. \tag{3}$$

Letting $\mu = 0$,

$\delta^\pi(X) = (1-\eta)X + \mu\eta \to X$ if $\eta \to 0$, that is, $\tau \to \infty$.

Moreover,

$r(\pi, \delta^\pi) = E(R(\theta, \delta^\pi)) = E(E(((1-\eta)X - \theta)^2|\theta)) = E(E(((1-\eta)X - \theta)^2|X)) = E(E((\mu_* - \theta)^2|X)) = E(Var(\theta|X)) =>$

$$r(\pi, \delta^\pi) = E(\tau^2\eta) = \tau^2\eta. \tag{4}$$

Since $R(\theta, \hat{\theta}) = E((\theta - X)^2) = \sigma^2$,

$$r(\pi, \hat{\theta}) = \sigma^2. \tag{5}$$

(4) and (5) yield

$r(\pi, \delta^\pi) - r(\pi, \hat{\theta}) = \tau^2\eta - \sigma^2 = \tau^2\frac{\sigma^2}{\sigma^2+\tau^2} - \sigma^2 = \sigma^2\frac{\tau^2-\sigma^2-\tau^2}{\sigma^2+\tau^2} = \frac{-\sigma^4}{\sigma^2+\tau^2}$.

$(r(\pi, \delta^\pi) - r(\pi, \hat{\theta}))\tau = \frac{-\sigma^4}{\sigma^2+\tau^2}\tau$.

$$\begin{aligned} \tau\frac{\sigma^4}{\sigma^2+\tau^2} =& \tau(r(\pi, \hat{\theta}) - r(\pi, \delta^\pi)) \\ =& \tau(r(\pi, \hat{\theta}) - r(\pi, d) + r(\pi, d) - r(\pi, \delta^\pi))) \\ \geq& \int_{\theta_o - b}^{\theta_o + b} \frac{c}{\sqrt{2\pi}} e^{-\frac{\theta^2}{2\tau^2}} d\theta + 0 \end{aligned} \tag{6}$$

by

$$\tau(r(\pi, \hat{\theta}) - r(\pi, d)) \geq \int_{\theta_o - b}^{\theta_o + b} \frac{c}{\sqrt{2\pi}} e^{-\frac{\theta^2}{2\tau^2}} d\theta. \tag{3}$$

Letting $\tau \to \infty$ in inequality (6) yields $0 \geq 2bc/\sqrt{2\pi} > 0$.

The contradiction implies that $\hat{\theta}$ is not inadmissible.

Case 2. Now assume σ is unknown, then the parameters are $\gamma = (\theta, \sigma)$. Again we shall suppose that $\hat{\theta}$ is inadmissible and show that it leads to a contradiction.

If $\hat{\theta}$ is inadmissible in such case, there exists an estimator d such that

$$R(\gamma, d) - R(\gamma, \hat{\theta}) \begin{cases} \leq 0 & \forall \ \gamma \\ < 0 & \text{for } \gamma = (\theta_o, \sigma_o) \end{cases}$$

The risk becomes $R(\gamma, d) = E_\gamma((d(X) - \theta)^2)$. It implies that

$$E_{\theta,\sigma}((d(X) - \theta)^2) - E_{\theta,\sigma}((X - \theta)^2) \begin{cases} \leq 0 & \forall \ (\theta, \sigma) = (\theta, \sigma_o) \\ < 0 & \forall \ (\theta, \sigma) = (\theta_o, \sigma_o) \end{cases}$$

That is, $\hat{\theta}$ is inadmissible when $\sigma = \sigma_o$ is fixed. It corresponds to the assumption in part one. It contradicts to the result in part one. Thus $\hat{\theta}$ is admissible when σ is unknown. □.

Notice that if X_1, ..., X_n are i.i.d. from $N(\theta, \sigma^2)$, then $\overline{X} \sim N(\theta, \sigma^2/n)$.
$\overline{X}$ is the MLE, the UMVUE,
consistent,
efficient,
admissible under the squared error loss.

Remark. Recall that assuming $X \sim bin(n, p)$, we had shown that
$\hat{p} = X/n$ is the MLE,
an MME,
the Bayes estimator under the loss $L(p, a) = \frac{(a-p)^2}{p(1-p)}$ w.r.t. the prior $U(0, 1)$.
the Bayes estimator under the loss $L(p, a) = (a - p)^2$ w.r.t. the prior $\frac{1}{p(1-p)}$.
admissible under the weighted squared error loss,
admissible w.r.t the squared error loss,
UMVUE,
consistent,
efficient.

Proposition 1. Suppose that $X \sim bin(n,p)$. Then
(1) $d(x) = \frac{x+\alpha}{n+\alpha+\beta}$ is the Bayes rule w.r.t. prior $\pi(p) = cp^{\alpha-1}(1-p)^{\beta-1}$, and $L(p,a) = (a-p)^2$;
and is admissible with respect to $L(p,a) = (a-p)^2$, where $\alpha, \beta \geq 0$
(2) $\hat{p} = X/n$ is the Bayes rule w.r.t. prior $\pi(p) = cp^{-\alpha-1}(1-p)^{-\beta-1}$ and $L(p,a) = (a-p)^2 p^\alpha (1-p)^\beta$; and is admissible w.r.t. $L(p,a) = (a-p)^2 p^\alpha (1-p)^\beta$, where $\alpha, \beta \geq -1$.
Proof. (1) Since $\pi(p) = cp^{\alpha-1}(1-p)^{\beta-1}$, $\pi(p|x) \sim$ "$beta(x+\alpha, n-x+\beta)$", as

$$f_{X|p}(x|p)\pi(p) \propto p^{x+\alpha-1}(1-p)^{n-x+\beta-1}.$$

The Bayes estimator is $\delta = E(p|X) = \frac{X+\alpha}{n+\alpha+\beta}$. Since the conditions in Theorem 1 hold, δ is admissible.
(2) Notice that one is more interested in whether X/n is admissible. Under the loss
$L(\theta,a) = (a-\theta)^2 p^\alpha (1-p)^\beta$, where $\alpha, \beta \geq -1$.
We have proved the special case of $\alpha = \beta \in \{0, 1\}$.
Let $\pi(p) = p^{-1-\alpha}(1-p)^{-1-\beta}$. Then $\pi(p|x) \propto p^{-1-\alpha+x}(1-p)^{-1-\beta+n-x}$, $p \in (0,1)$.
$E(L(p,a)|X=x) \propto \int_0^1 ((p-a)^2 p^\alpha (1-p)^\beta p^{-1-\alpha+x}(1-p)^{-1-\beta+n-x})dp$
$= \int_0^1 ((\frac{p-a)^2}{p(1-p)} p^x (1-p)^{n-x})dp.$
It can be viewed as the case of squared error loss with the posterior $1/(p(1-p))$, thus the Bayes estimator is X/n, provided that we need to check that the posterior risk is finite if $(\alpha, X) = (-1, 0)$ or $(\beta, X) = (-1, n)$.
$E(L(p,a)|X=x) \propto \int_0^1 ((p-a)^2 p^{-1+x}(1-p)^{-1+n-x})dp.$

12.3. Minimaxity.

Definition. A decision rule δ is called a minimax decision rule if

$$\sup_{\theta\in\Theta} R(\theta,\delta) = \inf_{d\in\mathcal{D}} \sup_{\theta\in\Theta} R(\theta,d), \tag{1}$$

where $\mathcal{D}$ is the collection of all nonrandomized decision rule. A decision rule δ is an equalizer rule if $R(\theta,\delta)$ is constant in θ.

Two typical methods for determining a minimax decision rule are given in the next two theorems.
Theorem 2. *If δ is a Bayes and equalizer rule, then it is minimax.*
Theorem 3. *If δ is admissible and is an equalizer rule, then it is minimax.*
Proof of Theorem 3. Suppose that the equalizer rule δ is admissible.
$\vdash$: If δ is not minimax then it leads to a contradiction.
By the 3 assumptions, $\exists$ a rule d such that $\sup_\theta R(\theta,d) < \sup_\theta R(\theta,\delta)$ (see (1)).
=> $R(\theta,d) \leq \sup_\theta R(\theta,d) < \sup_\theta R(\theta,\delta) = R(\theta,\delta)\ \ \forall\ \theta$. =>

$$R(\theta,d) < R(\theta,\delta)\ \ \forall\ \theta. \tag{2}$$

Then δ is inadmissible, contradicting the assumption that it is admissible. The contradiction implies that δ is minimax. □
Proof of Theorem 2. Let δ be an equalizer Bayes rule w.r.t. the prior π.
$\vdash$: If δ is not minimax, **it leads to a contradiction.**
By the 3 assumptions, $\exists$ a rule d such that Eq. (2) holds. It yields $r(\pi,d) < r(\pi,\delta)$ **Why ??**
contradicting the assumption that δ is the Bayes estimator w.r.t. π. The contradiction implies that δ is minimax. □

Example 1. Suppose that X_1, ..., X_n are i.i.d. from $N(\theta,\sigma^2)$, show that the MLE of θ is minimax under the loss $L = (a-\theta)^2$.
Proof. The MLE is $\hat{\theta} = \overline{X}$. It is admissible
and with constant risk σ^2/n **Why ??**
Thus it is minimax **by Theorem ?** □.
Remark. Under the set up in Example 1, the MLE $\hat{\theta} = \overline{X}$ is UMVUE, consistent, efficient, admissible and minimax (w.r.t. squared error loss).
Example 2. Suppose that $X \sim bin(n,p)$ and the loss function is $L = \frac{(a-p)^2}{p(1-p)}$. Show that the MLE is minimax.
Proof. $\hat{p}$ is the Bayes rule w.r.t the loss and the uniform prior. Moreover, $R(p,\hat{p}) = \frac{p(1-p)}{np(1-p)} = \frac{1}{n}$ and thus it is an equalizer rule. Consequently, it is minimax. □.

Remark. For $bin(n,p)$, the MLE $\hat{p}$ is UMVUE, consistent, efficient, and is admissible and minimax w.r.t loss $\frac{(p-a)^2}{p(1-p)}$. It yields the properties (7) and (8) of the edf.

Example 3. Suppose that $X \sim bin(n,p)$ and the loss function is $L = (a-p)^2$, is the MLE $\hat{p}$ minimax ?

Sol. $\vdash$: $\hat{p}$ is not minimax.

In view of Th 2, try to find an equalizer rule of the form $\tilde{p} = aX + b$, as $\hat{p}$ is the same form.

$R(\theta, \tilde{p}) = V(\tilde{p}) + (bias(\tilde{p}))^2$

$= V(aX+b) + (a(np) + b - p)^2$

$= a^2np(1-p) + (anp + b - p)^2 = -a^2np^2 + a^2np + b^2 - 2pb(1-an) + p^2(1-an)^2$

$= p^2 \underbrace{[-(a^2n) + (1-an)^2]}_{=0} + p \underbrace{[(a^2n) - 2b(1-an)]}_{=0} + b^2 = b^2$ (equalizer rule),

$-(a^2n) + (1-an)^2 = 0$

$=> -(a^2n) + 1 - 2an + (an)^2 = 0$

$=> a^2(-n+n^2) - 2an + 1 = 0$

$=> a = \frac{2n \pm \sqrt{4n^2+4n-4n^2}}{2(-n+n^2)} = \frac{2n\pm\sqrt{4n}}{2(n^2-n)} = \frac{1}{n\pm\sqrt{n}}$

$(a^2n) - 2b(1-an) = 0$ yields $b = \frac{1}{2}\frac{a^2n}{1-an}$

Thus $b = \frac{\pm\sqrt{1/n}}{2(1\pm\sqrt{1/n})}$.

$b = \frac{-\sqrt{1/n}}{2(1-\sqrt{1/n})} => \tilde{p}(0) = b < 0$.

$b = \frac{\sqrt{1/n}}{2(1+\sqrt{1/n})} => \tilde{p}(0) = b > 0$.

Which b to choose ?

$$a = \frac{1/n}{1+\sqrt{1/n}} \text{ and } b = \frac{\sqrt{1/n}}{2(1+\sqrt{1/n})}, \tag{3}$$

$R(p, \tilde{p}) = b^2 = \frac{1/n}{4(1+\sqrt{1/n})^2} < \frac{1}{4n} = \sup_{p\in[0,1]} R(p,\hat{p}) = \sup_{p\in[0,1]} p(1-p)/n$.

Q: Can we say that $\hat{p}$ is not minimax ??

Q: Can we say that $\tilde{p} = aX + b$ is minimax ?

In view of Th 2, need to know whether $\tilde{p}$ is Bayes.

Proposition 1. Suppose that $X \sim bin(n,p)$ and $L(p,\delta) = (\delta - p)^2$. $d(x) = \frac{x+\alpha}{n+\alpha+\beta}$ is Bayes estimator w.r.t. beta(α,β) distribution (and is admissible) for all $\alpha, \beta > 0$.

Corollary. Let α and β satisfy Eq. (3), $a = \frac{1}{n+\alpha+\beta}$ and $b = \frac{\alpha}{n+\alpha+\beta}$. *i.e.*, $\alpha = b/a$ and $\beta = ?$ Then $d = aX + b = \frac{x+\alpha}{n+\alpha+\beta}$ is equalizer and is admissible and minimax.

Remark. For $bin(n,p)$, the MLE $\hat{p}$ is UMVUE, consistent, efficient, and is admissible w.r.t. loss $(p-a)^2p^\alpha(1-p)^\beta$, $\alpha, \beta \geq -1$, and and is minimax w.r.t loss $\frac{(p-a)^2}{p(1-p)}$ but not $(p-a)^2$.

Homework Due Wednesday.

Review : 3 statistical inferences:

(1) point estimation, MLE, MME, Bayes, UMVUE, consistency, SLLN, CLT.

(2) testing hypothesis, NP, LRT test, size and level of a test.

(3) interval estimation. $\{\theta \in I\}$. Pivotal method, LRT method.

Point estimation.

Let X_1, ..., X_n be a random sample from F.

Recall that

(1) $N(\mu, \sigma^2)$: $f_{\mathbf{X}}(\mathbf{x}) \propto \exp(\sum_{i=1}^n X_i \frac{\mu}{\sigma^2} - \sum_{i=1}^n X_i^2 \frac{1}{2\sigma^2})$

(2) $Gamma(\alpha,\beta)$: $f_{\mathbf{X}} \propto \exp(-\sum_{i=1}^n X_i/\beta + (\alpha-1)\sum_{i=1}^n \log X_i)$

(3) $bin(1,p)$: $f_{\mathbf{X}} \propto \exp(\sum_{i=1}^n X_i \log \frac{p}{1-p})$

(4) $NB(r,p)$: $f_{\mathbf{X}} \propto \exp(\sum_{i=1}^n X_i \log(1-p)) \prod_{j=1}^n \binom{r+X_j-1}{X_j}$

(5) $Pois(\lambda)$: $f_{\mathbf{X}} \propto \exp(\sum_{i=1}^n X_i \log \lambda)$

The UMVUE of $E(X)$ is $\hat{\mu} = \overline{X}$. **Why ?**

(6) $beta(\alpha,\beta)$: $f_{\mathbf{X}} \propto \exp(\sum_{i=1}^n(\alpha-1)\log X_i + (\beta-1)\sum_{j=1}^n \log(1-X_i))$ The UMVUE of $E(X)$ is $E(\overline{X}|\overline{\log X}, \overline{log(1-X)})$. **Why ?**

(7) $U(0,b)$: The UMVUE of $E(X)$ is $\hat{\mu} = \frac{n+1}{n}X_{(n)}/2$, not $\overline{X}$. **Why ?**

A special case.

Let $\mathbf{X} \sim Multi(n, \mathbf{p})$, $\mathbf{p} = (p_1, ..., p_5)$. $f_{\mathbf{X}}(\mathbf{x}) = \binom{n}{x_1,x_2,x_3,x_4,x_5} \prod_{i=1}^5 p_i^{x_i}$.

Does it belong to the exponential family ?

$f_{\mathbf{X}}(\mathbf{x}) = \binom{n}{x_1,x_2,x_3,x_4,x_5} \exp(\sum_{i=1}^5 x_i \ln p_i)$

What is the MSS of $\mathbf{p}$?

$f_{\mathbf{X}}(\mathbf{x}) = \binom{n}{x_1,x_2,x_3,x_4,x_5} p_5^n \exp(\sum_{i=1}^4 x_i \ln(p_i/p_5))$.

What is the UMVUE of $\mathbf{p}$? **Why ?**

A3. As in Example 10.3.4, with $\mathbf{X} \sim Multinomial(n, p_1, ..., p_5)$. Set H_0: $p_1 = p_2 = p_5 = 0.01$, $p_3 = 0.5$ v.s. H_1: H_0 is not true.

a. Derive the likelihood ratio test for $n = 1$ and $n = 36$ with level $\alpha = 0.05$.

b. Give an estimate of $P(H_o|H_1)$ when $p_1 = p_2 = p_5 = 0.02$, $p_3 = 0.4$, $n = 36$, using simulation. Present the program.

c. Compute **(not estimate !)** $P(H_o|H_1)$ when $p_1 = p_2 = p_5$, $p_3 = 0.4$, $n = 1$.

Sol. a. Two ways to describe $\mathbf{X} \sim Multinomial(n, p_1, ..., p_5)$ when $n = 1$:

(1) $\mathbf{X} = (X_1, X_2, X_3, X_4, X_5)$,

$f_{\mathbf{X}}(\mathbf{x}) = \frac{n!}{x_1!x_2!x_3!x_4!x_5!} p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4} p_5^{x_5}$, ...??

(2) $f_Y(y) = p_1^{\mathbf{1}(y=1)} p_2^{\mathbf{1}(y=2)} p_3^{\mathbf{1}(y=3)} p_4^{\mathbf{1}(y=4)} p_5^{\mathbf{1}(y=5)} = \begin{cases} p_1 & \text{if } y = 1 \\ ... & \\ p_5 & \text{if } y = 5 \end{cases}$

(3) $\begin{array}{rccccc} y: & 1 & 2 & 3 & 4 & 5 \\ f_Y(y): & p_1 & p_2 & p_3 & p_3 & p_5 \end{array}$.

Which is more convenient ?

If $n = 1$, LRT: $\phi = \mathbf{1}(\lambda \le c)$ with $E_{\mathbf{p}}(\phi) \le 0.05$, $\mathbf{p}$ under H_0.

$$\lambda = \begin{cases} \frac{0.01}{1} & \text{if } X \in \{1,2,5\} \\ \frac{0.47}{1} & \text{if } X = 4 \\ \frac{0.5}{1} & \text{if } X = 3 \end{cases} = 0.01^{X_1+X_2+X_5} 0.47^{X_4} 0.5^{X_3}$$

=> $\phi = \mathbf{1}(Y \in \{1,2,5\}) = \mathbf{1}(X_1 + X_2 + X_5 = 1)$.

Details:

$f_{\mathbf{X}}(\mathbf{x}) = \frac{n!}{x_1!x_2!x_3!x_4!x_5!} p_1^{x_1} p_2^{x_2} p_3^{x_3} p_4^{x_4} p_5^{x_5}$,

$\hat{\mathbf{p}}_0 = (0.01, 0.01, 0.5, 0.47, 0.01)$,

$\hat{\mathbf{p}} = \mathbf{X}/n$ $(= (X_1, X_2, X_3, X_4, X_5)/n)$ **Why ?**

$X_i \sim bin(n, p_i)$,

If $n = 1$,

$\lambda = \frac{0.01^{X_1} 0.01^{X_2} 0.01^{X_5} 0.47^{X_4} 0.5^{X_3}}{X_1^{X_1} X_2^{X_2} X_3^{X_3} X_4^{X_4} X_5^{X_5}} = 0.01^{X_1+X_2+X_5} 0.47^{X_4} 0.5^{X_3}$

$\phi = \mathbf{1}(\lambda \le c)$,

$E(\phi) = P(\lambda \le c) \le 0.05$.

$\lambda = \begin{cases} 0.01 & \text{if } X_1 + X_2 + X_5 = 1 \\ 0.47 & \text{if } X_4 = 1 \\ 0.5 & \text{if } X_3 = 0.5 \end{cases}$

$c = 0.05$? 0.01 ? 0.02 ?

(If $n = 36$? ...... $\phi = \mathbf{1}(-2\ln\lambda \ge \chi^2_{??,0.05}$.)

(3) $\begin{array}{rccccc} y: & 1 & 2 & 3 & 4 & 5 \\ f_Y(y): & p_1 & p_2 & p_3 & p_3 & p_5 \\ \hat{p}_o: & 0.01 & 0.01 & 0.5 & 0.47 & 0.01 \\ \hat{p}_i: & \mathbf{1}(y=1) & \mathbf{1}(y=2) & \mathbf{1}(y=3) & \mathbf{1}(y=4) & \mathbf{1}(y=5) \\ \lambda: & 0.01 & 0.01 & 0.5 & 0.47 & 0.01 \end{array}$.

$\phi = \mathbf{1}(Y \in \{1,2,5\})$

c.

$P(H_0|H_1) = 1 - P(Y \in \{1,2,5\}) = 1 - P(X_1 + X_2 + X_3 = 1) = 1 - 3p$, $p \in [0, \frac{8}{30}]$, when $p_1 = p_2 = p_5 = p$, $p_3 = 0.2$.

If do not impose $p_1 = p_2 = p_5$, $p_3 = 0.4$, $P(H_0|H_1) = 1 - P(Y \in \{1,2,5\}) = 1 - p_1 - p_2 - p_5$, $p_i \ge 0$ and $\sum_{i=1}^5 p_i = 1$. It is a **function of** $(p_1, ..., p_5)$ **under** H_1.

Interval estimation.

Q: The covarage probability and confidence coefficient of $\{\theta \in I\}$?

$P(\theta \in I)$, $\inf_\theta P(\theta \in I)$.

Let X_1, ..., X_n be i.i.d. from $N(\mu, \sigma^2)$.

The confidence interval for μ is $\overline{X} \pm t_{\alpha/2,n-1} S/\sqrt{n}$. I = ? θ= ? Its confidence coefficient ?

$P(\overline{X} - t_{\alpha/2,n-1}S/\sqrt{n} < \mu < \overline{X} + t_{\alpha/2,n-1}S/\sqrt{n}) = P(|\frac{\overline{X}-\mu}{S/\sqrt{n}}| < t_{\alpha/2,n-1}) = 1 - \alpha$.

Let X_1, ..., X_n be i.i.d. from Exp(μ).

$P(\overline{X} - 1.96\overline{X}/\sqrt{n} < \mu < \overline{X} + 1.96\overline{X}/\sqrt{n}) \approx 0.95$ for given μ.

Does its confidence coefficient ≈ 0.95 ?

Not always.

$P(\overline{X} - 1.96\overline{X}/\sqrt{n} < \mu < \overline{X} + 1.96\overline{X}/\sqrt{n}) = P(\frac{\overline{X}}{\mu}(1 - 1.96/\sqrt{n}) < 1 < \frac{\overline{X}}{\mu}(1 + 1.96/\sqrt{n})) = P(\frac{1}{1.96+1/\sqrt{n}} < \frac{\overline{X}}{\mu} < \frac{1}{1-1.96/\sqrt{n}}) = P(\frac{n}{1+1.96/\sqrt{n}} < \frac{\sum_{i=1}^n X_i}{\mu} < \frac{n}{1-1.96/\sqrt{n}}) = \int_{\frac{n}{1+1.96/\sqrt{n}}}^{\frac{n}{1-1.96/\sqrt{n}}} \frac{x^{n-1}e^{-x}}{\Gamma(n)} dx$.

```
> n=100
> pgamma(n/(1-2/sqrt(n)),shape=n)- pgamma(n/(1+2/sqrt(n)),shape=n)
[1] 0.949306
```

Does its confidence coefficient ≈ 0.95 **?**

Example 2. Suppose that $X \sim bin(n,p)$, $n = 3$. Observe $X = 1$. $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n}$ yields $\frac{1}{3} \pm 1.96\sqrt{\frac{2}{27}}$. (a) **Is it a 95% CI for** p **?** (b) How about if $X = 1$ and $n \ge 20$?

Sol. Confidence coefficient of CI = $\inf_p P_p(p \in I)$. **???**

The CI is of the form

$$I = \begin{cases} 0 \pm 0 & \text{if } X = 0 \\ \frac{1}{3} \pm 1.96\sqrt{2/27} & \text{if } X = 1 \\ \frac{2}{3} \pm 1.96\sqrt{2/27} & \text{if } X = 2 \\ 1 \pm 0 & \text{if } X = 3 \end{cases} = \begin{cases} [0,0] & \text{if } X = 0 \\ [0, \frac{1}{3} + a] & \text{if } X = 1 \\ [\frac{2}{3} - a, 1] & \text{if } X = 2 \\ [1,1] & \text{if } X = 3 \end{cases} \quad (a \approx 0.53).$$

$\frac{1}{3} - a < 0 < \frac{2}{3} - a < \frac{1}{3} + a < 1 < \frac{2}{3} + a$.

The coverage probability is $P_p(p \in I)$

$$= \begin{cases} P_p(X = 0 \text{ or } 1) & \text{if } p = 0 \\ P_p(X = 1) & \text{if } p \in (0, \frac{2}{3} - a) \\ P_p(X = 1 \text{ or } 2) & \text{if } p \in [\frac{2}{3} - a, \frac{1}{3} + a] \\ P_p(X = 2) & \text{if } p \in (\frac{1}{3} + a, 1) \\ P_p(X = 2 \text{ or } 3) & \text{if } p = 1 \end{cases} = \begin{cases} (1-p)^3 & \text{if } p = 0 \\ 3p(1-p)^2 & \text{if } p \in (0, \frac{2}{3} - a) \\ 3p(1-p) & \text{if } p \in [\frac{2}{3} - a, \frac{1}{3} + a] \\ 3p^2(1-p) & \text{if } p \in (\frac{1}{3} + a, 1) \\ p^3 & \text{if } p = 1 \end{cases}$$

Confidence coefficient of CI = $\inf_p P_p(p \subset I) = 0$.

(b). The approximate 95% CI is $I = \begin{cases} \{0\} & \text{if } X = 0 \\ \frac{1}{n} \pm 1.96\sqrt{\frac{n-1}{n^3}} & \text{if } X = 1 \\ \frac{2}{n} \pm 1.96\sqrt{\frac{2(n-2)}{n^3}} & \text{if } X = 2 \\ \cdots & \cdots \end{cases}$

The coverage probability is

$$P_p(p \in I) = \begin{cases} P_p(X = 0 \text{ or } 1, \text{ or }) & \text{if } p = 0 \\ P_p(X \neq 0 \text{ and } X = 1, \text{ or...}) & \text{if } p \in (0,a), \text{ where } a < 1 \\ \cdots & \cdots \end{cases}$$

Confidence coefficient of the approximate 95% CI = $\inf_p P_p(p \subset I) = 0$.

Does its confidence coefficient ≈ 0.95 **?**

Homework solution.

Homework Solution Week 15

1. Answer the following questions:

$$\hat{F}(t) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}(X_i \le t) = \overline{Y}, \text{ where } Y_i = \mathbf{1}(X_i \le t)$$

X_1, ..., X_n are i.i.d. from Cauchy,

(A) $E(X) = ?$

$0, \infty$, DNE.

$f(x) = (\pi(1+(x)^2))^{-1}$ and $\int_{-\infty}^{\infty} xf(x)dx = 0$ **as** $xf(x)$ **is odd ??**

Remark. $\int g(x)dx$ exists $=>$ $\int_t^{\infty} |g(x)|dx < \infty$ for all t.

$\int_{-\infty}^{\infty} |x|f(x)dx = \frac{2}{\pi}\ln(1+x^2)|_0^{\infty} = \infty$.

(B) $\overline{X} \xrightarrow{a.s.} \mu_X$? (Yes, No, Not sure, explain).

No, as μ_X does not exist.

(C) $\sqrt{n}(\overline{X} - \mu_X) \xrightarrow{D} N(0, \tau^2)$? (Yes, No, Not sure, explain).

No, as μ_X does not exist.

(D) $\hat{F}(t) \xrightarrow{a.s.} F(t)$?

Yes, by SLLN, as $E(\hat{F}(t)) = F(t)$.

(E) $\sqrt{n}(\hat{F}(t) - F(t)) \xrightarrow{D} N(0, F(t)(1-F(t)))$?

Yes. By CLT, as $V(\hat{F}(t)) = F(t)(1-F(t))/n < \infty$ and $E(\hat{F}(t)) = F(t)$.

(f) $\sigma^2_{\hat{F}(t)} = ?$

$F(t)(1-F(t))/n$.

(g) $\hat{\sigma}^2_{\hat{F}(t)} = ?$

$\hat{F}(t)(1-\hat{F}(t))/n$.

(H) $\hat{\sigma}^2_{\hat{F}(t)} \to 0$?

Proof (1): Yes, as $\hat{\sigma}^2_{\hat{F}(t)} = \hat{F}(t)(1-\hat{F}(t))\frac{1}{n} \xrightarrow{a.s.} F(t) \times (1-F(t)) \times 0 = 0$ by the continuous mapping theorem with $g(x,z) = x(1-x)z$. **x, z = ? Is it OK ?**

Proof (2): Yes, as $|\hat{\sigma}^2_{\hat{F}(t)}| = |\frac{\hat{F}(t)(1-\hat{F}(t))}{n}| \leq \frac{1}{n} \to 0$.

(I) $n\hat{\sigma}^2_{\hat{F}(t)} \xrightarrow{a.s.}$?

$n\hat{\sigma}^2_{\hat{F}(t)} \xrightarrow{a.s.} (F(t)(1-F(t))$, ...

as $g(x) = x(1-x))$ **is cts and** $n\hat{\sigma}^2_{\hat{F}(t)} = g(\overline{Y})$**, where** $\overline{Y} = \hat{F}(t)$**.**

(J) $\sqrt{n}n(\hat{\sigma}^2_{\hat{F}(t)} - \sigma^2_{\hat{F}(t)}) \xrightarrow{D}$?

Delta method: Let $g(x) = x(1-x)$, then $\hat{\sigma}^2_{\hat{F}(t)} = \frac{1}{n}g(\hat{F}(t))$.

$\sqrt{n}n(\hat{\sigma}^2_{\hat{F}(t)} - \sigma^2_{\hat{F}(t)}) = \sqrt{n}(g(\hat{F}(t)) - g(F(t)))$.

$\sqrt{n}(g(\hat{F}(t)) - g(F(t))) \xrightarrow{D} N(0, \tau^2)$, $\tau^2 = (1-2F(t))^2F(t)(1-F(t))$.

(K) $E(\hat{\sigma}^2_{\hat{F}(t)}) = ?$

$= E(\hat{F}(t) - (\hat{F}(t))^2)/n = (F(t) - (F(t))^2)/n$ **??**

$E(\hat{\sigma}^2_{\hat{F}(t)}) = E(\overline{Y} - (\overline{Y})^2)/n$

$= (E(\overline{Y}) - (\sigma^2_{\overline{Y}} + (E(\overline{Y}))^2))/n$

$= \frac{1}{n}(F(t) - (\sigma^2_{\hat{F}(t)} + (E(\hat{F}(t))^2)))$

$= \frac{1}{n}[F(t) - ((F(t))^2 + F(t)(1-F(t))/n)]$

$= \frac{1}{n}[F(t)(1-F(t))(1+1/n)]$

(l) $V(\hat{\sigma}^2_{\hat{F}(t)}) = ?$

$V(\hat{\sigma}^2_{\hat{F}(t)}) = \frac{1}{n^2}V(\overline{Y} - (\overline{Y})^2)$

$= \frac{1}{n^2}[V(\overline{Y}) + V((\overline{Y})^2) - 2(E((\overline{Y})^3) - E(\overline{Y})E((\overline{Y})^2))]$.

$V((\overline{Y})^2) = E((\overline{Y})^4) - (E((\overline{Y})^2))^2$.

$E((\overline{Y})^2) = V(\overline{Y}) + (E(\overline{Y}))^2$=......

To compute $E((\overline{Y})^3)$, notice

$\overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$, $Y_i \sim bin(1,p)$ and $W = \sum_{i=1}^n Y_i \sim bin(n,p)$.

$E((\overline{Y})^3) = \frac{1}{n^3}E(W^3)$,

$M_W(t) = E(e^{Wt}) = (E(e^{Y_i t}))^n = (q + pe^t)^n$.

Consider either of these two methods:

(1) $E(W^3) = \frac{d^3 M_W(t)}{dt^3}|_{t=0}$.

(2) $E(W(W-1)(W-2)) = \sum_{k=0}^n k(k-1)(k-2)\binom{n}{k}p^k q^{n-k}$, ...

$E(W(W-1)(W-2)) = E(W^3) - 3E(W^2) + 2E(W)$.

$\sum_{k=0}^n k(k-1)(k-2)\binom{n}{k}p^k q^{n-k}$

$= \sum_{k=3}^n k(k-1)(k-2)\binom{n}{k}p^k q^{n-k}$

$= \sum_{k=3}^n n(n-1)(n-2)\frac{(n-3)!}{(k-3)!}p^k q^{n-k}$, ...

Similarly for $E((\overline{Y})^4)$.

$E((\overline{Y})^4) = \frac{1}{n^4}E(W(W-1)(W-2)(W-3)) - \cdots + 6E(W)]$

$= \frac{1}{n^4}M_W^{(4)}(0)$

$= \frac{np}{n^4}[1 - 7p + 7np + 12p^2 - 18np^2 + 6n^2p^2 - 6p^3 + 11np^3 - 6n^2p^3 + n^3p^3]$.

(m) $\sqrt{n}(\hat\sigma_{\hat F(t)} - \sigma_{\hat F(t)}) \xrightarrow{D}$?

Ans. 0, as

$\sqrt{n}(\hat\sigma_{\hat F(t)} - \sigma_{\hat F(t)}) = \frac{1}{\sqrt{n}}\sqrt{n}(\sqrt{\hat F(t)(1-\hat F(t))} - \sqrt{F(t)(1-F(t))}) = W_n Z_n$, where $W_n = \frac{1}{\sqrt{n}}$ and

$Z_n = \sqrt{n}(\sqrt{\hat F(t)(1-\hat F(t))} - \sqrt{F(t)(1-F(t))})$

$= \sqrt{n}(g(\hat F(t)) - g(F(t))) \xrightarrow{D} N(0, \tau^2)$

and $g(x) = \sqrt{x(1-x)}$.

A3. As in Example 10.3.4, with $\mathbf{X} \sim Multinomial(n, p_1, ..., p_5)$. Set

H_0: $p_1 = p_2 = p_5 = 0.01$, $p_3 = 0.5$ v.s. H_1: H_0 is not true.

a. Derive the likelihood ratio test for $n = 1$ and $n = 36$ with level $\alpha = 0.05$.

b. Give an estimate of $P(H_o|H_1)$ when $p_1 = p_2 = p_5 = 0.02$, $p_3 = 0.4$, $n = 36$, using simulation. Present the program.

c. Compute **(not estimate !)** $P(H_o|H_1)$ when $p_1 = p_2 = p_5$, $p_3 = 0.4$, $n = 1$.

Sol. a. Two ways to describe $\mathbf{X} \sim Multinomial(n, p_1, ..., p_5)$ when $n = 1$:

(1) $\mathbf{X} = (X_1, X_2, X_3, X_4, X_5)$,

$f_{\mathbf{X}}(\mathbf{x}) = \frac{n!}{x_1!x_2!x_3!x_4!x_5!}p_1^{x_1}p_2^{x_2}p_3^{x_3}p_4^{x_4}p_5^{x_5}$, ...??

(2) $f_Y(y) = p_1^{\mathbf{1}(y=1)}p_2^{\mathbf{1}(y=2)}p_3^{\mathbf{1}(y=3)}p_4^{\mathbf{1}(y=4)}p_5^{\mathbf{1}(y=5)} = \begin{cases} p_1 & \text{if } y = 1 \\ ... & \\ p_5 & \text{if } y = 5 \end{cases}$

(3) $\begin{array}{cccccc} y: & 1 & 2 & 3 & 4 & 5 \\ f_Y(y): & p_1 & p_2 & p_3 & p_3 & p_5 \end{array}$.

Which is more convenient ?

If $n = 1$, LRT: $\phi = \mathbf{1}(\lambda \le c)$ with $E_{\mathbf{p}}(\phi) \le 0.05$, $\mathbf{p}$ under H_0.

$$\lambda = \begin{cases} \frac{0.01}{1} & \text{if } Y \in \{1,2,5\} \\ \frac{0.47}{1} & \text{if } Y = 4 \\ \frac{0.5}{1} & \text{if } Y = 3 \end{cases} = 0.01^{X_1+X_2+X_5}0.47^{X_4}0.5^{X_3}$$

=> $\phi = \mathbf{1}(Y \in \{1,2,5\}) = \mathbf{1}(X_1 + X_2 + X_5 = 1)$.

Details:

$f_{\mathbf{X}}(\mathbf{x}) = \frac{n!}{x_1!x_2!x_3!x_4!x_5!}p_1^{x_1}p_2^{x_2}p_3^{x_3}p_4^{x_4}p_5^{x_5}$,

$\hat{\mathbf{p}}_0 = (0.01, 0.01, 0.5, 0.47, 0.01)$,

$\hat{\mathbf{p}} = \mathbf{X}/n$ $(= (X_1, X_2, X_3, X_4, X_5)/n)$ **Why ?**

$X_i \sim bin(n, p_i)$,

If $n = 1$,

$\lambda = \frac{0.01^{X_1}0.01^{X_2}0.01^{X_5}0.47^{X_4}0.5^{X_3}}{X_1^{X_1}X_2^{X_2}X_3^{X_3}X_4^{X_4}X_5^{X_5}} = 0.01^{X_1+X_2+X_5}0.47^{X_4}0.5^{X_3}$

$\phi = \mathbf{1}(\lambda \le c)$,

$E(\phi) = P(\lambda \le c) \le 0.05$.

$\lambda = \begin{cases} 0.01 & \text{if } X_1+X_2+X_5=1 \\ 0.47 & \text{if } X_4=1 \\ 0.5 & \text{if } X_3=1 \end{cases}$
$c = 0.05$? 0.01 ? 0.02 ?

(If $n = 36$? ...... $\phi = \mathbf{1}(-2\ln\lambda \geq \chi^2_{??,0.05}.)$

$$\begin{array}{llccccc} & y: & 1 & 2 & 3 & 4 & 5 \\ & f_Y(y): & p_1 & p_2 & p_3 & p_3 & p_5 \\ (3) & \hat{p}_o: & 0.01 & 0.01 & 0.5 & 0.47 & 0.01 \\ & \hat{p}_i: & \mathbf{1}(y=1) & \mathbf{1}(y=2) & \mathbf{1}(y=3) & \mathbf{1}(y=4) & \mathbf{1}(y=5) \\ & \lambda: & 0.01 & 0.01 & 0.5 & 0.47 & 0.01 \end{array}.$$

$\phi = \mathbf{1}(Y \in \{1,2,5\})$

c.
$P(H_0|H_1) = 1 - P(Y \in \{1,2,5\}) = 1 - P(X_1+X_2+X_5 = 1)$
$= 1 - 3p$, $p \in [0, \frac{6}{30}]$, when $p_1 = p_2 = p_5 = p$, $p_3 = 0.4$.
If do not impose $p_1 = p_2 = p_5$, $p_3 = 0.4$,
$P(H_0|H_1) = 1 - P(Y \in \{1,2,5\}) = 1 - p_1 - p_2 - p_5$, $p_i \geq 0$ and $\sum_{i=1}^5 p_i = 1$.
It is a **function of** $(p_1, ..., p_5)$ **under** H_1.

b. How to get one sample $(X_1, X_2, X_3) \sim Muiltnomial(7, p_1, p_2, p_3)$?
```
p=c(1,4,5)/10
x=rmultinom(1,7,p)
x[,]
[1] 0 6 1
```
How to get 20 samples $(X_1, X_2, X_3) \sim Muiltnomial(100, p_1, p_2, p_3)$?
```
x=rmultinom(20,100,p)
x[,1] = ?
```
Dimension of x ?

$\hat{P}(H_0|H_1) =$? by simulation:
Choose one $\mathbf{p}$ under H_1, generate data and do the LRT test. Repeat 20+ and get the average.
In your report of simulation, report the value of the parameters and sufficient statistics.

3. In each of the cases in the three problems (40, 41, 48), generate a sample of size 100 and construct the specified 95% CI in the problems. You should state your assumption and give the sufficient statistic.
```
?rnbinom
?rpois
?rmultinom
```

10.48. Let $\mathbf{U}_i$ $(= (X_i, Y_i))$'s be i.i.d. from $N(\vec{\mu}, \Sigma)$,
where (μ, Σ) **are parameters**. CI for $\theta = \mu_x/\mu_y$?
Sol. Two ways: (1) MLE, (2) $Z = X - \theta Y$.
(1) MLE: The MLE of the parameters are $(\hat{\mu}, \hat{\Sigma})$, where $\hat{\mu} = (\overline{X}, \overline{Y})$ and $\hat{\Sigma} = \overline{\mathbf{U}'\mathbf{U}} - \overline{\mathbf{U}}'\overline{\mathbf{U}}$.
The MLE of θ is $\hat{\theta} = \frac{\overline{X}}{\overline{Y}}$ $(= g(\overline{\mathbf{U}})$, where $g(x,y) = x/y)$.
$\bigtriangledown g = (1/y, -x/y^2)$
$\hat{\sigma}^2 = (1/\overline{Y}, -\overline{X}/\overline{Y})\hat{\Sigma}(1/\overline{Y}, -\overline{X}/\overline{Y})^t/n$. Thus, the CI of θ is $\hat{\theta} \pm 1.96\sqrt{\hat{\sigma}^2}$.
(2) Since Z_i's are i.i.d. from $N(0, \sigma^2)$, as $E(Z) = \mu_x - \theta\mu_y = 0$.
$T = \frac{\overline{Z}}{S_Z/\sqrt{n}} = \frac{\overline{Z}}{\hat{\sigma}_Z/\sqrt{n-1}} \sim t_{n-1}$ distribution.
Solve θ for $\frac{\overline{Z}}{\hat{\sigma}_Z/\sqrt{n-1}} = t_{n-1,\alpha/2}$.
$\overline{Z} = \overline{X} - \theta\overline{Y}$ and

$$\hat{\sigma}_Z^2 = \overline{Z^2} - (\overline{Z})^2 = \hat{\sigma}_X^2 - 2\theta(\overline{XY} - \overline{X}\cdot\overline{Y}) + \theta^2\hat{\sigma}_Y^2 \tag{1}$$

$$(\overline{X} - \theta\overline{Y})^2 = (t_{n-1,\alpha/2})^2(\hat{\sigma}_X^2 - 2\theta(\overline{XY} - \overline{X}\cdot\overline{Y}) + \theta^2\hat{\sigma}_Y^2)/(n-1)$$

This gives the endpoints of the CI for θ.

(3) From $-t_{n-1,\alpha/2}\hat\sigma_Z/\sqrt{n-1} \le \overline{X} - \theta\overline{Y} \le t_{n-1,\alpha/2}\hat\sigma_Z/\sqrt{n-1}$ yields

$$\frac{\overline{X}}{\overline{Y}} - \frac{t_{n-1,\alpha/2}\hat\sigma_Z/\sqrt{n-1}}{\overline{Y}} \le \theta \le \frac{\overline{X}}{\overline{Y}} + \frac{t_{n-1,\alpha/2}\hat\sigma_Z/\sqrt{n-1}}{\overline{Y}}$$

Is this a CI for θ ? (see Eq. (1)).
In report of your simulation, report the value of (μ, Σ) and $(\overline{\mathbf{U}}, \hat\Sigma)$.

2. Prove statement (11): The edf $\hat F$ is inadmissible w.r.t the loss function

$$L(F,a) = \int (F(t) - a(t))^2 dF(t)$$

and the parameter space being the collection of all continuous cdfs (Aggarwal (1955)). **Hint:** $\hat F$ is of the form

$$d(t) = \sum_{i=0}^{n} a_i \mathbf{1}(X_{(i)} \le t < X_{(i+1)}) \tag{1}$$

where $X_0 = -\infty$, $X_{(1)} < \cdots < X_{(n)}$ are order statistics of X_is and $X_{(n+1)} = \infty$. Compute $R(F,d)$ and find the one that minimizes $R(F,d)$ over all possible $d(\cdot)$ as in Eq. (1). You can try $n = 1$ first.

11. $\hat F$ is inadmissible w.r.t the loss function

$$L(F,a) = \int (F(t) - a(t))^2 dF(t)$$

and the parameter space being the collection of all continuous cdfs (Aggarwal (1955)).
Prove statement (11). **Hint:** $\hat F$ is of the form

$$d(t) = \sum_{i=0}^{n} a_i \mathbf{1}(X_{(i)} \le t < X_{(i+1)}) \tag{2}$$

where $X_0 = -\infty$, $X_{(1)} < \cdots < X_{(n)}$ are order statistics of X_is and $X_{(n+1)} = \infty$. Compute $R(F,d)$ and find the one that minimizes $R(F,d)$ over all possible $d(\cdot)$. You can try $n = 1$ first.

Proof of Part 11. Consider $n = 1$ first. $\hat F$ is of the form $d(t) = a + b\mathbf{1}(X \le t)$ or

$$d(t) = a_0\mathbf{1}(t < X) + a_1\mathbf{1}(X \le t) \tag{1}$$

$$\begin{aligned}
R(F,d) =& E(\int (d(t) - F(t)^2 dF(t)) \\
=& E(\int (a_0\mathbf{1}(t<x) + a_1\mathbf{1}(x \le t) - F(t))^2 dF(t)) \\
=& \int\int (a_0\mathbf{1}(t<x) + a_1\mathbf{1}(x \le t) - F(t))^2 dF(t)dF(x) \\
=& \int\int (a_0\mathbf{1}(t<x) + a_1\mathbf{1}(x \le t) - F(t))^2 dF(x)dF(t) \\
=& \int_0^1\int_0^1 (a_0\mathbf{1}(u<y) + a_1\mathbf{1}(y \le u) - u)^2 dydu \quad (y = F(x), \quad u = F(t)) \\
=& \int_0^1\int_0^1 (a_0\mathbf{1}(u<y) - u)^2 dy + \int_0^1 (a_1\mathbf{1}(y \le u) - u)^2 dydu \\
=& \int_0^1\int_u^1 (a_0 - u)^2 dy + \int_0^u (a_1 - u)^2 dydu \\
=& \int_0^1 (1-u)(a_0 - u)^2 + u(a_1 - u)^2 du
\end{aligned}$$

Now try to minimize the risk of this form. Taking derivatives w.r.t. a_0 and a_1 yields

$0 = \int_0^1 (1-u)(a_0-u)du$ and $0 = \int_0^1 u(a_1-u)du$. Thus

$$a_0 = \int_0^1 (1-u)udu / \int_0^1 (1-u)du = \frac{B(2,2)}{B(1,2)} = \frac{\Gamma(2)\Gamma(2)}{\Gamma(2+2)} \frac{\Gamma(1+2)}{\Gamma(1)\Gamma(2)} = 1/3$$

$$a_1 = \int_0^1 u^2du / \int_0^1 udu = B(3,1)/B(2,1) = 2/3$$

One can verify that (a_0, a_1) uniquely minimizes $R(F,d)$ of the form (1), thus $\hat{F}$ is inadmissible.

For arbitrary n,

$$d(t) = \sum_{i=0}^n a_i \mathbf{1}(X_{(i)} \le t < X_{(i+1)}) \tag{2}$$

where $X_0 = -\infty$, $X_{(1)} < \cdots < X_{(n)}$ are order statistics of X_is and $X_{(n+1)} = \infty$.

$$\begin{aligned}
R(F,d) &= E(\int (d(t) - F(t)^2 dF(t)) \\
&= E(\int (\sum_{i=0}^n a_i \mathbf{1}(X_{(i)} \le t < X_{(i+1)}) - F(t))^2 dF(t)) \\
&= \sum_{i=0}^n E(\int (a_i \mathbf{1}(X_{(i)} \le t < X_{(i+1)}) - F(t))^2 dF(t)) \\
&= n! \sum_{i=0}^n \int\int\cdots\int (a_i \mathbf{1}(x_{(i)} \le t < x_{(i+1)}) - F(t))^2 dF(t)dF(x_1)\cdots dF(x_n) \\
&= n! \sum_{i=0}^n \int\int\cdots\int_{x_1<\cdots<x_i\le t<x_{i+1}<\cdots x_n} (a_i - F(t))^2 dF(t)dF(x_1)\cdots dF(x_n) \\
&= n! \sum_{i=0}^n \int\cdots\int_{0<x_1<\cdots<x_i\le t<x_{i+1}<\cdots x_n<1} \int_0^1 (a_i - u)^2 dtdx_1\cdots dx_n \\
&= n! \sum_{i=0}^n \int_0^1 \int\cdots\int_{0<x_1<\cdots<x_i\le t<x_{i+1}<\cdots x_n<1} (a_i - t)^2 dx_1\cdots dx_n dt \\
&= \sum_{i=0}^n \binom{n}{i} \int_0^1 t^i(1-t)^{n-i}(a_i-t)^2 dt
\end{aligned}$$

Taking derivatives w.r.t. a_i and setting it to be zero yield

$$\int_0^1 t^i(1-t)^{n-i}(a_i-t)dt = 0$$

$$a_i = B(i+2, n-i+1)/B(i+1, n-i+1)$$

It can shown that $a_i = \frac{i+1}{n+2}$ uniquely minimizes $R(F,d)$ for form (2). Since $\hat{F}$ is of form d and $\hat{F} = \sum_{i=1}^n \frac{i}{n}\mathbf{1}(X_{(i)} \le t < X_{(i+1)})$. $\hat{F}$ is inadmissible. □.

Homework solutions for week 1

5.14. Let Z_i's be i.i.d. $N(0,1)$ and $X_j = \sigma_j Z_j + \mu_j$. Suppose that

$0 = Cov(\sum_{j=1}^n a_{ij}Z_j, \sum_{j=1}^n b_{rj}Z_j) => \sum_{j=1}^n a_{ij}Z_j \perp \sum_{j=1}^n b_{rj}Z_j)$,

then

$0 = Cov(\sum_{j=1}^n a_{ij}X_j, \sum_{j=1}^n b_{rj}X_j) => \sum_{j=1}^n a_{ij}X_j \perp \sum_{j=1}^n b_{rj}X_j)$.

Sol. Notice that (1) $Cov(\alpha X + a, \beta Y + b) = \alpha\beta Cov(X,Y)$ and

(2) $(\sum_{j=1}^n a_{ij}X_j, \sum_{j=1}^n b_{rj}X_j)$

$= (\sum_{j=1}^n a_{ij}(\sigma_j Z_j + \mu_j), \sum_{j=1}^n b_{rj}(\sigma_j Z_j + \mu_j))$

$= (\sum_{j=1}^n \alpha_{ij} Z_j + \underbrace{\sum_{j=1}^n a_{ij}\mu_j}_{constant}, \sum_{j=1}^n \beta_{rj} Z_j + \underbrace{\sum_{j=1}^n b_{rj}\mu_j}_{constant}).$

Thus it suffices to show that

$\sum_{j=1}^n \alpha_{ij} Z_j \perp \sum_{j=1}^n \beta_{rj} Z_j$.

Now $0 = Cov(\sum_{j=1}^n a_{ij} Z_j, \sum_{j=1}^n b_{rj} Z_j) = \sum_{j=1}^n a_{ij} b_{rj}$ and

$0 = Cov(\sum_{j=1}^n a_{ij} X_j, \sum_{j=1}^n b_{rj} X_j)$

$= Cov(\sum_{j=1}^n a_{ij}(\sigma_j Z_j + \mu_j), \sum_{j=1}^n b_{rj}(\sigma_j Z_j + \mu_j))$

$= Cov(\sum_{j=1}^n a_{ij}\sigma_j Z_j, \sum_{j=1}^n b_{rj}\sigma_j Z_j)$

$= Cov(\sum_{j=1}^n \alpha_{ij} Z_j, \sum_{j=1}^n \beta_{rj} Z_j)$.

Thus $\sum_{j=1}^n \alpha_{ij} Z_j \perp \sum_{j=1}^n \beta_{rj} Z_j$.

$=> \sum_{j=1}^n a_{ij} X_j \perp \sum_{j=1}^n b_{rj} X_j$.

5.38. (b) Let X_i's be i.i.d. $\sim X$. Write $S = S_n = \sum_{i=1}^n X_i$. Show that if $E(X) < 0$ then there is a $c \in (0,1)$ with $P(S_n > a) \le c^n$.

Sol. Counterexample. Let $X \sim U(-1, 0)$, $a = -2$ and $n = 1$,

then $P(S_n > a) = 1 \not\le c^n$, $\forall\ c \in (0,1)$.

Correction.

$M_X(t) = M_X(0) + M'_X(\xi)t$ (where $\xi \in [0,1]$)

$\approx 1 + tE(X) < 1$ (if $t \approx 0+$) **Why ??**

$P(S > a) \le e^{-at} M_S(t)$ if $t > 0$,

$= e^{-at}(M_X(t))^n = (\frac{M_X(t)}{e^{at/n}})^n = c^n$, where $c = \frac{M_X(t)}{e^{at/n}} \in (0,1)$ and $t \approx 0+$,

which is possible if $a \ge 0$ **or** $0 > a > \frac{t}{n}\log(M_X(t))$ **and** $t \approx 0+$**.**

5.42. (a) $X_i \sim beta(1, \beta)$, $\nu =$? so that $n^\nu(1 - X_{(n)}) \xrightarrow{D}$ some Y.

(b) If $X_i \sim Exp(1)$, find a sequence a_n so that $X_{(n)} - a_n \xrightarrow{D}$ some Y.

Sol. 1. $Y_n \to Y$ if $\lim_{n\to\infty} Y_n(\omega) = Y(\omega)\ \forall\ \omega \in \Omega$ (the sample space).

2. $Y_n \xrightarrow{a.s.} Y$ if $P(\{|Y_n - Y| \to 0\}) = 1$

3. $Y_n \xrightarrow{P} Y$ if $\lim_{n\to\infty} P(|Y_n - Y| \ge \epsilon) = 0\ \forall\ \epsilon > 0$.

4. $Y_n \xrightarrow{D} Y$ if $\lim_{n\to\infty} F_{Y_n}(t) = F_Y(t)$ for each cts point t of F_Y.

(b) $P(X_{(n)} - a_n \le t) = P(X_{(n)} \le t + a_n) = (1 - e^{-(t+a_n)})^n = (1 - \frac{e^{-t}}{e^{a_n}})^n = (1 - \frac{e^{-t}}{n})^n \to e^{-e^{-t}} = F_Y(t)$, if $e^{a_n} = n$. **Any restriction on** t **?**

(a) $P(n^\nu(1 - X_{(n)}) \le t) =$ ...... or

$P(n^\nu(1 - X_{(n)}) > t) = P(X_{(n)} < 1 - t/n^\nu) = \{1 - [1 - (1 - t/n^\nu)]^\beta\}^n = (1 - (t/n^\nu)^\beta)^n = (1 - \frac{t^\beta}{n^{\nu\beta}})^n \to e^{-t^\beta}$

if $t > 0$ **Why ? and** $\nu = 1/\beta$.

$F_{n^\nu(1-X_{(n)})}(t) \xrightarrow{D} F_Y(t) = \mathbf{1}(t \ge 0)(1 - e^{-t^\beta})$ **??**

$n^\nu(1 - X_{(n)}) \xrightarrow{D} Y$ (Weibull distribution) **??**

5.43. Proof of the Delta method. $\sqrt{n}(g(\overline{X}) - g(\mu)) \xrightarrow{D} N(0, \sigma_X^2(g'(\mu))^2)$ if

(1) $\mu = E(X)$,

(2) g' is continuous at μ and

(3) $g'(\mu) \ne 0$.

Proof. $\sqrt{n}(g(\overline{X}) - g(\mu)) = \sqrt{n}g'(\mu)(\overline{X} - \mu)$?

$\sqrt{n}(g(\overline{X}) - g(\mu)) = \sqrt{n}g'(\mu)(\overline{X} - \mu) + R$ ($R = remainder = o_p(\overline{X} - \mu)$) ?

$\sqrt{n}(g(\overline{X}) - g(\mu)) = g'(\xi)\sqrt{n}(\overline{X} - \mu)$, where ξ is between μ and $\overline{X}$?

Q: Is ξ random ?

$\forall\ \epsilon > 0$, $\exists\ \delta > 0$ such that $|g'(X_n) - g'(\mu)| \le \epsilon$ whenever $|X_n - \mu| \le \delta$.

$\{|g'(X_n) - g'(\mu)| \le \epsilon\} \supset \{|X_n - \mu| \le \delta\}$.

$\{|g'(\xi) - g'(\mu)| \le \epsilon\} \supset \{|\xi - \mu| \le \delta\} \supset \{|X_n - \mu| \le \delta\}$.

$P(|g'(\xi) - g'(\mu)| > \epsilon) \le P(|\overline{X} - \mu| > \delta) \to 0\ \forall\ \epsilon > 0$.

$P(|g'(\xi) - g'(\mu)| \geq \epsilon) \leq P(|\overline{X} - \mu| \geq \delta) \to 0 \; \forall \; \epsilon > 0.$

=> $g'(\xi) \xrightarrow{P} g'(\mu).$

$\sqrt{n}(g(\overline{X}) - g(\mu)) = \underbrace{g'(\xi)\sqrt{n}(\overline{X} - \mu) \xrightarrow{D} g'(\mu)Z}_{Y_n Z_n \xrightarrow{D} yZ}$, where $Z \sim N(0, \sigma_X^2)$.

$g'(\mu)Z \sim N(0, \sigma_X^2 (g'(\mu))^2).$

Q: $\sqrt{n}(\overline{X} - \mu) \xrightarrow{D} N(0, \sigma^2)$ => $\sqrt{n}\overline{X} \xrightarrow{D} N(\sqrt{n}\mu, \sigma^2)$ **???**

5.44. $\vdash$: If $p = 0.5$, $n(Y_n(1 - Y_n) - 0.25) \xrightarrow{D} \frac{-1}{4}\chi_1^2$.

Proof. Let $g(Y_n) = Y_n(1 - Y_n)$,

$n(g(Y_n) - g(0.5)) = -(\sqrt{n}(Y_n - 0.5))^2 = -0.5^2(\sqrt{n}\frac{Y_n - 0.5}{0.5})^2.$

Extra Question: If $Z_n \xrightarrow{D} Z \sim N(0, 1)$, how to show $Z_n^2 \xrightarrow{D} \chi_1^2$?

$F_{Z_n^2}(t) = P(Z_n^2 \leq t) = P(-\sqrt{t} \leq Z_n \leq \sqrt{t})$

$\to P(-\sqrt{t} \leq Z \leq \sqrt{t}) = P(Z^2 \leq t) = F_{Z^2}(t)$

=> $Z_n^2 \to \chi_1^2$ **??**

Additional problem. $\vdash$: If $X_n \xrightarrow{a.s.} X$ and g is continuous, then $g(X_n) \xrightarrow{a.s.} g(X)$.

Proof. Q: Which proof is correct ??

1. Since $\{X_n \to X\} \subset \{g(X_n) \to g(X)\}$,

$1 = P(\{X_n \to X\}) \leq P(\{g(X_n) \to g(X)\}) \leq 1.$ □

2. $\forall \; \epsilon > 0$, $\exists \; \delta > 0$ such that $|g(X_n) - g(X)| < \epsilon$ whenever $|X_n - X| \leq \delta$. Thus $\{|g(X_n) - g(X)| < \epsilon\} \supset \{|X_n - X| \leq \delta\}$.

=> $P(|g(X_n) - g(X)| < \epsilon) \geq P(|X_n - X| \leq \delta) \to 1.$

=> $P(|g(X_n) - g(X)| < \epsilon) \to 1 \; \forall \; \epsilon > 0.$

=> $\begin{cases} g(X_n) \xrightarrow{a.s.} g(X) & \textbf{???} \\ g(X_n) \xrightarrow{P} g(X) & \textbf{???} \end{cases}$

Remark. $P(|X_n - X| \geq \epsilon) \to 0 \; \forall \; \epsilon > 0$ iff $P(|X_n - X| \leq \epsilon) \to 1 \; \forall \; \epsilon > 0$.

$P(X_n \to X) = P(|X_n - X| \to 0)$

Additional problem. $\vdash$: If $X_n \xrightarrow{P} X$ and g is continuous, then $g(X_n) \xrightarrow{P} g(X)$.

Proof. Q: Which proof is correct ??

1. $\forall \; \epsilon > 0$, $\exists \; \delta > 0$ such that $|g(X_n) - g(X)| < \epsilon$ whenever $|X_n - X| \leq \delta$. Thus $\{|g(X_n) - g(X)| < \epsilon\} \supset \{|X_n - X| \leq \delta\}$.

=> $P(|g(X_n) - g(X)| < \epsilon) \geq P(|X_n - X| \leq \delta) \to 1.$

=> $P(|g(X_n) - g(X)| < \epsilon) \to 1 \; \forall \; \epsilon > 0.$

=> $P(|g(X_n) - g(X)| \geq \epsilon) \to 0 \; \forall \; \epsilon > 0.$

=> $g(X_n) \xrightarrow{P} g(X).$

2. Since X is a r.v., $\forall \; \eta > 0$, $\exists \; a > 0$ such that $P(|X| > a) < \eta$.

$\forall \; \epsilon > 0$, $\exists \; \delta > 0$ such that $|g(X_n) - g(X)| < \epsilon$ whenever $|X_n - X| \leq \delta$ and $|X| \leq a$. Thus $\{|g(X_n) - g(X)| < \epsilon\} \supset \{|X_n - X| \leq \delta, \; |X| \leq a\}$.

=> $P(|g(X_n) - g(X)| < \epsilon) \geq P(|X_n - X| \leq \delta) - \eta \; \forall \; \epsilon, \eta > 0.$

=> $P(|g(X_n) - g(X)| < \epsilon) \to 1 - \eta \; \forall \; \epsilon, \eta > 0.$

=> $P(|g(X_n) - g(X)| < \epsilon) \to 1$ letting $\eta \to 0$.

=> $P(|g(X_n) - g(X)| \geq \epsilon) \to 0 \; \forall \; \epsilon > 0.$

=> $g(X_n) \xrightarrow{P} g(X).$

Remark. $P(|X_n - X| \geq \epsilon) \to 0 \; \forall \; \epsilon > 0$ iff $P(|X_n - X| \leq \epsilon) \to 1 \; \forall \; \epsilon > 0$.

$P(X_n \to X) = P(|X_n - X| \to 0)$

9.17. CI for θ based on X_1, ..., X_n. (a) $f(x;\theta) = \mathbf{1}(x - \theta \in (-0.5, 0.5))$. (b) $F(X;\theta) = 2x^2/\theta^2\mathbf{1}(x \in (0, \theta))$.

Sol. (a) The MSS of θ is $(X_{(1)}, X_{(n)})$,
as $f_{\mathbf{X}}(\mathbf{x}) = \mathbf{1}(X_{(1)} > \theta - 0.5, X_{(n)} < \theta + 0.5)$.

$T = X_{(n)} - \theta$ is a pivotal.

$F_T(t) = (t + 0.5)^n\mathbf{1}(t \in -0.5, 0.5) + \mathbf{1}(t \geq 0.5)$.

$f_T(t) = n(t+0.5)^{n-1}$, $t \in (-0.5, 0.5)$, with mode at 0.5.
Choose $b = 0.5$ and $(a+0.5)^n = \alpha$. That is, $a = (\alpha)^{\frac{1}{n}} - 0.5$.
$a \le X_{(n)} - \theta \le 0.5$
$X_{(n)} - 0.5 \le \theta \le X_{(n)} - (\alpha)^{\frac{1}{n}} + 0.5$.
Question: Can we choose $(F_T(b), F_T(a)) = (1-\alpha/2, \alpha/2)$?

(b) $X_{(n)}$ is the MSS. $T = X_{(n)}/\theta$ is a pivatol.
$F_T(t) = (t)^{2n}\mathbf{1}(t \in (0,1)$.
$f_T(t) = 2nt^{2n-1}$, $t \in (0,1)$.
Choose $b = 1$ and $F_T(a) = a^{2n} = \alpha$. => $a = \alpha^{\frac{1}{2n}}$.
$a \le X_{(n)}/\theta \le 1$
$X_{(n)} \le \theta \le X_{(n)}/\alpha^{\frac{1}{2n}}$.

Additional. before Chapter 9
(b) Redo the folllowing problem and compute $P(H_0|H_1)$ explicitly:
Carry out the following simulation project.
1.b.1. Use R to generate 5 observations from $N(1,1)$. Now pretend that you only known that the data were from $N(\mu,\sigma)$ without knowing μ and σ, use t-test to test H_0: $\mu = \underbrace{0}_{=\mu_o}$ v.s. H_1: $\mu \ne 0$ with a size 0.1.

Sol. The LRT test is $\phi = \mathbf{1}(|T| > t_{n-1,0.05})$, where $T = \frac{\overline{X}-\mu_o}{S/\sqrt{n}}$ (see Ex 3 in §8.2)
Since $X \sim N(1,1)$,
$\overline{X} \sim N(1, 1/\sqrt{5})$,
$4S^2 \sim \chi^2(4) = Gamma(2,2)$,
$\overline{X} \perp S^2$.
$f_{S^2}(t) = \frac{1}{4}\frac{(t/4}{2^2}e^{-t/8}$, $t > 0$.
$P(H_0|H_1) = P(T \in [-t_{4,0.05}, t_{4,0.05}])$
$= P(\overline{X} \in [-t_{4,0.05}S/\sqrt{5}, t_{4,0.05}S/\sqrt{5}])$
$= \int_0^\infty \int_{-t_{4,0.05}\sqrt{y/5}}^{t_{4,0.05}\sqrt{y/5}} f_{\overline{X}}(x) f_{S^2}(y)dydx$
$= \int_{-t_{4,0.05}}^{t_{4,0.05}} f_T(y)dy$ where f_T is given in #8.35(b)
$\frac{\overline{X}-\mu_o}{S/\sqrt{n}}$ is called non-central t-distribution with parameter μ and df $n-1$.
It can be obtained numericall in R:

```
> n=5
> df =n-1
> sigma=1
> mu=1
> ncp = mu * sqrt(n)/sigma #ncp = (μ − μo)√n/σ.
> q=qt(0.05, df, ncp=0, lower.tail = F)
> b=pt(q, df, ncp, lower.tail = TRUE)
> a=pt(-q, df, ncp, lower.tail = TRUE)
> b-a
[1] 0.4200955 # = P(H0|H1)
> z=0
> m=10000
> for (i in 1:m) {
+ x=rnorm(n)+mu
+ y=t.test(x)
+ z=z+as.numeric(y$p.value > 0.1)
+ }
> z/m
[1] 0.4189 # = P̂(H0|H1)
```

[1] 0.4156 # $= \hat{P}(H_0|H_1)$
For $m = 100$, it results 0.36, 0.43, 0.44, 0.53, ...

Recall the SLLN. $\overline{X} \to E(X)$ a.s. X_i**= ??**
For submitting homework in simulation,

1. Put the command in a file called hw8.r
2. R −−vanilla < hw8.r > output8
3. Edit the file called output8 by answering the question in the problem.
4. mail qyu@math.binghamton.edu < output8

A.1. Discuss whether the following solutions are correct for
7.12. Compare the MLE $\hat\theta = \min\{\overline{X}, 1/2\}$ and the MME $\tilde\theta = \overline{X}$.

Sol (1). $MSE(\hat\theta) = \begin{cases} E((\overline{X}-\theta)^2) & \text{if } \hat\theta = \overline{X} \\ E((\theta-1/2)^2) & \text{if } \hat\theta = 1/2 \end{cases} = \begin{cases} \theta(1-\theta)/n & \text{if } \hat\theta = \overline{X} \\ (\theta-1/2)^2 & \text{if } \hat\theta = 1/2 \end{cases}$

$$MSE(\hat\theta) - MSE(\tilde\theta) = \begin{cases} 0 & \text{if } \hat\theta = \overline{X} \\ (\theta-1/2)^2 - \theta(1-\theta)/n & \text{if } \hat\theta = 1/2 \end{cases} \quad \begin{cases} > 0 & \text{if } \theta = 0 \\ < 0 & \text{if } \theta = 1/2 \\ ... & \text{otherwise} \end{cases}$$

Thus none of them is better than the other in terms of MSE.
Sol (2). $MSE(\tilde\theta) = \frac{\theta(1-\theta)}{n} = \sum_{i=0}^n (\frac{i}{n}-\theta)^2 \binom{n}{i}\theta^i(1-\theta)^{n-i}$.
$MSE(\hat\theta) = \sum_{i\le n/2}(\frac{i}{n}-\theta)^2\binom{n}{i}\theta^i(1-\theta)^{n-i} + \sum_{i>n/2}(\frac{1}{2}-\theta)^2\binom{n}{i}\theta^i(1-\theta)^{n-i}$.

$$MSE(\hat\theta) - MSE(\tilde\theta) = \sum_{i>n/2}[(\frac{1}{2}-\theta)^2 - (\frac{i}{n}-\theta)^2]\binom{n}{i}\theta^i(1-\theta)^{n-i} \begin{cases} = 0 & \text{if } \theta = 0 \\ < 0 & \text{if } \theta \in (0, 1/2] \end{cases}$$

Thus the MLE is better than the MME in terms of MSE.
Answer:

$$MSE(\hat\theta) = \begin{cases} E((\overline{X}-\theta)^2) & \text{if } \hat\theta = \overline{X} \\ E((\theta-1/2)^2) & \text{if } \hat\theta = 1/2 \end{cases} \tag{1}$$

Comment: Eq.(1) => $E((\hat\theta-\theta)^2)$ is a random variable. This is wrong !
$E((\hat\theta-\theta)^2)$ **is a function of θ.**
e.g. if $X \sim N(\mu, 1)$, then $E(X) = \mu$ is not a random variable.

$$E(X) = \begin{cases} \mu & \text{if } X \in A \text{ (a set)} \\ 2 & \text{if } \mu = 2 \\ \mu & \text{if } \theta = 2 \\ 2 & \text{if } \theta = \mu = 2 \end{cases}$$

A.2. Question related to #7.14. Recall $W \sim bin(1, p)$, with df. $f(t) = P(W = t) = \begin{cases} p & \text{if } t = 1 \\ q & \text{if } t = 0 \end{cases}$. Answer the following questions:
$f(t) = p$ if $t = 1$. Yes, No ?
$f(t) = p$ if $W = 1$. Yes, No ?
$P(W = t) = p$ if $W = 1$. Yes, No ?
$P(W = t) = p$ if $t = 1$. Yes, No ?
$P(\{\omega \in \Omega : W(\omega) = t\}) = P(W = t) = p$ if $W = 1$. Yes, No ?
$P(\{\omega \in \Omega : W(\omega) = t\}) = P(W = t) = p$ if $t = 1$. Yes, No ?
Notice that $Z = \begin{cases} X & \text{if } W = 1 \\ Y & \text{if } W = 0. \end{cases}$
$P(Z \le t) = 1 - P(Z > t) = 1 - P(X > t, Y > t) = 1 - P(X > t)P(Y > t)$.
$P(Z \le t, W = a) = P(X \le t, W = 1)$ if $a = 1$ Yes, No ?
$P(Z \le t, W = a) = P(X \le t, W = 1)$ if $W = 1$ Yes, No ?
$P(Z \le t, W = a) = P(X \le 1, W = 1)$ if $t = 1 = a$ Yes, No ?

$P(Z \le t, W = a) = P(X \le t, W = 1)$ if $Z = X$ Yes, No ?

Answer: $f(t) = P(W = t) = P(\omega \in \Omega : W(\omega) = t) = \begin{cases} p & \text{if } t = 1 \\ q & \text{if } t = 0 \end{cases}$ is not a random variable, but a function of t.

$$f(t) = \begin{cases} q & \text{if } t = 0 \\ f(t) & \text{if } W = 1 \\ f(t) & \text{if } \omega = 1 \\ f(x) & \text{if } t = x \\ f(t) & \text{if } a = 1 \end{cases}$$

A.3. What is the connection between A.1 and A.2 ?

Given random variable X,
$E(X)$ and $P(X \in A)$ are constant, not random variables.
They do not change accoring to values of X, as in
Sol(1) in A.1 and the statement "$f(t) = 0$ if $W = 1$".

Homework Solution

6.9b. Find MSS for θ, where $f(x;\theta) = e^{-(x-\theta)}$, $x > \theta$.

Sol. Two solutions:

1. $f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n e^{-x_i+\theta}\mathbf{1}(x_i > \theta)$
$= e^{-n\overline{x}}e^{n\theta}\mathbf{1}(x_{(1)} > \theta)$.

$T = X_{(1)}$ is MSS, as

$$\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{y})} = e^{-n(\overline{x}-\overline{y})}\frac{\mathbf{1}(x_{(1)} > \theta)}{\mathbf{1}(y_{(1)} > \theta)} \text{ is independent of } \theta \text{ iff } x_{(1)} = y_{(1)}. \tag{1}$$

2. $f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n e^{-x_i+\theta}\mathbf{1}(x_i > \theta)$.

$T(\mathbf{X}) = \mathbf{X}$ is MSS, as

$$\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{y})} = \frac{\prod_{i=1}^n e^{-x_i+\theta}\mathbf{1}(x_i > \theta)}{\prod_{i=1}^n e^{-y_i+\theta}\mathbf{1}(y_i > \theta)} \text{ is independent of } \theta \text{ iff } \mathbf{x} = \mathbf{y}. \tag{2}$$

Anything Wrong ??

Eq. (1) is correct, but needs justification as follows.

If $x_{(1)} = y_{(1)}$ then

$f_{\mathbf{X}}(\mathbf{x})/f_{\mathbf{X}}(\mathbf{y}) = e^{-n(\overline{x}-\overline{y})}\frac{\mathbf{1}(x_{(1)}>\theta)}{\mathbf{1}(y_{(1)}>\theta)} = e^{-n(\overline{x}-\overline{y})}$ is independent of θ. OW,

$$\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{y})} = e^{-n(\overline{x}-\overline{y})}\frac{\mathbf{1}(x_{(1)} > \theta)}{\mathbf{1}(y_{(1)} > \theta)} = \begin{cases} e^{-n(\overline{x}-\overline{y})} > 0 & \text{if } \theta = [x_{(1)} \wedge y_{(1)}] - 1 \\ \infty\mathbf{1}(y_{(1)} < x_{(1)}) & \text{if } \theta = \frac{x_{(1)}+y_{(1)}}{2} \end{cases}$$

where $0 \times \infty = 0$ and $\frac{0}{0} = 1$, $\frac{1}{0} = \infty$.

Eq. (2) is incorrect and a counterexample is as follows.

$f_{\mathbf{X}}(\mathbf{x})/f_{\mathbf{X}}(\mathbf{y}) = e^{-n(\overline{x}-\overline{y})}\frac{\mathbf{1}(x_{(1)}>\theta)}{\mathbf{1}(y_{(1)}>\theta)} = e^{-n(\overline{x}-\overline{y})}$ is independent of θ if $x_{(1)} = y_{(1)}$ and $x_{(2)} = y_{(2)} + 1$ and thus if $\mathbf{x} \ne \mathbf{y}$.

Thus $\mathbf{X}$ is not a MSS.

\# 6.9. (c) Let $X_1, ..., X_n \sim f(x-\theta)$, where $f(x-\theta) = \frac{\exp(-(x-\theta))}{(1+\exp(-(x-\theta)))^2}$. Show
$T = (X_{(1)}, ..., X_{(n)})$ is MSS.

Proof. WLOG, assume $x_1 < \cdots < x_n$ and $y_1 < \cdots < y_n$.

$\prod_{i=1}^n f(x_i;\theta) = \exp(-\sum_{i=1}^n x_i + n\theta)\prod_{i=1}^n(1+\exp(-(x_i-\theta)))^{-2}$

Since $\prod_{i=1}^n f(x_i;\theta)/f(y_i;\theta) = e^{n(\overline{y}-\overline{x})}\prod_{i=1}^n \frac{(1+\exp(-(y_{(i)}-\theta)))^2}{(1+\exp(-(x_{(i)}-\theta)))^2}$

$\prod_{i=1}^n f(x_i;\theta)/f(y_i;\theta)$ does not depend on θ iff $x_{(i)} = y_{(i)}$ $i \in \{1, ..., n\}$ **Done ?**

Need a proof !

$\prod_{i=1}^n f(x_{(i)};\theta)/f(y_{(i)};\theta) = c$;
iff $\prod_{i=1}^n \frac{(1+\exp(-(y_{(i)}-\theta)))^2}{(1+\exp(-(x_{(i)}-\theta)))^2} = c$;
iff $\prod_{i=1}^n \frac{(1+\exp(-(y_{(i)}-\theta)))}{(1+\exp(-(x_{(i)}-\theta)))} = \sqrt{c} \ \forall\ \theta$;
iff $\prod_{i=1}^n \frac{1+\exp(-y_{(i)})e^\theta}{1+\exp(-x_{(i)})e^\theta} = \sqrt{c} \ \forall\ \theta$;
iff $\prod_{i=1}^n \frac{(\exp(y_{(i)})+e^\theta)}{(\exp(x_{(i)})+e^\theta)} = \exp(n\overline{y} - n\overline{x})\sqrt{c} = a \ \forall\ \theta$;
$(a = \lim_{\theta\to\infty} \prod_{i=1}^n \frac{y_{(i)}+\theta}{x_{(i)}+\theta} = 1$, where $(t_i = e^{x_i},\ s_i = e^{y_i},\ \eta = e^\theta))$;
iff $\prod_{i=1}^n \frac{s_{(i)}+\eta}{t_{(i)}+\eta} = 1 \ \forall\ \eta, s_i, t_i > 0$;
iff $\prod_{i=1}^n (s_{(i)}+\eta) = \prod_{i=1}^n (t_{(i)}+\eta) \ \forall\ \eta, s_i, t_i > 0$
(both are polynomial of degree n in η);
iff $t_{(i)} = s_{(i)} \ \forall\ i$; iff $x_{(i)} = y_{(i)} \ \forall\ i$.
Previous proofs make use of two results in complex analysis:
(1) $\sum_{i=0}^n a_i x^i = \sum_{i=0}^n b_i x^i$ iff $a_i = b_i$ for all i.
(2) $\sum_{i=0}^n a_i x^i = c\prod_{i=1}^n (x - c_i)$, where $c_1, ..., c_n, c$ ares uniquely determined complex numbers.
Additional homework.
A1. Let $X \sim N(0,1)$,
$W \sim bin(2, 0.1)$,
$Y \sim bin(1, 0.5)$.
X, W and Y are independent.
$Z = \begin{cases} X & \text{if } Y = 1 \\ W & \text{if } Y = 0. \end{cases}$
$P(Z \le t) = P(X \le t, Y = 1) + P(W \le t, Y = 0)$ if $Y = 1$. Yes, No ?
$P(Z \le t, Y = a) = \begin{cases} P(X \le t, Y = 1) & \text{if } a = 1 \\ P(W \le t, Y = 0) & \text{if } a = 0 \end{cases} Yes, No?$
$= P(X \le t, Y = a = 1) + P(W \le t, Y = a = 0)$ Yes, No ?
$P(Z \le t, Y = a) \ne P(X \le t, Y = 1)$ if $Y = 1$. Yes, No ?
$P(Z \le t, Y = a) = P(X \le t, Y = 1)$ if $a = 1$. Yes, No ?
Are statements 3 and 4 equivalent ? Yes, No.
3. $\{w : w \in (0, 3)\}$ if $w > 2$
4. $\{w : w \in (0, 3), w > 2\}$
Are the sets in 3 and 7 the same ? Yes, No.
7. $\{u : u \in (0, 3)\}$ if $t > 2$
Are the sets in 3 and 8 the same ? Yes, No.
8. $\{x : x \in (0, t)\}$ if $t = 3$
9. $P(Z \le t) = P(\omega :\ Z(\omega) \le t) = P(\omega :\ X(\omega) \le t)$ if $Y = 1$. Yes, No ?

\# 6.5. X_1, ..., X_n are independent. $f_{X_i}(x) = \frac{\mathbf{1}_{(x\in(-i(\theta-1), i(\theta+1)))}}{2i\theta}$. A two-dimensional sufficient statistic for θ ?
Sol.

$$\begin{aligned}
\mathcal{L}(\theta) =& [\prod_{i=1}^n \frac{1}{2i\theta}]\mathbf{1}_{(X_i\in(-i(\theta-1), i(\theta+1))):\ i\in\{1,...,n\})} \\
=& [\prod_{i=1}^n \frac{1}{2i\theta}]\mathbf{1}_{(\frac{X_i}{i}\in(-(\theta-1), (\theta+1))):\ i\in\{1,...,n\})} \\
=& [\prod_{i=1}^n \frac{1}{2i\theta}]\mathbf{1}_{(\frac{X_i}{i}-1\in(-\theta, \theta)):\ i\in\{1,...,n\})} \\
=& [\prod_{i=1}^n \frac{1}{2i\theta}]\mathbf{1}_{(\max_i |\frac{X_i}{i}-1|\in(0,\theta)))}
\end{aligned}$$

Thus $T = (\max_i |\frac{X_i}{i} - 1|, 0)$. In fact, $\max_i |\frac{X_i}{i} - 1|$ is MSS for θ.
Solution to Additional Homework.

6.8. Let X_1, ..., X_n be i.i.d. $\sim f(x-\theta)$, where $f(\cdot)$ is a df. Show $T=(X_{(1)},...,X_{(n)})$ is MSS.

Sol. Discuss whether the following two solutions are correct (give your reasoning).

Solution 1. Let $f=\frac{1}{2}\exp(-|x|)$, then it is shown in Exercise 6.9.e that T is MSS for $f(t-\theta)$.

Solution 2. The statement is a wrong statement. It suffices to give a counterexample as follows. Let $X_i \sim N(\theta,1)$, then $X_i \sim f(x-\theta)$ with $f(x)=\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$. An MSS is $\overline{X}$, T is not a function of $\overline{X}$. Thus T is not MSS.

Correct solution. There are two interpretations of the problem.

(1) Only θ is a parameter, though f is an arbitrary given density. (T **is MSS for** θ).

(2) Both f and θ are parameters. (T **is MSS for** (f,θ)).

In case (1), it is a wrong statement. Thus Solution **2 but not 1** is correct.

In case (2), it is a correct statement. Thus Solution **1 but not 2** is correct.

It suffices to show that given $T(\mathbf{x})\neq T(\mathbf{y})$, we can find (f_1,θ_1) and (f_2,θ_2) such that $\frac{\prod_{i=1}^n f_1(x_{(i)}-\theta_1)}{\prod_{i=1}^n f_1(y_{(i)}-\theta_1)} \neq \frac{\prod_{i=1}^n f_2(x_{(i)}-\theta_2)}{\prod_{i=1}^n f_2(y_{(i)}-\theta_2)}$. In particular, let $f_1=f_2=\frac{1}{2}\exp(-|x|)$ and let θ_1 and θ_2 be as in #8.9.c. □

6.8. Let X_1, ..., X_n be i.i.d. $\sim f(x-\theta)$, where $f(\cdot)$ is a df. Show $T=(X_{(1)},...,X_{(n)})$ is MSS.

Sol. There are two interpretations of the problem.

(1) f is not a parameter, but θ is. (T **is MSS for** θ).

(2) Both f and θ are parameters. (T **is MSS for** (f,θ)).

In case (1), it is a wrong statement. **Counterexample.**
Let $X_i \sim N(\theta,1)$, then $X_i \sim f(x-\theta)$ with $f(t)=\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$. An MSS is $\overline{X}$, T is not a function of $\overline{X}$. Thus T is not MSS.

In case (2), it is a correct statement.

It suffices to show that given $T(\mathbf{x})\neq T(\mathbf{y})$, we can find (f_1,θ_1) and (f_2,θ_2) such that $\frac{\prod_{i=1}^n f_1(x_{(i)}-\theta_1)}{\prod_{i=1}^n f_1(y_{(i)}-\theta_1)} \neq \frac{\prod_{i=1}^n f_2(x_{(i)}-\theta_2)}{\prod_{i=1}^n f_2(y_{(i)}-\theta_2)}$. In particular, let $f_1=f_2$ and let θ_1 and θ_2 be as in #6.9.c or d, or e. □

6.9. e. $\vdash$: Let X_1, ..., $X_n \sim f=\frac{1}{2}e^{-|x-\theta|}$. $T=(X_{(1)},...,X_{(n)})$ is MSS.

Proof. $f_{\mathbf{X}|T}(\mathbf{x}|t)=\frac{1}{n!}$, thus T is sufficient by the definition.

It is easy to show that if $T(\mathbf{x})=T(\mathbf{y})$, then $\frac{f_{\mathbf{X}}(\mathbf{x};\theta)}{f_{\mathbf{X}}(\mathbf{y};\theta)}=1$ for each θ.

Need to show that if $T(\mathbf{x})\neq T(\mathbf{y})$, then there exist θ_1 and θ_2 such that

$$\frac{f_{\mathbf{X}}(\mathbf{x};\theta_1)}{f_{\mathbf{X}}(\mathbf{y};\theta_1)} \neq \frac{f_{\mathbf{X}}(\mathbf{x};\theta_2)}{f_{\mathbf{X}}(\mathbf{y};\theta_2)} \tag{1}$$

Discuss whether the following two approaches are correct.

Approach 1. If $T(\mathbf{x})\neq T(\mathbf{y})$, without loss of generality (WLOG), one can assume $x_{(i)}=y_{(i)}$ for $i\neq j$ and $x_{(i)}<y_{(i)}$ for $i=j$. Then

$$\frac{f_{\mathbf{X}}(\mathbf{x};\theta)}{f_{\mathbf{X}}(\mathbf{y};\theta)}=\prod_i \frac{e^{-|x_{(i)}-\theta|}}{e^{-|y_{(i)}-\theta|}}$$
$$=\frac{e^{-|x_{(i)}-\theta|}}{e^{-|y_{(i)}-\theta|}}=\begin{cases} e^{-x_{(i)}-y_{(i)}} & \text{if } \theta<x_{(j)} \\ e^{+x_{(i)}+y_{(i)}-2\theta} & \text{if } \theta\in(x_{(j)},y_{(j)}) \end{cases}$$

Letting $\theta_1=x_{(j)}-1$ and $\theta_2=(x_{(j)}+y_{(j)})/2$ yields Ineq. (1).

Approach 2. If $T(\mathbf{x})\neq T(\mathbf{y})$, then there exists $j\in\{1,...,n\}$ such that $x_{(i)}=y_{(i)}$ for $i<j$ and $x_{(i)}\neq y_{(i)}$ for $i=j$. Without loss of generality, assume that $x_{(j)}<y_{(j)}$. Let $t=y_{(j)}\wedge\min\{x_{(k)}:\ x_{(k)}>x_{(j)},\ k>j\}$ (where $x_{(n+1)}=\infty$).
WLOG, assume $x_{(j+1)}>x_{(j)}$ and $t=y_{(j)}\wedge x_{(j+1)}$.
Let $\theta\in[x_{(j)},t]$. Then

$$\frac{f_{\mathbf{X}}(\mathbf{x};\theta)}{f_{\mathbf{X}}(\mathbf{y};\theta)}=\prod_i \frac{e^{-|x_{(i)}-\theta|}}{e^{-|y_{(i)}-\theta|}}$$

$$= (\prod_{i<j} \frac{e^{-|x_{(i)}-\theta|}}{e^{-|y_{(i)}-\theta|}})(\prod_{i=j} \frac{e^{-|x_{(i)}-\theta|}}{e^{-|y_{(i)}-\theta|}})(\prod_{i>j} \frac{e^{-|x_{(i)}-\theta|}}{e^{-|y_{(i)}-\theta|}})$$

$$= \frac{e^{x_{(j)}-\theta}}{e^{-y_{(j)}+\theta}} (\prod_{i>j} \frac{e^{-(x_{(i)}-\theta)}}{e^{-(y_{(i)}-\theta)}})$$

$$= e^{x_{(j)}+y_{(j)}-2\theta} (\prod_{i>j} \frac{e^{-(x_{(i)})}}{e^{-(y_{(i)})}})$$

Letting $\theta_1 = x_{(j)}$ and $\theta_2 = t$, we get inequality (1). Thus T is MSS. □

Discuss whether the proofs are correct.

6.9(d) Let X_1, ..., $X_n \sim f(x-\theta)$, where $f(x) = \frac{1}{\pi(1+x^2)}$. Show $T = (X_{(1)}, ..., X_{(n)})$ is MSS.

Sol. $\prod_{i=1}^n f(x_{(i)};\theta)/f(y_{(i)};\theta) = c$;

iff $\prod_{i=1}^n (1+(x_{(i)}-\theta)^2)/(1+(y_{(i)}-\theta)^2) = c \to 1$ if $\theta \to \infty$

(both are polynomial of degree $2n$ in θ);

approach 1:

iff their coefficients are the same.

iff $x_{(j)} = y_{(j)}$ $\forall$ j.

approach 2: $(1+(x-\theta)^2) = (1-i(x-\theta))(1+i(x-\theta))$

$= (1-ix+i\theta)(1+ix-i\theta)$

$= -(i+x-\theta)(i-x+\theta)$.

$\prod_{i=1}^n f(x_{(i)};\theta)/f(y_{(i)};\theta) = 1$;

iff $\prod_{j=1}^n [-(i+y_{(j)}-\theta)(i-y_{(j)}+\theta)] = \prod_{j=1}^n [-(i+x_{(j)}-\theta)(i-x_{(j)}+\theta)]$ $\forall$ θ;

(both are polynomial of degree $2n$ in θ);

iff their $2n$ roots are the same

iff $x_{(j)} = y_{(j)}$ $\forall$ j.

prove by induction on j: WLOG, assume $x_1 < \cdots < x_n$ and $y_1 < \cdots < y_n$.

$j=1$. $\theta = i + y_1 => x_1 = y_1$.

$j = k$ $(< n)$. Assume $x_h = y_h$ for $h < k$.

$j = k+1$. By induction,

$\prod_{j=k+}^n [-(i+y_{(j)}-\theta)(i-y_{(j)}+\theta)] = \prod_{j=k+1}^n [-(i+x_{(j)}-\theta)(i-x_{(j)}+\theta)]$ $\forall$ θ;

$\theta = i + y_{k+1} => x_{k+1} = y_{k+1}$.

Homework Solution

6.10. Let X_i's be i.i.d. from $U(\theta, \theta+1)$. Show $T = (X_{(1)}, X_{(n)})$ is not complete.

A wrong proof:

(1) Let $g(T) = X_{(n)} - X_{(1)} - E(X_{(n)} - X_{(1)})$.

(2) Then $E(g(T)) = 0$ $\forall$ θ.

(3) But $g(T)$ is a non-zero function.

Thus T is not complete.

What is wrong ?

Should add "$g(T)$ is a statistic" in (1);

Replace (3) by "$P(g(T) = 0) = 1$ $\forall$ $\theta \in \Theta$."

Reason 1:

Let $T \sim Exp(1)$ and $g(T) = 1(X = 1)$ is non-zero. $P(g(T) = 0) = ??$

Reason 2. If it works, then we can show that

compete statistic T is not complete!!

Let X_1, ..., X_n be i.i.d. $\sim N(\theta, 1)$, then $T = \sum_{i=1}^n X_i$ is complete **Why ??**

$$f_X(x) \propto \exp(-\frac{1}{2}[x^2 - 2 \underbrace{\mu}_{w_1(\theta)} \underbrace{x}_{t_1(x)}])$$

$\{w_1(\theta) : \theta > 0\} = (0, \infty)$ contains an open set in $\mathcal{R}^1$.

Let $g(T) = T - E(T)$.

Then $E(g(T)) = 0 \ \forall\ \theta$.
But $P(g(T) = 0) = 0 < 1 \ \forall\ \theta$.
Thus T is not complete. **Anything wrong ?**
A Correct proof.
$E(X_{(1)}) = \int t f_{X_{(1)}}(t)dt = \theta + \frac{1}{n+1}$,
$E(X_{(n)}) = \int t f_{X_{(n)}}(t)dt = \theta + \frac{n}{n+1}$,
Since $E(X_{(n)} - X_{(1)}) = \frac{n-1}{n+1}$ is independent of θ,
let $g(T) = X_{(n)} - X_{(1)} - E(X_{(n)} - X_{(1)})$.
That is, $g(x, y) = y - x - \frac{n-1}{n+1}$.
Then $E(g(T)) = 0 \ \forall\ \theta$.
But $P(g(T) = 0) = 0 < 1 \ \forall\ \theta$.
Thus T is not complete. □.
#7.14. Let $X \perp Y$, $Y \sim f(x; \mu) \propto e^{-x/\mu}$, $x > 0$, $X \sim f(x; \lambda)$. We observe (Z, W), where $Z = \min(X, Y)$ and $W = \mathbf{1}_{(X \le Y)}$. If (Z_i, W_i), $i = 1, ..., n$ are i.i.d. from (Z, W), MLE of (μ, λ) ?
Sol. $\mathcal{L}(\theta) = \prod_{i=1}^n f_{Z,W}(Z_i, W_i) = ??$
$\mathcal{L}(\theta) = \prod_{i=1}^n [(\underbrace{f_{Z,W}(Z_i, 0)}_{=?})^{1-W_i} (\underbrace{f_{Z,W}(Z_i, 1)}_{=?})^{W_i}]$,

$$\text{(recall } f_W(t) = (f_W(0))^{1-t}(f_W(1))^t = \begin{cases} f_W(0) & \text{if } t = 0 \\ f_W(1) & \text{if } t = 1 \end{cases})$$

$f_{Z,W}(z, 0) = \frac{\partial}{\partial z} F_{Z,W}(z, 0)$.
$F_{Z,W}(z, 0) = P(X \wedge Y \le z, X > Y)$
$= P(Y \le z, X > Y)$
$= P(X > Y) - P(X > Y > z)$
$= P(X > Y) - \int_z^\infty \int_z^x f_X(x) f_Y(y) dy dx$

$$f_{Z,W}(z, 0) = -(-\frac{\partial z}{\partial z}) \int_z^z f_X(z) f_Y(y) dy - \int_z^\infty [\frac{\partial}{\partial z} \int_z^x f_X(x) f_Y(z) dy] dx \quad \textbf{why}?$$

$$\frac{\partial}{\partial z} \int_{a(z)}^{b(z)} g(x, z)dx = b'(z)g(b(z), z) - a'(z)g(a(z), z) + \int_{a(z)}^{b(z)} \frac{\partial}{\partial z} g(x, z)dx$$

$$f_{Z,W}(z, 0) = \int_z^z f_X(z) f_Y(y) dy + \int_z^\infty f_X(x) f_Y(z) dx$$
$$= \frac{1}{\mu} e^{-z/\mu} e^{-z/\lambda}, \ z > 0.$$

Likewise, $f_{Z,W}(z, 1) = \frac{1}{\lambda} e^{-z/\lambda} e^{-z/\mu}$, $z > 0$.

$$=> \quad f_{Z,W}(z, w) = \begin{cases} \frac{1}{\lambda} e^{-z/\lambda} e^{-z/\mu} & \text{if } w = 1 \\ \frac{1}{\mu} e^{-z/\lambda} e^{-z/\mu} & \text{if } w = 0 \end{cases}$$
$$= (\frac{1}{\lambda} e^{-z/\lambda} e^{-z/\mu})^w (\frac{1}{\mu} e^{-z/\lambda} e^{-z/\mu})^{1-w}, \quad z > 0.$$

$\mathcal{L}(\theta) = \prod_{i=1}^n f_{Z,W}(Z_i, W_i)$
$= \prod_{i=1}^n [(\frac{1}{\mu} e^{-Z_i(1/\mu + 1/\lambda)})^{1-W_i} (e^{-Z_i(1/\mu+1/\lambda)} \frac{1}{\lambda})^{W_i}]$
$= [(\frac{1}{\mu})^{(1-\overline{W})} e^{-\overline{Z}(1/\mu+1/\lambda)}) (\frac{1}{\lambda})^{\overline{W}}]^n$
$H = \frac{1}{n} \ln \mathcal{L}$
$= -(1 - \overline{W}) \ln\mu - \ln\overline{W} \ln\lambda - \overline{Z}(\frac{1}{\mu} + \frac{1}{\lambda})$
$= -(1 - \overline{W}) \ln\mu - \frac{\overline{Z}}{\mu} - \ln\overline{W} \ln\lambda - \frac{\overline{Z}}{\lambda})$
It suffices to maximize μ and λ seperately.
If $\overline{W} \notin \{0, 1\}$, then
$\frac{\partial H}{\partial \mu} = -\frac{1-\overline{W}}{\mu} + \overline{Z}/\mu^2 = 0 => \hat{\mu} = \overline{Z}/(1 - \overline{W})$,
as $H(0, \lambda) = H(\infty, \lambda) = -\infty$.

Similarly, $\hat\lambda = \overline{Z}/\overline{W}$.
Otherwise, if $\overline{W}=1$, the observations are all from X and thus $\hat\mu = \overline{X}$. $\hat\lambda = 1$ (or any number).
If $\overline{W}=0$, then $\hat\lambda = \overline{Y}$ and $\hat\mu = 1$ (or any number).
Recall $W \sim bin(1,p)$, with df. $f(t) = P(W=t) = \begin{cases} p & \text{if } t=1 \\ q & \text{if } t=0 \end{cases}$.

$f(t) = p$ if $t=1$. Yes, No ?
$f(t) = p$ if $W=1$. Yes, No ?
$P(W=t) = p$ if $W=1$. Yes, No ?
$P(W=t) = p$ if $t=1$. Yes, No ?
$P(\{\omega\in\Omega:\ W(\omega)=t\}) = P(W=t) = p$ if $W=1$. **No**
$P(\{\omega\in\Omega:\ W(\omega)=t\}) = P(W=t) = p$ if $t=1$. **Yes**

Notice that $Z = \begin{cases} X & \text{if } W=1 \\ Y & \text{if } W=0. \end{cases}$
$P(Z\le t) = 1-P(Z>t) = 1-P(X>t, Y>t) = 1-P(X>t)P(Y>t)$.

$P(Z\le t, W=a) = P(X\le t, W=1)$ if $a=1$ Yes, No ?
$P(Z\le t, W=a) = P(X\le t, W=1)$ if $W=1$ Yes, No ?
$P(Z\le t, W=a) = P(X\le 1, W=1)$ if $t=1=a$ Yes, No ?
$P(Z\le t, W=a) = P(X\le t, W=1)$ if $Z=X$ Yes, No ?

7.2. Let X_i's be i.i.d. from $\mathcal{G}(\alpha,\beta)$. $f(x)\propto x^{\alpha-1}e^{-x/\beta}$, $x>0$. MLE of (α,β) ?
Sol. (a) If α is given, the MLE of $\beta = \overline{X}/\alpha$.
(b) Three ways for numerical solutions of the MLE:

(1) Plot $y=\mathcal{L}(x)$, where $x\in\Theta$, **looking for ?**
(2) Plot $y=\frac{d\ln\mathcal{L}(x)}{dx}$ and $y=0$ **looking for ?**
(3) Newton-Raphson method

$$x^{new} = x^{old} - \left(\frac{d\ln\mathcal{L}(x)}{dx} \Big/ \frac{d^2\ln\mathcal{L}(x)}{dx^2}\right)\Big|_{x=x^{old}} \text{ untill } |x^{new}-x^{old}|<\epsilon.$$

$\mathcal{L} = \prod_{i=1}^n \frac{X_i^{\alpha-1}e^{-X_i/\beta}}{\Gamma(\alpha)\beta^\alpha}$
$= \frac{(\prod_i X_i)^{\alpha-1}}{\Gamma(\alpha)^n}\frac{\exp(-n\overline{X}/\beta)}{\beta^{n\alpha}}$
$= \frac{(\prod_i X_i)^{\alpha-1}}{\Gamma(\alpha)^n}\frac{\exp(-n\alpha)}{(\overline{X}/\alpha)^{n\alpha}}$
$= \frac{(\prod_i X_i)^{\alpha-1}}{\Gamma(\alpha)^n}(\frac{\alpha}{\overline{X}e})^{n\alpha}$
$= (\prod_i X_i)^{\alpha-1}(\frac{(\frac{\alpha}{\overline{X}e})^\alpha}{\Gamma(\alpha)})^n$.

```
# R program for hw7.2

x=c(22,23.9,20.9,23.8,25,24,21.7,23.8,22.8,23.1,23.1, 23.5,23,1)
n=length(x)
(mean(x)/sd(x))**2 # MME: α̃ = (X̄)²/σ̂².
    # as E(X) = αβ and V(X) = αβ².
[1] 12.87023
a=(1:170)/10 # possible range of α (in view of MME).
y=((prod(x))**(a-1))*((a/(exp(1)*mean(x)))**(a*n))/((gamma(a))**n)
(z=a[y==max(y)])
[1] 3.4 #MLE of α
    sum(x)/(n*z)
[1] 6.336134 # (MLE of β))
plot(a,y)
```

The second and third approach need to compute $\Gamma(\alpha)'$
$= \int_0^\infty t^{\alpha-1}(\ln t)e^{-t}dt = \frac{d}{da}\int_0^\infty t^{\alpha-1}(\ln t)e^{-t}dt$.
Why not x=c(22,23.9,20.9,23.8,25,24,21.7,23.8,22.8,23.1,23.1, 23.5,23,23) ?
Compare to x=c(22,23.9,20.9,23.8,25,24,21.7,23.8,22.8,23.1,23.1, 23.5,23,1)

MME: $\tilde{\alpha} = (\overline{X})^2/\hat{\sigma}^2 \approx 485.1$. It is too large for computing $\Gamma(\alpha)$.

Q: What else ??

Since $\tilde{\alpha} = 485.1 >> 100$, use Stirling's formula $n! \approx \sqrt{2\pi}n^{n+1/2}e^{-n}$.

If $\alpha = m$, then

$\mathcal{L} \approx (\prod_i X_i)^{m-1}(\frac{(\frac{m}{e\overline{X}})^m}{\sqrt{2\pi}(m-1)^{m-1/2}e^{-(m-1)}})^n$

$\approx (\prod_i X_i)^{m-1}((\frac{1}{\overline{X}})^m \frac{1}{\sqrt{2\pi}}(\frac{m}{m-1})^{m-1/2}m^{1/2}e^{-m+m-1})^n$

$\approx (\prod_i X_i)^{m-1}((\frac{1}{\overline{X}})^m \frac{1}{\sqrt{2\pi}}m^{1/2})^n$

$\approx (\prod_i X_i/(\overline{X})^n)^{m-1}((\frac{1}{\overline{X}})\sqrt{\frac{m}{2\pi}})^n$

```
#Use Stirling's formula
n=14
x=c(22,23.9,20.9,23.8,25,24,21.7,23.8,22.8,23.1,23.1, 23.5,23,23)
a=1:600
y=((prod(x)/mean(x)**n)**(a-1))*((sqrt(a/(2 *pi))/mean(x))**(n))
z=a[y==max(y)]
z
[1] 514
sum(x)/(n*z)
[1] 0.04496943
```

7.6. Let X_i's be i.i.d. from $f = x^{-2}\mathbf{1}(x > \theta)$. MME of θ ?

Sol. MME: $E(X^i) = \overline{X^i}$, i is an integer.

$(E(1/X) = \overline{1/X}$.

$E(1/X) = \int_\theta^\infty x^{-3}dx = \frac{1}{-2x}|_\theta^\infty = 1/(2\theta)$.

$\hat{\theta} = \overline{1/(2X)}$.

#7.10. $X_1, ..., X_n \sim f(x|\theta) = \frac{\alpha}{\beta^\alpha}x^{\alpha-1}\mathbf{1}_{(0<x<\beta)}$, $\theta = (\alpha, \beta)$. MLE of θ)

Solution: Note that $\mathcal{L}(\theta) = \frac{\alpha^n}{\beta^{n\alpha}}(\prod_{i=1}^n x_i)^{\alpha-1}\mathbf{1}(0 < x_{(1)}, x_{(n)} < \beta)$.

There are three ways to solve the problem.

The first way (bivariate):

$\frac{\partial \ln \mathcal{L}}{\beta} = \frac{-\alpha n}{\beta} < 0$, thus no stationary points.

The MLE must be on the boundary: $\alpha = 0$ or ∞, or $\beta = x_{(n)}$ or ∞, which are 4 straight line on the plane. Verify

θ :	$\alpha = 0$	$\alpha = \infty$	$\beta = \infty$	$\beta = x_{(n)}$
$\mathcal{L}(\theta)$:	0	0	0	*see below*

$\frac{\partial \ln \mathcal{L}}{\alpha} = \frac{n}{\alpha} - n\ln\beta + \ln\prod_{i=1}^n x_i = \frac{n}{\alpha} - n\ln x_{(n)} + \ln\prod_{i=1}^n x_i$ at $\beta = x_{(n)}$.

Thus $\frac{\partial \ln \mathcal{L}}{\alpha}|_{\beta=x_{(n)}} = 0$ yields the stationary point $\hat{\alpha} = \frac{1}{\ln x_{(n)} - \overline{\ln x}}$ on the line $\beta = x_{(n)}$.

Verify the boundary points and stationary point on the line $\beta = x_{(n)}$:

$\theta = (\alpha, x_{(n)})$:	$\alpha = 0$	$\alpha = \infty$	$\alpha = \hat{\alpha}$
$\mathcal{L}(\theta)$:	0	0	> 0

Thus $(\frac{1}{\ln x_{(n)} - \overline{\ln x}}, x_{(n)})$ is the MLE of θ.

The second way (one-by-one):

$\frac{\partial \ln \mathcal{L}}{\beta} = \frac{-\alpha n}{\beta} < 0$, thus for each α, the maximum of $\mathcal{L}(\alpha, \beta)$ over β is at $\beta = x_{(n)}$, denoted by $\hat{\beta}$. Thus it suffices to maximize $\mathcal{L}(\alpha, \hat{\beta})$ over α.

$\frac{\partial \ln \mathcal{L}(\alpha, \hat{\beta})}{\alpha} = 0$ yields $\hat{\alpha} = 1/(\ln x_{(n)} - \overline{\ln x})$.

Verify the boundary points and stationary point:

$\theta = (\alpha, x_{(n)})$:	$\alpha = 0$	$\alpha = \infty$	$\alpha = \hat{\alpha}$
$\mathcal{L}(\theta)$:	0	0	> 0

Thus $(\hat{\alpha}, x_{(n)})$ is the MLE of θ.

The third way (one-by-one):

For each β, $\frac{\partial \ln \mathcal{L}(\alpha,\beta)}{\alpha} = 0$ yields $\hat{\alpha} = \hat{\alpha}(\beta) = 1/(\beta - \overline{\ln x})$. Since $\frac{\partial^2 \ln \mathcal{L}(\alpha,\beta)}{\alpha^2} = -\frac{n}{\alpha^2} < 0$, for each β, $\mathcal{L}(\hat{\alpha}(\beta), \beta)$ reaches its maximum at $\alpha = \hat{\alpha}$.

Thus it suffices to maximize $\mathcal{L}(\hat{\alpha}(\beta), \beta)$ over β. Now

$$\mathcal{L}(\hat{\alpha}(\beta), \hat{\beta}) = (\frac{1}{\ln\beta - \overline{\ln x}})^n \beta^{-\frac{n}{\beta - \overline{\ln x}}} (\prod_{i=1}^{n} x_i)^{\frac{1}{b - \overline{\ln x}} - 1}.$$

It can be shown that $\frac{d \ln \mathcal{L}(\hat{\beta}(\beta),\beta)}{d\beta} < 0$, thus the maximum is obtained at $\beta = x_{(n)}$. That is the MLE of β is $\hat{\beta} = x_{(n)}$ and $\hat{\alpha}(\hat{\beta})$ is the MLE of α.

Homework Solutions, week 4

Additional Problem

Let X_1, ..., X_n be i.i.d. $\sim bin(1, p)$, and $p \sim U(0, 1)$. **Are X_i's i.i.d. ?**

Sol.

$f_{X_1}(x) = 0.5$ if $x \in \{0, 1\}$.

If $x_1 = \cdots = x_n = 1$.

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \int_0^1 \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} dp \\ &= \int_0^1 p^n dp = \frac{1}{n+1} \end{aligned}$$

Ans: No !

7.1. There are two MLEs, and one is $\hat{\theta} = \begin{cases} 1 & \text{if } x \in \{0, 1\} \\ 3 & \text{if otherwise.} \end{cases}$

#7.21. Assume $Y_i = x_i\beta + \epsilon_i$, where ϵ_i's are i.i.d. r.v. and x_i and β are constant. Compare the three estimators

$\hat{\beta}_1 = \sum_i x_i Y_i / \sum_i x_i^2$,

$\hat{\beta}_2 = \sum_i Y_i / \sum_i x_i$,

$\hat{\beta}_3 = \frac{1}{n} \sum_i \frac{Y_i}{x_i}$.

Sol.

$\sigma_1^2 = V(\sum_i x_i Y_i / \sum_i x_i^2) = \frac{\sigma^2}{n} / \overline{x^2}$,

$\sigma_2^2 = V(\sum_i Y_i / \sum_i x_i) = \frac{\sigma^2}{n} / (\overline{x})^2$,

$\sigma_3^2 = V(\frac{1}{n} \sum_i \frac{Y_i}{x_i}) = \frac{\sigma^2}{n} \overline{x^{-2}}$,

$\frac{1}{n} \sum_i (x_i - \overline{x})^2 = \overline{x^2} - (\overline{x})^2 \geq 0 \Rightarrow \overline{x^2} \geq (\overline{x})^2 \Rightarrow \sigma_1^2 \leq \sigma_2^2$.

$n^2 = (\sum_i x_i \frac{1}{x_i})^2 \leq \sum_i x_i^2 \sum_i \frac{1}{x_i^2} \Rightarrow \frac{1}{\overline{x^2}} \leq \overline{x^{-2}} \Rightarrow \sigma_1^2 \leq \sigma_3^2$.

Other relation ?

$1/(\overline{x})^2 \leq \overline{x^{-2}}$?

$1/(\overline{x})^2 \geq \overline{x^{-2}}$?

$1/(\overline{x})^2 \leq \overline{x^{-2}}$? Try $\overline{x} = 0+$ but $x_i \neq 0$.

$1/(\overline{x})^2 \geq \overline{x^{-2}}$? Try one $x_i = 0+$ but $\overline{x} \neq 0$.

7.23. If X_i's are a random sample from $N(\mu, \sigma^2)$, $S^2 \sim \frac{\sigma^2}{n-1} V = h(V)$, where $V \sim \chi^2(n-1)$. If $Y = \sigma^2$ has prior $f(y; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \frac{1}{y^{\alpha+1}} e^{\frac{-1}{\beta y}}$, $y > 0$. Bayes estimator of σ^2 ?

Sol. $S^2 = \frac{\sigma^2}{n-1} V = h(V)$ $(= T)$, where $V \sim \chi^2(n-1)$.

$h^{-1}(T) = \frac{n-1}{\sigma^2} T$.

Let $Y = \sigma^2$.

$f_{T|Y}(t|y) = f_{S^2|\sigma^2}(t|y) = f_V(h^{-1}(t)|J| = \frac{(\frac{n-1}{y} t)^{\frac{n-1}{2} - 1} e^{-(\frac{n-1}{y} t)/2}}{\Gamma(\frac{n-1}{2}) 2^{\frac{n-1}{2}}} \frac{n-1}{y}$, $t > 0$.

$\pi(y|t) = f_{\sigma^2|S^2}(y|t) \propto \frac{1}{y^{\frac{n-1}{2} - 1}} \exp(\frac{-1}{\frac{2}{(n-1)t} y}) y^{-1} \cdot y^{-\alpha-1} exp(\frac{-1}{\beta y})$

$= y^{-\alpha-1-\frac{n-1}{2}+1-1}\exp(\frac{-1}{\frac{2}{(n-1)t}y}+\frac{-1}{\beta y})$
$= \frac{1}{y^{\alpha^*+1}}\exp(\frac{-1}{\beta^* y})$, where ...
$E(Y) = \int_0^\infty t\frac{1}{\Gamma(\alpha)\beta^\alpha}\frac{1}{t^{\alpha+1}}e^{\frac{-1}{\beta t}}dt$
$= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha}\frac{1}{t^{\alpha-1+1}}e^{\frac{-1}{\beta t}}dt$
$= \frac{\Gamma(\alpha-1)\beta^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty \frac{1}{\Gamma(\alpha-1)\beta^{\alpha-1}}\frac{1}{t^{\alpha-1+1}}e^{\frac{-1}{\beta t}}dt$
$= \frac{1}{(\alpha-1)\beta}$.
$E(Y|T) = ...$

7.12. Compare the MLE $\hat\theta = \min\{\overline{X}, 1/2\}$ and the MME $\tilde\theta = \overline{X}$.
Sol. $MSE(\tilde\theta) = \frac{\theta(1-\theta)}{n} = \sum_{i=0}^n(\frac{i}{n}-\theta)^2\binom{n}{i}\theta^i(1-\theta)^{n-i}$.
$MSE(\hat\theta) = \sum_{i\le n/2}(\frac{i}{n}-\theta)^2\binom{n}{i}\theta^i(1-\theta)^{n-i} + \sum_{i>n/2}(\frac{1}{2}-\theta)^2\binom{n}{i}\theta^i(1-\theta)^{n-i}$.

$$MSE(\hat\theta) - MSE(\tilde\theta) = \sum_{i>n/2}[(\frac{1}{2}-\theta)^2 - (\frac{i}{n}-\theta)^2]\binom{n}{i}\theta^i(1-\theta)^{n-i}\begin{cases} = 0 & \text{if } \theta = 0 \\ < 0 & \text{if } \theta \in (0, 1/2] \end{cases}$$

Homework Solutions, week 5

§**7.57.** Let X_i's be i.i.d. $\sim bin(1,p)$. $h(p) = P(\sum_{i=1}^n X_i > X_{n+1})$. UMVUE of $h(p)$?
Sol. Let $\tilde p = \mathbf{1}_{(\sum_{i=1}^n X_i > X_{n+1})}$. $\hat p = E(\tilde p|T)$, where $T = \sum_{i=1}^{n+1} X_i \sim bin(n+1, p)$. Then $\hat p$ is the UMVUE of $h(p)$.
$E(\hat p) = P(\sum_{i=1}^n X_i > X_{n+1}) = h(p)$.
$\hat p = 0\cdot P(\tilde p = 0|T) + 1\cdot P(\tilde p = 1|T) = P(\tilde p = 1|T)$.
$P(\tilde p = 1|T = t) = ?$
If $t = 0$, $P(\tilde p = 1|T = t) = 0$.
If $t > 2$, $P(\tilde p = 1|T = t) = 1$.
If $t = 1$, $P(\tilde p = 1|T = t) = P(\sum_{i=1}^n X_i = 1, X_{n+1} = 0)/P(T = 1)$.
If $t = 2$, $P(\tilde p = 1|T = t) = P(\sum_{i=1}^n X_i = 2, X_{n+1} = 0)/P(T = 1)$.
The $\hat p = E(\tilde p|T) = \begin{cases} 0 & \text{if } T = 0 \\ \cdots & \cdots \\ 1 & \text{if } T > 2. \end{cases}$

Homework Solutions, week 6

D. Assume that X_1, ..., X_{100} are i.i.d. from $N(0,1)$. $T = \overline{X}\vee 0$, $Y = \mathbf{1}(T > 1)$. Check which of the following equations are correct. If so, given the explicit expressions of the density functions involved and complete the calculation; otherwise, make proper corrections based on the given density functions.
D.1 $E(Y-1) = \sum_x x f_{Y-1}(x)$
D.2 $E(Y-1) = \int (x-1) f_Y(x)dx$.
D.3 $E(Y-1) = \int (x\vee 0) - 1) f_{\overline{X}}(x)dx$.
D.4 $E(Y-1) = \int_1^\infty (x-1) f_T(x)dx$.
D.5 $E(Y-1) = \int\cdots\int (\mathbf{1}(\overline{x}\vee 0) - 1) f_{\mathbf{X}}(\mathbf{x})dx_1\cdots dx_n$.
Sol. Formula:

$$E(g(Y)) = \begin{cases} \sum_t g(t)f_Y(t) = \sum_t t f_{g(Y)}(t) & \text{if } Y \text{ and } g(Y) \text{ are discrete} \\ \cdot & \text{if } \cdot \\ \int g(t)f_Y(t)dt = \int t f_{g(Y)}(t)dt & \text{if } Y \text{ and } g(Y) \text{ are continuous} \end{cases}$$

$$\underbrace{Y-1}_{g(Y)} = \underbrace{\mathbf{1}(T>1)-1}_{g(T)} = \underbrace{\mathbf{1}((\overline{X}\vee 0) > 1) - 1}_{g(\overline{X})} = \underbrace{\mathbf{1}(\frac{X_1+\cdots+X_n}{n}\vee 0) > 1) - 1}_{g(\mathbf{X})}.$$

D.1 $Y \in \{0,1\}$, $Y-1 \in \{-1, 0\}$, $f_{Y-1}(t) = P(Y-1 = t)$.
$E(Y-1) = \sum_x x f_{Y-1}(x) = 0P(Y-1 = 0) - 1P(Y-1 = -1)$

$$= -P(Y=0) = -P(T\le 1) = -P((\overline{X}\vee 0)\le 1) = -(1 - P(\overline{X} > 1)) = -1 + 1 - \Phi(10)$$

$E(Y-1) \approx -1$.

D.2 $E(Y-1) = \sum_x (x-1) f_Y(x) = (0-1)P(Y=0) + (1-1)P(Y=1) \neq \int (x-1) f_Y(x) dx$, as Y is discrete.

D.3 $E(Y-1) = \int (\mathbf{1}((x \vee 0) > 1) - 1) f_{\overline{X}}(x) dx \neq \int ((x \vee 0) - 1) f_{\overline{X}}(x) dx$, where $\overline{X} \sim N(0, 1/100)$.

D.4 $E(Y-1) = \int (\mathbf{1}(x > 1) - 1) f_T(x) dx + (0-1) f_T(0) = \int_0^\infty f_T(x) dx - 1$
$\neq \int_1^\infty (x-1) f_T(x) dx$,
where

$$f_T(t) = 0.5^{\mathbf{1}(t=0)} (f_{\overline{X}}(t))^{\mathbf{1}(t>0)}$$

D.5 $E(Y-1) = \int \cdots \int (\mathbf{1}((\frac{\sum_{i=1}^n x_i}{n} \vee 0) > 1) - 1) f_{\mathbf{X}}(\mathbf{x}) dx_1 \cdots dx_n \neq \int \cdots \int (\mathbf{1}((\frac{\sum_{i=1}^n x_i}{n} \vee 0)) - 1) f_{\mathbf{X}}(\mathbf{x}) dx_1 \cdots dx_n$. $\blacksquare$

7.49.
1. Find an unbiased estimator of λ based only on $Y = \min(X_i)$
2. Find a better estimator than the one in part (a), prove it is better.
3. The following data are high stress failure times (in hours) of Kevlar/expoxy spherical vessels used in a sustained pressure environment on the space shuttle:

$$50.1, 70.1, 137.0, 166.9, 170.5, 152.8, 80.5, 123.5, 112.6, 148.5, 160.0, 125.4$$

Failure times are often modeled with the exponential distribution. Estimate the mean failure time using the stimators from part (a) and (b)

Sol. 1. $Y = X_{(1)} \sim Exp(\lambda/n)$. We have $E(nY) = \lambda$. The $\hat{\lambda} = nY$, as $E(nY) = n\lambda/n = \lambda$.

2. We know X_i belong to an exponential family. Therefore by Theorem 6.2.25 we know $\sum X_i$ is a sufficient and complete statistic. By Theorem 7.3.23 we know $\overline{X}$ is the best unbiased estimator. Theorefore it must be better than $T = nY$ since theorem 7.3.23 dictates the UMVUE in this situation is unique. $P(nX_{(1)} = \overline{X}) = 0$. Thus $Var(T) > Var(\overline{X})$.

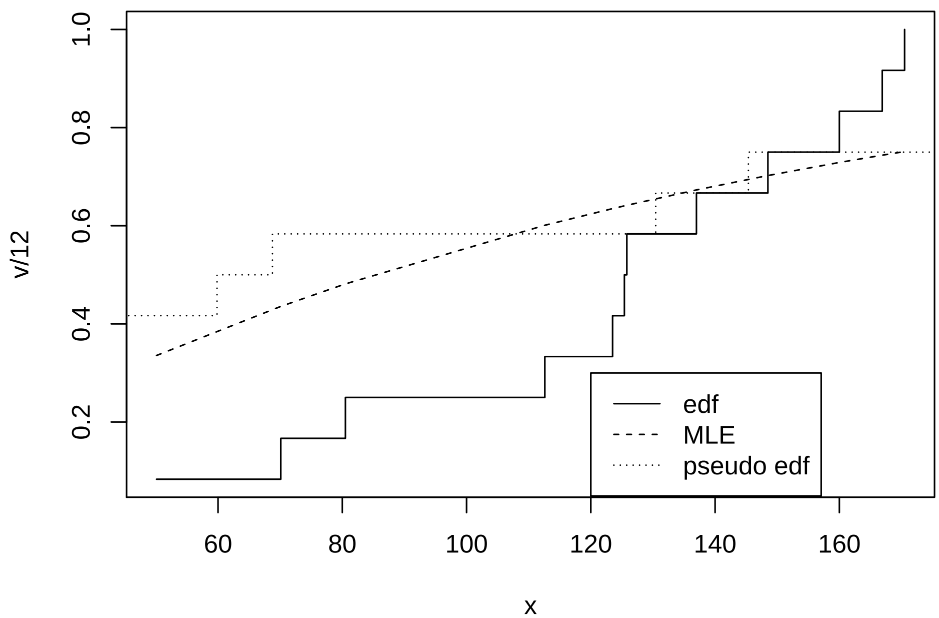
In fact, $V(T) = \lambda^2 < \lambda^2/n = V(\overline{X})$.

3. We have $nY = 601.2$ and $\overline{X} = 128.8$
Something looks fishy.

Compare the empirical distribution function edf to the MLE of cdf, where the edf $\hat{F}(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i \leq t)$.

Compare the edf of pseudo random numbers to the MLE of cdf.

From the figure, we can conclude that the data does not fit the expnential distribution. No wonder the two estimates differ so large !



7.46. X_1, X_2, X_3 i.i.d. from $U(\theta, 2\theta)$, $\theta > 0$.

a. MME $\hat{\theta} = \frac{2}{3}\overline{X}$.

b. MLE induces unbiased estimator $\tilde{\theta} = \frac{4}{7}X_{(3)}$.

c. Which of the two estimators can be improved by sufficiency ? How ?

Sol. **Which ?** $T = (X_{(1)}, X_{(3)})$ is sufficient for θ.

$E(\tilde{\theta}|T) = \tilde{\theta}$, thus $\tilde{\theta}$ cannot be improved.

$\vdash$: $E(\hat{\theta}|T) = \frac{2}{3}\frac{X_{(1)}+X_{(3)}}{2}$.

Two ways to prove:

$E(\hat{\theta}|T) = \frac{2}{3}E(\frac{X_1+X_2+X_3}{3}|T) = \frac{2}{3}E(X_1|T) = \frac{2}{3}\frac{X_{(1)}+X_{(3)}}{2}$.

$E(\hat{\theta}|T) = \frac{2}{3}E(\frac{X_{(1)}+X_{(2)}+X_{(3)}}{3}|T)$

$= \frac{2}{3\cdot 3}[E(X_{(1)} + X_{(3)}|T) + E(X_{(2)}|T)]$

$= \frac{2}{3\cdot 3}[(X_{(1)} + X_{(3)}) + \frac{X_{(1)}+X_{(3)}}{2}]$

$= \frac{2}{3}\frac{X_{(1)}+X_{(3)}}{2}$.

$$
\begin{aligned}
f_{X_{(2)}|X_{(1)},X_{(3)}}(x|y,z) =& \frac{f_{X_{(2)},X_{(1)},X_{(3)}}(x,y,z)}{f_{X_{(1)},X_{(3)}}(y,z)} \\
=& \frac{\frac{3!}{1!1!1!}(f_X(x))^1(f_X(y))^1(f_X(z))^1}{\frac{3!}{1!1!1!}(f_X(x))^1(F_X(z)-F_X(y))^1(f_X(z))^1} \\
=& \frac{\mathbf{1}(x \in (y,z))}{z-y}
\end{aligned}
$$

Thus $\hat{\theta}$ can be improved.

HOW ?

either compare $V(\hat{\theta})$ and $V(E(\hat{\theta}|T))$,
or show $E(V(\hat{\theta}|T)) > 0$, as $V(\tilde{\theta}) = +E(V(\tilde{\theta}|T))$.

$V(E(\hat{\theta})|T) = V(\frac{2}{3}\frac{X_{(1)}+X_{(3)}}{2})$.
$V(X_{(1)}) = \int_\theta^{2\theta} t^2 3(1-F_X(t))^{3-1} f_X(t)dt - (\theta+\frac{1}{4}\theta)^2$.
$V(X_{(3)}) = \int_\theta^{2\theta} t^2 3(F_X(t))^{3-1} f_X(t)dt - (\theta+\frac{3}{4}\theta)^2$.
$Cov(X_{(1)}, X_{(3)})$
$= \int_\theta^{2\theta}\int_\theta^y xy3\cdot 2f_X(x)(F_X(y)-F_X(x))^{3-2} f_X(y)dxdy - (\theta+\frac{1}{4}\theta)(\theta+\frac{3}{4}\theta)$.

Sol to d. Estimates based on data using R:

```
x=c(1.29,.86,1.33)
2*mean(x)/3
4*max(x)/7
```

7.51. Let X_i's be i.i.d. $N(\theta,\theta^2)$, $\theta>0$. $\mathcal{T}=\{T: T=a_1\overline{X}+a_2(cS)\}$, where $E(cS)=\theta$.
(a) Solve $T^* = argmin_{T\in\mathcal{T}} MSE(T)$.
(b) Show $MSE(T^*) < MSE(T^o)$ from #7.50
(c) Show that $MSE(T^*) < MSE(T^{*+})$, where $T^{*+} = T \vee 0$.
(d) Is θ a scale or location parameter ?

Sol. Since $\overline{X} \perp S^2$,
$MSE(T^*) = V(T^*) + (bias(T^*))^2 = a_1^2 V(\overline{X}) + a_1^2 V(cS) + ((a_1+a_2-1)\theta)^2$
$= \theta^2(a_1^2/n + a_2^2(c^2-1) + (a_1+a_2-1)^2) = \theta^2 g(a_1,a_2)$.
$\frac{\partial g}{\partial a_i} = 0$ yields
$a_1^* = \frac{n(c^2-1)}{(n+1)c^2-n}$ and $a_2^* = \frac{1}{(n+1)c^2-n}$.

Check: 1 point (a_1^*, a_2^*) and 4 boundary lines $a_i = \pm\infty$.
Since $g(a_1,a_2)\to\infty$ if $a_i\to\pm\infty$,
(a_1^*,a_2^*) is the unique minimum point and $T^* = a_1^*\overline{X} + a_2^*(cS)$.

(b) $MSE(T^*) < MSE(T^o)$, as
(1) $T^o \in \mathcal{T}$, where $a_2^o = a_2^*$ and $a_1^o = 1-a_2^o$,
(2) $P(T^*\neq T^o) = 1$,
(3) T^* is the unique minimum point in $\mathcal{T}$.

(c) **Find the correct solutions among the following approaches:**
(c.1) $MSE(T^*) = E((T^*-\theta)^2)$
$= E((T^*-\theta)^2\mathbf{1}(T^*<0)) + E((T^*-\theta)^2\mathbf{1}(T^*\geq 0))$
$\geq E((T^*-\theta)^2\mathbf{1}(T^*\geq 0)) = MSE(T^*\vee 0)$
(c.2) $MSE(T^*) = E((T^*-\theta)^2)$
$= E((T^*-\theta)^2\mathbf{1}(T^*<0)) + E((T^*-\theta)^2\mathbf{1}(T^*\geq 0))$
$> E((T^*-\theta)^2\mathbf{1}(T^*\geq 0))$ as $P(T^*<0)>0$
$= MSE(T^*\vee 0)$.
In fact, $P(T^*<0) = \int_0^\infty\int_{-\infty}^{-a_2^o y/a_1^o} f_{\overline{X}}(x) f_{cS}(y)dxdy > 0$,
as $f_{cS}(y)>0$ on $(0,\infty)$ and $f_{\overline{X}}(x)>0\ \forall\ x$.
(c.3) If $T^*\geq 0$, then $T^* = T^{*+}$, thus their MSE's are the same. Otherwise, $T^* < 0 = T^{*+}$,
then $(T^{*+}-\theta)^2 = \theta^2$ and $(T^*-\theta)^2 > \theta^2$, thus $MSE(T^*) > MSE(R^{*+})$.
(c.4) If $T^*<0$, $MSE(T^{*+}) = (0-\theta)^2 \leq E((T^*-\theta)^2)$ as $\theta>0$,
otherwise, they are the same.
(d) $\vdash$: θ is not a location parameter.
If θ is a location parameter, then it is possible that $\theta=-1$, but $\theta>0$. A contradiction.
$\vdash$: θ is a scale parameter.
$X\sim N(\theta,\theta^2)$,
$=> f_X(t) = \frac{1}{\theta\sqrt{2\pi}} e^{-(\frac{t}{\theta}-1)^2/2}$
$=> f_X(t) = \frac{1}{\theta} f_Y(t/\theta)$, where $f_Y(t) = \frac{1}{\sqrt{2\pi}} e^{-(t-1)^2/2}$, $Y\sim N(1,1)$

7.52. (2) Prove the rather remarkable identity $E(S^2|\overline{X}) = \overline{X}$ if $X\sim Poisson(\lambda)$.

That is, prove directly $E(S^2|\overline{X}) = \overline{X}$ if $X\sim Poisson(\lambda)$.

(3) Use completeness, form a general theorem ...

Possible approach: (1) Find $f_{S^2|\overline{X}}$= ? (2) $f_{\mathbf{X}|\overline{X}}$ = ? (3) first simplify $E(S^2|\overline{\mathbf{X}})$.

Recall $S^2 = \frac{n}{n-1}[\overline{X^2} - (\overline{X})^2]$.
Let $T = \sum_i X_i$, then
$E(S^2|\overline{X} = v) = E(S^2|T = t)$, where $t = nv$.

Consider the case $n = 2$.
$E(S^2|\overline{X} = v) = 2[E(\overline{X^2}|\overline{X} = v) - (v)^2]$.
$f_{X_1|X_1+X_2}(x|t) = P(X_1 = x, X_2 = t - x)/P(X_1 + X_2 = t)$
$= \frac{(e^{-\lambda}\lambda^x/x!)(e^{-\lambda}\lambda^{t-x}/(t-x)!)}{(e^{-2\lambda}(2\lambda)^t/t!)} = \binom{t}{x}0.5^t$
$E(X_1^2 + X_2^2|T = t) = 2E(X_1^2|T = t) = 2(tpq + (tp)^2) = 2(t/4 + t^2/4)$.
$E(S^2|\overline{X} = v) = 2[\frac{1}{2}2(t/4 +^2 /4) - v^2] = t/2 + t^2/4 - v^2 = v$, as $v = t/2$.
In general, $n \geq 2$.

$X_i|\sum_i X_i \sim bin(t, 1/n)$, $i = 1, ..., n$.
Let $t = nv$.
$E(S^2|\overline{X} = v) = E(S^2|\overline{X} = v)$
$= \frac{n}{n-1}E(\overline{X^2} - (\overline{X})^2|\overline{X} = v)$
$= \frac{n}{n-1}E(\overline{X^2} - (t/n)^2|\overline{X} = v)$
$= \frac{n}{n-1}E(\overline{X^2}|\overline{X} = v) - (t/n)^2$
$= \frac{n}{n-1}(E(X_1^2|\overline{X} = v) - (v)^2)$
$f_{X_1|\overline{X}}(x|v)$
$= \frac{P(X_1=x,\sum_{i=1}^n X_i=nv)}{P(\sum_{i=1}^n X_i=nv)}$
$= \frac{P(X_1=x,T=t)}{P(T=t)}$, where $t = nv$,
$= \frac{P(X_1=x,\sum_{i=2}^n X_i=t-x)}{P(T=t)}$
$= \binom{t}{x}(\frac{1}{n})^x(1 - \frac{1}{n})^{t-x}$, $x = 0, ..., t$.

$X_1|\sum_i X_i \sim bin(t, 1/n)$,

$E(S^2|\overline{X} = v)$
$= \frac{n}{n-1}(t\frac{1}{n}(1 - \frac{1}{n}) + (t/n)^2 - (t/n)^2)$
$= t/n = v$
Thus $E(S^2|\overline{X}) = \overline{X}$.
$E(S^2|T = t)$
$= \frac{n}{n-1}E(\overline{X^2} - (\overline{X})^2|T = t)$
$= \frac{n}{n-1}E(\overline{X^2} - (t/n)^2|T = t)$
$= \frac{n}{n-1}(E(X_1^2|T = t) - (t/n)^2)$
$= \frac{n}{n-1}(t\frac{1}{n}(1 - \frac{1}{n}) + (t/n)^2 - (t/n)^2)$
$= t/n$
(3) A general formula is $E(\hat{\tau}|W) = W$ if W is a sufficient and complete statistic and if $E(W) = \tau$ and $E(\hat{\tau}) = \tau$.
Q: Is $E(S^2|X_1)$ random variable ?
Is $E(S^2|X_1)$ a statistic ?
$E(S^2|T)$ **is a statistics if** T **is sufficient. Why ?**

7.59. X_i's are i.i.d. from $N(\mu, \sigma^2)$, UMVUE of σ^p, where $p > 0$.
Sol. It is known that $T = (\overline{X}, S^2)$ is suf and complete for $\theta = (\mu, \sigma^2)$ (due to exponential family). $Y = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n - 1)$.
$\chi^2(d) = G(d/2, 2)$.

$E(Y^{p/2}) = \int_0^\infty y^{p/2} \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} e^{-y/\beta} dy$
$= \frac{\Gamma(\frac{n-1}{2}+p/2)2^{\frac{n-1}{2}+p/2}}{\Gamma(\frac{n-1}{2})2^{\frac{n-1}{2}}} = \frac{\Gamma(\frac{n+p-1}{2})2^{p/2}}{\Gamma(\frac{n-1}{2})} = c.$
$\hat\sigma^p = \frac{((n-1)S^2)^{p/2}}{c}$ is UMVUE of σ^p.

7.44 $\overline{X} \sim N(\theta, 1/n)$,
$V((\overline{X})^2 - 1/n) = V((\overline{X})^2) = E((\overline{X})^4) - (E((\overline{X})^2))^2$.
$E((\overline{X})^4) = E((\overline{X})^3(\overline{X}-\theta)) + \theta E((\overline{X})^2(\overline{X}-\theta)) + \theta^2 E((\overline{X})^2)$
$= \frac{1}{n}[E(3(\overline{X})^2) + \theta E(2\overline{X}))] + \theta^2 E((\overline{X})^2)$
$= (\frac{3}{n} + \theta^2)E((\overline{X})^2) + \frac{2\theta}{n}E(\overline{X}))$.

$V((\overline{X})^2 - 1/n) = E((\overline{X})^4) - (E((\overline{X})^2))^2$
$= (\frac{3}{n} + \theta^2)E((\overline{X})^2) - (E((\overline{X})^2))^2 + \frac{2\theta}{n}E(\overline{X}))$
$= (\frac{3}{n} + \theta^2 - \frac{1}{n} - \theta^2)(\frac{1}{n} + \theta^2) + \frac{2\theta^2}{n}$
$= \frac{2}{n}(\frac{1}{n} + \theta^2) + \frac{2\theta^2}{n}$
$= \frac{2}{n^2} + \frac{4\theta^2}{n}$.

7.60 X_i's are i.i.d. from $G(\alpha, \beta)$ with α known. UMVUE of $\theta = 1/\beta$?
Sol. $T = \sum_i X_i$ is suf and complete.
$T \sim G(n\alpha, \beta)$,
$E(T) = n\alpha\beta$.
Try $Y = \frac{n\alpha}{T}$.
If $E(Y) = c/\beta$, then $\hat\theta = \frac{n\alpha}{c\sum_i X_i}$ is the UMVUE of θ.
In fact, $E(Y) = \int_0^\infty n\alpha t^{-1} \frac{t^{\alpha-1}e^{-t/\beta}}{\Gamma(\alpha)\beta^\alpha} dt$
$= \frac{n\alpha\Gamma(\alpha-1)\beta^{-\alpha-1}}{\Gamma(\alpha)\beta^\alpha}$.

Homework solutions for week 7

Additional questions:
Remark. For a test $\phi = \mathbf{1}(\hat\theta \in RR)$ with estimate $\hat\theta = t$ and H_0: $\theta = \theta_o$,

$$\text{the P-value} = \begin{cases} P(\hat\theta > t) & \text{if right-sided test } (H_1) \\ P(\hat\theta < t) & \text{if left-sided test } (H_1) \\ P(|\hat\theta - \theta_o| > |t - \theta_o|) & \text{if two-sided test} \end{cases}.$$

2. Carry out the following simulation project.
   2.1. Use R to generate 5 observations from $N(1,1)$. Now pretend that you only known that the data were from $N(\mu, \sigma)$ without knowing μ and σ, use t-test to test H_0: $\mu = 0$ v.s. H_1: $\mu \neq 0$ with a size 0.2. Record the P-value.
   What is a correct decision here (in terms of rejecting H_0 or not) ?
   Do you think that you will accept H_0 based on data ? Why ?
   2.2. Repeat procedure 2.1 100 times. That is, record 100 P-values.
   How many times, say z, would you reject H_0 ?
   Question: What does the number z tell you about $P(H_0|H_1)$?

Sol. 2.1. It is a correct decision to reject H_0,
but we may not reject as the test statistic is random.
2.2. One record is z=73.
z would tell me that $P(H_0|H_1) \approx 1 - z/100 = 0.27$. In fact $P(H_0|H_1) = P(\frac{|\overline{X}|}{S/\sqrt{n}} < t_{0.1,n-1}, \mu = 1) \approx$ 0.24.

2. Carry out the following simulation project.
   2.1. Use Splus to generate 5 observations from $N(1,1)$. Now pretend that you only known that the data were from $N(\mu, \sigma)$ without knowing μ and σ, use t-test to test H_0: $\mu = 0$ v.s. H_1: $\mu \neq 0$ with a size 0.2. Record the P-value. Splus commands are :
   $x < -rnorm(5) + 1$
   y=t.test(x)
   y$p.value

2.2. Repeat procedure 2.1 20 times. That is, record 20 P-values.

a. What is a correct decision here (in terms of rejecting H_0 or not) ?

b. How many times you will reject H_0 ?

c. **What does the number tell you about** $P(H_0|H_1)$ **?**

d. Do you think that you will accept H_0 based on data ? Why ?

Hint: *Figure out how to use t.test using* **?t.test** *in this case.* Notice t-test: $\phi = \mathbf{1}(|\frac{\overline{X}-\mu_0}{S/\sqrt{n}}| > t_{\alpha/2}$, where $\mu_0 = 0$ and $\alpha = 0.2$.

t.test(x, alternative="two.sided", mu=0) "greater", "less"

Answer:

a. Correct decision is to reject H_0, as the data are from $N(1, 1)$, not $N(0, 1)$.

b. reject H_0 15 times (in one study).

c. An estimate of $P(H_0|H_1)$ is $1 - 15/20$.

d. No. not necessary. Remember we only observe one sample in reality. Due to type II error, the t.test may suggest incorrectly to accept H_0.

3. Carry out the following simulation project.

3.1. Use Splus to generate 5 observations from $N(1, 1)$. Now pretend that you only known that the data were from $N(\mu, \sigma)$ without knowing μ and σ, use t-test to test H_0: $\mu = 1$ v.s. H_1: $\mu < 1$ with a size 0.05. **Figure out how to use t.test in this case.** Record the P-value.

3.2. Repeat procedure 3.1 20 times. That is, record 20 P-values.

a. What is a correct decision here (in terms of rejecting H_0 or not) ?

b. How many times you will reject H_0 ?

c. **What does the number tell you about** $P(H_1|H_0)$ **?**

d. Do you think that you will accept H_0 based on data ? Why ?

Answer:

a. Correct decision is to accept H_0, as the data are from $N(1, 1)$.

b. reject H_0 2 times (in one study).

c. An estimate of $P(H_1|H_0)$ is $2/20$.

d. No, due to type I error, the t.test may suggest incorrectly to reject H_0 approximately 1 time (20×0.05), though we shall not reject H_0, as the data are from $N(1, 1)$.

8.1 Solution without using LRT. 5 elements of a test:

1,2,3. H_0: $p = 0.5$, v.s. H_1: $p \neq 0.5$.

A natural estimate of p is X/n.

4. Test Statistic $\mathbf{1}_{(X/n \notin (a,b))}$, where $X \sim bin(n, p)$ and $n = 1000$.

RR $X/1000 \notin (a, b)$.

R

> p=2*pbinom(440,1000,0.5)

5. Since p-value < 0.0001, reject H_0.

8.2. In a city the number of auto accidents used to follow Poisson(15). If this year the # is 10, is it justified that the # dropped?

Ans. 5 elements of a test.

1. 2. H_0: $\mu = 15$, vs. H_1: $\mu < 15$.

Note that the sample size $n = 1$ and the observation is $X = 10$.

LRT:

MLE: $\hat{\mu}_0 = 15$, and $\hat{\mu} = X$.

$\lambda = ...$

3. test statistic reduces to $\phi = \mathbf{1}_{(\lambda \le c_o)} = \mathbf{1}_{(X \le c)}$.

4. α:

R

> round(ppois(5:11,15),2)

[1] 0.00 0.01 0.02 0.04 0.07 0.12 0.18

If we take $\alpha = 0.07$, then $\phi = \mathbf{1}_{(X \le 9)}$.

(If we take $\alpha = 0.01$, then $\phi = \mathbf{1}_{(X \le 6)}$.)

P-value= $P(X \le 10) = \sum_{i=0}^{10} e^{-15}(15)^i/i! = 0.12 > 0.05$.

5. Conclusion: It is not justified that the # dropped this year.

8.5. Suppose X_1, ..., X_n are iid with
$f_X(x) = \frac{\theta\nu^\theta}{x^{\theta+1}}\mathbf{1}(\nu \le x < \infty),\ \theta > 0,\ \nu > 0.$

a. Find the MLE of θ and ν.

b. Show that the LRT of
H_0: $\theta = 1$, ν unknown vs. H_1: $\theta \ne 1$, ν unknown,
is $\phi = \mathbf{1}(T \notin (c_1, c_2))$, where $T = \log\left[\frac{\prod_{i=1}^n X_i}{(\min_i X_i)^n}\right]$.

c. Show that under H_0, $2T$ has a χ^2 distribution.

Question: Why ask question c ?

Answer: Among the 5 elements of a test, we need to choose α, and for $\phi = \mathbf{1}_{(\lambda \le t)}$, we need to know $t = ?$ Otherwise, it is not a test.

One approach. Let $Y = logX$, then $h^{-1}(y) = e^y$ and $|J| = e^y$.
$f_Y(y) = f_X(e^y)e^y = \nu e^{-y}$, $y > \nu$.
Reorder $(Y_1, ..., Y_n)$ as $(\tilde{Y}_1, ..., \tilde{Y}_n)$ such that
$\tilde{Y}_1 = Y_{(1)}$ and $(\tilde{Y}_2, ..., \tilde{Y}_n)$ the rest Y_j's.
Then $f_{\tilde{Y}_1,...,\tilde{T}_n}(\mathbf{y}) = nf_Y(y_1)\prod_{j=2}^n f_Y(y_j)$, $y_1 < y_j$ for $j \ge 2$. **??? need a proof.**

Under certain condition, $f_Y(t) = \sum_i f_X(g_i^{-1}(t))|\frac{\partial g_i^{-1}}{\partial t}|$, where
(a) g_i is a 1-1 map from A_i to $g(A_i)$,
(b) $A_1, ..., A_k$ are disjoint and
(c) $P(X \in \cup_i A_i) = 1$.

What are A_i here ?

$\tilde{Y}_1 = Y_{(1)}$ and
$(\tilde{Y}_2, ..., \tilde{Y}_n) = (Y_2, ..., Y_n)$ if $Y_1 = Y_{(1)}$,
$(\tilde{Y}_2, ..., \tilde{Y}_n) = (Y_1, Y_3, ..., Y_n)$ if $Y_2 = Y_{(1)}$,
......
$(\tilde{Y}_2, ..., \tilde{Y}_n) = (Y_1, Y_2, ..., Y_{n-1})$ if $Y_n = Y_{(1)}$,

Sol.

a. $\mathcal{L}(\theta, \nu) = \theta^n\nu^{n\theta}(\prod_i X_i)^{-\theta-1}\mathbf{1}(\nu \le X_{(1)})$.

For each θ, $\mathcal{L} \uparrow$ in ν for $\nu \le X_{(1)}$. Thus $\mathcal{L}(\theta, \nu) \le \mathcal{L}(\theta, X_{(1)})$. That is, the MLE of ν is $\hat{\nu} = X_{(1)}$, which does not depend on θ. To find the MLE of θ, it suffices to maximize $\mathcal{L}(\theta, X_{(1)})$. $\frac{\partial \ln\mathcal{L}}{\partial\theta} = \frac{n}{\theta} + n\ln X_{(1)} - \sum_i \ln X_i = 0$ yields

$$\hat{\theta} = 1/(\overline{\ln X} - \ln X_{(1)}) = n/ln(\prod_i X_i/X_{(1)}^n) = n/T.$$

$\frac{\partial^2\ln\mathcal{L}}{\partial\theta^2} = -\frac{n}{\theta^2} < 0$ implies that $(\hat{\theta}, \hat{\nu})$ is the MLE of (θ, ν).

b. It is easy to show that the MLE under H_0 is $(\hat{\theta}^o, \hat{\nu}^o) = (1, X_{(1)})$. Thus

$$\lambda = (\hat{\theta})^{-n}(\frac{\prod_i X_i}{X_{(1)}^n})^{\hat{\theta}-1} = (n/T)^{-n}(e^T)^{\frac{n}{T}-1} = n^{-n}T^ne^{n-T}.$$

$$\phi = \mathbf{1}(\lambda \le c) = \mathbf{1}(T^ne^{-T} \le c_1)$$

Let $g(T) = \ln(T^ne^{-T})$, $g' = \frac{n}{T} - 1$ and $g'' = -\frac{n}{T^2} < 0$.

Thus g is concave down with maximum point $T = n$. It follows that

$$\phi = \mathbf{1}(\lambda \le c) = \mathbf{1}(T^ne^{-T} \le c_o) = \mathbf{1}(T \notin (c_1, c_2))$$

where $c_1e^{-c_1/n} = c_2e^{-c_2/n}$ and $P\{c_1 < T < c_2) = 1 - \alpha$.

8.5.c. Suppose X_1, ..., X_n are iid with
$f_X(x) = \frac{\theta\nu^\theta}{x^{\theta+1}}\mathbf{1}(\nu \le x < \infty),\ \theta > 0,\ \nu > 0.$ (Pareto Distribution).

Show that under H_0 ($\theta = 1$), $2T$ has a χ^2 distribution,

where $T = \log\left[\frac{\prod_{i=1}^n X_i}{(\min_i X_i)^n}\right]$.

Question: Why ask question c ?

Answer: Among the 5 elements of a test, we need to choose α, and for $\phi = \mathbf{1}_{(\lambda \le t)}$, we need to know $t =$? Otherwise, it is not a test.

Under H_0, $f(x) = \frac{\nu}{x^2}\mathbf{1}(x \ge \nu)$ and $F(x) = 1 - \frac{\nu}{x}$, if $x/\nu \ge 1$.

Notice that $Z = X/\nu$ is a pivatol, as $F_Z(t) = 1 - \frac{1}{t}$, $t \ge 1$.

$$T = \log(\frac{\prod_{i=1}^n X_i}{(X_{(1)})^n}) = \log\prod_{i=1}^n \frac{X_{(i)}}{X_{(1)}} = \underbrace{\log\prod_{i=1}^n \frac{Z_{(i)}}{Z_{(1)}}}_{Z_{(i)}=X_{(i)}/\nu} = \sum_{i=1}^n \log\frac{Z_{(i)}}{Z_{(1)}} = \sum_{i=2}^n Y_{(i)}$$

where $Y_{(i)} = \log(\frac{Z_{(i)}}{Z_{(1)}})$, $i \ge 1$. Notice that $Y_{(1)} = 0$.

$f_{X_{(2)},...,X_{(n)}|X_{(1)}}(x_2, ..., x_n|x_1) = \frac{f_{X_{(1)},...,X_{(n)}}(x_1,...,x_n)}{f_{X_{(1)}}(x_1)}$

$= \frac{n!f(x_1)\cdots f(x_n)}{\binom{n}{1}f(x_1)(1-F(x_1))^{n-1}}$

$= \frac{(n-1)!x_1^{n-1}}{(\prod_{i=2}^n x_i)^2}$, $\nu \le x_1 \le \cdots \le x_n$.

Notice that $f_{Z_{(i)}}$ does not depend on ν, as

$f_{Z_{(2)},...,Z_{(n)}|Z_{(1)}}(x_2, ..., x_n|x_1) = \frac{(n-1)!x_1^{n-1}}{(\prod_{i=2}^n x_i)^2}$, $1 \le x_1 \le \cdots \le x_n$.

Then $z_{(i)} = z_{(1)}e^{y_{(i)}}$, $i \ge 2$, $|J| = x_1^{n-1}\exp(\sum_{i=2}^n y_i)$ and

$f_{Y_{(2)},...,Y_{(n)}|Z_{(1)}}(y_2, ..., y_n|x_1)$

$= f_{Z_{(2)},...,Z_{(n)}|Z_{(1)}}(x_1e^{y_2}, ..., x_1e^{y_n}|x_1)|J|$

$= (n-1)!\exp(-\sum_{i=2}^n y_i)$, $0 \le y_2 \le \cdots \le y_n$.

Thus, $Y_{(2)}, ..., Y_{(n)}$ are order statistics of i.i.d. Y_2, ..., $Y_n \sim Exp(1)$.

$$T = \log(\frac{\prod_{i=1}^n X_i}{(X_{(1)})^n}) = \log(\frac{\prod_{i=1}^n Z_i}{(Z_{(1)})^n}) = \log(\frac{\prod_{i=2}^n Z_{(i)}}{(Z_{(1)})^{n-1}}) = \sum_{i=2}^n Y_{(i)} = \sum_{i=2}^n Y_i$$

$\sim Gamma(n-1, 1)$.

$\chi^2(m) = Gamma(m/2, 2)$.

Thus $2T \sim \chi^2(2(n-1))$.

Summary:

H_0: $\theta = 1$, vs. H_1: $\theta \ne 1$

$\alpha = 0.05 = E_{\theta_o}(\phi)$

$\phi = \mathbf{1}(\lambda \le a) = \mathbf{1}_{(\frac{W}{2}e^{-\frac{W}{2n}} \le c)} = \mathbf{1}_{(W \notin (a,b))}$, where

$W = 2T$, $\frac{a}{2}\exp(-\frac{a}{2n}) = \frac{b}{2}\exp(-\frac{b}{2n})$, and $F_W(b) - F_W(a) = 1 - \alpha$.

Question: If $n = 3$ and $2T = W = 8$. What is the conclusion ?

Ans: Need to find out (a, b) !

Note $2T \sim \chi^2(4)$, $E(2T) = 4$ and $V(2T) = 8$.

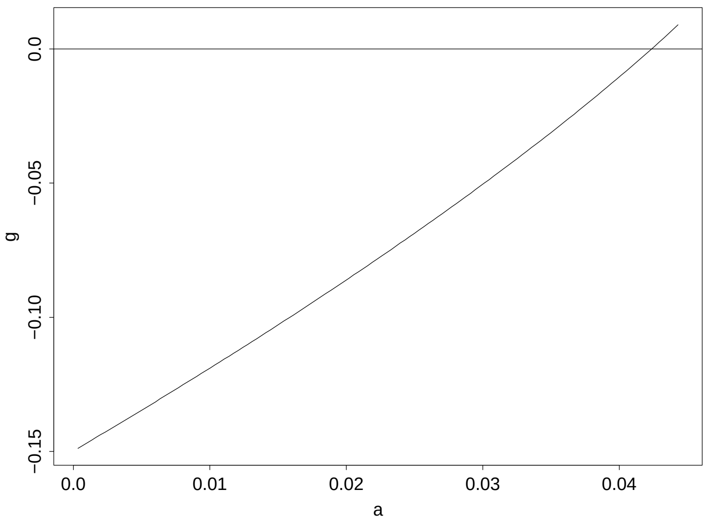
Solve numerically

$g(a) = \frac{a}{2}\exp(-\frac{a}{2n}) - \frac{b}{2}\exp(-\frac{b}{2n}) = 0$, where $b = F_W^{-1}(F_W(a) + 1 - \alpha)$.

R

```
x=(1:499)/10000 # probabilities in (0,0.05)
df = 4
a=qchisq(x) # t_1, quantile of Exponential at x
b=qchisq(pchisq(a,df)+0.95,df) # t_2,
g=(a/2)*exp(-(a/6))-(b/2)*exp(-(b/6))
```

plot(a,g,type="l")
abline(h=0) # check whether the curve cross the x-axis



(l=max(a[g<=0]))
(r=max(b[g<=0]))
pchisq(l,df)
[1] 0.7075065
[1] 20.48774
[1] 0.0496 # < 0.0499, make change if >0.0499
Ans: Do not reject H_0.
Additonal.
1. (a) **Under each of the assumptions in 8.5 and 8.7**, generate 10 observations from R, and do the tests.
   Remark. There are two issues:
      (1) How to generate 10 observations ?
      (2) How to determine RR or c in $\mathbf{1}(\lambda \le c)$?

8.5. Suppose X_1, ..., X_n are iid with
$f_X(x) = \frac{\theta\nu^\theta}{x^{\theta+1}}\mathbf{1}(\nu \le x < \infty)$, $\theta > 0$, $\nu > 0$, ν unknown, (Pareto Distribution).
The LRT of
H_0: $\theta = 1$ vs. H_1: $\theta \ne 1$.

(1) Data generation.
x=rnorm(10) # ?
x=rexp(10) ?
x=1/(1-runif(10)) # Pareto
Reason: $F^{-1}(X)$, where $X \sim U(0,1)$.
$f(x) = \frac{\theta v^\theta}{x^{\theta+1}}\mathbf{1}(x > v)$, $\theta > 0$ and $v > 0$.
Since H_0: $\theta = 1$, v.s. H_1: $\theta \ne 1$.
Select (θ, v), say $\theta = v = 1$.
$F(t) = \int_0^t f(x)dx = -1/x\big|_0^t = 1 - 1/t$, $t > 0$.
$F^{-1}(y) = \frac{1}{1-y}$, $y \in (0,1)$.

Reason: $F^{-1}(X)$, where $X \sim U(0,1)$.
(2) RR= ?
$\phi = \mathbf{1}(\lambda(W) \le c) = \mathbf{1}(W \notin (a,b))$,

where $\lambda = n^{-n}(\frac{W}{2})^n e^{n-(\frac{W}{2})}$,

$W = 2\ln\frac{\prod_{i=1}^n X_i}{(X_{(1)}^n} \sim \chi^2(2(n-1))$,

$\lambda(a) = \lambda(b)$, and

$F_W(b) - F_W(a) = 1-\alpha$.

Question: Some students use t.test(x). Can we ?

Are these right "

$\phi = \mathbf{1}(\lambda \le \underbrace{c}_{=0.05})$?

$\phi = \mathbf{1}(\lambda \le c) = \mathbf{1}(W \notin (\chi^2_{2n-2,0.975}, \chi^2_{2n-2,0.025}))$?

$\phi = \mathbf{1}(\lambda \le c) = \mathbf{1}(W > \chi^2_{2n-2,0.05})$?

The R program is given, but you need to provide data as in step (1)

Second approach:

$$\begin{aligned}
E(e^{2Tt}) =& E(\exp(2t\log(\frac{\prod_{i=1}^n X_i}{(X_{(1)})^n})))\\
=& E(\exp(\log([\frac{\prod_{i=1}^n X_i}{(X_{(1)})^n}]^{2t})))\\
=& E([\frac{\prod_{i=1}^n X_i}{(X_{(1)})^n}]^{2t})\\
=& E(E([\frac{\prod_{i=1}^n X_i}{(X_{(1)})^n}]^{2t}|X_{(1)}))\\
=& E(E([\frac{\prod_{i=2}^n X_{(i)}}{(X_{(1)})^{n-1}}]^{2t}|X_{(1)}))\\
=& E(E([\frac{\prod_{i=2}^n X_i}{(X_{(1)})^{n-1}}]^{2t}|X_{(1)})) \qquad \textbf{???}
\end{aligned}$$

Not easy to proceed.

Third approach: It is easy to simplify $2T$ as follows.

$$2T = \sum_{i=1}^n 2\ln X_i - 2n\ln\min_i X_i = \sum_{i=1}^n \big(2\ln X_i - \min_j(2\ln X_j)\big)$$

Define $Y_i = 2\ln X_i$ and $Z_i = Y_i - Y_{(1)}$.

$$2T = \sum_{i=1}^n (Y_i - Y_{(1)}) = \sum_{i=2}^n (Y_{(i)} - Y_{(1)}) = \sum_{i=2}^n Z_{(i)},$$

Since X_1, ..., X_n are i.i.d, so is Y_1, ..., Y_n. Note $Z_{(1)} = 0$. We shall show

$$Z_{(2)}, ..., Z_{(n)} \text{ have the same distribution as } U_{(1)}, ..., U_{(n-1)}, \qquad (1)$$

which are the order statistics of $(n-1)$ i.i.d. r.v.s from Gamma(1,2), and

$$2T = \sum_{i=2}^n Z_{(i)} = \sum_{i=1}^{n-1} U_{(i)} = \sum_{i=1}^{n-1} U_i \sim Gamma(n-1, 2) = \chi^2(2(n-1)).$$

To this end, notice that $Z_{(i)} = Y_{(i)} - Y_{(1)}$, $i = 1, ..., n$.

$$2T = h(Z_{(2)}, ..., Z_{(n)}) \text{ and } (Z_{(2)}, ..., Z_{(n)}) = H(Y_1, ..., Y_n).\ h = \textbf{??}\ \ H = \textbf{??}$$

$f_{\mathbf{Y}} \Rightarrow f_{\mathbf{Z}} \Rightarrow f_{2T}$. We first show that

$$Y_i \text{ has a density function } f_Y(y) = \frac{\theta\nu^\theta}{2} e^{-y\theta/2},\ y > 2\ln(\nu),\ \theta > 0, \tag{2}$$

and

$$f_Y(y) = \frac{\nu}{2} e^{-y/2},\ y > 2\ln(\nu) \text{ under } H_0. \tag{3}$$

To prove Eqs. (2) and (3), set $y = G(x) = 2\ln x$ and $G^{-1}(y) = e^{y/2}$.

$$f_Y(y) = f_X(G^{-1}(y))|\frac{dG^{-1}}{dy}| = \frac{\theta\nu^\theta}{2} e^{-y\theta/2},\ y > 2\ln(\nu) \text{ (which is (2))}.$$

Thus, under H_0, $\theta = 1$ and Eq. (3) holds.

Moreover, recall $Z_{(i)} = Y_{(i)} - Y_{(1)}$. $(Z_{(2)}, ..., Z_{(n)}) = H(Y_1, ..., Y_n)$.

Given $f_{\mathbf{Y}}$, we can find $f_{\mathbf{Z}}$ by two ways:

1. $f_{\mathbf{Z}}(\mathbf{z}) = f_{\mathbf{Y}}(H^{-1}(\mathbf{z}))|Jacobian|$ (not applicable!)
2. $F_{\mathbf{Z}}(\mathbf{z}) = P(H(\mathbf{Y}) \le \mathbf{z})$ and $f_{\mathbf{Z}} = F'_{\mathbf{Z}}$.

In method 2, we can either compute

$$P\{Z_{(2)} \le z_2, \cdots Z_{(n)} \le z_n\} \text{ directly}$$

or

$$P\{z_1 < Z_{(2)} \le z_2 < \cdots \le z_{n-1} < Z_{(n)} \le z_n\}$$

and then identify its distribution. We take the latter approach.
For $z_0 = 0 < z_1 < z_2 < \cdots < z_n$, We first compute an preliminary result.

$$\begin{aligned}
&P\{z_{j-1} < Y_j - y \le z_j, y < Y_j\} \\
=&P\{z_{j-1} < Y_j - y \le z_j, 0 < Y_j - y\} \\
=&P\{z_{j-1} < Y_j - y \le z_j\} && \text{(as } z_1 > 0\text{)} \\
=&\int_{z_{j-1}+y}^{z_j+y} \frac{\nu}{2} e^{-x/2} dx && (by\ (3)) \\
=&(e^{-z_{j-1}/2} - e^{-z_j/2})\nu e^{-y/2} && (4)
\end{aligned}$$

$$\begin{aligned}
&P\{z_1 < Z_{(2)} \le z_2 < \cdots \le z_{n-1} < Z_{(n)} \le z_n\} \\
=&P\{z_1 < Z_{(2)} \le z_2 < \cdots \le z_{n-1} < Z_{(n)} \le z_n, Y_{(1)} \in \{Y_1, ..., Y_n\}\} \\
=&\sum_{i=1}^n P\{z_1 < Z_{(2)} \le z_2 < \cdots \le z_{n-1} < Z_{(n)} \le z_n, Y_{(1)} = Y_i\} \\
& && \text{(mutually exclusive)} \\
=&nP\{z_1 < Z_{(2)} \le z_2 < \cdots \le z_{n-1} < Z_{(n)} \le z_n, Y_{(1)} = Y_1\} && \text{(as } Y_i\text{s are i.i.d.)} \\
=&nP\{z_1 < Z_{(2)} \le z_2 < ...z_{n-1} < Z_{(n)} \le z_n, Y_1 < Y_j,\ j \ge 2\} \\
=&n!P\{z_1 < Z_2 \le z_2, ..., z_{n-1} < Z_n \le z_n, Y_1 < Y_j,\ j \ge 2\} \\
=&n!P\{z_1 < Y_2 - Y_1 \le z_2, ..., z_{n-1} < Y_n - Y_1 \le z_n, Y_1 < Y_j,\ j \ge 2\} \\
=&n!E(\mathbf{1}\{z_1 < Y_2 - Y_1 \le z_2, ..., z_{n-1} < Y_n - Y_1 \le z_n, Y_1 < Y_j,\ j \ge 2\}) \\
=&n!E(E(\mathbf{1}\{z_1 < Y_2 - Y_1 \le z_2, ..., z_{n-1} < Y_n - Y_1 \le z_n, Y_1 < Y_j,\ j \ge 2\}|Y_1))
\end{aligned}$$

$$=n!\int (E(\mathbf{1}\{z_1 < Y_2 - y \le z_2, ..., z_{n-1} < Y_n - y \le z_n, y < Y_j,\ j \ge 2\}|Y_1 = y) f_{Y_1}(y) dy$$

$$=n!\int E(\mathbf{1}\{z_1 < Y_2 - y \le z_2, ..., z_{n-1} < Y_n - y \le z_n, y < Y_j,\ j \ge 2\}) f_{Y_1}(y) dy$$

by i.i.d.

$$=n!\int P\{z_1 < Y_2 - y \le z_2, ..., z_{n-1} < Y_n - y \le z_n, y < Y_j,\ j \ge 2\} f_Y(y) dy$$

$$=n!\int \prod_{j=2}^{n} P\{z_{j-1} < Y_j - y \le z_j, y < Y_j\} f_Y(y) dy \qquad \text{(as } Y_i\text{s are independent)}$$

$$=n!\int \prod_{i=2}^{n} \big((e^{-z_{j-1}/2} - e^{-z_i/2})\nu e^{-y/2}\big) f_Y(y) dy \qquad (by\ (4))$$

$$=n!\prod_{i=2}^{n}(e^{-z_{j-1}/2} - e^{-z_i/2}) \int \nu^{n-1} e^{-(n-1)y/2} f_Y(y) dy$$

$$=n!\prod_{i=2}^{n}(e^{-z_{j-1}/2} - e^{-z_i/2}) \int_{2\ln\nu}^{\infty} \nu^{n-1} e^{-(n-1)y/2} \frac{\nu}{2} e^{-y/2} dy$$

$$=(n-1)!\prod_{i=2}^{n}(e^{-z_{j-1}/2} - e^{-z_i/2}) \int_{2\ln\nu}^{\infty} \nu^{n} e^{-ny/2} d\frac{ny}{2}$$

$$=(n-1)!\prod_{i=2}^{n}(e^{-z_{j-1}/2} - e^{-z_i/2}) \int_{n\ln\nu}^{\infty} \nu^{n} e^{-t} dt \qquad (t = \tfrac{ny}{2})$$

$$=(n-1)!\prod_{i=2}^{n}(e^{-z_{j-1}/2} - e^{-z_i/2}) \int_{n\ln\nu}^{\infty} e^{-t+n\ln\nu} dt \qquad \nu^n = e^{n\ln v}$$

$$=(n-1)!\prod_{i=2}^{n}(e^{-z_{j-1}/2} - e^{-z_i/2})$$

$$=(n-1)!\prod_{i=2}^{n}\int_{z_{j-1}}^{z_i} \frac{1}{2} e^{-x/2} dx$$

Verify that if U_1, ..., U_{n-1} are iid from G(1,2), then

$$P\{U_{(i)} \in (z_i, z_{i+1}], i = 1, ..., n-1\} = (n-1)!\prod_{i=2}^{n}\int_{z_{j-1}}^{z_i} \frac{1}{2} e^{-x/2} dx.$$

In other words, statement (1) holds and thus
$2T = \sum_{i=2}^{n} Z_{(i)} \sim G(\frac{n-1}{2}, 2) = \chi^2(2n-2)$.

Additonal.

1. (a) **Under the assumptions in 8.7**, generate 10 observations from R, and do the tests.

8.7.(a) Find the LRT for testing H_0: $\theta \le 0$ v.s. H_1: $\theta > 0$, based on a sample X_1, ..., X_n from a density $f_X(x) = \frac{1}{\beta}e^{-\frac{x-\theta}{\beta}}$, $x > \theta$ and $\beta > 0$, where β and θ are unknown.

There are two issues:

(1) How to generate 10 observations ?

(2) How to determine RR ?

(1) Are these right ?

x=rnorm(10) # ?

x=rexp(10) ?

(2) RR

8.7.(a) Find the LRT for testing H_0: $\theta \le 0$ v.s. H_1: $\theta > 0$, based on a sample X_1, ..., X_n from a density $f_X(x) = \frac{1}{\beta}e^{-\frac{x-\theta}{\beta}}$, $x > \theta$ and $\beta > 0$, where β and θ are unknown.

Sol. $\mathcal{L}(\mu,\beta)=\beta^{-n}exp(-\sum_i(X_i-\theta)/\beta)\mathbf{1}(\theta\le X_{(1)})$
$=\beta^{-n}exp(-n(\overline{X}-\theta)/\beta)\mathbf{1}(\theta\le X_{(1)})$.

MLE under Θ: For each β, $\mathcal{L}\uparrow$ in θ for $\theta\le X_{(1)}$.

Thus the MLE of θ is always $\hat\theta=X_{(1)}$.
It is easy to check that
the MLE $\hat\beta=\overline{X}-X_{(1)}$ under Θ

MLE under Θ_0:

It turns out that the MLE $\hat\theta^o=\min\{\hat\theta,0\}$ and $\hat\beta^o=\overline{X}-\hat\theta^o$ under Θ_0.

$$\lambda=\begin{cases}1 & \text{if } \overline{X}\le 0,\\ (\frac{\overline{X}-X_{(1)}}{\overline{X}})^n & \text{otherwise.}\end{cases}=\begin{cases}1 & \text{if } \overline{X}\le 0,\\ (1-\frac{X_{(1)}}{\overline{X}})^n & \text{otherwise.}\end{cases} \tag{1}$$

Thus the LRT test is $\mathbf{1}(\lambda\le c)=\mathbf{1}((1-\frac{X_{(1)}}{\overline{X}})^n\le c \text{ and } \overline{X}>0)$.

Are we done ?

No ! We need to know $c=$?

In view of Eq. (1), if $c=1$, $E(\phi)=1>\alpha$. Thus, $c\in[0,1)$ and we can assume $\overline{X}>0$,

$(1-\frac{X_{(1)}}{\overline{X}})^n\le c\Leftrightarrow\frac{X_{(1)}}{\overline{X}}\ge c_1\Leftrightarrow\frac{\overline{X}}{X_{(1)}}\le c_2$

$\Leftrightarrow\frac{\sum_{i=1}^n X_i}{X_{(1)}}\le c_3\Leftrightarrow\frac{\sum_{i=1}^n X_{(i)}}{X_{(1)}}\le c_3$

$\Leftrightarrow\frac{\sum_{i=2}^n X_{(i)}}{X_{(1)}}\le c_4\Leftrightarrow\frac{\sum_{i=2}^n X_{(i)}-(n-1)X_{(1)}}{X_{(1)}}\le c_5$

$\Leftrightarrow\frac{\sum_{i=2}^n T_{(i)}-(n-1)T_{(1)}}{T_{(1)}+\theta}\le c_5$

where

$$T=\frac{X-\theta}{\beta}=h(X), \tag{2}$$

then $h^{-1}(t)=\beta t+\theta$,
$f_T(t)=f_X(\beta t+\theta)|(h^{-1}(t))'|=\frac{1}{\beta}e^{-t}\beta=e^{-t}$, $t>0$.

Notice that $\overline{T}>0$ w.p.1. Let

$$W=\sum_{i=2}^n T_{(i)}-(n-1)T_{(1)}. \tag{3}$$

Thus by Eq. (1), the LRT test is

$\phi=\mathbf{1}_{(\frac{W}{T_{(1)}+\theta}\le b)}$ **(first way)** or

$\phi=\mathbf{1}_{(\{W/b\le T_{(1)}\theta\})}$ **(second way)**.

It can be shown (see #8.5) that under H_o,

(1) $T_{(1)}\sim Gamma(1,1/n)$;
(2) $W\sim Gamma(n-1,1)$;
(3) $W\perp T_{(1)}$.

Thus we need to find $(f_W,f_{T_{(1)}})$ or $(F_W,F_{T_{(1)}})$ assuming $\theta\le 0$ due to H_0. Then find b by

the first way: $\alpha=\int_0^b f_{\frac{W}{T_{(1)}+\theta}}(x)dx=...$

the 2nd way: $\alpha=E(E(\mathbf{1}(W/b\le T_{(1)}+\theta)|W))$.

$$\begin{aligned}\alpha=&\sup_{\theta\le 0}E(E(\mathbf{1}(W/b\le T_{(1)}+\theta)|W))\\ =&\sup_{\theta\le 0}\int_0^\infty e^{-\frac{nw}{b}+\theta}f_W(w)dw\\ =&\int_0^\infty e^{-\frac{nw}{b}}f_W(w)dw\\ =&\int_0^\infty e^{-\frac{nw}{b}}\frac{w^{n-1-1}e^{-w}}{\Gamma(n-1)}dw\end{aligned}$$

$$= \int_0^\infty \frac{w^{n-1-1} e^{-w(1+\frac{n}{b})}}{\Gamma(n-1)} dw$$
$$= (1+\frac{n}{b})^{n-1} \textbf{A mistake. Should be } = (1+\frac{n}{b})^{-(n-1)}$$
$$b = \frac{n}{\alpha^{\frac{1}{n-1}} - 1} \textbf{Should be } b = (\frac{n}{\alpha^{\frac{-1}{n-1}} - 1})^{-1}$$

$\phi = \mathbf{1}(\lambda \le c) = \mathbf{1}(\frac{\sum_{i=2}^n X_{(i)} - (n-1)X_{(1)}}{X_{(1)}} \le \frac{n}{\alpha^{\frac{1}{n-1}}-1})$,

Wrong, should be $\mathbf{1}(\frac{\sum_{i=2}^n X_{(i)} - (n-1)X_{(1)}}{X_{(1)}} \le \frac{n}{\alpha^{\frac{-1}{n-1}}-1})$**, or**

$\phi = \mathbf{1}(\frac{\sum_{i=2}^n X_{(i)}}{X_{(1)}} \le \frac{n}{\alpha^{\frac{-1}{n-1}}-1} + n - 1)$

$\vdash$: (1) $T_{(1)} \sim Gamma(1, 1/n)$.

Notice: Let $T = X/\beta = h(X)$, then $h^{-1}(t) = \beta t$,

$f_T(t) = f_X(\beta t)|(h^{-1}(t))'| = \frac{1}{\beta} e^{-t} \beta = e^{-t}$, $t > 0$.

For $t > 0$, $P(T_{(1)} > t) = (P(T > t))^n = (e^{-t})^n$, thus ...

$P(Gamma(n, 1)$ r.v. $> t) = P(Poiss(t)$ r.v. $\le n-1)$.

$\vdash$: (2) $W \sim Gamma(n-1, 1)$ and (3) $W \perp T_{(1)}$.

Proof. Let $Y_1 = T_{(1)}$, $Y_j = T_{(j)} - T_{(1)}$, $j = 2, ..., n$. ($\mathbf{Y} = g(\mathbf{T})$) **?**

$T_{(1)} = Y_1$, $T_{(j)} = Y_j + Y_1$, $j = 2, ..., n$. ($\mathbf{Y} = g^{-1}(\mathbf{T})$) **?**

$f_{\mathbf{T}}(\mathbf{t}) = n! e^{-\sum_{i=1}^n t_i}$, $0 < t_1 < \cdots < t_n$.

$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{T}}(g^{-1}(\mathbf{y}))|J|$=......

$= n! \exp(-(ny_1 + y_2 + \cdots + y_n))$, y_1, $0 < y_2 < \cdots < y_n$.

Thus $Y_1 \perp (Y_2, ..., Y_n)$ and $(Y_2, ..., Y_n)$ are order statistics of n-1 Exp(1) r.v.'s.

$W = \sum_{i=2}^n Y_i \sim Gamma(n-1, 1)$.

Another proof. Let $U = \sum_{i=2}^n T_{(i)} = \sum_{i=2}^n X_{(i)}/\beta$ and $V = T_{(1)} = X_{(1)}/\beta$ then $W = V/U$. Thus WLOG, we can assume that $\beta = 1$ and $\theta = 0$.

Notice $f_{X_1} \Rightarrow f_{\mathbf{X}} = \prod_{i=1}^n f_{X_i}$.

$\Rightarrow f_{X_{(1)},...,X_{(n)}}(x_1, ..., x_n) = n! f_{\mathbf{X}}(x_1, ..., x_n)$, where $x_1 < \cdots < x_n$.

$\Rightarrow f_{U,V} \Rightarrow f_W$.

(1) $f_{U,V} = ?$ (2) $f_W = ?$

Typically two ways: (1) cdf (2) Jacobian.

We use cdf approach for $f_{U,V}$ as follows. $f_{U,V} = \frac{\partial^2 F_{U,V}}{\partial u \partial v}$ or $f_{U,V}(u, v) = -\frac{\partial^2 P(U \le u, V > v)}{\partial u \partial v}$. We use the latter one.

Typically, let $n = 3$.

Note $P(U \le u, V > v) = 0$ if $u < v(n-1)$. Thus we can assume $u > v(n-1)$.

$$\begin{aligned}
&P\{U \le u, V > v\} \\
=&P\{X_{(2)} + \cdots + X_{(n)} \le u, X_{(1)} > v\} \\
=&n! P\{X_2 + \cdots + X_n \le u, X_n > \cdots > X_2 > X_1 > v\} \\
=&n! P\{v < X_1 < X_2,\ X_2 < X_3 \wedge (u - X_3 - \cdots - X_n), \\
&\cdots, X_{n-1} < X_n \wedge (u - X_n), X_n \le u\} \\
=&n! P\{v < X_1 < X_2,\ X_2 < X_3 \wedge (u - X_3),\ X_3 \le u\} \\
=&n! P\{v < X_1 < X_2,\ X_2 < X_3,\ X_3 < u/2, X_3 \le u\} \\
&+ n! P\{v < X_1 < X_2,\ X_2 < u - X_3,\ X_3 \ge u/2, X_3 \le u\} \\
=&n! P\{v < X_1 < X_2 < X_3 < u/2\} \\
&+ n! P\{v < X_1 < X_2 < u - X_3,\ u/2 \le X_3 \le u\}
\end{aligned}$$

Should we further simplify ??

$$
\begin{aligned}
&= n!P\{v < X_1 < X_2 < X_3 < u/2\} \\
&\quad + n!P\{v < X_1 < X_2 < u - X_3,\ v < u - X_3 < u/2,\ u/2 \le X_3 \le u\} \\
&= n!P\{v < X_1 < X_2 < X_3 < u/2\} \\
&\quad + n!P\{v < X_1 < X_2 < u - X_3,\ X_3 < u - v,\ u/2 \le X_3 \le u\} \\
&= n!P\{v < X_1 < X_2 < X_3 < u/2\} \\
&\quad + n!P\{v < X_1 < X_2 < u - X_3,\ u/2 \le X_3 < u - v\} \\
&= n!(\int_v^{u/2}\int_v^{x_3} + \int_{u/2}^{u-v}\int_v^{u-x_3})\int_v^{x_2} e^{-x_1-x_2-x_3}dx_1dx_2dx_3 \\
&= n!(\int_v^{u/2}\int_v^{x_3} + \int_{u/2}^{u-v}\int_v^{u-x_3})[e^{-v-x_2-x_3} - e^{-2x_2-x_3}]dx_2dx_3
\end{aligned}
$$

$$
\begin{aligned}
&= 6(\int_v^{u/2}\int_v^{x_3} + \int_{u/2}^{u-v}\int_v^{u-x_3})[e^{-v-x_2-x_3} - e^{-2x_2-x_3}]dx_2dx_3 \\
&= 6\int_v^{u/2}\int_v^{x_3} e^{-v-x_2-x_3}dx_2dx_3 - 6\int_v^{u/2}\int_v^{x_3} e^{-2x_2-x_3}dx_2dx_3 \\
&\quad + 6\int_{u/2}^{u-v}\int_v^{u-x_3} e^{-v-x_2-x_3}dx_2dx_3 - 6\int_{u/2}^{u-v}\int_v^{u-x_3} e^{-2x_2-x_3}dx_2dx_3, \\
&\quad 0 < v < \frac{u}{n-1}.
\end{aligned}
$$

$$
\begin{aligned}
&f_{U,V}(u,v) = -6\frac{\partial^2}{\partial u \partial v}P\{U \le u, V > v\} \\
&= -6\frac{\partial}{\partial u}\{[-\int_v^{v} e^{-v-x_2-v}dx_2 - \int_v^{u/2} e^{-v-v-x_3}dx_3 \\
&\quad - \int_v^{u/2}\int_v^{x_3} e^{-v-x_2-x_3}dx_2dx_3] + [\int_v^{v} e^{-2x_2-v}dx_2 \\
&\quad + \int_v^{u/2} e^{-2v-x_3}dx_3] + [-\int_v^{v} e^{-v-x_2-(u-v)}dx_2 \\
&\quad - \int_{u/2}^{u-v} e^{-v-v-x_3}dx_3 - \int_{u/2}^{u-v}\int_v^{u-x_3} e^{-v-x_2-x_3}dx_2dx_3] \\
&\quad + [\int_v^{v} e^{-2x_2-(u-v)}dx_2 + \int_{u/2}^{u-v} e^{-2v-x_3}dx_3]\} \\
&= -6\frac{\partial}{\partial u}\{-\int_v^{u/2}\int_v^{x_3} e^{-v-x_2-x_3}dx_2dx_3 - \int_{u/2}^{u-v}\int_v^{u-x_3} e^{-v-x_2-x_3}dx_2dx_3\} \\
&= -6\{-(1/2)\int_v^{u/2} e^{-v-x_2-u/2}dx_2 - \int_v^{u-(u-v)} e^{-v-x_2-(u-v)}dx_2 \\
&\quad - (-1/2)\int_v^{u-u/2} e^{-v-x_2-u/2}dx_2\} - \int_{u/2}^{u-v} e^{-v-(u-x_3)-x_3}dx_3\} \\
&= 6\{(1/2)\int_v^{u/2} e^{-v-x_2-u/2}dx_2 - (1/2)\int_v^{u-u/2} e^{-v-x_2-u/2}dx_2 \\
&\quad + \int_{u/2}^{u-v} e^{-v-u}dx_3\} \\
&= 6\int_{u/2}^{u-v} e^{-v-u}dx_3 \\
&= 6(\frac{u}{2} - v)e^{-v-u},\ 0 < v < \frac{u}{2}.
\end{aligned}
$$

After finding $f_{U,V}$, we can find f_W again by two ways:

1. Let $(u,w)=g(u,v)$, then $f_{U,W}(u,w)=f_{U,V}(g^{-1}(u,w))|(\frac{\partial g^{-1}}{\partial(u,w)})_{2\times 2}| \Rightarrow f_W$.
2. $F_W(w)=\int\int_{\frac{u}{v}\le w} f_{U,V}(u,v)dudv$. $f_W(w)=F_W'(w)$.

   Now $(u,w)=(u,v/u)=g(u,v)$, thus $(u,v)=(u,uw)=g^{-1}(u,w)$.

$$\frac{\partial}{\partial(u,w)}g^{-1}=\begin{pmatrix}1 & w\\ 0 & u\end{pmatrix}$$

$$f_{U,W}(u,w)=6u(\frac{u}{2}-uw)e^{-uw-u},\ 0<w<1/2\ \&\ u>0.$$

$$\begin{aligned}f_W(w)=&\int_0^\infty 6u^2(\frac{1}{2}-w)e^{-u(w+1)}du\\ =&6(\frac{1}{2}-w)\int_0^\infty u^{3-1}e^{-u(w+1)}du\\ =&6(\frac{1}{2}-w)(2)(\frac{1}{w+1})^3,\ w\in(0,1/2).\end{aligned}$$

Summary: H_0: $\theta\le 0$, vs H_1: $\theta>0$

$\alpha=0.05$;

$\phi=\mathbf{1}_{(W\ge c)}$, where

$$0.05=\int_c^{\frac{1}{(n-1)}} f_W(w)dw.$$

If we set $n=4$, the derivation is more tedious.

$$\begin{aligned}&P\{U\le u, V>v\}\\ =&P\{X_{(2)}+\cdots+X_{(n)}\le u, X_{(1)}>v\}\\ =&n!P\{X_2+\cdots+X_n\le u, X_n>\cdots>X_2>X_1>v\}\\ =&n!P\{v<X_1<X_2,\ X_2<X_3\wedge(u-X_3-\cdots-X_n),\\ &\cdots, X_{n-1}<X_n\wedge(u-X_n), X_n\le u\}\\ =&n!P\{v<X_1<X_2,\ X_2<X_3\wedge(u-X_3-X_4),\ X_3<X_4\wedge(u-X_4), X_4\le u\}\\ =&n!P\{v<X_1<X_2,\ X_2<X_3,\ X_3<(u-X_4)/2, X_3<X_4,\ X_4\le u/2, X_4\le u\}\\ &+n!P\{v<X_1<X_2,\ X_2<u-X_3-X_4,\ X_3\ge(u-X_4)/2,\ X_3<X_4,\ X_4\le u/2, X_4\le u\}\\ &+n!P\{v<X_1<X_2,\ X_2<X_3,\ X_3<(u-X_4)/2, X_3<u-X_4,\ X_4>u/2, X_4\le u\}\\ &+n!P\{v<X_1<X_2,\ X_2<u-X_3-X_4,\ X_3\ge(u-X_4)/2,\ X_3<u-X_4,\ X_4>u/2,\ X_4\le u\}\\ =&n!P\{v<X_1<X_2,\ X_2<X_3,\ X_3<(u-X_4)/2, X_3<X_4\le u/2\}\\ &+n!P\{v<X_1<X_2,\ X_2<u-X_3-X_4,\ X_3\ge(u-X_4)/2,\ X_3<X_4\le u/2\}\\ &+n!P\{v<X_1<X_2,\ X_2<X_3,\ X_3<(u-X_4)/2,\ u/2<X_4\le u\}\\ &+n!P\{v<X_1<X_2,\ X_2<u-X_3-X_4,\ (u-X_4)/2\le X_3<u-X_4,\ u/2<X_4\le u\}\end{aligned}$$

Homework solutions for week 8

8.24. Is the LRT for the simple hypotheses equivalent to the one obtained by NPL ?

Sol. Let H_o: $\theta=\theta_o$ v.s. H_1: $\theta=\theta_1$.

The LRT test is $\phi=\mathbf{1}(\lambda\le c)$, where $\lambda=\lambda(x)=\frac{f(x;\theta_o)}{f(x;\theta_o)\vee f(x;\theta_1)}$.

Let $\alpha_o=E_{\theta_o}(\phi)$.

The MP test from NPL is

$$\phi_1=\begin{cases}1 & \text{if } f(x;\theta_1)>kf(x;\theta_o)\\ p(x) & \text{if } f(x;\theta_1)=kf(x;\theta_o)\\ 0 & \text{if } f(x;\theta_1)<kf(x;\theta_o)\end{cases} \quad (1)$$

where $k \geq 0$ and $E_{\theta_o}(\phi_1(X)) = \alpha$.

$$\phi_1 = \mathbf{1}(\frac{f(x;\theta_1)}{f(x;\theta_o)} > k) + p(x)\mathbf{1}(\frac{f(x;\theta_1)}{f(x;\theta_o)} = k) \text{ where } \frac{0}{0} = 1.$$

Two understandings of the question:

(a) Given ϕ, does $\phi = \phi_1$ for a ϕ_1 of form Eq. (1) ?

(b) Given ϕ_1, does $\phi_1 = \phi$ for a LRT ϕ ?

(a) Yes iff $E_{\theta_o}(\phi) = E_{\theta_o}(\phi_1)$. Need a proof.

(b) $\begin{cases} Yes & \text{if } P(f(\mathbf{X},\theta_1) = kf(\mathbf{X},\theta_0)) = 0 \\ \text{No in general} & \text{otherwise} \end{cases}$ Need a proof.

Proof in Case (a). 3 cases of "if": (1) $c = 0$, (2) $c \in (0, 1)$, (3) $c = 1$.
In case (a.1), $\lambda(x) \leq c = 0$ implies that
$f(x;\theta_o) = 0 < f(x;\theta_1)$ (*i.e.* $1/\lambda(x) = \infty$),
otherwise, x is not of concern **why ?**
Define $k = \infty = 1/0$ and $\infty \times 0 = 0$, then
$\phi = \mathbf{1}(\lambda = 0) = \phi_1$ with $k = \infty$ and $p(x) = 0$ and $\alpha_o = 0 = E_{\theta_o}(\phi_1)$.
In case (a.2), $\phi = \phi_1$ as in Eq. (1) with $k = 1/c > 1$ and $p(x) = 1$. It is a MP level α_o test.
In case (a.3), $\phi = \phi_1$ as in Eq. (1) with $k = 0$ ($\neq 1 = 1/c$) and $p(x) = 1$. $\alpha_o = 1$. It is a MP level α_o test.
Not of practical interest !
Proof in Case (b). The answer in general is "No". A counterexample for discrete case: $X \sim bin(1, p)$. H_0: $p = 0.5$ v.s. H_1: $p = 0.4$. $\phi_1 = \frac{1}{10}\mathbf{1}(X = 0)$ satisfies Eq. (1), but it is not a LRT ϕ.

#8.22. Let X_1 ,..., X_{10} be i.i.d. from bin(1,p). For testing H_0: $p = 1/2$ vs. H_1: $p = 1/4$. A UMP size 0.0547 test is $\phi = \mathbf{1}_{(\sum_i X_i \leq 2)}$.
Question c: For what α level does there exist a UMP test of H_0: $p = 1/2$ vs. H_1: $p = 1/4$?
Answer: Two points of view:

(1) Theoretical point of view: $\alpha \leq 1$.

(2) Pratical point of view: $\alpha < 1/2$.

Since $T = \sum_{i=1}^{10} X_i$ is the sufficient statistic and has $\uparrow$ MLR, the UMP test of level α is

$$\phi = \begin{cases} 1 & \text{if } T < i \\ \gamma & \text{if } T = i \\ 0 & \text{if } T > i, \end{cases} \tag{1}$$

where

$$\alpha = [\sum_{j<i} \binom{10}{j} + \gamma\binom{10}{i}]0.5^{10}, \quad i \in \{0, 1, 2, 3, 4, 5\}.$$

$\alpha \in [0, 0.5)$.

For instance, for $\alpha = 0.05$, take $i = 2$ and

$$\gamma = [\alpha - \sum_{j=0}^{1} \binom{10}{j} 0.5^{10}]/[\binom{10}{2}0.5^{10}] = 0.8933333$$

$$\phi = \begin{cases} 1 & \text{if } T < 2 \\ 0.8933333 & \text{if } T = 2 \\ 0 & \text{if } T > 2, \end{cases}$$

R commands:

```
x=0:5
round(pbinom(x,10,0.5),4)
[1] 0.0010 0.0107 0.0547 0.1719 0.3770 0.6230
(0.05-pbinom(1,10,0.5))/dbinom(2,10,0.5)
[1] 0.8933333
```

Question related to 8.29, where $X \sim Cauchy(\theta)$. If $X \in bin(10, \theta)$, let $\phi_i = \mathbf{1}(X \in RR_i)$, where $RR_1 = \{X \in [0, 2]$, $RR_2 = [0, 1]$ and $RR_3 = [0, 1.5]$.

1. $\phi_1 \neq \phi_2$ as $RR_1 \neq RR_2$. Yes, No, DNK.
2. $\phi_3 \neq \phi_2$ as $RR_3 \neq RR_2$. Yes, No, DNK.
3. $\phi_3 = \phi_2$ as $P((RR_3 \setminus RR_2) \cup (RR_2 \setminus RR_3) = 0$. Yes, No, DNK.
4. $\phi_1 \neq \phi_2$ as $P((RR_1 \setminus RR_2) \cup (RR_2 \setminus RR_1) > 0$. Yes, No, DNK.

8.29. Let $X \sim f(x; \theta) = \frac{1}{\pi(1+(x-\theta)^2}$. $\phi = \mathbf{1}(1 < X < 3)$.

(b) $\vdash$: ϕ is UMP level α test for H_o: $\theta = 0$ v.s. H_1: $\theta = 1$.

(c) $\vdash$: Prove of disprove: ϕ is the UMP level $E_{\theta_o}(\phi)$ test for testing H_o: $\theta \leq 0$ v.s. H_1: $\theta > 0$.

(d) What can be said about in general.

Proof. **(b)** Let $g(x) = f(x; 1)/f(x; 0) = \frac{1+x^2}{1+(x-1)^2}$. Since $g(1) = g(3) = 2$, by the NPL, a UMP level α test is $\phi_o = \mathbf{1}(g(X) > 2)$ with $\alpha = E_{\theta=0}(\phi_o)$.

$g(x) > 2 <=> \frac{1+x^2}{1+(x-1)^2} > 2 <=> 1 + x^2 - 2(1 + (x-1)^2) > 0 <=> 1 < x < 3$.

By the NPL, $\phi = \phi_o$ is the UMP test of level α.

(d). In genenal, we can show that a test of the form $\phi = \mathbf{1}(X \in (u, v))$ is a UMP level α test for testing H_o: $\theta = \theta_o$, v.s. H_1: $\theta = \theta_1 \neq \theta_o$, where (u, v) depends on (θ_o, θ_1) and α.

It suffices to show that $\phi = \mathbf{1}(X \in (u, v)) = \mathbf{1}(\frac{f(X;\theta_1)}{f(X;\theta_o)} > k)$ for some $k > 1$.

In fact, since $\alpha = E_{\theta_o}(\phi(X)) \downarrow 0$ as $k \uparrow \infty$,

$\frac{f(x;\theta_1)}{f(x;\theta_o)} = \frac{1+(x-\theta_o)^2}{1+(x-\theta_1)^2} > k \ (> 1)$.

$<=> 1 + (x - \theta_o)^2 - (1 + (x - \theta_1)^2)k > 0$

$<=> ax^2 + bx + c \equiv x^2(1 - k) - 2x(\theta_o - k\theta_1) + \theta_o^2 - k\theta_1^2 + 1 - k > 0$

$<=> \quad u < x < v$

where u and v are $\frac{-b \pm \sqrt{b^2-4ac}}{2a}$.

(c) $\vdash$: ϕ is not UMP level α test for H_o: $\theta \leq 0$ v.s. H_1: $\theta > 0$.

Proof 1. Take $\theta = 2$ and let

$$G_2(x) = f(x; 2)/f(x; 0) = \frac{1 + x^2}{1 + (x-2)^2} \qquad (1)$$

$G_2(1) = \frac{1+1}{1+1^2} \neq \frac{1+3^2}{1+1^2} = G_2(3)$

Thus ϕ is not UMP level α test.

Is it correct ?

Example: Suppose that $\Theta = \{0, 1\}$. $\alpha = 0.2$.

$f(x; \theta)$ is given by the table: $\begin{pmatrix} x: & 1 & 2 & 3 \\ f(x;1): & 0.3 & 0.4 & 0.3 \\ f(x;0): & 0.5 & 0.2 & 0.3 \end{pmatrix}$

Prove or disprove that $\phi = \mathbf{1}(1 < X < 3)$ is the UMP level 0.2 test.

Disroof. Let $G(x) = f(x; 1)/f(x; 0)$.

Then $G(1) = 3/5 \neq 1 = G(3)$. $\alpha = E_{\theta=0}(\phi)$.

Thus ϕ is not a UMP level α test.

Proof. ϕ is the UMP level 0.2 test by the NPL, as

$E_{\theta=0}(\phi) = 0.2$ and

$$\phi = \mathbf{1}(1 < X < 3) = \mathbf{1}(X = 2) = \begin{cases} 1 & \text{if } f(X;1) > 1.1f(X;0) \\ 0 & \text{if } f(X;1) < 1.1f(X;0). \end{cases}$$

What does the example mean ?

8.29 (c) (continued). How to correct proof 1?

Ans: Find the UMP level α test ϕ_2 with $\theta_1 = 2$ and show either $P(\phi(X) = \phi_2(X)) < 1$ or $\beta_\phi(2) < \beta_{\phi_2}(2)$.

In fact, by the conclusion in part (d), the UMP level α test is

$\phi_2 = \mathbf{1}(X \in (u, v))$ and $G(u) = G(v)$.

Recall $\phi = \mathbf{1}(X \in (1, 3))$.

Since $G_2(1) \neq G_2(3)$,

$(u,v) \neq (1,3)$ and $P(X \in A) > 0$ for each nonempty open interval A,

$P(\phi(X) \neq \phi_2(X)) > 0$.

Additional Homework solutions for week 8

1. The Weibull random variable has a pdf $f(x;\theta) = \theta x^{\theta-1} e^{-x^\theta}$, $x, \theta > 0$.
   (1) Find a MP test of size $\alpha = 0.1$ for testing
   H_0: $\theta = 1$ versus H_1: $\theta = 2$.
   (2) Compute the Type II Error probability.
   (3) If $X = 1.2$ is observed, what is your conclusion ?

Sol. H_0: $\theta = 1$, v.s. H_1: $\theta = 2$. $\alpha = 0.1$.

A MP test is $\phi = \mathbf{1}_{(r \geq k)}$, where $E(\phi(X)) = 0.1$ and

$r = \frac{f(x;\theta_1)}{f(x;\theta_0)} = 2xe^{-x^2+x} \begin{cases} \uparrow & \text{if } x \in (0,1) \\ \downarrow & \text{if } x \geq 1 \end{cases}$.

Thus $\phi = \mathbf{1}_{(X \in (a,b))}$, (or $\mathbf{1}_{(X \in [a,b])}$, or $\mathbf{1}_{(X \in (a,b])}$, or $\mathbf{1}_{(X \in [a,b))}$), and

$\int_a^b e^{-x} dx = 0.9$, *i.e.*,

$$e^{-a} - e^{-b} = 0.9. \tag{1}$$

$$2ae^{-a^2+a} - 2be^{-b^2+b} = 0. \tag{2}$$

Substituting $b = g(a)$ to Eq. (2) yielding

$$w(a) = 2ae^{-a^2+a} - 2g(a)e^{-(g(a))^2+g(a)} = 0. \tag{3}$$

```
x=(1:200)/100
w=w(x)
plot(x,w)
abline(h=0)
```

It turns out $(a,b) = (0.87, 1.13)$. Thus do not reject H_0, as $X = 1.2$.

Additonal.

1. The Weibull random variable has a pdf $f(x;\theta) = \theta x^{\theta-1} \exp(-x^\theta)$, $x, \theta > 0$.
   (1) Find a MP test of size $\alpha = 0.1$ for testing H_0: $\theta = 1$ versus H_1: $\theta = 2$.
   (2) Compute the Type II Error probability.
   (3) If $X = 1.2$ is observed, what is your conclusion ?

Sol. The LR

$$g(x) = \frac{f(x;2)}{f(x;1)} = 2xe^{x-x^2}$$

$$g' = 2(2x+1)(-x+1)$$

$$\begin{array}{lccccc} x: & 0 & & 1 & & 2 \\ g: & & \nearrow & max & \searrow & \\ & \nearrow & & & & \searrow \end{array}$$

$g(x) > k <=> x \in (a,b)$.

Thus the MP test of level α is $\phi = \mathbf{1}_{(X \in (a,b))}$, where

$$0.1 = \alpha = \int_a^b e^{-x} dx = e^{-a} - e^{-b}$$

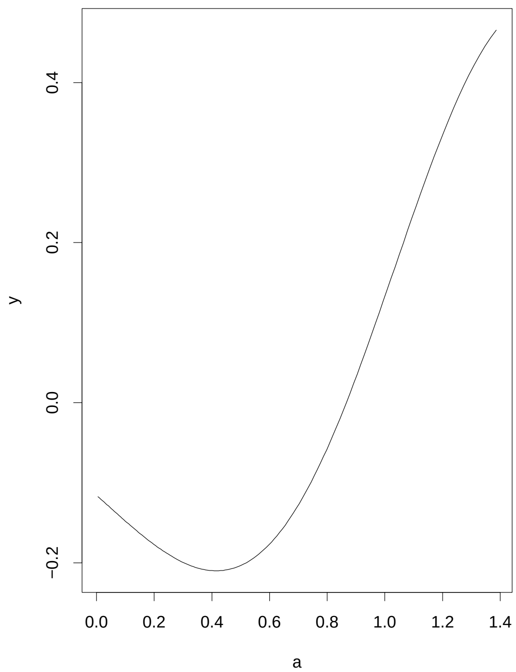
$$or\ 0.1 = pexp(b) - pexp(a).$$

$$g(a) = g(b) \quad <=> \quad ae^{a-a^2} = be^{b-b^2}$$

Solve the equations using R:

$$h(a) = ae^{a-a^2} - be^{b-b^2} = 0, \text{ where } b = -\ln(e^{-a} - 0.1)$$

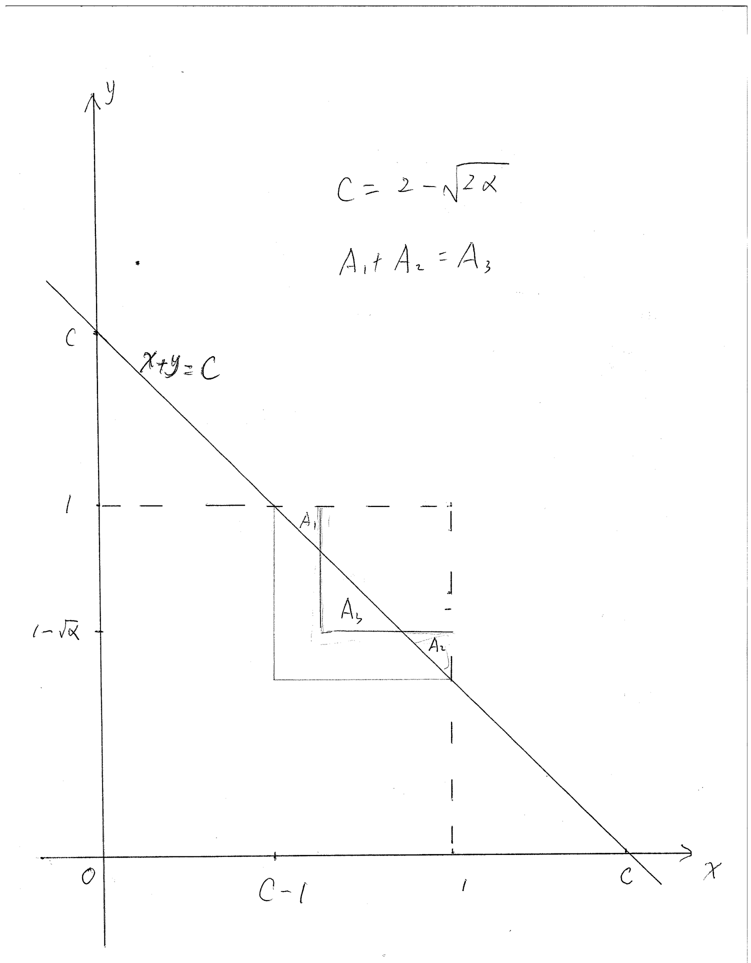
```
x=(1:150)/200
a=qexp(x)
b=-log(exp(-a)-0.1)
# or b= qexp(0.1+pexp(a))
y=a*exp(a-a*a)-b*exp(b-b*b)
plot(a,y,type="l")
max(a[y<=0])
max(b[y<=0])
[1] 0.8556661 [1] 1.12393
```



8.13. Let X_1, X_2 be i.i.d. $\sim U(\theta, \theta+1)$, $\phi_2 = \mathbf{1}(X_1 + X_2 > C)$, find a test which is more powerful than ϕ_2.
Sol. Note $C = 2 - \sqrt{2\alpha}$.
How to get a better test ?
By NPL, for testing H_o: $\theta = \theta_o = 0$ v.s. H_1: $\theta = \theta_1 = 1 - \sqrt{\alpha}$, the UMP test of size α is

$$\phi = \begin{cases} 1 & \text{if } f(x,y;\theta_1) = 1 > 0 = f(x,y;\theta_o) \ i.e., \ x \text{ or } y \geq 1 \\ 0 & \text{if } f(x,y;\theta_1) = 0 < 1 = f(x,y;\theta_o) \ i.e., \ x, \ y < 1 - \sqrt{\alpha} \end{cases}$$

with $E_{\theta_0}(\phi) = \alpha$. If $\alpha = 0.05$, $C - 1 = 0.68$ and $1 - \sqrt{\alpha} = 0.78$.



Candidate $\phi = \mathbf{1}(\mathbf{X} \in RR)$, where

$RR = \{X_1 \geq 1 \text{ or } X_2 \geq 1\} \cup \{X_1, X_2 \geq 1 - \sqrt{\alpha}\}$ **Why ?**

Are we done ?

No, we need a proof.

Consider 3 cases:

(1) $\theta \in [0, C-1]$;

(2) $\theta \in (C-1, 1-\sqrt{\alpha})$;

(3) $\theta \in [1-\sqrt{\alpha}, \infty)$.

Since $\phi \geq \phi_2$ on the region $\{X_1 \geq 1 \text{ or } X_2 \geq 1\}$, it is easy to check that $\beta_\phi(\theta) \geq \beta_{\phi_2}(\theta)\ \forall\ \theta$ in cases (1) and (3).

In case (2) $P(\phi = 1, X_1, X_2 < 1) = \alpha$, but $P(\phi_2 = 1, X_1, X_2 < 1) < \alpha$, thus $\beta_\phi(\theta) \geq \beta_{\phi_2}(\theta)\ \forall\ \theta$ in case (2).

For testing H_0: $\theta \in \Theta_0$ v.s. H_1: $\theta \in \Theta_0^c$,

LRT $\phi = \mathbf{1}(\lambda \leq c)$, where

$$\lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta|\mathbf{x})}{\sup_{\theta \in \Theta} L(\theta|\mathbf{x})} = \frac{L(\hat{\theta}_0|\mathbf{x})}{L(\hat{\theta}|\mathbf{x})}$$

$L(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i;\theta)$,

$\hat{\theta}$ is the MLE of θ under Θ,

$\hat{\theta}_0$ is the MLE of θ under Θ_0,

c is determined by $\alpha = \sup_{\theta \in \Theta_0} P(\lambda \leq c)$, or otherwise,

$$c = \sup\{t: \ \alpha \geq \sup_{\theta \in \Theta_0} P(\lambda \leq t)\}.$$

8.35.(c) Show that the pdf of the noncentral t dostribution T' has an MLR in its noncentrality parameter $\delta = |\mu|$.
Sol. $T = X/\sqrt{Y/\nu}$, $X \sim N(\mu, 1)$ and $Y \sim \chi^2(\nu)$. Let $|\mu_2 > |\mu_1|$

$$F_T(t, \delta) = \int_0^\infty \int_0^{t\sqrt{y/\nu}} \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2} dx \frac{y^{\nu/2-1}e^{-y/2}}{\Gamma(\nu/2)2^{\nu/2}} dy$$

$$f_T(t, \delta) = \int_0^\infty \sqrt{y/\nu} \frac{1}{\sqrt{2\pi}} e^{-(t\sqrt{y/\nu}-\mu)^2/2} \frac{y^{\nu/2-1}e^{-y/2}}{\Gamma(\nu/2)2^{\nu/2}} dy$$

$$\frac{f_T(t,\delta_2)}{f_T(t,\delta_1)} = \frac{\int_0^\infty \sqrt{y/\nu} \frac{1}{\sqrt{2\pi}} e^{-(t\sqrt{y/\nu}-\mu_2)^2/2} \frac{y^{\nu/2-1}e^{-y/2}}{\Gamma(\nu/2)2^{\nu/2}} dy}{\int_0^\infty \sqrt{y/\nu} \frac{1}{\sqrt{2\pi}} e^{-(t\sqrt{y/\nu}-\mu_1)^2/2} \frac{y^{\nu/2-1}e^{-y/2}}{\Gamma(\nu/2)2^{\nu/2}} dy}$$

$(\frac{f_T(t,\delta_2)}{f_T(t,\delta_1)})'_t$
$= \frac{1}{c^2}[\int_0^\infty y/\nu \frac{-(t\sqrt{y/\nu}-\mu_2)}{\sqrt{2\pi}} e^{-(t\sqrt{y/\nu}-\mu_2)^2/2} \frac{y^{\nu/2-1}e^{-y/2}}{\Gamma(\nu/2)2^{\nu/2}} dy$
$\int_0^\infty \sqrt{y/\nu} \frac{1}{\sqrt{2\pi}} e^{-(t\sqrt{y/\nu}-\mu_1)^2/2} \frac{y^{\nu/2-1}e^{-y/2}}{\Gamma(\nu/2)2^{\nu/2}} dy$
$- \int_0^\infty \sqrt{y/\nu} \frac{1}{\sqrt{2\pi}} e^{-(t\sqrt{y/\nu}-\mu_2)^2/2} \frac{y^{\nu/2-1}e^{-y/2}}{\Gamma(\nu/2)2^{\nu/2}} dy$
$\int_0^\infty y/\nu \frac{-(t\sqrt{y/\nu}-\mu_1)}{\sqrt{2\pi}} e^{-(t\sqrt{y/\nu}-\mu_1)^2/2} \frac{y^{\nu/2-1}e^{-y/2}}{\Gamma(\nu/2)2^{\nu/2}} dy] > 0$

$$\frac{\int_0^\infty y/\nu \frac{-(t\sqrt{y/\nu}-\mu_2)}{\sqrt{2\pi}} e^{-(t\sqrt{y/\nu}-\mu_2)^2/2} \frac{y^{\nu/2-1}e^{-y/2}}{\Gamma(\nu/2)2^{\nu/2}} dy}{\int_0^\infty \sqrt{y/\nu} \frac{1}{\sqrt{2\pi}} e^{-(t\sqrt{y/\nu}-\mu_2)^2/2} \frac{y^{\nu/2-1}e^{-y/2}}{\Gamma(\nu/2)2^{\nu/2}} dy}$$

$$> \frac{\int_0^\infty y/\nu \frac{-(t\sqrt{y/\nu}-\mu_1)}{\sqrt{2\pi}} e^{-(t\sqrt{y/\nu}-\mu_1)^2/2} \frac{y^{\nu/2-1}e^{-y/2}}{\Gamma(\nu/2)2^{\nu/2}} dy}{\int_0^\infty \sqrt{y/\nu} \frac{1}{\sqrt{2\pi}} e^{-(t\sqrt{y/\nu}-\mu_1)^2/2} \frac{y^{\nu/2-1}e^{-y/2}}{\Gamma(\nu/2)2^{\nu/2}} dy}$$

$$g(t) = \text{The ratio of the intergrants} = \frac{e^{-(t\sqrt{y/\nu}-\mu_2)^2/2}}{e^{-(t\sqrt{y/\nu}-\mu_1)^2/2}} = e^{\frac{1}{2}(2t\sqrt{y/\nu}-\mu_2-\mu_1)(\mu_2-\mu_1)}$$

If $\mu_2 > 0$, then $\mu_2 - \mu_1 > 0$, $g(t) \uparrow$ in t, and $\frac{f_T(t,\delta_2)}{f_T(t,\delta_1)} \uparrow$ in t too.
If $\mu_2 < 0$, then $\mu_2 - \mu_1 < 0$, $g(t) \uparrow$ in t, and $\frac{f_T(t,\delta_2)}{f_T(t,\delta_1)} \uparrow$ in t too.

Homework solutions for week 10

Additional problem.

3. Carry out the following simulation project.
3.1. Use Splus to generate 10 observations from $N(1, 4)$.
3.2. Now pretend that you only known that the data were from $N(\mu, \sigma)$ without knowing μ and construct a 80% confidence interval for μ.
3.3. Repeat Steps 3.1 and 3.2 100 times.
Splus commands are :

```
y=rep(0,200)
dim(y)=c(100,2)
z=qnorm(0.9)
for(i in 1:100) {
x=rnorm(10,1,2)
u=mean(x)
y[i,1]=u-z*2/sqrt(10)
```

```
y[i,2]=u+z*2/sqrt(10)
}
plot(1:100,y[1:100,2], xlim=c(0,100), ylim=c(-2,5))
par(new=T)
plot(1:100,y[1:100,1], xlim=c(0,100), ylim=c(-2,5))
for(i in 1:100) {
par(new=T)
plot(c(i,i),y[i,1:2],type="l", xlim=c(0,100), ylim=c(-2,5))
}
```

a. In the above approach, what assumption is made on σ ?
b. What is the proportion that the CIs contain the true mean?
c. What is the relation between the proportion and the confidence coefficient?
d. Compare the lengths of the interval between those in problems 2 and 3 and make comments on the lengths and discuss why there is a difference.

Sol. c. The proportion $\approx$ the confidence coefficient 80%.
d. The mean length of the CI in #2 is longer than the length of CI in # 3,
as we have more accurate information in # 3 than in # 2. In fact,
in #2 the 80% CI is $\overline{X} \pm 1.53S/\sqrt{5}$,
in #3 the 80% CI is $\overline{X} \pm 1.28\sigma/\sqrt{5}$ (noting $E(S^2) = \sigma^2$)).

```
x=(1:150)/200
a=qexp(x)
b=-log(exp(-a)-0.1)
y=a*exp(a-a*a)-b*exp(b-b*b)
plot(a,y,type="l")
max(a[y<=0])
max(b[y<=0])
```

9.2. Suppose $X \perp \overline{X}$.
$P(X \in [\overline{x} - 1.96/\sqrt{n}, \overline{x} + 1.96/\sqrt{n}])$
$= \Phi(\overline{x} + 1.96/\sqrt{n}) - \Phi(\overline{x} - 1.96/\sqrt{n})$?
$P(X \in [\overline{X} - 1.96/\sqrt{n}, \overline{X} + 1.96/\sqrt{n}])$
$= \Phi(\overline{X} + 1.96/\sqrt{n}) - \Phi(\overline{X} - 1.96/\sqrt{n})$?
$P(X \in [\overline{X} - 1.96/\sqrt{n}, \overline{X} + 1.96/\sqrt{n}])$
$= P(-1.96/\sqrt{n} \leq X - \overline{X} \leq 1.96/\sqrt{n}])$?
$X - \overline{X} \sim$?

$$P(X \in [\overline{X} - 1.96/\sqrt{n}, \overline{X} + 1.96/\sqrt{n}]) = P(\frac{|X - \overline{X}|}{\sqrt{1+1/n}}| \leq \frac{1.96/\sqrt{n}}{\sqrt{1+1/n}}).$$

Howmework week 9 and 10
Additonal.

1. A tire company guarantees that a particular brand of tire has a mean lifetime of **<u>42 thousand miles or more</u>**. A consumer test agency collected 10 observations as follows: 42, 36, 46, 43, 41, 35, 43, 45, 40, 39. Assume the lifetime has an exponential distribution. Use these data to determine a 95% and a 99% confidence intervals for the mean lifetime of a tire using the pivotal method and the LRT method, separately.

Sol. The LRT method depens on H_1.
H_1: $\mu \neq 42$ or $\mu < 42$ **???**
The LRT test leads to RR: $\{\overline{X} < a\}$.
The CI is $\frac{2\sum_{i=1}^n X_i}{\chi^2_{2n,\ \underbrace{1-\alpha}_{right-tail}}} < \mu$ or $(21.1, \infty)$.
The pivotal $T = 2\sum_{i=1}^n X_i/\mu \sim \chi^2_{2n}$. The CI is

$\frac{2\sum_{i=1}^n X_i}{\chi^2_{2n,\ \underbrace{1-\alpha}_{right-tail}}} < \mu$ (common sense),

or a shortest one, **but how ?**

It is not obvious to use T, as T leads to

$[\frac{2\sum_i X_i}{b}, \frac{2\sum_i X_i}{a}]$ with $F_T(b) - F_T(a) = 0.95$.

Let $W = 1/T$.

$F_W(w) = P(W \leq w) = P(T \geq 1/w) = 1 - F_T(1/w)$ and

$f_W(w) = f_T(1/w)1/w^2$, $w > 0$.

$T \sim \chi^2_{2n}$ with df $f_T(t) = \frac{t^{\frac{2n}{2}-1}e^{-t/2}}{\Gamma(n)2^n}$, $t > 0$.

The shortest CI is $[2\sum_{i=1}^n X_i a, 2\sum_{i=1}^n X_i b]$, where

$F_W(b) - F_W(a) = 1 - \alpha$ and $f_W(a) = f_W(b)$.

9.3. X_1, ..., X_n are i.i.d. from $F(x) = (\frac{x}{\beta})^\alpha$ if $x \in (0, \beta)$ and α is given. 95% upper confidence limit for β, where $X_{(n)} = 25$, $\alpha = 12.59$, $n = 14$.

Sol. $f(x;\beta) = (\alpha/\beta)(x/\beta)^{\alpha-1}$, $x \in (0, \beta)$.

H_0: $\beta = \beta_o$, v.s. H_1: $\beta \neq \beta_o$.

MLE: $\hat{\beta}_0 = \beta_o$, $\hat{\beta} = X_{(n)}$.

LRT leads to

Acceptance region : $\{\mathbf{x}: \ \lambda(\mathbf{x}) > c\}$, where

$\lambda = \frac{L(\mathbf{X};\hat{\theta}_o)}{L(\mathbf{X};\hat{\theta})} = \prod_{i=1}^n \frac{(\frac{X_i}{\beta_0})^\alpha}{(\frac{X_i}{X_{(n)}})^\alpha} = (\frac{X_{(n)}}{\beta_0})^{n\alpha}$.

Notice that $T = X_{(n)}/\beta$ is a pivotal random variable, as

$P(T \leq t) = t^{n\alpha}$ if $t \in [0, 1]$.

The acceptance region leads to

$P(a \leq T) = 0.95 = 1 - a^{n\alpha}$ **??** or

$P(a \leq T \leq 1) = 0.95 = 1 - a^{n\alpha}$ **??**

$a = 0.05^{\frac{1}{n\alpha}}$ and $b = 1$.

$a \leq \frac{X_{(n)}}{\beta} \leq 1$,

$\frac{X_{(n)}}{1} \leq \beta \leq \frac{X_{(n)}}{a}$.

95% CI for β is

$[25, 25.43]$ **or** $(0, 25.43]$ **??**

\# **9.4.** Use LRT to derive CI for λ, where

X_1, ..., $X_n \sim N(0, \sigma_X^2)$

Y_1, ..., $Y_m \sim N(0, \sigma_Y^2)$

$\lambda = \sigma_Y^2/\sigma_X^2$,

A2. Then generate a set of data with $n = 9$ and $m = 16$ from normal distributions and derive the LRT based 90% CI of λ using numerical method. **Report the sufficient statistics.**

Sol. H_0: $\lambda = \lambda_o$, v.s. H_1: otherwise.

a. MLE under Θ: $\hat{\sigma}_X^2 = \sum_i X_i^2/n$, $\hat{\sigma}_Y^2 = \sum_i Y_i^2/m$, $\hat{\lambda} = \hat{\sigma}_Y^2/\hat{\sigma}_X^2$.

MLE under H_0: $\tilde{\sigma}_X^2 = \frac{\sum_i X_i^2 + \sum_j Y_j^2/\lambda_o}{n+m}$, $\tilde{\sigma}_Y^2 = \lambda_o\tilde{\sigma}_X^2$, $\tilde{\lambda} = \lambda_o$,

$$\begin{aligned}
L =&(2\pi\sigma_X^2)^{-n/2}\exp(-\frac{1}{2}\sum_i X_i^2/\sigma_X^2)(2\pi\sigma_Y^2)^{-m/2}\exp(-\frac{1}{2}\sum_i Y_i^2/\sigma_Y^2) \\
=&(2\pi\sigma_X^2)^{-n/2}(2\pi\sigma_Y^2)^{-m/2}\exp(-\frac{1}{2}[\sum_i X_i^2/\sigma_X^2) + \sum_i Y_i^2/\sigma_Y^2]) \\
=&(2\pi\sigma_X^2)^{-n/2}(2\pi\sigma_Y^2)^{-m/2}\exp(-\frac{1}{2}\frac{\sum_i X_i^2 + \sum_i Y_i^2/\lambda}{\sigma_X^2})
\end{aligned}$$

$$
\begin{aligned}
LR \propto & \frac{(\tilde\sigma^2_X)^{-n/2}(\tilde\sigma^2_Y)^{-m/2}}{(\hat\sigma^2_X)^{-n/2}(\hat\sigma^2_Y)^{-m/2}}\\
=&(\frac{n}{n+m}\frac{\sum_i X_i^2+\sum_j Y_j^2/\lambda_o}{\sum_i X_i^2})^{-n/2}(\frac{m}{n+m}\frac{\sum_i X_i^2+\sum_j Y_j^2/\lambda_o}{\sum_i Y_i^2})^{-m/2}\\
\propto&(1+\frac{\sum_j Y_j^2/\lambda_o}{\sum_i X_i^2})^{-n/2}(\frac{\lambda_o\sum_i X_i^2}{\sum_j Y_j^2}+1)^{-m/2},\ \ T=\frac{\sum_j Y_j^2/\lambda_o}{\sum_i X_i^2}\\
\propto&(1+T)^{-n/2}(1+\frac{1}{T})^{-m/2}\\
=&(1+T)^{-n/2-m/2}(T)^{m/2}
\end{aligned}
$$

$(\ln LR)'_T=-\frac{n/2+m/2}{1+T}+\frac{m/2}{T}\begin{cases}=+\infty & \text{if } T=0+\\ <0 & \text{if } T\approx\infty \text{ but } T\neq\infty\end{cases}$
and has a unique zero point. Thus LR first ↑, then ↓

Acceptance region $a<T<b$, where $(LR(a)=LR(b))$.

Notice $F=\frac{\sum_{j=}^m Y_j^2/(m\sigma_Y^2)}{\sum_{i=1}^n X_i^2/(n\sigma_X^2)}=\frac{n}{m}T\sim F_{m,n}$.

Acceptance region can be written as

$$a<\frac{\overline{Y^2}}{\overline{X^2}\lambda_o}<b,\ where\ F_{m,n}(b)-F_{m,n}(a)=1-\alpha,\ LR(am/n)=LR(bm/n).$$

$CI:\ (\frac{\frac{\overline{Y^2}}{\overline{X^2}}}{b},\frac{\frac{\overline{Y^2}}{\overline{X^2}}}{a})$,

Question: How to find (a,b) numerically by R-program ?
Numerical method for given $\overline{Y^2}/\overline{X^2}$:

$$LR\propto g=(1+T)^{-(n+m)/2}(T)^{m/2}$$

Assign α, say $\alpha=0.05$,
$t=F_{m,n,1-\alpha}$,
$a=$(1:999)/1000
$a=t*a$
$p=F_{m,n}(a)$
$b=F_{m,n,(\alpha-p)}$ (critical value, *i.e.*, $1-F_{m,n}(b)=\alpha-p$)
solve $LR(am/n)-LR(bm/n)=0$.

Then $a<\frac{\overline{Y}^2}{\overline{X}^2\lambda}<b$ yields
the (LRT method) CI $\frac{\overline{Y}^2}{\overline{X}^2 b}<\lambda<\frac{\overline{Y}^2}{\overline{X}^2 a}$, where $LR(am/n)-LR(bm/n)=0$.

An approximate CI: $(\frac{\frac{\overline{Y^2}}{\overline{X^2}}}{F_{m,n,\alpha/2}},\frac{\frac{\overline{Y^2}}{\overline{X^2}}}{F_{m,n,1-\alpha/2}})$ if n and m are large.
But it is not the LRT CI !

```
R program for #9.4 m=10
n=15
x=rnorm(n)
y=rnorm(m,0,3)
f=mean(y**2)/mean(x**2)
f # [1] 10.5161
t=qf(0.05,m,n)
a=(1:999)/1000
a=t*a
p=pf(a,m,n)
b=qf(1-0.05+p,m,n)
```

```
r=m/n
y=(1+r*b)**(-(n+m)/2)*(r*b)**(m/2)
x=(1+r*a)**(-(n+m)/2)*(r*a)**(m/2)
z=y-x
u=z[z>=0]
s=min(u)
f/b[z==s]
f/a[z==s]
```

Simulation result: The 95% CI is [3.359162, 36.13342]

A different numerical method for given $\overline{Y^2}/\overline{X^2}$:

1. Generate a sequence of λ. For each λ: **(note λ is not LR)**
2. find c_λ for the RR: $LR \le c_\lambda$;
3. $\lambda \in$CI if λ is not rejected.

This can be done as follows:
Treat $\overline{Y^2}/\overline{X^2}$ as one of a or b and find the other one.
Compute the $P(a < T < b)$, if it is $< 1-\alpha$, then $p \in$ CI.

9.6. 90% CI of p based on two-tailed test with $X = X_1 \sim bin(n,p)$.

A3. Do #9.6. Then generate a r.v. from $bin(n,p)$ with your (n,p), **report the sufficient statistic values** and derive the LRT based 90% CI of p using numerical method.

Sol. H_0: $p = p_o$ v.s. H_1: $p \ne p_o$.
MLE: $\hat{p}_0 = p_o$ and $\hat{p} = X/n$.
$\lambda = \frac{p_o^X(1-p_o)^{n-X}}{(X/n)^X(1-X/n)^{n-X}}$.
LRT $\phi = \mathbf{1}_{(\lambda \le c)}$

$$\ln\lambda = x\ln p_o + (n-x)\ln(1-p_o) - x\ln x + x\ln n - (n-x)\ln(1-x/n)$$

$$(\ln\lambda)'_x = \ln p_o - \ln(1-p_o) - \ln(x/n) - x/x + \ln(1-x/n) + (n-x)\frac{1/n}{1-x/n}$$

$$= \ln\frac{p_o}{1-p_o} + \ln\frac{1-x/n}{x/n} = \begin{cases} + & \text{if } x = 0 \\ - & \text{if } x = n \end{cases}$$

RR: $X \le a$ or $X \ge b$
AR: $a < X < b$, where $\lambda(a) = \lambda(b)$.
Notice that in LRT $\phi = \mathbf{1}_{(\lambda \le c)}$, c depends on p_o, thus write

$$\phi = \mathbf{1}_{(\lambda \le c_{p_o})}.$$

where c_{p_o} satisfies that (1) $E_{p_o}(\phi) \le \alpha$,
(2) $E_{p_o}(\phi) + P_{p_o}(X \in (a, a+1]) > \alpha$ and (3) $E_{p_o}(\phi) + P_{p_o}(X \in [b-1, b)) > \alpha$.
Then a CI for p is $\{p : \frac{p^X(1-p)^{n-X}}{(X/n)^X(1-X/n)^{n-X}} > c_p\}$, where $P_p(\lambda \le c_p) \le \alpha$, $E_p(\phi) + P_p(X \in (a, a+1]) > \alpha$ and $E_p(\phi) + P_p(X \in [b-1, b)) > \alpha$.

In other words, the test is $\phi = \mathbf{1}_{(X \notin (a,b))}$, where
$\ln\lambda(a) = \ln\lambda(b)$,
$P_{p_o}(\lambda(X) \le \lambda(a)) \le \alpha$ and $P_{p_o}(\lambda(X) \le \lambda(a+1)) > \alpha$.
That is, observe $X = x$,
we do not reject H_o if the p-value $P_{p_o}(\lambda(X) \le \lambda(x)) > \alpha$.
$(1-\alpha)$ CI : $\{p : P_p(\lambda(X) > \lambda(x)) < 1-\alpha\}$.

The idea of deriving CI numerically is as follows.

1. Generate a sequence of p. For each p:
2. find c_p;
3. $p \in$CI if p is not rejected.

```
# R for 9.6.
n=10
x=rbinom(1,n,0.2)
x
[1] 8 # an observation for testing
p=(1:1000)/1001
z=rep(0,1000) # 1(p is accepted)
for (j in 1:1000){
y=1:(n-1)
g= y*log(p[j]) +(n-y)*log(1-p[j]) -y*log(y/n) -(n-y)*log(1-(y/n)) # ln λ
g=c(n*log(1-p[j]),g,n*log(p[j]))
u = 0
for(i in 1:(n+1)) {
if( g[i] > g[x]) u=u+dbinom(i,n,p[j])
}
if (u < 0.9) z[j] =1
}
min(p[z==1])
max(p[z==1])
CI: [0.5004995, 0.9450549]
# not a LRT CI
n=10
m=1000
t=0.4
x=rbinom(1,n,t)
x=8
q=(1:m)/m
p=1:m
for (i in 1:m){
y=binom.test(x,n,q[i])
p[i] = y$p.value
}
q=q[p>=0.1]
min(q)
max(q)
[1] 0.5
[1] 0.945
```

9.13 (b). Let $X \sim Beta(\theta, 1)$. Find a pivotal quantify and a CI of size $e^{-1/2} - e^{-1}$.

Sol. $f_X(x) = \theta x^{\theta-1}$, $x \in (0,1)$. Let $T = -log(X^\theta) = -\theta \log X$. $X = e^{-T/\theta}$, $\frac{dx}{dt} = \frac{1}{\theta}e^{-t/\theta}$, $f_T(t) = \theta exp(-(t/\theta)(\theta-1))\frac{1}{\theta}e^{-t/\theta} = e^{-t}$, $t > 0$.

$e^{-0.5} - e^{-1} = P(T \leq b) = 1 - e^{-b}$. $0 \leq -\theta \log(X) \leq b$. $0 \leq \theta \leq \frac{b}{-\log X}$.

It is the shortest CI with the confidence coefficience (due to $f_T(a) = f_T(b)$).

9.12. Assume X_1, ..., X_n are i.i.d. from $N(\theta, \theta)$. Construct a 95% CI for θ.

Sol. 3 approaches:

(1) Pivotal method,

(2) LRT method,

(3) $\overline{X} \pm t_{n-1,\alpha/2} S/\sqrt{n}$ **(not a good one), Why ??**

(1a) A pivotal $T = \frac{\overline{X}-\theta}{\sqrt{\theta/n}} \sim N(0,1)$. It leads to

$$|\frac{\overline{X}-\theta}{\sqrt{\theta/n}}| \leq 1.96 \text{ (and } \theta > 0).$$

or $(\overline{X}-\theta)^2 \le 1.96^2\theta/n$.
$\theta^2 - 2\overline{X}\theta + (\overline{X})^2 - 1.96^2\theta/n \le 0$.
$\theta^2 - 2(\overline{X} + \frac{1.96^2}{2n})\theta + (\overline{X})^2 \le 0$.
Two solutions to the equation: $\theta^2 - 2(\overline{X} + \frac{1.96^2}{2n})\theta + (\overline{X})^2 = 0$:
$\theta = \overline{X} + \frac{1.96^2}{2n} \pm c$, if $2\overline{X} + \frac{1.96^2}{2n} \ge 0$, where
$c = \sqrt{(\overline{X} + \frac{1.96^2}{2n})^2 - (\overline{X})^2}$.
$[\overline{X} + \frac{1.96^2}{2n} - c, \overline{X} + \frac{1.96^2}{2n} + c]$. **It is right ???**

$[0 \vee (\overline{X} + \frac{1.96^2}{2n} - c), 0 \vee (\overline{X} + \frac{1.96^2}{2n} + c)]$. **It is right ???**
Solutions 95% CI for θ:

$$\begin{cases} [0 \vee (\overline{X} + \frac{1.96^2}{2n} - c), 0 \vee (\overline{X} + \frac{1.96^2}{2n} + c)] & \text{if } 2\overline{X} + \frac{1.96^2}{2n} \ge 0 \\ \{0\} & \text{if } 2\overline{X} + \frac{1.96^2}{2n} < 0 \end{cases}$$

Q: Why $\{0\}$ is a 95% CI for θ in the latter case ?
Hint: Compute the coverage probability.
(1b) Another pivotal $Y = \frac{(n-1)S^2}{\theta} \sim \chi^2(n-1)$. **Not a good one**. Thus
$P(a \le Y \le b) = 1-\alpha$
=> 95% CI is $[(n-1)S^2/b, (n-1)S^2/a]$,
where $f_Y(a) = f_Y(b)$ gives the shortest CI of this type.
(2) LRT method
$f_{\mathbf{X}}(\mathbf{x};\theta) = \prod_{i=1}^n f(x_i;\theta) = \frac{1}{\sqrt{2\pi\theta/n}} e^{-\sum_{i=1}^n \frac{(x_i-\theta)^2}{2\theta}}$.
$\ln f_{\mathbf{X}}(\mathbf{x},\theta) = c + \frac{-n}{2}\ln\theta - \frac{n}{2}(\frac{\overline{x^2}}{\theta} - 2\overline{x} + \theta)$.
$\ln f_{\mathbf{X}}(\mathbf{x},\theta)' = -\frac{n}{2\theta} - \frac{n}{2}(-\frac{\overline{x^2}}{\theta^2} + 1) = 0$.
$\theta^2 + \theta - \overline{x^2} = 0$.
$\theta = \frac{-1 \pm \sqrt{1+4\overline{x^2}}}{2}$.
=> $\hat\theta = \frac{-1+\sqrt{1+4\overline{x^2}}}{2} > 0$.
Check: $\begin{matrix} \theta: & 0 & \hat\theta & \infty \\ \ln f_{\mathbf{X}}: & -\infty & finite & -\infty \end{matrix}$.
The MLE is $\hat\theta = \sqrt{\overline{x^2} + \frac{1}{4}} - 1/2$.
The derivation of the CI is more complicated and is skipped here.
We consider a special case as follows.

A4. 9.12. Generate a random variable X_1 from $N(\theta,\theta)$ using R ($n = 1$) and report the sufficient statistic value. Use the LRT approach to derive a 90% CI for θ by numerical method.
Sol. LRT based on $Y \sim N(\theta,\theta)$ with $n = 1$.
$f_Y(y,\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(y-\theta)^2}{2\theta}}$
$\ln f_Y(y,\theta) = c + \frac{-1}{2}\ln\theta - \frac{1}{2}(\frac{y^2}{\theta} - 2y + \theta)$.
$(\ln f_Y(y,\theta))' = -\frac{1}{2\theta} - \frac{1}{2}(-\frac{y^2}{\theta^2} + 1) = 0$.
$\theta^2 + \theta - y^2 = 0$.
$\theta = \frac{-1 \pm \sqrt{1+4y^2}}{2}$. =>

$$\hat\theta = \frac{-1 + \sqrt{1+4y^2}}{2} > 0. \qquad (1)$$

Check: $\begin{matrix} \theta: & 0 & \hat\theta & \infty \\ \ln f_Y: & -\infty & finite & -\infty \end{matrix}$.
The MLE is $\hat\theta$.

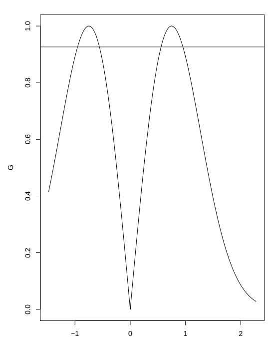
$$\lambda = \frac{\frac{1}{\sqrt{\theta_0}} e^{-\frac{(y-\theta_0)^2}{2\theta_0}}}{\frac{1}{\sqrt{\hat\theta}} e^{-\frac{(y-\hat\theta)^2}{2\hat\theta}}}$$

The 95% CI of θ induced by the LRT is $\{\theta : \frac{\frac{1}{\sqrt{\theta}} e^{-\frac{(y-\theta)^2}{2\theta}}}{\frac{1}{\sqrt{\hat\theta(y)}} e^{-\frac{(y-\hat\theta(y))^2}{2\hat\theta(y)}}} > c_\theta\}$, or

$$\{\theta : \frac{\sqrt{\hat\theta(y)}}{\sqrt{\theta}} \exp(-\frac{(y-\theta)^2}{2\theta} + \frac{(y-\hat\theta(y))^2}{2\hat\theta(y)}) > c_\theta\},$$

Notice that in # 9.23, the function $\lambda(x,\theta)$ is concave down,
but the function $\lambda(y,\theta)$ here may not be concave down,
and the acceptance region may not be an interval.
Sketch of a numerical solution for given $Y = y$:
1. Give a range of θ, say $(0, y + 3\sqrt{y})$ **(why ?)**
2. For each $\theta \in (0, y + 3\sqrt{y})$, it belongs to the CI if θ is not rejected.

This can be done in two ways:
a. Find c_θ for each θ. Check $\lambda(x,\theta) > c_\theta$?
b. Compute the p-value. Check p-value $> \alpha$?
Or Compute 1−p-value. Check 1−p-value $< 1-\alpha$?
If the acceptance region (AR) is $[a,b]$, and X is cts,$1-p = F(b) - F(a)$.
If the AR is $[a,b]$ and X is discrete, $1-p = F(b) - F(a-1)$.
If the AR is not an interval, $1-p = \int \mathbf{1}(\lambda(x,\theta) \geq \lambda(x_o,\theta))dF(x;\theta)$.



Graph of $y = \lambda(x,\theta)$ with $\theta = 4$ and observation $X = y = 0.557788$.

```
y=rnorm(1,1,1)
# y=1.457788
y
m=2000    # for Y
k=1000
p=(1:k)/100
z=rep(0,k)
u=(3:(m-3))/m   # partition (0,1) into 2000 parts
for (j in 1:k){
q=p[j]
t=sqrt(y*y+0.25)-0.5   # MLE based on observation
g=sqrt(t/q)*exp(-(y-q)**2/(2*q)+(y-t)**2/(2*t))   # λ
x=qnorm(u,q,sqrt(q))   # quantiles
t=sqrt(x*x+0.25)-0.5   # MLE for x
G=sqrt(t/q)*exp(-(x-q)**2/(2*q)+(x-t)**2/(2*t))    # λ
x=x[G>=g]
w=length(x)/m   # ∫ 1(λ(x,θ) ≥ λ(x_o,θ))dF(x;θ_o).
```

```
if (w < 0.95) z[j] =1
}
z    # check whether adjustment is need for p.
a=min(p[z==1])
b=max(p[z==1])
c(a,b)
[1] 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[38] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[75] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[112] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[149] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[186] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[223] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[260] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[297] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[334] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0
[371] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[408] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[445] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[482] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[519] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[556] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[593] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[630] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[667] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[704] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[741] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[778] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[815] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[852] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[889] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[926] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[963] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[1000] 0

[1] 0.04 3.61
```

(2) LRT based on sufficient statistic $Y = \overline{X} \sim N(\theta, \theta/n)$.

There is a mistake in the following approach.

$f_Y(y,\theta) = \frac{1}{\sqrt{2\pi\theta/n}} e^{-\frac{(y-\theta)^2}{2\theta/n}}$

$\ln f_Y(y,\theta) = c + \frac{-1}{2}\ln\theta - \frac{n}{2}(\frac{y^2}{\theta} - 2y + \theta)$.

$(\ln f_Y(y,\theta))' = -\frac{1}{2\theta} - \frac{n}{2}(-\frac{y^2}{\theta^2} + 1) = 0$.

$\theta^2 + \theta/n - y^2 = 0$.

$\theta = \frac{\frac{-1}{n} \pm \sqrt{\frac{1}{n^2} + 4y^2}}{2}$.

$=> \hat{\theta} = \frac{\frac{-1}{n} + \sqrt{\frac{1}{n^2} + 4y^2}}{2} > 0$.

Check: $\begin{array}{cccc} \theta: & 0 & \hat{\theta} & \infty \\ \ln f_Y: & -\infty & finite & -\infty \end{array}$.

The MLE is $\hat{\theta} = \sqrt{(\overline{X})^2 + \frac{1}{4n^2}} - \frac{1}{2n}$.

$$\lambda = \frac{\frac{1}{\sqrt{\theta_0}} e^{-\frac{(y-\theta_0)^2}{2\theta_0/n}}}{\frac{1}{\sqrt{\hat\theta}} e^{-\frac{(y-\hat\theta)^2}{2\hat\theta/n}}}$$

The 95% CI of θ induced by the LRT is $\{\theta : \frac{\frac{1}{\sqrt{\theta}} e^{-\frac{(y-\theta)^2}{2\theta/n}}}{\frac{1}{\sqrt{\hat\theta}} e^{-\frac{(y-\hat\theta)^2}{2\hat\theta/n}}} > c\}$, or

$$\{\theta : \frac{\sqrt{\hat\theta}}{\sqrt{\theta}} \exp(-\frac{(y-\theta)^2}{2\theta/n} + \frac{(y-\hat\theta)^2}{2\hat\theta/n}) > c\}, \text{ where } \hat\theta = \sqrt{y^2 + \tfrac{1}{4n^2}} - \tfrac{1}{2n}.$$

Sketch of a numerical solution for given $\overline{X} = y$:

1. Give a range of θ, say $(0, y + 3\sqrt{y/n})$.
2. For each $\theta \in (0, y + 3\sqrt{y/n})$, it belongs to the CI if θ is not rejected.

Remark. There is an error in the previous discussion unless $n = 1$. The sufficient statistics is not $\overline{X}$, but $(\overline{X}, \overline{X^2})$. Thus the derivation is not correct.

 # 9.23. $X_1, ..., X_n \sim Poisson(\theta)$, derive a 90% CI based on Formula (9.2.17) in page 435 and LRT.

x=c(155,104,66,50,36,40,30,35,42)

Sol. (1) Method from Formula (9.2.17) in page 435.

$$[\frac{1}{2n}\chi^2_{2y_o,1-\alpha/2}, \frac{1}{2n}\chi^2_{2(y_o+1),\alpha/2}].$$

Here $y_o = 558$. $\alpha = 0.1$

```
a=qchisq(0.05,2*y)/(2*n)
b=qchisq(0.95,2*(y+1))/(2*n)
c(a,b)
[1] 57.74689 66.49441
```

(2) LRT method. H_0: $\theta = \theta_o$ v.s. H_1: $\theta \neq \theta_o$

Sufficient and complete statistic $Y = \sum_i X_i \sim Poisson(n\theta)$.

$$f_Y(y;\theta) = e^{-n\theta}(n\theta)^y/y!$$

MLE: $\hat\theta = Y/n$, $\hat\theta_0 = \theta_o$.

$$\lambda = \frac{e^{-n\theta_o}(n\theta_o)^y}{e^{-y}y^y} = e^{y-n\theta_o}(n\theta_o/y)^y$$

$$\ln\lambda = y - n\theta_o + y\ln(n\theta_o) - y\ln y$$

$$(\ln\lambda)'_y = 1 + \ln(n\theta_o) - \ln y - 1$$

Thus λ first $\uparrow$ and then $\downarrow$.

$\phi = \mathbf{1}_{(y\notin(a,b))}$, where $\lambda(a) = \lambda(b)$ and $E_{\theta_o}(\phi) \leq \alpha$,

$E_{\theta_o}(\phi) + P_{\theta_o}(Y \in [a, a+1)) > \alpha$, $E_{\theta_o}(\phi) + P_{\theta_o}(Y \in (b-1, b]) > \alpha$.

The 95% CI of θ is $\{\theta : \lambda > c_\theta\}$.

$\theta \in$ CI if $\begin{cases} F_Y(y;\theta) - F(b-1,\theta) < 1-\alpha & \text{if } b < y, \\ F_Y(b;\theta) - F_Y(y-1;\theta) < 1-\alpha & \text{if } b \geq y \end{cases}$, where

$\{b, y\} = \{\min(A), \max(A)\}$ and

$A = \{Y : \log(\lambda(Y;\theta)) \geq \log(\lambda(y;\theta))\}$.

```
x=c(155,104,66,50,36,40,30,35,42)
y=sum(x)    #y=558
n=9       # sample size
k=1000
```

```
z=(y-200):(y+200)      # selected Y values
q=((-k):k)/10
q=q+y      # selected nθ range: y ± 100, separation by 0.1
v=q      # initialize v
for (i in 1:(2*k+1)) {
p=q[i]
g=y-p+y*log(p)-y*log(y)      # log(λ(y))
G=z-p+z*log(p)-z*log(z)      # log(λ(Y) for selected Y values)
t=z[G>=g]
m=length(t)
if (t[1] < 2) u=as.numeric(ppois(t[m],p)<0.95) else # left end of t = 0?
u=as.numeric(ppois(t[m],p)-ppois(t[1]-1,p)<0.95)
v[i]=u
}
v
x=q[v == 1]
x[c(1, length(x))]/n
[1] 57.04444 67.25556
```

Additional 8.2. (b) Redo the following problem and compute $P(H_0|H_1)$:
Carry out the following simulation project.

1.b.1. Use R to generate 5 observations from $N(1,1)$. Now pretend that you only known that the data were from $N(\mu,\sigma)$ without knowing μ and σ, use t-test to test H_0: $\mu = 0$ v.s. H_1: $\mu \neq 0$ with a size 0.2. Record the P-value. R commands are :

$x < -rnorm(5) + 1$
y=t.test(x)
y$p.value

What is a correct decision here (in terms of rejecting H_0 or not) ?

1.b.2. Repeat procedure 2.1 100 times. That is, record 100 P-values.

How many times, say z, would you reject H_0 ?

Question: What does the number z tell you about $P(H_0|H_1)$?

Sol. Under given condition, $X \sim N(1,1)$ and $\frac{\overline{X}}{S/\sqrt{n}} \sim$ non-central t-distribibtion with degree freedom $n = 5$ and non-central parameter
$(\mu\sigma^2/\sqrt{n} = 1 \times \sqrt{n})$. $P(H_o|H_1) = P(|\frac{\overline{X}}{S/\sqrt{n}}| < t_{n-1,\alpha/2}) = 0.2391$
(see R program below).
The number $z = 74$ tells us $1 - z/100 = 0.26 \approx P(H_o|H_1)$.
R:

```
n=5
q=qt(0.9,n-1)
pt(q, n-1, sqrt(n))-pt(-q, n-1, sqrt(n))
[1] 0.2391017
```

8.47. Consider two independent normal sample with equal variance. $\vdash$: The LRT for H_o^-: $\mu_X - \mu_y \leq -\delta$ v.s. H_1^-: $\mu_X - \mu_y > -\delta$ is
$\phi = \mathbf{1}(T^- > t_{m+n-2,\alpha})$, where $T^- = \frac{\overline{X}-\overline{Y}-(-\delta)}{\sqrt{(\frac{1}{m}+\frac{1}{n})S_p^2}}$

Sol. Since $\overline{X}$ and $\overline{Y}$ are the sufficient statistics of (μ_x, μ_y) and the parameter of interest is $\theta = \mu_x - \mu_y$, it suffices to consider the likelihood of $\overline{X} - \overline{Y} \sim N(\theta, \sigma^2(\frac{1}{m} + \frac{1}{n}))$. ......

Homework solutions for week 11

Additional. R project:

A.1. Generate n=9 observations from a $N(\mu_0, \sigma^2)$.
2. Perform a t-test for H_0: $\mu = \mu_0$ against H_1: $\mu \neq \mu_0$.
3. Repeat Steps 1 and 2 400 times, count the number m_k of rejections among the first k times for $k = 1$, 2, 3, ..., 400.
4. plot(m_k/k,k) for $k \in \{1, ..., 400\}$.

5. Make comment on $\{m_k/k\}_{k\geq 1}$ and α, the size of the test.

   Ans. to A.5: $m_k/k \xrightarrow{a.s.} \alpha$ **(**$= P(H_1|H_o)$ **by the SLLN.**

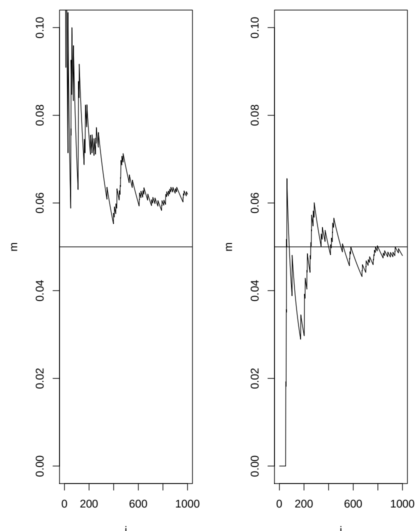
B.1. Generate n=9 observations from a discrete random variable X, with $X = a, b, c$ w.p. 1/6, 2/6, 3/6, and with mean and variance the same as in part A **(you need to determine** a, b, c **and check ?sample in R)**.

2. Perform a t-test for H_0: $\mu = \mu_0$ against H_1: $\mu \neq \mu_0$.
3. Repeat Steps 1 and 2 400 times, count the number m_k of rejections among the first k times for $k = 1$, 2, 3, ..., 400.
4. plot(m_k/k,k).
5. Make comment on $\{m_k/k\}_{k\geq 1}$ and α, the size of the test.

Ans.: $m_k/k \xrightarrow{a.s.} P(H_1|H_o)$ **by the SLLN. Notice that it is most likely that** $\alpha \neq P(H_1|H_o)$**, as the distribution is not normal.**

C. Let $n = 80$ and repeat A and B. Is there any difference on your comments. If there is one, why ?

Ans.: Under the assumption in A, compare the two curves of $(m_k/k, k)$**, the variation is smaller for** $n = 80$**, about** $\sqrt{9}/\sqrt{80} \approx 1/3$ **variation for** $n = 9$**.** $m_k/k \xrightarrow{a.s.} P(H_1|H_o)$ **by the SLLN. Notice that** $P(H_1|H_o) = \alpha$**, as the distribution is normal.**



Under the assumption in B $m_k/k \xrightarrow{a.s.} P(H_1|H_o)$ **by the SLLN. Notice that** $P(H_1|H_o) \approx \alpha$**, as the distribution is approximately normal by the CLT.**

```
myfun=function(n){
z=0
N=1000
m=1:N
for (i in 1:N){
x=rnorm(n)+1
y=t.test(x, mu=1)
z=z+(y$p.value<0.05)
m[i]=z/i
}
i=1:N
plot(i,m, xlim=c(0,N), ylim=c(0.0,0.10), type="l",lty=1)
abline(h=0.05)
}
makepsfile = function() {
ps.options(horizontal = F)
ps.options(height=9.0, width=6.5)
```

```
postscript("fighw.ps")
par(mfrow=c(1,2))
n=9
myfun(n)
n=80
myfun(n)
dev.off()
}
makepsfile()
```

Additional.

A6.

1. Generate n=4 observations from a discrete random variable X, with $X = a, b, c$ w.p. 1/6, 2/6, 3/6, and with mean and variance the same as in part A5 **(you need to determine a, b, c and check help(sample) in R)**.
2. Perform a t-test for H_0: $\mu = \mu_0$ against H_1: $\mu \neq \mu_0$ with $\alpha = 0.1$
3. Repeat Steps 1 and 2 400 times, count the number m_k of rejections among the first k times for $k = 1$, 2, 3, ..., 400.
4. Plot(k,m_k/k).
5. Do you believe the size of the test is 0.1 based on m_k/k ?
6. Compute the size of the test $P(H_1|H_0)$ here. Make comment on the relation between $\{m_k/k\}_{k\geq 1}$, α, and the size of the test.

Sol. (6) The size $P(H_1|H_0) = P(\frac{|\overline{X}-\mu_o|}{S/\sqrt{n}} > t_{\alpha/2,n-1}) = ?$

> qt(0.05,3) # **why ?**

[1] -2.353363

t-test: $\phi = \mathbf{1}(\frac{|\overline{X}-\mu_o|}{S/\sqrt{n}} > 2.35)$.

The size $P(H_1|H_0) = P(\frac{|\overline{X}-\mu_o|}{S/\sqrt{n}} > 2.35) = ?$

$P((X_1, ..., X_4) \in A) = \begin{cases} \int \cdots \int_A f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} & \text{if cts} \\ \sum_{\mathbf{x} \in A} f_{\mathbf{X}}(\mathbf{x}) & \text{if discrete} \end{cases}$. $\mathbf{X} =$ **?**

The size $P(H_1|H_0)$

$$= \sum_{x_1,x_2,x_3,x_4} \mathbf{1}(\frac{|\overline{x}-\mu_o|}{S/\sqrt{4}} > 2.35) f(x_1)f(x_2)f(x_3)f(x_4)$$

$$= \sum_{x_1,x_2,x_3,x_4} (1 - \mathbf{1}(x_1 = \cdots = x_4))\mathbf{1}(\frac{|\overline{x}-\mu_o|}{s/\sqrt{4}} > 2.35) \prod_{i=1}^{4} f(x_i) + \frac{1}{6^4} + \frac{1}{3^4} + \frac{1}{2^4}$$

$=$.......

where $f(x) = \begin{cases} 1/6 & \text{if } x = a \\ 2/6 & \text{if } x = b \\ 3/6 & \text{if } x = c, \end{cases}$ $E(X) = \mu_o$ and $V(X)= \sigma^2$ for (μ_o, σ) in A5.

Ans. $\alpha = 0.1 \neq P(H_1|H_0)$, size of the test.

$m_k/k = \overline{Z} = \overline{\mathbf{1}(reject\ correct\ H_o)} \xrightarrow{a.s.} P(H_1|H_0)$ due to SLLN

A8. Assume the assumption as in #10.1.

(a) Select a parameter (say $\theta_o = 0.5$), and generate a random sample of 25, say O, through
n=25
O= runif(n)
$O = F_X^{-1}(O)$, # convert F_X^{-1} to R formula.
where $F_X(t) = \mathbf{1}(t \in [-1, 1)) \int_{-1}^{t} \frac{1}{2}(1 + \theta y)dy + \mathbf{1}(t \geq 1)$.

(b) Derive the likelihood function $L(\theta, O)$ and the MLE using numerical method, *e.g.*, plotting $(\theta, L(\theta, O))$.

(c) Show that the MLE is consistent by verifying (A1) – (A5) in Theorem 1 of §10.1.

(d) Estimate the asymptotic variances of the MLE and the MME with given O. Which is smaller ?

Sol. (a) $f_X(x) = 0.5(1 + \theta x)$, $x \in (-1, 1)$, $|\theta| < 1$.
Recall that
$Y = F_X(X) \sim U(0, 1)$ if X is a cts r.v..
Thus pseudo random number can be derived by R through $F_X^{-1}(Y)$.
Here F_X^{-1} is given by
$y = \int_{-1}^{x} \frac{1+\theta t}{2} dt = (0.5t + 0.125t^2)|_{-1}^{x} = 0.125x^2 + 0.5x + 0.375$.
$0.125x^2 + 0.5x + 0.375 - y = 0$.
$x = \frac{-0.5 \pm \sqrt{0.5^2 - 4(0.125)(0.375-y)}}{2*0.125} = \frac{-0.5 \pm \sqrt{0.5(0.125+y)}}{0.25}$.

$$F_X^{-1}(y) = \frac{-0.5 + \sqrt{0.5(0.125 + y)}}{0.25}$$

```
n=25
y=runif(n)
x=4*(-0.5 + sqrt(0.5*(0.125+y)))
round(x,2)
[1] 0.35 0.80 0.64 0.93 0.94 0.88 0.46 0.07 0.99 0.97 -0.73 0.74
[13] -0.74 0.60 0.90 0.54 0.68 -0.75 0.02 0.30 -0.19 -0.94 -0.24 0.64
[25] 0.06
```

(b) MLE $\hat{\theta}$ = ?
Usual approaches:
(1) Solve $\frac{\partial \ln L(\mathbf{X};\theta)}{\partial \theta} = 0$ if feasible, such as $N(\mu, \sigma^2)$.
(2) Compare $L(\mathbf{x}; \theta)$, $\theta \in \Theta$ if the latter is finite, such as $\Theta = \{\theta_0, \theta_1\}$
Neither works here !

Two numerical methods:
(1) graph $\mathcal{L}(\theta)$ to find the maximum point, $\mathcal{L}(\theta) \propto \prod_{i=1}^{n}(1 + \theta x_i)$
(2) Newton-Raphson method.

$$\theta^{new} = \theta^{old} - \left(\frac{d\ln\mathcal{L}(\theta)}{d\theta} / \frac{d^2\ln\mathcal{L}(\theta)}{d\theta^2}\right)\Big|_{\theta=\theta^{old}} \text{ until } |\theta^{new} - \theta^{old}| < \epsilon.$$

```
t=-1+(1:2000)/1000
L=rep(0,2000)
for (i in 1:2000)
L[i]=prod(1+t[i]*x)
plot(t,L,lty="l") # bell-shape
plot(t[500:2000],L[500:2000],type="l")
h=max(L)
t[L==h]
[1] 0.517 # MLE of θ
```

(d) $\sigma^2_{\hat{\theta}}$ is not easy to derive.
But $\hat{\sigma}^2_{\hat{\theta}}$ can.
$\hat{\sigma}^2_{\hat{\theta}} = \frac{(\tau(\hat{\theta})')^2}{I_n(\hat{\theta})}$ or $(\tau(\hat{\theta})')^2/\hat{I}_n(\hat{\theta})$ or CLT or Delta method.
$\tau(\theta) = ??$
$I_n(\theta) = nI_1(\theta)$ $I_1(\theta) = E((\frac{\partial \ln f(X;\theta)}{\partial \theta})^2) = -E(\frac{\partial^2 \ln f(X;\theta)}{\partial \theta^2})$.
$\ln(X; \theta) = \ln[0.5(1 + \theta X)]$
$\hat{I}_n(\hat{\theta}) = \sum_{i=1}^{n}(\frac{X_i}{1+\hat{\theta}X_i})^2$
$I_1(\theta) = E((\frac{\partial \ln f(X;\theta)}{\partial \theta})^2)$

$= E((\frac{X}{1+\theta X})^2)$
$= \int_{-1}^{1} (\frac{t}{1+\theta t})^2 \frac{1+\theta t}{2} dt$
$= \int_{-1}^{1} \frac{t^2}{1+\theta t} dt/2$
$= \int_{-1}^{1} \frac{t^2}{\theta^{-1}+t} dt \frac{1}{2\theta}$
$= \int_{-1}^{1} \frac{t^2+2\theta^{-1}t+\theta^{-2}-2\theta^{-1}(t+\theta^{-1})+\theta^{-2}}{\theta^{-1}+t} dt \frac{1}{2\theta}$
$= \int_{-1}^{1} t + \theta^{-1} - 2\theta^{-1} + \frac{\theta^{-2}}{\theta^{-1}+t} dt \frac{1}{2\theta}$
$= \left[t^2/2 - \theta^{-1}t + \theta^{-2}\ln(t+\theta^{-1})\right]_{-1}^{1} \frac{1}{2\theta}$
$= \left[-2\theta^{-1} + \theta^{-2}\ln\frac{1+\theta^{-1}}{-1+\theta^{-1}}\right] \frac{1}{2\theta}$
$= -\theta^{-2} + 0.5\theta^{-3}\ln\frac{1+\theta}{1-\theta}$

$\hat{\sigma}^2_{\hat{\theta}} = 0.10$

MME: $E(X) = \int_{-1}^{1} x(0.5) + 0.5\theta x)dx = (\frac{x^2}{4} + \frac{x^3\theta}{6})|_{-1}^{1} = \theta/3$

$\theta/3 = \overline{X} \Rightarrow \check{\theta} = 3\overline{X}$.

$\hat{\sigma}^2_{\check{\theta}} = 9S^2/n = 9(\overline{X^2} - (\overline{X})^2)/(n-1)$.

$\hat{\sigma}^2_{\check{\theta}} = 0.13$.

Which of $\hat{\sigma}^2_{\hat{\theta}}$ and $\hat{\sigma}^2_{\check{\theta}}$ is smaller ?

Why ??

(c) Verifying (A1)-(A5) for $f(x;\theta) = 0.5(1+x\theta)$, $x, \theta \in (-1,1)$.

(A1) X_1, ..., X_n are i.i.d. with $f(\cdot;\theta_o)$, $\theta_o \in \Theta$;
(A1): It is true with $\theta_o = 0.5 \in (-1,1) = \Theta$.

(A2) $f(\cdot;\theta) \neq f(\cdot;\theta^*)$ $\forall$ $\theta \neq \theta^*$ and $\theta, \theta^* \in \Theta$ (identifiability);
(A2): $f(\cdot;\theta) = f(\cdot;\theta^*)$
$\Rightarrow 0 = f(0.2;\theta) - f(0.2;\theta^*) = 0.5(\theta - \theta^*) \times 0.2$
$\Rightarrow \theta = \theta^*$.

(A3) $\{x: \ f(x;\theta) > 0\}$ does not depend on θ and $\frac{\partial}{\partial\theta} f(x;\theta)$ exists;
(A3): $\{x: \ f(x;\theta) > 0\} = (-1,1)$ does not depend on θ
$\frac{\partial}{\partial\theta} f(x;\theta) = 0.5x$ if $x, \theta \in (-1,1)$.

(A4) Θ contains an open set O and $\theta_o \in O$;
(A4) $\Theta = (-1,1)$ is open and $\theta_o = 0.5 \in (-1,1)$.

(A5) $\tau = \tau(\theta)$ is a continuous function of θ.
(A5): $\tau(\theta) = \theta$ is continuous in θ.

Additional. A4 $\vdash$: If $Y_n \xrightarrow{D} a$ and $X_n \xrightarrow{D} X$, then $Y_n/X_n \xrightarrow{D} a/X$.

Proof. Disprove unless $P(X=0)=0$. In view of Slutsky's Theorem, it suffices to disprove
$Z_n = 1/X_n \xrightarrow{D} 1/X = Z$, unless $P(X=0)=0$.

Notice that t is a continuous point of a cdf $F_W(t)$ iff $P(W=t)=0$.
Define $1/0 = +\infty$.
$F_{Z_n}(0) = P(1/X_n \leq 0) = P(X_n < 0) + P(X_n = \infty) = F_{X_n}(0-)$.
If $t < 0$, then
$F_{Z_n}(t) = P(Z_n \leq t)$
$= P(1/X_n \leq t)$
$= P(X_n < 0 \ \& \ 1 \geq X_n t)$
$= P(X_n < 0 \ \& \ 1/t \leq X_n)$
$= P(1/t \leq X_n < 0)$
$= F_{X_n}(0-) - F_{X_n}(s-)$, where $s = 1/t$.
If $t > 0$, then
$F_{Z_n}(t) = P(Z_n \leq t)$
$= P(1/X_n \leq t)$
$= P(X_n < 0 \ \& \ 1/X_n \leq t) + P(X_n \geq 0 \ \& \ 1/X_n \leq t) + P(X_n = 0 \ \& \ 1/X_n \leq t)$

$= P(X_n < 0) + P(X_n > 0 \;\&\; X_n \geq 1/t)$
$= P(X_n < 0) + P(X_n \geq 1/t)$
$= F_{X_n}(0-) + 1 - F_{X_n}(s-)$.

$$F_{Z_n}(t) = \begin{cases} F_{X_n}(0-) - F_{X_n}(s-) & \text{if } t < 0 \text{ and } s = 1/t \\ F_{X_n}(0-) & \text{if } t = 0 \\ F_{X_n}(0-) + 1 - F_{X_n}(s-) & \text{if } t > 0 \text{ and } s = 1/t, \end{cases}$$
$$= \begin{cases} F_X(0-) - F_X(s-) & \text{if } t < 0 \text{ and } s = 1/t \\ F_X(0-) & \text{if } t = 0 \\ F_X(0-) + 1 - F_X(s-) & \text{if } t > 0 \text{ and } s = 1/t, \end{cases}$$

if s is a continuous point of F_X and $P(X = 0) = 0$. Notice that if $P(X = 0) = 0$, then 0 is a continuous point of F_X. Moreover, if $t \neq 0$, then t is a continuous point of F_X iff $s = 1/t$ is a continuous point of $F_{1/X}$. Moreover, $1/X_n \xrightarrow{D} 1/X$,

A second proof. $Z_n = 1/X_n \xrightarrow{D} 1/X = Z$.

Notice $g(x) = 1/x$ with the domain $\mathcal{R}$ is not a continuous function. If $P(X = 0) = 0$, by letting $X_* = X\mathbf{1}(X \neq 0) + 1(X = 0)$, WLOG, we can assume that $\{X = 0\} = \emptyset$. Then $g(x)$ with the domain $\mathcal{R} \setminus \{0\}$ is a continuous function. Thus $g(X_n) = Z_n = 1/X_n \xrightarrow{D} 1/X = Z$.

$\vdash$: Let $P(X = 0) > 0$, let $F_X(t) = \begin{cases} 0 & \text{if } t < 0 \\ 0.5 + 0.5t & \text{if } t \in [0, 1] \\ 1 & \text{if } t > 1 \end{cases}$ for n is even.

$$F_{X_n}(t) = \begin{cases} F_X(t) & \text{if } n \text{ is even} \\ \begin{cases} 0 & \text{if } t < -1/n \\ 0.5 & t \in [-1/n, 0] \\ 0.5 + 0.5t & \text{if } t \in (0, 1] \\ 1 & \text{if } t > 1 \end{cases} & \text{if } n \text{ is odd.} \end{cases}$$

$X_n \xrightarrow{D} X$.

Notice that $F_{1/X}(t) = 0 \;\forall\; t < 0$ and thus $F_{1/X}$ is continuous for all $t < 0$.

If $t < 0$, then

$F_{Z_n}(t) = F_{X_n}(0-) - F_{X_n}(s-)$, where $s = 1/t$.

If n is odd, $F_{Z_n}(t) = 0.5$ if $s = \frac{1}{t} < -1/n$,

Thus $1/X_n \xrightarrow{D} 1/X$ is not true.

 # 10.5. Let X_1, ..., X_n be i.i.d. sample from $N(\mu, \sigma^2)$, then

$$\sqrt{n}(\frac{1}{\overline{X}} - \frac{1}{\mu}) \xrightarrow{D} N(0, \tau^2). \tag{1}$$

This means that roughly speaking, $Var(T_n) \approx \tau^2$, where $T_n = \frac{\sqrt{n}}{\overline{X}}$.

Setting $g(\mu) = 1/\mu$, using delta method,

τ^2
$= g'(\mu)\sigma^2 g'(\mu)$
$= \frac{-1}{\mu^2}\sigma^2\frac{-1}{\mu^2}$
$= \sigma^2/\mu^4 < \infty$ if $\mu \neq 0$.

That is,

$$Var(T_n) \approx \sigma^2/\mu^4 < \infty. \tag{2}$$

(a) However,

$$Var(T_n) = \infty \;\; \forall \;\; n, \tag{3}$$

as $E(T_n^2) = \int \frac{n}{x^2\sqrt{2\pi\sigma^2/n}} e^{-\frac{(x-\mu)^2}{2\sigma^2/n}} dx$

$\geq \int_{-c}^{c} \frac{n}{x^2} dx \min\{\frac{1}{\sqrt{2\pi\sigma^2/n}} e^{-\frac{(x-\mu)^2}{2\sigma^2/n}} : \; x \in [-c, c]\} = \infty$.

Another proof: $\lim_{x\to 0} \frac{1}{x} = \infty \Rightarrow \int \frac{1}{x} f_{\overline{X}}(x)dx = \infty$ **????**

Let $f_T(t) = ct^2\mathbf{1}(|t| < 1)$. Then $\int \frac{1}{x^2} f_T(x)dx < \infty$.
Question: What do Eq. (2) and (3) mean ?
$Var(T_n) \approx \sigma^2/\mu^4$ is only in the sense of Eq. (1).
It does not really mean that
$Var(T_n) \to \tau$.
This example serves as a counterexample.
(b) $\vdash$: In view of Eq. (1), $\mu \neq 0$. If we delete interval $(-\delta, \delta)$ from the sample space (of $\overline{X}$), then

$$V(T_n) < \infty \tag{4}$$

Q: Does Eq. (4) contradict Eq. (3) ?

$$V(T_n\mathbf{1}(|\overline{X}| \geq \delta)) < \infty \tag{5}$$

Exercise 10.4. Go over it after #10.5
$Y_i = \beta X_i + \epsilon_i$, $i = 1, ..., n$. Observe (X_i, Y_i)s. Assume independence, $X_i \sim N(\mu_X, \tau^2)$, $\epsilon_i \sim N(0, \sigma^2)$. Approximate the mean and variance of
(a) $\check{\beta} = \sum_i X_iY_i/\sum_j X_j^2$;
(b) $\tilde{\beta} = \frac{\sum_i Y_i}{\sum_j X_i}$;
(c) $\hat{\beta} = \overline{Y/X}$.
Sol. Q: What does it mean ? (1) $\hat{\sigma}^2_{\hat{\beta}} \approx V(\hat{\beta})$? Or (2) $\frac{\hat{\beta}-\beta}{\hat{\sigma}_{\hat{\beta}}} \xrightarrow{D} N(0, 1)$.
Ans.: Most of the time, (2).
Q: Can we write $E(\epsilon\epsilon) \approx E(\epsilon)E(\epsilon) = 0$??
$\sum_i X_iY_i/\sum_j X_j^2 = \beta + \frac{\sum_i X_i\epsilon_i}{\sum_j X_j^2} = \beta + \overline{X\epsilon}/\overline{X^2}$.
About the formulas in page 245.

$$E(X/Y) \approx \mu_X/\mu_Y \text{ and } Var(X/Y) \approx (\frac{\mu_X}{\mu_Y})^2(\frac{\sigma_X^2}{\mu_X^2} + \frac{\sigma_Y^2}{\mu_Y^2} - 2\frac{Cov(X,Y)}{\mu_X\mu_Y})$$

Counterexample. Let $X = 1$ and $Y \sim Exp(1)$.
$E(X/Y) = \int_0^\infty y^{-1}e^{-y}dy = \infty \approx \frac{1}{1}$ **???**
It really means: $E(\frac{X_n}{Y_n}) \approx \mu_{X_n}/\mu_{Y_n}$, in the sense that

$$\sqrt{n}[(X_n/Y_n) - (\mu_{X_n}/\mu_{Y_n})] \xrightarrow{D} N(0, \sigma^2), \textbf{ if the latter equation holds.}$$

For instance, let X_1, X_2, ... be i.i.d. from $Exp(1)$. $Y_n = 1/\overline{X} = n/\sum_{i=1}^n X_i$, then

$$E(Y_n) = \int_0^\infty \frac{n}{y}\frac{y^{n-1}e^{-y}}{\Gamma(n)}dy = n\frac{\Gamma(n-1)}{\Gamma(n)} = \frac{n}{n-1}$$

Thus $E(Y_n) \approx 1 = \frac{E(1)}{E(X)}$ if $n \approx \infty$, but
$E(Y_1) = \infty \not\approx 1$.
$E(Y_2) = 2 \not\approx 1$.
$E(Y_{20}) = 1\frac{1}{19} \approx 1$.

$$E(\frac{\sum_i X_iY_i}{\sum_j X_j^2}) = \beta + \sum_i E(\epsilon_i(\frac{X_i}{\sum_j X_j^2})) = \beta + \sum_i \underbrace{E(\epsilon_i)E(\frac{X_i}{\sum_j X_j^2})}_{as\ \epsilon \perp X} = \beta.$$

To approximate variance, use the delta method.
$V(g(\hat{\theta})) \approx \bigtriangledown g(\hat{\theta})\hat{\Sigma}_{\hat{\theta}} \bigtriangledown g(\hat{\theta})$,

or $V(g(\mathbf{W})) \approx \triangledown g(\mathbf{w})^t COV(\mathbf{W}) \triangledown g(\mathbf{w})|_{\mathbf{W}=E(\mathbf{W})}$.
$\sum_i X_iY_i/\sum_j X_j^2 = \beta + \overline{\epsilon X}/\overline{X^2} = g(\overline{\epsilon X}, \overline{X^2})$,
where $g(v,x) = \beta + v/x$.

$$\sqrt{n}(\begin{pmatrix} \overline{\epsilon X} \\ \overline{X^2} \end{pmatrix} - \begin{pmatrix} 0 \\ \mu_X^2 + \sigma_X^2 \end{pmatrix}) \xrightarrow{D} N(0,\Sigma),$$

where $\Sigma = COV(\begin{matrix} \epsilon X \\ X^2 \end{matrix})$.

$$Var(\textstyle\sum_i X_iY_i/\sum_j X_j^2) \approx (1/x, -v/x^2) COV(\begin{matrix} \overline{\epsilon X} \\ \overline{X^2} \end{matrix}) \begin{pmatrix} 1/x \\ -v/x^2 \end{pmatrix} \Big|_{v=0, x=\mu_X^2+\sigma_X^2}$$

$$= (1/x, 0) COV(\begin{matrix} \overline{\epsilon X} \\ \overline{X^2} \end{matrix}) \begin{pmatrix} 1/x \\ 0 \end{pmatrix} \Big|_{v=0, x=\mu_X^2+\sigma_X^2}$$

$= \frac{1}{nx^2} V(\epsilon X)\big|_{x=\mu_X^2+\sigma_X^2}$
$= \frac{1}{nx^2} E((\epsilon X)^2)\big|_{x=\mu_X^2+\sigma_X^2}$
$= \frac{1}{nx^2} \sigma^2 x\big|_{x=\mu_X^2+\sigma_X^2}$

Question: $Var(\sum_i X_iY_i/\sum_j X_j^2) = \frac{1}{nx^2}\sigma^2 x\big|_{x=\mu_X^2+\sigma_X^2}$?

Why approximation ?

(1) It is difficult to compute $V(\hat\beta)$, where $\hat\beta = \sum_i X_iY_i/\sum_j X_j^2$;

(2) Need in $\frac{\hat\beta-\beta}{\hat\sigma^2_{\hat\beta}}$.

$Var(\frac{\sum_i X_iY_i}{\sum_j X_j^2}) \approx \frac{1}{nx^2}\hat\sigma^2 x\big|_{x=\overline{X^2}}$ and $\hat\sigma^2 = \overline{\hat\epsilon^2}$, where $\hat\epsilon = Y - \hat\beta X$.

$$Var(\frac{\sum_i X_iY_i}{\sum_j X_j^2}) \approx \frac{1}{n}(\frac{1}{x}, -\frac{v}{x^2})(\overline{ZZ'} - \overline{Z}(\overline{Z})') \begin{pmatrix} \frac{1}{x} \\ -\frac{v}{x^2} \end{pmatrix} \Big|_{v=\overline{\epsilon X}, x=\overline{X^2}, Z=\begin{pmatrix} \epsilon X \\ X^2 \end{pmatrix}} ?$$

$$Var(\frac{\sum_i X_iY_i}{\sum_j X_j^2}) \approx \frac{1}{n}(\frac{1}{x}, -\frac{v}{x^2})(\overline{ZZ'} - \overline{Z}(\overline{Z})') \begin{pmatrix} \frac{1}{x} \\ -\frac{v}{x^2} \end{pmatrix} \Big|_{v=\overline{\hat\epsilon X}, x=\overline{X^2}, Z=\begin{pmatrix} \hat\epsilon X \\ X^2 \end{pmatrix}} ?$$

10.4.b. $\frac{\overline{Y}}{\overline{X}} = \frac{\beta\overline{X}}{\overline{X}} + \frac{\overline{\epsilon}}{\overline{X}} = \beta + \frac{\overline{\epsilon}}{\overline{X}} = g(\overline{\epsilon}, \overline{X})$. $g(v,x) = ?$

$Var(\overline{Y}/\overline{X}) = (1/x, 0) COV(\begin{matrix} \overline{\epsilon} \\ \overline{X} \end{matrix}) \begin{pmatrix} 1/x \\ 0 \end{pmatrix} \Big|_{v=0, x=\mu_X}$

$Var(\overline{Y}/\overline{X}) = Var(\overline{\epsilon}/\overline{X}) \approx \frac{\sigma^2}{n\mu_X^2} \approx \frac{\hat\sigma^2}{n(\overline{X})^2}$ using delta method.

Question: Notice in 10.4. a&b $V(\frac{\overline{Y}}{\overline{X}}) \approx \frac{V(\overline{Y})}{(E\overline{X})^2}$, is it true in general ?

Counterexample. See #10.5. $X_i \sim N(\mu, \sigma_X^2)$.

$$\sqrt{n}(\frac{1}{\overline{X}} - \frac{1}{\mu}) \xrightarrow{D} N(0, \tau^2). \tag{1}$$

$V(1/\overline{X}) \approx \frac{\sigma_X^2}{n\mu_X^4}$ in the sense of Eq. (1) if Eq. (1) holds. Note that $V(\overline{1})/(E(\overline{X}))^2 = 0$.
In fact, it is proved in #10.5 that

$$V(1/\overline{X}) = \infty. \tag{2}$$

Let $Y_i = 1$,

$V(\frac{\overline{Y}}{\overline{X}}) = \frac{V(\overline{Y})}{(\mu_X)^2}$??

10.4c. $V(\overline{Y/X}) = ?$

Sol. Notice that $Y/X = \frac{X\beta+\epsilon}{X} = \beta + \frac{\epsilon}{X}$,

$V(\overline{Y/X}) = V(\overline{\epsilon/X})$.

Standard approach: CLT. $\sqrt{n}(\overline{Z}-E(Z))\xrightarrow{D}N(0,\tau^2)$, where $Z=Y/X$ and $\tau^2=V(\overline{Y/X})=V(\overline{\epsilon/X})=V(\epsilon/X)/n$.

Let $Z=Y/X$, then $V(\overline{Y/X})\approx\hat\sigma^2_Z/n$ and $\hat\sigma^2_Z=\overline{Z^2}-(\overline{Z})^2$. **???**

Note that the CLT does not work here, as $\sigma^2_{Y/X}$ is not finite.

Question: $Var(\overline{\epsilon/X})\approx\frac{\sigma^2}{n\mu^2}$ => $Var(\epsilon/X)\approx\frac{\sigma^2}{\mu^2}$ **??**

Remark. $V(\frac{\epsilon}{X})=E((\frac{\epsilon}{X})^2)=E(\epsilon^2)E(\frac{1}{X^2})=$ **?**

Counterexample. See #10.5. $X_i\sim N(\mu,\sigma^2_X)$ and Eq. (2).

Also, it can be checked by numerical calculation in R

mu=2.0

e=rnorm(100000)

x=rnorm(100000)+mu

var(e/x)/n

[1] 7.725 # v(e/x)= 5.725×100000

1/(mu*mu+1) # $=\frac{\sigma_e^2}{\mu_X^2}$

[1] 0.2

$\frac{\sigma_e^2}{\mu_X^2}=0.2\neq 7.725*n\approx V(\frac{e}{X})=E(\frac{e^2}{X^2})$.

$E(\epsilon^2)E(\frac{1}{X^2})=\infty$ **Why ?**

Ans. See Exercise 10.5a.

But if $X\sim N(\mu,1)$, then $E(1/\overline{X})$ does not exist, as $E(1/|X|)=\infty$ (see #10.5).

10.31.c. Prove: $\sqrt{n}(\hat p_1-\hat p_2)\xrightarrow{D}N(0,\tau^2)$.

Two independent bin(n_1,p_1) and $bin(n_2,p_2)$. $n=n_1+n_2$. $\min\{n_1,n_2\}\to\infty$.

Under H_o: $p_1=p_2=p$.

Let $Y_n=(\hat p_1,\hat p_2)'$, $q_i=1-p_i$, $\mathbf{p}=(p_1,p_2)'$, $\Sigma_{Y_n}=\begin{pmatrix}p_1q_1/n_1 & 0\\ 0 & p_2q_2/n_2\end{pmatrix}$ and $g(Y_n)=\hat p_1-\hat p_2$, Notice that

$\hat\Sigma_{Y_n}^{-1/2}(Y_n-\mathbf{p})\xrightarrow{D}N(0,I_{2\times 2})$ **Why ??**

Since $\hat v^2=(\bigtriangledown g(\hat\theta))^t\hat\Sigma_{Y_n}\bigtriangledown g(\hat\theta)=\frac{\hat p\hat q}{n_1}+\frac{\hat p\hat q}{n_2}$, where $\hat q_i=1-\hat p_i$.

$$\frac{g(Y_n)-g(\mathbf{p})}{\sqrt{\hat{v^2}}}=\frac{\hat p_1-\hat p_2}{\sqrt{\frac{1}{n_1}\hat p_1(1-\hat p_1)+\frac{1}{n_2}\hat p_2(1-\hat p_2)}}\xrightarrow{D}N(0,1)\quad \text{Why ??}$$

Homework Solutions, week 13

10.31(a,b). a. Two independent $bin(n_1,p_1)$ and $bin(n_2,p_2)$. H_0: $p_1=p_2$. Show a test has RR $\frac{(\hat p_1-\hat p_2)^2}{(\frac{1}{n_1}+\frac{1}{n_2})\hat p(1-\hat p)}>\chi^2_{1,\alpha}$. That is,

$$\frac{(\hat p_1-\hat p_2)^2}{(\frac{1}{n_1}+\frac{1}{n_2})\hat p(1-\hat p)}\xrightarrow{D}\chi^2_1.$$

Proof.

Proof 1. MLE: With two independent binomials, thus AS1-AS7 are satisfied. The MLE of $)p_1,p_2)$ is $(\hat p_1,\hat p_2)$.

$\hat p_1-\hat p_2$ is the MLE of p_1-p_2.

$\hat p_1-\hat p_2$ has mean 0 and variance $p(1-p)/n_1+p(1-p)/n_2$. Thus

$$\frac{\hat p_1-\hat p_2}{\sqrt{p(1-p)/n_1+p(1-p)/n_2}}\xrightarrow{D}N(0,1).$$

$$\frac{(\hat p_1-\hat p_2)^2}{(\frac{1}{n_1}+\frac{1}{n_2})p(1-p)}\xrightarrow{D}\chi^2_1.$$

$\hat p=\frac{S_1+S_2}{n}\xrightarrow{a.s.}p$ if $p_1=p_2=p$, where $n=n_1+n_2$.

$$\frac{\hat p(1-\hat p)}{p(1-p)}\frac{(\hat p_1-\hat p_2)^2}{(\frac{1}{n_1}+\frac{1}{n_2})\hat p(1-\hat p)}\xrightarrow{D}\chi^2_1$$

by Slutsky's theorem.

Proof 2. Suppose that $M_{X_n}(t) = E(e^{X_n t})$ and $M_X(t) = E(e^{Xt})$ exist.

$X_n \xrightarrow{D} X$ iff $M_{X_n}(t) \to M_X(t)$ for $t \in [0, c)$, $c > 0$.

Let $X_{n_1} = \frac{\hat{p}_1 - p}{\sqrt{p(1-p)/n_1}}$ and $Y_{n_2} = \frac{\hat{p}_2 - p}{\sqrt{p(1-p)/n_2}}$, then

$$X_{n_1} \xrightarrow{D} N(0,1)$$

$$Y_{n_2} \xrightarrow{D} N(0,1)$$

$M_{X_{n_1}}(t) \to exp(t^2/2)$, and $M_{Y_{n_2}}(t) \to exp(t^2/2)$.

$$U_n = \frac{\sqrt{\frac{1}{n_1}}X_{n_1} - \sqrt{\frac{1}{n_2}}Y_{n_2}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \ n = n_1 + n_2.$$

$$M_{U_n}(t) = E(\exp(\frac{\sqrt{\frac{1}{n_1}}X_{n_1} - \sqrt{\frac{1}{n_2}}Y_{n_2}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}t))$$

$$= E(\exp(\frac{\sqrt{\frac{1}{n_1}}X_{n_1}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}t))E(\exp(-\frac{\sqrt{\frac{1}{n_2}}Y_{n_2}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}t))$$

$$= E(\exp(X_{n_1}\frac{\sqrt{\frac{1}{n_1}}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}t))E(\exp(Y_{n_2}\frac{-\sqrt{\frac{1}{n_2}}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}t))$$

$$\to \exp((\frac{\sqrt{\frac{1}{n_1}}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}t)^2/2)\exp((-\frac{\sqrt{\frac{1}{n_2}}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}t)^2/2) = \exp(t^2/2).$$

That is, $U_n \xrightarrow{D} N(0,1)$. The rest is the same as in the MLE approach.

Proof 3. Assume $n_i/n = n_i/(n_1 + n_2) \to p_{i*} \in (0,1)$. Then

$$\sqrt{n}[(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)] = \sqrt{n/n_1}\sqrt{n_1}(\hat{p}_1 - p_1) - \sqrt{n/n_2}\sqrt{n_2}(\hat{p}_2 - p_2).$$

Let $g(x,y) = x - y$. Since

$$\sqrt{n}(\hat{p}_i - p_i) = \sqrt{n/n_i}\sqrt{n_i}(\hat{p}_i - p_i) \xrightarrow{D} N(0, p_i(1-p_i)/p_{i*}), \ i = 1, 2,$$

by Slutsky's theorem. Then by the delta method,

$$\sqrt{n}[(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)] = \sqrt{n}(g(\hat{p}_1, \hat{p}_2) - g(p_1, p_2))$$

$$\xrightarrow{D} N(0, p_1(1-p_1)/p_{1*} + p_2(1-p_2)/p_{2*}).$$

Thus

$$\sqrt{n}\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{p_1(1-p_1)/p_{1*} + p_2(1-p_2)/p_{2*}}} \xrightarrow{D} N(0,1).$$

The conclusion then follows from Slutsky' theorem and the assumption $p_1 = p_2 = p$, as $\hat{p}$ is the consistent estimator of $p_1 = p_2 = p$ and n_i/n is the consistent estimator of p_{i*}.

10.47. Shortest CI with form $(\frac{\chi^2_{d,1-\alpha-c}}{2\sum_i x_i}, \frac{\chi^2_{d,c}}{2\sum_i x_i})$, $c \in [0, \alpha]$?

Sol. Write $(\frac{t}{2\sum x}, \frac{b}{2\sum x}) = (\frac{\chi^2_{d,1-\alpha-c}}{2\sum_i x_i}, \frac{\chi^2_{d,c}}{2\sum_i x_i})$.
It suffices to minimize

$$l = b - t, \tag{1}$$

where $b = b(t)$ is determined by

$$F(b) - F(t) = 1 - \alpha. \tag{2}$$

Eq. (1) => $l'_t = b'_t - 1 = 0$ => $b'_t = 1$.
Eq. (2) => $F'(b)b'_t - F'(t)t'_t = 0$ => $f(b)b'_t - f(t) = 0$.
Thus $f(b) = f(t)$.
In other word, if $f(t) = f(b)$ and $F(b) - F(t) = 1 - \alpha$, then we have the shortest CI of the form given.

4. **Addtional A3.** 1. As in Example 10.3.4, set
H_0: $p_1 = p_2 = p_5$, $p_3 = 0.5$ v.s. H_1: H_0 is not true.
a. Derive the likelihood ratio test.
b. Give an estimate of $P(H_o|H_1)$ when $p_1 = p_2 = p_5$, $p_3 = 0$, $n = 100$, using simulation. Present the program.
m=1000
n=100
x=rmultinom(m,n, prob=c(1/8,1/8,0,5/8,1/8))
reject H_o if $-2log\lambda > \chi^2_{3,\alpha}$ **Why not** $\chi^2_{1,\alpha}$ **?**

Remark. In compute $-2\log(\lambda)$, use 0^0 rather $log(0^0) = 0 \log 0$ in R codes.
$P(H_0|H_1)$ can be estimated by frequency of acceptance **Why ?**
x=sample(c(1,2,3,4,5),100,replace=T,prob=c(1/8,1/8,0,5/8,1/8))
Can we estimate $P(H_1|H_o)$ with the previous codes ?

#10.41. Let X_1, ..., X_n be i.i.d. negative binomial(r,p).

a. Calculate Wilks' approximation (10.4.1) *i.e.*, the score function form, and show how to construct some approximate intervals with this expression.

b. Find an approximate $1-\alpha$ confidence interval for the mean of the negative binomial distribution. Show how to incorporate the continuity correction into your interval.

Sol. (a) Assuming r is known,
$S(p) = \frac{\partial}{\partial p} \log p^{nr}(1-p)^{n\overline{X}} = \frac{nr}{p} - \frac{n\overline{X}}{1-p}$,
$-\frac{\partial S(p)}{\partial p} = \frac{nr}{p^2} + \frac{n\overline{X}}{(1-p)^2}$.
$E(-\frac{\partial S(p)}{\partial p}) = \frac{nr}{p^2} + \frac{nrq/p}{(1-p)^2} = \frac{nr}{p^2} + \frac{nr}{qp} = \frac{nr}{p^2q}$.
It is easy to derive the CI is

$$\{p : \left|\frac{\frac{nr}{p} - \frac{n\overline{X}}{1-p}}{\sqrt{\frac{nr}{p^2q}}}\right| \le z_{\alpha/2}\} \tag{3}$$

Solving $\frac{(\frac{nr}{p} - \frac{n\overline{X}}{1-p})^2}{\frac{nr}{p^2p}} = z^2_{\alpha/2}$ yields

$$\frac{2\overline{X}r + z^2_{\alpha/2}\frac{r}{n} \pm \sqrt{(\frac{rz^2_{\alpha/2}}{n})^2 + \frac{4\overline{X}r}{n}z^2_{\alpha/2}(r+\overline{X})}}{2(\overline{X})^2}$$

Sol. (b) In both (a) and (b), if r is a parameter, estimate it by the MLE $\hat{r}$.
Recall $\mathcal{L}(p,r) = p^{nr}(1-p)^{\sum_{i=1}^n X_i} \prod_{j=1}^n \binom{r+X_j-1}{X_j}$.
The MLE of p: $\hat{p} = \frac{nr}{nr+n\overline{X}}$
due to $\mathcal{L}' = 0$ or $\frac{nr}{p} - \frac{n\overline{X}}{1-p} = 0$,

$$\hat{r} = \text{argmax}_r \mathcal{L}(\hat{p}, r) \tag{4}$$

Since $E(X) = rq/p$ and $V(X) = rq/p^2$, Solving p through $\theta = r(1-p)/p$, Eq.(3) yields

$$\frac{(\theta - \overline{X})^2 nr}{r\theta + \theta^2} \le z^2_{\alpha/2}$$

The $(1-\alpha)$ CI is $\frac{2rn\overline{X}+z_{\alpha/2}r\pm\sqrt{(2nr\overline{X}+z_{\alpha/2}r)^2-4(n^2r-z_{\alpha/2}n)r(\overline{X})^2}}{2(nr-z_{\alpha/2})}$.

Write the $(1-\alpha)$ CI as $[a\overline{X}+b, c\overline{X}+d]$.

Continuity correction:

$[a(\overline{X}-\frac{0.5}{n})+b, c(\overline{X}+\frac{0.5}{n})+d]$ **Why ?**

10.41.c. Use the data in #9.23 to construct an approximate 90% CI for the mean of the negative binomial. The data: x=c(155,104,66,50,36,40,30,35,42)

Sol. There are 9 data in #9.23. Thus it may not be appropriate to use the approximation.
However, if $Y_1 \sim bin(25,p)$, $P(Y_1 \le t) \approx N(25p, 25pq)$. **Why ??**
Since each X_i is large (that is, each corresponds to an experiment with more than 20 independent Bernoulli trials), thus there are large number of Bernoulli trials and it is fine to use normal approximation.
There are several ways:
(1) MLE $\hat\mu \pm 1.64\hat\sigma_{\hat\mu}/\sqrt{n}$,
(2) MME $\overline{X} \pm 1.64(\overline{X^2}-(\overline{X})^2)/\sqrt{n}$.
(3) $\overline{X} \pm 1.64\hat\sigma/\sqrt{n}$, where $\hat\sigma^2 = \overline{X}/\hat p$,
as $E(X)=rq/p$ and $V(X)=rq/p^2$,

The MLE of p: $\hat p = \frac{nr}{nr+n\overline{X}}$
due to $\mathcal{L}'=0$ or $\frac{nr}{p}-\frac{n\overline{X}}{1-p}=0$,
where r can be estimated by its MLE.

How to find the MLE of r ?
The likelihood $\mathcal{L}(p,r) = p^{nr}(1-p)^{\sum_{i=1}^n X_i}\prod_{j=1}^n\binom{r+X_j-1}{X_j}$.
Fixed r, $\mathcal{L}$ is maximized by
$\hat p = \frac{nr}{nr+n\overline{X}} = \frac{1}{1+\overline{X}/r}$ $(=g(\overline{X}))$,
due to $\mathcal{L}'=0$ or $\frac{nr}{p}-\frac{n\overline{X}}{1-p}=0$, **after checking !**
It is the MLE of p if r is given.
Otherwise, the MLE of r is
$\hat r = \text{argmax}_r\mathcal{L}(\hat p, r)$, where

$$\mathcal{L} = (\frac{nr}{nr+n\overline{X}})^{nr}(1-\frac{nr}{nr+n\overline{X}})^{\sum_{i=1}^n X_i}\prod_{j=1}^n\binom{r+X_j-1}{X_j}$$

Since $E(X)=rq/p$ and $V(X)=rq/p^2$,
the MLE $\hat\mu = \hat r\hat q/\hat p = \hat r(\frac{1}{\hat p}-1) = \hat r h(\overline{X})$.
In fact, $\hat\mu = \hat r(1+\overline{X}/\hat r - 1) = \overline{X}$.
$\sigma^2_{\hat\mu} = \hat r\hat q/\hat p^2 = \overline{X}/\hat p$.
So the CI of μ is $\overline{X} \pm z_{\alpha/2}\sqrt{\frac{\overline{X}}{\hat p n}}$.

$$\begin{aligned}\mathcal{L}(r+1)/\mathcal{L}(r) =& \frac{(\frac{1}{1+\overline{X}/(r+1)})^{n(r+1)}(\frac{\overline{X}/(r+1)}{1+\overline{X}/(r+1)})^{n\overline{X}}\prod_{j=1}^n\binom{(r+1)+X_j-1}{X_j}}{(\frac{1}{1+\overline{X}/r})^{nr}(\frac{\overline{X}/r}{1+\overline{X}/r})^{n\overline{X}}\prod_{j=1}^n\binom{r+X_j-1}{X_j}}\\
=&\frac{\frac{(\overline{X}/(r+1))^{n\overline{X}}}{(1+\overline{X}/(r+1))^{n(r+1)+n\overline{X}}}}{\frac{(\overline{X}/r)^{n\overline{X}}}{(1+\overline{X}/r)^{nr+n\overline{X}}}}r^{-n}\prod_{i=1}^n(r+X_i)\\
=&\frac{\frac{(\overline{X})^{n\overline{X}}}{(r+1+\overline{X})^{n(r+1)+n\overline{X}}}}{\frac{(\overline{X})^{n\overline{X}}}{(r+\overline{X})^{nr+n\overline{X}}}}r^{-n}\prod_{i=1}^n(r+X_i)\\
=&(\frac{r+\overline{X}}{r+1+\overline{X}})^{nr+n\overline{X}}(r+1+\overline{X})^{-n}r^{-n}\prod_{i=1}^n(r+X_i)\end{aligned}$$

No easy way to solve.

$\hat{\sigma}^2_{\hat{\mu}} = \hat{r}|h'(\hat{\mu})|\sqrt{\frac{\hat{r}\hat{q}}{n\hat{p}}}$

Eq.(3) yields

a. $\frac{\overline{X}-\mu_X}{\sigma_X/\sqrt{n}} = \frac{\overline{X}-\frac{rq}{p}}{\sqrt{\frac{rq}{p^2}}} => \{p: \; |\frac{\overline{X}-\frac{rq}{p}}{\sqrt{\frac{rq}{np^2}}}| \leq z_{\alpha/2}\}.$

$(\overline{X} - \frac{rq}{p})^2 = \frac{rq}{np^2} z^2_{\alpha/2}\}.$

$(p\overline{X} - rq)^2 = \frac{rq}{n} z^2_{\alpha/2}\}.$