MATH 448, MATHEMATICAL STATISTICS

Textbook: Mathematical statistics with applications (7th Ed.),

by Dennis D. Wackerly, W. Mendenhall III, R. L. Scheaffer

Chapters to be covered: 8-10, 16.
Classroom CW 213 MWF 10:20am-11:50am
Office: WH 132

Office hours: 7:00-8:00pm Monday and Tuesday through zoom
https://binghamton.zoom.us/j/8265526594?pwd=d3l6OGx1cmZ4M3cxZEJwVGd1RGcrUT09

Meeting ID: 826 552 6594

Passcode: 031320

Exams: 3 tests + final,

Feb 19 (M), Mar 18 (M), Apr. 15. (M)

Final: May 6 (M) 8:05pm-10:05pm LH 009 closed book

Homework: Due Wednesday in class, no late homework.

HW Solution: https://usermanual.wiki/Document/SolutionManualMathematicalStatistics WithApplications7thEditionWackerly.313163145/help

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 !!! It is a final exam for math 447. It is the format of the exams for Math 448. First do the exam, then grade it yourself carefully and hand in. The solution is on my website below. https://brainly.com/textbook-solutions/b-mathematical-statistics-applications-7th-edition-college-math-9780495110811

Quizzes: once a week, at the beginning of Friday class.

Grading Policy:

1. 10% hw +10% quiz +45% tests +35% final
2. Correction: If you make correction **at the next class** after I distribute the test in class, you can get 40% of the missing grades back. The correction should be on a different paper for the whole problem, not the incorrect statement. **No partial credits for correction. Can not ask me for how to make correction.**
3. A or A- = 85 +; C = 60 +.
   10+10+45*(0.3+0.4*0.7)+35*0.3=56

Syllabus: Prerequisites: MATH 447 with a grade of C or better.

Summarizing data by graphical and numerical methods, point estimation, consistency, bias, mean square error, confidence intervals, relative efficiency, sufficient statistics, minimum variance unbiased estimators, the method of moments, the method of maximum likelihood, hypothesis testing, type I and type II errors, lemma of Neyman-Pearson, Bayesian statistics.
Quiz on this Friday: 447 formula 1-15.

4/1(M) 4/24(W) No class. 4/22(M) No class after 1pm 4/25(Th) meet M class

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 fishes.

Catch m, say $m = 10$,

tag them and put them back.

Re-catch k fish, say $k = 10$,

X of them are tagged, say $X = 3$.

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

Answer: 1. $f(x; N) = P(X = x) = \frac{\binom{m}{x}\binom{N-m}{k-x}}{\binom{N}{k}}$, $x \in \{0, 1, ..., k \wedge m\}$, $k \vee m \leq N$.

2. Many methods to estimate N: MME, MLE, Bayes estimator, etc. *e.g.*

MME: Solve $\overline{X} = E(X) = km/N \Rightarrow \hat{N} = km/X = 33\frac{1}{3}$.

MLE: $\check{N} = 33$ (from google).

Or use R: (in a department computer, type)

```
R
> m=10 # of tagged fishes
> k=10 # recaptured fishes
> n=0:29 # untagged fishes in the pond
> (a=dhyper(3, m, n, k, log = FALSE))          phyper()     qhyper()     rhyper()
  [1] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
  [7] 0.00000000 0.00617030 0.02193885 0.04676438 0.07794064 0.11227163
  [13] 0.14697377 0.17998962 0.20998789 0.23623637 0.25844663 0.27663361
  [19] 0.29100419 0.30187504 0.30961542 0.31460922 0.31723096 0.31783178
  [25] 0.31673201 0.31421827 0.31054320 0.30592702 0.30055988 0.29460473
> n[a==max(a)]+10
  [1] 33                  # MLE
> n=0:10000
> a=dhyper(3, m, n, k, log = FALSE)
> n[a==max(a)]+10
  [1] 33
```

Q: Properties of these estimators ?

What is the good (or possibly best) estimator ?

What is the meaning of a good or best estimator ?

Typically, statistics deals with such problems:

Suppose that X_1, ..., X_n are i.i.d. from X, with cdf $F(x; \theta)$, where θ is unknown in Θ, try to find out:

1. $\theta = ?$ or $P(X \leq x) = ?$ (this is called *point estimation*).

   What is θ in the capture-recapture problem ?

2. $(a, b) = ?$ such that it is likely that $a \leq \theta \leq b$ (this is called *interval estimation*);
3. $\theta = \theta_o$? where θ_o is given. (This is called hypothesis testing).

In 448, we shall learn these concepts.

Chapter 8. Estimation

§8.1. Introduction.

Def. Denote $\mathbf{X} = (X_1, ..., X_n)$, where X_1, ..., X_n, i.i.d. from $X \sim F(x;\theta)$ $(= P(X \leq x|\theta))$.

θ is called the parameter of the distribution.

We call $\mathbf{X}$ a data set or observations from X.

The sample mean is $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

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) ∝ e^{-x/μ} = e^{-ρx}, x > 0).
> mean(x)
[1] 0.3559676
```

Remark. In the example, we observe $X_1 = 0.3163466$, $X_2 = 0.4865695$, $X_3 = -0.2163855$. We can say X_1, X_2 and X_3 are r.v.s,

but cannot say 0.3163466, 0.4865695, -0.2163855 are r.v.s.

They are numbers.

Def. A statistic is a function (or a formula) of a random vector or random variable, say $\mathbf{X}$, but does not depend on the parameter θ.

An estimator is a statistic used to guess the parameter θ.

An estimate is a value of the estimator.

Remark. Most of the time we let X (or Y, or Z) be r.v., x or y or t be value of X, say $X = x$ or $X = y$ or $X = t$. *e.g.*,

```
> (x=rnorm(3,0,1))
[1] 0.3163466 0.4865695 -0.2163855
```

we observe $X_1 = 0.3163466$ (or $X_1 = x$), $X_2 = 0.4865695$ (or $X_2 = y$),

then x and y are numbers, not r.v.s.

$E(X_1) = 0$? $E(0.3163466) = 0$? $E(x) = 0$?

Ex 1. Suppose that X_1, ..., X_n are i.i.d. from $N(\mu, 1)$. Let

(a) $X_1 + X_2$, (b) $X_1 + \mu$, (c) 2, (d) $\overline{X}$, (e) $X_2 + X_3^2 + 5$.

Which of them is a statistic ?

Ex. 2. Suppose that X_1, ..., X_n are i.i.d. from bin(1,p).

An estimator of p is $\overline{X}$, denoted by $\hat{p} = \overline{X}$.

If $(X_1, X_2, X_3) = (1, 0, 1)$, $n = 3$, then $\frac{2}{3}$ is an estimate of p, denoted by $\hat{p} = 2/3$.

Is $\overline{X}$ an estimator, or an estimate ?

Is $2/3$ an estimator, or an estimate ?

Is $\hat{p}$ an estimator, or an estimate ?

Remark. 1. Given a parameter θ, one can use $\hat{\theta}$ or $\tilde{\theta}$ or $\check{\theta}$ to denote its estimator.

2. An estimator $\hat{\theta}$ is a r.v. *e.g.* $\hat{\mu} = \overline{X}$, where X_1, ..., X_n are i.i.d..

Q: (1) $E(\hat{\mu}) = ?$ (2) $V(\hat{\mu}) = ?$ (3) $P(2 \le \hat{\mu} \le 5) = ?$ or $P(a \le \hat{\mu} \le b) = ?$

Possible Ans:

(1) $E(\hat{\mu}) = E(\overline{X}) = E(X)$ or $= \mu_X$? Yes, No, DNK.

How about $X \sim$ Cauchy distribution ?

$E(\overline{X}) = E(X)$ if it exists.

(2) $V(\hat{\mu}) = V(\sum_{i=1}^n X_i/n) = \sum_{i=1}^n V(X_i)/n^2 = \sigma_X^2/n$? Yes, No, DNK.

$$(3)\ P(a \le \hat{\mu} \le b) = \begin{array}{lll} (a) & F_{\hat{\mu}}(b) - F_{\hat{\mu}}(a) & ? \\ (b) & F_{\hat{\mu}}(b-) - F_{\hat{\mu}}(a) & ? \\ (c) & F_{\hat{\mu}}(b) - F_{\hat{\mu}}(a-) & ? \\ (c) & F_{\hat{\mu}}(b-) - F_{\hat{\mu}}(a-) & ? \end{array}$$

Formula 17. A cdf $F(t)$ ($= \underline{P(X \le t)}$,) satisfying

(1) $F(-\infty) = \underline{0}$, and $F(\infty) = \underline{1}$, (2) $F(x+) = \underline{F(x)}$, (3) $F(x)\ \underline{\uparrow}$.

Moreover, $F(b) - F(a) = \mathrm{P}(\underline{a < X \le b})$

Remark. Recall $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$. $\overline{X^2} = ?$ $\overline{1/X} = ?$

§8.2. The Bias and mean square error of point estimators

Def. Let $\hat{\theta}$ be a point estimator of a parameter θ.

If $E(\hat{\theta}) = \theta$, $\hat{\theta}$ is unbiased.

O.W. $\hat{\theta}$ is called a biased estimator, and

the bias of $\hat{\theta}$ is denoted by $B(\hat{\theta})$, $B(\hat{\theta}) = E(\hat{\theta}) - \theta$.

The mean square error of $\hat{\theta}$ is $\mathrm{MSE}(\hat{\theta}) = E((\hat{\theta} - \theta)^2)$

Formula: $MSE(\hat{\theta}) = (B(\hat{\theta}))^2 + V(\hat{\theta})$.

Proof. $\vdash$: $\mathrm{MSE}(\hat{\theta})(= E((\hat{\theta} - \theta)^2)) = (B(\hat{\theta}))^2 + V(\hat{\theta})$.

$$\begin{aligned}
& MSE(\hat{\theta}) \\
=& E((\hat{\theta} - \theta)^2) \\
=& E((\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2) \\
=& E[(\hat{\theta} - E(\hat{\theta}))^2 + 2(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta) + (E(\hat{\theta}) - \theta)^2] \qquad (a+b)^2 = a^2 + 2ab + b^2 \\
=& E(\hat{\theta} - E(\hat{\theta}))^2 + 2E((\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)) + (E((\hat{\theta}) - \theta)^2) \quad E(aX + bY) = aE(X) + bE(Y) \\
=& E(\hat{\theta} - E(\hat{\theta}))^2 + 2(E(\hat{\theta}) - E(\hat{\theta}))(E(\hat{\theta}) - \theta) + (E(\hat{\theta}) - \theta)^2 \quad \textbf{Why ?}
\end{aligned}$$

$=E((\hat\theta - E(\hat\theta))^2)+$? $+ (E(\hat\theta)-\theta)^2$

$=V(\hat\theta) + (B(\hat\theta))^2$

Ex. 1. Suppose that X_1, X_2 and X_3 are i.i.d. from $N(\mu, \sigma^2)$, and their observations are 1.408, 0.015, 0.050, thus $\overline{X} = 0.491$.
Let $T_1 = X_1 = 1.408$, $T_2 = \overline{X}$, $T_3 = \mu$, $T_4 = \overline{X} + 2 = 2.491$.

(A) Which of them is an estimate of μ ?

(B) Are those estimators unbiased ?

(C) MSE of those estimators = ?

Sol. (a) $T_1 = X_1$ is an estimator, 1.408 is an estimate.

$E(T_1) = E(X_1) = \mu$, thus T_1 is an unbiased estimator. $E(1.408) = ?$

bias= $B(T_1) = 0$,

MSE$(T_1) = V(X_1) = \sigma^2$.

(b) $T_2 = \overline{X}$ is an estimator,

$E(T_2) = E(\overline{X}) = \mu$. thus T_2 is an unbiased estimator.

bias= $B(T_2) = 0$,

MSE$(T_2) = V(X_1) = \sigma^2/3$.

(c) T_3 is not an estimator.

$E(T_3) = \mu$??

bias= $B(T_3) =$??

(d) T_4 is an estimator,

$E(T_4) = E(\overline{X}) + 2 = \mu + 2$, thus T_4 is an biased estimator.

bias= $B(T_4) = 2$,

MSE$(T_4) = V(T_4) + (B(T_4))^2 = V(\overline{X}+2) + (B(T_4))^2 = V(\overline{X}) + (B(T_4))^2$???

$= \sigma^2/3 + 2^2$.

Q: In Ex.1 above, do we know μ ?

Quiz this Friday: 447: 16-42. 448: [1].

Ex.2. Suppose that X_1, ..., X_n are i.i.d. from $N(\mu, \sigma^2)$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$.

a. Is S an unbiased estimator of σ ?

b. Find an unbiased estimator of σ.

Sol. Recall 447 formulae [23], [24] and [41]:

[23] $X \sim \mathcal{G}(\alpha, \beta)$. $f(x) = \underline{\frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}}$, if $x > \underline{0}$, $\mu = \underline{\alpha\beta}$, $\sigma^2 = \underline{\alpha\beta^2}$, $\Gamma(\alpha+1) = \underline{\alpha\Gamma(\alpha)}$

[24] $Exp(\lambda) = \underline{\mathcal{G}(1,\lambda)}$, $\chi^2(\nu) = \underline{\mathcal{G}(\frac{\nu}{2}, 2)}$

[41] If $Y \sim N(\mu, \sigma^2)$, $\frac{\overline{Y} - \mu_{\overline{Y}}}{\sigma_{\overline{Y}}} \sim \underline{N(0,1)}$, $\frac{(n-1)S^2}{\sigma^2} \sim \underline{\chi^2(n-1)}$, $\sqrt{n}\frac{\overline{Y}-\mu}{S} \sim \underline{t_{n-1}}$,

where $\mu_{\overline{Y}} = \underline{\mu}$, $\sigma^2_{\overline{Y}} = \underline{\sigma^2/n}$

[41]=> $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$; => $\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \overline{X})^2 \sim \chi^2(n-1)$.

[23]=> density of $\mathcal{G}(\alpha, \beta)$ is $f(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}$, $x > 0$.

[24]=> $f_{\chi^2(n-1)}(x) = \frac{x^{\frac{n-1}{2}-1}e^{-x/\beta}}{\Gamma(\frac{n-1}{2})\beta^{\frac{n-1}{2}}}$, $x > 0$.

$$
\begin{aligned}
E(S) =& E(\sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2}) \qquad = E(\sqrt{\tfrac{\sigma^2}{n-1}\tfrac{1}{\sigma^2}\textstyle\sum_{i=1}^n (X_i - \overline{X})^2}) \\
=& \sqrt{\frac{1}{n-1}}\sigma E(\sqrt{\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \overline{X})^2}) \\
=& \sqrt{\frac{1}{n-1}}\sigma E(\sqrt{Y}) \qquad Y \sim \chi^2(n-1) = G(\tfrac{n-1}{2}, 2) \\
=& \sqrt{\frac{1}{n-1}}\sigma \int_0^\infty \sqrt{y}\frac{y^{\alpha-1}e^{-y/\beta}}{\Gamma(\alpha)\beta^\alpha}dy \qquad (\alpha, \beta) =? \\
=& \sqrt{\frac{1}{n-1}}\sigma \int_0^\infty \sqrt{y}\frac{y^{\frac{n-1}{2}-1}e^{-y/2}}{\Gamma(\frac{n-1}{2})2^{\frac{n-1}{2}}}dy \\
=& \sqrt{\frac{1}{n-1}}\sigma \int_0^\infty \frac{y^{\frac{n}{2}-1}e^{-y/2}}{\Gamma(\frac{n-1}{2})2^{\frac{n-1}{2}}}dy \qquad \textit{why ?} \\
=& \sqrt{\frac{1}{n-1}}\sigma \int_0^\infty \frac{y^{\frac{n}{2}-1}e^{-y/2}}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}dy\frac{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}{\Gamma(\frac{n-1}{2})2^{\frac{n-1}{2}}} \qquad \textbf{why do this ??} \\
=& \sqrt{\frac{1}{n-1}}\sigma\frac{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}{\Gamma(\frac{n-1}{2})2^{\frac{n-1}{2}}} \qquad [23]\ \ \Gamma(\alpha+1) = \alpha\Gamma(?) \\
=& \sigma\sqrt{\frac{1}{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}2^{1/2} = \sigma\ ? \qquad \textbf{Is } S \textbf{ unbiased ?} \\
\textit{Let}\ \ \tilde{\sigma} =& \frac{1}{\sqrt{\frac{2}{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}}S. \text{ Then } \tilde{\sigma} \text{ is unbiased.} \\
\textit{Let}\ \ \sigma =& \frac{1}{\sqrt{\frac{2}{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}}S. \text{ Is it unbiased ???}
\end{aligned}
$$

Remark. The above statement may not be true if X_i's are not normal.

§**8.3 Some common unbiased point estimators**.

Ex.1. Suppose that X_1, ..., X_n are i.i.d. from X with mean μ_X,
and Y_1, ..., Y_n are i.i.d. from Y with mean μ_Y.
Unbiased estimators of μ_X, μ_Y and $\mu_X - \mu_Y$?

Sol. The unbiased estimator of μ_X is $\hat{\mu}_X = \overline{X}$,
The unbiased estimator of μ_Y is $\hat{\mu}_Y = \overline{Y}$,

The unbiased estimator of $\mu_X - \mu_Y$ is $\overline{X} - \overline{Y}$.

Reason: $E(\overline{X}) = E(X)$

$E(\overline{Y}) = E(Y)$

$E(\overline{X} - \overline{Y}) = E(\overline{X}) - E(\overline{Y}) = E(X) - E(Y)$.

Ex.2. Let $X \sim bin(n,p)$ and $Y \sim bin(m,\theta)$. Find the unbiased estimators of p, θ and $p-\theta$.

Sol. The unbiased estimators are $\hat{p} = X/n$, $\hat{\theta} = Y/m$ and $\hat{p} - \hat{\theta} = X/n + Y/m$.

Reason: $E(\hat{p}) = E(X/n) = np/n = p$.

$E(\hat{\theta}) = E(Y/m) = m\theta/m = \theta$.

$E(\hat{p} - \hat{\theta}) = E(\hat{p}) - E(\hat{\theta}) = p - \theta$.

Ex.3. Let X_1, ..., X_n be i.i.d. from X, with mean μ and variance σ^2.

Let $\hat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$ and let $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ $\quad (\hat{\sigma^2} = \frac{n-1}{n}S^2)$.

Is $\hat{\sigma^2}$ unbiased estimator of σ^2 ? Is S^2 unbiased estimator of σ^2 ?

Sol. $\hat{\sigma^2}$ is biased but S^2 is unbiased. The reason is as follows.

$$\begin{aligned}
\hat{\sigma^2} =& \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 \\
=& \frac{1}{n}\sum_{i=1}^n (X_i^2 - 2X_i\overline{X} + (\overline{X})^2) \\
=& \frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{2}{n}\sum_{i=1}^n X_i\overline{X} + (\overline{X})^2 && ??? \\
=& \frac{1}{n}\sum_{i=1}^n X_i^2 - 2\overline{X}\cdot\overline{X} + (\overline{X})^2 \\
=& \overline{X^2} - (\overline{X})^2.
\end{aligned}$$

$$\begin{aligned}
E(\hat{\sigma^2}) =& E(\overline{X^2}) - E((\overline{X})^2) \\
=& E(\overline{Y}) - E(Z^2) && (Y, Z) =? \\
=& E(Y) - (\sigma_Z^2 + (\mu_Z)^2) && [15]: \quad \sigma_Y^2 = E(Y^2) - \mu_Y^2 \\
=& E(X^2) - (\sigma_{\overline{X}}^2 + (\mu_{\overline{X}})^2) \\
=& E(X^2) - \sigma_{\overline{X}}^2 - (\mu_X)^2 \\
=& E(X^2) - (\mu_X)^2 - \sigma_{\overline{X}}^2 \\
=& \sigma_X^2 - \sigma_X^2/n \\
=& (1 - \frac{1}{n})\sigma_X^2 \\
=& \frac{n-1}{n}\sigma^2 && \text{Thus } \hat{\sigma^2} \text{ is a biased estimator of } \sigma^2. \\
S^2 =& \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2.
\end{aligned}$$

$$
\begin{aligned}
&= \frac{n}{n-1}(\overline{X^2} - (\overline{X})^2) \\
&= \frac{n}{n-1}\hat{\sigma^2} \\
E(S^2) &= E(\frac{n}{n-1}\hat{\sigma^2}) \\
&= \frac{n}{n-1}E(\hat{\sigma^2}) \\
&= \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2 \text{ Thus } S^2 \text{ is an unbiased estimator of } \sigma^2.
\end{aligned}
$$

Remark. Since $E(\hat{\sigma^2}) = \frac{n-1}{n}\sigma^2$, thus $S^2 = \frac{n}{n-1}\hat{\sigma^2}$ is unbiased estimator of σ^2. Recall in §8.2.

$$
\begin{aligned}
E(S) &= E(\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}) = \sqrt{\frac{1}{n-1}}\sigma E(\sqrt{\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2}) \\
&= \sqrt{\frac{1}{n-1}}\sigma E(\sqrt{Y}) \qquad\qquad Y \sim \chi^2(n-1) = G(\tfrac{n-1}{2}, 2) \\
&= \sqrt{\frac{1}{n-1}}\sigma \int_0^\infty \sqrt{y}\frac{y^{\frac{n-1}{2}-1}e^{-y/2}}{\Gamma(\frac{n-1}{2})2^{\frac{n-1}{2}}}dy \\
&= \sqrt{\frac{1}{n-1}}\sigma \int_0^\infty \frac{y^{\frac{n}{2}-1}e^{-y/2}}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}dy\frac{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}{\Gamma(\frac{n-1}{2})2^{\frac{n-1}{2}}} \qquad \textbf{why do this ??} \\
&= \sqrt{\frac{1}{n-1}}\sigma\frac{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}{\Gamma(\frac{n-1}{2})2^{\frac{n-1}{2}}} \\
&= \sigma\sqrt{\frac{2}{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \\
\textit{Let}\ \ \tilde{\sigma} &= \frac{1}{\sqrt{\frac{2}{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}}S. \text{ Then } \tilde{\sigma} \text{ is unbiased.}
\end{aligned}
$$

Q: Since $\tilde{\sigma}$ is unbiased estimator of σ, is $(\tilde{\sigma})^2$ an unbiased estimator of σ^2 **??**

$E(Y^2) = \sigma_Y^2 + \mu_Y^2$.

Formulae

1. Estimator of μ is $\overline{X}$ where $\overline{X} =$ ______________, Estimator of σ^2 is S^2, where $S^2 =$ ______________, **key:** $\underline{\sum_i X_i/n}$, $\underline{\frac{1}{n-1}\sum_{i=1}^n(X_i - \overline{X})^2}$.
2. An estimator $\hat{\theta}$ is unbiased if ______________, bias $B(\hat{\theta}) =$ ______________, MSE = __________________, (**key:** $\underline{E(\hat{\theta}) = \theta}$, $\underline{E(\hat{\theta}) - \theta}$, $\underline{V(\hat{\theta}) + (B(\hat{\theta}))^2}$,

§8.4. Evaluating the goodness of a point estimator.

Let X be a r.v. $X \sim f(x;\theta)$. Let $\hat{\theta}$ be an estimator of θ.

$\hat{\theta} - \theta =$ error of the estimator.

$P(|\hat{\theta}-\theta|=0)=0$ most of the time.

Thus it is often to consider error bound $b=2\sigma_{\hat{\theta}}$. That is,

$|\hat{\theta}-\theta|<b=2\sigma_{\hat{\theta}}$.

Ideally, if θ is the mean, $\hat{\theta}=X$ and $\sigma_{\hat{\theta}}$ is known, then

$$P(|\hat{\theta}-\theta|<2\sigma_{\hat{\theta}})=\begin{cases}0.9544 & \text{if } X\sim N(\mu,\sigma^2) \text{ (from the normal table)}\\ 1 & \text{if } X\sim U(0,2\theta)\\ 0.9502 & \text{if } X\sim Exp(\theta).\end{cases}$$

Reason: (1) If $X\sim N(\mu,\sigma^2)$ and $\sigma_{\hat{\theta}}=\sigma_X=\sigma$ is known, then

$P(|\hat{\theta}-\theta|<2\sigma_{\hat{\theta}}))$

$=P(|\hat{\theta}-\mu|<2\sigma)$

$=P(|\frac{\hat{\theta}-\mu}{\sigma}|<2)$ $\quad$ (*see* [22])

$=1-2\times 0.0228$ from the table in P.848

$=1-0.0456=0.9544$.

(2) If $X\sim U(0,2\theta)$, (see [21]). $E(X)=\frac{0+2\theta}{2}=\theta$, $\sigma^2=\frac{(2\theta+0)^2}{12}=\theta^2/3$

$2\sigma=\frac{2}{\sqrt{3}}\theta>1$

$P(|\hat{\theta}-\theta|<2\sigma_{\hat{\theta}}))$

$=P(|\hat{\theta}-\theta|<2\theta/\sqrt{3})$

$=P(\theta-2\theta/\sqrt{3}<\hat{\theta}<\theta+2\theta/\sqrt{3})$

$\geq P(0\leq\hat{\theta}\leq 2\theta)=1$.

(3) If $X\sim Exp(\theta)$ with $E(X)=\theta$, then $X\sim\Gamma(1,\theta)$, $\sigma=\theta$. $\quad$ (see [23]. [24]).

$P(|X-\theta|<2\sigma)=P(X<3\theta)=1-exp(-3)=1-0.04978707\approx 0.95$.

In general, by Tchebysheff's inequality, $P(|X-\theta|>2\sigma)\leq 1/2^2$. (see [14]).

Thus $P(|X-\theta|<2\sigma)\geq 0.75$.

But $\sigma_{\hat{\theta}}$ is often unknown.

Thus estimate it by $\hat{\sigma}_{\hat{\theta}}$, and $b=2\hat{\sigma}_{\hat{\theta}}$ called the 2-standard error (bound) (SE).

Ex. 1. A sample of $n=1000$ voters showed $Y=560$ in favor of A. Estimate p, the fraction of voters in the population favouring A and give a 2-standard-error bound to the estimate.

Sol. $\hat{p}=Y/n=560/1000=0.56$. $\sigma^2_{\hat{p}}=pq/n$.

$2\hat{\sigma}_{\hat{p}}=2\sqrt{\hat{p}\hat{q}/n}=2\sqrt{0.56\times 0.44/1000}\approx 0.03$.

Ex. 2. A comparison of durability of 2 types of car tires was obtained by road-testing samples of $n_1=n_2=100$ tires of each type.

$\overline{Y}_1=26400$ miles, $\overline{Y}_2=25100$ miles,

$S_1^2=1,440,000$ and $S_2^2=1,960,000$.

Estimate the difference in mean mileage to wear-out and place a 2-SE bound on the error.

Sol. $\theta=\mu_1-\mu_2$, $\hat{\theta}=\overline{Y}_1-\overline{Y}_2=1300$.

$2SD=2\sigma_{\hat{\theta}}=?$

$$\sigma^2_{\hat{\theta}}=\sigma^2_{\overline{Y}_1-\overline{Y}_2}=V(\overline{Y}_1-\overline{Y}_2)=V(\overline{Y}_1)+V(\overline{Y}_2) \qquad [34]$$

$$
\begin{aligned}
&= V(Y_1)/n_1 + V(Y_2)/n_2. \\
2SD &= 2\sqrt{V(Y_1)/n_1 + V(Y_2)/n_2} \\
2SE &= 2\sqrt{S_1^2/n_1 + S_2^2/n_2} \\
&= 2\sqrt{\frac{1440000 + 1960000}{100}} \\
&= 368.8
\end{aligned}
$$

§**8.5.** Confidence interval (CI).

Def. Suppose $P(\hat{\theta}_L \le \theta \le \hat{\theta}_U) = 1 - \alpha$. Then

$[\hat{\theta}_L, \hat{\theta}_U]$ is called a $100(1-\alpha)\%$ (2-sided) confidence interval (CI) of θ;

$[0, \hat{\theta}_U]$ is called a $100(1-\alpha)\%$ lower one-sided confidence interval (CI) of θ;

$[\hat{\theta}_L, \infty]$ is called a $100(1-\alpha)\%$ upper one-sided confidence interval (CI) of θ;

Meaning of the 95% CI for θ:

If one repeats 100 times, to construct the the 95% CI for θ, then about 95% of the times, $[\hat{\theta}_L, \hat{\theta}_U]$ will contain θ.

Ex. 1. If X_1, ..., X_{100} are i.i.d. from $N(\mu, 1)$, find a 95% CI for μ.

Sol. $[\overline{X} - 1.96/\sqrt{n}, \overline{X} + 1.96/\sqrt{n}]$ or written as $\overline{X} \pm 1.96/\sqrt{n}$.

Reason:
$$
\begin{aligned}
&P(\overline{X} - 1.96/\sqrt{n} < \mu < \overline{X} + 1.96/\sqrt{n}) \\
=&P(-1.96/\sqrt{n} < \mu - \overline{X} < 1.96/\sqrt{n}) \\
=&P(|\mu - \overline{X}| < 1.96 \times 1/\sqrt{n}) \\
=&P(|\frac{\overline{X} - \mu}{1/\sqrt{n}}| \le 1.96) \\
=&0.95
\end{aligned}
$$

Quiz on Friday. 447: 1-20. 448: 1-3.

Ex. 2. Suppose that we are to obtain a single observation Y from an exponential distribution with mean θ, say $Y \sim Exp(\theta)$. Use Y to construct a 90% CI for θ.

Sol. Try to obtain $P(a < \theta < b) = 0.9$ or to obtain $P(a \le \theta \le b) = 0.9$.

Idea: use a pivotal method:

(1) Find a pivotal function $Z = g(Y, \theta)$, such that Z is independent of θ;

(2) Solve $P(a < g(Y, \theta) < b) = 0.9$.

Let $Z = Y/\theta$ $(= g(Y, \theta))$. Then $Z \sim Exp(1)$ (to be proved later).

$$
\begin{aligned}
0.9 =&P(a \le Z \le b) \qquad\qquad = P(-\ln 0.95 \le Z \le -\ln 0.05) \\
=&P(a \le Y/\theta \le b) \qquad\qquad [17]\ in\ 447
\end{aligned}
$$

$$
\begin{aligned}
&= P(1/a \ge \theta/Y \ge 1/b)\\
&= P(Y/a \ge \theta \ge Y/b)
\end{aligned}
$$

then a 90% CI for θ is $[Y/b, Y/a]$, where $a = -ln0.95 = 0.05129$ and $b = -\ln 0.05 = 2.995732$.

Why ? [23], [24] $=> f = \frac{1}{\theta}e^{-x/\theta}$, $x > 0$. $=> g(Y,\theta) = \frac{Y}{\theta}$ is a pivatol function.

$$
\begin{aligned}
P(Z > t) &= P(Y/\theta > t) && t > 0\\
&= P(Y > \theta t)\\
&= \int_{\theta t}^{\infty} \frac{1}{\theta} e^{-y/\theta} dy\\
&= \int_{\theta t}^{\infty} e^{-y/\theta} d\frac{y}{\theta}\\
&= \int_{t}^{\infty} e^{-u} du && u =?\\
&= -e^{-u}\Big|_t^{\infty}\\
&= e^{-t}\\
P(-\ln 0.95 \le Z \le -ln0.05) &= P(-\ln 0.95 < Z < -ln0.05)\\
&= F_Z(b) - F_Z(a-) && [17]\ in\ 447\\
&= 1 - e^{-b} - (1 - e^{-a})\\
&= e^{-a} - e^{-b}\\
&= e^{-(-\ln 0.95)} - e^{-(-\ln 0.05)}\\
&= e^{\ln 0.95} - e^{\ln 0.05}\\
&= 0.95 - 0.05\\
&= 0.9
\end{aligned}
$$

Thus a 90% CI for θ is $[Y/b, Y/a] = [\frac{Y}{-\ln 0.95}, \frac{Y}{-ln0.05}]$.

```
> a=−log(0.95)
> b=−log(0.05)
> 1/a−1/b
    [1] 19.16192
```

The length of the 1st CI is $Y/a - Y/b = 19.16Y$.

$$
\begin{aligned}
\textbf{Another way: } P(0 \le Z \le -\ln 0.1) &= P(-\ln 1 \le Z \le -ln0.1)\\
&= F_Z(-ln0.1) - F_Z(-\ln 1)\\
&= 1 - e^{-(-\ln 0.1)} - (1 - e^{-(-\ln 1)})\\
&= 1 - e^{\ln 0.1} - (1 - e^{\ln 1})\\
&= 1 - 0.1 = 0.9
\end{aligned}
$$

Thus $[\frac{Y}{-\ln 0.1}, \frac{Y}{0}]$ is another 90% CI for θ, with a length ∞.

Q: Which of these two CIs is better ?

$$\textbf{The 3rd way: } P(-\ln 0.9 \le Z \le \infty) = P(-\ln 0.9 \le Z \le -ln 0) \qquad \ln 0 = ?$$
$$= F_Z(-ln0) - F_Z(-ln0.9)$$
$$= 1 - e^{-(-\ln 0)} - (1 - e^{-(-\ln 0.9)})$$
$$= 1 - e^{\ln 0} - (1 - e^{\ln 0.9})$$
$$= 0.9$$

Thus $[\frac{Y}{\infty}, \frac{Y}{-\ln 0.9})$ $(= [0, \frac{Y}{-\ln 0.9}]$ is a third 90% CI for θ, with a length 9.491222Y.

```
> c=-log(0.9)
> 1/c
  [1] 9.491222
```

$$\begin{pmatrix} a < Z < b & [Y/b, Y/a] & length \\ -\ln 0.95 < Z < -\ln 0.05 & [\frac{Y}{-\ln 0.05}, \frac{Y}{-\ln 0.95}] & 19.2 \\ -\ln 1 < Z < -\ln 0.1 & [\frac{Y}{-\ln 0.1}, \frac{Y}{-\ln 1}] & \infty \\ -\ln 0.9 < Z < -\ln 0 & [\frac{Y}{-\ln 0}, \frac{Y}{-\ln 0.9}] & 9.49 \end{pmatrix}$$

Q: Which of these 3 CIs is better ?

Comments.

Ex. 3. Suppose that $X \sim U(\theta, \theta+1)$. Construct a 95% CI for θ.

Sol. Let $Z = g(X, \theta) = X - \theta$. Then $Z \sim U(0, 1)$.

Reason: $F_X(t) = (t - \theta)$ if $t \in (\theta, \theta+1)$.

$P(Z \le t) \quad (if \;\; 0 < t < 1)$

$= P(X - \theta \le t)$

$= P(X \le \theta + t)$

$= \int_{-\infty}^{\theta+t} I(x \in (\theta, \theta+1))dx \qquad I(x \in B) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{otherwise} \end{cases}$

$= [\int_{-\infty}^{\theta} + \int_{\theta}^{\theta+t}] I(x \in (\theta, \theta+1))dx$

$= \int_{\theta}^{\theta+t} I(x \in (\theta, \theta+1))dx$

$= \int_{\theta}^{\theta+t} 1 dx I(t \in (0, 1)$

$= \begin{cases} ? & \text{if } t \le 0 \\ t & \text{if } t \in (0,1) \\ ? & \text{if } t \ge 1 \end{cases}$

$<=> F_X(t) = (t-\theta)I(0 < t-\theta < 1) + I(t-\theta \ge 1)$. $<=> F_Z(x) = xI(0 < x < 1) + I(x \ge 1)$.

$P(a < X - \theta < b) = 0.95$

=P($0.025 < X - \theta < 0.975$)

=P($0.025 - X < -\theta < -X + 0.975$)
=P($X - 0.025 > \theta > X - 0.975$)
=P($X - 0.975 \le \theta \le X - 0.025$)

$[X - 0.975, X - 0.025]$ is a 95% CI for θ.

Q: How about $[X-1, X-0.05]$ due to $P(0.05 < X - \theta < 1) = 0.95$?

How about $[X - 0.95, X]$ due to $P(0 < X - \theta < 0.95) = 0.95$?

$$\begin{pmatrix} a < Z = Y - \theta < b & [X-b, X-a] & length \\ 0.025 < Z < 0.975 & [X-0.975, X-0.025] & 0.95 \\ 0 < Z < 0.95 & [X-0.95, X] & 0.95 \\ 0.05 < Z < 1 & [X-1, X-0.05] & 0.95 \end{pmatrix}$$

Summary. There are 3 typical pivotal functions $Z = g(X, \theta)$:

Ex.1. $X \sim N(\mu, \sigma^2)$, $Z = X - \mu \sim N(0, \sigma^2)$ or $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.

Ex.2. $X \sim Exp(\theta)$ $(E(X) = \theta)$, $Z = g(X, \theta) = \frac{X}{\theta} \sim Exp(1)$.

Ex.3. $X \sim U(\theta, \theta + b)$, $Z = g(X, \theta) = \frac{X-\theta}{b} \sim U(0,1)$.

§**8.6. Large sample CI** for θ: $[\hat\theta_L, \hat\theta_U]$.

Exact CI $P(\hat\theta_L \le \theta \le \hat\theta_U) = 1 - \alpha$.

Large sample approximate CI $P(\hat\theta_L \le \theta \le \hat\theta_U) \approx 1 - \alpha$.

Most of the time (due to the CLT),

$Z = \frac{\hat\theta - \theta}{\sigma_{\hat\theta}} \sim N(0,1)$ approximately,

or $Z = \frac{\hat\theta - \theta}{\hat\sigma_{\hat\theta}} \sim N(0,1)$ approximately, **What is their difference ?**

then

$$\begin{aligned} & P(|Z| \le z_{\alpha/2}) && \approx 1 - \alpha \\ =& P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) \\ =& P(-z_{\alpha/2} \le \frac{\hat\theta - \theta}{\sigma_{\hat\theta}} \le z_{\alpha/2}) \\ =& P(-z_{\alpha/2}\sigma_{\hat\theta} \le \hat\theta - \theta \le z_{\alpha/2}\sigma_{\hat\theta}) \\ =& P(z_{\alpha/2}\sigma_{\hat\theta} \ge -\hat\theta + \theta \ge -z_{\alpha/2}\sigma_{\hat\theta}) \\ =& P(\hat\theta + z_{\alpha/2}\sigma_{\hat\theta} \ge \theta \ge \hat\theta - z_{\alpha/2}\sigma_{\hat\theta}) \\ =& P(\hat\theta - z_{\alpha/2}\sigma_{\hat\theta} \le \theta \le \hat\theta + z_{\alpha/2}\sigma_{\hat\theta}) \\ \approx& P(\hat\theta - z_{\alpha/2}\hat\sigma_{\hat\theta} \le \theta \le \hat\theta + z_{\alpha/2}\hat\sigma_{\hat\theta}). \end{aligned}$$

Thus an approximate CI for θ is $\hat\theta \pm z_{\alpha/2}\hat\sigma_{\hat\theta}$. In particular by Table 4 (p.848) about $N(0,1)$,

an approximate 90% CI is $\hat\theta \pm 1.645\hat\sigma_{\hat\theta}$;

an approximate 95% CI is $\hat\theta \pm ????\hat\sigma_{\hat\theta}$;

an approximate 99% CI is $\hat\theta \pm 2.57\hat\sigma_{\hat\theta}$.

An approximate one-sided CI for θ is

$[0, \hat{\theta} + z_\alpha \hat{\sigma}_{\hat{\theta}}]$ (upper bounded);

$[\hat{\theta} - z_\alpha \hat{\sigma}_{\hat{\theta}}, \infty)$ (lower bounded). Need to find $\hat{\theta}=$? $\sigma_{\hat{\theta}}=$? or $\hat{\sigma}_{\hat{\theta}}=$?

Recall Table 8.1 (p.397):

θ	*sample size(s)*	$\hat{\theta}$	$E(\hat{\theta})$	$\sigma_{\hat{\theta}}$	$\hat{\sigma}_{\hat{\theta}}$
μ	n	$\overline{Y}$	μ	$\sigma/\sqrt{n}$	$S/\sqrt{n}$
p	n	$\hat{p}$	p	$\sqrt{pq/n}$	$\sqrt{\hat{p}(1-\hat{p})/n}$
$p_1 - p_2$	n_1 *and* n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{p_1q_1/n_1 + p_2q_2/n_2}$	??
$\mu_1 - \mu_2$	n_1 *and* n_2	$\overline{Y}_1 - \overline{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$	??

Answer to the last question:

If $\sigma_1 = \sigma_2$ is assumed, ??= $\sqrt{S^2/n_1 + S^2/n_2}$, where

$$S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{\sum_{i=1}^{n_1}(Y_{1i} - \overline{Y}_1)^2 + \sum_{j=1}^{n_2}(Y_{2j} - \overline{Y}_2)^2}{n_1+n_2-2}.$$

Otherwise, ??=$\sqrt{S_1^2/n_1 + S_2^2/n_2}$.

Ex. 8.7. The shopping times of $n = 64$ randomly selected customers at a local market were recorded. The mean and variance of the 64 shopping times were 33 minutes and 256 minutes^2, respectively. Find a 90% CI for the true average shopping time per customer.

Sol. Formula: $\hat{\theta} \pm z_{\alpha/2}\hat{\sigma}_{\hat{\theta}}$ has the form: $\overline{Y} \pm z_{\alpha/2}S/\sqrt{n}$,

$\overline{Y} =$? $S = 256$? $\alpha/2 = 0.45$ or 0.05 ? $z_{\alpha/2} =$?

$\overline{Y} \pm 1.645 S/\sqrt{n}$

$33 \pm 1.645\sqrt{256/n}$

$33 - 1.645\sqrt{256/n} = 29.71$

$33 + 1.645\sqrt{256/n} = 36.29$

Thus a 90% CI for the true average shopping time per customer is $[29.71, 36.29]$ or 33 ± 3.29.

```
> 1.645*sqrt(256/64)
[1] 3.29
```

Q: Does the true average shopping time per customer $\mu \in [29.71, 36.29]$? Yes, No, DNK.

90 percents of the time, μ will be contained by a CI.

Ex. 8.8. Two brands of refrigerators, denoted by A and B, are each guaranteed for 1 year. In a random sample of 50 refrigerators of brand A, 12 were observed to fail before 1 year. In a random sample of 60 refrigerators of brand B, 12 were also observed to fail before 1 year. Estimate the true difference $(p_1 - p_2)$ between proportions of failures during the guarantee period, with confidence coefficient approximately 0.98.

Sol. Formula: $\hat{\theta} \pm z_{\alpha/2}\hat{\sigma}_{\hat{\theta}}$ has the form: $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{p_1q_1/n_1 + p_2q_2/n_2}$

Or $(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\hat{p}_1\hat{q}_1/n_1 + \hat{p}_2\hat{q}_2/n_2}$ **Q: Which is the formula to choose ?**

> qnorm(0.98) $\Phi^{-1}(0.98)$, $pnorm(2.05) = \Phi(2.05) = 0.98$, $\Phi(x)$ is the cdf of $N(0,1)$.

[1] 2.053749

> qnorm(0.99)

[1] 2.326348 **Which is correct one ?**

$\hat{p}_1 = 12/50$,

$\hat{p}_2 = 12/60$,

$z_{\alpha/2} = z_{0.01} = 2.33$.

$(0.24 - 0.2) \pm 2.33\sqrt{\frac{0.24*0.76}{50} + \frac{0.2*0.8}{60}}$

0.04 ± 0.1851.

Ans: The true difference $(p_1 - p_2)$ between proportions of failures during the guarantee period, with confidence coefficient approximately 0.98 is

0.04 ± 0.1851?

Or $[-.1451, 0.2251]$?

Ex. 3. A simulation study. Suppose $n = 100$ observations X_i's from $bin(1,p)$, where $p = 0.5$. Let $Y = \sum_{i=1}^n X_i$. A 80% approximate CI is $Y/n \pm 1.28\sqrt{\frac{\frac{Y}{n}(1-Y/n)}{n}}$. Note $Y \sim bin(100, 0.5)$.

```
> n=100
> Y=rbinom(1,100,0.5)
> p=Y/n
> c(Y/n-1.28*sqrt(p*(1-p)/n), Y/n+1.28*sqrt(p*(1-p)/n))
[1] [0.4863208, 0.6136792]                # Does it contain p ?
> Y=rbinom(1,100,0.5)
> p=Y/n
> c(Y/n-1.28*sqrt(p*(1-p)/n), Y/n+1.28*sqrt(p*(1-p)/n))
[1] [0.4160512, 0.5439488]                # Does it contain p ?
> Y=rbinom(1,100,0.5)
> p=Y/n
> c(Y/n-1.28*sqrt(p*(1-p)/n), Y/n+1.28*sqrt(p*(1-p)/n))
[1] [0.327568, 0.452432]                # Does it contain p ?
> Y=rbinom(1,100,0.5)
> p=Y/n
> c(Y/n-1.28*sqrt(p*(1-p)/n), Y/n+1.28*sqrt(p*(1-p)/n))
[1] [0.3764625, 0.5035375]                # Does it contain p ?
```

Summary: The simulation study shows that an 80% CI interval for p may or may not contain the true value of p (which is $p = 0.5$ in this example). However, if we repeat this procedure 100 times, roughly 80% of the time, the true value of p will be contained in the CIs.

Q: Suppose that 100 CIs were constructed.

1. Is it possible that the true value of p will be contained in the CIs all time ? Yes, Unlikely.
2. Is it possible that the true value of p will be contained in the CIs half of the time ? Y, U
3. Is it possible that the true value of p will be contained in the CIs 82% of the time ? Y, U

Q: How can we tell ?

```
> sqrt(0.8*0.2/100)
[1] 0.04
> pnorm(3)
[1] 0.9986501
> 3*sqrt(0.8*0.2/100)
[1] 0.12
```

$1 - 0.8 = 0.2 > 0.12$

$0.8 - 0.5 = 0.3 > 0.12$

§**8.7. Selecting the sample size.**

By the CLT

$$P(\frac{\hat{\theta}-\theta}{\hat{\sigma}_{\hat{\theta}}} \leq t) \approx \Phi(t) \text{ if } n \text{ is large.}$$

447 [42] $F_{\overline{Y}}(t) \approx \Phi(\underline{\frac{t-\mu_{\overline{Y}}}{\sigma_{\overline{Y}}}})$, where $\Phi(t)$ is the cdf of $\underline{N(0,1)}$

It leads to CI $\hat{\theta} \pm z_{\alpha/2}\hat{\sigma}_{\hat{\theta}}$.

Then L =length of the CI $= 2z_{\alpha/2}\hat{\sigma}_{\hat{\theta}}$.

error$=z_{\alpha/2}\hat{\sigma}_{\hat{\theta}}$.

Q: How to determine n for a given L or error.

Ideally, n is as large as possible due to [42] in 447.

Practically, n should not be so large, as it costs time and money.

Example 8.9. The reaction of an individual to a stimulus in a psychological experiment may take one of two forms: A & B. If an experimenter wishes to estimate the probability p that a person will react in manner A, how many people must be included in the experiment ? Here, we assume

1. the error =0.04,
2. $p \approx 0.6$,
3. error of estimate is less than 0.04 w.p. 0.9.

Sol.

$$\begin{aligned} error =& 0.04 \\ =& z_{\alpha/2}\sigma_{\hat{\theta}} \\ =& 1.645\sqrt{p(1-p)/n} \qquad or \ = 1.645\sqrt{\hat{p}(1-\hat{p})/n} \text{ ??} \end{aligned}$$

$$\begin{aligned}
&\approx 1.645\sqrt{0.6(0.4)/n}\\
&\approx \frac{1.645\sqrt{0.6(0.4)}}{\sqrt{n}} = 0.04\\
&\quad (1.645/0.04)^2 \times 0.24 \approx n\\
n &\approx 405.9. \qquad \text{Is it the final answer ?}
\end{aligned}$$

Ans. 406 people must be included in the experiment.

Ex. 8.10. An experimenter wishes to compare the effectiveness of 2 methods of training industrial employees to perform an assembly operation. The selected employees are to be divided into two groups of equal size. The 1st receives training method 1, and the 2nd receives training method 2. After training, each employee will perform the assembly operation and the length of assembly time will be recorded.

It is expected the measurements for both groups to have a range of approximately 8 minutes.
How many workers must be selected in each group,
if the difference in mean assembly times is to be correct within 1 minute with prob. 0.95 ?

Sol. Let $\theta = \mu_1 - \mu_2$.
Let $Z = \hat\theta = \overline{X} - \overline{Y}$, the difference in mean assembly time.
Z is to be correct within 1 minute

$$=> \begin{cases} |Z-\theta| = |\overline{X}-\overline{Y}-\theta| = |\hat\theta-\theta| = 1 = z_{\alpha/2}\hat\sigma_{\hat\theta} & ? \\ |Z-\theta| = |\overline{X}-\overline{Y}-\theta| = |\hat\theta-\theta| \le 1 = z_{\alpha/2}\hat\sigma_{\hat\theta} & ? \end{cases} \tag{1}$$

$$\begin{aligned}
\sigma_{\hat\theta} &= \sqrt{\sigma_X^2/n + \sigma_Y^2/n}\\
&= \frac{1}{\sqrt{n}}\sqrt{\sigma_X^2+\sigma_Y^2}\\
&= \frac{1}{\sqrt{n}}\sqrt{2\sigma_X^2} \ \text{ assuming } \sigma_X^2 = \sigma_Y^2,
\end{aligned} \tag{2}$$

as "It is expected the measurements for both groups to have a range of approximately 8 min."
=>

$$8 \approx 2\times 1.96\sigma_X = 2\times 1.96\sigma_Y \ => \ \sigma_X = \sigma_Y \approx 2. \tag{3}$$

$$(1),\ (2)\ \textit{and}\ (3) \ => \ 1 = 1.96\frac{1}{\sqrt{n}}\sqrt{\sigma_X^2+\sigma_X^2}.$$

=> $1 = 1.96\frac{1}{\sqrt{n}}\sqrt{2^2+2^2}$
$\quad n \approx 30.73.$

Ans: Each group needs 31 workers.

§8.8. Small-sample CI for μ and $\mu_1 - \mu_2$.

We have learned several types of CIs:

§8.5. A 95% CI for μ_X is $\begin{cases} \overline{X} \pm 1.96/\sqrt{n} & \text{if } X_i\text{'s are i.i.d.} \sim N(\mu_X, 1) \\ [\frac{X}{-\ln 0.975}, \frac{X}{-\ln 0.025}] & \text{if } X \sim Exp(\mu_X); \end{cases}$

for θ is $[X - 0.975, X - 0.025]$ if $X \sim U(\theta, \theta+1)$;

§8.6. If n and m are large, given X_1, ..., X_n, Y_1, .., Y_m, and $X_i's \perp Y_j's$, then

a $(1-\alpha)\%$ CI for μ_X is $\overline{X} \pm z_{\alpha/2}\hat{\sigma}/\sqrt{n}$,

for $\mu_X - \mu_Y$ is $\overline{X} - \overline{Y} \pm z_{\alpha/2}\sqrt{S_X^2/n + S_Y^2/m}$, (provided n, or n and $m \geq 20$.)

Q: How about n or $m < 20$?

Need stronger assumptions:

Case 1. X_1, ..., X_n are i.i.d. $\sim N(\mu_X, \sigma^2)$,

CI for μ_X: $\overline{X} \pm t_{\alpha/2,n-1}S_X/\sqrt{n}$,

Case 2. X_1, ..., X_n are i.i.d. $\sim N(\mu_X, \sigma^2)$, Y_1, ..., Y_m are i.i.d. $\sim N(\mu_Y, \sigma^2)$ and $X_i \perp Y_j$,

CI for $\mu_X - \mu_Y$: $\overline{X} - \overline{Y} \pm t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n} + \frac{1}{m}}$, where $S_p^2 = \frac{S_X^2(n-1)+S_Y^2(m-1)}{n+m-2}$.

Ex. 8.11. Suppose that 8 independent observations are obtained from $N(\mu, \sigma^2)$.

3005, 2925, 2935, 2965, 2995, 3005, 2937, 2905.

Construct a 95% CI for μ.

Sol. The 95% CI is $\overline{X} \pm t_{\alpha/2,n-1}S/\sqrt{n}$.

$n = 8$. $\overline{X} = \frac{\sum_{i=1}^n X_i}{n} = 2959$, $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$, and $S = 39.1$, $t_{0.025,7} = 2.365$.

$2959 \pm 2.365 * 39.1/\sqrt{8}$

A 95% CI for μ is 2959 ± 32.7

Ex. 8.12. Suppose that 2 sets of independent samples are obtained from $N(\mu_X, \sigma^2)$ and $N(\mu_Y, \sigma^2)$.

32, 37, 35, 28, 41, 44, 35, 31, 34,

35, 31, 29, 25, 34, 40, 27, 32, 31,

Construct a 95% CI for $\mu_1 - \mu_2$.

Sol. $\overline{X} - \overline{Y} \pm t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n} + \frac{1}{m}}$, where $S_p^2 = \frac{S_X^2(n-1)+S_Y^2(m-1)}{n+m-2}$.

$\overline{X} = 35.22$, $\overline{Y} = 31.56$,

$S_X^2 = 24.445$, $S_Y^2 = 20.027$,

$S_p^2 = \frac{24.445\times(9-1)+20.027\times(9-1)}{16} = 22.236$

$S_p = 4.716$,

$t_{0.025,n+m-2} = t_{\alpha/2,16} = 2.12$,

$35.22 - 31.56 \pm 2.12 * 4.716\sqrt{1/9 + 1/9}$

The 95% CI is 3.66 ± 4.71

How to remember the formula and derive it ?

$\vdash$: CI for μ_X: $\overline{X} \pm t_{\alpha/2,n-1}S_X/\sqrt{n}$.

448. Formula [5]

If (1) X_1, ..., $X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$.

(2) $Y_1, ..., Y_m$ are i.i.d. from $N(\mu_2, \underline{\sigma^2})$, (3) X_i's $\perp$ Y_j's,

then $T = \underline{\frac{\overline{X}-\mu_o}{S_x/\sqrt{n}}}, \sim \underline{t_{n-1}}$, ...

Thus $P(|\frac{\overline{X}-\mu_x}{S_X/\sqrt{n}}| \le t_{\alpha/2,n-1}) = 1-\alpha$ as t_{n-1} is symmetric,

$=P(|\overline{X}-\mu_x| \le t_{\alpha/2,n-1}S_X/\sqrt{n})$

$=P(-t_{\alpha/2,n-1}S_X/\sqrt{n} \le \overline{X}-\mu_x \le t_{\alpha/2,n-1}S_X/\sqrt{n})$

$=P(t_{\alpha/2,n-1}S_X/\sqrt{n} \ge -\overline{X}+\mu_x \ge -t_{\alpha/2,n-1}S_X/\sqrt{n})$

$=P(\overline{X}+t_{\alpha/2,n-1}S_X/\sqrt{n} \ge \mu_x \ge \overline{X}-t_{\alpha/2,n-1}S_X/\sqrt{n})$

$=P(\overline{X}-t_{\alpha/2,n-1}S_X/\sqrt{n} \le \mu_x \le \overline{X}+t_{\alpha/2,n-1}S_X/\sqrt{n})$.

=> a $100(1-\alpha)\%$ CI for μ_X is $\overline{X} \pm t_{\alpha/2,n-1}S_X/\sqrt{n}$.

Or $P(|\frac{\overline{X}-\mu_x}{S_X/\sqrt{n}}| \le t_{\alpha/2,n-1}) = 1-\alpha$

=> $|\frac{\overline{X}-\mu_x}{S_X/\sqrt{n}}| \le t_{\alpha/2,n-1}$

=> $|\overline{X}-\mu_x| \le t_{\alpha/2,n-1}S_X/\sqrt{n}$

=> $|\mu_X-\overline{X}| \le t_{\alpha/2,n-1}S_X/\sqrt{n}$

=> $\overline{X}-t_{\alpha/2,n-1}S_X/\sqrt{n} \le \mu_x \le \overline{X}+t_{\alpha/2,n-1}S_X/\sqrt{n}$.

$\vdash$: CI for $\mu_x - \mu_y$: $\overline{X}-\overline{Y} \pm t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n}+\frac{1}{m}}$, where $S_p^2 = \frac{S_X^2(n-1)+S_Y^2(m-1)}{n+m-2}$.

448. Formula [5]

If (1) $X_1, ..., X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$.

(2) $Y_1, ..., Y_m$ are i.i.d. from $N(\mu_2, \underline{\sigma^2})$, (3) X_i's $\perp$ Y_j's,

then $T = \underline{\frac{\overline{X}-\mu_o}{S_x/\sqrt{n}}}, \sim \underline{t_{n-1}}$,

$T = \frac{\overline{X}-\overline{Y}-(\mu_x-\mu_y)}{\hat{\sigma}_p\sqrt{1/n_x+1/n_y}} \sim \underline{t_{n+m-2}}$, where $\hat{\sigma} = \underline{\sqrt{\frac{(n-1)S_x^2+(m-1)S_y^2}{n+m-2}}}$,

$W = (n_x-1)S_x^2/\sigma^2 \sim \underline{\chi^2_{n-1}}$,

$F = S_x^2/S_y^2 \sim \underline{F_{n-1,m-1}}$.

Thus $\frac{\overline{X}-\overline{Y}-(\mu_x-\mu_y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{n+m-2}$

$P(|\frac{\overline{X}-\overline{Y}-(\mu_x-\mu_y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}}| \le t_{\alpha/2,n+m-2}) = 1-\alpha$

$= P(-t_{\alpha/2,n+m-2} \le \frac{\overline{X}-\overline{Y}-(\mu_x-\mu_y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \le t_{\alpha/2,n+m-2})$

$= P(-t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n}+\frac{1}{m}} \le \overline{X}-\overline{Y}-(\mu_x-\mu_y) \le t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n}+\frac{1}{m}})$

$= P(t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n}+\frac{1}{m}} \ge -(\overline{X}-\overline{Y})+(\mu_x-\mu_y) \ge -t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n}+\frac{1}{m}})$

$= P\big((\overline{X}-\overline{Y})+t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n}+\frac{1}{m}} \ge (\mu_x-\mu_y) \ge (\overline{X}-\overline{Y})-t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n}+\frac{1}{m}}\big)$

$= P\big((\overline{X}-\overline{Y})-t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n}+\frac{1}{m}} \le (\mu_x-\mu_y) \le (\overline{X}-\overline{Y})+t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n}+\frac{1}{m}}\big)$

$|\frac{\overline{X}-\overline{Y}-(\mu_x-\mu_y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}}| \le t_{\alpha/2,n+m-2}$

$|\frac{(\mu_x-\mu_y)-(\overline{X}-\overline{Y})}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}}| \le t_{\alpha/2,n+m-2}$

$|(\mu_x-\mu_y)-(\overline{X}-\overline{Y})| \le t_{\alpha/2,n+m-2}\ S_p\sqrt{\frac{1}{n}+\frac{1}{m}}$

$$(\overline{X}-\overline{Y})-t_{\alpha/2,n+m-2}\ S_p\sqrt{\frac{1}{n}+\frac{1}{m}}\le \mu_x-\mu_y\le(\overline{X}-\overline{Y})+t_{\alpha/2,n+m-2}\ S_p\sqrt{\frac{1}{n}+\frac{1}{m}}$$

448. Formula [5] is due to 447 Formulae:

[42] $F_{\overline{Y}}(t)\approxeq\Phi(\ \underline{\frac{t-\mu_{\overline{Y}}}{\sigma_{\overline{Y}}}}\ )$, where $\Phi(t)$ is the cdf of $\underline{N(0,1)}$

$$\frac{\overline{X}-\mu_x}{S_X/\sqrt{n}}\sim t_{n-1}$$

[41] If $Y\sim N(\mu,\sigma^2)$, $\frac{\overline{Y}-\mu_{\overline{Y}}}{\sigma_{\overline{Y}}}\sim\underline{N(0,1)}$, $\frac{(n-1)S^2}{\sigma^2}\sim\underline{\chi^2(n-1)}$, $\overline{Y}\perp S^2$, $\sqrt{n}\frac{\overline{Y}-\mu}{S}\sim\underline{t_{n-1}}$, where $\mu_{\overline{Y}}=\underline{\mu}$, $\sigma^2_{\overline{Y}}=\underline{\sigma^2/n}$

Thus (1) $Z_y=\frac{(m-1)S^2_Y}{\sigma^2_Y}\sim\underline{\chi^2(m-1)}$ and $Z_x=\frac{(n-1)S^2_X}{\sigma^2_X}\sim\underline{\chi^2(n-1)}$,

[44] If X_1 ___ X_2.

X_i's $\sim$:	$X_1+X_2\sim$:		
$\mathcal{G}(\alpha_i,\beta)$	______		$\underline{\mathcal{G}(\alpha_1+\alpha_2,\beta)}$
$\chi^2(v_i)$	______	**key:** $\perp$,	$\underline{\chi^2(v_1+v_2)}$
$Pois(\lambda_i)$	______		$\underline{Pois(\lambda_1+\lambda_2)}$
$N(\mu_i,\sigma_i^2)$	______		$\underline{N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)}$
$bin(n_i,p)$	______		$\underline{bin(n_1+n_2,p)}$

Since $X_i\perp Y_j$, we have $Z_x\perp Z_y$. By [44], $Z_x+Z_y\sim\chi^2(n-1+m-1)$.

Moreoer, since $\sigma_x=\sigma_y$,

$$Z_x+Z_y=\underline{\frac{(n-1)S_x^2+(m-1)S_y^2}{\sigma_x^2}}\sim\chi^2(n-1+m-1).$$

Furthermore, $\frac{\overline{X}-\overline{Y}-(\mu_x-\mu_y)}{\sigma_X}\sim N(0,1)$ and $\overline{X}-\overline{Y}\perp(n-1)S_x^2+(m-1)S_y^2$.

[20] Suppose that $Z\sim N(0,1)$, $X\sim\chi^2(u)$, $Y\sim\chi^2(v)$. If $Z\perp X$, $T=\underline{Z/\sqrt{X/u}}$, then $T\sim t_u$; If $X\perp Y$, $F=\underline{\frac{X/u}{Y/v}}$, then $F\sim F_{u,v}$ and $X+Y\sim\underline{\chi^2(u+v)}$,

§8.9. CI for σ^2.

In §8.8, we need that $\sigma_X=\sigma_Y=\sigma$ for the CI of $\mu_x-\mu_y$, assuming X_1, ..., $X_n\overset{i.i.d.}{\sim}N(\mu,\sigma^2)$. Thus we need to estimate σ^2 and to construct the CI for σ^2. It is given by

$$[\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}},\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}].$$

How to derive it ? A class exercie based on The 447 formulae:

[40] Let Y_1, ..., Y_n be a random sample of Y. $\overline{Y}=\underline{\frac{1}{n}\sum_{i=1}^n Y_i}$, $S^2=S_Y^2=\underline{\frac{1}{n-1}\sum_{i=1}^n(Y_i-\overline{Y})^2}$

[41] If $Y\sim N(\mu,\sigma^2)$, then $\frac{\overline{Y}-\mu_{\overline{Y}}}{\sigma_{\overline{Y}}}\sim\underline{N(0,1)}$, $\frac{(n-1)S^2}{\sigma^2}\sim\underline{\chi^2(n-1)}$, $\overline{Y}\perp S^2$, $\sqrt{n}\frac{\overline{Y}-\mu}{S}\sim\underline{t_{n-1}}$, where $\mu_{\overline{Y}}=\underline{\mu}$, $\sigma^2_{\overline{Y}}=\underline{\sigma^2/n}$

Thus $W=(n-1)S_x^2/\sigma^2\sim\ \underline{\chi^2_{n-1}}\ $,

$P(\chi^2_{\alpha/2,n-1}\ge W\ge\chi^2_{1-\alpha/2,n-1})=1-\alpha$

$= P(\chi^2_{\alpha/2,n-1} \geq \frac{(n-1)S_x^2}{\sigma^2} \geq \chi^2_{1-\alpha/2,n-1})$
$= P(1/\chi^2_{\alpha/2,n-1} \leq \frac{\sigma^2}{(n-1)S_x^2} \leq 1/\chi^2_{1-\alpha/2,n-1})$
$= P((n-1)S_x^2/\chi^2_{\alpha/2,n-1} \leq \sigma^2 \leq (n-1)S_x^2/\chi^2_{1-\alpha/2,n-1})$

$$(1-\alpha)(100\%) \text{ CI of } \sigma^2 \text{ is } [\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}].$$

How about $(1-\alpha)100\%$ CI of σ ?

$P((n-1)S_x^2/\chi^2_{\alpha/2,n-1} \leq \sigma^2 \leq (n-1)S_x^2/\chi^2_{1-\alpha/2,n-1})$
$= P(\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}} \leq \sigma \leq \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}}).$
$=>$ CI for σ is $[\sqrt{\frac{(n-1)}{\chi^2_{\alpha/2,n-1}}}, \sqrt{\frac{(n-1)}{\chi^2_{1-\alpha/2,n-1}}})].$

Ex. 8.13. An experimenter wanted to check the variability of measurements obtained by using equipment designed to measure the volume of an audio source. 3 independent measurements recorded by it for the same sound were 4.1, 5.2 and 10.2. Estimate σ^2 with confidence coefficient 0.9.

Sol. $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$, $n = ?$

X_i's: 4.1, 5.2 and 10.2.
$S^2 = 10.57$,
$\chi^2_{0.95,2} = 0.103$
$\chi^2_{0.05,2} = 5.991.$

```
> qchisq(c(0.05,0.95),2)
   [1] 0.1025866, 5.991465
```

An unbiased estimate of σ^2 is $S^2 = 10.57$,
A 90% CI for σ^2 is $[(n-1)S_x^2/\chi^2_{0.05,2}, (n-1)S_x^2/\chi^2_{0.95,2}] = [3.53, 205.24]$,
or $[0, (n-1)S_x^2/\chi^2_{0.10,2}] = [0, 100.3222]$. **Which is better ?**

Chapter 9. Properties of the point estimators and methods of estimation

§**9.2. Relative efficiency.** It is often that there can be many estimators of a parameter θ, say $\hat{\theta}_1$, ..., $\hat{\theta}_k$.

One property we like is the unbiasedness. It is possible that $\hat{\theta}_1$,..., $\hat{\theta}_k$ are all unbiased.

Then it is natural to select the one with smaller variance.

Def. 9.1. Given two unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ of a parameter θ, with variances $V(\hat{\theta}_1)$ and $V(\hat{\theta}_2)$, the efficiency of $\hat{\theta}_2$ relative to $\hat{\theta}_1$, is defined to be the ratio

$$eff(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}.$$

If $eff(\hat{\theta}_1, \hat{\theta}_2) = 1.8$, which is better ? $\hat{\theta}_1$ or $\hat{\theta}_2$?
If $eff(\hat{\theta}_1, \hat{\theta}_2) = 0.73$, which is better ? $\hat{\theta}_1$ or $\hat{\theta}_2$?

Def. Given $X_1 \leq ... \leq X_n$, the median is $\begin{cases} \text{the middle one} & \text{if } n \text{ is odd} \\ \text{the average of the middle two} & \text{if } n \text{ is even} \end{cases}$

For example, case A: 1, 3, 8, 4, 5. The median is ?

Case B: 1, 3, 8, 8. The median is ?

Formula: if n is large and $\tilde{\theta}$ is the median of i.i.d. observations X_1, ..., X_n, then

$$V(\tilde{\theta}) \approx 1.2533^2\sigma^2/n.$$

Thus $eff(\overline{X}, \tilde{\theta}) = 1.2533$, which is better ? median or $\overline{X}$?

Ex. 9.1. Suppose that Y_1, ..., $Y_n \overset{i.i.d.}{\sim} U(0, \theta)$. Two unbiased estimators are $\hat{\theta}_1 = 2\overline{Y}$ and $\hat{\theta}_2 = \frac{n+1}{n}Y_{(n)}$, where $Y_{(n)} = \max\{Y_1, ..., Y_n\}$. Find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.

Sol. 3 steps:

(1) Show both estimators are unbiased;

(2) Compute their variances;

(3) Find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$: $eff(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}$.

Step 1. $\vdash$: both estimators are unbiased.

$E(\hat{\theta}_2) = \frac{n+1}{n}E(Y_{(n)})$.

447 [6] $f_{Y_{(j)}}(t) = \binom{n}{j-1,1,n-j}$ ________ **key:** $\underline{(F(t))^{j-1}(f(t))^1(1-F(t))^{n-j}}$.

$$\binom{n}{j-1, 1, n-j} = ?$$

$$\binom{n}{k, m, h} = \frac{n!}{k!m!h!}.$$

$$\begin{aligned} f_{Y_{(n)}}(t) =& \binom{n}{n-1, 1, n-n}\underline{(F(t))^{n-1}(f(t))^1(1-F(t))^{n-n}} \\ =& \binom{n}{n-1, 1, n-n}(F(t))^{n-1}(f(t)) \\ =& \underbrace{\binom{n}{n-1, 1, 0}}_{=??}\frac{t^{n-1}}{\theta^n}. \end{aligned}$$

$$\binom{n}{n-1, 1, n-n} = \frac{n!}{(n-1)!1!0!} \qquad n! = ?, \quad 1! = ? \quad 0! = ?$$

$$E(Y_{(n)}) = \int t f_{Y_{(n)}}(t)dt$$

$$
\begin{aligned}
&= \int_0^\theta t \binom{n}{n-1,1,0} \frac{t^{n-1}}{\theta^n} dt \\
&= \int_0^\theta \frac{n!}{(n-1)!1!0!} \frac{t^n}{\theta^n} dt \\
&= \int_0^\theta \frac{n!}{(n-1)!} \frac{t^n}{\theta^n} dt \\
&= \int_0^\theta n \frac{t^n}{\theta^n} dt \\
&= (n/(n+1)) t^{n+1} \Big|_0^\theta / \theta^n \\
&= (n/(n+1)) \theta^{n+1}/\theta^n. \\
&= (n/(n+1))\theta. \\
E(\hat{\theta}_2) &= ((n+1)/n)(n/(n+1))\theta = \theta. \\
E(\hat{\theta}_1) &= E(2\overline{Y}) = 2E(\overline{Y}) = 2\frac{0+\theta}{2} = \theta.
\end{aligned}
$$

Thus both estimators are unbiased.
Step 2. $V(2\overline{Y}) = 4V(\overline{Y}) = 4\sigma^2/n = 4\frac{\theta^2}{12n}$

$$
V(\hat{\theta}_2) = E(\hat{\theta}_2^2) - \theta^2 = (\frac{n+1}{n})^2 \underbrace{E(Y_{(n)}^2)}_{=??} - \theta^2 \qquad \hat{\theta}_2 = \tfrac{n+1}{n} Y_{(n)}
$$

$$
\begin{aligned}
E(Y_{(n)}^2) &= \int t^2 f_{Y_{(n)}}(t) dt \\
&= \int_0^\theta t^2 \binom{n}{n-1,1,0} \frac{t^{n-1}}{\theta^n} dt \\
&= \int_0^\theta \frac{n!}{(n-1)!1!0!} \frac{t^{n+1}}{\theta^n} dt \\
&= \int_0^\theta n \frac{t^{n+1}}{\theta^n} dt \\
&= (n/(n+2)) t^{n+2} \Big|_0^\theta / \theta^n \\
&= (n/(n+2)) \theta^{n+2}/\theta^n \\
&= (n/(n+2))\theta^2
\end{aligned}
$$

$$
V(\hat{\theta}_2) = E(\hat{\theta}_2^2) - \theta^2 = (\frac{n+1}{n})^2 (n/(n+2))\theta^2 - \theta^2 = \theta^2(\frac{(n+1)^2}{n(n+2)} - 1) = \frac{\theta^2}{n(n+2)}.
$$

Step 3.

$$
eff(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}
$$

$$=\frac{\frac{\theta^2}{n(n+2)}}{\frac{\theta^2}{3n}}$$

$$=3/(n+2)\begin{cases} > 1 & ?? \\ \leq 1 & ?? \end{cases}$$

Q: Which is better ?

§9.3. Consistency. Let $\hat{\theta}_n$ be an estimator of θ based on i.i.d. observations X_1, ..., X_n. Ideally, we like

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \epsilon) = 0 \text{ or } \lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \epsilon) = 1 \ \ \forall\ \epsilon > 0. \tag{1}$$

Def. 9.2. An estimator $\hat{\theta}_n$ is said to be consistent if Eq.(1) holds. It is also said that $\hat{\theta}$ converges to θ in probability, denoted by $\hat{\theta}_n \xrightarrow{P} \theta$.
Remark: The difference between $\hat{\theta} \to \theta$ and $\hat{\theta}_n \xrightarrow{P} \theta$ can be seen from the next example: Suppose that X has a uniform distribution on the interval $[1, 2]$. Let

$\theta = 0$,
$\theta_o = \mathbf{1}(X = 1)$ and
$\hat{\theta}_n = \mathbf{1}(X \in [1, 1 + \frac{1}{n}])$.

Q: $\theta_o = \theta$?

$P(\theta_o = \theta) = 0$?
$P(\theta_o = \theta) = 1$?
$\hat{\theta}_n \to \theta$?
$\hat{\theta}_n \to \theta_o$?
$\hat{\theta}_n \xrightarrow{P} \theta_o$?
$\hat{\theta}_n \xrightarrow{P} \theta$?
Abusing notations, write $\hat{\theta}_n = \hat{\theta}$.

Th. 9.1. An unbiased estimator $\hat{\theta}_n$ for θ is consistent if $\lim_{n\to\infty} V(\hat{\theta}_n) = 0$.
Proof. [14] Tchebysheff's Inequality:

$$P(|X - \mu| > k\sigma) \underline{\leq 1/k^2}.$$

For each $k > 0$, letting $\epsilon = k\sigma_{\hat{\theta}}$, then $\sigma^2_{\hat{\theta}} \to 0$ **why ?**

$$\begin{aligned} \sigma_{\hat{\theta}} \to 0 \text{ and } \epsilon =& k\sigma_{\hat{\theta}} \to 0, \text{ as } n \to \infty, \\ P(|\hat{\theta} - \theta| > \epsilon) =& P(|\hat{\theta} - \theta| > k\sigma_{\hat{\theta}}) \\ \leq& 1/k^2 \ \forall\ k > 0 \end{aligned}$$

=> $\lim_{n\to\infty} P(|\hat{\theta} - \theta| > \epsilon) \leq 1/k^2 \ \forall\ k > 0$ and $\forall\ \epsilon > 0$.

=> $\lim_{n\to\infty} P(|\hat{\theta}-\theta| > \epsilon) = 0 \ \forall\ \epsilon > 0$.

Theorem 9.2. If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two estimators. $\hat{\theta}_i \xrightarrow{P} \theta_i$, $i = 1, 2$, then

(1) $\hat{\theta}_1 + \hat{\theta}_2 \xrightarrow{P} \theta_1 + \theta_2$

(2) $\hat{\theta}_1\hat{\theta}_2 \xrightarrow{P} \theta_1\theta_2$

(3) $\hat{\theta}_1/\hat{\theta}_2 \xrightarrow{P} \theta_1/\theta_2$ if $\theta_2 \neq 0$;

(4) $g(\hat{\theta}_1) \xrightarrow{P} g(\theta_1)$ if g is continuous at θ_1.

Q: Can we add in Th.9.2 $\hat{\theta}_1 - \hat{\theta}_2 \xrightarrow{P} \theta_1 - \theta_2$?? **why ?**

Ex. 9.2. Show that $\overline{Y}_n = \sum_{i=1}^n Y_i/n$ is a consistent estimator of μ_y if Y_1,..., Y_n are i.i.d. and σ_Y is finite.

Proof. Since $V(\hat{\mu}) = V(\overline{Y}) = \sigma_Y^2/n \to 0$, and $E(\overline{Y}) = \mu_Y$ ($\overline{Y}$ is unbiased),

by Th$\dot{9}$.1, $\hat{\mu} = \overline{Y}_n$ is consistent.

Ex. 9.3. Suppose that Y_1, ..., $Y_n \overset{i.i.d.}{\sim} Y$, $E(Y_i^k) = m_k$'s are finite for $k = 1, 2, 4$. Show that $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2$ is a consistent estimator of σ^2.

Proof. $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2$

$= \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^n (Y_i - \overline{Y})^2$

$= \frac{n}{n-1}[\frac{1}{n}\sum_{i=1}^n Y_i^2 - (\overline{Y})^2]$

$= \frac{n}{n-1}[\overline{Y_i^2} - (\overline{Y})^2]$

$\overline{Y} \xrightarrow{P} E(Y)$ by Ex.9.2.

$\overline{Y^2} \xrightarrow{P} E(Y^2)$ by Ex.9.2.

$(\overline{Y})^2 \xrightarrow{P} (E(Y))^2$, by (???) of Th.9.2.

$\frac{n}{n-1} \to 1$ or $\frac{n}{n-1} \xrightarrow{P} 1$??

So $S^2 = \frac{n}{n-1}[\overline{Y^2} - (\overline{Y})^2]$

$\xrightarrow{P} E(Y^2) - (E(Y))^2 = \sigma^2$ by (???) of Thereom 9.2

That is, S^2 is a consistent estimator of σ^2.

Th. 9.3. If $P(U_n \le t) \to \Phi(t)$, the cdf of $(N(0,1))$, and $W_n \xrightarrow{P} 1$, then $P(U_n/W_n \le t) \to \Phi(t)$.

Example 9.4. Let Y_1, ..., Y_n be a random sample from a distribution with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2$. Let $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2$. Show that the cdf of $\sqrt{n}\frac{\overline{Y}-\mu_Y}{S_n}$ converges to $\Phi(t)$, the cdf of N(0,1).

Sol. By the CLT, $P(\sqrt{n}\frac{\overline{Y}-\mu_Y}{\sigma_Y} \le t) \to \Phi(t)$.

$S_n^2 \xrightarrow{P} \sigma^2$.

$S_n^2/\sigma^2 \xrightarrow{P} 1$

$P(\sqrt{n}\frac{\overline{Y}-\mu_Y}{S_n} \le t) = P(\sqrt{n}\frac{\overline{Y}-\mu_Y}{\sigma}\frac{\sigma}{S_n} \le t) \to \Phi(t)$ by Theorem 9.3.

As applications, if n is large, an approximate CI for μ is $\overline{X} \pm z_{\alpha/2}S_n/\sqrt{n}$,

an approximate CI for p is $\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$.

§**9.4. Sufficiency.**

Data are often quite large and not convenient to handle. There is a way to simplify it without losing information about the parameter θ.

Def. 9.3. Let Y_1, ..., Y_n denote a random sample from a distribution with unknown parameter θ. Then the statistic $U = g(Y_1, ..., Y_n)$ is said to be sufficient for θ if the conditional distribution of Y_1, ..., Y_n given U, does not depend on θ, *i.e.*, $f_{Y_1,...,Y_n|U}(y_1, ..., y_n|u)$ does not depend on θ.

Ex. 1. Suppose that X_1, ..., X_n are i.i.d. from bin(1,p). Then $X_i \in \{0, 1\}$. Let $Y = \sum_{i=1}^n X_i$.
The distribution of Y is ?
The conditional distribution of $\mathbf{X} = (X_1, ..., X_n)$ for given Y, say $f_{\mathbf{X}|Y}(x_1, ..., x_n|y)$ is

$$\begin{aligned} f_{\mathbf{X}|Y}(x_1, ..., x_n|y) =& P(X_1 = x_1, ..., X_n = x_n)|Y = y) \\ =& \frac{P(X_1 = x_1, ..., X_n = x_n, Y = y)}{P(Y = y)} \\ =& \frac{\prod_{i=1}^n p_i^x (1-p)^{1-x_i} \mathbf{1}(\sum_{i=1}^n x_i = y)}{\binom{n}{y} p^y (1-p)^{n-y}} \\ =& \begin{cases} \frac{p^y(1-p)^{n-y}}{\binom{n}{y}p^y(1-p)^{n-y}} = \frac{1}{\binom{n}{y}} & \text{if } \sum_{i=1}^n x_i = y,\ x_i \in \{0, 1\} \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

which is independent of the parameter p.

The advantage of the sufficient statistic Y (such as in the above example) is that it simplifies the data if one just wants to make inference about θ. In Ex. 1 above, the original data is $(X_1, ..., X_n)$, and Y is a sufficient statistic. Y is much simpler than $(X_1, ..., X_n)$, in terms of recording and manuscripting (in particular if $n \geq 10^3$).

Def. 9.4. Let y_1, ..., y_n be the sample observations taken on corresponding r.v.s Y_1, ..., Y_n whose distribution depends on a parameter θ. Then the likelihood of the sample, denoted by
$$L(y_1, ..., y_n|\theta) \stackrel{def}{=} \begin{cases} \text{the joint probability of } y_1, ..., y_n & \text{if } Y_i\text{s are discrete.} \\ \text{the joint density of } y_1, ..., y_n & \text{if } Y_i\text{s are continuous r.v.s} \end{cases}$$

For simplification, we write $L(\theta) = L(y_1, ..., y_n|\theta) = L(\vec{y}|\theta)$.

Ex.1 (continued). The likelihood function of $(X_1, ..., X_n)$ for given observations $(x_1, ..., x_n)$, is $L(p) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i}$.

It is OK to write $L(p) = \prod_{i=1}^n p^{X_i} q_i^{1-X_i}$.

Th. 9.4. Let U be a statistic based on the random sample Y_1, ..., Y_n. Then U is a sufficient statistic for the estimation of θ iff $L(\theta)$ can be factored into two nonnegative functions

$$L(\theta) = g(u, \theta)h(y_1, ..., y_n)$$

where $g(u,\theta)$ is only the function of (u,θ) and $h()$ does not depend on θ.

Ex.1 (continued).

$$
\begin{aligned}
L(p) =& \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\
=& p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \\
=& p^y(1-p)^{n-y} \\
=& \underbrace{p^y(1-p)^{n-y}}_{g(y,p)} \times \underbrace{\mathbf{1}(y=\sum_{i=1}^{n} x_i)}_{h(x_1,\ldots,x_n)}.
\end{aligned}
$$

(a) Thus $Y=\sum_{i=1}^{n} X_i$ is sufficient. (b) Thus $y=\sum_{i=1}^{n} x_i$ is sufficient.

Which of (a) and (b) is a correct answer ?

Ex. 9.5. Let Y_1, ..., Y_n be i.i.d from the density function $f(y|\theta) = \begin{cases} \frac{1}{\theta}e^{-y/\theta} & \text{if } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$, where $\theta > 0$. Show that $\overline{Y}$ is a sufficient statistic for θ.

Sol. Two approaches:

$$
\begin{aligned}
L(\theta) =& \prod_{i=1}^{n} f(y_i|\theta) \\
=& \frac{e^{-y_1/\theta}}{\theta} \times \frac{e^{-y_2/\theta}}{\theta} \times \cdots \times \frac{e^{-y_n/\theta}}{\theta} \\
=& \frac{e^{-\sum_{i=1}^{n} y_i/\theta}}{\theta^n} \\
=& \frac{e^{-n\overline{y}/\theta}}{\theta^n} \\
=& \underbrace{\frac{e^{-n\overline{y}/\theta}}{\theta^n}}_{=g(\overline{y},\theta)} \underbrace{\times 1}_{=h(y_1,\ldots,y_n)}
\end{aligned}
\tag{1}
$$

Thus $\overline{Y}$ (? or $\overline{y}$?) is a sufficient statistic.

$$
\begin{aligned}
L(\theta) =& \prod_{i=1}^{n} f(y_i|\theta) = \frac{e^{-y_1/\theta}}{\theta}\mathbf{1}(y_1 \geq 0) \times \frac{e^{-y_2/\theta}}{\theta}\mathbf{1}(y_2 \geq 0) \times \cdots \times \frac{e^{-y_n/\theta}}{\theta}\mathbf{1}(y_n \geq 0) \\
=& \frac{e^{-\sum_{i=1}^{n} y_i/\theta}}{\theta^n}\mathbf{1}(y_{(1)} \geq 0) \quad (y_{(1)} = y_1??) \\
=& \underbrace{\frac{e^{-n\overline{y}/\theta}}{\theta^n}}_{=g(\overline{y},\theta)} \underbrace{\mathbf{1}(y_{(1)} \geq 0)}_{=h(y_1,\ldots,y_n)}.
\end{aligned}
\tag{2}
$$

Thus $\overline{Y}$ is a sufficient statistic.

Q: Are these two approaches both correct ?

Ex.3. Suppose that X_1, ..., X_n are i.i.d. from $U(a, b)$, find a sufficient statistic.
Sol. $X_i \sim f(x) = \frac{1}{b-a}\mathbf{1}(x \in [a, b])$.

$$\begin{aligned} L(a, b) =& \prod_{i=1}^{n} \frac{1}{b-a}\mathbf{1}(x_i \in [a, b]) \\ =& \frac{1}{(b-a)^n}\mathbf{1}(x_{(1)} \geq a)\mathbf{1}(x_{(n)} \leq b) \\ =& \underbrace{\frac{1}{(b-a)^n}\mathbf{1}(x_{(1)} \geq a)\mathbf{1}(x_{(n)} \leq b)}_{g(x_{(1)}, x_{(n)}, a, b)} \qquad h(x_1, ..., x_n) = ? \end{aligned}$$

A sufficient statistic is

$(x_{(1)}, x_{(n)})$?
$\vec{Y} = (X_{(1)}, X_{(n)})$?
$\vec{y} = (x_{(1)}, x_{(n)})$?
Class exercies.

Ex.4. Suppose that X_1, ..., X_n are i.i.d. from $U(0, b)$, find a sufficient statistic.
Sol. $X_i \sim f(x) = \frac{1}{b}\mathbf{1}(x \in [0, b])$.

$$\begin{aligned} L(a, b) =& \prod_{i=1}^{n} \frac{1}{b}\mathbf{1}(x_i \in [0, b]) \\ =& \frac{1}{(b)^n}\mathbf{1}(x_{(1)} \geq 0)\mathbf{1}(x_{(n)} \leq b) \\ =& \frac{1}{(b)^n}\mathbf{1}(x_{(1)} \geq 0)\mathbf{1}(x_{(n)} \leq b). \quad h(x_1, ..., x_n) = ? \quad sufficient\ statistic = ? \end{aligned}$$

Review

The practice test is in "homework solution (pdf file)".

Ex. R1. Suppose that X_1, ..., $X_n \overset{i.i.d.}{\sim} U(0, \theta)$.

1. Show that $\tilde{\theta} = 2\overline{X}$ is an unbiased estimator of θ.
2. Show that $\tilde{\theta}$ is a consistent estimator of θ.
3. Show that $\hat{\theta} = \frac{n+1}{n}X_{(n)}$ is an unbiased estimator of θ.
4. Show that $\hat{\theta}$ is a consistent estimator of θ.
5. Compute eff$(\hat{\theta}, \tilde{\theta})$.

Sol. 1. $\vdash$: $\tilde{\theta} = 2\overline{X}$ is an unbiased estimator of θ. **Class exercise.** Hint: $\frac{a+b}{2}$ *by* 447 [21]

$$\begin{aligned} E(\tilde{\theta}) =& E(2\overline{X}) \\ =& 2E(X) \end{aligned}$$

$$
\begin{aligned}
&=2 \times \frac{0+\theta}{2} && \tfrac{a+b}{2} \textit{ by } 447\ [21] \\
&=\theta.
\end{aligned}
$$

That is,$\tilde{\theta} = 2\overline{X}$ is an unbiased estimator of θ.

2. $\vdash$: $\tilde{\theta} = 2\overline{X}$ is a consistent estimator of θ.

By Tchebysheffs Inequality (447 [14]), it suffices to show $V(\tilde{\theta}) \to 0$ as $n \to \infty$. **Class exercise.**

$$
\begin{aligned}
V(\tilde{\theta}) &= V(2\overline{X}) && \tfrac{(b-a)^2}{12} \textit{ by } 447\ [21] \\
&=4V(\overline{X}) \\
&=4V(X)/n \\
&=4\theta^2/(12n) && \tfrac{(b-a)^2}{12} \textit{ by } 447\ [21] \\
&=\theta^2/(3n) \to 0 && \text{as } n \to \infty,
\end{aligned}
$$

Thus $\tilde{\theta}$ is consistent by Tchebysheffs Inequality (447 [14]).

3. $\vdash$: $\hat{\theta} = \frac{n+1}{n} X_{(n)}$ is an unbiased estimator of θ.

By 448 [6], $f(x) = \frac{n!}{(n-1)!1!(n-n)!}(\frac{x}{\theta})^{n-1}\frac{1}{\theta}(1-\frac{x}{\theta})^{n-n}$. **Class exercise:**

Thus $E(X_{(n)}) = \int_0^\theta x \frac{n!}{(n-1)!1!(n-n)!}(\frac{x}{\theta})^{n-1}\frac{1}{\theta}(1-\frac{x}{\theta})^{n-n} dx$

$$
\begin{aligned}
E(X_{(n)}) &= \int_0^\theta xn(\frac{x}{\theta})^{n-1}\frac{1}{\theta}dx \\
&= n\int_0^\theta (\frac{x}{\theta})^n dx \\
&= n\theta \int_0^\theta (\frac{x}{\theta})^n dx/\theta \\
&= n\theta \int_0^1 y^n dy \\
E(X_{(n)}) &= \frac{n\theta}{n+1} y^{n+1}\Big|_0^1 \\
&= \frac{n}{n+1}\theta
\end{aligned}
$$

$$
\textit{Then } \ E(\hat{\theta}) = E(\frac{n+1}{n}X_{(n)}) = \frac{n+1}{n}E(X_{(n)}) = \frac{n+1}{n}\frac{n}{n+1}\theta = \theta.
$$

Thus $\hat{\theta} = \frac{n+1}{n} X_{(n)}$ is an unbiased estimator of θ.

4. $\vdash$: $V(\hat{\theta}) \to 0$ and $\hat{\theta}$ is a consistent estimator of θ.

$$
\begin{aligned}
E(X_{(n)}^2) &= \int_0^\theta x^2 n(\frac{x}{\theta})^{n-1}\frac{1}{\theta}dx \\
&= n\int_0^\theta (\frac{x^{n+1}}{\theta^n})dx
\end{aligned}
$$

$$
\begin{aligned}
&= n\theta^2 \int_0^1 y^{n+1} dy \\
&= \frac{n\theta^2}{n+2} y^{n+2} \Big|_0^1 \\
&= \frac{n}{(n+2}\theta^2 \\
V(X_{(n)}) &= E(X_{(n)}^2) - (E(X_{(n)})^2 \\
&= \frac{n}{n+2}\theta^2 - (\frac{n}{n+1}\theta)^2 \\
&= [\frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2}]\theta^2 \\
&= [\frac{n}{(n+2)(n+1)^2}]\theta^2 \\
V(\hat{\theta}) &= V(\frac{n+1}{n} X_{(n)}) \\
&= (\frac{n+1}{n})^2 V(X_{(n)}) \\
&= (\frac{n+1}{n})^2 [\frac{n}{(n+2)(n+1)^2}]\theta^2 \\
&= ([\frac{1}{n(n+2)}]\theta^2 \to 0 \text{ as } n \to \infty
\end{aligned}
$$

Thus $\hat{\theta}$ is consistent by Tchebysheffs Inequality (447 [14]).

5. Compute $eff(\hat{\theta}, \tilde{\theta})$.

$$
eff(\hat{\theta}, \tilde{\theta}) = \frac{V(\tilde{\theta})}{V(\hat{\theta})} = \frac{\theta^2/(3n)}{\theta^2/(n(n+2)} = \frac{n+2}{3}.
$$

Ex. R2. Suppose that Y_1, ..., $Y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ (see 447 [22]). Let $n = 2k$ for $k = 1$, 2, ..., $\theta = \sigma^2$, and

$$
\hat{\theta} = \frac{1}{2k} \sum_{i=1}^{k} (Y_{2i} - Y_{2i-1})^2
$$

1. Show that $\hat{\theta}$ is unbiased estimator of σ^2.
2. Show that $\hat{\theta}$ is a consistent estimator of σ^2.
3. Compute $eff(\hat{\theta}, S^2)$.

Sol. 1. Show that $\hat{\theta}$ is unbiased estimator of σ^2.

$$
\hat{\theta} = \frac{1}{2k} \sum_{i=1}^{k} (Y_{2i} - Y_{2i-1})^2
$$

$$
\begin{aligned}
&= \frac{1}{2}[\sum_{i=1}^{k} \frac{Y_{2i}^2}{k} + \sum_{i=1}^{k} \frac{Y_{2i-1}^2}{k} - 2\frac{1}{k}\sum_{i=1}^{k}(Y_{2i}Y_{2i-1})] \\
E(\hat{\theta}) &= \frac{1}{2}[E(Y^2) + E(Y^2) - 2E(Y_1Y_2)] \\
&= \frac{1}{2}[E(Y^2) + E(Y^2) - 2E(Y_1)E(Y_2)] \\
&= \frac{1}{2}[E(Y^2) + E(Y^2) - 2(E(Y))^2] \\
&= E(Y^2) - (E(Y))^2 \\
&= \sigma^2.
\end{aligned}
$$

Thus $\hat{\theta}$ is unbiased estimator of σ^2.

2. Show that $\hat{\theta}$ is a consistent estimator of σ^2.

Since Y_{2i} and Y_{2i-1} are i.i.d., $Y_{2i} - Y_{2i-1} \sim N(\mu - \mu, \sigma^2 + \sigma^2)$ by 447 [44].

44. If X_1___X_2.

X_i's $\sim$:	$X_1 + X_2 \sim$:
$\mathcal{G}(\alpha_i, \beta)$	____________
$\chi^2(v_i)$	____________
$Pois(\lambda_i)$	____________
$N(\mu_i, \sigma_i^2)$	____________
$bin(n_i, p)$	____________

key: $\perp$, $\mathcal{G}(\alpha_1 + \alpha_2, \beta)$, $\chi^2(v_1 + v_2)$, $Pois(\lambda_1 + \lambda_2)$, $N(\mu_x + \mu_y, \sigma_1^2 + \sigma_2^2)$, $bin(n_1 + n_2, p)$

That is, $Y_{2i} - Y_{2i-1} \sim N(0, 2\sigma^2)$ and $\frac{Y_{2i}-Y_{2i-1}}{\sqrt{2}\sigma} \sim N(0,1)$.

$\vdash$: If $X \sim N(0,1)$, then $X^2 \sim \chi^2$.

$$
\begin{aligned}
P(X^2 \le t) &= P(-\sqrt{t} \le X \le \sqrt{t}) \\
&= \int_{-\sqrt{t}}^{\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx && \textit{by } 447\ [22] \\
&= 2\int_{0}^{\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\
&= 2\int_{0}^{\sqrt{t}} \frac{1}{x\sqrt{2\pi}} e^{-x^2/2} dx^2/2 \\
&= 2\int_{0}^{t/2} \frac{1}{\sqrt{2y}\sqrt{2\pi}} e^{-y} dy && \text{where } y = x^2/2 \\
&= \int_{0}^{t/2} \frac{1}{\sqrt{2y}\sqrt{2\pi}} e^{-2y/2} d(2y) \\
&= \int_{0}^{t} \frac{1}{\sqrt{u}\sqrt{2\pi}} e^{-u/2} du && \text{where } u = 2y \\
&= \int_{0}^{t} \frac{u^{0.5-1}}{\sqrt{2\pi}} e^{-u/2} du && \textit{see } 447\ [23]\ [24]
\end{aligned}
$$

Thus it is the $\chi^2(1)$ or $Gamma(1/2,2)$. Then

$$\begin{aligned}\hat{\theta} =&\frac{1}{2k}\sum_{i=1}^{k}(Y_{2i}-Y_{2i-1})^2\\ =&\frac{\sigma^2}{k}\sum_{i=1}^{k}\frac{(Y_{2i}-Y_{2i-1})^2}{2\sigma^2}\\ =&\frac{\sigma^2}{k}Z && (Z\overset{def}{=}\textstyle\sum_{i=1}^{k}\frac{(Y_{2i}-Y_{2i-1})^2}{2\sigma^2}\sim\chi^2(k))\end{aligned}$$

where $Z=\sum_{i=1}^{k}\frac{(Y_{2i}-Y_{2i-1})^2}{2\sigma^2}\sim\chi^2(k)$, the χ^2 with k degree freedoms, or $Gamma(k/2,2)$, with the mean $(k/2)2=k$ and the variance $(k/2)2^2$. Then $\hat{\theta}$ is an unbiased estimator of θ and $V(\hat{\theta})=(\frac{\sigma^2}{k})^2(2k)\to 0$ if $k\to\infty$. Thus $\hat{\theta}$ is consistent.

3. Compute $eff(\hat{\theta},S^2)$ **Class execise** Hint: $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$ 447 [41].

Ex. R3. Suppose that X_1, ..., $X_n\overset{i.i.d.}{\sim}U(\theta,1)$.

1. Show that $\tilde{\theta}=2\overline{X}-1$ is an unbiased estimator of θ.
2. Show that $\tilde{\theta}$ is a consistent estimator of θ.
3. Show that $\hat{\theta}=\frac{(n+1)X_{(1)}-1}{n}$ is an unbiase estimator of θ.
4. Show that $\hat{\theta}$ is a consistent estimator of θ.
5. Compute eff$(\hat{\theta},\tilde{\theta})$.

Sol. 1. $\vdash$: $\tilde{\theta}=2\overline{X}-1$ is an unbiased estimator of θ. **Class exercise.** Hint: $\frac{a+b}{2}$ *by* 447 [21]

$$\begin{aligned}E(\tilde{\theta}) =&E(2\overline{X}-1)\\ =&2E(X)-1\\ =&2\times\frac{\theta+1}{2}-1 && \tfrac{a+b}{2}\ \textit{by}\ 447\ [21]\\ =&\theta.\end{aligned}$$

That is,$\tilde{\theta}=2\overline{X}-1$ is an unbiased estimator of θ.

2. $\vdash$: $\tilde{\theta}=2\overline{X}-1$ is a consistent estimator of θ.

By Tchebysheffs Inequality (447 [14]), it suffices to show $V(\tilde{\theta})\to 0$ as $n\to\infty$. **Class exercise.**

$$\begin{aligned}V(\tilde{\theta}) =&V(2\overline{X}) && \tfrac{(b-a)^2}{12}\ \textit{by}\ 447\ [21]\\ =&4V(\overline{X})\\ =&4V(X)/n\\ =&4(1-\theta)^2/(12n) && \tfrac{(b-a)^2}{12}\ \textit{by}\ 447\ [21]\\ =&(1-\theta)^2/(3n)\to 0 && \text{as } n\to\infty,\end{aligned}$$

Thus $\tilde{\theta}$ is consistent by Tchebysheffs Inequality (447 [14]).

3. $\vdash$: $\hat{\theta} = \frac{(n+1)X_{(1)}-1}{n}$ is an unbiased estimator of θ.

By 448 [6], $f(x) = \frac{n!}{(1-1)!1!(n-1)!}(\frac{x-\theta}{1-\theta})^{1-1}\frac{1}{1-\theta}(1-\frac{x-\theta}{1-\theta})^{n-1}$. **Class exercise:**

Thus $E(X_{(1)}) = \int_\theta^1 x\frac{n!}{(1-1)!1!(n-1)!}(\frac{x-\theta}{1-\theta})^{1-1}\frac{1}{1-\theta}(1-\frac{x-\theta}{1-\theta})^{n-1}dx$

$$\begin{aligned}
E(X_{(1)}) &= \int_\theta^1 xn(\frac{x-\theta}{1-\theta})^{1-1}\frac{1}{1-\theta}(1-\frac{x-\theta}{1-\theta})^{n-1}dx \\
&= n\int_\theta^1 x\frac{1}{1-\theta}(\frac{1-\theta-(x-\theta)}{1-\theta})^{n-1}dx \\
&= n\int_\theta^1 x\frac{1}{1-\theta}(\frac{1-x}{1-\theta})^{n-1}dx \\
&= n(1-\theta)^{-n}\int_\theta^1 x(1-x)^{n-1}dx \qquad (1) \\
&= n(1-\theta)^{-n}\int_0^{1-\theta}(1-y)(y)^{n-1}dy \qquad y = 1-x \\
&= n(1-\theta)^{-n}(\frac{y^n}{n}-\frac{y^{n+1}}{n+1})\Big|_0^{1-\theta} \\
&= n(1-\theta)^{-n}(\frac{(1-\theta)^n}{n}-\frac{(1-\theta)^{n+1}}{n+1}) \\
&= n(\frac{1}{n}-\frac{(1-\theta)}{n+1}) \\
E(X_{(1)}) &= \frac{n\theta+1}{n+1} \\
\frac{(n+1)E(X_{(1)})-1}{n} &= \theta
\end{aligned}$$

Thus $\hat{\theta} = \frac{(n+1)X_{(1)}-1}{n}$ is an unbiased estimator of θ.

4. Show that $\hat{\theta}$ is a consistent estimator of θ. Need to find $V(\hat{\theta})$ ($V(X) = E(X^2) - (E(X))^2$).

$$\begin{aligned}
E(X_{(1)}^2) &= \int_\theta^1 x^2 n(\frac{x-\theta}{1-\theta})^{1-1}\frac{1}{1-\theta}(1-\frac{x-\theta}{1-\theta})^{n-1}dx \\
&= n(1-\theta)^{-n}\int_\theta^1 x^2(1-x)^{n-1}dx \qquad \textit{by Eq.}(1) \\
&= n(1-\theta)^{-n}\int_0^{1-\theta}(1-y)^2(y)^{n-1}dy \quad y = 1-x \\
&= n(1-\theta)^{-n}\int_0^{1-\theta}(1-2y+y^2)(y)^{n-1}dy \\
&= n(1-\theta)^{-n}(\frac{y^n}{n}-2\frac{y^{n+1}}{n+1}+\frac{y^{n+2}}{n+2})\Big|_0^{1-\theta} \\
&= n(1-\theta)^{-n}[\frac{(1-\theta)^n}{n}-2\frac{(1-\theta)^{n+1}}{n+1}+\frac{(1-\theta)^{n+2}}{n+2}] \\
E(X_{(1)}^2) &= n(\frac{1}{n}-2\frac{(1-\theta)}{n+1}+\frac{(1-\theta)^2}{n+2})
\end{aligned}$$

$$
\begin{aligned}
0 \le V(\hat\theta) =& V(\frac{(n+1)X_{(1)}-1}{n})\\
=&(\frac{(n+1)}{n})^2 V(X_{(1)})\\
=&(\frac{(n+1)}{n})^2 [E(X_{(1)}^2)-(E(X_{(1)}))^2]\\
=&(\frac{(n+1)}{n})^2 [n(\frac{1}{n}-2\frac{(1-\theta)}{n+1}+\frac{(1-\theta)^2}{n+2})-(\frac{n\theta+1}{n+1})^2]\\
=&(\frac{(n+1)}{n})^2 [(\frac{n}{n}-2\frac{(1-\theta)n}{n+1}+\frac{(1-\theta)^2 n}{n+2})-(\frac{n\theta+1}{n+1})^2]\\
=&(\frac{(n+1)}{n})^2 [(\frac{n}{n}-2\frac{(1-\theta)(n+1-1)}{n+1}+\frac{(1-\theta)^2(n+2-2)}{n+2})-(\frac{(n+1-1)\theta+1}{n+1})^2]\\
=&(\frac{(n+1)}{n})^2 [(1-2(1-\theta)+2\frac{(1-\theta)}{n+1}+(1-\theta)^2-\frac{(1-\theta)^2 2}{n+2}-(\theta+\frac{1-\theta}{n+1})^2]\\
=&(\frac{(n+1)}{n})^2 [(1-2(1-\theta)+(1-\theta)^2+2\frac{(1-\theta)}{n+1}-\frac{(1-\theta)^2 2}{n+2}-(\theta+\frac{1-\theta}{n+1})^2]\\
=&(\frac{(n+1)}{n})^2 [(1-(1-\theta))^2+2\frac{(1-\theta)}{n+1}-\frac{(1-\theta)^2 2}{n+2}-(\theta+\frac{1-\theta}{n+1})^2]\\
=&(\frac{(n+1)}{n})^2 [(\theta)^2+2\frac{(1-\theta)}{n+1}-\frac{(1-\theta)^2 2}{n+2}-\theta^2+2\theta(\frac{1-\theta}{n+1})-(\frac{1-\theta}{n+1})^2]\\
=&(\frac{(n+1)}{n})^2 [2\frac{(1-\theta)}{n+1}-\frac{(1-\theta)^2 2}{n+2}+2\theta(\frac{1-\theta}{n+1})-(\frac{1-\theta}{n+1})^2] \qquad \to 0\\
=&(\frac{(n+1)}{n})^2 [2\frac{(1-\theta)}{n+1}+2\theta(\frac{1-\theta}{n+1})-\frac{(1-\theta)^2 2}{n+2}-(\frac{1-\theta}{n+1})^2]\\
=&(\frac{(n+1)}{n})^2 [2(1+\theta)\frac{(1-\theta)}{n+1}-(\frac{1-\theta}{n+1})^2-\frac{(1-\theta)^2 2}{n+2}]\\
=&(\frac{(n+1)}{n})^2 [2\frac{(1-\theta^2)}{n+1}-(1-\theta)^2[(\frac{1}{n+1})^2+\frac{2}{n+2}]]\\
\to& 0 \text{ as } n\to\infty.
\end{aligned}
$$

Thus $\hat\theta$ is consistent by Tchebysheff's inequality.

5. Compute eff$(\hat\theta, \tilde\theta)$.

$$
eff(\hat\theta,\tilde\theta)=\frac{(1-\theta)^2/(3n)}{(\frac{(n+1)}{n})^2[2\frac{(1-\theta^2)}{n+1}-(1-\theta)^2[(\frac{1}{n+1})^2+\frac{2}{n+2}]]}.
$$

Review on consistency: Suppose that X has a uniform distribution on the interval $[1,2]$. Let

$\theta = 0$,

$\theta_o = \mathbf{1}(X=1)$ and

$\hat\theta_n = \mathbf{1}(X\in[1,1+\frac{1}{n}])$.

Q: $\theta_o = \theta$? Yes, No, DNK

$P(\theta_o = \theta) = 0$? Yes, No, DNK

$P(\theta_o = \theta) = 1$? Yes, No, DNK

$P(\hat{\theta}_n = \theta) = 1/n$? Yes, No, DNK

$\hat{\theta}_n \to \theta$? Yes, No, DNK

$\hat{\theta}_n \to \theta_o$? Yes, No, DNK

$\hat{\theta}_n \xrightarrow{P} \theta_o$? Yes, No, DNK

$\hat{\theta}_n \xrightarrow{P} \theta$? Yes, No, DNK

Summary on CI:

Large sample CI: for θ: $[\hat{\theta}_L, \hat{\theta}_U]$ based on $Z = \frac{\hat{\theta}-\theta}{\hat{\sigma}_{\hat{\theta}}} \sim N(0,1)$ approximately. **class exercise**

$$\begin{aligned}
1-\alpha \approx & P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) \\
= & P(-z_{\alpha/2} \le \frac{\hat{\theta}-\theta}{\hat{\sigma}_{\hat{\theta}}} \le z_{\alpha/2}) \\
= & P(\hat{\theta} - z_{\alpha/2}\hat{\sigma}_{\hat{\theta}} \le \theta \le \hat{\theta} + z_{\alpha/2}\hat{\sigma}_{\hat{\theta}})
\end{aligned}$$

Small sample CI:

There are 4 typical pivotal functions $Z = g(X, \theta)$:

1. $X \sim N(\mu, \sigma^2)$, $Z = X - \mu \sim N(0, \sigma^2)$ or $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$.
2. $X \sim Exp(\theta)$ $(E(X) = \theta)$, $Z = g(X, \theta) = \frac{X}{\theta} \sim Exp(1)$.
3. $X \sim U(\theta, \theta + b)$, $Z = g(X, \theta) = \frac{X-\theta}{b} \sim U(0,1)$.
4. For σ under i.i.d. $N(\mu, \sigma^2)$: : $W = (n-1)S_x^2/\sigma^2 \sim \underline{\ \chi^2_{n-1}\ }$,

R.4. Suppose that X_1,..., X_n are i.i.d. from $N(\mu, \sigma^2)$,
find an unbiased estimator of σ based on S (need to prove it).
Recall that $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ is an unbased estimator of σ^2.

Sol. Recall 447 [41], [23], [24]. If $Y \sim N(\mu, \sigma^2)$, $\frac{\overline{Y}-\mu_{\overline{Y}}}{\sigma_{\overline{Y}}} \sim \underline{N(0,1)}$, $\frac{(n-1)S^2}{\sigma^2} \sim \underline{\chi^2(n-1)}$, $\sqrt{n}\frac{\overline{Y}-\mu}{S} \sim \underline{t_{n-1}}$, where $\mu_{\overline{Y}} = \underline{\mu}$, $\sigma^2_{\overline{Y}} = \underline{\sigma^2/n}$

$$\begin{aligned}
\textit{Sol.} \qquad E(S) = & E(\sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2}) \\
= & \sqrt{\frac{1}{n-1}}\sigma E(\sqrt{\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \overline{X})^2}) \\
= & \sqrt{\frac{1}{n-1}}\sigma E(\sqrt{Y}) \qquad\qquad Y \sim \chi^2(n-1) = G(\tfrac{n-1}{2}, 2) \\
= & \sqrt{\frac{1}{n-1}}\sigma \int_0^\infty \sqrt{y}\frac{y^{\frac{n-1}{2}-1}e^{-y/2}}{\Gamma(\frac{n-1}{2})2^{\frac{n-1}{2}}}dy
\end{aligned}$$

$$=\sqrt{\frac{1}{n-1}}\sigma\int_0^\infty \frac{y^{\frac{n}{2}-1}e^{-y/2}}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}dy\frac{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}{\Gamma(\frac{n-1}{2})2^{\frac{n-1}{2}}} \quad \textbf{why do this ??}$$

$$=\sqrt{\frac{1}{n-1}}\sigma\frac{\Gamma(\frac{n}{2})2^{\frac{n}{2}}}{\Gamma(\frac{n-1}{2})2^{\frac{n-1}{2}}}$$

$$=\sigma\sqrt{\frac{2}{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}$$

$$Let\ \ \tilde\sigma = \frac{1}{\sqrt{\frac{2}{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}}S. \text{ Then } \tilde\sigma \text{ is unbiased.}$$

§9.5. The Rao-Blackwell Th. and Minimum-Variance Unbiased Estimator

Let $\hat\theta$ be an estimator of θ.

It is desirable that an estimator satisfies:

1. It is unbiased: $E(\hat\theta) = \theta$;
2. It is consistent: $\lim_{n\to\infty} P(|\hat\theta - \theta| > \epsilon) = 0 \ \forall\ \epsilon > 0$;
3. $eff(\hat\theta, \hat\theta_2) = \frac{V(\hat\theta_2)}{V(\hat\theta)} \geq 1\ \forall$ unbiased estimator $\hat\theta_2$.

Def. An estimator satisfying the above 3 properties is called the minimum variance unbiased estimator (MVUE).

Q: How to find an MVUE of θ ?

Ans: [11] If $X_1, ..., X_n$ are i.i.d. from $f(x;\theta) = exp\{\underline{\qquad\qquad} + g(\theta) + h(x)\}$, $\hat\gamma = G(\underline{\qquad\qquad})$ and $\underline{\qquad\qquad} = \gamma(\theta)$, then $\hat\gamma$ is the MVUE of γ.

key: $\underline{T(x)\psi(\theta)}$, $\underline{\sum_i T(X_i))}$, $\underline{E(\hat\gamma)}$,

This is due to

Theorem 9.5. (The Rao-Blackwell Th.) Let $\hat\theta$ be an unbiased estimator for θ such that $V(\hat\theta) < \infty$. If U is a sufficient statistic for θ. Define $\hat\theta^* = E(\hat\theta|U)$, then for all θ,

$E(\hat\theta^*) = \theta$ and $V(\hat\theta^*) \leq V(\hat\theta)$.

Recall the sufficiency:

Def. 9.3. Let Y_1, ..., Y_n denote a random sample from a distribution with unknown parameter θ. Then the statistic $U = g(Y_1, ..., Y_n)$ is said to be sufficient for θ if the conditional distribution of Y_1, ..., Y_n given U, does not depend on θ, *i.e.*, $f_{Y_1,...,Y_n|U}(y_1, ..., y_n|u)$ does not depend on θ.

The advantage of the sufficient statistic Y (such as in the above example) is that it simplifies the data if one just wants to make inference about θ.

Def. 9.4. Let y_1, ..., y_n be the sample observations taken on corresponding r.v.s Y_1, ..., Y_n whose distribution depends on a parameter θ. Then the likelihood of the sample, denoted by

$$L(y_1, ..., y_n|\theta) \overset{def}{=} \begin{cases} \text{the joint probability of } y_1, ..., y_n & \text{if } Y_i\text{s are discrete.} \\ \text{the joint density of } y_1, ..., y_n & \text{if } Y_i\text{s are continuous r.v.s} \end{cases}$$

For simplification, we write $L(\theta) = L(y_1, ..., y_n|\theta) = L(\vec y|\theta)$.

Th. 9.4. Let U be a statistic based on the random sample Y_1, ..., Y_n. Then U is a sufficient statistic for the estimation of θ iff $L(\theta)$ can be factored into two nonnegative functions

$$L(\theta) = g(u,\theta)h(y_1, ..., y_n)$$

where $g(u,\theta)$ is only the function of (u,θ) and $h()$ does not depend on θ.

Ex. 9.6 Let $Y \sim bin(m,p)$. Is $\hat{p} = Y/m$ an MVUE of p ?
Sol. $n = 1$. $f_Y(y;p) = \binom{m}{y} p^y (1-p)^{m-y}$.

$$\begin{aligned}
f_Y(y;p) =& exp(y\ln p + (m-y)\ln(1-p) + \ln\binom{m}{y}) \\
=& exp(\underbrace{y}_{T(y)} \underbrace{\ln\frac{p}{1-p}}_{\psi(y)} + \underbrace{m\ln(1-p)}_{g(p)} + \underbrace{\ln\binom{m}{y}}_{h(y)}) \qquad \text{Is } Y \text{ a sufficient statistic?} \\
=& exp(T(y)\psi(p) + g(p) + h(y))
\end{aligned}$$

$T(Y) = Y$, $E(T) = E(Y) = mp$, $\hat{p} = T/m = Y/m$. $E(\hat{p}) = p$. Thus $\hat{p}$ is MVUE of p.

Ex. 9.7. Suppose that Y_1, ..., $Y_n \overset{i.i.d.}{\sim} f = \frac{2y}{\theta} e^{-y^2/\theta}$, $y > 0$. MVUE of θ ?

Sol. $\quad n \geq 1, \quad$ $$\begin{aligned}
f =& \frac{2y}{\theta} e^{-y^2/\theta} \mathbf{1}(y > 0) \\
=& exp(\underbrace{y^2}_{T(y)} \underbrace{\frac{-1}{\theta}}_{\psi(\theta)} + \underbrace{\ln(2y)}_{h(y)} + \underbrace{-\ln\theta}_{g(\theta)}) \qquad \text{Is } Y^2 \text{ a sufficient statistic?}
\end{aligned}$$

$E(\sum_{i=1}^n T(Y_i)) = E(\sum_i Y_i^2) = nE(Y^2)=$?

$$\begin{aligned}
E(Y^2) =& \int_0^\infty y^2 \frac{2y}{\theta} e^{-y^2/\theta} dy \\
=& \int_0^\infty \frac{2y^2}{\theta} e^{-y^2/\theta} dy^2/2 \qquad dy^2 = 2ydy \\
=& \int_0^\infty \frac{y^2}{\theta} e^{-y^2/\theta} dy^2 \\
=& \int_0^\infty \frac{u}{\theta} e^{-u/\theta} du \qquad u =? \\
=& \Gamma(2)\theta \underbrace{\int_0^\infty \frac{u^{2-1}}{\Gamma(2)\theta^2} e^{-u/\theta} du}_{\textbf{why do this}?} \\
=& \theta
\end{aligned}$$

$\hat{\gamma} = G(\sum_{i=1}^n T(Y_i)) = \sum_i^n Y_i^2/n$.

$E(\hat{\gamma}) = \theta$.

$\hat{\gamma} = \sum_{i=1}^n Y_i^2/n = \overline{Y^2}$ is a MVUE of θ.

Ex. 9.9. Suppose that Y_1, ..., $Y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. Find the MVUE of (μ, σ^2).

Sol.
$$
\begin{aligned}
f =& \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{\sigma^2}} \\
=& \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2-2\mu x+\mu^2}{\sigma^2}} \\
=& exp(-\frac{x^2 - x(2\mu) + \mu^2}{\sigma^2} + \ln\frac{1}{\sqrt{2\pi\sigma^2}}) \\
=& exp(-\frac{x^2}{\sigma^2} + \frac{x(2\mu)}{\sigma^2} - \frac{\mu^2}{\sigma^2} + \ln\frac{1}{\sqrt{2\pi\sigma^2}}) \\
=& exp(-\underbrace{(x^2, x)}_{T(x)} \underbrace{(\frac{-1}{\sigma^2}, \frac{\mu}{\sigma^2})'}_{\psi(\theta)} - \frac{\mu^2}{\sigma^2} + \ln\frac{1}{\sqrt{2\pi\sigma^2}})
\end{aligned}
$$

$T(x) = (x^2, x)$, $\psi(\theta) = (\frac{-1}{\sigma^2}, \frac{\mu^2}{\sigma^2})$. **Is $T(X)$ a sufficient statistic ?**

$\hat{\gamma} = (\frac{n}{n-1}(\overline{X^2} - (\overline{X})^2), \overline{X})) = (\frac{1}{n-1}(\sum_{i=1}^n (X_i - \overline{X})^2), \overline{X})$,

$\gamma(\theta) = (\mu, \sigma^2)$,

$E(\sum_i T(X_i)) = n(\sigma^2 + \mu^2, \mu)$,

MVUE of μ is $\overline{X}$.

MVUE of σ^2 is $\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$.

Ex. 9.10. Suppose that Y_1, ..., $Y_n \overset{i.i.d.}{\sim} f = \frac{1}{\theta}e^{-y/\theta}$, $y > 0$. MVUE of $V(X)$?

Sol. $\sigma^2 = \theta^2$

$$f = \frac{1}{\theta} e^{-\frac{x}{\theta}} \mathbf{1}(x > 0)$$

$T(x) = x$, $E(X) = \theta$,

$Y = \sum_{i=1}^n X_i \sim G(n, \theta)$.

$E(Y^2) = \sigma_{Y^2}^2 + \mu_{Y^2}^2 = n\theta^2 + (n\theta)^2 = (n + n^2)\theta^2$.

$E(Y^2/(n + n^2) = \theta^2$.

$Y^2/(n + n^2) = (\sum_{i=1}^n X_i)^2/(n + n^2) = \frac{n}{n+1}(\overline{X})^2$.

$\hat{\gamma} = \frac{n}{n+1}(\overline{X})^2$ is the MVUE of θ^2.

§**9.6. The method of moments.**

Q: How to construct an estimator in general.

Ans. Two common methods:

(1) Method of Moments estimator (MME);

(2) Maximum likelihood estimator (MLE).

[12] An MME of θ, is the solution of θ to $\mu_i'(\theta) =$ ________ for k i's, where $\mu_i'(\theta) =$ ________, and k is the dimension of θ. **key:** $\underline{\overline{X^i}}$, $\underline{E(X^i)}$,

Ex. 9.11 Assuming $X_1, \ldots., X_n \overset{i.i.d.}{\sim} U(0, \theta)$, find an MME of θ.

Sol. Since $E(X) = \theta/2$,

set $\overline{X} = \hat{\theta}/2 \Rightarrow \hat{\theta} = 2\overline{X}$.

Q: Can we derive an MME of θ as follows ?

$$\text{Set } \overline{X} = E(X) = \theta/2 \Rightarrow \hat{\theta} = 2\overline{X}.$$

Q: Is $\hat{\theta}$ unbiased ?

$E(\hat{\theta}) = E(2\overline{X}) = 2\frac{\theta - 0}{2} = \theta$. Answer ?

Q: Is $\hat{\theta}$ consistent ?

$V(\hat{\theta}) = V(2\overline{X}) = 4V(X)/n = 4\frac{\theta^2}{12}/n \to 0$, by Tchebysheff's Inequality.

Actually letting $\epsilon = k\frac{\sigma}{\sqrt{3n}}$, $P(|\hat{\theta} - \theta| > k\frac{\sigma}{\sqrt{3n}}) \le 1/k^2$

Thus it is consistent.

Another proof: $\overline{X} \xrightarrow{P} E(X) = \theta/2$ by the law of large numbers (provided $V(X)$ exists).

$2\overline{X} \xrightarrow{P} 2\theta/2 = \theta$.

Q: Can we derive an MME of θ as follows ?

(1) Since

$$\begin{aligned} E(X^2) &= \int_0^\theta x^2 \frac{1}{\theta} dx \\ &= \frac{\frac{x^3}{3}}{\theta}\Big|_0^\theta \\ &= \theta^2/3, \\ set \;\; \overline{X^2} &= \tilde{\theta}^2/3 \quad \Rightarrow \quad \tilde{\theta} = \sqrt{3\overline{X^2}}. \end{aligned}$$

Or (2) Set $\overline{X^2} = E(X^2) = \int_0^\theta x^2 \frac{1}{\theta} dx = \frac{\frac{x^3}{3}}{\theta}\Big|_0^\theta = \theta^2/3 \Rightarrow \tilde{\theta} = \sqrt{3\overline{X^2}}$.

Is $\tilde{\theta}$ unbiased ?

$E(\tilde{\theta}) = \int_0^\theta \sqrt{3t} f_{\overline{X^2}}(t) dt$ is difficult to solve for us, so we ignore the answer.

Is $\tilde{\theta}$ consistent ?

Need to check whether $V(X^2)$ exists.

It suffices to show $E(X^4)$ exists, as $V(X^2) = E(X^4) - (E(X^2))^2$ and $E(X^2)$ exists.

$E(X^4) = \int_0^\theta x^4/\theta dx = \frac{x^5}{5\theta}\Big|_0^\theta = \theta^4/5$.

Thus $V(X^2)$ exists.

Then $\tilde{\theta} \to \sqrt{3\theta^2/3} = \theta$?

or $\tilde\theta \xrightarrow{P} \sqrt{3\theta^2/3} = \theta$?

Ex. 9.12. Assuming X_1, ..., $X_n \overset{i.i.d.}{\sim} Beta(\alpha,\beta)$, find an MME of (α,β).

Sol. 447. [25.] $X \sim beta(\alpha,\beta)$. $f(x) = \underline{\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}}$, if $x \underline{\in (0,1)}$, $\mu = \underline{\frac{\alpha}{\alpha+\beta}}$, where $B(\alpha,\beta) = \underline{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}}$

$$\begin{aligned}
Thus\ \ E(X) =& \frac{\alpha}{\alpha+\beta} \\
E(X^2) =& \int_0^1 x^2 \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} dx \\
=& \int_0^1 \frac{x^{\alpha+2-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} dx \\
=& \frac{B(\alpha+2,\beta)}{B(\alpha,\beta)} \quad ?? \qquad\qquad Why \\
=& \frac{\Gamma(\alpha+2)\Gamma(\beta)}{\Gamma(\alpha+\beta+2)} \Big/ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \\
=& \frac{(\alpha+1)\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)} \Big/ \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \\
=& \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}
\end{aligned}$$

Sketch hereafter: $Set \begin{cases} \overline{X^2} = \frac{\alpha}{\alpha+\beta}\frac{\alpha+1}{\alpha+\beta+1} \\ \overline{X} = \frac{\alpha}{\alpha+\beta} \end{cases}$ **What to do next ?**

$$=> \begin{cases} \overline{X} = \frac{\alpha}{\alpha+\beta} \\ \overline{X^2}/\overline{X} = \frac{\alpha+1}{\alpha+\beta+1} \end{cases}$$

$$=> \begin{cases} \overline{X}(\alpha+\beta) - \alpha = 0 \\ \overline{X^2}(\alpha+\beta+1) = \overline{X}(\alpha+1) \end{cases}$$

$$=> \begin{cases} (\overline{X}-1)\alpha + \beta\overline{X} = 0 & => \ \beta = \alpha(1-\overline{X})/\overline{X} \\ (\overline{X^2}-\overline{X})\alpha + \overline{X^2}\beta = \overline{X} - \overline{X^2} & => \ (\overline{X^2}-\overline{X})\alpha + \overline{X^2}\alpha(1-\overline{X})/\overline{X} = \overline{X} - \overline{X^2} \\ A\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \begin{pmatrix}0 \\ \overline{X} - \overline{X^2}\end{pmatrix} & \text{A=??} \end{cases} \qquad (2)$$

One way: $A\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \begin{pmatrix}0 \\ \overline{X} - \overline{X^2}\end{pmatrix} => \begin{pmatrix}a\\ \beta\end{pmatrix} = (A'A)^{-1}A'\begin{pmatrix}0 \\ \overline{X} - \overline{X^2}\end{pmatrix}$

2nd way from Eq. (2): $\alpha[(\overline{X^2}-\overline{X}) + \frac{(\overline{X^2}-\overline{X^2}\overline{X})}{\overline{X}}] = \overline{X} - \overline{X^2}$

$\hat\alpha = \frac{\overline{X}-\overline{X^2}}{\overline{X^2}-(\overline{X})^2}\overline{X}$

$\hat\beta = \frac{(\overline{X}-\overline{X^2})(1-\overline{X})}{\overline{X^2}-(\overline{X})^2}$

Remark. The MME of (α,β) is consistent, based on Theorems 9.1 and 9.2.

Th. 9.1. An unbiased estimator $\hat{\theta}_n$ for θ is consistent if $\lim_{n\to\infty} V(\hat{\theta}_n) = 0$.

Theorem 9.2. If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two estimators. $\hat{\theta}_i \xrightarrow{P} \theta_i$, $i = 1, 2$, then

$\hat{\theta}_1 + \hat{\theta}_2 \xrightarrow{P} \theta_1 + \theta_2$

$\hat{\theta}_1 \hat{\theta}_2 \xrightarrow{P} \theta_1 \theta_2$

$\hat{\theta}_1 / \hat{\theta}_2 \xrightarrow{P} \theta_1 / \theta_2$ if $\theta_2 \neq 0$;

$g(\hat{\theta}_1) \xrightarrow{P} g(\theta_1)$ if g is continuous at θ_1.

Is MME always unbiased ?

An unbiased estimator of σ^2 is $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$

The MME of σ^2 is $\overline{X^2} - (\overline{X})^2$ $(= \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2)$. Is it unbiased ?

Is S^2 the MVUE of σ^2 ?

Yes, if under $N(\mu, \sigma^2)$.

No, if under $Exp(\theta)$, then the MVUE of σ^2 is $\frac{n}{n+1}(\overline{X})^2$.

Class exercise (count half of the quiz today).

Q: Derive an MME of θ based on $\overline{X^{0.5}}$ if X_1, ..., X_n are i.i.d from $U(0, \theta)$.

Quiz on Friday: 447: [9]–[25], 448: [1]-[13]

Ex. 9.13. Assuming $X_1, ..., X_n \overset{i.i.d.}{\sim} Gamma(\alpha, \beta)$, find an MME of (α, β).

Sol. Note that $E(X) = \alpha\beta$ and $V(X) = \alpha\beta^2$.

Sketch: $\overline{X} = \alpha\beta$ and $\overline{X^2} = \alpha\beta^2 + (\alpha\beta)^2$

=> $\overline{X^2} = \overline{X}\beta + (\overline{X})^2$.

=> $\hat{\beta} = \frac{\overline{X^2} - (\overline{X})^2}{\overline{X}}$

and $\hat{\alpha} = \frac{(\overline{X})^2}{\overline{X^2} - (\overline{X})^2}$.

§9.7. The Method of Maximum Likelihood.

[13] Given a random sample X_1, ..., X_n from $f(x; \theta)$, their likelihood is $L(\theta) =$ __________, the MLE $\hat{\theta}$ of θ maximizes _________. If $g(\theta)$ is a _________ function of θ, the MLE of $g(\theta)$ is _________.

key: $\underline{\prod_i f(X_i; \theta)}$, $\underline{L(\theta)}$. $\underline{1-1}$, $\underline{g(\hat{\theta})}$,

Ex. 9.14. Given a random sample X_1, ..., X_n from $bin(1, p)$, find the MLE of p.

Sol. Two usual steps.

(1) solve $\frac{\partial \ln L}{\partial p} = 0$ to get $\hat{p}$;

(2) either check (2a) $\frac{\partial^2 \ln L}{\partial p^2} < 0$? or check

(2b) $\ln L$ at the boundary points: 0 and 1: whether $\ln L(a) < \ln L(\hat{p})$ and $\ln L(b) < \ln L(\hat{p})$.

$$\begin{aligned} L(\theta) =& \prod_{i=1}^{n} f(X_i; p) \\ =& \prod_{i=1}^{n} (p^{X_i}(1-p)^{1-X_i}) \\ =& p^{\sum_{i=1}^{n} X_i}(1-p)^{n-\sum_{i=1}^{n} X_i} \\ =& p^Y(1-p)^{n-Y} \qquad \text{where } Y = \textstyle\sum_{i=1}^{n} X_i \sim bin(?, ?) \\ \ln L =& Y \ln p + (n-Y)\ln(1-p) \\ (\ln L)'_p =& Y/p - (n-Y)/(1-p) = 0 \\ => \quad & Y(1-p) - (n-Y)p = 0 \\ => \quad & Y(1-p) - np + Yp = 0 \\ => \quad & Y = np \quad => \quad \begin{cases} p = Y/n = \overline{X} \ ? \\ \hat{p} = Y/n = \overline{X} \ ? \end{cases} \quad \text{which is correct ?} \end{aligned}$$

$(\ln L)''_p = -Y/p^2 - (n-Y)/(1-p)^2 < 0$. Thus $\hat{p} = \overline{X}$ is the MLE of p.

$\ln L(a) < \ln L(\hat{p})$?

$\ln L(b) < \ln L(\hat{p})$. ?

Ex. 9.15. Given a random sample Y_1, ..., Y_n from $N(\mu, \sigma^2)$, find the MLE of (μ, σ^2).

Sol.

$$\begin{aligned} L =& \prod_{i=1}^{n} f(Y_i; \mu, \sigma^2) \\ =& \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{Y_1-\mu}{2\sigma^2}} \times \cdots \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{Y_n-\mu}{2\sigma^2}} \\ =& (\frac{1}{\sqrt{2\pi\sigma^2}})^n e^{-\sum_{i=1}^{n} \frac{Y_i-\mu}{2\sigma^2}} \\ \ln L =& -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-\mu)^2 \\ \frac{\partial \ln L}{\partial \mu} =& \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i-\mu) = 0 \qquad => \qquad \hat{\mu} = \overline{Y} \\ \frac{\partial \ln L}{\partial \sigma^2} =& -\frac{n}{2}/\sigma^2 + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(Y_i-\mu)^2 = 0 \\ => \quad \sigma^2 =& \frac{1}{n}\sum_{i=1}^{n}(Y_i-\mu)^2 \\ => \quad \sigma^2 =& \frac{1}{n}\sum_{i=1}^{n}(Y_i-\overline{Y})^2 \end{aligned}$$

$$=\frac{1}{n}\sum_{i=1}^{n}(Y_i^2 - 2\overline{Y}Y_i + (\overline{Y})^2)$$

$$\overline{Y^2} - (\overline{Y})^2 \qquad = \text{Can it be simplified ?}$$

Need to check whether $(\mu, \sigma^2) = (\overline{Y}, \hat{\sigma}^2)$ is indeed the MLE:

(1) $\frac{\partial^2 \ln L}{\partial \mu^2}$, $\frac{\partial^2 \ln L}{\partial (\sigma^2)^2} \ln L$, ... or

(2) $\ln L$ at $\mu = \pm\infty$ and $\sigma^2 = 0$ and ∞.

It is more convenienct to check (2) here:

$$\ln L = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \mu)^2$$

$$= \begin{cases} -\infty & \text{if } \sigma^2 = \infty \\ -\frac{n}{2}\ln 0 - \frac{n}{2}\ln(2\pi) - \frac{1}{0+} = \infty - \infty \ \textbf{??} & \text{if } \sigma^2 = 0+ \\ -\infty & \text{if } \mu = \pm\infty. \end{cases} \quad (\ln 0+, \frac{1}{0+}) = \lim_{x\downarrow 0}(\ln x, \frac{1}{x}) = (-\infty, \infty)$$

Since $\lim_{x\downarrow 0} \frac{\ln x}{x^{-1}} = \lim_{x\downarrow 0} \frac{(\ln x)'}{(x^{-1})'} = \lim_{x\downarrow 0} \frac{(1/x)}{-(x^{-2})} = 0$,

$$-\ln(0+) - \frac{1}{0+} = -\infty.$$

Thus the MLE of (μ, σ^2) is $(\overline{Y}, \hat{\sigma^2})$, where $\hat{\sigma^2} = \overline{Y^2} - (\overline{Y})^2$.

Ex. 9.16. Given a random sample Y_1, ..., Y_n from $U(0, \theta)$, find the MLE of θ.

Sol.

$$L = \prod_{i=1}^{n} f(Y_i; \theta)$$

$$= \begin{cases} \frac{1}{\theta} \times \cdots \times \frac{1}{\theta} & \text{if } 0 \le Y_i \le \theta,\ i = 1, ..., n \\ 0 & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{\theta^n} & \text{if } 0 \le Y_{(1)} \le Y_{(n)} \le \theta \\ 0 & \text{otherwise} \end{cases}$$

$$\le \frac{1}{(Y_{(n)})^n}.$$

Thus the MLE of θ is $\hat{\theta} = Y_{(n)}$.

Remark. The usual approach of taking $\frac{d\ln L}{d\theta} = 0$ solve for the MLE does not work, as $\frac{d}{d\theta}\ln L = \frac{d}{d\theta}(-n\ln\theta) = -\frac{n}{\theta} \ne 0$

Invariance principle of the MLE: If g is a 1-1 function of θ and $\hat{\theta}$ is the MLE of θ then the MLE of $g(\theta)$ is $g(\hat{\theta})$.

Example 9.16 (continued) Find the MLE of $V(Y)$.

Sol. $V(Y) = \sigma_Y^2 = \theta^2/12$. Thus $\hat{\sigma}_Y^2 = Y_{(n)}^2/12$.

Example 9.15 (continued) Find the MLE of σ.

Sol. $V(Y) = \hat{\sigma}_Y^2 = \overline{Y^2} - (\overline{Y})^2$. Then the MLE of σ is $\hat{\sigma} = \sqrt{\overline{Y^2} - (\overline{Y})^2}$.

§9.8. Some large sample properties of the MLE.

Recall the CLT. $P(\overline{X} \leq t) \approx \Phi(\frac{t-\mu_{\overline{X}}}{\sigma_{\overline{X}}}) = \Phi(\frac{t-\mu_X}{\sigma_X/\sqrt{n}})$

Assuming $X_1, ..., X_n \overset{i.i.d.}{\sim} f(x;\theta)$, $\theta \in \mathcal{R}$, if $g'(\theta)$ is continuous, and $\hat{\theta}$ is the MLE of θ, then

$$P(g(\hat{\theta}) \leq t) \approx \Phi(\frac{t - g(\theta)}{\hat{\sigma}_{g(\hat{\theta})}}), \text{ where } \hat{\sigma}^2_{g(\hat{\theta})} = \left.\frac{(\frac{\partial}{\partial\theta}g(\theta))^2}{E(-\frac{\partial^2}{\partial\theta^2}\ln L(\theta))}\right|_{\theta=\hat{\theta}}.$$

An approximate CI for $g(\theta)$ is $g(\hat{\theta}) \pm z_{\alpha/2}\hat{\sigma}_{g(\hat{\theta})}$.

Ex. Let $X_1, ..., X_n$ be $\overset{i.i.d.}{\sim} bin(1,p)$. Construct an approximate $100(1-\alpha)\%$ CI for σ_X^2.
Sol. The MLE of p is $\hat{p} = \overline{X}$ (as derived before).
the MLE of σ^2 $(= g(p) = p(1-p))$ is $g(\hat{p}) = \overline{X}(1-\overline{X})$ by the invariance principle of the MLE.
To solve $\hat{\sigma^2}_{g(\hat{p})}$, need to find

$\frac{\partial}{\partial\theta}g(\theta)$? $L(\theta)$? $\frac{\partial^2}{\partial\theta^2}\ln L(\theta))$? $E(-\frac{\partial^2}{\partial\theta^2}\ln L(\theta))$? $\theta =$?

$$\begin{aligned}
L =& \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i} \\
=& p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i} \\
=& p^{n\hat{p}}(1-p)^{n(1-\hat{p})} \\
\ln L =& n\hat{p}\ln p + n(1-\hat{p})\ln(1-p) \\
\frac{\partial}{\partial p}\ln L =& \frac{n\hat{p}}{p} - \frac{n - n\hat{p}}{1-p} \\
\frac{\partial^2}{\partial p^2}\ln L =& -\frac{n\hat{p}}{p^2} - \frac{n-n\hat{p}}{(1-p)^2} \\
E(-\frac{\partial^2}{\partial p^2}\ln L) =& \frac{np}{p^2} + \frac{n-np}{(1-p)^2} \\
=& \frac{n}{p(1-p)} \\
\hat{\sigma}^2_{g(\hat{p})} =& \left.\frac{(\frac{\partial}{\partial\theta}g(\theta))^2}{E(-\frac{\partial^2}{\partial\theta^2}\ln L(\theta))}\right|_{\theta=\hat{\theta}} && g(p) = p(1-p) \\
=& \left.\frac{(g'(\theta))^2}{E(-\frac{\partial^2}{\partial\theta^2}\ln L(\theta))}\right|_{\theta=\hat{\theta}} && g' = 1-2p
\end{aligned}$$

$$
\begin{aligned}
&=(1-2\hat{p})^2/\frac{n}{\hat{p}(1-\hat{p})}\\
&=(1-2\hat{p})^2\hat{p}(1-\hat{p})/n
\end{aligned}
$$

The approximate 95% CI for $g(\hat{p})$ is $\hat{p}(1-\hat{p}) \pm 1.96\sqrt{(1-2\hat{p})^2\hat{p}(1-\hat{p})/n}$

Example 9.19. Assuming X_1, ..., $X_n \overset{i.i.d.}{\sim}$ Poisson $P(\lambda)$ with $f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$, $x = 0, 1, 2, ...$

Derive the MME of λ and $e^{-\lambda}$ $(= P(X = 0))$, and the MLE of λ and $e^{-\lambda}$.

Is the MLE of λ MVUE ?

Construct an approximate $100(1-\alpha)\%$ CI for the MLE of λ.

Sol. To solve the MME: Since $\mu = \lambda$, an MME of λ is $\hat{\lambda} = \overline{X}$.

Since $\sigma^2 = \lambda = E(X^2) - (E(X))^2$, another MME is $\tilde{\lambda} = \overline{X^2} - (\overline{X})^2$.

The MME of $e^{-\lambda}$ is ??

To solve the MLE:

$$
\begin{aligned}
L =&\prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{X_i}}{X_i!}\\
\text{ln}L =&\text{ln}\prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{X_i}}{X_i!}\\
=&\text{ln}\prod_{i=1}^{n} e^{-\lambda} + \text{ln}\prod_{i=1}^{n}\lambda^{X_i} - \text{ln}\prod_{i=1}^{n} X_i!\\
=&\text{ln}e^{-n\lambda} + \text{ln}\lambda^{\sum_{i=1}^{n} X_i} - \text{ln}\prod_{i=1}^{n} X_i!\\
=&-n\lambda + \sum_{i=1}^{n} X_i\text{ln}\lambda - \text{ln}\prod_{i=1}^{n} X_i!\\
\frac{d}{d\lambda}\text{ln}L =&-n + \sum_{i=1}^{n} X_i/\lambda \;\; (=0) \qquad\qquad => \hat{\lambda} = \overline{x}\\
\frac{d^2}{d\lambda^2}\text{ln}L =&-\sum_{i=1}^{n} X_i/\lambda^2 < 0,\\
&\textit{or check } L(0) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{X_i}}{X_i!}\Big|_{\lambda=0} =? \;\; \textit{and} \;\; \textit{check } L(\infty) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{X_i}}{X_i!}\Big|_{\lambda=\infty} =?
\end{aligned}
$$

Thus the MLE of λ is $\hat{\lambda} = \overline{X}$.

The MLE of e^{λ} is $e^{\overline{X}}$ by the invariance principle.

Q: Is $\hat{\lambda} = \overline{X}$ the MVUE of λ ?

$E(\hat{\lambda}) = E(\overline{X}) = \mu = \lambda$. Thus it is unbiased.

Need to show that $\hat{\lambda}$ is based on the sufficient statistic.

$$\begin{aligned}
L =& \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{X_i}}{X_i!} \\
=& \underbrace{\prod_{i=1}^{n} e^{-\lambda}}_{h(\lambda)} \underbrace{\prod_{i=1}^{n} \lambda^{X_i}}_{g(\vec{X},\lambda)} \underbrace{\frac{1}{\prod_{i=1}^{n} X_i!}}_{T(\vec{X})} \\
=& \underbrace{\prod_{i=1}^{n} e^{-\lambda}}_{h(\lambda)} \underbrace{\lambda^{\sum_{i=1}^{n} X_i}}_{g(\vec{X},\lambda)} \underbrace{\frac{1}{\prod_{i=1}^{n} X_i!}}_{T(\vec{X})} \\
=& \underbrace{\prod_{i=1}^{n} e^{-\lambda}}_{h(\lambda)} \underbrace{\lambda^{n\overline{X}}}_{g(\vec{X},\lambda)} \underbrace{\frac{1}{\prod_{i=1}^{n} X_i!}}_{T(\vec{X})}
\end{aligned}$$

Thus $\overline{X}$ is sufficient for λ and is unbiased. Thus it is the MVUE of λ.

Is $\hat{\lambda}$ consistent ?

Is $e^{-\overline{X}}$ MVUE ?

447 [44]: If X_1___X_2. **key:** $\perp$,

X_i's $\sim$:	$X_1 + X_2 \sim$:	
$\mathcal{G}(\alpha_i, \beta)$	______	$\mathcal{G}(\alpha_1 + \alpha_2, \beta)$
$\chi^2(v_i)$	______	$\chi^2(v_1 + v_2)$
$Pois(\lambda_i)$	______	$Pois(\lambda_1 + \lambda_2)$
$N(\mu_i, \sigma_i^2)$	______	$N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
$bin(n_i, p)$	______	$bin(n_1 + n_2, p)$

Thus $\sum_{i=1}^{n} X_i \sim Poisson(n\lambda)$ is a sufficient statistic for λ.

$E(e^{-\hat{\lambda}}) = e^{-\lambda}$?

It suffices to check $E(e^{-X}) = e^{-\lambda}$ first

$$\begin{aligned}
E(e^{-X}) =& \sum_{i=0}^{\infty} e^{-i} e^{-\lambda} \lambda^i / i! \\
=& \sum_{i=0}^{\infty} e^{-\lambda} (\lambda/e)^i / i! \\
=& \frac{e^{-\lambda}}{e^{-\lambda/e}} \sum_{i=0}^{\infty} e^{-\lambda/e} (\lambda/e)^i / i! \\
=& \frac{e^{-\lambda}}{e^{-\lambda/e}} \\
=& e^{-\lambda(1-1/e)} \\
E(e^{-\hat{\lambda}}) =& E(e^{-\sum_{i=1}^{n} X_i/n}) \\
=& \sum_{i=0}^{\infty} e^{-i/n} e^{-n\lambda} (n\lambda)^i / i!
\end{aligned}$$

$$
\begin{aligned}
=&\sum_{i=0}^{\infty} e^{-n\lambda}(e^{-1/n}n\lambda)^i/i! \\
=&\frac{e^{-n\lambda}}{e^{-e^{1/n}n\lambda}}\sum_{i=0}^{\infty} e^{-e^{-1/n}n\lambda}(e^{-1/n}n\lambda)^i/i! \qquad = \frac{e^{-n\lambda}}{e^{-e^{1/n}n\lambda}}\sum_{i=0}^{\infty} e^{-\lambda^*}(\lambda^*)^i/i! \\
=&e^{-(1-e^{-1/n})n\lambda} = e^{\lambda} \quad ?? \\
e^x =& \sum_{i=0}^{\infty} x^i/i! = 1 + x + x^2/2! + ...
\end{aligned}
$$

Is $e^{-\overline{X}}$ consistent ?

A CI for λ is

$\overline{X} \pm z_{\alpha/2}\hat{\sigma}_{\overline{X}}$,

$\overline{X} \pm z_{\alpha/2}\hat{\sigma}_X/\sqrt{n}$.

Notice that $\sigma_X^2 = \lambda$

$\overline{X} \pm z_{\alpha/2}\sqrt{\overline{X}}/\sqrt{n}$.

Review Problem 1. Suppose that X_1, ..., X_n are i.i.d. from $G(\alpha, \beta)$, the gamma distribution,

(a) Find the MME of (α, β).

(b) Is it consistent ?

Sol. Since $E(X) = \alpha\beta$ and $V(X) = \alpha\beta^2 = E(X^2) - (E(X))^2$,

Setting $\overline{X} = \hat{\alpha}\hat{\beta}$ and $\overline{X^2} - (\overline{X})^2 = \hat{\alpha}(\hat{\beta})^2$ yields

$\hat{\beta} = \frac{\overline{X^2}-(\overline{X})^2}{\overline{X}}$,

$\hat{\alpha} = \overline{X}/\hat{\beta} = \frac{(\overline{X})^2}{\overline{X^2}-(\overline{X})^2}$

Review Problem 2. Suppose that X_1, ..., X_n are i.i.d. from Poisson distribution with mean λ.

Find the MLE of $P(X = 1) = e^{-\lambda}\lambda$.

$\hat{P}(X = 1) = e^{-\hat{\lambda}}\hat{\lambda}$.

$E(\hat{P}(X = 1) = P(X = 1)$?

$$
\begin{aligned}
E(e^{-\hat{\lambda}}\hat{\lambda}) =&E(e^{-\sum_{i=1}^n X_i/n}\sum_{i=1}^{n} X_i/n) \\
=&\sum_{i=0}^{\infty} e^{-i/n}e^{-n\lambda}\frac{i}{n}(n\lambda)^i/i! \\
=&\sum_{i=0}^{\infty} \frac{i}{n}e^{-n\lambda}(e^{-1/n}n\lambda)^i/i!
\end{aligned}
$$

$$=\frac{e^{-n\lambda}}{e^{-e^{1/n}n\lambda}}\sum_{i=0}^{\infty}\frac{i}{n}e^{-e^{-1/n}n\lambda}(e^{-1/n}n\lambda)^i/i! \qquad = \frac{e^{-n\lambda}}{ne^{-e^{1/n}n\lambda}}\sum_{i=0}^{\infty} ie^{-\lambda^*}(\lambda^*)^i/i!$$

$$=e^{-(1-e^{-1/n})n\lambda}\frac{1}{n}e^{1/n}n\lambda \qquad E(X^*)$$

$$=e^{-(1-e^{-1/n})n\lambda}e^{1/n}\lambda$$

$$e^x = \sum_{i=0}^{\infty} x^i/i! = 1 + x + x^2/2! + ...$$

Ans: $E(\hat{P}(X=1) \neq P(X=1)$. Thus the MLE is not unbiased.

Class exercise. Suppose that X_1, ..., X_n are i.i.d. from Poisson distribution with mean λ. Find the MLE of $P(X \leq 3)$ and check whether it is unbiased.

Quiz on Friday: 447 9-42, 448: 1-17.

Chapter 10. Hypothesis Testing

§10.1. Introduction.

3 typical statistical inferences:

(1) estimation: $\theta = ?$

(2) Confidence interval: $I = [a, b] =?$ such that it is likely that $\theta \in I$.

(3) Hypothesis testing: $\theta = \theta_o$?

§10.2. Elements of a statistical test.

448 [15] The 5 elements of a test are (1) ________, (2) ________, (3) test statistic (4) ________, (5) ________, **key:** <u>H_o</u>, <u>H_a</u>, <u>*RR*</u>, <u>*Conclusion*</u>,

Remark. It is often to write H_1 instead of H_a.

Def. 10.1. Probability of type I error is the probability rejecting correct H_o, denoted by $\alpha = P(H_a|H_o)$. α is call the level of the test. Probability of type II error is the probability not rejecting incorrect H_o, denoted by $\beta = P(H_o|H_a)$,

448 [16] Probability of type I error is ________, Probability of type II error is ________, **key:** <u>$P(H_a|H_o)$</u>, <u>$P(H_o|H_a)$</u>.

Ex. 10.1. John claims that he will gain 50% of more of the voters in a city election. A random sample of $n = 15$ was taken, resulting Y people favor Jone. Describe the 5 elements of a test.

Sol. Let p =proportion of voters who likes John.

H_o: $p = 0.5$ (or $p \geq 0.5$) which one ?

H_a: $p < 0.5$.

Test statistic: Y=# of people who like John in a random sample of size n.

RR: $Y \leq y_o$, where y_o needs to be computed.

Conclusion: reject H_o or not. Choose one ! And write down what it means.

Remark. What is the intepretation of H_o: $p = 0.5$ v.s. H_1: $p < 0.5$?

It is to find out whether voters dislike John, as $p \geq 0.5$ is a question mark.

We need to learn how to determine y_o.

Q: If $y_o = 10$ people favor John, do you believe H_o ?

If $y_o = 0$ person favors John, do you believe H_o ?

If $y_o = 7 < n/2$ people favor John, do you believe H_o ?

If $y_o = 6 < n/2$ people favor John, do you believe H_o ?

Remark. The statistical issue is how to select y_o.

Ex. 10.1b. Under the set-up in Ex. 10.1,

(1) if we select $y_o = 15$, $P(H_a|H_o)$=? and $P(H_o|H_a)$=?

(2) if we select $y_o = 0$, $P(H_a|H_o)$=? and $P(H_o|H_a)$=?

(3) if we select $y_o = 2$, $P(H_a|H_o)$=? and $P(H_o|H_a)$=?

Sol. (1) $\alpha = 1$ and $\beta = 0$. **Why ?**

447 [2]. Axioms of probability: (1) $P(A) \geq \underline{0}$, (2) $P(S) = \underline{1}$. S=?

447 [7]. $P(\underline{\overline{A}})$ =$1 - P(A)$.

$\alpha = P(H_a|H_o) = P(Y \leq 15|p = 0.5)) = P(S) = 1.$

$\beta = P(H_o|H_1) = P(Y > 15|p < 0.5) = P(\emptyset) = 0.$

(2) $\alpha \approx 0$ and $\beta \approx 1 - 0 = 1$?

$\alpha \approx 0$ and $\beta = 1 - (1-p)^{15} \begin{cases} \approx 1 & \text{if } p \neq 0 \\ 0 & \text{if } p = 0 \end{cases}$. **Why ?**

$$\begin{aligned}
\alpha &= P(H_a|H_o) \\
&= P(Y \leq 0|p = 0.5)) \\
&= P(Y = 0) \\
&= \binom{n}{0} p^0 (1-p)^{n-0} \\
&= 0.5^{15} \\
&\approx 0.0003 \approx 0.
\end{aligned}$$

$$\begin{aligned}
\beta &= P(H_o|H_1) \\
&= P(Y > 0|p < 0.5) \\
&= 1 - P(Y = 0|p < 0.5) = ? \quad P(Y = 0|p < 0.5) = p^0(1-p)^{15} = 0??
\end{aligned}$$

$$\textbf{Ans}: \beta \begin{cases} \approx 1 - 0 = 1 & \text{if } p \in (0, 0.5) \\ = 1 - 1 = 0 & \text{if p=0} \end{cases}$$

(3) $P(H_a|H_o) = P(Y \leq 2|H_o) = [\sum_{i=0}^{2} \binom{n}{i} p^i (1-p)^{15-i}]\Big|_{p=0.5} = 0.004.$ $\quad \binom{n}{i} = \frac{n!}{i!(n-i)!}$

$$[\sum_{i=0}^{2} \binom{n}{i} p^i (1-p)^{15-i}]\Big|_{p=0.5}$$

$$=[\binom{n}{0} p^0 (1-p)^{15-0} + \binom{n}{1} p^1 (1-p)^{15-1} + \binom{n}{2} p^2 (1-p)^{15-2}]\Big|_{p=0.5}$$

$$
\begin{aligned}
&= 0.004\\
P(H_o|H_a) &= 1 - 0.004 = 0.996 \quad ???\\
P(H_o|H_a) &= P(Y > 2|H_a)\\
&= \sum_{i=3}^{15}\binom{n}{i}p^i(1-p)^{15-i}] \text{ is a function of } p < 0.5, \text{ which can be anything} < 0.996, \ e.g.
\end{aligned}
$$

```
> 1-round(pbinom(3,15,0.1),2)
    [1] 0.06
> 1-round(pbinom(3,15,0.2),2)
    [1] 0.35
> 1-round(pbinom(3,15,0.3),2)
    [1] 0.7
> 1-round(pbinom(3,15,0.4),2)
    [1] 0.91
```

Ex. 10.1c. Under the set-up in Ex. 10.1, if one tries to select $\alpha \approx 0.05$,
what are y_o ? $P(H_a|H_o)$ and $P(H_o|H_a)$? α ?

Sol. Use R program:

```
> round(pbinom(0:14,15,0.5),2)
    [1] 0.00 0.00 0.00 0.02 0.06 0.15 0.30 0.50 0.70 0.85 0.94 0.98 1.00 1.00 1.00
```

Ans: Select $y_o = 3$, that is, reject H_o if 3 or less out of 15 do not favor John.

$P(H_a|H_o) = P(Y \le y_o|H_o) = [\sum_{i=0}^{y_o}\binom{15}{i}p^i(1-p)^{15-i}]\Big|_{p=0.5} = 0.02 \le 0.05.$

$P(H_o|H_a) = 2\%$?

$\alpha = $?

Ex. 10.1d. Under the set-up in Ex. 10.1, if 30% of people likes John and one sets $y_o = 5$, what are $P(H_a|H_o)$ and $P(H_o|H_a)$?

Sol. $\alpha = P(H_a|H_o) = \sum_{i=0}^{5}\binom{15}{i}p^i(1-p)^{15-i}\Big|_{p=0.5} \approx 0.15$ (see pbinom() above),

$P(H_o|H_a) = P(Y > 5|H_a) = \sum_{i=6}^{15}\binom{n}{i}p^i(1-p)^{15-i}]\Big|_{p=0.3}$ Can we use pbinom above ?

$= 1 - \sum_{i=0}^{5}\binom{n}{i}p^i(1-p)^{15-i}]\Big|_{p=0.3} = 0.278$ **Which you prefer in exams ?**

Ex. 10.2. Under the set-up in Ex. 10.1, if 20% of people likes John and one still sets $y_o = 2$, what are $P(H_a|H_o)$ and $P(H_o|H_a)$?

Sol. $\alpha = P(H_a|H_o) = \sum_{i=0}^{2}\binom{15}{i}p^i(1-p)^{15-i}\Big|_{p=0.5} = 0.004$

and $P(H_o|H_a) = P(Y > 2|H_a) = \sum_{i=3}^{15}\binom{n}{i}p^i(1-p)^{15-i}]\big|_{p=0.2} = 0.60$

Ex. 10.3. Under the set-up in Ex. 10.1, if 10% of people likes John and one still sets $y_o = 2$, what are $P(H_a|H_o)$ and $P(H_o|H_a)$?

Sol. $\alpha = P(H_a|H_o) = \sum_{i=0}^{2}\binom{15}{i}p^i(1-p)^{15-i}\Big|_{p=0.5} = 0.004$

and $P(H_o|H_a) = P(Y > 2|H_a) = \sum_{i=3}^{15}\binom{n}{i}p^i(1-p)^{15-i}]\big|_{p=0.1} = 0.18$

```
> round(pbinom(2,15,0.5),4)
    [1] 0.004
> 1-round(pbinom(2,15,0.1),2)
    [1] 0.18
> 1-round(pbinom(2,15,0.2),2)
    [1] 0.6
> 1-round(pbinom(2,15,0.3),2)
    [1] 0.87
> 1-round(pbinom(2,15,0.4),2)
    [1] 0.97
> round(pbinom(3,15,0.5),2)
    [1] 0.02
```

Remark. $\begin{pmatrix} RR & Y\le 0 & Y\le 2 & Y\le 3 & Y\le 15 \\ \alpha & 0 & 0.004 & 0.02 & 1 \\ \beta & \begin{cases}\approx 1 & \text{if } p>0 \\ 0 & \text{if p=0}\end{cases} & (0,0.996) & 1-\sum_{i=0}^{3}\binom{15}{i}p^iq^{15-i} & 0 \end{pmatrix}$

$\alpha \uparrow <=> \beta \downarrow$ but $\alpha \ne 1-\beta$ in general.

Quiz on Friday: 447 9-42, 448: 1-17.

§10.3. Common large sample tests.

A large sample test for testing θ based on observation $\mathbf{X}$ is as follows.

Case :	(1)	(2)	(3)
H_o :	$\theta = \theta_o$		
H_a :	$\theta > \theta_o$	$\theta < \theta_o$	$\theta \ne \theta_o$
test statistic	$\hat\theta$		
Reject region	$\{\mathbf{X}: \hat\theta > \theta_o + z_\alpha\hat\sigma_{\hat\theta}\}$	$\{\mathbf{X}: \hat\theta < \theta_o - z_\alpha\hat\sigma_{\hat\theta}\}$	$\{\mathbf{X}: \lvert\hat\theta-\theta_o\rvert > z_{\alpha/2}\hat\sigma_{\hat\theta}\}$
Conclusion :			

Reason: Under certain assumptions,

$$P(H_1|H_o) = \begin{cases} P(\hat\theta > \theta_o + z_\alpha\hat\sigma_{\hat\theta}) \approx P(\frac{\hat\theta-\theta}{\sigma_{\hat\theta}} > z_\alpha) & \text{in case (1)} \\ P(\hat\theta < \theta_o - z_\alpha\hat\sigma_{\hat\theta}) \approx P(\frac{\hat\theta-\theta}{\sigma_{\hat\theta}} < -z_\alpha) & \text{in case (2)} \\ P(|\hat\theta-\theta_o| > z_{\alpha/2}\hat\sigma_{\hat\theta}) \approx P(\frac{|\hat\theta-\theta|}{\sigma_{\hat\theta}} > z_{\alpha/2}) & \text{in case (3)} \end{cases} \approx \alpha$$

448 [17] For a large sample test for H_o: $\theta = \theta_o$, a test statistic is $Z =$ ____________, a RR is Z________ if $\theta > \theta_o$; and a RR is ______________ if $\theta \ne \theta_o$; **key:** $\underline{\frac{\hat\theta-\theta_o}{\hat\sigma_{\hat\theta}}}$, $\underline{> z_\alpha}$, $\underline{|Z| > z_{\alpha/2}}$,

Remark. Test statistic can be either $\hat\theta$ or $\frac{\hat\theta-\theta_o}{\hat\sigma_{\hat\theta}}$.

Ex. 10.5. A vice president in charge of sales for a large corporation claims that salespeople are averaging no more than 15 sales contracts per week. As a check on his claim, $n = 36$

salespersons are selected at random, resulting $\overline{X} = 17$ (> 15) and $S^2 = 9$. Does the evidence contradicts the vice president's claim with $\alpha = 0.05$? **What is your instinct answer, as** $\overline{X} = 17 > 15$ **?**

Sol. The 5 components of a test:

1: H_o: $\mu = 15$ v.s.

2: H_a: $\mu \neq 15$

or H_a: $\mu > 15$,

or H_a: $\mu < 15$,

Which one ?

3. Test statistic: **which of the next 3 ?**

$\hat{\theta} = \overline{X}$?

or $\hat{\mu} = \overline{X}$?

or $Z = \frac{\hat{\mu}-\mu}{\hat{\sigma}_{\hat{\mu}}}$?

4. RR : $\hat{\mu} > \mu + z_\alpha \hat{\sigma}_{\hat{\mu}}$.

$$\begin{aligned} \overline{X} &> 15 + 1.645 S/\sqrt{n} \\ &\approx 15 + 1.645 * 3/\sqrt{36} \\ &= 15.82 \\ 17 &> 15.82? \end{aligned}$$

Or $\frac{\hat{\mu}-\mu}{\hat{\sigma}_{\hat{\mu}}} > 1.645$.

$$\begin{aligned} \frac{\hat{\mu}-\mu}{\hat{\sigma}_{\hat{\mu}}} &= \frac{17-15}{\sqrt{9}/\sqrt{36}} \\ &= 4 > 1.645 \end{aligned}$$

5. Conclusion:

(1) Reject H_o.

(2) The VP's claim is not correct. **Done ?**

(3) It seems that salespeople are averaging more than 15 sales contracts per week.

Comments: (2) may be ignored, but not (3).

Ex. 10.6. A machine in a factory must be repaired if it produces more than 10% defectives a day. A random sample of 100 items from the day's production contains 15 defectives and the supervisor says that the machine must be repaired as $15\% > 10\%$. Does the sample evidence support his decision ? Use a test with level 0.01.

What is your instinct answer, as $15\% > 10\%$ **?**

Sol. 1. H_o: $p = 0.1 = p_o$ v.s.

2. H_a: $p > 0.1$ or $p \neq 0.1$ or $p < 0.1$ **which one ?**

3. Test statistic: 3 possible ways:

$$\hat{p} = \overline{X}, \qquad \overline{X} = ?$$

$$Z = \frac{\hat{p} - p_o}{\sqrt{\frac{\hat{p}_o(1-\hat{p}_o)}{n}}}$$

$$Z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$$

Which is better ?

4. RR : $\hat{p} > p_o + z_\alpha \hat{\sigma}_{\hat{p}}$

$z_{0.01} \approx 2.32$

$\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

RR: $Z = \frac{\hat{p}-p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}} > z_\alpha = 2.32$

$Z = 1.667$

5. Conclusion: do not reject H_o ??

$p = 0.1$ that day, no need to repair the machine.

Ex. 10.7. A psychological study was conducted to compare the reaction times of men and women to a stimulus. Independent random samples of 50 men nd 50 women were employed in the experiment. The results are

men	$n_1 = 50$	$\overline{Y}_1 = 3.6$	$S_1^2 = 0.18$
women	$n_2 = 50$	$\overline{Y}_2 = 3.8$	$S_2^2 = 0.14$

Do the data suggest

a difference between the true mean reaction between men and women with $\alpha = 0.05$?

Sol. 1: H_o: $\mu_1 = \mu_2$

H_o: $\mu_1 - \mu_2 = 0$

2: v.s. H_a: $\mu_1 - \mu_2 \neq 0$.

H_a: $\mu_1 - \mu_2 > 0$.

H_a: $\mu_1 - \mu_2 < 0$.

3. Test statistic: $Z = \overline{Y}_1 - \overline{Y}_2 = -0.2$? or

$$Z = \frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} = -2.5 \ ?$$

4. RR: $|Z| > 1.96$

5. Conclusion: Reject H_o, **done ?**

there is a difference in the reaction between men and women.

In the next exam, formulas are 447 9-42 and 44; 448: 1-17. This week's homework due on Monday

In particular, [44] If X_1___X_2.

X_i's $\sim$:	$X_1+X_2\sim$:		
$\mathcal{G}(\alpha_i,\beta)$	______		$\underline{\mathcal{G}(\alpha_1+\alpha_2,\beta)}$
$\chi^2(v_i)$	______		$\underline{\chi^2(v_1+v_2)}$
$Pois(\lambda_i)$	______	**key:** $\perp$,	$\underline{Pois(\lambda_1+\lambda_2)}$
$N(\mu_i,\sigma_i^2)$	______		$\underline{N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)}$
$bin(n_i,p)$	______		$\underline{bin(n_1+n_2,p)}$

§10.4. Calculating $P(H_o|H_1)$ and finding the sample size for Z tests.

Given a test, say $I(Z\in RR)$, $\alpha=P(Z\in RR|H_o$ is true$)=P(H_1|H_o)=E(I(Z\in RR))H_o)$ and $\beta=P(Z\notin RR|H_1$ is true$)=P(H_o|H_1)$.

In this section, we shall study how to compute β for a given test and how to choose the sample size n in order to achieve given α and β, if the sample size n is large. $n\geq$??

Ex. 10.8. Recall the assumption in Ex.10.5: A vice president in charge of sales for a large corporation claims that salespeople are averaging no more than 15 sales contracts per week. As a check on his claim, $n=36$ salespersons are selected at random, resulting $\overline{X}=17$ and $S^2=9$. Suppose now H_o: $\mu=15=\mu_o$ v.s. H_a: $\mu=16$ rather than $\mu>15$. $\alpha=0.05$. $\beta=$?

Sol. Now H_o: $\mu=15$, v.s. H_a: $\mu=16$.

The test statistic is $Z=\frac{\overline{X}-\mu_o}{\hat\sigma/\sqrt{n}}$ or $\overline{X}$.

RR is $Z=\frac{\overline{X}-\mu_o}{\hat\sigma/\sqrt{n}}>1.645$, or $\overline{X}>15+z_\alpha s/\sqrt{n}$.

$$\begin{aligned}\beta =&1-P(RR) \qquad\qquad for\ \mu=16\\ =&P(\overline{X}\leq 15+z_\alpha s/\sqrt{n})\\ =&P(\overline{X}\leq 15+1.645*3/\sqrt{36})\\ =&P(\overline{X}\leq 15.82)\\ \approx&\Phi(\frac{15.82-16}{s/\sqrt{n}})\\ =&\Phi(\frac{-0.18}{3/6})\\ =&\Phi(-0.36)\\ =&0.3594\end{aligned}$$

```
> pnorm(-0.36)                R-code
[1] 0.3594236
```

Or check the normal table......

z	.00	$\cdots$	0.05	0.06	0.07	$\cdots$
0.1	$\cdots$					
$\vdots$						
0.3	.3821	$\cdots$	.3632	.3594	.3557	$\cdots$

In general, H_1: $\mu>15$, then

$$\beta=P(\overline{X}\leq 15.82)=\Phi(\frac{15.82-\mu}{s/\sqrt{n}})=\Phi(\frac{15.82-\mu}{3/6})\qquad(\mu>15). \tag{1}$$

Recall Ex. 10.1b. If $Y \sim bin(15, p)$, H_o: $p = 0.5$ v.s. H_1: $p < 0.5$.

$$P(H_a|H_o) = P(Y \leq 2|H_o) = [\sum_{i=0}^{2} \binom{n}{i} p^i (1-p)^{15-i}]\Big|_{p=0.5} = 0.004.$$

$$\begin{aligned} P(H_o|H_a) =& P(Y > 2|H_a) \\ =& \sum_{i=3}^{15} \binom{n}{i} p^i (1-p)^{15-i}] \text{ is a function of } p < 0.5, \text{ different from Eq.(1)} \end{aligned}$$

```
> 1-round(pbinom(3,15,0.1),2)
    [1] 0.06
> 1-round(pbinom(3,15,0.2),2)
    [1] 0.35
> 1-round(pbinom(3,15,0.3),2)
    [1] 0.7
> 1-round(pbinom(3,15,0.4),2)
    [1] 0.91
```

For a test with given α and β, one needs to find out n before carrying out data sample and doing the test. The formula is

$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_a - \mu_o)^2}$, provided that σ^2 is given.

Reason: Write H_o: $\mu = \mu_o$ v.s. H_a: $\mu = \mu_a < \mu_o$.

$$\begin{aligned} \alpha =& P(\overline{X} \geq \mu_o + z_\alpha \sigma/\sqrt{n}|H_o) \\ \beta =& P(\overline{X} < \mu_o + z_\alpha \sigma/\sqrt{n}|H_a) \\ =& P(\overline{X} - \mu_a < \mu_o + z_\alpha \sigma/\sqrt{n} - \mu_a|H_a) \\ =& P(\frac{\overline{X} - \mu_a}{\sigma/\sqrt{n}} \leq \frac{\mu_o + z_\alpha \sigma/\sqrt{n} - \mu_a}{\sigma/\sqrt{n}}|H_a) \\ =& \Phi(\frac{\mu_o + z_\alpha \sigma/\sqrt{n} - \mu_a}{\sigma/\sqrt{n}}) \\ =& \Phi(-z_\beta) \\ & \frac{\mu_o + z_\alpha \sigma/\sqrt{n} - \mu_a}{\sigma/\sqrt{n}} = -z_\beta \\ & \frac{\mu_o - \mu_a}{\sigma/\sqrt{n}} + z_\alpha = -z_\beta \\ & \frac{\mu_o - \mu_a}{\sigma/\sqrt{n}} = -z_\alpha - z_\beta \\ & \sqrt{n} = \frac{z_\alpha + z_\beta}{\mu_a - \mu_o} \sigma \\ & n = (\frac{z_\alpha + z_\beta}{\mu_a - \mu_o} \sigma)^2 \end{aligned}$$

Ex. 10.5 (continued. If $\beta = 0.05$ when $\mu_a = 16$, v.s. $\mu_o = 15$, what is n ?
Sol. $n = [\frac{1.645+1.645)}{(16-15)}\sigma]^2 = 3.29^2 * 9 = 97.4$. Thus $n \geq 98$.

§10.5. Relation between hypothesis testing procedure and CI

Consider large sample case (*i.e.* $n \geq 20$), with $\hat{\theta}$ is an estimator of θ. Under proper assumptions,

$$P(\hat{\theta} \leq t) \approx \Phi(\frac{t-\theta_o}{\hat{\sigma}_{\hat{\theta}}})$$

H_o: $\theta = \theta_o$ v.s. H_a: $\theta \neq \theta_o$

$$\begin{aligned} \text{Accept } H_o \text{ if } &|\frac{\hat{\theta}-\theta_o}{\hat{\sigma}_{\hat{\theta}}}| \leq z_{\alpha/2} \\ <=> &|\hat{\theta}-\theta_o| \leq z_{\alpha/2}\hat{\sigma}_{\hat{\theta}} \\ <=> & -z_{\alpha/2}\hat{\sigma}_{\hat{\theta}} \leq \hat{\theta}-\theta_o \leq z_{\alpha/2}\hat{\sigma}_{\hat{\theta}} \\ <=> &\hat{\theta}-z_{\alpha/2}\hat{\sigma}_{\hat{\theta}} \leq \theta_o \leq \hat{\theta}+z_{\alpha/2}\hat{\sigma}_{\hat{\theta}} \\ <=> &\theta_o \in [\hat{\theta}-z_{\alpha/2}\hat{\sigma}_{\hat{\theta}}, \hat{\theta}+z_{\alpha/2}\hat{\sigma}_{\hat{\theta}}] \qquad 2-sided\ CI \end{aligned}$$

H_o: $\theta = \theta_o$ v.s. H_a: $\theta > \theta_o$

$$\begin{aligned} \text{Accept } H_o \text{ if } &\frac{\hat{\theta}-\theta_o}{\hat{\sigma}_{\hat{\theta}}} \leq z_{\alpha} \\ <=> &\hat{\theta}-\theta_o \leq z_{\alpha}\hat{\sigma}_{\hat{\theta}} \\ <=> &\hat{\theta}-z_{\alpha}\hat{\sigma}_{\hat{\theta}} \leq \theta_o \\ <=> &\theta_o \in [\hat{\theta}-z_{\alpha}\hat{\sigma}_{\hat{\theta}}, \infty) \qquad upper-tail-CI \end{aligned}$$

H_o: $\theta = \theta_o$ v.s. H_a: $\theta < \theta_o$

$$\begin{aligned} \text{Accept } H_o \text{ if } &\frac{\hat{\theta}-\theta_o}{\hat{\sigma}_{\hat{\theta}}} \geq z_{\alpha} \\ <=> &\hat{\theta}-\theta_o \geq z_{\alpha}\hat{\sigma}_{\hat{\theta}} \\ <=> &\hat{\theta}-z_{\alpha}\hat{\sigma}_{\hat{\theta}} \geq \theta_o \\ <=> &\theta_o \in (-\infty, \hat{\theta}-z_{\alpha}\hat{\sigma}_{\hat{\theta}}] \qquad lower-tail-CI \end{aligned}$$

Thus in some sense, the hypothesis test procedure and CI are equivalent.

Quiz on Friday: 447: 1-25, 448: 1-20.

§10.6. Another way to report the result of a statistical test: Significance levels or p-values.

Def. 10.2. If W is a test statistic, the p-value or attained significant level, is the smallest level of significant α for which the observed data indicate that H_o should be rejected.

448 [19] The P-value is $\begin{cases} P(W \underline{\quad\quad} w)|H_o \text{ is correct}) & \text{if } H_a: \theta > \theta_o \\ P(W \underline{\quad\quad} w)|H_o \text{ is correct}) & \text{if } H_a: \theta < \theta_o \\ \underline{\quad} P(W \underline{\quad\quad} |w|)|H_o \text{ is correct}) & \text{if } H_a: \theta \neq \theta_o \end{cases}$

where W is the (Z or T) test statistic and w is the observed value of W. **key:** $\underline{\geq}$, $\underline{\leq}$, $\underline{2}$, $\underline{\geq}$.

Remark. Reject H_o if $p-$values $\leq \alpha$.

Ex.10.10. Suppose that $Y \sim bin(15, p)$. H_o: $p = 0.5$ v.s. H_a: $p < 0.5$ with $\alpha = 0.05$. Suppose that $Y = 3$ is observed. (I) Do the usual test, (II) Find the p-value.

Sol. (1) H_o: $p = 0.5$ v.s.

(2) H_a: $p < 0.5$ with $\alpha = 0.05$.

(3) Test statistic $Y = 3$.

(4) RR $Y \leq 3$ as

$P(Y \leq 4) \approx 0.059$ and $P(Y \leq 3) < 0.05$

```
> round( pbinom(0:14,15,0.5),3)
 [1] 0.000 0.000 0.004 0.018 0.059 0.151 0.304 0.500 0.696 0.849 0.941 0.982
 [13] 0.996 1.000 1.000
> round( pbinom(0:14,15,0.5),2)
 [1] 0.00 0.00 0.00 0.02 0.06 0.15 0.30 0.50 0.70 0.85 0.94 0.98 1.00 1.00
```

(5) reject H_o, that is, we conclude that $p < 0.5$.

(II) p-value$=P(Y \leq 3) \approx 0.018$.

Remark. In both ways, we reject H_o, but the p-value provides more information and we are more confident that H_o should be rejected, namely, $p < 0.5$.

Remark. $P(H_o|H_1)$ is called the probability of type II error;

$P(H_1|H_o)$ is called the probability of type I error;

α is called the level of the test H_o v.s. H_1.

It is often that $\alpha = P(H_1|H_o)$, such as under $N(0,1)$.

But in this example, the level $\alpha = 0.05 > 0.018 = P(H_1|H_o)$ the probability of type I error.

Class exercise. Under the assumptions in Ex.10.10.

If $Y = 2$, what is the p-value ? Do we reject H_o ?

If $Y = 8$, what is the p-value ? Do we reject H_o ?

Ex.10.11. A psychological study was conducted to compare the reaction times of men and women to a stimulus. Independent random samples of 50 men and 50 women were employed in the experiment. The results are $\begin{matrix} \textit{men} & n_1 = 50 & \overline{Y}_1 = 3.6 & S_1^2 = 0.18 \\ \textit{women} & n_2 = 50 & \overline{Y}_2 = 3.8 & S_2^2 = 0.14 \end{matrix}$ Let $\alpha = 0.05$.
For testing H_o: $\mu_1 - \mu_2 = 0$ v.s. H_a: $\mu_1 - \mu_2 \neq 0$.

$Z = \frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} = -2.5$

RR: $|Z| > 1.96$

Since $|Z| = 2.5 > 1.96$, reject H_o.

Then p-value= ?

448 [19] The P-value is $\begin{cases} P(W \geq w)|H_o \text{ is correct}) & \text{if } H_a: \ \theta > \theta_o \\ P(W \leq w)|H_o \text{ is correct}) & \text{if } H_a: \ \theta < \theta_o \\ \underline{2}P(W \geq |w|)|H_o \text{ is correct}) & \text{if } H_a: \ \theta \neq \theta_o \end{cases}$

where W is the (Z or T) test statistic and w is the observed value of W.

Which of the 3 is right choice here ?

Sol. $p - value = P(|Z| > | - 2.5|)$

$> \ 1 - pnorm(2.5)$ $\qquad = 2P(Z > 2.5)$

[1]0.006209665 $\qquad = 2 \times 0.0062 = 0.0124$

z	.00	.01	.02
$\vdots$	$\vdots$		
2.5	.0062	.0060	

Thus we reject H_o too. However, we are more confident to reject H_o and believe $\mu_1 \neq \mu_2$.

Ex. 3. Suppose that a Z-test for H_o: $\mu = 1$ v.s. H_a: $\mu < 1$ yields $Z = -1.5$. p-value=?

448 [19] The P-value is $\begin{cases} P(W \geq w)|H_o \text{ is correct}) & \text{if } H_a: \ \theta > \theta_o \\ P(W \leq w)|H_o \text{ is correct}) & \text{if } H_a: \ \theta < \theta_o \\ \underline{2}P(W \geq |w|)|H_o \text{ is correct}) & \text{if } H_a: \ \theta \neq \theta_o \end{cases}$

where W is the (Z or T) test statistic and w is the observed value of W.

Which of the 3 is right choice here ?

Sol. p-value= $P(Z < -1.5) = 0.0668$ from the normal table.

z	.00	$\cdots$	0.05	0.06	0.07	$\cdots$
1.4	$\cdots$					
$\vdots$						
1.5	.0668	$\cdots$				$\cdots$

. Or

```
> pnorm(-1.5)
[1] 0.0668072
```

Remark. Given p-value 0.0668,

we do not reject H_o at level $\alpha = 0.05$, but reject H_o at level $\alpha = 0.1$.

This is the advantage of reporting the p-value.

That is, if we reject H_o at level 0.1, we are risking the 10% probability to make wrong decision.

If we reject H_o at level 0.05 , we are risking the 5% probability to reject correct H_o.

§10.7. Some comments on the theory of hypothesis testing.

1. We consider 3 possible ways for H_o v.s. H_a.

   For example, regarding the difference between means μ_1 and μ_2.

   (1) H_o: $\mu_1 - \mu_2 = 0$ v.s. H_a: $\mu_1 - \mu_2 \neq 0$.

   (2) H_o: $\mu_1 - \mu_2 = 0$ v.s. H_a: $\mu_1 - \mu_2 > 0$.

   (3) H_o: $\mu_1 - \mu_2 = 0$ v.s. H_a: $\mu_1 - \mu_2 < 0$.

   Since naming μ_1 and μ_2 is somewhat arbitrary, we can ignore the 3rd way above.

   How to choose between first two ways ? It depends on the practical situations.

   If $\mu_1 > \mu_2$ suggests a large financial loss for us, then it is H_a. Otherwise, H_a is $\mu_1 - \mu_2 < 0$.

2. Why set H_o: $\mu_1 - \mu_2 = 0$ v.s. H_a: $\mu_1 - \mu_2 > 0$; not

   H_o: $\mu_1 - \mu_2 \leq 0$ v.s. H_a: $\mu_1 - \mu_2 > 0$?

   The answer is that either ways works. They leads to the same RR and conclusion.

   However, the second way is more complicated to compute α, thus at this course

   we choose the simple way.

3. If the test suggests H_a is false, we report that

   "do not reject H_o", rather than saying that we accept H_o,

   as H_o may still be wrong. We just do not have evidence to say that it is wrong.

4. Is it possible to set H_o: $\mu_1 - \mu_2 = 0$ v.s. H_a: $\mu_1 - \mu_2 = 3$?

   Ans: Yes, we can, if in the practical situation, we are comparing $\mu_1 - \mu_2 = 0$ v.s. $\mu_1 - \mu_2 = 3$.

   However, in most situation, we do not have 3 in mind.

5. Given α, say 0.05 for testing H_o: $\mu = 0$ v.s. H_1: $\mu > 0$.

   with Z test statistic where $Z \sim N(0,1)$,

   Both $\phi_1 = I(Z > 1.645)$ and $\phi_2 = I(|Z| > 1.96)$ have $\alpha = 0.05$.

   But their β values are different, *i.e.*, their $P(H_o|H_1)$ are different.

   Thus how to find a good level-α test is a theoretical issue.

   It is related to the most powerful test, in the sense to have the smallest $P(H_o|H_1)$.

   Comments on the correction of the 2nd test:

Typos in the 2nd test.

5. If (1) X_1, ..., X_n are i.i.d. from $N(\mu_1, \_\_\_)$, (2) Y_1, ..., Y_m are i.i.d. from $N(\mu_2, \_\_\_)$, and (3) X_i's ___ Y_j's, then

   5.1. $100(1-\alpha)\%$ CI for μ_1 is ____________,

   5.2. $100(1-\alpha)\%$ CI for $\mu_1 - \mu_2$ is ____________ $\hat{\sigma}_p\sqrt{\frac{1}{n}+\frac{1}{m}}$, where $\hat{\sigma}_p =$ _________

   5.3. $100(1-\alpha)\%$ CI for σ_x^2 is ____________

5. If (1) X_1, ..., X_n are i.i.d. from $N(\mu_x, \_\_\_)$, (2) Y_1, ..., Y_m are i.i.d. from $N(\mu_y, \_\_\_)$, and (3) X_i's ___ Y_j's, then $T = \underline{\frac{\overline{X}-\mu_x}{S_x/\sqrt{n}}}, \sim$ ____________,

   $T = \frac{\overline{X}-\overline{Y}-(\mu_x-\mu_y)}{\hat{\sigma}_p\sqrt{1/n_x+1/n_y}} \sim$ ____________, where $\hat{\sigma} =$ ________________,

$W = (n_x - 1)S_x^2/\sigma^2 \sim$ ________________, $F = S_x^2/S_y^2 \sim$ ________________,

B.1. Two steps in finding the MLE:

(1) $\frac{d\ln L(\theta)}{d\theta} = 0$ yields $\hat\theta$

(2) Check. Either check whether $(\ln L(\theta))'' < 0$??

Or check $L(\theta)$ at the boundary and $\hat\theta$ and compare them.

2. The MLE of θ is

$$\hat\theta = \frac{4}{\overline{\overline{X}}} = \frac{4n}{\sum_{i=1}^n X_i} = \frac{4n}{T}$$

$T = \sum_i X_i$ is $G(\alpha, \beta)$. Thus f_T is known !!

$E(1/T) = ?$

$$E(X) = \mu$$

$$E(g(X)) = g(\mu) \ ???$$

$$E(1/X) = 1/\mu \ ??? \qquad (g(x) = 1/x)$$

$$E(X^2) = \mu^2 \ ??? \qquad (g(x) = x^2)$$

447. [15] $$E(g(X)) = \begin{cases} \dots & \text{if discrete} \\ \dots & \text{if cts} \end{cases}$$

$V(1/T) = ??$

Need to compute $E(\frac{1}{T^2})$ and $(E(\frac{1}{T}))^2$, ...

$$\int_0^\infty t^k \frac{t^{\alpha-1}e^{-t/\beta}}{\beta^\alpha\Gamma(\alpha)}dt = \int_0^\infty \frac{t^{(\alpha+k)-1}e^{-t/\beta}}{\beta^\alpha\Gamma(\alpha)}dt \quad and \quad \Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$

B.2. 3 H_1s: $p_1 - p_2 \neq 0$, $p_1 - p_2 > 0$, $p_1 - p_2 < 0$.

Only one is correct. Often the data suggest the H_1.

$\hat\sigma_{\hat p_1 - \hat p_2}$:

$\sqrt{\hat p_1 \hat q_1/n_1 + \hat p_2 \hat q_2/n_2}$,

$\sqrt{pq/n_1 + pq/n_2}$, where $p = \frac{34+98}{112+260}$, as $p_1 = p_2$ under H_o.

§10.8. Small sample tests for μ and $\mu_1 - \mu_2$.

For large sample test about μ or $\mu_1 - \mu_2$, we use test statistic $Z = \frac{\overline{X} - \mu_o}{\hat\sigma_{\overline{X}}}$ or

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\hat\sigma_{\overline{X}-\overline{Y}}} = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}}$$

as $\sigma_X^2 \approx S_X^2$ and $\sigma^2_{\overline{X}-\overline{Y}} \approx \frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}$

In particular, for testing H_o: $\mu = \mu_o$, v.s. H_a: $\mu > \mu_o$, RR: $Z > z_\alpha$.

Q: What is the meaning of large sample size ?

For one sample test, $n \geq 20$;

For two sample test, $n_1 \geq 20$ and $n_2 \geq 20$.

Q: What to do if $n < 20$?

Recall that the CI for μ is $\overline{X} \pm t_{\alpha/2,n-1}S/\sqrt{n}$, based on $T = \frac{\overline{X}-\mu_o}{S/\sqrt{n}} \sim t_{n-1}$ distribution.

The CI for $\mu_1 - \mu_2$ is $\overline{X} - \overline{Y} \pm t_{\alpha/2,n_1+n_2-2}\hat{\sigma}_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

where $\hat{\sigma}_p^2 = \frac{(n_1-1)S_X^2+(n_2-1)S_Y^2}{n_1+n_2-2}$ based on $T = \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\hat{\sigma}_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$,

under the assumption that $\begin{cases} 1. \ X_1, ..., X_{n_1} \overset{i.i.d.}{\sim} N(\mu_X, \sigma^2) \\ 2. \ Y_1, ..., Y_{n_2} \overset{i.i.d.}{\sim} N(\mu_Y, \sigma^2) \\ 3. \ X_i \perp Y_j \ \forall \ i, j \end{cases}$

Small sample size test for μ under the assumption that $X_1, ..., X_n \overset{i.i.d.}{\sim} N(\mu, \sigma)$:

Case :	(1)	(2)	(3)
H_o :	$\mu = \mu_o$		
H_a :	$\mu < \mu_o$	$\mu > \mu_o$	$\mu \neq \mu_o$
test statistic	$T = \frac{\overline{X}-\mu_o}{S/\sqrt{n}}$		
Reject region	$T < -t_{\alpha,n-1}$	$T > t_{\alpha,n-1}$	$\|T\| > t_{\alpha/2,n-1}$
Conclusion :			

Ex.10.12. Suppose that muzzle velocities of eight shells tested with a new gunpowder, and yields $\overline{Y} = 2959$ and $S = 39.1$. The manufactory claims that the new gunpowder produces an average velocity of no less that 3000 feet/second. (A) Do the sample data provide sufficient evidence to contradict the claim at $\alpha = 0.025$? (B) Compute the p-value too.

Sol. (A) 1. and 2.: H_o: $\theta = 3000$, v.s. H_1: $\theta < 3000$ or H_1: $\theta > 3000$?

3. Test statistic: $T = \frac{\overline{X}-\mu_o}{S/\sqrt{n}} = \frac{2959-3000}{39.1/\sqrt{8}} = -2.966$

4. RR: $T < -2.365$, as $t_{0.025,8-1} = 2.365$.

$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$\cdots$	df
3.078				$\cdots$	1
$\vdots$					
		2.365	2.998	$\cdots$	7

5 Conclusion: Since $T = -2.966 < -2.365$, reject H_o **Done ?**

There is evidence that the velocity is less than 3000 feet/second.

Or the data do provide sufficient evidence to contradict the claim that the velocity of no less that 3000 feet/second.

(B) 448 [19] The P-value is $\begin{cases} P(W \underline{>} w)|H_o \text{ is correct}) & \text{if } H_a : \ \theta > \theta_o \\ P(W \underline{<} w)|H_o \text{ is correct}) & \text{if } H_a : \ \theta < \theta_o \\ \underline{2}P(W \underline{>} |w|)|H_o \text{ is correct}) & \text{if } H_a : \ \theta \neq \theta_o \end{cases}$

where W is the (Z or T) test statistic and w is the observed value of W.

The p-value using R-code is

```
> pt(2.996,7)
> 1- pt(2.996,7)
```

> pt(-2.996,7)

Which one is correct ?

[1] 0.01002765

Ex.10.12(c). For the same data above, if an independent agent is asked to check whether the new gunpowder produces an average velocity of 3000 feet/second at $\alpha = 0.025$, what is the answer ?

Sol. 1 and 2: H_o: $\theta = 3000$, v.s. H_1: $\theta \neq 3000$.

3. Test statistic: $T = \frac{\overline{X}-\mu_o}{S/\sqrt{n}} = \frac{2959-3000}{39.1/\sqrt{8}} = -2.966$

4. RR: $|T| > t_{0.0125,7} = ???$

$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$\cdots$	df
3.078				$\cdots$	1
$\vdots$					
	1.895	2.365	2.998	$\cdots$	7

How to continue ?

Since it is a 2-sided test, use R-codes to find out P-value of 2.996:

> 2*pt(-2.996,7)

The P-value is $2 * 0.01 = 0.02$.

5. Conclusion: Since P-value $= 0.02 < \alpha = 0.025$ reject H_o.

There is some evidence that the velocity is not 3000 feet/second.

Small sample test for comparing two population means:

Case :	(1)	(2)	(3)
H_o :	$\mu_1 - \mu_2 = D_o$		
H_a :	$\mu_1 - \mu_2 < D_o$	$\mu_1 - \mu_2 > D_o$	$\mu_1 - \mu_2 \neq D_o$
test statistic	$T = \frac{\overline{X}-\overline{Y}-D_o}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$		
Reject region	$T < -t_{\alpha,n-1}$	$T > t_{\alpha,n-1}$	$\|T\| > t_{\alpha/2,n-1}$
Conclusion :			

Ex.10.14. The workers on the assembling lines were trained using two different methods. Suppose that 2 sets of independent samples are obtained from $N(\mu_X, \sigma^2)$ and $N(\mu_Y, \sigma^2)$.

32, 37, 35, 28, 41, 44, 35, 31, 34,

35, 31, 29, 25, 34, 40, 27, 32, 31.

(A) Do the sample data provide sufficient evidence to indicate that there is a difference in true mean assembly times for those trained using these two methods at $\alpha = 0.05$?

(B) Compute the p-value too.

Sol. (A) From the given conditions, we have $n_1 = 9 = n_2$,

$\overline{X} = 35.22$,

$\overline{Y} = 31.56$,

$\sum_{i=1}^{n_1}(X_i - \overline{X})^2 = 195.56$,

$\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2 = 160.22$.

$S_p = \sqrt{\frac{\sum_{i=1}^{n_1}(X_i-\overline{X})^2+\sum_{i=1}^{n_2}(Y_i-\overline{Y})^2}{n_1+n_2}}$

1. H_o: $\mu_1 = \mu_2$ v.s.
2. H_a: $\mu_1 - \mu_2 \neq 0$
3. Test statistic: $T = \frac{\overline{X}-\overline{Y}}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{35.22-31.56}{\sqrt{\frac{195.56+160.22}{9+9-2}}\sqrt{1/9+1/9}} = 1.65$
4. Reject region :$|T| > t_{\alpha/2,n-2} = 2.12$
5. Conclusion: Since $|T| = 1.65 < 2.12$, do not reject H_o, there is no evidence to suggest that there is a difference in the two assembly times for those trained using the two methods.

(B) 448 [19] The P-value is $\begin{cases} P(W \geq w)|H_o \text{ is correct}) & \text{if } H_a: \ \theta > \theta_o \\ P(W \leq w)|H_o \text{ is correct}) & \text{if } H_a: \ \theta < \theta_o \\ \underline{2}P(W \geq |w|)|H_o \text{ is correct}) & \text{if } H_a: \ \theta \neq \theta_o \end{cases}$

where W is the (Z or T) test statistic and w is the observed value of W.

The p-value is

```
> 2*pt(1.65,16)
> 2*(1-pt(1.65,16))
> 2*pt(-1.65,16)
[1] 0.1184333
```

Ex.10.14 (c). Given the data as in Ex. 10.14, do the sample data provide sufficient evidence to indicate that the true mean assembly times for those trained using the first method is longer than the other one at $\alpha = 0.1$? **Class exercise.**

1. H_o: $\mu_1 = \mu_2$ v.s.
2. H_a: $\mu_1 - \mu_2 > 0$
3. Test statistic: $T = \frac{\overline{X}-\overline{Y}}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{35.22-31.56}{\sqrt{\frac{195.56+160.22}{9+9-2}}\sqrt{1/9+1/9}} = 1.65$

$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$\cdots$	*df*
3.078				$\cdots$	1
$\vdots$					
		2.365	2.998	$\cdots$	7
$\vdots$					
1.337	1.746	2.120		$\cdots$	16

4. Reject region :$T > t_{\alpha,n-1} = 1.746$
5. Conclusion: Since $T = 1.65 < 1.746$, reject H_o, there is some evidence to suggest that the true mean assembly times for those trained using the first method is longer than the other one at $\alpha = 0.1$.

447. [44] If $X_1 \underline{\quad\quad} X_2$. **key:** $\perp$,

X_i's $\sim$:	$X_1 + X_2 \sim$:	
$\mathcal{G}(\alpha_i, \beta)$	$\underline{\qquad\qquad}$	$\underline{\mathcal{G}(\alpha_1+\alpha_2, \beta)}$
$\chi^2(v_i)$	$\underline{\qquad\qquad}$	$\underline{\chi^2(v_1+v_2)}$
$Pois(\lambda_i)$	$\underline{\qquad\qquad}$	$\underline{Pois(\lambda_1+\lambda_2)}$
$N(\mu_i, \sigma_i^2)$	$\underline{\qquad\qquad}$	$\underline{N(\mu_x+\mu_y, \sigma_1^2+\sigma_2^2)}$
$bin(n_i, p)$	$\underline{\qquad\qquad}$	$\underline{bin(n_1+n_2, p)}$

These distributions are really from 4 distributions.

$G(\alpha, \beta)$, with df $f(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} \propto x^\alpha e^{-x/\beta}$, $x > 0$,

$Pois(\lambda)$, with df $f(x) = e^{-\lambda}\lambda^x/x! \propto \lambda^x$, $x = 0, 1, 2, ...$

$N(\mu,\sigma^2)$, $f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \propto e^{-\frac{x^2-2\mu x}{2\sigma^2}}$,
and $bin(n,p)$, with df $f(x)=\binom{n}{x}p^x(1-p)^{n-x} \propto (\frac{p}{1-p})^x$, $x=0,1,...,n$.
They belong to the exponential family.

Def. A family of distributions $\{f(x|\theta):\theta\in A\}$ $(A\subset\mathcal{R}^p)$ belongs to the exponential family if

$$f(x|\theta)=h(x)c(\theta)\exp(\sum_{i=1}^{k}w_i(\theta)t_i(x)). \qquad (k,\theta,h,c,w_i,t_i)$$

$bin(n,p)$.
$f(x|p)=\binom{n}{x}p^x(1-p)^{n-x}$, $x\in\{0,1,...,n\}$

$$f(x|p)=\underbrace{\binom{n}{x}\mathbf{1}_{(x\in\{0,1,...,n\})}}_{h(x)}\underbrace{(1-p)^n}_{c(\theta)}\exp(x\ln(\frac{p}{1-p}))$$

$k=?$ $\theta=?$ $t_i(x)=?$ $w_i(\theta)=?$
$N(\mu,\sigma^2)$.

$$\begin{aligned}f(x|\mu,\sigma)&=\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(x-\mu)^2}{2\sigma^2})\\&=\exp(-\frac{x^2-2\mu x+\mu^2}{2\sigma^2})\frac{1}{\sqrt{2\pi\sigma^2}}\\&=\underbrace{\exp(-\frac{\mu^2}{2\sigma^2})\frac{1}{\sqrt{2\pi\sigma^2}}}_{c(\theta)}\cdot\underbrace{1}_{h(x)}\exp(\underbrace{-\frac{1}{2\sigma^2}}_{w_1(\theta)}\underbrace{x^2}_{t_1(x)}+\underbrace{\frac{\mu}{\sigma^2}}_{w_2(\theta)}\underbrace{x}_{t_2(x)}).\end{aligned}$$

$k=?$ $\theta=?$ $t_i(x)=?$ $w_i(\theta)=?$
$G(\alpha,\beta)$, with df

$$\begin{aligned}f(x)&=\frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}I(x>0)\\&=\underbrace{\frac{1}{\beta^\alpha\Gamma(\alpha)}}_{c(\theta)}\underbrace{I(x>0)}_{h(x)}e^{(\alpha-1)\ln x-\frac{1}{\beta}x}\\&=\underbrace{\frac{1}{\beta^\alpha\Gamma(\alpha)}}_{c(\theta)}\underbrace{I(x>0)/x}_{h(x)}e^{\alpha\ln x-\frac{1}{\beta}x}\end{aligned}$$

Which is correct ?

$k=?$ $\theta=?$ $t_i(x)=?$ $w_i(\theta)=?$
$Pois(\lambda)$, with df

$$\begin{aligned}f(x)&=e^{-\lambda}\lambda^x/x!I(x=0,1,2,...)\\&=\underbrace{e^{-\lambda}}_{c(\theta)}\underbrace{(1/x!)I(x=0,1,2,...)}_{h(x)}e^{x\ln\lambda}\end{aligned}$$

$k=?$ $\theta=?$ $t_i(x)=?$ $w_i(\theta)=?$
The above expressions present sufficient statistic, which lead to MVUE.

For Pois(λ), $t(X) = X$ leads to $\sum_i X_i$ or $\overline{X}$. **Why ??**

$$\prod_i f(x_i) = \underbrace{e^{-n\lambda}}_{c(\theta)} \underbrace{\prod_i (1/x_i!) I(x_i = 0, 1, 2, ...)}_{h(\vec{x})} e^{\sum_i x_i \ln\lambda}$$

=> $\overline{X}$ is sufficient, and $E(\overline{X}) = \lambda$.

Thus $\overline{X}$ is MVUE of λ.

For bin(n,p),

$$f(x|p) = \underbrace{\binom{n}{x} \mathbf{1}_{(x\in\{0,1,\ldots,n\})}}_{h(x)} \underbrace{(1-p)^n}_{c(\theta)} \exp(\underbrace{x}_{t(x)} \ln(\frac{p}{1-p}))$$

the sufficient statistic is $\sum_i X_i$ due to $t(x) = x$ => $\sum_i x_i$ or $\overline{x}$.

$T = \overline{X}$. $E(T) = p$, thus $T = \overline{X}$ is a MVUE of p

For $N(\mu, \sigma^2)$,

$$f(x) = \underbrace{\exp(-\frac{\mu^2}{2\sigma^2})\frac{1}{\sqrt{2\pi\sigma^2}}}_{c(\theta)} \cdot \underbrace{1}_{h(x)} \exp(\underbrace{-\frac{1}{2\sigma^2}}_{w_1(\theta)} \underbrace{x^2}_{t_1(x)} + \underbrace{\frac{\mu}{\sigma^2}}_{w_2(\theta)} \underbrace{x}_{t_2(x)})$$

(X, X^2) is a sufficient statistic, it yields $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ or $(\overline{X}, \overline{X^2})$.

Thus $(\overline{X}, S^2)$ is the MVUE of (μ, σ^2), where $S^2 = \frac{n}{n-1}(\overline{X^2} - (\overline{X})^2)$

For $G(\alpha, \beta)$,

$$f(x) = \underbrace{\frac{1}{\beta^\alpha \Gamma(\alpha)}}_{c(\theta)} \underbrace{I(x > 0)/x}_{h(x)} e^{\alpha \ln x - \frac{1}{\beta}x}$$

$(X, \ln X)$ leads to $(\sum_i X_i, \sum_i \overline{\ln X_i})$ or $(\overline{X}, \overline{\ln X})$. Thus $(\overline{X}, \overline{\ln X})$ is sufficent.

Since $E(\ln X) \propto \int_0^\infty \ln x x^{\alpha-1} e^{-x/\beta} dx$ no simple expression,

for simplicity in 448, set α as a constant such as $\alpha = 4$ in the 2nd test.

Example 1. Let $f(x|\mu, \lambda) = \frac{1}{\lambda} e^{-\frac{x-\mu}{\lambda}}$, $x > \mu$, $\lambda > 0$.

Does $\{f(\cdot|\mu, \lambda) : \mu \in (-\infty, \infty), \lambda > 0\}$ belong to the exponential family?

Sol. Yes, as $f(x|\mu, \lambda) = \underbrace{1}_{h(x)} \underbrace{\frac{1}{\lambda} e^{\mu/\lambda}}_{c(\theta)} \exp(\underbrace{-\frac{1}{\lambda}}_{w_1(\theta)} \underbrace{x}_{t_1(x)})$.

Q: Is it correct ?

Ans: No, as $f(x|\mu, \lambda) = \mathbf{1}_{(x>\mu)} \frac{1}{\lambda} e^{\mu/\lambda} e^{-\frac{1}{\lambda}x} \neq h(x)c(\theta)\exp(\sum_{j=1}^k w_j(\theta)t_j(x))$.

It suffices to show that $\log \mathbf{1}(x > \mu) \neq \sum_{i=2}^2 w_i(\theta)t_i(x)$.

If $x > \mu$, $\quad 0 = \sum_{i=2}^2 w_i(\theta)t_i(x) = w_2(\theta)t_2(x)$.

Thus $w_2(\cdot) = 0$ or $t_2(\cdot) = 0$.

If $x < \mu$, $\quad -\infty = \sum_{i=2}^2 w_i(\theta)t_i(x) = w_2(\theta)t_2(x) = 0$. A contradiction.

Example 2. Let $f(x|\theta) = \begin{cases} p_1 & \text{if } x = 1 \\ p_2 & \text{if } x = 2 \\ p_3 & \text{if } x = 3, \end{cases}$ where $\theta = (p_1, p_2)$, $p_i \geq 0$ and $p_1 + p_2 + p_3 = 1$.
Does it belong to the exponential family ?

Sol. Yes. Let $y_i = \mathbf{1}_{(x=i)}$, $i = 1, 2, 3$. Then

$f(x|\theta) = p_1^{y_1} p_2^{y_2} p_3^{y_3}$ if $x \in \{1, 2, 3\}$.

$f(x|\theta) = \mathbf{1}_{(x\in\{1,2,3\})} \exp(y_1 \ln p_1 + y_2 \ln p_2 + y_3 \ln p_3)$ **(**k **= 3??)**

Why do not set $\theta = (p_1, p_2, p_3)$ **?**

$y_3 = 1 - y_1 - y_2$.
$f = \mathbf{1}_{(x\in\{1,2,3\})} \exp(y_1 \ln(p_1/p_3) + y_2 \ln(p_2/p_3)) \exp(\ln p_3)$
$h = ?$ $c = ?$ $w_i = ?$ $t_i = ?$

Example 1. Given data of size $n = 100$, solve the following problems related the data below.

> (Y=sort(X))

[1] 2.05 2.09 2.24 2.25 2.28 2.34 2.40 2.43 2.49 2.50 2.57 2.71 2.74 2.81 2.81
[16] 2.82 2.85 2.96 2.98 2.98 3.03 3.06 3.07 3.17 3.27 3.30 3.33 3.34 3.36 3.36
[31] 3.39 3.49 3.52 3.53 3.54 3.54 3.56 3.60 3.63 3.64 3.65 3.74 3.75 3.82 3.84
[46] 3.91 3.91 3.91 3.91 3.93 3.97 3.99 4.07 4.12 4.21 4.29 4.40 4.52 4.57 4.59
[61] 4.60 4.61 4.64 4.65 4.67 4.75 4.77 4.84 4.85 4.87 4.89 4.93 5.03 5.07 5.08
[76] 5.11 5.11 5.12 5.13 5.16 5.18 5.19 5.24 5.28 5.31 5.36 5.44 5.46 5.48 5.50
[91] 5.51 5.57 5.59 5.59 5.63 5.65 5.74 5.78 5.84 5.97

> mean(X)

[1] 4.0832

Assume $X \sim U(a, 6)$. Let $P = P(X > 3)$. **Derive**

(1) the MLE of a and the MLE $\hat{P}$,

(2) the density of $X_{(1)}$,

(3) $\sigma_{\hat{P}}$ as a function of a.

(4) Compute the MLE estimate of P. and $SE_{\hat{P}}$ based on the above data.

Sol. (1) Maximizing likelihood function over $a < b = 6$ yields the MLE is $\hat{a} = X_{(1)}$. **Proof:**
$L = \prod_{i=1}^n \frac{I(a \leq X_i \leq b)}{b-a}$
$= \frac{I(a \leq X_{(1)} \leq 6)}{(6-a)^n} \leq \frac{I(a = X_{(1)} \leq 6)}{(6-X(1))^n}$. $\hat{a} = X_{(1)}$.

Range of a: 2.05 ? 2 ? $(-\infty, 3)$?). $(-\infty, 6)$?) data independent.

Difference between the maximum likelihood estimator and the maximum likelihood estimate.

Estimator = estimate ?

$X_{(1)}$ is the estimator of a, and 2.05 is the MLE estimate of a based on the given data.

Since $P = \frac{6-3\vee a}{6-a} = \begin{cases} \frac{6-3}{6-a} & \text{if } a < 3 \\ 1 & \text{if } a \in [3,6] \end{cases}$, by invariance principle of MLE, we have

$$\text{MLE } \hat{P} = \frac{6-3\vee\hat{a}}{6-\hat{a}} = \begin{cases} \frac{6-3}{6-\hat{a}} & \text{if } \hat{a} < 3 \\ 1 & \text{if } \hat{a} \in [3,6] \end{cases} \quad \text{v.s. estimate of } \hat{P} \text{ is } 0.7594. \tag{1}$$

(2) $f_{X_{(1)}}(t) = \frac{n!}{1!(n-1)!} f(t) S^{n-1}(t), \qquad t \in (X_{(1)}, 6).$
(3) $\sigma^2_{\hat{P}} = E((\hat{P})^2) - (E(\hat{P}))^2$ (see Eq.(1)).

$$\sigma^2_{\hat{P}} = \sigma^2_{0.7594} = ??$$

If $\hat{a} \in [3,6]$, then $\hat{P} = 1$, then $\sigma^2_{\hat{P}} = 0$. **right ? wrong ? DNK**

If $a \in [3,6]$, then $\hat{a} \geq 3$ and $\hat{P} = 1$, thus $\sigma^2_{\hat{P}} = 0$.

Hence, $\sigma_{\hat{P}} = \begin{cases} 0 & \text{if } a \in [3,6] \\ \sqrt{E(\hat{P}^2) - (E(\hat{P}))^2} & \text{if } a < 3 \end{cases}$, where

for $a < 3$,
$$\begin{aligned}
E(\hat{P}) =& \int_a^3 \frac{3}{6-x} \times \frac{n!}{1!(n-1)!} \frac{1}{6-a} (\frac{6-x}{6-a})^{n-1} dx \\
&+ \int_3^6 1 \times \frac{n!}{1!(n-1)!} \frac{1}{6-a} (\frac{6-x}{6-a})^{n-1} dx \\
=& \frac{3n}{(6-a)^n} \int_a^3 (6-x)^{n-2} dx + \frac{n}{(6-a)^n} \int_3^6 (6-x)^{n-1} dx \\
=& \frac{-3n}{(n-1)(6-a)^n} (6-x)^{n-1}\Big|_a^3 - \frac{1}{(6-a)^n} (6-x)^n\Big|_3^6 \\
=& \frac{3n}{(n-1)(6-a)^n} [(6-a)^{n-1} - (6-3)^{n-1}] + \frac{3^n}{(6-a)^n}
\end{aligned}$$

$$\begin{aligned}
E((\hat{P})^2) =& \int_a^3 (\frac{3}{6-x})^2 \times \frac{n!}{1!(n-1)!} \frac{1}{6-a} (\frac{6-x}{6-a})^{n-1} dx \\
&+ \int_3^6 1^2 \times \frac{n!}{1!(n-1)!} \frac{1}{6-a} (\frac{6-x}{6-a})^{n-1} dx \\
=& \frac{3^2 n}{(6-a)^n} \int_a^3 (6-x)^{n-3} dx + \frac{n}{(6-a)^n} \int_3^6 (6-x)^{n-1} dx \\
=& \frac{-3^2 n}{(n-2)(6-a)^n} (6-x)^{n-2}\Big|_a^3 - \frac{1}{(6-a)^n} (6-x)^n\Big|_3^6 \\
=& \frac{3^2 n}{(n-2)(6-a)^n} [(6-a)^{n-2} - (6-3)^{n-2}] + \frac{3^n}{(6-a)^n}
\end{aligned}$$

```
n=100
a=2.05
(A=3*n/((n-1)*(6-a)**n)*( (6-a)**(n-1)- (6-3)**(n-1))+(3/(6-a))**n)
[1] 0.7671653
```

(B=3*3*n/((n-2)*(6-a)**n)*((6-a)**(n-2)- (6-3)**(n-2))+(3/(6-a))**n)
[1] 0.5886027
(s=sqrt(B-A*A))
[1] 0.00774954 => $2\hat\sigma_{\hat P} = 0.015$ for the given data.

(4) Compute the MLE estimate of P and $SE_{\hat P}$ based on the above data.
Sol. 0.76 ± 0.015,
(6-3)/(6-a) [1] 0.7594937
$2 * 0.0077 = 0.015$

§10.9. Testing hypotheses concerning variances

So far, the tests are about means. For large sample test about $\mu_1 - \mu_2$, we use Z-test statistic

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\hat\sigma_{\overline{X}-\overline{Y}}} = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\hat\sigma_X^2}{n_1} + \frac{\hat\sigma_Y^2}{n_2}}} = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}}}$$

as $\sigma_X^2 \approx S_X^2$, $\sigma_Y^2 \approx S_Y^2$ and $\sigma^2_{\overline{X}-\overline{Y}} \approx \frac{S_X^2}{n_1} + \frac{S_Y^2}{n_2}$.

On the otherhand, based on $T = \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\hat\sigma_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$,

under the assumption that $\begin{cases} 1.\ X_1, ..., X_{n_1} \overset{i.i.d.}{\sim} N(\mu_X, \sigma^2) \\ 2.\ Y_1, ..., Y_{n_2} \overset{i.i.d.}{\sim} N(\mu_Y, \sigma^2) \\ 3.\ X_i \perp Y_j\ \forall\ i, j \end{cases}$ and

$\hat\sigma_p^2 = \frac{(n_1-1)S_X^2+(n_2-1)S_Y^2}{n_1+n_2-2}$,

the small sample t-test for comparing two population means is

Case :	(1)	(2)	(3)
H_o :	$\mu_1 - \mu_2 = D_o$		
H_a :	$\mu_1 - \mu_2 < D_o$	$\mu_1 - \mu_2 > D_o$	$\mu_1 - \mu_2 \neq D_o$
test statistic	$T = \frac{\overline{X}-\overline{Y}-D_o}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$		
Reject region	$T < -t_{\alpha,n-1}$	$T > t_{\alpha,n-1}$	$\lvert T\rvert > t_{\alpha/2,n-1}$
Conclusion :			

An important assumption for the t-test is $\sigma_X = \sigma_Y$.

Thus one may need to test whether $\sigma_X^2 = \sigma_Y^2$.

This is the first testing problem about σ^2.

For this problem, the assumption is that $\begin{cases} 1.\ X_1, ..., X_{n_1} \overset{i.i.d.}{\sim} N(\mu_1, \sigma_1^2) \\ 2.\ Y_1, ..., Y_{n_2} \overset{i.i.d.}{\sim} N(\mu_2, \sigma_2^2) \\ 3.\ X_i \perp Y_j\ \forall\ i, j \end{cases}$

Then $(n_1 - 1)S_X^2/\sigma_1^2 \sim \chi^2(n_1 - 1)$,
$(n_2 - 1)S_Y^2/\sigma_2^2 \sim \chi^2(n_2 - 1)$,

$$F = \frac{\chi^2(n_1)}{n_1}\bigg/\frac{\chi^2(n_2)}{n_2} \sim F_{n_1,n_2}, \quad (\textit{peak of its density at } 1). \tag{10.9.1}$$

448. [20] Suppose that $Z \sim N(0,1)$, $X \sim \chi^2(u)$, $Y \sim \chi^2(v)$. If Z____X, $T =$ ____________, then $T \sim t_u$; If X____Y, $F =$ ____, then $F \sim F_{u,v}$ and $X + Y \sim$ ____________.

key: $\underline{\perp}$, $\underline{Z/\sqrt{X/u}}$, $\underline{\perp}$, $\underline{\frac{X/u}{Y/v}}$, $\underline{\chi^2(u+v)}$,

$H_o:$	$\sigma_1^2 = \sigma_2^2$		
$H_a:$	$\sigma_1 > \sigma_2$	$\sigma_1 < \sigma_2$	$\sigma_1 \neq \sigma_2$
$Case:$	(1)	(2)	(3)
$Test\ statistic$	F		$see\ Eq.(10.9.1)$
$RR:$	$F > F_{n_1,n_2,\alpha}$	$1/F > F_{n_2,n_1,\alpha}$	$F > F_{n_1,n_2,\alpha/2}$ or $F < F_{n_1,n_2,1-\alpha/2}$
$Conclusion:$			

Another type of problem is: $\sigma^2 = \sigma_o^2$ **?**

$H_o:$	$\sigma^2 = \sigma_o^2$		
$H_a:$	$\sigma > \sigma_o$	$\sigma < \sigma_o$	$\sigma \neq \sigma_o$
$Case:$	(1)	(2)	(3)
$Test\ statistic$	$\chi^2 = \frac{(n_1-1)S_X^2}{\sigma_o^2}$		
$RR:$	$\chi^2 > \chi^2_{\alpha,n_1-1}$	$\chi^2 < \chi^2_{1-\alpha,n_1-1}$	$\chi^2 < \chi^2_{1-\alpha/2,n_1-1}$ or $\chi^2 > \chi^2_{\alpha/2,n_1-1}$

Ex. 10.16. A company produces machined engine parts that are supposed to have a diameter variance no larger than 0.0002 (diameter in inches).
A random sample of 10 parts gave a sample variance of 0.0003. Conduct a test at $\alpha = 0.05$.

Sol. (1) H_o: $\sigma^2 = 0.0002$,

(2) H_a: $\sigma^2 \neq 0.0002$? H_a: $\sigma^2 > 0.0002$? H_a: $\sigma^2 < 0.0002$?

Key words: no larger than 0.0002, *i.e.*, $\sigma^2 \leq 0.0002$. Its opposite: $\sigma^2 > 0.0002$.

Moreover, $0.0002 < 0.0003 = S^2$, it is likely $\sigma^2 > 0.0002$.

(3) Test statistic: $\chi^2 = \frac{(n-1)S_X^2}{\sigma_o^2} = (10-1)0.0003/0.0002 = 13.5$

(4) RR $\chi^2 > \chi^2_{\alpha,n-1} = \chi^2_{0.05,9} = 16.919$

	$\chi^2_{0.1}$	$\chi^2_{0.05}$	$\chi^2_{0.025}$	$\cdots$	$d.f.$
Display χ^2 table.	$\vdots$				
		16.919			9

> qchisq(.05,9) ?

> qchisq(.95,9) ?

[1] 16.91898

(5) Conclusion: Since $\chi^2 = 13.5 < 16.919$, do not reject H_o,

no evidence to believe $\sigma^2 > 0.0002$.

Ex. 10.17. Under previous assumptions, find the P-value.

Sol. P-value $= P(RR)$, where RR$= \chi^2 > 16.919$.

> 1-pchisq(13.5,9)

[1] 0.1412558 P-value =0.1412558

Quiz on Friday: 447: 9-44, 448: 1-20.

Ex.10.18. An experimenter was convinced that the variability in his measuring equipment results in a standard deviation of 2. 16 measurements yielded $s^2 = 6.1$. Do the data disgree with his claim ? Determine the P-value for the test. What would you conclude if $\alpha = 0.05$.

Sol. (1) H_o: $\sigma = 2$,

(2) H_1: $\sigma \neq 2$? $\sigma > 2$? $\sigma < 2$?

Key words: results in a standard deviation of 2.

(3) Test statistic: $\chi^2 = \frac{(n-1)S_X^2}{\sigma_o^2} = (16-1)6.1/2^2 = 22.875$

(4) RR: $\chi^2 < \chi^2_{1-\alpha/2,n-1}$ or $\chi^2 > \chi^2_{\alpha/2,n-1}$.

> qchisq(0.025,15) [1] 6.262138

> qchisq(0.975,15) [1] 27.48839

Or get from χ^2 table...

(5) Conlusion ?

Do not reject H_o. The variability in the measuring equipment results in an SD of 2.

Remark. The meaning of the critical points:

For χ^2 distribution with degree of freedom ν

$$\begin{pmatrix} \textit{critical pts } x =: & 0 & \chi^2_{1-\alpha,\nu} & \nu & \chi^2_{\alpha,\nu} & \infty \\ P(\chi^2 > x) & 1 & 1-\alpha & \downarrow 0.5 & \alpha & 0 \end{pmatrix}$$

For F with degrees of freedom n_1 and n_2

$$\begin{pmatrix} \textit{critical pts } x =: & 0 & F_{n_1,n_2,1-\alpha} = 1/F_{n_2,n_1,\alpha} & 1 & F_{n_1,n_2,\alpha} & \infty \\ P(F > x) & 1 & 1-\alpha & 0.5 \pm \epsilon & \alpha & 0 \end{pmatrix}$$

Since $F = \frac{\chi^2(u)/u}{\chi^2(v)/v} \sim F_{u,v}$, $1/F = \frac{\chi^2(v)/v}{\chi^2(u)/u} \sim F_{v,u}$,

Ex.10.18(c) Compute the P-value in the example.

$\chi^2(\nu) \sim G(\frac{\nu}{2}, 2)$ with mean ν, the degree of freedom,

> 2*(1-pchisq(22.875,15)) as $22.875 > \nu = 15$

[1] 0.17366028

The P-value= 0.173

(5) Conclusion: Do not reject H_o even if $\alpha = 10\%$ or 15%, let along $\alpha = 0.05$.

Comments. For the χ^2 test with degree of freedom ν and with the test statistic $\chi^2 = y$,

$$\text{the P-value is obtained by the R codes } \begin{cases} 1 - pchisq(y,\nu) & \text{if } H_1\text{: } \sigma > \sigma_o \\ pchisq(y,\nu) & \text{if } H_1\text{: } \sigma < \sigma_o \\ 2 * pchisq(y,\nu) & \text{if } H_1\text{: } \sigma \neq \sigma_o \text{ and } y < \nu \\ 2*(1 - pchisq(y,\nu)) & \text{if } H_1\text{: } \sigma \neq \sigma_o \text{ and } y > \nu \end{cases} \quad (1)$$

448. [20] Suppose that $Z \sim N(0,1)$, $X \sim \chi^2(u)$, $Y \sim \chi^2(v)$. If Z____X, $T =$ ____________, then $T \sim t_u$; If X____Y, $F =$ ____, then $F \sim F_{u,v}$ and $X + Y \sim$ ____________.

key: $\perp$, $Z/\sqrt{X/u}$, $\perp$, $\frac{X/u}{Y/v}$, $\chi^2(u+v)$,

Ex. 10.14 (continued) Suppose that 2 sets of independent samples are obtained from $N(\mu_X, \sigma^2)$ and $N(\mu_Y, \sigma^2)$, respectively.

32, 37, 35, 28, 41, 44, 35, 31, 34,

35, 31, 29, 25, 34, 40, 27, 32, 31,

Do the sample data provide sufficient evidence to indicate that there is difference in true SD for those trained using the two methods at $\alpha = 0.05$? Compute the p-value too.

$H_o:$	$\sigma_1^2 = \sigma_2^2$		
$H_a:$	$\sigma_1 > \sigma_2$	$\sigma_1 < \sigma_2$	$\sigma_1 \neq \sigma_2$
$Case:$	(1)	(2)	(3)
$Test\ statistic$	F		$(see\ Eq.10.9.1)$
$RR:$	$F > F_{n_1,n_2,\alpha}$	$1/F > F_{n_2,n_1,\alpha}$	$F < F_{n_1,n_2,1-\alpha/2}$ or $F > F_{n_1,n_2,\alpha/2}$
$Conclusion:$			

Sol. $n_1 = 9 = n_2$,

$\overline{X} = 35.22$, $\overline{Y} = 31.56$,

$\sum_{i=1}^{n_1}(X_i - \overline{X})^2 = 195.56$,

$\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2 = 160.22$.

H_o: $\sigma_1 = \sigma_2$ v.s.

H_a: $\sigma_1 \neq \sigma_2$? H_a: $\sigma_1 > \sigma_2$? H_a: $\sigma_1 < \sigma_2$?

Key words: "difference in true SD"

Test statistic: $F = S_X^2/S_Y^2 = \frac{195.56}{160.22} = 1.22$

RR: $F > F_{0.025,8,8}$ or $F < F_{1-0.025,8,8}$ $\quad E(F) \approx 1$

Which one to find ?

$F > F_{0.025,8,8} \approx 4.433$ or $F < F_{1-0.025,8,8} = 1/4.433 = 0.2256$

```
> qf(0.95,8,8)
[1] 3.438101
> qf(0.975,8,8)
[1] 4.43326
```

Use F-table ......

Conclusion: Do not reject H_o, there is no evidence to suspect that there is a difference in the SD's for those trained using the two methods

Ex. 10.14(c). Compute the P-value in the example.

```
> pf(1.22,8,8)
[1] 0.6073314
```

p-value$= 2 * (1 - 0.607) \approx 0.8$.

Remark. For test statistic value $F = \frac{\chi^2(u)/u}{\chi^2(v)/v} = y$,

$$\text{P-value}=\begin{cases} pf(y,u,v) & \text{if } H_a\text{: } \sigma_1 < \sigma_2 \\ 1-pf(y,u,v) & \text{if } H_a\text{: } \sigma_1 > \sigma_2 \\ 2pf(y,u,v)? & \\ 2(1-pf(1/y,v,u))? & \\ pf(y,u,v)+(1-pf(1/y,v,u))? & \text{if } H_a\text{: } \sigma_1 \neq \sigma_2 \text{ and } y<1 \\ 2(1-pf(y,u,v))? & \\ 2pf(1/y,v,u) & \\ 1-pf(y,u,v)+pf(1/y,v,u) & \text{if } H_a\text{: } \sigma_1 \neq \sigma_2 \text{ and } y>1. \end{cases} \quad (2)$$

Ex. 10.19. Suppose that we wish to compare the variation in diameters of parts produced by the company with that produced by a competitor. Our company results in $S^2 = 0.0003$ with $n = 10$, and the competitor yielded $s_2^2 = 0.0001$ with $n = 20$. Do the data provide sufficient information to indicate a smaller variation in diameters for the competitor ? Test with $\alpha = 0.05$ and compute the P-value.

Sol: H_o: $\sigma_1 = \sigma_2$ v.s.

H_a: $\sigma_1 \neq \sigma_2$? H_a: $\sigma_1 > \sigma_2$? H_a: $\sigma_1 < \sigma_2$?

Key words: a smaller variation in diameters for the competitor

Test statistic: $F = \frac{S_X^2}{S_Y^2} = 3$

RR: $F > F_{9,19,0.05} = 2.42$

Conclusion: Reject H_o, the data provide sufficient information to indicate a smaller variation in diameters for the competitor.

P-value: use which formulat in Eq. (2) ?

```
> 1-pf(3,9,19)
[1] 0.02096038 P-value
```

Remark. Reconsider the problem of testing H_o: $\mu_X = \mu_Y$, we need to check

1. X_i's and Y_i' are indeed i.i.d.;
2. X_i and Y_i are indeed from $N(\mu, \sigma_i^2)$;
3. $\sigma_X = \sigma_Y$,

due to the assumption:

$T = \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\hat{\sigma}_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$,

under the assumption that $\begin{cases} 1.\ X_1, ..., X_{n_1} \overset{i.i.d.}{\sim} N(\mu_X, \sigma^2) \\ 2.\ Y_1, ..., Y_{n_2} \overset{i.i.d.}{\sim} N(\mu_Y, \sigma^2) \\ 3.\ X_i \perp Y_j \ \forall\ i, j \end{cases}$

In the future, we may learn how to check assumptions 1 and 2.

§10.10. Power of tests and the Neyman-Pearson Lemma.

A test consists of 5 elements:

H_o, say $\theta = \theta_o$ or $\theta \in \Theta_o$; $\Theta_o = ?$

H_a, say $\theta \neq \theta_o$, or $\theta \in \Theta_a$;

Test statistic;

RR;

Conclusion.

So far, Θ_o consists of only one element, *e.g.*, $\theta = \theta_o$, or $\theta = 0$ etc.

In such case, H_o is called a simple hypothesis.

In some examples, we have, *e.g.*, H_o: $p \geq 0.5$, then Θ_o is a composite hypothesis.

On the other hand, most of the time, we have H_a: $p < 0.5$ or $\mu \neq 0$ *etc.*, then H_a is a composite hypothesis. However there are cases that H_a is a simple hypothesis. Notice that Θ_a and Θ_o are two sets for the two hypotheses.

Def. 10.3. Let W be the test statistic and RR the rejection region. For a test of a hypothesis involving the value of the parameter θ, the power of the test, denoted by $\mathcal{P}(\theta)$, is

$$\mathcal{P}(\theta) = P(W \in RR \text{ when the parameter value is } \theta).$$

For simple hypotheses, $\mathcal{P}(\theta_o) = \alpha = P(H_1|H_o)$ and $\mathcal{P}(\theta_a) = 1 - \beta = 1 - P(H_o|H_1)$.

Theorem 10.1. The Neyman-Pearson Lemma. Suppose that we wish to test the simple hypotheses H_o: $\theta = \theta_o$ v.s. H_a: $\theta = \theta_a$, based on a random sample Y_1, ..., Y_n from a distribution with parameter θ. Let $L(\theta) = \prod_{i=1}^n f(Y_i; \theta)$. For a given α, the test that maximizes the power at θ_o has a RR determined by $\frac{L(\theta_o)}{L(\theta_a)} \leq k$, the value k is chosen so that the test has the desired value of α. Such a test is called the most powerful test (MP test) for H_o versus H_a.

Ex. 10.22. Suppose that the observation is $Y \sim f(y|\theta) = \theta y^{\theta-1}$, $0 < y < 1$.

Find the MP test with $\alpha = 0.05$ to test H_o: $\theta = 2$ v.s. H_a: $\theta = 1$.

Sol. Is H_o simple hypothesis ?

Is H_a simple hypothesis ?

Test statistic is Y.

$$\begin{aligned} RR: \quad \frac{L(\theta_o)}{L(\theta_a)} &= f(y|\theta_o)/f(y|\theta_a) \\ &= \frac{\theta_o y^{\theta_o - 1}}{\theta_a y^{\theta_a - 1}} \\ &= 2y \leq k \text{ where } y \in (0,1). \end{aligned}$$

$$RR: \quad Y \leq k/2 = y_o.$$

Need to find $k = ?$ or $y_o = ?$

$$\begin{aligned}
\alpha =& 0.05 \\
=& P(H_1|H_o) \\
=& P(RR,\ \theta = 2) \\
=& P(Y \leq y_o,\ \theta = 2) \\
=& \int_0^{y_o} 2y^{2-1} dy \qquad\qquad\qquad\qquad f = \theta y^{\theta-1} \\
=& y^2 \Big|_0^{y_o} \\
=& y_o^2 \\
=> \quad y_o =& \sqrt{0.05} = 0.2236.
\end{aligned}$$

RR: $Y \leq 0.2236$.

$\mathcal{P}(2) = ?$ $\qquad$ H_o: $\theta = 2$

$P(H_1|H_o) =$Probability of type I error $= ?$

$\mathcal{P}(1) = P(Y < y_o$ if $\theta = 1)$

$= \int_0^{y_o} 1y^{1-1}dy \qquad f = \theta y^{\theta-1}$

$= y_o = 0.2236.$ $\qquad$ v.s. $\mathcal{P}(2) = 0.2236^2 = 0.05.$

$\beta(1) =$probability of type II error

$= P(H_o|H_1)$

$= 1 - \mathcal{P}(1)$

$= 1 - y_o = 1 - 0.2236 = 0.7764.$

Q: What happen if RR is $Y > \sqrt{0.95}$?

$\alpha_2 = P(Y > \sqrt{0.95}|\theta = 2)$

$= \int_{\sqrt{0.95}}^1 2y^{2-1}dy$

$= y^2\Big|_{\sqrt{0.95}}^1$

$= 1 - 0.95$

$= 0.05$

$\mathcal{P}_2(1) = \int_{\sqrt{0.95}}^1 y^{1-1}dy$

$= 1 - \sqrt{0.95}$

$= 1 - 0.975 = 0.025 = P(Y < \sqrt{0.95}$ if $\theta = 1) < P(Y < 0.2236$ if $\theta = 1)$

$\beta_2(1) =$probability of type II error

$= 1 - \mathcal{P}_2(1)$

$= 0.975$

> 0.7764

$= \beta(1)$

That is, the test with RR $Y \le 0.2236$ is more powerful than the test with RR $Y \ge 0.975$.

Or the test with RR $Y \le 0.2236$ has smaller $P(H_o|H_1)$ than the test with RR $Y \ge 0.975$.

Q: How about MP test for a composite hypothesis test ?

Ans. N-P Lemma works if RR is the same for each pair of θ_o and θ_a, where θ_o is under H_o and θ_a is under H_a.

Ex. 10.23. Suppose that X_1, ..., X_n are i.i.d. from $N(\mu, \sigma^2)$, where σ^2 is known. Find the uniformly MP test of level α for testing H_o: $\mu = \mu_o$ v.s. H_a: $\mu > \mu_o$.

Sol. Let $\mu > \mu_o$.

$$\begin{aligned}
L(\mu) =& \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\
=& (\frac{1}{\sqrt{2\pi\sigma^2}})^n e^{-\sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2}} \\
k \ge \frac{L(\mu_o)}{L(\mu)} =& \frac{(\frac{1}{\sqrt{2\pi\sigma^2}})^n e^{-\sum_{i=1}^n \frac{(x_i-\mu_o)^2}{2\sigma^2}}}{(\frac{1}{\sqrt{2\pi\sigma^2}})^n e^{-\sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2}}} \\
=& e^{-\sum_{i=1}^n \frac{(x_i-\mu_o)^2}{2\sigma^2} + \sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2}} \\
=& e^{\sum_{i=1}^n \frac{2x_i(\mu_o-\mu)-\mu_o^2+\mu^2}{2\sigma^2}} \\
\ln k \ge& \sum_{i=1}^n \frac{2x_i(\mu_o-\mu)-\mu_o^2+\mu^2}{2\sigma^2} \\
2\sigma^2 \ln k \ge& \sum_{i=1}^n (2x_i(\mu_o-\mu)-\mu_o^2+\mu^2) \\
=& 2n\overline{x}(\mu_o-\mu) - n(\mu_o^2-\mu^2) \\
\overline{x} \ge& \frac{-2\sigma^2 \ln k - n(\mu_o^2-\mu^2)}{2n(-\mu_o+\mu)} \\
\overline{x} \ge& c \\
\alpha =& P(\overline{X} \ge c, \mu = \mu_o) = \Phi(\frac{c-\mu_o}{\sigma/\sqrt{n}})
\end{aligned}$$

Thus the UMP test has a RR $\overline{X} \ge c$, where $\frac{c-\mu_o}{\sigma/\sqrt{n}} = z_\alpha$, *i.e.* $c = \mu_o + z_\alpha \sigma/\sqrt{n}$.

448 [21] The MP test for H_o: $\theta = \theta_o$ v.s. H_a: $\theta = \theta_a$, the MP test has the RR satisfying: $\frac{L(\theta_o)}{L(\theta_a)}$____$k$ and $P_\theta(RR) = \alpha$ if $\theta =$ ____. **key:** $\underline{\le}$, $\underline{\theta_o}$,

Def. For composite hypothesis H_o, the size α of the test is

$$\alpha = \sup_{\theta \in H_o} \mathcal{P}_\theta = P(RR|\theta \in H_o)$$

The test is a level α_1 test if $\alpha \leq \alpha_1$.

Ex. 10.22(c): Suppose that the observation is $Y \sim f(y|\theta) = \theta y^{\theta-1}$, $0 < y < 1$. Find the size $\alpha = 0.05$ MP test for H_o: $\theta \geq 2$ v.s. H_a: $\theta < 2$.

Sol. There are two types of problem for testing hypothesis.

(1) Data are given, carry out the test by presenting the 5 elements of a test.

(2) Data are not given, present the first 4 elements of a test.

This example belongs to the 2nd case.

The RR is the same as the case H_o: $\theta = 2$ v.s. H_a: $\theta = 1$,

i.e. RR: $Y \leq y_o = \sqrt{0.05} = 0.2236$.

The proof is as follows. $P(H_1|H_o)$ is not uniquely defined in this case. Let $P_\theta = P(RR|\theta)$, then

$$\begin{aligned}\mathcal{P}_\theta = &\int_0^{y_o} \theta y^{\theta-1} dy \qquad\qquad \theta > 0\\ =&y_o^\theta \quad \text{decreases from 1 to 0, as } \theta \text{ increases from 0 to } \infty.\end{aligned}$$

$$\mathcal{P}_\theta = y_o^\theta \downarrow_0^1 \text{ as } \theta \to \infty \text{ from 0, and } \mathcal{P}_\theta \begin{cases} \downarrow_0^{0.05} & \text{if } \theta \uparrow_2^\infty \\ \uparrow_{0.05}^1 & \text{if } \theta \downarrow_0^2 \end{cases}$$

$P(H_1|\theta \in H_o) = 0.05^{\theta/2}$, where $\theta \geq 2$.

The size of the test is

$$\begin{aligned}\alpha =& \sup_{\theta \geq 2} \mathcal{P}_\theta(RR)\\ =&y_o^\theta\Big|_{\theta=2}\\ =&y_o^2\\ =&0.05 \qquad => \quad y_o = \sqrt{0.05}\end{aligned}$$

The level of the test is 0.05 ? or 0.1 ?

$P(H_o|\theta \in H_1) = 1 - 0.05^{\theta/2}$, where $\theta \in [0, 2)$.

Note that it is reasonable that RR is $Y < y_o$, as

$E(Y) = \int_0^1 y f(y|\theta) dy$

$= \int_0^1 y\theta y^{\theta-1} dy$

$= \int_0^1 \theta y^\theta dy$

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

$$\frac{\theta}{\theta+1}\Big|_{\theta \geq 2} \geq \frac{2}{3} > \frac{\theta}{\theta+1}\Big|_{\theta < 2}. \qquad\qquad Y \approx 0 \;<=>\; \theta \approx 0$$

$$E(Y|H_o) = E(Y|\theta \geq 2) \geq E(Y|\theta = 2) > E(Y|\theta < 2) = E(Y|H_1).$$

Remark. Let $X \sim U(-0.5+p, 0.5+p)$. H_o: $p=0$, v.s. H_a: $p \neq 0$. $\alpha = 0.05$. There is no UMP test. This can be shown as follows.

Notice that $E(X) = \frac{0.5+p+(-0.5+p)}{2} = p$.
If $X = 0$ or X is close to 0, we will believe that $p = 0$, *i.e.*,
a reasonable test for H_o: $p = 0$ is the one to reject $p = 0$ if X is far away from 0.
If we set $\alpha = 0.05$, then it is $\phi_1 = I(|X| > 0.475) = I(RR_1)$, as

$$\begin{aligned} E(\phi_1 | p=0) =& P(|X| > 0.475 | p=0) \\ =& \int_{-\infty}^{-0.475} 1dx + \int_{0.475}^{\infty} 1dx? \quad \int_{-0.5}^{-0.475} 1dx + \int_{0.475}^{0.5} 1dx? \\ =& 0.025 + 0.025 \\ =& 0.05. \end{aligned}$$

Consider another two tests: $(\phi_1 = I(|X| > 0.475) = I(RR_1))$

2. $\phi_2 = I(X > 0.45) = I(RR_2)$,
3. $\phi_3 = I(X < -0.45) = I(RR_3)$.

The size of the 3 tests are all $\alpha = 0.05$.

If $p = 0.1$, the powers $\begin{cases} \mathcal{P}_1(0.1) =? = P(X \in RR_1 \; if \; p = 0.1) = P(X \in (0.475, 0.6) = 0.125 \\ \mathcal{P}_2(0.1) =? = P(X \in RR_2 \; if \; p = 0.1) = P(X \in (0.45, 0.6) = 0.15 \\ \mathcal{P}_3(0.1) =? = P(X \in RR_3 \; if \; p = 0.1) = 0 \end{cases}$

thus ϕ_2 is more powerful than ϕ_1 and ϕ_3, and ϕ_1 is more powerful than ϕ_3,

If $p = -0.1$, the powers $\begin{cases} \mathcal{P}_1(-0.1) = P(X \in RR_1 \; if \; p = -0.1) = 0.125 \\ \mathcal{P}_2(-0.1) = P(X \in RR_2 \; if \; p = -0.1) = 0 \\ \mathcal{P}_3(-0.1) = P(X \in RR_3 \; if \; p = -0.1) = 0.15 \end{cases}$

thus ϕ_3 is more powerful than ϕ_1 and ϕ_2, and ϕ_2 is more powerful than ϕ_1,

ϕ_1 is the most reasonable test for H_o v.s. H_1 in this example, but no MP test!

Example 3. Suppose X_1, ..., X_n are i.i.d. from $f(x|\theta) = \theta e^{-x\theta}$, $x > 0$. Find the MP test for testing H_o: $\theta \leq 1$ with $\alpha = 0.1$.

Sol. H_o: $\theta \leq 1$,

H_1: $\theta > 1$.

To find the RR, the NP lemma needs to compute $L(\theta)$.

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} \theta exp(-X_i \theta) \\ &= \theta^n exp(-\sum_{i=1}^{n} X_i \theta) \end{aligned}$$

One may consider the sufficient statistic $Y = \sum_{i=1}^{n} X_i$, instead of X_1, ..., X_n.
Distribution of Y ?

1. Moment generating function method.

$$
\begin{aligned}
M_Y(t) =& E(exp(\sum_{i=1}^{n} X_i t)) \\
=& E(\prod_{i=1}^{n} exp(X_i t)) \\
=& (E(exp(X_1 t)))^n \\
=& (\int_0^\infty e^{xt}\theta e^{-\theta x} dx)^n \\
=& (\int_0^\infty \theta e^{-(\theta - t)x} dx)^n \\
=& (\frac{\theta}{\theta - t})^n
\end{aligned}
$$

Recall Gamma distribution W has mgf

$$
\begin{aligned}
M_W(t) =& E(e^{Wt}) \\
=& \int_0^\infty e^{xt} \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} dx \\
=& \int_0^\infty \frac{x^{\alpha-1}e^{-x(1/\beta - t)}}{\beta^\alpha \Gamma(\alpha)} dx \\
=& \frac{1}{(1/\beta - t)^\alpha \beta^\alpha} \\
=& (\frac{\theta}{\theta - t})^\alpha \text{ if } \theta = 1/\beta \\
=& \begin{cases} \frac{\theta}{\theta - t} & \text{if } \alpha = 1 \\ (\frac{\theta}{\theta - t})^n & \text{if } \alpha = n \end{cases}
\end{aligned}
$$

$X_1 \sim G(1, 1/\theta)$. $Y = \sum_{i=1}^n X_i \sim G(n, 1/\theta)$.
2. 447 [44] => $G(\alpha, \beta) + G(\alpha, \beta) = G(2\alpha, \beta)$.
3. Test statistic: $Y = \sum_{i=1}^n X_i$.
4. RR: $L(\theta) = \prod_{i=1}^n \theta e^{-x_i \theta}$.

$$
\begin{aligned}
\frac{L(\theta_o)}{L(\theta_1)} =& \frac{\theta_o^n t^{n-1} e^{-\theta_o t}}{\theta_1^n t^{n-1} e^{-\theta_1 t}} & f(t) = \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)} \\
=& \frac{e^{(\theta_1 - \theta_o)t}}{(\theta_1/\theta_o)^n} \le k & \theta_1 > \theta_o \\
& t \le c
\end{aligned}
$$

$$
RR: \quad Y = \sum_{i=1}^n X_i \le c, \text{ where } c \text{ satisfies}
$$

$$
0.05 = \int_0^c \frac{\theta^n t^{n-1} e^{-\theta t}}{\Gamma(n)} dt = \int_0^c \frac{t^{n-1} e^{-t}}{\Gamma(n)} dt \qquad at\ \theta = 1
$$

Quiz on Friday: 448 1-22, 447: 1-16.

§10.11. Likelihood Ratio test (LRT)

We shall introduce a test method called the Likelihood Ratio test (LRT). We first define some notiatons. So far, most of the time, we denote

H_o: $\theta = \theta_o$ v.s. H_1: $\theta \neq \theta_o$, or H_1: $\theta < \theta_o$, or H_1: $\theta > \theta_o$.

These can be written as

H_o: $\theta \in \Theta_o$ v.s. H_a: $\theta \notin \Theta_o$. where $\Theta_o \subset \Theta$.

Ex. 1. (a) H_o: $\theta = \theta_o$, v.s. H_a: $\theta \neq \theta_o$.

=> $\Theta_o = \{\theta_o\}$ and $\Theta = \mathcal{R}$.

(b) H_o: $\theta = \theta_o$, v.s. H_a: $\theta > \theta_o$.

=> $\Theta_o = \{\theta_o\}$ and $\Theta = [\theta_o, \infty)$.

448 [22] The Likelihood ratio test for H_o: $\theta \in \Theta_o$ v.s. H_a: $\theta \notin \Theta_o$ has a RR: $\{\lambda$____$k\}$, where

$\lambda =$ __________; $\hat\theta_o$ is the MLE under ____; $\hat\theta$ is the MLE under ____;

k satisfies $max\{P(RR): \ \theta \in \Theta_o\} =$ ____;

if n is large, then $-2\ln\lambda$ is approximated _______;

where $v =$ ___________; r and $r_o = \#$ of free parameters in Θ and in Θ_o, respectively.

key: $\underline{\leq}$, $\underline{\frac{L(\hat\theta_o)}{L(\hat\theta)}}$, $\underline{\Theta_o}$, $\underline{\Theta}$, $\underline{\alpha}$, $\underline{\chi^2(v)}$, $\underline{r - r_o}$

Ex. 10.24. Assume that X_1, ..., $X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$, H_o: $\mu = 0$ v.s. H_a: $\mu \neq 0$. LRT of size 0.05 ?

Sol. Nee to solve $\lambda = \frac{L(\hat\theta_o)}{L(\hat\theta)}$.

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

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

Step 1. Under Θ_o, MLE $\hat\theta_o$: $\hat\mu_o = 0$, $\hat\sigma_o^2 = \overline{X^2}$ because

$L(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}}$.

$\ln L(\theta_o) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{\sum_{i=1}^n X_i^2}{2\sigma^2}$.

$(\ln L)'_{\sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^n X_i^2}{2\sigma^4} = 0$

=> $\hat\sigma^2 = \overline{X^2}$

Check:

$\theta:$	$\sigma^2 = 0$	$\sigma^2 = \infty$	$\hat\sigma_o^2$
$L(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i)^2}{2\sigma^2}}:$	0	0	>0

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

$L(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}}$.

$\ln L = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{n\sum_i (X_i-\mu)^2}{2\sigma^2}$.

$(\ln L)'_{\sigma^2} = 0 \Rightarrow \hat\sigma^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2$

$(\ln L)'_{\mu} = 0 \Rightarrow \hat\mu = \overline{X}$. a—arbitrary

Check:

$\theta = (\mu, \sigma):$	$(a, 0)$	(a, ∞)	$(\overline{X}, \hat\sigma_o^2)$	$(-\infty, a)$	(∞, a)
$L(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}}:$	0	0	>0	0	0

By Steps 1 and 2,

$$\begin{aligned}\lambda =&L(\hat{\theta}_o)/L(\hat{\theta})\\ =&\frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat{\sigma}_o^2}}e^{-\frac{X_i^2}{2\hat{\sigma}_o^2}}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat{\sigma}^2}}e^{-\frac{(X_i-\hat{\mu})^2}{2\hat{\sigma}^2}}}\\ =&(\frac{\hat{\sigma}^2}{\hat{\sigma}_o^2})^{n/2}\frac{e^{-n/2}}{e^{-n/2}}\\ =&(\frac{\overline{X^2}-(\overline{X})^2}{\overline{X^2}})^{n/2}\end{aligned}$$

RR: $\lambda \le k$

$<=>$ $(\frac{\overline{X^2}-(\overline{X})^2}{\overline{X^2}})^{n/2} \le k \quad (<1)$

$<=>$ $\frac{\overline{X^2}-(\overline{X})^2}{\overline{X^2}} \le k_2 \quad (<1)$

$<=>$ $1-\frac{(\overline{X})^2}{\overline{X^2}} \le k_2 \quad (<1)$

$<=>$ $\frac{(\overline{X})^2}{\overline{X^2}} \ge k_2$

$<=>$ $\frac{\overline{X^2}}{(\overline{X})^2} \le k_3$

$<=>$ $\frac{\overline{X^2}}{(\overline{X})^2} - 1 \le k_4$

$<=>$ $\frac{\overline{X^2}-(\overline{X})^2}{(\overline{X})^2} \le k_4$

$<=>$ $\frac{\frac{n-1}{n}(\overline{X^2}-(\overline{X})^2)}{(\overline{X})^2} \le k_5$

$<=>$ $\frac{(\overline{X})^2}{\frac{n-1}{n}(\overline{X^2}-(\overline{X})^2)} \ge 1/k_5$ v.s. $(\frac{\overline{X^2}-(\overline{X})^2}{\overline{X^2}})^{n/2} \le k$

$<=>$ $\frac{|\overline{X}|}{S/\sqrt{n}} \ge t_{\alpha/2,n-1}$ RR for the LRT test.

Remark. In this example, we have the exact distribution of the LRT, not an approximate one. Thus no need to use approximated χ^2 distribution.

If we do use approximation, then $r_o = 1$ and $r = 2$. $-2\ln(\frac{\overline{X^2}-(\overline{X})^2}{\overline{X^2}})^{n/2} \approx \chi^2(1)$, but n should be large.

Ex. 24(c). Assume that X_1, ..., $X_n \overset{i.i.d.}{\sim} N(\mu,\sigma^2)$, $\theta = (\mu,\sigma^2)$, H_o: $\mu = 0$ v.s. H_a: $\mu > 0$. LRT of size 0.05 ?

Sol. Solve $\lambda = L(\hat{\theta}_o)/L(\hat{\theta}) \le k$.

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

$\Theta = \{(\mu,\sigma^2): \ \mu \ge 0, \ \sigma > 0\}$,

Under Θ_o, MLE $\hat{\theta}_o$: $\hat{\mu}_o = 0$, $\hat{\sigma}_o^2 = \overline{X^2}$ for the same reason as Ex.10.24.

Under Θ,

$$L(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(X_i-\mu)^2}{2\sigma^2}}$$

$$
\begin{aligned}
\ln L =& \frac{n}{2}\ln(2\pi\sigma^2) - \frac{\sum_{i=1}^n (X_i-\mu)^2}{2\sigma^2} \\
(\ln L)'_\mu =& \sum_{i=1}^n (X_i-\mu)/\sigma^2 = 0 && => \quad \hat\mu = \overline{X} \\
\textbf{Check } (\ln L)''_\mu =& -\sum_{i=1}^n 1/\sigma^2 < 0 && \ln L \text{ is concave down}
\end{aligned}
$$

$$=> \hat\mu = \overline{X} \text{ ? or } \hat\mu = \overline{X} \vee 0 = \max\{\overline{X}, 0\} \text{ ? } \textbf{Why ?}$$

$$
\begin{pmatrix}
Cases: & & 0 \leq \overline{X} & & | & & \overline{X} < 0 & & \\
\mu: & 0 & \overline{X} & \infty & | & -\infty & \overline{X} & 0 & \infty \\
L(\theta): & + & < L(\overline{X}) & 0 & | & ignore & ignore & > 0 & 0
\end{pmatrix}
$$

$$
\begin{aligned}
(\ln L)'_{\sigma^2} = 0 &=> \sigma^2 = \tfrac{1}{n}\textstyle\sum_{i=1}^n (X_i-\mu)^2 \\
&=> \hat\sigma^2 = \tfrac{1}{n}\textstyle\sum_{i=1}^n (X_i-\hat\mu)^2
\end{aligned}
$$

$$
\text{Check: } \begin{pmatrix}
\sigma: & & \sigma^2 = 0 & \sigma^2 = \infty & \hat\sigma^2 \\
L(\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} & : & 0 & 0 & > 0
\end{pmatrix}
$$

Thus under Θ, MLE $\hat\theta$: $\hat\mu = \overline{X} \vee 0$, and $\hat\sigma^2 = \frac{1}{n}\sum_{i=1}^n (X_i - (\overline{X} \vee 0))^2$

$$
\begin{aligned}
\lambda =& \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat\sigma_o^2}} e^{-\frac{(X_i-\hat\mu_o)^2}{2\hat\sigma_o^2}}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat\sigma^2}} e^{-\frac{(X_i-\hat\mu)^2}{2\hat\sigma^2}}} && \hat\mu_o = 0 \\
=& (\frac{\hat\sigma^2}{\hat\sigma_o^2})^{n/2} \frac{exp(\frac{-\sum_{i=1}^n x_i^2}{2\sum_{i=1}^n x_i^2/n})}{exp(-\frac{\sum_{i=1}^n (x_i-\hat\mu)^2}{2\sum_{i=1}^n (x_i-\hat\mu)^2/n})} \\
=& \frac{(\prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat\sigma_o^2}}) e^{-n/2}}{(\prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat\sigma^2}}) e^{-n/2}} = (\frac{\hat\sigma^2}{\hat\sigma_o^2})^{n/2} && (1) \\
-2\ln\lambda =& -2\ln(\frac{\hat\sigma^2}{\hat\sigma_o^2})^{n/2} \approx \chi^2(r-r_o) \text{ if } n \text{ is large. } (r, r_o) = ??
\end{aligned}
$$

If $n < 20$, then we can derive the exact distribution of λ as follows.

$$
\begin{aligned}
\hat\sigma^2 &= \tfrac{1}{n}\textstyle\sum_{i=1}^n (X_i - (\overline{X} \vee 0))^2 \\
&= \tfrac{1}{n}\textstyle\sum_{i=1}^n [X_i^2 - 2(\overline{X} \vee 0)X_i + (\overline{X} \vee 0)^2] \\
&= \tfrac{1}{n}\textstyle\sum_{i=1}^n X_i^2 - 2\tfrac{1}{n}\sum_{i=1}^n (\overline{X} \vee 0)X_i + \tfrac{1}{n}\sum_{i=1}^n (\overline{X} \vee 0)^2 \\
&= \tfrac{1}{n}\textstyle\sum_{i=1}^n X_i^2 - 2(\overline{X} \vee 0)\overline{X} + (\overline{X} \vee 0)^2 \\
&= \tfrac{1}{n}\textstyle\sum_{i=1}^n X_i^2 - 2(\overline{X} \vee 0)^2 + (\overline{X} \vee 0)^2 \\
&= \overline{X^2} - (\overline{X} \vee 0)^2
\end{aligned}
$$

$$\lambda = L(\hat\theta_o)/L(\hat\theta)$$

$$=(\frac{\hat\sigma^2}{\hat\sigma_o^2})^{n/2} \qquad \text{by Eq. (1) in last page}$$
$$=(\frac{\overline{X^2}-(\overline{X}\vee 0)^2}{\overline{X^2}})^{n/2} \qquad \sigma_o^2=\overline{X^2}$$
$$\lambda=\begin{cases}(\frac{\overline{X^2}-(\overline{X}\vee 0)^2}{\overline{X^2}})^{n/2} & \text{if } \overline{X}>0\\ 1 & \text{otherwise}\end{cases} \le k \; for \; RR, \quad v.s. \;\; \lambda=\frac{\overline{X^2}-(\overline{X})^2}{\overline{X^2}}\le k \text{ for } H_1:\mu\ne 0$$

Notice that $P(\lambda=1)=0.5=P(\lambda<1)$ under H_o.

RR: $\lambda\le k<1$

$<=> (\frac{\overline{X^2}-(\overline{X}\vee 0)^2}{\overline{X^2}})^{n/2}\le k \quad (<1)$

$<=> \frac{\overline{X^2}-(\overline{X}\vee 0)^2}{\overline{X^2}}\le k_1 \quad (<1)$

$<=> 1-\frac{(\overline{X}\vee 0)^2}{\overline{X^2}}\le k_1 \quad (<1)$

$<=> \frac{(\overline{X}\vee 0)^2}{\overline{X^2}}\ge k_2 \; (\in(0,1))$

$<=> \frac{\overline{X^2}}{(\overline{X}\vee 0)^2}\le k_3 \; (\in(1,\infty))$

$<=> \frac{\overline{X^2}}{(\overline{X}\vee 0)^2}-1\le k_4 \; (\in(0,\infty))$

$<=> \frac{\overline{X^2}-(\overline{X}\vee 0)^2}{(\overline{X}\vee 0)^2}\le k_4$

$<=> \frac{(\overline{X}\vee 0)^2}{\frac{n}{n-1}(\overline{X^2}-(\overline{X}\vee 0)^2)}\ge k_5 \; (\in(0,\infty))$

$<=> \frac{\overline{X}\vee 0}{S/\sqrt{n}}\ge\sqrt{k_5}$ For H_1: $\mu\ne 0$, the RR is $\frac{|\overline{X}|}{S/\sqrt{n}}\ge t_{\alpha/2,n-1}$.

$<=> \frac{\overline{X}\vee 0}{S/\sqrt{n}}\ge t_{\alpha,n-1}$?? or $\frac{\overline{X}\vee 0}{S/\sqrt{n}}\ge t_{\alpha/2,n-1}$??

$<=> \frac{\overline{X}}{S/\sqrt{n}}\ge t_{\alpha,n-1}$ RR for the LRT test.

Remark. In this example, we also have the exact distribution of the LRT, not an approximate one. Thus no need to use approximated χ^2 distribution.

Ex.10.25. Suppose that an engineer wishes to compare the number of complaints per week filed by union stewards for two different shifts at a plant. 100 independent observations yield $\overline{x}=20$ for shift 1 and $\overline{y}=22$ for shift 2. Suppose that the number of complaints per week on the ith shift has the Poisson distribution with mean θ_i, for $i=1, 2$. Use the LRT method to test H_o: $\theta_1=\theta_2$ v.s. H_1: $\theta_1\ne\theta_2$ with $\alpha\approx 0.01$.

Sol. Under Θ, the MLE of $\theta_1=\overline{x}$, the MLE of $\theta_2=\overline{y}$. # of parameters $r=$??

Under H_o, the MLE of $\theta_1=\theta_2$, thus $\hat\theta_o=(\overline{x}+\overline{y})/2$. # of parameters $r_o=$?? The likelihood

$$\begin{aligned}L(\theta)=L(\theta_1,\theta_2)=&\prod_{i=1}^{100}e^{-\theta_1}\frac{\theta_1^{x_i}}{x_i!}\prod_{i=1}^{100}e^{-\theta_2}\frac{\theta_2^{y_i}}{y_i!}\\ \propto&\prod_{i=1}^{100}e^{-\theta_1}\theta_1^{x_i}\prod_{i=1}^{100}e^{-\theta_2}\theta_2^{y_i}\\ =&e^{-100\theta_1}\theta_1^{\sum_{i=1}^{100}x_i}e^{-100\theta_2}\theta_2^{\sum_{i=1}^{100}y_i}\end{aligned}$$

$$
\begin{aligned}
&= e^{-n\theta_1}\theta_1^{n\overline{x}}e^{-n\theta_2}\theta_2^{n\overline{y}} && n = 100\\
L(\hat{\theta}_o) &= e^{-2n\hat{\theta}_o}\hat{\theta}_o^{n(\overline{x}+\overline{y})}\\
&= e^{-\sum_i (x_i+y_i)}\left(\frac{\sum_i (x_i+y_i)}{2n}\right)^{\sum_i (x_i+y_i)}\\
L(\hat{\theta}) &= e^{-\sum_i x_i}\left(\frac{\sum_i x_i}{n}\right)^{n\overline{x}}e^{-\sum_i y_i}\left(\frac{\sum_i y_i}{n}\right)^{n\overline{y}}\\
\lambda = \frac{L(\hat{\theta}_o)}{L(\hat{\theta})} &= \frac{\left(\frac{\sum_i (x_i+y_i)}{2n}\right)^{n\overline{x}+n\overline{y}}}{\left(\frac{\sum_i x_i}{n}\right)^{n\overline{x}}\left(\frac{\sum_i y_i}{n}\right)^{n\overline{y}}} && \overline{x} = 20 \ \& \ \overline{y} = 22\\
&= \frac{21^{100(20+22)}}{20^{100(20)}22^{100(22)}}\\
-2\ln\lambda &= 9.53
\end{aligned}
$$

H_o: $\theta_1 = \theta_2$ v.s.
H_1: $\theta_1 \neq \theta_2$
Test statistic: $-2\ln\lambda$.
RR: $\lambda \leq k$,
$-2\ln\lambda \geq \chi^2_{0.1,1} = g$, $v = r - r_o$. $(r_o, r) = ??$
$g = 6.635$.
> qchisq(0.99,1)
[1] 6.634897
> 1-pchisq(9.53,1)
[1] 0.002021401
$k = 6.635$. $-2\ln\lambda = 9.53 > k = 6.635$
Conclusion: Do reject H_o. There is a difference in the number of complaints per week filed by union stewards for two different shifts at a plant.

Remark. In this example, we can only use the approximate distribution of $-2\ln\lambda$. Also $n = 100$.

Quiz on Friday: 448: [1]-[22] all 447

15. $Y = g(X)$. $E(g(X)) = \begin{cases} \sum_y y f_Y(y) & \text{dis} \\ \int y f_Y(y)dy & \text{cts} \end{cases} = \begin{cases} \underline{\sum_x g(x)f_X(x)} & \text{dis} \\ \underline{\int g(x)f_X(x)dx} & \text{cts} \end{cases}$,
16. The mgf of X is $M(t) = \underline{E(e^{Xt})}$, $\frac{d^kM(t)}{dt^k}\big|_{t=0} = \underline{E(X^k)}$
23. $X \sim \mathcal{G}(\alpha, \beta)$. $f(x) = \underline{\frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha}}$, if $x > \underline{0}$, $\mu = \underline{\alpha\beta}$, $\sigma^2 = \underline{\alpha\beta^2}$, $\Gamma(\alpha+1) = \underline{\alpha\Gamma(\alpha)}$
24. $Exp(\lambda) = \underline{\mathcal{G}(1,\lambda)}$, $\chi^2(\nu) = \underline{\mathcal{G}(\frac{\nu}{2}, 2)}$
40. Let Y_1, ..., Y_n be a random sample of Y. $\overline{Y} = \underline{\frac{1}{n}\sum_{i=1}^n Y_i}$, $S^2 = S_Y^2 = \underline{\frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y})^2}$
1. Estimator of μ is $\overline{X}$ where $\overline{X} =$ ______________, Estimator of σ^2 is S^2, where $S^2 =$ ______________, **key:** $\underline{\sum_i X_i/n}$, $\underline{\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2}$.

17. For a large sample test for H_o: $\theta = \theta_o$, a test statistic is $Z =$ ____________, a RR is Z________ if $\theta > \theta_o$; and a RR is ______________ if $\theta \neq \theta_o$; **key:** $\underline{\frac{\hat{\theta}-\theta_o}{\hat{\sigma}_{\hat{\theta}}}}$, $\underline{> z_\alpha}$, $\underline{|Z| > z_{\alpha/2}}$,

18. Sample size for an upper-tail α-level test is $n = ($______________$)^2$ **key:** $\underline{\frac{(z_\alpha+z_\beta)\sigma}{\mu_a-\mu_o}}$,

20. Suppose that $Z \sim N(0,1)$, $X \sim \chi^2(u)$, $Y \sim \chi^2(v)$. If Z____X, $T =$ ____________, then $T \sim t_u$; If X____Y, $F =$ ____, then $F \sim F_{u,v}$ and $X + Y \sim$ ____________. **key:** $\underline{\perp}$, $\underline{Z/\sqrt{X/u}}$, $\underline{\perp}$, $\underline{\frac{X/u}{Y/v}}$, $\underline{\chi^2(u+v)}$,

21. The MP test for H_o: $\theta = \theta_o$ v.s. H_a: $\theta = \theta_a$, the MP test has the RR satisfying: $\frac{L(\theta_o)}{L(\theta_a)}$____$k$ and $P_\theta(RR) = \alpha$ if $\theta =$ ____. **key:** $\underline{\leq}$, $\underline{\theta_o}$,

44. If X_1___X_2.

X_i's $\sim$:	$X_1 + X_2 \sim$:	key: $\underline{\perp}$,
$\mathcal{G}(\alpha_i, \beta)$	________	$\underline{\mathcal{G}(\alpha_1+\alpha_2, \beta)}$
$\chi^2(v_i)$	________	$\underline{\chi^2(v_1+v_2)}$
$Pois(\lambda_i)$	________	$\underline{Pois(\lambda_1+\lambda_2)}$
$N(\mu_i, \sigma_i^2)$	________	$\underline{N(\mu_x+\mu_y, \sigma_1^2+\sigma_2^2)}$
$bin(n_i, p)$	________	$\underline{bin(n_1+n_2, p)}$

Chapter 16. Introduction to Bayesian Methods for Inference

16.1. Under the assumption that X_1, ..., X_n are i.i.d. from $f(x;\theta)$, $\theta \in \Theta$, where θ is an unknown parameter, constant (not random),
we have learned 3 methods to estimate unknown parameter, say θ:

unbiased estimator,

MME,

MLE.

In this section, we study a new estimator: Bayes estimator, under
the 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),

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

Homework 16.1.1. 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. from f_X ? Prove or disprove it through the assumption as follows.

$f(x|\theta)$ is the density of $bin(1,p)$, and $p \sim U(0,1)$.

Remark 16.1. Two ways to compute the Bayes estimator:

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

They lead to the same estimator.

The second method is often simpler in derivation.

Example 16.1. Let X_1, ..., X_n be a random sample from $bin(k,\theta)$,
$\theta \sim beta(\alpha,\beta)$ with $\pi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$, $\theta \in [0,1]$, where (k,α,β) is known.
Bayes estimator of θ ?

25. $X \sim beta(\alpha,\beta)$. $\underline{f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}}$, if $\underline{x \in (0,1)}$, $\mu = \underline{\frac{\alpha}{\alpha+\beta}}$, where $B(\alpha,\beta) = \underline{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}}$

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

Method 1. Based on $\mathbf{X}$.

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

$$\begin{aligned}
\pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x},\theta)}{f_{\mathbf{X}}(\mathbf{x})} \\
&= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{f_{\mathbf{X}}(\mathbf{x})} \\
&\propto f(\mathbf{x}|\theta)\pi(\theta) && \text{as } f_X \text{ does not depend on } \theta \\
&= \prod_{i=1}^n \binom{k}{x_i}\theta^{\sum_i x_i}(1-\theta)^{nk-\sum_i x_i}\frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} \\
&\propto \theta^{\sum_i x_i}(1-\theta)^{nk-\sum_i x_i}\theta^{\alpha-1}(1-\theta)^{\beta-1}\textbf{(main trick!!)} \\
&= \theta^{\sum_i x_i+\alpha-1}(1-\theta)^{kn-\sum_i x_i+\beta-1} && (1)
\end{aligned}$$

$$\begin{aligned}
\text{Thus } \theta|(\mathbf{X}=\mathbf{x}) &\sim beta(\sum_i x_i+\alpha, nk-\sum_i x_i+\beta) && (2) \\
\theta|(\mathbf{X}=\mathbf{x}) &\sim beta(a,b).
\end{aligned}$$

Q: What is the meaning of Eq. (2) if $n=0$?
The Bayes estimator is

$$\begin{aligned}
\hat{\theta} &= E(\theta|\mathbf{X}) \\
&= \frac{a}{a+b} \quad \textbf{why}? \\
&= \frac{\sum_i X_i+\alpha}{nk+\alpha+\beta} \\
&= \frac{1}{nk+\alpha+\beta}\frac{nk\sum_i X_i}{nk} + \frac{1}{nk+\alpha+\beta}(\alpha+\beta)\frac{\alpha}{\alpha+\beta} \\
&= \frac{nk}{nk+\alpha+\beta}\frac{\sum_i X_i}{nk} + \frac{\alpha+\beta}{nk+\alpha+\beta}\frac{\alpha}{\alpha+\beta} \\
&= r\frac{\sum_{i=1}^n X_i}{nk} + (1-r)\frac{\alpha}{\alpha+\beta} \qquad \approx \begin{cases} MLE & \text{if } r\approx 1 \text{ or } n\approx\infty \\ E(\theta) & \text{if } r\approx 0 \text{ or } n=0, \end{cases} && (3)
\end{aligned}$$

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 the sufficient statistic $T = \sum_i X_i$.

$T|\theta \sim bin(nk, \theta)$? Yes, No, DNK

or $T \sim bin(nk, \theta)$? Yes, No, DNK

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

Additional HW:

448 [22] The Likelihood ratio test for H_o: $\theta \in \Theta_o$ v.s. H_a: $\theta \notin \Theta_o$ has a RR: $\{\lambda \_\_\_\_ k\}$, where

$\lambda =$ __________; $\hat{\theta}_o$ is the MLE under ____; $\hat{\theta}$ is the MLE under ____;

k satisfies $max\{P(RR):\ \theta \in \Theta_o\} =$ ____;

if n is large, then $-2\ln\lambda$ is approximated _______;

where $v =$ ___________; r and $r_o = \#$ of free parameters in Θ and in Θ_o, respectively.

key: $\underline{\leq}$, $\underline{\frac{L(\hat{\theta}_o)}{L(\hat{\theta})}}$, $\underline{\Theta_o}$, $\underline{\Theta}$, $\underline{\alpha}$, $\underline{\chi^2(v)}$,

24(c).

$$\begin{aligned}
\lambda = \frac{L(\hat{\theta}_o)}{L(\hat{\theta})} &= \frac{\left(\frac{\sum_i (x_i+y_i)}{2n}\right)^{n\overline{x}+n\overline{y}}}{\left(\frac{\sum_i x_i}{n}\right)^{n\overline{x}}\left(\frac{\sum_i y_i}{n}\right)^{n\overline{y}}} \qquad \overline{x} = 20\ \&\ \overline{y} = 22 \\
&= \frac{21^{100(20+22)}}{20^{100(20)}22^{100(22)}} \leq k =? \qquad by\ 448[22] \\
-2\ln\lambda &\sim \chi^2(2-1) \\
-2\ln\lambda &= 9.53
\end{aligned}$$

H_o: $\theta_1 = \theta_2$ v.s.

H_1: $\theta_1 \neq \theta_2$

Test statistic: λ or $-2\ln\lambda$.

RR: $\lambda \leq k$? Yes, No, DNK

$-2\ln\lambda \leq \chi^2_{0.05,1}$? Yes, No, DNK

$-2\ln\lambda \geq \chi^2_{0.05,1}$? Yes, No, DNK

Remark 16.2. The tricks $\propto$ are only applied to typical density functions. It does not apply non-standard cases as follows.

Example 16.2. Suppose that $X|\theta \sim bin(2,\theta)$ and θ has prior $\pi(p) = p$, $p \in \{0.2, 0.8\}$. Find the Bayes estimator of θ.

Sol. The Bayes estimator is $E(\theta|X)$. In particular

$\hat\theta = E(\theta|X = x) = \sum_\theta \theta\pi(\theta|x)$

$\quad = 0.2\pi(0.2|x) + 0.8\pi(0.8|x)$, $x \in \{0, 1, 2\}$

Need to find out $\pi(\cdot|x)$ $(=\frac{f(x,\theta)}{f_X(x)})$, given $\begin{cases} f(x|\theta) = \binom{2}{x}\theta^x(1-\theta)^{2-x}, \ x \in \{0,1,2\} \\ \pi(\theta) = \theta, \ \theta \in \{0.2, 0.8\} \end{cases}$

$$\begin{aligned}
f(x,\theta) =& \binom{2}{x}\theta^x(1-\theta)^{2-x}\theta && = f(x|\theta)\pi(\theta) \\
=& \binom{2}{x}\theta^{x+1}(1-\theta)^{2-x}, && x = \textbf{??} \ \ \theta = \textbf{??} \\
f_X(x) =& \sum_\theta f(x,\theta) \\
=& f(x, 0.2) + f(x, 0.8) \\
=& \binom{2}{x}0.2^{x+1}(1-0.2)^{2-x} + \binom{2}{x}0.8^{x+1}(1-0.8)^{2-x}, \ x \in \{0,1,2\}. \\
\pi(\theta|x) =& f(x,\theta)/f_X(x) \\
=& \frac{\binom{2}{x}\theta^{x+1}(1-\theta)^{2-x}}{\binom{2}{x}0.2^{x+1}0.8^{2-x} + \binom{2}{x}0.8^{x+1}0.2^{2-x}} && f_X \text{ is needed !} \\
=& \frac{\theta^{x+1}(1-\theta)^{2-x}}{0.2^{x+1}0.8^{2-x} + 0.8^{x+1}0.2^{2-x}}, \ x \in \{0,1,2\}, \ \theta \in \{0.2, 0.8\}. \\
\hat\theta =& 0.2\pi(0.2|x) + 0.8\pi(0.8|x) \\
=& 0.2 * \frac{0.2^{x+1}(1-0.2)^{2-x}}{0.2^{x+1}0.8^{2-x} + \binom{2}{x}0.8^{x+1}0.2^{2-x}} + 0.8 * \frac{0.8^{x+1}(1-0.8)^{2-x}}{0.2^{x+1}0.8^{2-x} + 0.8^{x+1}0.2^{2-x}} \\
=& \frac{0.2^{x+2}(0.8)^{2-x}}{0.2^{x+1}0.8^{2-x} + 0.8^{x+1}0.2^{2-x}} + \frac{0.8^{x+2}(0.2)^{2-x}}{0.2^{x+1}0.8^{2-x} + 0.8^{x+1}0.2^{2-x}} \\
=& \begin{cases} \frac{0.2^2(0.8)^2}{0.2^1 0.8^2 + 0.8^1 0.2^2} + \frac{(0.8)^2(0.2)^2}{0.2^1 0.8^2 + 0.8^1 0.2^2} & \text{if } x = 0 \\ \frac{2*0.2^3(0.8)^1}{2*0.2^2 0.8^1 + 2*0.8^2 0.2^1} + \frac{2*0.8^3(0.2)^1}{2*0.2^2 0.8^1 + 2*0.8^2 0.2^1} & \text{if } x = 1 \\ \frac{0.2^4}{0.2^3 + 0.8^3} + \frac{0.8^4}{0.2^3 + 0.8^3} & \text{if } x = 2 \end{cases} \\
=& \begin{cases} 2(0.8)(0.2) & \text{if } x = 0 \\ 0.2^2 + 0.8^2 & \text{if } x = 1 \\ \frac{0.2^4 + 0.8^4}{0.2^3 + 0.8^3} & \text{if } x = 2 \end{cases} \\
\hat\theta = E(\theta|X = x) =& \begin{cases} 0.32 & \text{if } x = 0 \\ 0.68 & \text{if } x = 1 \\ 0.192782 & \text{if } x = 2 \end{cases}
\end{aligned}$$

448 [24] Under the Bayes model, conditional on θ, Y_1, ..., Y_n are i.i.d. with $f(y|\theta)$, and $\theta \sim g(\theta)$. The posterior df is $g(\theta|\underline{y}) = \underline{\frac{\prod_{i=1}^{n} f(y_i|\theta)g(\theta)}{f_{\underline{Y}}(\underline{y})}}$, where $\underline{y} = (y_1, ..., y_n)$, the Bayes estimator of $h(\theta)$ is $\hat{h} = \underline{E(h(\theta)|\underline{y})}$,

Example 16.3. Suppose that X_1, ..., X_n are 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. In [24], $h(\theta)=\theta$ in Ex. 16.3 here. Let $\mathbf{X}=(X_1,..,X_n)$.

447 [22]. $X\sim N(\mu,\sigma^2)$. $f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $\underline{\frac{X-\mu}{\sigma}\sim N(0,1)}$

Two ways: (1) $E(\theta|\mathbf{X})$ and (2) $E(\theta|T(\mathbf{X}))$, where $T(\mathbf{X})$ is a sufficient statistic.

Which to choose ?

A sufficient statistic is $Y=\sum_{i=1}^n X_i$. $Y|\theta\sim N(n\theta,n\sigma^2)$.

Another sufficient statistic is $T=\overline{X}$. $T|\theta\sim N(\theta,\sigma^2/n)$. **Which is more convenient ?**

$$E(\theta|T=t)=\int\theta\,\underbrace{\pi(\theta|t)}d\theta. \qquad \text{Method (2)}$$

$$\begin{aligned}
\pi(\theta|t)&=\frac{f(t,\theta)}{f_T(t)}=\frac{f(t|\theta)\pi(\theta)}{f_T(t)}=\ \textbf{??}\\
&\propto f(t|\theta)\pi(\theta) \qquad \textbf{(main trick)}\\
&\propto \exp\big(-\frac{1}{2}\frac{(t-\theta)^2}{\sigma^2/n}\big)\exp\big(-\frac{1}{2}\frac{(\theta-\mu)^2}{\tau^2}\big)\\
&=\exp\big(-\frac{1}{2}\frac{t^2-2t\theta+\theta^2}{\sigma^2/n}-\frac{1}{2}\frac{\theta^2-2\theta\mu+\mu^2}{\tau^2}\big)\\
&\propto \exp\big(-\frac{1}{2}\frac{-2t\theta+\theta^2}{\sigma^2/n}-\frac{1}{2}\frac{\theta^2-2\theta\mu}{\tau^2}\big)\\
&=\exp\big(-\frac{1}{2}\frac{\theta^2}{\sigma^2/n}-\frac{1}{2}\frac{\theta^2}{\tau^2}+\frac{1}{2}\frac{2t\theta}{\sigma^2/n}+\frac{1}{2}\frac{2\theta\mu}{\tau^2}\big) \quad =e^{-a\theta^2+b\theta} \qquad (4)\\
&=\exp\big(-\frac{1}{2}\{\underbrace{[\frac{1}{\sigma^2/n}+\frac{1}{\tau^2}]}_{\frac{1}{\sigma_*^2}}\theta^2-2\underbrace{[\frac{t}{\sigma^2/n}+\frac{\mu}{\tau^2}]}_{\frac{\mu_*}{\sigma_*^2}}\theta\}\big)\propto e^{-(\frac{\theta-\mu_*}{\sigma_*})^2/2}
\end{aligned}$$

$$f_Z(x)=\frac{1}{\sqrt{2\pi\sigma_*^2}}exp(-(\frac{x-\mu_*}{\sigma_*})^2/2)=\frac{1}{\sqrt{2\pi\sigma_*^2}}exp(-\frac{1}{2}[\frac{x^2}{\sigma_*^2}-2\frac{\mu_*}{\sigma_*^2}x+\frac{\mu_*^2}{\sigma_*^2})\propto e^{-ax^2+bx}$$

(kernel of f_Z for $N(\mu_*,\sigma_*^2)$).

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

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

σ_*^2= **? ??** and μ_*=**? ??**

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

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

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

Method (1) $E(\theta|\mathbf{X}=\mathbf{x}) = \displaystyle\int \theta \underbrace{\pi(\theta|\mathbf{x})} d\theta.$

$$
\begin{aligned}
\pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{f_{\mathbf{X}}(\mathbf{x})} = \textbf{??} \\
&\propto f(\mathbf{x}|\theta)\pi(\theta) && \textbf{(main trick)} \\
&\propto \exp\big(-\frac{1}{2}\sum_i \frac{(x_i-\theta)^2}{\sigma^2}\big)\exp\big(-\frac{1}{2}\frac{(\theta-\mu)^2}{\tau^2}\big) \\
&= \exp\big(-\frac{1}{2}\sum_i \frac{(x_i-\theta)^2}{\sigma^2} - \frac{1}{2}\frac{(\theta-\mu)^2}{\tau^2}\big) \\
&\propto \exp\big(-\frac{1}{2}\sum_i \frac{-2x_i\theta+\theta^2}{\sigma^2} - \frac{1}{2}\frac{\theta^2-2\theta\mu}{\tau^2}\big) \\
&= \exp\big(\sum_i \frac{2x_i\theta-\theta^2}{2\sigma^2} - \frac{\theta^2-2\theta\mu}{2\tau^2}\big) \\
&= \exp\big(\sum_i \frac{x_i\theta}{\sigma^2} - \theta^2(\frac{n}{2\sigma^2}+\frac{1}{2\tau^2}) + \frac{\theta\mu}{\tau^2}\big) \\
&= \exp\big(\frac{\overline{x}\theta}{\sigma^2/n} - \theta^2(\frac{n}{2\sigma^2}+\frac{1}{2\tau^2}) + \frac{\theta\mu}{\tau^2}\big) && \text{same as Eq. (4).}
\end{aligned}
$$

Remark 16.3. It is interesting to notice the following fact again.

In Example 16.3, the Bayes estimator is

$$
\begin{aligned}
\hat{\theta} &= \frac{\frac{\overline{X}}{\sigma^2/n}+\frac{\mu}{\tau^2}}{\frac{1}{\sigma^2/n}+\frac{1}{\tau^2}} \\
&= \frac{\frac{1}{\sigma^2/n}}{\frac{1}{\sigma^2/n}+\frac{1}{\tau^2}}\overline{X} + \frac{\frac{1}{\tau^2}}{\frac{1}{\sigma^2/n}+\frac{1}{\tau^2}}\mu \\
&= r\overline{X} + (1-r)\mu \\
&\approx \begin{cases} \overline{X} & \text{if } n \text{ is large or } r \approx 1 \\ E(\theta) & \text{if } r \approx 0 \text{ or } n = 0 \end{cases} && \text{(see Eq. (4))}
\end{aligned}
$$

a weighted average of the MLE $\overline{X}$ and the prior mean $E(\theta)$.

Additional homework problems:

A1. **Homework 16.1.1.** 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. from f_X ? Prove or disprove it through the assumption as follows.

$f(x|p)$ is the density of $bin(1,p)$, and $p \sim U(0,1)$.

Sol. $\vdash$: X_i's are **not** i.i.d. from f_X.

It suffices to give a counter-example, say $n=2$ and $f_{X_1,X_2}(1,1) \neq f_{X_1}(1)f_{X_2}(1)$ as follows.

$$\begin{aligned}
f_{X_1}(1) = f_{X_2}(1) =& \int_0^1 f(x,p)dp & = \int f(x,p)dp \ ? \\
=& \int_0^1 f(x|p)\pi(p)dp & = \int f(x|p)\pi(p)dp? \\
=& \int_0^1 p^x(1-p)^{1-x}dp & = \int p^x(1-p)^{1-x}dp \ ? \\
=& \int_0^1 p^1(1-p)^{1-1}dp \\
=& \int_0^1 p^{2-1}(1-p)^{1-1}dp \\
=& B(2,1) \\
=& \frac{\Gamma(2)\Gamma(1)}{\Gamma(3)} \\
=& 1/2 \\
f_{X_1,X_2}(1,1)) =& f_{\mathbf{X}}(\mathbf{x}) = & \int_0^1 f(\mathbf{x}|p)\pi(p)dp \\
=& \int_0^1 \prod_{i=1}^2 \binom{1}{x_i} p^{x_i}(1-p)^{1-x_i}dp \\
=& \int_0^1 p^{\sum_{i=1}^2 x_i}(1-p)^{2-\sum_i x_i}dp \\
=& \int_0^1 p^2(1-p)^{2-2}dp \\
=& \int_0^1 p^{3-1}(1-p)^{1-1}dp \\
=& B(3,1) = \frac{\Gamma(3)\Gamma(1)}{\Gamma(4)} \\
=& \frac{2}{3*2} \\
\neq& (\frac{1}{2})^2 \\
=& f_{X_1}(1)f_{X_2}(1)
\end{aligned}$$

A2. 448 [22] The Likelihood ratio test for H_o: $\theta \in \Theta_o$ v.s. H_a: $\theta \notin \Theta_o$ has a RR: $\{\lambda \underline{\leq k}\}$, where $\lambda = \underline{\frac{L(\hat{\theta}_o)}{L(\hat{\theta})}}$, $\hat{\theta}_o$ is the MLE under $\underline{\Theta_o}$, $\hat{\theta}$ is the MLE under $\underline{\Theta}$, k satisfies $max\{P(RR): \ \theta \in \Theta_o\} = \underline{\alpha}$;
if n is large, then $-2\ln\lambda$ is approximated $\underline{\chi^2(v)}$, where $v = \underline{r - r_o}$;
r and $r_o = \#$ of free parameters in Θ and in Θ_o, respectively.

24(c).

$$\lambda = \frac{L(\hat{\theta}_o)}{L(\hat{\theta})} = \frac{(\frac{\sum_i (x_i+y_i)}{2n})^{n\overline{x}+n\overline{y}}}{(\frac{\sum_i x_i}{n})^{n\overline{x}}(\frac{\sum_i y_i}{n})^{n\overline{y}}} \qquad \overline{x} = 20 \ \& \ \overline{y} = 22$$

$$= \frac{21^{100(20+22)}}{20^{100(20)}22^{100(22)}} \leq k = ? \qquad by \ 448[22]$$

$$-2\ln\lambda \sim \chi^2(2-1)$$

$$-2\ln\lambda = 9.53$$

H_o: $\theta_1 = \theta_2$ v.s.
H_1: $\theta_1 \neq \theta_2$
Test statistic: λ or $-2\ln\lambda$.
RR: $\lambda \leq k$? Yes, No, DNK — by 448.[22] Yes.
$-2\ln\lambda \leq \chi^2_{0.1,1}$? Yes, No, DNK — No, as $\lambda \leq k$
$-2\ln\lambda \geq \chi^2_{0.1,1}$? Yes, No, DNK — Yes.

Example 16.1(c). Let X_1, ..., X_n be a random sample from $bin(k,\theta)$,
$\theta \sim beta(\alpha,\beta)$ with $\pi(\theta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$, $\theta \in [0,1]$, where $(n,k,\alpha,\beta) = (10,1,3,4)$.
Let $\hat\theta$ be the Bayes estimator of θ and $\check\theta$ the MLE. Find the MSE of $\hat\theta$ and $\check\theta$.

447[25] $X \sim beta(\alpha,\beta)$. $f(x) = \underline{\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}}$, if $x \in \underline{(0,1)}$, $\mu = \underline{\frac{\alpha}{\alpha+\beta}}$, where $B(\alpha,\beta) = \underline{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}}$

448 [24] Under the Bayes model, conditional on θ, Y_1, ..., Y_n are i.i.d. with $f(y|\theta)$, and $\theta \sim g(\theta)$. The posterior df is $g(\theta|\underline{y})$ = ____________________, where $\underline{y} = (y_1,...,y_n)$ the Bayes estimator of $h(\theta)$ is $\hat h$ = ______________

key: $\underline{\frac{\prod_{i=1}^n f(y_i|\theta)g(\theta)}{f_{\underline{Y}}(\underline{y})}}$, $\underline{E(h(\theta)|\underline{y})}$,

Sol. Notice that $E(h(\theta)|y) = E(\theta|\mathbf{X})$ in Example 1, where $h(\theta) = \theta$ and $\mathbf{X} = (X_1,...,X_n)$.
The Bayes estimator is $\hat\theta = \frac{\sum_{i=1}^n X_i+\alpha}{n+\alpha+\beta}$.
Recall that the MLE is $\check\theta = \overline{X}$.

$$\begin{aligned}
\text{If } \theta \text{ is not random, then } MSE(\check\theta) =& E((\check\theta-\theta)^2) \qquad\qquad = V(\check\theta) + (B(\check\theta))^2\\
=& E((\overline{X}-\theta)^2)\\
=& V(\overline{X})\\
=& V(X)/n\\
=& pq/n\\
=& \theta(1-\theta)/n.
\end{aligned}$$

However, if θ is random, then

$$\begin{aligned}
MSE(\check\theta) =& E((\check\theta-\theta)^2) \qquad\qquad = V(\check\theta) + (B(\check\theta))^2\\
=& E(E((\check\theta-\theta)^2|\theta))\\
=& E(V(\check\theta|\theta)) \qquad\qquad E(\check\theta|\theta) = E(\overline{X}|\theta) = \theta\\
=& E(V(\overline{X}|\theta))\\
=& E(\theta(1-\theta)/n)\\
=& \frac{1}{n}E(\theta-\theta^2)\\
=& \frac{1}{n}[\frac{\alpha}{\alpha+\beta} - E(\theta^2)]\\
E(\theta^2) =& \int_0^1 x^2\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}dx\\
=& \int_0^1 \frac{x^{\alpha+2-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}dx\\
=& \frac{B(\alpha+2,\beta)}{B(\alpha,\beta)}\int_0^1 \frac{x^{\alpha+2-1}(1-x)^{\beta-1}}{B(\alpha+2,\beta)}dx\\
=& \frac{\Gamma(\alpha+2)\Gamma(\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha+2+\beta)\Gamma(\alpha)\Gamma(\beta)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\Gamma(\alpha+2)\Gamma(\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha+2+\beta)\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{(\alpha+1)\alpha\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta)}{(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}
\end{aligned}
$$

$$
\begin{aligned}
MSE(\check{\theta}) &= \frac{1}{n}[\frac{\alpha}{\alpha+\beta} - E(\theta^2)] \\
&= \frac{1}{n}[\frac{\alpha}{\alpha+\beta} - \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)}] \\
&= \frac{1}{n}[\frac{\alpha}{\alpha+\beta}(1 - \frac{(\alpha+1)}{(\alpha+\beta+1)}] \\
&= \frac{1}{n}[\frac{\alpha}{\alpha+\beta}\frac{\beta}{\alpha+\beta+1}] \\
&= \frac{1}{10}\frac{12}{7*8} = \frac{3}{140}
\end{aligned}
$$

$$
\begin{aligned}
MSE(\hat{\theta}) &= E((\hat{\theta}-\theta)^2) \\
&= E((\frac{\sum_{i=1}^n X_i + \alpha}{n+\alpha+\beta} - \theta)^2) \\
&= E(E((\hat{\theta}-\theta)^2|\theta)) \\
&= E(V(\hat{\theta}|\theta) + (B(\hat{\theta}|\theta))^2) \\
&= E(V(\frac{\sum_{i=1}^n X_i + \alpha}{n+\alpha+\beta}|\theta)) + E((E(\frac{\sum_{i=1}^n X_i + \alpha}{n+\alpha+\beta}|\theta) - \theta)^2) \\
&= \frac{E(V(\sum_{i=1}^n X_i|\theta))}{(n+\alpha+\beta)^2} + E((\frac{n\theta+\alpha}{n+\alpha+\beta} - \theta)^2) \\
&= \frac{E(n\theta(1-\theta))}{(n+\alpha+\beta)^2} + E((\frac{\alpha - \theta(\alpha+\beta)}{n+\alpha+\beta})^2) \\
&= \frac{E(n\theta(1-\theta))}{(n+\alpha+\beta)^2} + E((\frac{\alpha}{\alpha+\beta} - \theta)^2)(\frac{\alpha+\beta}{n+\alpha+\beta})^2 \\
&= \frac{E(n(\theta-\theta^2))}{(n+\alpha+\beta)^2} + V(\theta)(\frac{\alpha+\beta}{n+\alpha+\beta})^2 \\
&= \frac{nE(\theta) - nE(\theta^2) + (\alpha+\beta)^2E(\theta^2) - (\alpha+\beta)^2(E(\theta))^2}{(n+\alpha+\beta)^2} \\
&= \frac{nE(\theta) + ((\alpha+\beta)^2 - n)E(\theta^2) - (\alpha+\beta)^2(E(\theta))^2}{(n+\alpha+\beta)^2}
\end{aligned}
$$

$$
E(\theta) = \frac{\alpha}{\alpha+\beta} = 3/7, \ and \ \ E(\theta^2) = \frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)} = \frac{3*4}{7*8} = 3/14
$$

$$
MSE = \frac{10*3/7 + 39*3/14 - 49*(3/7)^2}{17^2}
$$

A 2nd way.

$$
\begin{aligned}
\hat{\theta} =& \frac{n}{n+\alpha+\beta}\frac{\sum_i X_i}{n} + \frac{\alpha+\beta}{n+\alpha+\beta}\frac{\alpha}{\alpha+\beta} \\
=& r\frac{\sum_{i=1}^n X_i}{n} + (1-r)\frac{\alpha}{\alpha+\beta} \\
MSE(\hat{\theta}) = & E((\hat{\theta}-\theta)^2) \\
=& E((r\frac{\sum_{i=1}^n X_i}{n} + (1-r)\frac{\alpha}{\alpha+\beta} - \theta)^2) \\
=& E((r(\frac{\sum_{i=1}^n X_i}{n} - \theta) + (1-r)(\frac{\alpha}{\alpha+\beta} - \theta))^2) \\
=& r^2 E((\frac{\sum_{i=1}^n X_i}{n} - \theta)^2) + (1-r)^2 E((\frac{\alpha}{\alpha+\beta} - \theta)^2) \\
& + 2r(1-r)E((\frac{\sum_{i=1}^n X_i}{n} - \theta)(\frac{\alpha}{\alpha+\beta} - \theta)) \\
=& r^2 E((\frac{\sum_{i=1}^n X_i}{n} - \theta)^2) + (1-r)^2 V(\theta) \\
& + 2r(1-r)E(E((\frac{\sum_{i=1}^n X_i}{n} - \theta)(\frac{\alpha}{\alpha+\beta} - \theta)|\theta)) \\
=& r^2 E((\frac{\sum_{i=1}^n X_i}{n} - \theta)^2) + (1-r)^2 V(\theta) \\
& + 2r(1-r)E((\frac{\alpha}{\alpha+\beta} - \theta)E((\frac{\sum_{i=1}^n X_i}{n} - \theta)|\theta)) \\
=& r^2 E((\frac{\sum_{i=1}^n X_i}{n} - \theta)^2) + (1-r)^2 V(\theta) \\
=& r^2 \frac{1}{n}\frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)} + (1-r)^2(\frac{(\alpha+1)\alpha}{(\alpha+\beta+1)(\alpha+\beta)} - (\frac{\alpha}{\alpha+\beta})^2) \\
=& \cdots
\end{aligned}
$$

Chapter 11. Linear Regression.

11.1. Introduction. So far, we assume that Y_1, ..., Y_n are i.i.d. random variable from some distribution, *e.g.*,

Y_1, ..., Y_n are i.i.d. from $N(\mu, \sigma^2)$,

Y_1, ..., Y_n are i.i.d. from binomial.

In this chapter we assume the case that

Y_1, ..., Y_n are independent, but not identically distributed, and x_1, ..., x_n are constant.

e.g.,

1. $Y_i = \beta_o + \beta_1 x_i + \epsilon_i$, where ϵ_1, ..., ϵ_n are i.i.d. from $N(0, \sigma^2)$;
2. $Y_i = \beta_o + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$, where ϵ_1, ..., ϵ_n are i.i.d. from $N(0, \sigma^2)$;
3. $Y_i = \beta_o + \beta_1 x_i + \cdots + \beta_p x_i^p + \epsilon_i$, where ϵ_1, ..., ϵ_n are i.i.d. from $N(0, \sigma^2)$;

Def. (x_i, Y_i)'s satisfy a simply linear regression model, if they satisfy $E(Y_i|x_i) = \beta_o + \beta_1 x_i$.
In Case 2 and 3, we say that (x_i, Y_i) satisfies a linear regression model of degree p.

In case 1, $y = \beta_o + \beta_1 x$ is a straight line, (x_i, Y_i)'s are points around the straight line.

In case 2, $y = \beta_o + \beta_1 x + \beta_2 x^2$ is a quadratic curve,

(x_i, Y_i)'s are points around the curve $y = \beta_o + \beta_1 x + \beta_2 x^2$.

In case 3, $y = \beta_o + \beta_1 x + \cdots + \beta_p x^p$ is a polynomial curve,

(x_i, Y_i)'s are points around the curve $y = \beta_o + \beta_1 x + \cdots + \beta_p x^p$.

Question: Difference between

linear regression and elementary statistics?

§**1. In elementary statistics**.

We have a random sample of observations:

$Y_1, ..., Y_n \overset{iid}{\sim} F$, the cumulative distribution function (cdf),

with the density function f.

$f(y) = f_o(y; \theta)$, $\theta \in \Theta$, the parameter space, and

f_o is given except for θ,

e.g., $f(y; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$, $\theta = (\mu, \sigma)$.

There are three statistical inferences:

(1) $\theta = ?$ or $E(Y) = ?$ or $F = ?$

(2) $\theta = \underbrace{\theta_o}_{given}$?

(3) $(L, R) = ?$ such that θ is likely in (L, R)

In (1), we try to estimate θ by $\hat{\theta}$ through

the maximum likelihood estimator (MLE)

or other estimators,

or to estimate $E(Y_1)$ or $F(y)$, say $\mu(\theta)$ or $F(y; \theta)$, by

$\hat{\mu} = \overline{Y}$ (or $\mu(\hat{\theta})$)

or by $F(y; \hat{\theta})$

In (2) we try to test hypotheses H_o: $\theta = \theta_o$.

In (3) we try to construct a confidence interval (CI).

§2. Usual task for linear regression analysis:

1. Assume a proper probability model, that is, a linear regression model
   $Y_i = \alpha + \beta_1 X_i + \cdots + \beta_p X_i^p + \epsilon_i$, $\epsilon_i \sim N(0, \sigma^2)$.
2. Carry out statistical analysis.

(1) $(\alpha, \beta) = ?$

(2) $(\alpha, \beta) = (a, b)$? (where (a, b) is given).

(3) (L, R)= ? for β (or for α) so that β is likely in (L, R).

§11.2. The linear regression model

Data: (X_i, Y_i), $i = 1$, ..., n.

$X_i \in \mathcal{R}^p$ — covariate vector,

$Y_i \in \mathcal{R}^1$ —- response variable.

For example, in **whiteside** data,

Y = gas consumption (in 1000 cubic feet units),

X = (external temperature, before or after insulation wall was installed)$'$.

Our main interest is Y, e.g., we want to know the distribution of Y, the gas assumption. It is certainly helpful if we know the external temperature and other information.

Thus we look for a relation between X and Y, say $Y = g(X)$. One can guess a function g by plotting the data. See the plot of (temp, gas) (and its program in ch2) and another plot of (X, Y).

The first plotting suggests a linear relation

The second suggests a quadratic or exponential curve.

Both the linear curve and quadratic curves can be represented by linear regression models.

$$E(Y|temp) = \beta_0 + \beta_1 temp \text{ or } E(Y|X) = \beta_0 + \beta_1 X$$

and

$$E(Y|X) = \beta_0 + \beta_1 X + \beta_2 X^2.$$

Possible model assumptions:

1. (X_1, Y_1), ..., (X_n, Y_n) are observations,
2. $Y_i = \alpha + \beta' X_i + \epsilon_i$, $i = 1$, ..., n, where β, $X_i \in \mathcal{R}^p$,
3. ϵ_is are uncorrelated, with mean 0 and variance σ^2
4. ϵ_is are i.i.d. from ϵ,
5. $\epsilon \sim N(0, \sigma^2)$.

6. X_is are constant (not random variables).
Since $Y_i = (\alpha, \beta') \begin{pmatrix} 1 \\ X_i \end{pmatrix} + \epsilon_i$
or $Y_i = \theta' Z_i + \epsilon_i$,
abusing notations, one can simply replace assumption 2 by
2^*. $Y_i = \beta' X_i + \epsilon_i$, where β, $X_i \in \mathcal{R}^p$.

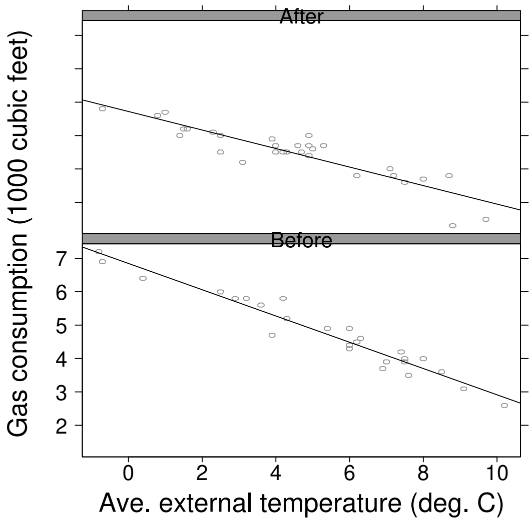

Figure 2.1. Plotting of whiteside data

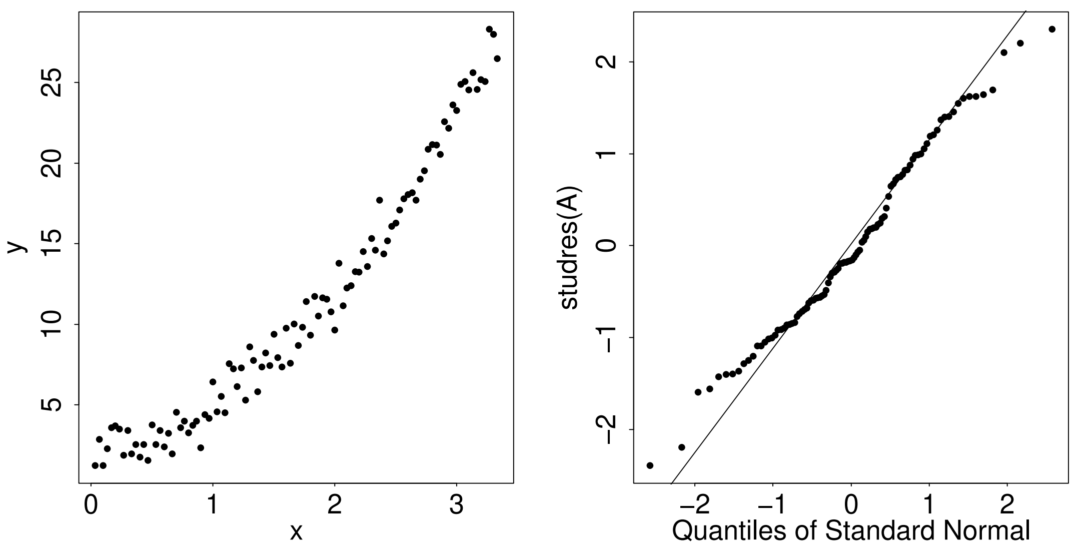

Figure 2.2. Plotting of (x,y)

Least square estimator

So far, we have learnt

maximum likelihood estimator (MLE),

method of moment estimator (MME),

Unbiased estimator,

Bayes estimator.

In linear regression, the most common estimator is the least square estimator (LSE),

$$\hat{\gamma} = argmin_{\gamma \in \mathcal{R}^p} \sum_{i=1}^{n} (Y_i - \gamma' X_i)^2 \text{ under the model } Y_i = \gamma' X_i + \epsilon_i,$$

where γ and X_i are $p \times 1$ vectors, and (Y_i, X_i)'s are observations, but not ϵ_i's.

Some simple cases:

(1) $Y_i = \alpha + \epsilon_i$. Then $X_i = 1$ and $\gamma = \alpha$;

(2) $Y_i = \beta x_i + \epsilon_i$. Then $X_i = x_i$ and $\gamma = \beta$;

(3) $Y_i = \alpha + \beta x_i + \epsilon_i$. Then $X_i = (1, x_i)'$ and $\gamma = (\alpha, \beta)'$.

(4) $Y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$. Then $X_i = (1, x_i, x_i^2)'$ and $\gamma = (\alpha, \beta_1, \beta_2)'$.

Derivation of the LSE:

Case (1) $Y_i = \alpha + \epsilon_i$;

$L = \sum_i (Y_i - \alpha)^2$, called the loss function,

$L' = -\sum_i 2(Y_i - \alpha) = 0$.

$L'' = -\sum_i 2(-1) > 0$

=> $\hat{a} = \overline{Y}$ is the minimum point.

Thus the LSE of α is $\hat{\alpha} = \overline{Y}$.

Remark. The MLE maximizes the likelihood L and the LSE minimizes the loss L.

Case (2) $Y_i = \beta x_i + \epsilon_i$;

$L = \sum_i (Y_i - \beta x_i)^2$,

$L' = -\sum_i 2x_i(Y_i - \beta x_i) = 0$

=> $\sum_i x_i Y_i = \sum_i \beta x_i^2$

=> $\hat{\beta} = \sum_i x_i Y_i / \sum_i x_i^2$.

$L'' = -\sum_i 2(-x_i^2) > 0$

Thus the LSE of β is $\hat{\beta} = \overline{xY}/\overline{x^2}$.

Case (3) $Y_i = \alpha + \beta x_i + \epsilon_i$.

$L = \sum_i (Y_i - \alpha - \beta x_i)^2$,

$\frac{dL}{d\alpha} = -\sum_i 2(Y_i - \alpha - \beta x_i) = 0$

$=> \sum_i (Y_i - \beta x_i) = \sum_i \alpha$

$=> \hat\alpha = \overline{Y} - \beta \overline{x}$.

$L = \sum_i (Y_i - \overline{Y} + \beta \overline{x} - \beta x_i)^2$.

$\frac{\partial L}{\partial \beta} = -\sum_i 2x_i (Y_i - \overline{Y} + \beta \overline{x} - \beta x_i) = 0$

$=> \hat\beta = \sum_i x_i (Y_i - \overline{Y}) / \sum_i x_i (x_i - \overline{x})$.

Check: $\begin{array}{cccccc} (\alpha,\beta): & \alpha=-\infty & \alpha=\infty & \beta=-\infty & \beta=\infty & (\hat\alpha,\hat\beta) \\ L: & \infty & \infty & \infty & \infty & finite \end{array}$

Thus the LSE of β is $\hat\beta = \overline{xY}/\overline{x^2}$ and the LSE of α is $\hat\alpha = \overline{Y} - \hat\beta \overline{x}$.

The LSE in Case (3) can be solved by a matrix form as follows. Write

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \alpha + x_1\beta + \epsilon_1 \\ \vdots \\ \alpha + x_n\beta + \epsilon_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \quad or\ \mathbf{Y} = \mathbf{X}\gamma + \mathbf{e}$$

where $\mathbf{Y}' = (Y_1, ..., Y_n)$, $\mathbf{X}' = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix}$ and $\mathbf{e}' = (\epsilon_1, ..., \epsilon_n)$.

The solution is $\hat\gamma = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, provided that

$$\mathbf{X} \text{ is of full rank. } \textbf{What does it mean ?} \tag{1}$$

Ans: Rank of $\mathbf{X} = p$, in particular $p = 2$ in Case (3).

Notice that Rank of $\mathbf{X} \le \min\{n, p\}$. In Case (3),

$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix}$

$(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} \sum_i x_i^2 & -\sum_i x_i \\ -\sum_i x_i & n \end{pmatrix} / (n\sum_i x_i^2 - (\sum_i x_i)^2)$.

$(\begin{pmatrix} a & b \\ c & d \end{pmatrix})^{-1} = \frac{\begin{pmatrix} d & -c \\ -b & a \end{pmatrix}}{|\begin{pmatrix} a & b \\ c & d \end{pmatrix}|} \qquad (|A| = \text{determinant of } A)$

$\mathbf{X}'\mathbf{Y} = \begin{pmatrix} \sum_i Y_i \\ \sum_i x_i Y_i \end{pmatrix}$. $\begin{pmatrix} \hat\alpha \\ \hat\beta \end{pmatrix} = \frac{\begin{pmatrix} \sum_i x_i^2 \sum_j Y_j - \sum_i x_i \sum_j x_j Y_j \\ -\sum_i x_i \sum_j Y_j + n\sum_i x_i Y_i \end{pmatrix}}{n\sum_i x_i^2 - (\sum_i x_i)^2} = \begin{pmatrix} \frac{\overline{X^2}\overline{Y} - \overline{X}\cdot\overline{XY}}{\overline{X^2} - (\overline{X})^2} \\ \frac{\overline{XY} - \overline{X}\cdot\overline{Y}}{\overline{X^2} - (\overline{X})^2} \end{pmatrix}$

Example 11.1. Given data $\vec{x} = (-2, -1, 0, 1, 2)'$ and $\vec{y} = (0, 0, 1, 1, 3)'$. Find the LSE based

(a) $y_i = \alpha + \epsilon_i$; (b) $y_i = \beta x_i + \epsilon_i$; (c) $y_i = \alpha + \beta x_i + \epsilon_i$;

Sol. (a) the LSE of α is $\hat{\alpha} = \overline{y} = (0 + 0 + 1 + 1 + 3)/5 = 1$.

(b) The LSE of $\hat{\beta} = \overline{xy}/\overline{xx} = (1 + 2 * 3)/(4 + 1 + 0 + 1 + 4) = 7/10 = 0.7$

(c) Use the matrix form: $\hat{\gamma} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

$\mathbf{Y} = (0, 0, 1, 1, 3)'$,

$$\mathbf{X}' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{pmatrix},$$

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & -1 & 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$$

$$= \begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix},$$

$$(\mathbf{X}'\mathbf{X})^{-1} = ? \qquad (\begin{pmatrix} a & b \\ c & d \end{pmatrix})^{-1} = \frac{\begin{pmatrix} d & -c \\ -b & a \end{pmatrix}}{|\begin{pmatrix} a & b \\ c & d \end{pmatrix}|} \qquad (|A| = \text{determinant of } A)$$

$$= \begin{pmatrix} 0.2 & 0 \\ 0 & 0.1 \end{pmatrix}$$

$$\mathbf{X}'\mathbf{Y} = \begin{pmatrix} \sum_i Y_i \\ \sum_i x_i Y_i \end{pmatrix}.$$

$$= \begin{pmatrix} 5 \\ 7 \end{pmatrix},$$

$$\hat{\gamma} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

$$= \begin{pmatrix} 0.2 & 0 \\ 0 & 0.1 \end{pmatrix} \begin{pmatrix} 5 \\ 7 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 0.7 \end{pmatrix}.$$

448 [23] Suppose (x_i, Y_i)'s are independent, $\mu_{x_i}(\theta) = E(Y_i)$, the LSE of θ, denoted by $\hat{\theta}$, minimizes $\underline{\hspace{4cm}}$.

If $\mu_{x_i}(\theta) = \beta_0 + \beta_1 x_i$ and $V(Y_i) = \sigma^2$, the LSE $\hat{\beta}_0 = \underline{\hspace{3cm}}$,
$\hat{\beta}_1 = \underline{\hspace{3cm}}$, $S_{xy} = \underline{\hspace{3cm}}$, $E(\hat{\beta}_0) = \underline{\hspace{1.5cm}}$, $E(\hat{\beta}_1) = \underline{\hspace{1.5cm}}$,
$\sigma^2_{\hat{\beta}_0} = \underline{\hspace{3cm}}$, $\sigma^2_{\hat{\beta}_1} = \underline{\hspace{3cm}}$, an unbiased estimator of σ^2 is $\hat{\sigma}^2$=$\underline{\hspace{1.5cm}}$,

key: $\underline{\sum_{i=1}^n [Y_i - \mu_{x_i}(\theta)]^2}$, $\underline{\overline{Y} - \hat{\beta}_1\overline{x}}$, $\underline{S_{xy}/S_{xx}}$, $\underline{n(\overline{xY} - \overline{x}\overline{Y})}$, $\underline{\beta_0}$, $\underline{\beta_1}$,
$\underline{\sigma^2\overline{x^2}/S_{xx}}$, $\underline{\sigma^2/S_{xx}}$, $\underline{\frac{1}{n-2}(S_{yy} - \hat{\beta}_1 S_{xy})}$,

Remark. [23] is not for the model $Y = \alpha + \epsilon$ or $Y = \beta X + \epsilon$.

Properties of the LSE.

The linear regression model assumes

$$Y_i = X_i'\gamma + \epsilon_i$$

where ϵ_is are i.i.d. random variables with mean zero and variance σ^2.

The LSE satisfies

1. $E(\hat\gamma) = \gamma$ under assumptions 1, 2^*, 3 and 6,
2. $V(\hat\gamma) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ under assumptions 1, 2^*, 3, 4 and 6.
   Reason: Under assumptions 1, 2^*, 3 and 6,
   $E(\hat\gamma) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\gamma)$,
   $Cov(\hat\gamma) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Cov(\mathbf{e})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$,
   and $Cov(\mathbf{e}) = \sigma^2 I_n$ if assumption 4 also holds.
3. If $\epsilon_i \sim N(0,\sigma^2)$, then the LSE $\hat\gamma$ is the MLE.
   That is, it maximizes

$$\begin{aligned} L(\alpha,\beta) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(Y_i-\alpha-\beta x_i)^2}{2\sigma^2}) \\ &\propto \exp(-\frac{\sum_i (Y_i-\alpha-\beta x_i)^2}{2\sigma^2}) \\ <=> \text{ minimizes } &\frac{\sum_i (Y_i-\alpha-\beta x_i)^2}{2\sigma^2} \\ &\propto \sum_i (Y_i-\alpha-\beta x_i)^2 \end{aligned}$$

23. Suppose (x_i, Y_i)'s are independent, $\mu_{x_i}(\theta) = E(Y_i)$, the LSE of θ, denoted by $\hat\theta$, minimizes ________________.
    If $\mu_{x_i}(\theta) = \beta_0 + \beta_1 x_i$ and $V(Y_i) = \sigma^2$, the LSE $\hat\beta_0 =$ ____________,
    $\hat\beta_1 =$ ____________, $S_{xy} =$ ____________, $E(\hat\beta_0) =$ _____, $E(\hat\beta_1) =$ _____,
    $\sigma^2_{\hat\beta_0} =$ ____________, $\sigma^2_{\hat\beta_1} =$ ____________, an unbiased estimator of σ^2 is $\hat\sigma^2$=_____,
    key: $\underline{\sum_{i=1}^n [Y_i - \mu_{x_i}(\theta)]^2}$, $\underline{\overline{Y} - \hat\beta_1\overline{x}}$, $\underline{S_{xy}/S_{xx}}$, $\underline{n(\overline{xY} - \overline{x}\overline{Y})}$, $\underline{\beta_0}$, $\underline{\beta_1}$,
    $\underline{\sigma^2\overline{x^2}/S_{xx}}$, $\underline{\sigma^2/S_{xx}}$, $\underline{\frac{1}{n-2}(S_{yy} - \hat\beta_1 S_{xy})}$,
    If $Y_i \sim$ ________________, then $\hat\beta_j \sim$ ________________, $\frac{n-2}{\sigma^2}\hat\sigma^2 \sim$ _____,
    $\hat\sigma^2, \hat\beta_0, \hat\beta_1$ are ________________, ________________ $\sim t_{n-2}$,
    key: $\underline{N(\beta_0 + \beta_1 x_i, \sigma^2)}$, $\underline{N(\beta_j, \sigma^2_{\hat\beta_j})}$, $\underline{\chi^2(n-2)}$, $\underline{independent}$, $\underline{\frac{\hat\beta_j - \beta_j}{\hat\sigma_{\hat\beta_j}}}$,

Final Exam cover all formulae in 447 & 448.