1. (30 pts) Suppose that one can only observe one \( \mathbf{X} \sim \text{Multinomial}(n, p_1, p_2, p_3) \), where \( n = 2 \).

Set
\[
H_0: \ p_1 = 0.01, \ p_2 = 0.05 \text{ v.s. } H_1: \ H_0 \text{ is not true.}
\]

a. How many values of \( \mathbf{X} \) are there and what are they?

b. Simplify \( f_{\mathbf{X}}(\mathbf{y}) \) for each of these distinct values \( \mathbf{y} \) of \( \mathbf{X} \).

c. Derive the likelihood ratio test with level \( \alpha = 0.05 \).

d. Compute \( P(H_1|H_0) \) and \( P(H_0|H_1) \).

2. (40 pts total) Let \( \hat{F}(t) = \frac{1}{n} \sum_{i=1}^{n} 1(X_i \leq t) \), where \( X_i \)’s are i.i.d. from \( \mathbf{X} \) with the density function
\[
f(x) = c/|x|^3, \ |x| \geq 1.
\]

2.1 (13 pts) Compute (you must show your work!)

(a) \((c, \mu_X, \sigma_X, F(t))\).

(b) A proper estimator of \( \sigma_Z^2 \), where \( Z_n = n\hat{\sigma}^2_{F(3)} \), assuming \( F(t) \) is unknown.

2.2. Circle one answer. **No need to justify.** (±3 each; DNK (do not know) 0 pt).

(c) Is \( \hat{F}(t) \) the UMVUE of \( F(t) \)? (Yes, No, DNK).

(d) \( \bar{X} \overset{a.s.}{\rightarrow} \mu_X \) (Yes, No, DNK).

(e) \( (\bar{X} - \mu_X)/\sigma_X \overset{D}{\rightarrow} N(0, 1) \) (Yes, No, DNK).

(f) \( \hat{F}(t) \overset{a.s.}{\rightarrow} F(t) \) (Yes, No, DNK).

(g) \( \hat{F}(t) - F(t) \overset{D}{\rightarrow} N(0, \sigma^2_{F(t)}) \) (Yes, No, DNK).

(h) \( \frac{\hat{\sigma}^2_{F(0)} - \sigma^2_{F(0)}}{\sigma^2_{F(0)}} \overset{D}{\rightarrow} N(0, 1) \) by the delta method? (Yes, No, DNK).

(i) \( \frac{\hat{\sigma}^2_{F(3)} - \sigma^2_{F(3)}}{\sigma^2_{F(3)}} \overset{D}{\rightarrow} N(0, 1) \) (Yes, No, DNK).

(j) \( \hat{\sigma}^2_{F(t)} \overset{a.s.}{\rightarrow} 0 \) (Yes, No, DNK).

(k) \( n\hat{\sigma}^2_{F(t)} \overset{D}{\rightarrow} F(t)(1 - F(t)) \) (Yes, No, DNK).

3. (30 pts) Suppose that there are two independent random samples: \( X_1, \ldots, X_n \) i.i.d. \( \sim N(0, \sigma^2_X) \), and \( Y_1, \ldots, Y_n \) i.i.d. \( \sim N(0, \sigma^2_Y) \), where \( \sigma^2_X \) and \( \sigma^2_Y \) are unknown. Define \( \gamma = \sigma^2_Y/\sigma^2_X \).
(1) If $n = 25$, construct an approximate 95% CI for $\gamma$ based on Wald-type statistic.
(2) What is its coverage probability? You can use R code as a solution.
Solutions.
1. Suppose that one can only observe one $\mathbf{X} \sim \text{Multinomial}(n, p_1, p_2, p_3)$, where $n = 2$. Set $H_0$: $p_1 = 0.01, p_2 = 0.05$ v.s. $H_1$: $H_0$ is not true.
   
a. How many values of $\mathbf{X}$ are there and what are they?
b. Simplify $f_\mathbf{X}(\mathbf{y})$ for each of these distinct values $\mathbf{y}$ of $\mathbf{X}$.
c. Derive the likelihood ratio test with level $\alpha = 0.05$.
d. Compute $P(H_1|H_0)$ and $P(H_0|H_1)$.

Sol. $\mathbf{X} \in \{(2, 0, 0), (0, 2, 0), (0, 0, 2), (1, 1, 0), (1, 0, 1), (0, 1, 1)\}$.

$\hat{p} = \mathbf{X}/2$. Write $f(\mathbf{y}, \mathbf{p}) = f_\mathbf{X}(\mathbf{y}, \mathbf{p})$.

$$
\lambda = \frac{f(\mathbf{X}, \mathbf{p}_0)}{f(\mathbf{X}, \hat{\mathbf{p}})} = \frac{2!0.01^y_10.05^y_20.94^y_3}{2!(y_1/2)^{y_1}(y_2/2)^{y_2}(y_3/2)^{y_3}}
$$

$$= \begin{cases} 
0.01^2/1! & \text{if } \mathbf{X} = (2, 0, 0), \\
0.01 * 0.05/0.5^2 & \text{if } \mathbf{X} = (1, 1, 0), \\
0.01 * 0.94/0.5^2 & \text{if } \mathbf{X} = (1, 0, 1), \\
0.05^2/1! & \text{if } \mathbf{X} = (0, 2, 0), \\
0.05 * 0.94/0.5^2 & \text{if } \mathbf{X} = (0, 1, 1) \\
0.94^2/1! & \text{if } \mathbf{X} = (0, 0, 2), 
\end{cases}
$$

$$4\lambda = \begin{cases} 
0.01^2/0.5^2 & \text{if } \mathbf{X} = (2, 0, 0), \\
0.01 * 0.05 & \text{if } \mathbf{X} = (1, 1, 0), \\
0.01 * 0.94 & \text{if } \mathbf{X} = (1, 0, 1), \\
0.05^2/0.5^2 & \text{if } \mathbf{X} = (0, 2, 0), \\
0.05 * 0.94 & \text{if } \mathbf{X} = (0, 1, 1) \\
0.94^2/0.5^2 & \text{if } \mathbf{X} = (0, 0, 2), 
\end{cases}
$$

$\phi = 1(\lambda < 0.01) = 1(\mathbf{X} \notin \{(0, 0, 2), (0, 1, 1)\})$.

$$P(H_1|H_0) = 1 - p_3^2 - 2p_2p_3 = 1 - 0.8836 - 0.094 = 0.0224,$$

$$P(H_0|H_1) = p_3^2 + 2p_2p_3, p_2, p_3 \in [0, 1] \text{ and } p_2 + p_3 \in [0, 1], \text{ but } \mathbf{p} \neq (0.01, 0.05, 0.94).$$
3. Suppose that there are two independent random samples: \( X_1, \ldots, X_n \) i.i.d. \( \sim N(0, \sigma_X^2) \), and \( Y_1, \ldots, Y_n \) i.i.d. \( \sim N(0, \sigma_Y^2) \), where \( \sigma_X^2 \) and \( \sigma_Y^2 \) are unknown. Define \( \gamma = \sigma_Y^2 / \sigma_X^2 \).

(1) If \( n = 25 \), construct an approximate 95% CI for \( \gamma \) based on Wald-type statistic.

(2) What is its coverage probability? You can use R code as a solution.

Sol. \( \hat{\gamma} = \frac{\hat{\sigma}_Y}{\hat{\sigma}_X} \) is called a Wald-type statistic if \( \theta \) is known and if

\[
P(|\hat{\theta} - \theta| \leq z_{\alpha/2}) \approx 1 - \alpha
\]  

(1) Let \( g(x, y) = y/x \). Then \( \gamma = g(\sigma_X^2, \sigma_Y^2) \). \( \frac{\partial g}{\partial(x, y)} = (-y/x^2, 1/x) \),

\[X^2/\sigma_X^2 \sim \chi^2(1)\]

\[
\Sigma = \text{Cov}(\left(\frac{X^2}{Y^2}\right)) = \left(\begin{array}{cc} 2\sigma_X^4 & 0 \\ 0 & 2\sigma_Y^4 \end{array}\right)/n.
\]

\[
\sigma_\gamma^2 \approx \left(\frac{\sigma_Y^2}{\sigma_X^4}, 1/\sigma_X^2\right) \left(\begin{array}{cc} 2\sigma_X^4 & 0 \\ 0 & 2\sigma_Y^4 \end{array}\right) \left(\frac{\sigma_Y^2}{\sigma_X^4}, 1/\sigma_X^2\right)'/n = (2\gamma^2 + 2\gamma^2)/n = 4\gamma^2/n
\]

\[\hat{\gamma} \pm 1.96 \times \frac{\hat{\gamma}}{\sqrt{n}}. (1 \pm 0.784)\hat{\gamma}\]

Coverage probability \( P(0.216\hat{\gamma} < \gamma < 1.784\hat{\gamma}) = P(F_{25,25} \in (\frac{1}{1.784}, \frac{1}{0.216})) \).

x=c(1.784,0.216)
x=1/x
pf(x[2],25,25)- pf(x[1],25,25)
0.922495

Some comments:

a. \( \hat{\sigma}_Y^2 = \frac{1}{n} \sum_i (Y_i - \bar{Y})^2 \) or \( S_Y^2 \) ? where \( S_Y^2 = \frac{1}{n-1} \sum_i (Y_i - \bar{Y})^2 \)?

b. Several types of CI for \( \gamma = \sigma_X^2 / \sigma_Y^2 \) based on \( \hat{\gamma} \).

b.1. \( \hat{\gamma}_1 = \frac{\hat{\sigma}_Y^2}{\hat{\sigma}_X^2} \sim F_{n-1,n-1} \), where \( \hat{\gamma}_1 = S_Y^2 / S_X^2 \);

b.2. \( \hat{\gamma} = \frac{\bar{Y}^2}{\bar{X}^2} \sim F_{n,n} \), where \( \hat{\gamma} = \frac{\bar{Y}^2}{\bar{X}^2} \);

b.3. Wald-type: \( \hat{\gamma} - \gamma = g(\bar{X}^2, \bar{Y}^2) - g(\sigma_X^2, \sigma_Y^2) \approx g'(\sigma_X^2, \sigma_Y^2)(\left(\frac{\bar{X}^2}{\bar{Y}^2}\right) - \left(\frac{\sigma_X^2}{\sigma_Y^2}\right)) \),

where \( \sqrt{n}(\bar{X}^2, \bar{Y}^2) - (\sigma_X^2, \sigma_Y^2)) \xrightarrow{D} N((0,0), \Sigma) \) (by the CLT).

b.4. Compute \( \sigma_\gamma^2 \) directly and \( \frac{\hat{\gamma} - \gamma}{\sigma_\gamma} \xrightarrow{D} N(0, 1) \).

\[\sigma_\gamma^2 = E((\bar{X}^2/\bar{Y}^2)^2) - (E(\bar{X}^2)^2)/E(\bar{Y}^2)^2) \approx \left(\frac{E(\bar{X}^2)}{E(\bar{Y}^2)}\right)^2 \text{?? Is it the correct answer?}
\]

(need to compute \( \sigma_{\bar{X},25}^2 \) or use formula \( \frac{2d_1^2 + d_2 - 2}{d_1(d_2 - 2)^2(d_2 - 4)} \).

Did not justify why Wald-type statistic work.

Counterexample. \( X_i \)'s \( \sim U(0, \theta) \), MLE does not achieve the CR lower bound and is not asymptotically normal. \( \hat{\theta} = X_{(n)} \) and \( P(\hat{\theta} \leq t) = (F_X(t))^n \).

2. Let \( \tilde{F}(t) = \frac{1}{n} \sum_{i=1}^n 1(X_i \leq t) \), where \( X_i \)'s are i.i.d. from \( X \) with the density function \( f(x) = c/|x|^3, |x| \geq 1 \).
2.1 Compute (you must show your work!)

(a) \((c, \mu_X, \sigma_X, F(t))\).

\(c = 1, E(X) = 0, \sigma_X = \infty.\)

\(F(t) = \begin{cases} \frac{1}{2t^2} & \text{if } t < -1 \\ 1/2 & \text{if } t \in [-1, 1] \\ 1 - \frac{1}{2t^2} & \text{if } t > 1. \end{cases} \)

(b) A proper estimator of \(\sigma_Z^2_n\), where \(Z_n = n\hat{\sigma}_{F(3)}^2\), assuming \(F(t)\) is unknown. Since \(\sigma_{\hat{F}(t)}^2 = F(t)(1 - F(t))/n\), set \(g(t) = t - t^2, g'(t) = 1 - 2t.\)

If \(t = 3, V(\hat{F}(t)(1 - \hat{F}(t))) \approx (1 - 2F(t))^2\)

\(\hat{\sigma}_2 = \hat{\hat{V}}(Z_n) = (1 - 2\hat{F}(t))^2\hat{F}(t)(1 - \hat{F}(t))/n.\)

**another way:** \(V(\hat{F}(t)(1 - \hat{F}(t))) = V(\hat{F}(t)) + V((\hat{F}(t))^2) - 2Cov(\hat{F}(t), (\hat{F}(t))^2) = \ldots\)

use mgf or directed method.

\(V(\hat{F}(t))^2 = E(((\sum_i W_i)^2/n^2)^2 - (E((\sum_i W_i)^2/n^2)^2, W = 1(X \leq t).\)

\(\text{mgf } E(e^{Wt}) = q + pe^t, \text{ where } p = F(t).\)

very tedious! not suggested.

**Different question:** \(\hat{\sigma}_W = ? \) where \(W_n = \hat{\sigma}_{F(3)}\sqrt{n} - \sqrt{F(3)}, \text{ and } h(t) = \sqrt{t - t^2}.\)

\(h'(t) = -\frac{1}{2}(t - t^2)^{-1/2}(1 - 2t).\)

\((h'(t))^2(t - t^2) = \frac{1}{4}(1 - 2t)^2,\)

\(SD(\sqrt{\hat{F}(t)(1 - \hat{F}(t)))} \approx \frac{1}{2}(1 - 2\hat{F}(t))/\sqrt{n}.\)

2.2. Circle one answer. No need to justify.

(c) Is \(\hat{F}(t)\) the UMVUE of \(F(t)\) ? (Yes, )

(d) \(\frac{\hat{X}}{\sigma_X} \xrightarrow{a.s.} \mu_X\) ? (Yes, ). SLLN

(e) \(\frac{\hat{X} - \mu_X}{\sigma_X^2} \xrightarrow{D} N(0, 1)\) ? (No, ). CLT is not applicable as \(\sigma_X^2 = \infty.\)

(f) \(\hat{F}(t) \xrightarrow{a.s.} F(t)\) ? (Yes, ). SLLN

(g) \(\hat{F}(t) - F(t) \xrightarrow{D} N(0, \sigma_{\hat{F}(t)}^2)\) ? (No ). \(\sigma_{\hat{F}(t)}^2 = F(t)S(t)/n, 1/n \to 1/n\)

(h) \(\frac{\hat{\sigma}_{\hat{F}(0)}^2 - \sigma_{\hat{F}(0)}^2}{\hat{\sigma}_{V(\hat{F}(0))}^2} \xrightarrow{D} N(0, 1)\) ? (No) (if use Wald type statistics as \(1 - 2F(0) = 0\).

\(\text{Yes} \) if use exact formula for \(\sigma_{\hat{F}(0)}^2.\)

(i) \(\frac{\hat{\sigma}_{\hat{F}(0)} - \sigma_{\hat{F}(0)}}{\sigma_{\hat{F}(0)}} \xrightarrow{D} N(0, 1)\) ? (Yes, ).

(j) \(\hat{\sigma}_{\hat{F}(t)}^2 \to 0\) ? (No).

(k) \(n\hat{\sigma}_{\hat{F}(t)}^2 \xrightarrow{a.s.} F(t)(1 - F(t))\) ? (No).