1. (Stat 401) Suppose you have a bag that contains three six-sided dice: two of the dice are fair (all six sides have 1/6 probability), but the third die is irregular, having probabilities given in the following table:

<table>
<thead>
<tr>
<th>Face</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(Face)</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.25</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Suppose further that these three dice look and feel identical. Randomly choose a die from the bag, roll this die three times, and let $X$ be the number of 5s observed.

(a) What is the probability of seeing two 5s?
(b) If $X = 2$ was observed, what is the probability the irregular die was rolled?

2. (Stat 411) Let $X_1, \ldots, X_n$ be an iid sample from a Poisson distribution with parameter $\theta > 0$.

(a) Suppose your measurement device breaks down and you can only observe if the Poisson counts are positive or not; that is, instead of $X_1, \ldots, X_n$ you observe only the corrupted data

$$ Y_i = \begin{cases} 1 & \text{if } X_i > 0 \\ 0 & \text{if } X_i = 0 \end{cases}, \quad i = 1, \ldots, n. $$

Find the MLE $\tilde{\theta}_n$ for $\theta$ using only the corrupted data $Y_1, \ldots, Y_n$.

(b) Find the asymptotic distribution of $\tilde{\theta}_n$ in part (a).

(c) Find the asymptotic relative efficiency of $\tilde{\theta}_n$ in (a) compared to the MLE $\hat{\theta}_n$ based on the full data. Denote this quantity by $\text{are}_\theta(\tilde{\theta}_n, \hat{\theta}_n)$.

i. Explain how $\text{are}_\theta(\tilde{\theta}_n, \hat{\theta}_n)$ reveals the loss of information resulting from the data corruption.

ii. How does $\text{are}_\theta(\tilde{\theta}_n, \hat{\theta}_n)$ change as $\theta$ moves from small to large values? Explain this behavior intuitively.

---

**Solutions**

1. (a) No matter the die chosen, the distribution of the number of 5s is binomial; but the success probabilities are different for each die. For a regular die, the distribution of $X$ is $\text{Bin}(3, 1/6)$; for the irregular die, the distribution of $X$ is $\text{Bin}(3, 1/4)$. Since the different dice form a partition of the sample space, we can use the total probability formula to evaluate $P(X = 2)$. If $I$ is the event that an irregular die is chosen, then

$$ P(X = 2) = 2P(X = 2 \mid I^c)P(I^c) + P(X = 2 \mid I)P(I) = (2/3)(\frac{2}{3})(1/6)^2(5/6)^1 + (1/3)(\frac{3}{2})(1/4)^2(3/4)^1 = 0.093. $$
(b) We know \( P(X = 2 \mid I) \) but we want \( P(I \mid X = 2) \); this is a job for Bayes theorem:

\[
P(I \mid X = 2) = \frac{P(X = 2 \mid I)P(I)}{P(X = 2)} = \frac{0.14 \cdot 1/3}{0.093} = 0.502.
\]

2. (a) The corrupted data \( Y_1, \ldots, Y_n \) are iid \( \text{Ber}(\eta) \), where \( \eta = P(\theta | X_1 > 0) = 1 - e^{-\theta} \), a function of \( \theta \). The function is one-to-one, so we can express \( \theta \) in terms of \( \eta \): \( \theta = -\log(1 - \eta) \). It is well-known that \( \hat{\eta}_n = \overline{Y} \) is the MLE for \( \eta \). Since \( \theta \) is a continuous function of \( \eta \), it follows that the MLE of \( \theta \) based on the corrupted data is \( \hat{\theta}_n = -\log(1 - \hat{\eta}_n) = -\log(1 - \overline{Y}) \).

(b) The asymptotic distribution of \( \overline{Y} \) is normal, according to the CLT. Specifically, \( \sqrt{n}(\overline{Y} - \eta) \) is approximately \( \mathcal{N}(0, \eta(1 - \eta)) \) when \( n \to \infty \). Recall that \( \hat{\eta}_n \) is a nice function of \( \overline{Y} \), so we can obtain its asymptotic distribution via the Delta Theorem. Let \( g(y) = -\log(1 - y) \), so that \( \hat{\theta}_n = g(\overline{Y}) \). Then

\[
\sqrt{n}(\hat{\theta}_n - \theta) = \sqrt{n}(g(\overline{Y}) - g(\eta))
\rightarrow \mathcal{N}(0, [g'(\eta)]^2 \eta(1 - \eta))
= \mathcal{N}(0, [1/(1 - \eta)]^2 \eta(1 - \eta))
= \mathcal{N}(0, \eta/(1 - \eta))
= \mathcal{N}(0, e^\theta - 1).
\]

(c) The MLE of \( \theta \) based on the full data is \( \hat{\theta}_n = \overline{X} \), and its variance is \( \theta/n \). So the asymptotic relative efficiency of \( \hat{\theta}_n \) to \( \hat{\theta}_n \) is

\[
\text{are}_\theta(\hat{\theta}_n, \hat{\theta}_n) = \frac{\text{asymptotic variance of } \hat{\theta}_n}{\text{asymptotic variance of } \hat{\theta}_n} = \frac{e^\theta - 1}{\theta}.
\]

i. Since \( e^\theta - 1 > \theta \) for all \( \theta > 0 \), the asymptotic relative efficiency is greater than 1, suggesting that \( \hat{\theta}_n \) is less efficient than \( \hat{\theta}_n \), a consequence of the data corruption.

ii. For \( \theta \approx 0 \), the asymptotic relative efficiency is close to 1; for large \( \theta \), the asymptotic relative efficiency is large. The intuition is that the corruption is less costly when \( \theta \) is small; specifically, if \( \theta \) is close to zero, then the full data set will likely contain mostly zeros and ones, just like the corrupted data set. On the other hand, when \( \theta \) is large, there will be a very large difference between the full and corrupted data, so the loss of efficiency will be greater.
1. (STAT401) Suppose $X_1$ and $X_2$ are independent, and $X_i$ has p.m.f.

$$f(x_i) = \lambda_i e^{-\lambda_i}, \quad x_i \geq 0, \quad i = 1, 2,$$

where $\lambda_i > 0$. Define $T = \min(X_1, X_2)$.

(i) Show that for $t \geq 0$, $P(T > t) = e^{-(\lambda_1 + \lambda_2)t}$.

(ii) Find $E(T)$.

(iii) Find $P(T = X_1)$.

Solution:

(i) $P(T > t) = P(X_1 > t, X_2 > t) = P(X_1 > t)P(X_2 > t) = e^{-(\lambda_1 + \lambda_2)t}$.

(ii) From (i), $F_T(t) = P(T \leq t) = 1 - e^{-(\lambda_1 + \lambda_2)t}$, and then the p.d.f. of $T$ is

$$\frac{dF_T(t)}{dt} = (\lambda_1 + \lambda_2)e^{-(\lambda_1 + \lambda_2)t}.$$

Therefore, $E(T) = \int_0^\infty t(\lambda_1 + \lambda_2)e^{-(\lambda_1 + \lambda_2)t}dt = (\lambda_1 + \lambda_2)^{-1}$.

(Remarks: A simpler way is that $E(T) = \int_0^\infty P(T > t)dt = (\lambda_1 + \lambda_2)^{-1}$.)

(iii) Simple calculation yields that

$$P(T = X_1) = P(X_1 \leq X_2) = \int_0^\infty \int_{x_1}^\infty \lambda_1 \lambda_2 e^{-(\lambda_1 x_1 + \lambda_2 x_2)} dx_2 dx_1$$

$$= \int_0^\infty \lambda_1 e^{-\lambda_1 x_1} e^{-\lambda_2 x_1} dx_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$
2. (STAT411) Suppose that $X_1,\ldots,X_n$ are i.i.d. from the c.d.f.,

$$F(x) = P(X \leq x | \alpha, \beta) = (x / \beta)^\alpha, \text{ if } 0 \leq x \leq \beta,$$

where the parameters $\alpha$ and $\beta$ are positive.

(i) Find a two-dimensional sufficient statistics for $(\alpha, \beta)$.
(ii) Find the maximum likelihood estimator of $\alpha$ and $\beta$.

Solution:

(i) The p.d.f. is $f(x) = \alpha x^{\alpha-1} \beta^{-\alpha}$, if $0 \leq x \leq \beta$, and the likelihood is then

$$L(\alpha, \beta) = \alpha^n \left( \prod_{i=1}^{n} x_i \right)^{\alpha-1} \beta^{-n\alpha} I_{[0,\beta]}(\max \{x_i\}) I_{[0,\infty]}(\min \{x_i\}).$$

By Factorization theorem, the sufficient statistics for $(\alpha, \beta)$ is $(\prod x_i, \max \{x_i\})$.

(ii) First, note that for any fixed $\alpha$, $L(\alpha, \beta)$ is 0 when $\beta < \max \{x_i\}$, and a decreasing function of $\beta$ as long as $\beta \geq \max \{x_i\}$. This implies that

$$L(\alpha, \beta) \leq L(\alpha, \max \{x_i\})$$

for any $\alpha$, so the m.l.e of $\beta$ is $\hat{\beta} = \max \{x_i\}$, and then it suffices to maximize $L(\alpha, \hat{\beta})$.

Next, the log-likelihood is

$$l(\alpha, \hat{\beta}) = \log L(\alpha, \hat{\beta}) = n \log \alpha + (\alpha - 1) \sum \log x_i - n\alpha \log \hat{\beta},$$

and its first derivative w.r.t. $\alpha$ is

$$\frac{\partial}{\partial \alpha} l(\alpha, \hat{\beta}) = \frac{n}{\alpha} + \sum \log x_i - n \log \hat{\beta}.$$

Setting it equal to zero yields that

$$\hat{\alpha} = \frac{n}{n \log \max \{x_i\} - \sum \log x_i} = \left( \log \max \{x_i\} - n^{-1} \sum \log x_i \right)^{-1}.$$

The second derivative is $-n\alpha^{-2} \leq 0$, so this is the m.l.e. of $\alpha$. 
3. (STAT411) Suppose that $X_1, ..., X_n$ are independently Bern($p_i$) random variables with

$$p_i = \frac{e^{a_i\beta}}{1 + e^{a_i\beta}}, \quad i = 1, \ldots, n,$$

where $a_1, \ldots, a_n$ are known positive constants and $\beta$ is an unknown parameter.

(i) Show that $T = \sum a_i X_i$ is a sufficient statistic for $\beta$, and compute $E(T)$ and $Var(T)$.

(ii) Construct the size $\alpha$ UMP test for testing

$$H_0 : \beta = 0 \quad vs \quad H_1 : \beta > 0.$$

**Solution:**

(i) The likelihood is

$$\prod p_i^{x_i} (1 - p_i)^{1-x_i} = \prod \frac{e^{a_i\beta x_i}}{1 + e^{a_i\beta}} = \prod \frac{e^{\beta T}}{(1 + e^{a_i\beta})}.$$ 

By factorization theorem, $T$ is a sufficient statistic for $\beta$. Next,

$$E(T) = \sum a_i E(X_i) = \sum a_i p_i,$$

$$Var(T) = \sum a_i^2 Var(X_i) = \sum a_i^2 p_i (1 - p_i).$$ 

(ii) For any $\beta_1 > 0$, consider simple test $H_0 : \beta = 0 \quad vs \quad H_1 : \beta = \beta_1$. By N-P Lemma, the most powerful test is to reject $H_0$ if

$$\Lambda = \frac{2^{-n}}{\prod (1 + e^{a_i\beta})} = \frac{\prod (1 + e^{a_i\beta})^\beta}{\prod (1 + e^{a_i\beta})} \leq k,$$

which is equivalent to reject $H_0$ if $T \geq k'$, where $k'$ is a constant so that $P_{H_0}(T \geq k') = \alpha$. Since the rejection region does not depend on $\beta_1$, it is a size $\alpha$ UMP test for testing

$$H_0 : \beta = 0 \quad vs \quad H_1 : \beta > 0.$$
Stat 401, Problem, Spring 2012:

Let $X_1, \ldots, X_n$ be iid random variables with common pdf

$$f(x) = \begin{cases} e^{-(x-\theta)} & x > \theta \\ 0 & \text{elsewhere} \end{cases}$$

where the parameter $\theta$ is a real number. Let $Y_n = \min\{X_1, \ldots, X_n\}$.

(a) Find the cdf of $Y_n$.
(b) Show that $Y_n \rightarrow \theta$ in probability as $n$ goes to infinity.

Stat 401, Solution, Spring 2012:

(a) For $y > \theta$,

$$P[X_1 > y] = \int_y^\infty e^{-(x-\theta)} = e^{\theta-y}.$$ 

Let $F_n$ be the cdf of $Y_n$. Then for any $y > \theta$,

$$F_n(y) = P[Y_n \leq y] = P[\min\{X_1, \ldots, X_n\} \leq y] = 1 - P[\min\{X_1, \ldots, X_n\} > y] = 1 - P[X_1 > y, \ldots, X_n > y] = 1 - (e^{\theta-y})^n = 1 - e^{n(\theta-y)}.$$ 

Note that $F_n(y) = 0$, if $y \leq \theta$.

(b) Proof: For any $\epsilon > 0$, since $Y_n = \min\{X_1, \ldots, X_n\} > \theta$, then

$$P[|Y_n - \theta| \geq \epsilon] = P[Y_n \geq \theta + \epsilon] = 1 - F_n(\theta + \epsilon) = e^{n(\theta-\theta-\epsilon)} = e^{-n\epsilon} \rightarrow 0$$ 

as $n$ goes to $\infty$. By the definition, $Y_n \rightarrow \infty$ in probability.
Four different fabrics are examined on a Martindale wear tester that can compare four materials in a single run (block). The weight loss (in milligrams) from five runs in measured and the following results are obtained:

<table>
<thead>
<tr>
<th>Fabric</th>
<th>Run (Block)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>36</td>
<td>17</td>
<td>30</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>38</td>
<td>18</td>
<td>39</td>
<td>40</td>
<td>25</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>36</td>
<td>26</td>
<td>41</td>
<td>38</td>
<td>28</td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>30</td>
<td>17</td>
<td>34</td>
<td>33</td>
<td>21</td>
</tr>
</tbody>
</table>

Denote the observations by $Y_{ij}$, $i=1,2,3,4$; $j=1,2,3,4,5$. We find that $\sum_{i=1}^{4} \sum_{j=1}^{5} Y_{ij}^2 = 19280$; $\sum_{i=1}^{4} \sum_{j=1}^{5} Y_{ij} = 602$; $Y_1 = 138$, $Y_2 = 160$, $Y_3 = 169$, $Y_4 = 135$; $Y_{a1} = 140$, $Y_{a2} = 78$, $Y_{a3} = 144$, $Y_{a4} = 141$, $Y_{b} = 99$.

(a) Obtain the appropriate ANOVA table and test whether there are differences among treatment means. You may need the critical F-value: $F(0.05; 3, 12) = 3.49$.

(b) Test whether there are differences among the block means. You may need the critical F-value: $F(0.05; 4, 12) = 3.26$. Has the blocking arrangement been useful? Why?

Stat 481, Solution, Spring 2012:

(a) $SSTO = 19280 - (602)^2/20 = 1159.8$;
$SSTreat = (138^2+160^2+169^2+135^2)/5 - (602)^2/20 = 165.8$;
$SSBlock = (140^2+78^2+144^2+141^2+99^2)/4 - (602)^2/20 = 905.3$;
$SSError = SSTO – SSTreat – SSBlock = 88.7$.

The appropriate ANOVA table is

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>165.8</td>
<td>3</td>
<td>55.3</td>
<td>7.5</td>
</tr>
<tr>
<td>Block</td>
<td>905.3</td>
<td>4</td>
<td>226.3</td>
<td>30.6</td>
</tr>
<tr>
<td>Error</td>
<td>88.7</td>
<td>12</td>
<td>7.4</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1159.8</td>
<td>19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The F-statistic for treatment effects is $7.5 > 3.49$. Therefore there are significant differences among the fabrics.

(b) The F-statistic for block effects is $30.6 > 3.26$. Therefore there are significant differences among the block means. The blocking arrangement is very useful. If blocks are ignored, the resulting F-statistic for treatment differences is 0.89 which is insignificant. Blocking has increased the precision of the comparison.
A random sample of observations \( \{X_1, \ldots, X_n\} \) is drawn from a population with unknown median \( M \), where its CDF function is assumed to be continuous and strictly increasing. The hypothesis to be tested concerns the value of the population median

\[ H_0 : M = M_0 \text{ vs } H_1 : M > M_0 \]

where \( M_0 \) is a specified value. Define \( \theta = P(X > M_0) \).

(1). Find an appropriate test statistic, and give its distribution under a general \( \theta \).

(2). For a sample with size \( n = 16 \), find the critical region given a significance level \( \alpha = 0.05 \).

(3). Let \( X \sim N(\mu = M, 4) \). The hypothesized median is given \( M_0 = 26 \), then calculate the power at \( M_1 = 27.7 \). [Hint: calculate \( \theta_1 = P(X > M_0 | M_1) \) first.]

**Solution:**

(1). Sign test statistic for Median

\[
K = \# \{X_i > M_0, i = 1, \ldots, n\} = \sum_{i=1}^{n} I_{\{X_i > M_0\}}
\]

\( K \sim \text{Binomial}(n, \theta) \), \( \theta = P(X > M_0) \)

(2). Under \( H_0 : M = M_0 \) then \( \theta = P(X > M_0) = 0.5 \), then \( K \sim B(n = 16, \theta = 0.5) \). Right-sided critical region is \( \{K > k_\alpha\} \). Use Binomial Table, we have \( \{K \geq 12\} \) with nominal significance level 0.05. Exact test size is 0.0384.

(3). Standardization based on \( X \sim N(\mu = M_1, 4) \):

\[
\theta_1 = P(X > 26 | M_1 = 27.7) = P \left( Z > (26 - 27.7) / \sqrt{4} \right)
\]

\[ = P(Z > -0.85) = 0.80 \]

Then \( K \sim B(n = 16, \theta = 0.8) \), and

\[
\text{Power}(M_1 = 27.7) = P(K \geq 12 | \theta_1 = 0.8) = 1 - 0.2018 = 0.7982.
\]
Try to fit data \( \{(x_i, Y_i), i = 1, ..., n\} \) with a simple linear regression model \( Y_i = \beta_0 + \beta_1 x_i \), where i.i.d. errors \( \varepsilon_i \sim N(0, \sigma^2) \).

(1). Based on the least square criterion (loss function), derive the normal equation and then the least squares estimates for the intercept and slope, \( \hat{\beta}_0, \hat{\beta}_1 \).

(2). Calculate the variance of linear coefficient estimator \( \hat{\beta}_1 \), \( \text{Var}(\hat{\beta}_1) \), and covariance of the two estimators, \( \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \).

[Hint: Calculate the variance components based on the expression of the two estimators
\[
\hat{\beta}_0 = \sum_{i=1}^{n} k_i Y_i, \quad \hat{\beta}_1 = \sum_{i=1}^{n} c_i Y_i, \quad \text{where} \quad c_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, k_i = \frac{1}{n} - c_i \cdot \bar{x};
\]
or based on the variance matrix of the estimator vector \( \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \).

Solution:

(1). Least square estimation:

\[
Q(\hat{\beta}_0, \hat{\beta}_1) = \min_{\beta_0, \beta_1} \{Q(\beta_0, \beta_1)\} = \min_{\beta_0, \beta_1} \left\{ \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)^2 \right\}
\]

\[
\begin{align*}
\frac{\partial Q}{\partial \beta_0} &= 0 \\
\frac{\partial Q}{\partial \beta_1} &= 0 \\
\Rightarrow \quad \sum_{i=1}^{n} Y_i &= n \beta_0 + \beta_1 \sum_{i=1}^{n} x_i \\
\sum_{i=1}^{n} Y_i x_i &= \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2
\end{align*}
\]

\[
\Rightarrow \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad \hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}
\]
or

\[
Q(\hat{\beta}) = \min_{\beta} Q(\beta) = \min_{\beta} \{ (Y - X \beta)'(Y - X \beta) \}
\]

\[
\frac{\partial Q}{\partial \beta} = 0 \quad \Rightarrow \quad (X'X) \hat{\beta} = X'Y
\]

\[
\Rightarrow \quad \hat{\beta} = (X'X)^{-1} X'Y
\]
Method 1. Variance matrix of the estimator vector $\hat{\beta}$

\[
\text{Var} \left( \hat{\beta} \right) = \text{Var} \left( \left( X'X \right)^{-1} X'Y \right) = \left( X'X \right)^{-1} X' \text{Var}(Y) X \left( X'X \right)^{-1} \\
= \left( X'X \right)^{-1} X' \cdot \sigma^2 I_n \cdot X \left( X'X \right)^{-1} = \sigma^2 \left( X'X \right)^{-1}
\]

Note that

\[
X'X = \left( \frac{n}{\sum_{i=1}^{n} x_i} \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i^2 \right), \quad \left( X'X \right)^{-1} = \frac{1}{n S_{xx}} \left( \frac{\sum_{i=1}^{n} x_i^2}{-\sum_{i=1}^{n} x_i} n \right)
\]

\[
\therefore \text{Var} \left( \hat{\beta}_1 \right) = \frac{\sigma^2}{S_{xx}}, \text{Cov} \left( \hat{\beta}_0, \hat{\beta}_1 \right) = -\frac{\sigma^2 \bar{x}}{S_{xx}}.
\]

Method 2. The coefficient estimator

\[
\hat{\beta}_1 = \sum_{i=1}^{n} c_i Y_i, \text{ where } c_i = \frac{(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
\]

and we can show that $\sum_{i=1}^{n} c_i = 0, \sum_{i=1}^{n} c_i^2 = 1/S_{xx}$.

\[
\text{Var} \left( \hat{\beta}_1 \right) = \text{Var} \left( \sum_{i=1}^{n} c_i Y_i \right) = \sum_{i=1}^{n} c_i^2 \text{Var}(Y_i) = \frac{\sigma^2}{S_{xx}}
\]

since $Y_i's$ are independent, $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$.

Covariance

\[
\text{Cov} \left( \hat{\beta}_0, \hat{\beta}_1 \right) = \text{Cov} \left( \bar{y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1 \right) = \text{Cov} \left( \bar{y}, \hat{\beta}_1 \right) - \text{Cov} \left( \bar{x}, \hat{\beta}_1 \hat{\beta}_1 \right)
\]

and

\[
\text{Cov} \left( \bar{y}, \hat{\beta}_1 \right) = \text{Cov} \left( \sum_{i=1}^{n} Y_i/n, \sum_{i=1}^{n} c_i Y_i \right) = \sum_{i=1}^{n} \frac{c_i}{n} \text{Var}(Y_i) = \sum_{i=1}^{n} \frac{c_i \sigma^2}{n} = 0.
\]

\[
\therefore \text{Cov} \left( \hat{\beta}_0, \hat{\beta}_1 \right) = -\bar{x} \text{Var} \left( \hat{\beta}_1 \right) = -\frac{\bar{x} \sigma^2}{S_{xx}}.
\]
**Question.** Consider the following sampling design based on a finite labeled population of $N = 10$ units:

<table>
<thead>
<tr>
<th>Sample (s)</th>
<th>P(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 2, 3, 7)</td>
<td>0.35</td>
</tr>
<tr>
<td>(2, 4, 5, 8)</td>
<td>0.25</td>
</tr>
<tr>
<td>(4, 6, 7, 9)</td>
<td>0.30</td>
</tr>
<tr>
<td>(1, 3, 6, 10)</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Suppose the sampling design has been implemented and a sample has actually been drawn. Let it be $s_1 = (1, 2, 3, 7)$.

(a) Justify that the Horvitz-Thompson Estimator [HTE] serves as one homogeneous linear unbiased estimator for the population mean $\bar{Y}$ of a study variable ‘$Y$’. Explicitly write down the form of the HTE for the chosen sample.

(b) Construct one more homogeneous linear unbiased estimator for the population mean $\bar{Y}$ and write down its form for the same chosen sample.

(c) Argue that for neither of the two estimators proposed by you, variance of the estimator can be unbiasedly estimated.

(d) If the purpose is to unbiasedly estimate the population variance and the sample drawn is the same as $s_1 = (1, 2, 3, 7)$, what would be your answer? Justify.

**Solution.**

We will use the following basic facts:

Fact # 1: For unbiased estimation of the population total or the population mean, we must necessarily have $\pi_i > 0$ for each $i$ where $\pi_i$ denotes the inclusion probability of the $i$th unit according to the given sampling design. It is readily verified that this holds for the sampling design under consideration.

Fact # 2: For unbiased estimation of the population variance or for unbiased estimation of the variance of an unbiased estimator, we must necessarily have $\pi_{ij} > 0$ for each pair of units $\{i,j\}$ where $\pi_{ij}$ denotes the joint inclusion of the units $i$ and $j$ according to the given sampling design. It is readily verified that $\pi_{1, 4} = \pi_{1, 5} = \pi_{1, 8} = \pi_{1, 9} = 0$ and so are many other pairs!
In view of Fact # 2, answers to Q3 and Q4 are immediate. We are not in a position to estimate the population variance nor the variance of the homogeneous linear unbiased estimators being proposed.

(a) \[
\text{HTE}(Y\text{BAR}) = \frac{1}{N} \times \text{sum} \left[ \frac{y_i}{\pi_i} \right] = \frac{[y_1/\pi_1 + y_2/\pi_2 + y_3/\pi_3 + y_7/\pi_7]}{10} \\
= \frac{[y_1/(0.45) + y_2/(0.60) + y_3/(0.45) + y_7/(0.65)]}{10} \\
= \frac{y_1}{45} + \frac{y_2}{60} + \frac{y_3}{45} + \frac{y_7}{65}. 
\]

(b) This time we start with a very general form of the homogeneous linear unbiased estimator as:

<table>
<thead>
<tr>
<th>Samples</th>
<th>y_1</th>
<th>y_2</th>
<th>y_3</th>
<th>y_4</th>
<th>y_5</th>
<th>y_6</th>
<th>y_7</th>
<th>y_8</th>
<th>y_9</th>
<th>y_10</th>
</tr>
</thead>
<tbody>
<tr>
<td>P[sample]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1, 2, 3, 7)</td>
<td>a</td>
<td>b</td>
<td>c</td>
<td>-</td>
<td>-</td>
<td>d</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2, 4, 5, 8)</td>
<td>-</td>
<td>e</td>
<td>-</td>
<td>f</td>
<td>g</td>
<td>-</td>
<td>h</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4, 6, 7, 9)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>i</td>
<td>-</td>
<td>j</td>
<td>k</td>
<td>-</td>
<td>l</td>
<td>-</td>
</tr>
<tr>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1, 3, 6, 10)</td>
<td>m</td>
<td>-</td>
<td>n</td>
<td>-</td>
<td>-</td>
<td>o</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>p</td>
</tr>
<tr>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conditions for unbiasedness are given by

(i) \[0.35 a + 0.10 m = 1/10;\]
(ii) \[0.35 b + 0.25 e = 1/10;\]
(iii) \[0.35 c + 0.10 n = 1/10;\]
(iv) \[0.25 f + 0.30 i = 1/10;\]
(v) \[0.25 g = 1/10;\]
(vi) \[0.30 j + 0.10 o = 1/10;\]
(vii) \[0.35 d + 0.30 k = 1/10;\]
Clearly, HTE corresponds to the special choice of the coefficients viz.,

\[ a = m = \frac{1}{10} \pi_1; \quad b = e = \frac{1}{10} \pi_2, \text{ etc etc} \]

We just have to make a different choice of the coefficients. For example, we can choose:

Etc etc.....

At the end, we write down the estimate based on the sample \( s_1 \).
Problem 1. Consider a sequence of items from a production process with each item being graded as good or defective. Suppose that a good item is followed by another good item with probability $a$. Similarly, a defective item is followed by another defective item with probability $b$. If the first item is defective, what is the probability that the second good item to appear is the fifth item?

Solution: The first item is defective and the second good item to appear is the fifth item.

$$= \{ X_1 = D, X_2 = G, X_3 = D, X_4 = D, X_5 = G \}$$
$$\cup \{ X_1 = D, X_2 = D, X_3 = G, X_4 = D, X_5 = G \}$$
$$\cup \{ X_1 = D, X_2 = D, X_3 = D, X_4 = G, X_5 = G \}$$

$$= (1 - b)(1 - a) \cdot b \cdot (1 - b)$$
$$+ b \cdot (1 - b)(1 - a)(1 - b)$$
$$+ b \cdot b \cdot (1 - b) \cdot a$$
Problem 2. A coin is tossed repeatedly until two successive heads appear. Find the mean number of tosses required.

Solution: Let $X_n$ be the cumulative number of successive heads. The state space of $X_n$ is $\{0, 1, 2, \ldots\}$ and the transition probability matrix is

$$P = \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

Let $T = \min\{n \geq 0 : X_n = 2\}$

$V_0 = E\{T | X_0 = 0\}$ and $V_1 = E\{T | X_0 = 1\}$

Then we have:

$$\begin{cases} V_0 = \frac{1}{2} V_0 + \frac{1}{2} V_1 + 1 \\ V_1 = \frac{1}{2} V_0 + 1 \end{cases}$$

Solving the above equation, we obtain $V_0 = 6$.
\( (\text{Stat 471 Linear and Nonlinear Programming}) \): A bank is attempting to determine where its assets should be invested during the current year. At present $500,000 is available for investments in bonds, home loans and auto loans and personal loans. The rate of return on bonds is 10\%. On auto loans it is 13\%. On home loans it is 16\% On personal loans it is 20\%. To avoid portfolio risks the bank has the following policy. Total investments in personal loans cannot exceed the investment in bonds. Investment in home loans cannot exceed investment in auto loans. At most 25\% of the investment can be in personal loans. Total investment in auto loans and personal loans put together is at most half the amount invested in bonds and home loans. Just formulate an LP reflecting the banks ultimate interest.

**Solution:**

Let \( B, A, H, P \) denote investments in bonds, auto loans, home loans and personal loans. The available total for investment is 500K. Thus the objective is to maximize the total rate of return with appropriate constraints that can be expressed as

\[
\text{max } 10B + 13A + 16H + 20P
\]

subject to

\[
\begin{align*}
B + A + H + P & \leq 500 \text{ available total for investment} \\
P - B & \leq 0 \text{ personal loans can’t be more than investment in bonds} \\
H - A & \leq 0 \text{ Home loans can’t exceed investment in auto loans} \\
3P - B - A - H & \leq 0 \text{ at most 25\% can be in personal loans} \\
2A + 2P - B - H & \leq 0 \text{ investment in auto loans and personal loans... home loans}
\end{align*}
\]

\( (\text{Stat 471 Linear and Nonlinear Programming}) \): A dance teacher for the Asian Indian community is tensed up with organizing a dance drama on the Indian epic Ramayana for CTU, the up coming music and dance festival. There are 5 characters in the drama and the 5 girls Asha, Banu, Chaya, Deepa and Ela who are to act in the drama will need the following hours of rehearsal time if given the characters of Rama, Sita, Hanuman, Vali and Ravana.

\[
\begin{pmatrix}
\text{Rama} & \text{Sita} & \text{Vali} & \text{Hanuman} & \text{Ravana} \\
\text{Asha} & 13 & 18 & 6 & 15 & 14 \\
\text{Banu} & 16 & 7 & 14 & 13 & 1 \\
\text{Chaya} & 7 & 8 & 11 & 9 & 6 \\
\text{Deepa} & 14 & 13 & 6 & 5 & 4 \\
\text{Ela} & 22 & 14 & 9 & 10 & 7
\end{pmatrix}
\]

Use Kuhn’s Hungarian Method to find the assignment of characters to appropriate girls that will minimize the total rehearsal time?

You get no credit if you do not use Hungarian algorithm.

**Solution:**

Subtracting the smallest entry in each row from the row entries and subtracting the smallest entry of each column from the column entries will not change the optimal assignment. Thus the optimal assignment for the original matrix above is the same as the optimal assignment
for the new matrix
\[
\begin{bmatrix}
7 & 12 & 0 & 9 & 8 \\
15 & 6 & 13 & 12 & 0 \\
1 & 2 & 5 & 3 & 0 \\
10 & 9 & 2 & 1 & 0 \\
15 & 7 & 2 & 3 & 0 \\
\end{bmatrix}
\]

Now on the new matrix subtracting the least element in each column from the columns yield the matrix
\[
\begin{bmatrix}
6 & 10 & 0 & 8 & 8 \\
14 & 4 & 13 & 11 & 0 \\
0 & 0 & 5 & 2 & 0 \\
9 & 7 & 2 & 0 & 0 \\
14 & 5 & 2 & 2 & 0 \\
\end{bmatrix}
\]

which too has the same optimal assignment. The following is a minimal deletion of the rows and columns containing 0’s. The deleted rows 1, 3, 4 and and column 5 and their entries are in bold face.
\[
\begin{bmatrix}
6 & 10 & 0 & 8 & 8 \\
14 & 4 & 13 & 11 & 0 \\
0 & 0 & 5 & 2 & 0 \\
9 & 7 & 2 & 0 & 0 \\
14 & 5 & 2 & 2 & 0 \\
\end{bmatrix}
\]

Kuhn’s Hungarian algorithm says that a new matrix is formed by computing \( \theta \) the smallest among among all the undeleted entries. The new matrix is formed as follows: From the current matrix above subtract \( \theta \) from all undeleted entries and add \( \theta \) to all entries deleted both along rows and columns and leave the rest unchanged. The new matrix has no fewer zero entries and has the same optimal assignment as the initial matrix. Thus we get the new matrix
\[
\begin{bmatrix}
6 & 10 & 0 & 8 & 10 \\
12 & 2 & 11 & 9 & 0 \\
0 & 0 & 5 & 2 & 7 \\
9 & 7 & 2 & 0 & 2 \\
12 & 3 & 0 & 0 & 0 \\
\end{bmatrix}
\]

Again the minimum deletion is 4 and it is to delete the 3rd row and last 3 columns. We get the new matrix with deleted row and column entries in bold face as follows:
\[
\begin{bmatrix}
6 & 10 & 0 & 8 & 10 \\
12 & 2 & 11 & 9 & 0 \\
0 & 0 & 5 & 2 & 7 \\
9 & 7 & 2 & 0 & 2 \\
12 & 3 & 0 & 0 & 0 \\
\end{bmatrix}
\]
The minimum of undeleted entries is 2 and subtracting 2 from all undeleted entries and adding 2 to all twice deleted entries we arrive at the new matrix

\[
\begin{bmatrix}
4 & 8 & 0^* & 8 & 10 \\
10 & 0^* & 11 & 9 & 0 \\
0^* & 0 & 7 & 4 & 9 \\
7 & 5 & 2 & 0^* & 2 \\
11 & 1 & 0 & 0 & 0^*
\end{bmatrix}
\]

The new matrix has an optimal assignment of 0’s in row 1, column 3; row 2, column 2; row 3 column 1; row 4, column 4; and row 5, column 5. Thus the optimal assignment of characters to girls is given as follows:

Asha → Vali
Banu → Sita
Chaya → Rama
Deepa → Hanuman
Ela → Ravana

(Stat 473 Game Theory:) Bob and Joe play the following game. With eyes closed Bob shows 3 or 4 fingers. With eyes closed Joe shows 4 or 5 fingers. With eyes open they check the total fingers shown. If this total is even, Bob wins from Joe a dollar amount equal to total fingers shown by both. If odd, Bob pays Joe this total amount. If Bob chooses 3 or 4 fingers with probability \( \frac{1}{5} \) and \( \frac{4}{5} \) respectively, what is best for Joe? Should he show 4 fingers or 5 fingers? If Bob chooses 3 fingers with probability \( p \) and chooses 4 fingers with probability \( 1 - p \), what is best for Joe? Find those \( p \)'s for which Joe chooses 4 fingers. Find those \( p \)'s for which Joe chooses 5 fingers.

Solution:

The following payoff matrix summarizes the payoff to Bob from Joe where Bob is the row player and Joe is the column player.

\[
\begin{pmatrix}
4 & 5 \\
3 & \begin{pmatrix}
-7 & 8 \\
4 & -9
\end{pmatrix} \\
4 & \begin{pmatrix}
-7 & 8 \\
4 & -9
\end{pmatrix}
\end{pmatrix}
\]

If Bob chooses 3 or 4 fingers with chance \( \frac{1}{5}, \frac{4}{5} \) respectively then in case Joe chooses 4 fingers then Bob’s net gain is \( \frac{1}{5},(-7) + \frac{4}{5}.(8) = 5 \). If Joe chooses 5 fingers, then the expected gain to Bob is \( \frac{1}{5},(8) + \frac{4}{5}.(-9) = -\frac{28}{5} \). Thus Joe prefers to choose 5 fingers. If Bob chooses 3 fingers with probability \( p \) and 4 fingers with probability \( (1 - p) \), then Bob’s expected payoff is \( -7p + 8(1 - p) = 8 - 15p \) in case Joe shows 4 fingers. Bob’s expected payoff is \( 8p - 9(1 - p) = 17p - 9 \) in case Joe shows 5 fingers. Joe would like to choose \( \min(8 - 15p, 17p - 9) \) in case he knows \( p \). For \( p < \frac{17}{32} \) Joe should choose 5 fingers. For \( p > \frac{17}{32} \) Joe should choose 4 fingers.