The Hong Kong University of Science & Technology

MATH3423 - Statistical Inference
Final Examination - Fall 2014/2015

Answer ALL Questions **Date: 12 December 2014**

Full marks: 80 + 10 for Bonus **Time Allowed: 3 hours**

- DO NOT open the exam paper until instructed to do so.
- It is a closed-book examination.
- Five questions are included in this paper.
- Give detailed explanation how to obtain the final answer. NO mark will be given if only the final answer is written down.
- Unless specified, numerical answers should be EITHER exact OR corrected to 6 decimal places.
- You may write on the both sides of the examination booklet.
- Cheating is a serious offense. Students caught cheating are subject to a zero score as well as additional penalties.

Name : ______________________

Student Number : ______________

Signature : ____________________

For marking use only:

Question No.	Marks	Out of
1		20
2		20
3		20
4		20
5		10

1. Let $X_1, ..., X_n$ be a random sample from the Bernoulli(θ), where θ is the unknown parameter.

(a) (**2 marks**) Find the complete and sufficient statistic for θ. Find its distribution.

Answer

$$\begin{aligned} f_X(x;\theta) &= \theta^x(1-\theta)^{1-x} \\ &= \exp\{x\log\theta + (1-x)\log\theta\} \\ &= \exp\left\{\log(1-\theta) + x\log\frac{\theta}{1-\theta}\right\} \end{aligned}$$

$$\therefore\ a(\theta) = \log(1-\theta), b(X) = 0, c(\theta) = \log(\frac{\theta}{1-\theta}), d(X) = X,$$

$\therefore$ Bin$(1,\theta)$ belongs to the exponential family and $\sum_{i=1}^n d(X_i) = \sum_{i=1}^n X_i$ is complete and sufficient for θ.

$$\begin{aligned} E(e^{t\sum X_i}) &= \prod_{i=1}^n E(e^{tX_i}) \\ &= \prod_{i=1}^n (1-\theta+\theta e^t) \\ &= (1-\theta+\theta e^t)^n \end{aligned}$$

$\Rightarrow\ \sum_{i=1}^n X_i\ \sim\ \text{Bin}(n\theta)$.

(b) (**3 marks**) Find the UMVUE for θ^2.

Answer

$$\begin{aligned} E((\sum_{i=1}^n X_i)^2) &= Var(\sum_{i=1}^n X_i)\ +\ (E(\sum_{i=1}^n X_i))^2 \\ &= n\theta(1-\theta)\ +\ n^2\theta^2 \\ &= n\theta\ +\ n(n-1)\theta^2 \\ &= nE(\frac{\sum_{i=1}^n X_i}{n})\ +\ n(n-1)\theta^2 \end{aligned}$$

$$\Rightarrow E\left(\frac{(\sum_{i=1}^n X_i)^2 - (\sum_{i=1}^n X_i)}{n(n-1)}\right)\ =\ \theta^2$$

(c) (**3 marks**) Find the CRLB for θ^2. Is the variance of the UMVUE for θ^2 equal to its CRLB? Explain in details.

Answer

$$
\begin{aligned}
\log f(X;\theta) &= X\log\theta + (1-X)\log(1-\theta) \\
\frac{\partial}{\partial\theta}\log f(X;\theta) &= \frac{X}{\theta} - \frac{1-X}{1-\theta} \\
\frac{\partial^2}{\partial\theta^2}\log f(X;\theta) &= \frac{-X}{\theta^2} - \frac{1-X}{(1-\theta)^2} \\
E\left[\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right] &= E\left[\frac{-X}{\theta^2} - \frac{1-X}{(1-\theta)^2}\right] \\
&= -\frac{1}{\theta^2}E[X] - \frac{1}{(1-\theta)^2}E[1-X] \\
&= -\frac{1}{\theta^2}\theta - \frac{1}{(1-\theta)^2}(1-\theta) \\
&= -\frac{1}{\theta(1-\theta)}
\end{aligned}
$$

$$
\Rightarrow \text{The CRLB} = \frac{(2\theta)^2}{-nE\left[\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right]} = \frac{4\theta^3(1-\theta)}{n}
$$

$$
\begin{aligned}
\frac{\partial}{\partial\theta}\sum_{i=1}^{n}\log f(x) &= \frac{(1-\theta)\sum_{i=1}^{n}x_i - (n-\sum_{i=1}^{n}x_i)\theta}{(1-\theta)\theta} \\
&= \frac{n}{\theta(1-\theta)}(\bar{x}-\theta)
\end{aligned}
$$

Hence, CRLB for θ^2 cannot be achieved.

(d) (**2 marks**) Find the limiting distribution of the maximum likelihood estimator for θ^2 as $n \to \infty$ by Delta method. What phenomenon do you observe?

Answer

$$
\hat{\theta} = \bar{x} \;\Rightarrow\; \widehat{\theta^2} = \bar{X}^2
$$

As $n \to \infty$,

$$
\begin{aligned}
\bar{X} &\to N\left(\theta,\ \frac{\theta(1-\theta)}{n}\right) \\
\Rightarrow \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}X_i - \theta\right) &\to N\left(0,\ \frac{\theta(1-\theta)}{n}\right) \\
\bar{X}^2 &\to N\left(\theta^2,\ \frac{4\theta^3(1-\theta)}{n}\right) \\
\Rightarrow \sqrt{n}\left(\bar{X}^2 - \theta^2\right) &\to N\left(0,\ \frac{4\theta^3(1-\theta)}{n}\right)
\end{aligned}
$$

As $n \to \infty$, the maximum likelihood estimator of θ^2 is unbiased, normally distributed and fully efficiency, i.e., its variance is equal to C-R lower bound.

(e) (**5 marks**) Find the UMVUE of $P(X_1 + X_2 + X_3 = 1)$.

Answer

$$P(X_1 + X_2 + X_3 = 1) \;=\; \binom{3}{1} \theta(1-\theta)^2 \;=\; 3\theta(1-\theta)^2$$

$$\begin{aligned} E(h(y)) &= 3\theta(1-\theta)^2 \\ \Rightarrow \sum_{y=0}^{n} h(y) \binom{n}{y} \theta^y (1-\theta)^{n-y} &= 3\theta(1-\theta)^2 \\ \Rightarrow \sum_{y=0}^{n} h(y) \binom{n}{y} \frac{1}{3}\, \theta^{y-1} (1-\theta)^{n-2-y} &= 1 \end{aligned}$$

Set $h(y) = 0$ for $y = 0,\ n,\ n-1$.

$$\Rightarrow \sum_{y=1}^{n-2} h(y) \binom{n}{y} \frac{1}{3}\, \theta^{y-1} (1-\theta)^{n-2-y} \;=\; 1$$

Let $x = y - 1$

$$\Rightarrow \sum_{x=0}^{n-3} h(x+1) \binom{n}{x+1} \frac{1}{3}\, \theta^{x} (1-\theta)^{n-3-x} \;=\; 1$$

$$\Rightarrow \sum_{x=0}^{n-3} h(x+1) \frac{\binom{n}{x+1}}{\binom{n-3}{x}} \frac{1}{3} \binom{n-3}{x} \theta^{x} (1-\theta)^{n-3-x} \;=\; 1$$

$$\Rightarrow h(x+1) \;=\; \frac{3\binom{n-3}{x}}{\binom{n}{x+1}}$$

$$\Rightarrow h(y) \;=\; \begin{cases} \dfrac{3\binom{n-3}{y-1}}{\binom{n}{y}} & \text{for } y = 1, 2, \ldots, n-2 \\ 0 & \text{for } y = 0, n, n-1 \end{cases}$$

(f) (**5 marks**) Find the maximum likelihood estimator for the variance of $\sum X_i$, i.e., $n\theta(1-\theta)$. Is it unbiased? Hence or otherwise, find the UMVUE for the variance of of $\sum X_i$.

Answer

The maximum likelihood estimator for the variance of $\sum X_i$ is $n\bar{X}(1-\bar{X})$.

$$\begin{aligned}
E(n\bar{X}(1-\bar{X})) &= E(\sum_{i=1}^{n} X_i(1-\bar{X})) \\
&= \frac{1}{n}E(\sum_{i=1}^{n} X_i(n-\sum_{i=1}^{n} X_i)) \\
&= \frac{1}{n}E(Y(n-Y)) \quad \text{where } Y=\sum_{i=1}^{n} X_i \\
&= \frac{1}{n}[E(nY)-E(Y^2)] \\
&= \frac{1}{n}[nE(Y)-Var(Y)-[E(Y)]^2] \\
&= \frac{1}{n}[n\cdot n\theta-n\theta(1-\theta)-n^2\theta^2] \\
&= \frac{1}{n}[n^2\theta+n\theta^2-n\theta-n^2\theta^2] \\
&= \theta(n-1)-\theta^2(n-1) \\
&= (n-1)\theta(1-\theta) \\
\therefore\ E[\frac{n^2}{n-1}\bar{X}(1-\bar{X})] &= n\theta(1-\theta)
\end{aligned}$$

Since $\frac{n^2}{n-1}\bar{X}(1-\bar{X})$ is function of complete sufficient statistic $\sum_{i=1}^{n} X_i$, it is the UMVUE for $n\theta(1-\theta)$.

2. Let $X_1, ..., X_n$ be a r.s. from the continuous uniform distribution in the interval $(\theta, 2\theta)$, $\theta \in (0, \infty)$.

   Hint:

$$f(y_1, y_n) = n(n-1)(y_n - y_1)^{n-2}/\theta^n \qquad \theta \le y_1 \le y_n \le 2\theta$$

   and

$$Cov(Y_1, Y_n) = \frac{\theta^2}{(n+1)^2(n+2)} .$$

   (a) (**3 marks**) Find the method of moments estimator, $\tilde{\theta}$, for θ. Is it unbiased? Hence or otherwise, find an unbiased estimator of θ as a function of $\tilde{\theta}$. What is its corresponding variance?

   Answer

$$f_X(x) = \frac{1}{2\theta - \theta} = \frac{1}{\theta}$$

   and

$$F_X(x) = \int_\theta^x \frac{1}{\theta}\, dt = \left[\frac{t}{\theta}\right]_\theta^x = \frac{x}{\theta} - 1 = \frac{1}{\theta}(x - \theta)$$

   Method of moments estimator for θ:

$$\begin{aligned} \widetilde{E(X)} &= \frac{1}{n}\sum_{i=1}^n X_i \\ \Rightarrow \frac{3\tilde{\theta}}{2} &= \bar{X} \\ \Rightarrow \tilde{\theta} &= \frac{2\bar{X}}{3} \end{aligned}$$

$$\begin{aligned} E(\tilde{\theta}) &= \theta \Rightarrow \text{Unbiased} \\ Var(\tilde{\theta}) &= \frac{\theta^2}{27n} \end{aligned}$$

   (b) (**3 marks**) Find $E(Y_1)$, where $Y_1 = \min(X_1, \ldots, X_n)$. Hence or otherwise, find an unbiased estimator of θ as a function of Y_1.

   Answer

$$\begin{aligned} f_{Y_1}(y_1) &= \frac{n!}{(1-1)!(n-1)!}[F_X(y_1)]^{1-1}[1 - F_X(y_1)]^{n-1} f_X(y_1) \\ &= n\left(1 - \frac{1}{\theta}(y_1 - \theta)\right)^{n-1}\frac{1}{\theta} \\ &= \frac{n}{\theta}\left[2 - \frac{y_1}{\theta}\right]^{n-1} \qquad \theta < y_1 < 2\theta \end{aligned}$$

$$
\begin{aligned}
\Rightarrow\ E(Y_1) &= \int_\theta^{2\theta} y_1 f_{Y_1}(y_1)\ dy_1 \\
&= \frac{n}{\theta}\int_\theta^{2\theta} y_1 \left(2-\frac{y_1}{\theta}\right)^{n-1}\ dy_1 \\
&= \frac{n}{\theta^n}\int_\theta^{2\theta} y_1(2\theta-y_1)^{n-1}\ dy_1 \\
&= \frac{n}{\theta^n}\int_\theta^{0} -(2\theta-z)z^{n-1}\ dz, \qquad\qquad \text{let } z=2\theta-y_1, dz=-dy_1 \\
&= \frac{n}{\theta^n}\int_0^{\theta} (-z^n+2\theta z^{n-1})\ dz \\
&= \frac{n}{\theta^n}\left[\frac{-1}{n+1}\ z^{n+1}+\frac{2\theta}{n}\ z^n\right]_0^\theta \\
&= \frac{n}{\theta^n}\left[\frac{-\theta^{n+1}}{n+1}+\frac{2\theta^{n+1}}{n}\right] \\
&= \frac{-n\theta}{n+1}+2\theta \\
&= \frac{n+2}{n+1}\ \theta \\
\Rightarrow\ U_a &= \frac{n+1}{n+2}Y_1
\end{aligned}
$$

(c) (**2 marks**) Find $E(Y_n)$, where $Y_n=\max(X_1,\ldots,X_n)$. Hence or otherwise, find an unbiased estimator of θ as a function of Y_n.

Answer

$$
\begin{aligned}
f_{Y_n}(y_n) &= \frac{n!}{(n-1)!(n-n)!}[F_X(y_n)]^{n-1}[1-F_X(y_n)]^{n-n}f_X(y_n) \\
&= n\left(1-\frac{1}{\theta}(y_n-\theta)\right)^{n-1}\frac{1}{\theta} \\
&= \frac{n}{\theta}\left[\frac{y_n}{\theta}-1\right]^{n-1} \qquad\qquad \theta<y_n<2\theta
\end{aligned}
$$

$$
\begin{aligned}
E(Y_n) &= \int_\theta^{2\theta} y_n f_{Y_n}(y_n)\ dy_n \\
&= \frac{n}{\theta}\int_\theta^{2\theta} y_n\left(\frac{y_n}{\theta}-1\right)^{n-1}\ dy_n \\
&= \frac{n}{\theta^n}\int_\theta^{2\theta} y_n(y_n-\theta)^{n-1}\ dy_n \\
&= \frac{n}{\theta^n}\int_0^{\theta}(z+\theta)z^{n-1}\ dz, \qquad\qquad \text{let } z=y_n-\theta, dz=dy_n \\
&= \frac{n}{\theta^n}\int_0^{\theta}(z^n+\theta z^{n-1})\ dz \\
&= \frac{n}{\theta^n}\left[\frac{1}{n+1}\ z^{n+1}+\frac{\theta}{n}\ z^n\right]_0^\theta \\
&= \frac{n}{\theta^n}\left[\frac{\theta^{n+1}}{n+1}+\frac{\theta^{n+1}}{n}\right] \\
&= \frac{n\theta}{n+1}+\theta \\
&= \frac{2n+1}{n+1}\ \theta \\
\Rightarrow\ U_b &= \frac{n+1}{2n+1}Y_n
\end{aligned}
$$

(d) (**9 marks**) Define the unbiased estimators of θ in parts (b) and (c) as U_a and U_b, respectively. Find a constant k so that the unbiased estimator, $kU_a+(1-k)U_b$, has the smallest variance. What is the variance of this unbiased estimator?

Answer

$$
\begin{aligned}
E(Y_1^2) &= \int_\theta^{2\theta} y_1^2 f_{Y_1}(y_1)\ dy_1 \\
&= \frac{n}{\theta}\int_\theta^{2\theta} y_1^2\left(2-\frac{y_1}{\theta}\right)^{n-1}\ dy_1 \\
&= \frac{n}{\theta^n}\int_\theta^{2\theta} y_1^2(2\theta-y_1)^{n-1}\ dy_1 \\
&= \frac{n}{\theta^n}\int_\theta^{0} -(2\theta-z)^2 z^{n-1}\ dz, \qquad\qquad \text{let } z=2\theta-y_1, dz=-dy_1 \\
&= \frac{n}{\theta^n}\int_0^{\theta}(z^{n+1}-4\theta z^n+4\theta^2 z^{n-1})\ dz \\
&= \frac{n}{\theta^n}\left[\frac{1}{n+2}\ z^{n+2}-\frac{4\theta}{n+1}\ z^{n+1}+\frac{4\theta^2}{n}\ z^n\right]_0^\theta \\
&= \frac{n}{\theta^n}\left[\frac{\theta^{n+2}}{n+2}-\frac{4\theta^{n+2}}{n+1}+\frac{4\theta^{n+2}}{n}\right] \\
&= \left[\frac{n}{n+2}-\frac{4n}{n+1}+4\right]\theta^2
\end{aligned}
$$

$$
\begin{aligned}
Var(Y_1) &= E(Y_1^2) - [E(Y_1)]^2 \\
&= \left(\frac{n}{n+2} - \frac{4n}{n+1} + 4\right)\theta^2 + \left[\left(\frac{n+2}{n+1}\right)\ \theta\right]^2 \\
&= \left(\frac{n}{n+2} - \frac{4n}{n+1} + 4 - \left(1 + \frac{1}{n+1}\right)^2\right)\theta^2 \\
&= \left(\frac{n}{n+2} - \frac{4n}{n+1} + 4 - 1 - \frac{2}{n+1} - \frac{1}{(n+1)^2}\right)\theta^2 \\
&= \left(\frac{n}{n+2} - \frac{4n}{n+1} + 3 - \frac{1}{(n+1)^2}\right)\theta^2 \\
&= \frac{n\theta^2}{(n+1)^2(n+2)}
\end{aligned}
$$

$$
\begin{aligned}
E(Y_n^2) &= \int_\theta^{2\theta} y_n^2 f_{Y_n}(y_n)\ dy_n \\
&= \frac{n}{\theta}\int_\theta^{2\theta} y_n^2 \left(\frac{y_n}{\theta} - 1\right)^{n-1}\ dy_n \\
&= \frac{n}{\theta^n}\int_\theta^{2\theta} y_n^2 (y_n - \theta)^{n-1}\ dy_n \\
&= \frac{n}{\theta^n}\int_0^{\theta} (z+\theta)^2 z^{n-1}\ dz, \qquad\qquad \text{let } z = y_n - \theta, dz = dy_n \\
&= \frac{n}{\theta^n}\int_0^{\theta} (z^{n+1} + 2\theta z^n + \theta^2 z^{n-1})\ dz \\
&= \frac{n}{\theta^n}\left[\frac{1}{n+2}\ z^{n+2} + \frac{2\theta}{n+1}\ z^{n+1} + \frac{\theta^2}{n}\ z^n\right]_0^\theta \\
&= \left[\frac{n}{n+2} + \frac{2n}{n+1} + 1\right]\theta^2
\end{aligned}
$$

$$
\begin{aligned}
Var(Y_n) &= E(Y_n^2) - [E(Y_n)]^2 \\
&= \left(\frac{n}{n+2} + \frac{2n}{n+1} + 1\right)\theta^2 - \left(\frac{2n+1}{n+1}\ \theta\right)^2 \\
&= \left(\frac{n}{n+2} + \frac{2n}{n+1} + 1 - \left(2 - \frac{1}{n+1}\right)^2\right)\theta^2 \\
&= \left(\frac{n}{n+2} + \frac{2n}{n+1} + 1 - 4 + \frac{4}{n+1} - \frac{1}{(n+1)^2}\right)\theta^2 \\
&= \left(\frac{n}{n+2} + \frac{2n+4}{n+1} - 3 - \frac{1}{(n+1)^2}\right)\theta^2 \\
&= \frac{n\theta^2}{(n+1)^2(n+2)}
\end{aligned}
$$

$$
Cov(Y_1, Y_n) = \frac{\theta^2}{(n+1)^2(n+2)}
$$

$$
\begin{aligned}
\Rightarrow Var(U_a) &= \left(\frac{n+1}{n+2}\right)^2 Var(Y_1) \\
&= \left(\frac{n+1}{n+2}\right)^2 \cdot \frac{n\theta^2}{(n+1)^2(n+2)} \\
Var(U_b) &= \left(\frac{n+1}{2n+1}\right)^2 Var(Y_n) \\
&= \left(\frac{n+1}{2n+1}\right)^2 \cdot \frac{n\theta^2}{(n+1)^2(n+2)} \\
Cov(U_a, U_b) &= \frac{(n+1)^2}{(n+2)(2n+1)} Cov(Y_1, Y_n) \\
&= \frac{(n+1)^2}{(n+2)(2n+1)} \cdot \frac{\theta^2}{(n+1)^2(n+2)}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& k^2 Var(U_a) + (1-k)^2 Var(U_b) + 2k(1-k)Cov(U_a, U_b) \\
= & \left\{ \frac{k^2 n}{(n+2)^2} + \frac{(1-k)^2 n}{(2n+1)^2} + \frac{2k(1-k)}{(n+2)(2n+1)} \right\} \cdot \frac{(n+1)^2\theta^2}{(n+1)^2(n+2)} \\
\Rightarrow \quad & \frac{kn}{(n+2)^2} - \frac{(1-k)n}{(2n+1)^2} + \frac{1-2k}{(n+2)(2n+1)} = 0 \\
\Rightarrow \quad & k = \frac{n+2}{5n+4}
\end{aligned}
$$

Therefore, the unbiased estimator has the smallest variance is

$$
\begin{aligned}
kU_a + (1-k)U_b &= \frac{n+2}{5n+4} U_a + \frac{2(2n+1)}{5n+4} U_b \\
&= \frac{n+1}{5n+4} Y_1 + \frac{2(n+1)}{5n+4} Y_n \\
&= \frac{n+1}{5n+4}(Y_1 + 2Y_n)
\end{aligned}
$$

$$
\Rightarrow Var\left(\frac{n+1}{5n+4}(Y_1 + 2Y_n)\right) = \frac{\theta^2}{(5n+4)(n+2)}
$$

(e) (**3 marks**) Does the UMVUE for θ exist? If yes, find it; if no, explain in details.

<u>Answer</u>

No. UMVUE for θ doesn't exist.

$$
E\left(\frac{n+1}{n+2} Y_1\right) = \theta \;\; \& \;\; E\left(\frac{n+1}{2n+1} Y_n\right) = \theta
$$

$$
\Rightarrow \quad E\left(\frac{n+1}{n+2} Y_1 - \frac{n+1}{2n+1} Y_n\right) = 0
$$

$$
\text{But} \quad \frac{n+1}{n+2} Y_1 - \frac{n+1}{2n+1} Y_n \neq 0
$$

Therefore, (Y_1, Y_n) is not complete.

3. Individuals were classified according to gender and according to whether or not they were color-blind as follows:

	Male	Female
Normal	x_{11}	x_{12}
Color-blind	x_{21}	x_{22}

Let $X = (X_{11}, X_{12}, X_{21}, X_{22}) \sim$ multinomial $(n, P_{11}, P_{12}, P_{21}, P_{22})$.

(a) Test the hypothesis $H_0 : P_{11} = \frac{p}{2}, P_{12} = \frac{p^2}{2} + pq, P_{21} = \frac{q}{2}, P_{22} = \frac{q^2}{2}$, where $q = 1 - p$, against $H_1 : (P_{11}, P_{12}, P_{21}, P_{22})$ takes any other value in $[0, 1]^4$ at the level of significance α.

i. (**4 marks**) Find the likelihood ratio statistic and then derive the approximate large sample likelihood ratio test.

Answer

The likelihood function is $L(P_{11}, P_{12}, P_{21}, P_{22}) = \text{constant} \cdot \prod_{i=1}^{2}\prod_{i=1}^{2} P_{ij}^{x_{ij}}$

Under H_0,

$$\begin{aligned} L_o &= \text{constant} \times \left(\frac{p}{2}\right)^{x_{11}} \left(\frac{p^2}{2} + pq\right)^{x_{12}} \left(\frac{q}{2}\right)^{x_{21}} \left(\frac{q^2}{2}\right)^{x_{22}} \\ &= \text{constant} \times p^{x_{11}}(p^2 + 2pq)^{x_{12}} q^{x_{21}} q^{2x_{22}} \end{aligned}$$

$$\begin{aligned} \ln(L_o) &= \text{constant} \times x_{11}\ln(p) + x_{12}\ln(p^2 + 2p(1-p)) + (x_{21} + 2\ x_{22})\ln(1-p) \\ &= \text{constant} \times (x_{11} + x_{12})\ln(p) + x_{12}\ln(2-p) + (x_{21} + 2\ x_{22})\ln(1-p) \end{aligned}$$

$$\begin{aligned} \frac{\partial \ln(L_o)}{\partial p} &= \frac{x_{11} + x_{12}}{p} - \frac{x_{12}}{2-p} - \frac{x_{21} + 2x_{22}}{1-p} = 0 \\ \Rightarrow &\quad (x_{11} + 2x_{12} + x_{21} + 2x_{22})p^2 - (3x_{11} + 2x_{21} + 4(x_{12} + x_{22}))p + 2(x_{11} + x_{12}) = 0 \end{aligned}$$

Let $x_1 = x_{11} + x_{12}$, $x_2 = x_{21} + 2x_{22}$

$$\Rightarrow (x_1 + x_2 + x_{12})p^2 - (3x_1 + x_{12} + 2x_2)p + 2x_1 = 0$$

$$\Rightarrow \hat{p} = \frac{(3x_1 + x_{12} + 2x_2) - \sqrt{(x_1 + x_{12} + 2x_2)^2 - 4x_1x_{12}}}{2(x_1 + x_2 + x_{12})}$$

The numerator of the likelihood ratio

$$\sup\{L(\theta, \mathbf{x}) : \theta \in \Theta_o\} = \text{constant } \times \prod_{i=1}^{2}\prod_{j=1}^{2} \hat{P}_{ij}^{x_{ij}}$$

where $\hat{P}_{11} = \frac{\hat{p}}{2}, \hat{P}_{12} = \frac{\hat{p}^2}{2} + \hat{p}(1-\hat{p}), \hat{P}_{21} = \frac{1-\hat{p}}{2}, \hat{P}_{22} = \frac{(1-\hat{p})^2}{2}$

The denominator of the likelihood ratio involves finding the MLE for θ

$$\sup\{L(\theta, \mathbf{x}) : \theta \in \Theta\} = \text{constant } \times \prod_{i=1}^{2}\prod_{j=1}^{2} \left(\frac{x_{ij}}{n}\right)^{x_{ij}}$$

$$\Rightarrow \quad \lambda(\mathbf{x}) = \prod_{i=1}^{2}\prod_{j=1}^{2}\left(\frac{n\hat{P}_{ij}}{x_{ij}}\right)^{x_{ij}}$$

The approximate large sample likelihood ratio test

$$C_1 \;=\; \left\{\mathbf{x} : 2\sum_{i=1}^{2}\sum_{j=1}^{2} x_{ij}\log\frac{x_{ij}}{n\hat{P}_{ij}} \geq \chi^2_\alpha(2)\right\} \qquad \text{for large} \quad n$$

ii. (**2 marks**) Write down the Pearson's goodness of fit test statistic and state the critical region for this test.

Answer

$$C_1 \;=\; \left\{\mathbf{x} : G = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(x_{ij} - n\hat{P}_{ij})^2}{n\hat{P}_{ij}} \geq \chi^2_\alpha(2)\right\} \qquad \text{for large} \quad n$$

(b) Suppose $x_{11} = 442, x_{12} = 514, x_{21} = 38, x_{22} = 6$. Perform the following tests at $\alpha = 0.05$. State clearly the hypothesis statements, value of test statistic, critical value and your conclusion for each test.

i. (**6 marks**) Test whether the null hypothesis $H_0 : P_{11} = \frac{p}{2}, P_{12} = \frac{p^2}{2} + pq, P_{21} = \frac{q}{2}, P_{22} = \frac{q^2}{2}$ is true by the two tests derived above.

Answer

Since $x_1 = 442 + 514 = 956$, $x_2 = 38 + 2*6 = 50$ and $x_{12} = 514$

$$\begin{aligned}\hat{p} &= \frac{(3x_1 + x_{12} + 2x_2) - \sqrt{(x_1 + x_{12} + 2x_2)^2 - 4x_1x_{12}}}{2(x_1 + x_2 + x_{12})} \\ &= \frac{3482 - 2\sqrt{124841}}{3040} \\ &= 0.912942\end{aligned}$$

	Male	Female
Normal	442 (456.471)	514 (496.210)
Color-blind	38 (43.529)	6 (3.78955)

The test statistics of tests derived above

$$\begin{aligned}2\sum_{i=1}^{2}\sum_{j=1}^{2} x_{ij}\log\frac{x_{ij}}{n\hat{P}_{ij}} &= 2.92128 \\ G = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(x_{ij} - n\hat{P}_{ij})^2}{n\hat{P}_{ij}} &= 3.08818\end{aligned}$$

Since both test statistics are smaller than $\chi^2_\alpha(2) = 5.991$, H_0 can't be rejected.

ii. (**4 marks**) Test the hypothesis that color blindness is independent of gender. No need to make the Yates's Correction.

Answer

	Male	Female	
Normal	442	514	956
Color-blind	38	6	44
	480	520	1000

$$
\begin{aligned}
G &= \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(x_{ij} - \frac{a_i b_j}{n})^2}{\frac{a_i b_j}{n}} \\
&= n\left(\sum_{i=1}^{2}\sum_{j=1}^{2} \frac{x_{ij}^2}{a_i b_j} - 1\right) \\
&= 27.1387 \ > \ \chi_0^2.05(1) = 3.841
\end{aligned}
$$

H_0 is rejected and conclude that color blindness and gender are not independent.

iii. (**4 marks**) Test whether the probabilities of color-blind individuals for male and female are equal by z test. No need to make the continuity correction.

Answer

Let P_1 and P_2 be the probabilities of color-blind individuals for male and female, respectively. Let P be the probability of color-blind individuals under H_0.

Then $\hat{P}_1 = \frac{38}{442+38}$, $\hat{P}_2 = \frac{6}{514+6}$ and $\hat{P} = \frac{38+6}{442+514+38+6}$.

$$
\begin{aligned}
z &= \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})(\frac{1}{n_1}+\frac{1}{n_2})}} \\
&= \frac{\frac{38}{442+38} - \frac{6}{514+6}}{\sqrt{\frac{38+6}{442+514+38+6}(1-\frac{38+6}{442+514+38+6})(\frac{1}{480}+\frac{1}{520})}} \\
&= 5.20949 \ > \ z_{0.025} = 1.96
\end{aligned}
$$

H_0 is rejected and conclude that the probabilities of color-blind individuals for male and female are not equal.

4. If $X_1, X_2, \ldots, X_n$ are independently and normally distributed with the same unknown mean μ but different known variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$.

(a) **(4 marks)** Find the maximum likelihood estimator of μ. Hence, find its distribution.

Answer

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\{-\frac{(x_i-\mu)^2}{2\sigma_i^2}\} \\
&= (2\pi)^{-n/2} \prod_{i=1}^{n} (\sigma_i^2)^{-1/2} \exp\{-\frac{1}{2}\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{\sigma_i^2}\} \\
\log L(\mu) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\log(\sigma_i^2) - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma_i^2}
\end{aligned}
$$

$$
\frac{\partial \log L(\mu)}{\partial \mu} = 0 \Rightarrow \hat{\mu} = \frac{\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}}
$$

$$
\begin{aligned}
X_i &\sim N(\mu, \sigma_i^2) \\
\frac{X_i}{\sigma_i^2} &\sim N\left(\frac{\mu}{\sigma_i^2}, \frac{1}{\sigma_i^2}\right) \\
\sum_{i=1}^{n} \frac{X_i}{\sigma_i^2} &\sim N\left(\mu\sum_{i=1}^{n}\frac{1}{\sigma_i^2}, \sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right) \\
\hat{\mu} = \frac{\sum_{i=1}^{n}\frac{X_i}{\sigma_i^2}}{\sum_{j=1}^{n}\frac{1}{\sigma_j^2}} &\sim N\left(\mu, \frac{1}{\sum_{j=1}^{n}\frac{1}{\sigma_j^2}}\right)
\end{aligned}
$$

(b) **(6 marks)** Construct the UMP test for testing $H_0 : \mu \le \mu_0$ against $H_1 : \mu > \mu_0$ at a significance level of α.

Answer

The MP test for testing $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1$, where $\mu_1 > \mu_0$ has the critical region

$$
C_1 = \left\{\mathbf{x} : \frac{f_{\mathbf{X}}(\mathbf{x}, \mu_0)}{f_{\mathbf{X}}(\mathbf{x}, \mu_1)} \le k\right\}
$$

$$
\begin{aligned}
\frac{f_{\mathbf{X}}(\mathbf{x}, \mu_0)}{f_{\mathbf{X}}(\mathbf{x}, \mu_1)} &= \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n}\frac{(x_i-\mu_0)^2}{\sigma_i^2} - \sum_{i=1}^{n}\frac{(x_i-\mu_1)^2}{\sigma_i^2}\right]\right\} \\
&= \exp\left\{\frac{1}{2}\left[\sum_{i=1}^{n}\frac{\mu_1^2-\mu_0^2-2x_i(\mu_1-\mu_0)}{\sigma_i^2}\right]\right\} \le k
\end{aligned}
$$

Take logarithm,

$$\begin{aligned}
\frac{1}{2}\left[\sum_{i=1}^{n}\frac{\mu_1^2-\mu_0^2-2x_i(\mu_1-\mu_0)}{\sigma_i^2}\right] &\leq \log k \\
\Longrightarrow \quad \frac{\mu_1-\mu_0}{2}\left[\frac{n(\mu_1+\mu_0)}{\sigma_i^2}-2\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}\right] &\leq \log k \\
\Longrightarrow \quad \frac{\displaystyle\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}}{\displaystyle\sum_{j=1}^{n}\frac{1}{\sigma_j^2}} &\geq K \\
\Longrightarrow \quad C_1 = \left\{\mathbf{x}:\frac{\displaystyle\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}}{\displaystyle\sum_{j=1}^{n}\frac{1}{\sigma_j^2}} \geq K\right\}
\end{aligned}$$

Since $\frac{\sum_{i=1}^{n}\frac{X_i}{\sigma_i^2}}{\sum_{j=1}^{n}\frac{1}{\sigma_j^2}} \sim N\left(\mu, \frac{1}{\sum_{j=1}^{n}\frac{1}{\sigma_j^2}}\right)$

$$\begin{aligned}
Pr\left(\frac{\sum_{i=1}^{n}\frac{X_i}{\sigma_i^2}}{\sum_{j=1}^{n}\frac{1}{\sigma_j^2}} \geq K;\mu_0\right) &= \alpha \\
\Longrightarrow \quad \frac{K-\mu_0}{\sqrt{\frac{1}{\sum_{j=1}^{n}\frac{1}{\sigma_j^2}}}} &= z_\alpha \\
\Longrightarrow \quad K = \mu_0 + z_\alpha\sqrt{\frac{1}{\sum_{j=1}^{n}\frac{1}{\sigma_j^2}}}
\end{aligned}$$

$$\therefore C_1 = \left\{\mathbf{x}:\frac{\displaystyle\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}}{\displaystyle\sum_{j=1}^{n}\frac{1}{\sigma_j^2}} \geq \mu_0 + z_\alpha\sqrt{\frac{1}{\sum_{j=1}^{n}\frac{1}{\sigma_j^2}}}\right\}$$

Since the critical value of C_1 doesn't depend on the value of μ under the alternative hypothesis, C_1 is the UMP test for testing $H_0:\mu=\mu_0$ against $H_1:\mu>\mu_0$.

$$
\begin{aligned}
& \sup\{Pr(\mathbf{X} \in C_1) : \mu \in \Theta_0\} \\
= \;& \sup\left\{ Pr\left(\frac{\sum_{i=1}^{n} \frac{X_i}{\sigma_i^2}}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}} \geq \mu_0 + z_\alpha \sqrt{\frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}}} \right) : \mu \leq \mu_0 \right\} \\
= \;& \sup\left\{ Pr\left(Z \geq \frac{\mu_0 + z_\alpha \sqrt{\frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}}} - \mu}{\sqrt{\frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}}}} \right) : \mu \leq \mu_0 \right\} \\
= \;& \sup\left\{ Pr\left(Z \geq z_\alpha \; + \; \frac{\mu_0 - \mu}{\sqrt{\frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}}}} \right) : \mu \leq \mu_0 \right\} \\
= \;& \alpha \qquad \text{when } \mu = \mu_0
\end{aligned}
$$

Thus, C_1 is the UMP test for testing $H_0 : \mu \leq \mu_0$ against $H_1 : \mu > \mu_0$.

(c) **(2 marks)** Based on the test in part (b), calculate the power of test at $\mu_1 = 1$, where $\mu_1 \in \Theta_1$, when $\alpha = 0.05$, $\mu_0 = 0$, $n = 10$, $\sigma_1^2 = \ldots = \sigma_5^2 = 1$ and $\sigma_6^2 = \ldots = \sigma_{10}^2 = 2$. Round the value to two decimal places before finding the probability.

Answer

$$
\begin{aligned}
& Pr(\mathbf{X} \in C_1 : \mu = \mu_1) \\
= \;& Pr\left(\frac{\sum_{i=1}^{n} \frac{X_i}{\sigma_i^2}}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}} \geq \mu_0 + z_\alpha \sqrt{\frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}}} : \mu = \mu_1 \right) \\
= \;& Pr\left(Z \geq \frac{\mu_0 + z_\alpha \sqrt{\frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}}} - \mu_1}{\sqrt{\frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}}}} \right) \\
= \;& Pr\left(Z \geq z_\alpha \; + \; \frac{\mu_0 - \mu_1}{\sqrt{\frac{1}{\sum_{j=1}^{n} \frac{1}{\sigma_j^2}}}} \right) \\
= \;& Pr\left(Z \geq -1.09361\right) \\
= \;& 0.8621
\end{aligned}
$$

(d) Assuming that $\mu = 0$ and all σ_j^2, for $j = 1, \dots, n$, are equal to σ^2 but unknown, consider another hypothesis testing problem with $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 \neq \sigma_0^2$ at the level of significance α.

i. (**4 marks**) Find the expression of the likelihood ratio statistic.

Answer

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{x_i^2}{2\sigma^2}\} \\
&= (2\pi\sigma^2)^{-n/2} \exp\{-\sum_{i=1}^{n} \frac{x_i^2}{2\sigma^2}\} \\
\log L(\sigma^2) &= -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{x_i^2}{2\sigma^2}
\end{aligned}
$$

Under H_a,

$$
\begin{aligned}
\log L(\sigma^2) &= -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{x_i^2}{2\sigma^2} \\
\frac{\partial \log L(\sigma^2)}{\partial \sigma^2} &= 0n \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = S_n^2
\end{aligned}
$$

$$
\begin{aligned}
\lambda(\mathbf{x}) &= \frac{\sup\{L(\theta, \mathbf{x}) : \theta \in \Theta_o\}}{\sup\{L(\theta, \mathbf{x}) : \theta \in \Theta\}} \\
&= \frac{(2\pi\sigma_0^2)^{-n/2} \exp\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n} x_i^2\}}{(2\pi\hat{\sigma}^2)^{-n/2} \exp\{-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n} x_i^2\}\}} \\
&= \left(\frac{S_n^2}{\sigma_0^2}\right)^{n/2} \exp\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n} x_i^2 + \frac{n}{2}\}
\end{aligned}
$$

ii. (**4 marks**) Hence, derive the exact likelihood ratio test at the significance level of α.

<u>Answer</u>

$$
\begin{aligned}
\lambda(\mathbf{x}) &\leq K \\
\Rightarrow \left(\frac{\sum_{i=1}^n x_i^2}{n\sigma_0^2}\right)^{n/2} \exp\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n x_i^2 + \frac{n}{2}\} &\leq K \\
\Rightarrow \left(\frac{\sum_{i=1}^n x_i^2}{\sigma_0^2}\right)^{n/2} \exp\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n x_i^2\} &\leq K
\end{aligned}
$$

Let $y = \dfrac{\sum_{i=1}^n x_i^2}{\sigma_0^2}$. Consider

$$
\begin{aligned}
g(y) &= x^{n/2}\exp\{-y/2\} \\
g'(y) &= \frac{n}{2}x^{\frac{n}{2}-1}e^{-\frac{n}{2}} \;-\; \frac{1}{2}x^{\frac{n}{2}}e^{-\frac{n}{2}} \\
&= x^{\frac{n}{2}-1}e^{-\frac{n}{2}}(\frac{n}{2}-\frac{x}{2}) \\
g'(y) = 0 &\Rightarrow x = n \\
&\Rightarrow g'(x) > 0 \text{ when } x < n \\
& g'(x) < 0 \text{ when } x > n
\end{aligned}
$$

$\therefore$ $g(x)$ attain maximum at $x = n$, decrease when $x > n$ and increase when $x < n$.

$$
\begin{aligned}
g(y) \leq K &\Rightarrow x \leq k_1 \text{ or } x \geq k_2 \\
\lambda(\mathbf{x}) \leq K &\Rightarrow \frac{\sum_{i=1}^n x_i^2}{\sigma_0^2} \leq k_1 \text{ or } \frac{\sum_{i=1}^n x_i^2}{\sigma_0^2} \geq k_2
\end{aligned}
$$

Under H_0, $\dfrac{\sum_{i=1}^n x_i^2}{\sigma_0^2} \sim \chi^2(n)$ and $Pr\left(\dfrac{\sum_{i=1}^n x_i^2}{\sigma_0^2} \leq k_1\right) = Pr\left(\dfrac{\sum_{i=1}^n x_i^2}{\sigma_0^2} \leq k_1\right) = \frac{\alpha}{2}$
$\Rightarrow k_1 = \chi^2_{1-\frac{\alpha}{2}}(n), \quad k_2 = \chi^2_{\frac{\alpha}{2}}(n)$

5. (**Bonus: 10 marks**) Consider a random sample of a fixed size n, $\{X_1, \ldots, X_n\}$, from a p.m.f. given by

$$p_{-1} = P(X_i = -1) = \frac{1-\theta}{2}, \quad p_0 = P(X_i = 0) = \frac{1}{2}, \quad p_1 = P(X_i = 1) = \frac{\theta}{2},$$

where $0 \le \theta \le 1$. Define $n_{-1} = \sum_{i=1}^n I_{\{X_i=-1\}}$, $n_0 = \sum_{i=1}^n I_{\{X_i=0\}}$, and $n_1 = \sum_{i=1}^n I_{\{X_i=1\}}$. Given that $(n_{-1}, n_0, n_1) \sim$ multinomial(n, p_{-1}, p_0, p_1).

Find the maximum likelihood estimator, $\hat{\theta}$, for θ. Find $E(\hat{\theta})$. Hence or otherwise, find an unbiased estimator for θ

Answer

Since

$$\ln L(\theta) = \ln c + n_{-1}\ln(1-\theta) - n_{-1}\ln 2 - n_0 \ln 2 + n_1 \ln\theta - n_1 \ln 2,$$

$$0 = \frac{d}{d\theta}\ln L(\theta)\Big|_{\theta=\hat{\theta}} = -\frac{n_{-1}}{1-\hat{\theta}} + \frac{n_1}{\hat{\theta}} \implies \hat{\theta} = \frac{n_1}{n_1+n_{-1}} \in [0,1],$$

which is a unique critical point and inside the parameter space of θ.

Also,

$$\frac{d^2}{d\theta^2}\ln L(\hat{\theta}) = -\frac{n_{-1}}{(1-\hat{\theta})^2} - \frac{n_1}{\hat{\theta}^2} < 0.$$

Hence, $\hat{\theta}$ is a MLE of θ.

Note that

$$\begin{aligned} E(\hat{\theta}) = E\Big(\frac{n_1}{n_1+n_{-1}}\Big) &= E_{n_1+n_{-1}}\Big[E\Big(\frac{n_1}{n_1+n_{-1}}\Big| n_1+n_{-1}\Big)\Big] \\ &= E_{n_1+n_{-1}}\Big[\frac{1}{n_1+n_{-1}}E\Big(n_1\Big|n_1+n_{-1}\Big)\Big]. \end{aligned}$$

Now we consider the conditional pmf of n_1 given n_1+n_{-1}.

Since $n_0 = \sum_{i=1}^n I_{\{X_i=0\}} \sim Bin(n, p_0)$,

$$n_{-1} + n_1 = n - n_0 \sim Bin(n, 1-p_0 = 1/2)$$

because $n - n_0 = n - \sum_{i=1}^n I_{\{X_i=0\}} = \sum_{i=1}^n [1 - I_{\{X_i=0\}}] = \sum_{i=1}^n I_{\{X_i\neq 0\}} \sim Bin(n, EI_{\{X_i\neq 0\}} = P(X_i \neq 0) = 1 - P(X_i = 0))$.

Since

$$\begin{aligned} P(n_1 = t|\, n_1+n_{-1} = s) &= \frac{P(n_1 = t, n_1+n_{-1} = s)}{P(n_1+n_{-1} = s)} \\ &= \frac{P(n_1 = t, n_{-1} = s-t, n_0 = n-s)}{P(n_1+n_{-1} = s)} \\ &= \frac{\binom{n}{s-t,\, n-s,\, t} p_{-1}^{s-t} p_0^{n-s} p_1^t}{\binom{n}{s}(\frac{1}{2})^s(\frac{1}{2})^{n-s}} \\ &= \frac{s!}{t!(s-t)!}\theta^t(1-\theta)^{s-t}. \end{aligned}$$

Thus, $n_1|n_1 + n_{-1} \sim Bin(n_1 + n_{-1}, \theta)$.

Hence, $E\Big(n_1\Big|n_1 + n_{-1}\Big) = (n_1 + n_{-1})\theta$ and $E(\hat{\theta}) = E_{n_1+n_{-1}}\Big[\dfrac{1}{n_1 + n_{-1}}(n_1 + n_{-1})\theta\Big] = \theta$.

********** **END** **********