Day 1, Problem 1.

Selected Qualifying problems —by Ou Zhao

1. Let X_1 and X_2 be independent discrete random variables with common mass function

$$P(X_i = x) = -\frac{\theta^x}{x\log(1-\theta)}, \quad x = 1, 2, \ldots,$$

where $\theta \in (0,1)$.

(a) Find the mean and variance of X_1.

(b) Let $I(\cdot)$ denote the usual indicator function. Show that $U(\mathbf{X}) = -I(X_1 = 1)$ is an unbiased estimator of $\tau(\theta) = \theta/\log(1-\theta)$.

(c) Argue that $T(\mathbf{X}) = T = X_1 + X_2$ is a complete sufficient statistic for θ.

(d) Find the uniform minimum variance unbiased estimator (UMVUE) of $\tau(\theta)$.

Solution:
(a) For $i = 1, 2$,

$$E[X_i] = \sum_{x=1}^{\infty} x\frac{\theta^x}{x\log(1-\theta)} = -\frac{1}{\log(1-\theta)}\sum_{x=1}^{\infty}\theta^x = -\frac{1}{\log(1-\theta)}\frac{\theta}{1-\theta}.$$

and similarly,

$$E[X_i^2] = -\frac{1}{\log(1-\theta)}\sum_{x=1}^{\infty} x\theta^x = -\frac{1}{\log(1-\theta)}\frac{\theta}{(1-\theta)^2}$$

then

$$\mathrm{Var}[X_1] = E[X_1^2] - (E[X_1])^2 = -\frac{1}{\log(1-\theta)}\frac{\theta}{(1-\theta)^2}\left(1 + \frac{\theta}{\log(1-\theta)}\right).$$

(b) It is unbiased because

$$E[U(\mathbf{X})] = E[-I(X_1 = 1)] = -P(X_1 = 1) = \frac{\theta}{\log(1-\theta)}.$$

(c)

Let $X = (X_1, X_2)$, then the joint density is given by

$$p_\theta(x) = P(X_1 = x_1, X_2 = x_2) = \frac{\theta^{x_1+x_2}}{x_1 x_2 \log^2(1-\theta)} = \frac{1}{x_1 x_2} \frac{1}{\log^2(1-\theta)} \exp\{(x_1+x_2)\log\theta\}$$

Let $T(X) = X_1 + X_2$, then we know T is complete and sufficient because $p_\theta(x)$ is an exponential family with full rank.

(d) Next, let us consider the density for T. For any $t = 2, 3, \ldots$

$$\begin{aligned}
P_T(t) = P(T(X) = t) &= E_\theta[P(X_1 + X_2 = t|X_2)] \\
&= \sum_{x_2=1}^{t-1} P(X_1 + X_2 = t|X_2 = x_2)P(X_2 = x_2) \\
&= \sum_{x_2=1}^{t-1} \frac{P(X_1 = t - x_2, X_2 = x_2)}{P(X_2 = x_2)} P(X_2 = x_2) \\
&= \sum_{x_2=1}^{t-1} \frac{\theta^t}{\log^2(1-\theta)} \frac{1}{x_2(t-x_2)} \\
&= \frac{\theta^t}{\log^2(1-\theta)} \frac{2}{t} \left(\sum_{i=1}^{t-1} \frac{1}{i} \right)
\end{aligned}$$

So the UMVU of $\theta/\log(1-\theta)$ based on T is given by

$$\delta(T) = E_\theta(\eta(X)|T).$$

When $T = t$ for $t = 2, 3, \ldots,$

$$\begin{aligned}
\delta(t) &= E_\theta(\eta(X)|X_1 + X_2 = t) \\
&= -P(X_1 = 1|X_1 + X_2 = t) \\
&= -\frac{P(X_1 = 1)P(X_2 = t-1)}{P(T = t)} \\
&= \frac{t}{2(1-t)(\sum_{i=1}^{t-1} \frac{1}{i})}.
\end{aligned}$$

so

$$\delta(T) = \frac{T}{2(1-T)(\sum_{i=1}^{T-1} \frac{1}{i})}$$

is the UMVUE of $\tau(\theta) = \theta/\log(1-\theta)$.

~~Problem 1~~ Day 1 Prob 2

Let A_k be the event of a match at the k^{th} draw, $k = 1, 2, \ldots, N$. If A is the event of a match, then

$$A = \bigcup_{k=1}^{N} A_k.$$

We seek $P(A)$.

If <u>with</u> replacement: A_k's are independent and $P(A_k) = \frac{N}{N}\frac{1}{N} = \frac{1}{N}$. Therefore,

$$P(A) = P\left(\bigcup_1^N A_k\right) = 1 - P\left(\bigcap_1^N A_k^c\right)$$

$$= 1 - \prod_1^N P(A_k^c)$$

$$= 1 - \prod_1^N \left[1 - \frac{1}{N}\right]$$

$$= 1 - \left(1 - \frac{1}{N}\right)^N$$

plug-in $N=5$

So as $N \to \infty$,

$$\to 1 - e^{-1}.$$

if without replacement: A_k's will not be independent. Need to use I-E Principle.

$$P(A) = P\left(\bigcup_{k=1}^{N} A_k\right)$$

$$= \sum_{1}^{N} P(A_k) - \sum_{i_1<i_2} P(A_{i_1}A_{i_2})$$

$$+ \sum_{i_1<i_2<i_3} P(A_{i_1}A_{i_2}A_{i_3}) - \cdots$$

$$+ (-1)^{N+1} P\left(\bigcap_{1}^{N} A_i\right).$$

Now,

$$P(A_k) = \frac{\binom{N}{1}(N-1)!\,(1)\,(N-1)!}{N!\,N!} = \frac{1}{N}.$$

$$P(A_{i_1}A_{i_2}) = \frac{\binom{N}{2}(N-2)!\,2!\,(1)(1)(N-2)!}{N!\,N!}$$

$$= \frac{\binom{N}{2}}{\binom{N}{2}} \frac{1}{N(N-1)} = \frac{1}{N(N-1)} = \frac{(N-2)!}{N!}$$

$$P(A_{i_1}A_{i_2}A_{i_3}) = \frac{\binom{N}{3}(N-3)!\,3!\,(1)(1)(1)(N-3)!}{N!\,N!} = \frac{(N-3)!}{N!}$$

etc.

$$P\left(\begin{matrix}\text{intersections}\\ \text{of } k\ A_i\text{'s}\end{matrix}\right) = \frac{(N-k)!}{N!}$$

$\vdots$

Therefore,

$$P(A) = N \cdot \frac{1}{N} - \binom{N}{2}\frac{(N-2)!}{N!} + \binom{N}{3}\frac{(N-3)!}{N!}$$

$$- \binom{N}{4}\frac{(N-4)!}{N!} + \cdots$$

$$+ (-1)^{N+1}\frac{1}{N!}$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \frac{1}{5!} - \cdots$$

$$+ (-1)^{N+1}\frac{1}{N!}$$

As $N \to \infty$,

$$\longrightarrow 1 - e^{-1}.$$

Same limiting value!

Day 1. Problem 3.

3.

(a) Define $\nu = \log\theta$. Then $\nu \sim N(\mu_0, \sigma_0^2)$. It follows that

$$\pi(\nu \mid \underset{\sim}{X}) = \frac{f(\underset{\sim}{X} \mid \nu)\,\pi(\nu)}{\int f(\underset{\sim}{X} \mid \nu)\,\pi(\nu)\,d\nu},$$

where the numerator is equal to

$$e^{-n\nu} I(0 < X_{(1)} < X_{(n)} < e^{\nu}) \times \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2\sigma_0^2}(\nu-\mu_0)^2\right\}$$

$$= \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{1}{2\sigma_0^2}\left\{\nu^2 - 2\sigma_0^2\left(\frac{\mu_0}{\sigma_0^2} - n\right)\nu\right\}\right] e^{-\frac{\mu_0^2}{2\sigma_0^2}}\, I(\nu > \log X_{(n)})\, I(X_{(1)} > 0)$$

$$= \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2\sigma_0^2}(\nu - \mu_0 + n\sigma_0^2)^2\right\} \cdot \exp\left(\frac{n^2}{2}\sigma_0^2 - n\mu_0\right) I(\nu > \log X_{(n)})\, I(X_{(1)} > 0),$$

and the denominator is equal to

$$\int_{\log X_{(n)}}^{\infty} \frac{1}{\sigma_0}\phi\left(\frac{\nu - \mu_0 + n\sigma_0^2}{\sigma_0}\right) d\nu \;\; \exp\left(\frac{n^2}{2}\sigma_0^2 - n\mu_0\right) I(X_{(1)} > 0)$$

$$= \Phi\left(\frac{\mu_0 - n\sigma_0^2 - \log X_{(n)}}{\sigma_0}\right) \exp\left(\frac{n^2}{2}\sigma_0^2 - n\mu_0\right) I(X_{(1)} > 0).$$

Therefore,

$$\pi(\nu \mid \underset{\sim}{X}) = \frac{\frac{1}{\sigma_0}\phi\left(\frac{\nu - \mu_0 + n\sigma_0^2}{\sigma_0}\right)}{\Phi\left(\frac{\mu_0 - n\sigma_0^2 - \log X_{(n)}}{\sigma_0}\right)} \cdot I(\nu > \log X_{(n)})\, \cancel{I(X_{(1)}>0)}.$$

#

(b) According to part (a),

$$\pi(\theta|\underline{x}) = \frac{\frac{1}{\sigma_0}\phi\left(\frac{\nu-\mu_0+n\sigma_0^2}{\sigma_0}\right)}{\Phi\left(\frac{\mu_0-n\sigma_0^2-\log X_{(n)}}{\sigma_0}\right)} I(\theta > X_{(n)})\cdot \frac{1}{\theta}$$

(assuming $\theta > X_{(n)}$) $\propto \frac{1}{\theta}\exp\left\{-\frac{1}{2\sigma_0^2}\left(\log\theta-\mu_0+n\sigma_0^2\right)^2\right\} \triangleq h(\theta)$.

Setting $\frac{dh(\theta)}{d\theta}=0$ yields

$$1+\frac{1}{\sigma_0^2}\left(\log\theta-\mu_0+n\sigma_0^2\right)=0.$$

Solving this for θ gives $\underset{\theta>0}{\operatorname{argmax}}\, h(\theta) = \exp\{\mu_0-(n+1)\sigma_0^2\}$.

Therefore, the Bayes estimator is

$$\hat{\theta}_B = \max\left\{X_{(n)},\ \exp\{\mu_0-(n+1)\sigma_0^2\}\right\}.$$

#

(c) Note that $\lim_{n\to\infty}\hat{\theta}_B = \lim_{n\to\infty} X_{(n)}$ and $X_{(n)}$ is a consistent estimator for θ.

Next one may focus on showing that $X_{(n)} \xrightarrow{P} \theta$.

The pdf of $X_{(n)}$ is $f_{X_{(n)}}(x|\theta) = \frac{n}{\theta^n}x^{n-1} I(0<x<\theta)$.

One can show that $E(X_{(n)}) = \frac{n}{n+1}\theta \xrightarrow{n\to\infty} \theta$

and $Var(X_{(n)}) = \frac{n}{(n+2)(n+1)^2}\theta^2 \xrightarrow{n\to\infty} 0$.

It follows that $X_{(n)} \xrightarrow{P} \theta$.

One can also use the Chebyshev's inequality to show $X_{(n)} \xrightarrow{P} \theta$. For every $\varepsilon > 0$,

$$P(|X_{(n)} - \theta| \geq \varepsilon) \leq \frac{E(X_{(n)} - \theta)^2}{\varepsilon^2}$$

$$= \frac{E(X_{(n)}^2) + \theta^2 - 2\theta \cdot \frac{n}{n+1}\theta}{\varepsilon^2}$$

$$= \frac{2}{(n+1)(n+2)} \cdot \frac{1}{\varepsilon^2}$$

$$\xrightarrow{n \to \infty} 0.$$

#

#4 This is an open-ended problem. No solution is given.

Day 1,
Problem 5
Solution

a) $E(Y_{ij}) = \mu + \alpha_i + \beta_j \; (= \mu_{ij})$

$\text{var}(Y_{ij}) = \sigma^2$

b) $\hat{Y}_{ij} = \bar{Y}_{i\cdot} + \bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot}$

$$= \frac{\sum_{j=1}^{b} Y_{ij}}{b} + \frac{\sum_{i=1}^{a} Y_{ij}}{a} - \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} Y_{ij}}{ab}$$

c) $$E(\hat{Y}_{ij}) = \frac{1}{b}\sum_{j=1}^{b} E(Y_{ij}) + \frac{1}{a}\sum_{i=1}^{a} E(Y_{ij}) - \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} E(Y_{ij})$$

$$= \frac{1}{b}\sum_{j=1}^{b}(\mu + \alpha_i + \beta_j) + \frac{1}{a}\sum_{i=1}^{a}(\mu + \alpha_i + \beta_j) - \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}(\mu + \alpha_i + \beta_j)$$

$$= \frac{1}{b}\left[b\mu + b\alpha_i + \sum_{j=1}^{b}\beta_j\right] + \frac{1}{a}\left[a\mu + \sum_{i=1}^{a}\alpha_i + a\beta_j\right] - \frac{1}{ab}\left[ab\mu + b\sum_{i=1}^{a}\alpha_i + a\sum_{j=1}^{b}\beta_j\right] = \mu + \mu - \mu + \alpha_i + \frac{1}{a}\sum_i \alpha_i - \frac{1}{a}\sum_i \alpha_i$$

$$+ \beta_j + \frac{1}{b}\sum_j \beta_j - \frac{1}{b}\sum_j \beta_j = \mu + \alpha_i + \beta_j = \mu_{ij}$$

d) $$\text{var}(\hat{Y}_{11}) = \text{var}\left\{\frac{1}{3}Y_{11} + \frac{1}{3}Y_{12} + \frac{1}{3}Y_{13} + \frac{1}{2}Y_{11} + \frac{1}{2}Y_{21} - \frac{1}{6}Y_{11} - \frac{1}{6}Y_{12} - \frac{1}{6}Y_{13} - \frac{1}{6}Y_{21} - \frac{1}{6}Y_{22} - \frac{1}{6}Y_{23}\right\}$$

$$= \text{var}\left\{\frac{4}{6}Y_{11} + \frac{1}{6}Y_{12} + \frac{1}{6}Y_{13} + \frac{2}{6}Y_{21} - \frac{1}{6}Y_{22} - \frac{1}{6}Y_{23}\right\}$$

$$= \frac{16}{36}\sigma^2 + \frac{1}{36}\sigma^2 + \frac{1}{36}\sigma^2 + \frac{4}{36}\sigma^2 + \frac{1}{36}\sigma^2 + \frac{1}{36}\sigma^2$$

$$= \frac{24}{36}\sigma^2 = \frac{2}{3}\sigma^2$$

Generally, $\text{var}(\hat{Y}_{ij}) = \frac{2}{3}\sigma^2$ when $a = 2$, $b = 3$

since all Y_{ij} have variance σ^2.

e) Both Y_{ij} and $\hat{Y}_{ij}$ are unbiased for μ_{ij}, but $\mathrm{var}(\hat{Y}_{ij}) = \frac{2}{3}\sigma^2 < \sigma^2 = \mathrm{var}(Y_{ij})$, so $\hat{Y}_{ij}$ is a better estimator.

f) $$\hat{Y}_{ij} \pm t_{(1-\alpha/2,\, df_{error})}\sqrt{\frac{2}{3}\, MSE}$$

where $$MSE = \sum_i \sum_j (Y_{ij} - \hat{Y}_{ij})^2 \Big/ [(a-1)(b-1)]$$
$$= \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot})^2 \Big/ [(a-1)(b-1)]$$

and $df_{error} = (a-1)(b-1)$

(Important to give correct formula and d.f. for estimator of σ^2)

Day 1,
Problem 6
Solution

a) $$X'W = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_n \end{bmatrix} \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & & w_n \end{bmatrix}$$

$$= \begin{bmatrix} w_1 & w_2 & \cdots & w_n \\ w_1X_1 & w_2X_2 & \cdots & w_nX_n \end{bmatrix}$$

$$\Rightarrow X'WX = \begin{bmatrix} w_1 & w_2 & \cdots & w_n \\ w_1X_1 & w_2X_2 & \cdots & w_nX_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum w_i & \sum w_iX_i \\ \sum w_iX_i & \sum w_iX_i^2 \end{bmatrix}$$

$$\Rightarrow (X'WX)^{-1} = \frac{1}{(\sum w_i)(\sum w_iX_i^2) - (\sum w_iX_i)^2} \begin{bmatrix} \sum w_iX_i^2 & -\sum w_iX_i \\ -\sum w_iX_i & \sum w_i \end{bmatrix}$$

and $$X'WY = \begin{bmatrix} \sum w_iY_i \\ \sum w_iX_iY_i \end{bmatrix}$$

$$\Rightarrow b_{w0} = \frac{\sum w_iX_i^2 \sum w_iY_i - \sum w_iX_i \sum w_iX_iY_i}{(\sum w_i)(\sum w_iX_i^2) - (\sum w_iX_i)^2}$$

$$b_{w1} = \frac{\sum w_i \sum w_iX_iY_i - \sum w_iX_i \sum w_iY_i}{(\sum w_i)(\sum w_iX_i^2) - (\sum w_iX_i)^2}$$

b) $w_i = \frac{1}{\sigma_i^2}$ and $\sigma_i^2 = kX_i$ for a constant k

$\Rightarrow w_i = \frac{1}{kX_i}$

Hence $$b_{w0} = \frac{\sum \frac{1}{kX_i} X_i^2 \sum \frac{1}{kX_i} Y_i - \sum \frac{1}{kX_i} X_i \sum \frac{1}{kX_i} X_i Y_i}{\left(\sum \frac{1}{kX_i}\right)\left(\sum \frac{1}{kX_i} X_i^2\right) - \left(\sum \frac{1}{kX_i} X_i\right)^2}$$

$$= \frac{\left(\frac{1}{k} \sum X_i\right)\left(\frac{1}{k} \sum \frac{Y_i}{X_i}\right) - \left(\frac{n}{k}\right)\left(\frac{1}{k} \sum Y_i\right)}{\left(\frac{1}{k} \sum \frac{1}{X_i}\right)\left(\frac{1}{k} \sum X_i\right) - \left(\frac{n}{k}\right)^2}$$

$$= \frac{\left(\sum X_i\right)\left(\sum \frac{Y_i}{X_i}\right) - n \sum Y_i}{\left(\sum \frac{1}{X_i}\right)\left(\sum X_i\right) - n^2} =$$

$$b_{w1} = \frac{\sum \frac{1}{kX_i} \sum \frac{1}{kX_i} X_i Y_i - \sum \frac{1}{kX_i} X_i \sum \frac{1}{kX_i} Y_i}{\left(\sum \frac{1}{kX_i}\right)\left(\sum \frac{1}{kX_i} X_i^2\right) - \left(\sum \frac{1}{kX_i} X_i\right)^2}$$

$$= \frac{\left(\frac{1}{k} \sum \frac{1}{X_i}\right)\left(\frac{1}{k} \sum Y_i\right) - \left(\frac{n}{k}\right)\left(\frac{1}{k} \sum \frac{Y_i}{X_i}\right)}{\left(\frac{1}{k} \sum \frac{1}{X_i}\right)\left(\frac{1}{k} \sum X_i\right) - \left(\frac{n}{k}\right)^2}$$

$$= \frac{\left(\sum \frac{1}{X_i}\right)\left(\sum Y_i\right) - n \sum \frac{Y_i}{Y_i}}{\left(\sum \frac{1}{X_i}\right)\left(\sum X_i\right) - n^2}$$

c) $$b_{w0} = \frac{18\left[\sum_{i=1}^{3} Y_i + \frac{1}{2}\sum_{i=4}^{6} Y_i + \frac{1}{3}\sum_{i=7}^{9} Y_i\right] - 9\sum_{i=1}^{9} Y_i}{(5.5)(18) - 81}$$

$$= \frac{9\sum_{i=1}^{3} Y_i - 3\sum_{i=7}^{9} Y_i}{18} = \frac{1}{2}\sum_{i=1}^{3} Y_i - \frac{1}{6}\sum_{i=7}^{9} Y_i$$

$$b_{w1} = \frac{5.5\sum_{i=1}^{9} Y_i - 9\left[\sum_{i=1}^{3} Y_i + \frac{1}{2}\sum_{i=4}^{6} Y_i + \frac{1}{3}\sum_{i=7}^{9} Y_i\right]}{(5.5)(18) - 81}$$

$$= \frac{-3.5\sum_{i=1}^{3} Y_i + \sum_{i=4}^{6} Y_i + 2.5\sum_{i=7}^{9} Y_i}{18}$$

$$= -\frac{7}{36}\sum_{i=1}^{3} Y_i + \frac{1}{18}\sum_{i=4}^{6} Y_i + \frac{5}{36}\sum_{i=7}^{9} Y_i$$

d) One possible solution:

(1) Calculate b_{w1} for observed $Y_1, Y_2, \ldots, Y_9$.

(2) Randomly permute observed $Y_1, Y_2, \ldots, Y_9$, forming $Y_1^*, Y_2^*, \ldots, Y_9^*$. Calculate b_{w1}^* based on permuted data

(3) Repeat (2) for all (or many) possible permutations obtaining $b_{w1}^{*(1)}, \ldots, b_{w1}^{*(N)}$.

(4) P-value for the test of $H_0: \beta_1 = 0$ vs. $H_a: \beta_1 \neq 0$

is $$P = \frac{1}{N}\left\{\# b_{w1}^* : |b_{w1}^*| \geq |b_{w1}|\right\}$$

- Other ~~distribution~~ - free solutions are possible, but should not rely on a normality assumption.