Stat 513
Homework Key 6

February 2016

REGULAR PROBLEMS

1. Suppose $Y_1, \ldots, Y_n$ are i.i.d. Bernoulli random variables with $Y_i \sim \mathcal{B}(1,p), p \in (0,1)$. Denote the target of inference by $\theta = Var(Y_i)$.

   (a) Derive a uniform minimum variance unbiased estimator(UMVUE) $\tilde{\theta}$ of θ.

   (b) Find the variance of $\tilde{\theta}$. Does it meet the Cramer-Rao lower bound for an unbiased estimator of θ?

   (c) Derive an exact probability distribution for $\tilde{\theta}$.

   (d) Derive an asymptotic distribution for $\tilde{\theta}$.

   Ans:

   (a) We know that $\bar{Y}$ is complete sufficient. An unbiased estimator of θ is the unbiased sample variance,

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

   As S^2 is a function of the complete sufficient statistic $\bar{Y}$ only, we have that $\tilde{\theta} = S^2$ is the UMVUE of θ.

   (b) The variance of $\tilde{\theta}$ can be based on the result of STAT 512 Homework 6, Problem 5. We found that

$$\begin{aligned} Var(S^2) &= \frac{1}{n}(\gamma - 3\sigma^4) + \frac{2}{n-1}\sigma^4 \\ &= \frac{1}{n}(\theta(1-3\theta) - 3\theta^2) + \frac{2}{n-1}\theta^2 \end{aligned}$$

   where $\gamma = p(1-p)(1-3p(1-p)) = \theta(1-3\theta)$ is the fourth central moment of the Bernoulli(p) distribution. We know that the CRLB is attainable for $g(p) = p$, so a BLUE only exists for linear functions of p. Since $\theta = p(1-p)$, our UMVUE $\tilde{\theta}$ does not attain the CRLB.

   (c) The exact distribution of $\tilde{\theta}$ is based on the sum of the Y_i. Since $T = \sum_{i=1}^{n} Y_i \sim \mathcal{B}(n,p)$, we can write its pmf as

$$\begin{aligned} Pr(\tilde{\theta} = s) &= Pr\left(\frac{n}{n-1}\frac{1}{n}T\left(1 - \frac{1}{n}T\right) = s\right) \\ &= Pr(nT - T^2 = (n-1)s) \\ &= Pr(T^2 - nT + (n-1)s = 0) \\ &= Pr\left(T \in \left\{\frac{n \pm \sqrt{n^2 - 4(n-1)s}}{2}\right\}\right) \\ &= 2 \times Pr\left(T = \frac{n + \sqrt{n^2 - 4(n-1)s}}{2}\right) \end{aligned}$$

   for $s \in \left\{\frac{t(n-t)}{n(n-1)} : t = 0, 1, \ldots, n/2\right\}$.

   (d) By the CLT, we have that

$$\sqrt{n}(\bar{Y} - p) \to_d \mathcal{N}(0, \theta).$$

   By the Delta Method with $g(x) = x - x^2$, we obtain that

$$\sqrt{n}(\tilde{\theta} - \theta) \to_d \mathcal{N}(0, V(\theta)))$$

   with $V(\theta) = (1-2p)^2\theta = (1-\theta^2)\theta$.

2. Suppose $Y_1, \dots, Y_n$ are i.i.d. normal random variables with $Y_i \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu \in (-\infty, \infty)$ and $\sigma^2 \in (0, \infty)$. Denote the target of inference by $\theta = \sigma^2$.

   (a) Derive a uniform minimum variance unbiased estimator(UMVUE) $\tilde{\theta}$ of θ.

   (b) Find the variance of $\tilde{\theta}$. Does it meet the Cramer-Rao lower bound for an unbiased estimator of θ?

   (c) Derive an exact probability distribution for $\tilde{\theta}$.

   (d) Derive an asymptotic distribution for $\tilde{\theta}$.

Ans:

(a) The complete sufficient statistic in this model is $(\bar{Y}, \sum_{i=1}^n Y_i^2)$. An unbiased estimator of θ is again the unbiased sample variance,

$$S^2 = \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^n (Y_i^2 - \bar{Y})^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n Y_i^2 - \bar{Y}^2\right).$$

As S^2 is a function of the complete sufficient statistic $\bar{Y}$ only, we have that $\tilde{\theta} = S^2$ is the UMVUE of θ.

(b) The variance of $\tilde{\theta}$ follows because

$$\tilde{\theta} \sim \frac{\sigma^2}{n-1}\chi^2_{n-1},$$

so that

$$Var(\tilde{\theta}) = \frac{\sigma^4}{(n-1)^2}2(n-1) = \frac{2\sigma^4}{n-1}.$$

The score for $\theta = \sigma^2$ is just

$$\mathcal{U}(\sigma^2) = \frac{1}{2\sigma^4}((Y-\mu)^2 - \sigma^2),$$

so we see that a BLUE can exist for θ in this case. The information for an observation is

$$\mathbb{J}(\theta) = E[\mathcal{U}^2(\theta)] = \frac{1}{2\sigma^4}.$$

Hence the CRLB is

$$\mathbb{J}^{-1}(\theta) = \frac{1}{n}2\sigma^4 < Var(\tilde{\theta}).$$

(c) The distribution of $\tilde{\theta}$ is a scaled χ^2_1 distribution,

$$\tilde{\theta} \sim \frac{\sigma^2}{n-1}\chi^2_1,$$

as seen from the course notes as well as Casella-Berger Theorem 5.3.1.

(d) An asymptotic distribution for $\tilde{\theta}$ follows from similar arguments from the Homework 1 key, where we derived an asymptotic joint distribution for the sample mean and variance. Applying that result, and letting γ_4 be the kurtosis of the $\mathcal{N}(\mu, \sigma^2)$,

$$\sqrt{n}(\tilde{\theta} - \theta) \to_d \mathcal{N}(0, V(\theta)).$$

Recall from STAT 512 Homework 4, Question 2 that $\gamma_4 = 3\sigma^4$ so that

$$V(\theta) = \gamma_4 - \sigma^4 = 3\sigma^4 - \sigma^4 = 2\theta^2.$$

3. Suppose $Y_1, \dots, Y_n$ are i.i.d. normal random variables with $Y_i \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu \in (-\infty, \infty)$ and $\sigma^2 \in (0, \infty)$. Denote the target of inference by $\theta = \mu^2$.

   (a) Derive a uniform minimum variance unbiased estimator(UMVUE) $\tilde{\theta}$ of θ.

   (b) Find the variance of $\tilde{\theta}$. Does it meet the Cramer-Rao lower bound for an unbiased estimator of θ?

   (c) Derive an exact probability distribution for $\tilde{\theta}$.

   (d) Derive an asymptotic distribution for $\tilde{\theta}$.

Ans:

(a) Recall that the unbiased sample variance we appealed to in problems 1 and 2 satisfies

$$E[S^2] = Var(Y_1) = E[Y_1^2] - \mu^2.$$

Hence, a natural choice for an unbiased estimator of $\theta = \mu^2$ is

$$\tilde{\theta} = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - S^2,$$

since

$$E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - S^2\right] = E[Y_1^2] - (E[Y_1^2] - \mu^2) = \mu^2.$$

However, as seen in the homework (as well as Casella-Berger Theorem 5.3.1), $\bar{Y}$ is independent of S^2 for a random sample from a normal distribution. We can express $\tilde{\theta}$ in a way that allows us to make use of this independence:

$$\begin{aligned}
\tilde{\theta} &= \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - S^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \bar{Y}^2 + \bar{Y}^2 - S^2 \\
&= \frac{n-1}{n} S^2 + \bar{Y}^2 - S^2 \\
&= \bar{Y}^2 - \frac{1}{n} S^2.
\end{aligned}$$

(b) The variance of $\tilde{\theta}$ is easily calculated since $\bar{Y}$ is independent of S^2,

$$Var(\tilde{\theta}) = Var(\bar{Y}^2) + \frac{1}{n^2} Var(S^2).$$

Now, $\bar{Y} \sim \mathcal{N}(\mu, \sigma^2/n)$, so $\bar{Y}^2 \sim \sigma^2/n\chi_1^2(\delta)$ where $\delta = \mu^2/2$. Additionally, $S^2 \sim \sigma^2/(n-1)\chi_{n-1}^2$. Hence, we can simplify the variance to

$$Var(\tilde{\theta}) = Var(\bar{Y}^2) + Var(S^2) = 2(1+\mu^2)\frac{\sigma^4}{n^2} + 2\frac{1}{n^2}\frac{\sigma^4}{(n-1)}.$$

Since the score for μ is just

$$\mathcal{U}(\mu) = \sigma^{-2}(Y_i - \mu),$$

only linear functions of μ admit a BLUE. Hence, $\tilde{\theta}$ will not obtain the CRLB.

(c) The exact distribution of $\tilde{\theta}$ is a convolution of the two χ^2 distributions detailed above,

$$\tilde{\theta} \sim \frac{\sigma^2}{n}\chi_1^2(\mu^2/2) - \frac{\sigma^2}{n-1}\chi_{n-1}^2.$$

(d) Since $n^{-1}S^2 \to_p 0$, we know that if

$$\bar{Y}^2 \to_d X,$$

then, by Slutsky's theorem

$$\sqrt{n}(\tilde{\theta} - \theta) = \sqrt{n}(\bar{Y}^2 - \theta)\frac{1}{n^{1/2}} S^2 \to_d X.$$

Additionally, we have from the CLT that

$$\sqrt{n}(\bar{Y} - \mu) \to_d \mathcal{N}(0, \sigma^2).$$

We apply the Delta Method with $g(x) = x^2$, finding

$$\sqrt{n}(\bar{Y}^2 - \mu^2) \to_d \mathcal{N}(0, 4\mu^2\sigma^2).$$

Hence, by our Slutsky's theorem argument,

$$\sqrt{n}(\tilde{\theta} - \theta) \to_d \mathcal{N}(0, 4\theta\sigma^2).$$

4. Suppose $Y_1, \ldots, Y_n$ are i.i.d. exponential random variables with $Y_i \sim \mathcal{E}(\mu)$ with mean $\mu \in (0, \infty)$. Denote the target of inference by $\theta = Var(Y_i) = \mu^2$.

   (a) Derive a uniform minimum variance unbiased estimator(UMVUE) $\tilde{\theta}$ of θ.

   (b) Find the variance of $\tilde{\theta}$. Does it meet the Cramer-Rao lower bound for an unbiased estimator of θ?

   (c) Derive an exact probability distribution for $\tilde{\theta}$.

   (d) Derive an asymptotic distribution for $\tilde{\theta}$.

Ans:

(a) We know that $\bar{Y}$ is complete sufficient. Also,

$$E\bar{Y}^2 = \frac{n+1}{n}\mu^2.$$

Hence by Lehman Scheffe theorem, the UMVUE is $\tilde{\theta} = \dfrac{n\bar{Y}^2}{n+1}$.

(b) $S = \sum_{i=1}^n Y_i \sim Gamma(n, \mu)$.

$$\begin{aligned} ES^k &= \int_0^\infty \frac{s^{k+n-1}e^{-s/\mu}}{\mu^n \Gamma(n)} ds \\ &= \mu^k \frac{(n+k-1)!}{(n-1)!} \int_0^\infty \frac{s^{k+n-1}e^{-s/\mu}}{\mu^{n+k}\Gamma(n+k)} ds \\ &= \mu^k \frac{(n+k-1)!}{(n-1)!}. \end{aligned}$$

Using the above relation for $k = 2, 4$ we get,

$$Var(S^2) = ES^4 - (ES^2)^2 = n(n+1)(4n+6)\mu^4.$$

$$Var(\tilde{\theta}) = Var(\frac{S^2}{n(n+1)}) = \frac{(4n+6)\mu^4}{n(n+1)}.$$

Since

$$f(y|\mu) = \frac{e^{-\frac{\sum_{i=1}^n y_i}{\mu}}}{\mu^n},$$

we have

$$\mathcal{U}(\mu) = \frac{-n}{\mu^2}(\bar{Y} - \mu). \tag{1}$$

Therefore we have a BRUE for μ. However, $\theta = \mu^2$ is not a linear function of μ. Therefore there is no BRUE, i.e. CRLB can not be attained by the UMVUE $\tilde{\theta}$.

(c) Let

$$\begin{aligned} g(s) &= \frac{s^2}{n(n+1)}. \\ \Rightarrow \tilde{\theta} = g(S) &= \frac{S^2}{n(n+1)} \\ g^{-1}(y) &= \sqrt{n(n+1)y} \\ g'(s) &= \frac{2s}{n(n+1)} \\ f_{\tilde{\theta}}(y) = f_{g(S)}(y) &= \frac{f_S(g^{-1}(y))}{g'(g^{-1}(y))} \\ &= \frac{n(n+1)e^{-\sqrt{n(n+1)y}/\mu}(n(n+1)y)^{n/2-1}}{2(n-1)!\mu^n} 1[y > 0]. \end{aligned}$$

(d) By CLT,

$$\sqrt{n}(\bar{Y} - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mu^2).$$

By Delta method,

$$\sqrt{n}(\bar{Y}^2 - \mu^2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 4\mu^4).$$

$\sqrt{n}(\bar{Y}^2 - \tilde{\theta}) = \bar{Y}^2 O_P(\frac{1}{\sqrt{n}}) = O_P(1)o_P(1) = o_P(1)$. Hence, by Slutskey's theorem,

$$\sqrt{n}(\tilde{\theta} - \mu^2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 4\mu^4).$$

5. Suppose $Y_1, \ldots, Y_n$ are i.i.d. exponential random variables with $Y_i \sim \mathcal{U}(0, \mu)$ with mean $\mu \in (0, \infty)$. Denote the target of inference by $\theta = Var(Y_i)$.

   (a) Derive a uniform minimum variance unbiased estimator(UMVUE) $\tilde{\theta}$ of θ.

   (b) Find the variance of $\tilde{\theta}$. Does it meet the Cramer-Rao lower bound for an unbiased estimator of θ?

   (c) Derive an exact probability distribution for $\tilde{\theta}$.

   (d) Derive an asymptotic distribution for $\tilde{\theta}$.

   Ans:

   (a) $Y_{(n)}$ is complete sufficient with PDF $n\mu^{-n}y^{n-1}I_{[0,\mu]}(y)$. Therefore

$$EY_{(n)}^k = \frac{n}{\mu^n} \int_0^\mu x^{n+k-1} dx \tag{2}$$

$$= \frac{n\mu^k}{n+k} \tag{3}$$

   Using the above we get that $EY_{(n)}^2 = \frac{n\mu^2}{n+2}$ and we know $\theta = Var(Y_i) = \frac{\mu^2}{12}$. Therefore $\tilde{\theta} = \frac{n+2}{12n} Y_{(n)}^2$ is the UMVUE.

   (b) Using (2) we get that

$$Var(\tilde{\theta}) = \frac{(n+2)^2\mu^4}{144n} \left(\frac{1}{n+4} - \frac{n}{(n+2)^2} \right) = \frac{\mu^4}{36n(n+4)}.$$

   We know that the information for μ is $\frac{n}{\mu^2}$. Therefore the CRLB is $\frac{4\mu^4}{n}$. For large n, $Var(\tilde{\theta}) \approx \frac{\mu^4}{36n^2} < \frac{4\mu^4}{n}$. This happens because uniform is not a regular family.

   (c) Let $c = \frac{12n}{n+1}$. Let us define

$$g(s) = \frac{s^2}{c}.$$

$$\Rightarrow \tilde{\theta} = g(Y_{(n)}) = \frac{{Y_{(n)}}^2}{c}$$

$$\Rightarrow g^{-1}(y) = \sqrt{cy}$$

$$\Rightarrow g'(s) = \frac{2s}{c}$$

$$\Rightarrow f_{\tilde{\theta}}(y) = f_{g(Y_{(n)})}(y) = \frac{f_{Y_{(n)}}(g^{-1}(y))}{g'(g^{-1}(y))}$$

$$= \frac{nc\mu^{-n}}{2} (\sqrt{nc})^{n-2} I_{[0,\mu]}(\sqrt{cy})$$

$$= \frac{6n^2\mu^{-n}}{(n+1)} (\sqrt{nc})^{n-2} I_{[0,\mu]}(\sqrt{cy})$$

   where $c = \frac{12n}{n+1}$.

(d)

$$
\begin{aligned}
&P(n(\mu - Y_{(n)}) \geq y) \\
&= \left(1 - \frac{x}{\mu n}\right)^{\frac{-n\mu}{x} \cdot \frac{-x}{\mu}} \to e^{-x/\mu}.
\end{aligned}
$$

Therefore

$$n(\mu - Y_{(n)}) \xrightarrow{\mathcal{D}} \mathcal{E}(\mu).$$

Let $g(\mu) = \frac{\mu^2}{12}$. Hence by Generalised Delta method,

$$n\left(\frac{\mu^2}{12} - \frac{Y_{(n)}^2}{12}\right) \xrightarrow{\mathcal{D}} g'(\mu)\mathcal{E}(\mu) \equiv \frac{\mu}{6}\mathcal{E}(\mu) \equiv \mathcal{E}(\mu^2/6).$$

Now

$$|n(\tilde{\theta} - \frac{Y_{(n)}^2}{12})| = |\frac{Y_{(n)}^2}{6n}| \xrightarrow{P} 0.$$

Therefore by Slutskey's theorem,

$$n\left(\frac{\mu^2}{12} - \tilde{\theta}\right) \xrightarrow{\mathcal{D}} \mathcal{E}(\mu^2/6).$$

6. Suppose $Y_1, \ldots, Y_n$ are i.i.d. exponential random variables with $Y_i \sim \mathcal{E}(\lambda)$ with hazard $\lambda \in (0, \infty)$. Denote the target of inference by $\theta = P(Y_i > 1)$.

(a) Derive a uniform minimum variance unbiased estimator(UMVUE) $\tilde{\theta}$ of θ.

(b) Find the variance of $\tilde{\theta}$. Does it meet the Cramer-Rao lower bound for an unbiased estimator of θ?

(c) Derive an exact probability distribution for $\tilde{\theta}$.

(d) Derive an asymptotic distribution for $\tilde{\theta}$.

Ans:

(a) $U = I[Y_1 > 1]$ is an unbiased estimator of θ. Recall the complete sufficient statistic is $S = \sum_{i=1}^n y_i$. Hence,

$$
\begin{aligned}
\tilde{\theta} &= E(U|S = s) \\
&= P(Y_1 > 1|S = s) \\
&= P(Y_1/S > 1/s|S = s)
\end{aligned}
$$

We will use the following result now:
Let $X \sim Gamma(\alpha, 1)$, $Y \sim Gamma(\beta, 1)$ where α, β are the shape parameters. If X and Y are independent, then

$$\frac{X}{X+Y} \sim Beta(\alpha, \beta).$$

Following the above result we get that

$$\frac{Y_1}{S} \sim Beta(1, n-1)$$

which is independent of λ, hence ancillary. Following Basu's theorem, $\frac{Y_1}{S}$ and S are independent. Therefore,

$$
\begin{aligned}
\tilde{\theta} &= P(Y_1/S > 1/s) \\
&= \frac{\Gamma(n)}{\Gamma(n-1)} 1[s > 1] \int_{1/s}^{1} (1-x)^{n-2} dx \\
&= \left(1 - \frac{1}{s}\right)^{n-1} 1[s > 1].
\end{aligned}
$$

Therefore the UMVUE is $\left(1 - \frac{1}{S}\right)^{n-1} 1[S > 1]$ where $S = \sum_{i=1}^n Y_i$.

(b) $S \sim Gamma(n, \lambda)$ with $f(s) = \dfrac{e^{-\lambda s} s^{n-1} \lambda^n}{\Gamma(n)}$.

$$
\begin{aligned}
E\tilde{\theta}^2 &= \int_1^\infty \left(1 - \frac{1}{s}\right)^{2n-2} \frac{e^{-\lambda s} s^{n-1} \lambda^n}{\Gamma(n)} ds \\
&= \int_1^\infty \frac{(s-1)^{2n-2}}{s^{n-1}} \frac{e^{-\lambda s} \lambda^n}{\Gamma(n)} ds.
\end{aligned}
$$

$$
Var(\tilde{\theta}) = \int_1^\infty \frac{(s-1)^{2n-2}}{s^{n-1}} \frac{e^{-\lambda s} \lambda^n}{\Gamma(n)} ds - \theta^2.
$$

From (1) we see that a BRUE for $\dfrac{1}{\lambda} = \mu$ exists. Since $\theta = e^{-\lambda}$ is not a linear function of $\dfrac{1}{\lambda}$, no BRUE exists. So the CRLB will not match with the variance of $\tilde{\theta}$.

(c) S follows $Gamma(n, \lambda)$. Let us define

$$
\begin{aligned}
& g(s) = \left(1 - \frac{1}{s}\right)^{n-1} 1[s > 1]. \\
& \Rightarrow \tilde{\theta} = g(S) = \left(1 - \frac{1}{s}\right)^{n-1} 1[s > 1]. \\
& \Rightarrow g^{-1}(y) = \frac{1}{1 - y^{1/(n-1)}} 1[y > 0] \\
& \Rightarrow g'(s) = \frac{(n-1)(s-1)^{n-2}}{s^n} 1[s > 1]. \\
& \Rightarrow f_{\tilde{\theta}}(y) = f_{g(S)}(y) = \frac{f_S(g^{-1}(y))}{g'(g^{-1}(y))} \\
& = \frac{\lambda^n}{(n-1)\Gamma(n)} e^{-\frac{\lambda}{1 - y^{1/(n-1)}}} y^{-\frac{n-2}{n-1}} (1 - y^{1/(n-1)})^{-(n+1)} 1[y > 0].
\end{aligned}
$$

(d) By CLT,

$$
\sqrt{n}\left(\bar{Y} - \frac{1}{\lambda}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{\lambda^2}\right).
$$

Therefore using delta method we get

$$
\sqrt{n}(e^{-1/\bar{Y}} - e^{-\lambda}) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \lambda^2 e^{-2\lambda}\right).
$$

$$
\sqrt{n}(\tilde{\theta} - e^{-1/\bar{Y}}) = \frac{1}{\sqrt{n}\bar{Y}} + o_P(1/(\sqrt{n}\bar{Y})).
$$

Since $\dfrac{1}{\sqrt{n}\bar{Y}}$ is $o_P(1)$, the above term is $o_P(1)$. Therefore by Slutskey's theorem,

$$
\sqrt{n}(\tilde{\theta} - e^{-\lambda}) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \lambda^2 e^{-2\lambda}\right).
$$

MORE INVOLVED PROBLEMS

7. Suppose $Y_1, \ldots Y_n$ are i.i.d. Bernoulli random variables with $Y_i \sim \mathcal{B}(p), p \in (0, \infty)$. Denote the target of inference by p. Suppose further that the data are sample sequentially with the following (fairly stupid) group sequential stopping rule with potential analyses after $n_1 = 1$ and $n_2 = 2$ observations have been observed:

   - Y_1 is sampled.
   - If $Y_1 = 1$, we stop sampling at $n_1 = 1$ and define $(M, s) = (1, Y_1)$.
   - If $Y_1 = 0$, we sample $Y_2, \ldots, Y_1 1$ and then stop sampling at $n_2 = 11$. We define $(M, s) = (2, \sum_{i=2}^{11} Y_i)$.

   Note that under this scheme, M is our (random) stopping time, and S. In group sequential terminology, the "continuation set" at the first analysis is $C_1 = \{0\}$.

   (a) Derive the exact distribution for statistic (M, S).

   (b) Show that the exact distribution is a curved exponential family distribution.

   (c) Find a minimal sufficient statistic for p.

   (d) Show that minimal sufficient statistic for p is a complete sufficient statistic.

   (e) Find a uniform minimum variance unbiased estimator (UMVUE) $\tilde{p}$ of p.

   (f) Find the maximum likelihood estimator (MLE) $\hat{p}$ of p, and show that it is a biased estimator.

   (g) Compare the MSE of the MLE $\hat{p}$to the MSE of the UMVUE $\tilde{p}$.

   (h) Derive a "bias adjusted estimator" $\check{p}$ such that

$$E\left[\frac{S}{n_M}|p = \check{p}\right] = \hat{p}$$

Ans:

(a) We observe that when $M = 1, S = 1$ and when $M = 2, S = \sum_{i=2}^{11} Y_i \sim Bin(10, p)$.

$$f(s, m) = p^{2-m}\left((1-p)\binom{10}{s}p^s(1-p)^{10-s}\right)^{m-1}.$$

(b)

$$f(s, m) = \binom{10}{s}^{m-1} e^{(sm-s+2-m)\log p + (m-1)(11-s)\log(1-p)} \tag{4}$$

Therefore we see that (M, S) has an exponential family distribution. Now we have only one parameter p but clearly $SM - S + 2 - M$ and $(M-1)(11-S)$ are linearly independent. Therefore it is a curved exponential family. Let us denote $T_1 = SM - S + 2 - M$ and $T_2 = (M-1)(11-s)$.

(c) Clearly T_1 and T_2 are minimal sufficient. Hence (M, S) is also minimal sufficient since T_1 and T_2 are $1-1$ functions of M and S.

(d) We have to show that (M, S) is complete. Now $M \in \{1, 2\}$ and $S \in \{1, \ldots, 10\}$. Suppose, if possible $\exists$ a no-zero function $h : \{1, 2\} \times \{1, \ldots, 10\} \to \mathbb{R}$ such that for all $p \in (0, 1)$

$$Eh(M, S) = 0$$
$$E[E[h(M, S)|M]] = 0 \qquad pE[h(1, S)|M = 1] + (1-p)E[h(2, S)|M = 2] = 0$$
$$ph(1, 1) + (1-p)Eh(2, X) = 0$$

where $X \sim Bin(10, p)$. Letting $p \to 1$ we get that

$$h(1, 1) = 0.$$

Therefore

$$(1-p)Eh(2, X) = 0$$

for all p. Since X is complete for the family $X \sim Bin(10, p)$ the above holds only if $h(2, s) = 0$ for $s \in \{1, \ldots, 10\}$. Hence $h(m, s) = 0$ for all (m, s) such that $f(s, m) > 0$. Therefore (M, S) is complete.

(e) $2 - M \sim Bin(1, p)$. Therefore $E(2 - M) = p$. Since it is a function of the complete sufficient statistic it is also the UMVUE.

(f) From (4) we get that

$$l(p) = (sm - s + 2 - m)\log p + (11m - 11 - sm + s)\log(1-p) + c$$

where c is a constant term independent of p. Hence,

$$\mathcal{U}(p) = \frac{sm - s + 2 - m}{p} - \frac{11m - 11 - sm + s}{1-p}.$$

Solving for $\mathcal{U}(p) = 0$, we get that

$$\hat{p} = \frac{SM - S + 2 - M}{10M - 9}.$$

We also observe that it is equivalent as writing

$$\hat{p} = \frac{S}{n_M} \tag{5}$$

$$E\hat{p} = EE[\hat{p}|M] = p.1 + (1-p)\frac{10p}{11}.$$

since when $M = 1, S = 1$ and when $M = 2, S = \sum_{i=2}^{11} Y_i \sim Bin(10, p)$. Hence

$$E\hat{p} = \frac{21p - 10p^2}{11} \neq p. \tag{6}$$

(g) $MSE(\hat{p}) = Var(\hat{p}) + (E\hat{p} - p)^2$,

$$(E\hat{p} - p)^2 = (\frac{10p(1-p)}{11})^2.$$

$$Var(\tilde{p}) = E\hat{p}^2 - (E\hat{p})^2$$
$$E\hat{p}^2 = E[E\hat{p}^2|M]$$
$$= E[E(\frac{S}{n_M})|M]] = p.1 + (1-p)\frac{10p(1-p) + 100p^2}{121}$$

Therefore,

$$Var(\hat{p}) = \frac{p(-100p^3 + 330p^2 - 361p + 131)}{121}$$
$$MSE(\hat{p}) = \frac{130p^3 - 261p^2 + 131p}{121}$$
$$MSE(\tilde{p}) = Var(\tilde{p}) = p(1-p)$$
$$MSE(\hat{p}) - Var(\tilde{p}) = \frac{10p(1-p)(1-13p)}{121}$$

When $p < \frac{1}{13}$, $MSE(\hat{p}) > MSE(\tilde{p})$. When $p > \frac{1}{13}$, $MSE(\hat{p}) < MSE(\tilde{p})$. When $p = \frac{1}{13}$, $MSE(\hat{p}) = MSE(\tilde{p})$.

(h) From (5) we get that $\frac{S}{n_M}$ is the MLE. Hence

$$E\left[\frac{S}{n_M}|p = \check{p}\right] = E[\hat{p}|p = \check{p}]$$

Therefore following (6) we get that

$$E[\hat{p}|p = \check{p}] = \frac{21\check{p} - 10\check{p}^2}{11}.$$

Let

$$\hat{p} = \frac{21\check{p} - 10\check{p}^2}{11}$$
$$\Rightarrow 10\check{p}^2 - 21\check{p} + 11\hat{p} = 0$$

therefore,

$$\check{p} = \frac{21 \pm \sqrt{441 - 440\hat{p}}}{20}.$$

Clearly

$$\frac{21 + \sqrt{441 - 440\hat{p}}}{20} > 1.$$

and

$$0 = \frac{21 - \sqrt{441 - 440.0}}{20} < \frac{21 - \sqrt{441 - 440\hat{p}}}{20} < \frac{21 - \sqrt{441 - 440.1}}{20} = 1.$$

Therefore the only feasible solution is

$$\check{p} = \frac{21 - \sqrt{441 - 440\hat{p}}}{20}.$$