MATH 503: Mathematical Statistics
Dr. Kimberly F. Sellers, Instructor
Homework 4 Solutions

1. Let $X_1, X_2, \ldots, X_n$ represent a random sample from the discrete distribution having the pmf

$$f(x;\theta) = \begin{cases} \theta^x(1-\theta)^{1-x} & x = 0,1; 0 < \theta < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Show that $Y_1 = \sum_{i=1}^n X_i$ is a complete sufficient statistic for θ. Find the unique function of Y_1 that is the UMVUE of θ.

Solution:

$$\begin{aligned} f(x;\theta) &= \theta^x(1-\theta)^{1-x} = \exp[\ln\{\theta^x(1-\theta)^{1-x}\}] = \exp[x\ln\theta + (1-x)\ln(1-\theta)] \\ &= \exp\left[x\ln\left(\frac{\theta}{1-\theta}\right) + \ln(1-\theta) + 0\right], \end{aligned}$$

so $f(x)$ has the form of an exponential family, where $k(x) = x \Rightarrow Y = \sum_{i=1}^n k(X_i) = \sum_{i=1}^n X_i$ is complete sufficient for θ. Further,

$$E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n \theta = n\theta,$$

so $\bar{x} = \frac{\sum_{i=1}^n X_i}{n}$ is unbiased for $\theta \Rightarrow \bar{x}$ is a UMVUE for θ by the Lehmann-Scheffé Theorem.

2. Show that the first order statistic $X_{(1)}$ of a random sample of size n from the distribution having pdf $f(x;\theta) = e^{-(x-\theta)}$, $\theta < x < \infty$, $-\infty < \theta < \infty$, zero elsewhere, is a complete sufficient statistic for θ. Find the unique function of this statistic which is the UMVUE of θ.

Solution: $X_1, X_2, \ldots, X_n$ are iid with pdf $f(x;\theta) = e^{-(x-\theta)} \cdot I_{(\theta,\infty)}(x)$ and $F(x) = 1 - e^{-(x-\theta)}$ for $x > \theta$.

$$\prod_{i=1}^n f(x_i;\theta) = \prod_{i=1}^n e^{-(x_i-\theta)} \cdot I_{(\theta,\infty)}(x_i) = e^{-\sum x_i + n\theta} I_{(\theta,\infty)}(x_{(1)}) = e^{-\sum x_i} e^{n\theta} I_{(\theta,\infty)}(x_{(1)}),$$

i.e. By the Neymann-Fisher Factorization Theorem (NFFT) with $k_1(x_{(1)};\theta) = e^{n\theta} I_{(\theta,\infty)}(x_{(1)})$ and $k_2(\boldsymbol{x}) = e^{-\sum x_i}$, $Y = X_{(1)}$ is a sufficient statistic of θ whose density function is

$$f_{X_{(1)}}(x) = n[1 - (1 - e^{-(x-\theta)})]^{n-1} e^{-(x-\theta)} = ne^{-n(x-\theta)}, \qquad x > \theta$$

with

$$
\begin{aligned}
E(g(x_{(1)})) = \int_\theta^\infty g(x)\cdot ne^{-n(x-\theta)}dx = n\int_\theta^\infty g(x)\cdot e^{-n(x-\theta)}dx &\doteq 0 \\
\int_\theta^\infty g(x)\cdot e^{-n(x-\theta)}dx &= 0. \qquad (1)
\end{aligned}
$$

Substituting $y = x - \theta$ into Equation (**??**), we get that

$$
\begin{aligned}
\int_0^\infty g(y+\theta)\cdot e^{-ny}dy &= 0 \\
\Rightarrow \int_0^\infty h(y)\cdot e^{-ny}dy &= 0
\end{aligned}
$$

is the form of a LaPlace transform. The only function $h(y)$ that satisfies the above function is $h(y) = g(y+\theta) = 0$. Therefore, $Y = X_{(1)}$ is complete sufficient.

Further,

$$
\begin{aligned}
E(X_{(1)}) &= \int_\theta^\infty xne^{-n(x-\theta)}dx = n\int_\theta^\infty xe^{-n(x-\theta)}dx \\
&= n\int_0^\infty (y+\theta)e^{-ny}dy \qquad \text{(letting } y = x-\theta\text{)} \\
&= n\int_0^\infty ye^{-ny}dy + n\theta\int_0^\infty e^{-ny}dy \\
&= n\left(\frac{\Gamma(2)}{n^2}\right)\int_0^\infty \frac{n^2}{\Gamma(2)}y^{2-1}e^{-ny}dy + n\theta\left(\frac{1}{n}\right)\int_0^\infty ne^{-ny}dy \\
&= \frac{1}{n} + \theta,
\end{aligned}
$$

so $Y = X_{(1)} - \frac{1}{n}$ is the UMVUE for θ by the Lehmann-Scheffé Theorem.

3. Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution with pdf $f(x;\theta) = \theta x^{\theta-1}$, $0 < x < 1$, zero elsewhere, and $\theta > 0$.

   (a) Show that the geometric mean, $(X_1X_2\cdots X_n)^{1/n}$ of the sample is a complete sufficient statistic for θ.

   (b) Find the MLE of θ. Note that it is a function of this geometric mean.

Solution:

(a)

$$
\begin{aligned}
f(x;\theta) = \theta x^{\theta-1} = \exp[\ln(\theta x^{\theta-1})] &= \exp[\ln(\theta) + (\theta-1)\ln(x)] \\
&= \exp[\ln(\theta) + \theta\ln(x) - \ln(x)],
\end{aligned}
$$

is an exponential family, where $q(\theta) = \ln(\theta)$, $p(\theta) = \theta$, $K(x) = \ln(x)$, and $S(x) = -\ln(x)$. Thus, $Y = \sum_{i=1}^n K(x_i) = \ln(x_i)$ is a complete sufficient statistic for θ. Noting that such estimates are not unique, however, we can represent the complete sufficient statistic as a function of Y, namely

$$
Z = \exp\left(\frac{Y}{n}\right) = \exp\left(\frac{1}{n}\sum_{i=1}^n \ln(x_i)\right) = \exp\left(\ln(X_1X_2\cdots X_n)^{1/n}\right) = (X_1X_2\cdots X_n)^{1/n},
$$

thus $Z = (X_1X_2\cdots X_n)^{1/n}$ is also a complete sufficient statistic of θ.

(b) The likelihood function is $L(\theta; x_1, \ldots, x_n) = \theta^n (\prod_{i=1}^n x_i)^{\theta-1}$, thus

$$\begin{aligned} \ln L(\theta; \boldsymbol{x}) &= n\ln(\theta) + (\theta-1)\sum_{i=1}^n \ln x_i \\ \frac{\partial}{\partial\theta}\ln L(\theta; \boldsymbol{x}) &= \frac{n}{\theta} + \sum_{i=1}^n \ln x_i = 0. \end{aligned} \tag{2}$$

Backsolving for θ in Equation (**??**) yields $\hat{\theta} = \frac{-n}{\sum_{i=1}^n \ln x_i} = \frac{-1}{\frac{1}{n}\sum_{i=1}^n \ln x_i} = \left(\frac{1}{n}\sum_{i=1}^n \ln x_i\right)^{-1} = (-\ln Z)^{-1}$, thus $\hat{\theta}$ is a function of the geometric mean, $Z = (X_1 X_2 \cdots X_n)^{1/n}$.

4. Let $X_1, X_2, \ldots, X_n$, $n > 2$, be a random sample from a binomial distribution $\mathrm{b}(1, \theta)$.

   (a) Show that $Y_1 = X_1 + X_2 + \ldots + X_n$ is a complete sufficient statistic for θ.

   (b) Find the function $\phi(Y_1)$ which is the UMVUE of θ.

   Solution:

   (a)

$$\begin{aligned} f(x) &= \theta^x(1-\theta)^{1-x} = \exp[\ln\{\theta^x(1-\theta)^{1-x}\}] = \exp[x\ln\theta + (1-x)\ln(1-\theta)] \\ &= \exp\left[x\ln\left(\frac{\theta}{1-\theta}\right) + \ln(1-\theta) + 0\right] \end{aligned}$$

is an exponential family where $k(x) = x$, $p(\theta) = \ln\left(\frac{\theta}{1-\theta}\right)$, $q(\theta) = \ln(1-\theta)$, and $s(x) = 0$, thus $Y_1 = \sum_{i=1}^n k(x_i) = \sum_{i=1}^n x_i$ is complete sufficient for θ.

   (b) $E(Y_1) = E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n \theta = n\theta$, therefore $\varphi(Y_1) = \frac{Y_1}{n} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$ is unbiased for θ. Thus, $\phi(Y_1) = \frac{Y_1}{n}$ is the UMVUE of θ by the Lehmann-Scheffé Theorem.

5. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that is $\mathrm{N}(0, \sigma^2 = \theta)$.

   (a) Show that $Y = \sum_{i=1}^n X_i^2$ is a complete sufficient statistic for θ.

   (b) Find the UMVUE of θ^2.

   Solution:

   (a) The normal density function

$$f(x) = \frac{1}{\sqrt{2\pi\theta}}e^{-\frac{1}{2\sigma}x^2} = \exp\left[-\frac{1}{2}\ln(2\pi\theta) - \frac{1}{2\theta}x^2 + 0\right]$$

has an exponential family form where $q(\theta) = -\frac{1}{2}\ln(2\pi\theta)$, $p(\theta) = -\frac{1}{2\theta}$, $k(x) = x^2$, $s(x) = 0$, thus $Y = \sum_{i=1}^n k(x_i) = \sum_{i=1}^n x_i^2$ is complete sufficient for θ.

   (b) To find the UMVUE of θ^2, let's first consider $E(Y) = E\left(\sum_{i=1}^n X_i^2\right) = \sum_{i=1}^n E(X_i^2)$ where

$$E(X_i^2) = \mathrm{Var}(X_i) + E^2(X_i) = \theta + 0^2 = \theta.$$

Next, we consider $E(Y^2) = \text{Var}(Y) + E^2(Y)$, where $Y = \sum_{i=1}^n x_i^2$. $X_1, X_2, \ldots, X_n \sim \text{N}(0, \theta)$ iid implies that

$$\begin{aligned}
X_1, X_2, \ldots, X_n &\sim \text{N}(0,\theta) \text{ iid} \\
\frac{X_1}{\sqrt{\theta}}, \frac{X_2}{\sqrt{\theta}}, \ldots, \frac{X_n}{\sqrt{\theta}} &\sim \text{N}(0,1) \text{ iid} \\
\frac{X_1^2}{\theta}, \frac{X_2^2}{\theta}, \ldots, \frac{X_n^2}{\theta} &\sim \chi_1^2 \text{ iid} \\
\sum_{i=1}^n \frac{X_i^2}{\theta} = \frac{\sum_{i=1}^n X_i^2}{\theta} &\sim \chi_n^2, \qquad (3)
\end{aligned}$$

therefore $E\left(\frac{\sum_{i=1}^n X_i^2}{\theta}\right) = \frac{1}{\theta}E(Y) = n$ because of Equation (**??**), thus $E(Y) = n\theta$. Meanwhile, $\text{Var}\left(\frac{\sum_{i=1}^n X_i^2}{\theta}\right) = \frac{1}{\theta^2}\text{Var}(Y) = 2n$ by Equation (**??**), therefore $\text{Var}(Y) = 2n\theta$. Thus,

$$E(Y^2) = 2n\theta^2 + (n\theta)^2 = 2n\theta^2 + n^2\theta^2,$$

so $\frac{Y^2}{2n+n^2}$ is unbiased for θ^2 thus, by the Lehmann-Scheffé Theorem, $\frac{Y^2}{2n+n^2}$ is UMVUE for θ^2.

6. Let $X_1, \ldots, X_n$ are iid $\text{N}(\mu, 1)$ random variables. Find the MVUE of $\theta = \mu^2$.

Solution: By definition, the X_is have pdf $f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\mu)^2}$, so

$$\begin{aligned}
\prod_{i=1}^n f(x_i) &= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (x_i-\mu)^2} = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\left(\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + \mu^2\right)} \\
&= \underbrace{\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n x_i^2}}_{k_2(\boldsymbol{x})} \cdot \underbrace{e^{n\bar{x}\mu - \frac{n}{2}\mu^2}}_{k_1(\bar{x};\mu)}
\end{aligned}$$

where, by the Neymann-Fisher Factorization Theorem (NFFT), we find that $Y = \bar{X}$ is a sufficient statistic of μ. Further,

$$E(\bar{X}^2) = \text{Var}(\bar{X}) + E^2(\bar{X}) = \frac{1}{n} + \mu^2$$

therefore $\bar{X}^2 - \frac{1}{n}$ is unbiased for μ^2. Thus, by the Rao-Blackwell Theorem, $\bar{X}^2 - \frac{1}{n}$ is MVUE for $\theta = \mu^2$.

Alternatively, we can further note that

$$\begin{aligned}
\prod_{i=1}^n f(x_i) &= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (x_i-\mu)^2} \\
&= \exp\left[\ln\left(\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (x-\mu)^2}\right)\right] \\
&= \exp\left[-\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^n (x_i-\mu)^2\right],
\end{aligned}$$

where

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i-\mu)^2 &= \sum_{i=1}^{n}[(x_i-\bar{x})+(\bar{x}-\mu)]^2 \\
&= \sum_{i=1}^{n}[(x_i-\bar{x})^2+(\bar{x}-\mu)^2+2(x_i-\bar{x})(\bar{x}-\mu)] \\
&= \sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu)^2 \\
&= \sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}^2-2\bar{x}\mu+\mu^2) \\
&= \sum_{i=1}^{n}(x_i-\bar{x})^2+n\bar{x}^2-2n\bar{x}\mu+n\mu^2,
\end{aligned}
$$

therefore

$$
L(\mu,x)=\prod_{i=1}^{n}f(x_i)=\exp\left[-\frac{n}{2}\ln(2\pi)-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2-\frac{n}{2}\bar{x}^2+n\bar{x}\mu-\frac{n}{2}\mu^2\right]
$$

has the form of an exponential family where $s(x)=-\frac{n}{2}\ln(2\pi)-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2-\frac{n}{2}\bar{x}^2$, $k(\bar{x})=\bar{x}$, $p(\mu)=n\mu$, $q(\mu)=\frac{n}{2}\mu^2$ so $Y=k(\bar{X})=\bar{X}$ is a complete sufficient statistic for μ. Since $E(\bar{X}^2)=\frac{1}{n}+\mu^2$, this implies that $\bar{X}^2-\frac{1}{n}$ is unbiased for μ^2 and thus, by the Lehmann-Scheffé Theorem, $\bar{X}^2-\frac{1}{n}$ is actually UMVUE for $\theta=\mu^2$.