Name: Carson Slater
Course: Theory of Statistics 2
Instructor: Dr. Harvill

Homework 1

1.

Let $X_1, X_2, ..., X_n \overset{\text{iid}}{\sim} f(x|\boldsymbol{\theta})$ such that $f(x|\boldsymbol{\theta})$ belongs to an exponential family, given by

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})\exp\left\{\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\right\},$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_d), d \leq k$. By the factorization theorem, a statistic $T(\boldsymbol{X})$ is a sufficient statistic for $\boldsymbol{\theta}$ if and only if there exists function $g(t|\theta)$ and $h(\boldsymbol{x})$ such that, for all sample points $\boldsymbol{x}$ and parameter points $\boldsymbol{\theta}$, $f(\boldsymbol{x}|\boldsymbol{\theta}) = g(T(\boldsymbol{x})|\boldsymbol{\theta})h(\boldsymbol{x})$. Let $f(\boldsymbol{x}|\boldsymbol{\theta})$ by the joint distribution of $X_1, X_2, ..., X_n$. We then have, by properties of independent random variables,

$$\begin{aligned}
f(\boldsymbol{x}|\boldsymbol{\theta}) &= \prod_{j=1}^{n} f(x_j|\boldsymbol{\theta}) \\
&= \left(\prod_{j=1}^{n} h(x_j)c(\boldsymbol{\theta})\right)\exp\left\{\sum_{j=1}^{n}\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x_j)\right\} \\
&= \left(\prod_{j=1}^{n} h(x_j)\right)(c(\boldsymbol{\theta}))^n \exp\left\{\sum_{i=1}^{k} w_i(\boldsymbol{\theta})\sum_{j=1}^{n} t_i(x_j)\right\} \\
&= \left(\prod_{j=1}^{n} h(x_j)\right)(c(\boldsymbol{\theta}))^n \exp\left\{w_1(\boldsymbol{\theta})\sum_{j=1}^{n} t_1(x_j)\right\}\ldots\exp\left\{w_k(\boldsymbol{\theta})\sum_{j=1}^{n} t_k(x_j)\right\}.
\end{aligned}$$

This implies $f(\boldsymbol{x}|\boldsymbol{\theta}) = g(T(\boldsymbol{x})|\boldsymbol{\theta})h(\boldsymbol{x})$ where,

$$\begin{cases}
h(\boldsymbol{x}) = \prod_{j=1}^{n} h(x_j) \\
g(T(\boldsymbol{x})|\boldsymbol{\theta}) = (c(\boldsymbol{\theta}))^n \exp\left\{\sum_{i=1}^{k} w_i(\boldsymbol{\theta})\sum_{j=1}^{n} t_i(x_j)\right\},
\end{cases}$$

which yields

$$T(\boldsymbol{X}) = \left(\sum_{j=1}^{n} t_1(X_j), \sum_{j=1}^{n} t_2(X_j), \ldots, \sum_{j=1}^{n} t_k(X_j)\right)$$

as a sufficient statistic for $\boldsymbol{\theta}$ by the factorization method.

2.

a.

Let $\boldsymbol{x} = (X_1, X_2, \ldots, X_n)$ be the result of $N = n$ independent, identically distributed Bernoulli(θ) trials, and let $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$. We then know that the PMF of the sum of n Bernoulli(θ) trials can be written as

$$\begin{aligned}
P(T = t) &= P(\sum_{i=1}^{n} X_i = t) \\
&= \sum_{\text{all combinations}} P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \quad \text{(since } X_i\text{'s IID)} \\
&= \binom{n}{\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\
&= \binom{n}{\sum_{i=1}^{n} x_i} \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \\
&= \binom{n}{t} \theta^{t}(1-\theta)^{n-t} \\
&= f(\boldsymbol{x}|\theta)
\end{aligned}$$

So $\sum_{i=1}^{n} X_i \sim \text{Bernoulli}\left(n, \sum_{i=1}^{n} x_i\right)$ Using the ratio of two sample points, $\boldsymbol{x}$, and $\boldsymbol{y}$, where $t_{\boldsymbol{x}} = T(\boldsymbol{X})$ and $t_{\boldsymbol{y}} = T(\boldsymbol{Y})$

$$\begin{aligned}
\frac{f(t_{\boldsymbol{x}}|\theta)}{f(t_{\boldsymbol{y}}|\theta)} &= \frac{\binom{n_1}{t_{\boldsymbol{x}}}\theta^{t_{\boldsymbol{x}}}(1-\theta)^{n_1 - t_{\boldsymbol{x}}}}{\binom{n_2}{t_{\boldsymbol{y}}}\theta^{t_{\boldsymbol{y}}}(1-\theta)^{n_2 - t_{\boldsymbol{y}}}} \\
&= \frac{\binom{n_1}{t_{\boldsymbol{x}}}}{\binom{n_2}{t_{\boldsymbol{y}}}} \theta^{t_{\boldsymbol{x}} - t_{\boldsymbol{y}}}(1-\theta)^{(n_1-n_2)-(t_{\boldsymbol{x}}-t_{\boldsymbol{y}})}.
\end{aligned}$$

$\dfrac{f(t_{\boldsymbol{x}}|\theta)}{f(t_{\boldsymbol{y}}|\theta)}$ is constant as a function with respect to θ if $T(\boldsymbol{X}) = T(\boldsymbol{Y})$ and $n_1 = n_2$. So then the pair (X, N) is minimally sufficient. Also, the distribution of N is not contingent on θ, as the distribution of N is for some probability p_i at each $i = 1, 2, \ldots$ Therefore, N is an ancillary statistic for θ.

b.

To show that $\frac{X}{N}$ is unbiased, we can find

$$\begin{aligned}
E\left[\frac{X}{N}\right] &= \sum_{j}^{\infty}\sum_{i=1}^{n}\left(\frac{x}{n}\right)p_j\binom{n}{x}\theta^x(1-\theta)^{n-x} \\
&= \sum_{j}^{\infty}\frac{p_j}{n}\underbrace{\sum_{i=1}^{n}x_i\binom{n}{x_i}\theta^x(1-\theta)^{n-x_i}}_{E[X|N=n]=n\theta} \\
&= \sum_{j}^{\infty}\frac{p_j}{n}n\theta \\
&= \theta.
\end{aligned}$$

Additionally,

$$\begin{aligned}
\text{Var}\left[\frac{X}{N}\right] &= E\left[\left(\frac{X}{N}\right)^2\right] - E\left[\frac{X}{N}\right]^2 \\
&= \sum_{j}^{\infty}\frac{p_j}{n^2}\underbrace{\sum_{i=1}^{n}x_i^2\binom{n}{x_i}\theta^x(1-\theta)^{n-x_i}}_{\text{Var}[X|N]+E[X|N]^2} -\theta^2 \\
&= \sum_{j=1}^{\infty}\left(\frac{p_j}{n^2}(n\theta(1-\theta)\right) - \theta^2 \\
&= \sum_{j=1}^{\infty}\frac{p_j\theta(1-\theta)}{n} + \theta^2 - \theta^2 \\
&= E[N^{-1}]\theta(1-\theta).
\end{aligned}$$

Name: Carson Slater
Course: Theory of Statistics 2
Instructor: Dr. Harvill

3.

Let $X_1, \dots, X_n \overset{iid}{\sim} \mathrm{N}(\theta, a\theta^2)$ a is known and constant, and $\theta > 0$.

a.

We have that this family of distributions can be written as

$$\begin{aligned} f(x|\theta) &= \frac{1}{\sqrt{2\pi a\theta^2}} \exp\left\{-\frac{(x-\theta)^2}{2a\theta^2}\right\} \\ &= \frac{1}{\sqrt{2\pi a\theta^2}} \exp\left\{\frac{-x^2 + 2x\theta - \theta^2}{2a\theta^2}\right\} \\ &= \frac{1}{\sqrt{2\pi a}\,\theta} \exp\left\{\frac{-x^2}{2a\theta^2} + \frac{x}{a\theta} - \frac{1}{2a}\right\} \end{aligned}$$

We can write this normal family as an exponential family, denoting the following components,

$$\begin{aligned} h(x) &= \frac{1}{\sqrt{2\pi}} I_{(-\infty,\infty)}(x) & c(\theta) &= \frac{1}{\sqrt{a}\theta} \exp\left\{-\frac{1}{2a}\right\} \\ t_1(x) &= -\frac{x^2}{2} & w_1(\theta) &= \frac{1}{a\theta^2} \\ t_2(x) &= x & w_2(\theta) &= \frac{1}{a\theta}. \end{aligned}$$

Here we observe that the functions w_1 and w_2 are both only functions of vectors that have one element, which is θ (one-dimensional). This shows that the parameter space does not contain a two-dimensional open set, but rather a one-dimensional open set where $\theta > 0$.

b.

By the factorization theorem, we can show that $T(\boldsymbol{X}) = (n\bar{x}, \sum_{i=1}^n x_i^2)$ is a sufficient statistic for θ. This is because the joint density function of $\boldsymbol{X}$ is

$$\begin{aligned} f(\boldsymbol{x}|\theta) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi a}\,\theta} \exp\left\{-\frac{x_i^2}{2a\theta^2} + \frac{x_i}{a\theta} - \frac{1}{2a}\right\} \\ &= \frac{1}{\sqrt{2\pi a}\,\theta} \exp\left\{-\frac{\sum_{i=1}^n x_i^2}{2a\theta^2} + \frac{\sum_{i=1}^n x_i}{a\theta} - \frac{n}{2a}\right\}, \end{aligned}$$

and those are the functions of x that are inseparable from θ. We also know that no sufficient statistic is unique, and that any one-to-one function of a sufficient statistic is also sufficient. Let one-to-one function

$$\boldsymbol{g}(T(\boldsymbol{X})) = \left(\frac{t_1(\boldsymbol{X})}{n}, \frac{t_2(\boldsymbol{X}) - 2t_1(\boldsymbol{X})^2 + t_1(\boldsymbol{X})^2/n}{n-1}\right),$$

where t_1 and t_2 are the respective elements of statistic $T(\boldsymbol{X})$. Because sufficient statistics are not unique, and that one-to-one functions of sufficient statistics are sufficient, we have that $\boldsymbol{g}(T(\boldsymbol{X})) = (\bar{X}, S^2)$. So $\bar{X}$ and S^2 are sufficient statistics for θ.

Name: Carson Slater
Course: Theory of Statistics 2
Instructor: Dr. Harvill

To show that $(\bar{X}, S^2)$ are not complete, we let $h(\cdot)$ be such that $E[h(T)] = 0 \; \forall \; \theta > 0$ and for all $T(\boldsymbol{X})$, which would imply $P(h(x) = 0) = 1$. Now suppose, for contradiction, that $h(T(\boldsymbol{X})) = \frac{n\bar{x}^2}{a+n} - \frac{S^2}{a}$. This implies that

$$\begin{aligned}
E\left[\frac{n\bar{x}^2}{a+n} - \frac{S^2}{a}\right] &= \frac{n}{a+n}E[\bar{x}^2] - \frac{1}{a}E[S^2] \\
&= \frac{n}{a+n}\left(\text{Var}[\bar{x}] + E[\bar{x}]^2\right) - \frac{\sigma^2}{a(n-1)}E[\frac{(n-1)S^2}{\sigma^2}] \\
&= \frac{n}{a+n}\left(a\theta^2 + \theta^2\right) - \frac{a(n-1)\theta^2}{a(n-1)} \quad \left(\text{Because} \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}\right) \\
&= \frac{\theta^2 n}{a+n}(a+1) - \theta^2 \\
&= \theta^2 - \theta^2 \\
&= 0.
\end{aligned}$$

So there exists an $h(\cdot)$ where $E\left[\frac{n\bar{x}^2}{a+n} - \frac{S^2}{a}\right] = 0$, but $E[h(T(\boldsymbol{X}))] = 0$ only here. So because $E[h(T(\boldsymbol{X}))] = 0$ does not hold for all values of $\theta > 0$ and $x \in \mathbb{R}$, then $T(\boldsymbol{X})$ is not a complete family with respect to θ.

Name: Carson Slater
Course: Theory of Statistics 2
Instructor: Dr. Harvill

4.

Let $X_1, \ldots, X_n$ be IID with a Geometric(θ) distribution. Using the factorization theorem, we can find a sufficient statistic for θ. Using the factorization theorem, we can show that $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ is a sufficient statistic for θ. We have it that the joint PMF of X is

$$\begin{aligned} f(\boldsymbol{x}|\theta) &= \prod_{i=1}^n \theta(1-\theta)^{x_i - 1} \\ &= \theta^n (1-\theta)^{\sum_{i=1}^n x_i - n} \\ &= \theta^n \frac{(1-\theta)^{\sum_{i=1}^n x_i}}{(1-\theta)^n}. \\ \implies f(\boldsymbol{x}|\theta) &= g(T(\boldsymbol{x})|\theta)h(\boldsymbol{x}), \end{aligned}$$

where $h(\boldsymbol{x}) = I_{(1,2,\ldots)}(\boldsymbol{x})$ and $g(T(\boldsymbol{x})|\theta) = \theta^n \frac{(1-\theta)^{T(\boldsymbol{X})}}{(1-\theta)^n} = \frac{\theta^n}{(1-\theta)^n}(1-\theta)^{\sum_{i=1}^n x_i}$. So $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$ is a sufficient statistic for θ by the factorization theorem.

To find a family of distributions for $T(\boldsymbol{X}) = \sum_{i=1}^n X_i$, we can also rewrite the distribution function for $X_i \overset{iid}{\sim}$ Geometric(θ) as,

$$f(x|\theta) = \theta(1-\theta)^{x-1}$$

So this joint PDF is an exponential family, and be written using the following components,

$$\begin{aligned} h(x) &= I_{(-\infty,\infty)}(\boldsymbol{x}) & c(\theta) &= \left(\frac{\theta}{1-\theta}\right) \\ t(x) &= x & w(\theta) &= \ln(1-\theta). \end{aligned}$$

Since the family is an exponential family, where $\theta \in \mathbb{R}$, where $d = k = 1$, then $\sum_{i=1}^n x_i$ is a complete family.

5.

Consider the following distributions, where $X = 0, 1, 2$ is a random variable.

	$P(X=0)$	$P(X=1)$	$P(X=2)$	
Distribution 1	p	$3p$	$1-4p$	$0 < p < 0.25$
Distribution 2	p	p^2	$1-p-p^2$	$0 < p < 0.5$

To show that the family of distributions of X is complete, let $T(\boldsymbol{X}) = X$ be a sufficient statistic for parameter p. Consider distribution 1. Let $g_1(x)$ be any function of X such that at each $X = 0, 1, 2$, $E_p[g_1(x)|p] = 0$ for all $0 < p < 0.25$. It follows that

$$\begin{aligned} 0 &= E_p[g_1(x)|p] \\ &= \sum_{i=0}^{2} g_1(x_i)P(x_i) \\ &= g_1(0)p + g_1(1)3p + g_1(2)(1-4p) \\ \iff 0 &= p(g_1(0) + 3g_1(1) + 4g_1(2)) - g_1(2). \end{aligned}$$

Here, $p > 0$, but $g_1(x)$ does not have to equal zero at each X for all p. A counterexample would be if $g_1(0) = 1$, $g_1(1) = -\frac{1}{3}$, and $g_1(2) = 0$. Hence, this family of distributions is not complete, as $P(g_1(x) = 0) \neq 1$.

Now consider distribution 2. Let $g_2(x)$ be any function of X such that at each $X = 0, 1, 2$, $E_p[g_2(x)|p] = 0$ for all $0 < p < 0.5$. It follows that

$$\begin{aligned} 0 &= E_p[g_2(x)|p] \\ &= \sum_{i=0}^{2} g_2(x_i)P(x_i) \\ &= (g_2(1) - g_2(2))p^2 + (g_2(0) - g_2(2))p + g_2(2). \end{aligned}$$

Since $0 < p < 0.5$, the only way this equality would be true is if $g_2(0) = g_2(1) = g_2(2) = 0 \implies P(g_2(x) = 0) = 1$. Hence, distribution 2 is a complete distribution, with X as a sufficient statistic for p.

Name: Carson Slater
Course: Theory of Statistics 2
Instructor: Dr. Harvill

6.

Let $X_1, X_2, \dots, X_n$ be an IID random sample from the PDF $f(x|\mu) = e^{-(x-\mu)}$.

a.

Let $Y = T(\boldsymbol{x}) = X_{(1)} = \min(X_i)$. By properties of order statistics, we have it that the PDF of Y is

$$\begin{aligned} f(y|\mu) &= n\left(1 - e^{\mu}\int_{\mu}^{y}\frac{1}{e^t}dt\right)^{n-1} e^{\mu-y} \\ &= ne^{n(\mu-y)} \qquad -\infty < \mu < y < \infty, \end{aligned}$$

which cannot be written as an exponential family, as the support is a function of the parameter μ. Instead, using the factorization theorem, we can write $f(y|\mu) = g(T(\boldsymbol{x})|\mu)h(\boldsymbol{x})$, where

$$g(T(\boldsymbol{x})|\mu) = ne^{n(\mu-y)}I_{(\mu,\infty)}(\min(X_i)) \qquad\qquad h(\boldsymbol{x}) = 1.$$

Hence $X_{(1)}$ is a sufficient statistic for μ.

Now let $g : \langle\mu, \infty\rangle \mapsto \mathbb{R}$ be a function such that $E[g(Y)] = 0$ for all $\mu \in \mathbb{R}$. This also implies that

$$\begin{aligned} 0 &= E[g(Y)] \\ &= \frac{\partial}{\partial\mu}E[g(Y)] \\ &= \frac{\partial}{\partial\mu}\int_{\mu}^{\infty} g(y)e^{n(\mu-y)}dy \qquad \left(\frac{\partial}{\partial\mu} = ne^{n\mu}, \text{then divide by } n \text{ on both sides.}\right) \\ &= e^{n\mu}\frac{\partial}{\partial\mu}\int_{\mu}^{\infty} g(y)e^{-ny}dy + \underbrace{e^{n\mu}\int_{\mu}^{\infty} g(y)e^{-ny}dy}_{\text{(Recall } E[g(y)]=0)} \\ &= e^{n\mu}\frac{\partial}{\partial\mu}\int_{\mu}^{\infty} g(y)e^{-ny}dy \end{aligned}$$

Substituting $t = -y$ and $dt = -dy$, we can rewrite the integral as

$$\begin{aligned} 0 &= e^{n\mu}\int_{-\infty}^{-\mu} g(-t)e^{nt}dt. \\ &= (e^{n\mu})g(-t)(e^{-\mu n}) \qquad \left(\text{By Fundamental Thm. of Calculus, also divided by } -\frac{1}{n} \text{ on both sides.}\right) \\ &= g(y). \quad \text{(Substituted back } y = -t) \end{aligned}$$

So $g(y)$ must be equal to zero for all $\mu < y < \infty$ and $-\infty < \mu < \infty$. Therefore $Y = X_{(1)}$ is a complete statistic for μ.

b.

We have that $X_{(1)}$ is a complete sufficient statistic, but we can use the ratio of $f(\boldsymbol{x}|\mu)$ and $f(\boldsymbol{y}|\mu)$, where $\boldsymbol{x}$ and $\boldsymbol{y}$ are random samples, to show $X_{(1)}$ is a complete minimally sufficient statistic using Bahadur's Theorem. To show a minimal sufficient statistic exists we have

$$\begin{aligned}\frac{f(\boldsymbol{x}|\mu)}{f(\boldsymbol{x}|\mu)} &= \frac{\exp\{n\mu - \sum_{i=1}^n x_i\}}{\exp\{n\mu - \sum_{i=1}^n y_i\}} \\ &= \frac{\exp\{\sum_{i=1}^n x_i\}}{\exp\{\sum_{i=1}^n y_i\}}\end{aligned}$$

Which implies a minimal sufficient statistic exists, since $\dfrac{f(\boldsymbol{x}|\mu)}{f(\boldsymbol{x}|\mu)}$ is constant as a function of μ when $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ for $T(\boldsymbol{X}) = \boldsymbol{x} = \boldsymbol{y} = T(\boldsymbol{X})$. So based on Bahadur's Theorem, since a minimally sufficient statistic exists, we know $X_{(1)}$ is a complete, minimally sufficient statistic. Now consider S^2,

$$\begin{aligned}S^2 &= \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \frac{1}{n-1}\sum_{i=1}^n [(x_i - \mu) - (\bar{x} - \mu))]^2 \quad \text{(Adding a `well chosen' } \mu) \\ &= \frac{1}{n-1}\sum_{i=1}^n (z - \bar{z})^2 \quad (\text{Let } z = \bar{x} - \mu)\end{aligned}$$

Since z and $\bar{z}$ have sampling distributions that do not depend on μ, we can say that S^2 is ancillary for μ. Therefore, by Basu's Theorem, $X_{(1)}$ and S^2 are independent.

Name: Carson Slater
Course: Theory of Statistics 2
Instructor: Dr. Harvill

7.

Let T_1 be sufficient and T_2 be a minimally sufficient statistic for θ. Also, let U be an unbiased estimate of θ. Also, let $U_1 = E[U|T_1]$ and $U_2 = E[U|T_2]$.

a.

We know that

$$\begin{aligned} U_2 &= E[U|T_2] \\ &= E[E[U_1|T_1]|T_2] \\ &= E[E[U_1|T_2]|T_1] \quad (T_1 \text{ is a function of } T_2\text{, so if } T_2 \text{ is given, } T_1 \text{ is known.}) \\ &= E[U_1|T_2]. \quad (\text{By definition of } U_1) \end{aligned}$$

b.

Using formulas for conditional variance, we have that

$$\begin{aligned} \text{Var}[U_1] &= \text{Var}(E[U_1|T_2]) + \underbrace{E[\text{Var}(U_1|T_2)]}_{\text{Always } (+)} \\ &\geq \text{Var}(E[U_1|T_2]) \\ &= \text{Var}(E[E[U_1|T_1]|T_2]) \\ &= \text{Var}[U_2]. \end{aligned}$$