Stat 411 – Review problems for Exam 2 ***Solutions***

1. (a) Geometric distribution: $L(\theta) = \theta^n(1-\theta)^{\sum_{i=1}^n X_i - n}$, so $T = \sum_{i=1}^n X_i$ is a (minimal) sufficient statistic for θ.

   (b) Truncated scale exponential distribution: $L(\theta) = e^{-(1/\theta)\sum_{i=1}^n X_i}/\{\theta(1-e^{-7/\theta})\}^n$, so $T = \sum_{i=1}^n X_i$ is a (minimal) sufficient statistic for θ.

   (c) Gamma distribution: $L(\theta) = (\prod_{i=1}^n X_i)^{\theta_1 - 1} e^{-(1/\theta_2)\sum_{i=1}^n X_i}/\{\theta_2^{\theta_1}\Gamma(\theta_1)\}^n$, so $T = (\prod_{i=1}^n X_i, \sum_{i=1}^n X_i)$ is a (jointly minimal) sufficient statistic for (θ_1, θ_2).

   (d) Location-scale exponential distribution: $L(\theta) = (1/\theta_2)^n e^{-\sum_{i=1}^n X_i + n\theta} I_{[\theta_1,\infty)}(X_{(1)}$, so $T = (X_{(1)}, \sum_{i=1}^n X_i)$ is a (jointly minimal) sufficient statistic for (θ_1, θ_2).

2. A naive unbiased estimator of θ^2 is $X_1 X_2$. Also, $T = \sum_{i=1}^n X_i$ is a complete sufficient statistic. Use the Rao–Blackwell strategy to get the MVUE.

$$\begin{aligned}
\mathsf{E}_\theta(X_1X_2 \mid T = t) &= \mathsf{P}_\theta(X_1 = 1, X_2 = 1 \mid T = t) \\
&= \frac{\mathsf{P}_\theta(X_1 = 1, X_2 = 1, X_3 + \cdots + X_n = t-2)}{\mathsf{P}_\theta(T = t)} \\
&= \frac{\theta \cdot \theta \cdot \binom{n-2}{t-2}\theta^{t-2}(1-\theta)^{n-t}}{\binom{n}{t}\theta^t(1-\theta)^{n-t}} \\
&= \binom{n-2}{t-2} \div \binom{n}{t} \\
&= \frac{t(t-1)}{n(n-1)}.
\end{aligned}$$

   This is unbiased for θ^2 and a function of T; therefore, it's the unique MVUE.

3. The sample total $T = \sum_{i=1}^n X_i$ and, hence $\overline{X} = T/n$, is a complete sufficient statistic for θ in the Poisson problem. A reasonable guess for the MVUE of θ^2 is $\overline{X}^2$. Its expected value is $\mathsf{E}_\theta(\overline{X}^2) = \mathsf{V}_\theta(\overline{X}) + \mathsf{E}_\theta(\overline{X})^2 = \theta/n + \theta^2$. So $\overline{X}^2$ isn't the MVUE. But what about $\overline{X}^2 - \overline{X}/n$? The expected value is

$$\mathsf{E}_\theta(\overline{X}^2 - \overline{X}/n) = \mathsf{E}_\theta(\overline{X}^2) - \mathsf{E}_\theta(\overline{X}/n) = \theta/n + \theta^2 - \theta/n = \theta^2.$$

   Since $\overline{X}^2 - \overline{X}/n$ is an unbiased estimator of θ^2 based on the complete sufficient statistic, the Lehmann–Scheffe theorem ensures that it's the MVUE.

4. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathsf{N}(\theta, 1)$. From the Homework, we know that $\overline{X}^2 - 1/n$ is the MVUE of $\theta^2 > 0$. But there's nothing to say that $|\overline{X}| < n^{-1/2}$ can't happen; therefore, it is possible that the MVUE of θ^2 is negative.

5. (a) Negative binomial (with fixed r) is an exponential family:

$$\begin{aligned}
f_\theta(x) &= \binom{r+x-1}{r-1}\theta^r(1-\theta)^x \\
&= \exp\Big\{\log(1-\theta)x + \log\binom{r+x-1}{r-1} + r\log\theta\Big\}.
\end{aligned}$$

Since $K(x) = x$ in this case, the complete sufficient statistic for θ based on an iid sample $X_1, \ldots, X_n$ is $T = \sum_{i=1}^n X_i$.

(b) The two-parameter gamma distribution is an exponential family:

$$\begin{aligned} f_\theta(x) &= \frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} x^{\theta_1 - 1} e^{-x/\theta_2} \\ &= \exp\{(\theta_1 - 1)\log x + (-1/\theta_2)x - \theta_1 \log\theta_2 - \log\Gamma(\theta_1)\}. \end{aligned}$$

This is a two-parameter exponential family with $p_1(\theta) = \theta_1 - 1$, $p_2(\theta) = -1/\theta_2$, $K_1(x) = \log x$ and $K_2(x) = x$. Therefore, the (jointly) complete and sufficient statistic for (θ_1, θ_2) is $T = (\sum_{i=1}^n \log X_i, \sum_{i=1}^n X_i)$.

6. It can be shown, using moment-generating functions, for example, that $T = \sum_{i=1}^n X_i$ has a $\mathsf{Pois}(n\theta)$ distribution. Therefore, T has PMF

$$e^{-n\theta}(n\theta)^t/t! = \exp\{\theta t + t\log n - \log t! - n\theta\}.$$

This is of the exponential family form with $p(\theta) = \theta$, $K(t) = t$, $S(t) = t\log n - \log t!$ and $q(\theta) = -n\theta$.

7. Following the hint, if $X_i \sim \mathsf{Unif}(\theta, \theta+1)$, then we may write $X_i = \theta + Z_i$, where $Z_i \sim \mathsf{Unif}(0,1)$. Similarly, $X_{(1)} = \theta + Z_{(1)}$ and $X_{(n)} = \theta + Z_{(n)}$. We know that $Z_{(1)} \sim \mathsf{Beta}(1, n)$ and $Z_{(n)} \sim \mathsf{Beta}(n, 1)$. Factorization theorem shows that $T = (X_{(1)}, X_{(n)})$ is a sufficient statistic for θ. However, since

$$\mathsf{E}_\theta(X_{(1)}) = \theta + \frac{1}{n+1} \quad \text{and} \quad \mathsf{E}_\theta(X_{(n)}) = \theta + \frac{n}{n+1},$$

if we set $h(T) = X_{(n)} - X_{(1)} - (n-1)/(n+1)$, then for any θ

$$\mathsf{E}_\theta[h(T)] = \mathsf{E}_\theta(X_{(n)}) - \mathsf{E}_\theta(X_{(1)}) - \frac{n-1}{n+1} = \theta + \frac{n}{n+1} - \theta - \frac{1}{n+1} - \frac{n-1}{n+1} = 0.$$

We've found a non-constant function $h(t)$ such that $\mathsf{E}_\theta[h(T)] = 0$ for all θ; therefore, T is not complete.

8. In Homework 06, we showed that $X_{(1)}$ is a complete sufficient statistic for θ. This is also a location parameter problem, so $X_{(n)} - X_{(1)}$ is an ancillary statistic. By Basu's theorem, $X_{(1)}$ and $X_{(n)} - X_{(1)}$ are independent. Therefore,

$$0 = \mathsf{C}_\theta(X_{(1)}, X_{(n)} - X_{(1)}) = \mathsf{C}_\theta(X_{(1)}, X_{(n)}) - \mathsf{V}_\theta(X_{(1)}),$$

which implies $\mathsf{C}_\theta(X_{(1)}, X_{(n)}) = \mathsf{V}_\theta(X_{(1)})$. Because it's a location parameter problem, we can further write $X_{(1)} = \theta + Z_{(1)}$, where $Z_{(1)}$ is the minimum of n iid $\mathsf{Exp}(1)$ random variables. It can be shown (via CDF calculations) that $Z_{(1)} \sim \mathsf{Exp}(1/n)$ and, furthermore,

$$\mathsf{C}_\theta(X_{(1)}, X_{(n)}) = \mathsf{V}_\theta(X_{(1)}) = \mathsf{V}_\theta(\theta + Z_{(1)}) = 1/n^2.$$

9. Write $X_i = \theta Z_i$ where $Z_i \sim \mathsf{N}(0,1)$. This is an exponential family distribution so $T = X_1^2 + \cdots + X_n^2$ is a complete sufficient statistic; moreover, since it's a scale parameter problem, $U = X_1^2/(X_1^2 + \cdots + X_n^2)$ is an ancillary statistic. The trick is to look at

$$\mathsf{E}_\theta(X_1^2) = \mathsf{E}_\theta\Big\{\frac{X_1^2}{X_1^2 + \cdots X_n^2} \cdot (X_1^2 + \cdots + X_n^2)\Big\}.$$

The right-hand side can be factored as a product of expected values by Basu's theorem, i.e., Basu says T and U are independent, so $\mathsf{E}_\theta(UT) = \mathsf{E}_\theta(U)\mathsf{E}_\theta(T)$. Then $\mathsf{E}_\theta(U)$ is the quantity we want, and it equals

$$\mathsf{E}_\theta\Big\{\frac{X_1^2}{X_1^2 + \cdots X_n^2}\Big\} = \frac{\mathsf{E}_\theta(X_1^2)}{\mathsf{E}_\theta(X_1^2 + \cdots + X_n^2)} = \frac{\theta^2}{n\theta^2} = \frac{1}{n}.$$