MATH 503: Mathematical Statistics
Dr. Kimberly F. Sellers, Instructor
Homework 5 Solutions

1. **Let $X_1, X_2, \ldots, X_n$ be a random sample from each of the following distributions involving the parameter θ. In each case, find the MLE of θ and show that it is a sufficient statistic for θ and hence a minimal sufficient statistic.**

   (a) Binomial(1,θ), where $0 < \theta < 1$.

   (b) Poisson with mean $\theta > 0$.

   (c) Gamma with $\alpha = 3$ and $\beta = \theta > 0$.

   (d) N$(\theta, 1)$ where $-\infty < \theta < \infty$.

   (e) N$(0, \theta)$ where $0 < \theta < \infty$.

   Solutions:

   (a)

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

   which implies that $\hat{\theta} = \bar{x}$ is the MLE. $\bar{X}$ is sufficient because

$$\prod_{i=1}^n f(x_i) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} = \theta^{n\bar{x}}(1-\theta)^{n-n\bar{x}} = \underbrace{\left(\frac{\theta}{1-\theta}\right)^{n\bar{x}}(1-\theta)^n}_{k_1(\bar{x};\theta)} \cdot \underbrace{1}_{k_2(\boldsymbol{x})},$$

   i.e. the Neymann-Fisher Factorization Theorem (NFFT) applies, hence $\bar{X}$ is minimal sufficient.

   (b)

$$\begin{aligned}
f(x) &= \frac{e^{-\theta}\theta^x}{x!}, x = 0, 1, 2, \ldots \\
L(\theta; \boldsymbol{x}) &= \frac{e^{-n\theta}\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n (x_i!)} \\
\ln L(\theta; \boldsymbol{x}) &= -n\theta + \left(\sum_{i=1}^n x_i\right)\ln\theta - \sum_{i=1}^n \ln(x_i!) \\
\frac{\partial}{\partial\theta}\ln L(\theta; \boldsymbol{x}) &= -n + \frac{\sum_{i=1}^n x_i}{\theta} = 0,
\end{aligned}$$

which implies that $\hat{\theta} = \bar{x}$. $\bar{X}$ is sufficient because

$$\prod_{i=1}^{n} f(x_i) = e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}\left(\frac{1}{\prod_{i=1}^{n}(x_i!)}\right) = \underbrace{e^{-n\theta}\theta^{n\bar{x}}}_{k_1(\bar{x};\theta)} \cdot \underbrace{\left(\frac{1}{\prod_{i=1}^{n}(x_i!)}\right)}_{k_2(\boldsymbol{x})},$$

i.e. the Neymann-Fisher Factorization Theorem (NFFT) applies, hence $\bar{X}$ is minimal sufficient.

(c)

$$\begin{aligned}
f(x) &= \frac{1}{\Gamma(3)\theta^2} x^{3-1} e^{-x/\theta} = \frac{1}{2\theta^3} x^2 e^{-x/\theta}; x > 0 \\
L(\theta; \boldsymbol{x}) &= \left(\frac{1}{2\theta^3}\right)^n \left(\prod_{i=1}^{n} x_i\right)^2 e^{-\sum_{i=1}^{n} x_i/\theta} \\
\ln L(\theta; \boldsymbol{x}) &= -n\ln(2\theta^3) + 2\sum_{i=1}^{n} \ln(x_i) - \frac{\sum_{i=1}^{n} x_i}{\theta} = -\ln 2 - 3n\ln\theta + 2\sum_{i=1}^{n}\ln x_i - \frac{\sum_{i=1}^{n} x_i}{\theta} \\
\frac{\partial}{\partial\theta}\ln L(\theta; \boldsymbol{x}) &= \frac{-3n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} = 0,
\end{aligned}$$

which implies that $\hat{\theta} = \frac{\bar{x}}{3}$. $Y = \frac{\bar{X}}{3}$ is sufficient because

$$\begin{aligned}
\prod_{i=1}^{n} f(x_i) &= \left(\frac{1}{2\theta^3}\right)^n \left(\prod_{i=1}^{n} x_i\right)^2 e^{-\sum_{i=1}^{n} x_i/\theta} = \left(\frac{1}{2\theta^3}\right)^n \left(\prod_{i=1}^{n} x_i\right)^2 e^{-3n\bar{X}/(3\theta)} \\
&= \underbrace{\left(\frac{1}{2\theta^3}\right)^n e^{-3nY/\theta}}_{k_1(Y=\frac{\bar{X}}{3};\theta)} \cdot \underbrace{\left(\prod_{i=1}^{n} x_i\right)^2}_{k_2(\boldsymbol{x})},
\end{aligned}$$

i.e. the Neymann-Fisher Factorization Theorem (NFFT) applies, hence $\hat{\theta} = \frac{\bar{X}}{3}$ is minimal sufficient.

(d)

$$\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}, -\infty < x < \infty \\
L(\theta; \boldsymbol{x}) &= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2} \\
\ln L(\theta; \boldsymbol{x}) &= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2 \\
\frac{\partial}{\partial\theta}\ln L(\theta; \boldsymbol{x}) &= \sum_{i=1}^{n}(x_i-\theta) = 0
\end{aligned}$$

implies that $\hat{\theta} = \bar{x}$. $\bar{X}$ is sufficient because

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

i.e. the Neymann-Fisher Factorization Theorem (NFFT) applies.

(e)

$$\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{-1}{2\theta}x^2}, -\infty < x < \infty \\
L(\theta; \boldsymbol{x}) &= (2\pi\theta)^{-n/2} e^{-\frac{-1}{2\theta}\sum_{i=1}^n x_i^2} \\
\ln L(\theta; \boldsymbol{x}) &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\theta - \frac{1}{2\theta}\sum_{i=1}^n x_i^2 \\
\frac{\partial}{\partial\theta}\ln L(\theta; \boldsymbol{x}) &= -\frac{n}{2\theta} + \frac{\sum_{i=1}^n x_i^2}{2\theta^2} = 0
\end{aligned}$$

results in $\hat{\theta} = \bar{x^2}$ as the MLE. $\bar{x^2}$ is sufficient because

$$\prod_{i=1}^n f(x_i) = (2\pi\theta)^{-n/2} e^{-\frac{1}{2\theta}\sum_{i=1}^n x_i^2} = \underbrace{(2\pi\theta)^{-n/2} e^{-\frac{1}{2\theta}n\bar{x^2}}}_{k_1(\bar{x^2};\theta)} \cdot \underbrace{1}_{k_2\boldsymbol{x}}$$

thus the Neymann-Fisher Factorization Theorem (NFFT) holds, hence $\bar{X^2}$ is minimal sufficient.

2. **Let $Y_1 < Y_2 < Y_3 < Y_4$ denote the order statistics of a random sample of size $n = 4$ from a distribution having pdf $f(x;\theta) = \frac{1}{\theta}, 0 < x < \theta$, zero elsewhere, where $0 < \theta < \infty$. Argue that the complete sufficient statistic, Y_4 for θ, is independent of each of the statistics $\frac{Y_1}{Y_4}$ and $\frac{Y_1+Y_2}{Y_3+Y_4}$.**

Solution:

$$\begin{aligned}
f(x) &= \frac{1}{\theta}, 0 < x < \theta; 0 < \theta < \infty \\
F(x) &= \frac{x}{\theta}, 0 < x < \theta \\
L(\theta; \boldsymbol{x}) &= \prod_{i=1}^n f(x_i;\theta) = \frac{1}{\theta^4}\prod_{i=1}^n I_{(0,\theta)}(x_i) = \frac{1}{\theta^4}\prod_{i=1}^n I_{(x_i,\infty)}(\theta) = \frac{1}{\theta^4} I_{(x_{(4)},\infty)}(\theta) = \frac{1}{\theta^4} I_{(Y_4,\infty)}(\theta),
\end{aligned}$$

i.e. $L(\theta;\boldsymbol{x}) = \frac{1}{\theta^4}, y_4 < \theta < \infty$, where $Y_4 = X_{(4)}$ is the maximum order statistic. By definition, $L(\theta;\boldsymbol{x})$ is a decreasing function wrt θ in this support space, hence $L(\theta;\boldsymbol{x})$ is maximized at $\hat{\theta} = Y_4$, thus $\hat{\theta} = Y_4$ is the MLE of θ. Further, Y_4 is sufficient by the Neymann-Fisher Factorization Theorem where

$$\prod_{i=1}^n f(x_i;\theta) = \underbrace{\frac{1}{\theta^4} I_{(0,\theta)}(Y_4)}_{k_1(Y_4;\theta)} \cdot \underbrace{1}_{k_2(\boldsymbol{x})}.$$

Because the MLE is a sufficient statistic, we thus know that it is minimal sufficient.

Claim: Y_4 is complete.
Proof: Y_4 has the cdf $F_{Y_4}(y) = F^4(y) = \left(\frac{y}{\theta}\right)^4$ and pdf $f_{Y_4}(y) = 4\left(\frac{y}{\theta}\right)^3\left(\frac{1}{\theta}\right) = \frac{4}{\theta^4}y^3$, $0 < y < \theta$. Consider $g(Y_4)$ such that $E(g(Y_4)) = \int_0^\theta g(y)\frac{4}{\theta^4}y^3 \mathrm{d}y = \frac{4}{\theta^4}\int_0^\theta g(y)y^3\mathrm{d}y = 0 \therefore \int_0^\theta g(y)y^3\mathrm{d}y = 0$. Differentiating both sides with respect to y implies that $g(\theta)\theta^3 = 0$ where $\theta > 0$, thus $g(\theta) = 0$. Thus, Y_4 is complete.

Thus, Y_4 is complete and minimal sufficient statistic.

Claim: $\frac{Y_1}{Y_4}$ and $\frac{Y_1+Y_2}{Y_3+Y_4}$ are scale-invariant statistics.
Proof: For any d,

$$\begin{aligned}\frac{dY_1}{dY_4} &= \frac{Y_1}{Y_4} \quad \text{and} \\ \frac{dY_1+dY_2}{dY_3+dY_4} &= \frac{d(Y_1+Y_2)}{d(Y_3+Y_4)} = \frac{Y_1+Y_2}{Y_3+Y_4}. \qquad \blacksquare\end{aligned}$$

Claim: Scale-invariant statistics are ancillary.
Proof: Following the hint, let $X_i = \theta W_i \Rightarrow W_i = \frac{X_i}{\theta}$ and $\mathrm{d}W_i = \frac{\mathrm{d}X_i}{\theta}$. By univariate transformation,

$$f_X(x) = f_W\left(\frac{x}{\theta}\right)\cdot\frac{1}{\theta} = \frac{1}{\theta} \therefore f_W\left(\frac{x}{\theta}\right) = f_W(w) = 1; 0 < w < 1,$$

i.e. W does not depend on θ. Meanwhile, consider statistic

$$Z = u(Y_1, Y_2, Y_3, Y_4) = u(\theta W_1, \theta W_2, \theta W_3, \theta W_4) = u(W_1, W_2, W_3, W_4)$$

which doesn't depend on θ on the W's distribution doesn't depend on θ, thus Z is ancillary. $\blacksquare$

Y_4 is complete (and minimal) sufficient, and $\frac{Y_1}{Y_4}$ and $\frac{Y_1+Y_2}{Y_3+Y_4}$ respectively are ancillary. Thus, by Basu's Theorem, Y_4 is independent of $\frac{Y_1}{Y_4}$, and Y_4 is independent of $\frac{Y_1+Y_2}{Y_3+Y_4}$, respectively.

3. **Let $Y_1 < \cdots < Y_n$ be the order statistics of a random sample from a $\mathrm{N}(\theta, \sigma^2)$, $-\infty < \theta < \infty$, distribution. Show that the distribution of $Z = Y_n - \bar{X}$ does not depend on θ. Thus $\bar{Y} = \sum_{i=1}^n Y_i/n$, a complete sufficient statistic for θ, is independent of Z.**

Solution: Let $X_1, \ldots, X_n$ be a random sample from a $\mathrm{N}(\theta, \sigma^2)$ distribution, i.e. $f(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\theta)^2}$; and consider the corresponding order statistics $Y_1, \ldots, Y_n$. By definition, $X_i = \theta + W_i$ implies that $W_i = X_i - \theta$ where $W_i \sim \mathrm{N}(0, \sigma^2)$ which doesn't depend on θ, so similarly $Y_i = \theta + W_{(i)}$ where $W_{(i)}$ are ordered statistics of $W_1, \ldots, W_n$ where the distribution of $W_{(i)}$ likewise doesn't depend on θ.

$Z = Y_n - \bar{X}$ is a location-invariant statistic because for any d,

$$(Y_n+d) - (\overline{X+d}) = (Y_n+d) - \frac{1}{n}\sum_{i=1}^n (X_i+d) = (Y_n+d) - \frac{\sum_{i=1}^n X_i + nd}{n} = Y_n + d - \frac{\sum_{i=1}^n X_i}{n} - d = Y_n - \bar{X}.$$

In particular, $Z = Y_n - \bar{X}$ such that $(Y_n - \theta) - \frac{\sum_{i=1}^n (X_i-\theta)}{n} = W_{(n)} - \frac{\sum_{i=1}^n W_i}{n} = W_{(n)} - \bar{W}$ where the distribution of the Ws does not depend on θ, so Z (which is a function of the Ws) does not depend on θ, hence Z is ancillary, therefore (by Basu's Theorem), Z is independent of $\bar{Y}$, which is complete sufficient for θ.

Proof that $\bar{Y}$ is complete sufficient for θ: $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$ where $Y_1, \ldots, Y_n$ are order statistics of $X_1, \ldots, X_n$. We know that $\bar{X}$ is complete sufficient for θ because

$$f(x) = \exp\left[\underbrace{-\ln(\sqrt{2\pi\sigma^2}) - \frac{1}{2\sigma^2}x_i^2}_{S(x)} + \underbrace{\frac{n\theta}{\sigma^2}}_{p(\theta)}\underbrace{\frac{x}{n}}_{K(x)} + \underbrace{-\frac{\theta^2}{2\sigma^2}}_{q(\theta)}\right]$$

has the form of an exponential family. Hence $Y = \sum_{i=1}^n K(X_i) = \sum_{i=1}^n \frac{X_i}{n} = \bar{X} = \bar{Y}$ is complete sufficient for θ.

4. **Let $X_1, X_2, \ldots, X_n$ be iid with the distribution $\mathbf{N}(\theta, \sigma^2)$, $-\infty < \theta < \infty$. Prove that a necessary and sufficient condition that the statistics $Z = \sum_{i=1}^n a_i X_i$ and $Y = \sum_{i=1}^n X_i$, a complete sufficient statistic for θ, are independent is that $\sum_{i=1}^n a_i = 0$.**

   Solution: $Y = \sum_{i=1}^n X_i$ is a complete sufficient statistic for θ because

$$\begin{aligned} f(x) &= \exp\left[-\ln(\sqrt{2\pi\sigma^2}) - \frac{1}{2\sigma^2}(x^2 - 2\theta x + \theta^2)\right] \\ &= \exp\left[\underbrace{-\ln(\sqrt{2\pi\sigma^2}) - \frac{1}{2\sigma^2}x^2}_{S(x)} + \underbrace{\frac{\theta}{\sigma^2}}_{p(\theta)}\underbrace{x}_{K(x)} + \underbrace{\frac{-\theta^2}{2\sigma^2}}_{q(\theta)}\right] \end{aligned}$$

   is an exponential family so $Y = \sum_{i=1}^n K(X_i) = \sum_{i=1}^n X_i$ is complete sufficient for θ. Thus it remains to show that $Z = \sum_{i=1}^n a_i X_i$ is ancillary if and only if $\sum_{i=1}^n a_i = 0$.

   Because $X_1, \ldots, X_n \sim N(\theta, \sigma^2)$, $Z = \sum_{i=1}^n a_i X_i$ is also normally distributed with

$$\begin{aligned} E\left(\sum_{i=1}^n a_i X_i\right) &= \sum_{i=1}^n a_i E(X_i) = \sum_{i=1}^n a_i \theta = \theta \sum_{i=1}^n a_i \\ \text{Var}\left(\sum_{i=1}^n a_i X_i\right) &= \sum_{i=1}^n a_i^2 \text{Var}(X_i) = \sum_{i=1}^n a_i^2 \sigma^2 = \sigma^2 \sum_{i=1}^n a_i^2. \end{aligned}$$

   In order for Z to be ancillary, we need the distribution (i.e. the expected value) to not depend on θ, which occurs if and only if $\theta \sum_{i=1}^n a_i = 0 \Leftrightarrow \sum_{i=1}^n a_i = 0$.

5. **Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with pdf $f(x;\theta) = \frac{1}{2}\theta^3 x^2 e^{-\theta x}$, $0 < x < \infty$, zero elsewhere, where $0 < \theta < \infty$.**

   (a) Find the MLE of θ. Is it unbiased? Hint: find the pdf of $Y = \sum_{i=1}^n X_i$ and then compute $E(\hat{\theta})$.

   (b) Argue that Y is a complete sufficient statistic for θ.

   (c) Find the UMVUE of θ.

   (d) Show that $\frac{X_1}{Y}$ and Y are independent.

   (e) What is the distribution of $\frac{X_1}{Y}$?

   Solutions:

   (a)

$$\begin{aligned} f(x) &= \frac{1}{2}\theta^3 x^2 e^{-\theta x}, 0 < x < \infty \\ L(\theta; x) &= \left(\frac{1}{2}\right)^n \theta^{3n} \left(\prod_{i=1}^n x_i\right)^2 e^{-\theta \sum_{i=1}^n x_i} \\ \ln L(\theta; x) &= -n \ln 2 + 3n \ln \theta + 2\sum_{i=1}^n \ln(x_i) - \theta \sum_{i=1}^n x_i \\ \frac{\partial \ln L(\theta; x)}{\partial \theta} &= \frac{3n}{\theta} - \sum_{i=1}^n x_i = 0 \end{aligned}$$

implies that $\hat{\theta} = \frac{3n}{\sum_{i=1}^n X_i}$ is the MLE.

Consider $Y = \sum_{i=1}^n X_i$. By definition, $X_i \sim \text{Gamma}\left(3, \frac{1}{\theta}\right)$ iid, therefore $Y \sim \text{Gamma}\left(3n, \frac{1}{\theta}\right)$. $E(\hat{\theta}) = E\left(\frac{3n}{\sum_{i=1}^n X_i}\right) = 3nE\left(\frac{1}{Y}\right)$ where

$$E\left(\frac{1}{Y}\right) = \int_0^\infty \frac{1}{y}\frac{\theta^{3n}}{\Gamma(3n)}y^{3n-1}e^{-\theta y}\text{d}y = \frac{\theta^{3n}}{\Gamma(3n)}\frac{\Gamma(3n-1)}{\theta^{3n-1}} = \frac{\theta}{3n-1},$$

thus $E(\hat{\theta}) = \frac{3n\theta}{3n-1} \neq \theta$, i.e. $\hat{\theta}$ is not unbiased.

(b) $Y \sim \text{Gamma}\left(3n, \frac{1}{\theta}\right)$ has the pdf

$$\begin{aligned} f(y) &= \frac{\theta^{3n}}{\Gamma(3n)}y^{3n-1}e^{-\theta y}, 0 < y < \infty \\ &= \exp\left[\underbrace{3n\ln(\theta) - \ln(\Gamma(3n))}_{q(\theta)} + \underbrace{(3n-1)\ln(y)}_{S(y)} + \underbrace{-\theta}_{p(\theta)}\underbrace{y}_{K(y)}\right] \end{aligned}$$

which is an exponential family, so Y is complete sufficient for θ.

(c) Consider the statistic, $Z = \frac{3n-1}{3n}\hat{\theta}$:

$$E\left(\frac{3n-1}{3n}\hat{\theta}\right) = \frac{3n-1}{3n}\cdot 3nE\left(\frac{1}{Y}\right) = (3n-1)\frac{\theta}{3n-1} = \theta,$$

thus $Z = \frac{3n-1}{3n}\hat{\theta} = \frac{3n-1}{\sum_{i=1}^n X_i}$ is UMVUE of θ by the Lehmann-Scheffé Theorem (because we showed that $Y = \sum_{i=1}^n X_i$ is complete sufficient for θ in Part (b)).

(d) Depending on which version of Basu's Theorem you use (CB only requires complete sufficiency, while HMC requires complete minimal sufficiency), we first show that Y is minimal sufficient for θ (we've already shown that Y is complete sufficient): Y is minimal sufficient because

$$\frac{f(y)}{f(w)} = \frac{\frac{\theta^{3n}}{\Gamma(3n)}y^{3n-1}e^{-\theta y}}{\frac{\theta^{3n}}{\Gamma(3n)}w^{3n-1}e^{-\theta w}} = \left(\frac{y}{w}\right)^{3n-1}e^{-\theta(y-w)}$$

is constant wrt θ iff $Y = W$, therefore Y is minimal sufficient for θ. Meanwhile, $\frac{X_1}{Y}$ is a scale-invariant statistic thus it is ancillary, thus (by Basu's Theorem) $\frac{X_1}{Y}$ is independent of Y.

(e) Consider the following bivariate transformation: let

$$\begin{cases} A = \frac{X_1}{Y} \\ B = Y \end{cases} \Rightarrow \begin{cases} X_1 = AY = AB \\ Y = B \end{cases} \quad \text{thus } J = \begin{vmatrix} b & 0 \\ 0 & 1 \end{vmatrix} = b$$

thus the joint density function $g(a,b) = f_{X_1,Y}(ab,b)b = f_{X_1}(ab)f_{Y-X_1}(b-ab)b$ where

$$Y - X_1 = \sum_{i=2}^n X_i \sim \text{Gamma}\left(3(n-1), \frac{1}{\theta}\right),$$

i.e.

$$\begin{aligned}
g(a,b) &= \frac{1}{\Gamma(3)\left(\frac{1}{\theta}\right)^3}(ab)^{3-1}e^{-\theta ab}\frac{1}{\Gamma(3n-3)\left(\frac{1}{\theta}\right)^{3n-3}}(b-ab)^{(3n-3)-1}e^{-\theta(b-ab)}b \\
&= \frac{\theta^{3n}}{\Gamma(3)\Gamma(3n-3)}a^2b^2b^{3n-4}(1-a)^{3n-4}e^{-\theta b}b \\
&= \frac{\theta^{3n}}{2\Gamma(3n-3)}a^2(1-a)^{3n-4}b^{3n-1}e^{-\theta b}, b>0, 0<a<1 \\
\therefore g(a) &= \frac{\theta^{3n}}{2\Gamma(3n-3)}a^2(1-a)^{3n-4}\int_0^\infty b^{3n-1}e^{-\theta b}\mathrm{d}b \\
&= \frac{\Gamma(3n)}{\Gamma(3)\Gamma(3n-3)}a^{3-1}(1-a)^{(3n-3)-1}, 0<a<1
\end{aligned}$$

i.e. $A = \frac{X_1}{Y} \sim \text{Beta}(3, 3n-3)$.

6. **If $X_1, \ldots, X_N$ are iid Binomial(n,p) random variables, find the UMVUE of $\theta = p^n = P(X_1 = n)$.**

Solution: Noting that $\theta = p^n = P(X_1 = n)$ is the parameter of interest, let $V = \begin{cases} 1, & X_1 = n \\ 0, & \text{otherwise} \end{cases}$. By definition, $E(V) = P(X_1 = n) = \theta = p^n$, i.e. V is unbiased for θ. Meanwhile, $X_i \sim \text{Binomial}(n,p)$ iid, where

$$\begin{aligned}
f(x;p) &= \binom{n}{x}p^x(1-p)^{n-x} \\
&= \exp\left[\underbrace{\ln\binom{n}{x}}_{S(x)} + \underbrace{x}_{K(x)}\underbrace{\ln\left(\frac{p}{1-p}\right)}_{r(p)} + \underbrace{n\ln(1-p)}_{q(p)}\right]
\end{aligned}$$

is an exponential family, thus $T = \sum_{i=1}^n K(X_i) = \sum_{i=1}^n X_i$ is complete sufficient. Then, by the Rao-Blackwell Theorem, $E(V \mid T = t)$ is MVUE for θ where, by the Lehmann-Scheffé Theorem, this statistic is UMVUE for θ.

The form of $E(V \mid T = t)$ is provided below.

$$\begin{aligned}
E(V \mid T = t) &= P\left(X_1 = n \mid \sum_{i=1}^n X_i = t\right) = \frac{P\left(X_1 = n, \sum_{i=1}^N X_i = t\right)}{P\left(\sum_{i=1}^N X_i = t\right)} \\
&= \frac{P\left(X_1 = n, \sum_{i=2}^N X_i = t-n\right)}{P\left(\sum_{i=1}^N X_i = t\right)} = \frac{P(X_1 = n)P\left(\sum_{i=2}^N X_i = t-n\right)}{P\left(\sum_{i=1}^N X_i = t\right)} \\
&= \frac{p^n\binom{(N-1)n}{t-n}p^{t-n}(1-p)^{(N-1)n-(t-n)}}{\binom{Nn}{t}p^t(1-p)^{Nn-t}} = \frac{\binom{(N-1)n}{t-n}}{\binom{Nn}{t}}
\end{aligned}$$

because $\sum_{i=2}^N X_i \sim \text{Binomial}((N-1)n, p)$ and $\sum_{i=1}^N X_i \sim \text{Binomial}(Nn, p)$.