MA204: Mathematical Statistics

Tutorial 6

T6.1 UMVUE and Efficient Estimator

6.1.1 UMVUE

An estimator of the parameter $\theta \in \Theta$ is called a *uniformly minimum variance unbiased estimator* (UMVUE) if it is unbiased and has the smallest variance among all unbiased estimators of θ.

6.1.2 Efficient estimator

An unbiased estimator for θ is an *efficient estimator* if it has variance equal to the Cramér–Rao lower bound.

6.1.3 Their relationship

Obviously, an efficient estimator for θ is a UMVUE for θ. However, a UMVUE for θ does not necessarily imply that it is an efficient estimator for θ.

Example T6.1 (A special beta distribution). Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Beta}(\theta, 1)$ with pdf $f(x; \theta) = \theta x^{\theta-1} I(0 < x < 1)$, where $\theta > 0$.

(a) Find the MLE of the mean parameter $\mu = \theta/(\theta + 1)$.

(b) Prove that the cdf of the population r.v. $X \sim f(x; \theta)$ is $F(x; \theta) = x^\theta$ for $x \in (0, 1)$.

(c) Define $T \hat{=} -\sum_{i=1}^n \log X_i$. Prove that T/n is an unbiased estimator of $\tau(\theta) = 1/\theta$.

(d) Find the Fisher information $I_n(\theta)$.

(e) Prove that T/n is an efficient estimator of $\tau(\theta)$; hence T/n is the unique UMVUE for $\tau(\theta)$.

Solution: (a) [**see Example 3.29** on the page 152 of the Textbook] The likelihood function of θ is $L(\theta) = \prod_{i=1}^n \theta x_i^{\theta-1}$, so that the log-likelihood function is

$$\ell(\theta) = n \log \theta + (\theta - 1) \sum_{i=1}^n \log x_i.$$

Let $0 = \ell'(\theta) = n/\theta + \sum_{i=1}^n \log x_i$, the MLE of θ is given by

$$\hat{\theta} = \frac{n}{-\sum_{i=1}^n \log X_i} = \frac{n}{T}.$$

Thus, the MLE of μ is

$$\hat{\mu} = \frac{\hat{\theta}}{\hat{\theta}+1} = \frac{n}{n+T}.$$

(b) The cdf of X is

$$F(x;\theta) = \int_0^x f(t;\theta)\mathrm{d}t = \int_0^x \theta t^{\theta-1}\mathrm{d}t = \int_0^x \mathrm{d}t^\theta = t^\theta \Big|_0^x = x^\theta, \quad x \in (0,1). \tag{6.1}$$

(c) From (4.3) on page 165 of the Textbook, we have

$$-2\sum_{i=1}^n \log F(X_i;\theta) \sim \chi^2(2n)$$

for any continuous cdf; hence

$$-2\sum_{i=1}^n \log F(X_i;\theta) \overset{(6.1)}{=} -2\theta \sum_{i=1}^n \log X_i = 2\theta T \sim \chi^2(2n). \tag{6.2}$$

We have $E(2\theta T) = 2n$, i.e.,

$$E\left(\frac{T}{n}\right) = \frac{1}{\theta} = \tau(\theta), \tag{6.3}$$

indicating that T/n is an unbiased estimator of $\tau(\theta) = 1/\theta$.

On the other hand, from (6.2), $\text{Var}(2\theta T) = 2 \times 2n$, i.e.,

$$\text{Var}\left(\frac{T}{n}\right) = \frac{1}{n\theta^2}. \tag{6.4}$$

In particular, in (6.3) and (6.4) let $n = 1$, we obtain

$$E(-\log X_1) \;=\; \frac{1}{\theta} = \tau(\theta) \quad \text{and} \quad \text{Var}(-\log X_1) = \frac{1}{\theta^2}. \tag{6.5}$$

(d) Let $X \sim f(x;\theta) = \theta x^{\theta-1}$, $x \in (0,1)$. Then, from (3.24) in the Textbook, we have

$$\begin{aligned} I(\theta) &= E\left\{\frac{\mathrm{d}\log f(X;\theta)}{\mathrm{d}\theta}\right\}^2 = E\left(\frac{1}{\theta} + \log X\right)^2 = E\left[-\log X_1 - \tau(\theta)\right]^2 \\ &\overset{(6.5)}{=} \text{Var}(-\log X_1) \overset{(6.5)}{=} \frac{1}{\theta^2} \end{aligned}$$

while the method on page 152 of the Textbook is much easier; and hence

$$I_n(\theta) = nI(\theta) = \frac{n}{\theta^2}.$$

(e) Now, T/n is an unbiased estimator of $\tau(\theta)$, and

$$\text{Var}\left(\frac{T}{n}\right) \overset{(6.4)}{=} \frac{1}{n\theta^2} = \frac{\{\tau'(\theta)\}^2}{I_n(\theta)},$$

i.e., the variance attains the CR lower bound. Then T/n is an efficient estimator of $\tau(\theta)$, and hence T/n is the unique UMVUE for $\tau(\theta)$. ‖

T6.2 Sufficiency

6.2.1 Definition

Let $\mathbf{x} = (X_1, \ldots, X_n)^\top$ and $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ be their realizations. A statistic $T(\mathbf{x})$ is a *sufficient statistic* of θ if the conditional distribution of $X_1, \ldots, X_n$, given $T = t$, does not depend on θ for any value of t. For discrete cases, this means

$$\Pr\{X_1 = x_1.\ldots, X_n = x_n; \theta \mid T(\mathbf{x}) = t\} = h(\boldsymbol{x})$$

does not depend on θ.

6.2.2 Factorization theorem

A statistic $T(\mathbf{x})$ is a sufficient statistic of θ if and only if the joint pdf (or pmf) can be written in the form

$$f(x_1, \dots, x_n; \theta) = f(\boldsymbol{x}; \theta) = g(T(\boldsymbol{x}); \theta) \times h(\boldsymbol{x}),$$

where $h(\boldsymbol{x})$ does not depend on θ, $g(T; \theta)$ is a function of both T and θ, and it depends on $x_1, \dots, x_n$ only through T.

6.2.3 p-parameter exponential family

For the **one-parameter exponential family**, please see Exercise 3.19 on page 161 of the Textbook. Let $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top \in \boldsymbol{\Theta}$ be a parameter vector. If a pmf or pdf $f(x; \boldsymbol{\theta})$ can be expressed as

$$f(x; \boldsymbol{\theta}) = a(\boldsymbol{\theta}) b(x) \exp\left[\sum_{j=1}^{p} c_j(\boldsymbol{\theta}) d_j(x)\right], \quad -\infty < x < \infty,$$

for some specific functions $a(\cdot, \dots, \cdot)$, $b(\cdot)$, $c_j(\cdot, \dots, \cdot)$ and $d_j(\cdot)$, then $f(x; \boldsymbol{\theta})$ is said to belong to the *p-parameter exponential family*.

(a) Show that the pdf of $N(\mu, \sigma^2)$ belongs to the two-parameter exponential family.

(b) Show that the pdf of Beta(θ_1, θ_2) belongs to the two-parameter exponential family.

(c) Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} f(x; \boldsymbol{\theta})$, where $f(x; \boldsymbol{\theta})$ belongs to the p-parameter exponential family. Show that $\sum_{i=1}^n d_1(X_i), \dots, \sum_{i=1}^n d_p(X_i)$ is a set of jointly sufficient statistics of $\boldsymbol{\theta}$.

Example T6.2 (Revisited **28.1• Remarks on Example 3.28**). One student asked me that since $\mathcal{U} = \{\hat{\theta}\colon E(\hat{\theta}) = \theta\}$, why $\#\mathcal{U} = 1$? For example, let $W_i \mathrel{\hat{=}} E(X_i | T)$ for $i = 1, \dots, n$, where $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$ and $T = \sum_{i=1}^n X_i$ is sufficient for θ, we have

$$E(W_i) = E(X_i) = \theta \quad \text{and} \quad \text{Var}(W_i) \leqslant \text{Var}(X_i) = \theta(1 - \theta).$$

Clearly, W_i is a function of T and W_i is an unbiased estimator of θ, so all $W_i \in \mathcal{U}$. Similarly, we have $E[(X_i + X_j)/2|T] \in \mathcal{U}$, and so on. How to explain these phenomena?

Solution: From Example 3.22 on pages 140–141, we have

$$\Pr\{X_1 = x_1, \ldots, X_n = x_n | T = t\} = 1 \Big/ \binom{n}{t}, \quad t = \sum_{i=1}^{n} x_i, \quad x_i \in \{0, 1\},$$

so that

$$1 = \sum_{x_1, \ldots, x_n \in \{0,1\}} \binom{n}{x_1 + \cdots + x_n}^{-1}. \tag{6.6}$$

We can show that

$$X_i | (T = t) \sim \text{Bernoulli}(t/n), \quad i = 1, \ldots, n, \tag{6.7}$$

or $X_i|T \sim \text{Bernoulli}(T/n) = \text{Bernoulli}(\bar{X})$. Thus,

$$W_i = E(X_i|T) = \bar{X} \quad \text{and} \quad E[(X_i + X_j)/2|T] = \bar{X},$$

so $\mathcal{U} = \{\bar{X}\}$ and $\#\mathcal{U} = 1$.

Proof of (6.7): We only prove that $X_n|(T = t) \sim \text{Bernoulli}(t/n)$. Since X_n only takes value 0 or 1, we only need to prove

$$\Pr(X_n = 1 | T = t) = t/n = \bar{x}. \tag{6.8}$$

Now, we obtain

$$\begin{aligned}
& \Pr\{X_n = x_n = 1 | T = t\} \\
= & \sum_{x_1, \ldots, x_{n-1} \in \{0,1\}} \Pr\{X_1 = x_1, \ldots, X_n = x_n | T = t\} \\
= & \sum_{x_1, \ldots, x_{n-1} \in \{0,1\}} \binom{n}{x_1 + \cdots + x_n}^{-1} \\
= & \sum_{x_1, \ldots, x_{n-1} \in \{0,1\}} \binom{n}{x_1 + \cdots + x_{n-1} + 1}^{-1}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{x_1,\dots,x_{n-1}\in\{0,1\}} \left[\frac{n(n-1)!}{(x_1+\cdots+x_{n-1}+1)!(n-1-x_1-\cdots-x_{n-1})}\right]^{-1} \\
&= \frac{x_1+\cdots+x_{n-1}+1}{n} \sum_{x_1,\dots,x_{n-1}\in\{0,1\}} \binom{n-1}{x_1+\cdots+x_{n-1}}^{-1} \\
&\overset{(6.6)}{=} \frac{x_1+\cdots+x_{n-1}+x_n}{n}\cdot 1 \\
&= \bar{x},
\end{aligned}
$$

implying (6.8). ‖

Example T6.3 (Revisited Example 3.21). One student asked me that since $\mathcal{U} = \{\hat{\theta}\colon\ E(\hat{\theta}) = \theta\}$, why $\#\mathcal{U} = 1$? For example, define $W_i \hat{=} E(X_i|T)$ for $i = 1, \dots, n$, where $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim}$ Poisson(θ) and $T = \sum_{i=1}^n X_i$ is sufficient for θ, we have

$$
E(W_i) = E(X_i) = \theta \quad \text{and} \quad \mathrm{Var}(W_i) \leqslant \mathrm{Var}(X_i) = \theta.
$$

Clearly, W_i is a function of T and W_i is an unbiased estimator of θ, so all $W_i \in \mathcal{U}$. Similarly, we have $E[(X_i + X_j)/2|T] \in \mathcal{U}$, and so on. How to explain these phenomena?

Solution: From Example 3.21 on page 140, we have

$$
\Pr\{X_1 = x_1, \dots, X_n = x_n|T = t\} = \binom{t}{x_1, \dots, x_n}\cdot\frac{1}{n^t},
$$

i.e.,

$$
(X_1, \dots, X_n)|(T = t) \sim \text{Multinomial}(t; 1/n, \dots, 1/n).
$$

Thus, $X_i|(T = t) \sim \text{Binomial}(t, 1/n)$ and $W_i = E(X_i|T) = T/n = \bar{X}$. so $\mathcal{U} = \{\bar{X}\}$ and $\#\mathcal{U} = 1$. ‖

T6.3 Completeness

6.3.1 Definition

A statistic $T(\mathbf{x})$ is said to be *complete* if

$$E[h(T)] = 0, \forall\, \theta \in \Theta \quad \Rightarrow \quad \Pr\{h(T) = 0\} = 1, \forall\, \theta \in \Theta,$$

where the function $h(T)$ is a statistic.

6.3.2 Lehmann–Scheffé theorem

Let $T(\mathbf{x})$ be a complete sufficient statistic of θ. If $g(T)$ is an unbiased estimator of $\tau(\theta)$, then $g(T)$ is the unique UMVUE for $\tau(\theta)$.

Example T6.4 (An exponential distribution). Let $X_1, \ldots, X_n$ be a random sample from the exponential distribution with density

$$f(x;\theta) = \theta \mathrm{e}^{-\theta x}, \quad x > 0,\ \theta > 0.$$

(a) Show that $T = T(\mathbf{x}) = \sum_{i=1}^n X_i$ is a sufficient statistic of θ.

(b) Show that T is complete for θ.

(c) Prove that $\bar{X}$ is the unique UMVUE for $\tau_1(\theta) = 1/\theta$.

(d) Find the unique UMVUE for θ.

(e) Find the unique UMVUE for $\tau_2(\theta) = \mathrm{e}^{-K\theta} = \Pr(X_1 > K)$, where $K > 0$ is a given constant.

Solution: (a) The joint pdf of $X_1, \ldots, X_n$ is

$$f(\boldsymbol{x};\theta) = \prod_{i=1}^n \theta \mathrm{e}^{-\theta x_i} = \theta^n \mathrm{e}^{-\theta T} \times 1,$$

therefore, T is sufficient for θ, and $T \sim \text{Gamma}(n, \theta)$.

(b) Assume that a function $h(T)$ satisfies

$$E[h(T)] = \int_0^\infty h(t) \frac{\theta^n}{\Gamma(n)} t^{n-1} \mathrm{e}^{-\theta t} \mathrm{d}t = 0$$

for $\theta > 0$ iff

$$\int_0^\infty h(t) t^{n-1} \mathrm{e}^{-\theta t} \mathrm{d}t = 0 \quad \text{for all } \theta > 0.$$

Since $t^{n-1}\mathrm{e}^{-\theta t} > 0$ for all $t > 0$ and $\theta > 0$, we have $h(t) = 0$ for all $t > 0$ and hence T is complete for θ.

(c) $E(T) = n/\theta$ so that $E(T/n) = E(\bar{X}) = 1/\theta = \tau_1(\theta)$. Thus, according to the Lehmann–Scheffé theorem, $\bar{X}$ is the unique UMVUE for $\tau_1(\theta) = 1/\theta$.

(d) To find the UMVUE of θ, one might suspect that the estimator is of the form c/T, where c is a constant which may depend on n. Now we calculate

$$\begin{aligned} E\left(\frac{c}{T}\right) &= \int_0^\infty \frac{c}{t} \cdot \frac{\theta^n}{\Gamma(n)} t^{n-1} \mathrm{e}^{-\theta t} \mathrm{d}t = \frac{c\theta^n}{\Gamma(n)} \int_0^\infty t^{n-2} \mathrm{e}^{-\theta t} \mathrm{d}t \\ &= \frac{c\theta^n}{\Gamma(n)} \cdot \frac{\Gamma(n-1)}{\theta^{n-1}} = \frac{c\theta}{n-1} = \theta, \end{aligned}$$

if let $c = n - 1$. Thus, $(n-1)/T$ is an unbiased estimator of θ, so $(n-1)/T$ the unique UMVUE of θ for $n > 1$.

(e) Similar to Example 3.20 in the Textbook, we can construct a Bernoulli random variable $\xi \mathrel{\hat{=}} I(X_1 > K)$, so we have

$$E(\xi) = \Pr(X_1 > K) = \mathrm{e}^{-K\theta} = \tau_2(\theta),$$

i.e., ξ is an unbiased estimator of $\tau_2(\theta)$. Based on (3.30) on page 146 of the Textbook, we know that

$$g(T) \mathrel{\hat{=}} E[\xi | T]$$

is also an unbiased estimator of $\tau_2(\theta)$. Thus, according to the Lehmann–Scheffé theorem, $g(T)$ is the unique UMVUE for $\tau_2(\theta)$.

Finding $g(T)$: In the follows, we want to find

$$
\begin{aligned}
g(t) &= E[\xi|T=t] = \Pr(\xi = 1|T=t) \\
&= \Pr(X_1 > K|T=t) = \int_K^\infty f_{X_1|T=t}(x_1|t)\mathrm{d}x_1,
\end{aligned} \tag{6.9}
$$

where $f_{X_1|T=t}(x_1|t)$ denotes the conditional density of X_1 given $T=t$.

Note that the definition of a density is

$$
f(x) = F'(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Pr(x < X \leqslant x + \Delta x)}{\Delta x},
$$

approximately, we have $f(x)\Delta x \approx \Pr(x < X \leqslant x + \Delta x)$.

Let $\mathrm{Gamma}(\cdot|n,\theta)$ denote the pdf of $\mathrm{Gamma}(n,\theta)$ distribution, then,

$$
\begin{aligned}
f_{X_1|T=t}(x_1|t)\Delta x_1 &= \frac{f_{X_1,T=t}(x_1,t)\Delta x_1 \Delta t}{f_T(t)\Delta t} \\
&\approx \frac{\Pr(x_1 < X_1 \leqslant x_1 + \Delta x_1,\ t < T \leqslant t + \Delta t)}{\mathrm{Gamma}(t|n,\theta)\Delta t} \\
&= \frac{\Pr(x_1 < X_1 \leqslant x_1 + \Delta x_1) \cdot \Pr(t - x_1 < \sum_{i=2}^n X_i \leqslant t - x_1 + \Delta t)}{\mathrm{Gamma}(t|n,\theta)\Delta t} \\
&= \frac{\theta \mathrm{e}^{-\theta x_1}\Delta x_1 \cdot \mathrm{Gamma}(t - x_1|n-1,\theta)\Delta t}{[1/\Gamma(n)]\theta^n t^{n-1}\mathrm{e}^{-\theta t}\Delta t} \\
&= \frac{\theta \mathrm{e}^{-\theta x_1} \cdot [1/\Gamma(n-1)]\theta^{n-1}(t-x_1)^{n-2}\mathrm{e}^{-\theta(t-x_1)}\Delta x_1 \Delta t}{[1/\Gamma(n)]\theta^n t^{n-1}\mathrm{e}^{-\theta t}\Delta t} \\
&= (n-1)\cdot \frac{(t-x_1)^{n-2}}{t^{n-1}}\Delta x_1,
\end{aligned} \tag{6.10}
$$

for $x_1 < t$ and $n > 1$. Thus, (6.9) becomes

$$
\begin{aligned}
g(t) &\overset{(6.9)}{=} \int_K^\infty f_{X_1|T=t}(x_1|t)\mathrm{d}x_1 \overset{(6.10)}{=} \int_K^t (n-1)\cdot \frac{(t-x_1)^{n-2}}{t^{n-1}}\mathrm{d}x_1 \\
&= \frac{n-1}{t^{n-1}} \int_K^t (t-x_1)^{n-2}\mathrm{d}x_1
\end{aligned}
$$

$$= \frac{1}{t^{n-1}} \int_0^{t-K} (n-1) y^{n-2} \mathrm{d}y \qquad [\text{let } y = t - x_1]$$

$$= \frac{1}{t^{n-1}} \cdot y^{n-1} \Big|_0^{t-K} = (1 - K/t)^{n-1} \cdot I(t > K),$$

so that

$$g(T) = (1 - K/T)^{n-1} \cdot I(T > K). \qquad \|$$

Example T6.5 (A geometric distribution). Let $X_1, \ldots, X_n$ be a random sample from the geometric distribution with density

$$f(x;\theta) = \theta(1-\theta)^x, \quad x = 0, 1, \ldots, \infty;\ 0 < \theta < 1.$$

(a) Show that $T = T(\mathbf{x}) = \sum_{i=1}^n X_i$ is a sufficient statistic of θ.

(b) Show that $T \sim \text{NBinomial}(n, \theta)$, whose density is

$$g(t;\theta) = \binom{n+t-1}{t} \theta^n (1-\theta)^t, \quad t = 0, 1, \ldots, \infty.$$

Hint: For a positive integer n,

$$(x+a)^{-n} = \sum_{y=0}^{\infty} (-1)^y \binom{n+y-1}{y} x^y a^{-n-y}, \quad \text{for } |x| < a.$$

(c) Show that T is complete for θ.

(d) Find the UMVUE for $\tau(\theta) = (1-\theta)/\theta$ by the Lehmann–Scheffé Theorem. Show that it is also an efficient estimator for $\tau(\theta)$.

(e) Show that $I_{\{0\}}(X_j)$ is an unbiased estimator for θ.

(f) Find the UMVUE for θ.

Solution: (a) The joint pmf of $X_1, \ldots, X_n$ is

$$f(\boldsymbol{x};\theta) = \prod_{i=1}^n \theta(1-\theta)^{x_i} = \theta^n (1-\theta)^T \times 1,$$

indicating that T is sufficient for θ.

(b) The mgf of X_i $(i = 1, \ldots, n)$ is

$$\begin{aligned} M_{X_i}(t) &= E(\mathrm{e}^{tX_i}) = \sum_{x=0}^{\infty} \mathrm{e}^{tx}\theta(1-\theta)^x \\ &= \theta \sum_{x=0}^{\infty} [\mathrm{e}^t(1-\theta)]^x = \frac{\theta}{1-\mathrm{e}^t(1-\theta)}, \quad t < -\log(1-\theta). \end{aligned}$$

So the mgf of T is

$$M_T(t) = \prod_{i=1}^{n} M_{X_i}(t) = \left[\frac{\theta}{1-e^t(1-\theta)}\right]^n, \quad t < -\log(1-\theta).$$

On the other hand, let $Y \sim \text{NBinomial}(n, \theta)$. The mgf of Y is

$$\begin{aligned} M_Y(t) &= E(\mathrm{e}^{tY}) = \sum_{y=0}^{\infty} \mathrm{e}^{ty} \binom{n+y-1}{y} \theta^n (1-\theta)^y \\ &= \theta^n \times \sum_{y=0}^{\infty} \binom{n+y-1}{y} [\mathrm{e}^t(1-\theta)]^y = \theta^n \times [1-\mathrm{e}^t(1-\theta)]^{-n} \\ &= \left[\frac{\theta}{1-\mathrm{e}^t(1-\theta)}\right]^n, \qquad t < -\log(1-\theta). \end{aligned}$$

Since $M_T(t) = M_Y(t)$, $T(\mathbf{x}) \sim \text{NBinomial}(n, \theta)$.

(c) Assume that $E[h(T)] = 0$, we have

$$\sum_{t=0}^{\infty} h(t) \binom{n+t-1}{t} \theta^n (1-\theta)^t = 0,$$

or

$$\sum_{t=0}^{\infty} \binom{n+t-1}{t} h(t)(1-\theta)^t = 0, \quad 0 < \theta < 1. \tag{6.11}$$

The equation (6.11) is a polynomial of $(1-\theta)$ and $(1-\theta)$ must be nonzero. The fact that it equals to zero implies that all its coefficients are zero, i.e.,

$$\binom{n+t-1}{t} h(t) = 0, \quad t = 0, 1, \ldots, \infty.$$

Hence, $h(T) = 0$, i.e. $\Pr\{h(T) = 0\} = 1$. Therefore, $T(\mathbf{x})$ is complete for θ.

(d) We have

$$E(T) = \left.\frac{\mathrm{d}M_T(t)}{\mathrm{d}t}\right|_{t=0} = \left.\frac{n\mathrm{e}^t\theta^n(1-\theta)}{[1-\mathrm{e}^t(1-\theta)]^{n+1}}\right|_{t=0} = n \cdot \frac{1-\theta}{\theta} = n \cdot \tau(\theta).$$

Denote $\bar{X} = T(\mathbf{x})/n$. It implies that $E(\bar{X}) = \tau(\theta)$, and thus $\bar{X}$ is an unbiased estimator for $\tau(\theta)$. Because $T(\mathbf{x})$ is sufficient and complete, according to the Lehmann–Scheffé Theorem, $\bar{X}$ is the unique UMVUE for $\tau(\theta) = (1-\theta)/\theta$.

$\bar{X}$ is an efficient estimator for $\tau(\theta)$: To prove that $\bar{X}$ is an efficient estimator for $\tau(\theta)$, we need to show that $\mathrm{Var}(\bar{X})$ equals to the Cramér–Rao lower bound. Since

$$E(T^2) = \left.\frac{\mathrm{d}^2M_T(t)}{\mathrm{d}t^2}\right|_{t=0} = \left.\frac{n\mathrm{e}^t\theta^n(1-\theta)[1-n\mathrm{e}^t(1-\theta)]}{[1-\mathrm{e}^t(1-\theta)]^{n+2}}\right|_{t=0} = \frac{n(1-\theta)[1-n(1-\theta)]}{\theta^2},$$

we have

$$\mathrm{Var}(T) = E(T^2) - [E(T)]^2 = \frac{n(1-\theta)}{\theta^2}.$$

Hence,

$$\mathrm{Var}(\bar{X}) = \frac{\mathrm{Var}(T)}{n^2} = \frac{1-\theta}{n\theta^2}.$$

The log-likelihood function is

$$\ell(\theta; \mathbf{x}) = \log f(\mathbf{x}; \theta) = n \log \theta + T \log(1-\theta).$$

Thus, the Fisher information is

$$I_n(\theta) = E\left[-\frac{\mathrm{d}^2\ell(\theta; \mathbf{x})}{\mathrm{d}\theta^2}\right] = E\left[\frac{n}{\theta^2} + \frac{T}{(1-\theta)^2}\right] = \frac{n}{\theta^2(1-\theta)}.$$

Since $\tau'(\theta) = -1/\theta^2$, the Cramér–Rao lower bound is

$$\upsilon(\theta) = \frac{[\tau'(\theta)]^2}{I_n(\theta)} = \frac{1-\theta}{n\theta^2}.$$

Since $\mathrm{Var}(\bar{X}) = \upsilon(\theta)$, $\bar{X}$ is an efficient estimator for $\tau(\theta) = (1-\theta)/\theta$.

(e) Let

$$I_{\{0\}}(X_j) = \begin{cases} 1, & \text{if } X_j = 0, \\ 0, & \text{otherwise.} \end{cases}$$

From the pmf of X_j, we obtain that

$$E[I_{\{0\}}(X_j)] = 1 \times \Pr(X_j = 0) = f(0;\theta) = \theta.$$

Therefore, $I_{\{0\}}(X_j)$ is an unbiased estimator for θ.

(f) Let

$$T = \sum_{i=1}^{n} X_i \quad \text{and} \quad T_{-j} = \sum_{i=1, i\neq j}^{n} X_i = T - X_j.$$

From (b), we know that $T \sim \text{NBinomial}(n, \theta)$ and $T_{-j} \sim \text{NBinomial}(n-1, \theta)$. So

$$\Pr(X_j = 0 \mid T = t) = \frac{\Pr(X_j = 0, T = t)}{\Pr(T = t)}$$

$$= \frac{\Pr(X_j = 0,\ T - X_j = T_{-j} = t)}{\Pr(T = t)}$$

$$= \frac{\Pr(X_j = 0) \cdot \Pr(T_{-j} = t)}{\Pr(T = t)} \qquad \text{[since } X_j \text{ and } T_{-j} \text{ are independent]}$$

$$= \frac{\theta \cdot \binom{n+t-2}{t} \theta^{n-1} (1-\theta)^t}{\binom{n+t-1}{t} \theta^n (1-\theta)^t} = \frac{n-1}{t+n-1}.$$

Let

$$g(T) = E[I_{\{0\}}(X_j) \mid T] = \Pr\{X_j = 0 \mid T\} = \frac{n-1}{T+n-1}.$$

Since $E[g(T)] = E\{E[I_{\{0\}}(X_j) \mid T]\} = E[I_{\{0\}}(X_j)] = \theta$, $g(T)$ is an unbiased estimator for θ. Because T is sufficient and complete, according to the Lehmann–Scheffé Theorem,

$$g(T) = \frac{n-1}{\sum_{i=1}^{n} X_i + n - 1}$$

is the unique UMVUE for θ. ||