

HW1 (Due 4/08/20)

1. Prove $Var[X] = Var(E[X|Y] + E(Var[X|Y]))$ (Conditional variance identity)

$X = S(x)$ is an unbiased estimator of $q(\theta)$. $Y = T(x)$ is a sufficient statistic for θ .

Let $E[X] = \mu_X$ is a constant. Using Law of iterated Expectation, $E(E[X|Y]) = E[X] = \mu_X$.

$$\begin{aligned} Var[X] &= E[(X - \mu_X)^2] = E\{(X - E[X|Y] + E[X|Y] - \mu_X)^2\} \\ &= E\{(X - E[X|Y])^2\} + E\{(E[X|Y] - \mu_X)^2\} + 2E\{(X - E[X|Y])(E[X|Y] - \mu_X)\} \\ &= E\{E[(X - E[X|Y])^2|Y]\} + Var(E[X|Y]) + 0 \\ &= E(Var[X|Y]) + Var(E[X|Y]) \end{aligned}$$

Proof $E\{(X - E[X|Y])(E[X|Y] - \mu_X)\} = 0$

For $E[X|Y]$ is a constant,

$$\begin{aligned} E[(X - E[X|Y])(E[X|Y] - \mu_X)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (X - E[X|Y])(E[X|Y] - \mu_X) f(X, Y) dX dY \\ &= \int_{-\infty}^{\infty} (E[X|Y] - \mu_X) \left\{ \int_{-\infty}^{\infty} (X - E[X|Y]) f(X, Y) dX \right\} dY \\ &= \int_{-\infty}^{\infty} (E[X|Y] - \mu_X) \cdot 0 dY = 0 \end{aligned}$$

Proof $\int_{-\infty}^{\infty} (X - E[X|Y]) f(X, Y) dX = 0$

$$\begin{aligned} \int_{-\infty}^{\infty} (X - E[X|Y]) f(X, Y) dX &= \int_{-\infty}^{\infty} X f(X, Y) dX - E[X|Y] \int_{-\infty}^{\infty} f(X, Y) dX \\ &= \int_{-\infty}^{\infty} X f(X|Y) g(Y) dX - E[X|Y] g(Y) \\ &= g(Y) E[X|Y] - E[X|Y] g(Y) = 0 \end{aligned}$$

Thus $Var[X] = E(Var[X|Y]) + Var(E[X|Y])$

2. $Cov(X, Y) \leq \sqrt{Var(X)Var(Y)}$.

BY Cauchy-Schwarz inequality $|EXY| \leq E|XY| \leq \sqrt{E|X|^2 E|Y|^2}$, then

$$|E(X - \mu_X)(Y - \mu_Y)| \leq \sqrt{E(X - \mu_X)^2 E(Y - \mu_Y)^2} \implies Cov(X, Y) \leq \sqrt{Var(X)Var(Y)}$$

Or for $\sqrt{Var(X)Var(Y)} \geq 0$,

$$|Corr(X, Y)| = \left| \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \right| \leq 1 \implies Cov(X, Y) \leq \sqrt{Var(X)Var(Y)}$$

3. Problem 3.4.22 (Bickel and Doksum). *Regularity Conditions are Needed for the Information Inequality.* Let $X \sim U(0, \theta)$ be the uniform distribution on $(0, \theta)$. Note that $\log p(x, \theta)$ is differentiable for all $\theta > x$, that is, with probability 1 for each θ , and we can thus define moments of $T = \partial/\partial\theta \log p(X, \theta)$. Show that, however,

$$(i) \ E\left(\frac{\partial}{\partial\theta} \log p(X, \theta)\right) = -\frac{1}{\theta} \neq 0$$

For $p(x, \theta) = \frac{1}{\theta}, 0 < x < \theta$,

$$E\left(\frac{\partial}{\partial\theta} \log p(X, \theta)\right) = E\left(\frac{\partial}{\partial\theta} \log \frac{1}{\theta}\right) = E\left(-\frac{\partial}{\partial\theta} \log \theta\right) = E\left(-\frac{1}{\theta}\right) = -\frac{1}{\theta} \neq 0$$

$$(ii) \ Var\left(\frac{\partial}{\partial\theta} \log p(X, \theta)\right) = 0 \text{ and } I(\theta) = ET^2 = VarT + (ET)^2 = 1/\theta^2.$$

$$I(\theta) = E\left(\left[\frac{\partial}{\partial\theta} \log p(X, \theta)\right]^2\right) = E\left(\left[-\frac{1}{\theta}\right]^2\right) = \frac{1}{\theta^2}$$

$$Var\left(\frac{\partial}{\partial\theta} \log p(X, \theta)\right) = E\left(\left[\frac{\partial}{\partial\theta} \log p(X, \theta)\right]^2\right) - E\left(\frac{\partial}{\partial\theta} \log p(X, \theta)\right)^2 = \frac{1}{\theta^2} - \frac{1}{\theta^2} = 0$$

$$(iii) \ 2X \text{ is unbiased for } \theta \text{ and has variance } (1/3)\theta^2 < (1/I(\theta)) = \theta^2.$$

Let $Y = 2X, \frac{dx}{dy} = \frac{1}{2}, p_Y(y, \theta) = p(x, \theta) \left|\frac{dx}{dy}\right| = \frac{1}{2\theta}, 0 < y < 2\theta$

$$E_\theta Y = \int_0^{2\theta} \frac{y}{2\theta} dy = \frac{1}{2\theta} \cdot \frac{1}{2} y^2 \Big|_0^{2\theta} = \theta$$

$E[2X] - \theta = 0$, hence $2X$ is unbiased for θ .

$$Var[Y] = \frac{(2\theta)^2}{12} = \frac{1}{3}\theta^2 < \frac{1}{I(\theta)} = \theta^2$$

This indicates that the Cramer-Rao Theorem is not applicable to this pdf.

Theorem 3.4.1. (Information Inequality) and Corollary 3.4.1. need two regularity assumptions on the family $\{P_\theta : \theta \in \Theta\}$:

- (I) The set $A = \{x : p(x, \theta) > 0\}$ does not depend on $\theta, \forall x \in A, \theta \in \Theta, \partial/\partial\theta \log p(X, \theta)$ exists and is finite.

support depend on θ

- (II) If T is any statistic such that $E\theta(|T|) < \infty, \theta \in \Theta$, then the operations of integration and differentiation by θ can be interchanged in $\int T(x)p(x, \theta)dx$.

(the mean of score has to be always 0)

When I and II fail to hold, or even I and II are satisfied and UMVUE of $\psi(\theta)$ exist, the variance of the best estimate is not equal to the bound in many situations.

4. Let X be a $N(\theta, 1)$ variable and consider the estimate $T_{a,b}(X) = aX + b$ of θ .

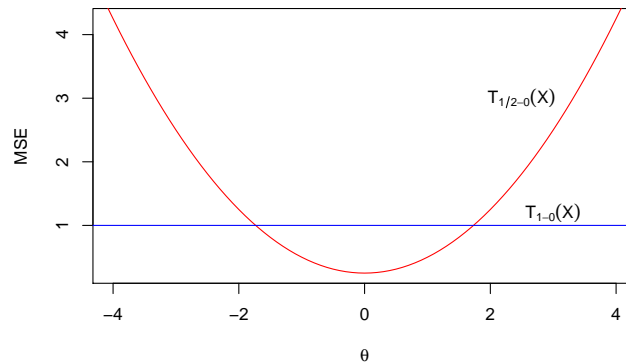
a. Calculate the MSE of $T_{a,b}(X)$. Use the form in Proposition 1.3.1. (Bickel and Doksum).

$$\begin{aligned} \text{MSE}(T_{a,b}(X)) &= \text{Bias}^2(T_{a,b}(X)) + \text{Var}(T_{a,b}(X)) = (E(aX + b) - \theta)^2 + \text{Var}(aX + b) \\ &= (aE X + b - \theta)^2 + a^2 \text{Var}(X) = (a\theta + b - \theta)^2 + a^2 \end{aligned}$$

b. Plot MSE of $T_{1/2,0}(X)$ and of the “natural” estimate $T_{1,0}(X) = X$ as a function of θ and show that neither estimate improves the other for all θ .

$$\text{MSE}(T_{1/2,0}(X)) = (\frac{1}{2}\theta - \theta)^2 + (\frac{1}{2})^2 = \frac{1}{4}\theta^2 + \frac{1}{4}$$

$$\text{MSE}(T_{1,0}(X)) = (\theta - \theta)^2 + 1 = 1$$



$\text{MSE}(T_{1,0}(X)) = 1, \forall \theta \in \mathbb{R}$ is a horizontal line, which doesn't depend on θ . $\text{MSE}(T_{1/2,0}(X))$ is a parabola opening upward and depend on θ .

$\text{MSE}(T_{1/2,0}(X))$ crosses $\text{MSE}(T_{1,0}(X))$ and is smaller than $\text{MSE}(T_{1,0}(X))$ only when $\theta \in (-\sqrt{3}, \sqrt{3})$. Hence they do not improve each other.

c. Is there any estimate of the form $aX + b$ which improves on X for all θ ?

Only when $a = 1$, $\text{MSE}(T_{1,b}(X)) = b^2 + 1$ is free of θ and it reaches the lower bound when $b = 0$. There is not any estimate of the form $aX + b$ which improves on X for all θ .

d. Show that X is the only unbiased estimate of the form $aX + b$.

$$\text{Bias } T_{a,b}(X) = a\theta + b - \theta = 0 \iff \theta = \frac{b}{1-a}$$

Only when $T_{1,0}(X) = X$, $\text{Bias}(T_{a,b}(X)) = 0, \forall \theta \in \mathbb{R}$, the form $aX + b$ is an unbiased estimator of θ .

By Rao-Blackwell Theorem, $T_{1,0}(X)$ is a sufficient statistic for θ , X is a function of T . Hence it is the only unbiased estimate of the form $aX + b$.

not proof unique

5. Let X_1, \dots, X_n be a sample from a population with mean μ and variance σ^2 both of which are unknown. Let $T(\underline{X}) = \sum_{i=1}^n c_i X_i$.

a. Show that T is unbiased for μ , if and only if $\sum_{i=1}^n c_i = 1$.

$$E[T(\underline{X})] = E\left[\sum_{i=1}^n c_i X_i\right] = \sum_{i=1}^n c_i E[X_i] = \mu \sum_{i=1}^n c_i = \mu \iff \sum_{i=1}^n c_i = 1, \forall \mu \in R$$

Hence T is unbiased for μ , if and only if $\sum_{i=1}^n c_i = 1$.

b. Show that \bar{X} has uniformly smallest variance among all unbiased estimates of this form.

$T(\underline{X}) = \sum_{i=1}^n c_i X_i$ is a sufficient statistic for μ when $\sum_{i=1}^n c_i = 1$.

Let $S(\underline{X}) = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, $E[S(\underline{X})] = E[\frac{1}{n} T(\underline{X})] = \mu$. $S(\underline{X})$ is an unbiased estimator of μ .

By Rao-Blackwell Theorem, $Var[\tilde{S}(\underline{X})] \leq Var[S(\underline{X})]$, $\forall \mu \in R$, $\tilde{S}(\underline{X}) = E[S(\underline{X})|T(\underline{X})]$ is a better unbiased statistic for μ .

Since $S(\underline{X}) = \frac{1}{n} T(\underline{X})$, $Var[S(\underline{X})|T(\underline{X})] = 0$ then $\tilde{S}(\underline{X}) = S(\underline{X}) = \bar{X}$ is the UMVUE of μ when $\sum_{i=1}^n c_i = 1$.

HW2. (Due 4/15/20)

1. Let X_1, \dots, X_n be a sample from $U(\theta_1, \theta_2)$ where θ_1 and θ_2 are unknown.

a. Show that $T(\underline{X}) = (\min(X_1, \dots, X_n), \max(X_1, \dots, X_n))$ is sufficient.

Let $X_{(1)} = \min X_i, X_{(n)} = \max X_i, f_X(x|\theta_1, \theta_2) = \frac{1}{\theta_2 - \theta_1},$

$$f(\mathbf{x}|\theta_1, \theta_2) = \prod_{i=1}^n \frac{1}{\theta_2 - \theta_1} \mathbf{1}_{[\theta_1, \theta_2]}(x_i) = \frac{1}{(\theta_2 - \theta_1)^n} \cdot \mathbf{1}_{x_{(1)} \in [\theta_1, \infty)} \cdot \mathbf{1}_{x_{(n)} \in (-\infty, \theta_2]}$$

$\mathbf{1}_{x_{(1)} \in [\theta_1, \infty)}$ is free of θ_2 , $\mathbf{1}_{x_{(n)} \in (-\infty, \theta_2]}$ is free of θ_1 . By Factorization Theorem, $T(\underline{X})$ are sufficient statistic

b. Assuming that $T(\underline{X})$ is complete, find a UMVUE of $(\theta_1 + \theta_2)/2$.

$$F_X(x|\theta_1, \theta_2) = \frac{x - \theta_1}{\theta_2 - \theta_1}$$

Let $Y = \frac{X - \theta_1}{\theta_2 - \theta_1} \sim U(0, 1), f_Y(y) = 1, F_Y(y) = y$

$$f_{Y_{(1)}}(y) = \frac{n!}{(1-1)!(n-1)!} f_Y(y) [F_Y(y)]^{1-1} [1 - F_Y(y)]^{n-1} = n(1-y)^{n-1}, 0 \leq y \leq 1$$

$$f_{Y_{(n)}}(y) = \frac{n!}{(n-1)!(n-n)!} f_Y(y) [F_Y(y)]^{n-1} [1 - F_Y(y)]^{n-n} = ny^{n-1}, 0 \leq y \leq 1$$

$Y_{(1)} \sim \text{Beta}(1, n), Y_{(n)} \sim \text{Beta}(n, 1).$

$$E[Y_{(1)}] = E\left[\frac{X_{(1)} - \theta_1}{\theta_2 - \theta_1}\right] = \frac{E[X_{(1)}] - \theta_1}{\theta_2 - \theta_1} = \frac{1}{n+1}$$

$$E[Y_{(n)}] = E\left[\frac{X_{(n)} - \theta_1}{\theta_2 - \theta_1}\right] = \frac{E[X_{(n)}] - \theta_1}{\theta_2 - \theta_1} = \frac{n}{n+1}$$

$$E\left[\frac{X_{(1)} + X_{(n)}}{2}\right] = \frac{1}{2} \left(\frac{\theta_2 - \theta_1}{n+1} + \theta_1 + \frac{n(\theta_2 - \theta_1)}{n+1} + \theta_1 \right) = \frac{\theta_1 + \theta_2}{2}$$

By Lehmann-Scheffe theorem $\frac{X_{(1)} + X_{(n)}}{2}$ is a UMVUE of $(\theta_1 + \theta_2)/2$.

2. Let X_1, \dots, X_n be a sample from a $\Gamma(p, \lambda)$ population where both p and λ are unknown. Find the UMVUE of p/λ .

For Gamma Distribution, $E(\bar{X}) = E(X) = p/\lambda$. \bar{X} is an unbiased estimator of $g(p, \lambda) = p/\lambda$.

$$f_{p,\lambda}(\mathbf{x}) = \exp\left[\underbrace{-\lambda \sum_{i=1}^n x_i + (p-1) \sum_{i=1}^n \log x_i}_{\eta(p,\lambda)T(\mathbf{x})} - \underbrace{n \log\left(\frac{\lambda^p}{\Gamma(p)}\right)}_{B(p,\lambda)}\right] \underbrace{\mathbf{1}_{\{x \in (0, \infty)\}}}_{h(\mathbf{x})}$$

Gamma is a exponential family. $T = (\sum_{i=1}^n X_i, \sum_{i=1}^n \log X_i)$ is a complete sufficient statistic for (λ, p) where $x \in \mathcal{X} \subset \mathbb{R}_+^2$.

By Lehman-Scheffe Theorem, \bar{X} is the UMVUE of p/λ .

3. Let $\underline{X} = X_1, \dots, X_n$ be a sample from $N(\mu, \sigma^2)$. Show that if $n \geq 2$, \underline{X} though sufficient is not complete. Hint: Consider $g(\underline{X}) = X_2 - X_1$.

$$f_{\mu}(x_{1:n}) = \exp\left[\underbrace{\frac{\mu}{\sigma^2}}_{\eta(\mu)} \underbrace{\sum_{i=1}^n x_i}_{T(x)} - \underbrace{n\left(\frac{\mu^2}{2\sigma^2} + \ln(\sqrt{2\pi}\sigma)\right)}_{B(\mu)} \underbrace{\exp\left[-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2}\right]}_{h(x)}\right] \mathbf{1}_{\{x \in \mathbb{R}\}}$$

$T(X_{1:n}) = \sum_{i=1}^n X_i$ is sufficient for μ .

$g(\underline{X}) = X_2 - X_1$ is a function of T where $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ given by $(x_1, x_2) \mapsto x_1 - x_2$

If $E_{\mu}g(T) = 0$ for all μ , $g(X_1, X_2) \sim N(0, 2\sigma^2)$.

$P_{\mu}(g(T) = 0) = 0 \neq 1 \forall \mu$ since the random variables are continuous.

Thus the statistic $T(\underline{X})$ is not complete for all μ .

- Proof $g \sim N(0, 2\sigma^2)$

Let $Y_1 = -X_1 \sim N(-\mu, \sigma^2)$ and $Y_2 = X_2 \sim N(\mu, \sigma^2)$ are independent random variables with moment generating functions $M_{Y_1}(t) = \exp(-\mu t + \sigma^2 t^2/2)$ and $M_{Y_2}(t) = \exp(\mu t + \sigma^2 t^2/2)$.

Then the moment generating function of the random variable $g(\underline{X}) = Y_2 + Y_1$ is given by

$M_G(t) = M_{Y_1}(t)M_{Y_2}(t) = \exp(\mu t + \sigma^2 t^2/2 - \mu t + \sigma^2 t^2/2) = \exp(\sigma^2 t^2)$. Hence $g \sim N(0, 2\sigma^2)$.

- Or by Location family ancillary statistic.

Let $X_1 = Z_1 + \mu$, $X_2 = Z_2 + \mu$. $Z_1, Z_2 \sim N(0, \sigma^2)$ is free of μ . $g(\underline{X}) = X_2 - X_1$.

$$F_G(g|\mu) = P_{\mu}(G \leq g) = P_{\mu}(X_2 - X_1 \leq g) = P_{\mu}((Z_2 + \mu) - (Z_1 + \mu) \leq g) = P_{\mu}(Z_2 - Z_1 \leq g)$$

The C.D.F. of $g(\underline{X})$ does not depend on μ . $g(\underline{X})$ is an ancillary statistic and is not a complete sufficient statistic.

HW3. (Due 4/22/20)

1. Suppose that X_1, \dots, X_n is a sample from a population with density $f(x, \theta) = \theta x^{\theta-1}, 0 < x < 1, \theta > 0$, and let $T(\underline{X}) = (-1/n) \sum_{i=1}^n \log X_i$. (Note that $X_i \sim \text{Beta}(\theta, 1)$.)

- a. Use Theorem 1.6.2 (Bickel and Doksum) to show that $E(T(\underline{X})) = 1/\theta, \text{Var}(T(\underline{X})) = 1/n\theta^2$.

By Theorem 1.6.2, X is distributed according to

$$q_\theta(\mathbf{x}) = \exp \left[\underbrace{-n(\theta-1)}_{\eta(\theta)} \underbrace{\left(-\frac{1}{n} \sum_{i=1}^n \log(x_i)\right)}_{T(x)} - \underbrace{(-n \log(\theta))}_{A(\eta)} \right] \underbrace{\prod_{i=1}^n \mathbf{1}_{\{x_i \in (0,1)\}}}_{h(x)}, x \in \mathcal{X} \subset \mathbb{R}^q \quad \square$$

$$\eta = -n(\theta-1) \implies \theta = \frac{n-\eta}{n} \implies A(\eta) = -n \log(\theta) = -n \log(n-\eta) + n \log(n)$$

η is an interior point of the natural parameter space ε . □

$$E(T(\underline{X})) = A'(\eta) = \frac{n}{n-\eta} = \frac{n}{n+n(\theta-1)} = \frac{1}{\theta}$$

$$\text{Var}(T(\underline{X})) = A''(\eta) = \frac{n}{(n-\eta)^2} = \frac{n}{[n+n(\theta-1)]^2} = \frac{1}{n\theta^2} \quad \blacksquare$$

- b. Show that $I_1(\theta) = 1/\theta^2$ and that $T(\underline{X})$ is an UMVUE of $q(\theta) = 1/\theta$.

$\{P_\theta\}$ is an exponential family and $\eta(\theta)$ has a nonvanishing continuous derivative on Θ , then I and II hold. (Proposition 3.4.1.)

$$I_1(\theta) = -E \left(\frac{\partial^2}{\partial \theta^2} \log p(X, \theta) \right) = -E \left(\frac{\partial^2}{\partial \theta^2} [(\theta-1) \log(x) + \log(\theta)] \right) = -E \left(\frac{\partial}{\partial \theta} [\log(x) + \frac{1}{\theta}] \right) = \frac{1}{\theta^2} \quad \square$$

$\text{Var}_\theta(T(X))$ is finite and $0 < I(\theta) < \infty$ for all θ . $E_\theta(T(X)) = \psi(\theta)$ is differentiable. (Theorem 3.4.1)

X_1, \dots, X_n is a sample from a population with density $f(x, \theta) = \theta x^{\theta-1}, 0 < x < 1, \theta > 0$, $I(\theta) = nI_1(\theta)$ (Proposition 3.4.2)

$$\text{CRLB} = \frac{(\frac{\partial}{\partial \theta} (E(T(X))))^2}{nI_1(\theta)} = \frac{(\frac{\partial}{\partial \theta} (\frac{1}{\theta}))^2}{n \frac{1}{\theta^2}} = \frac{1}{n\theta^2}$$

- By Theorem 3.4.2.

$\{P_\theta\}$ is a one-parameter exponential family of the form $p(x, \theta) = h(x) \exp\{\eta(\theta)T(x) - B(\theta)\}$ □

$T(X) = -\frac{1}{n} \sum_{i=1}^n \log(x_i)$ is a function of natural sufficient statistic $\sum_{i=1}^n \log(x_i)$ □

$\eta(\theta) = -n(\theta-1)$ has a continuous nonvanishing derivative on θ □

Then $T(X)$ achieves the information inequality bound and is a UMVUE of $E_\theta(T(X)) = 1/\theta$. ■

- Or By Lehman-Scheffe Theorem?

$\text{Beta}(\theta, 1)$ is a canonical 1-parameter exponential family. $T(\underline{X}) = -\frac{1}{n} \sum_{i=1}^n \log(x)$ is a function of complete sufficient statistic for θ , $T(\underline{X})$ is the UMVUE of θ .

2. Do problem 1 with $f(x, \theta)$ replaced by the Weibull density $f(x, \theta) = cx^{c-1}\theta \exp(-\theta x^c), x > 0, \theta > 0, c > 0$, and T replaced by $T(\underline{X}) = (1/n) \sum_{i=1}^n X_i^c$.

By Theorem 1.6.2, X is distributed according to

$$f_{\theta}(\mathbf{x}) = \exp \left[\underbrace{-n\theta}_{\eta(\theta)} \underbrace{\left(\frac{1}{n} \sum_{i=1}^n x_i^c \right)}_{T(\mathbf{x})} - \underbrace{(-n \log(\theta))}_{A(\eta)} \underbrace{\prod_{i=1}^n c x_i^{c-1} \mathbf{1}_{\{x_i > 0\}}}_{h(\mathbf{x})} \right] \quad \square$$

$$\eta = -n\theta \implies \theta = \frac{-\eta}{n} \implies A(\eta) = -n \log(\theta) = -n \log(-\eta) + n \log(n)$$

η is an interior point of the natural parameter space ε . □

$$E(T(\underline{X})) = A'(\eta) = \frac{-n}{\eta} = \frac{-n}{-n\theta} = \frac{1}{\theta}$$

$$\text{Var}(T(\underline{X})) = A''(\eta) = \frac{n}{\eta^2} = \frac{n}{(-n\theta)^2} = \frac{1}{n\theta^2} \quad \blacksquare$$

$\{P_{\theta}\}$ is an exponential family and $\eta(\theta)$ has a nonvanishing continuous derivative on Θ , then I and II hold. (Proposition 3.4.1.)

$$I_1(\theta) = -E \left(\frac{\partial^2}{\partial \theta^2} \log p(X, \theta) \right) = -E \left(\frac{\partial^2}{\partial \theta^2} [-\theta x^c + \log(\theta)] \right) = -E \left(\frac{\partial}{\partial \theta} [-x^c + \frac{1}{\theta}] \right) = \frac{1}{\theta^2} \quad \square$$

$\text{Var}_{\theta}(T(X))$ is finite and $0 < I(\theta) < \infty$ for all θ . $E_{\theta}(T(X)) = \psi(\theta)$ is differentiable. (Theorem 3.4.1)

X_1, \dots, X_n is a sample from a population with density $f(x, \theta) = \theta x^{\theta-1}, 0 < x < 1, \theta > 0$, $I(\theta) = nI_1(\theta)$ (Proposition 3.4.2)

$$\text{CRLB} = \frac{(\frac{\partial}{\partial \theta}(E(T(X))))^2}{nI_1(\theta)} = \frac{(\frac{\partial}{\partial \theta}(\frac{1}{\theta}))^2}{n \frac{1}{\theta^2}} = \frac{1}{n\theta^2}$$

- By Theorem 3.4.2.

$\{P_{\theta}\}$ is a one-parameter exponential family of the form $p(x, \theta) = h(x) \exp\{\eta(\theta)T(x) - B(\theta)\}$ □

$T(X) = \frac{1}{n} \sum_{i=1}^n x_i^c$ is a function of natural sufficient statistic $\sum_{i=1}^n x_i^c$ □

$\eta(\theta) = -n\theta$ has a continuous nonvanishing derivative on θ □

Then $T(X)$ achieves the information inequality bound and is a UMVUE of $E_{\theta}(T(X)) = 1/\theta$. ■

- Or By Lehman-Scheffe Theorem

When c is known, $Weibull(\theta, c)$ is a canonical 1-parameter exponential family. $T(\underline{X}) = \frac{1}{n} \sum_{i=1}^n x_i^c$ is a function of complete sufficient statistic for θ , $T(\underline{X})$ is the UMVUE of θ .

3. Show that the information inequality bound is unchanged by a smooth reparametrization. That is, suppose $P_{\theta}, \theta \in \Theta$ satisfies I and II. Let h be a function from Θ to \mathbb{R} such that h_0 is a continuous and nonvanishing on Θ . Let $\eta = h(\theta)$ and define $Q_{\eta} = P_{\theta}$. Then the information inequality bound obtained from Q_{η} evaluated at $\eta = h(\theta)$ is the same as the bound obtained from P_{θ} .

Hint: $\frac{\partial}{\partial \eta} \log q(x, \eta) = \left[\frac{\partial}{\partial \theta} \log p(x, \theta) \right] \frac{\partial \theta}{\partial \eta}, \frac{\partial}{\partial \eta} E_{\eta}(T(X)) = \left(\frac{\partial}{\partial \theta} \psi(\theta) \right) \frac{\partial \theta}{\partial \eta}$, where $\psi(\theta) = E_{\theta}(T(X))$.

$$Q_{\eta} = P_{\theta} \implies \frac{\partial}{\partial \eta} Q_{\eta} = \frac{\partial}{\partial \theta} P_{\theta} \frac{\partial \theta}{\partial \eta}$$

$$\left[\frac{\partial}{\partial \eta} E_{\eta}(T(X)) \right]^2 = \left[\left(\frac{\partial}{\partial \theta} \psi(\theta) \right) \frac{\partial \theta}{\partial \eta} \right]^2 = \left[\left(\frac{\partial}{\partial \theta} E_{\theta}(T(X)) \right) \right]^2 \left(\frac{\partial \theta}{\partial \eta} \right)^2$$

$$\begin{aligned}
 I(\eta) &= E \left[\left(\frac{\partial}{\partial \eta} \log q(x, \eta) \right)^2 \right] = E \left[\left(\left\{ \frac{\partial}{\partial \theta} \log p(x, \theta) \right\} \frac{\partial \theta}{\partial \eta} \right)^2 \right] = E \left[\left(\frac{\partial}{\partial \theta} \log p(x, \theta) \right)^2 \right] \left(\frac{\partial \theta}{\partial \eta} \right)^2 \\
 \frac{\left[\frac{\partial}{\partial \eta} E_\eta(T(X)) \right]^2}{I(\eta)} &= \frac{\left[\left(\frac{\partial}{\partial \theta} E_\theta(T(X)) \right)^2 \right] \left(\frac{\partial \theta}{\partial \eta} \right)^2}{E \left[\left(\frac{\partial}{\partial \theta} \log p(x, \theta) \right)^2 \right] \left(\frac{\partial \theta}{\partial \eta} \right)^2} = \frac{\left[\frac{\partial}{\partial \theta} E_\theta(T(X)) \right]^2}{I(\theta)}
 \end{aligned}$$

Hence, the information inequality bound obtained from Q_η evaluated at $\eta = h(\theta)$ is the same as the bound obtained from P_θ

HW4. (Due 4/29/20)

1. Let X_1, \dots, X_n be a sample from a population with frequency distribution $f(1, \theta) = (1/3)(1 - \theta)$, $f(2, \theta) = (1/3)(1 + c\theta)$, $f(3, \theta) = (1/3)[1 + \theta(1 - c)]$, where c is a constant. Give two frequency substitution estimates of θ and compute their asymptotic variances.

Let $N_1 = np_1$, $N_2 = np_2$, $N_3 = np_3$, $\sum_{i=1}^3 N_i = n$

$$\frac{3(p_2 - p_1)}{c+1} = \frac{(1+c\theta) - (1-\theta)}{c+1} = \theta \implies T_1 = \frac{3}{c+1} \left(\frac{N_2}{n} - \frac{N_1}{n} \right)$$

$$\frac{3(p_1 - p_3)}{c-2} = \frac{(1-\theta) - (1+\theta(1-c))}{c-2} = \theta \implies T_2 = \frac{3}{c-2} \left(\frac{N_1}{n} - \frac{N_3}{n} \right)$$

$$\frac{3(p_2 - p_3)}{2c-1} = \frac{(1+c\theta) - (1+\theta(1-c))}{2c-1} = \theta \implies T_3 = \frac{3}{2c-1} \left(\frac{N_2}{n} - \frac{N_3}{n} \right)$$

The frequency substitution estimates of θ are $T_{1,2,3} = h(\frac{\mathbf{N}}{n})$ satisfies $h(\mathbf{p}(\theta)) = \theta$ for all $\theta \in \Theta$ where $\mathbf{p}(\theta) = (p_1, p_2, p_3)^T = (f(1, \theta), f(2, \theta), f(3, \theta))^T$

$h(\mathbf{p}(\theta))$ are 3 continuous functions at $\mathbf{p}(\theta)$ and are differentiable.

- By Delta Method (Theorem 5.4.1),

$$\sqrt{n} \left(h\left(\frac{\mathbf{N}}{n}\right) - \theta \right) \xrightarrow{\mathcal{D}} N(0, \sigma^2(\theta, h(\mathbf{p}(\theta))))$$

$$\sigma^2(\theta, h(p_1, p_2)) = \left(\left[\frac{3}{c+1} \right]^2 \frac{1+c\theta}{3} + \left[\frac{-3}{c+1} \right]^2 \frac{1-\theta}{3} \right) - \left(\frac{3}{c+1} \frac{1+c\theta}{3} + \frac{-3}{c+1} \frac{1-\theta}{3} \right)^2 = \frac{6-3(1-c)\theta}{(c+1)^2} - \theta^2$$

$$\sigma^2(\theta, h(p_1, p_3)) = \left(\left[\frac{3}{c-2} \right]^2 \frac{1-\theta}{3} + \left[\frac{-3}{c-2} \right]^2 \frac{1+(1-c)\theta}{3} \right) - \left(\frac{3}{c-2} \frac{1-\theta}{3} + \frac{-3}{c-2} \frac{1+(1-c)\theta}{3} \right)^2 = \frac{6-3c\theta}{(c-2)^2} - \theta^2$$

$$\sigma^2(\theta, h(p_2, p_3)) = \left(\left[\frac{3}{2c-1} \right]^2 \frac{1+c\theta}{3} + \left[\frac{-3}{2c-1} \right]^2 \frac{1+(1-c)\theta}{3} \right) - \left(\frac{3}{2c-1} \frac{1+c\theta}{3} + \frac{-3}{2c-1} \frac{1+(1-c)\theta}{3} \right)^2 = \frac{6+3\theta}{(2c-1)^2} - \theta^2$$

- Note: (Example 1.6.7; Example 2.1.3; Example 2.1.4; Example 2.2.8; Example 2.3.3; Example 3.4.4; Example 3.4.7; Example 4.1.1; Example 4.3.1; Section 5.4.1)

By Theorem 5.3.2

$$\sqrt{n} \left(h\left(\frac{\mathbf{N}}{n}\right) - h(\mathbf{p}(\theta)) \right) = \sqrt{n} \sum_{j=1}^3 \frac{\partial h}{\partial p_j}(\mathbf{p}(\theta)) \left(\frac{N_j}{n} - p(x_j, \theta) \right) + \mathcal{O}_p(1)$$

Using the definition of N_j

$$\sum_{j=1}^3 \frac{\partial h}{\partial p_j}(\mathbf{p}(\theta)) \left(\frac{N_j}{n} - p(x_j, \theta) \right) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^3 \frac{\partial h}{\partial p_j}(\mathbf{p}(\theta)) (\mathbf{1}_{(x_i=x_j)} - p(x_j, \theta))$$

Thus, $\sqrt{n} (h(\frac{\mathbf{N}}{n}) - h(\mathbf{p}(\theta)))$ is asymptotically normal with mean 0,

Its asymptotic variance is

$$\sigma^2(\theta, h) = \text{Var}_\theta \left[\sum_{j=1}^3 \frac{\partial h}{\partial p_j}(\mathbf{p}(\theta)) \mathbf{1}_{(x_1=x_j)} \right] = \sum_{j=1}^3 \left[\frac{\partial h}{\partial p_j}(\mathbf{p}(\theta)) \right]^2 p(x_j, \theta) - \left[\sum_{j=1}^3 \frac{\partial h}{\partial p_j}(\mathbf{p}(\theta)) p(x_j, \theta) \right]^2$$

- Or by the property of Multi-nomial Distribution

$$\text{Var}(N_1) = n(p_1)(1 - p_1) = \frac{n}{3^2}(1 - \theta)(2 + \theta)$$

$$\text{Var}(N_2) = n(p_2)(1 - p_2) = \frac{n}{3^2}(1 + c\theta)(2 - c\theta)$$

$$\text{Var}(N_3) = n(p_3)(1 - p_3) = \frac{n}{3^2}(1 + (1 - c)\theta)(2 - \theta + c\theta)$$

$$\text{Cov}(N_1, N_2) = -np_1p_2 = -\frac{n}{3^2}(1 - \theta)(1 + c\theta)$$

$$\text{Cov}(N_1, N_3) = -np_1p_3 = -\frac{n}{3^2}(1 - \theta)(1 + (1 - c)\theta)$$

$$\text{Cov}(N_2, N_3) = -np_2p_3 = -\frac{n}{3^2}(1 + c\theta)(1 + (1 - c)\theta)$$

$$n\text{Var}T_1 = \frac{3^2(\text{Var}N_1 + \text{Var}N_2 - 2\text{Cov}(N_1, N_2))}{(c+1)^2n} = \frac{(1-\theta)(2+\theta) + (1+c\theta)(2-c\theta) + 2(1-\theta)(1+c\theta)}{(c+1)^2} = \frac{6-3(1-c)\theta}{(c+1)^2} - \theta^2$$

$$n\text{Var}T_2 = \frac{3^2(\text{Var}N_1 + \text{Var}N_3 - 2\text{Cov}(N_1, N_3))}{(c-2)^2n} = \frac{(1-\theta)(2+\theta) + (1+(1-c)\theta)(2-\theta+c\theta) + 2(1-\theta)(1+(1-c)\theta)}{(c-2)^2} = \frac{6-3c\theta}{(c-2)^2} - \theta^2$$

$$n\text{Var}T_3 = \frac{3^2(\text{Var}N_2 + \text{Var}N_3 - 2\text{Cov}(N_2, N_3))}{(2c-1)^2n} = \frac{(1+c\theta)(2-c\theta) + (1+(1-c)\theta)(2-\theta+c\theta) + 2(1+c\theta)(1+(1-c)\theta)}{(2c-1)^2} = \frac{6+3\theta}{(2c-1)^2} - \theta^2$$

2. Let X_1, \dots, X_n be a sample from a population with the density $f(x, \theta) = \theta(\theta + 1)x^{\theta-1}(1 - x), 0 < x < 1, \theta > 0$.

- a. Show that $T_n = \frac{2\bar{X}}{1-\bar{X}}$ is a method of moments estimate of θ .

For $X \sim \text{Beta}(\theta, 2)$,

$$EX = \frac{\theta}{\theta + 2} \stackrel{\text{set}}{=} \frac{1}{n} \sum_i X_i = \bar{X} \implies \hat{\theta}_{MOM} = \frac{2\bar{X}}{1-\bar{X}}$$

$T_n = \frac{2\bar{X}}{1-\bar{X}}$ is a method of moments estimate of θ .

- b. Show that $\frac{\sqrt{n}(T_n - \mu_n(\theta))}{\sigma_n(\theta)} \rightarrow N(0, 1)$ in distribution, where $\mu_n(\theta) = \theta, \sigma_n^2(\theta) = \theta(\theta + 2)^2/2(\theta + 3)$.

$$E[\bar{X}] = \mu = \frac{\theta}{\theta+2}, n\text{Var}[\bar{X}] = \sigma^2 = \frac{2\theta}{(\theta+2)^2(\theta+3)}. \text{ By the Central Limit Theorem, } \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{\mathcal{D}} N(0, 1)$$

$$\text{Let } T_n = \frac{2\bar{X}}{1-\bar{X}} = \frac{2}{1-\bar{X}} - 2 = h(\bar{X}).$$

$$h(\mu) = \frac{2}{1-\mu} - 2 \text{ is a continuous function at } \frac{\theta}{\theta+2} \text{ and differentiable.}$$

$$h'(\mu) = \frac{2}{(1-\mu)^2} = \frac{2}{(1-\frac{\theta}{\theta+2})^2} = \frac{(\theta+2)^2}{2} \neq 0 \text{ and exist.}$$

By Delta Method

$$\sqrt{n}(h(\bar{X}) - h(\mu)) = \sqrt{n}(T_n - \frac{2}{1-\mu} + 2) = \sqrt{n}(T_n - \frac{2}{1-\frac{\theta}{\theta+2}} + 2) = \sqrt{n}(T_n - \theta)$$

$$\sigma^2[h'(\mu)]^2 = \frac{2\theta}{(\theta+2)^2(\theta+3)} \left[\frac{(\theta+2)^2}{2} \right]^2 = \frac{\theta(\theta+2)^2}{2(\theta+3)}$$

$$\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{D}} N(0, \frac{\theta(\theta+2)^2}{2(\theta+3)}) \implies \frac{\sqrt{n}(T_n - \theta)}{\sqrt{\frac{\theta(\theta+2)^2}{2(\theta+3)}}} \xrightarrow{\mathcal{D}} N(0, 1)$$

Therefore, $\frac{\sqrt{n}(T_n - \mu_n(\theta))}{\sigma_n(\theta)} \rightarrow N(0, 1)$ in distribution, where $\mu_n(\theta) = \theta, \sigma_n^2(\theta) = \frac{\theta(\theta+2)^2}{2(\theta+3)}$

c. Show that T_n is not efficient by calculating the information bound.

T_n is an unbiased estimator of θ . $Var[T_n] = \frac{1}{n} \sigma_n^2(\theta) = \frac{\theta(\theta+2)^2}{2n(\theta+3)}$

$$\log f(x|\theta) = \log \theta + \log(\theta + 1) + (\theta - 1) \log x + \log(1 - x)$$

$$\frac{\partial^2}{\partial \theta^2} \log f(x|\theta) = \frac{\partial}{\partial \theta} \left(\frac{1}{\theta} + \frac{1}{\theta+1} + \log x \right) = -\frac{1}{\theta^2} - \frac{1}{(\theta+1)^2}$$

$$I(\theta) = -E\left[\frac{\partial^2}{\partial \theta^2} \log f(\mathbf{x}|\theta)\right] = -E\left[-\frac{1}{\theta^2} - \frac{1}{(\theta+1)^2}\right] = \frac{2\theta^2 + 2\theta + 1}{\theta^2(\theta+1)^2}$$

$$CRLB = \frac{1}{nI(\theta)} = \frac{\theta^2(\theta+1)^2}{n(2\theta^2 + 2\theta + 1)}$$

$$\frac{CRLB}{Var[T_n]} = \frac{\frac{\theta^2(\theta+1)^2}{n(2\theta^2 + 2\theta + 1)}}{\frac{\theta(\theta+2)^2}{2n(\theta+3)}} = \frac{2\theta(\theta+3)(\theta+1)^2}{(\theta+2)^2(2\theta^2 + 2\theta + 1)} = \frac{2\theta^4 + 10\theta^3 + 14\theta^2 + 6\theta}{2\theta^4 + 10\theta^3 + 17\theta^2 + 12\theta + 4} < 1$$

Thus, T_n is not efficient.

HW5. (Due 5/13/20)

1. Let $X_n \text{ Bin}(n, p)$, $p \in (0, 1)$. Example 5.2.2. Section 5.3

- Recall the moment-generating function

X_n be binomial with n trials and success probability p . $X_n = \sum_{i=1}^n X_i$ where the X_i are iid Bernoulli with success probability p

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} \cdot \exp \left\{ x \ln\left(\frac{p}{1-p}\right) + n \ln(1-p) \right\}$$

$\eta = \ln\left(\frac{p}{1-p}\right)$ is an interior point of ε , By Theorem 1.6.2 the moment-generating function of $T(X)$ exist and is given by

$$M_X(t) = \exp[A(t + \eta) - A(\eta)] = [pe^t + 1 - p]^n$$

$$M'_X(t) = \frac{n!}{(n-1)!} [pe^t + 1 - p]^{n-1} pe^t$$

$$M''_X(t) = \frac{n!}{(n-2)!} [pe^t + 1 - p]^{n-2} (pe^t)^2 + \frac{n!}{(n-1)!} [pe^t + 1 - p]^{n-1} pe^t$$

$$M'''_X(t) = \frac{n!}{(n-3)!} [pe^t + 1 - p]^{n-3} (pe^t)^3 + \frac{3n!}{(n-2)!} [pe^t + 1 - p]^{n-1} (pe^t)^2 + \frac{n!}{(n-1)!} [pe^t + 1 - p]^{n-1} pe^t$$

$$M''''_X(t) = \frac{n!}{(n-4)!} [pe^t + 1 - p]^{n-4} (pe^t)^4 + \frac{6n!}{(n-3)!} [pe^t + 1 - p]^{n-3} (pe^t)^3 + \frac{7n!}{(n-2)!} [pe^t + 1 - p]^{n-1} (pe^t)^2 + \frac{n!}{(n-1)!} [pe^t + 1 - p]^{n-1} pe^t$$

$$\mu_1 = E[X^1] = np$$

$$\mu_2 = E[X^2] = n(n-1)p^2 + np$$

$$\mu_3 = E[X^3] = n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np$$

$$\mu_4 = E[X^4] = n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np$$

- a. Find the UMVUE of p^2 .

Each X_i 's are iid Bernoulli random variables. $(x_i \in \mathbb{N}^+)$ is the natural sufficient statistic for p .

Binomial distribution is a 1-parameter exponential family. When n is known, the total number of successes X_n is the complete sufficient statistic for p .

Let $T(X_n)$ is a function of c.s.s,

$$E[T(X_n)] = E\left[\frac{X_n^2 - X_n}{n(n-1)}\right] = \frac{E[X_n^2] - E[X_n]}{n(n-1)} = \frac{n(n-1)p^2 + np - np}{n(n-1)} = p^2$$

$T(X_n)$ is an unbiased estimator for p^2 , By Lehmann Scheffe Theorem, $T = \frac{X_n^2 - X_n}{n(n-1)}$ the UMVUE of p^2 .

- b. Find the asymptotic distribution of your estimator found in a.

$$E\left[\frac{X_n}{n}\right] = \mu = p, n\text{Var}\left[\frac{X_n}{n}\right] = \sigma^2 = p(1-p).$$

By the Central Limit Theorem, $\sqrt{n}\left(\frac{X_n}{n} - p\right) \xrightarrow{D} N(0, p(1-p))$

$$\text{Let } T_n = \frac{X_n^2 - X_n}{n(n-1)} = \frac{n\left(\frac{X_n}{n}\right)^2 - \frac{X_n}{n}}{n-1} = h\left(\frac{X_n}{n}\right).$$

$h(\mu) = \frac{n\mu^2 - \mu}{n-1}$ is a continuous function at p and differentiable;

$h'(\mu) = \frac{2n\mu - 1}{n-1} \neq 0$ when $\mu \neq \frac{1}{2n}$; $h''(\mu) = \frac{2n}{n-1} \neq 0$; $h'''(\mu) = 0$.

In the case $np = 1/2$, this does not give the asymptotic distribution.

By Delta Method

$$\sqrt{n}\left(h\left(\frac{X}{n}\right) - h(\mu) - \frac{1}{2n}h''(\mu)\sigma^2\right) = \sqrt{n}\left(h\left(\frac{X}{n}\right) - \frac{np^2 - p}{n-1} - \frac{1}{2n}\frac{2n}{n-1}(p - p^2)\right) = \sqrt{n}(T - p^2)$$

$$\lim_{n \rightarrow \infty} \sigma^2 [h'(\mu)]^2 = (p - p^2) \lim_{n \rightarrow \infty} \left[\frac{2np - 1}{n - 1} \right]^2 = 4p^3(1 - p)$$

Therefore, the asymptotic distribution of $T(\frac{X_n}{n})$ is

$$\sqrt{n} \left(T - p^2 \right) \xrightarrow{\mathcal{D}} N(0, 4p^3(1 - p)) \quad \blacksquare$$

Try using Multivariate Central Limit Theorem. It might be unnecessary.

$$E[X_n] = \mu_1 = np, E[X_n^2] = \mu_2 = n(n-1)p^2 + np.$$

$$\mu_4 = E[X_n^4] = n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + np \text{ is finite.}$$

By Multivariate Central Limit Theorem, $\sqrt{n}((X_n, X_n^2) - (\mu_1, \mu_2)) \xrightarrow{\mathcal{D}} N(0, \Sigma)$

$$\text{where } \Sigma = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_1^2) \\ \text{Cov}(X_1, X_1^2) & \text{Var}(X_1^2) \end{bmatrix} = \begin{bmatrix} \mu_2 - \mu_1^2 & \mu_3 - \mu_1\mu_2 \\ \mu_3 - \mu_1\mu_2 & \mu_4 - \mu_2^2 \end{bmatrix}$$

$$\text{Var}[X_1] = \mu_2 - \mu_1^2 = [n(n-1)p^2 + np] - [n^2p^2] = np(1-p)$$

$$\text{Cov}[X_1, X_1^2] = \mu_3 - \mu_1\mu_2 = [n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np] - [np][n(n-1)p^2 + p]$$

$$= -2n(n-1)p^3 + n(2n-3)p^2 + np$$

$$\text{Var}[X_1^2] = \mu_4 - \mu_2^2 = [n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 + 7n(n-1)p^2 + p] - [n(n-1)p^2 + np]^2$$

$$= -2n(n-1)(2n-3)p^4 + 4n(n-1)(n-3)p^3 + n(6n-7)p^2 + np$$

$$J(\mu_1, \mu_2) = \left(\frac{\partial f}{\partial t_1}, \frac{\partial f}{\partial t_2} \right) \Big|_{(t_1, t_2) = (\mu_1, \mu_2)} = [-1 \quad 1], f(t_1, t_2) = t_2 - t_1$$

$$J\Sigma J' = [-1 \quad 1] \begin{bmatrix} \mu_2 - \mu_1^2 & \mu_3 - \mu_1\mu_2 \\ \mu_3 - \mu_1\mu_2 & \mu_4 - \mu_2^2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \mu_4 - \mu_2^2 - 2(\mu_3 - \mu_1\mu_2) + \mu_2 - \mu_1^2$$

$$= 2n(n-1)p^2(2np-3p+1)(1-p) = \eta^2$$

Let $T(X_n, X_n^2) = \frac{X_n^2 - X_n}{n(n-1)} = \frac{f(X_n, X_n^2)}{n-1}$, where $f(u, v) = v - u$. By Delta Method,

$$\sqrt{n}(T - p^2) = \sqrt{n} \left(\frac{X_n^2 - X_n}{n(n-1)} - \frac{(n-1)p^2 + p - p}{n-1} \right) = \frac{1}{n(n-1)} \sqrt{n} (f(X_n, X_n^2) - f(\mu_1, \mu_2))$$

$$E\left[\frac{1}{n(n-1)} \sqrt{n} (f(X_n, X_n^2) - f(\mu_1, \mu_2))\right] = 0$$

$$\lim_{n \rightarrow \infty} n \text{Var}\left[\frac{1}{n(n-1)} \sqrt{n} (f(\bar{X}_n, \bar{X}_n^2) - f(\bar{\mu}_1, \bar{\mu}_2))\right] = \lim_{n \rightarrow \infty} \frac{\eta^2}{n(n-1)^2} = \lim_{n \rightarrow \infty} \frac{2p^2(2np-3p+1)(1-p)}{n-1} = 4p^3(1-p)$$

Therefore, $\sqrt{n}(T - p^2) = \frac{\sqrt{n}}{n-1} (f(X_n, X_n^2) - f(\mu_1, \mu_2)) \xrightarrow{\mathcal{D}} N(0, 4p^3(1-p))$

- c. Find an MLE of p and use it to find an MLE of p^2 . Find the asymptotic distribution of the MLE of p^2 . Is it asymptotically efficient?

The likelihood function to be maximized for binomial distribution is given as

$$L(p) = \binom{n}{X_n} p^{X_n} (1-p)^{n-X_n}$$

$$l(p) = \log \binom{n}{X_n} X_n \log(p) - X_n \log(1-p) + n \log(1-p)$$

$$\frac{\partial l(p)}{\partial p} = \frac{X_n}{p} + \frac{X_n}{1-p} - \frac{n}{1-p} \stackrel{\text{set}}{=} 0 \implies \hat{p}_{MLE} = \frac{X_n}{n}$$

$$E[\hat{p}_{MLE}] = \frac{E[X_n]}{n} = \frac{np}{n} = p, \text{Var}[\hat{p}_{MLE}] = \frac{\text{Var}[X_n]}{n^2} = \frac{np(1-p)}{n^2} = \frac{p-p^2}{n}$$

In the Bernoulli case $I(p) = \frac{1}{p(1-p)}, 0 < p < 1$.

$$\text{the estimating sequence } \hat{p} = \begin{cases} c, & \text{if } X = 0, n, \\ \frac{X_n}{n}, & \text{if } 0 < X < n \end{cases}$$

Under the Cramer-Rao conditions for asymptotic normality, $\sqrt{n}(\hat{p}_{MLE} - p) \xrightarrow{\mathcal{L}} N(0, I(p)^{-1}) = N(0, p(1-p))$

$$E[\hat{p}_{MLE}^2] \approx \lim_{n \rightarrow \infty} (\text{Var}[\hat{p}_{MLE}] + E[\hat{p}_{MLE}]^2) = \lim_{n \rightarrow \infty} \left(\frac{p-p^2}{n} + p^2 \right) = p^2$$

Let the asymptotically unbiased estimator $E[\hat{p}_{MLE}^2] \approx p^2 = \psi(p), \psi'(p) = 2p$

$$n\text{Var}[\hat{p}_{MLE}^2] \approx n(\psi'(p))^2 \text{Var}\left[\frac{X_n}{n}\right] = 4np^2 \cdot \frac{p-p^2}{n} = 4p^3(1-p)$$

$$\sqrt{n}(\hat{p}_{MLE}^2 - p^2) \xrightarrow{a} N(0, 4p^3(1-p))$$

$$\begin{aligned} I(p) &= -E \left[\frac{\partial^2}{\partial p^2} \log f(x, p) \right] = -E \left[\frac{\partial}{\partial p} \left(\frac{X_n}{p} - \frac{X_n - n}{1-p} \right) \right] \\ &= \frac{E[X_n]}{p^2} + \frac{n - E[X_n]}{(1-p)^2} = \frac{np}{p^2} + \frac{n - np}{(1-p)^2} = \frac{n}{p(1-p)} \end{aligned}$$

$$\text{CRLB} = \frac{(\psi'(p))^2}{I(p)} = \frac{4p^2}{\frac{n}{p(1-p)}} = \frac{4p^3(1-p)}{n}$$

$$\frac{\text{CRLB}}{\text{Var}_p[\hat{p}_{MLE}^2]} = \frac{\frac{4p^3(1-p)}{n}}{\frac{4p^3(1-p)}{n}} = 1$$

The the MLE of p^2 is efficient.

HW6. (Due 5/27/20)

1. Define $(Z_i, \Delta_i) \equiv (\min(X_i, Y_i), 1_{\{X_i < Y_i\}})$, where X_i 's and Y_i 's are statistically independent, and $X_i \stackrel{iid}{\sim} f_\lambda(x) = \lambda e^{-\lambda x} 1_{R^+}(x)$, $\lambda > 0$; $Y_i \stackrel{iid}{\sim} g_{\mu_0}(y) = \mu_0 e^{-\mu_0 y} 1_{R^+}(y)$, $\mu_0 > 0$, where μ_0 is known.

$$\begin{aligned}
 P(\Delta = 1) &= P(X \leq Y) = \int_0^\infty \int_0^y f_{X,Y}(x,y) dx dy \\
 &\stackrel{X \perp Y}{=} \int_0^\infty \int_0^y \lambda e^{-\lambda x} \cdot \mu_0 e^{-\mu_0 y} dx dy = \int_0^\infty \mu_0 e^{-\mu_0 y} \left(\int_0^y \lambda e^{-\lambda x} dx \right) dy = \int_0^\infty \mu_0 e^{-\mu_0 y} (-e^{-\lambda x} \Big|_0^y) dy \\
 &= \int_0^\infty \mu_0 e^{-\mu_0 y} (1 - e^{-\lambda y}) dy = \int_0^\infty \mu_0 e^{-\mu_0 y} dy - \frac{\mu_0}{\mu_0 + \lambda} \int_0^\infty (\mu_0 + \lambda) e^{-(\mu_0 + \lambda)y} dy \\
 &= 1 - \frac{\mu_0}{\mu_0 + \lambda} (-e^{-(\mu_0 + \lambda)y} \Big|_0^\infty) = 1 - \frac{\mu_0}{\mu_0 + \lambda} = \frac{\lambda}{\mu_0 + \lambda}
 \end{aligned}$$

Then, $P(\Delta = 0) = 1 - \frac{\lambda}{\mu_0 + \lambda} = \frac{\mu_0}{\mu_0 + \lambda}$

$\Delta \sim \text{Bernoulli}(\frac{\lambda}{\mu_0 + \lambda})$; $f_\Delta(\delta) = (\frac{\lambda}{\mu_0 + \lambda})^\delta (\frac{\mu_0}{\mu_0 + \lambda})^{1-\delta}$

$D_n = \sum_{i=1}^n \Delta_i \sim \text{Bino}(n, \frac{\lambda}{\mu_0 + \lambda})$ □

$E[D_n] = \frac{n\lambda}{\mu_0 + \lambda}$;

$V[D_n] = \frac{n\lambda}{\mu_0 + \lambda} (1 - \frac{\lambda}{\mu_0 + \lambda}) = \frac{n\lambda\mu_0}{(\mu_0 + \lambda)^2}$

$E[D_n^2] = V[D_n] + (E[D_n])^2 = \frac{n\lambda\mu_0}{(\mu_0 + \lambda)^2} + (\frac{n\lambda}{\mu_0 + \lambda})^2 = \frac{n^2\lambda^2 + n\lambda\mu_0}{(\mu_0 + \lambda)^2}$ □

.....

$$\begin{aligned}
 F_Z(z) &= P(X \leq z \cup Y \leq z) = 1 - P(X \geq z \cap Y \geq z) \stackrel{X \perp Y}{=} 1 - P(X \geq z)(Y \geq z) \\
 &= 1 - \int_z^\infty \lambda e^{-\lambda x} dx \cdot \int_z^\infty \mu_0 e^{-\mu_0 y} dy = 1 - (-e^{-\lambda x} \Big|_z^\infty) \cdot (-e^{-\mu_0 y} \Big|_z^\infty) = 1 - e^{-(\lambda + \mu_0)z}
 \end{aligned}$$

$f(z) = (\lambda + \mu_0) e^{-(\lambda + \mu_0)z}$; $Z \sim \text{Expo}(\lambda + \mu_0) = \text{Gamma}(1, (\lambda + \mu_0)^{-1})$

$T_n = \sum_{i=1}^n Z_i \sim \text{Gamma}(n, (\lambda + \mu_0)^{-1})$ □

$$\begin{aligned}
 E[T_n^{-1}] &= \int_0^\infty w^{-1} \frac{(\lambda + \mu_0)^n}{\Gamma(n)} w^{n-1} e^{-(\lambda + \mu_0)w} dw \\
 &= \frac{(\lambda + \mu_0) \Gamma(n-1)}{\Gamma(n)} \underbrace{\int_0^\infty \frac{(\lambda + \mu_0)^{n-1}}{\Gamma(n-1)} w^{n-1-1} e^{-(\lambda + \mu_0)w} dw}_{=1} = \frac{\lambda + \mu_0}{n-1} \\
 E[T_n^{-2}] &= \int_0^\infty w^{-2} \frac{(\lambda + \mu_0)^n}{\Gamma(n)} w^{n-1} e^{-(\lambda + \mu_0)w} dw \\
 &= \frac{(\lambda + \mu_0)^2 \Gamma(n-2)}{\Gamma(n)} \underbrace{\int_0^\infty \frac{(\lambda + \mu_0)^{n-2}}{\Gamma(n-2)} w^{n-2-1} e^{-(\lambda + \mu_0)w} dw}_{=1} = \frac{(\lambda + \mu_0)^2}{(n-1)(n-2)}
 \end{aligned}$$

Or by $T_n^{-1} \sim \text{Inverse-Gamma}(n, (\lambda + \mu_0)^{-1})$,

$E[T_n^{-1}] = \frac{(\lambda + \mu_0) \Gamma(n-1)}{\Gamma(n)} = \frac{\lambda + \mu_0}{n-1}$;

$E[T_n^{-2}] = \frac{(\lambda + \mu_0)^2 \Gamma(n-2)}{\Gamma(n)} = \frac{(\lambda + \mu_0)^2}{(n-1)(n-2)}$

$$\text{Var}[T_n^{-1}] = E[T_n^{-2}] - (E[T_n^{-1}])^2 = \frac{(\lambda+\mu_0)^2}{(n-1)(n-2)} - \left(\frac{\lambda+\mu_0}{n-1}\right)^2 = \frac{(\lambda+\mu_0)^2}{(n-1)^2(n-2)} \quad \square$$

.....

$$\begin{aligned} P(\Delta = 1, Z \leq z) &= P(X \leq Y \cap Z \leq z) = P(X \leq Y \cap X \leq z) \\ &= \int_0^z \int_x^\infty f_{X,Y}(x,y) dy dx = \int_0^z \lambda e^{-\lambda x} \left(\int_x^\infty \mu_0 e^{-\mu_0 y} dy \right) dx \\ &= \int_0^z \lambda e^{-\lambda x} (-e^{-\mu_0 y}|_x^\infty) dx = \int_0^z \lambda e^{-(\mu_0+\lambda)x} dx = \frac{\lambda}{\mu_0+\lambda} \int_0^z (\mu_0+\lambda) e^{-(\mu_0+\lambda)x} dx \\ &= \frac{\lambda}{\mu_0+\lambda} (-e^{-(\mu_0+\lambda)x}|_0^z) = \frac{\lambda}{\mu_0+\lambda} (1 - e^{-(\mu_0+\lambda)z}) = P(\Delta = 1) \cdot P(Z \leq z) \\ P(\Delta = 0, Z \leq z) &= P(Y \leq X \cap Z \leq z) = P(Y \leq X \cap Y \leq z) \\ &= \int_0^z \int_y^\infty f_{X,Y}(x,y) dx dy = \int_0^z \mu_0 e^{-\mu_0 y} \left(\int_y^\infty \lambda e^{-\lambda x} dx \right) dy \\ &= \int_0^z \mu_0 e^{-\mu_0 y} (-e^{-\lambda x}|_y^\infty) dy = \int_0^z \mu_0 e^{-(\mu_0+\lambda)y} dy = \frac{\mu_0}{\mu_0+\lambda} \int_0^z (\mu_0+\lambda) e^{-(\mu_0+\lambda)y} dy \\ &= \frac{\mu_0}{\mu_0+\lambda} (-e^{-(\mu_0+\lambda)y}|_0^z) = \frac{\mu_0}{\mu_0+\lambda} (1 - e^{-(\mu_0+\lambda)z}) = P(\Delta = 0) \cdot P(Z \leq z) \end{aligned}$$

Therefore, $Z \perp \Delta$. □

.....

For $\sum_{i=1}^n \Delta_i \perp \sum_{i=1}^n Z_i$

$$\begin{aligned} E[\hat{\lambda}_n] &= E\left[\frac{D_n}{T_n}\right] = E[D_n]E[(T_n)^{-1}] = \frac{n\lambda}{\mu_0+\lambda} \cdot \frac{\lambda+\mu_0}{n-1} = \frac{n\lambda}{n-1} \\ E[\hat{\lambda}_n^2] &= E[(D_n)^2]E[(T_n)^{-2}] = \frac{n^2\lambda^2+n\lambda\mu_0}{(\mu_0+\lambda)^2} \cdot \frac{(\lambda+\mu_0)^2}{(n-1)(n-2)} = \frac{n^2\lambda^2+n\lambda\mu_0}{(n-1)(n-2)} \\ V[\hat{\lambda}_n] &= E[\hat{\lambda}_n^2] - (E[\hat{\lambda}_n])^2 = \frac{n^2\lambda^2+n\lambda\mu_0}{(n-1)(n-2)} - \left(\frac{n\lambda}{n-1}\right)^2 = \frac{n\lambda(n\lambda+n\mu_0-\mu_0)}{(n-1)^2(n-2)} \quad \square \end{aligned}$$

.....

a. Show that the likelihood function for λ based on $(Z_1, \Delta_1), \dots, (Z_n, \Delta_n)$ is proportional to $L(\lambda) \sim \lambda^{D_n} e^{-\lambda T_n}$, where $D_n \equiv \sum_{i=1}^n \Delta_i$ and $T_n \equiv \sum_{i=1}^n Z_i$.

For $Z \perp \Delta$

$$f(z, \delta) = \begin{cases} f_\Delta(\delta) \cdot f_Z(z) = \frac{\lambda}{\mu_0+\lambda} (\mu_0+\lambda) e^{-(\mu_0+\lambda)z} = \lambda e^{-(\mu_0+\lambda)z}, & X \leq Y \\ f_\Delta(\delta) \cdot f_\Delta(\delta) = \frac{\mu_0}{\mu_0+\lambda} (\mu_0+\lambda) e^{-(\mu_0+\lambda)z} = \mu_0 e^{-(\mu_0+\lambda)z}, & X > Y \end{cases}$$

For μ_0 is known,

$$\begin{aligned} L(z_i, \delta_i, \lambda) &= \prod_{i=1}^n \left(\frac{\lambda}{\mu_0+\lambda} \right)^{\delta_i} \left(\frac{\mu_0}{\mu_0+\lambda} \right)^{1-\delta_i} (\mu_0+\lambda) e^{-(\mu_0+\lambda)z_i} = \left(\frac{\lambda}{\mu_0+\lambda} \right)^{\sum \delta_i} \left(\frac{\mu_0}{\mu_0+\lambda} \right)^{n-\sum \delta_i} (\mu_0+\lambda)^n e^{-(\mu_0+\lambda) \sum z_i} \\ &= \lambda^{\sum_{i=1}^n \delta_i} \mu_0^{n-\sum_{i=1}^n \delta_i} e^{-(\mu_0+\lambda) \sum_{i=1}^n z_i} = \mu_0^{n-D_n} e^{-\mu_0 T_n} \cdot \lambda^{D_n} e^{-\lambda T_n} \\ L(\lambda) &\propto \lambda^{D_n} e^{-\lambda T_n} \quad \blacksquare \end{aligned}$$

b. Find the MLE of $\lambda, \hat{\lambda}_n$.

$$l(\lambda) = \log C + D_n \log \lambda - \lambda T_n$$

$$\frac{\partial}{\partial \lambda} l(\lambda) = \frac{1}{\lambda} D_n - T_n \stackrel{\text{set}}{=} 0$$

$$\frac{\partial^2}{\partial \lambda^2} l(\lambda) = \frac{-D_n}{\lambda^2} < 0, \forall D_n > 0$$

$$\text{Therefore, } \hat{\lambda}_n = \frac{D_n}{T_n} \quad \blacksquare$$

c. Show that $\hat{\lambda}_n$ is a consistent estimator of λ .

$$\text{By Law of Large Number, } \frac{D_n}{n} \xrightarrow{\mathcal{P}} E[\Delta] = \frac{\lambda}{\mu_0 + \lambda}$$

$$\text{By the same way, } \frac{T_n}{n} \xrightarrow{\mathcal{P}} E[Z] = \frac{1}{\lambda + \mu_0};$$

$$y = \frac{1}{x} \text{ is a continuous function at } x = \frac{1}{\lambda + \mu_0}, \lambda > 0, \mu_0 > 0. \text{ Then, } \frac{n}{T_n} \xrightarrow{\mathcal{P}} \lambda + \mu_0$$

$$\text{Then, } \hat{\lambda}_n = \frac{D_n}{n} \frac{n}{T_n} \xrightarrow{\mathcal{P}} \frac{\lambda}{\mu_0 + \lambda} (\lambda + \mu_0) = \lambda, \text{ for all } \lambda,$$

Thus, the sequence of r.v.'s $\hat{\lambda}_n$ is consistent for λ ■

d. Find the joint asymptotic distribution of (D_n, T_n) and use this to find the limiting distribution of $\sqrt{n}(\hat{\lambda}_n - \lambda)$.

$$E[D_n] = \frac{n\lambda}{\mu_0 + \lambda}, \text{Var}[D_n] = \frac{n\lambda\mu_0}{(\mu_0 + \lambda)^2}$$

$$E[T_n^{-1}] = \frac{\lambda + \mu_0}{n-1}, \text{Var}[T_n^{-1}] = \frac{(\lambda + \mu_0)^2}{(n-1)^2(n-2)}$$

$$\text{For } D_n \perp T_n, \text{Cov}(D_n, T_n^{-1}) = 0$$

$$\text{By Multivariate Central Limit Theorem, } \sqrt{n}((D_n, T_n^{-1}) - (\frac{n\lambda}{\mu_0 + \lambda}, \frac{\lambda + \mu_0}{n-1})) \xrightarrow{\mathcal{D}} N(0, \Sigma)$$

$$\text{where } \Sigma = \begin{bmatrix} \text{Var}(D_n) & \text{Cov}(D_n, T_n^{-1}) \\ \text{Cov}(D_n, T_n^{-1}) & \text{Var}(T_n^{-1}) \end{bmatrix} = \begin{bmatrix} \frac{n\lambda\mu_0}{(\mu_0 + \lambda)^2} & 0 \\ 0 & \frac{(\lambda + \mu_0)^2}{(n-1)^2(n-2)} \end{bmatrix}$$

$$J(\frac{n\lambda}{\mu_0 + \lambda}, \frac{\lambda + \mu_0}{n-1}) = \left(\frac{\partial f}{\partial t_1}, \frac{\partial f}{\partial t_2} \right) \Big|_{(t_1, t_2) = (\mu_1, \mu_2)} = \begin{bmatrix} \frac{n-1}{n} \frac{\lambda + \mu_0}{n-1} & \frac{n-1}{n} \frac{n\lambda}{\mu_0 + \lambda} \end{bmatrix} = \begin{bmatrix} \frac{\lambda + \mu_0}{n} & \frac{(n-1)\lambda}{\mu_0 + \lambda} \end{bmatrix}, f(t_1, t_2) = \frac{n-1}{n} t_1 t_2$$

$$J\Sigma J' = \begin{bmatrix} \frac{\lambda + \mu_0}{n} & \frac{(n-1)\lambda}{\mu_0 + \lambda} \end{bmatrix} \begin{bmatrix} \frac{n\lambda\mu_0}{(\mu_0 + \lambda)^2} & 0 \\ 0 & \frac{(\lambda + \mu_0)^2}{(n-1)^2(n-2)} \end{bmatrix} \begin{bmatrix} \frac{\lambda + \mu_0}{n} \\ \frac{(n-1)\lambda}{\mu_0 + \lambda} \end{bmatrix} = \frac{\lambda\mu_0}{n} + \frac{\lambda^2}{n-2}$$

$$\lim_{n \rightarrow \infty} n \text{Var} \left[f(D_n, T_n^{-1}) - f\left(\frac{n\lambda}{\mu_0 + \lambda}, \frac{\lambda + \mu_0}{n-1}\right) \right] = \lim_{n \rightarrow \infty} \left(\lambda\mu_0 + \frac{n\lambda^2}{n-2} \right) = \lambda\mu_0 + \lambda^2$$

$$\text{Therefore, } \sqrt{n}(\hat{\lambda}_n - \lambda) = \sqrt{n}\left(\frac{D_n}{T_n} - \lambda\right) \xrightarrow{\mathcal{D}} N(0, J\Sigma J') = \mathcal{D} \rightarrow N(0, \lambda^2 + \lambda\mu_0) \quad \blacksquare$$

The previous results can confirm this:

$$\lim_{n \rightarrow \infty} E[\hat{\lambda}_n] = \lim_{n \rightarrow \infty} \frac{n\lambda}{n-1} = \lambda$$

$$\lim_{n \rightarrow \infty} n \text{Var}[\hat{\lambda}_n] = \lim_{n \rightarrow \infty} n \frac{n\lambda(n\lambda + n\mu_0 - \mu_0)}{(n-1)^2(n-2)} = \lambda^2 + \lambda\mu_0$$

HW7. (Due 6/03/20)

1. Consider the family of exponential distributions with unknown mean, $f(x; \theta) = \frac{1}{\theta} e^{-x/\theta}$, where $x > 0, \theta > 0$.

(a) Find a central $1 - \alpha$ asymptotic CI for θ based on the sample average.

$$f(\mathbf{x}; \theta) = \prod_{i=1}^n \frac{1}{\theta} e^{-x_i/\theta} = \exp \left[-\frac{1}{\theta} \sum_{i=1}^n x_i - n \log \theta \right]$$

$$T(\mathbf{x}) = \sum_{i=1}^n x_i \text{ is a complete sufficient statistic, } \eta = -\frac{1}{\theta}, \theta = -\frac{1}{\eta}, A(\eta) = -n \log(-\eta), \varepsilon = \mathbb{R}_-, \dot{\varepsilon} = \mathbb{R}_-$$

$$E[\bar{X}] = \frac{1}{n} \dot{A}(\eta) = \frac{-n}{n\eta} = \theta$$

$$\text{Var}[\bar{X}] = \frac{1}{n^2} \ddot{A}(\eta) = \frac{n}{n^2 \eta^2} = \frac{\theta^2}{n}$$

$$\hat{\eta} = \frac{-1}{\bar{X}} \in \varepsilon \text{ is the unique MLE. Then, } \hat{\theta} = \bar{X} \text{ is the unique MLE of } \theta$$

$$\bar{X} \xrightarrow{P} E[X] = \theta, \hat{\theta} = \bar{X} \text{ is consistent for } \theta.$$

$$\text{CRLB} = \frac{1}{I(\theta)} = \frac{1}{E[(\frac{\partial}{\partial \theta} \log p(\theta))^2]} = \frac{1}{E[\frac{(x-\theta)^2}{\theta^4}]} = \frac{1}{\frac{1}{\theta^2}} = \theta^2$$

By Theorem 10.1.12 Asymptotic efficiency of MLEs (Casella and Berger), $\sqrt{n}(\bar{X} - \theta) \xrightarrow{D} N(0, \theta^2)$

In practice, we replace θ^2 by $\hat{\theta}^2 = \bar{X}^2$

$$P(|\sqrt{n}(\bar{X} - \theta)| \leq Z_{\alpha/2}) = 1 - \alpha$$

$$P(\bar{X} - Z_{\alpha} \frac{\bar{X}}{\sqrt{n}} \leq \theta \leq \bar{X} + Z_{\alpha} \frac{\bar{X}}{\sqrt{n}}) = 1 - \alpha$$

95% Asymptotic Confidence Interval for θ is $\bar{X} \pm Z_{\alpha} \frac{\bar{X}}{\sqrt{n}}$ ■

(b) Use a variance-stabilizing transform.

Make a transform with $f(\theta) = \log \theta, f'(\theta) = \frac{1}{\theta}$, then $\sqrt{n}(\log \bar{X} - \log \theta) \xrightarrow{D} N(0, \theta^2 \cdot \frac{1}{\theta^2}) = N(0, 1)$

$$P(|\sqrt{n}(\log \bar{X} - \log \theta)| \leq Z_{\alpha}) = 1 - \alpha$$

$$P(\log \bar{X} - \frac{Z_{\alpha}}{\sqrt{n}} \leq \log \theta \leq \log \bar{X} + \frac{Z_{\alpha}}{\sqrt{n}}) = 1 - \alpha$$

95% Asymptotic Confidence Interval for $\log \theta$ is $\log \bar{X} \pm \frac{Z_{\alpha}}{\sqrt{n}}$

For θ , that is $\bar{X} \exp[\pm \frac{Z_{\alpha}}{\sqrt{n}}]$ ■

2. Simulate 30 independent exponential random variables with mean $\theta = 2$. Compare the confidence intervals in problem1 with a bootstrap percentile confidence interval. Do the same calculation for various values of θ of your choice. Comment on the differences or similarities.

theta: 2 size: 30 alpha:0.05	LCL	Mean	UCL	length
Asym Normal	1.1703	1.8225	2.4746	1.3043
Log Transform	1.2743	1.8225	2.6066	1.3323
Bootstrap Perc	1.2506	1.8225	2.4783	1.2277

The result shows that, based on Delta Method, the Confidence Interval is right shift after log transformation. The reason is that the true exponential distribution is right skewness, not symmetric around mean value. The Asymptotic Confidence Interval is biased.

- Various θ

The exponential distribution has a positive skewness that does not rely upon the value of the parameter θ .

$$E[X^3] = \int_0^{\infty} x^3 f(x) dx = 6\theta^3$$

$$\text{Skewness is } E\left[\frac{(X-\mu)^3}{\sigma^3}\right] = \frac{1}{\theta^3} (E[X^3] - 3\theta E[X^2] + 3\theta^2 E[X] - \theta^3) = 6 - 4 = 2$$

This test proves that various values of θ doesn't change the right-shift trend and distance. The coverage probabilities of Log transform always win but cannot reach 95%.

Table 2: size=30, Rep=10⁴, alpha=0.05

theta	Odds of Asym Normal CI	Odds of Log-Trans CI
0.001	0.9373	0.9485
0.0032	0.9319	0.9466
0.01	0.9313	0.9441
0.0316	0.9366	0.9491
0.1	0.9353	0.9497
0.3162	0.9364	0.9495
1	0.9299	0.9451
3.162	0.934	0.9474
10	0.9359	0.9472
31.62	0.9291	0.9443
100	0.9329	0.9484
316.2	0.9355	0.9481
1000	0.9314	0.9466

- Various sample size

The asymptotic distribution of sample average holds when sample size is large enough.

When setting a small sample size, the coverage probability will decline, the odds of asymptotic normal CI declines faster than the variance stabilizing transform.

When increasing the sample size, the two methods tend to have similar performance. The first 95% coverage probability happens when sample size is 500.

Table 3: theta=2, Rep=10⁴, alpha=0.05

n	Odds of Asym Normal CI	Odds of Log-Trans CI
5	0.8752	0.9369
10	0.9027	0.9404
15	0.9148	0.9426
30	0.9316	0.9475
60	0.9396	0.9456
100	0.9402	0.9464
500	0.9497	0.9506
1000	0.9501	0.9498

- Various α level

The coverage probability changes with α setting. We found that, with higher confidence level, the performance of Log-transform CI is better than asymptotic normal CI.

Table 4: size=30, Rep= 10^4 , theta=2

alpha	Odds of Asym Normal CI	Odds of Log-Trans CI
0.3	0.6958	0.6947
0.2	0.7911	0.7916
0.15	0.8479	0.853
0.1	0.8906	0.8969
0.05	0.9299	0.9446
0.01	0.971	0.9881

Above tests show that variance stabilizing transform has higher coverage probability than asymptotic normal method, especially when we need a higher confidence level and can only afford limited sample size.

- Bootstrap inference

The plots with various sample size show similar results by the three methods.

Comparing the lower bound, bootstrap inference (black line) is close to Log transform (red line).

Comparing the upper bound, bootstrap inference is close to the asymptotic normality method (green line).

In practice, we don't use infinitely large dataset. Is a new confidence interval combined with two methods (asymptotic normality based and log transformed) legitimate?

Another question is, in which situations, this method is appropriate? When residuals have non-constant variance? Or we already know the data follow an asymmetric distribution?

n	LCL.N	LCL.logN	LCL.B	UCL.N	UCL.logN	UCL.B
5	0.1401	0.4723	0.3009	2.129	2.726	1.969
10	0.6781	0.9596	1.033	2.889	3.315	2.632
15	0.9342	1.14	1.036	2.848	3.137	3.126
30	0.8227	0.8957	0.7983	1.74	1.832	1.94
60	1.284	1.335	1.354	2.154	2.214	2.114
100	1.641	1.678	1.652	2.441	2.483	2.46

