Economics 203A Homework #7 Anton Cheremukhin

December 5, 2005

Exercise 1 *Prove that the sample mean of a random sample of size n from $N(\mu, \sigma^2)$ with σ^2 known is an efficient estimator of μ*

Proof. $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $Var\overline{X} = Var[\frac{1}{n}\Sigma X_i] = \frac{1}{n^2}\Sigma Var X_i = \frac{\sigma^2}{n}$, $I_n(\mu) = nE[s(X_i,\mu)^2] = nE\left[\left(\frac{\partial \log f(X_i,\mu)}{\partial\mu}\right)^2\right] = nE\left[\left(\frac{\partial(-\frac{(X_i-\mu)^2}{2\sigma^2} - \log\sqrt{2\pi}\sigma)}{\partial\mu}\right)^2\right] = nE\left[\frac{(X_i-\mu)^2}{\sigma^4}\right] = \frac{n}{\sigma^2}$

Hence $Var\overline{X} = \frac{\sigma^2}{n} = \frac{1}{I_n(\mu)}$ which means it is an efficient estimator of μ. ■

Exercise 2 *Prove that the sample mean of a random sample of size n from $b(1,\theta)$ is an efficient estimator of θ*

Proof. $X_i \overset{iid}{\sim} b(1,\theta)$, $p \equiv \theta$ $\quad Var\overline{X} = Var[\frac{1}{n}\Sigma X_i] = \frac{1}{n^2}\Sigma Var X_i = \frac{p}{n}(1-p)$

$I_n(p) = -nE\left[\frac{\partial^2 \log f(X_i,p)}{\partial p^2}\right] = -nE\left[\frac{\partial}{\partial p}\frac{\partial(\log[p^{x_i}(1-p)^{1-x_i}])}{\partial p}\right] =$

$= -nE\left[\frac{\partial}{\partial p}\left(X_i\frac{\partial \log p}{\partial p} + (1-X_i)\frac{\partial \log(1-p)}{\partial p}\right)\right] = -nE\left[\frac{\partial}{\partial p}\left(X_i\frac{1}{p} + (1-X_i)\frac{1}{1-p}\right)\right]$

$= -nE\left[-X_i\frac{1}{p^2} + (1-X_i)\frac{1}{(1-p)^2}\right] = -n\left[-p\frac{1}{p^2} + (1-p)\frac{1}{(1-p)^2}\right] = \frac{n}{p(1-p)}$

Hence $Var\overline{X} = \frac{p}{n}(1-p) = \frac{1}{I_n(p)}$ which means it is an efficient estimator of θ. ■

Exercise 3 *Let X have a gamma distribution with $\alpha = 4$ and $\beta = \theta$.*
(a) Find the Fisher information $I(\theta)$.
(b) If $X_1, \ldots, X_n$ is a random sample from this distribution, show that the MLE is an efficient estimator of θ.

Proof. $X_i \overset{iid}{\sim} \Gamma(4,\theta)$, $\quad Var\overline{X} = Var[\frac{1}{n}\Sigma X_i] = \frac{1}{n^2}\Sigma Var X_i = \frac{4\theta^2}{n}$

$I_n(\theta) = -nE\left[\frac{\partial^2 \log f(X_i,\theta)}{\partial\theta^2}\right] = -nE\left[\frac{\partial}{\partial\theta}\frac{\partial(\log[\frac{1}{6\theta^4}X_i{}^3\exp(-\frac{X_i}{\theta})])}{\partial\theta}\right] =$

$= -nE\left[\frac{\partial}{\partial\theta}\left(\frac{\partial[-4\log\theta - \frac{X_i}{\theta} + C]}{\partial\theta}]\right)\right] = -nE\left[\frac{\partial}{\partial\theta}\left(-\frac{4}{\theta} + \frac{X_i}{\theta^2}]\right)\right] = -nE\left[\frac{4}{\theta^2} - 2\frac{X_i}{\theta^3}\right] =$

$= -n\left[\frac{4}{\theta^2} - \frac{2}{\theta^3}4\theta\right] = \frac{4n}{\theta^2}$

$\hat{\theta}_{MLE} = \arg\max_\theta[\sum \log[\frac{1}{6\theta^4}X_i{}^3\exp(-\frac{X_i}{\theta})]] = \left\{\theta | \sum(-\frac{4}{\theta} + \frac{X_i}{\theta^2}) = 0\right\} = \frac{\overline{X}}{4}$

Hence $Var\frac{\overline{X}}{4} = \frac{1}{16}\frac{4\theta^2}{n} = \frac{1}{I_n(\theta)}$ which means $\hat{\theta}_{MLE}$ is an efficient estimator of θ. ■

Economics 203A Homework #8 Anton Cheremukhin

December 5, 2005

Exercise 1 *Let* $X_1, \ldots, X_{10}$ *be a random sample from* $N(0, \sigma^2)$. *Find a best critical region of size* $\alpha = 5\%$ *for testing* $H_0 : \sigma^2 = 1$ *against* $H_1 : \sigma^2 = 2$. *Is this a best critical region for testing against* $H_1 : \sigma^2 = 4$? *Against* $H_1 : \sigma^2 = \sigma_1^2 > 1$?

Proof. $\frac{L(\sigma_1;x_1,\ldots,x_n)}{L(\sigma_0;x_1,\ldots,x_n)} = \frac{\left(1/\sqrt{2\pi\sigma_1^2}\right)^n \exp[-\left(\sum_i {x_i}^2\right)/2\sigma_1^2]}{\left(1/\sqrt{2\pi\sigma_0^2}\right)^n \exp[-\left(\sum_i {x_i}^2\right)/2\sigma_0^2]} = \left(\frac{\sigma_0}{\sigma_1}\right)^n \exp\left[\frac{\sigma_1^2-\sigma_0^2}{2\sigma_0^2\sigma_1^2}\sum_i x_i^2\right]$

if $\sigma_1 > \sigma_0$ then $\frac{L(\sigma_1;x_1,\ldots,x_n)}{L(\sigma_0;x_1,\ldots,x_n)} \geq k \Leftrightarrow \sum_i x_i^2 \geq c, \qquad \frac{\sum_{i=1}^n x_i^2}{\sigma^2} \sim \chi^2(n)$

Critical region $C = \{x | \sum_i x_i^2 > \sigma_0^2 \cdot \chi^2_{0.05}(10)\} = \{x | \sum_i x_i^2 > 18.307\}$ is independent of σ_1.

Hence this is the best critical region for any $\sigma_1 > \sigma_0$. ■

Exercise 2 *Let* $X_1, \ldots, X_n$ *be a random sample from a distribution with pdf* $f(x) = \theta x^{\theta-1}$, $0 < x < 1$, *zero elsewhere. Show that the best critical region for testing* $H_0 : \theta = 1$ *against* $H_1 : \theta = 2$ *takes the form*$\{(x_1, \ldots, x_n) : \prod_{i=1}^n x_i \geq c\}$

Proof. $\frac{L(\theta_1;x_1,\ldots,x_n)}{L(\theta_0;x_1,\ldots,x_n)} = \frac{\Pi_{i=1}^n(\theta_1 x^{\theta_1-1})}{\Pi_{i=1}^n(\theta_0 x^{\theta_0-1})} = \left(\frac{\theta_1}{\theta_0}\right)^n \left(\prod_{i=1}^n x_i\right)^{\theta_1-\theta_0}$

if $\theta_1 > \theta_0$ then $\frac{L(\theta_1;x_1,\ldots,x_n)}{L(\theta_0;x_1,\ldots,x_n)} \geq k \Leftrightarrow \prod_{i=1}^n x_i \geq c$ ■

Exercise 3 *Let* $X_1, \ldots, X_n$ *be a random sample from* $N(\theta, 100)$. *Find a best critical region of size* $\alpha = 5\%$ *for testing* $H_0 : \theta = 75$ *against* $H_1 : \theta = 78$.

Proof. $\frac{L(\sigma_1;x_1,\ldots,x_n)}{L(\sigma_0;x_1,\ldots,x_n)} = \exp\frac{1}{2\sigma^2}\left[\sum_i (x_i-\theta_0)^2 - \sum_i (x_i-\theta_1)^2\right] = \exp\frac{\theta_1-\theta_0}{\sigma^2}\sum_i (x_i - \frac{\theta_0+\theta_1}{2})$

if $\theta_1 > \theta_0$ then $\frac{L(\theta_1;x_1,\ldots,x_n)}{L(\theta_0;x_1,\ldots,x_n)} \geq k \Leftrightarrow \overline{x} = \frac{1}{n}\sum_i x_i \geq c, \qquad \sqrt{n}\frac{\overline{x}-\theta}{10} \sim N(0,1)$

Critical region $C = \{x | \overline{x} > 75 + \frac{10}{\sqrt{n}} u_{0.05}\} = \{x | \overline{x} > 75 + \frac{16.45}{\sqrt{n}}\}$ is independent of θ_1.

Hence this is the best critical region for any $\theta_1 > \theta_0 = 75$. ■

Economics 203A Homework #9 Anton Cheremukhin

December 5, 2005

Exercise 1 *Let $X_1, \ldots, X_n$ be i.i.d. $N(0, \theta)$. Show that $\sum_{i=1}^{n} X_i^2$ is a sufficient statistic for θ*

Proof. $f(x, \theta) = \exp[\frac{-\sum_i X_i{}^2}{2\theta^2}] \left(\frac{1}{\sqrt{2\pi\theta^2}}\right)^n = k_1(u(x), \theta) k_2(\theta)$
By factoriztion theorem $u(x) = \sum_{i=1}^{n} X_i^2$ is a sufficient statistic. ■

Exercise 2 *Prove that the sum of the observations of a random sample of size n from a Poisson distribution with mean θ is a sufficient statistic for θ.*

Proof. $X_i \sim Poisson(\theta) \Rightarrow u(x) = \Sigma_{i=1}^{n} x_i \sim \Sigma_{i=1}^{n} Poisson(\theta) = Poisson(\Sigma_{i=1}^{n}\theta) = Poisson(n\theta)$
$\frac{f(x,\theta)}{g(u(x),\theta)} = \frac{\prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!}}{\frac{(n\theta)^{(\Sigma_{i=1}^{n} x_i)} e^{-n\theta}}{(\Sigma_{i=1}^{n} x_i)!}} = \frac{1}{n^{\Sigma_{i=1}^{n} x_i}} \frac{(\Sigma_{i=1}^{n} x_i)!}{\prod_{i=1}^{n} x_i!} = h(x) \quad \Rightarrow$ Sufficient by definition. ■

Exercise 3 *Let $X_1, \ldots, X_n$ be i.i.d. from a distribution with pdf $f(x) = (1-\theta)^x \theta$, $x = 0, 1, 2, \ldots$, zero elsewhere. Show that $\sum_{i=1}^{n} X_i$ is a sufficient statistic for θ*

Proof. $f(x, \theta) = \prod_{i=1}^{n} (1-\theta)^{X_i} \theta = (1-\theta)^{\Sigma_{i=1}^{n} X_i} \theta^n = k_1(u(x), \theta) k_2(\theta)$
By factoriztion theorem $u(x) = \sum_{i=1}^{n} X_i$ is a sufficient statistic. ■

Exercise 4 *Let $X_1, \ldots, X_n$ be i.i.d. from a beta distribution with parameters $\alpha = \theta$ and $\beta = 2$. Show that the product $\prod_{i=1}^{n} X_i$ is a sufficient statistic for θ*

Proof. $f(x, \theta) = \prod_{i=1}^{n} \frac{\Gamma(\theta+2)}{\Gamma(\theta)\Gamma(2)} x_i^{\theta-1} (1-x_i)^{2-1} = \left[(\Pi_{i=1}^{n} x_i)^{\theta-1} - (\Pi_{i=1}^{n} x_i)^{\theta}\right] [\theta(\theta+1)]^n = k_1(u(x), \theta) k_2(\theta$
By factoriztion theorem $u(x) = \Pi_{i=1}^{n} X_i$ is a sufficient statistic. ■

Exercise 5 *In Exercises 1, 2, 3, and 4, show that the MLE of θ is a function of the sufficient statistic for θ.*

Proof. 1) $\hat{\theta}_{ML} = \arg\max_\theta [-\frac{\Sigma_{i=1}^{n} X_i{}^2}{2\theta^2} - n \log \theta] = \{\theta | \frac{\Sigma_{i=1}^{n} X_i{}^2}{\theta^3} = \frac{n}{\theta}\} = \sqrt{\frac{1}{n} \Sigma_{i=1}^{n} X_i{}^2} = \sqrt{\frac{1}{n} u(x)}$
2) $\hat{\theta}_{ML} = \arg\max_\theta [\Sigma_{i=1}^{n} X_i \log \theta - n\theta] = \{\theta | \frac{\Sigma_{i=1}^{n} X_i}{\theta} = n\} = \frac{1}{n} \Sigma_{i=1}^{n} X_i = \frac{1}{n} u(x)$
3) $\hat{\theta}_{ML} = \arg\max_\theta [\Sigma_{i=1}^{n} X_i \log(1-\theta) + n\theta] = \{\theta | \frac{\Sigma_{i=1}^{n} X_i}{1-\theta} = n\} = 1 - \frac{1}{n} \Sigma_{i=1}^{n} X_i = 1 - \frac{1}{n} u(x)$
4) $\hat{\theta}_{ML} = \arg\max_\theta [\log \theta + \log(\theta+1) + \theta \frac{\log \Pi_{i=1}^{n} x_i}{n}] = \{\theta | \frac{1}{\theta} + \frac{1}{\theta+1} = -\frac{\log \Pi_{i=1}^{n} x_i}{n}\}$
Hence, $\frac{1}{\hat{\theta}_{ML}} + \frac{1}{\hat{\theta}_{ML}+1} + \frac{\log u(x)}{n} = 0$ and $\hat{\theta}_{ML}$ is a function of the sufficient statistic. ■

Exercise 6 *Let $X_1, \ldots, X_n$ be i.i.d. from a distribution with pdf $f(x) = \theta \exp(-\theta x)$, $x > 0$, zero elsewhere. Show that $Y = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ. Prove that $\frac{n-1}{Y}$ is the unbiased minimum variance estimator of θ*

Proof. $f(x,\theta) = \prod_{i=1}^{n} \theta \exp(-\theta x_i) = \exp(-\theta \Sigma_{i=1}^{n} x_i + n \log \theta) = \exp[p(\theta)K(X) + q(\theta)]$ - a regular exponential p.d.f. Hence $Y = K(X) = \Sigma_{i=1}^{n} X_i$ is a complete sufficient statistic.

$E[X] = \int_0^\infty x\theta \exp(-\theta x)\, dx = \theta \int_0^\infty x e^{-x\theta}\, dx = \frac{1}{\theta}$

$E[X^2] = \theta \int_0^\infty x^2 e^{-x\theta}\, dx = 2 \int_0^\infty x e^{-x\theta}\, dx = \frac{2}{\theta^2}$ $\quad Var[X] = E[X^2] - E[X]^2 = \frac{1}{\theta^2}$

$E[Y] = \frac{n}{\theta}$ $\quad Var[Y] = \frac{n}{\theta^2}$

Y (as a sum of exponential distributions) has a gamma distribution with

$\{\alpha\beta = \frac{n}{\theta}, \alpha\beta^2 = \frac{n}{\theta^2}\} \Leftrightarrow \{\alpha = n, \beta = \frac{1}{\theta}\}$

$E[\frac{1}{Y}] = \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-2} e^{-x/\beta} dx = \frac{\Gamma(\alpha-1)\beta^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} = \frac{1}{\beta(\alpha-1)} = \frac{\theta}{n-1}$

Hence $\frac{n-1}{Y}$ is an unbiased estimator of θ.

By Lehman-Scheffe Theorem it is a unique unbiased minimum variance estimator of θ. ■