22S:194 Statistical Inference II
Homework Assignments

Luke Tierney

Spring 2003

Assignment 1

Problem 6.3
Problem 6.6

Due Friday, January 31, 2003.

Problem 6.9
Problem 6.10

Due Friday, January 31, 2003.

Problem 6.14
Problem 6.20

Due Friday, January 31, 2003.

Solutions

6.3 The joint density for the data is

$$\begin{aligned} f(x_1,\ldots,x_n|\mu,\sigma) &= \frac{1}{\sigma^n}\exp\{-\sum(x_i-\mu)/\sigma\}\prod 1_{(\mu,\infty)}(x_i) \\ &= \frac{1}{\sigma^n}\exp\left\{-\frac{n}{\sigma}\overline{x}+\frac{n\mu}{\sigma}\right\}1_{(\mu,\infty)}(x_{(1)}) \end{aligned}$$

So, by the factorization criterion, $(\overline{X}, X_{(1)})$ is sufficient for (μ,σ).

6.6 The density of a single observation is

$$\begin{aligned} f(x|\alpha,\beta) &= \frac{1}{\Gamma(\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta} \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\exp\left\{(\alpha-1)\log x-\frac{1}{\beta}x\right\} \end{aligned}$$

This is a two-parameter exponential family, so $(T_1,T_2)=(\sum\log X_i,\sum X_i)$ is sufficient for (α,β). Since $\prod X_i$ is a one-to-one function of $\sum\log X_i$, the pair $(\prod X_i,\sum X_i)$ is also sufficient for (α,β).

6.9 a. Done in class already.

b. $\Theta = \mathbb{R}, \mathcal{X} = \mathbb{R}^n$, and

$$f(x|\theta) = e^{-\sum x_i + n\theta} 1_{(\theta,\infty)}(x_{(1)})$$

Suppose $x_{(1)} = y_{(1)}$. Then

$$f(x|\theta) = \exp\{\sum y_i - \sum x_i\} f(y|\theta)$$

for all θ. So $k(x,y) = \exp\{\sum y_i - \sum x_i\}$ works.
Suppose $x_{(1)} \neq y_{(1)}$. Then for some θ one of $f(x|\theta), f(y|\theta)$ is zero and the other is not. So no $k(x,y)$ exists.
So $T(X) = X_{(1)}$ is minimal sufficient.

c. $\Theta = \mathbb{R}, \mathcal{X} = \mathbb{R}^n$. The support does not depend on θ so we can work with ratios of densities. The ratio of densities for two samples x and y is

$$\begin{aligned} \frac{f(x|\theta)}{f(y|\theta)} &= \frac{e^{-\sum(x_i-\theta)}}{e^{-\sum(y_i-\theta)}} \frac{\prod(1+e^{-(y_i-\theta)})^2}{\prod(1+e^{-(x_i-\theta)})^2} \\ &= \frac{e^{-\sum x_i}}{e^{-\sum y_i}} \frac{\prod(1+e^{-y_i}e^{\theta})^2}{\prod(1+e^{-x_i}e^{\theta})^2} \end{aligned}$$

If the two samples contain identical values, i.e. if they have the same order statistics, then this ratio is constant in θ.
If the ratio is constant in θ then the ratio of the two product terms is constant. These terms are both polnomials of degree $2n$ in e^{θ}. If two polynomials are equal on an open subset of the real line then they are equal on the entire real line. Hence they have the same roots. The roots are $\{-e^{x_i}\}$ and $\{-e^{y_i}\}$ (each of degree 2). If those sets are equal then the sets of sample values $\{x_i\}$ and $\{y_i\}$ are equal, i.e. the two samples must have the same order statistics.
So the order statistics $(X_{(1)}, \ldots, X_{(n)})$ are minimal sufficient.

d. Same idea:

$$\frac{f(x|\theta)}{f(y|\theta)} = \frac{\prod(1+(y_i-\theta)^2)}{\prod(1+(x_i-\theta)^2)}$$

If the two samples have the same order statistics then the ratio is constant. If the ratio is constant for all real θ then two polynomials in θ are equal on the complex plane, and so the roots must be equal. The roots are the complex numbers

$$\theta = x_j \pm i, \theta = y_j \pm i$$

with $i = \sqrt{-1}$. So again the order statistics are minimal sufficient.

e. $\Theta = \mathbb{R}, \mathcal{X} = \mathbb{R}^n$. The support does not depend on θ so we can work with ratios of densities. The ratio of densities for two samples x and y is

$$\frac{f(x|\theta)}{f(y|\theta)} = \exp\{\sum |y_i - \theta| - \sum |x_i - \theta|\}$$

If the order statistics are the same then the ratio is constant. Suppose the order statistics differ. Then there is some open interval I containing no x_i and no y_i such that $\#\{x_i > I\} \neq \#\{y_i > I\}$. The slopes on I of $\sum |x_i - \theta|$ and $\sum |y_i - \theta|$ as functions of θ are

$$n - 2(\#\{x_i > I\}), n - 2(\#\{y_i > I\})$$

So $\sum |y_i - \theta| - \sum |x_i - \theta|$ has slope

$$2(\#\{x_i > I\} - \#\{y_i > I\}) \neq 0$$

and so the ratio is not constant on I.

So again the order statistic is minimal sufficient.

6.10 To thos that the minimal sufficient statistic is not complete we need to fins a function g that is not identically zero but has expected value zero for all θ. Now

$$E[X_{(1)}] = \theta + \frac{1}{n+1}$$
$$E[X_{(n)}] = \theta + \frac{n}{n+1}$$

So $g(X_{(1)}, X_{(n)}) = X_{(n)} - X_{(1)} - \frac{n-1}{n+1}$ has expected value zero for all θ but is not identically zero for $n > 1$.

6.14 $X_1, \ldots, X_n$ are *i.i.d.* from $f(x-\theta)$. This means $Z_i = X_i - \theta$ are *i.i.d.* from $f(z)$. Now

$$\widetilde{X} = \widetilde{Z} + \theta$$
$$\overline{X} = \overline{Z} + \theta$$

So $\widetilde{X} - \overline{X} = \widetilde{Z} - \overline{Z}$ is ancillary.

6.20 a. The joint density of the data is

$$f(x_1, \ldots, x_n|\theta) = 2^n \left(\prod x_i\right) 1_{(0,\theta)}(x_{(n)}) \frac{1}{\theta^{2n}}$$

$T = X_{(n)}$ is sufficient (and minimal sufficient). $X_{(n)}$ has density

$$f_T(t|\theta) = 2nt^{2n-1}\frac{1}{\theta^{2n}} 1_{(0,\theta)}(t)$$

Thus
$$0 = \int_0^\theta g(t)\frac{2n}{\theta^{2n}}t^{2n-1}dt$$
for all $\theta > 0$ means
$$0 = \int_0^\theta g(t)t^{2n-1}dt$$
for almost all $\theta > 0$, and this in turn implies $g(t)t^{2n-1} = 0$ and hence $g(t) = 0$ for all $t > 0$. So $T = X_{(n)}$ is complete.

b. Exponential family, $T(X) = \sum \log(1 + X_i)$,
$$\{w(\theta) : \theta \in \Theta\} = (1, \infty)$$
which is an open interval.

c. Exponential family, $T(X) = \sum X_i$,
$$\{w(\theta) : \theta \in \Theta\} = \{\log\theta : \theta > 1\} = (0, \infty)$$
which is an open interval.

d. Exponential family, $T(X) = \sum e^{-X_i}$,
$$\{w(\theta) : \theta \in \Theta\} = \{-e^\theta : \theta \in \mathbb{R}\} = (-\infty, 0)$$
which is an open interval.

e. Exponential family, $T(X) = \sum X_i$,
$$\{w(\theta) : \theta \in \Theta\} = \{\log\theta - \log(1-\theta) : 0 \le \theta \le 1\} = [-\infty, \infty]$$
which contains an open interval.

Assignment 2

Problem 7.6
Problem 7.11

Due Friday,February 7 , 2003.

Problem 7.13
Problem 7.14

Due Friday, February 7, 2003.

Solutions

7.6 The joint PDF of the data can be written as

$$f(x|\theta) = \theta^n \prod x_i^{-2} 1_{[\theta,\infty)}(x_{(1)})$$

a. $X_{(1)}$ is sufficient.

b. The likelihood increases up to $x_{(1)}$ and then is zero. So the MLE is $\widehat{\theta} = X_{(1)}$.

c. The expected value of a single observation is

$$E_\theta[X] = \int_\theta^\infty x\theta\frac{1}{x^2}dx = \theta\int_\theta^\infty \frac{1}{x}dx = \infty$$

So the (usual) method of moments estimator does not exist.

7.11 a. The likelihood and log likelihood are

$$L(\theta|x) = \theta^n \left(\prod x_i\right)^{\theta-1}$$
$$\log L(\theta|x) = n\log\theta + (\theta-1)\sum \log x_i$$

The derivative of the log likelihood and its unique root are

$$\frac{d}{d\theta}L(\theta|x) = \frac{n}{\theta} + \sum \log x_i$$
$$\widehat{\theta} = -\frac{n}{\sum \log x_i}$$

Since $\log L(\theta|x) \to -\infty$ as $\theta \to 0$ or $\theta \to \infty$ and the likelihood is differentiable on the parameter space this root is a global maximum.

Now $-\log X_i \sim \text{Exponential}(1/\theta) = \text{Gamma}(1, 1/\theta)$. So $-\sum \log X_i \sim \text{Gamma}(n, 1/\theta)$. So

$$\begin{aligned} E\left[-\frac{n}{\sum \log X_i}\right] &= n\int_0^\infty \frac{\theta^n}{x\Gamma(n)} x^{n-1} e^{-\theta x} dx \\ &= n\frac{\Gamma(n-1)}{\Gamma(n)}\theta = \frac{n}{n-1}\theta \end{aligned}$$

and

$$E\left[\left(\frac{n}{\sum \log X_i}\right)^2\right] = n^2 \frac{\Gamma(n-2)}{\Gamma(n)}\theta^2 = \frac{n^2}{(n-1)(n-2)}\theta^2$$

So

$$\text{Var}(\widehat{\theta}) = \theta^2 \frac{n^2}{n-1}\left(\frac{1}{n-2} - \frac{1}{n-1}\right) = \frac{\theta^2 n^2}{(n-1)^2(n-2)} \sim \frac{\theta^2}{n} \to 0$$

as $n \to \infty$.

b. The mean of a single observation is

$$E[X] = \int_0^1 \theta x^\theta dx = \frac{\theta}{\theta+1}$$

So

$$\overline{X} = \frac{\theta}{\theta+1}$$

is the method of moments equation, and

$$\begin{aligned} \widetilde{\theta}\overline{X} + \overline{X} &= \widetilde{\theta} && \text{or} \\ \widetilde{\theta}(\overline{X} - 1) &= -\overline{X} && \text{or} \\ \widetilde{\theta} &= \frac{\overline{X}}{1 - \overline{X}} \end{aligned}$$

We could use the delta method to find a normal approximation to the distribution of $\widehat{\theta}$. The variance of the approximate disrtibtion is larger than the variance of the MLE.

7.13 The likelihood is

$$L(\theta|x) = \frac{1}{2}\exp\{-\sum |x_i - \theta|\}$$

We know that the sample median $\widetilde{X}$ minimizes $\sum |x_i - \theta|$, so $\widehat{\theta} = \widetilde{X}$. The minimizer is unique for odd n. For even n any value between the two middle order statistics is a minimizer.

7.14 We need the joint "density" of W, Z:

$$\begin{aligned} P(W = 1, Z \in [z, z+h)) &= P(X \in [z, z+h), Y \geq z+h) + o(h) \\ &= h\frac{1}{\lambda}e^{-z/\lambda}e^{-z/\mu} + o(h) \\ &= h\frac{1}{\lambda}e^{-z\left(\frac{1}{\lambda}+\frac{1}{\mu}\right)} + o(h) \end{aligned}$$

and, similarly,

$$P(W = 0, Z \in [z, z+h)) = h\frac{1}{\mu}e^{-z\left(\frac{1}{\lambda}+\frac{1}{\mu}\right)} + o(h)$$

So

$$f(w, z) = \lim_{h \downarrow 0} \frac{1}{h} P(W = w, Z \in [z, z+h)) = \frac{1}{\lambda^w \mu^{1-w}} e^{-z\left(\frac{1}{\lambda}+\frac{1}{\mu}\right)}$$

and therefore

$$f(w_1, \ldots, w_n, z_1, \ldots, z_n | \lambda, \mu) = \frac{1}{\lambda^{\sum w_i} \mu^{n - \sum w_i}} e^{-\sum z_i \left(\frac{1}{\lambda}+\frac{1}{\mu}\right)}$$

Since this factors,

$$f(w_1, \ldots, w_n, z_1, \ldots, z_n | \lambda, \mu) = \frac{1}{\lambda^{\sum w_i}} e^{-\sum z_i/\lambda} \frac{1}{\mu^{n - \sum w_i}} e^{-\sum z_i/\mu}$$

it is maximized by maximizing each term separately, which produces

$$\begin{aligned} \widehat{\lambda} &= \frac{\sum z_i}{\sum w_i} \\ \widehat{\mu} &= \frac{\sum z_i}{n - \sum w_i} \end{aligned}$$

In words, if the X_i represent failure times, then

$$\widehat{\lambda} = \frac{\text{total time on test}}{\text{number of observed failures}}$$

Assignment 3

Problem 7.22
Problem 7.23

Due Friday, February 14, 2003.

Problem 7.33

Due Friday, February 14, 2003.

Problem 7.38
Problem 7.39

Due Friday, February 14, 2003.

Solutions

7.22 We have

$$\overline{X}|\theta \sim N(\theta, \sigma^2/n)$$
$$\theta \sim N(\mu, \tau^2)$$

a. The joint density of $\overline{X}, \theta$ is

$$f(\overline{x}, \theta) = f(\overline{x}|\theta) f(\theta) \propto \exp\left\{-\frac{n}{2\sigma^2}(\overline{x}-\theta)^2 - \frac{1}{2\tau^2}(\theta-\mu)^2\right\}$$

This is a joint normal distribution. The means and variances are

$$E[\theta] = \mu \qquad E[\overline{X}] = E[\theta] = \mu$$
$$\text{Var}(\theta) = \tau^2 \qquad \text{Var}(\overline{X}) = \text{Var}(\theta) + \frac{\sigma^2}{n} = \tau^2 + \frac{\sigma^2}{n}$$

The covariance and correlation are

$$\text{Cov}(\theta, \overline{X}) = E[\overline{X}\theta] - \mu^2 = E[\theta^2] - \mu^2 = \mu^2 + \tau^2 - \mu^2 = \tau^2$$
$$\rho = \frac{\tau^2}{\sqrt{(\tau^2 + \sigma^2/n)\sigma^2/n}}$$

b. This means that the marginal distribution of $\overline{X}$ is $N(\mu, \tau^2 + \sigma^2/n)$.

c. The posterior distribuiton of θ is

$$\begin{aligned} f(\theta|\overline{x}) &\propto \exp\left\{-\frac{n}{2\sigma^2}(\overline{X}-\theta)^2 - \frac{1}{2\tau^2}(\theta-\mu)^2\right\} \\ &\propto \exp\left\{\left(\frac{n}{\sigma^2}\overline{x}+\frac{1}{\tau^2}\mu\right)\theta - \frac{1}{2}\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)\theta^2\right\} \end{aligned}$$

This is a normal distribution with mean and variance

$$\begin{aligned} \mathrm{Var}(\theta|\overline{X}) &= \left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1} = \frac{\tau^2\sigma^2/n}{\tau^2+\sigma^2/n} \\ E[\theta|\overline{X}] &= \mathrm{Var}(\theta|\overline{X})\left(\frac{n}{\sigma^2}\overline{X}+\frac{1}{\tau^2}\mu\right) = \frac{\tau^2}{\tau^2+\sigma^2/n}\overline{X} + \frac{\sigma^2/n}{\tau^2+\sigma^2/n}\mu \end{aligned}$$

7.23 We have $S^2|\sigma^2 \sim \mathrm{Gamma}((n-1)/2, 2\sigma^2/(n-1))$ and

$$f(\sigma^2) = \frac{1}{\Gamma(\alpha)\beta^\alpha}\frac{1}{(\sigma^2)^{\alpha+1}}e^{-1/(\beta\sigma^2)}$$

The posterior distribution $\sigma^2|S^2$ is therefore

$$\begin{aligned} f(\sigma^2|s^2) &\propto \frac{1}{(\sigma^2)^{(n-1)/2}}e^{-s^2(n-1)/(2\sigma^2)}\frac{1}{(\sigma^2)^{\alpha+1}}e^{-1/(\beta\sigma^2)} \\ &= \mathrm{IG}(\alpha+(n-1)/2, (1/\beta+(n-1)s^2/2)^{-1}) \end{aligned}$$

If $Y \sim \mathrm{IG}(a,b)$, then $V = 1/Y \sim \mathrm{Gamma}(a,b)$. So

$$\begin{aligned} E[Y] = E[1/V] &= \int_0^\infty \frac{1}{v}\frac{1}{\Gamma(a)b^a}v^{a-1}e^{-v/b}dv \\ &= \frac{1}{b\Gamma(a)}\int_0^\infty z^{a-2}e^{-z}dz \\ &= \frac{\Gamma(a-1)}{b\Gamma(a)} = \frac{1}{b(a-1)} \end{aligned}$$

So the posterior mean of σ^2 is

$$E[\sigma^2|S^2] = \frac{1/\beta+(n-1)S^2/2}{\alpha+(n-3)/2}$$

7.33 From Example 7.3.5 the MSE of $\widehat{p}_B$ is

$$\begin{aligned}
E[(\widehat{p}_B - p)^2] &= \frac{np(1-p)}{(\alpha+\beta+n)^2} + \left(\frac{np+\alpha}{\alpha+\beta+n} - p\right)^2 \\
&= \frac{np(1-p)}{(\sqrt{n/4}+\sqrt{n/4}+n)^2} + \left(\frac{np+\sqrt{n/4}}{\sqrt{n/4}+\sqrt{n/4}+n} - p\right)^2 \\
&= \frac{np(1-p) + (np+\sqrt{n/4} - p(\sqrt{n}+n))^2}{(\sqrt{n}+n)^2} \\
&= \frac{np(1-p) + (\sqrt{n/4} - p\sqrt{n})^2}{(\sqrt{n}+n)^2} \\
&= \frac{n}{(\sqrt{n}+n)^2}\left(p(1-p) + (1/2-p)^2\right) \\
&= \frac{n}{(\sqrt{n}+n)^2}\left(p - p^2 + 1/4 + p^2 - p)^2\right) \\
&= \frac{n/4}{(\sqrt{n}+n)^2}
\end{aligned}$$

which is constant in p.

7.38 a. The population density is

$$\theta x^{\theta-1} = \theta x^{-1} e^{\theta \log x}$$

So $T(X) = \frac{1}{n}\log X_i$ is efficient for $\tau(\theta) = E_\theta[\log X_1]$.

$$\begin{aligned}
\tau(\theta) &= \int_0^1 \log x \theta x^{\theta-1} dx \\
&= -\int_0^\infty y\theta e^{-\theta y} dy = 1/\theta
\end{aligned}$$

b. The population density is

$$\frac{\log\theta}{\theta-1}\theta^x = \frac{\log\theta}{\theta-1} e^{x\log\theta}$$

So

$$\begin{aligned}
\sum \log f(x_i|\theta) &= n(\log\log\theta - \log(\theta-1)) + \sum x_i \log\theta \\
\sum \frac{\partial}{\partial\theta}\log f(x_i|\theta) &= n\left(\frac{1}{\theta\log\theta} - \frac{1}{\theta-1} + \frac{\overline{x}}{\theta}\right) \\
&= \frac{n}{\theta}\left(\frac{1}{\log\theta} - \frac{\theta}{\theta-1} + \overline{x}\right) = \frac{n}{\theta}\left(\overline{x} - \left(\frac{\theta}{\theta-1} - \frac{1}{\log\theta}\right)\right)
\end{aligned}$$

So $\overline{X}$ is efficient for $\tau(\theta) = \frac{\theta}{\theta-1} - \frac{1}{\log\theta}$

7.39 Done in class.

Assignment 4

Problem 7.44
Problem 7.48

Due Friday, February 21, 2003.

Problem 7.62
Problem 7.63
Problem 7.64

Due Friday, February 21, 2003.

Solutions

7.44 $X_1, \ldots, X_n$ are *i.i.d.* $N(\theta, 1)$. $W = \overline{X} - 1/n$ has

$$E[W] = \theta^2 + \frac{1}{n} - \frac{1}{n} = \theta^2$$

Since $\overline{X}$ is sufficient and complete, W is the UMVUE of θ. The CRLB is

$$\frac{(2\theta)^2}{I_n(\theta)} = \frac{(2\theta)^2}{n} = \frac{4\theta^2}{n}$$

Now

$$\begin{aligned}
E[\overline{X}^2] &= \theta^2 + \frac{1}{n} \\
E[\overline{X}^4] &= E[(\theta + X/\sqrt{n})^4] \\
&= \theta^4 + 4\theta^3 \frac{1}{\sqrt{n}} E[Z] + 6\theta^2 \frac{1}{n} E[Z^2] + 4\theta \frac{1}{n^{3/2}} E[Z^3] + \frac{1}{n^2} E[Z^4] \\
&= \theta^4 + 6\theta^2 \frac{1}{n} + \frac{3}{n^2}
\end{aligned}$$

So

$$\begin{aligned}
\mathrm{Var}(W) = \mathrm{Var}(\overline{X}^2) &= \theta^4 + 6\theta^2 \frac{1}{n} + \frac{3}{n^2} - \theta^4 - \frac{1}{n^2} - \frac{2}{n}\theta^2 \\
&= \frac{4}{n}\theta^2 + \frac{2}{n^2} > \frac{4}{n}\theta^2
\end{aligned}$$

7.48 a. The MLE $\widehat{p} = \frac{1}{n}\sum X_i$ has variance $\text{Var}(\widehat{p}) = \frac{p(1-p)}{n}$. The information is

$$\begin{aligned} I_n(p) &= -E\left[\frac{\partial^2}{\partial\theta^2}\left(\sum X_i \log p + \left(n - \sum X_i\right)\log(1-p)\right)\right] \\ &= -E\left[\frac{\partial}{\partial\theta}\left(\frac{\sum X_i}{p} - \frac{n-\sum X_i}{1-p}\right)\right] \\ &= -\left(-\frac{np}{p^2} - \frac{n-np}{(1-p)^2}\right) = \frac{n}{p} + \frac{n}{1-p} = \frac{n}{p(1-p)} \end{aligned}$$

So the CRLB is $\frac{p(1-p)}{n}$.

b. $E[X_1X_2X_3X_4] = p^4$. $\sum X_i$ is sufficient and complete.

$$\begin{aligned} E[W|\sum X_i = t] &= P(X_1 = X_2 = X_3 = X_4 = 1|\sum X_i = t) \\ &= \begin{cases} 0 & t < 4 \\ \frac{P(X_1=X_2=X_3=X_4=1, \sum_5^n X_i = t-4)}{P(\sum_1^n X_i = t)} & t \geq 4 \end{cases} \\ &= \begin{cases} 0 & t < 4 \\ \frac{p^4\binom{n-4}{t-4}p^{t-4}(1-p)^{n-t}}{\binom{n}{t}p^t(1-p)n-t} & t \geq 4 \end{cases} \\ &= \frac{t(t-1)(t-2)(t-3)}{n(n-1)(n-2)(n-3)} \end{aligned}$$

So the UMVUE is (for $n \geq 4$)

$$\frac{\widehat{p}(\widehat{p}-1/n)(\widehat{p}-2/n)(\widehat{p}-3/n)}{(1-1/n)(1-2/n)(1-3/n)}$$

No unbiased estimator eists for $n < 4$.

7.62 a.

$$\begin{aligned} R(\theta, a\overline{X}+b) &= E_\theta[(a\overline{X}+b-\theta)^2] \\ &= a^2\text{Var}(\overline{X}) + (a\theta+b-\theta)^2 \\ &= a^2\frac{\sigma^2}{n} + (b-(1-a)\theta)^2 \end{aligned}$$

b. For $\eta = \frac{\sigma^2}{n\tau^2+\sigma^2}$,

$$\delta_\pi = E[\theta|X] = (1-\eta)\overline{X} + \eta\mu$$

So

$$\begin{aligned} R(\theta, \delta_\pi) &= (1-\eta)^2\frac{\sigma^2}{n} + (\eta\mu - \eta\theta)^2 = (1-\eta)^2\frac{\sigma^2}{n} + \eta^2(\mu-\theta)^2 \\ &= \eta(1-\eta)\tau^2 + \eta^2(\mu-\theta)^2 \end{aligned}$$

c.

$$\begin{aligned}
B(\pi,\delta_\pi) &= E[E[(\theta-\delta_\pi)^2|X]]\\
&= E[E[(\theta-E[\theta|X])^2|X]]\\
&= E[\mathrm{Var}(\theta|X)] = E\left[\frac{\sigma^2\tau^2}{\sigma^2+n\tau^2}\right] = \frac{\sigma^2\tau^2}{\sigma^2+n\tau^2} = \eta\tau^2
\end{aligned}$$

7.63 From the previous problem, when the prior mean is zero the risk of the Bayes rule is

$$R(\theta,\delta^\pi) = \frac{\tau^4+\theta^2}{(1+\tau^2)^2}$$

So for $\tau^2 = 1$

$$R(\theta,\delta^\pi) = \frac{1}{4} + \frac{1}{4}\theta^2$$

and for $\tau^2 = 10$

$$R(\theta,\delta^\pi) = \frac{100}{121} + \frac{1}{121}\theta^2$$

With a smaller τ^2 the risk is lower near the prior mean and higher far from the prior mean.

7.64 For any $a = (a_1,\dots,a_n)$

$$E[\sum L(\theta_i,a_i)|X=x] = \sum E[L(\theta_i,a_i)|X=x]$$

The independence assumptions imply that (θ_1, X_i) is independent of $\{X_j : j \neq i\}$ and therefore.

$$E[L(\theta_i,a_i)|X=x] = E[L(\theta_i,a_i)|X_i=x_i]$$

for each i. Since δ^{π_i} is a Bayes rule for estimating θ_i with loss $L(\theta_i,a_i)$ we have

$$E[L(\theta_i,a_i)|X_i=x_i] \geq E[L(\theta_i,\delta^{\pi_i}(X_i))|X_i=x_i] = E[L(\theta_i,\delta^{\pi_i}(X_i))|X=x]$$

with the final equality again following from the independence assumptions. So

$$\begin{aligned}
\sum E[L(\theta_i,a_i)|X_i=x_i] &\geq \sum E[L(\theta_i,\delta^{\pi_i}(X_i))|X=x]\\
&= E[\sum L(\theta_i,\delta^{\pi_i}(X_i))|X=x]
\end{aligned}$$

and therefore

$$E[\sum L(\theta_i,a_i)|X=x] \geq E[\sum L(\theta_i,a_i)|X_i=x_i]$$

for all a, which implies that $\delta^\pi(X) = (\delta^{\pi_1}(X_1),\dots,\delta^{\pi_n}(X_n))$ is a Bayes rule for estimating $\theta = (\theta_1,\dots,\theta_n)$ with loss $\sum L(\theta_i,a_i)$ and prior $\prod \pi_i(\theta_i)$.

Assignment 5

Problem 8.5
Problem 8.6

Due Friday, February 28, 2003.

Solutions

8.5 a. The likelihood can be written as

$$L(\theta,\nu) = \frac{\theta^n \nu^{n\theta}}{\prod x_i^{\theta+1}} 1_{[\nu,\infty)}(x_{(1)})$$

For fixed θ, this increases in ν for $\nu \le x_{(1)}$ and is then zero. So $\widehat{\nu} = x_{(1)}$, and

$$\begin{aligned} L^*(\theta) = \max_\nu L(\theta,\nu) &= \theta^n \prod \left(\frac{x_{(1)}}{x_i}\right)^\theta \frac{1}{\prod x_i} \\ &\propto \theta^n e^{-\theta T} \end{aligned}$$

So $\widehat{\theta} = n/T$.

b. The likelihood ratio criterion is

$$\Lambda(x) = \frac{L^*(1)}{L^*(\widehat{\theta})} = \frac{e^{-T}}{\left(\frac{n}{T}\right)^n e^{-n}} = \text{const} \times T^n e^{-T}$$

This is a unimodal function of T; it increases from zero to a maximum at $T = n$ and then decreases back to zero. Therefore for any $c > 0$

$$R = \{x : \Lambda(x) < c\} = \{x : T < c_1 \text{ or } T > c_2\}$$

c. The conditional density of $X_2, \ldots, X_n$, given $X_1 = x$ and $X_i \ge X_1$, is

$$\begin{aligned} f(x_2,\ldots,x_n|x_1, x_i \ge x_1) &= \frac{f(x_1)\cdots f(x_n)}{f(x_1)P(X_2 > X_1|X_1 = x_1)\cdots P(X_n > X_1|X_1 = x_1)} \\ &= \frac{f(x_2)\cdots f(x_n)}{P(X_2 > x_1)\cdots P(X_n > x_1)} \end{aligned}$$

and

$$P(X_i > y) = \int_y^\infty \frac{\theta\nu^\theta}{x^{\theta+1}} dx = \frac{\nu^\theta}{y^\theta}$$

So

$$f(x_2, \ldots, x_n | x_1, x_i \geq x_1) = \theta^{n-1} \prod_{i=2}^{n} \frac{x_1^\theta}{x_i^{\theta+1}} 1_{\{x_i > x_1\}}$$

Let $Y_1 = X_i / x_1$, $i = 2, \ldots, n$. Then

$$\begin{aligned} f_Y(y_2, \ldots, y_n | x_1, x_i > x_1) &= x_1^{n-1} f(y_2 x_1, \ldots, y_n x_1 | x_1, x_i > x_1) \\ &= \frac{\theta^{n-1}}{y_2^{\theta+1}, \ldots, y_n^{\theta+1}} \end{aligned}$$

i.e. $Y_2, \ldots, Y_n$ are *i.i.d.* with density $\theta / y^{\theta+1}$, and $T = \log Y_2 + \cdots + \log Y_n$. If $Z = \log Y$, then

$$f_Z(z) = f_Y(y) \frac{dy}{dz} = \frac{\theta}{e^{(\theta+1)z}} e^z = \theta e^{-\theta z}$$

and thus $T | \{X_1 = x_1, X_i > X_1\} \sim \text{Gamma}(n-1, 1/\theta)$. By symmetry, this means that $T | X_{(1)} \sim \text{Gamma}(n-1, 1/\theta)$, which is indepentent of $X_{(1)}$, so T has this distribution unconditionaly as well.

For $\theta = 1$,

$$\begin{aligned} T &\sim \text{Gamma}(n-1, 1) \\ 2T &\sim \text{Gamma}(n-1, 2) = \chi^2_{n-1} \end{aligned}$$

8.6 a. The likelihood ratio criterion is

$$\begin{aligned} \Lambda &= \frac{\left(\frac{n+m}{\sum X_i + \sum Y_i}\right)^{n+m} e^{-n-m}}{\left(\frac{n}{\sum X_i}\right)^n e^{-n} \left(\frac{n}{\sum Y_i}\right)^m e^{-m}} \\ &= \frac{(n+m)^{n+m}}{n^n m^m} \frac{(\sum X_i)^n (\sum Y_i)^m}{(\sum X_i + \sum Y_i)^{n+m}} \end{aligned}$$

The test rejects if this is small.

b. The likelihood ratio criterion is of the form $\Lambda = \text{const} \times T^n (1-T)^m$. So the test rejects if T is too small or too large.

c. Under H_0, $T \sim \text{Beta}(n, m)$.

Assignment 6

Problem 8.14
Problem 8.17

Due Friday, March 7, 2003.

Problem 8.15
Problem 8.25

Due Friday, March 7, 2003.

Problem 8.28
Problem 8.33

Due Friday, March 7, 2003.

Solutions

8.14 Use $R = \{x : \sum x_i > c\}$. $\alpha = 0.01$ means

$$0.01 = P(\sum X_i > c | p = 0.49) \approx P\left(Z > \frac{c - 0.49}{\sqrt{0.49 \times 0.51}}\sqrt{n}\right)$$

So

$$\frac{c - 0.49}{\sqrt{0.49 \times 0.51}}\sqrt{n} = 2.326$$

$\beta = 0.99$ implies

$$0.99 = P(\sum X_i > c | p = 0.51) \approx P\left(Z > \frac{c - 0.51}{\sqrt{0.49 \times 0.51}}\sqrt{n}\right)$$

So

$$\frac{c - 0.51}{\sqrt{0.49 \times 0.51}}\sqrt{n} = -2.326$$

So

$$\begin{aligned} c - 0.49 &= 2.326\sqrt{0.49 \times 0.51}\frac{1}{\sqrt{n}} \\ c - 0.51 &= -2.326\sqrt{0.49 \times 0.51}\frac{1}{\sqrt{n}} \end{aligned}$$

or

$$\begin{aligned} \sqrt{n} \times 0.002 &= 2 \times 2.326\sqrt{0.49 \times 0.51} \\ n &= (100)^2 \times (2.326)^2 \times 0.49 \times 0.51 \\ &= 13521 \end{aligned}$$

8.17 For $X_1, \ldots, X_n$ *i.i.d.* Beta$(\mu, 1)$

$$\begin{aligned} L(\mu|x) &= \mu^n \prod x_i^{\mu-1} \\ &= \mu^n e^{\mu \sum \log x_i} e^{-\sum \log x_i} \end{aligned}$$

So

$$\begin{aligned} \widehat{\mu} &= -\frac{n}{\sum \log x_i} \\ L(\widehat{\mu}|x) &= \left(-\frac{n}{\sum \log x_i}\right)^n \exp\{-n - \sum \log x_i\} \end{aligned}$$

and

$$\begin{aligned} \Lambda(x) &= \frac{\left(-\frac{n+m}{\sum \log x_i + \sum \log y_i}\right)^{n+m} \exp\{-n - m - \sum \log x_i - \sum \log y_i\}}{\left(-\frac{n}{\sum \log x_i}\right)^n \exp\{-n - \sum \log x_i\} \left(-\frac{m}{\sum \log y_i}\right)^m \exp\{-m - \sum \log y_i\}} \\ &= \frac{(n+m)^{n+m}}{n^n m^m} T^n (1-T)^m \end{aligned}$$

So

$$\{\Lambda < c\} = \{T < c_1 \text{ or } T > c_2\}$$

for suitable c_1, c_2.

Under H_0, $-\log X_i, -\log Y_i$ are *i.i.d.* exponential, so $T \sim \text{Beta}(n, m)$.

To find c_1 and c_2 either cheat and use equal tail probabilities (the right thing to do by symmetry if $n = m$), or solve numerically.

8.15

$$\begin{aligned} L(\sigma^2|x) &= \frac{1}{(\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum x_i^2\right\} \\ \frac{L(\sigma_1^2)}{L(\sigma_0^2)} &= \left(\frac{\sigma_0}{\sigma_1}\right)^{n/2} \exp\left\{\frac{1}{2} \sum x_i^2 \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\right\} \end{aligned}$$

If $\sigma_1 > \sigma_2$ then this is increasing in $\sum x_i^2$. So

$$L(\sigma_1^2)/L(\sigma_0^2) > k$$

for some k if and only if $\sum X_i^2 > c$ for some c.

Under $H_0 : \sigma = \sigma_0$, $\sum X_i^2/\sigma_0^2 \sim \chi_n^2$, so

$$c = \sigma_0^2 \chi_{n,\alpha}^2$$

8.25 a. For $\theta_2 > \theta_1$

$$\frac{g(x|\theta_2)}{g(x|\theta_1)} = \frac{\exp\left\{-\frac{(x-\theta_2)^2}{2\sigma^2}\right\}}{\exp\left\{-\frac{(x-\theta_1)^2}{2\sigma^2}\right\}} = \text{const} \times \exp\{x(\theta_2 - \theta_1)/\sigma^2\}$$

This in increasing in x since $\theta_2 - \theta_1 > 0$.

b. For $\theta_2 > \theta_1$

$$\frac{g(x|\theta_2)}{g(x|\theta_1)} = \frac{\theta_2^x e^{-\theta_2}/x!}{\theta_1^x e^{-\theta_1}/x!} = \text{const} \times \left(\frac{\theta_2}{\theta_1}\right)^x$$

which is increasing x since $\theta_2/\theta_2 > 1$.

c. For $\theta_2 > \theta_1$

$$\frac{g(x|\theta_2)}{g(x|\theta_1)} = \frac{\binom{n}{x}\theta_2^s(1-\theta_2)^{n-x}}{\binom{n}{x}\theta_1^s(1-\theta_1)^{n-x}} = \text{const} \times \left(\frac{\theta_2/(1-\theta_2)}{\theta_1/(1-\theta_1)}\right)^x$$

This is increasing in x since $\theta/(1-\theta)$ is increasing in θ.

8.28 a.

$$\frac{f(x|\theta_2)}{f(x|\theta_1)} = e^{\theta_1-\theta_2}\frac{(1+e^{x-\theta_1})^2}{(1+e^{x-\theta_2})^2} = \text{const} \times \left(\frac{e^{\theta_1}+e^x}{e^{\theta_2}+e^x}\right)^2$$

Let

$$\begin{aligned} g(y) &= \frac{A+y}{B+y} \\ g'(y) &= \frac{B+y-A-y}{(B+y)^2} = \frac{B-A}{(B+y)^2} \end{aligned}$$

Then $g'(y) \geq 0$ if $B \geq A$. So we have MLR in x.

b. Since the ratio is increasing in x, the most powerful test is of the form $R = \{x > c\}$. Now

$$F_\theta(x) = 1 - \frac{1}{1+e^{x-\theta}}$$

So for $H_0 : \theta = 0$ and $\alpha = 0.2 = 1/(1+e^c)$, so

$$\begin{aligned} 1 + e^c &= 5 \\ e^c &= 4 \\ c &= \log 4 = 1.386 \end{aligned}$$

The power is

$$\beta(1) = \frac{1}{1+e^{\log(4)-1}} = \frac{1}{1+4/e} = 0.405$$

c. Since we have MLR, the test is UMP. This is true for any θ_0. This *only* works for $n = 1$; otherwise there is no one-dimensional sufficient statistic.

8.33 a.

$$P(Y_1 > k \text{ or } Y_n > 1|\theta = 0) = P(Y_1 > k|\theta = 0) = (1-k)^n = \alpha$$

So $k = 1 - \alpha^{1/n}$.

b.

$$\begin{aligned}\beta(\theta) &= P(Y_n > 1 \text{ or } Y_1 > k|\theta)\\ &= P(Y_n > 1|\theta) + P(Y_1 > k \text{ and } Y_n \leq 1)\\ &= \begin{cases} 1 & \theta > 1\\ 1 - (1-\theta)^n + (1 - \max\{k, \theta\})^n & \theta \leq 1\end{cases}\end{aligned}$$

c.

$$f(x|\theta) = 1_{(\theta,\infty)}(Y_1)1_{(-\infty,\theta+1)}(Y_n)$$

Fix $\theta' > 0$. Suppose $k \leq \theta'$. Then $\beta(\theta') = 1$.

Suppose $k > \theta'$. Take $k' = 1$ in the NP lemma. Then

$$\begin{aligned} f(x|\theta') < f(x|\theta_0) \Rightarrow &\qquad 0 < Y_1 < \theta' < k, \text{ so } x \notin R\\ f(x|\theta') > f(x|\theta_0) \Rightarrow &\qquad 1 < Y_n < \theta' + 1, \text{ so } x \in R\end{aligned}$$

So R is a NP test for any θ'.

So R is UMP.

d. The power is one for all n if $\theta > 1$.

Assignment 7

Problem 8.31
Problem 8.34

Due Friday, March 14, 2003.

Problem 8.49
Problem 8.54

Due Friday, March 14, 2003.

Problem 8.55
Problem 8.56

Due Friday, March 14, 2003.

Solutions

8.31 a. The joint PMF of the data is

$$f(x|\lambda) = \frac{\lambda^{\sum x_i} e^{-n\lambda}}{\prod x_i!}$$

For $\lambda_2 > \lambda_1$,

$$\frac{f(x|\lambda_2)}{f(x|\lambda_1)} = \left(\frac{\lambda_2}{\lambda_1}\right)^{\sum x_i} e^{n(\lambda_1 - \lambda_2)}$$

is increasing in $\sum x_i$, so has MLR. So a test which rejects the null hypothesis if $\overline{X} > c$ is UMP of its size.

b. If $\lambda = 1$, then $\overline{X} \sim \text{AN}(1, 1/n)$, so $c \approx 1 + z_\alpha/\sqrt{n}$.
If $\lambda = 2$, then $\overline{X} \sim \text{AN}(2, 2/n)$, so

$$\begin{aligned} P(\overline{X} > 1 + z_\alpha/\sqrt{n}|\lambda = 2) &\approx P\left(Z > \left(\frac{z_\alpha}{\sqrt{n}} - 1\right)\sqrt{\frac{n}{2}}\right) \\ &= P\left(Z > \frac{z_\alpha}{\sqrt{2}} - \sqrt{\frac{n}{2}}\right) \end{aligned}$$

For $\alpha = 0.05$, $z_\alpha = 1.645$.
For $\beta(2) = 0.9$,

$$\frac{z_\alpha}{\sqrt{2}} - \sqrt{\frac{n}{2}} = -z_{0.1} = -1.282$$

so

$$n = (z_\alpha + \sqrt{2} z_{1-\beta})^2 = (1.645 + \sqrt{2} \times 1.282)^2 = 11.27$$

so this suggest using $n = 12$.

It might be better to use the variance-stabilizing transformation $\sqrt{X}$. Either way, use of the CLT is a bit questionable.

8.34 a. $T \sim f(t-\theta)$, $T_0 \sim f(t)$. So

$$P(T > c|\theta) = P(T_0 + \theta > c)$$

This is increasing in θ.

b. First approach: MLR implies that $T > c$ is UMP of θ_1 against θ_2 for size $\alpha = P(T > c|\theta_1)$. Since $\phi'(t) \equiv \alpha$ is size α and $\beta'(t) = E[\phi'(T)|\theta] = \alpha$ for all θ, UMP implies that $\beta(\theta_2) \geq \beta'(\theta_2) = \alpha = \beta(\theta_1)$.

Second approach: Let $\alpha = P(T > c|\theta_1)$, $h(t) = g(t|\theta_2)/g(t|\theta_1)$. Then

$$\begin{aligned} P(T > c|\theta_2) - \alpha &= E[\phi(T) - \alpha|\theta_2] \\ &= \frac{E[(\phi(T) - \alpha)h(T)|\theta_1]}{E[h(T)|\theta_1]} \\ &\geq \frac{h(c)E[\phi(T) - \alpha|\theta_1]}{E[h(T)|\theta_1]} \\ &= \frac{h(c)(\alpha - \alpha)}{E[h(T)|\theta_1]} \\ &= 0 \end{aligned}$$

Can also go back to the NP proof.

8.49 a. The p-value is $P(X \geq 7|\theta = 0.5) = 1 - P(X \leq 6|\theta = 0.5) = 0.171875$. This can be computed in R with

```
> 1 - pbinom(6,10,0.5)
[1] 0.171875
```

b. The p-value is $P(X \geq 3|\lambda = 1) = 1 - P(X \leq 2|\lambda = 1) = 0.0803014$. This can be computed in R with

```
> 1 - ppois(2, 1)
[1] 0.0803014
```

c. If $\lambda = 1$ then the sufficient statistic $T = X_1 + X_2 + X_3$ has a Poisson distribution with mean $\lambda_T = 3$. The observed value of T is $t = 9$, so the p-value is $P(T \geq 9|\lambda_T = 3) = 1 - P(T \leq 8|\lambda_T = 3) = 0.001102488$.

8.54 a. From problem 7.22 the posterior distribution of $\theta|x$ is normal with mean and variance

$$\begin{aligned} E[\theta|X=x] &= \frac{\tau^2}{\tau^2+\sigma^2/n}\overline{x} \\ \operatorname{Var}(\theta|X=x) &= \frac{\tau}{1+n\tau^2/\sigma^2} \end{aligned}$$

So

$$P(\theta \leq 0|x) = P\left(Z \leq -\frac{(\tau^2/(\tau^2+\sigma^2/n))\overline{x}}{\sqrt{\tau^2/(1+n\tau^2/\sigma^2)}}\right) = P\left(Z \geq \frac{\tau}{\sqrt{(\sigma^2/n)(\tau^2+\sigma^2/n)}}\overline{x}\right)$$

b. The p-value is

$$P(\overline{X} \geq \overline{x}|\theta=0) = P\left(Z \geq \frac{1}{\sigma/\sqrt{n}}\overline{x}\right)$$

c. For $\tau = \sigma = 1$ the Bayes probability is larger than the p-value for $\overline{x} > 0$ since

$$\frac{1}{\sqrt{(\sigma^2/n)(1+\sigma^2/n)}} < \frac{1}{\sqrt{1/n}}$$

d . As $n \to \infty$,

$$\frac{\tau}{\sqrt{(\sigma^2/n)(\tau^2+\sigma^2/n)}}\overline{x} = \frac{1}{\sqrt{(\sigma^2/n)(1+\sigma^2/(\tau^2 n))}}\overline{x} \to \frac{1}{\sigma/n}$$

and therefore $P(\theta \leq 0|x)$ converges to the p-value.

8.55 The risk functions for these tests are given by

$$\begin{aligned} R(\theta,\delta) &= \begin{cases} b(\theta_0-\theta)(1-\beta(\theta)) & \text{for } \theta < \theta_0 \\ c(\theta-\theta_0)2\beta(\theta) & \text{for } \theta \geq \theta_0 \end{cases} \\ &= \begin{cases} b(\theta_0-\theta)P(Z+\theta > \theta_0 - z_\alpha) & \text{for } \theta < \theta_0 \\ c(\theta-\theta_0)2P(Z+\theta \leq \theta_0 - z_\alpha) & \text{for } \theta \geq \theta_0 \end{cases} \\ &= \begin{cases} b(\theta_0-\theta)(1-\Phi(\theta_0-\theta-z_\alpha) & \text{for } \theta < \theta_0 \\ c(\theta-\theta_0)2\Phi(\theta_0-\theta-z_\alpha) & \text{for } \theta \geq \theta_0 \end{cases} \end{aligned}$$

where Z is a stnadard normal random variable and Φ is the standard normal CDF.

8.56 The risk function for a test δ and zero-one loss is

$$R(\theta,\delta) = \begin{cases} 1-\beta(\theta) & \text{for } \theta \in \Theta_1 \\ \beta(\theta) & \text{for } \theta \in \Theta_0 \end{cases}$$

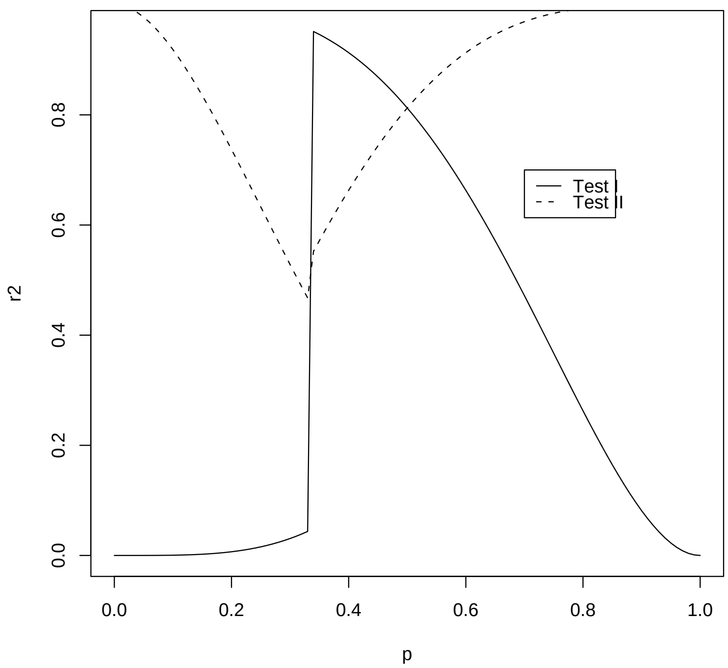
For the two tests in the problem this produces

$$R(p,\delta_I) = \begin{cases} 1 - P(X \leq 1|p) & \text{for } p > 1/3 \\ P(X \leq 1|p) & \text{for } p \leq 1/3 \end{cases}$$

and

$$R(p,\delta_I) = \begin{cases} 1 - P(X ge 41|p) & \text{for } p > 1/3 \\ P(X \geq 4|p) & \text{for } p \leq 1/3 \end{cases}$$

A graph of the risk functions is



Test II has lower risk for large and small values of p; Test I has lower risk for p between 1/3 and approximately 5.5.

These risk function graphs were produced in R using

```
b <- pbinom(1, 5, p)
b1 <- pbinom(1, 5, p)
b2 <- 1-pbinom(3, 5, p)
r1<-ifelse(p <= 1/3, b1, 1-b1)
r2<-ifelse(p <= 1/3, b2, 1-b2)
plot(p,r2,type="l")
lines(p,r1, lty = 2)
legend(0.7,0.7,c("Test I", "Test II"), lty=c(1,2))
```

Assignment 8

Problem 9.1
Problem 9.2

Due Friday, March 28, 2003.

Problem 9.4

Due Friday, March 28, 2003.

Problem 9.12
Problem 9.13

Due Friday, March 28, 2003.

Solutions

9.1 Since $L \leq U$,

$$\{L \leq \theta \leq U\}^c = \{L > \theta\} \cup \{U < \theta\}$$

Also

$$\{L > \theta\} \cap \{U < \theta\} = \emptyset$$

So

$$P(\{L \leq \theta \leq U\}^c) = P(L > \theta) + P(U < \theta) = \alpha_1 + \alpha_2$$

and thus

$$P(L \leq \theta \leq U) = 1 - \alpha_1 - \alpha_2$$

9.2 This can be interpreted conditionally or unconditionally. Given $X_1, \ldots, X_n$,

$$\begin{aligned} P(X_{n+1} \in \overline{x} \pm 1.96/\sqrt{n}|\overline{x}) &= P(Z \in \overline{x} - \theta \pm 1.96/\sqrt{n}) \\ &\leq P(Z \in 0 \pm 1.96/\sqrt{n}) \\ &< 0.95 \qquad \text{if } n > 1 \\ &\leq 0.95 \qquad \text{if } n = 1 \end{aligned}$$

Equality holds only if $n = 1$ and $\overline{x} = \theta$.

Unconditionally,

$$\begin{aligned} P(X_{n+1} \in \overline{X} \pm 1.96/\sqrt{n}) &= P(Z - \overline{Z} \in 0 \pm 1.96/\sqrt{n} \\ &= P\left(Z' \in 0 \pm \frac{1.96}{\sqrt{n}\sqrt{1+1/n}}\right) \\ &= P(Z' \in 0 \pm 1.96/\sqrt{n+1}) \\ &< 0.95 \qquad \text{for } n \geq 1 \end{aligned}$$

9.4 In this problem

$$X_i \text{ are } i.i.d.\ N(0,\sigma_X^2)$$
$$Y_i/\sqrt{\lambda_0} \text{ are } i.i.d.\ N(0,\sigma_X^2)$$

a.

$$\begin{aligned}\Lambda &= \frac{\left(\frac{n+m}{\sum X_i^2+\sum Y_i^2/\lambda_0}\right)^{(n+m)/2}}{\left(\frac{n}{\sum X_i^2}\right)^{n/2}\left(\frac{m}{\sum Y_i^2/\lambda_0}\right)^{m/2}}\\ &= CT^{n/2}(1-T)^{m/2}\end{aligned}$$

with

$$T = \frac{1}{1+\frac{\sum Y_i^2}{\lambda_0\sum X_i^2}} = \frac{1}{1+\frac{m}{n}F}$$

So $\Lambda < k$ if and only if $F < c_1$ or $F > c_2$.

b. $F/\lambda_0 \sim F_{m,n}$. Choose c_1, c_2 so $c_1 = F_{m,n,1-\alpha_1}$, $c_2 = F_{m,n,\alpha_2}$, $\alpha_1+\alpha_2=\alpha$, and $f(c_1)=f(c_2)$ for

$$f(t) = \left(\frac{1}{1+\frac{m}{n}t}\right)^{n/2}\left(1-\frac{1}{1+\frac{m}{n}t}\right)^{m/2}$$

c.

$$\begin{aligned}A(\lambda) &= \{X,Y : c_1 \le F \le c_2\}\\ &= \left\{X,Y : c_1 \le \frac{\sum Y_i^2/m}{\lambda\sum X_i^2/n} \le c_2\right\}\\ C(\lambda) &= \left\{\lambda : \frac{\sum Y_i^2/m}{c_2\sum X_i^2/n} \le \lambda \le \frac{\sum Y_i^2/m}{c_1\sum X_i^2/n}\right\}\\ &= \left[\frac{\sum Y_i^2/m}{c_2\sum X_i^2/n}, \frac{\sum Y_i^2/m}{c_1\sum X_i^2/n}\right]\end{aligned}$$

This is a $1-\alpha$ level CI.

9.12 All of the following are possible choices:

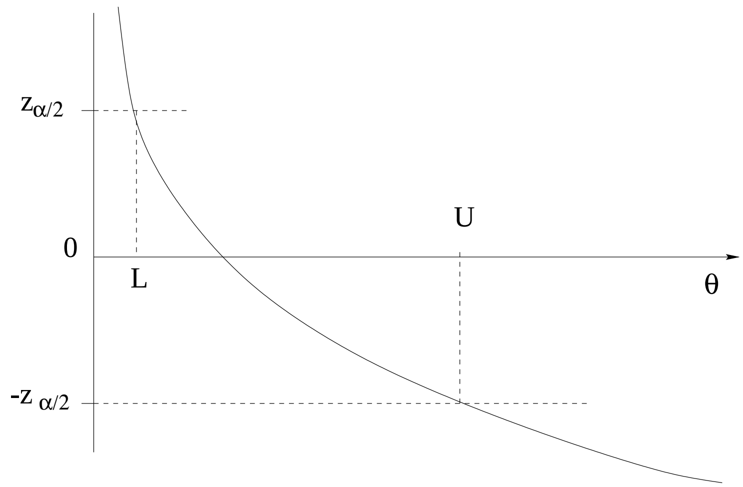
1. $\sqrt{n}(\overline{X}-\theta)/S \sim t_{n-1}$
2. $(n-1)S^2/\theta \sim \chi^2_{n-1}$
3. $\sqrt{n}(\overline{X}-\theta)/\sqrt{\theta} \sim N(0,1)$

The first two produce the obvious intervals.

For the third, look for those θ with

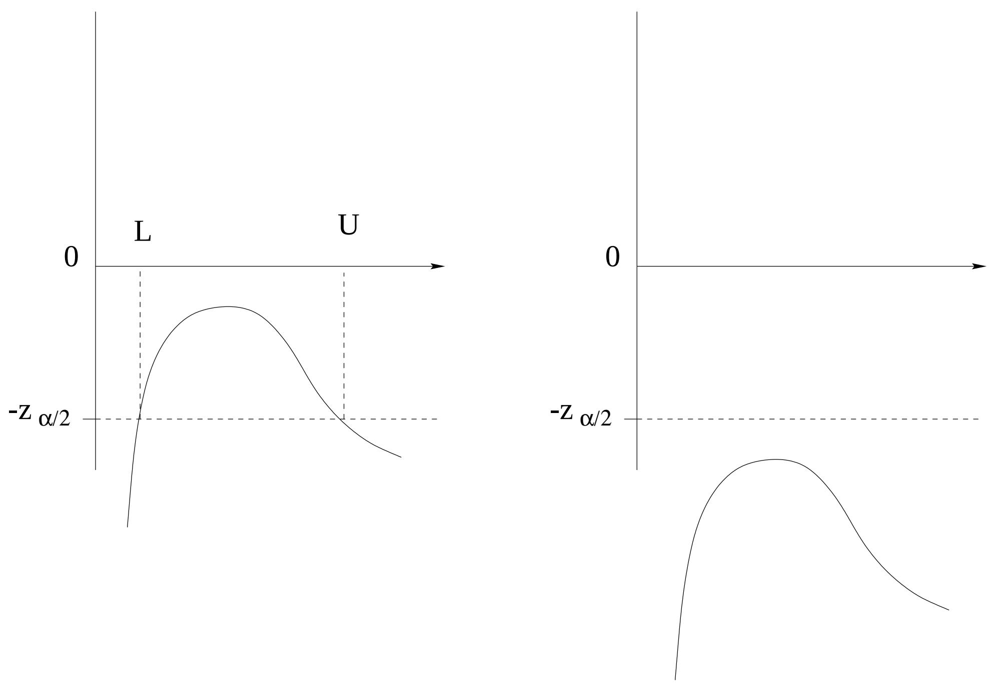
$$-z_{\alpha/2} < Q(X,\theta) = \frac{\overline{X}-\theta}{\sqrt{\theta/n}} < z_{\alpha/2}$$

If $\overline{x} \ge 0$, then $Q(x,\cdot)$ is decreasing, and the confidence set is an interval.



If $\overline{x} < 0$, then $Q(X, \theta)$ is negative with a single mode. If $\overline{x}$ is large enough (close enough to zero, then the confidence set is an interval, corresponding to the two solutions to $Q(x, \theta) = -z_{\alpha/2}$.

If $\overline{x}$ is too small, then there are no solutions and the confidence set is empty.



9.13 a. $U = \log X \sim \exp(1/\theta)$, $\theta U \sim \exp(1)$. So

$$\begin{aligned} P(Y/2 < \theta < Y) &= P(1/2 < \theta U < 1) \\ &= e^{-1} - e^{-1/2} = 0.239 \end{aligned}$$

b. $\theta U \sim \exp(1)$.

$$
\begin{aligned}
P(-\log(1-\alpha/2) < \theta U < -\log(\alpha/2)) &= 1-\alpha \\
P\left(\frac{-\log(1-\alpha/2)}{U} < \theta < \frac{-\log(\alpha/2)}{U}\right) &= 1-\alpha \\
[-\log(1-\alpha/2)Y, -\log(\alpha/2)Y] &= [0.479Y, 0.966Y]
\end{aligned}
$$

c. The interval in b. is a little shorter,

$$
\frac{b}{a} = \frac{0.487}{0.5}
$$

though it is not of optimal length.

Assignment 9

Problem 9.27
Problem 9.33

Due Friday, April 4, 2003.

Problem 10.1

Due Friday, April 4, 2003.

Solutions

9.27 a. The posterior density is

$$\begin{aligned}\pi(\lambda|X) &\propto \frac{1}{\lambda^n}e^{-\sum x_i/\lambda}\frac{1}{\lambda^{a+1}}e^{-1/(b\lambda)}\\ &= \frac{1}{\lambda^{n+a+1}}e^{-[1/b+\sum x_i]/\lambda}\\ &= \mathrm{IG}(n+a,[1/b+\sum x_i]^{-1})\end{aligned}$$

The inverse gamma density is unimodal, so the HPD region is an interval $[c_1, c_2]$ with c_1, c_2 chosen to have equal density values and $P(Y > 1/c_1) + P(Y < 1/c_2) = \alpha$, with $Y \sim \mathrm{Gamma}(n + a, [1/b + \sum x_i]^{-1})$.

.b The distribution of S^2 is $\mathrm{Gamma}((n-1)/2, 2\sigma^2/(n-1))$. The resulting posterior density is therefore

$$\begin{aligned}\pi(\sigma^2|s^2) &\propto \frac{(s^2)^{(n-1)/2-1}}{(\sigma^2/(n-1))^{(n-1)/2}}e^{-(n-1)s^2/\sigma^2}\frac{1}{(\sigma^2)^{a+1}}e^{-1/(b\sigma^2)}\\ &\propto \frac{1}{(\sigma^2)^{(n-1)/2+a+1}}e^{-[1/b+(n-1)s^2]/\sigma^2}\\ &= \mathrm{IG}((n-1)/2+a,[1/b+(n-1)s^2]^{-1})\end{aligned}$$

As in the previous part, the HPD region is an interval that can be determined by solving a corresponding set of equations.

c. The limiting posterior distribution is $\mathrm{IG}((n-1)/2, [(n-1)s^2]^{-1})$. The limiting HPD region is an interval $[c_1, c_2]$ with $c_1 = (n-1)s^2/\chi^2_{n-1,\alpha_1}$ and $c_2 = (n-1)s^2/\chi^2_{n-1,1-\alpha_2}$ where $\alpha_1 + \alpha_2 = \alpha$ and c_1, c_2 have equal posterior density values.

9.33 a. Since $0 \in C_a(x)$ for all a, x,

$$P_{\mu=0}(0 \in C_a(X)) = 1$$

For $\mu < 0$,

$$\begin{aligned} P_\mu(\mu \in C_a(X)) &= P_\mu(\min\{0, X - a\} \leq \mu) \\ &= P_\mu(X - a \leq \mu) = P_\mu(X - \mu \leq a) = 1 - \alpha \end{aligned}$$

if $a = z_\alpha$. For $\mu > 0$,

$$\begin{aligned} P_\mu(\mu \in C_a(X)) &= P_\mu(\max\{0, X - a\} \geq \mu) \\ &= P_\mu(X + a \geq \mu) = P_\mu(X - \mu \geq -a) = 1 - \alpha \end{aligned}$$

if $a = z_\alpha$.

b. For $\pi(\mu) \equiv 1$, $f(\mu|x) \sim N(x, 1)$.

$$\begin{aligned} P(\min\{0, x - a\} \leq \mu \leq \max\{0, x - a\}|X = x) &= P(x - a \leq \mu \leq x + a|X = x) \\ &= 1 - 2\alpha \end{aligned}$$

if $a = z_\alpha$ and $-z_\alpha \leq x \leq z_\alpha$. For $a = z_\alpha$ and $x > z_\alpha$,

$$\begin{aligned} P(\min\{0, x - a\} \leq \mu \leq \max\{0, x - a\}|X = x) &= P(-x \leq Z \leq a) \\ &\to P(Z \leq z) = 1 - \alpha \end{aligned}$$

as $x \to \infty$.

10.1 The mean is $\mu = \theta/3$, so the method of moments estimator is $W_n = 3\overline{X}_n$. By the law of large numbers $\overline{X}_n \xrightarrow{P} \mu = \theta/3$, so $W_n = 3\overline{X}_n \xrightarrow{P} \theta$.

Assignment 10

Problem 10.3
Problem 10.9 (but only for $e^{-\lambda}$; do not do $\lambda e^{-\lambda}$)

Due Friday, April 11, 2003.

Problem: Find the approximate joint distribution of the maximum likelihood estimators in problem 7.14 of the text.

Due Friday, April 11, 2003.

Problem: In the setting of problem 7.14 of the text, suppose $n = 100$, $\sum W_i = 71$, and $\sum Z_i = 7802$. Also assume a smooth, vague prior distribution. Find the posterior probability that $\lambda > 100$.

Due Friday, April 11, 2003.

Solutions

10.3 a. The derivative of the log likelihood is

$$\begin{aligned}\frac{\partial}{\partial\theta}\left(\frac{n}{2}\log\theta - \frac{\sum(X_i-\theta)^2}{2\theta}\right) &= -\frac{n}{2\theta} + \frac{\sum(X_i-\theta)^2}{1\theta^2} + \frac{2\sum(X_i-\theta)}{2\theta}\\ &= -\frac{n}{2\theta} + \frac{\sum X_i^2 - n\theta^2}{2\theta^2} = n\frac{W-\theta-\theta^2}{2\theta^2}\end{aligned}$$

So the MLE is a root of the quadratic equation $\theta^2+\theta-W=0$. The roots are

$$\theta_{1,2} = \frac{1}{2} \pm \sqrt{\frac{1}{4}+W}$$

The MLE is the larger root since that represents a local maximum and since the smaller root is negative.

b. The Fisher information is

$$\begin{aligned}I_n(\theta) &= -nE\left[\frac{\partial}{\partial\theta}\left(\frac{\theta^2+\theta-W}{2\theta^2}\right)\right]\\ &= \frac{nE[W]}{\theta^3} - \frac{n}{2\theta^2} = n\frac{E[W]-\theta/2}{\theta^3}\\ &= n\frac{\theta^2+\theta-\theta/2}{\theta^3} = n\frac{\theta+1/2}{\theta^2}\end{aligned}$$

So $\widehat{\theta} \sim \mathrm{AN}(\theta, \frac{\theta^2}{n(\theta+1/2)})$.

10.9 (but only for $e^{-\lambda}$; do not do $\lambda e^{-\lambda}$)

The UMVUE of $e^{-\lambda}$ is $V_n = (1-1/n)^{n\overline{X}}$ and the MLE is $e^{-\overline{X}}$. Since

$$\sqrt{n}(V_n - e^{-\overline{X}}) = \sqrt{n}O(1/n) = O(1/\sqrt{n})$$

both $\sqrt{n}(V_n - e^{-\lambda})$ and $\sqrt{n}(e^{\overline{X}} - e^{-\lambda})$ have the same normal limiting distribution and therefore their ARE is one.

In finite samples one can argue that the UMVUE should be preferred if unbiasedness is deemed important. The MLE is always larger than the UMVUE in this case, which might in some contexts be an argument for using the UMVUE. A comparison of mean square errors might be useful.

For the data provided, $e^{-\overline{X}} = 0.0009747$ and $V_n = 0.0007653$.

Problem: Find the approximate joint distribution of the maximum likelihood estimators in problem 7.14 of the text.

Solution: The log-likelihood is

$$\ell(\lambda, \mu) = -\sum w_i \log \lambda - (n - \sum w_i) \log \mu - \sum z_i \left(\frac{1}{\lambda} + \frac{1}{\mu}\right)$$

So

$$\begin{aligned}
\frac{\partial}{\partial \lambda}\ell(\lambda,\mu) &= -\frac{\sum w_i}{\lambda} + \frac{\sum z_i}{\lambda^2} \\
\frac{\partial}{\partial \mu}\ell(\lambda,\mu) &= -\frac{n - \sum w_i}{\mu} + \frac{\sum z_i}{\mu^2} \\
\frac{\partial^2}{\partial \lambda^2}\ell(\lambda,\mu) &= \frac{\sum w_i}{\lambda^2} - \frac{2\sum z_i}{\lambda^3} \\
\frac{\partial^2}{\partial \mu^2}\ell(\lambda,\mu) &= \frac{n - \sum w_i}{\mu^2} - \frac{2\sum z_i}{\mu^3} \\
\frac{\partial^2}{\partial \lambda \partial \mu}\ell(\lambda,\mu) &= 0 \\
E[W_i] &= \frac{\mu}{\lambda + \mu} \\
E[Z_i] &= \frac{\lambda\mu}{\lambda + \mu}
\end{aligned}$$

So

$$\begin{aligned}
E\left[-\frac{\partial^2}{\partial \lambda^2}\ell(\lambda,\mu)\right] &= 2\frac{n}{\lambda^2}\frac{\mu}{\lambda+\mu} - \frac{n}{\lambda^2}\frac{\mu}{\lambda+\mu} = \frac{n}{\lambda^2}\frac{\mu}{\lambda+\mu} \\
E\left[-\frac{\partial^2}{\partial \mu^2}\ell(\lambda,\mu)\right] &= \frac{n}{\mu^2}\frac{\lambda}{\lambda+\mu}
\end{aligned}$$

and thus

$$\begin{bmatrix} \widehat{\lambda} \\ \widehat{\mu} \end{bmatrix} \sim \text{AN}\left(\begin{bmatrix} \lambda \\ \mu \end{bmatrix}, \begin{bmatrix} \frac{\lambda^2(\lambda+\mu)}{n\mu} & 0 \\ 0 & \frac{\mu^2(\lambda+\mu)}{n\lambda} \end{bmatrix} \right)$$

Problem: In the setting of problem 7.14 of the text, suppose $n = 100$, $\sum W_i = 71$, and $\sum Z_i = 7802$. Also assume a smooth, vague prior distribution. Find the posterior probability that $\lambda > 100$.

Solution: The MLE's are

$$\begin{aligned} \widehat{\lambda} &= \frac{\sum Z_i}{\sum W_i} = 109.89 \\ \widehat{\mu} &= \frac{\sum Z_i}{n - \sum W_i} \end{aligned}$$

The observed information is

$$\widehat{I}_n(\widehat{\lambda}, \widehat{\mu}) = \begin{bmatrix} \frac{\sum W_i}{\widehat{\lambda}^2} & 0 \\ 0 & \frac{n - \sum W_i}{\widehat{\mu}^2} \end{bmatrix}$$

Thus the marginal posterior distribution of λ is approximately

$$N(\widehat{\lambda}, \widehat{\lambda}^2 / \sum W_i) = N(109.89, 170.07)$$

So

$$P(\lambda > 100 | X) \approx P\left(Z > \frac{100 - 109.89}{\sqrt{170.07}} \right) \approx 0.7758$$

Assignment 11

Problem: Let $X_1, \dots, X_n$ be a random sample from a Pareto$(1, \beta)$ distribution with density $f(x|\beta) = \beta/x^{\beta+1}$ for $x \geq 1$. Find the asymptotic relative efficiency of the method of moments estimator of β to the MLE of β.

Due Friday, April 18, 2003.

Problem: Let $X_1, \dots, X_n$ be *i.i.d.* Poisson(λ) and let $W = e^{-\overline{X}}$. Find the parametric bootstrap variance $\text{Var}^*(W)$ and show that $\text{Var}^*(W)/\text{Var}(W) \xrightarrow{P} 1$ as $n \to \infty$.

Due Friday, April 18, 2003.

1. Let $X_1, \dots, X_n$ be a random sample that may come from a Poisson distribution with mean λ. Find the sandwich estimator of the asymptotic variance of the MLE $\widehat{\lambda} = \overline{X}$.
2. Let $g(x) = e^{-x}$ for $x > 0$ be an exponential density with mean one and let $f(x|\theta)$ be a $N(\theta, 1)$ density. Find the value θ^* corresponding to the density of the form $f(x|\theta)$ that is closest to g in Kullback-Liebler divergence.

Due Friday, April 18, 2003.

Solutions

Problem: Let $X_1, \dots, X_n$ be a random sample from a Pareto$(1, \beta)$ distribution with density $f(x|\beta) = \beta/x^{\beta+1}$ for $x \geq 1$. Find the asymptotic relative efficiency of the method of moments estimator of β to the MLE of β.

Solution: The mean is finite and $E[X] = \beta/(\beta - 1)$ if $\beta > 1$. So the method of moments estimator is $\widehat{\beta}_{\text{MM}} = \overline{X}/(\overline{X} - 1)$ if $\overline{X} > 1$ and undefined otherwise. The variance is finite and $\text{Var}(X) = \frac{\beta}{(\beta-1)^2(\beta-2)}$ if $\beta > 2$. So for $\beta > 2$ the central limit theorem implies that

$$\sqrt{n}(\overline{X} - \beta/(\beta - 1)) \xrightarrow{\mathcal{D}} N\left(0, \frac{\beta}{(\beta - 1)^2(\beta - 2)}\right)$$

Since $\widehat{\beta}_{\text{MM}} = g(\overline{X})$ with $g(x) = x/(x - 1)$ and $g'(x) = -1/(x - 1)^2$, we have $g'(\beta/(\beta - 1)) = -(\beta - 1)^2$ and the delta method shows that

$$\sqrt{n}(\widehat{\beta}_{\text{MM}} - \beta) \xrightarrow{\mathcal{D}} N\left(0, \frac{g'(\beta/(\beta - 1))^2\beta}{(\beta - 1)^2(\beta - 2)}\right) = N(0, \beta(\beta - 1)^2/(\beta - 2))$$

The MLE is $\widehat{\beta} = n/(\sum \log X_i)$ and the Fisher information is $I_n(\beta) = n/\beta^2$. So the asymptotic relative efficiency of the method of moments estimator to the MLE is

$$\text{ARE}(\widehat{\beta}_{\text{MM}}, \widehat{\beta}) = \frac{\beta^2}{\beta(\beta-1)^2/(\beta-2)} = \frac{\beta(\beta-2)}{(\beta-1)^2}$$

for $\beta > 2$. For $\beta \leq 2$ the method of moments estimator will exist for large n but will not he $\sqrt{n}$-consistent; it makes sense to say the asymptotic relative efficiency is zero in this case.

Problem: Let $X_1, \ldots, X_n$ be *i.i.d.* Poisson(λ) and let $W = e^{-\overline{X}}$. Find the parametric bootstrap variance $\text{Var}^*(W)$ and show that $\text{Var}^*(W)/\text{Var}(W) \xrightarrow{P} 1$ as $n \to \infty$.

Solution: Using the MGF of the Poisson distribution we have

$$\begin{aligned}
E[W|\lambda] &= M_{\sum X_i}(-1/n) = \exp\{n\lambda(e^{-1/n} - 1)\} \\
E[W^2|\lambda] &= M_{\sum X_i}(-2/n) = \exp\{n\lambda(e^{-2/n} - 1)\}
\end{aligned}$$

The variance of W is therefore

$$\begin{aligned}
\text{Var}(W|\lambda) &= E[W^2|\lambda] - E[W|\lambda]^2 \\
&= \exp\{n\lambda(e^{-2/n} - 1)\} - \exp\{2n\lambda(e^{-1/n} - 1)\} \\
&= \exp\{2n\lambda(e^{-1/n} - 1)\}(\exp\{n\lambda(e^{-2/n} - 2e^{-1/n} + 1)\} - 1) \\
&= \exp\{2n\lambda(e^{-1/n} - 1)\}(\exp\{n\lambda(e^{-1/n} - 1)^2\} - 1) \\
&= \lambda b_n g(a_n\lambda, b_n\lambda)
\end{aligned}$$

with

$$\begin{aligned}
g(x,y) &= \begin{cases} e^{2x}\frac{e^y - 1}{y} & \text{if } y \neq 0 \\ e^{2x} & \text{if } y = 0 \end{cases} \\
a_n &= n(e^{-1/n} - 1) \\
b_n &= n(e^{-1/n} - 1)^2 = \frac{a_n^2}{n}
\end{aligned}$$

The bootstrap variance is

$$\begin{aligned}
\text{Var}^*(W) &= \text{Var}(W|\lambda = \overline{X}) \\
&= \exp\{2n\overline{X}(e^{-1/n} - 1)\}(\exp\{n\overline{X}(e^{-1/n} - 1)^2\} - 1) \\
&= \overline{X} b_n g(a_n\overline{X}, b_n\overline{X})
\end{aligned}$$

Now g is continuous, $a_n \to 1$ and $b_n \to 0$. So by the law of large numbers, Slutsky's theorem, and the continuous mapping theorem

$$\frac{\text{Var}^*(W)}{\text{Var}(W|\lambda)} = \frac{\overline{X} g(a_n\overline{X}, b_n\overline{X})}{\lambda g(a_n\lambda, b_n\lambda)} \xrightarrow{P} \frac{\lambda g(\lambda, 0)}{\lambda g(\lambda, 0)} = 1$$

Problem:

1. Let $X_1, \ldots, X_n$ be a random sample that may come from a Poisson distribution with mean λ. Find the sandwich estimator of the asymptotic variance of the MLE $\widehat{\lambda} = \overline{X}$.
2. Let $g(x) = e^{-x}$ for $x > 0$ be an exponential density with mean one and let $f(x|\theta)$ be a $N(\theta, 1)$ density. Find the value θ^* corresponding to the density of the form $f(x|\theta)$ that is closest to g in Kullback-Liebler divergence.

Solution:

1. For the Poisson distribution

$$\begin{aligned} \frac{\partial}{\partial\lambda} \log f(X|\lambda) &= \frac{X - \lambda}{\lambda} \\ \frac{\partial^2}{\partial\lambda^2} \log f(X|\lambda) &= -\frac{X}{\lambda^2} \end{aligned}$$

So the sandwich estimator of the asymptotic variance of the MLE is

$$\widehat{\text{Var}}(\sqrt{n}(\widehat{\lambda} - \lambda) = \frac{\frac{1}{n}\sum\left(\frac{\partial}{\partial\lambda}\log f(X_i|\widehat{\lambda})\right)^2}{\left(\frac{1}{n}\sum\frac{\partial^2}{\partial\lambda^2}\log f(X_i|\widehat{\lambda})\right)^2} = \frac{1}{n}\sum(X_i - \overline{X})^2$$

2. The Kullback-Liebler divergence from any distribution with density g to a $N(\theta, 1)$ distribution is

$$\text{KL}(g, f) = \int \log\frac{g(x)}{f(x|\theta)} g(x)dx = \text{const} + \frac{1}{2}\int (x - \theta)^2 g(x)dx$$

This is minimized by $\theta^* = E_g[X]$; for this particular choice of g this means $\theta^* = 1$.

Assignment 12

Problem 10.30 (b)

Due Friday, April 25, 2003.

Problem: Consider the setting of Problem 10.31. Derive an expression for $-2\log\Lambda$, where Λ is the likelihood ratio test statistic, and find the approximate distribution of this quantity under the null hypothesis.

Due Friday, April 25, 2003.

Problem 10.38

Due Friday, April 25, 2003.

Solutions

10.30 (b) For the Huber M-estimator $\psi(-\infty) = -k$ and $\psi(\infty) = k$, so $\eta = 1/(1+1) = 1/2$ and the breakdown is 50%.

The formula for the breakdown given in this problem is only applicable to monotone ψ functions. For redescending ψ functions the estimating equation need not have a unique root. To resolve this one can specify that an estimator should be determined using a local root finding procedure starting at, say, the sample median. In this case the M-estimator inherits the 50% breakdown of the median. See Huber, pages 53–55, for a more complete discussion.

Problem: Consider the setting of Problem 10.31. Derive an expression for $-2\log\Lambda$, where Λ is the likelihood ratio test statistic, and find the approximate distribution of this quantity under the null hypothesis.

Solution: The restricted likelihood corresponds to n_1+n_2 Bernoulli trials with S_1+S_2 successes and common success probability p, so the MLE of p is $\widehat{p} = (S_1+S_2)/(n_1+n_2)$. The unrestricted likelihood consists of two independent sets of Bernoulli trials with success probabilities p_1 and p_2, and the correpsonding MLS's are $\widehat{p}_1 = S_1/n_1$ and $\widehat{p}_2 = S_2/n_2$. The likelihood ratio statistic is therefore

$$\Lambda = \frac{\widehat{p}^{S_1+S_2}(1-\widehat{p})^{F_1+F_2}}{\widehat{p}_1^{S_1}(1-\widehat{p}_1)^{F_1}\widehat{p}_2^{S_2}(1-\widehat{p}_2)^{F_2}} = \left(\frac{\widehat{p}}{\widehat{p}_1}\right)^{S_1}\left(\frac{\widehat{p}}{\widehat{p}_2}\right)^{S_2}\left(\frac{1-\widehat{p}}{1-\widehat{p}_1}\right)^{F_1}\left(\frac{1-\widehat{p}}{1-\widehat{p}_2}\right)^{F_2}$$

and

$$-2\log\Lambda = 2\left(S_1\log\frac{\widehat{p}_1}{\widehat{p}} + S_2\log\frac{\widehat{p}_2}{\widehat{p}} + F_1\log\frac{1-\widehat{p}_1}{1-\widehat{p}} + F_2\log\frac{1-\widehat{p}_2}{1-\widehat{p}}\right)$$

The restricted parameter space under the null hypothesis is one-dimensional and the unrestricted parameter space is two-dimensional. Thus under the null hypothesis the distribution of $-2\log\Lambda$ is approximately χ^2_1.

10.38 The log likelihood for a random sample from a Gamma(α,β) distribution is

$$\log L(\beta) = n\left(-\log\Gamma(\alpha) - \alpha\log\beta + (\alpha-1)\frac{1}{n}\sum\log X_i - \overline{X}/\beta\right)$$

So the score function is

$$V_n(\beta) = \frac{\partial}{\partial\beta}\log L(\beta) = n\left(-\frac{\alpha}{\beta} + \frac{\overline{X}}{\beta^2}\right) = n\frac{\overline{x}-\alpha\beta}{\beta^2}$$

and the Fisher information is

$$I_n(\beta) = -nE\left[\frac{\alpha}{\beta^2} - \frac{2\overline{X}}{\beta^3}\right] = n\frac{2\alpha\beta}{\beta^3} - n\frac{\alpha}{\beta^2} = n\frac{\alpha}{\beta^2}$$

So the score statistic is

$$\frac{V_n(\beta)}{\sqrt{I_n(\beta)}} = \sqrt{n}\frac{\overline{X}-\alpha\beta}{\sqrt{\alpha}\beta} = \sqrt{n}\frac{\overline{X}-\alpha\beta}{\sqrt{\alpha\beta^2}}$$

Assignment 13

Problem 10.41

Due Friday, May 2, 2003.

Problem: Let $x_1, \ldots, x_n$ be constants, and suppose

$$Y_i = \beta_1(1 - e^{-\beta_2 x_i}) + \varepsilon_i$$

with the ε_i independent $N(0.\sigma^2)$ ramdom variables.

a. Find the normal equations for the least squares estimators of β_1 and β_2.

b. Suppose β_2 is known. Find the least squares estimator for β_1 as a function of the data and β_2.

Due Friday, May 2, 2003.

Problem: Let $x_1, \ldots, x_n$ be constants, and suppose

$$Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

Let y^* be a constant and let let x^* satisfy

$$y^* = \beta_0 + \beta_1 x^*$$

that is, x^* is the value of x at which the mean response is y^*.

a. Find the maximum likelihood estimator $\widehat{x}^*$ of x^*.

b. Use the delta method to find the approximate sampling distribution of $\widehat{x}^*$.

Due Friday, May 2, 2003.

Solutions

10.41 This problem should have stated that r is assumed known.

a. The log likelihood for p is

$$\log L(p) = \text{const} + n(r \log p + \overline{x} \log(1-p))$$

The first and second partial derivatives with respoct to p are

$$\frac{\partial}{\partial p} = \frac{nr}{p} - \frac{n\overline{x}}{1-p}$$
$$\frac{\partial^2}{\partial p^2} = -\frac{nr}{p^2} - \frac{n\overline{x}}{(1-p)^2}$$

So the Fisher information is $I_n(p) = \frac{nr}{p^2(1-p)}$ and the score test statistic is

$$\sqrt{n}\frac{\frac{nr}{p} - \frac{n\overline{x}}{1-p}}{\sqrt{\frac{nr}{p^2(1-p)}}} = \sqrt{\frac{n}{r}}\left(\frac{(1-p)r + p\overline{x}}{\sqrt{1-p}}\right)$$

b. The mean is $\mu = r(1-p)/p$. The score statsitic can be written in terms of the mean as

$$\sqrt{\frac{n}{r}}\left(\frac{(1-p)r + p\overline{x}}{\sqrt{1-p}}\right) = \sqrt{n}\frac{\mu - \overline{x}}{\sqrt{\mu + \mu^2/r}}$$

A confidence interval is give by

$$C = \left\{\mu : \left|\sqrt{n}\frac{\mu - \overline{x}}{\sqrt{\mu + \mu^2/r}}\right| \leq z_{\alpha/2}\right\}$$

The endpoints are the solutions to a quatriatic,

$$U, L = \frac{r(8\overline{x} + z_{\alpha/2}^2) \pm \sqrt{rz_{\alpha/2}^2}\sqrt{16r\overline{x} + 16\overline{x}^2 + rz_{\alpha/2}^2}}{8r - 2z_{\alpha/2}^2}$$

To use the continuity corection, replace $\overline{x}$ with $\overline{x} + \frac{1}{2n}$ for the upper end point and $\overline{x} - \frac{1}{2n}$ for the lower end point.

Problem: Let $x_1, \dots, x_n$ be constants, and suppose

$$Y_i = \beta_1(1 - e^{-\beta_2 x_i}) + \varepsilon_i$$

with the ε_i independent $N(0.\sigma^2)$ ramdom variables.

a. Find the normal equations for the least squares estimators of β_1 and β_2.
b. Suppose β_2 is known. Find the least squares estimator for β_1 as a function of the data and β_2.

Solution:

a. The mean response is $\mu_i(\beta) = \beta_1(1 - e^{-\beta_2 x_i})$. So the partial derivatives are

$$\begin{aligned}\frac{\partial}{\partial\beta_1}\mu_i(\beta) &= 1 - e^{-\beta_2 x_i}\\ \frac{\partial}{\partial\beta_2}\mu_i(\beta) &= \beta_1(1 - e^{-\beta_2 x_i})x_i\end{aligned}$$

So the normal equations are

$$\begin{aligned}\sum_{i=1}^n \beta_1(1 - e^{-\beta_2 x_i})^2 &= \sum_{i=1}^n (1 - e^{-\beta_2 x_i})Y_i\\ \sum_{i=1}^n \beta_1^2(1 - e^{-\beta_2 x_i})^2 x_i &= \sum_{i=1}^n \beta_1(1 - e^{-\beta_2 x_i})x_i Y_i\end{aligned}$$

b. If β_2 is known then the least squares estimator for β_1 can be found by solving the first normal equation:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n (1 - e^{-\beta_2 x_i})Y_i}{\sum_{i=1}^n (1 - e^{-\beta_2 x_i})^2}$$

Problem: Let $x_1, \ldots, x_n$ be constants, and suppose

$$Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

Let y^* be a constant and let let x^* satisfy

$$y^* = \beta_0 + \beta_1 x^*$$

that is, x^* is the value of x at which the mean response is y^*.

a. Find the maximum likelihood estimator $\widehat{x}^*$ of x^*.
b. Use the delta method to find the approximate sampling distribution of $\widehat{x}^*$.

Solution: This prolem should have explicitly assumed normal errors.

a. Since $x^* = (y^* - \beta_1)/\beta_2$, the MLE is

$$\widehat{x}^* = \frac{y^* - \widehat{\beta}_1}{\widehat{\beta}_2}$$

by MLE invariance.

b. The partial derivatives of the function $g(\beta_1, \beta_2) = (y^* - \beta_1)/\beta_2$ are

$$\begin{aligned}\frac{\partial}{\partial\beta_1}g(\beta_1, \beta_2) &= -\frac{1}{\beta_2}\\ \frac{\partial}{\partial\beta_2}g(\beta_1, \beta_2) &= -\frac{y^* - \beta_1}{\beta_2^2}\end{aligned}$$

So for $\beta_2 \neq 0$ the variance of the approximate sampling distribution is

$$\begin{aligned}
\widehat{\text{Var}}(\widehat{x}^*) &= \nabla g \sigma^2 (X^T X)^{-1} \nabla g^T \\
&= \frac{\frac{1}{n\beta_2^2}\sum x_i^2 - 2\overline{x}\frac{y^*-\beta_1}{\beta_2^3} + \frac{(y^*-\beta_1)^2}{\beta_2^4}}{\sum (x_i - \overline{x})^2} \\
&= \frac{\frac{1}{n\beta_2^2}\sum (x_i - \overline{x})^2 + \frac{(y^*-\beta_1-\beta_2\overline{x})^2}{\beta_2^4}}{\sum (x_i - \overline{x})^2} \\
&= \frac{1}{n\beta_2^2} + \frac{(y^*-\beta_1-\beta_2\overline{x})^2}{\beta_2^4 \sum (x_i - \overline{x})^2}
\end{aligned}$$

So by the delta method $\widehat{x}^* \sim \text{AN}(x^*, \widehat{\text{Var}}(\widehat{x}^*))$. The approximation is reasonably good if β_2 is far from zero, but the actual mean and variance of $\widehat{x}^*$ do not exist.