BIOSTAT 602 Biostatistical Inference
Homework 07

Ashton Baker

Tuesday March 13, 2017

1. Let $X_1, \dots X_n$ be i.i.d. observations from a gamma distribution with pdf

$$f_X(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \qquad x>0, \quad \alpha,\beta>0$$

   (a) Show that the sample arithmetic mean $T(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^n X_i$ and sample geometric mean $S(\mathbf{X}) = \left(\prod_{i=1}^n X_i\right)^{1/n}$ are jointly sufficient and complete for α, β.
   Solution. To prove sufficiency, consider the sample likelihood

$$\begin{aligned}\mathcal{L}(\alpha,\beta|\mathbf{x}) &= \prod_{i=1}^n \left(\frac{1}{\Gamma(\alpha)\beta^\alpha} x_i^{\alpha-1} e^{-x_i/\beta}\right) \\ &= \frac{1}{\Gamma(\alpha)^n \beta^{n\alpha}} \left(\prod_{i=1}^n x_i\right)^{\alpha-1} e^{-\frac{1}{\beta}\sum_{i=1}^n x_i}\end{aligned}$$

   This leaves a function which only depends on $\theta = \{\alpha,\beta\}$ via the statistics $T(\mathbf{X})$ and $S(\mathbf{X})$. So T and S are jointly sufficient for α, β.
   Since the pdf of the gamma distribution can be written

$$f_X(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{-1} \exp\left[-\frac{1}{\beta}x + \alpha \log x\right],$$

   the gamma distribution belongs to the exponential family of distributions with $t_1(x) = x$, $t_2(x) = \log x$, and therefore by Theorem 6.2.25,

$$U(\mathbf{X}) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n \log(X_i)\right) = \left(\sum_{i=1}^n X_i, \log\prod_{i=1}^n X_i\right)$$

   is sufficient for α, β. And since $(S(\mathbf{X}), T(\mathbf{X})) = \left(\frac{1}{n}\sum_{i=1}^n X_i, \left(\prod_{i=1}^n X_i\right)^{1/n}\right)$ is itself a function of $\mathbf{X}$ only in terms of $U(\mathbf{X})$, S and T are also jointly complete for α, β.

   (b) find the UMVUE for $(\alpha\beta)^n$.

   (c) Show that T and S/T are independent random variables.

2. Let $X_1, \dots, X_n$ be i.i.d. observations from an Inverse Gaussian distribution $IG(\mu,\lambda)$ with pdf

$$f_x(x|\mu,\lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left[-\lambda(x-\mu)^2/(2\mu^2 x)\right] \qquad x>0, \quad \mu,\lambda>0$$

   (a) Show that the sample arithmetic mean $T(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^n X_i$ and sample harmonic mean $S(\mathbf{X}) = \frac{n}{\sum_{i=1}^n (1/X_i)}$ are jointly sufficient and complete for μ, λ.

   (b) Show that the MLEs for μ, λ are

$$\hat{\mu} = \overline{X}, \quad \hat{\lambda} = \frac{n}{\sum_{i=1}^n \left(\frac{1}{X_i} - \frac{1}{\overline{X}}\right)}.$$

   You need not check for global optimality, but must verify that the MLEs are local maximizers and fall inside the parameter space.
   Solution. First, we calculate $\ell(\mu,\lambda|\mathbf{x})$:

$$\begin{aligned}\mathcal{L}(\mu,\lambda|\mathbf{x}) &= \prod_{i=1}^n \left(\frac{\lambda}{2\pi x_i^3}\right)^{1/2} \exp\left[-\lambda(x_i-\mu)^2/(2\mu^2 x_i)\right] \\ \ell(\mu,\lambda|\mathbf{x}) &= \sum_{i=1}^n \log\left[\left(\frac{\lambda}{2\pi x_i^3}\right)^{1/2} \exp\left[-\lambda(x_i-\mu)^2/(2\mu^2 x_i)\right]\right] \\ &= \sum_{i=1}^n \left[\frac{1}{2}\log\left(\frac{\lambda}{2\pi}\right) - \frac{3}{2}\log x_i - \lambda(x_i-\mu)^2/(2\mu^2 x_i)\right]\end{aligned}$$

Then $\hat{\mu}_{MLE}$ is a value of μ such that $\partial\ell(\mu,\lambda|\mathbf{x})/\partial\mu = 0$. So we can determine this by solving:

$$\begin{aligned}
0 &= \frac{\partial}{\partial\mu}\sum_{i=1}^{n}\left[\frac{1}{2}\log\left(\frac{\lambda}{2\pi}\right) - \frac{3}{2}\log x_i - \lambda(x_i-\mu)^2/(2\mu^2 x_i)\right] \\
0 &= \sum_{i=1}^{n}\left[0 - 0 + \frac{\lambda(x_i-\mu)}{\mu^3}\right] \\
0 &= \frac{\lambda}{\mu^3}\left(-n\mu + \sum_{i=1}^{n} x_i\right) \\
\mu &= \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{X}
\end{aligned}$$

(c) Using the fact that $n\lambda/\hat{\lambda}$ has a χ^2_{n-1} distribution, find the bias and MSE of $1/\hat{\lambda}$ as an estimator of $1/\lambda$.

(d) Find the bias and MSE of $\hat{\lambda}$ and MSE of $\hat{\lambda}$ as an estimator of λ for $n > 5$. You can use the fact that if $Y \sim \chi^2_k$, then

$$\begin{aligned}
\mathbb{E}\left[1/Y\right] &= \frac{1}{k-2}, \quad k > 2 \\
\mathrm{Var}\left(1/Y\right) &= \frac{2}{(k-2)^2(k-4)}, \quad k > 4
\end{aligned}$$

3. The following are related to Poisson distribution.

(a) Let X be a single observation from a Poisson(λ) distribution. Show that the only unbiased estimator (and hence UMVUE) of $\exp(-2\lambda)$ is $T(X) = (-1)^X$. Is this a reasonable estimator?

Solution. Let $T(X)$ be an unbiased estimator of $\exp(-2\lambda)$. Then $\mathbb{E}\left[T(X)\right] = e^{-2\lambda}$, so

$$\begin{aligned}
\sum_{x=0}^{\infty} T(x)e^{-\lambda}\frac{\lambda^x}{x!} &= e^{-2\lambda} \\
e^{-\lambda}\sum_{x=0}^{\infty} T(x)\frac{\lambda^x}{x!} &= e^{-2\lambda} \\
\sum_{x=0}^{\infty} T(x)\frac{\lambda^x}{x!} &= e^{-\lambda}
\end{aligned}$$

Then, we can represent $e^{-\lambda}$ as a Taylor series to obtain

$$\sum_{x=0}^{\infty} T(x)\frac{\lambda^x}{x!} = \sum_{x=0}^{\infty}(-1)^x\frac{\lambda^x}{x!}.$$

The uniqueness of the Taylor series requires that $T(X) = (-1)^X$, so this is the only estimator which is unbiased for $\exp(-2\lambda)$.

I would say this is *not* a reasonable estimator because the quantity $\exp(-2\lambda)$ is always positive, and yet $(-1)^X$ can easily be negative. It's not reasonable for an estimator to fall outside the parameter space.

(b) Let $X_1, X_2, \ldots, X_n$ be an i.i.d. random sample from a Poisson(λ) distribution, where $n \geq 3$. Show that the only unbiased estimator (and hence UMVUE) of $\exp(-2\lambda)$ is

$$T(\mathbf{X}) = \left(1 - \frac{2}{n}\right)^{\sum X_i}$$

4. Let $X_1, \ldots, X_n$ be an i.i.d. random sample from a Bernoulli(p) distribution, where $n \geq 3$. find the UMVUE for p^3.

Solution. Given that the pmf of a Bernoulli distribution can be written $f_X(x|p) = p^x(1-p)^{1-x}$, the sample pmf is

$$\begin{aligned}\mathcal{L}(p|\mathbf{x}) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum x_i}(1-p)^{\sum(1-x_i)} \\ &= p^{\sum x_i}(1-p)^{n-\sum x_i}\end{aligned}$$

which depends on $\mathbf{x}$ only in terms of $\sum_{i=1}^{n} x_i$. So $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is sufficient for p.

We can also represent the pmf of a Bernoulli distribution as

$$\begin{aligned}f_x(x|p) &= p^x(1-p)^{1-x} \\ &= p^x(1-p)^{-x}(1-p)^1 \\ &= \left(\frac{p}{1-p}\right)^p (1-p) \\ &= \exp\left(\log\left(\frac{p}{1-p}\right)x + \log(1-p)\right)\end{aligned}$$

so the Bernoulli distribution is a member of the exponential family, and

$$T(X) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} 1\right) = \left(\sum_{i=1}^{n} X_i, n\right)$$

is a complete statistic for p by Theorem 6.2.25.