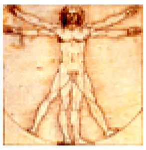

DOTTORATO DI RICERCA IN METODOLOGIA DELLE SCIENZE SOCIALI

Introduzione all'inferenza statistica
a.a. 2008-2009

Soluzione esercizi assegnati tratti da Casella e Berger

6.1 By the Factorization Theorem, $|X|$ is sufficient because the pdf of X is

$$f(x|\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} = \frac{1}{\sqrt{2\pi}\sigma} e^{-|x|^2/2\sigma^2} = g(|x|\,|\,\sigma^2) \cdot \underbrace{1}_{h(x)}.$$

6.2 By the Factorization Theorem, $T(X) = \min_i(X_i/i)$ is sufficient because the joint pdf is

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} e^{i\theta - x_i} I_{(i\theta,+\infty)}(x_i) = \underbrace{e^{in\theta} I_{(\theta,+\infty)}(T(\mathbf{x}))}_{g(T(\mathbf{x})|\theta)} \cdot \underbrace{e^{-\Sigma_i x_i}}_{h(\mathbf{x})}.$$

Notice, we use the fact that $i > 0$, and the fact that all x_is $> i\theta$ if and only if $\min_i(x_i/i) > \theta$.

6.3 Let $x_{(1)} = \min_i x_i$. Then the joint pdf is

$$f(x_1, \ldots, x_n|\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma} e^{-(x_i-\mu)/\sigma} I_{(\mu,\infty)}(x_i) = \underbrace{\left(\frac{e^{\mu/\sigma}}{\sigma}\right)^n e^{-\Sigma_i x_i/\sigma} I_{(\mu,\infty)}(x_{(1)})}_{g\left(x_{(1)}, \Sigma_i x_i | \mu, \sigma\right)} \cdot \underbrace{1}_{h(\mathbf{x})}.$$

Thus, by the Factorization Theorem, $\left(X_{(1)}, \sum_i X_i\right)$ is a sufficient statistic for (μ, σ).

6.17 The population pmf is $f(x|\theta) = \theta(1-\theta)^{x-1} = \frac{\theta}{1-\theta} e^{\log(1-\theta)x}$, an exponential family with $t(x) = x$. Thus, $\sum_i X_i$ is a complete, sufficient statistic by Theorems 6.2.10 and 6.2.25. $\sum_i X_i - n \sim$ negative binomial(n, θ).

6.32 In the Formal Likelihood Principle, take $E_1 = E_2 = E$. Then the conclusion is $\mathrm{Ev}(E, x_1) = \mathrm{Ev}(E, x_2)$ if $L(\theta|x_1)/L(\theta|x_2) = c$. Thus evidence is equal whenever the likelihood functions are equal, and this follows from Formal Sufficiency and Conditionality.

6.35 Let 1 = success and 0 = failure. The four sample points are $\{0, 10, 110, 111\}$. From the likelihood principle, inference about p is only through $L(p|\mathbf{x})$. The values of the likelihood are 1, p, p^2, and p^3, and the sample size does not directly influence the inference.

7.6 a. $f(\mathbf{x}|\theta) = \prod_i \theta x_i^{-2} I_{[\theta,\infty)}(x_i) = \left(\prod_i x_i^{-2}\right) \theta^n I_{[\theta,\infty)}(x_{(1)})$. Thus, $X_{(1)}$ is a sufficient statistic for θ by the Factorization Theorem.

b. $L(\theta|\mathbf{x}) = \theta^n \left(\prod_i x_i^{-2}\right) I_{[\theta,\infty)}(x_{(1)})$. θ^n is increasing in θ. The second term does not involve θ. So to maximize $L(\theta|\mathbf{x})$, we want to make θ as large as possible. But because of the indicator function, $L(\theta|\mathbf{x}) = 0$ if $\theta > x_{(1)}$. Thus, $\hat{\theta} = x_{(1)}$.

c. $\mathrm{E}\,X = \int_\theta^\infty \theta x^{-1}\,dx = \theta \log x|_\theta^\infty = \infty$. Thus the method of moments estimator of θ does not exist. (This is the Pareto distribution with $\alpha = \theta$, $\beta = 1$.)

7.9 This is a uniform$(0,\theta)$ model. So $\mathrm{E}\,X = (0+\theta)/2 = \theta/2$. The method of moments estimator is the solution to the equation $\tilde{\theta}/2 = \bar{X}$, that is, $\tilde{\theta} = 2\bar{X}$. Because $\tilde{\theta}$ is a simple function of the sample mean, its mean and variance are easy to calculate. We have

$$\mathrm{E}\,\tilde{\theta} = 2\mathrm{E}\,\bar{X} = 2\mathrm{E}\,X = 2\frac{\theta}{2} = \theta, \quad \text{and} \quad \mathrm{Var}\,\tilde{\theta} = 4\mathrm{Var}\,\bar{X} = 4\frac{\theta^2/12}{n} = \frac{\theta^2}{3n}.$$

The likelihood function is

$$L(\theta|\mathbf{x}) = \prod_{i=1}^n \frac{1}{\theta} I_{[0,\theta]}(x_i) = \frac{1}{\theta^n} I_{[0,\theta]}(x_{(n)}) I_{[0,\infty)}(x_{(1)}),$$

where $x_{(1)}$ and $x_{(n)}$ are the smallest and largest order statistics. For $\theta \geq x_{(n)}$, $L = 1/\theta^n$, a decreasing function. So for $\theta \geq x_{(n)}$, L is maximized at $\hat{\theta} = x_{(n)}$. $L = 0$ for $\theta < x_{(n)}$. So the overall maximum, the MLE, is $\hat{\theta} = X_{(n)}$. The pdf of $\hat{\theta} = X_{(n)}$ is nx^{n-1}/θ^n, $0 \leq x \leq \theta$. This can be used to calculate

$$\mathrm{E}\,\hat{\theta} = \frac{n}{n+1}\theta, \quad \mathrm{E}\,\hat{\theta}^2 = \frac{n}{n+2}\theta^2 \quad \text{and} \quad \mathrm{Var}\,\hat{\theta} = \frac{n\theta^2}{(n+2)(n+1)^2}.$$

$\tilde{\theta}$ is an unbiased estimator of θ; $\hat{\theta}$ is a biased estimator. If n is large, the bias is not large because $n/(n+1)$ is close to one. But if n is small, the bias is quite large. On the other hand, $\mathrm{Var}\,\hat{\theta} < \mathrm{Var}\,\tilde{\theta}$ for all θ. So, if n is large, $\hat{\theta}$ is probably preferable to $\tilde{\theta}$.

7.11 a.

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_i \theta x_i^{\theta-1} = \theta^n \left(\prod_i x_i\right)^{\theta-1} = L(\theta|\mathbf{x}) \\ \frac{d}{d\theta}\log L &= \frac{d}{d\theta}\left[n\log\theta + (\theta-1)\log\prod_i x_i\right] = \frac{n}{\theta} + \sum_i \log x_i. \end{aligned}$$

Set the derivative equal to zero and solve for θ to obtain $\hat{\theta} = (-\frac{1}{n}\sum_i \log x_i)^{-1}$. The second derivative is $-n/\theta^2 < 0$, so this is the MLE. To calculate the variance of $\hat{\theta}$, note that $Y_i = -\log X_i \sim$ exponential$(1/\theta)$, so $-\sum_i \log X_i \sim$ gamma$(n, 1/\theta)$. Thus $\hat{\theta} = n/T$, where $T \sim$ gamma$(n, 1/\theta)$. We can either calculate the first and second moments directly, or use the fact that $\hat{\theta}$ is inverted gamma (page 51). We have

$$\begin{aligned} \mathrm{E}\frac{1}{T} &= \frac{\theta^n}{\Gamma(n)}\int_0^\infty \frac{1}{t} t^{n-1} e^{-\theta t}\,dt = \frac{\theta^n}{\Gamma(n)}\frac{\Gamma(n-1)}{\theta^{n-1}} = \frac{\theta}{n-1}. \\ \mathrm{E}\frac{1}{T^2} &= \frac{\theta^n}{\Gamma(n)}\int_0^\infty \frac{1}{t^2} t^{n-1} e^{-\theta t}\,dt = \frac{\theta^n}{\Gamma(n)}\frac{\Gamma(n-2)}{\theta^{n-2}} = \frac{\theta^2}{(n-1)(n-2)}, \end{aligned}$$

and thus

$$\mathrm{E}\,\hat\theta = \frac{n}{n-1}\theta \quad \text{and} \quad \mathrm{Var}\,\hat\theta = \frac{n^2}{(n-1)^2(n-2)}\theta^2 \to 0 \text{ as } n \to \infty.$$

b. Because $X \sim \text{beta}(\theta, 1)$, $\mathrm{E}X = \theta/(\theta+1)$ and the method of moments estimator is the solution to

$$\frac{1}{n}\sum_i X_i = \frac{\theta}{\theta+1} \Rightarrow \tilde\theta = \frac{\sum_i X_i}{n - \sum_i X_i}.$$

7.12 $X_i \sim$ iid Bernoulli(θ), $0 \le \theta \le 1/2$.

a. method of moments:

$$\mathrm{E}X = \theta = \frac{1}{n}\sum_i X_i = \bar X \quad \Rightarrow \quad \tilde\theta = \bar X.$$

MLE: In Example 7.2.7, we showed that $L(\theta|\mathbf{x})$ is increasing for $\theta \le \bar x$ and is decreasing for $\theta \ge \bar x$. Remember that $0 \le \theta \le 1/2$ in this exercise. Therefore, when $\bar X \le 1/2$, $\bar X$ is the MLE of θ, because $\bar X$ is the overall maximum of $L(\theta|\mathbf{x})$. When $\bar X > 1/2$, $L(\theta|\mathbf{x})$ is an increasing function of θ on $[0, 1/2]$ and obtains its maximum at the upper bound of θ which is $1/2$. So the MLE is $\hat\theta = \min\{\bar X, 1/2\}$.

b. The MSE of $\tilde\theta$ is $\mathrm{MSE}(\tilde\theta) = \mathrm{Var}\,\tilde\theta + \mathrm{bias}(\tilde\theta)^2 = (\theta(1-\theta)/n) + 0^2 = \theta(1-\theta)/n$. There is no simple formula for $\mathrm{MSE}(\hat\theta)$, but an expression is

$$\begin{aligned}\mathrm{MSE}(\hat\theta) &= \mathrm{E}(\hat\theta-\theta)^2 = \sum_{y=0}^{n}(\hat\theta-\theta)^2\binom{n}{y}\theta^y(1-\theta)^{n-y} \\ &= \sum_{y=0}^{[n/2]}\left(\frac{y}{n}-\theta\right)^2\binom{n}{y}\theta^y(1-\theta)^{n-y} + \sum_{y=[n/2]+1}^{n}\left(\frac{1}{2}-\theta\right)^2\binom{n}{y}\theta^y(1-\theta)^{n-y},\end{aligned}$$

where $Y = \sum_i X_i \sim \text{binomial}(n, \theta)$ and $[n/2] = n/2$, if n is even, and $[n/2] = (n-1)/2$, if n is odd.

c. Using the notation used in (b), we have

$$\mathrm{MSE}(\tilde\theta) = \mathrm{E}(\bar X-\theta)^2 = \sum_{y=0}^{n}\left(\frac{y}{n}-\theta\right)^2\binom{n}{y}\theta^y(1-\theta)^{n-y}.$$

Therefore,

$$\begin{aligned}\mathrm{MSE}(\tilde\theta) - \mathrm{MSE}(\hat\theta) &= \sum_{y=[n/2]+1}^{n}\left[\left(\frac{y}{n}-\theta\right)^2 - \left(\frac{1}{2}-\theta\right)^2\right]\binom{n}{y}\theta^y(1-\theta)^{n-y} \\ &= \sum_{y=[n/2]+1}^{n}\left(\frac{y}{n}+\frac{1}{2}-2\theta\right)\left(\frac{y}{n}-\frac{1}{2}\right)\binom{n}{y}\theta^y(1-\theta)^{n-y}.\end{aligned}$$

The facts that $y/n > 1/2$ in the sum and $\theta \le 1/2$ imply that every term in the sum is positive. Therefore $\mathrm{MSE}(\hat\theta) < \mathrm{MSE}(\tilde\theta)$ for every θ in $0 < \theta \le 1/2$. (Note: $\mathrm{MSE}(\hat\theta) = \mathrm{MSE}(\tilde\theta) = 0$ at $\theta = 0$.)

8.1 Let $X = \#$ of heads out of 1000. If the coin is fair, then $X \sim \text{binomial}(1000, 1/2)$. So

$$P(X \geq 560) = \sum_{x=560}^{1000} \binom{1000}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x} \approx .0000825,$$

where a computer was used to do the calculation. For this binomial, $\mathrm{E}X = 1000p = 500$ and $\mathrm{Var}\, X = 1000p(1-p) = 250$. A normal approximation is also very good for this calculation.

$$P\{X \geq 560\} = P\left\{\frac{X - 500}{\sqrt{250}} \geq \frac{559.5{-}500}{\sqrt{250}}\right\} \approx P\{Z \geq 3.763\} \approx .0000839.$$

Thus, if the coin is fair, the probability of observing 560 or more heads out of 1000 is very small. We might tend to believe that the coin is not fair, and $p > 1/2$.

8.2 Let $X \sim \text{Poisson}(\lambda)$, and we observed $X = 10$. To assess if the accident rate has dropped, we could calculate

$$P(X \leq 10 |\, \lambda = 15) = \sum_{i=0}^{10} \frac{e^{-15}\, 15^i}{i!} = e^{-15}\left[1{+}15{+}\frac{15^2}{2!}{+}\cdots{+}\frac{15^{10}}{10!}\right] \approx .11846.$$

This is a fairly large value, not overwhelming evidence that the accident rate has dropped. (A normal approximation with continuity correction gives a value of .12264.)

8.12 a. For $H_0: \mu \leq 0$ vs. $H_1: \mu > 0$ the LRT is to reject $\mathrm{H_0}$ if $\bar{x} > c\sigma/\sqrt{n}$ (Example 8.3.3). For $\alpha = .05$ take $c = 1.645$. The power function is

$$\beta(\mu) = P\left(\frac{\bar{X}{-}\mu}{\sigma/\sqrt{n}} > 1.645{-}\frac{\mu}{\sigma/\sqrt{n}}\right) = P\left(Z > 1.645{-}\frac{\sqrt{n}\mu}{\sigma}\right).$$

Note that the power will equal .5 when $\mu = 1.645\sigma/\sqrt{n}$.

b. For $H_0\colon \mu = 0$ vs. $H_A: \mu \neq 0$ the LRT is to reject H_0 if $|\bar{x}| > c\sigma/\sqrt{n}$ (Example 8.2.2). For $\alpha = .05$ take $c = 1.96$. The power function is

$$\beta(\mu) = P\left(-1.96 - \sqrt{n}\mu/\sigma \leq Z \leq 1.96 + \sqrt{n}\mu/\sigma\right).$$

In this case, $\mu = \pm 1.96\sigma/\sqrt{n}$ gives power of approximately .5.