STA 200B HW9

8.5

4. Suppose that $X_1, \dots, X_n$ form a random sample from the normal distribution with unknown mean μ and known variance σ^2. How large a random sample must be taken in order that there will be a confidence interval for μ with confidence coefficient 0.95 and length less than 0.01σ?

solution Since X_i are iid normal, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a standard normal distribution. So

$$P\left(-1.96 < \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} < 1.96\right) = 0.95, \quad \text{rewritten as}$$
$$P\left(\bar{X}_n - \frac{1.96\sigma}{\sqrt{n}} < \mu < \bar{X}_n + \frac{1.96\sigma}{\sqrt{n}}\right) = 0.95.$$

Therefore $(\bar{X}_n - 1.96\sigma/\sqrt{n}, \bar{X}_n + 1.96\sigma/\sqrt{n})$ will be a confidence interval for μ with confidence coefficient 0.95. The length of this interval is $3.92\sigma/\sqrt{n}$ which should be less than 0.01σ. This means $n > 153664$ (or $n \geq 153665$).

6. Suppose that $X_1, \dots, X_n$ form a random sample from the exponential distribution with unknown mean μ. Describe a method for constructing a confidence interval for μ with a specified confidence coefficient $\gamma\,(0 < \gamma < 1)$. Hint: Determine constants c_1 and c_2 such that $\Pr[c_1 > (1/\mu)\sum_{i=1}^{n} X_i < c_2] = \gamma$.

solution The exponential distribution with mean μ is the same as the gamma distribution with $\alpha = 1$ and $\beta = 1/\mu$. Therefore, by property 3 in page 44 of lecture notes, $\sum_{i=1}^{n} X_i$ will have the gamma distribution with parameters $\alpha = n$ and $\beta = 1/\mu$. The p.d.f is proportional to $x^{n-1}e^{-x/\mu}$. In turn, it follows that $\sum_{i=1}^{n} X_i/\mu$ has a p.d.f proportional to $(\mu x)^{n-1}e^{-x} \sim x^{n-1}e^{-x}$ and therefore follows a gamma distribution with $\alpha = n$ and $\beta = 1$.

From definition of χ^2 distribution, $2\sum_{i=1}^n X_i/\mu$ follows χ^2 distribution with $2n$ degrees of freedom. Consider any pair of numbers $0 \le q_1 \le 1$ and $0 \le q_2 \le 1$ such that $q_2 - q_2 = \gamma$. For example, $q_1 = (1-\gamma)/2$ and $q_2 = (1+\gamma)/2$. Let c_1 and c_2 be $1/2$ times the q_1 and q_2 quantiles of χ^2_{2n}. It now follows that

$$\Pr(\frac{1}{c_2}\sum_{i=1}^n X_i < \mu < \frac{1}{c_1}\sum_{i=1}^n X_i) = \gamma.$$

Supplementary Exercises 8.9

6. Suppose that $X_1, \ldots, X_n$ form a random sample from an unknown probability distribution P on the real line. Let A be a given subset of the real line, and let $\theta = P(A)$. Construct an unbiased estimator of θ, and specify its variance.

solution Let $\hat{\theta}_n$ be the proportion of the n observations that lie in the set A. Since each observation has probability θ of lying in A, the observations can be thought of as forming n Bernoulli trials, each with probabiltiy θ of success. Hence, $E(\hat{\theta}_n) = \theta$ and $\mathrm{Var}(\hat{\theta}_n) = \theta(1-\theta)/n$.

8. Suppose that $X_1, \ldots, X_{n+1}$ form a random sample from a normal distribution, and let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $T_n = [\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2]^{1/2}$. Determine the value of a constant k such that the random variable $k(X_{n+1} - \bar{X}_n)/T_n$ will have a t distribution.

solution $X_{n+1} - \bar{X}_n$ has the normal distribution with mean 0 and variance $(1+1/n)\sigma^2$. Hence, the distribution of $[n/(n+1)]^{1/2}(X_{n+1} - \bar{X}_n)/\sigma$ is a standard normal distribution. Also, nT_n^2/σ^2 has an independent χ^2 distribution with $n-1$ degrees of freedom. Thus, the following ratio has the t distribution with $n-1$ degrees of freedom. That is,

$$\frac{\left(\frac{n}{n+1}\right)^{1/2}(X_{n+1} - \bar{X}_n)/\sigma}{\left[\frac{nT_n^2}{(n-1)\sigma^2}\right]^{1/2}} = \left(\frac{n-1}{n+1}\right)^{1/2}\frac{X_{n+1} - \bar{X}_n}{T_n}$$

It can be seen that $k = [(n-1)/(n+1)]^{1/2}$.

10. Suppose that $X_1, \ldots, X_n$ form a random sample from the normal distribution with mean μ and variance σ^2, where both μ and σ^2 are unknown. A confidence interval for μ is to be constructed with confidence coefficient 0.90. Determine the smallest value of n such that the expected squared length of this interval will be less than $\sigma^2/2$.

solution The endpoints of the confidence interval are $\bar{X}_n - c\sigma'/n^{1/2}$, $\bar{X}_n + c\sigma'/n^{1/2}$. Therefore, $L^2 = 4\sigma'^2/n$. Since as shown in Exercise 8.7.6 $E(\sigma'^2) = \sigma^2$, the expected squared length of the confidence interval is $E(L^2) = 4c^2\sigma^2/n$, where c is found from the table of the t distribution with $n-1$ degrees of freedom in the back of the book under the .95 column (to give probability .90 between $-c$ and c). We must compute the value of $4c^2/n$ for various values of n and see when it is less than $1/2$. For $n = 23$, it is found that $c_{22} = 1.717$ and the coefficient of σ^2 in $E(L^2)$ is $4(1.717)^2/23 = .512$. For $n = 24$, $c_{23} = 1.714$ and the coefficient of σ^2 is $4(1.714)^2/24 = .490$. Hence, $n = 24$ is the required value.

12. Suppose that $X_1, \ldots, X_n$ form a random sample from the normal distribution with unknown mean μ and unknown variance σ^2. Construct an upper confidence limit $U(X_1, \ldots, X_n)$ for σ^2 such that

$$\Pr[\sigma^2 < U(X_1, \ldots, X_n)] = 0.99.$$

solution Let c denote the .01 quantile of the χ^2 distribution with $n-1$ degress of freedom; i.e., $\Pr(V < c) = .01$ if V has the specified χ^2 distribution. Therefore,

$$\Pr(\frac{S_n^2}{\sigma^2} > c) = .99$$

or, equivalently,

$$\Pr(\sigma^2 < S_n^2/c) = .99.$$

Hence, $U = S_n^2/c$.

14. Suppose that $X_1, \ldots, X_n$ form a random sample from the Poisson distribution with unknown mean θ, and let $Y = \sum_{i=1}^n X_i$.

a. Dtermine the value of a constant c such that the estimator e^{-cY} is an unbiased estimator of $e^{-\theta}$.
b. Use the information inequality to obtain a lower bound for the variance of the unbiased estimator found in part (a).

solution (a) Since Y has a Poisson distribution with mean $n\theta$, it follows that

$$\begin{aligned} E(e^{-cY}) &= \sum_{y=0}^{\infty} \frac{e^{-cy}e^{-n\theta}(n\theta)^y}{y!} = e^{-n\theta}\sum_{y=0}^{\infty}\frac{(n\theta e^{-c})^y}{y!} \\ &= e^{-n\theta}e^{n\theta e^{-c}} = e^{n\theta[e^{-c}-1]}. \end{aligned}$$

Setting this expectation to $e^{-\theta}$, we get $c = \log(\frac{n}{n-1})$.

(b)

$$\begin{aligned} & f(x|\theta) = \frac{e^{-\theta}\theta^x}{x!} \\ & \log f(x|\theta) = -\theta + x\log(\theta) - \log(x!) \\ & \frac{d\log f(x|\theta)}{d\theta} = \frac{x}{\theta} - 1 \\ & I(\theta) = \mathrm{Var}(\frac{d\log f(X|\theta)}{d\theta}) = \frac{\mathrm{Var}(X)}{\theta^2} = \frac{1}{\theta}. \end{aligned}$$

Since $m(\theta) = e^{-\theta}$, $m'(\theta) = -e^{-\theta}$. Hence, by Cramér-Rao bound,

$$\mathrm{Var}(e^{-cY}) \geq \frac{\theta e^{-2\theta}}{n}.$$

Additional Problem

1. Consider an i.i.d. sample from Uniform$(0,\theta)$, $\theta > 0$ of size n. Determine the dependency of the MSE of the estimator $2\bar{X}$ on the sample size n. Is this estimator consistent for θ?

solution By $E(2\bar{X}) = \theta$, it is unbiased. It follows that

$$\mathrm{MSE}(2\bar{X}) = \mathrm{Var}(2\bar{X}) = \frac{4}{n}\mathrm{Var}(X_1) = \frac{4}{n}\frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

It converges to 0 as $n \to \infty$, so the estimator converges to θ in quadratic mean and is therefore consistent.

2. Show that the choice $\gamma_1 = (1-\gamma)/2$, $\gamma_2 = (1+\gamma)/2$ gives the shortest length interval in (b) p.51 of the lecture notes, and therefore is the best choice.

solution Denote the p.d.f of t distribution by $f(x)$. First note it is sysmetric at 0 and decreasing when $x > 0$. The length of the interval is $(t_{\gamma_2,n-1} - t_{\gamma_1,n-1})\frac{\tilde{\sigma}}{\sqrt{n}}$. Then we only need to prove given $\gamma_2 - \gamma_1 = \gamma$, $\gamma_1^* = (1-\gamma)/2$, $\gamma_2^* = (1+\gamma)/2$ minimizes $t_{\gamma_2,n-1} - t_{\gamma_1,n-1}$. W.L.O.G let's assume $\gamma_1 > \gamma_1^*$, $\gamma_2 > \gamma_2^*$. Then we have

$$\int_{t_{\gamma_1,n-1}}^{t_{\gamma_2,n-1}} f(x)dx = \gamma = \int_{t_{\gamma_1^*,n-1}}^{t_{\gamma_2^*,n-1}} f(x)dx \Rightarrow \int_{t_{\gamma_1^*,n-1}}^{t_{\gamma_1,n-1}} f(x)dx = \int_{t_{\gamma_2^*,n-1}}^{t_{\gamma_2,n-1}} f(x)dx.$$

By $t_{\gamma_2^*,n-1} = -t_{\gamma_1^*,n-1}$, $f(y) < f(t_{\gamma_2^*,n-1}) = f(t_{\gamma_1^*,n-1}) < f(x)$ for $y \in (t_{\gamma_2^*,n-1}, t_{\gamma_2,n-1})$ and $x \in (t_{\gamma_1^*,n-1}, t_{\gamma_1,n-1})$. This implies $\int_{t_{\gamma_1^*,n-1}}^{t_{\gamma_1,n-1}} f(x)dx > (t_{\gamma_1,n-1} - t_{\gamma_1^*,n-1})f(t_{\gamma_1^*,n-1})$ and $\int_{t_{\gamma_2^*,n-1}}^{t_{\gamma_2,n-1}} f(x)dx < (t_{\gamma_2,n-1} - t_{\gamma_2^*,n-1})f(t_{\gamma_2^*,n-1})$. Combined with $\int_{t_{\gamma_1^*,n-1}}^{t_{\gamma_1,n-1}} f(x)dx = \int_{t_{\gamma_2^*,n-1}}^{t_{\gamma_2,n-1}} f(x)dx$ and $f(t_{\gamma_2^*,n-1}) = f(t_{\gamma_1^*,n-1})$ we get $t_{\gamma_2,n-1} - t_{\gamma_2^*,n-1} > t_{\gamma_1,n-1} - t_{\gamma_1^*,n-1}$, which gives $t_{\gamma_2,n-1} - t_{\gamma_1,n-1} > t_{\gamma_2^*,n-1} - t_{\gamma_1^*,n-1}$.

3. Consider an i.i.d. sample from Uniform$(-\theta,\theta)$, $\theta > 0$, of size n. Show that the estimator $\sqrt{|X_{(1)}X_{(n)}|}$ is consistent for θ.

solution For any $\epsilon > 0$

$$\Pr(|X_{(1)} + \theta| > \epsilon) = \Pr(X_{(1)} > -\theta + \epsilon) = \prod_{i=1}^{n} \Pr(X_i > -\theta + \epsilon) = (1 - \frac{\epsilon}{2\theta})^n \to 0$$
$$\Pr(|X_{(n)} - \theta| > \epsilon) = \Pr(X_{(n)} < \theta - \epsilon) = \prod_{i=1}^{n} \Pr(X_i < \theta - \epsilon) = (1 - \frac{\epsilon}{2\theta})^n \to 0.$$

So we have $X_{(1)} \to -\theta$ in probability and $X_{(n)} \to \theta$ in probability and therefore $\sqrt{|X_{(1)}X_{(n)}|} \to \theta$ in probability.