Tutorial Sheet 3
Question 1

1. Prove or disprove : The sample standard deviation S of a random sample of size n is an unbiased estimator of the population standard deviation σ. (Hint : Any random variable has nonzero variance.)

Note that $V(S) = E(S^2) - \left(E(S)\right)^2$.
If S is an unbiased estimator of σ, then $E(S) = \sigma$.
But $E(S^2) = \sigma^2$. Hence $V(S) = 0$.
This can't hold, as S is a random variable.
Thus statement is disproved.

Question 2

2. A population X has $E[X] = 0, E[X^2] = 1$ and $E[X^4] = 4$. For a random sample of size 2, find the standard error of S^2.

Solution : Standard error of $S^2 = V(S^2) = \mathrm{E}(S^4) - \mathrm{E}(S^2)^2$

$$S^2 = \frac{2\sum_{i=1}^{2} {X_i}^2 - \left(\sum_{i=1}^{2} X_i\right)^2}{2}.$$

$$= \frac{2(X_1^2 + X_2^2) - (X_1^2 + X_2^2 + 2X_1X_2)}{2} = \frac{(X_1 - X_2)^2}{2}.$$

$$\therefore S^4 = \frac{X_1^4 - 4X_1^3X_2 + 6X_1^2X_2^2 - 4X_1X_2^3 + X_2^4}{4}.$$

By independence of X_1, X_2, $E(S^4) = \frac{2E(X^4)+6\left(E(X^2)\right)^2}{4} = \frac{7}{2}.$

Solution 2

$$(E(S^2))^2 = (\sigma^2)^2 = 1.$$

$$\therefore \text{Standard error of } S^2 = \frac{7}{2} - 1 = \frac{5}{2}$$

Question 3

The deaths of the patients with a disease occur in Poisson process with the average rate of α per day. Since the start of monitoring, the 1st death occurs after 11 hours, 2nd after 20 hours, 3rd after 26 hours and the 4th after 30 hours. Using this information, find the estimate for α using (a) method of moments, (b) method of maximum likelihood.

Solution : Let T = the time in hours between successive deaths due to the disease. Then its pdf is

$$f(t) = \begin{cases} \dfrac{\alpha e^{-\alpha t/24}}{24}, & \text{if } t > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Solution 3

(a) $E(T) = \frac{1}{\frac{\alpha}{24}} \Longrightarrow \alpha = \frac{24}{E(T)}$, therefore

method of moments estimate $\hat{\alpha} = \frac{24}{M_1} = \frac{96}{30} = 3.2.$

(b) Likelihood function is

$$L(\alpha) = \frac{\alpha^4}{24^4} e^{-\frac{30\alpha}{24}}.$$

$$L'(\alpha) = \frac{\alpha^3 e^{-\frac{30\alpha}{24}}}{24^5}(96 - 30\alpha) = 0 \Longrightarrow \hat{\alpha} = \frac{96}{30}$$

Question 4

1. Suppose the population X is a continuous uniform distribution on the interval $[a, 2a]$ where $a > 0$ is the parameter of the distribution. Find the maximum likelihood estimator of a using a random sample of size n.

Solution : $f(x) = \begin{cases} \frac{1}{a}, & a < x < 2a, \\ 0, & \text{elsewhere} \end{cases}$

$a < x_i < 2a$ for all $i = 1, 2, \dots, n$ if and only if $\frac{\max x_i}{2} < a < \min x_i$.

Yo

Solution 4

$$\therefore L(a) = \begin{cases} \dfrac{1}{a^n}, & \dfrac{\max x_i}{2} \leq a \leq \min x_i \\ 0, & \text{elsewhere.} \end{cases}$$

As likelihood function is decreasing on $\left[\frac{\max x_i}{2}, \min x_i\right]$, the estimator of a is where maximum value of $L(a)$occurs, i.e. $\hat{a} = \frac{\max x_i}{2}$.

Question 5

$$f(x|\theta) = \begin{cases} 1, & \text{for} \quad \theta \le x \le \theta + 1 \\ 0, & \text{otherwise} \end{cases}$$

We will see that the MLE for θ is not unique.

Proof: In this example, the likelihood function is

$$L(\theta) = \begin{cases} 1, & \text{for} \quad \theta \le x_i \le \theta + 1 \quad (i = 1, \cdots, n) \\ 0, & \text{otherwise} \end{cases}$$

The condition that $\theta \le x_i$ for $i = 1, \cdots, n$ is equivalent to the condition that $\theta \le \min(x_1, \cdots, x_n)$. Similarly, the condition that $x_i \le \theta + 1$ for $i = 1, \cdots, n$ is equivalent to the condition that $\theta \ge \max(x_1, \cdots, x_n) - 1$. Therefore, we can rewrite the the likelihood function as

$$L(\theta) = \begin{cases} 1, & \text{for} \quad \max(x_1, \cdots, x_n) - 1 \le \theta \le \min(x_1, \cdots, x_n) \\ 0, & \text{otherwise} \end{cases}$$

Thus, we can select any value in the interval $[\max(x_1, \cdots, x_n) - 1, \min(x_1, \cdots, x_n)]$ as the MLE for θ. Therefore, the MLE is not uniquely specified in this example.

Question 6 (a)

$$E(X) = \int_{-\infty}^{\infty} x f(x|\sigma)dx = \int_{-\infty}^{\infty} x \frac{1}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right) dx = 0.$$

Thus, if we try to solve equation $E(X) = \overline{X}$, we will not get the estimator, because $E(X)$ does not contain the unknown parameter σ.

Now, let us calculate the second order theoretical moment, we have $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x|\sigma)dx = \int_{-\infty}^{\infty} x^2 \frac{1}{2\sigma} \exp\left(-\frac{|x|}{\sigma}\right) dx$

$$\begin{aligned}
&= \int_0^{\infty} x^2 \frac{1}{\sigma} \exp\left(-\frac{|x|}{\sigma}\right) dx = \sigma^2 \int_0^{\infty} y^2 e^{-y} dy \quad (\text{Let } x = \sigma y) \\
&= -\sigma^2 \int_0^{\infty} y^2 de^{-y} = -\sigma^2 y^2 e^{-y} \big|_0^{\infty} + \sigma^2 \int_0^{\infty} 2ye^{-y} dy \\
&= 0 - 2\sigma^2 \int_0^{\infty} yde^{-y} = -2\sigma^2 ye^{-y} \big|_0^{\infty} + 2\sigma^2 \int_0^{\infty} e^{-y} dy \\
&= 0 + 2\sigma^2 e^{-y} \big|_0^{\infty} = 2\sigma^2
\end{aligned}$$

The second order sample moment is:

$$m_2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2.$$

Solving the equation $\mu_2 = m_2$, i.e. $2\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$, we can obtain the estimate of σ:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{2n}}$$

From this example, we can see that we also want to choose a k, such that $E(X^k)$ is a function of the unknown parameters. Only when it contains the unknown parameters, can we solve the equation. This is our second rule for selecting k.

Question 6 (b)

Solution: The log-likelihood function is

$$l(\sigma) = \sum_{i=1}^{n}\left[-\log 2 - \log\sigma - \frac{|X_i|}{\sigma}\right]$$

Let the derivative with respect to θ be zero:

$$l'(\sigma) = \sum_{i=1}^{n}\left[-\frac{1}{\sigma} + \frac{|X_i|}{\sigma^2}\right] = -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}|X_i|}{\sigma^2} = 0$$

and this gives us the MLE for σ as

$$\hat{\sigma} = \frac{\sum_{i=1}^{n}|X_i|}{n}$$

Again this is different from the method of moment estimation which is

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}X_i^2}{2n}}$$

Question 7

Suppose the true average growth μ of one type of plant during a 1-year period is identical to that of a second type, but the variance of growth for the first type is σ^2, whereas for the second type, the variance is $4\sigma^2$. Let $X_1, X_2, \ldots, X_m$ be *m* independent growth observations on the first type [so $E[X_i] = \mu, V[X_i] = \sigma^2$], and let $Y_1, Y_2, \ldots, Y_n$ be *n* independent growth observations on the second type [so $E[Y_i] = \mu, V[Y_i] = 4\sigma^2$].

a. Show that for any δ between 0 and 1, the estimator $\hat{\mu} = \delta\bar{X}$+(1-$\delta$)$\bar{Y}$ is unbiased for μ.
b. For fixed *m* and *n*, compute $V[\hat{\mu}]$ and then find the value of δ that minimizes $V[\hat{\mu}]$.

Solution 7

(a) $\mathrm{E}(\hat{\mu}) = E(\delta\bar{X} + (1-\delta)\bar{Y})$

$$= \delta E(\bar{X}) + (1-\delta)E(\bar{Y})$$

$$= \delta\mu + (1-\delta)\mu = \mu.$$

Therefore $\hat{\mu}$ is an unbiased estimator of μ.

Equating its derivative with respect to δ to 0,

You

Solution 7

(a) $E(\hat{\mu}) = E(\delta\bar{X} + (1-\delta)\bar{Y})$

$$= \delta E(\bar{X}) + (1-\delta)E(\bar{Y})$$

$$= \delta\mu + (1-\delta)\mu = \mu.$$

Therefore $\hat{\mu}$ is an unbiased estimator of μ.

(b) $V(\hat{\mu}) = \delta^2 V(\bar{X}) + (1-\delta)^2 V(\bar{Y})$ (as $\bar{X}, \bar{Y}$ are indep.)

$$= \left(\frac{1}{m}\delta^2 + 4\frac{1}{n}(1-\delta)^2\right)\sigma^2.$$

Equating its derivative with respect to δ to 0,

$$\frac{2\delta\sigma^2}{m} + \frac{8(1-\delta)\sigma^2}{n} = 0. \qquad \therefore \delta = \frac{4m}{4m+n}.$$

Question 8

The **mean squared error** of an estimator $\hat{\theta}$ (of θ) is $MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$. If $\hat{\theta}$ is unbiased, then $MSE(\hat{\theta}) = V(\hat{\theta})$ but in general $MSE(\hat{\theta}) = V(\hat{\theta}) + (bias)^2$. Consider the estimator $\hat{\sigma}^2 = KS^2$ where $S^2 =$ sample variance. What value of K minimizes the mean squared error of this estimator when the population distribution is normal?

[*Hint:* It can be shown that $E\left((S^2)^2\right) = \frac{(n+1)\sigma^4}{(n-1)}$. In general, it is difficult to find $\hat{\theta}$ to minimize $MSE(\hat{\theta})$, which is why we look only at unbiased estimators and minimize $V(\hat{\theta})$.]

Using the observations 0.5, 1, -1.5, 2 from the standard normal distribution, find an unbiased estimate, the maximum likelihood estimate and the estimate KS^2 for the value of K which minimizes the mean squared error.

Solution 8

$$V(KS^2) = K^2V(S^2) = K^2\left[E\left((S^2)^2\right) - \left(E(S^2)\right)^2\right]$$

$$= K^2\left[\frac{n+1}{n-1}\sigma^4 - \sigma^4\right]$$

$$= \frac{2K^2}{n-1}\sigma^4.$$

$$Bias(KS^2) = \mathrm{E}(KS^2) - \sigma^2 = (K-1)\sigma^2.$$

$$\therefore MSE(KS^2) = \frac{2K^2}{n-1}\sigma^4 + (K-1)^2\sigma^4.$$

To minimize mean squared error, solve $\frac{dMSE(KS^2)}{dK} = 0$.

$$K = \frac{n-1}{n+1}.$$