IE 605: Engineering Statistics

Solutions of tutorial 6

Solution 1

Let X_1, X_2, X_3, X_4, X_5 be a sample from $N(0, 4)$.

To find: $\mathbb{P}\left\{\sum_{i=1}^{5} X_i^2 \geq 5.75\right\}$.

$$\frac{X_i}{2} \sim N(0,1) \implies \frac{X_i^2}{4} \sim \chi_1^2$$

$$\sum_{i=1}^{5} \frac{X_i^2}{4} \sim \chi_5^2 \qquad \text{since } X_i's \text{ are independent}$$

$$\mathbb{P}\left\{\sum_{i=1}^{5} X_i^2 \geq 5.75\right\} = \mathbb{P}\left\{\frac{\sum_{i=1}^{5} X_i^2}{4} \geq \frac{5.75}{4}\right\}$$

$$= 1 - \mathbb{P}\left\{\chi_5^2 \leq 1.4375\right\} = 0.07983235$$

Solution 2

Let $X_1, X_2, \ldots, X_n$ be a random sample from $Poisson(\lambda)$.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The expectation of is given by,

$$\mathbb{E}\left[\bar{X}\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[X_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \lambda \qquad \text{since, } x_i \sim Poisson(\lambda) \implies \mathbb{E}\left[X_i\right] = \lambda$$

$$\mathbb{E}\left[\bar{X}\right] = \lambda.$$

$$var(\bar{X}) = var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$$

$$
\begin{aligned}
&= \frac{1}{n^2} var\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n} var(X_i) + \sum_{i,j=1, i\neq j}^{n} cov(X_i, X_j)\right) \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n} var(X_i)\right) \text{ since } X_i's \text{ are independent} \\
&= \frac{1}{n^2}\left(\sum_{i=1}^{n} \lambda\right) \\
&= \frac{\lambda}{n} \\
S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\
\mathbb{E}\left[S^2\right] &= \lambda \qquad \text{Check!} \\
var(S^2) &= \frac{1}{n}\left(\mu_4 - \frac{(n-3)\mu_2^2}{n-1}\right) \qquad \text{Check!}
\end{aligned}
$$

For Poisson distribution, $\mu_2 = \lambda$ and $\mu_4 = \lambda(3\lambda + 1)$. Putting these values we get

$$
\begin{aligned}
var(S^2) &= \frac{1}{n}\left(\frac{\lambda(3\lambda+1)(n-1) - (n-3)\lambda^2}{n-1}\right) \\
&= \frac{1}{n}\left(\frac{3(n-1)\lambda^2 + (n-1)\lambda - (n-3)\lambda^2}{n-1}\right) \\
var(S^2) &= \frac{1}{n}\left(\lambda + \frac{2n\lambda^2}{n-1}\right) \\
&> var(\bar{X}).
\end{aligned}
$$

Solution 3

For $n = 2$, the p.d.f. of $f_{X_{(1)}}(x)$ is given by,

$$
f_{X_{(1)}}(x) = \frac{1}{\beta(1,2)}[1 - F(x)]f(x); \quad -\infty < x < \infty
$$

where, $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-x^2}{2\sigma^2}}$

$$
\begin{aligned}
\mathbb{E}\left[X_{(1)}\right] &= \int_{-\infty}^{\infty} x.f_{X_{(1)}}(x)dx(1) \\
&= 2\int_{-\infty}^{\infty} 1 - F(x)]f(x)dx
\end{aligned}
$$

We have $\log f(x) = -\log(\sqrt{2\pi}\sigma) - \frac{x^2}{2\sigma^2}$

Differentiating w.r.t. x we get,

$$\frac{f'(x)}{f(x)} = -\frac{x}{\sigma^2}$$

$$\Rightarrow \int x f(x) dx = -\sigma^2 \int f'(x) dx = -\sigma^2 f(x) (2)$$

Integrating (1) by parts and using (2), we get

$$\begin{aligned}\mathbb{E}\left[X_{(1)}\right] &= 2.[(1-F(x))(-\sigma^2 f(x))]_{-\infty}^{\infty} \\ &\quad - 2\int_{-\infty}^{\infty} (-\sigma^2 f(x))(-f(x)) dx \\ &= -2\sigma^2 \int_{-\infty}^{\infty} [f(x)]^2 dx = -\frac{1}{\pi}\int_{-\infty}^{\infty} e^{\frac{-x^2}{\sigma^2}} dx \\ &= -\frac{1}{\pi}\frac{\sqrt{\pi}}{(1/\sigma)} \quad \left(\text{Since} \int_{-\infty}^{\infty} e^{-a^2x^2} dx = \frac{\sqrt{\pi}}{a}\right) \\ &= -\sigma/\sqrt{\pi}.\end{aligned}$$

Solution 4

The probability distribution of X (or Y or Z) is

$$f(x) = \begin{cases} \frac{1}{a} & \text{if } 0 < x < a \\ 0 & \text{otherwise.} \end{cases}$$

Thus the cumulative distribution of function of $f(x)$ is given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{a} & \qquad\qquad \text{if } 0 < x < a \\ 1 & \text{if } x \geq a. \end{cases}$$

Since $W = min\{X, Y, Z\}$, W is the first order statistic of the random sample X, Y, Z. Thus, the density function of W is given by

$$\begin{aligned} g(w) &= \frac{3!}{0!1!2!}[F(w)]^0 f(w)[1-F(w)]^2 \\ &= 3f(w)[1-F(w)]^2 \\ &= 3\left(1-\frac{w}{a}\right)^2\left(\frac{1}{a}\right) \end{aligned}$$

$$= \frac{3}{a}\left(1-\frac{w}{a}\right)^2$$

Thus, the pdf of W is given by

$$g(w) = \begin{cases} \frac{3}{a}\left(1-\frac{w}{a}\right)^2 & \text{if } 0 < w < a \\ 0 & \text{otherwise.} \end{cases}$$

The expected value of W is

$$\begin{aligned} E\left[\left(1-\frac{W}{a}\right)^2\right] &= \int_0^a \left(1-\frac{w}{a}\right)^2 g(w)dw \\ &= \int_0^a \left(1-\frac{w}{a}\right)^2 \frac{3}{a}\left(1-\frac{w}{a}\right)^2 dw \\ &= \int_0^a \frac{3}{a}\left(1-\frac{w}{a}\right)^4 dw \\ &= \frac{3}{5}\left[\left(1-\frac{w}{a}\right)^5\right]_0^a \\ &= \frac{3}{5} \end{aligned}$$

Solution 5

To find the distribution of W, we need the joint distribution of the random variable $\left(X_{(n)}, X_{(1)}\right)$. The joint distribution of $\left(X_{(n)}, X_{(1)}\right)$ is given by

$$h(x_1, x_n) = n(n-1)f(x_1)f(x_n)[F(x_n) - F(x_1)]^{n-2}$$

where $x_n \geq x_1$ and $f(x)$ is the probability density function of X. To determine the probability distribution of the sample range W, we consider the transformation

$$U = X_{(1)}, W = X_{(n)} - X_{(1)}$$

which has an inverse

$$X_{(1)} = U, X_{(n)} = U + W.$$

The Jacobian of this transformation is

$$J = \det \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = 1$$

Hence the joint density of (U, W) is given by

$$g(u, w) = |J|h(x_1, x_n) = n(n-1)f(u)f(u+w)[F(u+w) - F(u)]^{n-2}$$

where $w \geq 0$. Since $f(u)$ and $f(u+w)$ are simultaneously nonzero if $0 \leq u \leq 1$ and $0 \leq u + w \leq 1$. Hence $f(u)$ and $f(u+w)$ are simultaneously nonzero if

$0 \leq ule1 - w$. Thus, the probability of W is given by

$$\begin{aligned}
j(w) &= \int_{-\infty}^{\infty} g(u, w) du \\
&= \int_{-\infty}^{\infty} n(n-1) f(u) f(u+w) [F(u+w) - F(u)]^{n-2} du \\
&= n(n-1) w^{n-2} \int_{0}^{1-w} du \\
&= n(n-1)(1-w) w^{n-2}
\end{aligned}$$

where $0 \leq w \leq 1$

Solution 6

Let X be a random variable with an $F_{p,q}$ distribution, i.e. $X = \frac{U/p}{V/q}$, where U and V are two independent chi-square variates with p and q d.f. respectively.

We will derive the p.d.f. of X.

$$X = \frac{U/p}{V/q} \Rightarrow \frac{p}{q} X = \frac{U}{V}$$

$\frac{U}{V}$ being the ratio of two independent chi-square variates with p and q d.f. respectively is a $\beta_2(p/2, q/2)$ variate (Beta- Second kind) [**CHECK**]. Hence the probability density function of X is given by

$$f_X(x) = \frac{(\frac{p}{q})^{\frac{p}{2}}}{B(\frac{p}{2}, \frac{q}{2})} \frac{x^{(p/2)-1}}{\left[1 + \frac{p}{q} x\right]^{\frac{(p+q)}{2}}}, \quad 0 \leq x < \infty$$

The mean and variance of X.

$$\begin{aligned}
\mu_r' = \mathbb{E}\left[X^r\right] &= \int\limits_0^{\infty} x^r f_X(x) dx \\
&= \frac{(\frac{p}{q})^{\frac{p}{2}}}{B(\frac{p}{2}, \frac{q}{2})} \int\limits_0^{\infty} x^r \frac{x^{(p/2)-1}}{\left[1 + \frac{p}{q} x\right]^{\frac{(p+q)}{2}}} dx
\end{aligned}$$

To evaluate the integral, put $\frac{p}{q} x = y$ so that $dx = \frac{q}{p} dy$

$$\mu_r' = \frac{(\frac{p}{q})^{\frac{p}{2}}}{B(\frac{p}{2}, \frac{q}{2})} \int\limits_0^{\infty} \frac{\left(\frac{q}{p} y\right)^{r+(p/2)-1}}{[1+y]^{\frac{(p+q)}{2}}} \frac{q}{p} dy$$

$$
\begin{aligned}
&= \frac{(\frac{q}{p})^r}{B(\frac{p}{2},\frac{q}{2})} \int_0^\infty \frac{y^{r+(p/2)-1}}{[1+y]^{(p/2+r)+(q/2-r)}} dy \\
&= \frac{(\frac{q}{p})^r}{B(\frac{p}{2},\frac{q}{2})} B(r+p/2, q/2-r) \quad \text{Check}
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\mu_1'(\text{ mean }) &= \frac{q}{q-2}, \quad q>2 \quad \text{Verify} \\
\mu_2' &= \frac{q^2(p+2)}{p(q-2)(q-4)}, \quad q>4 \quad \text{Verify} \\
\mu_2(\text{ Variance }) &= \frac{2q^2(q+p-2)}{p(q-2)^2(q-4)}, \quad q>4 \quad \text{Verify}
\end{aligned}
$$

To show that $\frac{\frac{p}{q}X}{1+\frac{p}{q}X}$ has a beta distribution with parameters $p/2$ and $q/2$.

Note that $Z = \frac{p}{q}X$

has the pdf of Z as

$$
f_Z(z) = \frac{\Gamma[(p+q)/2]}{\Gamma(p/2)\Gamma(q/2)} \frac{z^{(p/2)-1}}{(1+z)^{(p+q)/2}}, \quad z>0
$$

If $u = z/(1+z)$, then $z = u/(1-u)$, $dz = (1-u)^{-2}du$, and the pdf of $U = Z/(1+Z)$ is

$$
\begin{aligned}
g_U(u) &= \frac{\Gamma[(p+q)/2]}{\Gamma(p/2)\Gamma(q/2)} \left(\frac{u}{1-u}\right)^{(p/2)-1} \frac{1}{(1-u)^{-(p+q)/2}} \frac{1}{(1-u)^2} \\
&= \frac{\Gamma[(p+q)/2]}{\Gamma(p/2)\Gamma(q/2)} u^{(p/2)-1}(1-u)^{(q/2)-1}, u>0 \\
&= \frac{1}{Beta(p/2,q/2)} u^{(p/2)-1}(1-u)^{(q/2)-1}, u>0 \\
\implies \frac{\frac{p}{q}X}{1+\frac{p}{q}X} &\sim Beta(p/2,q/2).
\end{aligned}
$$

Solution 7

We have,

$$
\begin{aligned}
f(x) &= 1; \quad 0 \le x \le 1 \\
F(x) &= \mathbb{P}\{X \le x\} = \int_0^x f(u)du = x
\end{aligned}
$$

Let $n = 2m + 1$ (odd), where m is a positive integer ≥ 1. Then median observation is $X_{(m+1)}$. Taking $r = (m + 1)$ in $X_{(r)}$, the pdf of median $X_{(m+1)}$ is given by,

$$
\begin{aligned}
f_{m+1}(x) &= \frac{1}{\beta(m+1, m+1)} x^m (1-x)^m \\
\mathbb{E}\left[X_{(m+1)}\right] &= \frac{1}{\beta(m+1, m+1)} \int_0^1 x.x^m (1-x)^m dx \\
&= \frac{\beta(m+2, m+1)}{\beta(m+1, m+1)} \\
&= \frac{1}{2} \qquad \text{Check} \\
\mathbb{E}\left[X^2_{(m+1)}\right] &= \int_0^1 x^2 f_{m+1}(x) dx \\
&= \frac{m+2}{2(2m+3)} \qquad \text{Check}
\end{aligned}
$$

Therefore,

$$
Var(X_{(m+1)}) = \mathbb{E}\left[X^2_{(m+1)}\right] - [\mathbb{E}\left[X_{(m+1)}\right]]^2 = \frac{1}{4(n+2)} \quad \text{Check.}
$$

Solution 8

To show: Sample standard deviation is not unbiased, but is consistent.

Let $X_1, \ldots, X_n$ be n random sample drawn from a population with mean μ and population variance σ^2. We define,

Sample mean:= $\bar{X} = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$.

Sample variance:= $S_n^2 = \frac{1}{n-1} (\sum\limits_{i=1}^{n} X_i - \bar{X})^2$

Sample standard deviation:= $S_n = \sqrt{\frac{1}{n-1} (\sum\limits_{i=1}^{n} X_i - \bar{X})^2} = \sqrt{S_n^2}$.

We know the sample variance defined S_n^2 is an unbiased estimator of population variance σ^2 (CHECK!) and the sample standard deviation is given as $S_n = \sqrt{S_n^2}$.

- To check: Sample standard deviation is an unbiased estimator of σ or not.

  We know,

$$
\begin{aligned}
& Var(S_n) \geq 0 \\
\implies & \mathbb{E}\left[S_n^2\right] - (\mathbb{E}\left[S_n\right])^2 \geq 0 \\
\implies & \mathbb{E}\left[S_n^2\right] \geq (\mathbb{E}\left[S_n\right])^2 \\
\implies & \mathbb{E}\left[S_n\right] \leq \sqrt{\mathbb{E}\left[S_n^2\right]} \qquad \text{we consider positive part only since } S_n \geq 0. \\
\implies & \mathbb{E}\left[S_n\right] \leq \sigma
\end{aligned}
$$

This implies that sample standard deviation S_n is a biased estimator of population standard deviation.

- To check: Sample standard deviation is an consistent estimator of σ or not.

  First we will show that S_n^2 is a consistent estimator of σ^2.

$$\mathbb{P}\left\{|S_n^2 - \sigma^2| \geq \epsilon\right\} \leq \frac{\mathbb{E}\left[(S_n^2 - \sigma^2)^2\right]}{\epsilon^2} = \frac{Var(S_n^2)}{\epsilon^2}$$

  If $var(S_n^2) \to 0$ as $n \to \infty$ then $\frac{Var(S_n^2)}{\epsilon^2} \to 0$

$$\implies \mathbb{P}\left\{|S_n^2 - \sigma^2| \geq \epsilon\right\} \to 0 \text{ as } n \to \infty$$
$$\implies S_n^2 \to \sigma^2 \text{ in probability.}$$

  Therefore, S_n^2 is a consistent estimator of σ^2.

  By invariance property of consistent estimator which states that:

  If T is a consistent estimator of θ and if f is a continuous function then $f(T)$ is a consistent estimator of $f(\theta)$.

  So by invariance property we can say that if S_n^2 is a consistent estimator of σ^2 then $S_n = \sqrt{S_n^2}$ is a consistent estimator of σ as square-root is a continuous function.

Solution 9

$$\begin{aligned}
P(Z > z) &= \sum_{x=1}^{\infty} P(Z > z|x)P(X = x) \\
&= \sum_{x=1}^{\infty} P(U_1 > z, ..., U_x > z|x)P(X = x) \\
&= \sum_{x=1}^{\infty} \prod_{i=1}^{x} P(U_i > z)P(X = x) \quad \text{(by independence of the } U_i\text{'s)} \\
&= \sum_{x=1}^{\infty} P(U_i > z)^x P(X = x) \\
&= \sum_{x=1}^{\infty} (1 - z)^x \frac{1}{(e-1)x!} \\
&= \frac{1}{(e-1)} \sum_{x=1}^{\infty} \frac{(1-z)^x}{x!} \\
&= \frac{e^{1-z} - 1}{e - 1}, \quad 0 < z < 1
\end{aligned}$$

Solution 10

The solution for three parts are as follows:

a. $\sum_{i=1}^{3}(\frac{X_i-i}{i})^2 \sim X_3^2$

b. $(\frac{X_i-1}{i})/\sqrt{\sum_{i=1}^{3}(\frac{X_i-i}{i})^2/2} \sim t_2$

c. Square the random variable in part (b).

Solution 11

Use $f_X(x) = 1/\theta, F_X(x) = x/\theta, 0 < x < \theta$. Let $Y = X_{(n)}, Z = X_{(1)}$.

$$\begin{aligned} f_{(Z,Y)}(z,y) &= \frac{n!}{0!(n-2)!0!}\frac{1}{\theta}\frac{1}{\theta}(\frac{z}{\theta})^0(\frac{y-z}{\theta})^{n-2}(1-\frac{y}{\theta})^0 \\ &= \frac{n(n-1)}{\theta^n}(y-z)^{n-2}, \quad 0 < z < y < \theta \end{aligned}$$

Now let $W = Z/Y, Q = Y$. Then $Y = Q, Z = WQ$ and $|J| = q$. Therefore

$$\begin{aligned} f_{(W,Q)}(w,q) &= \frac{n(n-1)}{\theta^n}(q-wq)^{n-2}q \\ &= \frac{n(n-1)}{\theta^n}(1-w)^{n-2}q^{n-1}, \quad 0 < w < 1, 0 < q < \theta \end{aligned}$$

The joint pdf factors into functions of w and q, and, hence, W and Q are independent.

Solution 12

Assume p and q are even. $F_{p,q} = \frac{U/p}{V/q}$ where $U \sim \chi_p^2$ and $V \sim \chi_q^2$ and are independent. As chi-squared distribution with p degrees of freedom is Gamma distributed with parameter $\alpha = p/2$ and $\lambda = 1/2$, we get $U \sim Gamma(p/2, 1/2)$ and $V \sim Gamma(q/2, 1/2)$. As $p/2$ and $q/2$ are integers, we have $\sum_{i=1}^{p/2} X_i \sim Gamma(p/2, 1/2)$ and $\sum_{i=1}^{q/2} Y_i \sim Gamma(p/2, 1/2)$, where $X_1, X_2, \ldots, X_{p/2}, Y_1, Y_2, \ldots, Y_{q/2}$ are iid RVs $\sim Exp(1/2)$. Hence we have

$$F_{p,q} = \frac{q}{p}\frac{\sum_{i=1}^{p/2} X_i}{\sum_{j=1}^{q/2} Y_j}$$

Finally, we know that $X_1, X_2, \ldots, X_{p/2}, Y_1, Y_2, \ldots, Y_{q/2}$ can be generated from uniform random variables. If $Z_1, Z_2, Z_{p/2}, W_1, W_2, W_{q/2}$ are iid $\sim Unif(0, 1)$. We have $X_i = -2\log(1 - Z_i)2$ for all $i = 1, 2, , p/2$ and $Y_j = -2\log(1 - W_j)$ for all $j = 1, 2, , q/2$. Thus we have

$$F_{p,q} = \frac{q}{p} \frac{\sum_{i=1}^{p/2} 2\log(1-Z_i)}{\sum_{j=1}^{q/2} 2\log(1-W_i)}$$