Parametric Statistics-Recitation 4 (Solutions)

Exercise 1

If X has the uniform distribution on the interval $[a, b]$, what is the value of the fifth central moment of X ?

Solution

Since the uniform distribution p.d.f is symmetric with respect to its mean $\mu = \frac{a+b}{2}$, it follows that $E[(X - \mu)^5] = 0$.

Exercise 2

If X has the uniform distribution on the interval $[a, b]$, write a formula for every even central moment of X.

Solution

The mean of X is $\mu = \frac{a+b}{2}$, so the 2kth central moment of X is the mean of $(X - \frac{a+b}{2})^{2k}$. Note that $Y = X - \frac{a+b}{2}$ has a uniform distribution in $[-\frac{b-a}{2}, \frac{b-a}{2}]$. Also if we define $Z = \frac{2Y}{b-a}$, note that Z has a uniform distribution in $[-1, 1]$. So

$$E[Y^{2k}] = \left(\frac{b-a}{2}\right)^{2k} E[Z^{2k}]$$

.
Also

$$E[Z^{2k}] = \int_{-1}^{1} \frac{z^{2k}}{2} dz = \frac{1}{2k+1}$$

So the 2kth central moment of X is

$$E[Y^{2k}] = \left(\frac{b-a}{2}\right)^{2k} \cdot \frac{1}{2k+1}$$

Exercise 3

Suppose that 20 percent of the students who took a certain test were from school A and that the arithmetic average of their scores on the test was 80 . Suppose also that 30 percent of the students were from school B and that the arithmetic average of their scores was 76. Suppose, finally, that the other 50 percent of the students were from school C and that the arithmetic average of their scores was 84 . If a student is selected at random from the entire group that took the test, what is the expected value of her score?

Solution

If X denotes the score of the selected student, then

$$E(X) = E[E(X|School)] = 0.2 \cdot 80 + 0.3 \cdot 76 + 0.5 \cdot 84 = 80.8$$

Exercise 4

Suppose that a person's score X on a mathematics aptitude test is a number in the interval $(0, 1)$ and that his score Y on a music aptitude test is also a number in the interval $(0, 1)$. Suppose also that in the population of all college students in the United States, the scores X and Y are distributed in accordance with the following joint p.d.f.:

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y) & \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) If a college student is selected at random, what predicted value of his score on the music test has the smallest M.S.E.?

(b) What predicted value of his score on the mathematics test has the smallest M.A.E.?

Solution

(a) The prediction is the mean of Y:

$$E(Y) = \int_0^1 \int_0^1 y \cdot \frac{2}{5}(2x+3y)dxdy = \frac{3}{5}$$

(b) The prediction is the median of X. First we find the marginal p.d.f of X:

$$f_1(x) = \int_0^1 \frac{2}{5}(2x+3y)dy = \frac{1}{5}(4x+3)$$

The median m is s.t

$$\int_0^m \frac{1}{5}(4x+3)dx = \frac{1}{2}$$

Therefore, $4m^2 + 6m - 5 = 0$ and $m = \frac{\sqrt{29}-3}{4}$.

Exercise 5

In the previous exercise, are the scores of college students on the mathematics test and the music test positively correlated, negatively correlated, or uncorrelated?

Solution

First,

$$E(XY) = \int_0^1 \int_0^1 xy \cdot \frac{2}{5}(2x+3y)dxdy = \frac{1}{3}$$

Next the marginal p.d.f f_1 of X was found in exercise 4. Therefore,

$$E(X) = \int_0^1 xf_1(x)dx = \int_0^1 \frac{1}{5}x(4x+3)dx = \frac{17}{30}$$

Furthemore it was found that $E(Y) = \frac{3}{5}$. So

$$Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{17}{51} - \frac{17}{50} < 0$$

Therefore X and Y are negative correlated.

Exercise 6

Three men A, B, and C shoot at a target. Suppose that A shoots three times and the probability that he will hit the target on any given shot is $1/8$, B shoots five times and the probability that he will hit the target on any given shot is $1/4$, and C shoots twice and the probability that he will hit the target on any given shot is $1/2$. What is the expected number of times that the target will be hit?

Solution

Let N_A, N_B, N_C denote the number of times each man hits the target. Then

$$E(N_A + N_B + N_C) = 3 \cdot \frac{1}{8} + 5 \cdot \frac{1}{4} + 2 \cdot \frac{1}{2} = \frac{21}{8}$$

Exercise 7

Suppose that the random variables $X_1, \ldots, X_n$ form n Bernoulli trials with parameter p. Determine the conditional probability that $X_1 = 1$, given that

$$\sum_{i=1}^{n} X_i = k \quad (k = 1, \ldots, n)$$

Solution

$$Pr\left(X_1 = 1 | \sum_{i=1}^{n} X_i = k\right) = \frac{Pr\left(X_1 = 1, \sum_{i=1}^{n} X_i = k\right)}{Pr\left(\sum_{i=1}^{n} X_i = k\right)} = \frac{Pr\left(X_1 = 1, \sum_{i=2}^{n} X_i = k-1\right)}{Pr\left(\sum_{i=1}^{n} X_i = k\right)}$$

Since $X_1, ..., X_n$ are independent, it follows that X_1 and $\sum_{i=2}^{n} X_i$ are independent. Therefore the final expression can be rewritten as

$$\frac{Pr(X_1 = 1) Pr\left(\sum_{i=2}^{n} X_i = k-1\right)}{Pr\left(\sum_{i=1}^{n} X_i = k\right)}$$

The sum $\sum_{i=2}^{n} X_i$ has the binomial distribution with parameters $n-1$ and p and the sum $\sum_{i=1}^{n} X_i$ has the binomial distribution with parameters n and p. Therefore,

$$Pr\left(\sum_{i=2}^{n} X_i = k-1\right) = \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} = \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}$$

and

$$Pr\left(\sum_{i=1}^{n} X_i = k\right) = \binom{n}{k} p^{k} (1-p)^{n-k}$$

Also $Pr(X_1 = 1) = p$. So we conclude that

$$Pr\left(X_1 = 1 | \sum_{i=1}^{n} X_i = k\right) = \frac{\binom{n-1}{k-1} p^{k} (1-p)^{n-k}}{\binom{n}{k} p^{k} (1-p)^{n-k}} = \frac{k}{n}$$

Exercise 8

In a clinical trial with two treatment groups, the probability of success in one treatment group is 0.5 , and the probability of success in the other is 0.6 . Suppose that there are five patients in each group. Assume that the outcomes of all patients are independent. Calculate the probability that the first group will have at least as many successes as the second group.

Solution

Let X be the number of successes in the group with probability 0.5 of success. Let Y be the number of successes in the group with probability 0.6 of success. We know that $X \sim Bin(5, 0.5)$ and $Y \sim Bin(5, 0.6)$. We want $Pr(X \geq Y)$. There are 36 possible (X, Y) pairs and we need the sum of the probabilities of the 21 of them for which $X \geq Y$. We shall calculate the probabilities of the 15 other ones and subtract the total from 1. Since X and Y are independent we have that $Pr(X = x, Y = y) = Pr(X = x)Pr(Y = y)$ and each probability can calculated from the p.f of binomial. So the desired probability is 0.4957.

Exercise 9

Suppose that in a certain book there are on the average λ misprints per page and that misprints occurred according to a Poisson process. What is the probability that a particular page will contain no misprints?

Solution

The p.f of Poisson distribution with mean λ is $f(x; \lambda) = \frac{\lambda^x}{x!}e^{-\lambda}$. So $Pr(X = 0) = f(0; \lambda) = e^{-\lambda}$.

Exercise 10

Suppose that X_1 and X_2 are independent random variables and that X_i has the Poisson distribution with mean $\lambda_i (i = 1, 2)$. For each fixed value of $k (k = 1, 2, \ldots)$, determine the conditional distribution of X_1 given that $X_1 +\ X_2 = k$.

Solution

For $X = 0, .., k$ we have

$$Pr(X_1 = x | X_1 + X_2 = k) = \frac{Pr(X_1 = x, X_1 + X_2 = k)}{Pr(X_1 + X_2 = k)} = \frac{Pr(X_1 = x, X_2 = k - x)}{Pr(X_1 + X_2 = k)}$$

and since X_1 and X_2 are independent we have that

$$Pr(X_1 = x | X_1 + X_2 = k) = \frac{Pr(X_1 = x)Pr(X_2 = k - x)}{Pr(X_1 + X_2 = k)}$$

Now we know by theorem that the sum of 2 Poisson distributions is a Poisson distribution with mean the sum of their means. So

$$Pr(X_1 = x) = \frac{\lambda_1^x}{x!}e^{-\lambda_1}$$

$$Pr(X_2 = k - x) = \frac{\lambda_2^{k-x}}{(k - x)!}e^{-\lambda_2}$$

$$Pr(X_1 + X_2 = k) = \frac{(\lambda_1 + \lambda_2)^k}{k!}e^{-(\lambda_1+\lambda_2)}$$

So we have that

$$Pr(X_1 = x | X_1 + X_2 = k) = \frac{k!}{x!(k - x)!}\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^x \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{k-x} = \binom{k}{x} p^x (1 - p)^{k-x}$$

So $X_1 | X_1 + X_2$ is a binomial distribution with parameters k and $p = \frac{\lambda_1}{\lambda_1+\lambda_2}$.

Exercise 11

Suppose that X has the geometric distribution with parameter p. Show that for every nonnegative integer k, $\Pr(X \geq k) = (1-p)^k$.

Solution

$$Pr(X \geq k) = \sum_{x=k}^{\infty} p(1-p)^x = p(1-p)^k \sum_{x=k}^{\infty} p(1-p)^{x-k}$$

Let $i = x - k$, then

$$Pr(X \geq k) = p(1-p)^k \sum_{i=0}^{\infty} (1-p)^i = p(1-p)^k \frac{1}{1-(1-p)} = (1-p)^k$$

Exercise 12

If the temperature in degrees Fahrenheit at a certain location is normally distributed with a mean of 68 degrees and a standard deviation of 4 degrees, what is the distribution of the temperature in degrees Celsius at the same location?

Solution

If X denotes the temperature in degrees Fahrenheit and Y denotes the temperature in degrees Celsius then $Y = \frac{5}{9}(X - 32)$. Since Y is a linear function of X, then Y will also have a normal distribution with

$$E(Y) = \frac{5}{9}(68 - 32) = 20$$

and

$$Var(Y) = \left(\frac{5}{9}\right)^2 \cdot 16 = \frac{400}{81}$$

Exercise 13

Find the 0.25 and 0.75 quantiles of the Fahrenheit temperature at the location mentioned in the previous exercise.

Solution

We know that $F \sim N(68, 4^2)$. So $\frac{F-68}{4} = Z \sim N(0,1)$. For the first quantile it must be $P(Z \leq z_1) = 0.25$ and for the third $P(Z \leq z_3) = 0.75$. So $z_1 = -0.6745$ and $z_3 = 0.6745$. There fore $f_i = 4z_i + 68$ and $f_1 = 65.302$, $f_3 = 70.698$.

Exercise 14

If a random sample of 25 observations is taken from the normal distribution with mean μ and standard deviation 2 , what is the probability that the sample mean will lie within one unit of μ ?

Solution

We know that $E(\overline{\mathrm{X}_{25}}) = \mu$ and $Var(\overline{\mathrm{X}_{25}}) = \frac{\sigma^2}{25} = \frac{4}{25}$. Therefore if we let $Z = \frac{5(\overline{\mathrm{X}_{25}} - \mu)}{2}$ then $Z \sim N(0,1)$. So

$$Pr(|\overline{\mathrm{X}_{25}} - \mu| \leq 1) = Pr(|Z| \leq 2.5) = Pr(-2.5 \leq Z \leq 2.5) = Pr(Z \leq 2.5) - Pr(Z \leq -2.5) = 0.9876$$

Exercise 15

Suppose that a random sample of size n is to be taken from the normal distribution with mean μ and standard deviation 2. Determine the smallest value of n such that

$$\Pr\left(\left|\bar{X}_n - \mu\right| < 0.1\right) \geq 0.9$$

Solution

If we let $Z = \frac{\sqrt{n}(\overline{\mathrm{X}_{\mathrm{n}}} - \mu)}{2}$, then $Z \sim N(0,1)$. Therefore

$$Pr(|\overline{\mathrm{X}_{\mathrm{n}}} - \mu| < 1) = Pr(|Z| < 0.05\sqrt{n}) \geq 0.9$$

So

$$0.05\sqrt{n} \geq 1.645$$

The smallest integer n which satisfies this inequality is $n = 1083$.

Exercise 16

Suppose that n items are being tested simultaneously, the items are independent, and the length of life of each item has the exponential distribution with parameter β. Determine the expected length of time until three items have failed.

Solution

The length of time Y_1 until one component fails has the exponential distribution with parameter $n\beta$. Therefore, $E(Y_1) = \frac{1}{n\beta}$. The additional length of time Y_2 until a second component fails has the exponential distribution with parameter $(n-1)\beta$. Therefore, $E(Y_1) = \frac{1}{(n-1)\beta}$. Similarly, $E(Y_3) = \frac{1}{(n-2)\beta}$. The total time until three components fail is $Y_1 + Y_2 + Y_3$ and

$$E(Y_1 + Y_2 + Y_3) = \left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2}\right)\frac{1}{\beta}$$

Exercise 17

Suppose that a certain examination is to be taken by five students independently of one another, and the number of minutes required by any particular student to complete the examination has the exponential distribution for which the mean is 80 . Suppose that the examination begins at 9:00 A.M. Determine the probability that at least one of the students will complete the examination before 9:40 A.M.

Solution

The time Y_1 until one of the students completes the examination has the exponential distribution with parameter $5\beta = \frac{5}{80} = \frac{1}{16}$. Therefore

$$Pr(Y_1 < 40) = 1 - e^{-\frac{40}{16}} = 0.9179$$