<div align="center">

Take home exam;
Due: Wednesday September 24

</div>

Exercise 1

Let $\alpha \in (-1, 1)$. The random variables X, Y have the joint pdf $f_{XY}(x, y) = 0.25(1 + \alpha xy)$, $x, y \in [-1, 1]$; $f_{XY}(x, y) = 0$ *otherwise*

A) Find the marginal pdf's of X and Y.

B) Find EX, EY, V ar(X), V ar(Y), Cov(X, Y) and Corr(X, Y). Are the random variables X and Y independent?

Solution

A) $f_X(x) = \int_{-1}^{1} f(x, y) \, dy = 0.5 = f_Y(y)$

B) $EX = \int_{-1}^{1} 0.5 \, x \, dx = 0 = EY$

$Var(X) = \int_{-1}^{1} 0.5 \, x^2 \, dx = \dfrac{1}{3} = Var(Y)$

$Cov(X, Y) = E(X - \overline{X})(Y - \overline{Y}) = E(XY) = \int_{-1}^{1} \int_{-1}^{1} xy * 0.25(1 + \alpha xy) \, dxdy = \dfrac{\alpha}{9}$

$Corr(X, Y) = \dfrac{Cov(X, Y)}{\sigma_x \sigma_y} = \dfrac{\alpha}{3}$

X and Y are independent because f(0,0) is equal $f_X(0) * f_Y(0)$

Exercise 2

A taxi was involved in a hit and run accident at night. There are two taxi companies in the city, namely Black Taxis and White Taxis. We know that 95% of the taxis in the city are Black and 5% are White. There is a witness to the accident and, according to the witness, the taxi involved in the accident was White. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 90% of the time and failed 10% of the time. Calculate the probability that the taxi involved was White.

Solution

P(W) = 0.05 * 0.9 + 0.95 * 0.1 = $\dfrac{7}{50}$

P(W|W) = $\dfrac{0.05 * 0.9}{\dfrac{7}{50}} = \dfrac{9}{28}$

Exercise 3

We say that the random variable X has a Pareto distribution with parameter 4 if its pdf is given by $f_X(x) = kx^{-4} \; for \; x \geq 1 \; and \; f_X(x) = 0$ otherwise, where k is some positive constant.

(a) Find the constant k.

(b) Find the corresponding cdf $F_X(x)$

(c) Find $\mu_X = EX$ and $\sigma_X^2 = Var(X)$

(d) What can you say about the skewness and kurtosis of the random variable X?

(e) Find $P(|X - \mu_X| < 1.5\sigma_X)$.

(f) Use Chebyshev's inequality to estimate $P(|X - \mu_X| < 1.5\sigma_X)$ and compare this estimate with the exact value found in the previous part.

Solution

(a) We know that $\int_1^\infty f_X(x)\ dx = 1$.

$$-\frac{k}{3x^3}\Big|_1^\infty = 1$$

Hence k=3.

(b) $\text{CDF} = \int_1^x f_X(t)\ dt = 1 - \frac{1}{x^3}$

(c) $\text{EX} = \int_1^\infty x * f_X(x)\ dx = \frac{3}{2}$

$\text{Var(x)} = EX^2 - (EX)^2 = \int_1^\infty x^2 * f_X(x)\ dx - \frac{9}{4} = \frac{3}{4}$

(d) Skewness and kurtosis are infinite because the integral of the third and the forth moments are infinite with the power -4 of Pareto distribution.

(e) $P(|x - \frac{3}{2}| < \frac{3\sqrt{3}}{4}) = P(\frac{3}{2} - \frac{3\sqrt{3}}{4} < x < \frac{3\sqrt{3}}{4} + \frac{3}{2}) = F_X(\frac{3\sqrt{3}}{4} + \frac{3}{2}) - F_X(\frac{3}{2} -$

$\frac{3\sqrt{3}}{4}) = F_X(\frac{3\sqrt{3}}{4} + \frac{3}{2})$ because x is greater than or equal to one by the task.

It is equal to $1 - 1 + \dfrac{64}{(3\sqrt{3} + 6)^3} = 0.0456009$

(f) According to Chebyshev's inequality the bound of probability is $1 - \dfrac{\sigma_X^2}{(\dfrac{3\sqrt{3}}{4})^2} =$

$1 - \dfrac{1}{\sqrt{3}} = 0.4226497$

Exercise 4 Let X_1, X_2 ... be a sequence of independent random variables with the pdf $f_X(x) = 0.5\ for\ 0 \leqslant x \leqslant 2\ and\ f_X(x) = 0$ otherwise (uniform distribution on the interval $[0, 2]$).

(a) Find the asymptotic distribution of the sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

(b) Find the asymptotic distribution of the exponent of the sample mean $e^{\overline{X}_n}$

(c) What is the asymptotic distribution of $n(e^{\overline{X}_n} - e)^2$

Solution

(a) $E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}E(\sum_{i=1}^{n} X_i) = EX$

$EX = \int_0^2 x * 0.5 \ dx = 1$

$\sigma^2 = EX^2 - (EX)^2 = \frac{1}{3}$

According to CLT, $\sqrt{n}(\sum_{i=1}^{n} x_i - \mu) \backsim N(0, \sigma^2)$.

$\frac{1}{n}\sum_{i=1}^{n} X_i \backsim N(1, \frac{1}{3n})$

(b) Using delta method we can find that g(x) = e^x and so:

$\sqrt{n}(g(\overline{X}_n) - g(0)) \backsim N(0, \sigma^2(g'(Q))^2)$

Thus $e^{\overline{X}_n} \backsim N(e, \frac{e^2}{9n^2})$

(c) We can find that we have find the square of $\sqrt{n}(g(\overline{X}_n) - g(0))$

We also know that if $\sqrt{n}(g(\overline{X}_n) - g(0)) \backsim N(0, 1)$ than the square of this value $\backsim \chi_1^2$

Thus $\frac{3n^2(e^{\overline{X}_n} - e)}{e^2} \backsim \chi_1^2$

And finally we have that $n(e^{\overline{X}_n} - e) \backsim \frac{e^2\chi_1^2}{3n}$

Exercise 5

The owner of a small business is developing a strategic plan for the business and hence, wants to develop some estimates of next year's income. Let $X denote the weekly income. From last year's records we obtain the estimates: $\mu_X = 2000 \ and \ \sigma_X = 250$

(a) Using these estimates and the CLT, estimate the probability that the annual income ($A) will exceed $110,000; that is, find P $(A \geqslant 110, 000)$.

(b) Find an interval [A1,A2] (with mid-point 104,000) such that P($A_1 \leqslant A \leqslant A_2$)=0.9. (such an interval may be described as a 90% prediction interval for A.

(c) (Extra credit). Discuss any difficulties associated with applying the CLT to this problem.

Solution

(a) According to CLT we know that A has distribution N(2000, 250)

And so the probability that random variable A will be greater than or equal to 110000 is 0.322206

(b) We know that this interval is is equal to 52 * X ± 1.65 * 52 * SE(X)

$$\text{SE(X)} = \frac{\sigma}{\sqrt{n}} = 57$$

And so our interval is [101036, 106964]

(c) CLT means that there should be a lot of measures to distributions goes to normal. 52 is not enough for this assumption and so our probabilities are not very truthy.

Exercise 6

Let $X_1, ..., X_n$ be a sequence of i.i.d. random variables with common probability density function $f(x, \lambda) = 2\lambda x e^{-\lambda x^2}$ for x > 0; f(x,λ) = 0 otherwise, where λ > 0 is some parameter.

(a) Find the maximum likelihood estimator of λ

(b) What is the Cramer-Rao lower bound for the variance of any un- biased estimator of λ ?

(c) What is the asymptotic distribution of the maximum likelihood estimator?

Solution

(a)

(b)

(c)