Lecture 22:
Continuous Random Variables - Part V

Satyajit Thakor
IIT Mandi

Covariance and correlation

- Example: Suppose that a fair coin is tossed three times so that there are eight equally likely outcomes, and that the r.v. X is the number of heads obtained in the first and second tosses and the r.v. Z is the number of heads obtained in the second and third tosses. Find $\mathrm{Cov}(X, Z)$ and $\mathrm{Corr}(X, Z)$.

– Let's first find the joint PMF & marginals

$X=0, Z=0$: TTT
$X=1, Z=0$: HTT
$X=2, Z=0$: –

$X=0, Z=1$: TTH
$X=1, Z=1$: HTH, THT
$X=2, Z=1$: HHT

$X=0, Z=2$: – , $X=1, Z=2$: HHT, $X=2, Z=2$: HHH

$z \backslash x$	0	1	2	$p_Z(z)$
0	1/8	1/8	0	1/4
1	1/8	1/4	1/8	1/2
2	0	1/8	1/8	1/4
$p_X(x)$:	1/4	1/2	1/4	

Covariance and correlation

- Now, $\operatorname{Cov}(X,Z) = E(XZ) - E(X)E(Z)$

$$E(X) = E(Z) = 0 + 1 \cdot \tfrac{1}{2} + 2 \cdot \tfrac{1}{4} = 1.$$

$$\operatorname{Var}(X) = \operatorname{Var}(Z) = E(X^2) - E(X)^2 = 1.5 - 1 = 0.5$$

$$E(XZ) = \sum_{x=0}^{2} \sum_{z=0}^{2} x \cdot z \cdot p_{XZ}(x,z)$$

$$= 1 \cdot 1 \cdot \frac{1}{4} + 1 \cdot 2 \cdot \left(\frac{1}{8} + \frac{1}{8}\right) + 2 \cdot 2 \cdot \frac{1}{8}$$

$$= \frac{1}{4} + \frac{1}{2} + \frac{1}{2} = 1.25.$$

$$\Rightarrow \operatorname{Cov}(X,Z) = 1.25 - 1 = 0.25.$$

$$\Rightarrow \operatorname{Corr}(X,Z) = \frac{\operatorname{Cov}(X,Z)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Z)}} = \frac{.25}{\sqrt{.5 \times .5}} = 0.5.$$

Continuous random variables

- Example: A researcher plants 12 seeds whose germination times in days are independent exponential distributions with $\lambda = 0.31$.
- (a) What is the probability that a given seed germinates within five days?

- Recall: If X has exp. distribution then,

$$f_X(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases} \quad \text{and} \quad F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0. \end{cases}$$

- Let X be germination time in days.
- We have X exponentially distributed with $\lambda = 0.31$.
- Hence, $P(X \leq 5) = F_X(5) = 1 - e^{-0.31 \cdot 5}$

$$\approx 0.7878$$

Continuous random variables

- (b) What are the expectation and variance of the number of seeds germinating within five days?

- We define Y_i as follows:

$$Y_i = \begin{cases} 1 & \text{if } i^{th} \text{ seed germinates in 5 days} \\ 0 & \text{" " " does not germinate in 5 days.} \end{cases}$$

- Hence, $P(Y_i = 1) = 0.7878$ and $P(Y_i = 0) = 1 - P(Y_i = 1)$.

- Let $Y = \sum_{i=1}^{12} Y_i$: Y is no. of seeds germinating in 5 days.

- Note that Y is binomially distributed with $n = 12$, $p = 0.7878$.

- Hence, $E(Y) = n \cdot p = 12 \cdot 0.7878 = 9.45$.

$$Var(Y) = np(1-p) = 12 \cdot 0.7878 \cdot 0.2122 = 2.01.$$

Continuous random variables

- (c) What is the probability that no more than nine seeds have germinated within five days?

- We need to find $P(Y \leq 9)$.

$$P(Y \leq 9) = \sum_{y=0}^{9} P(Y = y)$$

$$= \sum_{y=0}^{9} \binom{12}{y} p^{y} (1-p)^{12-y}$$

$$\approx 0 + 0 + 0 + 0.0001 + 0.0008 + 0.0047$$
$$+ 0.0202 + 0.0642 + 0.1489 + 0.2457$$

$$= 0.4845.$$

Continuous random variables

- So far, we have discussed joint and conditional distributions, independence, covariance and correlation for discrete random variables.
- These parameters can also be extended for continuous random variables.
- Joint PDF of continuous r.v.s X and Y: the joint probability density function is a function $f_{X,Y}(x,y)$ such that $f_{X,Y}(x,y) \geq 0$ and

$$\int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1.$$

- That is, the joint PDF is non-negative and the area under the joint PDF curve is 1.

Continuous random variables

- Example: a mining company obtains samples of ore from the location and measures their zinc content and their iron content. Suppose that the r.v. X is the zinc content of the ore, taking values between 0.5 and 1.5, and that the r.v. Y is the iron content of the ore, taking values between 20.0 and 35.0. Furthermore, suppose that their joint PDF is

$$f_{X,Y}(x,y) = \frac{39}{400} - \frac{17(x-1)^2}{50} - \frac{(y-25)^2}{10000}$$

for $0.5 \leq x \leq 1.5$ and $20.0 \leq y \leq 35.0$.

- Check whether $f_{X,Y}$ is a valid PDF:

– First, check whether $f_{X,Y}(x,y) \geq 0$:

$$f_{X,Y}(x,y) = \frac{39}{400} - \frac{17(x-1)^2}{50} - \frac{(y-25)^2}{10000}, \quad .5 \leq x \leq 1.5, \; 20 \leq y \leq 35$$

$$\geq \frac{39}{400} - \frac{17(0.5)^2}{50} - \frac{(10)^2}{10000} = 0.0025 \geq 0.$$

Continuous random variables

- Now, check whether the area under the curve is 1.

$$\int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} f_{X,Y}(x,y)\, dy\, dx$$

$$= \int_{x=0.5}^{1.5} \int_{y=20}^{35} \frac{39}{400} - \frac{17(x-1)^2}{50} - \frac{(y-25)^2}{10000}\, dy\, dx$$

$$= \int_{x=0.5}^{1.5} \left[\frac{39y}{400}\Big|_{20}^{35} - \frac{17(x-1)^2}{50} \cdot y \Big|_{20}^{35} - \frac{(y-25)^3}{3 \cdot 10000}\Big|_{20}^{35} \right] dx$$

$$= \int_{x=0.5}^{1.5} \left[\frac{39 \cdot 15}{400} - \frac{17 \cdot 15 \cdot (x-1)^2}{50} - \frac{375}{10000} \right] dx$$

$$= \frac{39 \cdot 15 \cdot x}{400}\Big|_{0.5}^{1.5} - \frac{17 \cdot 15}{50} \cdot \frac{(x-1)^3}{3}\Big|_{0.5}^{1.5} - \frac{375}{10000} \cdot x \Big|_{0.5}^{1.5}$$

$$= \frac{39 \cdot 15}{400} - \frac{17 \cdot 15}{50} \cdot \frac{1}{12} - \frac{375}{10000} = 1.$$

Continuous random variables

- What is the probability that a randomly chosen sample of ore has a zinc content between 0.8 and 1.0 and an iron content between 25 and 30?

- We want to find $P(0.8 \leq X \leq 1, 25 \leq Y \leq 30)$.

$$P(0.8 \leq X \leq 1, 25 \leq Y \leq 30)$$

$$= \int_{x=0.8}^{1} \int_{y=25}^{30} f_{XY}(x,y)\, dx\, dy$$

$$= 0.092 \quad (\text{Homework: verify})$$

- That is, about 9% of the ore has mineral levels $0.8 \leq X \leq 1, 25 \leq Y \leq 30$.