BITS Pilani
Pilani Campus

INTRODUCTION TO STATISTICAL METHODS

ISM Team

BITS Pilani
Pilani Campus

INTRODUCTION TO STATISTICAL METHODS

WEBINAR 2 : 07.12.2021

WEBINAR 2

A real valued function defined on the sample space.

- Random Variables $f: S \to R$

S . R

Dis

Cont.

- Joint Distribution X, Y
- Family of Random Variables

Dis cont

1.

2.

3.

WEBINAR 2

Discrete - Sum
Continuous - $\int$

innovate achieve lead

1. Check whether the following can serve as probability distributions: → Prob

a. $f(x) = \dfrac{x-2}{2}; \quad x = 1,2,3,4$ — Not valid Prob $x=1,\ f(1) = -\frac{1}{2}$ (-ve)

b. $f(x) = \begin{cases} 2e^{-2x}; & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$ $\geq 0 \quad > 0$

→ Valid

1. $f(x) \geq 0$ for every $x \geq 0$

2. $\int_{-\infty}^{\infty} f(x)\,dx = 1$

$$\int_0^{\infty} 2e^{-2x}\,dx = \left[2\frac{e^{-2x}}{-2} \right]_0^{\infty} = 1$$

$$\left[-e^{-\infty} - (-1) \right] = 1$$

Note:

WEBINAR 2

$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots$

$<$ or $\leq$

2. A random variable X is having probability distribution function is given by

$$f(x) = \frac{e^{-k}k^x}{x!},\ x = 0,1,2,3\ldots \text{with } \boxed{k = 2}$$

$\frac{e^{-\lambda}\lambda^x}{x!}$

a. Is it a valid distribution? Justify. If valid then find the following

b. P(x > 2)

c. P(1 < x < 3)

(i) $f(x) \geq 0$

(ii) $\sum_{x=0}^{\infty} \frac{e^{-2}2^x}{x!} = \frac{e^{-2}2^0}{0!} + \frac{e^{-2}2^1}{1!} + \frac{e^{-2}2^2}{2!} + \cdots$

$= e^{-2}\left[1 + \frac{2}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} \cdots\right] = e^{-2}(e^2)$

$= 1$

Valid.

a) $P(x>2) = P(x = 3, 4, \cdots)$

$= 1 - P(X = 0, 1, 2)$

$= 1 - [P(x=0) + P(x=1) + P(x=2)]$

$= 1 - \left[\frac{e^{-2}2^0}{0!} + \frac{e^{-2}2^1}{1!} + \frac{e^{-2}2^2}{2!}\right] = 1 - 5e^{-2}$

$= P[x = 2]$

$= \frac{e^{-2}2^2}{2!}$

WEBINAR 2

$x = 2,\ 0.1$

cumulative - till

$f(x)$ $p(x) \rightarrow P(X=x)$

$p(a) \rightarrow P(X=a)$

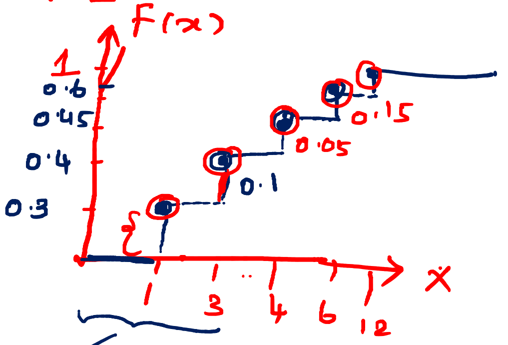
3. An insurance company offers its policyholders a number of different premium payment options. For a randomly selected policyholder, let of months between successive payments. The cdf of X is as follows

non-decreasing

$F(x) = P[X \le x]$

$$F(x) = \begin{cases} 0 & x < 1 \\ .30 & 1 \le x < 3 \\ .40 & 3 \le x < 4 \\ .45 & 4 \le x < 6 \\ .60 & 6 \le x < 12 \\ 1 & 12 \le x \end{cases}$$

$X = 1, 3, 4, 6, 12$



a. What is the PMF of X?
b. Compute $P(3 \le X \le 6)$
c. Obtain E(x) and Variance of X.

PMF - discrete

PdF - Continuous

a. $f(x) = \begin{cases} 0.3 & x = 1 \\ 0.1 & x = 3 \\ 0.05 & x = 4 \\ 0.15 & x = 6 \\ 0.4 & x = 12 \\ 0 & \text{otherw.} \end{cases}$

x	0	1	2	3
$f(x)$	0.1	0.4	0	0.5
$F(x)$	0.1	0.5	0.5	1

b. $P[X = 3, 4, 6]$

$= 0.1 + 0.05 + 0.15$

$= 0.3$

WEBINAR 2

$0 \le p(x) \le 1$

$E(x)$

$X = \{ \cancel{1, 3, \ldots} \ldots \}$

$\{0, 1, 2, 3, \ldots\}$

innovate achieve lead

3. An insurance company offers its policyholders a number of different premium payment options. For a randomly selected policyholder, let of months between successive payments. The cdf of X is as follows

$$F(x) = \begin{cases} 0 & x < 1 \\ .30 & 1 \le x < 3 \\ .40 & 3 \le x < 4 \\ .45 & 4 \le x < 6 \\ .60 & 6 \le x < 12 \\ 1 & 12 \le x \end{cases}$$

a. What is the PMF of X?
b. Compute P(3≤ X ≤6)
c. Obtain E(x) and Variance of X.

x	1	3	4	6	12
$p(x)$	0.3	0.1	0.05	0.15	0.4
$x\,p(x)$	0.3	0.3	0.2	0.9	4.8
$x^2 p(x)$	. . .				

$E(x) = \sum x\,p(x) = 6.5$

$Var(x) = E(x^2) - (E(x))^2 = 22.75$

WEBINAR 2

Note: Discrete $x = 0, 1, 2, 3,$ $x < 2, \Rightarrow \boxed{0, 1}$

Continuous. $0 < x < \infty$.

$x < 2,$ $x \leq 2$ $\int_0^{2}$ not $\int_0^{1} x$

4. Let X denote the temperature at which a certain chemical reaction takes place. Suppose that X has pdf → continuous.

$$f(x) = \begin{cases} \frac{1}{9}(4 - x^2) & -1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

$$f(x) = \begin{cases} \frac{1}{9}(4 - x^2), & -1 \leq x \leq 2 \\ 0 & \text{otherwise.} \end{cases}$$

a. Compute P(0≤ X ≤1)

b. Obtain E(x) and Variance of X

Average.

Mean

expected value

1.9999.

$P(0 \leq x \leq 1)$

$= \int_0^1 \frac{1}{9}(4 - x^2)\,dx.$

$= \frac{1}{9}\left[4x - \frac{x^3}{3}\right]_0^1 = \frac{11}{27}$

$= \frac{1}{9}\Big[$

$E(x) \to$ expected value

$= \int x\,p(x) = \int_{-1}^{2} x\frac{1}{9}(4 - x^2)\,dx = \int_{-1}^{2} \frac{1}{9}(4x - x^3)\,dx.$

$E(x) = \frac{1}{4},$

$E(x^2) = \int x^2\,p(x)\,dx = 3/5.$

$Var(x) = \frac{43}{80}$

$E(x^2) - (E(x))^2$

WEBINAR 2

Tossing coins

X : No. of heads/tails

Y : No. of tosses.

innovate achieve lead

5. A service station has both self-service and full-service islands. On each island, there is a single regular unleaded pump with two hoses. Let X denote the number of hoses being used on the self-service island at a particular time and let Y denote the number of hoses on the full-service island in use at that time. The joint probability mass function of X and Y is given below:

X	Y		
	0	1	2
0	0.10	0.04	0.02
1	0.08	0.20	0.06
2	0.06	0.14	0.30

a. Find the marginal probability mass function of X and Y

b. Give the verbal description the event ($X \neq 0$ and $Y \neq 0$) and compute the probability of this event.

c. Find $P(X=1|Y=2)$ and $P(Y=2|X=1)$

WEBINAR 2

$P(x \neq 0 \text{ and } y \neq 0)$

$= 1 - P(x=0 \text{ or } y=0)$

$\rightarrow P(A/B) = \dfrac{P(A \cap B)}{P(B)}$

innovate achieve lead

Marginal PMF

X	Y: 0	1	2	$f_X(x)$
0	0.10	0.04	0.02	0.16
1	0.08	0.20	0.06	0.34
2	0.06	0.14	0.30	0.5

$f_Y(y)$

$$f_X(x) = \begin{cases} 0.16 & x=0 \\ 0.34 & x=1 \\ 0.5 & x=2 \\ 0 & \text{oth.} \end{cases}$$

$$f_Y(y) = \begin{cases} 0.24 & y=0 \\ 0.38 & y=1 \\ 0.38 & y=2 \\ 0 & \text{oth.} \end{cases}$$

a. Find the marginal probability mass function of X and Y

b. Give the verbal description the event (X≠0 and Y≠0) and compute the probability of this event.

$x=1,2$ $\quad y=1,2$ at least

$1 - P(x=0 \overset{\text{and}}{,} y=0)$

$1 - 0.1 = 0.9$

c. Find P(X=1|Y=2) and P(Y=2|X=1)

$$P(x=1/y=2) = \frac{P(x=1 \text{ and } y=2)}{P(y=2)} = \frac{0.06}{0.38} = 0.157 \simeq 0.16$$

$$P(y=2/x=1) = \frac{P(x=1, y=2)}{P(x=1)} = \frac{0.06}{0.34} = 0.18$$

different

WEBINAR 2

Exercise. "Midsem."

innovate achieve lead

Consider the following Joint distribution of two random variables X and

$p(x,y)$

X \ Y	1	2	3	4	5	6	$f_X(x)$
1	0	0	2k	4k	4k	6k	16 K
2	4k	4k	8k	8k	8k	8k	40 K
3	2k	2k	k	k	0	2k	8 K
$f_Y(y)$	6k	6k	11 k	13 k	12 k	16k.	64 K

$P(X \leq 2)$

a. For what value(s) of k it is a valid distribution : 64 k = 1

$$k = \frac{1}{64}$$

b. Find Marginal Distribution of X and Y

$$f_X(x) = \begin{cases} 16K & x=1 \\ 40K & x=2 \\ 8K & x=3 \\ 0 & \text{otherwise.} \end{cases}$$

$$f_Y = \begin{cases} 6K & y=1 \\ 6K & y=2 \\ \vdots & \\ 16K & y=6 \end{cases}$$

c. Find $P(X \leq 2) = \sum_{x=0,1,2} \sum_{\forall y} P(x,y)$

$= 56\ K$

d. Find $P(X \leq 2 / \ Y = 2)$

$$= \frac{P[X \leq 2 \text{ and } Y=2]}{P(Y=2)}$$

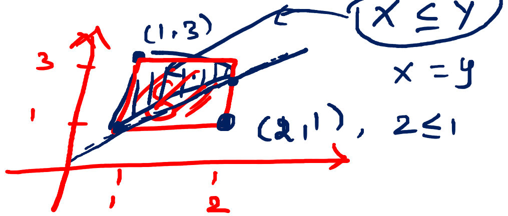
e. Find $P(X \leq 3 / \ Y \leq 2)$

$$= \frac{P[X \leq 3 \cap Y \leq 2]}{P[Y \leq 2]}$$

The joint probability mass function of the two random variables (X, Y) is given by

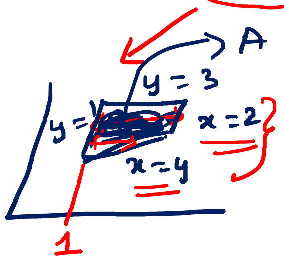
midsem

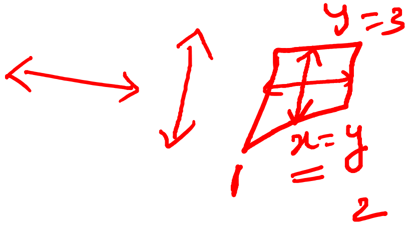
$$f(x,y)=\begin{cases}\frac{1}{5}(3x-y), & 1\le x\le 2,\ 1\le y\le 3\\ 0, & otherwise\end{cases}$$



a. Find the P(X ≤ Y)

b. Find the marginal density functions of X and Y





$$P(X\le Y)=\iint_A \frac{1}{5}(3x-y)\,dx\,dy$$

need to do by two parts of integration

$$=\int\int_1 \frac{1}{5}(3x-y)\,dx\,dy$$

changing the order the integration

$$=\int_1^2\int_x^3 \frac{1}{5}(3x-y)\,dy\,dx$$

$$= 19/30$$

c. Are X and Y independent?

d. Find E(XY)

Steps

$$\int_1^2\int_x^3 \frac{1}{5}(3x-y)\,dydx$$

$$\int_x^3 \frac{1}{5}(3x-y)\,dy=\frac{1}{5}\left(-\frac{9}{2}+9x\right)-\frac{x^2}{2}$$

$$=\int_1^2 \frac{1}{5}\left(-\frac{9}{2}+9x\right)-\frac{x^2}{2}dx$$

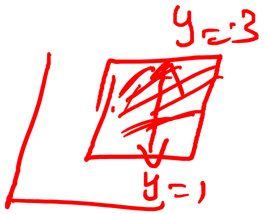
$$\int_1^2 \frac{1}{5}\left(-\frac{9}{2}+9x\right)-\frac{x^2}{2}dx=\frac{19}{30}$$

$$=\frac{19}{30}$$

WEBINAR 2

The joint probability mass function of the two random variables (X, Y) is given by

$$f(x,y) = \begin{cases} \frac{1}{5}(3x-y), 1 \le x \le 2, 1 \le y \le 3 \\ 0, otherwise \end{cases}$$



a. Find the P(X ≤ Y)

b. Find the marginal density functions of X and Y

$$f_Y(y) = \int_1^2 f(x,y)\,dx$$

$$= \frac{1}{5}\left(\frac{9}{2} - y\right)$$

$$f_X(x) = \int_1^3 f(x,y)\,dy$$

$= f(x)$ in terms of x.

$$f_X(x) = \frac{1}{5}\left[3x-1\right]$$

c. Are X and Y independent?

$f(x,y) \neq f(x)\,f(y)$

Not

d. Find E(XY)

$$= \int_1^3\int_1^2 \frac{1}{5}(3x-y)\,dx\,dy$$

$$= 3$$

$$\int_1^3\int_1^2 \frac{1}{5}xy(3x-y)\,dx\,dy$$

$$\int_1^2 \frac{1}{5}xy(3x-y)\,dx = \frac{y(-3y+14)}{10}$$

$$= \int_1^3 \left(\frac{y(-3y+14)}{10}\right)dy$$

$$\int_1^3 \left(\frac{y(-3y+14)}{10}\right)dy = 3$$

$$= 3$$

WEBINAR 2

The joint probability mass function of the two random variables (X, Y) is given by

$$f(x,y) = \begin{cases} \frac{1}{5}(3x - y), 1 \le x \le 2, 1 \le y \le 3 \\ 0, otherwise \end{cases}$$

a. Find the $P(X \le Y)$

b. Find the marginal density functions of X and Y

c. Are X and Y independent?

d. Find E(XY)

$$\int_1^3 \int_1^2 \frac{xy(3x-y)}{5} dxdy$$

$$\int_1^2 \frac{xy(3x-y)}{5} dx = \frac{y(-3y+14)}{10}$$

$$= \int_1^3 \left(\frac{y(-3y+14)}{10}\right) dy$$

$$\int_1^3 \left(\frac{y(-3y+14)}{10}\right) dy = 3$$

$$= 3$$

WEBINAR 2

innovate achieve lead

It has been claimed that in 60% of all solar-heat installations the utility bill is reduced by at least one-third. Accordingly, what are the probabilities that the utility bill will be reduced by at least one-third in

a. four of five installations;
b. at least four of five installations
c. atmost four of five installations?

x

Binomial $(n, p) = {}^nC_x \, p^x (1-p)^{n-x}$.

Solution (a) Substituting $x = 4$, $n = 5$, and $p = 0.60$ into the formula for the binomial distribution, we get

$$b(4; 5, 0.60) = \binom{5}{4}(0.60)^4(1-0.60)^{5-4}$$
$$= 0.259$$

(b) Substituting $x = 5$, $n = 5$, and $p = 0.60$ into the formula for the binomial distribution, we get

$$b(5; 5, 0.60) = \binom{5}{5}(0.60)^5(1-0.60)^{5-5}$$
$$= 0.078$$

and the answer is $b(4; 5, 0.60) + b(5; 5, 0.60) = 0.259 + 0.078 = 0.337$. ■

WEBINAR 2

not avail. $\leftarrow n$ - large.
p - small

innovate achieve lead

For health reasons, homes need to be inspected for radon gas which decays and produces alpha particles. One device counts the number of alpha particles that hit its detector. To a good approximation, in one area, the count for the next week follows a Poisson distribution with mean 1.3. Determine

"λ" average.

a. the probability of exactly one particle next week.
b. the probability of one or more particles next week.
c. the probability of at least two but no more than four particles next week.

Solution Unlike the binomial case, there is no choice of a fixed Bernoulli trial here because one can always work with smaller intervals.

(a) $P(X = 1) = \dfrac{\lambda^1 e^{-\lambda}}{1!} = \dfrac{1.3\, e^{-1.3}}{1} = .3543$

Alternatively, using Table 2W, $F(1, 1.3) - F(0, 1.3) = 0.627 - 0.273 = 0.354$

(b) $P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1.3} = 0.727$

(c) $P(2 \leq X \leq 4) = F(4, 1.3) - F(1, 1.3) = 0.989 - 0.627 = 0.362$

WEBINAR 2

A computing system manager states that the rate of interruptions to the internet service is 0.2 per week. Use the Poisson distribution to find the probability of

a. one interruption in 3 weeks
b. at least two interruptions in 5 weeks
c. at most one interruption in 15 weeks.

$\lambda = 0.2 / 1\text{wk.}$

$\lambda = 0.4 / 2\text{wk.}$

$\lambda = 1.0 / 5\text{wk.}$

$\frac{e^{-\lambda} \lambda^{x}}{x!}$

Solution Interruptions to the network occur randomly and the conditions for the Poisson distribution initially appear reasonable. We have $\lambda = 0.2$ for the expected number of interruptions in one week.

In terms of the cumulative probabilities,

(a) with $\lambda = (0.2) \cdot 3 = 0.6$, we get

$$F(1; 0.6) - F(0; 0.6) = 0.878 - 0.549$$
$$= 0.329$$

(b) With $\lambda = (0.2) \cdot 5 = 1.0$, we get

$$1 - F(1; 1.0) = 1 - 0.736$$
$$= 0.264$$

(c) With $\lambda = (0.2) \cdot 15 = 3.0$ we get

$$F(1; 3.0) = 0.199$$

WEBINAR 2

innovate achieve lead

The Poisson Approximation to the Binomial Distribution:

11. It is known that 5% of the books bound at a certain bindery have defective bindings. Find the probability that 2 of 100 books bound by this bindery will have defective bindings using
    a. the formula for the binomial distribution;
    b. the Poisson approximation to the binomial distribution

n - large.
p - 5%.

$\underbrace{100\, C_5\, (p)^5 (1-p)^{95}}$

Solution (a) Substituting $x = 2$, $n = 100$, and $p = 0.05$ into the formula for the binomial distribution, we get

$$b(2; 100, 0.05) = \binom{100}{2}(0.05)^2(0.95)^{98} = 0.081$$

(b) Substituting $x = 2$ and $\lambda = 100(0.05) = 5$ into the formula for the Poisson distribution, we get

$$f(2; 5) = \frac{5^2 \cdot e^{-5}}{2!} = 0.084$$

It is of interest to note that the difference between the two values we obtained (the error we would make by using the Poisson approximation) is only 0.003. [Had we used Table 2W instead of using a calculator to obtain e^{-5}, we would have obtained $f(2; 5) = F(2; 5) - F(1; 5) = 0.125 - 0.040 = 0.085$.]

WEBINAR 2

Let X, the grade of a randomly selected student in a test of a ISM course, be a normal random variable. A professor is said to grade such a test on the curve if he finds the average μ and the standard deviation σ of the grades and then assigns letter grades according to the following table.

Range of the grade	$X \geq \mu+\sigma$	$\mu \leq X < \mu+\sigma$	$\mu-\sigma \leq X < \mu$	$\mu-2\sigma \leq X < \mu-\sigma$	$X < \mu-2\sigma$
Letter grade	A	B	C	D	F

Suppose that the professor of the probability course grades the test on the curve. Determine the percentage of the students who will get A, B, C, D, and F, respectively.

WEBINAR 2

Range of the grade	$X \geq \mu+\sigma$	$\mu \leq X < \mu+\sigma$	$\mu-\sigma \leq X < \mu$	$\mu-2\sigma \leq X < \mu-\sigma$	$X < \mu-2\sigma$
Letter grade	A	B	C	D	F

Solution: By the fact that $(X-\mu)/\sigma$ is standard normal,

Ref normal table

$$P(A) = P(X \geq \mu+\sigma) = P\left(\frac{X-\mu}{\sigma} \geq 1\right) = 1 - \Phi(1) \approx 0.1587,$$

$$P(B) = P(\mu \leq X < \mu+\sigma) = P\left(0 \leq \frac{X-\mu}{\sigma} < 1\right) = \Phi(1) - \Phi(0) \approx 0.3413,$$

$P(C) = P(\mu - \sigma \leq X < \mu) = P\left(-1 \leq \frac{X-\mu}{\sigma} < 0\right) = \Phi(0) - \Phi(-1)$

$$= 0.5 - 0.1587 \approx 0.3413,$$

$P(D) = P(\mu - 2\sigma \leq X < \mu - \sigma) = P\left(-2 \leq \frac{X-\mu}{\sigma} < -1\right) = \Phi(-1) - \Phi(-2)$

$$= 0.1587 - 0.0228 \approx 0.1359,$$

$P(F) = P(X < \mu - 2\sigma) = P\left(\frac{X-\mu}{\sigma} < -2\right) = \Phi(-2) \approx 0.0228.$

Therefore, approximately 16% should get A, 34% B, 34% C, 14% D, and 2% F. If an instructor grades a test on the curve, instead of calculating μ and σ, he or she may assign A to the top 16%, B to the next 34%, and so on. ♦

WEBINAR 2

Exercise.

Midsem' July 2024

BITSAT is conducted every year by BITS for admission to three campuses. Among the eligible students the average score is 320 with a standard deviation of 40. What is the probability that when a random score is drawn, it ranges from 300 to 340? (Assume that scores follows normal distribution)

μ σ

THANKS