STATISTICS 2

NOTES BY
GAGNEET
KAUR

Date
September 21, 2021

classmate
Date
Page

WEEK 3

Expected Value of R.V.

Suppose X is a discrete random variable with range T_X and PMF f_X. The expected value of X, denoted $E[X]$, is defined as

$$E[X] = \sum_{t \in T_X} t \, f_X(t) = \sum_{t \in T_X} t \cdot P(X = x)$$

assuming above sum exists.

- other names : mean of X
- $E[X]$ may or may not belong to the range of X
- $E[X]$ has the same units as X.

Examples : Easy Summation given PMF

(1) $X \sim \text{Bernoulli}(p)$

$E[X] = 0(1-p) + 1(p) = p$

(2) $X \sim \text{Uniform}\{1,2,3,4,5,6\}$

$E[X] = 1 \cdot 1/6 + 2 \cdot 1/6 + 3 \cdot 1/6 + 4 \cdot 1/6 + 5 \cdot 1/6 + 6 \cdot 1/6 = 3.5$

(3) Value of lottery ticket (in Rs.) is $\left\{ \overset{1/1000}{200}, \overset{27/1000}{20}, \overset{972/1000}{0} \right\}$

$$E[x] = 200 \cdot \frac{1}{1000} + 20 \cdot \frac{27}{1000} + 0 \cdot \frac{972}{1000} = ₹\, 0.56$$

(4) Change in temperature (in °C) is $\left\{ \overset{1/5}{-2}, \overset{1/5}{-1}, \overset{1/2}{0}, \overset{1/10}{1} \right\}$

$$E[x] = -2 \cdot \frac{1}{5} + (-1) \cdot \frac{1}{5} + 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{10} = -0.5°C$$

Example : Uniform $\{a, a+1, \ldots, b\}$

$X \sim$ Uniform $\{a, a+1, \ldots, b\}$

$$E[x] = a \cdot \frac{1}{b-a+1} + (a+1)\frac{1}{b-a+1} + \ldots + b \cdot \frac{1}{b-a+1}$$

How to simplify sum?

Identity : $a + (a+1) + \ldots + b = (b-a+1)\left(\frac{a+b}{2}\right)$

$$\therefore \quad E[x] = \frac{a+b}{2}$$

$X \sim$ Geometric (p)

$$E[x] = \sum_{t=1}^{\infty} t(1-p)^{t-1} p = \frac{1}{p}$$

$$\boxed{E[x] = \frac{1}{p}}$$

$X \sim \text{Poisson}(\lambda)$

$$E[X] = \sum_{t=0}^{\infty} t \, e^{-\lambda} \frac{(\lambda)^t}{t!} = \lambda \qquad \boxed{E[X] = \lambda}$$

$X \sim \text{Binomial}(n, p)$

$$E[X] = \sum_{t=0}^{n} t \cdot {}^nC_t \, (p)^t (1-p)^{n-t} = np \qquad \boxed{E[X] = np}$$

HOW TO SIMPLIFY SUMS ?

(1) Difference Eqn (DE) : $a_{t+1} - r a_t = b_t \quad (r \neq 1)$

$$\sum_{t=1}^{n} a_t = \frac{a_1 - r a_n}{1-r} + \frac{1}{1-r} \cdot \sum_{t=1}^{n-1} b_t$$

(2) Geometric Progression (GP) : $a_{t+1} - r a_t = 0 \quad (r \neq 1)$

$$\sum_{t=1}^{n} a_t = \frac{a_1 - r a_n}{1-r} \xrightarrow[n \to \infty]{|r| < 1} \frac{a_1}{1-r}$$

(3) Exponential Fn :

$$\sum_{t=0}^{\infty} e^{-\lambda} \frac{\lambda^t}{t!} = 1$$

(4) Binomial formula :

$$\sum_{k=0}^{n} {}^nC_k \, a^k b^{k-1} = (a+b)^n$$

Date September 22, 2021

Properties of Expected Value

I. CONSTANT R·V and POSITIVE R·V

$$E[X] = \sum_{t \in T_X} t \cdot P(X=t)$$

(1) Consider a constant c as random variable X with $P(X=c) = 1$

$$E[c] = c \cdot 1 = c$$

(2) Suppose X takes only non-negative values, i.e. $P(X \geq 0) = 1$. Then,

$$E[X] \geq 0$$

II. EXPECTED VALUE OF A FUNCTION OF R·V

Theorem : Suppose $X_1, \dots, X_n$ have joint PMF $f_{X_1 \dots X_n}$ with range of X_i denoted T_{X_i}.
Let $g : T_{X_1} \times \dots \times T_{X_n} \to \mathbb{R}$ be a function, & let $Y = g(X_1, \dots, X_n)$ have range T_Y and PMF f_Y. Then,

$$E[g(X_1, \dots, X_n)] = \sum_{t \in T_Y} t \cdot f_Y(t) = \sum_{t_i \in T_{X_i}} g(t_1, \dots, t_n) f_{X_1 \dots X_n}(t_1, \dots, t_n)$$

→ We have seen how to find f_Y, PMF of a function of multiple R·V.

⇒ The above theorem says: To find $E[Y]$, you do not always need f_Y. The joint PMF of $X_1, \dots, X_n$ can be used directly.

Examples for Illustration

(1) $X \sim \{\overset{1/5}{-2}, \overset{1/5}{-1}, \overset{1/5}{0}, \overset{1/5}{1}, \overset{1/5}{2}\}$, $g(x) = X^2 \sim \{\overset{1/5}{0}, \overset{2/5}{1}, \overset{2/5}{4}\}$

$$E[g(x)] = 0 \cdot \frac{1}{5} + 1 \cdot \frac{2}{5} + 4 \cdot \frac{2}{5} = 2$$

$$E[g(x)] = (-2)^2 \cdot \frac{1}{5} + (-1)^2 \cdot \frac{1}{5} + 0 \cdot \frac{1}{5} + (1)^2 \cdot \frac{1}{5} + (2)^2 \frac{1}{5} = 2$$

(2) $(X, Y) \sim$ Uniform $\{(0,0), (1,0), (0,1), (1,1), (-1,1), (1,-1)\}$

$$g(x,y) = X^2 + XY + Y^2 \sim \{\overset{1/6}{0}, \overset{4/6}{1}, \overset{1/6}{3}\}$$

$$E[g(x,y)] = 0 \cdot 1/6 + 1 \cdot 4/6 + 3 \cdot 1/6 = 7/6$$

$$E[g(x,y)] = 0\left(\frac{1}{6}\right) + 1\left(\frac{1}{6}\right) + 1\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 1\left(\frac{1}{6}\right) + 1\left(\frac{1}{6}\right) = \frac{7}{6}$$

III. LINEARITY OF EXPECTED VALUE

(1) $E[cX] = c \cdot E[X]$ for a R.V. X and a constant c.

Proof: $E[cX] = \sum_{t \in T_X} ct\, f_X(t) = c \sum_{t \in T_X} t\, f_X(t) = c \cdot E[X]$

(2) $E[X+Y] = E[X] + E[Y]$ for any two random variables X, Y.

<u>Proof</u>: $$E[X+Y] = \sum_{t_1 \in T_X,\, t_2 \in T_Y} (t_1 + t_2)\, f_{XY}(t_1, t_2)$$

$$= \sum_{t_1 \in T_X,\, t_2 \in T_Y} t_1 \cdot f_{XY}(t_1, t_2) + \sum_{t_1 \in T_X,\, t_2 \in T_Y} t_2 \cdot f_{XY}(t_1, t_2)$$

$\Rightarrow$ $E[X+Y] = E[X] + E[Y]$

(3) $E[aX + bY] = a E[X] + b E[Y]$

\# One of the most useful properties of expected value!

<u>Illustrative Examples</u>

1. $(X, Y) \sim$ Uniform $\{(0,0), (1,0), (0,1), (1,1), (-1,1), (1,-1)\}$

$$g(X,Y) = X^2 + XY + Y^2 \sim \{\overset{1/6}{0}, \overset{4/6}{1}, \overset{1/6}{3}\}$$

$E[g(X,Y)] = 7/6$

$\therefore$ $E[g(X,Y)] = E[X^2 + XY + Y^2] = E[X^2] + E[XY] + E[Y^2] = 7/6$

$(E[X^2] = 4/6\ ;\ E[XY] = -1/6\ ,\ E[Y^2] = 4/6)$

2. $X \sim f_X$, $Y \sim f_Y$, the joint PMF f_{XY} is not given.
Can you compute $E[X+Y]$? Yes!

$E[X+Y] = E[X] + E[Y] = \sum_{t \in T_X} t \cdot f_X(t) + \sum_{t \in T_Y} t \cdot f_Y(t)$

Note: Expected value of the form $E[g(X) + h(Y)]$ can be computed with marginal PMFs, and it does not depend on the joint PMF.

PROBLEMS :

(1) Throw a fair die twice : What is the expected value of the sum of the two numbers seen ?

X : first no.
Y : second no.

$E[X+Y] = E[X] + E[Y] = 3.5 + 3.5 = 7$

(2) Expected Value of Binomial (n, p) :

Suppose $Y \sim$ Binomial (n, p).

$E[Y] = \sum_{k=0}^{n} k \cdot {}^nC_k \, p^k (1-p)^{n-k} = np$

- Such summations may be difficult to simplify.
- Alternative method using linearity of expected value

Let $X_1, \dots, X_n$ be iid Bernoulli (p). Then, $E[X_i] = p$

and

$$Y = X_1 + \dots + X_n \sim \text{Binomial}(n, p)$$

So, $E[Y] = E[X_1] + \dots + E[X_n] = np$

Zero Mean Random Variable

A random variable X with $E[X] = 0$ is said to be a zero-mean random variable.

<u>Translation of a random variable</u>

$X + c$, where c is a constant, is a "translated" version of X.

→ Range of $X+c$ is $\{t+c : t \in T_X\}$, which is translated version of T_X, the range of X

→ $P(X+c = t+c) = P(X=t)$ & the PMF is 'translated' as well.

<u>Translation by expected value</u>

$E[X]$ is a constant.

$Y = X - E[X]$ is translated version of X & $E[Y] = 0$

So, $X - E[X]$ is a zero-mean random variable.

Problem: Balls in Bins

Suppose 10 balls are thrown uniformly at random into 3 bins. What is the expected no. of empty bins?

$$X_i = \begin{cases} 1, & \text{if bin } i \text{ is empty} \\ 0, & \text{otherwise} \end{cases} \quad , i = 1, 2, 3$$

$$P(X_i = 1) = \left(\frac{2}{3}\right)^{10} \qquad P(X_i = 0) = 1 - \left(\frac{2}{3}\right)^{10}$$

$$Y = \text{no. of empty bins} = X_1 + X_2 + X_3$$

$$\therefore \; E[Y] = E[X_1] + E[X_2] + E[X_3] = 3 \cdot \left(\frac{2}{3}\right)^{10}$$

Date September 23, 2021

Variance

Variance & Standard Deviation

- The variance of a random variable X, denoted Var(X), is defined as

$$Var(X) = E\left[(X - E[X])^2\right]$$

- The standard deviation of X, denoted SD(X), is defined as

$$SD(X) = +\sqrt{Var(X}$$

→ Variance is expected value of the random variable $(X - E[X]^2)$.

- $Var(X) = \sum_{t \in T_X} (t - E[X])^2 \cdot P(X = t)$

- Variance is non-negative, & std. deviation is well defined.

- Units of SD(X) are same as units of X.

→ Intuitively, the more the 'spread' in T_X, the more will be the Var(X).

Example : Throw a die

$X \sim$ Uniform $\{1, 2, 3, 4, 5, 6\}$

→ $E[X] = 3.5$

→ $Var(X) = (1-3.5)^2 \cdot \frac{1}{6} + (2-3.5)^2 \cdot \frac{1}{6} + \ldots + (6-3.5)^2 \cdot 1/6$

$= 35/12 = 2.916$

→ $SD(X) = \sqrt{Var(X)} = \sqrt{2.916} = 1.7078$

Given PMF of X with a small range, variance & std. dev. can be readily computed.

PROPERTIES : Scaling and Translation

Let X be a random variable. Let a be a constant real number.

1. $Var(aX) = a^2 \cdot Var(X)$

2. $SD(aX) = |a| \cdot SD(X)$

3. $Var(X+a) = Var(X)$

4. $SD(X+a) = SD(X)$

Alternative Formula for Variance

The variance of a random variable is given by

$$Var(X) = E[X^2] - E[X]^2$$

\# $E[X] \rightarrow$ first moment
$E[X^2] \rightarrow$ second moment
$Var(X) \rightarrow$ second central moment

SUM & PRODUCT of Independent R.V

For any 2 random variables X & Y (independent or dependent),

$$E[X+Y] = E[X] + E[Y].$$

- Suppose X and Y are <u>independent</u> random variables Then,

(1) $E[XY] = E[X] \cdot E[Y]$

(2) $Var(X+Y) = Var(X) + Var(Y)$

---

(3) $Var(X-Y) = Var(X) + Var(-Y)$
$= Var(X) + (-1)^2 \cdot Var(Y)$
$= Var(X) + Var(Y)$

Variance of Common Distributions

Distribution	Expected Value	Variance
Bernoulli (p)	p	$p(1-p)$
Binomial (n, p)	np	$np(1-p)$
Geometric (p)	$1/p$	$(1-p)/p^2$
Poisson (λ)	λ	λ
Uniform $\{1, \ldots, n\}$	$(n+1)/2$	$(n^2-1)/12$

Standardised Random Variables

Definition: A random variable X is said to be standardised if

$$E[X] = 0 \quad , \quad Var(X) = 1$$

Theorem: Let X be a random variable. Then,

$$Y = \frac{X - E[X]}{SD(X)}$$

is a standardised R.V.

September 24, 2021

Date
Page

Covariance

Definition: Suppose X and Y are random variables on the same probability space. The covariance of X and Y, denoted $Cov(X,Y)$ is defined as

$$Cov(X,Y) = E\left[(X-E[X])(Y-E[Y]) \right]$$

\# $Cov(X,Y)$ is positive.
→ $(X-E[X])(Y-E[Y])$ tends to be positive
→ When X is above / below its average, Y tends to be correspondingly above / below its average.

\# $Cov(X,Y)$ is negative
→ $(X-E[X])(Y-E[Y])$ tends to be negative
→ When X is above / below its average, Y tends to be correspondingly below / above its average.

\# $Cov(X,Y) = 0$
→ X and Y are said to be 'uncorrelated'

Example: Computing Covariance

$Y \backslash X$	-1	0	1	f_Y
-1	$1/15$	$2/15$	$2/15$	$1/3$
0	$2/15$	$1/15$	$2/15$	$1/3$
1	$2/15$	$2/15$	$1/15$	$1/3$
$f(x)$	$1/3$	$1/3$	$1/3$	

$E[X] = 0$
$E[Y] = 0$

$$Cov(X,Y) = E\left[(X - E[X])(Y - E[Y])\right]$$

$$\Rightarrow Cov(X,Y) = (-1)(-1)\,1/15 + (-1)(1)\,2/15 + (-1)(1)\,2/15 + (1)(1)\,1/15$$

$$\Rightarrow Cov(X,Y) = \frac{-2}{15}$$

PROPERTIES

1. $Cov(X,X) = Var(X)$
   → Proof: $Cov(X,X) = E\left[(X - E[X])(X - E[X])\right] = E\left[(X - E[X])^2\right]$

2. $Cov(X,Y) = E[XY] - E[X]\cdot E[Y]$
   → Proof: $Cov(X,Y) = E\left[(X - E[X])(Y - E[Y])\right]$
   $= E\left[XY - X\cdot E[Y] - Y\cdot E[X] + E[X]\cdot E[Y]\right]$

3. Covariance is symmetric, $Cov(X,Y) = Cov(Y,X)$

4. Covariance is a 'linear' quantity
   → $Cov(X, aY + bZ) = a\,Cov(X,Y) + b\,Cov(X,Z)$
   → $Cov(aX + bY, Z) = a\,Cov(X,Z) + b\,Cov(Y,Z)$

COVARIANCE & INDEPENDENCE

1. If X and Y are independent, X and Y are uncorrelated.
   i.e. $Cov(X,Y) = 0$

2. Conversely, not true. If X & Y are uncorrelated, they may be dependent.

Date
Page

Correlation Coefficient

Definition: The correlation coefficient or simply correlation of two random variables X & Y, denoted $\rho(X,Y)$, is defined as

$$\rho(X,Y) = \frac{Cov(X,Y)}{SD(X) \cdot SD(Y)}$$

\# RESULT : $-SD(X) \cdot SD(Y) \leq Cov(X,Y) \leq SD(X) \cdot SD(Y)$

→ simplify $E\left[\left(\frac{X-E[X]}{SD(X)} + \frac{Y-E[Y]}{SD(Y)}\right)^2\right] \geq 0$ to get lower bound.

→ Simplify $E\left[\left(\frac{X-E[X]}{SD(X)} - \frac{Y-E[Y]}{SD(Y)}\right)^2\right] \geq 0$ to get upper bound.

\# Using the above, $-1 \leq \rho(X,Y) \leq 1$

PROPERTIES

1. $\rho(X,Y)$
   - → One no. that summarizes 'trend' between two R.V.
   - → Dimensionless qty.

2. $\rho(X,Y)$ is close to zero
   - → X and Y are close to being uncorrelated
   - → No clear trend b/w X and Y.

3. $\rho(X,Y) = 1$ or $\rho(X,Y) = -1$

→ There exist $a \neq 0$ and b so that $Y = aX + b$ with probability 1

→ Y is linear fn of X.

4. $|\rho(X,Y)|$ is close to one

→ X and Y are strongly correlated

→ increase in X is likely to match up with an increase in Y.

September 25, 2021

classmate
Date
Page

Bounds on Probabilities using mean & variance

NOTATION FOR MEAN & VARIANCE

1. $\mu \rightarrow$ denotes the mean $E[X]$
2. $\sigma^2 \rightarrow$ denotes the variance $Var(X)$
3. $\sigma \rightarrow$ denotes the standard dev. $SD(X)$.

STANDARD UNITS IN STATISTICS

- Consider a random variable X with mean μ and variance σ^2.
- In an experiment, X may take a value that is close to μ or away from μ
- $X-\mu$: measures the distance of X from the mean μ.
  - $\rightarrow$ could be +ve or -ve
- Standard units : The no. of standard deviations that a realization of a random variable is away from the mean.

→ We expect $X-\mu$ to fall between $-c\sigma$ and $c\sigma$ for a small value of c.

→ In other words, we expect X to fall between $\mu - c\sigma$ and $\mu + c\sigma$.

MARKOV's INEQUALITY

Theorem : Let X be a discrete random variable taking non-negative values with a finite μ. Then,

$$P(X \geq c) \leq \frac{\mu}{c}$$

Proof :

$$\mu = \sum_{t \in T_X} t \cdot P(X=t) = \sum_{t<c} t \cdot P(X=t) + \sum_{t \geq c} t \cdot P(X=t)$$

Since the first sum is non-negative,

$$\mu \geq \sum_{t \geq c} t \cdot P(X=t) \geq \sum_{t \geq c} c \cdot P(X=t) = c \sum_{t \geq c} P(X=t) = c \cdot P(X \geq c)$$

$$\Rightarrow \mu \geq c \cdot P(X \geq c) \Rightarrow P(X \geq c) \leq \mu/c$$

CHEBYSHEV's INEQUALITY

Theorem : Let X be a discrete R.V. with a finite mean μ and a finite variance σ^2. Then,

$$P(|X-\mu| \geq k\sigma) \leq 1/k^2$$

Proof: Apply Markov's inequality to $(X-\mu)^2$.

Other forms:

$\Rightarrow$ $P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}$, $P((X-\mu)^2 \geq k^2\sigma^2) \leq 1/k^2$

$\Rightarrow$ $P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - 1/k^2$

$\Rightarrow$ $P(X \geq \mu + k\sigma) + P(X \leq \mu - k\sigma) \leq 1/k^2$

$\rightarrow$ $P(X \geq \mu + k\sigma) \leq \frac{1}{k^2}$, $P(X \leq \mu - k\sigma) \leq \frac{1}{k^2}$

COMPARE ACTUAL & CHEBYSHEV

Let X be a random variable with finite mean μ and variance σ^2.

$$P(|X-\mu| \geq 2\sigma) \leq 1/4$$

(1) $X \sim \text{Binomial}(10, 0.5)$, $\mu = 5$, $\sigma = \sqrt{2.5} \approx 1.58$

$$P(|X-5| \geq 2\sigma) = P(X \in \{0,1,9,10\}) \approx 0.021$$

(2) $X \sim \text{Geometric}(1/4)$, $\mu = 4$, $\sigma \approx 3.46$

$$P(|X-4| \geq 2\sigma) = P(X \in \{11, 12, \ldots\}) \approx 0.056$$

WHAT DO MEAN & VARIANCE SAY ABOUT DISTRIBUTION?

- Mean μ, through Markov's inequality : bounds the probability that a non-negative random variable takes values much larger than the mean.

- Mean μ and standard dev. σ, through Chebyshev's inequality : bound the probability that X is away from μ by $k\sigma$.

- These are some of the most useful measures of expected centre and spread in practice.

→ Suppose the average no. of accidents decreases by 10000 per day across the country. Is that a 'significant' decrease?

→ If std. dev. of the no. of accidents is ~~un~~known, we can find how high 10000 is in terms of std. dev. to answer above question.

.