Week 4: Expected value

Subsection 1

Introduction

Summarizing data

- How to provide summary information about a large set of data?
  - Minimum value, maximum value, average value, median value
- Average value of a set of data is very useful in practice.
  - Average marks in a class exam
  - Run rate, batting average in cricket (and similar numbers in many sports)
- What does the average represent?
  - One number to represent a large set of numbers
  - Useful in comparisons and in so many scenarios
- Probability theory: expected value of a random variable is an important theoretical construct that represents the average value
  - It is used to express various types of summary information

Example: Casino math

Place bets on the sum seen when two die are rolled.

- Bets and returns
  - Under 7: get money back
  - Over 7: get money back
  - Equal to 7: get 4 times the money back

Example: Casino math

Place bets on the sum seen when two die are rolled.

- Bets and returns
  - Under 7: get money back
  - Over 7: get money back
  - Equal to 7: get 4 times the money back

Suppose you bet 1 unit of money.

Bet	Gain	Prob
Under 7	0	5/12
	-1	7/12
Over 7	0	5/12
	-1	7/12
Equal 7	$+3$	1/6
	-1	5/6

Example: Casino math

Place bets on the sum seen when two die are rolled.

- Bets and returns
  - Under 7: get money back
  - Over 7: get money back
  - Equal to 7: get 4 times the money back

Suppose you bet 1 unit of money.

Bet	Gain	Prob
Under 7	0	$5/12$
	-1	$7/12$
Over 7	0	$5/12$
	-1	$7/12$
Equal 7	$+3$	$1/6$
	-1	$5/6$

How did the casino decide on above returns? How to bet?

Expected value of a discrete random variable

Definition (Expected value)

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

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

assuming above sum exists.

Expected value of a discrete random variable

Definition (Expected value)

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

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

assuming above sum exists.

- Other names: mean of X, ~~average value~~ 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. $X \sim \text{Bernoulli}(p)$

$$X \in \{\overset{1-p}{0}, \overset{p}{1}\}$$

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

Examples: Easy summation given PMF

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

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

$1/6 \quad 1/6 \quad 1/6 \quad 1/6 \quad 1/6 \quad 1/6$

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

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

Examples: Easy summation given PMF

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

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

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

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

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

$$E[X] = 200\cdot\frac{1}{1000} + 20\cdot\frac{27}{1000} + 0\cdot\frac{972}{1000} = \text{Rs. } 0.56$$

Examples: Easy summation given PMF

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

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

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

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

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

$$E[X] = 200 \cdot \frac{1}{1000} + 20 \cdot \frac{27}{1000} + 0 \cdot \frac{972}{1000} = \text{Rs. } 0.56$$

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

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

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

$\frac{1}{b-a+1}$

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

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

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

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

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

How to simplify sum?

Identity: $a + (a+1) + \cdots + b = (b-a+1)\,\frac{a+b}{2}$

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

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

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

How to simplify sum?

Identity: $a + (a+1) + \cdots + b = (b-a+1)\,\frac{a+b}{2}$

$$E[X] = \frac{a+b}{2}$$

Examples: More involved summation

$(1-p)^{k-1}p$

1. $X \sim \text{Geometric}(p)$ $\quad X \in \{1, 2, \ldots, k, \ldots\}$

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

$$= p + 2(1-p)p + 3(1-p)^2 p + \cdots$$

Examples: More involved summation

1. $X \sim \text{Geometric}(p)$

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

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

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

Examples: More involved summation

1. $X \sim \text{Geometric}(p)$

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

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

$$E[X] = \sum_{t=0}^{\infty} t\ e^{-\lambda}\frac{\lambda^t}{t!}$$

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

$$E[X] = \sum_{t=0}^{n} t\ \overbrace{\binom{n}{t} p^t(1-p)^{n-t}}^{P(X=t)}$$

$$nC_t \leftrightarrow \binom{n}{t} = \frac{n!}{t!(n-t)!}$$

How to simplify sum?

1. Difference Equation (DE): $a_{t+1} - r\,a_t = \underbrace{b_t}_{\text{another sequence}}$ $(r \neq 1)$

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

How to simplify sum?

1. Difference Equation (DE): $a_{t+1} - r\,a_t = b_t$ $(r \neq 1)$

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

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

$$\sum_{t=1}^{n} a_t = \frac{a_1 - ra_n}{1-r} \quad \overset{|r|<1}{\underset{n\to\infty}{\Longrightarrow}} \quad \frac{a_1}{1-r}$$

How to simplify sum?

1. Difference Equation (DE): $a_{t+1} - r\,a_t = b_t$ $(r \neq 1)$

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

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

$$\sum_{t=1}^{n} a_t = \frac{a_1 - ra_n}{1-r} \quad \underset{n\to\infty}{\overset{|r|<1}{\Longrightarrow}} \quad \frac{a_1}{1-r}$$

3. Exponential Function:

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

How to simplify sum?

1. Difference Equation (DE): $a_{t+1} - r\,a_t = b_t$ $(r \neq 1)$

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

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

$$\sum_{t=1}^{n} a_t = \frac{a_1 - ra_n}{1-r} \quad \underset{n\to\infty}{\overset{|r|<1}{\Longrightarrow}} \quad \frac{a_1}{1-r}$$

3. Exponential Function:

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

4. Binomial formula: $\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = (a+b)^n$

$a = p$, $b = 1-p \Rightarrow \overset{RHS=1}{\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}} = 1$

Examples: More involved summation

1. $X \sim \text{Geometric}(p)$ (DE: $a_1 = p$, $r = 1 - p$, $b_t = r^t p$)

$E[X] = 1/p$

$$E[X] = \sum_{t=1}^{\infty} \underbrace{t(1-p)^{t-1}p}_{a_t} = 1/p$$

$a_{t+1} = (t+1)(1-p)^t p$

$a_{t+1} - (1-p)a_t = (1-p)^t p$

Examples: More involved summation

1. $X \sim$ Geometric(p) (DE: $a_1 = p$, $r = 1-p$, $b_t = r^t p$)

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

2. $X \sim$ Poisson(λ): $E[X] = \sum_{t=0}^{\infty} t \ e^{-\lambda} \frac{\lambda^t}{t!}$

$E[X]=\lambda$

$$E[X] = e^{-\lambda}\left(0 \cdot + \cancel{1} \cdot \frac{\lambda}{\cancel{1}!} + \cancel{2} \cdot \frac{\lambda^2}{\cancel{2}!_{1!}} + \cancel{3} \cdot \frac{\lambda^3}{\cancel{3}!_{2!}} + \cdots\right)$$

$$= \lambda e^{-\lambda}\left(\underbrace{1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \cdots}_{e^{\lambda}}\right) = \lambda$$

Examples: More involved summation

1. $X \sim$ Geometric(p) (DE: $a_1 = p$, $r = 1 - p$, $b_t = r^t p$)

$E[X] = 1/p$

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

2. $X \sim$ Poisson(λ): $E[X] = \sum_{t=0}^{\infty} t \; e^{-\lambda} \frac{\lambda^t}{t!}$

$E[X] = \lambda$

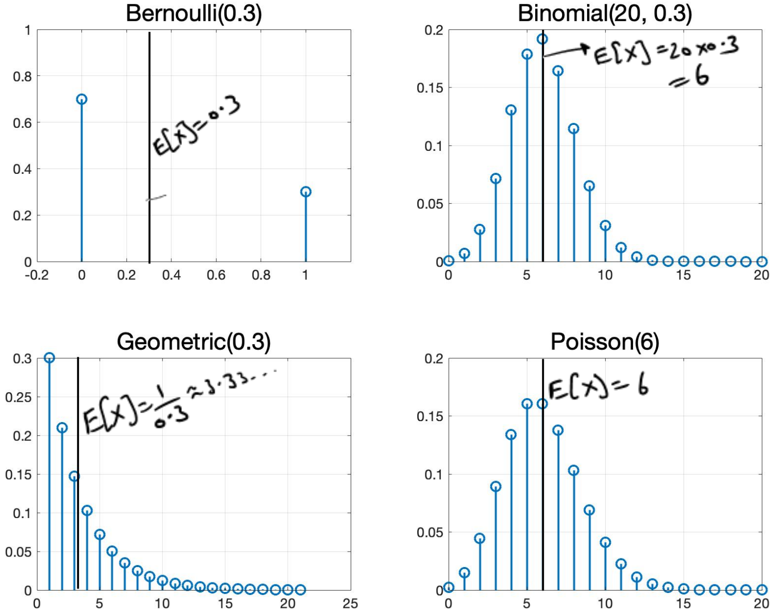
$$E[X] = e^{-\lambda}\left(0 \cdot +1 \cdot \frac{\lambda}{1!} + 2 \cdot \frac{\lambda^2}{2!} + 3 \cdot \frac{\lambda^3}{3!} + \cdots\right)$$
$$= \lambda e^{-\lambda}\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \cdots\right) = \lambda$$

3. $X \sim$ Binomial(n, p): $E[X] = \sum_{t=0}^{n} t \binom{n}{t} p^t (1-p)^{n-t}$

$E[X] = np$

$$E[X] = \sum_{t=1}^{n} \frac{\cancel{t}\; n!}{\underset{(t-1)!}{\cancel{t!}}(n-t)!} p^t (1-p)^{n-t} = np \sum_{t=1}^{n} \frac{(n-1)!}{(t-1)!(n-t)!} p^{t-1}(1-p)^{n-t}$$
$$= np\underline{(p + (1-p))^{n-1}} = np$$

PMF and expected value



Subsection 2

Properties of expected value

Example: Casino math

Bet	Gain	Prob
Under 7	0	5/12
	−1	7/12
Over 7	0	5/12
	−1	7/12
Equal 7	+3	1/6
	−1	5/6

Suppose you bet 1 unit of money with probability p_1 each on "Under 7" or "Over 7" and probability $1 - 2p_1$ on "Equal to 7". What is the expected gain?

$$\overbrace{p_1\cdot\frac{5}{12}\times 0 + p_1\cdot\frac{7}{12}\times(-1)}^{\text{Under 7}} + \overbrace{p_1\cdot\frac{5}{12}\times 0 + p_1\cdot\frac{7}{12}\times(-1)}^{\text{Over 7}} + (1-2p_1)\cdot\frac{1}{6}\cdot 3 + (1-2p_1)\cdot\frac{5}{6}(-1)$$

$$= -\frac{7p_1}{6} - (1-2p_1)\cdot\frac{2}{6} = \frac{-7p_1-2+4p_1}{6} = \frac{-(2+3p_1)}{6}$$

Negative

Constant random variable and positive random variable

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

Constant random variable and positive random variable

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

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

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

Constant random variable and positive random variable

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

1. Consider a constant c as a 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,

$$E[X] \geq 0$$

Expected value of a function of random variables

Theorem (Expected value of a function)

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

$$E[g(X_1, \ldots, X_n)] = \overbrace{\sum_{t \in T_Y} t\, f_Y(t)}^{E[Y]} = \sum_{t_i \in T_{X_i}} g(t_1, \ldots, t_n) f_{X_1 \cdots X_n}(t_1, \ldots, t_n).$$

Expected value of a function of random variables

Theorem (Expected value of a function)

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

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

- We have seen how to find f_Y, PMF of a function of multiple random variables
- The above theorem says: To find $E[Y]$, you do not always need f_Y. The joint PMF of $X_1, \ldots, X_n$ can be used directly.
- Simple sounding property with far-reaching consequences!

Examples for illustration

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

Examples for illustration

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

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

Examples for illustration

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

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

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

$g(x)$ $\quad$ $P(X=-2)$ $\quad$ $\sum_t t^2 P(X=t)$

Examples for illustration

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

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

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

2. $(X, Y) \sim \text{Uniform}\{\overset{1/6}{(0,0)}, \overset{1/6}{(1,0)}, \overset{1/6}{(0,1)}, \overset{1/6}{(1,1)}, \overset{1/6}{(-1,1)}, \overset{1/6}{(1,-1)}\}$

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

x	y	$g(x,y)$	Pr
0	0	0	1/6
1	0	1	1/6
0	1	1	1/6
1	1	3	1/6
-1	1	1	1/6
1	-1	1	1/6

Examples for illustration

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

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

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

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

$$E[g(X, Y)] = 0 \cdot \frac{1}{6} + 1 \cdot \frac{4}{6} + 3 \cdot \frac{1}{6} = 7/6$$

Examples for illustration

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

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

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

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

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

$$E[g(X, Y)] = 0 \cdot \tfrac{1}{6} + 1 \cdot \tfrac{4}{6} + 3 \cdot \tfrac{1}{6} = 7/6$$

$g(1,-1)$ $P(X=1, Y=-1)$

$$E[g(X, Y)] = 0 \cdot \tfrac{1}{6} + (1) \cdot \tfrac{1}{6} + (1) \cdot \tfrac{1}{6} + (3) \cdot \tfrac{1}{6} + (1) \cdot \tfrac{1}{6} + (1) \cdot \tfrac{1}{6} = 7/6$$

$g(0,0)$ $P(X=0, Y=0)$

Linearity of expected value

1. $E[cX] = cE[X]$ for a random variable X and a constant c

Proof:

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

Linearity of expected value

1. $E[cX] = cE[X]$ for a random variable X and a constant c

Proof:

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

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

Proof:

$$\begin{aligned} E[X+Y] &= \sum_{t_1 \in T_X, t_2 \in T_Y} (t_1 + t_2) f_{XY}(t_1, t_2) \\ &= \sum_{t_1 \in T_X, t_2 \in T_Y} t_1\, f_{XY}(t_1, t_2) + \sum_{t_1 \in T_X, t_2 \in T_Y} t_2\, f_{XY}(t_1, t_2) \\ &= E[X] + E[Y] \end{aligned}$$

Linearity of expected value

1. $E[cX] = cE[X]$ for a random variable X and a constant c

Proof:

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

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

Proof:

$$\begin{aligned} E[X+Y] &= \sum_{t_1 \in T_X, t_2 \in T_Y} (t_1 + t_2) f_{XY}(t_1, t_2) \\ &= \sum_{t_1 \in T_X, t_2 \in T_Y} t_1\, f_{XY}(t_1, t_2) + \sum_{t_1 \in T_X, t_2 \in T_Y} t_2\, f_{XY}(t_1, t_2) \\ &= E[X] + E[Y] \end{aligned}$$

$$E[aX + bY] = aE[X] + bE[Y]$$

One of the most useful properties of expected value!

Illustrative examples

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

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

Illustrative examples

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

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

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

Illustrative examples

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

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

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

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

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

Illustrative examples

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

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

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

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

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

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

Illustrative examples

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

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

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

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

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

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

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

Illustrative examples

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

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

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

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

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

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

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

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.

Problem: Throw a fair die twice

What is the expected value of the sum of the two numbers seen?

X: first number $\quad$ Y: second number

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

Problem: Expected value of Binomial(n, p)

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

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

- Such summations may be difficult to simplify

Problem: Expected value of Binomial(n, p)

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

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

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

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(p). Then, $E[X_i] = p$ and

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

Problem: Expected value of Binomial(n, p)

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

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

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

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli(p). Then, $E[X_i] = p$ and

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

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

Zero-mean random variable

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

Zero-mean random variable

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

Translation of a random variable

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

- Range of $X + c$ is $\{t + c : t \in T_X\}$, which is a translated version of T_X, the range of X
- $P(X + c = t + c) = P(X = t)$ and the PMF is "translated" as well

Zero-mean random variable

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

$E[X+c] = E[X] + c$

Translation of a random variable

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

- Range of $X + c$ is $\{t + c : t \in T_X\}$, which is a translated version of T_X, the range of X
- $P(X + c = t + c) = P(X = t)$ and the PMF is "translated" as well

Translation by expected value

Important: $E[X]$ is a constant.

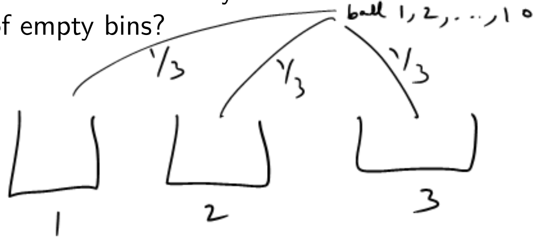
$Y = X - E[X]$ is a translated version of X and $E[Y] = 0$.

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

"centering"

Problem: Balls on bins

Suppose 10 balls are thrown uniformly at random into 3 bins. What is the expected number of empty bins?



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

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

$$Y = \#\text{ empty bins} = X_1 + X_2 + X_3$$

"Finding f_Y is a little hard"

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

Subsection 3

Variance

Motivation

How good is expected value in describing a random variable?

- Let $X \sim \{\overset{1}{10}\}$, $Y \sim \{\overset{1/2}{9}, \overset{1/2}{11}\}$, $Z \sim \{\overset{1/2}{0}, \overset{1/2}{20}\}$.

$$E[X] = E[Y] = E[Z] = 10$$

However, the three random variables are quite different!

Motivation

How good is expected value in describing a random variable?

- Let $X \sim \{\overset{1}{10}\}$, $Y \sim \{\overset{1/2}{9}, \overset{1/2}{11}\}$, $Z \sim \{\overset{1/2}{0}, \overset{1/2}{20}\}$.

$$E[X] = E[Y] = E[Z] = 10$$

However, the three random variables are quite different!

Center and spread

- Expected value: represents "center" of a random variable
- We need some indicator of "spread" about the expected value

Variance and standard deviation

Definition (Variance and standard deviation)

The variance of a random variable X, denoted $\mathrm{Var}(X)$, is defined as

$$\mathrm{Var}(X) = E[\underbrace{(X - \overbrace{E[X]}^{\text{some constant}})^2}_{\text{function of } X}].$$

The standard deviation of X, denoted $\mathrm{SD}(X)$, is defined as

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

Variance and standard deviation

Definition (Variance and standard deviation)

The variance of a random variable X, denoted $\mathrm{Var}(X)$, is defined as

$$\mathrm{Var}(X) = E[(X - E[X])^2].$$

The standard deviation of X, denoted $\mathrm{SD}(X)$, is defined as

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

- Variance is expected value of the random variable $(X - E[X])^2$.
  - $\mathrm{Var}(X) = \sum_{t \in T_X} (t - E[X])^2 P(X = t)$
  - Variance is non-negative, and standard deviation is well-defined.
  - Units of $\mathrm{SD}(X)$ are same as units of X.
- Intuitively, the more the "spread" in T_X, the more will be $\mathrm{Var}(X)$

Example

$X \sim \{\overset{1}{10}\}$, $Y \sim \{\overset{1/2}{9}, \overset{1/2}{11}\}$, $Z \sim \{\overset{1/2}{0}, \overset{1/2}{20}\}$

$$E[X] = E[Y] = E[Z] = 10$$

Variances and standard deviations are different!

Example

$$X \sim \{\overset{1}{10}\},\ Y \sim \{\overset{1/2}{9}, \overset{1/2}{11}\},\ Z \sim \{\overset{1/2}{0}, \overset{1/2}{20}\}$$

$$E[X] = E[Y] = E[Z] = 10$$

Variances and standard deviations are different!

$$\mathrm{Var}(X) = (\overset{t \in T_X}{10} - \overset{E[X]}{10})^2 \cdot 1 = 0,\ \mathrm{SD}(X) = 0$$

Example

$$X \sim \{\overset{1}{10}\},\ Y \sim \{\overset{1/2}{9}, \overset{1/2}{11}\},\ Z \sim \{\overset{1/2}{0}, \overset{1/2}{20}\}$$

$$E[X] = E[Y] = E[Z] = 10$$

Variances and standard deviations are different!

$$\mathrm{Var}(X) = (10 - 10)^2 \cdot 1 = 0,\ \mathrm{SD}(X) = 0$$

$t \in T_Y$ $E[Y]$ $P(Y=t)$

$$\mathrm{Var}(Y) = (9 - 10)^2 \cdot \frac{1}{2} + (11 - 10)^2 \cdot \frac{1}{2} = 1,\ \mathrm{SD}(X) = 1$$

Example

$X \sim \{\overset{1}{10}\},\ Y \sim \{\overset{1/2}{9}, \overset{1/2}{11}\},\ Z \sim \{\overset{1/2}{0}, \overset{1/2}{20}\}$

$$E[X] = E[Y] = E[Z] = 10$$

Variances and standard deviations are different!

$$\mathrm{Var}(X) = (10 - 10)^2 \cdot 1 = 0,\ \mathrm{SD}(X) = 0$$

$$\mathrm{Var}(Y) = (9 - 10)^2 \cdot \frac{1}{2} + (11 - 10)^2 \cdot \frac{1}{2} = 1,\ \mathrm{SD}(X) = 1$$

$$\mathrm{Var}(Z) = (0 - 10)^2 \cdot \frac{1}{2} + (20 - 10)^2 \cdot \frac{1}{2} = 100,\ \mathrm{SD}(X) = 10$$

Example: Throw a die

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

$$E[X] = 3.5$$

Example: Throw a die

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

$$E[X] = 3.5$$

$$\begin{aligned}
\text{Var}(X) &= (1-3.5)^2 \cdot \frac{1}{6} + (2-3.5)^2 \cdot \frac{1}{6} + (3-3.5)^2 \cdot \frac{1}{6} \\
&\quad + (4-3.5)^2 \cdot \frac{1}{6} + (5-3.5)^2 \cdot \frac{1}{6} + (6-3.5)^2 \cdot \frac{1}{6} \\
&= \frac{35}{12} = 2.916\ldots \\
\text{SD}(X) &= 1.7078\ldots
\end{aligned}$$

$\sqrt{\frac{35}{12}}$

Example: Throw a die

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

$$E[X] = 3.5$$

$$\begin{aligned}
\text{Var}(X) &= (1-3.5)^2 \cdot \frac{1}{6} + (2-3.5)^2 \cdot \frac{1}{6} + (3-3.5)^2 \cdot \frac{1}{6} \\
&\quad + (4-3.5)^2 \cdot \frac{1}{6} + (5-3.5)^2 \cdot \frac{1}{6} + (6-3.5)^2 \cdot \frac{1}{6} \\
&= \frac{35}{12} = 2.916\ldots \\
\text{SD}(X) &= 1.7078\ldots
\end{aligned}$$

Given PMF of X with a small range, variance and standard deviation can be readily computed.

Properties: Scaling and translation

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

1. $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$
2. $\mathrm{SD}(aX) = |a|\mathrm{SD}(X)$
3. $\mathrm{Var}(X + a) = \mathrm{Var}(X)$
4. $\mathrm{SD}(X + a) = \mathrm{SD}(X)$

Pf. $E\left[(X+a-\underbrace{E[X+a]})^2\right]$

$= E\left[(X+\cancel{a}-E[X]-\cancel{a})^2\right]$

Properties: Scaling and translation

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

1. $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$
2. $\mathrm{SD}(aX) = |a|\mathrm{SD}(X)$
3. $\mathrm{Var}(X + a) = \mathrm{Var}(X)$
4. $\mathrm{SD}(X + a) = \mathrm{SD}(X)$

- Contrast with similar properties for $E[X]$

Alternative formula for variance

Theorem

The variance of a random variable X is given by

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

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

$$E[X]^2 = \left(\sum_{t \in T_X} t P(X=t) \right)^2$$

Alternative formula for variance

Theorem

The variance of a random variable X is given by

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

Proof:

- $\text{Var}(X) = E[(X - E[X])^2] = E[X^2 - 2E[X]X + E[X]^2]$
- By linearity of expected value,

$$\text{Var}(X) = E[X^2] - 2E[E[X]X] + E[E[X]^2]$$

- Since $E[X]$ is a constant,

$$\text{Var}(X) = E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2$$

Alternative formula for variance

Theorem

The variance of a random variable X is given by

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

Proof:

- $\mathrm{Var}(X) = E[(X - E[X])^2] = E[X^2 - 2E[X]X + E[X]^2]$
- By linearity of expected value,

$$\mathrm{Var}(X) = E[X^2] - 2E[E[X]X] + E[E[X]^2]$$

- Since $E[X]$ is a constant,

$$\mathrm{Var}(X) = E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2$$

$E[X]$, $E[X^2]$: first and second moment, $\mathrm{Var}(X)$: second central moment

Sum and product of independent random variables

For any two random variables X and Y (independent or dependent),

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

Sum and product of independent random variables

For any two random variables X and Y (independent or dependent),

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

Suppose X and Y are **independent** random variables. Then, more is true.

1. $E[XY] = E[X]E[Y]$

Sum and product of independent random variables

For any two random variables X and Y (independent or dependent),

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

Suppose X and Y are **independent** random variables. Then, more is true.

1. $E[XY] = E[X]E[Y]$
2. $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Proof

$$E[XY] = \sum_{\substack{t_1 \in T_X,\\ t_2 \in T_Y}} t_1 t_2 \underbrace{P(X=t_1, Y=t_2)}_{X \,\&\, Y \text{ indep: }} = P(X=t_1)P(Y=t_2)$$

$$= \sum_{t_1 \in T_X} \left(\sum_{t_2 \in T_Y} t_1 t_2 P(X=t_1) P(Y=t_2) \right)$$

over t_2

factoring IS crucial

$$= \sum_{t_1 \in T_X} \left(t_1 P(X=t_1)\right) \underbrace{\left(\sum_{t_2 \in T_Y} t_2 P(Y=t_2) \right)}_{E[Y]}$$

$$= E[Y] \underbrace{\left(\sum_{t_1 \in T_X} t_1 P(X=t_1) \right)}_{E[X]}$$

$$= E[X]E[Y] \longrightarrow$$ Not always true if X & Y are dependent

Sum and product of independent random variables

For any two random variables X and Y (independent or dependent),

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

Suppose X and Y are **independent** random variables. Then, more is true.

1. $E[XY] = E[X]E[Y]$
2. $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ Pf: Exercise.

Example: Sum of two dice (independent)

What is the variance of the sum of two dice?

X: first die $\quad Y$: second die

X & Y: independent

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

$$= 2\sqrt{\frac{35}{12}} = \sqrt{\frac{35}{3}}$$

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$\{1,\ldots,n\}$	$(n+1)/2$	$(n^2-1)/12$

① $X \sim \left\{ \overset{1-p}{0}, \overset{p}{1} \right\}$ $E[X] = p$ $Var(X) = p - p^2$

$E[X^2] = 1 \cdot p = p$ $= p(1-p)$

② $X \sim \text{Binomial}(n,p)$ $E[X] = np$

$X = X_1 + \cdots + X_n$ $Var(X) = Var(X_1) + \cdots + Var(X_n)$

$X_i \sim \underline{\underline{iid}}$ Bernoulli(p) $= np(1-p)$ (more involved)

$$E[X^2] = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = \underbrace{np(1-p)}_{Var(X)} + \underbrace{(np)^2}_{E[X]^2}$$

Working

Standardised random variables

Definition

A random variable X is said to be standardised if

$$E[X] = 0, \mathrm{Var}(X) = 1.$$

Standardised random variables

Definition

A random variable X is said to be standardised if

$$E[X] = 0, \mathrm{Var}(X) = 1.$$

Theorem

Let X be a random variable. Then,

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

is a standardised random variable.

$$E[Y] = \frac{1}{SD(X)}\left(E[X] - E[X]\right) = 0 \qquad \mathrm{Var}(Y) = \mathrm{Var}\Bigg(\frac{X}{\underbrace{SD(X)}_{\text{constant}}}\Bigg) = \frac{\mathrm{Var}(X)}{SD(X)^2} = 1$$

→ PMF of Y is "very similar" to PMF of X.

Existence of expected value and variance

- There are random variables s.t. $E[X]$ goes to ∞ (or $-\infty$)
  - $X \sim \{\overset{1/2}{1}, \overset{1/4}{2}, \overset{1/8}{4}, \ldots, \overset{1/2^n}{2^{n-1}}, \ldots\}$
  - $E[X^2]$ will go to ∞ in this case and $\mathrm{Var}(X)$ is ill-defined
- There are random variables s.t. $E[X]$ is not well-defined
  - $X \sim \{\overset{1/2}{1}, \overset{1/4}{-2}, \overset{1/8}{4}, \ldots, \overset{1/2^n}{(-2)^{n-1}}, \ldots\}$
  - $E[X^2]$ will go to ∞ in this case and $\mathrm{Var}(X)$ is ill-defined
- There are random variables s.t. $E[X]$ is finite, but $E[X^2]$ goes to ∞
  - $\mathrm{Var}(X)$ goes to ∞ too $\quad X \in \{1, 2, 3, \ldots\} \quad P(X=k) \propto \frac{1}{k^3}$
- In this course, we will generally consider well-behaved random variables with finite mean and variance

Subsection 4

Covariance and correlation

Motivation

Consider the following two joint PMFs for two random variables X and Y.

f_{XY}	$X=0$	$X=1$	f_Y
$Y=0$	$1/\cancel{2}4$	$1/\cancel{2}4$	$1/2$
$Y=1$	$1/\cancel{2}4$	$1/\cancel{2}4$	$1/2$
f_X	$1/2$	$1/2$	

f_{XY}	0	1	f_Y
0	0	$1/2$	$1/2$
1	$1/2$	0	$1/2$
f_X	$1/2$	$1/2$	

Motivation

Consider the following two joint PMFs for two random variables X and Y.

f_{XY}	0	1
0	$1/\cancel{2}4$	$1/\cancel{2}4$
1	$1/\cancel{2}4$	$1/\cancel{2}4$

f_{XY}	0	1
0	0	1/2
1	1/2	0

- Same marginal PMFs in both cases. So, mean and variance are the same for X and Y.
- However, the two cases are very different
  - X and Y are independent in one case
  - Value of X determines the value of Y in the other

Motivation

Consider the following two joint PMFs for two random variables X and Y.

f_{XY}	0	1
0	1/2	1/2
1	1/2	1/2

f_{XY}	0	1
0	0	1/2
1	1/2	0

- Same marginal PMFs in both cases. So, mean and variance are the same for X and Y.
- However, the two cases are very different
  - X and Y are independent in one case
  - Value of X determines the value of Y in the other

How to summarize relationship between random variables?

Covariance

Definition (Covariance)

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

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

Covariance

Definition (Covariance)

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

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

- $\mathrm{Cov}(X, Y)$ 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

Covariance

Definition (Covariance)

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

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

- $\mathrm{Cov}(X, Y)$ 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
- $\mathrm{Cov}(X, Y)$ 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

Covariance

Definition (Covariance)

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

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

- $\mathrm{Cov}(X, Y)$ 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
- $\mathrm{Cov}(X, Y)$ 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
- $\mathrm{Cov}(X, Y) = 0$
  - X and Y are said to be "uncorrelated"

Examples: positive and negative covariance

1. X: height of a person, Y: weight of a person
    - A higher value of X tends to result in a higher value for Y
    - We expect $\text{Cov}(X, Y)$ to be positive

Examples: positive and negative covariance

1. X: height of a person, Y: weight of a person
   - A higher value of X tends to result in a higher value for Y
   - We expect $\text{Cov}(X, Y)$ to be positive
2. X: Rainfall during monsoon, Y: Farmer debt
   - A higher value of X tends to result in a lower value for Y
   - We expect $\text{Cov}(X, Y)$ to be negative

Examples: positive and negative covariance

1. X: height of a person, Y: weight of a person
   - A higher value of X tends to result in a higher value for Y
   - We expect $\text{Cov}(X, Y)$ to be positive
2. X: Rainfall during monsoon, Y: Farmer debt
   - A higher value of X tends to result in a lower value for Y
   - We expect $\text{Cov}(X, Y)$ to be negative
3. X: Runs in an IPL over, Y: Wickets in the same over
   - We expect $\text{Cov}(X, Y)$ to be negative

Example: Computing covariance

	$X=-1$	$X=0$	$X=1$	f_Y
$Y=-1$	$1/15$	$2/15$	$2/15$	$1/3$
$Y=0$	$2/15$	$1/15$	$2/15$	$1/3$
$Y=1$	$2/15$	$2/15$	$1/15$	$1/3$
f_X	$1/3$	$1/3$	$1/3$	

$E[Y]=0$

$E[X]=0$

$$\mathrm{Cov}(X,Y)=E\big[(X-E[X])(Y-E[Y])\big] = (-1)(-1)\cdot\frac{1}{15}+(-1)(1)\cdot\frac{2}{15}$$

$$+(-1)(1)\cdot\frac{2}{15}+(1)(1)\cdot\frac{1}{15}$$

$$=\frac{-2}{15}$$

Properties

1. $\text{Cov}(X, X) = \text{Var}(X)$
   - Proof: $\text{Cov}(X, X) = E[(X - E[X])(X - E[X])]$

Properties

1. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
   - Proof: $\mathrm{Cov}(X, X) = E[(X - E[X])(X - E[X])]$
2. $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$
   - Proof: $\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY - XE[Y] - YE[X] + E[X]E[Y]]$

Properties

1. $\text{Cov}(X, X) = \text{Var}(X)$
   - Proof: $\text{Cov}(X, X) = E[(X - E[X])(X - E[X])]$
2. $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$
   - Proof: $\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY - XE[Y] - YE[X] + E[X]E[Y]]$
3. Covariance if symmetric, $\text{Cov}(X, Y) = \text{Cov}(Y, X)$

Properties

1. $\text{Cov}(X, X) = \text{Var}(X)$
   - Proof: $\text{Cov}(X, X) = E[(X - E[X])(X - E[X])]$
2. $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$
   - Proof: $\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY - XE[Y] - YE[X] + E[X]E[Y]]$
3. Covariance if symmetric, $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
4. Covariance is a "linear" quantity
   1. $\text{Cov}(X, aY + bZ) = a\text{Cov}(X, Y) + b\text{Cov}(X, Z)$
   2. $\text{Cov}(aX + bY, Z) = a\text{Cov}(X, Z) + b\text{Cov}(Y, Z)$

Covariance and independence

1. If X and Y are independent, X and Y are uncorrelated, i.e. $\text{Cov}(X, Y) = 0$
   - Proof: $\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0$

Covariance and independence

1. If X and Y are independent, X and Y are uncorrelated, i.e. $\mathrm{Cov}(X, Y) = 0$
   - Proof: $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0$
2. If X and Y are uncorrelated, they may be dependent.

Example:

Dependent, but uncorrelated

	$X=-1$	$X=0$	$X=1$	f_Y
$Y=0$	$1/8$	$1/4$	$1/8$	$1/2$
$Y=1$	$1/4$	$0 \neq \frac{1}{2}\cdot\frac{1}{4}$	$1/4$	$1/2$
f_X	$3/8$	$1/4$	$3/8$	

$E[Y] = 1/2$

$E[X] = 0$

$E[XY] = 0$

$E[XY] = E[X]E[Y]$

Correlation coefficient

Definition (Correlation coefficient)

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

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

Correlation coefficient

Definition (Correlation coefficient)

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

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

- Result: $-\text{SD}(X)\text{SD}(Y) \leq \text{Cov}(X, Y) \leq \text{SD}(X)\text{SD}(Y)$
  - Simplify $E\left[\left(\frac{X-E[X]}{\text{SD}(X)} + \frac{Y-E[Y]}{\text{SD}(Y)}\right)^2\right] \geq 0$ to get lower bound
  - Simplify $E\left[\left(\frac{X-E[X]}{\text{SD}(X)} - \frac{Y-E[Y]}{\text{SD}(Y)}\right)^2\right] \geq 0$ to get upper bound

Correlation coefficient

Definition (Correlation coefficient)

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

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

- Result: $-\text{SD}(X)\text{SD}(Y) \leq \text{Cov}(X, Y) \leq \text{SD}(X)\text{SD}(Y)$
  - Simplify $E\left[\left(\frac{X - E[X]}{\text{SD}(X)} + \frac{Y - E[Y]}{\text{SD}(Y)}\right)^2\right] \geq 0$ to get lower bound
  - Simplify $E\left[\left(\frac{X - E[X]}{\text{SD}(X)} - \frac{Y - E[Y]}{\text{SD}(Y)}\right)^2\right] \geq 0$ to get upper bound
- Using the above, $-1 \leq \rho(X, Y) \leq 1$

Correlation and random variables

1. $\rho(X, Y)$
   - One number that summarizes "trend" between two random variables
   - Dimensionless quantity.

Correlation and random variables

1. $\rho(X, Y)$
   - One number that summarizes "trend" between two random variables
   - Dimensionless quantity.
2. $\rho(X, Y)$ is close to zero
   - X and Y are close to being uncorrelated
   - No clear trend between X and Y

Correlation and random variables

1. $\rho(X, Y)$
   - One number that summarizes "trend" between two random variables
   - Dimensionless quantity.
2. $\rho(X, Y)$ is close to zero
   - X and Y are close to being uncorrelated
   - No clear trend between 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 a linear function of X.

Correlation and random variables

1. $\rho(X, Y)$
   - One number that summarizes "trend" between two random variables
   - Dimensionless quantity.
2. $\rho(X, Y)$ is close to zero
   - X and Y are close to being uncorrelated
   - No clear trend between 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 a linear function 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

Problem

Consider the following joint PMF, where $-1/4 \leq x \leq 1/4$.

	$X = 0$	$X = 1$	f_Y
$Y = 0$	$1/4 - x$	$1/4 + x$	$1/2$
$Y = 1$	$1/4 + x$	$1/4 - x$	$1/2$
f_X	$1/2$	$1/2$	

$E[Y] = 1/2$
$E[Y^2] = 1/2$

$E[XY] = \frac{1}{4} - x$

$E[X] = 1/2, \; E[X^2] = 1/2$

$\mathrm{Cov}(X,Y) = \frac{1}{4} - x - \frac{1}{2} \cdot \frac{1}{2} = -x$

$\mathrm{Var}(X) = \frac{1}{4}, \; \mathrm{Var}(Y) = \frac{1}{4}$

$\rho(X,Y) = \frac{-x}{\frac{1}{2} \cdot \frac{1}{2}} = -4x$

Problem

Consider the following joint PMF, where $-1/4 \le x \le 1/4$.

	$X = c$	$X = c+1$	f_Y
$Y = d$	$1/4 - x$	$1/4 + x$	$1/2$
$Y = d+1$	$1/4 + x$	$1/4 - x$	$1/2$
f_X	$1/2$	$1/2$	

$E[Y] = d + 1/2$

$E[X] = c + 1/2$

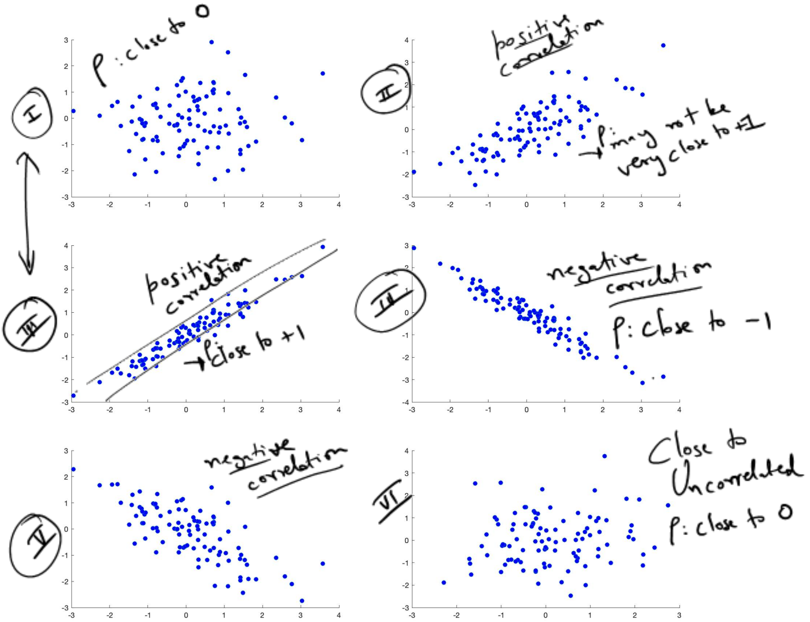
$$E[XY] = cd\left(\tfrac{1}{4} - x\right) + c(d+1)\left(\tfrac{1}{4} + x\right) + (c+1)d\left(\tfrac{1}{4} + x\right) + (c+1)(d+1)\left(\tfrac{1}{4} - x\right)$$

$$= cd + \frac{c}{2} + \frac{d}{2} + \frac{1}{4} - x$$

$\mathrm{Cov}(X,Y) = -x$

$\rho(X,Y) = -4x$

Scatter plots and correlation



Subsection 5

Bounds on probabilities using mean and variance

Notation for mean and variance

- If there is no confusion about the random variable X,
  - μ will denote the mean $E[X]$
  - σ^2 will denote the variance $\mathrm{Var}(X)$
  - σ will denote the standard deviation $\mathrm{SD}(X)$

Notation for mean and variance

- If there is no confusion about the random variable X,
  - μ will denote the mean $E[X]$
  - σ^2 will denote the variance $\mathrm{Var}(X)$
  - σ will denote the standard deviation $\mathrm{SD}(X)$
- If there are multiple random variables under discussion,
  - μ_X will denote the mean $E[X]$
  - σ_X^2 will denote the variance $\mathrm{Var}(X)$
  - σ_X will denote the standard deviation $\mathrm{SD}(X)$

What does mean say about the distribution?

Suppose the average marks in a course is $50/100$.

- What fraction of students will have marks ≥ 50?

What does mean say about the distribution?

Suppose the average marks in a course is $50/100$.

- What fraction of students will have marks ≥ 50?
  - It cannot be 0. Why?

What does mean say about the distribution?

Suppose the average marks in a course is $50/100$.

- What fraction of students will have marks ≥ 50?
  - It cannot be 0. Why?
- What fraction of students will have marks ≥ 80?

What does mean say about the distribution?

Suppose the average marks in a course is $50/100$.

- What fraction of students will have marks ≥ 50?
  - It cannot be 0. Why?
- What fraction of students will have marks ≥ 80?
  - It cannot be 1. Why?

What does mean say about the distribution?

Suppose the average marks in a course is $50/100$.

- What fraction of students will have marks ≥ 50?
  - It cannot be 0. Why?
- What fraction of students will have marks ≥ 80?
  - It cannot be 1. Why?
  - Can it be 0.9? No.

What does mean say about the distribution?

Suppose the average marks in a course is $50/100$.

- What fraction of students will have marks ≥ 50?
  - It cannot be 0. Why?
- What fraction of students will have marks ≥ 80?
  - It cannot be 1. Why?
  - Can it be 0.9?

What does mean say about the distribution?

Suppose the average marks in a course is $50/100$.

- What fraction of students will have marks ≥ 50?
  - It cannot be 0. Why?
- What fraction of students will have marks ≥ 80?
  - It cannot be 1. Why?
  - Can it be 0.9?

Average says something about the distribution of marks!

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.
  - Could be positive or negative

$\frac{X-\mu}{\sigma}$

- Standard units: The number 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 $\mu + c\sigma$

Examples

- Throw a pair of die, $X =$ sum of the two numbers
  - $\mu = 7$, $\sigma \approx 2.42$
  - $P(|X - \mu| \leq \sigma) = P(4.58 \leq X \leq 9.42) = P(X \in \{5, 6, 7, 8, 9\} = 2/3$
    - ⋆ So, $P(|X - \mu| > \sigma) = 1/3$
  - $P(|X - \mu| > 2\sigma) = P(X \in \{2, 12\}) = 2/36 \approx 0.056$

Examples

- Throw a pair of die, $X =$ sum of the two numbers
  - $\mu = 7$, $\sigma \approx 2.42$
  - $P(|X - \mu| \leq \sigma) = P(4.58 \leq X \leq 9.42) = P(X \in \{5, 6, 7, 8, 9\} = 2/3$
    - ★ So, $P(|X - \mu| > \sigma) = 1/3$
  - $P(|X - \mu| > 2\sigma) = P(X \in \{2, 12\}) = 2/36 \approx 0.056$
- $X \sim \text{Uniform}\{1, 2, \ldots, 100\}$
  - $\mu = 50.5$, $\sigma \approx 28.9$
  - $P(|X - \mu| > \sigma) = 58/100$
  - $P(|X - \mu| > 2\sigma) = 0$

Markov's inequality

Theorem (Markov's inequality)

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

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

$c = 100\mu$

$P(X \geqslant 100\mu) \leq \frac{1}{100} = 0.01$

Markov's inequality

Theorem (Markov's inequality)

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

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

Proof:

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

Markov's inequality

Theorem (Markov's inequality)

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

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

Proof:

- $\mu = \sum_{t \in T_X} t\, P(X = t) = \sum_{t<c} t\, P(X = t) + \sum_{t \geq c} t\, P(X = t)$
- Since the first sum is non-negative,

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

Chebyshev's inequality

Theorem (Chebyshev's inequality)

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

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Chebyshev's inequality

Theorem (Chebyshev's inequality)

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

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Proof:

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

Chebyshev's inequality

Theorem (Chebyshev's inequality)

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

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Proof:

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

Other forms:

Non-negative Markov

- $P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2}$, $P(\underline{(X-\mu)^2} \geq \underline{k^2\sigma^2}) \leq \frac{1}{k^2}$
- $P(\underline{\mu-k\sigma} < \underline{X} < \underline{\mu+k\sigma}) \geq 1-\frac{1}{k^2}$ $\longleftarrow$ complement of $|X-\mu| \geq k\sigma$
- $P(X \geq \mu+k\sigma) + P(X \leq \mu-k\sigma) \leq \frac{1}{k^2}$ $|X-\mu| \geq k\sigma \rightarrow X \geq \mu+k\sigma$ (or) $X \leq \mu-k\sigma$
  - $P(X \geq \mu+k\sigma) \leq \frac{1}{k^2}$, $P(X \leq \mu-k\sigma) \leq \frac{1}{k^2}$

Compare actual and Chebyshev

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

$$P(|X - \mu| \geq 2\sigma) \leq \frac{1}{4}$$

Compare actual and Chebyshev

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

$$P(|X-\mu| \geq 2\sigma) \leq \frac{1}{4}$$

1. $X \sim \text{Binomial}(10, 1/2)$, $\mu = 5$, $\sigma = \sqrt{2.5} \approx 1.58$

$$P(|X-5| \geq 2\sigma) = P(X \in \{0,1,9,10\}) \approx 0.021 \leq 1/4$$

Compare actual and Chebyshev

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

$$P(|X - \mu| \geq 2\sigma) \leq \frac{1}{4}$$

1. $X \sim \text{Binomial}(10, 1/2)$, $\mu = 5$, $\sigma = \sqrt{2.5} \approx 1.58$

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

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

$$P(|X - 4| \geq 2\sigma) = P(X \in \{11, 12, \ldots\}) \approx 0.056 < 1/4$$

What do mean and 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.

What do mean and 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 deviation σ, through Chebyshev's inequality: bound the probability that X is away from μ by $k\sigma$.

What do mean and 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 deviation 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~~ number of accidents decreases by 10000 per day across the country. Is that a "significant" decrease?
  - If standard deviation of the number of accidents is known, we can find how high 10000 is in terms of standard deviations to answer above question.

What do mean and 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 deviation 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 number of accidents decreases by 10000 per day across the country. Is that a "significant" decrease?
    - If standard deviation of the number of accidents is known, we can find how high 10000 is in terms of standard deviations to answer above question.
- There are a lot of important theoretical implications as well. We will see these later on.