Chapter 4: Random Variables

Random variables

Example: Suppose we randomly select 2 people one after another from the football game between Penn State and Temple to interview them. Suppose we know that at the stadium, a person is a PSU fan with probability 0.8 and a Temple fan with probability 0.2, and assume independence.

Let X be the number of PSU fans among the selected 2.

$S = \{PP, PT, TP, TT\}$

$\Rightarrow S_X = \{0,1,2\}$

$\Rightarrow S_Y = \{0,1\}$

$X =$ # of PSU fans among the 2 selected

$\Rightarrow \boxed{X = \{0,1,2\}}$

$P(X=0) = P(\underbrace{TT}_{T_1 \cap T_2}) = P(T_1) \cdot P(T_2) = 0.2 \times 0.2 = \boxed{0.04}$

$P(X=1) = P(\underbrace{PT}_{P_1 \cap T_2}) + P(\underbrace{TP}_{T_1 \cap P_2}) = 0.2 \times 0.8 + 0.8 \times 0.2 = 0.16 \times 2 = \boxed{0.32}$

$P(X=2) = P(\underbrace{PP}_{P_1 \cap P_2}) = 0.8 \times 0.8 = \boxed{0.64}$

$$Y = \begin{cases} 1 & \text{if atleast one of the 2 is a PSU fan} \\ 0 & \text{otherwise} \end{cases}$$

$P(Y=0) = 0.04$

$\downarrow$

None of them case

$P(Y=1) = 0.32 + 0.64 = \boxed{0.96}$

Random variable is a rule that associates a number with each outcome of the experiment.

$S = \{PP, PT, TP, TT\}$
$X = 2, 1, 1, 0$
$Y = 1, 1, 1, 0$

Random Variables : X, Y, Z, H, W

Values of these variables : x, y, z, h, w

The *sample space* or *support* of a random variable is the collection of its possible values.

of a random variable

$\boxed{S_X = \{\text{support}\} \text{ of random variable } X}$

$$\boxed{P(-\infty < X < \infty) = 1}$$

We know the *probability distribution* of a random variable X, if for any interval (a, b) on the real line we know $P(a < X < b)$.

a 8

$$\boxed{P(a \le X \le 8) = P(X \in [a, 8]) = \sum_{a \le x_i \le 8} p(x_i)}$$

RV → discrete, continuous

2 types of random variables:

- discrete

If its support S_X is finite or countably infinite

large but can be counted

- continuous

If its support S_X is an interval or union of intervals

reaction time $(0, \infty)$

not a fixed value but an interval

Discrete random variables → Rest of Ch-4 ; Continuous in Ch5

Let X be a discrete random variable with the support $S = \{x_1, x_2, x_3, \ldots\}$.

The *probability mass function (pmf)* of X → Probability at a particular value

$p(x) = P(X=x)$ for all $x \in x_1, x_2 \ldots$

Probability of an element x

So,

$p(x_i) = P(X = x_i) \quad (x_i \in S_x)$

$p(x) = 0 \quad (x_i \notin S_x)$

New defn ⇒ $S_x = \{x : p(x) > 0\}$ ⇒ $p(x)$ is a valid pmf iff

1. $0 \le p(x) \le 1$ for any x_i
2. $\sum_{x_i \in S} p(x_i) = 1$

Example 1: Suppose a family has 3 children. Also assume that girls and boys are equally likely and the gender of a child has nothing to do with the gender of the next child. Let X be the number of girls family has. What is the pmf of X?

S={ggg, ggb, gbg, gbb, bgg, bgb, bbg, bbb}

x	0	1	2	3
$P(X = x)$ or $p(x)$	1/8	3/8	3/8	1/8

$p(0) = P(X=0) = \frac{1}{8}$ → $\frac{{}^3C_0}{2^3}$

$p(1) = P(X=1) = \frac{3}{8}$ → $\frac{{}^3C_1}{2^3}$

$p(2) = P(X=2) = \frac{3}{8}$ → $\frac{{}^3C_2}{2^3}$

$p(3) = P(X=3) = \frac{1}{8}$ → $\frac{{}^3C_3}{2^3}$

1/2 1/2 1/2

$P(X=0) = \frac{1}{8}$

1/2 1/2 1/2

$P(X=1) = 3 \times \left(\frac{1}{8}\right)$

. . .

$\boxed{2 \times 2 \times 2 = 8}$

Example 2: Consider rolling a fair die. Let X be the number on the upper surface. What is the pmf of X?

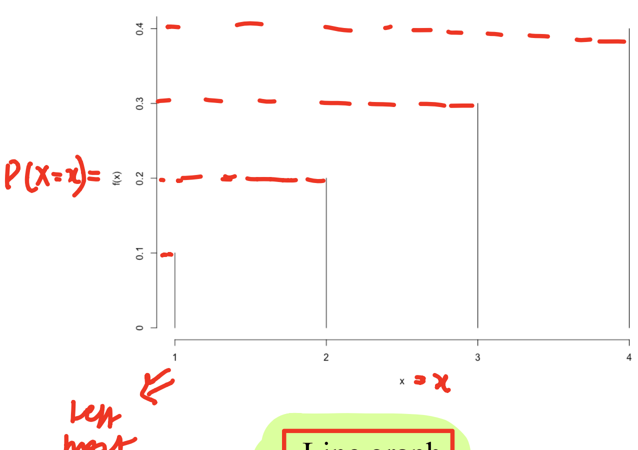
x	1	2	3	4	5	6
$P(X = x)$	1/6	1/6	1/6	1/6	1/6	1/6

NOTE: If the pmf is a constant over the support of a discrete random variable X, then we say X has a ***discrete uniform distribution***.

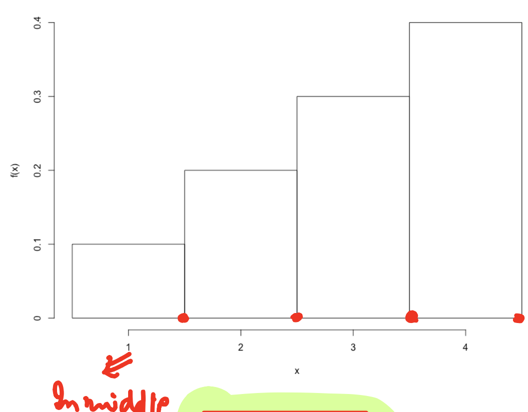
Visualizing a pmf

Example: Consider the below probability distribution

x	1	2	3	4
$P(X = x)$	0.1	0.2	0.3	0.4



Line graph



Histogram

The *cumulative distribution function (cdf)* of X

cdf of RV X is $\boxed{F(x) = P(X \le x)}$

$$\boxed{\begin{array}{l} p_X(x) = P(X = x) \\ F_X(x) = P(X \le x) \end{array}}$$
⇒ Better notation

Properties of the CDF:

①

- $F(x)$ is a non-decreasing function
- $\lim_{x\to-\infty} F(x) = 0 = F(-\infty)$ $\lim_{x\to\infty} F(x) = 1 = F(\infty)$
- $P(a < X \le b) = F(b) - F(a)$.
- If X is a discrete rv with sample space S, then cdf $F(x)$ is a jump/step function, with jumps occurring at the elements of the support S, and flat regions of $F(x)$ where X takes no values. Moreover, the size of the jump at each x from S equals $P(X = x)$.
- $P(X > a) = 1 - F(a)$

(cdf of x)

x (pmf of x)

∵ first element

Support $S_X = \{x_1, x_2, x_3, \dots\}$

- $p(x_1) = F(x_1)$, $p(x_i) = F(x_i) - F(x_{i-1})$ for $i = 2,3,\dots$
- $F(a) = P(X \le a) = \sum_{x_i \le a} p(x_i)$
- $P(a < X \le b) = F(b) - F(a) = \sum_{a < x_i \le b} p(x_i)$
- $P(a \le X \le b) = \sum_{a \le x_i \le b} p(x_i)$

Example 3: Joe plays a game of chance for money. In a round of this game,

- he wins \$12 with probability 0.4, \$33 with probability 0.2, and \$50 with probability 0.1
- he loses \$25 with probability 0.3.

Let X represent Joe's result after a round of this game. Then the pmf of X is given by:

x	-25	12	33	50
$p(x) = P(X = x)$	0.3	0.4	0.2	0.1

a) ~~What value of c makes it a valid pmf?~~ Is it a valid pmf?
b) Find the cdf of X. Plot the cdf.
c) What is the probability that Joe's result is \$33?
d) What is the probability that Joe's result is at most \$33?
e) What is the probability that Joe's result is less than \$33?

a) first check if all entries of $\boxed{0 \le p_x \le 1}$ & $\boxed{\sum p_x = 1}$

Now,

b) $F(-25) = P(X \le -25) = \boxed{0.3 = P(X=-25)}$

$F(12) = P(X \le 12) = p(-25) + p(12) = \boxed{0.7}$

d) $F(33) = P(X \le 33) = P(X=-25) + P(X=12) + P(X=33) = \boxed{0.9}$

e) $F(50) = P(X \le 50) = \boxed{1}$

$\boxed{p(33) = F(33) - F(12)}$ holds for all

$\Downarrow$

Property

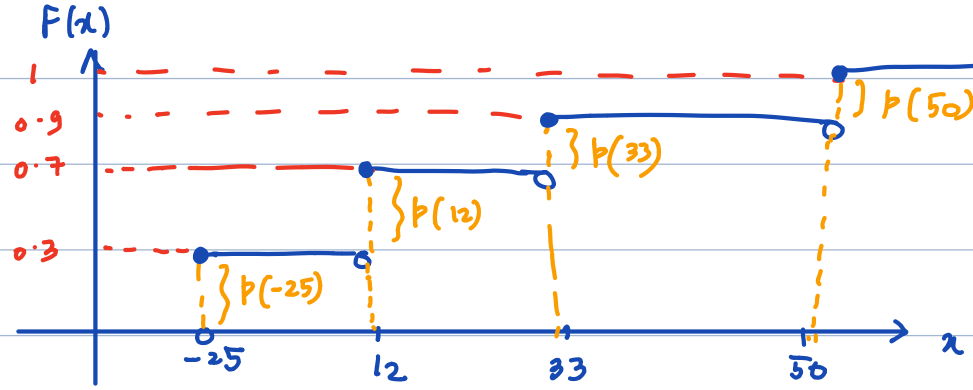
x	-25	12	33	50
$p(x)$	0.3	0.4	0.2	0.1
$F(x)$	0.3	0.7	0.9	1

$0\ (x < -25)$ | $0.3\ (-25 \le x < 12)$ | $0.7\ (12 \le x < 33)$ | $0.9\ (33 \le x < 50)$ | $1\ (x \ge 50)$

$F(-3) = P(X \le -3)$
$= P(X = -25)$
$= 0.3$

Simplify similarly | Just add as shown

Graph



c) $P(X=33) \overset{\text{pmf}}{=} p(33) = \boxed{0.2}$

$\overset{\text{cdf}}{=} F(33) - F(12) = 0.9 - 0.7 = \boxed{0.2}$

$P(X=33)$ also $= P(12 < X \leq 33) = F(33) - F(12) = \boxed{0.2}$

Also, if Joe's result is less than \$40, what's the probability that he didn't lose money?

$$P(X \geq 0 \mid X < 40) = \frac{P((X\geq 0) \cap (X<40))}{P(X<40)} = \frac{0.4+0.2}{0.3+0.4+0.2} = \frac{0.6}{0.9} = \boxed{0.67}$$

i)

ii) Using pmf ⇒ $\dfrac{P(0 \leq X < 40)}{P(X<40)}$

$$= \frac{p(12)+p(33)}{p(12)+p(33)+p(-25)} = \boxed{0.67}$$

iii) Using cdf ⇒

$$\frac{F(33) - F(-25)}{F(33)} = \boxed{0.67}$$

Parameters of Discrete Distributions

Example 3 (cont.): Let X represent Joe's result after a round of this game. What is the average result?

$P(X=33)$

i) $\overset{pmf}{=} p(33) = \boxed{0.2}$

ii) $\overset{cdf}{=} F(33) - F(12)$

$= 0.9 - 0.7 = \boxed{0.2}$

$\approx P(12 < X \le 33) = F(33) - F(12)$
$= \boxed{0.2}$

$P(X \le 33)$

i) $\overset{pmf}{=} p(-25) + p(12) + p(33)$
$= \boxed{0.9}$

ii $\overset{cdf}{=} F(33) = \boxed{0.9}$

$\approx P(-\infty < X \le 33) = F(33) - F(-\infty) = \boxed{0.9}$

$P(X<33)$

i) $\overset{pmf}{=} p(-25) + p(12) = \boxed{0.7}$

ii) $\overset{cdf}{=} P(X \le 12) = F(12) = \boxed{0.7}$

$\approx P(-\infty < X \le 12)$
$= F(12) - \underbrace{F(-\infty)}_{=0}$
$= F(12) = \boxed{0.7}$

Let X be a discrete random variable with pmf $p(x)$ and support $S_X = \{x_1, x_2, x_3, \dots\}$

The *expected value (expectation/mean/average)* of X

$$\boxed{E(X) = \mu_X = \mu_X = \sum_x x \cdot p(x)}$$

$$\boxed{\text{Average} = \sum x \cdot p(x) \quad \text{for all } x}$$

$\Rightarrow$ Draw another column in the table & find $E(X)$

Example 4: What is the expected value of a discrete random variable X with the pmf $p(x) = \frac{c}{x^2}$, $x = 1,2,3, \ldots$, where c is a constant?

$$p(x) = \frac{c}{x^2} \qquad (x = 1, 2, 3 - - -)$$

$$E(x) = \sum_{x=1}^{\infty} x \cdot \frac{c}{x^2} = c \sum_{x=1}^{\infty} \frac{1}{x} = c \cdot \infty = \infty$$

$\Downarrow$

Expectation is not finite

$\therefore$ x does not have expectation

or, expectation of x does not exist

Now say,

$$I(A) = \begin{cases} 1, & \text{if event } A \text{ occurs} \\ 0, & \text{if } A \text{ doesn't occur} \end{cases}$$

$\downarrow$

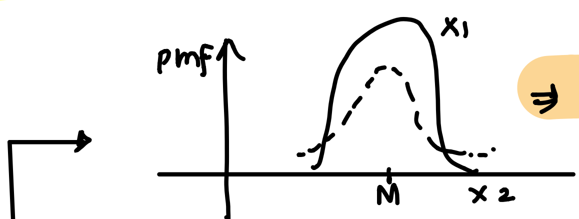
Indicator r.v.

$\rightarrow$ ⊛ Probability is just a special case of expectation.

I_A	0	1
$P_{I(A)}(i)$	$1-P(A)$	$P(A)$

$$\Rightarrow \quad E(I_A) = 0 \cdot (1-P(A)) + 1 \cdot P(A) = P(A)$$

⊛ Expected value $(E(X))$ has units of X only.



$\Rightarrow$ Both have same centered value, i.e mean of M, but X_1 is less variant $\because$ less space $\therefore$ more precise

How much variability in probability distribution?

Variance

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

Average squared deviation from mean is σ (variance)

$$\Rightarrow \sigma_X^2 = \sum_x (x-\mu)^2 \cdot p(x) = E(X^2) - \mu_x^2 = \left(\sum_x x^2 p(x)\right) - \mu_x^2$$

$\circledast$ Units $\Rightarrow$ (units of X)^2

Standard deviation

$$\boxed{\sigma_x = \sqrt{Var(X)}}$$

$\Rightarrow$ $\circledast$ Units = units of X

$\circledast$ SD & variance are always non-negative

$Var(X) \geq 0$

$\sigma_X \geq 0$

Example 3 (cont.): Find the variance and the standard deviation of Joe's result X.

Variance $\Rightarrow E(X^2) = \sum_x x^2 \cdot p(x) = (-25)^2(0.3) + (12)^2(0.4) + (33)^2(0.2) + (50)^2(0.1) = 712.9$

now,

$\sigma_X = E(X^2) - \mu_x^2$ calculate from the given data using formula

$\sigma_X = 712.9 - (8.9)^2$

$\Rightarrow \boxed{\sigma_X^2 = 633.7}$ $\$^2$

SD $\Rightarrow$ $\sqrt{\sigma_X^2} = 25.17$ \$

⊛ Proof of different formulae

$$Var(X) = \sigma_X^2 = E\left[(X-\mu_X)^2\right] = E\left[X^2 - 2\mu X + \mu^2\right]$$

$$= E(X^2) + E(-2\mu X) + E(\mu^2)$$

$$= E(X^2) - 2\mu \underbrace{E(X)}_{\mu} + \mu^2$$

$$\Rightarrow E(X^2) - 2\mu^2 + \mu^2 = \boxed{E(X^2) - \mu^2}$$

Expectation of a Function of Random Variable

X	$p(x)$
$h(X)$	r.v
$E(h(x))$	

Let $g(\cdot)$ and $h(\cdot)$ be real valued functions.

- The expected value of $h(X)$

1. In general, $E\big(h(X)\big) = \sum_x h(x)\cdot p(x)$

2. If $h(X)$ is a linear function, i.e. $h(X) = aX + b$ (1. is still valid but there is something simpler)

$E(ax+b) = a\cdot E(x)+b$

$Var(aX+b) = a^2\cdot Var(X) = a^2\cdot \sigma_X^2$

(SD) $\Rightarrow \sigma_{ax+b} = |a|\cdot \sigma_X$

(b does not affect the spread ∵ it is just shifting of graph)

(a does affect ∵ it decides the height)

- For any constants c and d and functions h and g

$$\boxed{E\big(c\,h(x) + d\,g(x)\big) = c\,E(h(x)) + d\,E(g(x))}$$

Example 3 (cont.): Zoe's donation Y (in \$) depends on Joe's result X. Find Zoe's expected donation, if Y is

a) $16 - 0.3X$
b) $e^{X/9}$

$E(Y) = ?$

• also calculate $Var(Y)$

a) Donation $Y = 16 - 0.3X = h(X)$ (say) $(a = -0.3,\ b = 16)$

$E(Y) = E(h(x)) = E(16-0.3X) = 16 - 0.3\,E(x) = 16 - 0.3(8.9) = \boxed{13.33\$}$

$Var(Y) = ?$

b) $E(Y) = E(e^{X/9}) = E(g(X))$ (say) $= \sum_x g(x)\cdot p(x)$

$= \sum_x e^{x/9}\cdot p(x) = e^{-\frac{25}{9}}\cdot(0.3) + e^{12/9}(0.4) + e^{33/9}(0.2) + e^{50/9}(0.1) = \boxed{35.22\$}$

$\neq e^{E(X)/9} = e^{8.9/9} = \boxed{2.69\$}$ wrong

$Var(Y) = ?$

∴ Second distribution makes bigger donations.

SOLUTIONS:

Variance, standard Dev.

(a) Since $16 - 0.3X$ is a linear function of X, $Var(16 - 0.3X) = (-0.3)^2 Var(X) = 0.09 * 633.7 = 57.03$. And $\sigma_{16-0.3X} = |-0.3|\sigma_X = 0.3 * 25.17 = 7.55$ (or $\sqrt{57.03}$).

Variance, standard Dev.

(b) $e^{X/9}$ is <u>not</u> a linear function of X $\Rightarrow$ use $Var(Y) = E(Y^2) - (E(Y))^2$. In class we found $E(Y) = 35.22$, so need $E(Y^2)$. Note $Y^2 = (e^{X/9})^2 = e^{2X/9}$.
Then $E(Y^2) = E(e^{2X/9}) = \sum x e^{2x/9} p_X(x) = e^{2*(-25)/9} * 0.3 + e^{2*12/9} * 0.4 + e^{2*33/9} * 0.2 + e^{2*50/9} * 0.1 = 7002.9$ $\Rightarrow$ the variance $Var(Y) = E(Y^2) - (E(Y))^2 = 7002.9 - 35.22^2 = 5762.45$ and the standard deviation $\sigma_Y = \sqrt{5762.45} = 75.91$.

Moments of random variable X

first moment

Recall: the mean $\mu = E(X) = E(X^1) \Rightarrow (r=1) \Rightarrow E(X^r)$

second moment about μ

the variance $\sigma^2 = E((X-\mu)^2) = E(X^2) - \mu^2 = \boxed{r=2 \ \& \ b=\mu}$

$E(X^2) \Rightarrow (r=2)$

For any positive integer r,

$r \in \mathbb{N}$ $\quad [X^r = \text{a function}]$

- the r-th moment of X

$$\boxed{E(X^r) = \sum_x x^r p(x)}$$

- the r-th moment of X about constant b

$$\boxed{E((X-b)^r) = \sum_x (x-b)^r p(x)}$$

Example 3 (cont.): Find the third moment of Joe's result X.

$$E(X^3) = \sum_x x^3 \cdot p(x) = (-25)^3(0.3) + (12)^3(0.4) + (33)^2(0.2) + (50)^3(0.1) = \boxed{15691.1}$$

Also, find the 3rd moment of X about its mean

$$E((X-8.9)^3) = \boxed{-1933.39}$$

Moment #	Physical sig.
1	Centre of distribution
2	Spread of distribution
3	Skew of distribution
4	...

Revision Notes

- $X = \{x_1, x_2 \ldots\}$ ⇒ all possible values of RV X

  $\boxed{S_X}$ support ↓ RV ⇒ X

  $\boxed{P(X=x) = p(x) \,,\; x \in x_1, x_2 \ldots}$ →

  pmf

  $\boxed{F(x) = P(X \le x) \,,\; x \in x_1, x_2 \ldots}$

  cdf

$\lim_{x \to -\infty} F(x) = 0$	$\lim_{x \to \infty} F(x) = 1$	$P(a < X \le b) = F(b) - F(a)$
$P(X > a) = 1 - F(a)$		

- Valid pmf ⇒ $0 \le p(x) \le 1$ & $\sum p(x) = 1$

- $\boxed{\text{units of } X}$

  $E(X) = \mu_x = \sum x \cdot p(x)$ ⇒ Expected value / Mean / Average

- Variance $\sigma_x^2 = E((X - \mu)^2) = E(X^2) - \mu_x^2 = Var(X)$

- Mean $\sigma_X = \sqrt{Var(X)}$

- $E(c\, h(X) + d\, g(X)) = c\, E(h(X)) + d\, E(g(X))$

- $E(h(X)) = \sum h(x)\cdot p(x)$

- $Var(aX+b) = a^2 \cdot Var(X) = a^2 \cdot \sigma_X^2$

- $E(a\,h(X)+b) = a\,E(h(X))$

- $SD(aX+b) = \sqrt{Var X} = |\sigma_X^2| = |a|\cdot |Var(X)|$

- $Var(Y) = \sigma_Y^2 = E(Y^2) - (E(Y))^2$

  $\Downarrow$

  non linear

- rth moment about a constant b, $E((X-b)^r) = \sum (x-b)^r \cdot p(x)$