Chapter 3 Discrete Random Variables and Distributions

3.1 Random Variables

A random variable is a rule (function) that associates a number with each outcome in the sample space S.

We often write RV as an abbreviation.

Ex:
- Roll a die, $S = \{1,2,3,4,5,6\}$

  X is the square of what we roll, $(X(\omega) = \omega^2)$

  ex: if we roll 5, then $X = 25$

- Flip a coin, $S = \{H, T\}$

  $X = 1$ if we get Heads

  $X = 0$ if we get Tails

Bernoulli RVs: A random variable whose possible values are 0 and 1 is called a Bernoulli Random Variable.

Ex:
- Flip a coin, X is 1 if Heads, 0 if Tails
- Roll a die, X is 1 if we get 5 or 6

  X is 0 if we get 1, 2, 3, or 4

Discrete and Continuous Random Variables:

A random variable with finitely many or countably infinitely many (1, 2, 3, 4, 5, 6, 7, ...) possible values is called a discrete random variable.

A random variable X whose possible values are an interval on the number line (or maybe a disjoint union of intervals) and which satisfies $P(X=c) = 0$ for all possible values c is called a continuous random variable.

Ex:

Discrete	Continuous
counting things ex: # students in class # people at an address	Temperature, any real number in some range
	Time, Age 43.218 years
characteristics/categorical eye color (blue, brown, green, ...)	Measurements, Height, Length, Volume, Area

3.2 Probability Distributions for Discrete RVs:

A discrete random variable has either finitely many, or countably infinitely many, 1, 2, 3, ...

possible values. It has some probability to be each of these values.

For example, say we flip a fair coin twice, $S = \{HH, HT, TH, TT\}$.

Let X be the number of heads flipped.

$$p(0) = P(X=0) = \frac{1}{4}$$
$$p(1) = P(X=1) = \frac{2}{4} = \frac{1}{2}$$
$$p(2) = P(X=2) = \frac{1}{4}$$

The probability mass function (pmf) or distribution of a discrete RV X is

the function $p(x)$ defined by $p(x) = P(X=x)$ for all x.

If we add up $p(x)$ over all possible values x of X, we get

$$\sum_{\substack{\text{all possible} \\ \text{values } x}} p(x) = 1$$

Example 1: Six boxes of components are ready to be shipped by a certain supplier. The number of defective components in each box is

box	1	2	3	4	5	6
number defective	0	2	0	1	2	0

A box is picked at random. Let X be the number of defective components.

Find the pmf for X.

Possible values for X: 0, 1, 2

$$p(0) = \frac{3}{6} = \frac{1}{2}$$
$$p(1) = \frac{1}{6}$$
$$p(2) = \frac{2}{6} = \frac{1}{3}$$

A parameter of a probability distribution

Say we have a coin that when flipped comes up Heads with probability α, and tails with probability $1-\alpha$.

Let X be 1 if the coin is Heads and 0 if the coin is Tails.

Find the pmf for X.

$$p(0) = P(X=0) = 1-\alpha$$
$$p(1) = P(X=1) = \alpha$$

This gives the family of Bernoulli (α) random variables.

The Cumulative Distribution Function:

The pmf tells us $P(X=x)$.

Often we want to know $P(X \le x)$.

The Cumulative Distribution Function (CDF) of a discrete RV X with pmf $p(x)$ is the function

$$F(x) = P(X \le x) = \sum_{y:\, y \le x} p(y).$$

Example 2: There are different sized USB flash drive in a box, with either 2 GB, 4 GB, 8 GB, 16 GB, 32 GB, or 64 GB of storage. Let Y be the size of a randomly selected USB drive in the box. The pmf of Y is

y	2	4	8	16	32	64
$P(y)$	0.05	0.05	0.1	0.3	0.3	0.2

a) What is $P(Y \le 8)$? What is $P(Y < 8)$?

b) Find the CDF for Y and graph it.

c) If we need to install an OS that requires 10 GB of space, what is the probability the random drive will have enough space available?

d) What is $P(4 \le Y \le 16)$?

Solution: a) $P(Y \le 8) = P(Y=2) + P(Y=4) + P(Y=8)$

$$= 0.05 + 0.05 + 0.1 = \boxed{0.2}$$

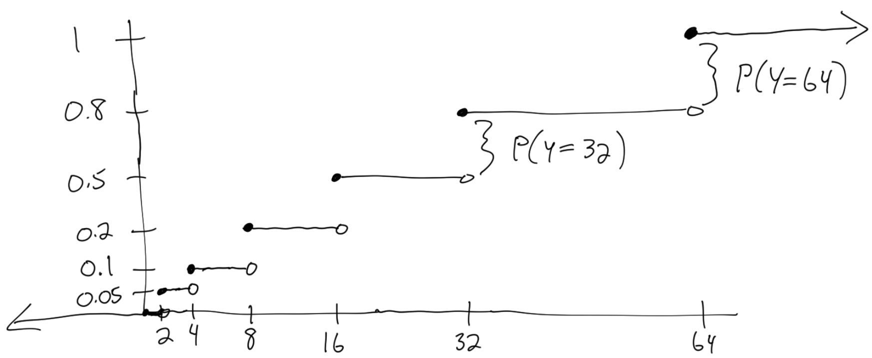
$$P(Y<8) = P(Y=2) + P(Y=4) = 0.05 + 0.05 = \boxed{0.1}$$

$$= P(Y \in \{2,4\})$$

b) CDF for Y,

y	2	4	8	16	32	64
$p(y)$	0.05	0.05	0.1	0.3	0.3	0.2

$$F(y) = \begin{cases} 0 & \text{if } y < 2 \\ 0.05 & \text{if } 2 \le y < 4 \\ 0.1 & \text{if } 4 \le y < 8 \\ 0.2 & \text{if } 8 \le y < 16 \\ 0.5 & \text{if } 16 \le y < 32 \\ 0.8 & \text{if } 32 \le y < 64 \\ 1 & \text{if } y \ge 64 \end{cases}$$



c) $P(Y > 10) = 1 - P(Y \le 10) = 1 - 0.2 = \boxed{0.8}$

OR $P(Y=16) + P(Y=32) + P(Y=64) = \boxed{0.8}$

d) $P(4 \le Y \le 16) = P(Y=4) + P(Y=8) + P(Y=16) = \boxed{0.45}$

3.3 Expected Value:

ex: Roll Die, X = square of what we roll

$S = \{1,2,3,4,5,6\}$, $D = \{1,4,9,16,25,36\}$

Let X be a discrete RV with set of possible values D and pmf $p(x)$.

The expected value or mean value of X is

$$E[X] = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Example 1: At an arcade, you can play a game to win tickets. From experience you know you have a 20% chance to win 300 tickets, a 30% chance to win 100 tickets, and a 50% chance to win 10 tickets. What is the expected value of the number of tickets you win?

Let X be the number of tickets we win.

$$E[X] = 0.2 \cdot 300 + 0.3 \cdot 100 + 0.5 \cdot 10 = 60 + 30 + 5 = \boxed{95}$$

Sometimes we want to know the expected value of some function $h(X)$.

Say X has pmf:

x	1	2	3
$p(x)$	.6	.3	.1
x^2	1	4	9

How can we find $E[X^2]$?

$$E[X] = 1 \cdot 0.6 + 2 \cdot 0.3 + 3 \cdot 0.1 = 1.5$$

$$E[X^2] = 1^2 \cdot 0.6 + 2^2 \cdot 0.3 + 3^2 \cdot 0.1 = 0.6 + 1.2 + 0.9 = \boxed{2.7}$$

If X is a random variable with possible values D and pmf $p(x)$, then for any function h,

$$E[h(X)] = \sum_{x \in D} h(x) \cdot p(x)$$

Example 2: Let X be the number of Heads we get by flipping a fair coin twice.

What is $E\left[\frac{1}{1+X}\right]$?

$h(x) = \frac{1}{1+x}$

$S = \{HH, HT, TH, TT\}$

x	0	1	2
$p(x)$	0.25	0.5	0.25
$h(x) = \frac{1}{1+x}$	1	$\frac{1}{2}$	$\frac{1}{3}$
	$h(0)$	$h(1)$	$h(2)$

$$E\left[\frac{1}{1+X}\right] = 1 \cdot 0.25 + \frac{1}{2} \cdot 0.5 + \frac{1}{3} \cdot 0.25$$
$$= 0.25 + 0.25 + \frac{1}{12}$$
$$= \boxed{\frac{7}{12}}$$

Variance: The expected value describes the center of a probability distribution.

The variance describes the spread.

Let X have pmf $p(x)$ and expected value μ. The variance of X is

$$Var(X) = V(X) = \sigma_X^2 = \sum_D (x-\mu)^2 \cdot p(x) = E[(X-\mu)^2]$$

The standard deviation of X is

$$SD(X) = \sigma_X = \sqrt{\sigma_X^2}$$

$$Var(X) = \sum_D (x-\mu)^2 \cdot p(x)$$

Example 3: Let X be the number of Heads we get by flipping a fair coin twice.

Find $Var(X)$ and $SD(X)$.

x	0	1	2
$p(x)$	0.25	0.5	0.25

$\{HH, HT, TH, TT\}$

$$\mu = 0(0.25) + 1(0.5) + 2(0.25) = 1$$

$$Var(X) = (0-1)^2 \cdot 0.25 + (1-1)^2 \cdot 0.5 + (2-1)^2 \cdot 0.25$$
$$= 1 \cdot 0.25 + 0 + 1 \cdot 0.25 = \boxed{0.5}$$

$$SD(X) = \boxed{\sqrt{0.5}} = \boxed{0.707}$$

Shortcut Formula for Variance:

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

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

Example 4: Let X be the internet speed of a randomly selected apartment in a certain complex.

Say X has pmf

x (in mbps)	0	10	25	40
$p(x)$	0.2	0.3	0.4	0.1

Find $Var(X)$.

$$E[X] = 0(0.2) + 10(0.3) + 25(0.4) + 40(0.1) = 17$$

$$E[X^2] = 0^2(0.2) + 10^2(0.3) + 25^2(0.4) + 40^2(0.1) = 440$$

$$V(X) = E[X^2] - (E[X])^2 = 440 - 17^2 = \boxed{151}$$

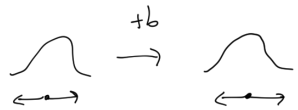
$$SD(X) = \sqrt{151} \approx \boxed{12.3}$$

Expected Value and Variance of a Linear Function

Let a, b be real numbers.

$$E[aX+b] = a\,E[X] + b$$

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



Pf: $$E[aX+b] = \sum_D (ax+b)\cdot p(x) = \sum_D ax\,p(x) + \sum_D b\,p(x)$$

$$= a\underbrace{\sum_D x\,p(x)}_{E[X]} + b\underbrace{\sum_D p(x)}_{1} = a\,E[X] + b\cdot 1.$$

$$Var(aX+b) = \sum_D \left(ax+b - E[aX+b]\right)^2 p(x)$$

Idea: $$\sum_D (ax - a\,E[X])^2\,p(x)$$

$$a^2 \sum_D (x-\mu)^2 p(x) = a^2\,Var(X)$$

Example 5: On a chess site, the average player Blitz rating is 1090 with standard deviation 400. Let X be a randomly chosen player's rating.

(a) What are the new mean and sd if the site adds 100 points to each rating?

$$E[X+100] = 1090 + 100 = \boxed{1190}$$

$$Var(X+100) = (400)^2 = \boxed{160000}$$

$$SD(X+100) = \boxed{400}$$

$$E[aX+b] = a\,E[X] + b$$

$$V(aX+b) = a^2\,V(X)$$

$$SD(aX+b) = |a|\,SD(X)$$

(b) What if they double the scores and then add 100 points?

$$E[2X+100] = 2\cdot 1090 + 100 = 2180 + 100 = \boxed{2280}$$

$$V(2X+100) = 2^2\cdot Var(X) = 2^2\cdot 160000 = \boxed{640000}$$

$$SD(2X+100) = 2\cdot SD(X) = 2\cdot 400 = \boxed{800}$$