LECTURE: EXPECTATION AND VARIANCE

1. Method of Indicators

Recall:

The **expected value** of a discrete random variable Y is

$$E(Y) = \sum_{y} yP(Y = y)$$

Example 1:

One evening, n customers dine at a restaurant. Each gives their hat to a concierge. After dinner, the concierge gives the hats back to the customers in a random order, i.e. each customer receives one of the hats at random.

What is the expected number of customers who get their own hat back?

Let X be the number of customers who get their own hat back.

We *could* use the definition of expected value, but then we would have to calculate $P(X = i)$, the probability that i customers get their hats back, which is very hard.

Instead, we will use the **method of indicators** to solve this problem.

Definition:

An **indicator random variable** is a random variable I which only takes the values 0 and 1.

It is used to indicate if an event takes place, so $I = 1$ if the event happens, and $I = 0$ if not, like an ON/OFF switch.

STEP 1: First, let's number the customers $1, 2, \ldots, n$

For $i = 1, \ldots, n$, let X_i indicate if customer i gets their own hat back:

$$X_i = \begin{cases} 1 & \text{customer } i \text{ gets their own hat back} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Then } X = X_1 + X_2 + \cdots + X_n$$

This makes sense, because we are adding a 1 whenever a customer gets their own hat back to get the total number of customers who get their hat back.

STEP 2:

$$E(X) = E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$$

By the definition of expected value:

$$E(X_i) = 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1) = P(X_i = 1)$$

Here $P(X_i = 1)$ is the probability that customer i gets their own hat back. Since the hats are distributed at random, and there are n hats to distribute, we must have $P(X_i = 1) = 1/n$. Thus,

$$E(X_i) = \frac{1}{n} \quad \text{for } i = 1, 2, \cdots, n$$

$$\text{Hence } E(X) = \underbrace{\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}}_{n \text{ times}} = 1$$

The key to this method is that $E(X_i)$ does not depend on i, it is exactly the same for all n customers.

Why does this make sense? Mathematically there is no distinction between the n customers. Imagine the customers lining up to leave the restaurant. The person at the front of the line is handed a hat at random, after which they leave. This is repeated until all customers have left. If we swap any two customers in line, nothing should change. The expected number of customers who receive their own hat should remain the same.

Note: How do you know when to use the method of indicator random variables and linearity of expectation? There is no magical formula for this, but it is useful when

(1) You are looking for an *expected value* involving a group of people or objects

(2) You have a symmetric group of people or objects, i.e. you can swap them around without affecting the result.

2. Functions of Random Variables

If Y is a random variable and $g(y)$ is a real-valued function, then $g(Y)$ is another random variable.

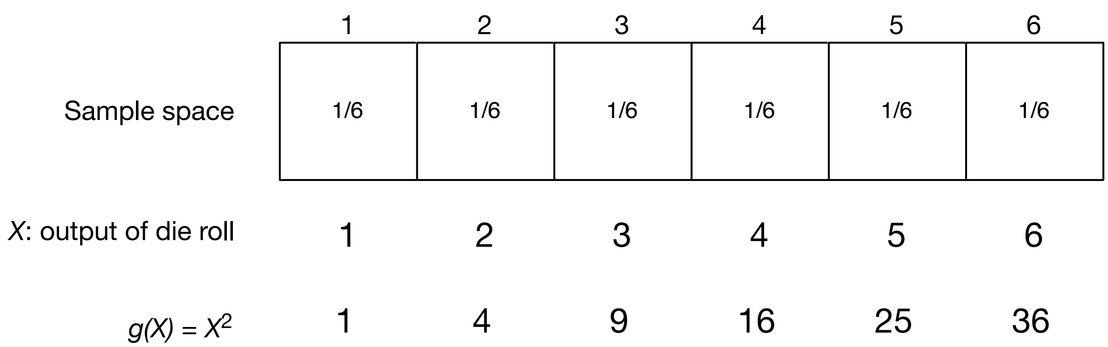
To evaluate $g(Y)$, just take the output of Y and run it through g, as the following example shows:

Example 2:

Let X be the output of a standard six-sided die, and let $g(x) = x^2$
Then $g(X)$ is also a discrete random variable.

In other words, you roll a die and square the output

First, let's show X and $g(X)$ graphically:



x	$g(x)$	$p(x)$
1	1	1/6
2	4	1/6
3	9	1/6
4	16	1/6
5	25	1/6
6	36	1/6

The expected value of a function of discrete random variable is computed in a similar fashion to the expected value of a random variable.

Fact:

Let Y be a discrete random variable with probability function $p(y)$, and let g be a real-valued function.

Then the expected value of the random variable $g(Y)$ is given by:

$$E[g(Y)] = \sum_{\text{all } y} g(y)\, p(y)$$

This is again a weighted average, but this time we are taking the weighted average of all possible values of $g(Y)$

Example 3:

Let X be the output of a standard six-sided die, and let $g(x) = x^2$. Find $E[g(X)] = E[X^2]$

We can use the formula for the expected value of a function of a random variable, together with the pmf table above to get:

$$\begin{aligned} E[g(X)] &= \sum_{\text{all } x} g(x)\, p(x) \\ &= 1\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 9\left(\frac{1}{6}\right) + 16\left(\frac{1}{6}\right) + 25\left(\frac{1}{6}\right) + 36\left(\frac{1}{6}\right) \\ &= \frac{1}{6}(1+4+9+16+25+36) \\ &= \frac{91}{6} \approx 15.17 \end{aligned}$$

Expected Value of a Constant:

If c is a constant, then $E(c) = c$

Combining this with the linearity of expectation, we get:

Expected Value of aY+b:

If Y is a random variable, and a and b are constants, then

$$E(aY + b) = aE(Y) + b$$

3. Variance

Just like the expected value gives the average of a random variable X, the variance describes how much it "spreads"

Example 4: (Motivation)

Let X be the number of heads when you flip a coin 6 times

Let Y be the outcome you get when you roll a 7−sided die with sides $0, 1, \cdots, 6$

Notice that X and Y both take on the same values: $0, 1, \cdots, 6$

On the one hand, $P(Y = y) = 1/7$ for all y

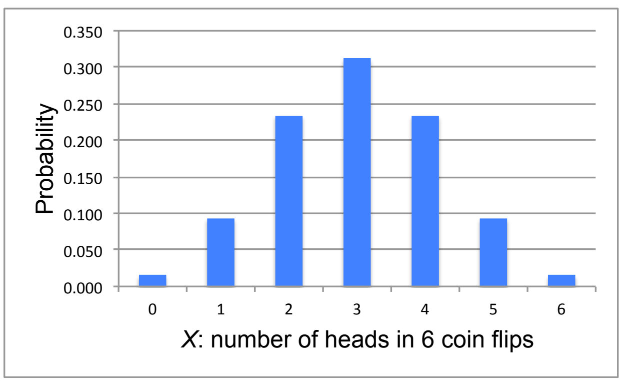
On the other hand, since there are 2^6 possible coin flips, and for the favorable events you choose x heads out of 6, we get

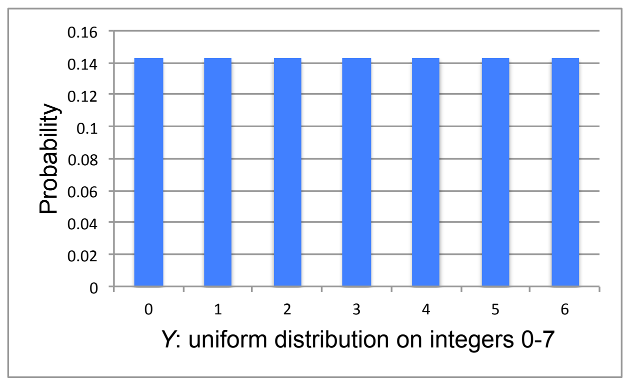
$$P(X = x) = \frac{\binom{6}{x}}{2^6}$$

Putting both pmfs in tables we get:

x	$p(x)$
0	1/64
1	6/64
2	15/64
3	20/64
4	15/64
5	6/64
6	1/64

y	$p(y)$
0	1/7
1	1/7
2	1/7
3	1/7
4	1/7
5	1/7
6	1/7





Despite having the expected value/mean, the two distributions are very different. X is relatively centered about the mean, while Y is all "spread out". We want a way to quantify this amount of "spread".

There are many ways to do this. For example, we could use the average distance from the mean. Statisticians have settled on a slightly different measure of "spread": the variance.

Definition: (Variance)

Let Y be a discrete random variable with probability function $p(y)$, and let $\mu = E(Y)$, then

$$\text{Var}(Y) = E[(Y-\mu)^2] = \sum_{\text{all } y} (y-\mu)^2\, p(y)$$

The **standard deviation** of a random variable is $\sigma = \sqrt{\text{Var}(Y)}$

In other words, the variance of a random variable is the expected value of the squared-difference from the mean

Example 5:

Calculate $\text{Var}(X)$ and $\text{Var}(Y)$, where X and Y are as above

The mean of both random variables is 3, so we have:

$$\begin{aligned}
\text{Var}(X) &= \sum_{\text{all } x} (x-3)^2\, p(x) \\
&= (0-3)^2\frac{1}{64} + (1-3)^2\frac{6}{64} + (2-3)^2\frac{15}{64} + (3-3)^2\frac{20}{64} + (4-3)^2\frac{15}{64} \\
&+ (5-3)^2\frac{6}{64} + (6-3)^2\frac{1}{64} \\
&= \frac{(9)(1)}{64} + \frac{(4)(6)}{64} + \frac{(1)(15)}{64} + \frac{(0)(20)}{64} + \frac{(1)(15)}{64} + \frac{(4)(6)}{64} + \frac{(9)(1)}{64} \\
&= \frac{96}{64} = \frac{3}{2}
\end{aligned}$$

$$
\begin{aligned}
\text{Var}(Y) &= \sum_{\text{all } y} (y-3)^2\, p(y) \\
&= (0-3)^2\frac{1}{7} + (1-3)^2\frac{1}{7} + (2-3)^2\frac{1}{7} + (3-3)^2\frac{1}{7} + (4-3)^2\frac{1}{7} \\
&+ (5-3)^2\frac{1}{7} + (6-3)^2\frac{1}{7} \\
&= \frac{1}{7}(9+4+1+0+1+4+9) \\
&= \frac{28}{7} = 4
\end{aligned}
$$

As predicted from the histograms, Y has a much higher variance.

As you can see, it is a pain to calculate variances using the definition above! Luckily there is a much easier way to do this, known as:

Magic Variance Formula:

$$\text{Var}(Y) = E[Y^2] - \mu^2 \quad \text{where } \mu = E[Y]$$

Why?

$$
\begin{aligned}
\text{Var}(Y) &= E[(Y-\mu)^2] \\
&= E(Y^2 - 2\mu Y + \mu^2) \\
&= E(Y^2) - 2\mu E(Y) + E(\mu^2) \\
&= E(Y^2) - 2\mu^2 + \mu^2 \\
&= E(Y^2) - \mu^2
\end{aligned}
$$

where we used the fact that $E(Y) = \mu$, and $E(c) = c$

Warning: The variance is **NOT** linear! In fact:

Variance of aY + b:

$$\mathrm{Var}(aY+b) = a^2\,\mathrm{Var}(Y)$$

This makes sense because shifting a random variable by b does not affect its spread.

Why?

STEP 1: On the one hand

$$E[(aY+b)^2] = E(a^2Y^2 + 2abY + b^2) = a^2E(Y^2) + 2abE(Y) + b^2$$

STEP 2: On the other hand, using linearity of expectation, we get:

$$[E(aY+b)]^2 = (aE(Y)+b)^2 = a^2E(Y)^2 + 2bE(Y) + b^2$$

STEP 3: Using the Magic Variance Formula, we subtract these two:

$$\begin{aligned}\mathrm{Var}(aY+b) &= E[(aY+b)^2] - [E(aY+b)]^2 \\ &= \left(a^2E(Y^2) + 2abE(Y) + b^2\right) - \left(a^2E(Y)^2 + 2bE(Y) + b^2\right) \\ &= a^2\left(E(Y^2) - [E(Y)]^2\right) \\ &= a^2\,\mathrm{Var}(Y)\end{aligned}$$

Warning: In general, $\mathrm{Var}(X+Y) \neq \mathrm{Var}(X) + \mathrm{Var}(Y)$

Which makes sense, why should $(X+Y)^2 \neq X^2+Y^2$ in the first place?

But this **is** true if the random variables are independent (see later)

Fact:

If $X_1, X_2, \dots, X_n$ are **independent** random variables, then

$$\mathrm{Var}(X_1 + X_2 + \cdots + X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n)$$