Discrete Random Variables

Owen G. Ward

2021-05-11

Introduction

We have already seen random variables, which associates a number to each outcome of an experiment.

If you roll two dice, the sum of those rolls is a random variable.

Similarly, the number of coin flips required until you observe a head is also a random variable.

Potentially, this could be infinite.

Discrete Random Variables

A random variable is discrete if the number of values it can take is finite or countably infinite.

Countably infinite means it takes on values like the natural numbers, $0, 1, 2, \ldots,$.

For a discrete random variable, we want to be able to compute the probability it takes on a specific value,

$$p(X = x).$$

For a die roll if X is the number you get

$$P(X = x) = p_X(x) = \frac{1}{6}, \qquad x = 1, 2, 3, 4, 5, 6$$

We call this function $p_X(x)$ the **probability mass function** or p.m.f of a random variable.

For Y the number of heads when we flip a coin 3 times then

$$p_Y(0) = \frac{1}{8}, p_Y(1) = \frac{3}{8}, p_Y(2) = \frac{3}{8}, p_Y(3) = \frac{1}{8}.$$

For this example easy to count the outcomes, not as obvious if we flip a coin 10 or 100 times.

For a discrete random variable the p.m.f gives us all the information we need to compute more complex quantities.

A p.m.f satisfies some important properties:

- $p_X(x) \geq 0 \ \forall x$
- $\sum_x p_X(x) = 1$, where this sum is over all the values X can take.

Bernoulli Random Variable

Suppose X_B is a random variable which takes value 1 with probability p and 0 with probability $1-p$. What is the pmf for X_B? How do we verify it is a pmf?

If we have an event A then we can compute

$$P(X \in A) = \sum_{x \in A} P(X = x) = \sum_{x \in A} p_X(x).$$

If A is the event roll $X \leq 2$ on a dice then $A = \{X : X \leq 2\}$ and

$$P(A) = p_X(x = 1) + p_X(x = 2) = 1/3.$$

Another important property linked to the p.m.f is the **cumulative distribution function** or c.d.f.

This is defined as $F_X(x) = P(X \leq x)$.

So, for a random variable X, then $F_X(2)$ is the probability X takes a value less than or equal to 2.

We can get this by summing up the probability of X taking every possible value less than or equal to 2.

The cdf has some further properties.

If we plot it it is flat, with jumps at the possible values X can take. The size of those jumps corresponds to the probability of each value.

It is also non decreasing function with $F_X(x) = 0$ when $x \to -\infty$ and $F_X(x) = 1$ when $x \to \infty$.

CDF of Number of Heads in 3 Coin Flips

This is $F_Y(y) = P(Y \leq y)$.

Properties of Random Variables

Expectation

We are often interested in some sort of average value of a random variable. If you toss a coin many times, how many heads would you "expect" to see.

There is a mathematical definition of this, known as the **expectation** of a random variable.

Mathematically it is given by

$$\mathbb{E}(X) = \sum_x x p_X(x),$$

where this sum is over all possible values of the random variable.

If we roll a die with outcome X, then the possible values are $1, 2, 3, 4, 5, 6$ and $p_X(x) = 1/6$ for each.

As such, we can compute

$$\mathbb{E}(X) = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{21}{6}.$$

For Y the number of heads from flipping a coin 3 times then the expected number of heads is

Expectation has some nice properties we can use to answer more complex questions.

Suppose Z is the sum of the roll of 2 Dice. What is $\mathbb{E}(Z)$?

This can take on the values $2, 3, \ldots, 12$ with there being 36 possible combinations.

This gives

$$\mathbb{E}(Z) = \frac{1}{36}(2(1) + 3(2) + 4(3) + 5(4) + 6(5)$$
$$+7(6) + 8(5) + 9(4) + 10(3) + 11(2) + 12(1)) = 7.$$

Is there a quicker way to do this?

If X_1 is the roll of the first die and X_2 the second then $\mathbb{E}(X_1) + \mathbb{E}(X_2) = \mathbb{E}(Z)$.

This is actually the case for any two random variables,

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

This also holds more generally, such as

$$\mathbb{E}(3X + Y) = 3E(X) + \mathbb{E}(Y) \qquad \mathbb{E}(4X - 5Y) = 4\mathbb{E}(X) - 5\mathbb{E}(Y).$$

Expectation is a linear operation.

What is the expectation of a constant?

Expectation of a random variable.

We actually have

$$\mathbb{E}(g(X)) = \sum_x g(x)p_X(x).$$

For example, if X is a dice roll and $g(x) = x^2$.

Then $\mathbb{E}(X^2) = 15.166$ and note this is an example of the rule that $\mathbb{E}(X^2)$ is almost never equal to $\mathbb{E}(X)^2$.

Variance

To quantify how X varies around its expectation $\mathbb{E}(X)$ we define the **variance** of a random variable.

This is given by

$$\sigma^2 = Var(X) = \mathbb{E}(X - \mathbb{E}(X))^2,$$

which can also be written as

$$Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

We can use this for the variance of a dice X to get

$$Var(X) = 15.1666 - (3.5)^2 = 2.92.$$

Variance is not a linear operation like expectation. For example

$$Var(2X) = 4Var(X) \qquad\qquad Var(X+2) = Var(X).$$

More generally

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

for any positive or negative a, b and X and Y are independent. Also have $Var(X) \geq 0$ and that $\sigma = \sqrt{(}Var(X))$ is the **standard deviation** of X.

The variance of a constant is zero.

When X and Y are not independent then

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y),$$

where

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

For any random variable you can compute all these properties once you know the pmf. We will see a few common pmfs and how to compute these things.

Binomial Random Variable

Suppose we flip a fair coin 5 times independently. Let p be the probability that a single flip lands on heads. We want to get the pmf of X, the number of heads.

We know that each coin flip is independent so the probability of getting 3 heads in the first 3 throws is

$$p \cdot p \cdot p \cdot (1-p) \cdot (1-p).$$

Here we only care about the number of heads, not the order. We saw before that to choose 3 from 5 can be done in $\binom{5}{3}$ ways.

So when we account for all the orderings we get

$$P(X=3) = \binom{5}{3} p^3 (1-p)^2$$

So for a general problem, we want to count the number of successes from n trials, where the probability of a success each time is p

Then

$$p_X(x = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

We can write this as

$$X \sim Bin(n, p).$$

and say X follows a Binomial distribution with parameters n, p.

The Binomial distribution counts the number of successes in a fixed number of trials.

There are 4 conditions for data to follow a Binomial distribution:

- The trials are independent
- The number of trials, n, is fixed
- You can get either a success or failure for each trial.

Each trial has the same probability of a success, p.

Example

Suppose we have $X \sim Bin(5, p = 0.2)$. What is $P(X = 3)$?

What is $P(X \leq 3)$?

Recap

Mean and Variance of a Binomial

If $X \sim Bin(n, p)$, what is $\mathbb{E}(X)$? What about $Var(X)$?

To compute this, let's first get the answer when $n = 1$.

It can be shown that when $X \sim Bin(n, p)$ we can actually write

$$X = X_1 + X_2 + \ldots + X_n,\ X_i \sim Bin(1, p), i = 1, 2, \ldots, n$$

So this means

$$\mathbb{E}(X) = \mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n) = p + p + \ldots + p = np.$$

Similarly, these are independent so

$$Var(X) = p(1-p) + \ldots + p(1-p) = np(1-p).$$

We can also get these by using the formula directly, using some properties of combinations.

Uses of the Variance

The variance keeps us an idea of how much observations will vary around their expectation.

In many cases, very likely a value will be with 3 standard deviations of its mean. Will see a simulation of this.

Example

The National Vaccine Information Center estimates that 90% of Americans have had chickenpox by the time they reach adulthood.

- Suppose we take a random sample of 100 American adults. Is the use of the binomial distribution appropriate for calculating the probability that exactly 97 out of 100 randomly sampled American adults had chickenpox during childhood?

- Calculate the probability that exactly 97 out of 100 randomly sampled American adults had chickenpox during childhood.

- What is the probability that at most 3 out of 10 randomly sampled American adults have not had chickenpox?

Poisson Random Variable

Suppose we have a random variable X which can take on the values $0, 1, 2, ...$ For example, we want to say the number of people in a store is X.

The **Poisson** distribution is commonly used for data like this, with pmf

$$P(Y = k) = p_Y(y = k) = \frac{\lambda^k}{k!}e^{-\lambda},$$

and parameter λ.

So if $Y \sim Poisson(\lambda)$ then

$$P(Y = 0) = e^{-\lambda},\ P(Y = 1) = \lambda e^{-\lambda}, ...$$

Mean and Variance of Poisson

Can show that if $Y \sim Poisson(\lambda)$ then

$$\mathbb{E}(Y) = \lambda,\ Var(Y) = \lambda.$$

The Poisson is actually related to the Binomial distribution. For a Binomial random variable with large n and small p then this becomes a Poisson random variable in the limit, with $np \to \lambda$.

x,y	x=0	x=1	x=2	x=3
y=1	0.04	0.23	0.05	0.10
y=2	0.12	0.15	0.02	0.05
y=3	0.10	0.09	0.01	0.04

Example

Suppose you have a book with 600 pages. The probability of a mistake on each page is $p = 0.005$ and each page is independent. What is the probability there is exactly two pages with an error?

Using $X \sim Binom(n = 600, p = 0.005)$ we get $P(X = 2) = 0.22423$.

Using an approximate Poisson distribution with $\lambda = 600 \times 0.005 = 3$ then for $Y \sim Poisson(\lambda = 3)$,

$$P(Y = 2) = \frac{\lambda^2}{2!}e^{-\lambda} = 0.22402,$$

a very similar value.

Example

A very skilled court stenographer makes one typographical error (typo) per hour on average.

- What probability distribution is most appropriate for calculating the probability of a given number of typos this stenographer makes in an hour?
- What are the mean and the standard deviation of the number of typos this stenographer makes?

- Would it be considered unusual if this stenographer made 4 typos in a given hour?
- Calculate the probability that this stenographer makes at most 2 typos in a given hour.

Joint Distributions

We have seen methods for distributions for single random variables. Often we actually want to look at the distribution of two random variables **jointly**.

For example, the distribution of a persons weight and height.

For two discrete random variables X, Y the joint pmf is

$$p_{X,Y}(x, y) = P(X = x, Y = y).$$

For discrete cases this will often take the form of tables describing the pmf.

Want to answer questions like:

- $P(X = 0)$ (marginal distribution)
- $P(X = 0|Y = 1)$ (conditional distribution)
- $\mathbb{E}(XY)$ (joint expectation)
- Whether X is independent of Y?

Marginal Distribution

We can get the marginal distribution of X by summing over all possible values of Y.

$$P(X = x) = \sum_y P(X = x, Y = y).$$

So for $P(X = 0)$ we get

$$P(X = 0) = P(X = 0, Y = 1) + P(X = 0, Y = 2) + P(X = 0, Y = 3).$$

Conditional Distribution

Here we use Bayes rule again to get

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

For $P(X = 0|Y = 1)$

Expectation of Joint Distributions

For a function $g(X, Y)$ we have

$$\mathbb{E}g(X, Y) = \sum_{x,y} g(x, y)p_{X,Y}(x, y).$$

For $g(X, Y) = XY$ then this gives

Independence of Joint Random Variables

We have that two random X, Y are independent if and only if

$$p_{X,Y}(x, y) = p_X(x)p_Y(y),$$

for all pairs x, y.

For the distribution above, are X and Y independent?