9/20/2022

Discrete probability distributions

Random variables come in two numerical flavors: discrete and continuous, and therefore so do probability distributions. We'll look at the discrete case in this lecture and the continuous case in the next.

Discrete distributions can be presented either as a formula, or in a table.

x	0	1	2	3	4	5	6
$p(x)$	.05	.10	.15	.25	.20	.15	.10

The table above shows an example of a table display of a discrete random variable.

An example with a piecewise formula could look like this, one with two outcomes:

$$p(x,\alpha) = \begin{cases} \alpha, & if\ x = 0 \\ 1-\alpha, & if\ x = 1 \\ 0, & otherwise \end{cases}$$

In this example the variable can only take on either the value 0 or 1, and α is some parameter whose value is between 0 and 1.

Recall the rules of probability that our probability function must obey:

- All individual probabilities must be between 0 and 1
- All probabilities must add up to 1

To verify the table or formula represents a probability distribution, check these two properties. If one value is missing from the table, use the complement rule to find it.

A distribution like this of values for individual outcomes of a random variable is sometimes referred to as a probability mass function or probability density function (pmf or pdf). We can also construct cumulative mass functions or cumulative density functions as we discussed when dealing with frequency tables.

The discrete pdf in the table above would become:

x	0	1	2	3	4	5	6
$p(x)$	0.05	0.15	0.30	0.55	0.75	0.90	1

The piecewise function would become:

$$F(x;\alpha) = \begin{cases} 0, & if\ x < 0 \\ \alpha, & if\ 0 \le x < 1 \\ 1, & if\ x \ge 1 \end{cases}$$

Typically, pdfs in function form take lower case function names and cdfs take capital function names.

We can find descriptive numerical values (parameters) for these distributions just as we can for observations.

The mean is also called the expected value and notated as either $E(X)$ or μ. The expected value is calculated similarly to a weighted average where the probabilities are the weights. The formula is a little simpler since the sum of the probabilities must add to 1.

$$E(X) = \sum_{i=1}^{n} x_i p(x_i)$$

For the pdf in the table this calculation is:

$$E(X) = 0(0.05) + 1(0.1) + 2(0.15) + 3(0.25) + 4(0.2) + 5(0.15) + 6(0.1) = 3.3$$

The median is the value where the cumulative distribution reaches or exceeds 50%. For this pdf, that value is 3.

The variance formula for the same distribution:

$$V(X) = \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i)$$

The standard deviation is, of course, the square root of the variance. $\sigma = \sqrt{V(X)}$.

There is an alternative formula for calculating the variance which uses the second moment.

The first moment is just the regular expected value. The second moment is similar but replaces x_i in the formula with x_i^2. (By extension, we can calculate third and fourth moments, etc. by changing the powers in a corresponding way for each moment.)

$$E(X^2) = \sum_{i=1}^{n} x_i^2 p(x_i)$$

Then the variance can be calculated as $V(X) = \sigma^2 = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$

This formula requires slightly fewer operations so it is generally preferred.

$$E(X^2) = 0^2(0.05) + 1^2(0.1) + 2^2(0.15) + 3^2(0.25) + 4^2(0.2) + 5^2(0.15) + 6^2(0.1) = 13.5$$

$$V(X) = 13.5 - (3.3)^2 = 2.61$$

$$\sigma \approx 1.616$$

I leave it to you to confirm that this produces the same result as the initial version of the variance formula.

There are a couple of other statistics we can calculate using moments. There is a skewness statistic based on a formula similar to the variance but with cubes. The kurtosis formula uses the 4th moment. We'll say more about these and their calculations in the next lecture. I've included links in the references if you want to explore them before then.

Let's look at a problem where we might want to apply an expected value to a discrete distribution.

Suppose that you attend a raffle that is fundraising for your kid's school. The PTA is selling raffle tickets for $15 each. They are selling 500 tickets, and are giving the following prizes away: one 1st place prize for $1000, one 2nd place prize for $500, a 3rd place prize for $100, and five $10 prizes for 4th place. What is the expected value of purchasing a ticket?

Begin by building your probability table. Since you are paying to play, the net prizes are the values stated minus the cost of buying the ticket.

x	**985**	**485**	**85**	**-5**	**-15**
$p(x)$	$\frac{1}{500}$	$\frac{1}{500}$	$\frac{1}{500}$	$\frac{5}{500}$	$\frac{492}{500}$

The expected value then is $985\left(\frac{1}{500}\right) + 485\left(\frac{1}{500}\right) + 85\left(\frac{1}{500}\right) - 5\left(\frac{5}{500}\right) - 15\left(\frac{492}{500}\right) = \$ - 11.70$

This means that for every ticket you buy, you can expect to lose $11.70 on average. This is good news for the PTA since that means they will make $11.70 per ticket.

What happens to the mean and variance if we transform the random variable with a linear operation? Such as $Y = aX + b$?

$$E(Y) = E(aX + b) = aE(X) + b$$

We can just transform the mean the same way.

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

A shift does not change the spread and so it doesn't change the variance, but the scalar multiple of X does change the variance by the square of the multiple. The standard deviation would square linearly except that the scaling factor will be positive.

$$st.dev(Y) = st.dev(aX + b) = |a|st.dev(X)$$

Standard deviation cannot be negative.

A Bernoulli random variable is a variable that has only 0 or 1 as an outcome. Any variable that has only two outcomes can be designated a Bernoulli random variable just by renaming the outcomes to either 0 or 1. For instance, a coin flip: call heads 1 and call tails 0.

Binomial distribution

There are several special discrete distributions that show up relatively frequently. We are going to look briefly at a couple of these. The most important is probably the binomial distribution (followed by Poisson).

For a random variable to be distributed by the binomial distribution it must have the following properties:

- There is a sequence of experiments called trials that are fixed in number (n).
- Each trial results in only two possible outcomes. Success is called 1, and failure is called 0.
- Each trial is independent (the outcome of one trial does not affect any that follow).
- The probability of success (p) with each trial is fixed.

The probability calculated is the probability of achieving a certain number of outcomes in the set of trials, for instance, the probability of getting 6 heads in 10 coin flips.

The binomial probability formula (pdf) is:

$$P(X = x) = B(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Sometimes this formula is written as

$$B(x; n, p) = \binom{n}{x} p^x q^{n-x}$$

Where q is understood to mean $1 - p$.

The two values n and p are parameters; x is the number of successes. The parentheses at the beginning is the combination rule $\binom{n}{x} = \frac{n!}{(n-x)!x!}$.

We saw how to calculate the probability of 6 head in 10 flips if the coin is fair using counting rules. But what if the coin is not fair? Then we can use the binomial formula. Suppose the probability of heads is actually 60% and not 50%? Then

$$B(6; 10, 0.6) = \binom{10}{6} (0.6)^6 (0.4)^4 \approx 0.2508$$

Example. Suppose the probability that any one person getting into a car puts on a seatbelt is 90%. An experimenter watches 20 people get into the car.
What is the probability that exactly 18 people put on their seatbelts?

$$B(18; 20, 0.9) = \binom{20}{18} (0.9)^{18} (0.1)^2 \approx 0.2852$$

What is the probability that more than 18 people will put on their seatbelts?
This means that either 19 or 20 people put them on?

$$B(19;20,0.9) + B(20;20,0.9) = \binom{20}{19}(0.9)^{19}(0.1)^1 + \binom{20}{20}(0.9)^{20}(0.1)^1 \approx 0.3917$$

Many technology programs also have cumulative distribution (cdfs) of these common probability distribution functions. Using that we can find the same number by using the cdf:

$$P(X > 18) = 1 - B_{cdf}(18;20,0.9)$$

Some distributions will have nice formulas for the cdf, but the binomial really doesn't. But you can let the technology do the extra computation behind the scenes, especially if there are a large number of cases.

Take care with the inequalities in these expressions. They can be quite tricky because $P(X \geq 18) \neq P(X > 18)$ for all discrete distributions.

While it's possible to calculate the mean and variance using our previous formulas, special distributions like this can derive general formulas for any distribution of this type. We won't derive them since many of the proofs here require working with series notation, and in some later cases, infinite series, which are not covered in the prerequisites for this course. We will show the results.

The expected value of a binomial experiment is $E(X) = \mu = np$
The variance of a binomial experiment is $V(X) = \sigma^2 = npq = np(1-p)$
The standard deviation is derived from the variance in the usual way.

Other discrete distributions

The **hypergeometric distribution** has the following properties:

- There is a finite population of N individuals.
- The variable is Bernoulli (outcomes are success or failure).
- A sample of n individuals is selected without replacement (each subset of size n is equally likely).

The pdf is:

$$P(X = x) = h(x;n,M,N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

Where x is the number of successes, n is the sample size, M is the number of "successes" in the whole population, and N is the population size.

The formula is derived from counting methods discussed in the last lecture.

Suppose you are randomly selecting from a population of 30 rocks samples, 10 of which are gneiss and the remaining 20 are sandstone. You randomly select 5 rocks. What is the probability that none of them are gneiss?

$$P(X = x) = h(0; 5,10,30) = \frac{\binom{10}{0}\binom{20}{5}}{\binom{30}{5}} \approx 0.1088$$

The expected value of the hypergeometric distribution is $E(X) = n\left(\frac{M}{N}\right)$.
The variance is $V(X) = \left(\frac{N-n}{N-1}\right) n\left(\frac{M}{N}\right)\left(1 - \frac{M}{N}\right)$.

Note: The expression $\left(\frac{N-n}{N-1}\right)$ is sometimes called the finite population correction factor.
Without it, notice the similarities to the binomial distribution if you replace $\frac{M}{N}$ with p.
In fact, you can use the binomial distribution to approximate the hypergeometric distribution if the sample size is small enough relative to the population, usually <5%. In that scenario, the change to the probabilities after each draw is small enough that it can be neglected in most applications.

The **negative binomial distribution** has the following properties:

- The experiment is a sequence of independent trials.
- The variable is Bernoulli (i.e. outcomes are success or failure)
- The probability is constant between trials.
- The experiment continues until r successes are achieved.

The formula for the pdf of a negative binomial probability is:

$$P(X = x) = nb(x; r, p) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x$$

Where x is the number of failures until r successes, p is the constant probability for each trial, and the r is the desired number of successes.

Sometimes this distribution is expressed in terms of the total number of trials, $x + r$ instead.

An example application of this distribution is a family that would like to have at least one girl and won't stop having children until that happens.

The expected value of this distribution is $E(X) = \frac{r(1-p)}{p}$.
The variance of this distribution is $V(X) = \frac{r(1-p)}{p^2}$.

One common variation of this distribution is when $r = 1$ (as in the example above). Then the distribution is referred to as the **geometric distribution** whose formula is

$$g(x; p) = (1 - p)^x p$$

Where x is the number of failures before achieving success. You can use the expected value and variance formulas from the negative binomial distribution here, just with $r = 1$.

The **Poisson distribution** is a little different than our previous cases, but is still a discrete distribution because it counts the number of events that occur in a given period of time (not in a number of trials).

The pdf for the Poisson distribution is given by:

$$P(X = x) = p(x; \mu) = \frac{e^{-\mu}\mu^{x}}{x!}$$

Where x is the number of occurrences, and μ is the mean number of occurrences in a given time frame. The value x must be an integer since it is a count, but μ need not be.

The Poisson distribution models what is called a Poisson process. We can think of this in two different ways: either the count of events in a fixed period of time, or the time it takes between events. The latter will be covered in the next lecture since time is a continuous variable, and it follows the exponential distribution.

Suppose that you are given the mean number of events in an hour: say, 18 customers come to your drive up window, on average, at a particular hour of the day. You are asked to find the probability of exactly 3 customers in the next 20 minutes. You can scale the μ parameter linearly to account for the different time frame. Since 20 minutes is $\frac{1}{3}$ of an hour, we can expect $\frac{18}{3} = 6$ customers in 20 minutes. We can then calculate the probability with that value.

$$p(3; 6) = \frac{e^{-6}6^{3}}{3!} \approx 0.0892$$

The mean and variance of a Poisson distribution are the same: $E(X) = V(X) = \mu$

I've included a link in the references below in case you want to read more about Poisson.

In the next lecture we'll talk about the continuous probability distribution case. Time to refresh those calculus skills. You are going to need them. And we'll briefly dip our toes in multivariable calculus while we are at it. Buckle up!

References:

1. https://faculty.ksu.edu.sa/sites/default/files/probability_and_statistics_for_engineering_and_the_sciences.pdf
2. https://assets.openstax.org/oscms-prodcms/media/documents/IntroductoryStatistics-OP_i6tAI7e.pdf
3. https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm
4. https://www.randomservices.org/random/expect/Skew.html
5. https://en.wikipedia.org/wiki/Sim%C3%A9on_Denis_Poisson