PLSC 502: "Statistical Methods for Political Research"

Random Variables
September 22, 2016

Random Variables

Last time, we discussed probability theory, which we expressed in terms of a variable X. We defined X as a set of realizations of some process, which in turn is governed by rules of probability regarding potential outcomes in the sample space.

The variables we were talking about have been what are called *random variables*, in that they all have a theoretical probability distribution. In a sense, a random variable is a set of a specific kind. As we noted before, broadly speaking, there are two kinds of random variables: *discrete* and *continuous*.

- *Discrete* variables can take on any one of several distinct, mutually-exclusive values.
  - Either *finite* or *countably infinite*.
  - e.g. Congressperson's ideology score $\{0, 1, 2, 3..., 100\}$.
- *Continuous* can take on *any* value in its range.
  - Uncountably infinite number of elements
  - E.g. time, or temperature, or $Y \in [0, 1]$.
  - Most continuous variables are measured discretely.

The intuition of probability on the two types is identical; the only thing that is different is the math (since the latter require us to use a little calculus...).

The Probability Density Function (PDF)

If X is some random variable, then the probability that $X =$ some value x in the range of X defines the *probability density function* (PDF). That is, the PDF is the function that maps the values of X to some associated probability of their occurrence.

For *discrete* random variables, we write

$$f(X) = \Pr(X = x) \, \forall \, x \tag{1}$$

Note that, thanks to our rules of probability, we know (e.g.) that

$$\begin{aligned} \sum_x \Pr(X = x) &= \sum_x f(X) \\ &= 1.0. \end{aligned}$$

That is, the probabilities of all (discrete) events sum to 1.0.

Continuous Random Variables

As I note above, a probability function $f(\cdot)$ defined on a set of real numbers S is called *continuous* if there exists a function $f(x)$ such that for any closed interval $[a, b]$ is $\Pr(a < x \leq b) = \int_b^a f(x)dx$.

This means that, for a function to be a continuous probability function it must satisfy two properties.

- $f(x) \geq 0$ for all x
- $\int_{-\infty}^{\infty} f(x)dx = 1$

For a *continuous* probability density function $f(X)$, the probability associated with any given point is zero.[1] Instead, we consider the probability that X takes on a value in some range from (say) A to B:

$$\Pr(a < X \leq b) = \int_a^b f(x)\,dx \tag{2}$$

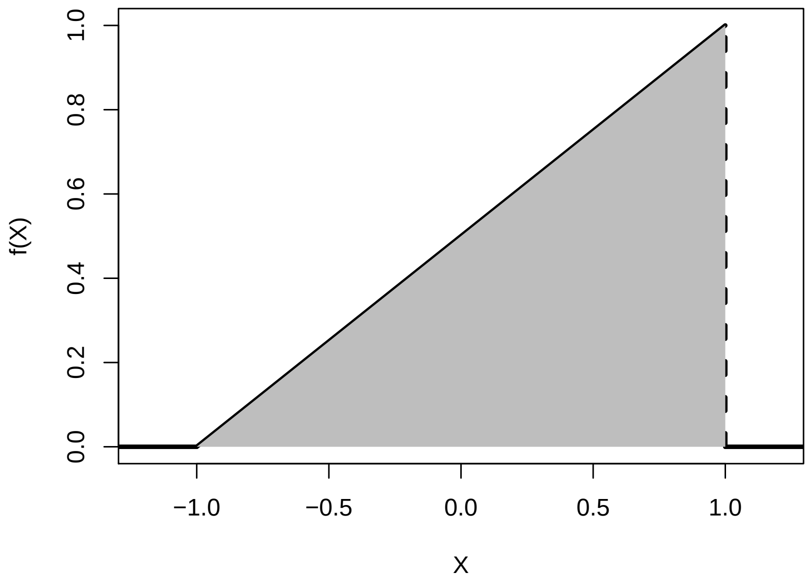
where $f(X)$ is the *probability density function* for (continuous) X. Note that the rules of probability above mean that, for a continuous PDF $f(X)$, $\int_{-\infty}^{\infty} f(X)\,dX = 1.0$.

An Example

Consider the function:

$$f(x) = \begin{cases} \frac{1}{2}(x+1) & \text{for} -1 \leq y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

[1]This is because, for a truly continuous variable, there is an infinitesimally small chance that it will take on precisely some value. So, for example, the probability that the temperature in State College, PA right now is 76.2301198034812... degrees Fahrenheit is, for all practical purposes, zero.



How do we know if this equation is a PDF? To be so, it must satisfy two properties, the first being that $f(x) \geq 0$ for all x. We can verify this visually.

Next, to be a PDF the following must also be true:

$$\Pr(-\infty \leq x \leq \infty) = \Pr(-1 \leq x \leq 1) = \int_{-1}^{1} f(x)dx = 1$$

To verify this we must evaluate the following integral:

$$\begin{aligned}
\Pr(-1 \leq Y \leq 1) &= \int_{-1}^{1} \frac{1}{2}(x+1)dx \\
&= \frac{1}{2}\left(\frac{x^2}{2} + x\right)\Big|_{-1}^{1} \\
&= \frac{1}{2}\left(\frac{1^2}{2} + 1\right) - \frac{1}{2}\left(\frac{-1^2}{2} - 1\right) \\
&= 0.75 - (-0.25) \\
&= 1.
\end{aligned}$$

This confirms that our function $f(X)$ is a PDF.

The Cumulative Distribution Function (CDF)

If X is at least ordinal, we may be interested in the probability that X will take on a value (say) less than or equal to than some value (say, x) in its range.

In the discrete case, this is equal to:

$$\begin{aligned}
\Pr(X \leq x) &= \sum_{X \leq x} \Pr(X = x) \\
&= 1 - \sum_{X > x} \Pr(X = x) \qquad (3)
\end{aligned}$$

This is the *cumulative distribution function* (CDF), which is sometimes written using capital letters: $f(X) \rightarrow F(X)$.

For continuous variables, we want to know the probability that $X \leq x$, i.e.

$$\Pr(X \leq x) \equiv F(x) = \int_{-\infty}^{x} f(t)\, dt \qquad (4)$$

More generically, the CDF is the indefinite integral of the PDF:

$$F(X) = \int f(t)\, dt \qquad (5)$$

These expositions tell us something:

- *The first derivative of the CDF (taken with respect to X) is the PDF*, and
- *the integral of the PDF is the CDF...*

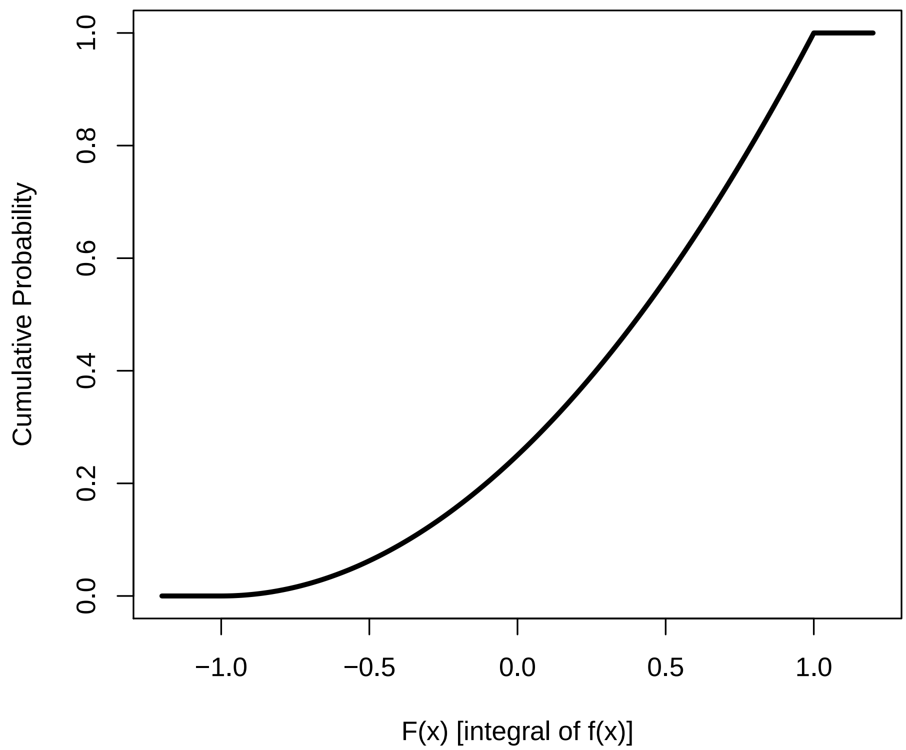
Moreover, if we write the CDF as $F(X)$, note that:

- $0 \leq F(X) \leq 1$.
- $\Pr(X < k) = 1$ - F(k).
- $\Pr(a < X \leq b) = F(b) - F(a)$.
- $F(-\infty) = 0$.
- $F(\infty) = 1$.

If we reconsider our example function $f(X)$, above, then the corresponding CDF is

$$
\begin{aligned}
F(X) &= \int f(t)dt \\
&= \int \frac{1}{2}(X+1)dX \\
&= \frac{1}{2}\left(\frac{1}{2}X^2 + X\right) + c.
\end{aligned}
$$

Think of this as the "area under the curve" going from -1 to 1. We can see what this looks like by plotting the function over a range of values for X:



Note that because the PDF is increasing in x, the rate at which probability "accumulates" increases as we go from left to right.

Beyond this, we need to get into some more specific aspects of random variables, including their characteristics (i.e., moments, etc.) and then to some specific distributions.

Expected Values, Variance, and Moments

Let X be a real-valued, interval-level random variable, where $f(x)$ is X's PDF and $F(x) = \int_{-\infty}^{x} f(x)\, dX$ is its associated CDF.

Expected Value

The expected value of X, denoted $\mathrm{E}(X)$:

- is equal to the probability-weighted mean of the values of that variable;
- can be thought of as Σ(Value$\times$Probability) for all possible outcomes;
- is often denoted μ.

The expected value of a variable is similar to the *mean* (it can be thought of as the mean of a theoretical probability distribution). In fact, it's almost exactly like the mean in the way its calculated...

For a discrete variable,

$$\mathrm{E}(X) = \sum x f(x) \tag{6}$$

- The sum of each possible value, times its probability of occurrence.
- $\mathrm{E}(X)$ need not be an actual value taken on by X...
- Example:
  - Flip a coin twice, and record the number of heads that come up...
  - What is the expected number of heads?
  - Probability distribution:

0 Heads	Prob. = .25	Prob×Value = .25×0	=	0
1 Head	Prob. = .50	Prob×Value = .50×1	=	.50
2 Heads	Prob. = .25	Prob×Value = .25×2	=	.50
		$\sum$	=	1.0

In the continuous case:

$$\mathrm{E}(X) = \int X \times f(x)dx \tag{7}$$

The intuition is the same as for the discrete case: $\mathrm{E}(X)$ is a "typical" value, gained by "summing" across values of X...

In our example function, what is the expected value (mean) of $f(X)$?

$$
\begin{aligned}
\mathrm{E}[f(X)] &= \int_{-1}^{1} X\left(\frac{X+1}{2}\right) dX \\
&= \int_{-1}^{1} \frac{1}{2}(X^2+X)\, dX \\
&= \frac{1}{2}\int_{-1}^{1} x^2\, dX + \frac{1}{2}\int_{-1}^{1} X\, dX \\
&= \frac{1}{2}\left(\frac{X^3}{3}+c_1\right)\Big|_{-1}^{1} + \frac{1}{2}\left(\frac{X^2}{2}+c_2\right)\Big|_{-1}^{1} \\
&= \frac{1}{2}\left(\frac{X^3}{3}+\frac{X^2}{2}+c_3\right)\Big|_{-1}^{1} \\
&= \frac{1}{2}\left[\left(\frac{1^3}{3}+\frac{1^2}{2}+c_3\right)-\left(\frac{-1^3}{3}+\frac{-1^2}{2}+c_3\right)\right] \\
&= \frac{1}{3}
\end{aligned}
$$

Properties of Expected Values

Note that expectations have lots of useful properties. In particular, for any real-valued X and $f(x)$,

- $\mathrm{E}(c) = c$ (The expectation of a constant is the constant).
- $\mathrm{E}(cX) = c\mathrm{E}(X)$ (The expectation of a variable and a constant is the constant times the expectation of the variable).
- $\mathrm{E}(x+y+z) = \mathrm{E}(x)+\mathrm{E}(y)+\mathrm{E}(z)$ (summation).
- If $g(x)$ is some function of x, then

$$
\begin{aligned}
\mathrm{E}[g(x)] &= \sum[g(x)\times \mathrm{Prob}(X=x)]\,\forall x \text{ (discrete case)} \\
&= \int g(x)f(x)\,dx \text{ (continuous case)}
\end{aligned}
$$

- This includes a constant function, so $\mathrm{E}(ax) = a\mathrm{E}(x)$.
- These two mean that in the case of a linear function $g(x) = a+bx$, $\mathrm{E}(a+bx) = a+b\mathrm{E}(x)$.

We'll come back to expected values later, when we discuss distributions.

Variance of a Random Variable

In addition to central tendency, we are also interested in a variable's *dispersion*. As we noted before, in general,

$$\mathrm{Var}(X) = \mathrm{E}[(x-\mu)^2] \tag{8}$$

In the discrete case, we have...

$$\mathrm{Var}(X) = \sum (x-\mu)^2 f(x) \tag{9}$$

The continuous case is:

$$\mathrm{Var}(X) = \int (x-\mu)^2 f(x)\, dx \tag{10}$$

We can simplify this further...

$$\begin{aligned}
\mathrm{E}[(x-\mu)^2] \equiv \sigma^2 &= \mathrm{E}[x^2 - 2x\mu + \mu^2] \\
&= \mathrm{E}(x^2) - 2\mu\mathrm{E}(x) + \mathrm{E}(\mu^2) \\
&= \mathrm{E}(x^2) - 2\mu^2 + \mu^2 \\
&= \mathrm{E}(x^2) - \mu^2 \\
&\equiv \left(\int x^2 f(x)\, dx - \mu^2\right)
\end{aligned} \tag{11}$$

This can be a useful formula...

- It implies that the expectation of the square of a variable X is $\mathrm{E}(X^2) = \sigma^2 + \mu^2$.
- We often write the variance as σ^2, and the positive square root of it (the standard deviation) as σ.

Properties of Variance

- $\mathrm{Var}(c) = 0$ (the variance of a constant is zero).
- $\mathrm{Var}(a + bx) = b^2\mathrm{Var}(x)$ (note that a drops out here...).

Example Again

For our example above: What is the variance of $f(X) = \frac{1}{2}(x+1)$ for $-1 < x < 1$?

We know that $\mu = 1/3$; which means:

$$
\begin{aligned}
\mathrm{E}[(x-\mu)^2] \equiv \sigma^2 &= \int_{-1}^{1} X^2 f(x)\,dx - \mu^2 \\
&= \int_{-1}^{1} \frac{1}{2}x^2(x+1)\,dx - \left(\frac{1}{3}\right)^2 \\
&= \frac{1}{2}\left(\frac{x^4}{4}+c_1\right)\Big|_{-1}^{1} + \frac{1}{2}\left(\frac{x^3}{3}+c_2\right)\Big|_{-1}^{1} - \frac{1}{9} \\
&= \frac{1}{2}\left(\frac{X^4}{4}+\frac{X^3}{3}+c_3\right)\Big|_{-1}^{1} - \frac{1}{9} \\
&= \frac{1}{2}\left[\left(\frac{1^4}{4}+\frac{1^3}{3}+c_3\right) - \left(\frac{-1^4}{4}+\frac{-1^3}{3}+c_3\right)\right] - \frac{1}{9} \\
&= \frac{19}{72}\ (\approx 0.2639).
\end{aligned}
$$

Back to Moments, for a Moment...

We discussed *moments* last week. As with empirical distributions, the kth moment of a random variable is

$$M_k = \mathrm{E}(X^k) \tag{12}$$

where k is a positive integer. For a random variable X, the kth moment exists iff

$$\mathrm{E}(|X|^k) < \infty.$$

For a continuous variable X, then, this can be written as

$$\mathrm{E}(|X|^k) = \int_{-\infty}^{\infty} |x|^k f(x)dx < \infty \tag{13}$$

If X is bounded (e.g., if X is a proportion/percentage, or a count of "successes" out of a finite number of trials,etc.) so that $\Pr(a \leq X \leq z)$, then all moments exist. They might also exist even if X is unbounded, however; moreover, if the kth moment exists (but the $\{k+1, k+2, ...\}$th do not) then it can be shown that all lower moments exist as well.

As we discussed before, we are usually interested in the *central moments*, defined as:

$$\mu_k = \mathrm{E}[(X-\mu)^k] \tag{14}$$

This means that the first central moment of any variable X (if it exists) is equal to zero, the second is the variance, the third is skewness, etc. Eq. (14) makes clear that, for any *symmetrical* variable X, the odd-numbered central moments will be zero (since the positive and negative values of X will "cancel each other out").

Moment-Generating Functions

For any real-valued t, the function

$$\psi(t) = \mathrm{E}[\exp(tX)] \tag{15}$$

is known as the *moment-generating function* (MGF) for X. If X is continuous, then we can write (15) as

$$\begin{aligned}
\psi(t) &= \int_{-\infty}^{\infty} \exp(tx) f(x) dx \\
&= \int_{-\infty}^{\infty} \left(1 + tx + \frac{t^2 x^2}{2!} + \ldots\right) f(x) dx \\
&= 1 + t\mathrm{E}(X) + \frac{t^2 \mathrm{E}(X^2)}{2} + \ldots \\
&= 1 + tM_1 + \frac{t^2 M_2}{2} + \ldots \qquad (16)
\end{aligned}$$

That is, we can write the tth MGF as an infinite sum of functions of the t moments of X.

It can be shown that (15) exists for all values of t if X is bounded (as above); for unbounded variables, (15) may exist for some values of t but not for others. Note, however, that (15) always exists at $t = 0$, since

$$\begin{aligned}
\psi(0) &= \mathrm{E}[\exp(0)] \qquad (17) \\
&= 1
\end{aligned}$$

As it turns out, though, the MGF in the *neighborhood* of zero is far more interesting...

MGFs have at least two very useful characteristics. First, the formulation above suggests that the MGF of a random variable can be (re)written as a function of its CDF:[2]

$$\psi(t) = \int_{-\infty}^{\infty} \exp(tx)\, d\, F(x). \tag{18}$$

That is, the tth MGF is the integral of $\exp(tx)$ with respect to the CDF. Second, so long as the MGF exists in an open interval around $t = 0$, we can express the moments of X as functions of it, via the relationship in (16):

[2] This uses something known as a *Riemann-Stieltjes integral*; that will *not* be on the final...

$$\begin{aligned}
\left.\frac{\partial^k \psi(t)}{\partial^k t}\right|_{t=0} &= \left.\frac{\partial^k \mathrm{E}[\exp(tX)]}{\partial^k t}\right|_{t=0} \\
&= \mathrm{E}\left[\left.\frac{\partial^k \exp(tX)}{\partial^k t}\right|_{t=0}\right] \\
&= \mathrm{E}\{[X^k \exp(tX)]|_{t=0}\} \\
&= \mathrm{E}(X^k) \qquad (19)
\end{aligned}$$

This "works" because we can differentiate $\psi(t)$ with respect to t an arbitrary number of times at $t = 0$. Thus, the first moment of X is the same thing as the first derivative of $\psi(t)$ with respect to t, evaluated at $t = 0$ (and is equal to $\mathrm{E}(X)$); the second is $\mathrm{E}(X^2)$, etc. We'll discuss these a bit more when we get into our classes on various probability distributions.