Vanderbilt University Political Science - Stats I
Fall 2024 - Prof. Jim Bisbee

Lecture 4

0.1 Continuous random variables

- We often deal with random variables that can take on an uncountably infinite number of values. These RVs are known as *continuous* random variables.

- The reason we care: it's impossible to assign nonzero probabilities to all the uncountably infinite points on an interval while satisfying that they all sum to 1. Thus the notion of $p(y)$ from discrete world is irrelevant here. We must develop a different method to describe the probability distribution of a continuous RV.

- Let's begin by defining the cumulative distribution function (or cdf, or "distribution function") of the RV Y as $F(y)$, where

$$F(y) \equiv P(Y \leq y) \text{ for } -\infty < y < \infty.$$

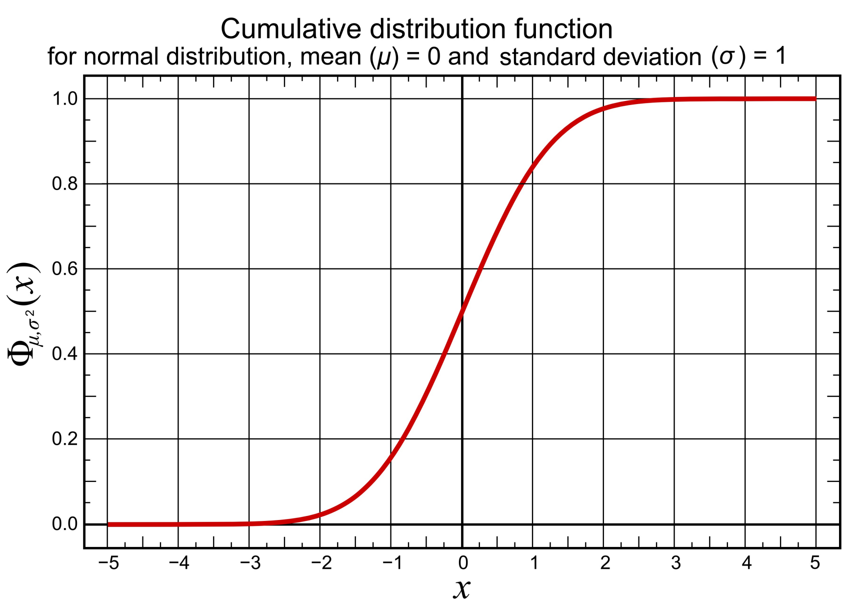

Figure 1: CDF

- A cdf has the following properties:

$$
\begin{aligned}
F(-\infty) &\equiv \lim_{y\to\text{-}\infty} F(y) = 0 \\
F(\infty) &\equiv \lim_{y\to\infty} F(y) = 1 \\
& F(y) \text{ is a nondescreasing function of } y\text{, which means that} \\
y_1 &< y_2 \Longrightarrow F(y_1) \le F(y_2).
\end{aligned}
$$

- Both discrete and continous RVs have cdfs. See your text [page 158] for an example of how to develop a cdf of an RV with a binomial distribution . A RV Y with cdf $F(y)$ is said to be **continuous** if $F(y)$ is continuous for -$\infty < y < \infty$. By contrast, the cdf's of discrete RVs are always **step** functions: they have discontinuities separating the possible values of y that they can take on.

- Remember that we talked earlier about the difficulty of assigning probabilities to points? Well with continuous RVs, we don't. In fact if Y is a continuous RV,

$$P(Y = y) = 0 \forall \text{ real numbers } y.$$

- Sounds weird, but isn't. When we move to the world of continuous RVs, the differences between values of y become infinitely small, making the chance that we see any one particular value of y zero.

  - For example, what's the probability of observing a temperature of 50.73093764 degrees Fahrenheit on October 2 in New York City? Now add 10 additional random digits to this number. Do it again. It doesn't take long to get to values for which it would be quite likely we would never see even if we were to measure the temperature on all October 2s that ever exist. And even *those* values can be made more precise.

- Instead, we get at this notion with the idea of **density**. Define the function $f(y)$ as the derivative of f:

$$f(y) \equiv \frac{dF(y)}{dy} = F'(y).$$

- Wherever the derivative exists, $f(y)$ is the *probability density function* (pdf, 'density function') for the RV Y. [NOTE: The following diagram should be labeled "The density function"

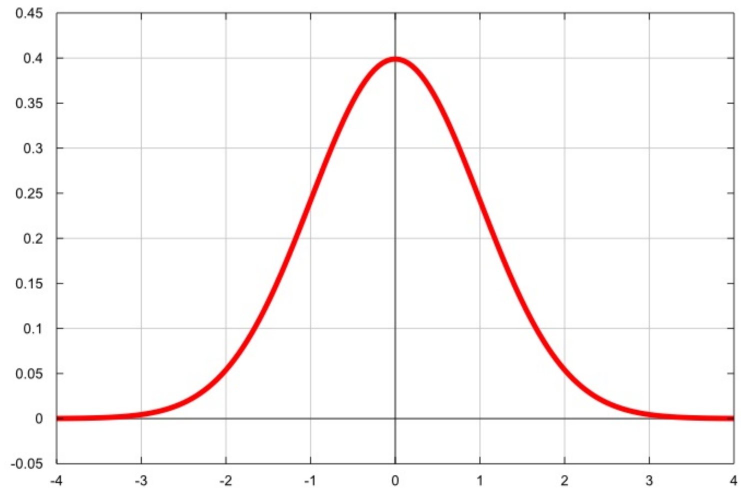

Figure 2: The density function

- How to think about this intuitively:
  - Note that where the cdf is changing rapidly (has a steep slope), the density is larger. Where it is changing slowly (has a flatter slope), the density is smaller.

- Having defined $f(y) \equiv \frac{dF(y)}{dy}$, we therefore can write

$$F(y) = \int_{-\infty}^{y} f(t)dt,$$

where $f(\cdot)$ is the pdf and t is a placeholder, the variable of integration. The pdf $f(\cdot)$ has the following properties:

$$\begin{aligned} f(y) &\geq 0 \forall y, \text{-}\infty < y < \infty. \\ \int_{-\infty}^{\infty} f(y)dy &= 1. \end{aligned}$$

- Be sure to work through the examples in your book to get a sense of the math of cdf's and pdf's.

- Now what if we want to find the probability that the random variable Y falls in a certain interval, e.g. $P(a \leq Y \leq b)$? It is the area under the density function in this interval [draw]

$$\begin{aligned} P(a &< Y \leq b) = P(Y \leq b) - P(Y \leq a) \\ &= F(b) - F(a) \\ &= \int_a^b f(y)dy. \end{aligned} \tag{1}$$

- In the continuous RV case, note that $P(a < Y < b) = P(a < Y \leq b) = P(a \leq Y < b) = P(a \leq Y \leq b)$. Why?

- Note that this is not necessarily the case for discrete RVs, because for those kinds of RVs, $P(y \leq a)$ is not necessarily equal to $P(y < a)$.

0.2 Expected values for continuous random variables

- Remember that in the case of discrete RV, we wrote

$$E(Y) \equiv \sum_y yp(y) \text{ and } E[g(Y)] = \sum_y g(y)p(y)$$

The math of expectations for continuous RVs is very similar:

$$\begin{aligned} E(Y) &\equiv \int_{-\infty}^{\infty} yf(y)dy \\ E[g(Y)] &= \int_{-\infty}^{\infty} g(y)f(y)dy \end{aligned}$$

- Furthermore:

$$\begin{aligned} VAR(Y) &\equiv \int_{-\infty}^{\infty} (y-\mu)^2 f(y)dy \\ &= E(Y^2) - \mu^2. \text{ [Exercise.]} \end{aligned}$$

- Reassuringly, this is the same as in the discrete case, as we showed last time in class.

0.3 Theoretical models of continuous probability distributions

- We'll look at two continuous RVs in detail:
  - The **Uniform**
  - The **Normal**
- And three distributions related to the Normal that we will use constantly in statistical tests:
  - * the **Chi-squared** (χ^2) distribution
    * the **t-distribution**
    * the **F idstribution**

0.4 The Uniform distribution

- Consider a random variable that can take on any value in an interval between two values, and these values are equiprobable.
  - e.g., the date of the election called by a prime minister who must call an election at some point in her five-year term.
  - the length of a student essay (in words) that is required to be between 1,000 and 2,000 words.
  - the length of a Tweet, which must be between 1 and 148 characters.
- (Hold off for a moment on scrutinizing whether these values are really all equiprobable; we'll get to that in a minute.)
- (Let's also hold off for a moment at balking at the fact that these are actually examples of discrete random variables, rather than continuous RVs.)
- We can represent the density function of such an RV like this (change a and b to the thetas):
- This flat density function represents the fact that all values in the interval are equiprobable.

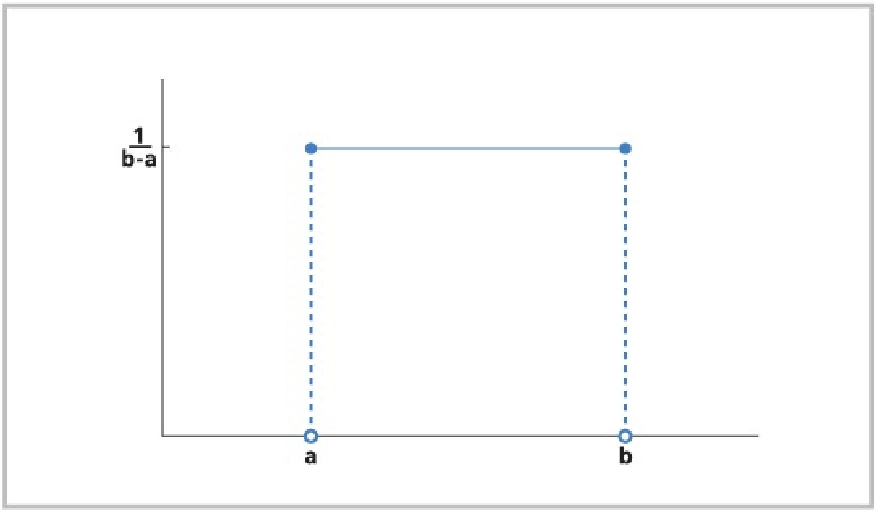

Figure 3: Uniform

- This leads to a pdf that looks like this:

$$f(y) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, & \theta_1 \leq y \leq \theta_2 \\ 0, & \text{elsewhere.} \end{cases}$$

- How do we get to the pdf? Simple geometry.
  - We know the rectangular area under the distribution function must equal 1.
  - We know the length of the base of the rectangle: it's $\theta_2 - \theta_1$.
  - Therefore its height h must solve $(\theta_2 - \theta_1) h = 1$ and thus $h = \frac{1}{\theta_2 - \theta_1}$.
- Note that when we "stretch out" the interval of a Uniform RV, its density gets smaller and smaller. Can you explain the intuition for this?
- Let's derive the cdf of the Uniform:

$$\begin{aligned} F(y) &= \int_{-\infty}^{y} f(t)dt \\ &= \int_{\theta_1}^{y} \frac{1}{\theta_2 - \theta_1} dt \\ &= \left. \frac{t}{\theta_2 - \theta_1} \right|_{\theta_1}^{y} = \frac{y - \theta_1}{\theta_2 - \theta_1} \end{aligned}$$

- Now let's show that $\mu = E(Y) = \frac{\theta_1 - \theta_2}{2}$.

$$
\begin{aligned}
E(Y) &\equiv \int y f(y)\, dy \\
&= \int_{\theta_1}^{\theta_2} y \frac{1}{\theta_2 - \theta_1} dy \\
&= \frac{1}{\theta_2 - \theta_1} \left. \frac{y^2}{2} \right|_{\theta_1}^{\theta_2} = \frac{(\theta_2)^2 - (\theta_1)^2}{2(\theta_2 - \theta_1)} \\
&= \frac{\theta_1 + \theta_2}{2}.
\end{aligned}
$$

- Of course, this is quite intuitive. The expected value of a Uniform RV should be at the point that divides the interval in half!

- A final result [proof will be on your HW]:

$$\sigma^2 = VAR(Y) = \frac{(\theta_2 - \theta_1)^2}{12}$$

- OK, now let's deal with the quibbles. When we think about the examples I mentioned earlier, we know enough about them that they are not properly modeled as having all their values equiprobable. [Go through examples.]

- However, what if we began examining a social process knowing nothing about it?
  - Or more to the point, we wanted to be completely *agnostic* about how the values might be distributed?
  - Assuming that the process follows a Uniform distribution is a good way to do this. And in fact, that's exactly what we do in a lot of statistical modeling. We also do this in a lot of formal modeling when we want to be agnostic about the beliefs an actor might have about the value of something. In both cases, we call this ″Uniform prior beliefs,″ or just a ″Uniform prior.″

0.5 The Normal distribution

- Many empirical distributions are closely approximated by a distribution that is:

- symmetric
- has positive (non-zero) probability for all possible values of y
- and is "bell shaped": specifically it has inflection points at one standard deviation away from its mean.

- In fact, it can be shown that:
  - in repeated random samples from a population
  - the means of these samples
  - are distributed around the population mean, μ, as described by a density function that we can actually write down.
  - The proof of this phenomenon is called the Central Limit Theorem, and this density function is defined as the Normal density function.
- We won't be proving the Central Limit Theorem. In a few lectures, I will instead be giving you empirical evidence for its existence.
- So unlike the other PMFs and PDFs we've looked at so far, we won't derive this function. It requires learning a lot of mathematics that I don't think you'll find particularly relevant.
- Instead, you will take it on faith that a RV Y has a Normal probability distribution iff, for $\sigma > 0$ and $-\infty < y < \infty$, the density function of Y is

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(y-\mu)^2}{2\sigma^2}}, \text{ for } -\infty < y < \infty.$$

- Draw on board, noting μ on x-axis (axis is labeled 'y'), y-axis labeled '$f(y)$', height of curve labeled $f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(y-\mu)^2}{2\sigma^2}}$.
- Note that the Normal has two parameters: μ and σ.
- As usual, if empirical values in population are distributed Normal, then

$$E(Y) = \mu \text{ and } VAR(Y) = \sigma^2.$$

- So based on what we learned earlier, what is $P(a \leq Y \leq b)$ if Y is distributed Normal?

$$P(a \leq Y \leq b) = \int_a^b f(y)dy = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(y-\mu)^2}{2\sigma^2}} dy.$$

- [Shade in area of graph.]

- A closed-form expression for this interval does not exist. Evaluating requires numerical methods. Commands are available in R (you'll learn them).

- Typically (and you'll recall this from any statistics class!), we **standardize** a Normally distributed variable so that our units are measured in terms of standard deviations. That is we transform that normal RV Y into the standard normal RV Z :

$$Z \equiv \frac{Y - \mu}{\sigma}.$$

- The RV Z is itself distributed Normal with mean zero and standard deviation one. Generally wecan standardize any Normal variable without losing any relevant information. The original distribution is now expressed in units of the standard deviation of the original Normal RV.

- If we standardize a Normal, its pdf becomes simpler:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

- We use the pdf and cdf of the standard Normal so much that we have special symbols for them:

  - We write $\phi(z)$ ("or little phi of z") to denote the pdf of the standard Normal evaluated at $Z = z$
  - and $\Phi(z) = \Pr(Z \leq z)$, (or "big phi of z") to denote the cdf of the standard Normal evaluated at $Z = z$.

0.6 Three distributions related to the standard Normal

- – There are three distributions related to the Normal that we will use constantly in statistical tests. It's not much use to go into detail about them here; I just want you to be aware that they all derived from the Normal. We will encounter them again soon.
    * the **Chi-squared** (χ^2) distribution [where Y is the sum of the squares of a series of standard Normal RVs]
    * the **t-distribution** [where Y is the ratio of a standard Normal RV / the square root of a chi-squared RV]
    * the **F-distribution** [where Y is the ratio of two chi-squared RVs]