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

Lecture 2

0.1 The Normal distribution

- – One more thing: the Normal distribution.
    - * Bell shaped.
    - * Unimodal, symmetric.
    - * Many variables, empirically, are distributed in a way that approximates the Normal distribution. E.g.: height.
    - * When this is the case, we can use a rule of thumb called Tchebysheff's Theorem to describe the distribution:
        - · the interval $\overline{y} \pm s$ contains about 68% of the observations
        - · the interval $\overline{y} \pm 2s$ contains about 95% of the observations; and
        - · the interval $\overline{y} \pm 3s$ contains about all of the observations.
    - * Example: the distribution of the height of American females approximates the Normal distribution. Its mean, $\overline{y}$, is 63.5 and its s.d., $s = 3$. Your cousin's best friend's new roommate is a woman who is 69.5 inches tall. She is taller than what proportion of females in the American population? {draw on board}
    - * Because the distribution of the mean of independently drawn observations from any population tends toward the Normal when the sample size gets large, we think a lot about the Normal. This fact also creates lots of empirical Normal distributions. More to come.
    - * However, do NOT make the mistake of assuming all distributions are Normal. Not e.g.: internet usage. Not e.g.: number of children.
  - – Inevitably, we find ourselves talking about how means and dispersion change over time and/or comparing groups. This begins to invite multivariate statistics.
  - – But for now, that's it for descriptive statistics with regard to one variable. Pretty simple

stuff, really. But social scientists spend a lot of time discussing and displaying descriptive statistics. Pay attention to talks in the department; you'll see.

  - Note what we haven't done: talked about samples versus populations. We've taken the data as given, and have resisted making inferences to a population.
  - That's our next big step. But to get there, we need to ascend through the basics of probability theory.

0.2 Probability theory: the logic

- We are all familiar with the process of moving from populations to samples. I have a 52-card deck that contains 13 spades. What's the probability of drawing a (sample) spade at random from a perfectly shuffled deck?
  - Obviously, it's 1/4.
  - We know this because we know the distribution of spades in the "population," and we have been precise about the sampling process.
- Well, we're now going to make the trip in reverse.
  - Now, we know the sample. We know its central tendency and its dispersion.
  - We make precise (and justified) assumptions about the sampling process.
  - These allow us to make very good guesses about the population's central tendency and dispersion using the sample's central tendency and dispersion.
  - We need the tools of the population-to-sample process to understand the sample-to-population process.
  - So let's learn them!

0.3 A probabilistic model for an experiment

- In probability theory, we use the term **experiment** to refer to *the process by which an observation is made*. An **observation** is a quantity of interest:
  - the price of a stock

- the number of experimental subjects who choose "A" instead of "B"
- the proportion of Pew survey respondents who approve of Biden's job as president
- the proportion of Gallup ""
- in this usage, experiments don't just happen in labs, but the term is helpful because it gets us thinking about social processes in the experimental context.

- Experiments have one or more outcomes called **events**.
  - Experiment: gubernatorial election in an imaginary country that has 101 voters; majority rule, everyone votes for Candidate A or B.
  - Possible events (list): a. Candidate A wins. b. Candidate B wins. c. Candidate A wins with 76 votes. d. Candidate A wins with 56 votes. e. Candidate A wins in a landslide ($\equiv$ 67 voters or greater). How about some others? (In this example, we do not care who voted for whom; we care only about the aggregate result.)
  - Note that the first event can be decomposed into [how many] other events?
  - How about the last event?
  - a, b and e are **compound** events. By contrast, c. and d. are **simple** events. Let's formalize this. Because certain concepts from set theory will be helpful for expressing relationships amoung events, we will also associate a distinct point–a **sample point–** with each simple event. We use E_i to refer to the simple event or sample point i. Draw on board: circle S with dots denoting Es.
  - The **sample space** S associated with an experiment is the set consisting of all possible sample points. How many simple events are in S for the hypothetical election ? (102; list on board with a table)
  - In set notation, we write $S = \{E_1, E_2, ...E_{102}\}$, where E_i denotes candidate A's number of votes plus one.
  - This sample space consists of a countable (finite) number of sample points. By definition, it is a **discrete sample space.**
  - Simple events / sample points are mutually exclusive.

- By contrast, compound events are sets of sample points. The compound event A, "A wins by landslide" occurs if and only if one of the events $E_{68}...E_{102}$ occurs. Thus $A = \{E_{68}...E_{102}\}$.
  - A simple event E_i is included in compound event A if and only if A occurs whenever E_i occurs.
  - Now we can be more specific about what an **event** is: it is a collection of sample points (a subset of S).
  - We can also consider S an event in itself. It occurs whenever the experiment occurs.

- We're now ready to construct a probabilistic model for an experiment with a discrete sample space. We do so by *assigning a number,* $P(A)$, *to each event* A in S. Of course, we can't do this willy-nilly. We do so that *three axioms of probability* hold:
  - Axiom 1: $P(A) \geq 0$.
  - Axiom 2: $P(S) = 1$. S occurs every time the experiment is performed.
  - Axiom 3: For any sequence of pairwise mutually exclusive events $A_1, A_2, A_3, ... A_n$, it must be the case that P($A_1 \cup A_2 \cup A_3 ... \cup A_n) = \sum_{i=1}^{n} P(A_i)$. That is, the relative frequency of the union of these mutually exclusive events (that is, an event made up of these events) equals the sum of their relative frequencies.
  - When we assign numbers $P(A)$ to events in this way, the numbers are defined as the **probabilities of A.**
  - There are deep-thought ideas about what probability actually means. But the most intuitive way to think about probability of an event A is the proportion of the time event A would occur if we were to repeat the experiment, in the exact same way, infinitely many times.

0.4 Calculating the Probability of an Event of Interest: overview

- The three axioms of probability put bounds on the probabilities we assign to events of interest. Now we need to talk about how to assign probabilities in a rigorous way. Your text discusses two such methods; often either can be used to assign probabilities in a given situation. The first is the *sample point* method. We won't be discussing this here in class; but it incorporates many of the techniques you learned regarding probability in Math Camp. The second is the *event composition* method. Because this method illustrates many of the laws of probability that you will use throughout your quantitative analysis classes, we will spend a bit of time reviewing this method.

0.5 The event-composition method

- In short, the *event-composition* method proceeds by decomposing and composing event A into *unions* and *intersections* of events with conveniently calculated probabilities. Four tools help you do this:
  - * the definitions of conditional probability and independence
    * multiplicative and additive laws
    * the probability of an event and its complement
    * the law of total probability and Bayes' Rule
- Before proceeding, recall that
  - $A \cap B$ "A intersection B" is the compound event where *both* A and B happen. (AND)
  - $A \cup B$ "A union B" is the compound event where either A or B happens, or both. (OR)
  - [Diagram on board]
- We require definitions of [put these on left-hand board for easy referral]:
  - **conditional probability**: $P(A|B) = \frac{P(A \cap B)}{P(B)}$.
    * Note that this is a *definition* that corresponds with notions of probability. As such it cannot be proven, but it can been shown that the definition corresponds with commonsense notions of probability. It has the intuitive meaning that $P(A|B)$ is the probability that both A and B occur given as a proportion of the probability that B occurs. The same is true for the following definition...

- **independence**: For two events A and B, if *any one* of the following conditions holds:

$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$
$$P(A \cap B) = P(A)P(B)$$

, then A and B are said to be *independent* events. If *none* of these three conditions holds, then the events are said to be *dependent*.

- We also require two *laws.* The first is about the *intersection* of two or more events; the second is about the *union* of two or more events.

  - **multiplicative law** (*intersection*):

    * $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$ OR
    * $P(A \cap B) = P(A)P(B)$ if $A,\ B$ are independent events.
    * Note this law follows directly from our definition of conditional probability.
    * can be extended to 3 events:
    *

$$\begin{aligned} P(A \cap B \cap C) &= P((A \cap B) \cap C) \text{ [associative property]} \\ &= P(A \cap B)P(C|A \cap B)\text{.[applying the multiplicative law once]} \quad (1) \\ &= P(A)P(B|A)P(C|A \cap B)\text{.[applying the law again]} \end{aligned}$$

    * can be extended to k events: $P(A_1 \cap A_2 \cap ... \cap A_k) = P(A_1)P(A_2|A_1)...P(A_k|A_1 \cap A_2 \cap ... \cap A_{k-1})$.

  - **additive law** (*union*):

    * $P(A \cup B) = P(A) + P(B) - P(A \cap B)$,
    * where if A, B mutually exclusive, $P(A \cap B) = 0$.
    * why? Look at Venn Diagram. We're avoiding "double-counting" the intersection of events A and B.:ftbpF3.8873in2.2286in0ptFigure

– A proof:

$$\begin{aligned} A\cup B &= A\cup(\tilde{}A\cap B)\text{, and } A,(\tilde{}A\cap B) \text{ are mutually exclusive events. Also,}\\ B &= (\tilde{}A\cap B)\cup(A\cap B)\text{, and } (\tilde{}A\cap B),(A\cap B) \text{ are mutually exclusive events. Then}\\ P(A\cup B) &= P(A)+P(\tilde{}A\cap B) \text{ and } P(B)=P(\tilde{}A\cap B)+P(A\cap B) \text{ [by Axiom 3] and thus}\\ P(A\cup B) &= P(A)+P(B)-P(A\cap B).\blacksquare \end{aligned}$$

- it's helpful to remember that the probability of **complementary events** is such that $P(A) = 1 - P(\sim A)$. Proof is easy:

$$\begin{aligned} S &= A\cup\tilde{}A \text{ (since } A \text{ is an event in sample space } S\text{)}\\ P(S) &= P(A)+P(\tilde{}A) \text{ (by Axiom 3, since } A,\tilde{}A \text{ mutually exclusive)}\\ 1 &= P(A)+P(\tilde{}A) \text{ (by Axiom 2)}\\ P(A) &= 1-P(\sim A).\blacksquare \end{aligned}$$

- So let's run through an example using the event composition method.
  – I randomly assign a group of 16 students into 3 teams of 6, 5 and 5 students. 11 of the students are male.
  – What is $P(A)$ =the probability that the team of six students (call this "Team 1") is entirely male?
  – First, note that the event "all members of Team 1 are male" is equivalent to the event "the first six assigned students are all male."
    * Let's decompose this event into simpler events A through F:
    * A: the first student picked is male.
    * B: the second student picked is male.
    * F: the sixth student picked is male.
  – The event of interest occurs if and only if all events A through F occur. It is thus the intersection of all these events: $A\cap B\cap C\cap D\cap E\cap F$. We therefore want to find $P(A\cap B\cap C\cap D\cap E\cap F)$.

- What should we do?
  - Re-write using multiplicative law:

$$
\begin{aligned}
P(A_1 \cap A_2 \cap ... \cap A_k) &= P(A_1)P(A_2|A_1)...P(A_k|A_1 \cap A_2 \cap ... \cap A_{k-1}), \text{ so} \\
P(A_1 \cap A_2 \cap ... \cap A_6) &= P(A_1)P(A_2|A_1)...P(A_k|A_1 \cap A_2 \cap ... \cap A_5)
\end{aligned}
$$

- What is the probability that the first student picked is male? $\frac{11}{16}$.

$$P(A_1 \cap A_2 \cap ... \cap A_6) = \frac{11}{16}P(A_2|A_1)...P(A_k|A_1 \cap A_2 \cap ... \cap A_5)$$

- Given A1, what's the probability that the second student picked is male? $\frac{10}{15}$.

$$P(A_1 \cap A_2 \cap ... \cap A_6) = \frac{11}{16} \cdot \frac{10}{15} \cdot ...P(A_6|A_1 \cap A_2 \cap ... \cap A_5)$$

- And so forth:

$$P(A_1 \cap A_2 \cap ... \cap A_6) = \frac{11}{16} \cdot \frac{10}{15} \cdot \frac{9}{14} \cdot \frac{8}{13} \cdot \frac{7}{12} \cdot \frac{6}{11} = \frac{3}{52}$$

- In many instances (including exams), we are happy to write this in factor notation. How might we do this?
  - Note that the product of the numerators is $11 \cdot 10 \cdot ... \cdot 2 \cdot 1$ with the final five multiplicands "shaved off."
  - This is just $\frac{11 \cdot 10 \cdot ... \cdot 2 \cdot 1}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}$, or $\frac{11!}{5!}$.
  - Similarly, the denominator can be written $\frac{10!}{16!}$.
  - So another way to write this is:

$$P(A_1 \cap A_2 \cap ... \cap A_6) = \frac{11!}{5!} \cdot \frac{10!}{16!}$$

0.6 The law of total probability and Bayes' Rule

- finally, the *law of total probability* and *Bayes' Rule* can be helpful in the event-composition approach to assigning probabilities to events.

- recall that we are working with a discrete sample space S, composed of simple events.
  - we can of course also view S as a union of k mutually exclusive subsets:
    * $S = B_1 \cup B_2 \cup ... \cup B_k$;
    * $B_i \cap B_j = \varnothing, \forall i \neq j$.
    * A collection of sets $\{B_1, B_2 ... B_k\}$ such that (1) their union is equivalent to S and (2) are themselves mutually exclusive is said to be a **partition** of S.
    * If A is a subset of S, it may be **decomposed** as the union of its intersections with each of the partitions of S as follows: $A = (A \cap B_1) \cup (A \cap B_2) \cup ...(A \cap B_k)$. Draw Figure 2.12 {p. 71} on board as illustration.
    * ftbpF4.782in3.3931in0ptintersection.tif
  - The partitioning and decomposing notions yield the following:
    * If the collection $\{B_1, B_2 ... B_k\}$ is a partition of S such that $P(B_i) > 0$ for $i = 1, 2, ...k$, then the **Law of Total Probability** states:

$$P(A) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$$

    * Proof {brief}:
      - If $i \neq j$, the intersections $A \cap B_i$ and $A \cap B_j$ do not overlap. Put formally, their intersection is the empty set, because

$$\begin{aligned} (A \cap B_i) \cap (A \cap B_j) &= A \cap (B_i \cap B_j) \quad \text{[distributive law]} \\ &= A \cap \oslash \\ &= \oslash. \end{aligned}$$

        The events $A \cap B_i$ and $A \cap B_j$ are thus mutually exclusive.

· Because A is the union of all these mutually exclusive events, we can write

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + ... + P(A \cap B_k) \quad \text{[additive law]} \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + ...P(A|B_k)P(B_k) \text{ [muliplicative law]} \\ &= \sum_{i=1}^{k} P(A|B_i)P(B_i) \quad \blacksquare. \end{aligned}$$

· Why do we care? Because there are lots of instances where it's easier to calculate $P(A|B_i)$ than $P(A)$. In other cases, we intrinsically care about $P(A|B_i)$. I'll show you this in a minute, but first let's derive one additional important result.

* If the collection $\{B_1, B_2...B_k\}$ is a partition of S such that $P(B_i) > 0$ for $i = 1, 2, ...k$, then

$$\begin{aligned} P(B_j|A) &= \frac{P(A \cap B_j)}{P(A)} \text{ [definition of conditional probability]} \\ &= \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{k} P(A|B_i)P(B_i)} \text{ [definition again (num), law of total probability (denom)]} \end{aligned}$$

* This is **Bayes' Rule**.
* Let's write this in a simple case, where there are only two partitions in the sample space:

itbphF4.0938in2.4641in0inFigure

* We then have:

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)}$$

* or more simply (writing B_1 as simply B):

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\tilde{}B)P(\tilde{}B)}$$

* So classic example (2.124, page 73):

$$\begin{aligned} P(Dem|Favor) &= \frac{P(Favor|Dem)P(Dem)}{P(Favor|Dem)P(Dem)+P(Favor|REP)P(REP)} \\ &= \frac{(.7)\,(.6)}{(.7)\,(.6)+(.3)\,(.4)} \\ &= 7/9 \\ &= .\bar{7} \end{aligned}$$

* Note our process of determining the probability $P(Dem|Favor)$. We used the definition of conditional probability (which undergirds Bayes' Rule) to decompose the set (DEM $\cap$ FAVOR).

– Another classic example (check if there's time) Ex. 2.125, p. 73.

* Begin by having students write $P\,(Disease|Pos)$ according to the definition of Bayes' Rule.
* What information do we need to know?
  · $P\,(Pos|Disease) = .90$
  · $P\,(\tilde{}Pos|\tilde{}Disease) = .90$
  · $P\,(Disease) = .01$
* And so:

$$\begin{aligned} P\,(Disease|Pos) &= \frac{P(Pos|Disease)P(Disease)}{P(Pos|Disease)P(Disease)+P(Pos|\tilde{}Disease)P(\tilde{}Disease)} \\ &= \frac{(.9)(.01)}{(.9)(.01)+(.10)(.99)} \\ &= \frac{.009}{.009+.099} = \frac{1}{12} \end{aligned}$$

* Put this in perspective:
  · What is the probability of a false positive? $P\,(Pos|\tilde{}Disease) = .10 = \frac{1}{10}$
  · What is the probability of a false negative? $P\,(\tilde{}Pos|Disease) = .10 = \frac{1}{10}$
* What affects the reliability of this test?