Probability

Paul Schrimpf

January 23, 2018

Contents

References

- Wooldridge (2013) appendix B
- Stock and Watson (2009) chapter 2
- Linton (2017) chapters 1-5
- Abbring (2001) sections 2.1-2.3
- Diez, Barr, and Cetinkaya-Rundel (2012) chapter 2
- Grinstead and Snell (2003) chapters 1-7

These lectures are mostly based on Wooldridge, so that will generally be the best reference to read. Stock and Watson (2009) chapter 2 is more detailed and assumes less prior knowledge than Wooldridge (2013).

The other references are freely available alternatives that cover similar material. Linton (2017) is an excellent, recently published book available online through UBC library. Linton (2017) is a bit more technical mathematically than what we will cover in this course, but I think it would still be useful. If you read Linton (2017), you can safely ignore the parts about measure theory and Sigma-algebras, and the use of linear algebra later. Abbring (2001) is somewhat similar to Linton (2017). More introductory references are Diez, Barr, and Cetinkaya-Rundel (2012) and Grinstead and Snell (2003).

Probability

- Purpose: system for quantifying chance and making predictions about future events
- Interpretations:
  - Relative frequency of an event in many repeated trials
  - Subjective assessment of the likelihood of an event

---

1 Definitions

The properties of probability are fairly intuitive, and you can work with probabilities without worrying too much about how probability is formally defined. However, if we want to be mathematically rigorous, we should carefully define what we mean by probability. This slide formally defines probability. Do not worry if it seems very abstract.

Basic definitions

- *Random experiment*: procedure that has well-defined set of outcomes and "could" be infinitely repeated
- *Sample space*: set of possible outcomes of an experiment, $S = \{a_1, a_2, ..., a_J\}$
- *Event*: any subset of sample space, $A \subseteq S$
- *Probability*: function from all subsets of the sample space, S, to $[0, 1]$ such that
  1. $P(S) = 1$
  2. $1 \geq P(A) \geq 0$ for all $A \subseteq S$
  3. If $A_1, A_2, ...$ are disjoint events then $P(A_1 \cup A_2 \cup ...) = P(A_1) + P(A_2) + ...$
- *Random variable*: function from S to a number

An example of a random experiment is rolling a dice. In this case the sample space is $S = \{1, 2, 3, 4, 5, 6\}$. Some events are: get a $1 = \{1\}$, get more than 4, $\{5, 6\}$, etc.

A random variable is a numeric representation of a random experiment. For example, our experiment could be flipping a coin. Then $S = \{\text{heads}, \text{tails}\}$. A function assigns a number to each element of S, so one example of a random variable is $X = \begin{cases} 1 & \text{if heads} \\ 0 & \text{if tails} \end{cases}$. Another is $Y = \begin{cases} 50 & \text{if heads} \\ -257 & \text{if tails} \end{cases}$. A third example is $Z = \begin{cases} 1 & \text{if heads} \\ 1 & \text{if tails} \end{cases}$.

Sets

Sets appear throughout mathematics. A **set** is any well-defined collection of objects. We will usually denote sets by capital letters. The things in a set are generally called **elements** of the set. For a sample space, the elements are the possible outcomes of an experiment. There is some notation related to sets that is useful. We list the elements of set inside braces, so for example the set of the colors red and green, could be written as $\{\text{red}, \text{green}\}$.

We can combine two sets to create a new one by taking the union or intersection of the sets. The **union** of set A and set B is written $A \cup B$ and is the set of everything in either A or B or both. The **intersection** of set A and B is written $A \cap B$ and is the set of everything in both A and B. Two sets are **disjoint** if their intersection is empty.

The set with nothing in it is called the **empty set** and is denoted by $\emptyset$.

If every element of A is also an element of B, then we say that A is a subset of B and denote this by $A \subseteq B$. If there is also something in B that is not in A (i.e. A and B are not the same), then we say that A is a strict subset of B and write it as $A \subset B$.

2 Properties

From the definition of probability on the previous slide, we can derive some intuitive properties. Each of these properties are fairly obvious. The point of showing that they follow from the previous definition is to verify that the formal definition agrees with our informal idea of probability. If they didn't agree, it would tell us that there is something wrong with our formal definition.

Properties of probability

1. $P(\emptyset) = 0$
2. $P(A^c) = 1 - P(A)$
3. $A \subseteq B$ implies $P(A) \leq P(B)$
4. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Conditional probability is very important for making decisions and interpreting our observations of the world. Often, we would like to know about something that we do not directly observe, but we do observe some related information. For example, to forecast GDP next quarter, which is something we can not observe today, we can use current information about GDP, inflation, consumer confidence, etc. To make a forecast, we do not care so much about the unconditional probability that GDP tomorrow is high, but about the probability that GDP tomorrow is high given what we observe today.

Conditional probability

- *Conditional probability*: $\text{P}(A|B) = \text{P}(A \cap B)/\text{P}(B)$
  - Satisfies axioms of unconditional probability (and so also has properties on previous slide)
- *Bayes' Rule* $\text{P}(A|B) = \frac{\text{P}(B|A)\text{P}(A)}{\text{P}(B)}$
- A is *independent* of B iff $\text{P}(A \cap B) = \text{P}(A)\text{P}(B)$, denote as $A \perp\!\!\!\perp B$
  - $A \perp\!\!\!\perp B$ implies $\text{P}(A|B) = \text{P}(A)$ and $\text{P}(B|A) = \text{P}(B)$.

Proof of Bayes' rule: definition of conditional probability $\text{P}(A \cap B) = \text{P}(A|B)\text{P}(B)$, and switching role of A and B, $= \text{P}(B|A)\text{P}(A)$. Rearrange to get conclusion.

Bayes' Rule allows us to switch between the $\text{P}(A|B)$ and $\text{P}(B|A)$. It is very useful. One common use is for interpreting the results of screening tests. For example, suppose there is a test for some disease. From clinical trials, we would likely know the probability of correctly testing positive conditional on having the disease, $\text{P}(+|\text{disease})$, and the probability of falsely testing positive, $\text{P}(+|\text{healthy})$. Suppose the unconditional probability of having the disease if $\text{P}(\text{disease})$. From this, we can use Bayes' rule to calculate the probability of having the disease conditional on testing positive as:

$$\begin{aligned}\text{P}(\text{disease}|+) =& \frac{\text{P}(+|\text{disease})\text{P}(\text{disease})}{\text{P}(+)} \\ =& \frac{\text{P}(+|\text{disease})\text{P}(\text{disease})}{\text{P}(+|\text{disease})\text{P}(\text{disease}) + \text{P}(+|\text{healthy})(1 - \text{P}(\text{disease}))}.\end{aligned}$$

When A and B are independent, then knowing whether B happens does not give you any information about the probability of A occurring.

3 Random variables

3.1 Discrete

Discrete random variables

- Recall: a *random variable* is a function from S to a number
- A random variable is *discrete* if it can only take countably many different values
- E.g. $X \in \{x_1, ..., x_m\}$
- *Probability mass function (PMF)* of X, $p_i = \text{P}(X = x_i)$ gives the probability that X equals each of its possible values
- *Cumulative distribution function (CDF)*: $F(x) = \text{P}(X \leq x) = \sum_{i=1}^{m} p_i 1\{x_i \leq x\}$

3.2 Continuous

Continuous random variables

- A random variable is *continuous* if it takes on any single real value with zero probability
- Set of possible values is uncountably infinite (e.g. real line or line segment)
- CDF is continuous and differentiable
- *Probability density function (PDF)*: the derivative of the CDF

$$f(x) = \frac{dF}{dx}(x) \text{ and } F(x) = \int_{-\infty}^{x} f(t)dt$$

3.3 Bivariate distributions

Bivariate distributions

- Discrete:
  - Joint PMF $f_{X,Y}(x,y) = \mathrm{P}(X = x, Y = y)$
  - Marginal PMF $f_X(x) = \mathrm{P}(X = x) = \sum_{y \in \text{all values of } Y} \mathrm{P}(X = x, Y = y)$
  - Conditional PMF $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
- Continuous:
  - joint CDF $F_{X,Y}(x,y) = \mathrm{P}(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,t)dudt$
  - joint PDF $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y)$
  - Marginal CDF for X, $F_X(x) = \mathrm{P}(X \leq x) = F_{X,Y}(x, \infty)$
  - Marginal PDF for X, $f_X(x) = \frac{dF_X}{dx}(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$
  - Conditional PDF $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$

Independence

- Random variables X, Y are independent if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ for all x, y
- $X \perp\!\!\!\perp Y$ implies $f_{X|Y}(x|y) = f_X(x)$

3.4 Expectations

We will use expectations a lot, so it's important to understand them and know their properties. The fact that the expectation is linear and the law of iterated expectations will be especially useful.

Expectation

- The *expectation* of X is

$$\mathrm{E}[X] = \begin{cases} \sum_{i=1}^{m} x_i f_X(x_i) & \text{if } X \text{ discrete} \\ \int_{-\infty}^{\infty} x f_X(x) dx & \text{if } X \text{ continuous} \end{cases}$$

  - Is a constant
  - a.k.a. expected value, population average, population mean, first moment

- Expectation is linear:

$$\mathrm{E}[aX + bY] = a\mathrm{E}[X] + b\mathrm{E}[Y]$$

- Expectation of function g:

$$\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

Variance and other moments

- *kth moment* of X: $\mathrm{E}[X^k]$
- *kth central moment* of X: $\mathrm{E}\left[(X - \mathrm{E}[X])^k\right]$
- *Variance* is the 2nd central moment

$$\begin{aligned} \mathrm{Var}(X) =& \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \\ =& \mathrm{E}[X^2] - \mathrm{E}[X]^2 \end{aligned}$$

- $\mathrm{Var}(a + bX) = b^2 \mathrm{Var}(X)$
- *Standard deviation* is $\sqrt{\mathrm{Var}(X)}$

Covariance

- *Covariance* of X and Y:

$$\begin{aligned}\text{Cov}(X,Y) =& \text{E}\left[(X-\text{E}[X])\,(Y-\text{E}[Y])\right]\\ =& \text{E}[XY]-\text{E}[X]\text{E}[Y]\\ =& \text{E}[(X-\text{E}[X])\,Y] = \text{E}[X\,(Y-\text{E}[Y])]\end{aligned}$$

- $\text{Cov}(X,X) = \text{Var}(X)$
- $\text{Cov}(a_1+b_1X+c_1Y, a_2+b_2X+c_2Y) = b_1b_2\text{Var}(X)+c_1c_2\text{Var}(Y)+(b_1c_2+c_1b_2)\text{Cov}(X,Y)$
- $\text{Var}(a+bX+cY) = b^2\text{Var}(X)+c^2\text{Var}(Y)+2bc\text{Cov}(X,Y)$
- $\text{Var}\left(\sum_{i=1}^N b_iX_i\right) = \sum_{i=1}^N\left(\sum_{j=1}^N b_ib_j\text{Cov}(X_i,X_j)\right)$

We will often need to calculate the variance of sums like in the last two bullet points. You should make sure you're comfortable with them. It's not too much algebra to derive them from the definition of variance. Let's do so now.

$$\begin{aligned}\text{Var}(a+bX+cY) =& \text{E}\left[(a+bX+cY-\text{E}[a+bX+cY])^2\right] && \text{definition of variance}\\ =& \text{E}\left[(a+bX+cY-(a+b\text{E}[X]+c\text{E}[Y]))^2\right] && \text{linearity of expectation}\\ =& \text{E}\left[(b(X-\text{E}[X])+c(Y-\text{E}[Y]))^2\right] && \text{rearranging terms}\\ =& \text{E}\left[b^2(X-\text{E}[X])^2+c^2(Y-\text{E}[Y])^2+2bc(X-\text{E}[X])(Y-\text{E}[Y])\right] && \text{expanding the square}\\ =& b^2\text{E}[(X-\text{E}[X])^2]+c^2\text{E}[(Y-\text{E}[Y])^2]+2bc\text{E}\left[(X-\text{E}[X])(Y-\text{E}[Y])\right] && \text{linearity of expectation}\\ =& b^2\text{Var}(X)+c^2\text{Var}(Y)+2bc\text{Cov}(X,Y) && \text{defination of variance}\end{aligned}$$

The last bullet with the summation follows from similar logic.

Correlation

- *Correlation* of X and Y is $\text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$
- *Cauchy-Schwartz inequality*: $|\text{Cov}(X,Y)| \leq \sqrt{\text{Var}(X)\text{Var}(Y)}$
  - Implies $-1 \leq \text{Corr}(X,Y) \leq 1$
- $\text{Corr}(X,Y) = \pm 1$ iff $Y = a+bX$

3.5 Conditional expectation

Conditional expectation

- *Conditional expectation* of Y given $X = x$:

$$\mathrm{E}[Y|X = x] = \begin{cases} \sum_{i=1}^{m} y_i f_{Y|X}(y_i|x) & \text{if discrete} \\ \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy & \text{if continuous} \end{cases}$$

- $\mathrm{E}[Y|X]$ is a function of X
- Has all properties of unconditional expectation
- Properties:
  - $\mathrm{E}[g_1(X) + g_2(X)Y|X] = g_1(X) + g_2(X)\mathrm{E}[Y|X]$
  - *Law of iterated expectations*: $\mathrm{E}_X\left[\mathrm{E}_{Y|X}[Y|X]\right] = \mathrm{E}[Y]$
  - $\mathrm{E}\left[X\left(Y - \mathrm{E}[Y|X]\right)\right] = 0$

Conditional variance

- *Conditional variance*: $\mathrm{Var}(Y|X) = \mathrm{E}\left[\left(Y - \mathrm{E}[Y|X]\right)^2 |X\right]$
- Relation to variance:

$$\mathrm{Var}(Y) = \mathrm{E}_X\left[\mathrm{Var}(Y|X)\right] + \mathrm{Var}\left(\mathrm{E}[Y|X]\right)$$

  - Analysis of variance (ANOVA)
  - $\mathrm{E}_X\left[\mathrm{Var}(Y|X)\right]$ is within-X variance
  - $\mathrm{Var}\left(\mathrm{E}[Y|X]\right)$ is between-X variance

---

References

References

Abbring, Jaap. 2001. "An Introduction to Econometrics: Lecture notes." URL http://jabbring.home.xs4all.nl/courses/b44old/lect210.pdf.

Diez, David M, Christopher D Barr, and Mine Cetinkaya-Rundel. 2012. *OpenIntro Statistics*. OpenIntro. URL http://www.openintro.org/stat/textbook.php.

Grinstead, Charles M and J. Laurie Snell. 2003. *Introduction to Probability*. American Mathematical Society. URL http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html.

Linton, Oliver B. 2017. *Probability, Statistics and Econometrics*. Academic Press. URL http://gw2jh3xr2c.search.serialssolutions.com/?sid=sersol&SS_jc=TC0001868500&title=Probability%2C%20statistics%20and%20econometrics.

Stock, J.H. and M.W. Watson. 2009. *Introduction to Econometrics, 2/E*. Addison-Wesley.

Wooldridge, J.M. 2013. *Introductory econometrics: A modern approach*. South-Western.