Math Camp for Economists:

Basic Probability Theory

Justin C. Wiltshire

Department of Economics
University of Victoria
Summer 2023

Overview of this lesson

In this lesson we're going to build up a lot of the knowledge required for other concepts:

- Basic set theory
- Probability spaces (and associated relevant concepts, like σ-algebras)
- Random variables (discrete and continuous)
- Probability functions
- Statistical independence

Sets: some definitions and notation

- A "set" is a collection of distinct objects called "elements"
  - $\rightarrow$ E.g. $A = \{a, b, c, d\}$, where a, b, c and d are the elements of the set A
- Elements of sets can themselves be sets
- A set with no elements is the "empty set", expressed as $\{\}$ or $\emptyset$
- A set with one element is called a "singleton"
- A "finite set" is a set with a finite number of elements
- An "infinite set" is a set that is not a finite set

Subsets: some definitions and notation

- A set A is a "subset" of a set B, $A \subseteq B$, if every element of A is also in B
  - $\rightarrow$ Then B is a "superset" of A
  - $\rightarrow$ $\emptyset$ is a subset of all sets
- Two sets are "equal" when they contain exactly and only the same elements
  - $\rightarrow$ $A = B$ if $A \subseteq B$ and $B \subseteq A$
- A set A is a "proper subset" of a set B, $A \subset B$, when $A \subseteq B$ but $A \neq B$
- The "power set" of a set A, $\mathcal{P}(A)$, contains all subsets of A, including $\emptyset$ and A
- A subset of a set A, $B \subseteq A$, is "closed" under an operation of A if the operation on members of B yields a member of B
  - $\rightarrow$ E.g. the natural numbers $\mathbb{N} \subset \mathbb{R}$ are closed under addition
- A subset of Euclidean space (e.g. $\mathbb{R}^n$) is a "convex set" if, given any two points in the subset, the subset contains the whole line segment that joins them

Intersections, unions, and complements of sets

- The "intersection" of two sets A and B, $A \cap B$, is the set containing all the elements of A that are <u>also</u> in B
  - $\rightarrow$ $A \cap B = \{x : x \in A \text{ and } x \in B\}$
  - The set of elements in both "*A and B*"
  - A and B are "disjoint" if they have no elements in common, $A \cap B = \emptyset$
- The intersection of n sets, $A_1, A_2, ..., A_n$, is written $\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap ... \cap A_n$
- The "union" of two sets A and B, $A \cup B$, is the set of all elements in "A or B"
  - $\rightarrow$ $A \cup B = \{x : x \in A \text{ or } x \in B\}$
- The union of n sets, $A_1, A_2, ..., A_n$, is written $\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup ... \cup A_n$
- The "complement" of a set A, A^c (or $\sim A$), is the set of all elements not in A
  - $\rightarrow$ $A \cup A^c = U$, where U is the universe of sets (or "universal set")
    - · Don't get into whether U exists. Let it be the union of all relevant sets
  - $\rightarrow$ $A \cap A^c = \emptyset$

De Morgan's laws

- $(A \cup B)^c = A^c \cap B^c$
- $(A \cap B)^c = A^c \cup B^c$

σ-algebra

Say we want to "measure" the N-dimensional volume of the subsets of a set X

- Recall: the power set $\mathcal{P}(X)$ (or 2^{Ω}) contains the complete set of subsets of X

  $\rightarrow$ E.g. $X = \{a, b, c\}$. Then $\mathcal{P}(X) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, X\}$
- Any collection of subsets of $\mathcal{P}(X)$, $\Sigma \subseteq \mathcal{P}(X)$, is a "family of subsets" over X

σ-algebra cont'd.

- $\Sigma \subseteq \mathcal{P}(X)$ is called a "σ-algebra" on X iff:
  (1) $X \in \Sigma$ (i.e. Σ is a nonempty collection of subsets of X)
  (2) $A \in \Sigma \implies A^c \in \Sigma$ (i.e. Σ is closed under complementation)
  (3) $A_i \in \Sigma, i \in \mathbb{N} \implies \bigcup_{i=1}^{\infty} A_i \in \Sigma$ (i.e. Σ is closed under countable unions)

Don't let this intimidate you. It's largely just a useful definition for later.

σ-algebra cont'd.

- $\Sigma \subseteq \mathcal{P}(X)$ is called a "σ-algebra" on X iff:
  (1) $X \in \Sigma$ (i.e. Σ is a nonempty collection of subsets of X)
  (2) $A \in \Sigma \implies A^c \in \Sigma$ (i.e. Σ is closed under complementation)
  (3) $A_i \in \Sigma, i \in \mathbb{N} \implies \bigcup_{i=1}^{\infty} A_i \in \Sigma$ (i.e. Σ is closed under countable unions)

Don't let this intimidate you. It's largely just a useful definition for later.

A few more things to note:

- (1) and (2) $\implies \emptyset \in \Sigma$ (as does the definition of subsets)
- (1) – (3) + De Morgan's laws $\implies$ Σ is closed under countable intersections
- If $\Sigma \subseteq \mathcal{P}(X)$ is a σ-algebra, then:
  a) $A \in \Sigma$ is called a Σ-"measurable set"
  b) (X, Σ) is called a "measurable space"

σ-algebra examples

For a set X

- $\{\emptyset, X\}$ is the smallest σ-algebra (or "trivial σ-algebra") on X
- The power set $\mathcal{P}(X)$ is the largest σ-algebra" on X
- In practice all good σ-algebra, $\mathcal{F}$, lie between the extremes: $\{\emptyset, X\} \subset \mathcal{F} \subset \mathcal{P}(X)$
  - $\rightarrow$ Usually can't measure *all* possible subsets of X

Exercise:

(1) If $A \subset X$, what is the smallest σ-algebra containing A?

(2) If $X = \{1, 2, 3\}$, what is a non-trivial σ-algebra?

σ-algebra examples

For a set X

- $\{\emptyset, X\}$ is the smallest σ-algebra (or "trivial σ-algebra") on X
- The power set $\mathcal{P}(X)$ is the largest σ-algebra" on X
- In practice all good σ-algebra, $\mathcal{F}$, lie between the extremes: $\{\emptyset, X\} \subset \mathcal{F} \subset \mathcal{P}(X)$

  $\rightarrow$ Usually can't measure *all* possible subsets of X

Exercise:

(1) If $A \subset X$, what is the smallest σ-algebra containing A?

$\rightarrow$ $\{\emptyset, A, A^c, X\}$

(2) If $X = \{1, 2, 3\}$, what is a non-trivial σ-algebra?

$\rightarrow$ There are several. One is $\mathcal{F} = \{\emptyset, \{1\}, \{2, 3\}, \{1, 2, 3\}\}$

Some more σ-algebra exercises

Exercise:

(1) Explain what a sigma algebra is to your classmate, and provide an example

(2) Have your classmate repeat the definition to you and provide you with an example

(3) Now both of you provide an example of a nonempty collection of subsets of a probability space that is *not* a σ-algebra

(4) How confident do you feel about your level of understanding of this concept?

Events

- An "event" B is a set of possible outcomes of an experiment
- Events A and B are "mutually exclusive" if they can't occur at the same time
  - $\rightarrow$ $A \cap B = \emptyset$
- Events A and B are "collectively exhaustive" if at least one of A or B <u>must</u> occur
  - $\rightarrow$ $A \cup B = \Omega$, where Ω is the "sample space" (the set of all possible outcomes)
  - $\rightarrow$ Note that if $(A \cup B) \subset \Omega$, then A and B are *not* collectively exhaustive

A "probability space"

- A triple $(\Omega, \mathcal{F}, P)$ that formally models a random experiment, comprising:
  (1) A "sample space", Ω, the set of possible outcomes
  (2) An "event space", $\mathcal{F}$, a σ-algebra on Ω of the events of interest
  (3) A "probability function", P, that assigns a "probability", $0 \leq p \leq 1$, to each event in $\mathcal{F}$

A probability function from Ω to the real numbers satisfies these axioms:

(1) $P(A) \geq 0$ for any event A

(2) If $A_1, A_2, \ldots$ are pairwise disjoint, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

→ $\sum$ is the "summation" operator. E.g. $\sum_{i=2}^{4} b_i = b_2 + b_3 + b_4$

(3) $P(\Omega) = 1$

When Ω is a countable set, write it as $\Omega = \{\omega_1, \omega_2, \ldots\}$, with every subset of Ω assigned a probability.
Then for any event A, $P(A) = \sum_{\omega_i \in A} P(\{\omega_i\})$

Some useful properties of probability functions

These properties follow from the axioms on the previous page:

- $P(A^c) = 1 - P(A)$
- $P(\emptyset) = 0$
- $0 \leq P(A) \leq 1$
- If $A \subset B$, then $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$

Some examples

Consider a single roll of a fair six-sided die

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Some examples of events in $\mathcal{F}$ might be:
  a) The simple event $\{6\}$ (6 is rolled)
     $\rightarrow$ if $\{6\}$ is in $\mathcal{F}$, then its complement must also be
  b) The complex event $\{1, 2\}$ (either 1 or 2 is rolled)
     $\rightarrow$ if $\{1, 2\}$ is in $\mathcal{F}$, then its complement must also be
- P maps each event in $\mathcal{F}$ to a probability. E.g.
  a) $\{6\}$ is mapped to $\frac{1}{6}$, and $\{1, 2, 3, 4, 5\}$ is mapped to $\frac{5}{6}$
  b) $\{1, 2\}$ is mapped to $\frac{2}{6} = \frac{1}{3}$, and $\{3, 4, 5, 6\}$ is mapped to $\frac{2}{3}$

Exercise:

(1) Explain a probability space to a classmate, then give a new example

(2) Have your classmate repeat the exercise

Random variables

- A "random variable" is a function mapping from a sample space to the real numbers
  - $\rightarrow$ It assigns a numerical value to each possible outcome of a random experiment
  - $\rightarrow$ If X is a r.v., then $X\colon \Omega \rightarrow \mathbb{R}$

Consider a single flip of a fair coin

- What is the sample space, Ω?

Random variables

- A "random variable" is a function mapping from a sample space to the real numbers
  - → It assigns a numerical value to each possible outcome of a random experiment
  - → If X is a r.v., then $X\colon \Omega \to \mathbb{R}$

Consider a single flip of a fair coin

- What is the sample space, Ω?
  - → $\Omega = \{H, T\}$
- Let the event space contain the two simple events $\{H\}$ and $\{T\}$ (this experiment would be pretty boring if we used the trivial σ-algebra! Why?)
- With a fair coin we know that $P(\emptyset) = 0$, $P(\{H\}) = P(\{T\}) = 0.5$, and $P(\Omega) = 1$

Does this define a r.v.? No! It defines a probability space. Importantly: $H, T \notin \mathbb{R}$

Random variables

Let $X \colon \Omega \to \mathbb{R}$ be a r.v. for one coin flip, where:

- $\Omega = \{H, T\} = \{\omega_1, \omega_2\}$
- $\mathcal{F} = \{\emptyset, \omega_1, \omega_2, \Omega\}$
- $P(\omega_1) = P(\omega_2) = 0.5$ (ignore the trivial $P(\emptyset = 0)$ and $P(\Omega = 1)$)
- Let $\mathcal{X} = \{0, 1\} =$ "range" of X such that $X(\omega) = \begin{cases} 0 \text{ if } \omega = H \\ 1 \text{ if } \omega = T \end{cases}$

Then the probability function P_X associated with X is:

- $P_X(X = x_i) = P(\{\omega_i \in \Omega \colon X(\omega_i) = x_i\}) \; \forall \; x_i \in \mathcal{X}$
- Then $P(X = 0) = P(X = 1) = 0.5$

Discrete and continuous random variables

- A “discrete” r.v. takes on a countable number of possible values
  - $\rightarrow$ Usually but not necessarily integers
- A “continuous” r.v. takes on an uncountably infinite number of possible values
  - $\rightarrow$ Usually a measurement
- A “random process” is an indexed collection of random variables

Some examples of discrete and continuous random processes

Description	Type	Potential Outcomes
Six-sided die roll	Discrete	?
Coin flip	?	?
Income of UVic graduate students	?	?

Some examples of discrete and continuous random processes

Description	Type	Potential Outcomes
Six-sided die roll	Discrete	1, 2, 3, 4, 5, 6
Coin flip	Discrete	Heads, Tails
Income of UVic graduate students	Continuous	$M \geq 0$

Probability mass function

A "probability mass function" (pmf), $P(X = x)$, is associated with a discrete r.v. X

- Yields the probability that X exactly equals a real value, $-\infty < x < \infty$, $\forall\ x \in X$
  - → E.g. our single coin flip example
- May also be written $p_X(x)$, $f_X(x)$, or even just $f(x)$ with just one r.v.
- The probabilities associated with all possible values must be non-negative and sum to 1
  - → $f_X(x) \geq 0\ \forall\ x$
  - → $\sum_x f_X(x) = 1$

Probability density function

A "probability density function" (pdf), $f_X(x)$, is associated with a continuous r.v. X

- Associates a probability with each <u>range</u> of realizations of X
  - → The integral of the pdf over the range (the area under the pdf over the range)
  - → E.g. For some a and b such that $a < b$, $P(a \leq X \leq b) = \int_a^b f_X(x)dx$
    - · $\int$ is the "integration" operator
  - → Think of $f_X(x)dx$ as the probability that X falls in the *tiny* interval $[x, x + dx]$
- The pdf of a r.v. X is the function that satisfies $F_X(x) = \int_{-\infty}^{x} f_X(u)du \ \forall \ x$ where, if $f(x)$ is continuous at x, $f_X(x) = \frac{dF_X(x)}{dx} = F'_X(x)$
  - → $F_X(x)$ is the "cumulative distribution function" (cdf) associated with X
- Also:
  - → $f_X(x) \geq 0 \ \forall \ x$
  - → $\int_x f_X(x)dx = 1$

Probability definitions

Let A and B be two events

- $P(A)$ is the "marginal" (or "unconditional") probability that A occurs
- $P(A \mid B)$ is the "conditional" probability that A occurs given that B occurs
  $\rightarrow$ $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ if $P(B) \neq 0$
- $P(A, B)$ is the "joint" probability that A and B both occur
  $\rightarrow$ It is the prob. of the intersection of A and B: $P(A, B) = P(A \cap B)$

Bayes' theorem

Let A and B be two events, and let $P(A) \neq 0$ and $P(B) \neq 0$

Using the definition of conditional probability we know:

- $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$
- $P(B \mid A) = \frac{P(A \cap B)}{P(A)}$

Then: $P(A \mid B) = ?$

Bayes' theorem

Let A and B be two events, and let $P(A) \neq 0$ and $P(B) \neq 0$

Using the definition of conditional probability we know:

- $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$
- $P(B \mid A) = \frac{P(A \cap B)}{P(A)}$

Then: $P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$

Statistical independence

Two events, A and B, are (statistically) "independent" if the occurrence of one does not affect the probability that the other occurs

- $P(A \mid B) = P(A)$
- $P(A \cap B) = P(A)P(B)$

Example: The outcome of each coin flip is independent of the others, with $P(H_s) = P(T_s) = 0.5 \; \forall \; s = 1, 2, ..., S$. Say you're going to flip twice ($S = 2$):

- $P(H_2 \mid T_1) = \, ?$
- $P(H_2 \cap T_1) = \, ?$

Statistical independence

Two events, A and B, are (statistically) "independent" if the occurrence of one does not effect the probability that the other occurs

- $P(A \mid B) = P(A)$
- $P(A \cap B) = P(A)P(B)$

Example: The outcome of each coin flip is independent of the others, with $P(H_s) = P(T_s) = 0.5 \ \forall \ s = 1, 2, ..., S$. Say you're going to flip twice ($S = 2$):

- $P(H_2 \mid T_1) = P(H_s) = 0.5$
- $P(H_2 \cap T_1) = P(H_s)P(T_s) = 0.5 \times 0.5 = 0.25$

Lab exercise 1

With a standard deck of 52 cards and a single draw, the probability pulling any particular card is $P(\text{any given card}) = \frac{1}{52}$

(1) What is the probability of drawing a Jack with your first draw?

(2) Say you draw a Jack, then return it to the deck and shuffle. What is the probability you draw another Jack with your second draw? Are the two events independent?

(3) Say, instead, that you drew a Jack with your first draw and kept it. What is the probability you draw another Jack with your second draw? Are the two events independent?

Lab exercise 2

Suppose you wake up one morning in Victoria to the radio telling you that the Blue Jays won last night. As your eyes clear, you look out the tiny window of your basement suite and see your landlord leaving for work with an umbrella. You're so tired, you can't remember what month it is. You wonder if you should dress for rain.

You turn to your phone to check the weather but, horror of horrors: your phone is dead! After checking the charger, you realize the power is out. Good thing you have a battery-powered radio alarm clock that woke you up! But with no power, your WiFi is down. With no WiFi and a dead phone, you don't know if you should dress for rain. That is you don't know the probability that it will rain given your landlord left with an umbrella, $P(R \mid U)$!

From past experience, you know the probability that it will rain on any given day in Victoria is $P(R) = 0.2$ and that your landlord takes an umbrella to work half the time: $P(U) = 0.5$. You also know your landlord takes an umbrella to work 80% of the days it rains: $P(U \mid R) = 0.8$.

(1) Assume the Blue Jays' successes are independent of whether it rains. What is $P(R \mid J_{win})$?

(2) Is your landlord's umbrella choice independent of whether it rains? How do you know?

(3) What is the probability it will rain given your landlord took an umbrella?

Lab exercise 3

Consider the experiment of flipping a coin twice

(1) What is the sample space? Use "H" to denote Heads and "T" to denote Tails

(2) Consider the r.v.s $X_1(\omega) = \begin{cases} 0 \text{ if } \omega \in \{HH, HT\} \\ 1 \text{ if } \omega \in \{TH, TT\} \end{cases}$ $X_2(\omega) = \begin{cases} 2 \text{ if } \omega \in \{HH\} \\ 1 \text{ if } \omega \in \{HT\} \\ 1 \text{ if } \omega \in \{TH\} \\ 0 \text{ if } \omega \in \{TT\} \end{cases}$

- $\rightarrow$ What are $\sigma(X_1)$ and $\sigma(X_2)$ (the σ-algebras for X_1 and X_2, respectively)?
- $\rightarrow$ Which r.v. is more informative? Why?

(3) What are the probability functions associated with each r.v.?

(4) If you had to guess, what would you say the values of X_1 and X_2 correspond to? Describe X_1 and X_2 colloquially. Do r.v.s need to have such intuitive interpretations?

Lab exercise 4

(1) What is a probability mass function (PMF)? Explain all the properties

(2) How does a PMF differ from a probability density function (PDF)? Explain the properties of a PDF

Further lab exercises (repeat from the lesson)

(1) If $A \subset X$, what is the smallest σ-algebra containing A?

(2) If $X = \{1, 2, 3\}$, what is a non-trivial σ-algebra?

(3) Explain what a sigma algebra is to your classmate, and provide an example

(4) Have your classmate repeat the definition to you and provide you with an example

(5) Now both of you provide an example of a nonempty collection of subsets of a probability space that is *not* a σ-algebra

(6) How confident do you feel about your level of understanding of this concept?

(7) Explain a probability space to a classmate, then give a new example

(8) Have your classmate repeat the exercise

Resources

This lecture drew notes from:

- Wikipedia (most things we've seen in quotations)
- The Bright Side of Mathematics YouTube channel
- Daisy Huang's 2006 notes on probability
- University of Warwick M.Sc. Financial Mathematics notes on probability