Welcome

- Welcome to **Statistics**, Summer 2023.
- **Goal of the course**: prepare students for 1st-year econometrics sequence.
- Structure
  - Lectures on Mondays, Wednesdays, and Fridays at 9-11h.
  - TA sessions on Fridays (?) with Bruno Daré.
  - Grading: statistics problem sets (10%), first exam (45%), second exam (45%).
  - Problem sets due one week later.
  - Main reference: Casella & Berger (2002) Statistical Inference

Probability Theory

Ricardo Dahis

PUC-Rio, Department of Economics

Summer 2023

"You can, for example, never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician." – Sherlock Holmes, The Sign of Four.

- **Goal of this lecture:** outline some basic ideas of probability theory that are crucial to the study of statistics.

Contents

Contents

In this section, we aim to study the following questions:

- What is a "probability"?
- Do probabilities always exist? (the answer is no)
- Can we define probabilities on any collection of events? (the answer is no)
- Under which conditions is it possible to assign probabilities? (we need σ-algebras)
- How to calculate probabilities when they exist?

probability spaces

- **Preview:** a **probability space** is a triplet $(\Omega, \mathcal{B}, \mathbb{P})$ where
  - Ω is a set
  - $\mathcal{B}$ is a σ-algebra on Ω
  - $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{B})$

- The basic interpretation of the components is as follows:
  - Ω is the *sample space*, which consists of all possible "states" relevant to the experiment – or set of all possible outcomes.
  - $\mathcal{B}$ is the collection of *events*, which is a collection of subsets of Ω to which we aim to assign a probability.
  - $\mathbb{P}$ is a probability measure, which assigns a number between 0 and 1 to events in $\mathcal{B}$ based on how likely the event is to occur.

Contents

fundamental sets

- **definition**: the set Ω collects all possible outcomes of a particular experiment, also called the *sample space*.

- **examples**:
  - Tossing a coin: $\Omega = \{H, T\}$
  - Gauging time between price changes: $\Omega = (0, \infty)$
  - Exam outcomes: $\Omega = [0, 10]$
  - The first is countable due to the one-to-one correspondence with integers, whereas the last two sets are uncountable

fundamental sets

- **definition**: An *event* is any collection of possible outcomes of an experiment, that is, any subset of Ω (including Ω itself).
- **example**: consider the experiment of selecting a random card from a deck and noting its suit: clubs (C), diamonds (D), hearts (H), spades (S).
- The sample space is

$$\Omega = \{C, D, H, S\}$$

- Possible events are

$$\begin{aligned} A &= \emptyset \\ B &= \{D, S\} \\ C &= \{C, D, H\} \\ D &= \{C, D, H, S\} = \Omega \end{aligned}$$

where $\emptyset$ represents the empty set.

elementary set operations

- **union** $A \cup B = \{x \,|\, x \in A \text{ or } x \in B\}$
- **intersection** $A \cap B = \{x \,|\, x \in A \text{ and } x \in B\}$
- **subtraction** $A - B = \{x \,|\, x \in A \text{ and } x \notin B\}$
- **complementation** $A^c = \{x \,|\, x \notin A\} = \Omega - A$
- **how to order and equate sets?**

$$A \subset B \quad \Leftrightarrow \quad x \in A \Rightarrow x \in B$$
$$A = B \quad \Leftrightarrow \quad A \subset B \Rightarrow B \subset A$$

how to combine elementary set operations?

- **commutativity**

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

- **associativity**

$$A \cup (B \cup C) = (A \cup B) \cup C$$
$$A \cap (B \cap C) = (A \cap B) \cap C$$

- **distributive laws**

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

- **DeMorgan's laws**

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

proof of the first distributive law

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

- formal proof requires to show that the sets in the RHS and in the LHS contain each other.
- (proof) ($\Rightarrow$): Let $x \in A \cap (B \cup C) = \{x \mid x \in A \text{ and } x \in (B \cup C)\}$, meaning that x is definitely either in B or in C. Given that x must also belong to A, we have that x is either in $A \cap B$ or in $A \cap C$. This implies that $x \in (A \cap B) \cup (A \cap C)$. ■
- (proof) ($\Leftarrow$): Now assume the latter, that is to say, that x is either in $A \cap B$ or in $A \cap C$. If $x \in (A \cap B)$, then x is both in A and in B. However, as x is in B, it must also belong to $B \cup C$, implying that $x \in A \cap (B \cup C)$. If, on the other hand, $x \in (A \cap C)$, by the same reasoning, it must also be in $A \cap (B \cup C)$, completing the proof. ■

extending to infinite collections...

- if set collection is countable,

$$\bigcup_{i=1}^{\infty} A_i = \{x \mid x \in A_i \text{ for some } i\}$$

$$\bigcap_{i=1}^{\infty} A_i = \{x \mid x \in A_i \text{ for every } i\}$$

- **example:** Let $S = (0, 1]$ and define $A_i = [\frac{1}{i}, 1]$. Then

$$\begin{aligned}
\bigcup_{i=1}^{\infty} A_i &= \bigcup_{i=1}^{\infty} \left[\frac{1}{i}, 1\right] = \left\{x \in (0,1] | x \in \left[\frac{1}{i}, 1\right] \text{ for some } i\right\} \\
&= \{x \in (0,1]\} = (0,1] \\
\bigcap_{i=1}^{\infty} A_i &= \bigcap_{i=1}^{\infty} \left[\frac{1}{i}, 1\right] = \left\{x \in (0,1] | x \in \left[\frac{1}{i}, 1\right] \text{ for all } i\right\} \\
&= \{x \in (0,1] | x \in [1,1]\} = \{1\}
\end{aligned}$$

extending to uncountable infinite collections...

- extends naturally to uncountable collections of sets

$$\bigcup_{a\in\Gamma} A_a = \{x \in S | x \in A_a \text{ for some } a\}$$
$$\bigcap_{a\in\Gamma} A_a = \{x \in S | x \in A_a \text{ for all } a\}$$

- we can take, for example, $\Gamma = \mathbb{R}$.

some final terminology

- A and B are mutually exclusive (or disjoint) if $A \cap B = \emptyset$
- $A_1, A_2, \ldots$ are pairwise disjoint if $A_i \cap A_j = \emptyset$ for all $1 \leq i \neq j \leq \infty$
- **definition**: If $A_1, A_2, \ldots$ are pairwise disjoint and such that $\cup_{i=1}^{\infty} A_i = S$, then we say that they form a partition of the sample space.
- for example, $A_i = [i, i+1)$ is a partition of $[1, \infty)$.
  - $[1,2), [2,3), [3,4), \ldots$
- partitions are very useful in that they allow us to divide the sample space into small, non-overlapping pieces.

Contents

σ-algebras

- **definition**: a σ-algebra on Ω is a collection of sets $\mathcal{B}$ of subsets of Ω such that
  (i) $\Omega \in \mathcal{B}$
  (ii) $A \in \mathcal{B} \Rightarrow A^c \in \mathcal{B}$
  (iii) $A_1, \ldots, A_n, \ldots \in \mathcal{B} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$

- It follows immediately that $\emptyset \in \mathcal{B}$ by (i)+(ii).

- The power set $P(\Omega)$ of a set Ω is a σ-algebra. (proof)

- An algebra is such that (iii') $A_1, \ldots, A_n \in \mathcal{B} \Rightarrow \bigcup_{i=1}^{n} A_i \in \mathcal{B}$, for $n \geq 1$.
  $\Longrightarrow$ every σ-algebra is an algebra (why?)

σ-algebras examples

Which of those are $\sigma-$algebras with $\Omega = \{1, 2, 3\}$?

(1) $\mathcal{B} = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$.

(2) $\mathcal{B} = \{\emptyset, \{1, 2\}, \{2, 3\}, \Omega\}$.

(3) $\mathcal{B} = \{\emptyset, \Omega\}$.

(4) $\mathcal{B} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \Omega\}$.

(5) $\mathcal{B} = \{\emptyset, \{1, 2\}, \{1, 3\}, \{2, 3\}, \Omega\}$

answer: (1), (3), (4)

- **proposition:** $\mathcal{B}$ is closed under countable intersections,

$$A_1, \ldots, A_n, \ldots \in \mathcal{B} \quad \Rightarrow \quad \bigcap_{i=1}^{\infty} A_i \in \mathcal{B}$$

- (proof) (ii) implies that $A_1^c, A_2^c, \ldots \in \mathcal{B}$ and hence it follows from (iii) that $\cup_{i=1}^{\infty} A_i^c \in \mathcal{B}$. From DeMorgan's law and (ii), we have $\cap_{i=1}^{\infty} A_i \in \mathcal{B}$ given that

$$\left(\bigcup_{i=1}^{\infty} A_i^c \right)^c = \bigcap_{i=1}^{\infty} A_i$$

$\blacksquare$

σ-algebras

- **proposition:** given a σ-algebra $\mathcal{B}$ on Ω, let I be an arbitrary non-empty index set and $\{\mathcal{B}_\alpha\}_{\alpha \in I}$ a family of σ-algebras on Ω. The collection $\cap_{\alpha \in I} \mathcal{B}_\alpha$ is a σ-algebra on Ω.

- (proof) The family of events $\cap_{\alpha \in I} \mathcal{B}_\alpha$ satisfies property (iii) because

$$
\begin{aligned}
A_1, A_2, \ldots \in \bigcap_{\alpha \in I} \mathcal{B}_\alpha \quad &\Rightarrow \quad A_1, A_2, \cdots \in \mathcal{B}_\alpha \text{ for all } \alpha \in I \\
&\Rightarrow \quad \bigcup_{i=1}^{\infty} A_i \in \mathcal{B}_\alpha \text{ for all } \alpha \in I \\
&\Rightarrow \quad \bigcup_{i=1}^{\infty} A_i \in \bigcap_{\alpha \in I} \mathcal{B}_\alpha
\end{aligned}
$$

■

- Given two σ-algebras $\mathcal{B}$ and $\mathcal{G}$, the collection of events $\mathcal{B} \cup \mathcal{G}$ is not necessarily a σ-algebra.

$$
\begin{aligned}
\mathcal{B} &= \{\emptyset, \Omega, \{1,2\}, \{3,4\}\} \\
\mathcal{G} &= \{\emptyset, \Omega, \{1\}, \{2,3,4\}\} \\
\mathcal{B} \cup \mathcal{G} &= \{\emptyset, \Omega, \{1,2\}, \{3,4\}, \{1\}, \{2,3,4\}\}
\end{aligned}
$$

σ-algebras

- **definition**: given a collection $\mathcal{C}$ of subsets of Ω, the σ-algebra $\sigma(\mathcal{C})$ on Ω **generated by** $\mathcal{C}$ is the smallest σ-algebra containing $\mathcal{C}$. It is the intersection of all σ-algebras on Ω which have $\mathcal{C}$ as a subclass.

- example: given any subset A of Ω, the smallest σ-algebra containing A is

$$\sigma(A) = \{\emptyset, \Omega, A, A^c\}$$

- example: suppose that $\Omega = \{1, 2, 3, 4\}$, and $\mathcal{C} = \{\{1\}, \{1, 3, 4\}\}$. Then

$$\sigma(\mathcal{C}) = \{\emptyset, \Omega, \{1\}, \{1, 3, 4\}, \{2, 3, 4\}, \{2\}, \{1, 2\}, \{3, 4\}\}$$

Borel σ-algebras

- **definition:** The Borel σ-algebra $\mathcal{B}(\Omega)$ on Ω is the σ-algebra on Ω generated by the family of all open sets,

$$\mathcal{B}(\Omega) \quad = \quad \sigma\left(\{A \subseteq \Omega |\ A \text{ is open}\}\right)$$

- The most important σ-algebras are the Borel σ-algebras on $\mathbb{R}^n$.
- proposition: For real numbers, we have the following equivalent characterizations:
  - (i) $\mathcal{B}(\mathbb{R}) = \sigma\left(\{(a,b) | a,b \in \mathbb{R}, a < b\}\right)$
  - (ii) $\mathcal{B}(\mathbb{R}) = \sigma\left(\{(a,b] | a,b \in \mathbb{R}, a < b\}\right)$
  - (iii) $\mathcal{B}(\mathbb{R}) = \sigma\left(\{[a,b] | a,b \in \mathbb{R}, a < b\}\right)$
  - (iv) $\mathcal{B}(\mathbb{R}) = \sigma\left(\{(-\infty,a] | a \in \mathbb{R}\}\right)$

Borel σ-algebras

- (proof of i) ($\Leftarrow$) An open interval (a, b) is an open set, hence

$$\{(a,b)|a,b \in \mathbb{R}, a < b\} \quad \subseteq \quad \{A \subseteq \mathbb{R}|\ A \text{ is open}\} \quad \subseteq \quad \mathcal{B}(\mathbb{R})$$

Thus

$$\sigma\left(\{(a,b)|a,b \in \mathbb{R}, a < b\}\right) \quad \subseteq \quad \mathcal{B}(\mathbb{R})$$

by definition of generated σ-algebras ("intersection of all σ-algebras on Ω which contain $\mathcal{C}$.") ■

- ($\Rightarrow$) Let $O \subseteq \mathbb{R}$ be an open set. Since every open set in $\mathbb{R}$ is a countable union of open intervals, there exists sequences of real numbers $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ with $a_k < b_k$ for all $k = 1, 2, \ldots$ such that

$$O \quad = \quad \bigcup_{n=1}^{\infty} (a_n, b_n)$$

Since a σ-algebra is closed under countable unions, it follows that

$$O \quad = \quad \bigcup_{n=1}^{\infty} (a_n, b_n) \quad \in \quad \sigma\left(\{(a,b)|a,b \in \mathbb{R}, a < b\}\right)$$

Borel σ-algebras

- we conclude that

$$\mathcal{B}(\mathbb{R}) \quad = \quad \sigma\left(\{A \subseteq \mathbb{R} | A \text{ is open}\}\right) \quad \subseteq \quad \sigma\left(\{(a,b) | a,b \in \mathbb{R}, a < b\}\right)$$

  which finishes the proof. ■

- Let $A \subseteq \mathbb{R}$ be a countable set. Then $A \in \mathcal{B}(\mathbb{R})$. (proof left as an exercise)

Contents

from set theory to probability theory...

- We have investigated set relations and σ-algebras. We can now link those sets to **probabilities**.

- example: when tossing a fair dice, the possible outcomes are 1, 2, 3, 4, 5 and 6. Therefore a natural choice for the sample space is

$$\Omega \quad = \quad \{1, 2, 3, 4, 5, 6\}.$$

We then associate an outcome $\omega \in \Omega$ with the number of black dots that turn up on the dice. As our collection of sets $\mathcal{B}$ to which we want to assign probabilities, lets assume that $\mathcal{B}$ is the power set of Ω, i.e., $\mathcal{B}$ consists of all possible subsets of Ω. Finally, we define a probability measure $\mathbb{P}$ for sets $A \in \mathcal{B}$ by

$$\mathbb{P}(A) \quad = \quad \frac{\text{number of elements in } A}{6}$$

With this definition,
  - $\mathbb{P}(\Omega) = 1$
  - $\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \cdots = \mathbb{P}(\{6\}) = 1/6$
  - if A and B are disjoint in $\mathcal{B}$, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

relative frequency and probability...

- **idea**: if we carry out an experiment a number of times, different outcomes may arise each time or some outcomes may repeat
- frequency of occurrence $\sim$ probability
  - relative frequency of any event is always between zero and one
  - fundamental law of statistics (a.k.a. Glivenko-Cantelli theorem) says that, as the number of experiments grows to infinity, the relative frequency of an event converges to its probability.
  - this is very appealing and intuitive, but involves some philosophical interpretations that we should perhaps not mess with, so we will take a more axiomatic approach.
    - P.S. see Dahis (2019) for a discussion about randomness and probability.
    - We do not observe the *true* probability. We make guesses, maybe based on axioms.
    - "Probability is a claim about a variable's frequency distribution".

Kolmogorov's axiomatic foundations

- **goal**: for each event A in the sample space Ω, we wish to associate with A a number $0 \leq \mathbb{P}(A) \leq 1$ that we will call the probability of A.

- The probability of A also has to satisfy some intuitive properties.

- **definition**: given a sample space Ω and a corresponding $\sigma-$algebra $\mathcal{B}$, a probability function $\mathbb{P}(\cdot)$ with domain $\mathcal{B}$ satisfies
  - (i) $\mathbb{P}(A) \geq 0$ for all $A \in \mathcal{B}$
  - (ii) $\mathbb{P}(\Omega) = 1$, and
  - (iii) $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ if $A_1, A_2, \ldots \in \mathcal{B}$ are pairwise disjoint

Kolmogorov's axiomatic foundations

- In view of the example above, why do we need to specify the collection of sets $\mathcal{B}$?
- In the dice example, specification of $\mathcal{B}$ was not crucial because the set of outcomes was **finite**.
- When the set of outcomes is **infinite**, e.g., the real numbers, things become much more difficult, and we will have to be much more formal to ensure that probabilities exist.

defining a legitimate probability function

- **theorem**: let $\mathcal{B}$ denote any σ-algebra of subsets of a finite sample space $S = \{s_1, s_2, \ldots, s_n\}$ and let $p_1, p_2, \ldots, p_n$ denote nonnegative numbers that sum to 1; for any $A \in \mathcal{B}$, define

$$\mathbb{P}(A) = \sum_{\{i \mid s_i \in A\}} p_i .$$

Then $\mathbb{P}(\cdot)$ is a **probability function** on $\mathcal{B}$.

- **proof**: (i) is true because $\mathbb{P}(A) = \sum_{\{i \mid s_i \in A\}} p_i \geq 0$ for any $A \in \mathcal{B}$ given that every $p_i \geq 0$; (ii) holds because $\mathbb{P}(S) = \sum_{\{i \mid s_i \in S\}} p_i = \sum_{i=1}^{n} p_i = 1$; and (iii) is true because, if $A_1, A_2, \ldots, A_n \in \mathcal{B}$ are pairwise disjoint events, then

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{\{j \mid s_j \in \bigcup_{i=1}^{n} A_i\}} p_j = \sum_{i=1}^{n} \sum_{\{j \mid s_j \in A_i\}} p_j = \sum_{i=1}^{n} \mathbb{P}(A_i)$$

by definition of $\mathbb{P}(A)$ and by disjointness of the A_i's given that the same p_j's appear exactly once on each side of the equality. ■

why do we need a collection of sets $\mathcal{B}$?

- Let A_1, A_2, ... be a countably infinite collection of disjoint sets in $\mathcal{B}$. We require that

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

  and $P(\Omega) = 1$.

- counterexample: consider the case of drawing a random number with equal probability from the interval $\Omega = [0, 1) \cap \mathbb{Q}$. $\mathbb{Q}$ is countable, so there are A_1, A_2, ..., such that $\Omega = \bigcup_{n=1}^{\infty} A_n$ and where each A_n contains a single element. It follows that

$$\mathbb{P}([0,1) \cap \mathbb{Q}) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

  The sum on the RHS is either 0 or ∞, which contradicts $\mathbb{P}([0,1) \cap \mathbb{Q}) = 1$. ■

- Counterexample also holds to **uncountably** infinite union of sets.

why do we need a collection of sets $\mathcal{B}$?

- **takeaway:** it is not straightforward to assign probabilities that satisfy Komolgorov's axioms when Ω is infinite.
- what sort of properties do we want the collection of sets $\mathcal{B}$ to which we can assign a probability to satisfy?
- It turns out that we have to define a **measure** and to require $\mathcal{B}$ to be a σ-algebra.

measure

- **definition**: Let $(S, \mathcal{S})$ be a measurable space, so that $\mathcal{S}$ is a σ-algebra on the set S. A **measure** defined on $(S, \mathcal{S})$ is a function $\mu : \mathcal{S} \to [0, \infty]$ that is countably additive, i.e., it is such that
  (i) $\mu(\emptyset) = 0$
  (ii) if $A_1, A_2, \dots \in \mathcal{S}$ is any sequence of pairwise disjoint sets, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu\left(A_n\right)$$

- The triplet $(S, \mathcal{S}, \mu)$ is called a **measure space**. Given a measure space, we say that μ is a **probability measure** if $\mu(S) = 1$.
  - The triplet $(S, \mathcal{S}, P)$ is called a **probability space**.

- We also say that μ is a finite measure if $\mu(S) < \infty$. μ is a σ-finite measure if there is a sequence $A_1, A_2, ... \in \mathcal{S}$ such that

$$\mu(A_n) < \infty \text{ for all } n \geq 1 \text{ and } \bigcup_{n=1}^{\infty} A_n = S$$

some mechanics of measures

- Let $(S, \mathcal{S}, \mu)$ be a measure space. Given any increasing sequence $A_1 \subseteq A_2 \subseteq \cdots \subseteq$ of sets in $\mathcal{S}$, we can define the limit of the sequence by

$$\lim_{n \to \infty} A_n = \bigcup_{n=1}^{\infty} A_n$$

Then

$$\mu\left(\lim_{n \to \infty} A_n\right) = \lim_{n \to \infty} \mu\left(A_n\right).$$

- Similarly, if $A_1 \supseteq A_2 \supseteq \cdots$ is a decreasing sequence of sets in $\mathcal{S}$, the limit is

$$\lim_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} A_n$$

In this case, if $\mu(A_1) < \infty$ then

$$\mu\left(\lim_{n \to \infty} A_n\right) = \lim_{n \to \infty} \mu(A_n).$$

- (proof) Given an increasing sequence $A_1 \subseteq A_2 \subseteq \cdots$ of sets in $\mathcal{S}$, let $B_1 = A_1$ and define recursively $B_n = A_n \backslash A_{n-1}$ for $n \geq 2$. By construction, the events $B_1, B_2, \ldots$ are pairwise disjoint,

$$A_n = \bigcup_{k=1}^{n} B_k \quad \text{and} \quad \bigcup_{n=1}^{\infty} A_n = \bigcup_{k=1}^{\infty} B_k.$$

As a consequence,

$$\begin{aligned}
\mu\left(\lim_{n\to\infty} A_n\right) &= \mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{k=1}^{\infty} B_k\right) \\
&= \sum_{k=1}^{\infty} \mu(B_k) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(B_k) \\
&= \lim_{n\to\infty} \mu\left(\bigcup_{k=1}^{n} B_k\right) = \lim_{n\to\infty} \mu(A_n)
\end{aligned}$$

■

The proof is similar for the decreasing case. (exercise!)

intuition for next slides

- In many cases it is hard to define a probability function on all sets A in a σ-algebra.
- Caratheodory's extension theorem shows that it is sufficient to define the probability measure on an *algebra* $\mathcal{A}$ instead. The probability measure is then uniquely defined on $\sigma(\mathcal{A})$, in a way consistent with its definition on $\mathcal{A}$.
  - The unique measure will be the *Lebesgue measure*.
- "Procedure"
  1. We will start from an algebra of intervals.
     - "An algebra is a collection of subsets closed under finite unions and intersections."
  2. Define the pre-measure with certain properties.
  3. The theorem implies there exists a unique "length" measure on the sigma algebra generated by the algebra.
     - "A sigma algebra is a collection closed under countable unions and intersections."

construction of a measure

(Caratheodory's Extension Theorem) Suppose that $\mathcal{A}$ is an **algebra** on a set Ω and let $\mathcal{S} = \sigma(\mathcal{A})$. Also, suppose that a pre-measure function $\mu_0 : \mathcal{A} \to [0, \infty)$ is countably additive and σ-finite in the sense that

(i) $\mu_0(\emptyset) = 0$ and

(ii) if $A_1, A_2, \cdots \in \mathcal{A}$ is any sequence of pairwise disjoint set such that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, then

$$\mu_0\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu_0(A_n)$$

(iii) there exists $A_1, A_2, \cdots \in \mathcal{A}$ such that $\mu_0(A_n) < \infty$ for all $n = 1, 2, \ldots$ and $\bigcup_{n=1}^{\infty} A_n = S$

then there exists a unique measure μ on $(S, \mathcal{S})$ such that

$$\mu(A) = \mu_0(A) \text{ for all } A \in \mathcal{A}.$$

construction of a measure

- Let $\mathcal{A}$ be the collection of sets $C \subseteq \mathbb{R}$ that admit the representation

$$C = (a_1, b_1] \cup \cdots \cup (a_k, b_k]$$

or

$$C = (-\infty, b_0] \cup (a_1, b_1] \cup \cdots \cup (a_k, b_k]$$

or

$$C = (a_1, b_1] \cup \cdots \cup (a_k, b_k] \cup (a_{k+1}, \infty)$$

or

$$C = (-\infty, b_0] \cup (a_1, b_1] \cup \cdots \cup (a_k, b_k] \cup (a_{k+1}, \infty)$$

for some $k \geq 1$ and reals $-\infty < b_0 < a_1 < b_1 < \cdots < a_k < b_k < a_{k+1} < \infty$. We can check that this collection is an algebra on $\mathbb{R}$ and that $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$.

construction of a measure

- Let $F : \mathbb{R} \to \mathbb{R}$ be increasing and right-continuous, i.e., for every $x \in \mathbb{R}$, the limit from the right $\lim_{y \to x^+} = F(x)$).
- We define the function $\mu_0 : \mathcal{A} \to [0, \infty)$ as follows

$$\mu_0(C) = \sum_{n=1}^{k} F(b_n) - F(a_n)$$

or

$$\mu_0(C) = F(b_0) - F(-\infty) + \sum_{n=1}^{k} F(b_n) - F(a_n)$$

or

$$\mu_0(C) = \sum_{n=1}^{k} F(b_n) - F(a_n) + F(\infty) - F(a_{k+1})$$

or

$$\mu_0(C) = F(b_0) - F(-\infty) + \sum_{n=1}^{k} F(b_n) - F(a_n) + F(\infty) - F(a_{k+1})$$

construction of a measure

- It is not trivial but it can be shown that μ_0 is countably additive.
- So, using the Caratheodory's Extension Theorem, μ_0 has a unique extension to a measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- Consider the function μ_0 and an algebra $\mathcal{A}$ with $F(x) = x$, which provides the existence of a unique measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that

$$\mu((a, b]) = b - a \ , \quad \text{for every } a < b$$

  This measure, which assigns a "length" to every Borel set in $\mathbb{R}$, is the Lebesgue measure.
- If we restrict the Lebesgue measure to Borel sets on $[0, 1)$, we obtain a probability measure $\mathbb{P}$ on $([0, 1], \mathcal{B}([0, 1)))$ and, in particular, $\mathbb{P}$ satisfies
  (i) If $A_1, A_2, \ldots$ are disjoint sets in $\mathcal{B}([0, 1))$, then $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$
  (ii) If A is a translation of $B \in \mathcal{B}([0, 1))$, then $\mathbb{P}(A) = \mathbb{P}(B)$.
  (iii) $\mathbb{P}([0, 1)) = 1$.

Contents

probability calculus

- (a) $\mathbb{P}(\emptyset) = 0$
- (b) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- (c) $\mathbb{P}(A) \leq 1$
- (d) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
- (e) $\mathbb{P}(A) \leq \mathbb{P}(B)$ if $A \subset B$
- (f) $\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap C_i)$ for any partition $C_1, C_2, \ldots$ (law of total probability)
- **inequalities**
  - Bonferroni's: $\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1$ (particular case)
  - Boole's: $\mathbb{P}(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

proofs

- (a) $\mathbb{P}(A) = \mathbb{P}(A \cup \emptyset)$ and $A \cap \emptyset = \emptyset$ by definition, and hence $\mathbb{P}(A) = \mathbb{P}(A) + \mathbb{P}(\emptyset)$ by (i) ■
- (b)+(c) $A \cup A^c = S$ and $A \cap A^c = \emptyset$ by definition, implying that $\mathbb{P}(S) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c) = 1$ by (ii)+(iii) ■
- (d) $\mathbb{P}(A \cup B) = \mathbb{P}\left(A \cup (B \cap A^c)\right) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c)$ given that $A \cap (B \cap A^c) = \emptyset$, and hence $B = (A \cap B) \cup (B \cap A^c)$ implies that $\mathbb{P}(B) - \mathbb{P}(A \cap B) = \mathbb{P}(B \cap A^c)$ ■
- (e) $B = A \cup (B \cap A^c)$ given that $A \cap B = A$, implying that $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \cap A^c) \geq \mathbb{P}(A)$ ■

yet one more proof

- (f) as $C_1, C_2, \ldots$ form a partition, $C_i \cap C_j = \emptyset$ for all $i \neq j$ and $S = \cup_{i=1}^{\infty} C_i$, so that

$$A = A \cap S = A \cap \left(\bigcup_{i=1}^{\infty} C_i \right) = \bigcup_{i=1}^{\infty} (A \cap C_i)$$

by the distributive law. However, the sets $A \cap C_i$'s are pairwise disjoints given that the C_i's are mutually exclusive and hence

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} (A \cap C_i) \right) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap C_i) \qquad \blacksquare$$

Boole's inequality proof

- (Boole's Inequality) Construct a disjoint collection $A_1^*, A_2^*, \ldots$ such that $\cup_{i=1}^{\infty} A_i^* = \cup_{i=1}^{\infty} A_i$, defining

$$\begin{aligned} A_1^* &= A_1 \\ A_i^* &= A_i \setminus \left(\bigcup_{j=1}^{i-1} A_j \right), \quad i = 2, 3, \ldots \end{aligned}$$

where $A \setminus B$ denotes the part of A that does not intersect with B, i.e., $A \setminus B = A \cap B^c$. It is immediate to see that $\cup_{i=1}^{\infty} A_i^* = \cup_{i=1}^{\infty} A_i$. Then

$$P\left(\bigcup_{i=1}^{\infty} A_i \right) = P\left(\bigcup_{i=1}^{\infty} A_i^* \right) = \sum_{i=1}^{\infty} P\left(A_i^* \right).$$

The last equality follows since A_i^* are disjoint (**verify**). Since, by construction, $A_i^* \subset A_i$, so $P\left(A_i^* \right) \leq P\left(A_i \right)$ and

$$\sum_{i=1}^{\infty} P\left(A_i^* \right) \leq \sum_{i=1}^{\infty} P\left(A_i \right)$$

establishing the desired result. ■

Contents

finite sample spaces, with equiprobable elements

- We have seen that these numbers known as **probabilities** exist under certain conditions, and also explored some of its properties. We will now see how to **calculate** those probabilities.
- The simplest case is when sample space is finite and its elements are equiprobable. Then the probability of observing a given event is equal to the proportion of elements in the event relative to the sample space.
- **examples**
  - (1) suppose interest lies on the event of observing a value above 4 in a die roll: there are only two values in the sample space that satisfy this condition, namely, $\{5, 6\}$, and hence the probability of this event is $2/6 = 1/3$.
  - (2) consider now flipping twice a coin and recording heads and tails, so $S = \{HH, HT, TH, TT\}$: probability of observing only one head is

$$\frac{\#\{HT, TH\}}{\#\{HH, HT, TH, TT\}} = 1/2$$

- In these cases, counting equiprobable outcomes is straightforward. But now consider these examples:
  - (1, lottery) In the N.Y. state lottery, a person chooses six numbers from 1, 2, ..., 44. What is the probability that she wins?
  - (2, US Open) In a single-elimination tournaments, players advance if they win the match. For 16 entrants, how many paths are there to victory?
- The trick is breaking down the problem in a series of simple tasks.

- **theorem:** If a job consists of k separate tasks, the i-th of which can be done in n_i ways, then the entire job can be done in $\prod_{i=1}^{k} n_i$ ways.
- (1, lottery) If sampling is **without replacement**, the first number can be chosen in 44 ways and the second in 43 ways. So there are $44 \times 43 = 1,892$ ways of choosing the first two numbers.
- (1, lottery) If sampling is **with replacement**, there are $44 \times 44 = 1,936$ ways of choosing the first two numbers.
- Determining whether sampling is done with or without **replacement** is a crucial feature to take into consideration.
- The other important feature is whether **ordering** matters.

counting

	without replacement	**with replacement**
ordered		
unordered		

counting

- **ordered, without replacement**. There are

$$44 \times 43 \times 42 \times 41 \times 40 \times 39 \quad = \quad \frac{44!}{38!} \quad = \quad 5,082,517,440$$

  possible tickets.

- **ordered, with replacement**. Variation of the lottery example. There are

$$44^6 \quad = \quad 7,256,313,856$$

  possible tickets.

- **unordered, without replacement**: we know the number of possible tickets when orderings are accounted for, so we should divide by the number of redundancies.

$$\frac{44 \times 43 \times 42 \times 41 \times 40 \times 39}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = \frac{44!}{6!38!} = 7,059,052$$

since there are 6! ways to reorder 6 numbers.

- this is also known as the **binomial coefficients**:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

should be read as "n choose r".

counting

- the last remaining case is **unordered, with replacement**.
- think that we randomly put 6 markers on the 44 numbers.

M		MM		M	M			M
1	2	3	$\cdots$	31	32	33	$\cdots$	44

- we have 6 markers and 43 divisions between the boxes: this setting could be similarly represented as

$$\mathrm{M}||\mathrm{MM}|\cdots|\mathrm{M}|\mathrm{M}||\cdots|\mathrm{M}$$

- there are $(43+6)! = 49!$ ways of reordering the Ms and bars, but we have to discount for the reordering among them. The final number is

$$\frac{49!}{6!43!} = 13,983,816.$$

- generally,

$$\binom{n+r-1}{r} = \frac{(n+r-1)!}{r!(n-1)!}$$

	without replacement	**with replacement**
ordered	permutation $\frac{n!}{(n-r)!}$	multiplication n^r
unordered	combination $\binom{n}{r}$	combination $\binom{n+r-1}{r}$

- **question**: what is the probability that at least two people in this room have a common birthday?
- **answer**: it is easier to calculate what is the probability that, in a group of n people, all have different birthdays. Let it be $\bar{p}$.

$$\begin{aligned} \bar{p} &= 1 \times \left(\frac{364}{365}\right) \times \left(\frac{363}{365}\right) \times \cdots \left(\frac{365-(n-1)}{365}\right) \\ &= \frac{365!}{365^n(365-n)!} \end{aligned}$$

- for $n = 20$, $\bar{p}$ is about 59%. for $n = 35$, $\bar{p}$ is about 19%.
- the probability that at least two people have a common birthday is then 1-$\bar{p}$.

application to poker

- **traditional game**: 5-card poker hand from a standard 52-card deck sampling without replacement from the deck
- whether ordered or unordered depends on the event of interest! (e.g., probability of an ace in the first two cards)
- traditional poker game does not depend on order
- sample space S consists of all 5-card possible hands: $\binom{52}{5} = 2,598,960$
- **assumption**: well-shuffled deck and cards randomly dealt
- $\mathbb{P}(4 \text{ aces}) = 48/2,598,960$ because there are only 48 different possible last cards
- $\mathbb{P}(4 \text{ of a kind}) = 13 \times 48/2,598,960$ because there are 13 ways to specify which denomination there will be four of

more on poker

- question: how to compute the probability of exactly one pair?
- answer: the number of hands with exactly one pair is

$$\underbrace{13}_{(a)} \times \underbrace{\binom{4}{2}}_{(b)} \times \underbrace{\binom{12}{3}}_{(c)} \times \underbrace{4^3}_{(d)} \quad = \quad 1,098,240$$

(a) number of ways to specify the denomination of the pair

(b) number of ways to specify two cards from that denomination

(c) number of ways to specify other three denominations

(d) number of ways to specify the other three cards from those denominations

$$\mathbb{P}(\text{exactly one pair}) \;=\; \frac{1,098,240}{2,598,960}$$

conditional probability

- every probability thus far has been unconditional, **though...** in many instances, we are in a position to update the sample space based on new information

- **example**: what is the probability of drawing four aces from a well-shuffled deck? counting arguments yield $\binom{52}{4} = 270,725$ distinct groups of 4 cards, but only one of these groups consists of 4 aces

- updating approach $4/52 \times 3/51 \times 2/50 \times 1/49 = 1/270,725$

- **definition**: if A and B are events in S and $\mathbb{P}(B) > 0$, then the conditional probability of A given B is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

more on conditional probabilities

- it is as if B were now the sample space and hence all further occurrences are calibrated with respect to their relation to B
- **remark**: if A and B are mutually exclusive, $A \cap B = \emptyset$ and hence $\mathbb{P}(A \cap B) = \mathbb{P}(A|B) = \mathbb{P}(B|A) = 0$
- for any B such that $\mathbb{P}(B) > 0$, the conditional probability $\mathbb{P}(\cdot|B)$ meets Kolmogorov's axioms
  - (may redefine sample space to B)
- **however...** conditional probabilities are slipperier than you think!

the death row puzzle

- **background story**: governor pardons one of three prisoners (A, B or C) from the death row at random and then informs warden.
- prisioner A asks the warden if B or C are to be executed. Warden reveals to A that B is to be executed
- **warden's thinking**: each prisoner has a 1/3 chance of pardon and, obviously, either B or C, if not both, must be executed and hence I have revealed nothing
- **A's reasoning**: given that B will be executed, either me or C will be pardoned and hence my chance has risen to 1/2

more on the death row puzzle

- **whose reasoning is correct?**
- let A, B and C denote the events that A, B, or C is pardoned. Then $P(A) = P(B) = P(C) = 1/3$.
- let b denote the event that the warden says that B will be executed.
- we must compute the conditional probability $\mathbb{P}(A|b)$ of A being pardoned given that the warden has revealed that B will be executed

prisoner pardoned	warden tells A	
A	B dies }	each with equal probability
A	C dies }	
B	C dies	
C	B dies	

still a bit more on the death row puzzle

$$\begin{aligned}\mathbb{P}(b) &= \mathbb{P}(b \cap A) + \mathbb{P}(b \cap B) + \mathbb{P}(b \cap C) \\ &= 1/3 \times 1/2 + 0 + 1/3 \times 1 \\ &= 1/6 + 0 + 1/3 = 1/2\end{aligned}$$

- **warden correctly calculates...**

$$\mathbb{P}(A|b) = \frac{\mathbb{P}(b \cap A)}{\mathbb{P}(b)} = \frac{1/6}{1/2} = 1/3$$

- **A instead computes...**

$$\mathbb{P}(A|B^c) = \frac{\mathbb{P}(B^c \cap A)}{\mathbb{P}(B^c)} = \frac{1/3}{2/3} = 1/2$$

- A's mistake: falsely interpreting the event b as the event B^c.
- See a more detailed explanation here.

Bayes' rule

- reexpressing the conditional probability formula provides a useful means to compute an intersection probability

$$\mathbb{P}(A \cap B) \;=\; \mathbb{P}(A|B)\mathbb{P}(B) \;=\; \mathbb{P}(B|A)\mathbb{P}(A)$$

- **Bayes' rule** to turn around conditional probabilities

$$\mathbb{P}(A|B) \;=\; \mathbb{P}(B|A)\,\frac{\mathbb{P}(A)}{\mathbb{P}(B)}$$

- if $A_1, A_2, \ldots$ is a partition of the sample space, then

$$\mathbb{P}(A_i|B) \;=\; \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\mathbb{P}(B)} \;=\; \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{\infty}\mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

Bayes' rule example

- example: Morse code uses "dots" and "dashes" to transmit messages, but often there are errors in transmission. Suppose that dots are sent with probability $\frac{3}{7}$ and dashes are sent with probability $\frac{4}{7}$, and with probability $\frac{1}{8}$ a dot is received as a dash, and vice-versa.
- If we observe a dot, what is the probability that a dot was sent?

$$\begin{aligned}\mathbb{P}(\text{dot sent}|\text{dot received}) &= \frac{\mathbb{P}(\text{dot sent} \cap \text{dot received})}{\mathbb{P}(\text{dot received})} \\ &= \mathbb{P}(\text{dot received}|\text{dot sent})\frac{\mathbb{P}(\text{dot sent})}{\mathbb{P}(\text{dot received})}\end{aligned}$$

and we know that $\mathbb{P}(\text{dot received}|\text{dot sent}) = 1 - \frac{1}{8} = \frac{7}{8}$ and $\mathbb{P}(\text{dot sent}) = \frac{3}{7}$. We now need to figure $\mathbb{P}(\text{dot received})$ out.

$$\begin{aligned}\mathbb{P}(\text{dot received}) &= \mathbb{P}(\text{dot received} \cap \text{dot sent}) + \mathbb{P}(\text{dot received} \cap \text{dash sent}) \\ &= \mathbb{P}(\text{dot received}|\text{dot sent})\mathbb{P}(\text{dot sent}) \\ &\quad +\mathbb{P}(\text{dot received}|\text{dash sent})\mathbb{P}(\text{dash sent}) \\ &= \frac{7}{8} \times \frac{3}{7} + \frac{1}{8} \times \frac{4}{7} = \frac{25}{56}\end{aligned}$$

- So $\mathbb{P}(\text{dot sent}|\text{dot received}) = \frac{21}{25}$.

statistical independence

- the occurrence of a particular event B sometimes brings about no information about another event A, and hence $\mathbb{P}(A|B) = \mathbb{P}(A)$

- Bayes' rule then yields

$$\mathbb{P}(B|A) \;=\; \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)} \;=\; \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(A)} \;=\; \mathbb{P}(B)$$

- **definition**: two events A and B are statistically independent if

$$\mathbb{P}(A \cap B) \;=\; \mathbb{P}(A)\mathbb{P}(B)$$

statistical independence

- **proposition:** if A and B are independent events, then the following pairs are also independent:
  (i) A and B^c;
  (ii) A^c and B;
  (iii) A^c and B^c.

- **proof:** we will show (i) only. We must show that $\mathbb{P}(A \cap B^c) = \mathbb{P}(A)\mathbb{P}(B^c)$. So

$$\begin{aligned}\mathbb{P}(A \cap B^c) &= \mathbb{P}(A) - \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) \\ &= \mathbb{P}(A)(1 - \mathbb{P}(B)) \\ &= \mathbb{P}(A)\mathbb{P}(B^c)\end{aligned}$$

$\blacksquare$

- **question:** if A and B are disjoint events, are they independent? (no, unless one is impossible)

independence between multiple events

- one of the reasons to employ the independence definition based on the intersection probability is to facilitate the extension to multiple events, **though...** we must be careful

- at first glance, it suffices to say that A, B, and C are independent if $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$

- **counterexample**: tossing two dice

- alternatively, we may think of defining independence between multiple events in terms of pairwise independence

- **counterexample**: permutations of letters

tossing two dice

- **sample space** $S = \{(1,1), \dots, (1,6), \dots, (6,1), \dots, (6,6)\}$
- **events**:
  - $A = \{(1,1), (2,2), \dots, (6,6)\}$ $\mathbb{P}(A) = 1/6$
  - $B = \{\text{sum between 7 and 10}\}$ $\mathbb{P}(B) = 1/2$
  - $C = \{\text{sum is 2 or 7 or 8}\}$ $\mathbb{P}(C) = 1/3$

$$\begin{aligned}\mathbb{P}(A \cap B \cap C) &= \mathbb{P}(\text{sum is 8, composed of doubles}) \\ &= 1/36 = 1/6 \times 1/2 \times 1/3 = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)\end{aligned}$$

- **however**, $\mathbb{P}(B \cap C) \neq \mathbb{P}(B)\mathbb{P}(C)$, for example, and hence $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ is not strong enough to ensure pairwise independence

letter permutation

- **equiprobable sample space**: 3! permutations of (a, b, c) along with triplets of each letter

$$S = \begin{Bmatrix} aaa & bbb & ccc \\ abc & bca & cba \\ acb & bac & cab \end{Bmatrix}$$

- **events**: $A_i = \{i\text{th place of the triplet is } a\}$ $\mathbb{P}(A_i) = 1/3$
  - $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_2 \cap A_3) = 1/9$, satisfying pairwise independence
  - $\mathbb{P}(A_1 \cap A_2 \cap A_3) = 1/9 \neq 1/27 = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)$, violating probability requirement

mutual independence

- **definition**: $A_1, \ldots, A_n$ are mutually independent if

$$\mathbb{P}\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j})$$

  for any subcollection $A_{i_1}, \ldots, A_{i_k}$

- **example**: three coin tosses
  - $S = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$
  - $H_i = \{i\text{th toss is } H\}$, $\mathbb{P}(H_i) = 1/2$

$$\mathbb{P}(H_1 \cap H_2) = \mathbb{P}(\{HHH\}, \{HHT\}) = 1/4 = \mathbb{P}(H_1)\mathbb{P}(H_2)$$
$$\mathbb{P}(H_1 \cap H_2 \cap H_3) = \mathbb{P}(\{HHH\}) = 1/8 = \mathbb{P}(H_1)\mathbb{P}(H_2)\mathbb{P}(H_3)$$

Contents

making life easier

- in many experiments, it is easier to deal with a summary variable than keeping track of the original probability structure

- **example**: yes-or-no referendum in a class of 50 students
  - sample space has 2^{50} ordered strings of 1s and 0s
  - X = number of 1s, hopefully captures the essence of the the matter, though... with a much smaller sample space

$$S_X = \{0, 1, \ldots, 50\}$$

- **definition**: a **random variable** $X : S \to E$ is a function mapping the sample space S into a measurable space E.
  - Understood that Ω is part of the probability space $(S, \mathcal{S}, P)$.
  - E is usually $\mathbb{R}$.

examples

experiment	random variable
toss two dice	$X =$ sum of the rolls
toss a coin 25 times	$X =$ number of heads
amount of fertilizer	$X =$ yield/acre

devil is in details

- we must also define a new sample space and check how should we proceed with the probability function **from** $\mathbb{P}(\cdot)$ on $\Omega = \{\omega_1, \ldots, \omega_n\}$ **to** $\mathbb{P}_X(\cdot)$ on $\Omega_X = \{x_1, \ldots, x_m\}$

- **device**: we observe $X = x_i$ if and only if the outcome of the random experiment is an $\omega_j \in \Omega$ such that $X(\omega_j) = x_i$

$$\mathbb{P}_X(X = x_i) \quad = \quad \mathbb{P}(\{\omega_j \in \Omega : \; X(\omega_j) = x_i\})$$

- Note that **random variables** are denoted with uppercase letters (X) and realized values or ranges are denoted with lowercase letters (x).

Kolmogorov is happy with induced probabilities

- (exercise 1.45) Show that the induced probability function satisfies the Kolmogorov axioms.
- proof: Let $\mathcal{X} = \{x_1, \dots, x_m\}$ be the range of the random variable X. $\mathcal{X}$ is finite, so define the set of all subsets of $\mathcal{X}$ as $\mathcal{B}$, a σ-algebra.

  (i) If $A \in \mathcal{B}$, then $\mathbb{P}_X(A) = \mathbb{P}\left(\cup_{x_i \in A}\{\omega_j \in \Omega : X(\omega_j) = x_i\}\right) \geq 0$ since $\mathbb{P}$ is a probability function.

  (ii) $\mathbb{P}_X(\mathcal{X}) = \mathbb{P}\left(\cup_{i=1}^m \{\omega_j \in \Omega : X(\omega_j) = x_i\}\right) = \mathbb{P}(\Omega) = 1.$

  (iii) If $A_1, A_2, \dots \in \mathcal{B}$ are pairwise disjoint then

$$\begin{aligned}\mathbb{P}_X\left(\bigcup_{k=1}^{\infty} A_k\right) &= \mathbb{P}\left(\bigcup_{k=1}^{\infty}\{\cup_{x_i \in A_k}\{\omega_j \in \Omega : X(\omega_j) = x_i\}\}\right) \\ &= \sum_{k=1}^{\infty} \mathbb{P}\left(\cup_{x_i \in A_k}\{\omega_j \in \Omega : X(\omega_j) = x_i\}\right) = \sum_{k=1}^{\infty} \mathbb{P}_X(A_k)\end{aligned}$$

- From now on, we can drop the distinction between P_X and P!

revisiting some examples

- **three coin tosses**

s	HHH	HHT	HTH	THH	TTH	THT	HTT	TTT
$X(s)$	3	2	2	2	1	1	1	0
$\mathbb{P}_X$	1/8		3/8			3/8		1/8

- **opinion poll**: recall that $S_X = \{0, 1, \ldots, 50\}$ and assume that each of the 2^{50} strings is equally likely, then the **induced probability** reads

$$\Pr_X(X = x) \quad = \quad \frac{\binom{50}{x}}{2^{50}}$$

(un)countable sample spaces

- previous illustrations have both finite S and finite S_X, for which the definition of $\mathbb{P}_X$ is straightforward
- **countable sample spaces**: still straightforward!
- **uncountable sample spaces**: we define the induced probability in a similar fashion. For any set $A \subset S_X$,

$$\mathbb{P}_X(X \in A) \quad = \quad \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\})$$

defines a legitimate probability function for which Kolmogorov's axioms hold.

support of a random variable

- **definition**: the support of a random variable X is the smallest closed set $R_X \subseteq \mathcal{B}$ such that $P_X(R_X) = 1$.
- **alternative definition in $\mathbb{R}^n$**: The support of a random variable X with values in $\mathbb{R}^n$ is the set $\{x \in \mathbb{R}^n | P_X(B(x, r)) > 0, \text{for all } r > 0\}$.
  - $B(x, r)$ denotes the ball with center at x and radius r.
- For discrete random variables, $R_X \equiv \{x \in \mathbb{R}^n | P(X = x) > 0\}$.
  - R_X is also countable.

Contents

- **definition**: we define the cumulative distribution function $F_X(x)$ of a random variable X as

$$F_X(x) \;=\; \mathbb{P}(X \leq x)$$

  for every $x \in \Omega_X$

- **example**: three coin tosses, $X =$ number of heads

$$F_X(x) \;=\; \begin{cases} 0 & \text{if} \quad -\infty < x < 0 \\ 1/8 & \text{if} \quad 0 \leq x < 1 \\ 1/2 & \text{if} \quad 1 \leq x < 2 \\ 7/8 & \text{if} \quad 2 \leq x < 3 \\ 1 & \text{if} \quad 3 \leq x < \infty \end{cases}$$

- $F(x)$ is a cdf if and only if

  (*i*) $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$

  (*ii*) $F_X(x)$ is a nondecreasing function of x

  (*iii*) $F_X(x)$ is right-continuous, $\lim_{x\downarrow x_0} F_X(x) = F_X(x_0)$

- Right-continuity is a consequence of the definition of the cdf. If we had defined $F(x) = P_X(X < x)$, then F would have been left-continuous.

example. . . tossing for a head

- the experiment consists of tossing a coin until a head appears
- X = number of tosses until first head
- p is the probability of a head on any given toss

$$
\begin{aligned}
\mathbb{P}(X = x) &= (1-p)^{x-1} p \\
&\Updownarrow \\
\mathbb{P}(X \leq x) &= \sum_{i=1}^{x} \mathbb{P}(X = i) \\
&= \sum_{i=1}^{x} (1-p)^{i-1} p \\
&= \frac{1-(1-p)^x}{1-(1-p)} p \\
&= 1-(1-p)^x
\end{aligned}
$$

for $x = 1, 2, \ldots$.

geometric distribution

$$F_X(x) \;=\; \mathbb{P}(X \leq x) \;=\; 1-(1-p)^x \qquad \begin{array}{l} 0 < p < 1 \\ x = 1, 2, \ldots \end{array}$$

- **let's check whether it indeed is a distribution**
  - $\lim_{x \to -\infty} F_X(x) = 0$ given that $F_X(x) = 0$ if $x < 0$
  - $\lim_{x \to \infty} F_X(x) = 1$ given that $\lim_{x \to \infty}(1-p)^x = 0$
  - $F_X(x)$ is nondecreasing given that we keep adding positive terms
  - $F_X(x)$ is right-continuous given that $\lim_{\epsilon \downarrow 0} F_X(x+\epsilon) = F_X(x)$
  - $F_X(x)$ is flat between nonnegative integers (discrete distribution)

continuous vs discrete distributions

- **definition**: a random variable is discrete if $F_X(x)$ is a step function of x, whereas it is continuous if $F_X(x)$ is a continuous function of x
- **example**: logistic distribution: $F_X(x) = \frac{1}{(1+e^{-x})}$
  - $\lim_{x\to-\infty} F_X(x) = 0$ given that $\lim_{x\to-\infty} e^{-x} = \infty$
  - $\lim_{x\to\infty} F_X(x) = 1$ given that $\lim_{x\to\infty} e^{-x} = 0$
  - $F_X(x)$ is nondecreasing given that $F'_X(x) = \frac{e^{-x}}{(1+e^{-x})^2} > 0$
  - $F_X(x)$ is not only right-continuous, but also continuous

identically distributed

- the nice thing about distribution functions is that they completely determine the probability structure of a random variable if $\mathbb{P}(\cdot)$ is defined only for events in the Borel σ-algebra on $\mathbb{R}$, defined as $\mathcal{B}$.

- **definition**: the random variables X and Y are identically distributed if, for every set $A \in \mathcal{B}$, $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$

- **equivalence theorem**: X and Y are identically distributed if and only if $F_X(x) = F_Y(x)$ for every $x \in \mathbb{R}$

- **proof:** $(\Rightarrow)$ Because X and Y are identically distributed, for any set $A \in \mathcal{B}$, $P(X \in A) = P(Y \in A)$. In particular, for every x, the set $(-\infty, x]$ is in $\mathcal{B}$, and

$$F_X(x) \;=\; \mathbb{P}(X \in (-\infty, x]) \;=\; \mathbb{P}(Y \in (-\infty, x]) = \; F_Y(x).$$

Contents

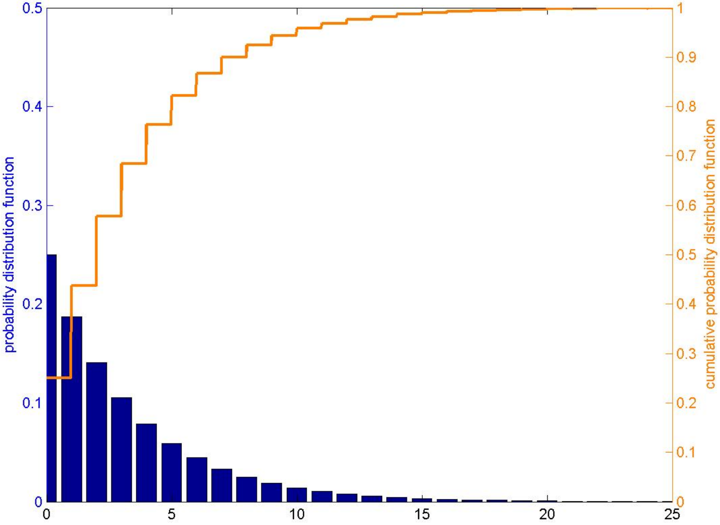
point probability

- **definition**: we define the probability mass function of a discrete random variable X as $f_X(x) = \mathbb{P}(X = x)$

- **example**: geometric distribution

$$
\begin{aligned}
f_X(x) = \mathbb{P}(X = x) &= \begin{cases} (1-p)^{x-1}p & \text{for } x = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases} \\
\mathbb{P}(a \leq X \leq b) &= \sum_{x=a}^{b} f_X(x) = \sum_{x=a}^{b} (1-p)^{x-1} p \\
\mathbb{P}(X \leq b) &= \sum_{x=1}^{b} f_X(x) = F_X(b)
\end{aligned}
$$

looks of the geometric distribution



how about continuous variables?

- **naïve calculation**: $\{X = x\} \subset [x - \epsilon \leq X \leq x]$ for any $\epsilon > 0$

$$\begin{aligned} \mathbb{P}(X = x) &\leq \mathbb{P}(x - \epsilon \leq X \leq x) = F_X(x) - F_X(x - \epsilon) \\ &\Rightarrow 0 \leq \mathbb{P}(X = x) \leq \lim_{\epsilon \downarrow 0}[F_X(x) - F_X(x - \epsilon)] = 0 \end{aligned}$$

  by the continuity of F_X.

- **definition**: we implicitly define the probability density function of a continuous random variable X as

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,\mathrm{d}u \text{ for all } u \in \mathbb{R}$$

- $\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b)$ given that $\mathbb{P}(X = x) = 0$

- moreover, the **fundamental theorem of calculus** ensures that, if $f_X(x)$ is continuous, then $f_X(x) = F'_X(x)$

requirements

- $f(x)$ is a pdf/pmf if and only if
  - (i) $f(x) \geq 0$ for all $x \in \mathbb{R}$
  - (ii) $\int_{-\infty}^{\infty} f(u)\,\mathrm{d}u = 1$
  - **proof**: if $f(x)$ is a pdf (or pmf), (i)+(ii) follow immediately from definition; in particular

$$1 = \lim_{x\to\infty} F(x) = \int_{-\infty}^{\infty} f(u)\,\mathrm{d}u$$

- the converse implication is equally easy to prove given that we may define $F(x)$ from $f(x)$ and then verify it indeed is a cdf

Contents

Reference:

- Casella and Berger, Ch. 1

Exercises:

- 1.4, 1.5, 1.13, 1.17, 1.18, 1.20, 1.21, 1.24, 1.33, 1.34, 1.36, 1.38, 1.41, 1.47, 1.51, 1.52, 1.55.