Introduction to Econometrics
Recitation 1

Bruno Furtado

September 14, 2020

1 Interpretations of Probability

What is the meaning of the probability of an event occurring? There are two main interpretations:

1. **Frequentist approach:** probability is the relative frequency of occurrence of the outcome when an experiment is repeated a "large" number of times.

   Example 1.1: Proportion of HEADS when flipping a fair coin n times converges to $1/2$ as $n \to \infty$.

   Problem: Some (most) experiments that we want to assign probabilities to are not repeatable, such as the outcome of an election or the effects of a monetary policy shock.

2. **Subjective (Bayesian) approach:** probability of an outcome reflects the degree of confidence that it will occur.

   Example 1.2: Assigning odds to different teams winning a championship game.

   Problem: there might not be agreement on what probability to assign to different events.

The mathematical theory of probability is built upon measure theory and based on certain axioms that are consistent with both interpretations above. This allows us to construct a useful theory without the need to commit to one or another particular interpretation.

2 Sample Spaces and Events

Let Ω be an arbitrary set. We call Ω the sample space, and it includes every possible outcome of the random experiment, denoted by $\omega \in \Omega$.

Example 2.1:

(a) Tossing a coin: $\Omega = \{H, T\}$;

(b) Rolling two 6-sided die: $\Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$;

(c) Drawing a point from the interval $[a, b]$: $\Omega = \{\omega : a \leq \omega \leq b\}$. This is actually also a good model to describe the outcome of an infinite sequence of coin tosses (see Billingsley (1995, section 1)).

Each $A \subseteq \Omega$ is called an event. These are the objects we want to be able to assign probabilities to, with the interpretation that the probability of $A \subseteq \Omega$ is the likelihood that any of the outcomes $\omega \in A$ takes place.

Example 2.2:

(a) Coin toss: $\{H\}, \{T\}, \{H, T\}, \emptyset$.

(b) Rolling two 6-sided die: $\{(1,6)\}, \{(6,6),(1,1)\}, \{(n,m) \in \{1,2,3,4,5,6\}^2 : n+m = 7\}, \ldots$

(c) Randomly choosing a point from $[0,1]$: $\{1/3\}, [1/3, 1/2], (2/3, 1], \ldots$

We want to define probabilities for as many subsets of Ω as possible. Unfortunately, it is not always possible to define a probability on 2^Ω. [1] Let $\mathcal{F} \subseteq 2^\Omega$ be a class of subsets of Ω. In order for it to be possible to assign probabilities to all $A \subseteq \mathcal{F}$, we need $\mathcal{F}$ to be a σ-algebra (also sometimes called a σ-field).

Definition (σ-algebra). A class $\mathcal{F}$ of subsets of Ω is called a σ-algebra if

1. $\Omega \in \mathcal{F}$.
2. $A \in \mathcal{F} \implies A^c \in \mathcal{F}$.
3. $A_1, A_2, \ldots \in \mathcal{F} \implies \bigcup_{n=1}^\infty A_n \in \mathcal{F}$.

It follows directly from the definition that:

1′. $\emptyset \in \mathcal{F}$, since $\emptyset = \Omega^c$.

3′. $A_1, A_2, \ldots \in \mathcal{F} \implies \bigcap_{n=1}^\infty A_n \in \mathcal{F}$, since, applying properties 2 and 3, $\bigcap_{n=1}^\infty A_n = \left(\bigcup_{n=1}^\infty A_n^c\right)^c \in \mathcal{F}$.

Example 2.3:

(a) $\{\emptyset, \Omega\}$ is always a σ-algebra.

(b) 2^Ω is always a σ-algebra.

(c) Take any $A \subset \Omega$. Then $\{A, A^c, \Omega, \emptyset\}$ is a σ-algebra.

Sometimes, we want to make sure that it is possible to assign probabilities to a specific collection of subsets $\mathcal{C} \subseteq 2^\Omega$ (e.g., all intervals of $\mathbb{R}$, all countable subsets of some sample space Ω etc.) However, $\mathcal{C}$ may not be a σ-algebra, in which case we want to define the smallest σ-algebra that contains it. This is called the σ-algebra *generated* by $\mathcal{C}$.

Definition (Generated σ-algebra). Let $\mathcal{C}$ be a class of subsets of Ω. The σ-algebra generated by $\mathcal{C}$ is the smallest σ-algebra (with the $\subseteq$ order) which contains $\mathcal{C}$, that is:

$$\sigma(\mathcal{C}) = \bigcap \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-algebra and } \mathcal{C} \subseteq \mathcal{F}\}.$$

To realize that this is a well-defined mathematical object, notice that since 2^Ω is a σ-algebra, there always exists at least one σ-algebra containing $\mathcal{C}$. Besides, the arbitrary intersection of σ-algebras is a σ-algebra.

In fact, let $\{\mathcal{F}_\theta : \theta \in \Theta\}$ be an arbitrary collection of σ-fields, where Θ is an arbitrary set of indices. Define:

$$\mathcal{F} = \bigcap_{\theta \in \Theta} \mathcal{F}_\theta$$

We want to show that $\mathcal{F}$ satisfies properties 1-3. First, for all $\theta \in \Theta$ we have $\Omega \in \mathcal{F}_\theta$, thus $\Omega \in \bigcap_{\theta\in\Theta} \mathcal{F}_\theta = \mathcal{F}$. Next, take any $A \in \mathcal{F}$. Then by definition $A \in \mathcal{F}_\theta$, $\forall \theta \in \Theta$. Since $\mathcal{F}_\theta$ are σ-fields, $A^c \in \mathcal{F}_\theta$ for every $\theta \in \Theta$. Therefore, $A^c \in \bigcap_{\theta\in\Theta} \mathcal{F}_\theta = \mathcal{F}$. Finally, let $\{A_n\}_{n\geq 1}$ be such that $A_n \in \mathcal{F}$, $\forall n \geq 1$. Then for all $\theta \in \Theta$ and $n \geq 1$, we have $A_n \in \mathcal{F}_\theta$, which implies that $\bigcup_{n=1}^\infty A_n \in \mathcal{F}_\theta$ since each $\mathcal{F}_\theta$ is a σ-algebra. Therefore, $\bigcup_{n=1}^\infty A_n \in \bigcap_{\theta\in\Theta} \mathcal{F}_\theta = \mathcal{F}$. This shows that $\mathcal{F}$ is a σ-field.

Definition (Borel σ-algebra). Let (Ω, τ) be a topological space, where $\tau = \{I \subseteq \Omega : I \text{ is open}\}$. The Borel σ-algebra for Ω, $\mathcal{B}(\Omega)$, is defined as the smallest σ-algebra containing all the open subsets of Ω:

$$\mathcal{B}(\Omega) = \bigcap \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-algebra and } \tau \subseteq \mathcal{F}\}.$$

In other words,

$$\mathcal{B}(\Omega) = \sigma(\tau)$$

Example 2.4: Let $\Omega = \mathbb{R}$ and $\mathcal{I}$ be the set of all open intervals of $\mathbb{R}$, i.e., $\mathcal{I} = \{(a, b) : -\infty \leq a < b \leq \infty\}$. Then, it turns out that $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{I})$.

[1] The symbol 2^Ω denotes the *power set* of Ω, that is, the set of all subsets of Ω.

3 Probability Measures

In probability theory, we view a probability measure as a particular case of the mathematical concept of a measure.

Definition (Measure). Let Ω be an arbitrary set and $\mathcal{F} \subseteq 2^\Omega$ be a σ-algebra. A *measure* μ is a function $\mu : \mathcal{F} \to \mathbb{R}$ satisfying the following properties:

1. $\mu(A) \geq 0$ for all $A \in \mathcal{F}$;
2. $\mu(\emptyset) = 0$;
3. If the sequence $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{F}$ is such that $A_i \cap A_j = \emptyset$ for all $i \neq j$, then

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

Example 3.1 (Lebesgue Measure (ℓ)): Let $\Omega = \mathbb{R}$ and define the collection of all finite unions of intervals:

$$\mathcal{F}_0 = \left\{ \bigcup_{j=1}^{n} I_j \in \mathbb{R} : I_j = (a_j, b_j],\ -\infty \leq a_j < b_j \leq \infty,\ I_k \cap I_j = \emptyset\ \forall j \neq k,\ n \in \mathbb{N} \right\}$$

Unfortunately, $\mathcal{F}_0$ is not a σ-algebra, but we can still define a "pre-measure" function $\lambda : \mathcal{F}_0 \to \mathbb{R}$ by

$$\lambda(A) = \sum_{j=1}^{n} (b_j - a_j)$$

for all $A = \bigcup_{j=1}^{n} (a_j, b_j] \in \mathcal{F}_0$. This function coincides with our intuitive idea of how to measure lengths (neat!) Can we define a measure on the larger collection $\mathcal{F} \equiv \sigma(\mathcal{F}_0)$ – the σ-algebra generated by $\mathcal{F}_0$ – that coincides with this intuitive definition whenever we are just trying to measure a union of intervals? Yes! By Carathéodory's Extension Theorem there exists a *unique* measure $\ell : \mathcal{F} \to \mathbb{R}$ such that $\ell(A) = \lambda(A)$ for all $A \in \mathcal{F}_0$. We call ℓ the Lebesgue measure.

If $\mu(\Omega) = M < \infty$, we call μ a *finite measure*. If it is possible to cover the whole sample space Ω with a countable collection of sets of finite μ measure, then we say that μ is σ*-finite*.

Definition (σ-finite Measure). A measure $\mu : \mathcal{F} \to \mathbb{R}$ is called σ-finite if there exists a countable collection $A_1, A_2, \ldots \in \mathcal{F}$ such that $\Omega = \bigcup_{i=1}^{\infty} A_i$ and $\mu(A_i) < \infty$ for all $i \in \mathbb{N}$.

Remark: Notice that a measure being σ-finite does not imply that it is finite. In particular, the Lebesgue measure ℓ on $\mathbb{R}$ is σ-finite but not finite. Let $J_n = (-n-1, -n] \cup (n, n+1] \in \mathcal{B}(\mathbb{R})$. Then $\mathbb{R} = \bigcup_{n=0}^{\infty} J_n$ and $\ell(J_n) = 2$ for all $n \in \mathbb{N}$. However $\ell(\mathbb{R}) = \infty$.

The following theorem, which states that we can uniquely extend a measure from an algebra to a σ-algebra, is one of the most important in measure theory, so it is worth mentioning, if only for curiosity's sake.

Theorem 3.1 (Carathéodory's Extension Theorem) *Let $m : \mathcal{F}_0 \to \mathbb{R}$ be a measure on a algebra $\mathcal{F}_0$ of subsets of Ω. If m is σ-finite, there exists a unique extension of m to the sigma algebra generated by $\mathcal{F}_0$, that is, there exists a unique $\mu : \sigma(\mathcal{F}_0) \to \mathbb{R}$ such that $\mu(A) = m(A)$ for all $A \in \mathcal{F}_0$.*[2]

Proof: *For a complete proof, see Ash and Doléans-Dade (2000, chapter 1).* ∎

As illustrated by example 3.1, the Caratheodory Extension Theorem is quite useful when defining measures on σ-algebras.

We now turn our attention to a specific class of finite measures, called probability measures, for which the measure of the whole sample space is normalized to 1.

[2]In fact, we only need to start with a measure on a semi-algebra, not an algebra, to get a unique extension to the generated σ-algebra. I'll not provide a formal definition of a semi-algebra, but the collection of all intervals in $\mathbb{R}$ is an example. Check out, for instance, Ok (2018, chapter 2).

Definition (Probability Measure). Let Ω be an arbitrary set and $\mathcal{F} \subseteq 2^\Omega$ be a σ-algebra. If $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ is a *measure* and $\mathbb{P}(\Omega) = 1$, then it is called a *probability measure*.

A *measure space* is a tuple $(\Omega, \mathcal{F}, \mu)$ such that $\mathcal{F}$ is a σ-algebra on Ω and $\mu : \mathcal{F} \to \mathbb{R}$. If $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ is a probability measure, we call $(\Omega, \mathcal{F}, \mathbb{P})$ a *probability space*.

Proposition 3.1 (Basic Properties of Probability) *Let $\mathbb{P}$ be a probability measure on a σ-algebra $\mathcal{F}$. Then the following properties hold:*

(1) $A, B \in \mathcal{F} \implies \mathbb{P}(B) = \mathbb{P}(A \cap B) + P(A^c \cap B)$;

(2) $A \in \mathcal{F} \implies \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$;

(3) ***Monotonicity:*** $A, B \in \mathcal{F}$ *and* $A \subseteq B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$;

(4) $0 \leq \mathbb{P}(A) \leq 1 \ \forall A \in \mathcal{F}$;

(5) ***Inclusion-exclusion formula:*** *For all* $A_1, \dots, A_n \in \mathcal{F}$,

$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} \mathbb{P}(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n+1} \mathbb{P}(A_1 \cap \cdots \cap A_n)$$

(6) ***Lower continuity:*** *if* $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{F}$ *with* $A_n \subseteq A_{n+1} \ \forall n \geq 1$, *and* $\lim_{N \to \infty} \bigcup_{n=1}^{N} A_n = A \in \mathcal{F}$, *then*

$$\lim_{n \to \infty} \mathbb{P}(A_n) = \mathbb{P}(A)$$

(7) ***Upper continuity:*** *if* $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{F}$ *with* $A_{n+1} \subseteq A_n \ \forall n \geq 1$, *and* $\lim_{N \to \infty} \bigcap_{n=1}^{N} A_n = A \in \mathcal{F}$, *then*

$$\lim_{n \to \infty} \mathbb{P}(A_n) = \mathbb{P}(A)$$

(8) ***Boole's inequality:*** *If* $A_n \in \mathcal{F}$, $\forall n \geq 1$, *and* $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$, *then:*

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

Proof:

(1) *Notice that* $B = (B \cap A) \cup (B \cap A^c)$, *and* $(B \cap A)$ *and* $(B \cap A^c)$ *are disjoint.*

(2) *Since* $\Omega = A \cup A^c$ *and* $A \in \mathcal{F}$,

$$\begin{aligned} &\mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A^c) \\ \iff\ &1 = \mathbb{P}(A) + \mathbb{P}(A^c) \\ \iff\ &\mathbb{P}(A^c) = 1 - \mathbb{P}(A) \end{aligned}$$

(3) *Exercise*

(4) *Follows from monotonicity and* $\emptyset \subseteq A \subseteq \Omega$ *for all* $A \in \mathcal{F}$.

(5) *We will proceed by induction. For* $n = 2$, *you can prove it as an exercise that* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

By the induction hypothesis, we assume the formula holds for $n-1$. Then, for n,

$$\begin{aligned}
\mathbb{P}\left[\left(\bigcup_{i=1}^{n-1} A_i\right) \cup A_n\right] =& \mathbb{P}\left(\bigcup_{i=1}^{n-1} A_i\right) + \mathbb{P}(A_n) - \mathbb{P}\left[\left(\bigcup_{i=1}^{n-1} A_i\right) \cap A_n\right] \\
=& \sum_{i=1}^{n-1} \mathbb{P}(A_i) - \sum_{1\le i<j\le n-1} \mathbb{P}(A_i \cap A_j) + \sum_{1\le i<j<k\le n-1} \mathbb{P}(A_i \cap A_j \cap A_k) \\
& + \cdots + (-1)^n \mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) + \mathbb{P}(A_n) - \mathbb{P}\left[\left(\bigcup_{i=1}^{n-1} A_i\right) \cap A_n\right] \\
=& \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1\le i<j\le n-1} \mathbb{P}(A_i \cap A_j) + \sum_{1\le i<j<k\le n-1} \mathbb{P}(A_i \cap A_j \cap A_k) \\
& + \cdots + (-1)^n \mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) - \mathbb{P}\left[\bigcup_{i=1}^{n-1} (A_i \cap A_n)\right] \qquad (*)
\end{aligned}$$

Now we again apply the induction hypothesis, this time to the last term in $(*)$, obtaining

$$\begin{aligned}
\mathbb{P}\left[\bigcup_{i=1}^{n-1} (A_i \cap A_n)\right] =& \sum_{i=1}^{n-1} \mathbb{P}(A_i \cap A_n) - \sum_{1\le i<j\le n-1} \mathbb{P}(A_i \cap A_j \cap A_n) \\
& + \sum_{1\le i<j<k\le n-1} \mathbb{P}(A_i \cap A_j \cap A_k \cap A_n) + \cdots + (-1)^n \mathbb{P}(A_1 \cap \cdots \cap A_n)
\end{aligned}$$

Substituting the expression above into $(*)$ and rewriting the sums, we get

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) =& \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1\le i<j\le n} \mathbb{P}(A_i \cap A_j) + \sum_{1\le i<j<k\le n} \mathbb{P}(A_i \cap A_j \cap A_k) \\
& + \cdots + (-1)^{n+1} \mathbb{P}(A_1 \cap \cdots \cap A_n)
\end{aligned}$$

(6) *Define $B_1 = A_1$ and $B_i = A_i \setminus A_{i-1}$, $\forall i \ge 2$. Then, $\forall k \ne l$, $B_k \cap B_l = \emptyset$. In fact, if $l > k$, then $B_l = A_l \setminus A_{l-1}$ and $B_k = A_k \setminus A_{k-1} \subseteq A_{l-1}$. Also notice that $A_n = \bigcup_{i=1}^n B_i$ $\forall n \ge 1$ and $A = \bigcup_{i=1}^\infty B_i$. Therefore, by σ-additivity,*

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(7) *Let $A = \bigcap_{n=1}^\infty A_n = \left(\bigcup_{n=1}^\infty A_n^c\right)^c$. Since $A \in \mathcal{F}$, we have $A^c = \bigcup_{n=1}^\infty A_n^c \in \mathcal{F}$. By part (6), $\mathbb{P}(A_n^c) \xrightarrow{n\to\infty} \mathbb{P}(A^c)$. Thus,*

$$1 - \mathbb{P}(A_n) \xrightarrow{n\to\infty} 1 - \mathbb{P}(A)$$

Therefore,

$$\mathbb{P}(A_n) \xrightarrow{n\to\infty} \mathbb{P}(A)$$

(8) *Exercise.* ∎

References

Ash, R. B. and C. Doléans-Dade (2000). *Probability and measure theory.*

Billingsley, P. (1995). *Probability and Measure - Third Edition.*

Ok, E. A. (2018). *Probability Theory with Economic Applications* (Manuscript ed.).