Probability spaces and measures

Robert J. Carroll

Last revised: 5 May 2026

---

1 Motivation

Almost every formal model in political economy involves uncertainty somewhere. A voter has a noisy signal about a candidate's quality. A politician faces an unknown state of the world she is trying to learn. An electoral demand function carries a stochastic shock. A pivotal-voter analysis aggregates over many independent vote realizations. A Bayesian game's equilibrium requires posterior beliefs over types. To talk about uncertainty rigorously — to compute, prove, or even consistently state claims about probabilities — you need a *probability space*, and the apparatus of measure theory is what makes probability spaces work cleanly when the state space is uncountable (a continuum of voter types, a real-valued shock, a continuous policy space).

The cardinality handout left a pointer here. There we noted that working with *all* subsets of an uncountable space is not a sensible thing to want: $\mathcal{P}(\mathbb{R})$ is too rich to admit a translation-invariant countably additive measure (*Vitali sets*, the *Banach–Tarski* paradox in $\mathbb{R}^3$, both products of the axiom of choice acting on the gap between $|\mathbb{R}|$ and $|\mathcal{P}(\mathbb{R})|$). The fix is to work with a chosen *sub-collection* of subsets — a *σ-algebra* — that is rich enough to contain every set we care about and well-behaved enough that the standard set operations stay inside it. A *measure* is then a function from the σ-algebra to $[0, \infty]$ that respects countable additivity, and a *probability measure* is a measure of total mass 1.

The framework is essentially due to Kolmogorov (1933), and although the apparatus is more elaborate than the discrete-probability picture taught in introductory courses, the substantive content is very small: "probability is a non-negative, countably additive set function with total mass one." The work is in pinning down which sets it is defined on. This handout develops the basic vocabulary — σ-algebras, measures, probability measures, Lebesgue measure on $\mathbb{R}$, independence, conditional probability — and then the next handout will use it to define random variables and expectations, with the third handout in the cluster covering convergence theorems and the law of large numbers / central limit theorem.

2 σ-algebras

What is the right collection of subsets to assign probabilities to? The naive answer — assign probabilities to every subset — fails for uncountable spaces, in a precise way the previous section flagged. The fix is to commit to a sub-collection that is closed under the operations we care about: complements ("the event did not happen"), unions ("either event happened"), and countable unions ("at least one of countably many events happened"). The combination is called a σ-algebra.

Definition 1. A *σ-algebra* on a set Ω is a collection $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ such that:

1. $\Omega \in \mathcal{F}$;

2. if $A \in \mathcal{F}$, then $A^c := \Omega \setminus A \in \mathcal{F}$ (closed under complement);
3. if $A_1, A_2, A_3, \ldots \in \mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ (closed under countable unions).

The pair $(\Omega, \mathcal{F})$ is called a *measurable space*, and members of $\mathcal{F}$ are *measurable sets* or, in the probability context, *events*.

By De Morgan's laws and (1)–(3), σ-algebras are also closed under countable intersections and contain the empty set. The "σ" is the standard notation for "countable" (versus "finite"); a *finitely additive algebra* drops the countability and is sometimes useful, but the workhorse is the countable version.

Example 2 (Standard σ-algebras)**.**

- *The trivial σ-algebra* $\{\emptyset, \Omega\}$. Every set has either "probability 0" or "probability 1" — no information.
- *The discrete σ-algebra* $\mathcal{P}(\Omega)$. Every subset is measurable. This is the right choice when Ω is countable (no Vitali pathologies on countable spaces); for uncountable Ω, it is generally too rich.
- *The Borel σ-algebra* $\mathcal{B}(\mathbb{R})$ on $\mathbb{R}$ is the smallest σ-algebra containing every open interval. It is the standard σ-algebra for working with the real line: it is rich enough to contain every open set, every closed set, every countable intersection of opens, every countable union of closeds, and so on, but it omits the pathological non-measurable sets.
- *The product σ-algebra* $\mathcal{F} \otimes \mathcal{G}$, defined as the smallest σ-algebra on $\Omega_1 \times \Omega_2$ containing every measurable rectangle $A \times B$ with $A \in \mathcal{F}, B \in \mathcal{G}$. The natural σ-algebra on a product space.

Proposition 3. *For any collection $\mathcal{S} \subseteq \mathcal{P}(\Omega)$, there is a smallest σ-algebra containing $\mathcal{S}$, written $\sigma(\mathcal{S})$.*

Proof. The intersection of any family of σ-algebras on Ω is itself a σ-algebra (each axiom is closed under arbitrary intersection). The intersection of all σ-algebras containing $\mathcal{S}$ — a non-empty family, since $\mathcal{P}(\Omega)$ is one — is therefore a σ-algebra, contains $\mathcal{S}$, and is contained in every σ-algebra containing $\mathcal{S}$. So it is the smallest such. □

The Borel σ-algebra above is the standard example: $\mathcal{B}(\mathbb{R}) = \sigma(\{(a, b) : a, b \in \mathbb{R}, a < b\})$. Equivalently, $\mathcal{B}(\mathbb{R})$ is generated by the half-open intervals $(-\infty, x]$ for $x \in \mathbb{R}$ — a useful characterization when working with cumulative distribution functions in the next handout.

3 Measures and probability measures

A σ-algebra fixes which subsets we can talk about; a *measure* is the function that actually assigns numbers to those subsets — sizes, probabilities, weights of importance. In a model with a discrete population of voters the natural measure is the proportion of the population in each subset. In a continuous voter-type model the natural measure is Lebesgue measure restricted to the unit interval. In a Bayesian game the natural measure is the prior over the type space. Each of these is a probability measure, and the work of this section is to spell out what makes a function on a

σ-algebra count as a measure at all. The defining property, beyond non-negativity, is *countable additivity*: the size of a countable disjoint union is the sum of the sizes. This is what separates measure theory from naive size-assignment, and it is where the technical content of any concrete measure construction lives.

Definition 4. A *measure* on a measurable space $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \to [0, \infty]$ such that:

1. $\mu(\emptyset) = 0$;
2. (countable additivity) for any countable collection $A_1, A_2, A_3, \ldots \in \mathcal{F}$ of pairwise disjoint sets, $\mu\left(\bigsqcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$.

A measure is *finite* if $\mu(\Omega) < \infty$, and a *probability measure* if $\mu(\Omega) = 1$. The triple $(\Omega, \mathcal{F}, \mu)$ is a *measure space*, and when μ is a probability measure we write $(\Omega, \mathcal{F}, \mathbb{P})$ and call it a *probability space*.

The Kolmogorov axioms of probability are these definitions specialized to the case $\mu = \mathbb{P}$ with $\mathbb{P}(\Omega) = 1$: non-negativity, normalization, countable additivity. There is no further axiomatic content.[1]

Example 5 (Standard measures)**.**

- *Counting measure*: on any $(\Omega, \mathcal{P}(\Omega))$, $\mu(A) = |A|$ for finite A, and $\mu(A) = \infty$ for infinite A. The natural "size" on a countable space.
- *Point mass* (or *Dirac measure*) δ_{ω_0}: $\delta_{\omega_0}(A) = 1$ if $\omega_0 \in A$, and 0 otherwise. The probability measure of a deterministic outcome.
- *Discrete probability measure*: any countable convex combination $\sum_n p_n \delta_{\omega_n}$ with $p_n \geq 0$ and $\sum p_n = 1$. The standard "Ω is countable, with probability mass function $p_n = \mathbb{P}(\{\omega_n\})$" picture.
- *Lebesgue measure* λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$: the formalization of length, treated in the next section.

Proposition 6 (Basic measure properties)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $A, B \in \mathcal{F}$ and $A_1, A_2, \ldots \in \mathcal{F}$.*

1. *(Monotonicity) If $A \subseteq B$, then $\mu(A) \leq \mu(B)$.*
2. *(Subadditivity) $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$ (countable union, no disjointness assumption).*
3. *(Continuity from below) If $A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$, then $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n)$.*

[1] The choice of *countable* additivity over mere finite additivity has been disputed at the foundations of Bayesian statistics. Bruno de Finetti argued that finite additivity is the philosophically defensible axiom (it follows from coherent betting behavior on finitely many propositions), with countable additivity an additional commitment that one cannot justify on the same grounds. The mathematical price of giving up countable additivity is severe — the limit theorems (LLN, CLT, dominated convergence, Fubini) do not survive in their familiar forms — so the working practice in probability theory is essentially universally Kolmogorov. The de Finetti tradition does survive in some corners of subjective Bayesian decision theory, where the axioms of choice under uncertainty (Savage, Anscombe–Aumann) are typically only finitely additive on the underlying state space. The political-economy implications are mostly invisible: every applied probabilistic model in political science assumes countable additivity. Billingsley (1995) and Durrett (2019) are standard graduate-level references for the Kolmogorov framework.

4. *(Continuity from above) If $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$ and $\mu(A_1) < \infty$, then $\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n)$.*

Proof sketch. Monotonicity: $B = A \sqcup (B \setminus A)$ (disjoint), so $\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)$. Subadditivity: rewrite $\bigcup A_n$ as a disjoint union $A_1 \sqcup (A_2 \setminus A_1) \sqcup (A_3 \setminus (A_1 \cup A_2)) \sqcup \cdots$, and bound each disjoint piece by the corresponding A_n. Continuity from below: rewrite $\bigcup A_n$ as the disjoint union $A_1 \sqcup (A_2 \setminus A_1) \sqcup (A_3 \setminus A_2) \sqcup \cdots$, and apply countable additivity; the partial sums telescope to $\mu(A_n)$. Continuity from above: apply continuity from below to the complementary sequence $A_1 \setminus A_n$ (which is increasing) and use $\mu(A_1) < \infty$ to subtract. □

These four properties are the bread and butter of measure-theoretic arguments, and most calculations with probability measures unfold by applying one of them. The finiteness condition in (4) is essential: without it, $\mathbb{R} = \bigcap_n (-\infty, n]$ has counterexamples to continuity from above for non-finite μ.

4 Lebesgue measure on $\mathbb{R}$

For continuous models — a continuum of voter types on $[0, 1]$, an interval of policy positions, a real-valued shock, a continuous time variable — the operative measure is *Lebesgue measure*: the formalization of "length" that extends the obvious assignment $\lambda([a, b]) = b - a$ to a much wider class of subsets of $\mathbb{R}$. The construction is technically demanding; we sketch the result without proving it.

Theorem 7 (Existence and uniqueness of Lebesgue measure)**.** *There exists a unique measure λ on the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ such that $\lambda([a, b]) = b - a$ for every closed bounded interval $[a, b]$. This λ is called* Lebesgue measure*, and it is translation-invariant: $\lambda(A + x) = \lambda(A)$ for every $A \in \mathcal{B}(\mathbb{R})$ and $x \in \mathbb{R}$.*

The construction proceeds in three steps: define the obvious length λ^* on intervals; extend λ^* to the *outer measure*, defined for arbitrary subsets of $\mathbb{R}$ by infimum over countable covers; then restrict to the class of subsets that interact well with the outer measure (the *Carathéodory-measurable* sets). The class of Carathéodory-measurable subsets turns out to be a σ-algebra strictly larger than $\mathcal{B}(\mathbb{R})$ but smaller than $\mathcal{P}(\mathbb{R})$, and the restriction of λ^* to it is the Lebesgue measure.[2]

A few specific values one should know.

- $\lambda(\{x\}) = 0$ for every $x \in \mathbb{R}$. (A singleton is contained in $[x, x + \epsilon]$ for every $\epsilon > 0$, so its measure is at most ϵ for every ϵ.) By countable additivity, every countable subset of $\mathbb{R}$ has Lebesgue measure 0. In particular, $\lambda(\mathbb{Q}) = 0$.
- $\lambda(\mathbb{R}) = \infty$. Lebesgue measure on $\mathbb{R}$ is not finite; it is *σ-finite* (there is a countable cover of $\mathbb{R}$ by sets of finite measure: $\mathbb{R} = \bigcup_n [-n, n]$).

[2]The technical heart of the construction is *Carathéodory's extension theorem*: an outer measure (a non-negative, monotone, countably subadditive set function defined on *all* subsets) restricts to a measure on the σ-algebra of subsets that "split additively" with respect to the outer measure — meaning $A \in \mathcal{F}_C$ iff $\lambda^*(E) = \lambda^*(E \cap A) + \lambda^*(E \cap A^c)$ for every E. The theorem is sharp and constructive, and the same machinery used to build Lebesgue measure can be used to build product measures on $\mathbb{R}^n$, Lebesgue–Stieltjes measures arising from cumulative distribution functions in the next handout, and more general measures on locally compact groups. Folland (1999) chapter 1 is the canonical careful treatment; Billingsley (1995) chapter 2 is the standard probability-theory presentation. For most political-economy applications, the existence of Lebesgue measure is invoked as a black box; the construction is used only inside the foundational theory and rarely in everyday applications.

- Restricted to a finite interval $[a, b]$, the measure $\lambda/(b-a)$ is the *uniform probability measure* on $[a, b]$ — the right object for "the voter's type is drawn uniformly from $[a, b]$."

The pathological side of the story.

Theorem 8 (Vitali)**.** *There exists a subset of* $\mathbb{R}$ *that is not Lebesgue-measurable.*

The construction uses the axiom of choice: define an equivalence relation on $[0, 1]$ by $x \sim y$ iff $x - y \in \mathbb{Q}$, and let V be a set of representatives for the equivalence classes (one element per class). Then translates $V + q$ for $q \in \mathbb{Q} \cap [-1, 1]$ are disjoint, and their union covers $[0, 1]$ and is contained in $[-1, 2]$. By translation-invariance, all translates have the same measure m. Countable additivity then forces both $\sum_q m \leq 3$ and $\sum_q m \geq 1$, so m is neither 0 (else the sum is $0 \ngeq 1$) nor positive (else the sum is $\infty \nleq 3$): no consistent value exists. Hence V is not measurable.

For applied probability work the moral is that the choice of σ-algebra is non-trivial: $\mathcal{B}(\mathbb{R})$ (or its completion, the Lebesgue σ-algebra) is the working choice, $\mathcal{P}(\mathbb{R})$ is too rich. Every subset one writes down explicitly in an applied context will be Borel; the non-measurable sets exist only via the axiom of choice and are not encountered in practice. But the technical infrastructure is what underwrites the practice.

5 Independence and conditional probability

Two of the most-used probabilistic objects in political-economy modeling are *independence* (independent voter shocks, independent draws from a population, independent error terms across constituencies) and *conditional probability* (posterior beliefs in a Bayesian signaling model, conditional vote-share distributions given an outcome on a leading indicator, updated estimates after a fresh poll). Independence is the formal version of "knowing one event happened gives no information about another." Conditional probability is the formal version of "updating beliefs given that something happened." Together they are the workhorses of every probabilistic model that has any temporal or informational structure, and they show up almost the moment one moves beyond static one-shot probability.

Definition 9. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$. The *conditional probability of A given B* is

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

For fixed B with $\mathbb{P}(B) > 0$, the function $A \mapsto \mathbb{P}(A \mid B)$ is itself a probability measure on $(\Omega, \mathcal{F})$ (verify: non-negativity, normalization, countable additivity all transfer).

Definition 10. Two events $A, B \in \mathcal{F}$ are *independent* (under $\mathbb{P}$) if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$. A finite collection $A_1, \ldots, A_n$ is *(mutually) independent* if for every subcollection $\{A_{i_1}, \ldots, A_{i_k}\}$,

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k}).$$

A countable family is independent if every finite subfamily is.

The product condition is what one really wants: pairwise independence (every pair satisfies the product condition) is strictly weaker than mutual independence, and both come up in different applications. The classical example: roll two fair dice; let $A =$ "first die is odd," $B =$ "second die is odd," $C =$ "the sum is odd." These three events are pairwise independent but not mutually independent (knowing any two forces the third).

The relationship between conditional probability and independence is direct: A and B are independent iff $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ (when $\mathbb{P}(B) > 0$) — knowing B gives no information about A. The *multiplication rule* $\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A \mid B)$ extends to longer chains: $\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1) \cdots \mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1})$, the *chain rule*.

Theorem 11 (Bayes' rule)**.** *For $A, B \in \mathcal{F}$ with $\mathbb{P}(A), \mathbb{P}(B) > 0$,*

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)}.$$

More generally, if $A_1, \ldots, A_n$ partition Ω (each with positive probability), then for any i,

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}{\sum_{j=1}^{n} \mathbb{P}(B \mid A_j)\mathbb{P}(A_j)}.$$

Proof. Both forms follow from the definition of conditional probability and the law of total probability ($\mathbb{P}(B) = \sum_j \mathbb{P}(B \mid A_j)\mathbb{P}(A_j)$ for a partition $\{A_j\}$). □

Bayes' rule is the workhorse of every Bayesian-updating argument in political economy. Posterior over politician types given a signal; updated beliefs about an election outcome given new polling; learning about an unknown state given an observation. The formal apparatus is one line; the substantive content lives in choosing the prior $\mathbb{P}(A_i)$ and the likelihood $\mathbb{P}(B \mid A_i)$ in a way that captures the modeling situation.

6 What's next

This is the first handout of the probability-and-measure cluster. The next two pick up directly:

- *Random variables and expectations.* A random variable is a measurable function from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ — a real-valued quantity defined on the probability space. The *distribution* of a random variable is the pushforward measure on $\mathbb{R}$, and the *expectation* is the Lebesgue integral with respect to $\mathbb{P}$. These objects, plus conditional expectation, are the next handout.
- *Convergence and limit theorems.* Once we have random variables and expectations, the basic limit theorems — the law of large numbers (sample averages converge to expectations), the central limit theorem (sums of independent shocks have approximately normal distributions), and their cousins — can be stated and proved. These are the workhorse aggregation theorems behind every empirical political-science argument that scales from individual units to electorates.

For broader treatments of measure-theoretic probability, see Billingsley (1995), Durrett (2019), or Williams (1991); the last is the gentlest first reading. For the measure-theoretic foundations themselves (Carathéodory extension, Lebesgue measure construction, L^p spaces), Folland (1999) is the standard.

7 Exercises

Exercise 12. Show that the intersection of any family of σ-algebras on the same set Ω is itself a σ-algebra. Conclude that for any $\mathcal{S} \subseteq \mathcal{P}(\Omega)$, the smallest σ-algebra containing $\mathcal{S}$ is well-defined.

Exercise 13. Show that the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is generated by each of the following classes:

1. the open intervals $\{(a, b) : a < b\}$;
2. the closed intervals $\{[a, b] : a < b\}$;
3. the half-open rays $\{(-\infty, x] : x \in \mathbb{R}\}$.

(The third characterization is the one used to define the cumulative distribution function of a real-valued random variable in the next handout.)

Exercise 14. Prove monotonicity and subadditivity for measures (Proposition 6 parts (1) and (2)) directly from the axioms.

Exercise 15. Prove continuity from below (Proposition 6 part (3)). Where in your proof does countable additivity enter? Why does the analogous claim fail without countable additivity (i.e., for merely finitely additive "measures")?

Exercise 16 (A continuous voter-type space)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the unit interval $[0, 1]$ with the Borel σ-algebra and Lebesgue measure (so $\mathbb{P} = \lambda$ restricted to $[0, 1]$, normalizing automatically since $\lambda([0, 1]) = 1$). Interpret Ω as a continuum of voter types. Compute $\mathbb{P}([0.4, 0.6])$, $\mathbb{P}([0, 0.3] \cup [0.7, 1])$, and $\mathbb{P}(\mathbb{Q} \cap [0, 1])$. The last computation reveals something substantive about the Lebesgue measure of countable sets; state it.

Exercise 17 (Independent voter shocks)**.** A binary electorate has n voters, each independently voting for L with probability p and for R with probability $1 - p$. Let X_i be the indicator that voter i votes for L. Specify the probability space (sample space, σ-algebra, probability measure). Compute $\mathbb{P}(X_1 = 1, X_2 = 1)$ and verify that $\{X_1 = 1\}$ and $\{X_2 = 1\}$ are independent events. Compute the probability that exactly k voters vote for L.

Exercise 18. A fair die is rolled twice. Let $A =$ "first roll is odd," $B =$ "second roll is odd," $C =$ "the sum of the two rolls is odd." Show that A, B, C are pairwise independent but not mutually independent. (*This is the canonical small example separating the two notions.*)

Exercise 19 (Bayesian updating in a signaling model)**.** A politician is one of two types: *competent* ($T = G$) with prior probability $\mathbb{P}(T = G) = 0.5$, or *incompetent* ($T = B$). The politician sends a signal $S \in \{H, L\}$ (high or low). The conditional probabilities are $\mathbb{P}(S = H \mid T = G) = 0.8$ and $\mathbb{P}(S = H \mid T = B) = 0.3$. Given the signal $S = H$, compute the posterior $\mathbb{P}(T = G \mid S = H)$ via Bayes' rule. Now do the same for $S = L$. Comment on which signal is more informative and why.

Exercise 20. Show that for any $A \in \mathcal{F}$ with $0 < \mathbb{P}(A) < 1$, the events A and A^c are *not* independent. Give the intuition: why is the negation of an event always informative about the event?

Exercise 21 (Vitali set, sketched)**.** Sketch the construction of a non-Lebesgue-measurable subset of $[0, 1]$, following the outline in §4. Identify which step uses the axiom of choice and which step uses translation-invariance of Lebesgue measure. Why does the same construction not give a non-measurable subset of a countable set?

References

Billingsley, Patrick (1995). *Probability and Measure*. 3rd ed. New York: Wiley.

Durrett, Rick (2019). *Probability: Theory and Examples*. 5th ed. Cambridge: Cambridge University Press.

Folland, Gerald B. (1999). *Real Analysis: Modern Techniques and Their Applications*. 2nd ed. New York: Wiley.

Williams, David (1991). *Probability with Martingales*. Cambridge: Cambridge University Press.