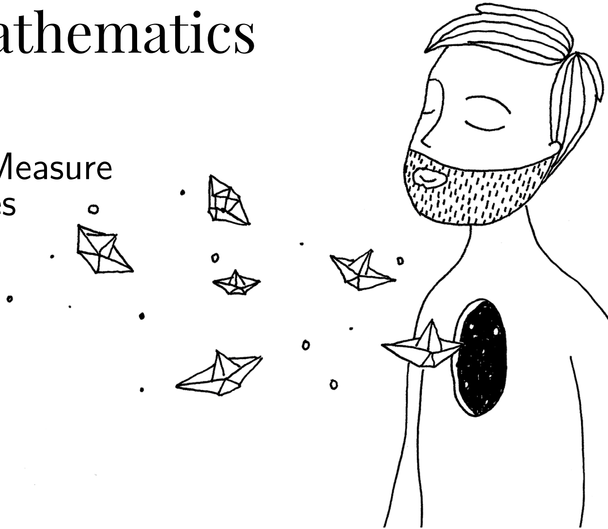
4509 - Bridging Mathematics

Lecture 10: Foundations of Measure Theory and Random Variables



PAULO FAGANDINI

Objectives

- Understand why elementary probability on the power set $\mathcal{P}(\Omega)$ fails on uncountable sample spaces.
- Build up σ-algebras, the Borel σ-algebra $\mathcal{B}(\mathbb{R})$, and a **measure** as a set function assigning size; formalize the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as the unit-mass case.
- Define random variables as measurable maps and contrast Lebesgue with Riemann integration.

The spine: on uncountable outcomes you literally cannot assign a probability to every subset, a paradox forces you to restrict to a well-behaved collection, the σ-**algebra**. A **measure** then assigns size on that collection, and probability is the unit-mass case. A random variable is a **measurable map**; expectation is a **Lebesgue integral** (slice the range, not the domain); and a **sub-σ-algebra is information**. This is the ground the whole econometrics block stands on.

Why naive probability breaks on $\mathbb{R}$

On $\Omega = [0, 1]$ we would want $P : \mathcal{P}([0, 1]) \to [0, 1]$ with

1. **length is probability:** $P([a, b]) = b - a$;
2. **translation invariance:** shifting a set does not change its probability, $P(A + c) = P(A)$;
3. **countable additivity:** for disjoint A, B, $P(A \cup B) = P(A) + P(B)$, and more generally for a countable disjoint family $\{A_i\}_{i\geq 1}$,

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

Why naive probability breaks on $\mathbb{R}$

The point-mass paradox. Draw a number uniformly from $[0, 1]$. What is the probability of hitting *exactly* 0.5? Requirement (1) with $a = b$ forces $P(\{x\}) = b - b = 0$ for *every* point x.

Yet the interval as a whole is the union of all its points, and $P([0, 1]) = 1$. In discrete probability we would just add outcomes up:

$$P([0,1]) = \sum_{x\in[0,1]} P(\{x\}) = \sum_{x\in[0,1]} 0.$$

The culprit: additivity is promised only for *countably* many pieces (requirement 3), and $[0, 1]$ has *uncountably* many points. You cannot add uncountably many zeros and expect 1. So we assign probability to intervals and their countable combinations, not to arbitrary collections of points. That restricted domain is a σ-algebra.

σ-algebras and the Borel sets

Definition

$\mathcal{F} \subseteq \mathcal{P}(\Omega)$ is a σ-**algebra** if

1. $\Omega \in \mathcal{F}$;
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$;
3. $\{A_i\}_{i \geq 1} \subseteq \mathcal{F} \Rightarrow \bigcup_i A_i \in \mathcal{F}$.

The events you are allowed to ask about: closed under "not", "or" (countably), and hence "and". A tidy, self-consistent question set.

σ-algebras: practice

Exercise

1. Let $\Omega = \{1, 2, 3\}$. For each collection, decide whether it is a σ-algebra, and if not, name the axiom that fails.
   - $\mathcal{F}_1 = \{\emptyset, \Omega\}$;
   - $\mathcal{F}_2 = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$;
   - $\mathcal{F}_3 = \{\emptyset, \{1\}, \{2\}, \Omega\}$.
2. A coin, $\Omega = \{H, T\}$. Write down *every* σ-algebra on Ω. How many are there?
3. A die, $\Omega = \{1, \ldots, 6\}$. Build the *smallest* σ-algebra containing the event $\{1\}$ ("the number is a one"). Which sets must it contain?

Part 3 is exactly the move behind the Borel sets: start from the events you care about, then close up under complement and countable union.

The Borel σ-algebra

On a finite Ω the power set $\mathcal{P}(\Omega)$ already works, so no machinery is needed. The point-mass paradox showed $\mathcal{P}([0,1])$ — and hence $\mathcal{P}(\mathbb{R}) \supseteq \mathcal{P}([0,1])$, since $[0,1]$ is already uncountable — is too big. We want the smallest domain that still contains every interval.

The **Borel σ-algebra** $\mathcal{B}(\mathbb{R})$ is the smallest σ-algebra containing all open intervals; its members are essentially every set you can name.

Why this matters: if we define "length" $= b - a$ on simple intervals, measure theory guarantees a unique, paradox-free extension to *all* Borel sets. That object is **Lebesgue measure**, which we define next; probability will be the special case where the total size is 1.

Measurable spaces and measures

A σ-algebra turns Ω into a **measurable space** $(\Omega, \mathcal{F})$: the arena of legal questions. A **measure** supplies the second ingredient, a size for each legal set.

Definition

Given a measurable space $(\Omega, \mathcal{F})$, a **measure** is a function $\mu : \mathcal{F} \to [0, \infty]$ with

1. $\mu(\emptyset) = 0$;
2. **countable additivity:** for disjoint $\{A_i\} \subseteq \mathcal{F}$, $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$.

The σ-algebra axioms are exactly the closure properties any measure needs of its domain; the measure is the separate second ingredient we add now. The triplet $(\Omega, \mathcal{F}, \mu)$ is a **measure space**.

A menu of measures

One measurable space carries many measures, each its own mathematics:

- **Counting measure:** $\mu(A) = \#A$, the number of points in A. On a discrete Ω this is what elementary probability sums over.
- **Lebesgue measure** λ on $\mathcal{B}(\mathbb{R})$: the *unique* measure with $\lambda([a, b]) = b - a$. The paradox-free object the previous slide promised; it extends "length" to every Borel set.
- **Dirac point mass** δ_{x_0}: $\delta_{x_0}(A) = 1$ if $x_0 \in A$, else 0. All the size sits at a single point.

A **probability measure** is just a measure whose total size is one. That single extra requirement, $\mu(\Omega) = 1$, is the whole content of the next slide.

Probability spaces

A **probability space** is a measure space with total mass one. Spelled out in Kolmogorov's axioms:

Definition (Kolmogorov)

A **probability space** is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathcal{F}$ is a σ-algebra on Ω and $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfies

1. $\mathbb{P}(A) \geq 0$ for every event $A \in \mathcal{F}$;
2. $\mathbb{P}(\Omega) = 1$;
3. **countable additivity:** for a countable disjoint family $\{A_i\}$,

$$\mathbb{P}\left(\bigcup_i A_i\right) = \sum_i \mathbb{P}(A_i).$$

Ω is what can happen, $\mathcal{F}$ what you can ask, $\mathbb{P}$ how likely.

Sub-σ-algebras and information

A **sub-σ-algebra** $\mathcal{G} \subseteq \mathcal{F}$ is the events you can currently distinguish: coarser means less information, finer means more.

A **filtration** $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$ models information accumulating over time.

Tomorrow, "conditioning on information" will literally mean projecting onto what a sub-σ-algebra can see.

Random variables

Definition

$X : \Omega \to \mathbb{R}$ is a **random variable** if preimages of Borel sets are events:

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F} \quad \forall B \in \mathcal{B}(\mathbb{R}),$$

equivalently $\{X \leq x\} \in \mathcal{F}$ for all x.

- **Pushforward (distribution of X):** $P_X(B) = \mathbb{P}(X^{-1}(B))$.
- **CDF:** $F_X(x) = \mathbb{P}(X \leq x) = P_X((-\infty, x])$.

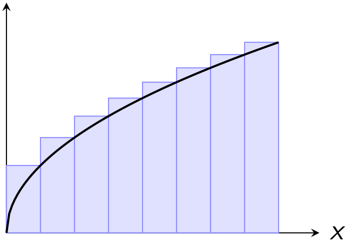
"Measurable" just means every question about X ("is $X \leq 3$?") is a legitimate event.
The pushforward lets you forget Ω and work on $\mathbb{R}$ with F_X.

Riemann versus Lebesgue

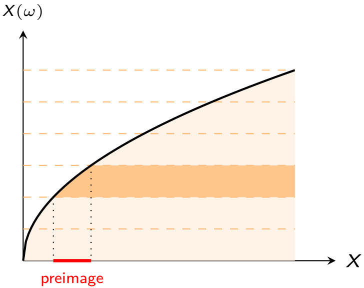
- **Riemann:** chop the *domain* into vertical strips, sum height times width.
- **Lebesgue:** chop the *range* into horizontal strips; for each level, measure the preimage and weight by it.

Riemann asks "how tall is the function here?"; Lebesgue asks "how much of the space maps into this height band?" The second works even when the function is wildly discontinuous: the Dirichlet function $\mathbf{1}_{\mathbb{Q}}$ has no Riemann integral but Lebesgue integral 0. This is the integral that defines expectation.

Riemann vs. Lebesgue, in a picture



Riemann: partition the *domain*



Lebesgue: partition the *range*

Riemann asks how tall the function is over each domain strip; Lebesgue fixes a height band and *measures the preimage* $\{\omega : X(\omega) \in [y, y+dy]\}$ that lands in it. Weighting those preimages by $\mathbb{P}$ instead of by length is what turns this integral into an expectation.

The expected value

Definition

The **expected value** of a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ is its Lebesgue integral against $\mathbb{P}$,

$$E[X] = \int_\Omega X(\omega)\, d\mathbb{P}(\omega) = \int_{-\infty}^{\infty} x\, dF_X(x) \quad \text{(a Lebesgue–Stieltjes integral)},$$

defined whenever $E[|X|] < \infty$.

The same number seen from two sides: integrate over outcomes ω weighted by $\mathbb{P}$, or over values x weighted by the distribution F_X (the pushforward lets you forget Ω). Because it is a Lebesgue integral, $E[X]$ stays well-defined even for wildly discontinuous X, which is why it survives the messy distributions econometrics throws at it.

Recovering the formulas you already know

The abstract $\int x\, dF_X(x)$ collapses to elementary calculus in the two "well-behaved" cases:

- **Density (continuous) case.** If X has a density f, i.e. the distribution P_X is absolutely continuous w.r.t. Lebesgue measure, then $f = \dfrac{dF_X}{dx}$ and

$$E[X] = \int_{-\infty}^{\infty} x\, dF_X(x) = \int_{-\infty}^{\infty} x\, f(x)\, dx,$$

  an ordinary Riemann integral.
- **Discrete case.** If F_X jumps by p_i at atoms x_i, then dF_X places mass p_i there and $E[X] = \sum_i x_i\, p_i$.

The substitution "$dF_X = f(x)\, dx$" economists and physicists write is exactly this: f is the **Radon–Nikodym derivative** $dP_X/d\lambda$, legitimate precisely when a density exists. The Lebesgue–Stieltjes integral is the umbrella; the Riemann integral and the discrete sum are its two specializations.

Expectation properties

- **Linearity:** $E[aX + bY] = aE[X] + bE[Y]$.
- **Monotonicity:** $X \leq Y$ almost surely $\Rightarrow E[X] \leq E[Y]$.
- **Jensen:** g convex $\Rightarrow g(E[X]) \leq E[g(X)]$.

Jensen is why $\mathrm{Var}(X) = E[X^2] - (E[X])^2 \geq 0$, and why risk aversion (concave utility) gives $E[u(X)] \leq u(E[X])$, the risk premium. Tomorrow expectation becomes an inner product $\langle X, Y \rangle = E[XY]$, and the linear algebra of Lecture 3 returns.

References

- G. Casella and R. L. Berger, *Statistical Inference*, Ch. 1 (probability theory).
- B. E. Hansen, *Probability and Statistics for Economists*, Ch. 1 and 2.
- P. Billingsley, *Probability and Measure*, Ch. 1 and 2.
- F. Hayashi, *Econometrics*, Ch. 2 (large-sample foundations).