Foundations of Probability Theory: a Sketch

Probability Spaces

Kolmogorov set down a clear mathematical foundation for probability theory in 1933. The basic ingredient is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where

- Ω is the set of all possible outcomes ω.
- $\mathcal{F}$ is a σ-*field* (or σ-*algebra*): a collection of subsets (=events) $A \subset \Omega$ such that
    i) $\Omega \in \mathcal{F}$
    ii) if $A \in \mathcal{F}$ then $A^c = \Omega \setminus A \in \mathcal{F}$
    iii) if $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcup_n A_n \in \mathcal{F}$
- $\mathbb{P}$ is a *probability measure* on $(\Omega, \mathcal{F})$: a mapping $\mathbb{P} : \mathcal{F} \to [0,1]$ such that
    i) $\mathbb{P}[\Omega] = 1$ (normalizing condition)
    ii) if $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint then $\mathbb{P}\left[\bigcup_n A_n\right] = \sum_n \mathbb{P}[A_n]$.

The pair $(\Omega, \mathcal{F})$ is a *measurable space*; $(\Omega, \mathcal{F}, \mathbb{P})$ is a *probability space.*

Typical construction of a σ-field: let $\mathcal{Z}$ be a collection of subsets $A \subset \Omega$. The σ-field *generated by* $\mathcal{Z}$ is the smallest σ-field containing $\mathcal{Z}$:

$$\sigma(\mathcal{Z}) := \bigcap_{\substack{\mathcal{G}\ \sigma\text{-field} \\ \mathcal{G} \supset \mathcal{Z}}} \mathcal{G} \quad \text{(this is a } \sigma\text{-field).}$$

Examples:

i) $\mathcal{Z} = \{B_1, B_2, \dots\}$ a countable partition of Ω. Then

$$\sigma(\mathcal{Z}) = \left\{ \bigcup_{i \in I} B_i | I \text{ subset of } \mathbb{N} \right\}.$$

The B_is are called *atoms*.

ii) $\Omega = \mathbb{R}^d$. The *Borel σ-field* is

$$\mathcal{B}(\mathbb{R}^d) := \sigma(\{\text{open sets in } \mathbb{R}^d\}).$$

Random Variables

Let $(E, \mathcal{E})$ be a measurable space (the "state space", e.g. $E = \mathbb{R}, \mathbb{R}^d, C[0,1], \dots$). A map $X : \Omega \to E$ is an *(E-valued) random variable* (or *($\mathcal{F}/\mathcal{E}$)-measurable)* if

$$X^{-1}(B) \in \mathcal{F} \quad \text{for all} \quad B \in \mathcal{E}.$$

The σ-field *generated by* X is the smallest σ-field on Ω that makes X measurable:

$$\sigma(X) := \bigcap_{\substack{\mathcal{G} \ \sigma\text{-field} \\ X \ \mathcal{G}/\mathcal{E}\text{-measurable}}} \mathcal{G} = X^{-1}(\mathcal{E}).$$

For $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ we have $\sigma(X) = \sigma(\{X \le x\}, x \in \mathbb{R})$; and $X : \Omega \to \mathbb{R}$ is a random variable if and only if $\{X \le x\} \in \mathcal{F}$ for all $x \in \mathbb{R}$. The *distribution function* of X is

$$F(x) = \mathbb{P}\left[X \le x\right], \quad x \in \mathbb{R}.$$

A map $Y : \Omega \to \mathbb{R}$ is $\sigma(X)$-measurable if and only if $Y = f(X)$ for some measurable function $f : \mathbb{R} \to \mathbb{R}$ (this also hold for more general state spaces).

Independence

A collection $\mathcal{G}_1, \mathcal{G}_2, \dots \subset \mathcal{F}$ of σ-fields is *independent* if for every choice $A_i \in \mathcal{G}_i,\ i = 1, 2, \dots$, the events $A_1, A_2, \dots$ are independent:

$$\mathbb{P}\left[A_1 \cap A_2 \cap \dots \cap A_n\right] = \mathbb{P}[A_1] \dots \mathbb{P}[A_n] \quad \text{for all} \quad n \in \mathbb{N}.$$

The random variables $X_1, X_2, \dots$ are *independent* if $\sigma(X_1), \sigma(X_2), \dots$ are independent.

Expectation

Let X be a non-negative random variable. Then there exists a sequence (X_n) of *simple random variables* (taken only finitely many values) such that $0 \leq X_1 \leq X_2 \leq \ldots \leq X_n \uparrow X$ a.s. Then $0 \leq \mathbb{E}[X_1] \leq \mathbb{E}[X_2] \leq \ldots$ (the definition of $\mathbb{E}[X_n]$ is clear) and the limit

$$\mathbb{E}[X] := \lim_n \mathbb{E}[X_n] \quad (\textit{monotone convergence}!)$$

exists (can be $+\infty$). A real-valued random variable $X = X_+ - X_-$ is *integrable* ($X \in L^1$) if $\mathbb{E}[|X|] < \infty$, which holds if and only if $\mathbb{E}[X_+] < \infty$ and $\mathbb{E}[X_-] < \infty$, and then

$$\mathbb{E}[X] = \int_\Omega X d\mathbb{P} := \mathbb{E}[X_+] - \mathbb{E}[X_-]$$

is the *expectation* of X. Moreover,

$$\mathbb{E}[h(X)] = \int_\mathbb{R} h(x) dF(x) = \int_\mathbb{R} h(x) f(x) dx$$

For any measurable $h : \mathbb{R} \to \mathbb{R}$ such that $h(X) \in L^1$, where $F(x)$ is the distribution function and $f(x)$ the density function (if it exists) of X.

Fundamental Result for Modelling

Given distribution functions $F_1, F_2, \ldots$ one can always construct a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a sequence $X_1, X_2, \ldots$ of independent random variables such that X_n has distribution F_n for all $n \in \mathbb{N}$. (Simulation of i.i.d. random variables!)

Idea: realize X_n on $\Omega_n = ([0,1], \mathcal{B}[0,1], \textit{Leb})$, take infinite product $\Omega = \Omega_1 \times \Omega_2 \times \ldots$ (see Chapter 8 in D. Williams, Probability with Martingales, Cambridge University Press, 1995).

Conditional Expectation

Let $\mathcal{G} \subset \mathcal{F}$ be a σ-field, and X an $\mathcal{F}$-measurable, integrable random variable. We consider $\mathcal{G}$ as the collection of events that can be "observed" (the available information). What is the best prediction, say Y, of X given $\mathcal{G}$?

- If $\mathcal{G} = \{\Omega, \emptyset\}$ then $Y = \mathbb{E}[X]$.
- If $\mathcal{G}$ has finitely many atoms $A_1, \ldots, A_n$ (each with $\mathbb{P}[A_i] > 0$) then

$$Y = \frac{1}{\mathbb{P}[A_i]} \int_{A_i} X d\mathbb{P} \quad \text{on} \quad A_i.$$

In both cases we clearly have $\mathbb{E}[Y 1_A] = \mathbb{E}[X 1_A]$ for all $A \in \mathcal{G}$. This is in fact the defining property: Y is called *conditional expectation* of X with respect to $\mathcal{G}$ if

i) Y is $\mathcal{G}$-measurable

ii) $\mathbb{E}[Y 1_A] = \mathbb{E}[X 1_A]$ for all $A \in \mathcal{G}$.

We write $Y = \mathbb{E}[X|\mathcal{G}]$, and this is defined up to a $\mathbb{P}$-null set.

Theorem: The conditional expectation $\mathbb{E}[X|\mathcal{G}]$ exists.

The idea of the proof is as follows: (1) If $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$: $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is a closed subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ (these are Hilbert spaces with inner product $\langle U, V \rangle = \mathbb{E}[UV]$). Let Y be the ortogonal projection of X onto $L^2(\Omega, \mathcal{G}, \mathbb{P})$. Then $\mathbb{E}[Y 1_A] = \langle Y, 1_A \rangle = \langle X, 1_A \rangle = \mathbb{E}[X 1_A]$ for all $1_A \in L^2(\Omega, \mathcal{G}, \mathbb{P})(\Leftrightarrow A \in \mathcal{G})$ and we are done. (2) If $X \geq 0$: approximate X by simple random variables $X_n \uparrow X$ a.s. (the X_ns are in $L^2(\Omega, \mathcal{G}, \mathbb{P})$!) and obtain $\mathbb{E}[X|\mathcal{G}]$ as monotone limit of $\mathbb{E}[X_n|\mathcal{G}]$. (3) For general $X \in L^1$ write $X = X_+ - X_-$.

Here is a list of frequently used properties.

i) $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.

ii) If $X \in \mathcal{G}$ then $\mathbb{E}[X|\mathcal{G}] = X$.

iii) (Monotonicity) If $X \leq Y$ then $\mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$.

iv) (Linearity) $\mathbb{E}[aX + Y|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + \mathbb{E}[Y|\mathcal{G}]$.

v) (cDCT) If $|X_n| \leq V$, where $V \in L^1$, and $\mathbb{P}[X_n \to X] = 1$ then $X \in L^1$ and

$$\lim_n \mathbb{E}[X_n|\mathcal{G}] = \mathbb{E}[\lim_n X_n|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}].$$

vi) (cMCT) If $0 \leq X_1 \leq X_2 \leq \ldots$ then

$$\lim_n \mathbb{E}[X_n|\mathcal{G}] = \mathbb{E}[\lim_n X_n|\mathcal{G}] \quad (\leq \infty).$$

vii) (cFatou) If $X_n \geq 0$ then

$$\mathbb{E}[\liminf_n X_n|\mathcal{G}] \leq \liminf_n \mathbb{E}[X_n|\mathcal{G}].$$

viii) (Tower property) If $\mathcal{H} \subset \mathcal{G}$ is a σ-field then

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\ \mathcal{H}] = \mathbb{E}[X|\mathcal{H}].$$

ix) ("Taking out what is known") If Y is $\mathcal{G}$-measurable and $|XY|$ and $|X|$ are integrable then

$$\mathbb{E}[XY|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}].$$

x) (Independence) If $\mathcal{G}$ is independent of X then

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X].$$

Note: *By convention* all equalities between random variables hold a.s.

References

[1] W. Feller: An Introduction to Probability Theory and Its Applications. Vol. I. Wiley, New York (1968).

[2] W. Feller. An Introduction to Probability Theory and Its Applications.Vol. II. Wiley, New York.

[3] J.Jacod, P.Protter. Probability Essentials, Springer.

[4] D. Williams, Probability with Martingales, Cambridge University Press, 1995.