Lecture Notes: From Continuity of Probability to Distribution Functions

Graduate Random Processes

Throughout, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a probability space: Ω is the sample space, $\mathcal{F}$ is a σ-algebra of events, and $\mathbb{P} : \mathcal{F} \to [0,1]$ is a probability measure (countably additive, with $\mathbb{P}(\Omega) = 1$). These notes develop the measure-theoretic backbone we will use for the rest of the course: limiting behaviour of events, the rigorous notion of a random variable, measurability, and the distribution function.

1 Continuity of Probability

A probability measure is finitely additive by definition and countably additive by assumption. A convenient reformulation of countable additivity is that $\mathbb{P}$ behaves well under monotone limits of events.

Theorem 1.1 (Continuity of probability)**.** *Let* $(A_n)_{n\geq 1} \subseteq \mathcal{F}$*.*

(a) (From below) *If* $A_1 \subseteq A_2 \subseteq \cdots$*, then* $\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

(b) (From above) *If* $A_1 \supseteq A_2 \supseteq \cdots$*, then* $\mathbb{P}\Big(\bigcap_{n=1}^{\infty} A_n\Big) = \lim_{n\to\infty} \mathbb{P}(A_n)$.

Proof. (a) Disjointify the increasing sequence. Set $B_1 := A_1$ and, for $n \geq 2$, $B_n := A_n \setminus A_{n-1}$. The B_n are pairwise disjoint, they lie in $\mathcal{F}$, and for each N,

$$\bigcup_{n=1}^{N} B_n = A_N, \qquad \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n.$$

By countable additivity followed by the definition of an infinite series,

$$\mathbb{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \mathbb{P}\Big(\bigcup_{n=1}^{\infty} B_n\Big) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) = \lim_{N\to\infty} \sum_{n=1}^{N} \mathbb{P}(B_n) = \lim_{N\to\infty} \mathbb{P}(A_N).$$

(b) Pass to complements. If $A_n \downarrow$, then $A_n^c \uparrow$, and $\big(\bigcap_n A_n\big)^c = \bigcup_n A_n^c$. Applying part (a) to (A_n^c) and using $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$,

$$1 - \mathbb{P}\Big(\bigcap_n A_n\Big) = \mathbb{P}\Big(\bigcup_n A_n^c\Big) = \lim_{n\to\infty} \mathbb{P}(A_n^c) = \lim_{n\to\infty} \big(1 - \mathbb{P}(A_n)\big) = 1 - \lim_{n\to\infty} \mathbb{P}(A_n). \qquad \square$$

Remark 1.2. Finiteness of the measure is essential in part (b): we used $\mathbb{P}(A_n^c) = 1 - \mathbb{P}(A_n)$, which requires $\mathbb{P}(\Omega) < \infty$. For a general (infinite) measure, continuity from above can fail without the extra hypothesis $\mu(A_{n_0}) < \infty$ for some n_0. Since $\mathbb{P}(\Omega) = 1$ this subtlety never bites us here, but it is worth remembering.

2 The Limit Superior of a Sequence of Events

We often want to speak of events that occur "infinitely often." The right object is the limit superior.

Definition 2.1. For events $(A_n)_{n\geq 1}$, define

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k, \qquad \liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k.$$

Unwinding the quantifiers is illuminating. An outcome ω lies in $\limsup_n A_n$ iff for *every* n there is some $k \geq n$ with $\omega \in A_k$; that is, ω belongs to infinitely many of the A_n. For this reason we write

$$\limsup_{n\to\infty} A_n = \{A_n \text{ i.o.}\} \qquad (\text{“}A_n \text{ infinitely often”}).$$

Dually, $\omega \in \liminf_n A_n$ iff $\omega \in A_n$ for *all but finitely many* n, i.e. $\{A_n \text{ eventually}\}$. Both sets are in $\mathcal{F}$, being countable unions and intersections of events.

Remark 2.2. Writing $G_n := \bigcup_{k\geq n} A_k$, the sets G_n decrease to $\limsup_n A_n$. Continuity from above (Theorem 1.1) therefore gives $\mathbb{P}(\limsup_n A_n) = \lim_{n\to\infty}\mathbb{P}(G_n)$, a fact we exploit immediately below.

3 The Borel–Cantelli Lemma

Lemma 3.1 (First Borel–Cantelli)**.** *If* $\sum_{n=1}^{\infty}\mathbb{P}(A_n) < \infty$*, then* $\mathbb{P}\big(\limsup_n A_n\big) = 0$.

Proof. For every n, $\limsup_k A_k \subseteq \bigcup_{k=n}^{\infty} A_k$, so by monotonicity and countable subadditivity,

$$\mathbb{P}\Big(\limsup_k A_k\Big) \;\leq\; \mathbb{P}\Big(\bigcup_{k=n}^{\infty} A_k\Big) \;\leq\; \sum_{k=n}^{\infty}\mathbb{P}(A_k).$$

The right-hand side is the tail of a convergent series, hence tends to 0 as $n\to\infty$. As the left-hand side does not depend on n, it is 0. □

Note that the first lemma requires *no* independence: it is a pure consequence of subadditivity.

Lemma 3.2 (Second Borel–Cantelli)**.** *If the events* (A_n) *are independent and* $\sum_{n=1}^{\infty}\mathbb{P}(A_n) = \infty$*, then* $\mathbb{P}\big(\limsup_n A_n\big) = 1$.

Proof. It suffices to show $\mathbb{P}\big((\limsup_n A_n)^c\big) = 0$. By De Morgan, $(\limsup_n A_n)^c = \bigcup_n \bigcap_{k\geq n} A_k^c$, so it is enough to prove $\mathbb{P}\big(\bigcap_{k=n}^{\infty} A_k^c\big) = 0$ for each fixed n. Using independence of (A_k^c) and the elementary bound $1 - x \leq e^{-x}$ for $x \in [0,1]$, for any $m \geq n$,

$$\mathbb{P}\Big(\bigcap_{k=n}^{m} A_k^c\Big) = \prod_{k=n}^{m}\big(1-\mathbb{P}(A_k)\big) \leq \prod_{k=n}^{m} e^{-\mathbb{P}(A_k)} = \exp\Big(-\sum_{k=n}^{m}\mathbb{P}(A_k)\Big).$$

Letting $m\to\infty$, the exponent tends to $-\infty$ because the series diverges, so $\mathbb{P}\big(\bigcap_{k=n}^{\infty} A_k^c\big) = 0$ by continuity from above. A countable union of null sets is null, so $\mathbb{P}\big((\limsup_n A_n)^c\big) = 0$. □

3.1 Why independence in the second lemma is too strong

Independence is a heavy hypothesis, and the proof above uses it in a very particular way: only to factor $\mathbb{P}(\bigcap_k A_k^c) = \prod_k (1 - \mathbb{P}(A_k))$. The *conclusion* $\mathbb{P}(\limsup_n A_n) = 1$, however, survives under far weaker assumptions. Three points are worth making.

(i) Pairwise independence already suffices. One does not need the full (mutual) independence of the family; pairwise independence of (A_n) together with $\sum_n \mathbb{P}(A_n) = \infty$ still forces $\mathbb{P}(\limsup_n A_n) = 1$. This is a genuine weakening, since pairwise independence is strictly weaker than mutual independence.

(ii) No independence is needed at all — only control of correlations. The Kochen–Stone lemma states that if $\sum_n \mathbb{P}(A_n) = \infty$, then

$$\mathbb{P}\Big(\limsup_n A_n\Big) \;\geq\; \limsup_{N\to\infty} \frac{\big(\sum_{k=1}^N \mathbb{P}(A_k)\big)^2}{\sum_{j=1}^N \sum_{k=1}^N \mathbb{P}(A_j \cap A_k)}.$$

No independence appears anywhere; the hypothesis is purely a second-moment (correlation) condition. When the events are pairwise independent, the off-diagonal terms satisfy $\mathbb{P}(A_j \cap A_k) = \mathbb{P}(A_j)\mathbb{P}(A_k)$, so the denominator is

$$\Big(\sum_{k\leq N} \mathbb{P}(A_k)\Big)^2 - \sum_{k\leq N} \mathbb{P}(A_k)^2 + \sum_{k\leq N} \mathbb{P}(A_k),$$

and the correction terms are of lower order than $\big(\sum_{k\leq N} \mathbb{P}(A_k)\big)^2$. Hence the ratio tends to 1 and we recover $\mathbb{P}(\limsup_n A_n) = 1$. Thus independence was a convenient *sufficient* condition standing in for the real requirement: that the events not be too positively correlated.

(iii) Independence is not necessary. It is easy to build dependent families with $\sum_n \mathbb{P}(A_n) = \infty$ and $\mathbb{P}(\limsup_n A_n) = 1$, and equally to build dependent families where the sum diverges yet $\mathbb{P}(\limsup_n A_n) < 1$. The dichotomy 0/1 that independence produces is a feature of independence, not of the divergent-sum hypothesis. The takeaway: read the second lemma as a statement about correlation structure, with independence merely the cleanest special case.

4 Random Variables

Informally a random variable is a numerical readout of the random outcome. The rigorous definition demands that this readout interact correctly with the event structure $\mathcal{F}$.

Definition 4.1. A *random variable* on $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ that is $\mathcal{F}$-measurable, meaning

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F} \qquad \text{for every Borel set } B \in \mathcal{B}(\mathbb{R}).$$

Requiring the preimage of every Borel set is a lot to check directly. Fortunately it is enough to test on a generating class.

Proposition 4.2. *$X : \Omega \to \mathbb{R}$ is a random variable if and only if $\{X \leq x\} := X^{-1}\big((-\infty, x]\big) \in \mathcal{F}$ for every $x \in \mathbb{R}$.*

Proof sketch. Necessity is immediate since $(-\infty, x]$ is Borel. For sufficiency, the collection $\mathcal{G} := \{B \subseteq \mathbb{R} : X^{-1}(B) \in \mathcal{F}\}$ is a σ-algebra, because preimages commute with complements and countable unions. By hypothesis $\mathcal{G}$ contains all half-lines $(-\infty, x]$, which generate $\mathcal{B}(\mathbb{R})$; hence $\mathcal{B}(\mathbb{R}) \subseteq \mathcal{G}$. $\square$

Proposition 4.3 (Trivial σ-algebra forces constants)**.** *If $\mathcal{F} = \{\varnothing, \Omega\}$ is the trivial σ-algebra, then every random variable X is constant.*

Proof. By Proposition 4.2, for each $x \in \mathbb{R}$ the set $\{X \le x\}$ must belong to $\mathcal{F} = \{\varnothing, \Omega\}$, so it equals either $\varnothing$ or Ω. As x increases the sets $\{X \le x\}$ increase, so

$$S := \{x \in \mathbb{R} : \{X \le x\} = \Omega\}$$

is an up-set: if $s \in S$ and $x \ge s$ then $x \in S$. Put $c := \inf S$.

Fix $\omega \in \Omega$. For any $x > c$ there is $s \in S$ with $s \le x$, whence $\{X \le s\} = \Omega$ and thus $X(\omega) \le s \le x$; letting $x \downarrow c$ gives $X(\omega) \le c$. For any $x < c$ we have $x \notin S$, so $\{X \le x\} = \varnothing$, meaning $X(\omega) > x$; letting $x \uparrow c$ gives $X(\omega) \ge c$. Hence $X(\omega) = c$ for every ω.

Finally c is finite: if $c = +\infty$ then $\{X \le x\} = \varnothing$ for all x, contradicting that X is real-valued; if $c = -\infty$ then $\{X \le x\} = \Omega$ for all x, likewise impossible. Thus $X \equiv c$ for some $c \in \mathbb{R}$. $\square$

Remark 4.4. The moral: a random variable can only distinguish outcomes that $\mathcal{F}$ can distinguish. With no nontrivial events available, there is no "information" to encode, and X collapses to a constant. This previews the general principle that $\sigma(X)$ measures how much of $\mathcal{F}$ the variable X can resolve.

5 Borel Measurability

The Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is the smallest σ-algebra containing the open subsets of $\mathbb{R}$ (equivalently, generated by the open intervals, or by the half-lines $(-\infty, x]$). A function $f : \mathbb{R} \to \mathbb{R}$ is *Borel measurable* if $f^{-1}(B) \in \mathcal{B}(\mathbb{R})$ for every $B \in \mathcal{B}(\mathbb{R})$. Borel measurable functions are exactly those we may compose with random variables and still obtain random variables.

Proposition 5.1. *Every continuous function $f : \mathbb{R} \to \mathbb{R}$ is Borel measurable.*

Proof. Let $\mathcal{G} := \{B \subseteq \mathbb{R} : f^{-1}(B) \in \mathcal{B}(\mathbb{R})\}$. As in Proposition 4.2, $\mathcal{G}$ is a σ-algebra (preimages respect complements and countable unions). If $U \subseteq \mathbb{R}$ is open, then continuity of f makes $f^{-1}(U)$ open, hence Borel, so $U \in \mathcal{G}$. Therefore $\mathcal{G}$ contains all open sets, and since $\mathcal{B}(\mathbb{R})$ is the σ-algebra they generate, $\mathcal{B}(\mathbb{R}) \subseteq \mathcal{G}$. That is, $f^{-1}(B) \in \mathcal{B}(\mathbb{R})$ for all Borel B. $\square$

Remark 5.2. This is an instance of the "good sets" principle used repeatedly above: to prove a property holds for all Borel sets, show the sets enjoying it form a σ-algebra and contain a generating class. Continuity is far more than needed — piecewise continuous, monotone, and semicontinuous functions are all Borel measurable — but continuity is the case we use most.

6 The Cumulative Distribution Function

Definition 6.1. The *cumulative distribution function* (CDF) of a random variable X is

$$F_X : \mathbb{R} \to [0,1], \qquad F_X(x) := \mathbb{P}(X \le x) = \mathbb{P}\big(X^{-1}((-\infty, x])\big).$$

6.1 Why the half-open interval $(-\infty, x]$

The choice to close the interval at its right endpoint — to use $\{X \le x\}$ rather than $\{X < x\}$ — is a genuine convention, and it is the convention that makes F_X *right-continuous.* Consider a sequence $x_n \downarrow x$. Then

$$(-\infty, x_n] \ \downarrow\ (-\infty, x], \qquad \text{so} \qquad \{X \le x_n\} \ \downarrow\ \{X \le x\} = \bigcap_n \{X \le x_n\}.$$

By continuity from above (Theorem 1.1), $F_X(x_n) = \mathbb{P}(X \le x_n) \to \mathbb{P}(X \le x) = F_X(x)$. Thus closing the interval on the right is exactly what synchronises the endpoint convention with continuity from above and yields right-continuity.

Had we instead defined $F(x) = \mathbb{P}(X < x)$, the same argument with $x_n \uparrow x$ would produce a *left*-continuous function. Neither is wrong, but the $\{X \le x\}$ convention is standard; it also makes the increment over a half-open interval clean:

$$\mathbb{P}\big(a < X \le b\big) = F_X(b) - F_X(a),$$

which is the identity that drives the construction of the distribution of X in Section 7.

Left limits carry probabilistic meaning too. If $x_n \uparrow x$ then $\{X \le x_n\} \uparrow \{X < x\}$, so by continuity from below

$$F_X(x^-) := \lim_{y \uparrow x} F_X(y) = \mathbb{P}(X < x), \qquad F_X(x) - F_X(x^-) = \mathbb{P}(X = x).$$

Hence F_X is continuous at x precisely when $\mathbb{P}(X = x) = 0$, and its jumps locate the atoms of X.

6.2 Properties of CDFs

Proposition 6.2. *The distribution function $F = F_X$ satisfies:*
(F1) Monotonicity: $x \le y \implies F(x) \le F(y)$.
(F2) Right-continuity: $\lim_{y \downarrow x} F(y) = F(x)$ *for every* x.
(F3) Normalization at the ends: $\lim_{x \to -\infty} F(x) = 0$ *and* $\lim_{x \to +\infty} F(x) = 1$.

Proof idea. (F1) holds because $\{X \le x\} \subseteq \{X \le y\}$ when $x \le y$. (F2) is the right-continuity computation above. For (F3), take $x_n \downarrow -\infty$, so $\{X \le x_n\} \downarrow \varnothing$ and $F(x_n) \to 0$; take $x_n \uparrow +\infty$, so $\{X \le x_n\} \uparrow \Omega$ and $F(x_n) \to 1$; both use continuity of $\mathbb{P}$. □

These three properties are not merely necessary — they are exactly characterizing.

Theorem 6.3 (Characterization)**.** *If $F : \mathbb{R} \to [0, 1]$ satisfies* (F1)–(F3)*, then there exists a probability space and a random variable X on it with $F_X = F$. In fact F determines a unique probability measure on $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$.*

The existence and, crucially, the *uniqueness* in Theorem 6.3 are supplied by an extension theorem, to which we now turn.

7 Carathéodory's Extension Theorem

To construct the distribution of X from its CDF, we would like to *declare* the probabilities of the simplest sets and have them propagate to all Borel sets. The natural simple sets are the half-open intervals $(a, b]$, on which we want

$$\mathbb{P}\big((a, b]\big) = F(b) - F(a).$$

Finite disjoint unions of such intervals form an *algebra* $\mathcal{A}$: a collection containing $\mathbb{R}$ and closed under complements and *finite* unions (but not necessarily countable ones). On $\mathcal{A}$ the prescription above is unambiguous and countably additive whenever it can be. The question is whether it extends to the σ-algebra $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$ generated by these intervals.

Theorem 7.1 (Carathéodory extension, probabilistic form)**.** *Let $\mathcal{A}$ be an algebra of subsets of Ω and let $\mu : \mathcal{A} \to [0, 1]$ be a set function with $\mu(\Omega) = 1$ that is countably additive on $\mathcal{A}$ — that is, whenever $(E_n) \subseteq \mathcal{A}$ are pairwise disjoint and their union happens to lie in $\mathcal{A}$, one has $\mu(\bigcup_n E_n) = \sum_n \mu(E_n)$. Then μ extends to a probability measure on $\sigma(\mathcal{A})$, and this extension is unique.*

We will not prove the theorem, but the mechanism is worth stating: one builds an *outer measure* $\mu^*(E) = \inf \sum_n \mu(A_n)$ over countable covers $E \subseteq \bigcup_n A_n$ with $A_n \in \mathcal{A}$, shows the sets that "split" every test set additively (the Carathéodory-measurable sets) form a σ-algebra on which μ^* is countably additive, and checks that this σ-algebra contains $\mathcal{A}$. Uniqueness comes from a π–λ argument: two probability measures agreeing on the algebra $\mathcal{A}$ (which is closed under finite intersections) must agree on the generated σ-algebra.

Corollary 7.2 (Lebesgue–Stieltjes measure)**.** *For any F satisfying* (F1)–(F3) *there is a unique probability measure $\mathbb{P}_F$ on $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ with $\mathbb{P}_F\big((a, b]\big) = F(b) - F(a)$ for all $a < b$. Taking $X(\omega) = \omega$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_F)$ realises F as a CDF, proving Theorem 6.3.*

Remark 7.3. Right-continuity of F is exactly what guarantees the interval set function is countably additive on the algebra, so the half-open convention of Section 6.1 is not cosmetic: it is the hypothesis that lets Carathéodory's theorem apply. Uniqueness is the practical payoff — it says a distribution is *completely* determined by its CDF, so from now on we may specify a random variable's law simply by writing down a function F with properties (F1)–(F3).