Probability Notes

Duncan MacIntyre

October 28, 2023

Contents

1 Some measure theory

Gap!!! The discussion of topologies should probably be removed.

Definition 1.1. Let Ω be a set. A topology on Ω is a collection $\mathscr{A} \subset P(\Omega)$ that is closed under unions and finite intersections with $\Omega, \emptyset \in \mathscr{A}$. For $E \subset \Omega$, if $E \in \mathscr{A}$, we call E open, and if $E^C \in \mathscr{A}$, we call E closed.

For example, if (X, d) is a metric space, $\mathscr{A} = \{E | E = \cup_i N_{r_i}(x_i)\}$ is a topology on X.

Gap!!! The example should be proven.

Definition 1.2. Let Ω be a set and $\mathscr{S} \subset P(\Omega)$. Then the topology generated by $\mathscr{S}$ is

$$\mathscr{A} = \{E \subset \Omega : E \text{ is a union of finite intersections of sets in } \mathscr{S}.\}$$

Gap!!! The definition should be more explicit.

Gap!!! Examples and Furstenberg's theorem should be added here

Definition 1.3. Let Ω be a set. An algebra is a collection $\mathscr{A} \subset P(\Omega)$ that is closed under finite unions and compliments with $\Omega \in \mathscr{A}$. If an algebra is also closed under countable unions, we call it a σ-algebra.

Theorem 1.4. *Algebras are closed under finite intersections and σ-algebras are closed under countable intersections.*

Proof. Let $\mathscr{A}$ be an algebra and let $A_1, \dots, A_n \in \mathscr{A}$. Then $A_1^C, \dots, A_n^C \in \mathscr{A}$, so $(\cap_{i=1}^n A_i)^C = \cup_{i=1}^n A_i^C \in \mathscr{A}$, so $\cap_{i=1}^n A_i \in \mathscr{A}$. The proof for σ-algebras is similar. $\square$

Definition 1.5. Let $\mathscr{F}$ be a σ-algebra on a set Ω. A measure on $\mathscr{F}$ is a function $\mu : \mathscr{F} \to [0, \infty]$ such that $\mu(\emptyset) = 0$ and for all disjoint $(A_i)_{i \in \mathbb{N}} \in \mathscr{F}$ we have $\mu(\cup_i A_i) = \sum_i \mu(A_i)$. We call $(\Omega, \mathscr{F}, \mu)$ a measure space.

Gap!!! Add examples of measures and probability measures.

Theorem 1.6. *An intersection of σ-algebras is also a σ-algebra.*

Proof. Let $(\mathscr{A}_i)$ be uncountably many σ-algebras. Let $A \in \cap \mathscr{A}_i$. Then $A^C \in \mathscr{A}_i$ for all $\mathscr{A}_i$, so $A^C \in \cap \mathscr{A}_i$. Similarly, let $(A_j)_{j \in \mathbb{N}} \in \cap \mathscr{A}_i$. Then $\cup_j A_j \in \mathscr{A}_i$ for all $\mathscr{A}_i$, so $\cup_j A_j \in \cap \mathscr{A}_i$. $\square$

Theorem 1.6 motivates the following definition.

Definition 1.7. Let $\mathscr{A} \subset \Omega$. The σ-algebra generated by $\mathscr{A}$ is the intersection of all σ-algebras that contain $\mathscr{A}$. We denote it by $\sigma(\mathscr{A})$.

We can think of $\sigma(\mathscr{A})$ as the smallest σ-algebra containing $\mathscr{A}$.

Notation 1.8. A Borel σ-algebra is a σ-algebra generated by a topology. We will use $\mathscr{B}$ to denote the Borel σ-algebra of $\mathbb{R}$ that is generated by the usual open set topology.

Gap!!! This section needs to be continued with product spaces, the upper/lower continuity of measures, the construction of the Lebesgue measure, the Caratheodory theorem, and the Dynkin uniqueness theorem.

2 Probability spaces

Definition 2.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a measure space. If $\mathbb{P}(\mathcal{F}) = 1$, we call $\mathbb{P}$ a probability measure, we call $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space, and we call an element of $\mathcal{F}$ an event.

Notation 2.2. It is common to use a shorthand notation for events. We often write {condition} to mean $\{\omega \in \Omega \ : \ \text{condition}\}$ and we often write $\mathbb{P}(\text{condition})$ to mean $\mathbb{P}(\{\omega \in \Omega \ : \ \text{condition}\})$.

Theorem 2.3. *Let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of events.*

(a) If $A_1 \subset A_2 \subset A_3 \subset \cdots$ then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\cup_{n=1}^{\infty} A_n)$.

(b) If $A_1 \supset A_2 \supset A_3 \supset \cdots$ then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\cap_{n=1}^{\infty} A_n)$.

Proof.

(a) Let $A = \cup_{n=1}^{\infty} A_n$. Let $B_1 = A_1$ and for $n > 1$ let $B_n = A_n \setminus A_{n-1}$. Then the B_n are disjoint. Now, $A_n = \cup_{i=1}^{n} B_i$, so $\mathbb{P}(A_n) = \sum_{i=1}^{n} \mathbb{P}(B_n)$. Also $A = \cup_{n=1}^{\infty} B_n$ so $\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(B_i)$. This sum converges because the partial sums $\sum_{i=1}^{n} \mathbb{P}(B_i)$ are monotonically increasing and bounded above by 1. The remainder $R_n = \sum_{i=n+1}^{\infty} \mathbb{P}(B_n)$ has $R_n \to 0$ as $n \to \infty$. Finally, we see that $\mathbb{P}(A) - \mathbb{P}(A_n) = R_n \to 0$ as $n \to \infty$. Thus $\mathbb{P}(A_n) \to \mathbb{P}(A)$ as $n \to \infty$.

(b) This follows from the previous part by taking compliments.

$\square$

Definition 2.4. If $(A_i)_{i\in\mathbb{N}}$ are a sequence of events, "A_i occurring infinitely often" is the event

$$\{A_i \text{ i.o.}\} = \bigcap_{m\in\mathbb{N}} \bigcup_{n\geq m} A_n$$

and "A_i occurring eventually" is the event

$$\{A_i \text{ eventually}\} = \bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} A_n.$$

Remark. $\{A_i \text{ i.o.}\}$ is the set of elements $\omega \in \Omega$ that belong in infinitely many of the events (A_i) and $\{A_i \text{ eventually}\}$ is the set of elements $\omega \in \Omega$ that belong in all (A_i) after a certain point. We note that $\{A_i \text{ eventually}\} \subset \{A_i \text{ i.o.}\}$.

Definition 2.5. The characteristic function of an event A in Ω is the function $\mathbf{1}_A : \Omega \to \{0, 1\}$ that has $\mathbf{1}_A(t) = 1$ if $t \in A$ and $\mathbf{1}_A(t) = 0$ if $t \notin A$.

Theorem 2.6. *Let $(A_i)_{i\in\mathbb{N}}$ be a sequence of events.*

(a) $\mathbf{1}_{\{A_i \text{ i.o.}\}} = \limsup_{i\to\infty} \mathbf{1}_{A_i}$.

(b) $\mathbf{1}_{\{A_i \text{ eventually}\}} = \liminf_{i\to\infty} \mathbf{1}_{A_i}$.

Proof.

(a) Suppose $\mathbf{1}_{\{A_i \text{ i.o.}\}}(t) = 1$. Then $t \in \{A_i \text{ i.o.}\}$, so t is in infinitely many A_i; call these A_{α_i}. Then $\mathbf{1}_{A_{\alpha_i}}(t) = 1$ for all i, so $\lim_{i\to\infty} \mathbf{1}_{A_{\alpha_i}}(t) = 1$ and therefore $\limsup_{i\to\infty} \mathbf{1}_{A_i}(t) = 1$.

Now suppose instead $\limsup_{i\to\infty} \mathbf{1}_{A_i}(t) = 1$. Then there exists a subsequence $\{A_{\alpha_i}\}$ of $\{A_i\}$ and an $N \in \mathbb{N}$ so that $\left|\mathbf{1}_{A_{\alpha_i}}(t) - 1\right| < \frac{1}{2}$ for all $i \geq N$. Then for these i it must be that $\mathbf{1}_{A_{\alpha_i}}(t) = 1$ so $t \in A_{\alpha_i}$. Thus t is in infinitely many A_i so $t \in \{A_i \text{ i.o.}\}$ and $\mathbf{1}_{\{A_i \text{ i.o.}\}}(t) = 1$.

(b) Suppose $\mathbf{1}_{\{A_i \text{ eventually}\}}(t) = 1$. Then $t \in \{A_i \text{ eventually}\}$, so there exists an $N \in \mathbb{N}$ such that t is in all A_i for $i \geq N$. For these i we have $\mathbf{1}_{A_i}(t) = 1$, so $\lim_{i\to\infty} \mathbf{1}_{A_i}(t) = 1$. In particular $\liminf_{i\to\infty} \mathbf{1}_{A_i}(t) = 1$.

Now suppose instead $\liminf_{i\to\infty} \mathbf{1}_{A_i}(t) = 1$. Then there is no subsequence $\{A_{\alpha_i}\}$ of $\{A_i\}$ with $\lim_{i\to\infty}\{A_{\alpha_i}\}(t) = 0$, so there are finitely many A_i that do not contain t. It follows that there exists an $N \in \mathbb{N}$ such that for all $i \geq N$, $t \in A_i$. Therefore $t \in \{A_i \text{ eventually}\}$ and $\mathbf{1}_{\{A_i \text{ eventually}\}}(t) = 1$.

$\square$

Gap!!! The above proof could probably be shortened.

Theorem 2.7. *Any events A_n have $\mathbb{P}(A_n \text{ i.o.}) \geq \limsup_{n\to\infty} \mathbb{P}(A_n)$.*

Gap!!! Prove the theorem.

Gap!!! Add the Borel-Cantelli Lemma and the definition of the Borel σ-algebra on $\mathbb{R}$ here.

Definition 2.8. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Events $(A_i)_{i\in I} \in \mathscr{F}$ are called independent if for any finite $J \subset I$ we have $\mathbb{P}\left(\cap_{j\in J} A_j\right) = \prod_{j\in J} \mathbb{P}\left(A_j\right)$.

Theorem 2.9. *If (A_n) are independent and for all n we define either $B_n = A_n$ or $B_n = A_n^C$ then all (B_n) are independent.*

Gap!!! Prove the above theorem.

Gap!!! Add the Goldberg conjecture discussion here, as an example.

Definition 2.10. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Collections $(\mathscr{A}_i) \subset \mathscr{F}$ are called independent if for any sequence $(A_i) \in \mathscr{F}$ with $A_i \in \mathscr{A}_i$ the (A_i) are independent.

Gap!!! Add examples of independent and non-independent collections.

3 Random variables

Definition 3.1. Let $(\Omega_1, \mathscr{F}_1, \mu_1)$ and $(\Omega_2, \mathscr{F}_2, \mu_2)$ be measure spaces. A function $g : \Omega_1 \to \Omega_2$ is called measurable if for all $A \in \mathscr{F}_2$ we have $g^{-1}(A) \in \mathscr{F}_1$.

Remark. This is similar to the topological definition of continuous functions.

Definition 3.2. A random variable on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is a measurable function $\mathsf{X} : \Omega \to \mathbb{R}$, where we use the usual Borel σ-algebra as the σ-algebra on $\mathbb{R}$.

Notation 3.3. When writing down an event involving a random variable X, it is common to write X when we really mean $\mathsf{X}(\omega)$. For example, we might write $\mathbb{P}(\mathsf{X} = 1)$ to mean $\mathbb{P}\left(\{\omega \in \Omega : \mathsf{X}(\omega) = 1\}\right)$.

Gap!!! The notation remark above could be made more clear.

Gap!!! Add some examples of random variables.

Remark. We can form an equivalence relation on random variables by writing $\mathsf{X} \sim \mathsf{Y}$ if $\mathbb{P}(X = Y) = 1$, that is, if $\{\omega \in \Omega : \mathsf{X}(\omega) \neq \mathsf{Y}(\omega)\}$ has measure zero. Equivalent random variables have the same probabilistic properties. Some may prefer to think of "random variables" as these equivalence classes of functions. In this sense, it doesn't make sense to ask what a random variable's value is for a certain $\omega \in \Omega$ with $\mathbb{P}(\{\omega\}) = 0$; it only makes sense to look at the random variable's values on events with measure greater than zero.

Definition 3.4. The distribution function of a random variable X is the function $F_{\mathsf{X}} : \mathbb{R} \to \mathbb{R}$ given by $F(t) = \mathbb{P}(X \leq t)$.

We immediately see that distribution functions are monotonically increasing and right continuous (i.e., $\lim_{t \to s^+} F_{\mathsf{X}}(t) = F_{\mathsf{X}}(s)$ for all $s \in [0, 1)$). Also, F_{X} has a jump discontinuity at s if and only if $\mathbb{P}(\mathsf{X} = s) > 0$ (and if so, the jump is of size $\mathbb{P}(\mathsf{X} = s)$). There is an interesting theorem that provides something like a converse.

Theorem 3.5. *Let $F : \mathbb{R} \to [0, 1]$ be monotonically increasing and right continuous (i.e., $\lim_{t \to s^+} F(t) = F(s)$ for all $s \in [0, 1)$). Then there exists a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a random variable $\mathsf{X} : \Omega \to \mathbb{R}$ such that $F = F_{\mathsf{X}}$.*

Gap!!! Add the two theorem proofs.

Definition 3.6. Given random variables (X_i) let $\mathscr{A}_i = \{\mathsf{X}_i^{-1}(A) : A \in \mathscr{B}\}$. We say that (X_i) are independent if $(\mathscr{A}_i)$ are independent. (Here, $\mathscr{B}$ is the usual Borel σ-algebra on $\mathbb{R}$.)

Theorem 3.7. *Let (X_i) be random variables and let $\mathscr{A}_i = \{\{\mathsf{X}_i \leq t\} : t \in \mathbb{R}\}$. Suppose $(\mathscr{A}_i)$ are independent. Then (X_i) are independent.*

Gap!!! Prove the theorem.

Definition 3.8. The joint distribution function of random variables $\mathsf{X}_1, \ldots, \mathsf{X}_n$ is the function $F : \mathbb{R}^n \to \mathbb{R}$ given by

$$F(\langle t_1, \ldots, t_n \rangle) = \mathbb{P}(\mathsf{X}_i \leq t_i \; \forall \, i \in \{0, \ldots, n\}).$$

Gap!!! Add notes on stochastic domination, percolation, etc.

4 Sequences of random variables

Definition 4.1. Let $\mathsf{X}_n, \mathsf{X} : \Omega \to \mathbb{R}$ be random variables. We say that $\{\mathsf{X}_n\}$ converges almost surely to X and write $\mathsf{X}_n \xrightarrow{\text{a.s.}} \mathsf{X}$ if $\mathbb{P}(\{\omega \in \Omega : \mathsf{X}_n(\omega) \to \mathsf{X}(\omega)\}) = 1$. We say that $\{\mathsf{X}_n\}$ converges in probability to X and write $\mathsf{X}_n \xrightarrow{\mathbb{P}} \mathsf{X}$ if for all $\epsilon > 0$ we have $\mathbb{P}(|\mathsf{X}_n - \mathsf{X}| < \epsilon) \to 1$ as $n \to \infty$.

Definition 4.2. Let $\mathsf{X}_n : \Omega_n \to \mathbb{R}$ and $\mathsf{X} : \Omega \to \mathbb{R}$ be random variables. We say that $\{\mathsf{X}_n\}$ converges in distribution to X and write $\mathsf{X}_n \xrightarrow{\text{dist}} \mathsf{X}$ if $F_{\mathsf{X}_n}(t) \to F_{\mathsf{X}}(t)$ for all $t \in \mathbb{R}$ such that $\mathbb{P}(\mathsf{X} = t) = 0$.

Theorem 4.3. *Let $\mathsf{X}_n, \mathsf{X} : \Omega \to \mathbb{R}$ be random variables.*

(a) If $\mathsf{X}_n \xrightarrow{a.s.} \mathsf{X}$ then $\mathsf{X}_n \xrightarrow{\mathbb{P}} \mathsf{X}$.

(b) If $\mathsf{X}_n \xrightarrow{\mathbb{P}} \mathsf{X}$ then $\mathsf{X}_n \xrightarrow{dist} \mathsf{X}$.

Proof.

(a) Suppose $\mathsf{X}_n \xrightarrow{\text{a.s.}} \mathsf{X}$. Let $S = \{\omega \in \Omega : \mathsf{X}_n(\omega) \to \mathsf{X}(\omega)\}$. Then $\mathbb{P}(S) = 1$.

Let $\epsilon > 0$ and let $A_n^\epsilon = \{|\mathsf{X}_n - X| < \epsilon\}$. For all $\omega \in S$, there exists an N such that $|\mathsf{X}_n(\omega) - X(\omega)| < \epsilon$ for all $n \geq N$; in other words, if $n \geq N$ then $\omega \in A_n^\epsilon$. Thus $S \subset \{A_n^\epsilon \text{ eventually}\}$. Since $\mathbb{P}(S) = 1$, we must have $\mathbb{P}(\{A_n^\epsilon \text{ eventually}\}) = 1$.

...

(b) ...

$\square$

Gap!!! Finish the proof. Add counterexamples to show that converses not true.

Theorem 4.4. *(Uniqueness)*

(a) Suppose $\mathsf{X}_n \xrightarrow{\mathbb{P}} \mathsf{X}$ *and* $\mathsf{X}_n \xrightarrow{\mathbb{P}} \mathsf{Y}$*. Then* $\mathbb{P}(\mathsf{X} = \mathsf{Y}) = 1$*.*

(b) Suppose $\mathsf{X}_n \xrightarrow{dist} \mathsf{X}$ *and* $\mathsf{X}_n \xrightarrow{dist} \mathsf{Y}$*. Then* $F_\mathsf{X} = F_\mathsf{Y}$*.*

Proof.

(a) Let $\epsilon > 0$. We have $|\mathsf{X} - \mathsf{Y}| \leq |\mathsf{X}_n - \mathsf{X}| + |\mathsf{X}_n - \mathsf{Y}|$, so if $|\mathsf{X} - \mathsf{Y}| \geq \epsilon$ then either $|\mathsf{X}_n - \mathsf{X}| \geq \frac{\epsilon}{2}$ or $|\mathsf{X}_n - \mathsf{Y}| \geq \frac{\epsilon}{2}$. Thus

$$\{|\mathsf{X} - \mathsf{Y}| \geq \epsilon\} \subset \left\{|\mathsf{X}_n - \mathsf{X}| \geq \frac{\epsilon}{2}\right\} \cup \left\{|\mathsf{X}_n - \mathsf{Y}| \geq \frac{\epsilon}{2}\right\}$$

so

$$\mathbb{P}(|\mathsf{X} - \mathsf{Y}| \geq \epsilon) \leq \mathbb{P}\left(|\mathsf{X}_n - \mathsf{X}| \geq \frac{\epsilon}{2}\right) + \mathbb{P}\left(|\mathsf{X}_n - \mathsf{Y}| \geq \frac{\epsilon}{2}\right) \to 0$$

as $n \to \infty$. Therefore $\mathbb{P}(|\mathsf{X} - \mathsf{Y}| < \epsilon) = 1$.

Take $A_1 = \{|\mathsf{X} - \mathsf{Y}| < 1\}$ and for $n \geq 2$ take $A_n = A_{n-1} \cap \{|\mathsf{X} - \mathsf{Y}| < \frac{1}{n}\}$. Then all $\mathbb{P}(A_n) = 1$ and $A_1 \supset A_2 \supset A_3 \supset \cdots$. Let $A = \cap_n A_n$. Then by Theorem 2.3, $\mathbb{P}(A) = 1$. Also for all $\omega \in A$ we must have $\mathsf{X}(\omega) = \mathsf{Y}(\omega)$. We conclude that $\mathbb{P}(\mathsf{X} = \mathsf{Y}) = 1$.

(b) From Definition 4.2 it is immediately clear that $F_\mathsf{X} = F_\mathsf{Y}$ except where F_X is discontinuous. This can happen for at most countably many $t \in \mathbb{R}$ (because at each discontinuity there exists a unique rational not in the range of F_X). Then $\{s \in \mathbb{R} : F_\mathsf{X}(s) = F_\mathsf{Y}(s)\}$ is dense everywhere in $\mathbb{R}$. Thus for any $t \in \mathbb{R}$ where F_X is discontinuous, there exist $t_n \in \mathbb{R}$ with $t_n \to t$ and $t_n > t$ and $F_\mathsf{X}(t_n) = F_\mathsf{Y}(t_n)$ for all n. Recalling that F_X and F_Y are right-continuous, we must have $F_\mathsf{X}(t) = \lim_{n\to\infty} F_\mathsf{X}(t_n) = \lim_{n\to\infty} F_\mathsf{Y}(t_n) = F_\mathsf{Y}(t)$. This shows that $F_\mathsf{X} = F_\mathsf{Y}$ everywhere.

$\square$

Theorem 4.5. *Suppose* $\mathsf{X}_n \xrightarrow{\mathbb{P}} \mathsf{X}$*. Then there exists a subsequence* $\{\mathsf{X}_{n_k}\}$ *of* $\{\mathsf{X}_n\}$ *such that* $\mathsf{X}_{n_k} \xrightarrow{a.s.} \mathsf{X}$*.*

Gap!!! Prove the theorem.

Theorem 4.6. *(Skorohod's Representation Theorem)*
Suppose $\mathsf{X}_n \xrightarrow{dist} \mathsf{X}$*. Then there exist* Y_n, Y *such that all* $F_{\mathsf{X}_n} = F_{\mathsf{Y}_n}$ *and* $F_\mathsf{X} = F_\mathsf{Y}$ *and* $\mathsf{Y}_n \xrightarrow{a.s.} \mathsf{Y}$*.*

Gap!!! Prove the theorem.