Advanced Probability

October 10, 2018

Contents

0 Reviews

0.1 Measure spaces

Let E be a set. Let $\mathcal{E}$ be a set of subsets of E. We say that $\mathcal{E}$ is a σ-algebra on E if:
• $\phi \in \mathcal{E}$;
• $\mathcal{E}$ is closed under countable unions and complements.
In that case, $(E, \mathcal{E})$ is called a *measurable space.*

We call the elements of $\mathcal{E}$ *measurable sets.*

Let μ be a function $\mathcal{E} \to [0, \infty]$. We say μ is a measure if:
• $\mu(\phi) = 0$; • μ is countably additive: for all sequences (A_n) of disjoint elements of $\mathcal{E}$, then

$$\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$$

In that case, the triple $(E, \mathcal{E}, \mu)$ is called a *measure space.*

Given a topological space E, there is a smallest σ-algebra containing all the open sets in E. This is the *Borel σ-algebra of $\mathcal{E}$*, denoted $\mathcal{B}(E)$.

In particular, for the real line $\mathbb{R}$, we will just write $\mathcal{B} = \mathcal{B}(\mathbb{R})$ for simplicity.

0.2 Integration of measurable functions

Let $(E, \mathcal{E})$ and $(E', \mathcal{E}')$ be measurable spaces. A function $f : E \to E'$ is *measurable* if $f^{-1}(A) = \{x \in E : f(x) \in A\} \in \mathcal{E} \forall A \in \mathcal{E}'$.

If we refer to a measurable function f without specifying range, the default is $(\mathbb{R}, \mathcal{B})$.

Similarly, if we refer to f as a non-negative measurable function, then we mean $E' = [0, \infty]$, $\mathcal{E}' = \mathcal{B}([0, \infty])$.

It is worth notice that under this set of definitions, a non-negative measurable function might not be $\mathbb{R}$-measurable (since we allowed ∞).

We write $m\mathcal{E}^+$ for set of non-negative measurable functions.

Theorem. Let $(E, \mathcal{E}, \mu)$ be a measure space. There exists a unique map $\tilde{\mu} : m\mathcal{E}^+ \to [0, \infty]$ such that:
•(a) $\tilde{\mu}(1_A) = \mu(A)$ for all $A \in \mathcal{E}$, where 1_A is the indicator function;
•(b) $\tilde{\mu}(\alpha f + \beta g) = \alpha\tilde{\mu}(f) + \beta\tilde{\mu}(g)$ for all $\alpha, \beta \in [0, \infty)$, $f, g \in m\mathcal{E}^+$ (linearity);
•(c) $\tilde{\mu}(f) = \lim_{n\to\infty} \tilde{\mu}(f_n)$ for any non-decreasing sequence $(f_n : n \in \mathbb{N})$ in $m\mathcal{E}^+$ such that $f_n(x) \to f(x)$ for all $x \in E$ (monotone-convergence).

We'll only prove uniqueness. For existence, see II Probability and Measure notes.

From now on, write μ for $\tilde{\mu}$.
We'll call $\mu(f)$ the *integral* of f w.r.t. μ.
We also write $\int_E f d\mu = \int Ef(x)\mu(dx)$.

A *simple function* is a finite linear combination of indicator functions of measurable sets with positive coefficients, i.e. f is simple if

$$f = \sum_{k=1}^{n} \alpha_k 1_{A_k}$$

for some $n \geq 0$, $\alpha_k \in (0, \infty)$, $A_k \in \mathcal{E} \forall k = 1, ..., n$.

From (a) and (b), for f simple,

$$\mu(f) = \ sum_{k=1}^n \alpha_k \mu(A_k)$$

Also, if $f, g \in m\mathcal{E}^+$ with $f \leq g$, then $f + h = g$ where $h = g - f \cdot 1_{f<\infty} \in m\mathcal{E}^+$. Then since $\mu(h) \geq 0$, (b) implies $\mu(f) \leq \mu(g)$.

Take $f \in m\mathcal{E}^+$. Define for $x \in E$, $n \in \mathbb{N}$,

$$f_n(x) = \left(2^{-n} \lfloor 2^n f(x) \rfloor\right) \wedge n$$

where $\wedge$ means taking the minimum. Note that (f_n) is a non-decreasing sequence of simple functions that converges to f pointwise everywhere on E. Then by (c),

$$\mu(f) = \lim_{n\to\infty} \mu(f_n)$$

So we have shown uniqueness: μ is uniquely determined by the measure (provided that it exists, which we're not going to show).

When is $\mu(f)$ zero (for $f \in m\mathcal{E}^+$)? For measurable functions f, g, we say $f = g$ *almost everywhere* if

$$\mu(\{x \in E : f(x) \neq g(x)\}) = 0$$

i.e. they only disagree on a measure-zero set.

We can show, for $f \in m\mathcal{E}^+$, that $\mu(f) = 0$ if and only if $f = 0$ almost everywhere.

Let f be a measurable function. We say that f is *integrable* if $\mu(|f|) < \infty$.

Write $L^1 = L^1(E, \mathcal{E}, \mu)$ for the set of all integrable functions. We extend the integral to L^1 by setting $\mu(f) = \mu(f^+) - \mu(f^-)$, where

$$f^{\pm}(x) = 0 \vee (\pm f(x))$$

where $\vee$ means the maximum (so $f = f^+ - f^-$). Note that now f^+, f^- are both non-negative, with disjoint support. Then we can show that L^1 is a vector space, and $\mu : L^1 \to \mathbb{R}$ is linear.

Lemma. (Fatou's lemma)
Let $(f_n : n \in \mathbb{N})$ be any sequence in $m\mathcal{E}^+$. Then

$$\mu(\liminf_{n\to\infty} f_n) \leq \liminf_{n\to\infty} \mu(f_n)$$

The proof is a straight forward application of monotone convergence.
The only hard part is to remember which way the inequality is (consider a sliding block function to the right).

Theorem. (Dominated convergence)
Let $(f_n : n \in \mathbb{N})$ be a sequence of measurable functions on $(E, \mathcal{E})$. Suppose $f_n(x)$ converges pointwise as $n \to \infty$, with limit $f(x)$ say. Suppose further that $|f_n| \leq g$ for all n, for some integrable function g. Then f_n is integrable for all n, so is f, and $\mu(f_n) \to \mu(f)$ as $n \to \infty$.

Definition. We call a measure space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}(\Omega) = 1$ a *probability space*. In this setting, measurable functions correspond to random variables, measurable sets correspond to events, almost everywhere corresponds to almost surely, and the integral $\mathbb{P}(X)$ corresponds to the expectation $\mathbb{E}(X) = \int_\Omega X d\mathbb{P}$, sometimes written $\mathbb{E}_\mathbb{P}(X)$ if we need to specify the underlying measure.

1 Conditional expectation

Throughout this section we'll use the default probability space $(\Omega\mathcal{F}, \mathbb{P})$.

1.1 The discrete case

Suppose $(G_n : n \in \mathbb{N})$ is a sequence of disjoint set in $\mathcal{F}$ such that $\cup_n G_n = \Omega$ (so a partition of the space Ω). Let X be an integrable random variable. Set $\mathcal{G} = \sigma(G_n : n \in \mathbb{N})$, which in this case is $\{\cup_{n\in I} G_n : I \subseteq \mathbb{N}\}$, i.e. all countable unions of G_n. Define $Y = \sum_{n\in\mathbb{N}} \mathbb{E}(X|G_n)1_{G_n}$, where $\mathbb{E}(X|G_n) = \mathbb{E}(X1_{G_n})/\mathbb{P}(G_n)$, except in the case where $\mathbb{P}(G_n)$ we define LHS to be 0 as well). Now note that Y is $\mathcal{G}$-measurable, is integrable, and $\mathbb{E}(Y1_A) = \mathbb{E}(X1_A)$ for any $A \in \mathcal{G}$. We'll write $Y = \mathbb{E}(X|\mathcal{G})$ almost surely, and say Y is *a version of* conditional expectation of X given $\mathcal{G}$.

1.2 Gaussian case

Let (W, X) be a Gaussian (normal) random variable in $\mathbb{R}^2$. Take a coarser σ-algebra $\mathcal{G}$ generated by W, which is $\{\{W in B\} : B \in \mathcal{B}\}$. Consider for $a, b \in \mathbb{R}$, the random variable $Y = aW + b$. We can choose a, b so that $\mathbb{E}(Y - X) = a\mathbb{E}(W)+b-\mathbb{E}(X) = 0$, and $cov(Y-X, W) = a\, var(W)-cov(X, W) = 0$. Then Y is $\mathcal{G}$-measurable, is integrable, and $\mathbb{E}(Y1_A) = \mathbb{E}(X1_A)$ for all $A \in \mathcal{G}$. To see this, note $Y - X$ and W are independent (as their covariance is 0), and $A = \{W \in B\}$ for some $B \in \mathcal{B}$. So for $A \in \mathcal{G}$, $\mathbb{E}((Y - X)1_A) = \mathbb{E}(Y - X)\mathbb{P}(A) = 0$.

1.3 Conditional density functions

Let (U, V) be a random variable in $\mathbb{R}^2$ with density functoin $f(u, v)$, i.e.

$$\mathbb{P}((U, V) \in A) = \int_A f(u, v)dudv$$

Take $\mathcal{G} = \sigma(U) = \{\{U \in B\} : B \in \mathcal{B}\}$. Take a Borel measurable function h on $\mathbb{R}$ and set $X = h(V)$, assume $X \in L^1(\mathbb{P})$. Note U has density function

$$f(u) = \int_{\mathbb{R}} f(u, v)dv$$

Define the conditional density function

$$f(v|u) = f(u, v)/f(u)$$

where we define $0/0 = 0$.

Now set $Y = g(U)$, where

$$g(u) = \int_{\mathbb{R}} h(v) f(v|u) dv$$

Then g is a Borel-measurable functoin on $\mathbb{R}$ (not obvious), so Y is a $\mathcal{G}$-measurable random variable, and is integrable and for all $A = \{U \in B\} \in \mathcal{G}$, $\mathbb{E}(Y 1_A) = \mathbb{E}(X 1_A)$. To see this,

$$\begin{aligned}
\mathbb{E}(Y 1_A) &= \int_{\mathbb{R}} g(u) 1_B(u) f(u) du \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} h(v) f(v|u) dv 1_B(u) f(u) du \\
&= \mathbb{E}(X 1_A)
\end{aligned}$$

where at the last step we use Fubini's theorem (introduced later) to swap integrals, and note that we can combine $\int f(v|u)f(u)$ to get $f(u, v)$.

1.4 Product measure and Fubini's theorem

Take finite (or countably infinite) measure spaces $(E_1, \mathcal{E}_1, \mu_1)$ and $(E_2, \mathcal{E}_2, \mu_2)$. Write $\mathcal{E}_1 \otimes \mathcal{E}_2$ for the σ-algebra on $E_1 \times E_2$ generated by sets of the form $A_1 \times A_2$ where $A_i \in \mathcal{E}_i$ for $i = 1, 2$. We call $\mathcal{E}_1 \otimes \mathcal{E}_2$ the *product σ-algebra.*

Theorem. There exists a unique measure $\mu = \mu_1 \otimes \mu_2$ on $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ such that

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$$

for all $A_i \in \mathcal{E}_i$ for $i = 1, 2$.

Theorem. (Fubini's theorem)
Let f be a non-negative measurable function $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$. For $x_1 \in E_1$, define in the obvious way

$$f_{x_1}(x_2) = f(x_1, x_2)$$

Then f_{x_1} is $\mathcal{E}_2$-measurable for all $x_1 \in E_1$. Now define $f_1(x_1) = \mu_2(f_{x_1})$. Then f_1 is $\mathcal{E}_1$ measurable and $\mu_1(f_1) = \mu(f)$ (see part II Prob and Measure notes for the integrable case). Define $\hat{f}$ on $E_2 \times E_1$ by

$$\hat{f}(x_2, x_1) = f(x_1, x_2)$$

then we can show $\hat{f}$ is $\mathcal{E}_2 \otimes \mathcal{E}_1$-measurable, and

$$(\mu_2 \otimes \mu_1)(\hat{f}) = (\mu_1 \otimes \mu_2)(f)$$

So by Fubini,

$$\mu_2(f_2) = \hat{f}(\hat{f}) = \mu(f) = \mu_1(f_1)$$

with obvious notations. This means

$$\int_{E_2} \left(\int_{E_1} f(x_1, x_2) \mu_1(dx_1) \right) \mu_2(dx_2) = \int_{E_1} \left(\int_{E_2} f(x_1, x_2) \mu_2(dx_2) \right) \mu_1(dx_1)$$

Note that this also holds for just f integrable.