

Lecture 11: October 31

Lecturer: Persi Diaconis

Scribes: Kenneth Tay

11.1 Product Spaces

Definition 11.1 Let (X, \mathcal{X}) , (Y, \mathcal{Y}) be two measurable spaces. We can define the following:

- the **product space** $X \times Y = \{(x, y) : x \in X, y \in Y\}$,
- **projection mappings** $\pi_x : (x, y) \mapsto x$, $\pi_y : (x, y) \mapsto y$,
- the **product σ -algebra** $\mathcal{X} \times \mathcal{Y} := \sigma(\pi_x, \pi_y)$ (i.e. the σ -algebra generated by the projection mappings),
- the **cylinder sets** $\mathcal{C} = \{\pi_x^{-1}(A) \cup \pi_y^{-1}(B) : A \in \mathcal{X}, B \in \mathcal{Y}\}$,
- the **measurable rectangles** $\mathcal{P} = \{A \times B : A \in \mathcal{X}, B \in \mathcal{Y}\}$,
- $\mathcal{U} = \{\text{all finite disjoint unions of measurable rectangles}\}$.

With the set-up above, we have the following facts (Fact 3 being the most important one):

Proposition 11.2 1. \mathcal{P} is a π -system (and is, in fact, a semi-ring).

2. \mathcal{U} is the field generated by \mathcal{P} .

3. $\mathcal{X} \times \mathcal{Y} = \sigma(\mathcal{C}) = \sigma(\mathcal{P}) = \sigma(\mathcal{U})$.

To simplify notation, we will let $Z := X \times Y$, $\mathcal{Z} := \mathcal{X} \times \mathcal{Y}$.

11.1.1 Sections

Definition 11.3 Let $f : Z \rightarrow W$, where W is some other space. For fixed x , define $f_x : Y \rightarrow W$ where $f_x(y) = f(x, y)$. (We can define f_y similarly.)

For $A \subseteq Z$, define $A_x = \{y : (x, y) \in A\}$, $A_y = \{x : (x, y) \in A\}$.

If you think of A in 2-dimensional space, the A_x and A_y 's correspond to the vertical and horizontal slices of A respectively.

Proposition 11.4 Sectioning “commutes” with set operations:

- $(A_x)^c = (A^c)_x$,
- For $\{A_i\}_{i \in I}$ for arbitrary index set I , $\left(\bigcap_{i \in I} A_i\right)_x = \bigcap_{i \in I} A_x^i$.

Definition 11.5 For fixed $x \in X$, define the **inclusion map** $i_x : Y \rightarrow Z$, where $i_x(y) = (x, y)$.

Note that $f_x(y) = f \circ i_x(y)$, and $A_x = i_x^{-1}(A)$.

The following lemma shows that “sections preserve measurability”:

Lemma 11.6 (Sectioning Lemma) If $f : (Z, \mathcal{Z}) \rightarrow (W, \mathcal{W})$ is measurable, then $f_x : Y \rightarrow W$ is measurable.

If $A \in \mathcal{Z}$, then A_x, A_y are measurable.

Proof: Note that, for fixed x ,

- For all y , $\pi_x \circ i_x(y) \equiv x$. Thus $\pi_x \circ i_x(y)$ is a constant map, and so is measurable.
- $\pi_y \circ i_x(y) = y$ is the identity map and so is measurable.

Since the composition of i_x with each of the projections is measurable, therefore, for fixed x , $i_x : Y \rightarrow Z$ is measurable w.r.t. the product space.

Hence, $f_x = f \circ i_x$ and $A_x = i_x^{-1}(A)$ are measurable as well. ■

11.2 Kernels

Definition 11.7 Let (X, \mathcal{X}) , (Y, \mathcal{Y}) be measurable spaces. A **Markov kernel** is a function $K(x, B) : (X, \mathcal{X}) \rightarrow [0, 1]$ such that

1. For every fixed x , $K(x, \cdot)$ is a probability on (Y, \mathcal{Y}) , and
2. For every fixed B , the map $x \mapsto K(x, B)$ is measurable.

Here are some examples of kernels:

- Fix a probability μ on (Y, \mathcal{Y}) , and let $K(x, B) = \mu(B)$ for all x .
- Let $X = \{0, 1\}$, and let μ_0 and μ_1 be 2 probabilities on (Y, \mathcal{Y}) . Define $K(0, B) = \mu_0(B)$, $K(1, B) = \mu_1(B)$. These are **mixture models**.
- Let Θ be any set with an associated σ -algebra \mathcal{F}_Θ , and let (Y, \mathcal{Y}) be a measurable space. Let $\{P_\theta(dy)\}_{\theta \in \Theta}$ be a family of probability distributions on (Y, \mathcal{Y}) . Then $K(\theta, dy) = P_\theta(dy)$ is a Markov kernel. (This is the usual set-up in theoretical statistics.)
- Let $X = Y$. The **transition operator** for a Markov chain on X , $K(x, B)$, is the “chance” of landing in B after 1 step given that you start at x .

Then $K^2(x, B) = \int K(y, B)K(x, dy)$ is the chance of landing in B after 2 steps if starting at x .

Definition 11.8 Let μ be a probability on (X, \mathcal{X}) . Define a probability $\mu \times K$ on (Z, \mathcal{Z}) by

$$\mu \times K(A) := \int K(x, A_x) \mu(dx).$$

In order for this definition to make sense, we need to show 2 things:

1. If $A \in \mathcal{Z}$, then the function $x \mapsto k(x, A_x)$ is measurable. (If it isn't, the integral doesn't make sense.)

Proof: Let $\mathcal{G} = \{C \in \mathcal{Z} : \text{the function } x \mapsto k(x, C_x) \text{ is measurable}\}$. For a measurable rectangle $A \times B$,

$$K(x, (A \times B)_x) = \delta_A(x)K(x, B)$$

which is measurable. So \mathcal{G} contains the π -system of measurable rectangles.

We claim that \mathcal{G} is a λ -system:

- Clearly $X \times Y$ is a measurable rectangle, and so is in \mathcal{G} .
- If $C \in \mathcal{G}$, then $K(x, (C^c)_x) = K(x, (C_x)^c) = 1 - K(x, C_x)$, so \mathcal{G} closed under complements.
- if $\{C^i\}$, $i = 1, 2, \dots$ are disjoint sets in \mathcal{G} , then

$$K\left(x, \left(\bigcup_{i=1}^{\infty} C^i\right)_x\right) = K\left(x, \bigcup_{i=1}^{\infty} C_x^i\right) = \sum_{i=1}^{\infty} K(x, C_x^i),$$

which is measurable.

As such, \mathcal{G} is a λ -system which contains π -system of measurable rectangles. By the $\pi - \lambda$ Theorem, we have $\mathcal{G} = \mathcal{Z}$. ■

2. $\mu \times K$ is a probability on (Z, \mathcal{Z}) .

Proof:

- Since K is a non-negative function, $\mu \times K(A)$ is certainly non-negative for all A .
- $\mu \times K(\emptyset) = \int K(x, \emptyset) \mu(dx) = 0$, and $\mu \times K(Z) = \int K(x, Y) \mu(dx) = 1$.
- If $\{A^i\}$, $i = 1, 2, \dots$ are disjoint sets in \mathcal{Z} , then

$$\begin{aligned} \mu \times K\left(\bigcup_{i=1}^{\infty} A^i\right) &= \int K\left(x, \left(\bigcup_{i=1}^{\infty} A^i\right)_x\right) \mu(dx) \\ &= \int K\left(x, \bigcup_{i=1}^{\infty} A_x^i\right) \mu(dx) \\ &= \int \sum_{i=1}^{\infty} K(x, A_x^i) \mu(dx) \\ &= \sum_{i=1}^{\infty} \int K(x, A_x^i) \mu(dx) \\ &= \sum_{i=1}^{\infty} \mu \times K(A_i). \end{aligned}$$
■

The probability measure $\mu \times K$ can be interpreted in the following way: first pick x from μ , then pick y from $K(x, \cdot)$. Then the chance that (x, y) is contained in $A \times B$ is $\mu \times K(A \times B) = \int_A K(x, B) \mu(dx)$

Note that $\mu \times K$ is the only probability on (Z, \mathcal{Z}) satisfying the property above.

11.3 Fubini's Theorem

Consider the following set-up:

- (X, \mathcal{X}) , (Y, \mathcal{Y}) measurable spaces, with (Z, \mathcal{Z}) being the product space,
- μ a probability on (X, \mathcal{X}) ,
- $K(x, dy)$ a Markov kernel on (X, \mathcal{Y}) .

From the previous section, we know that $\mu \times K(A) = \int K(x, A_x) \mu(dx)$ is the unique probability on (Z, \mathcal{Z}) such that $\mu \times K(A \times B) = \int_A K(x, B) \mu(dx)$.

Theorem 11.9 (Fubini's Theorem) *If $f : Z \rightarrow \mathbb{R}$ is Borel-measurable and non-negative, then*

1. $x \mapsto \int f(x, y) K(x, dy)$ is \mathcal{X} -measurable, and

2.

$$\int f d(\mu \times K) = \int_X \left[\int_Y f(x, y) K(x, dy) \right] \mu(dx).$$

Proof:

Let $\mathcal{G} = \{f : Z \rightarrow \mathbb{R}_+ \text{ measurable, statements 1 and 2 are true}\}$.

Step 1: Indicator functions.

Let $f(x, y) = \delta_A(x, y)$ for $A \in \mathcal{Z}$. Then $\int \delta_A(x, y) K(x, dy) = K(x, A_x)$ which is measurable. By definition,

$$\int \delta_A d(\mu \times K) = \mu \times K(A) = \int K(x, A_x) \mu(dx).$$

Hence $\delta_A \in \mathcal{G}$.

Step 2: Simple functions. If $f, g \in \mathcal{G}$, then using the facts that (i) linear combinations of measurable functions are measurable, and (ii) integrals are linear, $af + bg \in \mathcal{G}$. Therefore all simple functions are in \mathcal{G} .

Step 3: All non-negative functions. If $f_n \in \mathcal{G}$, f_n simple and $f_n \nearrow f$, then $\lim f_n$ is measurable, and $\lim \int f_n(x, y) K(x, dy) = \int f(x, y) K(x, dy)$. So $f \in \mathcal{G}$ as well. Since any measurable non-negative f has a sequence of simple functions $\{f_n\}$ such that $f_n \nearrow f$, all non-negative measurable functions are in \mathcal{G} . ■

Two important special cases of Fubini's Theorem:

1. **Tonelli's Theorem:** For each x , let $K(x, B) = \nu(B)$ where ν is a probability on \mathcal{Y} . Then $\mu \times \nu$ is called the product measure, and for $f \geq 0$,

$$\int f d(\mu \times \nu) = \int_X \left[\int f(x, y) \nu(dy) \right] \mu(dx).$$

2. (Bayesian statistics) Suppose $\{P_\theta(dy)\}_{\theta \in \Theta}$ is a family of probabilities on (Y, \mathcal{Y}) . Let Θ have an associated σ -algebra \mathcal{F}_Θ , and let $\pi(d\theta)$ be a probability on \mathcal{F}_Θ (i.e. the prior distribution). These define a joint measure $\pi \times P$ on $\Theta \times \mathcal{Y}$. Let $m(B) = \int_\Theta P_\theta(B) \pi(d\theta)$ (i.e. the marginal distribution).

With this set-up, we can define the **posterior** as a Markov kernel $K(y, d\theta)$ which splits $\pi \times P$ the other way (via $m(dy)$), i.e. for every (A, B) ,

$$\int_A P_\theta(B) \pi(d\theta) = \int_B K(y, A) m(dy).$$