STATS310A - Lecture 10

Persi Diaconis
Scribed by Michael Howes

10/21/21

Contents

1 Product σ-algebras

Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be two measurable spaces. Let $X \times Y$ be the product *set* $X \times Y = \{(x, y) : x \in X, y \in Y\}$. Define the *projections* $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ by

$$\pi_X(x, y) = x \text{ and } \pi_Y(x, y) = y.$$

Definition 1. The *product σ-algebra* is the small σ-algebra on $X \times Y$ making π_X and π_Y measurable. We denote the product σ-algebra by $\mathcal{X} \times \mathcal{Y}$.

Definition 2. The *cyclinder sets* are sets of the form $\pi_X^{-1}(A)$ for $A \in \mathcal{X}$ or $\pi_Y^{-1}(B)$ for $B \in \mathcal{Y}$. We denote the class of cylinder sets by $\mathcal{C}$.

Definition 3. Let $\mathcal{P} = \{A \times B : A \in \mathcal{X}, B \in \mathcal{Y}\}$ be the class of *measurable rectangles*.

Note that $\mathcal{P}$ is a π-system and indeed a semi-ring. Define $\mathcal{U}$ to the set of finite disjoint unions of measurable rectangles. The collection U is a field.

Proposition 1. *With the notation as above*

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

Definition 4. If $A \subset X \times Y$ and $x \in X$, define

$$A_x = \{y : (x, y) \in A\} \subseteq Y.$$

For $y \in Y$, define

$$A_y = \{x : (x, y) \in A\} \subseteq X.$$

The sets A_x and A_y are called *sections* of A.

Definition 5. For a function $F : X \times Y \to W$, define $f_x : Y \to W$ and $f_y : X \to W$ by

$$f_x(y) = f(x, y) \text{ and } f_y(x) = f(x, y).$$

The maps f_x and f_y are again called *sections* of f.

Proposition 2. *Sections commute with set operations. That is*

- $(A^c)_x = A^c_x$,
- $\left(\bigcap_{i\in I} A^i\right)_x = \bigcap_{i\in I} A^i_x$,
- $\left(\bigcup_{i\in I} A^i\right)_x = \bigcup_{i\in I} A^i_x$,

where I is any index set.

Proposition 3. *If $A \in \mathcal{X} \times Y$, then $A_x \in \mathcal{Y}$ for all $x \in X$. If $f : X \times Y \to (W, \mathcal{F})$ is measurable, then $f_x : Y \to (W, \mathcal{F})$ and $f_y : X \to (W, \mathcal{F})$ are also measurable.*

Proof. Consider the collection

$$G = \{A \in \mathcal{X} \times \mathcal{Y} : A_x \in \mathcal{Y}\}.$$

Note that G contains the measurable rectangles since

$$(R \times S)_x = \begin{cases} \emptyset & \text{if } x \notin R, \\ S & \text{if } x \in R. \end{cases}$$

Thus in either case $(R \times S)_x \in \mathcal{Y}$. Since sections commute with set operations, G is a σ-algebra. Thus $\sigma(\mathcal{P}) = \mathcal{X} \times \mathcal{Y} \subseteq G$, as required.

Let A be a measurable subset of W. Then

$$\begin{aligned} f_x^{-1}(A) &= \{y : f_x(y) \in A\} \\ &= \{y : f(x,y) \in A\} \\ &= \{y : (x,y) \in f^{-1}(A)\} \\ &= (f^{-1}(A))_x. \end{aligned}$$

Since $f^{-1}(A) \in \mathcal{X} \times \mathcal{Y}$, we can conclude that $f_x^{-1}(A) = (f^{-1}(A))_x \in \mathcal{Y}$. $\square$

2 Measures on product spaces

Definition 6. Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be measurable spaces. A *Markov Kernel* is a function $K : X \times \mathcal{Y} \to [0,1]$ such that

(a) For all $x \in X$, $K(x, \cdot)$ is a probability measure on $(Y, \mathcal{Y})$.

(b) For all $B \in \mathcal{Y}$, $K(\cdot, B)$ is measurable.

We will write $K(x, dy)$ to mean that K is a Markov kernel $K : X \times \mathcal{Y} \to [0,1]$.

Example 1. Say ν is a probability measure on $(Y, \mathcal{Y})$, then $K(x, B) = \nu(B)$ is a Markov Kernel.

Example 2. If Θ is any set and $\mathcal{F}$ a σ-algebra on Θ, then a family of probabilities $\{\mathbb{P}_\theta(\cdot) : \theta \in \Theta\}$ on $(X, \mathcal{X})$ is a Markov kernel

$$K(\theta, B) = \mathbb{P}_\theta(B).$$

Example 3. If $X = Y$, then $k(x, dy)$ defines a *Markov chain* on X.

Definition 7. Let $(X, \mathcal{X})$ and $(Y, \mathcal{Y})$ be measurable spaces and $K(x, dy)$ a kernel and μ a probability on X. The product measure $\mu \times K$ is a set function on $\mathcal{X} \times \mathcal{Y}$ defined by

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

Proposition 4. *The mapping $x \mapsto K(x, A_x)$ is measurable and integrable. Furthermore $\mu \times K$ is a probability on $X \times Y$.*

Proof. Define
$$G = \{A \in \mathcal{X} \times \mathcal{Y} : x \mapsto K(x, A_x) \text{ is measurable}\}.$$
Note that G contains the measurable rectangles. This is because

$$K(x, (S \times R)_x) = \begin{cases} 0 & \text{if } x \notin S, \\ K(x, R) & \text{if } x \in S. \end{cases}$$
$$= \delta_S(x) K(x, R).$$

Thus $x \mapsto K(x, (S \times R)_x)$ is the product of two measurable functions and hence measurable. Thus G continas the π-system $\mathcal{P}$. We will now show that G is a λ-system. Note that $X \times Y \in G$, since $X \times Y \in \mathcal{P}$. Furthermore if $A \in G$, then

$$K(x, (A^c)_x) = K(x, A_x^c) = 1 - K(x, A_x),$$

and so $A^c \in G$. Finally if $(A^i)_{i=1}^\infty$ are disjoint, then $(A_x^i)_{i=1}^\infty$ are disjoint and hence

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

and thus $\bigcup_i A^i \in G$ since the limits of measurable functions are measurable. Thus G is a λ-system and it must contain $\sigma(\mathcal{P}) = \mathcal{X} \times \mathcal{Y}$ by the π-λ theorem.

To see that $\mu \times K$ is a probability measure one can use the monotone convergence theorem. □

Example 4. If $K(x, B) = \nu(B)$ then we write $\mu \times K$ as $\mu \times \nu$ and call $\mu \times \nu$ the product measure.

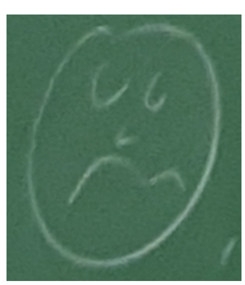
Example 5 (Decision theory/Bayesian statistics)**.** Given probability distributions $P = \{\mathbb{P}_\theta(\cdot)\}_{\theta \in \Theta}$ on $(X, \mathcal{X})$ and a probability π on Θ, $\pi \times P$ defines a probability on $\Theta \times X$. Define

$$m(B) = \int_\Theta \mathbb{P}_\theta(B) \pi(d\theta),$$

which is a probability distribution on $(X, \mathcal{X})$ called the *marginal distribution.* A *posterior* is a kernel $K(x, d\theta)$ on $X \times \mathcal{F}_\theta$ such that

$$\int_A P_\theta(B) \pi(d\theta) = \int_B K(x, A) \pi(dx),$$

for all $A \in \mathcal{F}_\theta$ and $B \in \mathcal{X}$. Unfortunately posteriors don't always exist.



We need topological conditions on X to be sure that posteriors exist (eg it suffices for X to be a complete seperable metric space). When things work out the objects of study are called regular conditional probabilities.

3 Fubinni's Theorem

Theorem 1. *Let* $(X, \mathcal{X})$ *and* $(Y, \mathcal{Y})$ *be measurable spaces. Let* $\mu(dx)$ *be a measure and* $K(x, dy)$ *be a kernel. The if* $f : X \times Y \to [0, \infty]$ *is measurable, then*

$$x \mapsto \int_Y f(x,y)K(x,dy),$$

is measurable on $(X, \mathcal{X})$ *and*

$$\int_{X\times Y} f(x,y)(\mu \times K)(dx,dy) = \int_X \left(\int_Y f(x,y)K(x,dy) \right) \mu(dx).$$

Proof. We will use a (1), (2), (3) argument. Let G be the set of all measurable $f : X \times Y \to \mathbb{R}^+$ such that the above two results hold. Suppose that $A \in \mathcal{X} \times Y$ and $f = \delta_A$. Then note that $\delta_A(x,y) = \delta_{A_x}(y)$ and so

$$\int_Y \delta_A(x,y)K(x,dy) = \int_Y \delta_{A_x}(y)K(x,dy) = K(x, A_x),$$

which is measurable. And furthermore

$$\begin{aligned}
\int_{X\times Y} \delta_A(x,y)(\mu \times K)(dx,dy) &= (\mu \times K)(A) \\
&= \int_X K(x, A_x)\mu(dx) \\
&= \int_X \left(\int_Y \delta_A(x,y)K(x,dy) \right) \mu(dx).
\end{aligned}$$

Thus $\delta_A \in G$. One can check that G is closed under linear combinations and monotone limits. Thus G contains all non-negative measurable functions. □

Remark 1. (a) We assumed $K(\cdot, B)$ and $\mu(\cdot)$ where probability measures. Everything works under the more general assumption that $K(\cdot, B)$ and $\mu(\cdot)$ are σ-finite.

(b) The textbook carefully works through the case when $K(x, dy) = \nu(dy)$.

(c) When applying Fubinni's theorem look out for functions that are both positive and negative. Everything works if $\int |f|(\mu \times K)(dx, dy) < \infty$.

(d) These results do not hold for finitely additive measures or non σ-finite measures.

(e) Measures on infinite products require care. You again need topology to deal with something like

$$\mu(x_1), K(x_1 dx_2), L((x_1, x_2), dx_3), \ldots$$