High-Dimensional Probability and Statistics

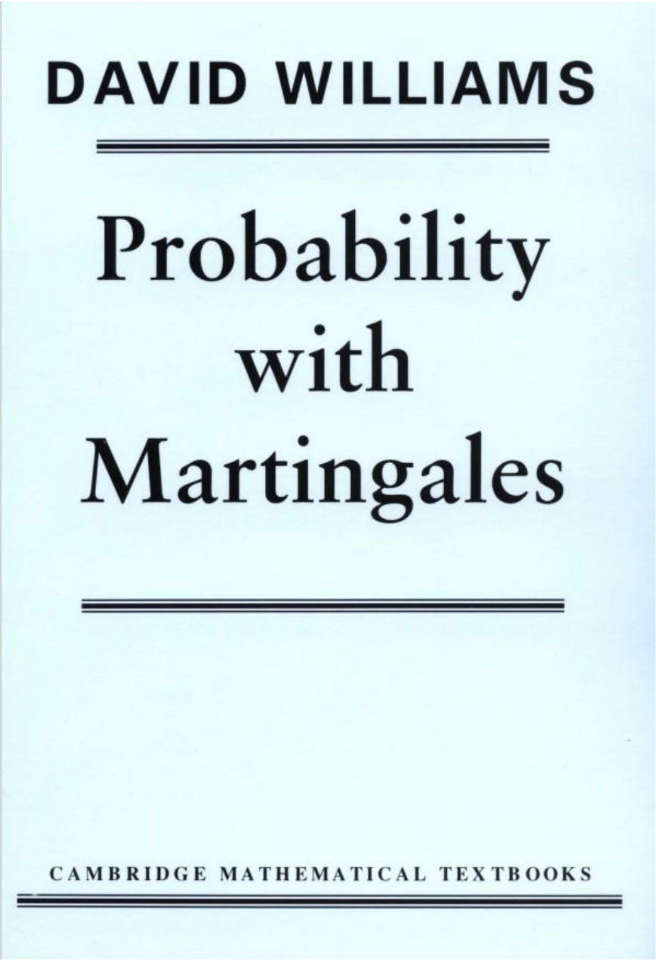
MATH/STAT/ECE 888: Topics in Mathematical Data Science

Sebastien Roch (Math+Stat)
UW-Madison

Fall 2021

Lecture 9 (09/27/21)

Today’s slides based on Williams (but results can be found in any graduate-level probability textbook)



See e.g. https://people.math.wisc.edu/~roch/grad-prob/

Facts About Conditional Expectation

Conditional Expectation: Definition

9.2. Fundamental Theorem and Definition (Kolmogorov, 1933)

► *Let* $(\Omega, \mathcal{F}, \mathsf{P})$ *be a triple, and* X *a random variable with* $\mathsf{E}(|X|) < \infty$. *Let* $\mathcal{G}$ *be a sub-*σ*-algebra of* $\mathcal{F}$. *Then there exists a random variable* Y *such that*

(a) Y *is* $\mathcal{G}$ *measurable,*

(b) $\mathsf{E}(|Y|) < \infty$,

(c) *for every set* G *in* $\mathcal{G}$ *(***equivalently, for every set** G **in some** π**-system which contains** Ω **and generates** $\mathcal{G}$*), we have*

$$\int_G Y \, d\mathsf{P} = \int_G X \, d\mathsf{P}, \qquad \forall G \in \mathcal{G}.$$

Moreover, if $\tilde{Y}$ *is another RV with these properties then* $\tilde{Y} = Y$, a.s., that is, $\mathsf{P}[\tilde{Y} = Y] = 1$. *A random variable* Y *with properties* (a)-(c) *is called* **a version of the conditional expectation** $\mathsf{E}(X|\mathcal{G})$ *of* X *given* $\mathcal{G}$, *and we write* $Y = \mathsf{E}(X|\mathcal{G})$, a.s.

Conditional Expectation: Definition cont'd

9.6. Agreement with traditional usage

The case of two RVs will suffice to illustrate things. So suppose that X and Z are RVs which have a joint probability density function (pdf)

$$f_{X,Z}(x,z).$$

Then $f_Z(z) = \int_{\mathbf{R}} f_{X,Z}(x,z)dx$ acts as a probability density function for Z. Define the *elementary conditional pdf* $f_{X|Z}$ of X given Z via

$$f_{X|Z}(x|z) := \begin{cases} f_{X,Z}(x,z)/f_Z(z) & \text{if } f_Z(z) \neq 0; \\ 0 & \text{otherwise.} \end{cases}$$

Let h be a Borel function on $\mathbf{R}$ such that

$$\mathsf{E}|h(X)| = \int_{\mathbf{R}} |h(x)| f_X(x)dx < \infty,$$

where of course $f_X(x) = \int_{\mathbf{R}} f_{X,Z}(x,z)dz$ gives a pdf for X. Set

$$g(z) := \int_{\mathbf{R}} h(x) f_{X|Z}(x|z)dx.$$

Then $Y := g(Z)$ *is a version of the conditional expectation of* $h(X)$ *given* $\sigma(Z)$.

Conditional Expectation: Definition cont'd

Proof. The typical element of $\sigma(Z)$ has the form $\{\omega : Z(\omega) \in B\}$, where $B \in \mathcal{B}$. Hence, we must show that

(a) $$L := \mathsf{E}[h(X)\mathrm{I}_B(Z)] = \mathsf{E}[g(Z)\mathrm{I}_B(Z)] =: R.$$

But

$$L = \int\int h(x)\mathrm{I}_B(z)f_{X,Z}(x,z)dxdz, \quad R = \int g(z)\mathrm{I}_B(z)f_Z(z)dz,$$

and result (a) follows from Fubini's Theorem. □

Independence

Let X and Y be two random variables. The *(joint) law* $\mathcal{L}_{X,Y}$ of the pair (X,Y) is the map

$$\mathcal{L}_{X,Y} : \mathcal{B}(\mathbf{R}) \times \mathcal{B}(\mathbf{R}) \to [0,1]$$

defined by

$$\mathcal{L}_{X,Y}(\Gamma) := \mathbf{P}[(X,Y) \in \Gamma].$$

8.4. Independence and product measure

Let X and Y be two random variables with laws $\mathcal{L}_X, \mathcal{L}_Y$ respectively and distribution functions F_X, F_Y respectively. Then the following three statements are equivalent:

(i) *X and Y are independent;*

(ii) $\mathcal{L}_{X,Y} = \mathcal{L}_X \times \mathcal{L}_Y$;

(iii) $F_{X,Y}(x,y) = F_X(x)F_Y(y)$;

moreover, if (X,Y) has 'joint' pdf $f_{X,Y}$ then each of (i)-(iii) is equivalent to

(iv) $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ *for* Leb $\times$ Leb *almost every* (x,y).

You do not wish to know more about this either.

Conditional expectation as least squares

9.4. Conditional expectation as least-squares-best predictor

► *If* $\mathsf{E}(X^2) < \infty$, *then the conditional expectation* $Y = \mathsf{E}(X|\mathcal{G})$ *is a version of the orthogonal projection (see Section 6.11) of* X *onto* $\mathcal{L}^2(\Omega, \mathcal{G}, \mathsf{P})$. *Hence,* Y *is the least-squares-best* $\mathcal{G}$*-measurable predictor of* X*: amongst all* $\mathcal{G}$*-measurable functions (i.e. amongst all predictors which can be computed from the available information),* Y *minimizes*

$$\mathsf{E}[(Y - X)^2].$$

Conditional Expectation Cheat Sheet

These properties are proved in Section 9.8. All X's satisfy $\mathsf{E}(|X|) < \infty$ in this list of properties. Of course, $\mathcal{G}$ and $\mathcal{H}$ denote sub-σ-algebras of $\mathcal{F}$. (The use of 'c' to denote 'conditional' in (cMON), etc., is obvious.)

(a) If Y is any version of $\mathsf{E}(X|\mathcal{G})$ then $\mathsf{E}(Y) = \mathsf{E}(X)$. (*Very* useful, this.)

(b) If X is $\mathcal{G}$ measurable, then $\mathsf{E}(X|\mathcal{G}) = X$, a.s.

(c) **(Linearity)** $\mathsf{E}(a_1 X_1 + a_2 X_2|\mathcal{G}) = a_1\mathsf{E}(X_1|\mathcal{G}) + a_2\mathsf{E}(X_2|\mathcal{G})$, a.s.
Clarification: if Y_1 is a version of $\mathsf{E}(X_1|\mathcal{G})$ and Y_2 is a version of $\mathsf{E}(X_2|\mathcal{G})$, then $a_1 Y_1 + a_2 Y_2$ is a version of $\mathsf{E}(a_1 X_1 + a_2 X_2|\mathcal{G})$.

(d) **(Positivity)** If $X \geq 0$, then $\mathsf{E}(X|\mathcal{G}) \geq 0$, a.s.

(e) **(cMON)** If $0 \leq X_n \uparrow X$, then $\mathsf{E}(X_n|\mathcal{G}) \uparrow \mathsf{E}(X|\mathcal{G})$, a.s.

(f) **(cFATOU)** If $X_n \geq 0$, then $\mathsf{E}[\liminf X_n|\mathcal{G}] \leq \liminf \mathsf{E}[X_n|\mathcal{G}]$, a.s.

(g) **(cDOM)** If $|X_n(\omega)| \leq V(\omega)$, $\forall n$, $\mathsf{E}V < \infty$, and $X_n \to X$, a.s., then

$$\mathsf{E}(X_n|\mathcal{G}) \to \mathsf{E}(X|\mathcal{G}), \quad \text{a.s.}$$

Conditional Expectation Cheat Sheet cont'd

(h) **(cJENSEN)** If $c : \mathbb{R} \to \mathbb{R}$ is convex, and $\mathsf{E}|c(X)| < \infty$, then

$$\mathsf{E}[c(X)|\mathcal{G}] \geq c(\mathsf{E}[X|\mathcal{G}]), \quad \text{a.s.}$$

Important corollary: $\|\mathsf{E}(X|\mathcal{G})\|_p \leq \|X\|_p$ for $p \geq 1$.

(i) **(Tower Property)** If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then

$$\mathsf{E}[\mathsf{E}(X|\mathcal{G})|\mathcal{H}] = \mathsf{E}[X|\mathcal{H}], \qquad \text{a.s.}$$

Note. We shorthand LHS to $\mathsf{E}[X|\mathcal{G}|\mathcal{H}]$ for tidiness.

(j) **('Taking out what is known')** If Z is $\mathcal{G}$-measurable and bounded, then

$$(*) \qquad \mathsf{E}[ZX|\mathcal{G}] = Z\mathsf{E}[X|\mathcal{G}], \qquad \text{a.s.}$$

If $p > 1$, $p^{-1} + q^{-1} = 1$, $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathsf{P})$ and $Z \in \mathcal{L}^q(\Omega, \mathcal{G}, \mathsf{P})$, then (*) again holds. If $X \in (\mathrm{m}\mathcal{F})^+$, $Z \in (\mathrm{m}\mathcal{G})^+$, $\mathsf{E}(X) < \infty$ and $\mathsf{E}(ZX) < \infty$, then (*) holds.

(k) **(Rôle of independence)** If $\mathcal{H}$ is independent of $\sigma(\sigma(X), \mathcal{G})$, then

$$\mathsf{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathsf{E}(X|\mathcal{G}), \quad \text{a.s.}$$

In particular, if X is independent of $\mathcal{H}$, then $\mathsf{E}(X|\mathcal{H}) = \mathsf{E}(X)$, a.s.

A Useful Fact [From Vershynin]

Lemma 6.1.2. *Let Y and Z be independent random variables such that $\mathbb{E}\,Z = 0$. Then, for every convex function F, one has*

$$\mathbb{E}\,F(Y) \leq \mathbb{E}\,F(Y + Z).$$

Proof This is a simple consequence of Jensen's inequality. First let us fix an arbitrary $y \in \mathbb{R}$ and use $\mathbb{E}\,Z = 0$ to get

$$F(y) = F(y + \mathbb{E}\,Z) = F(\mathbb{E}[y + Z]) \leq \mathbb{E}\,F(y + Z).$$

Now choose $y = Y$ and take expectations of both sides to complete the proof. (To check if you understood this argument, find where the independence of Y and Z was used!) $\square$