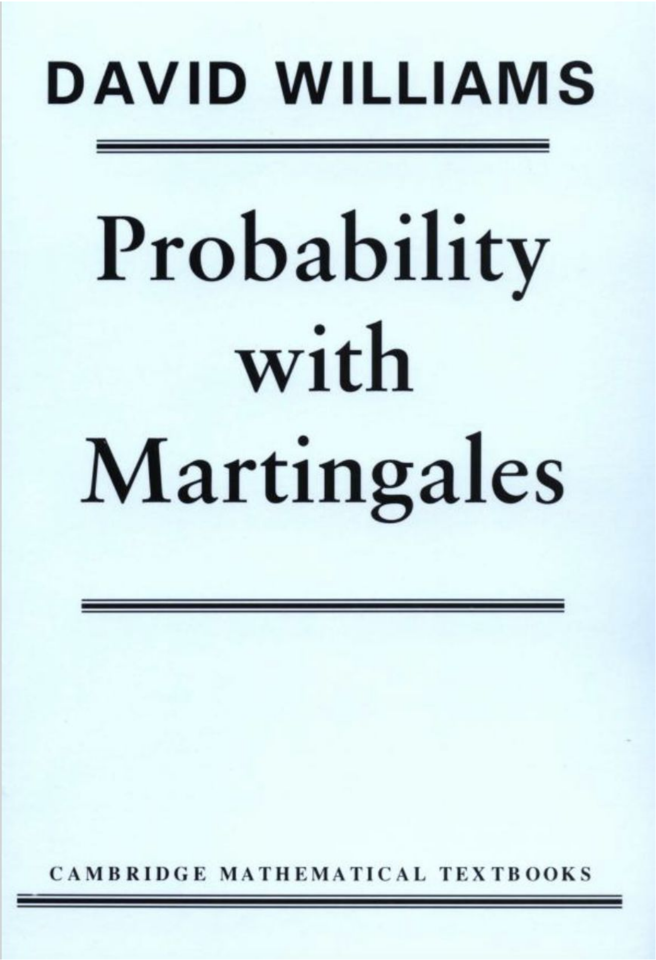
High-Dimensional Probability and Statistics

MATH/STAT/ECE 888: Topics in Mathematical Data Science
Sebastien Roch (Math+Stat)
UW-Madison
Fall 2021

Lecture 3 (09/13/21) & Lecture 4 (09/15/21)

Course website

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/

Important Probability Facts

Markov's inequality

6.4. Markov's inequality

Suppose that $Z \in \mathrm{m}\mathcal{F}$ *and that* $g : \mathbf{R} \to [0, \infty]$ *is* $\mathcal{B}$*-measurable and non-decreasing. (We know that* $g(Z) = g \circ Z \in (\mathrm{m}\mathcal{F})^+$*.) Then*

► $$\mathsf{E}g(Z) \geq \mathsf{E}(g(Z); Z \geq c) \geq g(c)\mathsf{P}(Z \geq c).$$

Examples: for $Z \in (\mathrm{m}\mathcal{F})^+$, $\quad c\mathsf{P}(Z \geq c) \leq \mathsf{E}(Z), \qquad (c > 0),$

for $X \in \mathcal{L}^1$, $\quad c\mathsf{P}(|X| \geq c) \leq \mathsf{E}(|X|) \quad (c > 0).$

► *Considerable strength can often be obtained by choosing the optimum* θ *for* c *in*

► $$\mathsf{P}(Y > c) \leq e^{-\theta c}\mathsf{E}(e^{\theta Y}), \qquad (\theta > 0, \quad c \in \mathbf{R}).$$

7.3. Chebyshev's inequality

As you know this says that *for* $c \geq 0$*, and* $X \in \mathcal{L}^2$,

$$c^2\mathsf{P}(|X - \mu| > c) \leq \mathrm{Var}(X), \qquad \mu := \mathsf{E}(X);$$

and it is obvious.

Lp norm

6.7. Monotonicity of $\mathcal{L}^p$ norms

►►For $1 \leq p < \infty$, we say that $X \in \mathcal{L}^p = \mathcal{L}^p(\Omega, \mathcal{F}, \mathbf{P})$ if

$$\mathsf{E}(|X|^p) < \infty,$$

and then we define

► $$\|X\|_p := \{\mathsf{E}(|X|^p)\}^{\frac{1}{p}}.$$

The monotonicity property referred to in the section title is the following:

►(a) *if* $1 \leq p \leq r < \infty$ *and* $Y \in \mathcal{L}^r$, *then* $Y \in \mathcal{L}^p$ *and*

$$\|Y\|_p \leq \|Y\|_r.$$

Cauchy-Schwarz inequality

6.8. The Schwarz inequality

►(a) *If* X *and* Y *are in* $\mathcal{L}^2$*, then* $XY \in \mathcal{L}^1$*, and*

$$|\mathsf{E}(XY)| \leq \mathsf{E}(|XY|) \leq \|X\|_2 \|Y\|_2.$$

The following is an immediate consequence of (a):

(b) *if* X *and* Y *are in* $\mathcal{L}^2$*, then so is* $X + Y$*, and we have the triangle law:*

$$\|X + Y\|_2 \leq \|X\|_2 + \|Y\|_2.$$

Modes of convergence

A13.1. Modes of convergence: definitions

Let $(X_n : n \in \mathbb{N})$ be a sequence of RVs and let X be a RV, all carried by our triple $(\Omega, \mathcal{F}, \mathsf{P})$. Let us collect together definitions known to us.

Convergence in probability

We say that $X_n \to X$ **in probability** if, for every $\varepsilon > 0$,

$$\mathsf{P}(|X_n - X| > \varepsilon) \to 0 \quad \text{as} \quad n \to \infty.$$

$\mathcal{L}^p$ convergence $(p \geq 1)$

We say that $X_n \to X$ in $\mathcal{L}^p$ if each X_n is in $\mathcal{L}^p$ and $X \in \mathcal{L}^p$ and

$$\|X_n - X\|_p \to 0 \qquad \text{as} \quad n \to \infty,$$

equivalently,

$$\mathsf{E}(|X_n - X|^p) \to 0 \qquad \text{as} \quad n \to \infty.$$

Modes of convergence cont'd

A13.2. Modes of convergence: relationships

Let me state the *facts.*

Convergence in probability is the weakest of the above forms of convergence.

(b) for $p \geq 1$,

$$(X_n \to X \text{ in } \mathcal{L}^p) \Rightarrow (X_n \to X \text{ in prob}).$$

No other implication between any two of our three forms of convergence is valid. But, of course, for $r \geq p \geq 1$,

(c) $(X_n \to X \text{ in } \mathcal{L}^r) \Rightarrow (X_n \to X \text{ in } \mathcal{L}^p)$.

More on Lp spaces

Vector-space property of $\mathcal{L}^p$

(b) Since, for $a, b \in \mathbf{R}^+$, we have

$$(a+b)^p \le [2\max(a,b)]^p \le 2^p(a^p + b^p),$$

$\mathcal{L}^p$ is obviously a vector space.

6.10. Completeness of $\mathcal{L}^p$ $(1 \le p < \infty)$

Let $p \in [1, \infty)$.

The following result (a) is important in functional analysis, and will be crucial for us in the case when $p = 2$. It is instructive to prove it as an exercise in our probabilistic way of thinking, and we now do so.

(a) *If (X_n) is a Cauchy sequence in $\mathcal{L}^p$ in that*

$$\sup_{r,s \ge k} \|X_r - X_s\|_p \to 0 \qquad (k \to \infty)$$

then there exists X in $\mathcal{L}^p$ such that $X_r \to X$ in $\mathcal{L}^p$:

$$\|X_r - X\|_p \to 0 \qquad (r \to \infty).$$

Note. We already know that $\mathcal{L}^p$ is a vector space. Property (a) is important in showing that $\mathcal{L}^p$ can be made into a *Banach space* L^p by a quotienting technique of the type mentioned at the end of the preceding section.

More on Lp spaces cont'd

Let (S, Σ, μ) be a measure space. Suppose that

► $$p > 1 \text{ and } p^{-1} + q^{-1} = 1.$$

Write $f \in \mathcal{L}^p(S, \Sigma, \mu)$ if $f \in \mathrm{m}\Sigma$ and $\mu(|f|^p) < \infty$, and in that case define

$$\|f\|_p := \{\mu(|f|^p)\}^{1/p}.$$

THEOREM

Suppose that $f, g \in \mathcal{L}^p(S, \Sigma, \mu)$, $h \in \mathcal{L}^q(S, \Sigma, \mu)$. *Then*

►(a) **(Hölder's inequality)** $fh \in \mathcal{L}^1(S, \Sigma, \mu)$ *and*

$$|\mu(fh)| \leq \mu(|fh|) \leq \|f\|_p \|h\|_q;$$

►(b) **(Minkowski's inequality)**

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

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 $\mathsf{E}(X|\mathcal{H}) = \mathsf{E}(X)$, $\quad$ a.s.