High-Dimensional Probability and Statistics

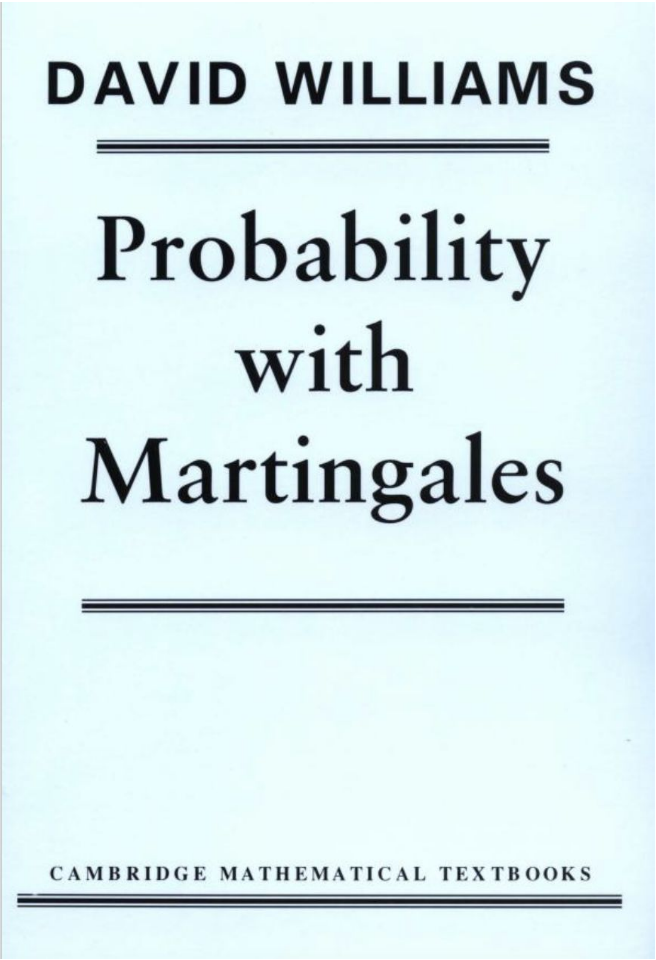
MATH/STAT/ECE 888: Topics in Mathematical Data Science

Sebastien Roch (Math+Stat)
UW-Madison

Fall 2021

Lecture 5 (09/17/21)

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/

More Probability Facts

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*

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

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)**

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

Orlicz spaces

2.7.1 A more general view: Orlicz spaces

Sub-gaussian distributions can be introduced within a more general framework of *Orlicz spaces*. A function $\psi : [0,\infty) \to [0,\infty)$ is called an *Orlicz function* if ψ is convex, increasing, and satisfies

$$\psi(0) = 0, \quad \psi(x) \to \infty \text{ as } x \to \infty.$$

For a given Orlicz function ψ, the Orlicz norm of a random variable X is defined as

$$\|X\|_\psi := \inf\left\{t > 0 : \ \mathbb{E}\,\psi(|X|/t) \le 1\right\}.$$

The *Orlicz space* $L_\psi = L_\psi(\Omega, \Sigma, \mathbb{P})$ consists of all random variables X on the probability space $(\Omega, \Sigma, \mathbb{P})$ with finite Orlicz norm, i.e.

$$L_\psi := \left\{X : \ \|X\|_\psi < \infty\right\}.$$

Orlicz spaces cont'd

Example 2.7.12 (L^p space)**.** Consider the function

$$\psi(x) = x^p,$$

which is obviously an Orlicz function for $p \geq 1$. The resulting Orlicz space L_ψ is the classical space L^p.

Example 2.7.13 (L_{ψ_2} space)**.** Consider the function

$$\psi_2(x) := e^{x^2} - 1,$$

which is obviously an Orlicz function. The resulting Orlicz norm is exactly the sub-gaussian norm $\|\cdot\|_{\psi_2}$ that we defined in (2.13). The corresponding Orlicz space L_{ψ_2} consists of all sub-gaussian random variables.

Jensen's inequality

►A function $c : G \to \mathbf{R}$, where G is an open subinterval of $\mathbf{R}$, is called **convex** on G if its graph lies below any of its chords: for $x, y \in G$ and $0 \leq p = 1 - q \leq 1$,

$$c(px + qy) \leq pc(x) + qc(y).$$

It will be explained below that c is automatically continuous on G. If c is twice-differentiable on G, then c is convex if and only if $c'' \geq 0$.

►*Important examples of convex functions:* $|x|, x^2, e^{\theta x} (\theta \in \mathbf{R})$.

THEOREM. Jensen's inequality

►► *Suppose that $c : G \to \mathbf{R}$ is a convex function on an open subinterval G of $\mathbf{R}$ and that X is a random variable such that*

$$\mathbf{E}(|X|) < \infty, \qquad \mathbf{P}(X \in G) = 1, \qquad \mathbf{E}|c(X)| < \infty.$$

Then

$$\mathbf{E}c(X) \geq c(\mathbf{E}(X)).$$