High-Dimensional Probability and Statistics

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

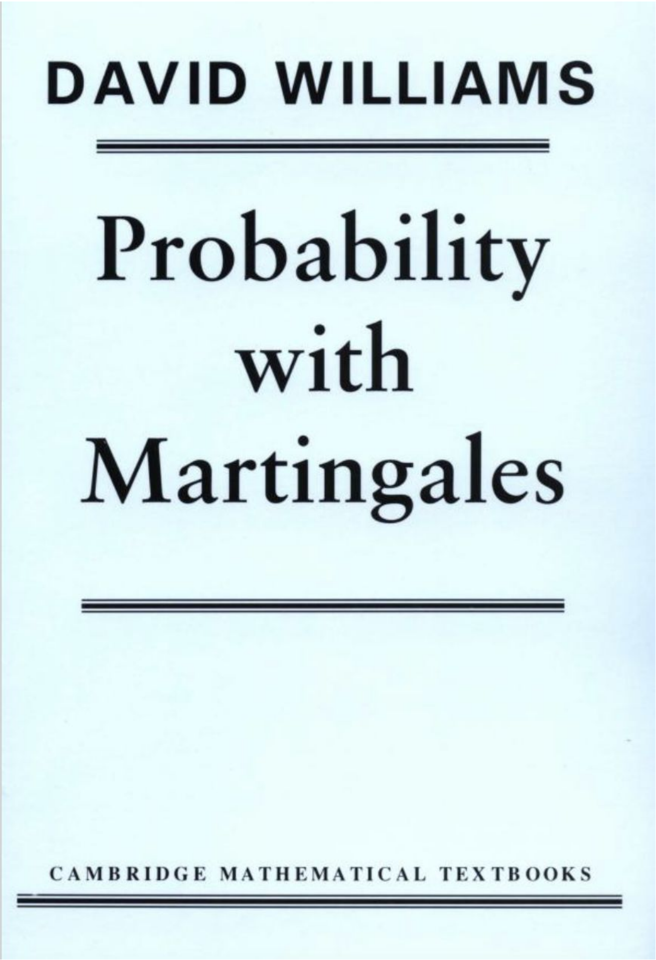
Sebastien Roch (Math+Stat)
UW-Madison

Fall 2021

Lecture 3 (09/13/21)

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

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

►*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.

Convergence in probability

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 $\varepsilon > 0$,

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

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*

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

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:*

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