Ma140a Probability
Lecture: Moments and Law of Large Numbers

Juan Pablo Vigneaux

Caltech

Department of Mathematics
California Institute of Technology
Pasadena, CA, USA
vigneaux@caltech.edu

August 3, 2024

Table of Contents

Table of Contents

Moments

Let $(\Omega, \mathcal{F}, P)$ be a probability triple and $X : \Omega \to \mathbb{R}$ a random variable. We denote by Λ_X the law of X; it is a measure on $\mathbb{R}$.

The p-th moment of X is

$$\mathbb{E}(X^p) := \int_\Omega X^p(\omega)\,\mathrm{d}P(\omega) = \int_\mathbb{R} x^p\,\mathrm{d}\Lambda_X(x). \tag{1}$$

When X^n is not integrable, one says that the moment does not exist.

Similarly, the *centered p-th moment* is

$$\mathbb{E}((X - \mathbb{E}X)^p) = \int_\mathbb{R} (x - \mathbb{E}X)^p\,\mathrm{d}\Lambda_X(x). \tag{2}$$

$\mathbb{E}X$ is called the *mean* of X (or average). $\mathbb{V}(X) := \mathbb{E}((X - \mathbb{E}X)^2)$ is the *variance* of X. Its square root $\sigma = \sqrt{\mathbb{V}(X)}$ is called *standard deviation*: it's a measure of "spread" of the distribution of X (in the same units as the mean).

$\mathcal{L}^p$ spaces

For $1 \leq p < \infty$ we say that $X \in \mathcal{L}^p \equiv \mathcal{L}^p(\Omega, \mathcal{F}, P)$ if X is a r.v. such that $\mathbb{E}|X|^p < \infty$. remark that $\mathcal{L}^p$ is a vector space.

We define $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$. It is a seminorm on $\mathcal{L}^p$, which means that

- it is homogeneous: for all $\lambda \in \mathbb{R}$ and $X \in \mathcal{L}^p$, $\|\lambda X\| = |\lambda| \, \|X\|$.
- It satisfies the triangular inequality (Minkowski's inequality): for all $X, Y \in \mathcal{L}^p$,

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p \, .$$

(The case $p = 2$ easily follows from the Cauchy-Schwarz inequality, see next slide.)

Remark that for any $p \geq 1$, $\|Y\|_p = 0$ iff $Y = 0$ almost everywhere. Let $\mathcal{N}_p$ denote the subset of $\mathcal{L}^p$ of functions that vanish almost everywhere. It is a vector space and the quotient $L^p = \mathcal{L}^p / \mathcal{N}_p$ is a normed vector space called *Lebesgue space*. One can prove that it's also complete (hence a Banach space).

Cauchy-Schwarz inequality

Theorem

If $X, Y \in \mathcal{L}^2$*, then* $XY \in \mathcal{L}^1$ *and*

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

Proof.

Remark first $|\mathbb{E}XY| \leq \mathbb{E}|XY|$, hence we can restrict to $X \geq 0$ and $Y \geq 0$. To ensure convergence of integrals, truncate first: $X_n = X \wedge n = \min(X, n)$ and $Y_n = Y \wedge n$.
For all $a, b \in \mathbb{R}$, one has

$$0 \leq \mathbb{E}(aX_n + bY_n)^2 = a^2\mathbb{E}X_n^2 + 2ab\mathbb{E}X_nY_n + b^2\mathbb{E}Y_n^2.$$

The resulting quadratic in a/b or b/a is nonnegative, and the desired inequality follows form considering the necessary nonpositivity of its discriminant:
$(2\mathbb{E}X_n\mathbb{E}Y_n)^2 \leq 4(\mathbb{E}X_n^2)(\mathbb{E}Y_n^2) \leq 4(\mathbb{E}X^2)(\mathbb{E}Y^2)$. Use the monotone convergence theorem to conclude. □

Monotonicity of L^p-norms

Theorem (Williams, 6.7)

Suppose $1 \le p \le r < \infty$. *If* $Y \in \mathcal{L}^r$, *then* $Y \in \mathcal{L}^p$ *and* $\|Y\|_p \le \|Y\|_r$.

Proof.

For $n \in \mathbb{N}$, we truncate X: let $X_n(\omega) = (|Y|(\omega) \wedge n)^p$. Let $c(x) = x^{r/p}$. Because both X_n and $c(X_n)$ are in $\mathcal{L}^1$ (since X_n is bounded) and c is convex, Jensen's inequality implies that

$$(\mathbb{E}X_n)^{r/p} \le \mathbb{E}X_n^{r/p} = \mathbb{E}((|Y| \wedge n)^r) \le \|Y\|_r^r .$$

Then use the monotone convergence theorem to conclude. □

This truncation method is very important.

Table of Contents

Independence means multiply

Theorem (cf. Williams 7.1; Tamuz 8.1)

Let $X, Y \in \mathcal{L}^1$ *be independent. Then* $X \cdot Y \in \mathcal{L}^1$ *and*

$$\mathbb{E}(X \cdot Y) = \mathbb{E}X \cdot \mathbb{E}Y.$$

In particular, if X *and* Y *are independent elements of* $\mathcal{L}^2$*, then*

$$\mathrm{Cov}(X,Y) := \mathbb{E}((X-\mathbb{E}X)(Y-\mathbb{E}Y)) = 0 \quad \textit{and} \quad \mathbb{V}(aX+bY) = a^2\mathbb{V}(X)+b^2\mathbb{V}(Y).$$

Proof.

Apply the standard machine: assuming X and Y positive, approximate them with an increasing sequence of simple functions. For indicators, equality follows from the definition of independence. □

Remark: for any measurable functions $f, g : \mathbb{R} \to \mathbb{R}$, $g(X)$ and $g(Y)$ are independent and $\mathbb{E}f(X)g(Y) = \mathbb{E}f(X)\mathbb{E}g(Y)$ will follow provided the integrability conditions hold.

Strong law in $\mathcal{L}^4$

Theorem (cf. Williams 7.2, Tamuz 8.3)

Suppose that $X_1, X_2, \dots$ are independent random variables, all with mean μ, and that there is a $K \in [0, \infty)$ such that $\mathbb{E}X_i^4 \leq K$ for all $i \geq 1$. Let $S_n = X_1 + \cdots + X_n$. Then $P(n^{-1}S^n \to \mu) = 1$. In other words: $S_n/n \to \mu$ almost surely.

Proof.

Suppose first the variables are centered: $\mathbb{E}X = 0$. Expand $\mathbb{E}S_n^4 = \mathbb{E}(X_1 + \cdots + X_n)^4$ and use the “independence means multiply” theorem to justify that most terms vanish, except those of the form X_i^4 and $X_i^2 X_j^2$ (one also needs monotonicity of norms to justify that finite fourth moment implies finite third and second moment).

Conclude that $\mathbb{E}S_n^4 \leq 3Kn^2$, hence that $\mathbb{E}(\sum_n (S_n/n)^4) < \infty$. It follows that $\sum_n (S_n/n)^4$ converges almost surely (a.s.) and therefore $S_n/n \to 0$ a.s. □

Table of Contents

More inequality

$(\Omega, \mathcal{F}, P)$ probability space, X r.v. defined on it.

Markov's inequality

If X is positive, then $P(X \geq c) \leq \frac{\mathbb{E}X}{c}$.

Proof: lower bound X by the simple function $cI_{\{X \geq c\}}$, then take expectations.

Chebyshev's inequality

$P(|X - \mathbb{E}X| \geq t) \leq \frac{\mathbb{V}(X)}{t^2}$.

Proof: Apply Markov's inequality to $|X - \mathbb{E}X|^2$.

Theorem (Weak law of large numbers, cf. Tamuz 9.1, also Williams 7.3)

Let $X_1, X_2, \ldots$ *be a sequence of independent real random variables in* $\mathcal{L}^2$*, such that* $\mathbb{E}X_n = \mu$ *and* $\mathbb{V}X_n \leq \sigma^2$ *for all* n*. Set* $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$*. Then for every* $\varepsilon > 0$ *and* $n \in \mathbb{N}^*$*,*

$$P(|Y_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon}.$$

Proof.

Follows from Chebyshev's inequality, since $\mathbb{V}(Y_n) = \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{V}(X_i) \leq \sigma^2/n$. □

By a truncation argument, the theorem also holds without supposing finiteness of the second moment, cf. Tamuz's Thm. 9.4.