Probabilistic Methods in Harmonic Analysis

Lecture 1: Basic Probability Theory

Ingo Witt
University of Göttingen

Summer School on
Dyadic Harmonic Analysis, Martingales, and Paraproducts

Bazaleti, Georgia
September 2-6, 2019

Real-valued random variables

Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. In particular, $\mathbb{P}(\Omega) = 1$.

- A real-valued random variable is a measurable function $X\colon \Omega \to \mathbb{R}$. That is, for any open set $U \subseteq \mathbb{R}$, one has $X^{-1}(U) \in \Sigma$.
- The probability distribution (or law) of X is the probability measure $\mathbb{P}_X$ on $\mathbb{R}$ given by $\mathbb{P}_X(S) = \mathbb{P}(X \in S)$ for $S \in \mathcal{B}(\mathbb{R})$.
  - The law $\mathbb{P}_X$ is often the most important information about X.
- The distribution function F_X of X is given by $F_X(\lambda) = \mathbb{P}(X \leq \lambda)$ for $\lambda \in \mathbb{R}$.
  - F_X is non-decreasing, continuous from the right, $F_X(-\infty) = 0$, and $F_X(\infty) = 1$.
  - The probability distribution is recovered as $\mathbb{P}_X = dF_X$.

Continuous versus discrete

- A real-valued random variable X is continuous if F_X is continuous.
- A real-valued random variable X has a density if $\mathrm{d}F_X \ll \mathrm{d}\lambda$. The density is given as $f_x = \mathrm{d}F_X/\mathrm{d}\lambda$. Then $\mathbb{P}(X \in S) = \int_S f_X(\lambda)\,\mathrm{d}\lambda$.
- A real-valued random variable is discrete if $\mathrm{d}F_X$ is purely discrete.

Examples

1. Continuous distributions
   - Normal (or Gaussian) distribution $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma > 0$:
     $f(\lambda) = \frac{1}{\sqrt{2\pi}\sigma}\,\mathrm{e}^{-\frac{(\lambda-\mu)^2}{2\sigma^2}}$.
2. Discrete distributions
   - Binomial distribution $B(n, p)$, $n \in \mathbb{N}$, $p \in (0, 1)$:
     $\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0, 1, \ldots, n$.
   - Bernoulli distribution $\mathrm{Bernoulli}(p) = B(1, p)$.

Some characteristica

We have the following characteristic quantities if defined:

- The expectation (or expected value) of X is $\mathbb{E}X = \int_\Omega X \, d\mathbb{P} = \int_{-\infty}^{\infty} \lambda \, dF_X(\lambda)$.
- The variance of X is $\mathbb{V}X = \mathbb{E}|X - \mathbb{E}X|^2 = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2$.
- The characteristic function of X is $\varphi_X(t) = \mathbb{E}\left[e^{itX}\right]$, $t \in \mathbb{R}$.

Examples

1. Continuous distributions
   - $N(\mu, \sigma^2)$: $\mathbb{E}X = \mu$, $\mathbb{V}X = \sigma^2$, $\varphi_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$.
2. Discrete distributions
   - $B(n, p)$: $\mathbb{E}X = np$, $\mathbb{V}X = np(1-p)$, $\varphi_X(t) = (1 - p + pe^{it})^n$.
   - Bernoulli(p): $\mathbb{E}X = p$, $\mathbb{V}X = p(1-p)$, $\varphi_X(t) = 1 - p + pe^{it}$.

Elementary inequalities

1. Markov's inequality Let X be non-negative. Then, for any $a > 0$, $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a}$.

Proof.

Let $A = \{X \geq a\}$. Then $\mathbb{P}(A) = \mathbb{E}(\chi_A) \leq \frac{1}{a}\,\mathbb{E}X$. □

2. Generalization: Let X be a random variable, φ be non-decreasing and non-negative on $[0, \infty)$, and $a \geq 0$ with $\varphi(a) > 0$. Then $\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}\varphi(|X|)}{\varphi(a)}$.
3. Chebyshev's inequality Let X be square-integrable. Then, for any $a > 0$, $\mathbb{P}(|X - \mathbb{E}X| \geq a) \leq \frac{\mathbb{V}X}{a^2}$.

Proof.

Apply Markov's inequality to $|X - \mathbb{E}X|^2$ at height a^2. □

1. Basic concepts
2. Independence
3. Rademacher functions

Independent random variables

- A sequence $\{A_j\}_{j=1}^{\infty}$ of events in Σ is said to be independent if, for any finite subset $J \subset \mathbb{N}$, $\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} \mathbb{P}(A_j)$.
- A sequence $\{\Sigma_j\}_{j=1}^{\infty}$ of σ-subalgebras of Σ is said to be independent if, whenever $A_j \in \Sigma_j$ for all j, then the sequence $\{A_j\}_{j=1}^{\infty}$ of events is independent.
- A sequence $\{X_j\}_{j=1}^{\infty}$ of random variables is said to be independent if the sequence $\{\sigma(X_j)\}_{j=1}^{\infty}$ of σ-subalgebras of Σ is independent.

Recall that, for a random variable X, $\sigma(X) = \{X^{-1}(S) \mid S \in \mathcal{B}(\mathbb{R})\}$ is smallest σ-subalgebra of Σ with respect to which X is measurable.

An equivalent characterization

Proposition

If X, Y, and XY are integrable and X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}X \cdot \mathbb{E}Y$.

Proof.

Prove this first when X and Y are simple (in a measure-theoretic sense), then approximate. □

Proposition

X and Y are independent if and only if $\mathbb{E}e^{i(tX+sY)} = \mathbb{E}e^{itX} \cdot \mathbb{E}e^{isY}$.

Modes of convergence

Let $\{X_j\}_{j=1}^{\infty}$ be a sequence of random variables and X be another random variable.

1. X_j converges in distribution (or weakly, or in law) to X if $\lim_{j\to\infty} F_{X_j}(\lambda) = F_X(\lambda)$ for all continuity points λ of F_X.
2. X_j converges in probability to X if, for any $\epsilon > 0$, $\lim_{j\to\infty} \mathbb{P}(|X - X_j| > \epsilon) = 0$.
3. X_j converges almost surely (or strongly) to X if $\mathbb{P}(\lim_{j\to\infty} X_j = X) = 1$.

Note that $X_j \xrightarrow{\text{a.s.}} X \implies X_j \xrightarrow{p} X \implies X_j \xrightarrow{d} X$.

Lemma

Suppose that $X_j \xrightarrow{p} X$. Then there exists a subsequence $\{j_k\}_{k\in\mathbb{N}}$ of $\mathbb{N}$ such that $X_{j_k} \xrightarrow{\text{a.s.}} X$.

Lemma (Borel-Cantelli)

Let $\{A_j\}_{j=1}^{\infty}$ a sequence of events. Then

$$\sum_{j=1}^{\infty} \mathbb{P}(A_j) < \infty \implies \mathbb{P}(A_j \text{ occurs infinitely often}) = 0.$$

Suppose in addition that the A_j are independent. Then

$$\sum_{j=1}^{\infty} \mathbb{P}(A_j) = \infty \implies \mathbb{P}(A_j \text{ occurs infinitely often}) = 1.$$

Proof (of the second part).

One has $\mathbb{P}\left(\bigcap_{n\le j\le N} A_j^c\right) = \prod_{n\le j\le N}(1-\mathbb{P}(A_j)) \le \prod_{n\le j\le N} \exp\left(-\mathbb{P}(A_j)\right) = \exp\left(-\sum_{n\le j\le N}\mathbb{P}(A_j)\right) \to 0$ as $N\to\infty$. Hence, $\mathbb{P}\left(\bigcup_{j\ge n} A_j\right) = 1$ for all n, which finishes the argument. □

Law of large numbers

Theorem

Let $\{X_j\}_{j\geq 1}$ be an i.i.d. sequence of integrable random variables and $S_N = \sum_{j=1}^{N} X_j$ for $N \geq 1$. Then:

1. (Weak form) $S_N/N \xrightarrow{p} \mathbb{E}X_1$,
2. (Strong form) $S_N/N \xrightarrow{\text{a.s.}} \mathbb{E}X_1$.

Proof.

(Weak form if X_1 is L^2) We can assume $\mathbb{E}X_1 = 0$. Then

$$\begin{aligned}\mathbb{P}(|S_N| > \epsilon N) &\leq \epsilon^{-2}N^{-2}\mathbb{E}\,|S_N|^2 \\ &= \epsilon^{-2}N^{-2}\left(\sum_{1\leq j\leq N}\mathbb{E}X_j^2 + 2\sum_{1\leq j<k\leq N}\mathbb{E}[X_jX_k]\right) \\ &= \epsilon^{-2}N^{-1}\mathbb{E}X_1^2 \to 0 \quad \text{as } N \to \infty. \qquad \square\end{aligned}$$

Central limit theorem (CLT)

Theorem

Let $\{X_j\}_{j\in\mathbb{N}}$ be an i.i.d. sequence of L^2 random variables with $\mathbb{E}X_1 = \mu$, $\mathbb{V}X_1 = \sigma^2$, where $\sigma > 0$. Then, for all $a < b$,

$$\lim_{N\to\infty} \mathbb{P}\left(a < \frac{S_N - N\mu}{\sqrt{N}\sigma} < b\right) = \frac{1}{\sqrt{2\pi}\sigma} \int_a^b e^{-\lambda^2/(2\sigma^2)}\, d\lambda.$$

Proof We can assume that $\mathbb{E}X_1 = 0$, $\mathbb{V}X_1 = 1$.
It suffices to show that, for all $\phi \in \mathscr{S}(\mathbb{R})$,

$$\mathbb{E}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\xi S_N/\sqrt{N}}\hat{\phi}(\xi)\, d\xi\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\lambda^2/2}\phi(\lambda)\, d\lambda.$$

as $N \to \infty$, where $\hat{\phi}(\xi) = \int_{-\infty}^{\infty} e^{-i\lambda\xi}\phi(\lambda)\, d\lambda$.

Continuation of the proof

By independence, we have $\mathbb{E}e^{i\xi S_N/\sqrt{N}} = \left(\varphi_{X_1}(\xi/\sqrt{N})\right)^N$.
Now use

$$e^{i\xi X_1} = 1 + i\xi X_1 - \frac{\xi^2}{2} X_1^2 - \xi^2 X_1^2 \int_0^1 (1-t)\left(e^{it\xi X_1} - 1\right) dt,$$

which implies, by taking expectations and using the dominated convergence theorem,

$$\mathbb{E}e^{i\xi S_N/\sqrt{N}} = \left(1 - \frac{\xi^2}{2N} + o(\xi^2/N)\right)^N \to e^{-\xi^2/2}$$

as $N \to \infty$ uniformly in ξ. Therefore, $\mathbb{E}(\dots)$ converges as $N \to \infty$ to

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\xi^2/2}\hat{\phi}(\xi)\, d\xi = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\lambda^2/2}\phi(\lambda)\, d\lambda. \qquad \square$$

Definition and first properties

For $j \in \mathbb{N}_0$, we define the jth Rademacher function to be

$$r_j(t) = \operatorname{sgn} \sin(2^j \pi t), \quad 0 \le t \le 1.$$

Using the right-continuous representative, we have

$$\begin{aligned}
&r_0(t) = 1, \quad 0 \le t \le 1, \\
&r_1(t) = 1, \quad 0 \le t < 1/2, \quad r_1(t) = -1, \quad 1/2 \le t \le 1, \\
&r_2(t) = 1, \quad 0 \le t < 1/4, \quad r_2(t) = -1, \quad 1/4 \le t \le 1/2, \\
&r_2(t) = 1, \quad 1/2 \le t < 3/4, \quad r_2(t) = -1, \quad 3/4 \le t \le 1, \quad \text{etc.}
\end{aligned}$$

Lemma

Given a sequence $\{\epsilon\}_{j=0}^{\infty} \subset \{+1, -1\}$,

$$\lambda_1 \left(\{ r_{j_1} = \epsilon_1, \ldots, r_{j_n} = \epsilon_{j_n} \} \right) = 2^{-n}.$$

In particular, the random variables $r_0, r_1, r_2 \ldots$ *are independent.*

The Walsh system

The Rademacher functions form an orthogonal system $\{r_j\}_{j\in\mathbb{N}_0}$ in $L^2([0,1],\mathrm{d}x)$. This system, however, is not complete.

A complete orthogonal system, into whom the Rademacher functions embed, is given by the system $\{W_k\}_{k\in\mathbb{N}_0}$ of Walsh functions defined as follows: Let k_j be the jth bit in the binary representation of k, starting with k_0 as the least significant bit. Then

$$W_k(t) = \prod_j r_j^{k_j}(t).$$

In particular, $r_j = W_{2^j}$.

Khintchine's inequality

Theorem

For $0 < p < \infty$, there are constants $0 < A_p < B_p < \infty$ such that, for any $\{a_j\} \in \ell^2$,

$$A_p \left\|\{a_j\}\right\|_{\ell^2} \le \left\|\sum_{j=0}^{\infty} a_j r_j\right\|_{L^p} \le B_p \left\|\{a_j\}\right\|_{\ell^2}.$$

Proof.

We can assume that $\{a_j\} \subset \mathbb{R}$ and that $a_j = 0$ for all, but finitely many $j \ge 1$.
Normalize $\|\{a_j\}\|_{\ell^2} = 1$.
Let $\rho > 0$. Then

$$\int_0^1 e^{\rho \sum_j a_j r_j(t)}\,dt = \prod_j \int_0^1 e^{\rho a_j r_j(t)}\,dt = \prod_j \frac{e^{\rho a_j} + e^{-\rho a_j}}{2} \le \prod_j e^{\rho^2 a_j^2/2} = e^{\rho^2/2},$$

where we have used the elementary inequality $\left(e^x + e^{-x}\right)/2 \le e^{x^2/2}$. □

Introduce $F(t) = \sum_j a_j r_j(t)$. Then $\int_0^1 e^{\rho|F(t)|}\,dt \le \int_0^1 e^{\rho F(t)}\,dt + \int_0^1 e^{-\rho F(t)}\,dt \le 2e^{\rho^2/2}$.
Therefore, for $\alpha > 0$,

$$e^{\rho\alpha}\lambda_1(\{|F| > \alpha\}) \le \int_0^1 e^{\rho|F(t)|}\,dt \le 2e^{\rho^2/2}.$$

We obtain $d_F(\alpha) \le 2e^{\rho^2/2-\rho\alpha}$ and, for $\rho = \alpha$,

$$d_F(\alpha) \le 2e^{-\alpha^2/2}.$$

It follows that

$$\|F\|_{L^p}^p = p\int_0^\infty \alpha^{p-1} d_F(\alpha)\,d\alpha \le 2p\int_0^\infty \alpha^{p-1}e^{-\alpha^2/2}\,d\alpha$$
$$= 2^{p/2}p\int_0^\infty \beta^{p/2-1}e^{-\beta}\,d\beta = 2^{p/2}p\,\Gamma(p/2).$$

This proves one of the inequalities with $B_p = \sqrt{2}p^{1/p}\Gamma(p/2)^{1/p}$. □

Remark

- The same proof works when one takes i.i.d. random variables $\{\epsilon_j\}_{j\ge1}$ with $\epsilon_1 \sim 2\,\text{Bernoulli}(1/2) - 1$ (instead of $\{r_j\}$).
- The best constants A_p, B_p are known in case $\{a_j\} \subset \mathbb{R}$ (Haagerup, 1982).