Concentration: Martingales

Master 2 Mathematics and Computer Science

Guillaume Aubrun, Aurélien Garivier, Rémi Gribonval

2020-2021

ENS DE LYON

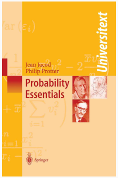
Table of contents

General information

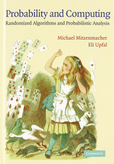
Webpage with information and resources:

`https://perso.ens-lyon.fr/aurelien.garivier/www.math.univ-toulouse.fr/_agarivie/concentration.html`

References for this lecture:



Probability Essentials / L'essentiel des probabilités
by J. Jacod and P. Protter



Probability and Computing
by M. Mitzenmacher and E. Upfal

Motivation

Strategic betting: if $X_1, \dots, X_n \sim \mathcal{U}\{-1, 1\}$ and $S_n = \sum_{i=1}^n X_i$, then $\mathbb{E}[S_n] = 0$ and $\mathbb{P}(|S_n| \geq x\sqrt{n}) \leq \exp(-x^2/2)$.
But what if you can choose $H_i = \phi_i(X_1, \dots, X_{i-1})$ and $Z_n = \sum_{i=1}^n H_i X_i$?

Pattern matching: in DNA sequences over $\mathcal{X} = \{A, C, T, G\}$, find "patterns" = abnormally frequent subsequences.
If $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{U}(\mathcal{X})$, if $p \in \mathcal{X}^k$ and $F_n = \sum_{i=0}^{n-k} \mathbb{1}\{\cap_{j=1}^k X_{i+j} = p_j\}$ then $\mathbb{E}[F_n] = (n-k+1)/4^n$, but $\mathbb{P}\big(F_n > \mathbb{E}[F_n] + x\big) \leq$?

Estimating the unseen: what is the probability that the next item is not in my sample?
X random variable on $\mathbb{N}$, sample $X_1, \dots, X_n$. Histogram $O_n(x) = \sum_{i=1}^n \mathbb{1}\{X_i = x\}$. The mass of the unseen $M_n = \sum_{x=0}^\infty \mathbb{P}(X = x)\mathbb{1}\{O_n(x) = 0\}$ can be estimated by the *Good-Turing estimator* $\hat{M}_n = n^{-1} \sum_{x \in \mathbb{N}} \mathbb{1}\{O_n(x) = 1\}$ since $\big|\mathbb{E}[M_n] - \mathbb{E}[\hat{M}_n]\big| \leq 1/n$. But how close are they whp?

$\implies$ need concentration not only the the mean of independent variables, but also for means of dependent variables, and for other functions of independent variables.

Conditional expectation

Probability measures

Universe $\mathcal{X}$ = set typically discrete or $\subset \mathbb{R}^d$

Sigma-field $\mathcal{A}$ of the *events*

- $\emptyset \in \mathcal{A}$ (and $\mathcal{X} \in \mathcal{A}$),
- $A \in \mathcal{A} \implies \bar{A} \in \mathcal{A}$
- if $\forall i \geq 1, A_i \in \mathcal{A}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ (and $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$)

$(\mathcal{X}, \mathcal{A})$ is called a *measurable space*

Probability $\mathbb{P} : \mathcal{A} \to [0, 1]$

- $\mathbb{P}(\mathcal{X}) = 1$
- whenever $\forall i, A_i \in \mathcal{A}$ and $\forall i \neq j, A_i \cap A_j = \emptyset$, $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

Ex 1: $\mathcal{X} = \{0, 1\}^n, \mathcal{A} = \mathcal{P}\big(\mathcal{X}\big), \mathbb{P}(x) = 2^{-n}$ for all $x \in \mathcal{X}$.
Ex 2: $\mathcal{X} = [0, 1], \mathcal{A} = \mathcal{B}([0, 1])$ the smallest sigma-field containing all open subsets (or all sub-intervals) of $\mathcal{X}$, $\mathbb{P} = \mathcal{U}[0, 1]$: $\mathbb{P}([a, b]) = b - a$ whenever $0 \leq a \leq b \leq 1$.
Ex 3: $\mathcal{X} = \mathbb{R}, \mathcal{A} = \mathcal{B}(\mathbb{R})$, $\mathbb{P}$ = standard gaussian distribution:

$$\mathbb{P}\big([a, b[\big) = \int_a^b \frac{\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}} dx$$

Random variables

Let $(\mathcal{X}, \mathcal{A}, \mathbb{P})$ be a probability space and $(\mathcal{Y}, \mathcal{B})$ be a measurable space.

Random variable $X : \mathcal{X} \to \mathcal{Y}$ such that $\forall B \in \mathcal{B}, X^{-1}(B) \in \mathcal{A}$.
If $\mathcal{Y} = \mathbb{R}$, most often $\mathcal{B} = \mathcal{B}(\mathbb{R})$ and X is called a *real random variable*.

Composition properties: the sum, product, composition, inf, sup etc. of random variables are random variables

Law of X = pushforward measure $P_X : \mathcal{B} \to [0,1]$ defined by
$P_X(B) = \mathbb{P}\big(X^{-1}(B)\big)$ is a probability on $(\mathcal{Y}, \mathcal{B})$

Generated sigma-field $\sigma(X) = \Big\{X^{-1}(B) : B \in \mathcal{B}\Big\}$

<u>Ex 1</u>: $\mathcal{X} = [0,1[\ \big(\text{and } A = \mathcal{B}([0,1])\big),\ \mathbb{P} = \mathcal{U}[0,1[, X = \mathbb{1}_{[0,1/2[}$
$\implies \sigma(X) = \Big\{\emptyset, [0,1[, [0,1/2[, [1/2,1[\Big\}$

Expectation

Let $(\mathcal{X}, \mathcal{A}, \mathbb{P})$ be a probability space

- for a simple rv $X = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$, $\mathbb{E}[X] = \sum a_i \mathbb{P}(A_i)$
- for a positive function $\mathbb{E}[X] = \sup \left\{ E[Y] : 0 \leq Y \leq X \text{ , Y simple} \right\}$
- for a real rv $X = X_+ - X_-$ where $X_+ = \max(X, 0)$ and $X_- = -\min(X, 0)$, if $\mathbb{E}[X_+] < \infty$ and $\mathbb{E}[X_-] < \infty$ then $\mathbb{E}[X] = \mathbb{E}[X_+] - \mathbb{E}[X_-]$

$\mathcal{L}^1$ = set of integrable rv = vector space on which $\mathbb{E}$ is a positive linear form

$L^1 = \mathcal{L}^1$ quotiented by $\mathbb{P}$-almost-sure equality, $X \in L^p$ if $|X|^p \in L^1$

Prop: Fatou, Monotone convergence, dominated convergence

Prop: L^2 with scalar product $\langle X, Y \rangle = \mathbb{E}[XY]$ is an Hilbert space

Prop: Expectation = best constant guess if $X \in L^2$,
$E[X] = \arg\min_{x \in \mathbb{R}} \mathbb{E}\left[(X - x)^2\right] \implies$ orthogonal projection onto $\mathbb{R} \subset L^2$

Characterization: $X : \Omega \to \mathbb{R}^d$ has law Q iff $\forall f \in \mathcal{C}_c(\mathbb{R}^d), \mathbb{E}\left[f(X)\right] = \int f dQ$.

Conditional expectation

Example: H = height, S = sex, W = weight of cats.

What is the best guess if we have side information Y = another rv?

$$\min_{\phi:\mathcal{Y}\to\mathcal{X}\mathrm{meas}} \mathbb{E}\left[\left(X-\phi(Y)\right)^2\right] = \min_{Z\ \sigma(Y)-\mathrm{meas}} \mathbb{E}\left[\left(X-Z\right)^2\right]$$

Prop: $\{\phi(Y) : \phi : \mathbb{R}\to\mathbb{R} \text{ measurable}\} = \{Z : \mathbb{R}\to\mathbb{R} \text{ which are } \sigma(Y) - \text{measurable}\}$

Orthogonal projection onto $L^2\left(\Omega,\sigma(Y),\mathbb{P}\right)$

Conditional expectation of $X \in L^2$ given a rv Y: the unique rv $\mathbb{E}[X|Y] \in L^2(\Omega,\sigma(Y),\mathbb{P})$ such that

$$\forall Z \in L^2(\Omega,\sigma(Y),\mathbb{P}),\ \mathbb{E}\left[\mathbb{E}[X|Y]Z\right] = \mathbb{E}[XZ]$$

Explicit formula in discrete case, depends only on $\sigma(Y)$, extended to $X \in L^1$ by density

Prop: positive, linear, $\mathbb{E}\left[\mathbb{E}[X|Y]\right] = \mathbb{E}[X]$, monotone cv, Fatou, dominated cv

Jensen's inequality: if $\phi : \mathbb{R}\to\mathbb{R}$ is convex, if $X, \phi(X) \in L^1$ then $\phi\left(\mathbb{E}[X|Y]\right) \le \mathbb{E}\left[\phi(X)|Y\right]$

Martingales

Definition

Filtration $(\mathcal{F}_n)_n$= increasing sequence of sigma-fields: $\forall n \geq 0, \mathcal{F}_n \subset F_{n+1} \subset \mathcal{A}$

Martingale $(X_n)_n$ such that $\forall n \geq 0, X_n \in L^1$, X_n is $\mathcal{F}_n$-measurable and $\forall m \leq n$, $\mathbb{E}[X_n|\mathcal{F}_m] = X_m$ a.s.

Ex 1: (Z_i) independent, integrable, centered variables, $\mathcal{F}_n = \sigma\big((X_m)_{m\leq n}\big)$ with $\mathcal{F}_0 = \{\emptyset, \Omega\} \implies X_n = \sum_{i=1}^n Z_i$ is a $(\mathcal{F}_n)_n$-martingale

Ex 1': Martingale transform: if for all $i \geq 1, H_i$ is $\mathcal{F}_{i-1}-$ measurable, $X_n = \sum_{i=1}^n H_i Z_i$ is a $(\mathcal{F}_n)_n$-martingale

Ex 2: for any filtration $(\mathcal{F}_n)_n$ and $X \in L^1$, $X_n = \mathbb{E}[X|\mathcal{F}_n]$ is a martingale

For example $\Omega = [0, 1[$, for all $n \geq 0$ and $0 \leq k \leq n$ let $I_{k,n} = \big[(k-1)2^{-n}, k2^{-n}\big[$ and $\mathcal{F}_n \setminus \mathcal{F}_{n-1} = \big\{\mathbb{1}_{I_{k,n}}, 1 \leq k \leq 2^n\big\}$. Then $X_n = \sum_{k=1}^n \left(2^n \int_{I_{k,n}} f\right) \mathbb{1}_{I_{k,n}}$.

Stopping times

Stopping time wrt filtration $(\mathcal{F}_n)_n$: rv $T : \Omega \to \mathbb{N} \cup \{+\infty\}$ such that $\forall n, \{T \leq n\} \in \mathcal{F}_n$

Examples: if X_n is $\mathcal{F}_n$-measurable, $T = \inf\left\{n : X_n \geq 10\right\}$

Generated sigma-field $\mathcal{F}_T = \left\{A \in \mathcal{A} : \forall n, A \cap \{T \leq n\} \in \mathcal{F}_n\right\}$

Martingale Stopping Theorem: $\mathbb{E}[X_T] = \mathbb{E}[X_0]$ and X_T is $\mathcal{F}_T$-measurable if either:

- T is bounded
- $\exists c$ such that $\forall n, |X_n| \leq c$ a.s.
- $\mathbb{E}[T] < \infty$ and $\exists c$ such that $\forall n, \mathbb{E}\Big[|X_{n+1} - X_n| \Big| \mathcal{F}_n\Big] \leq c$ a.s.

Doob's optional stopping: if S and T are two bounded stopping times such that $S \leq T$ a.s., then $\mathbb{E}\big[X_T | \mathcal{F}_S\big] = X_S$

Doob's maximal inequality for every $x > 0$,

$$\mathbb{P}\Big(\max_{1 \leq j \leq n} |X_j| \geq x\Big) \leq \frac{\mathbb{E}\big[|X_n|\big]}{x}$$

Example: ballot theorem

In an election, candidate A obtained a votes and candidate B obtained $b < a$ votes. Votes where counted one after another. What is the probability that, during the counting, candidate A was always ahead?

Let $n = a + b$ and for every $1 \leq k \leq n$, let S_k be the (positive or negative) number of votes by which A is leading after k votes are counted: $S_n = a - b$.

Prop: for all $0 \leq k < n$, $X_k := \frac{S_{n-k}}{n-k}$ is a martingale.

Let $T = \inf \left\{ k \in \{0, \dots, n\} : X_k = 0 \right\}$ if this set is not empty, or $T = n - 1$ otherwise (if A is always ahead). T is a bounded stopping time. In the first case, $X_T = 0$, while in the second case $X_T = S_1 = 1$. Hence,

$$\begin{aligned} \mathbb{E}[X_T] = \mathbb{E}[X_0] = \frac{\mathbb{E}[S_n]}{n} &= \frac{a-b}{a+b} \\ &= \mathbb{P}(T < n-1) \times 0 + \mathbb{P}(T = n-1) \times 1 \end{aligned}$$

$$\implies \mathbb{P}(T = n-1) = \frac{a-b}{a+b}$$

Ex: gambler's ruin problem.

Wald's equation

Th: Let $X_1, X_2, \ldots$ be nonnegative iid random variables, and let T be a stopping time. If $\mathbb{E}[T] < \infty$ and $\mathbb{E}[X] < \infty$, then

$$\mathbb{E}\left[\sum_{i=1}^{T} X_i\right] = \mathbb{E}[T]\,\mathbb{E}[X_1]\,.$$

Ex: strategic betting (see St Petersburg lottery)

Azuma-Hoeffding inequality

Th: Let $X_0, \dots, X_n$ be a martingale such that $\forall 1 \leq k \leq n, |X_k - X_{k-1}| \leq c_k$. Then for all $x > 0$,

$$\mathbb{P}\big(|X_n - X_0| > x\big) \leq 2 \exp\left(-\frac{x^2}{2\sum_{k=1}^{n} c_k^2}\right)$$

Application: dynamic betting

Mc-Diarmid's inequality

McDiarmid's inequality: If $X_1, \dots X_n$ are independent random variables on $\mathcal{X}$ and $f : \mathcal{X}^n \to \mathbb{R}$ is such that $\forall 1 \le i \le n, \forall x_1, \dots, x_n, x_i'$,

$$\left| f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_{i-1}, x_i', x_{i+1}, \dots, x_n) \right| \le c_i \ ,$$

then

$$\mathbb{P}\Big(\big| f(X_1, \dots, X_n) - \mathbb{E}\big[f(X_1, \dots, X_n) \big] \big| \ge x \Big) \le \exp\left(\frac{-2x^2}{\sum_{i=1}^n c_i^2} \right) \ .$$

Sanity check: $f(x) = \sum x_i$

Example: balls and bins, pattern-matching

Missing mass estimation: Good-Turing Estimator

X random variable on $\mathbb{N}$, sample $X_1, \ldots, X_n$.
Histogram $O_n(x) = \sum_{i=1}^{n} \mathbb{1}\{X_i = x\}$.

Mass of the unseen: $M_n = \sum_{x=0}^{\infty} \mathbb{P}(X = x)\mathbb{1}\{O_n(x) = 0\}$

Good-Turing estimator: $\hat{M}_n = n^{-1} \sum_{x \in \mathbb{N}} \mathbb{1}\{O_n(x) = 1\}$

Prop: $0 \leq \mathbb{E}[\hat{M}_n] - \mathbb{E}[M_n] \leq 1/n$

Prop: Concentration of $\hat{M}_n$: Mc-Diarmid's inequality!

Concentration of M_n: see negative association