Statistics and Data Analysis

Summary of Lecture 1: What you must know in Probability Theory

This document provides a summary of the main definitions and results of Lecture 1, together with short exercices which can be solved during the lecture. It is still recommended to carefully read Lecture 1!

1 Abstract random variables

1.1 Summary of the section

- Abstract random variable: measurable function $X : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathcal{E})$, where $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space and $(E, \mathcal{E})$ a measurable space.
- Law of a random variable: probability measure P on $(E, \mathcal{E})$ defined by $P(C) = \mathbb{P}(X \in C)$, for all $C \in \mathcal{E}$.
- Usually, two kinds of spaces: discrete (E is finite or countably infinite), continuous ($E = \mathbb{R}^d$).
- Real-valued random variables: $\mathbb{E}[X] = \int_{\omega \in \Omega} X(\omega) \mathrm{d}\mathbb{P}(\omega)$, $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
- Covariance between random variables: $\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. It is bilinear and symmetric, and $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$.
- Vector-valued random variables: $(\mathbb{E}[X])_i = \mathbb{E}[X_i]$, $(\mathrm{Cov}[X])_{i,j} = \mathrm{Cov}(X_i, X_j)$, $\mathrm{Var}(X) = \mathrm{tr}\,\mathrm{Cov}[X]$.

1.2 Training exercises

Exercise 1.1 (Properties of covariance matrices)**.** Let $X \in \mathbb{R}^d$ be such that $\mathbb{E}[\|X\|^2] < +\infty$ and set $K = \mathrm{Cov}[X]$.

1. For any $u \in \mathbb{R}^d$, show that $\mathrm{Var}(\langle u, X\rangle) = \langle u, Ku\rangle$.
2. For any $A \in \mathbb{R}^{k \times d}$ and $b \in \mathbb{R}^k$, compute $\mathbb{E}[AX + b]$ and $\mathrm{Cov}[AX + b]$.

Exercise 1.2. Show that if $X, Y \in \mathbb{R}$ are independent then $\mathrm{Cov}(X, Y) = 0$. What about the converse statement?

2 Discrete random variables

2.1 Summary of the section

- Probability mass function $(p_x)_{x \in E}$: $p_x = \mathbb{P}(X = x)$.
- Usual distributions:

	Symbol	Parameter	PMF	Support	Expectation	Variance
Bernoulli	$\mathcal{B}(p)$	$p \in [0, 1]$	$p^x(1-p)^{1-x}$	$x \in \{0, 1\}$	p	$p(1-p)$
Binomial	$\mathcal{B}(n, p)$	$p \in [0, 1]$, $n \in \mathbb{N}^*$	$\binom{n}{k} p^k (1-p)^{n-k}$	$k \in \{0, \dots, n\}$	np	$np(1-p)$
Geometric	$\mathcal{G}(p)$	$p \in (0, 1]$	$(1-p)^{k-1}p$	$k \in \mathbb{N}^*$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson	$\mathcal{P}(\lambda)$	$\lambda > 0$	$\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}$	$k \in \mathbb{N}$	λ	λ

2.2 Training exercises

Exercise 2.1 (What you read in the news)**.** In a famous newspaper article from 2011[1], two engineers claim that if p is the probability that one nuclear reactor has a serious accident during one year, then the probability that at least one serious accident occurs among N nuclear reactors, during M years, is $p \times N \times M$.

1. Applying this result with an estimated value $p = 4/14000$, the authors deduce that the probability to have at least one serious accident among the $N = 143$ currently working nuclear reactors in Europe, during the next $M = 30$ years, is equal to 1.23. What do you think of this statement?

2. With the same values for p, N and M, how would you correct this computation?

Exercise 2.2 (Unbiasing a coin toss, an exercise attributed to Von Neumann)**.** Assume that you have a random number generator which returns independent Bernoulli variables with an *unknown* parameter $p \in (0,1)$. How to use it to draw a Bernoulli random variable with parameter $1/2$?

3 Random variables with density

3.1 Summary of the section

- $X \in \mathbb{R}^d$ has density $p : \mathbb{R}^d \to [0,+\infty)$ if $\mathbb{P}(X \in B) = \int_B p(x)\mathrm{d}x$ for any measurable $B \subset \mathbb{R}^d$.
- Usual one-dimensional distributions:

	Symbol	Parameter	Density	Support	Expectation	Variance
Uniform	$\mathcal{U}([a,b])$	$a < b$	$\frac{1}{b-a}$	$x \in [a,b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential	$\mathcal{E}(\lambda)$	$\lambda > 0$	$\lambda \mathrm{e}^{-\lambda x}$	$x > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma	$\Gamma(a,\lambda)$	$a, \lambda > 0$	$\frac{\lambda^a}{\Gamma(a)} x^{a-1} \mathrm{e}^{-\lambda x}$	$x > 0$	$\frac{a}{\lambda}$	$\frac{a}{\lambda^2}$
Gaussian	$\mathcal{N}(\mu,\sigma^2)$	$\mu \in \mathbb{R}, \sigma^2 > 0$	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$x \in \mathbb{R}$	μ	σ^2

- If X, Y have densities p, q and are independent, $X+Y$ has density $p * q(z) = \int_{x \in \mathbb{R}^d} p(x)q(z-x)\mathrm{d}x$.
- Real-valued variables: CDF $F(x) = \mathbb{P}(X \leq x)$, quantile q_r such that $\mathbb{P}(X \leq q_r) = r$.
- Characteristic function: $\Psi_X(u) = \mathbb{E}[\mathrm{e}^{\mathrm{i}\langle u, X\rangle}]$, $u \in \mathbb{R}^d$.
- A random vector $X \in \mathbb{R}^d$ is Gaussian if for any $u \in \mathbb{R}^d$, $\langle u, X\rangle$ is Gaussian (with convention $\mu \sim \mathcal{N}(\mu, 0)$). If X is Gaussian then $\Psi_X(u) = \mathrm{e}^{\mathrm{i}\langle u,m\rangle - \frac{1}{2}\langle u, Ku\rangle}$, $m = \mathbb{E}[X]$, $K = \mathrm{Cov}[X]$. We denote $X \sim \mathcal{N}_d(m, K)$.

3.2 Training exercises

Exercise 3.1. 1. Show that if $X \sim \Gamma(a,\lambda)$ and $c > 0$ then $cX \sim \Gamma(a, \lambda/c)$.

2. Let $X \sim \Gamma(a,\lambda)$ and $Y \sim \Gamma(b,\lambda)$ be independent. Show that $X + Y \sim \Gamma(a+b, \lambda)$.

3. Let $X_1, \ldots, X_n$ be independent variables with law $\mathcal{E}(\lambda)$. Show that $\frac{1}{n}(X_1 + \cdots + X_n) \sim \Gamma(n, n\lambda)$.

Exercise 3.2. Let ϕ_r be the quantile or order r of $\mathcal{N}(0,1)$. What is the link between ϕ_r and ϕ_{1-r}?

Exercise 3.3. Let $G \sim \mathcal{N}(0,1)$.

[1] See the blog post https://www.afis.org/Nouveau-record-du-monde-de-probabilites, in French, for details.

1. Show that Ψ_G is C^1 on $\mathbb{R}$ and satisfies the differential equation

$$\begin{cases} \Psi'_G(u) + u\Psi_G(u) = 0, \\ \Psi_G(0) = 1. \end{cases}$$

2. Deduce that $\Psi_G(u) = \exp(-u^2/2)$.

3. Express the characteristic function of $X \sim \mathcal{N}(\mu, \sigma^2)$ in terms of Ψ_G.

4. Show that, if $X \sim \mathcal{N}(\mu, \sigma^2)$ and $Y \sim \mathcal{N}(\nu, \tau^2)$ are independent, then $X + Y \sim \mathcal{N}(\mu + \nu, \sigma^2 + \tau^2)$.

Exercise 3.4. If $X \sim \mathcal{N}_d(m, K)$ and $A \in \mathbb{R}^{k\times d}$, $b \in \mathbb{R}^k$, what is the law of $AX + b$?

4 Convergence and limit theorems

4.1 Summary of the section

- Convergence almost sure (a.s.): $\mathbb{P}(X_n \to X) = 1$; in probability: $\forall \epsilon > 0, \quad \mathbb{P}(\|X_n - X\| \geq \epsilon) \to 0$.
- Dominated convergence theorem: if $X_n \to X$ a.s. and $|X_n| \leq Y$ with $\mathbb{E}[Y] < +\infty$, then $\mathbb{E}[X_n] \to \mathbb{E}[X]$.
- Convergence in distribution: $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for any continuous and bounded $f : \mathbb{R}^d \to \mathbb{R}$.
- $X_n \to X$ in distribution in $\mathbb{R}^d \Leftrightarrow \Psi_{X_n}(u) \to \Psi_X(u)$ for all $u \in \mathbb{R}^d$.
- $X_n \to X$ in distribution in $\mathbb{R} \Leftrightarrow$ for any x such that $\mathbb{P}(X = x) = 0$, $\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$.
- (Strong) LLN: if $X_1, \ldots, X_n$ are iid and $\mathbb{E}[\|X_1\|] < +\infty$, then $\overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i \to \mathbb{E}[X_1]$, a.s.
- CLT: if $\mathbb{E}[\|X_1\|^2] < +\infty$, then $\sqrt{n}(\overline{X}_n - \mathbb{E}[X_1]) \to \mathcal{N}_d(0, K)$, in distribution, with $K = \mathrm{Cov}[X]$.

4.2 Training exercises

Exercise 4.1. Let U be a random variable uniformly distributed on $[0, 1]$. Define $T = -\ln U$.

1. By computing the distribution function of T, identify the law of this variable.

2. Deduce that $\mathbb{E}[\ln U]$ and $\mathrm{Var}(\ln U)$ exist and give their values.

Let $(U_n)_{n\geq 1}$ be a sequence of independent random variables, each with the uniform distribution on $[0, 1]$. For every $n \geq 1$ define

$$X_n = \left(\prod_{i=1}^n U_i\right)^{1/n}, \qquad Y_n = e^{\sqrt{n}} \left(\prod_{i=1}^n U_i\right)^{1/\sqrt{n}}.$$

3. Show that X_n converges almost surely and give its limit.

4. Let G be a standard normal random variable. Show that Y_n converges in distribution to a random variable expressed as a function of G.

Exercise 4.2. Show that, if $X_n \to X$ a.s., then $X_n \to X$ in probability. What do you think or know about the converse statement?

Exercise 4.3. Let $(U_n)_{n\geq 1}$ be a sequence of independent random variables with uniform distribution over $[0, 1]$. Let $M_n = \max_{1\leq i\leq n} U_i$.

1. Show that $M_n \to 1$, in probability.

2. Show that, for any $\omega \in \Omega$, the sequence $(M_n(\omega))_{n\geq 1}$ is nondecreasing. Deduce that $M_n \to 1$, a.s.

3. For any $x \geq 0$, compute $\lim_{n\to+\infty} \mathbb{P}(1 - M_n > x/n)$. Deduce that $n(1 - M_n)$ converges in distribution toward some limit X and describe the law of X.

Exercise 4.4 (Stronger convergence in the Central Limit Theorem)**.** Under the assumptions of the Central Limit Theorem, say in dimension $d = 1$ to make things simpler, it is a natural question to wonder whether there exists a random variable Z such that $Z_n := \sqrt{n}(\overline{X}_n - \mathbb{E}[X_1])$ converges to Z in probability. Notice that if such a variable exists, then necessarily $Z \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = \mathrm{Var}(X_1)$.

1. Set $Y_i := X_i - \mathbb{E}[X_1]$ and let $Z'_n := \frac{1}{\sqrt{n}} \sum_{i=n+1}^{2n} Y_i$. Show that Z'_n converges in distribution to some random variable Z' and explicit the law of Z'.
2. If Z_n converges in probability to some random variable Z, show that Z'_n converges in probability and express its limit in terms of Z.
3. What do you conclude?