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?

Correction of exercises

Correction of Exercise 1.1

1. $\operatorname{Var}(\langle u, X\rangle) = \operatorname{Var}\left(\sum_{i=1}^d u_i X_i\right) = \operatorname{Cov}\left(\sum_{i=1}^d u_i X_i, \sum_{j=1}^d u_j X_j\right) = \sum_{i,j=1}^d u_i u_j \operatorname{Cov}(X_i, X_j) = \langle u, Ku\rangle.$

2. $\mathbb{E}[AX+b] = A\mathbb{E}[X]+b$ and $\operatorname{Cov}[AX+b] = AKA^\top$.

Correction of Exercise 1.2 If X and Y are independent then $\operatorname{Cov}(X,Y) = \mathbb{E}[X-\mathbb{E}[X]]\mathbb{E}[Y-\mathbb{E}[Y]] = 0$. Conversely, if one takes $X \sim \mathcal{N}(0,1)$ and $Y = X^2$ then $\operatorname{Cov}(X,Y) = \mathbb{E}[X^3] = 0$, but X and Y are not independent, because for example $3 = \mathbb{E}[X^4] = \mathbb{E}[X^2 Y] \neq \mathbb{E}[X^2]\mathbb{E}[Y] = 1$.

Correction of Exercise 2.1

1. A probability larger than 1 is not possible.

2. The underlying assumption is that the $N \times M$ events 'one nuclear reactor has a serious accidents during one year' are independent, with identical probability p. So the probability that no incident occurs in $(1-p)^{NM}$, and thus the probability that at least one incident occurs is $1-(1-p)^{NM}$. Remark that in the $p \to 0$ limit, the Taylor expansion yields the formula $p \times N \times M$ used by the authors in the article — but this is only an approximation. With this corrected formula and for the values $p = 4/14000$, $N = 143$ and $M = 30$ one gets a probability of 0.7. This is still very large! However, both the estimation of p and the assumptions on the events are questionnable.

Correction of Exercise 2.2 Von Neumann's solution consists in throwing the coin twice at each toss: you thus obtain iid pairs $(X_i, Y_i)_{i\geq 1}$, which have law $\mathcal{B}(p)\otimes\mathcal{B}(p)$. You then keep the first toss for which $X_i \neq Y_i$, that is to say that you set $N = \inf\{i \geq 1 : X_i \neq Y_i\}$. Notice that N is a geometric random variable, with parameter $\mathbb{P}(X \neq Y) = \mathbb{P}(X=0, Y=1) + \mathbb{P}(X=1, Y=0) = 2p(1-p)$. Then it turns out that $X_N \sim \mathcal{B}(1/2)$. Indeed,

$$\begin{aligned}
\mathbb{P}(X_N = 1) &= \sum_{n=1}^{+\infty} \mathbb{P}(X_n = 1, n = N)\\
&= \sum_{n=1}^{+\infty} \mathbb{P}(X_1 = Y_1, \ldots, X_{n-1} = Y_{n-1}, X_n = 1, Y_n = 0) = \sum_{n=1}^{+\infty} \mathbb{P}(X_1 = Y_1)^{n-1}\mathbb{P}(X_1 = 1, Y_1 = 0),
\end{aligned}$$

where we have used the fact that the pairs (X_i, Y_i) are iid. Now, on the one hand,

$$\mathbb{P}(X_1 = Y_1) = \mathbb{P}(X_1 = 0, Y_1 = 0) + \mathbb{P}(X_1 = 1, Y_1 = 1) = (1-p)^2 + p^2,$$

while on the other hand, $\mathbb{P}(X_1 = 1, Y_1 = 0) = p(1-p)$. We deduce that

$$\mathbb{P}(X_N = 1) = p(1-p)\sum_{n=1}^{+\infty}\left((1-p)^2 + p^2\right)^{n-1} = \frac{p(1-p)}{1-((1-p)^2+p^2)} = \frac{1}{2}.$$

Correction of Exercise 3.1

1. Let $f : [0,+\infty) \to \mathbb{R}$ be measurable and bounded. For $c > 0$, setting $y = cx$,

$$\begin{aligned}
\mathbb{E}[f(cX)] &= \int_{x=0}^{+\infty} f(cx)\frac{\lambda^a}{\Gamma(a)} x^{a-1}\mathrm{e}^{-\lambda x}\mathrm{d}x\\
&= \int_{y=0}^{+\infty} f(y)\frac{\lambda^a}{\Gamma(a)}\left(\frac{y}{c}\right)^{a-1}\mathrm{e}^{-\lambda y/c}\frac{\mathrm{d}y}{c}\\
&= \int_{y=0}^{+\infty} f(y)\frac{(\lambda/c)^a}{\Gamma(a)} y^{a-1}\mathrm{e}^{-(\lambda/c)y}\mathrm{d}y,
\end{aligned}$$

and we recognise the density of the $\Gamma(a, \lambda/c)$ distribution in the right-hand side.

2. We use the fact that the density q of $X+Y$ is the convolution of the densities of X and of Y, given for $z>0$ by

$$q(z) = \int_{x\in\mathbb{R}} \mathbb{1}_{\{x>0\}} \frac{\lambda^a}{\Gamma(a)} x^{a-1}\mathrm{e}^{-\lambda x} \mathbb{1}_{\{z-x>0\}} \frac{\lambda^b}{\Gamma(b)} (z-x)^{b-1}\mathrm{e}^{-\lambda(z-x)}\mathrm{d}x$$
$$= \frac{\lambda^{a+b}\mathrm{e}^{-\lambda z}}{\Gamma(a)\Gamma(b)} \int_{x=0}^{z} x^{a-1}(z-x)^{b-1}\mathrm{d}x.$$

Setting $u = x/z$ in the integral, we get

$$\int_{x=0}^{z} x^{a-1}(z-x)^{b-1}\mathrm{d}x = \int_{u=0}^{1} (uz)^{a-1}((1-u)z)^{b-1} z\mathrm{d}u = z^{a+b-1}\mathrm{B}(a,b),$$

where the quantity

$$\mathrm{B}(a,b) = \int_{u=0}^{1} u^{a-1}(1-u)^{b-1}\mathrm{d}u$$

no longer depends on z. Now, since q is a probability density, it must satisfy $\int_{z=0}^{+\infty} q(z)\mathrm{d}z = 1$, that is to say

$$\frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)}\mathrm{B}(a,b)\int_{z=0}^{+\infty} z^{a+b-1}\mathrm{e}^{-\lambda z}\mathrm{d}z = 1.$$

But we know, from the expression of the density of the $\Gamma(a+b,\lambda)$ distribution, that

$$\int_{z=0}^{+\infty} z^{a+b-1}\mathrm{e}^{-\lambda z}\mathrm{d}z = \frac{\Gamma(a+b)}{\lambda^{a+b}}.$$

Equating both previous identities, we finally get the nice (and important) formula

$$\mathrm{B}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

which finally allows us to write

$$q(z) = \frac{\lambda^{a+b}}{\Gamma(a+b)} z^{a+b-1}\mathrm{e}^{-\lambda z},$$

which means that $X+Y \sim \Gamma(a+b,\lambda)$.

3. We notice that $\mathcal{E}(\lambda) = \Gamma(1,\lambda)$. As a consequence, using Question 2 recursively, we get $X_1+\cdots+X_n \sim \Gamma(n,\lambda)$, and then by Question 1, $\frac{1}{n}(X_1+\cdots+X_n) \sim \Gamma(n, n\lambda)$.

Correction of Exercise 3.2 Draw the density of $\mathcal{N}(0,1)$, place ϕ_r and ϕ_{1-r} on your picture, and understand why $\phi_{1-r} = -\phi_r$.

Correction of Exercise 3.3

1. Since $|\frac{\mathrm{d}}{\mathrm{d}u}\mathrm{e}^{\mathrm{i}uG}| = |G|$ and $\mathbb{E}[|G|] < +\infty$, the Leibniz differentiation theorem shows that Ψ_G is C^1 on $\mathbb{R}$, with

$$\Psi'_G(u) = \mathbb{E}\left[\mathrm{i}G\mathrm{e}^{\mathrm{i}uG}\right] = \int_{x\in\mathbb{R}} \mathrm{i}x\frac{\mathrm{e}^{\mathrm{i}ux-\frac{x^2}{2}}}{\sqrt{2\pi}}\mathrm{d}x.$$

On the other hand,

$$0 = \left[\mathrm{e}^{\mathrm{i}ux-\frac{x^2}{2}}\right]_{x=-\infty}^{+\infty} = \int_{x\in\mathbb{R}} \frac{\mathrm{d}}{\mathrm{d}x}\mathrm{e}^{\mathrm{i}ux-\frac{x^2}{2}}\mathrm{d}x = \int_{x\in\mathbb{R}} (\mathrm{i}u-x)\mathrm{e}^{\mathrm{i}ux-\frac{x^2}{2}}\mathrm{d}x,$$

so that $\Psi'_G(u) = -u\Psi_G(u)$. It is moreover direct that $\Psi_G(0) = 1$.

2. Solving the differential equation yields $\Psi_G(u) = \exp(-u^2/2)$.

3. We know that if $G \sim \mathcal{N}(0,1)$, then $X := \mu+\sigma G \sim \mathcal{N}(\mu,\sigma^2)$. As a consequence, $\Psi_X(t) = \mathbb{E}[\exp(\mathrm{i}tX)] = \mathbb{E}[\exp(\mathrm{i}t(\mu+\sigma G))] = \exp(\mathrm{i}t\mu)\Psi_G(\sigma t) = \exp(\mathrm{i}t\mu - \frac{\sigma^2t^2}{2})$.

4. The characteristic function of $X+Y$ writes $\Psi_{X+Y}(t) = \mathbb{E}[\mathrm{e}^{\mathrm{i}t(X+Y)}] = \Psi_X(t)\Psi_Y(t) = \exp(\mathrm{i}t(\mu+\nu) - \frac{(\sigma^2+\tau^2)t^2}{2})$, which proves the claimed statement.

Correction of Exercise 3.4 The characteristic function of $AX + b$ writes, for $u \in \mathbb{R}^k$, $\mathbb{E}[\exp(\mathrm{i}\langle u, AX + b\rangle)] = \exp(\mathrm{i}\langle u, b\rangle)\Psi_X(A^\top u) = \exp(\mathrm{i}\langle u, b\rangle)\exp(\mathrm{i}\langle A^\top u, m\rangle - \frac{1}{2}\langle A^\top u, KA^\top u\rangle) = \exp(\mathrm{i}\langle u, Am+b\rangle - \frac{1}{2}\langle u, AKA^\top u\rangle)$, so $AX + b \sim \mathcal{N}_k(Am + b, AKA^\top)$.

Correction of Exercise 4.1 1. For $t \geq 0$, $\mathbb{P}(T \leq t) = \mathbb{P}(-\ln U \leq t) = \mathbb{P}(U \geq e^{-t})$. Since $U \sim \mathcal{U}([0, 1])$,

$$\mathbb{P}(T \leq t) = 1 - e^{-t}.$$

We recognise the distribution function of the exponential law with parameter 1: $T \sim \mathrm{Exp}(1)$.

2. Note that $\ln U = -T$. Since $T \sim \mathrm{Exp}(1)$, $\mathbb{E}[T] = 1$ and $\mathrm{Var}(T) = 1$. Therefore

$$\mathbb{E}[\ln U] = -1, \qquad \mathrm{Var}(\ln U) = 1.$$

3. We write

$$X_n = \exp\left(\tfrac{1}{n}\sum_{i=1}^n \ln U_i\right) = \exp\left(-\tfrac{1}{n}\sum_{i=1}^n T_i\right),$$

where $T_i = -\ln U_i$ are i.i.d. $\mathrm{Exp}(1)$. By the strong law of large numbers,

$$\frac{1}{n}\sum_{i=1}^n T_i \xrightarrow{\text{a.s.}} \mathbb{E}[T_1] = 1.$$

Hence by continuity of $x \mapsto e^{-x}$, we have that $X_n \to e^{-1}$ almost surely.

4. We have

$$Y_n = \exp\left(\sqrt{n} - \tfrac{1}{\sqrt{n}}\sum_{i=1}^n T_i\right).$$

Let $S_n = \sum_{i=1}^n T_i$ and $G_n = \frac{S_n - n}{\sqrt{n}}$. By the CLT,

$$G_n = \frac{S_n - n}{\sqrt{n}} \xrightarrow{d} G \sim \mathcal{N}(0, 1).$$

Therefore by continuity of $x \mapsto e^{-x}$,

$$Y_n = e^{-G_n} \xrightarrow{d} e^{-G}.$$

So the limit law is that of e^{-G}, with $G \sim \mathcal{N}(0, 1)$.

Correction of Exercise 4.2 Let $\epsilon > 0$. We have $\mathbb{P}(\|X_n - X\| \geq \epsilon) = \mathbb{E}[\mathbb{1}_{\{\|X_n - X\| \geq \epsilon\}}]$. Since $X_n \to X$, a.s., the random variable $\mathbb{1}_{\{\|X_n - X\| \geq \epsilon\}}$ converges to 0, a.s. Moreover it is bounded, so the Dominated Convergence Theorem shows that its expectation converges to 0, which proves that $X_n \to X$ in probability.

Conversely, it is known that there are sequences which converge in probability but not a.s. However, any sequence which converges in probability admits an a.s. converging subsequence. See the course of *Probability Theory* for details.

Correction of Exercise 4.3

1. Let $\epsilon \in (0, 1)$. For any $n \geq 1$,

$$\mathbb{P}(M_n \leq 1 - \epsilon) = \mathbb{P}(\forall i \in \{1, \ldots, n\}, U_i \leq 1 - \epsilon) = \mathbb{P}(U_1 \leq 1 - \epsilon)^n = (1 - \epsilon)^n \to 0.$$

   Since $M_n \leq 1$, this shows that $M_n \to 1$ in probability.

2. For any $\omega \in \Omega$, $M_{n+1}(\omega) = \max(M_n(\omega), U_{n+1}(\omega)) \geq M_n(\omega)$. As a consequence, there exists $\ell(\omega) \leq 1$ such that $M_n(\omega) \to \ell(\omega)$. In other words, $M_n \to \ell$, a.s. But since we already know that $M_n \to 1$ in probability, we deduce that $\ell = 1$, a.s., and therefore that $M_n \to 1$, a.s.

3. Taking $\epsilon = x/n$ in the computation of Question 1, we get

$$\mathbb{P}(M_n < 1 - x/n) = \left(1 - \frac{x}{n}\right)^n = \exp\left(n \log\left(1 - \frac{x}{n}\right)\right) \to \exp(-x).$$

   We deduce that $n(1 - M_n)$ converges in distribution to $\mathcal{E}(1)$.

Correction of Exercise 4.4

1. Z'_n is $\sqrt{n}$ times the empirical mean of n iid centered random variables with finite variance σ^2, so by the Central Limit Theorem we have $Z'_n \to \mathcal{N}(0, \sigma^2)$ in distribution.

2. We have $Z_n + Z'_n = \sqrt{2}Z_{2n}$ so if $Z_n \to Z$ in probability we have $Z'_n = \sqrt{2}Z_{2n} - Z_n \to (\sqrt{2} - 1)Z$ in probability.

3. Under the assumption made above that $Z_n \to Z$ in probability, we deduce from the previous questions that necessarily, $(\sqrt{2} - 1)Z \sim \mathcal{N}(0, \sigma^2)$. But on the other hand, since $Z_n \to Z$ in distribution, we also have $Z \sim \mathcal{N}(0, \sigma^2)$. We deduce that for the assumption to hold true, it is necessary that $\sigma^2 = 0$. And it is straightforward to check that, conversely, if $\sigma^2 = 0$ then indeed $Z_n = 0$ converges in probability. As a conclusion, we have established that Z_n converges in probability if and only if the variables X_i are deterministic.