Exercises and solutions: Chapter 3 only

Sébastien Ott

October 13, 2025

***Exercise 3.1.** Show Lemma **??**.
Hint: where are the discontinuities of F_X? How does F_X behaves between these discontinuities?

Solution 3.1. First note that as $X \in \mathbb{Z}$ almost surely, F_X is constant on every interval of the form $(k, k+1)$, $k \in \mathbb{Z}$, and we have

$$P(X = k) = P\big(X \in [k, k+1)\big).$$

In particular, by the formula of total probability,

$$F_X(t) = \sum_{k=-\infty}^{\lfloor t \rfloor} P\big(X \in [k, k+1)\big) = \sum_{k=-\infty}^{\lfloor t \rfloor} P(X = k).$$

The discontinuity points of F_X are therefore $D \coloneqq \{k \in \mathbb{Z} : \ P(X = k) > 0\}$.

We first show that 1. implies 2.

If we assume that $X_n \xrightarrow{\text{Law}} X$, we have that for every $t \in \mathbb{R} \setminus D$, $F_{X_n}(t) \to F_X(t)$. Moreover, as $X \in \mathbb{Z}$ almost surely, one has that for every $k \in \mathbb{Z}$, and every $\epsilon \in (0,1)$,

$$P(X = k) = P(X \in (k - \epsilon, k + \epsilon]) = F_X(k + \epsilon) - F_X(k - \epsilon).$$

But, for $\epsilon > 0$, $k + \epsilon \notin D$ and $k - \epsilon \notin D$ so, on the one hand,

$$\begin{aligned}
\lim_{\epsilon \searrow 0} \lim_{n \to \infty} P(X_n \in (k - \epsilon, k + \epsilon)) &\le \lim_{\epsilon \searrow 0} \lim_{n \to \infty} P(X_n \in (k - \epsilon, k + \epsilon]) \\
&= \lim_{\epsilon \searrow 0} \lim_{n \to \infty} F_{X_n}(k + \epsilon) - F_{X_n}(k - \epsilon) \\
&= \lim_{\epsilon \searrow 0} F_X(k + \epsilon) - F_X(k - \epsilon) \\
&= \lim_{\epsilon \searrow 0} P(X = k) = P(X = k).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\lim_{\epsilon \searrow 0} \lim_{n \to \infty} P(X_n \in (k - \epsilon, k + \epsilon)) &\ge \lim_{\epsilon \searrow 0} \lim_{n \to \infty} P\big(X_n \in \big(k - \tfrac{\epsilon}{2}, k + \tfrac{\epsilon}{2}\big]\big) \\
&= \lim_{\epsilon \searrow 0} \lim_{n \to \infty} F_{X_n}\big(k + \tfrac{\epsilon}{2}\big) - F_{X_n}\big(k - \tfrac{\epsilon}{2}\big) \\
&= \lim_{\epsilon \searrow 0} F_X\big(k + \tfrac{\epsilon}{2}\big) - F_X\big(k - \tfrac{\epsilon}{2}\big) \\
&= \lim_{\epsilon \searrow 0} P(X = k) = P(X = k).
\end{aligned}$$

So, $\lim_{\epsilon \searrow 0} \lim_{n \to \infty} P(X_n \in (k - \epsilon, k + \epsilon)) = P(X = k)$.

We then show that 2. implies 1.

For $\epsilon > 0$, introduce

$$D_\epsilon = \bigcup_{k \in D} (k - \epsilon, k + \epsilon).$$

Here, our assumption is that for every $k \in \mathbb{Z}$,

$$\lim_{\epsilon \searrow 0} \lim_{n \to \infty} P\big(X_n \in (k - \epsilon, k + \epsilon)\big) = P(X = k).$$

Let $t \in \mathbb{R} \setminus D$. We need to show that $\lim_{n\to\infty} F_{X_n}(t) = F_X(t)$.

We first show that for any $\epsilon, \delta > 0$, there exist $K = K(\delta)$, $N = N(\epsilon,\delta) > 0$, such that for any $n \geq N$,

$$P(|X_n| > K) \leq \delta, \quad P(X_n \notin D_\epsilon) \leq \delta.$$

Indeed, Let K be such that $P(|X| > K-1) \leq \frac{\delta}{2}$. Let then $0 < \epsilon' < \epsilon, N$ be such that for $n \geq N$,

$$\sum_{k=-K}^{K} \left| P(X_n \in (k-\epsilon', k+\epsilon')) - P(X=k) \right| \leq \tfrac{\delta}{2}.$$

Then, for $n \geq N$,

$$P(|X_n| > K) = 1 - P(X_n \leq K) \leq 1 - \sum_{k=1-K}^{K-1} P(X_n \in (k-\epsilon, k+\epsilon))$$

$$\leq 1 + \tfrac{\delta}{2} - \sum_{k=1-K}^{K-1} P(X=k) = 1 + \tfrac{\delta}{2} - P(|X| \leq K-1) = \tfrac{\delta}{2} + P(|X| > K-1) \leq \delta.$$

Also,

$$\begin{aligned} P(X_n \notin D_\epsilon) &\leq P(X_n \notin D_{\epsilon'}) = 1 - P(X_n \in D_{\epsilon'}) \leq 1 - P(X_n \in D_{\epsilon'}, |X_n| \leq K) \\ &\leq 1 - \sum_{k=-K}^{K} P(X_n \in (k-\epsilon', k+\epsilon')) \leq 1 + \tfrac{\delta}{2} - \sum_{k=-K}^{K} P(X=k) \\ &= 1 + \tfrac{\delta}{2} - P(|X| \leq K) = \tfrac{\delta}{2} + P(|X| > K) \leq \tfrac{\delta}{2} + \tfrac{\delta}{2} = \delta. \end{aligned}$$

Now, on the one hand, for any $K > |t|$,

$$\begin{aligned} \lim_{n\to\infty} F_{X_n}(t) &\geq \lim_{\epsilon\searrow 0} \lim_{n\to\infty} P(X_n \leq t, X_n \in D_\epsilon, |X_n| \leq K+\epsilon) \\ &\geq \lim_{\epsilon\searrow 0} \lim_{n\to\infty} \sum_{k=-K}^{\lfloor t \rfloor} P(X_n \in (k-\epsilon, k+\epsilon)) = \sum_{k=-K}^{\lfloor t \rfloor} P(X=k). \end{aligned}$$

K being arbitrary, we can take $K \to \infty$ to obtain

$$\lim_{n\to\infty} F_{X_n}(t) \geq \sum_{k=-\infty}^{\lfloor t \rfloor} P(X=k) = F_X(t).$$

Let us now prove the reverse inequality. Let $\delta > 0$. By the observation made a few lines above, we have that there is $K = K(\delta) > 0$ such that

$$\begin{aligned} \lim_{n\to\infty} F_{X_n}(t) &\leq \lim_{\epsilon\searrow 0} \lim_{n\to\infty} \big(P(X_n \leq t, X_n \in D_\epsilon, |X_n| \leq K) + P(X_n \notin D_\epsilon) + P(|X_n| > K) \big) \\ &\leq 2\delta + \lim_{\epsilon\searrow 0} \lim_{n\to\infty} P(X_n \leq t, X_n \in D_\epsilon, |X_n| \leq K) \\ &\leq 2\delta + \lim_{\epsilon\searrow 0} \lim_{n\to\infty} \sum_{k=-K}^{\lfloor t \rfloor} P(X_n \in (k-\epsilon, k+\epsilon)) \\ &= 2\delta + \sum_{k=-K}^{\lfloor t \rfloor} P(X=k) \leq 2\delta + \sum_{k=-\infty}^{\lfloor t \rfloor} P(X=k) = 2\delta + F_X(t). \end{aligned}$$

$\delta > 0$ being arbitrary, we can take $\delta \to 0$ to obtain $\lim_{n\to\infty} F_{X_n}(t) \leq F_X(t)$.

Exercise 3.2.

1. Let $X_n \sim \text{Uni}([-1, \frac{1}{n}])$. Show that X_n converges in law towards a $\text{Uni}([-1,0])$ random variable
2. For $n \geq 1$, let $Y_n \sim \text{Bern}(1/n)$. Let X be a random variable with $E(X^2) < \infty$. Define $X_n = X + Y_n$. Show that X_n converges in probability towards X.
3. Let $X \sim \text{Bern}(1/2)$ and $X_n = 1 - X$ for $n \geq 1$. We have that $X_n \sim \text{Bern}(1/2)$ for all $n \geq 1$ (in particular, $X_n \xrightarrow{\text{Law}} X$ as $X_n \overset{\text{Law}}{=} X$). Show that X_n does not converge in probability towards X.

Solution 3.2.

1. The CDF of a $\text{Uni}([-1,0])$ is

$$F(t) = \begin{cases} 0 & \text{if } t \leq -1, \\ t+1 & \text{if } -1 < t < 0, \\ 1 & \text{if } t \geq 0. \end{cases}$$

In particular, it is continuous everywhere. Now, the CDF of X_n is

$$F_{X_n}(t) = \begin{cases} 0 & \text{if } t \leq -1, \\ \frac{n(t+1)}{n+1} & \text{if } -1 < t < \frac{1}{n}, \\ 1 & \text{if } t \geq \frac{1}{n}. \end{cases}$$

We need to check that for every $t \in \mathbb{R}$, $F_{X_n}(t) \to F(t)$ as $n \to \infty$. This is obvious for $t \leq -1$. Now, for $t > 0$, we have that for any $n \geq \frac{1}{t}$,

$$|F_{X_n}(t) - F(t)| = |1 - 1| = 0.$$

Remains to show the case $t \in (-1, 0)$. In this case,

$$|F_{X_n}(t) - F(t)| = \left|\tfrac{n(t+1)}{n+1} - (t+1)\right| = \tfrac{1}{n+1}\left|n(t+1) - (n+1)(t+1)\right| = \tfrac{t+1}{n+1} \xrightarrow{n\to\infty} 0.$$

2. We will show that X_n converges in L^2 towards X, which implies the convergence in law.

$$E\big((X_n - X)^2\big) = E\big((X + Y_n - X)^2\big) = E(Y_n^2) = \tfrac{1}{n} \xrightarrow{n\to\infty} 0.$$

3. We have that

$$P\big(|X - X_n| \geq 1\big) = P\big(|X - 1 + X| \geq 1\big) = P\big(|2X - 1| \geq 1\big) = 1,$$

as $X \in \{0, 1\}$ almost surely, and so $2X - 1 \in \{-1, 1\}$ almost surely, which implies $|2X - 1| = 1$ almost surely. Let $X \sim \text{Bern}(1/2)$ and $X_n = 1 - X$ for $n \geq 1$. In particular, X_n does no converge in probability towards X.

Exercise 3.3. This exercise is a prequel to the CLT, more precisely, to Theorem **??**. Let $p \in (0,1)$. Let $X_1, \ldots, X_n$ be an i.i.d. sequence of Bernoulli random variables with parameter p. Introduce

$$\tilde{S}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - p).$$

Show that $\tilde{S}_n \xrightarrow{\text{Law}} Z$, where $Z \sim \mathcal{N}(0, p(1-p))$.
Hint: remember Theorem **??**.

Solution 3.3. We will use Theorem ??. We first recall the moment generating function of $Z \sim \mathcal{N}(0, p(1-p))$ and of $X \sim \text{Bern}(p)$:

$$M_Z(t) = \exp\Big(\tfrac{p(1-p)t^2}{2}\Big), \quad M_X(t) = 1 + p(e^t - 1).$$

We want to prove that for any $t \in \mathbb{R}$,

$$M_{\tilde{S}_n}(t) \xrightarrow{n\to\infty} M_Z(t).$$

First, we have

$$\begin{aligned} M_{\tilde{S}_n}(t) &= E\Big(\exp\Big(t\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - p)\Big)\Big) = E\Big(\prod_{i=1}^n e^{t(X_i-p)/\sqrt{n}}\Big) \\ &= \prod_{i=1}^n E\big(e^{t(X_i-p)/\sqrt{n}}\big) = \prod_{i=1}^n e^{-tp/\sqrt{n}} E\big(e^{tX_i/\sqrt{n}}\big) = e^{-tp\sqrt{n}}\big(M_X(t/\sqrt{n})\big)^n, \end{aligned}$$

as the X_i's form an independent family of $\text{Bern}(p)$. This gives

$$M_{\tilde{S}_n}(t) = \exp\big(-tp\sqrt{n} + n\ln\big(M_X(t/\sqrt{n})\big)\big) = \exp\Big(-tp\sqrt{n} + n\ln\Big(1 + p(e^{t/\sqrt{n}} - 1)\Big)\Big).$$

Now, by a second order Taylor expansion around $z = 0$, we have

$$\ln(1+z) = z - \frac{z^2}{2} + o(z^2).$$

So,

$$\ln\Big(1 + p(e^{t/\sqrt{n}} - 1)\Big) = p(e^{t/\sqrt{n}} - 1) - \frac{p^2(e^{t/\sqrt{n}} - 1)^2}{2} + o\big(p^2(e^{t/\sqrt{n}} - 1)^2\big).$$

Now, for n large, another second order Taylor expansion gives

$$e^{t/\sqrt{n}} - 1 = \frac{t}{\sqrt{n}} + \frac{t^2}{2n} + o(n^{-1}).$$

Thus,

$$\begin{aligned} \ln\Big(1 + p(e^{t/\sqrt{n}} - 1)\Big) &= \frac{pt}{\sqrt{n}} + \frac{pt^2}{2n} + o(n^{-1}) - \frac{p^2t^2}{2n} + o(n^{-1}) + o(n^{-1}) \\ &= \frac{pt}{\sqrt{n}} + \frac{p(1-p)t^2}{2n} + o(n^{-1}) \end{aligned}$$

Plugging this in the expression for $M_{\tilde{S}_n}(t)$, we get

$$M_{\tilde{S}_n}(t) = \exp\Big(-tp\sqrt{n} + pt\sqrt{n} + \frac{p(1-p)t^2}{2} + o_n(1)\Big) = \exp\Big(\frac{p(1-p)t^2}{2} + o_n(1)\Big),$$

taking $n \to \infty$, $o_n(1)$ goes to 0, and we are left with

$$\lim_{n\to\infty} M_{\tilde{S}_n}(t) = \exp\Big(\tfrac{p(1-p)t^2}{2}\Big) = M_Z(t),$$

as wanted.

Exercise 3.4. This exercise studies a different regime than Exercise **??**. Here we will let the parameter of the Bernoulli's tend to 0 as n goes to ∞ and show that it changes the natural re-scaling (from $\sqrt{n}$ to 1), as well as the nature of the limiting random variable. Let $\lambda > 0$. For $n \geq 1$, let $p_n \in (0,1)$ be a sequence of numbers such that

$$\lim_{n\to\infty} np_n = \lambda,$$

and let $X_{n,1}, X_{n,2}, \ldots, X_{n,n} : \Omega \to \mathbb{R}$ be an i.i.d. family of random variable with law $\text{Bern}(p_n)$. Define

$$S_n = \sum_{i=1}^{n} X_{n,i}.$$

Show that $S_n \xrightarrow{\text{Law}} N$, where $N \sim \text{Poi}(\lambda)$.
*Hint: recall Lemma **??**, and look first at $\lim_{n\to\infty} P(S_n = k)$ for k fixed.*

Solution 3.4. By Lemma **??**, it is sufficient to show that for any $k \in \mathbb{N}$,

$$\lim_{n\to\infty} P(S_n = k) = e^{-\lambda}\tfrac{\lambda^k}{k!} = P(N = k).$$

Now, S_n is the sum of n Bernoulli random variables with parameter p_n, thus $S_n \sim \text{Bin}(n, p_n)$, so

$$P(S_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{1}{k!}(1-p_n)^n \prod_{i=1}^{k} p_n(n-k+i).$$

Now, we first have

$$(1-p_n)^n = \exp(n\ln(1-p_n)) = \exp\big(n\big(-p_n + o(n^{-1})\big)\big)$$

by a first order Taylor expansion, as $\lim_{n\to\infty} p_n n = \lambda$, so $p_n \leq C/n$ for some $C > 0$. We get,

$$\lim_{n\to\infty}(1-p_n)^n = \lim_{n\to\infty}\exp\big(n\big(-p_n + o(n^{-1})\big)\big) = \exp\Big(-\lim_{n\to\infty} np_n\Big) = e^{-\lambda}.$$

On the other hand, (recall k does not depend on n),

$$\lim_{n\to\infty}\prod_{i=1}^{k} p_n(n-k+i) = \prod_{i=1}^{k}\big(\lim_{n\to\infty} p_n(n-k+i)\big) = \prod_{i=1}^{k}\big(\lim_{n\to\infty} p_n n\big(1-\tfrac{k+i}{n}\big)\big) = \prod_{i=1}^{k}\lambda = \lambda^k.$$

Combining, we have

$$\begin{aligned}\lim_{n\to\infty} P(S_n = k) &= \lim_{n\to\infty}\frac{1}{k!}(1-p_n)^n\prod_{i=1}^{k} p_n(n-k+i)\\ &= \frac{1}{k!}\big(\lim_{n\to\infty}(1-p_n)^n\big)\Big(\lim_{n\to\infty}\prod_{i=1}^{k} p_n(n-k+i)\Big) = \frac{1}{k!}\lambda^k e^{-\lambda}\end{aligned}$$

which is exactly what we wanted.

Exercise 3.5. Show that

1. Convergence in probability implies convergence in law.

2. Convergence in L^p implies convergence in probability.

Solution 3.5. Let $X_1, X_2, \ldots$ be a sequence of random variables. Let X be a random variable.

1. Suppose $X_n \xrightarrow{\text{Proba}} X$. Let $t \in \mathbb{R}$ be such that F_X is continuous at t. We need to show that

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t).$$

First, for any $\epsilon > 0$,

$$\begin{aligned} F_{X_n}(t) &= P(X_n \le t) \\ &= P(X_n \le t,\ |X_n - X| < \epsilon) + P(X_n \le t,\ |X_n - X| \ge \epsilon) \\ &= P(X_n \le t,\ |X_n - X| < \epsilon,\ X \le t+\epsilon) + P(X_n \le t,\ |X_n - X| \ge \epsilon) \\ &\le P(X \le t+\epsilon) + P(|X_n - X| \ge \epsilon) \end{aligned}$$

where we used the formula of total probability and the (deterministic) fact that if $x \le t$ and $|x-y| \le \epsilon$, then $y \le t+\epsilon$. Thus,

$$\limsup_{n\to\infty} F_{X_n}(t) \le P(X \le t+\epsilon) + \limsup_{n\to\infty} P(|X_n - X| \ge \epsilon) = F_X(t+\epsilon)$$

as $X_n \xrightarrow{\text{Proba}} X$. Now, $\epsilon > 0$ is arbitrary, so we can take $\epsilon \searrow 0$ on both sides, and use continuity of F_X at t to obtain

$$\limsup_{n\to\infty} F_{X_n}(t) \le= \lim_{\epsilon\searrow 0} F_X(t+\epsilon) = F_X(t).$$

Let us prove the reverse inequality. We have

$$\begin{aligned} F_{X_n}(t) &= P(X_n \le t) \\ &\ge P(X_n \le t,\ |X_n - X| < \epsilon,\ X \le t-\epsilon) \\ &= P(|X_n - X| < \epsilon,\ X \le t-\epsilon) \\ &= P(X \le t-\epsilon) - P(|X_n - X| \ge \epsilon,\ X \le t-\epsilon) \\ &\ge P(X \le t-\epsilon) - P(|X_n - X| \ge \epsilon) \end{aligned}$$

where we used the same deterministic fact as before. As $X_n \xrightarrow{\text{Proba}} X$, we have that for any $\epsilon > 0$,

$$\liminf_{n\to\infty} F_{X_n}(t) \ge P(X \le t-\epsilon) - \liminf_{n\to\infty} P(|X_n - X| \ge \epsilon) = F_X(t-\epsilon).$$

As $\epsilon > 0$ is arbitrary, we can take $\epsilon \searrow 0$ on both sides and use continuity of F_X at t to obtain

$$\liminf_{n\to\infty} F_{X_n}(t) \ge \lim_{\epsilon\searrow 0} F_X(t-\epsilon) = F_X(t).$$

So, for every $t \in \mathbb{R}$ such that F_X is continuous at t, $F_X(t) \le \liminf_{n\to\infty} F_{X_n}(t) \le \limsup_{n\to\infty} F_{X_n}(t) \le F_X(t)$, and thus $\lim_{n\to\infty} F_{X_n}(t) = F_X(t)$, which is convergence in law.

2. Suppose $X_n \xrightarrow{L^p} X$. Then, by Chebychev's inequality (the moment version), one has that for any $\epsilon > 0$

$$P(|X - X_n| \geq \epsilon) \leq \frac{E\big(|X_n - X|^p\big)}{\epsilon^p}.$$

Now, as $X_n \xrightarrow{L^p} X$, $E\big(|X_n - X|^p\big) \to 0$ as $n \to \infty$. So, for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X - X_n| \geq \epsilon) = 0,$$

which is the definition of convergence in probability.

Exercise 3.6. Let $X_1, X_2, \ldots$ be a sequence of measurements on repetition of the same experiment (independent sequence of identically distributed random variables). What can we say on

$$S_n = \sum_{i=1}^{n} X_i,$$

as $n \to \infty$ in each of the following cases.

1. The experiment is picking an individual in the population uniformly at random, and X_i is the indicator function that the ith individual knows how to fix a leaking pipe.

2. The experiment is flipping a fair coin until getting "head". X_i is 10 to the power the number of tails before getting head in the ith repetition of the experiment (for example, you play a game where your lose is multiplied by 10 every times you get "tail").

3. We can assume that the X_i's are continuous random variables with common density
$$f(x) = 2\mathbb{1}_{[1,+\infty)}(x)\frac{1}{x^3}.$$

Solution 3.6. 1. The X_i are bounded random variables. In particular, $E(|X_1|) < \infty$, so the (strong) LLN implies that

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} E(X_1).$$

So, we know that S_n behaves like $nE(X_1)$ for n large. Moreover, still as the X_i's are bounded, $E(X_1^2) < \infty$ and thus the CLT implies that

$$\frac{S_n - nE(X_1)}{\sqrt{n}} \xrightarrow{\text{a.s.}} \mathcal{N}\big(0, \mathrm{Var}(X_1)\big).$$

In particular, we can approximate, for n large, the risk that S_n deviates from $nE(X_1)$ by more than $K\sqrt{n}$:

$$P\big(|S_n - nE(X_1)| \geq K\sqrt{n}\big) = P\Big(\big|\tfrac{S_n - nE(X_1)}{\sqrt{n}}\big| \geq K\Big) \approx \frac{2}{\sqrt{2\pi \mathrm{Var}(X_1)}} \int_K^\infty e^{-\frac{x^2}{2\mathrm{Var}(X_1)}}.$$

2. We have that $P(X_1 = 10^k) = 2^{-k-1}$. So,

$$E(|X_1|) = \sum_{k=0}^{\infty} 10^k 2^{-k-1} = \tfrac{1}{2} \sum_{k=0}^{\infty} 5^k = +\infty.$$

We thus have that the X_i's do not have a first or second moment, the LLN and the CLT thus say nothing about S_n.

3. We have that

$$E(|X_1|) = 2\int_1^{\infty} x \cdot x^{-3} dx = 2\int_1^{\infty} x^{-2} dx = 2[-x^{-1}]_1^{\infty} = 2,$$

so the X_i's have a first moment. The (strong) LLN thus tells us that

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} E(X_1).$$

So, as in the first case, we know that S_n behaves like $nE(X_1)$ for n large. But, unlike in the first case,

$$E(X_1^2) = 2\int_1^{\infty} x^2 \cdot x^{-3} dx = 2\int_1^{\infty} x^{-1} dx = 2[\ln(x)]_1^{\infty} = +\infty,$$

so the X_i's do not have a second moment. The CLT does not give any information in this case, and we thus cannot estimate the risk that S_n deviates from $nE(X_1)$!

Exercise 3.7. We perform the following experiment: we pick graphic cards uniformly at random in the NVIDIA production and test them. Let X_i be the random variable giving 0 is the ith card tested functions well, and 1 if there is a malfunction. Denote ρ the proportion of cards that have malfunction in the production. ρ is supposed to be unknown, so you cannot use it directly in computations.
In the following, you can use that if $Z \sim \mathcal{N}(0,1)$,

$$P(|Z| > 1.96) = 0.05, \quad P(|Z| > 2.31) = 0.04.$$

1. Propose (and justify your proposition) a method to "guess" ρ from the sequence $X_1, X_2, X_3, \ldots$.

2. Using Theorem **??** and some non-rigorous approximation, can you give a minimal number of tests to be performed if we want to have an estimation of ρ that is precise up to an error 0.01, with a risk a failing the estimation at most 0.05?

3. Using Theorem **??**, can you **rigorously** give a minimal number of tests to be performed if we want to have an estimation of ρ that is precise up to an error 0.01?

Solution 3.7. 1. We have that our random variables X_i are Bernoulli random variables. We can suppose them independent as being repeated experiment withour clear link between them (we can assume that we do not test twice the same card). As X_i gives 1 with probability ρ, we have $E(X_i) = \rho$ for all i's. A good way to guess ρ is to find a way to approximation of $E(X_1)$. This is provided by the sLLN:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} E(X_1).$$

We can thus take $\frac{1}{n}\sum_{i=1}^{n} X_i$ as an approximation of ρ as long as we take n large enough.

2. As we know nothing on ρ, we can only assume that the variance of X_i is given by some number σ^2 satisfying

$$\sigma^2 = \text{Var}(X_i) = \rho(1-\rho) \leq 1.$$

We want to find n such that

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_1)\Big| > 0.01\Big) \leq 0.05.$$

We can use that $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - E(X_i))$ is approximately a centred Gaussian random variable with variance σ^2. Thus,

$$\begin{aligned} P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} X_i - E(X_1)\Big| > 0.01\Big) &= P\Big(\Big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - E(X_i))\Big| > 0.01\sqrt{n}\Big) \\ &\approx P\big(|\mathcal{N}(0,\sigma^2)| > 0.01\sqrt{n}\big) \\ &= P\big(|\mathcal{N}(0,1)| > \tfrac{\sqrt{n}}{100\sigma}\big) \\ &\leq P\big(|\mathcal{N}(0,1)| > \tfrac{\sqrt{n}}{100}\big). \end{aligned}$$

As we know nothing about σ except $\sigma \leq 1$, we used the worst case scenario in the last line. Using the indication, we want n such that

$$\tfrac{\sqrt{n}}{100} \geq 1.96 \iff n \geq 38416.$$

3. Let $\tilde{S}_n = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - E(X_i))$. Let $Z \sim \mathcal{N}(0,1)$. We start as in the previous point and replace the use of $\approx$ by a suitable use of Theorem **??**. Note that $E(|X_i|^3) \leq 1$ is the only information we have on the third moment of the X_i's. We have that (again using a worst case scenario on σ)

$$\begin{aligned} P\big(|\tilde{S}_n| > \tfrac{\sqrt{n}}{\sigma 100}\big) &\leq P\big(|\tilde{S}_n| > \tfrac{\sqrt{n}}{100}\big) \\ &\leq P\big(\tilde{S}_n \leq -\tfrac{\sqrt{n}}{100}\big) + 1 - P\big(\tilde{S}_n \leq \tfrac{\sqrt{n}}{100}\big) \\ &\leq 2\tfrac{0.5\cdot E(|X_1|^3)}{\sqrt{n}} + P\big(Z \leq -\tfrac{\sqrt{n}}{100}\big) + 1 - P\big(Z \leq \tfrac{\sqrt{n}}{100}\big) \\ &\leq \tfrac{1}{\sqrt{n}} + P\big(|Z| > \tfrac{\sqrt{n}}{100}\big) \end{aligned}$$

where we used Theorem **??** in the third inequality. Using the hint, we can, for example, search for n such that

$$\tfrac{1}{\sqrt{n}} \leq 0.01 \text{ and } P\big(|Z| > \tfrac{\sqrt{n}}{100}\big) \leq 0.04.$$

For the second inequality, the given Gaussian tails yield $n \geq 231^2 = 53361$, and we can thus take

$$n = \max(10000, 53361) = 53361.$$