Convergence of Random Variables

Here, in this part, we will consider infinite sequences of random variables and their limiting (asymptotic) behavior. In fact, one of the subfield of the study in statistics is the Large Sample Statistics, or the Asymptotic Statistics, where the asymptotic behavior of statistics and tests are studied. The idea is that when one wants to consider a laarge sample, say, we have 10^{10} observations, then it is sometimes much easier not to consider what happen when $n = 10^{10}$, but to consider the limit when $n \to +\infty$. Say, if we will ask every person on Earth to choose an integer number from $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$, and denote the choice of n-th person by X_n, then the average number chosen will be

$$\bar{X}_N = \frac{X_1 + X_2 + ... + X_N}{N},$$

where $N \approx 7.608$bln as of 2018 March, the total population on the Earth[1]. Now, instead of calculating this fraction, we can calculate the limit of that: because of the LLN (see forthcoming), we will have that

$$\bar{X}_n \to \mathbb{E}(X_1) = \frac{1}{10} \cdot 0 + \frac{1}{10} \cdot 1 + \frac{1}{10} \cdot 2 + ... + \frac{1}{10} \cdot 9 = \frac{45}{10} = 4.5,$$

assuming that the choice of digits is uniform. Then we can state that $\bar{X}_N \approx 4.5$, since N is large enough[2].

In parametric statistics, one is considering a sample of size n, and estimating an unknown quantity (parameter) θ through a r.v. variable θ_n, which is an estimate for that unknown θ. Here θ_n depends on the sample size n (we will consider point estimation in the next part). Usually, we want to study the asymptotic behaviour of θ_n, the behaviour of θ_n for large samples: for "good" estimates we want to have that θ_n approaches θ in some sense.

Here, in this part we will define different types of convergences and investigate their properties and relationship.

We will assume that we are given a sequence of r.v. $X_1, X_2, ..., X_n, ...$ (infinite number of r.v.). Usually, we will assume that all these r.v. are defined on the same Experiment, on the same Probability Space. That is, mostly we will assume that $X_n = X_n(\omega)$, where $\omega \in \Omega$ for some fixed Ω. Only when defining the Convergence in Distributions notion, we will rest(skip?) this assumption.

As examples of sequences of r.v. we can consider:

Example, Sequences of R.V.:

- Let X_n be the daily closing price for one Facebook Inc. stock at the day n calculated from now. Then X_n will be a sequence of r.v. . What is the Sample Space behind each X_k? We can assume that the sample space is the set of all possible market scenarios for the rest of the world.

[1] Well, of course, newborn babies will not choose for your experiment numbers ⌣̈ But let's assume, for the effect of the presentation.

Btw, can you estimate how many people have ever lived on Earth?

[2] Well, mathematically, this is not correct, but in practice ... ⌣̈

Here one fact is notable: X_n-s will not be independent, the price tomorrow will depend somehow from the today's price. Here, in our course, we will mainly consider IID sequences, i.e. sequences X_k such that all X_k-s are Identically Distributed (have the same distribution) and are Independent.

- Assume we are tossing a coin infinitely many times[3], $X_n = 0$, if the n-th toss resulted in tails, and $X_n = 1$, if heads.
- Let X_n be the number of insurance claims for some insurance company for the day n calculated from today, and let Y_n be the claim size for that day. Then X_n and Y_n are sequences of r.v.

A naive way to think about the sequence of r.v.'s on the same probability space - say, we have (countably) infinite number of persons numbered by $1, 2, 3, ...$, watching after the result of the experiment. Everybody has his/her own function to calculate for any possible outcome (say, the Sample Space is $\Omega = [1, 6]$, the first person will calculate the square of ω, i.e. $X_1(\omega) = \omega^2$, the second person will calculate the inverse of ω, $X_2(\omega) = \frac{1}{\omega}$ etc.). The only unsure thing is - which outcome will happen. So before doing the experiment, the choice of every person is a r.v..

8.1 Convergence of a sequence of r.v.'s

Assume now we have a sequence of r.v. X_n, $n \in \mathbb{N}$, defined on the same Probability Space $(\Omega, \mathcal{F}, \mathbb{P})$. We want to define some convergence notions for that sequence.

Assume also X is a r.v. on the same space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 8.1. *We will say that $X_n \to X$ almost sure, and we will write $X_n \to X$ a.s. or $X_n \xrightarrow{a.s.} X$, if*

$$\mathbb{P}\Big(\omega \in \Omega : \lim_{n\to+\infty} X_n(\omega) = X(\omega)\Big) = 1,$$

or, for short,

$$\boxed{\mathbb{P}\Big(X_n \to X\Big) = 1}$$

Equivalently, we can write

$$X_n \xrightarrow{a.s.} X \qquad \text{iff} \qquad \mathbb{P}\Big(X_n \not\to X\Big) = 0.$$

You can think like this: X_n is a r.v., so it is a function of ω. If we fix an ω, then $X_n(\omega)$ becomes a numerical sequence. For this numerical sequence, we can consider its limit. If the limit exists at ω and is equal to $X(\omega)$, then ω is from the set $X_n \not\to X$. Now, the a.s. convergence idea is that the set (of all ω-s) where X_n does not tend to X is of probability 0.

Definition 8.2. *We will say that $X_n \to X$ in Probability, and we will write $X_n \xrightarrow{\mathbb{P}} X$, if*

$$\boxed{\textit{for any } \varepsilon > 0, \ \mathbb{P}\Big(|X_n - X| \geqslant \varepsilon\Big) \to 0, \qquad \text{when} \quad n \to \infty.}$$

Equivalently, we can write

$$X_n \xrightarrow{\mathbb{P}} X \qquad \text{iff} \qquad \mathbb{P}\Big(|X_n - X| < \varepsilon\Big) \to 1 \text{ for any } \varepsilon > 0.$$

Sometimes we will use the first form, with $\mathbb{P}\Big(|X_n - X| \geqslant \varepsilon\Big) \to 0$, and sometimes we will use the equivalent form $\mathbb{P}\Big(|X_n - X| < \varepsilon\Big) \to 1$.

Another notion of convergence is the Quadratic Mean convergence:

Definition 8.3. *We will say that $X_n \to X$ in Quadratic Mean or in L^2 (or in Mean Square Sense), and we will write $X_n \xrightarrow{L^2} X$ or $X_n \xrightarrow{qm} X$, if*

$$\mathbb{E}\Big((X_n - X)^2\Big) \to 0, \qquad \text{when} \quad n \to \infty.$$

In the above definitions, it is important that all random variables are defined on the same probability space. This is necessary to calculate $X_n(\omega)$ and $X(\omega)$ for the same ω.

On the other hand, for the CLT, this is too restrictive, since CLT works for a broad range of r.v.s, not necessarily defined on the same probability space. So we give another type of convergence notion, which will work also in that case. The idea is even if the random variables are defined on completely different spaces, their CDF's are just real valued functions defined on $\mathbb{R}$, so we can talk about the convergence of their CDF's.

So now we assume that X_n and X are arbitrary r.v.'s, not necessarily defined on the same probability space, and $F_{X_n}(x)$ and $F_X(x)$ are their CDF's, respectively.

Definition 8.4. *We will say that $X_n \to X$ in Distribution, and we will write $X_n \xrightarrow{D} X$, if*

$$F_{X_n}(x) \to F_X(x) \quad \text{when} \quad n \to \infty \ \ \textit{at any point of continuity } x \textit{ of } F_X(x).$$

Remark, Convergence in Distributions: The above definition says that $X_n \xrightarrow{D} X$ iff

$$\mathbb{P}(X_n \leqslant x) \to \mathbb{P}(X \leqslant x),$$

at any point of continuity of $\mathbb{P}(X \leqslant x)$. So convergence in distribution is the convergence of corresponding probabilities.

In general, if $X_n \xrightarrow{D} X$, then for many subsets $A \subset \mathbb{R}$, one will have

$$\mathbb{P}(X_n \in A) \to \mathbb{P}(X \in A).$$

In fact, the convergence will hold for all **continuity sets** A of random variable X, see https://en.wikipedia.org/wiki/Continuity_set. And, in particular, for any $a, b \in \mathbb{R}$ with $\mathbb{P}(X = a) = \mathbb{P}(X = b) = 0$, we will have

$$\mathbb{P}(a \leqslant X_n \leqslant b) \to \mathbb{P}(a \leqslant X \leqslant b), \qquad n \to +\infty.$$

Proposition 8.1. *Assume X_n is a sequence of r.v. and X is a r.v. . Then*

$$X_n \xrightarrow{D} X \qquad \text{iff} \qquad \mathbb{E}(h(X_n)) \to \mathbb{E}(h(X)) \quad \textit{for any continuous and bounded function } h.$$

The following chart represents the relationship between different types of convergences:

$$\left.\begin{array}{l}\text{Convergence in QM}\\ \text{Convergence AS}\end{array}\right. \Rightarrow \text{Convergence in Probability} \Rightarrow \text{Convergence in Distribution}$$

REMARK, RELATIONSHIP BETWEEN DIFFERENT TYPES OF CONVERGENCES: It can be proved, that the inverse implications in the above diagram are not true, in general.

!!! Give examples!

But, we can state some partial results:

a. If $X_n \xrightarrow{D} X$ and $X \equiv constant$, then $X_n \xrightarrow{P} X$;

b. If $X_n \xrightarrow{P} X$, then there is a subsequence of natural numbers n_k such that $X_{n_k} \xrightarrow{a.s.} X$;

c. If $X_n \xrightarrow{P} X$ and there exists a number M such that $\mathbb{P}(|X_n| \leqslant M) = 1$ for any n, then $X_n \xrightarrow{q.m.} X$.

EXAMPLE, CONVERGENCE IN PROBABILITY BUT NOT A.S.: This is a classical example of a sequence of r.v. X_n such that X_n tends to 0 in probability but not a.s. .

We consider the Sample Space $\Omega = [0,1]$ with a probability $\mathbb{P}([a,b]) = b-a$, for $[a,b] \subset \Omega$. We define the sequence X_n in the following way:

$$X_1(\omega) \equiv 1, \qquad X_2(\omega) = \begin{cases} 1, & \omega \in \left[0, \frac{1}{2}\right] \\ 0, & \text{otherwise} \end{cases} \qquad X_3(\omega) = \begin{cases} 0, & \omega \in \left[0, \frac{1}{2}\right] \\ 1, & \text{otherwise} \end{cases}$$

$$X_4(\omega) = \begin{cases} 1, & \omega \in \left[0, \frac{1}{2^2}\right] \\ 0, & \text{otherwise} \end{cases} \qquad X_5(\omega) = \begin{cases} 1, & \omega \in \left[\frac{1}{2^2}, \frac{2}{2^2}\right] \\ 0, & \text{otherwise} \end{cases}$$

$$X_6(\omega) = \begin{cases} 1, & \omega \in \left[\frac{2}{2^2}, \frac{3}{2^2}\right] \\ 0, & \text{otherwise} \end{cases} \qquad X_7(\omega) = \begin{cases} 1, & \omega \in \left[\frac{3}{2^2}, 1\right] \\ 0, & \text{otherwise} \end{cases}$$

and so on. In fact, to construct r.v. X_n for $n = 8, 9, 10, ..., 15$, we divide $[0,1]$ into 8 equal-length intervals and define

$$X_{2^3+k}(\omega) = \begin{cases} 1, & \omega \in \left[\frac{k}{2^3}, \frac{k+1}{2^3}\right] \\ 0, & \text{otherwise} \end{cases} \qquad \text{for all} \quad k = 0, 1, 2, ..., 7.$$

And, in general, for any $m = 0, 1, 2, ...$ and $k = 0, 1, 2, ..., 2^m - 1$

$$X_{2^m+k}(\omega) = \begin{cases} 1, & \omega \in \left[\frac{k}{2^m}, \frac{k+1}{2^m}\right] \\ 0, & \text{otherwise} \end{cases}$$

The graphs of the first few function (r.v.-s) X_k are give in Fig. ....

Now, first we prove that at any $\omega \in [0,1]$, $X_n(\omega)$ is divergent. To prove this, it is enough to note that the numerical sequence $X_n(\omega)$ consists of infinitely many 0-s and infinitely many 1-s, so it cannot converge. This, particularly, means that

$$\mathbb{P}(\omega \in \Omega | X_n(\omega) \text{converges}) = 0,$$

that is, X_n diverges a.s. (in fact, diverges surely, at any point).

On the other hand, we can see that X_n converges to 0 in the sense of Probabilities. To see this, we fix a positive $\varepsilon < 1$, and calculate $\mathbb{P}(|X_n - 0| \geqslant \varepsilon) \overset{X_n \text{is non-negative}}{=\!=} \mathbb{P}(X_n \geqslant \varepsilon)$. Clearly,

$$\mathbb{P}(X_1 \geqslant \varepsilon) = \mathbb{P}([0,1]) = 1, \quad \mathbb{P}(X_2 \geqslant \varepsilon) = \mathbb{P}([0, \frac{1}{2}]) = \frac{1}{2}, \quad \mathbb{P}(X_3 \geqslant \varepsilon) = \mathbb{P}([\frac{1}{2}, 1]) = \frac{1}{2},$$

$$\mathbb{P}(X_4 \geqslant \varepsilon) = \mathbb{P}([0, \frac{1}{2^2}]) = \frac{1}{2^2}, \mathbb{P}(X_5 \geqslant \varepsilon) = \mathbb{P}([\frac{1}{2^2}, \frac{2}{2^2}]) = \frac{1}{2^2}, \dots$$

and so on, so the sequence $\mathbb{P}(X_n \geqslant \varepsilon)$ has the form

$$1, \frac{1}{2}, \frac{1}{2}, \frac{1}{2^2}, \frac{1}{2^2}, \frac{1}{2^2}, \frac{1}{2^2}, \frac{1}{2^3}, \frac{1}{2^3}, \dots$$

and this clearly tends to 0. This means that $X_n \overset{\mathbb{P}}{\to} 0$.

In fact, for convergence a.s., most of the algebraic operations with sequences remain true. But this is **not true for the convergence in Distributions**.

Proposition 8.2. *(Uniqueness of a limit)*

a. *Assume* $X_n \xrightarrow{a.s.} X$ *and* $X_n \xrightarrow{a.s.} Y$. *Then* $X = Y$ *a.s., i.e.,* $\mathbb{P}(X = Y) = 1$ *(or, in other words,* $\mathbb{P}(X \neq Y) = 0$*).*

b. *Assume* $X_n \xrightarrow{\mathbb{P}} X$ *and* $X_n \xrightarrow{\mathbb{P}} Y$. *Then* $X = Y$ *a.s..*

c. *Assume* $X_n \xrightarrow{q.m.} X$ *and* $X_n \xrightarrow{q.m.} Y$. *Then* $X = Y$ *a.s..*

Remark, Non-uniquesness of the limit in the Convergence in Distributions case: It is important that the above proposition is not true for the convergence in Distributions, that is, it can happen that $X_n \xrightarrow{D} X$ and $X_n \xrightarrow{D} Y$, where $X \neq Y$ at most of the points, and even, $X \neq Y$ a.s.

For example, consider some r.v. with symmetric distribution around 0, e.g., $X \sim \mathcal{N}(0,1)$. Take $X_n = X$ for all n. Clearly, $X_n \xrightarrow{D} X$. But, also $X_n \xrightarrow{D} -X$. To see this, we need to prove that for any $x \in \mathbb{R}$

$$F_{X_n}(x) \to F_{-X}(x)$$

(as $-X$ is a continuous r.v., so F_{-X} is continuous everywhere). But it is easy to prove that $F_{-X}(x) = F_X(x)$ for any x, because of the symmetry of the distribution of X:

$$F_{-X}(x) = \mathbb{P}(-X \leqslant x) = \mathbb{P}(X \geqslant -x) \overset{\text{because of the symmetry}}{=\!=\!=} \mathbb{P}(X \leqslant x) = F_X(x).$$

Now, $X_n \xrightarrow{D} X$ and $X_n \xrightarrow{D} -X$. Maybe $X = -X$ a.s.? Of course, no! Because

$$\mathbb{P}(X = -X) = \mathbb{P}(2X = 0) = \mathbb{P}(X = 0) = 0,$$

so $X \neq -X$ a.s..

So we can have that the same sequence of r.v. X_n can have several (in fact, a lot of) different limits in the sense of Distributions convergence, but, important point is that **the distribution of all limiting r.v.'s will be the same**. Say, in our above case, X and $-X$ both have $\mathcal{N}(0,1)$ distribution.

Proposition 8.3. *Assume* $X_n \xrightarrow{a.s.(P)} X$ *and* $Y_n \xrightarrow{a.s.(P)} Y$. *Then* *a.* $X_n + Y_n \xrightarrow{a.s.(P)} X + Y$;

b. $X_n \cdot Y_n \xrightarrow{a.s.(P)} X \cdot Y$;

c. If $g \in C(\mathbb{R})$, *then* $g(X_n) \xrightarrow{a.s.(P)} g(X)$

Proposition 8.4. *Assume* $X_n \xrightarrow{L^2} X$ *and* $Y_n \xrightarrow{L^2} Y$. *Then* $X_n + Y_n \xrightarrow{L^2} X + Y$.

The same properties are not true, in general, for the convergence in Distributions. For example, in the general case, if $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} Y$, then not necessarily[4] $X_n + Y_n \xrightarrow{D} X + Y$.

Example: Assume $Z \sim \mathcal{N}(0,1)$, $X_n = -Z + \frac{1}{n}$, $Y_n = -X_n$. Then $X_n \xrightarrow{D} Z$ and $Y_n \xrightarrow{D} Z$, but $X_n + Y_n = 0$. ■

The next Proposition gives some properties we can use when dealing with the convergence in distributions.

Theorem 8.1 (Slutsky's Theorem)**.** *Assume* $X_n \xrightarrow{D} X$ *and* $Y_n \xrightarrow{P} c$, *where* $c \in \mathbb{R}$ *is a constant. Then*

a. $X_n + Y_n \xrightarrow{D} X + c$;

b. $X_n \cdot Y_n \xrightarrow{D} c \cdot X$.

Theorem 8.2 (Continuous Mapping Theorem)**.** *Assume* $X_n \xrightarrow{D} X$ *and* $g \in C(\mathbb{R})$. *Then* $g(X_n) \xrightarrow{D} g(X)$.

Sometimes, when using the convergence in distributions, we will write something like $X_n \xrightarrow{D} \mathcal{N}(\mu, \sigma^2)$ instead of writing $X_n \xrightarrow{D} X$ for some $X \sim \mathcal{N}(\mu, \sigma^2)$. This is because X is not important, its distribution is important in case of the convergence in Distributions.

8.2 Examples of convergence

Usually, one will have 2 types of examples of sequence of r.v., and we will give that 2 types of examples, explaining or checking their convergence.

The first type of examples are when we describe r.v. X_n **explicitly**, i.e., as a function of ω. So in that case we talk about a **concrete example of a sequence**. Usually, the explicit form of X_n will not be available: say, if X_n will be the claim size for the day n, then nobody will give a formula like $X_n(\omega) = 100n + \omega^2$ - even it is hard to describe what is ω here: usually, ω represents the scenario, and it will be non-numerical, and will be undescribable (??). So usually, this type of examples are introduced to explain the ideas of convergence notions.

The second type of examples concern r.v. X_n that are given by distribution. So we talk about the distribution of X_n, **without explicitly describing** X_n. Say, we are talking about $X_n \sim \mathtt{Bernoulli}(0.5)$: this means that for any fixed n, X_n is some function (defined on some Sample Space Ω, which is not shown anywhere), which takes the values 0 and 1 (and we are not specifying for which ω the value is 0, and for which - 1), with probabilities 0.5 both (so for "half" of the values[5] ω, the value $X_n(\omega)$ will be 0, and for all others will be 1). Of course, we can have different sets for taking the values 0 and 1 for different n.

Or, say, we will consider $X_n \sim \mathtt{Unif}\left(\left[0, \frac{1}{n}\right]\right)$. In this case we will calculate the limit of X_n in one of the senses above **irrespective** to the concrete realisation, concrete value of X_n. I am stressing

[4] The problem is that we need some information about the Joint Distribution of (X_n, Y_n).

[5] I am using here "half" of the values meaning that the probability of the set of that ω-s is 0.5.

this because we can have a lot of r.v.'s X_n having the same distribution, so $X_n \sim \mathtt{Unif}\left(\left[0, \frac{1}{n}\right]\right)$ is not unique[6]! But the convergence will hold in any case!

Now, the examples.

8.2.1 R.V. described Explicitly

EXAMPLE, CONVERGENCE OF A R.V. SEQUENCE: Assume we are rolling a fair die, $\Omega = \{1,2,3,4,5,6\}$ is the Sample Space, and assume X_n is calculating the result shown on the die divided by n, i.e.

$$X_n(\omega) = \frac{\omega}{n}, \qquad \omega \in \Omega.$$

We can describe this as: if the result will be 3, then, say, the first person is calculating $X_1 = \frac{3}{1} = 3$, the second person calculates $X_2 = \frac{3}{2}$, for the third one $X_3 = \frac{3}{3} = 1$ etc.

Let us prove that $X_n \to 0$ in all four senses. Let us denote $X \equiv 0$, and prove that $X_n \to X$ is all senses.

Almost Sure Convergence: To prove that $X_n \xrightarrow{a.s.} X$, we need to see for which $\omega \in \Omega$ we will have

$$X_n(\omega) \to X(\omega).$$

Now, if $\omega \in \Omega$ is fixed, then

$$X_n(\omega) = \frac{\omega}{n} \to 0 = X(\omega).$$

This means that for any ω, $X_n(\omega) \to X(\omega)$. Then the set $\{\omega \in \Omega : X_n(\omega) \to X(\omega)\} = \Omega$, so

$$\mathbb{P}(X_n \to X) = \mathbb{P}(\Omega) = 1.$$

Quadratic Mean Convergence: To prove that $X_n \xrightarrow{qm} X$, we need to prove that

$$\mathbb{E}\Big((X_n - X)^2\Big) \to 0.$$

To this end, we need to calculate the expected value on the left:

$$\mathbb{E}\Big((X_n - X)^2\Big) = \mathbb{E}\Big((X_n)^2\Big).$$

Now, let us build the distribution (PMF) of X_n:

Values of X_n	$\frac{1}{n}$	$\frac{2}{n}$	$\frac{3}{n}$	$\frac{4}{n}$	$\frac{5}{n}$	$\frac{6}{n}$
$\mathbb{P}(X_n = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Then,

$$\mathbb{E}\Big((X_n - X)^2\Big) = \mathbb{E}\Big((X_n)^2\Big) = \frac{1}{6}\cdot\left(\frac{1}{n}\right)^2 + \frac{1}{6}\cdot\left(\frac{2}{n}\right)^2 + \frac{1}{6}\cdot\left(\frac{3}{n}\right)^2 + \frac{1}{6}\cdot\left(\frac{4}{n}\right)^2 + \frac{1}{6}\cdot\left(\frac{5}{n}\right)^2 + \frac{1}{6}\cdot\left(\frac{6}{n}\right)^2 \to 0,$$

since each term goes to 0 as $n \to \infty$.

[6] Give explicit different examples of $X \sim \mathtt{Unif}[0,1]$, i.e., construct one or many probability spaces Ω and different formulas for $X(\omega)$ such that $X \sim \mathtt{Unif}[0,1]$

Convergence in Probability: Well, in fact, we can use the property that convergence a.s. (or in Quadratic Mean) implies Convergence in Probability. But let us prove the Convergence in Probability using the very definition. So to prove that $X_n \xrightarrow{\mathbb{P}} X$, we need to prove that

$$\text{for any } \varepsilon > 0,\ \mathbb{P}\Big(|X_n - X| \geqslant \varepsilon\Big) \to 0, \qquad \text{when} \quad n \to \infty.$$

Let us fix an $\varepsilon > 0$. Then,

$$\mathbb{P}\Big(|X_n - X| \geqslant \varepsilon\Big) \overset{X\equiv 0, X_n>0}{=\!=\!=} \mathbb{P}\Big(X_n \geqslant \varepsilon\Big) \overset{X_n(\omega)=\frac{\omega}{n}}{=\!=\!=} \mathbb{P}\left(\omega \in \Omega : \frac{\omega}{n} \geqslant \varepsilon\right).$$

Now if $n > \frac{6}{\varepsilon}$, then we will have $n > \frac{\omega}{\varepsilon}$ for any $\omega \in \Omega = \{1,2,3,4,5,6\}$. So no ω will satisfy $\frac{\omega}{n} \geqslant \varepsilon$. This means that for $n > \frac{6}{\varepsilon}$,

$$\mathbb{P}\left(\omega \in \Omega : \frac{\omega}{n} \geqslant \varepsilon\right) = 0 \to 0.$$

Accordingly,

$$\mathbb{P}\Big(|X_n - X| \geqslant \varepsilon\Big) = \mathbb{P}\left(\omega \in \Omega : \frac{\omega}{n} \geqslant \varepsilon\right) \to 0.$$

Convergence in Distributions: Again, we can use the property that Convergence in Probability implies Convergence in Distributions. But let us prove the Convergence in Distributions using the definition. To that end, we need to calculate the CDF's of X_n, $F_{X_n}(x)$ and of X, $F_X(x)$, and show that

$$F_{X_n}(x) \to F_X(x) \quad \text{when} \quad n \to \infty \ \text{ at any point of continuity } x \text{ of } F_X(x).$$

First, we have $X \equiv 0$, so

$$F_X(x) = \mathbb{P}(X \leqslant x) = \begin{cases} 1, & x \geqslant 0 \\ 0, & x < 0. \end{cases}$$

Note that the only discontinuity point of F_x is $x = 0$. So we need to check the above (CDF's) convergence at any point, except $x = 0$ (maybe we will have convergence at that point too, but this is not necessary to have Convergence in Distributions: it is enough to check for continuity points of F_X).

Now, about the CDF of X_n. Hope you remember how to construct CDF's for discrete random variables, and you can easily calculate from the PMF of X_n that

$$F_{X_n}(x) = \mathbb{P}(X_n \leqslant x) = \begin{cases} 0, & x < \frac{1}{n} \\ \frac{1}{6}, & \frac{1}{n} \leqslant x < \frac{2}{n} \\ \frac{2}{6}, & \frac{2}{n} \leqslant x < \frac{3}{n} \\ \frac{3}{6}, & \frac{3}{n} \leqslant x < \frac{4}{n} \\ \frac{4}{6}, & \frac{4}{n} \leqslant x < \frac{5}{n} \\ \frac{6}{6}, & \frac{5}{n} \leqslant x < \frac{6}{n} \\ 1, & x \geqslant \frac{6}{n} \end{cases}$$

Now, to prove that $F_{X_n}(x) \to F_X(x)$, for any $x \neq 0$, let us consider cases. Fix $x \neq 0$. If $x < 0$, then $F_{X_n}(x) = F_X(x) = 0$, for any n, so obviously, $F_{X_n}(x) = 0 \to 0 = F_X(x)$. The other case is when $x > 0$, and in this case we can calculate the value of F_X: $F_X(x) = 1$. Since x is fixed, then if $n > \frac{6}{x}$, we will have $x > \frac{6}{n}$, so we will have $F_{X_n}(x) = 1$, if $n > \frac{6}{x}$. Hence, $F_{X_n}(x) \to 1 = F_X(x)$.

Visually, the graphs of F_X and F_{X_n} for some values of n are given in Fig. 8.1-8.4. Hope it is visible for you that F_{X_n} approaches F_X at every point $x \neq 0$.

Joy and Happiness!

Extra Joy and Happiness: For the complete happiness, try to answer the following question: What about the point $x = 0$? Is it true that $F_{X_n}(0) \to F_X(0)$?

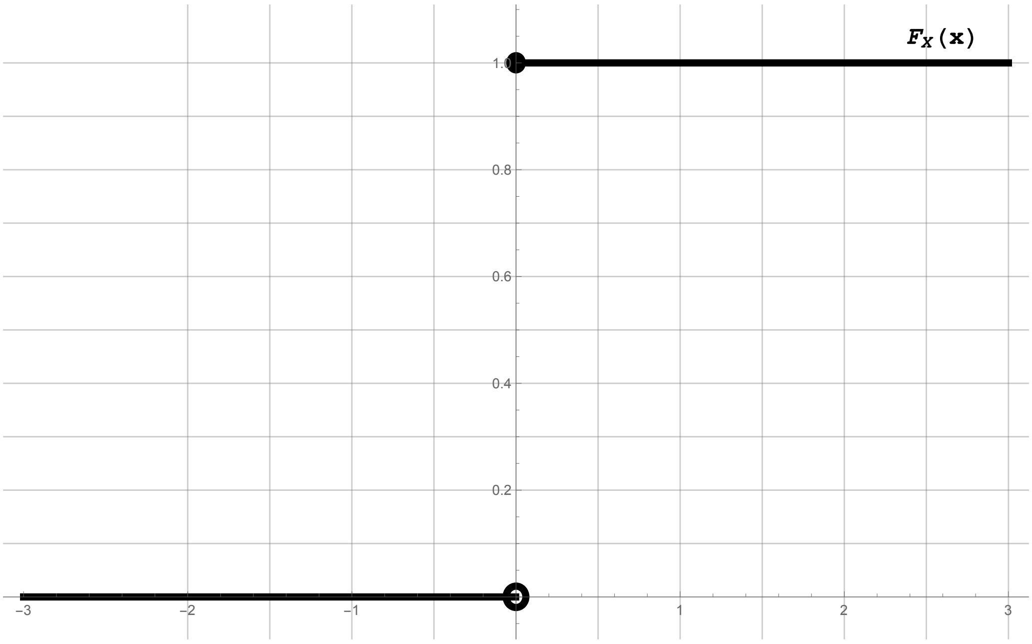

Fig. 8.1: Graph of the CDF of X, $F_X(x)$

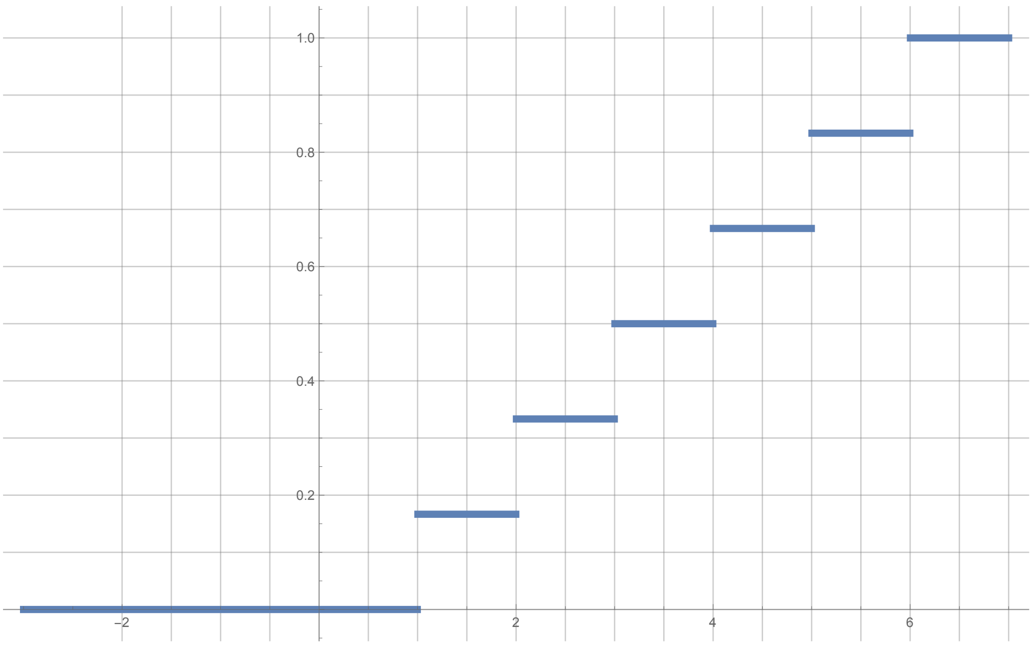

Fig. 8.2: Graph of the CDF of X_1, $F_{X_1}(x)$

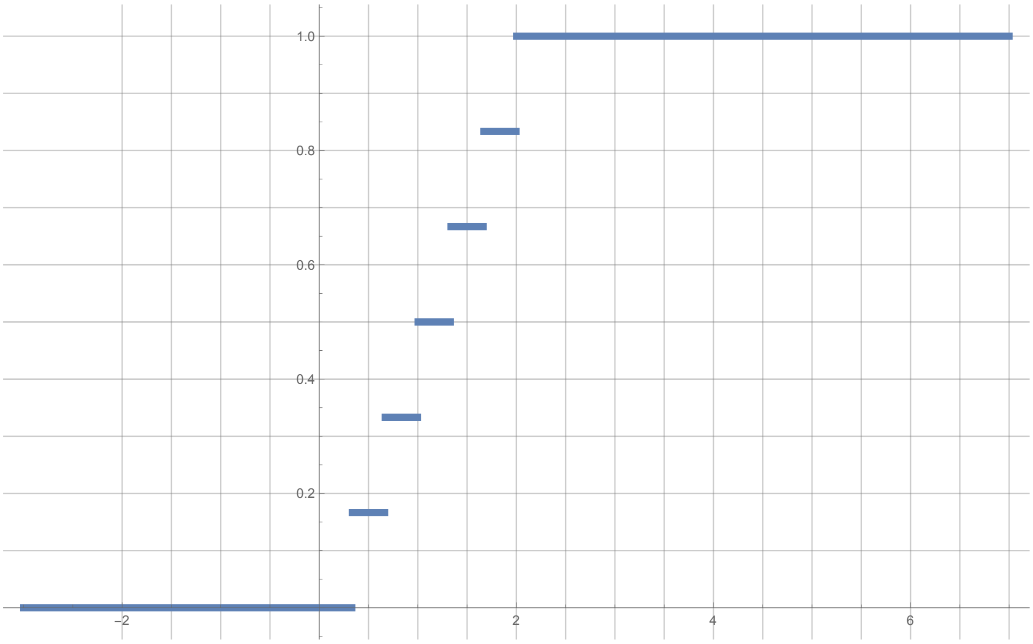

Fig. 8.3: Graph of the CDF of X_3, $F_{X_3}(x)$

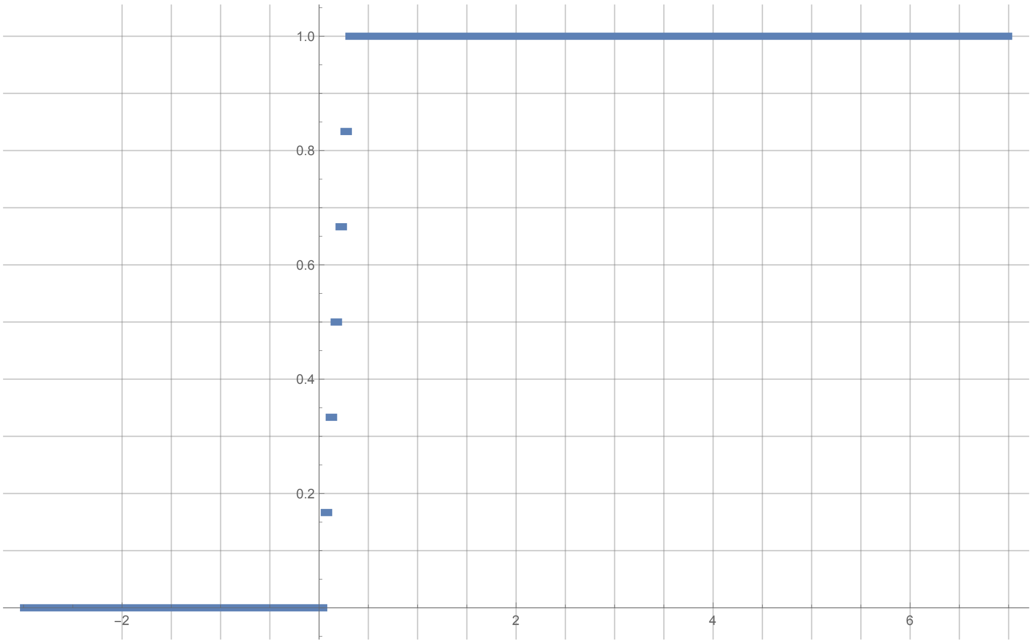

Fig. 8.4: Graph of the CDF of X_{20}, $F_{X_{20}}(x)$

Example, Convergence in different senses: Assume we are again rolling a fair die, $\Omega = \{1,2,3,4,5,6\}$, and assume X_n is given by

$$X_n(\omega) = \omega + \frac{n}{n+1}, \qquad \omega \in \Omega.$$

Then $X_n(\omega) \to X(\omega)$ in all four senses, where $X(\omega) = \omega + 1$.
Surely, you can solve this now!

EXAMPLE, CONVERGENCE IN DIFFERENT SENSES: We have considered the sequence $X_n(\omega) = \omega + \frac{1}{n}$, $\omega \in \Omega = [0, 1]$ (with the length as a probability) during our lecture. Consider also $X_n(\omega) = \frac{n \cdot \omega^2}{n+3} - \frac{2}{n}$.

EXAMPLE, CONVERGENCE IN DIFFERENT SENSES: Assume now our experiment is the following: we are choosing a point at random, uniformly, from[7] $\Omega = [0, 1]$. For any $n \in \mathbb{N}$, we define X_n to be

$$X_n(\omega) = \left(\omega + \frac{1}{n}\right)^2, \qquad \forall \omega \in [0, 1], \quad \forall n \in \mathbb{N}.$$

Now, intuitively, $X_n(\omega) \to X(\omega)$, where $X(\omega) = \omega^2$, $\omega \in \Omega$. Indeed, let us prove that the convergence holds in all senses.

Almost Sure Convergence: This is trivial, Calc 1: for a fixed $\omega \in \Omega$,

$$\lim_{n \to \infty} X_n(\omega) = \lim_{n \to \infty} \left(\omega + \frac{1}{n}\right)^2 = \omega^2 = X(\omega).$$

So even we have *sure* convergence: the convergence holds for *any* ω, $\{X_n \to X\} = \Omega$, so $\mathbb{P}(X_n \to X) = \mathbb{P}(\Omega) = 1$.

Quadratic Mean Convergence: We calculate

$$\mathbb{E}\Big((X_n - X)^2\Big) = \mathbb{E}\left(\left[\left(\omega + \frac{1}{n}\right)^2 - \omega^2\right]^2\right) = \mathbb{E}\left(\left[\frac{2\omega}{n} + \frac{1}{n^2}\right]^2\right)$$

Here we can calculate the expectation in several ways, see later calculations in other examples, but I want to do another trick: since $\omega \in [0, 1]$, then

$$\left[\frac{2\omega}{n} + \frac{1}{n^2}\right]^2 \leqslant \left[\frac{2}{n} + \frac{1}{n^2}\right]^2 \leqslant \left[\frac{2}{n} + \frac{1}{n}\right]^2 = \frac{9}{n^2}.$$

Then, by the monotonicity of expectation (that is, of $X \leqslant Y$ a.s., then $\mathbb{E}(X) \leqslant \mathbb{E}(Y)$),

$$0 \leqslant \mathbb{E}\Big((X_n - X)^2\Big) = \mathbb{E}\left(\left[\frac{2\omega}{n} + \frac{1}{n^2}\right]^2\right) \leqslant \frac{9}{n^2} \to 0.$$

Then, using the squeezing theorem, we will obtain $\mathbb{E}\Big((X_n - X)^2\Big) \to 0$.

Remark: In this case, we could calculate the expectation using the formula:

$$\mathbb{E}\Big((X_n - X)^2\Big) = \mathbb{E}\left(\left[\frac{2\omega}{n} + \frac{1}{n^2}\right]^2\right) = \int_0^1 \left[\frac{2\omega}{n} + \frac{1}{n^2}\right]^2 d\mathbb{P} = \int_0^1 \left[\frac{2\omega}{n} + \frac{1}{n^2}\right]^2 d\omega,$$

which is easy to calculate. Please find this formula in advanced Probability textbooks! See also another example below.

Convergence in Probability: We calculate, fixing $\varepsilon > 0$,

$$\mathbb{P}(|X_n - X| < \varepsilon) = \mathbb{P}\left(\left|\frac{2\omega}{n} + \frac{1}{n^2}\right| < \varepsilon\right) = \mathbb{P}\left(\omega \in [0,1] : \frac{2\omega}{n} < \varepsilon - \frac{1}{n^2}\right) = \mathbb{P}\left(\omega \in [0,1] : \omega < \frac{n\varepsilon}{2} - \frac{1}{2n}\right).$$

If n is large, then $\frac{n\varepsilon}{2} - \frac{1}{2n} > 1$, so for that large n-s we will have

$$\mathbb{P}(|X_n - X| < \varepsilon) = \mathbb{P}\left(\omega \in [0,1] : \omega < \frac{n\varepsilon}{2} - \frac{1}{2n}\right) = \mathbb{P}([0,1]) = 1 \to 1,$$

as $n \to \infty$.

Convergence in Distributions: Let us calculate the CDF of X_n first:

$$F_{X_n}(x) = \mathbb{P}(X_n \leqslant x) = \mathbb{P}(\{\omega \in [0,1] : X_n(\omega) \leqslant x\}) = \mathbb{P}(\{\omega \in [0,1] : \left(\omega + \frac{1}{n}\right)^2 \leqslant x\}) =$$

$$= \begin{cases} 0, & x < 0 \\ \mathbb{P}(\{\omega \in [0,1] : \omega + \frac{1}{n} \leqslant \sqrt{x}\}), & x \geqslant 0 \end{cases} = \begin{cases} 0, & x < 0 \\ \mathbb{P}(\{\omega \in [0,1] : \omega \leqslant \sqrt{x} - \frac{1}{n}\}), & x \geqslant 0 \end{cases}$$

$$= \begin{cases} 0, & x < 0 \\ 0, & 0 \leqslant \sqrt{x} < \frac{1}{n} \\ \mathbb{P}([0, \sqrt{x} - \frac{1}{n}]), & 0 \leqslant \sqrt{x} - \frac{1}{n} \leqslant 1 \\ 1, & \sqrt{x} - \frac{1}{n} > 1 \end{cases} = \begin{cases} 0, & x < \frac{1}{n^2} \\ \sqrt{x} - \frac{1}{n}, & \frac{1}{n^2} \leqslant x \leqslant (1 + \frac{1}{n})^2 \\ 1, & x > (1 + \frac{1}{n})^2 \end{cases}$$

In a similar fashion, $F_X(x)$ is equal to

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \sqrt{x}, & 0 \leqslant x \leqslant 1 \\ 1, & x > 1 \end{cases}$$

This F_X is continuous everywhere, so we need to check that $F_{X_n}(x) \to F_X(x)$ for any $x \in \mathbb{R}$. Clearly, this holds for $x \leqslant 0$, since $F_{X_n}(x) = F_X(x) = 0$ in that case. If, say $0 < x < 1$, then starting from some number, we will have $\frac{1}{n^2} \leqslant x \leqslant (1 + \frac{1}{n})^2$, so we will have for large n-s

$$F_{X_n}(x) = \sqrt{x} - \frac{1}{n} \to \sqrt{x}.$$

In a similar way, one can consider the case $x \geqslant 1$.

Example, Convergence in all senses: Consider again explicitly given r.v. sequence example: the Sample Space is $\Omega = [0,1]$, with usual length of the interval as a probability, i.e., $\mathbb{P}([a,b]) = b - a$, and consider the following sequence of r.v.:

$$X_n(\omega) = (1-\omega)^n, \qquad \omega \in \Omega, \quad n \in \mathbb{N}.$$

Almost Sure Convergence: Everybody who knows Calc 1 will readily state that

$$\lim_{n\to\infty} X_n(\omega) = \lim_{n\to\infty} (1-\omega)^n = \begin{cases} 0, & \omega \in (0,1] \\ 1, & \omega = 0. \end{cases}$$

At this point, we could take the right-hand side as the limit of $X_n(\omega)$ and denote by $X(\omega)$, and the proceed to prove that the convergence hold also in all other senses, besides this a.s. (in fact, sure) convergence. But probabilists and real analysts do the following: take $X(\omega) \equiv 0$. This function and the limit of X_n differ only at one point, at $\omega = 0$, so in the probability of one point is zero (recall that the probability of a set is the length of that set). This means that $X_n \to X$ almost surely. So everybody is happy with this new X - it is much simple than the actual limit, and $X_n \to X$ a.s. .

Quadratic Mean Convergence: Let us prove also that $X_n \xrightarrow{L^2} X$. To that end, we need to prove that

$$\mathbb{E}\Big((X_n - X)^2\Big) \to 0, \qquad n \to \infty,$$

so we need to calculate the expectation $\mathbb{E}\Big((X_n - X)^2\Big) = \mathbb{E}\Big((X_n)^2\Big)$. Our tool to calculate this is to find the PDF of X_n and then use it for calculations. So let us find the PDF of X_n.

First, we calculate the CDF of X_n (we need this also for the Convergence in Distributions study):

$$F_{X_n}(x) = \mathbb{P}(X_n \leqslant x) = \mathbb{P}(\omega \in [0,1] : (1-\omega)^n \leqslant x) = \begin{cases} 0, & x < 0 \\ \mathbb{P}(\omega \in [0,1] : 1-\omega \leqslant \sqrt[n]{x}), & x \geqslant 0 \end{cases} =$$

$$= \begin{cases} 0, & x < 0 \\ \mathbb{P}(\omega \in [0,1] : \omega \geqslant 1 - \sqrt[n]{x}), & x \geqslant 0 \end{cases} = \begin{cases} 0, & x < 0 \\ \mathbb{P}([1-\sqrt[n]{x}, 1]), & 0 \leqslant x \leqslant 1 \\ \mathbb{P}([0,1]), & x > 1 \end{cases} = \begin{cases} 0, & x < 0 \\ \sqrt[n]{x}, & 0 \leqslant x \leqslant 1 \\ 1, & x > 1 \end{cases} \tag{8.1}$$

Now, the PDF of X_n will be

$$f_{X_n}(x) = F'_{X_n}(x) = \begin{cases} \dfrac{1}{n} x^{\frac{1}{n}-1}, & 0 < x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$\mathbb{E}\Big((X_n - X)^2\Big) = \mathbb{E}\Big((X_n)^2\Big) = \int_{-\infty}^{+\infty} x^2 f_{X_n}(x)dx = \frac{1}{n} \cdot \int_0^1 x^2 \cdot x^{\frac{1}{n}-1}\, dx = \frac{1}{2n+1} \to 0.$$

Ura, dami i gospoda!

Remark, somewhat important: We haven't used the following formula to calculate the Expectation, but this is, actually, **the** definition of the Expectation:

$$\mathbb{E}\Big((X_n)^2\Big) = \mathbb{E}\Big((1-\omega)^{2n}\Big) = \int_0^1 (1-\omega)^{2n} d\mathbb{P} = \int_0^1 (1-\omega)^{2n} d\omega = -\int_0^1 (1-\omega)^{2n} d(1-\omega) = \frac{1}{2n+1}.$$

See some advanced Probability textbooks!

Convergence in Probability: You are correctly guessing that we are going to torture ourselves by proving that $X_n \xrightarrow{\mathbb{P}} X$, without using the fact that the convergence a.s. implies this.

Assume $\varepsilon > 0$ is fixed. We want to calculate

$$\mathbb{P}\Big(|X_n - X| < \varepsilon\Big) = \mathbb{P}\Big(|X_n| < \varepsilon\Big) = \mathbb{P}\Big(\omega \in [0,1] : (1-\omega)^n < \varepsilon\Big) = \mathbb{P}\Big(\omega \in [0,1] : (1-\omega) < \sqrt[n]{\varepsilon}\Big) =$$

$$= \mathbb{P}\Big(\omega \in [0,1] : \omega > 1 - \sqrt[n]{\varepsilon}\Big) = \begin{cases} \mathbb{P}([0,1]), & \varepsilon > 1 \\ \mathbb{P}((1-\sqrt[n]{\varepsilon}, 1]), & 0 < \varepsilon \leqslant 1 \end{cases} = \begin{cases} 1, & \varepsilon > 1 \\ \sqrt[n]{\varepsilon}, & 0 < \varepsilon \leqslant 1 \end{cases}$$

which tends to 1 as $n \to \infty$, for any fixed $\varepsilon > 0$. This means that $\mathbb{P}\Big(|X_n - X| \geqslant \varepsilon\Big) \to 0$, so X_n tends to X in Probability.

Convergence in Distribution: We have calculated above the CDF F_{X_n}, and from our previous examples we know the value of the CDF F_X. We need to prove that $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$, for any $x \neq 0$, for any continuity point of F_X.

From (8.1), this is clear for any $x \notin [0, 1]$. Also, if $x \in (0, 1]$, then

$$F_{X_n}(x) = \sqrt[n]{x} \to 1 = F_X(x), \qquad n \to \infty.$$

See the Fig. 8.5-8.8, and compare with the graph of the limit 8.1.

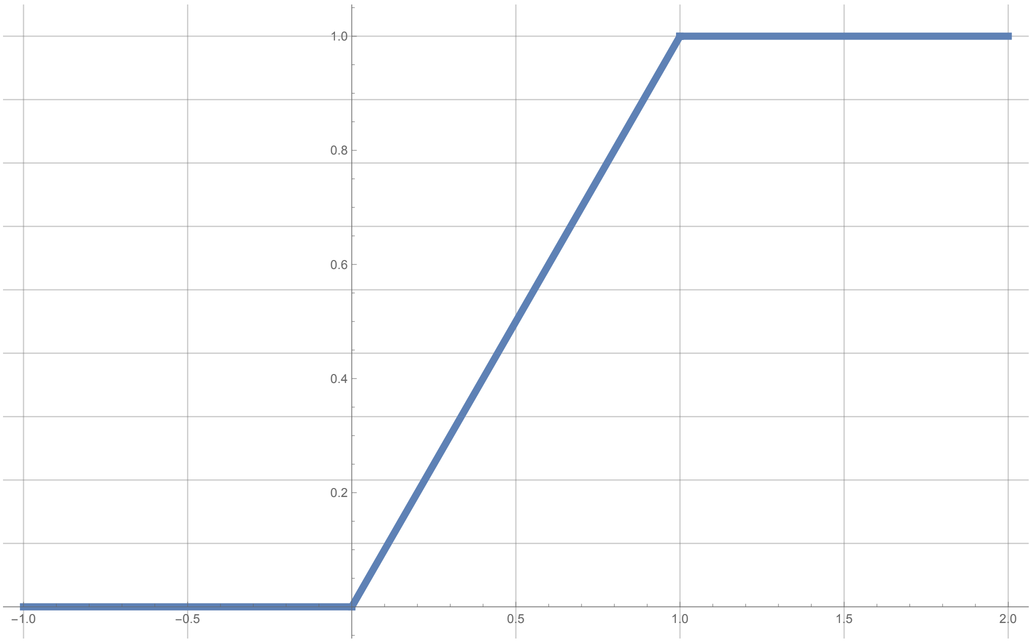

Fig. 8.5: Graph of the CDF of X_1, $F_{X_1}(x)$

EXAMPLE, CONVERGENCE IN DIFFERENT SENSES: Consider again our old friends $\Omega = [0, 1]$, $\mathbb{P}([a, b]) = b - a$ for $[a, b] \subset [0, 1]$, and define

$$X_n(\omega) = \begin{cases} 0, & \omega \in [0,1] \setminus \left\{ \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1 \right\} \\ 1, & \omega \in \left\{ \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1 \right\}. \end{cases}$$

Say, X_3 is equal to 1 at $\frac{1}{3}, \frac{2}{3}$ and $\frac{3}{3} = 1$, and is equal to 0 at any other point of $[0, 1]$. X_4 is equal to 1 at $\frac{1}{4}, \frac{2}{4}, \frac{3}{4}$ and $\frac{4}{4} = 1$, and is equal to 0 at any other point and so on.

Convergence in Almost Sure sense: We can prove that $X_n \to X$ a.s., where $X(\omega) \equiv 0$. Indeed, if $\omega \in [0, 1]$ is irrational, then $X_n(\omega) = 0 \to 0 = X(\omega)$. Now, $X_n \not\to X$ can happen only at rational points, but the length (probability) of the set of all rational points from $[0, 1]$ is 0. So

$$\mathbb{P}(X_n \to X) = 1.$$

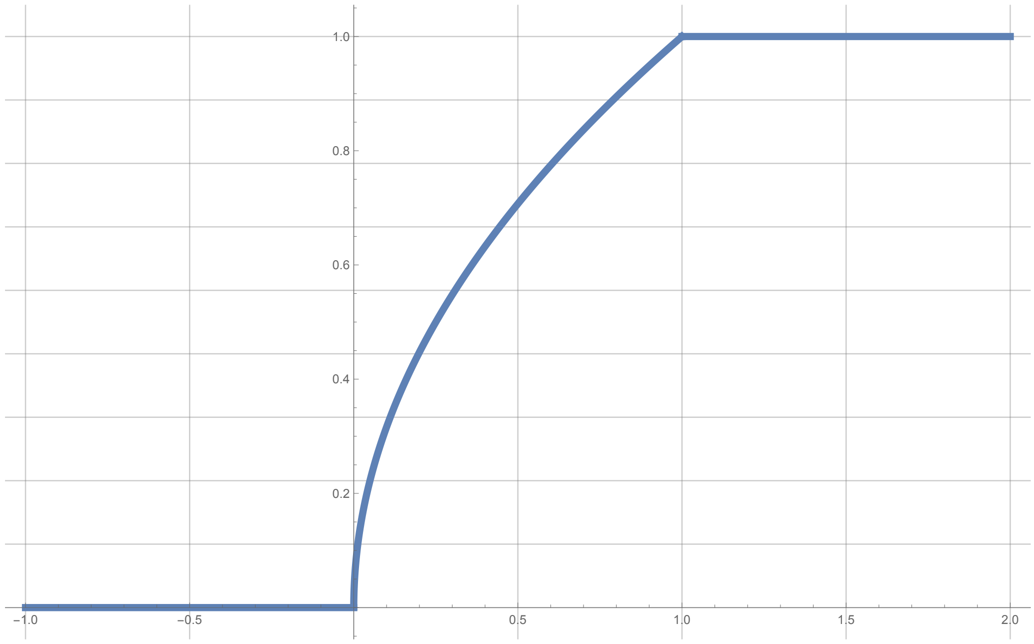

Fig. 8.6: Graph of the CDF of X_2, $F_{X_2}(x)$

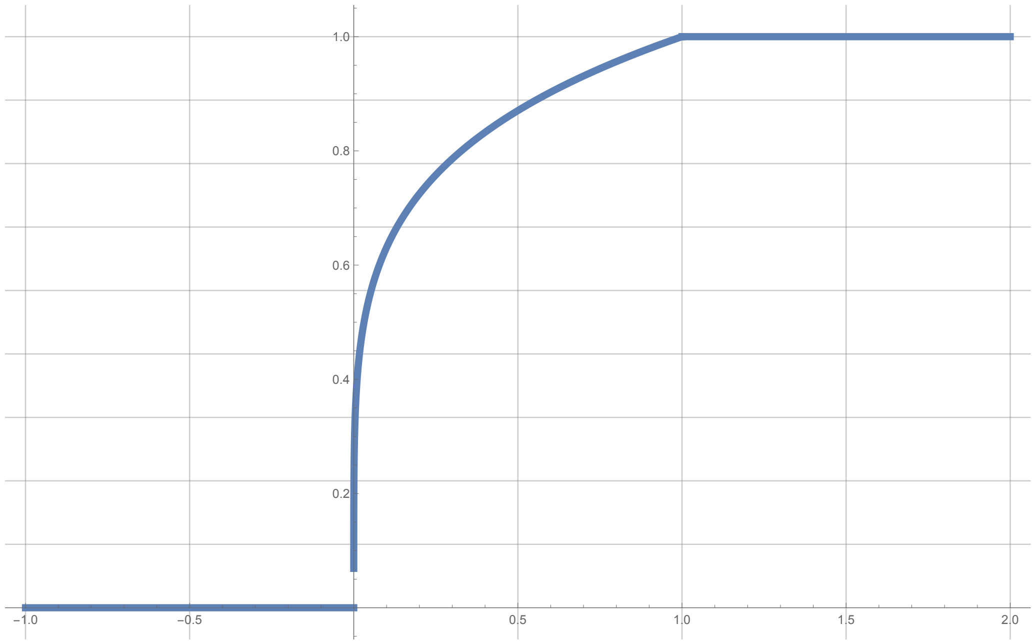

Fig. 8.7: Graph of the CDF of X_5, $F_{X_5}(x)$. In fact, here, and in the next figure, you see a jump at $x = 0$, but, in fact, the function is continuous. This is a drawback of computer software - it cannot give very high precision in this example to draw close to 0. Moral: do not trust computers much! Do not trust man either. Trust only good old friends. Say, Math ◡̈

Convergence in Probability: We fix $\varepsilon > 0$, and calculate

$$\mathbb{P}(|X_n - X| \geqslant \varepsilon) = \mathbb{P}(X_n \geqslant \varepsilon) = \begin{cases} 0, & \varepsilon > 1 \\ \mathbb{P}(X_n = 1), & \varepsilon \leqslant 1 \end{cases} = 0$$

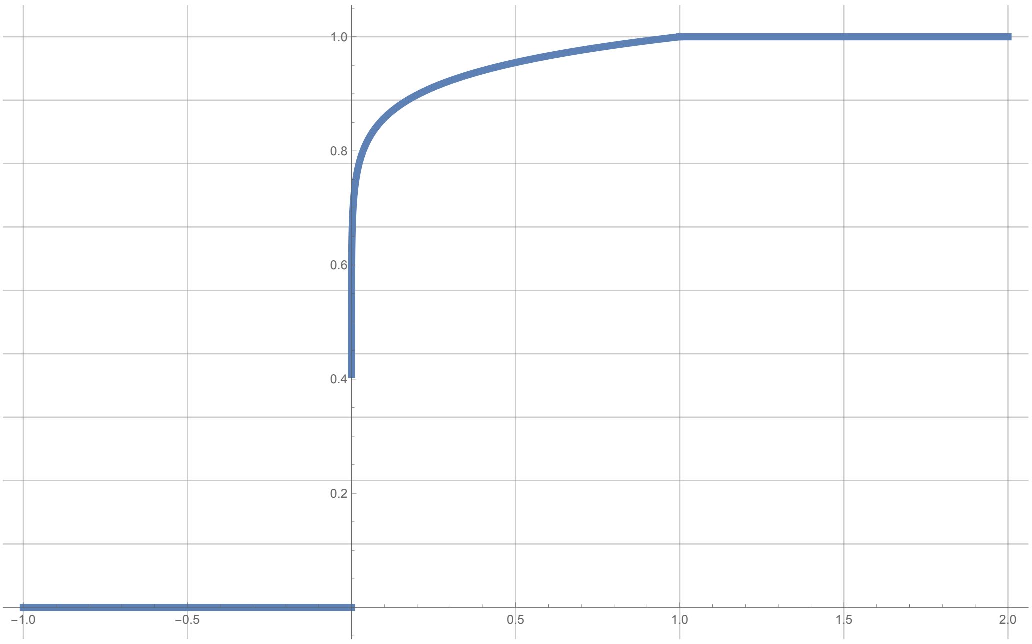

Fig. 8.8: Graph of the CDF of X_{15}, $F_{X_{15}}(x)$

Convergence in Quadratic Mean: We calculate $\mathbb{E}\Big((X_n - X)^2\Big) = \mathbb{E}\Big(X_n^2\Big)$ by considering that X_n is a discrete r.v. with value 0 with probability 1: $\mathbb{P}(X_n = 0) = 1$. So by the discrete r.v. expectation formula, $\mathbb{E}\Big(X_n^2\Big) = 0^2 \cdot \mathbb{P}(X_n = 0) = 0$.

We could do this by calculating first the CDF of X_n, F_{X_n}:

$$F_{X_n}(x) = \mathbb{P}(X_n \leqslant x) = \begin{cases} 0, & x < 0 \\ \mathbb{P}(X_n = 0), & 0 \leqslant x < 1 \\ 1, & x > 1 \end{cases} = \begin{cases} 0, & x < 0 \\ 1, & x \geqslant 0 \end{cases}$$

This means that X_n is a discrete r.v. with the value 0 with probability 1: $\mathbb{P}(X_n = 0) = 1$. The rest is as above.

Convergence in Distributions: This is simple, since, as we have calculated above,

$$F_{X_n}(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geqslant 0 \end{cases} = F_X(x) \qquad x \in \mathbb{R}.$$

Remark, Tactics to prove the convergence of a r.v. sequence, explicit case: In the case when we have our sequence $X_n(\omega)$ given explicitly, as a function of ω, then we can try to prove first the a.s. convergence. To that end, we consider $X_n(\omega)$ for a fixed ω and try to calculate its limit. This is the same as in the Real Analysis course we were calculating functional sequence pointwise limit $\lim f_n(x)$ or in the Calculus class we were calculating limits with parameters, say, the limit of x^n, when $n \to \infty$. If you obtain that $X_n \to X$ a.s. for some X, then the convergence hold also in the Probability and Distribution senses.

8.2.2 R.V. described through their Distribution

EXAMPLE, CONVERGENCE OF A SEQUENCE GIVEN BY THEIR DISTRIBUTIONS: First, very simple (and almost not random) example: assume $X_n = \frac{1}{n}$ with probability 1. Then $X_n \to 0$ in Distribution. If all X_n-s are defined on the same Probability Space, then also $X_n \to 0$ in Quadratic Mean and in Probability.

We can prove also the a.s. convergence. If we will denote by A_n the set, where $X_n \neq \frac{1}{n}$. By the condition above, $\mathbb{P}(A_n) = 0$. Denote now $A = \cup_{n=1}^{\infty} A_n$. Then $\mathbb{P}(A) = 0$ too. And in the complement of A_n, we will have that $X_n(\omega) = \frac{1}{n}$. Now, for any x of this type, we will have that $X_n(\omega) \to X(\omega)$ a.s.

EXAMPLE, CONVERGENCE OF A SEQUENCE GIVEN BY THEIR DISTRIBUTIONS: We consider the following experiment: we are tossing a coin infinitely many times (countably infinite, of course). Let X_n be 1, if n-th toss resulted in heads, and let X_n be 0 otherwise[8].

Now, is X_n a.s. convergent? Let us find the set of all points where X_n converges. If for some scenario X_n converges, then in that scenario either starting from some point on we will have heads shown, or we will have only tails shown[9]. And it can be shown that the set of all scenarios when X_n converges, is a null set (set of probability 0), so our X_n will not converge a.s..

Also, it will not converge in Probability (prove this!). But since all $X_n \sim \mathrm{Bernoulli}(0.5)$, then X_n will tend in Distribution to a r.v. X with $X \sim \mathrm{Bernoulli}(0.5)$. ■

EXAMPLE, CONVERGENCE IN DISTRIBUTION: Let $n \in \mathbb{N}$. Assume X_n is a discrete r.v. with

$$X_n \sim \mathrm{DiscreteUnif}\left(n; \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \dots, 1\right),$$

i.e.,

Values of X_n	$\frac{1}{n}$	$\frac{2}{n}$	$\frac{3}{n}$	...	1
PMF of X_n	$\frac{1}{n}$	$\frac{1}{n}$	$\frac{1}{n}$	...	$\frac{1}{n}$.

In fact, here we are not specifying the underlying probability space, so the r.v.'s X_n can be defined on different probability spaces.

Intuitively, the sequence X_n will tend in Distributions to the Standard Uniform Distributed r.v.. if $X \sim \mathrm{Unif}([0,1])$, then it is easy to prove that $X_n \xrightarrow{D} X$. Say, you can run the R code given below to see the picture.

In this example, we cannot talk about the a.s. or in Probability convergence of X_n, since they can be defined on different probability spaces.

R CODE, PREVIOUS EXAMPLE CODE: Here we want to draw the above distribution CDFs.

```
par(mfcol = c(1,2))

n <- 20
x <-(1:n)/n
```

```
f <- ecdf(x)
plot(f, xlim = c(-0.5,1.5), lwd = 2, main = "n = 20", ylab = "CDF")
#Now the Uniform
par(new = T)
plot(punif, xlim = c(-0.5,1.5), col = "red", lwd = 3, main = "n = 20",  ylab = "CDF")
par(new = F)

n <- 50
x <-(1:n)/n
f <- ecdf(x)
plot(f, xlim = c(-0.5,1.5), lwd = 2, main = "n = 50", ylab = "CDF")
#Now the Uniform
par(new = T)
plot(punif, xlim = c(-0.5,1.5), col = "red", lwd = 3, main = "n = 50",  ylab = "CDF")
par(new = F)
```

Example, Convergence in Probability and Distributions: Assume X_n is a Discrete r.v. with the following PMF:

X_n	$5 - \frac{1}{n^2}$	n
$\mathbb{P}(X_n = x)$	$1 - \frac{1}{n}$	$\frac{1}{n}$.

Then $X_n \xrightarrow{\mathbb{P}} X$ and $X_n \xrightarrow{D} X$, where $X \equiv 5$ (of course, the last statement follows from the former one, but let us prove independently).

Remark: When dealing with the Convergence in Probability, we need to assume that all X_n-s are defined on the same Probability Space. In fact, since X, the limiting r.v., is constant, then we can calculate $\mathbb{P}(|X_n - X| \geqslant \varepsilon)$ even if X_n-s are defined on different Probability Spaces.

Convergence in Probability: We fix $\varepsilon > 0$ and calculate the probability $\mathbb{P}(|X_n - X| \geqslant \varepsilon)$. To that end, we can see that the r.v. $|X_n - X|$ will take the values $\frac{1}{n^2}$ and $|n-5|$ with probabilities, $1 - \frac{1}{n}$ and $\frac{1}{n}$, respectively, i.e. $|X_n - X|$ has the distribution

$\vert X_n - X\vert$	$\frac{1}{n^2}$	$\vert n-5\vert$
$\mathbb{P}(\vert X_n - X\vert = x)$	$1 - \frac{1}{n}$	$\frac{1}{n}$.

We want to see when we can have $|X_n - X| \geqslant \varepsilon$. If $n > \frac{1}{\sqrt{\varepsilon}}$, then we will have $\frac{1}{n^2} < \varepsilon$, so $|X_n - X| \geqslant \varepsilon$ can happen only if $|X_n - X| = |n-5|$, and $|n-5| \geqslant \varepsilon$, and the probability of that is $\frac{1}{n}$. This means that

$$\mathbb{P}(|X_n - X| \geqslant \varepsilon) \overset{n > \frac{1}{\sqrt{\varepsilon}}, |n-5| \geqslant \varepsilon}{=\!=\!=} \mathbb{P}(|X_n - X| = |n-5|) = \frac{1}{n} \to 0.$$

Convergence in Distribution: The CDF of X is

$$F_X(x) = \begin{cases} 0, & x < 5 \\ 1, & x \geqslant 5 \end{cases}$$

The only discontinuity point of F_X is $x = 5$. The CDF of X_n is, for $n \geqslant 5$,

$$F_{X_n}(x) = \begin{cases} 0, & x < 5 - \frac{1}{n^2} \\ 1 - \frac{1}{n}, & 5 - \frac{1}{n^2} \leqslant x < n \\ 1, & x \geqslant n. \end{cases}$$

We need to prove that $F_{X_n}(x) \to F_X(x)$ for any $x \neq 5$. Say, $x > 5$. Then, starting from some number on, we will have $5 - \frac{1}{n^2} \leqslant x < n$, i.e.

$$F_{X_n}(x) = 1 - \frac{1}{n} \to 1 = F_X(x).$$

If $x < 5$, then starting from some number on, we will have $x < 5 - \frac{1}{n^2}$, and then

$$F_{X_n}(x) = 0 \to 0 = F_X(x).$$

Try to draw the graphs to see the convergence visually!

Remark: Note that in this case, we will not have a convergence in Quadratic Mean. This is because the distribution of $(X_n - X)^2$ will be

$(X_n - X)^2$	$\frac{1}{n^4}$	$(n-5)^2$
$\mathbb{P}((X_n - X)^2 = x)$	$1 - \frac{1}{n}$	$\frac{1}{n}$.

and

$$\mathbb{E}((X_n - X)^2) = \left(1 - \frac{1}{n}\right) \cdot \frac{1}{n^4} + \frac{1}{n} \cdot (n-5)^2 \to +\infty.$$

Example, Convergence in Probability and Distributions: Assume $X_n \sim \mathcal{N}\left(0, \frac{1}{n}\right)$. Then $X_n \xrightarrow{D} 0$. ' In general, we cannot talk about the convergence in Probability or a.s. or L^2 means in this case, since X_n-s can be r.v.s defined in very different sample spaces. If we will assume that they are defined on the same Probability Space, then one can prove that $X_n \xrightarrow{P} 0$, and also $X_n \xrightarrow{q.m.} 0$.

Example, Convergence in Probability and Distributions: Assume $X_n \sim Unif\left[0, \frac{1}{n}\right]$. Then $X_n \xrightarrow{D} 0$. If all X_n are defined on the same Ω, say, $\Omega = [0, 1]$, then also $X_n \xrightarrow{P} 0$ and $X_n \xrightarrow{q.m.} 0$.

Example, Poisson as a limit of Binomial: Assume $X_n \sim Binomial(n, \frac{\lambda}{n})$ and $X \sim Pois(\lambda)$. Then $X_n \xrightarrow{D} X$.

REMARK, CONVERGENCE IN PROBABILITY: As we have noted above, except the case of the convergence in Distributions, for all other three types of convergence we need to have that all X_k-s and X are defined on the same Probability Space. This is to ensure that we can calculate, say $\mathbb{E}((X_n - X)^2)$ or $\mathbb{P}(|X_n - X| \geqslant \varepsilon)$. But, in fact, there is one particular case, when we can talk about the convergence of a sequence of r.v. X_k to X in the L^2 or in the Probability sense even when X_k and X are not defined on the same probability space. This case is when $X \equiv const$. In that case, for any $n \in \mathbb{N}$, we can calculate

$$\mathbb{E}((X_n - X)^2) \qquad \text{and} \qquad \mathbb{P}(|X_n - X| \geqslant \varepsilon)$$

and see if these numerical sequences tend to 0 or not.

8.3 The LLN and CLT

Assume now X_n is a sequence of IID r.v. . This means that all X_k-s have the same distribution. In particular, this means that if one of X_k-s has a finite expectation and variance, then all X_k-s have finite expectations and variances and

$$\mathbb{E}(X_i) = \mathbb{E}(X_j), \qquad \text{and} \qquad Var(X_i) = Var(X_j), \qquad \forall i, j \in \mathbb{N}.$$

The Idea of LLN and CLT In Probability Theory and Statistics, one of the important questions is to study the distribution and/or the behavior of either the sum

$$S_n = X_1 + X_2 + ... + X_n$$

or the average

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}.$$

In fact, these problems are not easy ones. We know from the Probability Theory that the PDF of $X + Y$ in case when X and Y are independent, is the convolution of the PDFs of X and Y. And calculation of the convolution is not an easy task. And, for S_n or X_n, we need to calculate $n - 1$ convolutions!

Let us see what we can say about S_n and $\overline{X}_n$, in general. Since X_k-s are IID, they have the same mean and variance, and let

$$\mathbb{E}(X_k) = \mu \qquad \text{and} \qquad Var(X_k) = \sigma^2.$$

Proposition 8.5. *If X_k are IID with the above expectation and variance, then*[10]

$$\mathbb{E}(S_n) = n \cdot \mu, \qquad \mathbb{E}(\overline{X}_n) = \mu;$$

$$Var(S_n) = n \cdot \sigma^2, \qquad Var(\overline{X}_n) = \frac{\sigma^2}{n}.$$

Proof. Obvious, we will meet this many times! □

Say, what information is giving this proposition about $\overline{X}_n$? If n is large, then the possible values of $\overline{X}_n$ are around μ, since $\mathbb{E}(\overline{X}_n) = \mu$, and are very concentrated around μ, since $Var(\overline{X}_n) = \frac{\sigma^2}{n}$ is small (n is large).

But, of course, having only the expectation and variance of S_n and X_n, is giving us just a very partial information about their distributions. Sometimes, for some particular cases, we can exactly find the distributions of S_n and/or $\overline{X}_n$.

[10]We can read the assertion $\mathbb{E}(X_n) = \mu$ as *the mean of the means is the mean* ◡̈

Proposition 8.6. *a. If* $X_k \sim \mathcal{N}(\mu, \sigma^2)$, $k = 1, ..., n$, *are independent, then*

$$S_n = X_1 + ... + X_n \sim \mathcal{N}(n \cdot \mu, n \cdot \sigma^2) \qquad \text{and} \qquad \overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

b. If $X_k \sim \mathrm{Bernoulli}(p)$, $k = 1, ..., n$ *are independent, then*

$$S_n = X_1 + ... + X_n \sim \mathrm{Binom}(n, p);$$

c. If $X_k \sim \mathrm{Binom}(m, p)$, $k = 1, ..., n$, *are independent, then*[11]

$$S_n = X_1 + ... + X_n \sim \mathrm{Binom}(n \cdot m, p).$$

d. If $X_k \sim \mathrm{Geom}(p)$, $k = 1, ..., n$, *are independent, then*[12]

$$S_n = X_1 + ... + X_n \sim \mathrm{NegBin}(n, p)$$

e. If $X_k \sim \mathrm{Pois}(\lambda)$, $k = 1, ..., n$, *are independent, then*

$$S_n = X_1 + ... + X_n \sim \mathrm{Pois}(n \cdot \lambda).$$

f. If $X_k \sim \mathrm{Gamma}(\alpha, \beta)$, $k = 1, ..., n$, *are independent, then*

$$S_n = X_1 + ... + X_n \sim \mathrm{Gamma}(n \cdot \alpha, \beta).$$

In particular, if $X_k \sim \mathrm{Exp}(\lambda)$, $k = 1, ..., n$, *are independent, then*

$$S_n = X_1 + ... + X_n \sim \mathrm{Gamma}(n, \lambda).$$

But, unfortunately, the distribution of S_n or $\overline{X}_n$ cannot be calculated in simple terms for many cases. And here the we can use the LLN and/or CLT. The Law of Large Numbers (LLN) is giving the asymptotic behavior of $\overline{X}_n$, and the Central Limit Theorem describes the asymptotic distribution of S_n and $\overline{X}_n$.

Theorem 8.3 (The Strong Law of Large Numbers)**.** *If* $X_1, ..., X_n, ...$ *is a sequence of IID r.v. with* $\mathbb{E}(|X_1|) < +\infty$ *and if* $\mathbb{E}(X_i) = \mu$, *then*

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n} \to \mu \qquad a.s.$$

Theorem 8.4 (The Weak Law of Large Numbers)**.** *If* $X_1, ..., X_n, ...$ *is a sequence of IID r.v. with finite expectation* $\mathbb{E}(X_i) = \mu$ *and variance* $\mathrm{Var}(X_i) = \sigma^2$, *then*

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n} \xrightarrow{P} \mu.$$

[11] The distribution of $\overline{X}_n$ can be described, but is not of standard ones. Just divide the values of S_n by n.

[12] The Negative Binomial Distribution is the number of failures before the n-th success when doing $\mathrm{Bernoulli}(p)$ trials, see https://en.wikipedia.org/wiki/Negative_binomial_distribution.

R CODE, LLN:

```
#LLN
n <- 1000 #number of r.v.'s
expect <- 0.6 #this will be the expectation of each random variable
X <- rbinom(n, 1, expect) #generating n samples from the same distribution
S <- cumsum(X) #calculating the cumulative sum: S = (X_1, X_1+X_2, X_1+X_2+X_3,...)
p <- S/(1:n) #This will produce p=(X_1/1, (X_1+X_2)/2, (X_1+X_2+X_3)/3,...)
plot(p, type = "l")
abline(expect,0, col = "red", lwd = 2) #giving in red the limit
```

The result is given in Fig. 8.9.

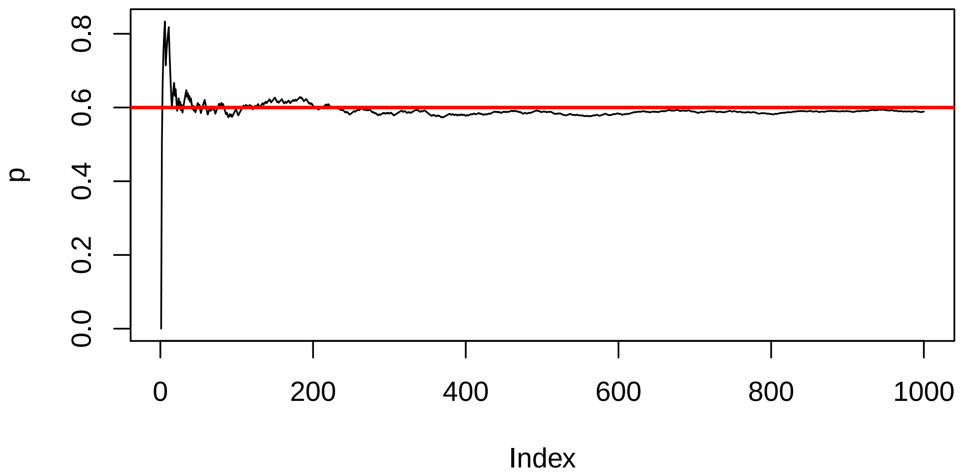

Fig. 8.9: The Law of Large Numbers

REMARK, USING LLN TO CALCULATE PROBABILITIES BY SIMULATIONS: Great and interesting thing is that we can use the LLN to calculate probabilities, areas, volumes, one and multi-dimensional integrals by Simulations. The method is called the Monte-Carlo Method, and is widely used in many applications.

To describe the ideas, assume we want to calculate the probability $\mathbb{P}(a \leqslant X \leqslant b)$ for some continuous r.v. X. Assume we can generate random numbers from the distribution of X. How to use this to calculate the mentioned probability?

We consider the indicator function $\mathbb{1}_{[a,b]}(x)$. If we will consider the r.v. $\mathbb{1}_{[a,b]}(X)$, and calculate the expected value $\mathbb{E}(\mathbb{1}_{[a,b]}(X))$, then we will get

$$\mathbb{E}(\mathbb{1}_{[a,b]}(X)) = \int_{-\infty}^{+\infty} \mathbb{1}_{[a,b]}(x) f_X(x)dx = \int_a^b f(x)dx = \mathbb{P}(a \leqslant X \leqslant b).$$

So we need to calculate the expected value $\mathbb{E}(\mathbb{1}_{[a,b]}(X))$. And we can approximate this by the LLN! Ura! Here is the method: by the LLN, if we will take r.v.s $X_1, X_2, ..., X_n$ that are IID with the same distribution as X, then

$$\frac{\mathbb{1}_{[a,b]}(X_1) + \mathbb{1}_{[a,b]}(X_2) + \ldots + \mathbb{1}_{[a,b]}(X_n)}{n} \to \mathbb{E}(\mathbb{1}_{[a,b]}(X)).$$

, so

$$\mathbb{E}(\mathbb{1}_{[a,b]}(X)) \approx \frac{\mathbb{1}_{[a,b]}(X_1) + \mathbb{1}_{[a,b]}(X_2) + \ldots + \mathbb{1}_{[a,b]}(X_n)}{n}.$$

The numerator in the fraction on the RHS is just the number of X_k-s in $[a,b]$. So basically,

$$\mathbb{P}(a \leqslant X \leqslant b) = \mathbb{E}(\mathbb{1}_{[a,b]}(X)) \approx \frac{\#\text{ of } X_k \text{ in } [a,b]}{n}.$$

And the code will go as follows: we generate $x_1, \ldots, x_n$ from the distribution of X, calculate the number of x_k-s in $[a,b]$, and divide this number to the total number of elements generated, n. This ratio will give an approximation for the probability $\mathbb{P}(a \leqslant X \leqslant b)$. Here is the code:

R code, Using LLN to calculate Probabilities by Simulations: Assume $X \sim \mathcal{N}(0,1)$, and we want to calculate (approximately) the probability

$$\mathbb{P}(0.3 \leqslant X \leqslant 1.4).$$

The actual probability is equal to

$$\mathbb{P}(0.3 \leqslant X \leqslant 1.4) = \frac{1}{\sqrt{2\pi}} \cdot \int_{0.3}^{1.4} e^{-x^2/2} dx,$$

and calculation of the last integral cannot be done easily - we need to use some numerical technique. Here we will use the idea above:

```
#Probabilities by Simulations
n <- 5000
a <- 0.3
b <- 1.4
x <- rnorm(n)
inside <- sum((x<=b) & (x>=a))
prob <- inside/n
real.prob <- pnorm(b)-pnorm(a)
```

I love Math! Well, of course, there is some hidden thing here - we need to be able to generate normal random variables. But since **R** is doing this for us, we can happily use this opportunity to calculate some complicated integrals $\ddot\smile$

Remark, Using LLN to calculate Integrals: As we have mentioned above, we can use the LLN and Monte Carlo Simulations to calculate integrals. Let me explain ideas for a 1D integral calculation. Assume we have a positive continuous function $g : [a,b] \to \mathbb{R}$, and we want to calculate approximately

$$\int_a^b g(x) dx.$$

Nothing probabilistic yet. But let us make this probabilistic, let us write this integral as an expected value for some r.v.. To that end, consider a r.v. $X \sim \mathrm{Unif}[a,b]$. Then the PDF of X will be $f_X(x) = \frac{1}{b-a}$ for $x \in [a,b]$ and $f_X(x) = 0$, for $x \notin [a,b]$. Now, consider the r.v. $g(X)$. We know that

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx = \frac{1}{b-a} \cdot \int_a^b g(x) dx,$$

so

$$\int_a^b g(x) dx = (b-a) \cdot \mathbb{E}(g(X)).$$

Hopplya!! Magic! Non-probabilistic, our old and good friend Calculus integral is an expected value of some random variable! Now, to calculate the expectation, we use the LLN: we take IID r.v. $X_1, X_2, ..., X_n \sim \mathrm{Unif}[a,b]$, and then

$$\frac{g(X_1) + g(X_2) + ... + g(X_n)}{n} \to \mathbb{E}(g(X)),$$

so

$$\int_a^b g(x) dx = (b-a) \cdot \mathbb{E}(g(X)) \approx (b-a) \cdot \frac{g(X_1) + g(X_2) + ... + g(X_n)}{n}.$$

Well, maybe one is not using this method for 1D or 2D integrals, because in this cases we have well-developed theory for numerical calculation of integrals, and many nice methods, with proven error estimation and analysis. But for multidimensional integrals, this simulations method will give some good results compared to other numerical methods.

The above method is implemented below.

R code, LLN for Integral Calculation:

```
#Integral by Simulations
#Say, we want to integrate the function x^2*sin(x^4) over [0,3]
g <- function(x){
  return(x^2*sin(x^4))
}
a <- 0
b <- 3
n <- 10000
x <-runif(n, min = a, max = b)
approx_int <- (b-a)*mean(g(x))
int <- integrate(g,a,b) #R-s native integral calculator
abs(int$value - approx_int) #absolute difference between R-s native integral calculator's value
```

Remark, LLN for Integral calculation, 2D version: Explain here how to calculate

$$\int_a^b g(x) dx$$

by using 2D ideas - calculation of the area under the curve $y = g(x)$ in $x \in [a,b]$.

REMARK, LLN, AGAIN: Assume X_k's are IID with the mean $\mathbb{E}(X_k) = \mu$ and $Var(X_k) = \sigma^2$. The LLN states that

$$\frac{X_1 + X_2 + ... + X_n}{n} \to \mu$$

(in Probability or in the a.s. sense), and an easy fact from the Proposition 8.5 says that

$$\mathbb{E}\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \mu$$

Do you see what is the difference between the LLN (in some form) and that very easy fact?

The latter one means that the means $\overline{X}_n$ are equal to the mean μ in the mean ⌣̈. So it says that the values of $\overline{X}_n$ are around μ, and if we will generate a lot of values for $\overline{X}_n$, the average of that values will be close to μ. Here n is fixed, the number of generated values of $\overline{X}_n$ is large.

The former says that **almost all** values of $\overline{X}_n$ are close to μ, if n is large enough.

This theorems state that random variables $\overline{X}_n$ become more and more concentrated around μ, around their mean. To give more accurate information, to give the asymptotic distribution of $\overline{X}_n$, we use the CLT:

Theorem 8.5 (The Central Limit Theorem). *Let X_n be a sequence of iid r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$. Denote*

$$\overline{X}_n = \frac{X_1 + X_2 + ... + X_n}{n}, \qquad n \in \mathbb{N}.$$

Then if we will standardize (normalize)[13] *$\overline{X}_n$, i.e. if we will denote*

$$Z_n = \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}},$$

then

$$Z_n \xrightarrow{D} Z$$

for some $Z \sim \mathcal{N}(0, 1)$.

REMARK, CLT: The conclusion of the CLT is that

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} Z$$

with $Z \sim \mathcal{N}(0, 1)$, which means that for any $a < b$,

$$\mathbb{P}(a \leqslant \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leqslant b) \to \mathbb{P}(a \leqslant Z \leqslant b) = \frac{1}{\sqrt{2\pi}} \cdot \int_a^b e^{-x^2/2} dx,$$

[13]If X is a r.v., then by its standardization we mean creating another r.v. by shifting and scaling X, which will have an expectation 0 and variance 1. If we will denote

$$Y = \frac{X - \mathbb{E}(X)}{SD(X)} = \frac{X - \mathbb{E}(X)}{\sqrt{Var(X)}},$$

then

$$\mathbb{E}(Y) = 0 \qquad \text{and} \qquad Var(Y) = 1.$$

so for large n,

$$\mathbb{P}\left(a \leqslant \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leqslant b\right) \approx \frac{1}{\sqrt{2\pi}} \cdot \int_a^b e^{-x^2/2} dx,$$

In the case when $Z_n \xrightarrow{D} Z$ for some $Z \sim \mathcal{N}(0,1)$, one uses the notation

$$Z_n \xrightarrow{D} \mathcal{N}(0,1).$$

So the CLT states that

$$\frac{\sqrt{n} \cdot (\overline{X}_n - \mu)}{\sigma} \xrightarrow{D} \mathcal{N}(0,1),$$

so

$$\sqrt{n} \cdot (\overline{X}_n - \mu) \xrightarrow{D} \sigma\mathcal{N}(0,1),$$

or

$$\sqrt{n} \cdot (\overline{X}_n - \mu) \xrightarrow{D} \mathcal{N}(0,\sigma^2).$$

If we will use a little bit not rigorous math terms, then we can write this as

$$\sqrt{n} \cdot (\overline{X}_n - \mu) \approx \mathcal{N}(0,\sigma^2),$$

which further can be written in the form[14]

$$\overline{X}_n - \mu \approx \frac{1}{\sqrt{n}} \cdot \mathcal{N}(0,\sigma^2) = \mathcal{N}\left(0, \frac{\sigma^2}{n}\right),$$

and then

$$\overline{X}_n \approx \mu + \mathcal{N}\left(0, \frac{\sigma^2}{n}\right) = \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

So, CLT states that if $\{X_n\}$ is a sequence of IID r.v., then the asymptotic distribution of $\overline{X}_n$ is

$$\overline{X}_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \qquad \text{for large } n,$$

that is, for large n, the approximate distribution of $\overline{X}_n$ is Normal with parameters μ and $\frac{\sigma^2}{n}$, and this is *independent of the actual distribution of* X_k*-s!*

Remark, CLT, in other form: Sometimes people use the CLT not for the sequence $\overline{X}_n$, but for

$$S_n = X_1 + X_2 + ... + X_n.$$

The process is just as above: we first standardize S_n: since $\mathbb{E}(S_n) = n \cdot \mathbb{E}(X_1) = n \cdot \mu$, and $Var(S_n) = n \cdot Var(X_1) = n\sigma^2$, then the r.v. Z_n defined by

$$Z_n = \frac{S_n - n\mu}{\sqrt{n} \cdot \sigma}$$

will have[15] an expectation 0 and variance 1:

$$\mathbb{E}(Z_n) = 0 \qquad \text{and} \qquad Var(Z_n) = 1.$$

[14]We write $k \cdot \mathcal{N}(0,\sigma^2) = \mathcal{N}(0, k^2\sigma^2)$ in the sense that if $X \sim \mathcal{N}(0,\sigma^2)$, then $k \cdot X \sim \mathcal{N}(0, k^2 \cdot \sigma^2)$

Then the CLT will then give:

$$Z_n \xrightarrow{D} \mathcal{N}(0,1).$$

If we will use some not rigorous notations like above, we can write

$$S_n = n \cdot \overline{X}_n \approx n \cdot \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) = \mathcal{N}\left(n \cdot \mu, n \cdot \sigma^2\right) \qquad \text{for large } n,$$

that is,

$$S_n \approx \mathcal{N}\left(n \cdot \mu, n \cdot \sigma^2\right) \qquad \text{for large } n.$$

This means that the asymptotic distribution of the sum S_n is Normal with the mean $n \cdot \mu$ and variance $n \cdot \sigma^2$.

R code, CLT:

```
#The Central Limit Theorem
no_of_simulations <- 5000
sample_size <- 5000
m <- c() # in m we will keep the values of normalized means

#Say, we will generate from Binom(10,0.4)
mu <- 10*0.4 #the expected value for a Binom(10,0.4) r.v
sig <- sqrt(10*0.4*(1-0.4)) #the standard deviation for a Binom(10,0.4) r.v

#Uncomment, if you want to generate from Unif[4,10]
#mu <- (4+10)/2 #The Expexted Values for Unif[4,10] r.v
#sig <- sqrt((10-4)^2/12) #the standard deviation for a Unif[4,10] r.v

for(i in 1:no_of_simulations){
  x <- rbinom(sample_size, size = 10, prob = 0.4) # we generate a sample from the Binom(10, 0.4
  #x <- runif(sample_size, min = 4, max = 10) # we generate a sample from the Unif[4,10]
  m[i] <- (sum(x) - sample_size* mu)/(sqrt(sample_size)*sig)
}
#plotting the results
hist(m, breaks = seq(min(m)-0.2, max(m)+0.2, by = 0.2), col = "lightcyan", freq = F, xlim = c(-
par(new = T)
curve(dnorm, xlim = c(-3,3), ylim = c(0,0.45), col = "red", lwd = 2)
```

R code, CLT, Another Interpretation: Here we give another method to demonstrate the CLT, and also to calculate probabilities by simulations. The idea is the following: according to CLT, for any a, b we need to have

$$\mathbb{P}(a \leqslant \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leqslant b) \approx \frac{1}{\sqrt{2\pi}} \cdot \int_a^b e^{-x^2/2} dx,$$

for a large n. We want to check this using **R**.

We will calculate the probability on the LHS using the following idea: we will take an observation $x_1, \dots, x_n$, calculate the value of $\overline{X}_n$, and then $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$. And we will repeat this process and obtain different values for $\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}$. And to calculate the probability that this quantity is between a and b, we will calculate how many of our generated values are in $[a,b]$, and divide that number to the total number of experiments (simulations). This ratio will be close to the probability above, if the number of simulations is large enough. Here is the code:

```
## Second method to check CLT Result, with probabilities

no_of_simulations <- 1000
sample_size <- 1000
standardized_x <- c()
#we will use gamma-distributed r.v.s
g_shape = 2 #parameters of gamma-distrib
g_scale = 1 #parameters of gamma-distrib
mu <- g_shape*g_scale # the expectation of gamma distrib
sig <- sqrt(g_shape*g_scale^2) # the standard deviation of the gamma distrib

for (i in 1:no_of_simulations){
  x <- rgamma(sample_size, shape = g_shape, scale = g_scale)
  #we will use this time the averages form of the CLT
  standardized_x[i] <- sqrt(sample_size)*(mean(x)-mu)/sig
}

#We want to calculate the probability that the standardized_x is in [a,b]
a <- -0.4
b <- 0.6
N <- sum((standardized_x<=b) & (standardized_x>=a))
prob <- N/no_of_simulations #Probability by Simulations
prob_with_normal <- pnorm(b)-pnorm(a) #Probability calculated by the Standard Normal CDF
prob
prob_with_normal
```

Remark, Prop, LLN and CLT: I am in love with the LLN, CLT and beautiful things like that, that's why I want to give the general idea again. This time, for $\overline{X}_n$ only. So what info we have obtained from Proposition 8.5, LLN and CLT about $\overline{X}_n$? They are gradually giving us the following info:

Proposition 8.5:

$$\mathbb{E}\left(\frac{X_1+X_2+...+X_n}{n}\right)=\mu \qquad \text{and} \qquad Var\left(\frac{X_1+X_2+...+X_n}{n}\right)=\frac{\sigma^2}{n};$$

LLN:

$$\frac{X_1+X_2+...+X_n}{n}\to\mu$$

in the Probability or in the a.s. sense;

CLT:

$$\frac{X_1+X_2+...+X_n}{n}\approx\mathcal{N}\left(\mu,\frac{\sigma^2}{n}\right) \qquad \text{for large } n$$

Remark, CLT, once again: Let me give the CLT for both cases simultaneously. Assume X_n be a sequence of IID r.v. with finite expectation $\mu = \mathbb{E}(X_i)$ and variance $\sigma^2 = Var(X_i)$. The steps of CLT are as follows:

Step	For the sum $S_n = X_1 + X_2 + ... + X_n$	For the Sample Mean $\overline{X_n} = \frac{X_1 + X_2 + ... + X_n}{n} = \frac{S_n}{n}$
1. Standardization (Normalization)	$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{Var(S_n)}}$	$Z_n = \frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{Var(\overline{X}_n)}}$
2. Calculation	$\mathbb{E}(S_n) = n \cdot \mu, Var(S_n) = n \cdot \sigma^2$	$\mathbb{E}(\overline{X}_n) = \mu, Var(\overline{X}_n) = \frac{\sigma^2}{n}$
3. The Value of Z_n simplified	$Z_n = \frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma}$	$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$
4. Conclusion	$Z_n = \frac{S_n - n \cdot \mu}{\sqrt{n} \cdot \sigma} \xrightarrow{D} \mathcal{N}(0, 1)$	$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} \mathcal{N}(0, 1)$

I syo!

8.4 Supplement

Let us give examples of Identically Distributed (ID) random variables:

Example, ID Random Variables: Let us construct several r.v. with $Bernoulli(0.5)$ distribution.

- The experiment is a one toss of a fair coin. Then $\Omega = \{H, T\}$. Take $X_1(H) = 0$ and $X_1(T) = 1$. Then $X_1 \sim Bernoulli(0.5)$;
- The experiment is again a one toss of a fair coin. Then $\Omega = \{H, T\}$. Take $X_2(H) = 1$ and $X_2(T) = 0$. Then $X_2 \sim Bernoulli(0.5)$;
- The experiment is a toss of a two fair coins (or a double toss of a one coin). Then $\Omega = \{HH, HT, TH, TT\}$. Let X_3 be the indicator of different sides in the results, i.e. $X_3(HH) = X_3(TT) = 0$ and $X_3(HT) = X_3(TH) = 1$. Then $X_3 \sim Bernoulli(0.5)$;
- The experiment is a roll of a fair die. Then $\Omega = \{1, 2, 3, 4, 5, 6\}$. Now, let X_4 be 0, if the result will show odd number, and let it be 1, if the result is an even number. That is, $X_4(1) = X_4(3) = X_4(5) = 0$ and $X_4(2) = X_4(4) = X_4(6) = 1$. Then, clearly, $X_4 \sim Bernoulli(0.5)$;
- The experiment is again a roll of a fair die. Then $\Omega = \{1, 2, 3, 4, 5, 6\}$. Now, let X_5 be 0, if the result is $\leqslant 3$, and let it be 1, if the result is > 3. That is, $X_5(1) = X_5(2) = X_5(3) = 0$ and $X_5(4) = X_5(5) = X_5(6) = 1$. Then, clearly, $X_5 \sim Bernoulli(0.5)$;
- The experiment is to pick a random number (uniformly) from $[0, 1]$. The probability measure is the length. Define $X_6(\omega) = 0$, if $\omega \in [0, 0.5)$ and $X_6(\omega) = 1$, if $\omega \in [0.5, 1]$. Then $X_7 \sim Bernoulli(0.5)$.

- The experiment is again to pick a random number (uniformly) from $[0,1]$. The probability measure is the length. Define $X_7(\omega) = 0$, if $\omega \in [0, 0.2) \cup (0.7, 1]$ and $X_7(\omega) = 1$, if $\omega \in [0.2, 0.7]$. Then $X_7 \sim \mathtt{Bernoulli}(0.5)$.

Another Supplement:

EXAMPLE, EXPLICITLY GIVEN R.V. AND ITS DISTRIBUTION: Assume X_n is given explicitly: let $\Omega = [0,1]$, and the probability is given through $\mathbb{P}([a,b]) = b - a$ for any $[a,b] \subset \Omega$. Let X_n be the following r.v.:

$$X(\omega) = \begin{cases} 4, & \omega \in [0, 0.3] \\ -3, & \omega \in (0.3, 1]. \end{cases}$$

What can be said about the distribution of X?

Clearly, X takes only 2 values, -3 and 4, so X is discrete. The PMF of X will be:

$$\mathbb{P}(X = -3) = \mathbb{P}(\omega \in (0.3, 1]) = 1 - 0.3 = 0.7, \qquad \text{and} \qquad \mathbb{P}(X = 4) = \mathbb{P}(\omega \in [0, 0.3]) = 0.3 - 0 = 0.3.$$

So we can write the distribution of X in the following form:

Values of X	-3	4
$\mathbb{P}(X = x)$	0.7	0.3

So, in fact, the distribution of X is very familiar to us.

But think like this: if we will have the distribution of X in the table form above, we, unfortunately, cannot say anything about the particular values $X(\omega)$, we cannot reconstruct $X(\omega)$ (even we cannot know where is running ω, i.e., what is the Sample Space Ω).