Seminar 4

Recap of axiomatic definition of probability

A probability space is the following tuple: $(\Omega, \mathcal{F}, \mathbb{P})$.

- **Sample space** Ω
- **Set of events** $\mathcal{F}$
- **Probability measure** $\mathbb{P}$

Set of events is $\mathcal{F} \subset 2^{\Omega}$ (σ-algebra), such that

1. $\Omega \in \mathcal{F}$
2. If $A \in \mathcal{F}$, then $\overline{A} \in \mathcal{F}$ (closed under complement operation)
3. If $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ (closed under countable union operation)

The pair $(\Omega, \mathcal{F})$ is called a measurable space. Set A is called measurable if $A \in \mathcal{F}$.

Probability measure is $\mathbb{P} : \mathcal{F} \to \mathbb{R}_{+}$, such that

1. $\mathbb{P}(\Omega) = 1$
2. If $A_1, A_2, \ldots \in \mathcal{F}$ and $A_i \cap A_j = \varnothing$ for $i \neq j$, then $\mathbb{P}\left(\bigcup_{k=0}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)$ (σ-additivity)

Example 1

Let

- $\Omega = (0, 1]$
- $\mathcal{F} = 2^{(0,1]}$
- $\mathbb{P}(A) = \frac{k}{n}$, where k is the number of points like $\frac{i}{n}, i \in \{1, \ldots, n\}$ in A

We can check that all the necessary conditions are satisfied:

1. $\Omega = (0, 1] \subset 2^{(0,1]} = \mathcal{F}$
2. If $A \subset (0, 1] \in \mathcal{F}$, then $\overline{A} \subset (0, 1] \in \mathcal{F}$ as well
3. If $A_1, A_2, \ldots \subset (0, 1] \in \mathcal{F}$, then all their elements are in $(0, 1]$, and thus the union $\bigcup_{k=1}^{\infty} A_k \subset (0, 1] \in \mathcal{F}$
4. $\mathbb{P}(\Omega) = \mathbb{P}((0, 1]) = \frac{n}{n} = 1$
5. If A_1 and $A_2 \in \mathcal{F}$ and $A_1 \cap A_2 = \varnothing$, then
   $\frac{k_1 + k_2}{n} = \mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) = \frac{k_1}{n} + \frac{k_2}{n}$

Example 2

Let

- $\Omega = (0, 1]$
- $\mathcal{F}$ is the set of all half-intervals $(a, b]$ in $(0, 1]$
- $\mathbb{P}((a, b]) = b - a$ (length of half-interval)

In this case, $\mathcal{F}$ is not a σ-algebra, because union of half-intervals is not necessarily a half-interval.

We need to add something else to $\mathcal{F}$. We can add all finite unions of half-intervals, then $\mathcal{F}$ will be an algebra, but still not σ-algebra. We can **not** build such $\mathcal{F}$ by hand, but it exists according to the following theorem.

Theorem 1: Let $\mathcal{A}$ be some set of subsets of set A, then exists a minimal σ-algebra $\sigma(\mathcal{A})$, which contains $\mathcal{A}$. It means that this $\sigma(\mathcal{A})$ will be a part of any larger σ-algebra, that contains $\mathcal{A}$.

Theorem 2 (Caratheodory Theorem): Let probability measure $\mathbb{P}$ be defined on algebra $\mathcal{A}$ and σ-additive on it. Then $\mathbb{P}$ can be extended to $\sigma(\mathcal{A})$ uniquely.

This gives us the following result: The measure $\mathbb{P}$, defined on half-intervals in $(0, 1]$ as $\mathbb{P}((a, b]) = b - a$, can be uniquely extended to a minimal σ-algebra containing such half-intervals. Then it will be a probability measure.

Lebesgue measure

What we just defined is called **probabilistic Lebesgue measure** $\lambda((a, b]) = \mathbb{P}((a, b]) = b - a$. Next, we can define non-probabilistic measure.

A mapping $\mu : \mathcal{F} \to [0, +\infty)$ is called a **measure** if it is additive and σ-additive (we simply ignore the $\mathbb{P}(\Omega) = 1$ property).

So we defined λ on $(0, 1]$. We can naturally extend it to $(n, n + 1]$: the measure of set $A \subset 2^{(n,n+1]}$ will be equal to measure of set $B \subset 2^{(0,1]}$ obtained by shifting the set.

Finally, for a general set $A \subset \mathbb{R}$, define

$$\lambda(A) = \sum_{n \in \mathbb{Z}} \lambda\left(A \cap (n, n + 1]\right)$$

This is the **Lebesgue measure on real line**, what we usually call length.

Borel σ-algebra

Minimal σ-algebra $\mathcal{B}(A)$ that contains all open subsets of A is called **Borel σ-algebra**.

Lemma: Borel σ-algebra of subsets of $(0, 1]$ coincides with minimal σ-algebra, containing all half-intervals.

Since we extended Lebesgue measure to the whole real line, we can find measure of every set in $\mathcal{B}(\mathbb{R})$.

Measurable mapping

Consider

- Two measurable spaces $(X, \mathcal{F}_X)$ and $(Y, \mathcal{F}_Y)$
- A mapping $T : X \to Y$
- A measurable set $A \in \mathcal{F}_X$

Full pre-image of A under T is then

$$T^{-1}(A) = \{x \in X | T(x) \in A\}$$

Full pre-image of A under T can also be measurable: $T^{-1}(A) \in \mathcal{F}_X$, but not necessarily. If for any measurable set A its full pre-image under T is measurable, we say that T is a **measurable mapping**.

Random variables

Consider probabiliy space $(\Omega, \mathcal{F}, \mathbb{P})$. A **random variable** is a measurable function $X : \Omega \to \mathbb{R}$ from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

It means that the pre-image of any set A in $\mathcal{B}(\mathbb{R})$ belongs to $\mathcal{F}$:

$$\forall A \in \mathcal{B}(\mathbb{R}) \implies X^{-1}(A) \in \mathcal{F}$$

Think of event $B \in \mathcal{B}(\mathbb{R})$ and its pre-image $A \in \mathcal{F}$. Naturally, the probability that random variable X lies in set B is the same as probability of event A:

$$\mathbb{P}(X \in B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(A)$$

Example 3

Let

- $\Omega = [0, 1]$
- $\mathcal{F} = \mathcal{B}([0, 1])$
- $X_1(\omega) = \omega$

In order to check if X_1 will be a random variable, we need to verify that X_1 is a measurable function from $([0, 1], \mathcal{B}([0, 1]))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

It will be such measurable function, because the pre-image of any set $[a, b] \in \mathcal{B}(\mathbb{R})$ lies in $\mathcal{B}([0, 1])$:

$$X_1^{-1}([a, b]) = [a, b] \cap [0, 1]$$

Example 4

Let

- $\Omega = [0, 1]$
- $\mathcal{F} = \{\Omega, \varnothing\}$
- $X_1(\omega) = \omega$

In order to check if X_1 will be a random variable, we need to verify that X_1 is a measurable function from $([0, 1], \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

It will **not** be such measurable function, because the pre-image of e.g. $[0, 1/2] \in \mathcal{B}(\mathbb{R})$ does not lie in $\mathcal{F}$:

$$X_1^{-1}([0, 1/2]) = [0, 1/2] \notin \{[0, 1], \varnothing\}$$

Independence of random variables

A **minimal σ-algebra generated by random variable** X is the minimal σ-algebra containing pre-images of all borel sets:

$$\sigma(X) = \sigma\{X^{-1}(B), B \in \mathcal{B}(\mathbb{R})\} = \{X^{-1}(B), B \in \mathcal{B}(\mathbb{R})\}$$

Random variables X and Y are called independent if their minimal generated σ-algebras are independent. This means that any events $A \in \sigma(X)$ and $B \in \sigma(Y)$ should be independent:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Let's say A is pre-image of some borel event B_1 and B is the pre-image of a different borel event B_2. Then, $\mathbb{P}(A) = \mathbb{P}(X^{-1}(B_1)) = \mathbb{P}(X \in B_1)$ and similarly for B_2, so finally,

$$\mathbb{P}(X \in B_1, X \in B_2) = \mathbb{P}(X \in B_1)\mathbb{P}(X \in B_2)$$

Distribution of a random variable

Consider probability sapce $(\Omega, \mathcal{F}, \mathbb{P})$ and random variable $X : \Omega \to \mathbb{R}$. We will call the image μ of measure $\mathbb{P}$ through the mapping X **the distribution** (or distribution law) of X :

$$\mu(A) = \mathbb{P}(X^{-1}(A))$$

We will write $X \sim \mu$.

Three types of distributions

Lebesgue theorem: Let ν be Lenesgue measure on $\mathbb{R}$ and μ be any probabilistic measure, then $\mu = \mu_d + \mu_s + \mu_{ac}$, where

- μ_d is **discrete measure**, i.e. it is concentrated on a countable set of points.
- μ_s is **singular measure**, i.e. exists measurable set S such that $\nu(S) = 0$ and $\mu_s(\overline{S}) = 0$ and $\forall x \in \mathbb{R}\, \mu_s(\{x\}) = 0$.
- μ_{ac} is **absolutely continuous measure**, i.e. from $\nu(A) = 0$ follows $\mu_{ac}(A) = 0$ for any measurable set A. By **Radon-Nikodim theorem**, it is equivalent to the existence of a non-negative measurable function $f : \mathbb{R} \to \mathbb{R}$ called **probability density function**, such that $\mu_{ac}(A) = \int_A f(x)dx$.

Because the distributions are defined through the measure, any probability distribution may be viewed as a mixture of three base types: discrete, singular and continuous.

Normally though, the distributions fall into just one category. Also, you never encounter singular distributions in practice.

Example 5

Consider event $A \in \mathcal{F}$ and a random variable $X = \mathbb{Ind}_A$, an indicator:

$$\mathbb{Ind}_A(x) = \begin{cases} 1, x \in A, \\ 0, \text{else} \end{cases}$$

$$\mathbb{P}(X = 1) = \mathbb{P}(A) = p$$

$$\mathbb{P}(X = 0) = 1 - \mathbb{P}(A) = 1 - p$$

We say that X follows **Bernoulli distribution** with parameter p and write $X \sim Be(p)$.

We will call $\mathbb{P}_X(\omega)$ a **probability mass function** (PMF).

Bernoulli trial scheme

Previously we have worked with independent events that were happening in one probability space. But sometimes we want to have multiple trials, where for every trial the probability space is known, but we are interested in the probability space covering all the trials at once. We can achieve it via direct product of probability spaces.

If all probability spaces are the same and equal to:

- $\Omega = \{0, 1\}$
- $\mathcal{F} = \{\varnothing, 0, 1, \Omega\}$
- $\mathbb{P}(1) = p$ and $\mathbb{P}(0) = 1 - p$

Then we call such experiment a **Bernoulli trial scheme**, and the probability space of it is:

- $\Omega = \{(i_1, \ldots, i_n), i_j \in \{0, 1\}\}$
- $\mathcal{F} = 2^{\Omega}$
- $\mathbb{P}(i_1, \ldots, i_n) = p^{\mathrm{num}\, j \text{ such that } i_j = 1}(1 - p)^{\mathrm{num}\, j \text{ such that } i_j = 0}$

Example 6

Consider $X_1, \ldots, X_n \sim Be(p)$ independent random variables. Then $Y = \sum_{k=1}^{n} X_k$ follows **Binomial distribution** with parameters n and p, $Y \sim Bi(n, p)$. $\mathbb{P}(Y = k) = ?$

Solution 6

$$\mathbb{P}(Y = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Example 7

Consider X and Y independent $\mathbb{Z}$-valued random variables. $\mathbb{P}(X + Y = k) = ?$

Solution 7

$$\mathbb{P}(X + Y = k) = \sum_{m} \mathbb{P}(X = m)\mathbb{P}(Y = k - m)$$

Example 8

Let $X \sim Bi(n, p)$ and $Y \sim Bi(m, p)$ be independent. What is the distribution of $Z = X + Y$?

Solution 8

$$\mathbb{P}(X + Y = k) = \sum_j \binom{n}{j} p^j (1-p)^{n-j} \binom{m}{k-j} p^{k-j}(1-p)^{m-k+j} = p^k(1-p)^{n+m-}$$

$$Z \sim Bi(n+m, p)$$

Cumulative distribution function

Note that distribution of a discrete distribution of a random variable X is uniquely defined by its PMF $\mathbb{P}(X = x_i)$. In general, we define the distribution using cumulative distribution function (CDF):

$$F_X(x) = \mathbb{P}(X < x)$$

It has the following properties:

- F_X is non-decreasing
- $\lim\limits_{x \to -\infty} F_X(x) = 0$
- $\lim\limits_{x \to +\infty} F_X(x) = 1$
- F_X if left continuous

Interesting enough, the converse is also true. Any function that conforms to the properties above defines some probability distribution on $\mathbb{R}$ and this relation is unique.

Probability density function

- If X has a discrete distribution, then F_X has a countable number of jumps $p_i = \mathbb{P}(X = x_i)$ and at $x = x_i$ it is continuous
- If X has absolutely continuous distribution, then F_X is differentiable a.e. and can be recovered from its derivative:

$$F_X(x) = \int\limits_{-\infty}^{x} f_X(t)dt$$

where $f_X(t)$ is the probability density function and $f_X(t) = F_X'(x)$ a.e.

Example 9

We say that random variable X is distributed uniformly on $[a, b]$ and write $X \sim U([a, b])$
if

$$f_X(x) = \begin{cases} \frac{1}{b-a}, a \leqslant x \leqslant b, \\ 0, \text{else} \end{cases}$$
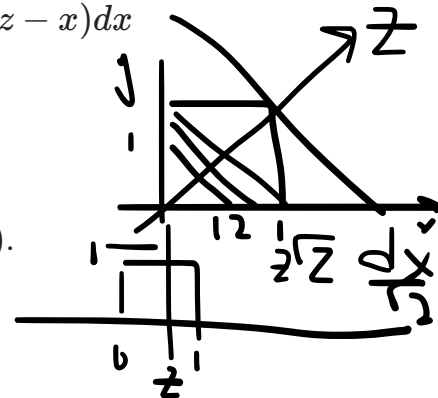
Example 10

Consider X and Y independent random variables with PDFs f_X and f_Y respectively.
Then, their sum $Z = X + Y$ has absolutely continuous distribution with density

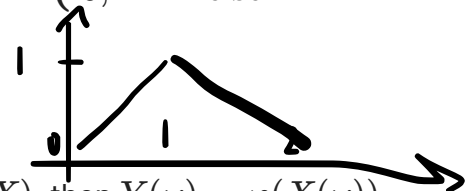$$f_Z(z) = \int f_X(x) f_Y(z - x) dx$$

Example 11

Let $X, Y \sim U([0, 1])$ and $Z = X + Y$. Find $f_Z(z)$.

Solution 11

$$f_Z(\underline{z}) = \int_0^1 \underline{f_X(x)} \underline{f_Y(z - x)} dx = \int_0^1 f_Y(\underline{z - x}) dx = \begin{cases} z, & 0 \leqslant z \leqslant 1, \\ 2 - z, & 1 \leqslant z \leqslant 2, \\ 0, & \text{else} \end{cases}$$

Functions of random variables

Random variables transform like functions, i.e. if $Y = \underline{\varphi(X)}$, then $Y(\omega) = \varphi(X(\omega))$.

Example 12

Let $\underline{X}$ be a random variable with CDF $\underline{F_X}$ and PDF $\underline{f_X}$. Find CDF and PDF of
$Y = \underline{a}X + b$.

loc-scale transform

Solution 13

If $a > 0$:

$$\underline{F_Y(y)} = \mathbb{P}(\underline{Y} < \underline{y}) = \mathbb{P}(a\underline{X} + \underline{b} < \underline{y}) = \mathbb{P}\left(X < \frac{y-b}{a}\right)$$

$$f_Y(y) = F_Y'(y) = \frac{1}{|a|}\underline{f_X}\left(X = \frac{y-b}{a}\right)$$

In general for a smooth φ, the density will be:

$$\underline{f_Y(\underline{y})} = \sum_{\varphi(\underline{x})=y} \frac{f_X(x)}{|\varphi'(x)|}$$

Example 13

Let X be a **normally distributed** random variable with parameters m and σ^2:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

Find PDF of $Y = X^2$.

Solution 13