Special Mathematics Lecture
Introduction to probability

Table of content

Probability

I Events and probabilities

I.1 Basic definitions

Aim : describe a non-predictable experiment, like throwing a die = dice.

Omega

Def: Sample space Ω is the set of all possible outcomes of an experiment. An element $\omega \in \Omega$ is called an elementary event.

$\Rightarrow$ the set Ω contains all possible elementary events.

Example : Throwing one die : $\Omega = \{1, 2, 3, 4, 5, 6\}$

and $\omega = 4$ if we get .

Sometimes, not all elementary events are interesting, some should be put together.

$\rightsquigarrow$ List of interesting events :

odd result — even result

max result.

Examples : $\{1, 3, 5\}$ and $\{2, 4, 6\}$

$\{1, 2, 3\}$ and $\{4, 5\}$ and $\{6\}$

small result — medium result

The list of interesting events is made of subsets

subset of

of Ω. If $A \subset \Omega$, $B \subset \Omega$, then

union

$\omega \in (A \cup B)$ means ω belongs either to A or to B,

intersection

$\omega \in (A \cap B)$ means ω belongs to A and B,

complement Ω without A

$\omega \in A^c := \Omega \setminus A$ means ω does not belong to A,

symmetric difference

$\omega \in A \,\Delta\, B$ means ω belongs to A or to B,

but not to both.

Also : $\Omega \setminus (A \cap B) = \Omega \setminus A \cup \Omega \setminus B$ check!

An arbitrary family of subsets of Ω is often not interesting, we require some compatibility relations between the elements of this family.

<u>Def</u>: A collection $\mathcal{F}$ of subsets of Ω is called an <u>event space</u> on Ω if:

1) $\mathcal{F}$ is not empty
2) If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ — stability of $\mathcal{F}$ for complement
3) If $\{A_i\}_{i\in\mathbb{N}} \subset \mathcal{F}$, then $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F}$

stability of $\mathcal{F}$ for countable unions

<u>Remarks</u>: 1) It follows that $\phi, \Omega \in \mathcal{F}$ (since — empty set

$\Omega = A \cup A^c$ and $\phi = \Omega^c$

2) $\mathcal{F}$ is stable for countable intersections

check!

we also say that $\mathcal{F}$ is closed for countable intersections

(One can use that if $A, B \in \mathcal{F}$, then A^c, B^c and $A^c \cup B^c$ belong to $\mathcal{F}$, and then

$(A \cap B)^c = A^c \cup B^c \in \mathcal{F} \Rightarrow ((A\cap B)^c)^c = A \cap B \in \mathcal{F}$.)

Example : Always with 1 die :

$\mathcal{F} = \{ \emptyset, \{1,2\}, \{3,4\}, \{5,6\}, \{1,2,3,4\}, \{1,2,5,6\},$

$\{3,4,5,6\}, \{1,2,3,4,5,6\}\}$.

We would like now to associate a weight to each element of $\mathcal{F}$, it means a positive number, with some compatibility conditions :

Def A map ($\equiv$ function) $\mathbb{P} : \mathcal{F} \to [0,1]$ is called a probability measure on $(\Omega, \mathcal{F})$

if 1) $\mathbb{P}(\Omega) = 1$, $\mathbb{P}(\emptyset) = 0$

2) if $\{A_i\}_{i\in\mathbb{N}} \subset \mathcal{F}$ are disjoint ($\equiv A_j \cap A_k = \emptyset$ whenever $j \neq k$), then

$$\mathbb{P}\Big(\underbrace{\bigcup_{i\in\mathbb{N}} A_i}_{\in \mathcal{F} \text{ by the stability for countable union}}\Big) = \sum_i \mathbb{P}(A_i)$$

We say that $\mathbb{P}$ is countably additive

Remark: $\mathbb{P}$ is only defined on the list of interesting events (namely on $\mathcal{F}$) and not on all subsets of Ω.

Properties: For any $A, B \in \mathcal{F}$:

1) $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$

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

3) if $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$

check!

Example: (Power set) If $\Omega = \{\omega_1, \dots, \omega_N\}$ finite set

we often consider $\mathcal{F} = \{$ all subsets of $\Omega \}$ in this case, it is called the <u>power set</u> of Ω

and define $\underset{\mathcal{F}}{\mathbb{P}(\underset{\cap}{A})} := \frac{1}{N} \underbrace{|A|}$, number of element of A

then $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$.

check!

Summing up :

<u>Def</u> : <u>A probability space</u> is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, with Ω an non-empty set of possible outcomes, $\mathcal{F}$ an event space on Ω (interesting events), and $\mathbb{P}$ a probability measure on $(\Omega, \mathcal{F})$.

<u>Remark</u> If Ω is finite, one often considers the power set of Ω for $\mathcal{F}$, and many simple problems reduce in counting the number of ways some events may occur.

I.2 Conditional probabilities

<u>Question</u> : If we have already a partial information about the outcome of the experiment, how can one implement this knowledge ?

<u>Example</u> : Throwing a die and knowing that the outcome is an even number.

In the sequel $(\Omega, \mathcal{F}, \mathbb{P})$ is always a probability space.

<u>Def</u>: If $A, B \in \mathcal{F}$ and $\mathbb{P}(B) > 0$, the probability of A given ($\equiv$ knowing) B is given by
$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

read as a conditional probability, the prob. of A knowing B.

<u>Remarks</u>: 1) $\mathbb{P}(B|B) = 1$ always!

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

$$\Longrightarrow \quad \mathbb{P}(B|A) = \mathbb{P}(A|B)\,\frac{\mathbb{P}(B)}{\mathbb{P}(A)} \quad \text{if } \mathbb{P}(A) \neq 0$$

<u>Example</u>: For the power set of $\Omega = \{1,2,3,4,5,6\}$ and the probability measure $\mathbb{P}(A) = \frac{1}{6}|A|$, if

set of even numbers

$B = \{2,4,6\}$ and $A = \{2\}$, one has

$$\mathbb{P}(A|B) = \mathbb{P}(\{2\}|\{2,4,6\}) = \frac{\mathbb{P}(\{2\} \cap \{2,4,6\})}{\mathbb{P}(\{2,4,6\})} = \frac{\mathbb{P}(\{2\})}{3/6}$$
$$= \frac{1/6}{3/6} = \underline{\underline{\frac{1}{3}}}$$

probability of getting 2 if we know that the outcome is an even number.

Thm: If $B \in \mathcal{F}$ and $\mathbb{P}(B) > 0$, then $(\Omega, \mathcal{F}, \mathbb{Q})$ is a probability space, with

$$\mathbb{Q}(A) := \mathbb{P}(A \mid B) \qquad \forall A \in \mathcal{F}.$$

Framework when a partial information is known.

Def: The sets $A, B \in \mathcal{F}$ are <u>independent</u> if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$. They are dependent otherwise.

Remark: If $A, B \in \mathcal{F}$ are independent, then

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\,\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A), \text{ and}$$

$$\mathbb{P}(B \mid A) = \dots = \mathbb{P}(B).$$

⚠ The intermediate steps require $\mathbb{P}(B) > 0$ or $\mathbb{P}(A) > 0$, but this is not required in the notion of independence.

Def: A **partition** of Ω is a collection $\{B_i\}_{i\in\mathbb{N}} \subset \mathcal{F}$ such that $B_j \cap B_k = \emptyset$ and $\bigcup_i B_i = \Omega$.

Think about a puzzle!

Thm (partition): If $\{B_i\}_i$ is a partition of Ω, and if $A \in \mathcal{F}$, then $\mathbb{P}(A) = \sum_{i\in\mathbb{N}} \mathbb{P}(A|B_i)\,\mathbb{P}(B_i)$.

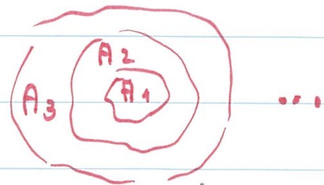
Proof ?

Thm (Bayes) If $\{B_i\}_{i\in\mathbb{N}}$ is a partition of Ω, and $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$, then

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\,\mathbb{P}(B_j)}{\underbrace{\sum_i \mathbb{P}(A|B_i)\,\mathbb{P}(B_i)}_{=\mathbb{P}(A)}}$$

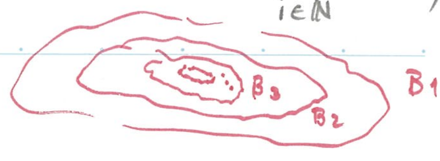
Very useful in the applications.

Continuity of probability measure:



Thm: 1) If $\{A_i\}_{i\in\mathbb{N}} \subset \mathcal{F}$, $A_i \subset A_{i+1}$ and $A = \bigcup_{i\in\mathbb{N}} A_i$, then $\mathbb{P}(A) = \lim_{i\to\infty} \mathbb{P}(A_i)$.

2) If $\{B_i\}_{i\in\mathbb{N}} \subset \mathcal{F}$, $B_{i+1} \subset B_i$ and $B = \bigcap_{i\in\mathbb{N}} B_i$, then $\mathbb{P}(B) = \lim_{i\to\infty} \mathbb{P}(B_i)$.



II Discrete valued random variables

Example : Consider again a die, with

$\Omega = \{1, 2, 3, 4, 5, 6\}$, $\mathcal{F} := P(\Omega)$, and $\mathbb{P}$

↑ power set of Ω

the natural measure $\mathbb{P}(A) = \frac{1}{6} \# A$.

We gamble : 3 \$ if the outcome is 6

0 \$ if the outcome is 4 or 5

More abstractly:

-1 \$ if the outcome is 1, 2, or 3.

It means that we create a function $X : \Omega \to \mathbb{Z}$

set of all integers

with $X(1) = X(2) = X(3) = -1$

$X(4) = X(5) = 0$

$X(6) = 3$

Any element of Ω is associated to

a number in $\mathbb{Z}$ (or in $\mathbb{R}$ later).

can be more general than $\mathbb{Z}$, but less general than $\mathbb{R}$.

Def : A <u>discrete valued random variable</u> on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that

notation

1) $X(\Omega) =$ image of Ω by $X \equiv \mathrm{Im}(X)$

$= \{ X(\omega) \mid \omega \in \Omega \}$ is <u>countable</u>

2) $\forall x \in \mathbb{R}, \; \{ \omega \in \Omega \mid X(\omega) = x \} \in \mathcal{F}$

$\Leftrightarrow \quad \underline{X^{-1}(x)} \in \mathcal{F} \quad \forall x \in \mathbb{R}.$

the meaning of this notation is precisely

$\{ \omega \in \Omega \mid X(\omega) = x \}$

Remember that <u>countable</u> means that it can be indexed by $\mathbb{N}$ (you can enumerate the elements).

In the previous example, X is a discrete valued random variable.

Remarks :

1) If X, Y are discrete valued random variables,

so are $X+Y$ and XY. Can you prove it?

all subsets of Ω

2) If $\mathcal{F}$ is the power set $P(\Omega)$ of Ω,

then any function from Ω to a countable subset of

$\mathbb{R}$ is a discrete valued random variable.

it means that if $\mathcal{F} = P(\Omega)$, the condition 2) of the previous definition is always satisfied.

random variable

Notation : ⚠ The set $\{\omega \in \Omega \mid X(\omega) = x\}$

values in $\mathbb{R}$

will be denoted by $\{X = x\}$

So far we have not used the probability

measure $\mathbb{P}$. In the next definition, it

plays an essential role.

<u>Def</u> : Given a discrete valued random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the <u>probability mass function</u> (= pmf) of X, also called <u>probability distribution</u> of X, is the function

short notation

$$p_X : \mathbb{R} \to [0,1] \,, \quad p_X(x) := \mathbb{P}(X = x)$$
$$= \mathbb{P}(\{X = x\}) .$$

much simpler than $X : \Omega \to \mathbb{R}$

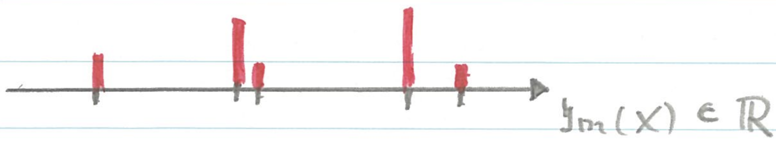
<u>Remarks</u> : 1) $p_X(x) = 0$ if $x \notin \mathrm{Im}(X)$

2) $$\sum_{x \in \mathrm{Im}(X)} p_X(x) = \sum_{x \in \mathrm{Im}(X)} \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = x\})$$

$$= \mathbb{P}\Big(\bigcup_{x \in \mathrm{Im}(X)} \{\omega \in \Omega \mid X(\omega) = x\} \Big) = \mathbb{P}(\Omega) = 1.$$

3) We are familiar with this picture

the values sum up to 1.



In fact this picture contains all the useful information!

In fact, once the probability distribution of a random variable is known, the prob. space $(\Omega, \mathcal{F}, \mathbb{P})$ is no more necessary, as shown here:

finite set, or $\mathbb{N}$

Given a countable set $S = \{s_i\}_{i\in I} \subset \mathbb{R}$ and $\{\pi_i\}_{i\in I} \subset [0,1]$ with $\sum_{i\in I} \pi_i = 1$,

$\exists\, (\Omega, \mathcal{F}, \mathbb{P})$ a probability space, and X a

there exists

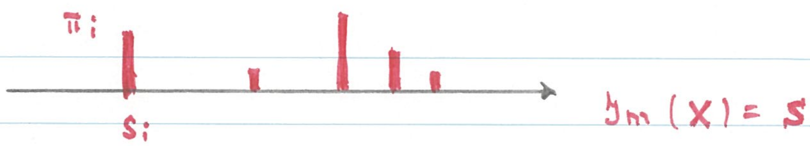
discrete valued random variable such that

$$P_X(s_i) = \pi_i \quad \forall\, i \in I, \quad P_X(s) = 0 \text{ if } s \notin S.$$

power set

For this, one can take $\Omega := S$, $\mathcal{F} := \mathcal{P}(\Omega)$

and $\mathbb{P}(\underset{\in \mathcal{F}}{A}) := \sum_{i \mid s_i \in A} \pi_i$, and the random variable $X : \Omega \to \mathbb{R}$, $X(s_i) := s_i$



Thus a discrete valued random variable is fully determined by the values $\{s_i\}_{i\in I}$ it takes with a probability $\{\pi_i\}_{i\in I}$.

<u>Remark</u> : The random variable defined by $X(\omega) = \omega$ is very convenient, but can lead to some confusion. In general : $X : \Omega \to \mathbb{R}$ with $\Omega \neq \mathbb{R}$. Thus $X(\omega) = \omega$ is often meaningless. However, if $\Omega \subset \mathbb{R}$ and if we consider the discrete valued random variable $X(\omega) = \omega$, then one can speak about the probability of ω (which is meaningless in general). Here, it means $\mathbb{P}(X = \omega) =$

$$= \mathbb{P}(\{\omega' \in \Omega \mid X(\omega') = \omega\}) = \mathbb{P}(\{\omega\}).$$

because of the special choice $X(\omega) = \omega$

It is this identification which gives us, when throwing a die, that $\mathbb{P}(5) = \frac{1}{6}$. This is correct because we have identified ⚄ with 5, and $\Omega =$ {⚀, ⚁, ⚂, ⚃, ⚄, ⚅} with $\{1,2,3,4,5,6\} \subset \mathbb{R}$. Thus $\mathbb{P}(X=5) = \mathbb{P}($\{⚄\}$) = \frac{1}{6}$.

X(⚀) = 1 ...

Examples (based on the previous observation)

1) **Bernoulli (probability) distribution** (parameter $p \in [0,1]$)

$S = \{0, 1\}$, $\pi_1 = p \in [0,1]$, $\pi_0 = 1-p =: q$

only 2 values — probability of 1 (success) — probability of 0

2) **Binomial distribution** (parameter $p \in [0,1]$)

$S = \{0, 1, 2, \dots, n\}$ $\qquad p \in [0,1]$, $q := 1-p$

$$\pi_k := \binom{n}{k} p^k q^{n-k} \qquad k = 0, 1, \dots, n$$

binomial coefficient $\binom{n}{k} = \dfrac{n!}{k!\,(n-k)!}$

One has $\displaystyle\sum_{k=0}^{n} \pi_k = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+k)^n = 1.$

π_k = prob. of k success (1) in n independent experiment, each of them having a probability p of success. This corresponds to n independent Bernoulli experiments.

3) Poisson distribution (parameter $\lambda > 0$)

$S = \{0, 1, 2, 3, \ldots\}$ $\qquad \pi_k := \frac{1}{k!} \lambda^k e^{-\lambda}$.

One has $\sum_{k=0}^{\infty} \pi_k = \sum_{k=0}^{\infty} \frac{1}{k!} \lambda^k e^{-\lambda} = e^{\lambda} e^{-\lambda} = 1$.

If λ represents the average number of a certain event in an interval of time T, then π_k is the probability that this event takes place k time during this interval of time T, assuming that the system has no memory and that 2 events can not take place simultaneously.

4) Geometric distribution ⚠ 2 slightly different definitions exist.

→ parameter $p \in [0,1]$

$S = \{1, 2, 3, 4, \ldots\}$

$\pi_k := p\, q^{k-1}$ $\qquad$ One has $\sum_{k=1}^{\infty} p\, q^{k-1} = p \frac{1}{1-q} = 1$

probability that the first occurrence of a success requires k independent trials of a Bernoulli experiment.

Even if probability distributions are useful for $\underline{1}$ random variable, one needs the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ for deeper arguments.

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a discrete valued random variable X. If $g : \mathbb{R} \to \mathbb{R}$ is a function, then $g \circ X : \Omega \to \mathbb{R}$ defined by $(g \circ X)(\omega) := g(X(\omega))$ $\forall \omega \in \Omega$ is again a discrete value random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.

$g \circ X$ is called a <u>function of the random variable X</u>, and its probability distribution is given by

$$P_{g\circ X}(y) = \mathbb{P}(g \circ X = y) = \mathbb{P}(\{\omega \in \Omega \mid g(X(\omega)) = y\})$$

$$= \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = g^{-1}(y)\})$$

$$= \sum_{x \in g^{-1}(y)} P_X(x)$$

<u>Remark</u>: there is no condition on g because X is discrete valued

Def: For a discrete valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, its <u>mean</u> $\equiv$ <u>expected value</u> $\equiv$ <u>expectation</u> or average value is defined by

$$\mathbb{E}(X) := \sum_{x \in \mathrm{Im}(X)} x \, \mathbb{P}(X = x) = \sum_{x \in \mathrm{Im}(X)} x \, p_X(x)$$

whenever $\displaystyle\sum_{x \in \mathrm{Im}(X)} |x| \, p_X(x) < \infty$.

This is called the absolute convergence

Example: When throwing a die, with

$X(\omega) = \omega$, $\omega = 1, 2, \dots, 6$, and $\mathbb{P}(A) = \frac{1}{6} \# A$,

then $\mathbb{E}(X) = 1 \cdot \frac{1}{6} + 2 \frac{1}{6} + 3 \frac{1}{6} + \dots + 6 \frac{1}{6} = \frac{21}{6} = \frac{7}{2}$.

Thm: For $g : \mathbb{R} \to \mathbb{R}$, one has

$$\mathbb{E}(g(X)) = \sum_{x \in \mathrm{Im}(X)} g(x) \, p_X(x)$$

whenever $\displaystyle\sum_{x \in \mathrm{Im}(X)} |g(x)| \, p_X(x) < \infty$.

Proof?

Properties :

1) $\mathbb{E}(aX+b) = a\,\mathbb{E}(X) + b \quad \forall\, a,b \in \mathbb{R}$

2) If $\mathbb{P}(X \geq 0) = 1$ and $\mathbb{E}(X) = 0$, then

$\quad\; \| $

$\mathbb{P}(\{\omega \in \Omega \mid X(\omega) \geq 0\}) = 1$ $\qquad$ $\mathbb{P}(X = 0) = 1.$

The dispersion of X around its mean is given by:

Def : The variance $\operatorname{var}(X)$ of a discrete valued random variable is defined by

small proof

$$\operatorname{var}(X) := \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

whenever this quantity is finite.

Suppose now that $B \in \mathcal{F}$ with $P(B) > 0$. If we know that B occurs, this affects the expectation of X. One speaks about a conditional expectation :

Def : If X is a discrete valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and if $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, the conditional expectation of X knowing B is given by

$$\mathbb{E}(X \mid B) := \sum_{x \in \mathrm{Im}(X)} x \, \mathbb{P}(X = x \mid B)$$

$$= \sum_{x \in \mathrm{Im}(X)} x \, \frac{\mathbb{P}(\{\omega \in \Omega \mid X(\omega) = x\} \cap B)}{\mathbb{P}(B)}$$

whenever it converges absolutely.

Thm (Partition thm)

If $\{B_i\}_i$ is a partition of Ω with $\mathbb{P}(B_i) > 0$ $\forall i$, then $\mathbb{E}(X) = \sum_i \mathbb{E}(X \mid B_i) \, \mathbb{P}(B_i)$.

Proof ?

III Multivariate discrete distributions

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $X_1, X_2, \ldots, X_n$ be discrete valued random variables on it.

<u>Def</u>: The <u>joint probability mass function</u>, or <u>joint probability distribution</u>, is the function $p_{X_1,\ldots X_n} : \mathbb{R}^n \to [0,1]$ defined by

$$p_{X_1,\ldots,X_n}(\underbrace{x_1,\ldots,x_n}_{\in \mathbb{R}^n}) = \mathbb{P}(\{\omega\in\Omega \mid X_j(\omega) = x_j \ \forall j\})$$
$$\equiv \mathbb{P}(X_j = x_j,\ j\in\{1,\ldots,n\}).$$

Clearly $p_{X_1,\ldots,X_n}(x_j,\ldots,x_n) = 0$ if $x_j \notin X_j(\Omega)$ for some j, and also

$$\sum_j \sum_{x_j\in X_j(\Omega)} p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = 1.$$

proof as in Chapter II

it means that if $x_j \notin X_j(\Omega)$ for at least one j, then the expression is 0.

Observe that the individual probability distribution can be recovered by summing over the other variables, namely

$$p_{X_j}(x_j) = \sum_{k \neq j} \sum_{x_k \in X_k(\Omega)} p_{X_1,\dots,X_n}(x_1, \dots x_n).$$

also called the marginal mass function of X_j

Consider now a function $g : \mathbb{R}^n \to \mathbb{R}$. One can define a new discrete valued random variable $g(X_1, \dots, X_n) : \Omega \to \mathbb{R}$ by

$$g(X_1, \dots X_n)(\omega) := g(X_1(\omega), \dots, X_n(\omega)).$$

<u>Thm</u>: The expectation of $g(X_1, \dots, X_n)$ is given by

$$\mathbb{E}(g(X_1, \dots, X_n)) = \sum_j \sum_{x_j \in X_j(\Omega)} g(x_1, \dots, x_n)\, p_{X_1,\dots X_n}(x_1, \dots, x_n)$$

whenever the sum converges absolutely.

We have already seen this for $n=1$ in Chapter II

<u>Exercise</u>: Check that $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$.

Recall from Chapter I that 2 events $A, B \in \mathcal{F}$ are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$.

The next definition is partially related to this, but for random variables.

<u>Def</u>: Let $X_1, \dots, X_n$ be discrete valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

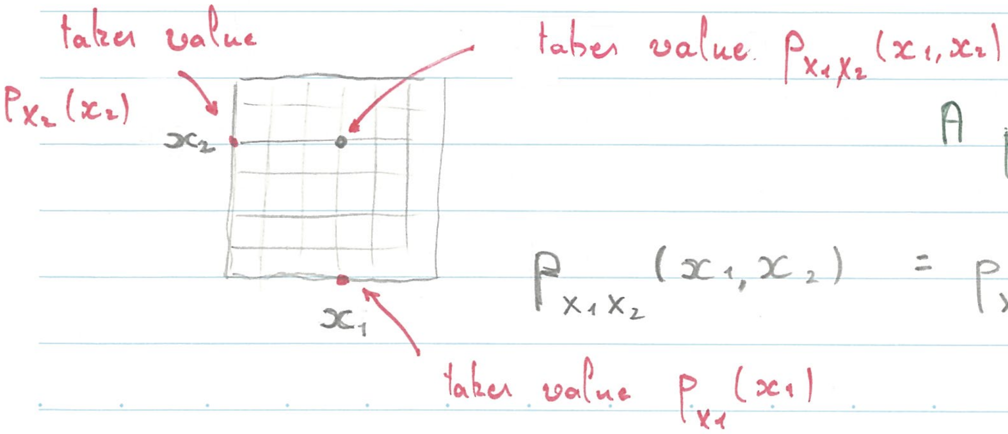
They are <u>independent</u> if $\forall x_j \in \mathbb{R}$ one has

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{j=1}^{n} \mathbb{P}(X_j = x_j)$$

which can also be written as

$$p_{X_1, \dots, X_n}(x_1, \dots x_n) = \prod_{j=1}^{n} p_{X_j}(x_j)$$

" the joint probability distribution is equal to the product of the individual probability distribution"



A picture for this?

$$p_{X_1 X_2}(x_1, x_2) = p_{X_1}(x_1)\; p_{X_2}(x_2)$$

<u>Example :</u> If $i, j \in \{0, 1, 2, \dots\}$ and

$$p_{X,Y}(i,j) = \frac{1}{i!\,j!} \lambda^i \mu^j e^{-(\lambda+\mu)} \qquad \text{with } \lambda, \mu > 0$$

then $p_{X,Y}(i,j) = \frac{1}{i!} \lambda^i e^{-\lambda} \quad \frac{1}{j!} \mu^j e^{-\mu}$

$$= p_X(i) \quad p_Y(j) \quad \leftarrow \text{product of 2 Poisson distributions}$$

which means that X and Y are 2 independent Poisson distributions.

<u>Thm :</u> If $X_1, \dots, X_n$ are independent discrete valued random variables, then

$$\mathbb{E}(\underbrace{X_1 X_2 \dots X_n}_{\text{the product, which is a random variable}}) = \prod_{j=1}^{n} \mathbb{E}(X_j) \quad \swarrow \text{product of expectations}$$

$$= \sum_{j} \sum_{x_j \in X_j(\Omega)} x_1 x_2 \dots x_n \, p_{X_1 \dots X_n}(x_1, \dots, x_n)$$

⚠ The converse is no true, namely the above equality does not imply that the random variables are independent.

<u>Remarks</u> : 1) The converse is true if we consider more functions, namely : X, Y are independent discrete valued random variables if and only if

$$\mathbb{E}(g(X)f(Y)) = \mathbb{E}(f(X))\,\mathbb{E}(g(Y)) \qquad \forall f,g : \mathbb{R} \to \mathbb{R}$$

whenever these expectations exist absolutely.

2) We can have that (X_i, X_j) are independent for all $i,j \in \{1, \dots, n\}$, but this does not imply that $X_1, \dots, X_n$ are independent, see [GW] p 48 Ex 3.6.2.

Consider X, Y discrete valued random variables. Since $Z := X+Y$ is also a discrete valued random variable (take $g : \mathbb{R}^2 \to \mathbb{R}$, $g(x,y) = x+y$) we can compute

⚠ in the next computation we do not assume that X, Y are independent

$$p_Z(z) := \mathbb{P}(Z=z)$$

$$= \mathbb{P}(\{\omega \in \Omega \mid X(\omega) + Y(\omega) = z\})$$

$$= \mathbb{P}(\{\omega \in \Omega \mid Y(\omega) = z - \underbrace{X(\omega)}_{\text{call it } x}\})$$

$$= \mathbb{P}\Big(\bigcup_{x \in X(\Omega)} \{\omega \in \Omega \mid X(\omega) = x \text{ and } Y(\omega) = z - x\}\Big)$$

Since these sets have empty intersections

$$= \sum_{x \in X(\Omega)} \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = x \text{ and } Y(\omega) = z - x\})$$

$$= \sum_{x \in X(\Omega)} \mathbb{P}(X = x, Y = z - x).$$

Now if X and Y are independent one gets:

<u>Thm</u> (Convolution formula) If X and Y are independent discrete valued random variables, then

$$\mathbb{P}(X+Y=z) = \sum_{x \in X(\Omega)} \mathbb{P}(X=x)\,\mathbb{P}(Y=z-x)$$

which also reads

$$p_{X+Y}(z) = \sum_{x \in X(\Omega)} p_X(x)\, p_Y(z-x). \quad (*)$$

One aplication of this formula for $X(\Omega) = \{0,1\}$ is given in the next exercise.

Exercise :

Show that the sum of n independent random variable having the Bernoulli distribution with parameter p is a binomial distribution with parameter n and p.

with Bernoulli dist.

Proof by induction : $\qquad Z = \sum_{i=1}^{n} X_i \qquad X_i(\Omega) = \{0,1\}$

$n=2$: $\mathbb{P}(Z=0) = \mathbb{P}(X_1=0)\,\mathbb{P}(X_2=0) + \mathbb{P}(X_1=1)\,\underbrace{\mathbb{P}(X_2=-1)}_{=0}$

we use ⊛

$$= \mathbb{P}(X_1=0)\,\mathbb{P}(X_2=0) = (1-p)^2 =: q^2$$

$$\mathbb{P}(Z=1) = \mathbb{P}(X_1=0)\,\mathbb{P}(X_2=1) + \mathbb{P}(X_1=1)\,\mathbb{P}(X_2=0) = 2pq$$

$$\mathbb{P}(Z=2) = \mathbb{P}(X_1=0)\,\underbrace{\mathbb{P}(X_2=2)}_{=0} + \mathbb{P}(X_1=1)\,\mathbb{P}(X_2=1) = p^2$$

binomial (2,p)

which gives $\mathbb{P}(Z=k) = \dfrac{2!}{k!\,(2-k)!}\,p^k\,q^{2-k}$

Assume the statement is true for $Y = \sum_{i=1}^{n-1} X_i$ and let us check it for $Z = Y + X$. One has

we use ⊛

$$\mathbb{P}(Z=k) = \mathbb{P}(X+Y=k) = \mathbb{P}(X=0)\,\mathbb{P}(Y=k) + \mathbb{P}(X=1)\,\mathbb{P}(Y=k-1)$$

$$= q\,\frac{(n-1)!}{k!\,(n-1-k)!}\,p^k\,q^{n-1-k} + p\,\frac{(n-1)!}{(k-1)!\,(n-1-k+1)!}\,p^{k-1}\,q^{n-1-k+1}$$

$$= \frac{(n-1)!}{k!(n-1-k)!} p^k q^{n-k} + \frac{(n-1)!}{(k-1)!(n-k)!} p^k q^{n-k}$$

$$= \frac{n!}{k!(n-k)!} p^k q^{n-k} \Big(\underbrace{\frac{n-k}{n} + \frac{k}{n}}_{=1} \Big)$$

$$= \frac{n!}{k!(n-k)!} p^k q^{n-k} = \text{binomial}(n, p).$$

Thus the statement has been proved for any n.

We introduce a special type of discrete valued random variables:

<u>Def</u> For any $A \in \mathcal{F}$, the <u>indicator function of the event A</u> is the discrete valued random variable $\mathbb{1}_A$ defined by $\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$.

$\mathbb{1}_A$ takes only 2 values: 0 and 1

Observe then that $\mathbb{E}(\mathbb{1}_A) = \sum_{x \in \mathbb{1}_A(\Omega)} x \, \mathbb{P}(\mathbb{1}_A = x)$

$= \mathbb{P}(\mathbb{1}_A = 1) = \mathbb{P}(\{\omega \in \Omega \mid \omega \in A\}) = \mathbb{P}(A)$.

Also $\mathbb{1}_{A \cap B} = \mathbb{1}_A \mathbb{1}_B$ and $\mathbb{1}_A + \mathbb{1}_{A^c} = 1$.

check

constant function 1

check

Finally $\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_{A\cap B}$ from which one deduces $\mathbb{P}(A\cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A\cap B)$.

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

Exercise : For $A, B \in \mathcal{F}$, A and B are independent if and only if $\mathbb{1}_A$ and $\mathbb{1}_B$ are independent random variables

$\mathbb{P}(\mathbb{1}_A = x, \mathbb{1}_B = y) = \mathbb{P}(\mathbb{1}_A = x)\,\mathbb{P}(\mathbb{1}_B = y)$.

Exercise : 3.37 p 46, very instructive.

IV Probability generating function

very useful tool.

Consider a sequence of numbers $\{\alpha_j\}_{j\in\mathbb{N}} \subset \mathbb{R}$.

One way to encode the information about this sequence is to define the power series

$f(x) := \sum_{j\in\mathbb{N}} \alpha_j x^j$. If this series converges for $|x|$ small enough, then $\alpha_j := \frac{1}{j!} f^{(j)}(0)$
$\uparrow$ jth derivative

Another option is to define $g(x) := \sum_{j\in\mathbb{N}} \frac{1}{j!} \alpha_j x^j$,

and then $\alpha_j := g^{(j)}(0)$.

Thus, if the radius of convergence of these power series is strictly positive, then the relation between the sequence and the power series is bijective.

Consider now a discrete valued random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $X(\Omega) = \{x_0, x_1, x_2, \dots\} \subset \mathbb{R}$.

$\uparrow$ it is a countable set

Set $p_j = \mathbb{P}(X = x_j) \equiv \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = x_j\})$.

One has $p_j \geq 0$ and $\sum_j p_j = 1$.

Def: In the framework introduced above, the probability generating function of X is the function $G_X : (-R, R) \to \mathbb{R}$, with

$$G_X(s) := \sum_{j=0} p_j s^j ,$$

the upper bound can be ∞ or a finite number if $X(\Omega)$ is finite

where R is the radius of convergence of the series.

Remarks: 1) $R \geq 1$ since for $|s| \leq 1$ one has $\sum_{j=0} |p_j s^j| \leq \sum_{j=0} p_j = 1 < \infty$,

2) $G_X(0) = p_0$ and $G_X(1) = 1$

Thm (uniqueness) Let X, Y be 2 discrete valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, then $G_X = G_Y$ if and only if $X = Y$.

if we consider on the maximal domain of definition.

Back to the examples of chapter II : integer valued random variable

Here $X(\Omega) = \{x_0, x_1, \dots\}$ with $x_j = j$.

Bernoulli distribution : $G_X(s) := p_0 s^0 + p_1 s^1 = q + ps$

always $q := 1-p$ $\qquad = \underline{(1-p) + ps}$.

Binomial distribution : $G_X(s) := \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} s^j$

$= \underline{(q+ps)^n}$.

Poisson distribution : $G_X(s) := \sum_{j=0}^{\infty} \frac{1}{j!} \lambda^j e^{-\lambda} s^j$

$= e^{-\lambda} e^{\lambda s} = \underline{e^{\lambda(s-1)}}$.

Geometric distribution : $G_X(s) = \sum_{j=1}^{\infty} p q^{j-1} s^j = sp \sum_{j=0}^{\infty} (sq)^j$

$= \underline{\frac{sp}{1-sq}}$ if $|s| < \frac{1}{q}$.

The probability generating function is a convenient tool for computing several quantities related to the random variable X.

Def: For $k \in \mathbb{N}^*$, the <u>k-moment</u> of the random variable X is given by $E(X^k)$, if it exists.

k^{th} derivative

Remark: For any $k \in \mathbb{N}$, $G_X^{(k)}(s)$ always exists for $s < 1$ since all derivatives of a power series always exist inside the radius of convergence.
However, the limit $\lim_{s \nearrow 1} G_X^{(k)}(s)$ might or might not exist, it depends on G_X.

Thm For an integer valued random variable X (i.e. $x_j = j$) and for $k \in \mathbb{N}^*$, one has

$$\lim_{s \nearrow 1} G_X^{(k)}(s) = \mathbb{E}\big(X(X-1)\dots(X-k+1)\big)$$

we accept that both can be ∞.

Examples: • $\lim_{s \nearrow 1} G'_X(s) = \mathbb{E}(X)$

$$\mathbb{E}(X^2) = \mathbb{E}(X(X-1) + X) = \lim_{s \nearrow 1} \left(G''_X(s) + G'_X(s)\right)$$

if there limit exist

Proof : For $|s|<1$

$$G_X^{(k)}(s) = \frac{d^k}{ds^k} \sum_{j=0} p_j s^j = \sum_{j=k} j(j-1)\dots(j-k+1) p_j s^{j-k}$$

$$= \sum_{j=k} \underbrace{j(j-1)\dots(j-k)}_{=: g_k(j)} s^{j-k} \mathbb{P}(X=j)$$

$$= \mathbb{E}\left(g_k(X) s^{X-k} \right).$$

Thus $\lim_{s \nearrow 1} G_X^{(k)}(s) = \lim_{s \nearrow 1} \mathbb{E}\left(g_k(X) s^{X-k}\right)$

$$= \mathbb{E}(g_k(X))$$

if the limit exits.

One could be a little bit more cautious.

$$= \mathbb{E}(X(X-1)\dots(X-k+1))$$

$\square$

Since both side are positive and increasing with $s \nearrow 1$, one accepts that both can be infinity.

Note that $G_X(s) = \sum_j s^j \mathbb{P}(X=j) = \mathbb{E}(s^X)$.

In addition to the convolution formula for 2 independent discrete valued random variables, one has :

Thm : If X, Y are independent integer valued random variables, then $G_{X+Y}(s) = G_X(s)\, G_Y(s)$.

Proof: $G_{X+Y}(s) = \mathbb{E}(s^{X+Y}) = \mathbb{E}(s^X s^Y)$

$= \mathbb{E}(s^X)\,\mathbb{E}(s^Y) = G_X(s)\,G_Y(s)$. $\square$

by independence

More generally, if $X_1, \dots, X_n$ are independent integer valued random variables, then

$$G_{X_1 + \dots + X_n} = G_{X_1} G_{X_2} \cdots G_{X_n}.$$

And what happens if n itself is random?

Thm (Random sum formula). Let $N, X_1, X_2, \dots$ be independent integer valued random, with all X_j identically distributed, and let G_X denote the common probability generating function. Then the sum $S = X_1 + X_2 + \dots + X_N$ has probability generating function $s \to G_S(s) = G_N(G_X(s))$.

Proof :

$$G_S(s) = \mathbb{E}(s^S) = \mathbb{E}(s^{X_1+X_2+\dots+X_N})$$

$$= \sum_{n=0} \mathbb{E}(s^{X_1+\dots+X_N} \mid N=n)\ \mathbb{P}(N=n)$$

partition thm, p 21 of the notes

$$= \sum_{n=0} \mathbb{E}(s^{X_1+\dots+X_n})\ \mathbb{P}(N=n)$$

independence

$$= \sum_{n=0} \mathbb{E}(s^X)^n\ \mathbb{P}(N=n)$$

$$= \sum_{n=0} G_X(s)^n\ \mathbb{P}(N=n)$$

$$= G_N(G_X(s)). \qquad \square$$

Aplication : One has

$$G_S(s)' = G_N(G_X(s))' = G_N'(G_X(s))\ G_X'(s)$$

$$\Rightarrow G_S'(1) = G_N'(\underbrace{G_X(1)}_{=1})\ G_X'(1) = G_N'(1)\ G_X'(1)$$

$$\Rightarrow \underline{\mathbb{E}(S) = \mathbb{E}(N)\ \mathbb{E}(X)}.$$

<u>Example</u> : One room contains 20 independent student, each staying with a probability p and leaving with a probability q. Each has a random number of electronic devices following a Poisson distribution of parameter 2. What is the probability generating function of the total number of electronic devices in the room at the end of the lecture ?

The number N of student remaining in the room follows a binomial distribution, and thus

p 33

$G_N(s) = (q + ps)^{20}$. Let $X_i \sim \text{Poisson}(2)$ be the number of electronic devices for student i. Then

p 33

$G_X(s) = e^{2(s-1)}$. If $S = X_1 + X_2 + \dots + X_N$, then

$$G_S(s) = G_N(G_X(s)) = (q + pG_X(s))^{20}$$

$$= (q + pe^{2(s-1)})^{20}.$$

Everything can be computed from this expression.

V Distribution function and density function

In this chapter we consider more general random variables, namely "discrete valued" will be dropped. For such random variables, the set $\{X = x\}$ will often have no weight, we shall replace it by $\{X \leq x\}$. It means that the probability mass function is mainly for discrete valued random variables.

<u>Def</u> A <u>random variable</u> X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that

$$\{X \leq x\} \equiv \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}, \quad \forall x \in \mathbb{R}.$$

<u>Remark</u>: If X is discrete valued, the conditions $\{X = x\} \in \mathcal{F}$ $\forall x \in \mathbb{R}$ and $\{X \leq x\} \in \mathcal{F}$ $\forall x \in \mathbb{R}$ are equivalent (a proof?), and thus the above definition includes discrete valued random variables.

Similar to P_X we have:

Def : For any random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$, the distribution function, or cumulative distribution function, is the function $F_X : \mathbb{R} \to [0,1]$ given by $F_X(x) := \mathbb{P}(X \leq x) \equiv \mathbb{P}(\underbrace{\{\omega \in \Omega \mid X(\omega) \leq x\}}_{\in \mathcal{F}, \text{ by assumption on } X})$ $\forall x \in \mathbb{R}$.

Clearly, the function F_X is increasing (1) (but not always strictly increasing) since $F_X(x) \leq F_X(y)$ if $x < y$.

Also $\lim\limits_{x \to -\infty} F_X(x) = 0$ (2)

$\lim\limits_{x \to \infty} F_X(x) = 1$ (3)

} a proof ?

and $F_X(X+\varepsilon) \xrightarrow{\varepsilon \searrow 0} F_X(x)$ (4) (this is called continuity from the right)

For property (4), consider for any $n \in \mathbb{N}^*$,

$A_n := \{\omega \in \Omega \mid X(\omega) > x + \frac{1}{n}\}$.

Then $\{A_n\}_{n\in\mathbb{N}^*}$ satisfies $A_n \subset A_{n+1}$ (increasing family), and $A_n^c = \{\omega\in\Omega \mid X(\omega) \leqslant x + \frac{1}{n}\}$, and $\bigcup_{n\in\mathbb{N}^*} A_n = \{\omega\in\Omega \mid X(\omega) > x\}$.

Thus $\lim_{n\to\infty} F_X(x+\frac{1}{n}) = \lim_{n\to\infty} \mathbb{P}(A_n^c) = \lim_{n\to\infty}(1-\mathbb{P}(A_n))$

$= 1 - \lim_{n\to\infty}\mathbb{P}(A_n) \underset{\text{Notes, page 9}}{=} 1 - \mathbb{P}(\bigcup_{n\in\mathbb{N}^*} A_n)$

$= 1 - \mathbb{P}(X > x) = \mathbb{P}(X\leq x) = F_X(x).$ ⌋

<u>Remarks</u>: 1) $\mathbb{P}(a < X \leq b) = \mathbb{P}(X\leq b) - \mathbb{P}(X\leq a)$

$= F_X(b) - F_X(a)$, and this could be different from $\mathbb{P}(a \leq X \leq b)$ in general. ⚠

2) We have seen (Notes p. 14) that a probability mass function determined entirely a discrete valued random variable. The same is true for the cumulative distribution function, namely:

For any function $F : \mathbb{R} \to [0,1]$ satisfying ① - ④, one can associate a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable X such that

$F(x) = \mathbb{P}(X \leq x)$.

Def: A random variable X is (absolutely) continuous if there exists $f_X : \mathbb{R} \to [0, \infty)$ "measurable" such that $F_X(x) = \int_{-\infty}^{x} f_X(y)\, dy \quad \forall x \in \mathbb{R}$

f_X is called the probability density function.

(the notion of measurability gives a meaning to)

Remark: If X is (a.) continuous, then

$\mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a \leq X \leq b)$

$= \mathbb{P}(a < X < b) = \int_a^b f_X(y)\, dy.$

Thus, any $f : \mathbb{R} \to [0, \infty)$, measurable and satisfying $\int_{\mathbb{R}} f(y)\,dy = 1$ defines a distribution function F by $F(x) := \int_{-\infty}^{x} f(y)\,dy$.

The conditions ① - ④ are automatically satisfied.

Several examples of probability density function are provided in an appendix.

For a discrete valued random variable X, $g(X)$ is also a discrete valued random variable, for any $g : \mathbb{R} \to \mathbb{R}$. What about the general case?

In this situation, $g(X)$ is a random variable if $\{\omega \in \Omega \mid g(X(\omega)) \leq x\} \in \mathcal{F}$, $\forall x \in \mathbb{R}$.

This is satisfied if g is continuous, and more generally is g is Borel measurable.

this notion belongs to measure theory, we do not discuss it.

Another natural question is about the (a.) continuity of $g(X)$ if X is (a.) continuous ?

One answer is provided below, but it is not the most general situation.

<u>Thm</u>: Let X be an (a.) continuous random variable with probability density function f_X. Let $g : \mathbb{R} \to \mathbb{R}$ be strictly increasing and differentiable.

↳ ⇒ invertible

Then $Y := g(X)$ is an (a.) continuous random variable with probability density function

$$f_Y(y) = f_X(g^{-1}(y))\,(g^{-1})'(y).$$

<u>Proof</u>: Since $\mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) \overset{\text{because of invertibility}}{=} \mathbb{P}(X \leq g^{-1}(y))$

$= \int_{-\infty}^{g^{-1}(y)} f_X(z)\,dz$ one infers that

$f_Y(y) = \frac{d}{dy}\mathbb{P}(Y \leq y) = f_X(g^{-1}(y))\,(g^{-1})'(y)$. In other words

$$\mathbb{P}(Y \leq y) = \int_{-\infty}^{y} f_X(g^{-1}(x))\,(g^{-1})'(x)\,dx = \int_{-\infty}^{g^{-1}(y)} f_X(z)\,dz$$

same expression

$\uparrow z = g^{-1}(x)$, $dz = (g^{-1})'(x)\,dx$

$\square$

Note that strict monoticity of g is not really necessary, as shown here:

<u>Example:</u> Consider X, f_X as before, and $g: \mathbb{R} \to \mathbb{R}$, $g(x) = x^2$. Then for $y \geq 0$

$$\mathbb{P}(g(X) \leq y) = \mathbb{P}(X^2 \leq y) =$$

$$= \mathbb{P}(\{\omega \in \Omega \mid X(\omega)^2 \leq y\})$$

$$= \mathbb{P}(\{\omega \in \Omega \mid -\sqrt{y} \leq X(\omega) \leq \sqrt{y}\})$$

$$= \int_{-\sqrt{y}}^{\sqrt{y}} f_X(z)\, dz$$

$$\Rightarrow \quad f_Y(y) = \frac{d}{dy} \mathbb{P}(g(X) \leq y)$$

$$= f_X(\sqrt{y}) \frac{1}{2} \frac{1}{\sqrt{y}} - (-) f_X(-\sqrt{y}) \frac{1}{2} \frac{1}{\sqrt{y}}$$

$$= \frac{1}{2} \frac{1}{\sqrt{y}} \left(f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right).$$

As in the discrete case, in the (a.) continuous case we set:

Mean : $\mathbb{E}(X) = \int_{\mathbb{R}} x \, f_X(x) \, dx$

whenever it converges absolutely

For suitable function g (whenever $g(X)$ is an (a.) continuous random variable) one has

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}} g(x) \, f_X(x) \, dx$$

whenever it converges absolutely.

Variance : $\operatorname{var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right)$

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

whenever this expression is well defined.

Examples (based on the density functions provided in annex).

1) Uniform distribution on (a,b) : $\mathbb{E}(X) = \frac{1}{b-a}\int_a^b x \, dx$

$$= \frac{1}{b-a} \frac{1}{2}(b^2 - a^2) = \frac{1}{2}(a+b).$$

2) Exponential distribution : $E(X) = \int_0^{\infty} x \lambda e^{-\lambda x} \, dx$

$= \int_0^{\infty} \frac{1}{\lambda} y e^{-y} \, dy = \frac{1}{\lambda}$. Also

$E(X^2) = \int_0^{\infty} x^2 \lambda e^{-\lambda x} \, dx = \dots = \frac{2}{\lambda^2}$.

3) Normal distribution $N(\mu, \sigma^2)$:

$\mathbb{E}(X) = \dots = \mu$, $var(X) = \sigma^2$

4) Cauchy distribution : $\mathbb{E}(X) = \frac{1}{\pi} \int_{\mathbb{R}} x \frac{1}{1+x^2} \, dx$

$= ?$ this expression is not absolutely integrable!

The mean of the Cauchy distribution does not exist.

For a nice aplication, look at the example in [GW] p 77 : The needle of George Louis Leclerc, comte de Buffon !

VI Multivariate distributions and independence

We investigate multivariate distributions, like in Chapter III, but for general random variables. The framework is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and n random variables $X_1, \dots, X_n$.

Def: The joint probability is the map

$F_{X_1,\dots,X_n} : \mathbb{R}^n \to [0,1]$ given by

$$F_{X_1,\dots,X_n}(x_1,\dots,x_n) := \mathbb{P}(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n).$$

Clearly: $\lim\limits_{x_1 \to -\infty} \dots \lim\limits_{x_n \to -\infty} F_{X_1,\dots X_n}(x_1,\dots,x_n) = 0$

and $\lim\limits_{x_1 \to \infty} \dots \lim\limits_{x_n \to \infty} F_{X_1,\dots X_n}(x_1,\dots,x_n) = 1$.

Also $F_{X_j}(x_j) = \underbrace{\lim\limits_{x_1 \to \infty} \dots \lim\limits_{x_n \to \infty}}_{j \text{ omitted}} F_{X_1,\dots,X_n}(x_1,\dots,x_n)$

marginal distribution function.

Remark: One gets the joint probability for less random variables by considering $x_j \to \infty$ for the other variables.

<u>Def</u>: The random variables $X_1, \dots, X_n$ are <u>independent</u> if $F_{X_1,\dots,X_n} = \prod_{j=1}^{n} F_{X_j}$, or equivalently if $\mathbb{P}(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{j=1}^{n} \mathbb{P}(X_j \leq x_j)$.

We say that the joint probability factorizes.

<u>Remark</u>: We can use the notation adopted from $\mathbb{R}^n$, namely $\underline{X} = (X_1, \dots, X_n)$, $\underline{x} = (x_1, \dots, x_n)$,

underlined, or bold

and then $F_{\underline{X}}(\underline{x}) = F_{X_1,\dots X_n}(x_1, \dots, x_n)$.

<u>Exercise</u>: Check that $X_1, \dots, X_n$ are independent random variables if and only if

$$F_{\underline{X}}(\underline{a} < \underline{x} \leq \underline{b}) = \prod_{j=1}^{n} F_{X_j}(a_j < x_j \leq b_j), \quad \forall \underline{a}, \underline{b} \in \mathbb{R}^n$$

↑ interpretation componentwise.

<u>Def</u>: $X_1, \dots, X_n$ are <u>jointly (a.) continuous</u> if

$$F_{\underline{X}}(\underline{x}) = \int_{-\infty}^{x_1} dy_1 \cdots \int_{-\infty}^{x_n} dy_n \, f_{\underline{X}}(y_1, \dots, y_n)$$

for some $f_{\underline{X}} : \mathbb{R}^n \to [0, \infty)$, measurable.

↖ joint probability density function

Remarks: 1) Sometimes one writes $f_{\underline{X}} = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F_{\underline{X}}$ but this is correct only if one knows that $X_1, \dots, X_n$ are jointly (a.) continuous, otherwise it is not correct.

2) If $X_1, \dots, X_n$ are jointly (a.) continuous, then $F_{\underline{X}}(\underline{a} < \underline{x} \leq b) = \int_{a_1}^{b_1} dy_1 \dots \int_{a_n}^{b_n} dy_n\, f_{\underline{X}}(y_1, \dots y_n)$

and $\leq$ can also be used, one gets the same quantity.

This can be generalized to any (measurable) subset of $\mathbb{R}^n$: $\mathbb{P}(\underline{X} \in A) = \int \dots \int_A dy_1 \dots dy_n\, f_{\underline{X}}(y_1, \dots y_n)$.

measurable

3) Any $f: \mathbb{R}^n \to [0, \infty)$ with $\int \dots \int_{\mathbb{R}^n} dy_1 \dots dy_n\, f(y_1, \dots y_n) = 1$ defines a joint probability density function for some random variables $X_1, \dots, X_n$ on a prob. space.

Observe that if $X_1, \dots X_n$ are jointly (a.) continuous,

then $f_{X_j}(x_j) = \frac{d}{dx_j} F_{X_j}(x_j)$

p 48 $\rightarrow$ $= \frac{d}{dx_j} \underbrace{\lim_{x_1 \to \infty} \dots \lim_{x_n \to \infty}}_{x_j \text{ omitted}} F_{\underline{X}}(\underline{x})$

$= \frac{d}{dx_j} \int_{-\infty}^{\infty} dy_1 \dots \int_{-\infty}^{x_j} dy_j \dots \int_{-\infty}^{\infty} dy_n \; f_{\underline{X}}(y_1, \dots, y_n)$

$= \int_{-\infty}^{\infty} dy_1 \underset{j \text{ omitted}}{\dots} \int_{-\infty}^{\infty} dy_n \; f_{\underline{X}}(y_1, \dots, x_j, \dots, y_n)$

Thus, the <u>marginal density function</u> can be obtained from $f_{\underline{X}}$ by integrating over $\mathbb{R}$ the other variables.

One can also prove: one proof?

<u>Thm</u>: Let $X_1, \dots, X_n$ be jointly (a.) continuous random variables. They are independent if and only if $f_{\underline{X}} = \prod_{j=1}^{n} f_{X_j}$ the joint probability density function factorizes.

Let us now consider the sum of two (a.) continuous random variables, as done on p. 26 in the discrete case.

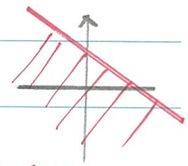
Consider X, Y two jointly (a.) continuous random variables, with joint probability density function $f_{X,Y}$.

Set $Z := X + Y$ which is again a random variable.

z is a fixed real number

Then $P(Z \leq z) = \mathbb{P}(X + Y \leq z)$

$$= \iint_A dx\, dy\, f_{X,Y}(x,y) \qquad A := $$



with $A := \{(x,y) \in \mathbb{R}^2 \mid x + y \leq z\}$

$$= \int_{-\infty}^{\infty} dx \int_{-\infty}^{z-x} dy\, f_{X,Y}(x,y)$$

change of variable $y' := y + x$

$$= \int_{-\infty}^{\infty} dx \int_{-\infty}^{z} dy'\, f_{X,Y}(x, y' - x)$$

$$= \int_{-\infty}^{z} dy \left[\int_{-\infty}^{\infty} dx\, f_{X,Y}(x, y - x) \right]$$

$$\Rightarrow \mathbb{P}(Z \leq z) = \int_{-\infty}^{z} f_Z(y)\, dy \quad \text{with } f_Z(y) = \int_{-\infty}^{\infty} dx\, f_{X,Y}(x, y-x).$$

<u>Thm</u> (convolution formula): If X, Y are independent (a.) continuous random variables, then $Z := X + Y$ is an (a.) continuous random variable, with $f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx$

also written $f_Z = f_X * f_Y$ convolution, $\forall z \in \mathbb{R}$.

Exercise : Use the previous formula to check that

1) if X, Y are independent random variable having normal distribution $N(0,1)$, then $X+Y$ has normal distribution $N(0,2)$

mean variance

2) if X, Y are independent random variables having a chi-squared distribution of degree n and m respectively, then $X+Y$ is a chi-squared distribution of degree $n+m$.

Consider $X_1, \dots, X_n$ jointly (a.) continuous random variables, and let $\phi : D^{\subset \mathbb{R}^n} \to R^{\subset \mathbb{R}^n}$ a diffeomorphism of class C^1, with

bijective, of class C^1, and with inverse of class C^1.

$D := \overline{\{ \underline{x} \in \mathbb{R}^n \mid f_{\underline{X}}(\underline{x}) > 0 \}}$ and

the closure of the set on which $f_{\underline{X}}$ is not 0.

$R := \underline{\phi(D)}$

image of D by ϕ

We consider new random variables, a function of the initial random variables.

Set $(Y_1, \dots, Y_n) := \phi(X_1, \dots, X_n)$, then

$Y_1, \dots, Y_n$ are jointly (a.) continuous, and

$$f_{Y_1,\dots Y_n}(y_1, \dots y_n) = \begin{cases} f_{X_1,\dots X_n}(\phi^{-1}(y_1, \dots, y_n)) \det J_{\phi^{-1}}(y_1, \dots, y_n) & \text{if } (y_1, \dots y_n) \in R \\ 0 & \text{if } (y_1, \dots, y_n) \notin R \end{cases}$$

(*)

determinant of Jacobian matrix

with $J_{\phi^{-1}}(y_1, \dots y_n) = \left(\frac{\partial \phi_i^{-1}}{\partial y_j}(y_1, \dots, y_n) \right)$

The equality (*) is called the Jacobian formula.

In simpler words, when one defines new random variables from a family of jointly (a.) continuous random variables, then the new family is also (a.) continuous, and one has an expression for the new joint density.

Look at some simple examples on p. 94-95 of [GW], and do 1 or 2 exercises... ♡

Conditional probability : Let X, Y be jointly (a.) continuous random variables, and suppose that $X \in (x, x+\delta)$ with $\delta > 0$ (partial knowledge).

Then $\mathbb{P}(Y \leq y \mid X \in (x, x+\delta)) =$

$$= \frac{\mathbb{P}(Y \leq y \text{ and } X \in (x, x+\delta))}{\mathbb{P}(X \in (x, x+\delta))}$$ we suppose it $\neq 0$

$$= \frac{\int_{-\infty}^{y} ds \int_{x}^{x+\delta} dt\, f_{X,Y}(t,s)}{\int_{-\infty}^{\infty} ds \int_{x}^{x+\delta} dt\, f_{X,Y}(t,s)} = \frac{\int_{-\infty}^{y} ds \int_{x}^{x+\delta} dt\, f_{X,Y}(t,s) \cdot \frac{1}{\delta}}{\int_{x}^{x+\delta} dt\, f_X(t) \cdot \frac{1}{\delta}}$$

$$\xrightarrow{\delta \searrow 0} \frac{\int_{-\infty}^{y} ds\, f_{X,Y}(x,s)}{f_X(x)}.$$ we suppose it $\neq 0$

Thus one gets the following definition :

Def : In the above framework, the conditional density function of Y given $X = x$, denoted by $f_{Y|X}(\cdot\,|x)$ is given by $$f_{Y|X}(y|x) := \frac{f_{X,Y}(x,y)}{f_X(x)}$$ for any x such that $f_X(x) > 0$.

Clearly, if X and Y are independent, $f_{Y|X}(\cdot|x) = f_Y$ since knowing $X = x$ does not provide any information on Y.

Once more (see p 46 for $n = 1$)

<u>Thm</u> If $X_1, \dots, X_n$ are jointly (a.) continuous, and if $g : \mathbb{R}^n \to \mathbb{R}$ is such that $g(X_1, \dots, X_n)$ is an (a.) continuous random variable, then

$$\mathbb{E}(g(X_1, \dots, X_n)) = \int_{\mathbb{R}^n} g(\underline{x}) \, f_{\underline{X}}(\underline{x}) \, d\underline{x}$$

whenever it converges absolutely.

<u>Remark</u>: The special case $g(x,y) := x + y$ leads to $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ even if X and Y are dependent. On the other hand $g(x,y) = xy$ leads to $\mathbb{E}(XY)$ but this is not equal to $\mathbb{E}(X)\mathbb{E}(Y)$ in general, except if X and Y are independent.

<u>Def :</u> The <u>conditional expectation of Y given $X = x$</u>, denoted by $\mathbb{E}(Y|X=x)$ is given if $f_X(x) > 0$ by

$$\int_{-\infty}^{\infty} dy\, y\, f_{Y|X}(y|x) = \int_{-\infty}^{\infty} dy\, y\, \frac{f_{X,Y}(x,y)}{f_X(x)} .$$

Note then that $\mathbb{E}(Y) = \int_{-\infty}^{\infty} \mathbb{E}(Y|X=x)\, f_X(x)\, dx$.

<u>Possible report</u> : Bivariate normal distribution [GW] p 100

<u>Exercise</u> (related to 1 (a.) continuous random variable, as on page 46) :

If X is (a.) continuous with distribution function F_X and probability density function f_X satisfying $f_X(x) = 0$ for $x < 0$, show that $\mathbb{E}(X) = \int_0^{\infty} (1 - F_X(x))\, dx$.

VII Moments and moment generating functions

VII.1 : Technicalities

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, X a random variable, and $F_X : \mathbb{R} \to [0,1]$ its distribution function $\left(F_X(x) := \mathbb{P}(X \leq x) \right)$.

Thm (~ Lebesgue decomposition thm)

measurable

$\exists !\ F_{pp}, F_{ac}, F_{sc} : \mathbb{R} \to [0,1]$ with

$$F_X = \alpha_{pp} F_{pp} + \alpha_{ac} F_{ac} + \alpha_{sc} F_{sc},$$

$\alpha_{pp}, \alpha_{ac}, \alpha_{sc} \in [0,1]$ and $\alpha_{pp} + \alpha_{ac} + \alpha_{sc} = 1$

and • $F_{pp}(x) := \sum_{y \leq x} p(y)$ a countable sum

pure point part of F and $\sum_{y \in \mathbb{R}} p(y) = 1$

a density function

• $F_{ac}(x) := \int_{-\infty}^{x} f(y)\, dy$ $f : \mathbb{R} \to [0, \infty)$

absolutely continuous part of F and $\int_{-\infty}^{\infty} f(y)\, dy = 1$

singular continuous part of F • $F_{sc} : \mathbb{R} \to [0,1]$ continuous, differentiable almost everywhere, with $F_{sc}'(x) = 0$ whenever it exits, $\lim_{x \to -\infty} F_{sc}(x) = 0$, $\lim_{x \to \infty} F_{sc}(x) = 1$.

Thus, any random variable can be seen as a linear combination of a discrete valued random variable, an (a.) continuous random variable, and a singular continuous random variable

quite rare

Then, if $\varphi : \mathbb{R} \to \mathbb{R}$ is a continuous (or more generally measurable) function, we set

$$\mathbb{E}(\varphi(X)) = \text{"}\lim\text{"} \sum_j \big(F_X(x_{j+1}) - F_X(x_j)\big)\, \varphi(y_j) \quad , \; x_j \leq y_j \leq x_{j+1} \quad ⊛$$

if it converges absolutely, and one ends up with

$$\mathbb{E}(\varphi(X)) = \sum_{x \in X(\Omega)} \varphi(x)\, p_X(x) \quad \text{if } X \text{ is discrete valued}$$

$$\mathbb{E}(\varphi(X)) = \int_{-\infty}^{\infty} \varphi(x)\, f_X(x)\, dx \quad \text{if } X \text{ is (a.) continuous.}$$

If X is singular continuous, there is no other formula except ⊛ for the expectation.

VII. 2 : Moments, variance and covariance

For any $k \in \mathbb{N}^*$, the k^{th} moment of X is defined by $\mathbb{E}(X^k)$, if it exists (special case $g(x) = x^k$)

Remark: The sequence $\{\mathbb{E}(X^k)\}_{k \in \mathbb{N}^*}$ does not determine X uniquely in general, but it does if the power series $\sum_{k \in \mathbb{N}^*} \frac{1}{k!} t^k \, \mathbb{E}(X^k)$ has a strictly positive radius of convergence.

Recall that if $\mu := \mathbb{E}(X)$, then $\mathrm{var}(X) = \mathbb{E}\big((X-\mu)^2\big)$ and one infers that $\mathrm{var}(X) = 0$ iff $\mathbb{P}(X = \mu) = 1$ which means that X is the constant r.v. equal to μ.

Thus, $\mathrm{var}(X)$ gives an idea of the dispersion of X around μ. Observe also that

$\mathrm{var}(aX) = a^2 \, \mathrm{var}(X)$ ...

non-linear, but quadratic

The standard deviation is defined by $\sqrt{\mathrm{var}(X)}$.

If X, Y are random variables, then if everything exists

$$\mathrm{var}(X+Y) = \mathbb{E}\Big(\big(X+Y-\underbrace{\mathbb{E}(X+Y)}_{=\mathbb{E}(X)+\mathbb{E}(Y)}\big)^2\Big)$$

$$= \mathbb{E}\Big(\big((X-\mathbb{E}(X)) + (Y-\mathbb{E}(Y))\big)^2\Big)$$

$$= \mathrm{var}(X) + \mathrm{var}(Y) + 2\underbrace{\mathbb{E}\Big((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\Big)}_{:=\,\mathrm{cov}(X,Y)}$$

directly from covariance

and $\mathrm{cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

In summary, $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\,\mathrm{cov}(X,Y)$

with $\mathrm{cov}(X,Y) = 0$ if X and Y are independent.

Def The correlation coefficient of the random variables X, Y is defined by (whenever it exists)

$$\rho(X,Y) := \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)^{1/2}\,\mathrm{var}(Y)^{1/2}}.$$

Properties :

1) $\rho(aX+b, cY+d) = \rho(X,Y) \quad \forall a,c \neq 0$

2) $-1 \leq \rho(X,Y) \leq 1$ with

$\rho(X,Y) = 0$ if X, Y are independend $\equiv$ uncorrelated

$\rho(X,Y) = 1$ iff $Y = \alpha X + \beta$, with $\alpha > 0$

$\rho(X,Y) = -1$ iff $Y = \alpha X + \beta$, with $\alpha < 0$.

Somebody for the proof? use (and prove) the Cauchy Schwarz inequality $\mathbb{E}(XY)^2 \leq \mathbb{E}(X^2)\,\mathbb{E}(Y^2)$.

VII. 3 Moment generating function and inequalities

Recall that for integer valued random variables, one has defined $G_X(s) := \mathbb{E}(s^X)$ the probability generating function. $= \sum_{j=0} s^j \, \mathbb{P}(X=j)$.

In the general framework, this formula is not well defined, but one sets :

Def.: The moment generating function ($\equiv$ mgf) of a random variable X is the function (whenever it exists)

$$M_X : (-R, R) \to \mathbb{R}, \quad M_X(t) := \mathbb{E}(e^{tX}).$$

Note that $\quad M_X(t) = \sum_{x \in X(\Omega)} e^{tx} \, p_X(\overset{\mathbb{E}(X=x)}{x})$ if X is discrete valued

and $\quad M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) \, dx$ if X is (a.) continuous

Note that formally one has $M_X(t) = \mathbb{E}\left(\sum_{k=0}^{\infty} \frac{1}{k!} (sX)^k \right)$

$$= \sum_{k=0}^{\infty} \frac{1}{k!} s^k \, \mathbb{E}(X^k).$$

Thm If M_X exists on an interval around 0, ($\Leftrightarrow R > 0$), then for any $k \in \mathbb{N}$,

$$\mathbb{E}(X^k) = M^{(k)}(0).$$

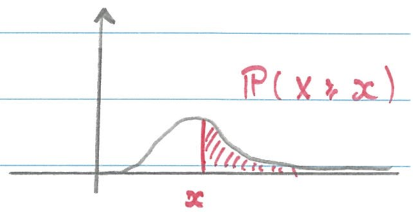
The proof is rather clear from, at least formally.

Note that if M_X exists on some interval around 0, then all moments of X exist, and M_X determines uniquely X.

Other properties are : ↙ Proof of these statements?

- $M_{aX+b}(t) = e^{tb} M_X(at)$,
- $M_{X+Y} = M_X M_Y$ if X and Y are independent,
- If $X \sim N(\mu, \sigma^2)$, then $M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$.

Now, let X be a random variable with $\mathrm{Im}(X) \geq 0$, which means $X(\Omega) \subset [0, \infty)$. Assume also that $\mathbb{E}(X) < \infty$. Then one has :



<u>Thm</u> (Markov inequality)

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}(X)}{x} \qquad \forall x > 0.$$

<u>Proof</u>: Clearly, for any $x \in \mathbb{R}$ and $\omega \in \Omega$ one has

indicator function, see p. 29

$$X(\omega) \geq \begin{cases} x & \text{if } X(\omega) \geq x \\ 0 & \text{if } X(\omega) < x \end{cases} = x \, \mathbb{1}_{A_x}(\omega),$$

where $A_x := \{\omega' \in \Omega \mid X(\omega') \geq x\}$. Thus

inequality between 2 functions

$X \geq x \mathbb{1}_{A_x}$. It then follows that $\mathbb{E}(X) \geq x \, \mathbb{E}(\mathbb{1}_{A_x})$
$= x \, \mathbb{P}(A_x) = x \, \mathbb{P}(X \geq x)$. $\square$

see p 29

As an application, for a random variable X, let us define the median $m \in \mathbb{R}$ which satisfies

$$\mathbb{P}(X < m) \leq \tfrac{1}{2} \leq \mathbb{P}(X \leq m)$$

⚠ m is not unique in general...

Then if $\operatorname{Im}(X) \geq 0$ one has

$$\frac{1}{2} \leq 1 - \mathbb{P}(X < m) = \mathbb{P}(X \geq m) \underset{\text{Markov}}{\leq} \frac{\mathbb{E}(X)}{m}$$

$$\Rightarrow \quad \underline{m \leq 2\,\mathbb{E}(X)}$$

This inequality is satisfied for all positive random variable

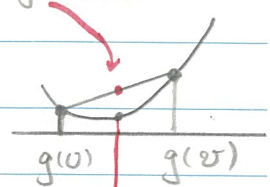
relation between the median and the mean

it is not necessary that g is differentiable

Recall now that a convex function $g : (a,b) \to \mathbb{R}$ satisfies

$$g((1-t)u + tv) \leq (1-t)\,g(u) + t\,g(v)$$

$\forall t \in [0,1]$ and $u, v \in (a,b)$.

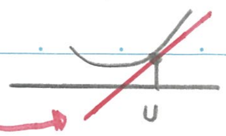


Thm (Supporting tangent thm)

It is not necessary that g is differentiable.

If $g : (a,b) \to \mathbb{R}$ is convex and $u \in (a,b)$, then

$$\exists\, \alpha \in \mathbb{R} \text{ s.t. } \quad g(x) \geq g(u) + \alpha(x-u) \quad \forall\, x \in (a,b)$$



An important and useful result:

Thm: Jensen's inequality

Let X be a random variable such that $\mathbb{E}(X)$ exists, and let $a, b \in \mathbb{R} \cup \{\pm\infty\}$ such that $\mathrm{Im}(X) \subset (a,b)$.

Let $g : (a,b) \to \mathbb{R}$ be measurable, convex and such that $\mathbb{E}(g(X))$ exists. Then $\underline{\mathbb{E}(g(X)) \geq g(\mathbb{E}(X))}$.

Proof: By the previous thm with $v = \mathbb{E}(X) \in (a,b)$, $\exists\, \alpha \in \mathbb{R}$ s.t. $g(x) \geq \underbrace{g(v)}_{cst} + \alpha(x - v)$, $\forall x$

By taking the expectation (check) one gets

$$\mathbb{E}(g(X)) \geq \underbrace{g(v)}_{= g(\mathbb{E}(X))} + \alpha \underbrace{\mathbb{E}(X - v)}_{= 0}$$

$$\Rightarrow \mathbb{E}(g(X)) \geq g(\mathbb{E}(X)). \qquad \square$$

VII.4 The characteristic function

The moment generating function does not exist for so many random variables, let's consider a function which always exists.

Def: The <u>characteristic function</u> ϕ_X of a random variable X is defined by

$$\phi_X(t) := \mathbb{E}(e^{itX}) = \mathbb{E}(\cos(tX)) + i\,\mathbb{E}(\sin(tX))$$

complex analysis comes into play!

Note that $\mathbb{E}(|e^{itX}|) = \mathbb{E}(1) = \text{"}\lim\text{"} \sum_j (F_X(x_{j+1}) - F_X(x_j))\, 1$

$$= \lim_{x \to \infty} F_X(x) = 1.$$

Thus, the expectation defining ϕ_X is absolutely convergent $\Rightarrow$ <u>ϕ_X always exists</u>.

If the moment generating function M_X exists, then

$\phi_X(t) = M_X(it)$. we have to give a meaning to this expression ...

The relation between the moments of X and ϕ_X are :

<u>Thm</u> : If $\mathbb{E}(|X|^N)$ exists for some $N \in \mathbb{N}^*$, then

$$\phi_X(t) = \sum_{k=0}^{N} \frac{1}{k!} (it)^k \, \mathbb{E}(X^k) + o(t^N)$$

for $t \to 0$.

⚠ The fact that ϕ_X always exists does not imply the existence of the moments of X. For example for the Cauchy distribution $\phi_X(t) = e^{-|t|}$. It turns out that this function is not differentiable at 0, and no moment exists.

<u>Properties</u> : • If $a, b \in \mathbb{R}$, $\phi_{aX+b}(t) = e^{itb} \phi_X(at)$

• If X, Y independent, $\phi_{X+Y} = \phi_X \phi_Y$

• ϕ_X determines uniquely F_X (but it is not simple)

• If X is (a.) continuous, $f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_X(t) \, dt$

density — Fourier inversion formula.

VIII The main limit theorems

Whenever something converges, one has to describe in which sense it converges.

Idea: If we throw a die many times, we expect the averge obtained to converge to 3.5. But How and why?

Def: A sequence $Z_1, Z_2, \ldots$ of random variables converges in mean square to the random variable Z_∞ if $\mathbb{E}((Z_n - Z_\infty)^2) \xrightarrow{n \to \infty} 0$.

We write $Z_n \xrightarrow{n \to \infty} Z_\infty$ in mean square.

Thm: (Mean square law of large numbers)

Let $X_1, X_2, \ldots$ be a sequence of iid ($\equiv$ independent and identically distributed) random variables, with mean $\mu < \infty$ and variance $\sigma^2 < \infty$. Then $\frac{1}{n}(X_1 + X_2 + \ldots + X_n) \xrightarrow{n \to \infty} \mu$ in mean square.

discrete r.v. taking only the value μ.

Proof: Set $S_n := \sum_{j=1}^{n} X_j$, then

$$\mathbb{E}\left(\frac{1}{n} S_n\right) = \frac{1}{n}\mathbb{E}(S_n) = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}(X_j) = \frac{1}{n} n\mu = \mu,$$

$$\text{and} \quad \mathbb{E}\left(\left(\frac{1}{n}S_n - \mu\right)^2\right) \overset{\text{because of the previous line}}{=} \operatorname{var}\left(\frac{1}{n}S_n\right)$$

$$= \frac{1}{n^2}\operatorname{var}\left(\sum_{j=1}^{n} X_j\right) \underset{\text{independence}}{=} \frac{1}{n^2}\sum_{j=1}^{n}\operatorname{var}(X_j)$$

$$= \frac{1}{n^2}\, n\, \sigma^2$$

$$= \frac{1}{n}\sigma^2 \xrightarrow{n \to \infty} 0. \qquad \square$$

Let us consider another type of convergence:

Def: A sequence $Z_1, Z_2, \ldots$ of random variables converges <u>in probability</u> to the random variable Z_∞ if $\forall \varepsilon > 0$, $\mathbb{P}(|Z_n - Z_\infty| > \varepsilon) \xrightarrow{n \to \infty} 0$.

We write $Z_n \xrightarrow{n \to \infty} Z_\infty$ in probability.

It turns out that the convergence in mean square is stronger than the convergence in probability, as shown in the next statement:

<u>Proposition</u> : If $Z_n \xrightarrow{n\to\infty} Z_\infty$ in mean square, then $Z_n \xrightarrow{n\to\infty} Z_\infty$ in probability.

Before the proof, let us consider :

<u>Lemma</u> : (Chebyshev's inequality)

If X is a random variable with $\mathbb{E}(X^2) < \infty$, then $\mathbb{P}(|X| \geq x) \leq \frac{\mathbb{E}(X^2)}{x^2}$ $\forall x > 0$.

<u>Proof</u> : Apply Markov's inequality to X^2 :

$$\mathbb{P}(|X| \geq x) = \mathbb{P}(X^2 \geq x^2) \leq \frac{\mathbb{E}(X^2)}{x^2}. \quad \square$$

<u>Proof of proposition</u> : Apply Chebyshev's inequality to $X := Z_n - Z_\infty$ and $\varepsilon > 0$:

$$\mathbb{P}(|Z_n - Z_\infty| > \varepsilon) \leq \mathbb{P}(|Z_n - Z_\infty| \geq \varepsilon) \leq \frac{\mathbb{E}((Z_n - Z_\infty)^2)}{\varepsilon^2}.$$

Since $\mathbb{E}((Z_n - Z_\infty)^2) \xrightarrow{n\to\infty} 0$, one infers that

$$\mathbb{P}(|Z_n - Z_\infty| > \varepsilon) \xrightarrow{n\to\infty} 0. \quad \square$$

Remark : The implication $Z_n \xrightarrow{n\to\infty} Z_\infty$ in probability $\Rightarrow Z_n \xrightarrow{n\to\infty} Z_\infty$ in mean square **does not hold**.

For example, consider the discrete random variables Z_n given by $\mathbb{P}(Z_n = 0) = 1 - \frac{1}{n}$ and $\mathbb{P}(Z_n = n) = \frac{1}{n}$.

Then $Z_n \xrightarrow{n\to\infty} 0$ in probability but not in mean square.

random variable taking only the value 0

Please, check!

Remark: Chebyshev's inequality can also be rewritten as $\mathbb{P}(|X - \mathbb{E}(X)| \geq x) \leq \frac{\mathrm{var}(X)}{x^2}$.

If $X_1, X_2, \ldots$ are iid random variables, with mean μ and variance $\sigma^2 \neq 0$ (but both well defined), then $S_n = X_1 + X_2 + \ldots + X_n \sim n\mu$ (not a precise statement) by the mean square law of large number, but what about $S_n - n\mu$?

We shall work in the <u>standardized version</u> of S_n, namely for $Z_n := \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{var}(S_n)}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$.

Recall also that the normal distribution $N(0,1)$ has density $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

<u>Thm</u>: Central limit theorem

Let $X_1, X_2, \dots$ be iid random variables, with mean μ and variance $\sigma^2 \neq 0$ well defined. Set $S_n := \sum_{i=1}^{n} X_i$ and $Z_n := \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Then for any $x \in \mathbb{R}$,

$$\mathbb{P}(Z_n \leq x) \xrightarrow{n \to \infty} \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy.$$

This is a type of convergence not seen so far...

<u>Def</u>: A sequence $Z_1, Z_2, \dots$ of random variables converges <u>in distribution</u>, or <u>weakly</u>, to Z_∞ if

$$\mathbb{P}(Z_n \leq x) \xrightarrow{n \to \infty} \mathbb{P}(Z_\infty \leq x) \quad \forall x \in C_\infty$$

distribution function

with $C_\infty = \{x \in \mathbb{R} \mid F_{Z_\infty} \text{ is continuous at } x\}$.

We write $Z_n \xrightarrow{n \to \infty} Z_\infty$ weakly.

Thus, the central limit theorem is a weak convergence of $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ to $N(0,1)$ with $C_\infty = \mathbb{R}$.

Remark: If $Z_n \xrightarrow{n\to\infty} Z_\infty$ is probability, then $Z_n \xrightarrow{n\to\infty} Z_\infty$ weakly, and the converse is false in general. However, if Z_∞ is a constant random variable, then the 2 convergences are equivalent.

One use of the characteristic function (apparently a difficult proof)

Thm (Continuity theorem)

Let $Z_1, Z_2, \ldots$ be a sequence of random variables with characteristic functions $\phi_1, \phi_2, \ldots$. Then $Z_n \xrightarrow{n\to\infty} Z_\infty$ weakly if and only if

$\phi_n(t) \xrightarrow{n\to\infty} \phi_\infty(t)$ for all $t \in \mathbb{R}$.

Based on the previous statement one can provide a weaker version of the law of large number, without any information on the variance:

Thm (Weak version of the law of large numbers)

Let $X_1, X_2, \dots$ be a sequence of iid random variables with mean μ. Then $\frac{1}{n}\left(\sum_{j=1}^{n} X_j\right) \xrightarrow{n \to \infty} \mu$ in probability.

Proof: If $U_n := \frac{1}{n} S_n = \frac{1}{n} \sum_{j=1}^{n} X_j$, then (following X)

characteristic function

$$\phi_{U_n}(t) = \left(\phi_{\frac{1}{n}X}(t)\right)^n = \left(\phi_X\left(\tfrac{1}{n}t\right)\right)^n =$$

by independence — by property of ϕ

because $E(|X|)$ exists

$$= \left(1 + \frac{it\mu}{n} + o\left(\frac{t}{n}\right)\right)^n$$

property of exp. function

$$\xrightarrow{n \to \infty} e^{it\mu} = \phi_\mu(t)$$

characteristic function for constant random variable μ

Thus $\phi_{U_n}(t) \xrightarrow{n\to\infty} \phi_\mu(t) \quad \forall t$

by continuity thm

$\Leftrightarrow U_n \xrightarrow{n \to \infty} \mu$ weakly

$\Leftrightarrow U_n \xrightarrow{n \to \infty} \mu$ in probability.

because μ is a cst random variable.

$\square$

Proof of the central limit theorem :

Recall that $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) =: \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} U_i$

with $U_i = X_i - \mu$, $\mathbb{E}(U_i) = 0$, and $\mathbb{E}(U_i^2) = \mathrm{var}(X_i) = \sigma^2$.

Then $\phi_{Z_n}(t) = \phi_{\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n} U_i}(t) = \left(\phi_{\frac{1}{\sigma\sqrt{n}} U}(t)\right)^n$ independence — U satisfying $\mathbb{E}(U) = 0$, $\mathbb{E}(U^2) = \sigma^2$

$$= \left(\phi_U\left(\frac{1}{\sigma\sqrt{n}} t\right)\right)^n$$

$$= \left(1 + 0 - \frac{1}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 \sigma^2 + o\left(\left(\frac{t}{\sqrt{n}}\right)^2\right)\right)^n$$

$$= \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \xrightarrow{n \to \infty} e^{-\frac{t^2}{2}}$$

characteristic function of $N(0,1)$

Thus $\phi_{Z_n}(t) \longrightarrow \phi_{N(0,1)}(t) \quad \forall t$

$\Longleftrightarrow \quad Z_n \xrightarrow{n \to \infty} N(0,1)$ weakly ($\equiv$ in distribution).

continuity theorem

$\square$

Large deviation : Consider $(X_i)_i$ a sequence of iid random variables, with mean μ and variance σ^2.

Then $S_n = \sum_{i=1}^{n}$ has mean $n\mu$ and variance $n\sigma^2$.

The standard deviation of S_n is $\sqrt{n}\,\sigma$.

Thus, it is unlikely that S_n will deviate from $n\mu$ by more than σn^{α} for some $\alpha > \frac{1}{2}$. Such unlikely event are called large deviation.

Consider a random variable X with $\mathbb{E}(X) = 0$ and suppose that $M_X(t) = \mathbb{E}(e^{tX})$ exists for $|t| < \delta$. (moment generating function)

Consider $(X_i)_i$ a sequence of iid random variables following X, and $S_n := \sum_{j=1}^{n} X_j$. Since the function $x \longrightarrow e^{tx}$ is strictly increasing (for any $t > 0$), one infers that $S_n > an$ ($a \in \mathbb{R}_+$)

$\iff e^{tS_n} > e^{tan}$, which implies that

(positive random variable)

$$\mathbb{P}(S_n > an) = \mathbb{P}(e^{tS_n} > e^{tan})$$

$$\leqslant \mathbb{P}(e^{tS_n} \geqslant e^{tan}) \leqslant \frac{\mathbb{E}(e^{tS_n})}{e^{tan}} = \left(\frac{\mathbb{E}(e^{tX})}{e^{ta}}\right)^n$$

(by independence)

(Markov's inequality)

$$= \left(\frac{M_X(t)}{e^{ta}}\right)^n, \quad \forall t \in (0,\delta) \text{ and } a \in \mathbb{R}.$$

Thus, one has obtained

$$\mathbb{P}(S_n \geq an) \leq \left(\frac{M_X(t)}{e^{ta}}\right)^n$$

with the l.h.s. independent of t. It follows that

$$\mathbb{P}(S_n \geq an) \leq \left(\inf_{t\in(0,\delta)} \frac{M_X(t)}{e^{ta}}\right)^n .$$

always > 0

Now, if we set $\Lambda(t) := \ln\left(\overline{M_X(t)}\right)$, one has

$$\frac{M_X(t)}{e^{ta}} = e^{-ta} M_X(t) = e^{-ta} e^{\ln M_X(t)} = e^{-(ta - \Lambda(t))},$$

and then $\inf_t \frac{M_X(t)}{e^{ta}} = \inf_t e^{-(ta - \Lambda(t))}$

$$= e^{-\sup_t (ta - \Lambda(t))} = e^{-\Lambda^*(a)}$$

with $\Lambda^*(a) := \sup_{t\in(0,\delta)} (at - \Lambda(t))$. Legendre transform of Λ

With these definitions one has

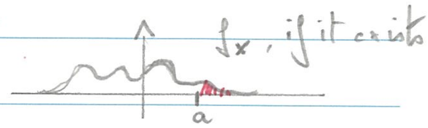
$$\mathbb{P}(S_n \geq an) \leq \left(e^{-\Lambda^*(a)}\right)^n$$

$$\Leftrightarrow \quad \ln(\mathbb{P}(S_n \geq an)) \leq \ln\left(\left(e^{-\Lambda^*(a)}\right)^n\right) = n(-\Lambda^*(a))$$

$$\Leftrightarrow \quad \frac{1}{n}\ln(\mathbb{P}(S_n \geq an)) \leq -\Lambda^*(a).$$

Thm (large deviation theorem)

Let $(X_i)_i$ be a sequence of iid random variables following X, with $\mathbb{E}(X) = 0$ and assume that $M_X(t)$ exists for $|t| < \delta$, $\delta > 0$. Let $a > 0$ such that $\mathbb{P}(X > a) > 0$

(

)

Then $\Lambda^*(a) > 0$ and

$$\frac{1}{n} \ln \left(\mathbb{P}(S_n \geqslant an) \right) \xrightarrow{n \to \infty} -\Lambda^*(a)$$

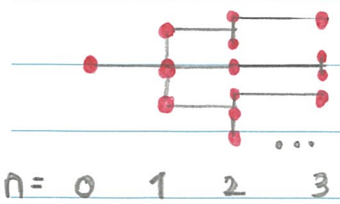
This means that $\mathbb{P}(S_n > an)$ decays to 0 as $e^{-\Lambda^*(a) n}$, as $n \to \infty$.

IX Branching processes

discrete time reproduction process

(So far, time has not played any role. In subsequent applications, it will be taken into account.)

Let us consider some objects, called <u>nomads</u>, living for 1 unit of time and having some children following a distribution $\{p_k\}_{k=0}^{\infty}$. The number of children of each nomad is independent of all other nomads. We start with 1 nomad at time $n=0$.



Such a process is called a <u>branching process</u>.

Let us denote by Z_n the number of nomads at time n. On the picture: $Z_0 = 1$, $Z_1 = 3$, $Z_2 = 6$, ...

Z_n is a integer valued random variable, and there are some natural questions: What is the mean and the variance of Z_n, what is the probability mass function of Z_n (see notes p. 13), what is the probability that mankind is extinct by time n, or will ultimately become extinct?

Clearly, $Z_0 = 1$, $\mathbb{P}(Z_0 = 1) = 1$

$$\mathbb{P}(Z_1 = k) = p_k \quad \text{for any } k$$

but already for Z_2 it is more complicated since it depends on Z_1, and so on ...

Let us denote by C the random variable satisfying $\mathbb{P}(C = k) = p_k$ $\forall k$. The notation C_j is for any random variable following C. Then

$Z_1 = C$, $Z_2 = C_1 + C_2 + \ldots + C_{Z_1}$, and
$Z_n = C_1 + C_2 + \ldots + C_{Z_{n-1}}$. (*) the independence is used here

Recall that a random sum of random variables has been considered in Chapter $\underline{\text{IV}}$ (notes p 36) with the probability generating function. Let us set

$$G_n(s) := \mathbb{E}\left(s^{Z_n}\right) = \sum_{k=0}^{\infty} s^k \, \mathbb{P}(Z_n = k)$$

and

$$G(s) \equiv G_c(s) := \sum_{k=0}^{\infty} s^k \, p_k .$$

<u>Thm</u>: For any $n \in \mathbb{N}^*$ one has

$$G_0(s) = s, \quad G_1(s) = G(s), \quad \text{and} \quad G_n(s) = G_{n-1}(G(s))$$

n^{th} iterate of G $\quad = G(G(G\ldots(G(s))\ldots\ldots)).$

<u>Proof</u>: G_0 and G_1 are quite clear.

The formula $G_n(s) = G_{n-1}(G(s))$ follows from the random sum formula (thm p 36) and ⊛.

By iteration $G_{n-1}(G(s)) = G_{n-2}(G(G(s)))\ldots$

$= \underset{\overset{\shortparallel}{G}}{G_1}(G(G(\ldots(G(s))\ldots))).$ $\square$

Recall that the probability generating function allows us to compute all moments. If we set $\mu = \mathbb{E}(C) = \sum k\, p_k$ (the mean number of children for each nomad), then one has:

<u>Thm</u>: $\mathbb{E}(Z_n) = \mu^n$

<u>Proof</u>: Recall that $\mathbb{E}(Z_n) = G_n'(s)\Big|_{s=1} =$ see p 34

$= G_{n-1}'(G(s))\, G'(s)\Big|_{s=1} = G_{n-1}'(1)\, \mu$ ↑ since $G(1) = 1$

$= \mu\, \mathbb{E}(Z_{n-1})$. Since $\mathbb{E}(Z_1) = \mathbb{E}(C) = \mu$, one deduces that $\mathbb{E}(Z_n) = \mu^n$. $\square$

Thus $\mathbb{E}(Z_n) \xrightarrow{n \to \infty} \begin{cases} 0 & \text{if } \mu < 1 \\ 1 & \text{if } \mu = 1 \\ \infty & \text{if } \mu > 1 \end{cases}$ critical case.

which is not a surprising result!

<u>Exercise</u>: If $\mu = \mathbb{E}(C)$ and $\sigma^2 = var(C)$ show that $var(Z_n) = \begin{cases} \sigma^2 \mu^{n-1} \dfrac{\mu^n - 1}{\mu - 1} & \text{if } \mu \neq 1 \\ n\sigma^2 & \text{if } \mu = 1. \end{cases}$

What about extinction? It is asymptotically the case if $\mu < 1$, but what about $\mu \geq 1$?

Let us set $\mathbf{e} := \mathbb{P}(Z_n = 0 \text{ for some } n \geq 1)$

extinction probability

How to find $\mathbf{e}$?

Remark: We haven't mentioned $(\Omega, \mathcal{F}, \mathbb{P})$ for a long time, because quite often it is not necessary. Here, Ω would correspond to all possible configurations of children, at all generations, and $\mathbb{P}$ could be computed with $\{p_k\}$. Obviously, this basic approach would be possible, but very complicated. However, it is good to keep in mind that it exists.

Let $E_n := \{Z_n = 0\}$ the event that nomads are extinct at the n^{th}-generation, and set $e_n := \mathbb{P}(E_n)$.

Then $\bigcup_n E_n = \{ Z_n = 0 \text{ for some } n \}$, and

observe that $E_n \subset E_{n+1}$ $\forall n$, which implies

that $e_n \leq e_{n+1} \leq 1$. In addition we have

the extinction takes place at any generation.

$$e = \lim_{n \to \infty} e_n$$

Remark: If $p_0 = 0$, then $e = 0$ since children

exist at all generation.

Thm (Extinction probability thm)

The extinction probability e is given by the

smallest non-negative solution of the equation $x = G(x)$.

by def de $G_n = G_{Z_n}$

Proof: Observe that $e_n = \mathbb{P}(Z_n = 0) = G_n(0)$. Then,

since $G_n(s) = G(G(\ldots(G(s))\ldots)) = G(G_{n-1}(s))$,

we infer that $e_n = G_n(0) = G(G_{n-1}(0)) = G(e_{n-1})$

because there is 1 nomad

for any $n = 1, 2, \ldots$, with the initial condition $e_0 = 0$.

Since $e_n \xrightarrow{n \to \infty} e$ and G is continuous on $[0,1]$, one deduces $e = G(e)$.

To show that it is the smallest, assume that $\eta \geq 0$ satisfies $\eta = G(\eta)$. Observe firstly that G is increasing (since all polynomials are increasing on $[0,1]$). Then:

$$e_1 = G_1(0) = G(0) \leq G(\eta) = \eta,$$

$$e_2 = G(e_1) \overset{\text{previous line}}{\leq} G(\eta) = \eta$$

$$\vdots$$

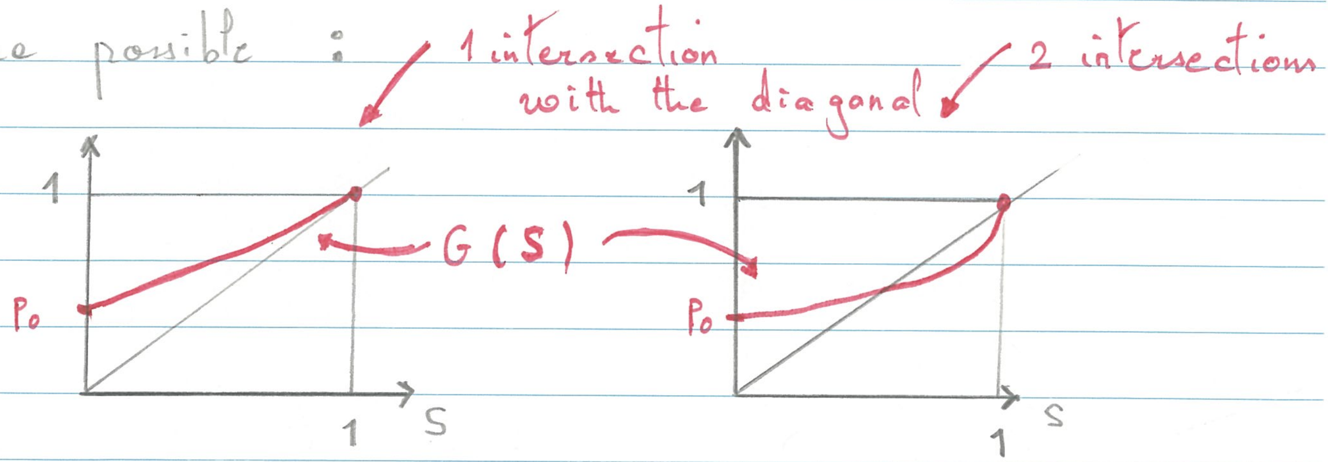
$$e_{n+1} = G(e_n) \leq G(\eta) = \eta \quad \forall n$$

$$\Rightarrow \quad e = \lim_{n \to \infty} e_n \leq \eta. \qquad \square$$

Thm (Extinction / survival thm)

If $p_1 = 1$, then $e = 0$ (boring case). If $p_1 < 1$, then $e = 1$ if $\mu \leq 1$ and $\exists\, e \in [p_0, 1)$ if $\mu > 1$.

Proof: The case $p_1 = 1$ is clear. Thus we suppose $p_1 < 1$. If $p_0 = 0$, then $\mu > 1$ and $e = 0$. Thus the statement is true in this case. Assume then that $p_0 > 0$. Recall also that on $[0,1]$ the function G is continuous, increasing, and convex ($G''(s) \geq 0$). In addition, $G(0) = p_0$ and $G(1) = 1$. Thus 2 pictures are possible:

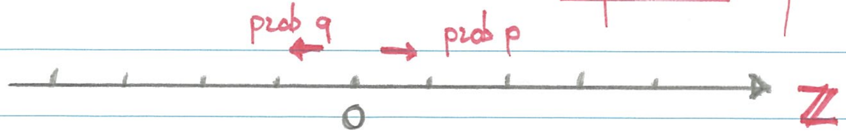


Here ↑, $\mu = G'(1) \leq 1$, and only solution $1 = G(1)$ $\Rightarrow$ $e = 1$.

In the second picture $\mu = G'(1) > 1$, and $\exists!\, x \in (p_0, 1)$ such that $x = G(x)$. Thus, this x is e, which means that $\exists!$ $e < 1$. An extinction is possible even for $\mu > 1$. □

X Random walk (1D)

discrete time, discrete space process



$p + q = 1$

$p, q \geq 0$

For each time $n \in \mathbb{N}$, a particle can either move to the left with probability q, or to the right with probability p, independently of the previous jump or of the position.

Let S_n be the random variable denoting the position of the particle at time n. Then

$$S_{n+1} = \begin{cases} S_n + 1 & \text{with prob. } p \\ S_n - 1 & \text{with prob. } q \end{cases}$$

and $S_n = S_0 + X_1 + X_2 + \dots + X_n$, with

S_0: initial position

$X_1, X_2, \dots, X_n$ independent discrete valued random variable taking the value -1 with prob. q and $+1$ with prob. p. The sequence $S_0, S_1, S_2, \dots$ is called a 1D random walk.

The random walk is symmetric if $p = q = \frac{1}{2}$.

For the next result we denote by

$u_n := \mathbb{P}(S_n = S_0)$ ← prob. that the particle is at the initial position at time n.

<u>Thm</u> One has $u_n = 0$ if n is odd,

$u_{2m} = \binom{2m}{m} p^m q^m \quad \forall m \in \mathbb{N}$

binomial coefficient: $\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (k \leq n)$

<u>Proof:</u> wlog, we assume $S_0 = 0$. For n odd, the result is clear. For $n = 2m$,

$S_{2m} = \sum_{j=1}^{2m} X_j$, and $S_{2m} = 0$ if and only if m of the X_j are equal to $+1$, and m to -1. There are $\binom{2m}{m}$ ways of choosing m (unordered) elements from a set of $2m$ elements. The probability of each of these associated with all +1 associated with all −1

realization is $p^m q^m$. Thus,

$$\mathbb{P}(S_{2m} = S_0) = \binom{2m}{m} p^m q^m. \quad \square$$

In this proof we have done the sum of the probabilities of all paths ending at 0 after 2m jumps.

Let us still look at another approach: observe that $\frac{1}{2}(X_j+1)$ has a <u>Bernoulli distribution</u>, of parameter p, see notes p 16. Since the sum of n independent Bernoulli distributions is a <u>binomial distribution</u> of parameters p and n, see notes page 28, 29, we get that

$$\frac{1}{2}(S_n + n) = \sum_{j=1}^{n} \frac{1}{2}(X_j+1) = \text{binomial}(n,p)$$

$$\Leftrightarrow \quad S_n = 2\,\text{binomial}(n,p) - n. \qquad \text{if}$$

we assume again that $S_0 = 0$,

we fix an arbitrary position

$$\mathbb{P}(S_n = k) = \mathbb{P}(2\,\text{binomial}(n,p) - n = k)$$

$$= \mathbb{P}\left(\text{binomial}(n,p) = \tfrac{1}{2}(n+k)\right)$$

$$= \binom{n}{\frac{1}{2}(n+k)} p^{\frac{1}{2}(n+k)} q^{n-\frac{1}{2}(n+k)}$$

$$= \binom{n}{\frac{1}{2}(n+k)} p^{\frac{1}{2}(n+k)} q^{\frac{1}{2}(n-k)}$$

if $\frac{1}{2}(n+k)$ is an integer between 0 and n

$= 0$ if $\frac{1}{2}(n+k)$ is not an integer between 0 and n.

note that previous theorem is a special case of this result, but we could also have stated the thm more generally.

<u>Def</u>: A random walk is <u>recurrent</u> if it revisits its initial position with prob. 1, or <u>transient</u> otherwise.

this definition also holds in dimension $D \geq 2$, and it is in fact more interesting than the 1D case.

Let us assume that $S_0 = 0$, and observe that $\mathbb{E}(X_j) = -1q + 1p = p - q$

$$\mathrm{var}(X_j) = \mathbb{E}(X_j^2) - \mathbb{E}(X_j)^2 = (-1)^2 q + 1^2 p - (p-q)^2$$

$$= \underset{(p+q)^2}{1} - (p-q)^2 = (p+q)^2 - (p-q)^2 = 4pq < \infty.$$

Thus, by the mean square law of large numbers see notes p 69, one has $\frac{1}{n} S_n \xrightarrow{n \to \infty} p - q$ in mean square. Then, if $p - q > 0$, the particle tends to move to the right, while if $p - q < 0$, it tends to go to the left. If $p = q = \frac{1}{2}$, then $\frac{1}{n} S_n \xrightarrow{n \to \infty} 0$ in mean square.

<u>Thm</u> If $S_0 = 0$,

$$\mathbb{P}(S_n = 0 \text{ for some } n) = 1 - |p - q|.$$

The 1D random walk is recurrent iff $p = q = \frac{1}{2}$.

名古屋大学大学院多元数理科学研究科

In the following proof, we use that there exists an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

<u>Proof</u>: Let $A_n := \{ S_n = 0 \}$ the set of all random walks satisfying $S_n = 0$, and

$B_n := \{ S_n = 0 \text{ but } S_k \neq 0 \ \forall\, 1 \leq k \leq n-1 \}$ the set of random walks returning at 0 at time n for the first time. Then one has

$$\mathbb{P}(A_n) = \sum_{k=1}^{n} \mathbb{P}(A_n \cap B_k)$$

if ever it already passes at 0 before time n

$$= \sum_{k=1}^{n} \mathbb{P}(B_k)\, \mathbb{P}(A_{n-k})$$

because of the independence of each step

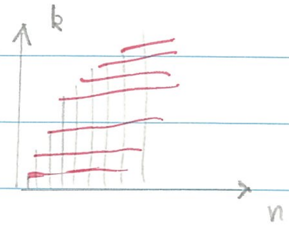
Let us set $u_n := \mathbb{P}(A_n)$ and $f_n := \mathbb{P}(B_n)$, for first

then one has $\boxed{u_n = \sum_{k=1}^{n} f_k \, u_{n-k}.}$ for $n \in \mathbb{N}^*$.

We shall now introduce a generating function :

Consider :

$$\sum_{n=1}^{\infty} u_n s^n = \sum_{n=1}^{\infty} \sum_{k=1}^{n} f_k \, u_{n-k} \, s^n$$



$$= \sum_{n=1}^{\infty} \sum_{k=1}^{n} f_k s^k \, u_{n-k} \, s^{n-k}$$

$$= \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} f_k s^k \, u_{n-k} \, s^{n-k}$$

$$= \underbrace{\sum_{k=1}^{\infty} f_k s^k}_{=: F(s)} \underbrace{\sum_{m=0}^{\infty} u_m s^m}_{U(s)}$$

$$\Rightarrow \quad U(s) - \underset{\uparrow \, u_0 \text{ initial condition}}{1} = F(s)\, U(s)$$

$$\Leftrightarrow \quad F(s) = 1 - \frac{1}{U(s)} .$$

But u_n has been computed in the first thm so $U(s)$ is computable, and we have

$$U(s) = (1 - 4pqs^2)^{-1/2}$$

↑ we don't prove it

$\Rightarrow F(s) = 1 - \sqrt{1 - 4pqs^2}$. Thus $\mathbb{P}(S_n = 0$ for some $n \geq 1)$

$$= \mathbb{P}\left(\bigcup_{j=1}^{\infty} B_j\right) \overset{\text{they are disjoint sets}}{=} \sum_{j=1}^{\infty} f_j = F(1) \quad , \text{ with}$$

$$F(1) = 1 - \sqrt{1 - 4pq} = 1 - \sqrt{(p-q)^2} = 1 - |p - q| . \qquad \square$$

$$1 = (p+q)^2$$

<u>Remark</u>: Consider $p = q = \frac{1}{2}$, and let us set $T := \min\{n \geq 1 \mid S_n = 0\}$

We have just shown that $\mathbb{P}(T < \infty) = 1$, but observe that

$$\mathbb{E}(T) = \sum_{n=1}^{\infty} n f_n = \lim_{s \nearrow 1} \sum_{n=1}^{\infty} (n f_n s^{n-1}) s$$

$$= \lim_{s \nearrow 1} F'(s) = \lim_{s \nearrow 1} -\frac{1}{2}(1-s^2)^{-\frac{1}{2}} (-2s)$$

$$= \infty .$$

It means that a symmetric walk is certain to return to its initial position, but the expected value for the time before it returns is infinite ($\equiv$ larger than any given number).

What about higher dimension ??

Random walks with boundaries

↳ the gambler's ruin problem.

Consider 2 players A, B. A has a \$ and B has $N-a$ \$, i.e. the total capital is $N \geq 1$. A coin is flipped repeatedly, and comes with heads with prob. p, and tails with prob q (always $p+q=1$). At each heads B gives 1\$ to A, and at each tails A gives 1\$ to B. The game ends when either A or B has no more money.

This game corresponds to a random walk on $\{0,1,2,\dots,N\}$, with a start at $a \in \{0,\dots,N\}$ and an end when either 0 or N is reached. 0 and N are called <u>absorbing barriers</u>.

The position of the particle on $\{0, \dots, N\}$ corresponds to the money owned by A : A wins if the particle reached N, and B wins if the particle reaches 0.

Given $a \in \{0, \dots, N\}$ we denote by $\omega(a)$ the probability that A wins ($\equiv$ N is reached).

<u>Thm</u> : $$\omega(a) = \begin{cases} \dfrac{(q/p)^a - 1}{(q/p)^N - 1} & \text{if } p \neq q \\ a/N & \text{if } p = q = \frac{1}{2} \end{cases}.$$

If $N \to \infty$, then A can not win, only B can win (when 0 is reached), but what is the probability than B wins?

<u>Corollary</u> : If $N = \infty$ and $a \in \mathbb{N}$, then

$$\pi(a) := 1 - \omega(a) = \begin{cases} (q/p)^a & \text{if } p > q \\ 1 & \text{if } p \leq q \end{cases}.$$

Proof: If $p \neq q$

$$1 - w(a) = \frac{(q/p)^N - 1 - (q/p)^a + 1}{(q/p)^N - 1} = \frac{(q/p)^N - (q/p)^a}{(q/p)^N - 1}$$

$$\xrightarrow{N \to \infty} \begin{cases} \left(\frac{q}{p}\right)^a & \text{if } q/p < 1 \iff q < p \\ 1 & \text{if } q/p > 1 \iff q > p. \end{cases}$$

If $p = q = \frac{1}{2}$, $1 - w(a) = 1 - \frac{a}{N} \longrightarrow 1$ for $N \to \infty$. □

Proof of the thm: Observe that if $a = 0$ or $a = N$, then the statement is correct ($w(a) = 0$ or $w(a) = 1$). Thus assume that $a \neq 0$ or N.

Let H be the event that the first flip is a heads.

By the partition thm (see notes p9) one has

$$w(a) = \mathbb{P}(\text{A wins}) \quad \text{A has a \$}$$

$$= \mathbb{P}(\underbrace{\text{A wins} \mid H}_{\text{A has now } a+1 \text{ \$}})\,\mathbb{P}(H) + \mathbb{P}(\underbrace{\text{A wins} \mid H^c}_{\text{A has now } a-1 \text{ \$}})\,\mathbb{P}(H^c)$$

$$= w(a+1)\,p + w(a-1)\,q$$

$$\iff p\,w(a+1) - w(a) + q\,w(a-1) = 0. \quad (*)$$

$\circledast$ is a difference equation of second order, with the boundary condition $\omega(0) = 0$ and $\omega(N) = 1$. ~~an~~ there exist technics for solving this equation. Since $\circledast \Leftrightarrow \omega(a+1) - \frac{1}{p}\omega(a) + \frac{q}{p}\omega(a-1) = 0$, one ends up solving $x^2 - \frac{1}{p}x + \frac{q}{p} = 0$

$$x = \frac{\frac{1}{p} \pm \sqrt{\left(\frac{1}{p}\right)^2 - 4\left(\frac{q}{p}\right)}}{2} = \frac{1}{2p}\left(1 \pm \underbrace{\sqrt{1-4pq}}_{=|p-q|}\right)$$

$= |p-q|$ see p 94

$$= \frac{1}{2p}\left(\overset{p+q}{1} \pm |p-q|\right) = 1 \text{ or } \frac{q}{p}.$$

Then $\omega(a) = \begin{cases} \alpha + \beta\left(\frac{q}{p}\right)^a & \text{if } p \neq q \\ \alpha + \beta a & \text{if } p = q = \frac{1}{2} \end{cases}$ for $\alpha, \beta \in \mathbb{R}$.

check that it satisfies $\circledast$

For the boundary conditions $\omega(0) = 0 \Leftrightarrow \begin{cases} \alpha = -\beta & \text{if } p \neq q \\ \alpha = 0 & \text{if } p = q \end{cases}$

and $\omega(N) = 1 \Leftrightarrow \begin{cases} \beta\left(\left(\frac{q}{p}\right)^N - 1\right) = 1 & \text{if } p \neq q \\ \beta N = 1 & \text{if } p = q \end{cases}$

$$\Leftrightarrow \begin{cases} \beta = \dfrac{1}{\left(\frac{q}{p}\right)^N - 1} & \text{if } p \neq q \\ \beta = \frac{1}{N} & \text{if } p = q \end{cases}$$

$$\Rightarrow \omega(a) = \begin{cases} \dfrac{\left(\frac{q}{p}\right)^a - 1}{\left(\frac{q}{p}\right)^N - 1} & \text{if } p \neq q \\ \dfrac{a}{N} & \text{if } p = q. \end{cases}$$

$\square$

XI Random process in continuous time

We shall consider a family of integer valued random variables indexed by $t \in [0, \infty)$: $\{N_t\}_{t \geq 0}$.

Example : Messages received in my letter box

email ≡ line ≡ FB !

Let N_t denote the # messages received up to time t, with the following assumptions :

1) N_t is a random variable taking values in $\mathbb{N}$.

integer valued random variable

2) $N_0 = 0$ initial condition

3) $N_s \leq N_t \quad \forall\, s \leq t$, increasing

4) Independence : for $0 \leq s < t$, the number of messages received in $(s, t]$ is independent of the number received before s,

5) Arrival rate : $\exists\, \lambda > 0$ called the arrival rate such that for h small enough

$$\mathbb{P}(N_{t+h} = n+1 \mid N_t = n) = \lambda h + o(h)$$

$$\mathbb{P}(N_{t+h} = n \mid N_t = n) = 1 - \lambda h + o(h)$$

not equivalent because it does not say anything on $N_{t+h} > n+1$ ...

The last 2 condition are a precise formulation of the statement that the rate of arrival is linear in h, for small h.

With the 2 condition in 5) one can infer the probability of receiving 2 or more messages in the interval $(t, t+h)$:

$$\mathbb{P}(N_{t+h} \geq n+2 \mid N_t = n) = 1 - \mathbb{P}(N_{t+h} \in \{n, n+1\} \mid N_t = n)$$

$$= 1 - \mathbb{P}(N_{t+h} = n \mid N_t = n) - \mathbb{P}(N_{t+h} = n+1 \mid N_t = n)$$

$$= 1 - (1 - \lambda h + o(h)) - (\lambda h + o(h)) = o(h)$$

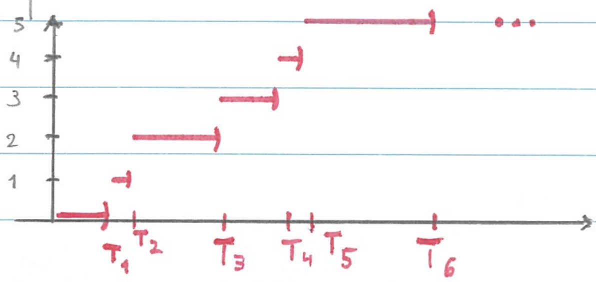
for h small enough. The 2 condition of 5) are used

<u>Def</u>: A <u>Poisson process with rate $\lambda > 0$</u> is a family of random variables $\{N_t\}_{t \geq 0}$ satisfying the above 5 conditions.

Such a model can be used for many purpose

For example, this model is used to describe the arrival of customers in shops/restaurants, or for the radioactive emission of particles.

Typical graph for $\{N_t\}_{t\geq 0}$:



Note that T_n and T_{n+1} could be equal.

Let T_i denote the time of arrival of the i^{th} message, i.e. $T_i := \inf\{t \mid N_t \geq i\}$. Then one has $0 = T_0 \leq T_1 \leq T_2 \leq T_3 \ldots$ and $N_t = i$ if and only if $t \in [T_i, T_{i+1})$. Thus $\{T_i\}_{i\in\mathbb{N}}$ is a sequence of random variables which determine $\{N_t\}_{t\geq 0}$ entirely, since

$$N_t = \max\{n \mid T_n \leq t\}.$$

The sequence $\{T_i\}_{i\in\mathbb{N}}$ is the "inverse process" of $\{N_t\}_{t\geq 0}$.

<u>Theorem</u> : For each $t > 0$, the random variable N_t has the Poison distribution with parameter λt, that is for any $k \in \mathbb{N}$, $\mathbb{P}(N_t = k) = \frac{1}{k!}(\lambda t)^k e^{-\lambda t}$.

probability of having received k messages at time t.

It directly follows that $\mathbb{E}(N_t) = \lambda t$, $var(N_t) = \lambda t$.

<u>Proof</u>: The idea is to get a differential equation, and to solve it. Let us set $P_k(t) := \mathbb{P}(N_t = k)$, fix $t > 0$ and consider h small enough. Then one has:

$$P_k(t+h) = \mathbb{P}(N_{t+h} = k)$$

partition thm

$$= \sum_{i=0}^{k} \mathbb{P}(N_{t+h} = k \mid N_t = i)\ \mathbb{P}(N_t = i)$$

$$= \mathbb{P}(N_{t+h} = k \mid N_t = k)\ \mathbb{P}(N_t = k) + \mathbb{P}(N_{t+h} = k \mid N_t = k-1)\ \mathbb{P}(N_t = k-1)$$
$$+ o(h) \longleftarrow \text{all the other terms}$$

$$= (1 - \lambda h + o(h))\, P_k(t) + (\lambda h + o(h))\, P_{k-1}(t) + o(h).$$

It follows that

$$P_k(t+h) - P_k(t) = \lambda h \left(P_{k-1}(t) - P_k(t) \right) + o(h)$$

$$\Leftrightarrow \frac{P_k(t+h) - P_k(t)}{h} = \lambda \left(P_{k-1}(t) - P_k(t) \right) + \frac{1}{h} o(h)$$

$$\xrightarrow{h \to 0} \underline{P_k'(t) = \lambda \left(P_{k-1}(t) - P_k(t) \right)}, \quad \text{for } k = 1, 2, \dots .$$

For $k = 0$, one has

$$\mathbb{P}(N_{t+h} = 0) = \mathbb{P}(N_{t+h} = 0 \mid N_t = 0)\, \mathbb{P}(N_t = 0)$$

$$= (1 - \lambda h)\, \mathbb{P}(N_t = 0) + o(h)$$

$$\Rightarrow \underline{P_0'(t) = -\lambda P_0(t)}.$$

for t for k

Thus, we obtain a **system of differential - difference equations** together with the boundary condition $N_0 = 0$, which is equivalent to $\begin{cases} P_0(0) = 1 \\ P_k(0) = 0 \text{ if } k \neq 0. \end{cases}$

Then, we can solve this system either by induction over k, or with a generating function, or by checking the proposed solution. $\square$

Recall that $T_i := \inf\{t \mid N_t \geq i\}$, and let us set

$X_i := T_i - T_{i-1}$ for $i = 1, 2, 3 \dots$

inter-arrival time

⚠ Observe that $\{N_t\}_{t\geq 0}$ is a continuous set of integer valued random variables, while $\{X_i\}_{i\in\mathbb{N}}$ is a discrete set of continuous random variable.

<u>Theorem</u>: In a Poisson process with rate λ, the inter-interval times $X_1, X_2, \dots$ are <u>independent</u> random variables, each following the <u>exponential distribution</u> with parameter λ, namely

$$\mathbb{P}(X_i \leq t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 - e^{-\lambda t} & \text{if } t \geq 0. \end{cases}$$

probability cumulative distribution function

We do not provide the proof of this result, let us just mention that it mainly follows from 4) on p. 100.

Def: A positive random variable X has the lack of memory if $\mathbb{P}(X > u+v \mid X > u) = \mathbb{P}(X > v)$ $\forall\, u, v \geq 0$. "The random variable does not remember how old it is."

Thm: A (a.) continuous random variable X has the lack of memory if and only if X follows the exponential distribution.

Proof: Let $\lambda > 0$, $u, v \geq 0$ and X following the exponential distribution. Then

$$\mathbb{P}(X > u+v \mid X > u) = \frac{\mathbb{P}(X > u+v \text{ and } X > u)}{\mathbb{P}(X > u)} = \frac{\mathbb{P}(X > u+v)}{\mathbb{P}(X > u)}$$

$$= \frac{e^{-\lambda(u+v)}}{e^{-\lambda u}} = e^{-\lambda v} = \mathbb{P}(X > v).$$

Conversely, set $G(u) = \mathbb{P}(X > u)$ for $u \geq 0$.

By the condition of lack of memory: $\mathbb{P}(X > u+v \mid X > u)$

$$= \frac{\mathbb{P}(X > u+v)}{\mathbb{P}(U > u)} = \frac{G(u+v)}{G(u)} \overset{\text{assumption}}{=} G(v) \quad \Rightarrow \quad G(u+v) = G(u)\, G(v)$$

with G non increasing. The only solution is $G(u) = e^{-\lambda u}$ for some $\lambda > 0$. $\square$

As a consequence of the previous 2 theorems, each inter-interval time X_j has the lack of memory. Conversely, if $\{X_j\}_{j\in\mathbb{N}}$ is a family of i.i.d. random variables following the exponential distribution with parameter $\lambda>0$, and if we set

$$T_n := \sum_{j=1}^{n} X_j \quad \text{and} \quad N_t := \max\{n \mid T_n \le t\},$$

then $\{N_t\}_{t\geq 0}$ is a Poisson process with rate λ.

<u>Application</u> : population growth (simple birth process)

Time $t=0$, $I\in\mathbb{N}$ amoebas in a pond. They can multiply themselves. Each cell division occurs at random time with the rules: $\exists\, \lambda>0$ s.t.

1) Prob. that 1 amoeba divides in $(t, t+h]$: $\lambda h + o(h)$,

2) Prob. that 1 amoeba does not divide in $(t,t+h]$: $1-\lambda h + o(h)$.

Note that the divisions are assumed to be independent in space and in time, and have no interaction with other amoebas.

Let M_t be the number of amoebas at time t. Then

$\mathbb{P}(M_{t+h} = k \mid M_t = k) = \mathbb{P}(\text{no division})$

presence of k independent amoebas at time t

$= (1 - \lambda h + o(h))^k = 1 - \lambda k h + o(h)$.

Also $\mathbb{P}(M_{t+h} = k+1 \mid M_t = k) = \mathbb{P}(1 \text{ division})$

$= \binom{k}{1} (\lambda h + o(h)) (1 - \lambda h + o(h))^{k-1}$

number of possibilities of having 1 division

$= k \lambda h + o(h)$ since $\binom{k}{1} = \dfrac{k!}{1!(k-1)!} = k$,

and $\mathbb{P}(M_{t+h} = k+2 \mid M_t = k) =$

$= 1 - \mathbb{P}(M_{t+h} \in \{k, k+1\} \mid M_t = k)$

$= 1 - (1 - \lambda k h) - \lambda k h + o(h) = o(h)$.

⇝ By following the approach of p 103 - 104 one gets :

Thm : If $M_0 = I$, then for $t > 0$,

$$\mathbb{P}(M_t = k) = \binom{k-1}{I-1} e^{-\lambda I t} (1 - e^{-\lambda t})^{k-1}$$ for $k \in \mathbb{N}$, $k \geq I$.

In this process, the number of amoebas can only increase. A more realistic model consists in a birth and death process:

1) division process, with rate $\lambda > 0$ as before
2) death process, with death rate $\mu > 0$.

⇝ For each amoeba, for the interval $(t, t+h]$, knowing that the amoeba lives at time t:

- death with prob. $\mu h + o(h)$,
- single division with prob. $\lambda h + o(h)$,
- no change, with prob. $(1 - \lambda h + o(h))(1 - \mu h + o(h)) = 1 - (\lambda + \mu) h + o(h)$,
- any other possibilities, with prob. $o(h)$.

Then, if we set $M_t = \#$ amoebas at time t, and $P_k(t) = \mathbb{P}(M_t = k)$, one can get ...

$$P_k'(t) = \lambda(k-1)\,P_{k-1}(t) - (\lambda+\mu)k\,P_k(t) + \mu(k+1)\,P_{k+1}(t)$$

with the convention $P_{-1}(t) = 0$ and the boundary condition

$$P_k(0) = \begin{cases} 1 & \text{if } k = I \\ 0 & \text{if } k \neq I \end{cases}.$$

For solving this differential - difference system, a method with generating function has to be used, see [GW] p 194 ...

Note that Chapters $\overline{\underline{IX}}$, $\overline{\underline{X}}$ and $\overline{\underline{XI}}$ can be studied in the more general setting of Markov processes. Still many beautiful subject to discover ...

Thank you