GEOSCIENCES
CENTRE DE GÉOSCIENCES

Probability for Geostatistics

ARMINES

Why probability ?

Probability theory $\Leftrightarrow$ mathematical analysis of random experiments

Random experiment

Trial that produces different outcomes when repeated indefinitely under the same controllable conditions

★ Randomness $\approx$ the uncontrolled, unexplained, unknown

Why probability ?

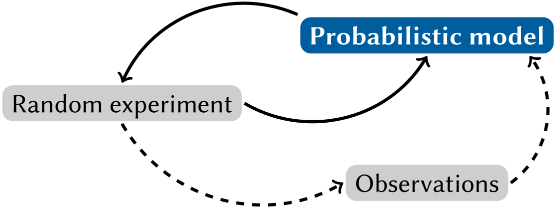
Probability theory ⇔ mathematical analysis of random experiments

Random experiment

Trial that produces different outcomes when repeated indefinitely under the same controllable conditions

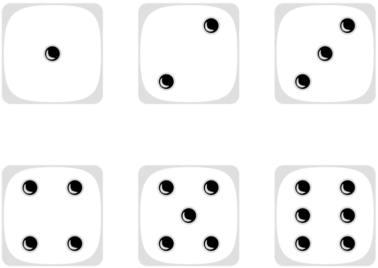
★ Randomness ≈ the uncontrolled, unexplained, unknown

In practice :



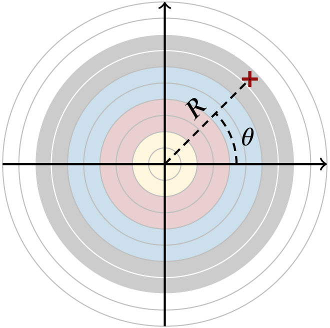
Two examples

Dice roll



Random : resulting number

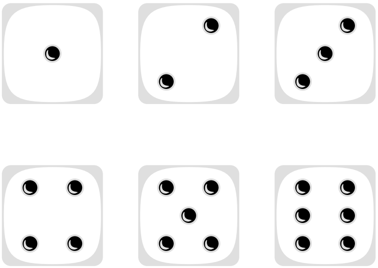
Archery



Random : R and θ

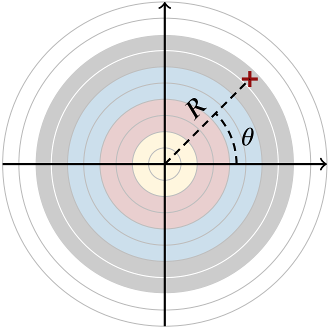
Two examples

Dice roll



Random : resulting number

Archery



Random : R and θ

Goal : **measure** the chance of realization of all sets of possible outcomes

Outline

1. **Measuring the uncertain**
   - How to construct a probability measure
   - Getting started with probability : basic rules
   - Getting to the point : random variables
2. **How to depict distributions of random variables**
   - Cumulative distribution functions (cdf)
   - Probability density and mass functions (pdf, pmf)
   - Dealing with dependence
3. **Momentous moments**
   - Expected value
   - Variance and covariance
   - Conditional moments

I

Measuring the uncertain

What to we need to measure the uncertain?

Goal Build a function $\boldsymbol{P}$ that associates a grade to each type of outcome of a random experiment

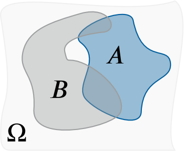
$$\boldsymbol{P}(\textit{type of outcome}) = \textit{grade}$$

The higher the grade, the more likely the type of outcome

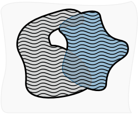
To build this mathematical function we need

- To list all possible outcomes of the experiment
- To specify **which** types of outcomes we want to grade (measure)
- To decide **how** we want to grade (quantify) them

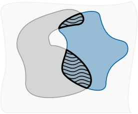
Prerequisites - Rudiments of set theory



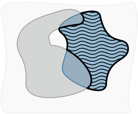
- Union: $A \cup B := \{x \in \Omega : x \in A \textbf{ or } x \in B\}$
- Intersection: $A \cap B := \{x \in \Omega : x \in A \textbf{ and } x \in B\}$
- Relative complement: $A \setminus B := \{x \in \Omega : x \in A \textbf{ but } x \notin B\}$
- Absolute complement: $A^c := \Omega \setminus A$
- Symmetric difference: $A \Delta B := (A \cup B) \setminus (A \cap B)$



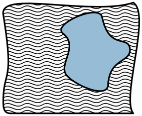
$A \cup B$



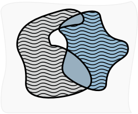
$A \cap B$



$A \setminus B$



A^c



$A \Delta B$

Inventory all outcomes - Universe

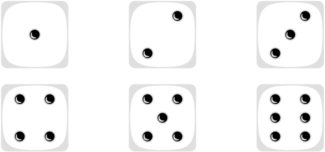
Universe

The universe (or sample space) is the set made of all theoretically possible outcomes of the studied random experiment; it is denoted by Ω

Universe

The universe (or sample space) is the set made of all theoretically possible outcomes of the studied random experiment; it is denoted by Ω

Dice roll

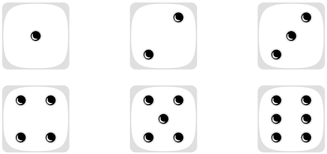


$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Universe

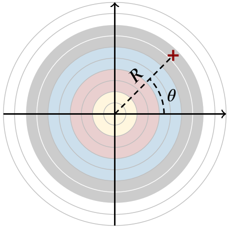
The universe (or sample space) is the set made of all theoretically possible outcomes of the studied random experiment; it is denoted by Ω

Dice roll



$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Archery



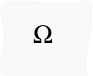
$$\Omega = [0, 122] \times [-\pi, \pi)$$

Specify what to measure - Events

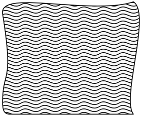
Ω

Set $\mathcal{A}$ made of all that can be measured on Ω?

Specify what to measure - Events

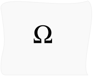


Set $\mathcal{A}$ made of all that can be measured on Ω?

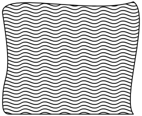


(i) *(Something happens)* $\Omega \in \mathcal{A}$

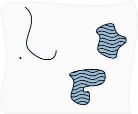
Specify what to measure - Events



Set $\mathcal{A}$ made of all that can be measured on Ω?

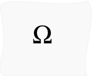


(i) *(Something happens)* $\Omega \in \mathcal{A}$

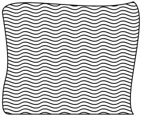


(ii) *(This or that happens)* $\mathcal{A}$ closed under countable unions :

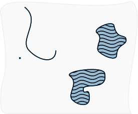
$$\left\{\forall\, n \in \mathbb{N} : A_n \in \mathcal{A}\right\} \quad \Rightarrow \quad \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$$



Set $\mathcal{A}$ made of all that can be measured on Ω?

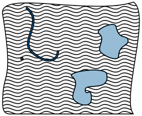


(i) *(Something happens)* $\Omega \in \mathcal{A}$



(ii) *(This or that happens)* $\mathcal{A}$ closed under countable unions :

$$\left\{\forall\, n \in \mathbb{N} \,:\, A_n \in \mathcal{A}\right\} \quad \Rightarrow \quad \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$$



(iii) *(This does not happen)* $\mathcal{A}$ closed under complementation :

$$A \in \mathcal{A} \quad \Rightarrow \quad A^c \in \mathcal{A}$$

Specify what to measure - Events

σ-algebra

Any set $\mathcal{A} \subset \mathcal{P}(\Omega)$ (power set of Ω) that fulfills conditions ***(i)*** *–* ***(iii)*** *is called a σ-algebra or a σ-field ("tribu" in French)*

★ $(\Omega, \mathcal{A})$ called a **measure space**

Specify what to measure - Events

σ-algebra

Any set $\mathcal{A} \subset \mathcal{P}(\Omega)$ (power set of Ω) that fulfills conditions **(i)** *–* **(iii)** *is called a σ-algebra or a σ-field ("tribu" in French)*

★ $(\Omega, \mathcal{A})$ called a **measure space**

Event

Let $(\Omega, \mathcal{A})$ be a measure space, then $A \in \mathcal{A}$ is called an event

Ex. $\emptyset$: impossible event $\quad$ Ω : sure event $\quad$ $\omega \in \Omega$: elementary/atomic event

★ $\mathcal{A}$ made of subsets of Ω : $\quad A \in \mathcal{A} \quad \Rightarrow \quad A \subset \Omega$

Choose how to quantify - Probability

Probability measure

A probability measure on the measure space $(\Omega, \mathcal{A})$ *is an application*

$$\boldsymbol{P}: \begin{array}{ccc} \mathcal{A} & \longrightarrow & [0,1] \\ A & \longmapsto & \boldsymbol{P}(A) \end{array}$$

such that (i) $\boldsymbol{P}(\Omega) = 1$ *and*

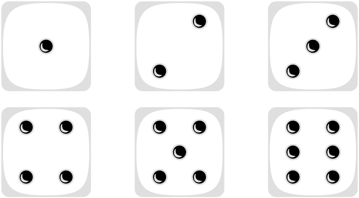
(ii) for all sequences $(A_n)_{n\in\mathbb{N}}$ *of pairwise disjoint (incompatible) events,*

$$\boldsymbol{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \boldsymbol{P}(A_n) \qquad (\sigma\text{-additivity})$$

- $(\Omega, \mathcal{A}, \boldsymbol{P})$ called a **probability space**
- Probability measures cannot be defined on any measure space !

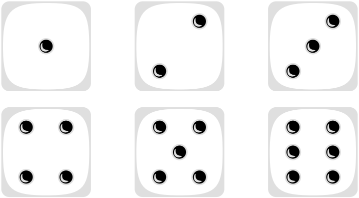
Choose how to quantify - Examples

Dice roll



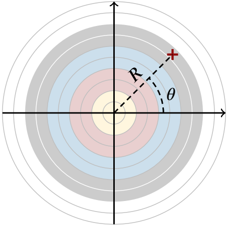
- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- $\mathcal{A} = \mathcal{P}(\Omega)$
- $\boldsymbol{P}(A) = \dfrac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)} \qquad A \in \mathcal{A}$

Dice roll



- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- $\mathcal{A} = \mathcal{P}(\Omega)$
- $\boldsymbol{P}(A) = \dfrac{\mathrm{Card}(A)}{\mathrm{Card}(\Omega)} \quad A \in \mathcal{A}$

Archery



- $\Omega = [0, 122] \times [-\pi, \pi)$
- $\mathcal{A} = \mathcal{B}(\Omega)$ *(Borel algebra)*
- $\boldsymbol{P}(A) = \dfrac{\mathrm{Area}(A)}{\mathrm{Area}(\Omega)} \quad A \in \mathcal{A}$

Measuring the uncertain - Summary

We needed

- A list of all possible outcomes ➨ *universe* Ω
- The types of outcomes to grade ➨ *events* listed in the *σ-algebra* $\mathcal{A}$
- A function to grade the types of outcomes ➨ *probability measure* $\boldsymbol{P}$

We have decided that

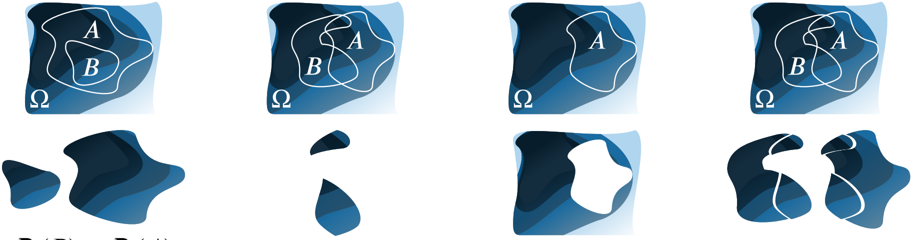
- Grades shall be a number between 0 and 1
- There is always something happening: $\boldsymbol{P}(\Omega) = 1$
- For two *incompatible events* A and B, the probability that either one of them happens is the sum of their probabilities

Given this setting, what else can we say about probability measures ?

Getting started - Basic properties

$(\Omega, \mathcal{A}, \boldsymbol{P})$ probability space $\qquad$ $A, B \in \mathcal{A}$ events

- Inclusion : $\quad B \subset A \quad \Rightarrow \quad \boldsymbol{P}(B) \leq \boldsymbol{P}(A)$
- Intersection : $\quad \boldsymbol{P}(A \cap B) \leq \boldsymbol{P}(A) \wedge \boldsymbol{P}(B)$
- Mutual exclusivity : $\quad A \cap B = \emptyset \quad \Rightarrow \quad \boldsymbol{P}(A \cap B) = 0$
- Absolute complement : $\quad \boldsymbol{P}(A^c) = 1 - \boldsymbol{P}(A)$
- Union : $\quad \boldsymbol{P}(A \cup B) = \boldsymbol{P}(A) + \boldsymbol{P}(B) - \boldsymbol{P}(A \cap B)$



$\boldsymbol{P}(B) \leq \boldsymbol{P}(A)$ $\qquad$ $\boldsymbol{P}(A \cap B)$ $\qquad$ $\boldsymbol{P}(A^c)$ $\qquad$ $\boldsymbol{P}(B) \quad \boldsymbol{P}(A)$

Getting started - (In)dependence

$(\Omega, \mathcal{A}, \boldsymbol{P})$ probability space events $\in \mathcal{A}$

Independent events

Two events A and B are said to be independent iff $\boldsymbol{P}(A \cap B) = \boldsymbol{P}(A)\, \boldsymbol{P}(B)$

★ A and B independent $\Leftrightarrow$ B has no influence on A and vice versa

Getting started - (In)dependence

Independent events

Two events A and B are said to be independent iff $\boldsymbol{P}(A \cap B) = \boldsymbol{P}(A)\,\boldsymbol{P}(B)$

★ A and B independent $\Leftrightarrow$ B has no influence on A and vice versa

Exercise

Consider a deck of 52 cards.

- Pick one card at random, put it back, shuffle and pick a second one.
  - 1^o What would be an adequate probability space for this experiment?
  - 2^o Are getting a jack and a queen independent events?
- How would you answer these questions if both cards were picked simultaneously?

Conditioning on an event

The conditional probability of an event A given that an event B with non-null probability has occurred is

$$\boldsymbol{P}(A \mid B) = \frac{\boldsymbol{P}(A \cap B)}{\boldsymbol{P}(B)}$$

Getting started - Conditioning

Conditioning on an event

The conditional probability of an event A given that an event B with non-null probability has occurred is

$$\boldsymbol{P}(A \mid B) = \frac{\boldsymbol{P}(A \cap B)}{\boldsymbol{P}(B)}$$

- $A \in \mathcal{A} \mapsto \boldsymbol{P}(A \mid B) \in [0, 1]$ probability measure
- extension to $\boldsymbol{P}(A \mid B)$ when $\boldsymbol{P}(B) = 0$: disintegration theorem
- A and B independent $\Rightarrow$ $\boldsymbol{P}(A \mid B) = \boldsymbol{P}(A)$ and $\boldsymbol{P}(B \mid A) = \boldsymbol{P}(B)$ if $\boldsymbol{P}(A), \boldsymbol{P}(B) \neq 0$

Getting started - Two major theorems

$(\Omega, \mathcal{A}, \boldsymbol{P})$ probability space events $\in \mathcal{A}$

Law of total probability

Let $(B_n)_{n\in\mathbb{N}}$ *be a partition of* Ω. *For any event* A $\boldsymbol{P}(A) = \sum_{n\in\mathbb{N}} \boldsymbol{P}\left(A \cap B_n\right)$

If in addition $\boldsymbol{P}\left(B_n\right) \neq 0$ *for all* $n \in \mathbb{N}$, *then* $\boldsymbol{P}(A) = \sum_{n\in\mathbb{N}} \boldsymbol{P}\left(A \mid B_n\right) \boldsymbol{P}\left(B_n\right)$

$(\Omega, \mathcal{A}, \boldsymbol{P})$ probability space events $\in \mathcal{A}$

Law of total probability

Let $(B_n)_{n\in\mathbb{N}}$ *be a partition of* Ω. *For any event* A $\boldsymbol{P}(A) = \sum_{n\in\mathbb{N}} \boldsymbol{P}\left(A \cap B_n\right)$

If in addition $\boldsymbol{P}\left(B_n\right) \neq 0$ *for all* $n \in \mathbb{N}$, *then* $\boldsymbol{P}(A) = \sum_{n\in\mathbb{N}} \boldsymbol{P}\left(A \mid B_n\right) \boldsymbol{P}\left(B_n\right)$

Bayes' theorem

For any two events A *and* B *such that* $\boldsymbol{P}(A) \neq 0$ *and* $\boldsymbol{P}(B) \neq 0$ *:*

$$\boldsymbol{P}(A \mid B) = \frac{\boldsymbol{P}(B \mid A)\, \boldsymbol{P}(A)}{\boldsymbol{P}(B)}$$

Main focus - Random elements

$(\Omega, \mathcal{A}, \boldsymbol{P})$ probability space — events $\in \mathcal{A}$

★ The experiment is often too complex to make $(\Omega, \mathcal{A}, \boldsymbol{P})$ explicit

$\Rightarrow$ **Assume** its existence and focus on the quantities of interest

Main focus - Random elements

$(\Omega, \mathcal{A}, \boldsymbol{P})$ probability space — events $\in \mathcal{A}$

★ The experiment is often too complex to make $(\Omega, \mathcal{A}, \boldsymbol{P})$ explicit
$\Rightarrow$ **Assume** its existence and focus on the quantities of interest

Random element (RE) - Intuition

Basically, a random element is an object with a priori unpredictable characteristics, e.g.

- *the shape, size and location of a potatoe drawn with closed eyes on a piece of paper,*
- *the color of a future baby cat,*
- *the total number rolled with 2 dices, etc.*

★ The experiment is often too complex to make $(\Omega, \mathcal{A}, \boldsymbol{P})$ explicit

$\Rightarrow$ **Assume** its existence and focus on the quantities of interest

Random element (RE) - Intuition

Basically, a random element is an object with a priori unpredictable characteristics, e.g.

- *the shape, size and location of a potatoe drawn with closed eyes on a piece of paper,*
- *the color of a future baby cat,*
- *the total number rolled with 2 dices, etc.*

★ Mathematically, random elements can be viewed as functions from a complex probability space to a more interesting or convenient one

Main focus - Random elements

The experiment is often too complex to make $(\Omega, \mathcal{A}, \boldsymbol{P})$ explicit

$\Rightarrow$ **Assume** its existence and focus on the quantities of interest

Random element (RE) - Formal definition

Let $(\mathcal{X}, \mathcal{B})$ be a measure space. A random element is an application $X : \Omega \to \mathcal{X}$ such that $\forall\, B \in \mathcal{B}$ $\quad X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}\quad$ *is an event ($\in \mathcal{A}$)*

Main focus - Random elements

$(\Omega, \mathcal{A}, \boldsymbol{P})$ probability space | events $\in \mathcal{A}$

★ The experiment is often too complex to make $(\Omega, \mathcal{A}, \boldsymbol{P})$ explicit
$\Rightarrow$ **Assume** its existence and focus on the quantities of interest

Random element (RE) - Formal definition

Let $(\mathcal{X}, \mathcal{B})$ be a measure space. A random element is an application $X : \Omega \to \mathcal{X}$ such that $\forall\, B \in \mathcal{B}$ $\quad X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}\quad$ *is an event ($\in \mathcal{A}$)*

- $\mathcal{X} \subset \mathbb{R}$ countable (ex. $\{0,1\}, \mathbb{N}, \mathbb{Z}, \mathbb{Q}$) : X discrete random variable (RV)
- $\mathcal{X} \subseteq \mathbb{R}$ **un**countable : X real-valued RV
- $\mathcal{X} \subseteq \mathbb{R}^d$ **un**countable : X random vector (RVec)

Main focus - Random elements

$(\Omega, \mathcal{A}, \boldsymbol{P})$ probability space — events $\in \mathcal{A}$

★ The experiment is often too complex to make $(\Omega, \mathcal{A}, \boldsymbol{P})$ explicit
$\Rightarrow$ **Assume** its existence and focus on the quantities of interest

Random element (RE) - Formal definition

Let $(\mathcal{X}, \mathcal{B})$ be a measure space. A random element is an application $X : \Omega \to \mathcal{X}$ such that $\forall\, B \in \mathcal{B}$ $\quad X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\}\quad$ is an event ($\in \mathcal{A}$)

- $\mathcal{X} \subset \mathbb{R}$ countable (ex. $\{0, 1\}, \mathbb{N}, \mathbb{Z}, \mathbb{Q}$) : X discrete random variable (RV)
- $\mathcal{X} \subseteq \mathbb{R}$ **un**countable : X real-valued RV
- $\mathcal{X} \subseteq \mathbb{R}^d$ **un**countable : X random vector (RVec)

★ Random elements extract the desired information from $(\Omega, \mathcal{A}, \boldsymbol{P})$

Main focus - Distribution of a RE

Distribution of a random element

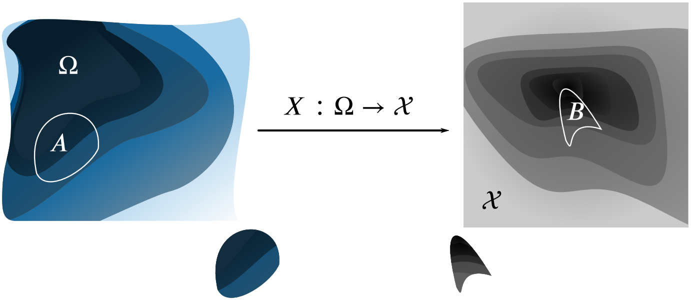
The distribution $\boldsymbol{P}_X$ of a random element X is the application

$$\boldsymbol{P}_X : \begin{array}{ccc} \mathcal{B} & \longrightarrow & [0,1] \\ B & \longmapsto & \boldsymbol{P}\left(X^{-1}(B)\right) \end{array}$$

- $\boldsymbol{P}_X$ also called the *pushforward measure* of $\boldsymbol{P}$ by X
- $(\mathcal{X}, \mathcal{B}, \boldsymbol{P}_X)$ probability space
- **Notation shortcut:** $\boldsymbol{P}(X \in B) := \boldsymbol{P}\left(X^{-1}(B)\right) = \boldsymbol{P}_X(B)$

Main focus - Pushforward measure

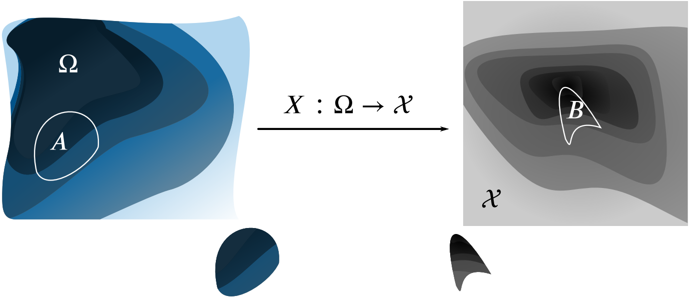
$(\Omega, \mathcal{A}, \boldsymbol{P})$ probability space $(\mathcal{X}, \mathcal{B})$ measure space RE $\Omega \to \mathcal{X}$



$$\boldsymbol{P}(A) = \boldsymbol{P}\left(X^{-1}(B)\right) = \boldsymbol{P}_X(B)$$

Main focus - Pushforward measure

$(\Omega, \mathcal{A}, \boldsymbol{P})$ probability space $\quad (\mathcal{X}, \mathcal{B})$ measure space $\quad$ RE $\Omega \to \mathcal{X}$



$$\boldsymbol{P}(A) = \boldsymbol{P}\left(X^{-1}(B)\right) = \boldsymbol{P}_X(B)$$

Exercise

Let X be the random variable that gives the total number rolled with 2 dices. What is the probability that X equals 8 ?

II

How to depict distributions of random variables

Cdf - Random variables

Cumulative distribution function (cdf)

The cdf of a random variable X is the application

$$F_X : \begin{array}{ccc} \mathbb{R} & \longrightarrow & [0,1] \\ x & \longmapsto & \boldsymbol{P}(X \leq x) \end{array}$$

Cdf - Random variables

Cumulative distribution function (cdf)

The cdf of a random variable X is the application $F_X : \begin{array}{ccc} \mathbb{R} & \longrightarrow & [0,1] \\ x & \longmapsto & \boldsymbol{P}(X \leq x) \end{array}$

- F_X non-decreasing and càdlàg (*"continue à droite, limite à gauche"*)
- $F_X(x) \xrightarrow[x \to -\infty]{} 0$ and $F_X(x) \xrightarrow[x \to +\infty]{} 1$

Cdf - Random variables

Cumulative distribution function (cdf)

The cdf of a random variable X is the application

$$F_X : \begin{array}{ccc} \mathbb{R} & \longrightarrow & [0,1] \\ x & \longmapsto & \boldsymbol{P}(X \leq x) \end{array}$$

- F_X non-decreasing and càdlàg (*"continue à droite, limite à gauche"*)
- $F_X(x) \xrightarrow[x \to -\infty]{} 0$ and $F_X(x) \xrightarrow[x \to +\infty]{} 1$

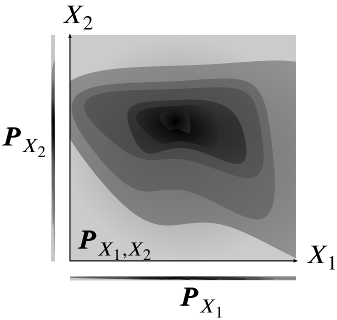
Exercise

Let X be the random variable that gives the total number rolled with 2 dices. What is its cdf ?

Cdf - Random vectors

Let $X = (X_1, \dots, X_d)$ be a RVec $(d \in \mathbb{N}^*)$

- $\boldsymbol{P}_X$ **joint distribution**
- $\boldsymbol{P}_{X_1}, \dots, \boldsymbol{P}_{X_d}$ **1-d marginal distributions**
- $\forall\, k \in \mathbb{N}^* : \left\{ \boldsymbol{P}_{X_{j_1}, \dots, X_{j_k}} : (j_1, \dots, j_k) \in [\![1, d]\!]^k \right\}$ ***k*-d marginal distributions**



★ **Notation:** $\boldsymbol{P}\left(X_1 \in B_1, \dots, X_d \in B_d\right)$

$= \boldsymbol{P}\left(X_1 \in B_1 \text{ and } \dots \text{ and } X_d \in B_d\right)$

$= \boldsymbol{P}\left(X_1^{-1}(B_1) \cap \dots \cap X_d^{-1}(B_d)\right)$

$= \boldsymbol{P}\,(X \in B) = \boldsymbol{P}_X\,(B)$

with $B = \{x_1 \in B_1 \text{ and } \dots \text{ and } x_d \in B_d\}$

Cdf - Random vectors

Cumulative distribution function

The cdf of a RVec $X = (X_1, \dots, X_d)$ is the application

$$F_X : \begin{array}{ccc} \mathbb{R}^d & \longrightarrow & [0,1] \\ x = (x_1, \dots, x_d) & \longmapsto & \boldsymbol{P}\left(X_1 \leq x_1, \dots, X_d \leq x_d\right) \end{array}$$

Cdf - Random vectors

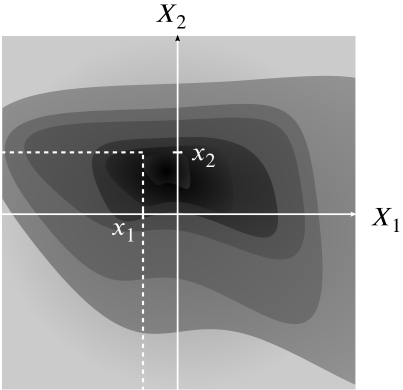
Cumulative distribution function

The cdf of a RVec $X = (X_1, \ldots, X_d)$ is the application

$$F_X : \begin{array}{ccc} \mathbb{R}^d & \longrightarrow & [0,1] \\ x = (x_1, \ldots, x_d) & \longmapsto & \boldsymbol{P}\left(X_1 \leq x_1, \ldots, X_d \leq x_d\right) \end{array}$$

- F_X non-decreasing and càdlàg for each of its variables
- $\forall\, j \in [\![1, d]\!] : F_X(x) \xrightarrow[x_j \to -\infty]{} 0$ and $F_X(x) \xrightarrow[x_1, \ldots, x_d \to +\infty]{} 1$
- $F_{X_1}, \ldots, F_{X_d}$ marginal cdfs

Cdf - Random vectors

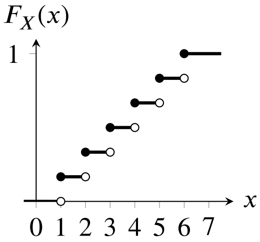


$F_X(x_1, x_2)$

Dice roll

If X is the result of a roll then $\forall\, x \in \mathbb{R}$:

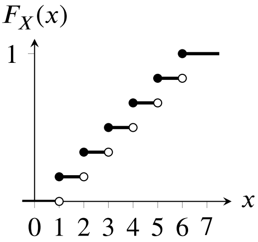
$$F_X(x) = \sum_{k=1}^{6} \frac{1}{6}\,\mathbf{1}\{k \le x\}$$



Dice roll

If X is the result of a roll then $\forall\, x \in \mathbb{R}$:

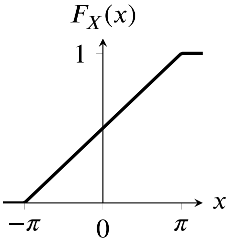
$$F_X(x) = \sum_{k=1}^{6} \frac{1}{6}\,\mathbf{1}\{k \leq x\}$$



Archery

If X is the angle of a blind shot, then $\forall\, x \in \mathbb{R}$:

$$F_X(x) = \frac{x+\pi}{2\,\pi}\,\mathbf{1}\{-\pi \leq x < \pi\} + \mathbf{1}\{x \geq \pi\}$$



Pdf/Pmf - Discrete univariate case

Probability mass function

The probability mass function (pmf) of a discrete RV X is the application

$$p_X : x \in \mathbb{R} \mapsto \begin{cases} \boldsymbol{P}(X = x) & \textit{if } x \in \mathcal{B}, \\ 0 & \textit{if } x \notin \mathcal{B}. \end{cases}$$

Probability mass function

The probability mass function (pmf) of a discrete RV X is the application

$$p_X : x \in \mathbb{R} \mapsto \begin{cases} \boldsymbol{P}(X = x) & \textit{if } x \in \mathcal{B}, \\ 0 & \textit{if } x \notin \mathcal{B}. \end{cases}$$

- $\forall\, x \in \mathbb{R}: \quad F_X(x) = \sum_{u \in \mathbb{R}} p_X(u) \mathbf{1}\{u \leq x\}$

Probability mass function

The pmf of a discrete RVec $X = (X_1, \ldots, X_d)$ is the application

$$p_X : (x_1, \ldots, x_d) \in \mathbb{R}^d \mapsto \begin{cases} \boldsymbol{P}\left(X_1 = x_1, \ldots, X_d = x_d\right) & \text{if } (x_1, \ldots, x_d) \in \mathcal{B}, \\ 0 & \text{if } (x_1, \ldots, x_d) \notin \mathcal{B}. \end{cases}$$

Pdf/Pmf - Discrete multivariate case

Probability mass function

The pmf of a discrete RVec $X = (X_1, \dots, X_d)$ is the application

$$p_X : (x_1, \dots, x_d) \in \mathbb{R}^d \mapsto \begin{cases} \boldsymbol{P}\left(X_1 = x_1, \dots, X_d = x_d\right) & \textit{if } (x_1, \dots, x_d) \in \mathcal{B}, \\ 0 & \textit{if } (x_1, \dots, x_d) \notin \mathcal{B}. \end{cases}$$

- $\forall\, x = (x_1, \dots, x_d) \in \mathbb{R}^d :$ $\quad F_X(x) = \sum_{u \in \mathbb{R}^d} p_X(u)\, \mathbf{1}\left\{u_1 \leq x_1, \dots, u_d \leq x_d\right\}$
- p_X exists $\Rightarrow$ its marginals exist and for all $j \in [\![1, d]\!], \forall\, x \in \mathbb{R}$:

$$p_{X_j}(x) = \sum_{u \in \mathbb{R}^{d-1}} p_X(u_1, \dots, u_{j-1}, x, u_{j+1}, \dots, u_d)$$

Pdf/Pmf - Continuous univariate case

Probability density function

A RV X has a probability density function (pdf) f_X if for all $B \in \mathcal{B}$ we can write

$$\boldsymbol{P}_X(B) = \int_B f_X(x)\, dx$$

Pdf/Pmf - Continuous univariate case

Probability density function

A RV X has a probability density function (pdf) f_X if for all $B \in \mathcal{B}$ we can write

$$\boldsymbol{P}_X(B) = \int_B f_X(x)\,dx$$

- $f_X(x)$ gives the marginal weight of the value $x \in \mathbb{R}$ with respect to $\boldsymbol{P}_X$
- $\forall\, x \in \mathbb{R}: \quad F_X(x) = \int_{-\infty}^{x} f_X(u)\,du \quad$ and $\quad f_X(x) = F'_X(x)$

Pdf/Pmf - Continuous multivariate case

Probability density function

A RVec $X = (X_1, \ldots, X_d)$ *has a (joint) pdf* f_X *if for all* $B \in \mathcal{B}$ *we can write*

$$\boldsymbol{P}_X(B) = \int_B f_X(x)\, dx$$

Probability density function

A RVec $X = (X_1, \ldots, X_d)$ has a (joint) pdf f_X if for all $B \in \mathcal{B}$ we can write

$$\boldsymbol{P}_X(B) = \int_B f_X(x)\, dx$$

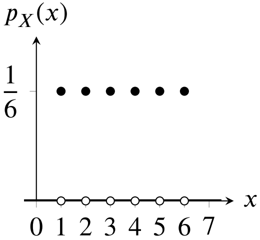
- $\forall\, x = (x_1, \ldots, x_d) \in \mathbb{R}^d$: $F_X(x) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_d} f_X(u_1, \ldots, u_d)\, du_d \ldots du_1$
- f_X exists $\Rightarrow$ its marginals exist and for all $j \in [\![1, d]\!], \forall\, x \in \mathbb{R}$:

$$f_{X_j}(x) = \int_{\mathbb{R}^{d-1}} f_X(u_1, \ldots, u_{j-1}, x, u_{j+1}, \ldots, u_d)\, du_d \ldots du_{j+1}\, du_{j-1} \ldots du_1$$

Dice roll

If X is the result of a roll then $\forall x \in \mathbb{R}$:

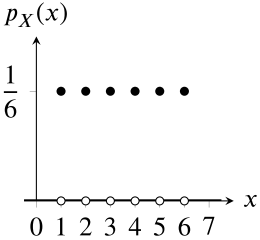
$$p_X(x) = \boldsymbol{P}(X = x) = \frac{1}{6}\mathbf{1}\{x \in \{1, \ldots, 6\}\}$$



Dice roll

If X is the result of a roll then $\forall\, x \in \mathbb{R}$:

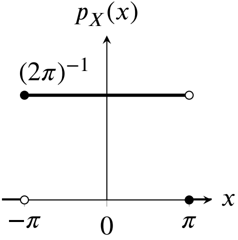
$$p_X(x) = \boldsymbol{P}\,(X = x) = \frac{1}{6}\,\mathbf{1}\{x \in \{1, \dots, 6\}\}$$



Archery

If X is the angle of a blind shot, then $\forall\, x \in \mathbb{R}$:

$$f_X(x) = \frac{1}{2\,\pi}\,\mathbf{1}\{-\pi \leq x < \pi\}$$



Conditional distribution

Conditional pdf

Assume that the RVec (X, Y) has joint pdf $f_{X,Y}$ with marginals f_X and f_Y.

The conditional pdf of X given that Y equals $y \in \mathbb{R}$ with $f_Y(y) \neq 0$ is defined for all $x \in \mathbb{R}$ as

$$f_{X|Y=y}(x) := \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

★ In the discrete case, replace pdfs by pmfs

Conditional distribution

Conditional pdf

Assume that the RVec (X, Y) has joint pdf $f_{X,Y}$ with marginals f_X and f_Y.

The conditional pdf of X given that Y equals $y \in \mathbb{R}$ with $f_Y(y) \neq 0$ is defined for all $x \in \mathbb{R}$ as

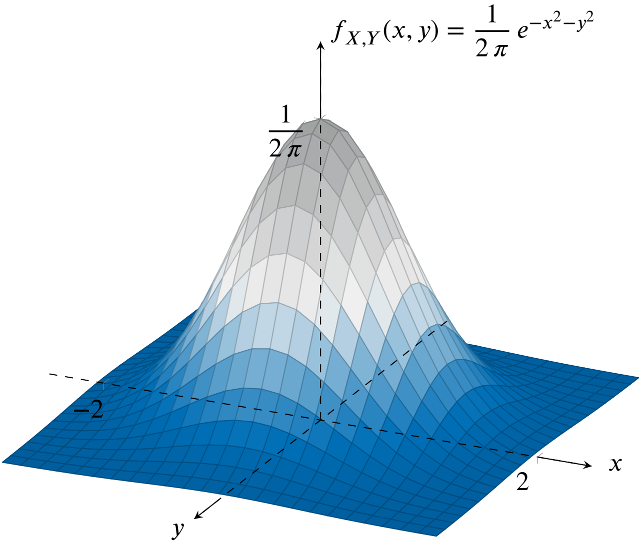
$$f_{X|Y=y}(x) := \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

- In the discrete case, replace pdfs by pmfs

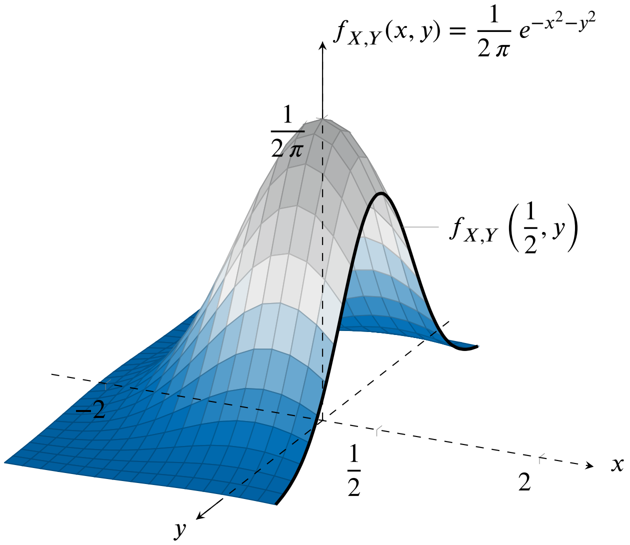
- X, Y independent $\Leftrightarrow$ $\forall (x, y) \in \mathbb{R}^2 : f_{X,Y}(x, y) = f_X(x)\, f_Y(y)$

  $\Rightarrow$ when $f_X(x), f_Y(y) \neq 0$: $f_{X|Y=y}(x) = f_X(x)$ and $f_{Y|X=x}(y) = f_Y(y)$

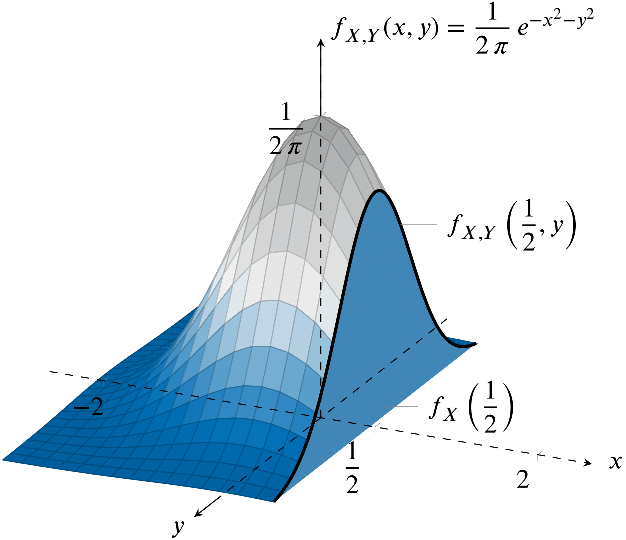
Conditional distribution



Conditional distribution



Conditional distribution



Conditional distribution - Discrete case

Law of total probability

Let X,Y be discrete RVs : $$\boldsymbol{P}(X = x) = \sum_{y \in \mathbb{R}} \boldsymbol{P}(X = x, Y = y) \quad \text{for all } x \in \mathbb{R}.$$

If in addition Y is valued in $\mathcal{Y} \subset \mathbb{R}$ and $\boldsymbol{P}(Y = y) > 0$ for all $y \in \mathcal{Y}$:

$$\boldsymbol{P}(X = x) = \sum_{y \in \mathcal{Y}} \boldsymbol{P}(X = x \mid Y = y)\, \boldsymbol{P}(Y = y)$$

Conditional distribution - Discrete case

Law of total probability

Let X,Y be discrete RVs : $\boldsymbol{P}(X = x) = \sum_{y \in \mathbb{R}} \boldsymbol{P}(X = x, Y = y)$ *for all* $x \in \mathbb{R}$.

If in addition Y *is valued in* $\mathcal{Y} \subset \mathbb{R}$ *and* $\boldsymbol{P}(Y = y) > 0$ *for all* $y \in \mathcal{Y}$ *:*

$$\boldsymbol{P}(X = x) = \sum_{y \in \mathcal{Y}} \boldsymbol{P}(X = x \mid Y = y)\, \boldsymbol{P}(Y = y)$$

Bayes' theorem

Let X,Y be discrete RVs. For all $(x, y) \in \mathbb{R}^2$ *such that* $\boldsymbol{P}(X = x), \boldsymbol{P}(Y = y) \neq 0$ *:*

$$\boldsymbol{P}(Y = y \mid X = x) = \frac{\boldsymbol{P}(X = x \mid Y = y)\, \boldsymbol{P}(Y = y)}{\boldsymbol{P}(X = x)}$$

Conditional distribution - Continuous case

Law of total probability

Let (X, Y) be a RVec with joint pdf $f_{X,Y}$: $\quad f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y)\, dy \quad \forall\, x \in \mathbb{R}.$

If in addition Y is valued in $\mathcal{Y} \subset \mathbb{R}_+$ and $f_Y(y) > 0$ for all $y \in \mathcal{Y}$:

$$f_X(x) = \int_{\mathcal{Y}} f_{X|Y=y}(x)\, f_Y(y)\, dy$$

Conditional distribution - Continuous case

Law of total probability

Let (X, Y) be a RVec with joint pdf $f_{X,Y}$: $\quad f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y)\, dy \quad \forall\, x \in \mathbb{R}.$

If in addition Y is valued in $\mathcal{Y} \subset \mathbb{R}_+$ and $f_Y(y) > 0$ for all $y \in \mathcal{Y}$:

$$f_X(x) = \int_{\mathcal{Y}} f_{X|Y=y}(x)\, f_Y(y)\, dy$$

Bayes' theorem

Let (X, Y) be a RVec with joint pdf $f_{X,Y}$. For all $(x, y) \in \mathbb{R}^2$ such that

$$f_X(x), f_Y(y) \neq 0 : \quad f_{Y|X=x}(y) = \frac{f_{X|Y=y}(x)\, f_Y(y)}{f_X(x)}$$

Momentous moments

Moments - Expectation

Expectation

The expectation of a RV X is $\boldsymbol{E}(X) = \int_{\mathbb{R}} x\, \boldsymbol{P}_X(dx) \in \mathbb{R}$ *when the integral is absolutely convergent*

★ Expectation ≈ average behavior of X

Moments - Expectation

Expectation

The expectation of a RV X is $\boldsymbol{E}(X) = \int_{\mathbb{R}} x \, \boldsymbol{P}_X(dx) \in \mathbb{R}$ *when the integral is absolutely convergent*

★ Expectation $\approx$ average behavior of X

- X has pdf f_X : $\boldsymbol{E}(X) = \int_{\mathbb{R}} x \, f_X(x) \, dx$
- X has pmf p_X : $\boldsymbol{E}(X) = \sum_{x \in \mathbb{R}} x \, p_X(x)$
- $\forall \, B \in \mathcal{B}(\mathbb{R}) : \boldsymbol{E}\left[\mathbb{1}\{X \in B\}\right] = \boldsymbol{P}_X(B) = \boldsymbol{P}(X \in B)$
- $X \in \mathbb{R}^d \quad \Rightarrow \quad \boldsymbol{E}(X) = \left(\boldsymbol{E}(X_1), \ldots, \boldsymbol{E}(X_d)\right) \in \mathbb{R}^d$

Properties : for any random variables X, Y with existing expectations

- $\forall\, \alpha, \beta \in \mathbb{R}: \quad \boldsymbol{E}\,(\alpha\, X + \beta\, Y) = \alpha\, \boldsymbol{E}\,(X) + \beta\, \boldsymbol{E}\,(Y)$ (linearity)
- X and Y independent $\Rightarrow \quad \boldsymbol{E}\,(X\, Y) = \boldsymbol{E}\,(X)\; \boldsymbol{E}\,(Y)$ $(\not\Leftarrow)$
- $g(X)$ valued in a measure space $\Rightarrow \quad \boldsymbol{E}\,(g(X)) = \int_{\mathbb{R}} g(x)\, \boldsymbol{P}_X(dx)$
- ϕ convex function on $\mathbb{R}$ $\Rightarrow \quad \phi\left(\boldsymbol{E}\,(X)\right) \leq \boldsymbol{E}\,(\phi(X))$ (Jensen's inequality)
- X nonnegative, $a > 0$ $\Rightarrow \quad \boldsymbol{P}\,(X \geq a) \leq \dfrac{\boldsymbol{E}\,(X)}{a}$ (Markov's inequality)

Moments - Variance

Variance

The variance of a RV X is $\boldsymbol{V}(X) = \boldsymbol{E}\left(\left(X - \boldsymbol{E}(X)\right)^2\right)$ *when the expectations exist (we can also write $Var(X)$)*

- Variance $\approx$ average deviation to the average behavior of X
- $\sigma(X) := \sqrt{\boldsymbol{V}(X)}$ **standard deviation** of X

Moments - Variance

Variance

The variance of a RV X is $\boldsymbol{V}(X) = \boldsymbol{E}\left(\left(X - \boldsymbol{E}(X)\right)^2\right)$ *when the expectations exist (we can also write $Var(X)$)*

- Variance $\approx$ average deviation to the average behavior of X
- $\sigma(X) := \sqrt{\boldsymbol{V}(X)}$ **standard deviation** of X

- $\boldsymbol{V}(X) = \boldsymbol{E}\left(X^2\right) - \boldsymbol{E}(X)^2$
- $\forall \alpha, \beta \in \mathbb{R}: \quad \boldsymbol{V}(\alpha X + \beta) = \alpha^2 \boldsymbol{V}(X)$
- $\forall a > 0: \quad \boldsymbol{P}\left(\left|X - \boldsymbol{E}(X)\right| \geq a\right) \leq \dfrac{\boldsymbol{V}(X)}{a^2}$ (Chebyshev's inequality)

Moments - Covariance

Covariance

The covariance of a RVec $X = (X_1, \dots, X_d)$ is defined as the $d \times d$ matrix

$$\Sigma^X := \boldsymbol{E}\left(\left(X - \boldsymbol{E}(X)\right)\left(X - \boldsymbol{E}(X)\right)^T\right)$$

when the expectations exist

- Covariance $\approx$ average deviation to independence of X

  RVs X and Y independent $\Rightarrow Cov(X, Y) = 0$ ⚠ $\not\Leftarrow$

Moments - Covariance

Covariance

The covariance of a RVec $X = (X_1, \dots, X_d)$ is defined as the $d \times d$ matrix

$$\Sigma^X := \boldsymbol{E}\left(\left(X - \boldsymbol{E}(X)\right)\left(X - \boldsymbol{E}(X)\right)^T\right)$$

when the expectations exist

Covariance $\approx$ average deviation to independence of X

RVs X and Y independent $\Rightarrow Cov(X, Y) = 0$ $\not\Leftarrow$

- Σ^X positive, semi-definite, symmetric
- $\forall (i, j) \in [\![1, d]\!]^2 : \quad \Sigma^X_{i,j} =: Cov(X_i, X_j)$ and $\Sigma^X_{i,i} = \boldsymbol{V}\left(X_i\right)$
- $\Sigma^X = \boldsymbol{E}\left(X\, X^T\right) - \boldsymbol{E}(X)\, \boldsymbol{E}(X)^T$

Conditional moments

X, Y RVs with joint pdf $f_{X,Y}$ plus finite expectations and variances

$y \in \mathbb{R}$ such that $f_Y(y) \neq 0$

Conditional moments

X, Y RVs with joint pdf $f_{X,Y}$ plus finite expectations and variances

$y \in \mathbb{R}$ such that $f_Y(y) \neq 0$

Conditional expectation : $\boldsymbol{E}\,(X \mid Y = y) = \int_{\mathbb{R}} x\, f_{X|Y=y}(x)\, dx$

- $\boldsymbol{E}\,(X \mid Y) : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ real-valued RV
- $\boldsymbol{E}\,(X) = \boldsymbol{E}\left(\boldsymbol{E}\,(X \mid Y)\right)$ (law of total expectation)

X, Y RVs with joint pdf $f_{X,Y}$ plus finite expectations and variances

$y \in \mathbb{R}$ such that $f_Y(y) \neq 0$

Conditional expectation : $\boldsymbol{E}(X \mid Y = y) = \int_{\mathbb{R}} x\, f_{X|Y=y}(x)\, dx$

- $\boldsymbol{E}(X \mid Y) : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ real-valued RV
- $\boldsymbol{E}(X) = \boldsymbol{E}\left(\boldsymbol{E}(X \mid Y)\right)$ (law of total expectation)

Conditional variance : $\boldsymbol{V}(X \mid Y = y) = \boldsymbol{E}\left(\left(X - \boldsymbol{E}(X \mid Y = y)\right)^2 \,\middle|\, Y = y\right)$

- $\boldsymbol{V}(X \mid Y) : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ real-valued RV
- $\boldsymbol{V}(X) = \boldsymbol{E}\left(\boldsymbol{V}(X \mid Y)\right) + \boldsymbol{V}\left(\boldsymbol{E}(X \mid Y)\right)$ (law of total variance)

Conditional moments

X, Y, Z RVs with joint pdf $f_{X,Y,Z}$ plus finite expectations and variances

$z \in \mathbb{R}$ such that $f_Z(z) \neq 0$

Conditional covariance :

$$Cov\,(X, Y \mid Z = z) = \boldsymbol{E}\left(\big(X - \boldsymbol{E}\,(X \mid Z = z)\big)\,\big(Y - \boldsymbol{E}\,(Y \mid Z = z)\big)\,\Big|\,Z = z\right)$$

- $Cov(X, Y \mid Z) : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ real-valued RV
- $Cov(X, Y) = \boldsymbol{E}\,(Cov(X, Y \mid Z)) + Cov\,\big(\boldsymbol{E}\,(X \mid Z), \boldsymbol{E}\,(Y \mid Z)\big)$

  (law of total covariance)

Common distributions

- Binomial distribution
- Poisson distribution
- Uniform distribution
- Exponential distribution
- Gaussian distribution
- Multivariate Gaussian distribution

Common distributions - Binomial

$$X \sim \mathcal{B}(n,p) \qquad (n,p) \in \mathbb{N}^* \times [0,1] \qquad X : (\Omega, \mathcal{A}) \to \left([\![0,n]\!], \mathcal{P}([\![0,n]\!]) \right)$$

$$\forall\, x \in \mathbb{R} : \quad F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k \, (1-p)^{n-k} \, \mathbf{1}\left\{x \in [0,n)\right\} + \mathbf{1}\left\{x \geq n\right\}$$

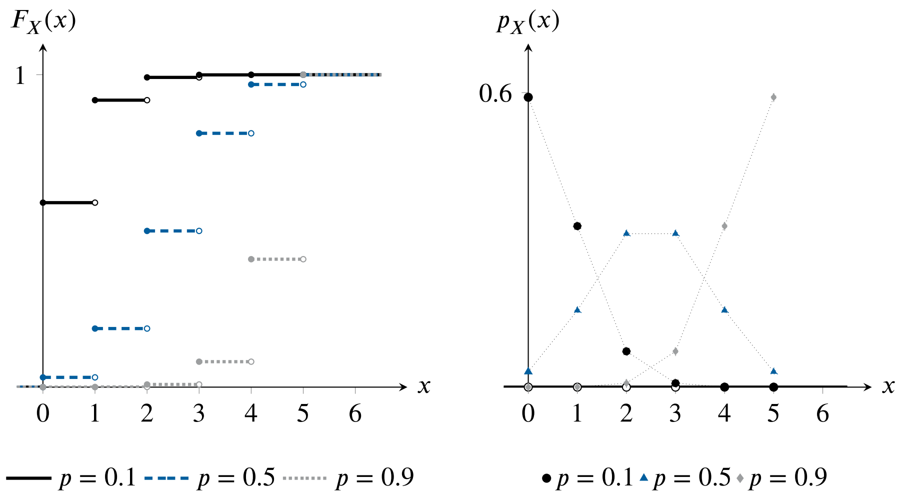
$$p_X(x) = \binom{n}{x} p^x \, (1-p)^{n-x} \, \mathbf{1}\left\{x \in [\![0,n]\!]\right\}$$

$$\boldsymbol{E}\,(X) = n\,p \qquad \boldsymbol{V}\,(X) = n\,p\,(1-p)$$

$$n = 1 \quad \Rightarrow \quad X \sim \mathcal{B}(p) \text{ Bernoulli distribution}$$

Common distributions - Binomial

$$X \sim \mathcal{B}(5, p)$$



$X \sim \mathcal{P}(\lambda)$ $\lambda \in \mathbb{R}_+^*$ $X : (\Omega, \mathcal{A}) \to (\mathbb{N}, \mathcal{P}(\mathbb{N}))$

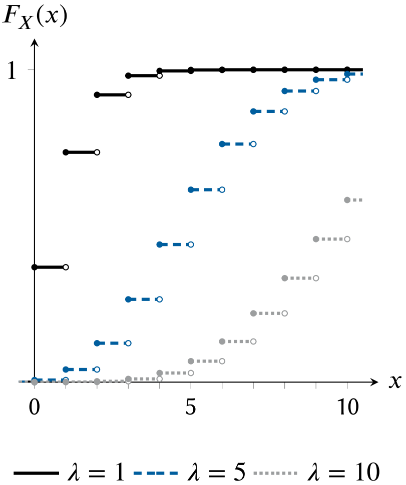
$$\forall\, x \in \mathbb{R} : \quad F_X(x) = \frac{\Gamma\left(\lfloor x+1 \rfloor, \lambda\right)}{\lfloor x \rfloor !} \mathbf{1}\left\{x \geq 0\right\}$$

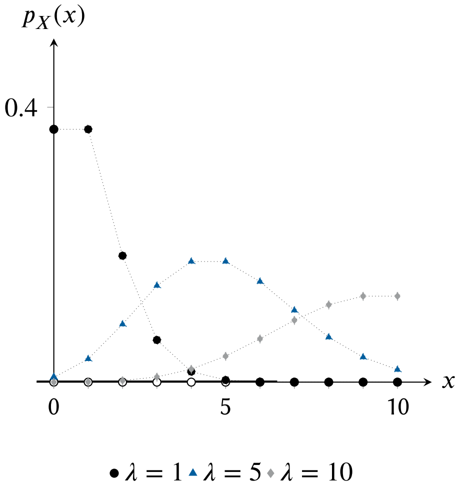
$$p_X(x) = \frac{e^{-\lambda}\,\lambda^x}{x!} \mathbf{1}\left\{x \in [\![0, n]\!]\right\}$$

$\boldsymbol{E}\,(X) = \lambda$ $\boldsymbol{V}\,(X) = \lambda$

Common distributions - Poisson

$$X \sim \mathcal{P}(\lambda)$$





Common distributions - Uniform

$X \sim \mathcal{U}_{[a,b]}$ $\quad$ $(a,b) \in \mathbb{R}^2,\, a < b$ $\quad$ $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$

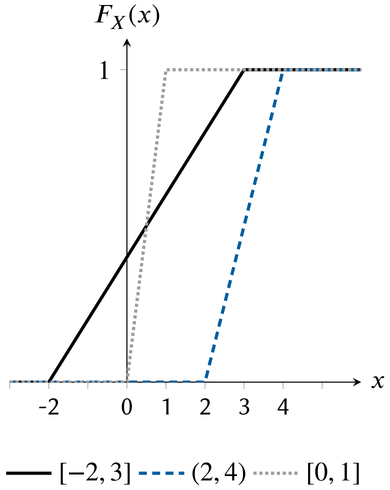
$$\forall\, x \in \mathbb{R} : \quad F_X(x) = \frac{x-a}{b-a}\,\mathbf{1}\left\{x \in [a,b]\right\} + \mathbf{1}\left\{x > b\right\}$$

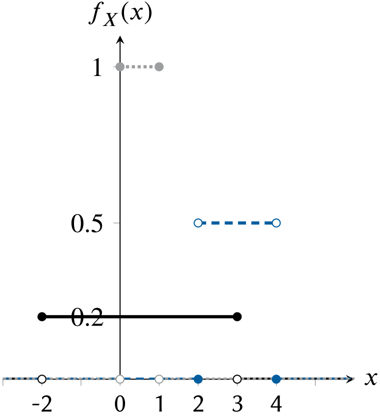
$$f_X(x) = \frac{1}{b-a}\,\mathbf{1}\left\{x \in [a,b]\right\}$$

$$\boldsymbol{E}\,(X) = \frac{a+b}{2} \qquad \boldsymbol{V}\,(X) = \frac{(b-a)^2}{12}$$

Common distributions - Uniform

$$X \sim \mathcal{U}_{[a,b]}$$





Common distributions - Exponential

$X \sim \mathcal{E}(\lambda)$ $\qquad \lambda \in \mathbb{R}_+^*$ $\qquad X : (\Omega, \mathcal{A}) \to (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$

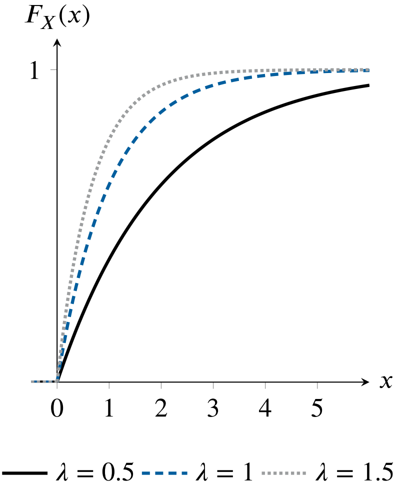
$$\forall\, x \in \mathbb{R} : \quad F_X(x) = 1 - e^{-\lambda\, x}\, \mathbf{1}\,\{x \geq 0\}$$

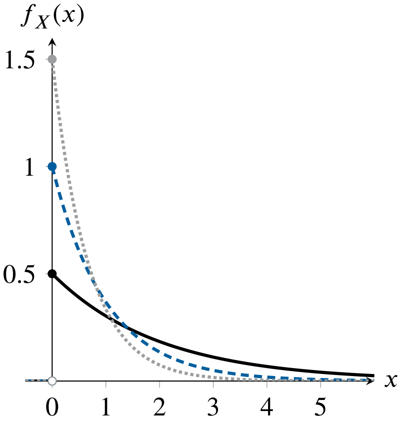
$$f_X(x) = \lambda\, e^{-\lambda\, x}\, \mathbf{1}\,\{x \geq 0\}$$

$$\boldsymbol{E}\,(X) = \frac{1}{\lambda} \qquad \boldsymbol{V}\,(X) = \frac{1}{\lambda^2}$$

Common distributions - Exponential

$$X \sim \mathcal{E}(\lambda)$$





Common distributions - Gaussian

$X \sim \mathcal{N}(\mu, \sigma^2)$ $\quad$ $(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+^*$ $\quad$ $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$

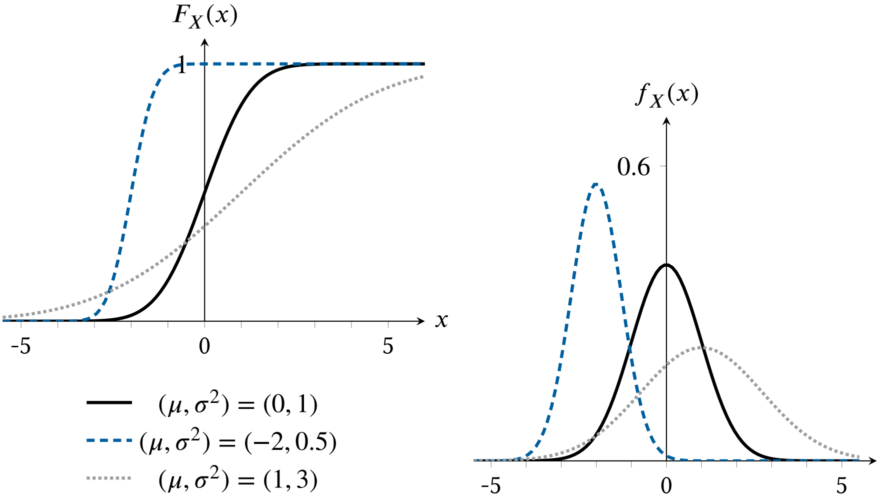
$$\forall\, x \in \mathbb{R} : \quad F_X(x) = \frac{1}{2}\left(1 + \operatorname{erf}\frac{x-\mu}{\sigma\sqrt{2}}\right)$$

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

$\boldsymbol{E}\,(X) = \mu$ $\quad$ $\boldsymbol{V}\,(X) = \sigma^2$

Common distributions - Gaussian

$$X \sim \mathcal{N}(\mu, \sigma^2)$$



Common distributions - Multi. Gaussian

$$X \sim \mathcal{N}(\mu, \Sigma) \qquad (\mu, \Sigma) \in \mathbb{R}^d \times \mathcal{M}_{d \times d}(\mathbb{R}_+^*),\, d \in \mathbb{N}^* \qquad X : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$$

$$\forall\, x \in \mathbb{R}^d : \quad f_X(x) = \frac{1}{|\Sigma|^{1/2}\,(2\,\pi)^{d/2}} \exp\left\{ \frac{1}{2}\,(x-\mu)^T\,\Sigma^{-1}\,(x-\mu) \right\}$$

$$\boldsymbol{E}\,(X) = \mu \qquad\qquad \boldsymbol{V}\,(X) = \Sigma$$

Common distributions - Multi. Gaussian

$$X \sim \mathcal{N}\left(\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}0.25 & 0.3\\0.3 & 2\end{pmatrix}\right)$$