Chapter 2

Probability theory

- Probability space
- Random variable
- Expectation and moments
- Classical probability distributions
- Independence
- Conditional expectation
- Stochastic convergences

Probability space

- Random situation
- Events
  - Intuitively
  - σ-field of events
- Probability
  - Axioms
  - Consequences
  - Probability space
  - Finite equiprobable model

Random situation

= physical situation for which several outcomes are possible

Set of possible outcomes : Ω

Events

Intuitively : any subset of Ω

For an observed outcome $\omega \in \Omega$, the event A occurs iff $\omega \in A$

Particular events

- the impossible event : $\emptyset$
- the sure event : Ω

Taking any subset of Ω as an event is not convenient

- mathematically : if Ω is non denumerable, taking every subset of Ω as an event may lead to some contradiction
- financially : it is sometimes useful to consider the set of events at time t as the available information up to time t

Furthermore, we have to authorize elementary set operations : "or" is $\cup$, "and" is $\cap$, …

σ-field (or σ-algebra) **of events**

= Set $\mathcal{F}$ of subsets of Ω such that

- $\emptyset \in \mathcal{F}$
- If $A \in \mathcal{F}$, then $\bar{A} \in \mathcal{F}$
- If $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{F}$, then

$$A_1 \cup A_2 \cup \ldots \cup A_n \cup \ldots = \bigcup_i A_i \in \mathcal{F}$$

Consequences

- $\Omega \in \mathcal{F}$
- $A_1 \cap A_2 \cap \ldots \cap A_n \cap \ldots \in \mathcal{F}$

Examples

- $\mathcal{F} = \{\emptyset, \Omega\}$
- $\mathcal{F} = \{\emptyset, \Omega, A, \bar{A}\}$
- ...

Theorem

Given a subset (non necessarily a σ-field) $\mathcal{G}$ of $\mathcal{F}$, there exist a unique smallest σ-field containing $\mathcal{G}$: the σ-field generated by $\mathcal{G}$, denoted $\sigma(\mathcal{G})$

Exercice : in the general case describe the σ-field generated by two events $\{A, B\}$

Note

For 2 σ-fields $\mathcal{F}$ and $\mathcal{G}$ on Ω, the relation $\mathcal{G} \subset \mathcal{F}$ means "the information in $\mathcal{F}$ is finer (more precise) than the one in $\mathcal{G}$"

Probability

= measure, for an event, of its tendency to occur

Axioms (Kolmogorov)

(K1) $\forall A \in \mathcal{F},\ \Pr(A) \geq 0$

(K2) $\forall A_1, \dots, A_n, \dots \in \mathcal{F}$, if the events are (pairwise) disjoint,

$$\Pr\left(\bigcup_i A_i\right) = \sum_i \Pr(A_i)$$

(K3) $\Pr(\Omega) = 1$

Consequences

$$A \subset B \quad \Longrightarrow \quad \Pr(A) \leq \Pr(B)$$
$$0 \leq \Pr(A) \leq 1$$
$$\Pr(\emptyset) = 0$$
$$\Pr(\bar{A}) = 1 - \Pr(A)$$
$$\Pr(A \cup B) = \Pr(\mathrm{A}) + \Pr(\mathrm{B}) - \Pr(\mathrm{A} \cap \mathrm{B})$$

Probability space

= probability triple : $(\Omega, \mathcal{F}, \Pr)$

Finite equiprobable model

If Ω is finite $(\Omega = \{\omega_1, \dots, \omega_n\})$ and equiprobable $(\Pr\{\omega_j\} = 1/n\ \ \forall j)$, then

$$\Pr(A) = \frac{\#(A)}{n} = \frac{\#(A)}{\#(\Omega)}$$

Random variable

- Definitions
  - Intuitively
  - Mathematically
  - Borel sets of $\mathbb{R}$
  - σ-field generated by a r.v.
- Probability law
  - Ideally
  - Cumulative distribution function
- Types of r.v.
  - Discrete
  - Continuous
  - Mixed
- Random vector
  - Borel sets of $\mathbb{R}^m$
  - Definition
  - Joint cumulative distribution function
  - Types of random vectors

Definitions

Intuitively : a variable whose value depends on the result of a random situation

$$X : \Omega \to \mathbb{R} : \omega \mapsto X(\omega)$$

Furthermore, expressions like "$X \in E$" must be, for "reasonable" E, an event

Mathematically, a r.v. is a function from Ω to $\mathbb{R}$ that is $\mathcal{F}$-measurable : for every borelian set E of $\mathbb{R}$,

$$X^{-1}[E] = \{\omega : X(\omega) \in E\} = [X \in E]$$

is an element of $\mathcal{F}$.

The support of a r.v. = the set of possible values of this r.v. :

$$X[\Omega] = \{X(\omega) : \omega \in \Omega\}$$

Borel sets of $\mathbb{R}$

= denumerable unions of intervals (bounded or not ; closed, open or semi-interval) and their complementaries

Notation : $\mathcal{B}$

Property : $\mathcal{B}$ is the σ-field on $\mathbb{R}$ generated by

$$\{[a; b[\ : a < b\}$$

σ-field generated by a r.v.

= the smallest sub-σ-field of $\mathcal{F}$ that contains every event of the form $[X \in E]$ with $E \in \mathcal{B}$

Notation : $\sigma(X)$

Probability law

Ideally, $\Pr[X \in E]$ for every $E \in \mathcal{B}$

Cumulative distribution function

$$F_X(t) = \Pr[X \leq t]$$

Properties :

$0 \leq F(t) \leq 1$ for every t

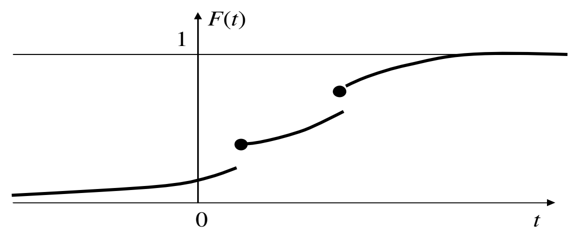
$F(t)$ is a non decreasing function

$\lim_{t\to-\infty} F(t) = 0$

$\lim_{t\to+\infty} F(t) = 1$

$F(t)$ is continuous to the right

⊕ : it is possible to construct the probability law from the c.d.f.



Types of r.v.

Discrete

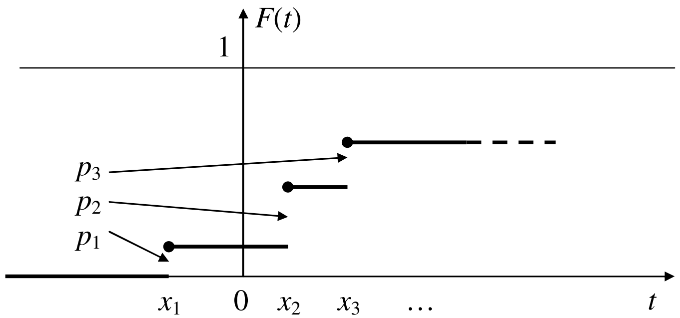
The support $X[\Omega]$ is finite or denumerable :

$$X \sim \begin{pmatrix} x_1 & \dots & x_n & \dots \\ p_1 & \dots & p_n & \dots \end{pmatrix}$$

with $\Pr[X = x_j] = p_j > 0$ and $\sum p_i = 1$

Probability law : $\Pr[X \in E] = \sum_{\{i : x_i \in E\}} p_i$

C.d.f. :



Continuous

The support is a non denumerable set (generally an interval) and, $\forall x, \Pr[X = x] = 0$

The probabilities are continuously distributed via a density function $f_X(x) \geq 0$, with

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

$$\Pr[x < X \leq x + h] \approx f(x) \cdot h$$

Probability law : $\Pr[X \in E] = \int_E f(x)dx$

C.d.f. :

$$F(t) = \int_{-\infty}^{t} f(x)dx$$

is a continuous function and, if it is derivable,

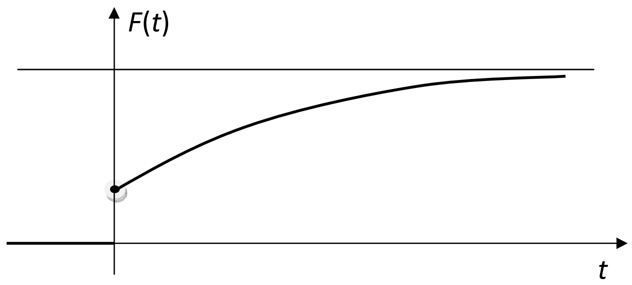
$$F'(t) = f(t)$$

Mixed

Is a mix of discrete and continuous

Example : the r.v. C represents the cost for the company of an insurance policy

- for >0 cost, it is easier to consider the cost as a continuous r.v.
- but, $\Pr[C=0]>0$



Random vector

Borel sets of $\mathbb{R}^m$

= denumerable unions of intervals (bounded or not ; closed, open or semi-interval), like

$$]a_1;b_1]\times\{a_2\}\times]-\infty;b_3]\times\ldots\times[a_m;\ +\infty[$$

and their complementaries

Notation : $\mathcal{B}_m$

Property : $\mathcal{B}_m$ is the σ-field on $\mathbb{R}^m$ generated by

$$\prod_{j=1}^{m}[a_j;b_j[\;=[a_1;b_1[\times\ldots\times[a_m;b_m[$$

Definition

A random vector $(X_1, \dots, X_m)$ is a function from Ω to $\mathbb{R}^m$, that is $\mathcal{F}$-measurable : for every borelian set E of $\mathbb{R}^m$,

$$(X_1, \dots, X_m)^{-1}[E] = \{\omega : (X_1(\omega), \dots, X_m(\omega)) \in E\} = [(X_1, \dots, X_m) \in E]$$

is an element of $\mathcal{F}$.

The support of a random vector = the set of possible values of this random vector :

$$(X_1, \dots, X_m)[\Omega] = \{(X_1(\omega), \dots, X_m(\omega)) : \omega \in \Omega\}$$

Joint cumulative distribution function

$$F_{X_1,\dots,X_m}(t_1, \dots, t_m) = \Pr([X_1 \le t_1] \cap \dots \cap [X_m \le t_m])$$

Properties

$0 \le F(t_1, \dots, t_m) \le 1$ for every $(t_1, \dots, t_m)$

$F(t_1, \dots, t_m)$ is a non decreasing function of each variable t_j

$\lim_{t_j \to -\infty} F(t_1, \dots, t_m) = 0 \quad (j = 1, \dots, m)$

$\lim_{\substack{t_1 \to +\infty \\ \dots \\ t_m \to +\infty}} F(t_1, \dots, t_m) = 1$

⊕ : it is possible to construct the probability law $\Pr[(X_1, \dots, X_m) \in E]$ for any $E \in \mathcal{B}_m$ from the c.d.f.

Types of random vectors

- Discrete : the support $(X_1, \dots, X_m)[\Omega]$ is finite or denumerable, with

$$\Pr([X_1 = x_1] \cap \dots \cap [X_m = x_m]) = p_{x_1,\dots,x_m}$$

(and $\sum_{x_1} \dots \sum_{x_m} p_{x_1,\dots,x_m} = 1$)

Probability law : for any $E \in \mathcal{B}_m$

$$\Pr[\,(X_1, \dots, X_m) \in E] = \sum_{\{(x_1,\dots,x_m)\in E\}} p_{x_1,\dots,x_m}$$

- Continuous : the probabilities are continuously distributed via a density function $f_{(X_1,\dots,X_m)}(x_1, \dots, x_m) \geq 0$, with

$$\int_{-\infty}^{+\infty} dx_1 \dots \int_{-\infty}^{+\infty} dx_m\, f(x_1, \dots, x_m) = 1$$

and

$$\Pr\begin{pmatrix}[x_1 < X_1 \leq x_1 + dx_1] \cap \dots \\ \cap\, [x_m < X_m \leq x_m + dx_m]\end{pmatrix} \approx f(x_1, \dots, x_m) \cdot dx_1 \dots\, dx_m$$

Probability law : for any $E \in \mathcal{B}_m$

$$\Pr[\,(X_1, \dots, X_m) \in E] = \int \dots \int f(x_1, \dots, x_m)\, dx_1 \dots dx_m$$

(the integral is taken over the set E)

Joint c.d.f. :

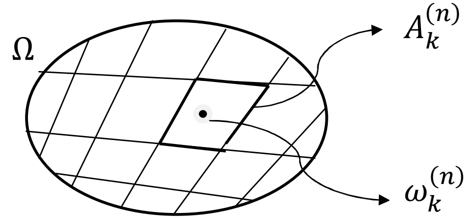
$$F(t_1, \dots, t_m) = \int_{-\infty}^{t_1} dx_1 \dots \int_{-\infty}^{t_m} dx_m\, f(x_1, \dots, x_m)$$

Expectation and moments

- Expectation
- Moments
- Variance
- Shape parameters
  - Skewness
  - Kurtosis
- Covariance and correlation
- Moment generating function
- Inequalities
  - Jensen's inequality
  - Markov's inequality
  - Chebyshev's inequality

Expectation

Generalization of the notion of integral



For every n,

- subdivisions of Ω : $\left\{A_1^{(n)}, A_2^{(n)}, \ldots, A_n^{(n)}\right\}$
- choice of $\omega_k^{(n)} \in A_k^{(n)} \quad (k = 1, 2, \ldots, n)$
- $p^{(n)} = \max\left\{\Pr\left(A_1^{(n)}\right), \ldots, \Pr\left(A_n^{(n)}\right)\right\}$

We then define

$$E(X) = \int_\Omega X(\omega)\, dPr(\omega)$$

$$= \lim_{\substack{n\to\infty \\ p^{(n)}\to 0}} \sum_{k=0}^{n} X\left(\omega_k^{(n)}\right) \cdot \Pr\left(A_k^{(n)}\right)$$

Interpretation : the expectation of a r.v. is a parameter localization (= wheighted mean of X)

Particular notation : mean = $E(X) = \mu$

For particular r.v.,

- Discrete r.v. : $E(X) = \sum x_i p_i$
- Continuous r.v. : $E(X) = \int_{-\infty}^{+\infty} x\, f_X(x)dx$
- Positive r.v. : $E(X) = \int_0^{+\infty} [1 - F(t)]dt$

A common notation : $E(X) = \int_{-\infty}^{+\infty} t\, dF_X(t)$

Generalization : for any function g,

$$E(g(X)) = \sum_i g(x_i)p_i = \int_{-\infty}^{+\infty} g(x)\, f_X(x)dx$$

Properties

a)The expectation is a linear operator

$$E(aX + bY + c) = aE(X) + bE(Y) + c$$

b) $E(XY) = ?$

Moments

Absolute moments

$$\mu_k' = E\left(X^k\right) \qquad k = 1, 2, \dots$$

Relative (or centered) moments

$$\mu_k = E\left((X - \mu)^k\right) \qquad k = 1, 2, \dots$$

In particular,

$$\mu_1' = E(X) = \mu$$
$$\mu_1 = 0$$

Variance

$$var(X) = \sigma^2 = \mu_2 = E((X-\mu)^2)$$

Developing,

$$\begin{aligned} var(X) &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - E^2(X) \end{aligned}$$

Interpretation : the variance is a dispersion parameter

Properties

a) $$\begin{aligned} var(aX+b) &= E\left(\left((aX+b)-(a\mu+b)\right)^2\right) \\ &= E((aX-a\mu)^2) \\ &= a^2 \cdot var(X) \end{aligned}$$

b) $var(X+Y) = ?$

Standard deviation : $\sigma = \sqrt{var(X)}$

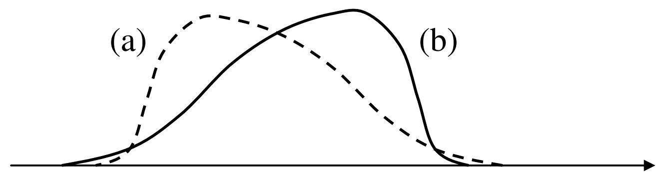
Shape parameters

Skewness

$$\gamma_1(X) = \frac{\mu_3}{\mu_2^{3/2}} = \frac{E((X-\mu)^3)}{\sigma^3}$$

Interpretation : γ_1 is a number without dimension and its sign indicates the type of dissymetry :

$\gamma_1(a) > 0$ $\gamma_1(b) < 0$

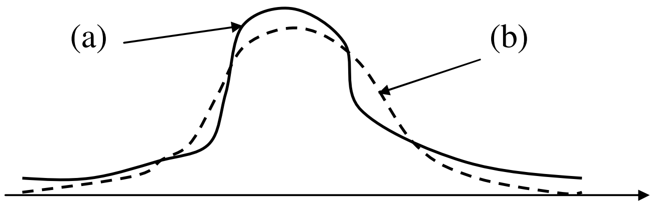


Kurtosis

$$\gamma_2(X) = \frac{\mu_4}{\mu_2^2} - 3 = \frac{E((X-\mu)^4)}{\sigma^4} - 3$$

Interpretation : γ_2 is a number without dimension and its value is indicative of the fatness of the distribution tails :

$$\gamma_2(a) > \gamma_2(b)$$



Covariance and correlation

Covariance

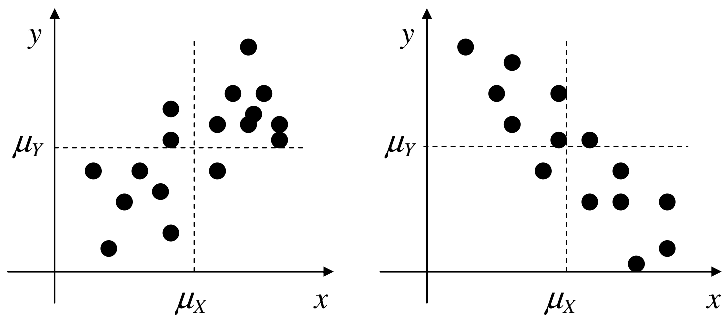
$$cov(X,Y) = \sigma_{X,Y} = E\big((X-\mu_X)(Y-\mu_Y)\big)$$

Developing,

$$\begin{aligned} cov(X,Y) &= E(XY - \mu_X Y - X\mu_Y + \mu_X\mu_Y) \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

Interpretation : the covariance measures (with its sign) the degree of linear dependence between two r.v. X and Y :

$cov(X,Y) > 0$ $\qquad$ $cov(X,Y) < 0$



Properties

a) Linearity

$$cov(aX + b, cY + d) = ac \cdot cov(X, Y)$$

$$cov(X + Y, Z) = cov(X, Z) + cov(Y, Z)$$

b) $\quad E(XY) = E(X) \cdot E(Y) + cov(X, Y)$

$$var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$$

Correlation coefficient

$$corr(X, Y) = \rho_{X,Y} = \frac{cov(X, Y)}{\sigma_X \cdot \sigma_Y}$$

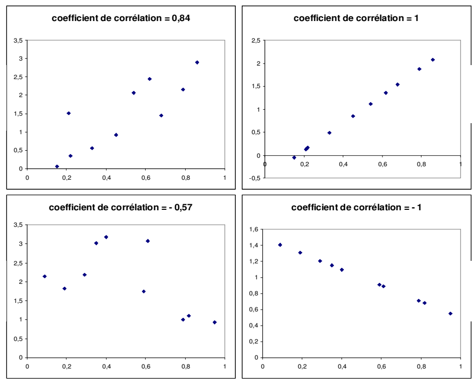
Interpretation : the correlation coefficient is a number without dimension that has the same interpretation as the covariance

Properties

a) $-1 \leq \rho \leq 1$

b) $\quad \rho_{X,Y} = \pm 1$ iff there exists a perfect linear relationship between X and Y

$\rho = 0{,}84$ $\quad$ $\rho = 1$



$\rho = -0{,}57$ $\quad$ $\rho = -1$

Moment generating function

$$m_X(t) = E(e^{tX})$$

For practical calculations,

$$m(t) = \sum_i e^{tx_i} p_i = \int_{-\infty}^{+\infty} e^{tx} f(x) dx$$

If we can derive "under the E sign",

$$m_X^{(k)}(t) = E\left(X^k e^{tX}\right)$$

Property

$$m_X^{(k)}(0) = E\left(X^k\right)$$

Inequalities

Jensen's inequality

If h is a convex function, then

$$E\big(h(X)\big) \geq h\big(E(X)\big)$$

[and the inverse inequality for a concave function]

Proof : for any x_0, there exists a straight line $y = ax + b$ such that

$$\begin{cases} h(x_0) = ax_0 + b \\ h(x) \geq ax + b \quad \forall x \end{cases}$$

Replacing x and x_0 respectively by X and $E(X)$, we get

$$\begin{cases} h\big(E(X)\big) = aE(X) + b \\ h(X) \geq aX + b \end{cases}$$

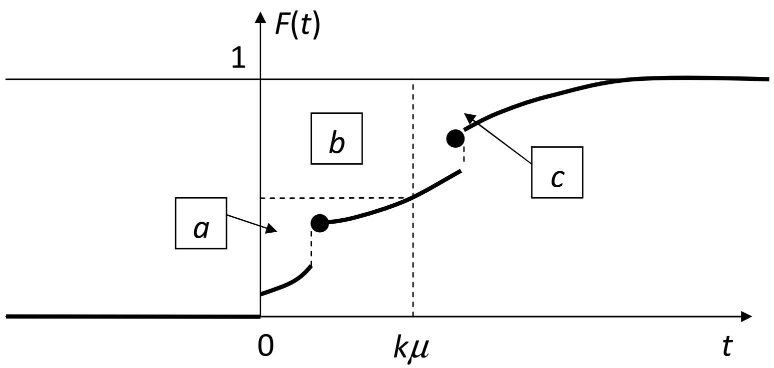
and then $E\big(h(X)\big) \geq aE(X) + b = h\big(E(X)\big)$

Markov's inequality

If X is a positive r.v. with mean μ, then

$$\Pr[X \geq k\mu] \leq \frac{1}{k} \qquad \forall k > 0$$

Proof :



$$\begin{aligned} b &= k\mu \cdot \Pr[X \geq k\mu] \\ &\leq a + b + c \\ &= \int_0^{+\infty} [1 - F(t)] dt \\ &= \mu \end{aligned}$$

Chebyshev's inequality

If X is a r.v. with mean μ and standard deviation σ, then

$$\Pr[|X - \mu| \geq h\sigma] \leq \frac{1}{h^2} \qquad \forall h > 0$$

Proof : the r.v. $Y = (X - \mu)^2$ is positive and $E(Y) = \sigma^2$

So, by Markov's inequality,

$$\Pr[(X - \mu)^2 \geq h^2\sigma^2] \leq \frac{1}{h^2}$$

Classical probability distributions

- Uniform distribution
  - Definition
  - Cumulative distribution function
  - Moments
- Normal distribution
  - Definition
  - Moments
  - Properties
  - Moment generating function
- Multinormal distribution
  - Definition
  - Properties
- Log-normal distribution
  - Definition
  - Density function
  - Moments
- Binomial distribution
  - Definition
  - Moment generating function
  - Moments
- Poisson distribution
  - Definition
  - Moment generating function
  - Moments
- Exponential distribution
  - Definition
  - Cumulative distribution function
  - Moments
  - Property

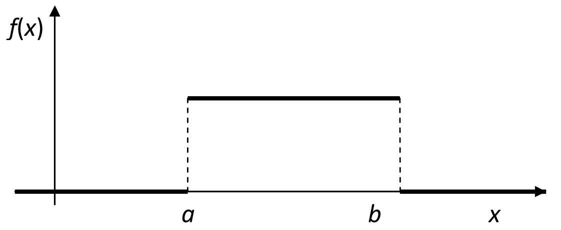
Uniform distribution

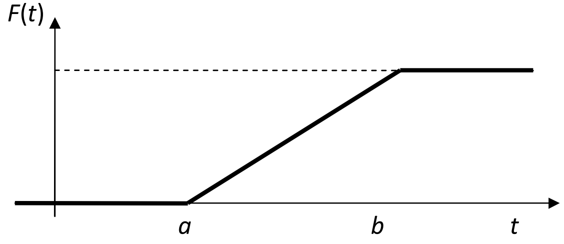
Definition : $X \sim \mathcal{U}(a;b)$ if $X[\Omega] = [a;b]$, $a < b$
and

$$f_X(x) = \frac{1}{b-a} \cdot \mathbf{1}_{[a;b]}(x)$$

Cumulative distribution function

$$F_X(t) = \begin{cases} 0 & if \quad t < a \\ \dfrac{t-a}{b-a} & if \quad a \le t \le b \\ 1 & if \quad t > b \end{cases}$$





Moments

$$E\left(X^k\right) = \frac{1}{b-a}\int_a^b x^k \, dx = \frac{b^{k+1} - a^{k+1}}{(k+1)(b-a)}$$

In particular,

$$E(X) = \frac{a+b}{2}$$

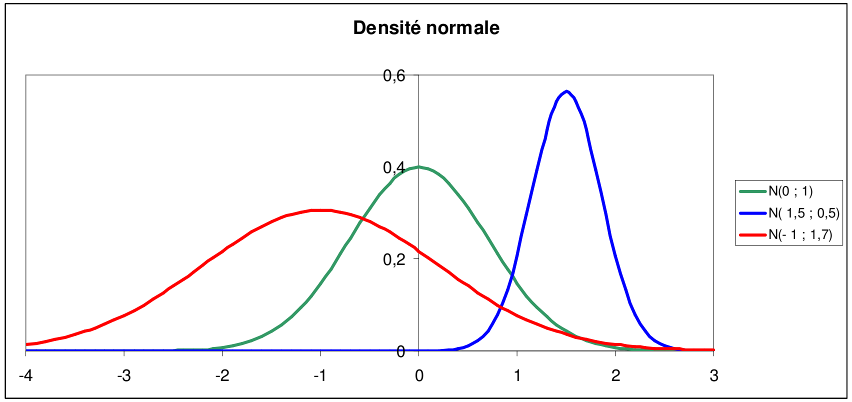
$$var(X) = \frac{(b-a)^2}{12}$$

Normal distribution

Definition : $X \sim \mathcal{N}(\mu; \sigma^2)$ if
$X[\Omega] = \mathbb{R}$, $\mu \in \mathbb{R}, \sigma > 0$ and

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

It is a density function (Poisson integral)



Moments

- Moments of odd order

$$E\left((X-\mu)^{2k+1}\right)$$
$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (x-\mu)^{2k+1} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$
$$= \frac{\sigma^{2k+2}}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} y^{2k+1}\, e^{-\frac{y^2}{2}} dy$$
$$= 0$$

Consequences

$$E(X) = \mu$$

(and then $E\left((X-\mu)^{2k+1}\right) = \mu_{2k+1}$)

$$\gamma_1 = 0$$

- Moments of even order

$$\mu_{2k} = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty}(x-\mu)^{2k}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$$
$$= \frac{\sigma^{2k+1}}{\sqrt{2\pi}\sigma}\int_{-\infty}^{+\infty} y^{2k}\, e^{-\frac{y^2}{2}}dy$$
$$= \frac{\sigma^{2k}}{\sqrt{2\pi}} I_k$$

$$I_k = \int_{-\infty}^{+\infty} y^{2k}\, e^{-y^2/2}dy$$
$$= \int_{-\infty}^{+\infty} y^{2k-1}\cdot ye^{-y^2/2}dy$$
$$= \int_{-\infty}^{+\infty} y^{2k-1}\cdot\left(-e^{-y^2/2}\right)'dy$$
$$= \left[-y^{2k-1}\cdot e^{-y^2/2}\right]_{y\to-\infty}^{y\to+\infty}$$
$$+(2k-1)\int_{-\infty}^{+\infty} y^{2k-2}\, e^{-y^2/2}dy$$
$$= (2k-1)\cdot I_{k-1}$$

$$I_k = (2k-1)I_{k-1}$$
$$= \cdots$$
$$= (2k-1)(2k-3)\ldots 1\cdot I_0$$
$$= \frac{(2k)!}{2^k k!}\cdot I_0$$

$$\mu_{2k} = \frac{\sigma^{2k}}{\sqrt{2\pi}} I_k$$
$$= \frac{\sigma^{2k}}{\sqrt{2\pi}}\cdot\frac{(2k)!}{2^k k!}\cdot I_0$$
$$= \sigma^{2k}\frac{(2k)!}{2^k k!}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy$$

So,

$$\mu_{2k} = \sigma^{2k}\frac{(2k)!}{2^k k!}$$

Consequences

$$var(X) = \mu_2 = \sigma^2$$

$$\mu_4 = \sigma^4\cdot 3 \qquad \Rightarrow \qquad \gamma_2 = 0$$

Properties

a) If $X \sim \mathcal{N}(\mu; \sigma^2)$, then

$$aX + b \sim \mathcal{N}(a\mu + b; a^2\sigma^2)$$

(the normal law is stable)

b) $X \sim \mathcal{N}(\mu; \sigma^2) \quad \Leftrightarrow \quad Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0; 1)$

(standard normal r.v.)

Classical notations

$$f_Z(x) = \varphi(x) \qquad F_Z(t) = \Phi(t)$$

Moment generating function

$$m(t) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{tx} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{1}{2\sigma^2}\left[x^2 - 2\mu x + \mu^2 - 2t\sigma^2 x\right]} dx$$

$$\begin{aligned}
x^2 - 2\mu x + \mu^2 - 2t\sigma^2 x \\
&= x^2 - 2x(\mu + t\sigma^2) + \mu^2 \\
&= x^2 - 2x(\mu + t\sigma^2) + (\mu + t\sigma^2)^2 \\
&\qquad -(\mu + t\sigma^2)^2 + \mu^2 \\
&= \left(x - (\mu + t\sigma^2)\right)^2 - 2t\mu\sigma^2 - t^2\sigma^4 \\
&= \left(x - (\mu + t\sigma^2)\right)^2 - 2\sigma^2\left(t\mu + \frac{t^2\sigma^2}{2}\right)
\end{aligned}$$

$$m(t) = e^{t\mu + \frac{t^2\sigma^2}{2}} \cdot \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{x-(\mu+t\sigma^2)}{\sigma}\right)^2} dx$$

$$m(t) = e^{t\mu + \frac{t^2\sigma^2}{2}}$$

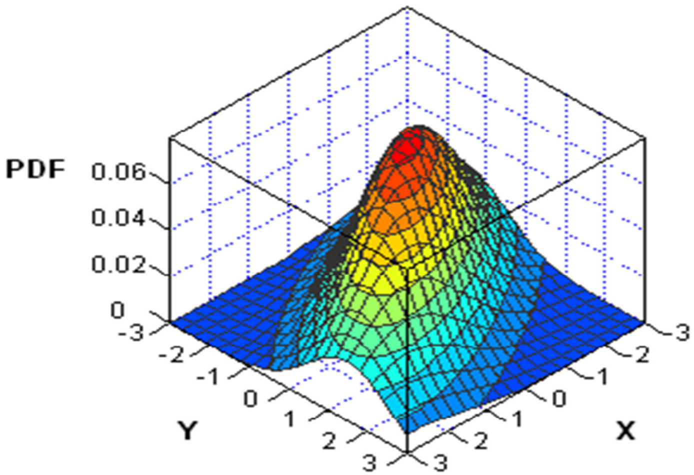
Multinormal distribution

Definition

A random vector $X = (X_1, X_2, \dots, X_m)$ is multinormal if any non trivial linear combination of its components is a normal r.v. :

For any $\alpha_1, \alpha_2, \dots, \alpha_m$, (at least one α_k is $\neq 0$), then

$$\sum_{k=1}^{m} \alpha_k X_k \sim \mathcal{N}$$



Properties

a) A random vector X is multinormal if and only if there exist

- a real vector μ
- a positive defined matrix V

such that the joint density function is given by

$$f_X(x) = f_{X_1, X_2, \dots, X_m}(x_1, x_2, \dots, x_m)$$
$$= \frac{1}{\sqrt{(2\pi)^m \cdot dtm(V)}} \cdot exp\left[-\frac{1}{2}(x-\mu)^t V^{-1}(x-\mu)\right]$$

where μ is the mean vector and V is the variance-covariance matrix :

$$\mu_k = E(X_k)$$
$$(V)_{jk} = cov(X_j, X_k)$$

b) The probability law of the random vector X is uniquely determined by the parameters μ and V

c) If the components of a multinormal random vector are uncorrelated, then they are independent

Proof : consider the density with a diagonal matrix V

d) If the components of a random vector are normally distributed, then the vector is non necessarily multinormal

Counter-example : if X is a normal r.v., consider the random vector

$$\begin{pmatrix} X \\ -X \end{pmatrix}$$

Log-normal distribution

Definition : $X \sim \mathcal{LN}(\mu; \sigma^2)$ if

$X[\Omega] = \mathbb{R}_0^+,\ \mu \in \mathbb{R}, \sigma > 0,$ and

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

Density function

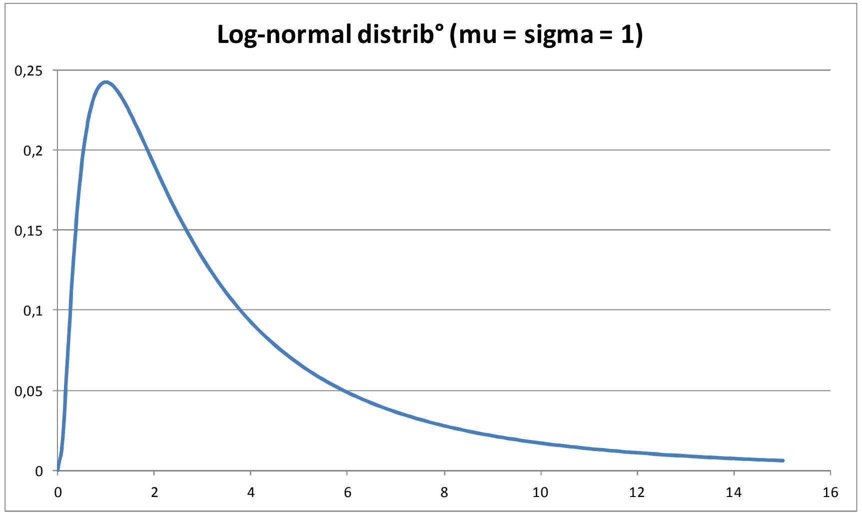
For $t > 0$,

$$\Pr[X \leq t] = \Pr[\ln X \leq \ln t] = F_N(\ln t)$$

And, for $x > 0$,

$$f_X(x) = \left(F_N(\ln x)\right)'_x$$
$$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right] \cdot \frac{1}{x}$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right] \cdot \mathbf{1}_{\mathbb{R}_0^+}(x)$$



Moments

$$E\left(X^k\right) = \frac{1}{\sqrt{2\pi}\sigma} \int_0^{+\infty} x^k \exp\left[-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2\right] \frac{dx}{x}$$

By using $y = \ln x$,

$$E\left(X^k\right) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{ky} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right] dy$$
$$= E\left(e^{kY}\right)$$

where $Y \sim \mathcal{N}(\mu; \sigma^2)$

$$E\left(X^k\right) = m_Y(k) = e^{k\mu + \frac{k^2\sigma^2}{2}}$$

In particular,

$$E(X) = e^{\mu + \frac{\sigma^2}{2}}$$

$$E(X^2) = e^{2\mu + 2\sigma^2}$$
$$\Longrightarrow \quad var(X) = e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right)$$

Binomial distribution

Definition : $X \sim \mathcal{B}(n; p)$ if $X[\Omega] = \{0, 1, \dots, n\}$, $n \in \mathbb{N}$, $p \in [0; 1]$ $(q = 1 - p)$ and

$$\Pr[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

It is a probability law (Newton formula)

Moment generating function

$$m(t) = \sum_{k=0}^{n} e^{tk} \cdot \binom{n}{k} p^k q^{n-k}$$
$$= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k q^{n-k}$$
$$= (pe^t + q)^n$$

Moments

Derivatives of the m.g.f.

$$m'(t) = npe^t (pe^t + q)^{n-1}$$
$$m''(t) = npe^t (npe^t + q)(pe^t + q)^{n-2}$$
$$\dots$$

$$\mu'_1 = E(X) = m'(0) = np$$
$$\mu'_2 = E(X^2) = m''(0) = np(np + q)$$
$$\dots$$

$$E(X) = np$$
$$var(X) = npq$$
$$\gamma_1(X) = \frac{q - p}{\sqrt{npq}}$$
$$\gamma_2(X) = \frac{1 - 6pq}{npq}$$

Poisson distribution

Definition : : $X \sim \mathcal{P}(\lambda)$ if $X[\Omega] = \mathbb{N}$, $\lambda > 0$ and

$$\Pr[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}$$

It is a probability law (expansion of e^{λ})

Moment generating function

$$m(t) = \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!}$$

$$= e^{\lambda(e^t - 1)}$$

Moments

Derivatives of the m.g.f.

$$m'(t) = \lambda e^t e^{\lambda(e^t-1)}$$
$$m''(t) = \lambda e^t (1 + \lambda e^t) e^{\lambda(e^t-1)}$$
$$\dots$$

$$\mu_1' = E(X) = m'(0) = \lambda$$
$$\mu_2' = E(X^2) = m''(0) = \lambda(1 + \lambda)$$
$$\dots$$

$$E(X) = \lambda$$
$$var(X) = \lambda$$
$$\gamma_1(X) = \frac{1}{\sqrt{\lambda}}$$
$$\gamma_2(X) = \frac{1}{\lambda}$$

Exponential distribution

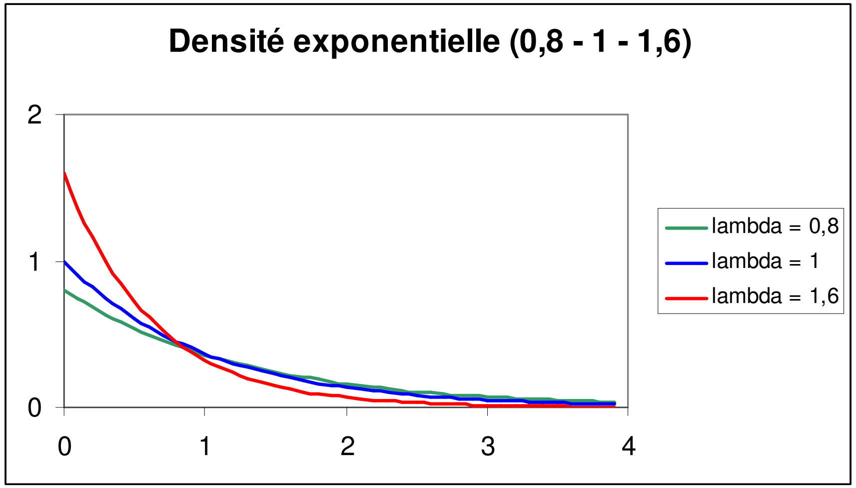
Definition : : $X \sim \mathcal{E}(\lambda)$ if $X[\Omega] = \mathbb{R}^+$, $\lambda > 0$ and

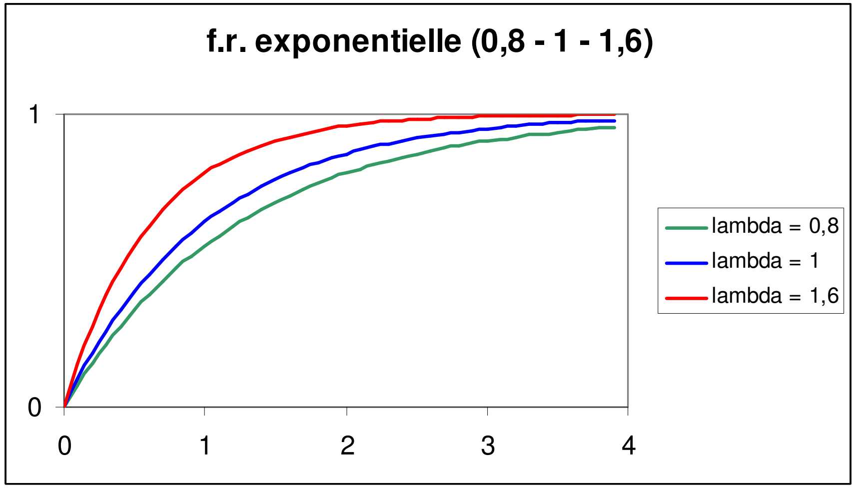
$$f_X(x) = \lambda e^{-\lambda x} \cdot \mathbf{1}_{\mathbb{R}^+}(x)$$

It is a probability law

Cumulative distribution function

$$F_X(t) = \left(1 - e^{-\lambda t}\right) \cdot \mathbf{1}_{\mathbb{R}^+}(t)$$





Moments

$$E(X^k) = \lambda \int_0^{+\infty} x^k e^{-\lambda x}\, dx = \frac{k!}{\lambda^k}$$

In particular,

$$E(X) = \frac{1}{\lambda}$$

$$var(X) = \frac{1}{\lambda^2}$$

Property

The exponential r.v. has "no memory" : for $s, t > 0$,

$$\Pr([X > s + t] | [X > s]) = \frac{\Pr[X > s + t]}{\Pr[X > s]} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \Pr[X > t]$$

Independence

- Conditional probability
- Independence
  - Independence of two events
  - Independence of two sub-σ-fields
  - Independence of two r.v.
- Properties

Conditional probability

Let A and B be elements of $\mathcal{F}$

Probability of A in the restricted set of possible outcomes B, denoted by $\Pr(A|B)$

$$\begin{cases} \Pr(A|B) = k \cdot \Pr(A \cap B) \\ \Pr(B|B) = 1 \end{cases}$$

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Independence

Independence of two events

The probability of A is not affected by the occurrence of B :

$$\Pr(A|B) = \Pr(A)$$

Definition : $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$

Independence of two sub-σ-fields

Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two sub-σ-fields of $\mathcal{F}$

$\mathcal{F}_1$ and $\mathcal{F}_2$ are independent if, for every $E_1 \in \mathcal{F}_1$ and $E_2 \in \mathcal{F}_2$, E_1 and E_2 are independent :

$$\Pr(E_1 \cap E_2) = \Pr(E_1) \cdot \Pr(E_2)$$

Independence of two r.v.

The r.v. X_1 and X_2 are independent if $\sigma(X_1)$ and $\sigma(X_2)$ are independent

Property

The r.v. X_1 and X_2 are independent if and only if

$$\Pr([X_1 \leq t_1] \cap [X_2 \leq t_2]) \\ = \Pr[X_1 \leq t_1] \cdot \Pr[X_2 \leq t_2]$$

i.e.

$$F_{X_1,X_2}(t_1,t_2) = F_{X_1}(t_1) \cdot F_{X_2}(t_2)$$

Properties

(without proofs)

a) If X and Y are independent, then

$$cov(X,Y) = 0$$
$$E(XY) = E(X) \cdot E(Y)$$
$$var(X+Y) = var(X) + var(Y)$$

The reciprocal is not true :

	X		
Y	-1	0	1
0	$1/4$	.	$1/4$
1	.	$1/2$	.

- The two r.v. are not independent (why ?)
- $E(XY) = E(X) \cdot E(Y) = 0$

b) The r.v. $X_1, \dots, X_m$ are independent iff

$$F_{X_1,\dots,X_m}(t_1, \dots, t_m) = F_{X_1}(t_1) \cdot \dots \cdot F_{X_m}(t_m)$$

c) The r.v. $X_1, \dots, X_m$ are independent iff

$$f_{X_1,\dots,X_m}(t_1, \dots, t_m) = f_{X_1}(t_1) \cdot \dots \cdot f_{X_m}(t_m)$$

d) If he r.v. $X_1, \dots, X_m$ are independent, then

$$m_{X_1+\dots+X_m}(t) = m_{X_1}(t) \cdot \dots \cdot m_{X_m}(t)$$

e1) If $X_1, \dots, X_m$ are independent r.v. with $X_j \sim \mathcal{B}(n_j; p)$, then

$$\sum_{j=1}^{m} X_j \sim \mathcal{B}\left(\Sigma n_j; p\right)$$

e2) If $X_1, \dots, X_m$ are independent r.v. with $X_j \sim \mathcal{P}(\lambda_j)$, then

$$\sum_{j=1}^{m} X_j \sim \mathcal{P}\left(\Sigma \lambda_j\right)$$

e3) If $X_1, \dots, X_m$ are independent r.v. with $X_j \sim \mathcal{N}(\mu_j; \sigma_j^2)$, then

$$\sum_{j=1}^{m} X_j \sim \mathcal{N}\left(\Sigma \mu_j; \Sigma \sigma_j^2\right)$$

Conditional expectation

- w.r.t. an event
  - Intuitively
  - Definition
  - Property
- w.r.t. a partition of Ω
  - Definition
  - w.r.t. a discrete r.v.
  - Property
- w.r.t. a σ-field (general case)
  - Definition
  - w.r.t. a r.v.
  - Rules for handling the conditional expectation
  - Projection property
- Conditional variance
  - Definition
  - Properties

w.r.t. an event

Let us consider a r.v. X such that $E(|X|)$ is finite

Intuitively

Let A be an event with $\Pr(A) > 0$

If X is discrete, we want to define

$$E(X|A) = \sum_k x_k \Pr([X = x_k]|A)$$

We can introduce the conditional c.d.f.

$$F_X(t|A) = \Pr([X \leq t]|A)$$

that has the same properties as the ordinary c.d.f. and "define"

$$E(X|A) = \int_{-\infty}^{+\infty} t dF_X(t|A)$$

Definition

Let us consider the indicator r.v. of the event A

$$\mathbf{1}_A : \omega \mapsto \begin{cases} 1 & if\ \omega \in A \\ 0 & if\ \omega \notin A \end{cases}$$

We define

$$E(X|A) = \frac{E(X \cdot \mathbf{1}_A)}{\Pr(A)}$$

Coherence with the intuitive definition for a discrete X ?

$$X \cdot \mathbf{1}_A : \omega \longmapsto \begin{cases} 0 & if\ \omega \notin A \\ x_k & if\ \omega \in A\ \ and\ \ X(\omega) = x_k \end{cases}$$

so that

$$E(X \cdot \mathbf{1}_A) = 0 + \sum_k x_k \Pr([X = x_k] \cap A)$$

Property

$$E(X \cdot \mathbf{1}_A) = E(E(X|A) \cdot \mathbf{1}_A)$$

Proof

The r.h.s. is equal to

$$E(X|A) \cdot E(\mathbf{1}_A) = E(X|A) \cdot \Pr(A) = E(X \cdot \mathbf{1}_A)$$

w.r.t. a partition of Ω

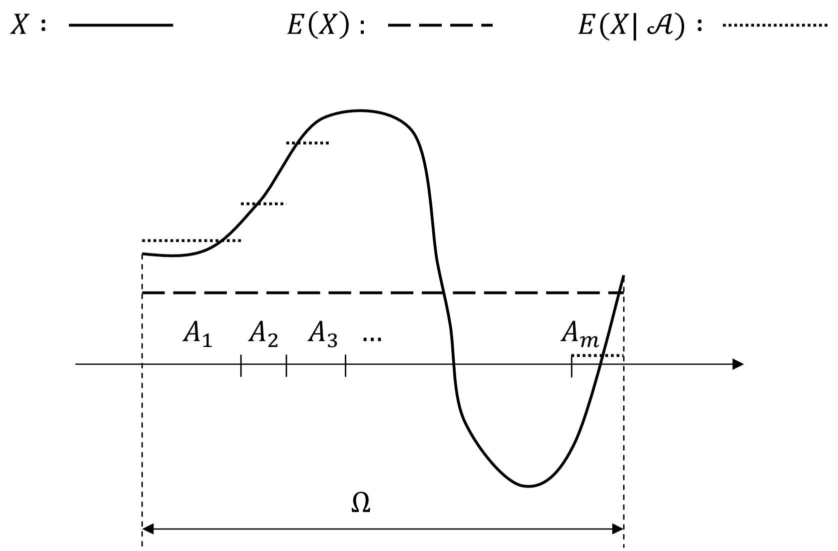
Definition

Let $\mathcal{A} = \{A_1, \dots, A_n, \dots\}$ a (discrete) partition of Ω with $\Pr(A_i) > 0 \quad \forall i$

We define the conditional expectation as the r.v.

$$E(X|\,\mathcal{A}) \;:\; \omega \longmapsto E(X|A_k) \qquad if\ \omega \in A_k$$

Graphical representation for $\Omega \subset \mathbb{R}$:



w.r.t. a discrte r.v.

Let Y be a discrete r.v.

We define the conditional expectation as the r.v.

$$E(X|Y) \;:\; \omega \mapsto E(X|[Y = y_k]) \qquad if\ Y(\omega) = y_k$$

Note : $E(X|Y)$ does not depend directly on the values y_k of Y, but on the generated partition $\{[Y = y_k] : k = 1,2, \dots\}$ of Ω

Property

For any union A of some elements of the partition $\mathcal{A} = \{A_1, \dots, A_n, \dots\}$, then

$$E(X \cdot \mathbf{1}_A) = E(E(X|\mathcal{A}) \cdot \mathbf{1}_A)$$

Proof

Denoting (k) the index values in the union A, the r.v. $E(X|\mathcal{A}) \cdot \mathbf{1}_A$ is defined by

$$\omega \mapsto \begin{cases} 0 & if\ \omega \notin A \\ E(X|A_k) & if\ \omega \in A_k\ for\ some\ (k) \end{cases}$$

And the r.h.s. is equal to

$$0 + \sum_{(k)} E(X|A_k) \cdot \Pr(A_k) = \sum_{(k)} E\left(X \cdot \mathbf{1}_{A_k}\right)$$
$$= E\left(X \cdot \sum_{(k)} \mathbf{1}_{A_k}\right) = E(X \cdot \mathbf{1}_A)$$

w.r.t. a σ-field (general case)

Definition

Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$

We define the conditional expectation as the r.v., denoted by $E(X|\mathcal{G})$ such that $\sigma\big(E(X|\mathcal{G})\big) \subset \mathcal{G}$ and

$$\forall A \in \mathcal{G} \qquad E(X \cdot \mathbf{1}_A) = E(E(X|\mathcal{G}) \cdot \mathbf{1}_A)$$

It is possible to proof that such a r.v. exists and is "unique" : there may exist several r.v. Z and Z' satisfying this property : $\sigma(Z)$, $\sigma(Z') \subset \mathcal{F}$ and
$$\forall A \in \mathcal{G}, E(X \cdot \mathbf{1}_A) = E(Z \cdot \mathbf{1}_A) = E(Z' \cdot \mathbf{1}_A)$$
but then, $\Pr[Z \neq Z'] = 0$

Thus, from now on, we would have to consider

- r.v. defined outside of an event with null probability (a "version" of the r.v.)
- equalities "almost sure" between r.v.

w.r.t. a r.v.

Let Y be a r.v.

We define the conditional expectation as the r.v.

$$E(X|Y) = E\big(X|\sigma(Y)\big)$$

Note : as $E(X|\mathcal{G})$ is $\mathcal{G}$-measurable, $E(X|Y)$ is a function of Y.

Rules for handling the conditional expectation

(R0) If $X \geq 0$, then $E(X|\mathcal{G}) \geq 0$

(R0') If $X_1 \leq X_2$, then $E(X_1|\mathcal{G}) \leq E(X_2|\mathcal{G})$

(R1) The conditional expectation is a linear operator :

$$E(aX + bY + c|\mathcal{G}) = aE(X|\mathcal{G}) + bE(Y|\mathcal{G}) + c$$

Proof : for any $A \in \mathcal{G}$,

$$\begin{aligned} & E\big((aX + bY + c) \cdot \mathbf{1}_A\big) \\ &= aE(X \cdot \mathbf{1}_A) + bE(Y \cdot \mathbf{1}_A) + cE(\mathbf{1}_A) \\ &= aE(E(X|\mathcal{G}) \cdot \mathbf{1}_A) + bE(E(Y|\mathcal{G}) \cdot \mathbf{1}_A) + cE(\mathbf{1}_A) \\ &= E\big((aE(X|\mathcal{G}) + bE(Y|\mathcal{G}) + c) \cdot \mathbf{1}_A\big) \end{aligned}$$

(R2) $E\big(E(X|\mathcal{G})\big) = E(X)$

Proof : definition with $A = \Omega$

(R3) If X and $\mathcal{G}$ are independent [$\equiv$ $\sigma(X)$ and $\mathcal{G}$ independent], then

$$E(X|\mathcal{G}) = E(X)$$

Proof : for any $A \in \mathcal{G}$,

$$E(X \cdot \mathbf{1}_A) = E(X) \cdot E(\mathbf{1}_A) = E(E(X) \cdot \mathbf{1}_A)$$

(R4) If $\sigma(X) \subset \mathcal{G}$ [X is $\mathcal{G}$-measurable], then

$$E(X|\mathcal{G}) = X$$

(X is considered as a constant w.r.t. $\mathcal{G}$)

Proof : X is a $\mathcal{G}$-measurable r.v. for which the definition is satisfied

(R5) Generalization of (R4) "taking out what is known" : if $\sigma(X) \subset \mathcal{G}$, then for any r.v. Y,

$$E(XY|\mathcal{G}) = X \cdot E(Y|\mathcal{G})$$

(R6) Tower property : if $\mathcal{H}$ is a sub-σ-field of $\mathcal{G}$, then

$$E(E(X|\mathcal{G})|\,\mathcal{H}) = E(X|\mathcal{H})$$

Proof : for any $A \in \mathcal{H}$,

$$\begin{aligned} E\{E[E(X|\mathcal{G})|\mathcal{H}] \cdot \mathbf{1}_A\} &= E\{E[E(X|\mathcal{G}) \cdot \mathbf{1}_A|\mathcal{H}]\} \\ &= E\{E[E(X \cdot \mathbf{1}_A|\mathcal{G})|\mathcal{H}]\} \\ &= E\big(E(X \cdot \mathbf{1}_A|\mathcal{G})\big) \\ &= E(X \cdot \mathbf{1}_A) \end{aligned}$$

But

$$\begin{aligned} E(E(X|\mathcal{H}) \cdot \mathbf{1}_A) &= E\big(E(X \cdot \mathbf{1}_A|\mathcal{H})\big) \\ &= E(X \cdot \mathbf{1}_A) \end{aligned}$$

(R7) Generalization of (R3) : if X is independent of $\mathcal{G}$ and if Y is $\mathcal{G}$-measurable, then

$$E(h(X,Y)|\mathcal{G}) = E(E_X(h(X,Y))|\mathcal{G})$$

where $E_X\big(h(X,Y)\big)$ means that

- we fix Y, and
- we take the expectation w.r.t. X

(without proof)

(R8) Jensen inequality : if h is a convex function, then

$$E(h(X)|\mathcal{G}) \geq h\big(E(X|\mathcal{G})\big)$$

Proof : for any x_0, there exists a straight line $y = ax + b$ such that

$$\begin{cases} h(x_0) = ax_0 + b \\ h(x) \geq ax + b \qquad \forall x \end{cases}$$

Replacing x and x_0 respectively by X and $E(X|\mathcal{G})$, we get

$$\begin{cases} h\big(E(X|\mathcal{G})\big) = aE(X|\mathcal{G}) + b \\ h(X) \geq aX + b \qquad (*) \end{cases}$$

Taking conditional expectation of (*),

$$E(h(X)|\mathcal{G}) \geq aE(X|\mathcal{G}) + b = h\big(E(X|\mathcal{G})\big)$$

Projection property

This property shows that $E(X|\mathcal{G})$ is an "updated version of $E(X)$", given the information in $\mathcal{G}$

Let $L^2(\mathcal{G})$ be the collection of r.v. Y such that $\sigma(Y) \subset \mathcal{G}$ and $E(Y^2)$ is finite (more than $E(|Y|)$ finite)

Projection property : If X is such that $E(X^2)$ is finite, then $E(X|\mathcal{G})$ is the element of $L^2(\mathcal{G})$ which is closest to X in the mean square sense :

$$\min_{Y \in L^2(\mathcal{G})} E((X - Y)^2) = E\left(\big(X - E(X|\mathcal{G})\big)^2\right)$$

Proof : for any $Y \in L^2(\mathcal{G})$,

$$\begin{aligned} E((X - Y)^2) &= E((X - E(X|\mathcal{G}) + E(X|\mathcal{G}) - Y)^2) \\ &= E\left(\big(X - E(X|\mathcal{G})\big)^2\right) \\ &\quad + E((E(X|\mathcal{G}) - Y)^2) \\ &\quad + 2E\big[\big(X - E(X|\mathcal{G})\big) \cdot (E(X|\mathcal{G}) - Y)\big] \end{aligned}$$

But

$$
\begin{aligned}
&E\big[\big(X - E(X|\mathcal{G})\big)\cdot(E(X|\mathcal{G}) - Y)\big] \\
&\quad = E\big\{E\big[\big(X - E(X|\mathcal{G})\big)\cdot(E(X|\mathcal{G}) - Y)|\mathcal{G}\big]\big\} \\
&\quad = E\big\{(E(X|\mathcal{G}) - Y)\cdot\ E\big[\big(X - E(X|\mathcal{G})\big)|\mathcal{G}\big]\big\} \\
&\quad = \ E\big\{(E(X|\mathcal{G}) - Y)\cdot\big[\big(E(X|\mathcal{G}) - E(X|\mathcal{G})\big)\big]\big\} \\
&\quad = 0
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&E((X - Y)^2) \\
&\quad = \ E\left(\big(X - E(X|\mathcal{G})\big)^2\right) + E((E(X|\mathcal{G}) - Y)^2) \\
&\quad \geq E\left(\big(X - E(X|\mathcal{G})\big)^2\right)
\end{aligned}
$$

And we have equality for $Y = E(X|\mathcal{G})$

Conditional variance

Definition

$$
var(X|\mathcal{G}) = E\left(\big(X - E(X|\mathcal{G})\big)^2\ |\mathcal{G}\right)
$$

Properties

- $var(X|\mathcal{G}) = E(X^2|\mathcal{G}) - E^2(X|\mathcal{G})$

$$
\begin{aligned}
var(X|\mathcal{G}) &= E(X^2\ |\mathcal{G}) - 2E(X\cdot E(X|\mathcal{G})\ |\mathcal{G}) \\
&\quad + E(E^2(X|\mathcal{G})\ |\mathcal{G}) \\
&= E(X^2\ |\mathcal{G}) - 2\ E(X|\mathcal{G})\cdot E(X|\mathcal{G}) \\
&\quad + E^2(X|\mathcal{G})
\end{aligned}
$$

- $var(X) = E\big(var(X|\mathcal{G})\big) + var\big(E(X|\mathcal{G})\big)$

$$
E\big(var(X|\mathcal{G})\big) = E(X^2) - E\big(E^2(X|\mathcal{G})\big)
$$

$$
\begin{aligned}
var\big(E(X|\mathcal{G})\big) &= E\big(E^2(X|\mathcal{G})\big) - E^2\big(E(X|\mathcal{G})\big) \\
&= E\big(E^2(X|\mathcal{G})\big) - E^2(X)
\end{aligned}
$$

Stochastic convergences

- Definitions
  - Almost sure convergence
  - Convergence in quadratic mean
  - Convergence in probability
  - Convergence in distribution
- Properties
- Limit theorems and approximations
  - Law of large numbers
  - Central limit theorem
  - Approximations of the binomial law

Definitions

What does "$X_n \longrightarrow X$" mean ?

Almost sure convergence : $X_n \xrightarrow{a.s.} X$

$$\Pr\left[\lim_{n \longrightarrow \infty} X_n = X\right] = 1$$

Convergence in quadratic mean : $X_n \xrightarrow{q.m.} X$

$$\lim_{n \longrightarrow \infty} E((X_n - X)^2) = 0$$

Convergence in probability : $X_n \xrightarrow{pr} X$

$$\forall \varepsilon > 0, \quad \lim_{n \longrightarrow \infty} \Pr[|X_n - X| > \varepsilon] = 0$$

Convergence in distribution : $X_n \xrightarrow{d} X$
(or convergence in law, or weak convergence)

$$\forall t \ : \ F_X(t) \ continuous, \quad \lim_{n \to \infty} F_{X_n}(t) = F_X(t)$$

Properties

(without proofs)

a) The convergence in distribution is equivalent to these two statements :

- For any continuous and bounded function h,

$$\lim_{n\to\infty} E\big(h(X_n)\big) = E\big(h(X)\big)$$

- $\lim_{n\to\infty} m_{X_n}(t) = m_X(t) \quad \forall t$

b) $X_n \xrightarrow{a.s.} X$

$$\Rightarrow \quad X_n \xrightarrow{pr} X \quad \Rightarrow \quad X_n \xrightarrow{d} X$$

$X_n \xrightarrow{q.m.} X$

Limit theorems and approximations

(without proofs)

Law of large numbers

If $X_1, X_2, \dots, X_n, \dots$ is a sequence of i.i.d. r.v. with finite mean μ, then, when $n \to \infty$,

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{a.s.} \mu$$

Particular case : let A be an event and $f_n(A)$ the proportion of occurrences of A for n independent realizations of the random situation ; then,

$$f_n(A) \xrightarrow{a.s.} \Pr(A)$$

Central limit theorem

If $X_1, X_2, \ldots, X_n, \ldots$ is a sequence of i.i.d. r.v. with finite mean μ and variance σ^2, then, when $n \to \infty$,

$$\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\,\sigma} = \frac{\frac{1}{n}(X_1 + \cdots + X_n) - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathcal{N}(0; 1)$$

Interpretation : with the former hypotheses, if n is "sufficiently large", then

$$X_1 + \cdots + X_n \sim \mathcal{N}(n\mu;\ n\sigma^2)$$

Approximations of the binomial law

a) Poisson approximation

If $n \to \infty$, $p \to 0$ and $np \to \lambda$ (> 0), then

$$\mathcal{B}(n; p) \xrightarrow{d} \mathcal{P}(\lambda)$$

b) Normal approximation

If $n \to \infty$ and fixed p, then

$$\frac{\mathcal{B}(n; p) - np}{\sqrt{np(1-p)}} \xrightarrow{d} \mathcal{N}(0; 1)$$

Interpretation : with the former hypotheses, if n is "sufficiently large" and p not too close to 0 and 1, then

$$\mathcal{B}(n; p) \sim \mathcal{N}\big(np; np(1-p)\big)$$