Conditional probability and independence

Given two events E, A:

Conditional probability of the E given A:

$$P(E|A) = P(E \cap A)/P(A),$$

provided $P(A) > 0$.

E is independent of A iff $P(E|A) = P(E)$.

For more than two events, say E, A, B:

- if $P(E|A) = P(E)$ we can **only** say A is independent of E
- to have **complete independence** of E, A, B we need to have

$$\left\{\begin{aligned} P(E \cap A) &= P(E)P(A), \quad P(A \cap B) = P(A)P(B), \\ P(E \cap B) &= P(E)P(B) \\ P(E \cap A \cap B) &= P(E)P(A)P(B). \end{aligned}\right.$$

Random variables

The triple $(\mathcal{S}, \mathcal{A}, P)$ is called probability space and is all we need to compute the probability of any event.

However,

- $\mathcal{S}$ is an abstract set, i.e. it contains objects of any kind (faces of a die, faces of a coin, etc.)
- in statistics we deal with data, i.e. numbers s.t. 1.2 kW/h, 10 defective items

Thus, the question: How do we conjugate sample spaces and events to data?

$\Downarrow$

Answer: by the concept of a random variable (r.v.).

Random variables (cont'd)

Random variable: mapping $X : \mathcal{S} \to \mathbb{R}$ that assigns a real number $X(s)$ to s, for all $s \in \mathcal{S}$.

Example 5 (Single die problem)

When we say "the probability of an odd number equals 1/2", we are using the r.v.

$$X = \begin{cases} 1 & \text{if } s = \text{⚀} \\ 2 & \text{if } s = \text{⚁} \\ \vdots & \\ 6 & \text{if } s = \text{⚅} \end{cases}$$

Probability of a random variable

In a triple $(\mathcal{S}, \mathcal{A}, P)$, to each $s \in \mathcal{S}$ there is associated a probability, through P.

Thus, there is a probability associated to each $X(s)$ and, if we have a subset B of reals, using P we can compute the **probability that** X **takes values in** B.

Formally, let $B \subseteq \mathbb{R}$ then

$$P(X \in B) = P(\{s : X(s) \in B\}).$$

X is continuous if $P(X = x) = 0$ for all $x \in \mathbb{R}$.
X is discrete if $P(X = x) > 0$ for all $x \in \mathbb{R}$ in the range of X.[6]

[6] If X is continuous, its rage is uncountable; if X is discrete, its range is countable; X can also be mixed, discrete and continuous.

The probability (density) function

Let $\mathcal{X} = \{x_1, x_2, \ldots\}$, be the range of X

Probability (density) function (pdf) of X: $p(x) = P(X = x)$, $\forall x \in \mathcal{X}$.

Example 6 (Singe die problem)

Let X, s.t. $X = -1$ if the die shows less than three dots, $X = 0$ if it shows three dots and $X = 1$ if it shows more than three dots. We have

$$\begin{aligned} P(X=-1) &= P(\{s : X(s) = -1\}) = P(\{\text{⚀}, \text{⚁}\}) \\ &= P(\{\text{⚀}\}) + P(\{\text{⚁}\}) = 1/3, \end{aligned}$$

and the pdf of X is $p(x) = \begin{cases} 1/3 & \text{if } x = -1 \\ 1/6 & \text{if } x = 0 \\ 1/2 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$.

$\{1,2\}$

$\{3\}$

$\{4,5,6\}$

The distribution function

(Cumulative) distribution function (df) of X: $\boxed{F(x) = P(X \leq x)}$, $\forall x \in \mathbb{R}$.[7]

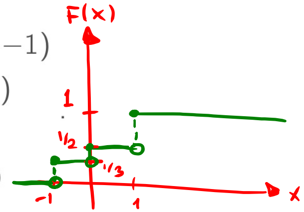
$F(+\infty) = 1$
Always
$F(-\infty) = 0$

Example 7 (Example 6 cont'd)

Considering that

$$F(-1) = P(X \leq -1) = 1/3, \quad F(0) = P(X \leq 0) = 3/6, \quad \text{then}$$

$$F(x) = \begin{cases} 0 & \text{if } x \in (-\infty, -1) \\ 1/3 & \text{if } x \in [-1, 0) \\ 3/6 & \text{if } x \in [0, 1) \\ 1 & \text{if } x \in [1, \infty) \end{cases}$$



Thus F is right-continuous and has jumps at $-1, 0, 1$ and is continuous everywhere.

[7] Yes, F is defined for all $x \in \mathbb{R}$.

Continuous r.v.

For a continuous r.v. if there exits a function $f : \mathbb{R} \to \mathbb{R}_{\geq 0}$, s.t.

$$\boxed{\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1,}$$

and for every $a \leq b$,

$$\boxed{P(a < X < b) = \int_a^b f(x)\mathrm{d}x.}$$

then f is the pdf of X (probability density function)

The df of X: $F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t, \quad \forall x \in \mathbb{R}.$

pdf and df are related:

$$f(x) = \partial F(x)/\partial x, \text{ at all continuity points of } F.$$

Properties of a df

Given a function F how can we be sure it is a df?

The df has the following properties:

(i) F is nondecreasing;

(ii) F is continuous from the right; (vedi grafico esempio)

(iii) $\lim_{x \to -\infty} F(x) = 0$;

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

Some further properties:

(a) $P(X = x) = F(x) - F(x^-)$, where $F(x^-) = \lim_{y \to x^-} F(y)$ since it's not necessarily continuous from the left

(b) $P(x < X \leq y) = F(y) - F(x)$

(c) $P(X > x) = 1 - F(x)$;

(d) $F(b) - F(a) = P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b)$, for a continuous X.

The quantile function

For a rv X with df F, the quantile function or inverse df is defined by

$$Q(p) = \inf\{x : F(x) \geq p\}, \quad \forall p \in (0, 1).$$

smallest x that has F(x) > p

$Q(1/4)$ is called the first quartile,
$Q(1/2)$ is the second quartile or the median,
$Q(3/4)$ is third quartile. In general

$$\xi_p = Q(p), \quad \text{for any} \quad p \in (0, 1),$$

is called the pth quantile of X.[8]

[8] We may assign $Q(0) = -\infty$ and $Q(1) = \infty$.

Notable rv's

- Bernoulli: $f(x) = \theta^x(1-\theta)^{1-x}$, $x = 0, 1$, $\theta \in [0, 1]$, notation: $\mathrm{Ber}(p)$;
- Binomial: $f(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$, $x = 0, 1, \ldots, n$, $\theta \in [0, 1]$, notation: $\mathrm{Bin}(n, p)$;
- Negative Binomial: $f(x) = \binom{x+r-1}{x}\theta^r(1-\theta)^x$, $x, r \in \mathbb{Z}_{\geq 0}$, notation: $\mathrm{NegBin}(r, p)$; If $r = 1$ it's called geometrical dist.
- Poisson: $f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$, y$\in \mathbb{Z}_{\geq 0}$, $\lambda > 0$, notation $\mathrm{Poi}(\lambda)$;

DISC

CONT

- Gaussian: $f(x) = e^{-\frac{(x-\mu)^2}{2\sigma^2}}/(\sqrt{2\pi}\sigma)$, $x, \mu \in \mathbb{R}$, $\sigma^2 > 0$, notation: $\mathrm{N}(\mu, \sigma^2)$;
- Exponential: $f(x) = \lambda e^{-\lambda x}$, $y \in \mathbb{R}_{\geq 0}$, $\lambda > 0$, notation: $\mathrm{Exp}(\lambda)$;
- Gamma: $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}$, $y \in \mathbb{R}_{\geq 0}$, $\alpha > 0, \lambda > 0$, notation: $\mathrm{Ga}(\alpha, \lambda)$
- Weibull: $f(x) = \frac{\alpha}{\beta}x^{\alpha-1}e^{-\frac{x^\alpha}{\beta}}$, $y \in \mathbb{R}_{\geq 0}$, $\alpha > 0$, $\beta > 0$, notation: $\mathrm{Wei}(\alpha, \beta)$;
- Uniform: $f(x) = (b-a)^{-1}\mathbf{1}_{[a,b]}$, $a, b \in \mathbb{R}$, $a < b$ notation: $\mathrm{Unif}(a, b)$;

(NOT FOR MODELLING)

We write $X \sim F$ to say that 'X is distributed as F'.

- Chi-squared distribution: If $Z_i \sim N(0,1)$, $i = 1, \ldots, n$, and Z_i's are independent[9] then

$$Z_1^2 + \cdots + Z_n^2 \sim \chi_n^2,$$

n is called degrees of freedom;

- t distribution: if $Z \sim \mathrm{N}(0,1)$ and $U \sim \chi_\nu^2$, with Z, U independent, then

$$Z/\sqrt{U/\nu} \sim t_\nu,$$

, ν is called degrees of freedom.

- F distribution: if $U_1 \sim \chi_{n_1}^2$ and $U_2 \sim \chi_{n_2}^2$, with U_1, U_2 independent, then

$$(U_1/n_1)/(U_2/n_2) \sim F_{n_1,n_2},$$

and n_1, n_2 are the numerator and denominator degrees of freedom, resp.

[9] More on independence of rv's later

Moments

For $g : \mathbb{R} \to \mathbb{R}$, the expectation of $g(X)$ is

$$\boxed{E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x,}$$

provided $\int_{-\infty}^{\infty} |g(x)|f(x)\mathrm{d}x$ exists and is finite.

Examples

- $g(x) = x$: $E(X) = \mu_X$ the expectation of X;
- $g(x) = (x - c)^n$, $c \in \mathbb{R}$, $n \in \mathbb{N}$: nth moment of X about c

$$E[(X - c)^n] = \int_{-\infty}^{\infty} (x - c)^n f(x)\mathrm{d}x,$$

  provided the integral exists;
- $g(x) = (x - E(X))^n$: nth central moment; for $n = 2$, we get the variance of X, denoted $\mathrm{var}(X) = \sigma_X^2$.

Transformation of a r.v.

Applying a function g to a r.v. X, leads to another r.v. $Y = g(X)$.

If X is discrete, the Y is discrete and

$$f_Y(y) = P(Y = y) = P(\{s : g(X(s)) = y\}).$$

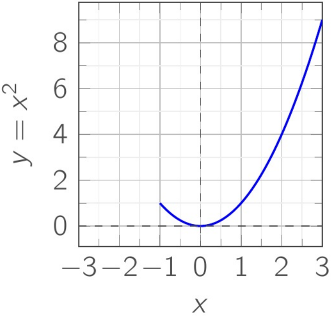
If X is continuous the procedure is more difficult, we have to:

- find $B_y = \{x : g(x) \leq y\}$, for each y in the range of Y;
- find the df $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(B_y) = \int_{B_y} f(x)\mathrm{d}x$;
- take the derivative, i.e. $f_Y(y) = F'_Y(y)$.

Transformation of a rv

Example 11

Let $X \sim \mathrm{Unif}(-1, 3)$, we compute the pdf of $Y = X^2$. Consider the graph of $g(x) = x^2$.



Now $Y \in [0, 9]$, for $x \in [-1, 1]$, $g(x) \in [0, 1]$, whereas for $x > 1$ $g(x)$ is bijective. We have two cases:

(1) $0 \leq y \leq 1$: $B_y = (-\sqrt{y}, \sqrt{y})$ and
$F_Y(y) = (1/4) \int_{B_y} \mathbf{1}_{[-1,3]} \mathrm{d}x = \sqrt{y}/2$;

(2) $y > 1$: $B_y = (-1, \sqrt{y})$,
$F_Y(y) = (\sqrt{y} + 1)/4$.

The pdf can be found by differentiating F_Y, taking care of the two cases.

Example 12 (Example 11 cont'd)

Let's compute $E(Y)$ and $\text{var}Y$. We have that

$$f_Y(y) = \begin{cases} 1/(4\sqrt{y}) & \text{if } 0 < y \le 1 \\ 1/(8\sqrt{y}) & \text{if } 1 < y \le 9. \end{cases}$$

Then

$$\begin{aligned} E(Y) &= \int_0^9 y f_Y(y)\mathrm{d}y = \int_0^1 y/(4\sqrt{y})\mathrm{d}y + \int_1^9 y/(8\sqrt{y})\mathrm{d}y \\ &= \int_0^1 \sqrt{y}/4\mathrm{d}y + \int_1^9 \sqrt{y}/8\mathrm{d}y = 7/3. \end{aligned}$$

$$\begin{aligned} E(Y^2) &= \int_0^9 y^2 f_Y(y)\mathrm{d}y = \int_0^1 y^2/(4\sqrt{y})\mathrm{d}y + \int_1^9 y^2/(8\sqrt{y})\mathrm{d}y \\ &= \int_0^1 y^{3/2}/4\mathrm{d}y + \int_1^9 y^{3/2}/8\mathrm{d}y = 61/5. \end{aligned}$$

Thus $\text{var}(Y) = E(Y^2) - E(Y)^2 = 61/5 - 49/9 = 794/45.$

Inverse tranform

Suppose Y is a continuous rv with distribution F_Y.

It can be shown that F_Y is continuous and bijective with inverse $F^{-1} = Q$.

Furthermore, if $X = \text{Unif}(0, 1)$, then $Q(X) \sim F_Y$.
This fact is useful when we want to draw random values from F_Y.
Indeed, if we

(a) draw a number p uniformly in (0,1)

(b) set $y = Q(p)$,

y is a random value from F_Y. This is knwon as the **inverse tranfrom sampling** method.