Advanced statistical methods and Bayesian inference in scientific research

Lecture 2

W. Dębski

18.03.2024

Probability theory - Axioms

The structure $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{p})$ called the **probability space**:

Ω	-	space of all elementary events
$\mathcal{F}$	-	set of all subsets of Ω
$\mathcal{A} \in \mathcal{F}$	-	are called **events**,
$\mathbf{p}()$	-	function $\Omega \to R^+$.

Probability theory - Axioms

Let Ω be a set (sample space)
Let $\mathcal{F}$ be a σ-algebra of all subsets of Ω . (event space)

σ-algebra of all subsets of Ω is a family of Ω's subsets such that

- ✦ $\Omega \in \mathcal{F}$
- ✦ $A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$ where $\bar{A}$ denotes compactness of A
- ✦ for all $A_i \in \mathcal{F}$ $\bigcup_i A_i \in \mathcal{F}$

Let $\mathbf{p}(\cdot)$ be a function: $\Omega \to R^+$ (probability measure)

a) $\mathbf{p}(\varnothing) = 0$

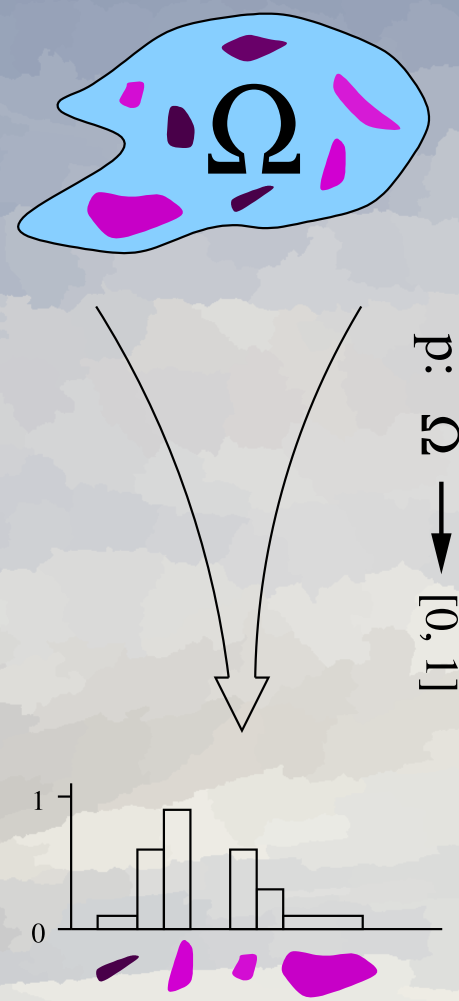
b) $\mathbf{p}(\mathcal{A}) \geq 0$ for all $\mathcal{A} \in \mathcal{F}$

c) $\mathbf{p}(\bigcup_i A_i) = \sum_i \mathbf{p}(A_i)$ for any $A_i, A_j \in \mathcal{F}$

$A_i \cap A_j = \varnothing$ if $i \neq j$

If $\mathbf{p}(\Omega) = 1$ than $\mathbf{p}()$ is called probability

Probability - graphical representation



Interpretations

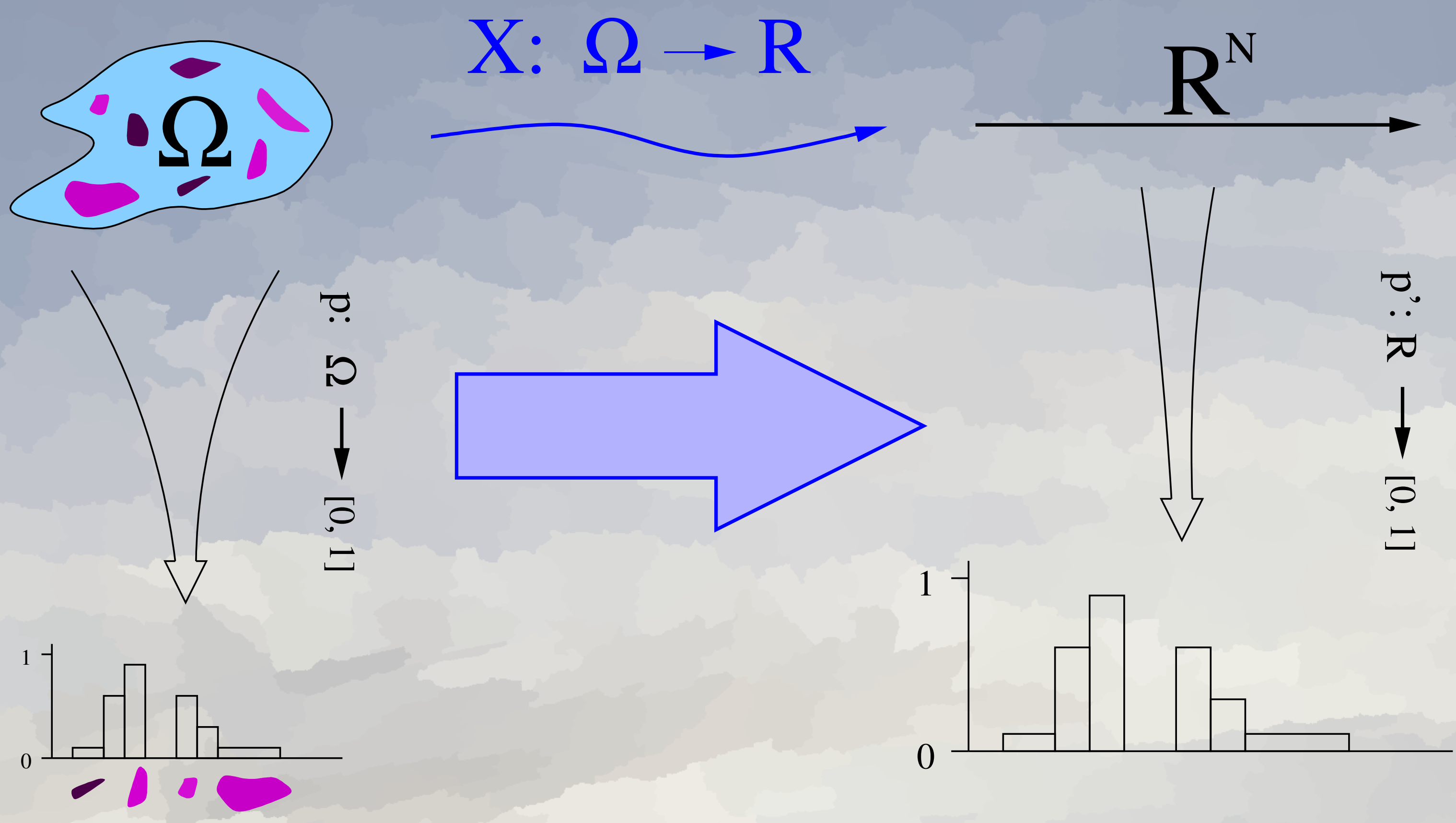
- Physical approach

  Physical (objective, frequency) probabilities, are associated with random physical systems like rolling dice and radioactive atoms. In such systems, a given type of event tends to occur at a given relative frequency, in a long run of trials. Physical probabilities are suppose to explain these stable frequencies.

- Bayesian approach

  Bayesian (evidential) probability can be assigned to any statement even when no random process is involved, as a way to represent its subjective plausibility, or the degree to which the statement is supported by the available evidence. On most accounts, evidential probabilities are considered to be degrees of belief.

Random variables



Random variables - math definition

A random variable $\mathbf{X}$ is a measurable function

$$\mathbf{X}: \quad \Omega \Longrightarrow \mathcal{E}$$

from a sample space Ω to a measurable space $\mathcal{E}$.

The probability that $\mathbf{X}$ takes on a value in $\mathcal{A} \subset \mathcal{E}$ reads

$$P'(\mathbf{X} \in \mathcal{A}) = P(\omega \in \Omega \mid \mathcal{X}(\omega) \in \mathcal{A})$$

Random variables

- physical parameters describing system
- noisy measurement data $\mathbf{d} = \mathbf{d}^{th} + noise$
- occurrence of physical processes (catalogs)
- time evolution of physical system (stochastic process)
- spatial under-sampling (e.g. Geo-statistics)
- approximate theoretical predictions
- missing information on a system

Discrete and continuous probability of random variables

Let consider a random variable X with real scalar values.

If values of X form a discrete set (countable) X is called discrete random value and corresponding probability distribution $\mathbf{p}(x)$ is discrete-domain function. Classically such pdf is represented by finite bin-width histograms.

If X can take continuous values X is called continuous random value. In such a case probability distribution $\mathbf{p}(x)$ is continuous function and cannot directly be described by "histograms" since probability of X taking given value X_o is strictly speaking zero (it is a set of measure zero).

Both situation can however be describe uniformly introducing cumulative distribution function (cpd):

$$F_X(x) = P(X < x)$$

Cumulative distribution and probability density function

Usually (continuous case) cpd is differentiable so one can define probability density function

$$\mathbf{p}(x) \ = \ \frac{d}{dx} F_X(x)$$

than any probability that, e.g

$$a < X < b$$

can be represented as

$$P(a < X < b) \ = \ \int_a^b \mathbf{p}(x) \mathbf{d}x$$

Multidimensional pdf

In the simplest cases random variables are scalar-valued functions.

- energy of the body
- temperature
- electric charge

but can also take more complex values:

- wind direction
- location/position
- velocity

Marginal pdf

In generally we will assumes that $X \in R^N$ is multi-dimensional vector of real values In such cases one can define marginals

- 1-dimensional by integrating out all but one component of X

$$\mathbf{p}_i(x_i) \; = \int\limits_{x \neq x_i} \mathbf{p}(\mathbf{x})\, \mathbf{dx}$$

- 2-dimensional by integrating out all but two components of X

$$\mathbf{p}_{ij}(x_i, x_j) \; = \int\limits_{x \neq x_i, x_j} \mathbf{p}(\mathbf{x})\, \mathbf{dx}$$

Meta-characteristic of pdf - Evidence

Normalization of pdf:

$$\mathbf{Z} = \int \mathbf{p}(x)\mathbf{dx}$$

$$\mathbf{p}(x) \longrightarrow \frac{1}{Z}\,\mathbf{p}(x)$$

If X depends on set of parameters $\mathcal{S}$ then

$$\boxed{\mathbf{p}(x) \;=\; \tfrac{1}{Z[\mathcal{S}]}\, f(\mathbf{x};\mathcal{S})}$$

Meta-characteristic of pdf -Entropy

Yet another “global” description of pdf $\mathbf{p}(x)$.

$$H(x) = \int \ln\left(\mathbf{p}(\mathbf{x})\right) \; \mathbf{p}(\mathbf{x})\mathbf{dx}$$

but $\mathbf{p}(\mathbf{x})$ IS NOT dimensionless so better

$$H(x) = \int \ln\left(\frac{\mathbf{p}(\mathbf{x})}{\mathbf{p_o}(\mathbf{x})}\right) \; \mathbf{p}(\mathbf{x})\mathbf{dx}$$

Describing *pdf*

Full pdf brings all available information on given random variable X we have in hand. However, sometime it is more useful to characterize $\mathbf{p}(x)$ by a finite set of numbers rather than use the full pdf.

The most popular approach is to use moments of $\mathbf{p}(x)$

$$\mu'_n = \int x^n \mathbf{p}(x) \mathbf{d}x$$

or central moments

$$\mu_n = \int (\mathbf{x} - \mathbf{x}^{avr})^n \mathbf{p}(x) \mathbf{d}x$$

Such description we shall call a point-like method.

Mode - maximum likelihood point

The mode value $\mathbf{x}^{ml}$ is the value of random variable X for which its density function $\mathbf{p}(\cdot)$ reads absolute maximum

$$\mathbf{x}^{ml} : \mathbf{p}(\mathbf{x}^{ml}) \;=\; \max$$

Warning:

Mode estimator $\mathbf{x}^{ml}$ is very sensitive to presence of outliers, or other noise in $\mathbf{p}(x)$!

Average (mean) - arithmetic

An average(mean) is a numeric quantity representing the center of $\mathbf{p}(\cdot)$

$$E(X) = \mathbf{x}^{avr} = \int_{-\infty}^{\infty} x\, \mathbf{p}(x)\, dx$$

multidimensional case

$$E(X_i) = \bar{X}_i = \int \cdots \int x_i\, \mathbf{p}(\mathbf{x})\, \mathbf{d}x$$

Let observe

$$E(X_i) = \bar{X}_i = \int x_i \mathbf{p}_i(x_i)\, \mathbf{d}x_i$$

where $\mathbf{p}_{\{i\}}(\cdot)$ is marginal distribution

Average value - comments

There are many (infinite) versions of the average values:

- harmonic average: $E_{1/x}(X) = \int \frac{1}{x} \mathbf{p}(x)\, \mathbf{d}x$
- root mean square: $E_{rms}(X) = \sqrt{\int x^2 \mathbf{p}(x)\, \mathbf{d}x}$
- power average: $E_m(X) = \left(\int x^m \mathbf{p}(x)\, \mathbf{d}x\right)^{1/m}$
- f - mean:

$$E_f(X) = f^{-1}\left(\int f(x)\, \mathbf{p}(x)\, \mathbf{d}x\right)$$

Mediana

Mediana $\mathbf{x}^{me}$ is such a value of random variable X that

$$P(X \leq \mathbf{x}^{me}) \;=\; P(X \geq \mathbf{x}^{me}) \;= \frac{1}{2}$$

if $\mathbf{p}(\;)$is continuous

$$\int_{-\infty}^{\mathbf{x}^{me}} \mathbf{p}(x)\;\mathbf{d}x \quad = \quad \int_{\mathbf{x}^{me}}^{\infty} \mathbf{p}(x)\;\mathbf{d}x$$

This is the very robust estimator !

Higher order moments dispersion

$$\sigma^2 = \int (\mathbf{x} - \mathbf{x}^{avr})^2 \, \mathbf{p}(x) \, \mathbf{d}x$$

Multi-dimensional extension - **covariance matrix**

$$\sigma_{ij} = \int \cdots \int \left(x_i - x_i^{avr}\right) \left(x_j - x_j^{avr}\right) \mathbf{p}(x) \, \mathbf{d}x$$

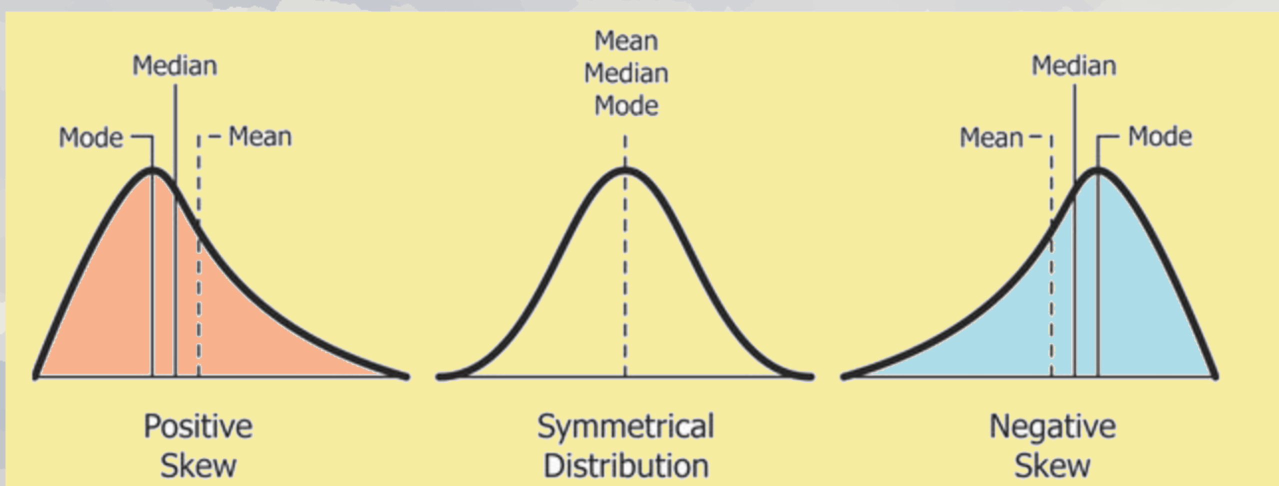
and correlation matrix

$$r_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}^2}\sqrt{\sigma_{jj}^2}} \qquad \Rightarrow -1 \leq r_{ij} \leq 1$$

Higher order moments - skewness

Skewness is a measure of the asymmetry of the probability distribution with respect to $\mathbf{x}^{avr}$

$$\gamma_1 = \frac{1}{\sigma^3} \int (x - \mathbf{x}^{avr})^3 \, \mathbf{p}(x) \mathbf{d}x$$



(after Wikipedia)

Higher order moments - kurtosis

Kurtosis is a measure of the how long is the "tail" of the probability distribution of a random variable.

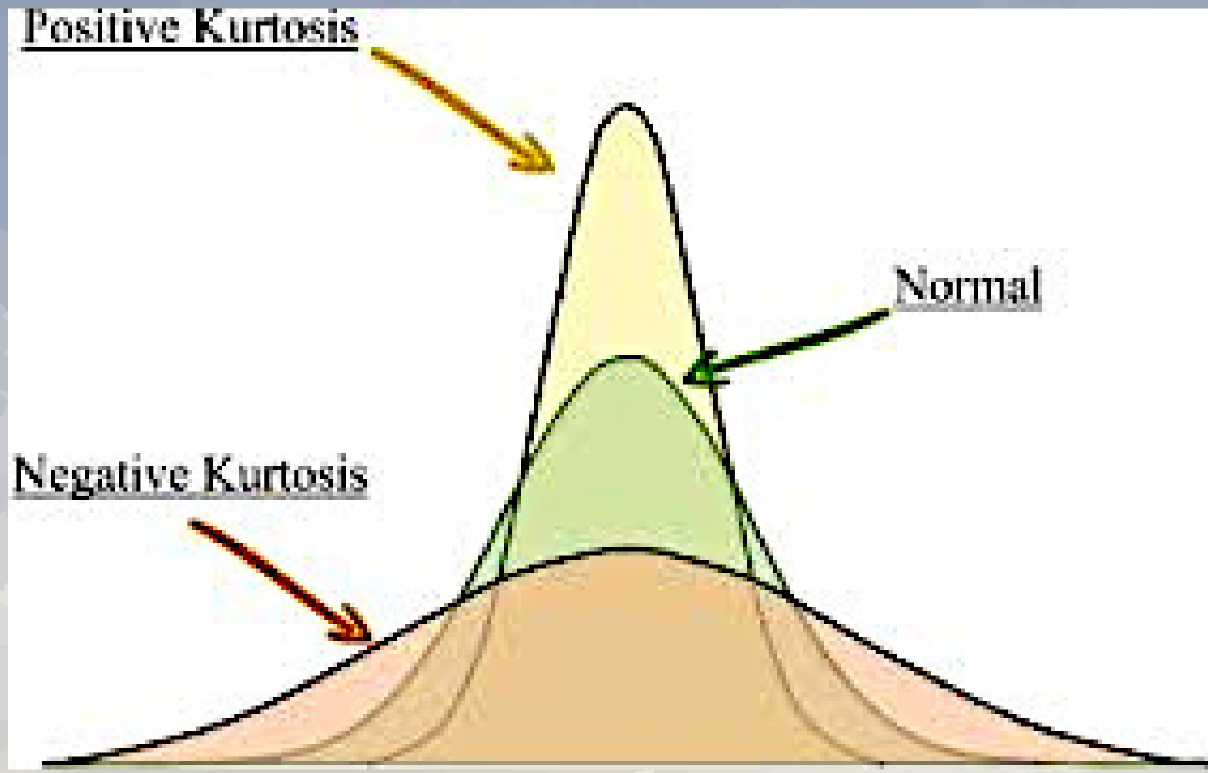
$$\sigma_4 = \frac{1}{\sigma^4} \int (x - \mathbf{x}^{avr})^4 \, \mathbf{p}(x) \mathbf{d}x$$

More exactly only data which are "outside the region of $\mathbf{x}^{avr}$" contribute significantly to kurtosis. This means that kurtosis is a measures of an existence of outliers only;

It is bounded

$$\sigma_4 \geq \sigma^4 (\gamma_1^2 + 1)$$

Kurtosis - illustration (after Wikipedia)



Characteristic function

The characteristic function of any real valued random variable X is defined as

$$\phi_X(t) = \int e^{itx} \mathbf{p}(x) \mathbf{d}x$$

It can be generalized for stochastic process $X(s)$

$$\phi_X(t) = \int \exp\left(i \int_R t(s) X(s) ds \right) \mathbf{p}(x) \mathbf{d}x$$

Characteristic function - some properties

Let

$$X_1, X_2, \cdots X_N$$

be independent random variables and

$$a_1, a_2, \cdots a_N$$

be constants. Let define new random variable

$$Y = \sum_{i=0}^{N} a_i X_i$$

then

$$\phi_Y(t) = \phi_{X_1}(a_1 t)\phi_{X_2}(a_2 t)\ldots\phi_{X_N}(a_N t)$$

Characteristic function - some properties

Let X, and Y be two random variables with characteristic functions

$$\phi_X(t), \text{ and } \phi_Y(t)$$

Than, X and Y are independent if and only if

$$\phi_{X,Y}(s,t) = \phi_X(s)\phi_Y(t)$$

Characteristic function - moments

If characteristic function $\phi_X(t)$ can be differentiated n times then

$$\mu'_n \;=\; i^{-n} \left[\frac{d^n}{dt^n}\phi_X(t)\right]_{t=0}$$

Example

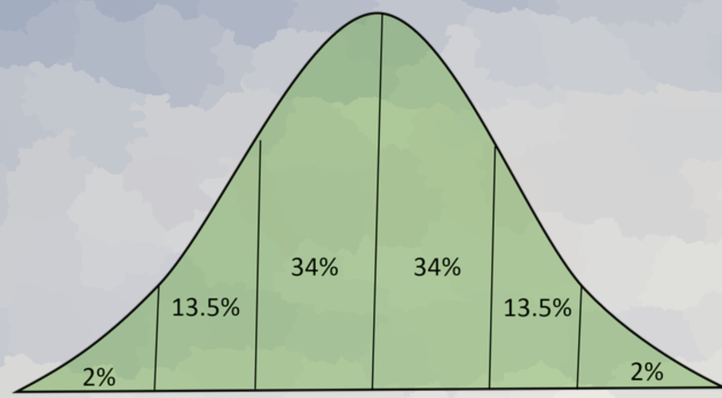
Let random variable X is Gaussian one. Then

$$\phi_X(t) \;=\; \exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right)$$

$$E[X] = \int x\mathbf{p}(x)\mathbf{d}x \;=\; i^{-1} \left[\frac{d}{dt}\phi_X(t)\right]_{t=0} \;=\; \mu$$

Quantiles

Quantiles are cut points dividing the range of a probability distribution into continuous intervals with equal probabilities (Wikipedia)



The quantile function is one way of prescribing a probability distribution, and it is an alternative to the probability density function (pdf) or probability mass function, the cumulative distribution function (cdf) and the characteristic function (Wikipedia).

See you next week ...