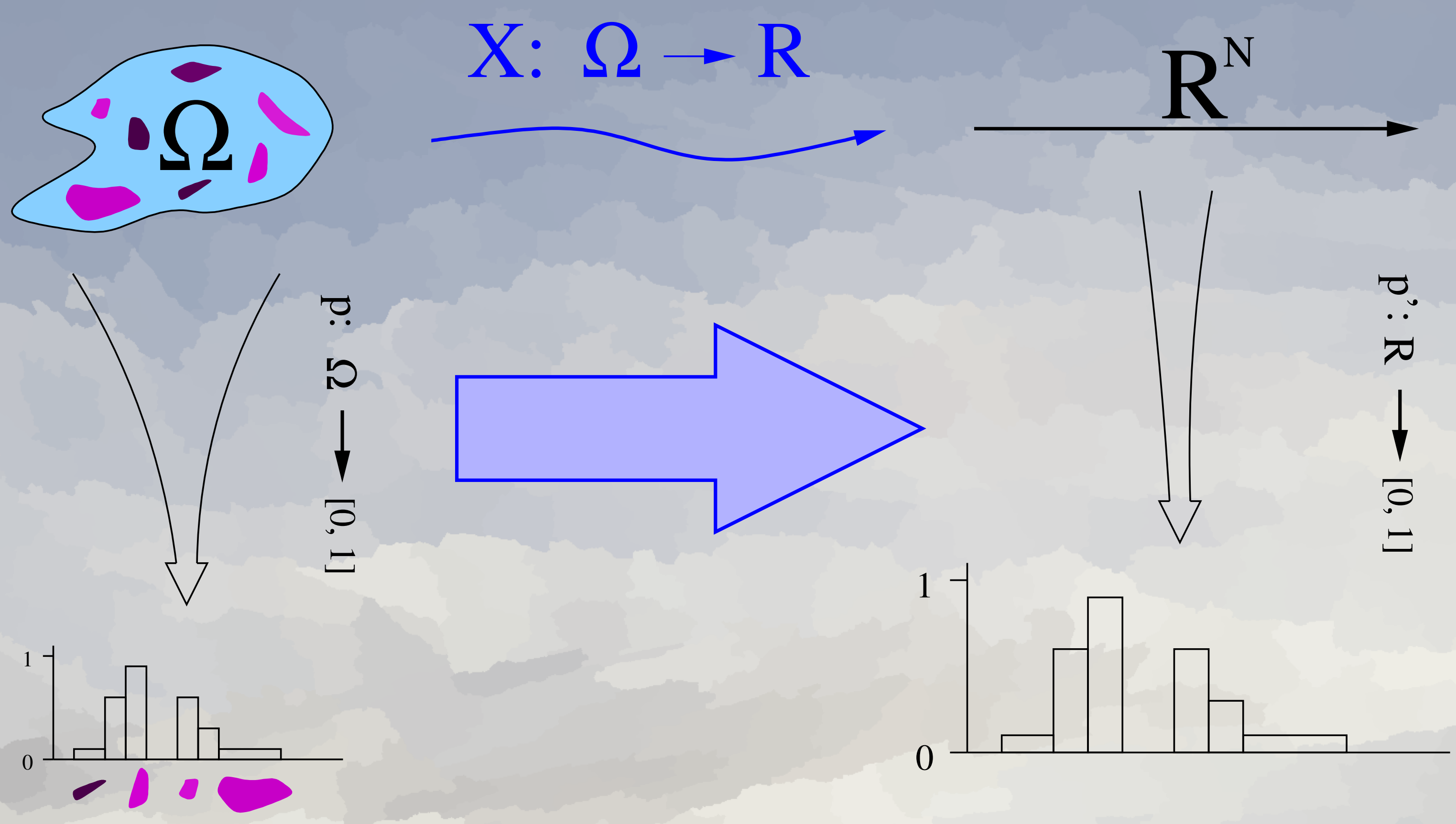
Advanced statistical methods and Bayesian inference in scientific research

Lecture 3

W. Dębski

25.03.2024

Random variables



Random variables

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

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

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

Describing *pdf* (1)

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.

Changing variables

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

Basic fact: the above integral does not depend on used parameterization

$$x = y(x)$$

$$\mathbf{d}x = \left|\frac{\partial y}{\partial x}\right| \mathrm{d}y$$

$$\mu_n' = \int y^n \mathbf{p}(x)\big|_{x=y(x)} \left|\frac{\partial y}{\partial x}\right| \mathbf{d}y = \int y^n \mathbf{p}'(y)\mathbf{d}y$$

$$\mathbf{p}'(y) = \mathbf{p}(x) \left|\frac{\partial y}{\partial x}\right|$$

Random variable transformation

Let transform X to Y

$$X \Longrightarrow Y$$

Consistency requirement:

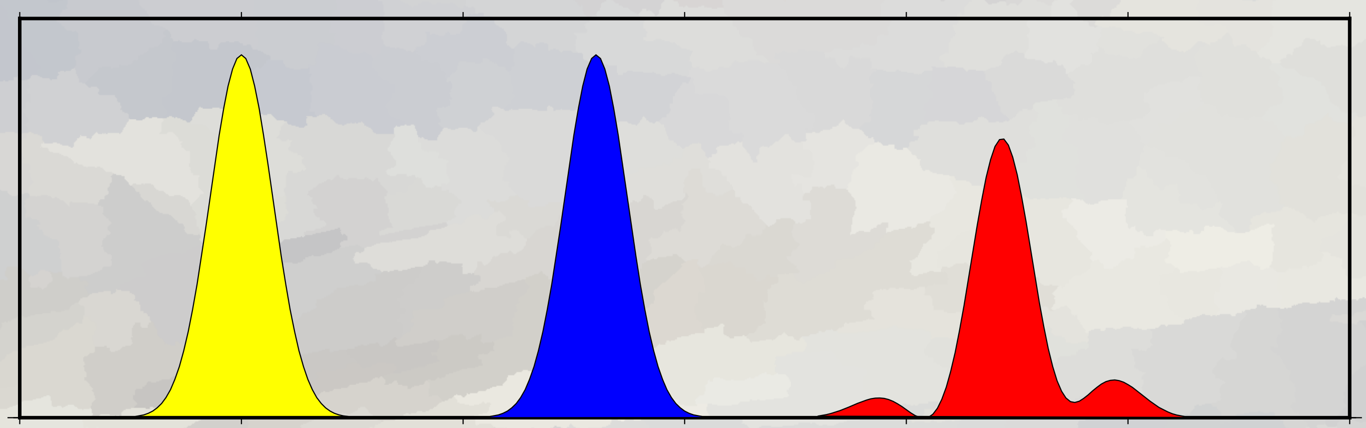
$$\forall_{\mathcal{A}} \;\; P_X(\mathcal{A}) \; = \; P_Y(\mathcal{B})$$

$$\boxed{\mathbf{p}(y) \; = \; \mathbf{p}(x) \left| \frac{\partial y}{\partial x} \right|}$$

Describing *pdf* (2)

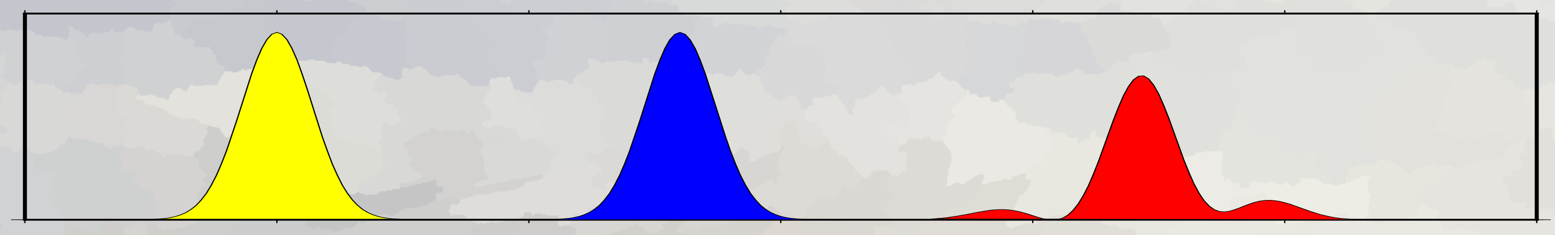
Moments μ' provides detailed characteristics of $p(x)$

However, quite often we need only a compact description of $p(x)$
- its shape rather than its full details



Meta-characteristic - Entropy

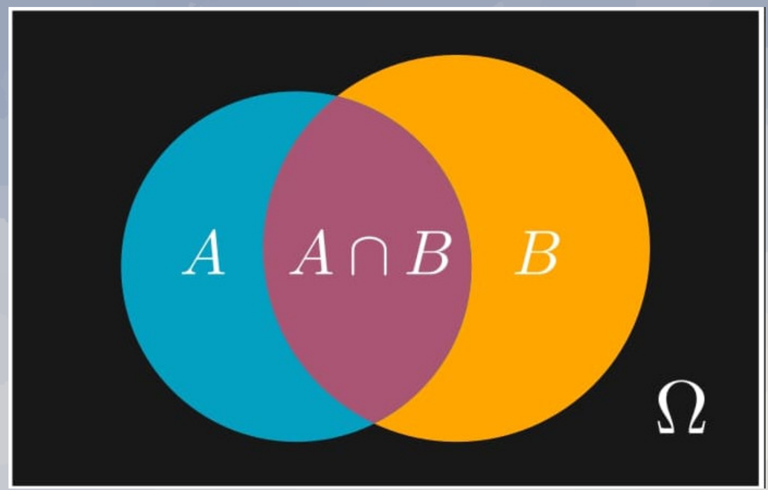
$$H[\mathbf{p}] = -\int \ln(\mathbf{p}(\mathbf{x}))\ \mathbf{p}(\mathbf{x})\mathbf{dx}$$



8.5 8.5 15

Conditional probability

Let us assume two events A and B



The conditional probability is defined as

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

If $P(B|A) = P(B)$ the events A and B are **independent**

Bayes theorem

$$P(A \cap B) = P(A|B) \times P(B)$$
$$P(A \cap B) = P(B|A) \times P(A)$$

$$P(B|A) \times P(A) = P(A|B) \times P(B)$$

$$\boxed{P(B|A) = \frac{P(A|B) \times P(B)}{P(A)}}$$

Bayes theorem - single variable X

- Variable X is measured for the first time but we know that its value should follow (approximately) $\rho(x)$

Question: how performed measurement allows us to constraint (update) the pdf distribution $p(x)$ describing X?

Bayes theorem - single variable X

$$P(B|A) \ = \ P(A|B) \times P(B)/P(A)$$

Interpretation:

B	$\rightarrow$	values of X
P(B)	$\rightarrow$	initial (*a priori*) pdf $\rho(x)$
$P(B\|A)$	$\rightarrow$	thought $p(x)$
A	$\rightarrow$	values of X drawn from $\rho(x)$
$P(A\|B)$	$\rightarrow$	how x depends on u with given x_o

$$p^{(1)}(x) \ = \ \kappa \, p(x|x_o, \rho) \times \rho$$

More explicite

$$p^{(1)}(x) \ = \ \kappa \int p(x|x_o, u) \times \rho(u) \mathbf{d}u$$

Bayes theorem - single variable X

If measurements are independent $P(A|B) = P(B)$

$$P(B|A) = \frac{P(A|B) \times P(B)}{P(A)} = \frac{P(A) \times P(B)}{P(A)} = P(B)$$

$$p^{(1)}(x) \backsim p(x|\rho) = \rho(x)$$

$$p^{(n)}(x) = \rho(x)$$

Conclusion (not quite precise):

If measurements are independent we get no refinement of $p(x)$ via Bayes theorem

Bayes theorem - two variables X, Y (1)

- Variable X is approximately known as $p^o(x)$.
  We measure another variable Y

Question: can knowledge of Y improves information about X ?

$$P(B|A) \; = \; P(A|B) \times P(B)/P(A) \quad X \rightarrow B \quad Y \rightarrow A$$

$$\boxed{p^+(x) \; = \; p(x|y) \; = \; \kappa \, P(y|x) \times p^o(x)}$$

Bayesian inversion - two variables X, Y (2)

- $x \rightarrow \mathbf{m} \qquad y \rightarrow \mathbf{d}$
- $p^o(x)$ - a *priori* distribution $\rho^{apr}(\mathbf{m})$
- $P(y|x)$ - likelihood function $L(\mathbf{d}, \mathbf{m})$
- $p^+(x)$ - a *posteriori* distribution $\sigma(\mathbf{m})$

$$\sigma(\mathbf{m}) \;=\; L(\mathbf{d}, \mathbf{m})\, \rho^{apr}(\mathbf{m})$$

Bayesian inversion - two variables X, Y (3)

Let assume that

$$\mathbf{d} \;=\; \mathbf{G} \cdot \mathbf{m}$$

and we have measured

$$\mathbf{d}^{obs} = \mathbf{d} + noise$$

assuming that noise has given statistics

$$p(\epsilon) \;=\; \rho(\mathbf{d} - \mathbf{d}^{obs})$$

$$L(\mathbf{d}, \mathbf{m}) \;=\rho(\mathbf{d}^{obs} - \mathbf{G} \cdot \mathbf{m})$$

$$\boxed{\sigma(\mathbf{m}) \;=\; \rho(\mathbf{d}^{obs} - \mathbf{G} \cdot \mathbf{m})\; \rho^{apr}(\mathbf{m})}$$

Mathematics of inference - Inference Space

$$(\mathcal{P}, \Sigma, \wedge)$$

where

$\mathcal{P}$ - parameter space

Σ - space of all probability distributions over $\mathcal{P}$

$\wedge(\cdot, \cdot)$ - joining operator: $\Sigma \times \Sigma \rightarrow \Sigma$

Joining information according to Tarantola

Two distributions describing **different** pieces of information about the same object

1. $p(x)$
2. $q(x)$

$$\zeta(x) \; = \; \mathbf{p} \wedge \mathbf{q}(\mathbf{x}) = \frac{\mathbf{p}(\mathbf{x})\,\mathbf{q}(\mathbf{x})}{\mu(\mathbf{x})}$$

$\mu(x)$ - non-informative probability

Non-informative distribution

$$q(x) \; = \; \mu(x)$$

$$\mathbf{p} \wedge \mathbf{q}(\mathbf{x}) = \frac{\mathbf{p}(\mathbf{x}) \; \mu(\mathbf{x})}{\mu(\mathbf{x})} \; = \; \mathbf{p}(\mathbf{x})$$

Essentially, $\mu(\cdot)$ can be arbitrary but usually is taken as *volumetric pdf*

Insufficient reasoning postulate (Laplace)

If x is a real unconstrained parameter then the state of lack of information on it is represented by a probability proportional to the volume (length) in $\mathcal{R}$.

$$p(A) = \int_A \mu(x)dx = vol(A),$$

where A stands for the volume (length) of the set A, ($A \subset \mathcal{R}$.

In a Cartesian coordinate system

$$\mu(x) = const.$$

Bounded parameters

Let x be a parameter whose values belong to a segment $\Omega = [a, b]$.

Next, let us introduce a new parameter x'

$$x' = g(x)$$

$g()$ - differentiable function

If $x' \in \mathcal{R}$ than $\mu(x') = const$

Using transformation properties of probability density function

$$\mu(x) \;=\; \left|\frac{dg}{dx}\right|$$

Example - positive defined parameters

Let f be a positive parameter (e.g. frequency) $f \in \mathcal{R}^+$

Then, one can take transformation function ($\mathcal{R}^+ \Rightarrow \mathcal{R}$)

$$x \;=\; g(f) = \ln(f)$$

What immediately leads to the non-informative pdf for f

$$\mu(f) \;=\; \frac{1}{f}$$

Comment: if assume $\mu(f) \;=\; const.$ we end up in contradiction with above
- Borrel's paradox

Example - positive defined parameters

However, we can prefer other parameter: period

$$T = 1/f$$

which represents exactly the same information like f.
Then, using transformation rule for parameter change

$$f \to T$$

We immediately get

$$\mu(T) \;=\; \mu(f(T)) \left|\frac{df}{dT}\right| \;=\; \frac{1}{T}$$

See you April, 8th