Advanced statistical methods and Bayesian inference in scientific research

Lecture 6

W. Dębski

6.05.2024

Probability (A)

- ✦ Physical approach
  Physical (objective, frequency) probabilities, are associated with random processes and systems like noisy/unprecised measurements, processes like radioactive atoms decay, or a systems of many (interacting?) elements. In such systems, a given type of event tends to occur at a given relative frequency, in a long run of trials/observations. Physical probability is a measure of these relative frequencies.

This definition undertakes (mixes) two different issues:

1. what probability is
2. how to construct it

Probability (B)

✦ Bayesian approach

Bayesian (evidential) probability is a kind of measure that can be assigned to any statement, observation, forecasting, etc. 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.

This definition:

1. tells what probability is but does not tell how to construct it
2. *expressis verbis* shows underlying assumption - insufficient reasoning principles

Probability difference

Above notions of "probability" actually defines two different objects and essentially we should use two different names for them. For traditional reasons it is not a case. The situation additionally complicates the fact, that both probabilities can/are described by the same mathematical structure (probabilistic space with e.g. Kolomogorov axioms). However, choice one of them, what is absolutely arbitrary determines hove given problems in hand is approach. For this reason the choice of Frequentists or Bayesian probability should be done appropriately to the task.

Probability differences - Example

Let assume we have N samples $X_1, X_2, \ldots X_N$ drawn from the normal distribution

$$\mathcal{N}(\theta, 1) = A \exp(-(x-\theta)^2)$$

We want to provide some sort of interval estimate C of θ

How should we proceed ???

Probability differences - Example (freq.)

1. We know that samples comes from $\mathcal{N}(\theta, 1)$ so we can construct (2σ) interval

$$C = \left(\bar{X}_N - \frac{1.96}{\sqrt{N}} , \bar{X}_N - \frac{1.96}{\sqrt{N}} \right)$$

2. then

$$P_\theta(\theta \in C) = 0.95$$

3. The solution is the statement about the interval C which is now a random variable because depends on "random" samples X_i. Unknown (thought) parameter θ is fixed (although unknown) parameter and we state nothing about it

4. Solving the problem relays on **constructing** interval C

Probability differences - Example (freq.)

What $P_\theta()$ means ?

Suppose that we repeat "experiment" K times

Nature sets θ_1 $\Rightarrow$Nature gives $\{X_i\}$ from $\mathcal{N}(\theta,1)$$\Rightarrow$statistician computes C_1

Nature sets θ_2 $\Rightarrow$Nature gives $\{X_i\}$ from $\mathcal{N}(\theta,1)$$\Rightarrow$statistician computes C_2

$\vdots$ $\Rightarrow$ $\vdots$ $\Rightarrow$ $\vdots$

Nature sets $\theta_K$$\Rightarrow$Nature gives $\{X_i\}$ from $\mathcal{N}(\theta,1)$$\Rightarrow$statistician computes C_2

We will find that C_i traps the parameters θ, 95% of the time

Probability differences - Example (Bayes.)

1. We treat θ as **unknown** parameter we wish to infer from "data " $\{X_i\}$ Since our knowledge about θ is limited we treat θ as a "random" (unknown) variable and describe it by appropriate probability

2. Since at start we know nothing about θ we describe it by a prior distribution $\rho(\theta)$

3. Inferring about θ goes by calculating *a posteriori* distribution having in hand the set of data $\{X_i\}$

$$\sigma(\theta) = \mathcal{L}(\theta|\{X_i\})\rho(\theta)$$

   using e.g. Bayes theorem

4. We construct the interval C such that

$$\int_C \sigma(\theta) d\theta = 0.95$$

Probability differences - Example (Bayes.)

The solution can be stated as

$$P(\theta \in C|\{X_i\}) \ = \ 0.95$$

- This time the solution is the statement about θ which now is treated as the random variable and knowledge about it is described by $\sigma(\theta)$
- $P(\theta \in C|\{X_i\}) \ = \ 0.95$ does not guarantee now that when experiment is repeated θ will be in C 95% of times
- The solution depends somehow on *a priori* - subjective $\rho(\theta)$

Probability differences - conclusion

- Frequentists
  - Frequentists inference is focused on constructing given procedure which assures that "frequency probability (like 0.95 here) is guaranteed
  - thought parameter is treated as fixed but unknown
  - solution incorporates usually various "hidden" assumptions (here, e.g. normality of $\{X\}$)
- Bayesian
  - this approach is a method for stating and updating possessed information about unknown parameters
  - inference means joining and analyzing possessed information
  - the method explicit manipulates information but does not create them !
  - can be used even if single experiment is available.

Example-II. Measurements

Voltage measurement

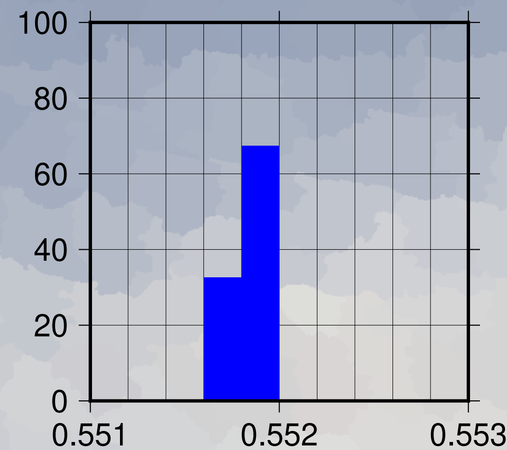
	0.5518 0.5518 0.5518 0.5518 0.5519 0.5518 0.5517 0.5518 0.5519 0.5518

What is the true value of the measured voltage?

$$V^{obs} = V^{true} + n$$

Example-II. Measurements

Naive (Frequentists) estimation: construct histogram of measured values



Conclusion: measuring accuracy at 95% confidence level

$$||n|| = \Delta V \ = \ 0.0002V$$

is apparently wrong !

Example-II. Measurements

Technical specification

Multimeter Technical Specifications

Function	Range	Resolution	Accuracy
DC Voltage (V)	2.5000V	0.0001V	±(0.05% +3)
	25.000V	0.001V	
	250.00V	0.01V	
	1000.0V	0.1V	
DC Voltage (mV)	25.000mV	0.001mV	
	250.00mV	0.01mV	

$$||n|| = \Delta V \;=\; 0.05\% * 2.5V + 0.0003 = 0.0015V$$

This example illustrates possible problems with "Frequentists" data-based construction of probability distribution

Example-II. Measurements

Technical specification is not fully informative !

Writing

$$V = V^{true} \pm \Delta V$$

we usually assume that noise n has a normal (Gaussian) distribution with zero average and variance estimated by ΔV.

Is it really true?

Example-II. Bayesian approach

The output of measurement V is a random variable due to existence of random noise. We wish to learn about it as much as possible

- *A priori distribution*

$$\rho(V) = const.$$

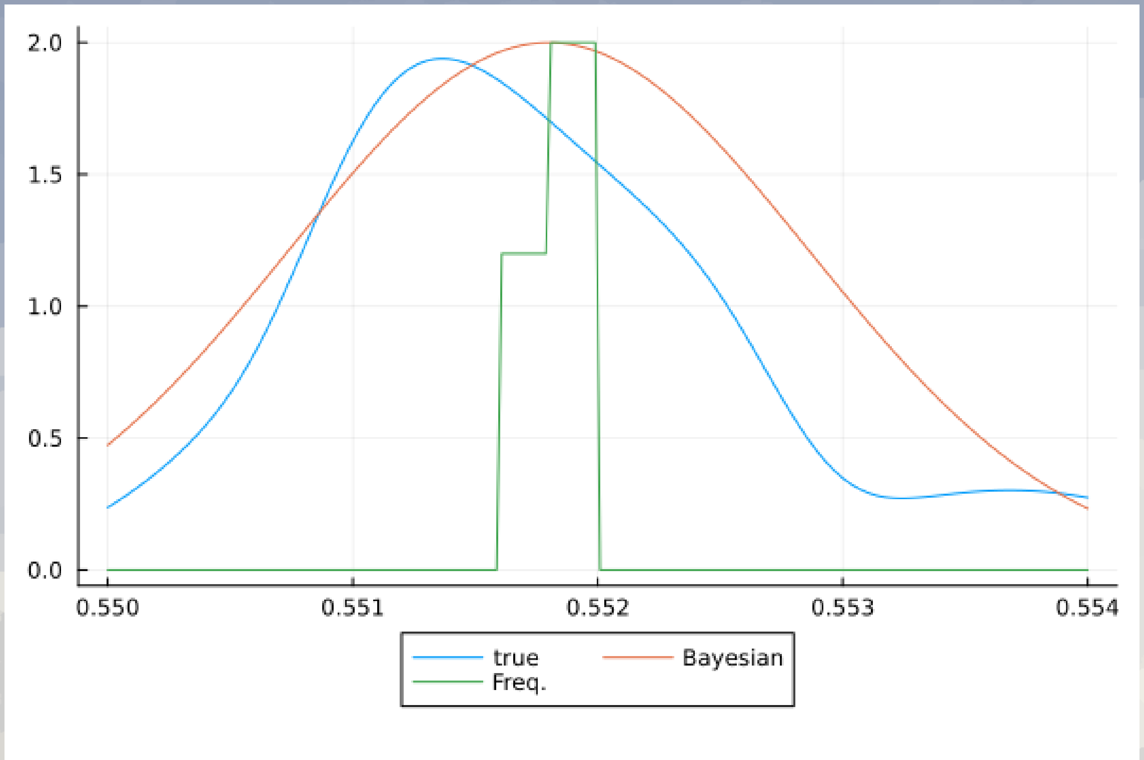
- we assume Gaussian noise with a variance provided by manufacturer:
  $\Delta V = 0.0005 * 2.5V + 0.0003 = 0.0015V$

- construct *a posteriori* based on measured values

$$\sigma(V) = \rho(V) \prod_i \exp\left(-\frac{(V_i - V)^2}{\Delta V^2}\right)$$

- explore it

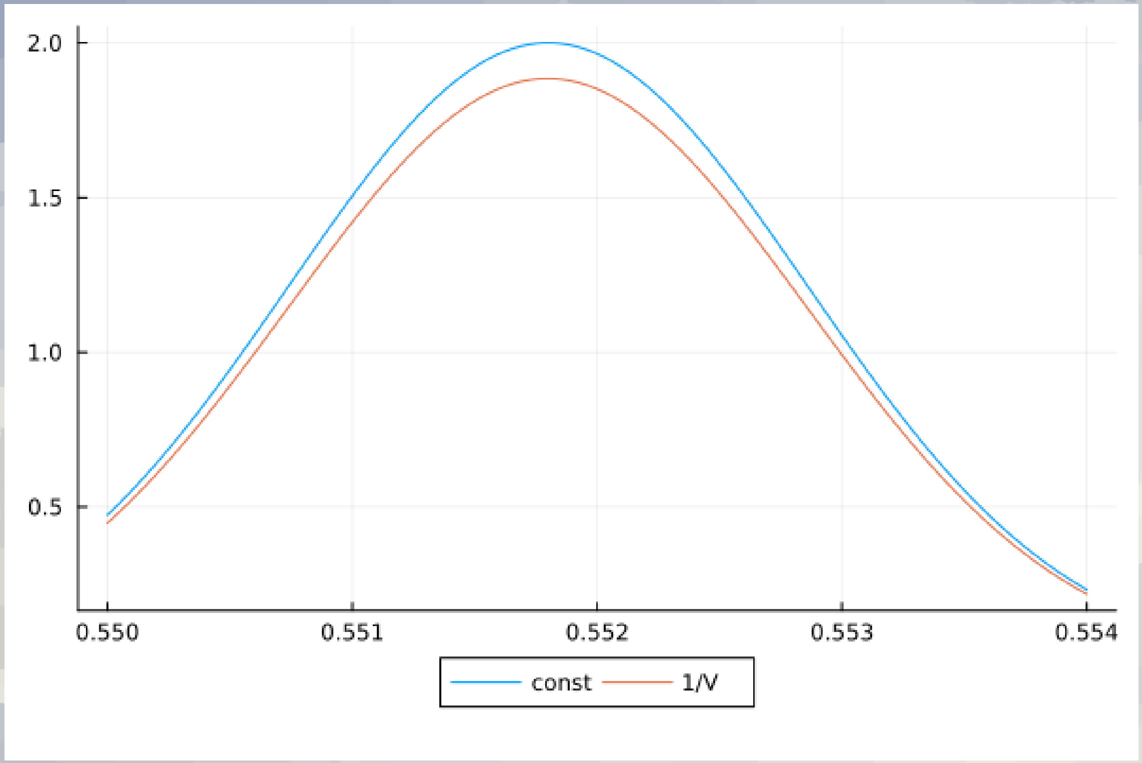
Example-II. conclusions



The output measurement is fully determined only if we provide the fully noise distribution !!!

Example-II. a priori dependences

Does *a priori* term influence final inference?



Conclusion

Inference about unknown parameters is just like a measurement corrupted by a noise term.

In both cases we want to learn about the parameter but our knowledge is always limited by many factors like, noise , limited number of data, lack of theoretical models, wrong initial expectations, etc.

We can quantified our vague or unprecised information about the parameter by means of probability distribution just like in case of noisy measurements.

See you May, 13th