Introduction to Probabilistic Reasoning

– inference, forecasting, decision –

Giulio D’Agostini

giulio.dagostini@roma1.infn.it

Dipartimento di Fisica

Università di Roma La Sapienza

– Part 3 –

“Probability is good sense reduced to a calculus” (Laplace)

Contents

- Summary of relevant formulae
- Propagating uncertainty the straight way
- Simple case of the Gaussian distribution :
  – when (and why) ML provides a good 'estimates
  – and why it can never tell something meaningfull about uncertainties
- the role of conjugate priors, with application to the Bernoulli trials ('binomial problem'): $\rightarrow$ beta pdf.
- Handling systematics, with the special case of two measurements with Gaussian response of the same detector, affected by a possible systematic error:
  – how the overall uncertainty increases
  – how the two results become correlated
- Examples with OpenBUGS (see web page) [$\rightarrow$ JAGS]

Important formulae

(For continous *uncertain variables*)

product rule: $f_{xy}(x,y\,|\,I) = f_{x|y}(x\,|\,y,I)f_y(y\,|\,I)$

$\rightarrow$ independence: $f_{xy}(x,y\,|\,I) = f_x(x\,|\,I)\,f_y(y\,|\,I)$

marginalizzation:
$$\int f_{xy}(x,y\,|\,I)\,dy = f_x(x\,|\,I)$$
$$\int f_{xy}(x,y\,|\,I)\,dx = f_y(y\,|\,I)$$

decomposition:
$$f_x(x\,|\,I) = \int f_{x|y}(x\,|\,y,I)f_y(y\,|\,I)\,dy$$
$$f_y(y\,|\,I) = \int f_{y|x}(y\,|\,x,I)f_x(x\,|\,I)\,dx$$

"weighted average"

Important formulae – continued

Bayes rule (using observation x and 'true value' μ and simplifying the notation)

$$\begin{aligned} f(\mu \,|\, x, I) &= \frac{(\mu, x \,|\, I)}{f(x \,|\, I)} \\ &= \frac{f(x \,|\, \mu, I) f(\mu \,|\, I)}{f(x \,|\, I)} \\ &= \frac{f(x \,|\, \mu, I) f(\mu \,|\, I)}{\int f(x \,|\, \mu, I) f(\mu \,|\, I)\, d\mu} \\ &\propto f(x \,|\, \mu, I) f(\mu \,|\, I) \end{aligned}$$

(In the following 'I' will be implicit in most cases)

Propagation of uncertainties

- “What has to do with *Bayesian*?”

Propagation of uncertainties

- “What has to do with *Bayesian*?”
  $\Rightarrow$ Think what ‘most’ of you have been taught in laboratory courses (typically):
  - Use probabilistic formulae for object that <u>are not</u> ‘random variables’ (in the frequentistic approach)

Propagation of uncertainties

- “What has to do with *Bayesian*?”
  $\Rightarrow$ Think what ‘most’ of you have been taught in laboratory courses (typically):
  - Use probabilistic formulae for object that are not ‘random variables’ (in the frequentistic approach)

  An insult to logic!!

Propagation of uncertainties

- "What has to do with *Bayesian*?"
  $\Rightarrow$ Think what 'most' of you have been taught in laboratory courses (typically):
  - Use probabilistic formulae for object that <u>are not</u> 'random variables' (in the frequentistic approach)

  An insult to logic!!
- In the Bayesian approach it is straightforward because true values <u>are</u> uncertain variables.

Propagation of uncertainties

- "What has to do with *Bayesian*?"
  $\Rightarrow$ Think what 'most' of you have been taught in laboratory courses (typically):
  - Use probabilistic formulae for object that <u>are not</u> 'random variables' (in the frequentistic approach)

  An insult to logic!!
- In the Bayesian approach it is straightforward because true values <u>are</u> uncertain variables.
  - For details see 2005 CERN lectures (nr. 4, pp. 17-28)
    `http://indico.cern.ch/conferenceDisplay.py`
    conferenceDisplay.py?confId=a043715

back to our slides

- a very simple inference in a gaussian model
- conjugate priors
- systematics

Simple case of Gaussian errors

$x \sim \mathcal{N}(\mu\,\sigma)$:

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

$$f(\mu\,|\,x, I) = \frac{\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right] f(\mu\,|\,I)}{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right] f(\mu\,|\,I)\,\mathrm{d}\mu}$$

IF $f(\mu\,|\,I) \approx$ const

$$f(\mu\,|\,x, I) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{(\mu-x)^2}{2\,\sigma^2}\right].$$

$\Rightarrow$ E$[\mu]$, $\sigma(\mu)$, etc.

Simple case of Gaussian errors

$x \sim \mathcal{N}(\mu\,\sigma)$:

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

$$f(\mu\,|\,x, I) = \frac{\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right] f(\mu\,|\,I)}{\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right] f(\mu\,|\,I)\,\mathrm{d}\mu}$$

IF $f(\mu\,|\,I) \approx$ const

$$f(\mu\,|\,x, I) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(\mu-x)^2}{2\,\sigma^2}\right].$$

$\Rightarrow$ E$[\mu]$, $\sigma(\mu)$, etc.

$\Rightarrow$ maximum of posterior $=$ maximum of likelihood $\surd$

Simple case of Gaussian errors

$x \sim \mathcal{N}(\mu\,\sigma)$:

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

$$f(\mu\,|\,x, I) = \frac{\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right] f(\mu\,|\,I)}{\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right] f(\mu\,|\,I)\,\mathrm{d}\mu}$$

IF $f(\mu\,|\,I) \approx$ const

$$f(\mu\,|\,x, I) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(\mu-x)^2}{2\,\sigma^2}\right].$$

$\Rightarrow$ E$[\mu]$, $\sigma(\mu)$, etc.

$\Rightarrow$ frequentistic equivalent of $\sigma(\mu)$?

Simple case of Gaussian errors

$x \sim \mathcal{N}(\mu\,\sigma)$:

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

$$f(\mu\,|\,x, I) = \frac{\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right]\,f(\mu\,|\,I)}{\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right]\,f(\mu\,|\,I)\,\mathrm{d}\mu}$$

IF $f(\mu\,|\,I) \approx$ const

$$f(\mu\,|\,x, I) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(\mu-x)^2}{2\,\sigma^2}\right].$$

$\Rightarrow$ E$[\mu]$, $\sigma(\mu)$, etc.

$\Rightarrow$ frequentistic equivalent of $\sigma(\mu)$? NO!

Simple case of Gaussian errors

$x \sim \mathcal{N}(\mu\,\sigma)$:

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

$$f(\mu\,|\,x, I) = \frac{\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right]\,f(\mu\,|\,I)}{\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right]\,f(\mu\,|\,I)\,\mathrm{d}\mu}$$

IF $f(\mu\,|\,I) \approx$ const

$$f(\mu\,|\,x, I) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(\mu-x)^2}{2\,\sigma^2}\right].$$

$\Rightarrow$ E$[\mu]$, $\sigma(\mu)$, etc.

$\Rightarrow$ frequentistic equivalent of $\sigma(\mu)$? NO! $\rightarrow$ 'prescriptions'!!

Prior conjugate

Since ever there have been computational problems:

$\rightarrow$ survive with approximations
e.g. "Gaussian approximation" of the posterior
(*somehow* a reflex of the Central Limit theorem)

Prior conjugate

Since ever there have been computational problems:

$\rightarrow$ survive with approximations
e.g. "Gaussian approximation" of the posterior
(*somehow* a reflex of the Central Limit theorem)

Another famous approximation is to chose a 'proper' shape for the prior:

$\rightarrow$ compromize between real beliefs and easy math!

Prior conjugate

Since ever there have been computational problems:

$\rightarrow$ survive with approximations
e.g. "Gaussian approximation" of the posterior
(*somehow* a reflex of the Central Limit theorem)

Another famous approximation is to chose a 'proper' shape for the prior:

$\rightarrow$ compromize between real beliefs and easy math!

$\rightarrow$ inferring p of binomial
$\rightarrow$ black/green board

Beta distribution

$X \sim$ Beta(r, s)):

$$f(x \,|\, \text{Beta}(r,s)) = \frac{1}{\beta(r,s)} x^{r-1}(1-x)^{s-1} \qquad \begin{cases} r,\; s > 0 \\ 0 \le x \le 1\,. \end{cases}$$

Denominator is just normalization, i.e.

$$\beta(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1}\, \mathrm{d}x\,.$$

$\rightarrow$ beta function, resulting in $\beta(r,s) = \frac{\Gamma(r)\,\Gamma(s)}{\Gamma(r+s)}$.

$\Rightarrow$ very flexible distribution

$\Rightarrow$ file beta_distribution.pdf

Uncertainties due to systematics

This is another subjects where the ’frequentistic’ approach fails miserably:

- no consistent theory, just ’prescriptions’
  1. add them linearly
  2. add the quadratically
  3. do 1 if small, 2 if large; . . .

Uncertainties due to systematics

This is another subjects where the ’frequentistic’ approach fails miserably:

- no consistent theory, just ’prescriptions’
  1. add them linearly
  2. add the quadratically
  3. do 1 if small, 2 if large; . . .

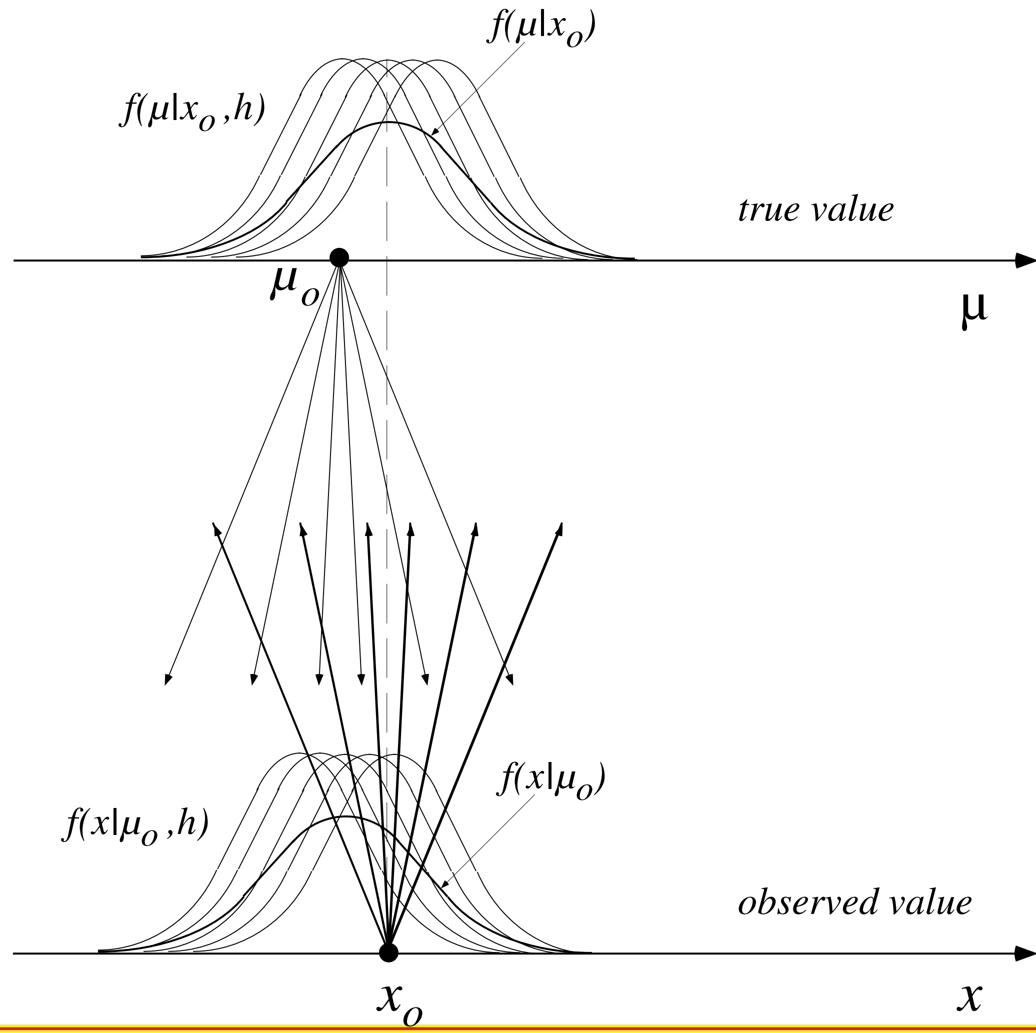
Straightforward in the bayesian approach:

→ influence quantities from which the final results may depend (calibrations constants, etc.) are characterized by an uncertain value, and hence we can attach to them probabilities, and use probability theory.
(This is what, in practice, also physicists wh think to adhere to fquentistic school finally do, as in error propagation.)

1. Conditional inference

Let's make esplicit one of the pieces of information:

$$f(\mu \,|\, x_0, I) = f(\mu \,|\, x_0, h, I_0)$$



1. Conditional inference

Let’s make esplicit one of the pieces of information:

$$f(\mu \,|\, x_0, I) = f(\mu \,|\, x_0, h, I_0)$$

then use a probability theory

$$f(\mu \,|\, x_0, I) = \int f(\mu \,|\, x_0, h, I_0)\, f(h \,|\, I_0)\, dh$$

Easy, logical, it ’does work’

try it!

2. Joint inference + marginalization

We can infer both μ and h from the data
(One piece of data, two inferred quantities? So what? they will just be 100% correlated)

$$f(\mu, h \,|\, x_0, I) \propto f(x_0 \,|\, \mu, h, I) \cdot f_0(\mu, h \,|\, I)$$

And then apply marginalization with respect to the so called *nuisance parameter*

$$f(\mu \,|\, x_0, I) = \int f(\mu\, h \,|\, x_0, I_0)\, f(h \,|\, I_0)\, dh$$

3. Raw result $\rightarrow$ corrected result

As third possibility, we might think at a raw result

- obtained at a fixed value of h (its nominal, 'best' value)
- only affected by uncertainties due to random errors (or all others, a part those coming from hour imperfect knpwledge about h $\;f(\mu_R\,|\,x_0, I)$

Then think to a corrections due to all possible values of h, consistently with out best knowledge:

$$\mu = g(\mu_R, h)$$

[$g(\mu_R, h)$ is not a pdf!]

$\rightarrow$ then use propagation of uncertainties

Example: $\mu = \mu_R + z$, where z is an offset known to be $0 \pm \sigma_z$

$\Rightarrow$ one of the assigned problems

Common systematic →correlation

As well understood, if measurements have a systematic effect in common, the results will become correlated.
As it happens when two quantities depend from a third one:

$$\begin{aligned}\mu_1 &= \mu_{R_1} + z \\ \mu_2 &= \mu_{R_2} + z\end{aligned}$$

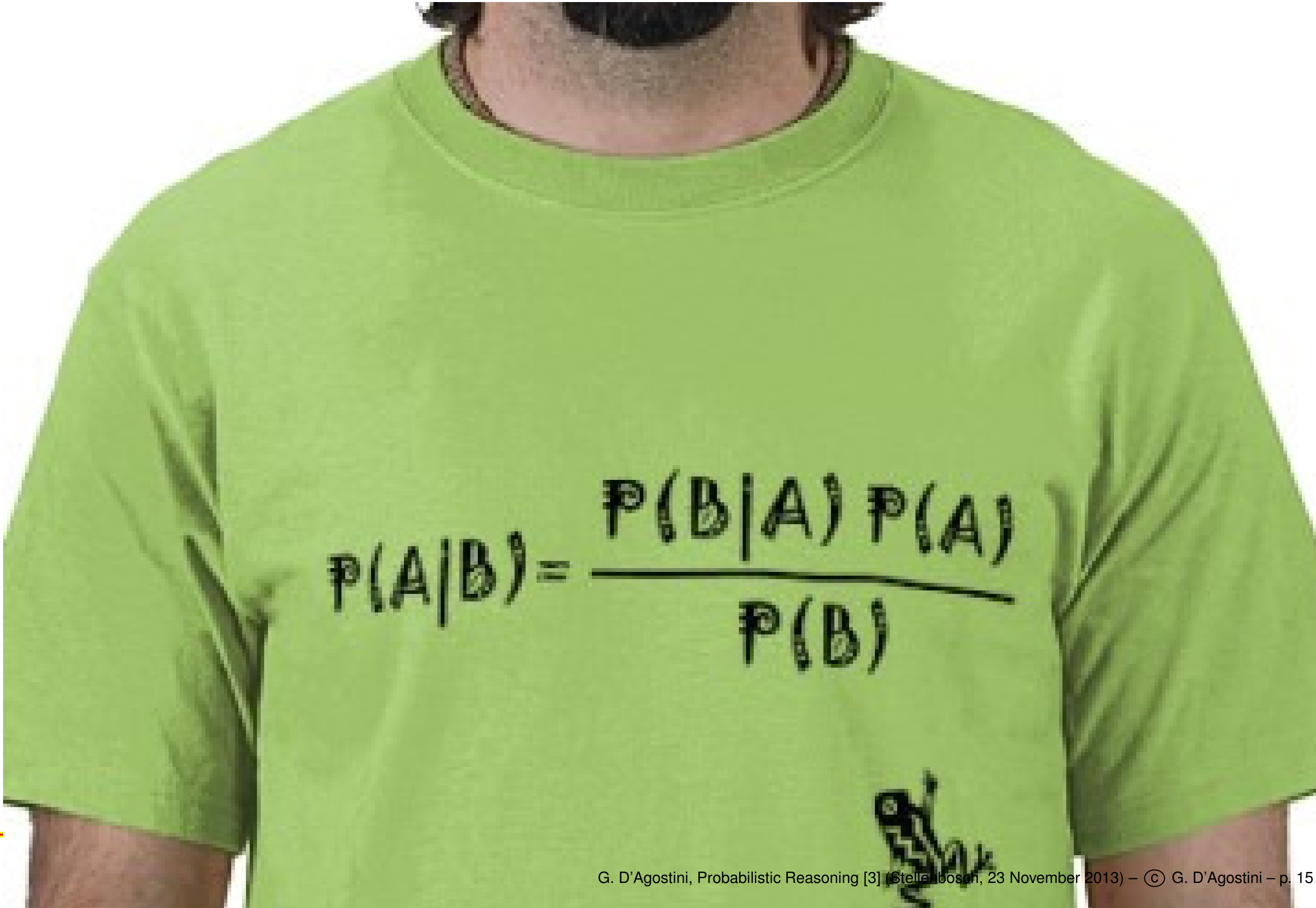
⇒ common uncertainty in z affects μ_1 and μ_2 in the same direction:

μ_1 and μ_2 will become positively correlated.

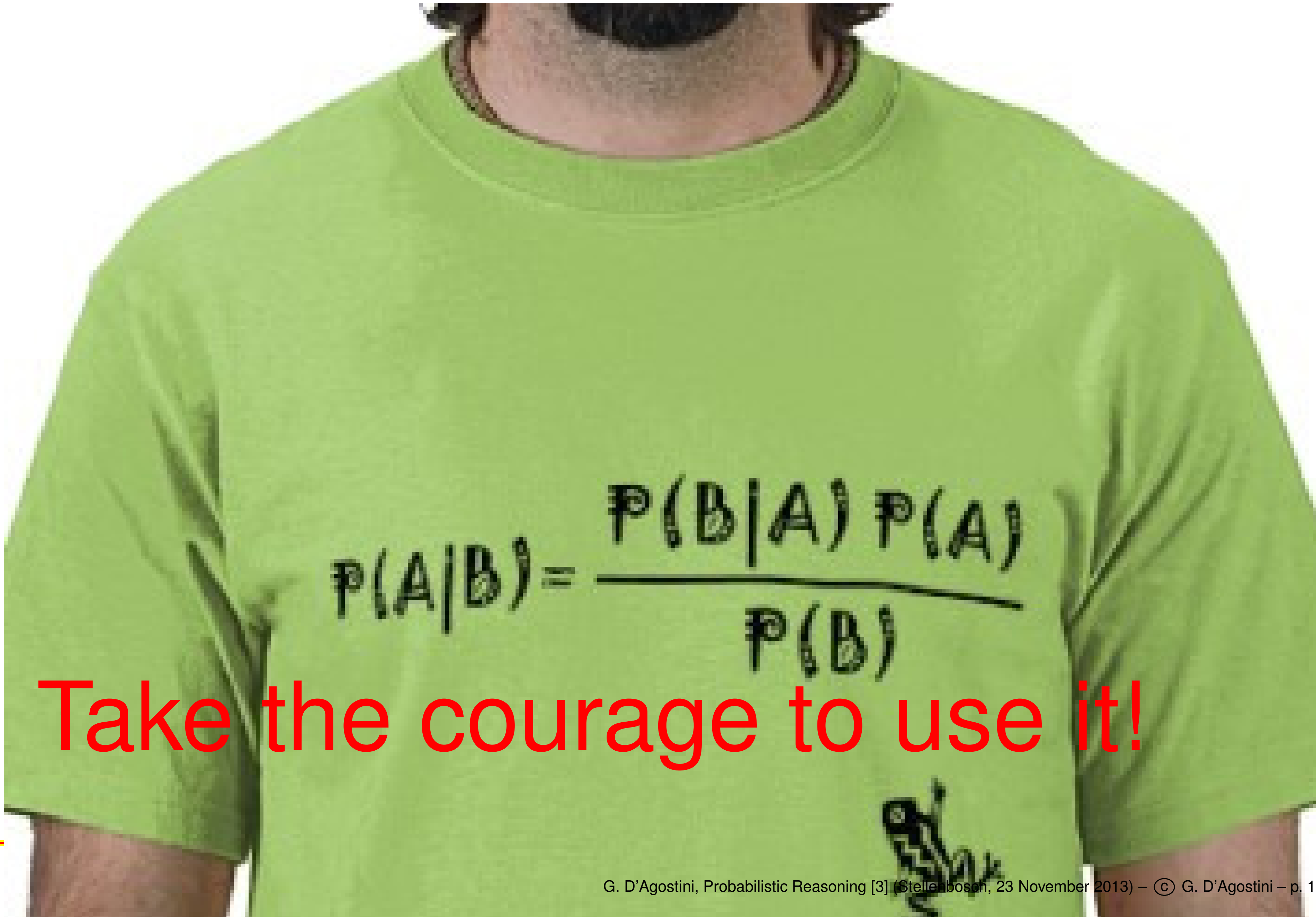
⇒ assigned problem

For a more detailed example, using '*reasoning 2*' see file em common_systematics.pdf (sections 6.8-6.10)

Conclusions

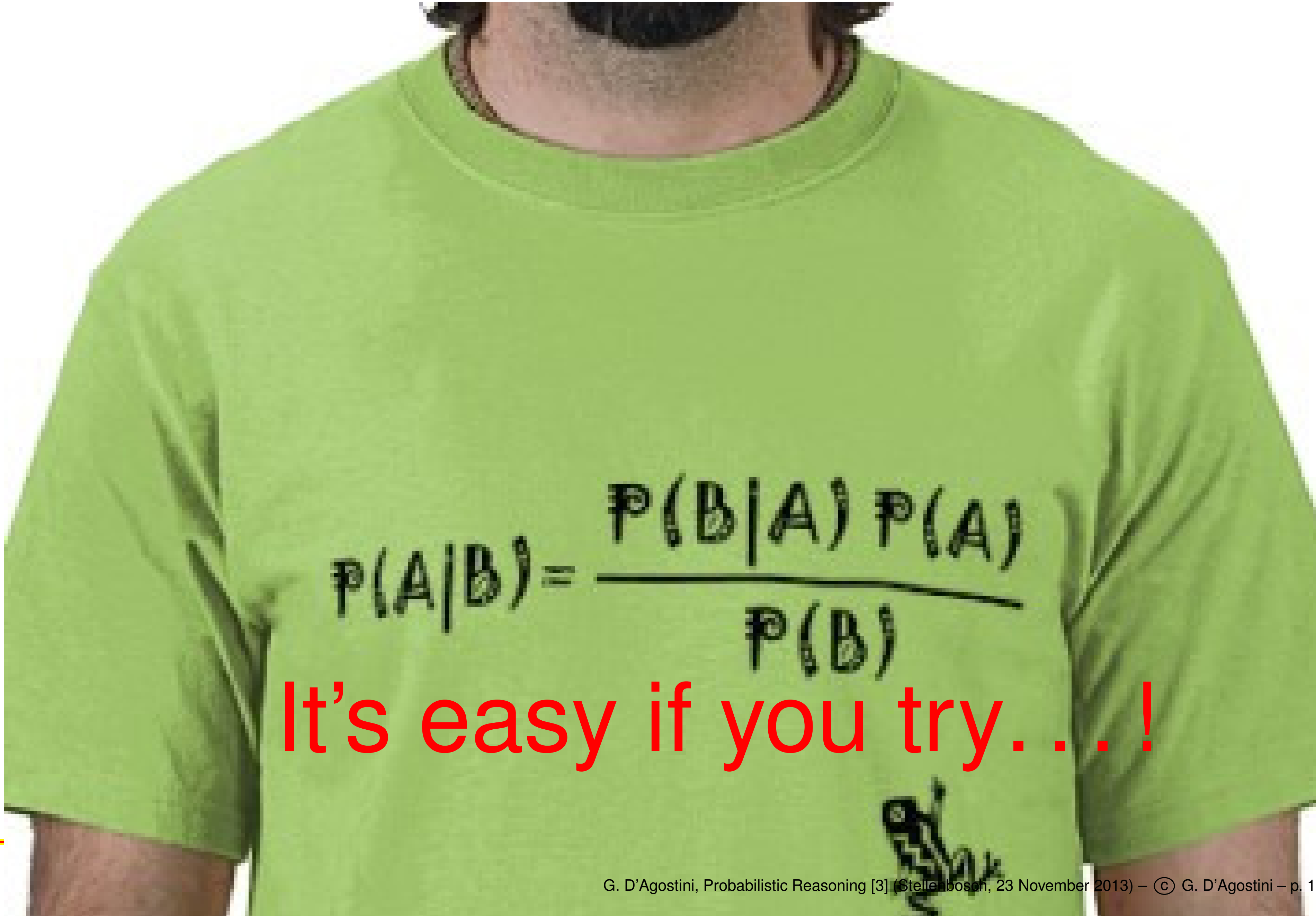


Conclusions



Take the courage to use it!

Conclusions



It's easy if you try. . . !