Advanced statistical methods and Bayesian inference in scientific research

Lecture 5

W. Dębski

29.04.2024

Bayes theorem - comments

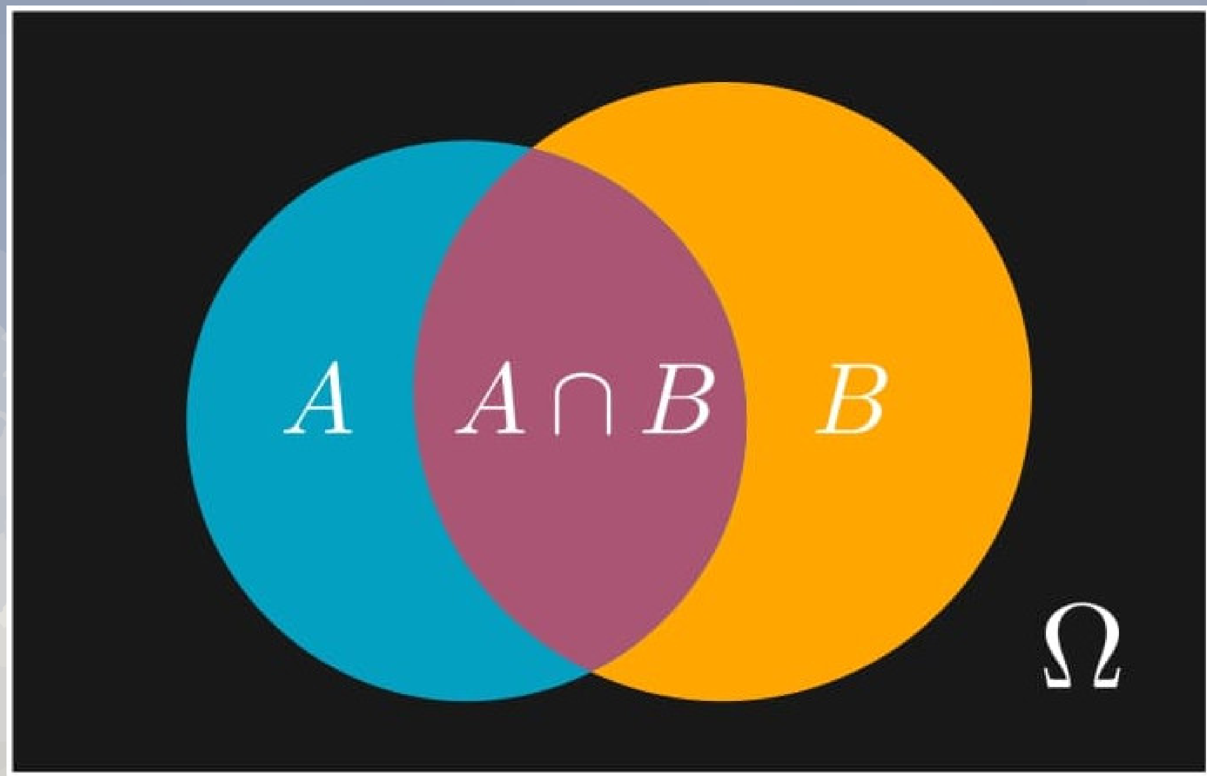
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

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

Bayes “theorem”

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

Bayes theorem - comments



Bayesian inference - random variable example

- Y - random variable we can measure
- X - random variable we are interested in
- we know X and Y are related by theory e.g. $Y = \bar{G}(X)$
- What can we say about X if we have measured $Y = y^o$?
- Can we evaluate its "accuracy" provided we know measurement uncertainties?

Bayesian inference - random variable example

- B - event that $Y \in [y_a, y_b] \quad\quad \Longrightarrow P(B)$
- A - random variable X have value in range $[x_a, x_b] \Longrightarrow P^{apr}(A)$
- X, Y related: conditional probability $P(Y|X)$ that theory predicts value $Y = y$ provided $X = x$ (possible “modelling” uncertainties)
- $P(X|Y)$ conditional probability of $X = x$ provided we have measured Y - object we are looking for

$$P^{post}(X) \;=\; P(X = x|Y = y^o) \;\sim\; P^{apr}(Y = y^o|X) \times P(X)$$

$$P^{apr}(X) \Longrightarrow P^{post}(X)$$

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: $\Sigma \times \Sigma \to \Sigma$

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

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

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

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

Using Inference space approach - example I

- random variable X is described by $\rho(x)$
- we perform another measurement of X

Question: how performed measurement constraints (update) the pdf distribution $p(x)$ describing X?

Assumption: noisy measurement with known noise characteristic $\psi(\cdot)$

$$x_o = x_o^{true} + n$$

$$X_o \Longrightarrow p(x) = \psi(x - x_o)$$

Using Inference space approach - example I

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

$$p^{(1)}(x) = \frac{1}{Z}\frac{\rho(x)\psi(x - x_o)}{\mu(x)}$$

$$Z = \int \rho(x)\psi(x - x_o)/\mu(x)dx$$

Comment:
Z (evidence) measure to what extend measurement is compatible with “apriori” $\rho(x)$

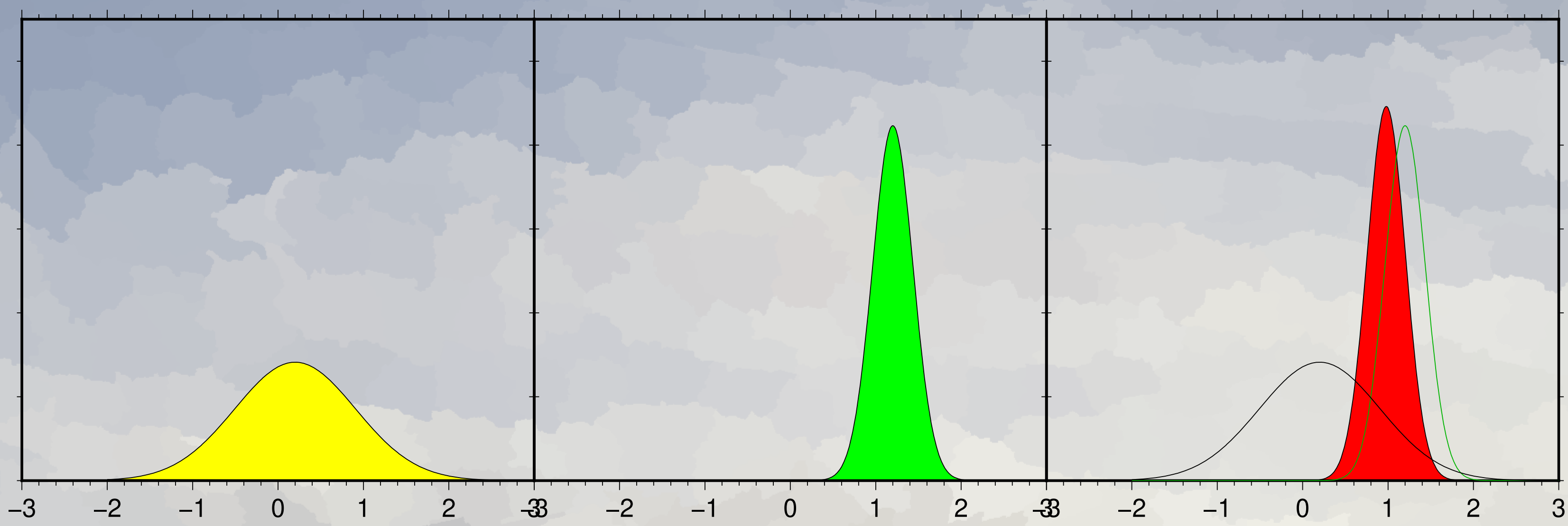
Using Inference space approach - example I

If another (independent) measurement is available

$$p^{(2)}(x) = \frac{1}{Z'} \frac{p^{(1)}(x)\psi(x - x_o)}{\mu(x)}$$

$$p^{(n)}(x) = \frac{1}{Z} \frac{p^{(n-1)}(x)\psi(x - x_o^{(n)})}{\mu(x)}$$

Using Inference space approach - example I



Using Inference space approach - example II

- random variable X is described by $\rho(x)$
- we perform a measurement of Y
- measurement errors are characterized by $\psi(y - y_o)$ distribution
- we know that X and Y are related as $Y \; = \; G(X) + \epsilon$
  and the relation is subjected to errors ϵ described by

$$\zeta(X, Y) \; = \; \zeta(Y - G(X))$$

Question:

how performed measurement constraints (update) knowledge of X ?

Using Inference space approach - example II

New "vectorized" random variable

$$X, Y \Longrightarrow Z := \begin{pmatrix} X \\ Y \end{pmatrix}$$

$$\mathcal{R} \times \mathcal{R} \Longrightarrow \mathcal{R}^2$$

$$\begin{aligned} \rho(x) &\rightarrow \rho'(z) = \rho(x)\mu(y) \\ \psi(y - y_o) &\rightarrow \psi'(z) = \psi(y - y_o)\mu(x) \\ \zeta(x, y) &\rightarrow \zeta'(z) \end{aligned}$$

Using Inference space approach - example II

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

$$p(z) = a \frac{\psi'(z)\zeta'(z)}{\mu(z)}$$

$$\sigma(z) = a \frac{p(z)\rho'(z)}{\mu(z)}$$

$$\sigma(z) = \frac{1}{Z} \frac{\zeta(z)\,\psi'(z)\,\rho'(z)}{\mu^2(z)}$$

Using Inference space approach - example II

Taking marginal integrals

$$\sigma(x) = \int_Y \sigma(z)\mathrm{d}y$$

$$\sigma(x) = \frac{1}{Z}\,\rho(x) \int\limits_Y \psi(y - y_o)\zeta(y - G(x))\mathrm{d}y$$

$$Z = \int\limits_X \int\limits_Y \rho(x)\psi(y - y_o)\zeta(y - G(x))\mathrm{d}y\mathrm{d}x$$

Using Inference space approach - example II

If the relation between X and Y is exactly known

$$\zeta(y - G(x)) \;=\; \delta(y - G(x))$$

$$\sigma(x) = \frac{1}{Z}\,\rho(x)\,\psi(y_o - G(x))$$

$$p^{(1)}(x) \;=\; \frac{1}{Z}\,\frac{\rho(x)\psi(x_o - x)}{\mu(x)}$$

Evidence

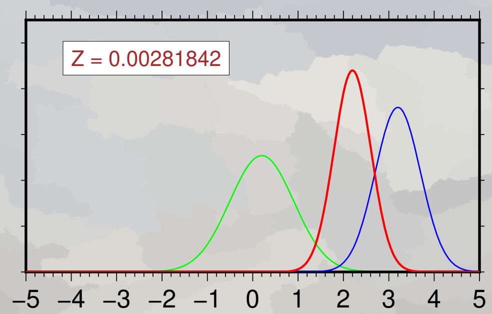
$$Z = \int_X \int_Y \rho(x)\psi(y - y_o)\zeta(y - G(x))\mathrm{d}y\mathrm{d}x$$

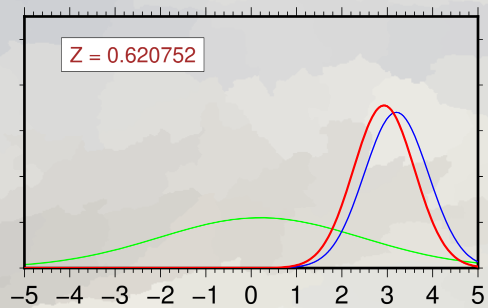
$$\mathcal{L}(x) = \int_Y \psi(y - y_o)\zeta(y - G(x))\mathrm{d}y$$

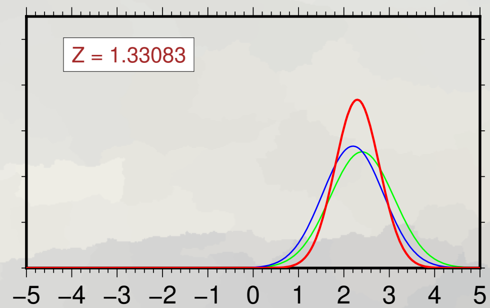
$$Z = \int_X \rho(x)\mathcal{L}(x)\mathrm{d}x$$

Evidence

- $\rho(x)$ - green
- $\mathcal{L}(x)$ - blue
- $\sigma(x) = \rho(x)\mathcal{L}(x)$ - red







Using Inference space approach - example III (model selection)

- two random variables $(X\,,Y)$
- three measurements: (X_i, Y_i^o) measurement errors: $\psi()$
- two different "theories" $Y = G(X), \quad Y = F(X)$
  (with adjustable parameters)
- does observation approve/falsify them?

Simplest answer: Let check which theory provides "best fit to Y^i

$$||Y - Y_i|| = \sqrt{\sum_i (Y - Y_i)^2}$$

Using Inference space approach - example III (model selection)

Physical example: movement of mass m

Physical parameters (random variables X,Y) - $(V,\ E_k)$

Theory G:

$$E_k = \frac{1}{2} m V^2$$

Theory F:

$$E_k = \frac{1}{2} m V^2 - \gamma \ln(V/V_r)$$

Measured values $(V_1, E_k^1),\ (V_2, E_k^2) \cdots$

Theory evaluation:

$$R_G \ = \sum_i \left(E_k^i - \frac{1}{2} m V_i^2 \right)^2$$

$$R_F \ = \sum_i \left(E_k^i - \frac{1}{2} m V_i^2 - \gamma \ln(V_i/V_r) \right)^2$$

Using Inference space approach - example III (model selection)

- "Fixed theory"

$$R_G = ||Y^o - G(X_i)||$$
$$R_F = ||Y^o - G(X_i)||$$

If $R_G < R_F$ G theory explain better data than F. *It is "better" one*

The only problem may appear if none of theory well fits data, i.e. R_G, R_F are very large

Using Inference space approach - example III (model selection)

Possible problems for parameter free theories:

- different complexity of theories

$$G(X) = a + b\,X$$
$$F(X) = a + b\,X + c\,X^2$$

$$||Y^o - G(X_i)|| = \min \qquad ||Y^o - F(X_i)|| = \min$$

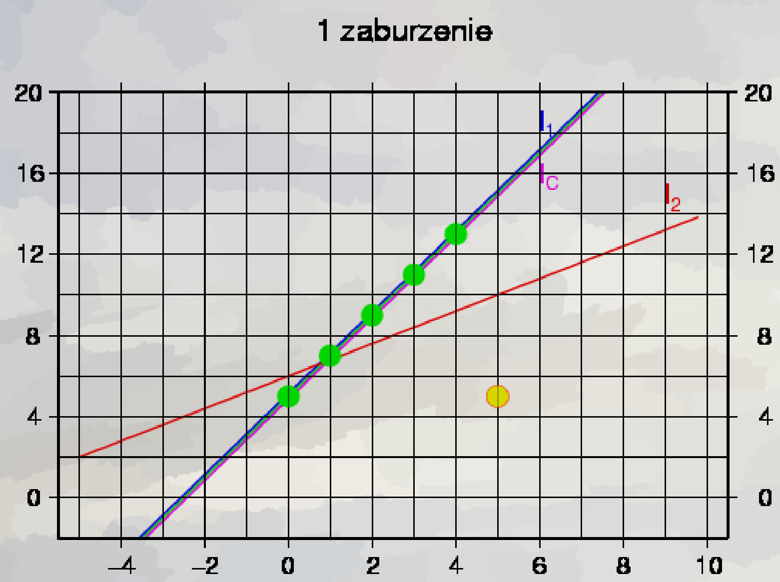
More complex theory will always gives "better fit" to data if we can adjust parameters

$$||Y^o - F(X_f)|| = 0 \qquad \text{while} \qquad ||Y^o - G(X_f)|| > 0$$

Using Inference space approach - example III (model selection)

Possible problems:

✦ noisy data



Using Inference space approach - example III (model selection)

Exhaustive solution: Bayesian inference

$$\sigma_G(\mathbf{a}) = \frac{1}{Z_G}\,\rho(x)\,\psi(Y^o - G(X;\mathbf{a}))$$

$$\sigma_F(\mathbf{a}) = \frac{1}{Z_F}\,\rho(x)\,\psi(Y^o - F(X;\mathbf{a}))$$

Solution:

Inspecting resulting posteriori $\sigma()$ distribution and measuring their “goodness”

Using Inference space approach - example III (model selection)

- Comparing values of evidence: Z_F Z_G.
  This does not take into account difference in theory complexity

- Calculate entropy

$$H[\sigma] = -\int \ln(\sigma)\, \sigma dx$$

Example: normal distribution

$$H[e^{-(x/\sigma)^2}] = \frac{1}{2} \ln(2\pi e \sigma^2)$$

$$H[\sigma_F] = \ln(Z_F) - \int_a \{\ln(\rho(a)) + \ln(\psi(a))\}\, \sigma(a) \mathrm{d}a$$

Using Inference space approach - example III (model selection)

Non-informative $\rho(a)$, for example: $\rho() = const$

$$H[\sigma_F] \;=\; \ln(Z_F) - \int_a \ln(\psi(a))\,\sigma(a)\mathrm{d}a$$

where $\psi()$ describes measurement errors.

$$\psi(x) = \exp(-||x||)$$

$$H[\sigma_F] \;=\; \ln(Z_F) + \int_a ||Y^o - G_F(X;\mathbf{a})||\,\sigma(a)\mathrm{d}a$$

Still no theory complexity is taken into account

Akaike information criterion

Let theory $G()$ contains n adjustable parameters

$$AIC = 2n - 2\ln(\bar{\mathrm{L}})$$

Extended version:

$$AIC[\sigma] = 2n - 2H[\sigma]$$

Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value

Bayesian information criterion

Let theory $G()$ contains n adjustable parameters and to construct $\sigma(a)$ we used k observational data

$$BIC = n \ln(k) - 2 \ln(\bar{\mathrm{L}})$$

Extended version:

$$BIC[\sigma] = n \ln(k) - 2H[\sigma]$$

Given a set of candidate models for the data, the preferred model is the one with the minimum BIC value

See you May, 6th