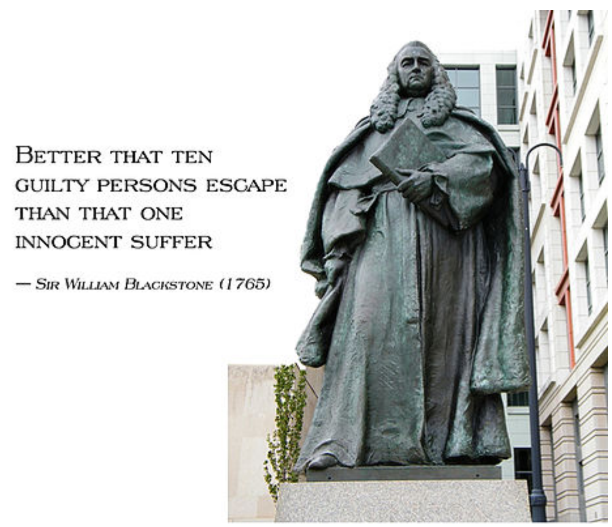
Slides 8: Bayesian inference (2)

- Bayesian estimation
- Loss function
- Posterior risk
- Credibility intervals
- Bayesian hypothesis testing
- Posterior odds
- A case study



Two Polish mathematicians, Tomasz Gliszczynski and Waclaw Zawadowski, set their university statistics classes to research the subject with the Belgian one euro coin. The test was carried out by spinning the coins on a table rather than tossing them in the air.

Out of $n = 250$ spins,

$x = 140$ showed the head of the Belgian monarch, King Albert, while $n - x = 110$ showed the one euro symbol.

The probability of heads p is the parameter of interest.

Zoom discussion.

The Bayesian approach treats p as a random variable. How can this make sense? Discuss this issue in groups of 3-4 students.

Think also of an appropriate prior distribution.

In the language of decision theory we are searching for an optimal action

$$\text{action } a = \{\text{assign value } a \text{ to unknown parameter } \theta\}.$$

The optimal a depends on the choice of the loss function $l(\theta, a)$. Bayes action minimises posterior risk

$$R(a|x) = \mathrm{E}\big(l(\Theta, a)|x\big)$$

so that

$$R(a|x) = \int l(\theta, a) h(\theta|x) d\theta \quad \text{or} \quad R(a|x) = \textstyle\sum_\theta l(\theta, a) h(\theta|x).$$

We consider two loss functions

1. Zero-one loss function: $l(\theta, a) = 1_{\{\theta \neq a\}}$
2. Squared error loss: $l(\theta, a) = (\theta - a)^2$

leading to two Bayesian estimators

1. $\hat{\theta}_{\mathrm{map}}$ maximum aposteriori probability
2. $\hat{\theta}_{\mathrm{pme}}$ posterior mean estimate

1. Using the zero-one loss function we find that the posterior risk is the probability of misclassification

$$R(a|x) = \sum_{\theta \neq a} h(\theta|x) = 1 - h(a|x).$$

It follows that to minimise the risk we have to maximise the posterior probability. We define $\hat{\theta}_{\text{map}}$ as the value of θ that maximises $h(\theta|x)$. Observe that with the uninformative prior, $\hat{\theta}_{\text{map}} = \hat{\theta}_{\text{mle}}$.

2. Using the squared error loss function we find that the posterior risk is a sum of two components

$$R(a|x) = \mathrm{E}((\Theta - a)^2|x) = \mathrm{Var}(\Theta|x) + [\mathrm{E}(\Theta|x) - a]^2.$$

Since the first component is independent of a, we minimise the posterior risk by putting

$$\hat{\theta}_{\text{pme}} = \mathrm{E}(\Theta|x) = \int \theta h(\theta|x) d\theta.$$

Posterior mean estimate.

Let x stand for the data in hand. For a 95% confidence interval formula

$$I_\theta = (a_1(x), a_2(x)),$$

the parameter θ is an unknown constant and a the confidence interval is treated as random

$$\mathrm{P}(a_1(X) < \theta < a_2(X)) = 0.95.$$

A credibility interval (or credible interval)

$$J_\theta = (b_1(x), b_2(x))$$

is treated as a nonrandom interval. A 95% credibility interval is computed as $\mathrm{P}(b_1 < \Theta < b_2) = 0.95$ using posterior distribution

$$\int_{b_1}^{b_2} h(\theta|x)d\theta = 0.95.$$

Question. Which one of the two explanations of 95% is more intuitive?

Example: IQ measurement

Given $n = 1$, we have $\bar{X} \sim \mathrm{N}(\mu; 10)$ and an exact 95% confidence interval for μ takes the form

$$I_\mu = 130 \pm 1.96 \cdot 10 = 130 \pm 19.6.$$

Posterior distribution of the mean is $\mathrm{N}(120.7; 8.3)$ and therefore a 95% credibility interval for μ is

$$J_\mu = 120.7 \pm 1.96 \cdot 8.3 = 120.7 \pm 16.3.$$

Bayesian hypotheses testing

Consider a choice between two simple hypotheses H_0: $\theta = \theta_0$ and H_1: $\theta = \theta_1$ given the likelihoods $f(x|\theta_0)$, $f(x|\theta_1)$ and prior probabilities

$$\mathrm{P}(H_0) = \pi_0, \quad \mathrm{P}(H_1) = \pi_1.$$

Decision should be taken depending on the following four cost values

	Decision	H_0 true	H_1 true
$x \notin \mathcal{R}$	Accept H_0	cost $= 0$	$c_1 =$ the error type II cost
$x \in \mathcal{R}$	Accept H_1	$c_0 =$ the error type I cost	cost $= 0$

For a given rejection region $\mathcal{R}$, the average cost is

$$c_0\pi_0 \mathrm{P}(X \in \mathcal{R}|H_0) + c_1\pi_1 \mathrm{P}(X \notin \mathcal{R}|H_1) = c_1\pi_1 + \int_{\mathcal{R}} \Big(c_0\pi_0 f(x|\theta_0) - c_1\pi_1 f(x|\theta_1)\Big)dx.$$

Now observe that

$$\int_{\mathcal{R}} \Big(c_0\pi_0 f(x|\theta_0) - c_1\pi_1 f(x|\theta_1)\Big)dx \geq \int_{\mathcal{R}^*} \Big(c_0\pi_0 f(x|\theta_0) - c_1\pi_1 f(x|\theta_1)\Big)dx,$$

where

$$\mathcal{R}^* = \{x : c_0\pi_0 f(x|\theta_0) < c_1\pi_1 f(x|\theta_1)\}.$$

It follows that the rejection region minimising the average cost is $\mathcal{R} = \mathcal{R}^*$. Thus the optimal decision rule is to reject H_0 for x such that

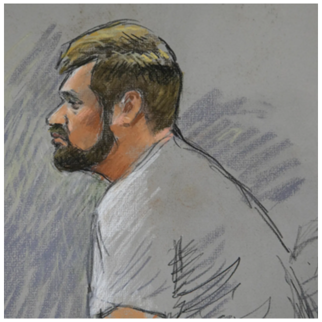
$$\frac{f(x|\theta_0)}{f(x|\theta_1)} < \frac{c_1\pi_1}{c_0\pi_0},$$

or in other terms, we reject H_0 for small values of the posterior odds

$$\frac{h(\theta_0|x)}{h(\theta_1|x)} < \frac{c_1}{c_0}.$$

The defendant "A" charged with rape, is a male of age 37 living in the area not very far from the crime place. The jury have to choose between two alternative hypotheses

$$H_0\text{: "A" is innocent, } H_1\text{: "A" is guilty.}$$



An uninformative prior probability

$$\pi_1 = \frac{1}{200000}, \text{ so that } \frac{\pi_0}{\pi_1} = 200000$$

takes into account the number of males who theoretically could have committed the crime without any evidence taken into account.

There were tree conditionally independent pieces of evidence

E_1: strong DNA match — evidence in favour of H_1

E_2: "A" is not recognised by the victim — evidence in favour of H_0

E_3: an alibi supported by "A"'s girlfriend — evidence in favour of H_0

The reliability of these pieces of evidence was quantified as

$\mathrm{P}(E_1|H_0) = \frac{1}{200{,}000{,}000}$, $\mathrm{P}(E_1|H_1)=1$, $\frac{\mathrm{P}(E_1|H_0)}{\mathrm{P}(E_1|H_1)} = \frac{1}{200{,}000{,}000}$

$\mathrm{P}(E_2|H_1) = 0.1$, $\mathrm{P}(E_2|H_0) = 0.9$, $\frac{\mathrm{P}(E_2|H_0)}{\mathrm{P}(E_2|H_1)} = 9$

$\mathrm{P}(E_3|H_1) = 0.25$, $\mathrm{P}(E_3|H_0) = 0.5$, $\frac{\mathrm{P}(E_3|H_0)}{\mathrm{P}(E_3|H_1)} = 2$

Then the posterior odds was computed as

$$\frac{\mathrm{P}(H_0|E)}{\mathrm{P}(H_1|E)} = \frac{\pi_0 \mathrm{P}(E|H_0)}{\pi_1 \mathrm{P}(E|H_1)} = \frac{\pi_0}{\pi_1} \frac{\mathrm{P}(E_1|H_0)}{\mathrm{P}(E_1|H_1)} \frac{\mathrm{P}(E_2|H_0)}{\mathrm{P}(E_2|H_1)} \frac{\mathrm{P}(E_3|H_0)}{\mathrm{P}(E_3|H_1)} = 0.018.$$

Thus we reject H_0 if the cost values are assigned so that

$$\frac{c_1}{c_0} = \frac{\text{cost for unpunished crime}}{\text{cost for punishing an innocent}} > 0.018.$$

Question. What would be your decision as a jury member?