Bayesian statistics, simulation and software

Module 3: Bayesian principle, binomial model and conjugate priors

Jesper Møller and Ege Rubak

Department of Mathematical Sciences
Aalborg University

Motivating example: Spelling correction

(Adapted from *Bayesian Data Analysis* by Gelman et al.)

- **Problem**: Someone types '*radom*'.
- **Question**: What did they mean to type? Random?

Ingredients:

- **Data** x: The observed word — *radom*.
- **Parameter of interest** θ: The correct word.

Comments: To solve this we need

- *background information* on which words are usually typed;
- an idea about how words are typically mistyped.

Bayesian idea

- **Data/observation model**: Conditional on θ (here the correct word), data x (here the typed word) is distributed according to a density

$$\pi(x|\theta) \propto L(\theta; x) \quad \leftarrow \text{the likelihood.}$$

- **Prior**: Prior knowledge (i.e. *before* collecting data) about θ is summaried by a density

$$\pi(\theta) \quad \leftarrow \text{the prior.}$$

- **Posterior**: The updated knowledge about θ *after* collecting data: The conditional distribution of θ given data x is

$$\pi(\theta|x) = \frac{\pi(x|\theta)\pi(\theta)}{\pi(x)}$$
$$\propto \pi(x|\theta)\pi(\theta)$$
$$(\text{“}posterior \propto likelihood \times prior''\text{)}.$$

Example: Prior

Google provides the following *prior probabilities* for three candidate words:

θ	$\pi(\theta)$
random	7.60×10^{-5}
radon	6.05×10^{-6}
radom	3.12×10^{-7}

Comments

- The relative high probability for the word *radom* is perhaps surprising: Name of a city in Poland (home to the biennial Radom Air Show) and the popular unofficial name for a semiautomatic 9 mm Para pistol of Polish design.
- Probably, in the context of writing a scientific report these prior probabilities would look different.

Example: Likelihood

Google's model of spelling and typing errors provides the following *conditional probabilities/likelihoods*:

θ	$\pi(x = \text{'radom'}\|\theta)$
random	0.00193
radon	0.000143
radom	0.975

Comments

- This is *not* a probability distribution!
- If one in fact intends to write 'radom' this actually happens in 97.5% of cases.
- If one intends to write either 'random' or 'radon' this is rarely misspelled 'radom'.

Example: Posterior

Combining the prior and likelihood we obtain the *posterior probabilities*:

$$\pi(\theta|x) = \frac{\pi(x|\theta)\pi(\theta)}{\pi(x)} \propto \pi(x|\theta)\pi(\theta)$$

θ	$\pi(x = \text{'radom'}|\theta)\pi(\theta)$	$\pi(\theta|x = \text{'radom'})$
random	1.47×10^{-7}	0.325
radon	8.65×10^{-10}	0.002
radom	3.04×10^{-7}	0.673

Conclusion

- With the given prior and likelihood the word 'radom' is twice as likely as 'random'.

Criticism

- The posterior probability for 'radom' may seem too high.
- Likelihood or prior to blame — or both?
- Likelihood is perhaps OK in this case.
- Prior depends on context — and hence might be “wrong”.

Binomial example

- **Data model**: Binomial, $X \sim \mathrm{B}(n,p)$, n known.

$$\pi(x|p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0,1,\ldots,n, \quad 0 \le p \le 1.$$

- **Prior**: Beta distribution,

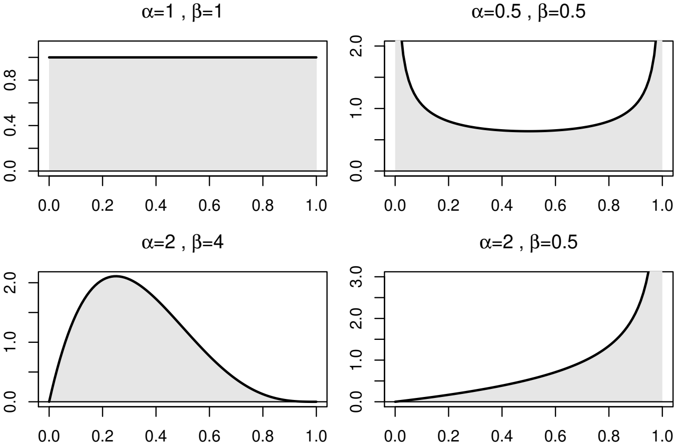
$$\pi(p) \sim \mathrm{Be}(\alpha,\beta),$$

  where we have to specify the parameters $\alpha > 0$ and $\beta > 0$, and the Beta distribution has density/pdf

$$\pi(p) = \begin{cases} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} p^{\alpha-1}(1-p)^{\beta-1} & \text{for } 0 \le p \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

If $\alpha = \beta = 1$, then $\pi(p) = 1$ for $0 \le p \le 1$ (the uniform distribution).

Beta distribution: Examples



Mean: $\mathbb{E}[p] = \dfrac{\alpha}{\alpha+\beta}$, Variance: $\mathbb{V}\mathrm{ar}[p] = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Binomial example — cont.

- **Data model**: $X \sim B(n, p)$.
- **Prior**: $\pi(p) \sim Be(\alpha, \beta)$, that is

$$\pi(p) = \begin{cases} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} p^{\alpha-1}(1-p)^{\beta-1} & \text{for } 0 \le p \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

- **Posterior**:

$$\begin{aligned} \pi(p|x) &\propto \pi(x|p)\pi(p) \\ &= \binom{n}{x} p^x (1-p)^{n-x} \quad \cdot \quad \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} p^{\alpha-1}(1-p)^{\beta-1} \\ &\propto p^{x+\alpha-1}(1-p)^{n-x+\beta-1} \\ &\sim Be(x+\alpha, n-x+\beta). \end{aligned}$$

Conjugate priors

In the binomial example: Both prior and posterior were beta distributions! Very convenient!

We say that the beta distribution is *conjugate*.

Definition: Conjugate priors

Let $\pi(x|\theta)$ be the data model. A class Π of prior distributions for θ is said to be *conjugate* for $\pi(x|\theta)$ if

$$\pi(\theta|x) \propto \pi(x|\theta)\pi(\theta) \in \Pi$$

whenever $\pi(\theta) \in \Pi$. That is, prior and posterior are in the same class of distributions.

Notice: Π should be a class of "tractable" distributions for this to be useful.

Posterior mean & variance

Posterior

$$\pi(p|x) \sim Be(x+\alpha, n-x+\beta).$$

Posterior mean

$$\mathbb{E}[p|x] = \frac{x+\alpha}{(x+\alpha)+(n-x+\beta)} = \frac{x+\alpha}{\alpha+\beta+n}.$$

If $n \gg \max\{\alpha, \beta\}$, then $\mathbb{E}[p|x] \approx \frac{x}{n}$ (the "natural" unbiased estimate).

Posterior variance

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[p|x] &= \frac{(x+\alpha)(n-x+\beta)}{(x+\alpha+n-x+\beta)^2(x+\alpha+n-x+\beta+1)} \\
&= \frac{(x+\alpha)(n-x+\beta)}{(\alpha+\beta+n)^2(\alpha+\beta+n+1)} \\
&= \frac{(\frac{x}{n}+\frac{\alpha}{n})(\frac{n-x}{n}+\frac{\beta}{n})}{(\frac{\alpha+\beta+n}{n})^2(\alpha+\beta+n+1)} \approx \frac{\frac{x}{n}\frac{n-x}{n}}{\alpha+\beta+n+1} \to 0 \quad \text{as } n \to \infty.
\end{aligned}$$

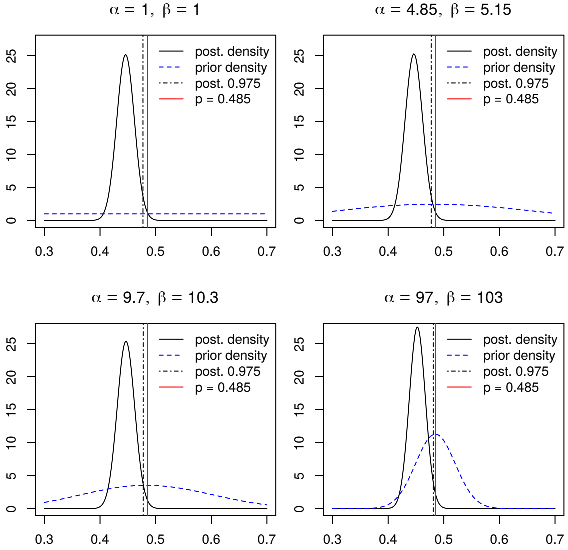
Example: Placentia Previa (PP)

- **Question**: Is the sex ratio different for PP births compared to normal births?
- **Prior knowledge**: 48.5% of new-borns are girls.
- **Data**: Of $n = 980$ cases of PP x=437 were girls (437/980=44.6%).
- **Data model**: $X \sim B(n, p)$.
- **Prior**: $\pi(p) \sim Be(\alpha, \beta)$.
- **Posterior**:

$$\begin{aligned}\pi(p|x) &\sim Be(x + \alpha, n - x + \beta) \\ &= Be(437 + \alpha, 543 + \beta)\end{aligned}$$

How to choose α and β, and what difference does it make?

Placenta Previa: Beta priors and posteriors



Conclusion (see the notes or Gelman et al. (2014))

For the different choices of priors, 95% posterior intervals (as defined later) for p do not contain 48.5%, which indicates that the probability for a female birth given placenta previa is lower than in the general population.