

Bayesian Analysis: Practical Sessions

Session 2: Posterior Distribution and Posterior Predictive Distribution.

Goals:

To Compute and draw:

- the posterior distribution
- the posterior predictive distribution

Exercise 2.1. Sèpia Verda: Estimating a (Poisson) frequency. The members of a cultural association called *La Sèpia Verda* don't know the expected number of weekly visitors to their web page. To find out, they register the number of weekly visitors in the last 10 weeks. This data can be found in the file *sepiaverda.txt*. The members of the association believe that the number of visitors will rarely fall under 5 or above 40.

- Choose the parameters of a conjugate prior distribution, and explain why you choose them (it might be useful to draw the prior predictive distribution to back your choice up).
- Draw in the same graph the prior distribution and the likelihood function.
- Draw the prior predictive distribution.
- Draw the posterior distribution, and give a point estimate and a 95% credible interval for the expected value of the number of weekly visitors.
- Compute the probability that the number of visitors next week will be lower than 10.
- Draw the posterior predictive distribution.
- Imagine that the members of the cultural association want to do a bet about the number of visitors next week. For what option will you bet?
 - $\tilde{y} < 15$
 - $15 \leq \tilde{y} < 25$
 - $25 \leq \tilde{y}$

Now, assume that the members of the association know nothing about the number of weekly visitors. They assume that all the possible values for the poisson's parameter, λ , are equally likely and hence use a flat prior, $\pi(\lambda)=1$ for $\lambda>0$, which is improper.

- Compute and draw in the same graph the prior distribution, the likelihood function and the posterior distribution.

Exercise 2.2. Asthma: Estimating a (binomial) proportion. A professional health worker from Sabadell needs to estimate the percentage of asthmatic people in that city. For this purpose a random sample of 200 citizens is taken, and 11 of them turn out to be asthmatic.

a) Choose the parameter's values for the prior distribution for next two scenarios:

a.1) Based on his experience, the professional health worker believes that the prevalence will be around 5%, and that it is very unlikely that it is larger than 20%.

a.2) There isn't any prior information about the prevalence of asthmatic people.

For every scenario:

- b) Draw the prior distribution, the likelihood function and the posterior distribution in the same graph.
- c) Give a point estimate and a 95% credible interval for the percentage of asthmatic people in Sabadell.
- d) Draw the prior and posterior predictive distribution.
- e) If the sample size was 10 instead of 200 citizens, and we had observed only one of them to be asthmatic, what would be different?

Exercise 2.3. Traffic accidents (Poisson). The weekly number of traffic accidents on *La Garriga's* highway has the $Poisson(\mu)$ distribution. Three students are going to count the number of accidents for each of the next eight weeks. Bru has no prior information, so he decides to use a $Gamma(0.01, 0.001)$ as a prior for μ , because it has a big variance. Clàudia lives in *La Garriga*, and she decides that a $Gamma(6.25, 2.5)$ captures well her prior information. Finally, Carles claims that he has no prior information, and so he will assume that all the possible values for μ are equally likely and hence use a flat prior, $\pi(\mu)=1$ for $\mu>0$, which is improper. The number of highway accidents during the next 8 weeks are: 3, 2, 0, 8, 2, 4, 6, 1.

- a) Draw these three prior distributions in the same graph.
- b) Draw the likelihood function
- c) Draw the three posterior distributions in the same graph.
- d) Think about the differences in using different priors.

Exercise 2.4. EuroVegas (Binomial). Two students are modeling their prior belief about the proportion of residents of *Baix Llobregat* that support building a macro casino in *Baix Llobregat*. Anna lives in *Gavà*, and her prior mean is 0.2 and her prior standard deviation is 0.08. Sam is an *Erasmus* student from Scotland, so he is not aware of the local feeling for or against the proposed casino. He decides to use a uniform prior. The two students take a random sample of $n=100$ *Baix Llobregat* residents and ask for their views of the casino. Out of the random sample $y=26$ said they support building a casino in *Baix Llobregat*.

- a) Draw these two prior distributions in the same graph.
- b) Draw the two posterior distributions in the same graph.
- c) Repeat a) and b) assuming that the data were: $n=1000$ and $y=260$, and think about the difference in using different sample sizes.