Obligatory assignment 1 MSA101/MVE197 autumn 2018

Petter Mostad

September 6, 2019

1. Consider the following probability densities defined for any parameter $\theta > 0$ by the probability density function

$$\pi(x \mid \theta) \propto_x x^2 \exp(-\theta x^3) \qquad \text{for } x \in [0, \infty).$$

   (a) Writing $\pi(x \mid \theta) = C(\theta)x^2 \exp(-\theta x^3)$, determine the function $C(\theta)$.

   (b) Using the probability density function found in (a), guess at a conjugate family of priors. Prove that this family is conjugate, and compute the formula for the posterior density for θ given a prior density from the family.

   (c) Compute the formula for the prior predictive density when using a prior from the conjugate class you found above.

2. A way to differentiate between different types of glass, for example to determine if two glass fragments come from the same glass source or not, is to measure the *Refreacive Index* (RI) of the glass. The figure below shows a histogram of values for RI for different types of glass. The values are taken from a forensic data base. As we can see, the RIs do not seem to follow a normal distribution. A possibility may then be to use a mixture of normal densities.

   We will now use the following example of such a mixture density: Using θ to denote RI, we use as a model

$$\pi(\theta) = \sum_{i=1}^{3} \gamma_i \operatorname{Normal}(\theta; \mu_i, \sigma_i^2)$$

   with the values

i	γ_i	μ_i	σ_i
1	0.33	1.5163	0.001
2	0.57	1.5197	0.001
3	0.10	1.5203	0.005

The measurement method for RI we use has a standard error of 0.001; in other words for a measurement x we have

$$\pi(x \mid \theta) = \text{Normal}(x; \theta, 0.001^2)$$

(a) Construct an R function that takes as input a vector of theta values, a vector of expectations of normal distributions, a vector of corresponding variances of those normal distributions, and a vector of probability weights. It should output the density values for the corresponding mixture distribution. Use this function to plot the density above for RI over a reasonable interval.

(b) Compute the formula for the prior predictive density for a measurement x. Plot this density.

(c) Find the formula for the posterior of θ given a value for x. Compute and make a plot of this posterior density when $x = 1.52083$. Also, estimate a 95% credibility interval for RI when x has this value.

(d) Given a value for x, compute the formula for the posterior predictive density of a new measurement x_{NEW} on the same piece of glass. Plot the density assuming $x = 1.52083$.

(e) Glass can be an important type of physical evidence in criminal cases. For example, if microscopic glass fragments found on the jacket of a suspect are sufficiently similar to the glass in a broken window at the scene of a burglary, it may count as important evidence for the prosecution. Specifically, specifying the two hypotheses

H_p :	The glass from the suspect and from the crime scene are from the same source
H_d :	The glass from the suspect and from the crime scene are from independent sources

the likelihood ratio

$$LR = \frac{\pi(\text{data} \mid H_p)}{\pi(\text{data} \mid H_d)}$$

is used to measure the strength of the evidence. If x_s and x_c are single RI measurements made on the glass from the suspect and the crime scene, respectively, obtain the formula for the likelihood ratio LR.

(f) Assuming $x_s = 1.52083$, compute and make a plot of LR as a function of possible values of x_c. Make a comment on how the plot should be interpreted.

(g) It is customary to make repeated measurements to improve accuracy: Assume n measurements were made on the glass from the suspect and m measurements were made on the glass from the crime scene. Describe how your analysis and results would change.

(h) OPTIONAL: The particular model for θ we have used in this exercise is obviously just a toy example. Discuss possibilities for how one can go about estimating such a model from a database like the one illustrated in the histogram below.

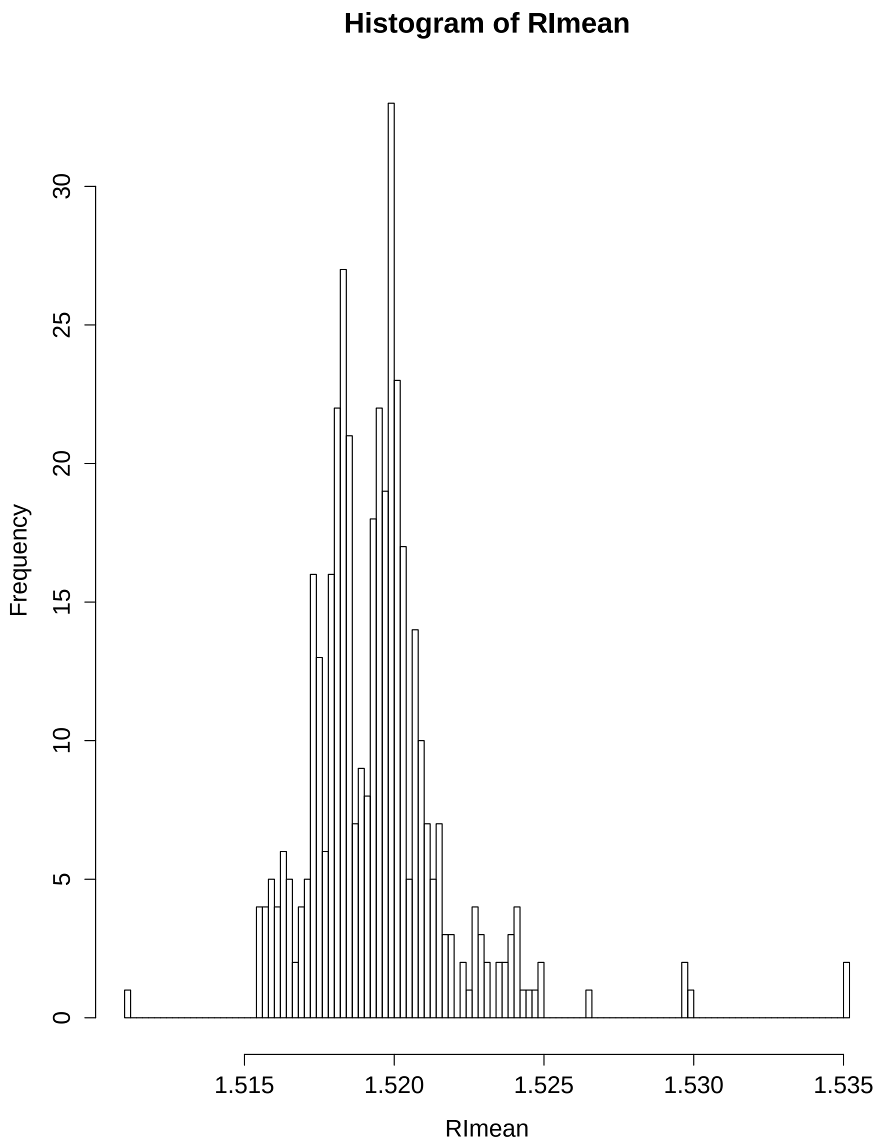

Figure 1: For each glass fragment in a database, the mean of repeated measurements of RI has been recorded. The results are shown in the histogram above.