Obligatory assignment 2 MSA101/MVE187, autumn 2020

Petter Mostad

September 23, 2020

In the answers to the questions below, you should include your R code so that I can check that you have done computations correctly.

1. Anton is building a windmill to produce electricity for his house. The wind speed at his house of course varies, but can be modelled as Gamma distributed with parmeters $\alpha = 1.5$ and $\beta = 0.1$, where the Gamma distibution has density

$$\text{Gamma}(v; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} v^{\alpha-1} \exp(-\beta v)$$

Anton has found that the output from his windmill at a given wind speed v is given by the function

$$P(v) = \begin{cases} 0 & v < 13 \\ \cos(\frac{\pi}{10}v - \frac{3\pi}{10}) + 1 & 13 \leq v < 23 \\ (1031 + 46v - v^2)/780 & 23 \leq v < 30 \\ 0 & 30 \leq v \end{cases}$$

(a) At any given time, the turbine may generate some power or no power. Calculate the probability that it delivers some power.

(b) Compute by simulation the expected value, together with an approximate 95% confidence interval for the simulation accuracy, for the amount of power generated by the wind turbine. Show how your results depend on the sample size you choose.

(c) Re-compute the expectation in (b) using discretization.

(d) Re-compute the expectation in (b) using the R function `integrate`.

(e) To improve on the accuracy in (b) you would like to use importance sampling. Make a plot that is relevant for choosing a proposal density (instrumental density), and make such a choice.

(f) Now repeat (b) using importance sampling with your chosen proposal density. Compare the results with those of (b).

2. In a study of the number of death notices in the London Times on each day in the interval 1910-1912 for women aged 80 and over, it was noticed that the number of deaths did not follow a Poisson distribution as was expected. The counts Y_i of days with i death notices is given in the table below:

Daily death notices (i)	0	1	2	3	4	5	6	7	8	9
Day count (Y_i)	162	267	271	185	111	61	27	8	3	1

(a) Assuming that the deaths each day follow a Poisson distribution, write down the likelihood function for the Poisson density λ in terms of the observations Y_i above. As a step on the way, it may be helpful to consider what would have happened if the data had contained the death count for each specific date in the period 1910-1912.

(b) Compute the maximum likelihood estimate for λ using the likelihood from (a). Plot the predicted counts using this λ together with the actual counts.

(c) As a small extension of the model above, we instead assume that, each day, there is a choice between two Poisson distributions, one with intensity λ_1 and the other with intensity λ_2. The probability for choosing λ_1 is p and the probability for choosing λ_2 is $1-p$. We now have a model with the three parameters, λ_1, λ_2, p. In order to work with this model, we add also variables $Z_0, Z_1, \dots, Z_9$, where Z_i is the (unobserved) count of days where λ_1 is used and where i deaths are observed. Thus $0 \leq Z_i \leq Y_i$ for $i = 0, \dots, 9$. Write down (a function proportional to) the likelihood for $Y_0, \dots, Y_9, Z_0, \dots, Z_9$ given the parameters λ_1, λ_2, p.

(d) We now assume we have a prior for the parameters p, λ_1, λ_2 that is a product of a uniform distribution on $[0,1]$ for p, a prior $\lambda_1 \sim \text{Gamma}(1,1)$ for λ_1, and a prior $\lambda_2 \sim \text{Gamma}(1,1)$ for λ_2. Write down the posterior[1] for p given all the other observations and variables.

(e) Find the posteriors for λ_1 given all the other variables and pararmeters, and similar for λ_2.

(f) Prove that the distribution of Z_i given all parameters, all $Y_0, \dots, Y_9$, and all other Z_i variables, is Binomial. Find the parameters of this distribution.

(g) Implemement a Gibbs sampler for the extended model above. Use the implementation to produce an approximate sample for the posterior distribution of the parameters p, λ_1, λ_2 given the data $y_0, \dots, y_9$. Visualize your results. Estimate the expected values of each of the parameters in this posterior.

[1] The expressions "write down the posterior" or "find the posterior" mean that you should write down the formula for a function that is proportional to the posterior density, and, in the cases where the posterior is a density in a known parametric family, identify this family and the correponding values of the parameters.