Obligatory assignment 2 MSA101/MVE187, autumn 2018

Petter Mostad

September 19, 2019

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

1. The amount of electrical power $P(v)$ produced by a wind turbine for a given wind speed v is described by the *power curve* of the turbine. For a 2 MW turbine we approximate the power curve with

$$P(v) = \begin{cases} 0 & v < 4 \\ \cos(v\frac{\pi}{11} + \frac{7\pi}{11}) + 1 & 4 \leq v < 15 \\ \frac{7}{4} + v\frac{1}{30} - v^2\frac{1}{900} & 15 \leq v < 25 \\ 0 & 25 \leq v \end{cases}$$

The wind speed at the location of the turbine will vary. Based on meteorological records for this location, we model the wind speed as gamma distributed with parameters $\alpha = 0.8$ and $\beta = 1/6$, where the gamma distibution has density

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

(a) Calculate the probability that the turbine delivers power at any given time, $\Pr(P(v) > 0)$.

(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) Compute the result of (b), now using numerical integration, either by using discretization or the R function `integrate`.

(d) 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.

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

(f) Use importance sampling to estimate the variance of the power production, $\mathrm{Var}(P(v))$: Start by writing down the integral you woud like to compute.

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. Compute the maximum likelihood estimate for λ. Plot the predicted counts using this λ together with the actual counts.

(b) 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 $Z_0, Z_1, \ldots, Z_9$, 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 $i = 0, \ldots, 9$. Write down (a function proportional to) the likelihood for $Y_0, \ldots, Y_9, Z_0, \ldots, Z_9$ given the parameters λ_1, λ_2, p.

(c) 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 $\lambda_1 \sim \mathrm{Gamma}(1,1)$ for λ_1, and a prior $\lambda_2 \sim \mathrm{Gamma}(1,1)$ for λ_2. Write down the posterior[1] for p given all the other observations and variables.

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

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

(f) Implmement 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 $y_0, \ldots, y_9$.

[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 distribution in a known parametric family, identify this family and the correponding values of the parameters.

Visualize your results. Estimate the expected values of each of the parameters in this posterior.

(g) (OPTIONAL) Use the expected values found above to simulate new data, i.e., new deaths, for the same number of days as in the real data. Compare the predictions you get to the predictions obtained with the model in (a), and to the real data.