Computer class 4 exercises

Richard Wilkinson

22 April 2016

Question 1

In the previous computer class, we considered a simple Bayesian inference problem where we wanted to learn about an unknown parameter p, to which we assigned a U[0,1] distribution. We were given data $x_1, \ldots, x_{10}$ which were independent $Bin(20, p)$ random variables, and we were told that $\sum_{i=1}^{10} x_i = 50$. Recall that the likelihood times the prior for this problem was proportional to

$$f_1(p) = p^{\sum x_i}(1-p)^{200-\sum_i x_i}.$$

- Using a $U[0, 1]$ distribution as the importance distribution, use importance sampling to generate a weighted sample
$$\{p_i, w_i\}_{i=1}^N$$
of particles and weights that approximates the posterior distribution.
- Calculate the posterior mean of p. Note that we can approximate any integral by a weighted sum. So for example,
$$E(p|x) = \int p\pi(p|x)dp \approx \frac{\sum w_i p_i}{\sum w_i}.$$
Alternatively, we can use the weighted version of statistical estimators in the Hmisc library, for example, `wtd.mean`. You may need to install Hmisc the first time you use it (`install.packages('Hmisc')`)
- We can resample the particles to get an unweighted sample of particles. To do this, first convert the weights into probabilities,
$$W_i = \frac{w_i}{\sum w_i}$$
and then sample from $\{p_i\}_{i=1}^N$ with replacement, picking particle i with probability W_i. Calculate the number of unique particles in your unweighted sample.
- Use the resampled particles to plot a histogram of the posterior distribution.
- Repeat the steps above using a $Beta(10, 30)$ distribution as the importance distribution.
- The variance of the importance weights is a useful measure of how successful a given importance distribution will be - we want the variance to be as small as possible. A related quantity that is often used is the effective sample size (ESS)
$$ESS = \frac{1}{\sum W_i^2}$$
where the W_i are the normalised weights ($\sum W_i = 1$). If all the weights are the same (i.e. they have zero variance), then the ESS = N, i.e. the sample is as effective as a sample of N unweighted particles. Whereas in the worst case where all the weights are 0 except for one which has $W = 1$, then the ESS=1, i.e., the sample is equivalent to a single unweighted sample. Calculate the ESS for your two importance distributions to see which gives a better sample.
- What choice of importance distribution would give the best possible ESS?

Question 2

This problem is described in the notes. Here we will work through the details.

Patients suffering from leukaemia are given a drug, 6-mercaptopurine (6-MP), and the number of days x_i until freedom from symptoms is recorded for patient i:

$$6^*, 6, 6, 6, 7, 9^*, 10^*, 10, 11^*, 13, 16, 17^*, 19^*, 20^*, 22, 23, 25^*, 32^*, 32^*, 34^*, 35^*,$$

where a * denotes censored observation. The time x to the event of interest follows a *Weibull* distribution:

$$f(x|\alpha,\beta) = \alpha\beta(\beta x)^{\alpha-1}\exp\{-(\beta x)^\alpha\}$$

for $x > 0$. For censored observations, we can show that

$$P(x > t|\alpha,\beta) = \exp\{-(\beta t)^\alpha\}.$$

We want to estimate the posterior mean of θ, and the posterior 5th and 95th percentiles.

Define d to be the number of uncensored observations and $\sum_u \log x_i$ to be the sum of logs of all uncensored observations. If we use the following prior distributions for α and β

$$f(\alpha) = 0.001\exp(-0.001\alpha), \qquad f(\beta) = 0.001\exp(-0.001\beta).$$

then we can show that the log of the posterior distribution is proportional to

$$\log f(\theta|x) \propto h(\theta) := d\log\alpha + \alpha d\log\beta + (\alpha-1)\sum_u \log x_i - \beta^\alpha \sum_{i=1}^n x_i^\alpha - 0.001\alpha - 0.001\beta + K,$$

where $\theta = (\alpha,\beta)^T$.

- Use importance sampling to estimate the posterior mean of α and β, using an $Exp(1)$ distribution for both parameters. Does this work well?

We will now use the Laplace approximation to design a better choice of the proposal density g.

- Obtain the posterior mode of θ, i.e., maximise $h(\theta)$ defined above. You can do this in R by writing a function to evaluate h and then using the `optim` command. Note that optim does minimization by default.
- Find the Hessian (matrix of second derivatives) of $h(\theta)$ at $\theta = m$, either by deriving it analytically, or estimating it using numerical differentiation (the `hessian` command in the numDeriv package works well),

$$M = \begin{pmatrix} \frac{\partial^2}{\partial\alpha^2}h(\theta) & \frac{\partial^2}{\partial\alpha\partial\beta}h(\theta) \\ \frac{\partial^2}{\partial\alpha\partial\beta}h(\theta) & \frac{\partial^2}{\partial\beta^2}h(\theta) \end{pmatrix}.$$

- Use an importance sampling algorithm to estimate the posterior mean and 5th and 95th percentiles of this distribution. Use a multivariate Gaussian distribution as your proposal, with mean m and covariance matrix $V = -M$. To simulate from a multivariate normal, you can either use the Cholesky decomposition of V, or use the mvtnorm package in R (you may need to install it using `install.packages('mvtnorm')` the first time you use this). Note that you will need to use `wtd.quantile` or resample the particles and use `quantile` to get the quantiles.
- Resample the particles to get an unweighted sample, and plot the posterior distribution of α and β.