MSc in Statistics and Operational Research
Bayesian Inference Project 2: Question 2

Modelling Winning Times of Scottish Hill Races

Analyse the data in the file "datahills.r". This file contains record times (in minutes) for 35 Scottish hill races, together with the length of the race (in miles) and elevation (in feet). The objective is to model the winning times in terms of distances and climbs of the races.

PART 1: A BAYESIAN MODEL FOR WINNING TIMES

A good starting point is Naismith's rule, which is used to calculate the length of time a hillwalk should take. The form of such a model is:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \epsilon_i, \qquad (1)$$

where y_i, x_{i1}, x_{i2} respectively denote time, distance and climb for the i^{th} race, and $\boldsymbol{\beta} = (\beta_1, \beta_2)$ is a parameter. Allowing for a Normal error distribution, we obtain that the observations $y_1, \ldots, y_n$ are independent, distributed according to

$$y_i|\boldsymbol{\beta}, \omega \sim N(\mathbf{x}_i^T \boldsymbol{\beta}, 1/\omega). \qquad (2)$$

To construct a Bayesian model, you also need to specify a prior distribution for the parameters $\boldsymbol{\beta}$ and ω. Take these parameters independent a priori with

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \sim N_2\left(\boldsymbol{\mu}_0, \mathbf{C}_0 = \text{Diag}\left(\frac{1}{\kappa_{01}}, \frac{1}{\kappa_{02}}\right)\right) \quad \text{and} \quad \omega \sim \text{Gamma}(\alpha_0, \lambda_0). \qquad (3)$$

In other words, the prior for $\boldsymbol{\beta}$ is bi-variate Normal with mean vector $\boldsymbol{\mu}_0 \in \mathbb{R}^2$ and covariance matrix $\mathbf{C}_0$. Since the covariance matrix is diagonal, β_1 and β_2 are independent a priori. ω is assigned a Gamma prior distribution.

1.1. Write down, up to a proportionality constant, the joint posterior density of $(\boldsymbol{\beta}, \omega)$.

1.2. Write down, up to a proportionality constant, the posterior densities of $\boldsymbol{\beta}$ and ω, conditioning on the other parameter (that is, $f(\boldsymbol{\beta}|y_1, \ldots, y_n, \omega)$ and $f(\omega|y_1, \ldots, y_n, \boldsymbol{\beta})$).

1.3. Using the expressions that you wrote down for the conditional posterior densities in question 1.2, show that the conditional posterior distributions are

$$\omega|\mathbf{y}, \boldsymbol{\beta} \sim \text{Gamma}\left(\alpha_0 + \frac{n}{2}, \lambda_0 + \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2}\right)$$

and

$$\boldsymbol{\beta}|\mathbf{y},\omega \sim N_2(\boldsymbol{\mu}_1, \mathbf{C}_1) \quad \text{with} \quad \mathbf{C}_1 = (\mathbf{C}_0^{-1} + \omega \mathbf{X}^T\mathbf{X})^{-1} \text{ and } \boldsymbol{\mu}_1 = \mathbf{C}_1(\mathbf{C}_0^{-1}\boldsymbol{\mu}_0 + \omega \mathbf{X}^T\mathbf{y}),$$

where

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \cdot \\ \cdot \\ \mathbf{x}_n^T \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{pmatrix}.$$

1.4. Using the conditional posterior distributions in the previous question, write a Gibbs sampling algorithm to compute the posterior distribution of $(\boldsymbol{\beta}, \omega)$ via simulation.

1.5. Program in R the Gibbs sampling algorithm developed in question 1.4. Choose hyperparameters for the prior distribution in such a way that the prior does not contain a lot of information (ie. choose moderately large prior variances for the parameters). Also, in the absence of prior information to the contrary, it seems natural to choose $\mu_{01} = \mu_{02}$ and $\kappa_{01} = \kappa_{02}$ in the prior distribution for $\boldsymbol{\beta}$.

Run the program for the hill races dataset, choosing a suitable burn-in period and a suitable number of draws to be used for inference. Display histograms of the marginal posterior distributions of β_1, β_2 and ω, as well as the following numerical summaries of their posterior distributions: minimun, 1st quartile, median, mean, 3rd quartile, maximum. Display a scatterplot of the joint posterior distribution of (β_1, β_2). What are your conclusions in terms of the effect of distance and climb on the winning time of races?

1.6. (Optional) Some of the observations may be regarded as outliers, *i.e.* not fitted well by the assumed model. A way to try to detect outliers is to see whether any of the observations y_i is particularly far from the value predicted by the model. For a race of distance x_{i1} and climb x_{i2}, a natural prediction of its winning time would be $\hat{y}_i = E[\beta_1|\mathbf{y}]x_{i1} + E[\beta_2|\mathbf{y}]x_{i2}$. For each of the races, compute $y_i - \hat{y}_i$ and plot the resulting numbers. Are there any races that appear out of line?

PART 2: A SAMPLING MODEL RESISTANT TO OUTLIERS

When there are outliers suspected in the data, the sensible thing to do is to take a closer look at the data and try and understand where the lack of fit may be coming from. There are several possibilities: e.g. the outliers may be due to errors in recording the data, or it may be that there were certain features corresponding to those observations that our model does not take into account. The latter may be solved e.g. by adding relevant explanatory variables to the model, or by changing other aspects of the model.

Now suppose that none of the above applies to your model: that is, there are no recording errors in the data that you know of, you have no additional explanatory variables that you could possibly use, and you can not think of a more suitable model specification.

In that case, a partial solution is to use a sampling distribution that is more robust to outliers than the Normal distribution used in (2). A distribution with thicker tails than the Normal distribution allows for larger departures from the specified mean in (1) and, therefore, outliers are likely to have less impact on the resulting posterior inference on β_1 and β_2. That is what we mean by a sampling distribution that is more resistant to outliers than the Normal distribution.

A distribution with this property is the Cauchy distribution. Using a Cauchy distribution leads to the following sampling density for observation i:

$$f(y_i|\boldsymbol{\beta},\omega) = \frac{\omega^{1/2}}{\pi}\frac{1}{1+\omega(y_i-\mathbf{x}_i^T\boldsymbol{\beta})^2}, \quad i=1,\ldots,n, \quad \text{independent.} \tag{4}$$

Consider the same prior distributions for $\boldsymbol{\beta}$ and ω used in Part 1.

2.1. Write down, up to a proportionality constant, the joint posterior density of $(\boldsymbol{\beta},\omega)$.

2.2. Write down, up to a proportionality constant, the conditional posterior densities of $\boldsymbol{\beta}$ and ω (that is, $f(\boldsymbol{\beta}|\mathbf{y},\omega)$ and $f(\omega|\mathbf{y},\boldsymbol{\beta})$). Is our new model amenable to Gibbs sampling? Explain why or why not.

2.3. Now consider the following idea: The Cauchy distribution in (4) can also be interpreted as a Normal distribution with an unknown (random) precision parameter. More specifically, assuming

$$y_i|\boldsymbol{\beta},\omega,z_i \sim N\left(\mathbf{x}_i^T\boldsymbol{\beta}, \frac{1}{\omega z_i}\right) \tag{5}$$

$$z_i|\boldsymbol{\beta},\omega \sim \text{Gamma}\left(\frac{1}{2},\frac{1}{2}\right) \tag{6}$$

is, in fact, equivalent to assuming the sampling model in (4). Thus, we can augment the parameters with the variables $z_1,\ldots,z_n$ (note that there is a variable z_i per observation) to obtain the posterior density

$$\begin{aligned} f\boldsymbol{\beta},\omega,z_1,\ldots,z_n|\mathbf{y}) \propto & \left\{\prod_{i=1}^{n}(\omega z_i)^{1/2}\exp\left(-\frac{\omega z_i}{2}(y_i-\mathbf{x}_i^T\boldsymbol{\beta})^2\right)z_i^{-1/2}e^{-\frac{1}{2}z_i}I[z_i>0]\right\} \\ & \times \exp\left(-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)^T\mathbf{C}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)\right)\omega^{\alpha_0-1}e^{-\lambda_0\omega}I[\omega>0] \end{aligned}$$

Find the conditional posterior distributions of $\boldsymbol{\beta}$, ω and z_i, $i=1,\ldots,n$ and, thus, write down a Gibbs sampling algorithm to compute the joint posterior distribution of $(\boldsymbol{\beta},\omega,z_1,\ldots,z_n)$.

2.4. Program the Gibbs sampler in the previous question using R .

2.5. How does the inference on β_1 and β_2 compare with that obtained in Part 1 where you used a Normal sampling model?

2.6 (Optional) Do the same as is question 1.6 and comment on the new results. Does the Cauchy model appear to be more resistant to outliers than the Normal model?

PART 3: METROPOLIS–HASTINGS FOR CAUCHY MODEL

In this part, consider again the same Cauchy sampling model as in Part 2 and the same prior distribution on $(\boldsymbol{\beta}, \omega)$ used throughout the project.

3.1. Instead of using data augmentation (like you did in Part 2), develop a Metropolis-Hastings algorithm that can handle the sampling model in (4) directly. Now consider β_1 and β_2 separately, so you need to start by finding the conditional posterior densities of β_1, β_2 and ω.

3.2. Code the Metropolis-Hastings algorithm in R .

3.3. Compare the posterior results for β_1 and β_2 with those obtained from the data augmentation algorithm in Part 2. Since both algorithms are computing exactly the same thing, your results should be virtually identical, are they?