Computer Intensive Methods Project 2

Nicolo Campolongo, Rendani Mbuvha

May 2016

1 Bayesian analysis of coal mine disasters - constructing a complex MCMC algorithm

In this problem we have data regarding British coal mine disasters per year, over the period $1851 - 1962$. In particular, we have different parameters to model the distribution of disasters over time in a continuous range: number of disasters each year $\boldsymbol{\tau} = (\tau_1, ..., \tau_n)$, time intervals $\mathbf{t} = (t_1, .., t_{d+1})$, disaster intensities in each interval $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_d)$ that have well defined probability distributions. Our goal is to split the whole period in time intervals of different size and, using Bayesian analysis, to estimate when there might have been abrupt changes in the rate of disasters per year.

Marginal posteriors

Using the Bayes theorem, we can get the expression of the joint distribution $f(\boldsymbol{\lambda}, \mathbf{t}, \theta|\boldsymbol{\tau})$ up to a normalizing constant:

$$\begin{aligned} f(\boldsymbol{\lambda}, \mathbf{t}, \theta|\boldsymbol{\tau}) &\propto f(\boldsymbol{\tau}|\boldsymbol{\lambda}, \mathbf{t}, \theta) f(\boldsymbol{\lambda}, \mathbf{t}, \theta) \\ &= f(\boldsymbol{\tau}|\boldsymbol{\lambda}, \mathbf{t}) f(\mathbf{t}) f(\boldsymbol{\lambda}|\theta) f(\theta) \end{aligned}$$

Then, we have the full complex expression:

$$f(\boldsymbol{\lambda}, \mathbf{t}, \theta|\boldsymbol{\tau}) \propto \exp\Big(-\sum_{i=1}^{d} \lambda_i (t_{i+1} - t_i) \Big) \prod_{i=1}^{d} \lambda_i^{n_i^{(\boldsymbol{\tau})}} \times \Big(\prod_{i=1}^{d} (t_{i+1} - t_i) \Big) \times \Big(\prod_{i=1}^{d} \Gamma_{\lambda_i}(2, \theta) \Big) \times \Big(\Gamma_{\theta}(2, \vartheta) \Big)$$

We are now going to illustrate how to get the different marginal distributions using the following expression of the Γ distribution:

$$\Gamma_x(\alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha - 1} e^{-\beta x}$$

Theta

$$
\begin{aligned}
f(\theta|\boldsymbol{\lambda},\mathbf{t},\boldsymbol{\tau}) \propto f(\boldsymbol{\lambda},\mathbf{t},\theta|\boldsymbol{\tau}) &\propto \Big(\prod_{i=1}^{d}\frac{\theta^2\lambda_i e^{\lambda_i\theta}}{\Gamma(2)}\Big)\times\frac{\vartheta^2\theta e^{-\vartheta\theta}}{\Gamma(2)}\\
&= \frac{\theta^{2d+1}}{(\Gamma(2))^{2d+1}}\Big(\prod_{i=1}^{d}\lambda_i e^{\lambda_i\theta}\Big)\vartheta^2 e^{-\vartheta\theta}\\
&\propto \theta^{2d+1}\Big(\prod_{i=1}^{d}\lambda_i\Big)\Big(\prod_{i=1}^{d}e^{-\lambda_i\theta}\Big)e^{-\vartheta\theta}\\
&\propto \theta^{2d+1}\exp\Big(-\theta\sum_{i=1}^{d}\lambda_i\Big)\exp(-\vartheta\theta)\\
&= \theta^{2d+1}\exp\Big(-\theta\Big(\vartheta+\sum_{i=1}^{d}\lambda_i\Big)\Big)\\
&\propto \Gamma_\theta\Big(2d+2,\vartheta+\sum_{i=1}^{d}\lambda_i\Big)
\end{aligned}
$$

$\boldsymbol{\lambda}$ intensities

$$
\begin{aligned}
f(\boldsymbol{\lambda}|\theta,\boldsymbol{\tau},\mathbf{t}) \propto f(\boldsymbol{\lambda},\mathbf{t},\theta|\boldsymbol{\tau}) &\propto \exp\Big(-\sum_{i=1}^{d}\lambda_i(t_{i+1}-t_i)\Big)\prod_{i=1}^{d}\lambda_i^{n_i(\tau)}\times\Big(\prod_{i=1}^{d}\Gamma_{\lambda_i}(2,\theta)\Big)\\
&= \prod_{i=1}^{d}\exp\Big(-\lambda_i(t_{i+1}-t_i)\Big)\lambda_i^{n_i(\tau)}\times\frac{\theta^2\lambda_i}{\Gamma(2)}\exp\Big(-\lambda_i\theta\Big)\\
&\propto \prod_{i=1}^{d}\lambda_i^{n_i(\tau)}\exp\Big(-\lambda_i(t_{i+1}-t_i)-\lambda_i\theta\Big)\\
&\propto \prod_{i=1}^{d}\Gamma_{\lambda_i}\Big(n_i(\tau)+2,t_{i+1}-t_i+\theta\Big)
\end{aligned}
$$

Breakpoints

$$
\begin{aligned}
f(\mathbf{t}|\theta,\boldsymbol{\tau},\boldsymbol{\lambda}) \propto f(\boldsymbol{\lambda},\mathbf{t},\theta|\boldsymbol{\tau}) &\propto \exp\Big(-\sum_{i=1}^{d}\lambda_i(t_{i+1}-t_i)\Big)\prod_{i=1}^{d}\lambda_i^{n_i(\tau)}\times\Big(\prod_{i=1}^{d}(t_{i+1}-t_i)\Big)\\
&\propto \prod_{i=1}^{d}\lambda_i^{n_i(\tau)}\exp\Big(\lambda_i(t_{i+1}-t_i)\Big)(t_{i+1}-t_i)
\end{aligned}
$$

Note: This last distribution is only valid in the case $t_1 < t_2 < ... < t_d < t_{d+1}$, otherwise it is 0.

Hybrid MCMC algorithm

In order to draw samples from the posterior $f(\theta, \boldsymbol{\lambda}, \mathbf{t}|\boldsymbol{\tau})$ we construct a hybrid MCMC algorithm. Indeed, while it is possible to sample directly from the conditional distributions $f(\theta|\boldsymbol{\lambda}, \mathbf{t}, \boldsymbol{\tau})$ and $f(\boldsymbol{\lambda}|\theta, \boldsymbol{\tau}, \mathbf{t})$ using Gibbs sampling, $f(\mathbf{t}|\theta, \boldsymbol{\tau}, \boldsymbol{\lambda})$ is more complex and requires a Metropolis-Hasting step. From the theory about MCMC, we know that the resulting chain still satisfies global balance and is thus a valid MCMC sampler.

Gibbs sampling

The Gibbs sampling algorithm is Markov Chain Monte Carlo algorithm for obtaining a sequence of observations, which in turn can be approximated from a multivariate probability distribution. We assume that we want to sample from a multivariate distribution f on X, where X is a space that can be divided into m blocks, i.e., $\mathbf{x} = (x1, ..., x_m) \in X$, where each block may be vector-valued itself. We denote by x^k the k-th component and by x^{-k} the remaining components and assume that is easy to simulate from the conditional distribution of x^k given all other components: $f_k(x^k|x^{-k})$ for all $k = 1, .., m$. We then simulate a sequence X_k, forming a Markov Chain on X, with the following:

$$\text{given } X_k, \text{ draw } X_{k+1}^n \sim f_n(x^n|x^{-n})$$

MH step

In the following we illustrate how to draw from $f(\mathbf{t}|\theta, \boldsymbol{\tau}, \boldsymbol{\lambda})$ using a MH step. Given d breakpoints, we draw a candidate breakpoint t_i using a proposal kernel and, after updating the number of disasters according to the new time intervals we calculate α:

for $i = 2 \leftarrow d$ **do**
  $\mathbf{t}^* \leftarrow \mathbf{t}_k$;
  draw $t_i^* \sim r(t_i^*|t_i)$;
  $\mathbf{t}^*(i) \leftarrow t_i^*$;
  if $t_i^* > \mathbf{t}_k(i-1)$ $t_i^* < \mathbf{t}_k(i+1)$ **then**
    update $n_i(\boldsymbol{\tau})$;
    set $\alpha(\mathbf{t}_k, \mathbf{t}^*) = 1 \wedge \frac{f(\mathbf{t}^*)}{f(\mathbf{t}_k)}$;
    draw $U \sim \mathrm{U}(0, 1)$;
    if $U \leq \alpha$ **then**
      $\mathbf{t}_{k+1}(i) = t_i^*$;
    else
      $\mathbf{t}_{k+1}(i) = t_i$;
    end
    $\mathbf{t}_{k+1}(i) = t_i$;
  else
end

In particular, in the MH step we use a *random walk proposal*. For each breakpoint t_i we generate a candidate t_i^* according to:

$$t_i^* = t_i + \epsilon, \text{ with } \epsilon \sim \mathrm{Unif}(-R, R), \; R = \rho(t_{i+1} - t_{i-1})$$

With this symmetric proposal we get the following:

$$\alpha(t_i, t_i^*) = 1 \wedge \frac{f(t_i^*) r(t_i|t_i^*)}{f(t_i) r(t_i^*|t_i)}$$

Since $r(t_i|t_i^*) = r(t_i^*|t_i)$, we get:

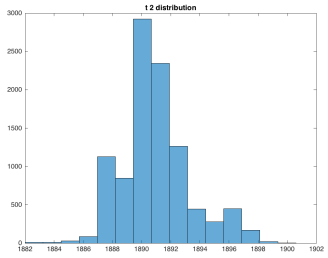
$$\alpha(t_i, t_i^*) = 1 \wedge \frac{f(t_i^*)}{f(t_i)}$$

Results

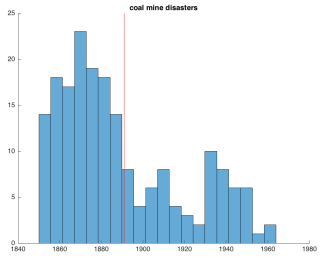
When defining the initial breakpoints, we choose them in order to define equally large time intervals. Furthermore, when estimating 1 and 2 breakpoints we draw 10000 samples using a *burn-in* of 1000, while with 3 and 4 breakpoints we use 15000 samples with no well defined *burn-in*. We notice in general that the choice of the hyperparameter ϑ does not really influence the posterior distribution, thus we always use the value $\vartheta = 1$. On the other hand, the value of ρ should be accurately set in order to get a reasonable acceptance rate. Results are described below.

One breakpoint

With one breakpoint, using $\rho = 0.1$ the average value of t_2 is 1891, as illustrated below:

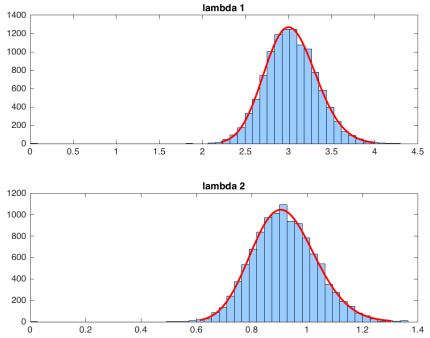


(a) t_2 sampled values (after *burn-in*)

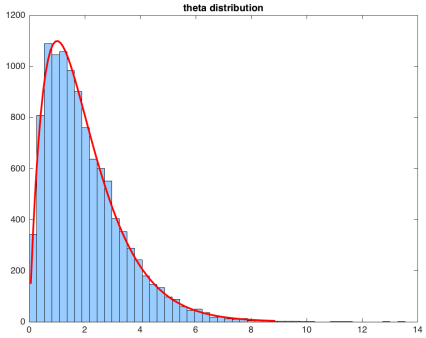


(b) Coal mine disasters with breakpoint at year 1891.

Also we get the following λ intensities and θ:



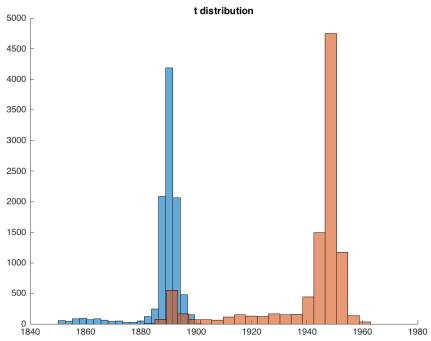
(c) λ_1 shows a mean value of ≈ 3 disasters/year, while for λ_2 we get ≈ 0.9 disasters/year.



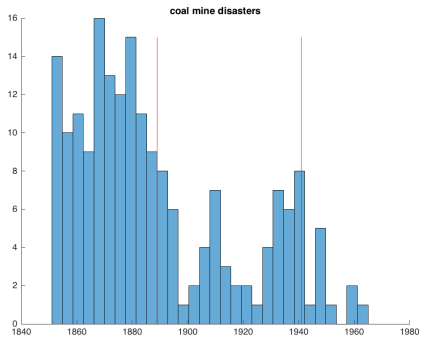
(d) θ distribution

Two breakpoints

With 2 breakpoints we still use the value of $\rho = 0.1$. However, having a "large" value ρ may cause two breakpoints to converge to the same year. In that case there would be no further updates since $t_i = t_i + 1$ and $f(\mathbf{t}^*) = 0$, so every candidate would be rejected. However, by running the algorithm, the 2 averaged breakpoints always fall in the intervals $[1889, .., 1892]$ and $[1941, .., 1946]$. Results for one simulation are shown below:



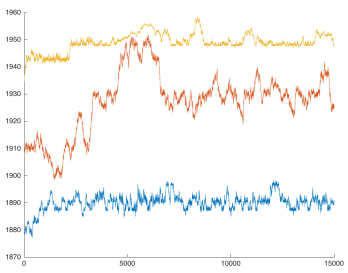
(e) Distribution of $\mathbf{t}$ samples

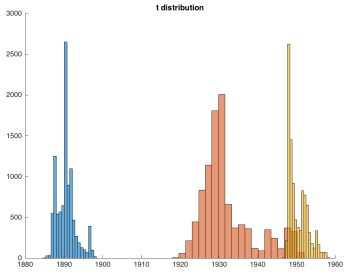


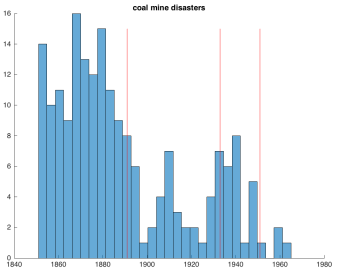
(f) The 2 breakpoints: $t_2 = 1889$ and $t_3 = 1941$.

Three breakpoints

When using 3 breakpoints, the value of ρ should be definitely reduced to make the algorithm work. Thus, we set $\rho = 0.03$. Moreover, in this case it is more difficult to identify different breakpoints. We ran various simulations and get quite different result. However, the most recurrent pattern is the one setting the first breakpoint in the interval $[1889, ..., 1892]$, while the second and the third identify the region between 1929 and 1946.



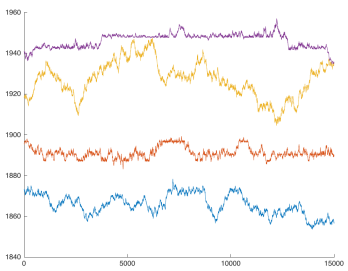


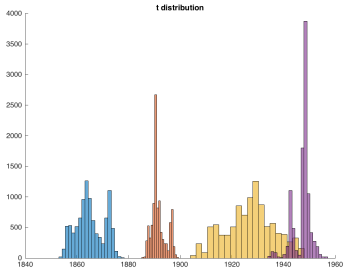


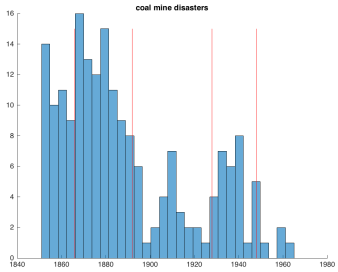
As can be noticed in the left-figure above, it takes more time for the chain to converge to the stationary distribution. Thus, we choose to draw 15000 samples and a *burn-in* of 7000. Results are still reasonable, showing the following tree breakpoints: $t_2 = 1889$, $t_3 = 1930$ and $t_4 = 1946$.

Four breakpoints

With 4 breakpoints we reduce again the value of ρ, setting it to 0.01.Wwe draw 15000 samples and set the *burn-in* to 5000.







The results in this case are $t_2 = 1866$, $t_3 = 1892$, $t_4 = 1928$ and $t_5 = 1948$. In general, when adding more breakpoints it is more difficult to get a good estimate.

Parametric bootstrap for the 100-year Atlantic wave

(a) Given the *Gumbel distribution* with the following distribution function:

$$F(x;\mu,\beta) = \exp\left(-\exp\left(-\frac{x-\mu}{\beta}\right)\right) \tag{1}$$

We proceed to finding the inverse $F^{-1}(u;\mu,\beta)$ by solving for the input variable x.

Now let, $U = F(x;\mu,\beta)$

$$\begin{aligned} U &= \exp\left(-\exp\left(-\frac{x-\mu}{\beta}\right)\right) \\ \log(U) &= -\exp\left(-\frac{x-\mu}{\beta}\right) \\ \log(\log(\frac{1}{U})) &= -\frac{x-\mu}{\beta} \\ x &= \mu - \beta\log(\log(\frac{1}{U})) \end{aligned}$$

$$\therefore F^{-1}(u;\mu,\beta) = \mu - \beta\log(\log(\frac{1}{U}))$$

(b) Parametric bootstrap is a technique that is used to create confidence intervals on parameters estimates of a distribution. In Parametric bootstrap we fit a parametric distribution to the given data - this is mostly done using using Maximum Likelihood estimators (MLE). From a distribution with these (solved) parameters we generate B bootstrap samples of size n, based on these samples we obtain new bootstrap parameter 're-estimates' which are used to create the distribution of deltas -using differences from the original estimates. In this case we used the Inversion Method to generate the bootstrap using the inverse distribution function derived in (a). Subtracting a specific percentile of these deltas from the original parameter estimates gives us an approximate confidence interval. The values for $\hat{\mu}$ and $\hat{\beta}$ where 4.1477 and 1.458 respectively. With $B = 10000$ and n=500 the resultant 95% confidence intervals for μ and β were as follows:

$$CI_\mu = (4.0093, 4.281)$$

and,

$$CI_\beta = (1.3851, 1.5886)$$

The algorithim for this parametric bootstrap is shown in Algorithm 1

calculate $\hat{\beta}, \hat{\mu}$;
for $i = 1 \leftarrow B$ **do**
  generate n $U(0,1)$ samples;
  generate bootstrap sample using $F^{-1}(u; \mu, \beta)$;
  calculate and store β_{boot}, μ_{boot};
end
Δ_β=sort$(\beta_{boot} - \hat{\beta})$;
Δ_μ=sort$(\mu_{boot} - \hat{\mu})$;
$CI_\beta = (\hat{\beta} - \Delta_\beta(ceil((1-\alpha/2) * B)), \hat{\beta} - \Delta_\beta(ceil((\alpha/2) * B))$;
$CI_\mu = (\hat{\mu} - \Delta_\mu(ceil((1-\alpha/2) * B)), \hat{\mu} - \Delta_\mu(ceil((\alpha/2) * B))$

Algorithm 1: Parametric Bootstrap

Figure 1 shows the distributions of the deltas for μ and β.

(c) Similar procedure as in (b) was performed to obtain the 95% one sided upper bound for 100 -year return value. The expected 100 year return value based on the original parameter estimates was 16.5436. With each set of bootstrap parameters the 100 -year return value was evaluated as :

$$F^{-}1(1 - 1/T; \mu_{boot}, \beta_{boot})$$

where $T = 3 \times 14 \times 100$ given 13 measurements in 3 months of each year .We then used the deltas for the return to create a 95% one-sided confidence interval for the upper-bound of the 100 year return value which found to be:

$$UB = 17.3067$$

We chose to provide an upper bound as this could be used when deciding to construct a barrier which can withstand very extreme wave heights for example. Figure 1 shows the distribution of the deltas for 100-year return value.

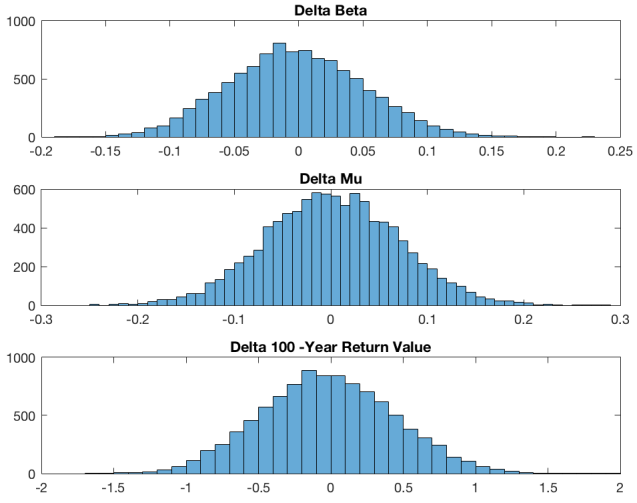


Figure 1: A Histogram of bootstrap Deltas for μ and β