Bayesian statistics, simulation and software

Module 6: The Gibbs sampler

Jesper Møller and Ege Rubak

Department of Mathematical Sciences
Aalborg University

The Gibbs sampler — the general algorithm

Aim: We want to sample $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ from a density $\pi(\boldsymbol{\theta})$, e.g. the prior or the posterior density (in the latter case, suppressing in the notation the dependence of the data x: $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}|x)$).
Assume $\theta_i \in \Omega_i \subseteq \mathbf{R}^{d_i}$ and $\boldsymbol{\theta} \in \Omega_1 \times \Omega_2 \times \cdots \times \Omega_k \subseteq \mathbf{R}^{d_1+d_2+\cdots+d_k}$
We can then generate an *approximate* sample from $\pi(\boldsymbol{\theta})$ (provided some technical conditions are satisfied) as follows:

Gibbs Sampler

- Choose initial value $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_k^{(0)})$.
- For $i = 1, 2, \ldots, t$
  1. Generate $\theta_1^{(i)} \sim \pi(\theta_1|\theta_2^{(i-1)}, \theta_3^{(i-1)}, \ldots, \theta_k^{(i-1)})$
  2. Generate $\theta_2^{(i)} \sim \pi(\theta_2|\theta_1^{(i)}, \theta_3^{(i-1)}, \ldots, \theta_k^{(i-1)})$

     ⋮

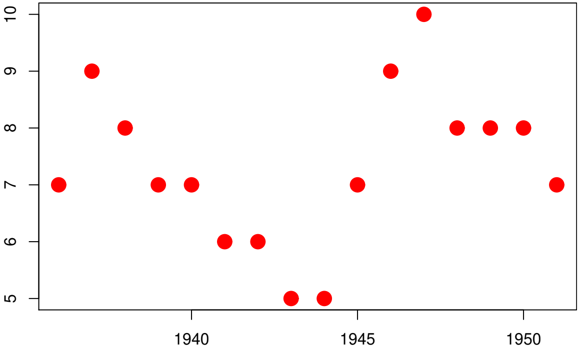
  k. Generate $\theta_k^{(i)} \sim \pi(\theta_k|\theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_{k-1}^{(i)})$

The higher i is the closer $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_k^{(i)})$ is to being a sample from $\pi(\boldsymbol{\theta})$.
When $d_1, \ldots, d_k$ are small, Gibbs sampling may be easy to use.

Example: Marriage rates in Italy

For the years 1936 to 1951 (16 years) the marriage rates per 1000 of the population in Italy have been observed. How do we compare marriage rates that occurred during WW2 to rates just before and after?

Data: $\mathbf{y} = (y_1, y_2, \ldots, y_{16})$.



Italian marriages: Model

Model: Conditional on (true) rates $\lambda_1, \lambda_2, \ldots, \lambda_{16}$ the observed rates $y_1, y_2, \ldots, y_{16}$ are independent and $y_i \sim Pois(\lambda_i)$:

- Joint density of data $\mathbf{y}$:

$$\pi(\mathbf{y}|\boldsymbol{\lambda}) = \prod_{i=1}^{16} \pi(y_i|\lambda_i) = \prod_{i=1}^{16} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}.$$

Italian marriages: Prior and hyper prior

Prior: Conditional on a *hyper parameter* $\beta > 0$ the rates $\lambda_1, \lambda_2, \ldots, \lambda_{16}$ are i.i.d. with $\lambda_i|\beta \sim Exp(\beta)$:

- The prior density of $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_{16})$ conditional on β is

$$\pi(\boldsymbol{\lambda}|\beta) = \prod_{i=1}^{16} \pi(\lambda_i|\beta) = \prod_{i=1}^{16} \beta \exp(-\beta\lambda_i).$$

As we are not sure which value the common parameter β should take, we assume a so-called *hyper prior* on β:

- $\beta \sim Exp(1)$, i.e. $\pi(\beta) = e^{-\beta}$ for $\beta > 0$.

Thus the prior density for $(\boldsymbol{\lambda}, \beta)$ is

$$\pi(\boldsymbol{\lambda}, \beta) = \pi(\beta)\pi(\boldsymbol{\lambda}|\beta) = e^{-\beta} \prod_{i=1}^{16} \beta \exp(-\beta\lambda_i), \qquad (\boldsymbol{\lambda}, \beta) \in (0, \infty)^{17}.$$

Posterior

Posterior density:

$$\begin{aligned}\pi(\boldsymbol{\lambda},\beta|\mathbf{y}) &\propto \pi(\mathbf{y}|\boldsymbol{\lambda},\beta)\pi(\boldsymbol{\lambda},\beta)\\ &= \left(\prod_{i=1}^{16}\pi(y_i|\lambda_i)\right)\left(\prod_{i=1}^{16}\pi(\lambda_i|\beta)\right)\pi(\beta)\\ &= \left(\prod_{i=1}^{16}\frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}\right)\left(\prod_{i=1}^{16}\beta e^{-\beta\lambda_i}\right)e^{-\beta}, \quad \lambda_1,\dots,\lambda_{16},\beta>0.\end{aligned}$$

This looks complicated. Therefore to explore the posterior we make use of a Gibbs sampler with low dimensional distributions – these are called full conditionals and are specified as follows.

Full conditionals — λ_i conditioned on anything else

- Let $\boldsymbol{\lambda}_{-i} = (\lambda_1, \ldots, \lambda_{i-1}, \lambda_{i+1}, \ldots, \lambda_{16})$, $i = 1, \ldots, 16$.
- The full conditional for λ_i has density

$$
\begin{aligned}
\pi(\lambda_i | \boldsymbol{\lambda}_{-i}, \mathbf{y}, \beta) &= \frac{\pi(\lambda_i, \boldsymbol{\lambda}_{-i}, \mathbf{y}, \beta)}{\pi(\boldsymbol{\lambda}_{-i}, \mathbf{y}, \beta)} \propto \pi(\boldsymbol{\lambda}, \mathbf{y}, \beta) \\
&= \left(\prod_{j=1}^{16} \pi(y_j | \lambda_j) \right) \left(\prod_{j=1}^{16} \pi(\lambda_j | \beta) \right) \pi(\beta) \\
&\propto \pi(y_i | \lambda_i) \pi(\lambda_i | \beta) \\
&= \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \cdot \beta e^{-\beta \lambda_i} \\
&\propto \lambda_i^{y_i + 1 - 1} e^{-\lambda_i (1 + \beta)} \\
&\sim Gamma(y_i + 1, (1 + \beta)^{-1}),
\end{aligned}
$$

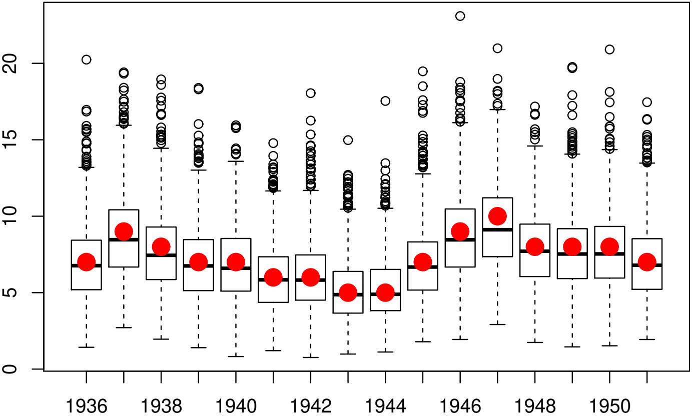
Full conditionals — β

- The full conditional for β has density

$$
\begin{aligned}
\pi(\beta|\boldsymbol{\lambda},\mathbf{y}) &\propto \left(\prod_{i=1}^{16}\pi(y_i|\lambda_i)\right)\left(\prod_{i=1}^{16}\pi(\lambda_i|\beta)\right)\pi(\beta)\\
&\propto \left(\prod_{i=1}^{16}\pi(\lambda_i|\beta)\right)\pi(\beta)\\
&= \left(\prod_{i=1}^{16}\beta e^{-\beta\lambda_i}\right)e^{-\beta}\\
&\propto \beta^{16+1-1}e^{-\beta(1+\sum_{i=1}^{16}\lambda_i)}\\
&\sim Gamma\left(17,\left(1+\sum_{i=1}^{n}\lambda_i\right)^{-1}\right).
\end{aligned}
$$

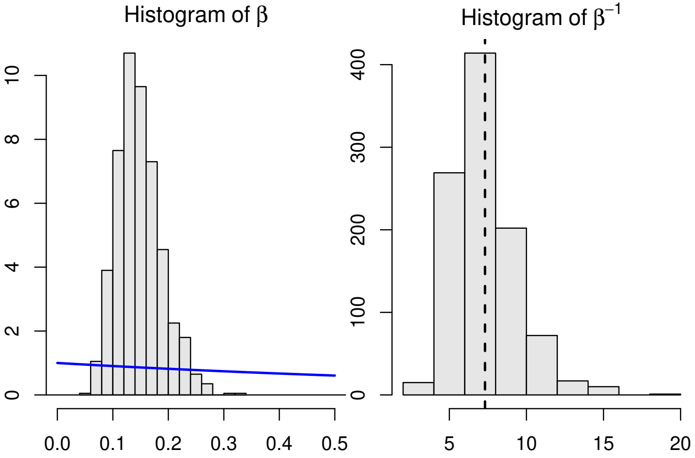
Posterior marriage rates: Boxplots

Although there is a clear trend of a drop during WW2 it is not extreme:



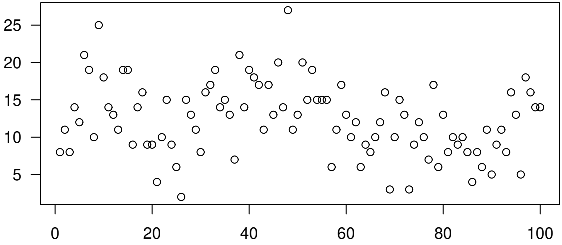
Posterior distribution of β

Note that β^{-1} is the prior mean of a marriage rate.



Example: Airport mishandling of luggage

Every hour the number of mishandled bags have been recorded:



Notation:

- Let $y_t \in \mathbb{N}_0$ denote the number of mishandled bags at time (hour) t.
- We consider two unknown cases/states at the airport: **'Normal'** or **'broken'**. Let $x_t \in \{1, 2\}$ denote the state of the airport at time t (1=normal, 2=broken).

Objective:

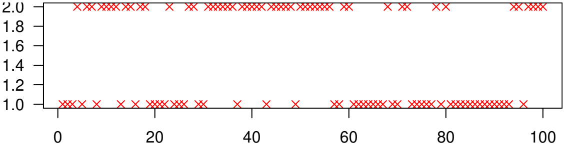
- Estimate the state of the airport at each hour.

Data model

- Conditional on $\mathbf{x} = (x_1, \ldots, x_{100})$ we assume the number of mishandled bags $y_1, \ldots, y_n$ are independent, and the conditional distribution of each $y_t|\mathbf{x}$ depends only on x_t: $y_t|\mathbf{x} \sim y_t|x_t$.
- Specifically, (based on previous knowledge) we assume
  - $y_t|x_t = 1 \sim Pois(10)$ Normal state
  - $y_t|x_t = 2 \sim Pois(15)$ Broken state

Note that the MLE of x_t (i.e., the most likely state according to the data model) is 1 if and only if $p(y_t|x_t = 1) > p(y_t|x_t = 2)$, that is,

$$\frac{e^{-10}10^{y_t}}{y_t!} > \frac{e^{-15}15^{y_t}}{y_t!} \quad \Leftrightarrow \quad \left(\frac{10}{15}\right)^{y_t} > e^{-5} \quad \Leftrightarrow \quad y_t > \frac{5}{\ln 15 - \ln 10}.$$

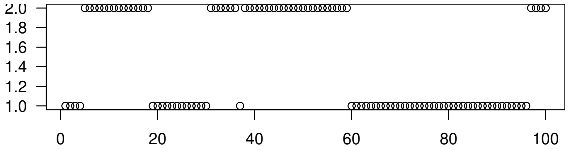


Prior

It is known that the airport tends to 'stick' in the same state. Thus the prior for $\mathbf{x}$ is assumed to be a Markov chain:

- $P(x_1 = 1) = P(x_1 = 2) = \frac{1}{2}$ (probabilities for initial state)
- $P(x_{t+1} = x_t | x_t) = 0.9$ (probability of staying)
- $P(x_{t+1} \neq x_t | x_t) = 0.1$ (probability of switching)

Example of a realisation from the prior:



Posterior

The posterior density is

$$\begin{aligned}\pi(\mathbf{x}|\mathbf{y}) &\propto \pi(\mathbf{y}|\mathbf{x})\pi(\mathbf{x})\\ &= \left(\prod_{t=1}^{100}\pi(y_t|x_t)\right)\left(\pi(x_1)\prod_{t=1}^{99}\pi(x_{t+1}|x_t)\right)\end{aligned}$$

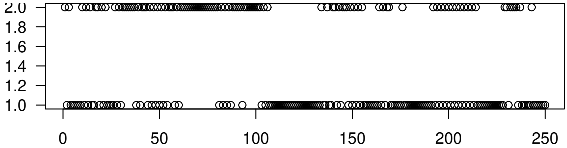
Thus we obtain a full conditional for each x_t:

$$\begin{aligned}\pi(x_1|y_1,\mathbf{x}_{-1}) &\propto \pi(y_1|x_1)\pi(x_1)\pi(x_2|x_1),\\ \pi(x_t|y_t,\mathbf{x}_{-t}) &\propto \pi(y_t|x_t)\pi(x_{t+1}|x_t)\pi(x_t|x_{t-1}) \quad \text{for } t=2,3,\ldots,99,\\ \pi(x_{100}|y_{100},\mathbf{x}_{-100}) &\propto \pi(y_{100}|x_{100})\pi(x_{100}|x_{99}).\end{aligned}$$

Since $x_t \in \{1,2\}$, it is easy to simulate x_t from this full conditional.

Posterior results

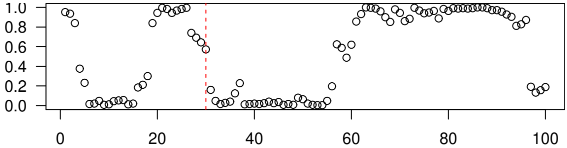
Example: Plot of e.g. x_{30} during $I = 250$ "sweeps" of the Gibbs sampler:



Estimate of the posterior probability that $x_{30} = 1$:

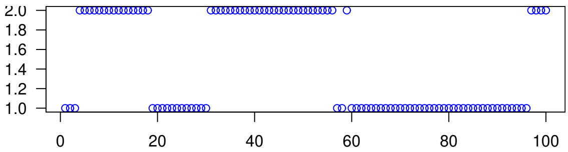
$$P(x_{30} = 1|\mathbf{y}) \approx \frac{1}{I} \sum_{i=1}^{I} 1[x_{30,i} = 1] = 57.2\%.$$

For all hours: Plot of posterior probabilities $P(x_t = 1|\mathbf{y})$, $t = 1, \ldots, 100$.

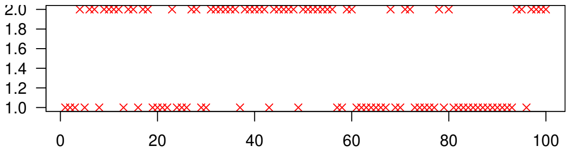


Comparison

MAP = the most likely state according to the posterior distribution:



Compare with the MLE (the most likely state using only the data model):



Which of these 3 plots do you like? (I like the plots of posterior probabilities together with the MAP!)