Recap: Bayesian Idea

- **Parameter of interest**: θ (e.g. mean height μ)
- **Data model**: Conditional on θ, data x is distributed according to pf/pdf $\pi(x)$:

$$\pi(x|\theta) \propto L(\theta; x) \quad \leftarrow \text{the likelihood}$$

- **Prior**: Prior knowledge (i.e. *before* collecting data) about θ is summaries by a pf/pdf,

$$\pi(\theta) \quad \leftarrow \text{the prior}$$

- **Posterior** : The updated knowledge about θ *after* collecting data: The conditional distribution of θ given data x:

$$\pi(\theta|x) = \frac{\pi(x|\theta)\pi(\theta)}{\pi(\theta)} \propto \pi(x|\theta)\pi(\theta)$$

$$\text{“}posterior = likelihood \times prior''$$

Recap: Normal model

Data: $\mathbf{X} = (X_1, X_2, \ldots, X_n)$.

Data model: $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \tau)$

$$\pi(\mathbf{x}|\mu,\tau) = \left(\frac{\tau}{2\pi}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2}\tau\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

We have considered two cases

1. Unknown mean μ, known precision τ
   - **Prior**: $\pi(\mu) = \mathcal{N}(\mu_0, \tau_0)$
   - **Posterior**: $\pi(\mu|x) = \mathcal{N}(\frac{n\tau\bar{x}+\tau_0\mu_0}{n\tau+\tau_0}, n\tau+\tau_0)$

2. Unknown precision τ, know mean μ
   - **Prior**: $\pi(\tau) = Gamma(\alpha, \beta)$
   - **Posterior**: $\pi(\tau|x) = Gamma(\frac{n}{2}+\alpha, \{\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2 + \frac{1}{\beta}\}^{-1})$

What if *both* mean and precision are unknown?

Normal example: mean and precision unknown

Assume both μ and τ unknown.

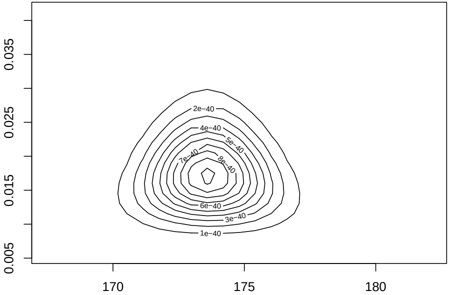
Prior: $\pi(\mu, \tau)$

Posterior: $\pi(\mu, \tau|\mathbf{x}) \propto \pi(\mathbf{x}|\mu, \tau)\pi(\mu, \tau)$.

Choice of prior: Assume μ and τ a priori independent, and normal and gamma, respectively. Specifically $\pi(\mu, \tau) = \pi(\mu)\pi(\tau)$, where

- $\pi(\mu) = \mathcal{N}(\mu_0, \tau_0)$
- $\pi(\tau) = Gamma(\alpha, \beta)$

Posterior density:



Normal example: mean and precision unknown

One question: What is the posterior marginal distribution of μ? It has density

$$\pi(\mu|\mathbf{x}) = \int \pi(\mu, \tau|\mathbf{x}) d\tau.$$

The integral is easy, but $\pi(\mu|\mathbf{x})$ is not a standard distribution.

Solution: Turn to simulations.

Simulation: Toy example

In the normal case, when μ is known, the posterior distribution of τ is

$$\pi(\tau|x) = Gamma\left(\frac{n}{2} + \alpha, \{\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2 + \frac{1}{\beta}\}^{-1}\right)$$

Assume that

- We *do not* know the mean and variance of $Gamma(\cdot, \cdot)$.
- We *cannot* integrate $\pi(\tau|x)$.
- We *can* simulate $\tau \sim \pi(\tau|x)$.

Now answer these questions:

- What is the posterior mean of τ?
- What is the posterior probability that $\tau > 0.025$?

Simulating an answer

Assume

- we have generated $\tau^{(1)}, \tau^{(2)}, \dots, \tau^{(t)} \overset{iid}{\sim} \pi(\tau|\mathbf{x})$.
- that h is a real function defined on $\mathbf{R}$.

An estimate of $\mathbb{E}[h(\tau)|\mathbf{x}]$ is given by

$$\frac{1}{t}\sum_{i=1}^{t} h(\tau^{(i)})$$

Answer to question 1: An estimate of the posterior mean is

$$\frac{1}{t}\sum_{i=1}^{t} \tau^{(i)}$$

Answer to question 2: Recall that $P(\tau > 0.025|\mathbf{x}) = \mathbb{E}[\mathbf{1}[\tau > 0.025]]$, hence an estimate of the probability is

$$\frac{1}{t}\sum_{i=1}^{t} \mathbf{1}[\tau^{(i)} > 0.025]$$

Unknown mean and precision: Simulating an answer

Setup: We now return to the original problem: Both μ and τ are unknown.

Problem: We could not say much, e.g. we could not recognise the marginal posterior of μ.

Can do: We know the *conditional* posterior distribution of μ given τ (and vice versa).

- $\pi(\mu|\tau, \mathbf{x}) = \mathcal{N}(\frac{n\tau\bar{x}+\tau_0\mu_0}{n\tau+\tau_0}, n\tau + \tau_0)$
- $\pi(\tau|\mu, \mathbf{x}) = Gamma(\frac{n}{2} + \alpha, \{\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2 + \frac{1}{\beta}\}^{-1})$

The idea is now to (attempt to) simulate $\pi(\mu, \tau|\mathbf{x})$ by alternating between

- simulating μ conditional on τ
- simulating τ conditional on μ

Later in the course we will show that approach in fact works (in a certain sense).

As an algorithm

- Choose initial values $\mu^{(0)}$ and $\tau^{(0)}$.
- For $i = 1, 2, \ldots, t$
  1. Conditional on $\tau^{(i-1)}$, generate

$$\mu^{(i)}|\mathbf{x}, \tau^{(i-1)} \sim \mathcal{N}\left(\frac{n\tau^{(i-1)}\bar{x} + \tau_0\mu_0}{n\tau^{(i-1)} + \tau_0}, n\tau^{(i-1)} + \tau_0\right)$$

  2. Conditional on $\mu^{(i)}$ generate

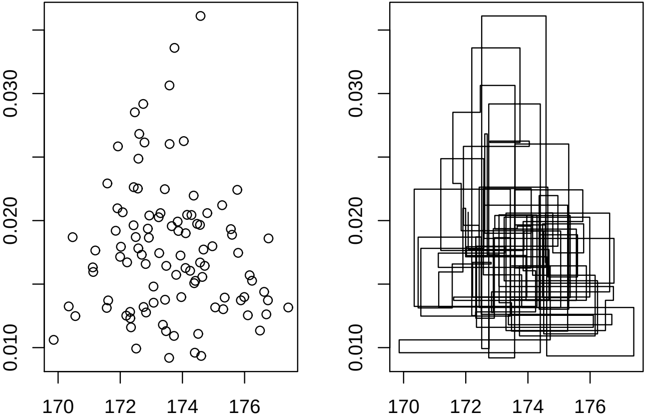
$$\tau^{(i)}|\mathbf{x}, \mu^{(i)} \sim Gamma(\frac{n}{2} + \alpha, \{\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu^{(i)})^2 + \frac{1}{\beta}\}^{-1})$$

- This algorithm generates a sequence of parameter pairs:

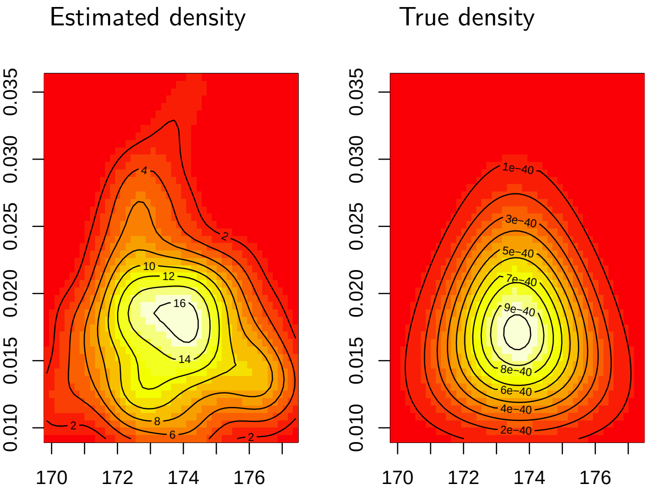
$$(\mu^{(0)}, \tau^{(0)}), (\mu^{(1)}, \tau^{(1)}), \ldots, (\mu^{(t)}, \tau^{(t)}),$$

- $(\mu^{(i)}, \tau^{(i)})$ is approximately a sample from the posterior $\pi(\mu, \tau|\mathbf{x})$.
- The higher i is, the better this approximation is.
- Algorithm is an example of a *Gibbs sampler*.
- We have generated a realisation of a Markov chain.

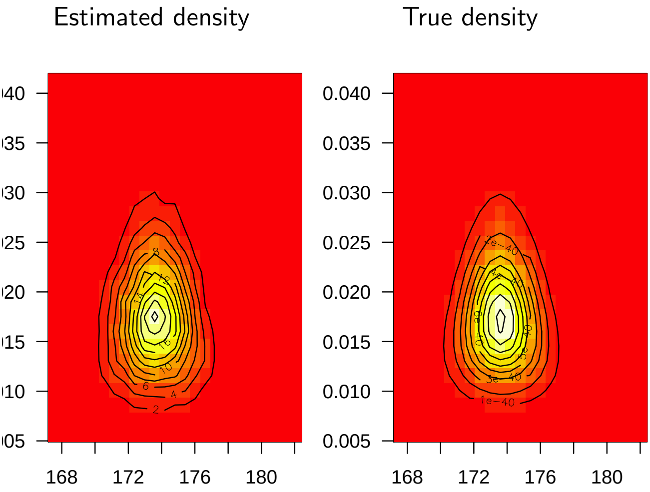
Simulated posterior distribution ($t = 100$)



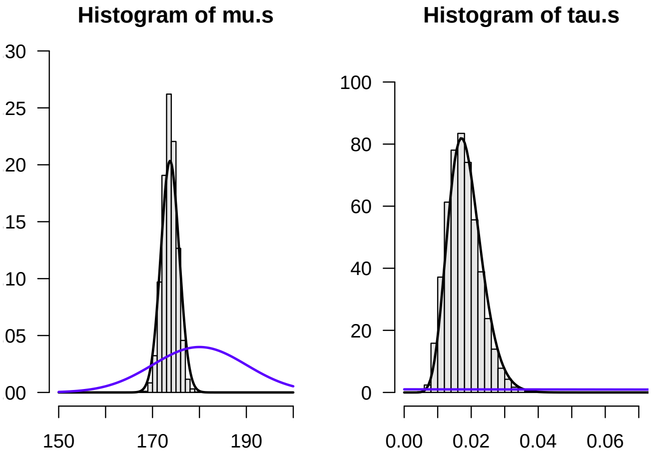
Simulated joint posterior distribution ($t = 100$)



Simulated joint posterior distribution ($t = 10,000$)



Marginal posterior distributions



The Gibbs sampler — The general algorithm

Aim: We want to sample $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ from a pdf/pf $\pi(\boldsymbol{\theta})$.

Assume $\theta_i \in \Omega_i \subseteq \mathbf{R}^{d_i}$. Then, $\theta \in \Omega_1 \times \Omega_2 \times \cdots \times \Omega_k \subseteq \mathbf{R}^{d_1+d_2+\cdots+d_k}$

We can now (under some conditions) generate an *approximate* sample from $\pi(\boldsymbol{\theta})$ 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 $\theta_2^{(i)} \sim \pi(\theta_2 | \theta_1^{(i)}, \theta_3^{(i-1)}, \ldots, \theta_k^{(i-1)})$

     $\vdots$

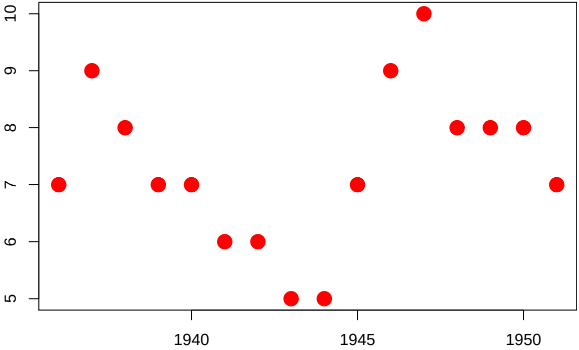
  k. Generate $\theta_k^{(i)} \sim \pi(\theta_k | \theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_{k-1}^{(i)})$

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 $\pi(\boldsymbol{\theta})$.

Example: Marriage rates in Italy!

For the years 1936 to 1951 (16 years) we have observed the marriage rates 1000 in Italy.

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 distribution of $\mathbf{y}$

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

- $\pi(y_i|\lambda_i) = Poisson(\lambda_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}$

Italian marriages: Prior

Prior: Conditional on *hyper parameter* β the rates $\lambda_1, \lambda_2, \ldots, \lambda_{16}$ are independent and $\lambda_i|\beta \sim Exp(\beta)$

- The prior distribution of the λ_is conditional on β:

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

- $\pi(\lambda_i|\beta) = Exp(\beta) = \beta e^{-\beta\lambda_i}$

As we a not sure which value the common parameter β should takes, we assume a *hyper prior* on β:

- A prior we assume that $\beta \sim Exp(1)$
  $\pi(\beta) = e^{-\beta}$.

Posterior

Conditional on the observed marriage rates what are the posterior distribution for the true rates?

Posterior:

$$\begin{aligned}
\pi(\boldsymbol{\lambda}, \beta|\mathbf{y}) &\propto \pi(\mathbf{y}|\boldsymbol{\lambda}, \beta)\pi(\boldsymbol{\lambda}, \beta) \\
&= \pi(\mathbf{y}|\boldsymbol{\lambda})\pi(\boldsymbol{\lambda}|\beta)\pi(\beta) \\
&= \prod_{i=1}^{16} \pi(y_i|\lambda_i) \prod_{i=1}^{16} \pi(\lambda_i|\beta)\pi(\beta) \\
&= \prod_{i=1}^{16} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \prod_{i=1}^{16} \beta e^{-\beta\lambda_i} e^{-\beta}
\end{aligned}$$

To explore the posterior we makes use of a Gibbs sampler. For this we need the full conditionals.

Full conditionals — λ_i

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

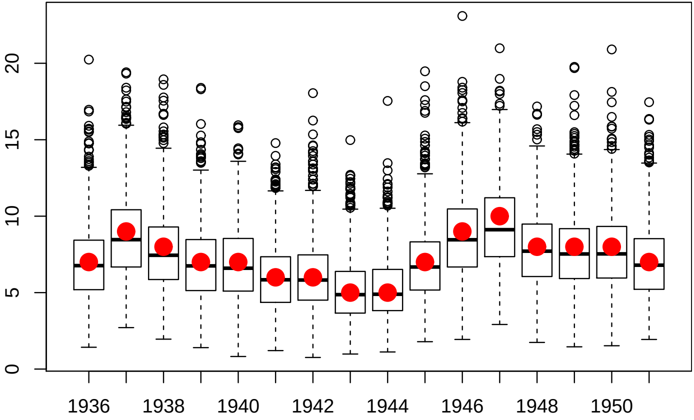
$$
\begin{aligned}
\pi(\lambda_i|\boldsymbol{\lambda}_{-i}, \mathbf{y}, \beta) &= \frac{\pi(\lambda_i, \boldsymbol{\lambda}_{-i}, \mathbf{y}, \beta)}{\pi(\boldsymbol{\lambda}_{-i}, \mathbf{y}, \beta)} \\
&\propto \prod_{i=1}^{16} \pi(y_i|\lambda_i) \prod_{i=1}^{16} \pi(\lambda_i|\beta)\pi(\beta) \\
&\propto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

Full conditionals — β

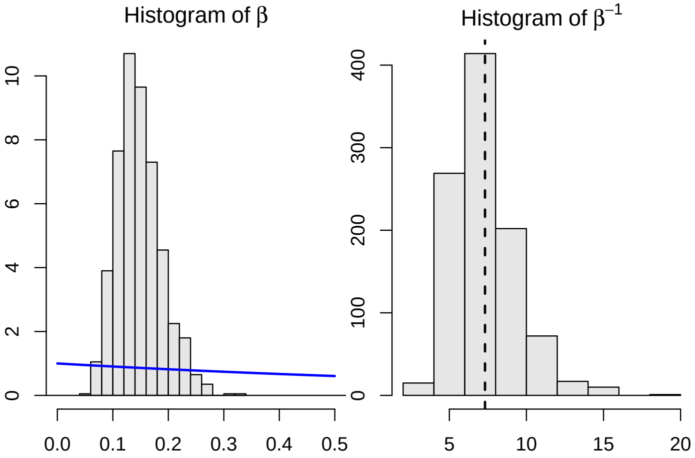
- Full conditional for β

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

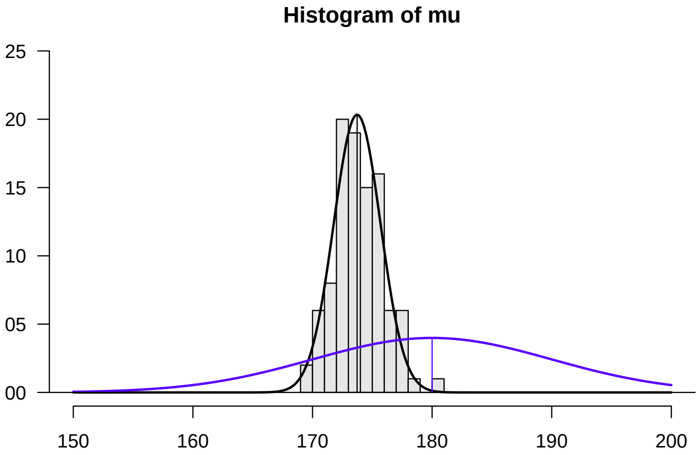
Posterior marriage rates



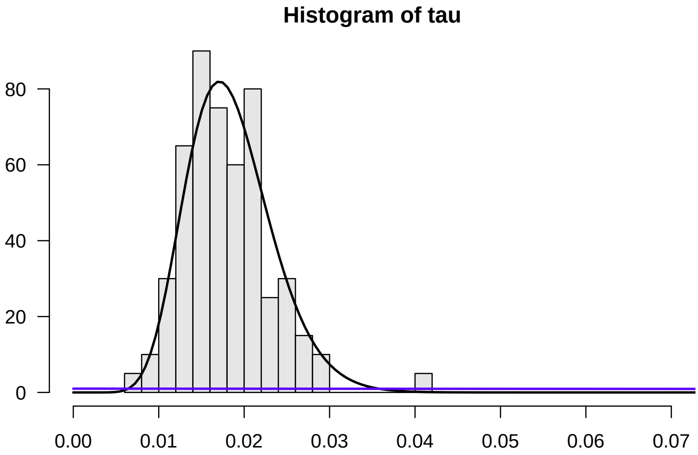
Posterior distribution of β



Known precision: Simulated posterior distribution

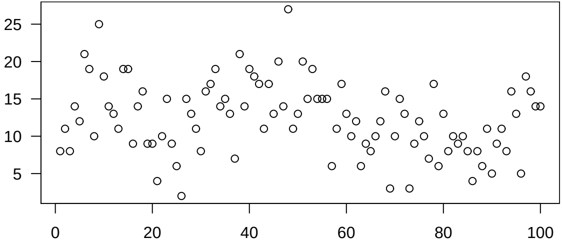


Known mean: Simulated posterior distribution



Example: Airport mishanling of luggage

Every hour the number of mishandled bags have been recorded:



The aiport is one of two states: **Normal** or **Broken**.

Notation:

- Let $y_t \in \mathbb{N}_0$ denote the number of mishandled bags at time t
- Let $x_t \in \{1, 2\}$ denote the state of the airport at time t (1=normal, 2=broken)

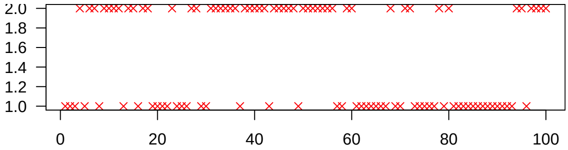
Objective:

- Estimate the state of the airport at each time point.

Mishandling: Data model

- Conditional on $\mathbf{x} = (x_1, \ldots, x_n)$ the number of mishandlings are independent, and the distribution of y_T only depends on x_t.
- The number of mishandlings is assumed to follow a Poisson distribution:
  - $y_t | x_t = 1 \sim Pois(10)$ Normal state
  - $y_t | x_t = 2 \sim Pois(15)$ Broken state

Most likely state according to data model:



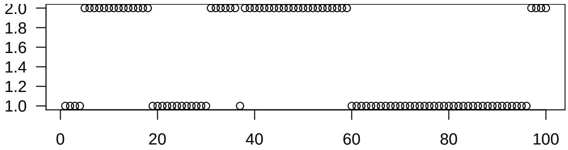
Mishandling: Prior

It is known that an airport tends to “stick” in the same state.

Hence we assume a Markov chain priort:

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

Example of realisation of prior:



Mishandling: Posterior

The posterior:

$$\pi(x|y) \propto \pi(y|x)\pi(x)$$
$$= \prod_{t=1}^{N} \pi(y_t|x_t)\pi(x_1) \prod_{t=1}^{N-1} \pi(x_{t+1}|x_t)$$

Full conditionals

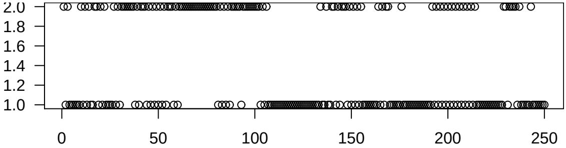
$$\pi(x_t|y_t, \mathbf{x}_{-t}) \propto \pi(y_t|x_t)\pi(x_{t+1}|x_t)\pi(x_t|x_{t-1})$$

Hence $x_t|y_t, \mathbf{x}_{-t}$ is a Bernoulli random variable:

$$\pi(x_t = i|y_t) = \frac{\pi(y_t|x_t = i)\pi(x_{t+1}|x_t = i)\pi(x_t = i|x_{t-1})}{\sum_{j=1}^{2} \pi(y_t|x_t = j)\pi(x_{t+1}|x_t = j)\pi(x_t = j|x_{t-1})}$$

Posterior results

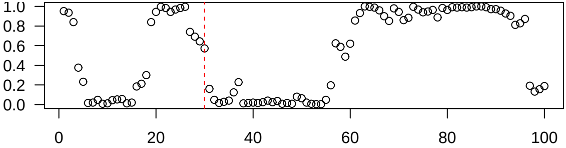
Plot of x_{30} during $I = 250$ "sweeps" of the Gibbs sampler:



Estimate of the posterior probability that $x_t = 1$:

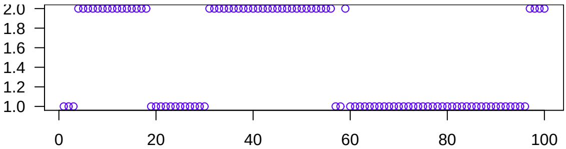
$$P(x_t = 1|\mathbf{y}) \approx \frac{1}{I}\sum_{i=1}^{I} 1[x_{t,i} = 1] = 57.2\%$$

Plot of the posterior probability for all times:



Comparison

Most likely state according to the posterior distribution



Compare this to the most likely state using only the data model: