Bayesian statistics, simulation and software

Module 5: Intro to simulation based inference

Jesper Møller and Ege Rubak

Department of Mathematical Sciences
Aalborg University

Recap: Bayesian Idea

- **Parameter of interest**: θ (e.g. mean height μ)
- **Data model**: Conditional on θ, data x is distributed according to the density

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

- **Prior**: Prior knowledge (i.e. *before* collecting data) about θ is summaries by a density

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

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

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

$$(\text{“}posterior \propto likelihood \times prior''\text{)}$$

Recap: Normal model

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

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

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

We have considered two cases

1. Unknown mean μ, known precision τ
   - **Prior**: $\pi(\mu) \sim \mathcal{N}(\mu_0, \tau_0)$
   - **Posterior**: $\pi(\mu|x) \sim \mathcal{N}(\frac{n\tau\bar{x}+\tau_0\mu_0}{n\tau+\tau_0}, n\tau+\tau_0)$
2. Unknown precision τ, know mean μ
   - **Prior**: $\pi(\tau) \sim Gamma(\alpha, \beta)$
   - **Posterior**: $\pi(\tau|x) \sim Gamma(\frac{n}{2}+\alpha, \{\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2+\frac{1}{\beta}\}^{-1})$

What if *both* mean and precision are unknown?

Normal example: mean and precision unknown

Assume both μ and τ are unknown.

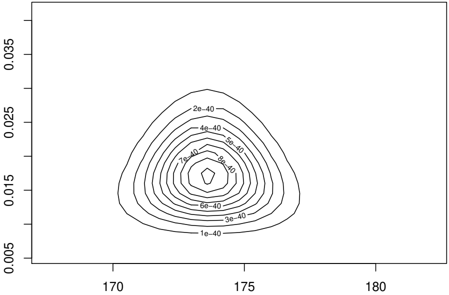
Prior: $\pi(\mu, \tau)$

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

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

Posterior density (note that a posteriori μ and τ are dependent!):



Normal example: mean and precision unknown

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

It has density

$$\begin{aligned}
\pi(\mu|\mathbf{x}) &= \int_0^\infty \pi(\mu,\tau|\mathbf{x})d\tau \propto \int_0^\infty \pi(\mu)\pi(\tau)\pi(\mathbf{x}|\mu,\tau)d\tau \\
&\propto \int_0^\infty \exp\left(-\frac{1}{2}\tau_0(\mu-\mu_0)^2\right)\tau^{\alpha-1}\exp(-\tau/\beta) \\
&\qquad \tau^{n/2}\exp\left(-\frac{1}{2}\tau\sum_{i=1}^n(x_i-\mu)^2\right)d\tau \\
&= \exp\left(-\frac{1}{2}\tau_0(\mu-\mu_0)^2\right)\Gamma\left(\frac{n}{2}+\alpha\right)\left(\frac{1}{2}\sum_{i=1}^n(x_i-\mu)^2+\frac{1}{\beta}\right)^{-(n/2+\alpha)}
\end{aligned}$$

where $\Gamma(\cdot)$ is the gamma-function. Maybe this derivation was "easy" but $\pi(\mu|\mathbf{x})$ is a non-standard distribution and it looks complicated!

Solution: Turn to simulations.

Simulation: Toy example

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

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

Imagine for the moment that we

- *do not* know the mean and variance of $Gamma(\cdot,\cdot)$,
- *cannot* integrate $\pi(\tau|x)$,
- *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 simulations $\tau^{(1)}, \tau^{(2)}, \dots, \tau^{(t)} \overset{iid}{\sim} \pi(\tau|\mathbf{x})$,
- h is a real function defined on $\mathbf{R}$ and we want $\mathbb{E}[h(\tau)|\mathbf{x}]$.

A natural estimate of $\mathbb{E}[h(\tau)|\mathbf{x}]$ is given by the *Monte Carlo estimate*:

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

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
$P(\tau > 0.025|\mathbf{x}) = \mathbb{E}(\mathbf{1}[\tau > 0.025]|\mathbf{x})$, 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).

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

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

- simulating μ conditional on τ,
- simulating τ conditional on μ.

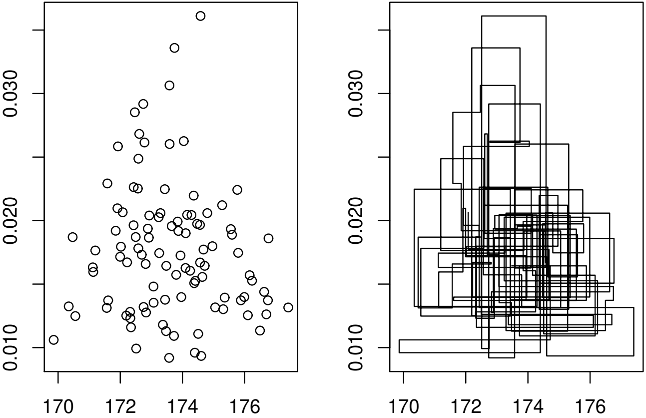
Later in the course we will show that this approach in fact works – it is an example of a so-called Gibbs sampler.

As an algorithm

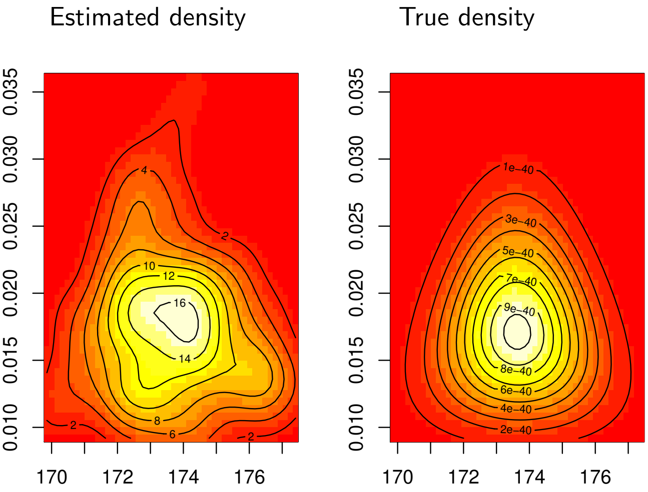
- Choose initial values $\mu^{(0)}$ and $\tau^{(0)}$ (actually in the updating scheme below $\mu^{(0)}$ is not used).
- For $i = 1, 2, \ldots, t$
  1. Conditional on $\tau^{(i-1)}$, generate
     $$\mu^{(i)}|\mathbf{x}, \tau^{(i-1)} \sim \mathcal{N}\left(\frac{n\tau^{(i-1)}\bar{x} + \tau_0\mu_0}{n\tau^{(i-1)} + \tau_0}, n\tau^{(i-1)} + \tau_0\right).$$
  2. Conditional on $\mu^{(i)}$ generate
     $$\tau^{(i)}|\mathbf{x}, \mu^{(i)} \sim Gamma\left(\frac{n}{2} + \alpha, \left\{\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu^{(i)})^2 + \frac{1}{\beta}\right\}^{-1}\right).$$

- This algorithm generates a sequence of parameter pairs:
  $$(\mu^{(0)}, \tau^{(0)}), (\mu^{(1)}, \tau^{(1)}), \ldots, (\mu^{(t)}, \tau^{(t)})$$
  is a realisation of a Markov chain.
- Fact: $(\mu^{(i)}, \tau^{(i)})$ is approximately a sample from the posterior $\pi(\mu, \tau|\mathbf{x})$.
- Fact: the higher i is, the better this approximation is.

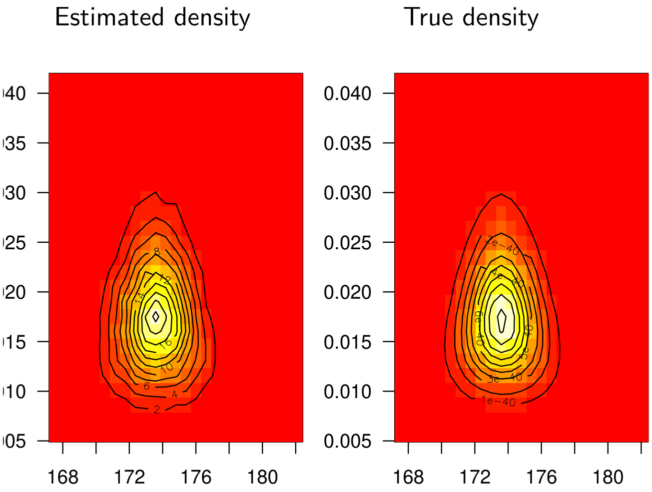
Simulated posterior distribution ($t = 100$)



Simulated joint posterior distribution ($t = 100$)



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



Marginal posterior distributions