Introduction to Bayesian analysis for medical studies
–
Part II: Bayesian computation

Boris Hejblum
https://borishejblum.science

Graduate School of Health and Medical Sciences
at the University of Copenhagen

June 11th, 2020

Introduction

Estimating the *posterior* distribution is often costly

Bayesian computational statistics

Computational aspects of Bayesian inference can get sophisticated but are key to its successful application

Numerical integration – I

Real world applications: $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)$
$\Rightarrow$ joint *posterior* distribution of all d parameters

⚠ hard to compute:

- complexe likelihood
- integrating constant $f(\boldsymbol{y}) = \int_{\Theta^d} f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}$
- ...

Analytical form rarely available
$\Rightarrow$ numerical computations: integral of d multiplicity
– difficult when d is big (numerical issues as soon as $d > 4$)

Numerical integration – II

Even dimension 1 can be tough !

Example :

Let $x_1,\ldots,x_n$ *iid* according to a Cauchy distribution $\mathscr{C}(\theta,1)$
with *prior* $\pi(\theta) = \mathcal{N}(\mu,\sigma^2)$ (μ and σ known)

$$\begin{aligned} p(\theta|x_1,\ldots,x_n) &\propto f(x_1,\ldots,x_n|\theta)\pi(\theta) \\ &\propto e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} \prod_{i=1}^{n}(1+(x_i-\theta)^2)^{-1} \end{aligned}$$

⚠ normalizing constant has no analytical form $\Rightarrow$ no analytical form for this *posterior* distibution

Marginal *posterior* distributions

Objective: draw conclusion based on the joint *posterior* distribution

⇒ probability of all possible values for each parameter (i.e. their marginal distribution – uni-dimensional)

⚠ Recovering all of the *posterior* density **numerically** requires the calculation of multidimensional integrals **for each possible value of the parameter**

⇒ a sufficiently precise computation seems unrealistic

Marginal *posterior* distributions

Objective: draw conclusion based on the joint *posterior* distribution

⇒ probability of all possible values for each parameter (i.e. their marginal distribution – uni-dimensional)

⚠ Recovering all of the *posterior* density **numerically** requires the calculation of multidimensional integrals **for each possible value of the parameter**

⇒ a sufficiently precise computation seems unrealistic

Algorithms based on **sampling simulations**
especially **Markov chain Monte Carlo** (MCMC)