Bayesian Statistics

Fabio Sigrist

ETH Zurich, Autumn Semester 2019

Today's topics

- Laplace approximation
- Independent Monte Carlo methods

Bayesian computation

- Many applications involve hierarchical models with many parameters where the posterior often
  - is high-dimensional
  - has complex dependencies
  - does not belong to a standard family of distributions
- In order to compute moments, quantiles, marginal densities or posterior predictive distributions, we need approximations

Laplace approximation

Laplace approximation

Laplace approximations are used to approximate integrals of the form

$$\int h(\theta)q(\theta)d\theta$$

where

- q is a possibly unnormalized smooth density which is concentrated around its mode $\theta_0 = \arg\max \log q(\theta)$
- h is an arbitrary smooth function

Laplace approximation

Expanding h into a first order Taylor series at the mode θ_0, we obtain

$$\int h(\theta)q(\theta)d\theta \approx h(\theta_0)q(\theta_0)(\det J(\theta_0))^{-1/2}(2\pi)^{p/2}$$

- $J(\theta)$ is the negative Hessian

$$J(\theta)_{ij} = -\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log q(\theta)$$

- The approximation error is of order $O(1/N)$

See blackboard for details

Laplace approximation

Assume we want to approximate a posterior expectation

$$\mathbb{E}(h(\theta) \mid x) = \frac{\int h(\theta)\pi(\theta)f(x \mid \theta)d\theta}{\int \pi(\theta)f(x \mid \theta)d\theta}$$

- We can use the Laplace approximation for the numerator and the denominator separately and take either $q(\theta) = f(x \mid \theta)$ or $q(\theta) = \pi(\theta)f(x \mid \theta)$
- If h is strictly positive, we can also take $q(\theta) = h(\theta)\pi(\theta)f(x \mid \theta)$ in the numerator
- If h is strictly positive and bounded away from 0 near θ_0, one can obtain an approximation error for the above ratio of $O(N^{-2})$

Laplace approximation

See blackboard for:

- Justification of approximation
- Bayes factor and BIC

Clicker question

Extensions of the Laplace approximation

Various **extensions** of the Laplace approximation are possible:

- If $q(\theta)$ is bimodal with modes at θ_0 and θ_1 that are well separated, we can use approximations around θ_0 and θ_1 separately
- We can also use higher order Taylor approximations

  See blackboard

Independent Monte Carlo methods

Monte Carlo methods and Bayesian statistics

- **Use of Monte Carlo methods in Bayesian statistics**:

  Approximate the posterior $\pi(\theta \mid x)$ and derived quantities such as $\mathbb{E}(h(\theta) \mid x)$ by simulating from $\pi(\theta \mid x)$

Introduction independent Monte Carlo methods

- Assume that $X \sim \pi$ where π is some distribution which we call **target distribution** in the following. Our goal is to approximate*

$$\mu_h = \mathbb{E}_\pi(h(X)) = \int h(x)\pi(x)dx$$

- If we have an i.i.d. sample $(X^1, \dots X^N)$ from π, the **law of large numbers** justifies the following approximation

$$\mu_h \approx \bar{h}_N := \frac{1}{N}\sum_{t=1}^{N} h(X^t)$$

*We use this notation in order to be consistent with the literature on Monte Carlo methods. Monte Carlo methods are not only used in Bayesian statistics but also in other fields. In Bayesian statistics, X corresponds to θ and π to $\pi(\theta \mid x)$

Convergence of independent Monte Carlo methods

Due to the **central limit theorem**, we have

$$\frac{\sqrt{N}\left(\bar{h}_N - \mu_h\right)}{\sqrt{\sum_{t=1}^{N}(h(X^t) - \bar{h}_N)^2/N}} \overset{\text{approx}}{\sim} \mathcal{N}(0,1)$$

- The **convergence rate $N^{-1/2}$** is independent of the dimension
- h does not need to be smooth (only $\int h(x)^2\pi(x)dx < \infty$ is needed)

See blackboard

Clicker question

Obtaining samples from π

How can we obtain i.i.d. samples X^t from π?

- All Monte Carlo algorithms we will see draw samples from a target distribution π on the basis of a sequence of i.i.d. uniform random variables (U^t)
- In applications, one uses pseudo-random numbers generated by a deterministic algorithm instead of truly random uniform sequences

Quantile transformation

If π is one-dimensional and we can compute the quantile function

$$F_\pi^{-1}(u) = \inf\{x; F_\pi(x) \geq u\}, \quad F_\pi(x) = \int_{-\infty}^{x} \pi(x')dx',$$

then $X^t = F_\pi^{-1}(U^t) \sim \pi$

- This is rarely feasible in Bayesian statistics

Illustration of quantile transformation

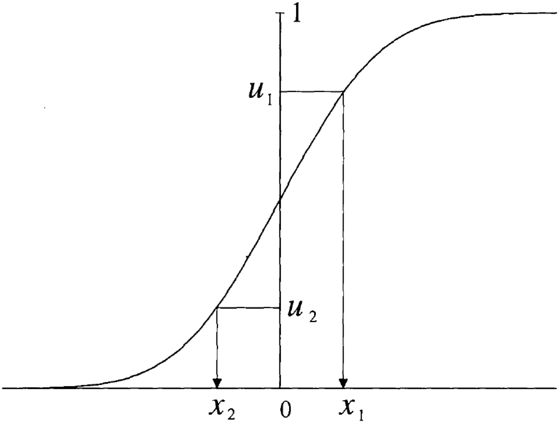

Figure: Quantile transformation for continuous F without flat sections.*

*Figure source: Glasserman (2003)

Illustration of quantile transformation

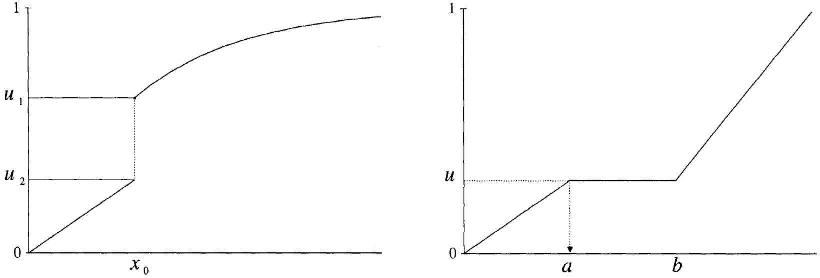

Figure: Quantile transformation for F with jumps or with flat sections.*

*Figure source: Glasserman (2003)