Computational Biology Group

D BSSE

ETH zürich

Introduction to Bayesian Statistics with R

3: Exercises

Jack Kuipers

28 November 2022

Exercise 3.1 - MCMC

For MCMC we can walk randomly and accept according the the MH ratio to eventually sample proportionally to any target distribution $p(x)$

```
# simple MCMC function
# n_its is the number of iterations
# start_x the initial position
# rw_sd is the sd of the Gaussian random walk
basicMCMC <- function(n_its = 1e3, start_x = 0, rw_sd = 1, ...) {
  xs <- rep(NA, n_its) # to store all the sampled values
  x <- start_x # starting point
  xs[1] <- x # first value
  p_x <- target_density(x, ...) # probability density at current value of x
  for (ii in 2:n_its) { # MCMC iterations
    x_prop <- x + rnorm(1, mean = 0, sd = rw_sd) # Gaussian random walk to propose next x
    p_x_prop <- target_density(x_prop, ...) # probability density at proposed x
    if (runif(1) < p_x_prop/p_x) { # MH acceptance probability
      x <- x_prop # accept move
      p_x <- p_x_prop # update density
    }
    xs[ii] <- x # store current position, even when move rejected
  }
  return(xs)
}
```

For example, if we want to sample from a Student-t distribution we can use the following target

```
target_density <- function(x, nu) {
  dt(x, nu) # student-t density
}
```

and run a short chain with $\nu = 5$

```
basicMCMC(nu = 5)
```

Examine the output MCMC chain for different lengths. How many samples would we need to get close to the Student-t distribution?

Use the samples to estimate (see description in Bonus Exercise 3.2)

$$\int \cos(t) f_5(t) \mathrm{d}t\,, \qquad f_\nu(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \rightarrow Student's\ t{-}distribution\ with\ \nu\ degree\ of\ freedom$$

Bonus Exercise 3.2 - Monte Carlo integration

NOTE: This exercise is an optional bonus for when you have sufficient free time.

Computing expectations can be applied to any continuous function

$$E[f(x)] = \int f(x)p(x)\mathrm{d}x$$

so that integrals where we recognise $p(x)$ as (proportional to) a probability distribution may be estimated with Monte Carlo methods since

$$E[f(x)] \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i)$$

for N random samples x_i sampled according to $p(x)$. Use samples from a Gaussian to estimate the following three integrals:

$$\int |x|\mathrm{e}^{-x^2}\mathrm{d}x\,, \qquad \int \sin(x)\mathrm{e}^{-x^2}\mathrm{d}x\,, \qquad \int \cos(x)\mathrm{e}^{-x^2}\mathrm{d}x$$

Reminder, the Gaussian has the following general form:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}\,.$$

Compare the estimated values to the exact values of the integrals. How does the accuracy depend on the sample size?

NOTE: For the comparison to the real values, you can integrate analytically or use `R`'s `integrate` function. The errors from the true value, like standard errors in general, decrease like $\frac{1}{\sqrt{N}}$.