MATH5835M Practical

Jochen Voss

2023-11-24

- The MATH5835M module is assessed by an examination (80%) and a practical (20%). This is the practical, worth 20% of your final module mark.
- You must hand in your solution via Gradescope by **Thursday, 14th December 2023, 2pm**.
- Reports must be typeset (not handwritten) and should be no more than 6 pages in length.
- Within reason you may talk to your friends about this piece of work, but you should not send R code (or output) to each other. Your report must be your own work.

Introduction

The distribution of wind speeds in a city follows a mixture of log-normal distributions: If we denote the wind speed (in miles per hour) by X, we have

$$\log(X) \sim \mathcal{N}(\mu_K, \sigma_K^2),$$

where K is random with

$$P(K = 1) = 0.1, \quad P(K = 2) = 0.4, \quad P(K = 3) = 0.5,$$

and the means and standard deviations are given by

$$\mu = \begin{pmatrix} 2.1 \\ 1.6 \\ -0.5 \end{pmatrix}, \quad \sigma = \begin{pmatrix} 0.55 \\ 0.7 \\ 0.6 \end{pmatrix}.$$

The density of the log-normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$ is

$$\varphi(x; \mu, \sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log(x) - \mu)^2}{2\sigma^2}\right)$$

for all $x \geq 0$. This function is available in R as `dlnorm()`. You can generate samples from the wind speed distribution as follows:

```
wind_speed <- function(n, w, mu, sigma) {
    k <- length(w)
    i <- sample(k, n, replace = TRUE, prob = w)
    rlnorm(n, mu[i], sigma[i])
}

w <- c(0.1, 0.4, 0.5)
mu <- c(2.1, 1.6, -0.5)
sigma <- c(0.55, 0.7, 0.6)
wind_speed(5, w, mu, sigma)
```

```
## [1]  7.266426  2.033405 12.911956  1.658656  8.007007
```

Task 1

We are interested in the probability that the wind speed is larger than 40 miles per hour.

- Use basic Monte Carlo estimation to estimate the probability that the wind speed is larger than 40 miles per hour.
- Determine the root mean squared error of your estimate.

Task 2

Our aim here is to find a more efficient estimator for the probability from task 1.

- Use importance sampling to estimate the probability that the wind speed is larger than 40 miles per hour.
- Compare the root mean squared error of your estimate with the one from Task 1.

Task 3

Now, for simplicity, assume that the wind speed follows a single log-normal distribution with parameters μ and σ:

$$\log(X) \sim \mathcal{N}(\mu, \sigma^2).$$

Furthermore, we assume that $\mu = 0$ is known and σ is unknown, random and exponentially distributed: $\sigma \sim \mathrm{Exp}(1)$.

On five different days we have observed the wind speeds $x_1 = 0.50$, $x_2 = 1.67$, $x_3 = 2.22$, $x_4 = 0.22$ and $x_5 = 4.36$. Using these data we want to learn about the unknown value σ. From Bayes rule we know that the conditional distribution of σ, given the data $(x_1, \ldots, x_5)$ has density

$$f(s) = \frac{1}{Z}\tilde{f}(s), \quad \tilde{f}(s) = \pi(s) \prod_{i=1}^{5} \varphi(x_i; 0, s),$$

where Z is the normalising constant, $\pi(s)$ is the density of the Exp(1) distribution, and $\varphi(x_i; 0, s)$ is the density of the log-normal distribution with parameters $\mu = 0$ and $\sigma = s$. In Bayesian statistics, π is called the *prior density* and f is called the *posterior density*.

We want to sample from the posterior density f using envelope rejection sampling.

- Produce a plot of $\tilde{f}(s)$ as a function of s.
- Choose an appropriate density $g(s)$ for the proposals for the rejection sampling method. Explain why your choice is appropriate.
- Show a plot of $\tilde{f}(s)/g(s)$, and from this determine an appropriate value for the constant c in the envelope rejection sampling algorithm.
- Implement the envelope rejection sampling algorithm and generate $N = 10,000$ samples from the posterior density f. Use these samples to plot a histogram.
- Compare the posterior distribution to the prior distribution, and comment on the differences you see.