Coursework Template

MSc in Statistics 2025/26, Imperial College London

Andrew Duncan, Oliver Ratmann

This Rmarkdown document is intended to provide a basic template that you may use throughout the course, and to introduce you to basic functionality of the open-source bookdown R package for writing Rmarkdown documents (Xie 2016).

Please note:

- Rmarkdown syntax for the title and document style;
- bookdown syntax to label and reference equations;
- bookdown syntax to referencing literate and provide a bib file;
- kable syntax to build clean and clear tables that can be referenced;
- Rmarkdown inline functions to collect and print all code used.

Sampling from the Cauchy distribution

In this question we seek to generate samples from the standard Cauchy distribution, with pdf given by

$$p(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}.(\#eq:1) \tag{1}$$

The cdf associated with @ref(eq:1) is given by

$$F(x) = \frac{1}{\pi}\arctan(x) + \frac{1}{2}.(\#eq:2) \tag{2}$$

Part a. We apply the inversion method as described by Graham and Talay (2013). We first compute the generalised inverse of the cdf @ref(eq:2). To this end, let $y \in [0, 1]$, then

$$
\begin{aligned}
y &= F(x), \quad \iff (\#eq:3a) && (3)\\
y &= \frac{1}{\pi}\arctan(x) + \frac{1}{2} \quad \iff && (4)\\
y - \frac{1}{2} &= \frac{1}{\pi}\arctan(x) \quad \iff && (5)\\
\pi\left(y - \frac{1}{2}\right) &= \arctan(x) \quad \iff && (6)\\
x &= \tan\left(\pi\left(y - \frac{1}{2}\right)\right). && (7)
\end{aligned}
$$

By the inversion method (Graham and Talay 2013), if $Y \sim U[0, 1]$, then

$$X = \tan\left(\pi\left(Y - \frac{1}{2}\right)\right), (\#eq:4) \tag{8}$$

Part b. The empirical cdf based on N samples $x_1, \ldots, x_N$ is defined to be

$$\widehat{F}(t) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}[x_i \leq t], \quad x \in \mathbb{R}.$$

To generate the empirical cdf in R, we perform the following steps:

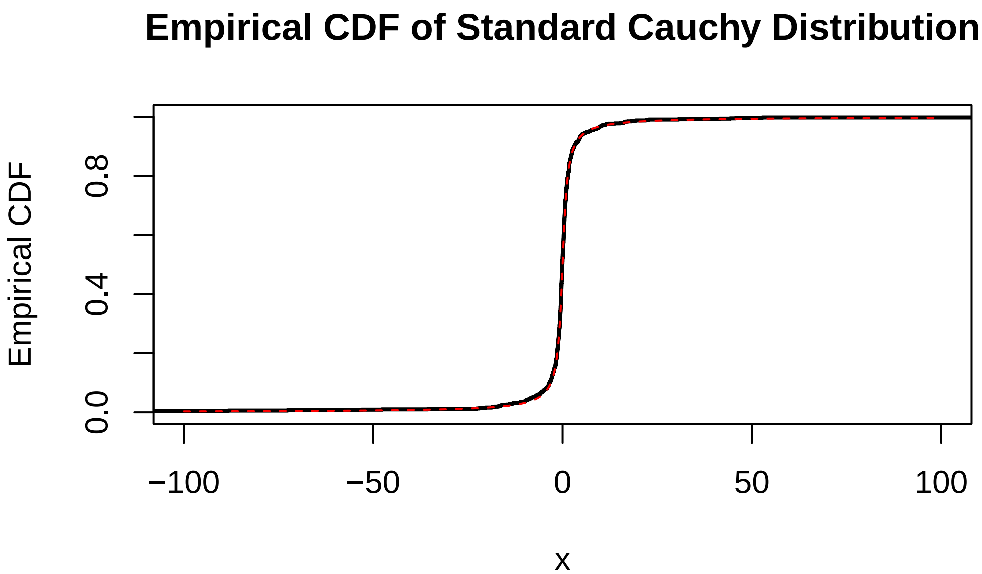
a. Generate N samples from the distribution using the inversion method, as detailed in the previous question.

b. Sort the samples in ascending order $x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_N$.
c. The empirical CDF at value x_i is given by i/N.

In the following figure we plot the empirical cdf based on 1000 samples (black line), and compare against the analytical cdf (red line).

Table 1: Empirical CDF

x	ecdf
0.5	0.65
1.0	0.75
10.0	0.97



Part c. We shall use the samples to compute the empirical cdf for the values $x = 0.5, 1, 10$.

Code appendix

Rather than re-paste all the code to the appendix, here is a trick which makes the markdown file output all the code (without) execution in the appendix, without any duplication.

Please keep in mind to format the code so that the entire code is clearly visible and does not run into the margins of the pdf version.

```
knitr::opts_chunk$set(
  collapse = TRUE,
  comment = "#>"
)
include_solutions <- TRUE
require(rmarkdown)
require(knitr)
require(kableExtra)
```

```r
# Put any library imports and other preamble here.

### Question 1a

generate_cauchy_samples <- function(n) {
  u <- runif(n)  # Generate 'n' uniform random numbers between 0 and 1
  samples <- tan(pi * (u - 0.5))  # Apply the inverse CDF formula to get Cauchy samples
  return(samples)
}

num_samples = 1000

# Generate samples from the standard Cauchy distribution
cauchy_samples <- generate_cauchy_samples(num_samples)


### Question 1b

# Sort the samples in ascending order
sorted_samples <- sort(cauchy_samples)

# Calculate the cumulative probabilities for each sample
cumulative_probs <- (1:num_samples) / num_samples

x_values <- seq(-100, 100,  length.out = 1000)
cdf_values <- pcauchy(x_values)

# Plot the empirical CDF
plot(sorted_samples,
     cumulative_probs,
     type = "s",
     xlab = "x",
     ylab = "Empirical CDF",
     main = "Empirical CDF of Standard Cauchy Distribution",
     lwd = 2, lty = 1, xlim = c(-100,100))
lines(x_values, cdf_values, type = "l",
      xlab = "x", ylab = "CDF", main = "CDF of Standard Cauchy Distribution",
      col = 'red', lty = 2)

### Question 1c

calculate_ecdf <- function(x) {
```

```r
    return(sum(sorted_samples <= x)/num_samples)
}

df <- data.frame(x = c(0.5, 1.0, 10.0),
                 ecdf = c(calculate_ecdf(0.5),
                          calculate_ecdf(1.0),
                          calculate_ecdf(10.0))
                 )
kbl(df,
    caption = 'Empirical CDF',
    label = 'ecdf',
    digits = 2) %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"), font_size = 12)
```

References

Graham, Carl, and Denis Talay. 2013. *Stochastic Simulation and Monte Carlo Methods: Mathematical Foundations of Stochastic Simulation*. Vol. 68. Springer Science & Business Media.

Xie, Yihui. 2016. *Bookdown: Authoring Books and Technical Documents with r Markdown*. CRC Press.