Exercise 1

TMA4300 Computer Intensive Statistical Models

Mads Adrian Simonsen, William Scott Grundeland Olsen

05 februar, 2021

Problem A: Stochastic simulation by the probability integral transform and bivariate techniques

1.

Let $X \sim \text{Exp}(\lambda)$, with the cdf

$$F_X(x) = 1 - e^{-\lambda x}, \quad x \geq 0.$$

Then the random variable $Y := F_X(X)$ has a $\text{Uniform}(0, 1)$ distribution. The probability integral transform becomes

$$Y = 1 - e^{-\lambda X} \Leftrightarrow X = -\frac{1}{\lambda}\log(1 - Y). \tag{1}$$

Thus, we sample Y from `runif()` and transform it using (1), to sample from the exponential distribution. Figure 1 shows one million samples drawn from the `generate_from_exp()` function defined in the code chunk below.

```
set.seed(123)

generate_from_exp <- function(n, rate = 1) {
  Y <- runif(n)
  X <- -(1 / rate) * log(1 - Y)
  X
}

# sample
n <- 1000000  # One million samples
lambda <- 4.32
x <- generate_from_exp(n, rate = lambda)

# plot
hist(x,
  breaks      = 80,
  probability = TRUE,
  xlim        = c(0, 2)
)
curve(dexp(x, rate = lambda),
  add = TRUE,
  lwd = 2,
```

```
  col = "red"
)
```

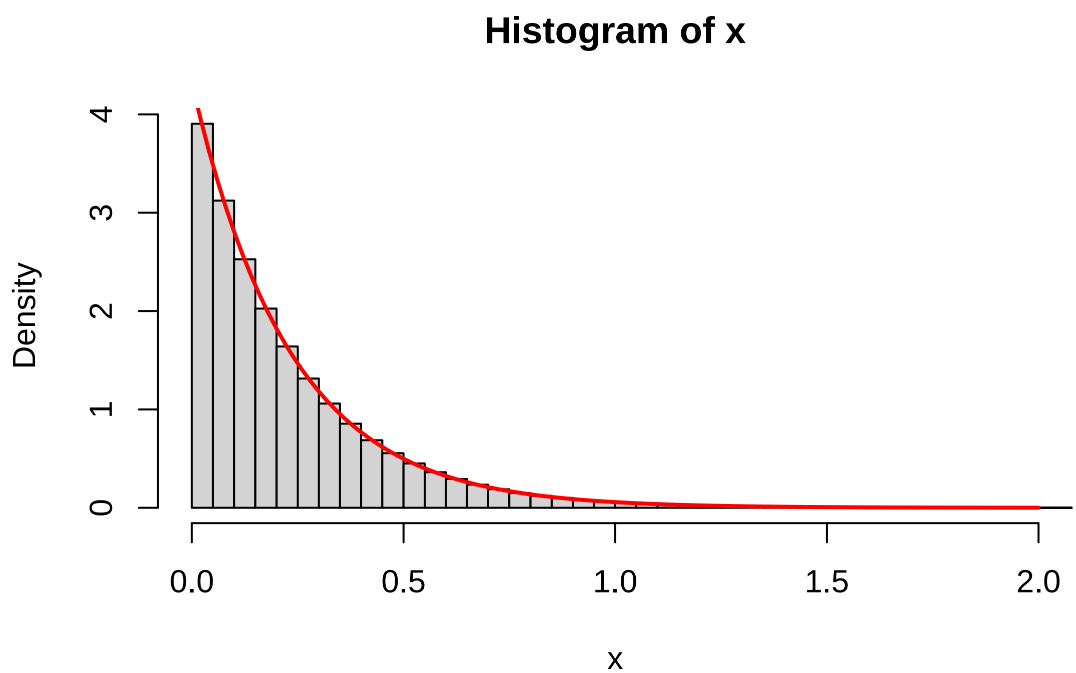

Figure 1: Normalized histogram of one million samples drawn from the exponential distribution, together with the theoretical pdf, with $\lambda = 4.32$.

2.

(a)

(b)

3.

(a)

(b)

(c)

4.

5

Problem B: The gamma distribution

1.

(a)

(b)

2.

(a)

(b)

3.

(a)

(b)

4.

5.

(a)

(b)

Problem C: Monte Carlo integration and variance reduction

1.

Let $X \sim \mathrm{N}(0,1)$, and $\theta = \Pr(X > 4)$. Let also $h(x) = I(x > 4)$, where $I(\cdot)$ is the indicator function. Then

$$\mathrm{E}[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x)\, dx = \int_{-\infty}^{\infty} I(x > 4) f_X(x)\, dx = \Pr(X > 4) = \theta.$$

Let $X_1, \dots X_n \sim \mathrm{N}(0,1)$ be a sample. Then the simple Monte Carlo estimator of θ is

$$\hat{\theta}_{\mathrm{MC}} = \frac{1}{n}\sum_{i=1}^{n} h(X_i),$$

with expectation

$$\mathrm{E}\left[\hat{\theta}_{\mathrm{MC}}\right] = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[h(X_i)\right] = \frac{1}{n}\sum_{i=1}^{n} \theta = \theta,$$

and sampling variance

$$\widehat{\mathrm{Var}}\left[\hat{\theta}_{\mathrm{MC}}\right] = \frac{1}{n^2}\sum_{i=1}^{n}\widehat{\mathrm{Var}}\left[h(X_i)\right] = \frac{1}{n}\widehat{\mathrm{Var}}[h(X)] = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(h(X_i) - \hat{\theta}_{MC}\right)^2.$$

Then the statistic

$$T = \frac{\hat{\theta}_{\mathrm{MC}} - \theta}{\sqrt{\widehat{\mathrm{Var}}\left[\hat{\theta}_{\mathrm{MC}}\right]}} \sim \mathrm{t}_{n-1},$$

and $t_{\alpha/2,\,n-1} = F_T^{-1}(1-\alpha/2)$, where $F_T^{-1}(\cdot)$ is the quantile function of the t_{n-1} distribution.

```r
#remove this-------------------------------------------------------------------------------
generate_from_exp <- function(n, rate = 1) {
  Y <- runif(n)
  X <- -(1 / rate) * log(Y)
  return(X)
}
std_normal <- function(n) {
  X1 <- pi * runif(n)   # n samples from Uniform(0, pi)
  X2 <- generate_from_exp(n, 1/2)   # n samples from Exponential(1/2)
  Z <- X2^(1/2) * cos(X1)   # Z ~ Normal(0, 1)
  return(Z)
}
#remove this-------------------------------------------------------------------------------

set.seed(321)
n <- 100000
x <- std_normal(n)
h <- function(x) {
  1 * (x > 4)
}

theta_MC <- (1 / n) * sum(h(x))  # Monte Carlo estimate of Pr(X > 4)

sampl_var_theta_MC <- sum((h(x) - theta_MC)^2) / (n * (n - 1))  # Sampling variance

t <- qt(0.05/2, df = n - 1, lower.tail = FALSE)  # quantile with 5% significance level
ci_MC <- theta_MC + t * sqrt(sampl_var_theta_MC) * c(-1, 1)  # Confidence Interval

list(theta_MC = theta_MC, confint = ci_MC)
```

```
## $theta_MC
## [1] 4e-05
##
## $confint
## [1] 8.008339e-07 7.919917e-05
```

2.

3.

(a)

(b)

Problem D: Rejection sampling and importance sampling

1.

2.

3.

4.