# Exercise 1

TMA4300 Computer Intensive Statistical Models

Mads Adrian Simonsen, William Scott Grundeland Olsen

29 januar, 2021

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

### Subproblem 1.

Let $X \sim \text{Exponential}(\lambda)$, with the cumulative density function

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

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

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

It is clear that if $U \sim \text{Uniform}(0,1)$, then $1 - U \sim \text{Uniform}(0,1)$, and therefore we may as well say that

$$X = -\frac{1}{\lambda}\ln(Y). \tag{1}$$

Thus, we sample $Y$ from `runif()` and transform it using Equation (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. It also shows the theoretical PDF of the exponential distribution with rate parameter $\lambda = 2$.

```r
#set.seed(123)

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

# sample
n <- 1000000  # One million samples
lambda <- 2
exp_samp <- generate_from_exp(n, rate = lambda)

# plot
ggplot() +
  geom_histogram(
```

```
    data = as.data.frame(exp_samp),
    mapping = aes(x = exp_samp, y = ..density..),
    binwidth = 0.05,
    boundary = 0
  ) +
  stat_function(
    fun = dexp,
    args = list(rate = lambda),
    aes(col = "Theoretical density")
  ) +
  ylim(0, lambda) +
  xlim(0, 2) +
  ggtitle("Dimulating from an exponential distribution") +
  xlab("x") +
  ylab("Density") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```

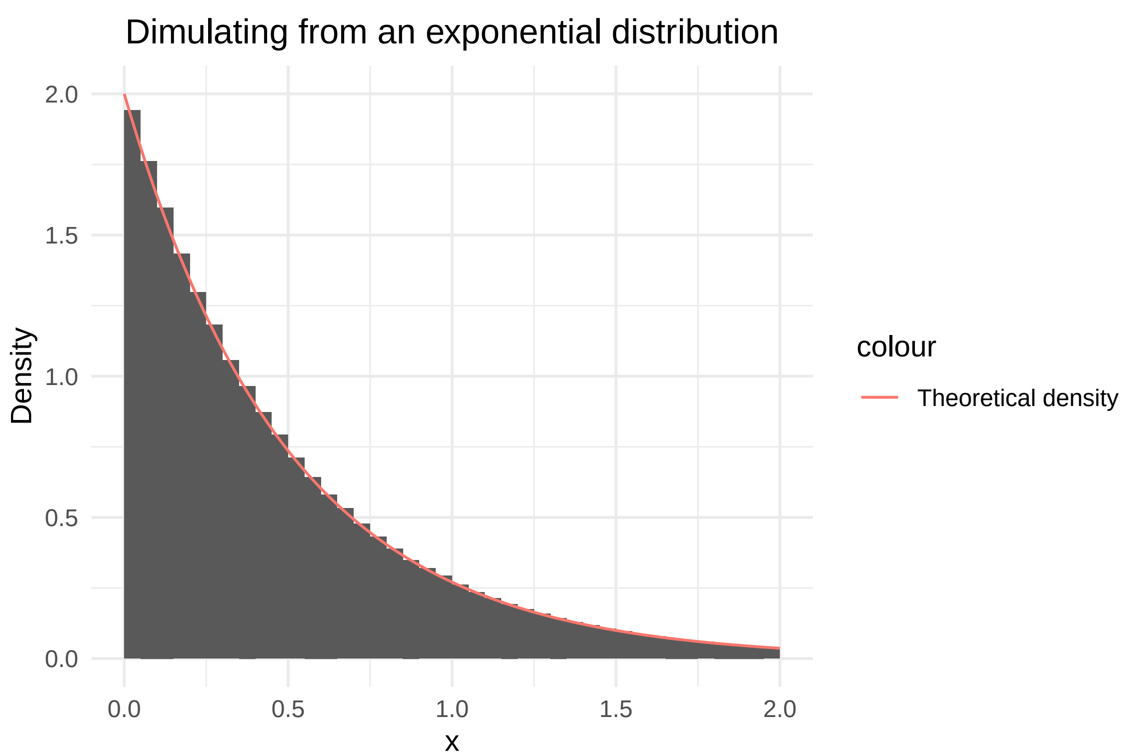


Figure 1: Normalized histogram of one million samples drawn from the exponential distribution, together with the theoretical PDF, with $\lambda = 2$.

Theoretically, the mean and variance of $X \sim \text{Exponential}(\lambda)$ is $\text{E}(X) = \lambda^{-1}$ and $\text{Var}(X) = \lambda^{-2}$. So for $\lambda = 2$ we would expect $\text{E}(X) = 1/2$ and $\text{Var}(X) = 1/4$. For the simulation we get the mean and variance as calculated in the code block below, showing what we would expect.

```
mean(exp_samp)
```

```
## [1] 0.5001832
```

```
var(exp_samp)
```

```
## [1] 0.2503854
```

## Subproblem 2.

### Subsubproblem (a)

We are considering the probability density function

$$g(x) = \begin{cases} cx^{\alpha-1} & \text{if } 0 < x < 1, \\ c\mathrm{e}^{-x} & \text{if } x \geq 1, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where $c$ is a normalizing constant and $\alpha \in (0, 1)$. If $x \leq 0$ the cumulative distribution function is zero. In the interval $0 < x < 1$ it becomes

$$G(x) = \int_{-\infty}^{x} g(\xi)\,\mathrm{d}\xi = \int_{0}^{x} c\xi^{\alpha-1}\,\mathrm{d}\xi = \frac{c}{\alpha}[\xi^{\alpha}]_0^x = \frac{c}{\alpha}x^{\alpha},$$

and finally for $x \geq 1$ we have

$$G(x) = \int_{-\infty}^{x} g(\xi)\,\mathrm{d}\xi = \int_{0}^{1} c\xi^{\alpha-1}\,\mathrm{d}\xi + \int_{1}^{x} c\mathrm{e}^{-\xi}\,\mathrm{d}\xi = \left[\frac{c}{\alpha}\xi^{\alpha}\right]_0^1 - \left[c\mathrm{e}^{-\xi}\right]_1^x = c\left(\frac{1}{\alpha} - \mathrm{e}^{-x} + \frac{1}{\mathrm{e}}\right),$$

for $\alpha \in (0, 1)$. That is, the cumulative density function is

$$G(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ \frac{c}{\alpha}x^{\alpha} & \text{if } 0 < x < 1, \\ c\left(\frac{1}{\alpha} - \mathrm{e}^{-x} + \frac{1}{\mathrm{e}}\right) & \text{if } x \geq 1. \end{cases}$$

In this case it is trivial to find $c$. We solve

$$1 = \int_{\mathbb{R}} g(x)\,\mathrm{d}x = \int_{0}^{1} cx^{\alpha-1}\,\mathrm{d}x + \int_{1}^{\infty} c\mathrm{e}^{-x}\,\mathrm{d}x = \frac{c}{\alpha} + \frac{c}{\mathrm{e}},$$

which gives that

$$c = \frac{\alpha\mathrm{e}}{\alpha + \mathrm{e}}.$$

Writing the cumulative density function using this as $c$ we obtain

$$G(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ \frac{\mathrm{e}}{\alpha+\mathrm{e}}x^{\alpha} & \text{if } 0 < x < 1, \\ 1 - \frac{\alpha}{\alpha+\mathrm{e}}\mathrm{e}^{1-x} & \text{if } x \geq 1, \end{cases}$$

for $\alpha \in (0, 1)$.

We may then find the inverse cumulative function. For $x \leq 0$ this is just zero, and for $0 < x < 1$, that is $0 < G(x) < \frac{\mathrm{e}}{\alpha+\mathrm{e}}$, we solve $x = \frac{\mathrm{e}}{\alpha+\mathrm{e}}y^{\alpha}$ for $y$ giving $G^{-1}(x) = \left(\frac{\alpha+\mathrm{e}}{\mathrm{e}}x\right)^{1/\alpha}$. Similarly for $x \geq 1$, that is $G(x) \geq 1 - \frac{\alpha}{\alpha+\mathrm{e}} = \frac{\mathrm{e}}{\alpha+\mathrm{e}}$, we solve $x = 1 - \frac{\alpha}{\alpha+\mathrm{e}}\mathrm{e}^{1-y}$ for $y$, such that

$$G^{-1}(x) = \begin{cases} \left(\frac{\alpha+\mathrm{e}}{\mathrm{e}}x\right)^{1/\alpha} & \text{if } 0 \leq x < \frac{\mathrm{e}}{\alpha+\mathrm{e}}, \\ \ln\left[\frac{\alpha\mathrm{e}}{(1-x)(\alpha+\mathrm{e})}\right] & \text{if } x \geq \frac{\mathrm{e}}{\alpha+\mathrm{e}}, \end{cases}$$

for $\alpha \in (0, 1)$.

### Subsubproblem (b)

```r
generate_from_piecewise <- function(n, alpha) {
  U <- runif(n)   # Generate n Uniform(0, 1) variables
  bound <- exp(1) / (alpha + exp(1))   # Boundary where G^(-1) changes
  left <- U < bound   # The left of the boundary
  U[left] <- (U[left] / bound)^(1 / alpha)   # Left CDF
  U[!left] <- 1 + log(alpha) - log(1 - U[!left]) - log(alpha + exp(1))   # Right CDF
  return(U)
}

# Sample
n <- 1000000  # One million samples
alpha <- 0.75
x <- generate_from_piecewise(n, alpha)

# The theoretically correct PDF
theo_PDF <- function(x, alpha) {
  const <- alpha * exp(1) / (alpha + exp(1))    # Normalizing constant
  func <- rep(0, length(x))
  left <- x > 0 & x < 1   # The PDF has one value for 0 < x < 1
  right <- x >= 1         # ... and one value for x >= 1
  func[left] <- const * x[left]^(alpha - 1)   # The value to the left
  func[right] <- const * exp(-x[right])       # The value to the right
  return(func)
}

# Plot
ggplot() +
  geom_histogram(
    data = as.data.frame(x),
    mapping = aes(x = x, y = ..density..),
    binwidth = 0.05,
    boundary = 0
  ) +
  stat_function(
    fun = theo_PDF,
    args = list(alpha = alpha),
    aes(col = "Theoretical density")
  ) +
  xlim(0, 5) +
  ggtitle("Simulating from g(x) given in Equation (2)") +
  ylab("Density") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```

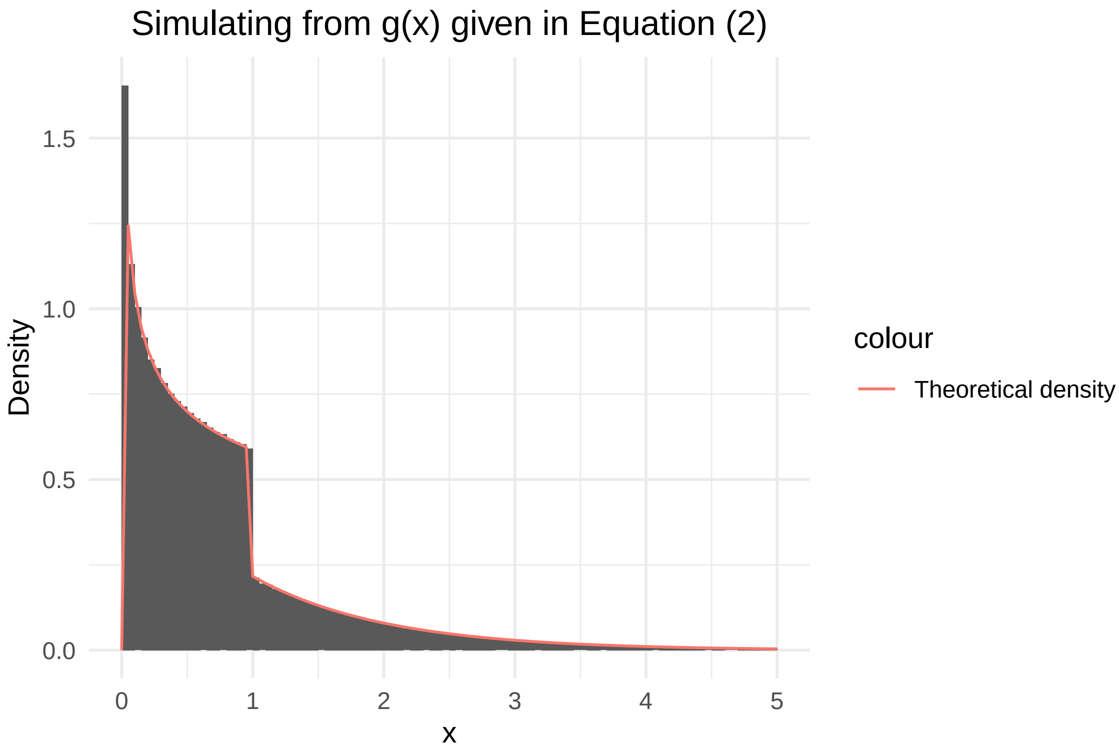


Figure 2: Normalized histogram of one million samples drawn from $g(x)$ given in Equation (2), together with the theoretical PDF, with $\alpha = 0.75$.

**3.**

**(a)**

**(b)**

**(c)**

**4.**

**5**

## Problem B: The gamma distribution

**1.**

**(a)**

**(b)**

**2.**

**(a)**

**(b)**

**3.**

**(a)**

**(b)**

**4.**



**5.**

**(a)**

**(b)**