Carey Expected Return R Project Test

Exercise 1

a)

We first write a `R` function that determines the value of the parameters μ and σ^2 from the expectation $\mathbb{E}\{X\}$ and the variance $\mathbb{V}\{X\}$ of a univariate lognormal random variable $X \sim LogN(\mu, \sigma^2)$:

```
ln_params <- function(exp, var) {
  sig_sq <- log(var / exp^2 + 1)
  mu <- log(exp) - sig_sq / 2
  params <- list("mu" = mu, "sig_sq" = sig_sq)
  return(params)
}
```

Note: We managed to derive these formulas by rearranging the expressions for the mean and variance of a log-normal r.v.

b)

Next, we use the function created in a) to determine the parameters μ and σ^2 such that $\mathbb{E}\{X\} = 3$ and $\mathbb{V}\{X\} = 5$:

```
params <- ln_params(3, 5)
ln_mu <- params$mu; ln_sig_sq <- params$sig_sq
ln_mu; ln_sig_sq
```

```
## [1] 0.8776959
```

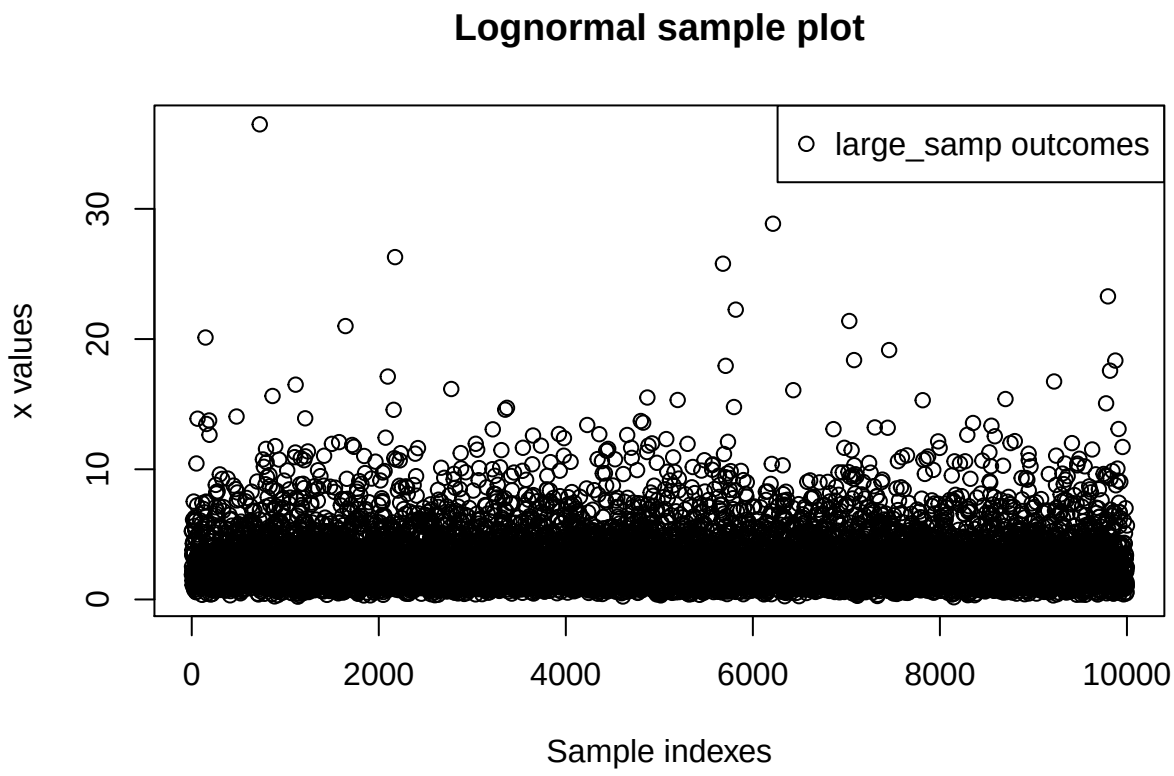
```
## [1] 0.4418328
```

Next, we generate a large sample (of size $n = 10,000$) from the distribution of X with such parameters:

```
n <- 10000
ln_sig <- sqrt(ln_sig_sq)
large_samp <- rlnorm(n, ln_mu, ln_sig)
```

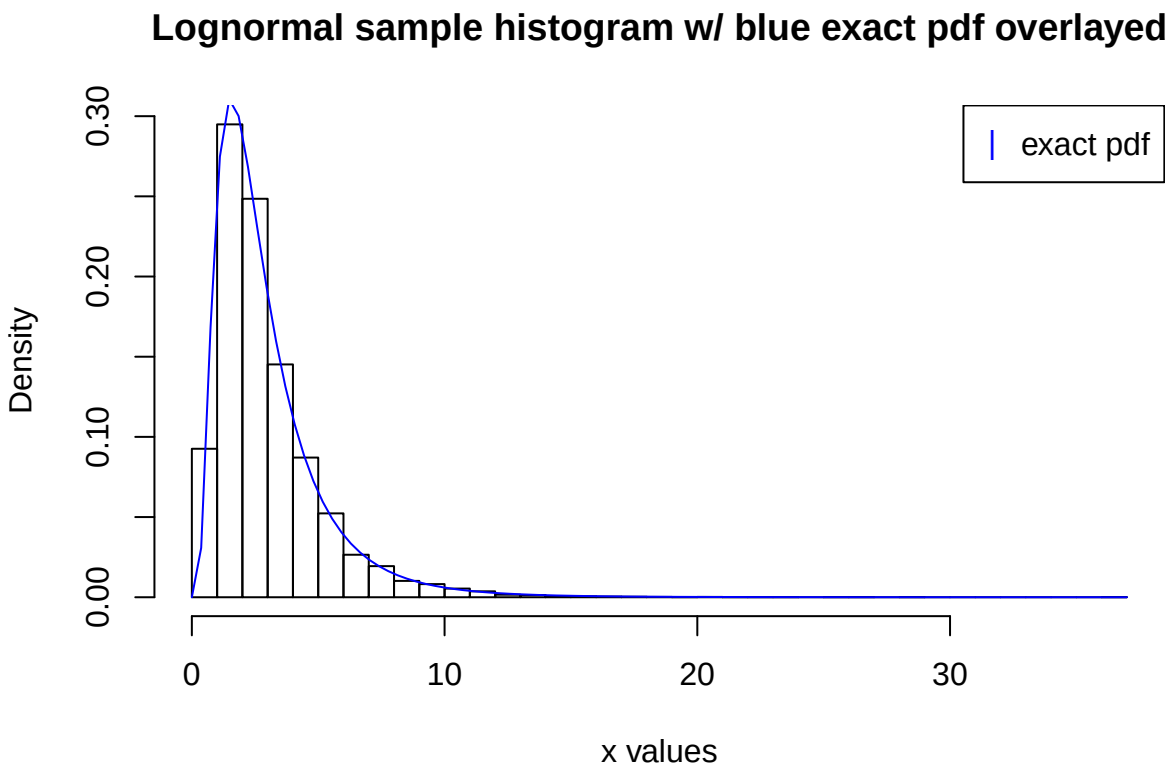
After that, we plot the sample as dots:

```
plot(large_samp, main = "Lognormal sample plot", xlab = "Sample indexes", ylab = "x values")
legend("topright", legend = "large_samp outcomes", pch = 1)
```



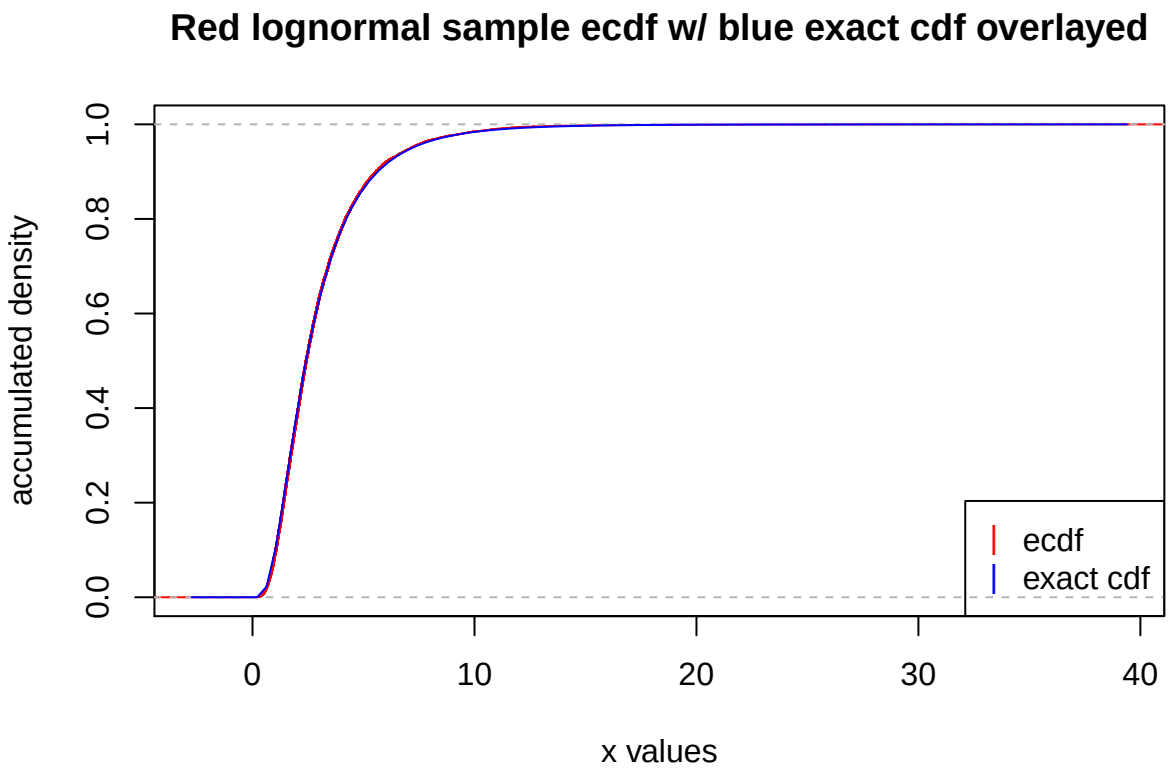
Next, we plot the histogram and overlay the the exact pdf for comparison:

```
hist(large_samp, 50, FALSE, main = "Lognormal sample histogram w/ blue exact pdf overlayed", xlab = "x
curve(dlnorm(x, meanlog = ln_mu, sdlog = ln_sig), add = TRUE, col = "blue")
legend("topright", legend = "exact pdf", col = "blue", pch = "|")
```



Finally, we plot the empirical cdf and overlay the exact cdf for comparison:

```
plot(ecdf(large_samp), main = "Red lognormal sample ecdf w/ blue exact cdf overlayed", xlab = "x values
curve(plnorm(x, meanlog = ln_mu, sdlog = ln_sig), add = TRUE, col = "blue")
legend("bottomright", legend = c("ecdf", "exact cdf"), col = c("red", "blue"), pch = "|")
```



Exercise 2

a)

First, we load in a library we will require for our simulation in b) and then proceed to write the function:

```r
# Load MASS library for use of mvrnorm function
library(MASS)

fit_locdisp_mlfp <- function(eps_t_mat, p_t_vec, nu, thresh) {
  # Extract length of vector and matrix, width of matrix
  t_bar <- length(p_t_vec)
  cols <- ncol(eps_t_mat)

  # 0. Initialize
  mu_vec <- vector("numeric", cols)
  for (t in c(1:t_bar)) {
    mu_vec <- mu_vec + p_t_vec[t] * t(eps_t_mat[t,])
  }
  sig_sq_hfp_mat <- matrix(rep(vector("numeric", cols), cols), cols, cols)
  for (t in c(1:t_bar)) {
    dif <- t(eps_t_mat[t,]) - mu_vec
    sig_sq_hfp_mat <- sig_sq_hfp_mat + p_t_vec[t] * t(dif) %*% dif
  }
  sig_sq_mat <- matrix(rep(vector("numeric", cols), cols), cols, cols)
```

```
  if (nu > 2) {
    sig_sq_mat <- ((nu - 2) / nu) * sig_sq_hfp_mat
  } else {
    sig_sq_mat <- sig_sq_hfp_mat
  }

  # Construct norm values to be updated after each run to check for convergence
  rel_euc_norm <- thresh * 2
  rel_fro_norm <- thresh * 2

  while ((rel_euc_norm > thresh) & (rel_fro_norm > thresh)) {
    # 1. Update weights and FP
    w_t_vec <- vector("numeric", t_bar)
    for (t in c(1:t_bar)) {
      dif <- t(eps_t_mat[t,]) - mu_vec
      w_t_vec[t] <- (nu + t_bar) / (nu + (dif %*% t(sig_sq_mat) %*% t(dif)))
    }
    sum_prod_p_w <- sum(p_t_vec * w_t_vec)
    q_t_vec <- vector("numeric", t_bar)
    for (t in c(1:t_bar)) {
      q_t_vec[t] <- p_t_vec[t] * w_t_vec[t] / sum_prod_p_w
    }

    # 2. Update output
    mu_vec_old <- mu_vec
    for (t in c(1:t_bar)) {
      mu_vec <- mu_vec + q_t_vec[t] * t(eps_t_mat[t,])
    }
    sig_sq_mat_old <- sig_sq_mat
    for (t in c(1:t_bar)) {
      dif <- t(eps_t_mat[t,]) - mu_vec
      sig_sq_hfp_mat <- sig_sq_hfp_mat + q_t_vec[t] * t(dif) %*% dif
    }

    # Update norm values to check for convergence; return to 1) if no convergence
    rel_euc_norm <- (norm(mu_vec - mu_vec_old)) / norm(mu_vec_old)
    rel_fro_norm <- (det(sig_sq_mat - sig_sq_mat_old)) / det(sig_sq_mat_old)
  }
  # 3. If convergence, output (mu_vec, sig_sq_vec)
  params <- list("mu_vec" = mu_vec, "sig_sq_mat" = sig_sq_mat)
  return(params)
}
```

b)

Next, we test the function:

```
num_runs <- 1000
df <- 100
conv_thresh <- 10^(-9)
num_runs_inv <- 1 / num_runs
p_t_bvn_vec <- rep(num_runs_inv, num_runs)
mu_input <- c(0, 0)
```

```r
Sigma_input <- cbind(c(1, 0), c(0, 1))
eps_t_bvn_mat <- mvrnorm(num_runs, mu_input, Sigma_input)
fit_locdisp_mlfp(eps_t_bvn_mat, p_t_bvn_vec, df, conv_thresh)
```

```
## $mu_vec
##             [,1]        [,2]
## [1,] -0.05412145 0.007199981
##
## $sig_sq_mat
##              [,1]         [,2]
## [1,]  0.934986422 -0.002462033
## [2,] -0.002462033  0.869252505
```

As these maximum likelihood estimates very closely resemble the vector of means and covariance matrix parameters we passed into the mvnorm function to perform our 1000 simulations from the standard bivariate normal distribution, this test provides evidence of the functionality of our fit_locdisp_mlfp function.