Midterm Project

Holden Ellis, Charlotte Huang

2026-02-08

Write Up

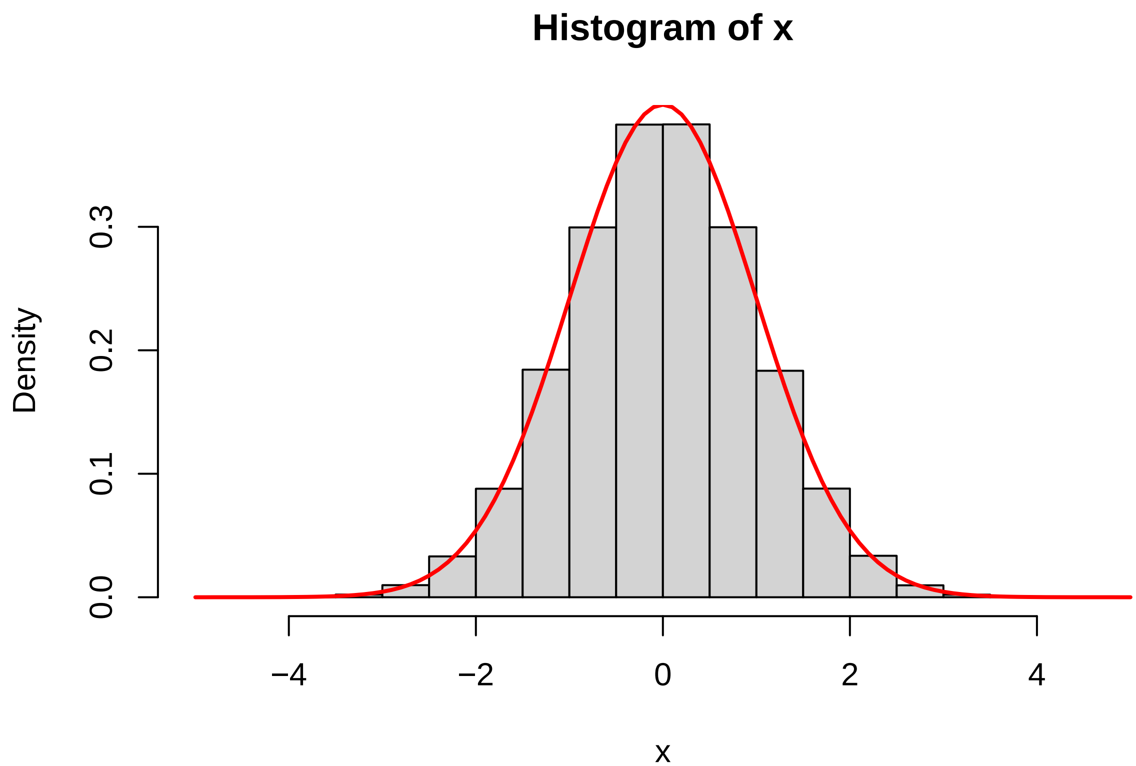
https://docs.google.com/document/d/1HZTEfuxbhKEsMSGoFqDvulxrU5hy9rIVp31Wf9el38g/edit?usp=sharing

```
source("Ziggurat.R")
source("BoxMuller.R") # if anyone wants to compare, much faster because of how R runs code
```

```
zigtable(function(x) { exp(-x) }, function(y) { -log(y) },8)
```

```
## $x
## [1] 0.0000000 0.4338930 0.7888599 1.1387148 1.5163812 1.9560291 2.5166676
## [8] 3.3489677
##
## $v
## [1] 0.1527383
```

```
x <- rnormzig(1000000)
hist(x, freq=FALSE)
curve(dnorm(x), add=TRUE, col="red", lwd=2) # looks good
```




```r
set.seed(777)

#testing: alpha = 0.01
samples1 <- rnormzig(4999)
print("4999 Samples")
```

```
## [1] "4999 Samples"
```

```r
shapiro.test(samples1) # no evidence against normal; p-value > alpha = 0.01
```

```
##
##  Shapiro-Wilk normality test
##
## data:  samples1
## W = 0.99954, p-value = 0.2852
```

```r
ks.test(samples1, "pnorm", mean=mean(samples1), sd=sd(samples1)) # same
```

```
##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  samples1
## D = 0.0083523, p-value = 0.8766
## alternative hypothesis: two-sided
```

```r
print("1,000,000 ziggurat samples")
```

```
## [1] "1,000,000 ziggurat samples"
```

```r
# testing 1 mil samples...
ks.test(x, "pnorm", mean=mean(x), sd=sd(x)) # rejected normality
```

```
## Warning in ks.test.default(x, "pnorm", mean = mean(x), sd = sd(x)): ties should
## not be present for the one-sample Kolmogorov-Smirnov test

##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  x
## D = 0.00046895, p-value = 0.9804
## alternative hypothesis: two-sided
```

```r
mean(x) # very close to 0
```

```
## [1] -0.0002673609
```

```r
sd(x)   # should be closer to 1
```

```
## [1] 1.000134
```

```r
print("1,000,000 rnorm samples")
```

```
## [1] "1,000,000 rnorm samples"
```

```r
# additional test:
x2 <- rnorm(1000000)
ks.test(x2, "pnorm", mean=mean(x2), sd=sd(x2)) # passed
```

```
##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  x2
## D = 0.00069352, p-value = 0.722
## alternative hypothesis: two-sided
```

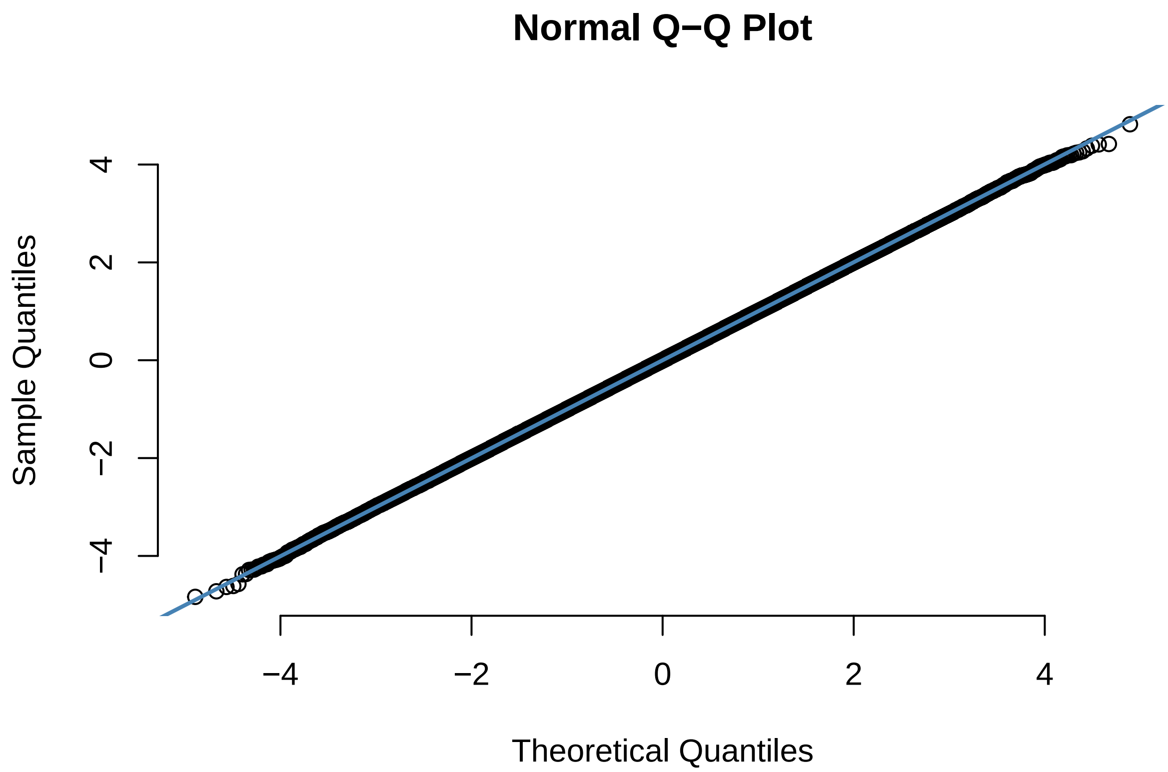
```r
mean(x2)  # basically 0
```

```
## [1] 0.000669734
```

```r
sd(x2)  # 1
```

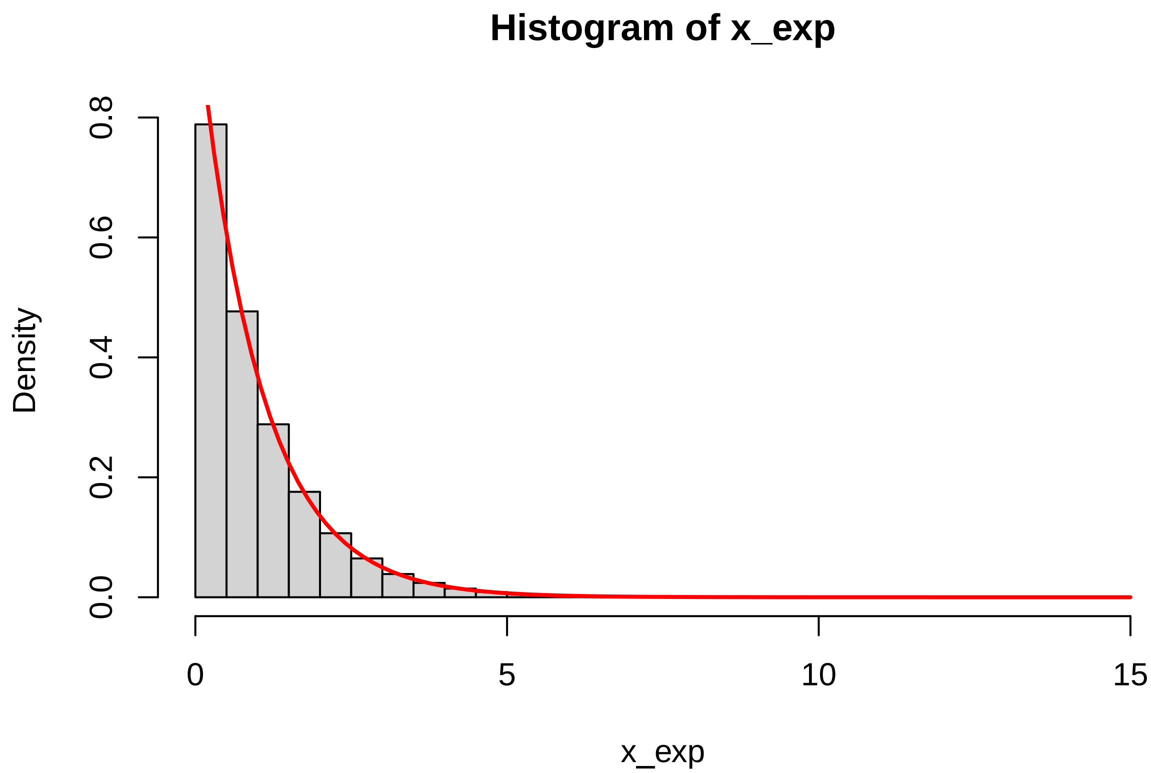
```
## [1] 1.001078
```

```r
# QQ plot
qqnorm(x, pch = 1, frame = FALSE)
qqline(x, col = "steelblue", lwd = 2) # basically normal
```

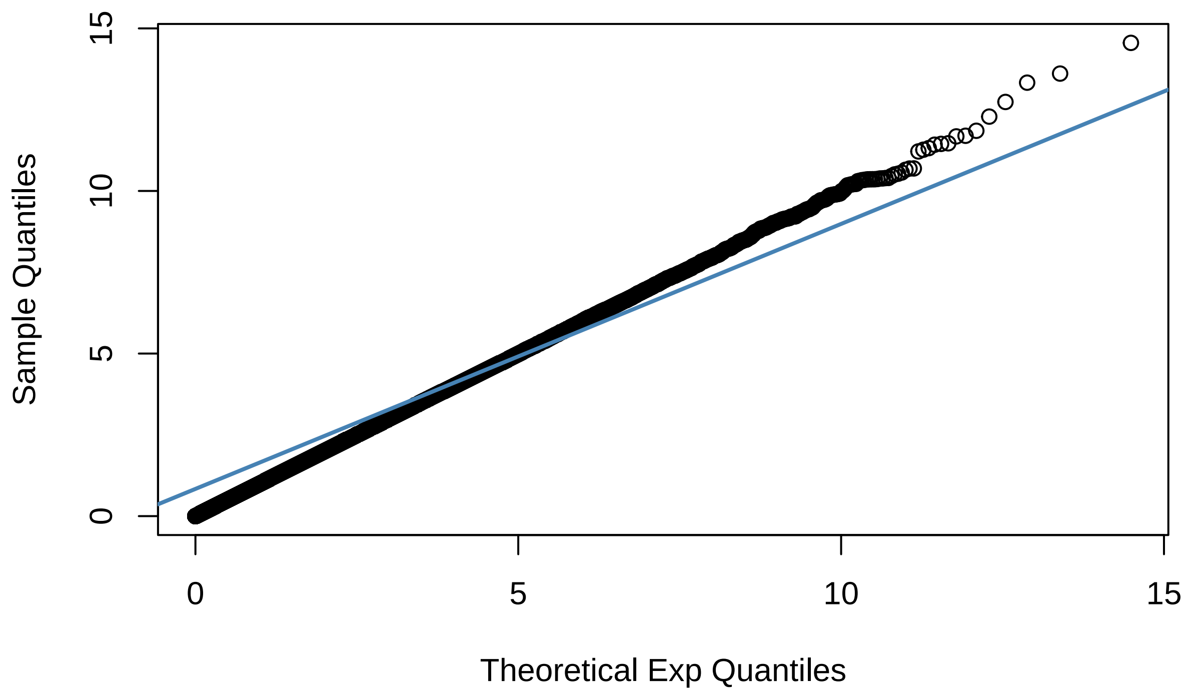


Assume we generate 100000 with Ziggurat method and they're actually normal, using rnorm() for now

```
set.seed(777)
x_exp <- rexpzig(1000000)
hist(x_exp, freq = FALSE)
curve(dexp(x), add=TRUE, col="red", lwd=2) # looks good
```




```r
qqplot(qexp(ppoints(length(x_exp)), rate = 1/mean(x_exp)),
       sort(x_exp),
       xlab = "Theoretical Exp Quantiles",
       ylab = "Sample Quantiles")
qqline(x_exp, col = "steelblue", lwd = 2) # basically normal
```




```r
print("1,000,000 Ziggurat samples")
```

```
## [1] "1,000,000 Ziggurat samples"
```

```r
mean(x_exp) #Should be close to 1
```

```
## [1] 0.9985969
```

```r
sd(x_exp) #Should be close to 1
```

```
## [1] 0.9979412
```

```r
ks.test(x_exp, "pexp", rate = mean(x_exp)) #large p-value -> fail to reject expoential distribution
```

```
##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  x_exp
## D = 0.0014393, p-value = 0.03175
## alternative hypothesis: two-sided
```

```r
print("1,000,000 rnorm samples")
```

```
## [1] "1,000,000 rnorm samples"
```

```r
# additional test:
x2_exp <- rexp(1000000)
ks.test(x2_exp, "pexp", rate = mean(x2_exp)) # passed
```

```
## Warning in ks.test.default(x2_exp, "pexp", rate = mean(x2_exp)): ties should
```

```
## not be present for the one-sample Kolmogorov-Smirnov test
##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  x2_exp
## D = 0.00096868, p-value = 0.3051
## alternative hypothesis: two-sided
```

```r
mean(x2_exp)  # basically 1
```

```
## [1] 1.000543
```

```r
sd(x2_exp)  # 1
```

```
## [1] 0.9995342
```

A chi-squared goodness of fit test (Knuth 1997) is used to check the distribution of the random variables generated. This test quantizes the horizontal axis of the probability density function into k bins and derives a single value as a quality metric from the determined actual and expected number of samples in each bin.

$\chi^2_{k-1} = \sum_{i=1}^{k} \frac{(Y_i - tp_i)^2}{tp_i}$

Where t = number of observation, p_i = probability that each observation falls into the the category i Y_i = the number of observation that actually do fall into the category i.

```r
set.seed(999)

n = 100000
samples <- rnorm(n)

chi_squared_test <- function(x, k){
  t <- length(x) # number of observations
  bins <- seq(min(x), 7, length.out = k+1) #200 bins spaced uniformly over data edge of min and max

  Yi <- hist(x, breaks = bins, plot = FALSE)$counts # observed counts
  # expected probabilites (p_i) of standard normal
  p_i <- numeric(k)
  for (i in 1:k) {
    p_i[i] <- pnorm(bins[i+1]) - pnorm(bins[i])
}
  # using computing chi square using equation from the paper
  chi_sq <- sum((Yi - t * p_i)^2 / (t * p_i))
  df <- k - 1
  return(data.frame(chi_square = chi_sq,
                    df = df))
}
```

```r
n = 100000
x <- rnorm(n)
k <- 200
results <- chi_squared_test(x, k)
```

```r
kable(results, caption = "Chi-squared value Ziggarut Method")
```

Table 1: Chi-squared value Ziggarut Method

chi_square	df
156.1863	199

The critical value for a chi-square test with 199 degrees of freedom at 95% confidence level ($\alpha = 0.05$) is 232.912. (using a look up table on https://www.medcalc.org/en/manual/chi-square-table.php)