Z-Squared CLT Demonstration

Michael Shehan

March 31, 2023

Given $Z_i \sim N(0,1)$ iid for $i = 1, 2, ..., n$,

$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2_{(n)}$$

We will demonstrate this property in two directions:

1) The sum of n Z^2 random variables should resemble the χ^2 distribution with n degrees of freedom. As degrees of freedom increase, we expect the distributions to approach $\sim N(\mu, \sigma^2)$
2) By Central Limit Theorem, the sum of n Z^2 and/or χ^2 random variables should approach N(0,1) distribution with large enough n when standardized.

Generating Z^2 Random Variables

We generate the simulation using 50 repetitions for each of 50 Z^2 random variables, $Z_1^2, Z_2^2, ..., Z_{50}^2$, where $Z_1, Z_2...Z_n$ iid $\sim N(0,1)$

```
# Create 50x50 matrix "z_sqs"
# let each column be a random variable Z^2(1), Z^2(2),... Z^2(50)
# with each row containing a z^2 value generated from rnorm, n = 50.

set.seed(777)

z_sqs <- matrix(nrow = 50, ncol = 50)

for(i in 1:50){
  z_sqs[,i] <- rnorm(n = 50, mean = 0, sd = 1)^2
}
```

Histogram Demonstration

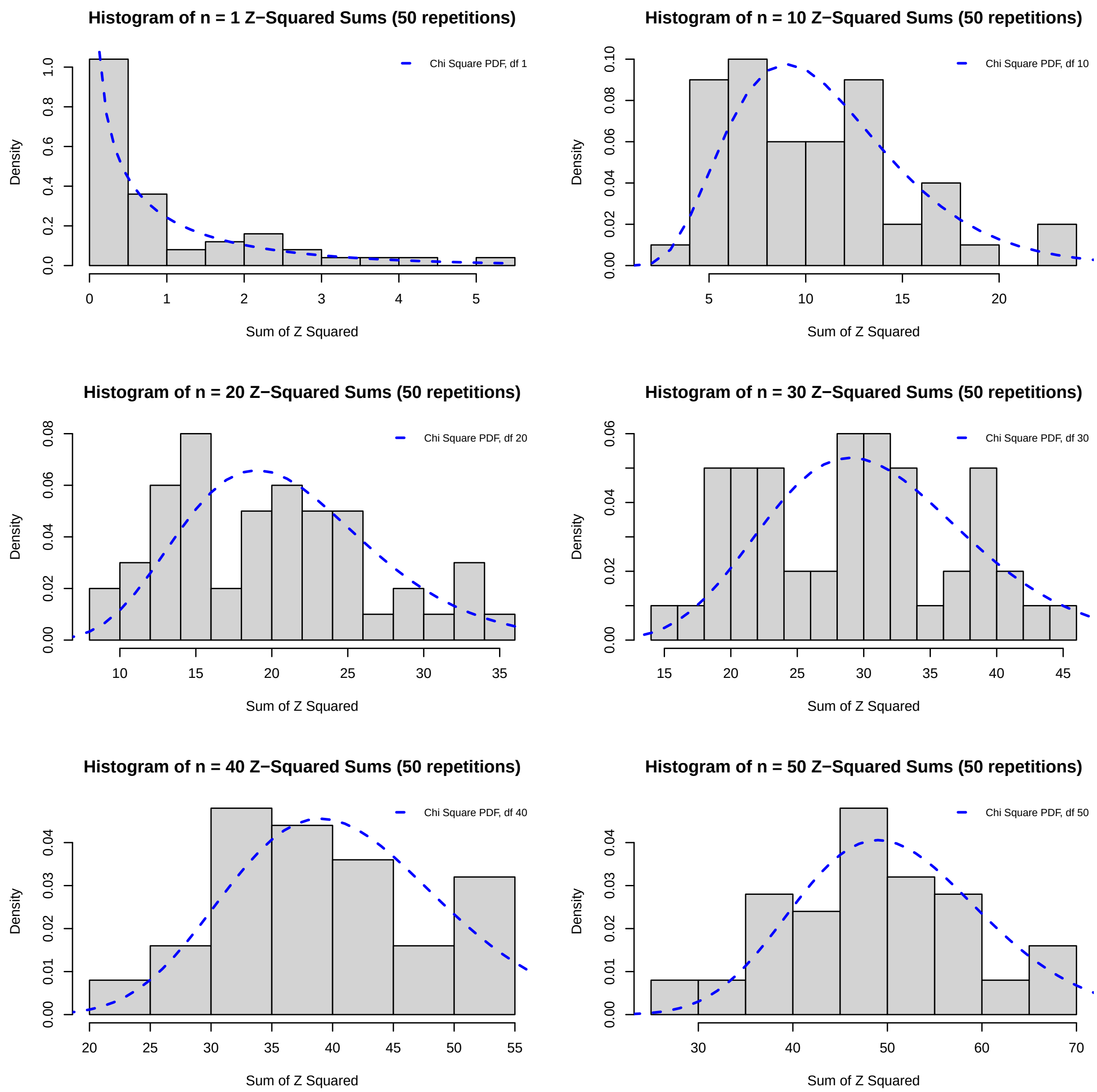
We plot histograms for n Z^2 sums, with the χ^2 PDF shown for corresponding n degrees of freedom.

We want to show sum of n Z^2 random variables $\sim \chi^2_{(df=n)}$

```
layout(matrix(c(1, 2, 3, 4, 5, 6), nrow=3, ncol=2, byrow=TRUE), widths = c(10,10))

xz = "Sum of Z Squared"

#hist df 1
```

```r
hist((z_sqs[,1]), breaks = 10, prob = T,
     main = "Histogram of n = 1 Z-Squared Sums (50 repetitions)",
     xlab = xz)
x <- seq(min(z_sqs[,1]),max(z_sqs[,1]), length = 50)
y <- dchisq(x, df = 1)
lines(x,y, col = "blue", lwd = 2, lty = "dashed")
legend(x = "topright",          # Position
       legend = "Chi Square PDF, df 1",  # Legend texts
       lty = 2,           # Line types
       col = "blue",            # Line colors
       lwd = 2,
       cex = .75,
       bty="n")

#hist of df 10
hist(rowSums(z_sqs[,1:10]),
     breaks = 10,
     prob = T,
     main = "Histogram of n = 10 Z-Squared Sums (50 repetitions)",
     xlab = xz)
lines(dchisq(0:35, 10), col = "blue", lwd = 2, lty = "dashed")
legend(x = "topright",          # Position
       legend = "Chi Square PDF, df 10",  # Legend texts
       lty = 2,           # Line types
       col = "blue",            # Line colors
       lwd = 2,
       cex = .75,
       bty="n")

#hist df 20
hist(rowSums(z_sqs[,1:20]),
     breaks = 10, prob = T,
     main = "Histogram of n = 20 Z-Squared Sums (50 repetitions)",
     xlab = xz)
lines(dchisq(0:35, 20), col = "blue", lwd = 2, lty = "dashed")
legend(x = "topright",          # Position
       legend = "Chi Square PDF, df 20",  # Legend texts
       lty = 2,           # Line types
       col = "blue",            # Line colors
       lwd = 2,
       cex = .75,
       bty="n")

#hist df 30
hist(rowSums(z_sqs[,1:30]), ylim = c(0, 0.06), breaks = 12,prob = T,
     main = "Histogram of n = 30 Z-Squared Sums (50 repetitions)",
     xlab = xz)
lines(dchisq(0:50, 30), col = "blue", lwd = 2, lty = "dashed")
legend(x = "topright",          # Position
       legend = "Chi Square PDF, df 30",  # Legend texts
       lty = 2,           # Line types
       col = "blue",            # Line colors
       lwd = 2,
```

```
       cex = .75,
       bty="n")

#hist df 40
hist(rowSums(z_sqs[,1:40]),
     breaks = 12, prob = T,
     main = "Histogram of n = 40 Z-Squared Sums (50 repetitions)",
     xlab = xz)
lines(dchisq(0:60, 40),col = "blue", lwd = 2, lty = "dashed")
legend(x = "topright",          # Position
       legend = "Chi Square PDF, df 40",  # Legend texts
       lty = 2,           # Line types
       col = "blue",            # Line colors
       lwd = 2,
       cex = .75,
       bty="n")

#hist df 50
hist(rowSums(z_sqs[,1:50]),
     breaks = 14,
     prob = T,
     main = "Histogram of n = 50 Z-Squared Sums (50 repetitions)",
     xlab = xz)
lines(dchisq(0:80, 50), col = "blue", lwd = 2, lty = "dashed")
legend(x = "topright",          # Position
       legend = "Chi Square PDF, df 50",  # Legend texts
       lty = 2,           # Line types
       col = "blue",            # Line colors
       lwd = 2,
       cex = .75,
       bty="n")
```



As expected, the histograms appear consistent with the corresponding χ^2 distributions.
Also, as we increase n, the distributions trend towards a normal shape, consistent with this property of the Chi Squared distribution.

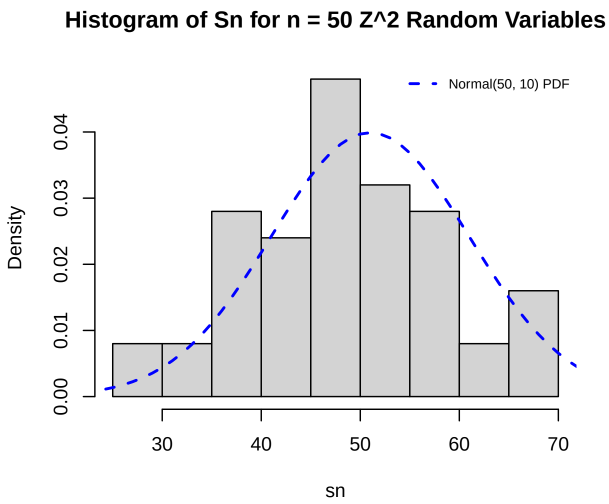
Approaching N(0,1) Distribution with Z^2 Random Variables

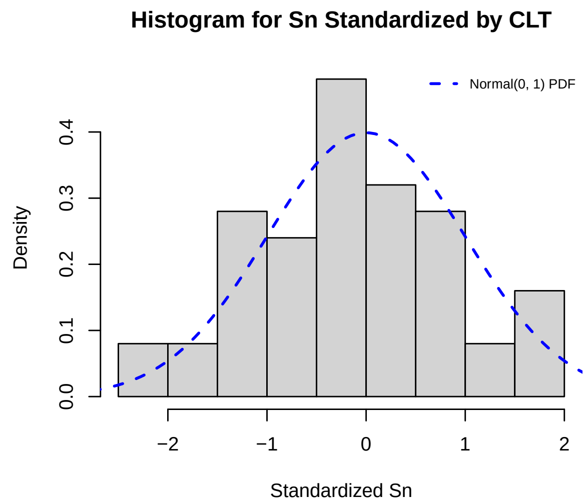
2) By Central Limit Theorem, the sum of n Z^2 and/or χ^2 random variables should approach N(0,1) distribution with large enough n when standardized.

We will examine the distribution of the sums of the n=50 Z^2 random variables we have already generated.

```
# from page 9 in course notes.
# let Sn = the sum of n = 50 iid random variables
```

```r
# each with mean = 1, variance = 2.
# then by Central Limit Theorem
# sn- * (n *u ) / sd * sqrt(n)
# should be aprroximately distributed as N(0,1) if n is large.


#let sn be the sum of random variables Z^2(1)...Z^2(50), each ~ ChiSquare(df = 1).
# we have 50 repetitions

sn <- rowSums(z_sqs)

# standardizing to ~N(0,1) using Central Limit Theorem
z2_std <- 0
for(i in 1:50){
  z2_std[i] <- (sn[i]-50)/sqrt(2*50)
}

#plot histograms

layout(matrix(c(1, 2), nrow=2, ncol=2, byrow=TRUE), widths = c(10,10))
hist(sn, main = "Histogram of Sn for n = 50 Z^2 Random Variables",
     prob = T)
lines(dnorm(0:100, mean = 50, sd = 10),
      col = "blue", lwd = 2, lty = "dashed")
legend(x = "topright",          # Position
       legend = "Normal(50, 10) PDF",  # Legend texts
       lty = 2,           # Line types
       col = "blue",            # Line colors
       lwd = 2,
       cex = .75,
       bty="n")

hist(z2_std,
     main ="Histogram for Sn Standardized by CLT",
     prob = T,
     xlab = "Standardized Sn")
lines(x=seq(-3,3, length=100),
      y = dnorm(x=seq(-3,3, length=100), mean = 0, sd = 1),
      col = "blue", lwd = 2, lty = "dashed")
legend(x = "topright",          # Position
       legend = "Normal(0, 1) PDF",  # Legend texts
       lty = 2,           # Line types
       col = "blue",            # Line colors
       lwd = 2,
       cex = .75,
       bty="n")
```






```r
qqnorm(z2_std) #plot qqnorm
qqline(z2_std)

#means and sds

mean(sn) #mean of Sn, expect 50
```

```
## [1] 48.26072
```

```r
sd(sn)^2 #variance of Sn, expected (2 * 50)
```

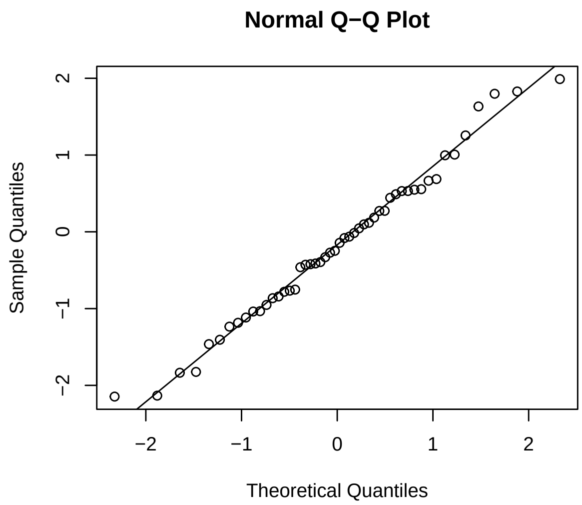
```
## [1] 101.2248
```

```r
mean(z2_std) #mean of Sn standardized by CLT (expected 0)
```

```
## [1] -0.1739279
```

```r
sd(z2_std)^2 #variance of Sn standardized by CLT (expected 1)
```

```
## [1] 1.012248
```



As expected, the distribution of sums of our 50 Z^2 random variables begins to take on a normal shape, and when standardized with Central Limit Theorem, is approximately distributed as N(0,1). This is supported by visual inspection of the QQ plot and the proximity of the expected vs. observed means and standard deviations.

In conclusion, we find these demonstrations support:

Given $Z_i \sim N(0,1)$ iid for $i = 1, 2, ..., n$,

$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2_{(n)}$$