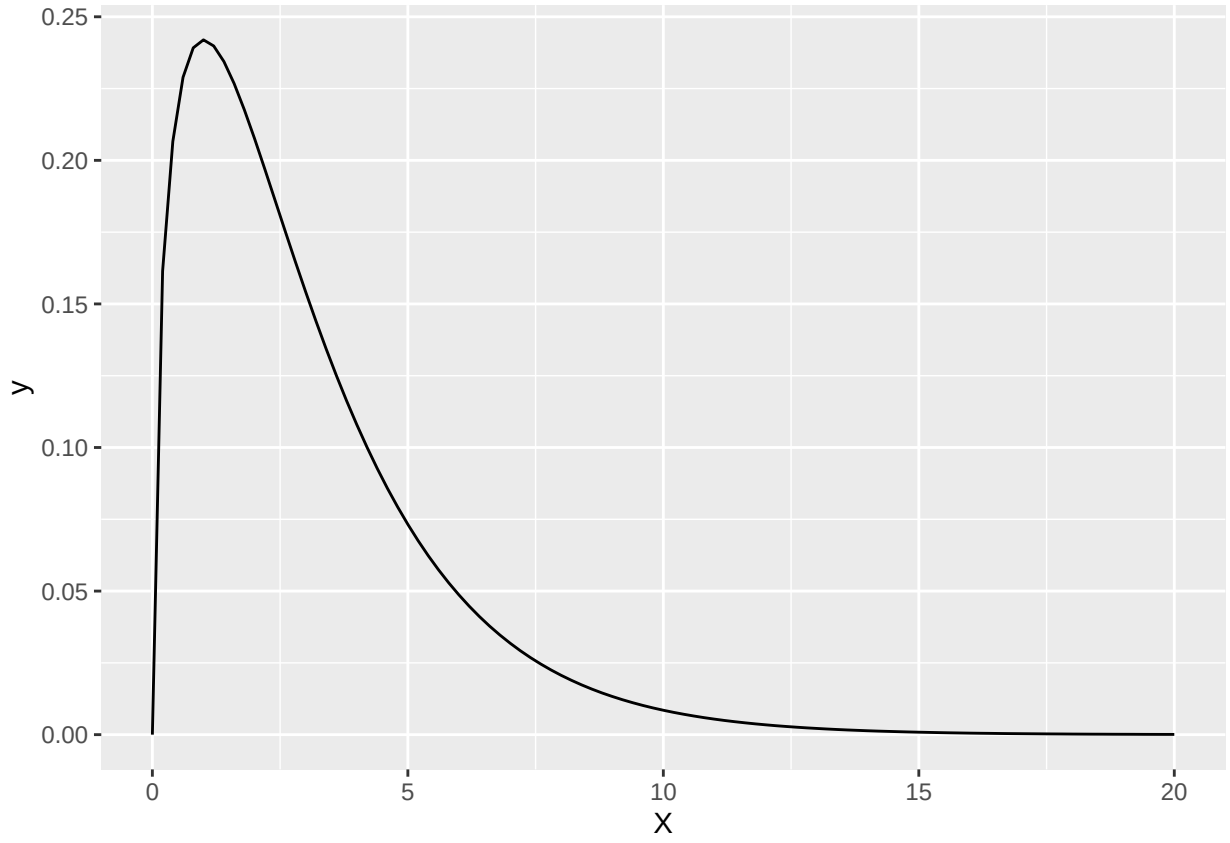
Homework 7

Will Scheib

```
set.seed(06072000)
```

Problem 1

```
ggplot(data.frame(X=c(0,20)), aes(x=X)) + stat_function(fun=dchisq, args=list(df=3))
```



The chi-squared distribution with 3 degrees of freedom is skewed considerably to the right.

Problem 2

Part a

```
ztest <- function(n, df, sd) {
  samp <- rchisq(n, df)
  sampmean <- mean(samp)
  zscore <- (sampmean-df)/(sd/sqrt(n))
  pval <- 2*pnorm(-abs(zscore))
  pval <= 0.05
}
monte_carlo <- function(n, df, sd, k, FUN) {
  results <- replicate(k, FUN(n, df, sd))
  mean(results)
}
sizes = c(9, 27, 51)
two_sidedz <- lapply(sizes, monte_carlo, 3, 2.45, 10000, ztest)
two_sidedz
```

```
## [[1]]
## [1] 0.0467
##
## [[2]]
## [1] 0.0504
##
## [[3]]
## [1] 0.0476
```

Part b

```
ztest2 <- function(n, df, sd) {
  samp <- rchisq(n, df)
  sampmean <- mean(samp)
  zscore <- (sampmean-df)/(sd/sqrt(n))
  pval <- pnorm(zscore)
  pval <= 0.05
}
left_sidedz <- lapply(sizes, monte_carlo, 3, 2.45, 10000, ztest2)
left_sidedz
```

```
## [[1]]
## [1] 0.0312
##
## [[2]]
## [1] 0.0352
##
## [[3]]
## [1] 0.0445
```

Part c

```
ztest3 <- function(n, df, sd) {
  samp <- rchisq(n, df)
  sampmean <- mean(samp)
  zscore <- (sampmean-df)/(sd/sqrt(n))
  pval <- pnorm(zscore, lower.tail=FALSE)
  pval <= 0.05
}
right_sidedz <- lapply(sizes, monte_carlo, 3, 2.45, 10000, ztest3)
right_sidedz
```

```
## [[1]]
## [1] 0.0615
##
## [[2]]
## [1] 0.0606
##
## [[3]]
## [1] 0.0572
```

Problem 3

```
resultsz <- cbind(two_sided=two_sidedz, left_sided=left_sidedz,
                  right_sided=right_sidedz)
row.names(resultsz) <- c("n=9", "n=27", "n=51")
resultsz
```

```
##      two_sided left_sided right_sided
## n=9  0.0467    0.0312     0.0615
## n=27 0.0504    0.0352     0.0606
## n=51 0.0476    0.0445     0.0572
```

As sample size increases, the test commits type 1 errors closer to 5% of the time. The error for the left sided test is very low at small sample sizes, whereas the error for the right sided test is higher at small sample sizes, and both normalize to 0.05 as n increases.

Problem 4

Part a

```
ttest <- function(n, df, sd) {
  samp <- rchisq(n, df)
  sampmean <- mean(samp)
  sampsd <- sd(samp)
  zscore <- (sampmean-df)/(sampsd/sqrt(n))
  pval <- 2*pt(-abs(zscore), n-1)
  pval <= 0.05
}
two_sidedt <- lapply(sizes, monte_carlo, 3, 2.45, 10000, ttest)
two_sidedt
```

```
## [[1]]
## [1] 0.0871
##
## [[2]]
## [1] 0.0715
##
## [[3]]
## [1] 0.0617
```

Part b

```
ttest2 <- function(n, df, sd) {
  samp <- rchisq(n, df)
  sampmean <- mean(samp)
  sampsd <- sd(samp)
  zscore <- (sampmean-df)/(sampsd/sqrt(n))
  pval <- pt(zscore, n-1)
  pval <= 0.05
}
left_sidedt <- lapply(sizes, monte_carlo, 3, 2.45, 10000, ttest2)
left_sidedt
```

```
## [[1]]
## [1] 0.1186
##
## [[2]]
## [1] 0.0876
##
## [[3]]
## [1] 0.0772
```

Part c

```
ttest3 <- function(n, df, sd) {
  samp <- rchisq(n, df)
  sampmean <- mean(samp)
  sampsd <- sd(samp)
  zscore <- (sampmean-df)/(sampsd/sqrt(n))
  pval <- pt(zscore, n-1, lower.tail=FALSE)
  pval <= 0.05
}
right_sidedt <- lapply(sizes, monte_carlo, 3, 2.45, 10000, ttest3)
right_sidedt
```

```
## [[1]]
## [1] 0.0173
##
## [[2]]
## [1] 0.0252
##
## [[3]]
## [1] 0.0295
```

Problem 5

```
resultst <- cbind(two_sided=two_sidedt, left_sided=left_sidedt,
                  right_sided=right_sidedt)
row.names(resultst) <- c("n=9", "n=27", "n=51")
resultst
```

```
##      two_sided left_sided right_sided
## n=9  0.0871    0.1186     0.0173
## n=27 0.0715    0.0876     0.0252
## n=51 0.0617    0.0772     0.0295
```

As sample size increases, the test commits type 1 errors closer to 5% of the time. The error for the left sided test is higher at small sample sizes, whereas the error for the right sided test is lower at small sample sizes, and both normalize to 0.05 as n increases.

Problem 6

Part a

```
incztest <- function(n, df, sd) {
  samp <- rchisq(n, df)
  sampmean <- mean(samp)
  sampsd <- sd(samp)
  zscore <- (sampmean-df)/(sampsd/sqrt(n))
  pval <- 2*pnorm(-abs(zscore))
  pval <= 0.05
}
two_sidedincz <- lapply(sizes, monte_carlo, 3, 2.45, 10000, incztest)
two_sidedincz
```

```
## [[1]]
## [1] 0.1231
##
## [[2]]
## [1] 0.08
##
## [[3]]
## [1] 0.0678
```

Part b

```
incztest2 <- function(n, df, sd) {
  samp <- rchisq(n, df)
  sampmean <- mean(samp)
  sampsd <- sd(samp)
  zscore <- (sampmean-df)/(sampsd/sqrt(n))
  pval <- pnorm(zscore)
  pval <= 0.05
}
left_sidedincz <- lapply(sizes, monte_carlo, 3, 2.45, 10000, incztest2)
left_sidedincz
```

```
## [[1]]
## [1] 0.1425
##
## [[2]]
## [1] 0.0996
##
## [[3]]
## [1] 0.0843
```

Part c

```
incztest3 <- function(n, df, sd) {
  samp <- rchisq(n, df)
  sampmean <- mean(samp)
  sampsd <- sd(samp)
  zscore <- (sampmean-df)/(sampsd/sqrt(n))
  pval <- pnorm(-zscore)
  pval <= 0.05
}
right_sidedincz <- lapply(sizes, monte_carlo, 3, 2.45, 10000, incztest3)
right_sidedincz
```

```
## [[1]]
## [1] 0.0279
##
## [[2]]
## [1] 0.0295
##
## [[3]]
## [1] 0.0318
```

Problem 7

```
resultsincz <- cbind(two_sided=two_sidedincz, left_sided=left_sidedincz,
                  right_sided=right_sidedincz)
row.names(resultsincz) <- c("n=5", "n=21", "n=48")
resultsincz
```

```
##      two_sided left_sided right_sided
## n=5  0.1231    0.1425     0.0279
## n=21 0.08      0.0996     0.0295
## n=48 0.0678    0.0843     0.0318
```

As sample size increases, even the incorrectly performed z test commits type 1 errors closer to 5% of the time. However, the pattern in the errors is similar to the t test in that the error for the left sided test is higher at small sample sizes, whereas the error for the right sided test is lower at small sample sizes, and both normalize to 0.05 as n increases.