# BIOS7345 Lab 1: Non-central distributions

*7 September 2018*

## Chi-Square distribution

### a)

Simulate 10,000 independent random observations from a Central Chi-Square distribution with 5 degrees of freedom

```
num_obs <- 10000  #number of observations
deg_free <- 5  #degrees of freedom
```

by first sampling from the standard normal distribution (i.e. $z_i$)

```
# first sample from the N(0,1)
z <- rnorm(n = deg_free, mean = 0, sd = 1)
```

and working with the sum of the square of the $z_i$'s.

```
# work with the sum of squared N(0,1)s
sumz2 <- sum(z^2)
```

Now, right a function `my_chi_sq()` that does the following:

1. takes in a parameter (`degrees_freedom`) for the desired degrees of freedom (duh),
2. has parameters for the mean and variance of the Normal (which defaults to the standard Normal), and
3. returns a single chi-squared random variable constructed as above.

```
# now write a function
my_chi_sq <- function(degrees_freedom, mu = 0, sigma = 1) {
    z <- rnorm(n = degrees_freedom, mean = mu, sd = sigma)
    sumz2 <- sum(z^2)
    return(sumz2)
}

# test your function
my_chi_sq(deg_free)
```

```
## [1] 2.43288
```

Use the `replicate()` function to create a vector of 10,000 random variable constructed using your `my_chi_sq()` function.

```
central_obs <- replicate(n = num_obs, expr = my_chi_sq(deg_free))
```

Plot the density of the observed chi-square observations with a range of 0 to 60 on the x-axis.

```
# plot the density of the observed chi-squared obs
central_plot <- ggplot(data = data.frame(central_obs), aes(x = central_obs)) +
    geom_density() + xlim(c(0, 60)) + ggtitle("Chi-Squared (df=5)")
```

Verify that this density plot agrees with the true density for $\chi^2_5$ using the built in `dchisq()` function.

```
# compare your density to the true density of a chi-squared(5)
central_plot <- central_plot + stat_function(fun = dchisq, args = list(df = 5),
    col = "red")
```

## b)

Simulate 10,000 independent random observations from a non-central chi-square distribution with 5 degrees of freedom and non-centrality parameter $\lambda = 2$

```r
num_obs <- 10000  #number of observations
deg_free <- 5  #degrees of freedom
lambda <- 2  #non-centrality parameter
```

by first sampling from the normal distribution (i.e. $y_i$), and working with the sum of the square of the $y_i$'s.

Recall from Rencher (5.22) that

$$\lambda = \frac{1}{2}\sum_{i=1}^{n} \mu_i^2,$$

where \$n = \$ the number of Normals you're summing up (which is also what?).

```r
# solve for mu using (5.22)
nc_mu <- sqrt(2 * lambda/deg_free)

# sample from the Normal distribution
y <- rnorm(n = deg_free, mean = nc_mu, sd = 1)

# work with the sum of squared Normals
sumy2 <- sum(y^2)
```

Now, we can use our handy dandy `my_chi_sq()` function from above to create a vector of 10,000 non-central chi squared observations.

```r
noncentral_obs <- replicate(n = num_obs, expr = my_chi_sq(deg_free, mu = nc_mu,
    sigma = 1))
```

Plot the density of the observed non-central chi-square observations with a range of 0 to 60 on the x-axis.

```r
# plot the observations
noncentral_plot <- ggplot(data = data.frame(noncentral_obs), aes(x = noncentral_obs)) +
    geom_density() + xlim(c(0, 60)) + ggtitle(TeX("Chi-Squared (df=5,$\\lambda$=2)"))
```

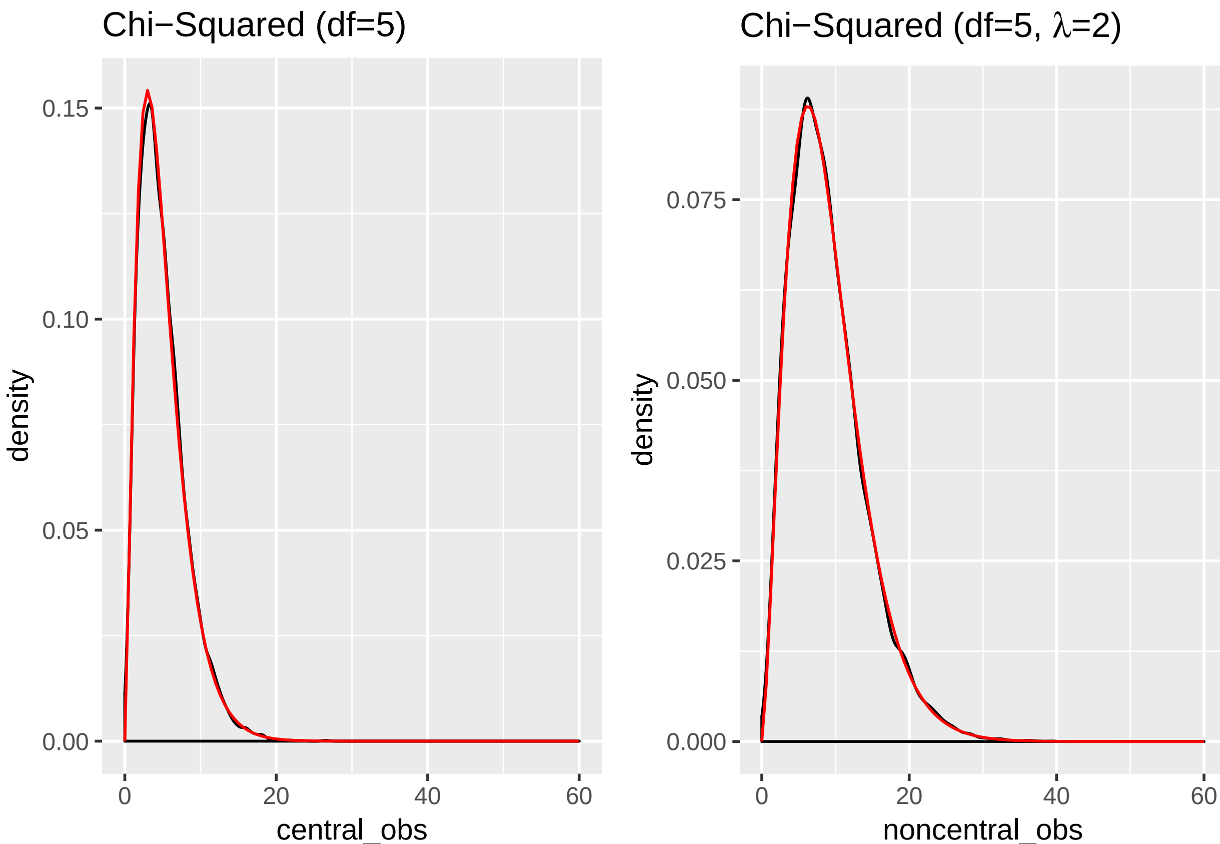
Verify that this density plot agrees with the true density for $\chi^2_{5,2}$ using the built in `dchisq()` function. Hint: $\lambda$ in `R` is 2 times $\lambda$ in class (don't ask me why).

```r
# compare your density to the true density of a chi-squared(5)
noncentral_plot <- noncentral_plot + stat_function(fun = dchisq, args = list(df = 5,
    ncp = 2 * lambda), col = "red")
```

## c)

Compare plots from **1a)** to **1b)**.

```r
ggarrange(central_plot, noncentral_plot)
```




How do the mean and variance compare between the central and non-central chi-square distributions?

Let's create a dataframe to keep track of these for the three distributions.

```
# make a table so we can view these side-by-side
tab <- data.frame(Dist = "Chi-Squared", Central = paste(round(mean(central_obs),
    3), " (", round(var(central_obs), 3), ")", sep = ""), NonCentral = paste(round(mean(noncentral_obs)
    3), " (", round(var(noncentral_obs), 3), ")", sep = ""))
```

## F distribution

### a)

Simulate 10,000 independent random observatios from a central F distribution with 2 and 20 degrees of freedom for the numerator and denominator, respectively.

```
n <- 10000  #number of observations
df_num <- 2  #degrees of freedom (numerator)
df_denom <- 20  #degrees of freedom (denominator)
```

Do this by first sampling from two independent chi-square distributions

```
num <- rchisq(n = 1, df = df_num)
denom <- rchisq(n = 1, df = df_denom)
```

and work with the ratio of the observations.

```
f <- (num/df_num)/(denom/df_denom)
```

Now, right a function `my_f()` that does the following:

1. takes in degrees of freedom parameters for the numerator and denominator (`degrees_freedom_num`, `degrees_freedom_denom`),
2. has a parameter for a noncentrality parameter (`ncp`) (which defaults to 0), and
3. returns a single F random variable constructed as above.

```
# now write a function
my_f <- function(degrees_freedom_num, degrees_freedom_denom, ncp = 0) {
    f <- (rchisq(n = 1, df = df_num, ncp = 2 * ncp)/df_num)/(rchisq(n = 1, df = df_denom)/df_denom)
    return(f)
}

# test your function
my_f(df_num, df_denom)
```

```
## [1] 1.132051
```

Use the `replicate()` function to create a vector of 10,000 random variable constructed using your `my_f()` function.

```
central_obs <- replicate(n = num_obs, expr = my_f(df_num, df_denom))
```

Plot the density of the observed F distribution with a range of 0 to 20 on the x-axis.

```
central_plot <- ggplot(data = data.frame(central_obs), aes(x = central_obs)) +
    geom_density() + xlim(c(0, 20)) + ggtitle("F (2,20)")
```

Verify that this density plot agrees with the true density for $F_{2,20}$ using the built in `df()` function.

```
# compare your density to the true density of a F(2,20)
central_plot <- central_plot + stat_function(fun = df, args = list(df1 = df_num,
    df2 = df_denom), col = "red")
```

## b)

Simulate 10,000 independent random observations from a non-central F distribution with 2 and 20 degrees of freedom for the numerator and denominator, respectively, and a non-centrality parameter $\lambda = 2$.

```
n <- 10000  #number of observations
df_num <- 2  #degrees of freedom (numerator)
df_denom <- 20  #degrees of freedom (denominator)
lambda <- 2  #non-centrality parameter
```

Do this by first sampling from two independent (possibly non-central) chi-square distributions (recalling that `R` parameterizes $\lambda$ differently)

```
num <- rchisq(n = 1, df = df_num, ncp = 2 * lambda)
denom <- rchisq(n = 1, df = df_denom)
```

and work with the ratio of the observations.

```
f <- (num/df_num)/(denom/df_denom)
```

Now, we can use our handy dandy `my_f()` function from above to create a vector of 10,000 non-central F observations.

```r
noncentral_obs <- replicate(n = num_obs, expr = my_f(df_num, df_denom, lambda))
```

Plot the density of the observed non-central chi-square observations with a range of 0 to 60 on the x-axis.

```r
# plot the observations
noncentral_plot <- ggplot(data = data.frame(noncentral_obs), aes(x = noncentral_obs)) +
    geom_density() + xlim(c(0, 20)) + ggtitle(TeX("F (2,20,$\\lambda$=2)"))
```

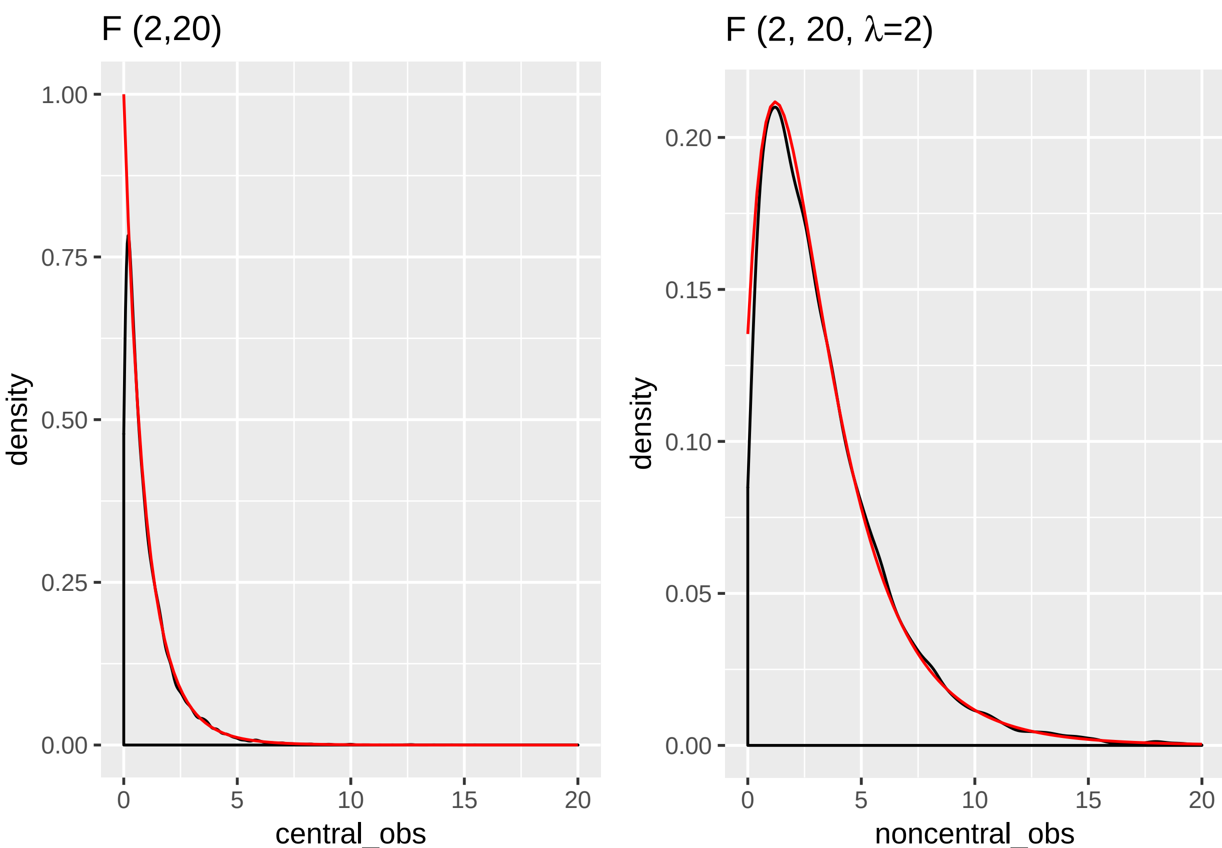
Verify that this density plot agrees with the true density for $F_{2,20}(2)$ using the built in `df()` function. Recall: $\lambda$ in `R` is 2 times $\lambda$ in class (don't ask me why).

```r
# compare your density to the true density of a chi-squared(5)
noncentral_plot <- noncentral_plot + stat_function(fun = df, args = list(df1 = df_num,
    df2 = df_denom, ncp = 2 * lambda), col = "red")
```

### c)

Compare the plots from **2a)** and **2b)**.

```r
ggarrange(central_plot, noncentral_plot)
```




How do the mean and variance compare between the central and non-central F distributions? Let's append a row to our dataframe for the F distribution.

```r
# make a table so we can view these side-by-side
tab <- rbind(tab, data.frame(Dist = "F", Central = paste(round(mean(central_obs),
    3), " (", round(var(central_obs), 3), ")", sep = ""), NonCentral = paste(round(mean(noncentral_obs)
    3), " (", round(var(noncentral_obs), 3), ")", sep = "")))
```

# T distribution

## a)

Simulate 10,000 independent random observations from a central T distribution with 2 degrees of freedom

```
num_obs <- 10000  #number of observations
deg_free <- 2  #degrees of freedom
```

by first sampling from the standard normal distribution (i.e. $z_i$)

```
# sample from the N(0,1)
z <- rnorm(n = 1, mean = 0, sd = 1)
```

and from an independent chi-square distribution,

```
u <- rchisq(n = 1, df = deg_free)
```

and working with the ratio of the two observations.

```
t <- z/sqrt(u/deg_free)
```

Now, right a function `my_t()` that does the following:

1. takes in a degrees of freedom parameter (`degrees_freedom`),
2. has a parameter for a noncentrality parameter (`ncp`) (which both default to 0), and
3. returns a single t random variable constructed as above.

```
# now write a function
my_t <- function(degrees_freedom, ncp = 0) {
    t <- rnorm(n = 1, mean = ncp, sd = 1)/sqrt(rchisq(n = 1, df = degrees_freedom)/degrees_freedom)
    return(t)
}

# test your function
my_t(deg_free)
```

```
## [1] 0.8491256
```

Use the `replicate()` function to create a vector of 10,000 random variable constructed using your `my_t()` function.

```
central_obs <- replicate(n = num_obs, expr = my_t(deg_free))
```

Plot the density of the constructed T observations with a range of -40 to 40 on the x-axis.

```
central_plot <- ggplot(data = data.frame(central_obs), aes(x = central_obs)) +
    geom_density() + xlim(c(-40, 40)) + ggtitle("t(2)")
```

Verify that this density plot agrees with the true density for $t_2$ using the built in `dt()` function.

```
# compare your density to the true density of a t(2)
central_plot <- central_plot + stat_function(fun = dt, args = list(df = deg_free),
    col = "red")
```

## b)

Simulate 10,000 independent random observations from a non-central T distribution with 2 degrees of freedom and non-centrality parameter $\mu = -3$

```r
n <- 10000  #number of observations
deg_free <- 2  #degrees of freedom
mu <- -3  #non-centrality parameter
```

by first sampling from the normal distribution (i.e. $y_i$)

```r
y <- rnorm(n = 1, mean = mu, sd = 1)
```

and from an independent chi-square distribution,

```r
u <- rchisq(n = 1, df = deg_free)
```

and working with the ratio of the two observations.

```r
t <- z/sqrt(u/deg_free)
```

Now, we can use our handy dandy `my_f()` function from above to create a vector of 10,000 non-central F observations.

```r
noncentral_obs <- replicate(n = num_obs, expr = my_t(deg_free, ncp = mu))
```

Plot the density of the observed T observations with a range from -40 to 40 on the x-axis.

```r
# plot the observations
noncentral_plot <- ggplot(data = data.frame(noncentral_obs), aes(x = noncentral_obs)) +
    geom_density() + xlim(c(-40, 40)) + ggtitle(TeX("t (2,$\\lambda$=2)"))
```

Verify that this density plot agrees with the true density for $F_{2,20}(2)$ using the built in `df()` function. Recall: $\lambda$ in `R` is 2 times $\lambda$ in class (don't ask me why).

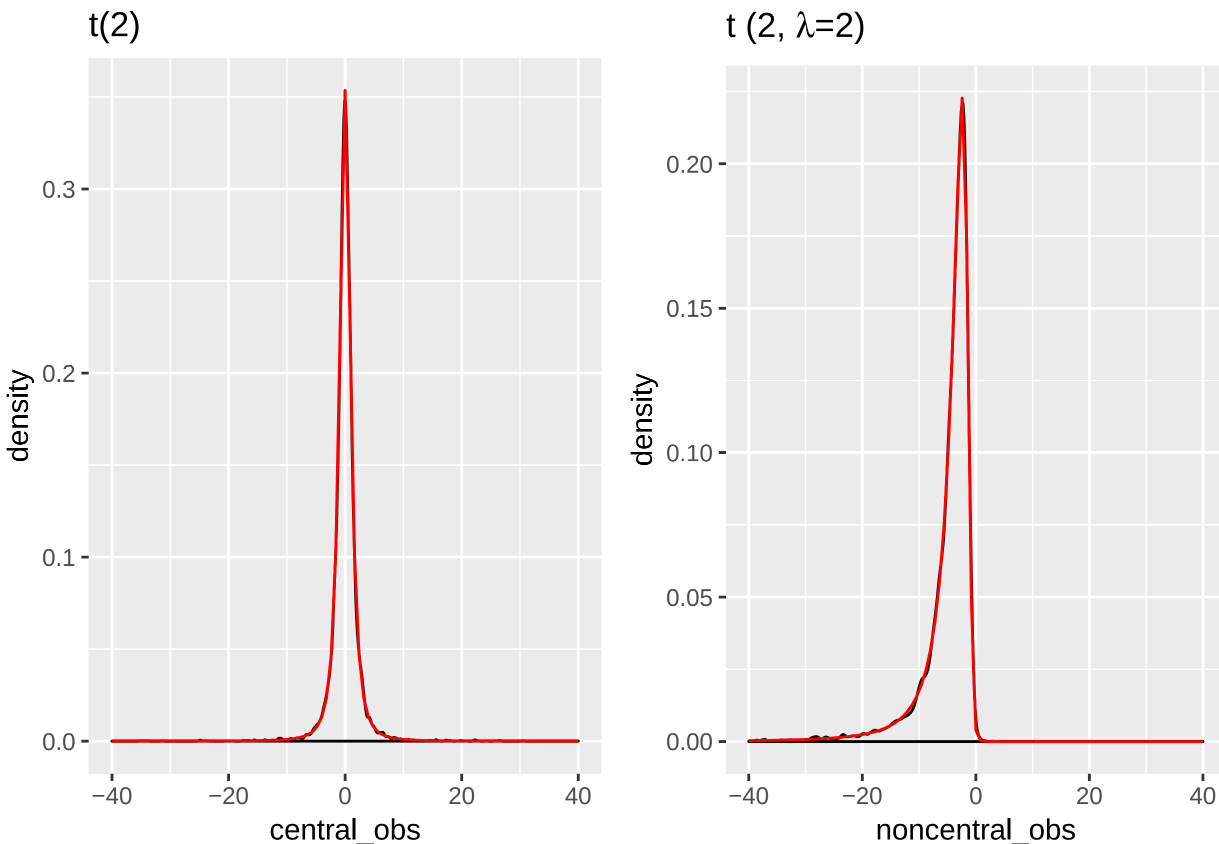
```r
# compare your density to the true density of a t(2,20)
noncentral_plot <- noncentral_plot + stat_function(fun = dt, args = list(df = deg_free,
    ncp = mu), col = "red")
```

Verify that this density plot agrees with a similar plot using the built-in `rt()` function with a non-centrality parameter.

### c)

Compare the plots from **3a)** and **3b)**.

```r
ggarrange(central_plot, noncentral_plot)
```




How do the mean and variance compare between the central and non-central t distributions? Let's append a row to our dataframe for the t distribution.

```
# make a table so we can view these side-by-side
tab <- rbind(tab, data.frame(Dist = "t", Central = paste(round(mean(central_obs),
    3), " (", round(var(central_obs), 3), ")", sep = ""), NonCentral = paste(round(mean(noncentral_obs),
    3), " (", round(var(noncentral_obs), 3), ")", sep = "")))
```

## Discussion

Use the `xtable()` function to print a PDF-friendly version of your table comparing the mean/variance between the central and non-central versions of these distributions.

```
xtable(tab, caption = "Comparing the mean/variance of central and non-central distributions.")
```

| | Dist | Central | NonCentral |
|---|---|---|---|
| 1 | Chi-Squared | 5.033 (9.888) | 8.951 (25.97) |
| 2 | F | 1.082 (1.51) | 3.327 (8.153) |
| 3 | t | -0.001 (7.775) | -5.264 (85.935) |

Table 1: Comparing the mean/variance of central and non-central distributions.

### Chi-squared distribution

- Mean and variance of non-central Chi-squared > mean and variance of central Chi-squared

## F distribution

- Mean and variance of non-central F $>$ mean and variance of central F

## t distribution

- Mean of the non-central t $<$ mean of central t
- Variance of non-central t $>$ variance of central t
- These depend on the value of $\mu$ (via a complicated formula)