STAT/BIOS 823

Probability Distributions

Homework 5.1

Directions

Use the attached `RMarkdown Probability Distributions.Rmd` to create a **pdf** report summarizing the common distributions discussed in this lesson. In addition, I have done the first few to show you my expectations for your report, and have given you some **leading questions to complete**.

Partial R-code for the following distributions have been given for you.

1. Bernoulli
2. Binomial
3. Hypergeometric: Qn 1
4. Poisson: Qn 2
5. Geometric: Qn 3
6. Negative Binomial: Qn 4
7. Normal: Qn 5
8. Exponential: Qn 6
9. Chi-square: Qn 7
10. Student's t: Qn 8
11. F: Qn 9 Optional
12. Beta Optional
13. Logistic Optional

All R-code and output must be clearly shown. Late submission will attract a penalty of 10 points per day after the due date.

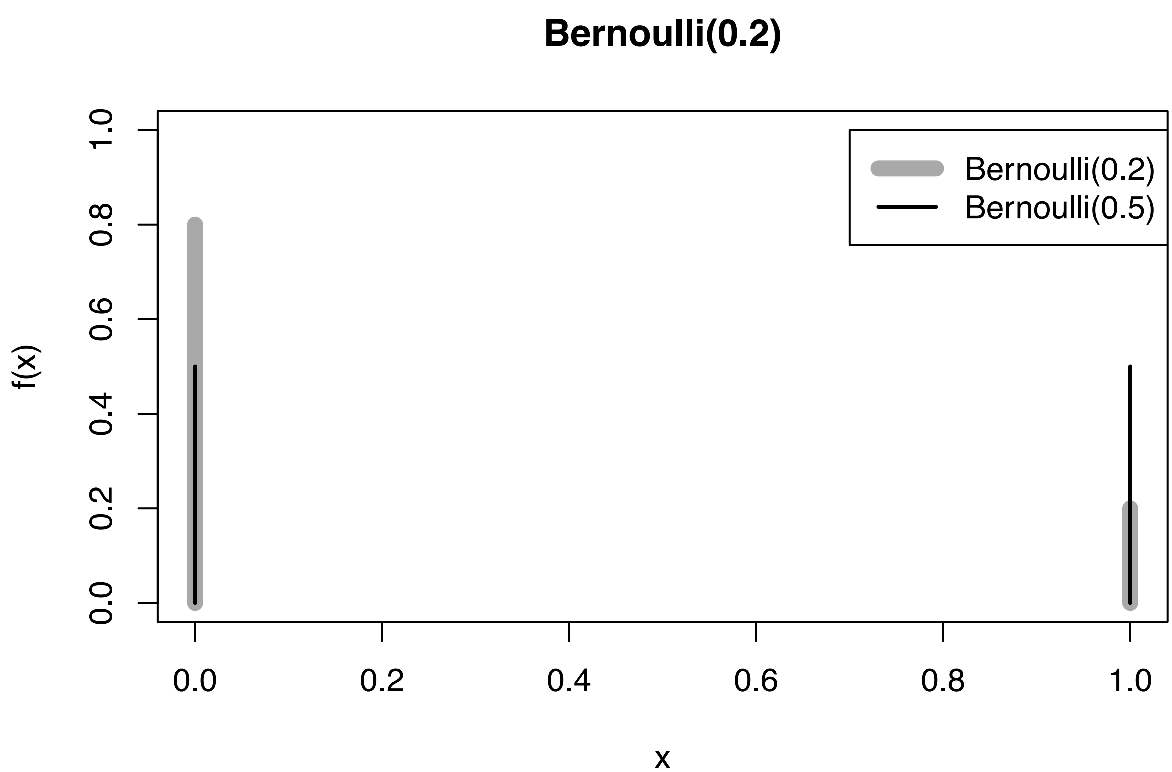
If you have any questions, please post them on the lesson discussion board.

1 Discrete Distributions

1.1 Bernoulli

The **Bernoulli distribution**, named for Jacob Bernoulli, assigns probability to the outcomes of a single Bernoulli experiment—one where the only possible outcomes can be thought of as a "success" or a "failure" (e.g., a coin toss). Here, the random variable x can take on the values 1 (success) with probability p, or 0 (failure) with probability $q = 1 - p$. The plot below contains the pmf of two Bernoulli distributions. The first (in gray) has a probability of success $p = 0.2$ and the second (in black) has a probability of success $p = 0.5$.

```r
x <- 0:1
plot(x, dbinom(x, 1, 0.2), type = "h", ylab = "f(x)",
    ylim = c(0, 1), lwd = 8, col = "darkgray",
    main = "Bernoulli(0.2)")
lines(x, dbinom(x, 1, 0.5), type = "h", lwd = 2,
    col = "black")
legend(0.7, 1, c("Bernoulli(0.2)", "Bernoulli(0.5)"),
    col = c("darkgray", "black"), lwd = c(8,
        2))
```



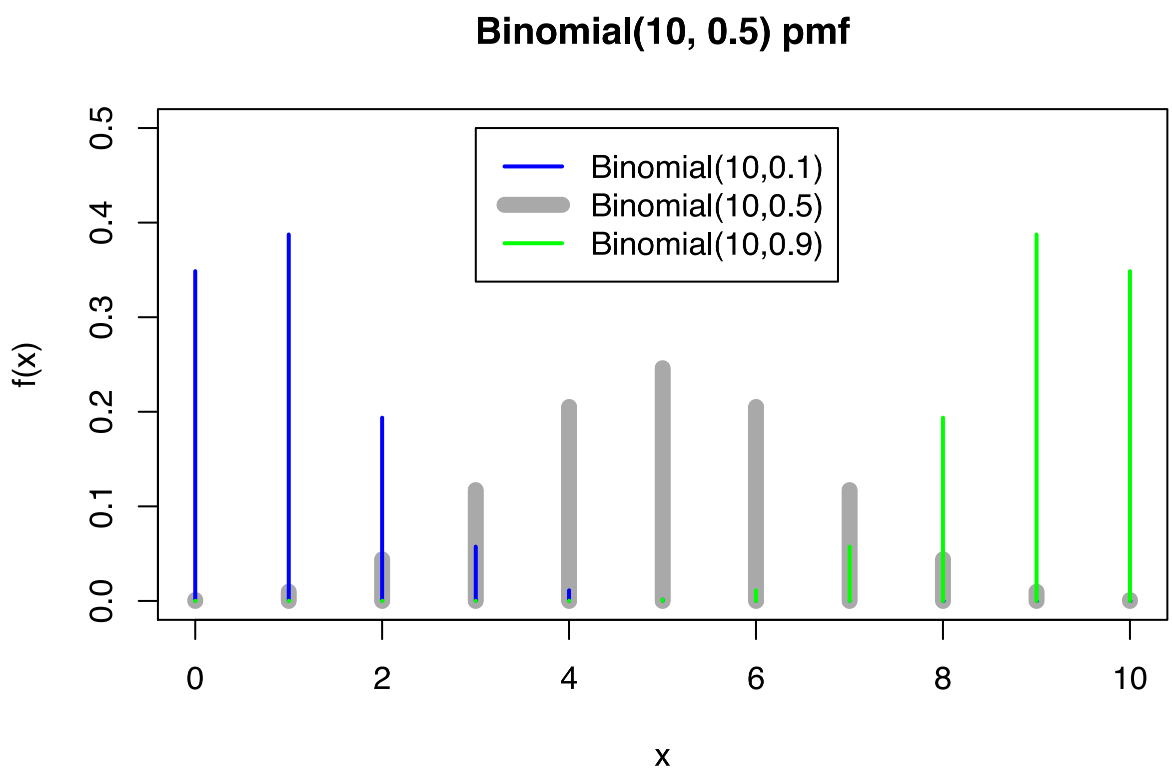
The Bernoulli experiment forms the foundation for many of the next discrete distributions.

1.2 Binomial

The **binomial distribution** applies when we perform n Bernoulli experiments and are interested in the total number of “successes” observed. The outcome here, $y = \sum x_i$, where $P(x_i = 1) = p$ and $P(x_i = 0) = 1 - p$. The plot below displays three binomial distributions, all for $n = 10$ Bernoulli trials: in gray, $p = 0.5$; in blue, $p = 0.1$; and in green, $p = 0.9$.

```r
x <- seq(0, 10, 1)
plot(x, dbinom(x, 10, 0.5), type = "h", ylab = "f(x)",
    lwd = 8, col = "dark gray", ylim = c(0,
        0.5), main = "Binomial(10, 0.5) pmf")
lines(x, dbinom(x, 10, 0.1), type = "h",
    lwd = 2, col = "blue")
lines(x, dbinom(x, 10, 0.9), type = "h",
    lwd = 2, col = "green")
legend(3, 0.5, c("Binomial(10,0.1)", "Binomial(10,0.5)",
```

```
    "Binomial(10,0.9)"), col = c("blue",
    "dark gray", "green"), lwd = c(2, 8,
    2))
```

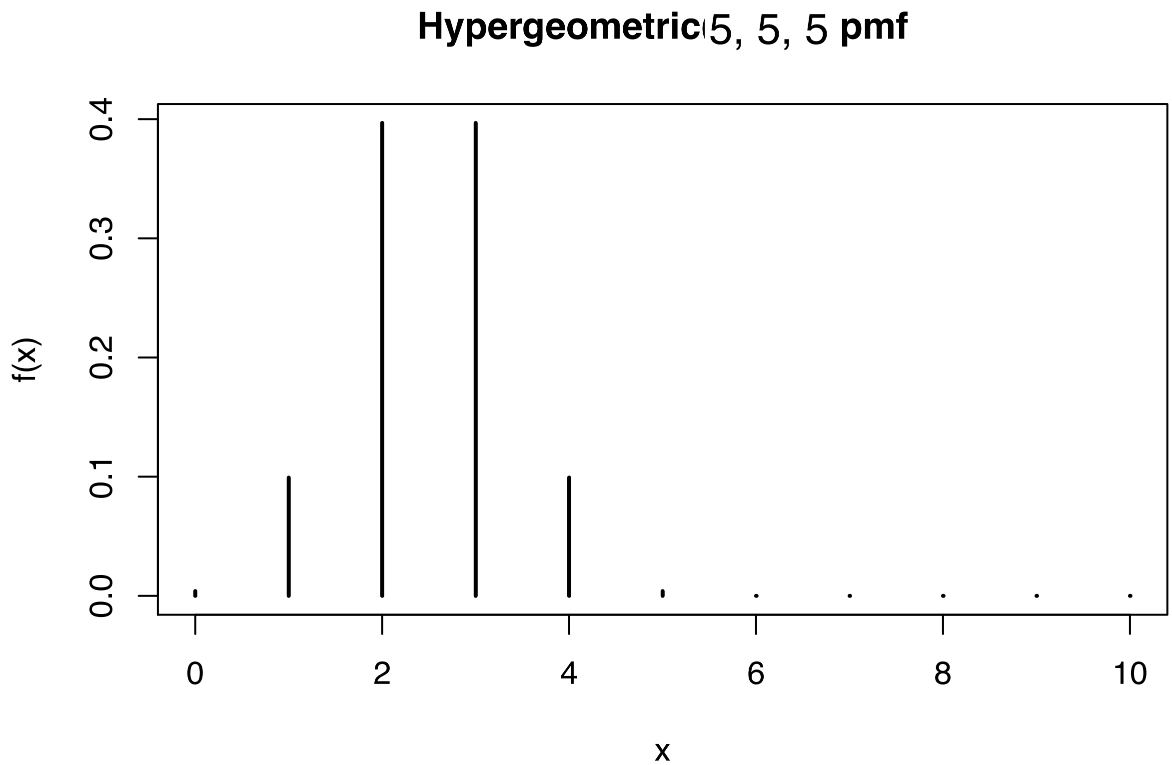


We can see the shifting of probability from low values for $p = 0.1$ to high values for $p = 0.9$. This makes sense, as it becomes more likely with $p = 0.9$ to observe a success for an individual trial. Thus, in 10 trials, more successes (e.g., 8, 9, or 10) are likely. For $p = 0.5$, the number of successes are likely to be around 5 (e.g., half of the 10 trials).

1.3 Hypergeometric

In the example I have below, I have set the number of balls in the urn to 10, 5 of which are white and 5 of which are black. I have also fixed the number of balls drawn from the urn to 5. Play around with the parameters and describe what you see.

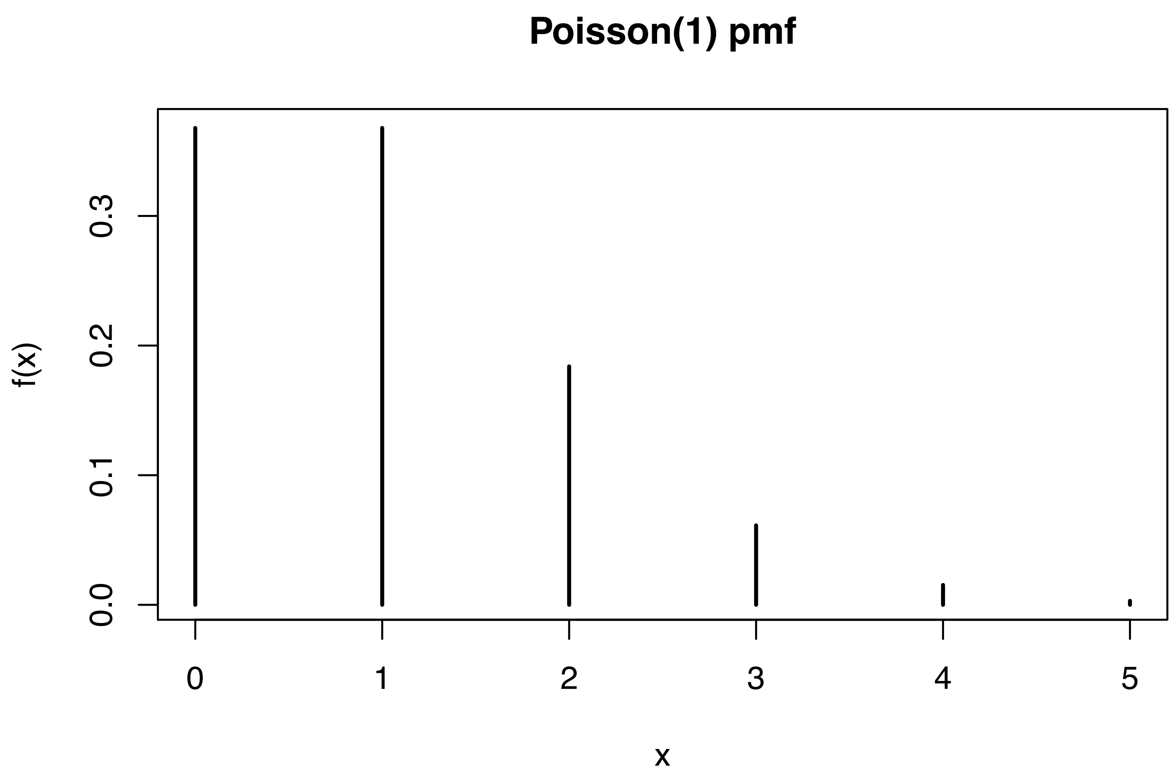
```
x <- seq(0, 10, 1)
plot(x, dhyper(x, 5, 5, 5), type = "h", ylab = "f(x)",
    lwd = 2, main = "Hypergeometric(5 , 5 ,5) pmf")
```



1.4 Poisson

What happens if you increase λ? To 2? To 3?

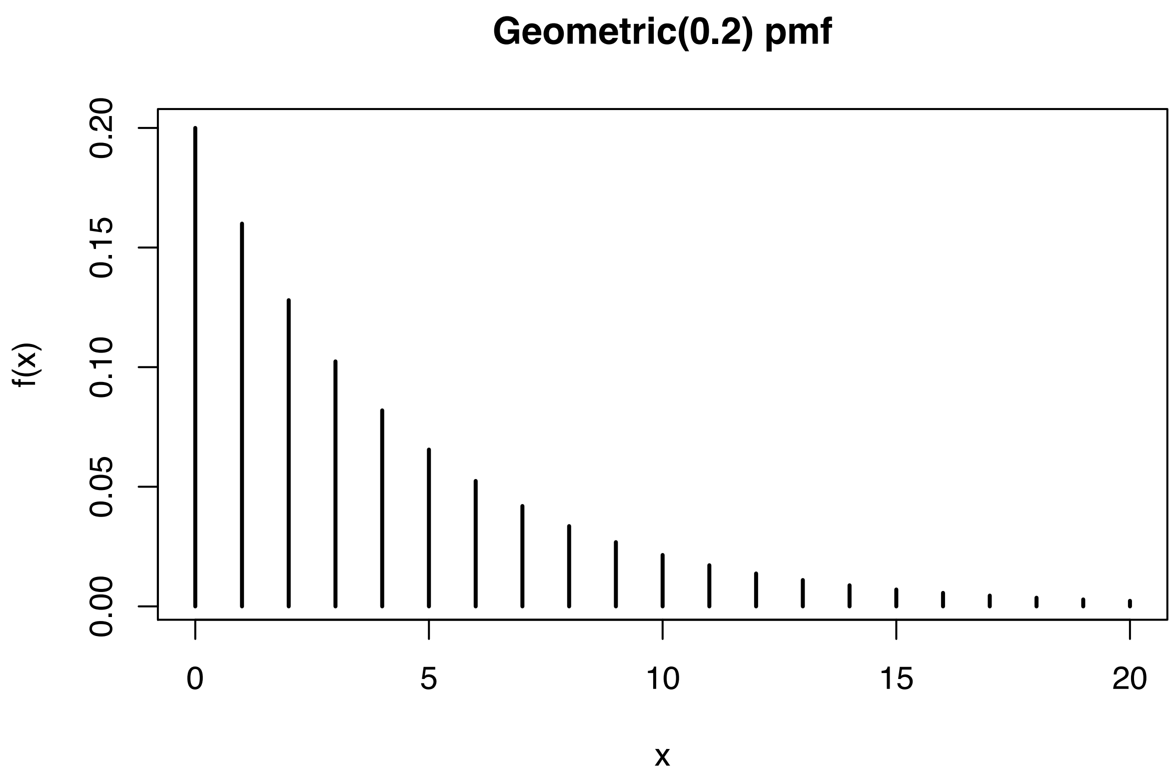
```
x <- seq(0, 5, 1)
plot(x, dpois(x, 1), type = "h", ylab = "f(x)",
    main = "Poisson(1) pmf", lwd = 2)
```



1.5 Geometric

What happens to the geometric distrbution if you vary p? Show me a few plots and explain.

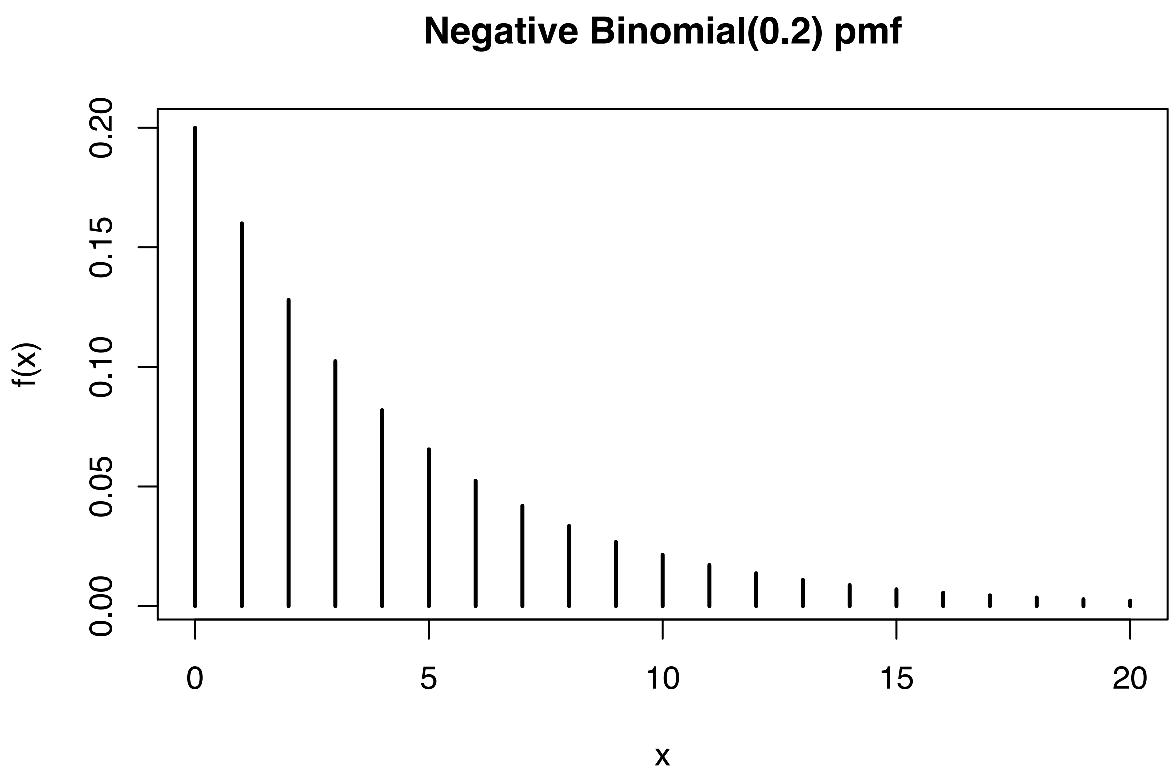
```
x <- seq(0, 20, 1)
plot(x, dgeom(x, 0.2), type = "h", ylab = "f(x)",
    lwd = 2, main = "Geometric(0.2) pmf")
```



1.6 Negative Binomial

The negative binomial I have below has set $r = 1$, so it's identical to the geometric above. Play around with r and see how it changes.

```
x <- seq(0, 20, 1)
plot(x, dnbinom(x, 1, 0.2), type = "h", ylab = "f(x)",
    lwd = 2, main = "Negative Binomial(0.2) pmf")
```

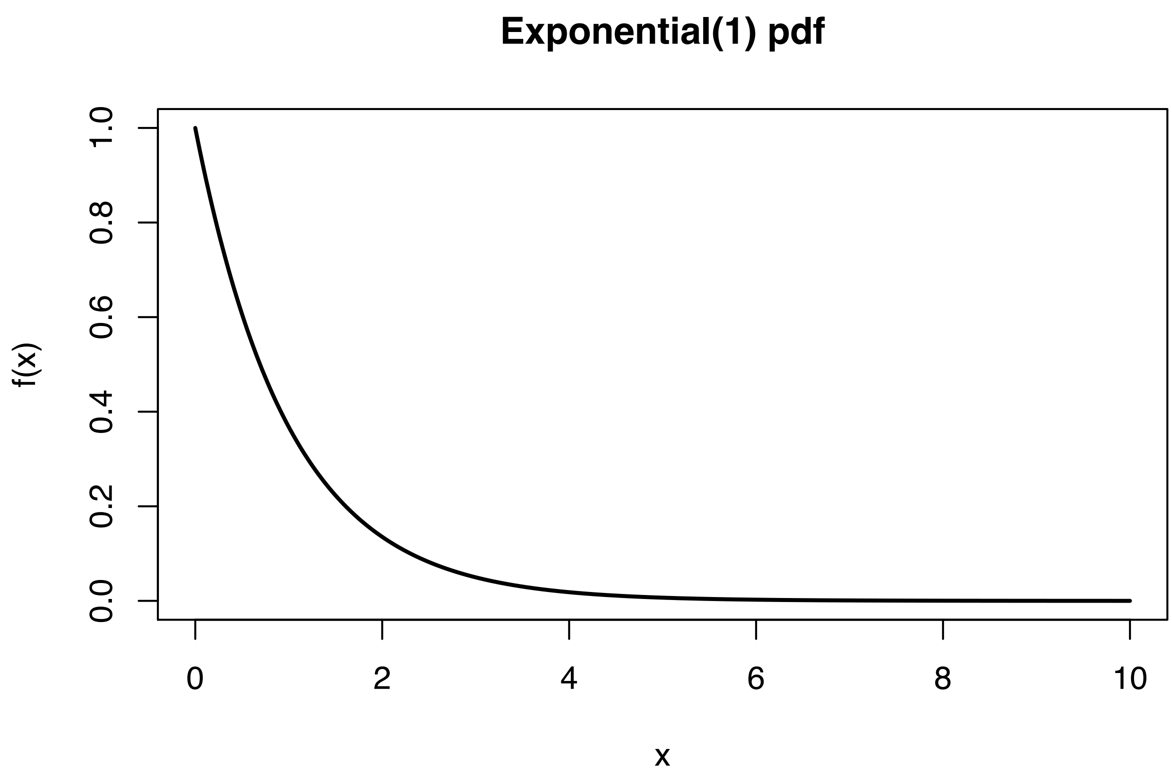


2 Continuous Distributions

2.1 Exponential

Vary λ and describe.

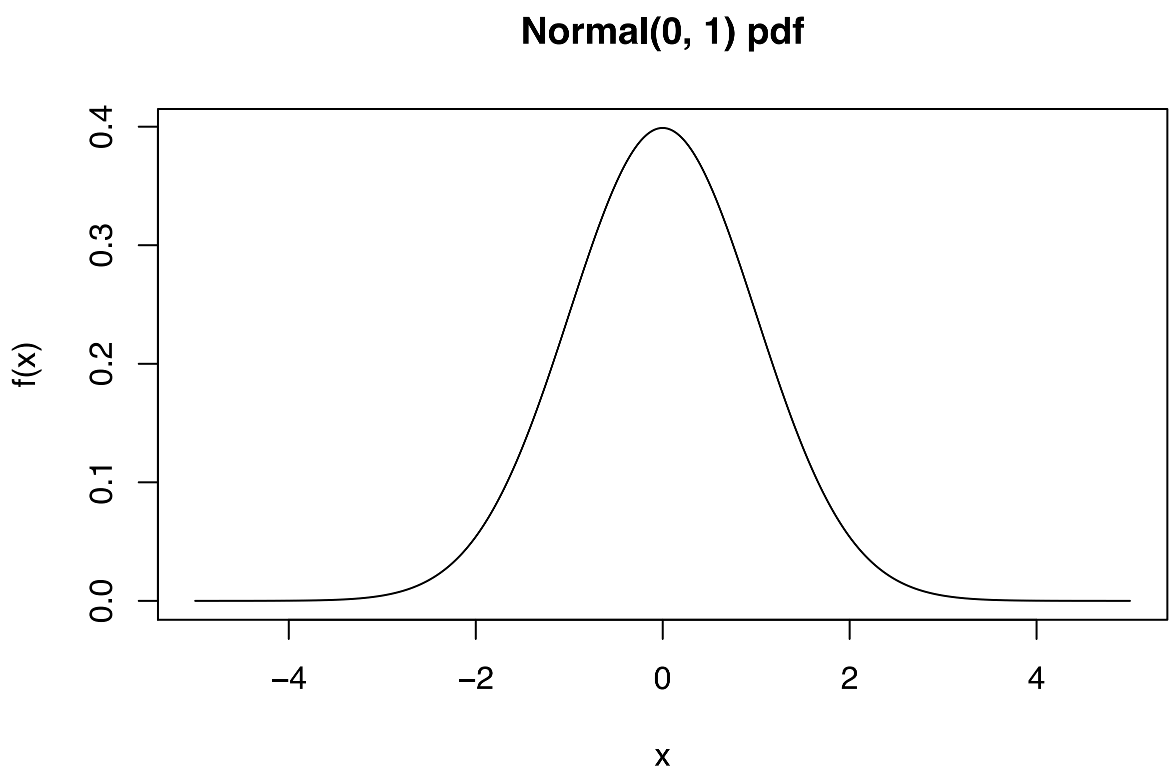
```
x <- seq(0, 10, 0.01)
plot(x, dexp(x, 1), type = "l", ylab = "f(x)",
    lwd = 2, main = "Exponential(1) pdf")
```



2.2 Normal

Vary σ and see how the distribution changes. If you make it too big, you may need to adjust the x-axis by making the sequence span a wider range than -5 to 5. You can use a trial-and-error approach to determing the proper limits for x for a given σ.

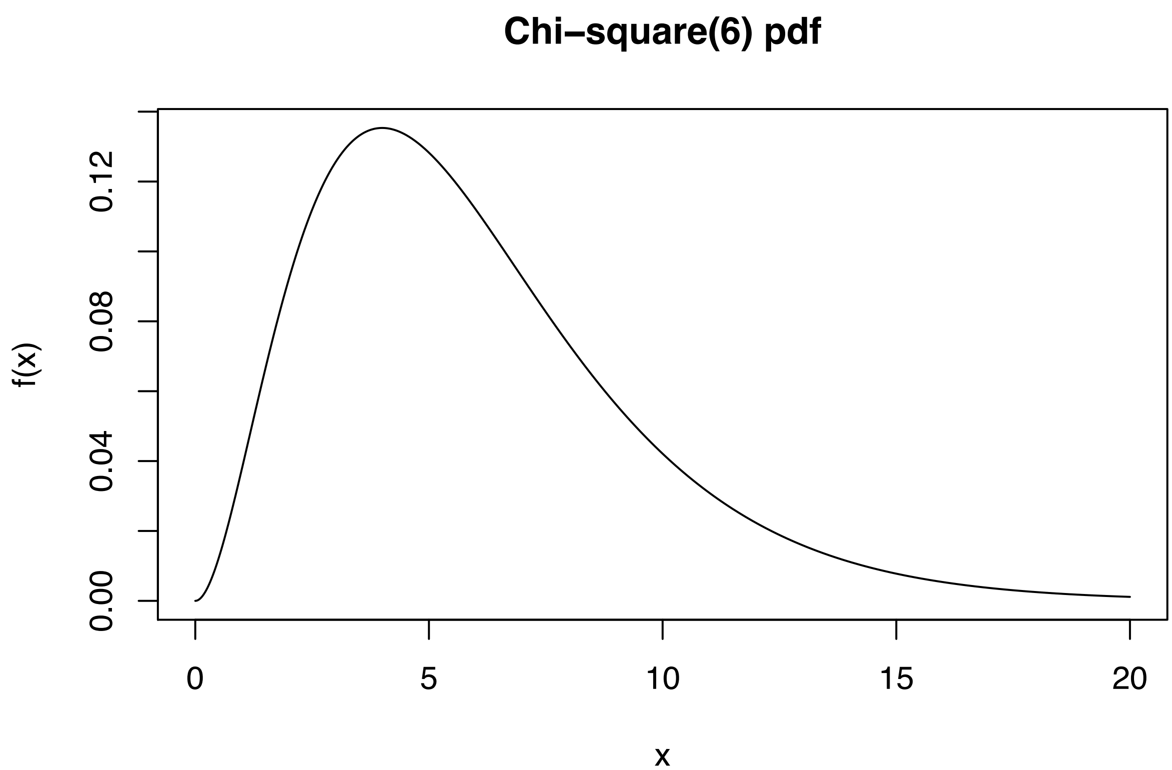
```
x <- seq(-5, 5, 0.01)
plot(x, dnorm(x, 0, 1), type = "l", ylab = "f(x)",
    main = "Normal(0, 1) pdf")
```



2.3 Chisquare

How do the degrees of freedom change the shape? Plot a few and explain.

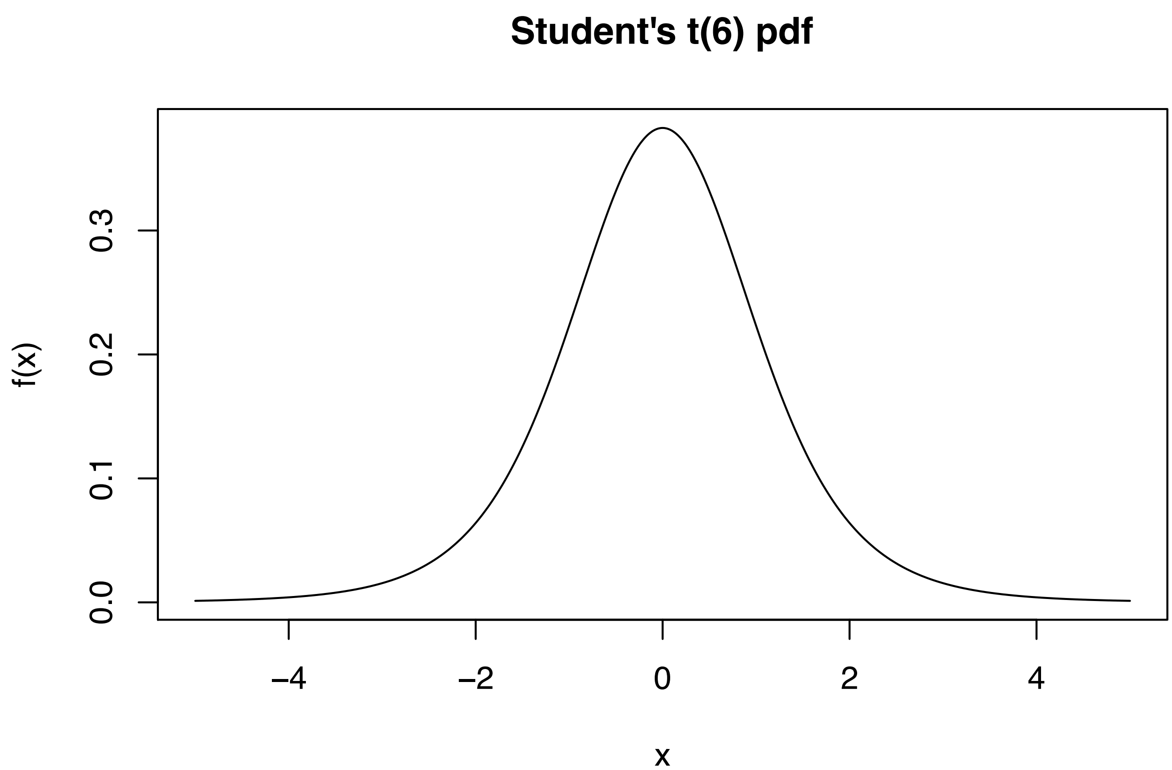
```
x <- seq(0, 20, 0.01)
plot(x, dchisq(x, 6), type = "l", ylab = "f(x)",
    main = "Chi-square(6) pdf")
```



2.4 Students t

How do the degrees of freedom change the shape? Plot a few and explain.

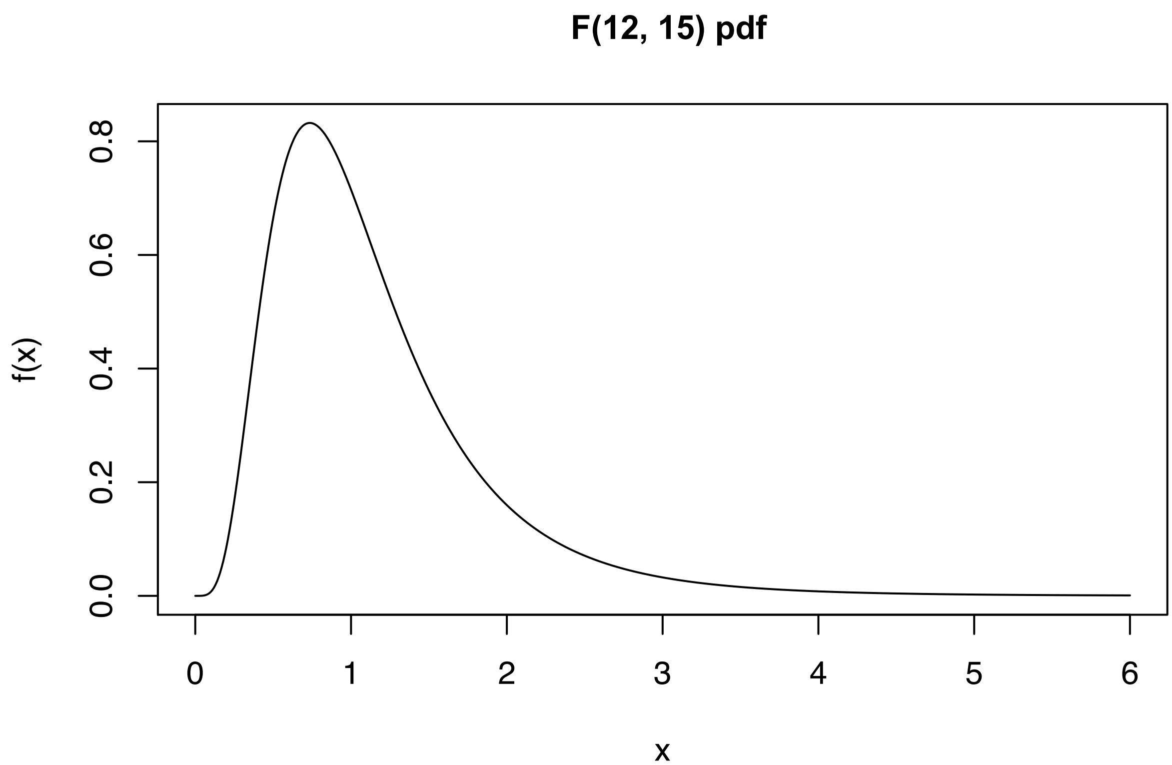
```
x <- seq(-5, 5, 0.01)
plot(x, dt(x, 6), type = "l", ylab = "f(x)",
    main = "Student's t(6) pdf")
```



2.5 F

How do the degrees of freedom (numerator and/or denominator) change the shape? Plot a few and explain.

```
x <- seq(0, 6, 0.01)
plot(x, df(x, 12, 15), type = "l", ylab = "f(x)",
    main = "F(12, 15)  pdf")
```


F(12, 15) pdf
f(x)
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
x