STA255 Week-3

Shahriar Shams

20/01/2020

Review of Week-2

- Random variable (Discrete random variable)
- Probability Mass Function(pmf) and Cumulative Distribution function (F)
- Some common discrete distributions
  - Bernoulli(p) $\rightarrow$ Binomial(n, p) $\rightarrow$ Poisson(λ)
  - Geometric(p) $\rightarrow$ Negative Binomial(r, p)

Learning goals

- Continuous random variable
- Probability density function (*pdf*)
- Cumulative Distribution function (F) (same as last week)
- Some common continuous distribution (Uniform, Exponential, Gamma, Normal)
- Quantile/Percentile of a distribution

Continuous Random Variable

There are several definitions of a *continuous random variable*. Some uses set theory and measure theory while some uses the distribution function.

Here is a simple one:

A *random variable* is *continuous* if it takes *infinite* number of values that are **not countable**.

Example:

- Height of a person.
  - It can be 170.66cm or can be 170.6666666.... cm
  - we have infinite numbers (uncountable) between 170cm and 171cm.
- Waiting time for the next bus.
  - It can be 5.5 minutes or 5.534678023... minutes
  - we have infinite numbers (uncountable) between 5.5 min and 5.6 min.
- Other examples: price of something, age, weight, area, volume, health care cost, income, expense etc.

Probability density Function (pdf)

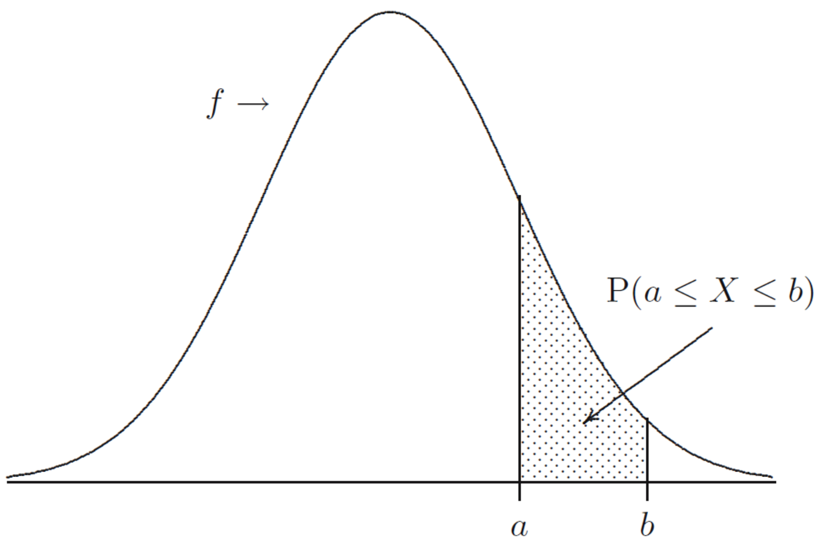
A *probability density function*, denoted by *f*, of a continuous random variable X is a function from $\mathbb{R}$ to $\mathbb{R}$ which satisfies the following conditions:

- $f(x) \geq 0$ for all x
- $\int_{-\infty}^{\infty} f(x)dx = 1$

For any numbers a and b with $a \leq b$

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Using a graph,

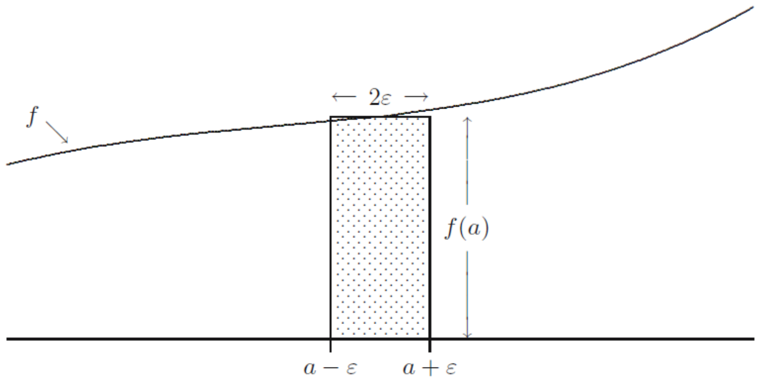


Some properties of *pdf* :

- $f(x) \neq P(X = x)$
  - $f(x)$ does not directly give probability.
  - The height at any point does not represent probability.
  - Re-visit the definition of pmf from last week.
  - pmf is a function from $\mathbb{R}$ to [0,1]
  - But *pdf* is a function from $\mathbb{R}$ to $\mathbb{R}$

- For a continuous random variable X, $P(X = a) = 0$ for any a.
  - For any positive ϵ

$$P(a - \epsilon \leq X \leq a + \epsilon) = \int_{a-\epsilon}^{a+\epsilon} f(x)dx \approx 2\epsilon f(a)$$



  - When $\epsilon \to 0$, the area also goes to 0 $\implies P(X = a) = 0$
  - Hence, $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$

Cumulative Distribution function (F)

Distribution function of a *continuous random variable* X with pdf f is defined as

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$$

Intuitively this gives,

$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = F(b) - F(a)$$

Definition of continuous random variable using distribution function:

A *random variable* is called *continuous* if it's distribution function $F(x)$ is *continuous every where.*

Obtaining $f(x)$ from $F(x)$

$f(x) = \frac{d}{dx}F(x)$

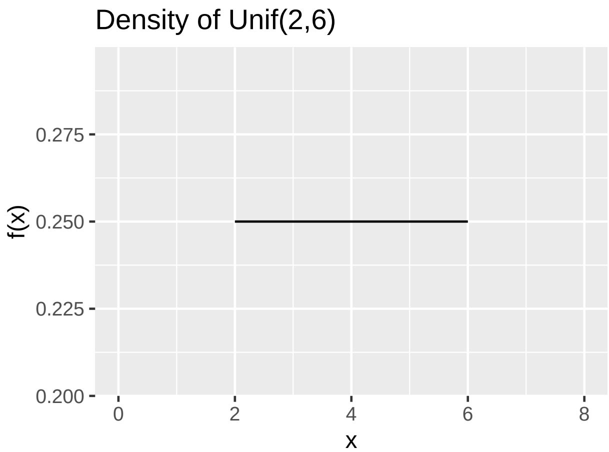
Some common continuous distributions

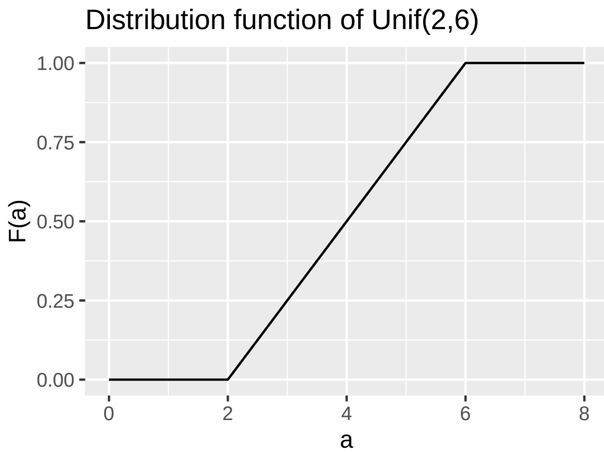
Uniform distribution:

A continuous random variable X has a *Uniform* distribution on the interval $[\alpha, \beta]$ if it's *pdf* is defined as

$$f(x) = \frac{1}{\beta - \alpha} \quad for \;\; \alpha \leq x \leq \beta$$

In notation, we say $X \sim U(\alpha, \beta)$ or $X \sim Unif(\alpha, \beta)$. Here α and β are the two parameters of the distribution.





Exercise: Find the cumulative distribution function of $U(\alpha, \beta)$

Exponential distribution

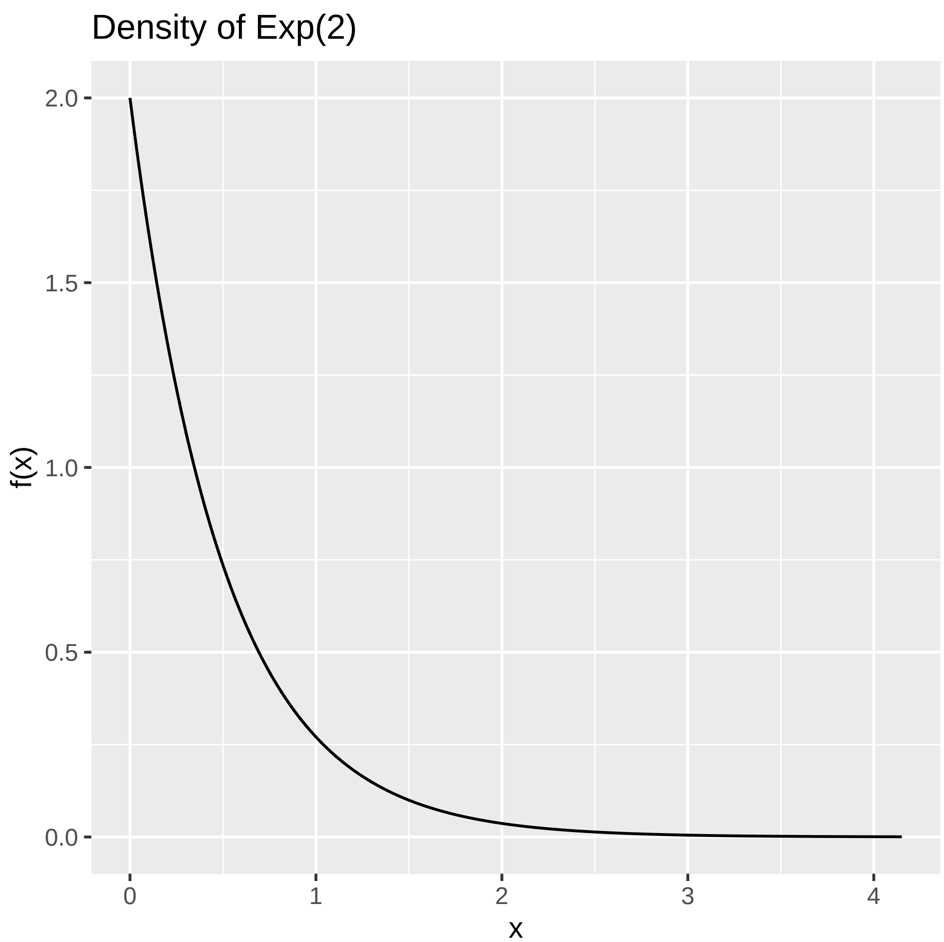
A continuous random variable X has an *Exponential* distribution with parameter λ if it's *pdf* is defined as

$$f(x) = \lambda e^{-\lambda x} \quad for \quad x \geq 0$$

In notation, we say $X \sim Exp(\lambda)$

Note: λ is the "rate parameter" and $\lambda > 0$

Density of a Exp($\lambda = 2$)



Exercise: Find the distribution function of $Exp(\lambda)$

$$f(x) = \lambda e^{-\lambda x} \quad , \quad x \geq 0$$

$$F(a) = \int_{-\infty}^{a} \lambda e^{-\lambda x} dx$$

$$= \lambda \int_{0}^{a} e^{-\lambda x} dx$$

$$= \lambda \left. \frac{e^{-\lambda x}}{-\lambda} \right|_{0}^{a}$$

$$= \left. - e^{-\lambda x} \right|_{0}^{a}$$

$$= -\left(e^{-\lambda a} - e^{-\lambda \cdot 0}\right)$$

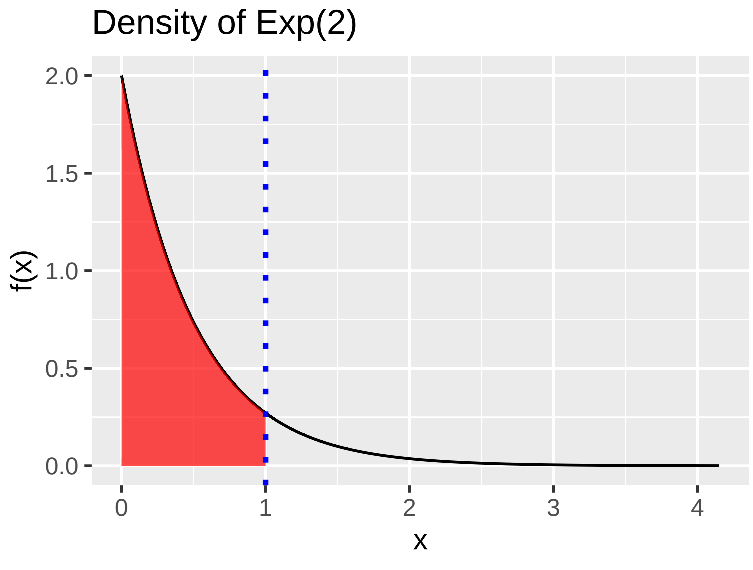
$$= -\left(e^{-\lambda a} - 1\right)$$

$$= 1 - e^{-\lambda a}$$

Example: for $\lambda = 2$ and $a = 1$ we get, $F(1) = 1 - e^{-2*1} = 0.8646647$

Visual of F(1) for Exp($\lambda = 2$)

Continuing with the density curve from the previous page, F(1) represents the following area (shaded in red).



The area shaded in red on the graph is equal to 0.8646647 which is the probability of getting a number less than 1 under a $Exp(\lambda = 2)$ distribution (ie. $P(X \leq 1)$)

Exponential distribution with a different parameter

- Exponential distribtuion is often written with a different parameter
- A continuous random variable X has an *Exponential* distribution with parameter β if it's *pdf* is defined as

$$f(x) = \frac{1}{\beta} e^{-x/\beta} \quad for \;\; x \geq 0$$

- The relation ship between the parameters of these two types of exponential is $\lambda = \frac{1}{\beta}$
- β or $1/\lambda$ is the mean of the distribution.

Gamma distribution:

A continuous random variable X has a *Gamma* distribution with parameter α and β if it's *pdf* is defined as

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \quad for \;\; x \geq 0, \; \alpha > 0 \; and \; \beta > 0$$

here, $\Gamma(\alpha)$ is known as the "gamma function". Here are few properties of the gamma function.

- Definition: $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$
- For any $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
- For any positive integer n, $\Gamma(n) = (n-1)!$
- $\Gamma(\frac{1}{2}) = \sqrt(\pi)$

In notation, we say $X \sim G(\alpha, \beta)$ or $X \sim Gamma(\alpha, \beta)$

Note: If we put $\alpha = 1$ in the Gamma distribution, it becomes an Exponential distribution $\implies$ Exponential is a special case of Gamma distribution.

$$Exp(\beta) \equiv Gamma(1, \beta)$$

Additive property of Gamma distribution

Adding two **independent** Gamma random variables with the **same sacond parameter** produces another Gamma random variable.

$$G(\alpha_1, \beta) + G(\alpha_2, \beta) = G(\alpha_1 + \alpha_2, \beta)$$

Chi-squared distribution

- A special case of Gamma distribution has been named as Chi-squared distribution.
- χ^2 symbolizes the chi-sq distribution.
- A χ^2 distribution with parameter v is equivalent to a Gamma distribution with parameter $v/2$ and $1/2$

$$\chi^2(v) \equiv Gamma(\frac{v}{2}, \frac{1}{2})$$

- Putting $\alpha = v/2$ and $\beta = 2$ in the pdf of the Gamma distribution(from the previous page) we get the pdf of a $\chi^2(v)$.

$$f(x) = \frac{1}{2^{v/2}\Gamma(v/2)} x^{(v/2)-1} e^{-x/2}$$

- the parameter of the χ^2 is often called the degrees of freedom.
- we will re-visit this deistribution in the second half of this course.

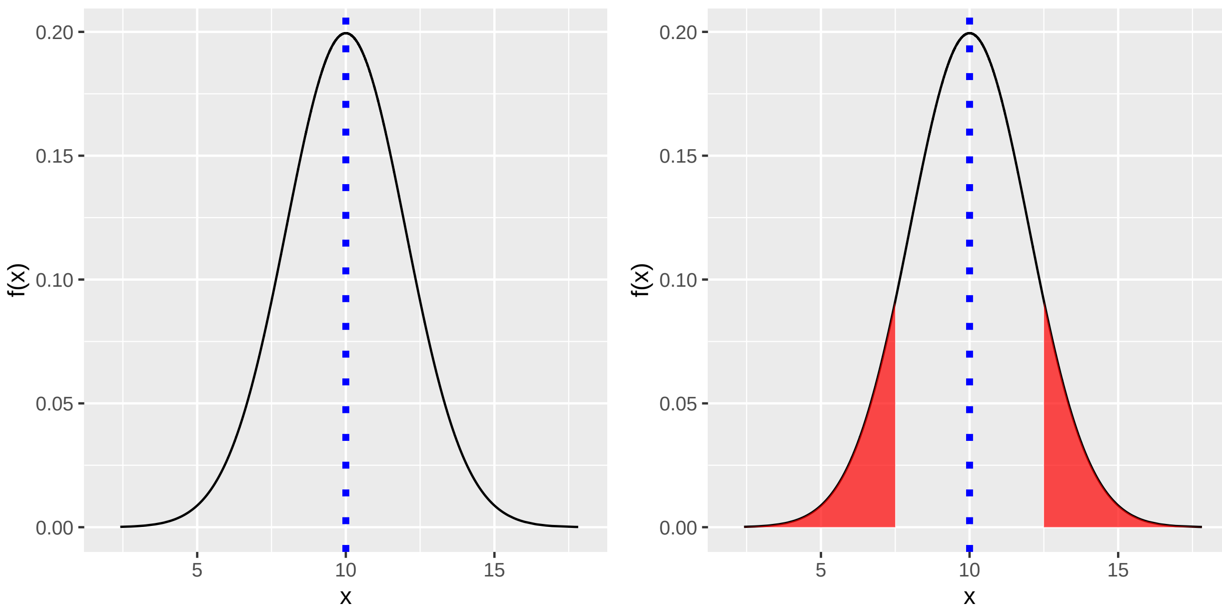
Normal distribution

A continuous random variable X has a *Normal* distribution with parameter μ and σ^2 if it's *pdf* is defined as

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad for \quad -\infty < x < \infty\,; \; \sigma^2 > 0$$

In notation, we say $X \sim N(\mu, \sigma^2)$

Density of a $N(\mu = 10, \sigma^2 = 4)$



Properties of Normal distribution:

- μ is the center of density (the dotted line on the graphs)
- σ^2 represents the spread of the density
- μ and σ^2 represent mean and variance of this distribution which we will learn next week.
- Normal density is symmetric around μ (the two shaded regions have the same probability)

Distribution function of $N(\mu, \sigma^2)$:

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \qquad for \quad -\infty < x < \infty$$

Unfortunately, this integration does not have a explicit solution. It is done in a different way through the help of a special case of this distribution.

Standard Normal Distribution

A Normal distribution with $\mu = 0$ and $\sigma^2 = 1$ is called the standard normal distribution.

Conventionally, a random variable following a standard normal distribution is denoted by Z

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad for \quad -\infty < z < \infty$$

In notation, we write $Z \sim N(0,1)$

Conventionally, the *pdf* of a standard normal is denoted using ϕ and the *distribution function* is denoted using Φ

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

and

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

Relationship between Normal and standard Normal

Let $X \sim N(\mu, \sigma^2)$.

Let Z be a transformation $Z = \frac{X-\mu}{\sigma}$

Then, $Z \sim N(0,1)$

Example: Suppose $X \sim N(\mu = 10, \sigma^2 = 4)$.

- $P(X > 10) =?$

- $P(X > 13) = P(\frac{X-\mu}{\sigma} > \frac{13-10}{2}) = P(Z > 1.5) = 1 - P(Z < 1.5) = 1 - \Phi(1.5)$

- $P(9 \leq X \leq 12) = P(\frac{9-10}{2} \leq \frac{X-\mu}{\sigma} \leq \frac{12-10}{2}) = P(-0.5 \leq Z \leq 1) = \Phi(1) - \Phi(-0.5)$

How to calculate $\Phi()$

- Way-1: Almost all the statistics text books have a table for standard normal probabilities at the back.
  - take a look at table A.3(page 792) of our current text book which gives the values of $\Phi(z)$ for different values of z.
- Way-2: In R, we can use a single command to calculate $\Phi(z)$ for any value of z
  - to calculate $\Phi(1.5)$, type *pnorm(1.5)*

```r
#Evaluating distribution function of standard normal at a=1.5
pnorm(1.5)
```

```
## [1] 0.9331928
```

For the examples from previous page,

- $P(X > 13) = 1 - \Phi(1.5) = 1 - 0.9331928 = 0.0668072$
- $P(9 \leq X \leq 12) = \Phi(1) - \Phi(-0.5) = 0.5328072$

```r
# using two pnorm() function in one line
pnorm(1) - pnorm(-0.5)
```

```
## [1] 0.5328072
```

- In the quiz/test/exam $\Phi(z)$ values will be provided to you for some relavent values of z.

Parcentile of a distribution

$\eta(p)$ is called the $100p^{th}$ percentile of a distribution if

$$F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(x)dx = p$$

- In plain language, $\eta(p)$ the value of x (a value on the x axis) at which the value of the cumulative distribution function is p.

Example: Let's revisit $Exp(\lambda = 2)$

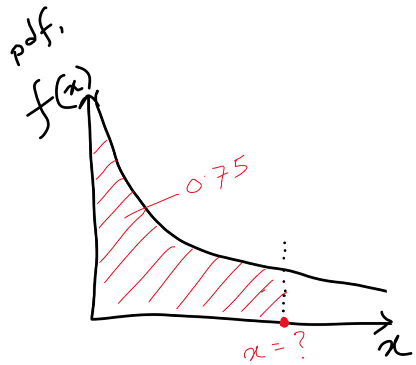
We have seen that F(1) = 0.8646647.

Hence, for $Exp(\lambda = 2)$, 0.86466647^{th} quantile or 86.46647^{th} percentile is 1.

Exercise: What is the 75^{th} percentile of $Exp(\lambda = 2)$ distribution?

$F(x) = 1 - e^{-2x} = 0.75 \implies x = ?$

On a graph:



Ans:

$$\begin{aligned} 1 - e^{-2x} &= 0.75 \\ \implies e^{-2x} &= 0.25 \\ \implies -2x &= ln(0.25) \\ \implies x &= \frac{ln(0.25)}{-2} \end{aligned}$$

Median

The 50^{th} *percentile* of any distribution is called the *Median* of that distribution.

Exercise: Find the median of a $Unif(\alpha, \beta)$ distribution.

median $\implies F(x) = \frac{1}{2}$

$$\begin{aligned} \int_{\alpha}^{x} \frac{1}{\beta - \alpha} dx &= \frac{1}{2} \\ \implies \frac{1}{\beta - \alpha} x|_{\alpha}^{x} &= \frac{1}{2} \\ \implies \frac{1}{\beta - \alpha}(x - \alpha) &= \frac{1}{2} \\ \implies x - \alpha &= \frac{\beta - \alpha}{2} \\ \implies x = \alpha + \frac{\beta - \alpha}{2} &= \frac{\alpha + \beta}{2} \end{aligned}$$

Note about chapter 4.5

- The distributions given in chapter 4.5 (Weibull/Lognormal/Beta) are not needed for this course.
- Please study them if you are planning to take advanced courses later on.

Note about chapter 4.6 and 4.7

We will come back to these two chapters at a later time in this term.

Homework

Chapter 4.1

1, 2, 4, 8, 10, 16

Chapter 4.2(Save these for next week)

17, 20, 25, 26, 29, 30, 32, 34, 36

Chapter 4.3

39, 40, 41, 42, 44, 46, 50, 53, 56, 58

Chapter 4.4

73, 74, 78