LECTURE 2B – UNIVARIATE CONTINUOUS DISTRIBUTIONS

Uniform, Exponential and Normal distributions

See Book Chapters 5 & 6

UNIVARIATE CONTINUOUS DISTRIBUTIONS

Uniform, Exponential and Normal distributions

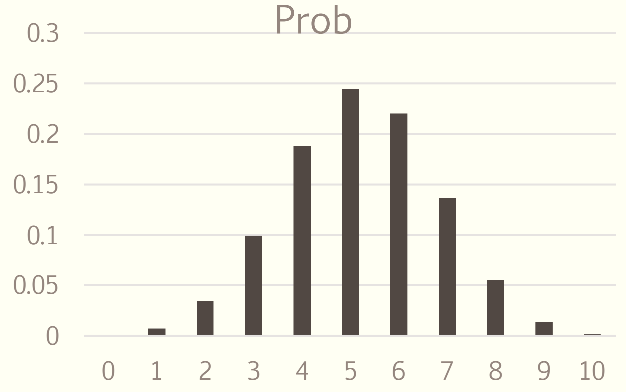
Continuous Random Variables

- Continuous random variables have many applications. Baseball batting averages, IQ scores, the length of time a long distance telephone call lasts, the amount of money a person carries, the length of time a computer chip lasts, and SAT scores are just a few. The field of reliability depends on a variety of continuous random variables.
- Whether a random variable is discrete or continuous depends on the problem you would like to model. For example:
  - If X is equal to the number of miles (to the nearest mile) you drive to work, then X is a discrete random variable. You count the miles. If X is the distance you drive to work, then you measure values of X and X is a continuous random variable.
  - If X is equal to the number of books in a backpack, then X is a discrete random variable. If X is the weight of a book, then X is a continuous random variable because weights are measured.
- How the random variable is defined is very important.

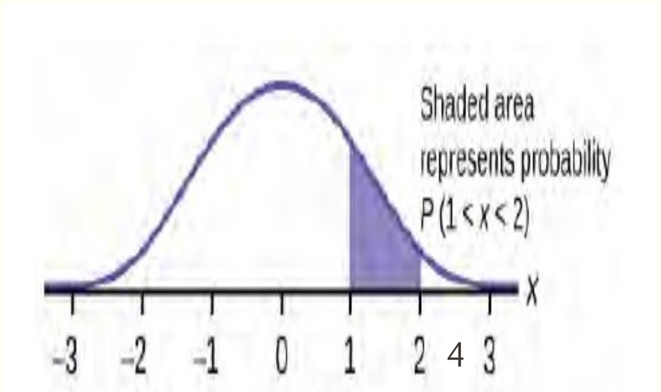
Continuous Random Variables – The Probability Density Function (PDF)

- Corresponding to the probability mass function (pmf) of a discrete random variable, we have a function called probability density function (pdf) for a continuous random variable, which is a curve.
- We use the symbol $f(x)$ to represent the pdf curve.
- Unlike the pmf for a discrete random variable, the pdf of a continuous random variable **does not give the probability**. That is, for a value of the random variable X, the pdf value **does not** represent a probability
- Instead, probability is given by area under the curve. i.e., **For continuous probability distributions, PROBABILITY = AREA UNDER THE PDF CURVE.**

pmf (probability mass function) of a discrete random variable



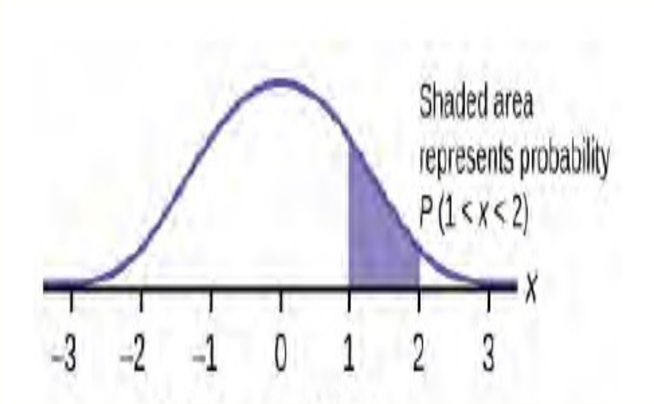
pdf (probability density function) of a continuous random variable



Continuous Random Variables – The Cumulative Density (Distribution) Function (CDF)

- To get the probabilities for a continuous random variable, we need the area under the pdf curve.
- The area under a pdf is given by the cumulative distribution function (**cdf**); so the cdf is used to evaluate probability. In the figure, the shaded area gives the probability $P(1 < X < 2) = P(1 \leq X \leq 2)$
- The symbol for the CDF is $F(x)$. $F(x)$ is the probability that the random variable is less than or equal to a value x. i.e., $F(x) = P(X \leq x)$. In the Figure, $P(1 \leq X \leq 2) = P(X \leq 2) - P(X \leq 1) = F(2) - F(1)$
- In general, calculus is needed to find the area under the curve (i.e., F(X)) for many probability density functions. That is, given a pdf function we need to use *integration* to get the cdf, because integration of a curve gives area under the curve. However, we will not use the calculus approach.
- We will find the area that represents probability by using geometry, *formulas for the cdf*, or *cdf tables.*

pdf (probability density function) of a continuous random variable



- $F(2) = P(X \leq 2)$ is the area from the left of -3 up to 2.
- $F(1) = P(X \leq 1)$ is the area from the left of -3 up to 1.
- So, the shaded area is $F(2) - F(1)$

Continuous Probability Distributions

- The important properties of continuous probability distributions are:
  - The outcomes are measured, not counted.
  - The entire area under the curve and above the x-axis is equal to one.
  - Probability is found for intervals of x values rather than for individual x values.
  - $P(x = c) = 0$ The probability that x takes on any single individual value is zero. The area below the curve, above the x-axis, and between $x = c$ and $x = c$ has no width, and therefore no area (area = 0). Since the probability is equal to the area, the probability is also zero.
  - $P(c < x < d)$ is the same as $P(c \leq x \leq d)$ because probability is equal to area.
  - $P(c \leq x \leq d)$ is the probability that the random variable X is in the interval between the values c and d. $P(c \leq x \leq d)$ is the area under the curve, above the x-axis, to the right of c and the left of d and is given by $P(X \leq d) - P(X \leq c) = F(d) - F(c)$.

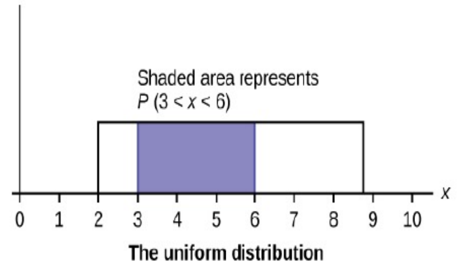


Figure 5.2 The graph shows a Uniform Distribution with the area between $x = 3$ and $x = 6$ shaded to represent the probability that the value of the random variable X is in the interval between three and six.

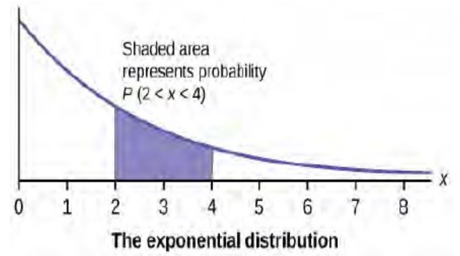


Figure 5.3 The graph shows an Exponential Distribution with the area between $x = 2$ and $x = 4$ shaded to represent the probability that the value of the random variable X is in the interval between two and four.

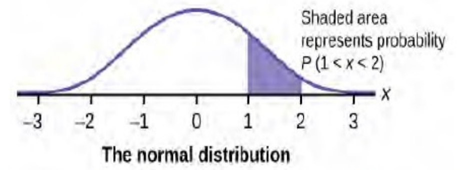


Figure 5.4 The graph shows the Standard Normal Distribution with the area between $x = 1$ and $x = 2$ shaded to represent the probability that the value of the random variable X is in the interval between one and two.

THE UNIFORM DISTRIBUTION

Continuous Distributions

Uniform Distribution

- The **continuous uniform distribution** is a distribution in which *all events represented by intervals of the same length are equally probable*. i.e., $P(a \le x \le a+1) = P(a + 1 \le x \le a+2)$. The uniform distribution is also called a *rectangular distribution*

  - $X \sim U(\mathbf{a}, \mathbf{b})$ where **a** = the lowest value of x and **b** = the highest value of x.

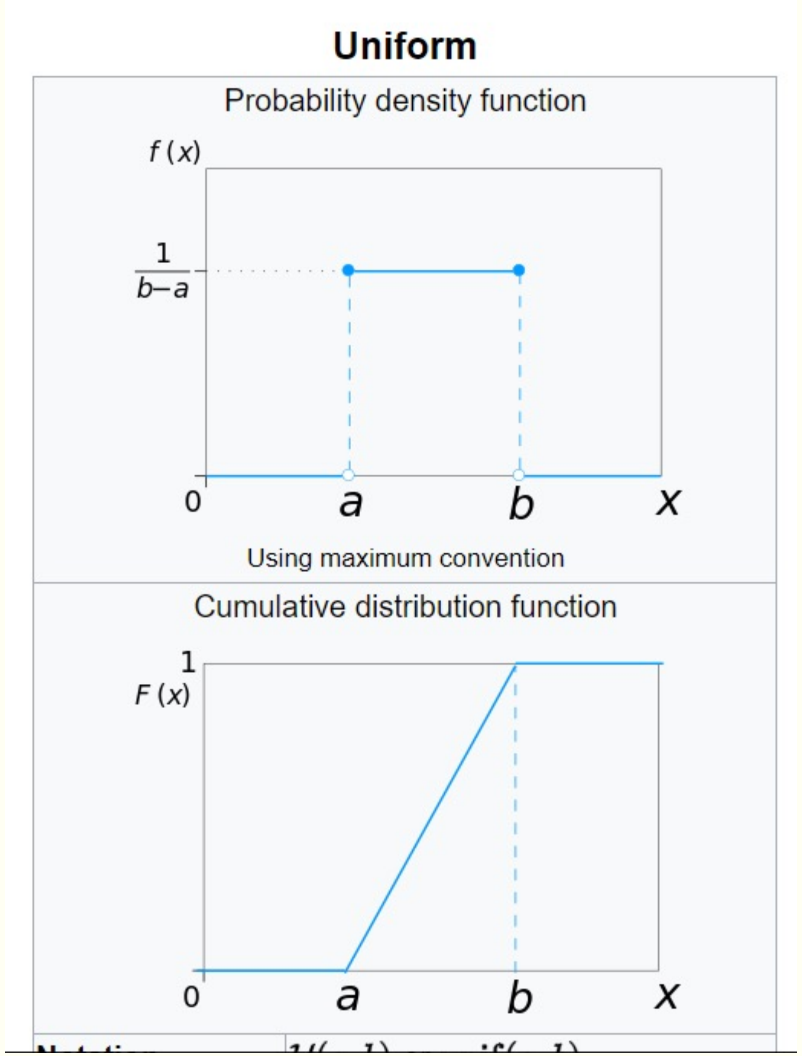
  - The probability density function (pdf) is:
    - $f(X) = 1/(\mathbf{b} - \mathbf{a})$ for $\mathbf{a} \le x \le \mathbf{b}$ and 0 **elsewhere**.

  - The cumulative distribution function (cdf) is:
    - $F(X) = (x - \mathbf{a}) / (\mathbf{b} - \mathbf{a})$ for $\mathbf{a} \le x \le \mathbf{b}$ which gives us the probability $\mathbf{P}(X \le x) = P(a \le X \le x)$.

  - Formulas for the theoretical mean and standard deviation are
    - Expected Value = $\mu = (\mathbf{a} + \mathbf{b})/2$ and

    - Standard Deviation = $\sigma = \sqrt{\frac{(\mathbf{b} - \mathbf{a})^2}{\mathbf{12}}} = (\mathbf{b} - \mathbf{a})/\sqrt{\mathbf{12}}$

  - The k^{th} percentile represents the value of the random variable X for which
$\mathbf{P}(X \le x) = F(X) = (x - \mathbf{a})/(\mathbf{b} - \mathbf{a}) = k/100$. i.e., $x = a + (b - a)$



Practice with calculator

Uniform Distribution

- Consider a random variable X ~ Uniform(2, 8.8)
  - **a** = 2 and **b** = 8.8, then P(**3** ≤ x ≤ **6**) = F(6) – F(3) = (6 – 2)/6.8 – (3 - 2)/6.8 = 3/6.8 = 0.441
  - Geometrically, in the above, the width of the rectangle is 3 and the height is 1/6.8 so the area = 3/6.8 = 0.441
  - The mean of the distribution is **μ** = (**a** + **b**)/2 = 10.8/2 = 5.4
  - The standard deviation **σ** = $\sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(8.8-2)^2}{12}}$ = 3.85
  - The **median** or **50th** percentile = x = a + (b – a) *0.5 = 2 + (8.8 – 2) *0.5 = 5.4
  - The **80th** percentile = x = a + (b – a) *0.8 = 2 + (8.8 – 2) *0.8 = 7.44
  - The **mode** is any value between **a** and **b** (inclusive)
  - Since the **mean** = **median**, we say that the distribution has a **skewness** of 0. That is, it is neither right- nor left-skewed. Its skewness = 0.

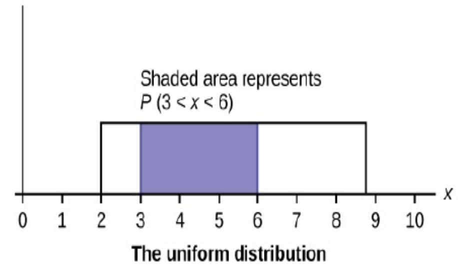
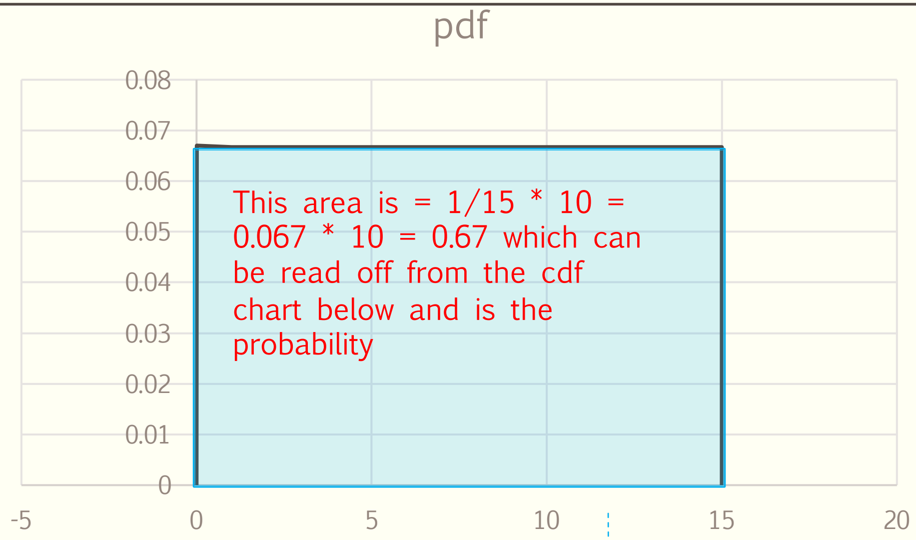


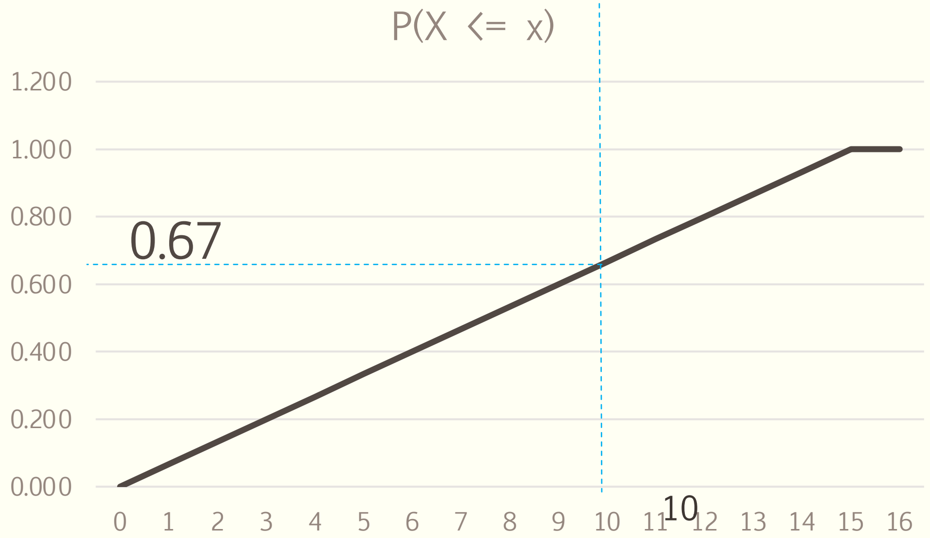
Figure 5.2 The graph shows a Uniform Distribution with the area between x = 3 and x = 6 shaded to represent the probability that the value of the random variable X is in the interval between three and six.

Practice with calculator

Book Example 5.4 – Page 316

- **Example:** The amount of time, in minutes, that a person must wait for a bus is uniformly distributed between zero and 15 minutes, inclusive. Here, a = 0 and b = 15
  - Probability "that the person waited for (**at most**) 10 minutes" = $P(X \leq 10) = F(10) = (10 - 0)/(15 - 0) = 0.67$
  - Probability "that the person waited *at least* 10 minutes" = $1 - P(X \leq 10) = 0.33$
  - Probability "that the person waited between 5 and 7 minutes" = Probability "that the person waited ≤ 7 minutes" – "Probability "that the person waited ≤ 5 minutes" = $P(X \leq 7) - P(X \leq 5) = F(7) - F(5) = 7/15 - 5/15 = 2/15$.
  - Probability "that the person has to wait for at least 5 more minutes" given "that the person has already waited for 4 minutes" =
    - P(" at least 5 more minutes"|"4 minutes")
    - $= P(X \geq 9)|X \geq 4)$
    - $= P((X \geq 4) \text{ and } (X \geq 9)) / P(X \geq 4)$
    - $= P(X \geq 9) / P(X \geq 4) = (6/15)/(11/15) = 6/11$
  - You can also look at it as follows: Since the person has already waited for 4 minutes, we have a new uniform distribution U (4, 15). In this distribution, $P(X \geq 9) = 6/11$.





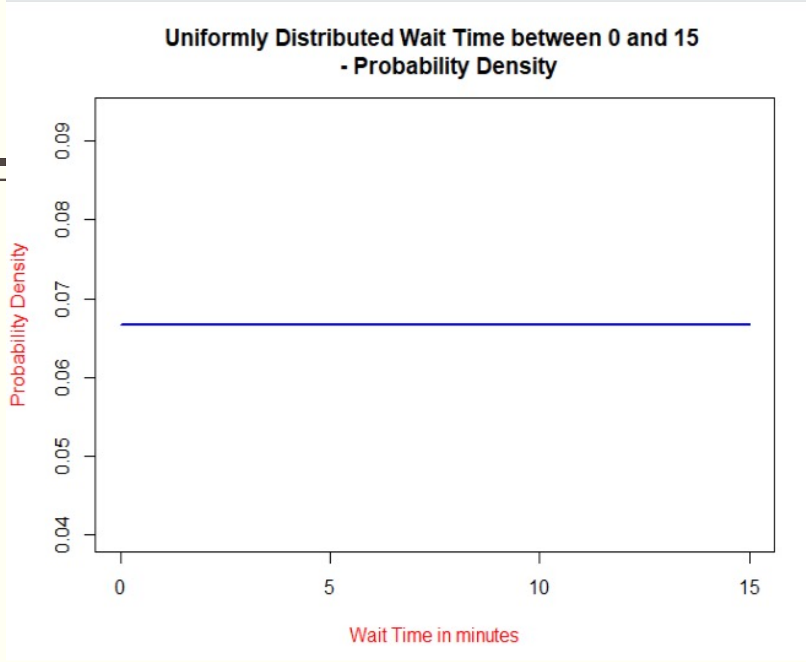
X	P(X <= x)
1	0.067
2	0.133
3	0.200
4	0.267
5	0.333
6	0.400
7	0.467
8	0.533
9	0.600
10	0.667
11	0.733
12	0.800
13	0.867
14	0.933
15	1.000

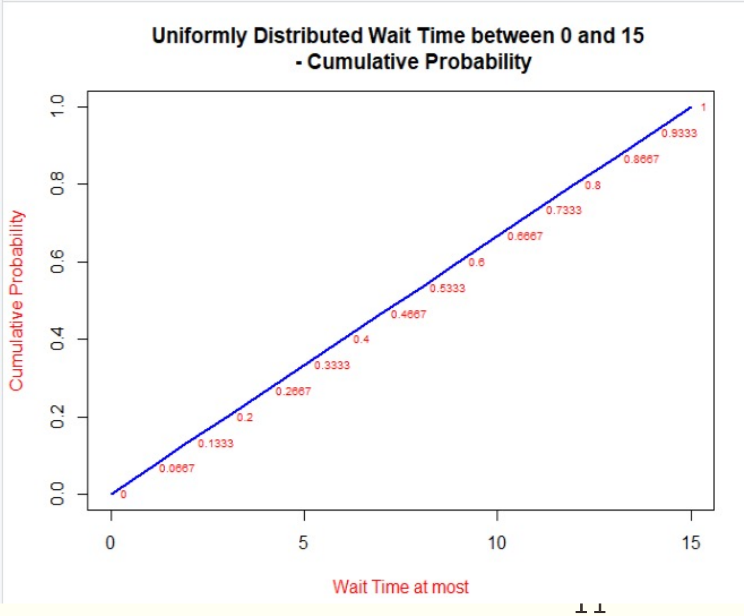
Uniform.R

R-commands: for X ~ U(low, high)

- *pdf = dunif(x, low, high)*
- *cdf = P(X <= x) = punif(x, low, high)*
- *k^{th} percentile/quantile = qunif(k/100, low, high)*

```
#table of probabilities for select values of X
nsize = 15
result <- vector("numeric", nsize+1)
cum_result <- vector("numeric", nsize+1)
for (i in 0:nsize) {
  result[i+1] <- dunif(i, 0, 15, log = FALSE)
  cum_result[i+1] <- punif(i, 0, 15, log = FALSE)
}
# Plot the probabilities
len_result <- length(result)
indx = len_result - 1
#
#Plot the probability mass function
plot((0:indx),result,
      xlim=c(0, 15),                                  #-- Specify Limits for the X-axis
      type = "l",
      main = "Uniformly Distributed Wait Time between 0 and 15
     - Probability Density",
      xlab = "Wait Time in minutes",
      ylab = "Probability Density",
      col  = "blue",
      col.lab ="red",
      lwd=2)
#
# Plot the cumulative probabilities
plot((0:indx),cum_result,
      xlim=c(0, 15),                                  #-- Specify Limits for the X-axis
      type = "l",
      main = "Uniformly Distributed Wait Time between 0 and 15
     - Cumulative Probability",
      xlab = "Wait Time at most",
      ylab = "Cumulative Probability",
      col  = "blue",
      col.lab ="red", lwd=2)
text((0:indx), cum_result[1:len_result], round(cum_result[1:len_result], 4), cex=0.6, pos=4, col="red")
#
```





Uniform.R

```r
# Probability that the person waited for (at most) 10 minutes
#Cumulative probability for i = 10
print(paste("Probability that the person waited for (at most) 10 minutes = ",
            round(punif(10, 0, 15, log = FALSE),4)))
#
# Probability that the person waited between 5 and 7 minutes
print(paste("Probability that the person waited between 5 and 7 minutes = ",
            round(punif(7, 0, 15, log = FALSE) - punif(5, 0, 15, log = FALSE),4)))
#
# 10th, 50th, and 90th percentiles
print(paste("The 10th percentile is ",qunif(0.10, 0, 15, lower.tail = TRUE, log.p = FALSE), " minutes"))
print(paste("The 50th percentile is ",qunif(0.50, 0, 15, lower.tail = TRUE, log.p = FALSE), " minutes"))
print(paste("The 90th percentile is ",qunif(0.90, 0, 15, lower.tail = TRUE, log.p = FALSE), " minutes"))
#
```

```
[1] "Probability that the person waited for (at most) 10 minutes =  0.6667"
```

```
[1] "Probability that the person waited between 5 and 7 minutes =  0.1333"
```

```
> # 10th, 50th, and 90th percentiles
> print(paste("The 10th percentile is ",qunif(0.10, 0, 15, lower.tail = TRUE, log.p = FALSE), " minutes"))
[1] "The 10th percentile is  1.5  minutes"
> print(paste("The 50th percentile is ",qunif(0.50, 0, 15, lower.tail = TRUE, log.p = FALSE), " minutes"))
[1] "The 50th percentile is  7.5  minutes"
> print(paste("The 90th percentile is ",qunif(0.90, 0, 15, lower.tail = TRUE, log.p = FALSE), " minutes"))
[1] "The 90th percentile is  13.5  minutes"
```

Practice with calculator

Book Problem 5.76 – Page 349

- According to a study by Dr. John McDougall of his live-in weight loss program at St. Helena Hospital, the people who follow his program lose between 6 and 15 pounds a month until they approach trim body weight. *Let's suppose that the weight loss is uniformly distributed.* We are interested in the weight loss of a randomly selected individual following the program for one month.

a. Define the random variable. X = weight loss of a randomly selected individual following the program for one month
b. X ~ Uniform(6, 15)
c. The pdf f(X) = 1/(**9**) for **6** ≤ x ≤ **15** and 0 **elsewhere**; the cdf F(X) = (x – 6)/9 for for **6** ≤ x ≤ **15** and 0 **elsewhere**;
d. μ = (15 + 6) / 2 = 10.5
e. σ = (15 - 6)/√12 = 2.598
f. Find the probability that the individual lost more than ten pounds in a month.
P(X > 10) = 1 – P(X <= 10) = 1 – F(10) = 1 – 4/9 = 5/9
g. Suppose it is known that the individual lost more than ten pounds in a month. Find the probability that he lost less than 12 pounds in the month.

One way: P(X ≤ 12 | X > 10). Since it is given X > 10, the lower limit becomes 10 and the upper limit is 15 for a new uniform distribution U(10, 15). In other words, the pdf becomes f(X) = 1/5.
P(X ≤ 12) in this new uniform distribution is: (12 – 10)/5 = 2/5.

Another way: P(X ≤ 12 | X > 10) = P((X ≤ 12) and (X > 10)) / P(X > 10) = P(10 < X ≤ 12) / P(X > 10) = (2/9) / (5/9) = 2/5

THE EXPONENTIAL DISTRIBUTION

Continuous Distributions

The Exponential Distribution

- The exponential distribution is often concerned with the amount of time until some specific event occurs, *though not always.*
  - The amount of time (beginning now) until an earthquake occurs has an exponential distribution
  - The time between packets arriving at a node on the internet is assumed to be exponential as well as the time taken by the node to service those packets
- The exponential distribution also represents instances where the value of something decays at a constant rate.
  - The lifetime of a bulb or a car battery are exponentially distributed *Field of reliability)
  - The value of the change in your pocket or purse approximately follows an exponential distribution.
- There is a close relationship between the Poisson distribution and the exponential distribution. If events follow a Poisson process (i.e., they occur continuously and independently at a constant rate), then the *time between the events* follows an exponential distribution.

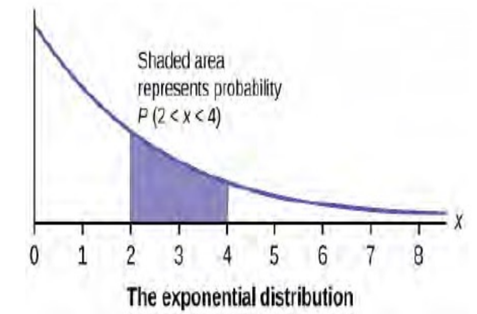


Figure 5.3 The graph shows an Exponential Distribution with the area between $x = 2$ and $x = 4$ shaded to represent the probability that the value of the random variable X is in the interval between two and four.

The Exponential Distribution

- The **exponential distribution** (also known as **negative exponential distribution**) is the probability distribution that describes the *time between events* in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate.

- In the Poisson process **λ** = average number of occurrences per time period.

- X ~ exp(**λ**) where **λ** > **0** is called the *rate parameter*. X takes on values in from 0 to ∞.

- The probability density function (**pdf**):
  $f(X) = \lambda e^{-\lambda x}$ for $\mathbf{0} \le x \le \infty$ and is 0 **elsewhere**.

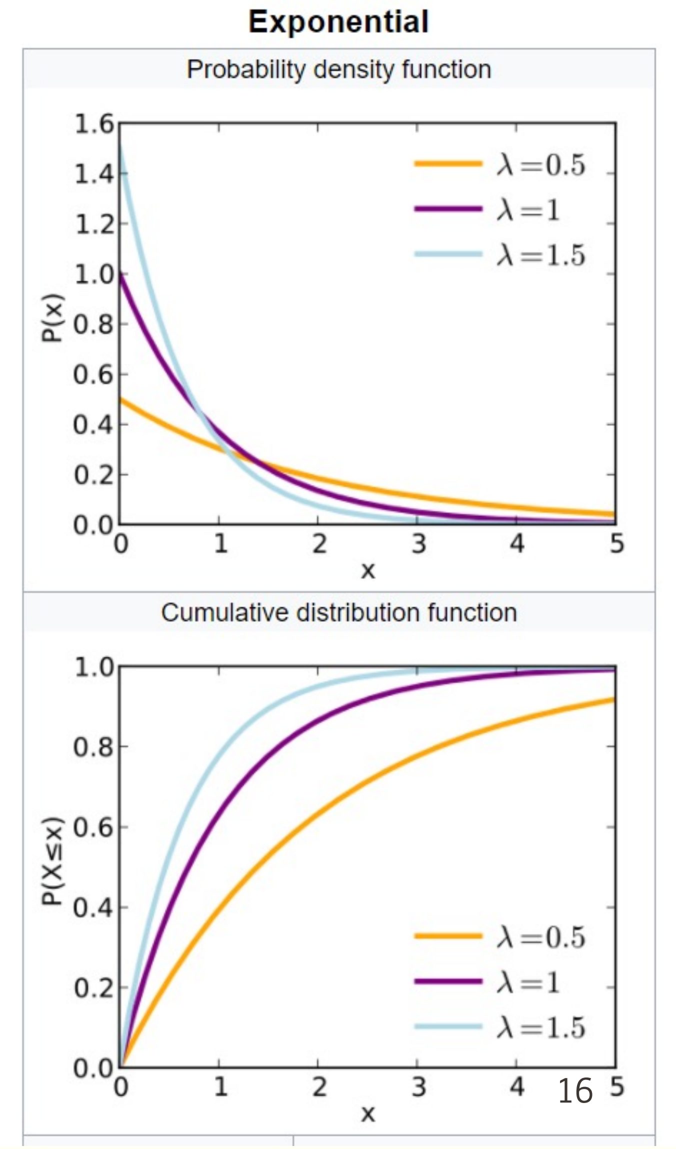
- The **cdf** is given by:
  $F(X) = 1 - e^{-\lambda x}$ for $\mathbf{0} \le x \le \infty$ which gives us the probability $P(X \le x)$.

- Formulas for the theoretical *mean* and *standard deviation* with units average time between occurrences.
  - $\mu = 1/\lambda$ and $\sigma = 1/\lambda$

- k^{th} **Percentile:** The value of the random variable X for which:

  $F(X) = 1 - e^{-\lambda x} = k/100$.
  i.e., $e^{-\lambda x} = 1 - k/100$, so $-\lambda x = \ln(1 - k/100)$ and $x = -\ln(1 - k/100)/\lambda$

- The *median* is given by $\ln(2)/\lambda = 0.693*\mu$, and *mode* = 0. This means **median is less than the mean and the distribution is right-skewed**, with a longer right tail. The skewness coefficient is 2 and independent of the parameter (**λ**).



Practice with calculator

Book Example 5.7 – Page 323

- **Example:** The amount of time spouses shop for anniversary cards can be modeled by an exponential distribution with the average amount of time equal to eight minutes.
- Average **time** between occurrences = 8 minutes which is the mean $\boldsymbol{\mu} = (1/\boldsymbol{\lambda})$. Therefore, $\boldsymbol{\lambda} = 1/8$ is the number of occurrences per minute.
- Probability "that the person spent (at most) 10 minutes" shopping for a card = $P(x \leq 10) = (1 - \mathbf{e}^{-\lambda x}) = 1 - \mathbf{e}^{-10/8} = 0.7135$
- Probability "that the person spent at least 8 minutes" = $1 - P(x \leq 8) = 1 - (1 - \mathbf{e}^{-8/8}) = \mathbf{e}^{-8/8} = 0.3679$
- Probability "that the person spent between 5 and 7 minutes"
  $= F(7) - F(5)$
  = Probability "that the person spent ≤ 7 minutes" – "Probability "that the person spent ≤ 5 minutes"
  $= 1 - e^{-7/8} - (1 - e^{-5/8})$
  $= e^{-5/8} - e^{-7/8} = 0.1184$
- **Percentile:** The k^{th} percentile represents the value of the random variable X for which $F(X) = 1 - \mathbf{e}^{-\lambda x} = k/100$.
  $x = -\ln(1 - k/100)/\boldsymbol{\lambda}$
- Hence the **median** is given by $-\ln(0.5)/\boldsymbol{\lambda} = -\ln(0.5)*8 = 5.545 = 0.6931*$**mean**
- The **mode** of an exponential distribution is 0

Exponential.R

R-commands: for X ~ exp(λ)

- *pdf = dexp(x, lambda)*
- *cdf = P(X <= x) = pexp(x, lambda)*
- *k^{th} percentile/quantile = qexp(k/100, lambda)*

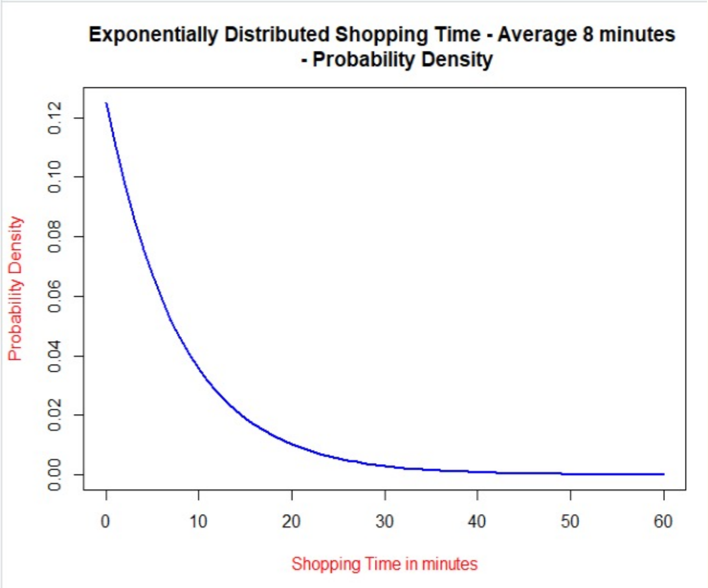
```
#
#Exponential Distribution
#Example: The amount of time spouses shop for anniversary cards can be
# modeled by an exponential distribution # with the average amount of time
#equal to eight minutes. Hence, lambda parameter = number of occurrences per minute = 1/8
# X = rv = time elapsed before a card is bought
# Let us plot the time taken to shop for a card over a time horizon of 60 minutes
#
#dexp(x, rate = 1, log = FALSE) - Not used in c
#pexp(q, rate = 1, lower.tail = TRUE, log.p = FALSE)
#qexp(p, rate = 1, lower.tail = TRUE, log.p = FALSE)
#
print(paste("Probability that the person spent (at most) 10 minutes shopping for a card is ",
            round(pexp(10, 0.125, lower.tail = TRUE, log= FALSE),4) ))
#
print(paste("Probability that the person spent between 5 and 7 minutes shopping for a card is ",
            round(pexp(7, 0.125, lower.tail = TRUE, log= FALSE) - pexp(5, 0.125, lower.tail = TRUE, log= FALSE),4) ))
```

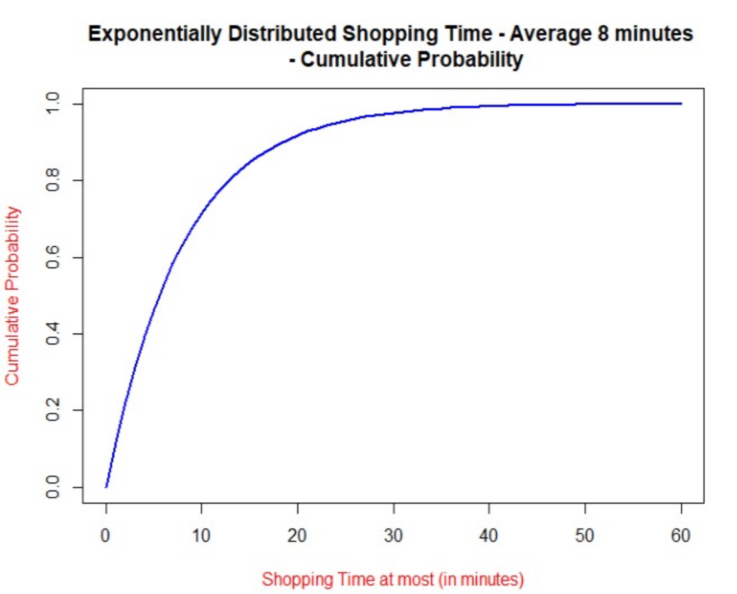
```
> print(paste("Probability that the person spent (at most) 10 minutes shopping for a card is ",
+             round(pexp(10, 0.125, lower.tail = TRUE, log= FALSE),4) ))
[1] "Probability that the person spent (at most) 10 minutes shopping for a card is  0.7135"
> #
> print(paste("Probability that the person spent between 5 and 7 minutes shopping for a card is ",
+             round(pexp(7, 0.125, lower.tail = TRUE, log= FALSE) - pexp(5, 0.125, lower.tail = TRUE, log= FALSE),4) ))
[1] "Probability that the person spent between 5 and 7 minutes shopping for a card is  0.1184"
```

Exponential.R

```
#table of probabilities for select values of X from 0 to 59.
nsize = 60
result <- vector("numeric", nsize+1)
cum_result <- vector("numeric", nsize+1)
#
for (i in 0:nsize) {
  result[i+1] <- round(dexp(i, 0.125, log = FALSE),4)
  cum_result[i+1] <- round(pexp(i, 0.125, lower.tail = TRUE, log = FALSE),4)
}
# Plot the probabilities
len_result <- length(result)
indx = len_result - 1
#Plot the probability mass function
plot((0:indx),result,
      type = "l",
      main = "Exponentially Distributed Shopping Time - Average 8 minutes
     - Probability Density",
      xlab = "Shopping Time in minutes",
      ylab = "Probability Density",
      col  = "blue",
      col.lab ="red",
      lwd=2)
# Plot the cumulative probabilities
plot((0:indx),cum_result,
      type = "l",
      main = "Exponentially Distributed Shopping Time - Average 8 minutes
     - Cumulative Probability",
      xlab = "Shopping Time at most (in minutes)",
      ylab = "Cumulative Probability",
      col  = "blue",
      col.lab ="red", lwd=2)
#
# 5th, 50th, and 95th percentiles
print(paste("The 5th percentile is ", round(qexp(0.05, 0.125, lower.tail = TRUE, log.p = FALSE),4)))
print(paste("The 50th percentile is ", round(qexp(0.50, 0.125, lower.tail = TRUE, log.p = FALSE),4)))
print(paste("The 95th percentile is ", round(qexp(0.95, 0.125, lower.tail = TRUE, log.p = FALSE),4)))
```

```
> # 5th, 50th, and 95th percentiles
> print(paste("The 5th percentile is ", round(qexp(0.05, 0.125, lower.tail = TRUE, log.p = FALSE),4)))
[1] "The 5th percentile is  0.4103"
> print(paste("The 50th percentile is ", round(qexp(0.50, 0.125, lower.tail = TRUE, log.p = FALSE),4)))
[1] "The 50th percentile is  5.5452"
> print(paste("The 95th percentile is ", round(qexp(0.95, 0.125, lower.tail = TRUE, log.p = FALSE),4)))
[1] "The 95th percentile is  23.9659"
```





The Exponential Distribution - Memoryless

- **Memoryless:** An exponentially distributed random variable X obeys the relation
  $P(X > s + t \mid X > s) = P(X > t)$, for all $s, t \geq 0$.
  Example: *The life of a light bulb is exponentially distributed. It has been burning for s hours. What is the probability that it burns for another t hours? It is the same as the probability that it burns for t hours!!!*
  But, **$P(X > s + t \mid X > s) = P((X > s + t) \text{ AND } (X > s)/P(X > s)$**
  So, **$P(X > s + t \mid X > s) = P(X > s + t)/P(X > s)$**
  Since this is = $P(X > t)$
  We have $P(X > s + t)/P(X > s) = P(X > t)$
  We can also say as $P(X > s + t) = P(X > s) * P(X > t)$ for memoryless

- Proof:
  - **We know that CDF**: $F(X) = P(X \leq x) = 1 - \mathbf{e^{-\lambda x}}$ for $\mathbf{0} \leq x \leq \infty$
  - So: **$P(X > s + t)$** $= 1 - (1 - e^{-(s+t)/\lambda}) = e^{-(s+t)/\lambda}$
    $P(X > s)$ $= 1 - (1 - e^{-s/\lambda}) = e^{-s/\lambda}$
  - Then: **$P(X > s + t)/P(X > s)$ =** $(e^{-(s+t)/\lambda})/(e^{-s/\lambda}) = e^{-t/\lambda}$ **= $P(X > t)$** so it is memoryless.

- That is:
  - Probability "that the person spends (at most) t more minutes" given "that the person has already spent (at least) s minutes" = P ("that the person spends (at least) t more minutes") = $e^{-t/\lambda}$
  - Note that it also works for **$P(X \leq s + t \mid X > s) = P(X \leq t)$,** for all $s, t \geq 0$ and is equal to $(1 - e^{-t/\lambda})$

- The exponential distribution and the geometric distribution are the only memoryless probability distributions.

The Exponential Distribution - Memoryless

- <u>Book Example 5.7 – Page 323</u>.
- <u>**Example:**</u> The amount of time spouses shop for anniversary cards can be modeled by an exponential distribution with the average amount of time equal to eight minutes.
- Average <u>**time**</u> between occurrences = 8 minutes which is the mean = 1/**λ**. Therefore, **λ** = 1/8.
- Suppose that a person has been shopping for at most 5 minutes. What is the probability that they will spend 2 more minutes?
- Probability that the "person spends 2 more minutes", given that the "person already spent 5 minutes" =
  - Probability that "person spends 5 minutes and 7 minutes"/ Probability that the "person spends (at most) 5 minutes"
    = (Probability that "person spends (at most) 7 minutes""/ Probability that the "person spends (at most) 5 minutes"
    = ($\mathbf{e}^{-7/8}$) / ($\mathbf{e}^{-5/8}$) = $\mathbf{e}^{-2/8}$ = Probability that the "person spends (at most) 2 minutes".

Book Problem 5.100 – Page 352

Practice with calculator

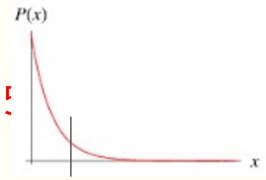
- A web site experiences traffic during normal working hours at a rate of 12 visits per hour (12/60 = 0.2/minute). Assume that the duration between visits has the *exponential distribution*. That is, the *X = duration between visits* ~ Expon(λ = 0.2/minute)
  a. Find the probability that the duration between two successive visits to the web site *is more than ten minutes*.

```
> print(paste("probability that the duration between two successive visits to the web site is more than 10 minutes",
+             1 - round(pexp(10, 0.2, lower.tail = TRUE, log = FALSE),4)))
[1] "probability that the duration between two successive visits to the web site is more than 10 minutes 0.1353"
```

  b. The top 25% of *durations between visits* are at least how long?

```
> print(paste("The top 25% of durations between visits are at least - in minutes",
+             round(qexp(0.75, 0.2, lower.tail = TRUE, log = FALSE),4)))
[1] "The top 25% of durations between visits are at least - in minutes 6.9315"
```

931



  c. Suppose that 20 minutes have passed since the last visit to the web site. What is the probability that the next visit will occur within the next 5 minutes?

```
> print(paste("probability that the duration of next visit is less than 5 minutess",
+             round(pexp(5, 0.2, lower.tail = TRUE, log = FALSE),4)))
[1] "probability that the duration of next visit is less than 5 minutess 0.6321"
```

= 1 - $e^{-\lambda x}$ = 1 – exp(-1) =

  d. Find the probability that less than 7 visits occur within a one hour period.

```
> print(paste("probability less than 7 visits in 1 hour",
+             round(ppois(7, 12, lower.tail = TRUE, log = FALSE),4)))
[1] "probability less than 7 visits in 1 hour 0.0895"
```

THE NORMAL DISTRIBUTION

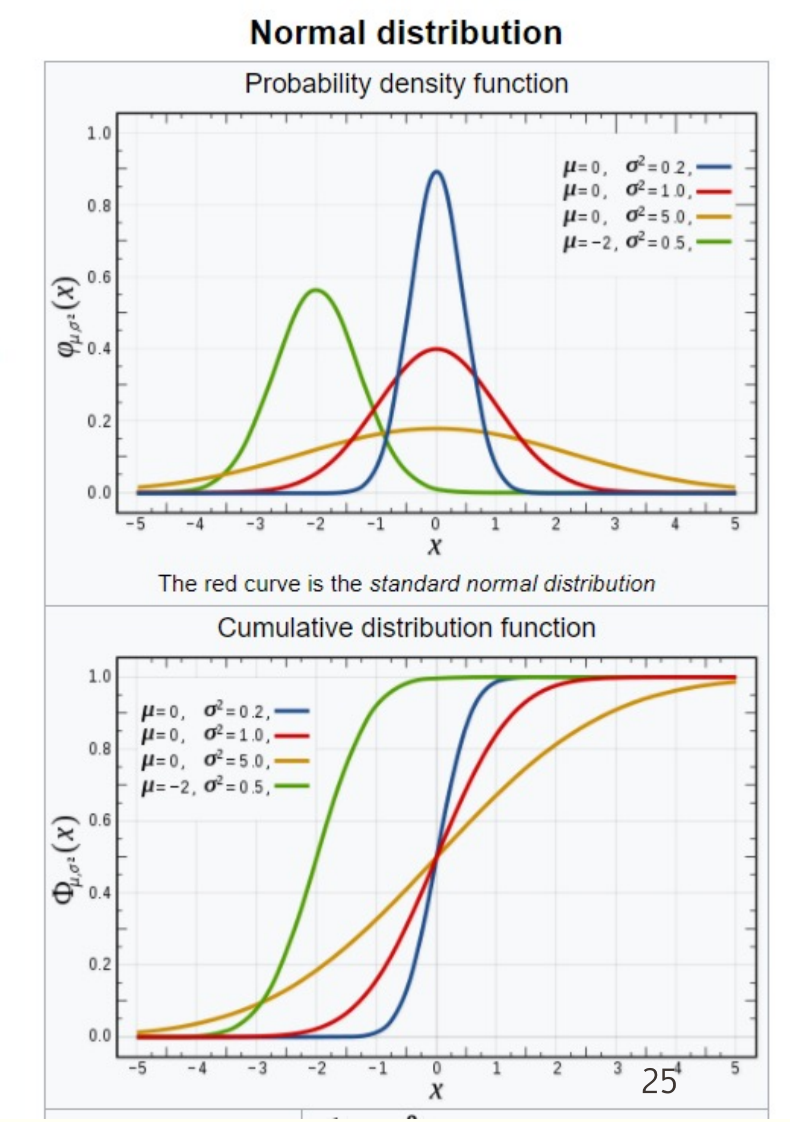
Continuous Distributions

The Normal Distribution

- The normal, a continuous distribution, is the most important of all the distributions. It is widely used and even more widely abused.
- Its graph is bell-shaped. You see the bell curve in almost all disciplines. Some of these include psychology, business, economics, the sciences, nursing, and, of course, mathematics. Some of your instructors may use the normal distribution to help determine your grade. Most IQ scores are normally distributed.
- The normal distribution is extremely important, but it cannot be applied to everything in the real world.

The Normal Distribution

- The normal distribution has two parameters (two numerical descriptive measures), the mean (**μ**) and the standard deviation (**σ**). If *X* is a quantity to be measured that has a normal distribution with mean (**μ**) and standard deviation (**σ**), we designate this by writing X ~ N(**μ**, **σ**)
- The probability density function (**pdf**) is a rather complicated function. **Do not memorize it**. It is not necessary.
- Pdf = f(X) = $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- The normal distribution is **symmetric** i.e., its mean = median. In fact, its mean = median = mode.
- Hence it has **skewness** = 0. In fact, the skewness of all distributions are measured relative to the normal distribution.
- For the Uniform and Exponential distributions, we had simple closed-form expressions for **cdf** that directly gave use the probabilities.
- For the normal, the **cdf** is very complicated and we cannot use it to directly give us the probability (area under the curve). We will use tables to get the cdf, after standardization.



Standardizing Random Variables

- The concept of **standardizing random variables** is important in many areas of statistics. Standardizing a variable means that you *subtract the mean* of the random variable *from every value of the random variable* and *divide each of the subtracted value by the standard deviation of the random variable.*

- Suppose a random variable X has a mean of ($\boldsymbol{\mu}$ =) 3 and a standard deviation of ($\boldsymbol{\sigma}$ =) 2. Then, the standardized random variable (usually denoted as **Z** or $\mathbf{Z_x}$ is given by)

  $Z = (X - \boldsymbol{\mu})/\boldsymbol{\sigma} = (X - \mathbf{3})/\mathbf{2}.$

- Note that because we subtract the mean of X from every value of X, the **mean of Z is always 0**.

- Similarly, since we divide the result by the standard deviation of X, the **standard deviation of Z is always 1**.

- This is true regardless of the type of random variable X.

- Further, since the mean and standard deviation are *constants*, the **type of distribution Z will be the same as X**. **That is,** $Z = (X - \boldsymbol{\mu})/\boldsymbol{\sigma} \sim \mathbf{N(0,\ 1)}$ **if** $X \sim N(\boldsymbol{\mu},\ \boldsymbol{\sigma})$
  - **Examples**:

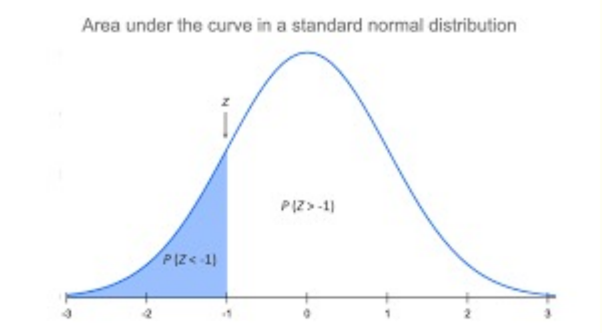
    If X is an *exponential* distribution then Z becomes a *standard exponential distribution*

    If X is a *binomial* distribution then Z becomes a *standard binomial distribution*

The Normal and Standard Normal Distributions

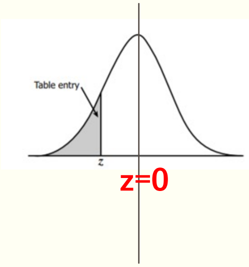
- If X ~ N(**μ**, **σ**) then the transformed variable **Z** = (X - **μ**)/**σ** ~ N(**0**, **1**). This is called the **standard (or standardized) normal distribution**.
- The Pdf of the normal distribution X with mean **μ** and standard deviation **σ** = f(X) = $\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- The mean of Z is 0 and its standard deviation is 1. Hence, The pdf of the standard normal distribution = f(Z) = $\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$
- Conversely, if **Z** = (X - **μ**)/**σ** then X = **μ** + **σ**Z
- For any normal random variate X it can be seen that:
  - F(x) = P(X ≤ x) = P(**μ** + **σz** ≤ x) = P(z ≤ (x - **μ**)/**σ**)
  - For example if X is normally distributed with mean of 4 and standard deviation of 2, to find P(X ≤ 3)

    F(3) = P(X ≤ 3) = P(Z ≤ (3 - 4)/2) = P(Z ≤ -0.5)
  - That is, P(X ≤ 3) = P(Z ≤ -0.5)



- The advantage of the standard normal distribution is that *<u>we only need the CDF of probabilities for Z</u>*. This table for the cdf of Z is called the *standard normal table.*

The Standard Normal Distribution

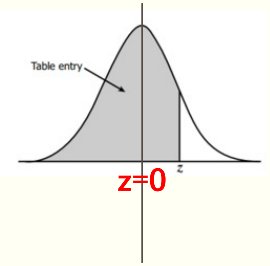


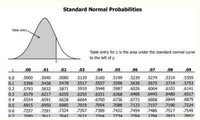
STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08691	.08534	.08379	.08226
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
-0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
-0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
-0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
-0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
-0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
-0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
-0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
-0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
-0.1	.46017	.45620	.45224	.44828	.44433	.44038	.43644	.43251	.42858	.42465
-0.0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997



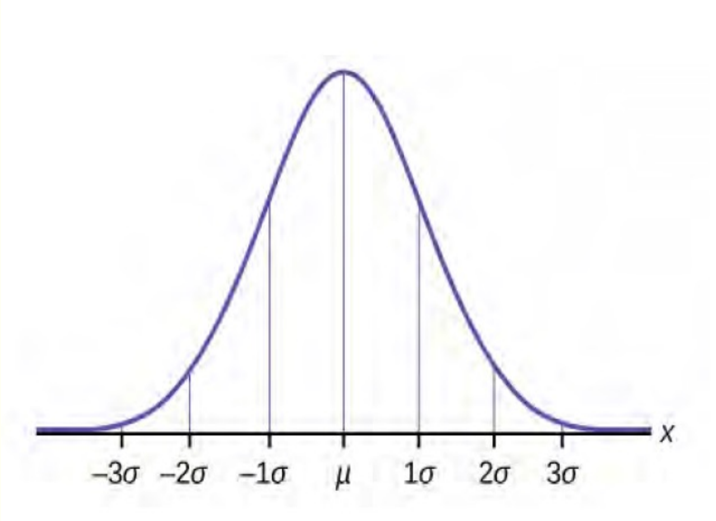
Practice with Table 

Normal to Standard Normal to Get probabilities

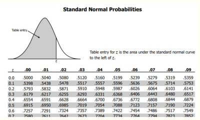
- **Example:**
  - Let X be a normal distribution with mean of 4 and standard deviation of 2. Use the **standard normal tables** to find:
  - $P(X \leq 5)$
    - $P(Z \leq (5 - 4)/2) = P(Z \leq 0.5); = F(0.5)$
    - From tables, the probability is 0.6915
  - $P(X > 10)$
    - $1 - P(X \leq 10) = 1 - P(Z \leq (10 - 4)/2) = 1 - P(Z \leq 3); = 1 - F(3)$
    - From tables, the probability is 1 - .99865 = 0.00135
  - $P(4 \leq X \leq 5)$
    - $P(0 \leq Z \leq 0.5) = F(0.5) - F(0)$
    - From tables, the probability is 0.6915 – 0.5 = 0.1915

Empirical Rule for the Normal Distribution

- **The Empirical Rule:** If X is a random variable and has a normal distribution with mean μ and standard deviation σ, then the **Empirical Rule** says the following:
  - About 68% of the x values lie between -1σ and $+1\sigma$ of the mean μ (within one standard deviation of the mean). The z-scores for $+1\sigma$ and -1σ are +1 and −1, respectively.
  - About 95% of the x values lie between -2σ and $+2\sigma$ of the mean μ (within two standard deviations of the mean). The z-scores for $+2\sigma$ and -2σ are +2 and −2, respectively.
  - About 99.7% of the x values lie between -3σ and $+3\sigma$ of the mean μ (within three standard deviations of the mean). The z-scores for $+3\sigma$ and -3σ are +3 and −3 respectively.



- Notice that almost all the x values lie within three standard deviations of the mean.
- The empirical rule is also known as the 68-95-99.7 rule.



Book Problem 6.80 – Page 387

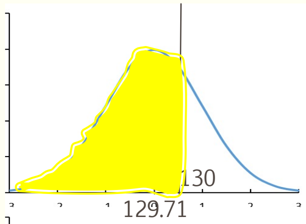
- Terri Vogel, an amateur motorcycle racer, averages 129.71 seconds per 2.5 mile lap (in a seven-lap race) with a standard deviation of 2.28 seconds. The distribution of her race times is normally distributed. We are interested in one of her randomly selected laps.

  a. In words, define the random variable *X*. X is Terri Vogel's race time in seconds per 2.5 miles

  b. X ~ Normal(129.71, 2.28)

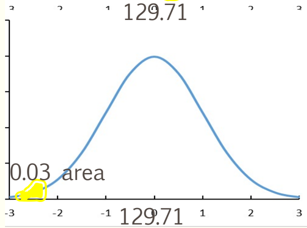
  c. Find the percent of her laps that are completed in less than 130 seconds.

  P(X ≤ 130) = P(Z ≤ (130 - 129.71)/2.28) = P(Z ≤ 0.1272) = F(0.1272) (from tables) = 0.55. So 55% of her laps are under 130 seconds.



  d. The fastest 3% of her laps are under _____.

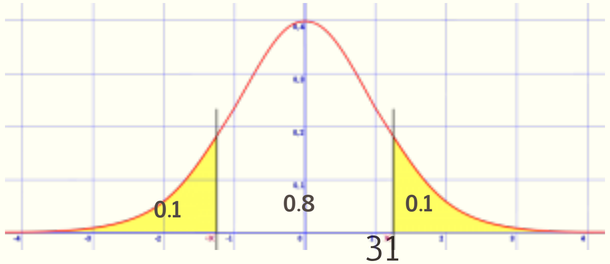
  We need x such that P(X ≤ x) = 0.03. We first find the z-value such that P(Z ≤ z-value) = 0.03.
  z-value = -1.88;
  Convert z-value to x as x = μ + σ*z-value
  = 129.71 + (-1.88*2.28) = 125.4235



  e. The middle 80% of her laps are from _______ seconds to _______ seconds.

  The distribution is symmetrical. So, we need z-value first such that P(Z ≤ z-value) = 0.10.
  The z-value (from tables) -1.28
  This corresponds to X-value of 129.71 + (-1.28*2.28) = 125.4235 = 126.79.
  The other end would therefore be 129.71 + (1.28*2.28) = 132.63.

  So the laps are from 126.79 seconds to 132.63 seconds.



Normal.R

- In R, we can work directly with the original distribution without converting to the standard normal (Z) distribution.
  - pnorm(q, mean, sd, lower.tail = TRUE, log.p = FALSE) where q is lower-tailed probability. It will return the X-value.
  - qnorm(p, mean, sd, lower.tail = TRUE, log.p = FALSE) where p is the X-value; it will return the lower-tail probability P(X <= x)

```
#Normal Distribution
#
# dnorm(x, mean = 0, sd = 1, log = FALSE)
# pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
# qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
#
#Terri Vogel, an amateur motorcycle racer, averages 129.71 seconds per 2.5 mile lap (in a seven-lap race)
#with a standard deviation of 2.28 seconds. The distribution of her race times is normally distributed.
#We are interested in one of her randomly selected laps.
#
#Find the percent of her laps that are completed in less than 130 seconds.
#
print(paste("Percent of her laps that are completed in less than 130 seconds = "
          ,round(pnorm(130, 129.71, 2.28, lower.tail = TRUE, log.p = FALSE), 4)*100))
#
#The fastest 3% of her laps are under
print(paste("The fastest 3% of her laps are under "
          ,round(qnorm(0.03, 129.71, 2.28, lower.tail = TRUE, log.p = FALSE), 4)))
#
#The middle 80% of her laps are from _______ seconds to _______ seconds.
#
f10 = round(qnorm(0.10, 129.71, 2.28, lower.tail = TRUE, log.p = FALSE), 4)
f90 = round(qnorm(0.90, 129.71, 2.28, lower.tail = TRUE, log.p = FALSE), 4)
print(paste("The middle 80% of her laps are from ",f10,"seconds to ", f90, " seconds." ))
```

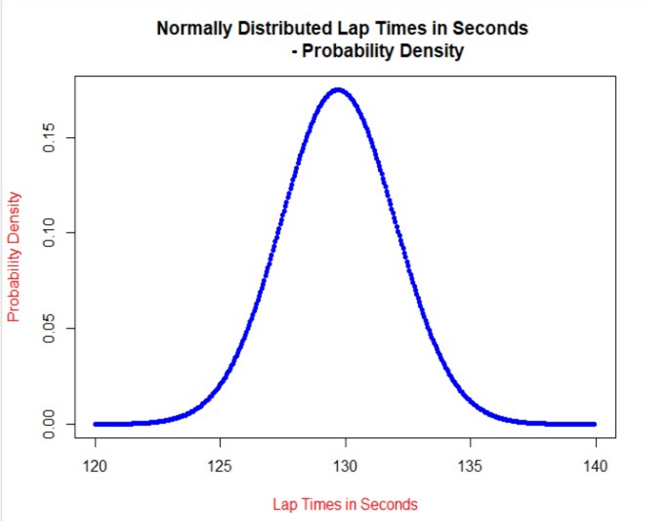
```
[1] "Percent of her laps that are completed in less than 130 seconds =  55.06"
```

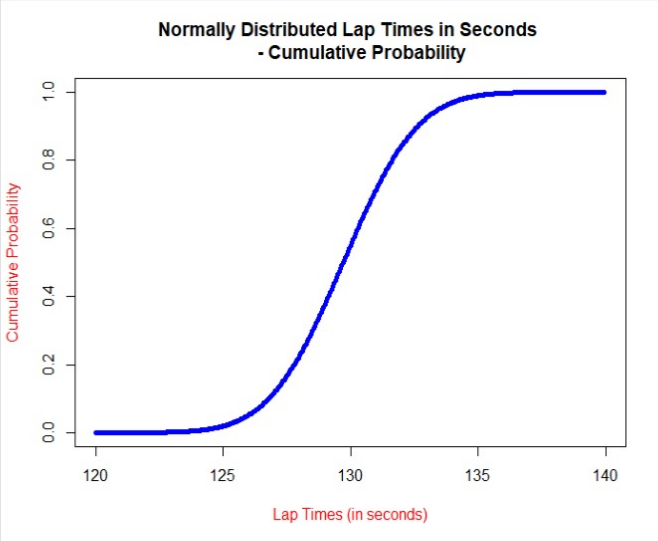
```
[1] "The fastest 3% of her laps are under  125.4218"
```

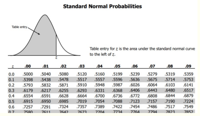
```
[1] "The middle 80% of her laps are from  126.7881 seconds to  132.6319  seconds."
```

Normal.R

```
#table of probabilities for select values of X from 120 to 140.
nsize = 400
result <- vector("numeric", nsize)
cum_result <- vector("numeric", nsize)
x <- vector("numeric", nsize)
xPx <- vector("numeric", nsize)
x2Px <- vector("numeric", nsize)
for (i in 1:nsize) {
  # We are creating 400 values of x at 0.05 intervals, so 120 to 140.
  x[i] <- 120 + ((i-1)/20)
  result[i] <- round(dnorm(x[i], 129.71, 2.28, log = FALSE), 4)
  cum_result[i] <- round(pnorm(x[i], 129.71, 2.28, lower.tail = TRUE, log.p = FALSE), 4)
}
#
# Plot the probabilities
len_result <- length(result)
indx = (len_result - 1)
#Plot the probability mass function
plot(x,result,                                                #-- Plot(x,y)
      xlim=c(120,140),                                        #-- set x-axis limits between 120 and 140
      type = "p",                                             #-- p is for plotting points
      pch=20,                                                 #-- pch=20 is small circles for points
      main = "Normally Distributed Lap Times in Seconds
              - Probability Density",
      xlab = "Lap Times in Seconds",
      ylab = "Probability Density",
      col  = "blue",
      col.lab ="red"
      )
#
# Plot the cumulative probabilities
len_result <- length(result)
indx = len_result - 1
plot(x,cum_result,
      xlim=c(120,140),
      type = "p",
      main = "Normally Distributed Lap Times in Seconds
     - Cumulative Probability",
      xlab = "Lap Times (in seconds)",
      ylab = "Cumulative Probability",
      col  = "blue",
      col.lab ="red",
      pch=20)
```





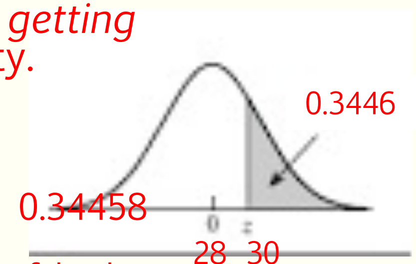
Practice with Table 

Book Problem 6.88 – Page 389

- Facebook provides a variety of statistics on its Web site that detail the growth and popularity of the site. On average, 28 percent of 18 to 34 year olds check their Facebook profiles before getting out of bed in the morning. Suppose this percentage follows a normal distribution with a standard deviation of five percent.

- The random variable X is "*percent of 18 to 34 year olds who check their Facebook profiles before getting out of bed in the morning*"." Note that the units of X is a percentage. Don't confuse with probability.

  a. Find the probability that the percent of 18 to 34-year-olds who check Facebook before getting out of bed in the morning is at least 30 percent.

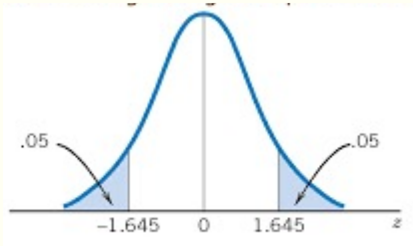
  

  X ~ N(28, 5); We need $P(X \geq 30)$ or $P(Z \geq (30 - 28)/5) = P(Z \geq 0.4) = 1 - P(Z <= 0.4) = 1 - .65542 = 0.34458$

  There is a 0.34458 probability that at least 30% of Facebook users check their profile before getting out of bed.

```
> print(paste("There is ",
+          round(1 - pnorm(30, 28, 5, lower.tail = TRUE, log.p = FALSE),4),
+       "probability that at least 30% of 18 to 34 year old Facebook users check their profile before getting out of bed"))
[1] "There is  0.3446 probability that at least 30% of 18 to 34 year old Facebook users check their profile before getting out of bed"
```

  b. Find the 95th percentile, and express it in a sentence.

  Here we are given the probability (quantile) and asked to find the value of the random variable.

  The 95th percentile means $P(X \leq x) = 0.95$.

  From tables, $P(Z \leq \text{z-value}) = 0.95$ means that z-value = + 1.645;

  So $x = \mu + \sigma * \text{z-value} = 28 + (5)*1.645 = 36.225$

```
> print(paste("There is a 95% probability that at most",
+             round(qnorm(0.95, 28, 5, lower.tail = TRUE, log.p = FALSE),4),
+             "% of 18 to 34 year old Facebook users check their profile before getting out of bed"))
[1] "There is a 95% probability that at most 36.2243 % of 18 to 34 year old Facebook users check their profile before getting out of bed"
```

ng out

The Normal Approximation to the Binomial Distribution

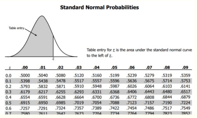
- The Normal distribution was shown (in the 1700's) to arise as an approximation to the binomial when the following conditions are met:

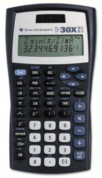
  - The number of trials (n) is large
  - X (the number of successes we are looking for) is moderately large
  - $np \geq 5$ and $nq \geq 5$
  - Then, for the equivalent Normal distribution: the Mean $\mu = np$ and Variance $= \sigma^2 = npq$

- Remember, also we saw earlier that:

  - The Poisson Distribution gives probabilities to the Binomial when n is large and p is small. That is, when we consider a large number of independent occurrences each with a small probability, then we can model the Poisson using the Binomial.

Practice with Tables and Calculator



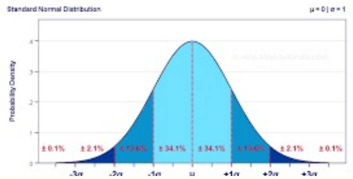


The Normal Approximation to the Binomial Distribution

- **Example**: We flip a coin 100 times (n = 100) and note that it only comes up heads 20% (p = 0.20) of the time. We will find the following using Binomial and Normal and Poisson and compare:
    - Binomial (n = 100, p = 0.20):
    - Poisson (λ = 20 per 100 trials):
    - Normal ($\mu = np = 20$ and $\sigma = \sqrt{npq} = 4$):
  - Probability Number of Heads between 12 and 28 in 100 trials: P(12 < X ≤ 28) = = P(X ≤ 28) - P(X ≤ 12)

```
> print(paste("Binomial Probability of X between 12 and 28 is", round(pbinom(28, 100, 0.2, log.p = FALSE ) - pbinom(12, 100, 0.2, log.p = FALSE ),4)))
[1] "Binomial Probability of X between 12 and 28 is 0.9547"
> print(paste("Poisson Probability of X between 12 and 28 is", round(ppois(28, 20, log.p = FALSE ) - ppois(12, 20, log.p = FALSE ),4)))
[1] "Poisson Probability of X between 12 and 28 is 0.9267"
> print(paste("Normal Probability of X between 12 and 28 is", round(pnorm(28, 20, 4, log.p = FALSE ) - pnorm(12, 20, 4, log.p = FALSE ),4)))
[1] "Normal Probability of X between 12 and 28 is 0.9545"
```

  - There is a 68% chance (centered around the *median*) that the number of heads is betwe




```
> print(paste("Binomial - There is a 68% chance that the number of heads is between", qbinom(0.16, 100, 0.2, log.p = FALSE )," and ",qbinom(0.84, 100, 0.2, log.p = FALSE )))
[1] "Binomial - There is a 68% chance that the number of heads is between 16  and  24"
> print(paste("Poisson -  There is a 68% chance that the number of heads is between", qpois(0.16, 20, log.p = FALSE ), " and ",qpois(0.84, 20, log.p = FALSE )))
[1] "Poisson -  There is a 68% chance that the number of heads is between 16  and  24"
> print(paste("Normal -  There is a 68% chance that the number of heads is between", qnorm(0.16, 20, 4, log.p = FALSE ), " and ",qnorm(0.84, 20, 4, log.p = FALSE )))
[1] "Normal -  There is a 68% chance that the number of heads is between 16.022168467161  and  23.977831532839"
```