

1. Graph the binomial distribution for $n = 10$ and $p = 0.1$ with the command `gbinom(10,0.1)`. Repeat this for $p = 0.2, 0.3, \dots, 0.9$. [Please do not include these graphs with your assignment.]

(a) How does the center of the distribution change as p changes?

Solution: The center moves to the right as p increases. (The mean is $10p$.)

(b) For which value of p is the distribution most strongly skewed right? left? most symmetric?

Solution: The distribution has the strongest skew right when p is small, 0.1 for this problem. The distribution has the strongest skew left when p is large, 0.9 for this problem. The distribution is perfectly symmetric when $p = 0.5$.

(c) For which value of p is the standard deviation the largest?

Solution: The standard deviation is largest when the graph is most spread out around the mean. This happens when $p = 0.5$. (You could use the formula to find $\sigma = \sqrt{np(1-p)}$ if you wanted to verify numerically what your eye tells you.)

2. Graph the binomial distribution for $n = 1$ and $p = 0.5$ with the command `gbinom(1,0.5)`. Repeat this for $n = 2, 4, 8, 16, 32, 64, 128$. [Please do not include these graphs with your assignment.]

(a) How does the center of the distribution change as n changes?

Solution: The center increases as n increases. (The mean is np .)

(b) Is this distribution skewed for any n ?

Solution: With $p = 0.5$, the distribution is perfectly symmetric for all n .

(c) What is the smallest n for which the distribution looks approximately normal? (There is no single correct answer.)

Solution: To my eye, when $n = 16$ I can see the bell shape appear, although it is somewhat apparent for smaller n and even more apparent for larger n . Many different answers could be correct.

(d) What happens to the range of values for which the probabilities are large enough to be visible as n increases?

Solution: The range of visible probabilities (the number of visible lines) increases as n increases.

(e) What happens to the range of values for which the probabilities are large enough to be visible over n as n increases?

Solution: The range of visible probabilities occupies a smaller and smaller portion of the entire possible range from 0 to n as n increases.

3. Graph the binomial distribution for $n = 1$ and $p = 0.1$ with the command `gbinom(1,0.1)`. Repeat this for $n = 2, 4, 8, 16, 32, 64, 128$. [Please do not include these graphs with your assignment.] About how large does n need to be before the distribution looks nearly symmetric and approximately normal? Compare your answer here to the answer in part (c) in the previous problem.

Solution: With $n = 16$, there is still a noticeable moderately strong skew. By the time $n = 64$, the skew has mostly disappeared and the visible probabilities look to be fairly symmetric.

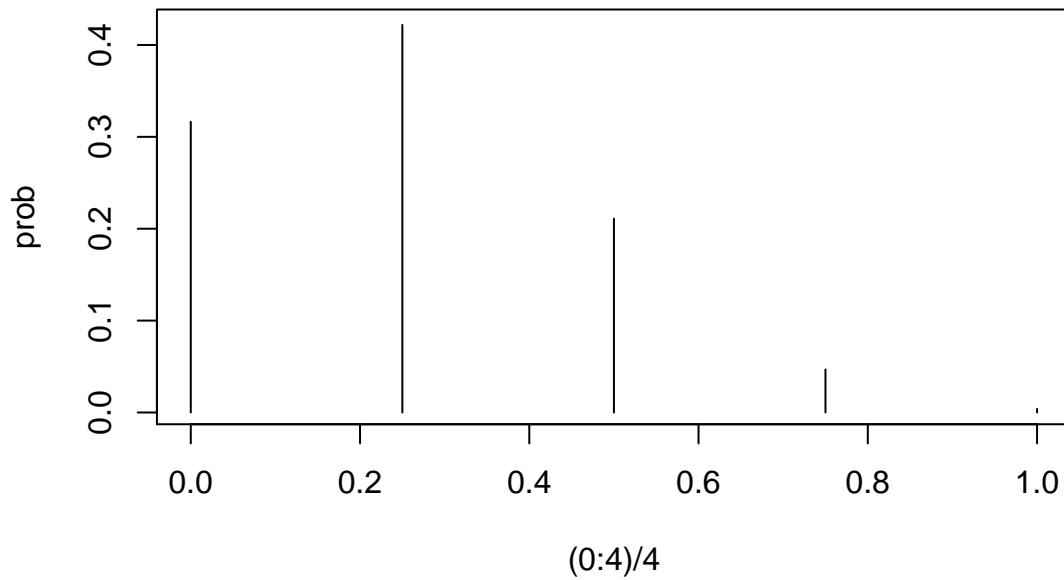
4. Exercise 5.3 (page 157).

Solution: In a forest, 25% of the white pines have blister rust. Four white pines are sampled and \hat{p} is the sample proportion with blister rust. Find and plot the probability distribution of \hat{p} .

```
> prob <- dbinom(0:4, 4, 0.25)
> prob
```

```
[1] 0.31640625 0.42187500 0.21093750 0.04687500 0.00390625
```

```
> plot((0:4)/4, prob, type = "h")
```



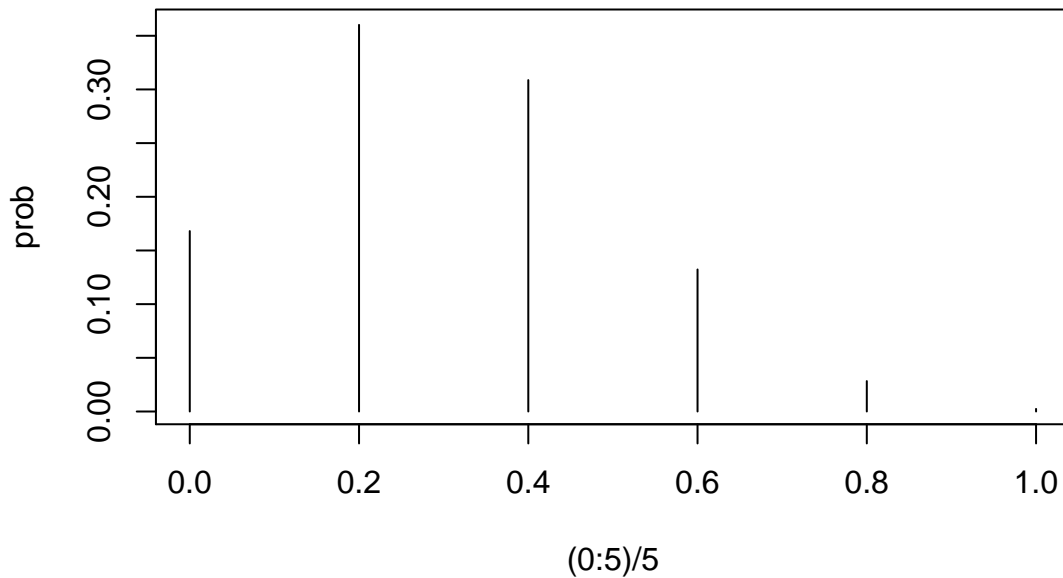
5. Exercise 5.6 (page 157).

Solution: In a population, 30% of the people have superior vision. The sample proportion with superior vision in a random sample of size five is \hat{p} . Find and plot the sampling distribution.

```
> prob <- dbinom(0:5, 5, 0.3)
> prob
```

```
[1] 0.16807 0.36015 0.30870 0.13230 0.02835 0.00243
```

```
> plot((0:5)/5, prob, type = "h")
```



The distribution has the same center (0.3) as Figure 5.3, but it is spread out more and has fewer possible values.

6. Exercise 5.16 (page 167). (The command `gnorm(145,22,prob=T,a=135,b=155)` draws a sketch for part (a) of this problem. The command `gnorm(145,22/sqrt(16),prob=T,a=135,b=155)` draws a sketch for part (b) of this problem.)

Solution: Heights of corn plants are normally distributed with a mean of 145 cm and a standard deviation of 22 cm.

(a) What percentage of the plants are between 135 and 155 cm? By hand,

$$\begin{aligned}
 \Pr\{135 < Y < 155\} &= \Pr\left\{\frac{135 - 145}{22} < \frac{Y - 145}{22} < \frac{155 - 145}{22}\right\} \\
 &= \Pr\{-0.45 < Z < 0.45\} \\
 &= 0.6736 - 0.3264 \\
 &= 0.3472
 \end{aligned}$$

Using R, we get this.

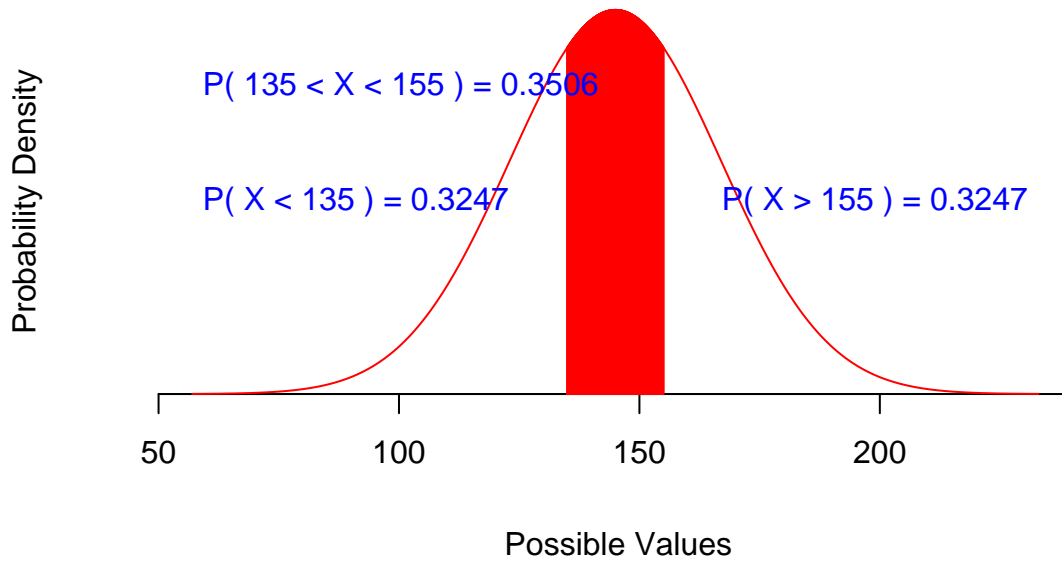
```
> pnorm(155, 145, 22) - pnorm(135, 145, 22)
```

```
[1] 0.3505637
```

Or we can graph it to see the answer.

```
> gnorm(145, 22, prob = T, a = 135, b = 155)
```

Normal Distribution
mu = 145 , sigma = 22



(b) What proportion of sample means with $n = 16$ would be between 135 and 155? By hand,

$$\begin{aligned} \Pr\{135 < \bar{Y} < 155\} &= \Pr\left\{\frac{135 - 145}{22/\sqrt{16}} < \frac{\bar{Y} - 145}{22/\sqrt{16}} < \frac{155 - 145}{22/\sqrt{16}}\right\} \\ &= \Pr\{-0.11 < Z < 0.11\} \\ &= 0.5438 - 0.4562 \\ &= 0.0876 \end{aligned}$$

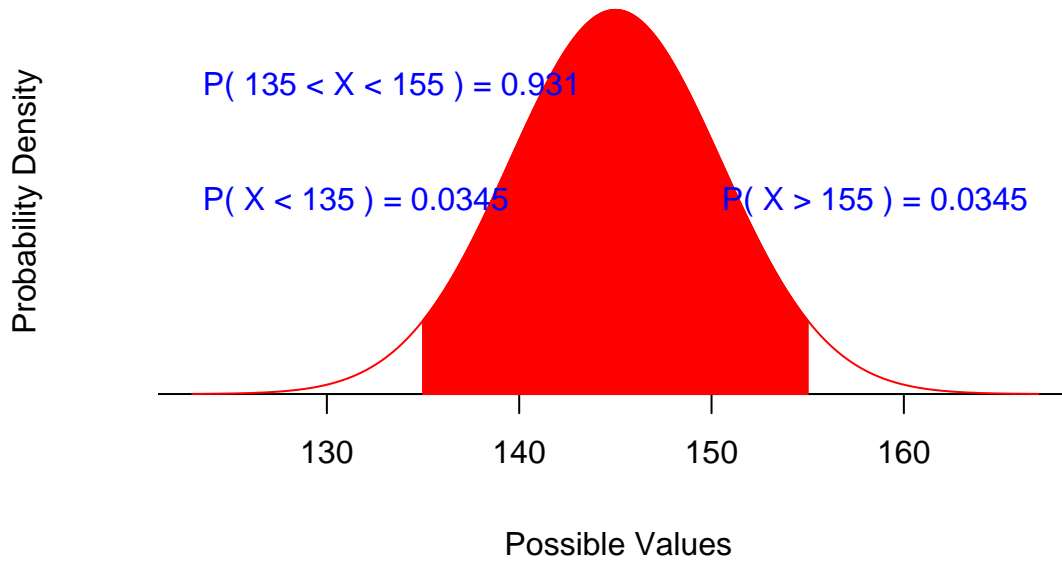
Using R, we get this.

```
> pnorm(155, 145, 22/sqrt(16)) - pnorm(135, 145, 22/sqrt(16))
[1] 0.9309637
```

Or we can graph it to see the answer.

```
> gnorm(145, 22/sqrt(16), prob = T, a = 135, b = 155)
```

Normal Distribution
mu = 145 , sigma = 5.5



- (c) What is $\Pr\{135 < \bar{Y} < 155\}$ if $n = 16$? The answer is the same as in part (b). This is just a different way to ask the same question.
- (d) What is $\Pr\{135 < \bar{Y} < 155\}$ if $n = 36$? By hand,

$$\begin{aligned} \Pr\{135 < \bar{Y} < 155\} &= \Pr\left\{\frac{135 - 145}{22/\sqrt{36}} < \frac{\bar{Y} - 145}{22/\sqrt{36}} < \frac{155 - 145}{22/\sqrt{36}}\right\} \\ &= \Pr\{-0.08 < Z < 0.08\} \\ &= 0.5319 - 0.4681 \\ &= 0.0638 \end{aligned}$$

Using R, we get this.

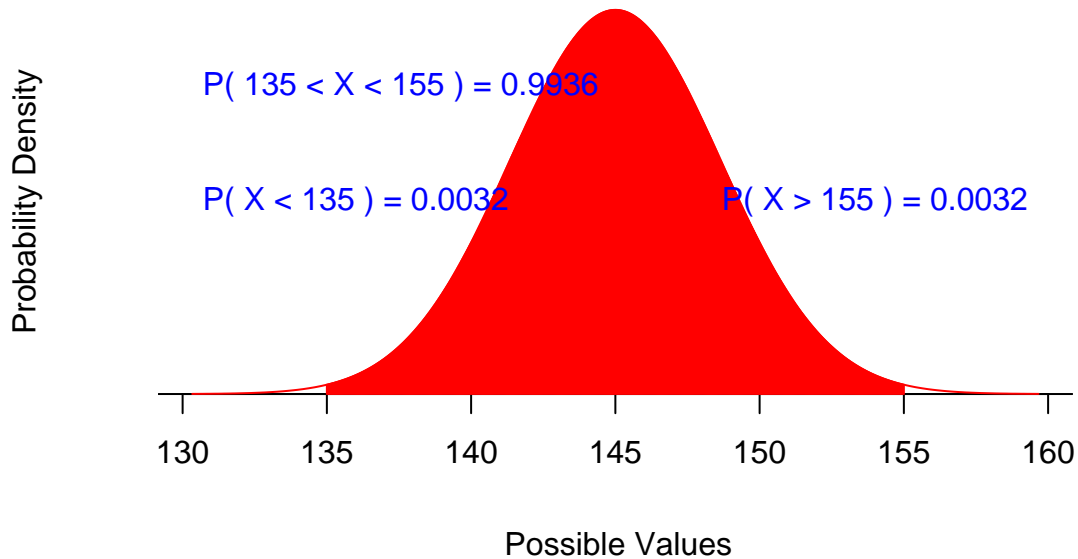
```
> pnorm(155, 145, 22/sqrt(36)) - pnorm(135, 145, 22/sqrt(36))
[1] 0.993614
```

Or we can graph it to see the answer.

```
> gnorm(145, 22/sqrt(36), prob = T, a = 135, b = 155)
```

Normal Distribution

mu = 145 , sigma = 3.66666666666667



7. Exercise 5.18 (page 167).

Solution: Lengths of individual fish are normally distributed with mean 54 mm and standard deviation 4.5 mm. A previous problem shows that 65.68% of the fish are between 51 and 60 mm in length. Four fish are randomly selected.

(a) What is the probability that all four fish have lengths between 51 and 60 mm?

The number of sampled fish with lengths between 51 and 60 mm is binomially distributed with $n = 4$ and $p = 0.6568$. The probability is $(0.6568)^4 = 0.186094088061338$. Or, using R,

```
> dbinom(4, 4, 0.6568)
```

```
[1] 0.1860941
```

(b) What is the probability that the mean fish length is between 51 and 60 mm?

The sample mean \bar{Y} is normally distributed because the population is. The standard deviation of the sampling distribution is $4.5/\sqrt{4} = 2.25$. The probability is:

$$\begin{aligned} \Pr\{51 < \bar{Y} < 60\} &= \Pr\left\{\frac{51 - 54}{4.5/\sqrt{4}} < \frac{\bar{Y} - 54}{4.5/\sqrt{4}} < \frac{60 - 54}{4.5/\sqrt{4}}\right\} \\ &= \Pr\{-0.33 < Z < 0.67\} \\ &= 0.7486 - 0.3707 \\ &= 0.3779 \end{aligned}$$

Using R,

```
> pnorm(60, 54, 4.5/sqrt(4)) - pnorm(51, 54, 4.5/sqrt(4))
```

```
[1] 0.9049584
```

8. The function `gmix` will graph the sampling distribution of the mean drawn from of a bimodal distribution where you can specify the sample size, the means, standard deviations, and the weight of the first mode. It also draws a normal distribution with the same mean and SE for comparison. For example, `gmix(1, 115, 25, 450, 50, 0.9)` will draw a graph similar to the one on page 170. A sampling distribution for $n = 1$ is just the population. The first mode is centered at

115 with an SD about 25. The second mode is centered at 450 with an SD about 50. About 90% of the probability is in the first mode. You can reproduce the later graphs by changing the first argument from 1 to whichever sample size you desire.

Consider the graphs of these two populations.

1. `gmix(1,115,25,450,50,0.9)`
2. `gmix(1,115,25,450,50,0.5)`

- (a) For each population, what is the smallest n in 5, 10, 15, 20, ... for which the sampling distribution appears to be unimodal? (In a unimodal distribution, there is a point where the function is increasing to the left of the point and decreasing to the right of the point.)

Solution: The first distribution is unimodal when $n = 35$. The second distribution is unimodal when $n = 20$.

- (b) For each population, what is the smallest n in 1, 2, 4, 8, 16, ... (doubling the sample size) for which you think that a normal approximation would be pretty good? (For example, when is the 90th percentile pretty close to that for the normal curve? When is the area within one SE close to that of the normal curve?)

Solution: The second distribution is approximately normal for n as small as 16. For the first, n would need to be at least 64, although there is still a noticeable skew, even when n is 128.

- (c) One distribution is approximately normal for a smaller n than the other. Make a guess as to why. What important characteristic distinguishes the two populations?

Solution: The first distribution is more strongly skewed. This means that n needs to be much larger before the distribution becomes approximately normal.