Pizza

Pizza. I just love it. Nothing is better than a slice. Nothing that is except perhaps a few question about the binomial distribution that deal with made up stories about pizza.

1 The binomial random variable: Background

The binomial random variable counts the number of succeses in n independent trials, where each trial has probability p of success.

There are two main numbers which categorize the binomial: n and p. With these we have the

Distribution That is

$$P(X = k) = \binom{n}{k} p^k (1-p)^{(n-k)}$$

Mean or expected value is $\mu = np$

Standard deviation $\sigma = \sqrt{np(1-p)}$.

We can ask questions about the binomial that involve computing a given probability such as $P(X = k)$ or a range of probabilities such as $P(X \leq k)$. Additionally, we might want to find random samples from the distribution.

For concreteness, let X count the number of heads in 100 tosses of a coin. Then X is binomial with $n = 100$ and $p = 1/2$.

We have:

- $P(X = 50)$

  ```
  > dbinom(50, 100, 1/2)
  ```

  ```
  [1] 0.07959
  ```

- $P(X \leq 60)$:

  ```
  > pbinom(60, 100, 1/2)
  ```

  ```
  [1] 0.9824
  ```

  (This includes the event $X = 60$.)

- A random sample of size 10

  ```
  > rbinom(10, 100, 1/2)
  ```

```
[1] 47 53 53 59 49 52 54 50 47 51
```

For each notice the function's name is found by combining either `d`, `p` or `r` combined with `binom`. As well, the values of n and p are needed, n first then p.

Question 1: Take a random sample of size 10 as was done above. Did you get the exact same 10 numbers? Did you expect to?

Question 2: Let X count the number of sixes in 100 rolls of a die. Find the following:

$$P(X \leq 20), \quad P(X = 16), \quad P(X > 12) = 1 - P(X \leq 12).$$

(For the last one, we use the *complement rule*: If A and B have no outcomes in common, but combine to give the entire set of outcomes then $P(A) + P(B) = 1$. In this case $A = \{X > 12\}$ and $B = \{X \leq 12\}$.)

Question 3: Let X be the number of heads in 25 tosses of a coin. Computing a probability like

$$P(X < 12)$$

is done with a command like

```
> pbinom(11, 25, 1/2)

[1] 0.345
```

Explain why? In particular why is that an 11 and not a 12.

Question 4: Historically, a pizza delivery person gets a tip at 80% of the houses she delivers to. In the next 25 deliveries, what is the probability of 15 or fewer tips?

Question 5: At the pizza counter, it is known that 30% of the customers that come in will order pepperoni. However, this does not mean that in the next 100 customers exactly 30 will order pepperoni. Let X be the number who do. Find

$$P(X = 30).$$

Question 6: The pizza restaurant boss knows that an order will be messed up with probability 0.02. If a night has 100 orders, let X be the number that get messed up. Find the following:

$$P(X = 2), \quad P(X = 0), \quad P(X \geq 3)$$

(For the latter, you might consider $P(X \leq 2)$ first. Why?)

Question 7: Suppose the number of pizzas sold each night, X, is binomially distributed with $n = 250$ and $p = .6$. Find the expected number, μ, of pizzas sold. What is the standard deviation, σ?

Now find (using your values of μ and σ)

$$P(X \leq \mu + \sigma)$$

What about (notice the $<$ and not $\leq$.)

$$P(X < \mu - \sigma)$$

The difference between the two of these answers is

$$P(\mu - \sigma \leq X \leq \mu + \sigma)$$

That is, X has a z-score between -1 and 1. What is this probability?

2 Plots

What does the distribution of the binomial look like? We can make a plot for a given n and p in the following manner.

Suppose $p = .3$ is the probability that a customer uses parmesan cheese on their pizza. Let X be the number of customers in the next 50 who use parmesan cheese. Then X is binomial with $n = 50$ and $p = 0.3$. A plot (FIgure 1) of the distribution can be produced with these commands:

```
> k = 0:50
> plot(k, dbinom(k, 50, 0.3), type = "h")
```

Question 8: Figure 1 shows that most of the action happens between $k = 5$ and $k = 30$. Replot, only this time use

```
> k = 5:30
```

Describe the shape of your plot. What value is most likely? What is the mean value?

Question 9: Suppose the number of empty seats at dinner time, X, has a binomial distribution with $n = 35$ and $p = .05$. Make a plot of the distribution of X.

What is the most likely value? What is the mean value?

3 Normal distribution

The Normal distribution is described by being bell shaped with mean μ and standard deviation σ. It too has built in R functions to answer questions about it. These are called `dnorm()`, `pnorm()` and `rnorm()`. Notice these are similar to those for the binomial. Of course to use these you specify the mean and standard deviation.

For example, suppose the amount of tomato sauce (in ounces) used on a large pizza pie, X, is normally distributed with mean 8 and standard deviation 1. Then we can compute

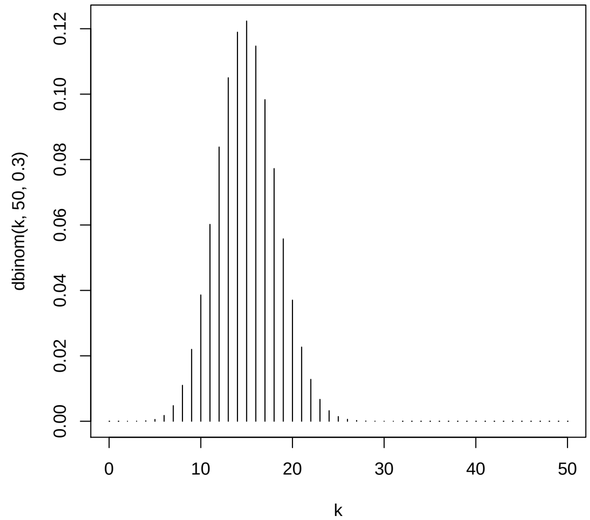

Figure 1: Plot of X ($n = 50, p = 0.3$).

- $P(X \leq 7)$

```
> pnorm(7, 8, 1)
```

```
[1] 0.1587
```

- $P(X > 9)$

```
> 1 - pnorm(9, 8, 1)
```

```
[1] 0.1587
```

- $P(6 \leq X \leq 7)$

```
> pnorm(7, 8, 1) - pnorm(6, 8, 1)
```

```
[1] 0.1359
```

- Finally, a random sample of size 10 can be produced with:

```
> rnorm(10, 8, 1)
```

```
[1] 7.801 8.660 7.833 6.935 8.548 5.630 8.299 8.825 9.271 8.014
```

Question 10: Suppose the amount of flour used in a pizza pie (in cups), X, is normally distributed with a mean of 3 and a standard deviation of 0.5. Find the following

$$P(X < 2), \quad P(X < 4), \quad P(2.5 < X < 3.5)$$

Question 11: Suppose the nightly amount of money a pizza restaurant takes in, Y, is normally distributed with $\mu = 2000$ and $\sigma = 400$. A new night manager is hired and on his first night the amount taken in is only $1,400. Find

$$P(Y \leq 1400)$$

Do you think the night manager is a cheat?

4 The normal and binomial on the same pizza?

The binomial and normal are related. Maybe you could guess from Figure 1 where it looked like the graph of the binomial is "bell-shaped." But which bell?

What happens when we plot a binomial random variable with n and p and match that with a normal distribution with $\mu = np$ and $\sigma = \sqrt{np(1-p)}$?

Question 12: Why do you think the normal has this choice of mean and standard deviation?

To illustrate, let's say X is the number of pepperoni slices on a pepperoni pizza. We assume X is binomial with $n = 40$ and $p = 0.6$. We can plot the binomial with

```
> p = 0.6
> n = 40
> k = 0:40
> plot(k, dbinom(k, n, p), type = "h")
```

To that we can *add* a normal curve with

```
> mu = n * p
> sigma = sqrt(n * p * (1 - p))
> curve(dnorm(x, mu, sigma), add = TRUE)
```

The result is the left plot in Figure 2. Notice how close the curve matches the tips of the lines that give the distribution of the binomial.

Question 13: Replot using this value for

```
> k = 15:35
```

Does the normal curve still track the binomial probabilities?

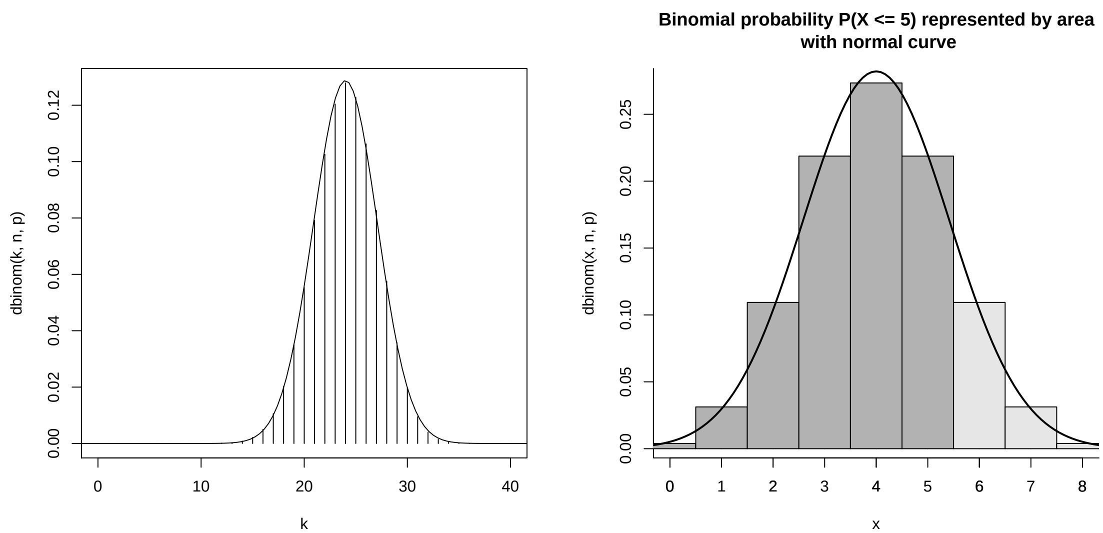

Figure 2: Binomial distribution with normal curve added. Right figure shows what is happening for a smaller n.

Question 14: Let X be binomial with $n = 40$ and $p = 0.6$, and Y be normal with mean $\mu = np$ and $\sigma = \sqrt{np(1-p)}$. Compare the following values

$$P(X \leq 30) \text{ to } P(Y \leq 30), \quad \text{and} \quad P(X \leq 22) \text{ to } P(Y \leq 22)$$

Are your answers similar. On the computer they are both just as easy to find, but by hand the ones with the normal distribution are much easier.

4.1 The continuity correction

The book uses a "continuity correction" when discussing the normal approximation. The right plot in Figure 2 shows where it comes from. Look at the $P(X \leq 5)$. This is found by adding the area in the boxes for $X = 0, 1, 2, 3, 4$ and 5. The area of these boxes is approximated by the area under the normal curve to the left of $5\ 1/2$ – where the box for $X = 5$ ends. So in formulas, a more accurate approximation is:

$$P(X \leq b) = P(Y \leq b + \frac{1}{2}), \quad \text{and } P(X \geq a) = P(Y \leq a - \frac{1}{2})$$

Question 15: Repeat the last exercise, only use this "continuity correction". Is the approximation more accurate?