Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They were written to help the students of STA 371G to review the topics discussed in class.*

Optional text: *Text in this color is optional to read.*

3.1   Binomial Distribution

If the normal distribution is the most important distribution in statistics, then the binomial distribution might be the one that is just behind it. We had already discussed the Bernoulli distribution in our first class when the concept of random variable was introduced. Recall that a Bernoulli random variable

$$X \sim \text{Bernoulli}(p)$$

has the value of $X = 1$ with probability p and the value of $X = 0$ with probability $1 - p$. We may intuitively consider X as the random outcome of flipping an unfair coin, which lands with "Head" with probability p, and lands with "Tail" with probability $1 - p$. The mean or expectation of a Bernoulli random variable $X \sim \text{Bernoulli}(p)$ can be calculated as

$$\mu = \mathbb{E}[X] = 1 \times p + 0 \times (1 - p) = p,$$

and the variance of $X \sim \text{Bernoulli}(p)$ can be calculated as

$$\sigma^2 = \text{Var}[X] = (1 - \mathbb{E}[X])^2 \times p + (0 - \mathbb{E}[X])^2 \times (1 - p) = p(1-p)^2 + (1-p)(-p)^2 = p(1-p).$$

Note that if $p = 1$ or $p = 0$, which means the coin would land with "Head" almost surely or with "Tail" almost surely, then the variance is $\sigma^2 = 0$, which is intuitive as the outcome has no variability if $p = 1$ or $p = 0$. The variance has its largest value when $p = 0.50$, which is also intuitive since when $p = 0.5$, the outcome has the highest variability as we are most uncertain whether we are going to get a "Head" or a "Tail" by flipping this fair coin. Thus the variance σ^2 (or its square root: the standard deviation σ) can be considered as a measure of variability.

The binomial distribution is a generalization of the Bernoulli distribution. Instead of flipping the unfair coin once, we are going to flip the unfair coin n times. The total number of "Heads" in these n coin flips would be a binomial random variable, denoted as

$$X \sim \text{Binomial}(n, p).$$

It is clear that X would have $n + 1$ possible random outcomes: from the smallest $X = 0$, which means zero "Heads" in n coin flips, to the largest $X = n$, which means n "Heads" in n coin flips. Since the coin flips are independent to each other, we can easily calculate the probability for $X = 0$ as

$$P(X = 0) = (1 - p)^n$$

and the probability for $X = 1$ as

$$P(X = n) = p^n.$$

For other values of X, one may show that the probability mass function of X can be expressed as

$$P(X = k) = \binom{n}{k} p^k (1-p)^n = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}, \qquad k = 0, 1, \cdots, n. \tag{3.1}$$

There is no need to remember this probability mass function in this class, I write it down just for the sake of completeness. You may follow me to derive the mean and variance of $X \sim \text{Binomial}(n, p)$ to enhance your understanding of the definitions of these two terms. The mean of $X \sim \text{Binomial}(n, p)$ can be expressed as

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{n} k P(X = k) = \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!((n-1)-k)!} p^k (1-p)^{(n-1)-k} \\
&= np. \tag{3.2}
\end{aligned}$$

Note that $P(Y = k) = \frac{(n-1)!}{k!((n-1)-k)!} p^k (1-p)^{(n-1)-k}$ is the probability mass function for $Y \sim \text{Binomial}(n-1, p)$, thus we have $\sum_{k=0}^{n-1} \frac{(n-1)!}{k!((n-1)-k)!} p^k (1-p)^{(n-1)-k} = \sum_{k=0}^{n-1} P(Y = k) = 1$. The variance of $X \sim \text{Binomial}(n, p)$ can be expressed as

$$\begin{aligned}
\text{Var}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\
&= \left(\sum_{k=0}^{n} k^2 P(X = k) \right) - n^2 p^2 \\
&= \left(\sum_{k=0}^{n} k(k-1) P(X = k) \right) + \left(\sum_{k=0}^{n} k P(X = k) \right) - n^2 p^2 \\
&= n(n-1)p^2 + np - n^2 p^2 \\
&= np(1-p). \tag{3.3}
\end{aligned}$$

We have just used the original definition of mean to find out the center of all possible random outcomes, and the definition of variance to find out how spread out the random outcomes are from their center. Now we are going to show a much simpler approach by using the summation of independent random variables. Recall that we have discussed in class that if the random variable X is a summation of n random variables, expressed as

$$X = \sum_{i=1}^{n} X_i = X_1 + X_2 + \cdots + X_n,$$

then its mean or expectation can be expressed as

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n], \tag{3.4}$$

and if X_i and X_j are **independent** to each other for $i \neq j$, then we further have

$$\text{Var}[X] = \sum_{i=1}^{n} \text{Var}[X_i] = \text{Var}[X_1] + \cdots + \text{Var}[X_n]. \tag{3.5}$$

As mentioned above, the Binomial distribution $X \sim \text{Binomial}(n, p)$ describes the number of "Heads" in n independent unfair coin flips, each of which lands with "Head" with probability p. Therefore, we can represent $X \sim \text{Binomial}(n, p)$ as the summation of n independent Bernoulli random variables $X_i \sim \text{Bernoulli}(p)$, expressed as

$$X = \sum_{i=1}^{n} X_i, \ \ X_i \sim \text{Bernoulli}(p)$$

Using (3.4), we have

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} p = np.$$

Using (3.5), we have

$$\text{Var}[X] = \sum_{i=1}^{n} \text{Var}[X_i] = \sum_{i=1}^{n} p(1-p) = np(1-p).$$

Using either approach, we find that the binomial random variable $X \sim \text{Binomial}(n, p)$ has the mean and variance as

$$\mathbb{E}[X] = np,$$

$$\text{Var}[X] = np(1-p).$$

3.1.1 Sampling From a Large Population

Flipping an unfair coin n times and records the number of "Heads" is one way to understand the binomial distribution. The other example is to survey n people randomly selected from a large population with a "Yes/No" question, where the population size is considerably larger than n; suppose that the proportion of people in the population that would answer the question with "Yes" is p and the proportion of people in the population that would answer the question with "No" is $1 - p$, then the number of "Yes" from a random sample of n people can be considered as a binomial random variable

$$X \sim \text{Binomial}(n, p).$$

For example, we have 4,438 BBA students in McCombs Business School in Fall 2013, about 87% of whom are Texas residents[1]. If we take a random sample of $n = 50$ McCombs BBA students, then the number of Texas residents from this sample is a discrete random variable X that takes its value from 0, 1, $\cdots$, 50. As $n = 50$ is considerably smaller than the population size of 4,438, we can consider the number of Texas resident X in a random sample of size $n = 50$ to be a binomial random variable as

$$X \sim \text{Binomial}(n = 50, p = 0.87).$$

In the 2012 United States presidential election, according to the the official Federal Election Commission report, out of a total of 7,993,851 votes in Texas, President Barack Obama received 3,308,124 votes. Thus if you randomly ask a Texas resident who casted his/her ballot that whether he/she voted for President Obama in 2012, then the probability for him/her to say "Yes" would be

$$p = \frac{3,308,124}{7,993,851} = 0.414 \text{ (or } 41.4\%).$$

---

[1]http://www.mccombs.utexas.edu/BBA/Prospective/Statistics

If we randomly survey $n = 1000$ Texas residents who casted their ballots in that election, then the number of "Yes" from these voters would be a random variable that is distributed as

$$X \sim \text{Binomial}(n = 1000, p = 0.414).$$

To describe the binomial distribution $X \sim \text{Binomial}(n, p)$, we can list all its possible random outcome $X = 0, 1, \cdots, n$ and their probabilities, which can be calculated with (3.1). Alternatively, if $n \times p > 5$ and $np(1 - p) > 5$ (which is usually satisfied when our sample size is not too small and neither p nor $1 - p$ are too close to zero), we can generally approximate $X \sim \text{Binomial}(n, p)$ with a normal distribution with mean np and variance $np(1 - p)$ (standard deviation $\sqrt{np(1 - p)}$). By doing so, we can use what we have learned about normal distributions in our previous classes to describe the binomial distribution.

3.2 The Normal Approximation to the Binomial Distribution

One may show that when n is not too small and p is neither too close to 1 nor 0, the probability mass function of the binomial distribution can be approximated with the probability density function of a normal distribution that has the same mean and variance as
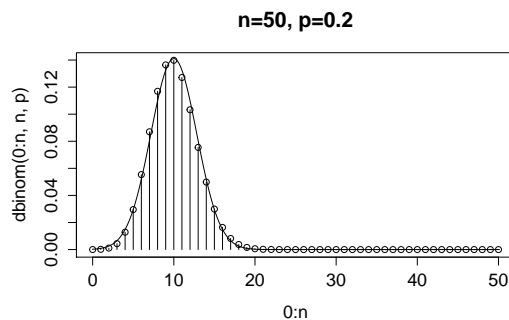
$$Y \sim \mathcal{N}(np, np(1 - p)).$$

Below we plot both the probability mass function of

$$X \sim \text{Binomial}(n = 50, p = 0.2)$$

and the probability density function of

$$Y \sim \mathcal{N}(np = 50 \times 0.2, np(1 - p) = 50 \times 0.2 \times (1 - 0.2)) = \mathcal{N}(10, 8).$$

The above figure (produced with R) shows that the probability density function of $Y \sim \mathcal{N}(10, 8)$ well matches the probability mass function of $X \sim \text{Binomial}(n = 50, p = 0.2)$. For example, the probability for $P(X = 10)$[2] is found to be

$$P(X = 10) = 0.1398,$$

while the area under the normal probability density curve[3] between

$$(10 - 0.5, 10 + 0.5) = (9.5, 10.5)$$

---

[2]In R: dbinom(10,50,0.2); In Excel: =BINOMDIST(10,50,0.2,FALSE).

[3]In R: pnorm(10.5,10,sqrt(8))-pnorm(9.5,10,sqrt(8)); In Excel: =NORMDIST(10.5,10,SQRT(8),TRUE)-NORMDIST(9.5,10,SQRT(8),TRUE).

is found to be
$$P(9.5 < Y < 10.5) = P(Y < 10.5) - P(Y < 9.5) = 0.1403.$$

With this approximation, we are now ready to answer the question: if we randomly survey 1000 Texas residents voted in 2012, how many of them had voted for President Obama? As analyzed earlier, we have

$$X \sim \text{Binomial}(n = 1000, p = 0.414).$$

As n is large and p is close to neither 0 nor 1, we can approximate it with a normal distribution as

$$X \sim \mathcal{N}(np = 1000 * 0.414, np(1 - p) = 1000 * 0.414 * 0.586) = \mathcal{N}(414, 15.6^2).$$

Therefore, recall that there is about 95% probability that a normal random variable deviates from its mean less than 2 standard deviation, thus we are 95% confident that

$$414 - 2 \times 15.6 < X < 414 + 2 \times 15.6.$$

You may notice that $X = 500$ is not within this 95% percent confidence interval, which indicates that $X \geq 500$ happens with a small probability. Let's standardize this value of $X = 500$ under the normal distribution $X \sim \mathcal{N}(414, 15.6^2)$ as

$$\frac{500 - 414}{15.6} = 5.51.$$

$X = 500$ is more than 5 standard deviation away towards the right of the mean! Now we can have a better idea about how extreme that $X = 500$ or larger is when $n = 1000$ and the population proportion is 0.414. This kind of concept about confidence interval and the measurement of how unlikely something is going to happen will be discussed more when we talk about linear regression.