Weekly Review 4

Last week we introduced a particular case of the Binomial Distribution. This week we gave a more general scenario where the binomial distribution is applicable. Mathematically speaking, suppose we perform a sequence of trials and they satisfy the following three criteria.

1. Each trial results in either a success or a failure (where success is just a generic term and is not necessarily something favourable in reality).
2. The probability of a success is \( p \) for each trial.
3. The trials are independent.

Then, let \( X \) be the number of successes out of \( n \) such trials, we have

\[
\Pr(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \; x = 0, \ldots, n,
\]

where

\[
\binom{n}{x} = nC_x = \frac{n!}{(n - x)!x!}
\]

(reads as \( n \) choose \( x \), or simply \( n \ C \ x \)) is called the binomial coefficient, which is the number of possible combinations of choosing \( x \) numbers out of \( n \) numbers, where the order of the \( x \) chosen numbers does not matter.

The upper case \( X \), called a random variable, is used to denote the result of an experiment (in this situation it is the number of successes and in other situations it may be the age, the IQ, the height, the blood pressure, etc.), whilst the lower case \( x \) is just a number and can be replaced by \( r, y, k, z, 5, 17 \) and so on. We usually write \( X \sim B(n, p) \), which reads “\( X \) follows a binomial distribution with \( n \) trials and the probability of a success is \( p \)”.

What I want you to know is when the binomial distribution is applicable. Please read Examples 4.3.1 – 4.3.4 in order to have a better understanding of the binomial distribution. You may not find it very relevant to clinical applications at this moment but you will see its relevance and importance in medical statistics later in this subject.

Then I very briefly introduced the Poisson distribution. If \( X \) follows a Poisson distribution with mean \( \lambda \) (the Greek letter that reads “lambda”), then

\[
\Pr(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \; x = 0, 1, 2, \ldots
\]
One commonly asked question is: when to use the Poisson distribution. The Poisson distribution is employed as a model when counts are made of events or entities that are distributed at random in space or time (see pages 104–107 of your textbook). Why is it relevant to medical students? Suppose you are planning a large scale study which requires 100 volunteers, and so in your proposal and calculation of budget, you have to estimate how long it takes for you to recruit these 100 volunteers. Can you do it in one week? Can you do it in one month? If from past experience we know that on the average we can recruit 5 volunteers per day, then on the average you need 20 days but the exact number of volunteers you can recruit in 20 days is random and the chance that it is 100 or more is

\[
\Pr(X = 100) + \Pr(X = 101) + \Pr(X = 102) + \cdots = \frac{\lambda^{100}e^{-\lambda}}{100!} + \frac{\lambda^{101}e^{-\lambda}}{101!} + \frac{\lambda^{102}e^{-\lambda}}{102!} + \cdots = 0.5133
\]

(which was obtained by using a computer software), where \( \lambda = 100 \) is the mean number of volunteers recruited per 20 days. Thus, a 20-day recruitment period is too optimistic because you have almost a half chance that you will not get enough volunteers. To play safe, you should plan a recruitment period so that the chance of getting 100 or more volunteers is at least, say, 0.95 and further calculation (by computer) based on the Poisson distribution suggested that 24 days would be sufficient for this purpose. (I was trying to explain why the Poisson distribution is relevant to you; I did not mean that you have to know the calculation I demonstrated in this paragraph.)

Then, I introduced the most important distribution in this subject, namely, the normal distribution, which is also known as the Gaussian distribution. One major difference between the normal distribution and the binomial or Poisson distribution is that it is a continuous distribution while the binomial and the Poisson are discrete.

Suppose \( X \sim N(\mu, \sigma^2) \). It is rather easy to calculate \( \Pr(X \leq x) \) by the following way:

1. Calculate \( z = (x - \mu)/\sigma \) (correct to two decimal places).
2. Turn to pages A-38 – A-39 (Table D) and check the entry corresponding to \( z \).

For example, if \( z = -1.37 \), then check the table on the left-hand side. First, find the row starting (and ending) with \(-1.30\), and then find the column with heading \(-0.07\), we get 0.0853. Thus, we write:

\[
\Pr(Z \leq -1.37) = 0.0853.
\]

Please do not write \(-1.37 = 0.0853\), which is ridiculous! Another example, if \( z = 2.74 \), check the table on the right-hand side. Find the row starting with 2.70 and then find the column with heading 0.04, we get 0.9969, and so

\[
\Pr(Z \leq 2.74) = 0.9969.
\]

However, please pay particular attention to the column headings on the left-hand side of the table. The first column corresponds to \(-0.09\) whilst the last column \(-0.00\). Thus, for
$z = -1.40$, you should find the row starting with $-1.40$ and then move to the last column instead of the first column and get 0.0808.

You also have to know two simple tricks:

(a) $\Pr(Z \geq 2.74) = \Pr(Z > 2.74) = 1 - \Pr(Z \leq 2.74) = 0.0031 = \Pr(Z \leq -2.74)$.

(b) $\Pr(-1.37 \leq Z \leq 2.74) = \Pr(Z \leq 2.74) - \Pr(Z \leq -1.37) = 0.9969 - 0.0853 = 0.9116$.

Please read Sections 4.6 and 4.7 carefully and repeatedly. It is very important that you have a good understanding of the normal distribution. This week I gave you Assignment 2, which provides you a lot of chances to practice the calculation of probabilities from normal distributions. The last two questions are not so straightforward.

This is the end of Chapter 4. Next week we will discuss Chapters 5 and 6 in such a way that I will move forth and back between Chapter 5 and Chapter 6 because it is more efficient and natural to introduce estimation of a parameter immediately after the sampling distribution of a statistic.

Cheers,
Sung Nok Chiu