Weekly Review 4

In the three hours of this week, we introduced the binomial, the hypergeometric, the geometric, the Pascal (negative binomial), and the Poisson distribution, concluded Chapter 5 by giving a meaning to subjective probability statements, and discussed the continuous uniform distribution and the normal (Gaussian) distribution. In the example class we worked on some problems related to the application of the general addition rule and the notion of conditional probability.

Last week we discussed the binomial distribution by considering the number of successes in a sequence of Bernoulli trials. Now, consider drawing balls from an urn containing a fixed number a of blue balls and another fixed number b of black balls. We can easily see that the binomial distribution applies when we *sample with replacement* (i.e. putting the sampled ball back to the urn), because drawing balls from the urn containing balls of two different colours with replacement forms a series of Bernoulli trials.

However, the binomial distribution does not apply when we *sample without replacement* (i.e. removing permanently the chosen ball from the urn). For such a sampling, it would be very difficult if we formulated the problem by considering a sequence of trials in which the outcomes of earlier trials would affect the probability of later trials. Nevertheless, in the context of drawing balls (and analogous scenarios), the sampling without replacement is in fact easy; it is the same as the Mark Six (our local lottery) example I used before. If we sample n balls from the urn without replacement, and if we denote by X the number of *successes* (say, a blue ball is a success) in the sampled n balls, then X follows a *hypergeometric distribution*

$$\Pr(X = x) = \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}, \qquad x = 0, \ldots, n, \ x \le a, \ \text{and } n - x \le b.$$

Let us go back to the case of sampling with replacement, i.e. we perform a sequence of Bernoulli trials. This time, however, instead of counting the number of successes in n (a fixed number known beforehand) trials, we just repeat the trials until some pre-specified condition is fulfilled. Denote by Y the number of trials until we get the first success and by Z the number of trials until we get the r^{th} success. Again using elementary arguments, we have derived in class that

$$\begin{aligned} \Pr(Y = y) &= (1-p)^{y-1}p, && y = 1, 2, \ldots, \\ \Pr(Z = z) &= \binom{z-1}{r-1} p^r (1-p)^{z-r}, && z = r, r+1, \ldots, \end{aligned}$$

and we say Y and Z follow the *geometric distribution* and the *negative binomial distribution* (or *Pascal distribution*), respectively.

The last section of Chapter 5 is devoted to the *Poisson distribution*, which has the following form:

$$\Pr(X = x) = \frac{\lambda^x \mathrm{e}^{-\lambda}}{x!} \qquad \text{for } x = 0,\, 1,\, 2,\, \ldots, \tag{1}$$

where λ is the parameter of the Poisson distribution and hence is a number that should be provided in (or can be obtained in some way from) the question. This distribution can be used to describe the number of occurrences of some random event (e.g. telephone calls, traffic accidents, customer arrivals) in an interval of time or a bounded space. It can be obtained (if you know calculus) as the limit of the binomial distribution by dividing the finite interval (or bounded space) into n equal parts with the assumptions that

1. the occurrences of the event are independent,
2. in any infinitesimally small part of the interval (space), the probability of more than one occurrence of the event is negligible,
3. the probability of one occurrence of the event in an infinitesimally small part of the interval (space) is p,

and then (if you know calculus) letting $n \to \infty$ and $p \to 0$ in such a way that $np = \lambda$ is a finite positive constant, which is the mean number of occurrences of the event in the finite interval or bounded space. Therefore, although the Poisson distribution itself is a distribution that has many applications having no direct connection with the binomial distribution, it can also be used as an approximation for the binomial distribution when n is large and p is small, i.e.

$$\binom{n}{x} p^x (1-p)^{n-x} \approx \frac{(np)^x \mathrm{e}^{-np}}{x!},$$

when $n \geq 100$ and $np < 10$. I have not yet mentioned this approximation in class but will do so when I discuss another approximation for the binomial distribution next week.

Considering Poisson as the limit of binomial, we could believe that the mean and the variance of the Poisson distribution are both λ, but it is also possible (again, if you know calculus) to derive them directly from the probability expression given in equation (1).

Note that using the Poisson distribution to describe the number of occurrences is often an assumption, rather than a mathematical result. However, usually it will be clear from the question that one must assume Poisson, e.g. when the only information (parameter) provided in the question is the mean number of occurrences (or the mean number of occurrences per unit time or per unit area).

For students who do not know calculus: the above "limiting" argument is not important; what is important is that you can identify which questions are asking you to use the Poisson distribution and you can calculate the numerical values by your scientific calculators: the calculation of $\mathrm{e}^{-\lambda}$ is the inverse of taking 'ln' (known as the 'natural log') of $-\lambda$.

To conclude Chapter 5, we discussed the meaning of fair games, in which the expected gain per game is zero (so that neither the host nor the guest has advantages), and then we offered an interpretation of *subjective probabilities* by considering the implicitly implied *betting odds* behind a subjective probability statement.

The random variables we considered in Chapter 5 concern counting: counting the number of successes, the number of trials or the number of occurrences. The values must be integers. However, if I take measurements, e.g. heights or weights, what I will get is not necessarily integer. Just a simple example, if we take a number randomly and uniformly (i.e. every number has the same chance to be chosen) from $[0,1]$ and call it X. What should the value $\Pr(X = 0.173690077)$ be?

The distributions appeared in Chapter 5 are therefore called *discrete*. The above example (a random number between 0 and 1) leads to the notion of a *continuous* distribution, and we can immediately notice that it does not make sense to talk about probability that a continuous X is equal to a particular number. We must formulate our question in some other way. A possibility is the *cumulative distribution function (c.d.f.)* (or simply called the *distribution function* in more advanced books) of a random variable X (either discrete or continuous), which is defined by

$$F(x) = \Pr(X \le x) \qquad \text{for all } x.$$

It is clear from its definition that (because it is cumulating probabilities, which are non-negative) the c.d.f. is a monotonically increasing function (meaning that $F(a) \le F(b)$ whenever $a < b$) satisfying that $F(-\infty) = 0$ and $F(\infty) = 1$. In particular, if X is *uniformly distributed* on $[a,b]$ (meaning that a random number between a and b is chosen in such a way that every number between a and b is equally likely), then

$$F(x) := \Pr(X \le x) = \begin{cases} 0, & x < a, \\[2ex] \dfrac{x-a}{b-a}, & a \le x \le b, \\[2ex] 1, & x > b. \end{cases}$$

To reflect the fact that every number has the same chance, we focus on the slope of $F(x)$.

The slope (or, for those who know calculus, the derivative) of $F(x)$ is

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b, \\ \\ 0, & \text{otherwise.} \end{cases}$$

The constant slope (derivative) within the interval $[a, b]$ tells us exactly the characteristic of the *uniform distribution*: every number within $[a, b]$ has the same (constant) chance to be chosen.

When $f(x)$ is called the derivative of $F(x)$, then it is very natura that $F(x)$ is called the anti-derivative of $f(x)$, and in calculus we also call it the integral of $f(x)$.

To get the probability for a continuous random variable from a given $f(x)$, we have to work out the anti-derivative (or the integral for those who know calculus) of $f(x)$, which actually is equal to the area under the curve of $f(x)$. As an analogue to *density* (mass per unit volume) in Physics, the function $f(x)$ is called the *probability density function (p.d.f.)*. If we want to get $\Pr(a \le X \le b)$, what we have to do is to calculate the area under the probability density function between a and b. Obviously, $\Pr(X = b) = \Pr(b \le X \le b) = 0$ because the area under the probability density function between b and b is zero. Therefore, for a continuous random variable X,

$$\Pr(X \le b) = \Pr(X < b) + \Pr(X = b) = \Pr(X < b).$$

Changing the p.d.f. from a constant to something else will lead to other interesting distributions. The most well-known continuous distribution is the *normal distribution*, also known as the *Gaussian distribution*, which is the most important distribution in this course.

I hope the non-math/non-stat majors will appreciate the mathematics in the last few weeks. Even if you don't, you need not worry too much. The previous few weeks form our appetizer only, and we have not yet started our main course, which will begin from Chapter 8, which will commence next week. (Of course, the most important part of a dinner is the dessert, which is our final examination!)

Enjoy your Assignment 2.

Cheers,
Heng Peng