Discrete probability distributions

In the chapter on probability we used the classical method to calculate the probability of various values of a random variable. In some cases, however, we may be able to develop a mathematical function for calculating such probabilities based on the definition of the random variable. In general, a random variable may be *discrete* or *continuous*. A random variable is discrete when it can assume countable values, and is continuous when it can take on values on a continuous scale. Usually, if you count it, it is discrete, and if you measure, it is continuous.

In this chapter, we will develop mathematical functions to calculate probabilities associated with discrete random variables. Accordingly, such a function is called *probability function*, *probability mass function*, or *probability distribution*. These functions will be directly driven from the definition of the random variable. For convenience, we shall denote a random variable by a capital letter, while the corresponding small letter will be used to denote a specific value of the random variable. For example, $X$ may represent the number of students passing a course. $x$ then, will denote a specific value of $X$. The probability function will be denoted by $p(x)$, where $p(x)$ will assign $P\{X = x\}$.

Probability functions must satisfy certain conditions that are directly related to the definition of a probability. These conditions are:

1. $p(x) \geq 0$. (Probabilities must be positive).
2. $\sum_{all \, x} p(x) = 1$. All possible outcomes (sample space) must have a probability of 1.

Associated with each random variable or distribution is a cumulative distribution function, $F(x)$, which assigns cumulative probabilities. $F(x) = \sum_{t \leq x} p(t)$ for $-\infty < t < \infty$. In the following section, we will define some standard discrete random variables and develop their probability distributions.

**Standard Discrete Random Variables**

The standard discrete probability distributions presented here are all based on the Bernoulli trial. The Bernoulli trial is any experiment or activity with two possible outcomes. For convenience, we will call these outcomes success and failure; with success being the desired outcome. For example, when shooting at a target, one may hit or miss. In this case success represents the outcome of hitting the target, while failure represents missing it. We will assume that the probability of success in the Bernoulli trial is $p$, and the probability of failure is $q$. Obviously, $p+q=1$, or $q=1-p$. Now, we may define a random variable $X$ as the number of successes in a Bernoulli trial. Obviously, $X$ is a discrete random variable with two possible values 0 and 1. We may define its probability function as:

$$p(x) = \begin{cases} p^x q^{1-x}, & x = 0, 1 \end{cases}.$$  

$p(x)$ is called the Bernoulli distribution and it clearly assigns the right probabilities and satisfies the two conditions above (i.e., $p(x)$ is always positive and $\sum_{x=0,1} p(x) = p + q = 1$). The Bernoulli distribution is rather very simple, but is a foundation for many powerful distributions.
The Binomial Distribution

If we assume that we have a series of independent Bernoulli trials, we may be interested in the number of successes in these trials. The word independent means that the probability of success does not change from one trial to another. The number of successes is clearly a discrete random variable \(X\) and in a series of \(n\) independent Bernoulli trials, it may assume any discrete value from 0 to \(n\) (i.e., 0, 1, 2, \(\ldots\), \(n\)).

For a given \(x\), we must have \(x\) successes and, accordingly, \(n-x\) failures. In other words, we need success and success and success \(\ldots\) etc. \(n\) times, failure and failure and failure \(\ldots\) etc. \(n-x\) times. In simpler terms, we are looking for the intersection of \(x\) successes and, accordingly, \(n-x\) failures. Since these are all independent events, they will happen with a probability \(p^xq^{n-x}\). Still, the \(x\) successes may be located in \(\binom{n}{x}\) ways within the \(n\) trials. Therefore,

\[
p(x) = \binom{n}{x} p^x q^{n-x}, x=0,1,\ldots,n.\]

The figure to the right explains how probabilities are calculated for 2 successes in 4 trials. 2 successes and 2 failures have a probability of \(p^2q^2\). This outcome may occur in 6 different ways leading to:

\[
P(X=4) = 6 \; p^2q^2.\]

Using our probability function, \(p(2) = \binom{4}{2} p^2q^{4-2} = 6 \; p^2q^2\). Because of the form of the probability function, it is called the binomial distribution. Notice that the Bernoulli distribution is a special case of the binomial when \(n = 1\).

The binomial distribution satisfies the two conditions above. It is always positive and

\[
\sum_{x=0}^{n} p(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (p+q)^n = (1)^n = 1.
\]

An interesting question is on average, how many successes are expected in such situation? This is called the expected value of \(X\) and is denoted by \(E[X]\). \(E[X]\) is the long-term average of \(X\), or simply the mean of \(x\) and may also be denoted by \(\mu\). If we made \(n\) independent Bernoulli trials and recorded the number of successes and repeated this process so many times, \(\mu\) or \(E[X]\) would be the average of the recorded successes. \(\mu\) or \(E[X]\), where \(X\) is any discrete random variable, may easily be calculated as \(\mu = E[X] = \sum_{all\; x's} xp(x)\), which simply weighs each value of \(X\) by its probability to get the average. In general, \(E[g(x)] = \sum_{all\; x's} g(x)p(x)\), where \(g(x)\) is any function of \(X\) and \(E[g(x)]\) is the long-term average of that function.

The following rules follow from the definition of \(E[X]\) and \(E[g(x)]\):

1. \(E[c] = c\), where \(c\) is a constant.
2. \(E[cX] = cE[X]\).
3. $E[X \pm Y] = E[X] \pm E[Y]$, $X$ and $Y$ are random variables.

For the Bernoulli trial, $\mu = E[X] = \sum_{x=0}^{1} xp(x) = 0(q) + 1(p) = p$

For the binomial distribution, $\mu = E[X] = \sum_{x=0}^{n} xp(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} = np$.

This result for the binomial distribution could easily be obtained by observing that the binomial is the summation of $n$ independent Bernoulli distributions, and from the third rule above, its mean would be the summation of the means, i.e., $np$.

$\mu$ or $E[X]$ is a measure of center. Another important measure is the variance, which measures how much the random variable varies around its mean. The variance is the average square distance from the mean and is denoted by $\sigma^2$. From its definition, $\sigma^2 = E[(X-\mu)^2] = E[(X-E[X])^2]$. From the expectation rules above, $\sigma^2 = E[X^2] - \mu^2$ or $\sigma^2 = E[X^2] - (E[X])^2$.

The following rules follow from the definition of $V[X]$:

1. $V[c] = 0$, $c$ is a constant
2. $V[cX] = c^2 V[X]$
3. $V[X \pm Y] = V[X] + V[Y]$, $X$ and $Y$ are independent random variables

For the Bernoulli distribution, $E[X^2] = \sum_{x=0}^{1} x^2 p(x) = 0^2(q) + 1^2(p) = p$, $V[X] = \sigma^2 = E[X^2] - (E[X])^2 = p - p^2 = p(1-p) = pq$.

Again, since the binomial is the summation of $n$ independent Bernoulli distributions, for the binomial distribution $\sigma^2 = npq$.

**Example:**

The probability that a patient survives a critical heart operation is 0.9. What is the probability that exactly 5 out of the next 7 patients having this operation survive?

**Solution:**

We may consider each operation as a Bernoulli trial with two possible outcomes: survival ($p=0.9$) and death ($q=0.1$). The 7 ($n$) operations are then a series of independent Bernoulli trials where we are interested in having 5 ($x$) successes, leading to the binomial distribution. We are assuming that the trials are independent because they are performed on different patients. However, if these operations are performed by the same team, their experience may lead to changing $p$ from operation (trial) to operation. Nevertheless, we will assume independence (constant $p$).

Using the binomial distribution, $P[X=5] = p(5) = \frac{7}{5} (0.9)^5 (0.1)^2 = (21)(0.505)(.01) = 0.124$
The Negative binomial (Pascal) Distribution

With a series of independent Bernoulli trials we may be interested in the number of trials (X) to obtain a specific number of successes (r). X is clearly a random variable that could be anywhere from r to ∞. In other words, we may be very lucky and succeed in every trial, and thus get the r successes in r trials. On the other hand, we may be unlucky and continue to try without getting the r successes. In general however, our last trial will always be a success (the rth success), otherwise we would have to continue, while the remaining r-1 successes could be located anywhere in the remaining x-1 trials in \( \binom{x-1}{r-1} \) ways. Each one of these locations has a probability of \( p^r q^{x-r} \) and thus the overall probability of needing x trials to obtain r successes is

\[
p(x) = \binom{x-1}{r-1} p^r q^{x-r}, x=r, r+1, r+2, \ldots, \infty.
\]

The figure to the right explains this development for the case of needing 4 trials to get 2 successes. Notice how the last trial is always the last success. The definition of this random variable is the opposite of the binomial case. Here the random variable is the number of trials to get specific successes, while in the binomial case, the random variable is the number of success in a specific number of trials. Therefore, the resulting distribution is called the negative binomial. It is also sometimes called the Pascal distribution after the person who developed it.

The mean and variance of the Pascal distribution are given by:

\[
\mu = \frac{r}{p}, \quad \sigma^2 = \frac{rq}{p^2}
\]

A special case of the Pascal distribution results when we are only interested in obtaining one success, i.e., \( r = 1 \). In this case, \( p(x) \) simplifies to

\[
p(x) = \binom{x-1}{0} p^1 q^{x-1} = pq^{x-1}, x=1, 2, 3, \ldots, \infty.
\]

This new random variable, the number of successes to obtain ONE success, is called the geometric random variable, and the resulting probability function is called the geometric distribution. The mean and variance of the Geometric distribution are given by:

\[
\mu = \frac{1}{p}, \quad \sigma^2 = \frac{q}{p^2}
\]

Observe that the Pascal random variable may be viewed as the summation of \( r \) independent Geometric random variables.

Example:

A scientist inoculates several mice, one at a time, with a disease germ until she finds 2 that have contracted the disease. If the probability of contracting the disease is 0.15, what is the probability that:

a. 8 mince are required?

b. Less than 5 mice are required?
Solution

a. We are looking to infect 2 mice (2 successes) and we are interested in needing to try 8 times to achieve this. This satisfies the Pascal distribution with \( r = 2, \ x = 8 \). Therefore,

\[
P(x=8) = p(8) = \binom{8-1}{2-1} p^2 q^{8-2} = \binom{7}{1} (0.15)^2 (0.85)^6 = (7)(0.0225)(0.37715) = 0.0594.
\]

b. Less than 5 may be satisfied by 2, 3, or 4, i.e., we are looking for \( P(x \leq 4) \), or \( F(4) \).

\[
F(4) = \sum_{x=2}^{4} p(x) = \sum_{x=2}^{4} \binom{x-2}{2-1} p^2 q^{x-2} = p + p^2 q + 2p^2 q^2 = 0.0225 + 0.0191 + 0.0325 = 0.0741.
\]

Example:

The probability that a student pilot passes the written test for his private pilot’s license is 0.7. Find the probability that a person passes the test:

a. on the third try
b. before the fourth try

Solution

Since the person needs to pass the test only once, we are looking for the number of trials to obtain one success, i.e. the Geometric distribution.

Notice that, in using the Geometric distribution, we are implicitly assuming that these trials are independent. In reality, however, one would expect a higher probability of passing the test on the second trial due to the learning effects. Nevertheless we will proceed with the independence assumption assuming that these effects are minimal.

a. \( x = 3, \ p(3) = p^1 q^2 = (0.7)^1 (0.3)^2 = 0.063 \). Notice that the probability is low because the person is much more likely to pass it on the first trial (0.7), or the second (0.21).

b. Before the fourth trial may be satisfied by the first or the second or the third, i.e., we are looking for \( P(x \leq 3) \), or \( F(3) \).

\[
F(3) = \sum_{x=1}^{3} p(x) = \sum_{x=1}^{3} pq^{x-1} = p + pq + pq^2 = 0.7 + 0.21 + 0.063 = 0.973.
\]

The Poisson Distribution

A very useful random variable and probability distribution results from the Poisson process in which we count discrete occurrences in a continuous interval. For example, we may be interested in counting the number of customers arriving to a service facility in a given time interval, or counting the number of white blood cells in a drop of blood, or counting the number of traffic accidents at a particular intersection over a week. We will assume that the rate of occurrence per unit time or measurement unit is a known constant, and will denote by \( \lambda \). To develop the Poisson distribution, we make the following assumptions:

1. The number of outcomes in a given interval or region is independent of the number that occurs at any other disjoint interval or region. That is to say that the Poisson process has no memory.
2. The probability that a single outcome will occur during a very short time interval or a very small region \((t)\) is proportional to \(t\).

3. The probability that more than one outcome will occur in \(t\) is negligible.

The Poisson distribution may be developed from the binomial distribution. For example, consider the case of counting the number of arrivals in a time interval \((t)\). If the rate of arrival is \(\lambda\) per time unit, then the average number of arrivals in \(t\) is \(\mu=\lambda t\). If we divide \(t\) to a huge number \((n)\) of infinitesimal intervals such that only one arrival can occur in each infinitesimal interval with a very small probability \((p)\), we may consider each such interval as a Bernoulli trial with only two possible outcomes: arrival (success) and no arrival (failure). Now, the series of infinitesimal intervals is a series of \((n)\) Bernoulli trials and we are interested in the number of arrival (successes) \(X\); i.e., \(X\) is binomially distributed with \(n \to \infty, p \to 0, \mu = np, X = 0, 1, 2, \ldots, \infty\).

In the binomial distribution, if we take the limit as \(n \to \infty\) and substitute \(p = \mu/n\), we get:

\[
\lim_{n \to \infty} p(x) = \lim_{n \to \infty} \binom{n}{x} \left(\frac{\mu}{n}\right)^x \left(1 - \frac{\mu}{n}\right)^{n-x} = \lim_{n \to \infty} \frac{n(n-1) \ldots (n-x+1)}{x!} \left(\frac{\mu}{n}\right)^x \left(1 - \frac{\mu}{n}\right)^{n-x} = 1(1- \frac{1}{n})^{n-x} \left(1 - \frac{\mu}{n}\right)^{n-x} = 1,
\]

However, \(\lim_{n \to \infty} 1(1- \frac{1}{n})^{n-x} = 1\), \(\lim_{n \to \infty} \left(1 - \frac{\mu}{n}\right)^{-x} = 1\),

\[
\lim_{n \to \infty} \left(1 - \frac{\mu}{n}\right)^n = \lim_{n \to \infty} \left[\left(1 + \frac{1}{-n/\mu}\right)^{-n/\mu}\right]^{-\mu} = e^{-\mu}
\]

\[
\therefore p(x) = \frac{e^{-\mu} \mu^x}{x!}, x = 1, 2, \ldots, \infty.
\]

The last equation is the distribution function of the Poisson random variable that satisfies the three conditions above. The mean of the Poisson distribution is \(\mu\) and the variance is \(\mu = \lambda t\).

**Example:**

The average number of oil tankers arriving each day at a certain port is known to be 1 per hour. The port can handle at most 12 tankers per 8-hour day. What is the probability on a given day tankers will have to be sent away?

**Solution:**

Let \(X\) represent the number of arrivals per day. \(X\) is Poisson distributed with a mean \(\mu = \lambda t = 1 \times 8 = 8\) tankers/day.

\[
P(X > 12) = 1 - P(X \leq 12) = 1 - F(12) = 1 - \sum_{x=0}^{12} p(x) = 1 - \sum_{x=0}^{12} \frac{e^{-\mu} \mu^x}{x!} = 1 - \sum_{x=0}^{12} \frac{e^{-8} 8^x}{x!} = 1 - e^{-8} = 1 - 0.936 = 0.064
\]

The cumulative results were obtained using Microsoft Excel, which has preprogrammed functions for all distributions covered here.
The Hypergeometric Distribution

All distributions described above were based on a series of independent Bernoulli trials. A situation in which we obtain a series of dependent Bernoulli trials comes about in sampling without replacement. For example, assume that we are interested in accepting a lot of 100 parts if it contains 5 or fewer defectives. We decide to take a sample of 10 parts, one by one, without replacement and count the number of defectives in the sample, and use that number to judge the quality of the lot. This inspection is a series of Bernoulli trials each with two possible outcomes: defective and not defective. If the lot contains 5 defectives, then the probability that the first part on the sample will be defective is \( \frac{5}{100} = 0.05 \). Now, the probability of the second part being defective depends on the results of the first part. If the first part is defective, we are left with 99 parts out of which only 4 are defective and accordingly, the probability of the second part being defective is \( \frac{4}{99} = 0.0404 \). If, however, the first part is not defective, we are left with 99 parts out of which 5 are defective, and accordingly, the probability of the second part being defective is \( \frac{5}{99} = 0.0505 \). Hence, we have a series of Bernoulli trials in which the probability of success in each trial depends on the results of all proceeding trials. We will now develop the probability function for this random variable, \( X \), defined as the number of items with a given characteristic in a sample \( n \) taken from a lot \( N \) that has \( k \) items possessing the desired characteristic. The diagram to the right depicts this situation. The \( x \) items in the sample must come from the \( k \) items in the lot, while the remaining items in the sample \( (n-x) \) will have to come from the remaining items in the lot \( (N-k) \). Using the classical approach,

\[
p(x) = \frac{\text{number of ways of choosing } x \text{ out of } k \text{ and } n-x \text{ out of } N-k}{\text{Number of ways } n \text{ out of } N}
\]

\[
\therefore p(x) = \binom{k}{x} \binom{N-k}{n-x} \frac{n}{N}, x = \max(0, n - N + k) \ldots \min(k, n).
\]

The range of \( X \) starts from either 0 or \( n- (N-K) \), whichever is larger, and ends at either \( n \) or \( k \), whichever is smaller. \( X \) is called a hypergeometric random variable and \( p(x) \) is called the hypergeometric distribution. The mean and variance of the hypergeometric distribution are given by

\[
\mu = \frac{nk}{N}, \quad \sigma^2 = \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N} \left(1 - \frac{k}{N}\right).
\]

Notice that if we replace \( k/N \) by \( p \), we get \( \mu = np \), and \( \sigma^2 = \frac{N-n}{N-1} \cdot np \), which are the mean and the variance of the binomial distribution if the term \( \frac{N-n}{N-1} \) is close to one, which is the case
when \( n \) is small relative to \( N \). Consequently, when \( n \) is small relative to \( N \), the binomial distribution with \( p = k/N \) may be used to approximate the hypergeometric distribution.

**Example:**

A committee of 5 members is formed by randomly choosing from a class of 40 students in which only 3 are freshmen. What is the probability that the committee will include:

a. exactly one freshman?

b. at most one freshman?

**Solution:**

a. Using the hypergeometric distribution with \( n = 5, N = 40, k = 3, x = 1 \), we find the probability of one freshman to be

\[
p(1) = \frac{\binom{3}{1} \binom{37}{4}}{\binom{40}{5}} = \frac{(3)(66045)}{658008} = 0.3011
\]

b. At most 1 is satisfied by 0 or 1.

\[
p(0) = \frac{\binom{3}{0} \binom{37}{5}}{\binom{40}{5}} = \frac{(1)(1435897)}{658008} = 0.6624
\]

\[
P(X \leq 1) = p(X = 0) + p(X = 1) = 0.6624 + 0.3011 = 0.9635, \text{ a very high probability as we expect the majority of committee members to be non-freshmen.}
\]