4. DISCRETE DISTRIBUTIONS

*When you fall in a river, you're no longer a fisherman; you're a swimmer.*

– Gene Hill

It seems as though most everything an engineering student does is based on numbers—finding reactions, flow rates, principal stresses, whatever. The outcome of an experiment need not be a number, though. When you flip a coin, the outcome is either “heads” or “tails.” When you draw a card, it is either a “club” or a “heart” or a “diamond” or a “spade.” It is often more convenient to represent outcomes as numbers, which is where the concept of a **random variable** comes in.

A **random variable** is a “function” that associates a unique numerical value with every outcome of an experiment. The value of the random variable will vary from trial to trial as the experiment is repeated. For example, if I toss a pair of dice, the number of dots that appear could be considered as a random variable that takes on values from 2 to 12. If I flip a coin, a result of “heads” could be assigned a value of 1 and “tails” a value of 0. If I’m measuring the speed of vehicles on Central Avenue, the speed itself can be used as the random variable because it’s already a number.

A **probability distribution** is a “function” that defines the probability of occurrence of every possible value that a random variable can take on.

For example, let’s consider the probabilities associated with tossing a pair of dice:

<table>
<thead>
<tr>
<th>$x_i$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
</table>

The probability distribution function, $p(x)$, can be plotted as a **probability curve**:
DISCRETE AND CONTINUOUS VARIABLES

There are two general types of probability distributions—*discrete* and *continuous*—and the distinction between them depends on the nature of the values that the random variable can take on.

A *discrete random variable* can only take on discrete (i.e., specific) values. For example, going back to the die tossing illustration, I can toss a 3, or I can toss a 4. I cannot, however, toss a 3.62. That's why we didn't connect the “dots” in the probability curve — the probability distribution is undefined for values other than 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, so it would misrepresent the probability distribution to connect the dots with lines or curves.

A *continuous random variable* takes on continuous values (i.e., real values). Distance and speed are two illustrations of continuous random variables.

To illustrate the fundamental difference between discrete and continuous random variables, suppose we are sampling the depth of a lake. What is the probability that the lake is exactly 9 feet deep (to an infinite number of zeros) at a particular location? The nature of *continuous* random variables (covered in the next chapter) is that the probability of their taking on any exact value is zero. We must therefore resort to other means when describing the probabilities associated with continuous random variables. We will do this by describing the probability that the random variable takes on a value within a certain range (e.g. a depth of nine feet or less, or a depth between 9.5 and 10.5 feet).

PROPERTIES OF DISCRETE DISTRIBUTIONS

There are a number of basic properties of statistical distributions, and they are similar (but not identical) for both discrete and continuous distributions.

The *probability distribution function* (also called the *probability mass function*) gives the probability that the random variable \( X \) will take on a value of \( x \) when the experiment is performed:

\[
p(x) = P(X = x)
\]

This equation may look a little weird if you don’t know the ground rules. In probability and statistics, we always use uppercase letters (\( X \)) as the names of random variables and we use the equivalent lowercase letter (\( x \)) to refer to some unspecified value that the variable can take on.

Note, too, that we use a lowercase \( p(\cdot) \) to denote the *probability mass function* but we use the uppercase \( P(\cdot) \) to denote the probability of occurrence (as in the last chapter). Confused yet?

By definition, \( p(x) \) is always a number between zero (never happens) and one (always happens):

\[
0 \leq p(x) \leq 1
\]

and, since every trial must have exactly one outcome,

\[
\sum x p(x) = 1
\]
Think about this. A sum less than one would imply that there’s always a possibility that there will be no outcome from a trial, as if the trial was never performed. That wouldn’t make any sense.

The **cumulative distribution function** gives the probability that the random variable $X$ will take on a value less than or equal to $x$ when the experiment is performed:

$$F(x) = P(X \leq x) = \sum_{x \leq x} p(x)$$

The **expected value** of a random variable is the probability-weighted average of the possible outcomes:

$$E(X) = \mu_x = \sum x \cdot p(x)$$

Huh? Think of it this way ... each time you conduct an experiment, you get a different outcome, but if you conduct the same experiment many, many times, the outcomes average out to the expected value. If you consider the set of all measured outcomes as the *population*, then $E(X)$ is just the population mean! That’s why we sometimes give it the symbol $\mu_x$.

Think back to tossing the coin. If we let $X = 1$ if the coin comes up “heads” and let $X = 0$ if the coin comes up “tails,” then the expected value of $X$ is 0.5 because “1” and “0” each come up 50% the time.

Another way to think of the expected value is as a return on investment. It’s the sum of all possible “rewards” weighted by the probability that you’ll actually see those rewards.

**Illustration**

Suppose an organization is having a raffle and selling tickets for $2.00 each. First prize is $1000, second prize is $500, and third prize is $250. All other entrants receive nothing. If 999 tickets have already been sold and you have the opportunity to buy the 1000th (and last) ticket, should you?

The expected return on your $2 investment can be computed as

$$E(X) = 250 \cdot p(250) + 500 \cdot p(500) + 1000 \cdot p(1000)$$

Now, what are the probabilities?

Well, there’s one chance in 1000 of winning $1000, so $p(1000) = 1/1000$. Likewise, there’s one chance in 1000 of winning $500, so $p(500) = 1/1000$. There’s also one chance in 1000 of winning $250, so $p(250) = 1/1000$. The other 997 tickets win nothing, so $p(0) = 997/1000$.

Have we just defined a probability distribution? You bet. We’ve enumerated all four possible outcomes ($0, 250, 500, 1000$) and assigned each a probability of occurrence between 0 and 1. Furthermore, the probabilities we’ve assigned add up to 1, which is a requirement of probability distributions:

$$\frac{997}{1000} + \frac{1}{1000} + \frac{1}{1000} + \frac{1000}{1000} = 1$$

Getting back to the problem, the expected return on our $2 investment is:
What does this mean? Can you have a return of $1.75? No, you can either have $1000, $500, $250, or $0. But if you kept buying tickets to similar raffles with the same odds of winning, then as the number of tickets approaches infinity, your average return would be $1.75. That’s how casinos make their money!

The expected value measures where the probability distribution is geometrically centered. It’s a measure of central tendency! There should therefore be a measure of dispersion to go along with it. And there is!

The variance of a probability distribution is a measure of the amount of variability in the distribution of the random variable, $X$, about its expected value.

For example, the earlier dice tossing distribution was centered at 7, but there were 11 possible outcomes that were not 7. In other words, there was some variability in the distribution. Suppose we had a situation where one die had 3 spots on all 6 faces and the other die had 4 spots on all 6 faces. There are still 12 possible outcomes, but the sum of the dots will always be 7. Now, both distributions have an expected value of 7, but they certainly aren't the same distribution! Variance, then helps us describe the difference in those distributions.

Mathematically, the variance is just the probability-weighted average of the squared deviations:

$$V(X) = \sigma_x^2 = \sum_x (x - \mu_x)^2 p(x)$$

This is actually the exact same definition we used for the population variance in the first chapter. There, each value $x$ represented one of $n$ members of the population, so each $x$ was weighted as $1/n$:

$$\sigma_x^2 = \frac{\sum_{i=1}^n (x_i - \mu_x)^2}{n}$$

Recalling that the expected value of a probability distribution is the probability-weighted average of the $x$ values, we can also write the variance as the expected value of the squared deviations:

$$V(X) = E\left[(x - \mu_x)^2\right]$$

Without proving it here, we can also calculate the variance as

$$V(X) = E(X^2) - \left[E(X)\right]^2$$

which is easier to compute than squaring a bunch of deviations.
BERNOULLI TRIALS

Heads or Tails? Win or Lose? Boy or Girl? Buttered Side Up or Buttered Side Down? It’s surprising how many times an experiment has just two possible outcomes. When there are only two possible outcomes to an experiment, we say that the experiment is a Bernoulli trial.

Actually, to be considered a Bernoulli trial, an experiment must meet three criteria:

1. There must be only 2 possible outcomes. One of these outcomes is called “success” and the other is called “failure”, even though one outcome may be just as desirable as the other.
2. Each outcome must have an invariant probability of occurring. The probability of success is usually denoted by $p$, and the probability of failure is denoted by $q = 1 – p$.
3. The outcome of each trial is completely independent of the outcome of any other trials.

There are a number of important probability distributions that describe compound events resulting from successive Bernoulli trials. The four we’ll study here are as follows:

- **Binomial Distribution** ................................. Gives the probability of exactly $x$ successes in $n$ successive trials
- **Negative Binomial Distribution** ....... Gives the probability that it will take exactly $n$ successive trials to produce exactly $x$ successes
- **Geometric Distribution** .............................. Gives the probability that it will take exactly $n$ successive trials to produce the first success
- **Poisson Distribution** ................................. Gives the probability of exactly $x$ successes over some continuous interval of time or space

**HOW TO DETERMINE WHICH DISTRIBUTION TO USE**

*There is a time in the life of every problem when it is big enough to see, yet small enough to solve.*

– Mike Leavitt

It's really not as hard as you think. Usually, you start off with a seemingly insurmountable word problem. First, “bash” the thing with the basic laws and axioms into a number of smaller problems. Keep doing this, until you see the definition of one of the discrete distributions.

With discrete distributions, there should never be any doubt as to which distribution to use. (Continuous distributions, however, are different). Let's examine the discrete distributions themselves and work through a few examples.

Just concentrate on the differences between the various distributions. You’ll get the hang of it quickly.
THE BINOMIAL DISTRIBUTION

Gives the probability of exactly $x$ successes in $n$ trials

A. Requirements

There must be $x$ successes and $(n-x)$ failures in the $n$ trials, but the order in which the successes and failures occur is immaterial.

B. Mathematical Relationships

1. General Equation:

$$p(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}$$

where

$n = \text{the number of trials}$

$x = \text{the number of successes}$

$p = \text{the probability of a success for any given trial}$

$q = 1 - p = \text{the probability of a failure for any given trial}$

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} = \text{the Binomial Coefficient}$$

The Binomial Coefficient is the possible number of arrangements of $n$ things taken $x$ at a time. For example, in how many ways can Coach Calipari pick a starting five from a team of nine players?

$n = 9$

$x = 5$

$$\binom{9}{5} = \frac{9!}{5!(9-5)!} = 126$$

2. Expectation – the expected (mean) number of successes in $n$ trials

$$E(X) = np$$ successes

3. Variance – the expected sum of the squared deviations from the mean

$$V(X) = npq$$
At this point, you might be asking “Why is it called the binomial distribution?” Well, the equation

\[ p(x) = \binom{n}{x} p^x q^{n-x} \]

actually describes an infinite number of probability distributions corresponding to all the possible values of \( n \) and \( p \). But once \( n \) and \( p \) have been defined, then the equation describes the probability of all \( n + 1 \) possible outcomes, from 0 successes in \( n \) trials to \( n \) successes in \( n \) trials. Furthermore, the \( n + 1 \) probabilities add up to exactly one.

C. Example

On the average, 30% of the vehicles on the southbound approach to an intersection turn right. What is the probability that none of the next 10 “arrivals” turn right?

Solution:

Define “trial” as a southbound approaching vehicle
Define “success” as a right turning vehicle
\( n = 10 \) “trials”
\( x = 0 \) “successes”
\( p = 0.3 = P(\text{success}) \)

\[ p(0) = P(X = 0) = \frac{10!}{0!(10 - 0)!} 0.3^0 0.7^{10-0} = 0.7^{10} = 0.0282 \]

This problem may also be worked using the basic laws and axioms.

What I want is the probability that the first car does not turn right and the second car does not turn right and the third car does not turn right and...and the 10th car does not turn right.

\[ p(\text{does not turn right}) = 1 - p(\text{turns right}) = 1 - 0.7 = 0.3 \]

Are these events independent? Well, unless there is some kind of parade, each car proceeds through the intersection in a direction related to its particular destination, which should be independent of the destinations of the other drivers.

Assuming all 10 events are independent, I can simply multiply their probabilities:

\[ p(0) = 0.7 \times 0.7 \times \ldots \times 0.7 = 0.7^{10} = 0.0282 \]

which is the same solution as before.
THE NEGATIVE BINOMIAL DISTRIBUTION

Gives the probability that it will take exactly $n$ trials to produce exactly $x$ successes

A. Requirements

The last ($n$th) trial must be a success, otherwise the $x$th success actually occurred on an earlier trial. This means that we must have $(x - 1)$ successes in the first $(n - 1)$ trials plus success on the $n$th trial.

B. Mathematical Relationships

1. General Equation:

$$p(n) = p[(x - 1) \text{ successes in } (n - 1) \text{ trials}] \times p(\text{success on the } n^{\text{th}} \text{ trial})$$

Using the Binomial Distribution,

$$p[(x - 1) \text{ successes in } (n - 1) \text{ trials}] = \binom{n - 1}{x - 1} p^{x-1} q^{n-x}$$

and

$$p(\text{success on the } n^{\text{th}} \text{ trial}) = p(\text{success on any trial}) = p$$

so,

$$p(n) = \binom{n - 1}{x - 1} p^{x-1} q^{n-x} p$$

or

$$p(n) = \binom{n - 1}{x - 1} p^{x} q^{n-x}$$

2. Expectation – the expected number of trials to produce $x$ successes

$$E(N) = \frac{x}{p} \text{ trials}$$

3. Variance – the expected sum of the squared deviations from the mean

$$V(N) = \frac{xq}{p^2}$$
The random variable here is $N$, the number of trials needed, not $X$. Once you specify values for $x$ and $p$, there are still an infinite number of possible outcomes, ranging from $n = x$ to $n = \infty$. Still, the equation

$$p(n) = \binom{n-1}{x-1} p^x q^{n-x}$$

is a probability distribution because it gives a probability of occurrence for all possible outcomes and, believe it or not, the probabilities sum to one despite there being an infinite number of them.

C. Example

In order to perform street improvements, it is sometimes necessary to get a R.O.W. acquisition from all property owners having property which fronts on the street. Typically 85% of all property owners are willing to sign the necessary paperwork with little or no argument. Condemnation proceedings must generally be taken against the other 15 percent in order to acquire the R.O.W.

On a given project, 40 parcels must be acquired in order for a road to be widened. As Project Manager, you know that condemnation proceedings can delay a project for months. What is the probability that all 40 property owners must be contacted before 2 of them fight the project in court?

Solution
THE GEOMETRIC DISTRIBUTION  
Gives the probability that it will take exactly \( n \) trials to produce the first success.

A. Requirements

The first success must occur on the \( n \)th trial, so we must have \((n - 1)\) failures in the first \((n - 1)\) trials plus success on the \( n \)th trial.

B. Mathematical Relationships

1. General Equation

Substituting \((x = 1)\) into the Negative Binomial equation:

\[
p(n) = \binom{n-1}{1-1} p^{1} q^{n-1} \Rightarrow p(n) = \frac{(n-1)!}{0!(n-1)!} p q^{n-1}
\]

or

\[
p(n) = pq^{n-1}
\]

2. Expectation – the expected number of trials to produce \( x \) successes

\[
E(N) = \frac{1}{p} \text{ trials per success}
\]

3. Variance – the expected sum of the squared deviations from the mean

\[
V(N) = \frac{q}{p^2}
\]

C. Example

Old-time mechanical slot machines have three cylinders with 128 stopping points on each cylinder for a total of 2,097,152 possible “outcomes.” Of these, 140,811 return at least some money. What is the probability that it will take 5 pulls before you get some money back?

Solution
THE POISSON DISTRIBUTION  

Gives the probability of exactly $x$ successes over some continuous interval of time or space.

With the Poisson distribution, we are concerned with the rate at which successes occur, or the number of successes per some “unit” (arrivals per 15 minutes, accidents per six months, potholes per mile, and so on). The Poisson distribution is still considered discrete because successes are discrete events, even though the interval used to define the success rate is continuous.

**A. Mathematical Relationships**

1. General Equation

   \[ x = \text{the number of successes in a (continuous) unit} \]

   \[ \lambda = \text{the average number of successes per unit} \]

   \[ p(x) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \]

   Note: The "unit" for both $\lambda$ and $x$ must be the same. For example, you can't have seconds associated with $\lambda$ and hours associated with $x$. In problem solving, the problem itself will define the “unit,” and the basis for it may be some off-the-wall interval, such as say, 17 days, rather than days or months or years.

2. Expectation – the expected number of successes in a unit

   \[ E(X) = \lambda \]

3. Variance

   \[ V(X) = \lambda \]

**B. Example**

An intersection has averaged 2 accidents per year over the past 10 years.

a. What is the probability that there will be 5 accidents at the intersection next year?

b. What is the probability that there will be 10 accidents at the intersection over the next two years?

c. What is the probability that there will be at least one accident at the intersection in the next 6 months?
Solutions
THE BINOMIAL – POISSON CONNECTION

The binomial distribution gives the probability of exactly \( x \) successes in \( n \) trials. If the probability of success (\( p \)) for any given trial is small and the number of trials (\( n \)) is large, the binomial distribution can be approximated by a Poisson distribution with \( \lambda = np \).

Why? Intuitively, if \( p \) is small and \( n \) is large, the long strings of “failures” between the infrequent “successes” start to look like continuous intervals rather than discrete events.

Generally speaking, the Poisson distribution provides a pretty good approximation of the binomial distribution as long as \( n > 20 \) and \( np < 5 \).

Example

Johnny is a good typist. On average, he makes just one typing mistake every hundred words. If you hire Johnny to type your 500-word essay on “Representations of Women in 19th Century Media,” what is the probability that there will be no errors? No more than two errors?
**DISCRETE DISTRIBUTIONS IN EXCEL**

Excel has built-in functions to analyze any of the discrete distributions we’ve covered here. For the Binomial distributions, Excel can give you both the value of the probability mass function,

\[ p(x) = \text{BINOMDIST}(x, n, p, \text{FALSE}) \]

and the value of the cumulative probability distribution,

\[ F(x) = \text{BINOMDIST}(x, n, p, \text{TRUE}) \]

The same is true of the Poisson distribution:

\[ p(x) = \text{POISSON}(x, \lambda, \text{FALSE}) \]

\[ F(x) = \text{POISSON}(x, \lambda, \text{TRUE}) \]

For some reason, you don’t get that option with the Negative Binomial distribution:

\[ p(n) = \text{NEGBINOMDIST}(n-x, x, p) \]

Yes, I know that \( n-x \) is the number of failures, and yes, I know that someone would have to be a real nitwit to do it this way, but that’s the way Uncle Bill’s nitwit programmers did it.

Excel doesn’t have a function for the Geometric distribution, but you can use the Negative Binomial distribution function and simply enter \( x = 1 \) for the number of successes.

Unfortunately, Excel does not automatically give you the mean or variance for any of the discrete distributions. You have to calculate them using the formulas presented here. Sorry about that!