Trials and successes

One event of importance is achieving **success** while trying to do something. In this case the basic unit is a **trial**, an activity (not event) that leads to success or to failure (either is an event). The distribution of successful trials can then be represented in two ways:

- as a random variable giving the number of trials leading to the first success;
- as a random variable giving the number of successes in a sequence of trials of a given length.

Both are **discrete variables**. The most common trials/successes model is called the **Bernoulli process** and is made of two distributions: **Geometric** and **Binomial**.

Geometric distribution

We perform some activity until we succeed (tossing coins, throwing darts, falling in love, etc.). Suppose we know the probability of success of each attempt; let it be p.

The random variable $\mathcal{G}$ is the set of possible numbers of attempts until the first success is achieved. Clearly, unless $p = 1$, the set $\mathcal{G}$ is infinite.

The probability of succeeding in the k^{th} attempt is equal to $k - 1$ failures (which happens with probability $(1 - p)^{k-1}$) followed by a success:

$$P(\mathcal{G} = k) = p(1 - p)^{k-1}$$

Note that this process is **memoryless**: having failed $n - 1$ times, we still have a probability p of succeeding in the n^{th} trial, no matter what the value of n is.

The mean and variance of the geometric distribution are:

$$\mu = \frac{1}{p}$$

$$\sigma^2 = \frac{1-p}{p^2}$$

This gives $\sigma = \frac{\sqrt{1-p}}{p}$.

When p is small, $\mu \approx \sigma$, a property of an interesting class of distributions.

You are an avid darts player. You established from past experience that you hit the bull's eye about 5 times per hundred attempts. Now your friend is offering you a wager: he will give you $20 if you score with one of the first 10 throws and you will give her $10 if you don't.

Should you accept?

The empirical estimate of p (probability of success in 1 trial) is 0.05 (5 out of 100). This gives the expected number of trials before the first success as 20.

At first, it looks unwise to accept the wager: for this random variable $\mu = 20$ and you are given only 10 chances.

But one should not forget that if you win, you get twice as much as pay when you lose. A naive person may even conclude that the wager is fair, because of the proportions profit/loss and μ / 10 are the same.

You lose if you fail to succeed 10 times in a row. The probability of failure being 0.95, failing 10 times occurs with probability $0.95^{10} = 0.59874$. Thus you win with probability **0.40126**.

Another way to calculate the same:

$$P(\mathcal{G} \leq 10) = \sum_{i=1}^{10} P(\mathcal{G} = i)$$

$$P(\mathcal{G} \leq 10) = \sum_{i=1}^{10} p(1-p)^{i-1} = p\sum_{i=1}^{10}(1-p)^{i-1}$$

$$\sum_{i=1}^{n}(1-p)^{i-1} = \frac{1-(1-p)^n}{1-(1-p)} = \frac{1-(1-p)^n}{p}$$

$$P(\mathcal{G} \leq 10) = 1-(1-p)^{10}$$

With $p = 0.05$ we get:

$$P(\mathcal{G} \leq 10) = 1 - 0.95^{10} = 0.40126$$

$$P(\mathcal{G} \leq 10) = 1 - 0.95^{10} = 0.40126$$

If the profit is \$20 and the loss is \$10, it is wise to accept the wager, because the average outcome is:

$$0.4 \times \$20 - 0.6 \times \$10 = +\$2$$

which implies an **average** profit of \$2 from this wager.

Note however that $\sigma = \frac{0.97}{0.05} = 19.5$ which, in this case, indicates a great element of risk.

Geometrically distributed variates

$$P(k) = p(1-p)^{k-1}$$

This is the probability that k trials will be needed to make an event that occurs with probability p happen. Note that the textbook defines it differently: they count the number of failures, not trials.

The following formula for a geometrically distributed variate G works (proof in the book; just add 1 for the successful trial):

$$G = \left\lceil \frac{\ln U}{\ln(1-p)} \right\rceil$$

Yes, both parts of the fraction are negative.

Binomial distribution

We perform some activity (tossing coins, throwing darts, falling in love, etc.) n times. Suppose we know the probability of success of each attempt; let it be p.

Assume that the random variable $\mathcal{X}$ gives the probability that k of the n trials were successful. Then $\mathcal{X}$ follows the **binomial distribution**:

$$p(\mathcal{X} = k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k}$$

For the binomial distribution with n trials:

$$\mu = np$$

$$\sigma^2 = np(1-p)$$

Note that μ does not have to be an integer.

Throwing dice

When a die is rolled, the probability of a 6 is, clearly, $p = \frac{1}{6}$. Thus, the average number of rolls needed to roll a 6 is $\mu = \frac{1}{p} = 6$.

Jim proposes a wager: if he rolls a 6 in no more than 4 rolls, you pay him \$10; otherwise he pays you \$10.

Should you accept?

The geometric distribution should be used.

$$P(x \leq 4) = 1 - \left(\frac{5}{6}\right)^4$$

$$P(x \leq 4) = 0.5177$$

How can it be? The mean is 6 and yet the majority of the probability is at 1–4. We must note that in the formula for the mean the probabilities are weighted:

$$\mu = \sum_{k=1}^{\infty} kp(1-p)^{k-1}$$

More dice throwing

The scenario is the same, but now Jim ponders on the effectiveness of a wager that he will roll one or more 6.

Jim considers a wager: he rolls 4 times and collects $8 for each 6 he rolls. He gives you $10 if no 6 s are rolled.

Should he offer the wager?

The binomial distribution should be used.

$$P_4(k) = \left(\frac{4!}{k!(k-4)!}\right)\frac{1}{6^k}\left(\frac{5}{6}\right)^{4-k}$$

$$P_4(0) = \left(\frac{5}{6}\right)^4 = 0.4823$$

$$P_4(1) = 4 \times \frac{1}{6} \times (\frac{5}{6})^3 = 0.386$$

$$P_4(2) = 6 \times \frac{1}{36} \times (\frac{5}{6})^2 = 0.116$$

$$P_4(3) = 4 \times \frac{1}{216} \times (\frac{5}{6}) = 0.016$$

$$P_4(4) = \frac{1}{1296} = 0.00077$$

The payoff is:

$$\$8 \times 0.386 + \$16 \times 0.116 + \$24 \times 0.016 + ... = \$5.333$$

which compares favourably with $P_4(0) \times \$10 = 4.823$.

This looks laborious and not quite worthy to be shown in class. But there is more to it...

The tedious formula evaluated was exactly the same as:

$$Payoff\ =\ D \times \sum_{k=1}^{n} kP_n(k)$$

where D is the unit payoff (\$8 in the calculations).

An observant student will notice that the sum resembles the formula for the mean:

$$\mu = \sum_{k=0}^{n} kP_n(k)$$

Hence, the formula is $Payoff = D \times \mu = D \times n \times p$. Here, $n = 4$ and $p = \frac{1}{6}$, so $Payoff\ =\ \frac{2D}{3}$.

To find out the minimum value of D to make Jim's new bet profitable for him, we solve the inequality:

$$\frac{2D}{3} > 0.4823 \times \$10$$

and we get: $D > \$7.234$.