The Binomial Probability Distribution

A binomial experiment is one that possesses the following properties:

1. The experiment consists of n repeated trials;

2. Each trial results in an outcome that may be classified as a success or a failure (hence the name, binomial);

3. The probability of a success, denoted by $p$, remains constant from trial to trial and repeated trials are independent.

The number of successes $X$ in $n$ trials of a binomial experiment is called a **binomial random variable**.

The probability distribution of the random variable $X$ is called a **binomial distribution**, and is given by the formula:

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

$n = \text{the number of trials}

x = 0, 1, 2, \ldots n

p = \text{the probability of success in a single trial}

q = \text{the probability of failure in a single trial}

(i.e. $q = 1 - p$)
Example: Hospital records show that of patients suffering from a certain disease, 75% die of it. What is the probability that of 6 randomly selected patients, 4 will recover?

This is a binomial distribution because there are only 2 outcomes (the patient dies, or does not).

Let $X = \text{number who recover}$. Here, $n = 6$ and $x = 4$. Let $p = 0.25$ (success - i.e. they live), $q = 0.75$ (failure, i.e. they die).

The probability that 4 will recover:

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

$$P(X) = \binom{6}{4} (0.25)^4 (0.75)^2 = 0.0329595$$

![Histogram of Probabilities of Death](image-url)

```r
rbinom(10,5,0.25)  # 2 1 1 1 1 1 3 4 0 2
```
The Poisson Probability Distribution

The Poisson Distribution was developed by the French mathematician Simeon Denis Poisson in 1837.

The Poisson random variable satisfies the following conditions:

1. The number of successes in two disjoint time intervals is independent.
2. The probability of a success during a small time interval is proportional to the entire length of the time interval.
3. Apart from disjoint time intervals, the Poisson random variable also applies to disjoint regions of space.

Applications

- the number of deaths by horse kicking in the Prussian army (first application)
- birth defects and genetic mutations
- rare diseases (like Leukemia, but not AIDS because it is infectious and so not independent) - especially in legal cases
- car accidents
- traffic flow and ideal gap distance
- number of typing errors on a page
- hairs found in McDonald's hamburgers
- spread of an endangered animal in Africa
- failure of a machine in one month
The probability distribution of a Poisson random variable $X$ representing the number of successes occurring in a given time interval or a specified region of space is given by the formula:

$$ P(X) = \frac{e^{-\mu} \mu^x}{x!} $$

$x = 0, 1, 2, 3...$

$\mu = \text{mean number of successes in the given time interval or region of space}$

**Example:** Twenty sheets of aluminum alloy were examined for surface flaws. The frequency of the number of sheets with a given number of flaws per sheet was as follows:

<table>
<thead>
<tr>
<th>Number of flaws</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

What is the probability of finding a sheet chosen at random which contains 3 or more surface flaws?
Probability = \( P(X \geq 3) \)

\[
= 1 - (P(x_0) + P(x_1) + P(x_2))
\]

\[
= 1 - \left( \frac{e^{-2.3} \cdot 2.3^0}{0!} + \frac{e^{-2.3} \cdot 2.3^1}{1!} + \frac{e^{-2.3} \cdot 2.3^2}{2!} \right)
\]

\[
= 0.40396
\]

We can see the predicted probabilities for each of "No flaws", "1 flaw", "2 flaws", etc. on this histogram.

[The histogram was obtained by graphing the following function for integer values of \( x \) only. Then the horizontal axis was modified appropriately.]

\[
\frac{e^{-2.3} \cdot 2.3^x}{x!}
\]

\[rpois(10,0.7)\quad # \quad 3 \quad 1 \quad 2 \quad 0 \quad 0 \quad 0 \quad 2 \quad 0 \quad 1 \quad 0 \quad 0\]
Normal Probability Distributions

The Normal Probability Distribution is very common in the field of statistics. Whenever you measure things like people's height, weight, salary, opinions or votes, the graph of the results is very often a normal curve.

A random variable $X$ whose distribution has the shape of a normal curve is called a **normal random variable**.

This random variable $X$ is said to be normally distributed with mean $\mu$ and standard deviation $\sigma$ if its probability distribution is given by

$$f(X) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
Properties of a Normal Distribution

1. The normal curve is symmetrical about the mean;
2. The mean is at the middle and divides the area into halves;
3. The total area under the curve is equal to 1;
4. It is completely determined by its mean and standard deviation

```r
rnorm(2, mean = 4, sd = 1)  # 5.062753 4.950963
```
**Example:** It was found that the mean length of 100 parts produced by a lathe was 20.05 mm with a standard deviation of 0.02 mm. Find the probability that a part selected at random would have a length between 20.03 mm and 20.08 mm.

\[ X = \text{length of part} \]

(a) 20.03 is 1 standard deviation below the mean;

\[
20.08 = \frac{20.08 - 20.05}{0.02} = 1.5 \text{ standard deviations above the mean}
\]

\[
P(20.03 < X < 20.08) = P(-1 < Z < 1.5)
\]

\[
= 0.3413 + 0.4332
\]

\[
= 0.7745
\]
Exponential Distributions

Some natural phenomena have a constant failure rate (or occurrence rate) property; for example, the arrival rate of cosmic ray alpha particles or Geiger counter tics. The exponential model works well for inter arrival times (while the Poisson distribution describes the total number of events in a given period). When these events trigger failures, the exponential life distribution model will naturally apply.

The general formula for the probability density function of the exponential distribution is

$$f(x) = \frac{1}{\beta} e^{-\frac{(x-\mu)}{\beta}} ; x \geq \mu, \beta > 0$$

where $\mu$ is the location parameter and $\beta$ is the scale parameter (the scale parameter is often referred to as $\lambda$ which equals $1/\beta$). The case where $\mu = 0$ and $\beta = 1$ is called the standard exponential distribution. The equation for the standard exponential distribution is

$$f(x) = e^x ; x \geq 0$$
Applications:

• The time until a radioactive particle decays, or the time between clicks of a geiger counter
• The time it takes before your next telephone call
• The time until default (on payment to company debt holders) in reduced form credit risk modeling
• Events occurring with a constant probability per unit length, such as the distance between mutations on a DNA strand, or between roadkills on a given road.
• Queuing theory, the service times of agents in a system (e.g. how long it takes for a bank teller etc. to serve a customer) are often modeled as exponentially distributed variables. (The inter-arrival of customers for instance in a system is typically modeled by the Poisson distribution in most management science textbooks.)
**Example:** Simulate air traffic on a small but busy airport with only one runway. In each unit of time one plane can land or one plane can take off, but not both. Planes arrive ready to land or to take off at random times, so at any given unit of time, the runway may be idle or a plane may be landing or taking off, and there may be several planes waiting either to land or take off. We therefore need two queues, called **landing** and **takeoff**, to hold these planes. It is better to keep a plane waiting on the ground than in the air, so a small airport allows a plane to take off only if there are no planes waiting to land. Hence, after receiving requests from new planes to land or take off, your simulation will first service the head of the queue of planes waiting to land, and only if the landing queue is empty will it allow a plane to take off. You will run the simulation through many units of time.

| rexp(2,0.1)     | # | 6.330843 | 8.554865 |
Weibull Distributions

The Weibull distribution is a continuous probability distribution. It is named after Waloddi Weibull, who described it in detail in 1951, although it was first identified by Fréchet (1927) and first applied by Rosin & Rammler (1933) to describe the size distribution of particles.

Applications:

- Survival analysis
- Reliability engineering and failure analysis
- Industrial engineering to represent manufacturing and delivery times
- Extreme value theory
- Weather forecasting to describe wind speed distributions, as the natural distribution often matches the Weibull shape
- Communications systems engineering
  - radar systems to model the dispersion of the received signals level produced by some types of clutters
  - modeling fading channels in wireless communications, as the Weibull fading model seems to exhibit good fit to experimental fading channel measurements
• General insurance to model the size of Reinsurance claims, and the cumulative development of Asbestosis losses
• In forecasting technological change (also known as the Sharif-Islam model)
• In hydrology the Weibull distribution is applied to extreme events such as annual maximum one-day rainfalls and river discharges. The picture below illustrates an example of fitting the Weibull distribution to ranked annually maximum one-day rainfalls showing also the 90% confidence belt based on the binomial distribution. The rainfall data are represented by plotting positions as part of the cumulative frequency analysis.
The probability density function of a Weibull random variable $x$ is:

$$f(x; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0, \\ 0 & x < 0, \end{cases}$$

where $k > 0$ is the **shape parameter** and $\lambda > 0$ is the **scale parameter** of the distribution.

```r
rweibull(2, shape = 1)  # 1.010639 1.367050
```