Continuous Probability Distribution

If a random variable is a continuous variable, its probability distribution is called a continuous probability distribution.

A continuous probability distribution differs from a discrete probability distribution in several ways:

- The probability that a continuous random variable will assume a particular value is zero.
- As a result, a continuous probability distribution cannot be expressed in tabular form.
- Instead, an equation or formula is used to describe a continuous probability distribution.

The equation used to describe a continuous probability distribution is called a probability density function (pdf). All probability density functions satisfy the following conditions:

- The random variable Y is a function of X; that is, \( y = f(x) \).
- The value of y is greater than or equal to zero for all values of x.
- The total area under the curve of the function is equal to one.

The charts below show two continuous probability distributions. The first chart shows a probability density function described by the equation \( y = 1 \) over the range of 0 to 1 and \( y = 0 \) elsewhere.

The next chart shows a probability density function described by the equation \( y = 1 - 0.5x \) over the range of 0 to 2 and \( y = 0 \) elsewhere. The area under the curve is equal to 1 for both charts.

The probability that a continuous random variable falls in the interval between a and b is equal to the area under the probability density function curve between a and b. For example, in the first chart above, the shaded area shows the probability that the random variable X will fall between 0.6 and 1.0. That probability is 0.40. And in the second chart, the shaded area shows the probability of falling between 1.0 and 2.0. That probability is 0.25.
Probability distributions are either continuous probability distributions or discrete probability distributions. A discrete distribution has a range of values that are countable. For example, the numbers on birthday cards have a possible range from 0 to 122 (122 is the age of Jeanne Calment the oldest person who ever lived). A continuous distribution has a range of values that are infinite, and therefore uncountable. For example, time is infinite: you could count from 0 seconds to a billion seconds...a trillion seconds...and so on, forever.

A discrete probability distribution is made up of discrete variables, while a continuous probability distribution is made up of continuous variables. The two types of distributions differ in several other ways.

Continuous probability distributions are expressed with a formula (a Probability Density Function) describing the shape of the distribution.

**Uniform Distribution**

The uniform distribution provides a model for selecting a point “at random” from an interval (a, b) on a real line. Suppose that a student has to catch a bus to reach his place of education. Suppose the public transport system is such that the bus promptly arrives at a bus stop at fifteen minutes interval. The student however arrives at a bus stop at a random time between bus arrivals. The waiting time, say X, can be modelled by continuous uniform distribution.

Definition: A continuous random variable X is said to have Uniform distribution on the interval (a, b) if its pdf is of the form

\[
f(x) = \frac{1}{b-a} \quad \text{for } a < x < b
\]

\[= 0; \quad \text{otherwise}
\]

We use the notation \(X \sim U(a, b)\).

You can write computer programs to generate random numbers. The random number generators are functions in the computer language and are based on programs that use uniform distribution defined on the interval \((0, 1)\).

A uniform distribution, also called a rectangular distribution, is a probability distribution that has constant probability.
This distribution is defined by two parameters, a and b:

- a is the minimum.
- b is the maximum.

The distribution is written as U(a,b).

The following graph shows the distribution with a=1 and b=3:

Like all probability distributions for continuous random variables, the area, under the graph, of a random variable is always equal to 1. In the above graph, the area is:

\[
A = l \times h = 2 \times 0.5 = 1.
\]

The expected value (i.e. the mean) of a uniform random variable X is:

\[
E(X) = \frac{1}{2} (a + b)
\]

“a” in the formula is the minimum value in the distribution, and “b” is the maximum value.

Mean = \( \frac{1}{2} \ a + b = \frac{1}{2} (1 + 3) = \frac{1}{2} (4) = 2 \)
The variance of a uniform random variable is:

\[ \text{Var}(x) = \frac{1}{12}(b-a)^2 \]

For the above image, the variance is \((1/12)(3 – 1)^2 = 1/12 * 4 = 1/3.\)

**Exponential Distribution**

Exponential distribution or negative exponential distribution represents a probability distribution to describe the time between events in a Poisson process. In Poisson process events occur continuously and independently at a constant average rate. Exponential distribution is a particular case of the gamma distribution.

**Probability density function**

Probability density function of Exponential distribution is given as:

**Formula**

\[
f(x; \lambda) = \begin{cases} 
\lambda e^{-\lambda x}, & \text{if } x \geq 0 \\
0, & \text{if } x < 0 
\end{cases}
\]

Where –

- \( \lambda = \) rate parameter.
- \( x = \) random variable.
Normal Distribution

A normal distribution is an arrangement of a data set in which most values cluster in the middle of the range and the rest taper off symmetrically toward either extreme. Height is one simple example of something that follows a normal distribution pattern: Most people are of average height; the numbers of people that are taller and shorter than average are fairly equal and a very small (and still roughly equivalent) number of people are either extremely tall or extremely short. Here’s an example of a normal distribution curve:

A graphical representation of a normal distribution is sometimes called a bell curve because of its flared shape. The precise shape can vary according to the distribution of the population but the peak is always in the middle and the curve is always symmetrical. In a normal distribution the mean, mode, and median are all the same.
Formula

\[ y = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma}} \]

Where –

- \( \mu = \) Mean
- \( \sigma = \) Standard Deviation
- \( \pi \approx 3.14159 \)
- \( e \approx 2.71828 \)
Example

Problem Statement:

A survey of daily travel time had these results (in minutes):

| 26 | 33 | 65 | 28 | 44 | 55 | 25 | 36 | 37 | 43 | 62 | 35 | 38 | 45 | 32 | 28 | 34 |

The Mean is 38.8 minutes, and the Standard Deviation is 11.4 minutes. Convert the values to z - scores and prepare the Normal Distribution Graph.

Solution:

The formula for z-score that we have been using:

$$z = \frac{x - \mu}{\sigma}$$

Where –

- $z$ = the "z-score" (Standard Score)
- $x$ = the value to be standardized
- $\mu$ = mean
- $\sigma$ = the standard deviation

To convert 26:

First subtract the mean: $26 - 38.8 = -12.8$,

Then divide by the Standard Deviation: $-12.8/11.4 = -1.12$

So 26 is -1.12 Standard Deviation from the Mean

Here are the first three conversions.

<table>
<thead>
<tr>
<th>Original Value</th>
<th>Calculation</th>
<th>Standard Score (z-score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>$(26 - 38.8)/11.4 = -1.12$</td>
<td>-1.12</td>
</tr>
<tr>
<td>33</td>
<td>$(33 - 38.8)/11.4 = -0.51$</td>
<td>-0.51</td>
</tr>
<tr>
<td>65</td>
<td>$(65 - 38.8)/11.4 = -2.30$</td>
<td>-2.30</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

And here they graphically represent:
Chi-Square Distribution

The chi-squared distribution (chi-square or $X^2$ - distribution) with degrees of freedom, $k$ is the distribution of a sum of the squares of $k$ independent standard normal random variables. It is one of the most widely used probability distributions in statistics. It is a special case of the gamma distribution.

![Chi-Square Distribution Graphs]

Chi-squared distribution is widely used by statisticians to compute the following:

- Estimation of Confidence interval for a population standard deviation of a normal distribution using a sample standard deviation.
- To check independence of two criteria of classification of multiple qualitative variables.
- To check the relationships between categorical variables.
- To study the sample variance where the underlying distribution is normal.
- To test deviations of differences between expected and observed frequencies.
- To conduct a The chi-square test (a goodness of fit test).
Probability density function

Probability density function of Chi-Square distribution is given as:

**Formula**

\[
  f(x; k) = \begin{cases} 
    \frac{k^{-\frac{x}{2}}e^{-\frac{x}{2}}}{2^{\frac{k}{2}} \Gamma\left(\frac{k}{2}\right)}, & \text{if } x > 0 \\
    0, & \text{if } x \leq 0 
  \end{cases}
\]

Where –

- \( \Gamma\left(\frac{k}{2}\right) \) = Gamma function having closed form values for integer parameter \( k \).
- \( x \) = random variable.
- \( k \) = integer parameter.

Cumulative distribution function

Cumulative distribution function of Chi-Square distribution is given as:

**Formula**

\[
  F(x; k) = \frac{\gamma\left(\frac{x}{2}, \frac{k}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \\
  = P\left(\frac{x}{2}, \frac{k}{2}\right)
\]

Where –

- \( \gamma(s, t) \) = lower incomplete gamma function.
- \( P(s, t) \) = regularized gamma function.
- \( x \) = random variable.
- \( k \) = integer parameter.
Student T Distribution

T-test is small sample test. It was developed by William Gosset in 1908. He published this test under the pen name of "Student". Therefore, it is known as Student's t-test. For applying t-test, the value of t-statistic is computed. For this, the following formula is used:

**Formula**

\[
    t = \frac{\text{Deviation from the population parameter}}{\text{Standard Error of the sample statistic}}
\]

Where –

- \( t \) = Test of Hypothesis.

**Test of Hypothesis about population**

**Formula**

\[
    t = \bar{X} - \frac{\mu}{s} \cdot \sqrt{n},
\]

\[
    \text{where } S = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}
\]

**Example**

**Problem Statement:**

An irregular sample of 9 qualities from an ordinary populace demonstrated a mean of 41.5 inches and the entirety of square of deviation from this mean equivalent to 72 inches. Show whether the supposition of mean of 44.5 inches in the populace is reasonable. (For \( v = 8, t_{0.05} = 2.776 \))

**Solution:**

\( \bar{x} = 45.5, \mu = 44.5, n = 9, \sum (X - \bar{X})^2 = 72 \)

Let us take the null hypothesis that the population mean is 44.5.

\( i.e. H_0 : \mu = 44.5 \) and \( H_1 : \mu \neq 44.5 \),

\[
    S = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}},
\]

\[
    = \sqrt{\frac{72}{9-1}} = \sqrt{\frac{72}{8}} = \sqrt{9} = 3
\]
Applying t-test:

$$|t| = \frac{\bar{X} - \mu}{s} \cdot \sqrt{n},$$

$$|t| = \frac{|41.5 - 44.5|}{3} \times \sqrt{9},$$

$$= 3$$

Degrees of freedom = $v = n - 1 = 9 - 1 = 8$. For $v = 8$, $t_{0.05}$ for two tailed test = 2.306. Since, the calculated value of $|t|$ > the table value of $t$, we reject the null hypothesis. We conclude that the population mean is not equal to 44.5.

Ref: (Link)

**F Distribution**

The F distribution (Snedecor’s F distribution or the Fisher-Snedecor distribution) represents continuous probability distribution which occurs frequently as null distribution of test statistics. It happens mostly during analysis of variance or F-test.
Probability density function

Probability density function of F distribution is given as:

**Formula**

\[
f(x; d_1, d_2) = \sqrt{\frac{(d_1 x)^{d_1} d_2^{d_2}}{(d_1 x + d_2)^{d_1 + d_2}}} x f\left(\frac{d_1}{2}, \frac{d_2}{2}\right)
\]

Where –

- \(d_1\) = positive parameter.
- \(d_2\) = positive parameter.
- \(x\) = random variable.

Cumulative distribution function

Cumulative distribution function of F distribution is given as:

**Formula**

\[
F(x; d_1, d_2) = I_{\frac{d_1 x}{d_1 x + d_2}} \left(\frac{d_1}{2}, \frac{d_2}{2}\right)
\]

Where –

- \(d_1\) = positive parameter.
- \(d_2\) = positive parameter.
- \(x\) = random variable.
- \(I\) = lower incomplete beta function.

Ref : (Link)