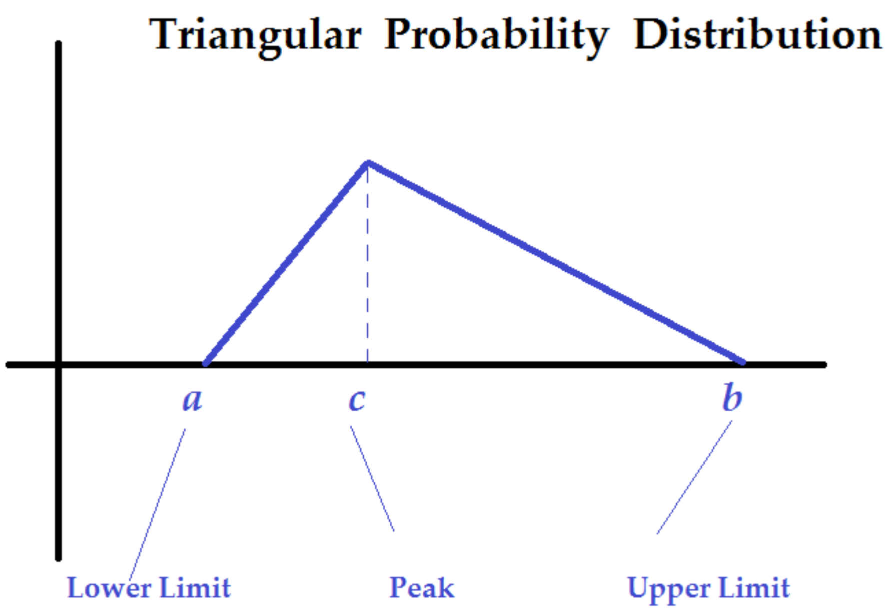
Triangular Probability Distribution

The triangular probability distribution (also called: "a lack of knowledge distribution") is a simplistic continuous model that is mainly used in situations when there is only limited sample data and information about a population. It is based on the knowledge of a minimum (a lower value), a maximum (an upper value), and a mode (peak) between those two values. For this reason, this distribution is very popular in simulation processes related to business decision models, project management models, financial models, and for modeling noises in digital audio and video data.



The probability density function ($\boldsymbol{pdf}$) of the triangular random variable $\boldsymbol{X}$ is given by:

$$\boldsymbol{f(x)} = \begin{cases} \frac{2}{(b-a)(c-a)}\,(x-a) & \text{if} \quad x \le c \\ \frac{2}{(b-a)(b-c)}\,(b-x) & \text{if} \quad x > c \end{cases} \qquad \textbf{(1)}$$

The following are some of the important numerical characteristics of the triangular distribution:

$$\boldsymbol{mean = E(x) = \mu = \frac{a+b+c}{3}} \quad \textbf{(2)}$$

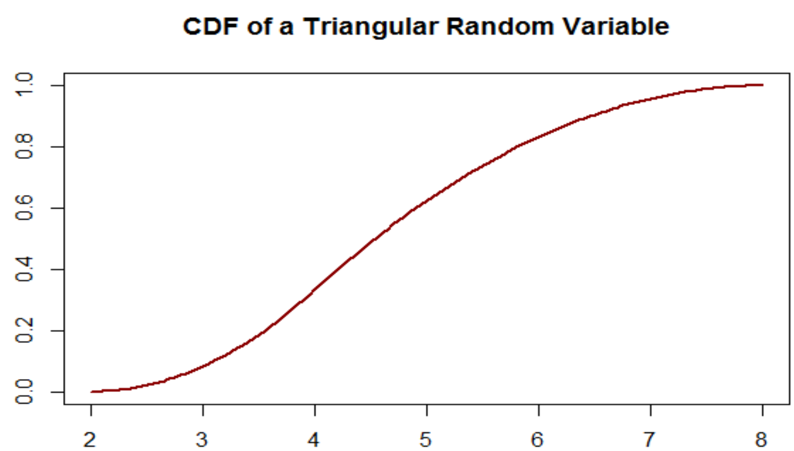
$$\boldsymbol{median = m = \begin{cases} a + \sqrt{0.5\,(b-a)(c-a)} & if \quad c < \frac{a+b}{2} \\ c & if \quad c = \frac{a+b}{2} \\ b - \sqrt{0.5\,(b-a)(b-c)} & if \quad c \geq \frac{a+b}{2} \end{cases}} \quad \textbf{(3)}$$

$$\boldsymbol{variance = \sigma^2 = \frac{a^2+b^2+c^2-ab-ac-bc}{18}} \quad \textbf{(4)}$$

The $\boldsymbol{cdf}$ (Cumulative density function) $\boldsymbol{P(X \leq x)}$ of the triangular random variable is:

$$\boldsymbol{F(x) = \begin{cases} \frac{1}{(b-a)(c-a)}\,(x-a)^2 & if \quad x < c \\ \frac{c-a}{b-a} & if \quad x = c \\ 1 - \frac{1}{(b-a)(b-c)}\,(b-x)^2 & if \quad x > c \end{cases}} \quad \textbf{(5)}$$

For example, the following is a display of the cumulative density function of a triangular random variable with a minimum value of 2, a maximum of 8, and with a peak at 4.



Random Number Generation of Triangular Random Variables:

The CDF expression given by formula (5) can be used to generate random values according to a specific triangular distribution. In this method, first a standard uniform random value r is created. This value is then used as a cumulative probability and replaces $F(x)$ in formula (5). The formula is then solved for the random variable x. The following rule describes this random number generation:

$$x = \begin{cases} a + \sqrt{r\,(b-a)(c-a)} & if\ r \le \frac{c-a}{b-a} \\ b - \sqrt{(1-r)(b-a)(b-c)} & if\ r > \frac{c-a}{b-a} \end{cases} \qquad \textbf{(6)}$$

Example:

In this example, we will simulate ten million triangular random values in R. We will then compare the numerical characteristics of this randomly generated set with the expected values.

1. Specify the specification of the triangular distribution:

```
> a<-2
> b<-8
> c<-4
```

2. Generate ten-million random values according to the standard uniform distribution:

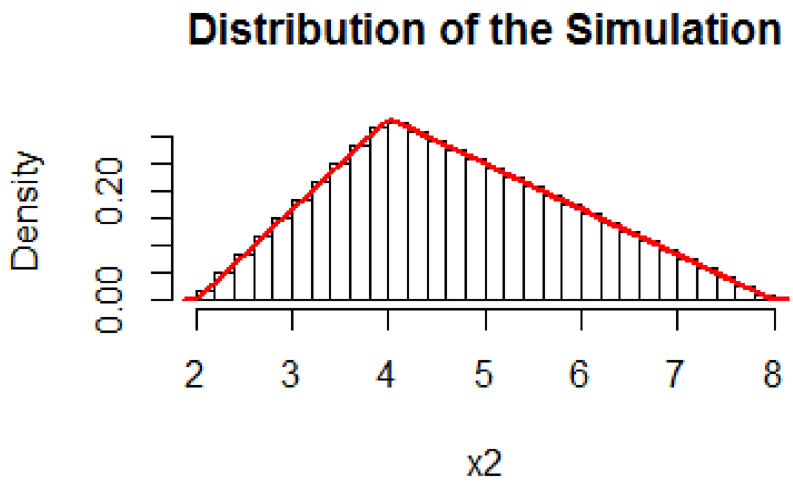
```
> N<-10^7
> r2<-runif(N)
```

3. Implement formula (6) to generate ten million triangular random values (labeled as x2):

```
> A<-a+sqrt((b-a)*(c-a)*r2)
> B<-b-sqrt((b-a)*(b-c)*(1-r2))
> C<-(c-a)/(b-a)
> x2<-ifelse(r2<C,A,B)
```

4. Create a relative frequency histogram along with the plot of the density function of the simulated values:

```
> hist(x2,freq=F,main="Distribution of the Simulation")
> lines(density(x2),lwd=2,col="red")
```



5. We will now compare the observed and the theoretical numerical characteristics of data:

```
> mean(x2)
[1] 4.666847


> (a+b+c)/3
[1] 4.666667


> median(x2)
[1] 4.536229

> ifelse(c<(a+b)/2,a+sqrt((b-a)*(c-a)/2), b-sqrt((b-a)*(b-c)/2))
[1] 4.44949

> sd(x2)
[1] 1.24714

> sqrt((a^2+b^2+c^2-a*b-a*c-b*c)/18)
[1] 1.247219
```

6. In fact, we can create a custom function in R that will allow for calculating the triangular cumulative probabilities:

```
ptriangular<-function(x,a,b,c)  {
  KK<-(1/((b-a)*(c-a)))*(x-a)^2
  PP<-1-(1/((b-a)*(b-c)))*(b-x)^2
  prob<-ifelse(x<c,KK,PP)
  return(prob)
}
```

For example, suppose that we wish to calculate $\boldsymbol{P(X > 5)}$. First observe that:

$$\boldsymbol{P(X > 5) = 1 - P(X \leq 5),}$$

And then use the "**ptriangular()**" function to calculate the above probability:

```
> 1-ptriangular(5,2,8,4)
[1] 0.375
```

7. We can also create a custom function in R that will handle the inverse problems; i.e., problems in which a cumulative probability (***P***) is used to calculate the corresponding value (***q***) of the triangular random variable:

```
qtriangular<-function(p,a,b,c) {
  QQ<-a+sqrt((b-a)*(c-a)*p)
  qq<-b-sqrt((b-a)*(b-c)*(1-p))
  q<-ifelse(p<(c-a)/(b-a),QQ,qq)
  return(q)
}
```

For example, the 95th percentile of the triangular distribution of our example can be determined as follows:

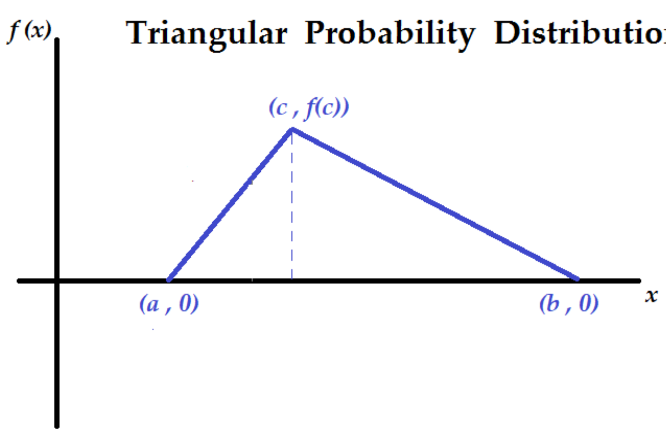
```
> qtriangular(0.95,2,8,4)
[1] 6.904555
```

Note that the "**qtriangular()**" function as defined above can also be used for the random number generation. The following code will generate 1000 triangular random values according to the triangular distribution of our example:

```
> x<-qtriangular(runif(1000),2,8,4)
```

Appendix

1. Derivation of the probability density function of the triangular probability distribution:



A primary condition for the pdf of the triangular distribution is that the total area under the curve should be equal to one unit:

$$\boldsymbol{Total\ Area} = \frac{\boldsymbol{base\ \times height}}{\mathbf{2}} = \frac{(\boldsymbol{b}-\boldsymbol{a})\boldsymbol{f}(\boldsymbol{c})}{\mathbf{2}} = \mathbf{1}$$

Therefore,

$$\boldsymbol{f}(\boldsymbol{c}) = \frac{\mathbf{2}}{\boldsymbol{b}-\boldsymbol{a}}$$

The pdf of the triangular distribution consists of two linear segments as described below:

(i) The line passing through the points $(a\,,0)$ and $(c\,,f(c))$:
The equation of this line is:

$$\boldsymbol{y} = \frac{\mathbf{2}}{(\boldsymbol{b}-\boldsymbol{a})(\boldsymbol{c}-\boldsymbol{a})}(\boldsymbol{x}-\boldsymbol{a})$$

(ii) The line passing through the points $(b\,,0)$ and $(c\,,f(c))$:
The equation of this line is:

$$y = \frac{2}{(b-a)(b-c)}(b-x)$$

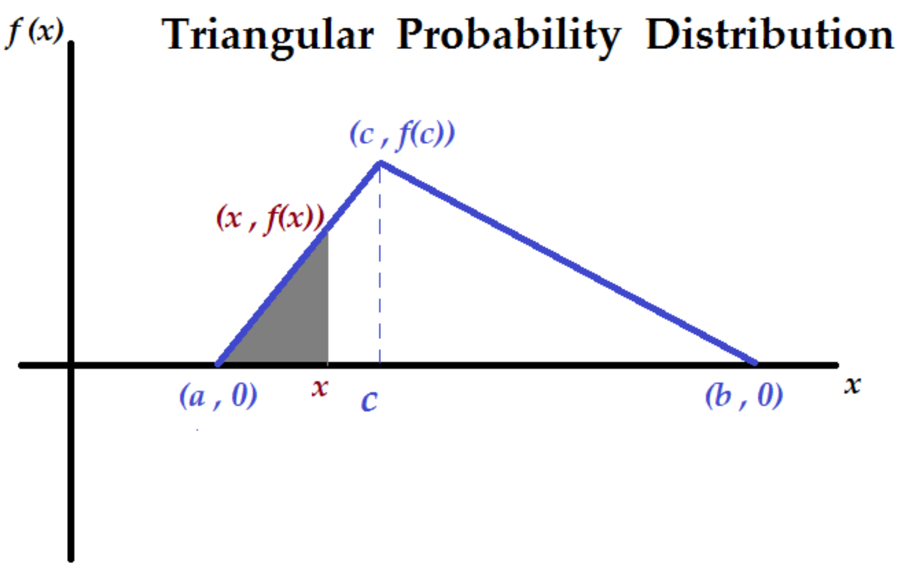
Therefore,

$$f(x) = \begin{cases} \frac{2}{(b-a)(c-a)}\,(x-a) & if \quad x \le c \\ \frac{2}{(b-a)(b-c)}\,(b-x) & if \quad x > c \end{cases}$$

2. Derivation of the cumulative probability function of the triangular probability distribution:

In general, given a value x, $a \le x \le b$, of the triangular random variable, there are three cases:

(i) $a \le x < c$



In this case, the cumulative probability function $F(x)$ is the area of the shaded region as shown above; which is a triangle with the length of its base being equal to $(x-a)$ and its height being equal to $f(x)$.

Therefore,

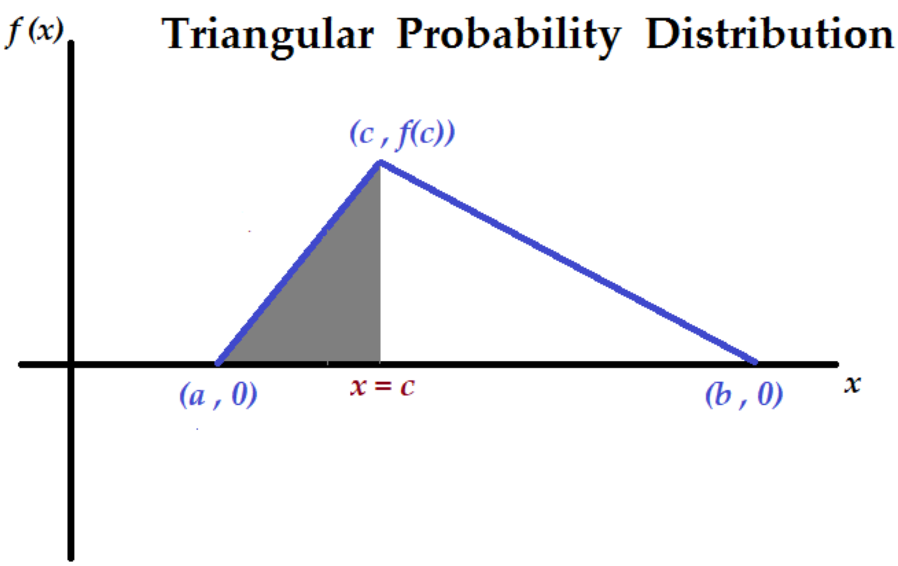
$$F(x) = Shaded\ Area = \frac{base\ \times height}{2} = \frac{(x-a)f(x)}{2}$$

Using the first expression for $f(x)$ and simplifying yields the following expression for $F(x)$

for the case when $a \leq x < c$:

$$F(x) = \frac{1}{(b-a)(c-a)}\ (x-a)^2$$

(ii) $x = c$



As in case (i), the cumulative probability function $F(x)$ is the area of the shaded region as shown above; which is a triangle with the length of its base being equal to $(c-a)$ and its height being equal to $f(c) = \frac{2}{b-a}$.

.

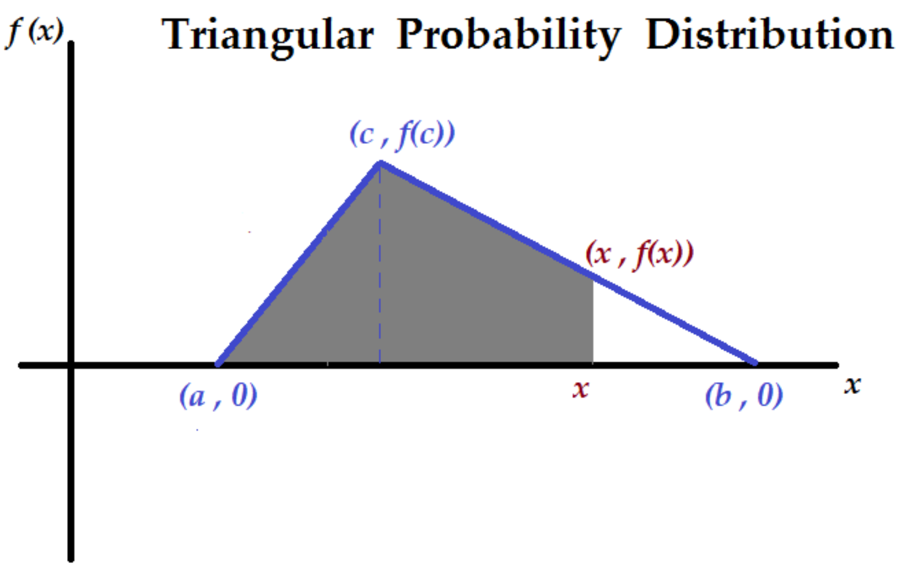
Therefore,

$$F(x) = Shaded\ Area = \frac{base\ \times height}{2} = \frac{(c-a)f(c)}{2}$$

The following expression is obtained for $\boldsymbol{F(x)}$ for the case when $\boldsymbol{x = c}$:

$$\boldsymbol{F(x) = \frac{(c-a)}{(b-a)}}$$

(iii) $\boldsymbol{c < x \leq b}$



In this case, the cumulative probability function is the shaded area of the polygon as shown above. To calculate this area geometrically, we may calculate the area of the triangle on the right tail (the triangle whose base is the segment from $\boldsymbol{x}$ to $\boldsymbol{b}$), and then subtract this area from the total area under the density curve. Therefore,

$$\boldsymbol{F(x) = 1 - \frac{(b-x)f(x)}{2}}$$

Plugging in for $\boldsymbol{f(x) = \frac{2}{(b-a)(b-c)}(b-x)}$ and simplifying, we obtain the following expression for $\boldsymbol{F(x)}$ for the case when $\boldsymbol{c < x \leq b}$:

$$\boldsymbol{F(x) = 1 - \frac{1}{(b-a)(b-c)}(b-x)^2}$$

In summary,

$$\boldsymbol{F(x)} = \begin{cases} \frac{1}{(b-a)(c-a)}\,(x-a)^2 & if \;\; x < c \\ \frac{c-a}{b-a} & if \;\; x = c \\ 1 - \frac{1}{(b-a)(b-c)}\,(b-x)^2 & if \;\; x > c \end{cases}$$

3. **Random number generation of a triangular random variable:**

As in any other random number generation, we may use the ***Inverse Transformation Method*** to generate a triangular random value. In this method, we first generate a standard uniform random value $\boldsymbol{r}$, and use this value as a cumulative probability $\boldsymbol{F(x)}$. We will then solve the resulting equation for the triangular random value $\boldsymbol{x}$. There may however be two different situations, depending on the value of $\boldsymbol{r}$:

(i) If $\boldsymbol{r} \le \frac{c-a}{b-a}$, then it results that the random value $\boldsymbol{x}$ should fall between the points $\boldsymbol{a}$ and $\boldsymbol{c}$. Therefore, we must solve the following equation for $\boldsymbol{x}$:

$$r = \frac{1}{(b-a)(c-a)}\,(x-a)^2$$

This equation is a quadratic equation, and only positive solutions are feasible:

$$(x-a)^2 = r(b-a)(c-a)$$

$$x - a = \sqrt{r(b-a)(c-a)}$$

$$x = a + \sqrt{r(b-a)(c-a)}$$

(ii) If $\boldsymbol{r} > \frac{c-a}{b-a}$, then it results that the random value $\boldsymbol{x}$ should fall between the points $\boldsymbol{c}$ and $\boldsymbol{b}$. Therefore, we must solve the following equation for $\boldsymbol{x}$:

$$r = 1 - \frac{1}{(b-a)(b-c)}\,(b-x)^2$$

This equation is a quadratic equation, and only positive solutions are feasible:

$$\frac{\mathbf{1}}{(\boldsymbol{b}-\boldsymbol{a})(\boldsymbol{b}-\boldsymbol{c})}(\boldsymbol{b}-\boldsymbol{x})^{2}=\mathbf{1}-\boldsymbol{r}$$

$$(\boldsymbol{b}-\boldsymbol{x})^{2}=(\mathbf{1}-\boldsymbol{r})(\boldsymbol{b}-\boldsymbol{a})(\boldsymbol{b}-\boldsymbol{c})$$

$$\boldsymbol{b}-\boldsymbol{x}=\sqrt{(\mathbf{1}-\boldsymbol{r})(\boldsymbol{b}-\boldsymbol{a})(\boldsymbol{b}-\boldsymbol{c})}$$

$$\boldsymbol{x}=\boldsymbol{b}-\sqrt{(\mathbf{1}-\boldsymbol{r})(\boldsymbol{b}-\boldsymbol{a})(\boldsymbol{b}-\boldsymbol{c})}$$

In summary, the following rule is used to generate a triangular random variable that takes values between $\boldsymbol{a}$ and $\boldsymbol{b}$, and has a peak at $\boldsymbol{c}$.

$$\boldsymbol{x}=\begin{cases}\boldsymbol{a}+\sqrt{\boldsymbol{r}\,(\boldsymbol{b}-\boldsymbol{a})(\boldsymbol{c}-\boldsymbol{a})} & \boldsymbol{if}\ \ \boldsymbol{r}\ \leq\dfrac{\boldsymbol{c}-\boldsymbol{a}}{\boldsymbol{b}-\boldsymbol{a}}\\ \boldsymbol{b}-\sqrt{(\mathbf{1}-\boldsymbol{r})(\boldsymbol{b}-\boldsymbol{a})(\boldsymbol{b}-\boldsymbol{c})} & \boldsymbol{if}\ \ \boldsymbol{r}>\dfrac{\boldsymbol{c}-\boldsymbol{a}}{\boldsymbol{b}-\boldsymbol{a}}\end{cases}$$