Modified Pert Simulation

Paulo Buchsbaum
June-9, 2012
(updated March-8, 2017)

Abstract

Distribution Pert is one of more useful functions for modeling and simulation in the companies. However, usually it is applied in a inadequate and mechanical way.

This article decodes Pert Distribution and related distributions (Modified Pert Distribution and Generalized Beta Distribution) stimulating a more intuitive understanding of this relation by analysis and graphical visualization.

Furthermore, it is suggested a method for specifying parameters in Modified Pert, more suitable for estimations inside the companies involved in simulation processes.

It's created a numerical method for determining the parameters of the Generalized Beta, in the proposed methodology.

Key words: Beta Distribution, Generalized Beta Distribution, Pert Distribution, Modified Beta Distribution, Simulation

I) Introduction

The Pert Distribution is one of the most important distributions for practical use in business, because it is widely used to generate random values within a range in the financial models and simulations in the area of processes and analysis in general.

Distribution Pert, as well as a variant called Modified Pert Distribution, is a particular case of Generalized Beta Distribution, encompassing a wide range of distributions with values within the defined range.

Initially this paper establish the relationship between the Modified Pert Distribution parameters and Generalized Beta Distribution parameters, in a complete and intuitive way, with a display of Pert families in graphical mode.

The article follows showing that the form normally used for simulation of the Pert Distribution is not well suited to real situations.

The article concludes by proposing an alternative way to specify the Modified Pert and develops a numerical method to proceed as suggested.

II) Beta Distribuition

The Generalized Beta Distribution, hereafter referred simply as Beta, is defined by four parameters, two of them (a and b, called *shape* parameters) define the make of the function, and other two define the range (Min and Max), within which there is possibility of having a value.

When Min=0 and Max=1, the Beta Distribution becomes the Standard Beta Distribution or just Beta Distribution and is suitable for modeling percentages, eg percentage of votes in an election.

It can be observed illustrations of some forms that Beta (http://en.wikipedia.org/wiki/Beta_distribution) takes in, as well as the kurtosis and CDF (cumulative distribution function) formulas, besides other informations.

This distribution has the property to display a frequency of values around the mode (most frequent value) that drops softly in both directions, in a configurable way, unlike the Triangular Distribution, which decreases sharply from mode to extremes values.

This feature makes Beta to be more realistic in practical applications than Triangular Distribution, because it is rare that a practical situation shows this kind of discontinuity.

The standard deviation (SD), mean and *shape* parameters in Beta are calculated from the following formulas:

Mean = min + (max - min) * (a/(a+b)) (1)
SD = Raiz(a/(a+b) * b/(a+b) * $(Max-Min)^2$ / (a+b+1)) (2)
Mode = Min + (a-1)/(a+b-2) * (Max - Min) (3)

The minimum and maximum parameters are intuitive since they represent the possible interval of occurrence of values.

To make sense a practical use in business, it is necessary that the *shape* parameters (a and b) are both above 1.

However, Beta has its practical use hampered because it is not intuitive directly estimate the *shape* parameters a and b.

III) Pert Distribution

The Pert Distribution is defined from the minimum (min), maximum (max) and mode. It is a subset from Beta where

Mean = (min + 4 * mode + max) / 6

The above formula which defines the mean, states exactly which Beta applies to the given situation. Without a extra definition as above it would not be possible to calculate the *shape* parameters a and b, because there are many different possibilities with the same mode and same extreme values (min and max).

The mean formula, where is a weighting where mode influences twice than the ends. Note that mean is different from mode. If the mode is closer to the minimum, the tail is longer to the maximum direction, bringing mean for the maximum side and vice versa.

This distribution is widely used to model project duration in PERT analysis, where its name originates.

In this model, the user specifies mode (most common value), minimum and maximum. From these data, the distribution is completely defined.

Pert can be used to estimate project duration, costs, margins, markups, turnover and, finally, many variables in the business world.

From these definitions we can calculate the *shape* parameters a and b, but let's leave this development to the next topic.

IV) Modified Pert Distribution

IV.1) Introduction

The statistician David Vose, owner of *Vose Software*, has proposed the Modified Pert Distribution, and he has deployed it in his ModelRisk software, a competitor to *Palisade*'s @Risk. It was also adopted by the Mathematica® from Wolfram. This distribution is more versatile for applications, because the mean is calculated in a more flexible way.

Mean = (min + λ * mode + max) / (λ +2) (4)

It is easy to see in this formula, the higher *lambda* (λ), the steeper the function in the mode neighbour (higher kurtosis), this feature becomes smaller the distance between mean and mode. it also makes the curve near the ends (minimum and maximum) less important and frequent.

Obviously Modified Pert becomes Pert when *lambda* (λ) is worth 4.

IV.2) ***Shape*** calculation basead on mean and extremes.

In order to estimate *lambda* (λ) as a function of a and b, we can consider Min = 0 and Max = 1, since this does not affect the shape of the distribution

Doing it in (1 in Mean = min + (max - min) * (a/(a + b))) and (3 - Mode = Min + (a-1)/(a+b-2) * (Max - Min)) we obtain:

Mean = a/(a +b) (5)
Mode = (a -1)/(a +b -2) (6)

Isolating lambda (λ) in (4 - Mean = (λ * mode + 1) / (λ +2)):

Mean * (λ +2) = (λ * Mode + 1)
(Mean - Mode) λ = 1 – 2* Mean
λ = [1 – 2 * Mean] / [Mean – Mode] (7)

Replacing the expressions to Mean (5) and Mode in (7):

λ = [1 – 2a/(a + b)] / [a/(a + b) – (a-1)/(a+b-2)]
λ = [(a + b)/(a +b) – 2a/(a +b)] /
[a(a +b-2) / [(a +b)(a +b-2)] – (a -1)(a +b) / [(a +b)(a +b-2)]]
λ = (b-a)/~~(a+b)~~ / [[a(a+b-2) - (a-1)(a+b)] / [~~(a+b)~~(a+b-2)]]
λ = (b-a) / [[a^2 + ab – 2a – a^2 - ab + a + b] / (a+b-2)]
λ = (b-a) / [[–2a + a + b] / (a+b-2)]
λ = (b-a) / [(b -a) / (a+b-2)]
λ = (a+b-2)

So:

a+b=λ (Lambda) + 2 (8)

The above result is amazing because it is an extreme simplification of two complex expressions.

It expresses that the Modified Pert Distribution encompasses the whole family of distributions, where the sum of *shapes* parameters are constant (*lambda* + 2). There is a clear similarity between the different curves that share the same *lambda*, as we will see at graphic in the next section.

Now let's compute a and b as a function of Min, Max, Mean e Mode.

From (1 - Mean = Min + (Max - Min) * (a/(a + b))): **(a +b)*

Mean (a + b) = Min (a + b) + (Max-Min) a
(Mean-Max) a = (Min - Mean) b
(Max-Mean) a = (Mean - Min) b
a = (Mean- Min) b / (Max-Mean) (9)

From (3 - Mode = Min + (a-1)/(a+b-2) * (Max - Min)): **(a +b-2)*

Mode (a+b-2) = Min (a+b-2) + (a-1) (Max-Min)
(Mode-Min-(Max-Min)) a = (Min-Mode) b + (-Max+Min-2*Min+2*Mode)
(Mode-Max) a = (Min-Mode) b - (Max - 2*Mode + Min) (10)

Replacing (9) in (10):

(Mode-Max) (Mean-Min) b / (Max-Mean) = (Min-Mode) b - (Max -2*Mode + Min)
(Mode-Max) (Mean-Min) b =
(Min-Mode) (Max-Mean) b - (Max-2*Mode +Min) (Max-Mean)
(~~Mode*Mean~~ + ~~Max*Min~~ –Mode*Min – Max*Mean ~~–Min*Max~~ ~~–Mode*Mean~~ +
Mean * Min + Max*Mode) b = - (Max-2*Mode +Min) (Max-Mean)
-Mode*Min – Max*Mean +Mean*Min + Max*Mode)b=-(Max-2*Mode +Min)(Max-Mean)
(Min-Max)(Mean-Mode) b = - (Max-2*Mode +Min)(Max-Mean)
b = (Max-2*Mode +Min) (Max-Mean)/[(Max-Min)(Mean-Mode)] (11)

Replacing (11) in (9) we obtain:

a = (Mean-Min) (Max-2*Mode +Min) ~~(Max-Mean)~~ / [(Max-Min) (Mean-Mode)]/~~(Max-Mean)~~
a = (Mean- Min) (Max-2*Mode +Min) / [(Max-Min) (Mean-Mode)] (12)

Adding (11) and (12):

a + b = (Max-2*Mode +Min)/ (Mean-Mode)

Calling a+b as SS (*shape* sum) and substituting in (11) and (12) we obtain:

a=(Mean-Min)/(Max-Min)*SS (13)
b=(Max-Mean)/(Max-Min)*SS (14)

It is as if a and b were the division of total into proportional parts to the ratio of the distance of the respective extremes to the mean of the distribution.

IV.3) Pert family visualization

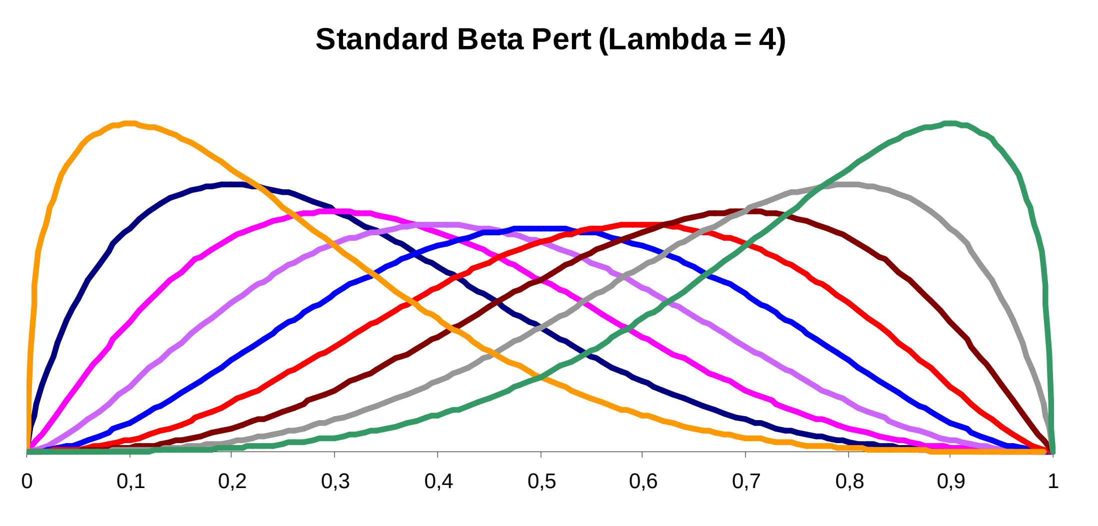
Below is shown a Pert family, mapped to Standard Beta with *lambda* = 4. It was plotted the value of the *Probability Distribution Function* (PDF) versus the values that can take the Beta variable.

As it is Standard Beta, the distribution can only take values between 0 and 1. The height of a point in PDF curve is proportional to the probability of distribution takes that value, more precisely, the immediate neighborhood of that value, since it makes no sense a probability of distribution taking a specific value.[1]

Mode is always the highest value of the curve, which represents the most probable value. In that case, mode can take any value between the two ends (0 and 1). By examinating the curves, it is clear that Beta degenerates, when mode nears the ends. This can only be imagined in the graph, because we only have used Mode values in the range between 0.1 and 0.9.

---

[1] It only makes sense talk in interval probability, ie probability that the distribution takes a value within a range which is given by the integral of the curve between two points from PDF. The integral of all PDF curve must always be 1, ie 100%, that covers all possibilities



Note the smoothness of distribution values near mode. This is a realistic behaviour of Beta, because natural phenomena do not have, in general, sharp changes. See also as at the least central curves, there is practically no values near the end in the longer tail.

When the mode is close to the middle between the two ends, the curve is symmetrical. Moving to the sides, there is a increasingly asymmetrical feature.

This versatility, different than symmetrical Normal distribution, is what becomes Beta Distribution as a so useful distribution to model many metrics from business world. It's a very common situation in which one needs assign a variable within a specified range, where the mode is closer to one of the two ends. For example, we can estimate an expected sales growth between 20% and 40%, being 25% the most likely growth (because a greater growth would depend on more uncertain actions).

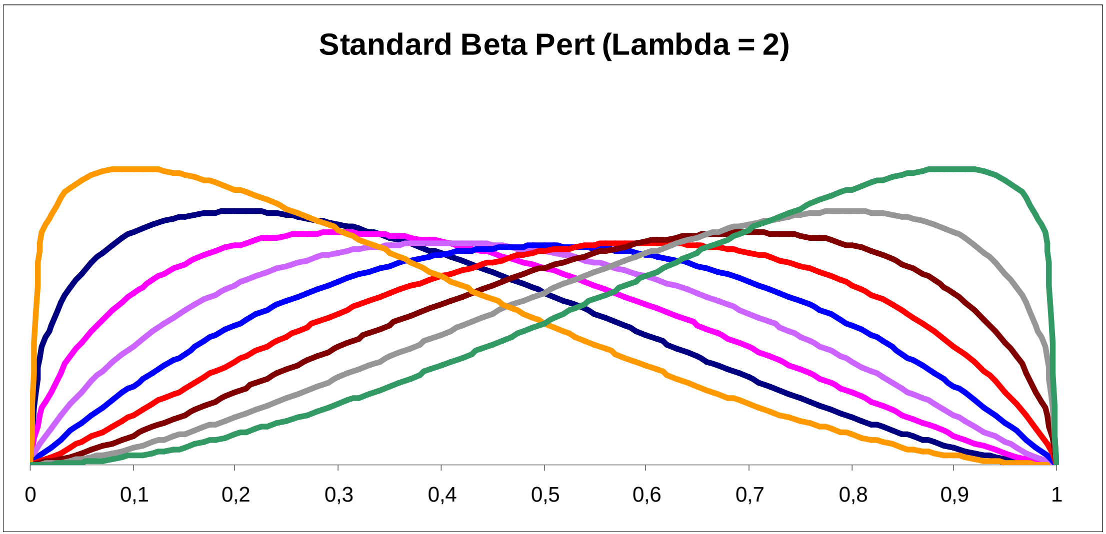
It is shown for *lambda* values above 4 that it is always possible to find a Normal Distribution fitting a symmetrical Beta (Mode = Mean). The approximation becomes accurate, the higher lambda value. However, the standard deviation is greater than the normal Beta and increasingly moves further Beta standard deviation.[2]

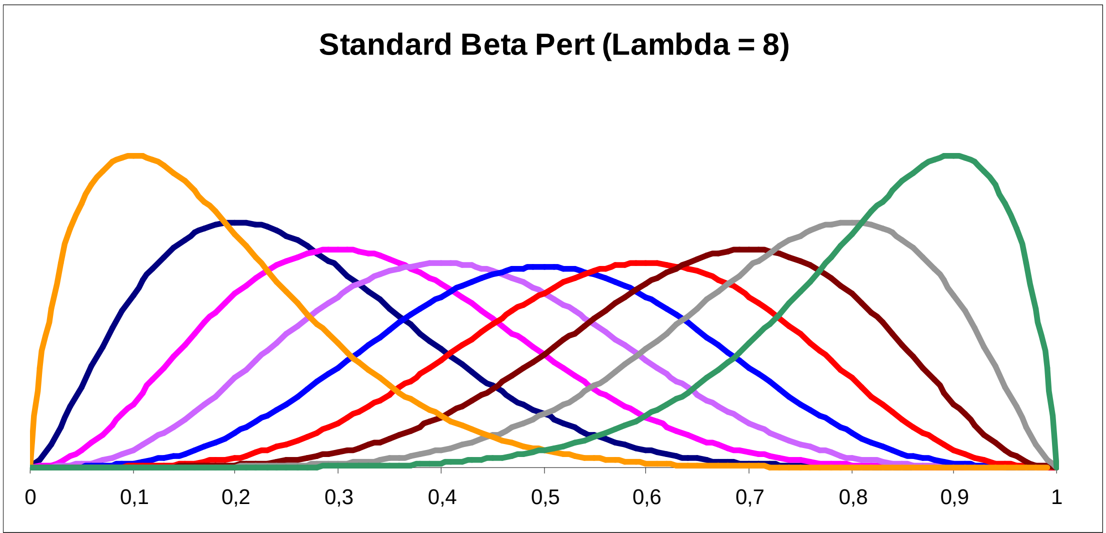
The mean, in turn, except when the mode is equidistant from the extremes, is always shifted toward the longer tail, which "attract" the mean.

Below we plot variations of two curve families. For *lambda* = 2, we have observed the flattening of the curves, resembling a horse saddle, with lower kurtosis. By other side, when *lambda* = 8, the curves are steeper, expressed by a higher kurtosis, making mode surroundings more abrupt. Note, in that case, as the middle valley is deeper than the curve with *lambda* = 4.

---

[2] The Normal Distribution is not a proper distribution to express variables that have a limited range of values. Some authors use Truncated Normal Distribution for this, but it creates a discontinuity in the ends that is not very realistic.





IV.4) <u>Shape calculation based on mode and extremes</u>

The formula (4) expresses the mean according to the mode and the extremes is repeated below.

Mean = (Min+Lambda*Mode+Max)/(Lambda+2)

Starting from the definition of the parameters a and b in (10) and (11) is

a=((Min+Lambda*Mode+Max)/(Lambda+2) -Min)/(Max-Min)*(Lambda+2)
a=((Min*(-1-Lambda) +Lambda*Mode+Max)/(Lambda+2)/(Max-Min)*(Lambda+2)
a=((Min*(-1-Lambda) +Lambda*Mode+Max)/(Max-Min) (15)
b=(Lambda+2) - ((Min*(-1-Lambda) +Lambda*Mode + Max)/(Max-Min)
b=(Lambda*Max+2*Max -Lambda*Min -2*Min +
Min*(1+Lambda) -Lambda* Mode- Max)/(Max-Min)
b=((Lambda+1)*Max - Min -Lambda* Mode)/(Max-Min) (16)

IV.5) <u>Standard deviation determination basead on mode and extremes</u>

The formula (2) expressing the SD based on *shapes*, minimum and maximum is repeated below.

SD = Raiz(a/(a+b) * b/(a+b) * (Max-Min)^2 / (a+b+1))

Substituting a*b in accordance with the values (10) and (11) is obtained:

a*b = (Max-Mean)*(Mean-Min)*((Max-2*Mode+Min)/((Max-Min)*(Mean-Mode)))^2 (17)

The piece (Mean - Max) * (Mean-Min) of the above formula can be expanded by using the average in the definition of (4):

(Max-Mean)*(Mean-Min) = ((Lambda+1)*Max - Lambda*Mode - Min) *(-(Lambda+1)*Min + Lambda*Mode + Max)/(Lambda+2)^2

The piece (Mean - Mode) of formula (17) can be expanded using the mean definition in (4):

Mean-Mode = (Min - 2*Mode + Max) / (Lambda + 2)

Substituting into above subexpressions (17) we have:

a*b = ((Lambda+1)*Max
- Lambda*Mode - Min)*(-(Lambda+1)*Min+Lambda*Mode+Max)/(Max-Min)^2

After we have placed the other components of the formula (2) to simplify:

a*b * (Max-Min)^2 / (a+b+1) =((Lambda+1)*Max - Lambda*Mode-Min)*
(-(Lambda+1)*Min + Lambda*Mode+Max)/(Lambda+3)
a*b * (Max-Min)^2 / (a+b+1) / (a+b)^2 = ((Lambda+1)*Max - Lambda*Mode - Min)*
(-(Lambda+1)*Min + Lambda*Mode + Max) / ((Lambda+3)*(Lambda+2)^2)

The left side has the formula corresponds to SD formula in (2), just adding square root and the right side is OK:

SD = Raiz (
((Lambda+1)*Max - Lambda*Mode - Min) *
(-(Lambda+1)*Min + Lambda*Mode + Max)/((Lambda+3)*(Lambda+2)^2)) (18)

IV.6) <u>Usual standard deviation formula based on extremes</u>

The standard deviation of Pert formula which is usually displayed is the following, even if some authors recognizing as an approximation:

DP = (Max - Min)/6 (19)

In Modified Pert the above formula is:

DP = (Max - Min)/(Lambda+2) (19a)

It is easily shown that this is a bad approximation:

Ron Davis (2008) uses the approximate formula in (19) to demonstrate a wrong relationship to the *shape* parameter calculation of Beta. Pleguezuele (1999) focuses a lot on above (19) approximation to deduce other properties, including until a cubic equation.

V) Suggested Specification for Modified Pert

V.1) Motivation

As noted above, the easiest way to define the Beta parameters is through Modified Pert, but this is not yet the ideal situation.

The problem is that the extreme values represent practically unattainable values because, in most of practical situations it shows that even the percentiles[3] near the ends (5º.). and (95º.) are still far from the theoretical extremes, especially on the side of longer tail.

Thus, a more viable proposition, it would produce a input specification that would include mode and two symmetrical percentiles, instead of mode and extremes

For example, you can specify that the user estimates the 5th percentile and 95th percentile, respectively as the point below where there is only 5% of cases and the point above it where there is only 5% of cases.

The reader can easily realize that it is much easier to train professionals to estimate unusual but possible ends, than estimate almost impossible values that would be the real extremes.

Thereafter, the system would calculate the ends, assuming that mean follows the formula (4).

The tool @Risk® from Palisade offers the possibility, through RiskPertAlt function, to input percentiles, but unfortunately this function does not allow a different *lambda* value than 4.

The example below, the cited @Risk® function, illustrates the point:

=RiskPertAlt(5%;10;"m. likely";15;95%;30)

In this case, the simulated variable has 5th percentile =10, 95th percentile = 30 and the most probable value is 15.

This distribution has the following parameters

a = 1,8
b = 4.2
Min = 7,9
Max = 43,5

Notice how the 95th percentile. (30) is distant from maximum. Even the 99th percentile. (34.5) is still far away. How do you expect that a professional, though trained in probability, can estimate the maximum if he virtually does not exist?

The idea, similar to the above example, is to specify *lambda*, P5, P95 and define the right Beta.

V.2) Beta parameters calculation

It's clearly useful to find some tool-independent method for estimating the Beta parameters from mode and percentiles.

The starting point is the l*ambda* selection, which tends to be more fixed in the model. The aimed *lambda* depends on, as we said above, the rate of decline in relation to the desired mode in both sides. The higher the value of lambda (and therefore from the sum of shape parameters a and b), the higher the mode, the steeper the fall, besides reducing the influence of the ends. *lambda* can define a family of functions that can be shared in all models.

---

[3] A Xo. percentile means that the stochastic variable is taked where X% of the theoretically possible values of this variable are below this value.

Anyway, the main problem is that these data do not allow input directly obtain the minimum and maximum of the Beta. The minimum is also known as *location* and the difference between the maximum and minimum is called *scale*.

Many p' statistical parameters from Beta Standard (minimum 0 and maximum 1) can be converted to the same p statistical parameter of Beta, for below equation:

p = Min + p' * (Max - Min)   (20)

The above statement also holds for mode, mean and percentiles.

Then one must find a relationship between the input data that is independent of *scale*. One proper ratio is expressed by the following:

Factor = (P95 - Mode) / (Mode - P5) (21)

One can easily check that this relation is independent of *location* and *scale*, that is, assumes the same value, whatever minimum and maximum values.

That Factor will be used to determine uniquely the relationship AShare involving the *shape* parameters

AShare = a /SS (22)

where

SS = a + b

We've chosen trying to determine AShare not a, since that property is more stable because it is not linked to the absolute value of the *shape* parameters.

From this point, there is already a *lambda* selected and therefore the value of the sum of *shape* parameters (SS) is lambda + 2 (see (12))

Next, it was built a data table in Excel® as a preparation for a curve fitting process, using Standard Beta. The original input value Mode is easier because we know that its value lies in the interval (0,1). For example, varying the mode between 0 and 1 with 0.05 step, excluding the extreme points, we've obtained nearly 200 points.

From mode, we compute a and b:

a = Mode*Lambda + 1
b = SS - a

Then It should be calculated the percentiles 5 and 90 through the inverse of Beta cumulative distribution function (CDF), in the case of Excel®, corresponds to the function BETAINV.

Finally, it generates the columns Factor (See (21)) and AShare (See (22))

These columns form respectively the independent variable (Factor) and a dependent variable (AShare) of some function that can be fitted to these points.

We've used to LABFIT® curve fitting tool (www.labfit.net), a brazilian commercial product that completely automates the curve fitting from hundreds of functions in its internal database. We've taken care to avoid deformation at the fitting extremes, by considering only the points where mininum(a, b) is greater than 1.2.

The tool have generated the *Harris* curve, function that has obtained the best fitting, according to the statistical error adopted:

Y = 1/(A+B*X**C)+D   (23)

The statistic parameter that the tool uses to sort the fitting functions is the *reduced chi-squared*, whose value was less than 1:1,000,000:

$$\chi_{Red}^2 = 1/(n-1) * \Sigma (O - E)^2 / \sigma^2$$

Where O is the observed data, E is the expected or theoric data, σ is the standard deviation of observed data and n is the number of observations.

For lambda = 6, we have A=1,6056, B=1,6056, C=1,3985 e D=0,18859

For lambda = 8, we have A=1,4257, B=1,4257, C=1,5843 e D=0,1493

Based on this parameter, the steps of determining the parameters of Beta are:

1) Calculate Factor of Beta specified by (21):

Factor = (P95 - Mode) / (Mode - P5)

2) Find AShare based on Factor in accordance with A, B, C and D parameters obtained for the aimed *lambda*. If *lambda* = 4, the equation (23) becomes:

AShare = 1/ (1,6056 + 1,6056 * Fator ^1,3985) + 0,18859

3) Calculate a and b:

a = AShare * SS
b = SS - b

4) Calculate Mode from Standard Beta Padrão:

Mode' = (a-1)/lambda

5) Calculate P5' from Standard Beta, using the inverse of the cumulative distribution function, which corresponds to BETAINV Excel® function

P5' = BETAINV(0,05;a;b)

6) Calculate the scale (Max-Min):

Scale = (P5 - Mode) /(P5' - Mode')

7) Calculate minimum:

Min = Mode - Mode' * Scale

8) Calculate maximum

Max = Min + Scale

The above process was used for *lambda* = 4 case, to compare with the tool @Risk®.

Mode = 15
P5=15
P95=30

The *chi-squared error* was reduced by less than 0.2%, except for the maximum (0.36%), what shows to be very applicable in practical situations, with the advantage that it can be calibrated for any *lambda* value.

VI) Conclusion

This article fulfills the mission of explaining Pert Distribution, justifying the logic of its generalization (Pert Distribution Modified) and also clarify its relationship with the Generalized Beta Distribution, highlighting the elegance of lambda (λ)parameter which added with 2 equals to the sum of Beta *shape* parameters.

The practical importance of the Beta Distribution or Pert Distribution is so great that professionals should use it with full awareness of what they are doing. This justifies the focus of this paper to give an intuitive and graphical view of Modified Pert, where each value of *lambda* defines a family of curves with similar shape.

In follow-up it has showed the weaknesses of conventional parametric Pert specification, demonstrating the difficulty of the user select minimum and maximum values, ending with a suggestion that replaces minimum and maximum with percentiles, maintaining the Mode specification.

The article concludes showing how to implement this new proposed specification using techniques of numerical analysis, that can be applied to any development environment.

We have achieved our goal if the reader gets a clearer view about Beta Pert family, reducing the dry and "black box" sensation, when one studies traditional statistics textbooks.

VII) Bibliography

DAVIS, Ron. "Teaching Project Simulation in Excel Using PERT-Beta Distributions", *Informs - Transactions on Education*, Vol. 8, No. 3, pp. 139–148, 2008

EVANS, Merran. *Statistical Distributions. 3ª.ed. New York* - John Wiley & Sons, 2000

JENSEN, Paul. "Beta Distribution", Operations Management/Industrial Engineering, 2004 http://www.me.utexas.edu/~jensen/ORMM/omie/computation/unit/project/beta.html

LAB Fit Curve Fitting Software (Nonlinear Regression and Treatment of Data Program) - 1999-2011

PALISADE. *Risk Analysis and Simulation Add-In for Microsoft® Excel*, 2010

PLEGUEZUELO, Rafael et alli. "The Parameters of the Classical Pert and Assessesment of its Success", *Central European Journal of Operations Research,* v.7, issue 3, pp 159-175, 1999

VOSE Software "Modified PERT distribution" - www.vosesoftware.com

WIKIPEDIA. "Beta Distribution" - http://en.wikipedia.org/wiki/Beta_distribution

WOLFRAM "Pert Distribution" http://reference.wolfram.com/mathematica/ref/PERTDistribution.html