# BetterEstimate: the (simple) math behind

Draft version 1.0 – October, 2006 (reviewed February, 2007)

Carlo Pescio, pescio@eptacom.net

## 1. Introduction

In [1], we introduced the BetterEstimate technique for project estimation. The technique borrows from traditional methods like the PERT technique of 3 estimates [2] and from more recent findings from Jorgensen [3] about increasing realism in expert estimates.
Under the BetterEstimate approach, experts will provide the following data for each task in a project:

- optimistic time
- probability of duration being smaller than the optimistic time
- pessimistic time
- probability of duration being larger than the pessimistic time

From those data, a probability distribution is determined; then, a most likely estimate is provided at the task level (straight from the probability distribution), together with a most likely estimate of effort at the project level (using Monte Carlo simulation). Prediction Intervals at project level can also be provided.

Note that unlike traditional methods, BetterEstimate does not ask the estimators for a most likely duration; instead, the most likely duration is calculated from the input data.

This paper presents the math behind the method. More exactly, a procedure to find a probability distribution from the data above is discussed, and an interesting insight on the relationship between the given times and probabilities is derived.

The math is relatively easy to follow, and requires only a basic understanding of probability theory.

## 2. Deriving a probability distribution

The first step in obtaining a probability distribution for a task is the selection of a distribution family. Although PERT adopted a *beta* distribution family, in this work we adopted the *triangular* distribution family, as it simplifies an already complex problem, yet provides good results.

We recall that the triangular probability distribution is a continuous distribution defined over the range [a..b] with probability density function:

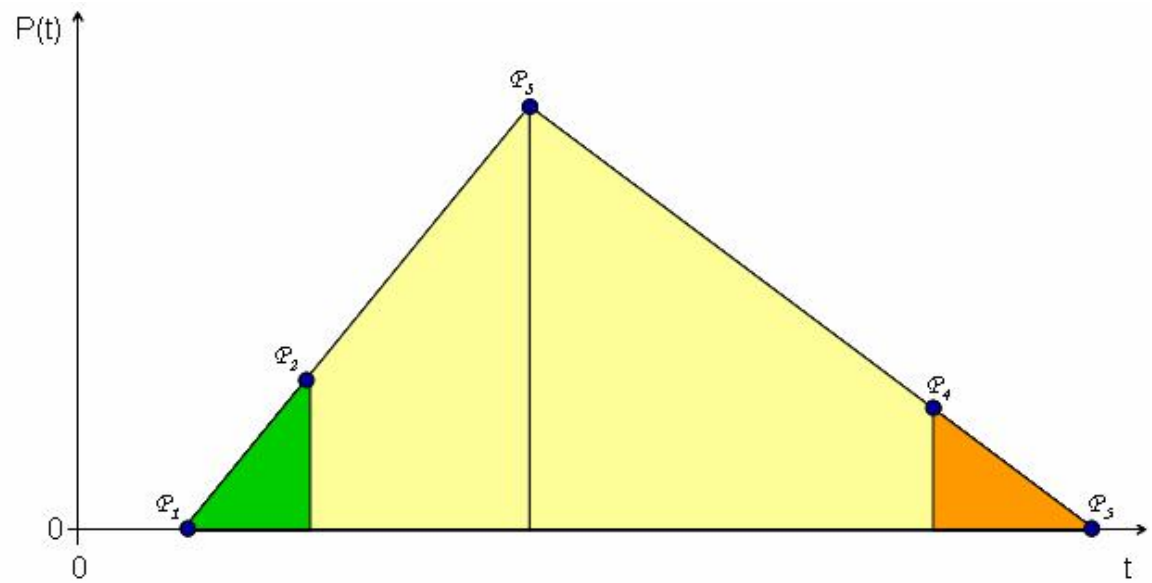
$$P(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & \text{for } a \le x \le c \\ \frac{2(b-x)}{(b-a)(b-c)} & \text{for } c \le x \le b \end{cases} \qquad (1)$$

and distribution function:

$$D(x) = \begin{cases} \frac{(x-a)^2}{(b-a)(c-a)} & \text{for } a \le x \le c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & \text{for } c \le x \le b \end{cases} \qquad (2)$$

Where c $\in$ [a..b] is the *mode*. The triangle in Fig.1 represents a triangular probability density:

**Fig. 1**




Assume each point *Pi* has coordinates ($t_i$, $y_i$), where $y_i = P(t_i)$. The data provided by the estimators in the BetterEstimate approach can be easily related to points and areas:

- The “optimistic” time is actually $t_2$ (in PERT, it would be $t_1$; this is a major difference between PERT and BetterEstimate).
- The “pessimistic” time is $t_4$.
- The probability of duration being smaller than the optimistic time ($P_o$) is the area of the green rectangle divided by the area of the outer rectangle ($P_1$ $P_3$ $P_5$).
- The probability of duration being higher than the pessimist time ($P_p$) is the area of the orange triangle divided by the area of the outer triangle.

Note that the area of the outer triangle, by the very definition of probability density, must be equal to 1. All the remaining data ($t_1$, $t_3$, $t_5$, $y_2$, $y_3$, $y_4$) are unknown, while $y_1$ and $y_5$ are obviously equal to 0.

Deriving a probability distribution boils down to finding $t_1$ and $t_5$ (the "actual" minimum and maximum duration), along with $t_3$ (the mode), so that we can calculate the expected duration, which for a triangular distribution is just the average of those 3 values: $(t_1 + t_3 + t_5) / 3$.

The problem may seem trivial, as many equations can be inferred from triangle similarity. For instance, $y_2 / y_3 = (t_2 - t_1) / (t_3 - t_1)$.
Unfortunately, this approach simply does not work. Indeed, the problem is under-constrained (we have more variables than equations), so some kind of heuristics will be needed to choose the "best" viable solution. Using triangle similarity leads to a quartic equation with complicated factors, which doesn't suggest any sensible heuristics.

**Note 1**:

Fig. 1 may suggest some inequalities that may not necessarily hold. More specifically, while by definition we can assume:

$t_1 <= t_2$
$t_4 <= t_5$
$t_2 <= t_4$

there is complete freedom on the value of $t_3$. While it is common to have

$t_2 <= t_3 <= t_4$

this is by no means a constraint on the triangular distribution, and we may also have

$t_3 <= t_2 <= t_4$

or

$t_2 <= t_4 <= t_3$

Both cases correspond to particular shapes of the triangle, with *P3* heavily shifted to the left or the right side.

Also, note that the green and orange triangle must not intersect; therefore:

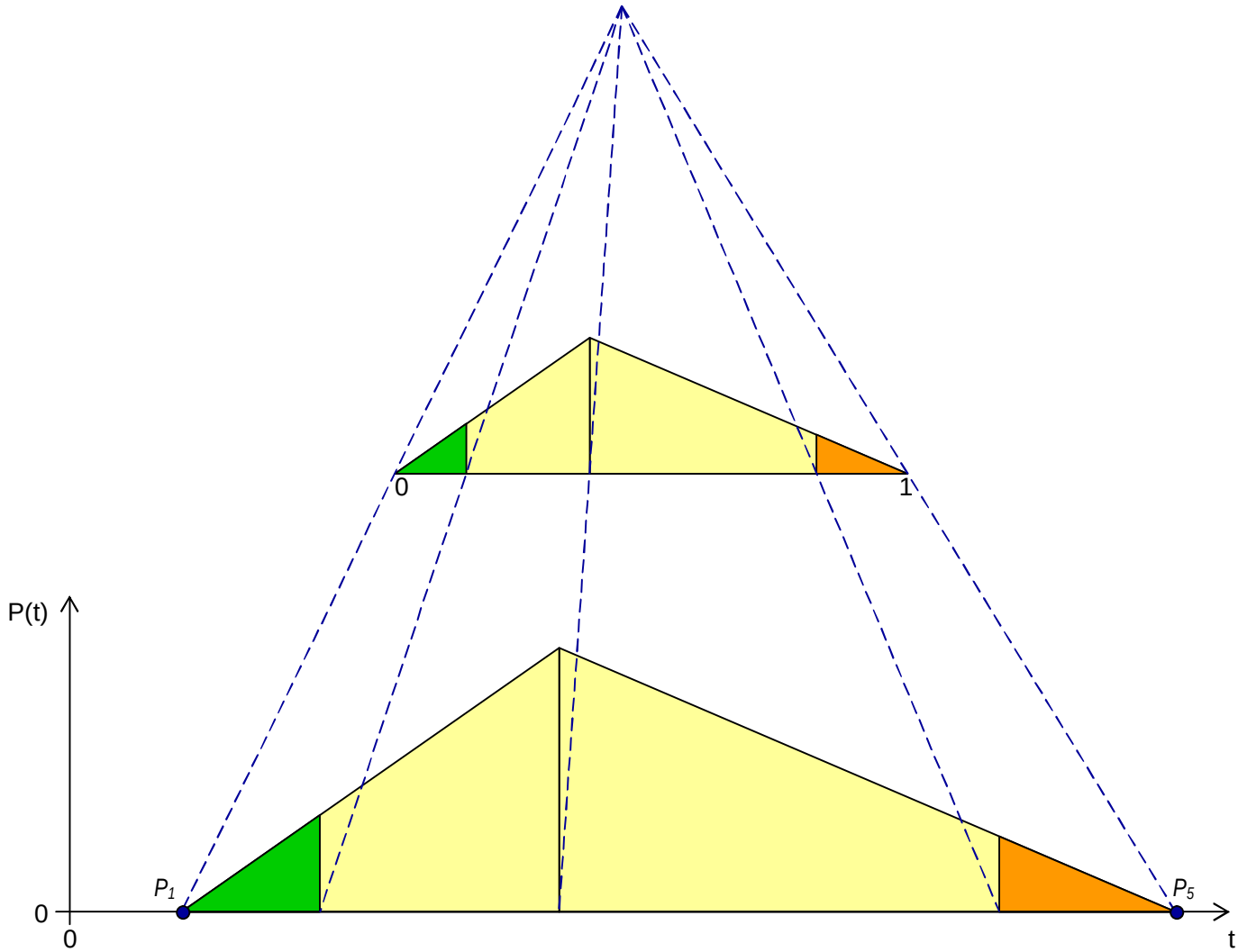
$P_o + P_p <= 1$

In most practical cases, a strict inequality will hold.

### 3. A “creative” approach

A different approach is to solve an easier problem instead, where some sensible heuristics can be applied, and then scale the results back to the original problem (see Fig. 2).

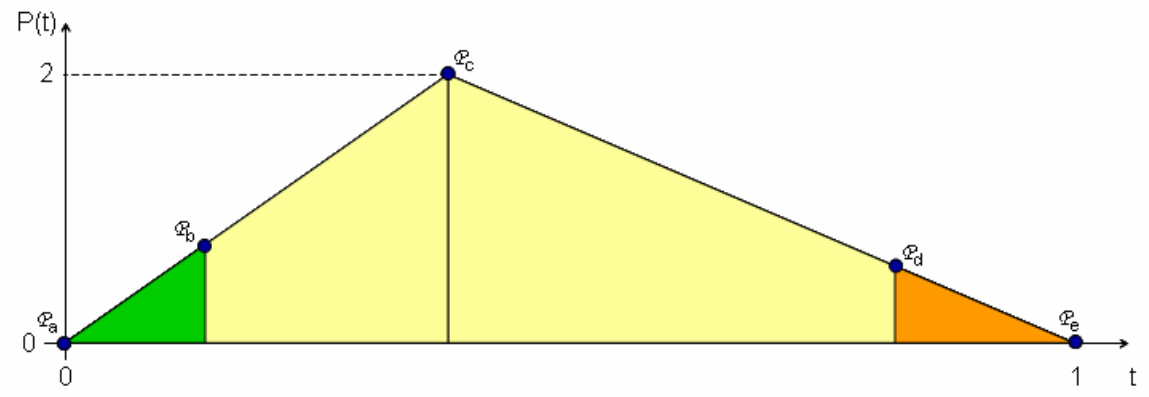
**Fig. 2**




The original triangle in Fig 1 can be considered as the linear scaling of a similar triangle defined over [0..1]. Of course, some data that was previously known (e.g. $t_2$) would be now unknown, while data previously unknown (e.g. $t_1$) would now be known (having been fixed to 0).

Linear scaling equations can be derived easily. Figure 3 is a zoom into the background triangle.

**Fig. 3**




From Fig. 1, Fig. 2 and Fig. 3, and considering that a linear scaling must (by definition) be defined by a linear relationship F( x ) = a x + b, we have:

$F(0) = b = t_1$
$F(t_b) = t_2$
$F(t_c) = t_3$ (3)
$F(t_d) = t_4$
$F(1) = a + b = t_5$

Note that the linear scaling must respect the constraint $t_1 >= 0$, therefore $b >= 0$.

We recall that $t_2$ and $t_4$ are known. If we consider $t_b$ and $t_d$ known (we'll derive them in the next paragraph), we can easily derive factors a and b:

$t_2 = a\ t_b + b$
$t_4 = a\ t_d + b$

Hence:

$a = (t_4 - t_2) / (t_d - t_b)$
$b = t_2 - a\ t_b =$
$t_2 - (t_4 - t_2) / (t_d - t_b)\ t_b =$ (4)
$(t_2\ t_d - t_2\ t_b - t_4\ t_b + t_2\ t_b) / (t_d - t_b) =$
$(t_2\ t_d - t_4\ t_b) / (t_d - t_b)$

Now, $b >= 0$ iff $t_2\ t_d - t_4\ t_b >= 0$ (as $t_d - t_b$ is always > 0 by definition). Therefore, the following inequality must hold:

$t_2\ t_d >= t_4\ t_b$

Since both $t_b$ and $t_2$ are > 0, this can be rewritten as

$t_d / t_b >= t_4 / t_2$

or, as we'll find more convenient later

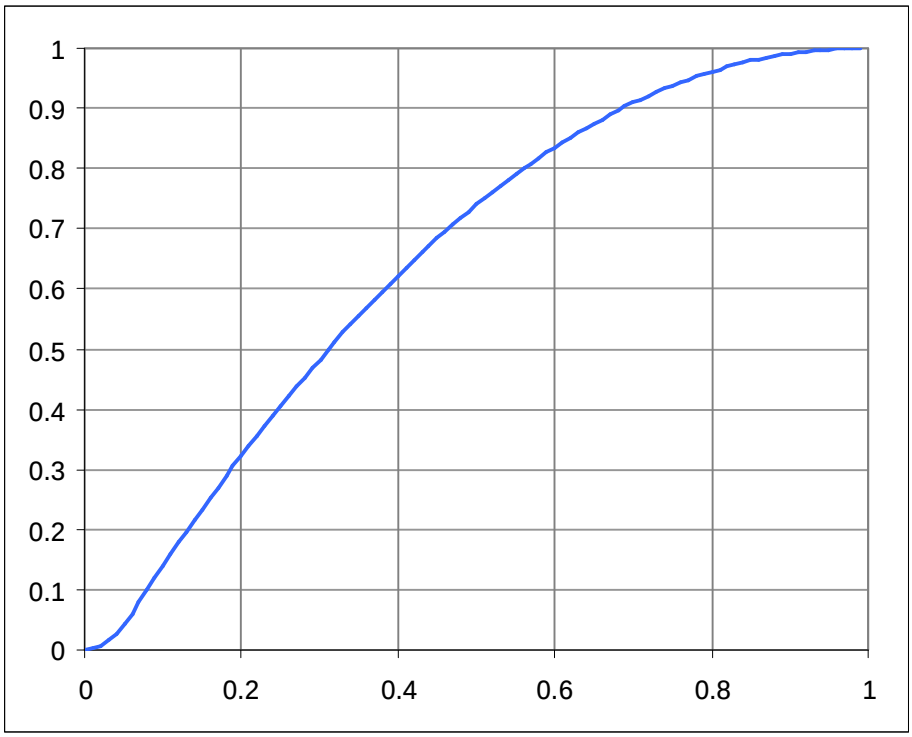
$t_4 / t_2 <= t_d / t_b$ (5)

This will lead to an interesting constraint on the probabilities $P_o$ and $P_p$ defined in paragraph 2.

## 4. Solving the problem in [0..1]

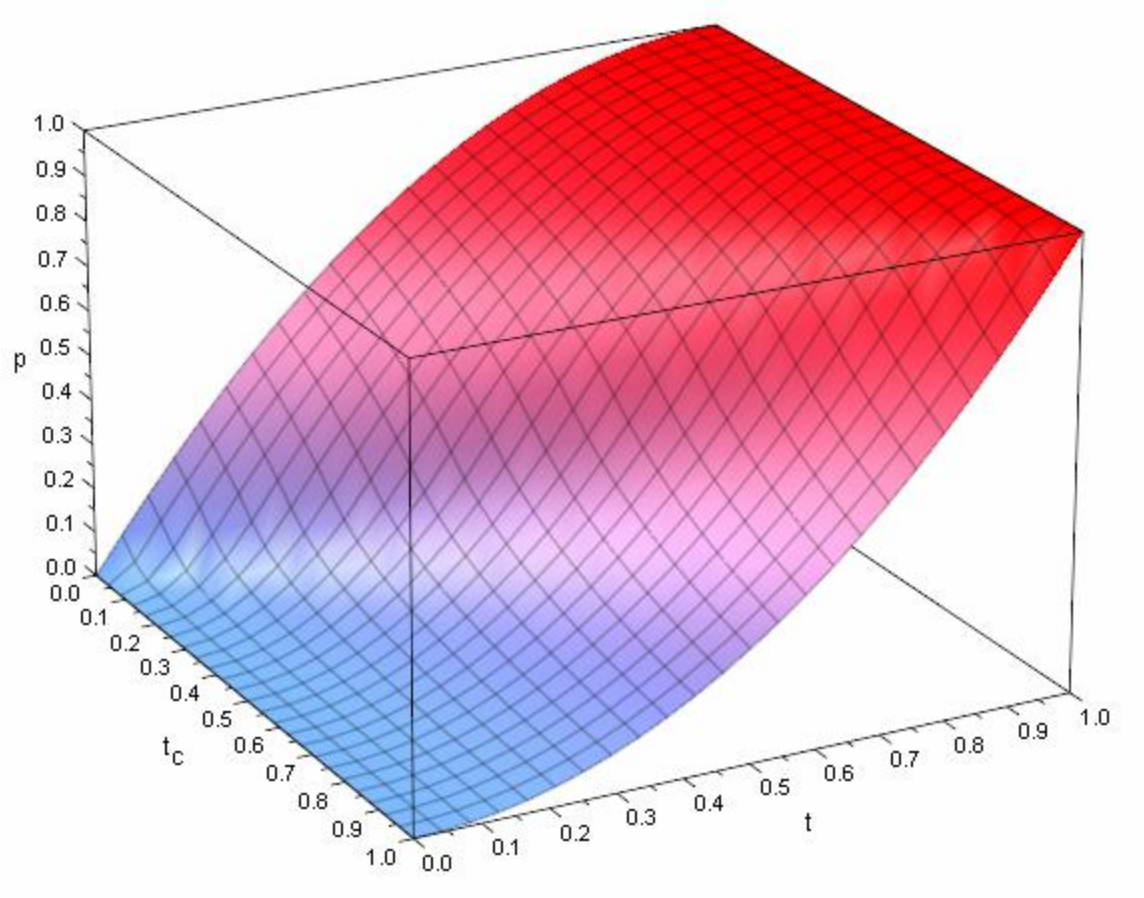
The exact meaning of “solving the problem for the triangle in [0..1]” is not necessarily obvious, and requires some explanation. Consider again the triangular probability density in Fig. 3, and the associated probability distribution in Fig. 4:

**Fig. 4**




Having fixed the extremities ($t_a$ and $t_e$) respectively to 0 and 1, the shape of the distribution is now entirely dependent on $t_c$. As we move $t_c$ from 0 to 1, we obtain a family of probability distributions, like in Fig. 5:

**Fig. 5**




(Note that once you fix a $t_c$, you get a single probability distribution as expected)

Now, in order for this triangle to be a scaled version of the original triangle in Fig 1, some constraints must hold:

$P_o = D( t_b )$
$P_p = 1 – D( t_d )$

where D is the probability distribution. Combining these constraints with equation (2), where a = 0, b = 1, $0 <= t_b <= t_d <=1$, $0 <= t_c <=1$, we obtain:

$$t_b = \begin{cases} \sqrt{P_o \cdot t_c} & P_o <= t_c \\ 1 - \sqrt{(1-P_o)\cdot(1-t_c)} & P_o > t_c \end{cases} \qquad (6)$$

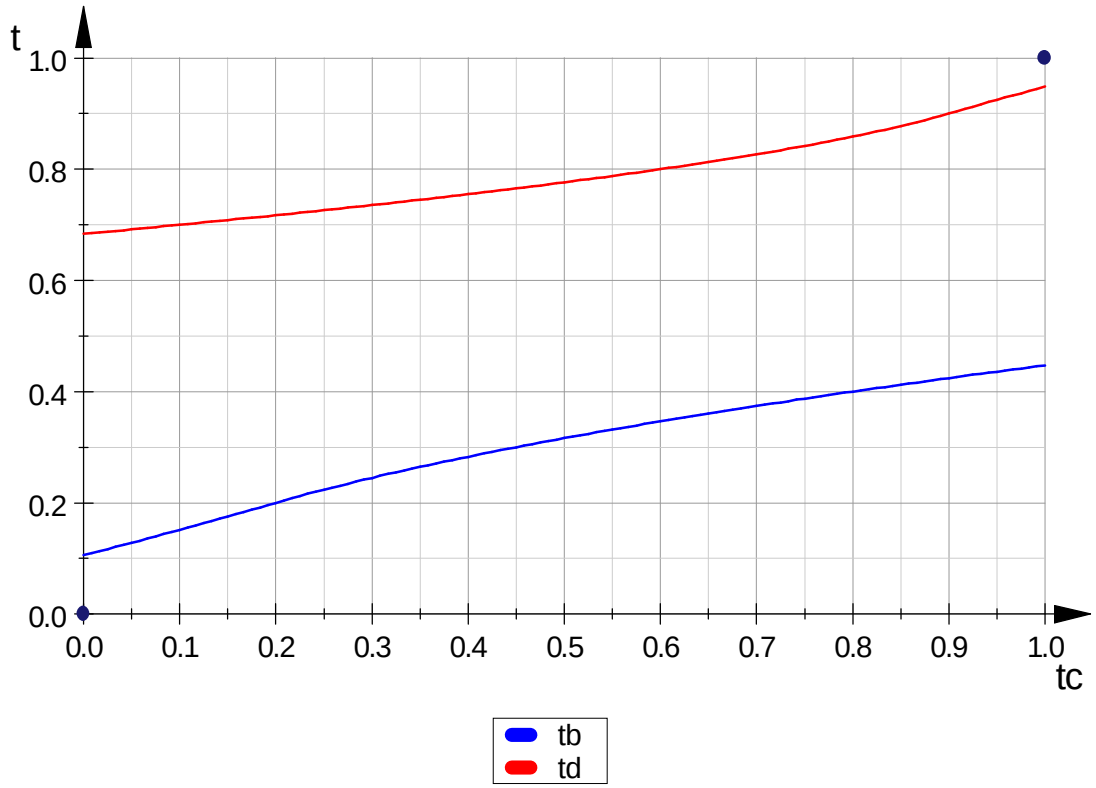
$$t_d = \begin{cases} \sqrt{t_c \cdot (1-P_p)} & P_p > 1- t_c \\ 1 - \sqrt{P_p \cdot (1-t_c)} & P_p <= 1- t_c \end{cases}$$

So, assuming for instance

$P_o = 0.2$
$P_p = 0.1$

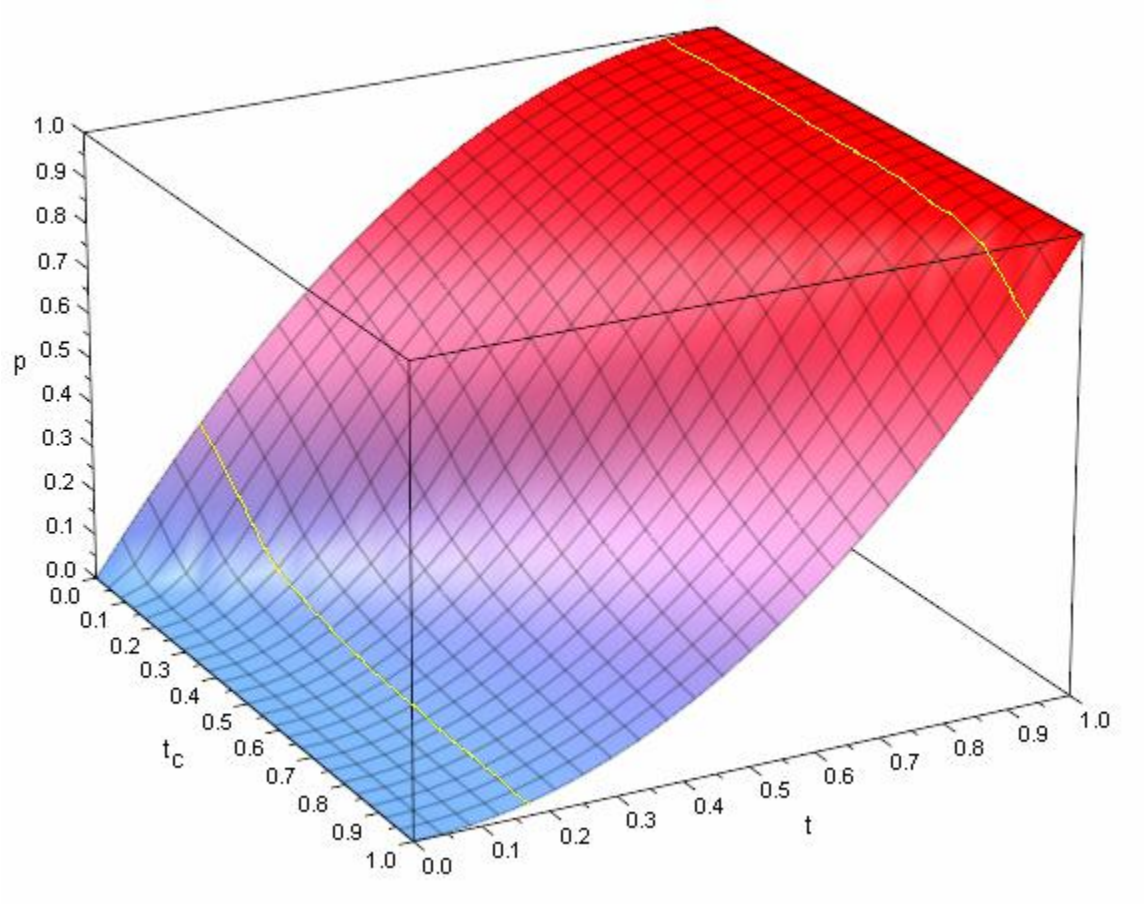
As $t_c$ moves in [0..1], $t_b$ and $t_d$ will move along the curves in Fig 6, as defined in (6):

**Fig. 6**




Note that these two curves can be seen also in Fig. 5, as highlighted in yellow in Fig. 7 for the chosen values of $P_o$ and $P_p$:

**Fig. 7**




Now, if we could define a criteria to choose the "best" $t_c$ (remember that $P_o$ and $P_p$ are known) we'll just have to put it back in (6), obtain $t_b$ and $t_d$, derive the linear scaling factors a and b as per equation (4), and then apply the linear scaling as per equation (3) to obtain the "best" probability distribution fitting the estimation data.

## 5. Some heuristics and the constraint on linear scaling

It is widely known that software developers routinely underestimate effort (see [1], [3] and their bibliography for more). Therefore, a sensible heuristic could be to compensate for underestimation.
It is also known (see [3] again) that estimators tend to provide very narrow ranges even in front of large uncertainty. Again, a sensible heuristic could be to compensate for a narrow $[t_2, t_4]$ provided by the expert.

When we consider the linear scaling, we have that $t_b$ must scale to $t_2$, and $t_d$ must scale to $t_4$.
Since $t_2$ and $t_4$ are fixed (provided by the estimator), we get the widest range for $[t_1, t_5]$ when $t_b$ and $t_d$ are closest to each other (therefore using the smallest portion of the [0..1] interval: this way, [0..1] will scale to the largest interval).

Therefore, a simple yet effective heuristics is to take $t_c$ in [0..1] such that $t_d - t_b$ reaches its minimum. Given the equations (6), this would lead to a relatively simple piecewise nonlinear minimum problem:

$$t_c = \min_{t \in [0..1]} \begin{cases} \sqrt{t \cdot (1 - P_p)} - \sqrt{P_o \cdot t} & P_o <= t \text{ AND } P_p >= 1- t \\ 1 - \sqrt{P_p \cdot (1 - t)} - \sqrt{P_o \cdot t} & P_o <= t \text{ AND } P_p <= 1- t \\ 1 - \sqrt{P_p \cdot (1 - t)} - \left(1 - \sqrt{(1 - P_o) \cdot (1 - t)}\right) & P_o > t \text{ AND } P_p <= 1- t \end{cases} \quad (7)$$

Note that a fourth condition, namely $P_o > t$ AND $P_p > 1- t$ is impossible because, by definition, $P_o + P_p <= 1$ (see paragraph 2, Note 1).

However, we must still respect the constraint $t_1 >= 0$. We recall from (5) that this can also be written as $t_4 / t_2 <= t_d / t_b$. Therefore, (7) must be rewritten by adding the following constraint:

$$\text{s.t. } \frac{t_4}{t_2} <= \frac{t_d}{t_b} \quad (8)$$

Solving (7) under constraint (8) is a relatively simple numerical problem, and we won't investigate it any further here (we just recall that $t_d$ and $t_b$ are defined in (6) as a function of $t_c$).

Also, once the probability distribution of each single task has been found, combining them using a Monte Carlo simulation is a straightforward and well-known process, which does not require further explanation.

## 6. Solution existence

The constraint (8) is actually a solution existence constraint. Given some input data, there might be no way to satisfy that constraint, that is, no solution in [0..1] that can properly scale back to the input data.

Unfortunately, (8) is not directly expressed in terms of BetterEstimate's input data. However, since (6) defines $t_b$ and $t_d$ in terms of $t_c$, $P_o$ and $P_p$, (8) can also be expressed in terms of $P_o$ and $P_p$. This is particularly interesting, as all the involved factors ($t_2$, $t_4$, $P_o$, $P_p$) would then be the input values of the BetterEstimate method.

It can be proven (see Appendix A) that $\frac{t_4}{t_2} <= \frac{t_d}{t_b}$ is equivalent to:

$$\frac{t_4}{t_2} <= \frac{1-\sqrt{Pp}}{1-\sqrt{1-Po}} \qquad (9)$$

Therefore, given the input data for BetterEstimate, that is, a pair ($t_2$, $t_4$) and a pair ($P_o$, $P_p$), we can solve the problem (that is, there is a triangular distribution fitting the data) iff

$$\frac{t_4}{t_2} <= \frac{1-\sqrt{Pp}}{1-\sqrt{1-Po}}$$

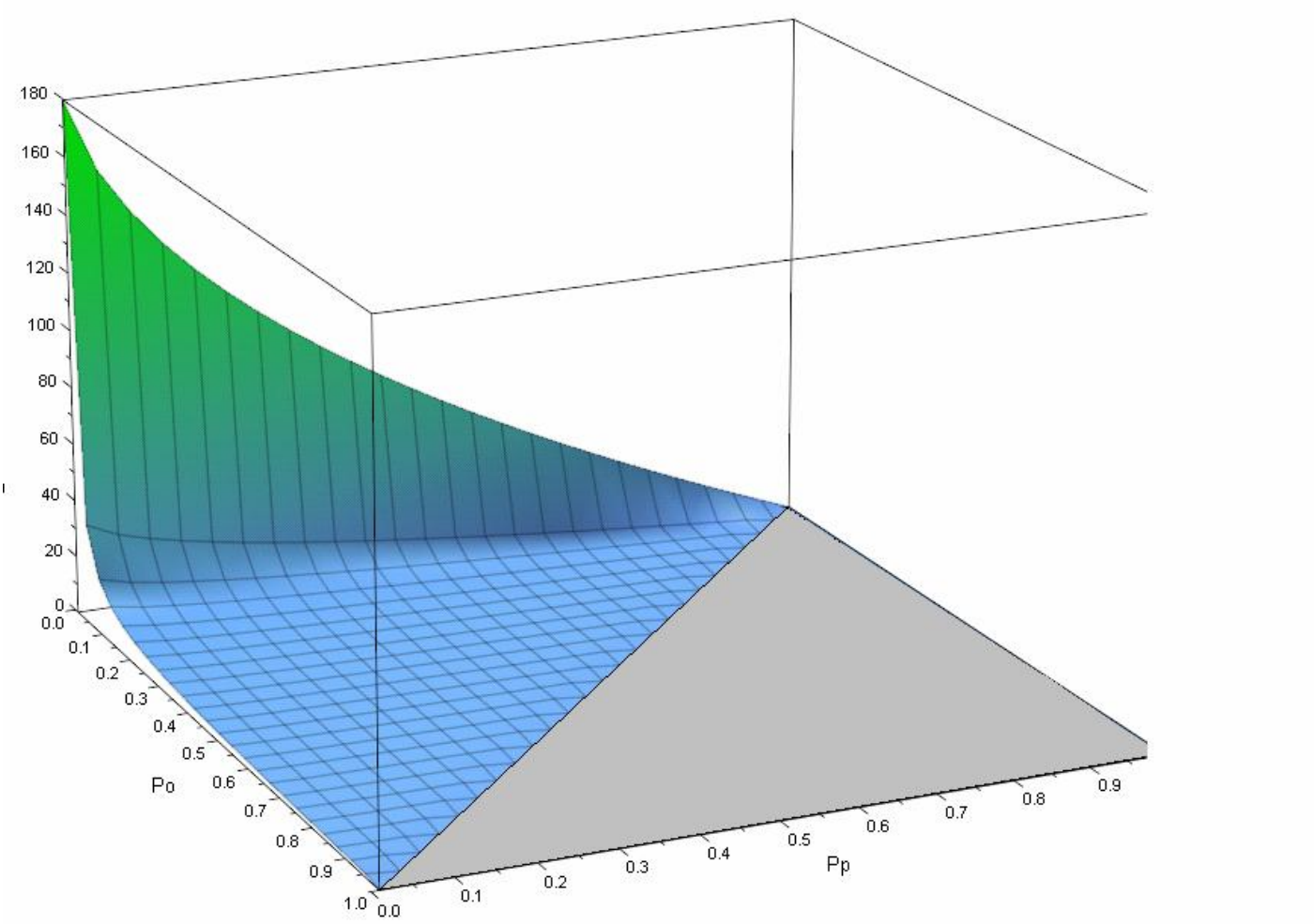
It makes then sense to investigate the function

$$M(P_o, P_p) = \frac{1-\sqrt{Pp}}{1-\sqrt{1-Po}} \qquad (10)$$

as $M(P_o, P_p)$ will effectively limit the maximum ratio $t_4 / t_2$ for which a solution can be found.

Fig. 8 provides a first look into the shape of $M(P_o, P_p)$

**Fig. 8**




Note that, according to what above, $M(P_o, P_p)$ is only defined when $P_o + P_p <= 1$.

The shape has interesting consequences, which we'll analyze in the next paragraph.

## 6. An interesting insight

In [1], we had to exclude some samples from Jørgensen's data set, as no triangular distribution could be found to fit the data. For instance, in the second data set, the following sample could not be used:

$t_2$= 4
$t_4$= 20
$P_o$ = 30%
$P_p$ = 10%

(also indicated, but not used by BetterEstimate, was a most likely time = 16; the actual time was 14).

In this case, $t_4$ / $t_2$ gives 5, while (10) gives 4.186, so (9) is not respected.

What this is actually telling us is that the estimated times are **relatively too distant for the given probabilities**. Let's analyze this statement better:

**- relatively too distant**
When estimates are given as a range [a, b], people tend to look at the width (b-a) as being an important factor (higher width representing higher uncertainty). However, (10) reminds us that we also have to look at the ratio (b/a) as an indicator of uncertainty, and that uncertainty must somehow be reflected in the given probabilities.

**- for the given probabilities**
In this case, although the ratio was quite big, the estimator had high confidence on the maximum value (90%) and lower confidence on the minimum value (70%). In fact, he also indicated a most likely value of 16: this suggests that the optimistic time was really optimistic, yet he indicated a 30% probability of requiring less than 4.

Generally speaking, a failure to find a probability distribution for the given data is usually caused by:

$t_2$ being relatively much smaller than $t_4$ (this usually indicates high optimism).

$P_o$ being too high (this usually indicates high optimism).

In this case, it is suggested that we review the estimate, either by increasing $t_2$ or decreasing $P_o$ (or both).

Alternatively, we may have been too pessimistic. Indeed, we can try to raise the value of (10) also by decreasing $t_4$ and/or decreasing $P_p$. This would only be appropriate if, upon consideration, the estimate is considered too pessimistic. Note, however, that decreasing $P_p$ gives us limited leverage in affecting the value of (10), again indicating that it's probably better to look for overoptimism first.

To clarify this point further, it is important to understand how decreasing $P_p$ or $P_o$ can influence (10).

Given the equation

$M(P_o, P_p) = th$

Where th is a threshold we want to reach (in practice, it will be $t_4$ / $t_2$), we begin by solving the equation for $P_o$ and (separately) for $P_p$:

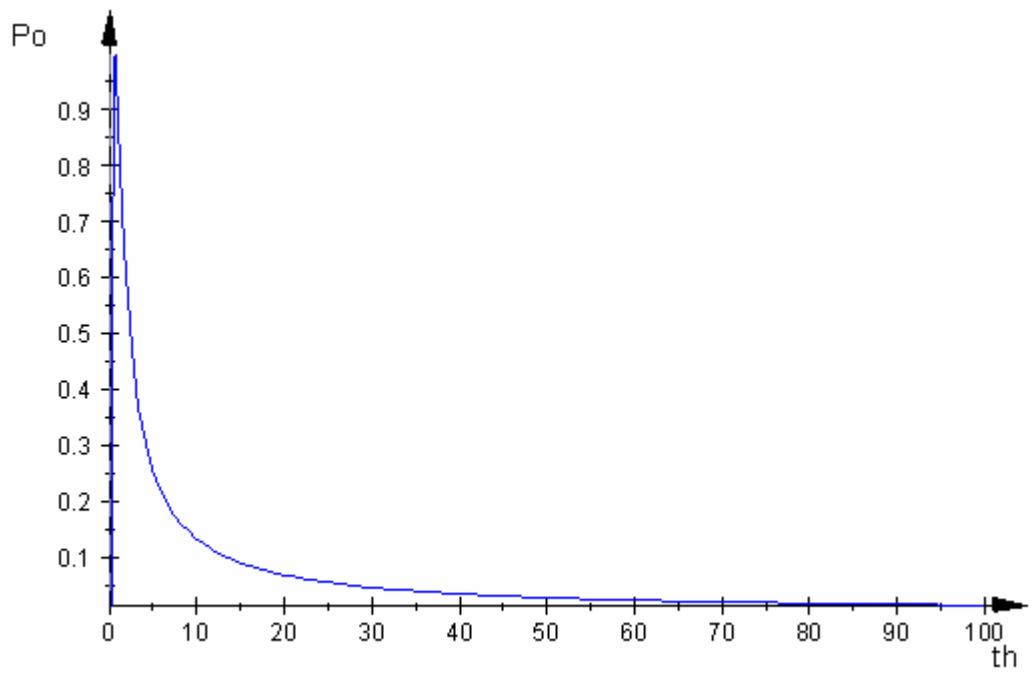
$$P_o(P_p, th) = 1 - \frac{(th + \sqrt{Pp} - 1)^2}{th^2} \qquad (11)$$

$$P_p(P_o, th) = (th \cdot (\sqrt{1 - Po} - 1) + 1)^2 \qquad (12)$$

We'll see that (11) is essentially hyperbolic in th, while (12) is obviously parabolic in th.

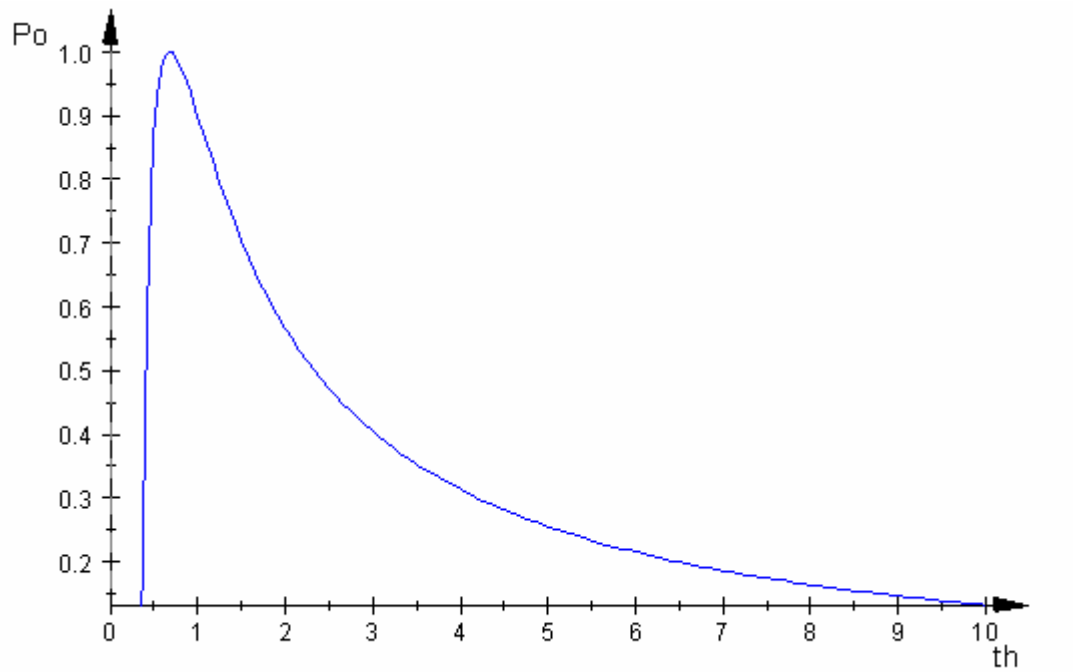
By fixing a value for $P_p$, $P_o(P_p, th)$ becomes obviously a function of th only. For instance, by fixing $P_p = 0.1$ (as in the excluded sample), we obtain the following function:

**Fig. 9**



The function is essentially hyperbolic, with the exception of the [0..1] range as visible here:

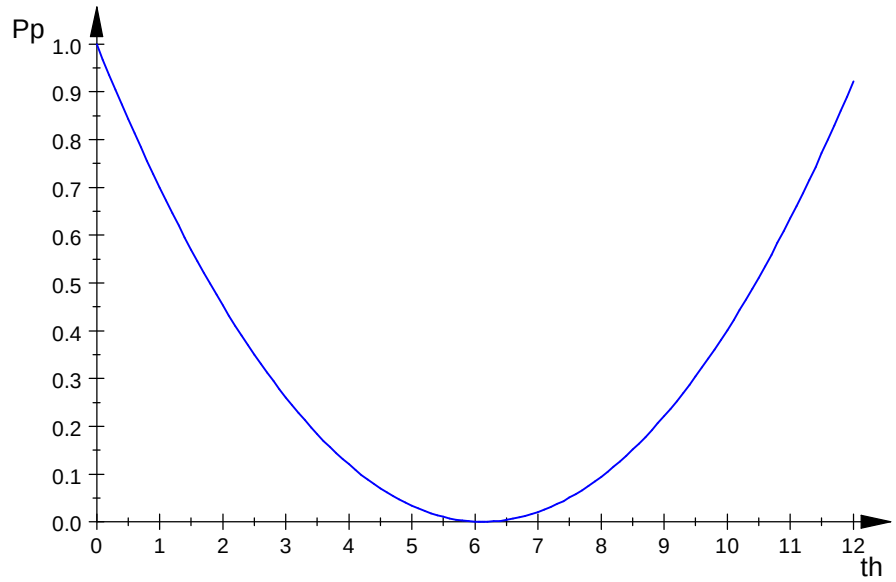
**Fig. 10**




Being hyperbolic, we can obtain very high values for th by decreasing Po. In our case, we could obtain th = 5 (therefore allowing a solution to exist) by just setting Po = 0.2548 (lowering the given 0.3 just a bit). Note that as revealed by Fig. 10, we have a large leverage in reaching high threshold values by simply adjusting Po.

Similarly, by fixing a value for $P_o$, $P_p( P_o, th )$ becomes obviously a function of th only. For instance, by fixing $P_o = 0.3$ (as in the excluded sample), we obtain the following function:

**Fig. 11**



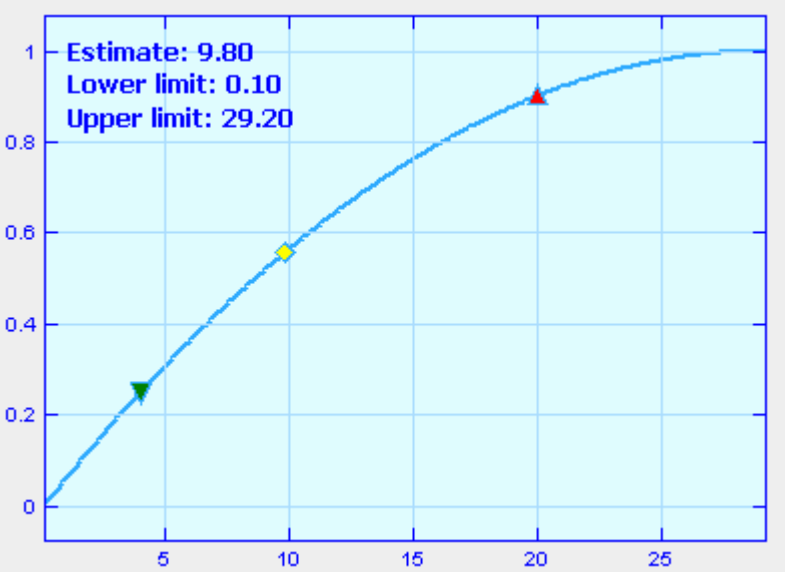

The function is parabolic, and since $P_p <= 1$, its domain is heavily constrained. We can see that, to raise the threshold up to 4, we would have to lower $P_p$ down to 0.0335 (from an initial 0.1). Also, note that while adjusting $P_o$ allows reaching very high thresholds, $P_p$ provides limited leverage (in this case, the maximum threshold is about 12.24).

It is important to understand that minimal adjustments of $P_o$ and/or $P_p$ may be enough to guarantee the existence of a solution, but that does not mean we are making a quality estimate. To better understand the concept, we can take a look at the “best” probability distribution fitting the data:

$t_2 = 4$
$t_4 = 20$
$P_o = 25\%$
$P_p = 10\%$

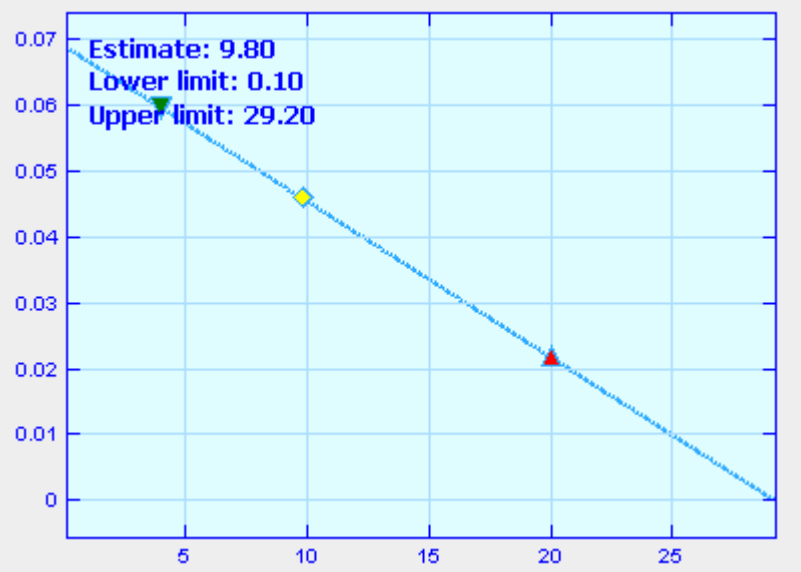
that is, a minimally readjusted input compared to the excluded sample.

**Fig. 12**




Here is the even more explicative (if unusually shaped) probability density:

**Fig. 13**




The resulting triangle is completely skewed. To obtain a reasonable density (not necessarily perfect, but reasonable), it would be better to keep $P_o$ and $P_p$ unchanged, but raise $t_2$ up to 9 (which is the first value that doesn't give a skewed triangle). In this case, BetterEstimate provides an estimated effort of 12.26, which is close enough to the actual 14.

Of course, this process can't be mechanical. A failure in finding a probability distribution, or a highly skewed distribution, is mostly a symptom that the input data are questionable. It's up to the estimator to critically review the input data: a tool can only provide assistance.

To summarize, when a probability distribution can't be found, we must review the estimate for over-optimism, and adjust $t_2$ and/or $P_o$ (they both provide good leverage). If the estimate is considered too pessimistic, is probably better to adjust $t_4$ (which provided a good leverage) than $P_p$ (which doesn't). Looking for excessively skewed distributions can also help to appreciate unbalanced estimates.

## 7. Conclusions and further work

We have presented the math behind the BetterEstimate method for probabilistic effort estimation. The math is relatively straightforward, except for the slightly creative approach needed to obtain a model amenable to heuristic reasoning. Obviously, users of the BetterEstimate tool do not need to understand the underlying math.

The approach relies partially on heuristics, aimed at compensating for over-optimism. This is an area that may require further investigation, as it has sometimes been observed that people tend to be over-optimistic on large tasks, but conservative on small tasks. It would generally be useful to apply the BetterEstimate method under controlled conditions, together with different estimation styles (e.g. top-down vs. bottom-up) and possibly derive better heuristics.

We also derived some relevant insight on the constraints that apply to the input values, and strategies to improve the estimate when a fitting probability distribution cannot be found. Here is an area where the existing tool could be significantly improved: today, a generic error message is provided, while the tool could better assist the user to refine and improve the estimate. This would be an interesting step toward computer-assisted expert estimation.

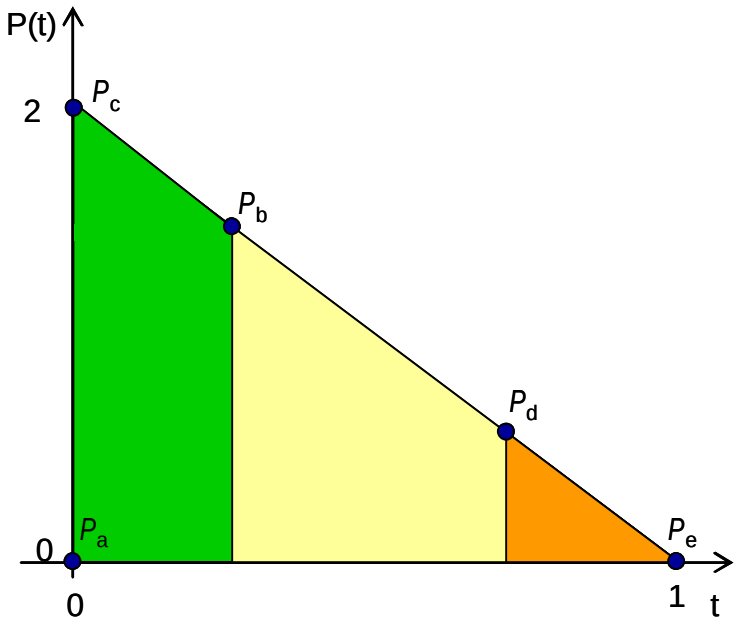
As highlighted in [1], the single most interesting area of research is probably the investigation of techniques to analytically derive the probabilities $P_o$ and $P_p$, using different sources of information like multiple experts (a-la Delphi), paired comparison, risk factors, and so on. Again, providing some kind of automatic support inside an integrated tool could improve our ability to estimate with higher realism and precision.

## 8. Appendix A

Proving that (8) is equivalent to (9) requires three distinct proofs, as (6) defines $t_b$ and $t_d$ in terms of $t_c$, $P_o$ and $P_p$ as piece-wise functions, but one combination is impossible, just like in (7).

As a formal proof will be rather long, we'll offer here a simpler, more intuitive explanation. The ratio $t_d$ / $t_b$ will reach its maximum value when $t_d$ is maximum and $t_b$ is minimum. In order for this to happen, given a $P_o$ and $P_p$ pair, $P_o$ must cover a near-vertical area (so that $t_b$ will be at is minimum) and $P_p$ must cover a near-horizontal area (so that $t_d$ will be at is maximum).
The (degenerate) condition under which this is true is the following:

**Fig. 14**




That is, the triangle is completely skewed, just like in Fig. 13. Please note that the base of the orange triangle (having area $P_p$) is 1- $t_d$, while the base of the green trapezoid (having area $P_o$) is $t_b$. By triangle similarity, we can see that:

(1- $t_d$) / 1 = $y_d$ / 2 **ó** $y_d$ = 2 (1- $t_d$)

$P_p = y_d (1- t_d) / 2 = (1- t_d)^2$

Hence $t_d = 1 - \sqrt{P_p}$

By the same reasoning, considering the complement triangle of the green trapezoid, we can show that:

$t_b = 1 - \sqrt{1 - P_o}$

As this is the limit, degenerate condition, it follows that in general:

$$t_d / t_b <= \frac{1 - \sqrt{P_p}}{1 - \sqrt{1 - P_o}}$$

QED.

## 9. Bibliography

[1] Carlo Pescio, “Realistic and Useful: Toward Better Estimates”, Draft Version 1.2.

[2] J.J. Moder, C.R. Phillips, and E.W. Davis, Project Management with CPM, PERT and Precedence Diagramming. Blitz Publishing, 1995.

[3] Magne Jørgensen, Realism in Assessment of Effort Estimation Uncertainty: It Matters How You Ask, IEEE Transactions on Software Engineering, April 2004.
*Online reference (draft)*
*www.simula.no/departments/ engineering/.artifacts/framingsubmit.pdf*

## Biography

Carlo Pescio is a consultant and mentor for several companies across Europe. He began programming in 1978 and graduated magna cum laude in Computer Science in 1991. He has designed software for medical devices, industrial process control, banking, finance, CAD, and several other fields. His recent interests focus on software design, project management for software-intensive systems and diagrammatic reasoning. He is as member of IEEE Computer Society, the IEEE Technical Council on Software Engineering and the ACM. Contact him at pescio@eptacom.net.