Advances in Social Sciences Research Journal – Vol.7, No.1
Publication Date: Jan. 25, 2020
DoI:10.14738/assrj.71.7720.
Luo, B. (2020). On the Robustness of PERT Fittings in Agricultural Yield Insurance. *Advances in Social Sciences Research Journal, 7(1)* 367-383.

On the Robustness of PERT Fittings in Agricultural Yield Insurance

Boyingzi Luo
School of Insurance,
Central University of Finance and Economics,
Beijing, China. P. R.

ABSTRACT
In agricultural insurance practice, risk and indemnity payment are often incurred from individual farmer's yield. However, high administration cost and data scarcity are simultaneously quite often seen, which form huge burdens for insurers to adequately rate insurance products. Under this circumstance, some methods that could be used to estimate farmers' yields, in particular, their distributions, are urgently needed. Among these methods, a so called PERT fitting technique often prevails due to its simplicity which only requires very little knowledge about the yield history, that is frequently implemented by both academics and practitioners. However, the very limited information used would sometimes cause severe bias, in other words, the reliability of this method is yet to be examined. In this paper, I used Monte Carlo experiments to test the robustness of PERT fittings under Var and CTE risk measures in different scenarios. The result proves that PERT method is indeed robust and trustworthy.

Key words: PERT fittings, Monte Carlo simulations, VaR, CTE.

INTRODUCTION

The main purpose of this research is to figure out some properties in fitting distributions for assumed populations with conventional PERT techniques, test their behaviors under different scenarios that are generated by controlling skewness and variance before evaluating their robustness in distribution fitting through properly designed experiments. In this research, I mainly focus on distributions derived from non-negative random variables and their fittings under actuarial context, which leads to the concentration to some particular risk measures, such as the left-tail Value at Risk and the corresponding Conditional Tail Expectation. Monte Carlo simulations will be the basic tool in random sampling from given distributions for its applicability and flexibility.

PERT DISTRIBUTION

PERT originally refers to "Program & Project Evaluation and Review Technique", which is, as a management tool, developed for the Program Evaluation Branch of the Special Projects of the Navy (U.S. Dept. of the Navy.(1958)). The entire process of conventional PERT includes a distribution fitting procedure that uses a special three-parameter general Beta distribution (Malcolm,D.G. et al. [1]).

In general, a Beta distribution on (0,1) with parameters $\alpha > 0$ and $\beta > 0$ has a density function defined as

$$f_{Beta(\alpha,\beta)}(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, \quad where \quad B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)},$$

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1}e^{-x}dx, \quad with \quad \alpha > 0.$$

Normally, if a random variable $X \sim Beta(\alpha,\beta)$ on $(0,1)$, by stretching its range with the form $Y = a + (b-a)X$, where $0 < a < b < \infty$, a non-negative random variable is obtained as so called general Beta distributed on (a,b). By transforming, a theoretic general Beta distribution has properties

$$\mu = min + \frac{\alpha}{\alpha+\beta}(max - min), \qquad (1.1)$$
$$mode = min + \frac{\alpha-1}{\alpha+\beta-2}(max - min), \quad for\ \alpha > 1,\ \beta > 1, \qquad (1.2)$$
$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \qquad (1.3)$$

In using general Beta to fit distributions, the usual methods are of two main kinds, maximum likelihood estimations (MLEs) and Moment/Quantile matchings. The former requires all information of each sample, whereas the latter requires some pre-chosen sample statistics. Under the circumstances where complete datasets (samples) are hard to acquire, moment/quantile matching methods seem to be more feasible due to its less stringent requirement for sample information. Moreover, among all moments/quantiles, μ, $mode$ and σ^2 are the most easily captured ones. In this case, one is capable of using any two of the above equations to solve $\hat{\alpha}$ and $\hat{\beta}$ by substituting μ, $mode$ and σ^2 by $\hat{\mu}$, $\widehat{mode}$ and $\hat{\sigma}^2$ respectively.

Under the context of its original use, PERT tackles to a triangular estimation problem, in which all data is categorised into three clusters, min, max and $mode$. The motivation is to transform uncertain problem into relatively certain one by categorising or ranking, instead of processing continuous stochastic information. In most cases where PERT is implemented, decision makers want to obtain some distribution information of concerned variables based on very simple point estimators instead of those for continuous behaviors. For instance, if an insurer needs to know what an individual farmer's yield distribution looks like with absence of the farmer's historical yield data, he could ask the farmer for his $\widehat{min}$, $\widehat{mode}$ and $\widehat{max}$ before conducting PERT fittings on them to reach the purpose. In this case, the inputs for PERT fitting are quite subjective and actually infers that: a continuous data set, without any loss of generality, is regarded as a combination of three clusters, and yields a sample group composed of only three values of observations. Indeed, in practice, with acquired $\widehat{mode}$, $\widehat{min}$, $\widehat{max}$ and the assumption that *the possibility for occasions that mode happening is 4 times than that of max plus min*. It is often estimated that

$$\hat{\mu} = \frac{\widehat{max} + 4\widehat{mode} + \widehat{min}}{6}, \qquad (1.4)$$
$$\hat{\sigma}^2 = \left(\frac{\widehat{max} - \widehat{min}}{6}\right)^2. \qquad (1.5)$$

Obviously, the above estimations closely relate to the assumption of *"4 times"*, and they are precise if population coincides with $general\ \ Beta(4,4)$. Actually, it could be explained in an easier way, that is, with the assumption of two independently distributed events $X_{optimitstc}$ and $X_{pessimestic}$ with symmetric assumptions *"mode happens 2 times that min"* and *"mode happens 2 times than max"*. Indeed, the constant 4 could be generalized as k if verification is needed. In this case, the choice of k becomes quite sensitive. Moreover, if a general Beta distribution has identical $mean$ and $mode$, which indicates unskewness and yields $\alpha = \beta$, the estimation (1.4) will be unbiased regardless of k, whereas (1.5) being biased in many cases even if the distribution is unskewed. Moreover, $\alpha \neq \beta$ could be used in conducting skewed fittings, for instance, as $\alpha + \beta = k$ suggested by Sasieni.

With the use of (1.2) and (1.4) or (1.2) and (1.5), $\hat{\alpha}$ and $\hat{\beta}$ for PERT distribution could be obtained. One should note that the original PERT is a very sophisticated process combining multiple phases, which is, in most cases, information about the concerned variable, elapsed time, for instance, is simply approximated by (1.4) and (1.5) (U.S. Dept. of the Navy. (1958)), instead of the fitted PERT distribution. In this research, I just concentrate on the PERT distribution, with nothing else relates to the whole program evaluation and review procedure.

POPULATION ASSUMPTIONS

In actuarial and risk assessment practices, the most commonly treated data is that generated from the class of non-negative random variables. For example life durations, farming yields and all sorts of claim payments etc. In this research, in a general way, I use random samples drawn from predetermined parametric populations as representations for non-negative datasets in real life applications. The reason for that is quite natural, because I can't simply deny the fact that a dataset comes from a parametric population, despite not knowing from what distribution exactly. As the fact, the choice of population distribution becomes essential under this context. I believe that a good choice must at least have the following properties:

1. It must be adequately representative to reality, its range, shape and other properties has to suit the research purposes.
2. It must be of great flexibility as it could be adjusted into multiple cases which are of concerns.
3. It must be easily operate, the relations of its parameters toward its moments and quantiles should be as simple and clear as possible.

By considering those in mind, I find modified Gamma distribution might be the best choice as it attains all the desired properties. Moreover, despite the fact that Gamma distribution is always right-skewed, a mirror image operation could be used to solve this problem, that is, by $min(x) + min(x) - x$ where x is a vector of samples from right skewed Gamma distributions. The converted samples have identical variance but opposite skewness against the original ones. Meanwhile, general Beta distribution seems to be an alternmative, but it is hard to operate as the relations of its moments and mode toward its parameters are of comparatively more sophisticated forms, besides, general Beta distribution requires a predetermined fixed range, which may lack variability to some extent.

In general, a modified Gamma distribution $m + X$ where m is a non-negative constant, X is a Gamma distributed random variable with shape parameter $\gamma > 0$ and scale parameter $\theta > 0$, posses a density function

$$f_{Gamma(\gamma,\theta)}(x) = \frac{1}{\Gamma(\gamma)\theta^{\gamma}} x^{\gamma-1} e^{-\frac{x}{\theta}},$$

$$with \quad mean = \gamma\theta + m, \quad variance = \gamma\theta^2, \quad skewness = \frac{2}{\sqrt{\gamma}}.$$

In this research, I assume that the population coincides with modified Gamma distributions, and I are concern about different scenarios generated from them by controlling population mean, variance and skewness. Kurtosis is not considered because it will make the research much more redundant, other information about the population will be ignored as well, as they are relatively less important than moments of the first three orders.

VaR AND CTE

Under actuarial context, censored data, such as claim payment, is usually of primary concerns to many practitioners. Because in many cases, tail information is of greater importance than that of center. In this research, I am going to tackle to both censored and uncensored data, which means they could either be realized loss or risk sets where loss comes from. In the first case, PERT method is simply implemented to loss data, whereas in the second case, data derived from risk sets with potential loss contained in is fitted. In this case, some additional methods should be undertook in order to separate actual loss from populations. Hence, I consider two widely used risk measures, Value at Risk (VaR) and Conditional Tail Expectation (CTE) to reach the purpose.

Unlike traditional definition on the right tail, I use a left tail definition of VaR, which is

$$VaR_p(X) = \inf\{x: Prob(X \leq x) \geq p\} = Q(p) = F^{(-1)}(p),$$

where p is the left tail probability (significance level), Q(p) as the p-quantile and $F^{(-1)}(p)$ *being the pseudo inverse of F.*
And corresponding left tail CTE defined as

$$CTE(p) = E[X|X \leq VaR_p(x)].$$

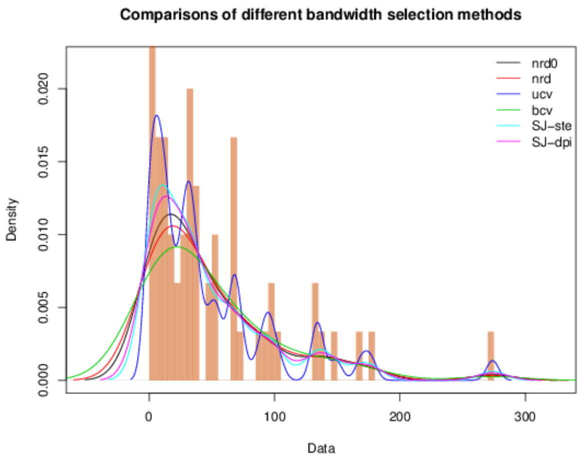
With these definitions, if population is regarded as a risk set with loss triggered in percentiles to its left tail, VaR, as an threshold, could be used as an indication of the severity to this risk, because claims will only occur with values less than it. Consequently, CTE infers to the expected loss of the given population.

PERT FITTINGS

Population Distribution Fitting

Here I implement PERT and general Beta fittings on populations with modified Gamma distributions under different combinations of skewness and variance. The experiments are conducted to use them fitting dataset composed of 10000 random samples drawn from a given modified Gamma populations. As the value $\widehat{mode}$ has to be estimated from the samples (with equally $\frac{1}{10000}$ empirical frequency each value), it has to be certain where the density reaches its peak. I use kernel density estimation (KDE) (Rosenblatt,M. [2], Parzen,E.[3]) to figure that, with the help of some *R* (a statistical software) packages.

In KDE, the choice of kernel function and bandwidth will both cause influence to the result. According to Wand,M.P. and Jones,M.C.[4], the loss of efficiency in estimation is small with the use of different kernel functions, but huge in distinct bandwidth selections. So in the latter sections, I will consistently use Gaussian kernel for its convenience and usability. Meanwhile, there are many commonly used methods in optimal bandwidth selection and their efficiencies are shown for a small-sample case with 60 samples drawn from a Gamma population.



$Gamma(1,50)$.

As shown in the figure, the method “ucv", which refers to “unbiased cross validation" (Bowman,A.W. (1984), Hall,P. et al. (1992)) seems to be the best in capturing changes of the empirical density to a small-size sample group, and it is a suitable method for a wide variety of sample groups. Therefore, I will use it to extract $\widehat{mode}$ in the latter. The empirical density estimated by KDE is a smoothed curve, without simple closed form expressions in most cases, but estimations for any values within the range could be conducted. In this case, I take 512 equally-distant points from $(sample(min), sample(max))$, obtain 512 estimations for the empirical density and get the $\widehat{mode}$ by picking the largest from them.

For PERT fitting, I use $\hat{\mu}$ derived from (1.4), combining with (1.1) and (1.2), I have

$$\frac{\widehat{mode}-\widehat{min}}{\widehat{max}-\widehat{min}} = \frac{\hat{\alpha}-1}{\hat{\alpha}+\hat{\beta}-2}, \qquad (2.1)$$

$$\frac{\widehat{max}-5\widehat{min}+4\widehat{mode}}{6(\widehat{max}-\widehat{min})} = \frac{\hat{\mu}-\widehat{min}}{\widehat{max}-\widehat{min}} = \frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}}, \qquad (2.2)$$

yields

$$6\left(\frac{\hat{\mu}-\widehat{min}}{\widehat{max}-\widehat{min}}\right) = \hat{\alpha},$$

$$6\left(\frac{\widehat{max}-\hat{\mu}}{\widehat{max}-\widehat{min}}\right) = \hat{\beta},$$

and

$$\hat{\alpha} + \hat{\beta} = 6.$$

This is exactly the case where $k = 6$ as in Sasieni (1986), and coincides with the method many statistical softwares implement (*@Risk* etc). The reason not to use (1.5) is intuitively because that it actually causes double bias as it is derived based on estimation (1.4).

In comparison, I also implement general Beta fitting and use (1.1) and (1.2) to solve the parameters. By this, I assume that the population mean is known, or perfectly estimated with $\hat{\mu} = \mu$. The reason for that is quite natural, because no matter how I modify the parameters in PERT distribution, it is still a special form of general Beta distributions. In result, somehow, this sort of general Beta fitting may be regarded as the “best" case of PERT fittings, which to some extent allows us to compare robustness among different PERT methods. In addition, I also assume that $\widehat{min}$ and $\widehat{max}$ equal to min and max in all cases.

By controlling γ, θ and m, I generate four populations $Gamma(2,25) + 303$, $Gamma(100,3.53)$, $Gamma(2,10) + 333$ and $Gamma(100,1.41) + 212$ which all have the same mean (353) but distinct skewness and variance, as the representations of four typical scenarios. However, population with negative skewness is spared because the fittings are symmetric (the right skewed estimation is symmetric in shape of the left skewed case) as all the fittings are of Beta forms.

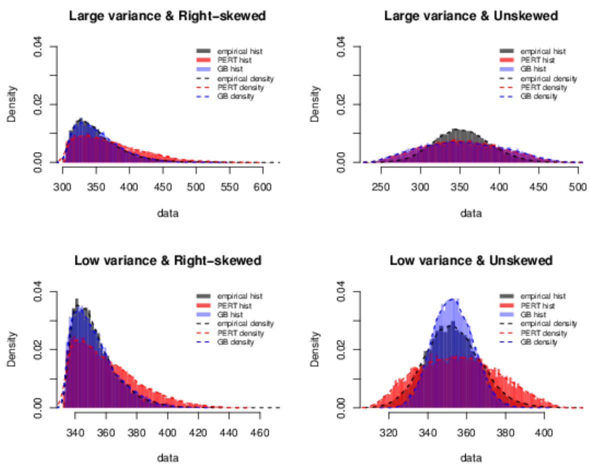


Figure 2: Densities of PERT and GB fittings in different scenarios.

The figure shows that in the right skewed cases, GB fittings present much more accurate results than PERT, as they precisely capture the peaks and tails of the empirical distributions while PERT fittings exhibit longer and heavier tails. The analogous results could be expected for left skewed cases. In the unskewed cases, where theoretic population skewness I implement ($\frac{2}{\sqrt{100}} = 0.2$) is used to approximate unskewness, the results from Figure 2 seem to reveal some different patterns. Theoretically, if the sample group is identically unskewed (when $n \to \infty$ and $\gamma \to \infty$), $\hat{\mu}$ would equal to $\widehat{mode}$, and yield $\hat{\alpha} = \hat{\beta}$ to both GB and PERT distributions. Moreover, under this circumstance, PERT would indeed become $GB(3,3)$ distribution with relation $\hat{\alpha} + \hat{\beta} = 6$, but GB fittings would be infeasible as (1.1) and (1.2) would be equivalent and only result in $\hat{\alpha} = \hat{\beta}$. Therefore more information, such as $\hat{\sigma}^2$, must be added in order to solve the parameters. Meanwhile, in GB fittings on (min, max), the parameters estimated from (1.1) and (1.2) are

$$\hat{\alpha} = \frac{(\hat{\mu}-min)(2\widehat{mode}-max-min)}{\widehat{mode}-\hat{\mu}}, \qquad (2.3)$$

$$\hat{\beta} = \frac{\hat{\alpha}(max-\hat{\mu})}{\hat{\mu}-min}. \qquad (2.4)$$

Theoretically, in GB distributions, I have

$$Skewness > 0 \Rightarrow mode < \mu \quad and \quad 2mode < min + max \Rightarrow \alpha > 0,$$
$$Skewness < 0 \Rightarrow mode > \mu \quad and \quad 2mode > min + max \Rightarrow \alpha > 0,$$
$$Skewness = 0 \Rightarrow mode = \mu = \frac{min + max}{2} \Rightarrow \alpha = \beta > 0.$$

These relations guarantee the positiveness (regularization properties) of parameters α and β in GB distributions. But when using it to conduct fittings, very often, the $\widehat{mode}$ and $\hat{\mu}$ derived from datasets would fail to follow those relations, thus leading to negative $\hat{\alpha}$ and $\hat{\beta}$ and resulting in invalidness. The same analysis could be applied to modified Gamma fitting which is widely used in actuarial practices. Because in modified Gamma fittings, relations $\hat{\gamma}\hat{\theta} + \hat{m} = \hat{\mu}$ and $(\hat{\gamma} - 1)\hat{\theta} + \hat{m} = \widehat{mode}$ are used, yield $\hat{\gamma} = \frac{\hat{\mu}-\hat{m}}{\hat{\mu}-\widehat{mode}}$ and $\hat{\theta} = \hat{\mu} - \widehat{mode}$ when $\hat{\gamma} \geq 1$, which guarantees the positiveness of these two parameters and may fail if datasets don't coincide with the requirement that $\hat{\mu} > \widehat{mdoe}$. Besides, more information is needed in estimating $\hat{m}$. It should be emphasized that these failures are irrelative with the size of datasets, but relating to the underlying parametric assumptions, which cannot be eliminated as long as I still use the

"μ & $mode$ $method$". However, PERT fitting is feasible almost everywhere as it only requires $min < mode < max$, which states that it has greater universality.

On the other hand, under some circumstances where datasets are closely unskewed, that is, when $|\hat{\mu} - \widehat{mode}|$ is very small. By (2.3), it is obvious that the estimations $\hat{\alpha}$ and $\hat{\beta}$ in GB fittings would be highly unstable as they are sensitive to any tiny change of the difference. Again I use dozens of homogeneous datasets, each contains 10000 random samples drawn from $Gamma(100,3.53)$ which has $skewness = 0.2$, implement GB and PERT fittings on them to test their robustness. Some typical results are displayed in Figure 3.

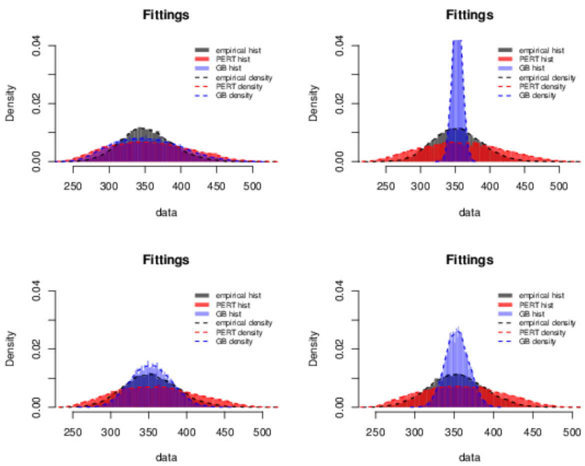


Figure 3: Densities of PERT and GB fittings under $Gamma(100, 3.53)$ populations.

The results prove that even under large and homogenous datasets, any tiny change in $|\hat{\mu} - \widehat{mode}|$ will be amplified in estimating $\hat{\alpha}$ and $\hat{\beta}$ in GB distributions, hence producing entirely distinct shapes, while some of them may cause devastating impact under actuarial context due to their significant underestimations to the empirical tails. However, in highly skewed datasets, as $|\hat{\mu} - \widehat{mode}|$ increases, the sensitivities of $\hat{\alpha}$ and $\hat{\beta}$ towards it will decline, therefore GB fittings under this circumstance will be more robust. So far, all the results I have obtained indicate that PERT seems to prevail in population distribution fitting under limited information of $\hat{\mu}$ and $\widehat{mode}$ for its robustness and universality.

Additionally, the impact of variance towards fittings is quite clear, which is, smaller variance shrinks the range of all empirical and fitted distributions, without changing the patterns of shapes and reciprocal positions of them. Consequently, in the latter, I will also treat population variance as a controlled variable.

VaR and CTE Distribution Fittings.

In actuarial practice, if the task is to use PERT to fit some certain kind of populations, for instance, loss or claim payment, though robust, the result may be relatively unacceptable as I have seen. But in other cases that the population to be fitted is some kind of risk set, which means, I don't care about all its information but only the tail-related one, the situation in consideration would be totally different. Contemporarily, I assume that loss is triggered at certain predetermined percentiles, which is indeed the case in many actuarial uses, then without any loss of generality, I could use the left tail VaR and CTE, which are defined previously, to examine the behaviors of "severity of loss" and "expected loss" in PERT fittings under different scenarios.

The idea is to firstly construct some situations close to reality. For this purpose, I build a collection composed of 2000 independent risk sets, (e.g, it could be a group of 2000 individuals facing potential loss in the left tail) who have identical underlying population distributions (modified Gamma distributions), which infers that all risk sets in this collection are highly homogenous. Then I use unbiased Monte Carlo method (By resetting the pseudo random number generator in each time of sampling to avoid circulations in long sequence) to draw 60 random samples from every risk set, which is to simulate the historical behaviors of each individual. At this stage, the sample quantity "60" is carefully selected for the reason that I hope to obtain a sample group with significant variabilities, simultaneously I don't want the assumed underlying distribution causes too much influence to the samples drawn from it. Ideally, our main purpose is to control skewness and variance to the whole collection in general manners, instead of making each sample set present obvious features of Gamma distribution. Then I implement PERT fittings to every risk set's 60 samples, by this, together with some calculations, I obtain two sets of 2-tuples

$\{(VaR^{i}{}_{empirical}, VaR^{i}{}_{PERT}), i = 1{:}2000\}, \{(CTE^{i}{}_{empirical}, CTE^{i}{}_{PERT}), i = 1{:}2000\}$, and integrate them in two distributions. I use p-quantile to calculate VaR by implementing a continuous median-unbiased quantile estimation method suggested by Hyndman,R,J. and Yanan Fan (1996). Meanwhile, under the assumption that every sample set posses a continuous density, CTE is equivalent to Tail Value at Risk (TVaR), which is

$$TVaR_p(X) = \frac{1}{p}\int_0^p VaR_\xi(X)d\xi, \qquad 0 < p < 1. \quad (2.5)$$

In order to reduce estimation error, I calculate CTE (TVaR) by implementing a numeric integration technique based on quadratic approximation procedure(Piessens,R. et al. [5]), combined with the quantile function estimation method as above.

As the same as in the previous section, I adjust γ, θ and m in the underlying modified Gamma populations to obtain scenarios with fixed variance and different skewness. Moreover, I could calculate VaR and CTE in the left skewed cases either with samples generated by mirror image operation from right skewed populations or simply calculating them on the right tail of right skewed populations. Moreover, I also add some comparisons between $p = 0.05$ and $p = 0.15$.

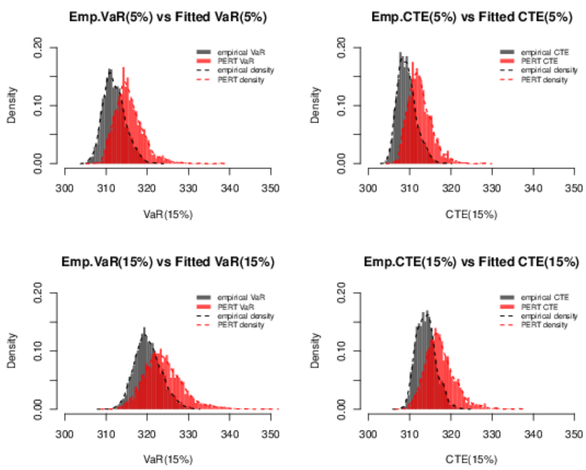


Figure 4: Densities of VaR and CTE fittings under population skeness=1,41, variance=1250.

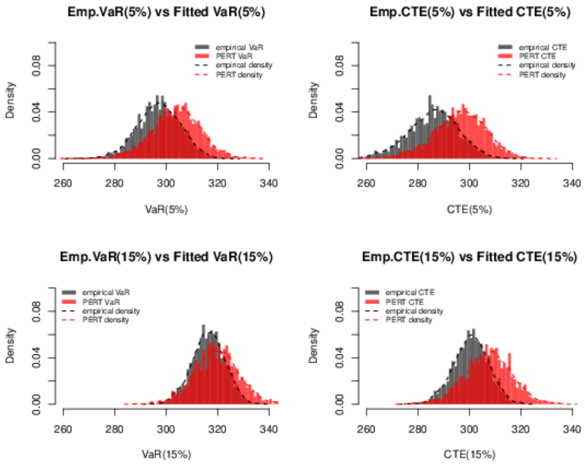


Figure 5: Densities of VaR and CTE fittings under population skeness=0.2, variance=1250.

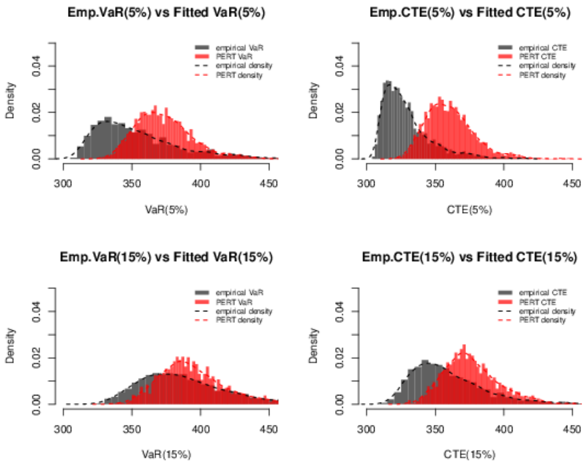


Figure 6: Densities of VaR and CTE fittings under population skeness=-1,41, variance=1250.

The results show that under all considered circumstances, PERT fittings in VaR and CTE present some overestimating patterns toward the empirical, and compared with population fittings, the fitted VaR and CTE distributions seem to be relatively closer to the empirical ones. This is due to the narrowing discrepancies as VaR and CTE only take information form left part of the original population distributions. Apparently, in every scenario, PERT fittings in both VaR and CTE display analogous shape toward the empirical ones, additionally with similar range and peak. Indeed, by (2.5), CTE is the direct result of VaR, consequently, its behavior would be determined by corresponding VaR fittings under $0 < \xi < p$.

The results also reveal a phenomenon that both empirical and PERT fittings in VaR and CTE tend to concentrate in distribution with increasing skewness. Intuitively, as skewness descending, left tail information obtained from every risk set within the collection, specifically under small significance level p, would be of greater variability due to the nature that samples are less gathering in the left tail when such tail is long and heavy. In result, this would cause larger range and variance with lower peak and kurtosis in both empirical and PERT fittings under VaR and CTE risk measures.

However, in the cases that $skewness = 0.2$ and $skewness = -1.41$, PERT fittings in VaR and CTE tend to converge to the empirical as p increasing from 0.05 to 0.15, while diverge in the case with $skewness = 1.41$. For detailed comparisons, I illustrate theoretic CDFs of modified Gamma populations together with those for corresponding PERT distributions under right, unskewed and left skewed cases where they are generated from $Gamma(2{,}25) + 303$,

$Gamma(100,3.53)$ and the left skewed version of $Gamma(2,25) + 303$ respectively. Note that the theoretic PERT distributions are established by a group of 100,000 samples drawn from modified Gamma populations in order to guarantee finite range. For this research, $p = 0.4$ will be fairly enough so that result is shown in Figure 7 with $0 < p < 0.4$.

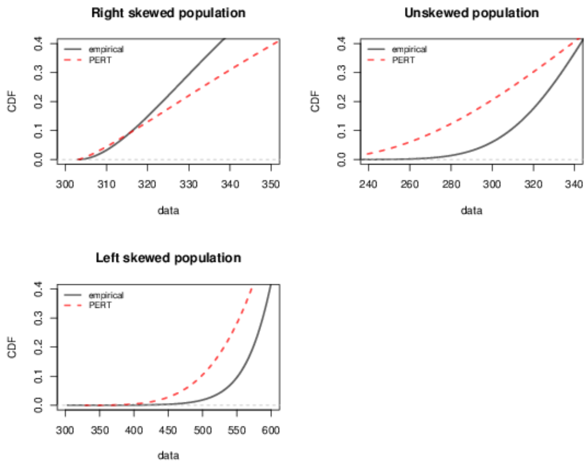


Figure 7: Theoretic CDFs of population and PERT distribution.

Figure 7 explains why PERT fittings diverge from $p = 0.05$ to $p = 0.15$ in right skewed cases while converge in the rest by the nature of theoretic CDFs. However, the theoretic reciprocal position between population and PERT distribution contradicts with the previous fitting results that PERT overestimates VaR and CTE towards empirical under all skewness. This actually give rise to another question, that is, How will sample size affect empirical and PERT distributions?

Obviously, as sample size increasing, more samples are collected which will affect $\widehat{mode}$, min and max. Let X_n be a sample group drwan from a modified Gamma population which has size n, clearly, $max(X_n)$ is non-decreasing with n, while $min(X_n)$ is non-increasing with n. Let $\Delta\lambda$ be the absolute increment of a variable λ caused by increment of sample size Δn, denote $min(X_n)$ and $max(X_n)$ by min and max. Theoretically, I have $\Delta max \geqslant \Delta min$ under right skewed populations, and vice versa. Meanwhile, I implement some random sampling simulations on a $n \times 2000$ data collection with sample size n of each risk set varies from 60 to 510 with 50 increment each time ($\Delta n = 50$), typical results under two populations $Gamma(2,25) + 303$ and $Gamma(100,3.53)$ are shown in Figure 8. I find that, in average, indeed $\Delta max \geqslant \Delta min$ and all Δmax, Δmin and $\Delta\widehat{mode}$ will approach to 0 as n increasing. Besides, sample size will cause slightly systematic impact on $\widehat{mode}$ for populations with $skewness \neq 0$, which is, a pattern that $|\widehat{skewness}|$ tends to be smaller than $|skewness|$ and gradually converge to it. Meanwhile, I need to elaborate that $\widehat{mode}$ is actually extracted with KDE method, instead coming from empirical dataset which is discrete.

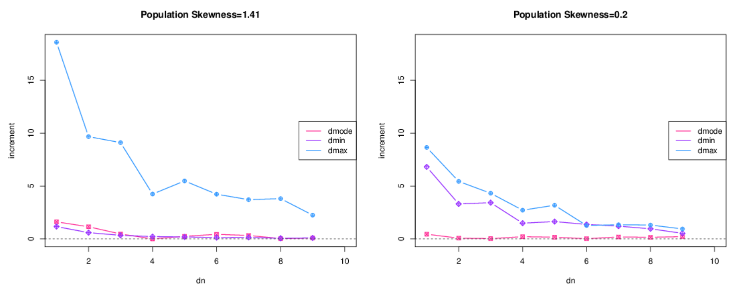


Figure 8: Δmax, Δmin and $\Delta\widehat{mode}$ generated by Δn.

Moreover, I simulate a density comparison among four sample sets, which contain 60, 200, 400 and 600 samples respectively from a $Gamma(2,25) + 303$ population, and implement corresponding PERT fittings toward them, results are shown in Figure 9 (Note that PERT densities are extracted from 2000 random samples from corresponding fitted PERT distributions). I find that, the results coincide with the findings displayed in Figure 8, with an obvious pattern that samples from small-sized dataset are less gathered in both right and left tails, while the peak ($\widehat{mode}$) deviate from the theoretic value which is larger. This is quite understandable, because under small sample size, samples are insufficiently clustering in the tails (presents shorter tails and enlarge the empirical densities in tails) where theoretic densities $f(x)$ are very low (approaching 0), whereas over-clustering on the right side of theoretic $mode$ where the theoretic probability $\int_{mode}^{max} f(x)dx$ is larger than $\int_{min}^{mode} f(x)dx$. Meanwhile, PERT fittings under small-sized samples tend to be more steep, which will be discussed later. If I carefully observe the trends in Figure 9, I may find that the reciprocal tail position that PERT lies below the empirical distributions in both tails might be the prior cause which would lead PERT overestimate the empirical VaR and CTE in small-sized sample set. (Similar trend could be found in low-skewed and left skewed populations).

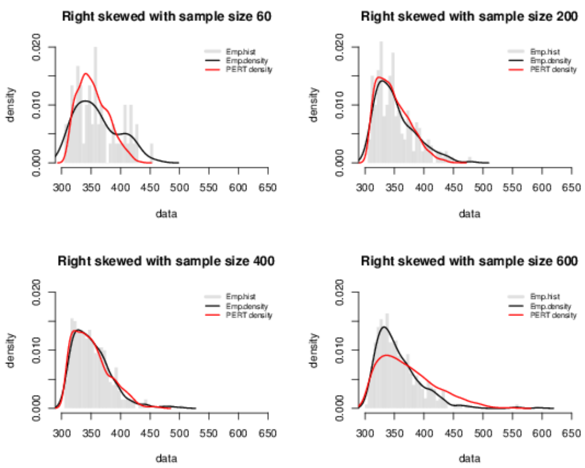


Figure 9: Density comparison between population and PERT distribution.

At this phase, I attempt to build a mathematical framework in illustrating the effect of Δn to PERT fitting. Intuitively, I have a parametric relation $\hat{\alpha} + \hat{\beta} = 6$ so that only $\hat{\alpha}$ is analysed, and contemporarily under right skewed populations with $\frac{mode-min}{max-min} \in (0,\frac{1}{2})$. Besides, I also assume $\frac{\widehat{mode}-min}{max-min} \in (0,\frac{1}{2})$ for any sample group generated from such populations. At first, by (2.1) and (2.2), I know that

$$\hat{\alpha} = \frac{2x_1x_2 - x_2}{x_1 - x_2}, \qquad (2.6)$$

where

$$x_1 = \frac{\widehat{mode}-min}{max-min}, \qquad (2.7)$$

$$x_2 = \frac{\frac{min+max+4\widehat{mode}}{6} - min}{max-min} = \frac{5}{6} - \frac{2}{3}\left(\frac{max-\widehat{mode}}{max-min}\right). \qquad (2.8)$$

Consequently, I have $0 < x_1 < x_2 < \frac{1}{2}$, so that

$$\frac{\partial\hat{\alpha}}{\partial x_1} = \frac{x_2(1-2x_2)}{(x_1-x_2)^2} > 0, \qquad \frac{\partial\hat{\alpha}}{\partial x_2} = \frac{x_1(2x_1-1)}{(x_1-x_2)^2} < 0. \qquad (2.9)$$

Let $x_2 = x_1 + \delta$ with $0 < \delta < \frac{1}{2}$, I obtain

$$\frac{\partial\hat{\alpha}}{\partial x_1} + \frac{\partial\hat{\alpha}}{\partial x_2} = \frac{-2\delta^2+(1-4x_1)\delta}{\delta^2},$$

therefore

$$\frac{\partial\hat{\alpha}}{\partial x_1} + \frac{\partial\hat{\alpha}}{\partial x_2} < 0, \quad if \quad x_1 > \frac{1}{4}, \quad or \quad x_1 \leqslant \frac{1}{4}, \quad \frac{1-4x_1}{2} \leqslant \delta < \frac{1}{2}, \quad (2.10)$$

$$\frac{\partial\hat{\alpha}}{\partial x_1} + \frac{\partial\hat{\alpha}}{\partial x_2} \geqslant 0, \quad if \quad x_1 \leqslant \frac{1}{4}, \quad 0 < \delta < \frac{1-4x_1}{2}. \quad (2.11)$$

(2.9), (2.10) and (2.11) reveal that, under cases where $x_1 \in (\frac{1}{4},\frac{1}{2})$ with $x_2 \in (x_1,\frac{1}{2})$ or $x_1 \in (0,\frac{1}{4}]$ with $x_2 \in [\frac{1-2x_1}{2},\frac{1}{2})$, the impact of the decrement in $\hat{\alpha}$ with Δx_2 outweighs the increment in $\hat{\alpha}$ with Δx_1, whereas this impact is reversed under cases where $x_1 \in (0,\frac{1}{4}]$ with $x_2 \in (x_1,\frac{1-2x_1}{2})$. Moreover,

$$\partial(\frac{\partial\hat{\alpha}}{\partial x_1} + \frac{\partial\hat{\alpha}}{\partial x_2})/\partial\delta = \frac{4x_1-1}{\delta^2}. \quad (2.12)$$

Combined with (2.10) and (2.11), (2.12) infers that, under cases $x_1 \in (\frac{1}{4},\frac{1}{2})$ with $x_2 \in (x_1,\frac{1}{2})$ or $x_1 \in (0,\frac{1}{4}]$ with $x_2 \in (x_1,\frac{1-2x_1}{2})$, the absolute difference between the impacts of $\hat{\alpha}$ caused by Δx_1 and Δx_2 would decline along with δ, whereas ascend with it if $x_1 \in (0,\frac{1}{4}]$, $x_2 \in [\frac{1-2x_1}{2},\frac{1}{2})$.

Now, I consider the impacts of Δn to x_1 and x_2. In fact, there are no explicit mathematical expressions to measure the relations of sample size n with x_1 and x_2, therefore I cannot straightly put partial on them because the relations are not continuous. In general, I assume that min and $\widehat{mode}$ are non-increasing functions of n, while max is a non-decreasing function of n, hence x_1 and x_2 are functions of n. Moreover, for simplicity, I assume that $\widehat{skewness} \uparrow skewness$ as $n \to \infty$, which will lead $\widehat{mode} \downarrow mode$ as $n \to \infty$. Denote $\frac{\partial x_1}{\partial n}$ and $\frac{\partial x_2}{\partial n}$ the respective increment of x_1 and x_2 generated by Δn, from (2.7) and (2.8), I have

$$\frac{\partial x_1}{\partial n} = \frac{\widehat{mode}-min-(\Delta\widehat{mode}-\Delta min)}{max-min+\Delta max+\Delta min} - \frac{\widehat{mode}-min}{max-min}, \quad (2.13)$$

$$\frac{\partial x_2}{\partial n} = -\frac{2}{3}(\frac{max-\widehat{mdoe}+\Delta max+\Delta\widehat{mode}}{max-min+\Delta max+\Delta min}) + \frac{2}{3}(\frac{max-\widehat{mode}}{max-min}). \quad (2.14)$$

Therefore the difference between them is

$$\frac{\partial x_1}{\partial n} - \frac{\partial x_2}{\partial n} = \frac{1}{3}(\frac{\widehat{mode}-3min+2max+3\Delta min+2\Delta max-\Delta\widehat{mode}}{max-min+\Delta max+\Delta min} - \frac{\widehat{mode}-3min+2max}{max-min}). \quad (2.15)$$

In order to figure out whether this difference is positive or negative, I use a simple mathematical relation, which is, for $a, b, c, d > 0$,

$$\frac{a+b}{c+d} > \frac{a}{c}, \quad if \quad \frac{a}{c} < \frac{b}{d}, \quad and\ vice\ versa. \quad (2.16)$$

In result, if $\Delta max + \Delta min \neq 0$, from (2.15) and (2.16), it is equivalent to compare

$$\frac{\widehat{mode}-3min+2max}{max-min} \quad with \quad \frac{3\Delta min+2\Delta max-\Delta\widehat{mode}}{\Delta max+\Delta min}. \quad (2.17)$$

Simultaneously the terms in (2.17) could be rewritten as

$$2\left(\frac{max-min}{max-min}\right)+\frac{\widehat{mode}-min}{max-min} \quad and \quad 2\left(\frac{\Delta max+\Delta min}{\Delta max+\Delta min}\right)+\frac{\Delta min-\Delta\widehat{mode}}{\Delta max+\Delta min}. \qquad (2.18)$$

Now I consider whether $\frac{\partial x_1}{\partial n}$ and $\frac{\partial x_2}{\partial n}$ are positive or negative, from (2.13), (2.14) and (2.16), this is equivalent to compare the terms within two pairs

$$\frac{\widehat{mode}-min}{max-min} \quad with \quad \frac{\Delta min-\Delta\widehat{mode}}{\Delta max+\Delta min}, \qquad \frac{max-\widehat{mode}}{max-min} \quad with \quad \frac{\Delta max+\Delta\widehat{mode}}{\Delta max+\Delta min}. \quad (2.19)$$

The truth is, I cannot rigorously prove which side is larger in these pairs due to the absence of corresponding explicit mathematical expressions. But considering this in a practical way, I may set some safe and reasonable assumptions that if $0 < \Delta min < \Delta max$, there are reltaions

$$\frac{\Delta min}{\Delta max+\Delta min} < \frac{\widehat{mode}-min}{max-min} < \frac{1}{2}, \qquad \frac{1}{2} < \frac{max-\widehat{mode}}{max-min} < \frac{\Delta max}{\Delta max+\Delta min}. \qquad (2.20)$$

In fact, regardless of the change in $\widehat{mode}$, for any right skewed parametric population which has continuous density f, mathematically I have $f(min) \downarrow 0$ and $f(max) \downarrow 0$ as $n \to \infty$, because empirical distribution will converge to theoretic population as more samples are gathered in. But it should be noted that, the convergent speed of min and max toward the theoretic minimum and maximum varies, with larger skewness indicating min converging faster than max, for the reason that new samples ought to cluster in greater proportion on the left tail of the empirical distribution. Indeed, it could be interpreted in another way, that is, under a fix sample size, if a Δn triggers the change of min, it should also trigger the change of max, while the opposite is not necessarily true. Therefore, combined with (2.13), (2.14) and (2.19), the extreme situations are

$$\frac{\Delta min}{\Delta max+\Delta min}=0, \ \frac{\Delta max}{\Delta max+\Delta min}=1, \quad for\ \Delta n,\ such\ that\ 0=\Delta min<\Delta max, \qquad (2.21)$$

$$\frac{\partial x_1}{\partial n}=\frac{\partial x_2}{\partial n}=0,$$

$$for\ \Delta n,\ such\ that\ \Delta min=\Delta max=0. \qquad (2.22)$$

In reality, (2.22) is a trivial case, while for Δn which will make both $\Delta min > 0$ and $\Delta max > 0$, which is indeed the case under small sample size, it is hard to compare the terms in the pairs of (2.19). But it may be derived from an unskewed population where theoretically all the terms in (2.20) should approximately equal to $\frac{1}{2}$, so that assumption (2.20) is reliable in a general manner. Moreover, if $\Delta\widehat{mode}$ is considered, undoubtedly I have $\frac{\Delta min-\Delta\widehat{mode}}{\Delta max+\Delta min} < \frac{\Delta min}{\Delta max+\Delta min}$ and $\frac{\Delta max}{\Delta max+\Delta min} < \frac{\Delta max+\Delta\widehat{mdoe}}{\Delta max+\Delta min}$, so that with (2.20), if $0 < \Delta min < \Delta max$, the relations in (2.19) will be

$$\frac{\widehat{mode}-min}{max-min} > \frac{\Delta min-\Delta\widehat{mode}}{\Delta max+\Delta min}, \qquad \frac{max-\widehat{mode}}{max-min} < \frac{\Delta max+\Delta\widehat{mode}}{\Delta max+\Delta min}. \qquad (2.23)$$

Consequently, this indicates $\frac{\partial x_1}{\partial n} < 0$ and $\frac{\partial x_2}{\partial n} < 0$, which means both x_1 and x_2 will decline as more samples are drawn. Additionally, by (2.21), it could be deduced from (2.15) to (2.18) that $\frac{\partial x_1}{\partial n} - \frac{\partial x_2}{\partial n} < 0$, which infers that the decrement $\frac{\partial x_1}{\partial n}$ is greater than the decrement $\frac{\partial x_2}{\partial n}$. Simulations designed as the same as above are conducted to examine these deductions, which are exhibited in Figure 10.

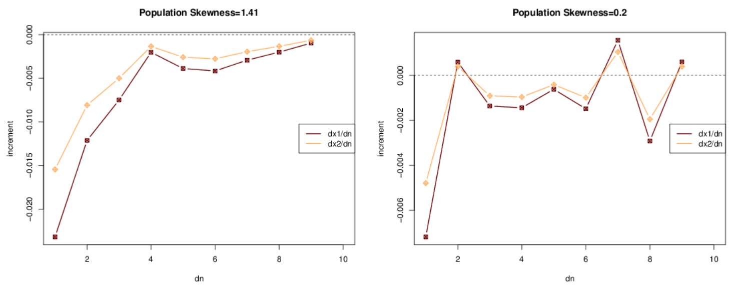


Figure 10: Increment of x_1 and x_2 generated by Δn.

The results in Figure 10 do support our deductions, the trend I depict in the previous discussion is very clear under large population skewness, while some fluctuations occur in the low skewness case, which is intuitively caused by the implementation of $\Delta n = 50$ that might be insufficient to extract the pattern when population is closely unskewed.

Particularly, I use a total differential form to express the impact of Δn to $\hat{\alpha}$, which is

$$d\hat{\alpha} = \frac{\partial\hat{\alpha}}{\partial x_1}\frac{\partial x_1}{\partial n}dn + \frac{\partial\hat{\alpha}}{\partial x_2}\frac{\partial x_2}{\partial n}dn.$$

Hence, with the previous results $\frac{\partial x_1}{\partial n} < 0, \frac{\partial x_2}{\partial n} < 0$ and $\frac{\partial x_1}{\partial n} - \frac{\partial x_2}{\partial n} < 0$, it is apparent that δ will ascend with Δn, together with (2.11), I obtain

$$d\hat{\alpha} = \frac{\partial\hat{\alpha}}{\partial x_1}\frac{\partial x_1}{\partial n}dn + \frac{\partial\hat{\alpha}}{\partial x_2}\frac{\partial x_2}{\partial n}dn < 0, \quad if \quad x_1 \leqslant \frac{1}{4}, \quad 0 < \delta < \frac{1-4x_1}{2}. \qquad (2.24)$$

Meanwhile, if $x_1 > \frac{1}{4}$, (2.24) would still hold because though both $\frac{\partial\hat{\alpha}}{\partial x_1}\frac{\partial x_1}{\partial n}$ and $\frac{\partial\hat{\alpha}}{\partial x_2}\frac{\partial x_2}{\partial n}$ are negative, the absolute discrepancies of $\frac{\partial\hat{\alpha}}{\partial x_1}$ and $\frac{\partial\hat{\alpha}}{\partial x_2}$ would be narrowing according to (2.12), while the discrepancies of $\frac{\partial x_1}{\partial n}$ and $\frac{\partial x_2}{\partial n}$ would consistently enlarge. However, in the case that $x_1 \leqslant \frac{1}{4}$ and $\frac{1-4x_1}{2} \leqslant \delta < \frac{1}{2}$, the general impact of Δn towards $\hat{\alpha}$ would be vague. Because in this case, the above discrepancies would all enlarge and I cannot be certain which one is more significant. But I should also notice that, for populations which have long right tails, x_1 will eventually reach to a very low level, thus widening the interval $[\frac{1-2x_1}{2}, \frac{1}{2})$, together with the fact that x_1 declines faster than x_2, x_2 will gradually lies into $[\frac{1-2x_1}{2}, \frac{1}{2})$, and the case becomes (2.24). So far, from the previous deductions, I believe that in general, $\hat{\alpha}$ in PERT fittings will tend to decrease as more samples are drawn. Moreover, this pattern will still be true for other parametric underlying populations with finite range. In addition, analogous analysis could be implement to unskewed or left skewed populations. Again I need to emphasize that, the above mathematical deductions are not absolute, instead, I only use them as tools in explaining the phenomenons of interest. Though rough, they are still useful in illustrating general patterns.

Some simulations are conducted as examinations for the deductions of the impact of Δn towards $\hat{\alpha}$. The experimental design of increasing sample size of each risk set from 60 to 510 is continuously used, with 10 distinct skewness varies from 1.41 to 0.14. The results shown in Figure 11 provide solid evidence for our analysis, with the phenomenon that $\hat{\alpha}$ decreases with Δn at decelerating speed.

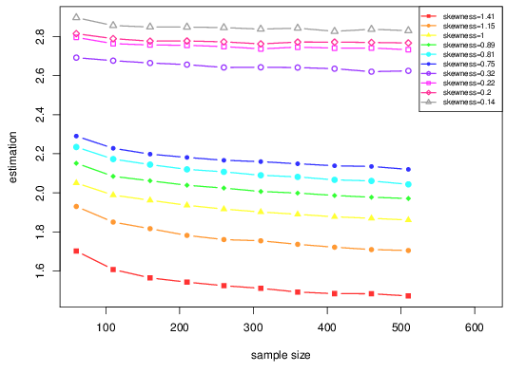


Figure 11: impact of Δn to $\hat{\alpha}$ under different population skewness.

Theoretically, the excess kurtosis (standard kurtosis minus 3) for a $Beta(\alpha,\beta)$ distribution is

$$ex.Kurtosis(Beta(\alpha,\beta)) = \frac{6[(\alpha-\beta)^2(\alpha+\beta+1)-\alpha\beta(\alpha+\beta+2)]}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)}.$$

For PERT distributions with $\hat{\alpha}+\hat{\beta}=6$, it is obvious that excess kurtosis is an increasing function of $\hat{\alpha}$ for $\hat{\alpha} \leqslant 3$ (right skewed cases), while decreasing with it when $\hat{\alpha} > 3$ (left skewed cases). Therefore, under right skewed cases, lower $\hat{\alpha}$ will make PERT distribution present more gentle shape in its tails (technically, with less outliers). In this case, VaR and CTE fittings will increase with more samples are drawn as $\hat{\alpha}$ increasing simultaneously. On the other hand, the empirical distribution will converge to the theoretic one at the same time. In the previous result (Figure 9), I find that empirical densities will drop in tails, and become more concentrate around the $\widehat{mode}$, hence, leading empirical VaR and CTE increasing as well. Therefore, there is no absolute and certain trend of the dynamics of PERT fittings toward empirical VaR and CTE when sample size enlarges. In fact, as sample size approaching to ∞, the reciprocal position of an underlying population and the corresponding PERT fitting will be determined by their theoretic levels (e.g, Figure 7).

However, the most important property is, PERT fitting to the assumed modified Gamma population in VaR and CTE risk measures is very robust under small sample size. I implement some simulations (Figure 12) in illustrating this robustness on an 60×2000 data collection under population assumptions $Gamma(2,25)+303$, $Gamma(100,3.53)$ and a mirror image version of $Gamma(2,25)+303$, which have skewness 1.41, 0.2 and -1,41 respectively. Moreover, I establish two sample statistics "Integrated Discrepancy (ID)" and "Overestimated Ratio (OR)" for this purpose. Denote $VaR^{(i)}$ the ith ordered value of the distribution of VaR, and I_i the indicator function for $I_i=1$, if $VaR^{(i)}_{PERT} \geqslant VaR^{(i)}_{empirical}$, $I_i=0$ otherwise. For VaR (Similarly for CTE), ID and OR are defined as

$$ID_{VaR}(n) = \frac{1}{n}\sum_{i=1}^{n}\left(VaR^{(i)}_{PERT} - VaR^{(i)}_{empirical}\right),$$

$$OR_{VaR}(n) = \frac{\sum_{i=1}^{n} I_i}{n}.$$

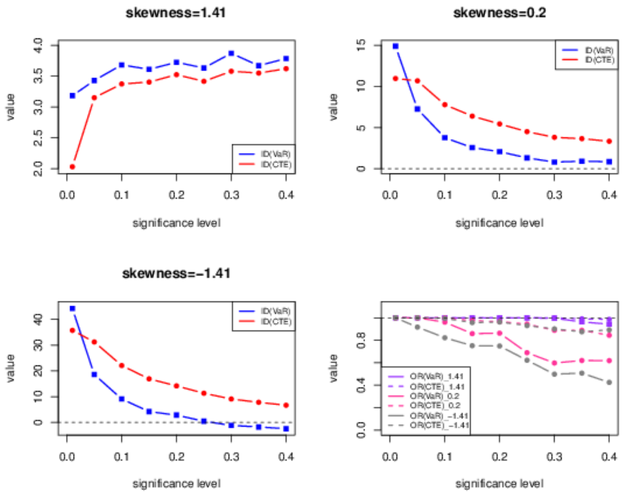


Figure 12: ID and OR statistics under different population skewness.

The results in Figure 12 reveal that under all population skewness, PERT fittings of VaR and CTE presents overestimated patterns toward the empirical with significance level varies from 0.01 to 0.4. The general dynamic behaviors in discrepancies between empirical and fitted concords with the theoretic properties (Figure 7). Moreover, the discrepancies will grow with skewness, which is a natural fact that the left tail distinction of empirical and PERT would be significantly enlarge with descending population skewness. In fact, theses results provide very valuable knowledge, particularly under risk assessment and actuarial context. Because it will overestimate the risk rather than underestimate it, more importantly, PERT fittings in small-sized samples would be rather better than large-sized samples as it is more robust, in this case, always overestimates whenever population skewness varies. This entirely different patterns against theoretic cases (Figure 7) could somehow be regarded as "Skewness free property".

Furthermore, all of the experiments are conducted on several assumed modified Gamma distributions, in this sense, I want to know if the properties I find would hold for wider range of population distributions. In fact, for any given right skewed population distribution, if sample size is relatively low, as depicted previously, samples would insufficiently cluster in the left tail, thus causing empirical left tail density larger than that of PERT, with PERT presents steep shape in this case ($\hat{\alpha}$ is relatively large). Therefore, for any underlying populations with arbitrary skewness, as long as the sample size is relatively low, there must be an interval of significance level (ξ_0, ξ_1) such that if p lies in it, the fitted VaR_p and CTE_p would be larger than the empirical VaR_p and CTE_p in general trend, whereas the interval (ξ_0, ξ_1) depends on the underlying population distribution and the sample size.

In addition, the assumption in PERT that $\hat{\alpha} + \hat{\beta} = k$ would impact the general shape of PERT distributions, with larger k generating more dense sample clusters around $mode$, thus enlarging the kurtosis. However, the choice of k is a quite subjective which often relies on the user's attitude towards the extent of concentration for a concerned variable. So far, there is no standardized method which could be used to adjust this assumption. In practice, the common way is to ask the individual risk takers for their thoughts, and use them to moderate the assumption.

CONCLUDING REMARKS

1. PERT method is very useful in distribution fitting procedure, with very loose requirement for sample information. The underlying assumption of it guarantees its

universality and validness for almost any sample set, and the fitting result is quite stable regardless of the underlying population distribution.

2. PERT method in CTE and VaR fittings would be more effective than straightly implement to a given sample set.
3. The parameter $\hat{\alpha}$ and $\hat{\beta}$ would change as sample size growing, which would lead the distribution consistently becomes more gentle, until min and max stop changing.
4. PERT method is fairly robustness in VaR and CTE fittings under small-sized sample set, for its “skewness free" and “distribution free" properties. In this case, it would be a very useful tool under actuarial context.
5. The uncertainty of $\widehat{mode}$ will cause some mismatching error in PERT fittings, but it could be reduced by the use of some modifications (e.g, using the average patterns of $\widehat{mode}$ in the whole data collection to adjust individual’s estimated $\widehat{mode}$).

References

Malcolm D G , Roseboom J H , Fazar C E C . *Application of a Technique for Research and Development Program Evaluation*[J]. Operations Research, 1959, 7(5):646-669.

Rosenblatt M . *Remarks on Some Nonparametric Estimates of a Density Function*[J]. The Annals of Mathematical Statistics, 1956, 27(3):832-837.

Parzen E . *On Estimation of Probability Density Function and Mode*[J]. The Annals of Mathematical Statistics, 1962, 33(3):1065-1076.

Hall P N , Wand M P . *On the Accuracy of Binned Kernel Density Estimators*[J]. Journal of Multivariate Analysis, 1995, 56(2):165-184.

Piessens R , Branders M . *Modified Clenshaw-CURtis method for the computation of Bessel function integrals*[J]. BIT, 1983, 23(3):370-381.