

An Approximate Method for Expansion of the Cumulative Distribution Function in the Asymptotic Tails

Lyuben D. Ivanov¹ and Justin Y-T. Wu²

Abstract: An efficient approximate method is developed for calculation of very small values of a cumulative distribution function (CDF) or probability of exceedance (POE) located in the asymptotic tails of a continuous distribution. Only three properly selected CDF points are needed to model each of the two tails, while a total of nine points are sufficient to interpolate and extrapolate to cover the entire range of the distribution. The approximated CDF is fast to calculate and is exact at the fitting points while providing smooth transitions from point to point as well as from the end points to the extreme tails. The method is most suitable when sufficient data is unavailable or difficult to compute. In general, the method has three areas of potential applications. The first is the calculation of the very small POE to characterize variables such as long-term extreme loads for risk analysis and design. The second area is calculation of the very small CDF value to characterize variables such as material strength and flaw size. The third case is calculating all ordinates of the entire probabilistic distribution, including CDF and PDF functions. Detailed formulae are provided and the performance demonstrated using commonly used distributions such as Weibull, Gaussian, Lognormal, etc.

Keywords: cumulative distribution function, CDF fitting, probability of exceedance, risk assessment, probabilistic method, reliability method

1 Introduction

In many engineering and scientific applications, very small CDF or POE at the extreme tails of a distribution is important for making risk-based decisions. For example, when designing highly reliable structures, the undesirable rare events are usually associated with the extreme tails of the interested design variables. Unfortunately, it is common that the available statistical data are insufficient to build the CDF or the POE in the asymptotic tails. In such a situation, one could use an

¹ Retired, formerly with American Bureau of Shipping, Houston – USA

² Comprel, Inc. Raleigh, NC 27614

approximate method that could allow for its expansion in those areas. To the authors' knowledge, efforts in this direction started decades ago (Brown and Tukey, 1946). Further, the work of Alfonse, Temple, Filzmoser and Holzer (2010), Anderson (2006), Feldman and Whitt (1998), Fialova and Jureckova (2004) Mandelbrot (1982) should be mentioned related to the "long tail" problem. For solution of this problem, the Pareto Tails Function is applied in different variations. The accuracy of the obtained results is reasonable but one should mention one of the basic requirements for the application of Pareto Tails Function – i.e., one should have at least two points in the area of the long tail. In addition, the methods developed are on a relatively high mathematical level and are not convenient for application in every day engineering work if corresponding computer programs are not available. In the paper, an attempt is made to develop a simple approximate method that does not require specialized computer program but provides reasonable accuracy tested against more accurate methods. It allows for calculating the CDF (or POE) in the area of tails where no data is available (in the proposed method, the data needed for the calculations are taken from the center of the probabilistic distribution). This problem is of primary importance in reliability/risk calculations, e.g. when the probability of exceedance of large ship's hull girder bending moments is to be calculated in the design stage or for an old ship. For aircrafts, 10^{-7} has been commonly referred to as the acceptable threshold for single flight probability of failure (Gallagher et al., 2005; TARAM handbook, 2011; FAA proposed risk assessment guidance, 2011).

Once the implemented error in the proposed method is revealed, it could be applied as a first approximation until more data is available in the probabilistic distribution tails.

2 Basis of the approximate method for expansion of a CDF in the area of very small tail probabilities

Three types of solutions are developed:

1. Expansion of the CDF into very small probabilities of exceedance of high values of the parameter under consideration.
2. Expansion of the CDF into very small probabilities of not exceeding small values of the parameter under consideration.
3. Complete restoration of the probabilistic distribution of the parameter under consideration, i.e., calculating the corresponding probability density function (PDF), cumulative distribution function (CDF) or probability of exceedance (POE).

To facilitate the development of the method, the whole region of possible realizations of the parameter under consideration was split into five sub-regions as shown in Fig. 1 where x stands for any parameter under consideration. The sign \approx is used in order to give some freedom to the user to use values that are close to the ordinates of the CDF (or POE) although not exactly equal to them (i.e., y_1, y_2, y_3 , etc). One should note that for the sake of brevity, the ordinates of the CDF are marked by “ y ” while those of the POE – by “ z ”. A total of nine data points is needed for the entire distributions, while only three points are needed for each of the tails. The derivation of the CDF formulas started from the fifth range, i.e., from the area of large x . Further, in all regions, four boundary conditions were used, the corresponding value of “ y ” or “ z ” at the boundaries and the two first derivatives, to ensure smooth transfer from one region into another one.

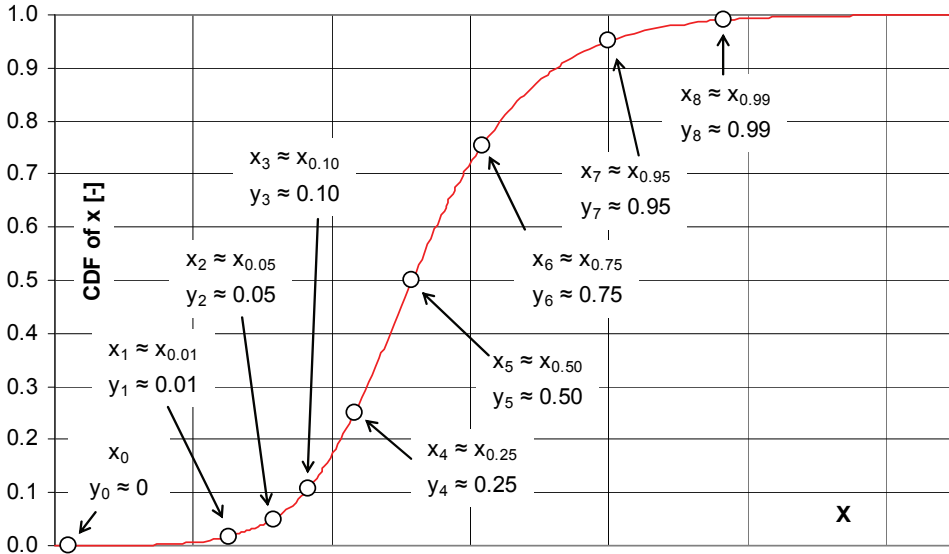


Figure 1: Cumulative distribution function built with eight points 1st region: $x_0 \leq x \leq x_2 \approx x_{0.05}$; 2nd region: $x_{0.05} \approx x_2 \leq x \leq x_3 \approx x_{0.10}$; 3rd region: $x_{0.10} \approx x_3 \leq x \leq x_5 \approx x_{0.50}$; 4th region: $x_{0.50} \approx x_5 \leq x \leq x_7 \approx x_{0.95}$; 5th region: $x_{0.95} \approx x_7 \leq x_{max}$

The proposed CDF model for Region 5 ($x_7 \leq x \leq x_{max}$), is:

$$y_V(x) = 1 - \exp \left\{ \ln(z_8) + \left[\ln(z_8) - \ln(z_7) \right] \frac{x - x_8}{x_8 - x_7} \right\} \quad (1)$$

which approaches unity exponentially for large x . While Eq. (1) has been found to work reasonably well with $y_8 = 0.99$, it also has been observed that for some

distributions, using $y_8 = 0.999$ or even larger values can provide better accuracies for extremely small POEs such as $1.e-08$. In general, it is recommended that the user should conduct a sensitivity study to select the most proper y_i values depending on the specific application requirements. In addition, two other forms for Region 5 have been studied, as presented in the Appendix. These two forms were found to provide excellent tail approximations for distributions close to Gaussian or Rayleigh.

For Region 4 ($x_5 \leq x \leq x_7$), the CDF model has the following nonlinear form with four constants.

$$y_{IV}(x) = 1 - \left(a + \frac{b}{x} + \frac{c}{x^2} + \frac{d}{x^3} \right) \quad (2)$$

For Region 3 ($x_5 \leq x \leq x_7$), the CDF is a cubic polynomial with four coefficients:

$$y_{III}(x) = \alpha_3 + \beta_3 x + \gamma_3 x^2 + \delta_3 x^3 \quad (3)$$

For Region 2 ($x_2 \leq x \leq x_3$), the CDF is also a cubic polynomial with four coefficients:

$$y_{II}(x) = \alpha_2 + \beta_2 x + \gamma_2 x^2 + \delta_2 x^3 \quad (4)$$

For the first region ($x_0 \leq x \leq x_2$), the CDF takes the form of Eq. (5) with one constant.

$$y_I(x) = y_2 \frac{(x - x_0)^{20J_2(x_2 - x_0)}}{(x_2 - x_0)^{20J_2(x_2 - x_0)}} \quad (5)$$

It is worth noting that it may be practically impossible to determine the smallest possible value of x , x_0 , in the first region. In such cases, it is recommended to determine it based on the available knowledge including the physics of the phenomenon under consideration. A simple parametric study could be performed to determine the sensitivity of the CDF or POE when different x_0 values are used.

If determination of the entire probabilistic distribution is not needed, one can use the equations only for the first and fifth region (see the appendix) to calculate, correspondingly, the tail CDF or POE. Except for symmetric probabilistic distributions (e.g., Gaussian distribution), it was found in this study that typically the CDF for the first region is more sensitive than the POE for the fifth region, i.e., even a small change of “ x ” in the first region could cause greater change of the ordinate of the CDF than for the POE in the fifth region (when the same relative change of “ x ” is applied). This is the reason why, in the proposed approach, the expansion of the

POE in the fifth region is calculated using only two percentiles of “x” while for the expansion of the CDF in the first region, three percentiles of “x” are used. One should emphasize that the accuracy of the approximate method also depends on the accuracy of the input data (i.e., the accuracy of the percentiles).

3 Expansion of A CDF into very small probabilities of exceedance of high values of the parameter under consideration

This is probably the most frequently met problem, especially when checking the structure’s strength against permissible stresses (bending moments, shear forces, etc.). Region No. 5 in Fig. 1 corresponds to this case. The calculations only require data for two ordinates of the CDF - y_7 and y_8 (the equations are given in the Appendix). To verify the results obtained by the approximate method, comparison was made against Monte Carlo simulation method (Kalos and Whitlock, 1986). For the calculations, a specialized computer program (Crystal Ball, Oracle Corporation) for Monte Carlo simulation was used. The comparison is shown in Fig. 2. The agreement between the results shown in Fig. 2 is obvious but this is not a proof for the accuracy of the approximate method. Therefore, more calculations were performed to compare the results obtained by the approximate method and results obtained from other (already known) probabilistic distributions, such as exponential, Weibull, Lognormal, etc.). A summary of the calculations is shown in Fig. 3 - Fig. 7. The calculated by Eq. 1 - Eq. 5 POEs are compared with the POEs for exponential, Weibull and lognormal distribution is shown in Fig. 3 - Fig. 5. The calculated by Eq. 1 - Eq. 5 POEs are compared with the POEs for Gaussian and Rayleigh distributions are shown in Fig. 6 and Fig. 7. The comparison is quite promising. One should mention here that for Gaussian and Rayleigh distribution, the effect of different coefficients of variance (COV) on the agreement between the POE obtained by the approximate method and the corresponding equations for the Gaussian and Rayleigh distribution was analyzed by a parametric study.

Based on it, an equation was derived for presentation of the POE for these two cases (see the Appendix). In order to determine the similarity between the original CDF and the Gaussian or Rayleigh CDFs, some of the properties of these two distributions could be utilized. For example, the Gaussian PDF is symmetric, i.e., if the pairs (y_1, y_8) ; (y_2, y_7) ; (y_4, y_6) are the same (or very close), it is very likely that the probabilistic distribution is a Gaussian one. As to the Rayleigh distribution, one can use its peculiar property that the COV is a fixed value (i.e., $COV = 0.523$).

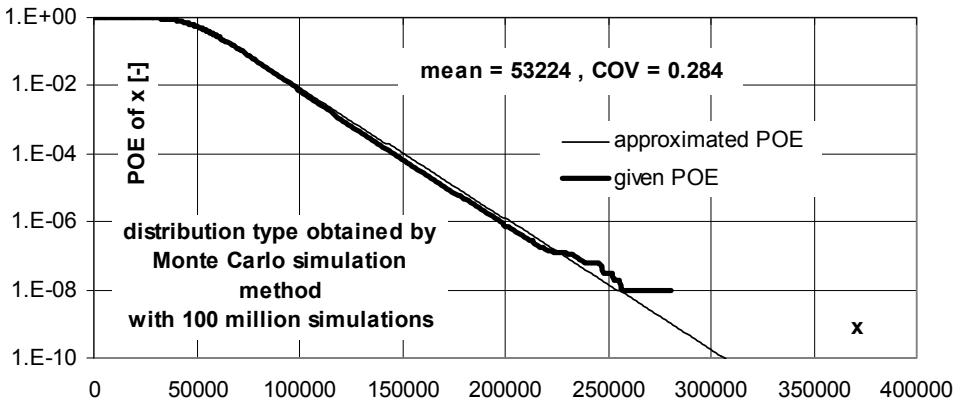


Figure 2: Comparison between the POE of x derived by Monte Carlo simulation (100 million simulations) and the approximate method

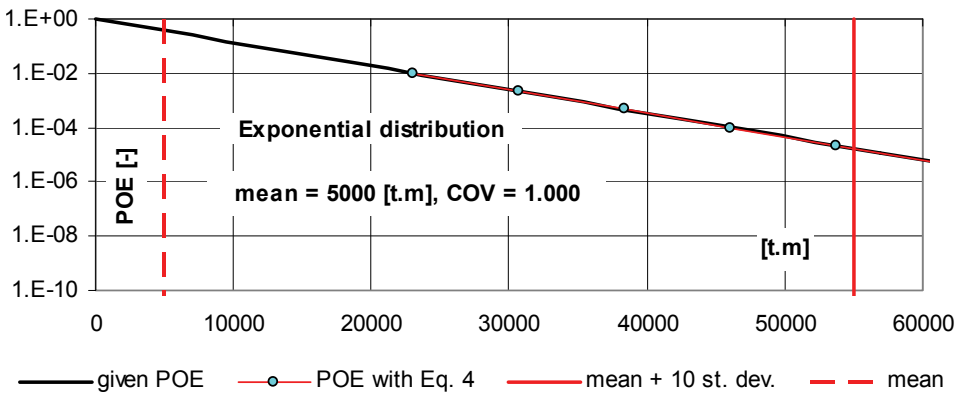


Figure 3: Comparison between the results for the POE obtained by the approximate method and the equations for exponential distribution

4 Expansion of A CDF into very small probabilities of not exceeding small values of the parameter under consideration

As in the previous section, the accuracy of the approximate method was tested against results for the CDF obtained with a specialized computer program for Monte Carlo simulation (see Fig. 8).

Comparison between the results for CDFs obtained by the approximate method and the formulas for known probabilistic distributions is shown in Fig. 9-13.

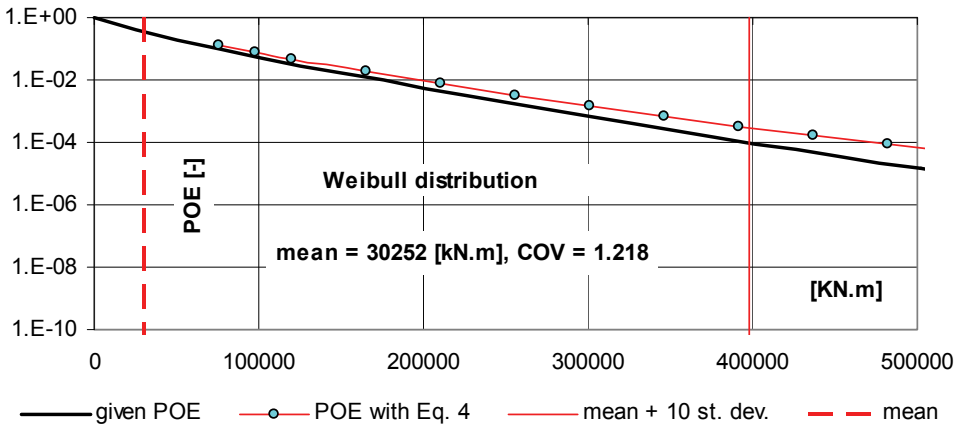


Figure 4: Comparison between the results for the POE obtained by the approximate method and the equations for Weibull distribution

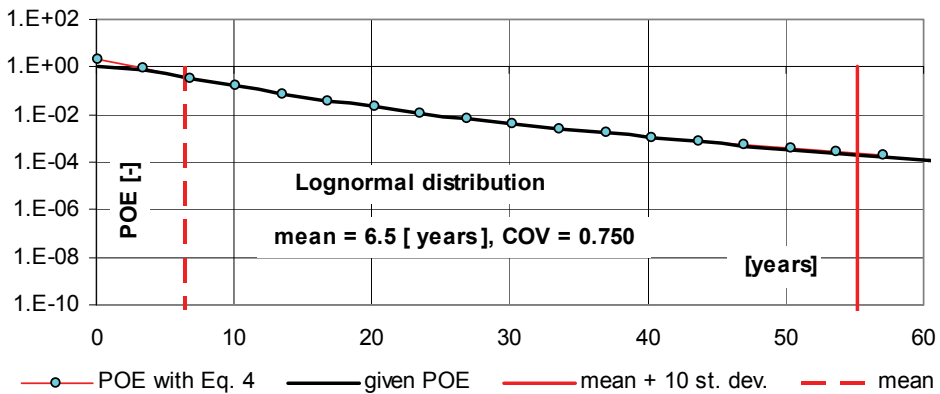


Figure 5: Comparison between the results for the POE obtained by the approximate method and the equations for Lognormal distribution

5 Calculating the ordinates of the complete probabilistic distribution of the parameter under consideration

There are cases, when the whole probabilistic distribution is needed. This can be done by the equations given in the APPENDIX. Their accuracy was tested against results obtained by a specialized computer program for Monte Carlo simulation. Fig. 14 illustrates the similarity between the PDF obtained by the specialized Monte Carlo computer program and the proposed method. One can observe the

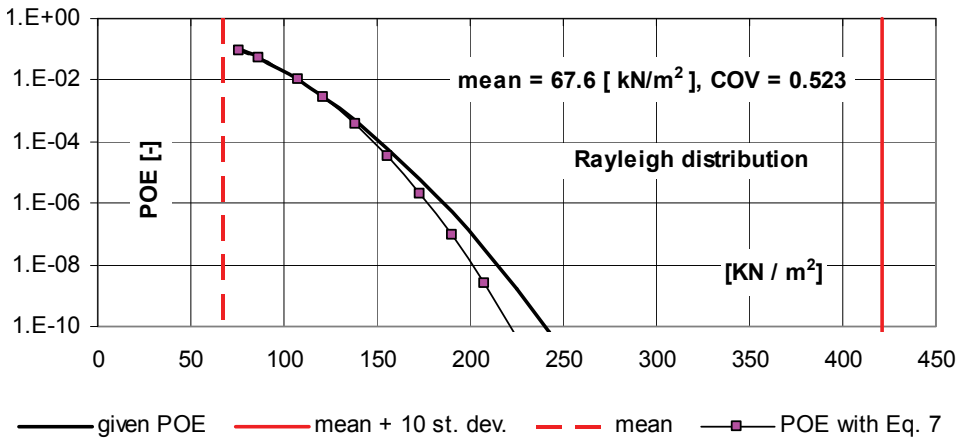


Figure 6: Comparison between the results for the POE obtained by the approximate method and the equations for Rayleigh distribution

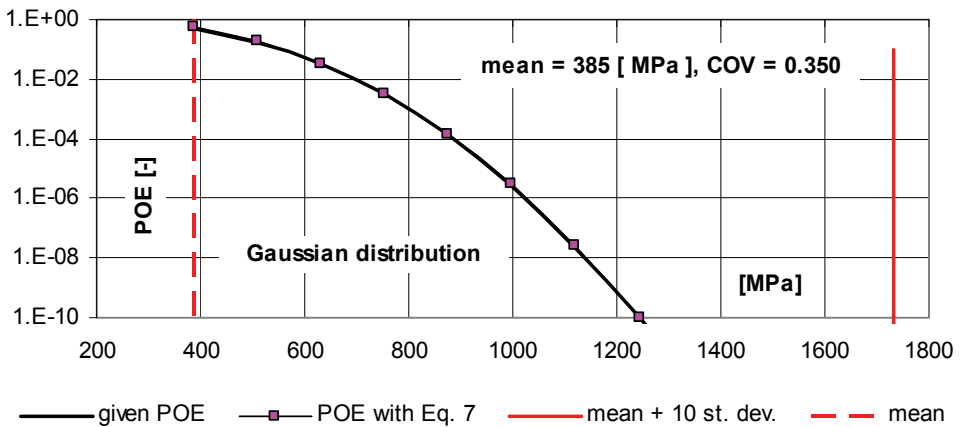


Figure 7: Comparison between the results for the POE obtained by the approximate method and the equations for Gaussian distribution

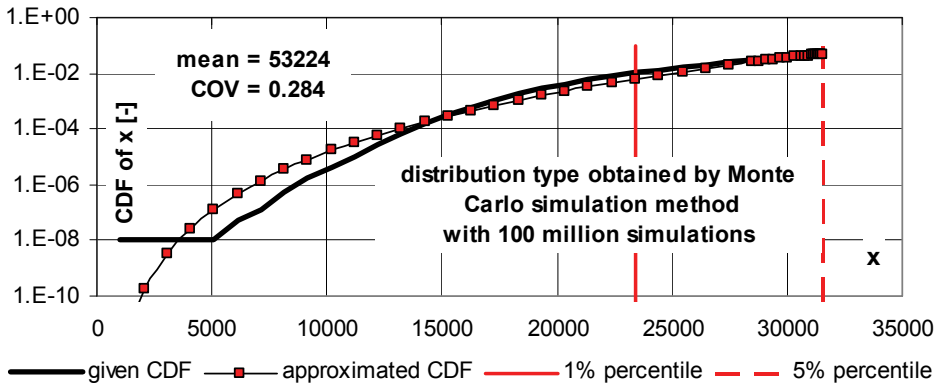


Figure 8: Comparison between the CDF of x derived by Monte Carlo simulation (100 million simulations) and the approximate method (for small values of x)

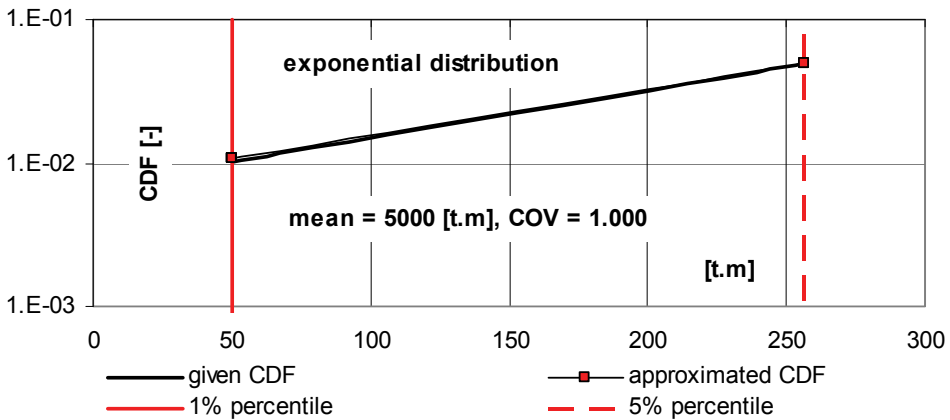


Figure 9: Comparison between the results for the CDF obtained by the approximate method and the equations for exponential distribution

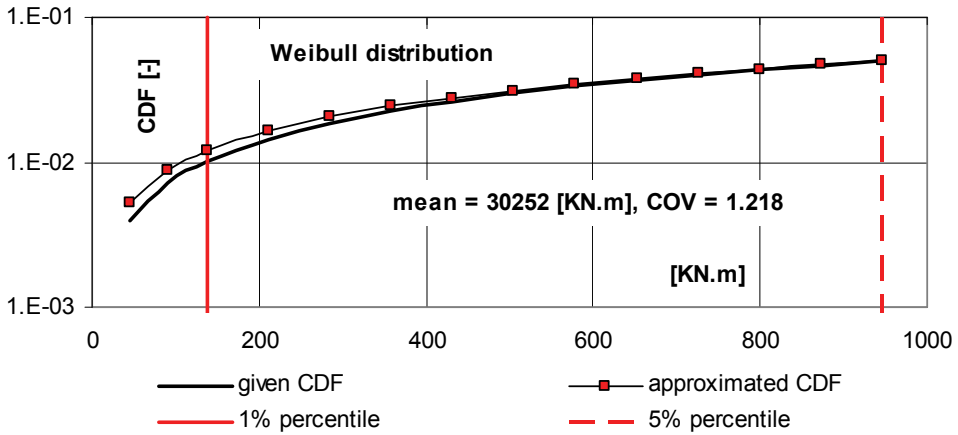


Figure 10: Comparison between the results for the CDF obtained by the approximate method and the Weibull distribution

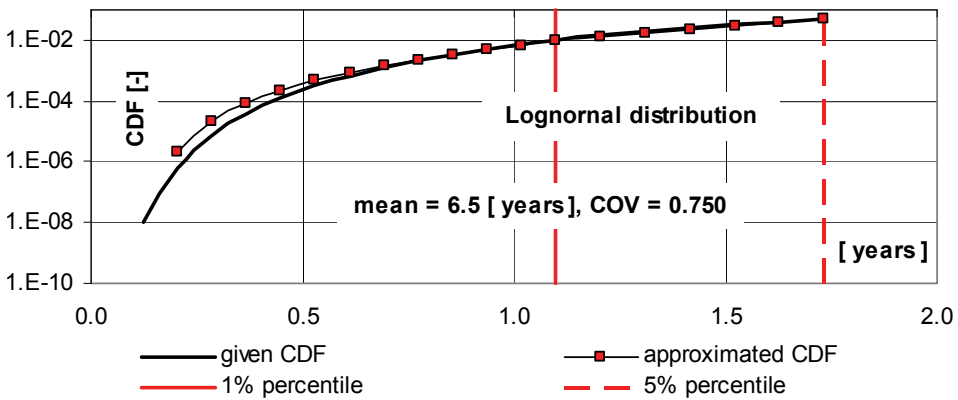


Figure 11: Comparison between the results for the CDF obtained by the approximate method and the equations for Lognormal distribution

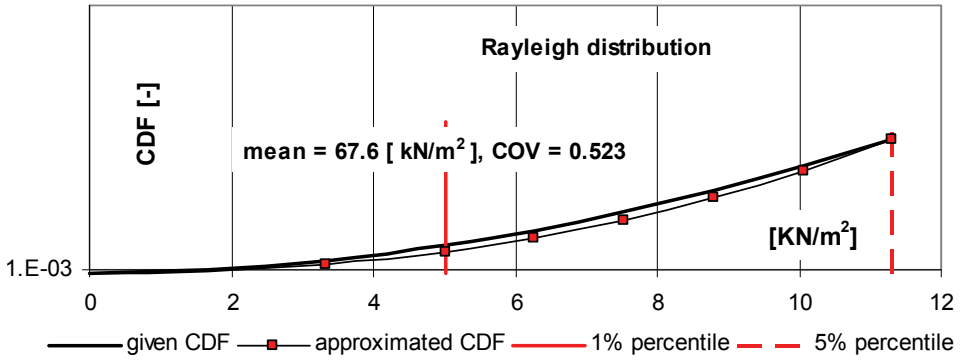


Figure 12: Comparison between the results for the CDF obtained by the approximate method and the equations for Rayleigh distribution

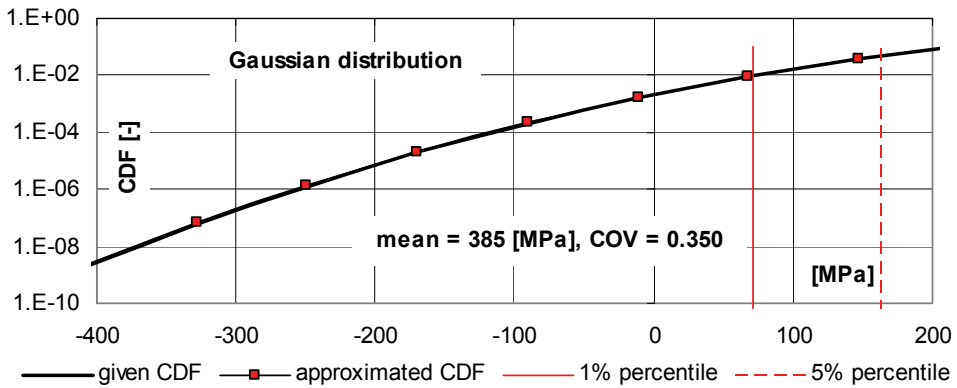


Figure 13: Comparison between the results for the CDF obtained by the approximate method and the equations for Gaussian distribution

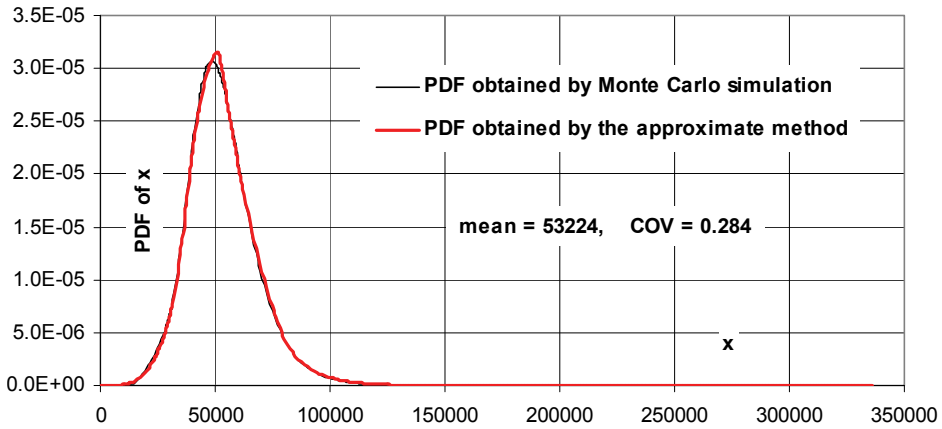


Figure 14: Calculated PDF by the approximate method and by Monte Carlo simulation with 100 million simulations

fact that the results are very close with slight differences in the area of the mode.

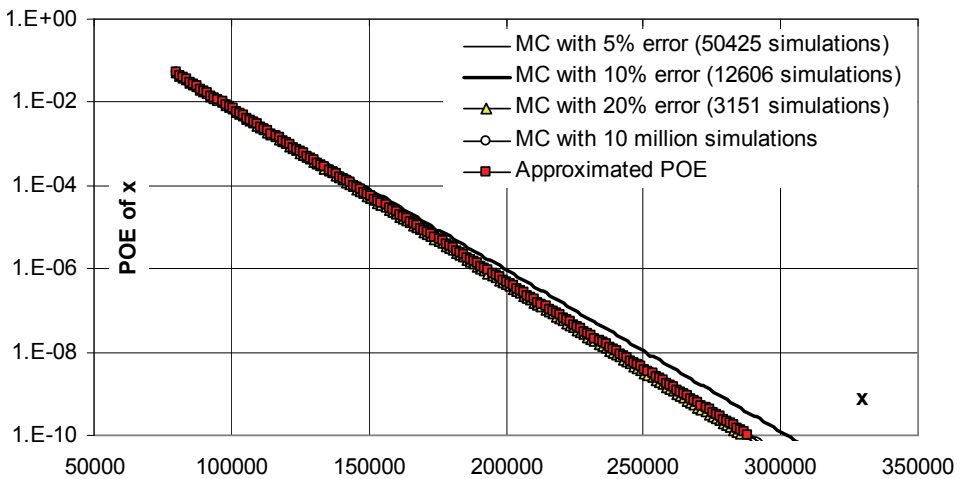


Figure 15: Effect of the accuracy of the input data derived by Standard Monte Carlo method on the POE of very large x

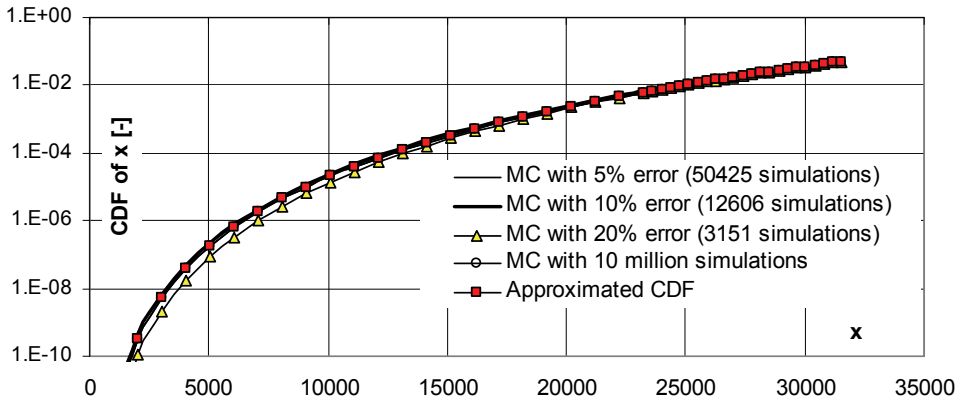


Figure 16: Effect of the accuracy of the input data derived by Standard Monte Carlo method on the CDF of very small x

6 Effect of the accuracy of the input data derived by standard Monte Carlo simulation method on the calculated CDF or POE in the asymptotic tails

To determine the sensitivity of the calculated by the approximate method CDF or POE in the asymptotic tails, a parametric study was performed with a different level of error of the input data. Four levels were selected: 5% error, 10% error, 20% error and a case with almost zero error (after 10 million simulations). The results are illustrated in Fig. 15 and Fig. 16. One can conclude from the graphs that the effect of a different level of error of the input data is not big. From a practical point of view, it means that there is no need to run the Standard Monte Carlo computer program with a very large number of simulations. Even a relatively small amount of simulations (in the example – 3151) could serve the purpose, especially in early design stages when many parameters are not yet accurately determined.

7 Conclusion

An approximate method is proposed for calculating the probability of exceeding a given value (or probability of not reaching it) in the asymptotic tail of a CDF. The accuracy of the method was tested against results obtained by results from more elaborate calculations with the Monte Carlo simulation method and also against known probabilistic distributions (Gaussian, Rayleigh, Weibull, lognormal, exponential). The accuracy of the approximate method is reasonable, especially for the asymptotic tail with large values of the parameter under consideration.

Acknowledgement: The authors would like to thank Prof. Boris Belinskiy from the University of Tennessee in Chattanooga for his help in developing the approximate method.

References

Alfons A, Temple M, Filzmoser P, Holzer J (2010): A comparison of robust methods for Pareto tail modeling in the case of Laeken indicators. Published in *Combining Soft Computing and Statistical Methods in Data Analysis* (editors C. Borgelt, G Gonzalez-Rodriguez, W Trutschnig, M A Lubiano, M A Gil, P Grzegorzewski, and O Hryniewicz), vol. 77 of *Advances in Intelligent and Soft Computing*, pages 17-24. Springer, Heidelberg.

Andersen C (2006): *The Long Tail: Why the Future of Business is Selling Less of More*, New York, Hyperion.

Crystal Ball® - Oracle Corporation, Santa Clara, California 95054

FAA (2011): *Proposed Risk Assessment Guidance*.

Feldman A, Whitt W (1998): Fitting Mixtures of Exponentials to Long-Tail Distributions to Analyze Network Performance Models, *Performance Evaluation*, vol. 31, pp. 245-279.

Fialova A, Jureckova J (2004): Estimating Pareto Tail Index Based on Sample Means, *REVSTAT Statistical Journal*, vol. 2, Number 1, pp. 76 – 100.

Gallagher, J.P., C.A. Babish, and J.C. Malas. (2005): Damage tolerant risk analysis techniques for evaluating the structural integrity of aircraft structures. Proceedings of the *Structural Integrity in Transportation Systems Symposium*, 11th International Conference on Fracture, March, in Turin, Italy.

Kalos M H, Whitlock P A (1986): *Monte Carlo Methods, vol. I: Basics*, John Wiley & Sons.

Mandelbrot B B (1982): *The Fractal Geometry of Nature*, W H Freeman & Co.

Transport Airplane (2011): *Risk Assessment Methodology (TARAM) Handbook*, FAA Transport Airplane Directorate ANM-100, Draft.

Appendix: Equations for Calculation of the CDF, POE or PDF

The full set of data necessary for solving the third problem formulated in Introduction section is given in the Table 1.

Fifth region: valid for $x_{0,95} \approx x_7 \leq x \leq x_{max}$

The following formulae are valid for any distribution type except Gaussian and

Table 1:

x:	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x:	x_0	$x_{0.01}$	$x_{0.05}$	$x_{0.10}$	x_{25}	x_{50}	$x_{0.75}$	$x_{0.95}$	$x_{0.99}$
y:	y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
y:	0.00	0.01	0.05	0.10	0.25	0.50	0.75	0.95	0.99
$z = 1 - y$	z_0	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8
$z = 1 - y$	1.00	0.99	0.95	0.90	0.75	0.50	0.25	0.05	0.01

z = ordinate of the cumulative distribution function; y = ordinate of the probability of exceedance

Rayleigh distribution:

$$POE(x) = z_V(x) = EXP \left\{ LN(z_8) + [LN(z_8) - LN(z_7)] \frac{x - x_8}{x_8 - x_7} \right\} \quad (6)$$

$$CDF(x) = y_V(x) = 1 - z_V(x) \quad PDF(x) = \frac{d[y_V(x)]}{dx} = - \frac{d[z_V(x)]}{dx} \quad (7)$$

$$\frac{d[z_V(x)]}{dx} = EXP \left\{ LN(z_8) + [LN(z_8) - LN(z_7)] \frac{x - x_8}{x_8 - x_7} \right\} \frac{LN(z_8) - LN(z_7)}{x_8 - x_7} \quad (8)$$

The following formulae are valid for Gaussian and Rayleigh distribution

$$POE(x) = z_V(x) = p \cdot EXP(-qx^\alpha) \quad CDF(x) = 1 - POE(x) = 1 - z_V(x) \quad (9)$$

$$PDF(x) = - \frac{d}{dx} [z_V(x)] \quad \frac{d}{dx} [z_V(x)] = -pq\alpha x^{\alpha-1} EXP(-qx^\alpha) \quad (10)$$

$$q = \frac{LN(z_8) - LN(z_7)}{x_7^\alpha - x_8^\alpha} \quad p = \frac{z_7}{EXP(-qx_7^\alpha)} \quad (11)$$

$$\alpha = 1.83 + \frac{0.28}{COV} \quad COV = \frac{\text{st. deviation}}{\text{mean value}}$$

Fourth region: valid for $x_{0.50} \approx x_5 \leq x \leq x_7 \approx x_{0.95}$

$$POE(x) = z_{IV}(x) = a + \frac{b}{x} + \frac{c}{x^2} + \frac{d}{x^3} \quad CDF(x) = 1 - POE(x) = 1 - z_{IV}(x) \quad (12)$$

$$PDF(x) = \frac{d}{dx} [CDF(x)] = - \frac{d}{dx} [z_{IV}(x)] \quad \frac{d}{dx} [z_{IV}(x)] = - \frac{b}{x^2} - \frac{2c}{x^3} - \frac{3d}{x^4} \quad (13)$$

$$J_5 = EXP \left\{ LN(z_8) + [LN(z_8) - LN(z_7)] \frac{x_7 - x_8}{x_8 - x_7} \right\} \frac{LN(z_8) - LN(z_7)}{x_8 - x_7} \quad (14)$$

$$B_4 = z_6 - z_5 + J_5 x_7^2 \left(\frac{1}{x_6} - \frac{1}{x_5} \right) \quad C_4 = \frac{1}{x_5} \left(\frac{2}{x_7} - \frac{1}{x_5} \right) - \frac{1}{x_6} \left(\frac{2}{x_7} - \frac{1}{x_6} \right) \quad (15)$$

$$D_4 = \frac{1}{x_5} \left(\frac{3}{x_7^2} - \frac{1}{x_5^2} \right) - \frac{1}{x_6} \left(\frac{3}{x_7^2} - \frac{1}{x_6^2} \right) \quad (16)$$

$$E_4 = z_7 - z_5 + J_5 x_7 \left(1 - \frac{x_7}{x_5} \right) - \frac{B_4}{C_4} \left[\frac{1}{x_5} \left(\frac{2}{x_7} - \frac{1}{x_5} \right) - \frac{1}{x_7^2} \right] \quad (17)$$

$$F_4 = \frac{1}{x_5} \left(\frac{3}{x_7^2} - \frac{1}{x_5^2} \right) - \frac{2}{x_7^3} - \frac{D_4}{C_4} \left[\frac{1}{x_5} \left(\frac{2}{x_7} - \frac{1}{x_5} \right) - \frac{1}{x_7^2} \right] \quad (18)$$

$$d = \frac{E_4}{F_4} \quad c = \frac{B_4 - d.D_4}{C_4} \quad (19)$$

$$b = - \left[J_5 x_7^2 + \frac{1}{x_7} \left(2c + \frac{3d}{x_7} \right) \right] \quad (20)$$

$$a = z_5 + J_5 \frac{x_7^2}{x_5} + \frac{1}{x_5^2 x_7} \left[c(2x_5 - x_7) + d \frac{3x_5^2 - x_7^2}{x_5 x_7} \right] \quad (21)$$

Third region: valid for $x_{0.10} \approx x_3 \leq x \leq x_5 \approx x_{0.50}$

$$CDF(x) = y_{III}(x) = \alpha_3 + \beta_3 x + \gamma_3 x^2 + \delta_3 x^3 \quad (22)$$

$$POE(x) = 1 - CDF(x) \quad (23)$$

$$PDF(x) = \frac{d}{dx} [CDF(x)] = \frac{d}{dx} [y_{III}(x)] \quad (24)$$

$$\frac{d}{dx} [y_{III}(x)] = \beta_3 + 2\gamma_3 x + 3\delta_3 x^2 \quad (25)$$

$$A_3 = \frac{1}{x_5 - x_4} \left(\frac{y_5 - y_3}{x_5 - x_3} - \frac{y_4 - y_3}{x_4 - x_3} \right) \quad (26)$$

$$B_3 = \frac{x_4(x_4 + x_3) - x_5(x_5 + x_3)}{x_5 - x_4} \quad (27)$$

$$\delta_3 = \frac{A_3(2x_5 - x_4 - x_3) + \frac{y_4 - y_3}{x_4 - x_3} - J_4}{B_3(x_4 + x_3 - 2x_5) + x_4(x_4 + x_3) + x_5^2 - 3x_5^2} \quad (28)$$

$$J_4 = \frac{d}{dx} [y_{IV}(x)] \text{ for } x = x_5 \quad (29)$$

$$\frac{d}{dx} [y_{IV}(x)] = - \frac{d}{dx} [z_{IV}(x)] \text{ for } x = x_5 \quad (30)$$

$$J_4 = \frac{b}{x_5^2} + \frac{2c}{x_5^3} + \frac{3d}{x_5^4} \quad (31)$$

$$\gamma_3 = A_3 + \delta_3 B_3 \quad (32)$$

$$\beta_3 = \frac{y_4 - y_3}{x_4 - x_3} - \gamma_3 (x_4 + x_3) - \delta_3 [x_4 (x_4 + x_3) + x_3^2] \quad (33)$$

$$\alpha_3 = y_3 - x_3 [\beta_3 + x_3 (\gamma_3 + \delta_3 x_3)] \quad (34)$$

Second region: valid for $x_{0.05} \approx x_2 \leq x \leq x_3 \approx x_{0.10}$

$$CDF(x) = y_{II}(x) = \alpha_2 + \beta_2 x + \gamma_2 x^2 + \delta_2 x^3 \quad (35)$$

$$POE(x) = 1 - CDF(x) \quad (36)$$

$$PDF(x) = \frac{d}{dx} [y_{II}(x)] = \beta_2 + 2\gamma_2 x + 3\delta_2 x^2 \quad (37)$$

$$A_2 = \frac{1}{x_3 - x_2} \left(\frac{y_3 - y_1}{x_3 - x_1} - \frac{y_2 - y_1}{x_2 - x_1} \right) \quad (38)$$

$$B_2 = \frac{x_2(x_2 + x_1) - x_3(x_3 + x_1)}{x_3 - x_2} \quad (39)$$

$$\delta_2 = \frac{A_2(2x_3 - x_2 - x_1) + \frac{y_2 - y_1}{x_2 - x_1} - J_3}{B_2(x_2 + x_1 - 2x_3) + x_2(x_2 + x_1) + x_1^2 - 3x_3^2} \quad (40)$$

$$J_3 = \frac{d}{dx} [y_{III}(x)] \text{ for } x = x_3 \quad (41)$$

$$\frac{d}{dx} y_{III}(x) = \beta_3 + 2\gamma_3 x + 3\delta_3 x^2 \quad (42)$$

$$J_3 = \beta_3 + 2\gamma_3 x_3 + 3\delta_3 x_3^2 \quad (43)$$

$$\gamma_2 = A_2 + \delta_2 B_2 \quad (44)$$

$$\beta_2 = \frac{y_2 - y_1}{x_2 - x_1} - \gamma_2 (x_2 + x_1) - \delta_2 [x_2 (x_2 + x_1) + x_1^2] \quad (45)$$

$$\alpha_2 = y_1 - x_1 [\beta_2 + x_1 (\gamma_2 + \delta_2 x_1)] \quad (46)$$

First region: valid for $x_0 \leq x \leq x_2 \approx x_{0.05}$

The following formulae are valid for any distribution type except Gaussian distribution:

$$CDF(x) = y_I(x) = y_2 \frac{(x - x_0)^{20J_2(x_2 - x_0)}}{(x_2 - x_0)^{20J_2(x_2 - x_0)}} \quad (47)$$

$$POE(x) = 1 - CDF(x) \quad (48)$$

$$PDF(x) = \frac{d}{dx} [y_I(x)] = \frac{20J_2(x_2 - x_0)y_2}{(x_2 - x_0)^{20J_2(x_2 - x_0)}} (x - x_0)^{20J_2(x_2 - x_0) - 1} \quad (49)$$

$$J_2 = \frac{d}{dx} [y_{II}(x)] \quad \text{for } x = x_2 \quad (50)$$

$$J_2 = \beta_2 + 2\gamma_2 x_2 + 3\delta_2 x_2^2 \quad (51)$$

$$J_2 = \frac{(y_3 - y_2)}{(x_3 - x_2)} \quad (52)$$

Eq. (52) is performed only for small values of x .

Due to symmetry of the Gaussian distribution, the calculation of the approximate CDF is to be carried out by the formulae for the POE in the fifth region in the following way:

Calculate the ordinates of the POE by the formulae for the fifth region.

For each of the used “ x ” in the calculation of the POE, find the corresponding “ x ” that is symmetrically located relative to the mean value of x , i.e.

$$\text{If } x_{i,f} = x_m + x_f \quad x_{i,a} = x_m - x_f \quad (53)$$

where $x_{i,f}$ = any x in the fifth region; x_m = mean value of x ; x_f = distance from x_m to $x_{i,f}$; $x_{i,a}$ = symmetric abscissa of $x_{i,f}$ when the axis of symmetry passes through the mean value of x .

Using the already calculated ordinates of the POE, build a graph with these ordinates but with the newly calculated symmetric x .

Once the POE is calculated, the derivation of the CDF and PDF can be determined by, e.g., Eq. (6) and Eq. (7).

Notes:

x_0 is to be determined based on the physics of the phenomenon under consideration. The calculations for the whole probabilistic distribution start from the CDF in the fifth region.

For the whole procedure, a simple EXCEL spreadsheet is developed. An example for application of this procedure to results obtained by specialized computer program for Monte Carlo simulation is shown in Fig. 14.