Tail Probability Calculator by Characteristic Function Inversion

[1]T. Duby, [2]G. Wimmer, [3]V. Witkovský

[1]Agilent Technologies Ltd, Yarnton, Oxford, Great Britain,
[2]Institute of Mathematics and Statistics, Masaryk University, Brno, Czech Republic,
Mathematical Institute, Slovak Academy of Sciences, Bratislava, Slovakia,
Faculty of Natural Sciences, Matej Bel University, Banská Bystrica, Slovakia,
[3]Institute of Measurement Science, Slovak Academy of Sciences, Bratislava, Slovakia,
Email: tomy@duby.co.uk

Abstract. *We are developing a computer program to calculate tail probabilities of random variables from their characteristic functions. This code written in Matlab and is an implementation of method described in [1]. We present some practical results. The need for such calculations appears naturally e.g. in metrology: in the problem of calibration [2] and/or in determination of the reference values in interlaboratory comparisons [3]. Currently, as suggested by GUM and its Supplement 1 [4], such problems are typically solved by propagation of distributions using a Monte Carlo method.*

Keywords: Characteristic Function, Inversion of Characteristic Function Tail Probability Calculation

1. Introduction

In advanced statistical analysis there is sometimes the need to deal with linear combination of *N* independent random variables. If these have probability density functions, pdf's, then the pdf of the linear combination is the convolution of the individual pdf's. This may be a computationally intensive task as it requires evaluation of an $N-1$ dimensional integral. And the evaluation of tail probability adds on top of this one more integration. (If $F_X(x)$ is the cumulative distribution function, cdf, of random variable *X*, then its tail probability is defined as $1-F_X(x)$.) Similarly, the Monte Carlo method [4, page 27] is computationally demanding.

Using characteristic functions of random variables to calculate the tail probability eliminates the need for numerical evaluation of an *n*-dimensional integral. The calculation reduces to evaluation of a single integral. Though it involves integrating of an oscillatory integrand over the whole real line, which has its own difficulties, in certain cases this evaluation can be efficiently performed.

In this paper we present the first results obtained from a computer program developed for efficient calculation of tail probabilities of a linear combination of continuous random variables.

2. Subject and Methods

Linear Combination of Random Variables

The following property of characteristic function makes it attractive for above mentioned calculations: Let $X_i, i=1,\ldots,N$ be independent random variables with characteristic functions $\varphi_{X_i}(t)$. Then the characteristic function of random variable Z, $\varphi_Z(t)$, defined as the linear combination of random variables X_i with real constant coefficients a_i, that is

$$Z=\sum_{i=1}^{N} a_i X_i, \tag{1}$$

is given by

$$\varphi_Z(t)=\prod_{i=1}^{N}\varphi_{X_i}(a_i t). \tag{2}$$

Inversion of the Characteristic Function

Over the years a number of methods were developed to obtain the cdf (or tail probability) from the characteristic function. Gil-Pelaez's inversion formula from 1951, [5] looks simple

$$F_X(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\Im\left(e^{-itx}\varphi_X(t)\right)}{t}\, dt \tag{3}$$

where $\Im(\cdot)$ is the imaginary part. As the above integral is an improper one, its numerical evaluation is the source of two kinds of errors: error due to truncation, ϵ_t, and error due numerical approximation of the truncated integral, ϵ_a. Experience has shown that extended (or compound) trapezoidal rule "tends to work well for oscillating integrals because the errors tend to cancel" [3, page 18]. This is the formula for extended trapezoidal rule

$$\int_a^b f(x)dx = h\left[\frac{f(a)}{2} + f(a+h) + \cdots f(a+(N-1)h) + \frac{f(b)}{2}\right] + \epsilon_a \tag{4}$$

where N is the number of intervals into which the original (finite) integration range is subdivided, $h = (b-a)/N$ is the step size. The error due to numerical approximation is

$$e_a = -\frac{(b-a)\, h^2}{12} f''(\xi),\ \ a < \xi < b, \tag{5}$$

provided the second derivative of function f exist.

R. Strawderman, [1] applied the theory of F. Stenger to evaluate of the integral in equation (3): Under certain conditions [7] when the integrand is analytic in a strip containing the path of integration, the error due to numerical approximation decreases exponentially with the step size h. Specifically, when function $g(z)$ is analytic on strip $D = \{z \in \mathbb{C}\colon |\Im(z)| < \pi/2\}$ then

$$\int_{-\infty}^{+\infty} g(x)dx = h\sum_{m=-\infty}^{+\infty} g(mh) + \epsilon_a \tag{6}$$ [1]

where

$$|\epsilon_a| < N_D(g)e^{-\frac{\pi^2}{h}} \tag{7}$$

and $N_D(g)$ is some number depending on function g [7, equation 1.8]. This method is applicable for those random variables for which the moment generating function exists on an analytic strip $[c, d]$, where $-\infty \le c < 0 < d \le +\infty$.

Method

We are developing a computer program called TPC for **T**ail **P**robability **C**alculator for efficient and high precision calculation of cumulative distribution function, cdf (or tail probability) of arbitrary continuous random variable. It is written in Matlab [8] as its *anonymous function* feature allows the user to his / hers function and pass it name as an argument.

TPC requires the following mandatory inputs:

1. The characteristic function of the random variable as an anonymous function
2. Values of x (scalar of vector) for which to calculate the cdf / tail probability.

The following inputs are optional

1. Width of the analytic strip as a vector of length 2. If the user does not provide it the program attempts to find it.
2. Expected accuracy of the tail probability calculation. Default is 10^{-8}.

[1] Note that though the integral in the left hand side is an improper one we cannot avoid the truncation error as the summation in the right hand side needs to be truncated.

The simplified flowchart of the TCP program is shown in Figure 1

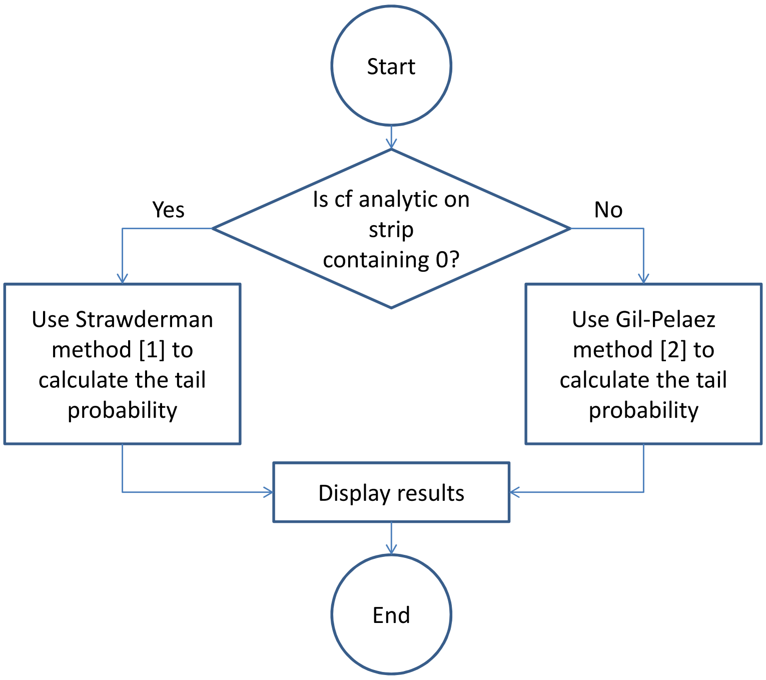

Figure 1: Overall flowchart of the TPC program.

3. Results

χ_1^2 distribution turned is a good case for testing this program: (1) its cdf is known in a closed formula, (2) its analytic strip is $[-\infty, 1/2]$, and (3) its characteristic function decays to zero as $t^{-\frac{1}{2}}$. Numerical results are presented in Table 1.

Table 1: Results of cumulative distribution function calculations of χ_1^2.

x	cdf of $\chi_1^2(x)$	Absolute error	No. of integrand subintervals
3	0.916735483336450	-4.4409e-16	243
10	0.998434597741998	-1.1102e-16	89
15	0.999892488823271	-1.1102e-16	58

A more complicated example can be found in [4, section 9.2.3]. Here the random variable X is the sum of four uniform random variables of unit standard deviation. In language of Matlab the characteristic function of X is defined as the anonymous function `fun923`:

```
bb = sqrt(3);
fun923 = @(t)cf_uniform(t,-bb,bb).^4;
```

where `cf_uniform(t,a,b)` is the author's Matlab function for cf of uniform distribution with limits `a` and `b`.[2] For this distribution the 95% coverage is known analytically [4, Appen-

[2] The standard deviation of uniform distribution with limits $[-a, a]$ is $\sigma = a/\sqrt{3}$.

dix E]. TPC gives for $x = 2\sqrt{3}\left[2 - \left(\frac{3}{5}\right)^{\frac{1}{4}}\right] = 3.879406741347821$ the tail probability is 0.025000000000001 giving a coverage of 94.9999999999998%.

4. Discussion

The computer program presented here based on [1] provides an efficient method to calculate tail probabilities by inversion of characteristic functions. For the sake of completeness the completed program will include methods that will calculate the tail probability in cases when assumptions of [1] do not hold: the random variable does not have a moment generating function.

Acknowledgements

The work was supported by the Slovak Research and Development Agency, grant APVV-0096-10.

References

[1] Strawderman R. Computing tail probabilities by numerical Fourier inversion: the absolutely continuous case. *Statistica Sinica*, 14: 175–201, 2004.

[2] Wimmer G, and Witkovský V. New procedure for calculating the uncertainty of one output quantity in calibration certificates. In *MEASUREMENT 2013, Proceedings of the 9th International Conference on Measurement 2013*. Smolenice, Slovakia, May 27−30, 2013. Submitted.

[3] Witkovský V, and Wimmer G. Method for evaluation of the key comparison reference value and its expanded uncertainty based on metrological approach. In *MEASUREMENT 2007, Proceedings of the 6th International Conference on Measurement 2007*. Smolenice, Slovakia, May 20−24, 2007, 26−29.

[4] Evaluation of measurement data — Supplement 1 to the “Guide to the expression of uncertainty in measurement” — Propagation of distributions using a Monte Carlo method, JCGM 101:2008.

[5] Gil-Pelaez J. Note on the inversion theorem, Biometrika, 38 (3/4): 481–482, 1951.

[6] Abate J and Whitt W, The Fourier-series method for inverting transforms of probability functions. *Queuing Systems*, 10: 5-88, 1992.

[7] Stenger F. Integration Formulae Based on the Trapezoidal Formula, *J. Inst. Maths Appliecs* 12: 103−114 1973.

[8] MathWorks, Inc., Natick, Massachusetts, U.S.A.