

## ARITHMETIC OPERATIONS ON INDEPENDENT RANDOM VARIABLES: A NUMERICAL APPROACH\*

SZYMON JAROSZEWICZ<sup>†</sup> AND MARCIN KORZEŃ<sup>‡</sup>

**Abstract.** Dealing with imprecise quantities is an important problem in scientific computation. Model parameters are known only approximately, typically in the form of probability density functions. Unfortunately there are currently no methods of taking uncertain parameters into account which would at the same time be easy to apply and highly accurate. An important special case is operations on independent random variables which occur frequently in obtaining confidence intervals for physical measurements and statistical estimators. In this paper we investigate the possibility of implementing arithmetic operations on independent random variables numerically. Equivalently, the problem can be viewed as propagating approximated probability density functions through arithmetic operations. We introduce a very broad family of distributions which is closed under the four arithmetic operations and taking powers. Furthermore we show that the densities of the distributions in the family can be effectively approximated using Chebyshev polynomials. A practical implementation is also provided, demonstrating the feasibility and usefulness of the approach. Several examples show applications in physical measurements, statistics, and probability theory, demonstrating very high numerical accuracy. These include an interesting problem related to combining independent measurements of a physical quantity, distributions of sample statistics of the Hill's estimator for tail exponents, generalized  $\chi^2$  distribution, and others. The results are usually extremely accurate, in some cases more accurate than specialized solutions available in statistical packages, and can be achieved with very little effort.

**Key words.** uncertainty propagation, arithmetic of random variables, numerical approximation

**AMS subject classifications.** 60-08, 65C20, 41A10, 26A15

**DOI.** 10.1137/110839680

**1. Introduction.** Arithmetic operations on independent random variables (i.r.v.'s) occur frequently in scientific computation. For example, obtaining confidence intervals for physical measurements often requires combining several imprecise values in order to obtain the desired quantity. Applications also arise frequently in probability theory and statistics, where most parametric tests are based on such distributions. For example, distributions of means of independent and identically distributed (i.i.d.) samples are obtained by adding the population's distribution several times and dividing the result by a constant. The Student's  $t$ -statistic is obtained by dividing a normally distributed random variable (r.v.) by a  $\chi^2$  distributed r.v. Such distributions are sometimes easy to compute (typically for normal populations), but in most cases the problem is intractable, and no closed form solutions exist.

The literature contains solutions for various special cases, but even if such solutions exist for the problem at hand, they can be exceedingly hard to apply. Take, for example, the noncentral  $t$ -distribution used to compute the power of the Student's  $t$ -test. Various formulas for its density are available in the literature but require infinite

---

\*Submitted to the journal's Methods and Algorithms for Scientific Computing section July 5, 2011; accepted for publication (in revised form) February 14, 2012; published electronically May 1, 2012. This work was supported by research grant N N516 068537 of the Polish Ministry of Science and Higher Education (Ministerstwo Nauki i Szkolnictwa Wyższego) from research funds for the period 2009–2011.

<http://www.siam.org/journals/sisc/34-3/83968.html>

<sup>†</sup>Institute of Computer Science, Polish Academy of Sciences, 01-248 Warsaw, Poland. Current address: National Institute of Telecommunications, 04-894 Warsaw, Poland (s.jaroszewicz@itl.waw.pl).

<sup>‡</sup>Faculty of Computer Science and Information Technology, West Pomeranian University of Technology, Zolnierska 49, 71-210 Szczecin, Poland (mkorzen@wi.zut.edu.pl).

summations or evaluation of confluent hypergeometric functions, posing severe numerical problems. Other applications include obtaining confidence intervals for physical measurements, where various quantities with known (or assumed) error distributions are combined using arithmetic operations to obtain the quantity of interest.

We propose a numerical solution to the problem: a software library called PaCAL (**P**rob**A**bilistic **C**ALculator), freely available for download at <http://pacal.sf.net>, which allows for computing with probability distributions as one does with ordinary variables. The distributions are represented using piecewise Chebyshev approximations. When arithmetic operations are performed, the approximation of the result is automatically computed and can be used in subsequent operations. Section 4 contains several illustrative examples which demonstrate the potential of the library.

We also analyze the theoretical aspects of such an implementation by addressing the problem of whether the results of arithmetic operations on approximable distributions remain approximable. More specifically, we present families which contain practically all distributions used in practice and which are closed under the four arithmetic operations and taking powers. Moreover, we show that distributions in those families can be effectively approximated using piecewise polynomial representations. We demonstrate that arithmetic with r.v.'s following distributions outside of our families can result in pathological densities which cannot be approximated; see Example 1.1.

In order to represent probability distributions numerically, several problems need to be addressed. The most important are infinite interpolation intervals and singularities (consider, for example, the  $\chi^2$  distribution with 1 degree of freedom). Our families include distributions with infinite supports and a finite number of singularities; effective methods of their approximation are described.

High accuracy of the computations and supporting an as wide as possible class of distributions have been the top priorities for us while developing PaCAL. The speed has not been neglected, but, together with the independence assumption, limits the application of the package to large scale problems. We demonstrate, however, a number of nontrivial practical applications where PaCAL gives very accurate results in a matter of seconds.

**1.1. Related work.** The most important work on arithmetic of r.v.'s is [19]. It describes a general approach to obtaining the resulting distributions analytically by means of Fourier and Mellin transforms. An algorithmic approach is also proposed with distributions represented using  $H$ -functions, a generalization of hypergeometric functions. Computing product and quotient distributions reduces to simple arithmetic operations on the parameters of those functions. Unfortunately, the family is not closed under addition. Another drawback is the difficulty in computing the values of  $H$ -functions numerically. The method proposed for obtaining actual densities involves series expansions based on moments, which is often not satisfactory. For example, it will not work unless the distribution has several finite moments. In fact, any reasonable family which includes only distributions with finite moments cannot be closed under division (the quotient of two normally distributed i.r.v.'s is a Cauchy r.v., which has no moments).

Newer work on arithmetic of r.v.'s can be found in [22], where approximations using Laguerre polynomials are used. Formulas for sums and products of i.r.v.'s in such a representation are given. However, as entire densities are approximated with a single expansion, the method is unlikely to handle discontinuities well. No theoretical results are provided on the closure properties of the representation. Later

chapters of [22] (and works by other authors) use so-called envelopes, that is, upper and lower bounds on the cumulative distribution functions (c.d.f.'s). Such methods have the advantage of working with dependent variables, but they offer only rough approximations of the distributions; the bounds become very loose after repeated operations. Our goal is, instead, to provide as accurate as possible results for the practically important case of independence.

In [12], the authors describe a library for arithmetic operations on i.r.v.'s based on symbolic computations using Maple. The functionality of this approach is very similar to that of PaCAL, but the methodology is very different. Symbolic computations are capable of providing exact results, but it is not always possible to express the resulting distributions symbolically.

Some work on arithmetic operations on r.v.'s can be found in the literature on measuring physical quantities. Typically, if errors are relatively small, normality assumption and the delta method are used; for larger errors Monte Carlo simulation combined with histogramming is recommended; see, e.g., [14]. Simulation can handle arbitrary arithmetic operations on (even dependent) r.v.'s, but the accuracy grows only with the square root of the number of samples taken, which means that usually only a few correct decimal digits can be obtained. The problem becomes much more severe in the tails of distributions.

As far as we know, the closeness of our proposed families of distributions under all arithmetic operations and its relation to approximability have not been published before. Some of the theoretical results we introduce (such as tail behavior) were already known for convolutions [9], but the distributions of products and quotients of r.v.'s are much less popular in the literature, and, to the best of our knowledge, most results presented here are either new or stronger than the previous ones. For example, [5] requires the distributions to have all finite moments, while our results are valid even for distributions with no moments. As far as we can tell, the general analysis of the behavior of singularities of probability density functions (p.d.f.'s) is also new and, except for specific cases, has not been previously discussed.

In this work we represent p.d.f.'s using piecewise Chebyshev interpolation with variable transformations to allow for infinite intervals and singularities. Our representation works for a very broad class of distributions including those for which no moments exist. The advantages of the Chebyshev representation are the existence of automatic approximation methods and excellent accuracy on well-behaved functions. Below we discuss other possible representations and compare them with our choice.

Probability density functions are frequently approximated using asymptotic expansions such as the Edgeworth series [16]. The series, however, is not guaranteed to converge even for common distributions, and, like other methods based on moments (or cumulants), the representation is not closed under division. Another approach involves spectral representations such as characteristic functions and Mellin transforms [9, 19], which have the property that arithmetic operations on i.r.v.'s reduce to standard arithmetic on spectral representations. Those methods are not suitable for a general purpose library, as spectral representations of even simple distributions are hard to represent numerically. Take, for example, the uniform distribution on  $[-1, 1]$ , trivially expressible as a piecewise polynomial, whose characteristic function is  $\sin(x)/x$ , which, due to its infinite periodic pattern, is difficult to approximate numerically.

One may also choose to represent c.d.f.'s, not p.d.f.'s. Using c.d.f.'s has some

advantages, such as being able to automatically handle points with nonzero measure; however, computing the results of arithmetic operations still requires p.d.f.'s which would have to be obtained by differentiation. Differentiation of Chebyshev polynomials (and other representations) is easy but reduces numerical accuracy.

Finally, we need to mention the main motivation for this work: the Chebfun package [1], which implements arithmetic on arbitrary functions by representing them as piecewise Chebyshev polynomials. While PaCAL borrows the main ideas from Chebfun, the contribution of this paper is still significant. By exploiting the specifics of p.d.f.'s, we are able to obtain an implementation which is more efficient and handles infinite intervals and singularities better than Chebfun. Moreover, we give theoretical guarantees which are impossible in the case of arbitrary functions.

**1.2. Preliminaries and notation.** We consider continuous r.v.'s, by which we mean real valued r.v.'s which possess p.d.f.'s. We do not deal with pathological cases such as the Cantor distribution or other singular distributions.

Let  $f, g$  be p.d.f.'s of i.r.v.'s  $X, Y$ , respectively. We consider r.v.'s  $X + Y, X - Y, X \cdot Y, X/Y$  formed by arithmetic operations on  $X$  and  $Y$ . The corresponding operations on their p.d.f.'s will be denoted, respectively, by  $f \oplus g, f \ominus g, f \odot g, f \oslash g$ . Note that, in the literature, the density of the sum is typically denoted using the convolution symbol  $*$ . We do not use this notation to maintain uniformity over all operations. The expressions  $f + g$  and  $f - g$  denote the pointwise sum and difference of real valued functions, and  $f^{(n)}$  denotes the  $n$ th derivative of  $f$ .

*Arithmetic operations on r.v.'s.* Let  $f$  and  $g$  be p.d.f.'s. The densities resulting from arithmetic operations on i.r.v.'s following those distributions are given by [19]

$$(1.1) \quad (f \oplus g)(t) = \int_{-\infty}^{+\infty} f(x)g(t-x) dx,$$

$$(1.2) \quad (f \ominus g)(t) = \int_{-\infty}^{+\infty} f(x)g(x-t) dx,$$

$$(1.3) \quad (f \odot g)(t) = \int_{-\infty}^{+\infty} f(x)g\left(\frac{t}{x}\right) \frac{1}{|x|} dx,$$

$$(1.4) \quad (f \oslash g)(t) = \int_{-\infty}^{+\infty} f(xt)g(x)|x| dx = \int_{-\infty}^{+\infty} f(x)g\left(\frac{x}{t}\right) \frac{|x|}{t^2} dx.$$

If  $u$  is a function consisting of  $n$  strictly monotone and differentiable pieces  $u_i$ , the distribution of  $u(X)$ , where  $X$  is an r.v. with density  $f$ , is given by [19]

$$(1.5) \quad p(x) = \sum_{i=1}^n f(u_i^{-1}(x)) \left| (u_i^{-1})^{(1)}(x) \right|,$$

where  $u^{-1}$  denotes the inverse of  $u$ , and, in agreement with our notation,  $u^{(1)}$  denotes the first derivative of  $u$ .

In the next section we present families of approximable distributions which are closed under arithmetic operations. The following example shows that approximability of the results should not be taken for granted, and that operations on seemingly benign distributions can lead to pathological results.

*Example 1.1* (quotient of approximable r.v.'s may not be approximable). Let the p.d.f.  $f$  be an infinite series of rectangular impulses centered at  $4^0, 4^1, \dots, 4^j, \dots$ , each of width 1 and height  $\frac{1}{2^{j+1}}$ . The function is approximable by a piecewise polynomial

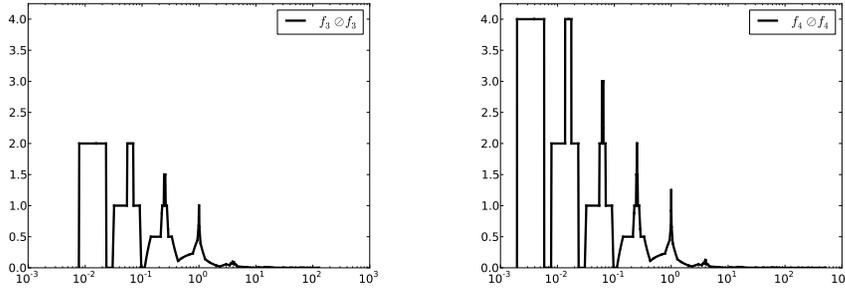


FIG. 1.1. Illustration for Example 1.1 for  $N = 3$  and  $N = 4$ . Logarithmic scale is used on the  $x$  axis.

function  $f_N$ , consisting of the first  $N$  impulses, with an absolute approximation error of  $\frac{1}{2^{N+1}}$ .

$f \circledast f$  has, however, an infinite number of singularities at points  $\frac{1}{4^{-k}}$ ,  $k = 0, 1, 2, \dots$ , and clearly cannot be approximated in the  $\|\cdot\|_\infty$  norm using piecewise polynomial functions. Indeed, consider the approximation  $f_N$  and let  $t = 4^{-k}$  be fixed:

$$\begin{aligned} (f_N \circledast f_N)(4^{-k}) &= \int f_N(x)f_N(x \cdot 4^{-k})|x| \, dx = \sum_{j=k}^{N-1} f_N(4^j)f_N(4^{j-k}) \cdot 4^j \\ &= \sum_{j=k}^{N-1} \frac{1}{2^{j+1}} \cdot \frac{1}{2^{j-k+1}} \cdot 4^j \\ &= \sum_{j=k}^{N-1} 2^{k-2} = (N - k)2^{k-2} \rightarrow \infty, \text{ as } N \rightarrow \infty. \end{aligned}$$

To see the second equality, notice that the integral is equal to a sum over overlapping impulses; the first two factors are the height of the resulting impulse and  $4^j$ , the value of  $x$  at its center. Figure 1.1, generated using PaCAL, provides a graphical illustration for  $N = 3, 4$ .

The rest of the paper is organized as follows. Section 2 presents the families of approximable distributions closed under arithmetic operations. Section 3 briefly discusses details of our implementation, and section 4 demonstrates its capabilities. Section 5 offers concluding remarks, and the appendix contains the proofs of theorems.

**2. Families of distributions closed under arithmetic operations.** We require p.d.f.'s and a specified number of their derivatives to be bounded and continuous except at a finite number of points at which they are *allowed* to diverge to infinity. We will call such points *potential singularities* of the density function. Note that the p.d.f. (or some of its derivatives) is not required to have a singularity at such a point; the only requirement is that outside those points the p.d.f. and a specified number of its derivatives must be continuous (and consequently bounded outside some arbitrary small neighborhoods of the potential singularities).

We now introduce the families  $\mathcal{F}_k$ ,  $k \in \mathbb{N}$ , of p.d.f.'s. We say that a p.d.f.  $f$  belongs to  $\mathcal{F}_k$  if there exists an  $n \in \mathbb{N}$  and points  $p_1, \dots, p_n \in \mathbb{R}$  (called *potential*

singularities of  $f$ ) such that the following assumptions hold for all  $0 \leq i \leq k$ :

A.1  $f^{(i)}$  is continuous everywhere except at the potential singularities  $p_1, \dots, p_n$ .

A.2  $|f^{(i)}(x)| = O\left(\frac{1}{|x-p_i|^{i+1}}\right)$  as  $x \rightarrow p_i$ , for every potential singularity  $p_i$  of  $f$ .

A.3  $|f^{(i)}(x)| = O\left(\frac{1}{|x|^{i+1}}\right)$  as  $x \rightarrow \pm\infty$ .

One of the main results of this paper, the following theorem, states that the families  $\mathcal{F}_k$  are closed with respect to all arithmetic operations and under taking powers of r.v.'s.

**THEOREM 2.1.** *Let  $f \in \mathcal{F}_k$  and  $g \in \mathcal{F}_k$  be p.d.f.'s with potential singularities  $p_1, \dots, p_n$  and  $q_1, \dots, q_m$ , respectively. Then*

1.  $f \oplus g \in \mathcal{F}_k$  with potential singularities at  $p_i + q_j$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ ;
2.  $f \ominus g \in \mathcal{F}_k$  with potential singularities at  $p_i - q_j$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ ;
3.  $f \odot g \in \mathcal{F}_k$  with potential singularities at 0 and at  $p_i q_j$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ ;
4.  $f \oslash g \in \mathcal{F}_k$  with potential singularities at 0 and at  $p_i/q_j$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ ,  $q_j \neq 0$ .

**THEOREM 2.2.** *Let  $r \in \mathbb{R}$ ,  $r \neq 0$ , and let  $X$  be an r.v. with p.d.f.  $f \in \mathcal{F}_k$  with potential singularities at  $p_1, \dots, p_n$ . Then the p.d.f. of  $X^r$  also belongs to  $\mathcal{F}_k$ , with potential singularities at 0 and at  $p_i^r$  for each  $p_i \neq 0$ . For  $r \notin \mathbb{Z}$  we assume that  $f(x) = 0$  for all  $x < 0$ .*

The proofs of these theorems and those of the remaining theorems can be found in the appendix. It is clear from Theorem 2.1 that the potential singularity at zero has a special meaning for multiplication and division: the sum and difference of bounded p.d.f.'s remain bounded, but products and quotients can introduce a singularity at zero (consider, e.g., the product of two independent normal r.v.'s). In fact, it is clear that p.d.f.'s having a single potential singularity at zero form useful subfamilies of  $\mathcal{F}_k$  closed under the four arithmetic operations and taking powers.

Assumptions A.1–A.3 are not severe, and the families are indeed very broad. Practically all standard distributions used in statistics belong to  $\mathcal{F}_\infty$ , including heavy tailed distributions such as Lévy or Pareto which do not have finite moments. The following result, showing that all asymptotically  $k$ -monotone distributions belong to  $\mathcal{F}$ , gives one justification of this fact.

**THEOREM 2.3.** *If for  $0 \leq i \leq k$ ,  $f^{(i)}(x)$  is monotone for  $x \geq M > 0$  and for  $x \leq -M$ ,  $M \in \mathbb{R}$ , then  $f$  satisfies assumption A.3 for  $k$ .*

An analogous result can be proved for monotonicity around singularities. Intuitively, the density needs to contain an infinite number of “bumps” (and the bumps need to be sufficiently high) if it is not to belong to  $\mathcal{F}_k$ .

It is also clear that  $\mathcal{F}_k$  is closed under linear transformations; if the p.d.f. of  $X$  is in  $\mathcal{F}_k$ , so is the p.d.f. of  $aX + b$  for  $a, b \in \mathbb{R}$ ,  $a \neq 0$ .

A note is needed here on how repeated computations affect the number of singularities in the density functions of their results. Theorem 2.1 implies that the number of singularities can increase with each operation. However, this rarely happens in practice. In the case of addition and subtraction, the singularities in the result become less steep with each operation and eventually disappear completely (typically after just one or two steps)<sup>1</sup> due to the smoothing properties of convolution. The same is true for multiplication and division, except for the singularity at zero which becomes steeper with each subsequent operation. Such a case is presented in section 4 in the example involving the product of uniformly distributed variables.

<sup>1</sup>For example, the p.d.f. of the sum of two independent  $\chi^2$  r.v.'s with 1 degree of freedom has no singularity, even though the p.d.f.'s of both arguments do.

**2.1. Approximation of probability density functions.** Additionally, it turns out that the p.d.f.'s belonging to  $\mathcal{F}_k$  can be effectively approximated. We first define a family  $\mathcal{F}_A$  of functions which will be used to approximate the p.d.f.'s; in short, we will use polynomials combined with transforms to handle infinite intervals. More formally, a function  $f$  belongs to the family  $\mathcal{F}_A$  if there exist  $a_1, \dots, a_n \in \mathbb{R}$ ,  $n \geq 1$ , and an  $\alpha \in \mathbb{N}$ ,  $\alpha \geq 1$ , such that

1.  $f(x) = P_0((-x + 1 + a_1)^{-1/\alpha})$  for  $x \leq a_1$ ,
2.  $f(x) = P_n((x + 1 - a_n)^{-1/\alpha})$  for  $x > a_n$ ,
3.  $f(x) = P_i(x)$  for  $x \in (a_i, a_{i+1}]$  for every  $1 \leq i < n$ ,

where  $P_i$  are polynomials. The expressions in points 1 and 2 come from the use of variable transform  $x = 1/u^\alpha$ ,  $\alpha \geq 1$ , which maps the interval  $(0, 1]$  onto  $[1, \infty)$ .

**THEOREM 2.4.** *Let  $f \in \mathcal{F}_k$  with potential singularities  $p_1, \dots, p_n$  for some  $k \geq 0$ . Then*

1. *for every  $\epsilon > 0$  and  $\delta > 0$  there exists a function  $\tilde{f} \in \mathcal{F}_A$  such that  $|f(x) - \tilde{f}(x)| \leq \epsilon$  for all  $x$  such that  $\min_j |x - p_j| > \delta$ ,*
2. *if  $k > 0$ , then  $P_i$  can be obtained using Chebyshev approximations,*
3. *if  $k > 1$ , then, in addition, the error of the Chebyshev approximation decreases as  $O(N^{-\min(\alpha-1, k)+1})$ , where  $N$  is the maximum degree of polynomials  $P_i$  used in the definition of  $\tilde{f}$ , and  $\alpha \geq 1$  is the constant used in the definition of  $\mathcal{F}_A$ .*

In other words, densities in  $\mathcal{F}_k$  can be approximated with polynomials (using transforms for infinite intervals) except for arbitrarily small neighborhoods around potential singularities. The larger the value of  $k$ , the easier the approximation. For  $k = 0$ , the theorem guarantees the existence of an approximation, but there is no guarantee on the rate of convergence, and Chebyshev expansions cannot be used to obtain it. If  $k > 0$ , efficient and numerically stable Chebyshev approximations are possible, and when  $k > 1$ , the convergence rate can be guaranteed, with the approximation error decreasing as  $O(N^{-k+1})$  for a sufficiently large value of  $\alpha$ . In practice, the distributions are defined using functions which are piecewise infinitely differentiable, so the convergence is very fast.

Let us comment here on approximation of functions with singularities. In practice one uses tricks which make this approximation easier. For example, one might approximate the function  $|x - p_i|f$ , which may not have a singularity, and multiply the approximation by  $1/|x - p_i|$  to regain the behavior around  $p_i$ . While such methods are useful in practice, the guarantees offered on the absolute error are not any stronger than those given in the theorem above. The handling of singularities in our implementation is described in the next section.

**3. A practical implementation.** In this section we describe a practical implementation of the proposed methodology. The library called **PaCAL (ProbAbilistic CALculator)** has been implemented in Python and is freely available at <http://pacal.sf.net>. We now briefly discuss implementation details such as the approximation and integration procedures used; full details are beyond the scope of this paper. Section 4 shows example applications of the package.

Overall, arithmetic operations on densities  $f$  and  $g$  are performed as follows. First, potential singularities of the result are obtained using Theorem 2.1. They determine the intervals on which the result will be approximated. For the approximation procedure to work, it suffices that the result be evaluated at a finite number of points. Each such evaluation is performed by directly applying numerical integration to one of the formulas (1.1)–(1.4). The two key components of the package are thus the

approximation and integration routines which will be described in the remaining part of this section.

*Approximation.* The distributions are represented as piecewise polynomial approximations of their p.d.f.'s. Each approximated piece of the function is called a *segment*. The endpoints of segments are placed at potential singularities and at plus and minus infinity. Without loss of generality we assume that each segment belongs to one of the following three kinds:

1. a segment on a closed interval with no singularities,
2. a segment on a closed interval with a singularity at one of its endpoints,
3. a segment with no singularities and with one of the endpoints at plus or minus infinity.

Segments of the first kind, which involve no singularities and no infinite endpoints, can be approximated directly, using Chebyshev approximations. It turns out that such an approximation can very easily be achieved using the so-called barycentric interpolation formula [2]

$$\tilde{f}(x) = \frac{\sum_{i=0}^N \frac{w_i}{x - x_i} f_i}{\sum_{i=0}^N \frac{w_i}{x - x_i}},$$

where  $x_i$ 's are interpolation points (nodes),  $f_i$  is the value of the interpolated function at  $x_i$ , and  $w_i$ 's are weights. We used the Chebyshev nodes of the first or second kind for which the weights  $w_i$  have a simple form [2]. In this case, the formula is equivalent to performing Chebyshev approximation using other, more expensive, methods. This approach is efficient and very stable numerically. In principle any other approximation method can be used, such as Fourier series.

*Infinite intervals.* Theorem 2.4 in section 2.1 gives a solution for approximating functions on infinite intervals; using the variable transform  $1/x^\alpha$ ,  $\alpha \in \mathbb{N}$ ,  $\alpha \geq 1$  guarantees that the error decreases as  $O(N^{-\min(\alpha-1, k)+1})$ , where  $N$  is the degree of the polynomial. In the literature [4], the value of  $\alpha = 1$  is typically used, but Theorem 2.4 shows that higher values can indeed be beneficial. In our implementation we use  $\alpha = 6$ . Additionally, in the tails of the distribution we approximate  $\log f(\exp(x))$ , the log-log transform of the p.d.f., in order to get good asymptotic behavior for very high values of  $x$ .

*Singularities.* Each of the approximated p.d.f.'s has associated with it a set of potential singularities. Outside of this set the function is well behaved, but at those points its behavior can be problematic.

First we need to determine whether a potential singularity is indeed a singularity and, if this is the case, take measures to represent it as accurately as possible. Unfortunately, we cannot determine in advance whether singularities in arguments' p.d.f.'s give rise to a singularity in the result of an arithmetic operation;<sup>2</sup> we thus try to determine the presence of a singularity numerically, by computing limits. The method we use to represent singularities works well even if there is no singularity at a given point, so the consequences of wrongly classifying a point as a singularity are not severe; typically only a small loss of accuracy occurs.

When a segment is believed to have a singularity at its left endpoint we approximate the log-log transform of  $f$  on an interval  $[p_i + \delta, b]$ , where  $p_i$  is the potential singularity (segment's left endpoint) and  $\delta$  is a small constant. We use  $\delta = 10^{-50}$

---

<sup>2</sup>The sum of two i.r.v.'s with p.d.f.'s  $0.25x^{-0.75}$  on  $[0, 1]$  has a singularity at zero, while the sum of two  $\chi^2$  distributions with 1 degree of freedom does not.

for  $p_i = 0$  and larger values otherwise. Singularities at the right endpoint are handled analogously. Note that this approach is quite different from what is typically suggested in the literature,<sup>3</sup> but it proved more robust in our case.

*Integration.* Numerical integration plays a central role in the implementation. It is used in (1.1)–(1.4) and in computation of moments and c.d.f.’s.

For finite intervals we use Clenshaw–Curtis and Fejér quadratures as described in [20, 21]. For integrals on infinite intervals and for integrals of functions involving singularities we use variable transformations to change the integral into a proper integral on a finite interval. Unfortunately, we cannot guarantee that the integrals can be computed for all densities in  $\mathcal{F}_k$ ; thus, even though we guarantee that the results of arithmetic operations can be approximated, we cannot guarantee that we can always obtain those approximations. However, this turns out to be possible in almost all practically useful cases.

It is easy to see that for integrals on semi-infinite intervals of functions such that  $f = O(1/x^\beta)$  as  $x \rightarrow \infty$ ,  $\beta > 1$ , the change of variable  $t = x^\alpha$ , where  $\alpha > 1/(\beta - 1)$ , leads to proper integrals with respect to  $t$ . Similarly, for improper integrals of a function with a singularity at zero such that  $f = O(1/x^\beta)$  as  $x \rightarrow 0$ ,  $0 < \beta < 1$ , the substitution  $t = x^\alpha$ , where  $\alpha > 1/(1 - \beta)$ , leads to a proper integral. Details can be found in [8, sections 2.12 and 3.1] and [3].

In our implementation, we use  $\alpha = 6$  for both infinite intervals and singularities, which is sufficient for most heavy tailed distributions like Lévy, Student’s  $t$ , or Pareto distribution with a broad range of parameters, as well as for distributions with “heavy singularities” like the  $\chi^2$  with 1 degree of freedom. For heavier tails the exponent needs to be increased.

*Error estimation.* A question arises regarding how to measure the accuracy of the results. Numerical interpolation and integration methods allow for error estimation, and such estimates are available in PaCAL. Our error estimation method for interpolation is based on comparing the interpolated values with function values provided to the interpolator at each new interpolation point; for integration we use the standard approach of taking the difference between successive approximations as the error estimate [18, 8].

However, we found that a very useful estimate of the accuracy is the integral of the density over the whole real line (the “zeroth” moment). Since we take no explicit steps to guarantee that the densities integrate to 1, the quantity  $|1 - \int_{-\infty}^{\infty} f|$  is a useful indicator of accuracy which takes into account the accumulation of error over subsequent operations. We call this quantity the  $\|\cdot\|_1$  error. We noticed that the  $\|\cdot\|_1$  error is a good indicator of the order of magnitude of other types of errors such as the  $\|\cdot\|_\infty$  error (see below) or the error in computation of the quantiles. Moreover, this type of error better corresponds to statistical applications, where quantities involving integrals of the density function (such as mean, median, and values of the c.d.f.) are more important than the density function itself.

If, in a given case, the resulting p.d.f. has a closed form expression, we report the maximum absolute difference, on a densely sampled set of points, between the expression and the density computed by PaCAL. We call this difference the  $\|\cdot\|_\infty$  error. Sometimes, statistics of the result, such as the mean or variance, are available and can be compared with values computed numerically. Since the mean and variance are computed using numerical integration, their accurate values are strong evidence

---

<sup>3</sup>Namely, recursively splitting the approximation interval or factoring the singularity out of the function.

that the density itself is also accurate. A rough verification is always possible using sampling and histograms.

**4. Experiments and examples.** In this section we present examples of how the library is used, some potential applications, and an experimental verification of its accuracy. Let us begin with a short example session:

```
from pacal import *
C = NormalDistr(0,1) / NormalDistr(0,1)
P = C * C; Q = C / C
P.plot()
show()
```

The code divides two normally distributed r.v.'s obtaining a Cauchy r.v. assigned to a variable  $C$ . Then the quotient  $Q$  and the product  $P$  of two Cauchy r.v.'s are computed<sup>4</sup> and their p.d.f.'s plotted. The result is shown in the upper half of Figure 4.1; note the singularity at zero. The distribution also features heavy tails. The lower half of the figure shows the difference between p.d.f.'s of  $P$  and  $Q$ , which should ideally be zero, as the product and quotient of Cauchy r.v.'s have the same distributions [19]. The difference is always below  $10^{-15}$  and often below  $10^{-16}$ . The  $\|\cdot\|_1$  error of  $C$  is about  $6 \cdot 10^{-16}$  and for  $P$  and  $Q$  in the range of  $10^{-13}$ . We can now compute various parameters of the distributions. For example, `P.cdf(-1)` and `Q.cdf(-1)` both yield the c.d.f. of the distribution at  $-1$  using numerical integration. The returned value (0.25) is accurate to 13 decimal places. We now show some examples which demonstrate the accuracy of the package on standard distributions used in the probability theory.

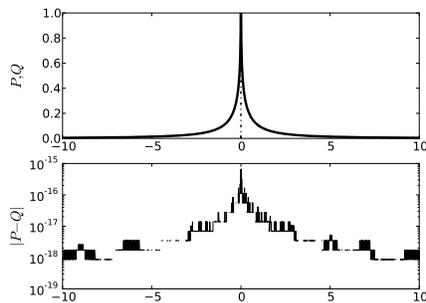


FIG. 4.1. *P.d.f.'s of the product and quotient of two independent Cauchy r.v.'s (computed as a quotient of independent normal r.v.'s). Difference between the two p.d.f.'s.*

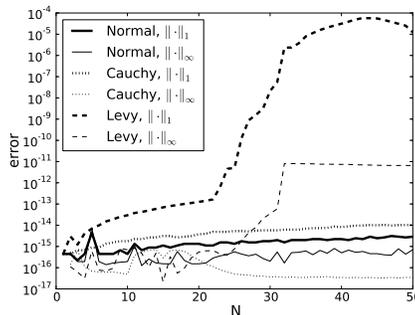


FIG. 4.2. *Errors of computed sums of  $N$  i.i.d. r.v.'s following some stable distributions.*

*Stable distributions.* We start with sums of stable distributions for which the closed forms of the results are known so we can compute the  $\|\cdot\|_\infty$  error. Figure 4.2 shows the accuracy for sums of  $N$  i.i.d. r.v.'s for three stable distributions: normal, Cauchy, and Lévy. The sums are generated by the following piece of code:

```
S_n = NormalDistr()
for i in range(N-1):
    S_n += NormalDistr()
```

<sup>4</sup>Independence is always assumed. For example, in the expression  $C * C$ , both uses of  $C$  are treated as independent r.v.'s.

In the first line, `NormalDistr` can be replaced by `CauchyDistr` or `LevyDistr` as appropriate. The accuracy remains very high for the normal and Cauchy distributions. For Lévy, the  $\|\cdot\|_1$  accuracy remains high for  $N$  up to around 20. For larger values, it degrades due to the very heavy tails of the distribution ( $O(x^{-1.5})$ ). However, even for  $N = 50$ , the  $\|\cdot\|_1$  error still corresponds to four correct decimal digits. The  $\|\cdot\|_\infty$  error is also quite low. It can also be seen that the  $\|\cdot\|_1$  error is indeed a good general indicator of accuracy. Computing the sum of 50 terms took 59.94 seconds for the Normal distribution, 74.61 seconds for the Cauchy distribution, and 33.15 seconds for the Lévy distribution.

*Sum of uniformly distributed r.v.'s.* Often, even when closed form expressions are available, they are numerically unstable. Take the sum of  $N$  i.r.v.'s uniformly distributed on  $[0, 1]$ . The closed form expression for the p.d.f. is given in [19] but is numerically unstable. For  $N = 50$ , the values obtained from the closed form expression blow up to the range of  $10^{10}$  (using double precision arithmetic), while PaCAL remains accurate and gives the mean of 25 and variance of 50/12, both with 13 correct decimal digits. Figure 4.3 shows how various types of error, including the error estimates provided by PaCAL, change with the number of terms in the sum. Exact values needed for computation of the  $\|\cdot\|_\infty$  error were obtained from the closed form solution using extended precision arithmetic. It is clear that the numerical error of the explicit formula grows exponentially, while PaCAL remains accurate. It can also be seen that PaCAL provides reasonable estimates of the  $\|\cdot\|_\infty$  error and, again, that the  $\|\cdot\|_1$  error is a good overall estimate of accuracy.

We have also compared the results with Mathematica, which allows for symbolic computation of convolutions. For the sum of 16 terms, it took about 190 seconds to obtain the closed form solution identical to the one given in [19] and used in our experiment. Numerical evaluation of the expression gave, of course, the same, exponentially growing, errors. For comparison, PaCAL took only 1.17 seconds to compute the sum for 16 terms, and about 9.78 seconds for 50 terms. Of course, in most cases closed form expressions do not exist, and the symbolic approach cannot be used at all.

*Product of uniformly distributed r.v.'s.* Another experiment demonstrating the accuracy of PaCAL involves the product of  $N$  i.r.v.'s uniformly distributed on  $[0, 1]$ . The closed form solution is  $\log^{N-1}(1/x)/(N-1)!$  and has a *logarithmic* singularity at zero, the steepness of which increases with  $N$ . For example, for  $N = 20$ , the range between zero and about  $2.87 \cdot 10^{-9}$  contains half of the total probability mass, while the support of the density function is  $[0, 1]$ . Such steep singularities typically pose severe numerical difficulties. Figure 4.4 shows the errors for the products of up to 20 uniformly distributed i.r.v.'s computed using PaCAL. While the relative  $\|\cdot\|_\infty$  error grows, several correct digits are available even for the difficult case of  $N = 20$ . Again, the error estimates are correct, if somewhat conservative, and the  $\|\cdot\|_1$  error is a good overall indicator. The computation took about 54.1 seconds.

*Product of two normal r.v.'s.* Figure 4.5 shows the product of two normally distributed r.v.'s, with variances equal to 1 and means, respectively, to 1, 2, and 3. All resulting distributions have singularities at zero, but for factors with higher means, the singularities contain very little probability mass (see the inset in the third chart). Nevertheless, PaCAL correctly detected the singularities and interpolated them accurately. Histograms are included in the figure for comparison.

**4.1. Physical measurements.** In this section we present applications to physical measurements, which frequently involve arithmetic operations on i.r.v.'s.

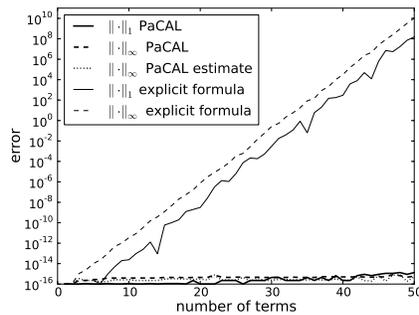


FIG. 4.3. Computation errors for the sum of  $N$  independent r.v.'s uniformly distributed on  $[0, 1]$ . The numerical errors for an explicit formula are also included.

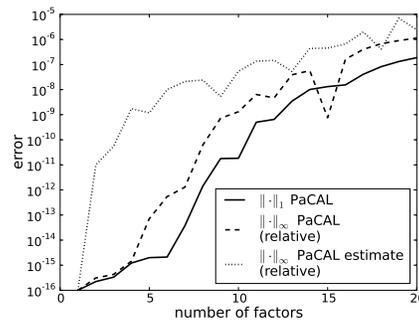


FIG. 4.4. Computation errors for the product of  $N$  i.r.v.'s uniformly distributed on  $[0, 1]$ .

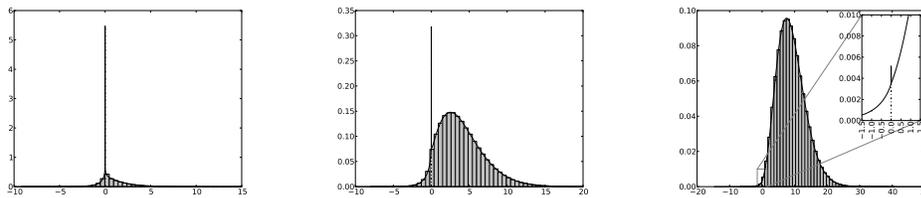


FIG. 4.5. Product of two independent normal r.v.'s with variances equal to 1 and means, respectively, to 1, 2, and 3. Note the singularities at zero.

*The coefficient of thermal expansion.* Consider measuring the coefficient of thermal expansion  $K$  by measuring the initial length  $L_a$  of an element, its final length  $L_b$ , and the temperature change  $\Delta T$ . All quantities are measured with some error: lengths with uniform error (e.g., using a digital meter) and the temperature with normal error (random noise is present in the measuring device). Suppose that  $L_a$  is uniformly distributed on the interval  $[9.0, 10.0]$ ,  $L_b$  uniformly distributed on  $[11.0, 12.0]$ , and  $\Delta T$  normally distributed with mean 2 and standard deviation of 1. The noise in the temperature measurement is thus large. The distribution of  $K = (L_b - L_a)/(L_a \Delta T) = (L_b/L_a - 1)/\Delta T$  (the expression is restructured to avoid statistical dependency between the numerator and the denominator) is shown in Figure 4.6. The distribution has a rather unexpected shape, very different from what a normal approximation might produce. The  $\|\cdot\|_1$  error of the result is  $8.79 \cdot 10^{-14}$ . Because the numerator is always positive we can compare the probability of  $K < 0$  with the probability of  $\Delta T < 0$ . Indeed, the expression  $K.\text{cdf}(0) - \text{NormalDistr}(1,2).\text{cdf}(0)$  gives the result  $-4.14 \cdot 10^{-14}$ .

*Combining independent measurements.* Suppose that we have two measuring instruments which differ in accuracy and error distributions. Suppose now that we want to combine the measurements of both devices such that the result is, in some sense, most accurate.

Suppose the errors are additive, so that we can ignore the measurement values themselves and concentrate only on the errors. Let  $E_1$  and  $E_2$  be r.v.'s representing the measurement errors of the two devices. Assume, for example, that  $E_1$  is uniformly distributed over  $[-1, 1]$  and  $E_2$  is normally distributed with zero mean and standard

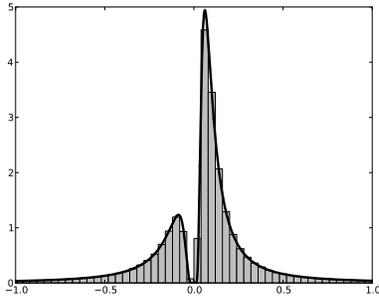


FIG. 4.6. The distribution of the coefficient thermal expansion  $K$ . A sampling-based histogram is included for illustration.

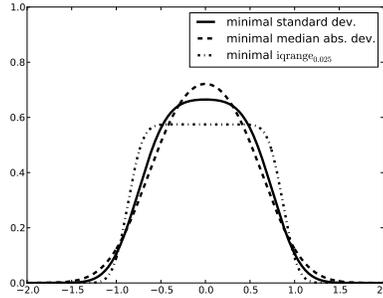


FIG. 4.7. Combining two independent measurements. Combined error distributions optimizing different measures of dispersion.

deviation 1 (this loosely corresponds to using a digital and an analog meter). Suppose further that the measurements will be combined by taking a weighted mean, in which case the combined error can be represented as  $E_\alpha = \alpha E_1 + (1 - \alpha)E_2$  for  $\alpha \in [0, 1]$  and the question is, Which value of  $\alpha$  is optimal?

The answer depends on which measure of dispersion is used for the result. We consider three measures: standard deviation  $\text{std}(E)$ , median absolute deviation  $\text{MAD}(E)$ , and inter-quantile range  $\text{iqrang}_{0.025}(E)$ , that is, the difference between 0.975-quantile and 0.025-quantile (of course, any other quantiles could be used). The last measure corresponds to obtaining an as tight as possible 95% confidence interval and, in many situations, is the most useful one.

The solution for the case of standard deviation is well known in measurement theory and is easy to derive [14]. Minimization of  $\text{iqrang}$  is more difficult as, in general, no closed form solutions exist. The  $\text{MAD}$  is also interesting as it is more robust than standard deviation and (as does  $\text{iqrang}$ ) exists for all distributions including power laws. Here, again, no general closed form solutions exist.

Using PaCAL we can easily find best values of  $\alpha$  by using the expression  $(\text{alpha}*E1 + (1-\text{alpha})*E2).\text{iqrang}(0.025)$  and plugging it into an arbitrary minimization algorithm to find the optimum value of  $\alpha$ . From a technical point of view, when our library is used, minimizing the  $\text{iqrang}$  or  $\text{MAD}$  is just as easy as minimizing the variance: simply replace the statistic to be computed in the expression passed to the minimizer. The same is true for other error distributions, as long as the problem is unimodal and the minimizer can obtain the solution.

Results of the optimization for the three dispersion measures are shown in Table 4.1. It can be seen that minimizing each dispersion measure gives visibly different results. Also, the proper choice of  $\alpha$  results in a much narrower confidence interval than combining the measurements for best variance. The difference in  $\text{MAD}$  is also quite significant. The results for standard deviation agree exactly with theoretical computations, and the  $\|\cdot\|_1$  errors are very small. Figure 4.7 shows error distributions for the three optimal values of  $\alpha$ .

**4.2. Statistical applications.**

*The generalized  $\chi^2$  distributions.* Adding the  $\chi^2$  r.v.'s with 1 degree of freedom poses interesting challenges. The p.d.f. of the  $\chi^2(1)$  distribution has a singularity at zero. After adding another independent  $\chi^2(1)$ -distributed r.v., we obtain the expo-

TABLE 4.1

Mixtures of measurements minimizing the three dispersion measures: *std*, *MAD*, and *iqrange*<sub>0.025</sub>.

	Minimal std	Minimal MAD	Minimal iqrange <sub>0.025</sub>
$E_\alpha$	$0.75E_1 + 0.25E_2$	$0.645E_1 + 0.355E_2$	$0.869E_1 + 0.130E_2$
std	<b>0.500000000001</b>	0.514511529168	0.518758649832
MAD	0.382868515019	<b>0.366016941933</b>	0.434874405386
iqrange <sub>0.025</sub>	1.83555677037	1.96758060442	<b>1.77536580577</b>
$\ \cdot\ _1$ error	0	$2.22 \cdot 10^{-16}$	$4.44 \cdot 10^{-16}$

nential distribution and the singularity disappears. However, after adding yet another such variable, the singularity at zero reappears in the first derivative of the density. Unless it is detected and handled, approximation will be poor. Overall the pattern repeats,  $\chi^2(n)$  has no singularities for even  $n$ , and for odd  $n$  a singularity appears in the  $((n-1)/2)$ th derivative. PaCAL handles the  $\chi^2$  distribution very well; we will, however, instead show a related, more interesting example.

The generalized  $\chi^2$  distribution [7] is defined as  $\lambda_0 X_0 + \sum_{i=1}^n \lambda_i X_i^2$ , where  $\lambda_i$  are real coefficients and  $X_i$  are normally distributed with mean  $\mu_i$  and variance 1. The distribution has several important applications in statistics, e.g., computing quadratic forms of normal variables. The p.d.f. of the distribution has no closed form representation, and approximations have to be used. Using PaCAL, it is very easy to work with the distribution. For example, when  $\lambda_0 = 0, \lambda_1 = 10, \lambda_2 = 1, \lambda_3 = 0.1, \mu_1 = \mu_2 = 1, \mu_3 = 0$ , the p.d.f. is shown in Figure 4.8. Notice the singularity in the first derivative at zero. After adding the initial normal term (with  $\lambda_0 = 2$  and  $\mu_0 = -5$ ), the singularity disappears (Figure 4.8). The integrals of both densities are within  $5 \cdot 10^{-16}$  from 1. To further assess the accuracy we computed means and variances of both distributions and compared them with theoretical values. We obtained means 22.1 and 17.1, respectively, and variances 606.02 and 610.02 correct up to 15 decimal places. Of course, one can compute any parameter of the distributions, such as quantiles and values of the c.d.f.

*Tail exponent estimation.* The next problem we discuss is that of obtaining distributions of sample statistics for an estimator of tail exponents. Suppose our data comes from a heavy tailed distribution with the p.d.f. behaving like  $x^{-\alpha}$  as  $x \rightarrow \infty$ , and we want to estimate the value of  $\alpha$  based on an i.i.d. sample of size  $N$ . A popular choice is Hill's estimator given by the following equation [6]:

$$\hat{\alpha} = 1 + N \left[ \sum_{i=1}^N \log \frac{X_i}{X_{\min}} \right]^{-1},$$

where  $X_i$  are the samples and  $X_{\min} > 0$  a threshold; only samples satisfying  $X_i \geq X_{\min}$  are used in the computation. Note that while we give no theoretical guarantees for functions of the r.v.'s such as the logarithm, several of them have been implemented and work very well in practice. Figure 4.9 shows the distribution of the estimator for samples of various sizes taken from the Pareto distribution with parameter 1 (the tail exponent equal to 2). It can be seen that the estimator is biased and that the bias decreases with sample size. Using PaCAL, we can easily obtain confidence intervals for the estimate; e.g., for  $N = 5$  and significance level of 0.05 we get an interval [1.5462, 3.5379].

*Noncentral  $t$ -distribution.* We now move to the noncentral  $t$ -distribution, which is used, e.g., to assess the power of the  $t$ -test. We computed the p.d.f. and the c.d.f.

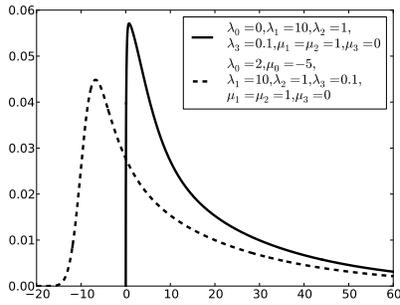


FIG. 4.8. The distribution of  $\lambda_0 X_0 + \sum_{i=1}^n \lambda_i X_i^2$ , where  $X_i \sim N(\mu_i, 1)$  (the generalized  $\chi^2$  distribution).

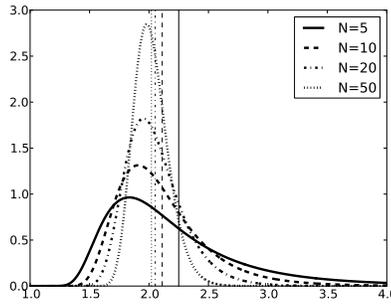


FIG. 4.9. The distribution of Hill's tail exponent estimator for an i.i.d. sample of size  $N$  drawn from a distribution with tail exponent 2. Vertical lines denote means of the estimator.

of this distribution for  $df = 7$  degrees of freedom and noncentrality parameter of  $\mu = 10.3$  for some example values of  $x$  using R, SAS, Mathematica, and PaCAL. The code used was very simple:

```
nonc_t = NormalDistr(mu, 1) / sqrt(ChiSquareDistr(df) / df)
print nonc_t.pdf(x), nonc_t.cdf(x)
```

Table 4.2 gives the computation errors. The exact values for comparison have been obtained using extended precision arithmetic and an expression for the density of the distribution involving hypergeometric functions.

TABLE 4.2

Error in computing the values of the p.d.f. and c.d.f. of the noncentral  $t$ -distribution with 7 degrees of freedom and noncentrality 10.3 using various statistical packages and PaCAL.

	$x = 5.1$		$x = -50$		$x = -1000$	
	p.d.f.	c.d.f.	p.d.f.	c.d.f.	p.d.f.	c.d.f.
PaCAL	6.94e-18	2.17e-19	5.22e-54	1.44e-53	8.26e-64	1.58e-62
SAS	3.21e-17	5.23e-17	4.42e-54	2.77e-54	— <sup>1</sup>	1.08e-62
R	1.70e-13	3.91e-13	1.55e-17 <sup>2</sup>	9.36e-14 <sup>2</sup>	5.14e-51 <sup>2</sup>	3.39e-14 <sup>2</sup>
Mathematica	1.04e-17	3.21e-17	1.54e-21	9.40e-40	5.14e-51	7.35e-49

<sup>1</sup>No value is returned due to error.  
<sup>2</sup>Decreased accuracy is signaled.

It can be seen that computing the noncentral  $t$ -distribution is not trivial; R and Mathematica fail for arguments further away from zero with error exceeding the computed value by several orders of magnitude. SAS works very well, except for a corner case in the computation of the p.d.f. PaCAL, despite being a generic package, did very well in all cases, always obtaining excellent accuracy (on par with SAS), showing that the package may be useful even for standard statistical distributions.

For the c.d.f., R uses an expansion into an infinite series of incomplete beta functions given in [17]. We believe that the expansion becomes inaccurate for large values of  $x$ . The example also shows that direct numerical integration can sometimes be preferable to more complicated approaches.

*Other examples.* We have run tests on several other examples which, due to space limitations, cannot be included in this paper. We have tested all examples (involving independent variables) from [19], as well as several standard relationships between distributions listed in statistical textbooks and Wikipedia, typically with excellent

results. Our tests included limiting of sums of i.i.d. variables, ratio distributions, etc.; see the PaCAL webpage and sources for more examples.

**5. Conclusions and future research.** This paper discussed a numerical approach to arithmetic on i.r.v.'s. Theoretical results, as well as a practical implementation, have been presented. On the theoretical side, very broad families of distributions have been introduced which are closed under the four arithmetic operations and under taking powers. The families are also shown to contain distributions which are approximable with polynomials to arbitrary accuracy. The practical implementation is shown to perform well even for difficult distributions with singularities and heavy tails. The accuracy is often close to machine precision. We have shown nontrivial applications to statistical inference and physical measurements.

There are several possible directions for future work. Some improvements can be made to the implementation such as better handling of singularities at nonzero locations (which currently incur a loss in accuracy) or the implementation of discrete distributions. On the theoretical side, it would be desirable to construct families of distributions closed also under taking functions of r.v.'s. An important case is closure under the logarithm and exponentiation.<sup>5</sup> A major topic of future research are operations on dependent r.v.'s. In the example involving the coefficient of thermal expansion it was possible to work around the dependency, but this is not always the case. Extension to random processes would also be very useful.

**Appendix: Proofs of the theorems.** We begin by proving results on properties of the densities in  $\mathcal{F}_k$ .

*Proof of Theorem 2.3.* We will give the proof only for  $x \rightarrow \infty$  and even  $i$ . The remaining cases are analogous. It is easy to see that  $f^{(i)}$  must be nonnegative and nonincreasing. We now have, for  $x \geq M$ ,

$$\begin{aligned} \|f\|_1 &\geq \int_{-\infty}^x f(y) dy = \int_{-\infty}^x dy \int_{-\infty}^y dz_1 \cdots \int_{-\infty}^{z_{i-1}} f^{(i)}(z_i) dz_i \\ &= \frac{1}{i!} \int_{-\infty}^x (x-y)^i f^{(i)}(y) dy \\ &\geq \frac{f^{(i)}(x)}{i!} \int_M^x (x-y)^i dy = \frac{f^{(i)}(x)}{i!(i+1)} (x-M)^{i+1} = \frac{f^{(i)}(x)}{(i+1)!} (x-M)^{i+1}, \end{aligned}$$

where we first express  $f$  as a repeated integral of  $f^{(i)}$ , replace the repeated integration with a single integral using the Cauchy formula for repeated integration [11], and apply the monotonicity of  $f^{(i)}$  to factor it out of the integral (recall that  $i$  is even so the integrand is nonnegative). Consequently

$$f^{(i)}(x) \leq \frac{(i+1)! \|f\|_1}{(x-M)^{i+1}} = O\left(\frac{1}{x^{i+1}}\right) \quad \text{as } x \rightarrow \infty. \quad \square$$

**LEMMA 6.1.** *If  $f \in \mathcal{F}_k$  with potential singularities  $p_1, \dots, p_n$ , then, for every  $\epsilon > 0$  and  $0 \leq i \leq k$ ,  $|f^{(i)}|$  is bounded on the set  $(-\infty, p_1 - \epsilon] \cup [p_1 + \epsilon, p_2 - \epsilon] \cup \dots \cup [p_{n-1} + \epsilon, p_n - \epsilon] \cup [p_n + \epsilon, \infty)$ .*

The lemma is an immediate consequence of the extreme value theorem and assumptions A.1 and A.3.

<sup>5</sup>The families  $\mathcal{F}_k$  are not closed under logarithms. To see this, construct a density consisting of rectangles, the height of which decreases as  $1/x$ .

Before proceeding to the proof of the main results, let us first restate basic results on convolutions [10, Propositions 8.8 and 8.10], which will frequently be used below.

LEMMA 6.2. *If  $f \in L^\infty$  and  $g \in L^1$ , then  $f \oplus g$  is bounded and uniformly continuous, and  $\|f \oplus g\|_\infty \leq \|f\|_\infty \|g\|_1$  (Young's inequality).*

LEMMA 6.3. *If  $f \in L^\infty$ ,  $g \in L^1$ , and  $f \in C^k$ , then  $(f \oplus g)^{(i)} = f^{(i)} \oplus g$ .*

Note that  $g$  need not be differentiable for the result to hold. However, the derivatives of  $f$  need to be continuous on the whole real line, which complicates our proofs as we assume only piecewise continuity. The strategy we will take is cutting the densities into smooth pieces to which Lemma 6.3 can be applied. The following result shows how such cuts can be obtained and gives some of their properties.

LEMMA 6.4. *Let  $f$  be a function. Let  $K = [p - \frac{\delta}{2}, p + \frac{\delta}{2}]$  and  $V = (p - \delta, p + \delta)$  for some  $p, \delta \in \mathbb{R}$  and  $\delta > 0$ . There is a function  $\psi_{p,\delta}$  such that*

1. *if  $f$  is  $k$  times continuously differentiable on  $V$ , then  $\psi_{p,\delta} f \in C_k$ , and for all  $0 \leq i \leq k$ ,*

$$(\psi_{p,\delta} f)^{(i)}(x) = \begin{cases} 0 & \text{for } x \notin V, \\ f^{(i)}(x) & \text{for } x \in K; \end{cases}$$

2. *if  $f$  is  $k$  times continuously differentiable on  $\mathbb{R} \setminus K$ , then  $(1 - \psi_{p,\delta})f \in C^k$ , and for all  $0 \leq i \leq k$ ,*

$$((1 - \psi_{p,\delta})f)^{(i)}(x) = \begin{cases} 0 & \text{for } x \in K, \\ f^{(i)}(x) & \text{for } x \notin V; \end{cases}$$

3. *if, in addition,  $|f^{(i)}(x)| = O(\frac{1}{|x-p|^{i+1}})$  for all  $0 \leq i \leq k$  on some set  $A$ , then  $|(\psi_{p,\delta} f)^{(i)}|$  and  $|((1 - \psi_{p,\delta})f)^{(i)}|$  are  $O(\frac{1}{|x-p|^{i+1}})$  on  $A$ . The bound is independent of  $\delta$ .*

*Proof.* By Theorem 1.4.1 in [13], there exists a function  $\psi_{p,\delta} \in C^\infty$  such that  $0 \leq \psi_{p,\delta} \leq 1$ ,

$$\psi_{p,\delta}(x) = \begin{cases} 0 & \text{for } x \notin V, \\ 1 & \text{for } x \in K, \end{cases}$$

and  $|(\psi_{p,\delta})^{(i)}(x)| = O(\frac{1}{\delta^i})$  on  $V \setminus K$ ; i.e. the derivatives of  $\psi_{p,\delta}$  are bounded, and the bound depends only on  $\delta$ . Furthermore, for all  $i > 0$ ,  $\psi_{p,\delta}^{(i)}$  is zero outside of  $V \setminus K$ . We will show that  $\psi_{p,\delta}$  satisfies the requirements of the lemma.

Since  $f$  is  $k$  times continuously differentiable on  $V$  and  $\psi_{p,\delta} \in C^\infty$ ,  $\psi_{p,\delta} f$  must be in  $C_k$ . The equality in part 1 follows, since  $\psi_{p,\delta} f$  is equal to  $f$  on  $K$  and is zero outside of  $K$ . The proof of part 2 is analogous.

For part 3 we need to split the set  $A$  into three parts. On  $A \cap K$ ,  $\psi_{p,\delta} f = 0$ , and the result is obvious. On  $A \cap (\mathbb{R} \setminus V)$ ,  $\psi_{p,\delta} f = f$ , and the bound follows from the assumption on  $f$ . The part on  $A \cap (V \setminus K)$  is more difficult.

For  $i = 0$  we have  $|\psi_{p,\delta} f| \leq |f| = O(\frac{1}{|x-p|})$ . Notice that since  $\frac{\delta}{2} \leq |x-p| \leq \delta$ , for  $x \in V \setminus K$ ,  $|\psi_{p,\delta}^{(i)}| = O(\frac{1}{\delta^i}) = O(\frac{1}{|x-p|^i})$  on  $A \cap (V \setminus K)$ . Using the Leibniz rule for the derivative of the product, we get for  $x \in A \cap (V \setminus K)$

$$|(\psi_{p,\delta} f)^{(i)}(x)| \leq \sum_{j=0}^i \binom{i}{j} |\psi_{p,\delta}^{(j)}(x)| \cdot |f^{(i-j)}(x)| = O(1/|x-p|^{i+1}).$$

The proof for  $1 - \psi_{p,\delta}$  is analogous. □

*Sum of i.r.v.'s.* We now proceed to prove Theorem 2.1 for addition. The idea is as follows: using Lemma 6.4, we will smoothly cut out problematic parts of the densities (e.g., containing singularities), demonstrate that the convolution of the remaining parts shows the desired behavior (e.g., continuity), and argue that the support of the convolution of the offending parts, which have been cut out, can be made arbitrarily small. In the next lemma, the offending parts correspond to  $f_1$  and  $g_1$ .

LEMMA 6.5. *Let  $f, g \in L^1$  and  $\|f\|_1, \|g\|_1 \leq 1$ . Further, let  $f_0, f_1, g_0, g_1$  be functions such that  $f = f_0 + f_1, g = g_0 + g_1, f_0, g_0 \in C^k$ . Then*

$$(6.1) \quad f \oplus g = f_0 \oplus g + f_1 \oplus g_0 + f_1 \oplus g_1.$$

If, in addition,  $t \notin \text{supp}(f_1 \oplus g_1)$ , then for all  $0 \leq i \leq k$ ,

$$(6.2) \quad (f \oplus g)^{(i)}(t) \leq \|f_0^{(i)}\|_\infty + \|g_0^{(i)}\|_\infty.$$

*Proof.* Equation (6.1) follows from distributivity of convolution. For  $t \notin \text{supp}(f_1 \oplus g_1)$ , the term  $f_1 \oplus g_1$  can be ignored, and the following statements hold:

$$(6.3) \quad \begin{aligned} (f \oplus g)^{(i)}(t) &= (f_0^{(i)} \oplus g + f_1 \oplus g_0^{(i)})(t) \leq \|f_0^{(i)} \oplus g + f_1 \oplus g_0^{(i)}\|_\infty \\ &\leq \|f_0^{(i)}\|_\infty \|g\|_1 + \|f_1\|_1 \|g_0^{(i)}\|_\infty \leq \|f_0^{(i)}\|_\infty + \|g_0^{(i)}\|_\infty, \end{aligned}$$

where the equality follows from Lemma 6.3 and the fact that  $f_0, g_0 \in C^k$ , the first inequality follows from the definition of the  $\|\cdot\|_\infty$  norm, the second inequality follows from the triangle inequality and Young's inequality, and the final inequality follows from the assumption on  $\|f\|_1$  and  $\|g\|_1$ .  $\square$

To prove Theorem 2.1(1), we need to show that if  $f$  and  $g$  satisfy assumptions A.1–A.3, so does  $f \oplus g$ . This is done in the following two lemmas.

LEMMA 6.6. *Let  $f, g$  be p.d.f.'s satisfying assumptions A.1 and A.3. Then  $f \oplus g$  satisfies A.3.*

*Proof.* Let  $t_0 > 0$  be a constant such that both  $f$  and  $g$  satisfy the bound in assumption A.3 on  $(-\infty, -t_0] \cup [t_0, \infty)$ . Take any  $t$  such that  $|t| > 8t_0$  and set  $p = q = 0, \delta = \frac{|t|}{4}$ . Let

$$(6.4) \quad f_0 = (1 - \psi_{p,\delta})f, \quad g_0 = (1 - \psi_{q,\delta})g, \quad f_1 = \psi_{p,\delta}f, \quad g_1 = \psi_{q,\delta}g,$$

where the  $\psi_{p,\delta}$  is given in Lemma 6.4. The reader should bear in mind that the functions depend on the chosen  $\delta$  (and thus on  $t$ ) even though it is not explicitly indicated. Since  $t$  (and thus  $\delta$ ) has been chosen large enough for  $f$  and  $g$  to satisfy the bound in assumption A.3 on the supports of  $f_0$  and  $g_0$ , respectively, by Lemma 6.4 we have  $f_0^{(i)}(x), g_0^{(i)}(x) = O(\frac{1}{|x|^{i+1}})$ , and since  $\text{supp}(f_0) = \text{supp}(g_0) = (-\infty, -\frac{|t|}{8}] \cup [\frac{|t|}{8}, \infty)$ , the bound attains its maximum at  $\pm \frac{|t|}{8}$ . Therefore  $\|f_0^{(i)}\|_\infty, \|g_0^{(i)}\|_\infty = O(\frac{1}{|t|^{\frac{i+1}{8}}})$ . Note that  $t \notin \text{supp}(f_1 \oplus g_1) = [-\frac{t}{2}, \frac{t}{2}]$ , and apply Lemma 6.5 to get  $(f \oplus g)^{(i)}(t) \leq \|f_0^{(i)}\|_\infty + \|g_0^{(i)}\|_\infty = O(\frac{1}{|t|^{\frac{i+1}{8}}})$ .  $\square$

LEMMA 6.7. *If p.d.f.'s  $f$  and  $g$  with potential singularities  $p_1, \dots, p_n$  and  $q_1, \dots, q_m$ , respectively, satisfy assumptions A.1 and A.2, then so does  $f \oplus g$  with potential singularities  $p_i + q_j, 1 \leq i \leq n, 1 \leq j \leq m$ .*

*Proof.* If  $f$  has no potential singularities, then  $f \in C^k$ , and for all  $0 \leq i \leq k, (f \oplus g)^{(i)} = f^{(i)} \oplus g$  (Lemma 6.3), which is bounded and continuous since  $f^{(i)}$  is bounded, and  $g \in L^1$  (Lemma 6.2). Thus  $f \oplus g$  satisfies A.1 and, since it has no potential singularities, trivially also A.2.

Let us now assume that  $f$  and  $g$  both have exactly one potential singularity, respectively,  $p$  and  $q$ . Fix a  $\delta > 0$ . Let  $f_0, f_1, g_0, g_1$  be defined as in (6.4). The reader should bear in mind that the functions depend on the chosen  $\delta$  even though it is not explicitly indicated. Notice that  $p, q$ , and  $\delta$  are such that supports of  $f_1$  and  $g_1$  contain the respective potential singularities of  $f$  and  $g$ , while  $f_0, g_0$  are well behaved, belonging to  $C^k$  by Lemma 6.4.

We begin by proving that  $f \oplus g$  satisfies assumption A.1. Using (6.1), we get  $f \oplus g = f_0 \oplus g + f_1 \oplus g_0 + f_1 \oplus g_1$ . Since  $f_0, g_0 \in C^k$ , by Lemmas 6.3 and 6.2,  $f_0^{(i)} \oplus g + f_1 \oplus g_0^{(i)}$  is continuous. On the other hand,  $f_1 \oplus g_1$  need not be bounded, continuous, or differentiable, but its support is contained in  $[p + q - 2\delta, p + q + 2\delta]$ . Since  $\delta$  can be made arbitrarily small,  $(f \oplus g)^{(i)}$  is continuous at every point different from  $p + q$ , which completes the proof for assumption A.1.

Let us now prove the asymptotic behavior around potential singularities. To this end, take any  $t \neq p + q$ , set  $\delta = \frac{|t-(p+q)|}{4}$ , and split  $f$  and  $g$  into  $f_0, f_1$  and  $g_0, g_1$ , respectively, using (6.4). Notice that  $\delta$  has been chosen such that after splitting  $f \oplus g$  according to (6.1),  $t$  lies outside  $\text{supp}(f_1 \oplus g_1)$ , and, according to Lemma 6.5,  $(f \oplus g)^{(i)}(t) \leq \|f_0^{(i)}\|_\infty + \|g_0^{(i)}\|_\infty$ .

We need to bound  $\|f_0^{(i)}\|_\infty$  as  $t \rightarrow p + q$ . Assume that  $t$  is close enough to  $p + q$  for assumption A.2 to hold for  $f$  on  $A = (p - \delta, p + \delta)$ . By Lemma 6.4,  $|f_0^{(i)}(x)| = O(\frac{1}{|x-p|^{i+1}})$  for  $x \in A$ . Outside of  $A$ ,  $f_0 = f$  and  $|f_0^{(i)}(x)|$  is bounded either by a constant (Lemma 6.1) or by  $O(\frac{1}{|x-p|^{i+1}})$  (assumption A.2). Clearly, for  $\delta$  small enough (that is, for  $t$  close enough to  $p+q$ ), the bound  $O(\frac{1}{|x-p|^{i+1}})$  will dominate. On the support of  $f_0$ , the bound is maximized for  $x = p \pm \frac{\delta}{2}$ . The same holds for  $g_0$ , and  $\|f_0^{(i)}\|_\infty + \|g_0^{(i)}\|_\infty = O(\frac{1}{\delta^{i+1}})$ . Substituting  $\delta = \frac{|t-(p+q)|}{4}$  completes the proof for this part.

If  $f$  has more than one potential singularity, we need to split  $f$  smoothly into  $f_0$  and  $f_1, \dots, f_n$  such that  $f = f_0 + f_1 + \dots + f_n$ ,  $f_0 \in C^k$ , and for  $1 \leq i \leq n$ , the support of  $f_i$  contains exactly one potential singularity  $p_i$ . The existence of such a split follows from [13, Theorem 1.4.5]. Analogously  $g$  is split into  $g_0$  and  $g_1, \dots, g_m$ , giving

$$(6.5) \quad f \oplus g = \sum_{i=0}^n \sum_{j=0}^m f_i \oplus g_j.$$

Each  $f_i, g_j$  has at most only one potential singularity and satisfies assumptions A.1 and A.2. By the discussion above, so does each  $f_i \oplus g_j$  with potential singularity at  $p_i + q_j$ . The sum (6.5) clearly satisfies those assumptions as well, with potential singularities  $p_i + q_j$ ,  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ .  $\square$

*Product of i.r.v.'s.* Our proof for the product will be based on the fact that  $X \cdot Y = \exp(\log X + \log Y)$ . We begin by proving some technical lemmas.

LEMMA 6.8. *Let  $p \in \mathbb{R}$ ,  $f(x) = O(\frac{1}{|x-p|^i})$  in some neighborhood  $[p-\delta, p) \cup (p, p+\delta]$  of  $p$ , and let  $g$  be a differentiable, continuous function on some interval  $[a, b]$ , such that  $g([a, b]) = [p - \delta, p + \delta]$  and  $0 < m < |g^{(1)}(y)|$  on  $[a, b]$ , where  $m$  is a constant. Then*

$$f(g(y)) = O(|y - p'|^{-i})$$

on  $[a, b] \setminus \{p'\}$ , where  $p' \in [a, b]$  satisfies  $p = g(p')$ .

*Proof.* First note that since  $g$  is invertible on  $[a, b]$ , such a  $p'$  always exists and is unique and equal to  $g^{-1}(p)$ . Since  $p' \in [a, b]$ , the mean value theorem states that for all  $y \in [a, b]$  we have  $g(y) - g(p') = g^{(1)}(\xi)(y - p')$  for some  $\xi \in [a, b]$ . Let us now substitute  $x = g(y)$  in the assumed bound on  $f(x)$ . Notice that for all  $y \in [a, b] \setminus \{p'\}$ ,  $g(y) \in [p - \delta, p) \cup (p, p + \delta]$ , so the bound can be applied. We get

$$f(g(y)) = O(|g(y) - g(p')|^{-i}) = O(|g^{(1)}(\xi)(y - p')|^{-i}) = O(|y - p'|^{-i}).$$

The last inequality follows from the assumption that  $|g^{(1)}(\xi)| > m$  for  $\xi \in [a, b]$ . □

The next lemma gives the  $n$ th derivatives of several nested functions.

LEMMA 6.9. *The following expressions hold for all  $n \geq 1$ :*

1.  $(f(e^x))^{(n)} = \sum_{i=1}^n \beta_i f^{(i)}(e^x) e^{ix}$ ,
2.  $(f(x^\alpha))^{(n)} = \sum_{i=1}^n \beta_i f^{(i)}(x^\alpha) x^{i\alpha-n}$ ,
3.  $(f(\log x))^{(n)} = \frac{1}{x^n} \sum_{i=1}^n \beta_i f^{(i)}(\log x)$ ,
4.  $(xf(x))^{(n)} = xf^{(n)}(x) + nf^{(n-1)}(x)$

for some constants  $\beta_i$  (different for every part of the lemma). In part 2, if  $\alpha \in \mathbb{N}$  and  $\lceil \frac{n}{i} \rceil \geq \alpha + 1$ , then  $\beta_i = 0$ .

*Proof.* The proof (except for the last part) will use Faà di Bruno's formula [15] for the  $n$ th derivative of the composition of two functions

$$(6.6) \quad (f(g))^{(n)} = \sum_{\pi \in P} f^{(|\pi|)}(g(x)) \prod_{B \in \pi} g^{(|B|)}(x),$$

where  $P$  is the set of all partitions of the set  $\{1, \dots, n\}$ ,  $|\pi|$  is the number of blocks in a partition  $\pi$ , and  $|B|$  is the size of partition block  $B$ .

For part 1, since all derivatives of  $e^x$  are equal to  $e^x$ , we have

$$(f(e^x))^{(n)} = \sum_{\pi \in P} f^{(|\pi|)}(e^x) \prod_{B \in \pi} e^x = \sum_{\pi \in P} f^{(|\pi|)}(e^x) e^{|\pi|x} = \sum_{i=1}^n \beta_i f^{(i)}(e^x) e^{ix},$$

where  $\beta_i$  is the number of partitions of  $\{1, \dots, n\}$  with  $i$  blocks. For part 2,

$$\begin{aligned} (f(x^\alpha))^{(n)} &= \sum_{\pi \in P} f^{(|\pi|)}(x^\alpha) \prod_{B \in \pi} (x^\alpha)^{(|B|)} = \sum_{\pi \in P} f^{(|\pi|)}(x^\alpha) \prod_{B \in \pi} c_{|B|} x^{\alpha-|B|} \\ &= \sum_{\pi \in P} f^{(|\pi|)}(x^\alpha) c_\pi x^{|\pi|\alpha-n} = \sum_{i=1}^n \beta_i f^{(i)}(x^\alpha) x^{i\alpha-n}, \end{aligned}$$

noting that  $\sum_{B \in \pi} |B| = n$  and denoting  $c_{|B|} = \alpha(\alpha - 1) \cdots (\alpha - |B| + 1)$ ,  $c_\pi = \prod_{B \in \pi} c_{|B|}$ ,  $\beta_i = \sum_{\pi \in P, |\pi|=i} c_\pi$ . If  $|\pi| = i$ , then there is a block  $B \in \pi$  with  $|B| \geq \lceil \frac{n}{i} \rceil$ . Clearly, if  $\alpha \in \mathbb{N}$  and  $\lceil \frac{n}{i} \rceil \geq \alpha + 1$ , then  $c_{|B|} = 0$ , and consequently  $c_\pi = 0$ . Since this is true for all partitions with  $i$  blocks,  $\beta_i$  must be equal to zero. The proof of part 3 is very similar to that of part 2 and is omitted. Part 4 follows from the Leibniz rule for differentiation. □

LEMMA 6.10. *Let  $f, g, h$  be functions. If  $g$  and  $h$  are  $n$  times continuously differentiable at some point  $p$  and  $f$  is  $n$  times continuously differentiable at  $g(p)$ , then  $h(x)f(g(x))$  is  $n$  times continuously differentiable at  $p$ .*

*Proof.* The continuity of  $(f(g))^{(i)}$  at  $p$  for  $0 \leq i \leq n$  follows from (6.6) after taking into account the continuity of composition and product of continuous functions. The continuity of  $[h(x)f(g(x))]^{(i)}$  at  $p$  now follows from application of the Leibniz rule for differentiation and the continuity of the product of continuous functions. □

LEMMA 6.11. Let  $|f^{(i)}(x)| = O(\frac{1}{|x-p|^{i+1}})$  in some neighborhood  $[p-\delta, p) \cup (p, p+\delta]$  of  $p$  for  $0 \leq i \leq n$ , and let  $g$  and  $h$  be  $n$  times continuously differentiable on some interval  $[a, b]$ , such that  $g([a, b]) = [p - \delta, p + \delta]$  and  $0 < m < |g^{(i)}(y)|$  on  $[a, b]$  for  $0 \leq i \leq n$ , where  $m$  is a constant. Then  $|[h(y)f(g(y))]^{(i)}| = O(|y - p'|^{-(i+1)})$  on  $[a, b] \setminus \{p'\}$ , where  $p' \in [a, b]$  satisfies  $p = g(p')$ .

*Proof.* Since  $g$  and  $h$  are  $n$  times continuously differentiable on  $[a, b]$ , they are also bounded on that interval. Applying (6.6) to  $(f(g))^{(i)}$  together with Lemma 6.8 and the fact that  $|g^{(i)}(y)|$  is upper bounded by a constant on  $[a, b]$  proves that  $(f(g))^{(i)}$  satisfies the conclusion of the lemma for  $0 \leq i \leq n$ . To see that multiplying by  $h$  does not change the conclusion, use the Leibniz rule for differentiation and take into account the boundedness of  $h^{(i)}$  on  $[a, b]$ .  $\square$

LEMMA 6.12. If  $f \in \mathcal{F}_k$  with potential singularities  $p_1, \dots, p_n$ , then  $e^x f(e^x)$  is in  $L^1$  and satisfies assumptions A.1 and A.2 for the same value of  $k$ , with potential singularities at  $\log p_i$  for each  $p_i > 0$ . Moreover,  $(e^x f(e^x))^{(i)} = O(1)$  for  $x \rightarrow \pm\infty$ ,  $0 \leq i \leq k$ .

*Proof.* Note that only the nonnegative part of the support of  $f$  affects the result, so without loss of generality we can assume that all potential singularities of  $f$  are nonnegative. Substituting  $u = e^x$ , we get  $\int_{-\infty}^{\infty} e^x f(e^x) dx = \int_0^{\infty} f(u) du$ , so  $e^x f(e^x) \in L^1$ . The fact that  $e^x f(e^x)$  satisfies A.1 and A.2 follows immediately from Lemmas 6.10 and 6.11.

To prove the asymptotic behavior at  $\pm\infty$ , let  $M$  be a constant such that the bounds in A.3 hold for  $f$  on  $[M, \infty)$ , and let  $\delta$  be a constant such that the bounds in A.2 hold for  $f$  on  $(0, \delta]$  (note that the bounds hold regardless of whether  $f$  has a potential singularity at zero). Thus  $|f^{(i)}(x)| = O(\frac{1}{x^{i+1}})$  on  $(0, \delta] \cup [M, \infty)$ , and for every  $0 \leq i \leq k$  we have  $|e^{(i+1)x} f^{(i)}(e^x)| = O(1)$  on  $(-\infty, \log \delta] \cup [\log M, \infty)$ . We now extend this result to  $(e^x f(e^x))^{(i)}$ . The case  $i = 0$  is trivial. For  $i > 0$ , denote  $z(x) = x f(x)$ , and use parts 1 and 4 of Lemma 6.9 to get

$$(6.7) \quad \begin{aligned} (e^x f(e^x))^{(i)} &= (z(e^x))^{(i)} = \sum_{j=1}^i \beta_j z^{(j)}(e^x) \cdot e^{jx} \\ &= \sum_{j=1}^i \beta_j e^{(j+1)x} f^{(j)}(e^x) + j \beta_j e^{jx} f^{(j-1)}(e^x). \end{aligned}$$

Since equation (6.7) is a weighted sum of terms of the form  $e^{(i+1)x} f^{(i)}(e^x)$ , the result follows.  $\square$

LEMMA 6.13. Let  $f$  and  $g$  be p.d.f.'s satisfying assumption A.1, such that  $f^{(i)}, g^{(i)} = O(1)$  as  $x \rightarrow \pm\infty$  for  $0 \leq i \leq k$ . Then  $(f \oplus g)^{(i)}(t) = O(1)$  as  $t \rightarrow \pm\infty$  for  $0 \leq i \leq k$ .

The proof is analogous to the proof of Lemma 6.6 and is omitted. Keep in mind that  $f, g \in L^1$  so the convolutions remain bounded. We are now ready to prove our main result for products of i.r.v.'s.

LEMMA 6.14. Let  $f$  and  $g$  be p.d.f.'s with potential singularities at  $p_1, \dots, p_n$  and  $q_1, \dots, q_m$ , respectively, satisfying assumptions A.1–A.3. Then  $f \odot g$  also satisfies assumptions A.1–A.3 and has potential singularities at 0 and at  $p_i q_j$  for  $1 \leq i \leq n$ ,  $1 \leq j \leq m$ .

*Proof.* Notice that it is sufficient to prove the result for nonnegative r.v.'s, that is, to assume that  $f(x), g(x) = 0$  for  $x < 0$ . Otherwise, due to distributivity of  $\odot$  with respect to pointwise addition of functions, each distribution is split into positive and negative parts which are multiplied separately and added; see [19] for details.

Using the nonnegativity of supports of  $f$  and  $g$  and substituting  $x = e^u$ , we get, for every  $t > 0$ ,

$$\begin{aligned}
 (f \odot g)(t) &= \int_0^{+\infty} \frac{1}{x} f\left(\frac{t}{x}\right) g(x) \, dx = \int_{-\infty}^{+\infty} e^{-u} f(e^{\log t - u}) g(e^u) e^u \, du \\
 (6.8) \qquad &= e^{-\log t} \int_{-\infty}^{+\infty} e^{\log t - u} f(e^{\log t - u}) e^u g(e^u) \, du = \frac{1}{t} h(\log t),
 \end{aligned}$$

where  $h(t) = (e^u f(e^u)) \oplus (e^u g(e^u))$ . By Lemmas 6.12 and 6.7,  $h$  satisfies assumptions A.1 and A.2 with potential singularities at  $\log p_i q_j$  for all  $p_i > 0, q_j > 0$ . Applying Lemmas 6.10 and 6.11 to  $\frac{1}{t} h(\log t)$ , we conclude that  $f \odot g$  satisfies assumptions A.1 and A.2 with potential singularities  $p_i q_j$  for all  $p_i > 0, q_j > 0$  outside the neighborhood of zero.

What remains to be demonstrated is asymptotic behavior of  $(f \odot g)(t)$  around zero (assumption A.2 for the potential singularity at zero) and at infinity (assumption A.3). We apply the Leibniz rule for differentiation of the product and Lemma 6.9(3):

$$\begin{aligned}
 (f \odot g)^{(i)}(t) &= \left[ \frac{1}{t} h(\log t) \right]^{(i)} = \sum_{j=0}^i \binom{i}{j} \left( \frac{1}{t} \right)^{(i-j)} (h(\log t))^{(j)} \\
 &= \sum_{j=0}^i \binom{i}{j} \frac{c_{i-j}}{t^{i-j+1}} (h(\log t))^{(j)} = \frac{c_i}{t^{i+1}} h(\log t) + \sum_{j=1}^i \binom{i}{j} \frac{c_{i-j}}{t^{i-j+1}} \frac{1}{t^j} \sum_{l=1}^j \beta_l h^{(l)}(\log t) \\
 (6.9) \qquad &= \frac{1}{t^{i+1}} \sum_{l=0}^i \gamma_l h^{(l)}(\log t),
 \end{aligned}$$

where  $\gamma_l$  and  $c_i$  are some constant coefficients. Let us now look at the terms  $h^{(l)}(\log t)$ . By Lemmas 6.12 and 6.13,  $h^{(l)}(t) = O(1)$  for  $t \rightarrow \pm\infty, 0 \leq l \leq k$ . As  $t \rightarrow \infty$ , we have  $\log t \rightarrow \infty$ , and  $|h^{(l)}(\log t)| = O(1)$ . The same is true when  $t \rightarrow 0^+$  and  $\log t \rightarrow -\infty$ . Multiplying by  $\frac{1}{t^{i+1}}$  gives the desired result.  $\square$

*Powers of r.v.'s.* Before completing the proof of Theorem 2.1 we need to prove Theorem 2.2.

*Proof of Theorem 2.2.* We consider only the case when  $x^r$  is an increasing function of  $x$ . When  $x^r$  is decreasing, the proof is analogous. When the function is not monotone, negative and positive values of  $x$  are considered separately and added using (1.5).

Let  $\alpha = 1/r$ . The p.d.f. of  $X^r$ , as given by (1.5), is  $\alpha x^{\alpha-1} f(x^\alpha)$ . By Lemmas 6.10 and 6.11 it satisfies assumptions A.1 and A.2, except the neighborhood of zero, where  $x^{\alpha-1}$  may have a singularity. To demonstrate the asymptotic behavior around zero and toward plus and minus infinity, use the Leibniz rule for the derivative of the

product, followed by application of Lemma 6.9(2), to get

$$\begin{aligned}
 (\alpha x^{\alpha-1} f(x^\alpha))^{(i)} &= \alpha \sum_{j=0}^i \binom{i}{j} (x^{\alpha-1})^{(i-j)} (f(x^\alpha))^{(j)} \\
 (6.10) \qquad \qquad \qquad &= \alpha \sum_{j=0}^i \binom{i}{j} c_{i-j} \sum_{l=1}^j \beta_l x^{\alpha(l+1)-(i+1)} f^{(l)}(x^\alpha)
 \end{aligned}$$

for some constants  $c_{i-j}$  and  $\beta_l$ . Assuming  $|f^{(l)}(x)| = O\left(\frac{1}{x^{l+r}}\right)$ , we get

$$x^{\alpha(l+1)-(i+1)} f^{(l)}(x^\alpha) = O\left(\frac{x^{\alpha(l+1)-(i+1)}}{x^{\alpha(l+1)}}\right) = O\left(\frac{1}{x^{i+1}}\right).$$

Since we assumed  $x^\alpha$  to be increasing, it maps the neighborhoods of zero, plus and minus infinity, into themselves, which means that the bound on  $|f^{(l)}(x^\alpha)|$  assumed above does indeed hold due to assumptions A.2 and A.3. Note that the bound also holds if  $f$  has no potential singularity at zero, since it must then be bounded together with its derivatives of order up to  $k$ .

Since assumptions A.2 and A.3 hold for each term in (6.10), they also hold for  $\alpha x^{\alpha-1} f(x^\alpha)$ .  $\square$

We are now ready to prove the main theorem of the paper.

*Proof of Theorem 2.1.* The part for addition follows from Lemmas 6.7 and 6.6, and the part for multiplication from Lemma 6.14. For subtraction it suffices to note that  $f \ominus g = f \oplus (g(-x))$ .

For division of two independent r.v.'s  $X$  and  $Y$ , write  $X/Y$  as  $X \cdot (Y^{-1})$ . Now, the density of  $Y^{-1}$  is in  $\mathcal{F}_k$  by Theorem 2.2 for  $r = -1$ . Applying the result for products shows that the density of  $X/Y$  is also in  $\mathcal{F}_k$ .  $\square$

*Approximation.* We now prove Theorem 2.4 on approximability of distributions in  $\mathcal{F}_k$ . Let us first recall some results from approximation theory.

LEMMA 6.15. *The following statements hold:*

1. *If  $f$  is continuous on  $[a, b]$ , then for every  $\epsilon > 0$  there exists a polynomial  $P$  such that  $|f - P| \leq \epsilon$  on  $[a, b]$ .*
2. *If  $f$  has a continuous derivative on  $[a, b]$ , then the polynomial  $P$  can be obtained using Chebyshev expansion.*
3. *Let  $f^{(i)}$  be continuous on  $[a, b]$  for all  $0 \leq i \leq k, k > 1$ . Then the error of the Chebyshev approximation decreases as  $O(N^{-k+1})$ , where  $N$  is the degree of the approximating polynomial.*

The first part follows from the Stone–Weierstrass theorem, the second follows from [18, Theorem 5.7], and the third is Theorem 5.14 in [18].

*Proof of Theorem 2.4.* Consider a p.d.f.  $f \in \mathcal{F}_k$  for some  $k \geq 0$  with potential singularities  $p_1 < \dots < p_n$ . We begin by defining an appropriate approximating function  $\tilde{f} \in \mathcal{F}_A$  (see section 2.1). Let  $a_i = p_i$  for  $1 \leq i \leq n$ , and take  $\alpha > k$ . For  $0 < i < n$ , let  $P_i$  be a polynomial approximation of  $f$  on  $[p_i + \delta, p_{i+1} - \delta]$ . The proof in this case follows immediately from Lemma 6.15. For the infinite intervals, let  $P_0$  be a polynomial approximation on  $[0, (1 + \delta)^{-1/\alpha}]$  of

$$h_0(u) = \begin{cases} f(a_1 + 1 - \frac{1}{u^\alpha}) & \text{for } 0 < u \leq 1, \\ 0 & \text{for } u = 0, \end{cases}$$

and let  $P_n$  be a polynomial approximation on  $[0, (1 + \delta)^{-1/\alpha}]$  of

$$h_n(u) = \begin{cases} f(a_n - 1 + \frac{1}{u^\alpha}) & \text{for } 0 < u \leq 1, \\ 0 & \text{for } u = 0. \end{cases}$$

The approximation intervals have been chosen such that they are mapped onto  $(-\infty, a_1 - \delta]$  and  $[a_n + \delta, \infty)$ , respectively. Note that the arguments of  $f$  are inverses of the expressions in the definition of  $\mathcal{F}_A$ . Thus, if  $P_0(u)$  approximates  $h_0(u)$  with error  $\epsilon$  for  $u \in [0, (1 + \delta)^{-1/\alpha}]$ , then, for every  $x \in (-\infty, a_0 - \delta]$ , we have

$$|\tilde{f}(x) - f(x)| = |\tilde{f}(a_1 + 1 + u^{-\alpha}) - f(a_1 + 1 + u^{-\alpha})| = |P_0(u) - h_0(u)| \leq \epsilon.$$

An analogous result holds for the positive tail. It now suffices to show, using Lemma 6.15, that a suitable  $P_0$  can be constructed for  $h_0$ . For this, let us find the maximum value of  $l$  for which  $h_0^{(l)}$  is continuous on  $[0, (1 + \delta)^{-1/\alpha}]$ . The only difficult part is continuity at 0, for which we need  $\lim_{u \rightarrow 0} h_0^{(l)}(u) = 0$ . Obviously  $l \leq k$ . The case  $l = 0$  is trivial. Using point 2 of Lemma 6.9 and assumption A.3, we get for  $1 \leq l \leq k$  and  $u \rightarrow 0$

$$h_0^{(l)}(u) = \sum_{j=1}^l \beta_j f^{(j)}(a_1 + 1 + u^{-\alpha}) u^{-j\alpha-l} = \sum_{j=1}^l \beta_j O(u^{\alpha(j+1)}) u^{-j\alpha-l} = O(u^{\alpha-l}),$$

and to get  $\lim_{u \rightarrow 0} h_0^{(l)}(u) = 0$  we need  $\alpha - l > 0$  and  $l \leq \alpha - 1$ . Thus  $h_0$  has at least  $\min(\alpha - 1, k)$  continuous derivatives, and the result follows from Lemma 6.15(3). The proof for the right tail is analogous.  $\square$

#### REFERENCES

- [1] Z. BATTLES AND L. N. TREFETHEN, *An extension of MATLAB to continuous functions and operators*, SIAM J. Sci. Comput., 25 (2004), pp. 1743–1770.
- [2] J.-P. BERRUT AND L. N. TREFETHEN, *Barycentric Lagrange interpolation*, SIAM Rev., 46 (2004), pp. 501–517.
- [3] J. P. BOYD, *Exponentially convergent Fourier-Chebyshev quadrature schemes on bounded and infinite intervals*, J. Sci. Comput., 2 (1987), pp. 99–109.
- [4] J. P. BOYD, *Chebyshev and Fourier Spectral Methods*, Dover, Mineola, NY, 2001.
- [5] P. L. BUTZER AND S. JANSCHKE, *A direct approach to the Mellin transform*, J. Fourier Anal. Appl., 3 (1997), pp. 325–376.
- [6] A. CLAUSET, C. R. SHALIZI, AND M. E. J. NEWMAN, *Power-law distributions in empirical data*, SIAM Rev., 51 (2009), pp. 661–703.
- [7] R. B. DAVIES, *Algorithm AS 155: The distribution of a linear combination of  $\chi^2$  random variables*, Appl. Statist., 29 (1980), pp. 323–333.
- [8] P. J. DAVIS AND P. RABINOWITZ, *Methods of Numerical Integration*, Academic Press, London, 1984.
- [9] W. FELLER, *An Introduction to Probability Theory and Its Applications*, Vol. II, John Wiley & Sons, New York, London, Sydney, 1971.
- [10] G. B. FOLLAND, *Real Analysis: Modern Techniques and Their Applications*, John Wiley & Sons, New York, 1984.
- [11] G. B. FOLLAND, *Advanced Calculus*, Prentice-Hall, Upper Saddle River, NJ, 2002.
- [12] A. G. GLEN, D. L. EVANS, AND L. M. LEEMIS, *APPL: A probability programming language*, Amer. Statist., 55 (2001), pp. 156–166.
- [13] L. HÖRMANDER, *The Analysis of Linear Partial Differential Operators*, Springer-Verlag, Berlin, 1990.
- [14] JCGM, *Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement (GUM)*, Technical report 100, Bureau International des Poids et Mesures, Sèvres, Cedex, France, 2008.

- [15] W. P. JOHNSON, *The curious history of Faà di Bruno's formula*, Amer. Math. Monthly, 109 (2002), pp. 217–234.
- [16] J. E. KOLASSA, *Series Approximation Methods in Statistics*, Springer-Verlag, New York, 2006.
- [17] R. V. LENTH, *Algorithm AS 243: Cumulative distribution function of the non-central  $t$  distribution*, Appl. Statist., 38 (1989), pp. 185–189.
- [18] J. C. MASON AND D. C. HANDSCOMB, *Chebyshev Polynomials*, Chapman & Hall/CRC, Boca Raton, FL, 2003.
- [19] M. D. SPRINGER, *The Algebra of Random Variables*, John Wiley & Sons, New York, Chichester, Brisbane, 1979.
- [20] L. N. TREFETHEN, *Is Gauss quadrature better than Clenshaw–Curtis?*, SIAM Rev., 50 (2008), pp. 67–87.
- [21] J. WALDVOGEL, *Fast construction of the Fejér and Clenshaw–Curtis quadrature rules*, BIT, 46 (2006), pp. 195–202.
- [22] R. C. WILLIAMSON, *Probabilistic Arithmetic*, Ph.D. thesis, Department of Electrical Engineering, University of Queensland, St. Lucia, Queensland, Australia, 1989.