U.S. Department of Commerce
National Oceanic and Atmospheric Administration
National Weather Service
National Centers for Environmental Prediction
5200 Auth Road
Camp Springs, MD 20746-4304

Office Note 468

Mathematical principles of the construction and characterization of a parameterized family of Gaussian mixture distributions suitable to serve as models for the probability distributions of measurement errors in nonlinear quality control

R. James Purser*
IM Systems Group, Rockville, Maryland

December 15, 2011

* email: jim.purser@noaa.gov

Abstract

Some of the nonGaussian distribution functions used to model observational error in nonlinear quality control are known to lead to undesirable multiple-minima in the cost function of the variational assimilation, which can potentially degrade the degree of convergence, and hence degrade the quality of the analysis itself in the most adverse cases. One remedy is to construct a system of plausible probability functions which, while possessing broad tails, still retain the Gaussian's property of convexity of the negative-logarithm of the density. Another remedy (to be dealt with in a sequel to this note) is to attempt to simulate, in the assimilation, the effect of including a loss model, which will tend to regularize the minimization problem even when the probabilities themselves do not all have convex negative-logarithms. This article employs methods of integral transforms to construct a system of probability models guaranteed to be expressible in the form of positive Gaussian mixtures, which we argue to be plausible in cases where the actual effective error is (as is so often the case) dominated by representation error. The simplest two-parameter form guarantees that the aforementioned convexity condition is met, while a further generalization extends the family beyond this limitation, but in a way in which the degree of violation is under the control of the third shape parameter. The archetypal example of the proposed new family is just the classical logistic or sech-squared distribution. It therefore seems appropriate to refer to the proposed system of distributions as the 'Super-logistic' family.

1. Introduction

The goal motivating this study is to find a systematic way of extending the Gaussian distribution in order to be able to model probability densities possessing tails more substantial than those we obtain with the Gaussian distribution. The Gaussian may seem to be a very natural distribution for describing errors from an idealized measuring instrument under well-controlled and repeatable conditions, but the reality of meteorological measurement practice is that the dominant portion of the effective error comes not from the inherent instrument noise, but from a combined variety of extraneous sources of what we call representation error (Lorenc 1986). The analysis is discretized using a finite resolution grid, so an important component of the representation error comes simply from the fluctuations of the atmospheric quantity that are real, but unresolved by the limited resolution of the analysis grid. Other sources of representation error come from the imperfections in the forward operator which determines what the theoretical measurement value should be, given the (assumed true) analysis state. Since the effective observation error ('ob-error') assumes a variance contingent upon factors that relate to the topical and contemporary, but imperfectly characterized, atmospheric conditions more than upon the standard instrument error, it is reasonable to take as a model for the measurement error an appropriately weighted mixture of the pure Gaussians that would each pertain to a particular ambient atmospheric condition. If these ambient conditions are themeselves distributed according to some (positive) probability density, then the mixture of Gaussians they imply must necessarily be of the positively-weighted kind, and this automatically ensures that the combined distribution of measurement errors will be of the heavy-tailed variety. There are other geophysical contexts, such as some climatological deviations, where short-tailed distributions have a role to play, but the emphasis of this note is specifically the representation of effective measurement errors in a realistic way, where heavy-tailed distributions are overwhelmingly the norm.

Based on the motivation given above, we should certainly demand that our Gaussian mixture be of the nonnegative kind – no negatively-weighted Gaussian components. We do not rule out the possibility that the centers of the participating Gaussians might *not* be coincidental. For example, when the errors tend to be relatively large, they may also tend to be skewed to one side of the mean, and when they are relatively small they may tend to lie on the other side. In this way, a generalized Gaussian mixture can exhibit a marked degree of skewness and asymmetry, which provides another sense in which the Gaussian is generalized (not only to possess heavy tails). However, we recognize that the relatively special generalization of the Gaussian which remains symmetric should serve as the reference point and anchor for the more exotic asymmetrical generalizations we shall eventually explore. We also emphasize that, in order for these extensions to be true generalizations, the Gaussian itself should belong to the extended family we are formulating, if only in the form of a limiting case. We limit the scope of our generalization to unimodal distributions, which are overwhelmingly the relevant category of distributions of measurement error in the context of atmospheric measurement (scatterometers might be considered one of the rare exceptions that would require a different special treatment). For a symmetrical unimodal distribution the centers of all the component Gaussians are collocated at the symmetry center of the composite distribution, which allows the resulting composite distribution to be written as a Laplace transform.

A closely related transform, which we shall refer to as the 'heat kernel transform' provides an interpretation of the Laplace transform representation in the form of an idealized physical analogue – the outcome of a spatially homogeneous diffusion process forced by a point source of heat injected at a rate measured as a function of age before the reference time, by an accompanying heating or forcing profile. The main advantage of this approach is that asymmetric distributions, that is, distributions exhibiting some degree of skewness that cannot be accommodated directly by the Laplace transform framework, can nevertheless be dealt with within the framework of the heat kernel transform by allowing mobility of the location of the effective heat source as a function of age. A secondary advantage is, of course, that the specification of any nonnegative heat profile, mobile or not, automatically guarantees that the resulting distribution at the reference time (as well as at all preceding times) is indeed a positive Gaussian mixture (the impulse response, or Green's function, of forced diffusion being a Gaussian).

A third important integral transform serving to characterize features of a probability distribution is the Fourier transform of the density. The application of this transform to the density function produces the characteristics function, which is also referred to as the moment generating function since the evaluation of its derivatives at the origin essentially yield the successive moments of the density distribution when these moments exist. The derivatives of the logarithm of the characteristic function correspondingly yield the successive cumulants, which play an especially important role in some of the asymptotic descriptions of distributions that facilitate their accurate numerical approximation. We shall see shortly how these integral transforms, together with the trivial transforms that express a change in variables, relate to one another.

The central challenge we address in this note is to single out from the multitude of possibilities a tightly constrained but compellingly natural family of densities describable as non-negative Gaussian mixtures with generally heavier tails than the Gaussian, yet accommodating the Gaussian as a limiting special case. The family must fulfill the requirement that the new distributions will serve as a convenient and trouble-free substitute for the Gaussian model in a variational data assimilation so that the progressive and continuous contamination of gross error in the measurements is largely solved through the optimization machinery of the variational assimilation itself. This method of Bayesian, or nonlinear quality control was proposed by Purser (1984) and has been developed since for use in operational data assimilation (Lorenc and Hammon 1988, Andersson and Järvinen 1999).

By trouble-free we allude to the problems associated with many existing probability models used in variational nonlinear quality control where their introduction into the variational assimilations has often led to multiple minima and a difficulty in locating the proper solutions during the cost function minimization. Problems of this kind are theoretically solved, regardless of the shapes of the error distributions of the measurements, by the formal inclusion within the variational problem of an appropriate so-called loss model, a topic whose deeper investigation we reserve for the companion note (Purser et al. 2012). Our approach here is, instead, to constrain the form of our principal family of generalized distributions to retain the crucial geometrical property that prevents multiple minima from ever appearing – namely, the *convexity* of the negative-logarithm of each probability density of the independent measurememt errors. For relatively large errors, we further narrow down our choices by requiring that the log-density falls off asymptotically according to a *power law*, $\approx |x|^c$. In the Gaussian case the power law behavior is just quadratic, of course, so the characteristic exponent, c, we seek for a heavy-tailed distribution must be less than two. But since it must also not fall below one if we are to retain convexity, we are naturally led to identify this critical limiting case, an asymptotic power law with unit exponent, $c = 1$, as a further restriction imposed on our family of densities. The simplest model of this type comprises the back-to-back exponential distribution, often referred to as the Laplace distribution (Johnson and Kotz, 1970) whose logarithm is exactly proportional to the absolute magnitude of the error itself. But the discontinuity of the derivative at the origin disqualifies this density from consideration as a natural model for measurement errors. Smoothing this logarithm by convolution with a suitable kernel would redeem the probability model, however, so the question we then face is: how should we choose the smoothing kernel? For technical reasons discussed in a later section, we desire a kernel whose tails decay asymptotically at least as fast as an exponential, rather than an algebraic rate; but the Gaussian, which fulfills this requirement, is found through an examination of the Laplace and heat transforms of the resulting density, to lead to a discontinuous heating profile that we can hardly accept as natural. Our choice of smoothing kernel is actually rather obvious – we choose whichever kernel emerges when we exponentiate the smoothed negative-absolute-value function. In effect, the archetypal exemplar of the heavy-tailed symmetric density function thus implicitly suggests itself through its dual role as smoothing kernel and as the product (after applying the exponential function) of that smoothing when starting out with the unsmoothed chevron function, $g(x) = -|x|$.

Known as the logistic or sech-squared distribution (Johnson and Kotz, 1970) the function that uniquely solves this prescription serves as the principal representative of our symmetric family of heavy-tailed distributions, and each alternative symmetric member of the family can then be obtained from it by raising its values to some positive exponent (and renormalizing). The limit where the exponent tends to zero recovers (after rescaling) the shape of the Laplace density; the limit where the exponent goes to infinity recovers (after rescaling) the Gaussian. Generalizing this one-parameter family to skewed densities is simply a matter of multiplying each appropriate symmetric member by an exponential function of the product of the error times a second parameter.

If we wish to extend the family to distributions whose far tails approximate a different positive power law for the negative-log-probability, we can still fix the representative symmetric member of this new branch by convolving the pure power function with the logistic smoothing kernel, and take the exponential of the resulting function. This will, once normalized, provide the standard representative of this new branch. Heavier or ligher tailed variants can then be obtained by raising this function to a power smaller than or greater than one respectively. When the power law governing the asymptotes of this generalized branch of the family implies a slowing rate of decline of the tail densities relative to the logistic density, then the logarithms of the distributions of this family will no longer possess the reassuring convexity property. In this case, the skewed generalizations cannot be obtained simply by multiplying by an exponential function because the result would be a function whose tail would eventually grow unboundedly on the side of positive skew. Instead, as we show in a later section, the properties of the Laplace and heat transforms come to the rescue in these cases and allow us to generalize the skewing procedure in a way that is natural and perfectly consistent with what is done to transform the logistic (or powers of the logistic) function into its skewed variants. The density resulting from this generalized skewing procedure remains a smooth continuous and positive Gaussian mixture as a direct consequence of the manner in which the procedure is formulated.

Owing to the importance for this development of the various integral transforms we have mentioned above, we give this topic some space in the following section. Section 3 explains the principles and reasoning guiding the criteria we adopt to prescribe the natural family of broad-tailed error distributions, confirming that the principal member of the branch whose asymptotes obey the unit-exponent power-law is indeed the logistic distribution, and goes on to provide a mathematical characterization of the functional forms of the densities of this branch, their important integral transforms and some asymptotic approximations they obey in the parameter range that brings these functions close to the Gaussian limiting form. Section 4 treats the skewed generalization of this family. Section 5 tackles the task of then extending the family consistently to include asymptote exponents that differ from the $c = 1$ possessed by the family of functions constructed in sections 3 and 4. Section 6 concludes with a discussion of these densities in the context of data assimilation and the implicit effective quality control of seriously errant observations that these functions provide.

2. Integral transforms, moments, and cumulants in the context of continuous Gaussian mixtures

(*a*) *Notation and transform definitions*

We shall employ a notation for the various interrelated density functions, their transforms, and their logarithms such that the suffix following the function name indicates the domain in which it applies. The doubly infinite domain $(-\infty, \infty)$ of the scalar error, x, will thus be denoted by the suffix, 'x', when we describe a probability density function, $f_{\mathsf{X}}(x)$, for example, while a function of this kind that is symmetric about the origin allows us to use the change of variable, $z = x^2$, and describe the same function equivalently by $f_{\mathsf{Z}}(z) = f_{\mathsf{X}}(\sqrt{z})$ on the one-sided domain $z \in [0, \infty)$. If we can assume that $f_{\mathsf{X}}(x)$ is the superposition (either discrete or continuous) of various Gaussians centered at $x = 0$, then we can equally assume that $f_{\mathsf{Z}}(z)$ has a Laplace transform representation:

$$f_{\mathsf{Z}}(z) = \int_0^\infty e^{-zs} F_{\mathsf{S}}(s)\, ds, \tag{2.1}$$

with the domain of the conjugate transform variable s taken always to be $s \geq 0$, and where $F_{\mathsf{S}}(s)$ may include impulsive (δ-function) components or other generalized functions (Schwartz 1966).

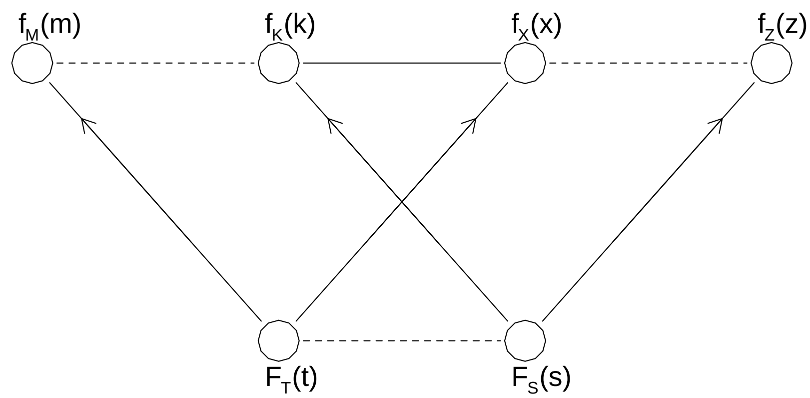

Figure 1. Interconnections among the domains linked through the various integral transforms discussed in Section 2. Dashed links indicate local change of variable transforms, while the integral transforms proper are shown in bold. Arrows show the directions in which the Laplace and heat transforms are guaranteed to provide a solution, f, for virtually all possible forcings, F, but transforms in the opposite direction depend upon a sufficiently well-behaved f.

If we make the substitution,

$$t = \frac{1}{4s}, \tag{2.2}$$

so that,

$$ds = -\frac{1}{4t^2}\, dt, \tag{2.3}$$

we can rewrite the Laplace transform in terms of what we shall call the heat kernel transform:

$$f_{\mathsf{X}}(x) = \int_0^\infty \left[\frac{1}{\sqrt{4\pi t}} \exp\left(\frac{-x^2}{4t} \right) \right] F_{\mathsf{T}}(t)\, dt, \tag{2.4}$$

where

$$F_{\mathsf{T}}(t) = \frac{\sqrt{\pi}}{2t^{3/2}} F_{\mathsf{S}}\left(\frac{1}{4t} \right). \tag{2.5}$$

If $\Phi(t)$ defines the Heaviside function (e.g., Schwartz 1966),

$$\Phi(t) = \begin{cases} 0 & : \quad t < 0 \\ 1 & : \quad t > 0 \end{cases}, \tag{2.6}$$

then the centered heat kernel,

$$W_{\mathsf{XT}}(x, t) = \frac{\Phi(t)}{\sqrt{4\pi t}} \exp\left(\frac{-x^2}{4t}\right), \tag{2.7}$$

appearing in (2.4) satisfies the (adjoint-) diffusion equation with unit diffusivity:

$$\frac{\partial}{\partial t} W_{\mathsf{XT}}(x, t) = \frac{\partial^2}{\partial x^2} W_{\mathsf{XT}}(x, t), \quad t > 0, \tag{2.8}$$

and integrates over x to unity for all $t > 0$.

The transform (2.4) is just a particular case of the general integral transform:

$$f_{\mathsf{A}}(a) = \int W_{\mathsf{AB}}(a, b) F_{\mathsf{B}}(b)\, db, \tag{2.9}$$

where the integral is over a suitable domain. Obviously, the Laplace transform itself fits this style with the kernel being defined:

$$W_{\mathsf{ZS}}(z, s) = \Phi(s) e^{-zs}, \tag{2.10}$$

but, less obviously, a local change of variables or a function rescaling can also be thought of as defining an integral transform of a relatively trivial kind in which the kernel comprises a generalized function. Thus, transforming from $f_{\mathsf{X}}(x)$ to $f_{\mathsf{Z}}(z)$ and from $F_{\mathsf{S}}(s)$ to $F_{\mathsf{T}}(t)$ we might legitimately construct transforms with the kernels,

$$W_{\mathsf{ZX}}(z, x) \equiv \Phi(z)\delta(x - \sqrt{z}), \tag{2.11a}$$

$$W_{\mathsf{TS}}(t, s) \equiv \frac{\sqrt{\pi}}{2t^{3/2}} \Phi(t)\delta(s - \frac{1}{4t}), \tag{2.11b}$$

respectively. We shall refer to trivial transforms of this kind as local transforms. Unlike the Laplace and heat transforms, which allow us always to go from the F source functions to the f functions, but not always the other way around, the local transforms are always reversible. But a very important nontrivial integral transform that is, for all practical purposes, always invertible is the Fourier transform. It enables us to construct from any given density, $f_{\mathsf{X}}(x)$, the corresponding characteristic, or moment generating function, $f_{\mathsf{K}}(k)$, and vice versa, according to the transform pair that we can choose to define generally,

$$f_{\mathsf{K}}(k) = \int_{-\infty}^{\infty} \exp(-ikx) f_{\mathsf{X}}(x)\, dx, \tag{2.12a}$$

$$f_{\mathsf{X}}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(ixk) f_{\mathsf{K}}(k)\, dk, \tag{2.12b}$$

or, in the special but important case of symmetric functions:

$$f_{\mathsf{K}}(k) = \int_{-\infty}^{\infty} \cos(kx) f_{\mathsf{X}}(x)\, dx, \tag{2.13a}$$

$$f_{\mathsf{X}}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \cos(xk) f_{\mathsf{K}}(k)\, dk. \tag{2.13b}$$

For any real and symmetric $f_{\mathsf{X}}(x)$, the function $f_{\mathsf{K}}(k)$ is then also a real and symmetric function, which suggests, by analogy to the relationship between x and z, that we try the change of variables transform, $m = k^2$, and thus $f_{\mathsf{M}}(m) = f_{\mathsf{K}}(\sqrt{m})$. Consider the form of the integral transform that expresses $f_{\mathsf{M}}(m)$ in terms of $F_{\mathsf{T}}(t)$. We combine the heat, and the Fourier transforms:

$$\begin{aligned} f_{\mathsf{M}}(m) &= \int_{x=-\infty}^{\infty}\int_{t=0}^{\infty} \frac{1}{\sqrt{4\pi t}} \exp\left(-i\sqrt{m}x - x^2/4t\right) F_{\mathsf{T}}(t)\, dt\, dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi t}} \exp\left(-\left(\frac{x}{\sqrt{4t}} + i\sqrt{mt}\right)^2\right) dx \int_0^{\infty} e^{-mt} F_{\mathsf{T}}(t)\, dt \\ &= \int_0^{\infty} e^{-mt} F_{\mathsf{T}}(t)\, dt. \end{aligned} \tag{2.14}$$

Thus, the same heating profile function $F_{\mathsf{T}}(t)$ that forces the heat kernel transform for $f_{\mathsf{X}}(x)$ also serves as the forcing of a Laplace transform for $f_{\mathsf{M}}(m) = f_{\mathsf{K}}(\sqrt{m})$. As a direct consequence of this duality we infer that $f_{\mathsf{K}}(k)$ must have a heat transform representation, with kernel $W_{\mathsf{XT}}(k, s)$ forced by,

$$\hat{F}_{\mathsf{S}}(s) = \frac{\sqrt{\pi}}{2s^{3/2}} F_{\mathsf{T}}\left(\frac{1}{4s}\right) = 2\pi F_{\mathsf{S}}(s), \tag{2.15}$$

or, equivalently, a transform with heat kernel,

$$W_{\mathsf{KS}}(k, s) = 2\pi W_{\mathsf{XT}}(k, s) \tag{2.16}$$

and with $F_{\mathsf{S}}(s)$ itself as the forcing.

A convenient diagram of all these transforms is shown in Fig. 1 with the local transforms shown as the dashed links and the nonlocal integral transforms shown as the solid links among the six nodes standing in for the domains we have subscripted, X, Z, S, T, M, K. Arrows show the direction in which the nonlocal transforms work when the inverse transforms are not guaranteed to.

When the Laplace transforms can be inverted, the inversion formula involves the complex Bromwich integral, for example:

$$F_{\mathsf{S}}(s) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{sz} f_{\mathsf{Z}}(z)\, dz, \tag{2.17}$$

where the contour in the *complex* z domain must pass rightward of all the finite poles and branch points of function $f_{\mathsf{Z}}(z)$ that is assumed to have been consistently analytically continued into the complex plane. If $c = 0$ satifies the aforementioned stipulation regarding the singularities, then the change of variable, $z = iw$, puts (2.17) into the standard form for a (complex) Fourier transform.

(*b*) *Convolution theorems*

Associated with each of the nonlocal integral transforms is a convolution theorem.

Theorem 1 (Convolution theorem for Fourier transforms)

If the functions $f_{\mathsf{X}}(x)$, $g_{\mathsf{X}}(x)$, $h_{\mathsf{X}}(x)$, are connected by transforms (2.12a) and (2.12b) to $f_{\mathsf{K}}(k)$, $g_{\mathsf{K}}(k)$, $h_{\mathsf{K}}(k)$, respectively, and if

$$h_{\mathsf{X}}(x) = \int_{-\infty}^{\infty} f_{\mathsf{X}}(x-x')g_{\mathsf{X}}(x')\,dx' \equiv f_{\mathsf{X}} \star g_{\mathsf{X}}(x), \tag{2.18}$$

then

$$h_{\mathsf{K}}(k) = f_{\mathsf{K}}(k)g_{\mathsf{K}}(k). \tag{2.19}$$

Conversely, if instead,

$$h_{\mathsf{K}}(k) = \int_{-\infty}^{\infty} f_{\mathsf{K}}(k-k')g_{\mathsf{K}}(k')\,dk', \tag{2.20}$$

then

$$h_{\mathsf{X}}(x) = 2\pi f_{\mathsf{X}}(x)g_{\mathsf{X}}(x). \tag{2.21}$$

□

Theorem 2 (Convolution theorem for Laplace transforms)

The convolution theorem for Laplace transforms states that, if $f_{\mathsf{Z}}(z)$ and $g_{\mathsf{Z}}(z)$ are respectively the transforms forced by $F_{\mathsf{S}}(s)$ and $G_{\mathsf{S}}(s)$, then

$$h_{\mathsf{Z}}(z) = f_{\mathsf{Z}}(z)g_{\mathsf{Z}}(z) = \int_{0}^{\infty} e^{-sz}H_{\mathsf{S}}(s)\,ds, \tag{2.22}$$

where

$$H_{\mathsf{S}}(s) = F_{\mathsf{S}} \star G_{\mathsf{S}}(s) \equiv \int_{0}^{s} F_{\mathsf{S}}(s')G_{\mathsf{S}}(s-s')\,ds'. \tag{2.23}$$

□

Theorem 3 (Convolution theorem for centered heat kernel transforms)

Suppose $f_{\mathsf{X}}(x)$ is the heat transform of $F_{\mathsf{T}}(t)$ and $g_{\mathsf{X}}(x)$ is the corresponding heat transform of $G_{\mathsf{T}}(t)$. Then if the function $h_{\mathsf{X}}(x)$, is the x-convolution of f_{X} and g_{X}:

$$h_{\mathsf{X}}(x) = f_{\mathsf{X}} \star g_{\mathsf{X}}(x) \equiv \int_{-\infty}^{\infty} f_{\mathsf{X}}(x-x')g_{\mathsf{X}}(x')\,dx', \tag{2.24}$$

we can derive $h_{\mathsf{X}}(x)$ as the centered heat transform of some $H_{\mathsf{T}}(t)$ where,

$$H_{\mathsf{T}}(t) = F_{\mathsf{T}} \star G_{\mathsf{T}}(t) \equiv \int_{0}^{t} F_{\mathsf{T}}(t-t')G_{\mathsf{T}}(t')\,dt'. \tag{2.25}$$

□

Theorems 1 and 2 are standard results readily obtained by factoring the exponential kernels; the heat kernel transform is not standard, so we indicate the proof of theorem 3 as follows.

Proof of theorem 3:

The proof stems from the fact that, for any two positive t_1 and t_2:

$$\int_{-\infty}^{\infty} W_{\mathsf{XT}}(x-x',t_1)W_{\mathsf{XT}}(x',t_2)\,dx' = W_{\mathsf{XT}}(x,t_1+t_2). \tag{2.26}$$

□

(*c*) *Moments and cumulants*

The qth moment of a distribution, $f_{\mathsf{X}}(x)$, is defined:

$$\chi_q = \int_{-\infty}^{\infty} x^q f_{\mathsf{X}}(x)\, dx. \tag{2.27}$$

It immediately follows from (2.12a) that

$$\chi_q = i^q \frac{d^q}{dk^q} f_{\mathsf{K}}(k)\Big|_{k=0}. \tag{2.28}$$

The moments of odd degree of a symmetric distribution vanish. Then,

$$\chi_{2q} = (-)^q \frac{d^{2q}}{dk^{2q}} f_{\mathsf{K}}(k)\Big|_{k=0} = (-)^q \frac{(2q)!}{q!} \frac{d^q}{dm^q} f_{\mathsf{M}}(m)\Big|_{m=0}. \tag{2.29}$$

When we consider that, in the context of $f_{\mathsf{X}}(x)$ being some Gaussian mixture probability model, then associated distributions $F_{\mathsf{S}}(s)$ and $F_{\mathsf{T}}(t)$ can be regarded themselves as probability densities; they quantify the preselection probability for the parameter (the s or t) that determines, in each realization, which Gaussian is picked as the *effective* population distribution for the final selection of the random variable associated with x. Thus, in synthesizing random variables possessing the distribution $f_{\mathsf{X}}(x)$ we are equally concerned with characterizing the preselection probability models, $F_{\mathsf{S}}(s)$ and $F_{\mathsf{T}}(t)$, and this characterization includes an understanding of their first few moments. Differentiating the Laplace transform formula q times at the origin:

$$\int_{0}^{\infty} s^q F_{\mathsf{S}}(s)\, ds = (-)^q \frac{d^q}{dz^q} f_{\mathsf{Z}}(z)\Big|_{z=0}, \tag{2.30}$$

reveals that $f_{\mathsf{Z}}(z)$ serves as an effective moment-generating function for the probability density described by $F_{\mathsf{S}}(s)$. By duality, $f_{\mathsf{M}}(m)$ likewise serves as an effective moment-generating function for the distribution, $F_{\mathsf{T}}(t)$. From this latter identification, combined with (2.29), we deduce:

Theorem 4

If we denote by χ_{2q} the $2q$th moment of the symmetric distribution $f_{\mathsf{X}}(x)$ composed as the Gaussian mixture described by the heat transform forced by $F_{\mathsf{T}}(t)$, and if we denote by τ_q the qth moment of $F_{\mathsf{T}}(t)$, i.e.,

$$\tau_q = \int_{0}^{\infty} t^q F_{\mathsf{T}}(t)\, dt, \tag{2.31}$$

then:

$$\chi_{2q} = \frac{(2q)!}{q!} \tau_q. \tag{2.32}$$

□

Remark:

The factor, $(2q)!/q!$ is sometimes written $(2q-1)!!2^q$, where, for positive odd $(2q-1)$, the double-factorial notation is used to express the stride-two products,

$$(2q-1)!! \equiv 1.3.5\ldots(2q-1). \tag{2.33}$$

The distribution of the sum of a set of independent random variables has density given by the convolution of the densities corresponding to this set, and therefore the resulting characteristic function is the product of the characteristic functions of this set (by the Fourier convolution theorem). While this allows one to obtain the moments of the resulting distribution from the successive derivatives of this product characteristic function, for the purposes of studying the asymptotic behavior of sums of large numbers of independent random variables, especially those that are identically distributed, an alternative combination of the moments that behaves additively is often more directly informative. Clearly such quantities are obtained when the function whose derivatives we take at $k = m = 0$ is the *logarithm* of the characteristic function, and the corresponding evaluations at each degree are then called the cumulants of the density distribution. As with the moments themselves, the odd-degree cumulants vanish for a symmetric distribution. If we denote the cumulants of $f_{\mathsf{X}}(x)$ by $\hat{\chi}_q$, so that, by analogy with (2.28) we have

$$\hat{\chi}_q = i^q \frac{d^q}{dk^q} \ln[f_{\mathsf{K}}(k)]\bigg|_{k=0} , \tag{2.34}$$

then, since, for symmetric densities, $\ln[f_{\mathsf{M}}(m)]$ also serves as the cumulant generating function for these $\hat{\chi}_q$ as well as for the cumulants, $\hat{\tau}_q$ of the corresponding $F_{\mathsf{T}}(t)$, the cumulant version of (2.32) in theorem 4 immediately follows.

The moments are expressed in terms of the cumulants by the succession of formulae:

$$\begin{aligned}
\chi_1 &= \hat{\chi}_1, && (2.35a)\\
\chi_2 &= \hat{\chi}_2 + \hat{\chi}_1^2, && (2.35b)\\
\chi_3 &= \hat{\chi}_3 + 3\hat{\chi}_1\hat{\chi}_2 + \hat{\chi}_1^3, && (2.35c)\\
\chi_4 &= \hat{\chi}_4 + 4\hat{\chi}_1\hat{\chi}_3 + 3\hat{\chi}_2^2 + 6\hat{\chi}_1^2\hat{\chi}_2 + \hat{\chi}_1^4, && (2.35d)\\
\chi_5 &= \hat{\chi}_5 + 5\hat{\chi}_1\hat{\chi}_4 + 10\hat{\chi}_2\hat{\chi}_3 + 10\hat{\chi}_1^2\hat{\chi}_3 + 15\hat{\chi}_1\hat{\chi}_2^2 + 10\hat{\chi}_1^3\hat{\chi}_2 + \hat{\chi}_1^5, && (2.35e)\\
&\dots \\
\chi_m &= \sum_{a_1+2a_2\dots+ma_m=m} M_3(m; a_1, \dots, a_m)[\hat{\chi}_1]^{a_1}[\hat{\chi}_2]^{a_2}\dots[\hat{\chi}_m]^{a_m}, && (2.35f)
\end{aligned}$$

where the Faà di Bruno coefficient in the general term is the combinatorial partition function, M_3, defined in Abramowitz and Stegun (1972, page 831) by:

$$M_3(m; a_1, \dots, a_m) = m!/(1!)^{a_1} a_1!(2!)^{a_2} a_2! \dots (m!)^{a_m} a_m! \, . \tag{2.36}$$

If we let $\tilde{\chi}_q$ denote the central moments about an origin for x chosen to make the first moment vanish, then, assuming the density is normalized, $\chi_0 = 1$:

$$\tilde{\chi}_q = \sum_{r=0}^{q} (-)^r \binom{q}{r} [\chi_1]^r \chi_{q-r}, \tag{2.37}$$

and the set of formulae defining the cumulants in terms of these central moments can be recovered recursively by symbolic manipulations inverting (2.35a) ..., etc., and substituting (2.37):

$$\hat{\chi}_1 = \chi_1, \tag{2.38a}$$

$$\hat{\chi}_2 = \tilde{\chi}_2, \tag{2.38b}$$
$$\hat{\chi}_3 = \tilde{\chi}_3, \tag{2.38c}$$
$$\hat{\chi}_4 = \tilde{\chi}_4 - 3\tilde{\chi}_2^2, \tag{2.38d}$$
$$\hat{\chi}_5 = \tilde{\chi}_5 - 10\tilde{\chi}_2\tilde{\chi}_3, \tag{2.38e}$$
$$\ldots$$

and so on. (More extensive tables of these results are given in Kendall and Steward 1977.) We shall later employ the cumulants of our densities to construct asymptotic approximations of them when they become sufficiently close to the Gaussian shape; for the limiting case of the Gaussian itself, the cumulants beyond degree 2 all vanish.

(*d*) *Heat kernel transforms when $f_{\mathsf{X}}(x)$ is asymmetric*

The nonnegative symmetric Gaussian mixture is obtained when $F_{\mathsf{T}}(t) \geq 0$. In order to accommodate skewness we preserve the feature that the heat source is a single point at any given time, but we allow this point mobility in time by ascribing to it a definite trajectory. Then the generalized heat kernel transform is defined:

$$f_{\mathsf{X}}(x) = \int_0^\infty \frac{1}{\sqrt{4\pi t}} \exp\left(\frac{-(x - \hat{x}(t))^2}{4t}\right) F_{\mathsf{T}}(t)\, dt, \tag{2.39}$$

where $\hat{x}(t)$ defines the trajectory location of the heat source at a time t in the past. (recall that t is to be thought of as the age characterizing this Gaussian component of the transform.) For the resulting function f_{X} to be a proper probability density, it must integrate to one; this is equivalent to requiring that the heating in the diffusion model, integrated over all ages, t, is also one:

$$\int_0^\infty F_{\mathsf{T}}(t)\, dt = 1. \tag{2.40}$$

We have established how the Gaussian might be generalized to heavy-tailed, and possibly skewed distributions having (like the Gaussian) infinite support over x, by modeling the construction of such distributions on problems in heat diffusion forced by a mobile point source of nonnegative heating acting over all past time. But in order to construct a useful parameterized family of distributions of this kind we must decide on the rules for the determination of the heating function, $F_{\mathsf{T}}(t)$, and the corresponding rules for the construction of the trajectory, $\hat{x}(t)$, in the cases of distributions with skewness. We seek to determine these rules in a way that seems most natural and self-consistent, as we explain in the next section.

3. A simple family of symmetric heavy-tailed distributions with advantageous features for the representation of measurement error

(*a*) *Guiding principles and a self-referential procedure for constructing the reference example of a smooth density with prescribed exponential-power-law tails*

The standard unit-variance Gaussian, which is generated by the centered heat kernel transform when $F_{\mathsf{T}}(t) = \delta(t - 1/2)$, has the convex function, $x^2/2$ for its negative-logarithm (ignoring

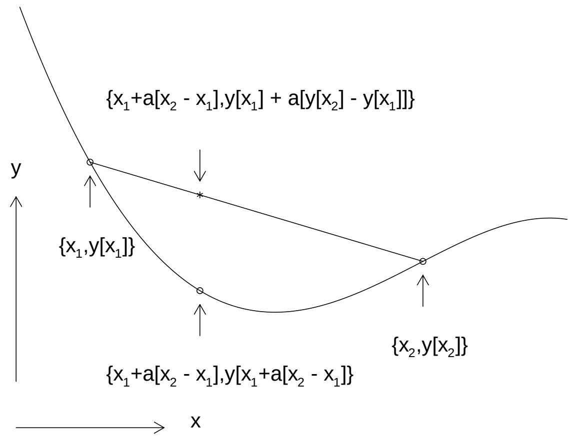

Figure 2. Schematic depiction of the property of convexity of a function $y(x)$. Although the function shown is convex to the left of x_2, it is clearly not convex in places sufficiently to the right of this point.

any additive constant). Convexity is the property, illustrated schematically in Fig. 2, that ensures that, for any pair of arguments, x_1 and x_2, and any number a in $[0,1]$, the convex function $y(x)$ obeys: $y(x_1+a(x_2-x_1)) \geq y(x_1)+a(y(x_2)-y(x_1))$. Convexity of negative-log-probability densities is a useful guarantee that any arbitrary multiplicative combination of these densities remain unimodal. (Essentially, we exploit the property that sums of convex functions are necessarily convex.) This is an important consideration in the data assimilation context where the posterior probability formed by the combination of independent measurements and background information *is* such a product. This suggests that, at least to begin with, we should seek heavy-tailed generalizations which preserve the property of convexity. The simple power-law behavior of the log-Gaussian form does not completely carry over in any smooth heavy-tailed distribution because the general exponent, c, in the log-probability needs to be smaller that the Gaussian's $c=2$ in order for the tails to be asymptotically heavier, which in turn means that the distribution profile possesses an undesirable singularity in its second derivative at the origin. However, it is perfectly acceptable to preserve the $c<2$ power law behavior asymptotically at large $|x|$, in which case, convexity can be preserved provided $c \geq 1$. The choice $c=1$ is therefore attractive because it provides maximal opportunity to fatten the tails while preserving convexity. This power law aymptotic behavior is taken as a desideratum as we take the first step towards a parameterized generalization of the distributions that describe observation errors.

Another property of the Gaussian family that we can seek to preserve in our initial generalization is the closure with respect to positive exponentiation; we can raise the function, e^{-x^2}, to any positive power and still recover a Gaussian, because the multiplication of the logarithm of this function by any positive number will preserve the original parabolic form. Likewise, for the symmetric heavy-tailed generalizations for each asymptotic shape index, c, we can hope to find a logarithmic form preserved under multiplication by an arbitrary positive factor, b. But we need to establish a systematic way to round the aforementioned singularity at the origin for all c, and in such a way that convexity is manifestly preserved for the cases $c \geq 1$ capable of supporting a convex negative-log-probability.

It is helpful to apply the integral transform formalism to the logarithm $g_{\mathsf{X}}(x)$ and $g_{\mathsf{Z}}(z)$ of the density function $f_{\mathsf{X}}(x)$ and $f_{\mathsf{Z}}(z)$, but in this case we must confront the obstacle that the range of the transforms is infinite. This difficulty is resolved by modifying the transform kernels, $\hat{W}_{\mathsf{ZS}}(z,s)=W_{\mathsf{ZS}}(z,s)-1$ and $\hat{W}_{\mathsf{XT}}(x,t)=W_{\mathsf{XT}}(x,t)-1$, which, being equivalent to evaluating the transforms, $g_{\mathsf{Z}}(z)-g_{\mathsf{Z}}(0)$ and $g_{\mathsf{X}}(x)-g_{\mathsf{X}}(0)$, enables finite results to be extracted even in some cases where the Laplace and heat transform forcings have infinite integrals. For the example of the unsmoothed power law formula characterized by an exponent, $c\in(0,2)$, then:

$$\begin{aligned} g_{\mathsf{X}}(x) &= \ln f_{\mathsf{X}}(x)=-|x|^c, && \text{(3.1a)}\\ g_{\mathsf{Z}}(z) &= \ln f_{\mathsf{Z}}(z)=-z^{c/2}, && \text{(3.1b)}\end{aligned}$$

and the modified Laplace transform representing this case is given by the following expression involving the gamma function.

Theorem 5

$$g_{\mathsf{Z}}(z)=-z^{c/2}=\int_0^\infty (e^{-sz}-1)\hat{G}_{\mathsf{S}}(s)\,ds, \tag{3.2}$$

with

$$\hat{G}_{\mathsf{S}}(s)=-\frac{1}{\Gamma(-c/2)}s^{-(1+c/2)}. \tag{3.3}$$

□

Proof of theorem 5

For $z>0$, $s>0$ and $c\in(0,2)$ let function J be defined:

$$J(z)=\int_0^\infty (1-e^{-sz})s^{-(1+c/2)}\,ds>0 \tag{3.4}$$

The two inequalities:

$$\begin{aligned} 1-e^{-sz} &< 1,\\ 1-e^{-sz} &< sz\end{aligned}$$

imply

$$\begin{aligned} J(z) &< \int_0^{1/z} sz s^{-(1+c/2)}ds+\int_{1/z}^\infty s^{-(1+c/2)}ds\\ &= \left(\frac{1}{1-c/2}+\frac{1}{c/2}\right)z^{c/2}\\ &\to 0,\quad \text{as } z\to 0.\end{aligned}$$

So the result is established at the end point $z\to 0$.

We take the z-derivative of $J(z)$ to produce the more standard Laplace transform:

$$\frac{dJ}{dz} = -\int_0^\infty e^{-sz}s^{-c/2}\,ds$$

$$
\begin{aligned}
&= -z^{c/2-1}\int_0^\infty e^{-s'} s'^{-c/2}\, ds' \\
&= -z^{c/2-1}\Gamma(1-c/2) \\
&= \frac{c}{2} z^{c/2-1}\Gamma\left(\frac{-c}{2}\right) = \frac{d}{dz} z^{c/2}\Gamma\left(\frac{-c}{2}\right),
\end{aligned}
$$

and the result is therefore established for all $z \geq 0$ by continuity.

□

Remarks

The fact that the coefficient, $-1/\Gamma(-c/2)$, is positive leads to the following theorem describing the Laplace transform of the exponential of the $g_{\mathsf{Z}}(z)$ of theorem 5:

Theorem 6

The function,

$$f_{\mathsf{Z}}(z) = \exp(g_{\mathsf{Z}}(z)) \equiv \exp(-z^{c/2}), \tag{3.5}$$

for $0 < c < 2$ has a Laplace tranform representation,

$$f_{\mathsf{Z}}(z) = \int_0^\infty e^{-sz} F_{\mathsf{S}}(s)\, ds, \tag{3.6}$$

with $F_{\mathsf{S}}(s) > 0 \quad \forall s > 0$.

□

Proof of theorem 6

From theorem 5 the positive, but monotonic-decreasing function, (3.3), can be split into two nonnegative parts:

$$\hat{G}_{\mathsf{S}}(s) = \hat{G}_{\mathsf{S}1}(s) + \hat{G}_{\mathsf{S}2}(s), \tag{3.7}$$

where,

$$\hat{G}_{\mathsf{S}2}(s) = \begin{cases} \hat{G}_{\mathsf{S}}(1) & : \quad s \leq 1 \\ \hat{G}_{\mathsf{S}}(s) & : \quad s \geq 1 \end{cases}, \tag{3.8}$$

has a finite integral, say I_2, over all $s > 0$ and is positive there (the integral of its complement, $\hat{G}_{\mathsf{S}1}(s)$, is infinite, however). Thus $\hat{g}_{\mathsf{Z}2}(z) = g_{\mathsf{Z}2}(z) + I_2$ is the *conventional* (unmodified) Laplace transform of $\hat{G}_{\mathsf{S}2}(s)$. Expanding the exponential as a series:

$$\exp[\hat{g}_{\mathsf{Z}2}(z)] = 1 + \hat{g}_{\mathsf{Z}2}(z) + \ldots \frac{1}{n!}\hat{g}^n_{\mathsf{Z}2}(z) \ldots \tag{3.9}$$

and invoking the Laplace transform convolution theorem 2 term by term for the powers of $\hat{g}_{\mathsf{Z}2}(z)$ proves that,

$$
\begin{aligned}
f_{\mathsf{Z}2}(z) &= \exp(-I_2)\exp[\hat{g}_{\mathsf{Z}2}(z)] \\
&= \int_0^\infty e^{-zs} F_{\mathsf{S}2}(s)\, ds
\end{aligned} \tag{3.10}
$$

with $F_{\mathsf{S}2}(s) > 0, \quad \forall s > 0$.

The infinite integral of $\hat{G}_{\mathsf{s}1}(s)$ introduces a subtlety that requires a slightly different treatment when we exponentiate and characterize $F_{\mathsf{s}1}(z)$. Consider the $\hat{s}$-parameterized modification to $\hat{G}_{\mathsf{s}1}(s)$ defined by modulating it with the Heaviside function,

$$\hat{G}_{\mathsf{s}1}(\hat{s}; s) = \Phi(s - \hat{s})\hat{G}_{\mathsf{s}1}(s), \tag{3.11}$$

so that the corresponding *unmodified* Laplace transform of it gives the finite integral for each $\hat{s} > 0$:

$$g_{\mathsf{z}1}(\hat{s}; z) = \int_0^\infty e^{-zs}\hat{G}_{\mathsf{s}1}(\hat{s}; s)\, ds = \int_{\hat{s}}^\infty e^{-zs}\hat{G}_{\mathsf{s}1}(s)\, ds. \tag{3.12}$$

Then, as before, we obtain a series expansion

$$\exp\left(g_{\mathsf{z}1}(\hat{s}, z) - g_{\mathsf{z}1}(\hat{s}; 0)\right) = F_{\mathsf{s}1}(\hat{s}; s) = \exp(-g_{\mathsf{z}1}(\hat{s}; 0)) \sum_{n=0}^{\infty} \frac{1}{n!} g_{\mathsf{z}1}^n(\hat{s}; z). \tag{3.13}$$

and infer, in this case, that $F_{\mathsf{s}1}(s) \geq 0 \quad \forall s > 0$. By continuity, this result must also hold in the limit, that is, $F_{\mathsf{s}1}(\hat{s}; s) \to F_{\mathsf{s}1}(s) \geq 0 \quad \forall s > 0$.

Now since $f_{\mathsf{z}}(z) = \exp[g_{\mathsf{z}}(z)] = f_{\mathsf{z}1}(z) f_{\mathsf{z}2}(z)$ the convolution theorem gives:

$$F_{\mathsf{s}}(s) = F_{\mathsf{s}1} \star F_{\mathsf{s}2}(s) \tag{3.14}$$

and, since $F_{\mathsf{s}2}(s) > 0$ and $F_{\mathsf{s}1}(s) \geq 0$, then $F_{\mathsf{s}}(s) > 0$ provided there is no $\hat{s} > 0$ such that $F_{\mathsf{s}1}(s) = 0 \quad \forall s < \hat{s}$. But we can be sure that there is no such positive $\hat{s}$ because, if there were, then $F_{\mathsf{s}}(s)$ itself would be subject to the some property and the function $f_{\mathsf{z}}(z)$ would then necessarily decrease faster than $e^{-z\hat{s}}$ as $z \to \infty$, which is clearly in contradiction with actual rate of decrease of $f_{\mathsf{z}}(z)$ given by (3.5) when $c < 2$. Thus, the result is established.

□

In the case where $c = 1$ the solution for the transform yielding $g_{\mathsf{z}}(z)$ is:

$$g_{\mathsf{z}}(z) \equiv -z^{1/2} = \int_0^\infty (e^{-sz} - 1)\frac{s^{-3/2}}{2\sqrt{\pi}}\, ds \tag{3.15}$$

and, in this special case, the corresponding transform yielding $f_{\mathsf{z}}(z)$ has an analytic expression:

$$f_{\mathsf{z}}(z) \equiv \exp(-z^{1/2}) = \int_0^\infty e^{-sz} \frac{1}{2\sqrt{\pi}s^{3/2}} \exp\left(-\frac{1}{4s}\right)\, ds, \tag{3.16}$$

as is shown in Appendix A through a direct application of the heat diffusion equation.

However, for the general choice of $c \in (0, 2)$, although we can guarantee that it is positive for all s, an expression for $F_{\mathsf{s}}(s)$ in elementary functions is not possible. In terms of the heat kernel transform, the case $c = 1$ translates to:

$$g_{\mathsf{x}}(x) \equiv -|x| = \int_0^\infty \frac{1}{\sqrt{4\pi t}} \left(\exp\left(-\frac{x^2}{4t}\right) - 1\right) 2\, dt, \tag{3.17}$$

i.e., a uniform heating profile of intensity 2 units. For the corresponding $f_{\mathsf{x}}(x)$ we have:

$$f_{\mathsf{x}}(x) \equiv \exp(-|x|) = \int_0^\infty \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{x^2}{4t}\right) 2\exp(-t)\, dt. \tag{3.18}$$

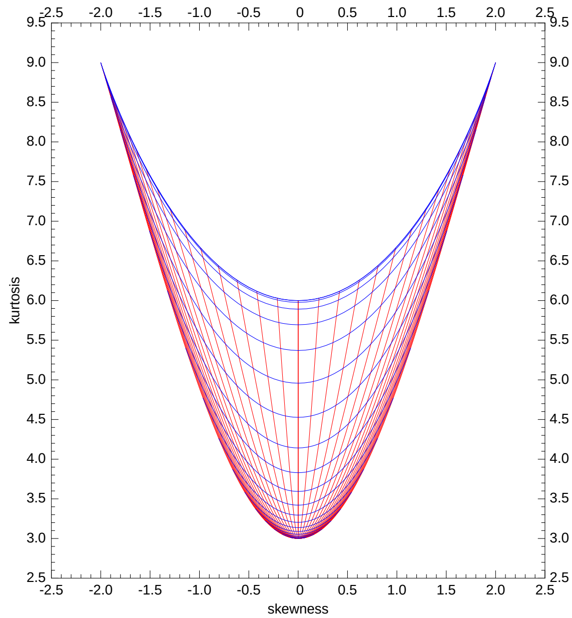

Figure 3. The locus of images in the skewness-kurtosis plane of a family of skewed and heavy-tailed distributions.

Perhaps the simplest way to smooth a function, $g_{\mathsf{X}}(x)$, is to convolve it with a Gaussian, which is equivalent to translating its corresponding heat profile function, $G_{\mathsf{T}}(t)$, by an appropriate fixed aging increment, Δt, of t. For example, for a convolution with the Gaussian, $(1/\sqrt{\pi})\exp(-x^2)$, a t-translation by an amount, $\Delta t = 1/4$, needs to be made. For the case $c = 1$ the resulting smoothed g_{X} becomes proportional to the integrated complementary error function, ierfc(x), which is defined in Abramowitz and Stegun (1972). But the salient feature of such a t-translation on the corresponding Laplace transform is to set a maximum value, $s = 1/(4\Delta t)$, beyond which the smoothed function's corresponding $G_{\mathsf{S}}(s)$ vanishes. Unfortunately, when this smoothed function $g_{\mathsf{X}}(x)$ is then exponentiated to get the corresponding smoothed $f_{\mathsf{X}}(x)$ we find that this $F_{\mathsf{S}}(s)$, while now positive for all $t > 0$, exhibits a rather unnatural rippled profile and cannot therefore be considered a compelling candidate for a natural heavy-tailed generalization of the Gaussian itself. In spite of its unnaturalness, this Gaussian-smoothed function $g_{\mathsf{X}}(x)$ does at least preserve convexity, but it is easy to see that convolution with any positive bell-shaped smoothing kernel must also preserve the convexity of $g_{\mathsf{X}}(x)$. A convexity-preserving generalization to an asymmetric distribution is obtained simply by adding to the symmetric $g_{\mathsf{X}}(x)$ a function linear in x (but not so steep that it would cause the positive-skew tail to grow with $|x|$ instead of decaying).

It is useful to know what domain of the skewness-kurtosis plane is covered in the probability model constructed in this way. Figure 3 shows this domain when the smoothing kernel used to obtain $g_{\mathsf{X}}(x)$ from $-|x|$ possesses exponentially decreasing tails. The upper segment of the boundary represents the distributions comprising the sub-family of back-to-back exponentials, sometimes referred to as asymmetric Laplace distributions (Kotz et al. 2001), and which includes the true Laplace distribution as the only symmetric example. The endpoints have a kurtosis of 9 and a skewness of ± 2. The lower segment of the boundary comprises the images in the skewness-kurtosis plane associated with the distributions formed from positive powers of the Gumbel (or Fisher-Tippett Type-1) distribution. The Gumbel distribution is,

$$f(x) = \exp(x - \exp(x)), \tag{3.19}$$

in the case of a negative skew, or,

$$f(x) = \exp(-x - \exp(-x)), \tag{3.20}$$

in the case of a positive skew. The Gaussian distribution itself is found at the minimum-kurtosis point on this bounding curve. It is noteworthy that, while all the distributions represented by the interior points of this region are expressible as Gaussian mixtures, the limiting cases mapping to the lower boundary, that is, the powers of Gumbel distributions, are not themselves expressible as the mobile point-forced Gaussian mixtures we have discussed. Nevertheless, this family of generally skewed, but minimal kurtosis distributions is found to bound a great many different families of Gaussian mixtures that, with controllable kurtosis and skewness, have a logarithm whose graph is also expressible as a convolution of the chevron function, $-|x|$, with *some* symmetric bell-shaped function that has exponentially diminishing tails. Having this family of exponentiated Gumbel distributions as bounding the limiting cases will be taken as an additional goal to guide us in our quest for the formulation of heavy-tailed generalizations of the Gaussian. This additional desideratum turns out to exclude those bell-shaped smoothing kernels whose tails diminish at only an algebraic rate (i.e., slower than exponential), such as the otherwise attractive functions proportional to $1/(1+x^2)^q$, for some $q > 1/2$. The fact that the logarithm of a symmetric member of this generalized family might be found as the convolution of $-|x|$ with a non-Gaussian bell-shaped function with exponential tails suggests an obvious recursive solution; let this bell-shaped convolving function be identical (apart form normalization) to the exponential of the *result* of this same convolution. Other symmetric members of this family can then be obtained from this self-consistent bell-shaped function by raising it to whichever positive power produces the desired exponential rate of decrease of the far tails.

We know that, since the asymptotic derivatives of any $g_{\mathsf{x}}(x)$ will be ∓ 1 at $\pm\infty$, and since the g_{x} that we want must be convex, a proper probability density (one that is non-negative and integrates to one) will always be found as half of the *second* derivative of this $g_{\mathsf{x}}(x)$. Let $f_{\mathsf{x}}(x)$ be this proper density associated with $g_{\mathsf{x}}(x)$. The simplest additional assumption is that:

$$f_{\mathsf{x}}(x) = \exp(g_{\mathsf{x}}(x)), \tag{3.21}$$

so that a defining condition for the basic representative $g_{\mathsf{x}}(x)$ is the differential equation:

$$g''_{\mathsf{x}}(x) = -2\exp(g_{\mathsf{x}}(x)). \tag{3.22}$$

A simple numerical iteration, which provides a computational solution, is worth describing.

For clarity, we shall temporarily drop the suffix x, it being understood that at each stage the iterations of g and f can, if needed, be decomposed as a heat transform. Let us start, for iteration zero, with a $g_{(0)}(x) = -|x|$ and, for *any* stage n of the iteration, define the corresponding

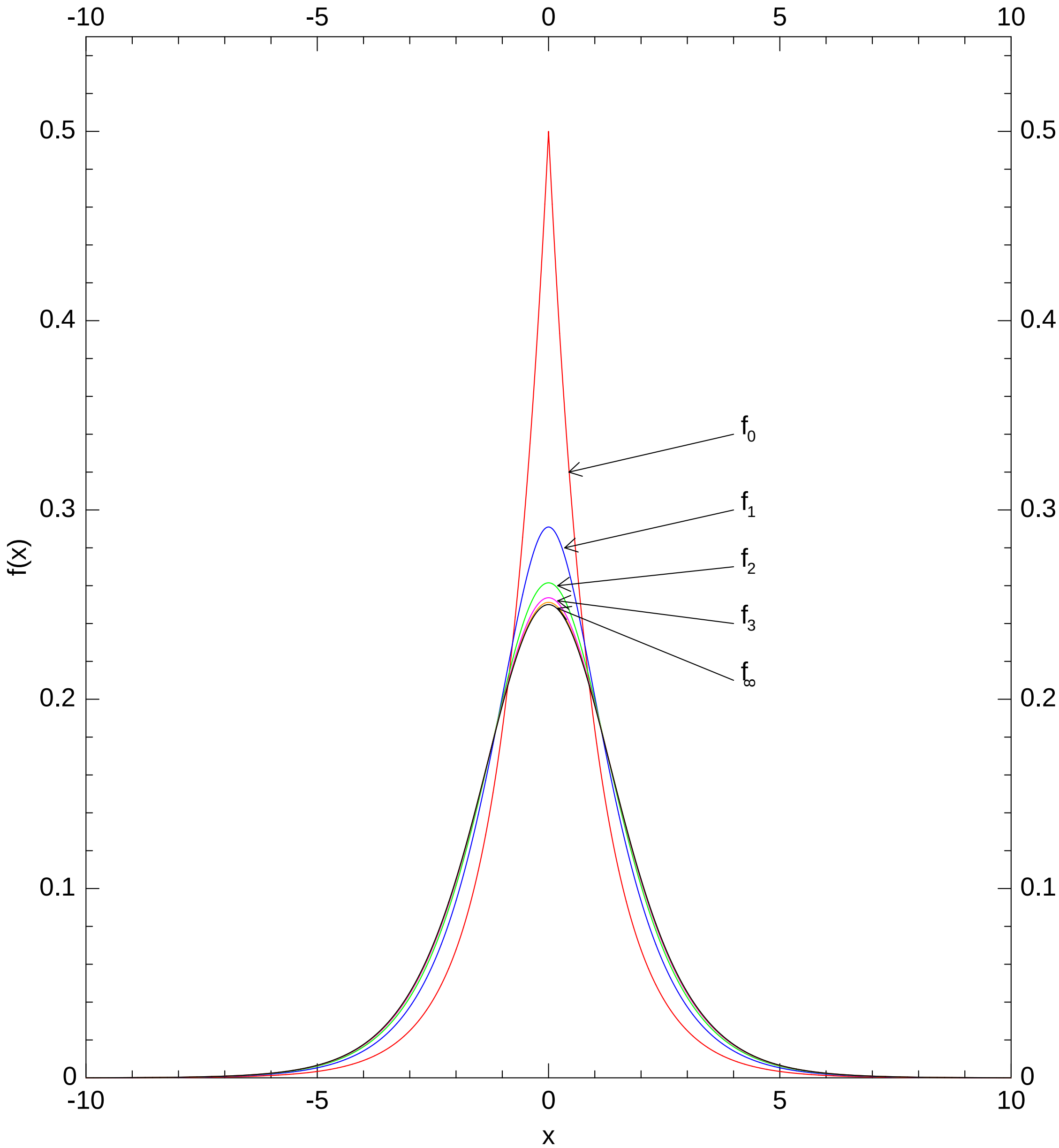

Figure 4. Successive iterations of the bell-shaped distribution, $f_{(n)}(x)$, converging to $f_{(\infty)}(x)$.

$f(x)$ by,

$$f_{(n)}(x) = \frac{\exp(g_{(n)}(x))}{\int_{-\infty}^{\infty} \exp(g_{(n)}(x'))\, dx'}. \tag{3.23}$$

Then we twice integrate the condition:

$$g''_{(n+1)}(x) = -2f_{(n)}(x), \tag{3.24}$$

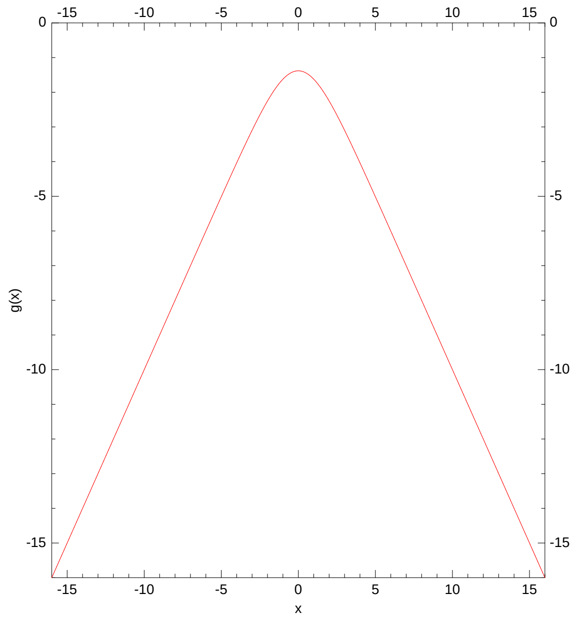

Figure 5. Final iteration of $g_{(n)}(x)$.

choosing the two integration constants such that

$$\lim_{x\to\pm\infty} g_{(n+1)}(x) + |x| = 0. \tag{3.25}$$

The first few iterations for $f_{(n)}(x)$ obtained by this procedure, together with its theoretical limit, are shown in Fig. 4. The final solution for $g(x)$ is shown in Fig. 5.

We deduce that the solution does indeed contain only positive contributions in its heat transform, or equivalently, in its Laplace transform under the usual $z = x^2$ substitution. This is because the starting iterate has this property and the subsequent iterations can only thereafter generate further positive Laplace transform contributions (the double integration of $g''_{(n+1)}$ can be regarded as a convolution of $2f_{(n)}$ with $g_{(0)}$; the operation of exponentiation is interpreted through its power series expansion and application of the convolution theorem for Laplace transforms). It remains to determine the analytic form of the converged solutions for $g(x) = g_{(\infty)}(x)$ and $f(x) = f_{(\infty)}$.

Theorem 7

The symmetrical solution to

$$g''(x) = -2\exp(g) \tag{3.26}$$

with the asymptotic limits as stated above is:

$$g_{(\infty)}(x) = \ln(f_{(\infty)}(x)) = \ln\left(\frac{1}{4\cosh^2(x/2)}\right) \equiv \ln\left(\frac{1}{4}\operatorname{sech}^2(x/2)\right). \tag{3.27}$$

□

Proof of theorem 7:

Multiply both sides by g' to obtain:

$$\frac{1}{2}[(g')^2]' = -2[\exp(g)]' \tag{3.28}$$

hence,

$$\begin{aligned} g' &= -2(\exp(g_0) - \exp(g))^{1/2} \\ &= 2\exp\left(\frac{g_0}{2}\right)[1 - \exp(g - g_0)]^{1/2}, \end{aligned} \tag{3.29}$$

for some constant g_0. Substituting

$$y = \exp\left(\frac{g - g_0}{2}\right), \tag{3.30}$$

so that

$$dg = \frac{2\,dy}{y}, \tag{3.31}$$

we obtain,

$$\int \frac{dy}{y(1-y^2)^{1/2}} = \exp\left(\frac{g_0}{2}\right)(x - x_0), \tag{3.32}$$

for some constant x_0. The integral on the left can be solved directly:

$$\int \frac{dy}{y(1-y^2)^{1/2}} = \operatorname{arccosh}(1/y), \tag{3.33}$$

and hence

$$\exp\left(\frac{g - g_0}{2}\right) = \operatorname{sech}\left(\exp\left(\frac{g_0}{2}\right)(x - x_0)\right) \tag{3.34}$$

or

$$g = g_0 + \ln\left[\operatorname{sech}^2\left(\exp\left(\frac{g_0}{2}\right)(x - x_0)\right)\right]. \tag{3.35}$$

Given that g must be symmetric about $x = 0$ and that, for large θ, $\operatorname{sech}^2(\theta) \approx \exp(-2\theta)/4$, the proper asymptotic behavior requires $x_0 = 0$ and $g_0 = \ln(1/4)$. Whence:

$$g(x) = \ln\left(\frac{1}{4}\operatorname{sech}^2\left(\frac{x}{2}\right)\right). \tag{3.36}$$

□

While this logistic or sech-squared form, $(1/4)\operatorname{sech}^2(x/2)$, supplies only a *single* representative of our family of heavy-tailed distributions, the possible alternative positive powers of the sech function allow us to control the degree of kurtosis while maintaining the property that the distribution remains a positive continuous Gaussian mixture (by theorem 6). A natural calibration of this shape-controlling exponentiation parameter, which we shall denote b, is such that $b = 1$ specifies the unadorned reference solution we have just obtained through the self-referential procedure defined and solved in theorem 7. Thus, the more general symmetric shapes of this family will be exemplified by the functions of the form,

$$f_{\mathsf{X}}(x) = \mathrm{sech}^{2b}(x/2), \tag{3.37}$$

which become proper probablity densities after normalization.

In order to ascertain precisely how the Gaussian mixture of the more general distribution is constituted, we can perform the appropriate inverse Laplace transform. By rewriting the new function in the equivalent way,

$$f_{\mathsf{X}}(x) = 2^{2b} \exp(-bx) \left(1 + \exp(-x)\right)^{-2b}, \tag{3.38}$$

transforming the variable to $z = x^2$ and correspondingly renaming the function:

$$f_{\mathsf{Z}}(z) = 2^{2b} \exp(-bz^{1/2}) \left(1 + \exp(-z^{1/2})\right)^{-2b} \tag{3.39}$$

we can expand it into an infinite series using the binomial theorem:

$$f_{\mathsf{Z}}(z) = 2^{2b} \sum_{n=0}^{\infty} \frac{(2b)_n(-)^n}{n!} \exp\left(-(b+n)z^{1/2}\right), \tag{3.40}$$

where the Pochhammer's symbol, $(c)_n$, (Abramowitz and Stegun 1972, p. 256) is defined for any real or complex c and any nonnegative integer n:

$$(c)_0 = 1, \quad (c)_1 = c, \quad \ldots, \quad (c)_n = c(c+1)\ldots(c+n-1) \equiv \frac{\Gamma(c+n)}{\Gamma(c)} \quad \ldots. \tag{3.41}$$

The individual exponential terms in the summation are each similar to the form given in (3.16) for which the inverse Laplace transform is known. This allows us to express the forcing function of the Laplace transform:

$$F_{\mathsf{S}}(s) = \frac{2^{2b-1}}{\sqrt{\pi} s^{3/2}} \sum_{n=0}^{\infty} \frac{(2b)_n(-)^n}{n!} (b+n) \exp\left(\frac{-(b+n)^2}{4s}\right). \tag{3.42}$$

(Solutions of this form can also be identified with derivatives of the theta functions and modified theta functions associated with the Jacobian elliptic functions; see Erdélyi and Bateman, 1954, Vol 1, for the example with $b = \frac{1}{2}$.) We can apply (2.5) to express the new distribution as a centered heat kernel integral (2.4) with:

$$F_{\mathsf{T}}(t) = 2^{2b+1} \sum_{n=0}^{\infty} \frac{(2b)_n(-)^n}{n!} (b+n) \exp\left(-(b+n)^2 t\right). \tag{3.43}$$

The numerical execution of approximations to the infinite summations in (3.43) and (3.42) can become problematic at large s, and small t, especially for $b \gg 1$. An alternative expression for the inverse Laplace transform comes from the standard Bromwich integral. Writing $z = iw$, this gives us:

$$F_{\mathsf{S}}(s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{iws} \mathrm{sech}^{2b}(\sqrt{iw}/2)\, dw \tag{3.44}$$

where the principal branch, $\Re(\sqrt{iw}) \geq 0$, is assumed. The corresponding integral for $F_{\mathsf{T}}(t)$ is obtained by applying to (3.44) the transformation (2.5), i.e.,

$$F_{\mathsf{T}}(t) = \frac{1}{4\sqrt{\pi}t^{3/2}} \int_{-\infty}^{\infty} e^{iw/4t} \mathrm{sech}^{2b}(\sqrt{iw}/2)\, dw \tag{3.45}$$

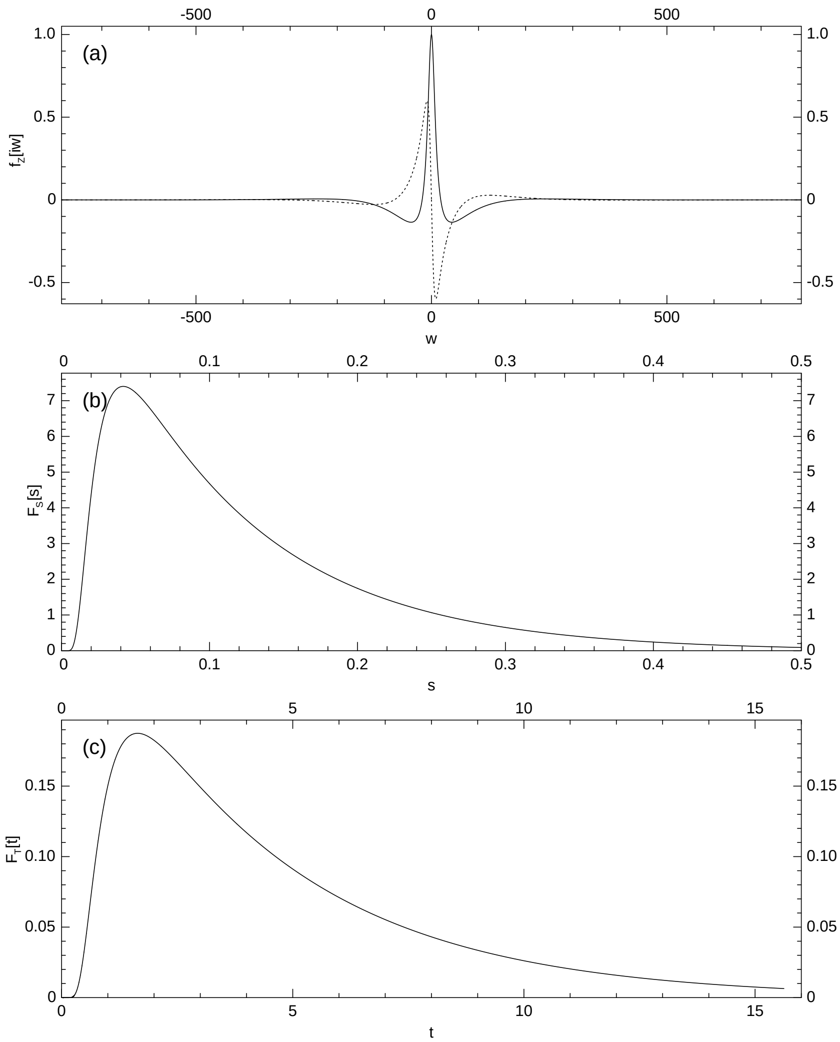

Figure 6. For the case where $b = 0.5$, the function $f_L(iw)$ is plotted in panel (a), where the solid graph is the real, the dotted graph the imaginary part. Panel (b) shows the corresponding forcing function, $F_{\mathsf{S}}(s)$, for the Laplace transform. Panel (c) shows the corresponding heating profile, $F_{\mathsf{T}}(t)$, in the heat kernel transform for $f_{\mathsf{X}}(\sqrt{z})$.

Figure 6 shows, in panel (a), the complex function $f_{\mathsf{Z}}(iw)$ plotted as real (solid graph) and imaginary (dotted graph) parts as functions of w in the case of $b = 0.5$. The corresponding $F_{\mathsf{S}}(s)$ is shown in panel (b) and the heating profile, $F_{\mathsf{T}}(t)$ for the heat transform is shown in panel (c). For this very small parameter b the forcing profiles $F_{\mathsf{S}}(s)$ and $F_{\mathsf{T}}(t)$ are both strongly skewed in the positive sense, despite the inverse relationship of their abscissae. (They happen

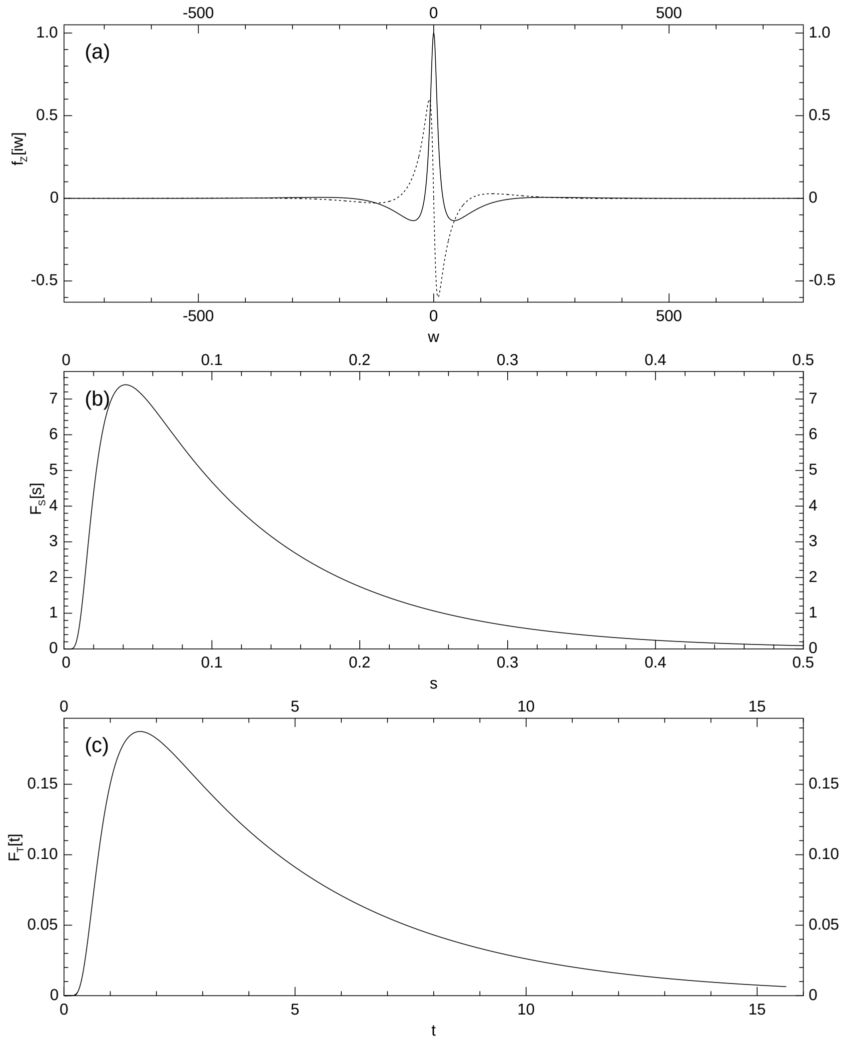

Figure 7. Like Fig.6, except with $b = 1$.

to share the same shape in this special case only, as a consequence of the sech function being its own Fourier transform.) But when the parameter $b = 1$, we obtain the corresponding set of graphs shown in Fig. 7, where we see a reduction in the skewness and the suggestion that these graphs are tending towards a more Gaussian profile. Visual corroboration is obtained when we examine the corresponding plots for a much larger value of the parameter, $b = 20$, which are provided in Fig. 8. The density function $f_{\mathsf{X}}(x)$ itself is, of course, also very close to being Gaussian, and, since the function $f_{\mathsf{Z}}(iw)$ is clearly becoming highly oscillatory for such large values of b, it is desirable to be able to characterize the asymptotic forms of the forcing profiles, $F_{\mathsf{S}}(t)$ and $F_{\mathsf{T}}(t)$ in the limit as $b \to \infty$ and thereby sidestep the inevitable practical numerical difficulties associated with either the summation method or the Fourier method as this limit is approached.

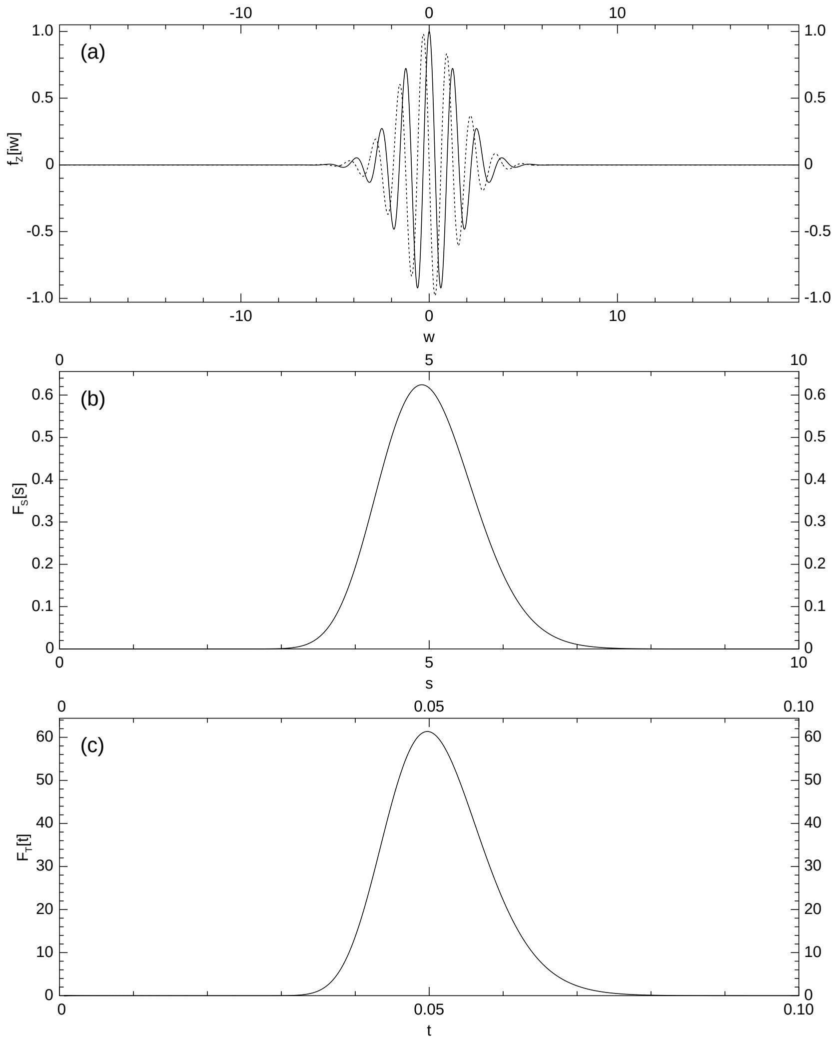

Figure 8. Like Fig.6, except with $b = 20$.

For very large values of the parameter, b, it is possible to approximate these integrals by exploiting the fact that the absolute magnitude of the integrand then peaks sharply at $z = iw = 0$ and becomes relatively insignificant beyond a small distance from this point. This allows the integrand, or more conveniently its logarithm, to be expanded in a locally convergent Taylor series about this point. Formally:

$$\begin{aligned} \ln(\mathrm{sech}^{2b}(\sqrt{z}/2)) &= -b\sum_{n=1}^{\infty}\frac{(2^{2n}-1)}{n(2n)!}B_{2n}z^n, \qquad (3.46a) \\ &= -\frac{bz}{4}+\frac{bz^2}{96}+\mathcal{O}(z^3) \end{aligned}$$

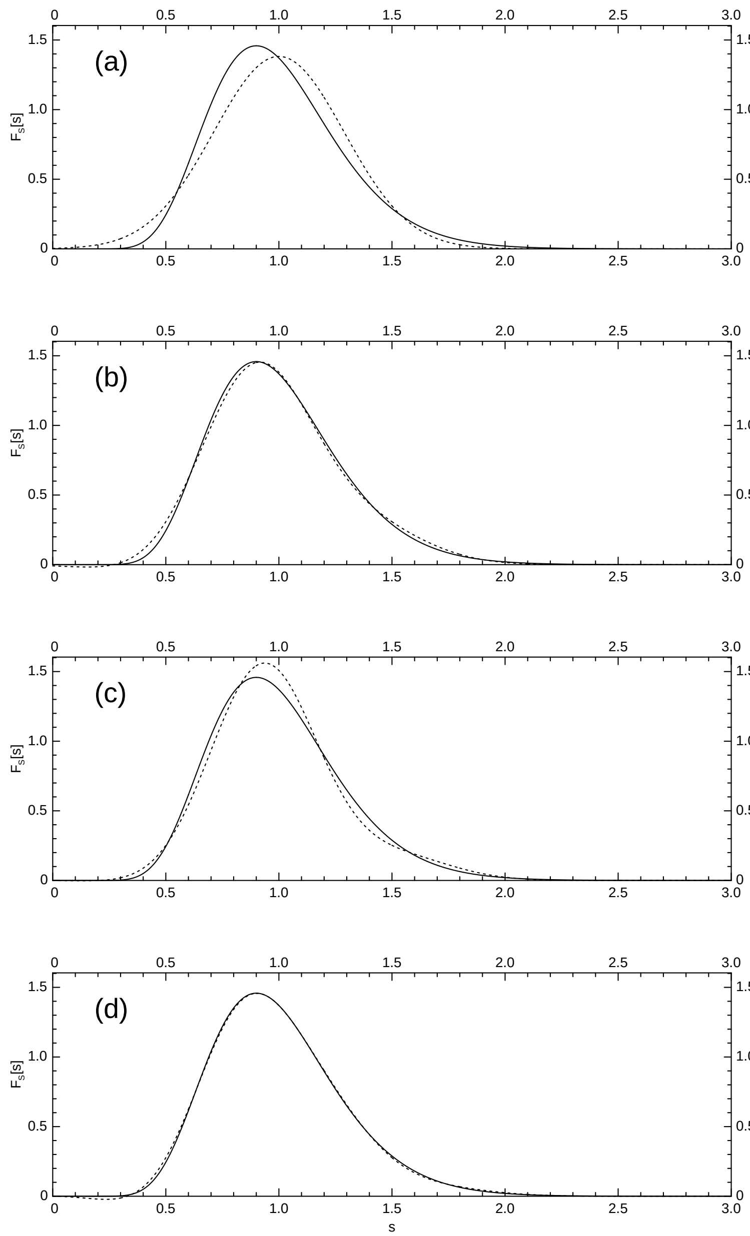

Figure 9. For the case of the parameter having an intermediate magnitude, $b = 2$, panel (a) shows $F_{\mathsf{S}}(s)$ together with the purely Gaussian approximation to it. Panels (b) and (c) show the comparison with the third-order and fourth-order Gram-Charlier expansion. Panel (d) shows the significant improvement in the fit obtained by using the fourth-order Edgeworth expansion instead. The exact $F_{\mathsf{S}}(s)$ is shown as a solid curve in each panel, while the approximation is shown dotted.

$$= \sum_{n=1}^{\infty} (-)^n \frac{\hat{\lambda}_n(b) z^n}{n!}, \tag{3.46b}$$

where B_{2n} denotes the Bernoulli number of index $2n$ (see Abramowitz and Stegun 1972, p.810, for a tabulation of these numbers) and $\hat{\lambda}_n(b)$, as defined above, are the cumulants of the

TABLE 1. Coefficients of the Hermite polynomials $H_n(x)$ up to degree $n = 9$.

n	0	1	2	3	4	5	6	7	8	9
0	1									
1		1								
2	−1		1							
3		−3		1						
4	3		−6		1					
5		15		−10		1				
6	−15		45		−15		1			
7		−105		105		−21		1		
8	105		−420		210		−28		1	
9		945		−1260		378		−36		1

distribution, $F_{\mathsf{S}}(s)$, as defined in section 2. For sufficiently large b it is these first two terms of the expansion (3.46a) that dominate, that is, a Gaussian form of f_{Z} when expressed in the w variable, and therefore in its transform, $F_{\mathsf{S}}(s)$, also. Taking the exponential of (3.46a) and dividing out the part responsible for the Gaussian enables us to express the perturbation factor formally:

$$\frac{\operatorname{sech}^{2b}(\sqrt{z}/2)}{\exp\left(-\frac{bz}{4} + \frac{bz^2}{96}\right)} = \exp\left(1 + \sum_{n=3}^{\infty} (-)^n \hat{\lambda}_n(b) z^n\right) \tag{3.47}$$

Expanding terms on the right in successive powers, $(-z)^n \equiv (d/ds)^n$, and truncating at some finite degree, n, enables the resulting series to be interpreted in the s-domain as a corrective differential operator applied to the Gaussian,

$$G_b(s) = \frac{1}{\sqrt{2\pi\lambda_2}} \exp\left(-\frac{1}{2}\frac{(s-\hat{\lambda}_1)^2}{\hat{\lambda}_2}\right). \tag{3.48}$$

Successive derivatives of a Gaussian are expressed by multiplying that Gaussian by Hermite polynomials (Koornwinder et al. 2010). For example, the Hermite polynomial of degree m is defined in the form most convenient for probability theory:

$$H_m(x) = (-)^m e^{x^2/2} \frac{d^m}{dx^m} e^{-x^2/2}. \tag{3.49}$$

Written as

$$H_m(x) = \sum_{j=0}^{m} H_{m,j} x^j, \tag{3.50}$$

the coefficients are found to be, for example:

$$H_{m,j} = \begin{cases} (-)^{m-j}(m-j-1)!!\binom{m}{j} & : \quad m-j \quad \text{even} \\ 0 & : \quad m-j \quad \text{odd} \end{cases}, \tag{3.51}$$

and satisfy the convenient recursion:

$$\begin{aligned} H_{0,j} &= \delta_{0j} \\ H_{m,j} &= H_{m-1,j-1} - (j+1)H_{m-1,j+1}, \quad m > 0. \end{aligned}$$

The coefficients of the first few of these polynomials are listed in Table 1. The resulting corrective series is known as the Gram-Charlier series (Kendall and Stuart 1977). If we let $\sigma = \sqrt{\hat{\lambda}_2}$, and rescale s about the mean, $s' = (s - \hat{\lambda}_1)/\sigma$, then the correction takes the form of a corrective factor, C, multiplying the approximating Gaussian:

$$F_{\mathfrak{S}}(s) \approx C \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}s'^2\right), \tag{3.52}$$

where, up to sixth order, the corrective factor is the expansion,

$$C = 1 + \frac{\hat{\lambda}_3}{3!\sigma^3}H_3(s') + \frac{\hat{\lambda}_4}{4!\sigma^4}H_4(s') + \frac{\hat{\lambda}_5}{5!\sigma^5}H_5(s') + \left(\frac{\hat{\lambda}_6}{6!\sigma^6} + \frac{\hat{\lambda}_3^2}{2!(3!)^2\sigma^6}\right)H_6(s'). \tag{3.53}$$

(Note that this expansion involves products of cumulants at terms at and beyond the sixth order.)

Figure 9 illustrates the effect of various approximations to $F_{\mathfrak{S}}(s)$ in the case of a parameter of intermediate magnitude, $b = 2$. The solid lines reproduce the true $F_{\mathfrak{S}}(s)$ in each panel while the approximations are shown dotted. In panel (a) we see the Gaussian approximation, which obviously cannot capture the pronounced asymmetry of the true distribution. The skewness is well captured by the third-order expansion shown in panel (b), but a small amount of negativity is introduced on the left (this erroneous side effect is greatly reduced in larger values of parameter b). Unfortunatly, application of the fourth-order Gram-Charlier expansion, shown in (c), makes the apparent fit noticeably worse, and higher orders of this expansion do not greatly improve matters. Such problems with the Gram-Charlier expansion are well known by statisticians, but an ingenious remedy was sought and found by Edgeworth (1904) in the context of approximations to standardized (rescaled to unit variance) sums of N independent identically-distributed random variables, where the objective is to characterize the approach to Gaussian form in greater detail than is given by the simple statement of the central limit theorem. In this case, while cumulants of the unstandardized distribution of sums combine additively, the standardization scales the magnitude of the mth cumulant relative to that of the single member by a factor of $1/N^{(m-2)/2}$. Edgeworth reordered, and subsequently truncated, the series according to successive powers of $1/\sqrt{N}$:

$$\begin{aligned} C_{(N)} \equiv 1 + \frac{\hat{\lambda}_3}{3!\sigma^3 N^{1/2}}H_3(s') + \frac{\hat{\lambda}_4}{4!\sigma^4 N}H_4(s') &+ \frac{\hat{\lambda}_5}{5!\sigma^5 N^{3/2}}H_5(s') \ldots \\ + \frac{\hat{\lambda}_3^2}{2!(3!)^2\sigma^6 N}H_6(s') &+ \frac{\hat{\lambda}_3\hat{\lambda}_4}{3!4!N^{3/2}}H_7(s') \ldots \\ &+ \frac{\hat{\lambda}_3^3}{3!(3!)^3 N^{3/2}}H_9(s') \ldots, \end{aligned} \tag{3.54}$$

whereupon, even taking $N = 1$, a different partial series is obtained when truncated beyond the first nontrivial (third order) term. The panel (d) of Fig. 9 shows the result of applying the fourth-order Edgeworth series approximation,

$$C_{(1)} \approx 1 + \frac{\hat{\lambda}_3}{6\sigma^3} H_3(s') + \left(\frac{\hat{\lambda}_4}{24\sigma^4} H_4(s') + \frac{\hat{\lambda}_3^2}{72\sigma^6} H_6(s') \right), \tag{3.55}$$

which vastly improves the overall fit (although the spurious negative amplitude close to the origin is made larger). The fit is found further improved upon including higher order terms of this Edgeworth series up to the practical limit of about eighth order. For closer approximations to the Gaussian, that is, with $b > 2$ we find that graphs of the function and its approximation at eighth order become visibly indistinguishable. However, we should be reminded that the series is asymptotic and not formally convergent at *any* b (Cramér 1926, 1928, 1946). Blinnikov and Moessner (1998), who discuss the advantages of the Edgeworth series over Gram-Charlier and other expansions in a very different spectroscopic context, also outline some efficient algorithms.

The heat transform forcing function, $F_{\mathsf{T}}(t)$, can also be approximated by asymptotic series. In this case, the Edgeworth series is constructed from the cumulants of $F_{\mathsf{T}}(t)$, which are also then given in terms of the even-degree cumulants of $f_{\mathsf{X}}(x)$ through the application of theorem 4 and, as derived in appendix B, of the even-degree cumulants $\hat{\chi}_{2q}$. In this case, we need to take $b \geq 10$, in order that the asymptotic Edgeworth expansion gives an adequate fit to the exact solution. Figure 10 shows some comparisons for this b with, again, the true solution shown solid in each panel and the approximation dotted. Panel (a) shows the Gaussian approximation, (b) shows the third-order Edgeworth, which essentially corrects for the skewness; (c) shows the fourth-order correction and (d) shows the excellent fit when the eighth-order expansion is applied.

4. Skewed generalizations of the basic heavy-tailed densities

It might be tempting to expect that, just as the sech-squared representative of the symmetric distribution was generated by the self-referential convolution procedure described in section 3, a similar procedure might be used to generalize this representative function to a skewed version when the symmetric starting function, $g_{(0)}(x) = -|x|$, is replaced by the asymmetric modification, $g_{(0)}(a; x) = -|x| + ax$ for a skewness-controlling parameter, $a \in (-1, 1)$. However, this is not recommended; the iteration needs to be stabilized by correcting *both* of the first two moments, χ_0 and χ_1 of the successive iterations of $f_{(n)}(a; x) = \exp(g_{(n)})$ and, by introducing a corrective x-displacement, we destroy any chance of acquiring a family of solutions that submit to simple analytic description; but what is worse, we cannot even guarantee that the resulting converged solution, $f_{(\infty)}(a; x)$, is expressible in the form of a mobile point-source heat kernel representation. We must therefore opt for another approach.

In order to induce a finite measure of skewness for any of the family of distributions $f_{\mathsf{X}}(x)$ of section 3 associated with the parameter b it is sufficient to add to the logarithm, $g_{\mathsf{X}}(x) = \ln[f_{\mathsf{X}}(x)]$, of each symmetric member of this family a component linear in x. This manifestly preserves the valuable property of convexity and, since it is equivalent to multiplying each symmetric distribution by an exponential function, each Gaussian component in a heat

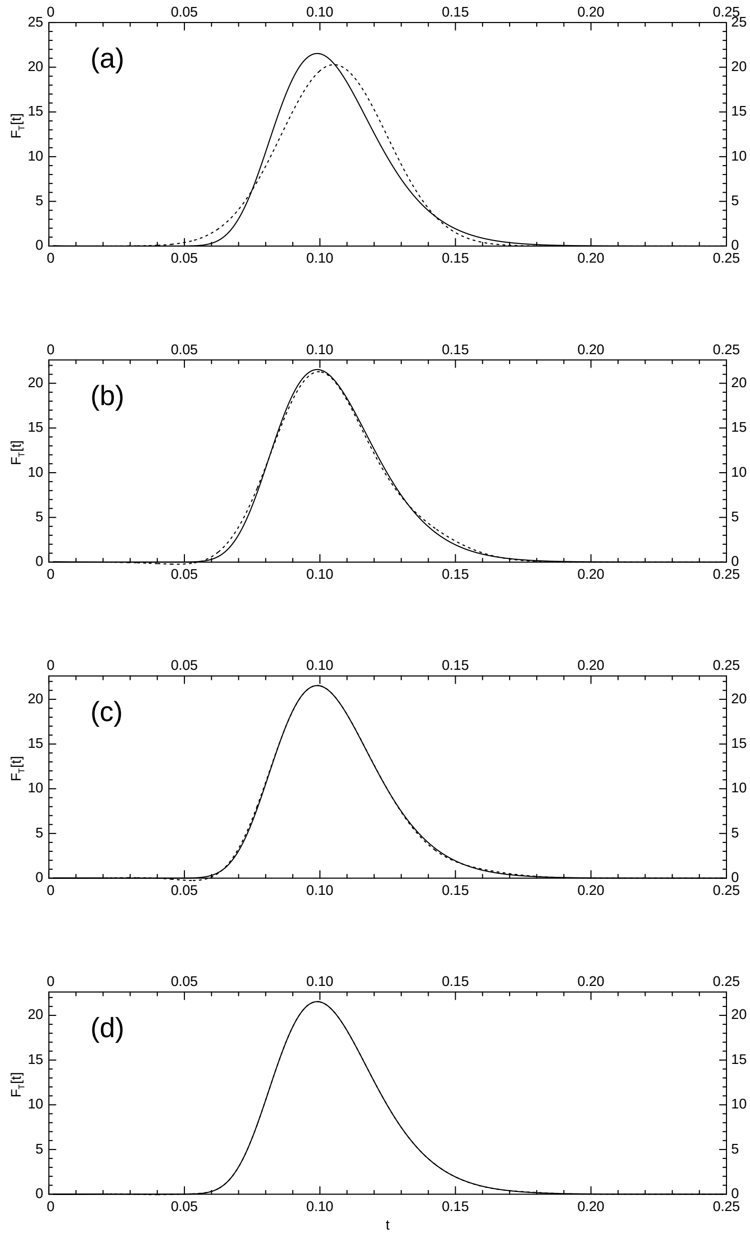

Figure 10. For the case of the parameter having an intermediate magnitude, $b = 10$, panel (a) shows $F_{\mathsf{T}}(t)$ (solid) together with the Gaussian approximation to it (dotted). In panels (b), (c) and (d) the Gaussian approximation is improved through the application of the Edgeworth series to orders 3, 4, and 8 respectively.

transform is changed only by an amplification and a displacement. In addition to the asymmetry parameter, a, we can include a positive width-scaling parameter, ω, so that the enlarged family of scaled (but still unnormalized) shapes might be expressed by:

$$f_{\mathsf{X}}(\omega, a, b; x) = \exp(abx)\,\mathrm{sech}^{2\omega b}\left(\frac{x}{2\omega}\right), \tag{4.1}$$

with $a \in (-1, 1)$. Then *all* of the moments of f_{X} exist, and possess the analytic expressions computed (for $\omega = 1$, since they are then trivially generalized to other ω) in Appendix B. The moments up to fourth degree are used in the calculation of kurtosis, enabling us to plot the images of each parameter pair, (a, b), in the skewness-kurtosis plane. The boundary of the range of such images is exactly that shown in Fig. 3, with the inscribed curvilinear grid being contours of constant a and of constant b. The lower bounding curve is approached in the limit $a \to \pm 1$ where the form of the distribution becomes progressively closer to the exponentiated Gumbel. Thus, we have fulfilled the desire to generalize the pure Gaussian in terms of a positive Gaussian mixture that preserves the convexity of the negative-log-probability, has the exponentiated Gumbel as limiting forms, has all moments finite and expressible analytically, and is achieved parsimoniously with just two parameters responsible for shape.

The alteration of the heat kernel can be understood by comparing the kernel with $a \neq 0$ at fixed x and t with the corresponding kernel (same b) for the symmetric case, $a = 0$, that we dealt with in section 3. Thus, for the case where the scale is standardized, $\omega = 1$, the integrand of the heat transform parameterized as $W_{\mathsf{XT}}(\omega, a, b; x, t)$ is modified according to:

$$
\begin{aligned}
W_{\mathsf{XT}}(1, a, b; x, t) F_{\mathsf{T}}(1, a, b; t) &\equiv e^{abx} W_{\mathsf{XT}}(1, 0, b; x, t) F_{\mathsf{T}}(1, 0, b; t) \\
&= \frac{1}{\sqrt{4\pi t}} \exp\left(abx - \frac{x^2}{4t}\right) F_{\mathsf{T}}(1, 0, b; t) \\
&= \frac{1}{\sqrt{4\pi t}} \exp\left(-\left[\left(\frac{x}{2\sqrt{t}} - ab\sqrt{t}\right)^2 - a^2 b^2 t\right]\right) F_{\mathsf{T}}(1, 0, b; t) \\
&= \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{(x - \hat{x}(t))^2}{4t}\right) A(t) F_{\mathsf{T}}(1, 0, b; t),
\end{aligned}
\tag{4.2}
$$

where the trajectory of the effective heat forcing is given by,

$$
\hat{x}(t) \equiv \hat{x}(1, a, b; t) = 2abt, \tag{4.3}
$$

and the amplification factor is:

$$
A(t) \equiv A(1, a, b; t) = e^{a^2 b^2 t}. \tag{4.4}
$$

In practice, we partition terms such that:

$$
W_{\mathsf{XT}}(1, a, b; x, t) = \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{(x - \hat{x}(1, a, b; t))^2}{4t}\right), \tag{4.5}
$$

and

$$
F_{\mathsf{T}}(1, a, b; t) = A(1, a, b; t) F_{\mathsf{T}}(1, 0, b; t). \tag{4.6}
$$

For a general scaling, ω, we can relate the mobile heat transform to that of the standardized case:

$$
f_{\mathsf{X}}(\omega, a, b; x) = \frac{1}{\omega} \int_0^\infty \frac{1}{\sqrt{4\pi t'}} \exp\left[-\frac{(x/\omega - \hat{x}(1, a, \omega b; t'))^2}{4t'}\right] F_{\mathsf{T}}(1, a, \omega b; t')\, dt'
$$

$$
\begin{aligned}
&= \int_0^\infty \frac{1}{\sqrt{4\pi\omega^2 t'}} \exp\left[-\frac{(x-\omega\hat{x}(1,a,\omega b;t'))^2}{4\omega^2 t'}\right] \frac{F_{\mathsf{T}}(1,a,\omega b;t')}{\omega}\,\omega^2 dt' \\
&= \int_0^\infty \frac{1}{\sqrt{4\pi t}} \exp\left[-\frac{(x-\omega\hat{x}(1,a,\omega b;t/\omega^2))^2}{4t}\right] \frac{F_{\mathsf{T}}(1,a,\omega b;t/\omega^2)}{\omega}\,dt \\
&= \int_0^\infty \frac{1}{\sqrt{4\pi t}} \exp\left[-\frac{(x-\hat{x}(\omega,a,b;t))^2}{4t}\right] F_{\mathsf{T}}(\omega,a,b;t)\,dt, \qquad (4.7)
\end{aligned}
$$

(4.8)

where

$$\hat{x}(\omega,a,b;t) = \omega\hat{x}(1,a,\omega b;t/\omega^2) = 2abt, \qquad (4.9)$$

and

$$
\begin{aligned}
F_{\mathsf{T}}(\omega,a,b;t) &= \frac{F_{\mathsf{T}}(1,a,\omega b;t/\omega^2)}{\omega} \\
&= A(1,a,\omega b;t/\omega^2)\frac{1}{\omega}F_{\mathsf{T}}(1,0,\omega b;t/\omega^2) \\
&= A(\omega,a,b;t)F_{\mathsf{T}}(\omega,0,b;t), \qquad (4.10)
\end{aligned}
$$

with

$$A(\omega,a,b;t) = A(1,a,\omega b;t/\omega^2) = e^{a^2b^2t} \equiv A(1,a,b;t). \qquad (4.11)$$

These results for the general $\hat{x}$ and A lead to the following.

Theorem 8

The amplification profile, $A(\omega,a,b;t)$ and the trajectory, $\hat{x}(\omega,a,b;t)$ are the same for all skewed members of this family of distributions, $f_{\mathsf{X}}(\omega,a,b;x)$, for which the corresponding log-probability densities, $g_{\mathsf{X}}(\omega,a,b;x)$, have the pair of asymptotes (apart from constant displacements) in common.

□

Proof of theorem 8

We find that the asymptotes (apart from additive constants coming from normalization and choice of bias) obey:

$$g_{\mathsf{X}}(\omega,a,b;x) \approx \begin{cases} -(1+a)b|x| & : \quad x \to -\infty \\ -(1-a)b|x| & : \quad x \to +\infty \end{cases}, \qquad (4.12)$$

and therefore define a and b. But we see from (4.9) and (4.11) that the trajectory, $\hat{x}$ and the amplification profile, A, also depend only upon a and b (and are then independent of ω).

□

Remarks

Since we have established that, for this Gaussian-mixture probability family, neither the trajectory function $\hat{x}$ nor the amplification factor, A, depend upon ω once a and b are given, we can drop ω from their list of parameters.

5. Systematic generalization of heavy-tailed distributions to include convexity control

In order to resolve the problem of how to choose the most natural heavy-tailed distribution scaling like $\exp(-|x|)$ at large $|x|$ we sought to find the function $f_{\mathsf{X}}(x)$ that, when convolved with $-|x|$, produced the logarithm, $g_{\mathsf{X}}(x)$, of that same $f_{\mathsf{X}}(x)$. In principle there is no reason why the same approach should not also work with an alternative more general scaling of the asymptotes and the application of the corresponding smoothing kernel implicitly being the exponential of the smoothed $-|x|^c$ profile. Arguments similar to those used in section 3 will continue to justify the assertion that the resulting broad-tailed distributions are still positive Gaussian mixtures. However, we do not need to seek a *new* smoothing kernel in this convoluted manner since the logistic distribution, in its role as a smoothing kernel, remains perfectly serviceable and, being already a well-characterized analytic function, will lead to a much simpler and more elegant generalization of our existing logistic-based family.

Recognizing the significance of the logistic distribution function as the common smoothing kernel for our generalized symmetric family, we shall adopt the notation for a scaled and normalized version of it:

$$\ell_{\mathsf{X}}(\omega; x) = \frac{1}{4\omega}\text{sech}^2\left(\frac{x}{2\omega}\right). \tag{5.1}$$

Our generalizations of f_{X} and g_{X} must carry the additional parameter, c, controlling the asymptotic exponent, and hence the degree of convexity. In the limit where the scaling parameter, $\omega \to 0$, i.e., no smoothing applied, we define:

$$g_{\mathsf{X}}(0, a, b, c; x) = \begin{cases} -[(1+a)b|x|]^c & : \quad x < 0 \\ -[(1-a)b|x|]^c & : \quad x > 0 \end{cases}, \tag{5.2}$$

The smoothed generalization for symmetric g_{X} then comes from the x-convolution with $\ell_{\mathsf{X}}(\omega; x)$, which we express symbolically:

$$g_{\mathsf{X}}(\omega, 0, b, c; \bullet) = g_{\mathsf{X}}(0, 0, b, c; \bullet) \star \ell_{\mathsf{X}}(\omega; \bullet) \tag{5.3}$$

or as an explicit integral,

$$g_{\mathsf{X}}(\omega, 0, b, c; x) = \int_0^\infty (bx')^c \left(\ell_{\mathsf{X}}(\omega; x - x') + \ell_{\mathsf{X}}(\omega; x + x')\right)\, dx'. \tag{5.4}$$

The variations of shape with parameters ω and b amount to little more than rescalings, since,

$$g_{\mathsf{X}}(\omega, 0, b, c; x) = (\omega b)^c g_{\mathsf{X}}(1, 0, 1, c; x/\omega). \tag{5.5}$$

We are assured that the symmetric $a = 0$ functions, g_{X}, defined in this way are positive Gaussian mixtures since ℓ_{X} is, and the g_{X} with $\omega = 0$ are, and the centers of the component Gaussians are all at the origin. It then follows that the corresponding exponentiated functions, $f_{\mathsf{X}} = \exp(g_{\mathsf{X}})$, are also positive Gaussian mixtures. However, we cannot as yet say anything about the asymmetrical cases where $a \neq 0$ since the foundation of the former proofs lay upon the assumption that the convolved Gaussians all share a common center of symmetry, which no longer pertains.

We can begin to make some progress for the restricted, but practically most interesting parameter range, $c<1$, and for the unsmoothed shapes that correspond to the limiting case $\omega=0$. Instead of looking at g_{X} we shift the focus to the actual density distribution, f_{X}, and establish the following result.

Theorem 9

For $c\in(0,1)$ and $a\in(-1,1)$ the function,

$$f_{\mathsf{X}}(0,a,b,c;x)=\begin{cases}\hat{e}_-(x)=\exp\left\{-[(1+a)b|x|]^c\right\} & : \quad x<0\\ \hat{e}_+(x)=\exp\left\{-[(1-a)b|x|]^c\right\} & : \quad x>0\end{cases}, \tag{5.6}$$

has a mobile heat transform representation,

$$f_{\mathsf{X}}(0,a,b,c;x)=\int_0^\infty \frac{1}{4\pi t}\exp\left(-\frac{[x-\hat{x}(t)]^2}{4t}\right)F_{\mathsf{T}}(0,a,b,c;t)\,dt, \tag{5.7}$$

for some trajectory, $\hat{x}(t)$, and some forcing profile, $F_{\mathsf{T}}(t)>0$ defined for all $t>0$.

□

Proof of theorem 9

By a direct adaptation of theorem 6 for Laplace transforms we can express $\hat{e}_\pm$ as positive continuous mixtures of exponentials:

$$\hat{e}_\pm(x)\equiv\hat{e}_\pm(x,0)=\int_0^\infty \hat{E}_\pm(u)\exp(\mp ux)\,du. \tag{5.8}$$

with $\hat{E}_\pm(u)>0$, where the more general functions,

$$\hat{e}_\pm(x,t)=\int_0^\infty \hat{E}_\pm(u)\exp(\mp ux-u^2t)\,du, \tag{5.9}$$

manifestly satisfy the homogeneous diffusion equation,

$$-\frac{\partial}{\partial t}\hat{e}_\pm(x,t)=\frac{\partial^2}{\partial x^2}\hat{e}_\pm(x,t). \tag{5.10}$$

At any $t\geq 0$ the two functions, $\exp(\mp ux-u^2t)$ have the property referred to by Widder (1941) as being 'completely convex' in x; their derivatives of each order n exist, are positive for even n, and are of the same sign for every odd n. Since this property is obviously inherited by any positive combination of such functions, it is possessed by both branches, $\hat{e}_\pm(x,t)$ at every $t\geq 0$ and at *every* x regardless of sign. The left branch increases monotonically with x from a zero asymptote at $x\to-\infty$, while the right branch increases monotonically in $-x$ from a zero asymptote at $x\to\infty$, so the graphs of the two curves are guaranteed to intersect at a unique $\hat{x}(t)$ at each t. A delta-function heat source of positive magnitude

$$F_{\mathsf{T}}(0,a,b,c;t)=\left(\frac{\partial\hat{e}_-(x,t)}{\partial x}-\frac{\partial\hat{e}_+(x,t)}{\partial x}\right)\Bigg|_{x=\hat{x}(t)} \tag{5.11}$$

is consistent with the *evolving* solution:

$$f_{\mathsf{X}}(0,a,b,c;x,t) = \begin{cases} \hat{e}_-(x,t) & : \quad x \le \hat{x}(t) \\ \hat{e}_+(x,t) & : \quad x \ge \hat{x}(t). \end{cases} \qquad (5.12)$$

to the forced diffusion equation:

$$\left(-\frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2}\right) f_{\mathsf{X}}(0,a,b,c;x,t) = F_{\mathsf{T}}(0,a,b,c;t)\delta(x-\hat{x}(t)). \qquad (5.13)$$

where the trajectory of the forcing, $\hat{x}(t)$, is implicitly defined by the intersection condition,

$$\hat{e}_-(\hat{x}(t)) = \hat{e}_+(\hat{x}(t)), \quad \text{for all } t > 0. \qquad (5.14)$$

□

We cannot extend this result to the complementary half of the range $c \in (1,2)$ using Laplace transforms in the same way. If there *is* a solution for a given $a > 1$ and b, an alternative strategy to finding it might be to express the $\hat{e}_\pm$ at each t as Taylor-Maclaurin series in x (about $x = 0$), and search for their intersection numerically, trusting that the intersection lies inside the radius of convergence. The required spatial derivatives can be obtained by first considering each branch $\hat{e}_+$ and $\hat{e}_-$ separately to be the respective halves of the solutions of their own *symmetric* heat transform problems. The principle will be illustrated for the branch, $\hat{e}_+$, and we will simplify the formulae by using the substitution, $b'_+ = (1-a)b$, whereupon we have:

$$\hat{e}_+(x,t) = f_{\mathsf{X}}(0,0,b'_+,c;x,t) = \int_0^\infty \frac{1}{\sqrt{4\pi t'}} \exp\left(-\frac{x^2}{4t'}\right) F_{\mathsf{T}}(0,0,b'_+,c;t+t')\,dt', \quad x \ge 0. \qquad (5.15)$$

(the Gaussian kernel $W_{\mathsf{XT}}(x,t')$ for the heat transform becomes the convolution kernel Green's function for this t-dependent solution of the diffusion equation.) From the evaluation at $x = 0$ we obtain,

$$\hat{e}_+(0,t) = \frac{1}{2}\int_0^\infty \frac{1}{\sqrt{\pi t'}} F_{\mathsf{T}}(0,0,b'_+,c;t+t')\,dt', \qquad (5.16)$$

while, from symmetry and the heat-flux condition at $x = 0$ we obtain,

$$\frac{\partial}{\partial x}\hat{e}_+(0,t) = -\frac{1}{2}F_{\mathsf{T}}(0,0,b'_+,c;t). \qquad (5.17)$$

But since for $x \ne 0$ the solution is known to satisfy the homogeneous diffusion equation, we can iterate this equation as often as we like and exploit the property that the t-derivative commutes with the convolution operator in (5.15), to deduce that,

$$\frac{\partial^{2n}}{\partial x^{2n}}\hat{e}_+(0,t) \equiv \frac{\partial^n}{\partial(-t)^n}\hat{e}_+(0,t) = \frac{1}{2}\int_0^\infty \sqrt{\pi t'} F_{\mathsf{T}}^{(n)}(0,0,b'_+,c;t+t')\,dt' \qquad (5.18a)$$

$$\frac{\partial^{2n+1}}{\partial x^{2n+1}}\hat{e}_+(0,t) \equiv \frac{\partial^n}{\partial(-t)^n}\frac{\partial \hat{e}_+(0,t)}{\partial x} = -\frac{1}{2}F_{\mathsf{T}}^{(n)}(0,0,b'_+,c;t), \qquad (5.18b)$$

where the derivatives in the *negative* t direction are denoted:

$$F_{\mathsf{T}}^{(n)}(0,0,b'_+,c;t) \equiv \frac{d^n}{d(-t)^n} F_{\mathsf{T}}(0,0,b'_+,c;t). \tag{5.19}$$

We can restate the formulae above rather elegantly as series of the fractional calculus operations with respect to $\{-t\}$, applied to F_{T}. The relevant definitions, which we explain in the following, are in this instance clesely related to the Weyl fractional integral transforms (Erdélyi and Bateman, Vol 2, 1954, Oldham and Spanier, 1974). Given a finite function, $H(t)$, decaying at a faster-than-algebraic rate as $t \to +\infty$, we can consistently define for *any* positive or negative integer, q, the qth-derivative. For $q \geq 0$ we have the conventional derivative,

$$H^{(q)}(t) \equiv \frac{d^q H(t)}{d(-t)^q}. \tag{5.20}$$

For $q < 0$ we can iterate successive integrals, which we express formally:

$$H^{(q)}(t) \equiv \int_0^\infty H^{(q+1)}(t+t')\,dt' \equiv \int_t^\infty H^{(q+1)}(t')\,dt'. \tag{5.21}$$

Then, regardless of the signs of q and r, we find:

$$\frac{d^{q+r}H}{d(-t)^{q+r}} \equiv \frac{d^q}{d(-t)^q}\frac{d^r H}{d(-t)^r}. \tag{5.22}$$

Fractional calculus consistently extends this identity to apply when q and r are arbitrary non-integer reals, by first defining:

$$H^{(r)} = \int_0^\infty \frac{t'^{-r-1}}{\Gamma(-r)} H(t+t')\,dt', \quad r \in (-1,0) \tag{5.23}$$

and generalizing this in integer steps, q, by assuming (5.22) to hold and applying either (5.20) or (recursively) (5.21). In particular, since $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, we find that we can formally express the value and all the x-derivatives of $\hat{e}_+$ at the origin by the single general formula:

$$\frac{\partial^n}{\partial x^n}\hat{e}_+(0,t) = (-)^n \frac{1}{2} F_{\mathsf{T}}^{(\frac{n-1}{2})}(0,0,b'_+,c;t). \tag{5.24}$$

Similarly, for the function branch, $\hat{e}_-(x,t)$, we define $b'_- = (1+a)b$ and, after allowance has been made for the fact that the x-derivatives we need are now those on the *negative* side of the origin (odd x-derivatives switch sign), we obtain, in the same way,

$$\frac{\partial^n}{\partial x^n}\hat{e}_-(0,t) = \frac{1}{2} F_{\mathsf{T}}^{(\frac{n-1}{2})}(0,0,b'_-,c;t). \tag{5.25}$$

Whether the Taylor-Maclaurin series converges for a sufficient radius to enable an intersection remains an open question, as does the existence of an intersection at all for large values of t, since, without the 'completely convex' property, we no longer have the guarantee that the analytically-continued $\hat{e}_-(x,t) - \hat{e}_+(x,t)$ even increases monotonically with x when $c > 1$.

Numerical experimentation could provide some insight into these questions of the existence of a heat-transform solution when $c > 1$.

Once we are in possession of the forcing, $F_{\mathsf{T}}(0, a, b, c; t)$, we can define the amplification profile for this $\omega = 0$ case:

$$A(a, b, c; t) = \frac{F_{\mathsf{T}}(0, a, b, c; t)}{F_{\mathsf{T}}(0, 0, b, c; t)}. \tag{5.26}$$

This quantity is defined for all t because $F_{\mathsf{T}}(0, 0, b, c; t)$ is positive for all positive $t = 1/4s$ by virtue of theorem 6. Since the trajectory, $\hat{x}$ and the amplification, A, were found, in the simpler $c = 1$ case, not to be functions of ω once a and b were defined (theorem 8), we make this property the defining principle for the generalization to $c \neq 1$ also. Thus, can we define,

$$F_{\mathsf{T}}(\omega, a, b, c; t) = A(a, b, c; t) F_{\mathsf{T}}(\omega, 0, b, c; t) > 0, \tag{5.27}$$

and construct the (unnormalized) distribution, $f_{\mathsf{X}}(\omega, a, b, c; x)$, according to the mobile heat transform,

$$f_{\mathsf{X}}(\omega, a, b, c; x) = \int_0^\infty \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{[x - \hat{x}(a, b, c; t)]^2}{4t}\right) F_{\mathsf{T}}(\omega, a, b, c; t)\, dt, \tag{5.28}$$

assured that it comprises a positive Gaussian mixture, as required.

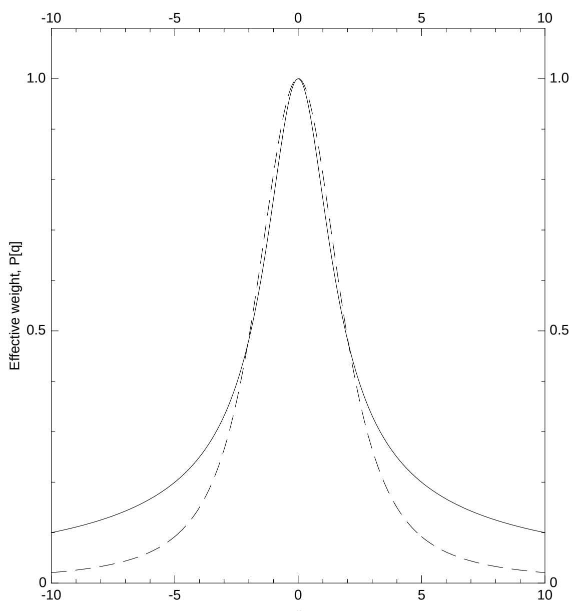

Figure 11. The solid curve shows the graph of the standardized effective weight function implied by the parameter combination $(a, b, c) = (0, b, 1)$, where the value of b does not affect the shape. For comparison, the dashed curve shows another symmetric case, but with parameters $(0, b, 1/25)$. The principal difference in shape occurs at the limbs where $O - A$ becomes relatively large in magnitude. The former graph has a profile that will never cause multimodality in the variational assimilation's cost function, and therefore should lead to a more robsut system.

6. Discussion and conclusion

Guided by a small set of basic, very simple principles, and motivated by the desire to achieve the benefits of nonlinear quality control without the hazards associated with multiple minima of the cost function, we have constructed an analytic family of symmetric distributions of measurement error based on positive powers of the sech function, and the asymmetric skewed generalizations of these obtained by multiplying by the exponential function of this error. Apart from the scaling parameter ω, this basic system of distributions requires just two shape parameters: the asymmetry, a, and the tail-broadness parameter, b. The symmetric ($a = 0$) examples will serve most purposes and will guarantee that the unimodality of the cost function will be preserved, even without including a formal loss model. In this case, the effective weight attributed to a given observation will not be a constant function of observation minus analysis, as it would be in the case of purely Gaussian errors, but, instead, should exhibit a bell-shaped modulation. To see the form of the effective weight, let us suppose that the observation residual for the model with parameter b is first rescaled by a factor, $\sqrt{b/2}\sigma$, so that the log-probability, apart from an additive constant, is just:

$$g_{\mathsf{x}}(x) \propto -2b \ln \left(\cosh \left[\frac{x}{\sqrt{2b}\sigma} \right] \right). \tag{6.1}$$

The contribution of this particular measurement to the gradient of the cost function when x is interpretted as the observation-minus-analysis is then

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{2b}{\sqrt{2b}\sigma} \tanh \left[\frac{x}{\sqrt{2b}\sigma} \right], \tag{6.2}$$

and the effective precision-weight of the observation is:

$$\frac{1}{x}\frac{\partial \mathcal{L}}{\partial x} = \frac{1}{\sigma^2} \mathcal{P} \left(\frac{x}{\sqrt{2b}\sigma} \right) \tag{6.3}$$

where the function $\mathcal{P}$ is defined:

$$\mathcal{P}(q) = \frac{1}{q} \tanh(q). \tag{6.4}$$

and has the form shown in the solid curve of Fig 11. For comparison, the dashed curve there shows the effective weight function for the case of the probability model whose third parameter $c = 1/25$, when scaled to make the graphs possess approximately the same width of their peaks. The impressive feature to observe here is that the enormous change in this third parameter (from 1 to 1/25) makes hardly any discernible difference to the shape of the weight curve near the origin, but makes a large and important change to its shape further out in the tails. In the former case, $c = 1$ will ensure continuing convexity in the cost function, whereas the smaller parameter, $c = 1/25$, will jeopardize this property. What we *do not* see in these graphs is the more squared-shouldered shapes typically found when the error model consists of the mixture of a Gaussian with a (very small amount of) broad uniform distribution (see, for examples Lorenc and Hammon 1988, Andersson and Järvinen 1999).

Although we do not show it, the effect of skewness in $f_{\mathsf{x}}(x)$ is to skew the effective weight function in the opposite direction. This is to be expected since, in the directions of positive skew, the large errors with this sign of ob-error are relatively more probable than equal errors of opposite sign, so an indication, from the $O - A$ statistic, $\equiv x$, that the error is of the sign tending to typify the larger error mangnitudes means that the correct response is then to distrust this measurement more than would be the case if the x had the opposite sign.

While this basic system of (a, b)-parameterized distributions is probably adequate to serve as the model for most species of scalar observational error encountered in practice, we have nevertheless extended the model in what we believe is a natural and rational way, so that the wings of the probability distribution are afforded some measure of shape control. For this, we introduce a third nontrivial shape parameter, the asymptote convexity, c, (which was implicitly $c = 1$ in the basic model) and prove that, for $c < 1$, at least, consistent smooth generalizations of our distributions continue to exist with asymmetry $a \in (-1, 1)$ and tail broadness, $b > 0$, occupying the same parameter range as hitherto. With this more general model, it should be possible to achieve a more exacting fit to actually inferred observation error statistics. However, in this case we are in danger of jeopardizing the valuable property of solution-uniqueness in the cost-function minimization unless the adoption of our more general error distributions is accompanied by an attempt to include the effects of a nontrivial loss model. This is a topic that will be treated in the companion article (Purser et al., 2012).

We have taken pains to ensure that each distribution of our proposed family is expressible as a continuous positive Gaussian mixture, defined by a path, or trajectory for each Gaussian's center (heat source) and a weight profile (effective heating rate, in the heat-diffusion analogy). This is advantageous for any tasks that require synthetic sampling from this distribution. A random number, t, generated in the first stage according to the weight distribution, $F_{\mathsf{T}}(t)$, can be used to condition a second sampling of a synthetic Gaussian random number scaled to have the standard deviation, $\sqrt{2t}$ and centered at $\hat{x}(t)$. The distribution of all the second random numbers, taken collectively, has the probability density intended. In pratice, the first stage of the process (the preselection of each particular t) can be relatively crude in detail, provided the first few moments of the sampling match those of $F_{\mathsf{T}}(t)$, since the randomness of the variables generated in the second stage tend to obscure minor defects in those generated in the first stage. An application of such synthetic random numbers might be the simulation of $O - A$ and $O - B$ statistics, so that the actual observational errors of a given type can be better mapped to the parameters, ω, a, b, c (and perhaps a bias parameter), of the proposed family.

Finally, it is convenient to give a name to the proposed new system of distributions. Although the original motivation was to find a simple and natural broad-tailed generalization of the Gaussian with certain desirable properties (being a positive Gaussian mixture, having infinite support, defined moments, significant limiting cases) it is clear from the development of this system that the *logistic* distribution plays as important a role as the Gaussian itself at all stages of the construction, both of the basic (a, b) family and its further (a, b, c) generalization. Moreover since the archetype of the new family with the simplest choice of parameters $[(a, b, c) = (0, 1, 1)]$ *is* the classical logistic density, it seems appropriate to refer to the family as a whole that generalizes it as the 'Super-Logistic' system of distributions.

Acknowledgment

The author is grateful to Dr. Vladimir Krasnopolsky for valuable discussions on integral transforms and for introducing to him the important Erdélyi and Bateman volumes. The author thanks Dr. Krasnopolsky and Dr. Robert Grumbine for their helpful reviews.

Appendix A

A verification of equation (3.16) using the heat diffusion equation

It is easily verified that,

$$f_{\mathsf{X}}(x,t) = \exp(-|x| - t), \tag{A.1}$$

satisfies the standard heat equation with a point source at the origin:

$$-\frac{\partial f_{\mathsf{X}}}{\partial t} - \frac{\partial^2 f_{\mathsf{X}}}{\partial x^2} = 2\exp(-t)\delta(x), \tag{A.2}$$

and therefore that $f_{\mathsf{X}}(x,0) = \exp(-|x|)$ is expressible as the heat kernel (Green's function) integral:

$$\exp(-|x|) = \int_0^\infty \frac{1}{\sqrt{4\pi t}} 2\exp(-t)\exp\left(\frac{-x^2}{4t}\right) dt. \tag{A.3}$$

By the changes variables, $x^2 = z$ and $s = 1/(4t)$, we obtain:

$$f_{\mathsf{Z}}(z) = \exp(-z^{1/2}) = \int_0^\infty F_{\mathsf{S}}(s) e^{-sz}\, ds, \tag{A.4}$$

with,

$$F_{\mathsf{S}}(s) = \frac{1}{2\sqrt{\pi} s^{3/2}} \exp\left(\frac{-1}{4s}\right). \tag{A.5}$$

Appendix B

Moments and cumulants for the distribution, $\exp(abx)\mathrm{sech}^{2b}(x/2)$

The parameterized unnormalized density function,

$$f_{\mathsf{X}}(a,b;x) = \exp(abx)\,\mathrm{sech}^{2b}(x/2), \tag{B.1}$$

with a shape determined by the two parameters, a and b, is the archetypal form of the heavy-tailed probability models we propose to model the errors of scalar meteorological observations. In order to obtain exact moments and cumulants, and hence to normalize the distribution and get the bias, variance, skewness and kurtosis, we need to be able to solve the integral for its moment-generating function, $f_{\mathsf{K}}(a,b;k)$.

First, we consider the change of variable, $w = e^{x/2}$, in the integral transform:

$$\begin{aligned} f_{\mathsf{K}}(a,b,k) &= \int_{-\infty}^{\infty} \exp([ab - ik]x)\mathrm{sech}^{2b}(x/2)\, dx \\ &= 2^{2b+1}\int_{-\infty}^{\infty} \frac{\exp((2[ab-ik]-1)x/2)}{(e^{x/2}+e^{-x/2})^{2b}} \frac{e^{x/2}}{2}\, dx \\ &= 2^{2b+1}\int_0^\infty \frac{w^{2[ab-ik]-1}}{(w+1/w)^{2b}}\, dw. \end{aligned} \tag{B.2}$$

By a second change of variable: $v = w^2/(w^2+1) \equiv 2e^{x/2}\mathrm{sech}(x/2)$ we find that

$$dv = \frac{2w\,dw}{(w^2+1)^2}, \tag{B.3}$$

and hence the result can be evaluated as a (complex) beta function:

$$\begin{aligned} f_{\mathsf{K}}(a,b;k) &= 2^{2b}\int_0^1 v^{b+ab-ik-1}(1-v)^{b-ab+ik-1}\,dv \\ &= 2^{2b}B\left(b+ab-ik,\, b-ab+ik\right) \\ &\equiv 2^{2b}\frac{\Gamma(b+ab-ik)\Gamma(b-ab+ik)}{\Gamma(2b)}. \end{aligned} \tag{B.4}$$

As usual, the moments χ_q can be found by repeatedly differentiating f_{K} of (B.4) with respect to k at the origin or, equivalently, by the appropriate combinations of the cumulants, $\hat{\chi}_q$, obtained by evaluating successive k-derivatives at the origin of the cumulant-generating function, $\ln(f_{\mathsf{K}}(a,b;k))$, which proves, in this case, to be much simpler. Since the cumulant-generating function is,

$$\ln[f_{\mathsf{K}}(a,b;k)] = \ln\left(\frac{2^{2b}}{\Gamma(2b)}\right) + \ln\left(\Gamma(b+ab-ik)\right) + \ln\left(\Gamma(b-ab+ik)\right), \tag{B.5}$$

and the qth cumulant is defined from it by:

$$\hat{\chi}_q = \frac{i^q d^q}{dk^q}\ln[f_{\mathsf{K}}(a,b;k)]\bigg|_{k=0}, \tag{B.6}$$

we find that,

$$\hat{\chi}_q = \psi_+^{(q-1)} + (-)^q\psi_-^{(q-1)}, \quad q>0, \tag{B.7}$$

where

$$\psi_\pm^{(q-1)} = \psi^{(q-1)}([1\pm a]b) \tag{B.8}$$

denote the evaluations of the polygamma function of index $q-1$ defined by (e.g., Abramowitz and Stegun, 1972, p. 260):

$$\psi^{(q-1)}(z) = \frac{d^q}{dz^q}\ln(z). \tag{B.9}$$

The following procedure has been found to provide accurate and reliable numerical evaluations of the polygamma function for small to moderate arguments. Since the pair of arguments $(1\pm a)b$ are real, an iterated application of the standard recurrence:

$$\psi^{(n)}(z+1) = \psi^{(n)}(z) + (-)^n n! z^{-(n+1)}, \tag{B.10}$$

will alway enable the desired result to be expressed in terms of the same polygamma function evaluated for an argument $z \in [\frac{1}{2}, \frac{3}{2}]$ But then, for any $|z|<1$ there exists a convergent series expansion in terms of Euler's constant, $\gamma \approx 0.5772\ldots$ (for the digamma function $\psi \equiv \psi^{(0)}$ only) and Riemann's zeta function:

$$\psi^{(n)}(1+z) = \begin{cases} -\gamma - \sum_{m=1}^{\infty}(-)^m \zeta(m+1) z^m & : \quad n=0 \\ (-)^{n+1} \sum_{m=0}^{\infty}(-)^m \frac{(n+m)!}{m!} \zeta(n+m+1) z^m & : \quad n>0 \end{cases}, \tag{B.11}$$

where the summation defining the zeta function,

$$\zeta(w) = \sum_{n=1}^{\infty} n^{-w}, \tag{B.12}$$

can be accelerated by replacing all but the first few (e.g., 10) terms by the approximating integral corrected by the asymptotic Euler-Maclaurin formula. For large arguments of the polygamma functions, an asymptotic expansion proves to be more accurate. Abramowitz and Stegun (1972) provide all the necessary formulae.

In the important case of the symmetric distributions, where $a = 0$ and the odd moments vanish, the cumulants simplify:

$$\hat{\chi}_q = \begin{cases} 2\psi^{(q-1)}(b) & : \quad q \text{ even} \\ 0 & : \quad q \text{ odd} \end{cases}. \tag{B.13}$$

References

Abramowitz, M., and I. A. Stegun 1972 *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.*, Dover, New York. 1046 pp.

Andersson, E., and H. Järvinen 1999 Variational quality control. *Quart. J. Roy. Meteor. Soc.,* **125,** 697–722.

Blinnikov, S., and R. Moessner 1998 Expansions for nearly Gaussian distributions. *Astron. Astrophys. Suppl. Ser.,* **130,** 193–205.

Cramér, H. 1926 On some classes of series used in mathematical statistics. *Skandinaviske Matematikercongres,* Copenhagen.

Cramér, H. 1928 On the composition of elementary errors. *Skand. Aktuartidskr.,* **11,** 13–74, 141–180.

Cramér, H. 1946 *Mathematical Methods of Statistics* Almqvist and Wiksells, Uppsala (reprinted by Princeton University Press, 1951).

Edgeworth, F. Y. 1904 The law of error.*Trans. Camb. Philos. Soc.,* **20,** 36–66, 113–141.

Erdélyi, A., and H. Bateman 1954 *Tables of Integral Transforms, Vol. I and II*, McGraw-Hill, New York.

Johnson, N. I., and S. Kotz 1970 *Continuous Univariate Distributions–2.* Houghton Mifflin Co., Boston.

Kendall, M., and A. Stuart 1977 *The Advanced Theory of Statistics, Vol. 1, 4th Edition* Macmillan, New York.

Koornwinder, T. H., R. S. C. Wong, R. Koekoek, R. F. Swarttouw 2010 *Orthogonal Polynomials.* In *NIST Handbook of Mathematical Functions* (Olver, F. W. J., and others, eds.) Cambridge University Press.

Kotz, S., T. J. Kozubowski, and K. Podgórski 2001 *The Laplace Distribution and Generalizations.* Birkhäuser, Boston.

Lorenc, A. C. 1986 Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.,* **112,** 1177–1194.

Lorenc, A. C., and O. Hammon 1988 Objective quality control of observations using Bayesian methods. Theory and a practical implementation. *Quart. J. Roy. Meteor. Soc.,* **114,** 515–543.

Oldham, K. V., and J. Spanier 1974 *The Fractional Calculus; Theory and Applications of Differentiation and Integration to Arbitrary Order* Academic Press.

Purser, R. J. 1984 A new approach to the optimal assimilation of meteorological data by iterative Bayesian analysis. *Preprint, AMS 10th Conference on Weather Forecasting and Analysis,* Clearwater Beach, Florida, pp 102–105.

Purser, R. J., X. Su, and M. de Pondeca 2012 Effect of the choice of loss function in nonlinear quality control, NOAA/NCEP Office Note, (manuscript in preparation).

Schwartz, L. 1966 *Mathematics for the Physical Sciences.* Addison-Wesley, Reading, MA (reprinted by Dover, New York, 2008).

Widder, D. V., 1941 *The Laplace Transform.* Princeton University Press (Reprinted by Dover, 2010.)