Applied Stochastic Models in Business and Industry

WILEY

RESEARCH ARTICLE OPEN ACCESS

Negative Probability

Nick Polson[1] | Vadim Sokolov[2]

[1]Booth School of Business, University of Chicago, Chicago, IL, USA | [2]Department of Systems Engineering and Operations Research, George Mason University, Fairfax, VA, USA

Correspondence: Vadim Sokolov (vsokolov@gmu.edu)

Received: 5 May 2024 | **Revised:** 28 October 2024 | **Accepted:** 4 November 2024

Keywords: Bayes rule | dual densities | Feynman | half coin | Heisenberg principle of uncertainty | negative probability | quantum computing | Wigner

ABSTRACT

Negative probabilities arise primarily in physics, statistical quantum mechanics, and quantum computing. Negative probabilities arise as mixing distributions of unobserved latent variables in Bayesian modeling. Our goal is to provide a link between these two viewpoints. Bartlett provides a definition of negative probabilities based on extraordinary random variables and properties of their characteristic function. A version of the Bayes rule is given with negative mixing weights. The classic half-coin distribution and Polya-Gamma mixing are discussed. Heisenberg's principle of uncertainty and the duality of scale mixtures of Normals is also discussed. A number of examples of dual densities with negative mixing measures are provided including the Linnik and Wigner distributions. Finally, we conclude with directions for future research.

1 | Introduction

Our paper was motivated by numerous conversations with Nozer Singpurwalla in 2022. Nozer had a keen interest in quantum probability and the foundations of statistical inference. For example [1], writes about Feynman's view that negative probabilities and subjective Bayes could explain how quantum systems violate Bell inequalities, and [2] solves a problem in particle physics. Nozer had a great sense of interesting problems that spanned many scientific fields and was fearless in his pursuit of such ideas. He had a lifelong interest in the foundations of statistics, see [3]. One of his favorite sayings about research was *one fine day we'll expect results!*

Many authors including [4–6, 7], use negative probabilities as a tool for explaining physical phenomena in quantum mechanics. As Dirac noted, "*negative energies and probabilities should not be considered as nonsense. They are well-defined concepts mathematically, like a negative of money.*" Quantum Bayesian Computation [8] uses negative probabilities which can help explain the collapse of the wave function, entanglement, and non-locality.

Eddington [9] considers the problem of a very large number of gas particles N with the same probability p of coordinates. Bernoulli's central limit theorem can be applied to find the "fluctuation" distribution of the number of particles in a given fixed volume. He shows that the whole fluctuation can be separated into two independent terms, one depending on the fluctuation of pN and the other on the fluctuation of N, which he distinguishes as *ordinary* and *extraordinary*. The extraordinary fluctuation is to be combined negatively and removed from the ordinary one. The ordinary component assumes that the gas extends uniformly without limit in all directions. But, an infinite extent of uniform gas is contrary to relativity theory. Hence, Eddington shows that *"this space-curvature is simply a way for taking the extraordinary fluctuation into account"*.

In memory of Nozer SingpurwallaNick Polson is Professor of Econometrics and Statistics at Chicago Booth: ngp@chicagobooth.edu. Vadim Sokolov is an Associate Professor at the Volgenau School of Engineering at George Mason University. vsokolov@gmu.edu. We would like to thank the referee for their very detailed and insightful comments. Including a version of the Bayes rule with negative probabilities.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Applied Stochastic Models in Business and Industry* published by John Wiley & Sons Ltd.

Bartlett provides a definition based on characteristic functions and extraordinary random variables. As [10] observes, negative probabilities must always be combined with positive probabilities to yield a valid probability distribution before any physical interpretation is admissible. To illustrate such random variables, we show that the classic negative probability half coin distribution is related to the Pólya-Gamma mixing distribution [11, 12]. The Linnik distribution [13] can be expressed as a Gaussian scale mixture but with negative mixing weights [14–17].

Mixtures of Exponential [18, 19] and Gaussian distributions have a long history in MCMC algorithms and hierarchical representations of distributions [20] and lead to EM algorithms for posterior mode and maximum likelihood inference. Our results build on this literature by extending the class of distributions to those with negative mixing weights.

From another perspective [21], provides a simple proof of the famous Bell's inequality with two applications of Hoeffding's inequality where Bell's theorem is related to statistical causality, see also probability bounds in [22].

The rest of the paper is organized as follows. Section 1.1 discusses two classic examples of [7]. Section 2 revisits the definition of negative probability and extraordinary random variables due to [10]. We consider an archetypal example of half-coin distribution due to [23]. Section 3 provides our results on new characterizations of scale mixture of Normals using dual densities [24, 25]. Bernstein's theorem for completely monotone functions is used to determine when the mixing weights are non-negative. Our work shows that many results in quantum mechanics are also related to the notion of dual densities and scale mixtures of normal. A number of examples, including the Linnik, the stable and the Wigner distribution are provided. Finally, Section 4 concludes with some directions for future research.

1.1 | Motivating Example

Feynman [7] provides the following simple example of negative probabilities. Feynman discusses the case with a conditional table for $p(\text{state} = j|E)$ for $j = (1, 2, 3)$ and $E = \{A, B\}$ with underlying base rates given by $p(A) = 0.7$, $p(B) = 0.3$. The conditional probability table (Table 1) has a negative entry with the usual constraint of summing to one. Specifically,

Notice that $p(\text{state} = 2|B) = 1.2 > 1$ in order to offset the negative conditional probability $p(\text{state} = 1|A) = -0.4$. The total of probabilities is still one, and we have a valid marginal probability distribution over the states.

TABLE 1 | Feynman's conditional probability table.

		Given A	Given B
State	1	0.3	−0.4
	2	0.6	1.2
	3	0.1	0.2

The observable marginal distributions form an ordinary random variable and are calculated as

$$p(\text{state} = 1) = p(\text{state} = 1|A)p(A) + p(\text{state} = 1|B)p(B)$$
$$= 0.7 \times 0.30.3 \times 0.4 = 0.09$$

Although $p(\text{state} = 2|B) = 1.2$, is allowed to be greater than one,

$$p(\text{state} = 2) = 0.7 \times 0.6 + 0.3 \times 1.2 = 0.78$$

which is still a valid probability. Similarly, we have

$$p(\text{state} = 3) = 0.7 \times 0.1 + 0.3.2 = 0.13$$

We can see that we have ordinary probabilities for the states. The key point is that the law of total probability still holds even though the mixing weights, which are unobserved (latent) are allowed to contain negative values.

Wigner [6] shows that in quantum theory the joint density function $P(x, p)$ of the location and momentum of a particle cannot be non-negative everywhere as it is always real and yet its integral over the whole space is zero. Hence, written as a convolution (a.k.a. Bayesian mixture model) in which mixing weights can be negative. Feynman provides the following concrete example: consider a particle diffusing in 1-dimension in a rod has probability $P(x, t)$ of being at x at time t and satisfies

$$\frac{\partial}{\partial t} P(x, t) = -\frac{\partial^2}{\partial x^2} P(x, t)$$

Suppose that at $x = 0$ and $x = \pi$ the rod has absorbers so that $P(x, t) = 0$ and let $P(x, 0) = f(x)$. What is $P(x, t)$ thereafter?

The solution is given by $P(x, t) = \sum_{n=-=1}^{\infty} p_n \sin(nx) e^{-n^2 t}$ where

$$f(x) = \sum_{n=1}^{\infty} p_n \sin(nx) \quad \text{and} \quad p_n = \frac{2}{\pi} \int f(x) \sin(nx) dx$$

This is a mixture with negative weights, and thus, is an extraordinary random variable. See also [26]

2 | Extraordinary Random Variables

Extraordinary random variables can generate ordinary random variables in two ways: as convolutions of probability measure, and through mixtures with negative weights. We consider each of these in turn.

2.1 | Convolutions

To fix notation. Let • denote an extraordinary random variable or generating function. Hence, the unobserved component $Z^\bullet$ will have density $f^\bullet(z)$ and generating function $G^\bullet_Z(s)$. Let $\phi_X(t) = E(e^{itX})$ denote the characteristic function of an ordinary random variable X. The generating function and Fourier transform (a.k.a. characteristic function) are related by

$$G_Z(s) = E(s^Z) \quad \text{and} \quad \phi_Z(t) := E(e^{itZ}) = G_Z(e^{it})$$

15264025, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/asmb.2910 by Readcube (Labtiva Inc.), Wiley Online Library on [01/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Negative probabilities arise as convolutions of probability measure. Imagine a random variable represented as a convolution

$$Y = X + Z^{\bullet}$$

where $Z^{\bullet}$ has an extraordinary probability distribution where X and $Z^{\bullet}$ are independent in the usual statistical sense. We can think of Y as observed, $Z^{\bullet}$ as a hidden and X the state of nature.

The generating function of the convolution $Y = X + Z^{\bullet}$ is a product, by independence, with

$$f_Y(y) = \int f_X(y-z) f_Z^{\bullet}(z) dz$$
$$G_Y(s) = G_X(s) G_{Z^{\bullet}}(s)$$

The existence of an ordinary random variable in the convolution case follows from the fundamental theorem [23] which we now describe in our notation with upper case symbols for the generating functions and • symbols marking any extraordinary variable.

Fundamental theorem. For every generating function $F^{\bullet}(s)$ of an extraordinary probability distribution there exist two probability generating functions G and H of ordinary non-negative distributions such that

$$F_Z^{\bullet}(s) G_X(s) = H_Y(s)$$

The sum of independent random variables leads to a product of their generating functions. Let L_1^+ denote the space of integrable densities. Then, $f \in L_1, g \in L_1^+, \exists f$ such that $f * g \in L_1^+$ where the convolution is given by

$$(f * g)(z) = \int g(z-x) f^{\bullet}(x) dx$$

Hence a law of total probability holds for convolutions with negative probabilities.

One can view the law of total probability as a convolution theorem for random variables. A natural generalization of Feynman's examples is convolutions with negative weights. The classic example is the half coin is connected with Pólya-Gamma mixing [20, 27]. See also [28, 29].

Half Coin. Let Y be a single toss of a coin with Bernoulli distribution, $Y \sim Ber(p)$. Then a full toss can be decomposed as a sum of two "half" coins [23].

$$G_Y(s) = G_X^{\bullet}(s) G_Z^{\bullet}(s)$$

Half-coins are extraordinary r.vs, so this is not an example of the fundamental theorem.

Specifically, the probability generating function is defined by the formula $f(z) = \sum_{n=1}^{\infty} p_n z^n$. The pdf of a fair coin is

$$f(z) = \frac{1}{2} + \frac{1}{2} z$$

If we assume that $\sum_{n=1}^{\infty} p_n = 1$ and $\sum_n |p_n| < \infty$ but drop the requirement fo non-negativity of its probabilities, we can define the half coin as having a pdf

$$f_{\frac{1}{2}}(z) = \sqrt{\frac{1}{2} + \frac{1}{2} z} = \frac{1}{\sqrt{2}} \left(1 + \frac{1}{2} s - \frac{1}{8} s^2 \ldots \right)$$

According to the Binomial theorem

$$\sqrt{\frac{1}{2} + \frac{1}{2} z} = \frac{1}{\sqrt{2}} \sum_{n=0}^{\infty} \binom{1/2}{n} z^n$$

where the coefficients are, with C_n the n-th Catalan number,

$$\binom{1/2}{n} = (-1)^{n-1} \frac{2C_{n-1}}{4^n} \quad \text{and} \quad C_n = \frac{1}{n+1} \binom{2n}{n}$$

Pólya-Gamma. The probability-generating function of the half-coin is related to that of the Pólya-Gamma distribution. Let $X \sim PG(b, 0)$. By definition, the moment generating function is

$$E\{e^{-2tX}\} = \frac{1}{\cosh^b(\sqrt{t})} = \frac{1}{(e^{\sqrt{t}} + e^{-\sqrt{t}})^b}$$

Letting $s = e^{-st}$, yields p.g.f.

$$E(s^X) = \left(\sqrt{s} + \frac{1}{\sqrt{s}} \right)^{-b}$$

Barndorff-Nielsen et al. [11] (Equation 3.6) gives the mixing density

$$f_X(u) = \sum_{k=0}^{\infty} \binom{-2\delta}{k} \frac{(\delta + k)}{B(\delta, \delta)} e^{-\frac{1}{2}(\delta+k)^2 u} \quad \delta > 0$$

Hence, we see the equivalence with the half-coin, where $b = -1/2$ and $\delta = 2$!

The negative factorial function can be written in the more usual way as

$$\binom{-2\delta}{k} = (-1)^k \binom{2\delta + k - 1}{k}$$

Bartlett's definition. Bartlett [10] provides a formal extension of Kolmogorov's mathematical probability as follows. He introduces extraordinary random variables through their characteristic functions. As Bartlett observes, negative probabilities must always be combined with positive ones to give an ordinary probability distribution before a physical interpretation is admissible. The following definitions of extraordinary random variables will be used throughout.

They are defined via their characteristic functions. In terms of Fourier transform (a.k.a. characteristic functions) we have

$$\phi_Y(t) = \phi_X(t) \phi_Z^{\bullet}(t)$$

Given ϕ_Y, ϕ_X, we would like to identify the mixing measure of the hidden variable Z. Solving for $\phi_Z(t)$ we have,

$$\phi_Z^{\bullet}(t) = \frac{\phi_Y(t)}{\phi_X(t)} = \phi_Y(t) \phi_X^{-1}(t)$$

1526402ￄ, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/asmb.2910 by Readcube (Labtiva Inc.), Wiley Online Library on [01/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

This has the same form as the convolution product above!

Notice that we write the characteristic function of Z in terms of a random variable, denoted by W. as follows

$$\phi^\bullet_Z(t) = \phi_Y(t)\phi^\bullet_W(t), \text{ with } \phi^\bullet_W(t) = \phi^{-1}_X(t)$$

However, the following identity holds

$$\phi_X(t)\phi^\bullet_W(t) = \phi_X(t)\phi^{-1}_X(t) = 1 = E(e^{it0})$$

$$X + W^\bullet \stackrel{P}{=} 0$$

Therefore, $W^\bullet$ will have an extraordinary probability distribution. That is it will take negative values in parts of its domain.

van Dantzig pair. If the functions $\phi_X(t)$ and $1/\phi_X(it)$ are both characteristic functions then we have a van Dantzig pair of random variables. This is similar to Bartlett's definition except we evaluate the reciprocal at $1/it$ rather than $1/t$. For applications, see [30] and [31].

Example. Let Y be the sum of tosses leading to a Binomial distribution $Bin(n, p)$. For $p < 0$ this is an extraordinary random variable. The reciprocal distribution takes the form of a negative Binomial distribution with a generating function given by $(p + qs)^{-n}$.

2.2 | Mixtures

In the case of mixtures, we have

$$f_Y(y) = \int f_{Y|Z}(y|z) f^\bullet_Z(z) dz$$

$$G_Y(s) = \int G_{Y|Z}(s|z) f^\bullet_Z(z) dz$$

An interesting class is scale mixtures of Gaussian with negative mixing weights. So far this class has received little attention in the literature relative to their ordinary mixture counterparts [27]. By construction, we have

$$Y = \sqrt{Z^\bullet} X \quad \text{where} \quad X \sim N(0, 1)$$

which leads to the class of densities of the form

$$f_Y(y) = \int_0^\infty \frac{1}{\sqrt{2\pi z}} e^{-\frac{1}{2}\frac{y^2}{z}} f^\bullet_Z(z) dz$$

with mixed weight mixing measure $f^\bullet_Z$.

2.3 | Bayes Rule for Extraordinary Random Variables

One would like to infer the distribution of the hidden variable $Z^\bullet$ given the observable Y, that is provide a Bayes rule for extraordinary random variables. We now do this.

We argue as follows. We have a likelihood $f(y|z, s)$ with two parameters $(Z, S^\bullet)$ with $S^\bullet$ being extraordinary. The joint prior to these parameters is

$$p(z, s) = p(z|s) g^\bullet(s)$$

We can then write the Bayes rule in two ways, with joint dependence on Z and $S^\bullet$, the other for Z conditioned on $S^\bullet$. The joint version is

$$f^\star(z, s|y) = \frac{f(y|z, s)p(z|s)g^\bullet(s)}{m(y)}$$

where

$$m(y) = \int\int f(y|z, s)p(z|s)g^\bullet(s) dz ds = \int m(y|s) g^\bullet(s) ds$$

and

$$m(y|s) = \int f(y|z, s)p(z|s) dz$$

The conditional version is

$$f^\star(z|y, s) = \frac{f(y|z, s)p(z|s)}{m(y|s)}$$

We can then rewrite the joint version as

$$\begin{aligned} f(z, s|y) &= \frac{f(y|z, s)p(z|s)g^\bullet(s)}{m(y)} \\ &= \frac{f(y|z, s)p(z|s)g^\bullet(s)}{m(y)} \frac{m(y|s)}{m(y|s)} \\ &= f(z|y, s) g^{\bullet\star}(s|y) \end{aligned}$$

where

$$g^{\bullet\star}(s|y) = \frac{m(y|s)}{m(y)} g^\bullet(s) = h(s|y) g^\bullet(s)$$

Thus updating the prior by adding conditioning on Y begin

$$\begin{aligned} p(z, s) &= p(z|s) g^\bullet(s) \\ f(z, s|y) &= f(z|y, s) g^{\bullet\star}(s|y) \end{aligned}$$

The specific case is $p(z|s) = se^{-sz}$, but the argument applies for any (ordinary) $p(z|s)$.

The dual nature of this class of densities is discussed in the next section where a Heisenberg principle of uncertainty for normal scale mixtures is given, see [24].

3 | Duality of Densities

The concept of dual densities was introduced in 1995 by Jack Good ([25]) as densities proportional to the moment generating functions (or characteristic functions when they exist) of a given density. This was further explored by Tilmann Gneiting, who established connections between the mixing distributions for dual densities that are also normal variance mixtures ([24]) and Nadarajah, who provides a list of dual densities for common continuous distributions [32, 33]. An important result from Gneiting's paper is as follows:

15264025, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/asmb.2910 by Readcube (Labtiva Inc.), Wiley Online Library on [01/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Consider the normal variance mixture $p(x) = \int (2\pi v)^{-1/2} e^{-x^2/2v} dF(v)$ and its characteristic function ϕ, see [34–36]. Then the dual density $\hat{p}$ is also a normal scale mixture with mixing density $\hat{f}$. The mixing density for the dual and the original (or, the primal) density are related by a simple formula:

$$\hat{f}(v) = \frac{1}{(2\pi)^{1/2} p(0) v^{3/2}} f\left(\frac{1}{v}\right), \ v > 0 \tag{1}$$

This gives a useful tool for constructing polynomially tailed (or slowly varying) priors that are also normal variance mixtures starting from a prior that is exponentially tailed (or rapidly varying). A classic example is the double-exponential or Laplace distribution whose dual is the Cauchy distribution. Since the Laplace can be written as a normal scale mixture with exponential mixing density, we can derive the mixing density for Cauchy. Some examples of exponentially-tailed densities with polynomially-tailed duals compiled from [32] and [24] are reported in Table 2:

Good and Gneiting [24, 25] introduced the concept of a dual density. Given a density $p(x)$, its dual density is given by its characteristic function

$$\phi_X(t) = \mathbb{E}(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} p(x) dx$$

appropriately normalized. The characteristic function is simply the Fourier transform (with sign reversal) of the probability density function. Some functions are invariant under this transform. For example, the characteristic function of normal is again normal. We call the two densities dual if each is proportional to the characteristics function of the other. The normal is it is dual.

The class of scale mixture of Normals with bounded density with mixing measure, F, defined on $(0, \infty)$ given y

$$p_X(x) = \int_0^{\infty} \frac{1}{\sqrt{2\pi v}} e^{-x^2/2v} dF(v)$$

has characteristic function

$$\phi_X(t) = \int_0^{\infty} e^{-vt^2/2} dF(v)$$

This follows from the characteristic function of a standard normal, namely

$$E(e^{itZ}) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi v}} e^{-x^2/2v} dx = e^{-vt^2/2}$$

When both p and ϕ are bounded and integrable, we have

$$p_X(0) = \int_0^{\infty} \frac{1}{\sqrt{2\pi v}} dF(v) < \infty$$

The dual density $\hat{p}$ is proportional to ϕ and is given by $\hat{p}(t) = \hat{p}(0)\phi(t)$,

Surprisingly, the dual density is also a scale mixture of normal, with density

$$\hat{p}_X(x) = \int_0^{\infty} \frac{1}{\sqrt{2\pi v}} e^{-x^2/2v} d\hat{F}(v)$$

where

$$\frac{1}{\sqrt{2\pi v}} d\hat{F}(v) = \frac{1}{2\pi p(0)} dF(1/v), \ v > 0$$

Then its dual characteristic function is given by

$$\begin{aligned}\phi_X(t) &= \int_{-\infty}^{\infty} \exp(itx) \int_0^{\infty} (2\pi v)^{-\frac{1}{2}} \exp(-x^2/2v) dF(v) \\ &= \int_0^{\infty} \exp\left(-\frac{1}{2} u\omega^2\right) f(u) du\end{aligned}$$

Therefore,

$$\hat{p}_X(t) = \int_0^{\infty} (2\pi v)^{-\frac{1}{2}} \exp(-x^2/2v) dG(v) \text{ where } g(v) = u^{-\frac{3}{2}} f(u^{-1})$$

TABLE 2 | Some common exponentially tailed densities that have polynomially tailed dual densities. The densities marked with asterisks also have a commonly known normal variance mixture representation.

Density$p(x)$	**Dual density $\hat{p}(x)$**	**Comments**
Exponential Power* (Special Case: Laplace, Normal)	Symmetric-stable(α) (Special case: Cauchy, Levy)	Symmetric stable distributions are heavy-tailed for $\alpha < 2$
Bessel function density*	Student's t	These densities are special cases of Generalized Hyperbolic distribution with parameters $(\lambda = -\frac{\nu+1}{2}, \delta^2 = 0, \kappa^2 = \nu)$ and $(\lambda = -\nu/2, \delta^2 = \nu, \kappa = 0)$ respectively
Gamma (shape=α, rate=λ)	$\hat{p}(x) = x^{-\alpha}$	
Laplace	$\hat{p}(x) \propto e^{ax}(1 + x^2)^{-1}$	Heavy-tailed if $a = 0$
Skew-Laplace I	$\hat{p}(x) \propto e^{ax}(c/(b + x) + 1/x)$	Heavy-tailed if $a = 0$
Skew-Laplace II	$\hat{p}(x) \propto \{\frac{cx}{x^2+a^2} + \frac{1}{b+x}\}$	Heavy-tailed
Fretchét	$\hat{p}(x) \propto \sqrt{x} K_{-1}(a\sqrt{x})$	Fretchét distribution has a lower exponential tail
Inverse Gaussian	$\hat{p}(x) \propto x^{1/4} K_{-\frac{1}{2}}(a\sqrt{x})$	
Linnik or α-Laplace distribution*	Generalized Cauchy	
$\alpha \in [0, 2)$	$\hat{p}(x) \propto (1 + \|x\|^{\alpha})^{-\beta}$	

15264025, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/asmb.2910 by Readcube (Labtiva Inc.), Wiley Online Library on [01/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Hence, the mixing measure can be obtained via inversion of a Laplace transform.

$$p_X(x) = \int_0^\infty \frac{1}{(2\pi v)^{1/2}} e^{-x^2/2v} dF(v)$$

$$\phi_X(t) = E(e^{itX}) = \int_0^\infty e^{-vt^2/2} dF(v)$$

Gneiting and Good [24, 25] shows that if p and $\hat{p}$ are normal scale mixtures,

$$\sigma_p \sigma_{\hat{p}} \geq 1 \Leftrightarrow p, \hat{p} \text{ are normal}$$

This follows from flopping between Fourier and Laplace transforms.

Therefore, a pair of dual densities $(p, \hat{p})$ follow a Heisenberg principle, when one learns something about p one has information about the other, but they both cannot be observed at the same time.

Another example of a Heisenberg principle of uncertainty is given by the Wigner distribution.

Wigner Distribution. [5] uncertainty principle asserts a limit to the precision with which position x and momentum p of a particle can be known simultaneously, namely, the standard deviations satisfy

$$\sigma_x \sigma_p \geq \frac{1}{2} h$$

where h is Planck's constant [6], exhibited a joint distribution function $f_\psi(x, p)$ for position and momentum however some of its values have to be negative and he asserts that "this cannot be interpreted as the simultaneous probability for coordinates and momentum" but can be used in calculations as an auxiliary mixture measure. For a unit vector, ψ, the Wigner distribution is defined as

$$f_\psi(x, p) = \frac{1}{2\pi} \int \psi\left(x + \frac{s}{2} h\right) \psi^\star\left(x - \frac{s}{2} h\right) e^{isp} ds$$

For a recent discussion on the Wigner distribution see [37], Wigner's quasi-probability distribution, which can be used to make predictions about quantum systems. Hudson [38] shows that for the Wigner quasi-probability density to be a true density is that the vssponding Schrödinger state function is the exponential of a quadratic polynomial (a 2-dim multivariate normal).

Mixture of Exponentials. A function $f(x)$ is completely monotone if and only if it can be represented as a Laplace transform of some distribution function $F(s)$ as

$$f(x) = \int_0^\infty e^{-sx} dF(s)$$

The function $p(\sqrt{x})$ is completely monotone if

$$(-1)^k \frac{d^k}{dx} p(\sqrt{x}) \geq 0 \ \forall k = 1, 2, 3, \ldots$$

Bernstein's theorem states that $p(x)$ is completely monotonic if and only if there is a unique measure G on $[0, \infty)$ such that $p(x) = \int_0^\infty e^{-x\lambda} dF(\lambda)$. Bernstein functions which include the class of scale mixtures of Normals. We have the representations

$$p(x) = \exp(-\phi(x)) = \sqrt{\frac{2}{\pi}} \int_0^\infty \sqrt{\lambda} \times \exp\left(-\lambda x^2\right) dF(\lambda) \ \Leftrightarrow \ p(\sqrt{x}) \text{ completely monotone}$$

This is essentially the Bernstein-Widder-Schoenberg theorem applied to $p(x) = \exp(-\phi(x))$.

The Cauchy-Laplace pair of distributions provides another example.

$$\frac{1}{2} e^{-|t|} = \int_{-\infty}^\infty e^{itx} \frac{1}{\pi} \frac{1}{1 + x^2} dx$$

The exponential power is a Gaussian mixture for $\alpha \in (0, 2]$ given by

$$\exp(-|t|^\alpha) = \int_0^\infty e^{-st^2/2} f(s) ds$$

where $f(s)$ can be identified as a positive α-stable r.v. with index $\alpha/2$. When $\alpha = 2$ we get the Cauchy/Laplace dual density pair described above.

Negative convolutions arise when $p(\sqrt{x})$ is not completely monotone and an example of this is given by

$$\frac{1}{\pi} \int_0^\infty e^{-tu} \sin(u^{1/2}) du = \frac{1}{2\pi^{1/2}} \frac{1}{t^{3/2}} e^{-1/4t} = \varphi_{1/2}(t)$$

where

$$e^{-|t|^{1/2}} = \int_0^\infty e^{-xt} \varphi_{1/2}(t) dt$$

Linnik Distribution. If we start with the Laplace transform identity for a Cauchy random variables

$$\frac{1}{1 + x^2} = \int_0^\infty e^{-tx} t^{-\frac{1}{2}} \sin(t) dt$$

Then under the transformation, $x \rightarrow \frac{1}{2} x^2$, this becomes a scale mixture of Normals representation for the Linnik distribution

$$\frac{1}{4 + x^4} = \int_0^\infty \sqrt{t} e^{-\frac{1}{2} tx^2} t^{-1} \sin(t/2) dt$$

If $h(\sigma) \propto \sigma^{-2} \sin(\sigma^{-2})$, we have

$$p_X(x) = \int_0^\infty \sigma^{-1} \phi(\sigma^{-1} x) h(\sigma) d\sigma$$

This result follows from the fact

$$\int_0^\infty \frac{1}{4 + x^4} = \frac{\pi}{8}$$

which can be calculated using identity

$$\frac{1}{4 + x^4} = \frac{1}{(x^2 - 2x + 2)(x^2 + 2x + 2)}$$

15264025, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/asmb.2910 by Readcube (Labtiva Inc.), Wiley Online Library on [01/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Similarly, when we have a scale mixtures of Gaussians [14, 17, 36],

$$\int_0^\infty N(0, t^{-1}C)W(t)dt = \frac{1}{1 + (x^T C^{-1} x)^2}$$

where the weights $W(t) = t^{-n/2}\sin(\frac{t}{2})$ can be negative, $x^T = (x_1, x_2, \dots, x_n)$. This provides an example with negative mixing weights [13, 14, 17].

The Linnik family for $0 < \alpha \leq 2$ is a scale mixture of Normals given by

$$\varphi(t) = E(e^{itX}) = \frac{1}{1 + |t|^\alpha}, \qquad \alpha \in (0, 2]$$

The mixing measure is given by

$$\int_0^\infty e^{-v\beta} \frac{e^{-|t|^\alpha v\beta}}{\Gamma(1 + 1/\beta)} dv = \frac{1}{(1 + |t|^\alpha)^{1/\beta}}$$

Hence, we have an ordinary mixing distribution for $\alpha \in (0, 2]$ whereas the case $\alpha = 4,\ \beta = 1$ above leads to extraordinary mixing.

4 | Discussion

Negative probabilities correspond to extraordinary random variables. They arise in many physical systems and quantum computing [8]. A related physical notion is that of dual densities which represents densities as characteristic functions rather than Laplace transforms (a.k.a. mixtures of exponential random variables). We provide many examples, including the Linnik family of distributions where certain cases lead to negative mixing weights.

Data Availability Statement

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

References

1. N. Singpurwalla , V. Volovoi , M. Brown , E. A. Pekoz , S. M. Ross , and W. Q. Meeker , *Is Reliability A New Science? In 10th International Conference on Mathematical Methods in Reliability* (France: Grenoble, 2017).

2. J. Landon , F. X. Lee , and N. D. Singpurwalla , “A Problem in Particle Physics and its Bayesian Analysis,” *Statistical Science* 26, no. 3 (2011): 352–368.

3. D. V. Lindley , *Making Decisions* (London: Wiley, 1985) ISBN 978-0-471-90803-6.

4. P. A. M. Dirac , “Bakerian Lecture - The physical interpretation of quantum mechanics,” *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences* 180, no. 980 (1942): 1–40.

5. W. Heisenberg , “Über die inkohärente Streuung von Röntgenstrahlen,” *Physikalishce Zeitschrift* 32 (1931): 740.

6. E. Wigner , “On the Quantum Correction For Thermodynamic Equilibrium,” *Physical Review* 40, no. 5 (1932): 749–759.

7. R. P. Feynman , “Negative Probability,” in *Quantum implications: essays in honour of David Bohm* (London: Routledge, 1987), 235–248.

8. N. Polson , V. Sokolov , and X. Jianeng , “Quantum Bayesian computation,” *Applied Stochastic Models in Business and Industry* 39, no. 6 (2023): 869–883.

9. S. A. Eddington , “The Combination of Relativity Theory and Quantum Theory,” *Communications of the Dublin Institute for Advanced Studies* (1943).

10. M. S. Bartlett , “Negative Probability,” *Mathematical Proceedings of the Cambridge Philosophical Society* 41 (1945): 71–73.

11. O. Barndorff-Nielsen , J. Kent , and M. Sørensen , “Normal Variance-Mean Mixtures and z Distributions,” *International Statistical Review / Revue Internationale de Statistique* 50, no. 2 (1982): 145–159.

12. N. G. Polson , J. G. Scott , and J. Windle , “Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables,” *Journal of the American Statistical Association* 108, no. 504 (2013): 1339–1349.

13. L. Devroye , “A note on Linnik's distribution,” *Statistics & Probability Letters* 9, no. 4 (1990): 305–306.

14. K.'a.-C. Chu , “Estimation and decision for linear systems with elliptical random processes,” *IEEE Transactions on Automatic Control* 18, no. 5 (1973): 499–505.

15. J. F. C. Kingman , “On Random Sequences With Spherical Symmetry,” *Biometrika* 59, no. 2 (1972): 492–494.

16. A. F. M. Smith , “On Random Sequences with Centred Spherical Symmetry,” *Journal of the Royal Statistical Society: Series B: Methodological* 43, no. 2 (1981): 208–209.

17. M. West , “On Scale Mixtures of Normal Distributions,” *Biometrika* 74, no. 3 (1987): 646–648.

18. D. J. Bartholomew , “Sufficient Conditions for a Mixture of Exponentials to Be a Probability Density Function,” *Annals of Mathematical Statistics* 40, no. 6 (1969): 2183–2188.

19. Diaconis and Ylvisaker, “Quantifying Prior Opinion,” in *Bayesian Statistics 2* (Amsterdam: North-Holland, 1985), 133–156.

20. N. G. Polson and J. G. Scott , “Mixtures, Envelopes and Hierarchical Duality,” *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 78, no. 4 (2015): 701–727.

21. R. D. Gill , “Statistics, Causality and Bell's Theorem,” *Statistical Science* 29, no. 4 (2014): 512–528.

22. J. Tian and J. Pearl , “Probabilities of Causation,” *Bounds and Identification* (2013): 287–313.

23. G. J. Székely , “Half of a coin: Negative probabilities,” *Wilmott Magazine* 50 (2005): 66–68.

24. T. Gneiting , “Normal Scale Mixtures and Dual Probability Densities,” *Journal of Statistical Computation and Simulation* 59, no. 4 (1997): 375–384.

25. I. J. Good , “Dual density functions,” *Journal of Statistical Computation and Simulation* 52, no. 2 (1995): 193–194.

26. P. Biane , “Matrix Valued Brownian Motion and a Paper by Polya,” in *Séminaire de Probabilités XLII* (Heidelberg: Springer, 2008), 171–185.

27. N. G. Polson , J. G. Scott , and J. Windle , “The Bayesian Bridge,” *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 76, no. 4 (2014): 713–733.

28. J. Navarro , “Stochastic comparisons of generalized mixtures and coherent systems,” *TEST. An Official Journal of the Spanish Society of Statistics and Operations Research* 25 (2016): 150–169.

29. G. Rabusseau and F. Denis , “Learning Negative Mixture Models by Tensor Decompositions,” 2014.

30. E. Lukacs , “A Survey of the Theory of Characteristic Functions,” *Advances in Applied Probability* 4, no. 1 (1972): 1–38.

15264025, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/asmb.2910 by Readcube (Labtiva Inc.), Wiley Online Library on [01/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

31. N. G. Polson , “Riemann Thorin van Dantzig Pairs, Wald Couples and Hadamard Factorisation,” arXiv preprint arXiv:1804.10043 (2021).

32. S. Nadarajah , “PDFs and Dual PDFs,” *American Statistician* 63, no. 1 (2009): 45–48.

33. C. S. Withers and S. Nadarajah , “Cumulants of Multinomial and Negative Multinomial Distributions,” *Statistics & Probability Letters* 87 (2014): 18–26.

34. D. F. Andrews and C. L. Mallows , “Scale Mixtures of Normal Distributions,” *Journal of the Royal Statistical Society: Series B: Methodological* 36, no. 1 (1974): 99–102.

35. B. P. Carlin and N. G. Polson , “Inference for Nonconjugate Bayesian Models Using the Gibbs Sampler,” *Canadian Journal of Statistics / La Revue Canadienne de Statistique* 19, no. 4 (1991): 399–405.

36. B. P. Carlin , N. G. Polson , and D. S. Stoffer , “A Monte Carlo approach to nonnormal and nonlinear state-space modeling,” *Journal of the American Statistical Association* 87, no. 418 (1992): 493–500.

37. Y. Gurevich and V. Vovk , “Negative probabilities,” 2020 *arXiv preprint arXiv.12393*.

38. R. L. Hudson , “When is the Wigner Quasi-Probability Density Non-Negative?,” *Reports on Mathematical Physics* 6, no. 2 (1974): 249–252.

15264025, 2025, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/asmb.2910 by Readcube (Labtiva Inc.), Wiley Online Library on [01/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License