Research Paper
Computational Methods for Differential Equations
http://cmde.tabrizu.ac.ir
Vol. *, No. *, *, pp. 1-33
DOI:10.22034/cmde.2026.73638.3773

Generating probability distributions from differential equations: A unified review of classical, modern, geometric, and generative frameworks

Mohamad Yamen Al-Mohamad, Ali Akbar Heydari*, and Hossein Jabbari Khamnei*
Department of Statistics, Faculty of Mathematics, Statistics and Computer Science, University of Tabriz, Tabriz, Iran.

Abstract

Many classical and modern probability laws are most naturally defined not by an explicit formula but as solutions of differential equations (DEs). This review organizes the DE-to-distribution literature into seven interconnected frameworks: (I) the Pearson system; (II) Fokker–Planck equations, including jump-diffusions and score-based generative models; (III) hazard-rate ODEs and frailty; (IV) fractional, delay, and integro-differential equations; (V) partial differential equations and Wasserstein flows; (VI) matrix- and tensor-variate equations; and (VII) information geometry. Two results are stated and proved in full: a verified closed-form solution of the generalized Pearson equation for all three root configurations, together with explicit support, normalizability, and tail-index conditions for Pearson Types I–VII; and a proof that Gamma frailty acting on a constant hazard yields the Lomax law exactly. New worked examples include a delay-differential equation whose normalized solution is a geometric mixture of shifted Erlang laws, and an integro-differential equation whose solution is a phase-type density. Every closed form is checked numerically against the governing equation, and all figures are reproducible from the supplementary code. The review closes with computational tools for implicitly defined densities, a decision table for choosing a framework, and open problems.

Keywords. Differential equations, Probability distributions, Fokker–Planck equation, Pearson system, Survival analysis, Fractional calculus, Information geometry, Score-based generative models.

2010 Mathematics Subject Classification. 60E05, 34A34, 35Q84, 60H10, 62N05.

Uncorrected

1. Introduction

Probability distributions are traditionally introduced via algebraic density formulas or explicit stochastic constructions. However, many of the most widely used distributions emerge naturally as solutions to differential equations (DEs). Defining a distribution via a DE implies that the shape of the distribution is governed by local dynamic forces—such as drift, diffusion, aging, memory, or geometric curvature—rather than by an arbitrary algebraic choice of functional form. Since Karl Pearson's seminal work in 1895 [52], DE-based distributions have become foundational in statistics, physics, reliability engineering, and modern machine learning. This review organizes the field into seven mathematically distinct but increasingly interconnected frameworks (Figure 1), illustrates each with both classical and newly derived distributions, and provides a roadmap for generating novel PDFs together with numerical illustrations.

Motivation. This DE-driven approach offers several distinct advantages:

- **Compact parameterization:** A single ODE can encode skewness, tail behavior, and support boundaries (Section 3).
- **Physical/stochastic interpretability:** Drift, diffusion, hazard, or memory effects map directly to equation terms (Sections 4 and 5).
- **Systematic extension:** Modifying the DE structure generates families of distributions with controlled, interpretable properties (Sections 6 and 7).

Received: 18 July 2026; Accepted: 19 November 2026.
∗ Corresponding authors. Emails: heydari@tabrizu.ac.ir, h_jabbari@tabrizu.ac.ir.

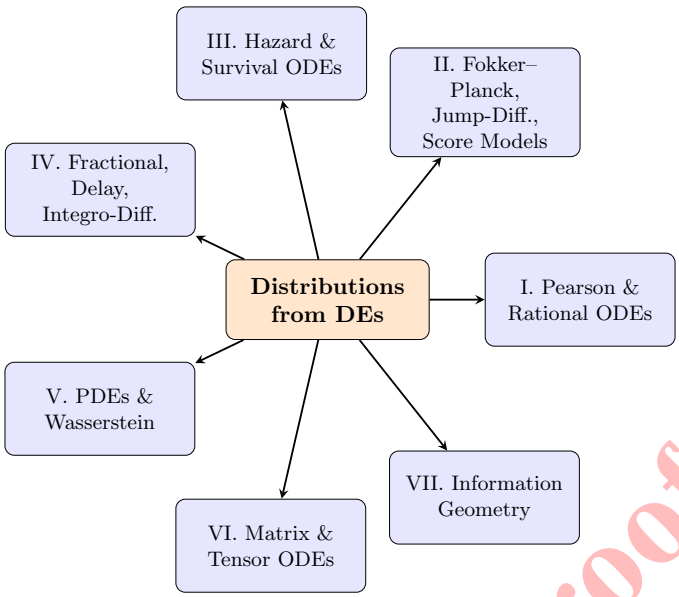


FIGURE 1. Seven interconnected frameworks for generating probability distributions from differential equations.

Uncorrected Proof

- **Forward compatibility with generative modeling:** The same steady-state and time-evolution equations that classically generate named distributions (Table 2) also underlie score-based diffusion models, the dominant paradigm in modern deep generative modeling (Section 4.6).

Contributions and Scope. Relative to earlier treatments, which typically address a single framework in isolation, this review makes the following contributions:

(1) We give a fully worked and numerically verified closed-form solution of the generalized Pearson ODE (Theorem 3.1) for all three discriminant cases (distinct real, double, and complex conjugate roots), and we derive from it a single statement of the support, normalizability, and tail-index conditions of Pearson Types I–VII (Corollary 3.2 and Table 1). In particular, every Pearson density with $b_2 \neq 0$ and unbounded support has a power-law tail of index a_1/b_2.
(2) We formally distinguish *genuine stationary densities* (solutions of $\partial_t p = 0$ under natural boundary conditions) from *fixed-time transition densities* of non-stationary processes such as geometric Brownian motion (Remark 4.2).
(3) We extend the Fokker–Planck framework to jump-diffusion processes driven by Lévy noise (Section 4.4), to quasi-stationary distributions of killed diffusions (Section 4.5), and to score-based generative models (Section 4.6), presented as the natural high-dimensional culmination of the steady-state philosophy.
(4) We give an explicit proof that Gamma-distributed frailty acting on a constant (exponential) baseline hazard generates the Pareto Type II (Lomax) distribution exactly (Theorem 5.2), and we show that the resulting hazard, like the log-logistic hazard, solves a Riccati-type ODE, so that frailty models fit naturally into the hazard-ODE framework.
(5) We add worked, fully normalized examples for the non-Markovian framework: the Mittag-Leffler density obtained from the fractional relaxation equation (Section 6.1), a delay-differential equation whose normalized solution is a geometric mixture of shifted Erlang laws (Proposition 6.1), and an integro-differential equation whose solution is a phase-type density (Proposition 6.2).
(6) Every closed-form example is accompanied by a numerical illustration in which the closed form is compared with an independent numerical solution of the governing equation (and, where possible, with Monte Carlo

samples). The parameter values are given in each caption, and the Python code that reproduces all figures and checks is provided as supplementary material.

Relation to existing reviews. Each framework has been surveyed separately: the Pearson system in the monographs of Ord [50] and Johnson, Kotz and Balakrishnan [38]; Pearson diffusions in [24]; hazard-based and frailty models in [35, 42]; fractional and anomalous-diffusion models in [44, 46]; and diffusion-based generative models in [70]. These works are either classical and univariate, or modern and focused on machine learning. The present review differs in three ways: (i) it treats all seven frameworks within one notation and one well-posedness/normalizability criterion (Theorem 2.1); (ii) it states the key closed forms (Theorems 3.1 and 5.2, Corollary 3.2, and Propositions 6.1 and 6.2) with complete proofs and numerical checks; and (iii) it ends with a practical selection procedure (Section 11). We do not claim that the individual classical distributions are new. What is new is the unified statement, the verified derivations, and the worked examples.

Related Work. The intersection of differential equations and probability theory has a rich history. Early work by Pearson [52] established the classification of continuous distributions via a first-order rational ODE [38, 50]; Forman and Sørensen [24] later identified the diffusion processes whose stationary laws are exactly the Pearson distributions (the *Pearson diffusions*), thereby linking Frameworks I and II. Concurrently, the development of stochastic processes by Fokker [23] and Planck [54] linked the time-evolution of PDFs to partial differential equations, with [59] remaining the standard reference. In the mid-20th century, survival analysis formalized the hazard rate, leading to ODE-based reliability models [43, 69].

In recent decades the focus has shifted toward complex and non-Markovian systems. Fractional calculus has been introduced to model anomalous diffusion and heavy-tailed phenomena [20, 44, 56], and delay differential equations have been used to capture feedback and multi-scale dynamics in the life sciences [62]. Nonlinear Fokker–Planck equations have been utilized in non-extensive statistical mechanics to derive Tsallis distributions [25, 55, 65, 66], and the classical theory of diffusions killed at a boundary gives rise to quasi-stationary distributions [16]. Numerically, the reliable integration of the ODEs underlying all of these frameworks rests on the classical theory of stiff and non-stiff solvers [31].

Most recently, the machine learning community has embraced DE-based distributions from the opposite direction: rather than solving a DE to *find* a distribution, one *learns* a DE whose solution matches data. Neural Ordinary Differential Equations [13] and Physics-Informed Neural Networks [57] allow flexible, data-driven fitting of implicitly defined distributions; normalizing flows [51, 58], including symmetry-preserving variants [22], and optimal transport [3, 67] establish deep connections between continuous-time dynamics and generative modeling. Score-based generative models [34, 63] and their subsequent refinement [41] make this connection precise: by Anderson's time-reversal theorem [4], the reverse-time dynamics of a diffusion process is determined by the score of its marginal densities, which can be learned directly from data [37]. Applications of these dynamical-systems tools now extend well beyond physics and finance, including stochastic chemical kinetics [28], molecular simulation [48], and the information-geometric analysis of learning dynamics itself [1]. Convex-analytic tools [10] underpin the normalization and identifiability arguments used throughout. This review unifies these historical and modern threads into a single, self-consistent framework.

Organization. Section 2 introduces a notation and states a general existence result for ODE-generated densities. Sections 3, 4, 5, 6, 7, 8, and 9 treat Frameworks I–VII in turn. Each framework section opens with a short paragraph *What is new here*, contains a multivariate extension and, where a closed form is available, a numerical illustration, and closes with a boxed *Practical summary* explaining when and how to use the framework. Section 10 develops computational inference for implicitly defined DE-based distributions. Section 11 compares the seven frameworks and gives a practical selection procedure. Section 12 discusses open problems, and Section 13 concludes.

Reading guide. Readers mainly interested in classical distribution theory may read Sections 2, 3, 5, and 6 together with Sections 4.1 and 4.2 and Section 11. Readers interested in modern generative and geometric modeling may read Section 2, Section 4.6, and Sections 7, 8, 9, and 10. The sections are written so that either route can be followed on its own.

2. Preliminaries and notation

Notation. The following conventions are used throughout.

Uncorrected Proof

- $f(x)$ or $p(x)$ denotes a probability density function (PDF) with respect to Lebesgue measure on its support $\mathcal{X} \subseteq \mathbb{R}^d$; an unnormalized solution of a generating equation is also written f, and its normalizing constant is $Z = \int_{\mathcal{X}} f$.
- ln denotes the natural logarithm; log is not used.
- For a nonnegative random variable X with density f, the *survival function* is $S(x) = \mathbb{P}(X > x) = \int_x^\infty f(t)\,\mathrm{d}t$ and the *hazard rate* (instantaneous failure rate) is $h(x) = f(x)/S(x) = -\frac{\mathrm{d}}{\mathrm{d}x} \ln S(x)$; $H(x) = \int_{x_0}^x h$ is the cumulative hazard.
- For a density p on $\mathbb{R}^d$, the *score function* is $\nabla_x \ln p(x)$, the gradient of the log-density with respect to the state variable (not with respect to parameters).
- For a diffusion $\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t$, μ is the drift and $\sigma^2(x)$ the *squared* diffusion coefficient (infinitesimal variance). Where a constant diffusion level is needed we write $\sigma^2(x) = 2D$, i.e., $D = \frac{1}{2}\sigma^2$ is the diffusion constant in the physicists' convention.
- In the Pearson equation (3.1), a_0, a_1 are the numerator coefficients and b_0, b_1, b_2 those of $Q(x) = b_0 + b_1 x + b_2 x^2$; the discriminant is $\Delta = b_1^2 - 4b_0 b_2$, so that $\Delta > 0$, $\Delta = 0$, and $\Delta < 0$ correspond to two distinct real roots, one double root, and two complex-conjugate roots of Q, respectively.

We write $\Gamma(\cdot)$ for the Euler gamma function, $B(a, b)$ for the beta function, and $E_\alpha(\cdot)$ for the one-parameter Mittag-Leffler function [29, 56]. All ODEs are assumed to hold on an open interval on which the relevant coefficient functions are continuous and the denominator (where it appears) is non-vanishing. When solutions are sought on a half-line $[x_0, \infty)$ or a finite interval, we implicitly refer to the appropriate classical existence theory for ordinary differential equations [6, 15, 31].

Because every framework in this review ultimately defines a density *implicitly*—through a generating equation rather than a closed formula—it is essential to state the minimal conditions under which such a definition is well posed. The following proposition provides the foundation.

Theorem 2.1 (Existence and uniqueness of ODE-generated log-densities)**.** *Let $I \subseteq \mathbb{R}$ be an open interval, $x_0 \in I$, and let $G : I \to \mathbb{R}$ be continuous. Consider the initial value problem*

$$\frac{\mathrm{d}}{\mathrm{d}x} \ln f(x) = G(x), \qquad \ln f(x_0) = c_0 \in \mathbb{R}. \tag{2.1}$$

Then Equation (2.1) possesses a unique solution on I, given by

$$\ln f(x) = c_0 + \int_{x_0}^x G(y)\,\mathrm{d}y.$$

Consequently $f(x) = \exp(\ln f(x)) > 0$ for all $x \in I$, and every solution is strictly positive. If, in addition,

$$Z := \int_{\mathcal{X}} f(x)\,\mathrm{d}x < \infty, \tag{2.2}$$

where $\mathcal{X} \subseteq I$ is the support of f, then

$$p(x) = \frac{f(x)}{Z},$$

defines a valid probability density function on $\mathcal{X}$.

Proof. Because G is continuous on I, it is integrable on every compact subinterval of I. Define

$$F(x) = c_0 + \int_{x_0}^x G(y)\,\mathrm{d}y, \qquad x \in I.$$

By the fundamental theorem of calculus, F is continuously differentiable on I and $F'(x) = G(x)$ for all $x \in I$. Moreover $F(x_0) = c_0$, hence F satisfies Equation (2.1). This establishes existence.

For uniqueness, let $\tilde{F}$ be any solution of Equation (2.1) on I. Then $(\tilde{F} - F)'(x) = G(x) - G(x) = 0$ for all $x \in I$, so $\tilde{F} - F$ is constant on I. From $\tilde{F}(x_0) = F(x_0) = c_0$ we obtain $\tilde{F} \equiv F$. Thus the solution is unique.

Exponentiating gives $f(x) = \exp(F(x)) > 0$ for all $x \in I$ because the exponential function is strictly positive on $\mathbb{R}$. Hence every solution of Equation (2.1) is automatically a positive function.

Uncorrected Proof

The constant Z defined in Equation (2.2) is the normalizing constant (partition function). Its finiteness depends on the asymptotic behavior of G as x approaches the boundary of $\mathcal{X}$. For instance, if $G(x) \sim -\alpha x$ with $\alpha > 0$ as $x \to \infty$, then f decays like a Gaussian and $Z < \infty$. If $G(x) \sim -c/x$ as $x \to 0^+$, then $f(x) \asymp x^{-c}$ near 0, which is integrable if and only if $c < 1$; likewise, if $G(x) \sim -c/x$ as $x \to \infty$, then $f(x) \asymp x^{-c}$ and the right tail is integrable if and only if $c > 1$. When $Z < \infty$, the function $p(x) = f(x)/Z$ is non-negative and integrates to one, thus a valid probability density. □

Remark 2.2. Theorem 2.1 separates the two main tasks when generating a density from an ODE:

(1) **Well-posedness.** Continuity of G is sufficient; the solution is automatically positive. No extra conditions (such as Lipschitz continuity) are needed because the right-hand side depends only on x, not on $\ln f$.
(2) **Integrability.** The finiteness of Z is the core statistical requirement. It imposes constraints on the parameters of G and determines the admissible support $\mathcal{X}$. In many classical families (Pearson, steady-state Fokker–Planck, hazard-rate ODEs) these constraints select the valid parameter region.

The simplicity of this separation underpins all frameworks reviewed in the subsequent sections.

Integrability and tail classification. The finiteness of the normalizing integral Z is intimately linked to the tail behavior of the unnormalized density $f(x)$. In the literature on extreme value theory [21], distributions are broadly classified according to their right-tail decay: exponential-type tails (e.g., normal, gamma) lead to convergent Z for all admissible parameter values, while power-law tails (e.g., Pareto, Student-t) require restrictions on the tail index for Z to be finite. In the DE-based framework, these conditions translate directly into constraints on the coefficients of the generating ODE.

Numerical normalization. When Z cannot be evaluated analytically, numerical integration becomes necessary. For one-dimensional densities on half-lines, Gauss–Laguerre quadrature or adaptive Gauss–Kronrod schemes are often efficient [19]. In higher dimensions, direct quadrature becomes intractable, and one resorts to Monte Carlo estimates or importance sampling [60]. These computational aspects are treated in depth in Section 10.

Connection to exponential families and convex analysis. Many ODE-generated densities belong to the class of exponential families [11], where the log-density is linear in the parameters and the log-partition function $\ln Z$ is a convex function of the parameters [8, 10]. This convexity guarantees that maximum likelihood estimation is a convex optimization problem. For more general DE-defined densities outside the exponential family, convexity is lost, and the integrability condition in Equation (2.2) may fail for a subset of the parameter space, so that any numerical optimization routine must be paired with an explicit check of integrability.

Further background references. Throughout this review we draw on standard texts for ordinary differential equations [6, 15, 31], fractional calculus and Mittag-Leffler functions [20, 29, 56], stochastic processes and the Fokker–Planck equation [49, 59], survival analysis [42], and information geometry [2]. These works provide the rigorous foundations for the constructions that follow.

3. Framework I: the Pearson system and rational ODEs

What is new here. We give a complete derivation of the three closed-form solutions of the Pearson equation (Theorem 3.1), verified against direct numerical integration, and we deduce from it one statement of the support, normalizability, and tail conditions of all Pearson types (Corollary 3.2 and Table 1). A consequence that is often overlooked is that Types IV–VII all have *power-law* tails with the same index a_1/b_2.

3.1. The Generalized Pearson ODE.

Karl Pearson (1895) [52] proposed that many continuous distributions satisfy a first-order linear ODE for the log-density:

$$\frac{1}{p(x)}\frac{\mathrm{d}p(x)}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\ln p(x) = -\frac{a_0 + a_1 x}{b_0 + b_1 x + b_2 x^2}. \tag{3.1}$$

This equation defines a rich family of distributions (Normal, Beta, Gamma, inverse Gamma, Beta-prime, Student's t, and Pearson Type IV) depending on the roots of the quadratic denominator $Q(x) = b_0 + b_1 x + b_2 x^2$ [38, 50].

Uncorrected Proof

Equation (3.1) is a special case of Equation (2.1) with the rational right-hand side $G = -(a_0+a_1x)/Q$, so Theorem 2.1 applies on every interval on which Q does not vanish.

Theorem 3.1 (Closed-form solution of the generalized Pearson ODE)**.** *Let $Q(x) = b_0+b_1x+b_2x^2$ have real coefficients with $b_2 \neq 0$, and let $\Delta = b_1^2 - 4b_0b_2$. On any open interval on which Q does not vanish, the solutions of*

$$\frac{\mathrm{d}}{\mathrm{d}x}\ln p(x) = -\frac{a_0 + a_1x}{b_0 + b_1x + b_2x^2},$$

are up to a positive multiplicative constant C, as follows.

(1) *If $\Delta > 0$, let $r_1 \neq r_2$ be the real roots of Q. Then*

$$p(x) = C\,|x - r_1|^{\alpha_1}\,|x - r_2|^{\alpha_2}, \qquad \alpha_i = -\frac{a_0 + a_1r_i}{b_2(r_i - r_j)}, \quad j \neq i.$$

(2) *If $\Delta = 0$, let $r = -b_1/(2b_2)$ be the double root. Then*

$$p(x) = C\,|x - r|^{-a_1/b_2}\exp\!\Big(\frac{a_0 + a_1r}{b_2\,(x - r)}\Big).$$

(3) *If $\Delta < 0$, write the complex-conjugate roots as $u \pm iv$ with $u = -b_1/(2b_2)$ and $v = \sqrt{4b_0b_2 - b_1^2}/(2|b_2|) > 0$. Then*

$$p(x) = C\,\big((x-u)^2 + v^2\big)^{-a_1/(2b_2)}\exp\Big[-\frac{a_0 + a_1u}{b_2\,v}\arctan\frac{x-u}{v}\Big] \tag{3.2}$$

$$= C\,\big((x-u)^2 + v^2\big)^{-a_1/(2b_2)}\exp\Big[-\frac{2a_0b_2 - a_1b_1}{2b_2^2\,v}\arctan\frac{x-u}{v}\Big].$$

In each case C is chosen so that $\int_{\mathcal{X}} p(x)\,\mathrm{d}x = 1$ on the support $\mathcal{X}$ described in Corollary 3.2.

Proof. Throughout we use the identity $a_0 + a_1x = a_1(x - s) + (a_0 + a_1s)$, valid for every real or complex s.

Case 1: $\Delta > 0$ (distinct real roots). Factor $Q(x) = b_2(x - r_1)(x - r_2)$ and seek constants with

$$-\frac{a_0 + a_1x}{b_2(x - r_1)(x - r_2)} = \frac{\alpha_1}{x - r_1} + \frac{\alpha_2}{x - r_2}.$$

Multiplying by $x - r_1$ and letting $x \to r_1$ gives $\alpha_1 = -(a_0+a_1r_1)/\big(b_2(r_1 - r_2)\big)$; the formula for α_2 follows by symmetry. Integrating,

$$\ln p(x) = \alpha_1 \ln|x - r_1| + \alpha_2 \ln|x - r_2| + \text{const},$$

and exponentiation gives $p(x) = C|x - r_1|^{\alpha_1}|x - r_2|^{\alpha_2}$. For later use we note that

$$\alpha_1 + \alpha_2 = \frac{-(a_0 + a_1r_1) + (a_0 + a_1r_2)}{b_2(r_1 - r_2)} = -\frac{a_1}{b_2}. \tag{3.3}$$

Case 2: $\Delta = 0$ (double root). Here $Q(x) = b_2(x - r)^2$ with $r = -b_1/(2b_2)$. Using the identity with $s = r$,

$$-\frac{a_0 + a_1x}{b_2(x - r)^2} = -\frac{a_1}{b_2}\,\frac{1}{x - r} - \frac{a_0 + a_1r}{b_2}\,\frac{1}{(x - r)^2}.$$

Since $\int (x - r)^{-2}\,\mathrm{d}x = -(x - r)^{-1}$, integration gives

$$\ln p(x) = -\frac{a_1}{b_2}\ln|x - r| + \frac{a_0 + a_1r}{b_2\,(x - r)} + \text{const},$$

which is the stated density.

Case 3: $\Delta < 0$ (complex-conjugate roots). Completing the square gives $Q(x) = b_2\big[(x - u)^2 + v^2\big]$ with $u = -b_1/(2b_2)$ and $v^2 = (4b_0b_2 - b_1^2)/(4b_2^2) > 0$. Using the identity with $s = u$,

$$-\frac{a_0 + a_1x}{b_2[(x - u)^2 + v^2]} = -\frac{a_1}{2b_2}\,\frac{2(x - u)}{(x - u)^2 + v^2} - \frac{a_0 + a_1u}{b_2}\,\frac{1}{(x - u)^2 + v^2}.$$

Uncorrected Proof

The two terms are integrated with

$$\int \frac{2(x-u)}{(x-u)^2+v^2}\,\mathrm{d}x = \ln\!\big((x-u)^2+v^2\big), \qquad \int \frac{\mathrm{d}x}{(x-u)^2+v^2} = \frac{1}{v}\arctan\frac{x-u}{v},$$

so that

$$\ln p(x) = -\frac{a_1}{2b_2}\ln\!\big((x-u)^2+v^2\big) - \frac{a_0+a_1u}{b_2\,v}\arctan\frac{x-u}{v} + \text{const.}$$

No linear term in x appears, because the numerator has been split exactly into a multiple of $Q'(x)/b_2 = 2(x-u)$ and a constant. Exponentiation gives the first form in Equation (3.2); the second follows from $a_0+a_1u = (2a_0b_2-a_1b_1)/(2b_2)$.

In all three cases differentiation of $\ln p$ recovers the right-hand side of Equation (3.1), and the formulas agree with a direct numerical integration of Equation (3.1) to relative accuracy 10^{-9} for the parameter values of Figure 2 (supplementary code). □

Corollary 3.2 (Support, normalizability, and tail index)**.** *Assume $b_2 \neq 0$ and let $\mathcal{X}$ be a maximal interval on which $Q \neq 0$ and $p > 0$.*

1. *If $\mathcal{X}$ is unbounded, then $p(x) \asymp |x|^{-a_1/b_2}$ as $|x| \to \infty$ in $\mathcal{X}$. Hence the tail is integrable if and only if $a_1/b_2 > 1$, and $\mathbb{E}|X|^k < \infty$ if and only if $k < a_1/b_2 - 1$.*
2. *At a finite endpoint that is a simple root r_i (Case 1) the density behaves like $|x-r_i|^{\alpha_i}$ and is integrable if and only if $\alpha_i > -1$.*
3. *At a double root r (Case 2) the density is integrable on (r,∞) if and only if $(a_0+a_1r)/b_2 < 0$ (and on $(-\infty,r)$ if and only if $(a_0+a_1r)/b_2 > 0$), together with $a_1/b_2 > 1$ for the infinite end.*

The resulting classification is summarized in Table 1.

Proof. (1) In Case 1, $|x-r_1|^{\alpha_1}|x-r_2|^{\alpha_2} \sim |x|^{\alpha_1+\alpha_2} = |x|^{-a_1/b_2}$ by Equation (3.3). In Case 2, $\exp\!\big((a_0+a_1r)/(b_2(x-r))\big) \to 1$, so $p \sim C|x|^{-a_1/b_2}$. In Case 3, $\arctan\!\big((x-u)/v\big) \to \pm\pi/2$, so the exponential factor tends to a positive constant and $p \asymp |x|^{-a_1/b_2}$. Equivalently, $G(x) = -(a_0+a_1x)/Q(x) \sim -(a_1/b_2)/x$, and the remark following Theorem 2.1 applies. The moment condition follows from $\int^\infty x^{k-a_1/b_2}\,\mathrm{d}x < \infty \iff k - a_1/b_2 < -1$. (2) is immediate from the Case 1 formula. (3) As $x \to r^+$, $\exp\!\big(\beta/(x-r)\big)$ with $\beta = (a_0+a_1r)/b_2$ tends to 0 faster than any power if $\beta < 0$ and to $+\infty$ if $\beta > 0$; the case $x \to r^-$ is symmetric. □

3.2. **Classification and Examples.** Table 1 lists the classical Pearson types with the conditions of Corollary 3.2. The Normal and Type III laws arise from the degenerate cases $b_2 = 0$, which are not covered by Theorem 3.1 but are integrated directly: for $b_2 = b_1 = 0$ one obtains $p \propto \exp\{-(a_0x + a_1x^2/2)/b_0\}$, i.e., $\mathcal{N}(-a_0/a_1,\, b_0/a_1)$; for $b_2 = 0 \neq b_1$, writing $a_0 + a_1x = \frac{a_1}{b_1}(b_0 + b_1x) + \big(a_0 - \frac{a_1b_0}{b_1}\big)$ gives $p \propto |x + b_0/b_1|^m e^{-(a_1/b_1)x}$ with $m = (a_1b_0 - a_0b_1)/b_1^2$, a shifted Gamma law with shape $m+1$ and rate a_1/b_1.

TABLE 1. The Pearson system in Equation (3.1): form of the solution (Theorem 3.1), support, and the conditions of Corollary 3.2. Here $r_1 < r_2$ are the real roots of Q, $\beta = (a_0+a_1r)/b_2$, $\nu = (a_0+a_1u)/(b_2v)$, and $m = (a_1b_0 - a_0b_1)/b_1^2$. "Tail" refers to the right tail (both tails for Types IV and VII).

Type (law)	Condition on Q	Density up to a constant	Support	Normalizable iff	Tail
Normal	$b_1 = b_2 = 0$	$\exp\{-(a_0x + \frac{1}{2}a_1x^2)/b_0\}$	$\mathbb{R}$	$a_1/b_0 > 0$	Gaussian
III (Gamma)	$b_2 = 0 \neq b_1$	$\lvert x + b_0/b_1\rvert^m\, e^{-(a_1/b_1)x}$	$x > -b_0/b_1 (b_1 > 0)$	$a_1/b_1 > 0,\ m > -1$	exponential
I (Beta)	$\Delta > 0$	$(x-r_1)^{\alpha_1}(r_2-x)^{\alpha_2}$	(r_1, r_2)	$\alpha_1 > -1,\ \alpha_2 > -1$	bounded
II (symm. Beta)	$\Delta > 0,\ b_1 = a_0 = 0$	$(r_2^2 - x^2)^{-a_1/(2b_2)}$	$(-r_2, r_2)$	$-a_1/(2b_2) > -1$	bounded
VI (Beta prime)	$\Delta > 0$	$(x-r_1)^{\alpha_1}(x-r_2)^{\alpha_2}$	(r_2, ∞)	$\alpha_2 > -1,\ a_1/b_2 > 1$	power, index a_1/b_2
V (inverse Gamma)	$\Delta = 0$	$(x-r)^{-a_1/b_2} e^{\beta/(x-r)}$	(r, ∞)	$\beta < 0,\ a_1/b_2 > 1$	power, index a_1/b_2
IV	$\Delta < 0$	$((x-u)^2+v^2)^{-a_1/(2b_2)} e^{-\nu \arctan\frac{x-u}{v}}$	$\mathbb{R}$	$a_1/b_2 > 1$	power, index a_1/b_2
VII (Student t)	$\Delta < 0,\ b_1 = a_0 = 0$	$(x^2+v^2)^{-a_1/(2b_2)}$	$\mathbb{R}$	$a_1/b_2 > 1$	power; t_ν with $\nu = a_1/b_2 - 1$

Three remarks help in reading Table 1. First, Types I and VI use the same Case 1 formula; they differ only in whether p is taken between the two roots or beyond one of them (the mirror-image support $(-\infty, r_1)$ is also possible

Uncorrected Proof

for Type VI). Second, the Type V law with support (r, ∞) is the inverse Gamma distribution with shape $a_1/b_2 - 1$ and scale $-\beta$, i.e., $p(x) \propto (x-r)^{-(a_1/b_2-1)-1} e^{\beta/(x-r)}$. Third, Types IV, V, VI and VII all have power-law tails with the *same* index a_1/b_2 (Corollary 3.2); the arctan factor of Type IV only produces skewness, since it tends to the constants $e^{\mp\nu\pi/2}$ in the two tails. Consequently the Pearson system is fully capable of producing Pareto-type tails, a point used in Section 11.

3.3. **Multivariate Extension: Copula-Pearson.** While the classical Pearson system is univariate, a multivariate extension can be constructed by specifying a Pearson ODE for each marginal and coupling them via a copula. For example, an Archimedean copula with generator ϕ yields a joint density

$$p(x_1, \ldots, x_d) = \Big[\prod_{i=1}^{d} p_i(x_i)\Big] c(F_1(x_1), \ldots, F_d(x_d)),$$

where F_i is the distribution function of the ith marginal and the copula density c is derived from ϕ. This allows each marginal to follow a different Pearson type, while the dependence structure is defined by the copula generator, which itself can be defined via an ODE (e.g., the Clayton generator $\phi(t) = (t^{-\theta} - 1)/\theta$ solves $\phi'(t) = -t^{-(1+\theta)}$, $\phi(1) = 0$). Alternatively, one may define a multivariate Pearson system by generalizing Equation (3.1) to a gradient equation $\nabla \ln p = -Ax + b$ with a positive-definite matrix A, which yields the multivariate Normal law $\mathcal{N}(A^{-1}b, A^{-1})$.

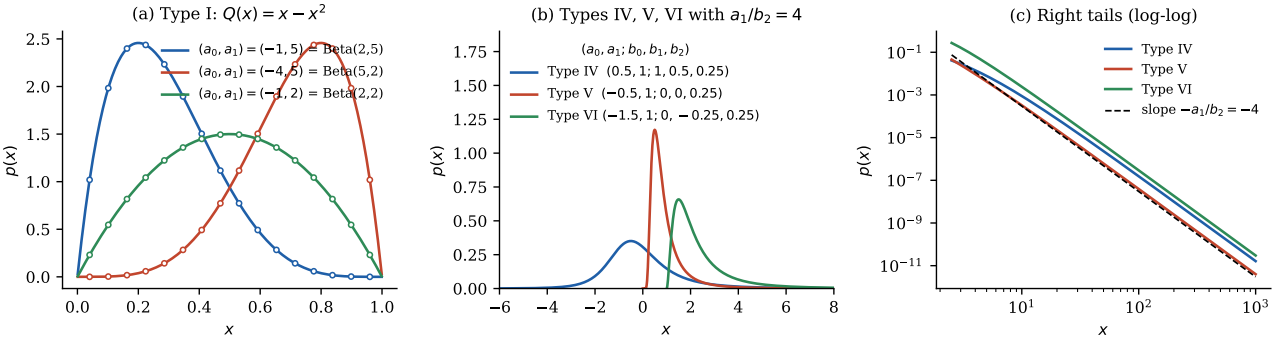


FIGURE 2. Densities generated by the Pearson equation (3.1). Lines: closed forms of Theorem 3.1, normalized by adaptive quadrature; circles: direct numerical integration of Equation (3.1) (Runge–Kutta, relative tolerance 10^{-10}). (a) Type I with $Q(x) = x - x^2$ ($b_0 = 0$, $b_1 = 1$, $b_2 = -1$) and $(a_0, a_1) = (-1, 5), (-4, 5), (-1, 2)$, which give Beta$(2, 5)$, Beta$(5, 2)$ and Beta$(2, 2)$. (b) Type IV $(a_0, a_1; b_0, b_1, b_2) = (0.5, 1; 1, 0.5, 0.25)$, Type V $(-0.5, 1; 0, 0, 0.25)$ (inverse Gamma with shape 3 and scale 2) and Type VI $(-1.5, 1; 0, -0.25, 0.25)$ (support $x > 1$). (c) The right tails of the three densities in (b) on log–log axes: all decay with the slope $-a_1/b_2 = -4$ predicted by Corollary 3.2.

3.4. **Numerical Illustration.** Figure 2(a) shows Beta densities obtained from Case 1 of Theorem 3.1; panel (b) shows the three unbounded types with a common tail index $a_1/b_2 = 4$, and panel (c) confirms the common power-law decay. In every panel the closed form and the numerical solution of Equation (3.1) coincide to plotting accuracy.

Practical summary: Framework I. *Use when* a univariate sample must be matched by its first four moments, or when a closed-form density with a clear support is needed. *How:* estimate $(a_0, a_1, b_0, b_1, b_2)$ by the method of moments or maximum likelihood; read the type from Δ and the support from the roots (Table 1); check $a_1/b_2 > 1$ whenever the support is unbounded. *Tail index* $= a_1/b_2$; finite moments of order $k < a_1/b_2 - 1$. *Limitations:* unimodal (or J/U-shaped) only; multivariate versions require copulas.

Uncorrected Proof

4. Framework II: Fokker–Planck, Jump-Diffusion, and Score-Based generative models

What is new here. We separate explicitly the genuine stationary laws of a diffusion from its fixed-time transition laws (Table 2 and Remark 4.2), give an explicit drift that produces the generalized Gamma law, state the nonlinear Fokker–Planck equation in the form whose steady state is the q-Gaussian, and connect the steady-state construction to quasi-stationary distributions and to score-based generative models.

4.1. From Langevin to Fokker–Planck.

If a particle's motion is governed by a Stochastic Differential Equation (SDE) in the Langevin form

$$\mathrm{d}X_t = \mu(X_t, t)\mathrm{d}t + \sigma(X_t, t)\mathrm{d}W_t, \tag{4.1}$$

where W_t is a standard Wiener process, μ is the drift, and σ^2 is the infinitesimal variance, the time evolution of its PDF $p(x, t)$ is governed by the Fokker–Planck (Kolmogorov forward) equation [49, 59]:

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}[\mu(x, t)p] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2(x, t)p]. \tag{4.2}$$

4.2. Steady-State Solutions.

Theorem 4.1 (Steady-state Fokker–Planck density)**.** *Suppose μ, σ are time-independent, $\sigma^2 > 0$ on the state interval, and the boundary conditions (natural or reflecting) force the probability current to vanish. Then any time-independent solution of Equation (4.2) has the form*

$$p(x) = \frac{C}{\sigma^2(x)} \exp\Big(\int^x \frac{2\mu(y)}{\sigma^2(y)}\mathrm{d}y\Big), \tag{4.3}$$

and it is a stationary probability *density if and only if the right-hand side is integrable over the state interval.*

Proof. Setting $\partial_t p = 0$ in Equation (4.2) and integrating once in x gives $-\mu p + \frac{1}{2}\partial_x[\sigma^2 p] = -J$, where J is the (constant) probability current. The boundary conditions force $J = 0$. The resulting first-order ODE $\frac{\mathrm{d}}{\mathrm{d}x}(\sigma^2 p) = 2\mu p$ is linear; writing $u = \sigma^2 p$ yields $u'/u = 2\mu/\sigma^2$, which integrates to $u(x) = C\exp(\int^x 2\mu/\sigma^2)$. Dividing by σ^2 gives Equation (4.3); the integrability statement is Theorem 2.1. □

Equation (4.3) is again of the form in Equation (2.1), with $G = (2\mu - (\sigma^2)')/\sigma^2$. In particular, if μ is linear and σ^2 is a polynomial of degree at most two, G is exactly the Pearson right-hand side of Equation (3.1); these are the *Pearson diffusions* of Forman and Sørensen [24], whose stationary laws are the Pearson distributions of Section 3. By designing specific drift and diffusion functions, one generates distributions rooted in physical or financial dynamics (Table 2).

Table 2. Distributions associated with the Fokker–Planck equation (4.2) with $\sigma^2(x) = 2D(x)$. Rows marked S are genuine stationary laws, i.e., normalizable solutions of Equation (4.3) and limits as $t \to \infty$. Rows marked T (Brownian motion and geometric Brownian motion) have no stationary probability density; the law listed is the fixed-time *transition* density $p(x, t \mid x_0)$ at time $t > 0$ (see Remark 4.2).

Process	Drift $\mu(x)$	$D(x) = \frac{1}{2}\sigma^2(x)$	Resulting law	S/T
Brownian motion	0	D	Normal $\mathcal{N}(x_0,\, 2Dt)$ at time t	T
Geometric BM	μx	Dx^2	Lognormal: $\ln X_t \sim \mathcal{N}\big(\ln x_0 + (\mu - D)t,\, 2Dt\big)$	T
Ornstein–Uhlenbeck	$-\theta x$	D	Normal $\mathcal{N}(0,\, D/\theta)$	S
Cox–Ingersoll–Ross	$\theta(\alpha - x)$	Dx	Gamma, shape $\theta\alpha/D$, rate θ/D (scaled χ^2)	S
Nonlinear-damping GBM	$D[(c{+}1)x - qs^{-q}x^{q+1}]$	Dx^2	Generalized Gamma $\propto x^{c-1}e^{-(x/s)^q}$	S
Pearson diffusion	$-\theta(x - m)$	$\theta\, Q(x)$	Pearson law with $a_0 + a_1 x = (x - m) + Q'(x)$ [24]	S

Uncorrected Proof

Remark 4.2 (Stationary densities versus transition densities)**.** Brownian motion and geometric Brownian motion (GBM) do not admit a normalizable stationary density. For Brownian motion, Equation (4.3) gives a constant, which is not integrable on $\mathbb{R}$; the Normal law in Table 2 is the transition density $p(x,t \mid x_0) = (4\pi Dt)^{-1/2}\exp\{-(x-x_0)^2/(4Dt)\}$, whose variance $2Dt$ grows without bound. For GBM, applying Equation (4.3) formally to $\mu(x) = \mu x$, $\sigma^2(x) = 2Dx^2$ gives $p(x) \propto x^{\mu/D-2}$, a power law that is not integrable at $x = 0$ and $x = \infty$ simultaneously; the Lognormal law is the transition density obtained from Itô's lemma applied to $\ln X_t$. By contrast, the Ornstein–Uhlenbeck and CIR rows are genuine equilibria. For the CIR process, Equation (4.3) gives $p(x) \propto x^{\theta\alpha/D-1}\exp(-\theta x/D)$, a Gamma law whose rate θ/D is fixed by the drift and diffusion coefficients, while the shape $\theta\alpha/D$ varies with α.

For the Pearson-diffusion row of Table 2, $(\sigma^2 p)' = 2\mu p$ with $\sigma^2 = 2\theta Q$ gives $(\ln p)' = -[(x-m)+Q'(x)]/Q(x)$, which is Equation (3.1) with $a_0 + a_1 x = (x-m) + Q'(x)$. Not every drift–diffusion pair yields a probability law. For example, the power-law drift $\mu(x) = -\gamma x^{2\beta-1}$ with $\frac{1}{2}\sigma^2(x) = Dx^{2\beta}$ gives $p(x) \propto x^{-(\gamma/D+2\beta)}$, a pure power law that is not normalizable on $(0,\infty)$ without truncation. A genuine Generalized Gamma law is obtained instead from the "nonlinear-damping" row of Table 2: with $\sigma^2(x) = 2Dx^2$ and $\mu(x) = D[(c+1)x - qs^{-q}x^{q+1}]$ one has $2\mu/\sigma^2 = (c+1)/x - qs^{-q}x^{q-1}$, and Equation (4.3) gives $p(x) \propto x^{c-1}\exp\{-(x/s)^q\}$, which is normalizable for all $c, q, s > 0$.

4.3. **Nonlinear Fokker–Planck and Tsallis Statistics.** To model systems with long-range interactions or nonextensive entropy, the diffusion term of the Fokker–Planck equation is made nonlinear [25, 55, 66]:

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial x}\big[kx\,p(x,t)\big] + D\frac{\partial^2}{\partial x^2}\big[p(x,t)^\nu\big], \qquad k, D > 0,\ \nu > 0, \tag{4.4}$$

where the linear restoring drift $-kx$ is needed for a normalizable steady state (without it, Equation (4.4) is the porous-medium equation, whose solutions spread indefinitely). Setting the current to zero, $kxp + D(p^\nu)' = 0$, and integrating gives

$$p(x) = \Big[A - \frac{(\nu-1)k}{2D\nu}\,x^2\Big]_+^{1/(\nu-1)},$$

with $A > 0$ fixed by normalization. With the entropic index $q = 2 - \nu$ this is the **Tsallis q-Gaussian**: for $q < 1$ ($\nu > 1$) it has compact support, for $q \to 1$ it tends to the Normal law, and for $1 < q < 3$ it has power-law tails $p \asymp |x|^{-2/(q-1)}$. The mechanism generalizes to arbitrary nonlinearities $p \mapsto \Phi(p)$, connecting nonlinear Fokker–Planck theory to generalized-entropy statistical mechanics [25, 65].

4.4. **Lévy-Driven Jump-Diffusion and Stable Distributions.** The Langevin equation (4.1) assumes continuous sample paths. Adding a pure-jump Lévy process L_t with Lévy measure $\nu(\mathrm{d}z)$ gives the jump-diffusion SDE $\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t + \mathrm{d}L_t$, whose forward Kolmogorov equation generalizes Equation (4.2) to the partial integro-differential equation [5]

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}[\mu p] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2 p] + \int_{\mathbb{R}\setminus\{0\}}\big[p(x-z,t) - p(x,t)\big]\,\nu(\mathrm{d}z), \tag{4.5}$$

where, for symmetric Lévy measures with $\int \min(1,|z|)\nu(\mathrm{d}z) = \infty$, the integral is understood as a principal value. When $\nu(\mathrm{d}z) = c\,|z|^{-1-\alpha}\mathrm{d}z$ with $0 < \alpha < 2$ (the symmetric α-stable Lévy measure) and $\mu = \sigma = 0$, Equation (4.5) reduces to the space-fractional diffusion equation $\partial_t p = -c'(-\Delta)^{\alpha/2}p$, whose fundamental solutions are the symmetric **α-stable distributions** with characteristic function $\varphi_t(k) = \exp(-c't|k|^\alpha)$. They have no closed-form density in general, but include the Cauchy law ($\alpha = 1$) and, as the limit $\alpha = 2$ (no jumps), the Gaussian law. This *space*-fractional construction is mathematically distinct from the *time*-fractional construction of Section 6.1, even though both produce heavy-tailed or anomalous behavior.

4.5. **Quasi-Stationary Distributions from Killed Diffusions.** If the diffusion in Equation (4.1) is killed (absorbed) upon reaching a boundary $\partial\mathcal{X}$, no proper stationary distribution exists on $\mathcal{X}$ in the sense of Theorem 4.1. Conditional on survival, however, the process typically converges as $t \to \infty$ to a *quasi-stationary distribution* (QSD)

Uncorrected Proof

ψ, characterized as the leading eigenfunction of the Fokker–Planck operator $\mathcal{L}^*$ with absorbing (Dirichlet) boundary conditions [16]:

$$\mathcal{L}^*\psi = -\lambda_1\psi, \qquad \mathcal{L}^*\psi := -\partial_x[\mu\psi] + \tfrac{1}{2}\partial_x^2[\sigma^2\psi], \tag{4.6}$$

where $\lambda_1 > 0$ is the smallest eigenvalue of $-\mathcal{L}^*$. For example, Brownian motion with $\sigma^2 = 2D$ killed at the endpoints of $(0, L)$ has $\psi(x) = \frac{\pi}{2L}\sin(\pi x/L)$ and $\lambda_1 = D\pi^2/L^2$. This reframes Framework II's generating mechanism from a zero-eigenvalue problem ($\mathcal{L}^*p = 0$ in Theorem 4.1) to a first-eigenvalue problem.

4.6. **Score-Based Generative Models as Learned Reverse-Time Diffusions.** The most consequential recent development connected to Framework II is the use of Equation (4.2) to define a high-dimensional generative model directly from data [34, 63, 70]. Consider a forward SDE that diffuses data $X_0 \sim p_0$ toward a tractable prior:

$$\mathrm{d}X_t = f(X_t, t)\,\mathrm{d}t + g(t)\,\mathrm{d}W_t, \qquad t \in [0, T], \tag{4.7}$$

whose marginal densities p_t satisfy the (now time-dependent, multivariate) Fokker–Planck equation (4.2). Anderson's time-reversal theorem [4], used in this setting by [63], shows that the reverse-time process

$$\mathrm{d}X_t = \big[f(X_t, t) - g(t)^2\nabla_x \ln p_t(X_t)\big]\,\mathrm{d}t + g(t)\,\mathrm{d}\bar{W}_t, \tag{4.8}$$

run backward in time from p_T to p_0 with a reverse-time Wiener process $\bar{W}_t$, has *exactly the same marginals* p_t as the forward process in Equation (4.7). The only unknown quantity in Equation (4.8) is the **score function** $\nabla_x \ln p_t(x)$, which is estimated by a neural network trained by (denoising) score matching [37]; sampling then proceeds by numerically integrating Equation (4.8) from a simple Gaussian prior. Subsequent work [41] reparametrizes the noise schedule and preconditioning to improve sample quality.

Remark 4.3. This construction is the natural high-dimensional culmination of the DE-to-distribution philosophy: instead of choosing μ, σ analytically to obtain a *named* closed-form density (Table 2), one fixes a simple f, g in Equation (4.7) and lets a neural network implicitly parameterize the score, so that the resulting implicitly defined p_0 can be as complex as a distribution over natural images. In one dimension with $f = 0$ the link is exact and elementary: the steady-state relation in Equation (4.3) with constant σ^2 says precisely that the score of the stationary law is $2\mu/\sigma^2$.

Corrected Proof

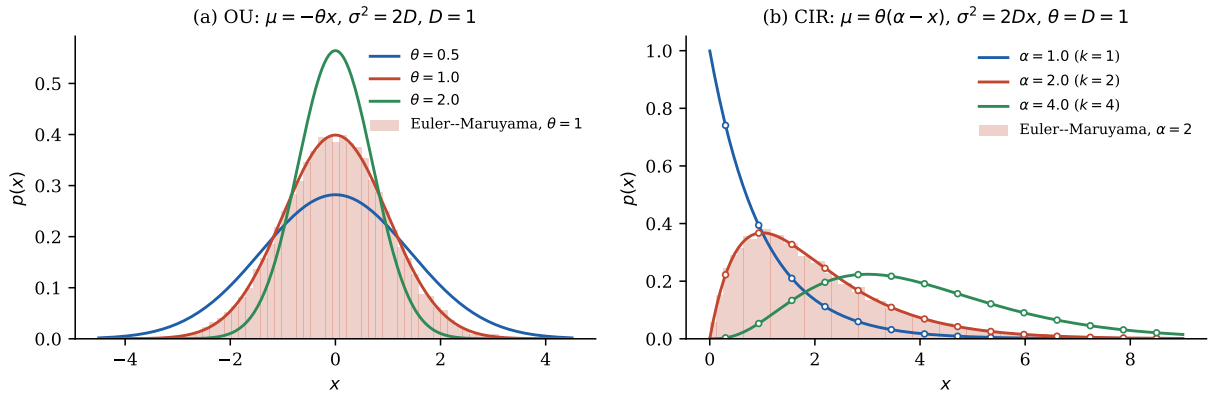


FIGURE 3. Genuine stationary densities from Theorem 4.1. (a) Ornstein–Uhlenbeck, $\mu(x) = -\theta x$, $\sigma^2 = 2D$ with $D = 1$ and $\theta \in \{0.5, 1, 2\}$: the stationary law is $\mathcal{N}(0, D/\theta)$. The histogram shows 2×10^4 Euler–Maruyama paths ($\theta = 1$, step 0.01, horizon 15). (b) Cox–Ingersoll–Ross, $\mu(x) = \theta(\alpha - x)$, $\sigma^2(x) = 2Dx$ with $\theta = D = 1$ and $\alpha \in \{1, 2, 4\}$, i.e., Gamma laws with rate $\theta/D = 1$ and shape $k = \theta\alpha/D \in \{1, 2, 4\}$. Lines: closed form; circles: numerical quadrature of Equation (4.3); histogram: 2×10^4 Euler–Maruyama paths ($\alpha = 2$, step 0.002, horizon 12).

4.7. **Numerical Illustration.** Figure 3 confirms that the densities obtained from Equation (4.3) are the long-run laws of the corresponding SDEs: the simulated histograms agree with the curves. As θ increases, the OU law concentrates; as α increases with θ/D fixed, the CIR law moves from the exponential ($k = 1$) toward a bell shape.

Practical summary: Framework II. *Use when* the data are long-run observations of a system with an identifiable drift (restoring force) and noise level, as in physics, finance, and population dynamics. *How:* specify μ and σ^2; compute Equation (4.3); verify integrability (Theorem 4.1) and that the law is truly stationary, not a transition law (Table 2); estimate parameters by likelihood of the stationary law or, for time series, of the transition density. Use jump terms for stable-type tails, the QSD for absorbed systems, and score-based models when the data are high-dimensional and no closed form is needed. *Limitations:* interpretability requires that μ, σ be physically motivated; learned scores are not interpretable.

5. Framework III: Hazard-Rate and Survival ODEs

What is new here. All entries of the hazard table are now exact ODEs: the log-logistic and Lomax hazards are shown to solve Riccati equations, and a genuine bathtub hazard is generated by a linear second-order ODE (the Siler model). Theorem 5.2 is illustrated by a Monte Carlo check.

5.1. **Survival ODEs.** In survival analysis and reliability, the *hazard rate* $h(x) \geq 0$ is the instantaneous failure rate at age x given survival up to x, i.e., $h(x) = \lim_{\delta \downarrow 0} \mathbb{P}(x < X \leq x + \delta \mid X > x)/\delta$ [42].

Theorem 5.1 (Hazard-density identity)**.** *For any density f with support $[x_0, \infty)$ and survival function $S(x) = \int_x^\infty f(t)\,\mathrm{d}t$, define the hazard $h(x) := f(x)/S(x)$. Then*

$$f(x) = h(x) \exp\Big(-\int_{x_0}^{x} h(t)\mathrm{d}t\Big), \qquad x \geq x_0. \tag{5.1}$$

Conversely, any locally integrable $h \geq 0$ with $\int_{x_0}^\infty h = \infty$ defines through Equation (5.1) a probability density on $[x_0, \infty)$.

Proof. By definition $h(x) = -S'(x)/S(x) = -\frac{\mathrm{d}}{\mathrm{d}x} \ln S(x)$. Integrating from x_0 to x and using $S(x_0) = 1$ gives $\ln S(x) = -\int_{x_0}^x h(t)\,\mathrm{d}t$. Hence $S(x) = \exp(-\int_{x_0}^x h)$, and $f = -S' = hS$ yields the stated expression. Conversely, $S := \exp(-\int_{x_0}^x h)$ decreases from 1 to 0 when $\int_{x_0}^\infty h = \infty$, so $f = -S'$ is a density. □

The converse part is what makes the hazard a convenient generator: normalization is automatic, and the only condition is divergence of the cumulative hazard. Instead of defining $h(x)$ algebraically, we define it as the solution to an ODE:

$$h'(x) = G(x, h(x); \theta), \tag{5.2}$$

or a higher-order analogue. The shape of the hazard function directly dictates the failure mode. Table 3 maps hazard ODEs to classical survival distributions; every row is an exact ODE whose solution is the stated hazard.

Table 3. Distributions generated via hazard-rate ODEs in Equation (5.2), with the hazard and the resulting density from Equation (5.1). All parameters are positive; α, β in the log-logistic row are its scale and shape.

Hazard ODE	Solution $h(x)$	Resulting law	Interpretation
$h'(x) = 0$	λ	Exponential	Constant failure rate (memoryless)
$h'(x) = \frac{k-1}{x} h(x)$	$\lambda k x^{k-1}$	Weibull	Wear-out ($k > 1$) or infant mortality ($k < 1$)
$h'(x) = \alpha h(x)$	$\lambda e^{\alpha x}$	Gompertz	Exponential aging (demography)
$h'(x) = \frac{\beta-1}{x} h(x) - h(x)^2$	$\frac{(\beta/\alpha)(x/\alpha)^{\beta-1}}{1+(x/\alpha)^\beta}$	Log-logistic	Rise then decline ($\beta > 1$)
$h'(x) = -h(x)^2/\alpha$	$\alpha/(\sigma + x)$	Lomax (Pareto II)	Decreasing; arises from Gamma frailty (Theorem 5.2)
$h''(x) = \gamma^2 [h(x) - c]$	$c + Ae^{-\gamma x} + Be^{\gamma x}$	Siler (bathtub)	Infant mortality, constant phase, and aging [61]

Uncorrected Proof

The log-logistic row follows from $\ln h = \text{const} + (\beta-1)\ln x - \ln\{1+(x/\alpha)^\beta\}$, whose derivative is $(\beta-1)/x - h(x)$ because $\frac{\mathrm{d}}{\mathrm{d}x}\ln\{1+(x/\alpha)^\beta\} = h(x)$. Thus the Weibull, log-logistic and Lomax hazards are all members of the Riccati family $h' = \frac{k}{x}h - \kappa h^2$, the quadratic term producing the eventual decline. In the Siler row, h has a bathtub shape when $A, B > 0$, with minimum at $x^* = \ln(A/B)/(2\gamma)$; the constant $c \geq 0$ is the background (accidental) hazard.

5.2. **Frailty and Mixture Models.** In survival analysis, unobserved heterogeneity (frailty) Z modifies the hazard multiplicatively: $h(x \mid Z) = Z\,h_0(x)$ for a baseline hazard h_0 [35]. The marginal (population-level) survival function is given by the Laplace transform of the frailty distribution evaluated at the baseline cumulative hazard $H_0(x) = \int_{x_0}^{x} h_0$:

$$S(x) = \mathbb{E}\big[e^{-Z H_0(x)}\big] = \mathcal{L}_Z(H_0(x)). \tag{5.3}$$

Theorem 5.2 (Gamma frailty on an exponential baseline yields Pareto)**.** *Let the baseline hazard be constant, $h_0(x) = 1/\sigma$ for $x \geq 0$ (so $H_0(x) = x/\sigma$), and let the frailty $Z \sim$ Gamma(shape $= \alpha$, rate $= 1$) across the population. Then the marginal law of the failure time is exactly the Pareto Type II (Lomax) law,*

$$S(x) = \Big(1+\frac{x}{\sigma}\Big)^{-\alpha}, \qquad f(x) = \frac{\alpha}{\sigma}\Big(1+\frac{x}{\sigma}\Big)^{-\alpha-1}, \qquad x \geq 0. \tag{5.4}$$

Proof. The Laplace transform of $Z \sim \text{Gamma}(\alpha, 1)$ is $\mathcal{L}_Z(s) = \int_0^\infty e^{-sz} z^{\alpha-1} e^{-z}\,\mathrm{d}z/\Gamma(\alpha) = (1+s)^{-\alpha}$ for $s > -1$. Substituting $H_0(x) = x/\sigma$ into Equation (5.3) gives $S(x) = (1+x/\sigma)^{-\alpha}$; differentiating $f = -S'$ gives the density. □

Remark 5.3. Theorem 5.2 shows that a constant (memoryless) individual hazard, when combined with Gamma-distributed frailty, produces a population-level hazard $h(x) = f(x)/S(x) = \alpha/(\sigma+x)$ that *decreases* with age: heavy-tailed, decreasing aggregate hazards can arise purely from mixing, without any individual mechanism for improving survival over time. This hazard solves the Riccati equation $h' = -h^2/\alpha$ of Table 3, so the frailty construction is itself a hazard ODE. The result is classical in the frailty literature [35]; we include the proof because it is the simplest exact example of a heavy (power-law, index α) tail generated by an ODE mechanism. If instead the frailty is normalized to have mean one, $Z \sim \text{Gamma}(\alpha, \alpha)$, the same argument gives $S(x) = (1 + x/(\alpha\sigma))^{-\alpha}$.

5.3. **Multivariate Extension: Shared Frailty and Copula Hazards.** Equation (5.3) extends naturally to correlated survival times via a *shared* frailty Z common to d individuals: $S(x_1, \ldots, x_d) = \mathcal{L}_Z\big(\sum_i H_{0,i}(x_i)\big)$. Written in terms of the marginal survival functions $S_i = \mathcal{L}_Z \circ H_{0,i}$, this is an Archimedean survival copula with generator $\mathcal{L}_Z^{-1}$; Gamma shared frailty generates the Clayton copula [35]. Alternatively, one may couple d marginal hazard ODEs in Equation (5.2) through cross-hazard terms $h_i'(x) = G_i(x, h_1, \ldots, h_d; \theta)$.

5.4. **Numerical Illustration.** Figure 4(a,b) show the classical monotone hazards; panel (c) shows the non-monotone shapes (rise-then-decline, decreasing, and bathtub) obtained from the ODEs in the last three rows of Table 3, and panel (d) shows that the Gamma-mixed exponential sample follows the Lomax law predicted by Theorem 5.2, with a much heavier tail than the baseline.

Practical summary: Framework III. *Use when* the variable is a lifetime, a failure time, or a time-to-event, possibly censored. *How:* choose the hazard shape from subject knowledge (monotone, unimodal, bathtub) using Table 3; normalization is automatic whenever $\int^\infty h = \infty$ (Theorem 5.1); fit by the censored-data likelihood $\prod_i h(x_i)^{\delta_i} S(x_i)$; add frailty (Equation (5.3)) to represent heterogeneity or to produce heavy tails. *Limitations:* support is a half-line; dependence between lifetimes requires shared frailty or copulas.

6. Framework IV: Fractional, Delay, and Integro-Differential Equations

What is new here. The three non-Markovian mechanisms of this framework (memory through a fractional derivative, a discrete delay, or a convolution kernel) are each worked out to an explicit, normalized probability law: the Mittag-Leffler density (Equation (6.2)), a geometric mixture of shifted Erlang laws generated by a delay equation (Proposition 6.1), and a phase-type density generated by an integro-differential equation (Proposition 6.2). Each closed form is checked against a numerical solution of its equation (Figures 5 and 6).

Uncorrected Proof

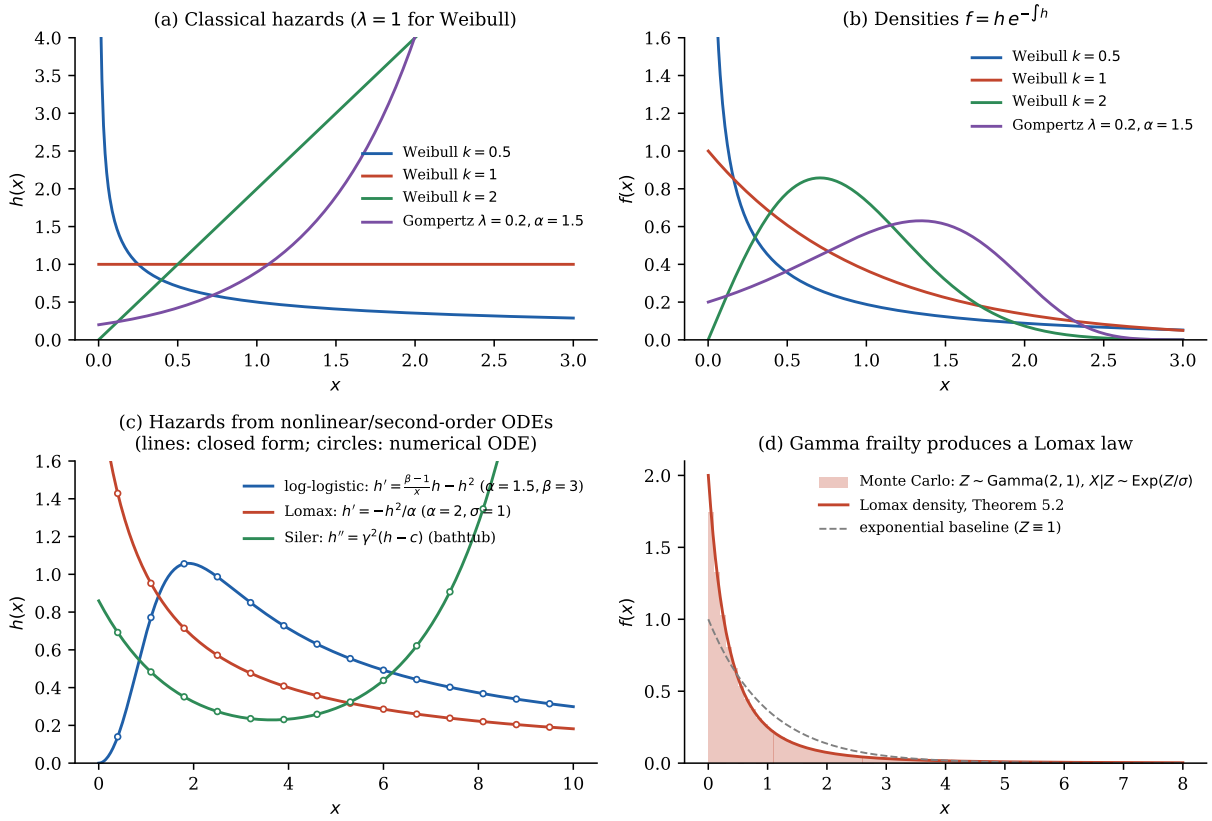


FIGURE 4. Hazard functions and the densities they generate through Equation (5.1). (a,b) Weibull hazards $h(x) = \lambda k x^{k-1}$ with $\lambda = 1$ and $k \in \{0.5, 1, 2\}$ (decreasing, constant, increasing) and the Gompertz hazard $h(x) = 0.2\, e^{1.5x}$; densities computed by numerical integration of the cumulative hazard. (c) Hazards from the nonlinear and second-order ODEs of Table 3: log-logistic ($\alpha = 1.5$, $\beta = 3$), Lomax ($\alpha = 2$, $\sigma = 1$) and Siler ($c = 0.05$, $A = 0.8$, $B = 0.01$, $\gamma = 0.6$). Lines: closed forms; circles: numerical solution of the ODE from the same initial values. (d) Monte Carlo check of Theorem 5.2: 2×10^5 draws of $Z \sim \text{Gamma}(2, 1)$ and $X \mid Z \sim \text{Exp}(\text{rate } Z/\sigma)$, $\sigma = 1$, compared with the Lomax density in Equation (5.4) and with the exponential baseline.

Uncorrected

Motivation. The ODEs of Frameworks I–III are local: the rate of change of f at x depends only on $f(x)$. Many lifetimes and waiting times, however, depend on the history of the process: a material relaxes with memory (viscoelasticity), a queue or a biological population reacts to its state a fixed time ago, and a system may alternate between hidden working and resting states. The three equations below encode such memory, and they produce laws that are heavy-tailed, multimodal, or of matrix-exponential type, which the local frameworks cannot produce.

6.1. **Fractional Calculus and the Mittag-Leffler Law.** Fractional ODEs introduce memory through non-integer derivatives [20, 56]. For $0 < \alpha < 1$ the Caputo derivative is

$$ {}^{C}D^{\alpha} f(x) = \frac{1}{\Gamma(1-\alpha)} \int_0^x \frac{f'(y)}{(x-y)^{\alpha}}\, \mathrm{d}y, $$

a weighted average of all past rates of change with the power-law memory kernel $(x-y)^{-\alpha}$. The fractional relaxation equation

$$ {}^{C}D^{\alpha}S(x) = -\lambda S(x), \qquad S(0)=1, \quad 0<\alpha\le 1,\ \lambda>0, \tag{6.1}$$

has the unique solution $S(x)=E_{\alpha}(-\lambda x^{\alpha})$, where $E_{\alpha}(z)=\sum_{k=0}^{\infty} z^{k}/\Gamma(\alpha k+1)$ is the one-parameter Mittag-Leffler function [29]. For $\alpha=1$ this is the exponential relaxation $e^{-\lambda x}$.

Probabilistic interpretation. For $0<\alpha\le 1$ the function $x\mapsto E_{\alpha}(-\lambda x^{\alpha})$ is completely monotone on $(0,\infty)$, decreases from 1 to 0, and is therefore a survival function [29, 53]. The solution of Equation (6.1) should thus be read as $S(x)=\mathbb{P}(X>x)$, and the corresponding **Mittag-Leffler density** is obtained by differentiation, using $\frac{\mathrm{d}}{\mathrm{d}x}E_{\alpha}(-\lambda x^{\alpha}) = -\lambda x^{\alpha-1}E_{\alpha,\alpha}(-\lambda x^{\alpha})$:

$$ f(x) = -S'(x) = \lambda\, x^{\alpha-1}E_{\alpha,\alpha}(-\lambda x^{\alpha}), \qquad x>0, \qquad E_{\alpha,\beta}(z)=\sum_{k=0}^{\infty}\frac{z^{k}}{\Gamma(\alpha k+\beta)}. \tag{6.2}$$

The density is automatically normalized, $\int_0^{\infty} f = S(0)-S(\infty)=1$, and its Laplace transform is $\mathbb{E}e^{-sX}=\lambda/(\lambda+s^{\alpha})$ [53]. As $x\to\infty$ and $0<\alpha<1$,

$$ S(x)\sim \frac{x^{-\alpha}}{\lambda\,\Gamma(1-\alpha)}, \qquad f(x)\sim \frac{\alpha\, x^{-\alpha-1}}{\lambda\,\Gamma(1-\alpha)}, $$

so the law interpolates between the exponential law ($\alpha=1$) and a Pareto-type tail of index α; in particular $\mathbb{E}X=\infty$ for $\alpha<1$. Samples can be drawn exactly as $X=\lambda^{-1/\alpha}(-\ln U)\big[\sin(\alpha\pi)/\tan(\alpha\pi V)-\cos(\alpha\pi)\big]^{1/\alpha}$ with U,V independent uniform on $(0,1)$ [26]. The Mittag-Leffler law is the waiting-time law of the continuous-time random walks that underlie anomalous diffusion [44, 46]; a rigorous treatment of well-posedness of Equation (6.1) is given in [20].

Time-fractional diffusion. Replacing ∂_t in the heat equation by a Caputo derivative of order $\beta\in(0,1)$, ${}^{C}\partial_t^{\beta}p = D\,\partial_x^2 p$ with $p(x,0)=\delta(x)$, gives a fundamental solution that is a symmetric probability density for every $t>0$, expressible through the Mainardi (M-Wright) function; it has a cusp at the origin, stretched-exponential tails, and subdiffusive variance $\mathbb{E}X_t^2 = 2Dt^{\beta}/\Gamma(1+\beta)$ [44, 45].

6.2. **Delay Differential Equations.** Delay differential equations (DDEs) model systems in which the current rate of change depends on a past state [62]:

$$ f'(x) = -a f(x) + b f(x-\tau), \qquad \tau>0. \tag{6.3}$$

Substituting $f(x)=e^{sx}$ gives the characteristic equation $s=-a+be^{-s\tau}$, a transcendental equation with infinitely many complex roots, so that general solutions are superpositions of damped oscillatory modes. Whether such a solution is a density must be checked: positivity can fail when $b<0$. For $0<b<a$ and a zero history, however, the solution is an explicit, normalized density.

Proposition 6.1 (A DDE-generated density)**.** *Let $0<b<a$ and $\tau>0$, and let f solve Equation (6.3) for $x>0$ with history $f(x)=0$ for $-\tau\le x<0$ and $f(0)=a-b$. Then*

$$ f(x) = (a-b)\sum_{n=0}^{\lfloor x/\tau\rfloor}\frac{b^{n}\,(x-n\tau)^{n}}{n!}\,e^{-a(x-n\tau)}, \qquad x\ge 0, \tag{6.4}$$

is a probability density on $[0,\infty)$. It is the law of $X=N\tau+\sum_{i=0}^{N}E_i$, where $E_0,E_1,\dots$ are i.i.d. exponential with rate a and N is independent with $\mathbb{P}(N=n)=(1-b/a)(b/a)^{n}$, $n=0,1,\dots$. In particular $\mathbb{E}X=\frac{1}{a}+\frac{b}{a-b}\big(\tau+\frac{1}{a}\big)$.

Proof. Let $g_n(x)=\frac{b^{n}(x-n\tau)^{n}}{n!}e^{-a(x-n\tau)}\mathbf{1}\{x\ge n\tau\}$. For $n\ge 1$, g_n is continuous and $g_n'(x)=-a g_n(x)+b\,g_{n-1}(x-\tau)$, while $g_0'=-a g_0$ on $x>0$. Summing over n shows that $f=(a-b)\sum_n g_n$ satisfies Equation (6.3) with the stated history and initial value (at the points $x=n\tau$ the derivative has different one-sided limits, as is usual for DDEs with

Uncorrected Proof

a discontinuous history), and the method of steps [62] shows that this solution is unique. Each term is nonnegative, so $f \geq 0$. Taking Laplace transforms in Equation (6.3) gives $s\hat{f}(s) - f(0) = -a\hat{f}(s) + be^{-s\tau}\hat{f}(s)$, hence

$$\hat{f}(s) = \frac{a-b}{s+a-b\,e^{-s\tau}}, \qquad \hat{f}(0) = 1,$$

so f integrates to one. On the other hand, with $\rho = b/a$,

$$\mathbb{E}e^{-sX} = \sum_{n\geq 0}(1-\rho)\rho^n e^{-sn\tau}\Big(\frac{a}{a+s}\Big)^{n+1} = \frac{(1-\rho)a}{a+s-b\,e^{-s\tau}} = \hat{f}(s),$$

which identifies the law of X. The mean follows from $\mathbb{E}N = b/(a-b)$. □

Proposition 6.1 has a transparent interpretation: an attempt (e.g., a transmission, a repair, or a cell-cycle phase) takes an exponential time with rate a and fails with probability b/a; each failure triggers a restart after a fixed delay τ. When $a\tau$ is large the density is *multimodal*, with modes near $n\tau + n/a$, $n = 0, 1, 2, \ldots$ (Figure 6(a)); the delay thus generates multimodality that no first-order ODE of Frameworks I–III can produce.

6.3. **Integro-Differential Equations and Phase-Type Distributions.** A third route to non-Markovian densities adds a convolution against a memory kernel $K \geq 0$:

$$f'(x) = -\lambda f(x) + \int_0^x K(x-y)\, f(y)\, \mathrm{d}y, \qquad x > 0. \tag{6.5}$$

The natural class of limits here is that of *phase-type* (PH) distributions [9, 47]. A PH law with m transient phases is the law of the absorption time of a continuous-time Markov chain with initial distribution $\boldsymbol{\pi}$ on the transient states and sub-generator T (nonnegative off-diagonal entries, nonpositive row sums, T invertible); its density and survival function are

$$f(x) = \boldsymbol{\pi}\, e^{Tx}\, \mathbf{t}, \qquad S(x) = \boldsymbol{\pi}\, e^{Tx}\mathbf{1}, \qquad \mathbf{t} = -T\mathbf{1}. \tag{6.6}$$

Equivalently, the row vector $\mathbf{p}(x) = \boldsymbol{\pi}e^{Tx}$ of phase probabilities solves the linear ODE system $\mathbf{p}' = \mathbf{p}T$, so PH densities are exactly the densities generated by linear constant-coefficient ODE systems with a Markovian (generator) structure. The following result makes the link with Equation (6.5) precise.

Proposition 6.2 (Integro-differential equations with exponential kernels)**.** *Let $K(u) = \sum_{k=1}^m c_k e^{-\mu_k u}$ with $c_k, \mu_k > 0$, and suppose $\kappa := \sum_{k=1}^m c_k/\mu_k < \lambda$. Then the solution of Equation (6.5) with $f(0) = \lambda - \kappa$ is the PH density in Equation (6.6) with $m+1$ phases, $\boldsymbol{\pi} = (1, 0, \ldots, 0)$ and*

$$T = \begin{pmatrix} -\lambda & q_1 & \cdots & q_m \\ \mu_1 & -\mu_1 & & \\ \vdots & & \ddots & \\ \mu_m & & & -\mu_m \end{pmatrix}, \qquad q_k = \frac{c_k}{\mu_k}.$$

In particular $f \geq 0$ and $\int_0^\infty f = 1$.

Proof. The matrix T is a sub-generator: its off-diagonal entries are nonnegative, rows $k \geq 1$ sum to zero, and row 0 sums to $-(\lambda - \kappa) < 0$, so absorption occurs only from phase 0, at rate $\lambda - \kappa$, and T is invertible. Let $\mathbf{p} = (p_0, \ldots, p_m)$ solve $\mathbf{p}' = \mathbf{p}T$, $\mathbf{p}(0) = \boldsymbol{\pi}$. The equations for $k \geq 1$, $p_k' = q_k p_0 - \mu_k p_k$, $p_k(0) = 0$, give $p_k(x) = q_k \int_0^x e^{-\mu_k(x-y)} p_0(y)\, \mathrm{d}y$. Substituting into $p_0' = -\lambda p_0 + \sum_k \mu_k p_k$ gives

$$p_0'(x) = -\lambda p_0(x) + \int_0^x \sum_k \mu_k q_k\, e^{-\mu_k(x-y)} p_0(y)\, \mathrm{d}y,$$

which is Equation (6.5) since $\mu_k q_k = c_k$. By linearity, $f = (\lambda - \kappa)p_0$ solves Equation (6.5) with $f(0) = \lambda - \kappa$, and $f = \boldsymbol{\pi}e^{Tx}\mathbf{t}$ because $\mathbf{t} = (\lambda - \kappa, 0, \ldots, 0)^\top$. It is a density because it is the density of an absorption time that is finite almost surely. □

Uncorrected Proof

In words, the system works (phase 0) until it either terminates, at rate $\lambda - \kappa$, or enters a pause of type k, at rate q_k; the pause lasts an exponential time with rate μ_k, after which work resumes. The kernel K records the memory induced by these hidden pauses. If $\kappa \geq \lambda$ the solution is not integrable. More generally, solutions of Equation (6.5) with exponential-sum kernels are *matrix-exponential* functions; they are densities only when positivity holds, and PH laws are the subclass with a Markovian representation [9]. PH distributions are dense (in the sense of weak convergence) in the class of all distributions on $(0, \infty)$, and they can be fitted to data by the EM algorithm [7]. Note that finite mixtures of exponentials alone are *not* dense, since all of them have decreasing densities; the Erlang (series) phases are essential.

Table 4 summarizes the distributions derived from these non-Markovian frameworks.

TABLE 4. Distributions from fractional, delay, and integro-differential equations.

Equation	**Solution / density**	**Distribution**	**Key feature**
Fractional relaxation (Equation (6.1))	$S = E_\alpha(-\lambda x^\alpha)$, $f = \lambda x^{\alpha-1} E_{\alpha,\alpha}(-\lambda x^\alpha)$	Mittag-Leffler	Power tail of index α; exponential for $\alpha = 1$
Time-fractional diffusion ${}^C\partial_t^\beta p = D\partial_x^2 p$	M-Wright (Mainardi) function	Symmetric M-Wright law	Cusp at 0; variance $\propto t^\beta$ (subdiffusion)
DDE with feedback (Equation (6.3)), $0 < b < a$	Equation (6.4)	Geometric mixture of shifted Erlang laws	Multimodal for large $a\tau$
Integro-differential (Equation (6.5)), exponential-sum kernel	$\boldsymbol{\pi} e^{Tx}\mathbf{t}$ (Proposition 6.2)	Phase-type	Dense in distributions on $(0, \infty)$; EM fitting

Uncorrected Proof

6.4. **Multivariate Extension.** Multivariate fractional relaxation replaces λ in Equation (6.1) with a matrix Λ, giving ${}^C D^\alpha \mathbf{S}(x) = -\Lambda \mathbf{S}(x)$ for a vector-valued function $\mathbf{S}$ of the scalar variable x, with matrix Mittag-Leffler solution $\mathbf{S}(x) = E_\alpha(-\Lambda x^\alpha)\mathbf{S}(0)$. When Λ is diagonalizable this reduces to independent scalar Mittag-Leffler components in the eigenbasis of Λ, while non-normal Λ induces cross-coupled memory effects. For $\alpha = 1$ and $\Lambda = -T^\top$ this is exactly the phase-probability system of the PH laws above, so fractional PH laws are obtained by the same substitution.

6.5. **Numerical Illustration.** Figure 5(a) shows the transition from exponential to power-law decay as α decreases, and panel (b) confirms that the density in Equation (6.2) matches exact samples. Figure 6(a) shows the multimodality generated by the delay, and Figure 6(b) the change from a single exponential ($c = 0$) to a two-phase law.

Practical summary: Framework IV. *Use when* waiting times or lifetimes show memory: power-law (Pareto-type) tails with infinite mean (Mittag-Leffler), regularly spaced multiple modes caused by retries or feedback delays (DDE, Proposition 6.1), or several hidden stages (phase-type, Proposition 6.2). *How:* interpret the solution of the equation as a survival function or density and check positivity and normalization (automatic in Equations (6.2), (6.4), and (6.6)); estimate α from the log–log slope of the empirical survival function, (a, b, τ) by maximum likelihood, and PH parameters by EM [7]. *Limitations:* Mittag-Leffler functions and DDE solutions require careful numerics (high-precision series or specialized algorithms); general DDE and integro-differential solutions need not be positive.

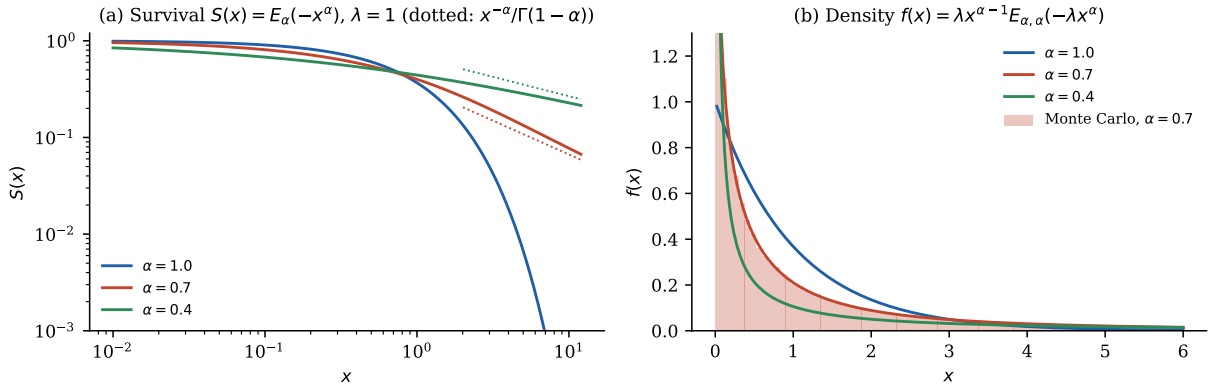


FIGURE 5. The Mittag-Leffler law generated by the fractional relaxation equation (6.1) with $\lambda = 1$ and $\alpha \in \{1, 0.7, 0.4\}$. The Mittag-Leffler functions are computed by summing their defining series in 80-digit arithmetic. (a) Survival functions $S(x) = E_\alpha(-x^\alpha)$ on log–log axes, with the power-law asymptotes $x^{-\alpha}/\Gamma(1-\alpha)$ (dotted); the exponential case $\alpha = 1$ decays faster than any power. (b) Densities in Equation (6.2), with a histogram of 3×10^5 exact samples for $\alpha = 0.7$.

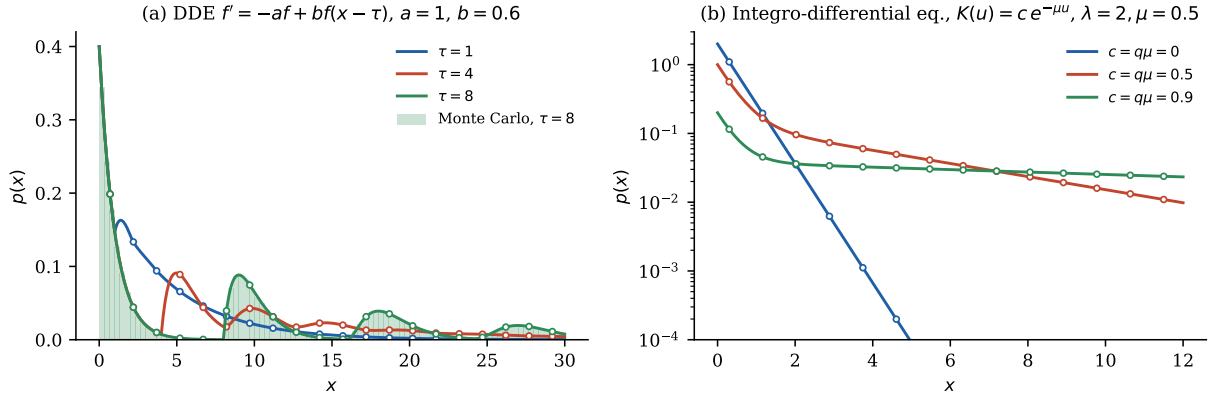


FIGURE 6. (a) Densities in Equation (6.4) generated by the delay equation (6.3) with $a = 1$, $b = 0.6$ and $\tau \in \{1, 4, 8\}$. Lines: closed form; circles: numerical solution of the DDE by the method of steps (Heun scheme, step 10^{-3}); histogram: 2×10^5 samples of $N\tau + \sum_{i=0}^{N} E_i$ for $\tau = 8$. (b) Phase-type densities generated by the integro-differential equation (6.5) with $K(u) = c\,e^{-\mu u}$, $\lambda = 2$, $\mu = 0.5$ and $c = q\mu \in \{0, 0.5, 0.9\}$ (log scale). Lines: $\boldsymbol{\pi} e^{Tx} \mathbf{t}$ from Proposition 6.2; circles: solution of the equivalent linear ODE system for $(f, \int_0^x e^{-\mu(x-y)} f(y)\mathrm{d}y)$. Larger c (more frequent pauses) produces a heavier, slowly decaying second exponential component.

7. Framework V: Partial Differential Equations and Wasserstein Flows

What is new here. In contrast to Frameworks I–IV, where the object of interest is a single limiting density, a PDE generates a *time-indexed family* of densities $\{p(\cdot, t)\}_{t>0}$ on $\mathbb{R}^d$. We make this distinction explicit, state the exact sense in which the Fokker–Planck equation is a Wasserstein gradient flow, and verify the heat-kernel family against a finite-difference solution.

Proof

7.1. **Spatiotemporal Distributions.** Moving beyond one-dimensional ODEs, PDEs generate multivariate and spatiotemporal distributions.

- **The heat equation:**

$$\frac{\partial p}{\partial t} = \Delta p, \qquad x \in \mathbb{R}^d,\ t > 0. \tag{7.1}$$

If the initial condition $p(x,0)$ is a Dirac delta at the origin, the solution of Equation (7.1) at time t is the heat kernel $p(x,t) = (4\pi t)^{-d/2}\exp\{-\|x\|^2/(4t)\}$, i.e., the multivariate Normal density $\mathcal{N}(0, 2tI_d)$. The heat equation therefore generates a *time-indexed family* $\{\mathcal{N}(0,2tI_d)\}_{t>0}$ of probability densities on $\mathbb{R}^d$, each of total mass one because Equation (7.1) conserves mass. Unlike the stationary constructions of Sections 3, 4, and 5, this family has no normalizable limit: Equation (7.1) has no nontrivial integrable steady state on $\mathbb{R}^d$, and $p(\cdot,t) \to 0$ uniformly as $t \to \infty$. For a general initial density p_0, $p(\cdot,t) = p_0 * \mathcal{N}(0,2tI_d)$ is the law of $X_0 + \sqrt{2}\,W_t$, which is exactly the forward (variance-exploding) noising process of the score-based models in Section 4.6.

- **The Schrödinger equation:** In quantum mechanics, the squared modulus $|\psi(x,t)|^2$ of the normalized wavefunction is a probability density. For the harmonic potential $V(x) = \frac{1}{2}\omega^2 x^2$ (with $\hbar = m = 1$), the stationary states give $|\psi_n(x)|^2 \propto H_n(\sqrt{\omega}x)^2 e^{-\omega x^2}$, where H_n are Hermite polynomials; for $n \geq 1$ these densities have n nodes (zero-probability points), a feature with no classical drift–diffusion analogue.

7.2. **Wasserstein Gradient Flows.** A modern and profound connection between PDEs and distributions arises in optimal transport [3, 67]. Many probability distributions can be viewed as the steady state of a Wasserstein gradient flow

$$\frac{\partial p}{\partial t} = \nabla \cdot \Big(p \nabla \frac{\delta \mathcal{F}[p]}{\delta p}\Big), \tag{7.2}$$

where $\mathcal{F}[p]$ is a free-energy functional and $\delta\mathcal{F}/\delta p$ its first variation. If $\mathcal{F}[p] = \int p \ln p + \int Vp$, i.e., the relative entropy with respect to the Gibbs measure $\propto e^{-V}$, then $\delta\mathcal{F}/\delta p = \ln p + 1 + V$ and Equation (7.2) becomes $\partial_t p = \nabla\cdot(p\nabla V) + \Delta p$, which is the Fokker–Planck equation (4.2) with drift $\mu = -\nabla V$ and $\sigma^2 = 2$; its steady state is $p \propto e^{-V}$ [39]. In one dimension this is Equation (4.3) with $\sigma^2 = 2$. If $\mathcal{F}[p]$ also includes a pairwise interaction energy $\frac{1}{2}\iint W(x-y)p(x)p(y)$, Equation (7.2) becomes the nonlinear McKean–Vlasov equation, whose steady states are self-organized spatial distributions used in swarming and consensus dynamics.

Table 5 summarizes distributions derived from spatial and temporal PDEs.

TABLE 5. Distributions derived from partial differential equations. "Family" indicates a time-indexed family $\{p(\cdot,t)\}_{t>0}$; "steady state" indicates a single limiting law.

PDE	Equation / condition	Resulting distribution	Application domain
Heat equation (7.1)	$\partial_t p = \Delta p$, $p(x,0) = \delta(x)$	Family $\mathcal{N}(0, 2tI_d)$	Spatial statistics, diffusion, noising in generative models
Schrödinger eq.	$\lvert\psi_n\rvert^2$, harmonic potential	Steady states $\propto H_n^2 e^{-\omega x^2}$ (nodal)	Quantum mechanics, nodal priors
Wasserstein flow (Equation (7.2))	$\mathcal{F} = \int p\ln p + \int Vp$	Steady state $\propto e^{-V}$ (Gibbs)	Sampling, Langevin MCMC
McKean–Vlasov	$\mathcal{F}$ with interaction W	Self-consistent steady states	Swarming, mean-field games
Reaction–diffusion	Turing-unstable steady state	Normalized periodic pattern density	Spatial ecology, materials

Uncorrected Proof

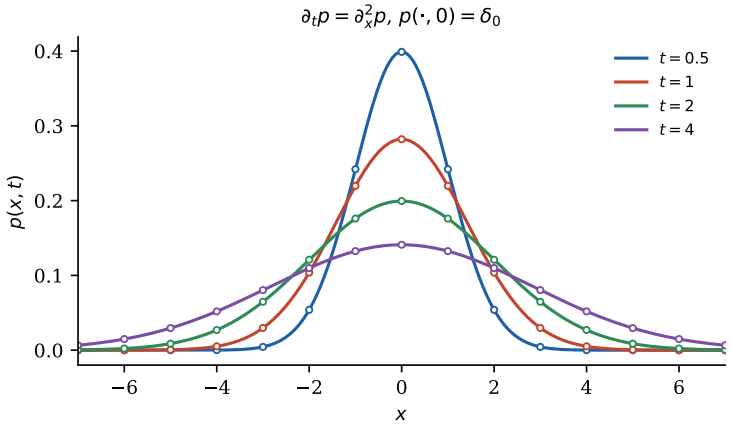


FIGURE 7. The time-indexed family generated by the heat equation (7.1) in one dimension, $p(x,t) = (4\pi t)^{-1/2}\exp(-x^2/4t)$, at $t \in \{0.5, 1, 2, 4\}$. Lines: closed form; circles: explicit finite-difference solution on $[-12, 12]$ with $\Delta x = 0.02$ and $\Delta t = 0.4\,\Delta x^2$, started from the exact profile at $t = 0.5$. Unlike the stationary densities of Figures 2, 3, and 4, the family never converges to a fixed shape: the variance $2t$ grows without bound.

7.3. **Numerical Illustration.** Figure 7 shows that the profile spreads and flattens monotonically while its total mass remains one, and that a standard numerical PDE solver reproduces the closed form.

Practical summary: Framework V. *Use when* the data are spatial or spatiotemporal (positions of particles, animals, or agents over time), or when a density on $\mathbb{R}^d$ is defined through an energy functional. *How:* write the model either as a PDE for $p(x,t)$ or as a free energy $\mathcal{F}$ and use Equation (7.2); decide whether the target is the time-indexed family or the steady state; fit by likelihood of snapshots (numerical PDE solution) or by simulation-based inference (Section 10.6). *Limitations:* normalization and likelihood evaluation are expensive in high dimension; steady states of nonlinear flows may be non-unique.

8. Framework VI: Matrix-Variate and Tensor Differential Equations

What is new here. We state the matrix analogue of the Pearson equation as an exact gradient equation for $\ln f(X)$ and give the correct gradient for each classical matrix-variate law (Table 6), including the matrix-variate t, which plays the role of Pearson Type VII.

8.1. **Matrix-Variate ODEs.** The Pearson and Fokker–Planck frameworks extend to random matrices using matrix calculus [30]. For a function of an $n \times p$ matrix X, let ∇_X denote the matrix of partial derivatives $\partial/\partial x_{ij}$. For symmetric positive-definite X we treat the entries as functionally independent and symmetrize the result; this convention does not affect the density recovered by integration. The matrix-variate analogue of Equation (3.1) is the gradient equation

$$\nabla_X \ln f(X) = -U^{-1}(X - M)V^{-1}, \tag{8.1}$$

with $n \times n$ and $p \times p$ positive-definite matrices U, V. Because the right-hand side is the gradient of $-\frac{1}{2}\operatorname{tr}\{U^{-1}(X - M)V^{-1}(X - M)^\top\}$, Equation (8.1) integrates to the matrix Normal density $f(X) \propto \exp\{-\frac{1}{2}\operatorname{tr}[U^{-1}(X - M)V^{-1}(X - M)^\top]\}$, i.e., $\operatorname{vec}(X) \sim \mathcal{N}(\operatorname{vec}(M), V \otimes U)$. Replacing the right-hand side by rational expressions in X (involving X^{-1} or $(I - X)^{-1}$) yields distributions on the cone of positive-definite matrices, exactly as the rational right-hand side of Equation (3.1) yields the univariate Pearson types. These laws are central in covariance estimation, MIMO communications, and multivariate time series (Table 6). The gradients listed in Table 6 follow from $\nabla_X \ln|X| = X^{-1}$, $\nabla_X \operatorname{tr}(AX) = A^\top$ and $\nabla_X \operatorname{tr}(AX^{-1}) = -X^{-1}A^\top X^{-1}$ for symmetric X.

Uncorrected Proof

TABLE 6. Matrix-variate distributions generated by gradient equations of the form in Equation (8.1); X is $p \times p$ symmetric positive definite except in the first and last rows, where X is $n \times p$. Parameterizations follow [30]; $\kappa > 0$ is a shape constant.

Structure	**Gradient** $\nabla_X \ln f(X)$	**Distribution**	**Use case**
Linear	$-U^{-1}(X-M)V^{-1}$	Matrix Normal	Multivariate spatiotemporal data
Inverse + constant	$\frac{n-p-1}{2}X^{-1} - \frac{1}{2}\Sigma^{-1}$	Wishart $W_p(n,\Sigma)$	Sample covariance matrices
Inverse + inverse-quadratic	$-\frac{\nu+p+1}{2}X^{-1} + \frac{1}{2}X^{-1}\Psi X^{-1}$	Inverse-Wishart $IW_p(\nu,\Psi)$	Bayesian covariance priors
Two inverse poles	$\left(a - \frac{p+1}{2}\right)X^{-1} - \left(b - \frac{p+1}{2}\right)(I-X)^{-1}$	Matrix-variate Beta, $0 < X < I$	Canonical correlations, MANOVA
Rational (Type VII analogue)	$-2\kappa\left[\Sigma + (X-M)\Omega^{-1}(X-M)^\top\right]^{-1}(X-M)\Omega^{-1}$	Matrix-variate t	Robust multivariate regression

The last row corresponds to $f(X) \propto |\Sigma + (X-M)\Omega^{-1}(X-M)^\top|^{-\kappa}$, the matrix-variate t law [30]; its gradient follows from $\mathrm{d}\ln|K| = \mathrm{tr}(K^{-1}\mathrm{d}K)$ with $K = \Sigma + (X-M)\Omega^{-1}(X-M)^\top$. It is the matrix counterpart of the Student-t (Pearson Type VII) law and, like it, has polynomially decaying tails.

8.2. **Tensor-Variate Extensions.** The matrix construction extends to tensor-valued data $\mathcal{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ by replacing the precision matrices U^{-1}, V^{-1} with a Kronecker-structured (Tucker-type) precision operator:

$$\nabla_{\mathrm{vec}(\mathcal{T})} \ln f(\mathcal{T}) = -\left(\Sigma_d^{-1} \otimes \cdots \otimes \Sigma_2^{-1} \otimes \Sigma_1^{-1}\right)\mathrm{vec}(\mathcal{T} - \mathcal{M}), \tag{8.2}$$

where Σ_k is a mode-k covariance factor and $\mathrm{vec}(\cdot)$ denotes column-major vectorization (the order of the Kronecker factors matches this convention). Integrating the linear gradient field in Equation (8.2) as for Equation (8.1) yields the **tensor Normal distribution**, whose Kronecker-structured covariance reduces the number of free covariance parameters from $O\left((\prod_k n_k)^2\right)$ in an unstructured Gaussian to $O\left(\sum_k n_k^2\right)$, the key practical motivation for this framework in multiway data. Inverse- and Beta-type analogues follow by the same substitutions used in Table 6, applied mode-wise.

Practical summary: Framework VI. *Use when* each observation is a matrix or a tensor (covariance matrices, images, multi-channel time series, spatio-temporal grids). *How:* choose the gradient structure in Table 6 from the support (all matrices, positive-definite cone, or $0 < X < I$) and the required tails; exploit Kronecker structure in Equation (8.2) to keep the number of parameters manageable; estimate by (flip-flop) maximum likelihood. *Limitations:* Kronecker separability is an assumption that should be tested; normalizing constants involve multivariate Gamma functions.

9. Framework VII: Information Geometry and Statistical Manifolds

What is new here. We make precise in what sense differential equations on a statistical manifold *generate* distributions: the geodesic ODE produces one-parameter families (paths) of laws joining two given laws, and the maximum-entropy principle produces exponential families whose log-densities satisfy first-order ODEs of Pearson type. Both are illustrated by a fully worked example for the Normal family (Equations (9.3) and (9.4), Figure 8).

Motivation. Frameworks I–VI generate a density as a function of the *data variable* x. Information geometry works one level higher, on the *parameter space* Θ of a family $\{p(\cdot \mid \theta)\}$, which it treats as a Riemannian manifold. This viewpoint answers practical questions that the other frameworks do not: how far apart two fitted models are, how

Uncorrected Proof

to interpolate or average between distributions (e.g., to build a path of intermediate laws for annealed importance sampling, tempering, or model morphing), and how to take parameter-invariant optimization steps (natural gradient).

9.1. **The Fisher–Rao Metric and the Geodesic Equation.** Information geometry views the space of probability distributions as a differentiable manifold [1, 2]. The local metric structure is given by the Fisher information matrix (FIM)

$$g_{ij}(\theta) = \mathbb{E}_\theta\Big[\frac{\partial \ln p(X \mid \theta)}{\partial \theta_i}\frac{\partial \ln p(X \mid \theta)}{\partial \theta_j}\Big]. \tag{9.1}$$

Given two members $p(\cdot \mid \theta_0)$ and $p(\cdot \mid \theta_1)$ of a family, the shortest path between them is a geodesic, i.e., a solution $\theta(t)$ of the second-order ODE

$$\frac{\mathrm{d}^2\theta^k}{\mathrm{d}t^2} + \Gamma^k_{ij}(\theta)\frac{\mathrm{d}\theta^i}{\mathrm{d}t}\frac{\mathrm{d}\theta^j}{\mathrm{d}t} = 0, \qquad \theta(0) = \theta_0,\ \theta(1) = \theta_1, \tag{9.2}$$

where $\Gamma^k_{ij} = \frac{1}{2}g^{kl}(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij})$ are the Christoffel symbols of the Fisher–Rao metric in Equation (9.1) (summation over repeated indices). The generated object is the one-parameter family of distributions $\{p(\cdot \mid \theta(t))\}_{t\in[0,1]}$, and its length is the Fisher–Rao distance.

Worked example: the Normal family. For $p(x \mid \mu, \sigma) = \mathcal{N}(\mu, \sigma^2)$, Equation (9.1) gives $g = \mathrm{diag}(1/\sigma^2,\, 2/\sigma^2)$, i.e., $\mathrm{d}s^2 = (\mathrm{d}\mu^2 + 2\,\mathrm{d}\sigma^2)/\sigma^2$. The nonzero Christoffel symbols are $\Gamma^\mu_{\mu\sigma} = \Gamma^\mu_{\sigma\mu} = -1/\sigma$, $\Gamma^\sigma_{\mu\mu} = 1/(2\sigma)$ and $\Gamma^\sigma_{\sigma\sigma} = -1/\sigma$, so Equation (9.2) becomes

$$\ddot\mu - \frac{2}{\sigma}\,\dot\mu\,\dot\sigma = 0, \qquad \ddot\sigma + \frac{1}{2\sigma}\,\dot\mu^2 - \frac{1}{\sigma}\,\dot\sigma^2 = 0. \tag{9.3}$$

In the coordinates $(\mu/\sqrt{2}, \sigma)$ the metric is $2\times$ the hyperbolic metric of the Poincaré half-plane, so the solutions of Equation (9.3) are vertical lines (μ constant) and semicircles centred on the axis $\sigma = 0$. The corresponding distance has the closed form [17]

$$d_{\mathrm{FR}}\big(\mathcal{N}(\mu_1,\sigma_1^2), \mathcal{N}(\mu_2,\sigma_2^2)\big) = \sqrt{2}\ \mathrm{arccosh}\Big(1 + \frac{(\mu_1-\mu_2)^2/2 + (\sigma_1-\sigma_2)^2}{2\sigma_1\sigma_2}\Big); \tag{9.4}$$

for $\mu_1 = \mu_2$ it reduces to $\sqrt{2}\,|\ln(\sigma_2/\sigma_1)|$. We checked numerically that solutions of Equation (9.3) lie on such semicircles (to 10^{-12}) and that their arc length agrees with Equation (9.4) (supplementary code). Along the Fisher–Rao geodesic joining two Normal laws with different means, the intermediate laws have a *larger* standard deviation than either endpoint (Figure 8(a)).

9.2. **Dual Connections, Exponential Families, and Maximum Entropy.** Amari introduced a one-parameter family of affine connections $\nabla^{(\alpha)}$; the Fisher–Rao (Levi-Civita) connection is $\alpha = 0$, while the exponential (e, $\alpha = +1$) and mixture (m, $\alpha = -1$) connections are dually flat [2]. Their geodesics are available in closed form and give two further ways to interpolate between laws p_0 and p_1:

$$e\text{-geodesic: } p_t \propto p_0^{1-t}\,p_1^{t}, \qquad m\text{-geodesic: } p_t = (1-t)p_0 + t p_1.$$

For an exponential family $p(x \mid \theta) = \exp\{\theta^\top T(x) - A(\theta)\}h(x)$, the e-geodesics are straight lines in the natural parameter θ, and the m-geodesics are straight lines in the expectation parameter $\eta = \mathbb{E}_\theta T(X) = \nabla A(\theta)$. The m-projection of an arbitrary law q onto the family, i.e., the minimizer of the Kullback–Leibler divergence $\mathrm{KL}(q \,\|\, p_\theta)$, is characterized by the moment-matching equation

$$\mathbb{E}_{\theta^*}[T(X)] = \nabla A(\theta^*) = \mathbb{E}_q[T(X)], \tag{9.5}$$

which, for q the empirical distribution, is the maximum likelihood equation.

The same exponential families arise from the *maximum entropy principle*: among all densities satisfying $\mathbb{E}T_k(X) = \tau_k$, $k = 1, \dots, K$, the entropy is maximized by $p(x) \propto h(x)\exp\{\sum_k \theta_k T_k(x)\}$, with θ determined by Equation (9.5). Differentiating shows that the maximum-entropy density solves the first-order ODE

$$\frac{\mathrm{d}}{\mathrm{d}x}\ln p(x) = \frac{h'(x)}{h(x)} + \sum_{k=1}^{K}\theta_k\,T_k'(x), \tag{9.6}$$

Uncorrected Proof

which is of the form in Equation (2.1). With $h \equiv 1$, the constraints $T = (x, x^2)$ give the Normal law, $T = (x, \ln x)$ the Gamma law (Pearson III), $T = (\ln x, \ln(1-x))$ the Beta law (Type I), and $T = (\ln x, \ln(1+x))$ the Beta-prime law (Type VI); in each case Equation (9.6) is the corresponding Pearson equation (3.1). Framework VII thus links the choice of sufficient statistics directly to the ODE of Framework I.

9.3. **Connections to Neural Network Parameter Spaces.** The geodesic construction of Equation (9.2) has a direct counterpart in modern machine learning: the Fisher information matrix in Equation (9.1) of a parametric model $p(x \mid \theta)$, with θ the weights of a neural network, defines the natural-gradient direction $\tilde{\nabla}_\theta \mathcal{L} = g(\theta)^{-1} \nabla_\theta \mathcal{L}$ used in second-order optimization; this is precisely the direction of steepest descent with respect to the Fisher–Rao metric rather than the Euclidean metric on parameter space [1]. Because the FIM of a deep network is typically both extremely high-dimensional and highly rank-deficient, practical algorithms use structured approximations to $g(\theta)^{-1}$ rather than solving Equation (9.2) exactly.

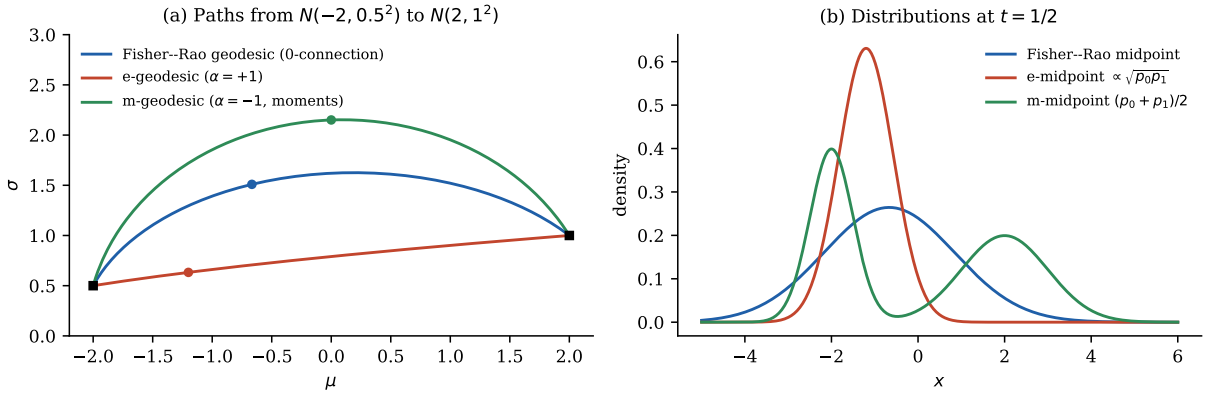


FIGURE 8. Three ways to interpolate between $p_0 = \mathcal{N}(-2, 0.5^2)$ and $p_1 = \mathcal{N}(2, 1^2)$. (a) Paths in the (μ, σ) half-plane: the Fisher–Rao geodesic (solution of Equation (9.3), a semicircle in $(\mu/\sqrt{2}, \sigma)$), the e-geodesic (straight line in natural parameters $(\mu/\sigma^2, -1/(2\sigma^2))$), and the mean and standard deviation of the m-geodesic (mixture); dots mark $t = 1/2$. (b) The three laws at $t = 1/2$: the Fisher–Rao and e-midpoints are Normal, whereas the m-midpoint is a bimodal mixture. The Fisher–Rao distance in Equation (9.4) between p_0 and p_1 is 4.12.

9.4. **Numerical Illustration.** Figure 8 shows that the choice of connection matters: the three interpolations have different shapes at the same parameter value $t = 1/2$. The m-path preserves mixture structure (useful for averaging predictive distributions), the e-path stays within the Normal family and concentrates mass between the modes (the geometric mean used in annealed importance sampling), and the Fisher–Rao path is the unique parameterization-invariant shortest path.

Practical summary: Framework VII. *Use when* one needs to compare, interpolate, or average fitted distributions, to derive a model from moment constraints (maximum entropy), or to optimize a model's parameters in a reparameterization-invariant way (natural gradient). *How:* compute the FIM in Equation (9.1); solve Equation (9.2) (closed form for Normal and other two-parameter families, numerically otherwise), or use the explicit e- and m-geodesics; choose sufficient statistics from the required support and tails and solve Equation (9.5). *Limitations:* Fisher–Rao geodesics are rarely available in closed form beyond low-dimensional families, and the FIM of large models must be approximated.

Uncorrected

10. Computational frameworks for implicit densities

A major hurdle in DE-based distributions is that the PDF $p(x;\theta)$ is often defined *implicitly* as the solution to an initial or boundary value problem. Unlike closed-form families (e.g., Gaussian, Gamma), evaluating $p(x;\theta)$ requires numerical integration, and computing gradients $\nabla_\theta \ln p(x;\theta)$ for Maximum Likelihood Estimation (MLE) or Bayesian inference introduces significant computational overhead. In this section, we delineate the modern computational toolkit required to handle these implicit densities.

10.1. Numerical ODE Solvers and Stiffness Analysis.

Every closed-form density discussed in Sections 3, 4, and 5 represents a rare case where the underlying ODE admits an analytic solution. In general, we must solve:

$$\frac{\mathrm{d}z}{\mathrm{d}x} = G(z, x; \theta), \quad z(x_0) = z_0, \tag{10.1}$$

where $z(x) = \ln p(x;\theta)$ or $z(x) = p(x;\theta)$. Even if an analytic form for $z(x)$ exists, the normalization constant $Z(\theta) = \int_{\mathcal{X}} \exp(z(x))\mathrm{d}x$ typically requires numerical quadrature.

Stiffness in Probability Dynamics. Standard explicit Runge–Kutta methods (e.g., RK45) [31] are efficient for non-stiff systems. However, DE-based densities often exhibit *stiffness*, characterized by widely separated eigenvalues of the Jacobian $J = \partial G/\partial z$. This arises routinely in:

- **Weibull Hazard Models:** For shape parameter $k \gg 1$, the hazard rate $h(x) \propto x^{k-1}$ changes rapidly near the mode, requiring impractically small step sizes for explicit stability.
- **Fractional Relaxation:** Discretizations of fractional ODEs in Equation (6.1) near the weak singularity at $x = 0$ introduce high-frequency components.
- **Multi-scale Fokker–Planck:** When the drift term dominates diffusion ($|\mu(x)| \gg \sigma^2(x)$ on the length scale of interest), the system becomes convection-dominated, leading to numerical oscillations unless upwind schemes or implicit solvers are used.

For such cases, implicit methods (e.g., Radau IIA, BDF) or exponential integrators are required [32].

10.2. Adjoint Sensitivity Methods for Gradient Computation.

In gradient-based optimization, we require $\nabla_\theta \mathcal{L}(\theta)$ where $\mathcal{L}(\theta) = \sum_{i=1}^{N} \ln p(x_i;\theta)$. Naïve backpropagation through the ODE solver steps has memory cost $O(N_{steps})$, which is prohibitive for high-accuracy solutions.

The Continuous Adjoint Method. Following [13], we treat the ODE solver as a continuous layer. Let $z(t)$ be the state satisfying $\dot{z} = g(z, t, \theta)$. We define the adjoint variable $a(t) = \partial\mathcal{L}/\partial z(t)$. The evolution of the adjoint is governed by the backward ODE:

$$\frac{\mathrm{d}a(t)}{\mathrm{d}t} = -a(t)^\top \frac{\partial g}{\partial z}(z(t), t, \theta), \quad a(T) = \frac{\partial \mathcal{L}}{\partial z(T)}. \tag{10.2}$$

The gradient with respect to parameters is then computed via:

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\theta} = -\int_T^0 a(t)^\top \frac{\partial g}{\partial \theta}(z(t), t, \theta)\,\mathrm{d}t, \tag{10.3}$$

where the integral is taken backward from $t = T$ to 0 (or, equivalently, as the solution of a differential equation from T to 0). This approach reduces memory complexity to $O(1)$ with respect to the number of integration steps; the same adjoint construction is standard for differential-algebraic systems in scientific computing [12].

Uncorrected Proof

Algorithm 1 Continuous adjoint sensitivity for MLE (after [13]).

Require: Data $\{x_i\}$, ODE dynamics $g(z,t,\theta)$, initial state z_0
Ensure: Gradient $\nabla_\theta \mathcal{L}$

1: **Forward Pass:** Solve $\dot{z} = g(z,t,\theta)$ from $t=0$ to T to obtain $z(T)$.
2: Compute loss $\mathcal{L}(z(T))$ and terminal adjoint $a(T) = \nabla_{z(T)}\mathcal{L}$.
3: **Backward Pass:** Define augmented state $\mathbf{s} = [z, a, \theta_{\text{grad}}]$ with initial conditions $\mathbf{s}(T) = [z(T), a(T), 0]$.
4: Solve the augmented ODE backward from $t=T$ to 0:

$$\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}t} = \begin{bmatrix} g(z,t,\theta) \\ -a^\top \dfrac{\partial g}{\partial z}(z,t,\theta) \\ -a^\top \dfrac{\partial g}{\partial \theta}(z,t,\theta) \end{bmatrix}.$$

5: The gradient is given by $\nabla_\theta \mathcal{L} = \theta_{\text{grad}}(0)$.

10.3. **Physics-Informed Neural Networks (PINNs).** For high-dimensional or fractional DEs where traditional grid-based solvers suffer from the curse of dimensionality, Physics-Informed Neural Networks (PINNs) [40, 57] offer a mesh-free alternative. We approximate the log-density $\ln p(x;\theta)$ using a neural network $u_\phi(x)$, ensuring positivity via $p_\phi(x) = \exp(u_\phi(x))$.

The network is trained by minimizing a composite loss function:

$$\mathcal{L}_{PINN}(\phi) = \lambda_{ODE}\mathcal{L}_{res} + \lambda_{BC}\mathcal{L}_{bc} + \lambda_{norm}\mathcal{L}_{norm} + \lambda_{data}\mathcal{L}_{data}, \tag{10.4}$$

where:

$$\mathcal{L}_{res} = \frac{1}{N_r}\sum_{j=1}^{N_r}\left\| \frac{\mathrm{d}u_\phi}{\mathrm{d}x}(x_j) - G(u_\phi(x_j), x_j;\theta)\right\|^2,$$

$$\mathcal{L}_{norm} = \left(\int_{\mathcal{X}} \exp(u_\phi(x))\mathrm{d}x - 1\right)^2,$$

$$\mathcal{L}_{bc} = \frac{1}{N_b}\sum_{j=1}^{N_b}\left(u_\phi(x_j^b) - u_j^b\right)^2,$$

$$\mathcal{L}_{data} = -\frac{1}{N}\sum_{i=1}^{N} u_\phi(x_i).$$

Here $\{x_j\}$ are collocation points, $\{(x_j^b, u_j^b)\}$ prescribed boundary or initial values of the log-density, and $\{x_i\}_{i=1}^N$ the observed data. *Sign and role of* $\mathcal{L}_{data}$. When $u_\phi = \ln p_\phi$ is a normalized log-density and the observations are unweighted, $\mathcal{L}_{data} = -\frac{1}{N}\sum_i \ln p_\phi(x_i)$ is the negative average log-likelihood; minimizing it is maximum likelihood estimation, and it estimates the cross-entropy $-\mathbb{E}_{\text{data}} \ln p_\phi$. The minus sign is therefore essential: $\mathcal{L}_{data}$ decreases as the model assigns more probability to the observed points. Because u_ϕ alone could be increased everywhere to lower $\mathcal{L}_{data}$, the term is meaningful only together with $\mathcal{L}_{norm}$ (or an exact normalization $u_\phi - \ln\int e^{u_\phi}$), which prevents the total mass from growing. For weighted data, the average is replaced by $-\sum_i w_i \ln p_\phi(x_i)/\sum_i w_i$. The integral in $\mathcal{L}_{norm}$ is approximated via Monte Carlo sampling or sparse quadrature. This formulation allows simultaneous estimation of the density shape and the parameters θ.

10.4. **Neural ODEs and Continuous Normalizing Flows.** Normalizing flows [51, 58] construct complex densities by transforming a simple base distribution $p_0(z_0)$ via an invertible map $z_T = f(z_0)$. Continuous Normalizing Flows (CNFs) parameterize this map using a Neural ODE [13]:

$$\frac{\mathrm{d}z(t)}{\mathrm{d}t} = f_\theta(z(t),t), \quad z(0) \sim p_0. \tag{10.5}$$

Uncorrected Proof

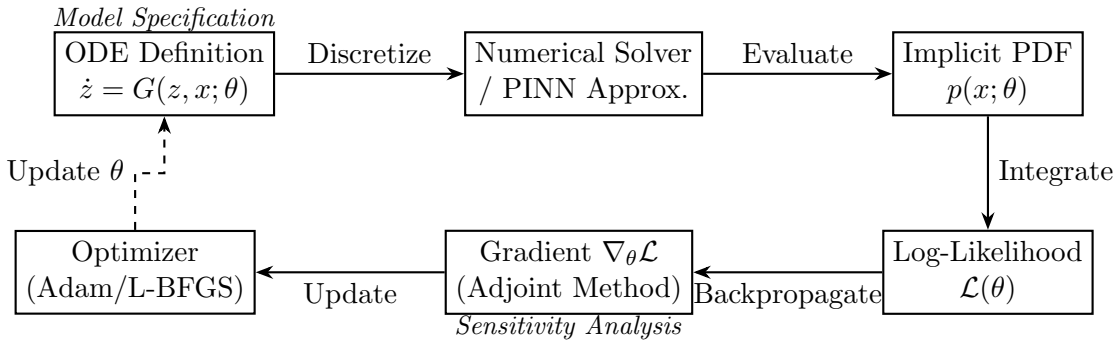


FIGURE 9. Computational workflow for inference in DE-based distributions. The ODE defines the density implicitly, which is evaluated via solvers or PINNs. Gradients are computed efficiently using adjoint methods.

The resulting density at time T is given by the change-of-variables formula:

$$\ln p_T(z_T) = \ln p_0(z_0) - \int_0^T \operatorname{tr}\Big(\frac{\partial f_\theta}{\partial z}(z(t), t)\Big)\mathrm{d}t. \tag{10.6}$$

Computing the trace of the Jacobian $\frac{\partial f_\theta}{\partial z}$ is expensive in high dimensions ($O(d^2)$). We employ Hutchinson's estimator [36] to approximate the trace with $O(d)$ cost:

$$\operatorname{tr}\Big(\frac{\partial f}{\partial z}\Big) = \mathbb{E}_{\epsilon\sim\mathcal{N}(0,I)}\big[\epsilon^\top \frac{\partial f}{\partial z}\epsilon\big],$$

which can be computed via automatic differentiation without forming the full Jacobian matrix.

Symmetry Constraints. If the target distribution is invariant under a symmetry group $\mathcal{G}$, we can constrain f_θ to be equivariant [22]. This reduces sample complexity and ensures the learned density respects physical laws exactly, analogous to choosing a Pearson type based on support constraints.

10.5. **Convex-Analytic Perspectives and Normalization.** For exponential family densities, the log-partition function $A(\theta)$ is convex [10, 68]. This implies that the negative log-likelihood is convex, guaranteeing a unique global optimum for MLE.

However, for general DE-based densities (e.g., Pearson types IV, VI), the mapping $\theta \mapsto p(x; \theta)$ is non-linear and non-convex. The normalization integral $Z(\theta)$ may not even be finite for all θ. We must therefore impose *integrability constraints* during optimization:

$$\Theta_{valid} = \Big\{\theta \in \mathbb{R}^d : \int_{\mathcal{X}} \exp\Big(\int_{x_0}^{x} G(y, \theta)\mathrm{d}y\Big)\mathrm{d}x < \infty\Big\}. \tag{10.7}$$

Optimization routines must project θ onto Θ_{valid} or use barrier methods to prevent divergence into regions where the density is non-normalizable.

10.6. **Likelihood-Free and Simulation-Based Inference.** For complex PDE-based models (e.g., McKean–Vlasov, Reaction-Diffusion), evaluating $p(x; \theta)$ is often intractable. However, simulating trajectories $X_t \sim P_\theta$ is usually cheap. In this regime, we employ Simulation-Based Inference (SBI) or Approximate Bayesian Computation (ABC) [18].

Instead of maximizing likelihood, we minimize a discrepancy metric between observed data $\mathcal{D}_{obs}$ and simulated data $\mathcal{D}_{sim}(\theta)$:

$$\mathcal{L}_{SBI}(\theta) = d\big(S(\mathcal{D}_{obs}), S(\mathcal{D}_{sim}(\theta))\big), \tag{10.8}$$

where $S(\cdot)$ are summary statistics and $d(\cdot, \cdot)$ is a distance metric (e.g., Wasserstein distance). Recent approaches avoid hand-crafted summaries altogether: a classifier trained to distinguish data simulated at θ from reference data estimates the likelihood ratio, which can then be used for posterior inference or MCMC [33, 64].

Uncorrected Proof

10.7. **Applications in Computational Biology.** The machinery described above has found important applications in the life sciences. In stochastic chemical kinetics, for instance in gene regulatory networks, the distribution of molecule counts evolves according to the Chemical Master Equation (CME), which can be simulated exactly [27] and, for large copy numbers, approximated by the chemical Langevin equation and its Fokker–Planck equation [28]; the rate constants θ then enter the drift and diffusion of Equation (4.2) and can be estimated from snapshot data with the adjoint and PINN methods above. PINNs that enforce the Fokker–Planck equation have been used to recover the drift and diffusion of an SDE from discrete particle observations [14], and normalizing flows have been trained to sample the equilibrium (Boltzmann) distributions $p \propto e^{-V/k_B T}$ of molecular systems, which are the stationary laws of overdamped Langevin dynamics [48].

11. Analysis and selection guidelines

Table 7 summarizes the key properties of the seven frameworks and offers a quick reference for choosing among them based on tail behavior, support, interpretability, and computational cost. The table highlights, for instance, that the Pearson system combines closed-form simplicity with the full range of tail behaviors (bounded, exponential, and power-law tails of index a_1/b_2 in Types IV–VII, Corollary 3.2), whereas fractional, delay, and integro-differential equations yield infinite-mean power laws, multimodal, and phase-type densities at the price of special-function or matrix-exponential computations. Frameworks II, V, and VII stand out for their ability to handle high-dimensional or spatiotemporal settings, making them particularly suitable for modern machine-learning and generative-modeling tasks.

Decision table. Table 8 turns Table 7 into a concise decision procedure: read the row that matches the data, and start from the suggested model.

To design a new DE-generated distribution, we propose the workflow illustrated in Figure 10. The procedure begins by choosing a framework (Table 8) and specifying the equation (polynomial order, drift and diffusion, fractional order, delay, or kernel). The equation is then solved, analytically or with a solver matched to its stiffness (Section 10.1), and the support and normalizing constant are determined; if $Z = \infty$, the tail parameters must be restricted or the support truncated. Finally the model is fitted and validated; if validation fails, the equation is revised. The following checklist makes the last two steps concrete.

Practitioner's checklist and evaluation criteria.

(1) *Well-posedness:* the right-hand side G is continuous on the proposed support (Theorem 2.1); for SDEs, the drift and diffusion satisfy Lipschitz and growth conditions.
(2) *Positivity and normalization:* $f > 0$ on the support and $Z < \infty$; state the admissible parameter region explicitly (e.g., $a_1/b_2 > 1$ for unbounded Pearson types, $\int^\infty h = \infty$ for hazards, $\kappa < \lambda$ in Proposition 6.2).
(3) *Stationary or transient:* for Fokker–Planck and PDE models, state whether the law is a stationary law or a fixed-time transition law (Table 2).
(4) *Numerical verification:* compare any closed form with a direct numerical solution of the equation and, where possible, with Monte Carlo samples, as in Figures 2, 3, 4, 5, 6, and 7.
(5) *Goodness of fit:* report the maximized log-likelihood with AIC/BIC for comparison with competing families, Kolmogorov–Smirnov or Anderson–Darling statistics, and probability integral transform (PIT) histograms or Q–Q plots.
(6) *Tail check:* compare the model's tail index (e.g., a_1/b_2 for Pearson, α for Mittag-Leffler and Lomax) with a nonparametric estimate from the data, such as the Hill estimator or the log–log slope of the empirical survival function [21].
(7) *Identifiability:* check that the Fisher information matrix is nonsingular at the estimate and that different equations do not give practically indistinguishable fits (Section 12).
(8) *Reproducibility:* report the solver, tolerances, quadrature rule, random seeds, and all parameter values.

This structured approach turns the principles discussed in the review into an actionable pipeline for practitioners.

Uncorrected Proof

TABLE 7. Comparative analysis of the seven DE-based distribution frameworks.

Framework	**Tail behavior**	**Support**	**Parameter meaning**	**Computation**	**Typical use**
I. Pearson	Bounded (I, II); exponential (Normal, III); power law of index a_1/b_2 (IV–VII)	Interval, half-line, or $\mathbb{R}$, set by the roots of Q	Moments, skewness, kurtosis	High (closed form, Table 1)	Moment matching, classical statistics
II. Fokker–Planck	Set by μ, σ; stable (power) tails via jumps	$\mathbb{R}$ or half-line	Drift and noise (physical if derived from a model)	Moderate (one integral); learned score in high d	Physics, finance, generative modeling
III. Hazard ODE	Light (Weibull $k > 1$, Gompertz) to power law (Lomax, via frailty)	$[x_0, \infty)$	Aging and failure mechanisms	High (explicit h, automatic normalization)	Reliability, survival analysis
IV. Fractional / delay / integro	Power law with infinite mean (Mittag-Leffler); matrix-exponential (PH)	Usually $(0, \infty)$	Memory, delay, hidden stages	Low to moderate (special functions, matrix exponentials)	Anomalous diffusion, retries, multistage systems
V. PDE / Wasserstein	Depends on t and on the energy $\mathcal{F}$	$\mathbb{R}^d$ or domains	Spatial forces, potentials	Low (PDE solvers)	Spatial statistics, swarms, ecology
VI. Matrix / tensor	Gaussian (matrix Normal) to polynomial (matrix t)	Matrices, PD cone, $0 < X < I$	Covariance structure	Moderate (Kronecker structure)	Multivariate and multiway data
VII. Information geometry	Inherited from the family	Parameter manifold	Information-theoretic	High for exponential families	Model comparison, interpolation, maximum entropy

Uncorrected Proof

12. Open Challenges and future directions

While DE-based distributions offer immense flexibility, several challenges remain at the frontier of statistical research:

- **The Curse of Normalization:** For high-dimensional PDEs (Section 7) or nonlinear Fokker–Planck equations (Section 4.3), calculating the partition function $Z = \int f(x)\mathrm{d}x$ is computationally intractable. Developing variational bounds or path-integral approximations for Z is a critical open problem.
- **Identifiability and Equifinality:** Different equations can yield identical or nearly identical PDFs. For example, the q-Gaussian steady state of the nonlinear Fokker–Planck equation (4.4) with $1 < q < 3$ is exactly a rescaled Student-t (Pearson Type VII) law with $(3 - q)/(q - 1)$ degrees of freedom, which is also the stationary law of a linear Pearson diffusion (Table 2); the stationary marginal alone cannot distinguish the two mechanisms. Rigorous information-theoretic criteria must be developed to select the "true" generative mechanism from marginal observations alone.

TABLE 8. Decision table for choosing a framework and a first model.

If the data are ...	Framework	First model to try	Estimation
Univariate, unimodal, fit by moments; tails bounded, exponential, or power law	I	Pearson type from Δ and the roots (Table 1)	Method of moments, then ML
Long-run states of a system with a restoring force and noise	II	OU, CIR, or Pearson diffusion (Table 2)	ML of stationary or transition density
Lifetimes or times-to-event, possibly censored	III	Weibull / log-logistic / Siler hazard; add frailty for heterogeneity (Table 3)	Censored-data ML
Waiting times with infinite mean or power-law survival	IV	Mittag-Leffler (Equation (6.2))	Log–log slope, then ML
Periodic multiple modes due to delays or retries	IV	DDE density (Proposition 6.1)	ML
Several hidden stages; need a flexible positive law	IV	Phase type (Proposition 6.2)	EM [7]
Spatial or spatiotemporal snapshots	V	Heat kernel, Gibbs law e^{-V}, McKean–Vlasov	PDE likelihood or SBI
Matrix- or tensor-valued observations	VI	Matrix Normal, Wishart, matrix t (Table 6)	Flip-flop ML
Comparing, averaging, or interpolating fitted laws	VII	Fisher–Rao, e-, or m-geodesic	Closed form or ODE solver
High-dimensional samples (images, molecules), no closed form required	II	Score-based diffusion model (Section 4.6)	Denoising score matching

Uncorrected Proof

- **Multivariate Coupling:** Extending 1D ODE frameworks to $d > 1$ dimensions without losing computational tractability is difficult. Copula-based ODE approaches (Sections 3.3 and 5.3) and Kronecker-structured tensor equations (Section 8.2) are promising but still nascent.
- **Bridging Classical and Learned Dynamics:** Score-based generative models (Section 4.6) show that a learned, implicit score function can substitute for an analytically chosen drift in Equation (4.2), but it remains an open question how to recover *interpretable*, closed-form structure from a trained score network when the underlying process is suspected to have simple, low-dimensional dynamics.
- **Integration with Deep Learning:** Embedding DE-based PDFs as the final layer of a Variational Autoencoder (VAE) or Normalizing Flow (Section 10.4) requires differentiable ODE solvers that strictly preserve the non-negativity and normalization constraints of probability theory.
- **Numerical Stability in Fractional and Jump Models:** Fractional/DDE PDFs (Section 6) and jump-diffusion laws with α-stable jumps (Section 4.4) require robust quadrature and careful handling of branch cuts, weak singularities, or highly oscillatory modes during the inference phase.

13. CONCLUSION

The generation of probability distributions from differential equations represents a profound synthesis of dynamical systems theory, statistical inference, and information geometry. By shifting the focus from algebraic formulas to differential generators, statisticians and physicists can build models that are not only mathematically rigorous but also physically and biologically interpretable.

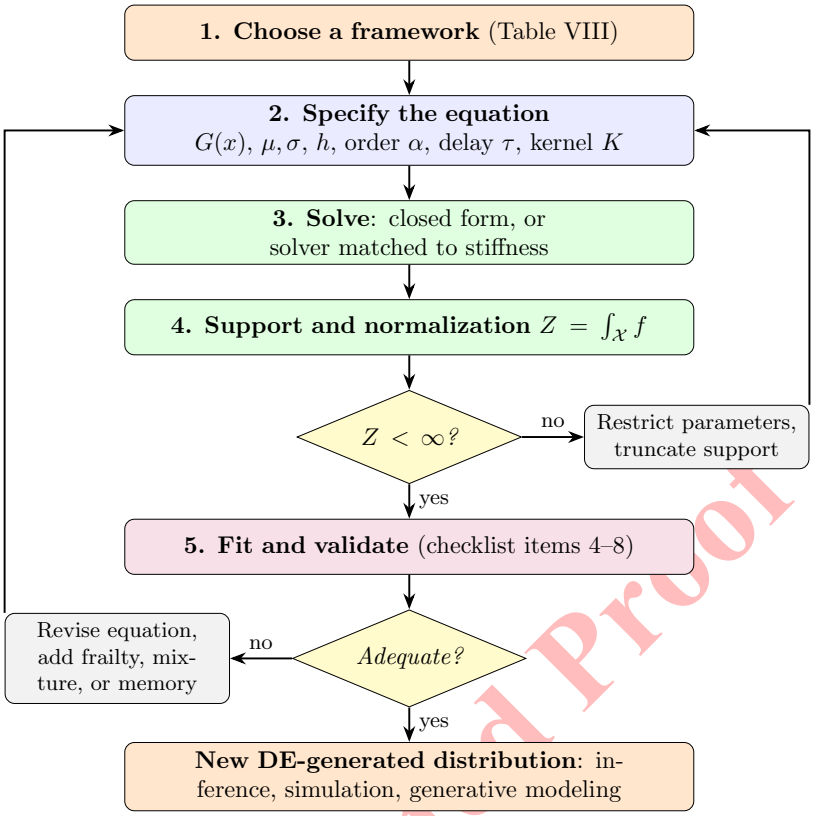


FIGURE 10. Simplified workflow for generating a new DE-based probability distribution. Orange: start and result; blue: specification; green: solution and normalization; purple: validation. Diamonds are decision points; both feedback loops return to the specification of the equation.

Uncorrected Proof

From the classical elegance of the Pearson system (Section 3), now placed on a fully worked and numerically verified closed-form footing with a single tail-index rule for all its types (Theorem 3.1 and Corollary 3.2), through frailty-generated heavy tails (Theorem 5.2), to the memory-rich dynamics of fractional, delay, and integro-differential equations with their explicit Mittag-Leffler, shifted-Erlang, and phase-type laws (Section 6), the spatial complexity of PDEs and Wasserstein flows (Section 7), the high-dimensional covariance and multiway modeling of matrix and tensor ODEs (Section 8), and the geometric foundations of statistical manifolds (Section 9), the DE framework provides a flexible and principled canvas for distribution design. Framework II in particular now spans the full arc of this review's thesis in miniature: the same steady-state equation that yields the textbook Gaussian and scaled chi-squared laws of Table 2 also underlies, via its time-reversed and learned form, the score-based generative models that currently define the state of the art in high-dimensional generative modeling (Section 4.6). As computational tools such as adjoint sensitivity methods, Physics-Informed Neural Networks, and stiffness-aware numerical solvers (Section 10) continue to mature, the barrier to fitting these implicitly defined distributions will keep receding, extending the mechanistic, equation-driven approach to statistical modeling into ever higher dimensions and more complex domains.

DATA AND CODE AVAILABILITY

No new data were analyzed. The Python scripts `make_figures.py` (reproducing Figures 2, 3, 4, 5, 6, 7, and 8) and `verify_results.py` (numerical verification of Theorems 3.1 and 5.2, Corollary 3.2, and Propositions 6.1 and 6.2,

Equations (6.2) and (9.4), and of the ODE representations in Tables 2 and 3) are provided as supplementary material. They use only NumPy, SciPy, Matplotlib and mpmath, and all random seeds are fixed.

References

[1] S. Amari, *Information Geometry and Its Applications*, Springer, Tokyo, 2016.

[2] S. Amari and H. Nagaoka, *Methods of Information Geometry*, American Mathematical Society and Oxford University Press, Providence, 2000.

[3] L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed., Birkhäuser, Basel, 2008.

[4] B. D. O. Anderson, *Reverse-time diffusion equation models*, Stochastic Process. Appl., *12*(3) (1982), 313–326.

[5] D. Applebaum, *Lévy Processes and Stochastic Calculus*, 2nd ed., Cambridge University Press, Cambridge, 2009.

[6] V. I. Arnold, *Ordinary Differential Equations*, Springer, Berlin, 1992.

[7] S. Asmussen, O. Nerman, and M. Olsson, *Fitting phase-type distributions via the EM algorithm*, Scand. J. Stat., *23*(4) (1996), 419–441.

[8] O. E. Barndorff-Nielsen, *Information and Exponential Families in Statistical Theory*, Wiley, Chichester, 1978.

[9] M. Bladt and B. F. Nielsen, *Matrix-Exponential Distributions in Applied Probability*, Springer, New York, 2017.

[10] J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization: Theory and Examples*, 2nd ed., Springer, New York, 2006.

[11] L. D. Brown, *Fundamentals of Statistical Exponential Families, with Applications in Statistical Decision Theory*, Institute of Mathematical Statistics, Hayward, 1986.

[12] Y. Cao, S. Li, L. Petzold, and R. Serban, *Adjoint sensitivity analysis for differential-algebraic equations: The adjoint DAE system and its numerical solution*, SIAM J. Sci. Comput., *24*(3) (2003), 1076–1089.

[13] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, *Neural ordinary differential equations*, Adv. Neural Inf. Process. Syst., *31* (2018), 6571–6583.

[14] X. Chen, L. Yang, J. Duan, and G. E. Karniadakis, *Solving inverse stochastic problems from discrete particle observations using the Fokker–Planck equation and physics-informed neural networks*, SIAM J. Sci. Comput., *43*(3) (2021), B811–B830.

[15] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations*, McGraw-Hill, New York, 1955.

[16] P. Collet, S. Martínez, and J. San Martín, *Quasi-Stationary Distributions: Markov Chains, Diffusions and Dynamical Systems*, Springer, Berlin, 2013.

[17] S. I. R. Costa, S. A. Santos, and J. E. Strapasson, *Fisher information distance: A geometrical reading*, Discrete Appl. Math., *197* (2015), 59–69.

[18] K. Cranmer, J. Brehmer, and G. Louppe, *The frontier of simulation-based inference*, Proc. Natl. Acad. Sci. USA, *117*(48) (2020), 30055–30062.

[19] P. J. Davis and P. Rabinowitz, *Methods of Numerical Integration*, 2nd ed., Dover Publications, Mineola, 2007.

[20] K. Diethelm, *The Analysis of Fractional Differential Equations*, Springer, Berlin, 2010.

[21] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin, 1997.

[22] M. Finzi, S. Stanton, P. Izmailov, and A. G. Wilson, *Generalizing convolutional neural networks for equivariance to Lie groups on arbitrary continuous data*, Proc. Mach. Learn. Res., *119* (2020), 3165–3176.

[23] A. D. Fokker, *Die mittlere Energie rotierender elektrischer Dipole im Strahlungsfeld*, Ann. Phys., *348*(5) (1914), 810–820.

[24] J. L. Forman and M. Sørensen, *The Pearson diffusions: A class of statistically tractable diffusion processes*, Scand. J. Stat., *35*(3) (2008), 438–465.

[25] T. D. Frank, *Nonlinear Fokker–Planck Equations: Fundamentals and Applications*, Springer, Berlin, 2005.

[26] D. Fulger, E. Scalas, and G. Germano, *Monte Carlo simulation of uncoupled continuous-time random walks yielding a stochastic solution of the space-time fractional diffusion equation*, Phys. Rev. E, *77*(2) (2008), 021122.

[27] D. T. Gillespie, *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem., *81*(25) (1977), 2340–2361.

Uncorrected Proof

[28] D. T. Gillespie, *The chemical Langevin equation*, J. Chem. Phys., *113*(1) (2000), 297–306.
[29] R. Gorenflo, A. A. Kilbas, F. Mainardi, and S. V. Rogosin, *Mittag-Leffler Functions, Related Topics and Applications*, Springer, Berlin, 2014.
[30] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*, Chapman and Hall/CRC, Boca Raton, 1999.
[31] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations I: Nonstiff Problems*, 2nd ed., Springer, Berlin, 1993.
[32] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*, 2nd ed., Springer, Berlin, 1996.
[33] J. Hermans, V. Begy, and G. Louppe, *Likelihood-free MCMC with amortized approximate likelihood ratios*, Proc. Mach. Learn. Res., *119* (2020), 4239–4249.
[34] J. Ho, A. Jain, and P. Abbeel, *Denoising diffusion probabilistic models*, Adv. Neural Inf. Process. Syst., *33* (2020), 6840–6851.
[35] P. Hougaard, *Analysis of Multivariate Survival Data*, Springer, New York, 2000.
[36] M. F. Hutchinson, *A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines*, Commun. Stat. Simul. Comput., *18*(3) (1989), 1059–1076.
[37] A. Hyvärinen, *Estimation of non-normalized statistical models by score matching*, J. Mach. Learn. Res., *6*(24) (2005), 695–709.
[38] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions, Volume 1*, 2nd ed., Wiley, New York, 1994.
[39] R. Jordan, D. Kinderlehrer, and F. Otto, *The variational formulation of the Fokker–Planck equation*, SIAM J. Math. Anal., *29*(1) (1998), 1–17.
[40] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, *Physics-informed machine learning*, Nat. Rev. Phys., *3*(6) (2021), 422–440.
[41] T. Karras, M. Aittala, T. Aila, and S. Laine, *Elucidating the design space of diffusion-based generative models*, Adv. Neural Inf. Process. Syst., *35* (2022), 26565–26577.
[42] J. P. Klein and M. L. Moeschberger, *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd ed., Springer, New York, 2003.
[43] L. Kleinrock, *Queueing Systems, Volume 1: Theory*, Wiley-Interscience, New York, 1975.
[44] F. Mainardi, *Fractional Calculus and Waves in Linear Viscoelasticity*, Imperial College Press, London, 2010.
[45] F. Mainardi, Y. Luchko, and G. Pagnini, *The fundamental solution of the space-time fractional diffusion equation*, Fract. Calc. Appl. Anal., *4*(2) (2001), 153–192.
[46] R. Metzler and J. Klafter, *The random walk's guide to anomalous diffusion: A fractional dynamics approach*, Phys. Rep., *339*(1) (2000), 1–77.
[47] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*, Johns Hopkins University Press, Baltimore, 1981.
[48] F. Noé, S. Olsson, J. Köhler, and H. Wu, *Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning*, Science, *365*(6457) (2019), eaaw1147.
[49] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, 6th ed., Springer, Berlin, 2003.
[50] J. K. Ord, *Families of Frequency Distributions*, Griffin, London, 1972.
[51] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, *Normalizing flows for probabilistic modeling and inference*, J. Mach. Learn. Res., *22*(57) (2021), 1–64.
[52] K. Pearson, *Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material*, Philos. Trans. R. Soc. Lond. A, *186* (1895), 343–414.
[53] R. N. Pillai, *On Mittag-Leffler functions and related distributions*, Ann. Inst. Stat. Math., *42*(1) (1990), 157–161.
[54] M. Planck, *Über einen Satz der statistischen Dynamik und seine Erweiterung in der Quantentheorie*, Sitzungsber. Preuss. Akad. Wiss., *24* (1917), 324–341.
[55] A. R. Plastino and A. Plastino, *Non-extensive statistical mechanics and generalized Fokker–Planck equation*, Physica A, *222*(1–4) (1995), 347–354.
[56] I. Podlubny, *Fractional Differential Equations*, Academic Press, San Diego, 1999.

Uncorrected Proof

[57] M. Raissi, P. Perdikaris, and G. E. Karniadakis, *Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations*, J. Comput. Phys., *378* (2019), 686–707.
[58] D. J. Rezende and S. Mohamed, *Variational inference with normalizing flows*, Proc. Mach. Learn. Res., *37* (2015), 1530–1538.
[59] H. Risken, *The Fokker–Planck Equation: Methods of Solution and Applications*, 2nd ed., Springer, Berlin, 1989.
[60] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed., Springer, New York, 2004.
[61] W. Siler, *A competing-risk model for animal mortality*, Ecology, *60*(4) (1979), 750–757.
[62] H. Smith, *An Introduction to Delay Differential Equations with Applications to the Life Sciences*, Springer, New York, 2011.
[63] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, *Score-based generative modeling through stochastic differential equations*, Proc. Int. Conf. Learn. Represent. (ICLR), 2021.
[64] O. Thomas, R. Dutta, J. Corander, S. Kaski, and M. U. Gutmann, *Likelihood-free inference by ratio estimation*, Bayesian Anal., *17*(1) (2022), 1–31.
[65] C. Tsallis, *Possible generalization of Boltzmann–Gibbs statistics*, J. Stat. Phys., *52*(1–2) (1988), 479–487.
[66] C. Tsallis and D. J. Bukman, *Anomalous diffusion in the presence of external forces: Exact time-dependent solutions and their thermostatistical basis*, Phys. Rev. E, *54*(3) (1996), R2197–R2200.
[67] C. Villani, *Optimal Transport: Old and New*, Springer, Berlin, 2009.
[68] M. J. Wainwright and M. I. Jordan, *Graphical models, exponential families, and variational inference*, Found. Trends Mach. Learn., *1*(1–2) (2008), 1–305.
[69] W. Weibull, *A statistical distribution function of wide applicability*, J. Appl. Mech., *18*(3) (1951), 293–297.
[70] L. Yang, Z. Zhang, Y. Song, S. Hong, R. Xu, Y. Zhao, W. Zhang, B. Cui, and M. H. Yang, *Diffusion models: A comprehensive survey of methods and applications*, ACM Comput. Surv., *56*(4) (2023), 1–39.

Uncorrected Proof