Chilean Journal of Statistics
Volume 15, Number 2, December 2024, Pages 147–168
DOI: 10.32372/chjs.15-02-03

Statistical Inference and Probability
Research Article

Likelihood function: Definition, examples, and numerical experiments

Flávio B. Gonçalves[1,*] and Pedro Franklin[2]

[1]Department of Statistics, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[2]Institute of Mathematics and Statistics, Universidade Federal de Uberlândia, Uberlândia, Brazil

(Received: 01 December 2023 · Accepted in final form: 24 October 2024)

Abstract

This article introduces a measure-theoretic definition of the likelihood function via Radon-Nikodym (RN) derivatives and establishes a new likelihood proportionality theorem showing that likelihoods obtained from any pair of dominating measures differ only by a parameter-free factor. This result validates the RN-based definition in light of the likelihood principle, particularly in settings where a single canonical dominating measure does not exist or multiple choices are natural—such as certain infinite-dimensional or missing-data problems. The role of continuous versions of RN derivatives is highlighted, demonstrating how continuity both ensures well-behaved, often unique likelihoods and aligns with Fisher's original intuition. The prior predictive measure is also examined as an alternative dominating measure in Bayesian contexts, while exponential families are shown to retain their defining exponential structure under any dominating measure. Collectively, these findings unify and refine fundamental measure-theoretic questions about likelihood, offering a rigorous framework for likelihood-based inference across a variety of statistical models.

Keywords: Continuous densities · Dominating measure · Likelihood principle · Proportional likelihood · Radon-Nikodym derivative · Statistical models

Mathematics Subject Classification: Primary 62A01 · Secondary 62A99.

1. Introduction

The concept of likelihood, introduced by Fisher (1921), predates the Kolmogorov (1933) formal axiomatization of probability and the Radon-Nikodym (RN) Theorem (Nikodym, 1930). Likelihood, in statistical inference, represents a function that quantifies the plausibility of a parameter value given observed data. Specifically, for a probability model $(\mathcal{X}, \mathcal{F}, P_\theta)$, where $\mathcal{X}$ is the sample space, $\mathcal{F}$ the σ-algebra of events, and P_θ a family of probability measures indexed by a parameter $\theta \in \Theta$, the likelihood function is traditionally expressed as $L(\theta; x) = P_\theta(x)$. In measure-theoretic terms, this corresponds to defining the likelihood function as the derivative of P_θ with respect to a σ-finite dominating measure, formalized in Lindley (1953), with earlier implicit formulations appearing in Halmos and Savage (1949). Fisher's intuitive definition laid the groundwork for this formalization.

*Corresponding author. Email: fbgoncalves@ufmg.br (F. Gonçalves)

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)
© Chilean Statistical Society – Sociedad Chilena de Estadística
http://www.soche.cl/chjs

The existence of multiple possible dominating measures introduces a fundamental issue: whether the choice of dominating measure affects inferential conclusions. According to the likelihood principle (LP), likelihood functions derived from distinct dominating measures should be proportional, in θ, almost surely. While this idea is widely recognized, systematic proofs and explicit discussions remain limited, particularly in infinite-dimensional settings. As Reid (2013) noted, the likelihood function is often described as the RN derivative of a probability measure with respect to a dominating measure. In some cases, the dominating measure is chosen as P_{θ_0} for a fixed $\theta_0 \in \Theta$. However, for infinite-dimensional parameter or probability spaces, selecting a suitable dominating measure poses significant challenges. Further studies on the mathematical properties of likelihood can be found in works such as Barndorff-Nielsen et al. (1976), Fraser and Naderi (1996), Fraser et al. (1997), Fraser and Naderi (2007), San Martín and González (2010), and Ruggeri (2010).

Building on this foundation, the present article addresses the proportionality of likelihood functions derived from different dominating measures and establishes a likelihood proportionality theorem. This theorem demonstrates that any two dominating measures yield likelihood functions differing only by a factor independent of the parameter, thereby aligning with the LP. In this article, the continuity of RN derivatives is explored as well, as continuous versions ensure well-behaved likelihood functions and reinforce Fisher's original intuition.

Beyond the proportionality theorem, this article extends to Bayesian settings, where prior predictive measures are considered as dominating measures. The analysis also highlights why exponential families retain their canonical form regardless of the chosen dominating measure. While traditional approaches predominantly focus on finite-dimensional models, we broaden the scope to include infinite-dimensional and non-parametric frameworks, which are increasingly relevant in modern Bayesian analysis. Specific examples in this context include missing data problems, Poisson processes (PP), mixtures of discrete and continuous components, as well as partially infinite-dimensional models, such as diffusions.

The structure of this article is as follows. Section 2 introduces the likelihood proportionality theorem, along with auxiliary results, and examines the role of continuous RN derivatives. In Section 3, we discuss prior predictive measures as those which are dominating in Bayesian models and the invariance of exponential families. In Section 4, some examples are provided to illustrate the challenges posed by the choice of dominating measure. Section 5 presents numerical experiments with simulated and real data. Conclusions are stated in Section 6, while Appendices A and B include results and proofs of the main theorems.

2. The likelihood function and Radon-Nikodym derivatives

This section provides the key theoretical results of the article, namely the likelihood proportionality theorem and a result on how continuous versions of RN derivatives help ensure well-defined likelihood functions.

2.1. *The likelihood proportionality theorem*

Let $(\Omega, \mathcal{F})$ be a measurable space, $(\Omega, \mathcal{F}, \mu)$ a measure space and $M(\Omega, \mathcal{F})$ denote the collection of all measurable functions $f\colon \Omega \to \mathrm{R}$.

Definition 2.1 A statistical model is a family of probability measures $\mathcal{P}$ on $(\Omega, \mathcal{F})$, namely $\mathcal{P} = \{P_\theta\colon \theta \in \Theta\}$, where each P_θ is a probability measure and Θ is an arbitrary index set. In the special case where $\Theta \subset \mathrm{R}^d$ for some fixed $d \in \mathrm{N}$, we say that $\mathcal{P}$ is a parametric model, with θ called a parameter and Θ the parameter space. Otherwise, we call $\mathcal{P}$ non-parametric.

A statistical model is called identifiable if the mapping $\theta \mapsto P_\theta$ is injective. For more details, see Lehmann (1986, Sec. 1.3) and Shao (2003, Ch. 5). In a typical inference problem, we wish to estimate or characterize a measure $P_{\theta^*} \in \mathcal{P}$ (the true or data-generating distribution) from observations drawn according to P_{θ^*}. The likelihood function quantifies how plausible each P_θ is, given the observed data. In measure-theoretic terms, we adopt the following definition.

Definition 2.2 Let $\mathcal{P} = \{P_\theta \colon \theta \in \Theta\}$ be a statistical model on $(\Omega, \mathcal{F})$, and let ν be any σ-finite measure that dominates $\mathcal{P}$ (meaning $\mathcal{P} \ll \nu$). For an observed point $\omega \in \Omega$, define the likelihood function associated with ν by $l_\nu(\theta; \omega) = \mathrm{d}P_\theta/\mathrm{d}\nu(\omega)$, for $\theta \in \Theta$, where $\mathrm{d}P_\theta/\mathrm{d}\nu$ is any chosen version of the RN derivative of P_θ with respect to ν. Different versions differ only on a set of ν-measure zero, so they yield the same likelihood values ν-almost everywhere; see, for example, Shao (2003, Sec. 2.2) for a measure-theoretic treatment.

As discussed, Fisher's motivation for constructing the likelihood function via RN derivatives is quite natural. However, a justification of this definition comes from the LP (Berger and Wolpert, 1988), which states that two likelihood functions containing the same information about P_θ must differ only by a factor free of θ. The LP can be specialized in various ways: for instance, to different data points ω_1 and ω_2, yielding $l(\theta; \omega_1) \propto_\theta l(\theta; \omega_2)$, meaning that these expressions differ by a factor independent of θ. The LP can also be applied to different experiments, where distinct measurable functions $f \in M(\Omega, \mathcal{F})$ are observed.

In this work, however, we focus on the version of the LP that deals specifically with distinct dominating measures. Specifically, let ν_1 and ν_2 be two σ-finite measures dominating the same family $\{P_\theta \colon \theta \in \Theta\}$. Denote the corresponding likelihoods by $\ell_{\nu_1}(\theta; \omega)$ and $\ell_{\nu_2}(\theta; \omega)$, or briefly $l_1(\theta; \omega)$ and $l_2(\theta; \omega)$. Then, under the LP, if $l_1(\theta; \omega) = h(\omega) l_2(\theta; \omega)$, for some function h that does not depend on θ, l_1 and l_2 yield the same inferences about θ –for instance, they have the same maximum-likelihood (ML) estimator–. Thus, we say l_1 and l_2 are proportional in θ.

Definition 2.2 is therefore validated if any choice of a σ-finite dominating measure ν yields likelihoods that are almost surely proportional in θ. Next, we introduce the notion of a minimal dominating measure.

Definition 2.3 For a family of probability measures $\mathcal{P} = \{P_\theta \colon \theta \in \Theta\}$, suppose there exists at least one σ-finite measure ν such that $\mathcal{P} \ll \nu$. Define $\Upsilon = \{\nu'$ is a σ-finite measure: $\mathcal{P} \ll \nu'\})$. If there is a measure $\lambda \in \Upsilon$ such that $\lambda \ll \nu'$ for all $\nu' \in \Upsilon$, we call λ a minimal dominating measure for the family $\mathcal{P}$. (In general, such a λ need not be unique, but any two minimal dominating measures are equivalent to each other.)

Proposition 2.4 Let $\mathcal{P} = \{P_\theta \colon \theta \in \Theta\}$ be a family of probability measures on $(\Omega, \mathcal{F})$, and assume that there is at least one σ-finite measure ν with $\mathcal{P} \ll \nu$. Then, there exists a minimal dominating measure for $\mathcal{P}$.

We now state a key lemma from Halmos and Savage (1949); see also Jorgensen and Labouriau (2012, p. 53).

Lemma 2.5 Let $\mathcal{P} = \{P_\theta \colon \theta \in \Theta\}$ be a family of probability measures on $(\Omega, \mathcal{F})$, and let ν be a σ-finite measure such that $\mathcal{P} \ll \nu$. Then, there exists a probability measure Q, dominated by ν, such that $\mathcal{P} \ll Q$. Specifically, $Q = \sum_{i=1}^{+\infty} c_i P_{\theta_i}$, where the c_i are nonnegative constants with $\sum_{i=1}^{+\infty} c_i = 1$, and each $P_{\theta_i} \in \mathcal{P}$.

Lemma 2.5 provides a powerful result for families of probability measures dominated by a single σ-finite measure, showing that a single probability measure Q can be constructed so that its support simultaneously covers all P_θ. This countable mixture device is crucial to the proof of our main theorem below, because it ensures that P_{θ_i}-almost sure implies P_θ-almost sure for any $\theta \in \Theta$.

For any measurable function f on $M(\Omega, \mathcal{F})$, we denote by $[f]_\mu$ its equivalence class under μ-almost everywhere equality, that is, the set of all g with $g = f$ μ-almost sure.

Theorem 2.6 Let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a family of probability measures on $(\Omega, \mathcal{F})$, and let ν_1 and ν_2 be two σ-finite measures such that $\mathcal{P} \ll \nu_1$ and $\mathcal{P} \ll \nu_2$. Then, there is a measurable set $A \subset \Omega$ with $P_\theta(A) = 1$ for all $\theta \in \Theta$, and there exist versions $f_{1,\theta} \in \left(\mathrm{d}P_\theta/\mathrm{d}\nu_1\right)_\nu$ and $f_{2,\theta} \in \left(\mathrm{d}P_\theta/\mathrm{d}\nu_2\right)_\nu$, for all $\theta \in \Theta$, together with a measurable function h, such that

$$f_{1,\theta}(\omega) = h(\omega) f_{2,\theta}(\omega), \quad \forall \theta \in \Theta, \forall \omega \in A. \tag{2.1}$$

Note that Equation (2.1) implies that $f_{1,\theta}(\omega)$ and $f_{2,\theta}(\omega)$ differ by a factor $h(\omega)$ that does not depend on θ. Equivalently, we have that $f_{1,\theta}(\omega) \propto_\theta f_{2,\theta}(\omega)$, for $\omega \in A$. Thus, Definition 2.2 (likelihood in terms of RN derivatives) is validated under the LP. In practical terms, any two dominating measures lead to likelihood functions that coincide up to a multiplicative factor free of θ so giving the same inference. Since A is a set of probability one under every $P_\theta \in \mathcal{P}$, this covers all θ, including the true parameter θ^*.

It is important to observe that Theorem 2.6 guarantees only that some versions of the RN derivatives (for each dominating measure) satisfy Equation (2.1). Other versions might fail to do so. This points to the importance of establishing nice or canonical versions of these densities—such as continuous versions—so that proportionality holds on the entire relevant subset of Ω. We examine this issue further in Subection 2.2.

In special cases, the sets of versions $\left(\mathrm{d}P_\theta/\mathrm{d}\nu_1\right)_\nu$ and $\left(\mathrm{d}P_\theta/\mathrm{d}\nu_2\right)_\nu$ each contain exactly one element (that is, they are singletons), thereby making Equation (2.1) hold automatically. For instance, if $\mathcal{P}$ is a family of purely discrete distributions with a common (or at least majorizing) counting measure ν, then the ratios $\mathrm{d}P_\theta/\mathrm{d}\nu$ are unique up to sets of ν-measure zero. Another simpler example is when the family $\mathcal{P}$ itself is countable (that is, Θ is countable and each P_θ is a distinct measure). Under certain additional conditions –like sharing a common support– any two versions of $\mathrm{d}P_\theta/\mathrm{d}\nu_1$ and $\mathrm{d}P_\theta/\mathrm{d}\nu_2$ must differ by a factor not depending on θ; see Proposition 2.7 for a detailed statement.

Note that the framework of Theorem 2.6 (and the version of the LP it implies) concentrates on the situation in which two likelihoods arise from the same model $\mathcal{P}$ on the same sample space Ω. Thus, they yield the same estimators (such as ML) and the same associated inference procedures whenever these depend only on the relative values of the likelihood in θ. This includes essentially all frequentist estimators that can be defined purely in terms of the likelihood function.

Proposition 2.7 Suppose that Θ is countable and, in addition, that the measures $\{P_\theta : \theta \in \Theta\}$ share a common support in the sense that there exists a set $S \subset \Omega$ with $\nu(S^c) = 0$ for some dominating measure ν. Then, any pair of versions $\{f_{1,\theta}, f_{2,\theta}\}$ of $\mathrm{d}P_\theta/\mathrm{d}\nu_1$ and $\mathrm{d}P_\theta/\mathrm{d}\nu_2$ respectively, satisfies the proportionality relation of Equation (2.1) on a set of full ν-measure.

Next, we relate Theorem 2.6 to the factorization theorem.

Proposition 2.8 Consider the setting of Theorem 2.6, so that $\mathcal{P}$, ν_1, ν_2, and a minimal dominating measure ν are given, and let Q be the measure from Lemma 2.5. Let T be a sufficient statistic for $\mathcal{P}$ with range space $(\mathcal{T}, \mathcal{B})$. Then, we have that:

(i) For each version $g^*_\theta \in \left(\mathrm{d}P_\theta/\mathrm{d}Q\right)_\nu$ on $(\Omega, \sigma(T))$ and $h_1 \in (\mathrm{d}Q/\mathrm{d}\nu_1)_\nu$ on $(\Omega, \mathcal{F})$, there exists a $\mathcal{B}$-measurable function g_θ such that $g^*_\theta = g_\theta \circ T$ and $f_{1,\theta} = (g_\theta \circ T) h_1$ is a version in $\left(\mathrm{d}P_\theta/\mathrm{d}\nu_1\right)_\nu$ for all $\theta \in \Theta$.

(ii) If $f_{1,\theta}$ and $f_{2,\theta}$ are constructed as in (i) for ν_1 and ν_2, respectively, from the same $g^*_\theta \in \left(\mathrm{d}P_\theta/\mathrm{d}Q\right)_\nu$, then there is a measurable set $A \subset \Omega$ with $\nu(A^c) = 0$ on which $f_{1,\theta}(\omega) \propto_\theta f_{2,\theta}(\omega)$, for all $\omega \in A$ and $\theta \in \Theta$.

Part (i) of Proposition 2.8 may be viewed as a strong form of the factorization theorem, since it asserts the density representation $\mathrm{d}P_\theta/\mathrm{d}\nu_1$ is valid jointly for all $\theta \in \Theta$ on the entire space Ω (up to a set of measure zero). From this perspective, the usual statement of the factorization theorem follows, because any two versions of $\mathrm{d}P_\theta/\mathrm{d}\nu_1$ that differ only on a ν_1-null set also differ on a ν-null set.

Note that Theorem 2.6 does not require Ω to be a separable or even metric space; it remains valid in general measure-theoretic contexts, as long as the family $\mathcal{P}$ can be dominated by the σ-finite measures ν_1, ν_2. The regularity conditions needed concern σ-finiteness and domination, rather than separability or metrizability of Ω.

2.2. *Continuous versions of Radon-Nikodym derivatives*

Earlier sections indicated that different versions of the RN derivative may fail to satisfy the proportionality condition stated in Equation (2.1) unless chosen with care. In practice, it is often desirable to obtain versions that automatically fulfill this proportionality across different dominating measures. This subsection explains why continuous RN derivatives are especially advantageous for ensuring such consistency, and in certain cases, they are even unique up to sets of measure zero.

A particularly valuable class of versions arises when the RN derivatives are continuous functions of $\omega \in \Omega$. As we will see, these continuous versions, when they exist, yield likelihood functions that differ only by a factor not depending on θ. Two results (Theorem 2.12 and Proposition 2.14) demonstrate this proportionality in various scenarios.

Historically, Piccioni (1982, 1983) proposed defining the likelihood specifically as a continuous version of the RN derivative, showing that under mild conditions it is unique. The author linked continuity to a limiting procedure reflecting Fisher's intuitive notion of likelihood. Likewise, Berger and Wolpert (1988) argued for (ν-almost everywhere) continuous densities to replicate the simplicity of the discrete setting in developing conditionality, sufficiency, and LPs. Moreover, continuity in θ of the likelihood function (which often follows from continuity in ω) is typically assumed in the classical regularity conditions for ML estimation.

Throughout this subsection, we assume that Ω is a separable metric space with distance d, and let $\mathcal{F}$ be the Borel σ-algebra induced by the corresponding topology $\boldsymbol{A}$. Intuitively, continuity in ω supports a more direct notion of likelihood. In purely discrete models, the likelihood is simply the probability assigned to the observed sample point ω_0. In continuous models, that idea can be approximated by shrinking neighborhoods around ω_0. Concretely, one fixes a sequence of nested sets $A_1 \supset A_2 \supset \cdots \supset A_n \supset \dots$ within the collection of open neighborhoods $\mathfrak{A}(\omega_0)$ of ω_0, and then considers the limit

$$\lim_{n\to+\infty} \frac{P_\theta(A_n)}{\nu(A_n)}. \tag{2.2}$$

Piccioni (1982) showed that this limit is finite for every such chain $\{A_n\}$ if and only if there is a version of $\mathrm{d}P_\theta/\mathrm{d}\nu$ that is continuous in ω. In that situation, the value of this continuous version at ω_0 coincides with the limit stated in Equation (2.2).

Continuity of the densities in ω almost guarantees that they satisfy the proportionality relation presented in Equation (2.1) when two different dominating measures ν_1 and ν_2 are used. In Theorem 2.12 and Proposition 2.14 below, we formalize this statement under certain assumptions. To do so, we need a few definitions that apply in general measure-theoretic contexts (not necessarily metric).

Definition 2.9 Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $A \in \mathcal{F}$ be a measurable set with $\mu(A) \geq 0$. We denote by $\mu\big|_A$ the restriction of μ to the measurable space $(A, \mathcal{F}(A))$, that is, for every $B \in \mathcal{F}(A)$, $\mu\big|_A(B) = \mu(B)$.

We now present a lemma guaranteeing that, if two σ-finite measures both dominate $\mathcal{P}$, it is possible to find a subset of Ω on which these measures become equivalent.

Lemma 2.10 Let $\mathcal{P} = \{P_\theta \colon \theta \in \Theta\}$ be a family of probability measures on $(\Omega, \mathcal{F})$, and let ν_1 and ν_2 be two σ-finite measures such that $\mathcal{P} \ll \nu_1$ and $\mathcal{P} \ll \nu_2$. Suppose Θ is nonempty. Then, there exists a measurable set $A \subseteq \Omega$ such that

(i) $P_\theta(A) = 1$ for all $\theta \in \Theta$,
(ii) $\nu_1\big|_A$ and $\nu_2\big|_A$ are equivalent measures, that is $\nu_1\big|_A \ll \nu_2\big|_A$ and $\nu_2\big|_A \ll \nu_1\big|_A$.

The construction of such a set A typically uses a countable mixture argument similar to Lemma 2.5, ensuring that all P_θ place full probability on A, and that within A, the measures ν_1 and ν_2 do not vanish in disjoint ways. In the following subsections, we restrict our focus to this set A —thereby making $\nu_1|_A$ and $\nu_2|_A$ equivalent– and then prove that continuous densities with respect to each measure remain proportional.

Definition 2.11 Consider a family of probability measures $\mathcal{P} = \{P_\theta \colon \theta \in \Theta\}$ on $(\Omega, \mathcal{F})$, where Θ is nonempty, and let ν_1 and ν_2 be two σ-finite measures with $\mathcal{P} \ll \nu_1$ and $\mathcal{P} \ll \nu_2$. We call a pair (A, ν) a dominating pair for the triple $(\mathcal{P}, \nu_1, \nu_2)$ if the following conditions hold:

(i) $A \in \mathcal{F}$ satisfies $\nu(A) = 1$, where $\nu = \sum_{i=1}^{+\infty} c_i P_{\theta_i}$ is a minimal dominating measure of $\mathcal{P}$ (see Proposition 2.4),
(ii) On the restricted measurable space $(A, \mathcal{F}(A))$, the three measures $\nu_1\big|_A, \nu_2\big|_A$, and $\nu\big|_A$ are all mutually equivalent (that is, each one is absolutely continuous with respect to each of the others).

By Proposition 2.4 and Lemma 2.10, such a dominating pair (A, ν) always exists whenever $\mathcal{P} \ll \nu_1$ and $\mathcal{P} \ll \nu_2$. Its importance lies in forcing a common footing on the set A, where ν_1, ν_2, and ν share the same support and are mutually absolutely continuous. This is crucial in establishing the proportionality of continuous versions of RN densities.

Theorem 2.12 Let (A, ν) be a dominating pair for $(\mathcal{P}, \nu_1, \nu_2)$. Suppose there exist continuous (in ω) RN derivatives $f_{1,\theta} \in \left(\mathrm{d}P_\theta|_A / \mathrm{d}\nu_1|_A\right)_{\nu|_A}$ and $f_{2,\theta} \in \left(\mathrm{d}P_\theta|_A / \mathrm{d}\nu_2|_A\right)_{\nu|_A}$ for all $\theta \in \Theta$. Then, for any $h \in \left(\mathrm{d}(\nu_2|_A)/\mathrm{d}(\nu_1|_A)\right)_{\nu|_A}$, there is a measurable set $B_h \subseteq A$ such that $P_\theta(B_h) = 1$ for all $\theta \in \Theta$, the function h is continuous on B_h, and $f_{1,\theta}(\omega) = h(\omega) f_{2,\theta}(\omega)$ for all $\theta \in \Theta$ and $\omega \in B_h$.

Theorem 2.12 shows that once we restrict to the common support A (where the three measures ν_1, ν_2, and ν agree up to sets of measure zero), continuous RN derivatives $\{f_{1,\theta}\}$ and $\{f_{2,\theta}\}$ for any two dominating measures necessarily differ by a multiplicative factor $h(\omega)$ not depending on θ. In short, continuous versions automatically satisfy the main proportionality requirement from Theorem 2.6.

Furthermore, if the measures ν_1 and ν_2 are locally finite in the sense of Appendix A (which also implies they are σ-finite), then such continuous versions, if they exist, are often unique (see Theorem 6.8 in Appendix A). In many classical models, it is straightforward to produce these continuous versions –for example, when $\Omega \subset \mathrm{R}^d$ and the densities are continuous functions of $\omega \in \Omega$.

Let S_ν be the support of a measure ν on $(\Omega, \mathcal{F})$; see Definition 6.3 in Appendix A.

Corollary 2.13 Suppose that ν_1 and ν_2 are LF measures on $(\Omega, \mathcal{F})$ and that $S_{\nu_1} = \Omega$. Assume further that there exist continuous versions of the RN derivatives $f_{1,\theta} \in \left(\mathrm{d}P_\theta/\mathrm{d}\nu_1\right)_\nu$ and $f_{2,\theta} \in \left(\mathrm{d}P_\theta/\mathrm{d}\nu_2\right)_\nu$ for all $\theta \in \Theta$, where ν is a minimal dominating measure for $\mathcal{P}$. Suppose also that $f_{1,\theta}(\omega) > 0$ and $f_{2,\theta}(\omega) > 0$ for all $\omega \in \Omega$ and $\theta \in \Theta$. Then, we get $f_{1,\theta}(\omega) \propto_\theta f_{2,\theta}(\omega), \quad \forall \omega \in \Omega, \forall \theta \in \Theta$.

Proposition 2.14 Let $\mathcal{P} = \{P_\theta \colon \theta \in \Theta\}$ be a family of probability measures and ν_1 and ν_2 be LF measures on $(\Omega, \mathcal{F})$ such that $\mathcal{P} \ll \nu_1$ and $\mathcal{P} \ll \nu_2$. Suppose (A, ν_3) is a dominating pair for $(\mathcal{P}, \nu_1, \nu_2)$, and let S_θ, S_1, S_2, S_3 be the supports of P_θ, ν_1, ν_2, and ν_3, respectively. Suppose there exists a continuous version (on S_θ) of $f_{2,\theta} \in (\mathrm{d}P_\theta|_{S_\theta}/\mathrm{d}\nu_2|_{S_\theta})_{\nu_1|_{S_\theta}}$, for all $\theta \in \Theta$, and a continuous version (on S_3) of $h \in (\mathrm{d}(\nu_2|_{S_3})/\mathrm{d}(\nu_1|_{S_3}))_{\nu_1|_{S_3}}$, then both $f_{2,\theta}$ and h are unique (on S_θ and S_3, respectively), and there exists a unique continuous version of $f_{1,\theta} \in (\mathrm{d}P_\theta|_{S_\theta}/\mathrm{d}\nu_1|_{S_\theta})_{\nu_1|_{S_\theta}}$ for each $\theta \in \Theta$. Moreover, defining $\Phi_\omega = \{\theta \in \Theta \colon \omega \in S_\theta\}$, we have $f_{1,\theta}(\omega) \propto_\theta f_{2,\theta}(\omega) \quad$ for every $\theta \in \Phi_\omega$.

Proposition 2.14 further refines the uniqueness concern: once a continuous version of $\mathrm{d}P_\theta/\mathrm{d}\nu_2$ is obtained, together with a continuous version of $\mathrm{d}\nu_2/\mathrm{d}\nu_1$, a corresponding unique continuous version of $\mathrm{d}P_\theta/\mathrm{d}\nu_1$ is thereby determined. In other words, continuity and the previous choice of the density with respect to ν_2 jointly force the density with respect to ν_1. This ensures that all these densities are consistent and satisfy the proportionality relation on their common supports.

3. Special cases

In this section, we consider two particular scenarios that highlight how a dominating measure can be naturally chosen or shown to be irrelevant to the final form of the likelihood function. First, we discuss the prior predictive measure in a Bayesian framework and identify conditions under which it serves as a valid dominating measure for the model. Then, we show that exponential families (traditionally defined by densities relative to a fixed measure) are, in fact, a property of the model itself and hence remain exponential families regardless of which dominating measure is chosen.

3.1. *The predictive measure as a dominating measure*

Izbicki et al. (2014) proposed a nonparametric methodology for density-ratio estimation, illustrating how to employ the prior predictive measure in the denominator of such a ratio. This yields an approximation to the likelihood function when the latter is not otherwise tractable. Below, we investigate conditions ensuring that this predictive measure can dominate the entire model $\mathcal{P}$.

Let X be a sample drawn from a parametric family $\mathcal{P} = \{P_\theta \colon \theta \in \Theta\} \subset \mathcal{M}(\mathcal{X})$, where $\Theta \subset \mathrm{R}^k$ and $\mathcal{X}$ is the sample space. Assume we have a nonzero prior R on Θ. Let $\mathcal{B}_\mathcal{X}, \mathcal{B}_\Theta$ be the respective σ-fields on $\mathcal{X}$ and Θ. Suppose that for each fixed $B \in \mathcal{B}_\mathcal{X}$, the function $\theta \mapsto P_\theta(B)$ is Borel measurable. Then, there is a unique probability measure P on $\mathcal{X} \times \Theta$ such that $P(B \times C) = \int_C P_\theta(B)\mathrm{d}R(\theta)$, $B \in \mathcal{B}_\mathcal{X}, C \in \mathcal{B}_\Theta$ (Shao, 2003, Ch. 4). The posterior distribution $P_{\theta|x}$ of θ given $X = x$ arises from Bayes' theorem whenever densities exist.

3.2. *Predictive measure and domination*

Assume $\mathcal{P}$ is dominated by a σ-finite measure ν, so that $f_\theta(x) = \mathrm{d}P_\theta/\mathrm{d}\nu(x)$ is a Borel function on $\mathcal{X}$. Define $m(x) = \int_\Theta f_\theta(x)\mathrm{d}R(\theta)$ and refer to m as the marginal density of X with respect to ν. If $m(x) > 0$ on a set of ν-full measure in $\mathcal{X}$, then $\mathrm{d}P_{\theta|x}/\mathrm{d}R(\theta) = f_\theta(x)/m(x)$ is well defined with the set $N = \{x \in \mathcal{X} \colon m(x) = 0\}$ playing a key role. On N, the likelihood (with respect to ν) is zero R-almost everywhere.

Next, define the predictive measure λ on $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ by $\lambda(A) = \int_A m(x)\nu(\mathrm{d}x), \quad A \in \mathcal{B}_{\mathcal{X}}$. The set N, where m vanishes, determines whether or not λ dominates each P_θ. The following results make this precise.

Proposition 3.1 The predictive measure λ does not depend on which σ-finite measure ν was used to dominate $\mathcal{P}$, aside from sets of measure zero.

Proposition 3.2 Let $m(x) > 0$ for all $x \in \mathcal{X}$. Then, the predictive measure λ dominates $\mathcal{P}$.

Note that, in many cases, it suffices to have $\nu(N) = 0$. However, if we only know that $P_\theta(N) = 0$ for R-almost every θ, we cannot guarantee that λ dominates $\mathcal{P}$.

Proposition 3.3 The predictive measure λ dominates P_θ if and only if $P_\theta(N) = 0$. Consequently, λ dominates the entire model $\mathcal{P}$ if and only if $P_\theta(N) = 0$ for all $\theta \in \Theta$.

The following result is often a practical criterion.

Proposition 3.4 If $M_\theta = \{x \in \mathcal{X}\colon f_\theta(x) > 0\}$ is the same set for all $\theta \in \Theta$, then $\mathcal{P} \ll \lambda$; that is, the predictive measure λ dominates $\mathcal{P}$.

These propositions confirm that if the marginal density m is positive on a set of full ν-measure, then λ can serve as a dominating measure for the model. This viewpoint justifies the practice in some Bayesian settings of using the prior predictive (or marginal) measure as the baseline for constructing likelihood functions, even if a dominating measure was given.

3.3. *Exponential families*

A parametric family $\mathcal{P} = \{P_\theta\colon \theta \in \Theta\}$, dominated by a σ-finite measure ν on $(\Omega, \mathcal{F})$, is called an exponential family if and only if there exist measurable functions $\eta\colon \Theta \to \mathrm{R}^p$ and $T\colon \Omega \to \mathrm{R}^p$ (for some fixed $p \in \mathrm{N}$), a nonnegative measurable function h, and a function $\xi\colon \Theta \to \mathrm{R}$ such that

$$\mathrm{d}P_\theta/\mathrm{d}\nu(\omega) = \exp((\eta(\theta))^\tau T(\omega) - \xi(\theta))h(\omega), \quad \omega \in \Omega, \tag{3.3}$$

where $\xi(\theta) = \log(\int_\Omega \exp((\eta(\theta))^\tau T(\omega))h(\omega)\mathrm{d}\nu(\omega))$. A thorough treatment of exponential families may be found in Jorgensen and Labouriau (2012).

In view of Equation (3.3), which explicitly depends on a particular dominating measure ν, it is relevant to investigate whether the notion of an exponential family is an intrinsic property of $\mathcal{P}$, or if it can vary with the choice of dominating measure ν. In particular, if ν is replaced by another σ-finite measure μ dominating $\mathcal{P}$, the form of $\mathrm{d}P_\theta/\mathrm{d}\mu$ may look distinct. Nonetheless, the following proposition confirms that the exponential-family structure remains invariant, regardless of which σ-finite measure is chosen to dominate $\mathcal{P}$.

Proposition 3.5 A parametric family $\mathcal{P}$ is an exponential family if and only if it admits a representation of the form stated in Equation (3.3) with respect to some dominating measure. Equivalently, if $\mathcal{P}$ is an exponential family, then for any σ-finite measure μ dominating $\mathcal{P}$, there exist measurable functions η, T, ξ, and a nonnegative function $h_\mu(\omega)$ such that $\mathrm{d}P_\theta/\mathrm{d}\mu(\omega) = \exp((\eta(\theta))^\tau T(\omega) - \xi(\theta))h_\mu(\omega)$, for $\omega \in \Omega, \forall\theta \in \Theta$. Hence, being an exponential family is a property of the model $\mathcal{P}$ itself, independent of the choice of dominating measure.

4. Examples

In this section, we provide several examples in which the choice of dominating measure is nontrivial or at least merits careful consideration. We illustrate how Theorem 2.6 (likelihood proportionality) guarantees a valid likelihood function once a suitable dominating measure is chosen, leading to the same inference in θ. We also highlight the relevance of continuous versions of RN derivatives, as discussed in Section 2.2, whenever such continuous densities are available.

4.1. *Finite-dimensional random variables*

A common setting for statistical models is a family $\mathcal{P}$ of probability measures on R^d (or a product of discrete and continuous components). Typical examples include the following:

- Independent and identically distributed univariate random variables,
- Multivariate distributions with both discrete and continuous coordinates,
- Hierarchical Bayesian models with mixture components.

Often, the default or natural dominating measure is the product of counting measures for discrete coordinates and Lebesgue measures for continuous coordinates. However, any probability measure with common support can also serve as a valid dominating measure, and Theorem 2.6 ensures that all such choices yield likelihoods proportional in θ almost surely.

A particularly instructive example is a point-mass mixture. Consider a random variable Y that, with probability $p < 1$, takes values $\{a_1, \ldots, a_m\} \subset \mathrm{R}$ at point masses, and otherwise assumes values in (subsets of) R continuously. More precisely, let $P\big(Y = a_i\big) = p_i > 0$, for $i \in \{1, \ldots, m\}$, and $\sum_{i=1}^{m} p_i = p < 1$, that is, the total point-mass probability is p, and the remaining probability $1 - p$ is allocated to a continuous component on R.

On the complementary event (probability $1 - p$), let Y have density f_j on $B_j \subset \mathrm{R}$ with weight q_j, $j \in \{1, \ldots, n\}$, so that $\sum_{j=1}^{n} q_j = 1 - p$. Gottardo and Raftery (2009) showed that the distribution P of Y is dominated by $\nu_1 + \nu_2$, where ν_1 is the counting measure and ν_2 is the Lebesgue measure, leading to

$$\frac{\mathrm{d}P}{d(\nu_1 + \nu_2)}(y) = \sum_{i=1}^{m} p_i 1_{\{a_i\}}(y) + \sum_{j=1}^{n} q_j f_j(y) 1_{B_j \setminus A}(y), \tag{4.4}$$

where $A = \{a_1, \ldots, a_m\}$ and $1_B(y)$ is the indicator function of $y \in B$. Ignoring the indicator terms in Equation (4.4) (that is, using an improper version of the RN derivative) can cause serious mis-specifications of the likelihood. By contrast, including those indicators correctly ensures that each point-mass and continuous portion is treated in its correct domain.

It is often desirable to have continuous versions of f_j on B_j. Whenever such continuity holds, it further ensures that likelihoods obtained from different dominating measures remain not only proportional in θ, but also coincide with the intuitive limit notion stated in Equation (2.2) around each observed point in the continuous region.

The result given in Gottardo and Raftery (2009) is more general, covering countable mixtures of mutually singular measures. For examples of point-mass mixture modeling in real-world data analysis, see Schmidt et al. (2017), Gonçalves et al. (2022), and Bhattacharjee and Chakraborty (2023).

4.2. *Missing data problems*

Consider a model $\mathcal{P} = \{P_\theta \colon \theta \in \Theta\}$ on $(\Omega, \mathcal{F})$, where Ω factors as $\Omega_1 \times \Omega_2$ and $\mathcal{F} = \sigma(\mathcal{F}_1 \times \mathcal{F}_2)$. In a typical missing data scenario, $\omega_1 \in \Omega_1$ is observed, whereas $\omega_2 \in \Omega_2$ is not. Often, the marginal density with respect to some measure on Ω_1 may be intractable, whereas the joint density on Ω is known or more manageable; see Gonçalves and Gamerman (2018) for an example. In such situations, a pseudo-likelihood is derived from the joint density of P_θ relative to a dominating measure on Ω, integrating out the unobserved portion ω_2. Under frequentism, this is often handled by the expectation-maximization (EM) or Monte Carlo EM algorithms, which require either integration or sampling with respect to the conditional distribution of $\omega_2 \mid \omega_1$. In a Bayesian framework (such as Markov chain Monte Carlo —MCMC—), it is also necessary to sample ω_2 from that conditional distribution.

Now suppose two distinct dominating measures, $\nu_1 = \nu_{1,1} \times \nu_{1,2}$ and $\nu_2 = \nu_{2,1} \times \nu_{2,2}$, are available for the same family $\mathcal{P}$. Denote the corresponding pseudo-likelihood functions by $\pi_{\theta,i}(\omega_1, \omega_2)$, where $\pi_{\theta,i}$ is the RN derivative with respect to ν_i. Then, we have $\pi_{\theta,i}(\omega_2 \mid \omega_1) \propto_{\omega_2} \pi_{\theta,i}(\omega_1, \omega_2)$, for $i \in \{1, 2\}$. By Theorem 2.6, these two likelihood expressions are proportional in θ (almost surely under each P_θ). However, that proportionality does not imply they are equal as functions of ω_2. Indeed, each measure ν_i induces a different conditional density $\pi_{\theta,i}(\omega_2 \mid \omega_1)$, potentially affecting the complexity of any integration or sampling procedure. In EM or MCMC contexts, a given dominating measure may be substantially simpler than another for sampling from the conditional distribution or for computing the relevant integrals.

Therefore, although Theorem 2.6 confirms that inference about θ is unaffected by which dominating measure is chosen, the practical implications for algorithmic complexity can be substantial. In particular, even though the resulting inference does not change, the computational cost may vary considerably if one dominating measure is notably more convenient for generating conditional samples than another.

4.3. *Poisson processes*

PPs are among the most widely used models for point patterns on a region $S \subset \mathrm{R}^d$, although they can be defined in more general measurable spaces (Kingman, 1993, Ch. 2). Let P_λ denote the law of a PP on S with intensity function λ. We now discuss two natural constructions for dominating P_λ.

A first construction is the following. Represent each realization ω of the PP as $(N, s_1, \ldots, s_N)$, where N is the random number of points and $s_j \in S$ are their locations. Factor its joint density as $\pi(N)\pi(s_1, \ldots, s_N \mid N)$ and use as dominating measure $\nu_3 = \sum_{k=0}^{+\infty} \nu_{3,k}$, where $\nu_{3,k} = \nu_1 \otimes (\nu_2)^k$, with ν_1 the counting measure on $\{0, 1, \ldots\}$ for N, and ν_2^k the k-dimensional Lebesgue measure for the locations. Hence, we obtain

$$\frac{\mathrm{d}P_\lambda}{\mathrm{d}\nu_3}(\omega) = \frac{1}{N!} \exp\left(-\int_S \lambda(s)\mathrm{d}s\right) \left(\int_S \lambda(s)\mathrm{d}s\right)^N \prod_{j=1}^{N} \frac{\lambda(s_j)}{\int_S \lambda(s)\mathrm{d}s}. \tag{4.5}$$

A second construction is the following. Alternatively, let ν be the law of another PP on S whose intensity function γ is strictly positive everywhere on S. In particular, it is possible to choose $\gamma \equiv 1$.

By the Jacod formula (Andersen et al., 1993, Corollary II.7.3), we obtain

$$\frac{\mathrm{d}P_\lambda}{\mathrm{d}\nu}(\omega) = \exp\left(-\int_S (\lambda(s) - \gamma(s))\mathrm{d}s\right) \prod_{j=1}^{N} \frac{\lambda(s_j)}{\gamma(s_j)}, \tag{4.6}$$

where N is again the number of points in the realization ω, and s_j its points.

It is straightforward to verify that Equations (4.5) and (4.6) are proportional in λ. Thus, for standard inference where an observed realization ω is available and the goal is to estimate λ, both formulations yield the same inferences. However, in more complex scenarios (such as partial observation of the process), the choice of dominating measure can be crucial, as it may drastically affect the computation of likelihoods or marginalizations (see Section 5.2 for details).

Continuity in the Skorokhod topology is as follows. If $S \subset \mathrm{R}$ and we view the realizations ω as càdlàg functions in the Skorokhod space D, then the density stated in Equation (4.6) can be shown to be continuous as a map from $\omega \in D$ to R. Since D is separable under the Skorokhod topology, standard results about continuous versions of likelihoods apply naturally in this framework.

4.4. *Diffusion processes*

Stochastic differential equations driven by Brownian motion, often called diffusions, are widely used in statistics to model continuous-time phenomena. Formally, a diffusion process $Y = \{Y_s\colon s \in [0,t]\}$ is the unique solution to an stochastic differential equation of the form $dY_s = a(Y_s, \theta)ds + \sigma(Y_s, \theta)dW_s, \quad s \in [0,t], \quad Y_0 = y_0$, where W_s is a standard Brownian motion, and a, σ satisfy mild regularity conditions ensuring existence and uniqueness (Kloeden and Platen, 1995). Typical trajectories of a diffusion are almost surely continuous but nowhere differentiable.

Despite their flexibility, diffusions pose major challenges for statistical inference: they evolve in infinite-dimensional spaces, often lack tractable transition densities. Then, under discrete observation, the exact likelihood is unavailable in closed form. Several exact or pseudo-likelihood methods (Beskos et al., 2006) aim to avoid discretization biases by working directly with the continuous-time likelihood of the entire path. However, these approaches demand a valid dominating measure. In fact, if the diffusion coefficient σ varies with θ over an uncountable set, distinct values of θ typically yield mutually singular measures on the path space. Hence, there may be no single σ-finite measure dominating all such laws (see also Gottardo and Raftery, 2009). (If Θ is countable, the family can be dominated by a countable sum of measures.)

A classical solution to this problem involves transforming the original diffusion paths so that they become dominated by a parameter-free measure. In a discrete-approximation context, Roberts and Stramer (2001) propose a pair of transformations; in continuous time, Beskos et al. (2006) develop exact algorithms based on a similar idea. Concretely, let $Y = \{Y_s\colon s \in [0,t]\}$ be a univariate diffusion observed at times $0 = t_0 < t_1 < \cdots < t_n = t$, with observations $y_0, \ldots, y_n$. We map (Y_s) into $(Y_{\text{obs}}, \dot{X})$, where $Y_{\text{obs}} = (y_0, \ldots, y_n)$ are the observed points, and $\dot{X}$ encapsulates the bridge segments between y_{i-1} and y_i.

First, we describe the Lamperti transform. Define $X_s = \eta(Y_s, \theta) = \int_0^{Y_s} (1/\sigma(u,\theta))\mathrm{d}u$, so that X_s solves an stochastic differential equation with unit diffusion coefficient and some drift $\alpha(X_s, \theta)$ derived from a and σ. Mapping each observation y_i to $x_i(\theta) = \eta(y_i, \theta)$ aligns the process into a standard scale. Now, we consider bridges with zero endpoints. For $s \in (t_{i-1}, t_i)$, define $\dot{X}_s = X_s - (1-(s-t_{i-1})/(t_i-t_{i-1}))x_{i-1}(\theta) - ((s-t_{i-1})/(t_i-t_{i-1}))x_i(\theta)$, with inverse φ_θ. Hence, each $\dot{X}_s$ starts and ends at 0 over the interval (t_{i-1}, t_i), which makes it dominated by a standard Brownian bridge measure.

Collecting $(Y_{\text{obs}}, \dot{X})$ yields a density with respect to the product measure $\nu^n \otimes \mathrm{W}^n$, where ν^n is the n-dimensional Lebesgue measure and W^n is a product of n standard Brownian-bridge laws over the subintervals.

Beskos et al. (2006, see Lemma 2) derived an explicit expression given by

$$\begin{aligned}\pi(Y_{\text{obs}}, \dot{X}) = {}& \prod_{i=1}^n \eta'(y_i;\theta)\phi\left(\tfrac{x_i(\theta)-x_{i-1}(\theta)}{\sqrt{t_i-t_{i-1}}}\right) \\ & \times \exp\left(\Delta A(x_0(\theta), x_n(\theta);\theta) - \int_0^t \left(\tfrac{\alpha^2+\alpha'}{2}\right)(\varphi_\theta(\dot{X}_s);\theta)\mathrm{d}s\right),\end{aligned} \tag{4.7}$$

where ϕ is the standard normal density, $\Delta A(x_0(\theta), x_n(\theta);\theta) = A(x_n(\theta);\theta) - A(x_0(\theta);\theta)$, and $A(u;\theta) = \int_0^u \alpha(z,\theta)\mathrm{d}z$.

Some continuity considerations are the following. If σ is continuously differentiable, it is typically possible to establish that Equation (4.7) is continuous in the sup norm on $C[0,t]$, the space of continuous paths on $[0,t]$, which is a separable metric space. In essence, the transformations described above provide a parameter-free dominating measure (the product of Brownian-bridge laws) for the augmented process $(Y_{\text{obs}}, \dot{X})$. This permits a valid likelihood in spite of the original family being mutually singular for distinct θ.

5. Numerical experiments

In this section, we present two sets of numerical experiments to illustrate the validity and practical impact of likelihood constructions under different dominating measures. The first set concerns point-mass mixtures, where we compare Monte Carlo estimates of the ML estimator using both simulated data and a real rainfall dataset from Porto Alegre, Brazil. The second set involves inhomogeneous PPs with artificial missing data, for which we implement a Monte Carlo EM algorithm to estimate the underlying intensity function. All experiments were carried out in `R` (R Core Team, 2024) on a MacBook Air (Apple M2 chip, 16 GB memory), and the datasets and scripts are publicly available online (Gonçalves and Franklin, 2023).

5.1. *Point-mass mixtures*

We first consider a point-mass mixture model with two components: a degenerate component at zero and a gamma distribution. Let p be the point-mass probability at zero and (α, β) the shape and rate parameters of the gamma distribution. The overall mixture likelihood is given by Equation (4.4), which we use to form the ML estimator.

We simulated $200,000$ replications of a sample of size $n = 1500$ from the mixture with $p = 0.6$, $\alpha = 0.5$, and $\beta = 0.05$. Each replication yielded an ML estimate of (p, α, β), with p simply the empirical proportion of zeros, and (α, β) estimated via numerical optimization. The runs were parallelized on 7 cores. Each replication took about 3.85×10^{-4} seconds on average. Figure 1 displays the empirical distribution of the ML estimators over these replications, with vertical lines at the true values.

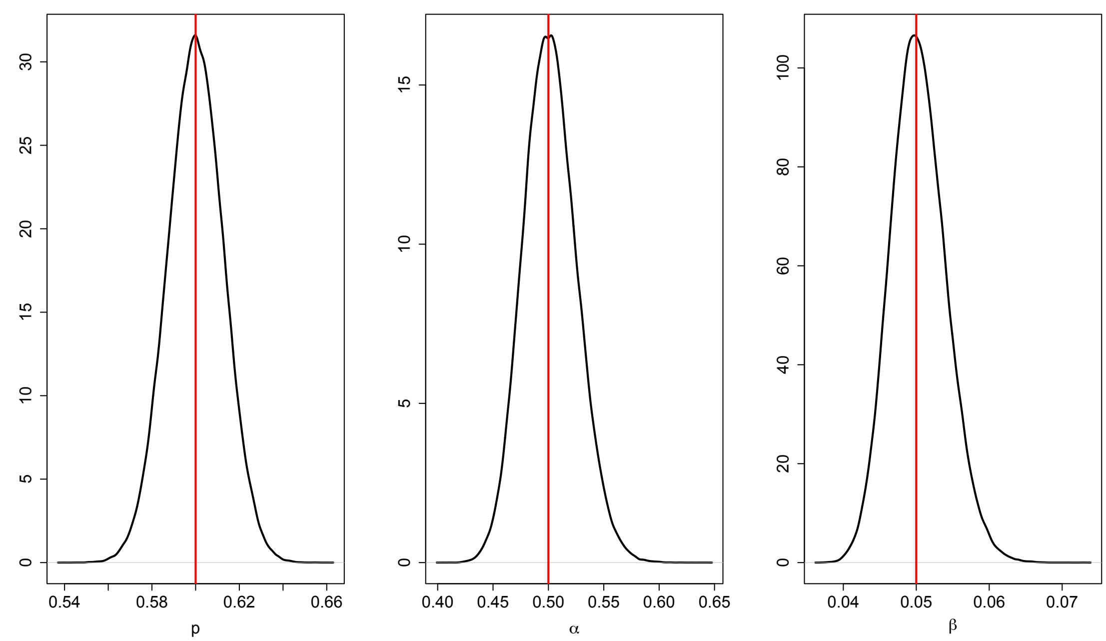

Figure 1. Empirical distribution of the ML estimator for (p, α, β) in the two-component point-mass mixture simulation. The vertical lines indicate the true parameter values.

Next, we fit the same point-mass-gamma mixture to daily rainfall data (in mm) from a station in Porto Alegre, Brazil, recorded from January 1, 2004 through 31 December 2023 (6585 daily measurements, 3880 of which are zero; a few days are missing). The ML estimates are $\widehat{p} = 0.589$, $\widehat{\alpha} = 0.5197$, $\widehat{\beta} = 0.0534$. Figure 2 shows a histogram of the positive rainfall values, overlaid with the fitted gamma density.

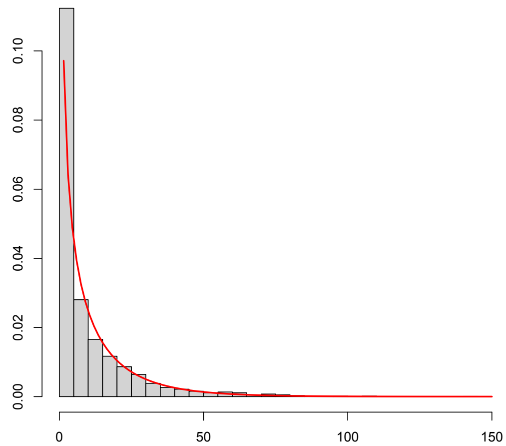

Figure 2. Histogram of positive daily rainfall and fitted gamma density. The ML estimates are $\widehat{\alpha} = 0.5197$, $\widehat{\beta} = 0.0534$.

5.2. *Poisson process with artificial missing data*

We next consider inhomogeneous PPs on $[0, 100]$ with intensity functions lacking closed-form integrals, making the usual likelihood intractable. We employ the representation in Equation (4.6) but introduce a missing-data formulation, effectively augmenting the latent structure of the process for easier computation.

We simulate 50 datasets from a PP Y on $[0, 100]$ with intensity function $\lambda_Y(s;\theta) = 4g(s;\theta)$, where

$$g(s;\theta) = \frac{1.1 + \sin(a_1 \sin(a_2 s + b_2)s + b_1)}{2.11},$$

so that $0.047 \leq g(s;\theta) \leq 0.995$, and $\theta = (a_1, b_1, a_2, b_2) = (0.045, 1.5707, 0.09, 0)$. To circumvent the intractable integral, we define a latent PP X on $[0, 100]$ with intensity $\lambda_X(s;\theta) = 4(1 - g(s;\theta))$. Given θ, X is independent of Y.

Then, the augmented likelihood function for (Y, X) can be written as

$$\begin{aligned} L(\theta; y, x) &\propto_\theta \exp(-\int_0^{100} \lambda_Y(s;\theta) + \lambda_X(s;\theta)\mathrm{d}s) \prod_{j=1}^{n} \lambda_Y(y_j;\theta) \prod_{j=1}^{m} \lambda_X(x_j;\theta) \\ &\propto_\theta \prod_{j=1}^{n} g(y_j;\theta) \prod_{j=1}^{m} (1 - g(x_j;\theta)), \end{aligned} \tag{5.8}$$

where $(y_1, \ldots, y_n)$ and $(x_1, \ldots, x_m)$ are the observed points of Y and X, respectively.

Then, we apply a Monte Carlo EM algorithm. In each E-step, we sample X conditional on the current θ', using Poisson thinning (Gonçalves and Gamerman, 2018), and compute the expected log-likelihood; in the M-step, we numerically maximize this expectation over θ. Each iteration (5,000 Monte Carlo samples, 7-core parallelization) takes about 40 seconds, mostly spent in the M-step. Figure 3 shows the true intensity function and the ML estimates across 50 replications. The mean and standard deviation of the estimated θ across the 50 runs are $(0.0453, 1.5727, 0.0886, 0.1003)$ and $(0.0019, 0.0852, 0.0019, 0.1431)$, respectively.

We also fit a similar PP model to the coal mining disaster data obtained from Jarrett (1979), which record 191 fatal explosions (killing 10 or more men) in Britain between 15 March 1851 and 22 March 1962. Rescale time to $[0, 111]$ years.

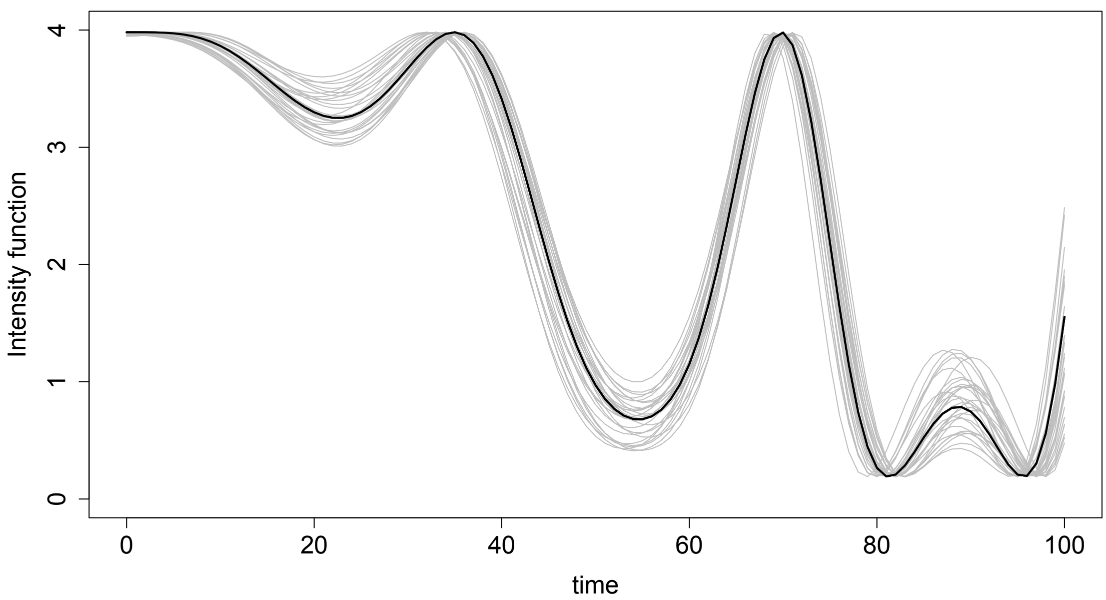

Figure 3. True intensity (black) and ML estimates (gray) across 50 simulated PPs under the augmented approach.

Specifically, we use

$$\lambda(s;\theta) = 0.3 + 3.2\Big(1 - \Phi\Big(\frac{s-\mu_1}{\sigma_1}\Big)\Big) + \frac{1.2\phi((s-\mu_2)/(\sigma_2))}{\phi(0)},$$

where Φ and ϕ are the standard normal distribution function and density, respectively, and $\lambda(s;\theta)/4.7$ plays the role of $g(s;\theta)$ in Equation (5.8). We run the same Monte Carlo EM scheme as before (5,000 samples in each E-step, 7-core parallelization), taking about 18 seconds per iteration. Figure 4 shows the resulting ML estimate of $\lambda(s;\theta)$. The fitted parameter values are $\widehat{\mu}_1 = 41.82$, $\widehat{\sigma}_1 = 16.54$, $\widehat{\mu}_2 = 86.30$, and $\widehat{\sigma}_2 = 7.00$.

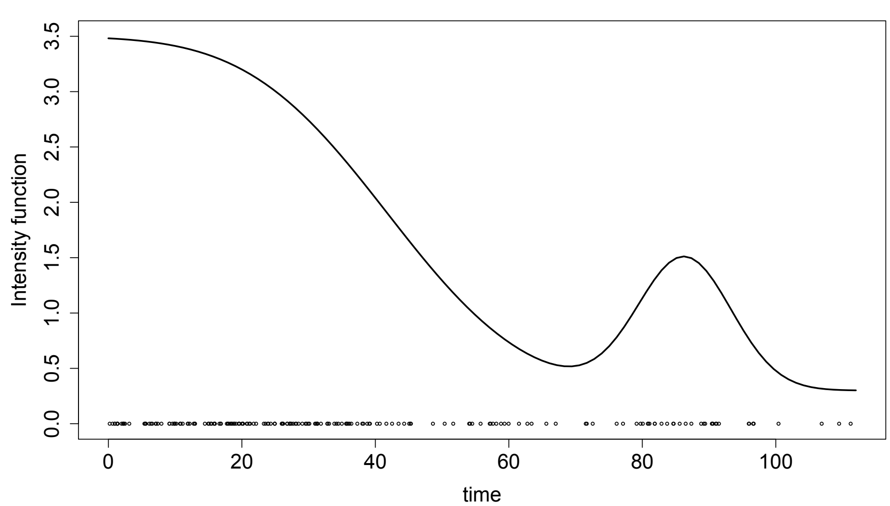

Figure 4. ML fit of the intensity function to the coal mining disaster data (points indicate observed events).

6. Conclusions

In this article, we have investigated the measure-theoretic foundations of likelihood theory, emphasizing a general definition of the likelihood function via Radon-Nikodym derivatives in both parametric and nonparametric settings. We introduced a likelihood proportionality theorem, establishing that any two dominating measures for a given statistical model yield likelihood functions that are almost surely proportional with respect to the parameter. This result validates the Radon-Nikodym-based definition of likelihood in accordance with the likelihood principle.

Beyond establishing proportionality in general, we offered a practical approach to identifying those versions of the Radon-Nikodym derivatives that guarantee the simplest form of proportionality. Specifically, we highlighted continuous versions of densities, showing that, under certain mild conditions, they are almost surely proportional and often unique. This underscores the consistency between continuous densities and the likelihood principle, and it can help practitioners obtain well-behaved likelihoods.

Despite clarifying how to select and handle dominating measures in many scenarios, our work leaves open some interesting questions about complex or infinite-dimensional models, in which no single σ-finite measure may dominate the entire family. Developing new methods to construct or approximate likelihoods in these challenging settings stands out as a fruitful direction for future research. Likewise, a deeper understanding of measure equivalences in high- or infinite-dimensional spaces would have broad implications for likelihood-based inference.

In conclusion, the results presented in this article establish a rigorous measure-theoretic foundation for likelihood. By employing Radon-Nikodym derivatives across a broad range of models, the proposed framework preserves the core principles of likelihood-based inference and provides a unified perspective that practitioners can reliably adopt in diverse statistical contexts.

Appendix A: Results and definitions

Most of the following definitions and results are adapted from Piccioni (1982). Propositions 6.6 and 6.9 are original to this article. Throughout, we assume Ω is a separable metric (or Lindelöf) space with Borel σ-algebra $\mathcal{F}$, so that references to neighborhoods and open sets are measure-theoretically meaningful.

Definition 6.1 A measure ν on $(\Omega, \mathcal{F})$ is called LF if, for every point $\omega \in \Omega$, there exists an open set (neighborhood) $U_\omega \subset \Omega$ containing ω such that $\nu(U_\omega) < +\infty$. Note that under our assumption that Ω is a Lindelöf space, every open cover admits a countable subcover. Hence, local finiteness implies σ-finiteness. In more general topological spaces, LF alone may not suffice to guarantee σ-finiteness.

Theorem 6.2 Let ν be an LF measure on $(\Omega, \mathcal{F})$ and Ω is Lindelof. Then, ν is σ-finite.

Definition 6.3 Let ν be a measure on $(\Omega, \mathcal{F})$. A point $\omega \in \Omega$ is called impossible for ν if there is a measurable neighborhood $U \subset \Omega$ of ω such that $\nu(U) = 0$. The support of ν, denoted S_ν, is the set of all points in Ω that are not impossible. Equivalently, $S_\nu = \bigcap_{\substack{U \subset \Omega \\ U \text{ open}, \nu(U)>0}} U^c$, that is, the smallest closed set on which ν is concentrated.

Remark 1 Consider the following points in Definition 6.3:

- If $\nu \equiv 0$ is the zero measure, then S_ν is empty by definition.
- If $\nu(\Omega) > 0$, then typically S_ν is nonempty and closed, and $\nu(S_\nu^c) = 0$.

Proposition 6.4 Let ν be an LF measure with $\nu(\Omega) > 0$. Then, $S_\nu \neq \varnothing$, that is, a nontrivial LF measure cannot be supported on the empty set.

Theorem 6.5 (Properties of support) If ν is an LF measure on $(\Omega, \mathcal{F})$ with $\nu(\Omega) \in (0, +\infty]$, then S_ν is a closed set and $\nu(S_\nu^c) = 0$. In particular, if $\nu(\Omega) < +\infty$, then $\nu(S_\nu) = \nu(\Omega)$.

Proposition 6.6 Let ν and μ be any two measures on $(\Omega, \mathcal{F})$, and let S_ν and S_μ denote their respective supports. If $\nu \ll \mu$, then $S_\nu \subseteq S_\mu$.

Remark 2 In Proposition 6.6, if ν is absolutely continuous with respect to μ, then a set of positive ν-measure must also have positive μ-measure, so ν cannot place mass outside the support of μ.

Theorem 6.7 (Piccioni, 1982) Let μ and ν be LF measures on $(\Omega, \mathcal{F})$ such that $\mu \ll \nu$ and $S_\mu = S_\nu = \Omega$. If there exists a continuous version of $\mathrm{d}\mu/\mathrm{d}\nu$ on Ω, then it is unique (up to a set of ν-measure zero).

Theorem 6.8 (A variant of uniqueness) Let μ and ν be LF measures on $(\Omega, \mathcal{F})$ with $\mu \ll \nu$. Suppose there is a continuous version of $\mathrm{d}\mu/\mathrm{d}\nu$ on the support S_μ. Then that continuous version is unique on S_μ (up to ν-null sets).

Remark 3 The property stated in Theorem 6.8 is particularly advantageous in settings where Ω is not compact, but continuity is only required on the region where μ concentrates its mass.

Auxiliary lemma for Lemma 2.10. Let $(\Omega, \mathcal{F}, \nu)$ be a measure space and $f\colon \Omega \to \mathrm{R}$ be a nonnegative measurable function. If $A \in \mathcal{F}$ satisfies $\nu(A) > 0$ and $f(\omega) > 0$ for all $\omega \in A$, then $\int_A f \mathrm{d}\nu > 0$. This is a direct consequence of Fubini-type arguments: since f never vanishes on A, the integral must be strictly positive.

Proposition 6.9 Let $\mathcal{P} = \{P_\theta\colon \theta \in \Theta\}$ be an exponential family as in the expression given in Equation (3.3) of the main text. For any $A \in \mathcal{F}$, define a measure λ on $(\Omega, \mathcal{F})$ by $\lambda(A) = \int_A h(\omega)\mathrm{d}\nu(\omega)$, where h is the base measure factor appearing in $\mathrm{d}P_\theta/\mathrm{d}\nu$. Then, we have that: (i) λ is σ-finite on $(\Omega, \mathcal{F})$; (ii) $\mathcal{P} \ll \lambda$, that is each P_θ is absolutely continuous with respect to λ; and (iii) $\mathrm{d}P_\theta/d\lambda(\omega) = \exp((\eta(\theta))^\tau T(\omega) - \xi(\theta))$ for all $\omega \in \Omega$.

Proposition 6.9 shows how to construct a new measure λ from the exponential-family base function $h(\omega)$. Because h is nonnegative, λ typically is σ-finite. The RN-derivative under this new measure then becomes a simpler exponential function of $\eta(\theta)T(\omega)$, illustrating the measure invariance of exponential family structure; see Proposition 3.5 in the main text.

Appendix B: Proofs

Proof [Proposition 2.4] Since $\Upsilon \neq \varnothing$, there is at least one $\nu \in \Upsilon$ such that $\mathcal{P} \ll \nu$. By Lemma 2.5 (Halmos-Savage), there exists a measure $\lambda = \sum_{i=1}^{+\infty} c_i P_{\theta_i}$ (for some probabilities c_i summing to 1) such that $\mathcal{P} \ll \lambda$. It is claimed that λ is a minimal dominating measure, that is $\lambda \ll \nu'$ for every $\nu' \in \Upsilon$. To see this, fix any $\nu' \in \Upsilon$. If $A \in \mathcal{F}$ satisfies $\nu'(A) = 0$, then $P_\theta(A) = 0$ for all $\theta \in \Theta$. In particular, we get $P_{\theta_i}(A) = 0$ for every i. Hence, we have $\lambda(A) = \sum_{i=1}^{+\infty} c_i P_{\theta_i}(A) = 0$, so $\lambda \ll \nu'$. Thus, we have λ is indeed minimal in Υ. ■

Proof [Theorem 2.6] By Proposition 2.4, there exists a minimal dominating measure ν for $\mathcal{P}$. Choose any versions $h_1 \in (\mathrm{d}\nu/\mathrm{d}\nu_1)_\nu, h_2 \in (\mathrm{d}\nu/\mathrm{d}\nu_2)_\nu, g_\theta \in (\mathrm{d}P_\theta/\mathrm{d}\nu)_\nu$ for all $\theta \in \Theta$. Define, for each $\theta \in \Theta$, $f_{1,\theta}(\omega) = g_\theta(\omega)h_1(\omega)$ and $f_{2,\theta}(\omega) = g_\theta(\omega)h_2(\omega)$. Clearly, $f_{1,\theta} \in (\mathrm{d}P_\theta/\mathrm{d}\nu_1)_\nu$ and $f_{2,\theta} \in (\mathrm{d}P_\theta/\mathrm{d}\nu_2)_\nu$. Let $A = \{\omega \in \Omega\colon h_2(\omega) > 0\}$.

Since h_2 is a version of $\mathrm{d}\nu/\mathrm{d}\nu_2$, we have $\nu(A^c) = 0$. Consequently, $P_\theta(A^c) = 0$ for all $\theta \in \Theta$. Then, define

$$h(\omega) = \begin{cases} h_1(\omega)/h_2(\omega), & \omega \in A; \\ 0, & \omega \in A^c. \end{cases}$$

Hence, for $\omega \in A$, $f_{1,\theta}(\omega) = h(\omega) f_{2,\theta}(\omega)$, for all $\theta \in \Theta$. On the set A^c, both sides vanish under P_θ. Thus, Equation (2.1) holds $\forall \theta \in \Theta$ on a set of probability one under each P_θ. ■

Proof [Proposition 2.7] Using the notation of Theorem 2.6, let $f_{1,\theta} \in (\mathrm{d}P_\theta/\mathrm{d}\nu_1)_\nu$ and $f_{2,\theta} \in (\mathrm{d}P_\theta/\mathrm{d}\nu_2)_\nu$. Then we can write $f_{1,\theta}(\omega) = (\mathrm{d}P_\theta/\mathrm{d}\nu)^{(1)}(\omega)(\mathrm{d}\nu/\mathrm{d}\nu_1)^{(1)}(\omega)$ and $f_{2,\theta}(\omega) = (\mathrm{d}P_\theta/\mathrm{d}\nu)^{(2)}(\omega)(\mathrm{d}\nu/\mathrm{d}\nu_2)^{(2)}(\omega)$, where $(\cdot)^{(1)}$ and $(\cdot)^{(2)}$ denote particular chosen versions. For each $\theta \in \Theta$, define $A_\theta = \{\omega \in \Omega\colon (\mathrm{d}\nu/\mathrm{d}\nu_1)^{(1)}(\omega) > 0, (\mathrm{d}\nu/\mathrm{d}\nu_2)^{(2)}(\omega) > 0, (\mathrm{d}P_\theta/\mathrm{d}\nu)^{(1)}(\omega) = (\mathrm{d}P_\theta/\mathrm{d}\nu)^{(2)}(\omega)\}$. By the RN Theorem, each version is defined ν-almost everywhere, so $\nu(A_\theta^c) = 0$. Define $B^c = \bigcup_{\theta\in\Theta} A_\theta^c$ and $B = \bigcap_{\theta\in\Theta} A_\theta$. Then $\nu(B^c) = 0$, so $P_\theta(B^c) = 0$ for each θ. On B, we have $(\mathrm{d}P_\theta/\mathrm{d}\nu)^{(1)}(\omega) = (\mathrm{d}P_\theta/\mathrm{d}\nu)^{(2)}(\omega)$, $(\mathrm{d}\nu/\mathrm{d}\nu_1)^{(1)}(\omega) > 0$, $(\mathrm{d}\nu/\mathrm{d}\nu_2)^{(2)}(\omega) > 0$. Hence, we get $f_{1,\theta}(\omega) = (\mathrm{d}P_\theta/\mathrm{d}\nu)^{(1)}(\mathrm{d}\nu/\mathrm{d}\nu_1)^{(1)} = (\mathrm{d}P_\theta/\mathrm{d}\nu)^{(2)}(\mathrm{d}\nu/\mathrm{d}\nu_2)^{(2)} = f_{2,\theta}(\omega)$ up to a ν-null set. Therefore, we have that $f_{1,\theta} \propto_\theta f_{2,\theta}$ on B. ■

Proof [Proposition 2.8] Let $\mathcal{P}, \nu_1, \nu_2, \nu$ be as in Theorem 2.6, and let Q be the measure from Lemma 2.5. Suppose T is a sufficient statistic for $\mathcal{P}$ with range $(\mathcal{T}, \mathcal{B})$. Then,

(i) Fixing $\theta \in \Theta$, choose any version $g_\theta^* \in (\mathrm{d}P_\theta/\mathrm{d}Q)_\nu$ that is measurable with respect to $\sigma(T)$, and pick $h_1 \in (\mathrm{d}Q/\mathrm{d}\nu_1)_\nu$. By Shao (2003, Sec. 1.4, Lemma 1.2), there exists a $\mathcal{B}$-measurable g_θ such that $g_\theta^* = g_\theta \circ T$. Since T is sufficient, we have $g_\theta \circ T \in (\mathrm{d}P_\theta/\mathrm{d}Q)_\nu$ on $(\Omega, \mathcal{F})$ (Lehmann, 1986, Sec. 2.6, Theorem 8). Define $f_{1,\theta}(\omega) = g_\theta(T(\omega))h_1(\omega)$. By the chain rule, we have $f_{1,\theta} \in (\mathrm{d}P_\theta/\mathrm{d}\nu_1)_\nu$.

(ii) Similarly, if $f_{2,\theta}$ is obtained from the same g_θ^* but with $h_2 \in (\mathrm{d}Q/\mathrm{d}\nu_2)_\nu$, then $f_{2,\theta}(\omega) = g_\theta(T(\omega))h_2(\omega)$. Let $A = \{\omega \in \Omega\colon h_1(\omega) > 0\}$. Then $\nu(A^c) = 0$, so on A, $f_{1,\theta}$ and $f_{2,\theta}$ differ by the factor $h_1(\omega)/h_2(\omega)$. In particular, we get $f_{1,\theta} \propto_\theta f_{2,\theta}$ on A.

■

Proof [Lemma 2.10]

(i) Let ν be a minimal dominating measure for $\mathcal{P}$, and $Q = \sum_i c_i P_{\theta_i}$ the measure from Lemma 2.5. Define $A_i = \{\omega\colon \mathrm{d}P_{\theta_i}/\mathrm{d}\nu(\omega) > 0\}$, where $A = \cup_i A_i$. Then, we have $P_{\theta_i}(A_i) = 1$ for all i, so $P_{\theta_i}(A) = 1$ for all i. Consequently, $Q(A) = 1$, but $Q(A^c) = 0$ implies $P_\theta(A^c) = 0$ for all $\theta \in \Theta$. Hence, $P_\theta(A) = 1$ for all θ.

(ii) Let $B \subset A$ be such that $\nu_2(B) = 0$. Suppose $\nu_1(B) > 0$. Since $B = \cup_i(A_i \cap B)$, there is some i_0 with $\nu_1(A_{i_0} \cap B) > 0$. By the auxiliary result from Appendix A (the nonvanishing integral argument), and the fact $\mathrm{d}P_{\theta_{i_0}}/\mathrm{d}\nu > 0$ on A_{i_0}, we have $\int_{A_{i_0}\cap B} \mathrm{d}P_{\theta_{i_0}}/\mathrm{d}\nu \mathrm{d}\nu_1 > 0$. Define $C = \{\omega\colon \mathrm{d}\nu/\mathrm{d}\nu_1(\omega) > 0\}$. Then, we get $\nu_1(A_{i_0} \cap B \cap C) > 0$, and another application of the same lemma shows $P_{\theta_{i_0}}(A_{i_0} \cap B) = \int_{A_{i_0}\cap B} \mathrm{d}P_{\theta_{i_0}}/\mathrm{d}\nu \mathrm{d}\nu/\mathrm{d}\nu_1 \mathrm{d}\nu_1 = \int_{A_{i_0}\cap B\cap C} \mathrm{d}P_{\theta_{i_0}}/\mathrm{d}\nu \mathrm{d}\nu/\mathrm{d}\nu_1 \mathrm{d}\nu_1 > 0$, a contradiction since $\nu_2(B) = 0$ should imply $P_{\theta_{i_0}}(B) = 0$. Therefore, $\nu_1(B) = 0$.

■

Proof [Theorem 2.12] Recall that (A, ν) is a dominating pair for $(\mathcal{P}, \nu_1, \nu_2)$. By definition, $\nu_1|_A, \nu_2|_A, \nu|_A$ are all equivalent, and $\nu(A) = 1$. Write $\dot{P}_\theta, \dot{\nu}_1, \dot{\nu}_2, \dot{\nu}$ for the restrictions of $P_\theta, \nu_1, \nu_2, \nu$ to $(A, \mathcal{F}(A))$, respectively. In particular, for each θ_i used in constructing ν, suppose we have continuous versions $f_{1,\theta_i} \in (\mathrm{d}\dot{P}_{\theta_i}/\mathrm{d}\dot{\nu}_1)_{\dot{\nu}}$ and $f_{2,\theta_i} \in (\mathrm{d}\dot{P}_{\theta_i}/\mathrm{d}\dot{\nu}_2)_{\dot{\nu}}$. Let $h \in (\mathrm{d}\dot{\nu}_2/\mathrm{d}\dot{\nu}_1)_{\dot{\nu}}$. We wish to prove that such continuous versions imply a proportion relation $(f_{1,\theta}, f_{2,\theta})$ differs by a factor $h(\omega)$ not depending on θ.

The proof has three main steps. First, we define a set $S_h \subset A$ where proportionality holds for all θ_i that state the measure ν, ensuring that $\dot{\nu}(S_h) = 1$. Then, we establish the continuity of h on S_h. Third, we demonstrate that the proportionality of the likelihood functions holds, for all θ, in a subset $B_h \subset S_h$, where $P_\theta(B_h) = 1$ for all θ. The detailed steps are as follows:

Step 1 —Constructing a set where the ratio is fixed for basis measures. For each index $i \in \mathrm{N}$, define $A_i = \{\omega \in A\colon f_{1,\theta_i}(\omega) = h(\omega) f_{2,\theta_i}(\omega))$. By the RN chain rule and the fact that f_{1,θ_i} and f_{2,θ_i} are each versions of $\mathrm{d}\dot{P}_{\theta_i}/\mathrm{d}\dot{\nu}_1$ and $\mathrm{d}\dot{P}_{\theta_i}/\mathrm{d}\dot{\nu}_2$, respectively, it follows that $\dot{\nu}(A_i^c) = 0$ for every i. Next, let $B_i = \{\omega \in A\colon f_{2,\theta_i}(\omega) > 0\}$, $B = \cup_{i=1}^{+\infty} B_i$, $D_h = \cap_{i=1}^{+\infty} A_i$, $S_h = D_h \cap B$. Since $\dot{\nu}(A_i^c) = 0$ for all i, we have $\dot{\nu}(D_h^c) = 0$. Also, $\dot{\nu}(B) = 1$ because for each θ_i, the set where $f_{2,\theta_i} = 0$ is negligible with respect to $\dot{\nu}$. Hence, we have $\dot{\nu}(S_h) = 1$. By absolute continuity, $P_\theta(S_h) = 1$ for all $\theta \in \Theta$.

Step 2 —Showing h is continuous on S_h. Let $\omega_0 \in S_h$ and take any sequence $\{\omega_n\} \subset S_h$ such that $\omega_n \to \omega_0$ in A. We must show $h(\omega_n) \to h(\omega_0)$. By construction, $\omega_0 \in D_h$ so belonging to every A_i. Further, there is some i_0 such that $\omega_0 \in B_{i_0}$, meaning $f_{2,\theta_{i_0}}(\omega_0) > 0$. On B_{i_0}, $h(\omega) = f_{1,\theta_{i_0}}(\omega)/f_{2,\theta_{i_0}}(\omega)$. Since $f_{2,\theta_{i_0}}$ is continuous and positive on $B_{i_0} \subset A$, it follows that $B_{i_0} \cap S_h$ is open in the subspace S_h. Thus, for sufficiently large n, $\omega_n \in B_{i_0} \cap S_h$, and so $h(\omega_n) = f_{1,\theta_{i_0}}(\omega_n)/f_{2,\theta_{i_0}}(\omega_n)$. But both $f_{1,\theta_{i_0}}$ and $f_{2,\theta_{i_0}}$ are continuous (by hypothesis). Hence, the ratio converges to the ratio at ω_0. Therefore h is continuous at ω_0. Since $\omega_0 \in S_h$ was arbitrary, h is continuous on all of S_h.

Step 3 —Proving the proportionality for all θ. For each $\theta \in \Theta$, define $B_\theta = \{\omega \in S_h\colon f_{1,\theta}(\omega) = h(\omega) f_{2,\theta}(\omega)\}$. Again by the RN chain rule, we get $\dot{\nu}(S_h \cap B_\theta^c) = 0$. The function $\omega \mapsto (f_{1,\theta} - h f_{2,\theta})(\omega)$ is continuous on S_h. Thus, we obtain B_θ is closed in the subspace S_h. Hence, we reach $B_h = \cap_{\theta\in\Theta} B_\theta$ is also closed in S_h. Since S_h is a subspace of a separable metric space, there is a countable subset $\{\theta_j\} \subset \Theta$ such that $B_h = \cap_{j=1}^{+\infty} B_{\theta_j}$. Moreover, $\dot{\nu}(B_\theta^c \cap S_h) = 0$ for every θ, so $\dot{\nu}(B_h^c \cap S_h) = 0$, which implies $\nu_1(B_h^c \cap S_h) = 0$ and so $P_\theta(B_h) = 1$ for each θ. On $B_h \subseteq S_h$, we get $f_{1,\theta}(\omega) = h(\omega) f_{2,\theta}(\omega), \quad \forall \theta \in \Theta, \omega \in B_h$. Hence, the proportionality relation holds $\dot{\nu}$-almost everywhere (and so P_θ-almost surely) for all θ. ∎

Proof [Corollary 2.13] Since $f_{1,\theta}$ and $f_{2,\theta}$ are strictly positive on Ω for each $\theta \in \Theta$, it follows that all P_θ, ν_1, ν_2 are pairwise equivalent. By Proposition 6.6, the supports satisfy $S_\theta = S_{\nu_1} = S_{\nu_2} = \Omega, \quad \forall \theta \in \Theta$. For each θ, define $h_\theta(\omega) = f_{1,\theta}(\omega)/f_{2,\theta}(\omega)$, $\omega \in \Omega$. Note that $h_\theta \in (\mathrm{d}\nu_2/\mathrm{d}\nu_1)_\nu$, and is continuous on Ω. Since ν_1, ν_2 are locally finite (LF), Theorem 6.8 guarantees that all the h_θ coincide on Ω. Denote this common continuous function by $h(\omega)$. Hence, we attain at $f_{1,\theta}(\omega) = h(\omega) f_{2,\theta}(\omega), \quad \forall \omega \in \Omega, \forall \theta \in \Theta$, which establishes the claimed proportionality. ∎

Proof [Proposition 2.14] Since $S_\theta \subset S_3$ by Proposition 6.6, simply define $f_{1,\theta}(\omega) = h(\omega) f_{2,\theta}(\omega), \quad \forall \omega \in S_\theta, \forall \theta \in \Theta$, where h is as in the statement of the proposition. Theorem 6.8 then ensures uniqueness of the versions $f_{1,\theta}, f_{2,\theta}$, and h. ∎

Proof [Proposition 3.1] Let μ be any σ-finite measure such that $\mathcal{P} \ll \mu$. Write $g_\theta(x) = \mathrm{d}P_\theta/\mathrm{d}\mu(x)$ and define $m^*(x) = \int_\Theta g_\theta(x)\mathrm{d}R(\theta)$. Then construct a measure ξ on $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ via $\xi(A) = \int_A m^*(x)\mu(\mathrm{d}x)$, $A \in \mathcal{B}_\mathcal{X}$. We claim $\lambda = \xi$. Indeed, for each $A \in \mathcal{B}_\mathcal{X}$, we have that

$$\begin{aligned}
\lambda(A) &= \int_A m(x)\nu(\mathrm{d}x) = \int_A \int_\Theta f_\theta(x) dR(\theta)\mathrm{d}\nu(x) \quad \text{(definition of } \lambda) \\
&\overset{(i)}{=} \int_\Theta \int_A f_\theta(x)\mathrm{d}\nu(x)\mathrm{d}R(\theta) = \int_\Theta P_\theta(A)\mathrm{d}R(\theta) \quad \text{(Fubini theorem)} \\
&= \int_\Theta \int_A g_\theta(x)\mu(\mathrm{d}x)\mathrm{d}R(\theta) \overset{(ii)}{=} \int_A \int_\Theta g_\theta(x)\mathrm{d}R(\theta)\mu(\mathrm{d}x) = \int_A m^*(x)\mu(\mathrm{d}x) = \xi(A),
\end{aligned}$$

where steps (i) and (ii) each follow by the Fubini theorem. Hence, $\lambda \equiv \xi$ except possibly on a set of measure zero, showing λ is independent of which ν or μ we started with (up to null sets). ∎

Proof [Proposition 3.2] If $m(x) > 0$ for all $x \in \mathcal{X}$, then $\{x\colon m(x) = 0\} = \varnothing$, so λ places full mass on all of $\mathcal{X}$. Since m was the pointwise sum of f_θ, each P_θ must also be absolutely continuous with respect to λ. In other words, λ dominates $\mathcal{P}$. ■

Proof [Proposition 3.3] First, if λ dominates P_θ, then $\lambda(N) = 0$ is immediate. Conversely, suppose $P_\theta(N) = 0$. Take any $A \in \mathcal{B}_\mathcal{X}$ with $\lambda(A) = 0$. We must show $P_\theta(A) = 0$. Note that

$$0 = \lambda(A) = \lambda(A \cap N^c) = \int_{A \cap N^c} m(x)\nu(\mathrm{d}x), \tag{6.9}$$

where m is the integrand defining λ. Since m is strictly positive on $A \cap N^c$, Equation (6.9) forces $\nu(A \cap N^c) = 0$. Hence, we get $P_\theta(A \cap N^c) = 0$, but $P_\theta(A) = P_\theta(A \cap N^c)$ by hypothesis ($P_\theta(N) = 0$). Thus, we get $P_\theta(A) = 0$. ■

Proof [Proposition 3.4] Define $M = \{x \in \mathcal{X} : f_\theta(x) > 0\}$ and take any $A \in \mathcal{B}_\mathcal{X}$ such that $\lambda(A) = 0$. We need to show that $P_\theta(A) = 0$ for all $\theta \in \Theta$. Since $\mathcal{P} \ll \nu$ and $P_\theta(A) = P_\theta(A \cap M)$ for all $\theta \in \Theta$, it suffices to prove that $\nu(A \cap M) = 0$. Suppose, for contradiction, that $\nu(A \cap M) > 0$. Since f_θ is strictly positive on $A \cap M$, we obtain

$$P_\theta(A) = P_\theta(A \cap M) = \int_{A \cap M} f_\theta(x)\nu(\mathrm{d}x) > 0, \quad \forall \theta \in \Theta. \tag{6.10}$$

On the other hand, applying Fubini's theorem, we get

$$\lambda(A) = \int_A m(x)\nu(\mathrm{d}x) = \int_A \int_\Theta f_\theta(x)\mathrm{d}R(\theta)\nu(\mathrm{d}x) = \int_\Theta P_\theta(A)\mathrm{d}R(\theta) = \int_\Theta P_\theta(A \cap M)\mathrm{d}R(\theta). \tag{6.11}$$

Since $R(\Theta) > 0$, it follows from Equations (6.10) and (6.11) that $\lambda(A) > 0$, contradicting our assumption that $\lambda(A) = 0$. Thus, we conclude that $\nu(A \cap M) = 0$, completing the proof. ■

Proof [Proposition 3.5] Suppose $\mathrm{d}P_\theta/\mathrm{d}\nu(\omega)$ is given by the exponential-family representation given in Equation (3.3) in the main text: $\mathrm{d}P_\theta/\mathrm{d}\nu(\omega) = \exp((\eta(\theta))^\tau T(\omega) - \xi(\theta))h_\nu(\omega)$. Consider the measure Q from Lemma 2.5, and let $q \in (\mathrm{d}Q/\mathrm{d}\nu)$. Since Q is minimal, $Q \ll \nu$. Without loss of generality, assume $q(\omega) > 0$ ν-almost everywhere. Define, for each $\theta \in \Theta$,

$$b_\theta(\omega) = \exp((\eta(\theta))^\tau T(\omega) - \xi(\theta))m(\omega), \quad \omega \in \Omega, \tag{6.12}$$

where $m(\omega) = h_\nu(\omega)/q(\omega)$. By the RN chain rule, we have

$$\exp((\eta(\theta))^\tau T(\omega) - \xi(\theta))h_\nu(\omega) = \mathrm{d}P_\theta/\mathrm{d}Q(\omega)q(\omega), \quad \nu\text{-almost everywhere.} \tag{6.13}$$

Combining Equations (6.12) and (6.13) shows $b_\theta(\omega) \equiv \mathrm{d}P_\theta/\mathrm{d}Q(\omega)\nu$-almost everywhere Hence, we have that $b_\theta \in (\mathrm{d}P_\theta/\mathrm{d}Q)$. Next, let μ be another σ-finite measure such that $\mathcal{P} \ll \mu$, and suppose $\mu \neq \nu$. By the minimality of Q, we have $Q \ll \mu$. Choose any version $s \in (\mathrm{d}Q/\mathrm{d}\mu)$. Define, for each θ, $p_\theta(\omega) = \exp((\eta(\theta))^\tau T(\omega) - \xi(\theta))h_\mu(\omega)$, for $\omega \in \Omega$, where $h_\mu(\omega) = m(\omega)s(\omega)$. By the RN chain rule again, it follows $p_\theta \in (\mathrm{d}P_\theta/\mathrm{d}\mu)$. This shows that any other dominating measure μ yields an exponential-family form with the same sufficient statistic T and natural parameter η (up to an adjusted normalizing function $\xi(\theta)$ and a modified base function h_μ). Thus, the exponential-family representation does not depend on the specific measure ν. ■

Proof [Theorem 6.8] By Proposition 6.6, we have $S_\mu \subseteq S_\nu$. Restrict μ and ν to S_μ. Applying the uniqueness theorem for the space (since now S_μ is the new universe and $\mu|_{S_\mu} \ll \nu|_{S_\mu}$), we conclude any continuous version of $\mathrm{d}\mu/\mathrm{d}\nu$ on S_μ is unique, up to measure-zero sets. ■

Statements

Acknowledgements

The authors thank Professor Gareth Roberts for insightful and stimulating discussions about this work. The authors also thank the referees and the editors and reviewers for comments that helped to improve the article.

Author contributions

Conceptualization: F.B.G., P.F.; data curation: F.B.G., P.F.; formal analysis: F.B.G., P.F.; investigation: F.B.G., P.F.; methodology: F.B.G., P.F.; writing —original draft: F.B.G., P.F.; writing —review and editing: F.B.G., P.F. All authors have read and agreed to the published version of the article.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Funding

Flávio Gonçalves is funded by FAPEMIG grant project APQ-01837-22, and CNPq grant project 308536/2023-1.

References

Andersen, P.K., Borgan, O., Gill, R.D., and Keiding, N., 1993. Statistical Models Based on Counting Processes. Springer, New York, NY, US.

Barndorff-Nielsen, O., Hoffmann-Jørgensen, J., and Pedersen, K., 1976. On the minimal sufficiency of the likelihood function. Scandinavian Journal of Statistics, 3, 115–127.

Berger, J.O. and Wolpert, R.L., 1988. The Likelihood Principle. Institute of Mathematical Statistics, Hayward, CA, US.

Bhattacharjee, S. and Chakraborty, S., 2023. Inflated log-Lindley distribution for modeling continuous data bounded in unit interval with possible mass at boundaries. Chilean Journal of Statistics, 14(2), 123–142.

Beskos, A., Papaspiliopoulos, O., Roberts, G.O., and Fearnhead, P., 2006. Exact and computationally efficient likelihood-based inference for discretely observed diffusion processes. Journal of the Royal Statistical Society B, 68, 333–382.

Fisher, R.A., 1921. On the "probable error" of a coefficient of correlation deduced from a small sample. Metron, 1, 3–32.

Fraser, D.A.S., McDunnough, P., Naderi, A., and Plante, A., 1997. From the likelihood map to Euclidean minimal sufficiency. Journal of Probability and Mathematical Statistics, 17, 223–230.

Fraser, D.A.S. and Naderi, A., 1996. On the Definition of Conditional Probability. In Brunner, E., and Denker, M. (eds.) Research Developments in Probability and Statistics. Springer, Utrecht, pp. 23–26.

Fraser, D.A.S. and Naderi, A., 2007. Minimal sufficient statistics emerge from the observed likelihood functions. International Journal of Statistical Sciences, 6, 55–61.

Gonçalves, F.B. and Gamerman, D., 2018. Exact Bayesian inference in spatio-temporal Cox processes driven by multivariate Gaussian processes. Journal of the Royal Statistical Society B, 80, 157–175.

Gonçalves, F.B., Venturelli, S.L.J., and Loschi, R.H., 2022. Flexible Bayesian modelling in dichotomous item response theory using mixtures of skewed item curves. British Journal of Mathematical and Statistical Psychology, 76, 69–86.

Gonçalves, F.B. and Franklin, P., 2023. Code and data repository for "Measure-theoretic definition of likelihood and related results". Available at https://github.com/fbambirra/likelihood_paper.

Gottardo, R. and Raftery, A.E., 2009. Markov chain Monte Carlo with mixtures of mutually singular distributions. Journal of Computational and Graphical Statistics, 17, 949–975.

Halmos, P.R. and Savage, L.J., 1949. Application of the Radon-Nikodym theorem to the theory of sufficient statistics. The Annals of Mathematical Statistics, 20, 225–241.

Izbicki, R., Lee, A.B., and Schafer, C.M., 2014. High-dimensional density ratio estimation with extensions to approximate likelihood computation. Proceedings of the 17th International Conference on Artificial Intelligence and Statistics, Reykjavic, Iceland, p. 33.

Jorgensen, B. and Labouriau, R., 2012. Exponential Families and Theoretical Inference. Lecture Notes, Department of Statistics, University of British Columbia, Vancouver, BC, Canada.

Jarrett, R.G., 1979. A note on the intervals between coal-mining disasters. Biometrika, 66, 191–193.

Kingman, J.F.C., 1993. Poisson Processes. Oxford University Press, New York, NY, US.

Kloeden, P. and Platen, E., 1995. Numerical Solution of Stochastic Differential Equations. Springer, New York, NY, US.

Kolmogorov, A., 1933. Grundbegriffe der Wahrscheinlichkeitsrechnung (in German). Springer, Berlin, Germany.

Lehmann, E.L., 1986. Testing Statistical Hypotheses. Springer, New York, NY, US.

Lindley, D.V., 1953. Statistical Inference. Journal of the Royal Statistical Society B, 15, 131–179.

Nikodym, O., 1930. Sur une généralisation des intégrales de M. J. Radon. Fundamenta Mathematicae (in French), 15, 131–179.

Piccioni, M., 1982. On the definition of likelihood in abstract spaces. Journal of the Franklin Institute, 313, 1–15.

Piccioni, M., 1983. Continuous versions of Radon-Nikodym derivatives as likelihood ratios. Systems and Control Letters, 2, 369–374.

R Core Team, 2024. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.r-project.org/.

Reid, N., 2013. Likelihood Formalities. STA3000 Lecture notes. Available at: http://www.utstat.utoronto.ca/reid/sta3000y/likelihood-formal.pdf.

Roberts, G.O. and Stramer, O., 2001. On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. Biometrika, 88, 603–621.

Shao, J., 2003. Mathematical Statistics. Springer, New York, NY, US.

Schmidt, A.M., Moraes, C.P., and Migon, H.S., 2017. Hierarchical mixture beta dynamic model of school performance in the Brazilian mathematical olympics for public schools. Chilean Journal of Statistics, 8(1), 3–24.

San Martín, E. and González, J., 2010. Bayesian identifiability: Contributions to an inconclusive debate. Chilean Journal of Statistics, 1(2), 69–91.

Ruggeri, F., 2010. Nonparametric Bayesian robustness. Chilean Journal of Statistics, 1(2), 51–68.

Disclaimer/Publisher's Note: The views, opinions, data, and information presented in all our publications are exclusively those of the individual authors and contributors, and do not reflect the positions of our journal or its editors. Our journal and editors do not assume any liability for harm to people or property resulting from the use of ideas, methods, instructions, or products mentioned in the content.