A detailed treatment of Doob's theorem

Jeffrey W. Miller

Harvard University, Department of Biostatistics

January 9, 2018

Abstract

Doob's theorem provides guarantees of consistent estimation and posterior consistency under very general conditions. Despite the limitation that it only guarantees consistency on a set with prior probability 1, for many models arising in practice, Doob's theorem is an easy way of showing that consistency will hold almost everywhere. In this article, we give a detailed proof of Doob's theorem.

1   Introduction

Doob (1949) established a remarkable theorem on consistency in the Bayesian setting. Roughly, Doob's theorem says that with probability 1, if the true parameter $\theta_0 \in \Theta$ is drawn from the prior Π, and $X_1, \ldots, X_n$ i.i.d. $\sim P_{\theta_0}$ for an identifiable model $\{P_\theta : \theta \in \Theta\}$, then

1. the posterior mean of $g(\theta)$ converges to $g(\theta_0)$ for any $g \in L^1(\Pi)$, and

2. the posterior distribution of θ is consistent, i.e., the posterior concentrates in neighborhoods of the true parameter θ_0,

as $n \to \infty$, under very general conditions. The theorem is stunning in its generality and elegance. The heart of the proof is a direct application of martingale convergence, however, the bulk of the proof involves showing an essential and nontrivial measurability result: that θ_0 is a *measurable* function of $(X_1, X_2, \ldots)$, i.e., that there is a measurable function f such that with probability 1, $\theta_0 = f(X_1, X_2, \ldots)$. In other words, given infinite data, the true parameter can be recovered in a measurable way.

Doob's theorem has been criticized on the grounds that it only guarantees consistency on a set with prior probability 1, and thus, if for example one chose a prior with all of its mass on a single point θ^*, then the theorem would apply but consistency would fail everywhere except at θ^* (Barron et al., 1999). While it is true that the theorem is not very useful for a poorly chosen prior, this misses the point, which is that for a well-chosen prior, consistency can be guaranteed on a very large set. For example, if Θ is a countable union of finite-dimensional spaces, then one can choose a prior with a density with respect to Lebesgue measure that is positive on all of Θ, and Doob's theorem can be used to guarantee consistency everywhere except on a set of Lebesgue measure zero, assuming identifiability. This approach was used by Nobile (1994) to prove consistency for finite mixture models with a prior on the number of components, under very general conditions.

However, this approach has the weakness that one cannot say for any given point whether consistency will hold, and in this respect the criticism of Doob's theorem is valid. Further, Doob's theorem is less useful when θ is infinite-dimensional, since the set on which consistency may fail can be very large, even for a seemingly well-chosen prior (Freedman, 1963). These limitations were the impetus for the beginnings of the modern approach to studying the consistency of nonparametric Bayesian procedures (Schwartz, 1965). Nonetheless, for many models arising in practice, Doob's theorem is an extremely easy method of guaranteeing that consistency will hold almost everywhere.

The original justification given by Doob (1949) was hardly even a sketch, providing only the outline of a proof. More complete proofs of Doob's theorem have been provided for special cases: Van der Vaart (1998) (Theorem 10.10) assumes Θ is a Euclidean space and proves the part of the theorem regarding posterior concentration. Ghosh and Ramamoorthi (2003) (Theorem 1.3.2) prove the posterior concentration part of the theorem, but they leave many details to the reader. Ghosal and van der Vaart (2017) (Theorem 6.9, 6.10) provide a proof for the general case, but their treatment is brief and many details are left to the reader.

Lijoi et al. (2004) prove an interesting result related to Doob's theorem in the context of posterior inference for a density function f_0 given i.i.d. samples $X_1, X_2, \ldots \sim f_0$. They show that if the prior has full Hellinger support then the posterior concentrates in Hellinger neighborhoods of an "essentially unique" random density $\tilde{g}$ with the property that the data are conditionally i.i.d. given $\tilde{g}$.

The purpose of this note is to provide a complete and detailed proof of Doob's theorem in full generality. The rest of the paper is organized as follows. In Section 2, we state the main results, and Section 3 contains the proofs of the main results. In Section 4, we prove the key measurability result that is used in the proofs. In Section 5, for completeness, we prove some supporting results.

2   Main results

Let $\widehat{\mathcal{X}}$ and $\widehat{\Theta}$ be complete separable metric spaces. Let $\mathcal{X} \subseteq \widehat{\mathcal{X}}$ and $\Theta \subseteq \widehat{\Theta}$ be Borel measurable subsets, and let $\mathcal{A}$ and $\mathcal{B}$ be the induced Borel σ-algebras of $\mathcal{X}$ and Θ, respectively. For each $\theta \in \Theta$, let P_θ be a probability measure on $(\mathcal{X}, \mathcal{A})$. Let Π be a probability measure on $(\Theta, \mathcal{B})$.

Assumption 2.1. *Assume*

  1. *$\theta \mapsto P_\theta(A)$ is measurable for every $A \in \mathcal{A}$, and*

  2. *$\theta \neq \theta' \Rightarrow P_\theta \neq P_{\theta'}$.*

In other words, condition (1) is that $\{P_\theta : \theta \in \Theta\}$ is a measurable family, and condition (2) is that θ is identifiable: if $\theta, \theta' \in \Theta$ such that $\theta \neq \theta'$ then there exists $A \in \mathcal{A}$ such that $P_\theta(A) \neq P_{\theta'}(A)$.

As indicated in the introduction, Doob's theorem has two parts: the first concerning consistent estimation (Theorem 2.2), and the second concerning posterior consistency (Theorem 2.4). A word on notation: we use θ for fixed values in Θ, and $\boldsymbol{\theta}$ for a random variable taking values in Θ. Let μ be the joint distribution of $((X_1, X_2, \ldots), \boldsymbol{\theta})$ defined by letting $\boldsymbol{\theta} \sim \Pi$ and $X_1, X_2, \ldots | \boldsymbol{\theta}$ i.i.d. $\sim P_{\boldsymbol{\theta}}$.

Theorem 2.2 (Doob's theorem for estimators). *Let $(X, \boldsymbol{\theta}) \sim \mu$, where $X = (X_1, X_2, \ldots)$. If $g : \Theta \to \mathbb{R}$ is a measurable function such that $\mathbb{E}|g(\boldsymbol{\theta})| < \infty$, and Assumption 2.1 holds, then*

$$\lim_{n \to \infty} \mathbb{E}(g(\boldsymbol{\theta}) \mid X_1, \ldots, X_n) = g(\boldsymbol{\theta}) \ \ a.s.[\mu].$$

Remark. The generalization to vector-valued $g : \Theta \to \mathbb{R}^k$ follows by applying the theorem to each coordinate separately. Also, note that the condition on g is equivalent to requiring $g \in L^1(\Pi)$.

The following corollary is equivalent to Theorem 2.2, but is presented from a frequentist perspective.

Corollary 2.3. *Suppose $g \in L^1(\Pi)$ and Assumption 2.1 holds. There exists $\Theta_0 \subseteq \Theta$ with $\Pi(\Theta_0) = 1$ such that for all $\theta_0 \in \Theta_0$, if $X_1, X_2, \ldots \sim P_{\theta_0}$ i.i.d. then*

$$\lim_{n \to \infty} \mathbb{E}(g(\boldsymbol{\theta}) \mid X_1, \ldots, X_n) = g(\theta_0) \ \ a.s.[P_{\theta_0}]$$

where the conditional expectation is computed under μ.

Theorem 2.4 (Doob's theorem for posterior consistency). *If Assumption 2.1 holds, then there exists $\Theta_1 \subseteq \Theta$ with $\Pi(\Theta_1) = 1$ such that the posterior is consistent at every $\theta_1 \in \Theta_1$. That is, for all $\theta_1 \in \Theta_1$, if $X_1, X_2, \ldots \sim P_{\theta_1}$ i.i.d. then for any neighborhood B of θ_1, we have*

$$\lim_{n \to \infty} \mathbb{P}(\boldsymbol{\theta} \in B \mid X_1, \ldots, X_n) = 1 \ \ a.s.[P_{\theta_1}]$$

where $\mathbb{P}(\boldsymbol{\theta} \in B \mid X_1, \ldots, X_n) = \mathbb{E}(I_B(\boldsymbol{\theta}) \mid X_1, \ldots, X_n)$ is computed under μ.

3 Proofs of the main results

First, we make some clarifying remarks and initial observations. Let $\widehat{\mathcal{X}}^\infty = \widehat{\mathcal{X}} \times \widehat{\mathcal{X}} \times \cdots = \{(x_1, x_2, \ldots) : x_1, x_2, \ldots \in \widehat{\mathcal{X}}\}$. Then $\widehat{\mathcal{X}}^\infty$ can be given a metric under which it is complete (as a metric space) and has the product topology. (For instance, if d is the metric on $\widehat{\mathcal{X}}$, then it can be verified that $D(x, y) = \sup_k \min\{1, d(x_k, y_k)\}/k$ is a metric on $\widehat{\mathcal{X}}^\infty$ satisfying these properties.) It follows that $\widehat{\mathcal{X}}^\infty$ is separable. Therefore, $\mathcal{X}^\infty = \mathcal{X} \times \mathcal{X} \times \cdots$ is a Borel measurable subset of a complete separable metric space, $\widehat{\mathcal{X}}^\infty$. Further, the induced Borel σ-algebra on $\mathcal{X}^\infty$ coincides with the product σ-algebra $\mathcal{A}^\infty$, which is generated by sets of the form

$$A_1 \times \cdots \times A_k \times \mathcal{X} \times \mathcal{X} \times \cdots = \{x \in \mathcal{X}^\infty : x_1 \in A_1, \ldots, x_k \in A_k\}$$

where $A_1, \ldots, A_k \in \mathcal{A}$ and $k \in \{1, 2, \ldots\}$. For each $\theta \in \Theta$, let P_θ^∞ be the product measure on $(\mathcal{X}^\infty, \mathcal{A}^\infty)$ obtained from P_θ — that is, by the Kolmogorov extension theorem, let P_θ^∞ be the unique probability measure on $(\mathcal{X}^\infty, \mathcal{A}^\infty)$ such that

$$P_\theta^\infty(A_1 \times \cdots \times A_k \times \mathcal{X} \times \mathcal{X} \times \cdots) = P_\theta(A_1) \cdots P_\theta(A_k)$$

for all $A_1, \ldots, A_k \in \mathcal{A}$ and all $k \in \{1, 2, \ldots\}$.

The joint distribution μ, defined in Section 2, is equal to the probability measure on $(\mathcal{X}^\infty \times \Theta, \mathcal{A}^\infty \otimes \mathcal{B})$ defined by

$$\mu(E) = \int_\Theta P_\theta^\infty(E_\theta)\, d\Pi(\theta)$$

for $E \in \mathcal{A}^\infty \otimes \mathcal{B}$, where $E_\theta = \{x \in \mathcal{X}^\infty : (x, \theta) \in E\}$ is the θ-section of E. In order for this definition to make sense, we need $\theta \mapsto P_\theta^\infty(E_\theta)$ to be a measurable function for all $E \in \mathcal{A}^\infty \otimes \mathcal{B}$; in Section 5, we prove that this is indeed the case, as long as condition (1) of Assumption 2.1 holds.

Similarly, giving $\widehat{\mathcal{X}}^\infty \times \widehat{\Theta}$ a metric that makes it a complete separable metric space with the product topology, it follows that $\mathcal{X}^\infty \times \Theta$ is a Borel measurable subset of $\widehat{\mathcal{X}}^\infty \times \widehat{\Theta}$, and the induced Borel σ-algebra on $\mathcal{X}^\infty \times \Theta$ coincides with the product σ-algebra $\mathcal{A}^\infty \otimes \mathcal{B}$.

Proof of Theorem 2.2. Since $\mathbb{E}|g(\boldsymbol{\theta})| < \infty$, $\mathbb{E}(g(\boldsymbol{\theta}) \mid X_1, \ldots, X_n)$ is a uniformly integrable martingale, and thus

$$\lim_{n \to \infty} \mathbb{E}(g(\boldsymbol{\theta}) \mid X_1, \ldots, X_n) = \mathbb{E}(g(\boldsymbol{\theta}) \mid X) \quad \text{a.s.}$$

by, e.g., Kallenberg (2002), 7.23. In Section 4, Theorem 4.1, we prove that there exists $f : \mathcal{X}^\infty \to \Theta$ measurable such that $f(x) = \theta$ for μ-almost all (x, θ). Defining $h = g \circ f$, we have $h : \mathcal{X}^\infty \to \mathbb{R}$ measurable such that $h(x) = g(\theta)$ a.e.$[\mu]$, and hence $h(X) = g(\boldsymbol{\theta})$ almost surely. Note that $\sigma(X) = \{A \times \Theta : A \in \mathcal{A}^\infty\} \subseteq \mathcal{A}^\infty \otimes \mathcal{B}$. Now, $h(X)$ is a version of $\mathbb{E}(g(\boldsymbol{\theta})|X)$, since it is $\sigma(X)$-measurable and for any $F \in \sigma(X)$ we have

$$\int_F g(\theta) d\mu(x, \theta) = \int_F h(x) d\mu(x, \theta).$$

Hence, $\mathbb{E}(g(\boldsymbol{\theta})|X) = h(X) = g(\boldsymbol{\theta})$ almost surely. $\qquad\square$

A minor technical point regarding the preceding proof: even though $\mathbb{E}(g(\boldsymbol{\theta})|X) = g(\boldsymbol{\theta})$ almost surely, $g(\boldsymbol{\theta})$ is not necessarily a version of $\mathbb{E}(g(\boldsymbol{\theta})|X)$ since $g(\boldsymbol{\theta})$ is not necessarily $\sigma(X)$-measurable. It is true that $h(x) = g(\theta)$ a.e.$[\mu]$ and h is Borel measurable, however, this is not sufficient to guarantee $\sigma(X)$-measurability of $g(\boldsymbol{\theta})$ since $(\mathcal{X}^\infty \times \Theta, \mathcal{A}^\infty \otimes \mathcal{B}, \mu)$ is not necessarily a complete measure space. When working in a probability measure space that is not necessarily complete, we follow the usual custom

of using "almost everywhere" (and "almost surely" for random variables) to signify that there exists a measurable set of measure 1 on which the condition under consideration holds.

Proof of Corollary 2.3. By Theorem 2.2, there exists a set $F \in \mathcal{A}^\infty \otimes \mathcal{B}$ of μ-probability 1 such that for all $(x, \theta) \in F$,

$$\lim_{n \to \infty} \mathbb{E}(g(\boldsymbol{\theta}) \mid x_1, \ldots, x_n) = g(\theta).$$

Then $1 = \mu(F) = \int_\Theta P_\theta^\infty(F_\theta) \, d\Pi(\theta)$, and therefore $\int_\Theta (1 - P_\theta^\infty(F_\theta)) d\Pi(\theta) = 0$. Since $1 - P_\theta^\infty(F_\theta) \geq 0$ for all θ, then $P_\theta^\infty(F_\theta) = 1$ a.e.[Π] by, e.g., Kallenberg (2002), 1.24. In other words, there exists $\Theta_0 \in \mathcal{B}$ with $\Pi(\Theta_0) = 1$ such that $P_\theta^\infty(F_\theta) = 1$ for all $\theta \in \Theta_0$. □

With this corollary in hand, we are well-positioned to prove the posterior consistency part of Doob's theorem.

Proof of Theorem 2.4. Let $\mathcal{C} \subseteq \mathcal{B}$ be a countable base for the topology of Θ, using the fact that Θ is a separable metric space. (For example, take the open balls of radius $1/k$ about every point in a countable dense subset, for all $k \in \{1, 2, \ldots \}$.) For each $C \in \mathcal{C}$, choose Θ_C according to Corollary 2.3 with $g(\theta) = I_C(\theta)$, so that $\Pi(\Theta_C) = 1$ and for each $\theta \in \Theta_C$,

$$\mathbb{E}(I_C(\boldsymbol{\theta}) \mid X_1, \ldots, X_n) \xrightarrow[n \to \infty]{} I_C(\theta) \quad \text{a.s.}$$

when $X_1, X_2, \ldots$ i.i.d. $\sim P_\theta$. Let

$$\Theta_1 = \bigcap_{C \in \mathcal{C}} \Theta_C.$$

Then $\Pi(\Theta_1) = 1$ since $\mathcal{C}$ is countable. Let $\theta_1 \in \Theta_1$ and let B be a neighborhood of θ_1. Since $\mathcal{C}$ is a base, there exists $C \in \mathcal{C}$ such that $\theta_1 \in C \subseteq B$. Therefore, if $X_1, X_2, \ldots \sim P_{\theta_1}$ i.i.d. then

$$\mathbb{E}(I_B(\boldsymbol{\theta}) \mid X_1, \ldots, X_n) \geq \mathbb{E}(I_C(\boldsymbol{\theta}) \mid X_1, \ldots, X_n) \xrightarrow[n \to \infty]{} I_C(\theta_1) = 1 \quad \text{a.s.}$$

since $I_B(\theta) \geq I_C(\theta)$ and $\theta_1 \in \Theta_1 \subseteq \Theta_C$; hence, $\mathbb{P}(\boldsymbol{\theta} \in B \mid X_1, \ldots, X_n) \to 1$ almost surely. □

4   Measurable recovery

In the proof of Theorem 2.2, we postponed a key step: the existence of f. Although it may at first appear to be a simple technical detail, this turns out to be surprisingly nontrivial — or at least, it seems to necessitate the use of a highly nontrivial result (Theorem 4.2).

Our proof below is modeled after Ghosh and Ramamoorthi (2003), who provided a proof of the posterior consistency part of Doob's theorem.

Theorem 4.1 (Measurable recovery)**.** *Under the setup of Sections 2 and 3, there exists $f : \mathcal{X}^\infty \to \Theta$ measurable such that $f(x) = \theta$ a.e.$[\mu]$.*

In other words, there exists $f : \mathcal{X}^\infty \to \Theta$ measurable such that there is some $F \in \mathcal{A}^\infty \otimes \mathcal{B}$ with $\mu(F) = 1$ satisfying $f(x) = \theta$ for all $(x, \theta) \in F$.

The difficult part of the theorem is ensuring that f is measurable. Otherwise, it is relatively simple, since if $X_1, X_2, \ldots \sim P_\theta$ i.i.d., then from the infinite sequence $X = (X_1, X_2, \ldots)$ one can almost surely recover P_θ (using the strong law of large numbers and separability of $\mathcal{X}$), and from P_θ one can recover θ (by identifiability, condition (2) of Assumption 2.1). This would determine a map from X to θ, but it comes with no guarantee of measurability. Although measurability is often a technical detail that can be assumed away, one should bear in mind that when dealing with conditional expectations, measurability is of the essence.

The following innocent-looking but deep theorem plays a central role in the proof.

Theorem 4.2 (Lusin, Souslin, Kuratowski)**.** *Let $\mathcal{U}$ and $\mathcal{V}$ be Borel measurable subsets of complete separable metric spaces, and give them the induced Borel σ-algebras. If $E \subseteq \mathcal{U}$ is measurable, and $g : \mathcal{U} \to \mathcal{V}$ is a measurable function such that the restriction $g|_E$ is one-to-one, then $g(E)$ is measurable and $g|_E^{-1} : g(E) \to \mathcal{U}$ is measurable.*

Proof. See Kechris (1995), Corollary 15.2, or Parthasarathy (2005), Corollary 3.3, for the proof. □

Proof of Theorem 4.1. Let $\mathcal{C} \subseteq \mathcal{A}$ be a countable collection that generates $\mathcal{A}$, contains $\mathcal{X}$, and is closed under finite intersections. (For example, let $\mathcal{C}_0$

be a countable base for $\mathcal{X}$, and let $\mathcal{C} = \{\mathcal{X}\} \cup \{A_1 \cap \cdots \cap A_k : A_1, \ldots, A_k \in \mathcal{C}_0, \ k = 1, 2, \ldots\}$.) Let

$$E = \bigcap_{A \in \mathcal{C}} F^A$$

where

$$F^A = \left\{ (x, \theta) \in \mathcal{X}^\infty \times \Theta : \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n I_A(x_i) \text{ exists and equals } P_\theta(A) \right\}.$$

We will abuse notation slightly and write π_x and π_θ for the projections from $\mathcal{X}^\infty \times \Theta$ to $\mathcal{X}^\infty$ and Θ respectively, i.e., $\pi_x(x, \theta) = x$ and $\pi_\theta(x, \theta) = \theta$. Note that π_x and π_θ are measurable (by the definition of the product σ-algebra). We will show that:

(a) E is measurable (i.e. $E \in \mathcal{A}^\infty \otimes \mathcal{B}$),

(b) $\mu(E) = 1$, and

(c) for any $x \in \pi_x(E)$, there is a unique $\theta \in \Theta$ such that $(x, \theta) \in E$.

Before proving (a), (b), and (c), let's see how to prove the result, assuming they are true. We will apply Theorem 4.2 with $\mathcal{U} = \mathcal{X}^\infty \times \Theta$, $\mathcal{V} = \mathcal{X}^\infty$, and $g = \pi_x$. Recall that $\mathcal{X}^\infty$ and $\mathcal{X}^\infty \times \Theta$ are Borel measurable subsets of complete separable metric spaces (in their respective Borel σ-algebras), and note that (c) means that the restriction $\pi_x|_E$ is one-to-one. Therefore, assuming (a) and (c), Theorem 4.2 implies that $\pi_x(E)$ is measurable and the map $\varphi : \pi_x(E) \to \mathcal{X}^\infty \times \Theta$ defined by $\varphi = \pi_x|_E^{-1}$ is measurable. Note that for any $(x, \theta) \in E$, we have $\varphi(x) = (x, \theta)$ by construction. It is straightforward to show that φ has a measurable extension $\tilde{\varphi} : \mathcal{X}^\infty \to \mathcal{X}^\infty \times \Theta$. Define $f : \mathcal{X}^\infty \to \Theta$ by $f(x) = \pi_\theta(\tilde{\varphi}(x))$. Then f is measurable and for any $(x, \theta) \in E$ we have $f(x) = \pi_\theta(\tilde{\varphi}(x)) = \pi_\theta(x, \theta) = \theta$. Along with (b), this will prove Theorem 4.1.

Now, we prove (a), (b), and (c).

(a) Let $A \in \mathcal{A}$. Since π_θ is measurable and $\theta \mapsto P_\theta(A)$ is measurable (by condition (1) in Assumption 2.1), then $(x, \theta) \mapsto P_\theta(A)$ is measurable. Similarly, for any $i \in \{1, 2, \ldots\}$, since π_x is measurable and $x \mapsto x_i$ is measurable then $(x, \theta) \mapsto I_A(x_i)$ is measurable. Hence, the function

$$h_n(x, \theta) := \frac{1}{n} \sum_{i=1}^n I_A(x_i) - P_\theta(A)$$

8

is measurable for any $n \in \{1, 2, \ldots\}$. In terms of h_n, the set F^A defined above becomes

$$F^A = \left\{ (x, \theta) \in \mathcal{X}^\infty \times \Theta : \lim_{n \to \infty} h_n(x, \theta) \text{ exists and equals } 0 \right\}.$$

It is straightforward to check that F^A is measurable, and therefore $E = \bigcap_{A \in \mathcal{C}} F^A$ is measurable.

(b) For $\theta \in \Theta$ and $A \in \mathcal{C}$, the sections $E_\theta = \{x \in \mathcal{X}^\infty : (x, \theta) \in E\}$ and $F_\theta^A = \{x \in \mathcal{X}^\infty : (x, \theta) \in F^A\}$ are measurable, since from the proof of (a) we know that E and F^A are measurable. Noting that

$$F_\theta^A = \left\{ x \in \mathcal{X}^\infty : \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n I_A(x_i) \text{ exists and equals } P_\theta(A) \right\},$$

the strong law of large numbers implies that for any $\theta \in \Theta$ and any $A \in \mathcal{C}$, we have $P_\theta^\infty(F_\theta^A) = 1$. Hence, $P_\theta^\infty(E_\theta) = 1$ for any $\theta \in \Theta$, since $E_\theta = \bigcap_{A \in \mathcal{C}} F_\theta^A$ (which follows from $E = \bigcap_{A \in \mathcal{C}} F^A$) and $\mathcal{C}$ is countable. Therefore,

$$\mu(E) = \int_\Theta P_\theta^\infty(E_\theta) \, d\Pi(\theta) = \int_\Theta d\Pi(\theta) = \Pi(\Theta) = 1.$$

(c) First, we show that $\theta \neq \theta' \Rightarrow E_\theta \cap E_{\theta'} = \varnothing$. Suppose $\theta, \theta' \in \Theta$ such that $\theta \neq \theta'$. Then $P_\theta \neq P_{\theta'}$ by identifiability (condition (2) of Assumption 2.1). This implies that there exists $A \in \mathcal{C}$ such that $P_\theta(A) \neq P_{\theta'}(A)$ (by, e.g., Kallenberg (2002), 1.17) since $\mathcal{C}$ is a π-system containing $\mathcal{X}$ that generates $\mathcal{A}$. Therefore, $F_\theta^A \cap F_{\theta'}^A = \varnothing$, and hence $E_\theta \cap E_{\theta'} = \varnothing$.

Now, to prove (c), suppose $x \in \pi_x(E)$. Let $\theta, \theta' \in \Theta$ such that $(x, \theta), (x, \theta') \in E$. (There is at least one such θ by the definition of $\pi_x(E)$.) Then $x \in E_\theta$ and $x \in E_{\theta'}$, and therefore $\theta = \theta'$ (by the preceding paragraph).

This proves (a), (b), and (c), completing the proof of Theorem 4.1. □

5   Supporting results

For completeness, in this section we provide some additional details.

5.1 Joint measures

Let $(\mathcal{X}, \mathcal{A})$ and $(\Theta, \mathcal{B})$ be measurable spaces, let P_θ be a probability measure on $(\mathcal{X}, \mathcal{A})$ for each $\theta \in \Theta$, and let Π be a probability measure on $(\Theta, \mathcal{B})$. We would like to define a joint probability measure ν on $(\mathcal{X} \times \Theta, \mathcal{A} \otimes \mathcal{B})$ by

$$\nu(E) = \int_\Theta P_\theta(E_\theta) \, d\Pi(\theta)$$

for all $E \in \mathcal{A} \otimes \mathcal{B}$, where $E_\theta = \{x \in \mathcal{X} : (x, \theta) \in E\}$ is the θ-section of E. However, in order for this definition to make sense, we need $\theta \mapsto P_\theta(E_\theta)$ to be a measurable function.

Lemma 5.1. *If $\theta \mapsto P_\theta(A)$ is measurable for every $A \in \mathcal{A}$, then $\theta \mapsto P_\theta(E_\theta)$ is measurable for every $E \in \mathcal{A} \otimes \mathcal{B}$.*

Proof. We apply the $\pi - \lambda$ theorem; see, e.g., Kallenberg (2002), 1.1. The collection of rectangles $\mathcal{R} = \{A \times B : A \in \mathcal{A}, \ B \in \mathcal{B}\}$ is a π-system. Let $\mathcal{L}$ be the collection of sets $E \in \mathcal{A} \otimes \mathcal{B}$ such that $\theta \mapsto P_\theta(E_\theta)$ is measurable. We will show that $\mathcal{L}$ is a λ-system containing $\mathcal{R}$. Since $\sigma(\mathcal{R}) = \mathcal{A} \otimes \mathcal{B}$ then $\mathcal{A} \otimes \mathcal{B} \subseteq \mathcal{L}$ will follow by the $\pi - \lambda$ theorem. This will prove the lemma.

To see that $\mathcal{R} \subseteq \mathcal{L}$, let $E \in \mathcal{R}$ and write $E = A \times B$ with $A \in \mathcal{A}, B \in \mathcal{B}$. Then

$$E_\theta = \left\{ \begin{array}{ll} A & \text{if } \theta \in B \\ \varnothing & \text{if } \theta \notin B, \end{array} \right.$$

so $P_\theta(E_\theta) = P_\theta(A) I_B(\theta)$, which is measurable as a function of θ. Hence, $E \in \mathcal{L}$.

To show that $\mathcal{L}$ is a λ-system, we must show that $\mathcal{X} \times \Theta \in \mathcal{L}$ and $\mathcal{L}$ is closed under increasing unions and proper differences. Since $\mathcal{R} \subseteq \mathcal{L}$, then $\mathcal{X} \times \Theta \in \mathcal{L}$ is immediate. If $E^1, E^2, \ldots \in \mathcal{L}$ such that $E^1 \subseteq E^2 \subseteq \cdots$, and $E = \bigcup_n E^n$, then $P_\theta(E_\theta) = P_\theta(\bigcup_n E_\theta^n) = \lim_{n \to \infty} P_\theta(E_\theta^n)$, which is measurable as a function of θ (being a limit of measurable functions). If $E, F \in \mathcal{L}$ such that $E \subseteq F$ then $P_\theta((F \backslash E)_\theta) = P_\theta(F_\theta \backslash E_\theta) = P_\theta(F_\theta) - P_\theta(E_\theta)$, which is measurable as a function of θ. $\square$

5.2 Measurable families are preserved under countable products

Now, consider the situation in Sections 2 and 3. Since Assumption 2.1 ensures that $\theta \mapsto P_\theta(A)$ is measurable for all $A \in \mathcal{A}$, Lemma 5.1 shows that $\theta \mapsto$

$P_\theta(E_\theta)$ is measurable for all $E \in \mathcal{A} \otimes \mathcal{B}$. However, since we would like to define the joint measure as

$$\mu(E) = \int_\Theta P_\theta^\infty(E_\theta) \, d\Pi(\theta)$$

on $(\mathcal{X}^\infty \times \Theta, \mathcal{A}^\infty \otimes \mathcal{B})$, what we really need is for $\theta \mapsto P_\theta^\infty(E_\theta)$ to be measurable for all $E \in \mathcal{A}^\infty \otimes \mathcal{B}$. If we can show that $\theta \mapsto P_\theta^\infty(A)$ is measurable for all $A \in \mathcal{A}^\infty$, then this will follow from Lemma 5.1 simply by replacing $(\mathcal{X}, \mathcal{A}, P_\theta)$ with $(\mathcal{X}^\infty, \mathcal{A}^\infty, P_\theta^\infty)$.

Lemma 5.2. *Under the setup of Sections 2 and 3, $\theta \mapsto P_\theta^\infty(A)$ is measurable for every $A \in \mathcal{A}^\infty$.*

Proof. This is another application of the indispensable $\pi - \lambda$ theorem. For our π-system, we take

$$\mathcal{R} = \{A_1 \times \cdots \times A_k \times \mathcal{X} \times \mathcal{X} \times \cdots : A_1, \ldots, A_k \in \mathcal{A}, \ k = 1, 2, \ldots \}.$$

Let $\mathcal{L}$ be the collection of sets $A \in \mathcal{A}^\infty$ such that $\theta \mapsto P_\theta^\infty(A)$ is measurable. We will show that $\mathcal{L}$ is a λ-system containing $\mathcal{R}$. Since $\sigma(\mathcal{R}) = \mathcal{A}^\infty$, then $\mathcal{A}^\infty \subseteq \mathcal{L}$ will follow by the $\pi - \lambda$ theorem, proving the desired result.

To see that $\mathcal{R} \subseteq \mathcal{L}$, note that if $A_1, \ldots, A_k \in \mathcal{A}$ then

$$P_\theta^\infty(A_1 \times \cdots \times A_k \times \mathcal{X} \times \mathcal{X} \times \cdots) = P_\theta(A_1) \cdots P_\theta(A_k),$$

which is measurable as a function of θ, since $\theta \mapsto P_\theta(A_i)$ is measurable for each $i = 1, \ldots, k$ by Assumption 2.1.

To show that $\mathcal{L}$ is a λ-system, we must show that $\mathcal{X}^\infty \in \mathcal{L}$ and $\mathcal{L}$ is closed under increasing unions and proper differences. Since $\mathcal{R} \subseteq \mathcal{L}$, we have $\mathcal{X}^\infty \in \mathcal{L}$. If $A_1, A_2, \ldots \in \mathcal{L}$ such that $A_1 \subseteq A_2 \subseteq \cdots$, and $A = \bigcup_i A_i$, then $P_\theta^\infty(A) = \lim_{i \to \infty} P_\theta^\infty(A_i)$, which is measurable as a function of θ (being a limit of measurable functions). Lastly, if $A, B \in \mathcal{L}$ such that $A \subseteq B$ then $P_\theta^\infty(B \setminus A) = P_\theta^\infty(B) - P_\theta^\infty(A)$ is measurable as a function of θ. $\square$

References

A. Barron, M. J. Schervish, and L. Wasserman. The consistency of posterior distributions in nonparametric problems. *The Annals of Statistics*, 27(2): 536–561, 1999.

J. L. Doob. Application of the theory of martingales. In *Actes du Colloque International Le Calcul des Probabilités et ses applications (Lyon, 28 Juin – 3 Juillet, 1948)*, pages 23–27. Paris CNRS, 1949. (Doob's original paper can be found at the end of a historical account entitled "Doob at Lyon", currently available at: `www.jehps.net/juin2009/Locker.pdf`.).

D. A. Freedman. On the asymptotic behavior of Bayes' estimates in the discrete case. *The Annals of Mathematical Statistics*, pages 1386–1403, 1963.

S. Ghosal and A. van der Vaart. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press, 2017.

J. K. Ghosh and R. V. Ramamoorthi. *Bayesian Nonparametrics*. Springer-Verlag, New York, 2003.

O. Kallenberg. *Foundations of Modern Probability (Second Edition)*. Springer-Verlag, New York, 2002.

A. S. Kechris. *Classical Descriptive Set Theory*. Springer-Verlag, New York, 1995.

A. Lijoi, I. Prünster, and S. G. Walker. Extending Doob's consistency theorem to nonparametric densities. *Bernoulli*, 10(4):651–663, 2004.

A. Nobile. *Bayesian Analysis of Finite Mixture Distributions*. PhD thesis, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA, 1994.

K. R. Parthasarathy. *Probability Measures on Metric Spaces*. American Mathematical Society, Providence, RI, 2005.

L. Schwartz. On Bayes procedures. *Probability Theory and Related Fields*, 4 (1):10–26, 1965.

A. W. Van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.