CONJUGATE PROJECTIVE LIMITS

Peter Orbanz

University of Cambridge

We characterize conjugate nonparametric Bayesian models as projective limits of conjugate, finite-dimensional Bayesian models. In particular, we identify a large class of nonparametric models representable as infinite-dimensional analogues of exponential family distributions and their canonical conjugate priors. This class contains most models studied in the literature, including Dirichlet processes and Gaussian process regression models. To derive these results, we introduce a representation of infinite-dimensional Bayesian models by projective limits of regular conditional probabilities. We show under which conditions the nonparametric model itself, its sufficient statistics, and – if they exist – conjugate updates of the posterior are projective limits of their respective finite-dimensional counterparts. We illustrate our results both by application to existing nonparametric models and by construction of a model on infinite permutations.

1. Introduction. Nonparametric Bayesian statistics revolves around a small number of fundamental models, including the Dirichlet process [16], Gaussian process [46, 51], beta process [23] and gamma process. All these models have conjugate posteriors [50]. Since most nonparametric Bayesian models are derived from such fundamental, conjugate models, virtually all nonparametric Bayesian inference is based directly or indirectly on conjugacy. The objective of this work is to study the shared properties of fundamental models and to characterize the class of models admitting conjugate posteriors.

By *nonparametric Bayesian model*, we refer to a Bayesian model on an infinite-dimensional parameter space [21, 24, 50]. We do not a priori distinguish between discrete models (e.g. Dirichlet processes) and continuous models (e.g. Gaussian process regression). Conjugate models such as the Gaussian and Dirichlet processes share another property, the existence of marginals in the exponential family. In the case of the Dirichlet process, there is a well-known connection between the two properties: Conjugacy of the nonparametric model can be derived directly from the conjugacy of the marginal, finite-dimensional Dirichlet priors [20]. We will show in the following how the vague but intuitively appealing link between conjugate posteriors and exponential family marginals in general nonparametric Bayesian models can be made precise. If an infinite-dimensional model is constructed from finite-dimensional marginal distributions, conjugacy of the marginals proves sufficient to guarantee a conjugate posterior of the nonparametric model.

The analysis of shared properties of models requires a shared representation, which leads us almost inevitably to projective limits, i.e. the representation of a stochastic process by its finite-dimensional marginal distributions [9]. Most representations used in Bayesian nonparametrics are adapted to specific models – examples include Lévy processes, stick-breaking constructions [49], transformed Poisson processes [17], and normalized completely random measures [25]. The advantages of such model-specific representations are that they emphasize useful properties of the model in question, as well as their simplicity – more general representations tend to come at the price of more technical subtleties involved in their application. Possible choices for more general representations of probability measures are densities, characteristic functions and projective limits. Densities are not applicable for nonparametric Bayesian models, both for lack of a suitable translation-invariant carrier measure on infinite-dimensional space, and because some important models (such as the Dirichlet process) are not dominated [47]. Characteristic functions are ill-suited for the questions considered here, since they do not live on the actual sample space.

A *projective limit* (also called an *inverse limit*) assembles an infinite-dimensional mathematical object from a family of finite-dimensional objects [7–9]. Projective limits of probability measures, i.e. Kolmogorov's extension theorem and its generalizations, are widely used in the construction of stochastic processes: A stochastic process with paths in an infinite-dimensional space is represented in terms of its finite-dimensional marginals [26]. Since a projective limit representation is not sufficient to express some important properties of sample paths, such as continuity of random functions or σ-additivity of random measures, we combine projective limit representations with the notion of a *pullback* under a suitable transformation mapping [19]. The pullback expresses almost sure properties of sample paths not expressible in terms of the projective limit.

Projective limits can be defined not only for measures, but also for sets, functions, and a wide variety of mathematical structures [7–9, 34]. This allows us to both define projective limits of conditional probabilities, and to apply the representation to sufficient statistics and other functions associated with a model. In this manner, we obtain a representation of a nonparametric Bayesian model in terms of a family of finite-dimensional "marginal" Bayesian models. The properties of the nonparametric model can be related directly to those of the parametric marginals. Application to the questions of sufficiency and conjugacy shows that both the sufficient statistics and the posterior updates of a nonparametric Bayesian model can be expressed in terms of their finite-dimensional counterparts. This result in particular establishes a large family of models – containing both the Gaussian and the Dirichlet process – which can be regarded as a nonparametric analogue of the exponential family, in a sense to be made precise in the ensuing discussion.

The results imply an approach to the construction from scratch of nonparametric Bayesian models on a wide range of domains. In this regard, an additional appeal of projective limits is the large number of such representations available in the mathematical literature, each of which may potentially be harvested for the purpose of Bayesian nonparametrics. Examples include the projective limit/pullback construction of continuous functions used in the construction of the Gaussian process [e.g. 2]; a variety of constructions of topological and algebraic objects discussed by Bourbaki [7, 8, 9]; the construction of random coagulation and fragmentation processes [4]; and recent constructions of infinite limits of permutations by Kerov *et al.* [29], and of graph limits by Lovász and Szegedy [37].

1.1. *Summary of Results.* Projective limits are, by themselves, not capable of expressing all properties of stochastic processes such as the Dirichlet and Gaussian process, and additional steps are required to obtain an applicable distribution. These steps and their formalization in the literature differ widely between models. Since our problem requires a unified formalism, we derive a representation in terms of a pullback of the projective limit under a measurable embedding [19]. Intuitively, the stochastic process of interest is represented by uniquely encoding each of its paths as a path of the projective limit process. The resulting representation is applicable to all important nonparametric Bayesian models.

Projective limits and pullbacks preserve a variety of properties of functions and set functions. For example, the projective limit and pullback obtained from injective functions are again injective functions. The same holds for continuous and measurable mappings, bijections, probability measures and regular conditional probabilities. Some of these facts are standard results, others are established in the following. In particular, we show:

(1) The countable projective limit of a projective family of probability kernels (regular conditional probabilities) on finite-dimensional spaces is a probability kernel on an infinite-dimensional space. The extensions theorems of Kolmogorov-Bochner and of Prokhorov both generalize along these lines (Theorem 1; Corollary 1). Similarly, the pullback of a probability kernel is again a probability kernel (Proposition 1).

A Bayesian model is defined by conditional probabilities. By application of the previous results to these conditionals, we obtain:

(2) A projective limit can be applied directly to finite-dimensional Bayesian models, resulting in infinite-dimensional Bayesian models on the corresponding projective limit spaces (Sec. 4.2). Pullbacks also preserve the structure of the Bayesian model (Sec. 4.3). Both operations commute with the computation of posteriors (Diagram (4.3)).

In other words, nonparametric Bayesian models can be directly constructed from finite-dimensional "marginal" Bayesian models. The construction is analogous to the construction of stochastic process measures by means of projective limits and pullbacks.

Since projective limits and pullbacks are applicable to measurable functions, they apply simultaneously to a model and its associated statistics.

(3) The projective limit of the sufficient statistics (resp. sufficient σ-algebras) of the marginal models is a sufficient statistic (resp. sufficient σ-algebra) of the infinite-dimensional projective limit model (Sec. 5). We also show that, if the sufficient σ-algebras of the marginals are minimal, the projective limit σ-algebra is again minimal sufficient. This holds even if the projective limit model is undominated (Proposition 3).

The great utility of conjugate Bayesian models is due to the representability of their posterior parameters as functions of the data and the model hyperparameters. We show that the structure and functional form of this update process carries over from the marginals to the nonparametric model.

(4) Projective limits and pullbacks of conjugate Bayesian models are conjugate, and in particular, the mapping to the posterior parameter of the infinite-dimensional model is the projective limit of the update mappings of the marginal models (Sec. 6). For the specific case in which the finite-dimensional marginals are conjugate exponential family models, we obtain a nonparametric analogue of the Diaconis-Ylvisaker representation [14] of conjugate parametric models (Corollary 2).

1.2. *Related Work.* The application of projective limits to statistical models was pioneered by Lauritzen [35, 36], to derive a family of parametric models which are defined by sequences (rather than averages) of sufficient statistics and generalize beyond exchangeable observations. In Lauritzen's work, the "dimensions" of the projective limit describe repeated observations from a parametric model, rather than dimensions of sample and parameter space as in our case. Nonetheless, if n observations in Lauritzen's "projective statistical fields" [36, Chapter IV] are interpreted as a sample of size n in a Bayesian nonparametric model, the projective limit aspects of Sec. 3 below can be regarded as an analogue of Lauritzen's projective fields for application to nonparametric Bayesian models.

Conjugate analysis in the finite-dimensional, parametric case, i.e. for dominated models, is the subject of a substantial literature [e.g. 12–14]. Bernardo and Smith [3] give a concise overview. It is also well known that almost all nonparametric Bayesian models are conjugate [50]; if the model is undominated, Bayes' theorem is not applicable, and conjugacy is often the only way to represent the posterior. Other models indirectly rely on conjugacy: The popular Dirichlet process mixture model [1, Example 4] does not have a conjugate posterior, but is amenable to Gibbs sampling only because the Dirichlet process law of the mixing measure is conjugate. However, conjugacy of nonparametric Bayesian models has not so far been analyzed as a structural property, in contrast to the parametric case. A notable exception is the special case of sequential independent increment processes, for which a class of models with exponential family marginals is discussed in detail by Küchler and Sørensen [31]. For models of this type, the existence of conjugate posteriors is studied by Magiera and Wilczyński [38].

1.3. *Outline.* We begin by developing a representation of stochastic processes suitable for our purposes in Sec. 2. Projective limits and pullbacks are then applied to conditional probabilities in Sec. 3, which facilitates their application to Bayesian models in Sec. 4. From the representation of nonparametric Bayesian models so obtained, we derive results on their sufficient statistics in Sec. 5, and on conjugate posteriors in Sec. 6. Two detailed examples in Sec. 7 illustrate the approach and results. Since projective limits of functions and pullbacks of measures are not commonly used in statistics, a brief summary of relevant facts is provided in Appendix A.

1.4. *Notation and Assumptions.* All random variables are in the following assumed to share an abstract probability space $(\Omega, \mathcal{A}, \mathbb{P})$ as common domain. We will frequently have to distinguish spaces of different dimensions, which are indexed by subscripts as $\mathcal{X}_{\mathrm{I}}$, $\mathcal{T}_{\mathrm{J}}$, etc. All mappings, σ-fields and other quantities on these spaces are indexed accordingly. We use superscripts $x_{\mathrm{I}}^{(j)}$ to denote elements of sequences or repetitive observations. For any measure ν, a superscript ν^* indicates the corresponding outer measure. Observations are generally assumed exchangeable. Topological spaces are assumed to be Polish spaces, i.e. complete, separable and metrizable. We refer to a measurable space as *standard Borel* if it is the Borel space generated by a Polish topology. As the underlying spaces are Polish, all conditional probabilities $P[X|\mathcal{C}]$ are required to be regular conditional probabilities (probability kernels).

2. Construction of Stochastic Processes. This section briefly surveys the construction of stochastic processes and introduces some relevant definitions. It assumes familiarity with the terminology of projective limits, which is used here in the sense of Bourbaki [7, 8, 9]. A more detailed summary of projective limits and pullbacks is given in Appendix A.

2.1. *Projective Limit Notation.* Let $(D, \preceq)$ be a partially ordered, directed set. *We assume D to be countable throughout.* Let $\langle \mathcal{X}_{\text{I}}, \mathcal{B}_{\text{I}}, f_{\text{JI}} \rangle_{\text{I} \preceq \text{J} \in \text{D}}$, or $\langle \mathcal{X}_{\text{I}}, \mathcal{B}_{\text{I}}, f_{\text{JI}} \rangle_{\text{D}}$ for short, be a projective system of topological measurable spaces indexed by D. That is, $\mathcal{X}_{\text{I}}$ are topological spaces, $\mathcal{B}_{\text{I}}$ their Borel σ-algebras, and $f_{\text{JI}} : \mathcal{X}_{\text{J}} \to \mathcal{X}_{\text{I}}$ are continuous generalized projection mappings satisfying (A.1). Denote by $\mathcal{X}_{\text{D}}$ the projective limit space. The mappings f_{JI} induce a family of unique generalized projection mappings $f_{\text{I}} : \mathcal{X}_{\text{D}} \to \mathcal{X}_{\text{I}}$. The space $\mathcal{X}_{\text{D}}$ is endowed with the smallest topology Top_{D} which makes all f_{I} continuous. Top_{D} is called the projective limit topology, and generates the projective limit Borel σ-algebra $\mathcal{B}_{\text{D}}$. A family $\langle P_{\text{I}} \rangle_{\text{D}}$ of probability measures on the spaces $\mathcal{X}_{\text{I}}$ is called *projective* if $f_{\text{JI}}(P_{\text{J}}) = P_{\text{I}}$ whenever $I \preceq J$. By the extension theorem of Kolmogorov and Bochner (Theorem 4), any projective family defines a unique probability measure P_{D} on $(\mathcal{X}_{\text{D}}, \mathcal{B}_{\text{D}})$ which satisfies $P_{\text{I}} = f_{\text{I}}(P_{\text{D}})$ for all $I \in D$. We refer to $P_{\text{D}} =: \varprojlim \langle P_{\text{I}} \rangle_{\text{D}}$ as the projective limit of $\langle P_{\text{I}} \rangle_{\text{D}}$, and to the measures P_{I} as the *marginals* of P_{D}. Intuitively, the measures P_{I} are probability distributions on finite-dimensional spaces, and P_{D} is a joint distribution of a stochastic process $\langle X_{\text{I}} \rangle_{\text{D}}$ on the infinite-dimensional space $\mathcal{X}_{\text{D}}$.

The projective limit space $\mathcal{X}_{\text{D}}$ is a subset of the product space $\prod_{I \in D} \mathcal{X}_{\text{I}}$. If pr_{I} denotes the canonical projection onto $\mathcal{X}_{\text{I}}$ in the product space, the canonical mappings f_{I} are the restrictions $f_{\text{I}} = \text{pr}_{\text{I}}|_{\mathcal{X}_{\text{D}}}$. It is often useful to regard the elements x_{D} of $\mathcal{X}_{\text{D}}$ as functions $x_{\text{D}} : D \to \cup_{I \in D} \mathcal{X}_{\text{I}}$, or more precisely, as functions on D taking values $x(I) \in \mathcal{X}_{\text{I}}$. In the context of nonparametric Bayesian estimation, the indices $I \in D$ may be thought of as covariates and the function values $x_{\text{I}} = x(I)$ as measurements, if X_{D} represents the observation space of the model. If X_{D} is a parameter space, continuous real-valued functions x_{D} may represent regressors, set functions x_{D} may represent density estimates, etc.

2.2. *Stochastic Processes.* A stochastic process is in general a collection $\langle X_{\text{I}} \rangle_{\text{D}}$ of random variables, indexed by an infinite set D. Hence, if

$P_D = \varprojlim \langle P_I \rangle_D$ is a projective limit measure with marginals P_I, the family $\langle X_I \rangle_D$ of random variables distributed according to the measures P_I is a stochastic process indexed by D. Conversely, any stochastic process can in principle be regarded as the projective limit of its marginals on suitably chosen subspaces. However, constructions of stochastic processes as projective limits have to address two fundamental technical problems:

(a) *Uncountable index sets.* An event $A \subset \mathcal{X}_D$ is measurable under P_D only if it depends on an at most countable subset $D' \subset D$ of coordinates [e.g. 5, Theorem 36.3]. In other words, unless D is countable, singletons are not measurable in the projective limit space, and the projective limit measure P_D is not useful for most applications.

(b) *Infinitary properties of sample paths.* If the spaces $\mathcal{X}_I$ in the projective system are finite-dimensional, the projective limit construction can only express properties of the random functions x_D that are *finitary*, such as non-negativity or monotonicity of real-valued functions, or finite additivity of set functions.

Problem (a) means, for example, that projective limits can directly define a useful measure on functions $\mathbb{Q} \to \mathbb{R}$, but not on functions $\mathbb{R} \to \mathbb{R}$, since the space $\mathbb{R}^{\mathbb{R}}$ of all functions $\mathbb{R} \to \mathbb{R}$ has uncountable dimension. Problem (b) implies, for example, that a projective limit construction of random set functions can define a sample space consisting of all charges (finitely additive probabilities), but not a sample space containing exactly all probability measures, which would require the projective limit to express countable additivity.

Both problems (a) and (b) can be jointly addressed in an elegant manner by means of pullbacks under suitable functions. Given a space $\mathcal{X}$, a measure space $(\mathcal{Y}, \mathcal{B}_{\mathcal{Y}}, \nu)$ and a function $\mathcal{J} : \mathcal{X} \to \mathcal{Y}$, the *pullback* of ν under $\mathcal{J}$ is the measure $\tilde{\nu}$ on $(\mathcal{X}, \mathcal{J}^{-1}\mathcal{B}_{\mathcal{Y}})$ satisfying $\mathcal{J}(\tilde{\nu}) = \nu$. The pullback measure ν is uniquely defined whenever the image $\mathcal{J}(\mathcal{X}) \subset \mathcal{Y}$ has full outer measure under ν, that is if $\nu^*(\mathcal{J}(\mathcal{X})) = \nu(\mathcal{Y})$ – see App. A.2 for more details. The most common example of a pullback is the restriction of a measure to a (possibly non-measurable) subspace, in which case $\mathcal{X} \subset \mathcal{Y}$ is an arbitrary subset and $\mathcal{J} : \mathcal{X} \hookrightarrow \mathcal{Y}$ the canonical inclusion map. The σ-algebra $\mathcal{J}^{-1}\mathcal{B}_{\mathcal{Y}}$ is then precisely the subspace σ-algebra $\mathcal{B}_{\mathcal{Y}} \cap \mathcal{X}$. Hence, if ν is a probability measure on $\mathcal{Y}$, and if the subspace has outer measure $\nu^*(\mathcal{X}) = 1$, the pullback $\tilde{\nu}$ is the restriction of ν to $(\mathcal{X}, \mathcal{B}_{\mathcal{Y}} \cap \mathcal{X})$.

To construct stochastic processes, we will specifically consider pullbacks under embedding maps. Let $\phi : \tilde{\mathcal{X}} \to \mathcal{X}$ be a mapping between topological spaces. Such a mapping is called an *embedding* if, regarded as a mapping onto its image, it is a homeomorphism. Analogously, we refer to ϕ as a *Borel embedding* if it constitutes a Borel isomorphism of its domain and its image $(\Gamma, \mathcal{B}(\mathcal{X}) \cap \Gamma)$. A definition of a stochastic process suitable for our questions in Bayesian nonparametrics is the following:

DEFINITION 1. Let $(\tilde{\mathcal{X}}, \mathcal{B}(\tilde{\mathcal{X}}), \tilde{P})$ be a topological measure space and $\langle \mathcal{X}_{\mathrm{I}}, \mathcal{B}_{\mathrm{I}}, P_{\mathrm{I}} \rangle_{\mathrm{D}}$ a projective system of standard Borel spaces with countable, directed index set D. Then $\tilde{P}$ is called a *countably representable stochastic process* if it is the pullback of the projective limit measure $P_{\mathrm{D}} := \varprojlim \langle P_{\mathrm{I}} \rangle_{\mathrm{D}}$ under a Borel embedding $\phi : \tilde{\mathcal{X}} \to \Gamma \subset \mathcal{X}_{\mathrm{D}}$.

An asymptotically identifiable model can have at most a countable number of degrees of freedom. Definition 1 therefore incorporates a restriction to sample paths of countable complexity: In a projective limit, the indices $I \in D$ can be thought of as dimensions or degrees of freedom; hence, the sample space $\tilde{\mathcal{X}}$ of a stochastic process with countably many degrees of freedom can be embedded into a suitably chosen projective limit space $\mathcal{X}_{\mathrm{D}}$ with countable index set.

The special case in which ν is a projective limit measure on an *uncountable* product space $\mathcal{Y} := \mathcal{X}_{\mathrm{D}}$, constructed from Euclidean spaces $\mathcal{X}_{\mathrm{I}} = \mathbb{R}^{\mathrm{I}}$, and $\mathcal{X}$ is e.g. the subset of continuous functions, is known in stochastic process theory as "Doob's separability theorem". In this case, the pullback $\tilde{\nu}$ is called a "separable modification" of ν [15]. The index set D is the set of all finite subsets of the "separant", a dense countable subset of $\mathbb{R}_+$. See also [5, Chapter 38].

The intuition that sample paths of $\tilde{P}$ (the elements of $\tilde{\mathcal{X}}$) are uniquely represented by their embeddings into $\mathcal{X}_{\mathrm{D}}$ can be helpful in establishing that a given mapping ϕ is a Borel embedding: Suppose that a measurable map ϕ is given. As a mapping onto its image, it is trivially surjective, so what remains to be established for Borel isomorphy is the existence of a measurable inverse. If the elements of $\tilde{\mathcal{X}}$ are uniquely represented by their embeddings, then ϕ is injective. In most settings, the mapping ϕ can be directly derived from a suitable representation result, such as the representation of continuous functions by their values on countable subsets as mentioned above, or the representation of measures by their values on a generating algebra of sets (by Carathéodory's extension theorem). If additionally both $\tilde{\mathcal{X}}$ and Γ are standard Borel spaces, Borel isomorphy follows automatically, since measurable bijections between standard Borel spaces are bimeasurable [26, Theorem A1.3].

EXAMPLE 1 (Dirichlet process). Suppose that $\tilde{P}$ is a Dirichlet process

DP (αG_0) over a standard Borel space $(V, \mathcal{B}_V)$. The spaces $\mathcal{X}_{\text{I}}$ can be chosen as finite-dimensional simplices $\triangle_{\text{I}} \subset \mathbb{R}^{\text{I}}$, indexed by measurable partitions $I = (A_1, \ldots, A_{|I|})$ of the space V. The marginals $P_{\text{I}}(X_{\text{I}})$ are Dirichlet distributions on the simplices. The projective limit is the space of all charges defined on a specific countable subalgebra $\mathcal{Q}$ which generates $\mathcal{B}_V$. The space $\tilde{\mathcal{X}}$ is the space of all probability measures on $\mathcal{B}_V$, and its image $\Gamma = \phi(\tilde{\mathcal{X}})$ is the set of probability measures on the subalgebra $\mathcal{Q}$. For a given measure $\tilde{x}$ on $\mathcal{B}_V$, the image $\phi(\tilde{x})$ is the restriction of $\tilde{x}$ to $\mathcal{Q}$. By the Carathéodory extension theorem, ϕ is injective. Whether P_{D} admits a pullback under ϕ depends on the parametrization of the marginals: If G_0 is a charge on $\mathcal{Q}$, and each Dirichlet marginal has parameter $\alpha \cdot f_{\text{I}}(G_0)$ for some fixed $\alpha > 0$, the Dirichlet distributions form a projective family. The projective limit satisfies $P_{\text{D}}^*(\Gamma) = 1$ if and only if G_0 is countably additive. Sec. 7.1 revisits this example in detail.

EXAMPLE 2 (Gaussian Process). To obtain a Gaussian process measure on the set $\tilde{\mathcal{X}} := C(\mathbb{R}_+, \mathbb{R})$ of continuous functions $\mathbb{R}_+ \to \mathbb{R}$, a projective limit is constructed as follows: Choose D as the set of all finite subsets I of $\mathbb{Q}_+$, ordered by inclusion, and define $\mathcal{X}_{\text{I}} := \prod_{i \in I} \mathbb{R}$. Let $f_{\text{JI}} := \text{pr}_{\text{JI}}$ be the coordinate projections in Euclidean space, and $\langle P_{\text{I}} \rangle_{\text{D}}$ a projective family of multivariate Gaussian distributions. The projective limit space is $\mathcal{X}_{\text{D}} = \mathbb{R}^{\mathbb{Q}_+}$, and the projective limit measure P_{D} can be regarded as a discrete-time Gaussian process indexed by $\mathbb{Q}_+$. We embed $\tilde{\mathcal{X}}$ into $\mathbb{Q}_+$ by means of the restriction map $\phi : \tilde{x} \mapsto \tilde{x}|_{\mathbb{Q}_+}$. The mapping ϕ is a Borel isomorphism as required in Def. 1: As a canonical inclusion map, ϕ is continuous and hence measurable. Since the representation of $\tilde{x}$ by its restriction is unique, ϕ is injective. The σ-algebra $\phi^{-1}\mathcal{B}_{\text{D}}$ induced by ϕ on $C(\mathbb{R}_+, \mathbb{R})$ coincides with the Borel σ-algebra generated by the topology of compact convergence [19, Section 454O]. Hence, $\tilde{\mathcal{X}}$ is standard Borel, and ϕ bimeasurable. The requirement $P_{\text{D}}^*(\tilde{\mathcal{X}}) = 1$ for the existence of the pullback measure is *not* generally satisfied for arbitrary Gaussian marginals P_{I}. It can, however, be related to the parameters of the marginals. A prototypical result is Kolmogorov's continuity theorem [2, Theorem 39.3]: If the expectation under P_{D} satisfies $\mathbb{E}[|X_i - X_j|^\alpha] \leq \gamma |i - j|^\beta$ for all $i, j \in \mathbb{Q}_+$ and any fixed $\alpha, \beta, \gamma \in \mathbb{R}_{>0}$, then $P_{\text{D}}^*(C(\mathbb{R}_+, \mathbb{R})) = 1$. An example to the contrary is obtained for marginals satisfying $\text{Cov}[X_i, X_j] = \delta_{ij}$. The resulting Gaussian white noise process is almost surely discontinuous, and hence $P_{\text{D}}^*(\tilde{\mathcal{X}}) \neq 1$.

3. Projective Limits of Conditional Probabilities. In this section, we apply the projective limit approach to conditional probabilities. By means of Theorem 1 below, a conditional probability on an infinite-dimensional space can be assembled as a projective limit of conditional probabilities on finite-dimensional spaces, in a similar manner as a probability measure can be specified as projective limits by means of Theorem 4.

3.1. *Construction Results.* Let $\langle \mathcal{X}_{\text{I}}, \mathcal{B}_{\text{I}}, f_{\text{JI}} \rangle_{\text{D}}$ be a projective system of standard Borel spaces. For each $I \in D$, let $P_{\text{I}}[X_{\text{I}}|\mathcal{C}_{\text{I}}]$ be a regular conditional probability on $(\mathcal{X}_{\text{I}}, \mathcal{B}_{\text{I}})$. More precisely, $X_{\text{I}} : \Omega \to \mathcal{X}_{\text{I}}$ is a random variable, $\mathcal{C}_{\text{I}} \subset \mathcal{A}$ is a σ-subalgebra on the abstract probability space Ω, and $P_{\text{I}}[\,.\,|\mathcal{C}_{\text{I}}](\,.\,) : \mathcal{B}_{\text{I}} \times \Omega \to [0,1]$ is a probability kernel.

The projections f_{JI} immediately generalize from probability measures to conditional probabilities by means of

$$(f_{\text{JI}} P_{\text{J}})[X_{\text{J}} \in \,.\,|\mathcal{C}_{\text{J}}] := P_{\text{I}}[X_{\text{I}} \in f_{\text{JI}}^{-1}\,.\,|\mathcal{C}_{\text{J}}]\;. \tag{3.1}$$

The projector acts only on the first argument of the probability kernel. To generalize the notion of a projective family, the second argument has to be taken into account as well: Consider a parametric family $P_{\text{I}}[X_{\text{I}}|\Theta_{\text{I}}]$, i.e. each $\mathcal{C}_{\text{I}}$ is generated by a parameter random variable Θ_{I}. Typically, if Θ_{J} parametrizes a high-dimensional random variable X_{J} and Θ_{I} a lower-dimensional variable X_{I}, we would assume the information contained in Θ_{I} to be a subset of the information contained in Θ_{J}. The concept can be expressed in very general terms by assuming that the σ-algebras $\mathcal{C}_{\text{I}}$ are ordered in accordance with the index set, i.e. $\mathcal{C}_{\text{I}} \subset \mathcal{C}_{\text{J}}$ whenever $I \preceq J$. In analogy to the index set, we refer to such an ordered family of σ-algebras as *directed*.

DEFINITION 2 (Projective family of conditional probabilities). Let $\langle \mathcal{C}_{\text{I}} \rangle_{\text{D}}$ be a directed family of σ-algebras. A family $\langle P_{\text{I}}[X_{\text{I}}|\mathcal{C}_{\text{I}}] \rangle_{\text{D}}$ of probability kernels on the the projective system $\langle \mathcal{X}_{\text{I}}, \mathcal{B}_{\text{I}}, f_{\text{JI}} \rangle_{\text{D}}$ is called *projective* if

$$(f_{\text{JI}} P_{\text{J}})[\,.\,|\mathcal{C}_{\text{J}}] =_{\text{a.e.}} P_{\text{I}}[\,.\,|\mathcal{C}_{\text{I}}] \qquad \text{whenever } I \preceq J\;. \tag{3.2}$$

Projectivity of conditionals is a stronger condition than projectivity of measures: We have $P(A) = \int_\Omega P[A|\mathcal{C}](\omega) d\mathbb{P}(\omega)$ for any $\mathcal{C} \subset \mathcal{A}$, and hence $P_{\text{J}}[f_{\text{JI}}^{-1} A_{\text{I}}|\mathcal{C}_{\text{J}}] =_{\text{a.e.}} P_{\text{I}}[A_{\text{I}}|\mathcal{C}_{\text{I}}](\omega)$ implies $P_{\text{J}}(f_{\text{JI}}^{-1} A_{\text{I}}) = P_{\text{I}}(A_{\text{I}})$. Therefore, projective conditionals imply projective measures, but the converse only holds under additional conditions (cf Lemma 2). If the conditional distributions of random variables X_{I} are projective given one directed family of σ-algebras, the same may be not true for another family, so the conditional projector is effectively parametrized by the family $\langle \mathcal{C}_{\text{I}} \rangle_{\text{D}}$.

THEOREM 1 (Projective limits of conditional probabilities). *Let E be a countable directed set. Let $\langle P_I[X_I|\mathcal{C}_I]\rangle_D$ be a projective family of probability kernels on a projective system $\langle \mathcal{X}_I, \mathcal{B}_I, f_{JI}\rangle_D$ of Polish measurable spaces. Then there exists a unique (up to equivalence) probability kernel, denoted $P_D[\,.\,|\mathcal{C}_D]$, which satisfies*

$$(f_I P_D)[\,.\,|\mathcal{C}_D] =_{a.e.} P_I[\,.\,|\mathcal{C}_I] \qquad \text{for all } I \in D\ , \tag{3.3}$$

and is measurable with respect to $\mathcal{C}_D := \sigma(\mathcal{C}_I; I \in D)$.

In analogy to probability measures, we refer to the conditionals $P_{\mathrm{I}}[X_{\mathrm{I}}|\mathcal{C}_{\mathrm{I}}]$ as the *marginal conditional probabilities* of $P_{\mathrm{D}}[X_{\mathrm{D}}|\mathcal{C}_{\mathrm{D}}]$, or *marginals* for short.

PROOF. The proof relies on the simple fact that measurability of mappings is preserved under projective limits (as is continuity [7, I.4.4]):

LEMMA 1. *Let $(\Omega, \mathcal{A})$ be a measurable space, $\langle \mathcal{X}_I, \mathcal{B}_I, f_{JI}\rangle_D$ a projective family of measurable spaces with projective limit $(\mathcal{X}_D, \mathcal{B}_D)$, and $\langle w_I : \Omega \to \mathcal{X}_I\rangle_D$ a projective family of measurable mappings. Then the projective limit $w_D := \varprojlim w_I$ is a measurable mapping $\Omega \to \mathcal{X}_D$.*

PROOF (LEMMA 1). Since $\langle w_{\mathrm{I}}\rangle_{\mathrm{D}}$ is projective, $w_{\mathrm{I}} \circ f_{\mathrm{I}} = f_{\mathrm{I}} \circ w_{\mathrm{D}}$. By measurability of w_{I} and f_{I}, the composition $f_{\mathrm{I}} \circ w_{\mathrm{D}}$ is $\mathcal{A}$-$\mathcal{B}_{\mathrm{I}}$-measurable for all $I \in D$. Since the canonical mappings f_{I} generate $\mathcal{B}_{\mathrm{D}}$, w_{D} is $\mathcal{A}$-$\mathcal{B}_{\mathrm{D}}$-measurable. □

The regular conditional probabilities $P_{\mathrm{I}}[X_{\mathrm{I}}|\mathcal{C}_{\mathrm{I}}]$ can be regarded as a family of random measures, i.e. as measurable mappings $P_{\mathrm{I}} : \Omega \to M(\mathcal{X}_{\mathrm{I}})$ defined by $\omega \mapsto P_{\mathrm{I}}[X_{\mathrm{I}}|\mathcal{C}_{\mathrm{I}}](\omega)$. To prove Theorem 1, we argue that this family is projective (in the sense of Lemma 9), with the desired conditional probability $P_{\mathrm{D}}[X_{\mathrm{D}}|\mathcal{C}_{\mathrm{I}}]$ as its projective limit. However, we have to account for the fact projectivity of the mappings holds only almost everywhere.

Denote by $M(\mathcal{X}_{\mathrm{I}})$ the set of probability measures on $\mathcal{X}_{\mathrm{I}}$. The continuous mappings f_{JI} induce, by means of $P_{\mathrm{J}} \mapsto f_{\mathrm{JI}}(P_{\mathrm{J}})$, a continuous projection $f_{\mathrm{JI}} : M(\mathcal{X}_{\mathrm{J}}) \to M(\mathcal{X}_{\mathrm{I}})$. With respect to these projectors, the measurable mappings $P_{\mathrm{I}} : \Omega \to M(\mathcal{X}_{\mathrm{I}})$ are projective almost everywhere: For any pair $I \preceq J$ of indices, (3.2) holds up to a null set $N_{\mathrm{JI}} \subset \Omega$ of exceptions. Write $N := \cup_{I \preceq J} N_{\mathrm{JI}}$ for the aggregate null set, $N^{\mathrm{C}} := \Omega \setminus N$ for its complement. The restricted mappings $P_{\mathrm{I}}|_{N^{\mathrm{C}}} : N^{\mathrm{C}} \to M(\mathcal{X}_{\mathrm{I}})$ form a projective family of $\mathcal{C}_{\mathrm{D}} \cap N^{\mathrm{C}}$-measurable mappings, and by Lemma 1 have a unique, measurable projective limit $P_{\mathrm{D}}^{\setminus \mathrm{N}} : N^{\mathrm{C}} \to M(\mathcal{X}_{\mathrm{D}})$. This mapping satisfies

$$(f_{\mathrm{I}} P_{\mathrm{D}}^{\setminus \mathrm{N}})(\omega) = f_{\mathrm{I}}(P_{\mathrm{D}}^{\setminus \mathrm{N}}(\omega)) = P_{\mathrm{I}}(\omega) \qquad \text{for all } \omega \in N^{\mathrm{C}}\ . \tag{3.4}$$

The first identity is due to the definition of projective limit mappings; the second follows by observing that, for any $\omega \in N^{\mathrm{C}}$, $\langle P_{\mathrm{I}}(\omega)\rangle_{\mathrm{D}}$ is a projective family of probability measures with projective limit measure $P_{\mathrm{D}}^{\setminus \mathrm{N}}(\omega)$.

Since $P_{\mathrm{D}}^{\setminus \mathrm{N}} : N^{\mathrm{C}} \to M(V)$ is $\mathcal{C}_{\mathrm{D}} \cap N^{\mathrm{C}}$-measurable and since $M(V)$ is Polish, $P_{\mathrm{D}}^{\setminus \mathrm{N}}$ has an extension to a measurable function $P_{\mathrm{D}} : \Omega \to M(V)$ [27, Theorem 12.2]. This function $P_{\mathrm{D}}(\omega) =: P_{\mathrm{D}}[X_{\mathrm{D}}|\mathcal{C}_{\mathrm{D}}](\omega)$ is a regular conditional probability on $\mathcal{X}_{\mathrm{D}}$, and satisfies (3.3) $\mathbb{P}$-almost everywhere. □

Like projective limits, pullbacks generalize from measures to conditional probabilities. A simple point-wise application of pullbacks to the measures $P[\,.\,|\mathcal{C}](\omega)$ for $\omega \in \Omega$ yields:

Proposition 1 (Pullback of regular conditional probabilities). *Let $P[X|\mathcal{C}]$ be a regular conditional probability on a standard Borel space $\mathcal{X}$. Let $\tilde{\mathcal{X}}$ be a Hausdorff space, $\phi : \tilde{\mathcal{X}} \to \mathcal{X}$ injective, and $\tilde{\mathcal{B}} := \phi^{-1}\mathcal{B}(\mathcal{X})$ the induced σ-algebra on $\tilde{\mathcal{X}}$. Denote by $\tilde{\Omega} \subset \Omega$ the set of all ω satisfying $P^*[\phi(\tilde{\mathcal{X}})|\mathcal{C}](\omega) = 1$. Then $\nu(A, \omega) := P[\phi(A)|\mathcal{C}](\omega)$ is a probability kernel on $\tilde{\mathcal{X}}$, and can be regarded as a regular conditional probability of the random variable $\tilde{X} := \phi^{-1} \circ X|_{\tilde{\Omega}}$, given $\mathcal{C} \cap \tilde{\Omega}$.*

Clearly, $\tilde{\Omega}$ may be empty. Since $\tilde{\mathcal{B}}$ is the σ-algebra induced by ϕ and ϕ is injective, the inverse ϕ^{-1} is automatically measurable with respect to $\mathcal{B}(\mathcal{X}) \cap \phi(\tilde{\mathcal{X}})$, so the restriction of the mapping $\phi^{-1} \circ X$ is indeed a valid $\tilde{\mathcal{X}}$-valued random variable on $\tilde{\Omega}$. The combination of Theorem 1 and Lemma 1 results in a two-stage approach to the construction of regular conditional probabilities, analogous to the two-stage construction of stochastic processes in the sense of Definition 1: First construct a suitable projective limit $P_{\mathrm{D}}[X_{\mathrm{D}}|\mathcal{C}_{\mathrm{D}}]$, and then pull back to a (possibly non-measurable) subspace $\tilde{\mathcal{X}} \subset \mathcal{X}_{\mathrm{D}}$, or to a space $\tilde{\mathcal{X}}$ embedded into $\mathcal{X}_{\mathrm{D}}$ by a Borel embedding ϕ.

Both steps can be combined into a single step under an additional assumption – namely that the embedding of $\tilde{\mathcal{X}}$, i.e. the image $\phi(\tilde{\mathcal{X}})$, is actually measurable in $\mathcal{X}_{\mathrm{D}}$. The extension result obtained for this case can be regarded as a conditional probability analogue of the well-known projective limit theorem of Prokhorov [9, IX.4.2], just as Theorem 1 is analogous to the extension theorems of Kolmogorov and Bochner.

Corollary 1 (Prokhorov extension). *Let $\langle \mathcal{X}_I, \mathcal{B}_I, f_{JI}\rangle_D$ be a countably indexed projective system of Polish measurable spaces, $\tilde{\mathcal{X}}$ a Hausdorff space, $\phi : \tilde{\mathcal{X}} \to \mathcal{X}_D$ continuous and injective, and require $\phi(\tilde{\mathcal{X}}) \in \mathcal{B}_D$. Let $\langle P_I[X_I|\mathcal{C}_I]\rangle_D$ be a projective family of probability kernels on $\mathcal{B}_I \times \Omega$. Define $\tilde{\Omega}$ to be the*

subset $\tilde{\Omega} \subset \Omega$ of all ω for which the family of measures $\langle P_I[\,.\,|\mathcal{C}_I](\omega)\rangle_D$ satisfies the following "Prokhorov condition":
For all $\varepsilon > 0$, there is a compact set $K \subset \tilde{\mathcal{X}}$ such that

$$P_I[\phi_I K|\mathcal{C}_I](\omega) > 1 - \varepsilon \qquad \text{for all } I \in D\,. \tag{3.5}$$

Then there is a unique (up to equivalence) probability kernel $\tilde{P}[\,.\,|\tilde{\mathcal{C}}_D](\omega)$ on $\mathcal{B}(\tilde{\mathcal{X}}) \times \tilde{\Omega}$ with the projective family as its marginals, i.e. $\phi_I \tilde{P}[\,.\,|\mathcal{C}_D] =_{a.e.} P_I[\,.\,|\mathcal{C}_I]$. This probability kernel

1. *is a Radon measure for each $\omega \in \tilde{\Omega}$;*
2. *is the pullback of $P_D[\,.\,|\mathcal{C}_D] = \varprojlim \langle P_I[\,.\,|\mathcal{C}_I]\rangle_D$ under ϕ, and hence a conditional probability given $\tilde{\mathcal{C}}_D = \mathcal{C}_D \cap \tilde{\Omega}$.*

In the following sections, we will derive a number of results on how certain statistical properties of conditional models are preserved under projective limits and pullbacks. Statement (2) of the Corollary means that the constructed probability kernel $\tilde{P}[\,.\,|\tilde{\mathcal{C}}_{\mathrm{D}}]$ can effectively be decomposed into the projective limit $P_{\mathrm{D}}[\,.\,|\mathcal{C}_{\mathrm{D}}]$ and a subsequent pullback, which makes all results in the following immediately applicable.

Proof. For almost all $\omega \in \tilde{\Omega}$, the measures $P_{\mathrm{I}}[\,.\,|\mathcal{C}_{\mathrm{I}}](\omega)$ form a projective family and satisfy the Prokhorov condition. Since the spaces $\mathcal{X}_{\mathrm{I}}$ are Polish, each of these measures is a Radon measure. By Prokhorov's theorem [9, IX.4.2], there is a unique Radon probability measure ν_ω on $\tilde{\mathcal{X}}$ satisfying $\phi_{\mathrm{I}}(\nu_\omega) = P_{\mathrm{I}}[\,.\,|\mathcal{C}_{\mathrm{I}}](\omega)$. By Theorem 4, there is also a unique projective limit probability kernel $P_{\mathrm{D}}[\,.\,|\mathcal{C}_{\mathrm{D}}]$ on $\mathcal{X}_{\mathrm{D}} = \varprojlim \langle \mathcal{X}_{\mathrm{I}}\rangle_{\mathrm{D}}$. Since $\phi = \varprojlim \langle \phi_{\mathrm{I}}\rangle_{\mathrm{D}}$, we have $P_{\mathrm{D}}[\,.\,|\mathcal{C}_{\mathrm{D}}](\omega) = \phi(\nu_\omega)$ for almost all $\omega \in \tilde{\Omega}$. The image $\phi(\tilde{\mathcal{X}})$ is measurable, and so $P^*_{\mathrm{D}}[\phi(\tilde{\mathcal{X}})|\mathcal{C}_{\mathrm{D}}](\omega) = \nu_\omega(\phi^{-1}\phi\tilde{\mathcal{X}}) = \nu_\omega(\tilde{\mathcal{X}})$. Therefore, the pullback under ϕ exists, and by uniqueness has to coincide with ν_ω almost everywhere. □

The induced conditional probabilities $\tilde{P}[\,.\,|\mathcal{C}_{\mathrm{D}}]$ on $\tilde{\mathcal{X}}$ are regular, since measurability in ω carries over from $\mathcal{X}_{\mathrm{D}}$ under the pullback. This is remarkable in so far as virtually no requirements are imposed upon the space $\tilde{\mathcal{X}}$ – in particular, the topology of $\tilde{\mathcal{X}}$ need not admit a countable subbase – and conditional probabilities on $\tilde{\mathcal{X}}$ need not be regular in general. In other words, much as the Radon regularity of measures on a space which supports non-Radon probability measures is induced by the marginals, so is regularity of the conditional.

3.2. *Criteria for Projectivity.* To construct a conditional stochastic process $P_{\mathrm{D}}[X_{\mathrm{D}}|\mathcal{C}_{\mathrm{D}}]$ by means of Theorem 1 will in practice require proof that a given family of conditional probabilities is projective. The following two results provide criteria to verify projectivity.

LEMMA 2 (Criterion 1). *Let the random variables X_I satisfy $f_{JI}X_J = X_I$, and let $\langle \mathcal{C}_I \rangle_D$ be a directed family of σ-algebras. Then the family $\langle P_I[X_I|\mathcal{C}_I] \rangle_D$ is conditionally projective if and only if the random variables satisfy the conditional independence relations*

$$X_I \perp\!\!\!\perp_{\mathcal{C}_I} \mathcal{C}_J \qquad \text{for all } I \preceq J \; . \tag{3.6}$$

PROOF. By Doob's criterion for conditional independence [26, Proposition 6.6], $X_{\mathrm{I}} \perp\!\!\!\perp_{\mathcal{C}_{\mathrm{I}}} \mathcal{C}_{\mathrm{J}}$ is equivalent to $\mathbb{P}(A|\mathcal{C}_{\mathrm{I}}, \mathcal{C}_{\mathrm{J}}) =_{\text{a.e.}} \mathbb{P}(A|\mathcal{C}_{\mathrm{I}})$, for all $A \in \sigma(X_{\mathrm{I}})$. Hence

$$\begin{aligned} P_{\mathrm{J}}[f_{\mathrm{JI}}^{-1}A_{\mathrm{I}}|\mathcal{C}_{\mathrm{J}}] \;&=\; \mathbb{P}(X_{\mathrm{J}}^{-1}f_{\mathrm{JI}}^{-1}A_{\mathrm{I}}|\mathcal{C}_{\mathrm{J}}) \overset{f_{\mathrm{JI}}X_{\mathrm{J}}=X_{\mathrm{I}}}{=} \mathbb{P}(X_{\mathrm{I}}^{-1}A_{\mathrm{I}}|\mathcal{C}_{\mathrm{J}}) \\ &\overset{\text{Doob}}{=} \mathbb{P}(X_{\mathrm{I}}^{-1}A_{\mathrm{I}}|\mathcal{C}_{\mathrm{J}}) = \mathbb{P}(X_{\mathrm{I}}^{-1}A_{\mathrm{I}}|\mathcal{C}_{\mathrm{I}}) = P_{\mathrm{I}}[A_{\mathrm{I}}|\mathcal{C}_{\mathrm{I}}] \; . \end{aligned} \tag{3.7}$$

□

We recall that projectivity of conditional probabilities $P_{\mathrm{I}}[X_{\mathrm{I}}|\Theta_{\mathrm{I}}]$ as in (3.2) implies projectivity of the corresponding unconditional measures $P_{\mathrm{I}} = X_{\mathrm{I}}(\mathbb{P})$. Lemma 2 gives a necessary and sufficient condition for the converse to hold as well: If we assume the σ-algebras $\mathcal{C}_{\mathrm{I}}$ to be generated by parameter variables Θ_{I}, (3.6) takes the form $X_{\mathrm{I}} \perp\!\!\!\perp_{\Theta_{\mathrm{I}}} \Theta_{\mathrm{J}}$. For a fixed I, the criterion demands that – given full knowledge of Θ_{I} – information about the parameters corresponding to any other dimensions will not change our mind about X_{I}. If this is true for any I, the family is conditionally projective. The lemma implies a similar result by Lauritzen [36, IV, 3.1] on sufficient statistics: Since (3.6) is a necessary condition, any sufficient statistics $\langle S_{\mathrm{I}} \rangle_{\mathrm{D}}$ satisfy $X_{\mathrm{I}} \perp\!\!\!\perp_{S_{\mathrm{I}}} S_{\mathrm{I}}$ if the family of models is known to be projective.

In practice, a candidate family of finite-dimensional conditionals can be expected to be defined by densities, with respect to some family $\langle \nu_{\mathrm{I}} \rangle_{\mathrm{D}}$ of carrier measures. Now consider the special case where the projective system consists of product spaces $\mathcal{X}_{\mathrm{I}} = \prod_{i \in I} \mathcal{X}_{\{\mathrm{i}\}}$. The carrier measures will then typically be product measures, and proving that the family is projective will involve an application of Fubini's theorem. The following criterion makes this step generic.

LEMMA 3 (Criterion 2). *Let* $\langle P_I[X_I|\Theta_I]\rangle_D$ *be a family of conditional probabilities on a projective system* $\langle \prod_{i\in I}\mathcal{X}_{\{i\}}, \otimes_{i\in I}\mathcal{B}_{\{i\}}, pr_{JI}\rangle_D$*, where each* $\mathcal{X}_I$ *is Polish. Require: (1) For all* $I \in D$*, the conditional density* p_I *of* $P_I[X_I|\Theta_I]$ *with respect to a carrier measure* ν_I *on* $\mathcal{X}_I$ *exists. (2) The carrier measures are product measures* $\nu_I = \otimes_{i\in I}\nu_{\{i\}}$*. Then the family* $P_I[X_I|\Theta_I]$ *of conditionals is projective if and only if*

$$\int_{\mathcal{X}_{J\setminus I}} p_J(x_J|\theta_J)d\nu_{J\setminus I}(x_{J\setminus I}) = p_I(x_I|pr_{JI}\theta_J) \qquad \textit{whenever } I \preceq J\ . \tag{3.8}$$

Note $I \preceq J$ means $I \subset J$ for product spaces, and hence $J \setminus I \in D$.

PROOF. First suppose condition (3.8) is satisfied. To show that the family is projective, we have to show

$$P_{\mathrm{J}}[\mathrm{pr}_{\mathrm{JI}}^{-1}\,.\,|\Theta_{\mathrm{J}} = \theta^{\mathrm{J}}] = P_{\mathrm{I}}[\,.\,|\Theta_{\mathrm{I}} = \mathrm{pr}_{\mathrm{JI}}\theta^{\mathrm{J}}] \tag{3.9}$$

for all θ^{J} up to a null set. Denote by $p_{\mathrm{I}}(x_{\mathrm{I}}|\theta_{\mathrm{I}})$ the conditional density of $P_{\mathrm{I}}[X_{\mathrm{I}}|\Theta_{\mathrm{I}}]$. By Fubini's theorem,

$$\begin{aligned}\int_{\mathrm{pr}_{\mathrm{JI}}^{-1}A_{\mathrm{I}}} p_{\mathrm{J}}(x_{\mathrm{J}}|\theta_{\mathrm{J}})d\nu_{\mathrm{J}}(x_{\mathrm{J}}) &= \int_{A_{\mathrm{I}}}\Big(\int_{\mathcal{X}_{\mathrm{J}\setminus\mathrm{I}}} p_{\mathrm{J}}(x_{\mathrm{I}}, x_{\mathrm{J}\setminus\mathrm{I}}|\theta_{\mathrm{J}})d\nu_{\mathrm{J}\setminus\mathrm{I}}(x_{\mathrm{J}\setminus\mathrm{I}})\Big)d\nu_{\mathrm{J}}(x_{\mathrm{J}})\\ &= \int_{A_{\mathrm{I}}} p_{\mathrm{I}}(x_{\mathrm{I}}|\mathrm{pr}_{\mathrm{JI}}\theta_{\mathrm{J}})d\nu_{\mathrm{J}}(x_{\mathrm{J}})\end{aligned} \tag{3.10}$$

for all $A_{\mathrm{I}} \in \mathcal{B}_{\mathrm{I}}$. Hence, $P_{\mathrm{J}}[\mathrm{pr}_{\mathrm{JI}}^{-1}A_{\mathrm{I}}|\Theta_{\mathrm{J}} = \theta_{\mathrm{J}}] = P_{\mathrm{I}}[A_{\mathrm{I}}|\Theta_{\mathrm{I}} = \mathrm{pr}_{\mathrm{JI}}\theta_{\mathrm{J}}]$, which establishes the "if" implication. Conversely, assume that the family is projective. Abbreviate $a(x_{\mathrm{I}},\theta_{\mathrm{J}}) := \int_{\mathcal{X}_{\mathrm{J}\setminus\mathrm{I}}} p_{\mathrm{J}}(x_{\mathrm{J}}|\theta_{\mathrm{J}})d\nu_{\mathrm{J}\setminus\mathrm{I}}(x_{\mathrm{J}\setminus\mathrm{I}})$. For all $A_{\mathrm{I}} \in \mathcal{B}_{\mathrm{I}}$,

$$\int_{A_{\mathrm{I}}} a(x_{\mathrm{I}},\theta_{\mathrm{J}})d\nu_{\mathrm{I}}(x_{\mathrm{I}}) = \int_{A_{\mathrm{I}}} p_{\mathrm{I}}(x_{\mathrm{I}}|\theta_{\mathrm{I}})d\nu_{\mathrm{I}}(x_{\mathrm{I}}) = \int_{A_{\mathrm{I}}} p(x_{\mathrm{I}}|\mathrm{pr}_{\mathrm{JI}}\theta_{\mathrm{J}})d\nu_{\mathrm{I}}(x_{\mathrm{I}}) \tag{3.11}$$

The first identity follows from (3.9), the second one from the fact that $a(\,.\,,\theta_{\mathrm{J}})$ is $\mathcal{B}_{\mathrm{I}}$-measurable by Tonelli's theorem. Since a and p integrate identically over all A_{I} and are $\mathcal{B}_{\mathrm{I}}$-measurable, $a(\,.\,,\theta_{\mathrm{J}}) = p_{\mathrm{I}}(\,.\,|\mathrm{pr}_{\mathrm{JI}}\theta_{\mathrm{J}})$ ν_{I}-a.s. □

4. Application to Bayesian Models. A Bayesian model is completely defined by a pair of conditional probabilities. The results of the previous section therefore provide the formal means of defining projective limits of Bayesian models. In conjunction with pullbacks under Borel embeddings, we can then represent nonparametric Bayesian models by projective families of finite-dimensional Bayesian models. Since the term "parametric" often implies a finite-dimensional or dominated model, we will use the term "parametrized" to describe a statistical model indexed by a parameter, regardless of whether the dimension of the parameter is finite or infinite.

4.1. *Parametrized and Bayesian Models.* We briefly recall the formal notion of model and parameter; a detailed discussion is given by Schervish [47, Ch. 1.5.5]. Let $X : \Omega \to \mathcal{X}$ be a random variable with values in a Polish space $\mathcal{X}$, such that $P^\infty = X^\infty(\mathbb{P})$ is exchangeable. Let $M(\mathcal{X})$ be the set of probability measures on $\mathcal{X}$, and denote by $F : \mathcal{X}^\infty \to M(\mathcal{X})$ the mapping induced by the empirical measure. Let T be a parametric index, i.e. a bimeasurable mapping from the image $(F \circ X^\infty)(\Omega) \subset M(\mathcal{X})$ onto a measurable space $(\mathcal{T}, \mathcal{B}_\mathcal{T})$. Then the derived random variable $\Theta := T \circ F \circ X^\infty$ is called a *parameter*, and we call the regular conditional probability $P[X|\Theta])$ a *parametrized model* of X. The model corresponds to a subset $\mathcal{P} := \{P[X|\Theta = \theta] \,|\, \theta \in \mathcal{T}\}$ of measures. In summary,

$$\Omega \xrightarrow{\quad X^\infty \quad} \mathcal{X}^\infty \xrightarrow{\quad F \quad} M(\mathcal{X}) \supset \mathcal{P} \xrightarrow{\quad T \quad} \mathcal{T} \,. \tag{4.1}$$

The assumption that $\mathcal{X}$ is Polish guarantees both the existence of regular conditional probabilities on $\mathcal{X}$ and the validity of de Finetti's theorem [26, Theorem 11.10]. The theorem implies a law of large numbers, which in turn guarantees convergence $\lim_n F_n(X^n) \to X(\mathbb{P})$ of the empirical measure in the weak* topology on $M(\mathcal{X})$, and hence ensures that F is well-defined.

The parameter random variable Θ induces an image measure $P^\theta = \Theta(\mathbb{P})$ on the parameter space $(\mathcal{T}, \mathcal{B}_\mathcal{T})$. For any given atomic random event $\omega \in \Omega$, the corresponding value $\theta = \Theta(\omega)$ of the parameter is completely determined by $X^\infty(\omega)$, as the image under $T \circ F$. The partial information about $\Theta(\omega)$ contained in a finite sample $X^n(\omega) = x^n$ can be conditioned upon as $P^\theta[\Theta|X^n = x^n]$, such that under suitable conditions the actual value $\theta = \Theta(\omega)$ is asymptotically recovered as $P^\theta[\Theta|X^n = x^n] \xrightarrow{n\to\infty} \delta_\theta$. In this context, $P^\theta(\Theta)$ is referred to as a *prior distribution*, $P[X|\Theta]$ as a *sampling model* or *likelihood*, and $P^\theta[\Theta|X]$ as the *posterior* under observation X. Additionally, the prior can be represented as a parametrized model $P^\theta[\Theta|Y = y]$, where Y is a *hyperparameter*. We refer to the whole system summarily as the *Bayesian model* defined by $P[X|\Theta]$ and $P^\theta[\Theta|Y]$. For our purposes, it is sufficient to assume that the prior is the "true" prior, i.e. the actual image measure under Θ. If the sampling model is dominated, Bayes' theorem is applicable, and the posterior can be represented by the density $\frac{p(x|\theta)}{p(x)}$ with respect to the prior. For undominated models, notably the Dirichlet process, some alternative to Bayes' theorem is required. In Bayesian nonparametrics, this alternative is usually conjugacy (Sec. 6).

4.2. *Application of Projective Limits.* Suppose that, for a projective system of sample spaces $\langle \mathcal{X}_\mathrm{I}, f_\mathrm{JI} \rangle_\mathrm{D}$, a parametrized model is given on each space: Each object in (4.1), except for the abstract probability space Ω, is equipped with an index I.

$$\begin{array}{ccccccc} & \overset{X_{\mathrm{J}}^{\infty}}{\nearrow} & \mathcal{X}_{\mathrm{J}}^{\infty} & \xrightarrow{F_{\mathrm{J}}} & M(\mathcal{X}_{\mathrm{J}}) \supset \mathcal{P}_{\mathrm{J}} & \xrightarrow{T_{\mathrm{J}}} & \mathcal{T}_{\mathrm{J}} \\ \Omega & & \downarrow f_{\mathrm{JI}} & & \downarrow f_{\mathrm{JI}} & & \vdots\, g_{\mathrm{JI}} \\ & \underset{X_{\mathrm{I}}^{\infty}}{\searrow} & \mathcal{X}_{\mathrm{I}}^{\infty} & \xrightarrow{F_{\mathrm{I}}} & M(\mathcal{X}_{\mathrm{I}}) \supset \mathcal{P}_{\mathrm{I}} & \xrightarrow{T_{\mathrm{I}}} & \mathcal{T}_{\mathrm{I}} \end{array} \tag{4.2}$$

The mappings $f_{\mathrm{JI}} : \mathcal{X}_{\mathrm{J}} \to \mathcal{X}_{\mathrm{I}}$ induce projections $f_{\mathrm{JI}} : \mathcal{X}_{\mathrm{J}}^{\infty} \to \mathcal{X}_{\mathrm{I}}^{\infty}$ and $f_{\mathrm{JI}} : M(\mathcal{X}_{\mathrm{J}}) \to M(\mathcal{X}_{\mathrm{I}})$. If $f_{\mathrm{JI}}\mathcal{P}_{\mathrm{J}} = \mathcal{P}_{\mathrm{I}}$, bimeasurability of T_{J} implies f_{JI} has a unique, measurable pushforward $g_{\mathrm{JI}} := T_{\mathrm{I}} \circ f_{\mathrm{JI}} \circ T_{\mathrm{J}}^{-1}$ on $\mathcal{T}_{\mathrm{J}}$. Hence, if the conditional probabilities $P_{\mathrm{I}}[\,.\,|\Theta_{\mathrm{I}}]$ defining the parametrized models $\mathcal{P}_{\mathrm{I}}$ are projective, we obtain $f_{\mathrm{JI}}P_{\mathrm{J}}[\,.\,|\Theta_{\mathrm{J}} = \theta_{\mathrm{J}}] = P_{\mathrm{I}}[\,.\,|\Theta_{\mathrm{I}} = g_{\mathrm{JI}}\theta_{\mathrm{J}}]$. By applying a projective limit to all spaces, mappings and conditionals which are indexed by I in (4.2), we obtain a projective limit system of the form (4.1) (with each quantity indexed by D, respectively). The resulting diagram again constitutes a parametrized model. We refer to this model as a *projective limit model* in the following.

The definition immediately carries over to Bayesian models: A projective system of Bayesian models is defined by three projective systems of standard Borel spaces $\langle \mathcal{X}_{\mathrm{I}}, \mathcal{B}(\mathcal{X}_{\mathrm{I}}), f_{\mathrm{JI}} \rangle_{\mathrm{D}}$, $\langle \mathcal{T}_{\mathrm{I}}, \mathcal{B}(\mathcal{T}_{\mathrm{I}}), g_{\mathrm{JI}} \rangle_{\mathrm{D}}$ and $\langle \mathcal{Y}_{\mathrm{I}}, \mathcal{B}(\mathcal{Y}_{\mathrm{I}}), h_{\mathrm{JI}} \rangle_{\mathrm{D}}$, and by projective families $\langle P_{\mathrm{I}}^{x}[X_{\mathrm{I}}|\Theta_{\mathrm{I}}] \rangle_{\mathrm{D}}$ and $\langle P_{\mathrm{I}}^{\theta}[\Theta_{\mathrm{I}}|Y_{\mathrm{I}}] \rangle_{\mathrm{D}}$ of conditional distributions. The uniqueness up to equivalence of the projective limit conditionals (Theorem 1) implies that the following diagram commutes:

$$\begin{array}{ccc} P_{\mathrm{D}}^{\theta}[\Theta_{\mathrm{D}}|Y_{\mathrm{D}}] & \xrightarrow{X_{\mathrm{D}}^{n} = x_{\mathrm{D}}^{n}} & P_{\mathrm{D}}^{\theta}[\Theta_{\mathrm{D}}|X_{\mathrm{D}}^{n}, Y_{\mathrm{D}}] \\ \downarrow g_{\mathrm{I}} & & \downarrow g_{\mathrm{I}} \\ P_{\mathrm{I}}^{\theta}[\Theta_{\mathrm{I}}|Y_{\mathrm{I}}] & \xrightarrow{X_{\mathrm{I}}^{n} = x_{\mathrm{I}}^{n}} & P_{\mathrm{I}}^{\theta}[\Theta_{\mathrm{I}}|X_{\mathrm{D}}^{n}, Y_{\mathrm{I}}] \end{array} \tag{4.3}$$

In other words, we obtain the same posterior regardless of whether we (i) take projective limits of the finite-dimensional models and then compute the infinite-dimensional posterior, or (ii) compute all finite-dimensional posteriors under marginal observations and take the projective limit.

4.3. *Application of pullbacks.* Pullbacks can, in principle, be applied to both parametrized and Bayesian models in a manner analogous to projective limits. However, the pullback in general results in a restriction of the abstract probability space, which formally has to be taken into account: If a probability measure $P = X(\mathbb{P})$ is pulled back under an injective map $\phi : \tilde{\mathcal{X}} \to \mathcal{X}$, the resulting random variable $\tilde{X} = \phi^{-1} \circ X$ is only defined on $\tilde{\Omega} := X^{-1}\phi(\tilde{\mathcal{X}})$. As a subset of the abstract probability space, $\tilde{\Omega}$ can always be assumed measurable, and we will for simplicity assume that it is not a null set. The corresponding restriction $\tilde{\mathbb{P}}$ of the abstract probability measure $\mathbb{P}$ is the conditional $\tilde{\mathbb{P}}(\,.\,) = \mathbb{P}(\,.\,|\tilde{\Omega})$, i.e. the abstract probability space underlying the pullback measure $\tilde{P} = \tilde{X}(\tilde{\mathbb{P}})$ is $(\tilde{\Omega}, \mathcal{A} \cap \tilde{\Omega}, \tilde{\mathbb{P}})$.

Consider the parametrized model $P[X|\Theta]$ described by (4.1). In this case, the entire diagram (4.1) may be pulled back to obtain

$$
\begin{array}{ccccccc}
\tilde{\Omega} & \xrightarrow{\tilde{X}^\infty} & \tilde{\mathcal{X}}^\infty & \xrightarrow{\tilde{F}} & M(\tilde{\mathcal{X}}) \supset \tilde{\mathcal{P}} & \xrightarrow{\tilde{T}} & \tilde{\mathcal{T}} \\
\Big\downarrow \mathcal{J}_\Omega & & \Big\downarrow \phi & & & & \Big\downarrow \mathcal{J}_{\mathcal{T}} \\
\Omega & \xrightarrow{X^\infty} & \mathcal{X}^\infty & \xrightarrow{F} & M(\mathcal{X}) \supset \mathcal{P} & \xrightarrow{T} & \mathcal{T}
\end{array}
\tag{4.4}
$$

The pullback is applicable only for those values $\Theta = \theta$ with $P^*[\tilde{\mathcal{X}}|\Theta = \theta] = 1$. Let $\tilde{\mathcal{T}} \subset \mathcal{T}$ be the set of such values. Denote the corresponding set of pullbacks $\tilde{\mathcal{P}}$. The restriction $\tilde{\Omega}$ induced by the pullback is represented in the diagram by the canonical inclusion mapping $\mathcal{J}_\Omega$. The mappings $\tilde{X}$, $\tilde{F}$ and $\tilde{T}$ are the restrictions of the mappings X, F and T to the respective restricted domains. Whenever $\theta \in \tilde{\mathcal{T}}$, we write $\tilde{P}[\,.\,|\tilde{\Theta} = \theta]$ for the pullback measure of $P[\,.\,|\Theta = \theta]$. This notation as a conditional probability is justified by the following lemma.

Lemma 4. *Let $(\tilde{A}, \theta) \mapsto \tilde{\nu}_\theta(\tilde{A})$ be the pullback of $P[X|\Theta]$.*

1. *The parameter $\tilde{\Theta} := \tilde{T} \circ \tilde{F} \circ \tilde{X}^\infty$ of the pullback model coincides with the restriction of the parameter Θ, that is, $\tilde{\Theta} = \Theta|_{\tilde{\Omega}}$ and $\tilde{\Theta}(\tilde{\Omega}) = \tilde{\mathcal{T}}$.*
2. *The function $(\tilde{A}, \omega) \mapsto \tilde{\nu}_{\Theta(\omega)}(\tilde{A})$ is a version of the regular conditional probability $\tilde{\mathbb{P}}[\tilde{X} \in \tilde{A}|\sigma(\tilde{\Theta})](\omega)$.*

Proof. (1) By the definition of $\tilde{X}^\infty$, $\tilde{F}$ and $\tilde{T}$ as pullbacks, $\tilde{\Theta} = \Theta|_{\tilde{\Omega}}$ holds if $\tilde{\Omega} \subset \mathrm{Dom}(\tilde{\Theta})$. Suppose $\omega \in \tilde{\Omega}$. Then $X^\infty(\omega) \in \tilde{\mathcal{X}}^\infty$, and the corresponding measure $P_\omega = F(X^\infty(\omega))$ concentrates on $\tilde{\mathcal{X}}$ in the sense that $P(A) = 1$ whenever $\tilde{\mathcal{X}} \subset A$. Since the outer measure of $\tilde{\mathcal{X}}$ is an infimum over such sets, $P^*_\omega(\tilde{\mathcal{X}}) = 1$ and hence $\Theta(\omega) \in \tilde{\mathcal{T}}$ and $\omega \in \mathrm{Dom}(\tilde{\Theta})$.

To show $\tilde{\Theta}(\tilde{\Omega}) = \tilde{\mathcal{T}}$, suppose $\theta \in \tilde{\mathcal{T}}$. Let $\nu(\,.\,) = P[\,.\,|\Theta = \theta]$. Then $\nu^*(\tilde{\mathcal{X}}) = 1$ and $\tilde{\nu} \in \tilde{\mathcal{P}}$, hence $\tilde{\Theta}^{-1}\{\theta\} \supset (\tilde{F} \circ \tilde{X}^\infty)^{-1}\{\tilde{\nu}\} \neq \emptyset$. This implies $\tilde{\Theta}(\tilde{\Omega}) = \tilde{\mathcal{T}}$, since $\tilde{\Theta}(\tilde{\Omega}) \subset \tilde{\mathcal{T}}$ by definition.
(2) For $A \in \mathcal{B}_x$ fixed, $\omega \mapsto \tilde{\nu}_{\Theta(\omega)}$ is the restriction $\tilde{\nu}_{\Theta(\,.\,)}(A \cap \tilde{\mathcal{X}}) = P[A|\Theta]|_{\tilde{\Omega}}$ of the integrable, $\sigma(\Theta)$-measurable function $P[A|\Theta](\,.\,)$. Since $\sigma(\tilde{\Theta}) = \sigma(\Theta) \cap \tilde{\Omega}$, the function $\omega \mapsto \tilde{\nu}_{\Theta(\omega)}(A \cap \tilde{\mathcal{X}})$ is $\sigma(\tilde{\Theta})$-measurable for every $A \in \mathcal{B}_x$. The pullback of an integrable function preserves the integral (cf. (A.5)). Hence,

the function is a conditional given $\sigma(\Theta)$, as for any $C \in \sigma(\Theta)$,

$$\int_{C\cap\tilde{\Omega}} \tilde{\nu}_{\Theta(\,.\,)}(A \cap \tilde{\mathcal{X}})d\tilde{\mathbb{P}}(\omega) = \int_C \mathbb{P}[X^{-1}A|\Theta](\omega)d\mathbb{P}(\omega) \\ = \mathbb{P}(A \cap C) = \tilde{\mathbb{P}}((X^{-1}A \cap \tilde{\Omega}) \cap (C \cap \tilde{\Omega}))\ . \tag{4.5}$$

□

For a Bayesian model, the pullback is consecutively applied to the sampling model and to the prior. The pullback of $P[X|\Theta]$ induces a restriction of the parameter space from $\mathcal{T}$ to $\tilde{\mathcal{T}}$. Lemma 4 guarantees that the induced random variable $\tilde{\Theta}$ is indeed the parameter variable of the resulting model. The prior family $P^{\theta}[\Theta|Y]$ can hence be pulled back under $\mathcal{J}_{\mathcal{T}}$. The pullback exists for all $y \in \mathcal{Y}$ with $P^{\theta,*}[\tilde{\mathcal{T}}|Y = y] = 1$, which in turn, by another applications of Lemma 4, induces a restriction of the hyperparameter space to $\tilde{\mathcal{Y}} \subset \mathcal{Y}$.

5. Sufficient Statistics. The purpose of this section is to show that the application of sufficient statistics commutes with the application of projective limits and pullbacks. If each element of a projective family of parametrized models admits a sufficient statistic, the projective limit of these functions is a sufficient statistic for the projective limit model. Similarly, the pullback of the sufficient statistic is a sufficient statistic for the pullback model.

Definition 3 (Sufficient statistic [22]). Let $P[X|\Theta]$ be a regular conditional probability. A σ-algebra $\mathcal{S} \subset \mathcal{A}$ is called *sufficient* for $P[X|\Theta]$ if there is a probability kernel $k : \mathcal{A} \times \Omega \to [0, 1]$, such that (i) $\omega \mapsto k(B, \omega)$ is $\mathcal{S}$-measurable for all $B \in \mathcal{B}_x$, and (ii)

$$P[B|\Theta, \mathcal{S}](\omega) = k(X^{-1}B, \omega) \qquad \mathbb{P}\text{-a.s.}, B \in \mathcal{B}(\mathcal{X})\ . \tag{5.1}$$

If $\mathcal{S}$ is sufficient and $(\mathcal{U}, \mathcal{B}(\mathcal{U}))$ is a measurable Polish space, then a measurable mapping $S : \mathcal{X} \to \mathcal{U}$ is called a *sufficient statistic* for $P[X|\Theta]$ if $\mathcal{S} = \sigma(S \circ X)$.

Theorem 2 (Sufficient σ-algebras and projective limits). *Consider a projective limit model $P_D[X_D|\Theta_D] = \varprojlim \langle P_I[X_I|\Theta_I]\rangle_D$. For each $I \in D$, let $\mathcal{S}_I \subset \mathcal{A}$ be a sufficient σ-algebra for $P_I[X_I|\Theta_I]$.*

1. *If $\langle \mathcal{S}_I\rangle_D$ is directed, $\mathcal{S}_D := \sigma(\mathcal{S}_I; I \in D)$ is sufficient for $P_D[X_D|\Theta_D]$.*
2. *Let $\langle \mathcal{U}_I, \mathcal{B}(\mathcal{U}_I), h_{JI}\rangle_D$ be projective measurable spaces and $S_I : \mathcal{X}_I \to \mathcal{U}_I$ projective measurable mappings. If each S_I is sufficient for $P_I[X_I|\Theta_I]$, a sufficient statistic for $P_D[X_D|\Theta_D]$ is given by $S_D := \varprojlim \langle S_I\rangle_D$.*

3. *Conversely, if any $\mathcal{C} \subset \mathcal{A}$ is sufficient for a projective limit model $P_D[X_D|\Theta_D]$, it is sufficient for all marginals $P_I[X_I|\Theta_I]$.*

PROOF. (1) By Doob's criterion for conditional independence [26, Proposition 6.6], (5.1) is equivalent to $X_{\text{I}} \perp\!\!\!\perp_{\mathcal{S}_{\text{I}}} \Theta_{\text{I}}$, which implies $X_{\text{I}} \perp\!\!\!\perp_{\mathcal{S}_{\text{D}}} \Theta_{\text{I}}$ for all $I \in D$. By Lemma 2, we additionally have $X_{\text{I}} \perp\!\!\!\perp_{\mathcal{S}_{\text{D}}} \Theta_{\text{J}}$ whenever $I \preceq J$. Since $\sigma(\Theta_{\text{D}}) = \sigma(\Theta_{\text{I}}; I \in D)$, we deduce $X_{\text{I}} \perp\!\!\!\perp_{\mathcal{S}_{\text{D}}} \Theta_{\text{D}}$. Similarly, $\sigma(X_{\text{D}}) = \sigma(X_{\text{I}}; I \in D)$ yields $X_{\text{D}} \perp\!\!\!\perp_{\mathcal{S}} \Theta_{\text{D}}$, which is just sufficiency of $\mathcal{S}_{\text{D}}$.
(2) Since the functions S_{I} are projective, the family $\langle \sigma(S_{\text{I}}) \rangle_{\text{D}}$ of σ-algebras is directed, and $\sigma(S_{\text{D}}) = \sigma(S_{\text{I}}; I \in D)$. The claim follows from (1).
(3) Let k_{D} be the kernel for which $\mathcal{C}$ and $P_{\text{D}}[X_{\text{D}}|\Theta_{\text{D}}]$ satisfy (5.1). We need to derive a suitable kernel k_{I} for each $I \in D$. By Lemma 2, the marginals satisfy $X_{\text{I}} \perp\!\!\!\perp_{\Theta_{\text{I}}} \Theta_{\text{J}}$, and hence $X_{\text{I}} \perp\!\!\!\perp_{(\Theta_{\text{I}}, \mathcal{C})} (\Theta_{\text{J}})$. Since trivially also $X_{\text{I}} \perp\!\!\!\perp_{(\Theta_{\text{I}}, \mathcal{C})} \mathcal{C}$, the chain rule yields $X_{\text{I}} \perp\!\!\!\perp_{(\Theta_{\text{I}}, \mathcal{C})} (\Theta_{\text{J}}, \mathcal{C})$. Again by Lemma 2, the latter implies

$$(f_{\text{I}} P_{\text{D}})[X_{\text{D}}|\Theta_{\text{D}}, \mathcal{C}] =_{\text{a.e.}} P_{\text{I}}[X_{\text{I}}|\Theta_{\text{I}}, \mathcal{C}] \ . \tag{5.2}$$

Since $\mathcal{C}$ is sufficient for $P_{\text{D}}[X_{\text{I}}|\Theta_{\text{D}}]$, the left-hand side of (5.2) is equal to a kernel $f_{\text{I}} k_{\text{D}}$. Hence, $k_{\text{I}} := f_{\text{I}} k_{\text{D}}$ is a $\mathcal{C}$-measurable kernel satisfying $k_{\text{I}}(A_{\text{I}}, \omega) =_{\text{a.e.}} P_{\text{I}}[A_{\text{I}}|\Theta_{\text{I}}, \mathcal{C}](\omega)$, which makes $\mathcal{C}$ sufficient for $P_{\text{I}}[X_{\text{I}}|\Theta_{\text{I}}]$. □

To make the construction of the sufficient statistics of a parametrized stochastic process fully compatible with the construction of the process itself requires an analogous result for pullbacks.

PROPOSITION 2 (Sufficiency and pullbacks). *Let $\tilde{P}[\tilde{X}|\tilde{\Theta}]$ be the pullback of a parametrized model $P[X|\Theta]$ under $\mathcal{J}_{\mathcal{X}} : \tilde{\mathcal{X}} \to \mathcal{X}$. If $S : \mathcal{X} \to \mathcal{U}$ is a sufficient statistic for $P[X|\Theta]$, then $\tilde{S} := S \circ \mathcal{J}_{\mathcal{X}}$ is sufficient for $\tilde{P}[\tilde{X}|\tilde{\Theta}]$.*

PROOF. We have to show that $\tilde{P}[\tilde{X}|\tilde{\Theta}]$ and $\tilde{S}$ satisfy (5.1) for a suitable kernel $\tilde{k}$, which we define as follows. Let $k(A, \omega)$ be the kernel (5.1) for S and $P[X|\Theta]$. The domain of $\tilde{S}$ is $\tilde{S}^{-1}\mathcal{U} = \tilde{\Omega}$. For any $A \in \sigma(X)$, define $\tilde{k}$ as $\tilde{k}(A \cap \tilde{\Omega}, \,.\,) := k(A, \,.\,)|_{\tilde{\Omega}}$. By definition, $\omega \mapsto \tilde{k}(A \cap \tilde{\Omega}, \omega)$ is measurable with respect to $\sigma(S \circ X) \cap \tilde{\Omega} = \sigma(\tilde{S} \circ \tilde{X})$. Hence, (5.1) is satisfied if the integral of $\tilde{k}$ matches that of the conditional probability for each set in $\sigma(\tilde{\Theta}, \tilde{X} \circ \tilde{S}) = \sigma(\Theta, X \circ S) \cap \tilde{\Omega}$. Let $\tilde{C} = C \cap \tilde{\Omega}$ be any such set. As pullbacks preserve integrals in the sense of (A.5),

$$\begin{aligned}\int_{\tilde{C}} \tilde{k}(\tilde{A}, \tilde{S} \circ \tilde{X}(\omega)) d\tilde{\mathbb{P}}(\omega) &= \int_C k(A, S \circ X(\omega)) d\mathbb{P}(\omega) = \mathbb{P}(X^{-1}A \cap C) \\ &= \tilde{\mathbb{P}}(X^{-1}A \cap C \cap \tilde{\Omega}) = \int_{\tilde{C}} \tilde{P}[\tilde{A}|\tilde{\Theta}, \tilde{S}](\omega) d\tilde{\mathbb{P}}(\omega) \ .\end{aligned}$$

Since $\tilde{S}$, $\tilde{k}$ and $\tilde{P}[\tilde{X}|\tilde{\Theta}]$ satisfy (5.1), $\tilde{S}$ is sufficient for $\tilde{P}[\tilde{X}|\tilde{\Theta}]$. □

We conclude this section with a result on minimality, i.e. the question whether a "smallest" sufficient σ-algebra exists for a given model. The concept is closely related to that of a minimal sufficient statistic – a sufficient statistic to which any statistic sufficient for the model can be reduced by transformation – but the two are not equivalent [33].

DEFINITION 4 (Minimal sufficient σ-algebra [33]). A σ-algebra $\mathcal{S}_0 \subset \mathcal{A}$ is called *minimal sufficient* for $P[X|\Theta]$ if it is sufficient, and if every other sufficient σ-algebra $\mathcal{C}$ satisfies:

$$\forall A \in \mathcal{S}_0 \exists C \in \mathcal{C} : \qquad P[A \triangle C | \Theta = \theta] = 0 \qquad \text{for all } \theta \in \mathcal{T} \,. \tag{5.3}$$

Intuitively, minimality captures the idea that any σ-algebra $\mathcal{C}$ can only be sufficient for the model if it contains all information contained in $\mathcal{S}_0$ (though this interpretation is inaccurate in the undominated case, as pointed out by Burkholder [10]). However, instead of demanding $\mathcal{S}_0 \subset \mathcal{C}$, and hence that every set in $\mathcal{S}_0$ is also in $\mathcal{C}$, we only require that each set in $\mathcal{S}_0$ be indistinguishable from a set in $\mathcal{C}$ with in the resolution of the model.

A minimal sufficient σ-algebra always exists if the model $P[X|\Theta]$ in question is dominated. In undominated models, a sufficient σ-algebra can – rather contrary to intuition – be contained in a *finer* σ-algebra which is *not* sufficient, and a minimal sufficient σ-algebra need not exist [10]. However, as the following theorem shows, existence is guaranteed if the model is constructed as a projective limit from dominated marginals. This implies, for example, that the Dirichlet process on the line admits a minimal sufficient σ-algebra, even though it is undominated.

PROPOSITION 3 (Minimal sufficiency). *Suppose that each σ-algebra $\mathcal{S}_I$ as specified in Theorem 2 is minimal sufficient for $P_I[X_I|\Theta_I]$. Then $\mathcal{S}_D = \varprojlim \langle \mathcal{S}_I \rangle_D$ is minimal sufficient for $P_D[X_D|\Theta_D] = \varprojlim \langle P_I[X_I|\Theta_I] \rangle_D$.*

PROOF. $\mathcal{S}_{\mathrm{D}}$ is sufficient by Theorem 2; we have to verify (5.3). Let $\mathcal{C} \subset \mathcal{A}$ be any sufficient σ-algebra for $P_{\mathrm{D}}[X_{\mathrm{D}}|\Theta_{\mathrm{D}}]$. By Theorem 2, $\mathcal{C}$ is sufficient for all $P_{\mathrm{I}}[X_{\mathrm{I}}|\Theta_{\mathrm{I}}]$, which implies that (5.3) is satisfied if $A \in \cup_{\mathrm{I}} \mathcal{S}_{\mathrm{I}}$. For the general case $A \in \mathcal{S}_{\mathrm{D}}$, observe that the set system $\cup_{\mathrm{I}} \mathcal{S}_{\mathrm{I}}$ is both an algebra and a generator of $\mathcal{S}_{\mathrm{D}}$. By the basic theorem on approximation of a measure on a subalgebra [2, Theorem 5.7]), any set $A \in \mathcal{S}_{\mathrm{D}}$ can hence be approximated by a sequence of sets $A_n \in \cup_{\mathrm{I}} \mathcal{S}_{\mathrm{I}}$ such that $\lim_n \mathbb{P}[A_n \triangle A | \Theta = \theta] = 0$. Since each A_n satisfies (5.3), there is a corresponding set $C_n \in \mathcal{C}$ such that $\mathbb{P}[A_n \triangle C_n | \Theta = \theta] = 0$. Then $\lim_n \mathbb{P}[A \triangle C_n | \Theta = \theta] = 0$, and therefore $\mathbb{P}[A \triangle \cup_n C_n | \Theta = \theta] = 0$. Since $\mathcal{C}$ is a σ-algebra, $\cup_n C_n \in \mathcal{C}$, and A satisfies (5.3) with $C := \cup_n C_n$. □

6. Conjugacy. The posterior of a Bayesian model is a regular conditional probability, and always exists if the model is defined on Polish spaces. However, since the abstract components of the model – the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and the random variables X and Θ – are not given explicitly, there is in general no way to deduce the posterior from the sampling distribution and the prior. The problem is solved by Bayes' theorem whenever the sampling distribution is dominated, i.e. if $P[X|\Theta]$ has a conditional density [47, Theorem 1.31]. This need not be the case in the infinite-dimensional setting of Bayesian nonparametrics. For a certain class of Bayesian models, so-called *conjugate* models, the posterior can be specified without appealing to Bayes' theorem. Virtually all nonparametric Bayesian models studied in the literature are of this type (see e.g. [50]).

Definition 5 (Conjugate Bayesian model). Let $P[X|\Theta]$ and $P^\theta[\Theta|Y]$ specify a Bayesian model. Let $(T^{(n)})_n$ be a family of measurable mappings $T^{(n)} : \mathcal{X}^n \times \mathcal{Y} \to \mathcal{W}$ with values in a Polish space. The family is called a *posterior index* of the model if there exists a probability kernel $k : \mathcal{B}_{\mathcal{T}} \times \mathcal{W} \to [0,1]$ such that

$$P^\theta[A|X^n = (x_1, \dots, x_n), Y = y] =_{\text{a.e.}} k(A, T^{(n)}(x_1, \dots, x_n, y)) \tag{6.1}$$

for every $A \in \mathcal{B}_{\mathcal{T}}$. A Bayesian model is called *conjugate* if there is a posterior index for which $\mathcal{W} \subset \mathcal{Y}$ and the associated kernel k satisfies

$$k(A, y') = P^\theta[A|Y = y'] \qquad \text{for all } y' \in \mathcal{Y}\,. \tag{6.2}$$

Apparently, any Bayesian model admits the identity $T^{(n)} := \mathrm{Id}_{\mathcal{X}^n \times \mathcal{Y}}$ as a trivial posterior index. By (6.2), a conjugate posterior is "in the same family" as the prior, a model property commonly referred to as *closure under sampling* [43].

The following theorem states that the posterior updates of a nonparametric Bayesian model have the same "functional form" as those of its finite-dimensional marginals. It also implies that conjugacy of an infinite-dimensional model requires conjugate marginals.

Theorem 3 (Conjugacy in projective limit models). *Let $\langle P_I[X_I|\Theta_I]\rangle_D$ and $\langle P_I^\theta[\Theta_I|Y_I]\rangle_D$ define a projective family of Bayesian models.*

1. *Let $(T_I^{(n)})_n$ be posterior indices and projective, i.e.*

$$T_I^{(n)} \circ (h_{JI} \otimes f_{JI}^n) = h_{JI} \circ T_J^{(n)} \qquad \text{for } I \in D, n \in \mathbb{N}. \tag{6.3}$$

Then the mappings $T_D^{(n)} := \varprojlim \langle T_I^{(n)} \rangle_D$ *form a posterior index of the projective limit model. If each marginal model is conjugate under* $(T_I^{(n)})_n$*, the projective limit model is conjugate under* $(T_D^{(n)})_n$.

2. *Conversely, let the projective limit be conjugate. If the canonical mappings* f_I*,* h_I *are surjective, the marginals of the model are closed under sampling. If additionally each of the mappings* f_I *and* h_I *is open or closed, the marginals are conjugate and their posterior indices are pushforwards of* $(T_D^{(n)})_n$ *satisfying*

$$T_I^{(n)} \circ (f_I^n \otimes h_I) = h_I \circ T_D^{(n)} \qquad \text{for } I \in D, n \in \mathbb{N}\,. \tag{6.4}$$

(The proof is given in App. B.) Consequently, a conjugate nonparametric Bayesian model can only be obtained from marginals which are closed under sampling. Dropping either of the two assumptions in the theorem – that the model is defined as a projective limit and that the canonical mappings be surjective – does not lift this restriction. If the model is not explicitly assumed to be a projective limit, a family of marginals can always be obtained by defining $\Theta_{\mathrm{I}} := g_{\mathrm{I}}\Theta_{\mathrm{D}}$ and $P_{\mathrm{I}}[X_{\mathrm{I}}|\Theta_{\mathrm{I}}] := \mathbb{E}[(f_{\mathrm{I}}P_{\mathrm{D}})[X_{\mathrm{D}}|\Theta_{\mathrm{D}}]|\sigma(\Theta_{\mathrm{I}})]$, etc. The components so obtained form projective families with the initial model as their limit, and the theorem is applicable. Similarly, if the canonical mappings are not assumed surjective, we simply obtain a more technical statement of the theorem which requires closure under sampling on the images of the canonical mappings. This generalization is trivial, since all measures used in the construction have to concentrate on these images. We also note, in the context of part (2), that the projectors pr_{I} in a countable product of Polish spaces are always open mappings.

Similar to projective limits, pullbacks preserve conjugacy:

PROPOSITION 4 (Pullbacks of conjugate models). *Let the Bayesian model specified by* $P[X|\Theta]$ *and* $P^{\theta}[\Theta|Y]$ *be conjugate, with posterior index* $(T^{(n)})_n$*. Let* $\tilde{P}[\tilde{X}|\tilde{\Theta}]$ *and* $\tilde{P}^{\theta}[\tilde{\Theta}|\tilde{Y}]$ *be the respective pullbacks under* $\mathcal{J}_{\mathcal{X}}$ *and* $\mathcal{J}_{\mathcal{T}}$*. Then the Bayesian model specified by* $\tilde{P}[\tilde{X}|\tilde{\Theta}]$ *and* $\tilde{P}^{\theta}[\tilde{\Theta}|\tilde{Y}]$ *is conjugate, with posterior index given by the pullbacks of* $(T^{(n)})_n$ *as*

$$\tilde{T}^{(n)} := \mathcal{J}_{\mathcal{Y}}^{-1} \circ T^{(n)} \circ (\mathcal{J}_{\mathcal{X}}^{n} \otimes \mathcal{J}_{\mathcal{Y}})\,. \tag{6.5}$$

PROOF. As an arbitrary subset of a Polish space, $\tilde{\mathcal{T}}$ is separable, but not in general Polish, and conditional probabilities are not guaranteed to be regular. Since the pullback is defined by restriction, which preserves measurability, $\tilde{P}^{\theta}(\tilde{\Theta}|\tilde{Y})$ nonetheless constitutes a well-defined regular conditional probability. Lemma 4 ensures that the spaces $\tilde{\mathcal{X}}$, $\tilde{\mathcal{T}}$ and $\tilde{\mathcal{Y}}$ all correspond to the same subset $\tilde{\Omega}$ of the abstract probability space. Equation (6.5) is an immediate consequence of the definitions of posterior indices and pullbacks of parametric models. □

As an example of the previous results, we consider one of the most widely used Bayesian nonparametric models, a Gaussian process model for regression under uniform measurement noise – see e.g. [51] for a discussion of the model's statistical properties, or [46] for a machine learning perspective. We construct the nonparametric model from parametric Gaussian-Gaussian conjugate models.

Example 3 (Gaussian process regression under white noise). Consider a regression problem on $[0, 1]$, in which measurements $x_s \in \mathbb{R}$ are recorded at covariate locations $s \in [0, 1]$. Assume each measurement to be a value θ_s corrupted by standard white noise $\varepsilon_s \sim \mathcal{N}(0, 1)$, that is, $x_s = \theta_s + \varepsilon_s$. Since the noise is discontinuous, the function space $\tilde{\mathcal{X}} := \mathcal{L}_2[0, 1]$ is a more adequate setting than the set of continuous functions considered in Example 2. Let $(e_i)_{i\in\mathbb{N}}$ be an orthonormal basis of $\mathcal{L}_2[0, 1]$. Any $\tilde{x} \in \mathcal{L}_2[0, 1]$ is uniquely representable as $\tilde{x} = \sum_i x_{\{i\}} e_i$, where $x_{\{i\}} = \langle \tilde{x}, e_i \rangle$. The mapping $\phi : \tilde{x} \mapsto (x_{\{i\}})_{i\in\mathbb{N}}$ is an isomorphism of the separable Hilbert spaces $\mathcal{L}_2[0, 1]$ and ℓ_2. Since $\ell_2 \subset \mathbb{R}^{\mathbb{N}}$, we choose the product projective limit $\mathcal{X}_{\mathrm{D}} = \mathbb{R}^{\mathbb{N}}$ with canonical mappings $f_{\mathrm{I}} := \mathrm{pr}_{\mathrm{I}}$. In the terminology of Definition 1, the set $\Gamma = \ell_2$ is the image of $\mathcal{L}_2[0, 1]$ under the Borel embedding ϕ.

To obtain a valid model on $\mathcal{L}_2[0, 1]$ as a pullback of the Gaussian projective limits on $\mathcal{X}_{\mathrm{D}}$ and $\mathcal{T}_{\mathrm{D}}$, we need to know that the models assign outer measure 1 to the subset $\Gamma = \ell_2$ (cf. Sec. 4.3). Gaussian processes with realizations in ℓ_2 – or, in our terminology, Gaussian projective limits which satisfy $P^*_{\mathrm{D}}(\ell_2) = 1$ – are characterized by a well-known result [32, Theorem 3.2]: Denote by $\mathcal{S}(\ell_2)$ the set of all positive definite Hermitian operators on ℓ_2 of "trace class", i.e. with finite trace $\mathrm{tr}(\Sigma) < \infty$. The Gaussian projective limit P_{D} satisfies $P^*_{\mathrm{D}}(\ell_2) = 1$ if and only if there are $m \in \ell_2$ and $\Sigma \in \mathcal{S}(\ell_2)$ such that

$$\mathbb{E}[X_{\mathrm{I}}] = f_{\mathrm{I}}(m) \qquad \text{and} \qquad \mathrm{Cov}[X_{\mathrm{I}}] = (f_{\mathrm{I}} \otimes f_{\mathrm{I}})(\Sigma) \; . \tag{6.6}$$

To define a Bayesian model, we choose $\tilde{\mathcal{T}} = \tilde{\mathcal{X}}$, $\mathcal{T}_{\mathrm{D}} = \mathcal{X}_{\mathrm{D}}$ and $g_{\mathrm{I}} = f_{\mathrm{I}}$. To define a projective family of priors, let $m \in \ell_2$, $\Sigma \in \mathcal{S}_{\ell_2}$, and define the measures $P^{\theta}_{\mathrm{I}}(\Theta_{\mathrm{I}}|Y_{\mathrm{I}})$ as the Gaussian measures on $\mathcal{T}_{\mathrm{I}} = \mathbb{R}^{\mathrm{I}}$ with means $g_{\mathrm{I}}(m)$ and covariance matrices $(g_{\mathrm{I}} \otimes g_{\mathrm{I}})(\Sigma)$. The hyperparameter spaces are therefore $\mathcal{Y}_{\mathrm{I}} := \mathbb{R}^{\mathrm{I}} \times \mathrm{Sym}(I, \mathbb{R})$, where $\mathrm{Sym}(I, \mathbb{R})$ is the symmetric cone of real-valued $|I| \times |I|$ s.p.d. matrices. The projector $J \to I$ on the latter deletes all rows and columns indexed by elements of $J \setminus I$. For the white-noise observation model, let $\mathbb{I} \in \mathcal{S}(\ell_2)$ be the identity operator. Each marginal is chosen as the Gaussian conditional $P_{\mathrm{I}}[X_{\mathrm{I}}|\Theta_{\mathrm{I}};\mathbb{I}]$, i.e. conditional on the random mean Θ_{I} for fixed unit covariance. The priors form a projective family of measures and the observation models, by Lemma 3, a projective family of conditional distributions. A conjugate posterior index of the model is given by

$$(6.7) \qquad T^{(1)}_{\mathrm{I}}: \ (x_{\mathrm{I}}, m_{\mathrm{I}}, \Sigma_{\mathrm{I}}) \ \mapsto \ \big((\Sigma_{\mathrm{I}} + \mathbb{I}_{\mathrm{I}})^{-1}(\Sigma_{\mathrm{I}} x_{\mathrm{I}} + m_{\mathrm{I}}), (\Sigma_{\mathrm{I}} + \mathbb{I}_{\mathrm{I}})^{-1}\Sigma_{\mathrm{I}}\big) \ .$$

Since the covariance of the observation model is the fixed identity matrix $\mathbb{I}_{\mathrm{I}}$, it is not a hyperparameter, and hence formally part of the definition of the mapping rather than an argument.

The posterior index $\tilde{T}^{(n)}$ of the nonparametric model can be constructed as follows. We define a candidate function which mimics the functional form of the finite-dimensional posterior indices $T^{(n)}_{\mathrm{I}}$: For $\tilde{x}, \tilde{m} \in \ell_2$ and $\tilde{\Sigma} \in \mathcal{S}(\ell_2)$, define a mapping $\ell_2 \times \ell_2 \times \mathcal{S}(\ell_2) \mapsto \ell_2 \times \mathcal{S}(\ell_2)$ as

$$(6.8) \qquad \tilde{T}^{(1)}: \ (\tilde{x}, \tilde{m}, \tilde{\Sigma}) \ \mapsto \ \big((\tilde{\Sigma} + \tilde{\mathbb{I}})^{-1}(\tilde{\Sigma}\tilde{x} + \tilde{m}), (\tilde{\Sigma} + \tilde{\mathbb{I}})^{-1}\tilde{\Sigma}\big) \ .$$

It is straightforward to verify that (1) the maps $T^{(1)}_{\mathrm{I}}$ are the finite-dimensional projections of $\tilde{T}^{(1)}$ under the projectors $f_{\mathrm{I}} \circ \phi$ and $h_{\mathrm{I}} \circ \phi$, and that (2) the family of maps $\langle T^{(1)}_{\mathrm{I}} \rangle_{\mathrm{D}}$ is projective (i.e. satisfies (A.3)). Therefore, $\tilde{T}^{(1)}$ indeed coincides with the unique projective limit map on the subspace $\ell_2 \times \ell_2 \times \mathcal{S}_{\ell_2}$ of the projective limit space. Since

$$(6.9) \qquad \tilde{T}^{(1)}(\ell_2 \times \ell_2 \times \mathcal{S}_{\ell_2}) \subset \ell_2 \times \mathcal{S}_{\ell_2} \ ,$$

we need not be concerned with the behavior of the projective limit map outside this subspace. By Theorem 3, the projective limit model is conjugate. As (6.9) shows, the posterior again assigns full outer measure to $\ell_2 \cong \mathcal{L}_2[0,1]$. By Theorem 4, the pullback of the model under ϕ is a conjugate Gaussian process model on ℓ_2. The sufficient statistics $S_{\mathrm{I}} = \mathrm{Id}_{\mathbb{R}^{\mathrm{I}}}$ of the marginals are trivially projective, and by Theorem 2 and Proposition 2, the pullback $\tilde{S} = \mathrm{Id}_{\ell_2}$ of their projective limit is a sufficient statistic of Gaussian process model.

Since conjugacy in parametric models is, with few exceptions, a property of exponential family models, we can interpret most conjugate nonparametric Bayesian models as infinite-dimensional analogues of the exponential family. Conjugate priors of exponential family models are characterized by a linear arithmetic in parameter space, as shown by Diaconis and Ylvisaker [14]. In particular, suppose that the marginals $P_{\mathrm{I}}[X_{\mathrm{I}}|\Theta_{\mathrm{I}}]$ and $P^{\theta}_{\mathrm{I}}[\Theta_{\mathrm{I}}|Y_{\mathrm{I}}]$ are exponential family marginals with canonical conjugate priors. With respect to suitable carrier measures on the spaces $\mathcal{X}_{\mathrm{I}}$ and $\mathcal{T}_{\mathrm{I}}$, the marginals are then defined by conditional densities

$$p^x_{\mathrm{I}}(x_{\mathrm{I}}|\theta_{\mathrm{I}}) = \frac{H_{\mathrm{I}}(x_{\mathrm{I}})}{Z_{\mathrm{I}}(\theta_{\mathrm{I}})} e^{\langle S_{\mathrm{I}}(x_{\mathrm{I}}),\theta_{\mathrm{I}}\rangle} \qquad p^\theta_{\mathrm{I}}(\theta_{\mathrm{I}}|\lambda, y_{\mathrm{I}}) = \frac{e^{\langle\theta_{\mathrm{I}},\gamma_{\mathrm{I}}\rangle - \lambda \log Z_{\mathrm{I}}(\theta_{\mathrm{I}})}}{K_{\mathrm{I}}(\lambda,\gamma_{\mathrm{I}})} \,. \tag{6.10}$$

The function $S_{\mathrm{I}} : \mathcal{X}_{\mathrm{I}} \to \mathcal{U}_{\mathrm{I}}$ is a sufficient statistic. Its range is a Polish topological vector space $\mathcal{U}_{\mathrm{I}}$, which contains the parameter space $\mathcal{T}_{\mathrm{I}}$ as a subspace, and is equipped with an inner product $\langle\,.\,,\,.\,\rangle$. H_{I} denotes a non-negative function, and Z_{I}, K_{I} are normalization functions. The prior is parametrized by $\lambda \in \mathbb{R}_+$, which determines concentration, and $\gamma_{\mathrm{I}} \in \mathrm{conv}(S_{\mathrm{I}}(\mathcal{X}_{\mathrm{I}}))$, the convex hull of the image of S_{I}.

In our previous notation, the Bayesian model defined by (6.10) has hyperparameter space $\mathcal{Y}_{\mathrm{I}} := \mathbb{R}_+ \times \mathcal{U}_{\mathrm{I}}$, with $y_{\mathrm{I}} = (\lambda, \gamma_{\mathrm{I}})$. The posterior under observations $x^n_{\mathrm{I}} = (x^{(1)}_{\mathrm{I}}, \ldots, x^{(n)}_{\mathrm{I}})$ is given by the density $p_{\mathrm{I}}(\theta_{\mathrm{I}}|T^{(n)}_{\mathrm{I}}(x^n_{\mathrm{I}}, y_{\mathrm{I}}))$, where the posterior index updates hyperparameters as $\lambda \mapsto \lambda + n$ and $\gamma_{\mathrm{I}} \mapsto \gamma_{\mathrm{I}} + \sum_k S_{\mathrm{I}}(x^{(k)}_{\mathrm{I}})$. Application of Theorems 3 and 4 results in an analogous representation for nonparametric models:

COROLLARY 2 (Exponential family marginals). *Let $\langle P_I[X_I|\Theta_I]\rangle_D$ be a projective family of exponential family models with sufficient statistics S_I, and let $\langle P^\theta_I[\Theta_I|Y_I]\rangle_D$ be the family of corresponding canonical conjugate priors. If the priors and the sufficient statistics both form projective families, the projective limit Bayesian model is conjugate with posterior index*

$$T^{(n)}_D(x^n_D, y_D) := (\lambda + n, \gamma_D + \sum_k S_D(x^{(k)}_D)) \,, \tag{6.11}$$

where $y_D = (\lambda, \gamma_D)$ and $S_D := \varprojlim \langle S_I\rangle_D$ is the sufficient statistic of $P_D[X_D|\Theta_D]$. Analogously, if the model is pulled back under $\phi : \tilde{\mathcal{X}} \to \mathcal{X}_D$ as in (4.4),

$$\tilde{T}^{(n)}(\tilde{x}^n, \tilde{y}) := (\lambda + n, \tilde{\gamma} + \sum_k \tilde{S}(\tilde{x}^{(k)})) \,, \tag{6.12}$$

is a conjugate posterior index of the pullback model.

An example of such a posterior is the Dirichlet process on the line with concentration α and base measure G_0, for which posterior parameters are updated under observations $v^{(1)}, \ldots, v^{(n)}$ as

$$(v^{(1)}, \ldots, v^{(n)}, \alpha \cdot G_0) \;\mapsto\; \frac{n}{n+\alpha} \sum_{k=1}^{n} \delta_{v^{(k)}} + \frac{\alpha}{n+\alpha} G_0 \,. \tag{6.13}$$

The next section covers this example in detail. The Gaussian process regression above is an instance of Corollary 2 as well, although our formulation in Example 3 uses the standard parametrization of the Gaussian, rather than an exponential family parametrization adapted to (6.12).

7. Examples. Two detailed construction examples are given to illustrate our results: The well-known Dirichlet process [16, 30], and a new nonparametric Bayesian model on the infinite symmetric group. The steps of both constructions are (i) the definition of projective systems to obtain $\mathcal{X}_{\mathrm{D}}$ and $\mathcal{T}_{\mathrm{D}}$, (ii) the definition of finite-dimensional priors and likelihoods for each $I \in D$ to define a projective limit Bayesian model, and (iii) a pullback step to ensure that the models concentrate on the desired subspace of interest – the set of probability measures and the infinite symmetric group, respectively. By means of the results in Secs. 5 and 6, sufficiency and conjugacy properties of the models can then be read off from the properties of the marginals.

7.1. *Dirichlet Process Priors.* In this example, $\tilde{P}^{\theta}$ is a Dirichlet process and $\tilde{P}$ its conjugate observation model. The domain of the Dirichlet process is assumed to be a Polish measurable space $(V, \mathcal{B}_V)$, i.e. random measures drawn from the process are convex combinations of the form $\theta_{\mathrm{D}} = \sum_{k \in \mathbb{N}} c_k \delta_{v_k}$ with $v_k \in V$.

7.1.1. *Projective System.* The finite-dimensional marginals will be Dirichlet and multinomial distributions. Ferguson [16] noted that a particularly intuitive way to index such distributions is to choose each $I \in D$ as a finite, measurable partition $I = (A_1, \dots, A_{|I|})$ of V. The $|I|$-dimensional Dirichlet distribution P_{I}^{θ} can then be interpreted as a random measure on the finite σ-algebra $\sigma(I)$ generated by the sets in I. Let $\mathcal{H}(\mathcal{B}_V)$ be the set of all finite partitions $I = (A_1, \dots, A_{|I|})$ with $A_i \in \mathcal{B}_V$. This set is itself not an adequate choice for D, since it is uncountable unless V is finite. However, since V is Polish, there exists a countable algebra $\mathcal{Q} \subset \mathcal{B}_V$ which generates $\mathcal{B}_V$. Any probability measure on $\mathcal{B}_V$ can, by Carathéodory's extension theorem, be unambiguously represented by its restriction to $\mathcal{Q}$. Bearing this in mind, we define $D := \mathcal{H}(\mathcal{Q})$ as the set of finite partitions with $A_i \in \mathcal{Q}$. A partial order on D is defined by $I \preceq J$ if and only if $I \cap J = J$, that is, if J is a refinement of the partition I.

For each index $I = (A_1, \dots, A_{m_{\mathrm{I}}})$ in D, the marginal spaces are chosen as the spaces corresponding to a Dirichlet-multinomial Bayesian model over m_{I} categories: Let the parameter space $\mathcal{T}_{\mathrm{I}}$ be the set of probability distributions on the σ-algebra generated by I, i.e. the unit simplex $\triangle_{\mathrm{I}} \subset \mathbb{R}^{\mathrm{I}}$. The hyperparameter space of a Dirichlet model on $\triangle_{\mathrm{I}}$ is $\mathcal{Y}_{\mathrm{I}} := \mathbb{R}_{>0} \times \triangle_{\mathrm{I}}$.

To define the observation spaces $\mathcal{X}_{\mathrm{I}}$, we interpret the sets $A_i \in I$ as categories or "bins" of a multinomial distribution. A sample in category A_i can be encoded as $\{X_{\mathrm{I}} = A_i\}$. Hence, X_{I} takes values in $\mathcal{X}_{\mathrm{I}} := I$. Both the topology $\mathrm{Top}_{\mathrm{I}}$ and Borel sets $\mathcal{B}_{\mathrm{I}}$ on $\mathcal{X}_{\mathrm{I}}$ are generated by the singleton events $\{A_1\}, \dots, \{A_{m_{\mathrm{I}}}\}$.

To define suitable projectors, consider a pair $I \preceq J$ of indices, where $I = (A_1, \dots, A_{m_{\mathrm{I}}})$ and $J = (A'_1, \dots, A'_{m_{\mathrm{J}}})$. Any set $A_i \in I$ is the union of some sets in J, hence $A_i = \cup_{j \in J_i} A'_j$ for some $J_i \subset [m_{\mathrm{J}}]$. Let $\theta_{\mathrm{J}} \in \triangle_{\mathrm{J}}$ be a finite probability distribution and $A'_j \in J$. We define

$$f_{\mathrm{JI}}(A'_j) = A_i \quad \text{for } j \in J_i, \quad \text{and} \quad (g_{\mathrm{JI}}\theta_{\mathrm{J}})_i := \sum_{j \in J_i} (\theta_{\mathrm{J}})_j \; . \tag{7.1}$$

In words, for any coarsening of a finite set of events J to I, f_{JI} maps A'_j to the coarser set containing it, and g_{JI} sums the corresponding probabilities. Since the model is of conjugate exponential family type, the projections $h_{\mathrm{JI}} : \mathcal{Y}_{\mathrm{J}} \to \mathcal{Y}_{\mathrm{I}}$ of hyperparameters are given by $h_{\mathrm{JI}} := \mathrm{Id}_{\mathbb{R}_+} \otimes g_{\mathrm{JI}}$. The families of mappings $f_{\mathrm{JI}} : J \to I$ and $g_{\mathrm{JI}} : \triangle_{\mathrm{J}} \to \triangle_{\mathrm{I}}$ are continuous, surjective and satisfy (A.1). It is straightforward to verify that $\langle \mathcal{X}_{\mathrm{I}}, \mathcal{B}(\mathcal{X}_{\mathrm{I}}), f_{\mathrm{JI}} \rangle_{\mathrm{D}}$ and $\langle \mathcal{T}_{\mathrm{I}}, \mathcal{B}(\mathcal{T}_{\mathrm{I}}), g_{\mathrm{JI}} \rangle_{\mathrm{D}}$ are projective systems of measurable Polish spaces. Properties of the hyperparameter spaces $\mathcal{Y}_{\mathrm{I}}$ carry over immediately from $\mathcal{T}_{\mathrm{I}}$.

What are the projective limit spaces $\mathcal{X}_{\mathrm{D}}$ and $\mathcal{T}_{\mathrm{D}}$ defined by the two systems? The set $\mathcal{X}_{\mathrm{D}}$ consists of all collections of the form

$$x_{\mathrm{D}} = \{C_{\mathrm{I}} \in I | I \in D, C_{\mathrm{I}} \supset C_{\mathrm{J}} \text{ whenever } I \preceq J\} \; . \tag{7.2}$$

Whereas a draw from X_{I} selects a single random set $C_{\mathrm{I}} \in I$, a draw from X_{D} selects one random set C_{I} for each I. A single, "smallest" set can be associated with each $x_{\mathrm{D}} \in \mathcal{X}_{\mathrm{D}}$ by defining $\lim x_{\mathrm{D}} := \cap_I C_{\mathrm{I}}$. Unlike the constituent sets C_{I}, the set $\lim x_{\mathrm{D}}$ is not in general an element of $\mathcal{Q}$, and we have $\mathcal{Q} \subset \{\lim x_{\mathrm{D}} | x_{\mathrm{D}} \in \mathcal{X}_{\mathrm{D}}\} \subset \mathcal{B}_V$. In particular, the proof of Lemma 6 below shows that the set $\{\lim x_{\mathrm{D}} | x_{\mathrm{D}} \in \mathcal{X}_{\mathrm{D}}\}$ contains all singleton $\{v\}$ for points $v \in V$, which are not contained in the countable set $\mathcal{Q}$. In analogy to the interpretation of X_{I} as an event $A_i \in I$, we can interpret $X_{\mathrm{D}} = x_{\mathrm{D}}$ as the event $\lim x_{\mathrm{D}} \in \mathcal{B}_V$.

The projective limit $\mathcal{T}_{\mathrm{D}}$ of parameter spaces is the set of all charges, i.e. of *finitely* additive probabilities on the algebra $\mathcal{Q}$. The space $(\mathcal{T}_{\mathrm{D}}, \mathcal{B}(\mathcal{T}_{\mathrm{D}}))$ contains the set $M(\mathcal{Q})$ of countably additive probability measures as a measurable subset [41, Proposition 9]. For any set function $G \in \mathcal{T}_{\mathrm{D}}$, the canonical maps $g_{\mathrm{I}} : \mathcal{T}_{\mathrm{D}} \to \triangle_{\mathrm{I}}$ are the evaluations $G \mapsto (G(A_1), \dots, G(A_{m_{\mathrm{I}}}))$. The fact that the space $M(\mathcal{Q})$ cannot directly be defined as a projective limit of finite-dimensional simplices is an example for the projective limit's ability to encode a finitary property (finite additivity), but not an infinitary one (countable additivity).

7.1.2. *Bayesian Model.* Each marginal Bayesian model is defined by a multinomial distribution $P_{\mathrm{I}}[X_{\mathrm{I}}|\Theta_{\mathrm{I}}]$ on m_{I} categories and by its conjugate prior $P^{\theta}_{\mathrm{I}}[\Theta_{\mathrm{I}}|Y_{\mathrm{I}}]$ on $\triangle_{\mathrm{I}}$. The Dirichlet distribution is a natural conjugate prior as in (6.10), with parameters $(\lambda, \gamma_{\mathrm{I}}) := (\lambda, \alpha G_{\mathrm{I}})$, where $\alpha \in \mathbb{R}_+$ controls concentration and an $G_{\mathrm{I}} \in \triangle_{\mathrm{I}}$ is the expected value. Since $\log Z_{\mathrm{I}}(\theta_{\mathrm{I}}) = 0$ for the Dirichlet distribution, the value of λ does not affect the model and is henceforth omitted. Though α controls the concentration of the model, it acts linearly on θ_{I}, in contrast to the nonlinear influence of λ on other conjugate priors. It is easy to show that the multinomial and Dirichlet families so defined form a projective family of Bayesian models if and only if the hyperparameters are chosen consistently as $\gamma_{\mathrm{I}} := \alpha \cdot g_{\mathrm{I}} G_0$ for a fixed $\alpha \in \mathbb{R}_+$ and some $G_0 \in \mathcal{T}_{\mathrm{D}}$.

7.1.3. *Pullback to $M(V)$.* What remains to be done is to ensure that the Dirichlet process prior $P^{\theta}_{\mathrm{D}}[\Theta_{\mathrm{D}}|Y_{\mathrm{D}} = y_{\mathrm{D}}]$ defines a measure on the set of probability measures.

LEMMA 5. *If V is Polish, the countable generating algebra $\mathcal{Q} \subset \mathcal{B}_V$ can be chosen such that, for any charge G_0 on $\mathcal{Q}$,*

$$P^{\theta,*}_{D}[M(\mathcal{Q})|Y_D = (\alpha, G_0)] = 1 \quad \Leftrightarrow \quad G_0 \in M(\mathcal{Q})\ . \tag{7.3}$$

(Proof: App. C) In other words, the prior concentrates on countably additive set functions if and only if its hyperparameter is countably additive. We obtain a corresponding concentration result for the sampling model:

LEMMA 6. *Define a relation $\phi \subset V \times \mathcal{X}_D$ by means of*

$$v \equiv_{\phi} x_D \qquad \Leftrightarrow \qquad \lim x_D = \{v\} \tag{7.4}$$

1. *ϕ is a mapping $V \to \mathcal{X}_D$, and a Borel embedding.*
2. *If $\theta_D \in M(V)$ is purely atomic, $P^*_D[\phi(V)|\Theta_D = \theta_D] = 1$.*

(Proof: App. C) Lemma 6 provides a suitable embedding ϕ for the pullback of $P_{\mathrm{D}}[\,.\,|\Theta_{\mathrm{D}}]$. For the prior, let $\mathcal{J}_{\mathcal{T}} : M(V) \to \mathcal{T}_{\mathrm{D}}$ be the mapping which takes a probability measure ν on $\mathcal{B}_V$ to its restriction on $\mathcal{Q}$. By the Carathéodory extension theorem, $\mathcal{J}_{\mathcal{T}}$ is injective. Since both $M(V)$ and the

Borel subset $M(\mathcal{Q}) \subset \mathcal{T}_{\text{D}}$ are standard Borel spaces, $\mathcal{J}_{\mathcal{T}}$ is a Borel embedding (cf Sec. 2.2). We set $\tilde{\mathcal{X}} := \{\{v\}|v \in V\}$, which we identify with V, and choose $\tilde{\mathcal{T}} = M(V)$ as parameter space and $\tilde{\mathcal{Y}} := \mathbb{R}_+ \times M(V)$ as hyperparameter space. We do not show here that draws from the Dirichlet process are almost surely discrete, and instead refer to [20].

Sufficient statistics of the marginal models can be defined as $S_{\text{I}} : I \to \triangle_{\text{I}}$ with $A_i \mapsto \delta_{\{A_i\}}$, i.e. A_i is mapped to a point mass at the singleton $\{A_i\} \in \mathcal{B}_{\text{I}}$. Define $\tilde{S} : V \to M(V)$ by $v \mapsto \delta_{\{v\}}$. For any $I \in D$, there is a unique $A_i \in I$ with $v \in A_i$, and hence $(g_{\text{I}} \circ \mathcal{J}_{\mathcal{T}}) \circ \tilde{S} = S_{\text{I}} \circ (f_{\text{I}} \circ \phi)$. Therefore, S_{I} is a pullback of the projective limit $S_{\text{D}} = \varprojlim \langle S_{\text{I}} \rangle_{\text{D}}$. By Theorem 2 and Proposition 2, $\tilde{S}$ is a sufficient statistic for the pullback model. By Theorem 3 and Proposition 4, the pullback model is conjugate. In summary:

COROLLARY 3. *The pullback Bayesian model defined by* $\tilde{P}[\tilde{X}|\tilde{\Theta}]$ *and* $\tilde{P}^{\theta}[\tilde{\Theta}|\tilde{Y}]$ *is a conjugate Bayesian model with hyperparameter space* $\mathbb{R}_{>0} \times M(V)$*, parameter space* $\{\nu \in M(V)|\nu \text{ discrete}\}$*, and sample space* V*. The posterior index under* n *observations* $\tilde{x}^{(1)}, \ldots, \tilde{x}^{(n)} \in V$ *is*

$$\alpha G_0 \quad \mapsto \quad \alpha G_0 + \sum_i \tilde{S}(\tilde{x}^{(i)}) \ , \tag{7.5}$$

and $\tilde{S}$ *is a sufficient statistics of the model.*

The measure $\tilde{P}^{\theta}[\tilde{\Theta}|\tilde{Y} = (\alpha, G_0)]$ is, in the terminology of nonparametric Bayesian statistics, a Dirichlet process with concentration α and base measure G_0.

7.2. *A Nonparametric Model on Permutations.* In this second example, the observations $\tilde{x}$ are elements of the infinite symmetric group $\mathbb{S}_\infty$, and the parameters are sequences $\tilde{\theta} \in \mathbb{R}^\infty$ satisfying a certain convergence condition. The infinite symmetric group is the set of all permutations of the set $\mathbb{N}$ which change an arbitrary but finite number of elements.

Models on such infinite permutations are of potential interest in two contexts: (1) Rank data is modeled by permutations, and a nonparametric approach to ranking problems motivates models on infinite permutations. In parametric rank data analysis, models for "partial" data, i.e. data in which part of each ranking is censored, are used to model "rank your favorite r items out of a total of n items" [18]. In this case, n is the order of the underlying symmetric group $\mathbb{S}_n$, and r the number of uncensored positions. Meilă and Bao [40] observe that positing a given set of n items to choose from makes most partial ranking tasks artificial. They suggest a nonparametric model on $\mathbb{S}_\infty$ to represent more realistic tasks ("rank your favorite r

movies", as opposed to "rank your favorite r out of these n movies").
(2) The cycles of an infinite permutation induce a partition of $\mathbb{N}$, and random permutations hence induce random partitions. The most prominent example of such a model is without doubt the Chinese Restaurant Process, proposed by Pitman and Dubins as a distribution on infinite random permutations with uniform marginals [42].

To construct a Bayesian model on $\mathbb{S}_\infty$ by means of a projective limit, we draw upon a beautiful construction recently proposed in representation theory by Kerov, Olshanski, and Vershik [28, 29]. This approach constructs an compactification $\mathfrak{S}$ of $\mathbb{S}_\infty$ as a projective limit of finite symmetric groups; Kerov *et al.* [29] refer to the elements of $\mathfrak{S}$ as "virtual permutations". We construct a Bayesian model by endowing each of the finite groups with a parametric model based on the Cayley distance, due to Fligner and Verducci [18]. We then give conditions under which the projective limit concentrates on the subset $\mathbb{S}_\infty$.

7.2.1. *Projective Limits of Symmetric Groups.* The projective limit of Kerov *et al.* [29] assembles the symmetric groups $\mathbb{S}_1, \mathbb{S}_2, \ldots$ sequentially, and we hence choose the index set $D = \{[n] | n \in \mathbb{N}\}$, ordered by inclusion. To define a projective system, we need a suitable notion of projection mappings. Given the choice of D, it is sufficient to consider mappings f_{JI} for $J = [n+1]$ and $I = [n]$, which we more conveniently denote $f_{n+1,n}$. Intuitively, the projection should remove the entry $n+1$ from permutations $\mathbb{S}_{n+1}$ – which raises the question of how to consistently delete $n+1$ from, say, $\binom{1\ 2\ 3\ 4}{3\ 1\ 4\ 2}$ to obtain a valid element of $\mathbb{S}_3$. An appropriate projector can be defined as follows. Any permutation $\pi \in \mathbb{S}_n$ admits a unique representation of the form

$$\pi = \sigma_{k_1}(1)\sigma_{k_2}(2)\cdots\sigma_{k_n}(n) \ , \tag{7.6}$$

where k_i are natural numbers with $k_i \leq i$, and $\sigma_i(j)$ denotes the transposition of i and j. Hence, the vector $(k_1, \ldots, k_n)$ is an encoding of π. Let ψ_n be the corresponding mapping defined by $\psi_n : \pi \mapsto (k_1, \ldots, k_n)$, which does not incur the same consistency problems as simple deletion of an entry of π. Due to the constraint $k_i \leq i$, the mapping is a bijection $\psi_n : \mathbb{S}_n \to \prod_{m \leq n}[m]$, and a homeomorphism of Polish spaces if both $\mathbb{S}_n$ and the image space are endowed with the discrete topology. On the encoding $\psi_n \pi$, we can easily define a natural projection by deleting the last element, i.e. as $f'_{n+1,n} : (k_1, \ldots, k_n, k_{n+1}) \mapsto (k_1, \ldots, k_n)$. The projectors $f_{n+1,n}$ are then chosen as the induced mappings on the groups, hence

$f_{n+1,n} := \psi_n^{-1} \circ f'_{n+1,n} \circ \psi_{n+1}$. The following diagram commutes:

$$\begin{array}{ccc} \mathbb{S}_{n+1} & \xrightarrow{\psi_{n+1}} & \prod_{m \leq n+1} [m] \\ \downarrow f_{n+1,n} & & \downarrow \mathrm{pr}_{[n]} \\ \mathbb{S}_n & \xrightarrow{\psi_n} & \prod_{m \leq n} [m] \end{array} \tag{7.7}$$

The projectors $f_{n+1,n}$ have a natural group-theoretic interpretation: They remove the element $n+1$ from the cycle containing it. Intuitively speaking, application of $\sigma_{k_1}(1), \ldots, \sigma_{k_n}(n)$ from the left consecutively constructs the cycles of $\pi \in \mathbb{S}_{n+1}$, pending insertion of the final element $n+1$ into its respective cycle. This last step is ommitted by deleting $\sigma_{k_{n+1}}(n+1)$. The definition of $f_{n+1,n}$ is consistent with the Chinese Restaurant Process [42]: The image measure of the CRP marginal distribution on $\mathbb{S}_{n+1}$ under $f_{n+1,n}$ is the CRP marginal on $\mathbb{S}_n$.

The projections $f_{n+1,n}$ determine $(X_{\mathrm{D}}, \mathrm{Top}_{\mathrm{D}}) := \varprojlim \langle \mathbb{S}_n, \mathrm{Top}_n, f_{n+1,n} \rangle_{\mathrm{D}}$. As a subset of $\prod_{n \in \mathbb{N}} \mathbb{S}_n$, the space X_{D} is countable, with a topology $\mathrm{Top}_{\mathrm{D}}$ inherited from the product topology. In this topology, X_{D} is totally disconnected and compact. In analogy to the finite groups, X_{D} is homeomorphic to the product space $\prod_{m \in \mathbb{N}} [m]$, under the encoding map $\psi_{\mathrm{D}} : (\sigma_{k_1}(1)\sigma_{k_2}(2)\cdots) \mapsto (k_1, k_2, \ldots)$. Unlike $\mathbb{S}_\infty$, the space X_{D} is not a group. Conversely, $\mathbb{S}_\infty$ is not compact, but rather a dense subset $\mathbb{S}_\infty \subset X_{\mathrm{D}}$. The projective limit X_{D} can thus be regarded as an abstract compactification of $\mathbb{S}_\infty$. Elements $\pi \in X_{\mathrm{D}}$ are representable in the form

$$\pi = \sigma_{k_1}(1)\sigma_{k_2}(2)\cdots \tag{7.8}$$

and can be interpreted as operations that iteratively permute pairs of elements ad infinitum. If and only if $\pi \in \mathbb{S}_\infty$, this process "breaks off" after a finite number n of steps, and the encoding of π is of the form $\psi(\pi) = (k_1, \ldots, k_n, n+1, n+2, \ldots)$.

7.2.2. *Distance-Based Models.* A widely used class of probability distributions on finite symmetric groups are location-scale models of the form

$$p(\pi|\theta, \pi_0) = \frac{1}{Z(\theta)} e^{-\theta d(\pi, \pi_0)} , \tag{7.9}$$

where d is a metric on $\mathbb{S}_n$. Such models are commonly referred to as *distance-based* models in the rank data literature. Fligner and Verducci [18] considered the intersection of this class with another type of model: Let $W^{(1)}, \ldots, W^{(k)}$

be a set of statistics on $\mathbb{S}_n$ such that the random variables $W^{(1)}(\pi), \dots, W^{(k)}(\pi)$ are independent if π is distributed uniformly. Define a parametric model on $\mathbb{S}_n$ as

$$p(\pi|\theta) := \frac{1}{Z(\theta)} \exp\Big(-\sum_{j=1}^{k} \theta^{(j)} W^{(j)}(\pi)\Big) \qquad \theta \in \mathbb{R}^k \; . \tag{7.10}$$

The moment-generating function $M(\theta)$ of this model is the product $M(\theta) = \prod_j M^{(j)}(\theta^{(j)})$ over the moment-generating functions $M^{(j)}$ of the variables $W^{(j)}(\pi)$. Hence, the partition function $Z(\theta)$ of the model factorizes as $Z(\theta) = \prod_j Z^{(j)}(\theta^{(j)}) = \prod_j M^{(j)}(-\theta^{(j)})$, and the statistics $W^{(j)}(\pi)$ are independent random variables if π is distributed either uniformly or according to $p(\pi|\theta)$. These models coincide with distance-based models of the form (7.9) whenever the metric $d(\pi, \pi_0)$, for π_0 the neutral permutation and π uniform on $\mathbb{S}_n$, is decomposable as a sum of independent random variables $d(\pi, \pi_0) = \sum_j W^{(j)}(\pi)$. The decomposable metrics considered in [18] are the Kendall metric and the Cayley metric. We will consider the Cayley metric d_{C} in the following, defined as the minimal number of (not necessarily adjacent) transpositions required to transform π into π'. For the neutral permutation π_0, this metric satisfies $d_{\mathrm{C}}(\pi, \pi_0) = (n - \#\mathrm{cycles}(\pi))$. Consequently, d_{C} can be decomposed into a sum of statistics which count the positions in π, but discount one element of each cycle. We hence choose

$$W^{(j)} := 1 - \mathbb{I}\{k_j = j\} = \begin{Bmatrix} 0 & j \text{ smallest element on its cycle} \\ 1 & \text{otherwise} \end{Bmatrix} \; . \tag{7.11}$$

The definition differs slightly[1] from the one given by Fligner and Verducci [18], who discount the largest element on each cycle instead. The smallest element is a more adequate choice in the context of nonparametric constructions, as it is well-defined for infinite cycles.

Independence of the variables $W^{(j)}$ is easily verified by constructing a uniform random permutation π by means of n iterations of the Chinese Restaurant Process: In step j, element j is inserted into the current permutation by uniformly sampling $U \in [j]$. If $U = j$, the element is placed on a new cycle. Otherwise, it is inserted to the immediate right of element U on the respective cycle. The variables $W^{(j)}$ are hence indeed independent and Bernoulli distributed. Since only $U = j$ creates a new cycle, the Bernoulli parameters are $\frac{j-1}{j}$.

The natural conjugate prior P_n^{θ} of the generalized Cayley model on $\mathbb{S}_n$ is given by the conditional measure $P_n^\theta[\Theta_n|Y_n]$ with density

$$p_n^\theta(\theta_n|\lambda,\gamma_n) := \frac{1}{K_n(\lambda,\gamma_n)}\exp\big(\sum_j \gamma^{(j)}\theta^{(j)} - \lambda\log Z_n(\theta_n)\big)\ . \tag{7.12}$$

By means of Lemma 3, we can show:

LEMMA 7. *$\langle P_n[\pi_n|\Theta_n]\rangle_D$ and $\langle P_n^\theta[\Theta_n|Y_n]\rangle_D$ are projective families of conditional distributions.*

(Proof: App. C) As limits of the two projective families, we obtain the regular conditional probabilities $P_{\mathrm{D}}[X_{\mathrm{D}}|\Theta_{\mathrm{D}}] = \varprojlim \langle P_n[X_n|\Theta_n]\rangle_{\mathrm{D}}$ on the projective limit space $\mathcal{X}_{\mathrm{D}} = \mathfrak{S}$ of virtual permutations, and $P_{\mathrm{D}}^\theta[\Theta_{\mathrm{D}}|Y_{\mathrm{D}}] = \varprojlim \langle P_n^\theta[\Theta_n|Y_n]\rangle_{\mathrm{D}}$ on the parameter space $\mathcal{T}_{\mathrm{D}} = \mathbb{R}^{\mathbb{N}}$.

7.2.3. *Pullback to $\mathbb{S}_\infty$.* For nonparametric applications, infinite random permutations are of particular interest, i.e. observations generated by the model should almost surely take values $X_{\mathrm{D}} \in \mathbb{S}_\infty$. The model constructed above can be guaranteed to concentrate on $\mathbb{S}_\infty$ by a suitable choice of pullbacks.

Because the pullback model is supposed to concentrate on $\mathbb{S}_\infty$, the Borel embedding ϕ should obviously be the canonical inclusion $\phi : \mathbb{S}_\infty \hookrightarrow \mathfrak{S}$. The form of the corresponding embedding $\mathcal{J}_\mathcal{T}$ on parameter space is less apparent: We observe that an infinite permutation π is in $\mathbb{S}_\infty$ if and only if the infinite sequence $(W^{(j)})_j$ contains only a finite number of ones. For a given parameter $\theta^{(j)}$, the probability that $W^{(j)}(\pi) = 1$ is

$$\Pr\{W^{(j)}(\pi) = 1\} = \frac{(j-1)e^{-\theta^{(j)}}}{1+(j-1)e^{-\theta^{(j)}}} =: q_j(\theta^{(j)})\ , \tag{7.13}$$

and we define a mapping $q : \mathbb{R}^{\mathbb{N}} \to (0,1)^{\mathbb{N}}$ by $q(\theta) := (q_j(\theta_j))_{j\in\mathbb{N}}$. By the Borel-Cantelli lemma, only a finite number of $W^{(j)}$ take value one if and only if $q(\theta) \in \ell_1$ (cf. proof of Lemma 8). Let $\mathcal{J}_{\ell_1}$ be the canonical inclusion $\ell_1(0,1) \hookrightarrow \mathbb{R}^{\mathbb{N}}$, and define $\mathcal{J}_\mathcal{T} := \mathcal{J}_{\ell_1} \circ q$. The pullback parameter space is therefore

$$\tilde{\mathcal{T}} := \mathcal{J}_\mathcal{T}^{-1}(0,1)^{\mathbb{N}} = \{\theta \in \mathbb{R}^{\mathbb{N}} | \|q(\theta)\|_{\ell_1} < \infty\}\ . \tag{7.14}$$

A pullback of the model under ϕ and $\mathcal{J}_\mathcal{T}$ is justified by the following concentration result:

LEMMA 8. *With ϕ and $\mathcal{J}_\mathcal{T}$ defined as above:*

1. *The mappings* ϕ *and* $\mathcal{J}_{\mathcal{T}}$ *are Borel embeddings.*
2. $P^*_D[\mathbb{S}_\infty|\Theta = \theta] = 1$ *if and only if* $q(\theta) \in \ell_1(0,1)$.
3. $P^{\theta,*}_D[\tilde{\mathcal{T}}|Y_D = (\lambda, \gamma_D)] = 1$ *if* $\gamma_D \in \tilde{\mathcal{T}}$.

The entire projective limit Bayesian model can therefore be pulled back under ϕ and $\mathcal{J}_{\mathcal{T}}$ in the sense of (4.4). The pullback model, given by conditionals $\tilde{P}[\tilde{X}|\tilde{\Theta}]$ and $\tilde{P}^\theta[\tilde{\Theta}|\tilde{Y}]$, is parametrized by hyperparameter sequences γ_{D} satisfying $\|q(\gamma_{\mathrm{D}})\|_{\ell_1} < \infty$. The parameter variable $\tilde{\Theta}$ almost surely takes values θ satisfying $\|q(\theta)\|_{\ell_1} < \infty$, and the observation variable satisfies $\tilde{X} \in \mathbb{S}_\infty$ almost surely. The sufficient statistics $S_{\mathrm{I}} = S_n$ of the finite-dimensional models, matching the exponential family representation (6.10), are simply given by $S^{(j)}(\pi) = -W^{(j)}(\pi)$. By Theorem 2 and Proposition 2, a sufficient statistic $\tilde{S} : \mathbb{S}_\infty \to \{0,1\}^{\mathbb{N}}$ of the pullback model is therefore given by the countable vector $\tilde{S}(\pi)$ with components $\tilde{S}^{(j)}(\pi) = \mathbb{I}\{k_j = j\} - 1$. By Theorems 3 and 4, the model is conjugate. In summary:

COROLLARY 4. *The pullbacks* $\tilde{P}[\tilde{X}|\tilde{\Theta}]$ *and* $\tilde{P}^\theta[\tilde{\Theta}|\tilde{Y}]$ *define a projective limit Bayesian model with hyperparameter space* $\mathbb{R}_{>0} \times \tilde{\mathcal{T}}$*, parameter space* $\tilde{\mathcal{T}}$*, and observation space* $\mathbb{S}_\infty$*. The sequence* $\tilde{S}$ *with components* $\tilde{S}^{(j)}(\pi) = \mathbb{I}\{k_j = j\} - 1$ *is a sufficient statistic of the model, and posterior updates under observations* $\pi^n = (\pi^{(1)}, \ldots, \pi^{(n)})$*, with* $\pi^{(j)} \in \mathbb{S}_\infty$*, are given by*

$$\tilde{T}^{(n)}(\pi^n, (\lambda, \tilde{\gamma})) = \Big(n + \lambda, \tilde{\gamma} + \sum_{j \in \mathbb{N}} \tilde{S}^{(j)}(\pi^{(j)})\Big) . \tag{7.15}$$

Intuitively, the parameter $\theta^{(j)}$ describes an element-wise concentration. If all elements of θ_{I} are negative in the finite-dimensional model, the expected value of $P_{\mathrm{I}}[X_{\mathrm{I}}|\Theta_{\mathrm{I}} = \theta_{\mathrm{I}}]$ is an anti-mode [18]. The larger the value of $\theta^{(j)}$, the higher the cost of deviation from the neutral permutation at position j. If such a deviation is observed in π, $W^{(j)}(\pi) = 1$, and (7.15) describes a decrease of the expected concentration at j in the posterior.

The definition of the sufficient statistics used here closely follows the customary presentation in the rank data literature. Alternatively, the model could be expressed (with a different partition function) in terms of sufficient statistics $S^{(j)}(\pi) = \mathbb{I}\{k_j = j\}$, which emphasizes the close relation of the model to the representation (7.6). Similarly, it may be useful to reparametrize the model by $\vartheta := q(\theta)$, such that concentration on $\mathbb{S}_\infty$ occurs for convergent parameter sequences. In its present form, the parameter sequence has to diverge instead – concentration on $\mathbb{S}_\infty$ requires $W^{(j)} = 0$ eventually, and since the variables $W^{(j)}$ are independent, this occurs almost surely only for diverging concentration parameters.

APPENDIX A: PROJECTIVE LIMITS AND PULLBACKS

Both projective limits (also called *inverse limits*) and pullbacks are standard techniques in pure mathematics, and projective limits of probability measures are widely used in probability theory. Since neither is a standard topic in statistics, though, this appendix provides a brief survey of some relevant definitions and results.

A comprehensive reference on general projective limits is Bourbaki's *Elements of mathematics*; see Bourbaki [7, 8] for projective limits of spaces and functions. Key references on projective limits of measures are Bourbaki [9], Rao [44, 45], Mallory and Sion [39], Choksi [11] and Schwartz [48]. On pullbacks of measures, cf. Fremlin [19, Vol. I]. Both projective limits and pullbacks are common topics in category theory [e.g. 34].

A.1. Projective Systems and their Limits. A projective limit assembles a mathematical object from a system of simpler objects. The assembled object may be an infinite-dimensional space constructed from finite-dimensional subspaces, a group constructed from subgroups, a measure assembled from its marginals, or a function defined by combining functions on subspaces. How the objects are "glued together" is defined by specifying a system of mappings, denoted f_{JI} in the following, which connect "larger" objects to "smaller" ones. These mappings generalize the notion of a projection in a product space. The notion of "larger" and "smaller" is defined in terms of a partial order on the set D of object indices. To admit a proper definition of a limit, and hence of an extension to infinity, the index set needs to be directed.

Let D be a set and $\preceq$ a partial order relation on D. The set is called *directed* if for any two elements $I, J \in D$, there is a $K \in D$ such that $I \preceq K$ and $J \preceq K$. Let $\{\mathcal{X}_{\text{I}}\}_{I \in D}$ be a family of sets indexed by a directed set D. Require that for any pair $I \preceq J$ in D, there is a mapping $f_{\text{JI}} : \mathcal{X}_{\text{J}} \to \mathcal{X}_{\text{I}}$ satisfying

$$f_{\text{II}} = \text{Id}_{\mathcal{X}_{\text{I}}} \quad \text{and} \quad f_{\text{KI}} = f_{\text{KJ}} \circ f_{\text{KI}} \qquad \text{whenever } I \preceq J \preceq K \;. \tag{A.1}$$

Then $\{\mathcal{X}_{\text{I}}, f_{\text{JI}} | I \preceq J \in D\}$, in short $\langle \mathcal{X}_{\text{I}}, f_{\text{JI}} \rangle_{\text{D}}$, is called a *projective system*. Define a space $\mathcal{X}_{\text{D}}$ as follows: Let $\{x_{\text{I}} | I \in D\}$ be a collection consisting of a single point each from the spaces $\mathcal{X}_{\text{I}}$, for which

$$x_{\text{I}} = f_{\text{JI}} x_{\text{J}} \qquad \text{whenever } I \preceq J \;. \tag{A.2}$$

Identify any such collection with a point x_{D}, and let $\mathcal{X}_{\text{D}}$ be the set of all such points. Then $\mathcal{X}_{\text{D}}$ is called the *projective limit* of the system. The functions $f_{\text{I}} : \mathcal{X}_{\text{D}} \to \mathcal{X}_{\text{I}}$ defined by $x_{\text{D}} \mapsto x_{\text{I}}$ are called *canonical mappings*.

The projective limit $\mathcal{X}_{\mathrm{D}}$ is a subset of the product space $\prod_{I \in D} \mathcal{X}_{\mathrm{I}}$. We write pr_{I} for the canonical projection $\mathrm{pr}_{\mathrm{I}} : \prod_{I \in D} \mathcal{X}_{\mathrm{I}} \to \mathcal{X}_{\mathrm{I}}$. The canonical mappings are just the restrictions $f_{\mathrm{I}} = \mathrm{pr}_{\mathrm{I}}|_{\mathcal{X}_{\mathrm{D}}}$ of the projections to the projective limit space. The product space may be interpreted as the set of all functions x with domain D that take values $x(I) \in \mathcal{X}_{\mathrm{I}}$. Consequently, the projective limit space is precisely the subset of those functions which commute with the mappings f_{JI}, in the sense that $x(I) = (f_{\mathrm{JI}} \circ x)(J)$ whenever $I \preceq J$.

If the spaces $\mathcal{X}_{\mathrm{I}}$ are endowed with additional structure, and if the canonical mappings f_{JI} are chosen to preserve this structure under preimages, a corresponding structure is induced on the projective limit space. Two examples relevant in the following are topological and measurable spaces. Suppose that each space $\mathcal{X}_{\mathrm{I}}$ carries a topology $\mathrm{Top}_{\mathrm{I}}$ and a σ-algebra $\mathcal{B}_{\mathrm{I}}$. The system $\langle \mathcal{X}_{\mathrm{I}}, \mathrm{Top}_{\mathrm{I}}, f_{\mathrm{JI}} \rangle_{\mathrm{D}}$ is called a *projective system of topological spaces* if each f_{JI} is $\mathrm{Top}_{\mathrm{J}}$-$\mathrm{Top}_{\mathrm{I}}$-continuous. The *projective limit topology* $\mathrm{Top}_{\mathrm{D}}$ is defined as $\mathrm{Top}_{\mathrm{D}} := \mathrm{Top}(f_{\mathrm{I}}; I \in D)$, the coarsest topology which makes all canonical mappings f_{I} continuous. Analogously, $\langle \mathcal{X}_{\mathrm{I}}, \mathcal{B}_{\mathrm{I}}, f_{\mathrm{JI}} \rangle_{\mathrm{D}}$ is *projective system of measurable spaces* if the f_{JI} are measurable, and $\mathcal{B}_{\mathrm{D}} := \sigma(f_{\mathrm{I}}; I \in D)$ is called the *projective limit σ-algebra*. If the σ-algebras are the Borel sets generated by the topologies $\mathrm{Top}_{\mathrm{I}}$, then $\mathcal{B}_{\mathrm{D}} = \sigma(\mathrm{Top}_{\mathrm{D}})$. The general theme is that the mappings f_{JI} are chosen to be compatible with the structure defined on the spaces $\mathcal{X}_{\mathrm{I}}$, and the projective limit structure is the one generated by the canonical maps f_{I}. In a similar manner, projective limits can be defined for a range of other structures, such as groups (with homomorphisms f_{JI}), etc.

Suppose now that two families of spaces $\langle \mathcal{X}_{\mathrm{I}} \rangle_{\mathrm{D}}$ and $\langle \mathcal{Y}_{\mathrm{I}} \rangle_{\mathrm{D}}$ are jointly indexed by the same directed set D, and connected by a family $\langle w_{\mathrm{I}} \rangle_{\mathrm{D}}$ of mappings. If the mappings commute with the projection maps, they define a projective limit mapping between the respective projective limit spaces.

LEMMA 9 (Projective limits of functions [8, III.7.2]). *Let $\mathcal{D}^x := \langle \mathcal{X}_I, f_{JI} \rangle_D$ and $\mathcal{D}^y := \langle \mathcal{Y}_I, g_{JI} \rangle_D$ be two projective systems with a common index set D. For each $I \in D$, let $w_I : \mathcal{X}_I \to \mathcal{Y}_I$. Require that the mappings satisfy*

$$g_{JI} \circ w_J = w_I \circ f_{JI} \,. \tag{A.3}$$

Then there exists a unique mapping $w_D : \mathcal{X}_D \to \mathcal{Y}_E$ such that $g_I \circ w_D = w_I \circ f_I$ for all I. In other words, the diagram on the right below commutes if and only if the diagram on the left commutes for all $I \preceq J \in D$:

$$\begin{array}{ccc} \mathcal{X}_J & \xrightarrow{w_J} & \mathcal{Y}_J \\ \downarrow f_{JI} & & \downarrow g_{JI} \\ \mathcal{X}_I & \xrightarrow{w_I} & \mathcal{Y}_I \end{array} \qquad\qquad \begin{array}{ccc} \mathcal{X}_D & \xrightarrow{w_D} & \mathcal{Y}_E \\ \downarrow f_I & & \downarrow g_I \\ \mathcal{X}_I & \xrightarrow{w_I} & \mathcal{Y}_I \end{array}$$

A number of useful properties of mappings are preserved under projective limits [7, 8]. If each w_{I} is injective or bijective, then so is w_{D}. Projective systems $\mathcal{D}^x$, $\mathcal{D}^y$ of topological spaces preserve continuity, i.e. w_{D} is $\mathrm{Top}^x_{\mathrm{D}}$-$\mathrm{Top}^y_{\mathrm{D}}$-continuous if and only if each w_{I} is continuous. Projective systems of measurable spaces preserve measurability (Lemma 1); projective systems of algebraic structures preserve homomorphy, etc. A notable exception is that w_{D} need not be surjective, even if all w_{I} are.

In a similar manner, projective limits can be defined for set functions, and in particular for probability measures P_{I}. The domains $\mathcal{X}_{\mathrm{I}}$ of the maps w_{I} above are replaced by the σ-algebras $\mathcal{B}_{\mathrm{I}}$, and the ranges $\mathcal{Y}_{\mathrm{I}}$ by $[0,1]$. We denote by $f_{\mathrm{JI}}(P_{\mathrm{J}})$ the image measure under projection, i.e. $f_{\mathrm{JI}}(P_{\mathrm{J}}) = P_{\mathrm{J}} \circ f_{\mathrm{JI}}^{-1}$.

THEOREM 4 (Kolmogorov; Bochner [6]). *Let $\langle \mathcal{X}_I, \mathcal{B}_I, f_{JI} \rangle_D$ be a projective system of Polish measurable spaces with countable index set D, and $\langle P_I \rangle_D$ a family of probability measures on these spaces. If the measures commute with projection, that is if $f_{JI}(P_J) = P_I$ whenever $I \preceq J$, there exists a uniquely defined probability measure P_D on the projective limit space $(\mathcal{X}_D, \mathcal{B}_D)$ such that $f_I(P_D) = P_I$ for all $I \in D$.*

The image measure $f_{\mathrm{JI}}(P_{\mathrm{J}})$ is referred to as a *marginal* of P_{J}, and whenever $\mathcal{X}_{\mathrm{I}} \subset \mathcal{X}_{\mathrm{J}}$ is exactly the subspace marginal of P_{J} on $\mathcal{X}_{\mathrm{I}}$. The theorem generalizes to the case of uncountable index sets, but then requires additional conditions to ensure $\mathcal{X}_{\mathrm{D}} \neq \emptyset$. The most commonly used condition is Bochner's "sequential maximality" [6]. Kolmogorov originally proved the theorem for product spaces, for which sequential maximality is automatically satisfied.

A.2. Pullbacks of Measures and Functions. Projective limit constructions of stochastic processes raise two problems: One is the effective restriction to countable index sets. The other is that a construction from finite-dimensional marginals can only express properties of the constructed random functions that are verifiable at finite subsets of points (such as non-negativity), but not infinitary properties (such as continuity, or countable additivity of set functions). Both problems can be addressed simultaneously by means of a *pullback*, defined via the following existence result.

LEMMA 10 (Pullback measure [19, Section 132G]). *Let $\mathcal{X}$ be a set, $(\mathcal{Y}, \mathcal{B}_y, \nu)$ be a measure space and $\mathcal{J} : \mathcal{X} \to \mathcal{Y}$ any function. If $\mathcal{J}(\mathcal{X})$ has full outer measure under ν, that is if $\nu^*(\mathcal{J}\mathcal{X}) = \nu(\mathcal{Y})$, there is a uniquely defined measure $\tilde{\nu}$ on $(\mathcal{X}, \mathcal{J}^{-1}\mathcal{B}_y)$ such that*

$$\tilde{\nu} \circ \mathcal{J}^{-1} = \nu \ . \tag{A.4}$$

The measure μ defined by (A.4) is called the *pullback* of ν under $\mathcal{J}$. If the pullback exists, ν can be represented as the image measure $\nu = \mathcal{J}(\tilde{\nu})$. The outer measure condition $\nu^*(\mathcal{J}(\mathcal{X})) = \nu(\mathcal{Y})$ ensures that the definition of $\tilde{\nu}$ by means of the assignment $\tilde{\nu}(\mathcal{J}^{-1}A) := \nu(A)$ is unambiguous: If $A, B \in \mathcal{B}_{\mathcal{Y}}$ are two sets, $\mathcal{J}^{-1}A = \mathcal{J}^{-1}B$ does not imply $A = B$. Hence, ν may assign different measures to A and B, in which case it is not possible to assign a consistent value to $\mathcal{J}^{-1}A = \mathcal{J}^{-1}B$ under the pullback. However, this problem does not occur on the image $\mathcal{J}(\mathcal{X})$, since $\mathcal{J}^{-1}A = \mathcal{J}^{-1}B$ *does* imply $(A \triangle B) \cap \mathcal{J}\mathcal{X} = \emptyset$. Thus, if $\nu^*(\mathcal{J}\mathcal{X}) = \nu(\mathcal{Y})$, any differences between A and B are consistently assigned measure zero.

The arguably most important application of pullbacks of measures is the restriction of a measure to a non-measurable subspace:

Example 4 (Restriction to subspaces). Let $\mathcal{X} \subset \mathcal{Y}$ be an arbitrary subspace, and ν a measure on $\mathcal{Y}$. If the subspace has full outer measure $\nu^*(\mathcal{X}) = \nu(\mathcal{Y})$, the measure ν has a uniquely defined pullback $\tilde{\nu}$ under the canonical inclusion map $\mathcal{X} \hookrightarrow \mathcal{Y}$. The measure ν lives on the the measurable space $(\mathcal{X}, \mathcal{B}_y \cap \mathcal{X})$, and assigns measure $\tilde{\nu}(A \cap \mathcal{X}) = \nu(A)$ to each intersection of a measurable set $A \in \mathcal{B}_{\mathcal{Y}}$ with $\mathcal{X}$. Hence, $\tilde{\nu}$ can be regarded as the restriction of ν to $\mathcal{X}$.

As for measures, pullbacks can be defined for functions. Let $\mathcal{J}_X : \tilde{\mathcal{X}} \to \mathcal{X}$ and $\mathcal{J}_Y : \tilde{\mathcal{Y}} \to \mathcal{Y}$ be two functions. A *pullback of a function* $f : \mathcal{X} \to \mathcal{Y}$ is any function $\tilde{f} : \tilde{\mathcal{X}} \to \tilde{\mathcal{Y}}$ for which the following diagram commutes:

$$\begin{array}{ccc} \tilde{\mathcal{X}} & \xrightarrow{\ \tilde{f}\ } & \tilde{\mathcal{Y}} \\ {\scriptstyle \mathcal{J}_X}\big\downarrow & & \big\downarrow{\scriptstyle \mathcal{J}_Y} \\ \mathcal{X} & \xrightarrow[\ f\ ]{} & \mathcal{Y} \end{array}$$

Conversely, if $\tilde{f}$ is given, any function f for which the diagram commutes is called a *pushforward* of $\tilde{f}$.

The definitions of pullbacks for measures and functions are compatible, in the sense that the simultaneous pullback of a measure and an integrable function under the same mapping preserves the integral: Let $\mathcal{Y} = \tilde{\mathcal{Y}} = \mathbb{R}$, and let $(X, \mathcal{C}, \nu)$ be a measure space such that $\mathcal{J}_X\tilde{\mathcal{X}}$ has full outer measure $\nu^*(\mathcal{J}_X\tilde{\mathcal{X}}) = \nu(\mathcal{X})$. Let f be $\mathcal{C}$-measurable, non-negative and ν-integrable. Then $\tilde{f}$ is $\mathcal{J}_X^{-1}\mathcal{C}$-measurable and $\tilde{\nu}$-integrable. Since ν is the image measure of $\tilde{\nu}$ under $\mathcal{J}_X$,

$$\int_{\mathcal{J}_X^{-1}C} \tilde{f}\tilde{\nu} = \int_C f d(\mathcal{J}_X\tilde{\nu}) = \int_C f d\nu \ . \tag{A.5}$$

APPENDIX B: PROOF OF THEOREM 3

PROOF OF (1). Let k_{I} be the probability kernel corresponding to $(T_{\text{I}}^{(n)})_n$ in (6.1). Then $\langle k_{\text{I}} \rangle_{\text{D}}$ is a projective family of probability kernels, and by Theorem 1, its projective limit k_{D} is a version of $P_{\text{D}}^{\theta}[\Theta_{\text{D}}|Y_{\text{D}}]$. To show that $(T_{\text{D}}^{(n)})_n$ is a valid posterior index, we have to verify for all $x^n \in \mathcal{X}_{\text{D}}^n$, $y \in \mathcal{Y}_{\text{D}}$

$$k_{\text{D}}(A, T_{\text{D}}^{(n)}(x^n, y)) =_{\text{a.e.}} P_{\text{D}}^{\theta}[A|X_{\text{D}}^n = x^n, Y_{\text{D}} = y] \;. \tag{B.1}$$

By construction, the following identities hold for all $I \in D$:

$$\begin{aligned} P_{\text{D}}^{\theta}[g_{\text{I}}^{-1}\,.\,|X_{\text{D}}^n = x^n, Y_{\text{D}} = y] &=_{\text{a.e.}} P_{\text{I}}^{\theta}[\,.\,|X_{\text{I}}^n = f_{\text{I}}^n x^n, Y_{\text{I}} = h_{\text{I}} y] \\ k_{\text{D}}(g_{\text{I}}^{-1}\,.\,, h_{\text{I}} T_{\text{D}}^{(n)}(x^n, y)) &=_{\text{a.e.}} k_{\text{I}}(\,.\,, T_{\text{I}}^{(n)}(f_{\text{I}}^n x^n, h_{\text{I}} y)) \end{aligned} \tag{B.2}$$

Conjugacy of the models in the projective family implies

$$k_{\text{I}}(\,.\,, T_{\text{I}}^{(n)}(f_{\text{I}}^n x^n, h_{\text{I}} y)) =_{\text{a.e.}} P_{\text{I}}^{\theta}[\,.\,|X_{\text{I}}^n = f_{\text{I}}^n x^n, Y_{\text{I}} = h_{\text{I}} y] \;. \tag{B.3}$$

Since these quantities are probability measures, application of Theorem 4 to both sides of the equation yields the terms on either side of (B.1), and almost sure equality in (B.1) follows from the uniqueness of projective limits.

The second assertion – conjugacy of the marginals implies conjugacy of the limit model under $(T_{\text{D}}^{(n)})_n$ – is an immediate consequence. □

For part (2), the existence of a posterior index for each marginal is established by means of the following lemma.

LEMMA 11. *Let $\mathcal{J}_{\mathcal{X}} : \tilde{\mathcal{X}} \to \mathcal{X}$ and $\mathcal{J}_{\mathcal{Y}} : \tilde{\mathcal{Y}} \to \mathcal{Y}$ be continuous functions between Polish spaces. Suppose that a measurable function $\tilde{f} : \tilde{\mathcal{X}} \to \tilde{\mathcal{Y}}$ is given. If $\mathcal{J}_{\mathcal{X}}$ is open or closed, there exists a pushforward of $\tilde{f}$ which is measurable.*

The proof of the lemma draws on the concept of a *selector* [e.g. 27]. For a given correspondence (equivalence relation) R on a product $A \times B$, a selector is function $\beta : A \to B$ with $f(a) \in R(a)$, i.e. an assignment which transforms the set-valued map $a \mapsto R(a)$ into a function, by selecting a single element of the set $R(a)$ for each a. By the axiom of choice, a selector always exists, but will usually be too complicated to be of any use. Under additional regularity conditions on the correspondence and the underlying spaces, the selection theorem of Kuratowski and Ryll-Nardzewski [27] guarantees the existence of a Borel-measurable selector. The correspondence of interest here is the preimage $\mathcal{J}_{\mathcal{X}}^{-1}$.

PROOF OF LEMMA 11. By the selection theorem [27, Theorem 12.16], a correspondence between Polish spaces admits a measurable selector if it is weakly measurable and takes non-empty closed values. We have to show that $\mathcal{J}_{\mathcal{X}}^{-1}$ satisfies these conditions. The upper inverse under the correspondence $\mathcal{J}_{\mathcal{X}}^{-1}$ of a set $A \subset \tilde{\mathcal{X}}$ is by definition $\{x \in \mathcal{X} | \mathcal{J}_{\mathcal{X}}^{-1} x \subset A\}$, which in this case is just $\mathcal{J}_{\mathcal{X}}(A)$. If $\mathcal{J}_{\mathcal{X}}$ is open, the upper inverse $\mathcal{J}_{\mathcal{X}}(A)$ of any open set $A \subset \tilde{\mathcal{X}}$ is in $\mathcal{B}(\mathcal{X})$, which makes the correspondence weakly measurable. Similaraly, if $\mathcal{J}_{\mathcal{X}}$ is closed, the upper inverse of any closed set is in $\mathcal{B}(\mathcal{X})$, hence $\mathcal{J}_{\mathcal{X}}^{-1}$ is measurable, and in particular weakly measurable since $\tilde{\mathcal{X}}$ is Polish. The singletons are closed, hence by continuity, $\mathcal{J}_{\mathcal{X}}^{-1} x$ is closed, and as a preimage non-empty. We note that the analoguous result for pullbacks instead of pushforwards follows *mutatis mutandis*. □

PROOF OF (2). Let k_{D} be the kernel corresponding to the posterior index $(T_{\mathrm{D}}^{(n)})_n$ which makes the projective limit model conjugate. The marginals form a projective system. Hence for any $I \in D$ and $A_{\mathrm{I}} \in \mathcal{B}(\mathcal{T}_{\mathrm{I}})$,

$$k_{\mathrm{D}}(g_{\mathrm{I}}^{-1} A_{\mathrm{I}}, h_{\mathrm{I}} T_{\mathrm{D}}^{(n)}(x_{\mathrm{D}}^n, y_{\mathrm{D}})) =_{\text{a.e.}} P_{\mathrm{I}}^{\theta}[A_{\mathrm{I}} | Y_{\mathrm{I}} = h_{\mathrm{I}} T_{\mathrm{D}}^{(n)}(x_{\mathrm{D}}^n, y_{\mathrm{D}})] \tag{B.4}$$

is a valid version of the posterior $P_{\mathrm{I}}^{\theta}[A_{\mathrm{I}} | X_{\mathrm{I}}^n = f_{\mathrm{I}}^n x_{\mathrm{D}}^n, Y_{\mathrm{I}} = h_{\mathrm{I}} y_{\mathrm{D}}]$. Since the mappings are surjective, any hyperparameter y_{I} and sample x_{I}^n is representable in this form, and the marginal model is closed under sampling since $h_{\mathrm{I}} T_{\mathrm{D}}^{(n)}(x_{\mathrm{D}}^n, y_{\mathrm{D}}) \in \mathcal{Y}_{\mathrm{I}}$. By the same identity, any measurable mappings $(T_{\mathrm{I}}^{(n)})_n$ satisfying (6.4) form a posterior index of the marginal model. A mapping $T_{\mathrm{I}}^{(n)}$ satisfies (6.4) if it is a pushforward of $T_{\mathrm{D}}^{(n)}$. If the canonical mappings are open or closed, the existence of such measurable mappings $T_{\mathrm{I}}^{(n)}$ follows from Lemma 11. □

APPENDIX C: PROOFS FOR SECTION 7

PROOF OF LEMMA 5. Fix any admissible metric d on V, and let $U \subset V$ be a dense, countable subset. Let $\mathcal{U}$ be the set of closed balls of the form

$$\mathcal{U} := \{\bar{B}(v, r) | v \in U, r \in \mathbb{Q}_+\} , \tag{C.1}$$

and let $\mathcal{Q}$ be the smallest algebra containing $\mathcal{U}$. For this choice of $\mathcal{Q}$, countable addivity of the random charge can be directly related to countable additivity of the parameter G_0 [41, Theorem 1], and we obtain

$$P_{\mathrm{D}}^{\theta,*}[M(\mathcal{Q}) | Y_{\mathrm{D}} = (\alpha, G_0)] = 1 \quad \Leftrightarrow \quad G_0 \in M(\mathcal{Q}) \tag{C.2}$$

□

Proof of Lemma 6. (1) *ϕ is a mapping:* We have to argue that, whenever the set $W \subset V$ is a singleton, there is exactly one $x_{\mathrm{D}} \in \mathcal{X}_{\mathrm{D}}$ with $\lim x_{\mathrm{D}} = W$. For any $x_{\mathrm{D}} = \{C_{\mathrm{I}} | I \in D\} \in \mathcal{X}_{\mathrm{D}}$, by definition, $\lim x_{\mathrm{D}} \subset C_{\mathrm{I}}$ for all I. Hence, every partition $I \in D$ contains exactly one set A_{I} with $v \in A_{\mathrm{I}}$. (Note that no such set need exist if W is not a singleton.) Therefore, $x_{\mathrm{D}} := \{A_{\mathrm{I}} | I \in D\}$ is the only element of $\mathcal{X}_{\mathrm{D}}$ satisfying $\lim x_{\mathrm{D}} = \{v\} = W$.
ϕ is measurable: Since the σ-algebra on $\mathcal{X}_{\mathrm{D}}$ is the projective limit $\mathcal{B}_{\mathrm{D}}$, ϕ is measurable if and only if each of the mappings $f_{\mathrm{I}} \circ \phi$ is measurable. For any $v \in V$, the image $(f_{\mathrm{I}} \circ \phi)(v) = A_{\mathrm{I}}$ is the unique set $A_{\mathrm{I}} \in I$ for which $v \in A_{\mathrm{I}}$. The preimage of $A_{\mathrm{I}} \in I$ is therefore simply

$$(f_{\mathrm{I}} \circ \phi)^{-1}\{A_{\mathrm{I}}\} = \{v \in V | v \in A_{\mathrm{I}}\} = A_{\mathrm{I}} \ , \tag{C.3}$$

and measurable since $A_{\mathrm{I}} \in \mathcal{Q} \subset \mathcal{B}_V$.
As a mapping onto its image, ϕ has a measurable inverse: By definition, ϕ is trivially injective. For measurability of the inverse on $\phi(V)$, we have to show $\phi(A) \in \mathcal{B}_{\mathrm{D}} \cap \phi(V)$ for every $A \in \mathcal{B}_V$, or equivalently, for every $A \in \mathcal{Q}$. For any $A \in \mathcal{Q}$, there is some $I \in D$ with $A \in I$, and hence $\{A\} \in \mathcal{B}_{\mathrm{I}}$. The singleton $\{A\}$ is the base of the cylinder $f_{\mathrm{JI}}^{-1}\{A\} = \{x_{\mathrm{D}} | \lim x_{\mathrm{D}} \subset A\} \in \mathcal{B}_{\mathrm{D}}$. Then $\phi(A) = f_{\mathrm{JI}}^{-1}\{A\} \cap \phi(V)$ and hence $\phi(A) \in \mathcal{B}_{\mathrm{D}} \cap \phi(V)$.

(2) Let θ_{D} be purely atomic of the form $\theta_{\mathrm{D}} = \sum_{i \in \mathbb{N}} c_i \delta_{v_i}$. We will show $P_{\mathrm{D}}(\{x_{\mathrm{D}}\} | \Theta_{\mathrm{D}} = \theta_{\mathrm{D}}) = c_i$ if $\lim x_{\mathrm{D}} = \{v_i\}$, that is, if $x_{\mathrm{D}} = \phi(v_i)$.

First observe that, for any $v \in V$, there is a decreasing sequence of sets $Q_n \in \mathcal{Q}$ with $\lim Q_n = \{v\}$. To see this, recall the definition of $\mathcal{Q}$: Since U is dense, there is a sequence $u_n \in U$ with $\lim u_n = v$ and $d(u_n, v) < \frac{1}{2n}$. Set $Q_n := \bar{B}(u_n, \frac{1}{2n})$. Hence, $v \in Q_n$ for all n, and $v \in \lim u_n$. On the other hand, $\bar{B}(u_n, \frac{1}{2n}) \subset \bar{B}(v, \frac{1}{n})$, and as the balls are compact, $\bar{B}(v, \frac{1}{n}) \searrow \{v\}$.

Consequently, there is a sequence $I_1 \preceq I_2 \preceq \dots$ of partitions in D such that $Q_n \in I_n$ for all n. In the representation $x_{\mathrm{D}} = \{C_{\mathrm{I}} | I \in D\}$, we therefore have $C_{\mathrm{I}_n} = Q_n$. For $x_{\mathrm{D}} = \phi(v_i)$,

$$P_{\mathrm{D}}[\{x_{\mathrm{D}}\} | \Theta_{\mathrm{D}} = \theta_{\mathrm{D}}] = \lim_{n \to \infty} P_{\mathrm{I}_n}[f_{\mathrm{I}_n}^{-1} Q_n | \Theta_{\mathrm{I}_n} = g_{\mathrm{I}_n} \theta_{\mathrm{D}}] = \lim_{n \to \infty} \theta_{\mathrm{D}}(Q_n) = c_i \ .$$

□

Proof of Lemma 7. To show that both $\langle P_n[\pi_n | \Theta_n] \rangle_{\mathrm{D}}$ and $\langle P_n^{\theta}[\Theta_n | Y_n] \rangle_{\mathrm{D}}$ are projective families of conditional distributions, we appeal to Lemma 3. First consider the models $P_n[\pi_n | \Theta_n]$. For $\pi_n = \sigma_{k_1}(1) \cdots \sigma_{k_n}(n)$, the preimage $f_{n+1,n}^{-1} \pi_n$ consists of the permutations $\pi_{n+1} = \sigma_{k_1}(1) \cdots \sigma_{k_n}(n) \sigma_m(n+1)$ for $m = 1, \dots, n+1$. For the sampling distributions, fix $\theta_{n+1} \in \mathcal{T}_{n+1}$, and let $\theta_n = \mathrm{pr}_{n+1,n}\theta_{n+1}$. Then

$$
\text{(C.4)} \quad
\begin{aligned}
P_{n+1}[f_{n+1,n}^{-1}\pi_n|\Theta_{n+1} = \theta_{n+1}] =& P_n[\pi_n|\Theta_n = \theta_n]\frac{(\sum_{m=1}^{n} e^{-\theta^{(n+1)}}) + 1}{1 + ne^{-\theta^{(n+1)}}} \\
=& P_n[\pi_n|\Theta_n = \theta_n]
\end{aligned}
$$

Lemma 3 requires a product space structure of the sample space and is thus not directly applicable on the groups $\mathbb{S}_n$. However, the encodings ψ_n map into a product space, and we may equivalently consider the image measures $\psi_n(P_n)$ on $\prod_{m\leq n}[m]$. By (C.4), the image measures under ψ_n satisfy

$$
\text{(C.5)} \qquad \mathrm{pr}_{n+1,n} \circ \psi_{n+1}(P_{n+1}[\,.\,|\Theta_{n+1} = \theta_{n+1}]) = \psi_n(P_n[\,.\,|\Theta_n = \theta_n])
$$

which establishes (3.8). By Lemma 3, the images form a projective family of conditional probabilities under the projections $\mathrm{pr}_{n+1,n}$, and hence by (7.7), so do $P_n[\pi_n|\Theta_n]$ under $f_{n+1,n}$.

For the priors, which are defined on the product spaces $\mathbb{R}^{n-1}$, Lemma 3 can be applied directly. We first note that, since $Z_n = \prod_j Z^{(j)}$, the partition function K_n factorizes as $K_n(\lambda, \gamma_n) = \prod_j K^{(j)}(\lambda, \gamma^{(j)})$. Therefore, the projection $(\mathrm{pr}_{n+1,n}P_{n+1}^{\theta})[\Theta_n|Y_n]$ has density

$$
\int \frac{p_n^{\theta}(\theta_n|\lambda, \gamma_n)e^{\theta^{(n+1)}\gamma^{(n+1)} - \lambda \log Z^{(n+1)}(\theta^{(n+1)})}}{K^{(n+1)}(\lambda, \gamma^{(n+1)})} d\theta^{(n+1)} = p_n^{\theta}(\theta_n|\lambda, \gamma_n) \ ,
$$

which establishes (3.8). Hence, $P_n^{\theta}[\Theta_n|\lambda, \gamma_n]$ is a projective family of conditionals by Lemma 3. □

PROOF OF LEMMA 8. (1) As a canonical inclusion, ϕ is an embedding and hence a Borel embedding. The mapping $q : \mathbb{R}^{\mathbb{N}} \to (0,1)^{\mathbb{N}}$ is injective and continuous, hence measurable. Its image $\ell_1(0,1)$ is a subset of the Polish space $(0,1)^{\mathbb{N}}$, and since convergence of a sequence in $(0,1)^{\mathbb{N}}$ is a measurable event in the tail σ-algebra, $\ell_1(0,1)$ is Borel and hence itself Polish. As a mapping onto its image, q is surjective, and as a measurable bijection between Polish spaces, it has a measurable inverse. Since $\mathcal{J}_{\ell_1}$ is again a canonical inclusion, the composition $\mathcal{J}_{\mathcal{T}} = \mathcal{J}_{\ell_1} \circ q$ is a Borel embedding.

(2) A virtual permutation π is an element of $\mathbb{S}_\infty$ if and only if $\sum_j W^{(j)}(\pi) < \infty$. If this is the case, all but a finite number of entries of π form their own cycle, and hence $\pi \in \mathbb{S}_\infty$. If the sum diverges, at least one cyclic set contains an infinite number of elements. The random variables $W^(j)(\pi)$ are independent under the model. Hence, by the Borel-Cantelli lemma, the sum

converges if and only if the sum of probabilities $\Pr\{W^{(j)}(\pi) = 1\}$ converges, i.e. if $q(\theta) \in \ell_1$, and hence if $\theta \in \tilde{\mathcal{T}}$.

(3) The random variables $\Theta_{\mathrm{D}}^{(j)}$ are independent given the hyperparameters. By the zero-one law, the event $\{\Theta_{\mathrm{D}} \in \tilde{\mathcal{T}}\} = \{q(\Theta_{\mathrm{D}}) \in \ell_1\}$ has probability either zero or one. The variables have expectation $\mathbb{E}[\Theta_{\mathrm{D}}^{(j)}] = \gamma_{\mathrm{D}}^{(j)}$. Each component of q by definition satisfies $q_j(t) \to 0$ if $t \to +\infty$. Hence, $\gamma_{\mathrm{D}} \in \tilde{\mathcal{T}} = q^{-1}(\ell_1)$ implies $\gamma_{\mathrm{D}}^{(j)} \to \infty$ as $j \to \infty$. Thus for any $\epsilon > 0$, the expectations satisify $\mathbb{E}[\Theta_{\mathrm{D}}^{(j)}] > \epsilon$ for a cofinite number of indices j, and $\Pr\{\Theta_{\mathrm{D}} \in \tilde{\mathcal{T}}\} = 1$. $\square$

REFERENCES

[1] Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric estimation. *Ann. Statist.*, **2**, 1152–1174.
[2] Bauer, H. (1996). *Probability Theory*. W. de Gruyter.
[3] Bernardo, J. M. and Smith, A. F. M. (1993). *Bayesian Theory*. J. Wiley & Sons.
[4] Bertoin, J. (2006). *Random Fragmentation and Coagulation Processes*. Cambridge University Press.
[5] Billingsley, P. (1995). *Probability and Measure*. J. Wiley & Sons.
[6] Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability*. University of California Press.
[7] Bourbaki, N. (1966). *Elements of Mathematics: General Topology I*. Hermann (Paris) and Addison-Wesley.
[8] Bourbaki, N. (1968). *Elements of Mathematics: Theory of Sets*. Hermann (Paris) and Addison-Wesley.
[9] Bourbaki, N. (2004). *Elements of Mathematics: Integration, Vol. I & II*. Springer.
[10] Burkholder, D. L. (1961). Sufficiency in the undominated case. *Ann. Math. Statist.*, **32**, 1191–1200.
[11] Choksi, J. R. (1958). Inverse limits of measure spaces. *Proc. Lond. Math. Soc.*, **3**, 321–342.
[12] Consonni, G. and Veronese, P. (2003). Enriched conjugate and reference priors for the Wishart family on symmetric cones. *Ann. Statist.*, **31**, 1491–1516.
[13] Dalal, S. R. and Hall, W. J. (1983). Approximating priors by mixtures of natural conjugate priors. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **45**, 278–286.
[14] Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *Ann. Statist.*, **7**(2), 269–281.
[15] Doob, J. L. (1953). *Stochastic Processes*. J. Wiley & Sons.
[16] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, **1**(2).
[17] Ferguson, T. S. and Klass, M. J. (1972). A representation of independent increment processes without gaussian components. *Ann. Math. Statist.*, **43**, 1634–1643.
[18] Fligner, M. A. and Verducci, J. S. (1986). Distance based ranking models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **48**(3), 359–369.
[19] Fremlin, D. H. (2000–2006). *Measure Theory*, volume I–V. Torres Fremlin.
[20] Ghosal, S. (2010). Dirichlet process, related priors and posterior asymptotics. In

N. L. Hjort *et al.*, editors, *Bayesian Nonparametrics*, pages 36–83. Cambridge University Press.
[21] Ghosh, J. K. and Ramamoorthi, R. V. (2002). *Bayesian Nonparametrics*. Springer.
[22] Halmos, P. R. and Savage, L. J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics. *Ann. Math. Stat.*, **20**, 225–241.
[23] Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.*, **18**, 1259–1294.
[24] Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G. (2010). *Bayesian Nonparametrics*. Cambrdige University Press.
[25] James, L. F., Lijoi, A., and Prnster, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scand. J. Stat.*, **36**, 76–97.
[26] Kallenberg, O. (2001). *Foundations of Modern Probability*. Springer, 2 edition.
[27] Kechris, A. S. (1995). *Classical Descriptive Set Theory*. Springer.
[28] Kerov, S. V., Olshanski, G. I., and Vershik, A. M. (1993). Harmonic analysis on the infinite symmetric group. *C. R. Math. Acad. Sci. Paris*, **316**, 773–778.
[29] Kerov, S. V., Olshanski, G. I., and Vershik, A. M. (2004). Harmonic analysis on the infinite symmetric group. *Invent. Math.*, **158**, 551–642.
[30] Kingman, J. F. C. (1975). Random discrete distributions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **37**, 1–22.
[31] Küchler, U. and Sørensen, M. (1997). *Exponential families of stochastic processes*. Springer.
[32] Kuo, H.-H. (1975). *Gaussian Measures in Banach Spaces*. Lecture Notes in Mathematics. Springer.
[33] Landers, D. and Rogge, L. (1972). Minimal sufficient σ-fields and minimal sufficient statistics. Two counterexamples. *Ann. Math. Statist.*, **43**, 2045–2049.
[34] Lane, S. M. (1998). *Categories for the Working Mathematician*. Springer.
[35] Lauritzen, S. L. (1984). Exreme point models in statistics (with discussion). *Scand. J. Stat.*, **11**, 65–91.
[36] Lauritzen, S. L. (1988). *Extremal Families and Systems of Sufficient Statistics*. Lecture Notes in Statistics. Springer.
[37] Lovász, L. and Szegedy, B. (2006). Limits of dense graph sequences. *J. Combin. Theory Ser. B*, **96**, 933–957.
[38] Magiera, R. and Wilczyński, M. (1991). Conjugate priors for exponential-type processes. *Statist. Probab. Lett.*, **12**, 379–384.
[39] Mallory, J. D. and Sion, M. (1971). Limits of inverse systems of measures. *Ann. Inst. Fourier (Grenoble)*, **21**, 25–57.
[40] Meilă, M. and Bao, L. (2008). Estimation and clustering with infinite rankings. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI)*.
[41] Orbanz, P. (2010). Projective limit random measures over Polish spaces. *Submitted*.
[42] Pitman, J. (2006). *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer.
[43] Raiffa, H. and Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Harvard University Press.
[44] Rao, M. M. (1971). Projective limits of probability spaces. *J. Multivariate Anal.*, **1**, 28–57.
[45] Rao, M. M. (2005). *Conditional Measures and Applications*. Chapaman & Hall, 2nd edition.
[46] Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press.
[47] Schervish, M. J. (1995). *Theory of Statistics*. Springer.

[48] Schwartz, L.-M. (1973). *Radon measures on arbitrary topological spaces and cylindrical measures*. Oxford University Press.

[49] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica*, **4**, 639–650.

[50] Walker, S. G., Damien, P., Laud, P. W., and Smith, A. F. M. (1999). Bayesian nonparametric inference for random distributions and related functions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **61**(3), 485–527.

[51] Zhao, L. H. (2000). Bayesian aspects of some nonparametric problems. *Ann. Statist.*, **28**, 532–552.