# Proofs

Construction of Nonparametric Bayesian Models from Parametric Bayes Equations

Peter Orbanz

## 1 Proof of Theorem 2

*Construction of the projective limit:* In the following, we have to explicitly treat the conditional $P_X^I(X^I|\Theta^I)$ as the function $P_X^I(A|\Theta^I)(\omega)$ for $A \in \mathcal{B}_x^I$ and $\omega \in \Omega$. As a function of $\omega$, the conditional is measurable w.r.t. the $\sigma$-algebra $\sigma(\Theta^I)$. As a regular conditional probability, the function $A \mapsto P_X^I(A|\Theta^I)(\omega)$ is a probability measure for $\mathbb{P}$-almost all $\omega \in \Omega$. The null set of exceptions will be denoted $N^I \subset \Omega$. Since the conditional probabilities are conditionally projective, we have $P_X^I(\,.\,|\Theta^I)(\omega) = P_X^J(\pi_{JI}^{-1}\,.\,|\Theta^J)(\omega)$ for almost all $\omega$. Again there is a null set of exceptions, which we will denote $N^{JI}$. Denote the union of all exceptions a $N := (\cup_I N^I) \cup (\cup_{I \subset J} N^{JI})$. As a countable union of null sets, $N$ is itself a null set. Now for any fixed $\omega \notin N$, the probability measures $P_X^I(\,.\,|\Theta^I)(\omega)$ form a projective family of measures in the sense of the Kolmogorov theorem. Application of the theorem yields a unique probability measure $\nu_\omega$ on $(\Omega_x^E, \mathcal{B}_x^E)$ for each $\omega \notin N$. Treat this collection of measures as a function $\nu(A, \omega) := \nu_\omega(A)$ for $\omega \notin N$, and set $\nu(A, \omega) := \delta_{X^E(\omega)}$ for $\omega \in N$, where $\delta_x$ denotes the Dirac measure concentrated at $x$. (The only purpose of the latter is to ensure that $v$ is a probability measure for every $\omega$; the choice of the Dirac measure is arbitrary.)
*$\mathcal{C}^E$-measurability:* The function $\nu(\,.\,,\,.\,)$ so obtained describes a conditional distribution of $X^E$ w.r.t. a $\sigma$-algebra $\mathcal{C}^E$ if we can show that $\omega \mapsto \nu(A, \omega)$ is $\mathcal{C}^E$-measurable for every $A \in \mathcal{B}_x^E$. This can be shown by means of the $\pi$-$\lambda$ theorem (also called the Dynkin lemma, [2]): First show that $\nu(A, \omega)$ is measurable for all $A$ in a generator of $\mathcal{B}_x^E$, and then deduce that this implies measurability for all $A$ by means of the $\pi$-$\lambda$ theorem. As a generator, we choose the "cylinder sets" $\mathcal{Z}^E = \{A \in \mathcal{B}_x^E | A = \pi_{EI}^{-1} A^I\}$, i.e. the set of all sets which are preimages under projection of some finite-dimensional event. Then $\mathcal{B}_x^E = \sigma(\mathcal{Z}^E)$, a fact used for example in the proof of the Kolmogorov theorem (cf [1]). For any $A \in \mathcal{Z}^E$, the function $\nu(A,\,.\,)$ is measurable: Since $\nu(\pi_{EI}^{-1} A^I, \omega) = P_X^I(A^I|\Theta^I)(\omega)$, the function $\omega \mapsto \nu(\pi_{EI}^{-1} A^I, \omega)$ is $\sigma(\Theta^I)$-measurable, and therefore $\mathcal{C}^E$-measurable as $\sigma(\Theta^I) \subset \mathcal{C}^E$. Let $\mathcal{L} \subset \mathcal{B}_x^E$ denote the system of all $A$ for which $\nu(A,\,.\,)$ is $\mathcal{C}^E$-measurable. In the sense of the $\pi$-$\lambda$ theorem, $\mathcal{L}$ is a $\lambda$-system: For $A = \Omega_x^E$, $\nu$ is constant hence measurable. Let $A \in \mathcal{L}$. Then $\nu(\complement A,\,.\,) = 1 - \nu(A,\,.\,)$, which is measurable. If $A_n \in \mathcal{L}$ is a pairwise disjoint sequence and $A' = \cup_n^\infty A_n$, then $\nu(A,\,.\,) = \lim_{n\to\infty} \sum_{i=1}^n \nu(A_n,\,.\,)$, which as a limit of measurable functions is measurable. It is well known that the cylinder sets $\mathcal{Z}^E$ form an algebra [1], so $\mathcal{Z}^E$ is in particular a $\pi$-system. Then by the $\pi$-$\lambda$ theorem,

$$\mathcal{B}_x^E = \sigma(\mathcal{Z}^E) = \mathcal{L} \subset \mathcal{B}_x^E\ . \tag{1}$$

In other words, the set of all sets $A$ for which $\omega \mapsto \nu(A, \omega)$ is $\mathcal{C}^E$-measurable is just $\mathcal{B}_x^E$. Therefore, $\nu(A, \omega)$ is a regular version of the conditional probability $P_X^E(A|\mathcal{C}^E)(\omega)$. By construction, its marginals are $\pi_{EI} P_X^E(\,.\,|\mathcal{C}^E)(\omega) = P_X^I(\,.\,|\sigma(\Theta^I))(\omega)$ almost everywhere.

*Interpreting $P_X^E(X^E|\mathcal{C}^E)$ as $P_X^E(X^E|\Theta^E)$:* Under the additional assumption $\pi_{JI}\Theta^J = \Theta^I$, define the variable $\Theta^E$ as $\Theta^E := \bigotimes_{i \in E} \Theta^{\{i\}}$. Then any conditional distribution given $\mathcal{C}^E$ can serve as a conditional given $\Theta^E$, since the $\sigma$-algebra $\sigma(\Theta^E)$ generated by $\Theta^E$ is just $\mathcal{C}^E$:

$$\begin{aligned}\sigma(\Theta^E) =& \Theta^{E,-1}(\mathcal{B}_\theta^E) = \Theta^{E,-1}\big(\cup_{I \in \mathcal{F}(E)} \sigma(\pi_{EI}^{-1}\mathcal{B}_\theta^I)\big) = \sigma\big(\cup_{I \in \mathcal{F}(E)} \Theta^{E,-1}\pi_{EI}^{-1}\mathcal{B}_\theta^I\big)\\ =& \sigma\big(\cup_{I \in \mathcal{F}(E)} \Theta^{I,-1}\mathcal{B}_\theta^I\big) = \sigma\big(\cup_{I \in \mathcal{F}(E)} \sigma(\Theta^{I,-1}\big) = \mathcal{C}^E\ .\end{aligned} \tag{2}$$

## 2 Proof of Theorem 3

*Proof of (1).* We have to construct a candidate for the probability kernel $k^E$, and show that $T$ is a posterior index for the projective limit posterior with kernel $k^E$. To this end we will show that the conditionals $P_\Theta^I(\Theta^I|T^I)$ are conditionally projective and define $k^E$ in terms of their projective limit. For each $I \in \mathcal{F}(E)$ and $A^I \in \mathcal{B}_\theta^I$, the function $\omega \mapsto k^I(A^I,\,.\,) \circ T^I \circ X^I(\omega)$ is $\sigma(T^I \circ X^I)$-measurable, and $k^I(A^I,\,.\,) \circ T^I \circ X^I(\omega) = P_\Theta^I(A^I|T^I)(\omega)$ a.e. Therefore, $k^I(A^I,\,.\,) \circ T^I \circ X^I$ is a version of $P_\Theta^I(A^I|T^I)$. The conditional probabilities $P_\Theta^I(\Theta^I|T^I)$ are conditionally projective:

$$\begin{aligned}P_\Theta^I(A^I|T^I = t^I) =& P_\Theta^I(A^I|X^I \in T^{I,-1}(t^I)) = P_\Theta^J(\pi_{JI}^{-1}A^I|X^J \in \pi_{JI}^{-1}T^{I,-1}(t^I))\\ =& P_\Theta^J(\pi_{JI}^{-1}A^I|X^J \in T^{J,-1}\pi_{JI}^{-1}(t^I)) = P_\Theta^J(\pi_{JI}^{-1}A^I|T^J \in \pi_{JI}^{-1}(t^I))\end{aligned} \tag{3}$$

Hence $P^{\mathrm{J}}_{\Theta}(\pi^{-1}_{\mathrm{JI}}A^{\mathrm{I}}|T^{\mathrm{J}}) = P^{\mathrm{I}}_{\Theta}(A^{\mathrm{I}}|T^{\mathrm{I}})$, which is just the definition of conditional projectiveness. By Theorem 2 there is an a.e.-unique projective limit of the form $P^{\mathrm{E}}_{\Theta}(\Theta^{\mathrm{E}}|\mathcal{C}^{\mathrm{E}})$, where $\mathcal{C}^{\mathrm{E}}$ is the $\sigma$-algebra

$$\mathcal{C}^{\mathrm{E}} := \sigma\big(\cup_{I\in\mathcal{F}(E)}\sigma(T^{\mathrm{I}})\big) \ . \tag{4}$$

It is straightforward to check that $\sigma(T) = \mathcal{C}^{\mathrm{E}}$, because $T$ satisfies Eq. (4). Therefore, the projective limit $P^{\mathrm{E}}_{\Theta}(\Theta^{\mathrm{E}}|\mathcal{C}^{\mathrm{E}})$ can serve as the conditional distribution $P^{\mathrm{E}}_{\Theta}(\Theta^{\mathrm{E}}|T)$. Now define a candidate for the kernel $k^{\mathrm{E}}$ as

$$k^{\mathrm{E}}(A,t) := P^{\mathrm{E}}_{\Theta}(A|T=t) \qquad \text{for all } A\in\mathcal{B}^{\mathrm{E}}_{\theta}, t\in\Omega^{\mathrm{E}}_{t} \ . \tag{5}$$

What remains to be shown is that $k^{\mathrm{E}}(A,T(x)) = P^{\mathrm{E}}_{\Theta}(A|X^{\mathrm{E}}=x)$ a.e. for all $A\in\mathcal{F}(E)$. If this identity can be shown to hold for $A\in\mathcal{Z}^{\mathrm{E}}$, then it holds for all $A$: Since $\sigma(\mathcal{Z}^{\mathrm{E}}) = \mathcal{B}^{\mathrm{E}}_{\theta}$, and since $\mathcal{Z}^{\mathrm{E}}$ is an algebra, the Carathéodory extension theorem is applicable to extend measures from $\mathcal{Z}^{\mathrm{E}}$ to $\mathcal{B}^{\mathrm{E}}_{\theta}$. Since the conditional probability $k^{\mathrm{E}}$ is a Markov kernel, the Carathéodory theorem can be applied pointwise in $x$. (For a conditional that is not a Markov kernel, the subset of exceptional points $x\in\Omega^{\mathrm{E}}_{x}$ on which the conditional is not unique depends on $A$. Over all $A$, these could then aggregate into a non-null set.) To show that the identity holds on $\mathcal{Z}^{\mathrm{E}}$, consider any $A\in\mathcal{Z}^{\mathrm{E}}$, i.e. there is some $I\in\mathcal{F}(E)$ such that $A = \pi^{-1}_{\mathrm{EI}}A^{\mathrm{I}}$. Then

$$\begin{aligned} k^{\mathrm{E}}(\pi^{-1}_{\mathrm{EI}}A^{\mathrm{I}}, t\in\pi^{-1}_{\mathrm{EI}}t^{\mathrm{I}}) =& P^{\mathrm{E}}_{\Theta}(\pi^{-1}_{\mathrm{EI}}A^{\mathrm{I}}|T\in\pi^{-1}_{\mathrm{EI}}t^{\mathrm{I}}) = P^{\mathrm{I}}_{\Theta}(A^{\mathrm{I}}|T^{\mathrm{I}}=t^{\mathrm{I}}) = P^{\mathrm{I}}_{\Theta}(A^{\mathrm{I}}|X^{\mathrm{I}}\in T^{\mathrm{I},-1}t^{\mathrm{I}}) \\ =& P^{\mathrm{E}}_{\Theta}(\pi^{-1}_{\mathrm{EI}}A^{\mathrm{I}}|X^{\mathrm{E}}\in\pi^{-1}_{\mathrm{EI}}T^{\mathrm{I},-1}t^{\mathrm{I}}) = P^{\mathrm{E}}_{\Theta}(\pi^{-1}_{\mathrm{EI}}A^{\mathrm{I}}|X^{\mathrm{E}}\in T^{-1}\pi^{-1}_{\mathrm{EI}}t^{\mathrm{I}}) \ , \end{aligned} \tag{6}$$

such that $k^{\mathrm{E}}(\pi^{-1}_{\mathrm{EI}}A^{\mathrm{I}}, T(x)) = P^{\mathrm{E}}_{\Theta}(\pi^{-1}_{\mathrm{EI}}A^{\mathrm{I}}|X^{\mathrm{E}}=x)$. By the Carathéodory theorem, this implies that $k^{\mathrm{E}}(\pi^{-1}_{\mathrm{EI}}A^{\mathrm{I}}, T(x)) = P^{\mathrm{E}}_{\Theta}(\pi^{-1}_{\mathrm{EI}}A^{\mathrm{I}}|X^{\mathrm{E}}=x)$, and hence $T$ is a posterior index for the projective limit posterior $P^{\mathrm{E}}_{\Theta}(\Theta^{\mathrm{E}}|X^{\mathrm{E}})$, and $k^{\mathrm{E}}$ is the probability kernel corresponding to $T$.

*Proof of (2).* To proof part (2), we have to show that the posterior index $T^{\mathrm{I}}$ and corresponding probability kernel $k^{\mathrm{I}}$ as specified in the theorem make each of the marginal Bayesian systems on the finite-dimensional subspaces $\Omega^{\mathrm{I}}_{x}$ conjugate. That is, we have to verify $k^{\mathrm{I}}(A^{\mathrm{I}}, T^{\mathrm{I}}(x^{\mathrm{I}})) = P^{\mathrm{I}}_{\Theta}(A^{\mathrm{I}}|X^{\mathrm{I}} = x^{\mathrm{I}})$. To this end, write

$$\begin{aligned} k^{\mathrm{I}}(A^{\mathrm{I}}, t^{\mathrm{I}}) =& k(\pi^{-1}_{\mathrm{EI}}A^{\mathrm{I}}, t\in\pi^{-1}_{\mathrm{EI}}t^{\mathrm{I}}) = P^{\mathrm{E}}_{\Theta}(\pi^{-1}_{\mathrm{EI}}A^{\mathrm{I}}|X^{\mathrm{E}}\in T^{-1}\pi^{-1}_{\mathrm{EI}}t^{\mathrm{I}}) = P^{\mathrm{I}}_{\Theta}(A^{\mathrm{I}}|X^{\mathrm{I}}\in\pi_{\mathrm{EI}}T^{-1}\pi^{-1}_{\mathrm{EI}}t^{\mathrm{I}}) \\ =& P^{\mathrm{I}}_{\Theta}(A^{\mathrm{I}}|X^{\mathrm{I}}\in\pi_{\mathrm{EI}}T^{\mathrm{I},-1}t^{\mathrm{I}}) \end{aligned} \tag{7}$$

Since, for each $x^{\mathrm{I}}$, there is some $t^{\mathrm{I}}$ such that $T^{\mathrm{I}}(x^{\mathrm{I}}) = t^{\mathrm{I}}$, this means:

$$k^{\mathrm{I}}(A^{\mathrm{I}}, t^{\mathrm{I}}) = P^{\mathrm{I}}_{\Theta}(A^{\mathrm{I}}|X^{\mathrm{I}}=x^{\mathrm{I}}) \qquad \Leftrightarrow \qquad x^{\mathrm{I}}\in T^{\mathrm{I},-1}(t^{\mathrm{I}}) \ , \tag{8}$$

and thus $P^{\mathrm{I}}_{\Theta}(A^{\mathrm{I}}|X^{\mathrm{I}} = x^{\mathrm{I}}) = k^{\mathrm{I}}(A^{\mathrm{I}}, T^{\mathrm{I}}(x^{\mathrm{I}}))$ as we had to show. In other words, the posterior $P^{\mathrm{I}}_{\Theta}(A^{\mathrm{I}}|X^{\mathrm{I}})$ is conjugate with posterior index $T^{\mathrm{I}}$ and probability kernel $k^{\mathrm{I}}$.


[1] H. Bauer. *Probability Theory*. W. de Gruyter, 1996.

[2] M. J. Schervish. *Theory of Statistics*. Springer, 1995.