Foundations of Nonparametric Bayesian Methods

Part II

Peter Orbanz

Overview: Today

1. Bayesian models
2. Construction of stochastic processes
3. Extension of conditional probabilities

Conditioning

Direct approach

Conditional probability of $X(\omega) \in A$ given that $X(\omega) \in B$:

$$\mu(A|B) := \frac{\mu(A \cap B)}{\mu(B)}$$

$\rightarrow$ no use if $\mu(B) = 0$ (think of Bayesian model on $\mathbb{R}^d$)

Abstract Conditional Probabilities

Measure-theoretic definition of conditionals is beyond scope of this talk.

In The Following

We ignore technical details and write $\mu(X|Y)$ or $\mu(A|Y)$ for the conditional probability of X (RV) or A (event) given Y.

Conditional Densities

If X, Y have joint density, $\mu(X|Y)$ has conditional density $p(x|y)$.

Bayesian Models

Parametric Family

Let $X : (\Lambda, \mathcal{A}) \to (\Omega_x, \mathcal{B}_x)$ and $\Theta : (\Lambda, \mathcal{A}) \to (\Omega_\theta, \mathcal{B}_\theta)$ be two random variables, and $\mu_X = X(\mathbb{P})$. Then the conditional distribution $\mu_X(X|\Theta)$ is called a *parametric family* of distributions (parameterized by $\theta \in \Omega_\theta$).

Bayesian model

If X observed and Θ unobserved, we call:

- $\mu_\Theta := \Theta(\mathbb{P})$ the *prior measure*
- $\mu_\Theta(\Theta|X)$ the *posterior measure*
- The overall model is called a Bayesian model.

Note: Not defined by a Bayes equation!

Bayes' Theorem

Problem:

Given the prior and the data, how can we determine the posterior? (Without exhaustive knowledge of $\mathbb{P}$, $\mathcal{A}$ etc)

Bayes Theorem

If the sampling model $\mu_X(X|\Theta)$ has density $p_{X|\theta}$, then:

$$\frac{d\mu_{\Theta|X}}{d\mu_{\Theta}}(\theta|x) = \frac{p_{X|\theta}}{\int p_{X|\theta} d\mu_{\theta}(\theta)}$$

for all x with $\int p_{X|\theta} d\mu_{\theta}(\theta) \notin \{0, \infty\}$.

Models With No Bayes Equation

For some models (e.g. DP) posterior $\ll$ prior *not* satisfied $\rightarrow$

Bayesian model, but no Bayes equation.

Bayesian Nonparametrics

Nonparametric Bayesian model

A Bayesian model with:

1. $\dim(\Omega_\theta) = \dim(\Omega_x) = +\infty$.
2. Model can be evaluated on partial observations.

Partial observation

Random quantity with d dimensions, only $m < d$ are observed.

Example: GP regression

GP draw is function f, but only finite number of values of f known.

Stochastic Process Models

Intuition

Stochastic process = ∞-dim probability distribution

Typical GP definition

"A Gaussian process is a probability distribution on an infinite collection of random variables X_t such that the marginal distribution for each finite subset $(t_1, \ldots, t_n)$ of indices is Gaussian."

$\rightarrow$ Existence? Uniqueness?

Stochastic Process Construction (1)

Stochastic process measure μ^{E}: Distribution of RV

$$X^{\mathrm{E}} : (\Lambda, \mathcal{A}) \to (\Omega^{\mathrm{E}}, \mathcal{B}^{\mathrm{E}})$$

- E: infinite index set (indexes entries of random vector)
- Ω_0: "one-dimensional" sample space
- $\Omega^{\mathrm{E}} := \prod_{i \in E} \Omega_0$
- Interpretation: μ^{E}-draws = mappings $x : E \to \Omega_0$

Projector

P_{JI}:= projection mapping $\Omega^{\mathrm{J}} \to \Omega^{\mathrm{I}}$ (for $I \subset J \subset E$)

Marginals

Marginal of μ^{J} on $\Omega^{\mathrm{I}} \subset \Omega^{\mathrm{J}}$:

$$\underbrace{(\mathrm{P}_{\mathrm{JI}}\mu^{\mathrm{J}})(A)}_{\text{on } \Omega^{\mathrm{I}}} := \underbrace{\mu^{\mathrm{J}}(\mathrm{P}_{\mathrm{JI}}^{-1}A)}_{\text{on } \Omega^{\mathrm{J}}}$$

marginals = projections of measures

Stochastic Process Construction (2)

Def: Projective family

Family $\{\mu^I | I \subset E \text{ finite}\}$ such that for all finite I, J with $I \subset J$:

$$\mathrm{P}_{JI}\mu^J = \mu^I$$

Note: If μ^E given, the finite-dim marginals $\mu^I := \mathrm{P}_{EI}\mu^E$ are a projective family.

Kolmogorov's Extension Theorem

If a family $\{\mu^I | I \subset E \text{ finite}\}$ of finite-dimensional measures is projective, there exists a unique measure μ^E on Ω^E with μ^I as its marginals.

Jargon: μ^E is called the *projective limit* of the μ^I.

Example: GP construction

Choice of components

- $\Omega_0 := \mathbb{R}$ and index set $E = \mathbb{R}$
- P_{JI}: Euclidean projector from $\mathbb{R}^{|J|}$ to $\mathbb{R}^{|I|}$.
- Marginal family: μ^{I} are $|I|$-dimensional Gaussians

Ensure marginals projective

- Start with mean function $m(\,.\,)$ and covariance $k(\,.\,,\,.\,)$.
- Note: $E = \mathbb{R}$, finite $I = \{t_1, \dots, t_{|I|}\} \subset \mathbb{R}$
- μ^{I} = Gaussian, mean $(m(t_1), \dots, m(t_{|I|}))$ and $\Sigma_{ij} = k(t_i, t_j)$

Apply Extension Theorem

GP measure μ^{E} exists and is unique.

Note: μ^{E} has mean m and covariance function k, but that is *not* an immediate consequence of theorem!

Extensions Theorem: Caveat

Problem

If dimension E is uncountable, the projective limit measure μ^{E} is basically useless.

Explanation

- Domain of μ^{E}: $\mathcal{B}^{\mathrm{E}}$ (generated by product topology)
- Sets in $\mathcal{B}^{\mathrm{E}}$: "axes-parallel" in all but countably many dimensions
- E uncountable $\rightarrow$ $\mathcal{B}^{\mathrm{E}}$ too coarse for meaningful modeling

A Note of Caution:

Problem is often neglected in literature.
Example: Original paper on the DP (Ferguson, 1973).

Uncountable Dimensions

Intuition:

Objects of interest *effectively* have countably many degrees of freedom.

Examples

- Continuous functions: Completely defined by values on dense subset (e.g. $\mathbb{Q}$ in $\mathbb{R}$)
- Probability measures: Completely defined by values on countable system of sets.

Strategies

1. Modify theorem to directly define measure on “interesting” space (eg space of continuous functions).
2. Use Kolmogorov theorem, than restrict μ^{E} to interesting subspace.

Extension of Conditional Probabilities

Motivation

Bayesian estimation deals with conditional probabilities or parametric families, rather than individual distributions.

Extension Result

Assumptions:

- E countable
- Conditionals on subspaces Ω^{I} satsify

$$\mu^{\mathrm{J}}(\mathrm{P}_{\mathrm{JI}}^{-1}\,.\,|\Theta^{\mathrm{J}}) = \mu^{\mathrm{I}}(\,.\,|\Theta^{\mathrm{I}}) \qquad \text{for } I \subset J$$

Then there is a conditional distribution $\mu^{\mathrm{E}}(X^{\mathrm{E}}|\Theta^{\mathrm{E}})$ on Ω^{E} with marginals $\mu^{\mathrm{I}}(\,.\,|\Theta^{\mathrm{I}})$.

Disclaimer

Result statement above neglects some technical details.

Conjugate Models

Definition 1

A likelihood and a family of priors are *conjugate* if all possible posteriors are elements of the prior family. ("Closure under sampling")

Definition 2

Likelihood and prior family are conjugate if there exists a measurable mapping of the form

$$\text{Prior parameters} \times \text{Data} \rightarrow \text{Posterior parameters}$$

In Exponential Family Models

Mapping T to posterior parameters:

$$(\lambda, y) \overset{T}{\mapsto} (\lambda + n, y + \sum_{i=1}^{n} S(x_i))$$

Conjugate Projective Limits

Extension Result: In Short

If mappings to posterior parameters satisfy projection relation, they define corresponding mapping for projective limit model.

In Detail

- $T^{\mathrm{I}}(x^{\mathrm{I}}, y^{\mathrm{I}})$ mappings to posterior parameters
- Fix y^{E} and write $T^{\mathrm{I}}_{y} = T^{\mathrm{I}}(\,.\,, \mathrm{P}_{\mathrm{EI}} y^{\mathrm{E}})$

If some mapping T^{E} satisfies

$$\mathrm{P}^{-1}_{\mathrm{EI}} \circ T^{\mathrm{I},-1}_{y} = T^{\mathrm{E},-1}_{y} \circ \mathrm{P}^{-1}_{\mathrm{EI}}$$

then T^{E} defines functional conjugacy for limit model.

For Exponential Family Marginals

If S^{E} sufficient for extension:

$$(\lambda, y^{\mathrm{E}}) \overset{T^{\mathrm{E}}}{\mapsto} (\lambda + n, y^{\mathrm{E}} + \sum_i S^{\mathrm{E}}(x^{\mathrm{E}}_i))$$

Projective Limits of Bayes Equations

$$
\begin{tikzcd}
\mu^{\mathsf{E}}(\Theta^{\mathsf{E}}|X^{\mathsf{E}}, Y^{\mathsf{E}}) \arrow[d, shift right, "\mathrm{P}_{\mathrm{EI}}"'] & \mu^{\mathsf{E}}(\Theta^{\mathsf{E}}|Y^{\mathsf{E}}) \arrow[l, "x_1^{\mathsf{E}},\dots,x_n^{\mathsf{E}}"', "T^{\mathsf{E}}"] \arrow[d, shift left, "\mathrm{P}_{\mathrm{EI}}"] \\
\mu^{\mathsf{I}}(\Theta^{\mathsf{I}}|X^{\mathsf{I}}, Y^{\mathsf{I}}) \arrow[u, shift right, "\varprojlim"'] & \mu^{\mathsf{I}}(\Theta^{\mathsf{I}}|Y^{\mathsf{I}}) \arrow[l, "x_1^{\mathsf{I}},\dots,x_n^{\mathsf{I}}"', "T^{\mathsf{I}}"] \arrow[u, shift left, "\varprojlim"]
\end{tikzcd}
$$

Model Constructions

Example Models

Marginals	Proj Limit	Observations
Bernoulli/beta	IBP/beta process	Binary arrays
Multin./Dirichlet	CRP/DP	Discrete dist.
Gauss/Gauss	GP/GP	cont. functions
Mallows/conj.	(exists)	Bijections of $\mathbb{N}$

Construction Recipe

- Choose finite-dimensional observation (eg permutations)
- Choose exponential family model on observations
- Choose canonical conjugate prior
- Check: Model and sufficient statistic projective

Warning: More difficult in uncountable dimensions.