Species Sampling Models and Posterior Consistency

Jaeyong Lee

Department of Statistics
Seoul National University

October 24, 2011

A Simple Nonparametric Problem

- Suppose $X_1, X_2, \ldots, X_n | F \sim F$ and

$$F \in M(\mathbb{R}) = \{ \text{ all probability measures on } \mathbb{R} \}.$$

- To tackle this nonparametric problem in a Bayesian way, we need a class of priors on $M(\mathbb{R})$ or a class of probability measures on the space of probability measures.
- The Dirichlet process and species sampling models are probability measures on $M(\mathbb{R})$ developed for this purpose.

Dirichlet Process on $\mathbb{R}$ (Ferguson 1973)

- Let α be a finite nonnull measure on $(\mathbb{R}, \mathcal{B})$, where $\mathbb{R}$ is the real line and $\mathcal{B}$ is the class of Borel sets.
- We say that the random probability measure P on $\mathbb{R}$ follows the Dirichlet process with parameter α, if for every partition $B_1, \ldots, B_k$ of $\mathbb{R}$ by Borel sets,

$$(P(B_1), \ldots, P(B_k)) \sim Dirichlet(\alpha(B_1), \ldots, \alpha(B_k)).$$

- Notation:

$$P \sim DP(\alpha).$$

Properties of Dirichlet process

$P \sim DP(\alpha)$ and $X_1, \ldots, X_n | P \sim P$.
Then,

- (Conjugacy) $P | X_1, \ldots, X_n \sim DP(\alpha + \sum_{i=1}^n \delta_{X_i})$.
- (Marginalization Property, Blackwell and MacQueen 1973) marginally $(X_1, X_2, \ldots)$ forms a Polya urn sequence:

$$\begin{aligned} X_1 &\sim \alpha / \alpha(\mathcal{X}) \\ X_{n+1} | X_1, \ldots, X_n &\sim \frac{\alpha + \sum_{i=1}^n \delta_{X_i}}{\alpha(\mathcal{X}) + n}, \ n \geq 1. \end{aligned}$$

- (Sethuraman's Representation)
  Let α be a finite measure on $\mathcal{X}$ and let

$$\begin{aligned} \theta_1, \theta_2, \ldots &\overset{iid}{\sim} Beta(1, \alpha(\mathcal{X})) \\ Y_1, Y_2, \ldots &\overset{iid}{\sim} \alpha/\alpha(\mathcal{X}) \end{aligned}$$

  and they are independent of each other. Define

$$\begin{aligned} p_1 &= \theta_1 \\ p_2 &= \theta_2(1-\theta_1) \\ &\cdots \\ p_n &= \theta_n \prod_{i=1}^{n-1} (1-\theta_i) \\ &\cdots \end{aligned}$$

  Then,

$$P = \sum_{i=1}^{\infty} p_i \delta_{Y_i} \sim DP(\alpha).$$

Historial Notes - Statistics Side

- The Dirichlet process remained only as a theoretical object until 1990s.
- After MCMC appeared on the stage, the Bayesian nonparametric statistics was popularized.
- The Dirichlet process was at the center stage of the Bayesian nonparametrics.
- The main reason for this is the marginalization property of the Dirichlet process with which the MCMC computation of the posterior can be done easily.

Historial Notes - Probability Side

- After Ferguson's works, probabilitists (Kingman, Pitman and more) used and extended the theory for genetic problems.
- The theory developed by probabilitists was largely neglected by the statistics community nearly 30 years.
- James and Ishwaran in early 2000s noted that this theory could be used in Bayesian nonparametric statistics.
- James, Ishwaran, Walker, Prünster, Lijoi, Mena, Müller, Quintana, ... and many more (and perhaps Lee) ... developed statistical methodologies and theory for statistics.

Species Sampling

- Imagine that we land on a planet where "no one has gone before". As we explore the planet, we encounter new species unknown to us.
- We record the names of species we encounter. If the species is new, we name it by picking an element from $\mathcal{X}$.

- Suppose $(X_1, X_2, \ldots)$ is an infinite sequence of such records.
- X_i : the species of the i th individual sampled.
- $\tilde{X}_j$: the jth distinct species appeared
- $k = k_n$: the number of distinct species appeared in $(X_1, \ldots, X_n)$
- $n_j = n_{jn}$: the number of times the jth species $\tilde{X}_j$ appears in $(X_1, \ldots, X_n)$
- $\mathbf{n} = (n_{1n}, n_{2n}, \ldots)$ or $(n_{1n}, n_{2n}, \ldots, n_{kn})$

Species Sampling Sequence

We call an exchangeable sequence $(X_1, X_2, \ldots)$ the species sampling sequence if

$$\begin{aligned} X_1 &\sim \nu \\ X_{n+1} | X_1, \ldots, X_n &\sim \sum_{j=1}^{k} p_j(\mathbf{n}_n) \delta_{\tilde{X}_j} + p_{k+1}(\mathbf{n}_n)\nu, \end{aligned}$$

where ν is a diffuse probability measure on $\mathcal{X}$, i.e. $\nu(\{x\}) = 0 \;\; \forall x \in \mathcal{X}$.

Remark. The Polya urn sequence is an example of species sampling sequence.

Prediction Probability Function

- A sequence of functions $(p_j,\ j = 1, 2, \ldots) : \mathcal{C} \to \mathbb{R}$ in the definition of species sampling sequence is called the prediction probability function (PPF).
- The PPF (p_j) satisfies

$$\begin{aligned} p_j(\mathbf{n}) &\geq 0 \\ \sum_{j=1}^{k(\mathbf{n})+1} p_j(\mathbf{n}) &= 1, \text{ for all } \mathbf{n} \in \mathbb{N}^*. \end{aligned}$$

- For a species sampling sequence (X_n), the corresponding prediction probability functions is defined as

$$\begin{aligned} p_j(\mathbf{n}) &= \mathbb{P}(X_{n+1} = \tilde{X}_j | X_1, \ldots, X_n), \quad j = 1, \ldots, k_n, \\ p_{k_n+1}(\mathbf{n}) &= \mathbb{P}(X_{n+1} \notin \{X_1, \ldots, X_n\} | X_1, \ldots, X_n). \end{aligned}$$

Species Sampling Model

- A sequence of random variables (X_n) is a species sampling sequence if and only if $X_1, X_2, \ldots | P$ is random sample from P where

$$P = \sum_{i=1}^{\infty} P_i \delta_{\tilde{X}_i} + R\nu \tag{1}$$

  for some sequence of positive random variables (P_i) and R such that $1 - R = \sum_{i=1}^{\infty} P_i \leq 1$, $(\tilde{X}_i)$ is a random sample from ν, and (P_i) and $(\tilde{X}_i)$ are independent.
- We call the directing random probability measure P in equation (1) the *species sampling model (or prior)* of the species sampling sequence (X_i).

Exchangeable Random Partition on $[n]$

- $[n] = \{1, 2, \ldots, n\}$, $n \in \mathbb{N} = \{1, 2, \ldots\}$
- (exchangeable random partition) A random partition Π_n of $[n]$ is called exchangeable, if for any permutation σ on $[n]$,

  $$\Pi_n \overset{d}{=} \sigma(\Pi_n),$$

  i.e., for any partition $\{A_1, A_2, \ldots, A_k\}$ of $[n]$,

  $$P(\Pi_n = \{A_1, A_2, \ldots, A_k\}) = P(\sigma(\Pi_n) = \{A_1, A_2, \ldots, A_k\}).$$

  Here, $\sigma(\Pi_n)$ is the partition formed from partition Π_n by applying permutation σ on $[n]$.

Exchangeable Partition Probability Function (EPPF)

- Π_n is an exchangeable random partition of $[n]$ if and only if for any partition $\{A_1, A_2, \ldots, A_k\}$ of $[n]$,

$$P(\Pi_n = \{A_1, A_2, \ldots, A_k\}) = p(|A_1|, |A_2|, \ldots, |A_k|),$$

  for some function p on $\mathcal{C}_n$ symmetric in its arguments, where $\mathcal{C}_n$ is the set of all compositions of n.
- (EPPF) The function p is called an EPPF of Π_n.

Exchangeable Random Partition on $\mathbb{N}$

- A sequence of random partition $\Pi_\infty = (\Pi_n)_{n\geq 1}$ is called an exchangeable random partition on $\mathbb{N}$ if
  - Π_n is an exchangeable random partition on $[n]$ for all n;
  - $\Pi_m = \Pi_{m,n}$ a.s. for all $1 \leq m \leq n < \infty$, where $\Pi_{m,n}$ is the partition of $[m]$ obtained by restricting Π_n to $[m]$.

EPPF on $\mathbb{N}$

- For $\mathbf{n} = (n_1, n_2, \ldots, n_k)$,

$$\begin{aligned}\mathbf{n}^{j+} &= (n_1, \ldots, n_{j-1}, n_j + 1, n_{j+1}, \ldots, n_k),\ 1 \leq j \leq k,\\ \mathbf{n}^{(k+1)+} &= (n_1, n_2, \ldots, n_k, 1).\end{aligned}$$

- A function $p : \cup_{l=1}^{\infty} \mathbb{N}^l \to [0,1]$ is called an EPPF of $\Pi_\infty = (\Pi_n)$ if
  - $p(1) = 1$;
  - for all $\mathbf{n} \in \cup_{l=1}^{\infty} \mathbb{N}^l$,

$$p(\mathbf{n}) = \sum_{j=1}^{k+1} p(\mathbf{n}^{j+}).$$

  - $p_n = p|_{\mathcal{C}_n}$ is the EPPF of Π_n for all n, where $\mathcal{C}_n$ is the set of $(n_1, \ldots, n_k)$ with $\sum_i n_i = n$.

Characterizations of SSM

- The distribution of species sampling model

$$F = \sum_j P_j \delta_{U_j} + (1 - \sum_j P_j)\nu,$$

  is characterized by
  - ν and the distribution of (P_j); or
  - ν and the distribution of Π_∞; or
  - ν and the EPPF (p) of Π_∞; or
  - ν and the PPF (p_j) of Π_∞.
- The species sampling model is characterized as a species sampling sequence.

Example: Pitman-Yor Process

- For a pair of real numbers (a, b) and a diffuse probability measure with either $0 \leq a < 1$ and $b > -a$ or $a < 0$ and $b = -ma$ for some $m = 1, 2, \ldots$, define

$$\begin{aligned} U_j &\stackrel{ind}{\sim} Beta(1-a, b+ja), j = 1, 2, \ldots \\ \tilde{X}_1, \tilde{X}_2, \ldots &\stackrel{iid}{\sim} \nu \end{aligned}$$

  and $(U_j) \perp (\tilde{X}_j)$.

- Construct $P_1, P_2, \ldots$ from U_is by the stick breaking process

$$\begin{aligned} P_1 &= U_1 \\ P_j &= (1-U_j)\ldots(1-U_{j-1}) \cdot U_j \ , \quad j = 2, 3, \ldots . \end{aligned}$$

- The random probability measure

$$P = \sum_{j=1}^{\infty} P_j \delta_{\tilde{X}_j}$$

  is called a Pitman-Yor process or $P \sim PY(a, b, \nu)$.
- Note $PY(0, \theta, \nu) = DP(\theta \cdot \nu)$.

- (EPPF of Pitman-Yor)

$$p^{a,b}(n_1, n_2, \ldots, n_k) = \frac{(\theta + a)_{k-1\uparrow a} \prod_{i=1}^{k} (1-a)_{n_i - 1\uparrow 1}}{(\theta + 1)_{n-1\uparrow 1}},$$

where $(x)_{n\uparrow c} = x(x+c)(x+2c)\cdots(x+(n-1)c)$.

- (PPF of Pitman-Yor)

$$p_j^{a,b}(n_1, n_2, \ldots, n_k) = \begin{cases} \frac{n_j - a}{n+b}, & j = 1, 2, \ldots, k \\ \frac{b+ka}{n+b}, & j = k+1. \end{cases}$$

Consistency Issue

- The class of species sampling models is a huge class of nonparametric priors with more flexibilities than the Dirichlet process and potentially the same computational ease.
- But, the asymptotic properties with the species sampling models are not well understood.
- In the simplest possible nonparametric model, does the species sampling model pass the test of the posterior consistency?

True Distribution

We assume

$$X_1, X_2, \ldots \sim \text{ iid } P_0,$$

where

$$P_0 = \sum_j q_j \delta_{z_j} + \lambda\mu,$$

where $z_j \in \mathcal{X}, q_1 \geq q_2 \geq \cdots \geq 0, \lambda = 1 - \sum_j q_j \leq 1$ and μ is a diffuse probability measure.

Let $\mathcal{Z} = \{z_1, z_2, \ldots\}$.

Model

In this talk, we consider the following model:

$$\begin{aligned} X_1, \ldots, X_n | P &\sim P, \\ P &\sim \mathcal{P}, \end{aligned}$$

where $\mathcal{P}$ is a species sampling prior.

Consistency of PY Process

Theorem

When the prior is $PY(a, b, \nu)$, the posterior is weakly consistent at P_0 if and only if any of the followings holds

(i) $a = 0$, *that is, a Dirichlet process prior,*

(ii) *when* $a > 0$, P_0 *is discrete or* $\mu = \nu$,

(iii) $a < 0$ *and* P_0 *is a mixture of at most* $m = |b/a|$ *degenerated measures.*

Some Remarks

- If P_0 is discrete, all the Pitman-Yor process priors with $0 \leq a < 1$ entail the consistent posteriors.
- If P_0 is continuous, the Dirichlet process is the only prior among the Pitman-Yor process priors which renders posterior consistency.
- The second part of condition (ii) means that the diffuse probability measure ν should be proportional to the continuous part μ of the true probability measure P_0. Thus, in order to get the consistency we should know the continuous part of the true measure a priori, which is unlikely in practical situations.
- The same result has been obtained by James (2008) independently.

Mixture Models

The story is different in the mixture models. Consider the following normal mixture model

$$\begin{aligned} X_i|\theta_i, h &\sim \text{ ind } N(\theta_i, h^2), \quad && i = 1, \ldots, n, \\ \theta_i|P &\sim \text{ iid } P, && i = 1, \ldots, n, \\ P &\sim \mathcal{P}, \\ h^2 &\sim \mu, \end{aligned}$$

where P and h are independent a priori.

Under certain conditions, the posterior is weakly (and strongly) consistent.

More Assumptions for General Theorem

- (Smoothness condition for predictive probability function) As $n \to \infty$,

$$S_n = S_n(\mathbf{n}) = \max_{1 \le i \le k} \sum_{j=1}^{k} \left| p_j(\mathbf{n}) - p_j(\mathbf{n}^{i+}) \right| \to 0, \quad P_0^{\infty} - a.s.$$

- (Separability condition for $\mathcal{Z}$, the support of the discrete part of P_0) There exists $\epsilon > 0$ such that for all $i \neq j$

$$d(z_i, z_j) > \epsilon,$$

where d is the metric of $\mathcal{X}$.

General Theorem

Assume the separability condition and the smoothness condition. The posterior is weakly consistent at P_0 if and only if

$$\lim_{n\to\infty} \sum_{j=1}^{k} |p_j(\mathbf{n}) - n_j/n| I(\tilde{X}_j \in \mathcal{Z}) = 0, \quad P_0^\infty - a.s. \qquad (2)$$

and one of the followings holds

(i) $p_{k+1}(\mathbf{n}) \to 0$ as $n \to \infty$, $\quad P_0^\infty$ − a.s.

(ii) P_0 is a mixture of a discrete probability measure and the diffuse measure ν.

Remarks

- Condition (2) says essentially that the conditional distribution of X_{n+1} given $X_1, \ldots, X_n$ behaves like the empirical distribution of $X_1, \ldots, X_n$.
- The smoothness condition for the predictive probability function $p_j(\mathbf{n})$ ensures a small change in $\mathbf{n}$ does not change $p_j(\mathbf{n})$ much.
- The condition $p_{k+1}(\mathbf{n}) \to 0$ as $n \to \infty$ is natural in the following sense. Since $p_{k+1}(\mathbf{n})$ is the predictive probability that X_{n+1} is sampled from ν, we expect that $p_{k+1}(\mathbf{n}) \to 0$ as $n \to \infty$, if the posterior consistency holds.
- Condition (ii) is satisfied by all discrete probability measures. Thus, all species sampling priors satisfying (2) are weakly consistent at every discrete probability measure.

Consistency Results for other Subclasses

- The N-IG process prior (Lijoi, Mena and Prünster, 2005) is consistent at all the discrete distributions, but inconsistent at all the continuous distributions except ν.
- The the prior with Poisson-Kingman partition, $PK(\rho_{a,b,c})$, is consistent at all discrete distributions, but inconsistent at all continuous distributions except $a = 0$ (DP case), where $\rho_{a,b,c}(x) = cx^{-a-1}e^{-bx}$ with $0 \leq a < 1, b \geq 0$ and $c > 0$.
- Under certain conditions, the Gibbs type prior is also consistent at all discrete distributions but inconsistent at all continuous except DP.