Posterior Consistency of Species Sampling Priors

Jaeyong Lee

`leejyc@gmail.com`

Seoul National University

jointly with

Gunho Jang and Sangyeol Lee

Some Notation

- Suppose $(X_1, X_2, \ldots)$ is a random sample with value in a complete separable metric space $\mathcal{X}$, population of various species.
- X_i : the species of the i th individual sampled.
- M_j : the first index of the jth species appeared
- $\tilde{X}_j = X_{M_j}$: the jth species appeared
- $k = k_n$: the number of different 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 $(n_{1n}, n_{2n}, \ldots, n_{kn})$
- $\mathbf{n}$ is an element of

$$\mathbb{N}^* = \bigcup_{k=1}^{\infty} \mathbb{N}^k,$$

where $\mathbb{N}$ is the set of positive integers.

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. $\nu(\{x\}) = 0 \ \forall x \in \mathcal{X}$.

Example : Polya Urn Sequence

Suppose $P \sim DP(\theta\nu)$, where $\theta > 0$ and ν is a probability measure and $X_1, X_2, \ldots | P \sim P$. Then, marginally $X_1, X_2, \ldots$ is a Polya urn sequence which satisfies

$$\begin{aligned} X_1 &\sim \nu \\ X_{n+1} | X_1, \ldots, X_n &\sim \sum_{j=1}^{k} \frac{n_j}{n+\theta} \delta_{\tilde{X}_j} + \frac{\theta}{n+\theta} \nu. \end{aligned}$$

Thus, the Polya urn sequence is a species sampling sequence.

Prediction Probability Function

- A sequence of functions $(p_j,\ j=1,2,\ldots):\mathbb{N}^*\to\mathbb{R}$ is called a sequence of prediction probability functions if

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

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

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

Example : Pólya Urn Sequence

$$p_j(n_1, \ldots, n_k) = \frac{n_j}{n+\theta} I(1 \le j \le k) + \frac{\theta}{n+\theta} I(j = k+1),$$

where $n = \sum_{i=1}^{k} n_i$.

Species Sampling Prior

- 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 prior (or process) of the species sampling sequence (X_i).

Exchangeable Partition Probability Function (EPPF)

- Let $[n] = \{1, 2, \ldots, n\}$.
- An exchangeable sequence (X_n) defines a random partition $\Pi_n = \{A_1, \ldots, A_k\}$ of $[n]$, where

$$A_i = \{j \in [n] : X_j = \tilde{X}_i\}.$$

Define a function $p : \mathbb{N}^* \longrightarrow [0, 1]$

$$p(\#A_1, \ldots, \#A_k) = \mathbb{P}(\Pi_n = \{A_1, \ldots, A_k\}).$$

- The function p is called the exchangeable partition probability function (EPPF) derived from the exchangeable sequence (X_n).

- An EPPF p derived from an exchangeable sequence (X_n) satisfies

$$
\begin{aligned}
p(1) &= 1 \\
p(\mathbf{n}) &= \sum_{j=1}^{k(\mathbf{n})+1} p(\mathbf{n}^{j+}), \quad \forall \mathbf{n} \in \mathbb{N}^*,
\end{aligned}
$$

  where $\mathbf{n}^{j+}$ is the same as $\mathbf{n}$ except that jth element is increased by 1.
- Conversely, every symmetric $p : \mathbb{N}^* \longrightarrow [0, 1]$ satisfying (2) is an EPPF of some exchangeable sequence.
- An EPPF p and a diffuse probability measure ν uniquely defines the distribution of a species sampling sequence.

Dirichlet Process (Sethuraman's Representation)

- Define

$$\begin{aligned} W_1, W_2, \ldots &\sim \text{i.i.d} \quad Beta(1, \theta) \\ \tilde{X}_1, \tilde{X}_2, \ldots &\sim \text{i.i.d} \quad \nu \end{aligned}$$

  and $(W_j) \perp (\tilde{X}_j)$.
- Construct $P_1, P_2, \ldots$ from W_is by the stick breaking process

$$\begin{aligned} P_1 &= W_1 \\ P_j &= (1 - W_1) \ldots (1 - W_{j-1}) \cdot W_j \ , \quad j = 2, 3, \ldots . \end{aligned}$$

Let

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

Then, $P \sim DP(\theta, \nu)$.

Pitman-Yor Process

- Define

$$\begin{aligned} W_j &\sim \text{i.i.d} \quad Beta(1-a, b+ja), j = 1, 2, \dots \\ \tilde{X}_1, \tilde{X}_2, \dots &\sim \text{i.i.d} \quad \nu \end{aligned}$$

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

- Construct $P_1, P_2, \dots$ from W_is by the stick breaking process

$$\begin{aligned} P_1 &= W_1 \\ P_j &= (1 - W_1) \dots (1 - W_{j-1}) \cdot W_j \ , \quad j = 2, 3, \dots. \end{aligned}$$

Let

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

Then, $P \sim PY(a, b, \nu)$, where ν is a diffuse probability measure and either $0 \leq a < 1$ and $b > -a$ or $a < 0$ and $b = -ma$ for some $m = 1, 2, \dots$
Note $PY(0, \theta, \nu) = DP(\theta \cdot \nu)$.

Posterior of PY Process

Suppose

$$\begin{aligned} X_1, \dots, X_n | P &\sim P, \\ P &\sim PY(a, b, \nu). \end{aligned}$$

Then,

$$P | X_1, \dots, X_n = \sum_{j=1}^{k} \tilde{P}_j \delta_{\tilde{X}_j} + \tilde{R}_k P_k,$$

where $(\tilde{P}_1, \dots, \tilde{P}_k, \tilde{R}_k)$ are independent of F_k and

$$\begin{aligned} (\tilde{P}_1, \dots, \tilde{P}_k, \tilde{R}_k) &\sim Dir(n_1 - a, \dots, n_k - a, b + ka) \\ P_k &\sim PY(a, b + ka, \nu). \end{aligned}$$

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 1. 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.

Mixture Models

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

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

where P and h are independent a priori.

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

Lemma 1

Under some general conditions, the followings hold.

(a) The posterior is weakly consistent at P_0 if and only if the followings hold for every P_0-continuity set U of $\mathcal{X}$

(i) $\lim_{n\to\infty} \mathbb{E}(P(U)|X_1,\ldots,X_n) = P_0(U), \quad P_0^\infty - a.s.,$

(ii) $\lim_{n\to\infty} \mathbb{V}ar(P(U)|X_1,\ldots,X_n) = 0, \quad P_0^\infty - a.s.$

where P_0^∞ is the infinite product of the true probability measure P_0 representing the probability measure of $(X_1, X_2, \ldots)$.

(b) If the posterior is weakly consistent at P_0, then for all open set O and closed set C of $\mathcal{X}$

(i) $\liminf_{n\to\infty} \mathbb{E}(P(O)|X_1,\ldots,X_n) \geq P_0(O), \quad P_0^\infty - a.s.,$

(ii) $\limsup_{n\to\infty} \mathbb{E}(P(C)|X_1,\ldots,X_n) \leq P_0(C), \quad P_0^\infty - a.s.$

Lemma 2

Suppose $X_1, X_2, \ldots, X_n$ are sampled from P_0. Let $\tilde{X}_1, \ldots, \tilde{X}_{k_n}$ be the distinct values among $X_1, \ldots, X_n$, $k_n^* = \sum_{j=1}^{k_n} I(\tilde{X}_j \notin \mathcal{Z})$ and

$$G_{k_n} = \frac{1}{k_n} \sum_{j=1}^{k_n} \delta_{\tilde{X}_j}.$$

Then,

(i) $\dfrac{k_n}{n} \to \lambda$ and $\dfrac{k_n^*}{n} \to \lambda$, $P_0^\infty - a.s.$,

(ii) $G_{k_n} \to \mu$, $P_0^\infty - a.s.$ if $\lambda > 0$.

Sketch of Proof

$$
\begin{aligned}
&\mathbb{E}(P(B)|X_1,\ldots,X_n)\\
&=\mathbb{E}[\mathbb{P}(X_{n+1}\in B|X_1,\ldots,X_n,\tilde{P}_1,\ldots,\tilde{P}_{k_n},\tilde{R}_{k_n})|X_1,\ldots,X_n]\\
&=\mathbb{E}[\sum_{j=1}^{k_n}\tilde{P}_j I(\tilde{X}_j\in B)+\tilde{R}_{k_n}\nu(B)|X_1,\ldots,X_n]\\
&=\sum_{j=1}^{k_n}\frac{n_{jn}-a}{b+n}I(\tilde{X}_j\in B)+\frac{b+ak_n}{b+n}\nu(B)\\
&=\frac{n}{b+n}F_n(B)-\frac{ak_n}{b+n}G_{k_n}(B)+\frac{b+ak_n}{b+n}\nu(B)\\
&\to P_0(B)-a\lambda\mu(B)+a\lambda\nu(B)\\
&=P_0(B)-a\lambda(\mu(B)-\nu(B)),
\end{aligned}
$$

where $F_n=\frac{1}{n}\sum_{j=1}^{n}\delta_{X_j}$ and G_{k_n} is defined earlier.

In summary,

$$\mathbb{E}(P(B)|X_1,\ldots,X_n) \to P_0(B) - \alpha\cdot\lambda\left(\mu(B)-\nu(B)\right) \quad a.s.$$

Thus,

$$\mathbb{E}(P(B)|X_1,\ldots,X_n) \to P_0(B) \quad a.s.$$

if and only if

$$\begin{aligned} \alpha &= 0 \\ \lambda &= 0 \\ \text{or } \mu(B) &= \nu(B). \end{aligned}$$

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

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

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.

Stronger Sufficient Condition

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

$$C_n = C_n(\mathbf{n}) = \sum_{j=1}^{k} \left| p_j(\mathbf{n}) - \frac{n_j}{n} \right| \to 0, \quad P_0^\infty - a.s. \quad \text{as} \quad n \to \infty. \qquad (3)$$

Remarks

- Condition (3) 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 smoothness condition is satisfied all the examples considered here.
- 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.

Pitman-Yor Process

Suppose $P \sim PY(a, b, \nu)$ with $0 \leq a < 1$ and $b > 0$. Since $p_j(\mathbf{n}) = \frac{n_j - a}{b+n}$ for $j = 1, \ldots, k$, and $p_{k+1}(\mathbf{n}) = (b + ka)/(b + n)$,

$$\sum_{j:\tilde{X}_j \in \mathcal{Z}} \left| p_j(\mathbf{n}) - \frac{n_j}{n} \right| = \sum_{j:\tilde{X}_j \in \mathcal{Z}} \left| \frac{n_j - a}{b + n} - \frac{n_j}{n} \right| = \frac{an(k - k^*) + b(n - k^*)}{n(b + n)} \to 0, \quad a.s.$$

Note $p_{k+1}(\mathbf{n}) \to a\lambda, P_0^\infty - a.s.$ Condition (i) is equivalent to $a = 0$ or $\lambda = 0$. Thus, the general theorem agrees with the theorem for the Pitman-Yor process.

Normalized Inverse-Gaussian (N-IG) Process

Lijoi, Mena and Prünster (2005) defined the N-IG process P by specifying the distribution of $(P(B_1), \ldots, P(B_k))$ for a partition $B_1, \ldots, B_k$ of $\mathcal{X}$ as the distribution of

$$(V_1, \ldots, V_k)/V,$$

where $V = V_1 + \cdots + V_k$ and

$$V_i \overset{\text{ind}}{\sim} IG(\theta\nu(B_i), 1), \quad i = 1, \ldots, k.$$

Here $IG(a, b)$ denotes the inverse-Gaussian distribution with parameter $a \geq 0$ and $b > 0$ whose density is

$$a(2\pi x^3)^{-1/2} \exp(-(a^2/x + b^2x)/2 + ab), \quad x > 0.$$

One can show the N-IG process is the species sampling prior with predictive distribution

$$\mathbb{P}(X_{n+1} \in B | X_1, \ldots, X_n) = w_{1,n} \sum_{j=1}^{k} (n_j - 1/2) \delta_{\tilde{X}_j}(B) + w_{0,n} \nu(B),$$

where

$$\begin{aligned} w_{0,n} &= \frac{\theta \int_1^\infty (1 - y^{-2})^n y^k e^{-\theta y} dy}{2n \int_1^\infty (1 - y^{-2})^{n-1} y^{k-1} e^{-\theta y} dy} \\ w_{1,n} &= \frac{\int_1^\infty (1 - y^{-2})^n y^{k-1} e^{-\theta y} dy}{n \int_1^\infty (1 - y^{-2})^{n-1} y^{k-1} e^{-\theta y} dy}. \end{aligned}$$

Consistency of N-IG Process

- The N-IG process prior is consistent at all the discrete distributions, but inconsistent at all the continuous distributions except ν.

Poisson-Kingman Partition $PK(\rho)$

- Let Λ with density ρ is the intensity measure of the Poisson process with

$$\int_0^1 x d\Lambda(x) < \infty \text{ and } \int_1^\infty d\Lambda(x) < \infty.$$

- Let $J_1, J_2, \ldots$ be the jump sizes of the Poisson point process with the intensity Λ.
- The normalized J_is, J_i/T, play the role of P_is in the species sampling prior.

- The EPPF of $PK(\rho)$ is given by

$$p(n_1, \ldots, n_k) = \frac{(-1)^{n-k}}{\Gamma(n)} \int_0^\infty u^{n-1} e^{-\psi(u)} \prod_{j=1}^k \psi_{n_j}(u) du$$

  where $\psi(u) = \int_0^\infty (1 - e^{-ux})\rho(x)dx$ and
  $\psi_m(u) = \frac{d^m \psi}{du^m}(u) = (-1)^{m-1} \int_0^\infty x^m e^{-ux} \rho(x) dx$ for $m = 1, 2, \ldots$.
- The predicted probability function $p_j(\mathbf{n})$ of $PK(\rho)$ is, for $j = 1, \ldots, k$,

$$p_j(\mathbf{n}) = -\frac{1}{n} \frac{\int_0^\infty u \frac{\psi_{n_j+1}(u)}{\psi_{n_j}(u)} u^{n-1} e^{-\psi(u)} \prod_{i=1}^k \psi_{n_i}(u) du}{\int_0^\infty u^{n-1} e^{-\psi(u)} \prod_{i=1}^k \psi_{n_i}(u) du}.$$

Consistency of $PK(\rho)$

- In this example, we consider

$$\rho_{a,b,c}(x) = cx^{-a-1}e^{-bx},$$

  where $0 \leq a < 1, b \geq 0$ and $c > 0$.
- $DP(\theta\nu)$ is equivalent to $PK(\rho_{0,1,\theta}, \nu)$.
- $PK(\rho_{a,b,c})$ is consistent at all discrete distributions, but inconsistent at all continuous distributions except $a = 0$.

Gibbs Partition

- An EPPF p is said to be of Gibbs form if

$$p(n_1, \ldots, n_k) = V_{n,k} \prod_{j=1}^{k} W_{n_j},$$

  for some nonnegative weights $W = (W_j)$ and $V = (V_{n,k})$.

- Assume $W_1 = V_{1,1} = 1$. Then, every Gibbs partition is represented by W_js and $V_{n,k}$s satisfying

$$W_j = \begin{cases} 1 & \text{if } j = 1, \\ \prod_{i=0}^{j-2}(b - a + bi) & j = 2, 3, \ldots \end{cases} \quad \text{and} \quad V_{n,k} = (bn - ak)V_{n+1,k} + V_{n+1,k+1}$$

  for some $b > 0$ and $a < b$.

- The predictive probability functions are, for $j = 1, \ldots, k$,

$$p_j(\mathbf{n}) = \frac{p(\mathbf{n}^{j+})}{p(\mathbf{n})} = \frac{V_{n+1,k}}{V_{n,k}} \frac{W_{n_j+1}}{W_{n_j}} = \frac{nbV_{n+1,k}}{V_{n,k}} \frac{n_j - a/b}{n}.$$

Consistency of Gibbs Form

- The species sampling prior generated by a Gibbs partition is consistent at all discrete probability measures if

$$nbV_{n+1,k}/V_{n,k} \to 1.$$

- Under the assumption $nbV_{n+1,k}/V_{n,k} \to 1$, the posterior is consistent at all continuous distributions if and only if $a = 0$ or the Dirichlet process.