Poisson Latent Feature Calculus for Generalized Indian Buffet Processes

Lancelot F. James

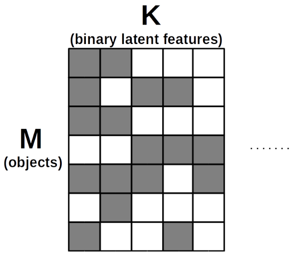
(paper from arXiv [math.ST] , Dec'14)

Discussion by: Piyush Rai

January 23, 2015

Overview of the paper

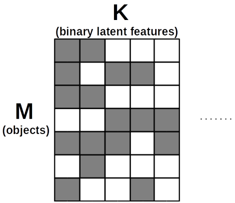
- Proposes generalized notions of the Indian Buffet Process (IBP)



- Allows principled extensions of the standard IBP to more general settings involving **non-binary latent features**
- Provides a unified and (perhaps) simpler/cleaner treatment than some other recent attempts to such problems
- Also allows extensions to **multivariate** latent features (each latent feature can be a vector - multinomial or a general multivariate vector)
- Based on ideas from Poisson Partition Calculus (proposed earlier by the same author)

The IBP

- A nonparametric Bayesian prior on random binary matrices $Z = [Z_1; \ldots; Z_M]$ with M rows and unbounded number of columns



- Prior denoted as $IBP(\theta)$: $\theta > 0$, with latent features $\tilde{\omega}_k, k = 1, 2, \ldots$
- Entry $(i, k) = 1$ if object i has latent feature k, zero otherwise
- Generative model: Customers (objects) selecting dishes (latent features)
  - Customer 1 selects Poisson(θ) dishes
  - Customer i selects each already selected dish k with prob. m_k/i (where m_k is # of previous customers who chose dish k), and Poisson(θ/i) new dishes

Generalization of the IBP

Standard $IBP(\theta)$: $Z_1, \ldots, Z_M \mid \mu$

- $\mu = \sum_{k=1}^{\infty} p_k \delta_{\tilde{\omega}_k}$
- Atoms $\tilde{\omega}_k$ drawn from a base measure B_0
- p_k: points from a Poisson random measure with a restricted mean intensity $\rho(s) = \theta s^{-1} \mathbb{I}_{\{0<s<1\}}$
- $Z_i = \sum_{k=1}^{\infty} b_{i,k} \delta_{\tilde{\omega}_k}$
- $b_{i,k} \sim Bernoulli(p_k)$

Generalized $IBP(\theta)$: $Z_1, \ldots, Z_M \mid \mu$

- $\mu = \sum_{k=1}^{\infty} \tau_k \delta_{\tilde{\omega}_k}$
- Atoms $\tilde{\omega}_k$ drawn from a base measure B_0
- τ_k: points from a Poisson random measure with general/unrestricted Lévy density $\rho(s \mid \omega)$
- $Z_i = \sum_{k=1}^{\infty} A_{i,k} \delta_{\tilde{\omega}_k}$
- $A_{i,k} \sim G_A(da \mid \tau_k)$

In the generalized IBP proposed here, $A_{i,k}$ isn't necessarily binary - can be a more general r.v. (or even vector valued)

Generalization of the IBP

Standard $IBP(\theta)$: $Z_1, \ldots, Z_M \mid \mu$

- $\mu = \sum_{k=1}^{\infty} p_k \delta_{\tilde{\omega}_k}$
- Atoms $\tilde{\omega}_k$ drawn from a base measure B_0
- p_k: points from a Poisson random measure with a restricted mean intensity $\rho(s) = \theta s^{-1} \mathbb{I}_{\{0<s<1\}}$
- $Z_i = \sum_{k=1}^{\infty} b_{i,k} \delta_{\tilde{\omega}_k}$
- $b_{i,k} \sim Bernoulli(p_k)$

Generalized $IBP(\theta)$: $Z_1, \ldots, Z_M \mid \mu$

- $\mu = \sum_{k=1}^{\infty} \tau_k \delta_{\tilde{\omega}_k}$
- Atoms $\tilde{\omega}_k$ drawn from a base measure B_0
- τ_k: points from a Poisson random measure with general/unrestricted Lévy density $\rho(s \mid \omega)$
- $Z_i = \sum_{k=1}^{\infty} A_{i,k} \delta_{\tilde{\omega}_k}$
- $A_{i,k} \sim G_A(da \mid \tau_k)$

In the generalized IBP proposed here, $A_{i,k}$ isn't necessarily binary - can be a more general r.v. (or even vector valued)

Can also think of $A_{i,k} = b_{i,k} A'_{i,k}$ where $b_{i,k}$ is binary and $A'_{i,k}$ is a general r.v. (akin to a "spike-and-slab" construction)

This Paper

Proposes IBP generalizations of the following form

- $\mu = \sum_{k=1}^{\infty} \tau_k \delta_{\tilde{\omega}_k}$
- τ_k: points from a Poisson random measure with Lévy density $\rho(s \mid \omega)$
- $Z_i = \sum_{k=1}^{\infty} A_{i,k} \delta_{\tilde{\omega}_k}$
- $A_{i,k} \sim G_A(da \mid \tau_k)$

IBP generalizations that work with any G_A and any ρ, so long as

- G_A admits a positive mass at 0 and ρ is a finite measure

The constructions rely on Poisson Partition Calculus (PPC)

Conjugacy not a requirement

No need for combinatorial arguments or long proofs

Notations

Recall $Z_1, \ldots, Z_M \mid \mu$ where $\mu = \sum_{k=1}^{\infty} \tau_k \delta_{\tilde{\omega}_k}$

Can express $\mu(d\omega) = \int_0^{\infty} sN(ds, d\omega)$

$N = \sum_{k=1}^{\infty} \delta_{\tau_k, \tilde{\omega}_k}$ is a Poisson random measure (PRM) with mean intensity

$$\mathbb{E}[N(ds, d\omega)] = \rho(s \mid \omega) ds B_0(d\omega) := \nu(ds, d\omega)$$

Some notations:

- N as PRM(ρB_0) or PRM(ν), or $\mathcal{P}(\nu)$
- μ is a completely random measure CRM(ρB_0) or CRM(ν)
- $Z_1, \ldots, Z_M \mid \mu$ as iid IBP$(G_A|\mu)$
- The marginal distribution of Z_i as IBP$(A|\rho B_0)$ or IBP(A, ν)

The General Scheme

- Generally apply the Poisson Calculus
- Describe $N|Z_1, \ldots, Z_M$
- Consequently describe $\mu|Z_1, \ldots, Z_M$
- Describe the marginal structure $\mathbb{P}(Z_1, \ldots, Z_M)$ via integration
- Provide descriptions for the marginal

$$Z = \sum_{k=1}^{\xi(\varphi)} X_k \delta_{\tilde{\omega}_k}$$

  and introduce variables (H_i, X_i) leading to tractable sampling for many ρ
- Describe $Z_{M+1}|Z_1, \ldots, Z_M$

Two Ingredients from Poisson Partition Calculus

- Laplace functional exponential change: updating a Poisson random measure

$$e^{-N(f)}\mathcal{P}(dN|\nu) = \mathcal{P}(dN|\nu_f)e^{-\Psi(f)}$$

where $\nu_f(ds, d\omega) = e^{-f(s,\omega)}\rho(s|\omega)dsB_0(d\omega)$ and $e^{-\Psi(f)} = \mathbb{E}_\nu[e^{-N(f)}]$ and

$$\Psi(f) = \int_\Omega \int_0^\infty (1 - e^{-f(s,\omega)})\rho(s|\omega)dsB_0(d\omega)$$

- Using ν_f, apply the moment measure calculation

$$\left[\prod_{i=1}^{L} N(dW_i)\right]\mathcal{P}(dN|\nu_f) = \mathcal{P}(dN|\nu_f, \mathbf{s}, \omega)\prod_{\ell=1}^{K} e^{-f(s_\ell,\omega_\ell)}\nu(ds_\ell, d\omega_\ell)$$

where $\mathcal{P}(dN|\nu_f, \mathbf{s}, \omega)$ corresponds to the law of the random measure

$$\tilde{N} + \sum_{\ell=1}^{K} \delta_{s_\ell,\omega_\ell} \qquad \text{where } \tilde{N} \text{ is also PRM}(\nu_f)$$

Joint Distribution

- Joint distribution of **A** (or, equivalently, of $Z = [Z_1; \dots ; Z_M]$)

$$\prod_{i=1}^{M}\prod_{j=1}^{\infty}[G_A(da_{i,j}|\tau_j)]^{\mathbb{I}_{\{a_{i,j}\neq 0\}}}[1-\mathbb{P}(A\neq 0|\tau_j)]^{1-\mathbb{I}_{\{a_{i,j}\neq 0\}}}$$

and setting $\pi_A(s) = \mathbb{P}(A \neq 0|s)$, the above becomes

$$e^{-\sum_{j=1}^{\infty}[-M\log(1-\pi_A(\tau_j))]}\prod_{i=1}^{M}\prod_{j=1}^{\infty}\left[\frac{G_A(da_{i,j}|\tau_j)}{1-\pi_A(\tau_j)}\right]^{\mathbb{I}_{\{a_{i,j}\neq 0\}}}$$

- After some further simplifications, the joint distribution of $((Z_1, \dots , Z_M), (J_1, \dots , J_K), N)$ where $(J_1 = s_1, \dots , J_K = s_K)$ are the unique jumps, becomes

$$\mathcal{P}(dN|\nu_{f_M}, \mathbf{s}, \omega)e^{-\Psi(f_M)}\prod_{\ell=1}^{K}[1-\pi_A(s_\ell)]^M\rho(s_\ell|\omega_\ell)B_0(d\omega_\ell)\prod_{i=1}^{M}h_{i,\ell}(s_\ell)$$

where $h_{i,\ell}(s) = [\frac{G_A(da_{i,\ell}|s)}{1-\pi_A(s)}]^{\mathbb{I}_{\{a_{i,\ell}\neq 0\}}}$ and
$\Psi(f) = \int_\Omega \int_0^\infty (1 - [1 - \pi_A(s)]^M)\rho(s \mid \omega)dsB_0(d\omega)$

Posterior and Marginal Distribution

- Integrating out N leads to the joint distribution of $((Z_i),(J_\ell))$

$$e^{-\Psi(f_M)}\prod_{\ell=1}^{K}[1-\pi_A(s_\ell)]^M\rho(s_\ell|\omega_\ell)B_0(d\omega_\ell)\prod_{i=1}^{M}h_{i,\ell}(s_\ell)$$

- $(J_\ell)|(Z_i)$ are conditionally independent with density

$$\mathbb{P}_{\ell,M}(J_\ell\in ds)\propto[1-\pi_A(s)]^M\rho(s|\omega_\ell)\prod_{i=1}^{M}h_{i,\ell}(s)ds$$

- .. and the marginal of $(Z_1,\ldots,Z_M)$ is

$$\mathbb{P}(Z_1,\ldots,Z_M)=e^{-\Psi(f_M)}\prod_{\ell=1}^{K}\left[\int_0^\infty[1-\pi_A(s)]^M\rho(s|\omega_\ell)\prod_{i=1}^{M}h_{i,\ell}(s)ds\right]B_0(d\omega_\ell)$$

Main Result

Notation: $\nu_{f_M}(ds, d\omega) := \rho_M(s|\omega)B_0(d\omega)ds$ where $\rho_M(s|\omega) = [1 - \pi_A(s)]^M \rho(s|\omega)$

Theorem: Suppose that $Z_1, \ldots, Z_M|\mu$ are iid IBP$(G_A|\mu)$, μ is CRM(ρB_0), then

- The posterior distribution of $N|Z_1, \ldots, Z_M$ is equivalent to the distribution of
$$N_M + \sum_{\ell=1}^{K} \delta_{J_\ell, \omega_\ell}$$
where N_M is PRM$(\rho_M B_0)$, and the distribution of (J_ℓ) is (cf, previous slide)
- The posterior distribution of $\mu|Z_1, \ldots, Z_M$ is equivalent to the distribution of
$$\mu_M + \sum_{\ell=1}^{K} J_\ell \delta_{\omega_\ell} \qquad \text{where } \mu_M \text{ is CRM}(\rho_M B_0)$$
- The marginal distribution of $(Z_1, \ldots, Z_M)$ is (cf, previous slide)

Proposition: Suppose $Z_1, \ldots, Z_M, Z_{M+1}|\mu$ are iid IBP$(G_A|\mu)$ then
$$Z_{M+1}|Z_1, \ldots, Z_M \sim \tilde{Z}_{M+1} + \sum_{\ell=1}^{K} A_\ell \delta_{\omega_\ell}$$
where $\tilde{Z}_{M+1}$ is IBP$(A, \rho_M B_0)$ and each $A_\ell|J_\ell = s$ has distribution $G_A(da|s)$

Special Cases: Bernoulli, Poisson, Negative-Binomial

- Bernoulli: $G_A(.|s) = Bernoulli(s)$

$$\mathbb{P}(A = 0|s) = 1 - \pi_A(s) = 1 - s$$

- Poisson: $G_A(.|s) = Poisson(bs)$

$$\mathbb{P}(A = 0|s) = 1 - \pi_A(s) = e^{-bs}$$

- Negative Binomial: $G_A(.|s) = NB(r, s)$ where r is the overdispersion param.

$$\mathbb{P}(A = 0|s) = 1 - \pi_A(s) = (1 - s)^r$$

- Setting $c_{\ell,M} = \sum_{i=1}^{M} a_{i,\ell}$, we have $\prod_{i=1}^{M} h_{i\ell}(s)$ given by

$$\left(\frac{s}{1-s}\right)^{c_{\ell,M}}, \frac{b^{c_{\ell,M}} s^{c_{\ell,M}}}{\prod_{i=1}^{M} a_{i,\ell}!}, \quad \text{and} \quad s^{c_{\ell,M}} \prod_{i=1}^{M} \binom{a_{i,\ell} + r - 1}{a_{i,\ell}}$$

for Bernoulli, Poisson, and NB cases, respectively (for the homogeneous case $\rho(s)$; the case $\rho(s|\omega)$ can be likewise worked out)

Describing posterior, marginal, and other quantities

- Specify $G_A(s)$
- Identify $\pi_A(s) = \mathbb{P}(A \neq 0|s)$
- For each $M = 0, 1, 2, \ldots$, define $\rho_M(s) = [1 - \pi_A(s)]^M \rho(s|\omega)$
- Specify $N_M \sim PRM(\rho_M B_0)$, $\mu_M \sim CRM(\rho_M B_0)$ and $\tilde{Z}_{M+1} \sim IBP(A, \rho_M B_0)$
- Identify $h_{i,\ell}(s)$ and if possible try to simplify $\prod_{i=1}^{M} h_{i,\ell}(s)$ where

$$h_{i,\ell}(s) = \left[\frac{G_A(da_{i,\ell}|s_\ell)}{1 - \pi_A(s_\ell)}\right]^{\mathbb{I}_{\{a_{i,\ell} \neq 0\}}}$$

- Use this to get the distribution of (J_ℓ) and the marginal of $(Z_1, \ldots, Z_M)$

Steps for sequential generalized IBP construction

- For each $M = 0, 1, 2, \ldots$ define $\rho_M(s) = [1 - \pi_A(s)]^M \rho(s)$
- Calculate and check
$$\varphi_{M+1}(\rho) = \int_0^\infty \pi_A(s)\rho_M(s)ds < \infty$$
- Then if $\tilde{Z}_{M+1}$ is IBP$(A, \rho_M B_0)$,
$$\tilde{Z}_{M+1} = \sum_{k=1}^{\xi(\varphi)} X_k \delta_{\tilde{\omega}_k}$$
where $\varphi := \varphi_{M+1}(\rho)$ and $\xi(\varphi)$ is a Poisson(φ) r.v., and (X_k) are taken from an IBP(A, ρ_M) process $((X_i, H_i))$
- $X_i|H_i = s$ has distribution $\frac{\mathbb{I}_{\{a \neq 0\}} G_A(da|s)}{\pi_A(s)}$ and H_i has marginal density $\frac{\pi_A(s)\rho_M(s)}{\varphi_{M+1}(\rho)}$
- X_i has marginal distribution $\mathbb{P}(X_i \in da) = \frac{\mathbb{I}_{\{a \neq 0\}} \int_0^\infty G_A(da|s)\rho_M(s)ds}{\varphi_{M+1}(\rho)}$ and the distribution of $H_i|X_i = a$ is given by $\mathbb{P}(H_i \in ds|X_i = a) = \frac{\mathbb{I}_{\{a \neq 0\}} G_A(da|s)\rho_M(s)ds}{\int_0^\infty G_A(da|v)\rho_M(v)dv}$

Generalized IBP: The Sequential Generative Process

Sequential generative process for $\mathrm{IBP}(A, \rho B_0)$:

- Customer 1 selects dishes and scores them according to $\mathrm{IBP}(A, \rho B_0)$ as
  - Draw $\mathrm{Poisson}(\varphi_1(\rho)) = J$ number of variables
  - Draw $((\omega_1, H_1), \ldots, (\omega_J, H_J))$ iid from B_0 and distribution for H_i with $M = 0$
  - Draw $X_i|H_i$, for $i = 1, \ldots, J$, or draw X_i directly (if so possible)
- After M customers have chosen K dishes $\omega_1, \ldots, \omega_K$, customer $M + 1$:
  - Selects/skips each existing dish ω_ℓ and scores (if selected) it according to A_ℓ, where $A_\ell|J_\ell$ has distribution $G_A(da|J_\ell)$. The probability that $A_\ell|J_\ell$ takes value zero is $1 - \pi_A(J_\ell)$
  - Also chooses and scores new dishes according to $\mathrm{IBP}(A, \rho_M B_0)$ process $\tilde{Z}_{M+1}$ with $\rho_M(s) = [1 - \pi_A(s)]^M \rho(s)$

Note: Specific cases when A is Poisson and Negative-Binomial are shown in the paper (sec 4.1 and 4.2)

Multivariate Generalizations

Multivariate CRM is also considered. Generalizes the IBP for vector-valued $A_{i,k}$

- Consider $\mu_0 = (\mu_j, j = 1, \ldots, q)$ with $\mu_j = \sum_{k=1}^{\infty} p_{j,k}\delta_{\tilde{\omega}_k}$ (note: all μ_j's assumed to the same base measure)
- $\mu_{\cdot} = \sum_{j=1}^{q} \mu_j = \sum_{k=1}^{\infty} p_{\cdot,k}\delta_{\tilde{\omega}_k}$ for $p_{\cdot,k} = \sum_{j=1}^{q} p_{j,k}$
- The multivariate CRM μ_0 can be constructed from a Poisson random measure on $(\mathbb{R}_+^q, \Omega)$

$$N := \sum_{k=1}^{\infty} \delta_{\mathbf{p}_{q,k},\tilde{\omega}_k} \qquad \text{where } \mathbf{p}_{q,k} = (p_{1,k}, \ldots, p_{q,k})$$

A simple case (sec 5.1 in the paper): Multinomial extension of the IBP. The joint probability of the $M \times K$ IBP "matrix" (each entry a multinomial vector)

$$\left[\prod_{\ell=1}^{K}\prod_{j=1}^{q}\left(\frac{p_{j,\ell}}{1-p_{\cdot,\ell}}\right)^{c_{j,\ell,M}}\right] e^{-M\sum_{j=1}^{\infty}[-\log(1-p_{\cdot,j})]}$$

.. akin to an "IBP with a condiment". All the theorems generalize for this case too. Other multivariate generalizations also possible (sec 5.2 in the paper)

Multivariate Generalizations

A more general case (sec 5.2):

Let $A_0 := (A_1, \ldots, A_v) \in \mathbb{R}^v$ with distribution $G_{A_0}(.|\mathbf{s}_q)$ where $\mathbf{s}_q \in \mathbb{R}_+^q$

$1 - \pi_{A_0} = \mathbb{P}(A_1 = 0, \ldots, A_v = 0|\mathbf{s}_q) > 0$

Define a vector-valued process $Z_0^{(i)} := (Z_1^{(i)}, \ldots, Z_v^{(i)})$ where $Z_j^{(i)} = \sum_{k=1}^{\infty} A_{j,k}^{(i)} \delta_{\tilde{\omega}_k}$

$$Z_{\cdot}^{(i)} = \sum_{j=1}^{v} Z_j^{(i)} = \sum_{k=1}^{\infty} A_{\cdot,k}^{(i)} \delta_{\tilde{\omega}_k}$$

For each (i, k), $A_{0,k}^{(i)} := (A_{1,k}^{(i)}, \ldots, A_{v,k}^{(i)})$ is independent $G_{A_0}(.|\mathbf{s}_{q,k})$ where $\mathbf{s}_{q,k}$ are vector-valued points of a PRM with intensity $\rho_q(.|\omega)$

We say that $Z_0^{(1)}, \ldots, Z_0^{(M)}|\mu_0$ are iid IBP$(G_{A_0}|\mu_0)$, with the likelihood being

$$e^{-\sum_{j=1}^{\infty}[-M \log(1-\pi_{A_0}(\mathbf{s}_{q,j}))]} \prod_{i=1}^{M} \prod_{j=1}^{\infty} \left[\frac{G_{A_0}(d\mathbf{a}_{0,j}^{(i)}|\mathbf{s}_{q,j})}{1 - \pi_{A_0}(\mathbf{s}_{q,j})} \right]^{\mathbb{I}_{\{a_{i,j} \neq 0\}}}$$

Thanks! Questions?