Hierarchical Dirichlet Process and Relative Entropy

Shui Feng
McMaster University

The 17th Workshop on Markov Processes and Related Topics

Beijing Normal University

Nov. 25–27, 2022

- Dirichlet Process
  - Background and Definition
  - Chinese Restaurant Process (Methphor)
  - LDPs
- Hierarchical Dirichlet Process
  - Background and Definition
  - Chinese Restaurant Franchise Process (Metaphor)
  - LDPs

Dirichlet Process

Background

Consideration a population of individuals of various types. To understand the population structure, one would like to know the followings.

- Total number of types in the populations (Classification).
- Frequency of each individual types.
- Number of types with the same frequencies.

A random sample X from the population is a random variable taking values of various types and the Dirichlet process is one model for the distribution of X.

Definition

Let S be a Polish space of types and ν_0 a probability on S. For any $\alpha > 0$, let $U_1, U_2, \dots$ be a sequence of iid random variables with common distribution $Beta(1, \alpha)$. Set

$$V_1 = U_1, V_n = (1 - U_1) \cdots (1 - U_{n-1}) U_n, \quad n \geq 2.$$

The **Dirichlet process** (first appeared in Ferguson [2]) with parameters α, ν_0 is the random measure

$$\Xi_{\alpha,\nu_0} = \sum_{i=1}^{\infty} V_i \delta_{\xi_i}.$$

where $\xi_1, \xi_2, \dots$ are i.i.d. with common distribution ν_0 and is independent of $\{V_i\}_{i \geq 1}$. Denote the law of Ξ_{α,ν_0} by Π_{α,ν_0}.

The constant α and the probability ν_0 are called the concentration parameter and the base measure of Π_{α,ν_0} respectively.

Chinese Restaurant Metaphor

- Consider a Chinese restaurant with infinite number of tables.
- Each table can sit infinite number of customers.
- Customers arrive at the restaurant independently.
- The first customer picks a table at random to sit.
- Subsequent customers will either sit at an occupied table with a probability proportional to the number of customers in the table or pick a new table with a probability proportional to α.
- Each table has a random label with distribution ν_0.

Chinese Restaurant Metaphor

Let X_n denote the table label of the nth customer. Then $X_1, \ldots, X_n, \ldots$ are random samples from Π_{α,ν_0}. More specifically,

- $X_1, \ldots, X_n, \ldots$ are i.i.d. given Ξ_{α,ν_0}.
- With probability one

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \longrightarrow \Xi_{\alpha,\nu_0}, \qquad n \to \infty.$$

LDPs

Theorem 1. *As α converges to infinity, we have*

$$\mathbb{V}_\alpha = (V_1, V_2, \ldots) \to \mathbf{0} = (0, 0, \ldots)$$
$$\Xi_{\alpha,\nu_0} \to \nu_0.$$

Define

$$\nabla = \{\mathbf{z} = (z_1, z_2, \ldots) : z_i \geq 0, \sum_{k=1}^{\infty} z_i \leq 1\}$$

$M_1(S)$ = set of probabilities on S equipped with the weak topology.

Theorem 2.

$$
\begin{aligned}
P\{\mathbb{V}_\alpha \in A\} &\approx \exp\{-\alpha \inf_{\mathbf{z}\in A} I_1(\mathbf{z})\} \\
P\{\Xi_{\alpha,\nu_0} \in B\} &\approx \exp\{-\alpha \inf_{\mu\in B} I_2(\mu)\}
\end{aligned}
$$

where

$$I_1(\mathbf{z}) = \log\left(\frac{1}{1-\sum_{i=1}^{\infty} z_i}\right), \ \mathbf{z} \in \nabla$$

and

$$I_2(\mu) = H(\nu_0|\mu), \ \mathrm{supp}(\mu) \subset \mathrm{supp}(\nu_0), \ \mu \in M_1(S)$$

$H(\nu_0|\mu)$ is the relative entropy.

Hierarchical Dirichlet Process

Background

- Consider a group of populations.
- Each population consists of individuals of various types.
- Different populations in the group share the same space of types.
- Each individual population can be viewed as functioning in a random environment.
- Our focus is on a random sample Y from an individual population.

Definition

The Hierarchical Dirichlet process (HDP) introduced in [3] is a non-parametric model for the study of data from group of populations.

- The prior for each population is a (level two) Dirichlet process
- All populations share the same base measure which itself is a draw from another (level one) Dirichlet process (hence the hierarchical structure).
- Given the common base measure, the Dirichlet processes for different populations are independent and identically distributed.

Definition

- Let $\mathbb{V}_\alpha = (V_1, V_2, \ldots)$ be defined as above.

- For $\beta > 0$ and any $n \geq 1$, let W_n be a Beta$(\beta V_n, \beta(1 - \sum_{k=1}^n V_k))$ random variable.

- The random variables $W_1, W_2, \ldots$ are conditionally independent given $\mathbb{V}_\alpha$.

- Define
$$Z_1 = W_1,\ Z_n = (1 - W_1) \cdots (1 - W_{n-1}) W_n,\ n \geq 2$$
and
$$\mathbb{Z}_{\alpha,\beta} = (Z_1, Z_2, \ldots).$$

Definition

The HDP with level two concentration parameter β, level one concentration parameter α, and base distribution ν_0 is the random measure

$$\Xi_{\alpha,\beta,\nu_0} \stackrel{d}{=} \Xi_{\beta,\Xi_{\alpha,\nu_0}} \stackrel{d}{=} \sum_{i=1}^{\infty} Z_i \delta_{\xi_i},$$

where $\stackrel{d}{=}$ denotes equality in distribution.

Question: What are the asymptotic behaviour associated with limiting procedures of large α and β?

Chinese Restaurant Franchise Metaphor

- Consider a Chinese restaurant franchise.
- Each restaurant is part of the restaurant chain.
- All restaurants share the same menus.
- Each individual restaurant has infinite number of tables
- Each table can sit infinite number of customers.
- Customers arrive at the restaurant independently.
- The first customer picks a table at random to sit.

- Subsequent customers will either sit at an occupied table with a probability proportional to the number of customers in the table or pick a new table with a probability proportional to θ.

- All customers sitting at the same table share the same dish.

- The dish has a random label that is distributed according to a level one Dirichlet process.

Chinese Restaurant Franchise Metaphor

Focusing on one restaurant. Let Y_n denote the dish label of the nth customer. Then $Y_1, \ldots, Y_n, \ldots$ are random samples from the Dirichlet process Π_{α,β,ν_0}.

- $Y_1, \ldots, Y_n, \ldots$ are i.i.d. given Ξ_{α,β,ν_0}.

- With probability one

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{Y_i} \longrightarrow \Xi_{\alpha,\beta,\nu_0}, \qquad n \to \infty.$$

LDPs

Assume that

$$\alpha \to \infty, \beta \to \infty, \alpha/\beta \to c > 0.$$

and set

$$\gamma = \alpha \vee \beta.$$

Theorem 3.

$$\mathbb{Z}_{\alpha,\beta} \to \mathbf{0} = (0, 0, \ldots)$$

$$\Xi_{\alpha,\beta,\nu_0} \to \nu_0.$$

LDPs

Theorem 4. (F. [1])

$$
\begin{aligned}
P\{\mathbb{Z}_{\alpha,\beta} \in C\} &\approx \exp\{-\gamma \inf_{\mathbf{z}\in C} I_3(\mathbf{z})\} \\
P\{\Xi_{\alpha,\beta,\nu_0} \in D\} &\approx \exp\{-\gamma \inf_{\mu\in D} I_4(\mu)\}.
\end{aligned}
$$

Let

$$
(a,b) = \begin{cases} (c,1) & c < 1 \\ (1,c^{-1}) & c > 1 \\ (1,1) & c = 1 \end{cases}
$$

Then the rate function I_3 has the form

$$I_3(\mathbf{z}) = \sup_{m \geq 1} S_m(z_1, \ldots, z_m)$$

with

$$S_m(z_1, \cdots, z_m) = \inf \Bigg\{ \sum_{i=1}^{m} \left(a \log \frac{1}{1-u_i} + b \prod_{j=1}^{i-1} (1-u_j) h(u_i, w_i) \right) :$$

$$u_i, w_i \in E, u_i < 1, i \geq 1, (w_1, \cdots, (1-w_1) \cdots (1-w_{m-1}) w_m) = (z_1, \cdots, z_m) \Bigg\}$$

and

$$\begin{aligned} h(u,w) &= u \log \frac{u}{w} + (1-u) \log \frac{1-u}{1-w}, \\ \prod_{j=1}^{i-1} (1-u_j) &= 1 \text{ for } i = 1. \end{aligned}$$

The rate function I_4 is given by

$$I_4(\mu) = \begin{cases} \inf_{\nu \in M_1(S), \mathrm{supp}(\nu) \subset \mathrm{supp}(\nu_0)} \{aH(\nu_0|\nu)\} + bH(\nu|\mu)\} & \mathrm{supp}(\mu) \subset \mathrm{supp}(\nu_0) \\ +\infty & \text{else} \end{cases}$$

Comparison of Rate Functions

$$\begin{aligned} I_3(\mathbf{z}) &< I_1(\mathbf{z}) \\ I_4(\mu) &< I_2(\mu). \end{aligned}$$

Multi-Levels

$$\Xi_{\alpha_1,\ldots.\alpha_{n+1},\nu_0} \stackrel{d}{=} \Xi_{\alpha_{n+1},\Xi_{\alpha_1,\ldots,\alpha_n,\nu_0}}.$$

References

[1] Feng, S.: Hierarchical Dirichlet process and relative entropy. ArXiv: 2210.13142, (2022).

[2] Ferguson, T.S.: A Bayesian analysis of some nonparametric problems. *Ann. Stat.*, **1**, (1973), 209–230.

[3] Teh, Y.W., Jordan, M.L., Beal, M.J. and Blei, D.M.: Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.*, **101**, (2006), 1566–1581.