COMS E6998-2, Probabilistic Models of Discrete Data: Scribe Note 10

Instructed by *Prof. David M. Blei*

lecture on April 8^{th}, 2016

Da Tang
Computer Science Department

This week, we discussed the paper by Zhou and Carin [1] on **Negative Binomial Process** (NBP) and its application. We started from introducing the Negative Binomial distribution and related stochastic processes (Poisson and Multinomial Processes). Then we discussed the NBP itself and its application's (actually it is NBP's extension's) application.

1 Negative Binomial Distribution

To discuss the Negative Binomial Process, we should first introduce the Negative Binomial distribution. In the paper [1], the authors pointed out two constructions for this distributions. For simplicity, I will just call them the "Gamma-Poisson" construction and "Poisson-Logarithmic" construction.

1.1 The Gamma-Poisson construction

Suppose we have a **Poisson distribution** with probability mass function as follows

$$f(m) = \frac{\lambda^m e^{-\lambda}}{m!}, m \in \mathbb{Z}^+. \tag{1}$$

We add some priors on λ and m: $\lambda \sim \text{Gamma}(r, \frac{p}{1-p})$, $m \sim \text{Pois}(\lambda)$. Marginalizing out λ gives out **Negative Binomial distribution**:

$$f(m|r,p) = \frac{\Gamma(r+m)}{m!\Gamma(r)}(1-p)^r p^m. \tag{2}$$

Here $r > 0$ is called the *dispersion parameter* and $p \in (0,1)$ is the probability of success for a single round of an experiment.

When r is an positive integer, the mass $f(m|r,p)$ has some special meaning that it could be view as the probability of having made exactly m successful trials before the r-th failure, if the trials are successful i.i.d with probability p.

1.2 The Poisson-Logarithmic construction

Alternatively, we could generate a Negative Binomial distribution in the following way: First we draw an integer l from a Poisson distribution: $l \sim \text{Pois}(-r\log(1-p))$. Then we draw l variables $u_t \sim \text{Log}(p)$ for

$t = 1, ..., l$ (i.i.d.), and then sum them to get $m = \sum_{t=1}^{l} u_t$. After that, we will have $m \sim \text{NB}(r, p)$. Here the distribution Log (**Logarithmic distribution**) is:

$$f(k) = -\frac{p^k}{k \log(1-p)}, k \in \mathbb{Z}^+. \tag{3}$$

The joint distribution of (m, l) given r and p is called the **Poisson-Logarithmic bivariate distribution**: where $m \in \mathbb{Z}^+$ and $l \in [1, m]$,

$$\text{PoisLog}(m, l|r, p) = \text{NB}(m|r, p)\text{CRT}(l|m, r). \tag{4}$$

Here **CRT** stands for Chinese Restaurant Table.

In [1], we know the equivalence between these two constructions:

Theorem 1 (The equivalence of two constructions for the Negative Binomial distribution)**.** *$PoisLog(m, l|r, p) = Pois(l|-r\log(1-p))SumLog(m|l, p)$, where $SumLog(m|l, p)$ is the distribution of m given l, p from $u_t \sim Log(p)$ for $t = 1, ..., l$ (i.i.d.), and $m = \sum_{t=1}^{l} u_t$.*

2 Poisson and Multinomial Processes

First, we define $X \sim \text{PP}(G)$, i.e. $X(A) \sim \text{Pois}(G(A))$ for any subset $A \subseteq \Omega$, where G is a complete random measure.

Also, we define $Y \sim \text{MP}(Y(\Omega), \frac{G}{G(\Omega)})$, i.e. $(Y(A_1), ..., Y(A_n)) \sim \text{Mult}(Y(\Omega), (\frac{G(A_1)}{G(\Omega)}, ..., \frac{G(A_n)}{G(\Omega)}))$ for any partition $(A_1, ..., A_n)$ of Ω, and $Y(\Omega) \sim \text{Pois}(G(\Omega))$.

Then we will have that X and Y are equivalent in distribution. With this property, we could see that Poisson Processes just unite the problems of count and mixture modeling perfectly.

3 Negative Binomial Process

Then we could define the Negative Binomial Process (NBP). Same as the Negative Binomial distribution, we also have two constructions for NBP.

3.1 The Gamma-Poisson construction

Say we draw a rate measure $G \sim \text{GaP}(\frac{J(1-p)}{p}, G_0)$. And then we draw $X_j \sim \text{PP}(G)$ for $j = 1, ..., J$, meaning that $X_j(\omega_k) \sim \text{Pois}(G(\omega_k))$ for each atom ω_k.

Define $X = \sum_{j=1}^{J} X_j$. Then we say that X is drawn from a Negative Binomial Process: $X \sim \text{NBP}(G_0, p)$. We have the property that $X(A) \sim \text{NB}(G(A), p)$ for any subset $A \subseteq \Omega$.

In class, we had an interesting discussion on the parameter J. It seems weird why (or how) we specify such a parameter. However, in theory, it does not matter much how we set the parameter J. Because that though the number of X_j increases as we increase J, the weights on the rate measure G will become smaller since $\frac{J(1-p)}{p}$ is larger.

3.2 The Poisson-Loagarithmic construction

Draw $K^+ \sim \text{Pois}(-\gamma_0 \log(1-p))$, where $\gamma_0 = G_0(\Omega)$ is the total mass. Define the new rate measure $g_0(d\omega) = \frac{G_0(d\omega)}{\gamma_0}$. Generate $(n_k, \omega_k) \sim \text{Log}(n_k|p) g_0(\omega_k)$ for $k = 1, ..., K^+$. Then $X = \sum_{k=1}^{K^+} n_k \delta_{\omega_k}$ comes from the Negative Binomial process $\text{NBP}(G_0, p)$.

We could show that this construction is equivalent to the construction in Section 3.1.

4 Extension of NBP for Mixed Membership Models

Draw $G \sim \text{GaP}(c, G_0)$, then draw $X_j \sim \text{NBP}(G, p_j)$ for $j = 1, ..., J$. This is called a **gamma-NB** process. It could be used in modeling grouped data and it can be reduced to an HDP. For mixed membership modeling, we could generate topics from the distribution G_0 and then the NBP will determine the distributions over topics. X_j (for $j = 1, .., J$) are documents and and they are drawn from the same rate measure. One key point of using NBP to model grouped data is that we could unite the traditional mixture models (or mixed membership models) like DP (or HDP) with count modeling together without having the over-dispersion issue of directly introducing Poisson to mixture models. This build a bridge for us to go from mixture modeling to count modeling. On the other hand, the new construction for NB and NBP will make the inference for them easier. We had some discussions on this, but did not go into details.

References

[1] MengChu Zhou and Lawrence Carin. Negative binomial process count and mixture modeling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(2):307–320, 2015.