Poisson Graphical Models

Pradeep Ravikumar
Carnegie Mellon University

Climate Informatics Workshop, 2016
Multivariate Count Data

- Climate Studies
- Spatial Incidence Data
- Case, disease incidence data
- Crime statistics
- Ad clicks
- Call-logs
- Document word counts
- Next generation sequencing
Multivariate Models

• Need **multivariate models** that can jointly model the multivariate count data

• Can be used to answer questions:
  
  • What is the likely activation level of gene X given the activation levels of genes Y and Z?
  
  • Given large counts for words “graphical” and “models” in word corpus, what are the likely counts for “machine” and “learning”?
Multivariate Models

• Need **multivariate models** that can jointly model the multivariate count data

• The dependencies among the multiple count-valued variables can be represented by a **graph**

• Such a dependency graph can be used for visualization, as well as scientific analyses:
  - discovering graph hubs (“biomarkers”, “potential drug targets”)
  - graph clusters (“regulatory pathways”), etc.
Multivariate Models: Networks

- Genomic networks characterizing multivariate models for microarray data
Multivariate Models for Multivariate Count Data

• For a single count-valued variable, most popular distribution is the Poisson:

\[ \mathbb{P}_{\text{Poiss}}(x | \lambda) = \frac{\lambda^x}{x!} \exp(-\lambda) \]

for count values \( x \in \{0, 1, 2, \ldots\} \), and where \( \lambda \) is the standard mean parameter.
Multivariate Models for Multivariate Count Data

• Key Question: How can we obtain multivariate extensions of the standard univariate Poisson distribution?
Multivariate Poisson

• Three classes of multivariate Poisson distributions
  1. Mixtures of independent Poissons
  2. Where marginals are Poisson
  3. Where conditionals are Poisson
Mixtures of
Independent Poissons
Mixtures of Independent Poissons

Mixture of Poissons

Multivariate Poissons

Mixture
Independent Poissons

- Independent multivariate Poisson distribution:

\[ P(x_1, \ldots, x_d | \lambda_1, \ldots, \lambda_d) := \prod_{i=1}^{d} P_{\text{Poiss}}(x_i | \lambda_i) \]

- Assumes the variables are independent i.e. no dependencies

- Not likely to hold in practice
Mixture of Ind. Poissons

With \( \mathbf{x} := (x_1, \ldots, x_d) \), \( \lambda := (\lambda_1, \ldots, \lambda_d) \):

\[
\mathbb{P}_{\text{MixedPoi}}(\mathbf{x}) = \int_{\mathbb{R}^d_+} g(\lambda) \prod_{i=1}^d \mathbb{P}_{\text{Poiss}}(x_i | \lambda_i) \, d\lambda
\]

- Given latent rate parameters, independent Poissons
- Marginalizing out latent rate parameters: Mixture of independent Poissons
Properties

\[
\mathbb{E}(x) = \mathbb{E}(\lambda), \\
\text{Var}(x) = \mathbb{E}(\lambda) + \text{Var}(\lambda).
\]

- Higher order moments also a function of lambda
- Permits both positive and negative dependencies by proper choice of mixing distribution \( g(\lambda) \)

If \( g(\lambda_1, \lambda_2) \) imposes a positive correlation between \( \lambda_1, \lambda_2 \), then \( x_1, x_2 \) are likely to be positively correlated as well.

For the Poisson case, let \( \lambda \in \mathbb{R}^d_{++} \) be a length \( d \) vector whose \( i \)-th element \( \lambda_i \) is the parameter of the Poisson distribution for \( x_i \). Now, given some mixing distribution \( g(\cdot) \), the family of Poisson mixture distributions is defined as

\[
P_{\text{MixedPoi}}(x) = \int_{\mathbb{R}^d_{++}} g(\lambda) d\lambda, \quad (5)
\]

where the domain of the joint distribution is any count-valued assignment (i.e. \( x_i \in \mathbb{Z}_+ \), \( 1 \leq i \leq d \)). While the probability density function (5) has the complicated form involving an integral, the mean and variance are known to be expressed succinctly as

\[
\mathbb{E}(x) = \mathbb{E}(\lambda), \\
\text{Var}(x) = \mathbb{E}(\lambda) + \text{Var}(\lambda). \quad (6)
\]
Choices of Mixing Distribution

- Log-normal
- Log-gamma
- Scaled gamma
Mixtures of Independent Poissons

- Caveats:
  - Inference via MLE is intractable
  - Functionals such as marginal, conditional distributions are typically intractable
Log-Normal Mixture Models

• Limited range of dependencies due to mixing of independent Poissons
Where marginals are Poisson
Where marginals are Poisson
Where marginals are Poisson: Additive Models
Additive Poisson Models

• Suppose:

\[ x_1' \sim \text{Pois}(\lambda_1) \]
\[ x_2' \sim \text{Pois}(\lambda_2) \]
\[ z \sim \text{Pois}(\lambda_0) \]

• Let:

\[ x_1 = x_1' + z \]
\[ x_2 = x_2' + z \]

• Since the sums of independent Poissons is Poisson, 
  \( x_1 \) is \textbf{Poisson} with rate \( \lambda_1 + \lambda_0 \),
  \( x_2 \) is \textbf{Poisson} with rate \( \lambda_2 + \lambda_0 \)
Additive Poisson Models

• $x_1, x_2$ are marginally Poisson

• What about their joint distribution?

$$
P_{\text{BiPoi}}(x_1, x_2 \mid \lambda_1, \lambda_2, \lambda_0)$$

$$= \exp(-\lambda_1 - \lambda_2 - \lambda_0) \frac{\lambda_1^{x_1}}{x_1!} \frac{\lambda_2^{x_2}}{2!} \sum_{z=0}^{\min(x_1, x_2)} \binom{x_1}{z} \binom{x_2}{z} z! \left(\frac{\lambda_0}{\lambda_1 \lambda_2}\right)^z.$$  

Wicksell 1916, M'Kendrick 1925, Campbell 1934

• Covariance($x_1, x_2$) = $\lambda_0$
Additive Poisson Models

- d-dimensional generalization:

\[ P_{\text{MulPoi}}(\mathbf{x}; \boldsymbol{\lambda}) = \exp \left( - \sum_{i=0}^{d} \lambda_i \right) \left( \prod_{i=1}^{d} \frac{\lambda_i^{x_i}}{x_i!} \right) \sum_{z=0}^{\min x_i} \left( \prod_{i=1}^{d} \binom{x_i}{z} \right) z! \left( \frac{\lambda_0}{\prod_{i=1}^{d} \lambda_i} \right)^z. \]

Additive Poisson Models

- Caveats:
  - Can only model positive dependencies
  - Complicated, intractable form in high dimensions
  - Inference is typically intractable
Where marginals are Poisson: Copula Models
Copulas

A copula is defined as a joint cumulative distribution, 

\[ C(\cdot) : [0, 1]^d \rightarrow [0, 1], \]

with uniform marginal distributions.
A copula is defined as a joint cumulative distribution,

\[ C(\cdot) : [0, 1]^d \to [0, 1], \]

with uniform marginal distributions.

Example: Gaussian Copula:

\[ C(u_1, \ldots, u_d) := H_R \left( H^{-1}(u_1), \ldots, H^{-1}(u_d) \right), \]

\( H^{-1}(\cdot) \): standard normal inverse cumulative distribution function,
\( H_R(\cdot) \): joint cumulative distribution function of \( \mathcal{N}(0, R) \),
where \( R \) is a correlation matrix.
Copula Model with Poisson Marginals

- Joint Cumulative Distribution with Poisson Marginals:

\[
G(x_1, x_2, \ldots, x_d \mid \theta) = C_\theta (F_1(x_1 \mid \lambda_1), \ldots, F_d(x_d \mid \lambda_d)),
\]

where \( F_i(x_i \mid \lambda_i) \) is the Poisson cumulative distribution function with parameter \( \lambda_i \) and \( \theta \) denotes the copula parameters.
Example: **Gaussian Copula Model with Poisson Marginals**

- Joint Cumulative Distribution with Poisson Marginals:

  \[ G(x_1, x_2, \cdots, x_d) = H_R \left( H^{-1}(F_1(x_1 | \lambda_1)), \cdots, H^{-1}(F_d(x_d | \lambda_d)) \right) \]

  Xue-Kun Song, 2000, Yahav and Shmueli, 2012, Cook et al., 2010.
Copula Models with Poisson Marginals

- Caveats:
  - When marginals are discrete, copula model parameters are not identifiable
  - Restrictive set of dependencies
  - Does not have a closed form density
Copula Models

• Negative-dependencies still place considerable mass at \([x,x]\)
Where conditionals are Poisson
Where conditionals are Poisson

Conditional Poisson

Conditionals are Poisson
Conditional Poissons

- Suppose the conditional distributions are Poisson:

\[ P(x_i | x_{-i}) = \exp\{\psi(x_{-i}) x_i - \log(x_i!) - \exp(\psi(x_{-i}))\} , \]

where \( x_{-i} \) is the set of all \( x_j \) except \( x_i \),
and \( \psi(x_{-i}) \) is any function that depends on rest of variables except \( x_i \).
Conditional Poissons

• Suppose the conditional distributions are Poisson:

\[
P(x_i \mid x_{-i}) = \exp\{\psi(x_{-i}) x_i - \log(x_i!) - \exp(\psi(x_{-i}))\},
\]

where \(x_{-i}\) is the set of all \(x_j\) except \(x_i\),
and \(\psi(x_{-i})\) is any function that depends on rest of variables except \(x_i\).

• Questions:

  • Does there exist a consistent joint?
  • If so, is it unique? What form does it take?
Conditional Poissons

- [Besag 74, Yang et al, 2015] When the conditionals are Poisson as earlier, then a consistent joint distribution does exist and takes the form:

\[
P(\mathbf{x} | \eta) = \exp \left\{ \sum_{C \in \mathcal{C}} \eta_C \prod_{i \in C} x_i - \sum_{i=1}^{d} \log(x_i!) - A(\eta) \right\},
\]

Poisson Graphical Model
Conditional Poissons

- [Besag 74, Yang, Ravikumar, Allen, Liu 2015] When the conditionals are Poisson as earlier, then a consistent joint distribution does exist.

Pairwise Case: with interaction factors of size at most two

\[
\mathbb{P}_{\text{PGM}}(\mathbf{x} | \boldsymbol{\eta}) = \exp \left\{ \sum_{i=1}^{d} \eta_i x_i + \sum_{(i,j) \in \mathcal{E}} \eta_{ij} x_i x_j - \sum_{i=1}^{d} \log(x_i!) - A_{\text{PGM}}(\boldsymbol{\eta}) \right\}
\]
Poisson Graphical Model

• [Besag 74] The Poisson Graphical Model distribution is not normalizable unless the interaction parameters are non-positive (i.e. zero or negative).

• Consequence: only allows negative dependencies!
Poisson Graphical Model

Variants

• Why does the Poisson graphical model only permit negative dependencies?

  • the interaction terms $x_i x_j$ scale quadratically $O(x^2)$, while log-base measure $-\log(x_i!)$ scales as $O(- x \log x)$

  • So if the interaction terms are positive, the net unnormalized measure goes to infinity
Why does the Poisson graphical model only permit negative dependencies?

- the interaction terms $x_i \times x_j$ scale quadratically $O(x^2)$, while log-base measure $-\log(x_i!)$ scales as $O(-x \log x)$

Three Approaches to address this:

1. Truncate domain (allow count values $\leq R$)
2. make the log-base measure term scale faster
3. make the interaction terms scale slower
PGM Variants

• Three Approaches to address this:

1. Truncate domain (allow count values $\leq R$)

2. make the log-base measure term scale faster

3. make the interaction terms scale slower
Square-Root PGM

• Suppose we modify the Poisson as follows:

\[ P(Z) = \exp \left( \theta \sqrt{Z} - \log(Z!) - A(\theta) \right) \]

• Specific form of sub-linear sufficient statistics
Square-Root PGM

• With the conditional distributions are set to the square-root Poisson defined earlier, there does exist a unique consistent joint with the form:

\[ P_{\text{SQR}}(x \mid \theta) = \exp\left\{ \theta^T \sqrt{x} + \sqrt{x}^T \Phi \sqrt{x} - \sum_i \log(x_i!) - A_{\text{SQR}}(\theta, \Phi) \right\} \]

Inouye, Ravikumar, Dhillon, 2016
SQR-PGM

• Allows both positive and negative dependencies unlike original PGM

• Caveat: Do not have closed form expressions for log-normalization constant
SQR-PGM

- Permits both strong negative, and strong positive dependencies
Software

• X-MRF

• R package available: https://cran.r-project.org/web/packages/XMRF

• Efficiently learns Conditional Poisson based Graphical Models (as well as some other classes of graphical models) even for very high-dimensional data
Summary

• Multivariate Poisson distributions come in three flavors: (1) mixture of independent Poissons, (2) where marginals are Poisson, and (3) where conditionals are Poisson

• Marginal Poisson models better at modeling positive dependencies

• Conditional Poisson Models better at modeling negative dependencies
  • Variants of Conditional Poisson Models can model both strong positive and negative dependencies
  • Inference scalable to high-dimensional settings
  • Representable as graphical models
Thank You!
PGM Variants

• Three Approaches to address this:

1. **Truncate domain (allow count values <= R)**
2. make the log-base measure term scale faster
3. make the interaction terms scale slower
Truncated Poisson

• Consider the following modification of Poisson:

1. Restrict domain to $X = \{0, 1, \ldots, R\}$.
2. Truncated Poisson Distribution:

   $$
P(Z) = \frac{\exp\left\{\theta Z - \log(Z!}\right\}}{\sum_{k=0}^{R} \exp\left\{\theta k - \log(k!)\right\}}.
   $$

3. Redistributes mass to all possible values, 0, 1, \ldots, R.
Truncated Poisson Graphical Model (TPGM)

• With the conditional distributions are set to the truncated Poisson defined earlier, there does exist a unique consistent joint with the form:

$$P_{\text{TPGM}}(x) = \exp\{\theta^T x + x^T \Phi x - \sum_i \log(x_i!) - A_{\text{TPGM}}(\theta, \Phi)\}.$$ 

Yang, Ravikumar, Allen, Liu 2013
TPGM: Caveats

1. Trade-off between \( R \) and the types of dependencies that can be modeled:
   - If \( R \) is small, (unequal) re-distribution of lot of prob. mass.
   - If \( R \) is large, stronger restrictions on \( \theta_{st} \) values.

2. Value of \( R \) has to be fixed apriori; TPGM thus models variables with finite domain.
PGM Variants

• Three Approaches to address this:

1. Truncate domain (allow count values $\leq R$)

2. make the log-base measure term scale faster

3. make the interaction terms scale slower
Quadratic PGM

• What if we modify the Poisson to use a quadratic log-base measure:

\[ P(Z) \propto \exp(\theta Z - Z^2) \]

• When the conditional distributions are set to the above, the unique consistent joint takes the form:

\[ P_{QPGM}(x) = \exp\{\theta^T x + x^T \Phi x - A_{QPGM}(\theta, \Phi)\}. \]

Yang, Ravikumar, Allen, Liu 2013
QPGM: Caveats

\[ P_{\text{QPGM}}(\mathbf{x}) = \exp\{\theta^T \mathbf{x} + \mathbf{x}^T \Phi \mathbf{x} - A_{\text{QPGM}}(\theta, \Phi)\}. \]

- Gaussian-esque thin tails rather than Poisson-esque thicker tails (i.e. ‘looks’ more Gaussian than Poisson)
Sublinear PGM

- Suppose we modify the Poisson as follows:

\[ P(Z) = \exp(\theta B(Z; R_0, R) - \log Z! - D(\theta)). \]

\[ B(x; R_0, R) = \begin{cases} 
  x & \text{if } x \leq R_0 \\
  \frac{1}{2(R-R_0)} x^2 + \frac{R}{R-R_0} x - \frac{R_0^2}{2(R-R_0)} & \text{if } R_0 < x \leq R \\
  \frac{R}{R+R_0} x - \frac{R_0^2}{2(R-R_0)} & \text{if } x \geq R 
\end{cases} \]
Sublinear Poisson Graphical Model (SPGM)

- With the conditional distributions are set to the sublinear Poisson defined earlier, there does exist a unique consistent joint with the form:

\[
P(X) = \exp \left\{ \sum_{s \in V} \theta_s B(X_s; R_0, R) + \sum_{(s,t) \in E} \theta_{st} B(X_s; R_0, R)B(X_t; R_0, R) \right. \\
- \left. \sum_{s \in V} \log X_s! - A(\theta, R_0, R) \right\}.
\]

Yang, Ravikumar, Allen, Liu 2013
Fixed-Length PGM

• Consider the PGM distribution conditioned on the sum of counts:

\[ P_{\text{FLPGM}}(x \mid \|x\|_1=L, \theta, \Phi) = \exp\{\theta^T x + x^T \Phi x - \sum_i \log(x_i!) - A_L(\theta, \Phi)\} \]

• Due to bounded domain, is normalizable for arbitrary interaction parameters: i.e. allows both positive, negative dependencies
Fixed-Length PGM

• Consider the PGM distribution conditioned on the sum of counts:

\[ \mathbb{P}_{\text{FLPGM}}(\mathbf{x} \mid \|\mathbf{x}\|_1=L, \theta, \Phi) = \exp\{\theta^T \mathbf{x} + \mathbf{x}^T \Phi \mathbf{x} - \sum_i \log(x_i!) - A_L(\theta, \Phi)\} \]

• Can be used to specify multivariate count-valued distribution:

\[ \mathbb{P}(\mathbf{x} \mid \theta, \Phi, \lambda) = \mathbb{P}(L \mid \lambda) \ \mathbb{P}_{\text{FLPGM}}(\mathbf{x} \mid \|\mathbf{x}\|_1=L, \theta, \Phi) \]
TPGM

- Strong negative, but limited positive dependencies