Multivariate Poisson models

Dimitris Karlis

Department of Statistics

Athens University of Economics

Limburg, October 2002

Outline

- Motivation
- Bivariate Poisson models
- Multivariate Poisson models
- Fully structure multivariate Poisson models
- Multivariate Poisson regression
- Multivariate Poisson mixtures
- Finite Multivariate Poisson mixtures
- Conclusions -Open problems

Motivation

Multivariate data are usually modelled via

- Multivariate Normal models
- Multinomial models (for categorical data)

What about multivariate count data?

- Small counts with a lot of zeros
- Normal approximation may not be adequate at all

Idea: Use multivariate Poisson models

Attractive idea but the models are computationally demanding.

Multivariate Count data

- Different type of crimes in different areas
- Purchases of different products
- Accidents (different types or in different time periods)
- Football data
- Different types of faults in production systems
- Number of faults in parts of a large system etc

Bivariate Poisson model

Let $X_i \sim Poisson(\theta_i)$, $i = 0, 1, 2$

Consider the random variables

$$\begin{aligned} X &= X_1 + X_0 \\ Y &= X_2 + X_0 \end{aligned}$$

$(X, Y) \sim BP(\theta_1, \theta_2, \theta_0)$,

Joint probability function given:

$$P(X = x, Y = y) = e^{-(\theta_1+\theta_2+\theta_0)} \frac{\theta_1^x}{x!} \frac{\theta_2^y}{y!} \sum_{i=0}^{min(x,y)} \binom{x}{i} \binom{y}{i} i! \left(\frac{\theta_0}{\theta_1\theta_2}\right)^i .$$

Properties of Bivariate Poisson model

- Marginal distributions are Poisson, i.e.

$$\begin{aligned} X &\sim Poisson(\theta_1 + \theta_0) \\ Y &\sim Poisson(\theta_2 + \theta_0) \end{aligned}$$

- Conditional Distributions : Convolution of a Poisson with a Binomial
- Covariance: $Cov(X, Y) = \theta_0$

  For a full account see Kocherlakota and Kocherlakota (1992) and Johnson, Kotz and Balakrishnan (1997)

Bivariate Poisson model (more)

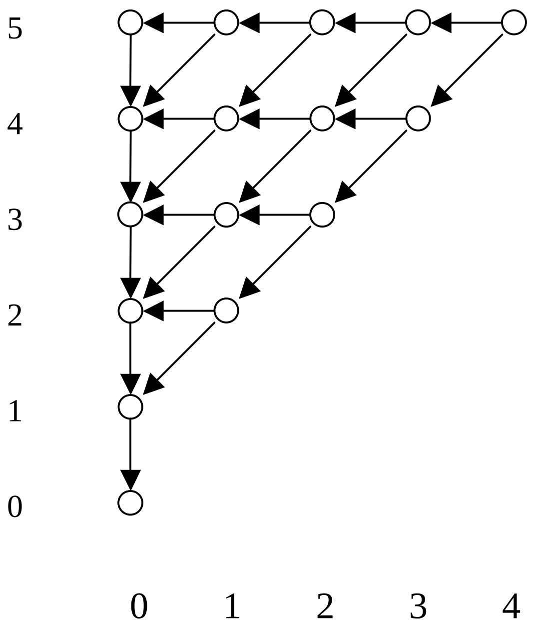
Limited use because of computational problems.

Recursive relationships:

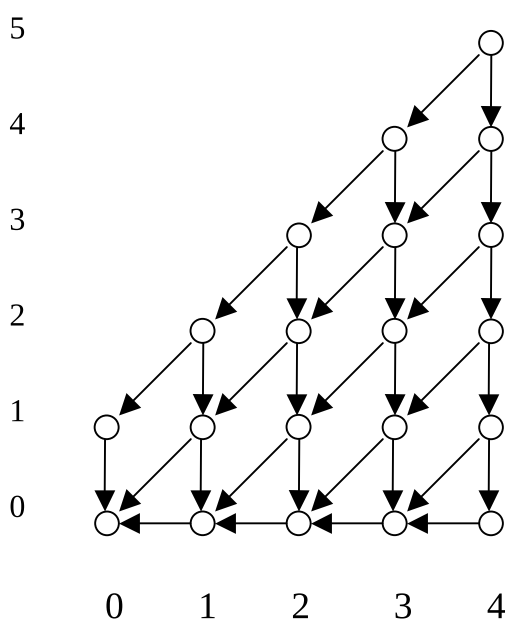
$$\begin{aligned} xP(x,y) = &\quad \theta_1 P(x-1,y) + \theta_0 P(x-1,y-1) \\ yP(x,y) = &\quad \theta_2 P(x,y-1) + \theta_0 P(x-1,y-1). \end{aligned} \tag{1}$$

with the convention that $P(x,y) = 0$, if $s < 0$.

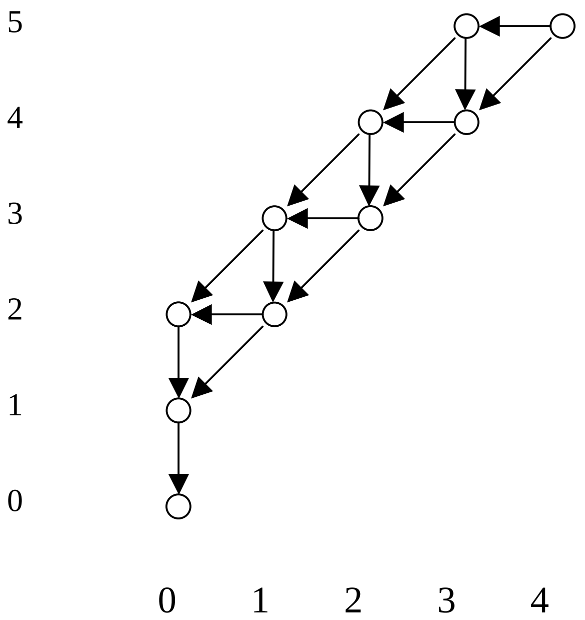
Need for "clever" use of thee relationships (See, e.g. Tsiamyrtzis and Karlis, 2002).



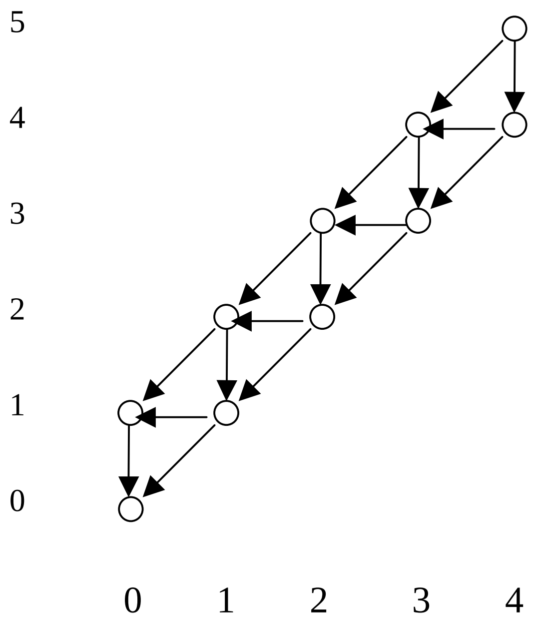
1a



1b



1c



1d

Figure 1

Bivariate Poisson model (estimation)

Various techniques:

- Moment method, Maximum likelihood, Even points etc (see, Kocherlakota and Kocherlakota, 1992).
- Recently: Bayesian estimation (Tsionas, 1999).

Bivariate Poisson regression model

$$
\begin{aligned}
(X_i, Y_i) &\sim BP(\theta_{1i}, \theta_{2i}, \theta_{0i}) \\
log(\theta_{ji}) &= \boldsymbol{x}_i' \boldsymbol{\beta}_j, \quad j = 0, 1, 2
\end{aligned}
$$

- Allows for covariate-dependent covariance.
- Separate modelling of means and covariance
- Standard estimation methods not easy to apply.
- Computationally demanding.
- Application of an easily programmable EM algorithm

Application of Bivariate Poisson regression model

Champions league data of season 2000/01

The model

$$
\begin{aligned}
(X,Y)_i &\sim BP(\lambda_{1i},\lambda_{2i},\lambda_{0i}) \\
\log(\lambda_{1i}) &= \mu + home + att_{h_i} + def_{g_i} \\
\log(\lambda_{2i}) &= \mu + att_{g_i} + def_{h_i}.
\end{aligned}
$$

Use of sum-to-zero or corner constraints

Interpretation

- the overall constant parameter specifies λ_1 and λ_2 when two teams of the same strength play on a neutral field.
- Offensive and defensive parameters are expressed as departures from a team of average offensive or defensive ability.

Application of Bivariate Poisson regression model (2)

Modelling the covariance term

$$\log(\lambda_{0i}) = \beta^{con} + \gamma_1 \beta^{home}_{h_i} + \gamma_2 \beta^{away}_{g_i}$$

γ_1 and γ_2 are dummy binary indicators taking values zero or one depending on the model we consider. Hence when $\gamma_1 = \gamma_2 = 0$ we consider constant covariance, when $(\gamma_1, \gamma_2) = (1, 0)$ we assume that the covariance depends on the home team only etc.

Results(1)

Table 1: Details of Fitted Models for Champions League 2000/01 Data ($^1H_0 : \lambda_0 = 0$ and $^2H_0 : \lambda_0 = constant$, B.P. stands for the Bivariate Poisson).

	Model Distribution	Model Details	Log-Lik	Param.	p.value	AIC	BIC
1	Poisson		-432.65	64		996.4	1185.8
		$\underline{\lambda_0}$					
2	Biv. Poisson	constant	-430.59	65	0.042^1	994.3	1186.8
3	Biv. Poisson	Home Team	-414.71	96	0.438^2	1024.5	1311.8
4	Biv. Poisson	Away Team	-416.92	96	0.655^2	1029.0	1316.2
5	Biv. Poisson	Home and Away	-393.85	127	0.151^2	1034.8	1428.8

Results(2)

Table 2:

Home	Away Goals						
	0	1	2	3	4	5	Total
0	10(17.3)	11(10.5)	5(4.2)	3(1.4)	0(0.4)	1(0.1)	30(33.9)
1	20(17.9)	17(14.8)	2(6.8)	3(2.5)	1(0.8)	0(0.2)	43(43.0)
2	14(12.8)	13(11.9)	6(6.1)	2(2.4)	0(0.8)	0(0.2)	35(34.2)
3	10 (7.6)	8 (7.6)	8(4.1)	2(1.7)	0(0.6)	0(0.2)	28(21.8)
4	3 (4.1)	4 (4.2)	3(2.4)	1(1.0)	1(0.4)	0(0.1)	12(12.2)
5	3 (2.0)	2 (2.2)	0(1.3)	1(0.5)	0(0.2)	0(0.1)	6 (6.3)
6	1 (1.0)	1 (1.1)	0(0.6)	0(0.3)	0(0.1)	0(0.0)	2 (3.1)
7	0 (0.4)	0 (0.5)	1(0.3)	0(0.1)	0(0.0)	0(0.0)	1 (1.3)
Total	61 (63.1)	56(52.8)	25 (25.8)	12(9.9)	2 (3.3)	1(0.9)	157 (155.8)*

Table 1: Estimated Parameters for 2000/01 Champions League Data.

		Poisson		Bivariate Poisson	
	Team	Att	Def	Att	Def
1	Anderlecht	0.23	0.32	0.30	0.40
2	Arsenal	0.09	-0.14	-0.01	-0.26
3	B.Munich	0.06	-0.87	0.09	-1.13
4	Barcelona	0.29	0.29	0.36	0.37
5	Besiktas	-0.73	0.63	-0.69	0.78
6	Deportivo	-0.19	-0.30	-0.24	-0.34
7	Dynamo Kyiv	-0.16	-0.25	-0.09	-0.17
8	Galatasaray	-0.03	0.10	-0.03	0.13
9	Hamburger SV	-0.05	0.25	-0.06	0.42
10	Heerenveen	-0.29	-0.22	-0.40	-0.25
11	Helsingborg	-0.43	0.23	-0.59	0.25
12	Juventus	0.13	0.62	0.07	0.77
13	Lazio	0.05	-0.20	0.05	-0.20
14	Leeds	0.04	-0.12	0.14	-0.03
15	Leverkusen	-0.05	0.42	-0.14	0.42
16	Lyon	0.46	-0.54	0.67	-0.55
17	Man.UND	0.27	-0.41	0.26	-0.60
18	Milan	-0.07	-0.23	-0.10	-0.31
19	Monaco	0.37	0.24	0.37	0.20
20	Olympiakos	0.22	-0.68	0.38	-0.77
21	PSG	0.27	0.12	0.32	0.15
22	PSV Eindhoven	0.15	-0.04	0.14	-0.07
23	Panathinaikos	-0.55	-0.16	-0.93	-0.26
24	Rangers	-0.03	-0.28	-0.04	-0.35
25	Real M.	0.41	0.08	0.46	0.07
26	Rosenborg	0.52	0.59	0.67	0.74
27	Shakhtar	0.16	0.69	0.23	0.92
28	Sparta	-0.52	0.36	-0.75	0.45
29	Spartak	-0.19	-0.35	-0.04	-0.26
30	Sporting	-0.59	0.51	-0.82	0.58
31	Sturm	-0.06	0.34	0.22	0.58
32	Valencia	0.22	-1.00	0.20	-1.68
	Other Parameters				
	Intercept	-0.05		-0.39	
	Home	0.64		0.79	
	λ_3	0.00		0.24	

Generalizing the model

Let $Y_i \sim Poisson(\theta_i)$, $i = 0, 1, \ldots, m$

Consider the random variables

$$\begin{aligned} X_1 &= Y_1 + Y_0 \\ X_2 &= Y_2 + Y_0 \\ &\cdots \\ X_m &= Y_m + Y_0 \end{aligned}$$

Then $(X_1, \ldots, X_m)$ jointly follow a multivariate Poisson distribution

Properties

- The joint probability function is given by

$$
\begin{aligned}
P(X) &= P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m) \\
&= exp(-\sum_{i=1}^{m} \theta_i) \prod_{i=1}^{m} \frac{\theta_i^{x_i}}{x_i!} \sum_{i=0}^{s} \prod_{j=1}^{m} \begin{pmatrix} x_j \\ i \end{pmatrix} i! \left(\frac{\theta_0}{\prod_{i=1}^{m} \theta_i} \right)^i .
\end{aligned}
$$

  where $s = min(x_1, x_2, \ldots, x_m)$.

- Marginally each X_i follows a Poisson distribution with parameter $\theta_0 + \theta_i$.
- Parameter θ_0 is the covariance between all the pairs of random variables.
- If $\theta_0 = 0$ then the variables are independent.

Problems

- Probability function too complicated for
  - even the calculation of the function (remedy: use of recurrence relationships)
  - for estimation purposes (remedy:use of an EM algorithm)
- Assumes common covariance for all pairs - unrealistic

The EM algorithm

Dempster *et al.* (1977), Meng and Van Dyk (1997), McLachlan and Krishnan (1997)

- Numerical method for finding ML estimates that offers a nice statistical perspective
- Missing data representation
- Can help in a variety of problems that can be presented as missing data problems

The EM mechanics

Let ϕ the vector of parameters of interest

Complete data $Y_i = (X_i, Z_i)$

- Observed part X_i
- Unobserved part Z_i

The steps are:

- **E-step**

  $Q(\phi \mid \phi_{(k)}) = E(logp(Y \mid \phi) \mid X, \phi_{(k)})$

  (the expectation is taken with respect to the conditional distribution $f(Y \mid X, \phi_{(k)})$ and

- **M-step** Maximize $Q(\phi \mid \phi_{(k)})$

  $\phi_{(k)}$ represents the vector of parameters after the k-th iteration

The EM in simple words

- **E-step**

  Estimate the missing part of the data using the data and the current values of the parameters and

- **M-step** Maximize the likelihood of the complete data using instead of the unobserved values their expectations from the E-step

Pros and Cons

Pros

- Easily programmable
- Estimates in the admissible range
- Byproducts of the algorithm have interesting statistical interpretation

Cons

- Slow Convergence
- Dependence on the initial values

The EM for the Multivariate Poisson

- Observed data: the vectors X_i
- Unobserved Data: the vectors Y_i

Note: Since we have m X_i's and $m+1$ Y_i's, in fact if we find one of the Y_i we get the rest easily

The EM for the Multivariate Poisson, (Karlis, 2002)

E-step: Using the data and the current estimates after the $k-th$ iteration $\theta^{(k)}$ calculate the pseudo-values

$$\begin{aligned} s_i &= E(Y_{0i} \mid X_i, t_i, \theta^{(k)}) = \\ &= \theta_0 t_i \frac{P(X_1 = x_{1i} - 1, X_2 = x_{2i} - 1, \ldots, X_m = x_{mi} - 1)}{P(X_i)} \end{aligned}$$

M-step: Update the estimates by

$$\theta_0{}^{(k+1)} = \frac{\sum_{i=1}^{n} s_i}{\sum_{i=1}^{n} t_i}, \qquad \theta_i{}^{(k+1)} = \frac{\bar{x}_i}{\bar{t}} - \theta_0{}^{(k+1)} \qquad i = 1, \ldots, m$$

If some convergence criterion is satisfied stop iterating otherwise go back to the *E-step* for one more iteration.

Interesting things

- Need for quick calculation of the probabilities
- It can be seen that for this model the EM algorithm is similar to the Newton-Raphson method
- Conditional distribution of Y_{0i}

$$f(Y_{0i} = y \mid x_i, \theta^{(k)}) = P(y) = \frac{\frac{\theta_0^y}{y!} \prod_{j=1}^{m} \frac{\theta_j^{-y}}{(x_{ji}-y)!}}{\sum_{y=0}^{s} \frac{\theta_0^y}{y!} \prod_{j=1}^{m} \frac{\theta_j^{-y}}{(x_{ji}-y)!}},$$

$y = 0, \ldots, min(x_i)$

An example-Accident Data

Accidents in 24 roads of Athens for the period 1987-1991

	Year					
Road	1987	1988	1989	1990	1991	length (km)
Akadimias	11	33	25	23	6	1.2
Alexandras	41	63	91	77	29	2.6
Amfitheas	5	35	44	21	13	2.4
...						
Peiraios	86	89	109	90	49	8.0
Sigrou	60	61	87	86	29	4.8
Estimated	$\hat{\theta}_1 = 4.902$	$\hat{\theta}_2 = 8.731$	$\hat{\theta}_3 = 11.795$			
Parameters	$\hat{\theta}_4 = 10.147$	$\hat{\theta}_5 = 2.517$	$\hat{\theta}_0 = 3.753$			

Restrictions

Important notes

- The above model is restrictive as it assumes the same covariance for all the pairs
- Almost all the applications of multivariate Poisson models imply this model.
- This model cannot generalize the idea of multivariate normal models

Extending the model

Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m)$ and $Y_i \sim Poisson(\theta_i)$, $i = 1, \ldots, m$. Then the general definition of multivariate Poisson models is made through the matrix $\mathbf{A}$ of dimensions $k \times m$, where the elements of the matrix are zero and ones and no duplicate columns exist.

Then the vector $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ defined as

$$\mathbf{X} = \mathbf{AY}$$

follows a multivariate Poisson distribution.

Complete Specification

$$\mathbf{A} = [A_1 \quad A_2 \quad \ldots \quad A_k]$$

where A_i is a matrix of dimensions $k \times \begin{pmatrix} k \\ i \end{pmatrix}$ where each column has exactly i ones and $k - i$ zeroes.

Example $k = 3$

$$A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad A_2 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad A_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

and then

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}$$

This correspond to

$$\begin{aligned} X_1 &= Y_1 + Y_{12} + Y_{13} + Y_{123} \\ X_2 &= Y_2 + Y_{12} + Y_{23} + Y_{123} \\ X_3 &= Y_3 + Y_{13} + Y_{23} + Y_{123} \end{aligned} \tag{2}$$

where all Y_i's, are independently Poisson distributed random variables with parameter θ_i, $i \in (\{1\}, \{2\}, \{3\}, \{12\}, \{13\}, \{23\}, \{123\})$

Note: Parameters θ_{ij} are in fact covariance parameters between X_i and X_j. Similarly θ_{123} is a common 3-way covariance parameter.

Other cases

Independent Poisson variables

Corresponds to the case $A = A_1$.

Example for $k = 3$

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

i.e. product of independent Poisson probability functions.

Full covariance structure

If we want to specify only up to 2-way covariances we take the form

$$\mathbf{A} = [A_1 \quad A_2]$$

Example $k = 3$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

This model is very interesting as it assumes different covariances between all the pairs and thus it resembles the multivariate normal model.

This correspond to

$$\begin{aligned} X_1 &= Y_1 + Y_{12} + Y_{13} \\ X_2 &= Y_2 + Y_{12} + Y_{23} \\ X_3 &= Y_3 + Y_{13} + Y_{23} \end{aligned}$$

where all Y_i's, are independently Poisson distributed random variables with parameter θ_i, $i \in (\{1\}, \{2\}, \{3\}, \{12\}, \{13\}, \{23\})$

the covariance matrix of (X_1, X_2, X_3) is now

$$Var(\mathbf{X}) = \begin{bmatrix} \theta_1 + \theta_{12} + \theta_{13} & \theta_{12} & \theta_{12} \\ \theta_{12} & \theta_2 + \theta_{12} + \theta_{23} & \theta_{23} \\ \theta_{13} & \theta_{23} & \theta_3 + \theta_{13} + \theta_{23} \end{bmatrix}$$

Properties

For the general model we have

$$E(X) = \mathbf{AM}$$

and

$$Var(X) = \mathbf{A\Sigma A^T}$$

where $\mathbf{M}$ and $\mathbf{\Sigma}$ are the mean vector and the variance covariance matrix for the variables $Y_0, Y_1, \ldots, Y_k$ respectively.

$\mathbf{\Sigma}$ is diagonal because of the independence of Y_i's and has the form

$$\mathbf{\Sigma} = diag(\theta_1, \theta_2, \ldots, \theta_m)$$

Similarly

$$\mathbf{M}^T = (\theta_1, \theta_2, \ldots, \theta_m)$$

Covariance Model

We concentrate on the 2-way full covariance model. A model with m variables has

$$m + \binom{m}{2}$$

parameters.

Bad news: The joint probability function has at least m summations!

Good news: One may use recurrence relationships (clearly need to find efficient algorithms to do so)

Example : a 3-variate full model

For example, consider again the case of the full trivariate Poisson model.

$$P(\boldsymbol{X}=x)=$$

$$\sum_{(y_{12},y_{13},y_{23})\in C}\frac{\exp\left(-\sum\theta_i\right)\theta_1^{x_1-y_{12}-y_{13}}\theta_2^{x_2-y_{12}-y_{23}}\theta^{x_3-y_{13}-y_{23}}\theta_{12}^{y_{12}}\theta_{13}^{y_{13}}\theta_{23}^{y_{23}}}{(x_1-y_{12}-y_{13})!(x_2-y_{12}-y_{23})!(x_3-y_{13}-y_{23})!y_{12}!y_{13}!y_{23}!},$$

where the summation is over the set $C \subset N^3$ defined as

$$C=[(y_{12},y_{13},y_{23})\in N^3:\{y_{12}+y_{13}\leq x_1\}\cup\{y_{12}+y_{23}\leq x_2\}\cup\{y_{13}+y_{23}\leq x_3\}\neq\emptyset].$$

Multivariate Poisson regression model with covariance structure

Let the vector $\boldsymbol{x}_i = (x_{1i}, x_{2i}, ..., x_{mi}), \quad i = 1, ..., n$, denotes the i-th available m-variate observation. Let $\mathcal{S} = R_1 \cup R_2$, where $R_1 = \{1, 2, \ldots, m\}$ and $R_2 = \{ij, \;\; i, j = 1, \;\; \ldots, m, \quad i < j\}$. The sets R_1 and R_2 contain the subscripts needed for the definition of the unobserved variables Y_i and the corresponding parameters θ_i, $i \in \mathcal{S}$. The model takes the form

$$\begin{aligned} \boldsymbol{X}_i &\sim m - Po(\boldsymbol{\theta} t_i), \quad i = 1, ..., n, \\ \ln \theta_{ji} &= \boldsymbol{z}_i' \beta_j, \quad i = 1, ..., n \end{aligned}$$

- t_i is an offset such as a population or an area and
- $\boldsymbol{\theta}$ is the vector of all the parameters, i.e. $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m, \theta_{12}, \ldots, \theta_{(m-1),m})$.
- β_j is the vector of regression coefficients for the j-th parameter and
- $\boldsymbol{z}_i$' is a vector of regressors not necessarily the same for all parameters

Example

Consider 3 variates X_1, X_2, X_3 that represent the purchase of 3 products. Jointly the y follow a 3-variate Poisson distribution with up yo 2-way covariance. Let z be another variable let say the sex. The model has the form

$$\begin{aligned}
(X_1, X_2, X_3)_i &\sim 3 - Poisson(\theta_{1i}, \theta_{2i}, \theta_{3i}, \theta_{12i}, \theta_{13i}, \theta_{23i}) \\
log(\theta_{1i}) &= a_1 + \beta_1 z_i \\
log(\theta_{2i}) &= a_2 + \beta_2 z_i \\
log(\theta_{3i}) &= a_3 + \beta_3 z_i \\
log(\theta_{12i}) &= a_4 + \beta_4 z_i \\
log(\theta_{13i}) &= a_5 + \beta_5 z_i \\
log(\theta_{23i}) &= a_6 + \beta_6 z_i
\end{aligned}$$

Perhaps one may use regressors only for the mean parameters $\theta_1, \theta_2, \theta_3$.

Estimation- ML estimation via EM algorithm

Denote as $\boldsymbol{\Theta}^{(j)} = {\boldsymbol{\beta}_1}^{(j-1)}, {\boldsymbol{\beta}_2}^{(j-1)}, \ldots, {\boldsymbol{\beta}_r}^{(j-1)}$, $r = 1, \ldots, k$ the vector of all the parameters after the j-th iteration.

- *E-Step:* Using the observed data and the current estimates after $j-1$ iterations $\boldsymbol{\Theta}^{(j-1)}$, calculate the pseudo values

$$\begin{aligned} s_{ir} &= E(\boldsymbol{Y}_{ir} \mid \boldsymbol{X}_i, t_i, \boldsymbol{\Theta}^{(j-1)}) \\ &= \frac{\displaystyle\sum_{y \in g^{-1}(x_i)} y_{ir} \prod_{r=1}^{k} Po(y_i \mid {\theta_{ir}}^{(j-1)} t_i)}{P(x_i \mid {\theta_{ir}}^{(j-1)}, t_i)}, \quad i = 1, , \ldots, n, r \in \mathcal{S} \end{aligned}$$

  where ${\theta_{ir}}^{(j-1)} = \exp({z_{ir}}' {\beta_r}^{(j-1)})$, $i = 1, ..., n$, $r = 1, ..., k$.

- *M-Step:* Update the vector $\boldsymbol{\beta}_r$ by fitting a Poisson regression on s_{ir}, $i = 1, \ldots, n$ and explanatory variables z_{ir}.

- If some convergence criterion is satisfied stop iterating,

Estimation- Bayesian estimation via MCMC algorithm (Karlis and Meligotsidou, 2002)

Closed form Bayesian estimation is impossible

Need to use MCMC methods

Implementation details

- Use the same data augmentation
- Jeffrey priors for regression coefficients
- The posterior distributions of β_r, $r = 1, \ldots, k$ are non-standard and, hence, Metropolis-Hastings steps are needed within the Gibbs sampler,

Application- Crime data

4 different types of crime (rapes , arson, manslaughter, smuggling)

Regressors only for the mean parameters (not to impose imposing so much structure)

Regressors used (socio-economic characteristics):

- the natural logarithm of the population in millions
- the Gross Domestic Product per capita in Euros for each prefecture (GDP),
- the unemployment rate of the prefecture (unem)
- a dummy variable to show whether the prefecture is at the borders of the country (borders)
- a dummy variable to show whether the prefecture has at least one city with population larger than 150 thousands habitants (city).

Results - EM algorithm

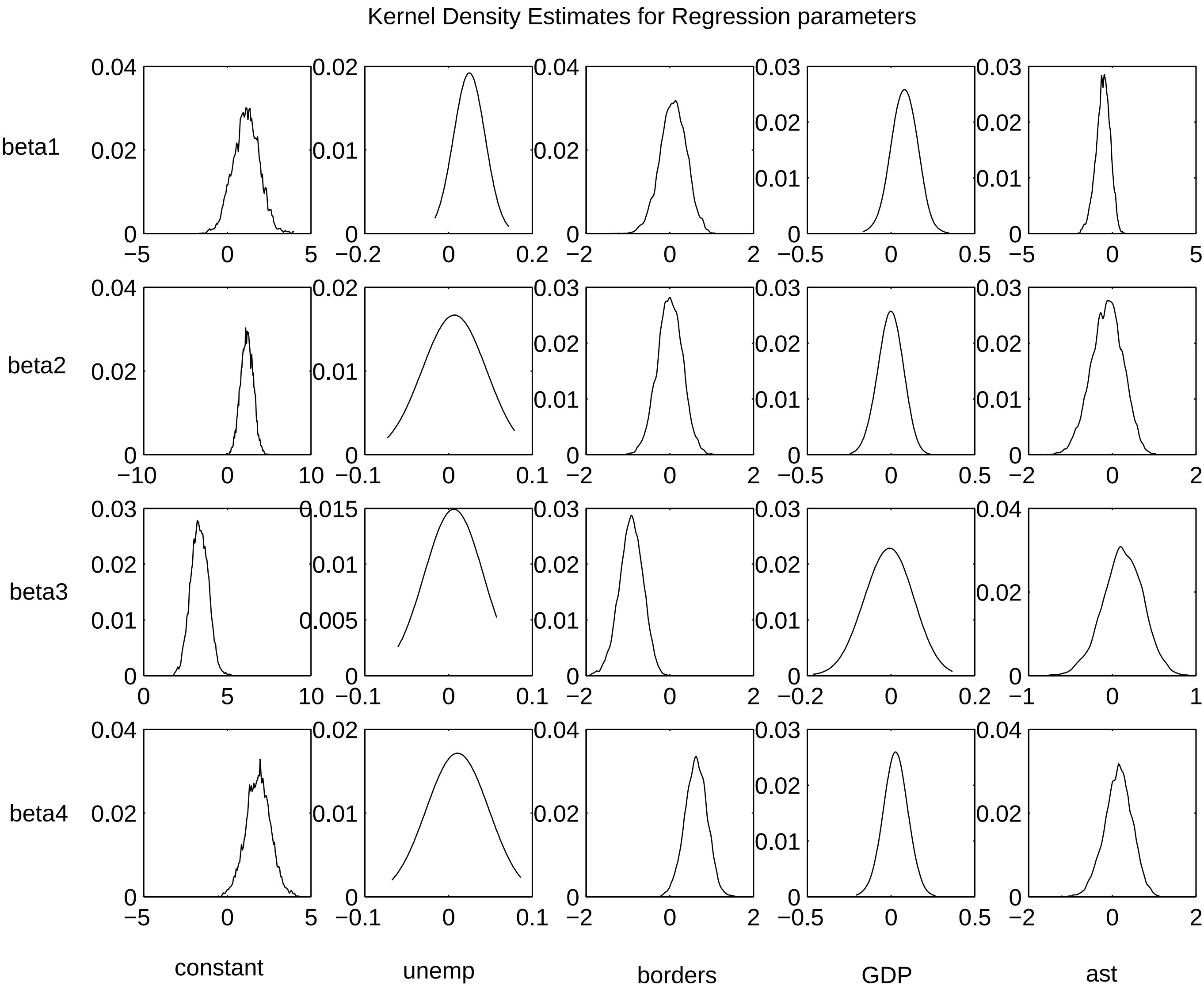
	θ_{12}	θ_{13}	θ_{14}	θ_{23}	θ_{24}	θ_{34}		
	0.8077	4.6115	1.0780	0.0000	0.0000	0.0001		
SE	0.0889	1.1127	0.2077	0.0898	0.1431	0.0892		
				Regression parameters				
	rapes		arsons		manslaughter		smuggling	
	β_1	S.E	β_2	SE	β_3	SE	β_4	SE
constant	0.7416	0.2349	2.5155	0.0306	3.2762	0.0383	1.3909	0.0614
pop	-3.5702	0.4895	-3.9422	0.0559	-2.8082	0.2156	4.7223	0.1354
GDP	0.0632	0.0318	0.0071	0.0357	0.0135	0.0135	0.0150	0.0554
unem	-0.0173	0.0285	-0.0046	0.0249	-1.0521	0.0380	0.8009	0.0408
borders	0.1586	0.0469	0.0440	0.0353	0.0386	0.0122	-0.0103	0.0203
city	-0.1902	0.0412	0.2721	0.0372	0.3750	0.0729	-0.2637	0.0345

Table 3: ML estimates derived via the EM algorithm for the crime data, when covariates were considered for the parameters θ_i, $i = 1, 2, 3, 4$.

Results - Bayesian estimation

	θ_{12}	θ_{13}	θ_{14}	θ_{23}	θ_{24}	θ_{34}		
	0.8845	4.7425	1.5177	0.0595	0.0543	0.1505		
SE	0.80573	2.02699	1.16568	0.17804	0.15937	0.39825		
				Regression parameters				
	rapes		arsons		manslaughter		smuggling	
	β_1	S.E	β_2	SE	β_3	SE	β_4	SE
constant	1.1975	0.86960	2.5197	0.75372	3.4696	0.52593	1.5347	0.83199
pop	-1.4777	3.00626	-3.6553	2.52460	-1.9301	1.68167	4.5748	2.66657
GDP	0.0482	0.02236	0.0055	0.02236	0.0053	0.01732	0.0123	0.02236
unem	-0.0003	0.34699	-0.0468	0.31289	-0.9884	0.28844	0.7077	0.29614
borders	0.0885	0.08185	0.0372	0.07348	0.0103	0.05000	-0.0274	0.08246
city	-0.3948	0.57079	0.2083	0.44215	0.3331	0.30463	-0.2892	0.44710

Table 4: Posterior summaries for the parameters of the model with covariates


Kernel Density Estimates for Regression parameters
beta1
beta2
beta3
beta4
constant
unemp
borders
GDP
ast

Mixtures of multivariate Poisson distribution

Several different ways to define such mixtures:

- Assume

$$\begin{aligned}(X_1, \ldots, X_m) &\sim m - Poisson(\alpha\boldsymbol{\theta}))\\ \alpha &\sim G(\alpha)\end{aligned}$$

- Assume

$$\begin{aligned}(X_1, \ldots, X_m) &\sim m - Poisson(\boldsymbol{\theta}))\\ \boldsymbol{\theta} &\sim G(\boldsymbol{\theta})\end{aligned}$$

- Part of the vector $\boldsymbol{\theta}$ varies, while some of he parameters remain constant. For example (in 2 dimensions)

$$\begin{aligned}(X_1, X_2) &\sim Biv.Poisson(\theta_1, \theta_2, \theta_0)\\ \theta_1, \theta_2 &\sim G(\theta_1, \theta_2)\end{aligned}$$

Dependence Structure

Consider the case

$$\begin{aligned}(X_1, \ldots, X_m \mid \boldsymbol{\theta}) &\sim m - Poisson(\boldsymbol{\theta}))\\ \boldsymbol{\theta} &\sim G(\boldsymbol{\theta})\end{aligned}$$

The unconditional covariance matrix is given by

$$Var(X) = \mathbf{ADA}^T$$

where $\mathbf{A}$ is the matrix used to construct the conditional variates from the original independent Poisson ones and

$$\mathbf{D} = \begin{bmatrix} Var(\theta_1) + E(\theta_1) & Cov(\theta_1, \theta_2) & \dots & Cov(\theta_1, \theta_m) \\ Cov(\theta_1, \theta_2) & Var(\theta_2) + E(\theta_2) & \dots & Cov(\theta_2, \theta_m) \\ & \dots & & \\ Cov(\theta_1, \theta_m) & \dots & & Var(\theta_m) + E(\theta_m) \end{bmatrix}$$

Example

$$
\begin{aligned}
(X_1, X_2) &\sim Biv.Poisson(\theta_1, \theta_2, \theta_3) \\
\theta_1, \theta_2, \theta_3 &\sim G(\theta_1, \theta_2, \theta_3)
\end{aligned}
$$

So,

$$Cov(X_1, X_2) = Var(\theta_3) + E(\theta_3) + Cov(\theta_1, \theta_2) + Cov(\theta_1, \theta_3) + Cov(\theta_2, \theta_3)$$

so, if initially the variables are uncorrelated, i.e. $\theta_3 = 0$ we have that

$$Cov(X_1, X_2) = Cov(\theta_1, \theta_2)$$

Important findings

Remark 1: The above formula imply that if the mixing distribution allows for any kind of covariance between the θ's then the resulting unconditional variables are correlated. Even in the case that one starts with independent Poisson variables the mixing operation can lead to correlated variables.

Remark 2: More importantly, if the covariance between the pairs (θ_i, θ_j) is negative the unconditional variables may exhibit negative correlation. It is well known that the multivariate Poisson distribution cannot have negative correlations, this is not true for its mixtures.

Remark 3: The covariance matrix of the unconditional random variables are simple expressions of the covariances of the mixing parameters and hence the moments of the mixing distribution. Having fitted a multivariate Poisson mixture model, one is able to estimate consistently the reproduced covariance structure of the data. This may serve as a goodness of fit index.

Finite Mixtures of multivariate Poisson distribution

If we assume that $\boldsymbol{\theta}$ can take only a finite number of different values finite multivariate Poisson mixture arise. The pf is given as

$$P(\boldsymbol{X}) = \sum_{j=1}^{k} p_j P(\boldsymbol{X} \mid \boldsymbol{\theta}_j)$$

where $P(\boldsymbol{X} \mid \boldsymbol{\theta}_j)$ denotes the pf of a multivariate Poisson distribution.

this model can be used for clustering multivariate count data Examples:

- Cluster customers of a shop according to their purchases in a series of different products
- Cluster areas according to the number of occurrences of different types of a disease etc

Interesting things

- Standard model-based clustering procedures can be applied. for example, estimation is feasible via EM algorithm, selection of the number of components can be used in a variety of criteria etc
- Since, mixing operation imposes structure is not a good idea to start with a model with a lot of covariance terms.
- Since we work with counts one may use the frequency table instead of the original observations. This speeds up the process and the computing time is not increased so much even if the sample size increases dramatically

Summary

- Multivariate Poisson model similar in nature to multivariate normal were considered.
- The model can be generalized to have quite large (but unnecessary) structure.
- Using up to 2-way covariance term suffice to describe most data sets
- Estimation can be accomplished via EM algorithms (or MCMC schemes from the Bayesian perspective)
- Multivariate Poisson regression models as well as multivariate clustering can be applied through these models

Open problem -Future and Ongoing research

- Need to speed up estimation, including quick calculation of the probabilities and improving the EM algorithm.
- Model selection procedure must be obtained that are suitable for the kind of data (e.g. selection of appropriate covariance terms)
- Finite mixtures of multivariate Poisson regressions
- Bayesian estimation for the finite mixture of multivariate Poisson model including selection for the number of components.