M2 PMA – Elsa TEULIERE – 2017-2018

Estimation of Poisson log-normale distributions using composite likelihood

1 Model and notations

We have n observations of a sample with the covariables, $(x^i)_{i\in\{1,...,n\}}$ and the data we want to analyse $(Y^i)_{i\in\{1,...,n\}}$. For each i, $Y^i \in \mathbb{N}^n$ and $x^i \in \mathbb{R}^p$

1.1 Poisson log-normale distribution

We suppose that the data follow a Poisson-log-normale distribution with parameters mu,Σ [1].
This means that given $Z^i \sim \mathcal{N}_d(0,\Sigma)$, for all $j \in \{1,...,d\}, Y^i_j \mid Z^i_j \sim \mathcal{P}(e^{x^i\mu_j + Z^i_j})$. All the dependency between the variables is encoded in the matrix Σ.
We can remind here some results about the Poisson-log-normale distribution :

1. The density of a Poisson-log-normale distribution :
   For $y \in \mathbb{N}^d$, $p_{(\mu,\Sigma|X)}(y) = \int_{(R)^d} \prod_{j=1}^d f_{exp(X\mu_j+\theta_j)}(y_j) g_{0,\Sigma}(\theta)\mathrm{d}\theta$ with X the covariables, $f_{\exp(X\mu_j+\theta_j)}$ the density function of a Poisson variable of parameters $\exp(X\mu_j+\theta_j)$ and $g_{0,\Sigma}(\theta)$ the density distribution of a $\mathcal{N}_d(0,\Sigma)$.

2. The moments of a Poisson-log-normale distribution :
   $\mathrm{E}(Y^i_j) = exp(x^i\mu_j + \frac{1}{2}\sigma_j^2)$
   $\mathrm{Var}(Y^i_j) = exp(x^i\mu_j + \frac{1}{2}\sigma_j^2) + (exp(x^i\mu_j + \frac{1}{2}\sigma_j^2))^2(exp(\sigma_j^2) - 1)$
   $(cov)(Y^i_j, Y^i_k) = exp(x^i\mu_j + \frac{1}{2}\sigma_j^2)exp(x^i\mu_k + \frac{1}{2}\sigma_k^2)(exp(\sigma_{jk} - 1)$.

1.2 Composite likelihood and M-estimators

1.2.1 Composite likelihood

We use the marginal composite likelihood, also called peudo-likelihood (see [4]).In our case we have $\mathcal{L}_C(Y^1,...,Y^n,\mu,\sigma \mid X) = \prod_{i=1}^n \prod_{k<j} p_{(\mu,\Sigma|X)}(Y^i_j, Y^i_k)$. As we do with the likelihhod, we can consider the log-composite likelihood : $\mathcal{CL}(Y^1,...,Y^n,\mu,\sigma \mid X) = \sum_{i=1}^n \sum_{k<j} \log(p_{(\mu',\Sigma',|X)}(Y^i_j, Y^i_k)$. Where $\mu' = (\mu_j,\mu_k)$ and $\Sigma' = (\sigma_{l,p})_{(l,p)\in\{j,k\}^2}$. We introduce the notation $\mathrm{H}_{(\mu,\Sigma|X)}(Y^i_j, Y^i_k) = \log(p_{(\mu',\Sigma'|X)}(Y^i_j, Y^i_k)$.

1.2.2 M-estimators theory and composite likelihood

Following the definition given by Van der Vaart [3], the log-composite likelihood can be considered as a M-estimator. We have $\mathrm{M}_n : \theta \mapsto \sum_{i=1}^n \mathrm{m}_n(\theta) = \sum_{k<j} \log(p_{(\mu',\Sigma',|X)}(Y^i_j, Y^i_k)$. We recall here the fundamental theorem of the theory of M-estimators :

Theorem 1. *Let (M_n) be a random sequence of functions and M a determinist set of function. If*

1. $\sup\limits_{\theta\in\Theta} \mid \mathrm{M}_n(\theta) - \mathrm{M}(\theta) \mid \xrightarrow{\mathbb{P}} 0$

2. *the maximum θ^* of M is unique.*

Then any sequence of estimators $\widehat{\theta}_n$ with $\mathrm{M}_n(\widehat{\theta}_n \geq \mathrm{M}_n(\theta^*) - \circ_p(1)$ converges in probability to θ^*.

2 Maximum of the composite likelihood

The composite likelihood allows us to solve the problem separetly for each marginal likelihood. Here we look for the maximum of $\mathrm{H}_{(\mu',\Sigma'|X)}(Y_1,Y_2)$ where Y is distributed as the observations. One have :

$$p_{(\mu',\Sigma'|X)}(Y_1,Y_2) = \int_{\mathbb{R}^2} \frac{\mathrm{e}^{Y_1(X\mu_1+z_1)}}{Y_1!} \frac{\mathrm{e}^{Y_2(X\mu_2+z_2)}}{Y_2!} \frac{\mathrm{e}^{-\exp(X\mu_1+z_1)}\mathrm{e}^{-\exp(X\mu_2+z_2)}}{2\pi\sqrt{\mid \det(\Sigma')\mid}} \mathrm{e}^{-\frac{1}{2}(z_1,z_2)\Sigma'^{-1}(z_1,z_2)^T} \mathrm{d}z_1\mathrm{d}z_2 \tag{1}$$

We note $\mathrm{h}_{(\mu',\Sigma'|X)}(z_1,z_2,Y_1,Y_2) = \frac{\mathrm{e}^{Y_1(X\mu_1+z_1)}}{Y_1!} \frac{\mathrm{e}^{Y_2(X\mu_2+z_2)}}{Y_2!} \frac{\mathrm{e}^{-\exp(X\mu_1+z_1)}\mathrm{e}^{-\exp(X\mu_2+z_2)}}{2\pi\sqrt{|\det(\Sigma')|}} \mathrm{e}^{-\frac{1}{2}(z_1,z_2)\Sigma'^{-1}(z_1,z_2)^T}$

$$\mathrm{h}_{(\mu',\Sigma'|X)}(z_1,z_2,Y_1,Y_2) = f_{\exp(X\mu_1+z_1)}(Y_1) f_{\exp(X\mu_2+z_2)}(Y_2) g_{0,\Sigma}(z_1,z_2) \tag{2}$$

We see here that the dependency in the variables of interest (μ_1, μ_2, $\sigma_{1,1}$, $\sigma_{2,2}$, $\sigma_{1,2}$) is split.

2.1 Using the results of the maximum likelihood for the composite maximum likelihood.

To maximize the composite likelihood, it is sufficient to maximize all the terms separetly, *i.e.* to maximize for each couple $(j,k) \in \{1,...,d\}$ the quantity $\frac{1}{n}\sum_{i=1}^n log(p_{(\mu,\Sigma|X)}(Y_j^i,Y_k^i))$. We observe that this quantity is just a regular maximum likelihood. We show that for each couple of variables the maximum-likelihood estimator is consistent. We then use this result in order to show the consistence of the estimator of the composite likelihood.

2.1.1 For one couple of variables

Consider a couple of random variables (Y_j,Y_k) of a vector of random variables $Y = (Y_i)_{i\in\{1...n\}}$ following a $\mathcal{PLN}(\mu,\Sigma)$-distribution, with $\mu = (\mu_i)_{i\in\{1...n\}}$ and $\Sigma = (\sigma_{il})_{(i,l)\in\{1...n\}^2}$. This couple is distributed as a $\mathcal{PLN}(\mu_{jk},\Sigma_{jk})$ law, with $\mu_{jk} = (\mu_j,\mu_k)$ and $\Sigma_{jk} = (\sigma_{jk})$. Since we are considering here covariables, it is a bit more complicated, and we note $\mathcal{PLN}_X(\mu,\Sigma)$ the Poisson-log-normale distribution given the vector of covariables X and $p_(\mu,\Sigma \mid X)$ its density function.

Theorem 2. *For each $(j,k) \in \{1...n\}^2$, $j \neq k$, the estimator $(\widehat{\mu}_n^{jk},\widehat{\Sigma}_n^{jk})$ constructed by maximizing the log-likelihood of the couple $(Y_j^i,Y_k^i)_{i\in\{1...n\}}$, given by $\sum_{i=1}^n \log(p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j^i,Y_k^i))$, is a consistent estimator of the correlation coefficient μ_{jk} and of the vector of variance-covariance Σ_{jk}.*

Proof. The idea of this proof is to use the theorem 1 for our composite likelihood. The proof will follow four steps to verify that the assumptions of the theorem are verified. We will first show the convergence in probability of our M-estimator to a function $\mathrm{M}_{(\mu_{jk},\Sigma_{jk})}$ then the existence of a unique maximum for this function $\mathrm{M}_{(\mu_{jk},\Sigma_{jk})}$. To do this we need to steps : at first we show that our model is identifiable and then, using the Kullback-divergence, that it has a maximum. Combinig this two steps will allow us to conclude the existence of a unique maximum. In the last part, we conclude, using the theroem 1 on the consistence of the estimator.

Step 1 : Convergence in probability.
Using the large number low, we do have that $\frac{1}{n}\sum_{i=1}^n \log(p_{(\mu,\Sigma|X)}(Y_j^i,Y_k^i))$ converge almost surely and so in probability to $\mathbb{E}_X[\mathbb{E}_{Y|X}[\log(p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j,Y_k))]]$.

Step 2 : Identifiability of the model.
We use the general definition of identifiability [2], namely that the family of probabilities $\mathcal{P} = \{\mathbb{P}_\theta, \theta \in \Theta\}$ is identifiable if the application $\theta \mapsto \mathbb{P}_\theta$ is injective. In our case $\mathcal{P} = \{(\mathcal{PLN}_X(\mu,\Sigma))_{X_in\mathcal{X}}; (\mu,\Sigma) \in \mathcal{M}_{p\times n}(\mathbb{R}) \times \mathcal{M}_n(\mathbb{R})\}$. To show the identifiability of the model we will use the fact that two variables

having the same distribution have the same moment. Consider $(\mu_{jk}, \Sigma_{jk}) \in \mathcal{M}_{p\times 2}(\mathbb{R}) \times \mathcal{M}_2(\mathbb{R})$ and $(\mu'_{jk}, \Sigma'_{jk}) \in \mathcal{M}_{p\times 2}(\mathbb{R}) \times \mathcal{M}_2(\mathbb{R})$, so that for every vector of covariables $X \in \mathcal{X}, \mathcal{PLN}_X(\mu_{jk}, \Sigma_{jk}) \sim \mathcal{PLN}_X(\mu'_{jk}, \Sigma'_{jk})$. We have :

$$\mathbb{E}_{(\mu_{jk},\Sigma_{jk}|X)}[Y_j] = \mathbb{E}_{(\mu'_{jk},\Sigma'_{jk}|X)}[Y_j] \tag{3}$$

$$\mathrm{Var}_{(\mu_{jk},\Sigma_{jk}|X)}[Y_j] = \mathrm{Var}_{(\mu'_{jk},\Sigma'_{jk}|X)}[Y_j] \tag{4}$$

$$\mathrm{Cov}_{(\mu_{jk},\Sigma_{jk}|X)}[Y_j, Y_k] = \mathrm{Cov}_{(\mu'_{jk},\Sigma'_{jk}|X)}[Y_j, Y_jk] \tag{5}$$

Using the formula of this moment we recall in the introduction, we find that $\sigma_{jj} = \sigma'_{jj}$, $\sigma_{jk} = \sigma'_{jk}$ and $X(\mu_j - \mu'_j) = 0$. The last condition has to be true for any vector of covariable X. We deduce from this that, we expect $\mathcal{X}^{\perp} = \{0\}$ for the model to be identifable, *ie.* that $\mathrm{rg}(\mathcal{X}) = p$.

Step 3 : Existence of a unique maximum for the $\mathrm{M}(\mu_{jk}, \Sigma_{jk})$-function.
What we consider is nothing else than the log-likelihood of two variables with a Poisson-log-normale distribution. We just shew that the model is identifiable. So there exist one and only one set of parameters $(\mu^*_{jk}, \Sigma^*_{jk})$ such that $(Y_j, Y_k) \sim \mathcal{PLN}_X(\mu^*_{jk}, \Sigma^*_{jk})$. We first can show that maximizing the log-likelihood is equivalent to minimize the Kullback divergence. We recall that the Kullback divergence for top distribution of density functions p_θ a,d p_{θ^*} is given by $\mathrm{D_{KL}}(p_{\theta^*} \parallel p_\theta) = \int p_{\theta^*}(x)\log(\frac{p_{\theta^*}(x)}{p_\theta(x)})\mathrm{d}x$. Since we have $\mathbb{E}_X[\mathbb{E}_{(\mu^*,\Sigma^*|X)}[\log(p_{(\mu^*_{jk},\Sigma^*_{jk}|X)}(Y_j, Y_k)) - \log(p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j, Y_k))]] = \mathbb{E}_X[\mathrm{D_{KL}}(p_{(\mu^*_{jk},\Sigma^*_{jk}|X)} \parallel p_{(\mu_{jk},\Sigma_{jk}|X)})]$. Since the Kullback-divergence is always positive, we see that maximizing the log-likelihood is equivalent to minimize the Kullback-divergence. The Kullback-divergence is only zero if the two distributions are the same. So we see that the M function is maximum for $(\mu_{jk}, \Sigma_{jk}) = (\mu^*_{jk}, \Sigma^*_{jk})$. So we can conclude that the function M only has one maximum, and this maximum is obtained for the parameters we are interested in.

Step 4 : Conclusion.
To summurize we have :

1. For any set of parameters (μ_{jk}, Σ_{jk}),
   $| \frac{1}{n}\sum_{i=1}^n \log(p_{(\mu,\Sigma|X)}(Y^i_j, Y^i_k)) - \mathbb{E}_X[\mathbb{E}_{Y|X}[\log(p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j, Y_k))]] | \xrightarrow{\mathbb{P}} 0$
2. The function $(\mu_{jk}, \Sigma_{jk}) \mapsto \mathbb{E}_X[\mathbb{E}_{Y|X}[\log(p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j, Y_k))]]$ only has one maximum for $(\mu_{jk}, \Sigma_{jk}) = (\mu^*_{jk}, \Sigma^*_{jk})$, the parameters of the Poisson-log-Normale distribution of the variables (Y_i, Y_j).

Applying theorem 1, we conclude that the sequence of estimators $(\widehat{\mu}^{jk}_n, \Sigma^{jk}_n)_{n\in\mathbb{N}}$ converges in probability to what we hope to approximate, namely $(\mu^*_{jk}, \Sigma^*_{jk})$. □

2.1.2 Generalization to the composite likelihood.

What we wnat to show is the consistence of the estimators obtained by maximizing the composite likelihood. We recall that the composite likelihood is given by : $\mathcal{CL}_{(\mu,\Sigma)|X}(Y^i) = \sum_{i=1}^n \sum_{j<k} log(p_{(\mu_{jk},\Sigma_{jk}|X)}(Y^i_j, Y^i_k)$.

Theorem 3. *The estimator $(\widehat{\mu}_n, \widehat{\Sigma}_n)$ constructed by maximizing the composite likelihood for $(Y_i)_{i\in\{1\ldots n\}}$ is a consistent estimator of the correlation coefficients μ and of the vector of variance-covariance Σ.*

Proof. The proof follows exactly the same path that we did for a couple of variables. We will again use theorem 1 and the results of theorem 2.

Step 1 : Convergence in probability. Again using the large number law we have :
$\frac{1}{n}\sum_{i=1}^n \sum_{j<k} log(p_{(\mu_{jk},\Sigma_{jk}|X)}(Y^i_j, Y^i_k) \xrightarrow{\mathbb{P}} \sum_{j<k} \mathbb{E}_X[\mathbb{E}_{Y|X}[\log(p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j, Y_k))]]$.
We note M the function $\mathrm{M} : (\mu, \Sigma) \mapsto \sum_{j<k} \mathbb{E}_X[\mathbb{E}_{Y|X}[\log(p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j, Y_k))]]$.

Step 2 : Identifiability of the model.
The family of model we consider here is $\mathcal{P} = \{(\mathcal{PLN}_X(\mu, \Sigma))_{X\in\mathcal{X}}; \mu \in \mathcal{M}_{pn}(\mathbb{R}), \Sigma \in \mathcal{M}_n(\mathbb{R})\}$. Using

the moment as we did with a Poisson log-normale distribution of a vector of two variables, but this time for n variables is sufficient to show the identifiability of the model.

Step 3 : Existence and uniqueness of the maximum for the function M.
Again here we have by the identifiability of the model, under the assumption that the data follow a Poisson log-normale distribution, that there exist one and only one set of parameters (μ^*, Σ^*) such as given a vector of covariables X, $(Y^i) \sim \mathcal{PLN}_X(\mu^*, \Sigma^*)$. As we did in the previous section we show that maximizing M is equivalent to minimize the function :
$(\mu, \Sigma) \mapsto \sum_{j<k} \mathbb{E}_X[\mathrm{D_{KL}}(p_{(\mu^*_{jk}, \Sigma^*_{jk}|X)} \parallel p_{(\mu_{jk}, \Sigma_{jk}|X)}]$.
We have here a finite sum of positives variables, so in order to minimize it, we can minimize each of the terms. So the above function is minimal for $\mu^* = (\mu^*_j)_{j\in\{1\ldots n\}}$ and Σ^* defined as follow : the term $(\sigma_{jk})_{j\neq k}$ of Σ^* is the term σ_{12} of Σ^*_{jk} and the term σ_{jj} of Σ^* is the term σ_{11} of Σ^*_{jk}.
Thus we have the existence of a maximum and its uniqueness.

Step 4 : Conclusion.
Applying theorem 1, we conclude that the sequence of estimators $(\widehat{\mu}_n, \widehat{\Sigma}_n)_{n\in\mathbb{N}}$ maximizing the composite likelihood function $(\mu, \Sigma) \mapsto \mathcal{CL}_{(\mu,\Sigma)|X}((Y^i)_{i\in\{1..n\}})$ converges in probability to what we hope to approximate, namely (μ^*, Σ^*), the parameters of our Poisson log-normale model. □

3 Approximation of the estimators

We now know that the estimators we consider are consitent, so it now worth to calculate them. To do so, we will use numerical approximations and so we have to calculate their gradients. As we saw previously, in order to estimate the maximum of the compsite likelihood, it is sufficient to estimate the maximum of the log-likelihood for each couple of variables. In order to do this, we have to calculate the gradient of the esitimators. Given a couple (j, k), we will calculate the derivative of it with respect to all the parameters.

3.1 With respect to the vector μ.

Let consider $i \in \{1...n\}$. We want to calculate $\partial_{\mu_i}\log(p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j, Y_k))$.

$$\partial_{\mu_i}\log(p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j, Y_k)) = \frac{\partial_{\mu_i}p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j, Y_k)}{p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j, Y_k)}$$

Since :

$$\begin{aligned}\partial_{\mu_i}p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j, Y_k) &= \mathbb{E}_{(Z_j,Z_k)}[\partial_{\mu_i}p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j, Y_k \mid Z_j, Z_k)] \\ &= \mathbb{E}_{(Z_j,Z_k)}[\partial_{\mu_i}p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j \mid Z_j)p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_k \mid Z_k)]\end{aligned}$$

if $i \neq j$ and $i \neq k$, this partial derivatives is equal to 0. Otherwise, taking for example $i = j$, one have :

$$\begin{aligned}\partial_{\mu_j}p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j, Y_k) &= \mathbb{E}_{(Z_j,Z_k)}[\partial_{\mu_j}p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j \mid Z_j)p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_k \mid Z_k)] \\ &= \mathbb{E}_{(Z_j,Z_k)}[X(Y_j p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j \mid Z_j) - (Y_j+1)p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j+1 \mid Z_j))p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_k \mid Z_k)] \\ &= X(Y_j p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j, Y_k) - (Y_j+1)p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j+1, Y_k))\end{aligned}$$

Where the second equality comes from the fact that $Y_j \mid Z_j$ is following a Poisson-distribution. We have :

$$\begin{aligned}\partial_{\mu_j}p_{\mu_j,\Sigma_{jj}|X}(Y_j \mid Z_j) &= \partial_{\mu_j}(\mathrm{e}^{Y_j(X\mu_j+Z_j)}\exp(-e^{X\mu_j+Z_j})\frac{1}{Y_j!}) \\ &= XY_j\frac{\mathrm{e}^{Y_j(X\mu_j+Z_j)}\exp(-e^{X\mu_j+Z_j})}{Y_j!} - X\frac{\mathrm{e}^{(Y_j+1)(X\mu_j+Z_j)}\exp(-e^{X\mu_j+Z_j})}{Y_j!} \\ &= X(Y_j p_{(\mu_j,\Sigma_{jj}|X)}(Y_j \mid Z_j) - (Y_j+1)p_{(\mu_j,\Sigma_{jj}|X)}(Y_j+1 \mid Z_j))\end{aligned}$$

So we conclude that

$$\partial_{\mu_i}\log(p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j,Y_k)) = X(Y_j - (Y_j+1)\frac{p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j+1,Y_k)}{p_{(\mu_{jk},\Sigma_{jk}|X)}(Y_j,Y_k))})$$

3.2 With respect to the matrix of variance - covariance.

We note $\sigma_{i,j}$ the terms of the matrix of varianve covariance. The term of this matrix are only present in the density function of the multivariate normale distribution wth which the parameters of the Poissons are taken. In order to calcul the derivatives we will in both cases follow the same path :

1. We calculate the derivatives of the density function of the bivariate normale distribution wrt the parameter of interest. By integrating under the $\int$ we deduce the derivative of the Poisson-log-normale density function.

2. We integrate by part to have a nice expression of this derivative.

We recall here that the density function of the Poisson log-Normal distribution with mean μ and matrix of variance-covariance Σ, given a vector of covariables X is given by :

$$p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j,y_k) = \int_{\mathbb{R}^2} p_{e^{X\mu_{jk}+z_1}}(y_j)p_{e^{X\mu_{jk}+z_2}}(y_k)h_{(0,\Sigma_{jk})}(z_j,z_k)\mathrm{d}z_j\mathrm{d}z_k$$

With $h_{(0,\Sigma_{jk})}$ the density function of the bivariate normale distribution of mean 0 and variance-covariance Σ_{jk} and p_β the density function of a Poisson distribution with parameters β. We finally recall that : $h_{(0,\Sigma_{jk})}(z_1,z_2) = \frac{1}{2\pi\sqrt{|\sigma_{jj}\sigma_{kk}-\sigma_{jk}^2|}}\mathrm{e}^{-\frac{1}{2}\frac{\sigma_{jj}z_1^2+\sigma_{kk}z_2^2-2\sigma_{jk}z_1z_2}{\sigma_{jj}\sigma_{kk}-\sigma_{jk}^2}}$.

3.2.1 wrt σ_{jj} and σ_{kk}

$$\partial_{\sigma_{jj}}h_{(0,\Sigma_{jk})}(z_1,z_2) = \frac{1}{2}[\frac{-\sigma_{kk}}{\sigma_{jj}\sigma_{kk}-\sigma_{jk}^2} + \frac{(\sigma_{jk}z_1-\sigma_{kk}z_2)^2}{(\sigma_{jj}\sigma_{kk}-\sigma_{jk}^2)^2}]h_{(0,\Sigma_{jk})}(z_1,z_2)$$

Since the bivariate normale distribution admit second order moments, we can apply the thoerem of derivation under $\int$ and deduce that :

$$\partial_{\sigma_{jj}}p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j,y_k) = \int_{\mathbb{R}^2} p_{e^{X\mu_{jk}+z_1}}(y_j)p_{e^{X\mu_{jk}+z_2}}(y_k)\frac{1}{2}[\frac{-\sigma_{kk}}{\sigma_{jj}\sigma_{kk}-\sigma_{jk}^2} + \frac{(\sigma_{jk}z_1-\sigma_{kk}z_2)^2}{(\sigma_{jj}\sigma_{kk}-\sigma_{jk}^2)^2}]h_{(0,\Sigma_{jk})}(z_j,z_k)\mathrm{d}z_j\mathrm{d}z_k$$

We will now try to find a nicer expression of this derivative. One can remark that $\partial_{z_2}(-\frac{1}{2}\frac{\sigma_{jj}z_1^2+\sigma_{kk}z_2^2-2\sigma_{jk}z_1z_2}{\sigma_{jj}\sigma_{kk}-\sigma_{jk}^2}) = \frac{\sigma_{jk}z_1-\sigma_{kk}z_2}{\sigma_{jj}\sigma_{kk}-\sigma_{jk}^2}$.

This gives the clue to integrate by part with respect to z_2. We then have :

$$\begin{aligned}\partial_{\sigma_{jj}}p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j,y_k) &= \int_{\mathbb{R}}[p_{e^{X\mu_{jk}+z_1}}(y_j)p_{e^{X\mu_{jk}+z_2}}(y_k)\frac{\sigma_{jk}z_1-\sigma_{kk}z_2}{\sigma_{jj}\sigma_{kk}-\sigma_{jk}^2}h_{(0,\Sigma_{jk})}(z_1,z_2)]_{-\infty}^{+\infty}\mathrm{d}z_1 \\ &- \int_{\mathbb{R}^2} y_k p_{e^{X\mu_{jk}+z_1}}(y_j)p_{e^{X\mu_{jk}+z_2}}(y_k)\frac{\sigma_{jk}z_1-\sigma_{kk}z_2}{\sigma_{jj}\sigma_{kk}-\sigma_{jk}^2}h_{(0,\Sigma_{jk})}(z_1,z_2)\mathrm{d}z_1\mathrm{d}z_2 \\ &- \int_{\mathbb{R}^2}(y_k+1)p_{e^{X\mu_{jk}+z_1}}(y_j)p_{e^{X\mu_{jk}+z_2}}(y_k+1)\frac{\sigma_{jk}z_1-\sigma_{kk}z_2}{\sigma_{jj}\sigma_{kk}-\sigma_{jk}^2}h_{(0,\Sigma_{jk})}(z_1,z_2)\mathrm{d}z_1\mathrm{d}z_2\end{aligned}$$

The derivatives of the density function of the Poisson distribution is obtained as we did in the previous section.

The first rht of this equality being null, we can again integrate by part wrt z_2 and we get :

$$\begin{aligned}\partial_{\sigma_{jj}}p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j,y_k) &= y_k^2 p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j,y_k) \\ &- ((y_k+1)^2 + y_k(y_k+1))p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j,y_k+1) \\ &+ (y_k+1)(y_k+2)p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j,y_k+2)\end{aligned}$$

We are calculating this gradients in order to approximate the composite log likelihood so we are in fact interested in $\partial_{\sigma_{jj}} \log p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k)$, which we now can easily calculate.

$$\begin{aligned}\partial_{\sigma_{jj}} p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k) = y_k^2 - ((y_k+1)^2 + y_k(y_k+1))\frac{p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k+1)}{p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k)} \\ + (y_k+1)(y_k+2)\frac{p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k+2)}{p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k)}\end{aligned}$$

Symetrically we have wrt σ_{kk} :

$$\begin{aligned}\partial_{\sigma_{kk}} p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k) = y_j^2 - ((y_j+1)^2 + y_j(y_j+1))\frac{p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j+1, y_k)}{p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k)} \\ + (y_j+1)(y_j+2)\frac{p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j+2, y_k)}{p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k)}\end{aligned}$$

3.2.2 wrt σ_{jk} :

$$\partial_{\sigma_{jk}} h_{(0,\Sigma)}(z_1, z_2) = [\frac{\sigma_{jk}}{\sigma_{jj}\sigma_{kk} - \sigma_{jk}^2} + \frac{-\sigma_{jk}(\sigma_{jj}z_1^2 + \sigma_{kk}z_2^2) + \sigma_{jk}z_1z_2 + \sigma_{jj}\sigma_{kk}z_1z_2}{(\sigma_{jj}\sigma_{kk} - \sigma_{jk}^2)^2}]h_{(0,\Sigma)}(z_1, z_2)$$

Again we can apply the theorem of derivating under the $\int$. We deduce that :

$$\begin{aligned}\partial_{\sigma_{jk}} p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k) = \int_{\mathbb{R}^2} p_{e^{X\mu_{jk}+z_1}}(y_j) p_{e^{X\mu_{jk}+z_2}}(y_k) \\ [\frac{\sigma_{jk}}{\sigma_{jj}\sigma_{kk} - \sigma_{jk}^2} + \frac{-\sigma_{jk}(\sigma_{jj}z_1^2 + \sigma_{kk}z_2^2) + \sigma_{jk}z_1z_2 + \sigma_{jj}\sigma_{kk}z_1z_2}{(\sigma_{jj}\sigma_{kk} - \sigma_{jk}^2)^2}]h_{(0,\Sigma_{jk})}(z_j, z_k)\mathrm{d}z_j\mathrm{d}z_k\end{aligned}$$

One can note that :

$$\begin{aligned}\partial_{z_1}(-\frac{1}{2}\frac{\sigma_{jj}z_1^2 + \sigma_{kk}z_2^2 - 2\sigma_{jk}z_1z_2}{\sigma_{jj}\sigma_{kk} - \sigma_{jk}^2})\partial_{z_2}(-\frac{1}{2}\frac{\sigma_{jj}z_1^2 + \sigma_{kk}z_2^2 - 2\sigma_{jk}z_1z_2}{\sigma_{jj}\sigma_{kk} - \sigma_{jk}^2}) = \\ \frac{-\sigma_{jk}(\sigma_{jj}z_1^2 + \sigma_{kk}z_2^2) + \sigma_{jk}z_1z_2 + \sigma_{jj}\sigma_{kk}z_1z_2}{(\sigma_{jj}\sigma_{kk} - \sigma_{jk}^2)^2}\end{aligned}$$

Which gives a clue that we could do again tow part integrations : one towards z_1 and one towards z_2. We will do a first integration by part with respect to z_1. We note that : $\partial_{z_1}(\frac{\sigma_{jk}z_1 - \sigma_{kk}z_2}{\sigma_{jj}\sigma_{kk} - \sigma_{jk}^2}h_{(0,\Sigma_{jk})}(z_1, z_2) = [\frac{\sigma_{jk}}{\sigma_{jj}\sigma_{kk} - \sigma_{jk}^2} + \frac{-\sigma_{jk}(\sigma_{jj}z_1^2 + \sigma_{kk}z_2^2) + \sigma_{jk}z_1z_2 + \sigma_{jj}\sigma_{kk}z_1z_2}{(\sigma_{jj}\sigma_{kk} - \sigma_{jk}^2)^2}]h_{(0,\Sigma)}(z_1, z_2)$. And taking the derivative of the poisson density with respect to its parameters allows us to conclude that :

$$\begin{aligned}\partial_{\sigma_{jk}} p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k) = \int_{\mathbb{R}} [p_{e^{X\mu_{jk}+z_1}}(y_j) p_{e^{X\mu_{jk}+z_2}}(y_k)\frac{\sigma_{jk}z_1 - \sigma_{kk}z_2}{\sigma_{jj}\sigma_{kk} - \sigma_{jk}^2}h_{(0,\Sigma_{jk})}(z_1, z_2)]_{-\infty}^{+\infty}\mathrm{d}z_2 \\ - \int_{\mathbb{R}^2}(y_j p_{e^{X\mu_{jk}+z_1}}(y_j) - (y_j+1)p_{e^{X\mu_{jk}+z_1}}(y_j+1))p_{e^{X\mu_{jk}+z_2}}(y_k)\frac{\sigma_{jk}z_1 - \sigma_{kk}z_2}{\sigma_{jj}\sigma_{kk} - \sigma_{jk}^2}h_{(0,\Sigma_{jk})}(z_1, z_2)\mathrm{d}z_1\mathrm{d}z_2\end{aligned}$$

The first right-hand term of the equality being equal to zero, we have, integrating by part again but this time wrt z_2, we have :

$$\begin{aligned}\partial_{\sigma_{jk}} p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k) = \int_{\mathbb{R}^2}(y_j p_{e^{X\mu_{jk}+z_1}}(y_j) - (y_j+1)p_{e^{X\mu_{jk}+z_1}}(y_j+1)) \\ (y_k p_{e^{X\mu_{jk}+z_2}}(y_k) - (y_k+1)p_{e^{X\mu_{jk}+z_2}}(y_k+1))h_{(0,\Sigma_{jk})}(z_1, z_2)\mathrm{d}z_1\mathrm{d}z_2\end{aligned}$$

What we can also write :

$$\begin{aligned}\partial_{\sigma_{jk}} p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k) = & y_k y_j p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k) - (y_k+1)y_j p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k+1) \\ & - (y_j+1)y_k p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k) + (y_k+1)(y_j+1)p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j+1, y_k+1))\end{aligned}$$

Again we are interested in $\partial_{\sigma_{jk}} \log p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k)$ which we can now easily calculate by dividing the above quantity by $p_{(\mu_{jk},\Sigma_{jk}|X)}(y_j, y_k)$.

References

[1] John Aitchison and CH Ho. The multivariate poisson-log normal distribution. *Biometrika*, 76(4):643–653, 1989.

[2] Vincent Rivoirard and Gilles Stoltz. *Statistique mathématique en action*. Vuibert, 2012.

[3] A. W. van der Vaart. *M- and Z-Estimators*, page 41–84. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[4] Cristiano Varin, Nancy Reid, and David Firth. An overview of composite likelihood methods. *Statistica Sinica*, pages 5–42, 2011.