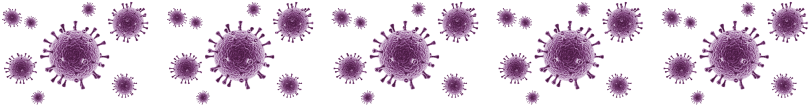
Modèles Linéaires Appliqués / Régression
Modèle de Poisson : Méthode des Marges

Arthur Charpentier

UQAM

Hiver 2020 - COVID-19 # 10



Méthode des Marges

Supposons que l'on prenne en compte ici deux classes de risques.
Tableau de contingence et biais minimial, Bailey (1963) et Mildenhall (1999)
On suppose que

$$N_{i,j} \sim \mathcal{P}(L_i C_j), \text{ i.e. } \boldsymbol{N} \sim \mathcal{P}(\boldsymbol{L}\boldsymbol{C}^\top)$$

voire

$$N_{i,j} \sim \mathcal{P}(e_{i,j} L_i C_j), \text{ i.e. } \boldsymbol{N} \sim \mathcal{P}(\boldsymbol{e}\boldsymbol{L}\boldsymbol{C}^\top)$$

L'estimation de $\boldsymbol{L} = (L_i)$ et de $\boldsymbol{C} = (C_j)$ se fait généralement de trois manières: par moindres carrés, par minimisation d'une distance (e.g. du chi-deux) ou par un principe de balancement (ou méthode des marges).

Méthode des Marges

Il est possible d'utiliser une méthode par moindres carrés (pondérée). On va chercher à minimiser la somme des carrés des erreurs, i.e.

$$D = \sum_{i,j} e_{i,j}(n_{i,j} - L_i C_j)^2$$

La condition du premier ordre donne ici

$$\frac{\partial D}{\partial L_i} = -2 \sum_j C_j e_{i,j}(n_{i,j} - L_i C_j) = 0$$

$$L_i = \frac{\sum_j C_j e_{i,j} n_{i,j}}{\sum_j e_{i,j} C_j^2} = \frac{\sum_j c_j y_{i,j}}{\sum_j e_{i,j} C_j^2}$$

L'autre condition du premier ordre donne

$$C_j = \frac{\sum_i L_i e_{i,j} n_{i,j}}{\sum_i e_{i,j} L_i^2} = \frac{\sum_i L_i y_{i,j}}{\sum_i e_{i,j} L_i^2}$$

On résoud alors ce petit systeme de maniere itérative (car il n'y a pas de solution analytique simple).

Méthode des Marges

Il est aussi possible d'utiliser une méthode basée sur la distance du chi-deux. On va chercher à minimiser

$$Q = \sum_{i,j} \frac{e_{i,j}(n_{i,j} - L_i C_j)^2}{L_i C_j}$$

Là encore on utilise les conditions du premier ordre, et on obtient

$$L_i = \left(\frac{\sum_j \left(\frac{e_{i,j} y_{i,j}^2}{C_j} \right)}{\sum_j e_{i,j} C_j} \right)^{\frac{1}{2}} \text{ et } L_j = \left(\frac{\sum_i \left(\frac{e_{i,j} y_{i,j}^2}{L_i} \right)}{\sum_i e_{i,j} L_i} \right)^{\frac{1}{2}}$$

où $y_{i,j} = e_{i,j} n_{i,j}$ (que l'on résout itérativement).

Méthode des Marges

Dans la méthode des marges – Bailey (1963) – formellement, on veut

$$\sum_j y_{i,j} = \sum_j e_{i,j} n_{i,j} = \sum_j e_{i,j} L_i C_j,$$

en somment sur la ligne i, pour tout i, ou sur la colonne j,

$$\sum_i y_{i,j} = \sum_i e_{i,j} n_{i,j} = \sum_i e_{i,j} L_i C_j,$$

La première, et la seconde, équation donnent respectivemnet

$$L_i = \frac{\sum_j y_{i,j}}{\sum_j e_{i,j} C_j} \text{ et } C_j = \frac{\sum_i y_{i,j}}{\sum_i e_{i,j} L_i}.$$

Cette solution... correspond à la régression de Poisson.