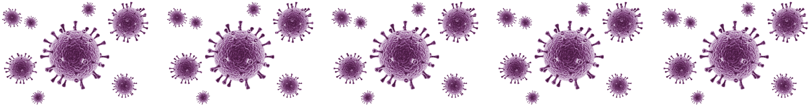
Modèles Linéaires Appliqués / Régression
Modèle de Poisson

Arthur Charpentier

UQAM

Hiver 2020 - COVID-19 # 7

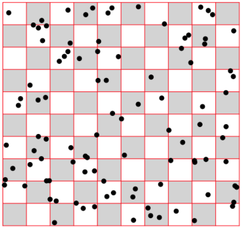


La loi des petits nombres

La loi de Poisson apparaît comme approximation de la loi binomiale quand $p \sim \lambda/n$

$$\binom{n}{k} p^k (1-p)^{n-k} \sim e^{-\lambda} \frac{\lambda^k}{k!}$$

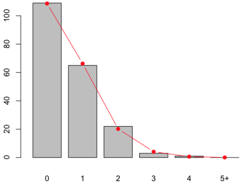
```
> data.frame(N,F=table(nb_cell),P=c
    (dpois(0:4,1),1-ppois(4,1)))
     N  F      P
1    0 36  36.78
2    1 39  36.78
3    2 16  18.39
4    3  7   6.13
5    4  2   1.53
6   5+  0   0.37
```



La loi des petits nombres

Nombre de soldats de cavaliers morts par ruade de cheval, entre 1875 et 1894, dans 10 corps (soit 200 corps annuels) Bortkiewicz (1898)

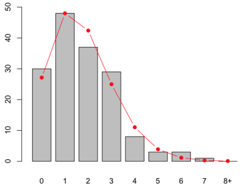
```
> data.frame(N,F=table(ruades),P=c(dpois(0:4,mean(ruades)),1-ppois(4,mean(ruades))))
    N   F        P
1   0 109   108.67
2   1  65    66.21
3   2  22    20.22
4   3   3     4.11
5   4   1     0.63
6  5+   0     0.08
```



La loi des petits nombres

Nombre d'ouragans, par an, Lévi & Partrat (1989)

```
> data.frame(N,F=table(hurricanes),
   P=c(dpois(0:4,mean(hurricanes))
   ,1-ppois(4,mean(hurricanes))))
     N   F       P
1    0  30   27.16
2    1  48   47.99
3    2  37   42.41
4    3  29   24.98
5    4   8   11.03
6    5   3    3.90
7    6   3    1.15
8    7   1    0.29
9   8+   0    0.08
```

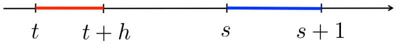


La loi des petits nombres

$$\mathbb{P}(N = y) = e^{-\lambda}\frac{\lambda^y}{y!}, \forall y \in \mathbb{N}.$$

Pour rappel, $\mathbb{E}(N) = \text{Var}(N) = \lambda \in \mathbb{R}_+$ (équidispersion)

$(N_t)_{t\geq 0}$ est un processus de Poisson homogène (de paramètre λ) s'il est à accroissements indépendants, et le nombre de sauts observés pendant la période $[t, t+h]$ suit une loi $\mathcal{P}(\lambda \cdot h)$.



$N_{s+1} - N_s \sim \mathcal{P}(\lambda)$ est indépendant de $N_{t+h} - N_t \sim \mathcal{P}(\lambda \cdot h)$.
Soit N_i la frénquence annulisée de sinistre pour l'assuré i, et supposons $N_i \sim \mathcal{P}(\lambda)$. Si l'assuré i a été observé pendant une période e_i, le nombre de sinistre observé est $Y_i \sim \mathcal{P}(\lambda \cdot e_i)$.

Maximum de Vraisemblance

Données $\{y_i\}$

$$\mathcal{L}(\lambda, \boldsymbol{y}) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$$

$$\log \mathcal{L}(\lambda, \boldsymbol{y}) = -\lambda + \sum_{i=1}^{n} y_i \log[\lambda] - \log \left(\prod_{i=1}^{n} y_i! \right)$$

qui donne la condition du premier ordre

$$\frac{\partial}{\partial \lambda} \log \mathcal{L}(\lambda, \boldsymbol{y}) = -n + \frac{1}{\lambda} \sum_{i=1}^{n} y_i \quad (= 0)$$

qui s'annule pour

$$\widehat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} y_i = \overline{y}$$

Maximum de Vraisemblance

De plus

$$\frac{\partial^2}{\partial\lambda^2}\log\mathcal{L}(\lambda,\boldsymbol{y}) = -\frac{1}{\lambda^2}\sum_{i=1}^n y_i \quad (<0)$$

(la log-vraisemblance est concave, $\widehat{\lambda}$ est bien un maximum)
L'information de Fisher est

$$I(\lambda) = -\mathbb{E}\left(\frac{\partial^2}{\partial\lambda^2}\log f_\lambda(Y)\right) = \frac{\mathbb{E}(Y)}{\lambda^2} = \frac{1}{\lambda}$$

$$I_n(\lambda) = -\mathbb{E}\left(\frac{\partial^2}{\partial\lambda^2}\log\mathcal{L}(\lambda,\boldsymbol{y})\right) = \frac{1}{\lambda^2}\sum_{i=1}^n\mathbb{E}(Y_i) = \frac{n}{\lambda} = n\cdot I(\lambda)$$

et $\sqrt{n}(\widehat{\lambda}-\widehat{\lambda}) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\lambda)$ quand $n\to\infty$.

Maximum de Vraisemblance

Données $\{(y_i, e_i)\}$

$$\mathcal{L}(\lambda, \boldsymbol{y}, \boldsymbol{e}) = \prod_{i=1}^{n} \frac{e^{-\lambda e_i}[\lambda e_i]^{y_i}}{y_i!}$$

$$\log \mathcal{L}(\lambda, \boldsymbol{y}, \boldsymbol{e}) = -\lambda \sum_{i=1}^{n} e_i + \sum_{i=1}^{n} y_i \log[\lambda e_i] - \log\left(\prod_{i=1}^{n} y_i!\right)$$

qui donne la condition du premier ordre

$$\frac{\partial}{\partial \lambda} \log \mathcal{L}(\lambda, \boldsymbol{y}, \boldsymbol{e}) = -\sum_{i=1}^{n} e_i + \frac{1}{\lambda} \sum_{i=1}^{n} y_i i \quad (= 0)$$

qui s'annule pour

$$\widehat{\lambda} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} e_i} = \sum_{i=1}^{n} \omega_i \frac{y_i}{e_i} \text{ avec } \omega_i = \frac{e_i}{\sum_{i=1}^{n} e_i}$$