

ECN 6338 Cours 3, annexe

Version avec le modèle poissonien

William McCausland

2025-01-27

Éléments de l'analyse maximum de vraisemblance

- ▶ Quantités pertinentes :
 - ▶ θ , un vecteur de paramètres inconnus,
 - ▶ $y = (y_1, \dots, y_T)$, un vecteur aléatoire des variables observables,
 - ▶ y° , le vecteur observé.
- ▶ Fonctions pertinentes :
 - ▶ $f(y|\theta)$, la densité conditionnelle des données (modèle),
 - ▶ $\mathcal{L}(\theta; y) = f(y|\theta)$, la vraisemblance,
 - ▶ $\mathcal{L}(\theta; y^\circ) = f(y^\circ|\theta)$, la vraisemblance réalisée.

Le modèle poissonien

- Supposez que les y_i sont iid Poisson avec moyenne $\theta > 0$.
- La fonction de masse de probabilité de y_i est

$$f(y_i|\theta) = e^{-\theta} \frac{\theta^{y_i}}{y_i!}.$$

- On observe le vecteur aléatoire $y = (y_1, \dots, y_n)$; la fonction de masse de probabilité de y est

$$f(y|\theta) = \prod_{i=1}^n f(y_i|\theta) = \prod_{i=1}^n e^{-\theta} \frac{\theta^{y_i}}{y_i!} = \left[\prod_{i=1}^n \frac{1}{y_i!} \right] e^{-n\theta} \theta^{\sum_{i=1}^n y_i}.$$

- Pour simplifier un facteur qui importe peu,

$$c \equiv \left[\prod_{i=1}^n \frac{1}{y_i!} \right].$$

Deux interprétations de la même expression

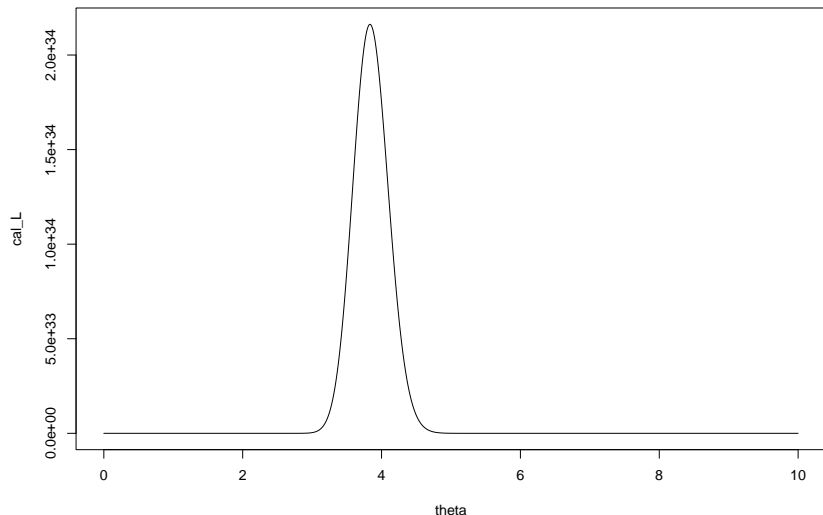
- ▶ L'expression :

$$f(y|\theta) = ce^{-n\theta} \theta^{\sum_{i=1}^n y_i} = \mathcal{L}(\theta; y).$$

- ▶ Deux interprétations :
 - ▶ Fonction de masse de probabilité $f(y|\theta)$.
 - ▶ Fonction de vraisemblance $\mathcal{L}(\theta; y)$.
- ▶ $f(y|\theta)$ donne, pour θ fixe, les probabilités relatives des séquences possibles (y_1, \dots, y_n) .
- ▶ $\mathcal{L}(\theta; y)$ donne, pour y fixe (notamment $y = y^\circ$) une note (ou évaluation) à chaque valeur θ pour la qualité de sa prévision des données observées.
- ▶ Soit $L(\theta; y) = \log \mathcal{L}(\theta; y)$, la log-vraisemblance.

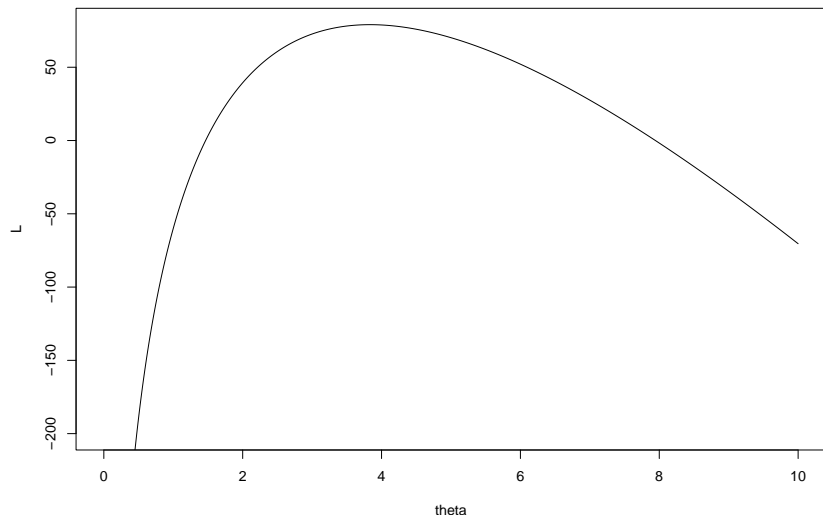
Vraisemblance poissonnienne pour $n = 60$, $\sum_{i=1}^n y_i = 230$

```
n = 60; somme_y = 230; theta = seq(0, 10, by=0.001)
cal_L = exp(-n*theta) * theta^somme_y
plot(theta, cal_L, type='l')
```



Log vraisemblance poissonnienne, $n = 60$, $\sum_{i=1}^n y_i = 230$

```
L = -n*theta + somme_y*log(theta)
plot(theta, L, type='l', ylim=c(-200, max(L)))
```



Maximum de la vraisemblance poissonienne

- Vraisemblance : $\mathcal{L}(\theta; y) = ce^{-n\theta} \theta^{\sum_{i=1}^n y_i}$.
- Log vraisemblance : $L(\theta; y) = \log c - n\theta + (\sum_{i=1}^n y_i) \log \theta$.
- Deux dérivées de la log vraisemblance :

$$\frac{\partial L(\theta; y)}{\partial \theta} = -n + \frac{\sum_{i=1}^n y_i}{\theta}$$

$$\frac{\partial^2 L(\theta; y)}{\partial \theta^2} = -\frac{\sum_{i=1}^n y_i}{\theta^2} < 0.$$

- La valeur $\hat{\theta}$ (souvent vue comme une variable aléatoire) qui maximise la vraisemblance et la log-vraisemblance est

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n y_i.$$

- Pour $n = 60$ et $\sum_{i=1}^n y_i = 230$, $\hat{\theta} = \frac{23}{6} \approx 3.833$.

Maximum de vraisemblance : conditions de régularité

- ▶ Définitions :
 - ▶ θ est le vecteur des paramètres ; Θ , l'ensemble de toutes les valeurs possibles de θ .
 - ▶ y est le vecteur (aléatoire) des données.
- ▶ Conditions informelles de régularité :
 1. Le modèle est correct pour une valeur $\theta = \theta_0 \in \Theta$.
 2. La vraie valeur θ_0 est dans l'intérieur de Θ .
 3. Identification :

$$\theta \neq \theta_0 \Rightarrow f(\cdot|\theta) \neq f(\cdot|\theta_0).$$

4. $L(\theta; y) \equiv \log f(y|\theta)$ a toujours un maximum global unique.
5. Le gradient de $L(\theta; y)$ (par rapport à θ) est toujours borné.
6. La matrice $\mathcal{I}(\theta)$ suivante (matrice d'information de Fisher) est définie positive:

$$\mathcal{I}(\theta) \equiv E_{y|\theta} \left[\frac{\partial L(\theta; y)}{\partial \theta^\top} \frac{\partial L(\theta; y)}{\partial \theta} \right].$$

Maximum de vraisemblance : résultats

Résultats : (Soit $\hat{\theta} \equiv \arg \max_{\theta} L(\theta; y)$, qui existe et est unique.)

1. $\hat{\theta} \rightarrow_p \theta_0$ (loi de grands nombres)
2. $\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow_d N(0, n\mathcal{I}(\theta_0)^{-1})$ (théorème central limite)
3. $\mathcal{I}(\theta) = E_{y|\theta} \left[-\frac{\partial^2 L(\theta; y)}{\partial \theta \partial \theta^\top} \right]$.

Problèmes restants :

1. Il faut trouver $\hat{\theta}$.
2. La variance asymptotique $\mathcal{I}(\theta_0)^{-1}$ de $\hat{\theta}$ dépend de θ_0 , qui est inconnu.
3. L'espérance dans les deux expressions pour $\mathcal{I}(\theta)$ sont difficiles à évaluer analytiquement.

Exemple poissonien

- ▶ Un cas rare où les calculs analytiques sont faisables.
- ▶ La matrice d'information de Fisher : ($E[y_i] = \theta$, $\text{Var}[y_i] = \theta$)

$$\mathcal{I}(\theta) = E_{y|\theta} \left[-\frac{\partial^2 L}{\partial \theta^2} \right] = E_{y|\theta} \left[\frac{\sum_{i=1}^n y_i}{\theta^2} \right] = \frac{n\theta}{\theta^2} = \frac{n}{\theta}.$$

- ▶ La variance de $\hat{\theta}$ (exacte, pas asymptotique) :

$$\text{Var}[\hat{\theta}] = \text{Var} \left[\frac{\sum_{i=1}^n y_i}{n} \right] = \frac{1}{n^2} n \text{Var}[y_i] = \frac{\theta}{n}.$$

- ▶ La variance dépend de θ inconnu, mais elle est souvent peu sensible à θ : pour $n = 60$, $\text{Var}[\hat{\theta}]$ est de $(0.2528)^2$ pour $\theta = 23/6 \approx 3.833$, $(0.2236)^2$ pour $\theta = 3$ et $(0.2739)^2$ pour $\theta = 4.5$.

Éléments de l'analyse bayésienne

- ▶ Quantités pertinentes :
 - ▶ θ , un vecteur de paramètres inconnus *aléatoire*
 - ▶ $y = (y_1, \dots, y_T)$, un vecteur aléatoire des variables observables,
 - ▶ y° , le vecteur observé.
- ▶ Fonctions pertinentes :
 - ▶ $f(y|\theta)$, la densité conditionnelle des données (modèle),
 - ▶ $\mathcal{L}(\theta; y^\circ) = f(y^\circ|\theta)$, la vraisemblance réalisé,
 - ▶ $f(\theta)$, la densité *a priori*,
 - ▶ $f(\theta, y)$, la densité conjointe,
 - ▶ $f(\theta|y)$, la densité *a posteriori*,
 - ▶ $f(y)$, la densité marginale des données,
 - ▶ $f(y^\circ)$, la vraisemblance marginale (un nombre).

Inférence bayésienne

- ▶ Par la règle de Bayes,

$$f(\theta|y^\circ) = \frac{f(\theta, y^\circ)}{f(y^\circ)} = \frac{f(\theta)f(y^\circ|\theta)}{f(y^\circ)} \propto f(\theta)f(y^\circ|\theta).$$

- ▶ $f(\theta)$ représente notre incertitude sur θ avant l'observation de y .
- ▶ $f(\theta|y^\circ)$ représente notre incertitude sur θ après qu'on observe $y = y^\circ$.
- ▶ Un point important à retenir : $f(\theta|y^\circ) \propto f(\theta, y^\circ)$.

Reprise et extension de l'exemple poissonien

- ▶ Si $y_i \sim \text{iid Po}(\theta)$, $f(y|\theta) = c e^{-n\theta} \theta^{\sum_{i=1}^n y_i}$.
- ▶ Mettons qu'on choisit la loi *a priori* $\theta \sim \text{Ga}(\alpha, \beta)$ sur $[0, \infty)$:

$$f(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}.$$

- ▶ La densité conjointe est

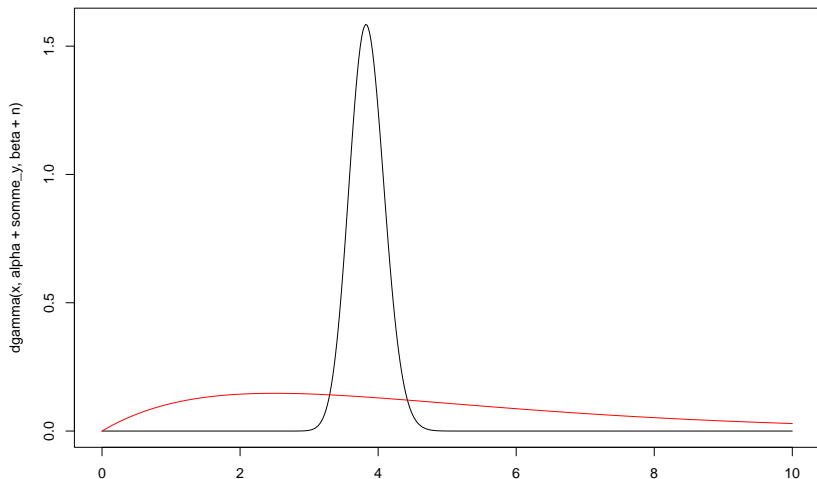
$$f(\theta, y) = f(\theta)f(y|\theta) = c \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \theta^{\alpha + \sum_{i=1}^n y_i - 1} \cdot e^{-(\beta + n)\theta}.$$

- ▶ La loi *a posteriori* doit être $\theta \sim \text{Ga}(\alpha + \sum_{i=1}^n y_i, \beta + n)$.
- ▶ La vraisemblance marginale est $f(\theta, y)/f(\theta|y)$:

$$f(y) = c \frac{\beta^\alpha}{(\beta + n)^{\alpha + \sum_{i=1}^n y_i}} \cdot \frac{\Gamma(\alpha + \sum_{i=1}^n y_i)}{\Gamma(\alpha)}.$$

Graphique pour l'exemple poissonien

```
n = 60; somme_y = 230; alpha=2; beta=0.4  
x = seq(0, 10, by=0.02)  
plot(x, dgamma(x, alpha+somme_y, beta+n), type='l')  
lines(x, dgamma(x, alpha, beta), col='red')
```



Exemple gaussien

- Considérez le modèle $y_t \sim \text{iid } N(\mu, h^{-1})$.
- Le vecteur de paramètres est $\theta = (\mu, h)$.
- Le vecteur d'observables est $y = (y_1, \dots, y_T)$.
- La densité des données est

$$\begin{aligned} f(y|\theta) &= \prod_{t=1}^T \sqrt{\frac{h}{2\pi}} \exp \left[-\frac{h}{2} (y_t - \mu)^2 \right] \\ &= \left(\frac{h}{2\pi} \right)^{T/2} \exp \left[-\frac{h}{2} \sum_{t=1}^T (y_t - \mu)^2 \right]. \end{aligned}$$

Exemple gaussien (suite)

- ▶ Mettons qu'on choisit une loi *a priori* où h et μ sont indépendents, avec

$$\mu \sim N(\bar{\mu}, \bar{\omega}^{-1}), \quad \bar{s}^2 h \sim \chi^2(\bar{\nu}),$$

où $\bar{\mu}$, $\bar{\omega}$, \bar{s} et $\bar{\nu}$ sont des hyperparamètres constants choisis par l'investigateur.

- ▶ La densité *a priori* est

$$f(\theta) \propto \exp \left[-\frac{\bar{\omega}}{2} (\mu - \bar{\mu})^2 \right] \cdot h^{(\bar{\nu}-2)/2} \exp \left[-\frac{1}{2} \bar{s}^2 h \right].$$

- ▶ La densité conjointe est

$$f(\theta, y) \propto h^{(\bar{\nu}+T-2)/2} \exp \left[-\frac{\bar{\omega}}{2} (\mu - \bar{\mu})^2 - \frac{h}{2} \left(\bar{s}^2 + \sum_{t=1}^T (y_t - \mu)^2 \right) \right].$$

L'intégration et les objectifs de l'analyse bayésienne

- ▶ Plusieurs problèmes d'inférence bayésienne ont, comme solution, une intégrale par rapport à la densité *a posteriori*.
- ▶ Exemple 1, estimation ponctuelle de θ_k sous perte quadratique:

$$\hat{\theta}_k = E[\theta_k | y^\circ] = \int \theta_k f(\theta | y^\circ) d\theta.$$

- ▶ Exemple 2, quantification de l'incertitude sur θ_k :

$$\text{Var}[\theta | y^\circ] = E[(\theta_k - E[\theta_k | y^\circ])^2 | y^\circ].$$

- ▶ Exemple 3, densité prédictive (valeurs de y_{T+1} sur une grille) :

$$f(y_{T+1} | y^\circ) = E[f(y_{T+1} | \theta, y^\circ) | y^\circ].$$

Preuve de l'exemple 3

$$\begin{aligned} E[f(y_{T+1}|y_1, \dots, y_T, \theta)|y_1, \dots, y_T] \\ &= \int f(y_{T+1}|y_1, \dots, y_T, \theta) f(\theta|y_1, \dots, y_T) d\theta \\ &= \int f(y_{T+1}, \theta|y_1, \dots, y_T) d\theta \\ &= f(y_{T+1}|y_1, \dots, y_T) \end{aligned}$$

Méthodes pour trouver $E[g(\theta)|y^\circ]$

- ▶ Calcul analytique : élégant, exacte, presque toujours insoluble.
- ▶ Simulation Monte Carlo indépendante :
 - ▶ Si on peut simuler $\theta^m \sim \text{iid } \theta|y^\circ$,

$$\frac{1}{M} \sum_{m=1}^M g(\theta^m) \rightarrow_p E[g(\theta)|y^\circ].$$

- ▶ Cependant, cette simulation est rarement faisable.
- ▶ Simulation Monte Carlo chaîne de markov (MCMC) :
 - ▶ On choisit un processus markovien avec densité de transition $f(\theta^m|\theta^{m-1})$ telle que la loi *a posteriori* $\theta|y^\circ$ est la loi stationnaire du processus. C'est à dire :

$$\theta^{m-1} \sim f(\theta|y^\circ) \Rightarrow \theta^m \sim f(\theta|y^\circ).$$

- ▶ Sous quelques conditions techniques, la loi de θ^m converge à la loi *a posteriori* et

$$\frac{1}{M} \sum_{m=1}^M g(\theta^m) \rightarrow_p E[g(\theta)|y^\circ].$$