MAP433 Statistique
Maximum de vraisemblance

Christophe Giraud

CMAP, Ecole Polytechnique

PC3

Le maximum de vraisemblance

L'estimateur du **maximum de vraisemblance est central** en statistiques, car:

- naturel et "systématique"
- propriétés d'optimalité
- invariant par reparamétrisation

L'intuition (1/2)

Observations: n comptages d'oiseaux $\mathrm{x}_1, \ldots, \mathrm{x}_n$.

Modèle: $X_1, \ldots, X_n$ i.i.d. de loi de Poisson de paramètre θ inconnu.

Estimation: On veut estimer θ par la valeur pour laquelle la probabilité des observations est la plus grande:

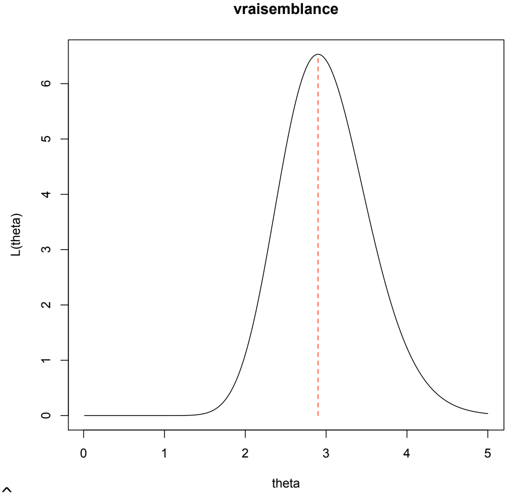
$$\boxed{\hat{\theta} \quad \text{maximise} \quad \mathbb{P}_\theta(X_1 = \mathrm{x}_1, \ldots, X_n = \mathrm{x}_n) =: L_{\mathrm{x}}(\theta)}$$

Calcul:

$$\begin{aligned} L_{\mathrm{x}}(\theta) &= \frac{e^{-\theta}\theta^{\mathrm{x}_1}}{\mathrm{x}_1!} \times \ldots \times \frac{e^{-\theta}\theta^{\mathrm{x}_n}}{\mathrm{x}_n!} \\ &= \frac{e^{-n\theta}\,\theta^{\mathrm{x}_1+\ldots+\mathrm{x}_n}}{\mathrm{x}_1!\ldots\mathrm{x}_n!} \end{aligned}$$

L'intuition (2/2)

$\texttt{data:}\ (\mathtt{x}_1, \ldots, \mathtt{x}_n) = (6, 4, 2, 2, 1, 5, 2, 0, 4, 3)$



on obtient $\hat{\theta} = 2.9$

Formalisation

Ecueil: dans le cas de variables continues

$$\mathbb{P}_\theta(X_1 = \mathrm{x}_1, \ldots, X_n = \mathrm{x}_n) = 0 \quad \text{pour tout } \theta \text{ !}$$

<u>ex:</u> $X_1, \ldots, X_n$ i.i.d. de loi $\mathcal{N}(\theta, 1)$

Notion adaptée : la densité
On suppose

$$d\mathbb{P}_\theta^X(x_1, \ldots, x_n) = f(\theta, x_1, \ldots, x_n)\, d\mu(x_1, \ldots, x_n)$$

avec μ mesure σ-finie **ne dépendant pas de** θ.

En général: $\mu =$ mesure de comptage ou mesure de Lebesgue

Vraisemblance

$$L_{\mathrm{x}} : \theta \to L_{\mathrm{x}}(\theta) := f(\theta, \underbrace{\mathrm{x}_1, \ldots, \mathrm{x}_n}_{\text{observations}})$$

Maximum de vraisemblance

$$\hat{\theta} = \operatorname*{argmax}_{\theta} L_{\mathrm{x}}(\theta)$$

Cas i.i.d.

Si $X_1, \ldots, X_n$ i.i.d. de loi $d\mathbb{P}_\theta^{X_i}(x) = p(\theta, x)\, d\mu(x)$, on a:

$$\begin{aligned} L_{\mathrm{x}}(\theta) &= p(\theta, \mathrm{x}_1) \times \ldots \times p(\theta, \mathrm{x}_n) \\ \text{et} \quad \hat{\theta} &= \operatorname*{argmax}_{\theta} \sum_{i=1}^{n} \log(p(\theta, \mathrm{x}_i)) \end{aligned}$$

Exemple

Observations: n mesures d'erreurs $\mathrm{x}_1, \dots, \mathrm{x}_n$.

Modèle: $X_1, \dots, X_n$ i.i.d. de loi de Laplace de paramètre θ inconnu.

$$p(\theta, x) = \frac{\theta}{2} e^{-\theta|x|} \quad \text{et} \quad d\mu(x) = dx.$$

Calcul:

$$\begin{aligned} \log(L_{\mathrm{x}}(\theta)) &= \sum_{i=1}^{n} \log(p(\theta, \mathrm{x}_i)) \\ &= -\theta \sum_{i=1}^{n} |\mathrm{x}_i| + n \log(\theta/2) \end{aligned}$$

maximum en

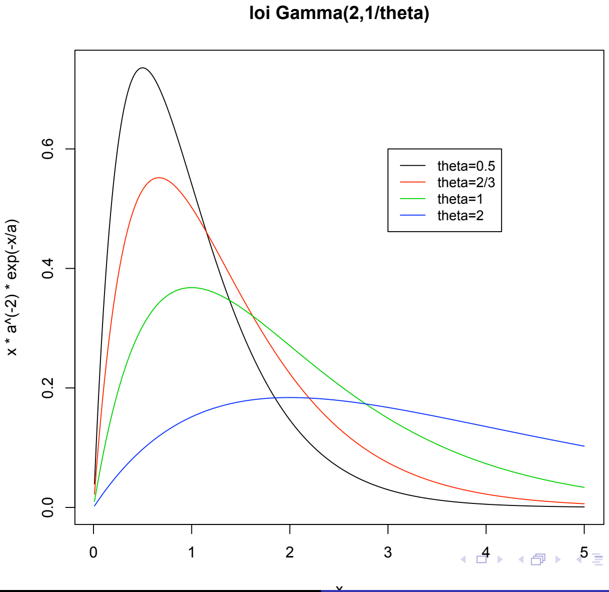
$$\hat{\theta} = \left(\frac{1}{n} \sum_{i=1}^{n} |\mathrm{x}_i| \right)^{-1}$$

Programme PC3

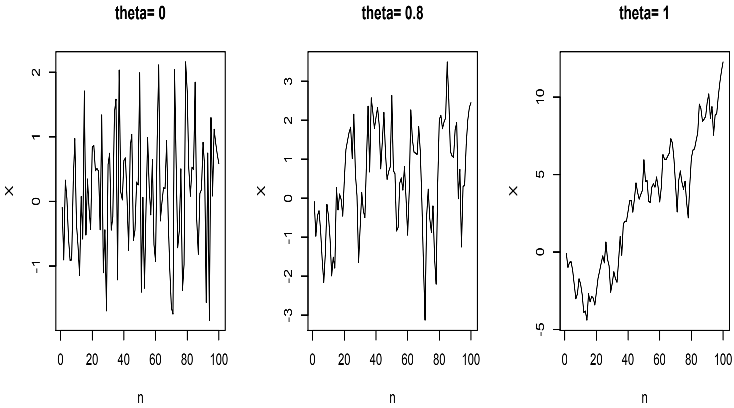
Programme aujourd'hui:
estimation par maximum de vraisemblance

1. cas continu
2. modèle exponentiel
3. cas non i.i.d. et vectoriel
4. cas discret

Loi Gamma$(2, 1/\theta)$



Modèle Autorégressif



Avec $\sigma^2 = 1$.