Paramètre unidimensionnel

Le modèle Poisson

$E_X = \mathbb{N}$, $\mu = \sum_{k\in\mathbb{N}} \delta_k$, $\Theta =]0, +\infty[$ et on a

$$L(\theta, x) = \exp(-\theta)\frac{\theta^x}{x!} = \exp(-\theta)\frac{1}{x!}\exp(x\ln\theta) = \beta(\theta)\xi(x)\exp\left(t(x)\alpha(\theta)\right)$$

$\beta(\theta) = \exp(-\theta), \alpha(\theta) = \ln\theta, \quad \xi(x) = 1/x!, t(x) = x$

Le modèle Gamma

$\theta = (a, \lambda) \in]0, +\infty[^2$ et μ la mesure de Lebesgue sur $E_X =]0, +\infty[$:

$$\forall x \in]0, \infty[, \quad L((a,\lambda), x) = \frac{\lambda^a}{\Gamma(a)}\exp(-\lambda x)x^{a-1} = \beta(\theta)\xi(x)\exp\left(\langle t(x), \alpha(\theta)\rangle\right)$$

$\beta((a,\lambda)) = \dfrac{\lambda^a}{\Gamma(a)}, \alpha((a,\lambda)) = (-\lambda, a-1), \quad \xi(x) = 1, t(x) = (x, \ln x)$

Paramètre multidimensionnel

Le modèle normal $\mathcal{N}(m, \sigma^2)$

$\theta = (m, \sigma^2) \in \mathbb{R}\times]0, \infty[$ et μ mesure le Lebesgue sur $\mathbb{R}$: $\forall x \in \mathbb{R}$

$$\begin{aligned} L((m,\sigma^2), x) &= \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left(-\frac{1}{2\sigma^2}(x-m)^2\right) \\ &= \beta(\theta)\xi(x)\exp\left(\langle t(x), \alpha(\theta)\rangle\right) \end{aligned}$$

$$\beta((m,\sigma^2)) = \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left(-\frac{m^2}{2\sigma^2}\right)$$

$$\alpha((m,\sigma^2)) = \left(\frac{m}{\sigma^2}, -\frac{1}{2\sigma^2}\right)$$

$$\xi(x) = 1 \quad t(x) = (x, x^2)$$

À partir de l'écriture (1) de la vraisemblance, on peut considérer la nouvelle mesure dominante $\nu = \xi.\mu$ et la vraisemblance devient

$$\forall(\theta, x) \in \Theta\times E_X, \quad L(\theta, x) = \frac{\exp\left(\langle t(x), \alpha(\theta)\rangle\right)}{\int_{E_X}\exp\left(\langle t(x), \alpha(\theta)\rangle\right)d\nu}$$

On obtient via une réécriture exponentielle du dénominateur

$$\forall(\theta, x) \in \Theta\times E_X, \quad L(\theta, x) = \exp\Big(\langle t(x), \alpha(\theta)\rangle - \ln c_\nu(\alpha(\theta)\Big)$$

avec

$$c_\nu(a) := \int_{E_X}\exp\left(\langle t(x), a\rangle\right)d\nu$$

Enfin, le reparamétrage $\lambda = \alpha(\theta)$ donne une vraisemblance de la forme :

$$\forall(\lambda, x) \in \alpha(\Theta)\times E_X, \quad L(\lambda, x) = \exp\left(\langle t(x), \lambda\rangle - \ln c_\nu(\lambda)\right)$$

Définition 4 (Forme canonique d'une famille exponentielle)

On dit que la famille $\{\mathbb{P}_\theta, \ \theta\in\Theta\}$ est une famille exponentielle de dimension p relativement à la mesure dominante μ si $\Theta\subset\mathbb{R}^p$ et s'il existe une application mesurable $t : E_X \to \mathbb{R}^p$ telle que

$$\forall\theta\in\Theta \quad c_\mu(\theta) := \int_{E_X}\exp\left(\langle t(x), \theta\rangle\right)d\mu(x) < +\infty$$

et une vraisemblance du modèle statistique s'écrive

$$\forall x \in E_X, \quad L(\theta, x) = \exp\left(\langle t(x), \theta\rangle - \Psi_\mu(\theta)\right)$$

avec $\Psi_\mu(\theta) = \ln c_\mu(\theta)$. L'ensemble $N_\mu := \{a\in\mathbb{R}^p, \ c_\mu(a) < +\infty\}$ est appelé l'ensemble naturel des paramètres (il contient Θ). La famille est dite complète si $N_\mu = \Theta$ et régulière s'il s'agit d'un ouvert.

Remarque 2

1. Le modèle statistique associé à une famille exponentielle (canonique) est automatiquement homogène relativement à la mesure dominante μ.
2. La forme de la vraisemblance d'une loi d'une famille exponentielle canonique relativement à μ implique que le support de la densité $L(\theta, \cdot)$, i.e. $\mathrm{supp}_{L(\theta,\cdot)} := \{x\in E_X \mid L(\theta, x) > 0\}$ doit être indépendant de θ.

Quelques exemples

Pour tous les exemples sous forme naturelle avec une fonction $\xi(x) = 1$, on dispose de la forme canonique via un reparamétrage approprié.

Modèle normal $\mathcal{N}(m, \sigma^2)$: le reparamétrage

$$\theta = (\theta_1, \theta_2) = \left(\frac{m}{\sigma^2}, -\frac{1}{2\sigma^2}\right), \quad \Theta = \mathbb{R} \times]-\infty, 0[$$

et on obtient

$$t(x) = (x, x^2), \quad \Psi_\mu(\theta) = \ln(\sqrt{\pi}) - \frac{1}{2}\left(\frac{{\theta_1}^2}{2\theta_2} + \ln(-\theta_2)\right)$$

Quelques exemples suite

Modèle de Poisson $\mathrm{Pois}(\lambda)$, $\lambda \in]0, +\infty[$:
on avait

$$\beta(\lambda) = \exp(-\lambda), \quad \alpha(\lambda) = \ln\lambda, \quad \xi(x) = 1/x!$$

d'où avec la mesure dominante $\mu = \sum_{x \in \mathbb{N}} \delta_x / x!$ et le reparamétrage $\theta = \ln\lambda$ avec $\Theta = \mathbb{R}$, on a

$$L(\theta, x) = \exp(x\theta - \exp\theta)$$

et $\Psi_\mu(\theta) = \exp\theta$, $t(x) = x$.

Quelques exemples suite

Modèle Binomial $\mathrm{Bin}(n, p)$ avec $p \in]0, 1[$: forme naturelle
Pour $\mu = \sum_{x=0}^n \delta_x$ on a

$$\begin{aligned} L(p, x) &= C_n^x p^x (1-p)^{n-x} \\ &= \exp(n\ln(1-p))\, C_n^x \exp\left(x \ln\left(\frac{p}{1-p}\right)\right) \end{aligned}$$

$\beta(p) = \exp(n\ln(1-p)), \alpha(p) = \ln(p/(1-p)),$
$t(x) = x, \xi(x) = C_n^x$

Forme canonique : on utilise la mesure dominante $\mu = \sum_{x=0}^n C_n^x \delta_x$ et le reparamétrage $\theta = \ln(p/(1-p))$, $\Theta = \mathbb{R}$

$$L(\theta, x) = \exp\left(x\theta - n\ln(1 + \exp\theta)\right)$$

$t(x) = x$ et $\Psi_\mu(\theta) = n\ln(1 + \exp\theta)$

Proposition 1 (admise [DC82])

1. *L'ensemble naturel des paramètres $N_\mu := \{a \in \mathbb{R}^p,\ c_\mu(a) < +\infty\}$ d'une famille exponentielle canonique est un convexe de $\mathbb{R}^p$ et la fonction $\Psi(\cdot)$ est convexe*
2. *Pour tout θ dans l'intérieur de N_μ, la fonction*

$$\Psi_\mu(\theta) = \ln c_\mu(\theta) = \ln \int_{E_X} \exp\left(\langle t(x), \theta\rangle\right) d\mu(x)$$

est indéfiniment dérivable et toute dérivation s'obtient par dérivation sous le signe somme.

Par exemple,

$$\begin{aligned} \nabla\Psi_\mu(\theta) &= \exp(-\Psi(\theta)) \int_{E_X} t(x) \exp\left(\langle t(x), \theta\rangle\right) d\mu(x) \\ &= \int_{E_X} t(x) L(\theta, x) d\mu(x) = \mathbb{E}_\theta[t(X)] \qquad (2) \\ \frac{\partial^2 \Psi_\mu}{\partial\theta_i \partial\theta_j}(\theta) &= \mathrm{Cov}_\theta(t_i(X), t_j(X)) = V_\theta(t(X))_{i,j} \qquad (3) \end{aligned}$$

- **Modèle Poisson :** $t(x) = x$ et $\Psi(\theta) = \exp\theta$ d'où

$$\begin{aligned} \mathbb{E}_\theta[X] &= \Psi'_\mu(\theta) = \exp\theta = \lambda \quad (\text{ car } \theta = \ln\lambda) \\ V_\theta(X) &= \Psi''_\mu(\theta) = \exp\theta = \lambda \end{aligned}$$

- **Modèle Binomial :** $t(x) = x$ et $\Psi(\theta) = n\ln(1+\exp\theta)$

$$\begin{aligned} \mathbb{E}_\theta[X] &= \Psi'_\mu(\theta) = n\frac{\exp\theta}{1+\exp\theta} = np \\ V_\theta(X) &= \Psi''_\mu(\theta) = n\,\frac{\exp\theta}{(1+\exp\theta)^2} \\ &= n\frac{\exp\theta}{1+\exp\theta}\,\frac{1}{1+\exp\theta} = np(1-p) \end{aligned}$$

- **Modèle normal** $\mathcal{N}(m,\sigma^2)$ **:** $t(x) = (t_1(x), t_2(x)) = (x, x^2)$ et $(\theta_1,\theta_2) = \left(\frac{m}{\sigma^2}, -\frac{1}{2\sigma^2}\right)$

$$\Psi_\mu(\theta) = \text{cte} - \frac{1}{2}\left(\frac{{\theta_1}^2}{2\theta_2} + \ln(-\theta_2)\right)$$

d'où avec $\nabla\Psi_\mu(\theta) = \mathbb{E}_\theta[t(X)] = (\mathbb{E}_\theta[t_1(X)], \mathbb{E}_\theta[t_2(X)])$

$$\begin{aligned} \mathbb{E}_\theta[t_1(X)] &= \mathbb{E}_\theta[X] = \frac{\partial\Psi_\mu}{\partial\theta_1} = -\frac{\theta_1}{2\theta_2} = m \\ \mathbb{E}_\theta[t_2(X)] &= \mathbb{E}_\theta[X^2] = \frac{\partial\Psi_\mu}{\partial\theta_2} = -\frac{1}{2}\left(-\frac{{\theta_1}^2}{2{\theta_2}^2} + \frac{1}{\theta_2}\right) \\ &= m^2 + \sigma^2 \end{aligned}$$

Échantillonnage

On a considéré un modèle statistique $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ pour une observation x d'une v.a. X.
En général, un jeu de données est constitué d'un ensemble d'observations $x_1, \ldots, x_n$, qui correspondent à n répétitions de la même expérience aléatoire sous-jacente et donnent ainsi n valeurs observées de X.
Ici, nous considérerons toujours que les répétitions se sont faites de manière indépendantes

Définition 5 (Modèle statistique d'un n-échantillon)

Les données $(x_1, \ldots, x_n)$ sont les valeurs observées de n variables aléatoires indépendantes et identiquement distribuées $(X_1, \ldots, X_n) \in (E_X)^n$, i.e. un n-échantillon de la loi $\mathbb{P}_\theta$. La loi de $(X_1, \ldots, X_n)$ est la mesure produit $\mathbb{P}_\theta^{\otimes n}$. Alors $\left((E_X)^n, \mathscr{E}^{\otimes n}, (\mathbb{P}_\theta^{\otimes n})_{\theta\in\Theta}\right)$ est le modèle statistique d'un n-échantillon

Vraisemblance des observations

Comme nous sommes placés dans un modèle statistique dominé par la mesure μ, on en déduit que le modèle d'un n-échantillon est également dominé par la mesure produit $\mu^{\otimes n}$.
D'où la vraisemblance associée à l'observation $x = (x_1, \ldots, x_n) \in (E_X)^n$ d'un n-échantillon est donnée par :

$$\begin{aligned} &L_n(\theta, x) \\ &= \prod_{i=1}^n \frac{d\mathbb{P}_\theta}{d\mu}(x_i) = \prod_{i=1}^n L_1(\theta, x_i) \\ &= \begin{cases} \prod_{i=1}^n f_{X_i}(\theta, x_i) & E_X \text{ continu et } \mu = \text{mesure de Leb.} \\ \prod_{i=1}^n \mathbb{P}_\theta(\{x_i\}) & E_X \text{ discret et } \mu = \text{mesure de comptage} \\ \equiv \prod_{i=1}^n \mathbb{P}_\theta\{X_i = x_i\} & \end{cases} \end{aligned}$$

n-échantillon d'une famille exponentielle

Si $\{\mathbb{P}_\theta, \theta \in \Theta\}$ forme une famille exponentielle (canonique) de lois de dimension p avec une statistique naturelle $t(x) \in \mathbb{R}^p$ et $\Psi_\mu(\theta)$, alors cela reste vrai pour le modèle associé à un n-échantillon. La statistique naturelle t_n et la fonction $\Psi_{\mu^{\otimes n}}(\theta)$ associées sont :

$$\forall x = (x_1, \ldots, x_n) \in \mathbb{R}^n, \quad t_n(x) = \sum_{i=1}^n t(x_i) \in \mathbb{R}^p \quad \Psi_{\mu^{\otimes n}}(\theta) = n\Psi_\mu(\theta)$$

Par exemple, le modèle normal $\mathcal{N}(m, \sigma^2)$ est de dimension 2 avec $\theta = \left(\frac{m}{\sigma^2}, -\frac{1}{2\sigma^2}\right)$ et $\forall x \in \mathbb{R},\ t(x) = (x, x^2)$. On en déduit :

$$L_n(\theta, x) = \exp\left(\langle t_n(x), \theta\rangle - \Psi_{\mathrm{Leb}^{\otimes n}}(\theta)\right)$$

$$t_n(x) = \left(\sum_{i=1}^n x_i, \sum_{i=1}^n {x_i}^2\right)$$

$$\Psi_{\mathrm{Leb}^{\otimes n}}(\theta) = n\Psi_{\mathrm{Leb}}(\theta) = n\left(\ln(\sqrt{\pi}) - \frac{1}{2}\left(\frac{{\theta_1}^2}{2\theta_2} + \ln(-\theta_2)\right)\right)$$

Information de Fisher

Pour faciliter la prise de décision : introduire des **mesures de discrimination entre deux lois** de la famille $\mathcal{P}$. L'information de Fisher répond à cette attente dans la mesure où l'on cherche le pouvoir discriminant du modèle statistique pour deux valeurs proches de θ.
Dans cette partie, Θ est supposé être un ouvert de $\mathbb{R}^p$ et le modèle $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ est homogène.

Définition 6 (Score)

Pour un modèle statistique $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$, on suppose que pour tout $\theta \in \Theta$ et pour $\mathbb{P}_\theta$-presque tout x, la fonction ***ln-vraisemblance*** *$\theta \mapsto \ln L(\theta, x)$ admette un gradient $\nabla \ln L(\theta, x)$.*
Pour tout $\theta \in \Theta$, on appelle ***score du modèle*** *au point θ, la v.a. $S(\theta)$ définie par $\nabla \ln L(\theta, X)$.*

Pour un paramètre θ unidimensionnel, le score se réduit à la v.a.

$$\frac{d}{d\theta} \ln L(\theta, X) = \frac{\frac{d}{d\theta} L(\theta, X)}{L(\theta, X)}$$

(À partir de l'hypothèse d'homogénéité) on vérifie que le score ne dépend ni de la vraisemblance, ni de la mesure dominante μ.

Définition 7 (Information de Fisher)

En reprenant les notations introduites dans la définition 6, on suppose que la fonction de score soit de $\mathbb{P}_\theta$-carré intégrable pour tout θ.
On appelle ***information de Fisher*** *du modèle statistique la fonction $I(\theta) : \theta \mapsto V_\theta(S(\theta)) = \mathbb{E}_\theta\left[(S(\theta) - \mathbb{E}_\theta[S(\theta)])(S(\theta) - \mathbb{E}_\theta[S(\theta)])^\top\right]$ définie sur Θ et à valeurs dans $\mathcal{L}^+$.*
Autrement dit, pour chaque θ il s'agit de la matrice de covariance relativement à $\mathbb{P}_\theta$ du score en θ

- une première source de variation liés à la dérivation d'une fonction strictement monotone de la vraisemblance
- une seconde source quantifiée par la variance de la v.a. score

Si $I(\theta)$ est grande alors la variabilité des probabilités du modèle au voisinage de $\mathbb{P}_\theta$ est importante : $\mathbb{P}_\theta$ est bien discriminée des autres

Modèle gaussien centré $\{\mathcal{N}(0, \theta)\}_{\theta\in]0,\infty[}$

La vraisemblance s'écrit

$$L(\theta, x) = \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{x^2}{2\theta}\right)$$

et

$$I(\theta) = \frac{1}{2\theta^2}.$$

Le modèle est d'autant plus informatif que θ est proche de 0. Cohérent avec l'idée de variance comme indicateur de la dispersion.

Une propriété importante qui sera donnée plus tard, est que l'information de Fisher donne une borne sur la variance des estimateurs.

Proposition 2

Pour un modèle de n-échantillon, on a

$$I_n(\theta) = nI(\theta).$$

On voit que l'information augmente avec la taille de l'échantillon.

Condition de dérivation sous le signe espérance

Pour toute fonction numérique $\{\mathbb{P}_\theta\}_{\theta\in\Theta}$-intégrable h

(CD1) $\quad \forall i = 1,\dots,p \qquad \dfrac{\partial}{\partial\theta_i}\mathbb{E}_\theta[h(X)L(\theta,X)] = \mathbb{E}_\theta\left[h(X)\dfrac{\partial L}{\partial\theta_i}(\theta,X)\right]$

Proposition 3

Sous la condition (CD1), on a

1. $\mathbb{E}_\theta[S(\theta)] = 0$, *i.e. la v.a. score est centrée ;*
2. *la matrice d'information vaut*

$$I(\theta) = \mathbb{E}_\theta\left[S(\theta)S(\theta)^\top\right] = \left(\mathbb{E}_\theta\left[\frac{\partial \ln L}{\partial\theta_i}(\theta,X)\frac{\partial \ln L}{\partial\theta_j}(\theta,X)\right]\right)_{1\le i,j\le p}$$

Dans le cas unidimensionnel, $I(\theta)$ se réduit au moment d'ordre deux du score.

Familles exponentielles

Ces conditions sont satisfaites pour les familles exponentielles régulières (c'est à dire $\Theta = N_\mu$ est un ouvert de $\mathbb{R}^p$) d'après un résultat similaire à celui de la Proposition 1 (voir [LR06, Th 2.7.1]).

On obtient alors à partir d'une représentation canonique (en utilisant également les formules (2-3))

$$I(\theta) = V_\theta(t(X)) \overset{(3)}{=} \left(\frac{\partial^2\Psi_\mu}{\partial\theta_i\partial\theta_j}(\theta)\right)_{i,j}.$$

Proposition 4 (Dépendance au paramétrage)

Pour un reparamétrage différentiable $\theta \mapsto \alpha(\theta)$, on a

$$I(\theta) = \left(\frac{\partial\alpha}{\partial\theta}\right) I(\alpha) \left(\frac{\partial\alpha}{\partial\theta}\right)^\top$$

Encore plus de régularité :

Condition de dérivation seconde sous le signe espérance

Pour une vraisemblance admettant des dérivées partielles secondes, pour toute fonction numérique $\{\mathbb{P}_\theta\}_{\theta\in\Theta}$-intégrable h

(CD2)

$$\forall i,j = 1,\dots,p \quad \frac{\partial^2}{\partial\theta_i\partial\theta_j}\mathbb{E}_\mu[h(X)L(\theta,X)] = \mathbb{E}_\mu\left[h(X)\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}(\theta,X)\right]$$

Proposition 5

Sous la condition (CD2), on a

(4) $$I(\theta) = -\left(\mathbb{E}_\theta\left[\frac{\partial^2 \ln L}{\partial\theta_i\partial\theta_j}(\theta,X)\right]\right)_{1\le i,j\le p}$$

Dans le cas unidimensionnel, l'expression (4) se réduit à

$$I(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2 \ln L}{\partial\theta^2}(\theta,X)\right]$$

Modèle binomial

La ln-vraisemblance s'écrit avec $\theta \in]0,1[$:

$$\ln L(\theta,x) = \ln C_n^x + x\ln\theta + (n-x)\ln(1-\theta)$$

d'où, à l'aide de la formule précédente, on obtient

$$I(\theta) = \frac{n}{\theta(1-\theta)}.$$

- L'information de Fisher est d'autant plus grande que θ est proche de 0 ou 1 (cohérent avec un dégré d'incertitude d'autant faible que θ est proche de 0 ou 1)
- On retrouve la forme associée à un n-échantillon de la loi de Bernoulli de paramètre θ.

Pour mesurer la distance entre deux lois, il existe d'autres indicateurs :

1. Distance en variation totale :

$$\begin{aligned} d_{\mathrm{TV}}(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2}) &= \sup_{A\in\mathscr{E}} |\mathbb{P}_{\theta_1}(A) - \mathbb{P}_{\theta_2}(A)| \\ &= \frac{1}{2}\sup_{A\in\mathscr{E}} \int_A |L(\theta_1,x) - L(\theta_2,x)| d\mu(x) \text{ (cas dominé)} \end{aligned}$$

2. « Distance » de Kullback-Leibler de $\mathbb{P}_{\theta_1}$ sur $\mathbb{P}_{\theta_2}$ (cas dominé) :

$$\begin{aligned} K[\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2}] &= \mathbb{E}_{\theta_1}\left[\ln \frac{L(\theta_1,X)}{L(\theta_2,X)}\right] \\ &= \int_{E_X} \ln\left(\frac{L(\theta_1,x)}{L(\theta_2,x)}\right) L(\theta_1,x)\, d\mu(x) \geq 0 \end{aligned}$$

et $K[\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2}] = 0$ ssi $\mathbb{P}_{\theta_1} = \mathbb{P}_{\theta_2}$. (mais pas de symétrie)

3. ...

Définition 8 (Statistique sur un modèle)

On appelle ***statistique*** *sur un modèle* $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$, *toute fonction mesurable* t *de* E_X *à valeurs dans* $t(E_X) \subset \mathbb{R}^p$ *et indépendante de* θ.
La valeur $t(x)$ *apparaît comme la réalisation de la v.a.* $T := t(X)$. *Par abus, on donne le nom de statistique aux deux quantités* t *et* T.

Définition 9 (Modèle image)

Pour une statistique t *définie sur le modèle statistique* $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$, *on appelle modèle image, le modèle* $(t(E_X), \mathcal{T}, (\mathbb{P}^t_\theta)_{\theta\in\Theta})$ *où* $\mathcal{T}$ *est l'ensemble de boréliens de* $t(E_X)$ *et* $\mathbb{P}^t_\theta$ $(\equiv \mathbb{P}^{t(X)}_\theta)$ *est la loi image de* $\mathbb{P}_\theta$ $(\equiv \mathbb{P}^X_\theta)$ *par* t.

Proposition 6

Si $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ *est un*

1. *modèle dominé par* μ *alors le modèle image est dominé par la mesure image* μ^t *de* μ *par* t *si celle-ci est* σ*-finie (ce qui sera toujours notre cas).*
2. *modèle homogène alors le modèle image est homogène.*

Information de Fisher d'uns statistique

L'information de Fisher d'une statistique d'un modèle statistique $(E_X, \mathscr{E}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ est définie comme l'information de Fisher du modèle image par t, à savoir l'information de Fisher (si elle existe) associée au modèle $(t(E_X), \mathcal{T}, \{\mathbb{P}^t_\theta\}_{\theta\in\Theta})$:

$$I^t(\theta) = V_\theta\left(\nabla \ln L^t(\theta, t(X))\right) \tag{5}$$

avec $L^t(\theta,\cdot)$ la vraisemblance dans le modèle image.

Lemme 1 (Score dans le modèle image)

Si le modèle (image) est homogène, on a pour tout θ

$$\nabla \ln L^t(\theta, t(X)) = \mathbb{E}_\theta[\nabla \ln L(\theta, X) \mid t(X)] \quad p.s.$$

Remarque 3

Le score dépend de θ et ne peut donc pas être une statistique.

Proposition 7 (Apport d'une statistique en termes d'inf. de Fisher)

Pour un modèle homogène, la matrice $I^t(\theta)$ *est inférieure ou égale à la matrice* $I(\theta)$, *i.e. la différence* $I(\theta) - I^t(\theta) \in \mathcal{L}^+$.

Autrement dit, l'information de Fisher portée par une statistique ne peut pas être supérieure à celle portée par le modèle initial.

Proposition 8

Deux statistiques indépendantes s *et* t *pour lesquelles l'information de Fisher est définie, vérifient*

$$I^{(s,t)}(\theta) = I^s(\theta) + I^t(\theta).$$

Cette propriété a déjà été utilisée pour le calcul de l'information de Fisher associé à un modèle de n-échantillon.