Divergences Empiriques

Hugo Harari-Kermadec

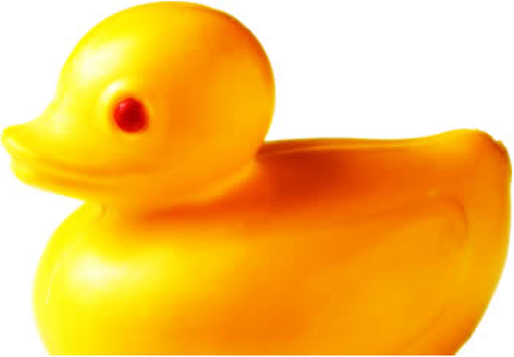
ENSAE

25 novembre 2008

1. Produit
   - Préambule
   - Une histoire de canards
   - Produit
   - Contraintes
2. Généralisation
3. Application et Biblio

On observe les tailles et les poids de 11 canards : $X_i = (T_i; P_i)$.
On veut la taille et le poids du "canard moyen" $\theta_M = (T_M; P_M)$.

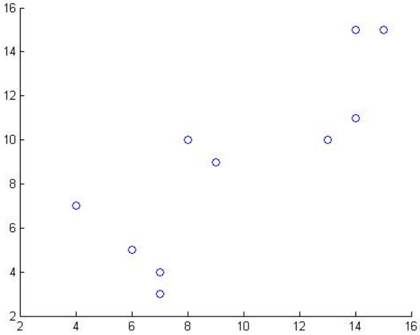
Taille



Poids

On observe les tailles et les poids de 11 canards : $X_i = (T_i; P_i)$.
On veut la taille et le poids du “canard moyen” $\theta_M = (T_M; P_M)$.

Taille



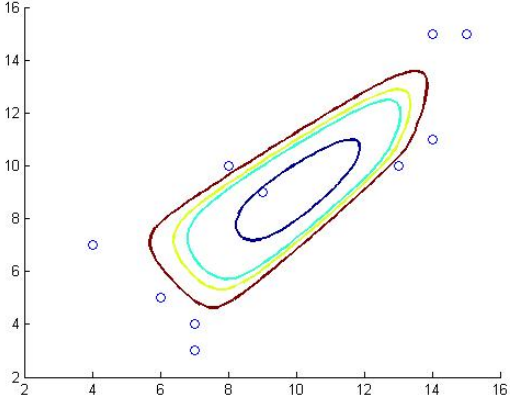
Poids

Produit :

Produit : $$\prod_i q_i$$

$$\textbf{Produit :} \sup_{\sum_i q_i X_i = \theta} \prod_i q_i$$

Produit : $\sup_{\sum_i q_i X_i = \theta} \prod_i q_i$



Lignes de niveau.

Définition (Ligne de niveau)

$$C_n(\eta) = \left\{ \theta = \sum_{i=1}^{n} q_i X_i \;\middle|\; -2\log\left(\prod_{i=1}^{n} q_i\right) \leq \eta \right\}$$

Définition (Ligne de niveau)

$$C_n(\eta) = \left\{ \theta = \sum_{i=1}^{n} q_i X_i \,\middle|\, 2\log\left(\prod_{i=1}^{n} \frac{1}{n}\right) - 2\log\left(\prod_{i=1}^{n} q_i\right) \leq \eta \right\}$$

Définition (Ligne de niveau)

$$
\begin{aligned}
C_n(\eta) &= \left\{ \theta = \sum_{i=1}^{n} q_i X_i \,\middle|\, 2\log\left(\prod_{i=1}^{n} \frac{1}{n}\right) - 2\log\left(\prod_{i=1}^{n} q_i\right) \leq \eta \right\} \\
&= \left\{ \quad \theta = X\mathbb{Q} \quad \middle| \quad -2n\sum_{i=1}^{n} \log\left(nq_i\right)/n \quad \leq \eta \right\}
\end{aligned}
$$

Définition (Ligne de niveau)

$$
\begin{aligned}
C_n(\eta) &= \left\{ \theta = \sum_{i=1}^{n} q_i X_i \;\middle|\; 2\log\left(\prod_{i=1}^{n} \frac{1}{n}\right) - 2\log\left(\prod_{i=1}^{n} q_i\right) \leq \eta \right\} \\
&= \left\{ \theta = X\mathbb{Q} \;\middle|\; -2n\sum_{i=1}^{n} \log\left(nq_i\right)/n \leq \eta \right\} \\
&= \left\{ \theta = X\mathbb{Q} \;\middle|\; 2nK(\mathbb{Q}, \mathbb{P}_n) \leq \eta \right\}
\end{aligned}
$$

Définition (Ligne de niveau)

$$
\begin{aligned}
C_n(\eta) &= \left\{ \theta = \sum_{i=1}^{n} q_i X_i \;\middle|\; 2\log\left(\prod_{i=1}^{n} \frac{1}{n}\right) - 2\log\left(\prod_{i=1}^{n} q_i\right) \leq \eta \right\} \\
&= \left\{ \theta = X\mathbb{Q} \;\middle|\; -2n\sum_{i=1}^{n} \log\left(nq_i\right)/n \leq \eta \right\} \\
&= \left\{ \theta = X\mathbb{Q} \;\middle|\; 2nK(\mathbb{Q}, \mathbb{P}_n) \leq \eta \right\}
\end{aligned}
$$

$$
\theta \in C_n(\eta) \Leftrightarrow \quad \exists \mathbb{Q},\ \theta = X\mathbb{Q} \ \text{ et } \ 2nK(\mathbb{Q}, \mathbb{P}_n) \leq \eta
$$

Définition (Ligne de niveau)

$$
\begin{aligned}
C_n(\eta) &= \left\{ \theta = \sum_{i=1}^{n} q_i X_i \,\middle|\, 2\log\left(\prod_{i=1}^{n}\frac{1}{n}\right) - 2\log\left(\prod_{i=1}^{n} q_i\right) \leq \eta \right\} \\
&= \left\{ \theta = X\mathbb{Q} \,\middle|\, -2n\sum_{i=1}^{n}\log\left(nq_i\right)/n \leq \eta \right\} \\
&= \left\{ \theta = X\mathbb{Q} \,\middle|\, 2nK(\mathbb{Q},\mathbb{P}_n) \leq \eta \right\}
\end{aligned}
$$

$$
\begin{aligned}
\theta \in C_n(\eta) &\Leftrightarrow \exists\mathbb{Q},\ \theta = X\mathbb{Q} \ \text{ et } \ 2nK(\mathbb{Q},\mathbb{P}_n) \leq \eta \\
&\Leftrightarrow \beta_n(\theta) = 2n \inf_{X(\mathbb{Q}-\mathbb{P}_n)=\theta-\overline{X}} K(\mathbb{Q},\mathbb{P}_n) \leq \eta
\end{aligned}
$$

On utilise la dualité pour résoudre l'optimisation sous contrainte :

On utilise la dualité pour résoudre l'optimisation sous contrainte :

$$\beta_n = 2n \inf_{X(\mathbb{Q}-\mathbb{P}_n)=\theta-\overline{X}} K(\mathbb{Q},\mathbb{P}_n)$$
$$= 2 \inf_{X(\mathbb{Q}-\mathbb{P}_n)=\theta-\overline{X}} \sum_{i=1}^{n} -\log\left(nq_i\right)$$

On utilise la dualité pour résoudre l'optimisation sous contrainte :

$$\begin{aligned}
\beta_n &= 2n \inf_{X(\mathbb{Q}-\mathbb{P}_n)=\theta-\overline{X}} K(\mathbb{Q},\mathbb{P}_n) \\
&= 2 \inf_{X(\mathbb{Q}-\mathbb{P}_n)=\theta-\overline{X}} \sum_{i=1}^{n} -\log\left(nq_i\right) \\
&= 2 \sup_{\lambda\in\mathbb{R}^p} \left\{ -\lambda'(\theta-\overline{X}) - \sum_i \log\left(1+\lambda'(X_i-\theta)\right) \right\}
\end{aligned}$$

car $(x \mapsto -\log(x))^* = (x \mapsto -1-\log(-x))$.

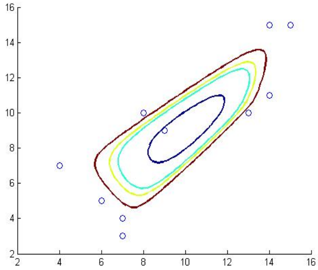
Information additionnelle : contraintes sur $\mathbb{Q}$

Si l'on connaît également les âges A_i et l'âge moyen A_0, on peut redresser notre échantillon :

Information additionnelle : contraintes sur $\mathbb{Q}$

Si l'on connaît également les âges A_i et l'âge moyen A_0, on peut redresser notre échantillon :

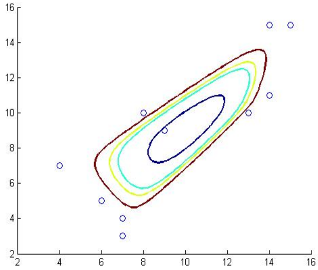
$$C_n(\eta) = \left\{ \theta = X\mathbb{Q} \middle| 2nK(\mathbb{Q}, \mathbb{P}_n) \leq \eta \right\}$$

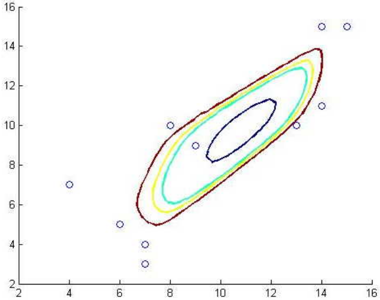


Information additionnelle : contraintes sur $\mathbb{Q}$

Si l'on connaît également les âges A_i et l'âge moyen A_0, on peut redresser notre échantillon :

$$C_n = \left\{ \theta = X\mathbb{Q} \middle| 2nK(\mathbb{Q}, \mathbb{P}_n) \leq \eta,\ A\mathbb{Q} = A_0 \right\}$$





On peut comparer les q_i au $\frac{1}{n}$ autrement que par le produit :

On peut comparer les q_i au $\frac{1}{n}$ autrement que par le produit :

$$C_n(\eta) = \left\{ \theta = \sum_{i=1}^n q_i X_i \middle| 2\log\left(\prod_{i=1}^n \frac{1}{n}\right) - 2\log\left(\prod_{i=1}^n q_i\right) \leq \eta \right\}$$

On peut comparer les q_i au $\frac{1}{n}$ autrement que par le produit :

$$C_n(\eta) = \left\{ \theta = \sum_{i=1}^{n} q_i X_i \,\middle|\, \sum_i \left(\frac{1}{n} - q_i \right)^2 \leq \eta \right\}$$

On peut comparer les q_i au $\frac{1}{n}$ autrement que par le produit :

$$\begin{aligned} C_n(\eta) &= \left\{ \theta = \sum_{i=1}^{n} q_i X_i \;\middle|\; \sum_i \left(\frac{1}{n} - q_i \right)^2 \leq \eta \right\} \\ &= \left\{ \theta = X\mathbb{Q} \;\middle|\; 2nK(\mathbb{Q}, \mathbb{P}_n) \leq \eta \right\} \end{aligned}$$

On peut comparer les q_i au $\frac{1}{n}$ autrement que par le produit :

$$
\begin{aligned}
C_n(\eta) &= \left\{ \theta = \sum_{i=1}^{n} q_i X_i \;\middle|\; \sum_i \left(\frac{1}{n} - q_i \right)^2 \leq \eta \right\} \\
&= \left\{ \theta = X\mathbb{Q} \;\middle|\; 2n||\mathbb{Q} - \mathbb{P}_n||_2^2 \leq \eta \right\}
\end{aligned}
$$

$$
\theta \in C_n(\eta) \Leftrightarrow \exists \mathbb{Q},\ \theta = X\mathbb{Q} \ \text{ et } \ ||\mathbb{Q} - \mathbb{P}_n||_2^2 \leq \eta
$$

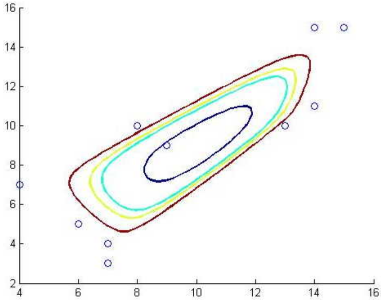
On peut comparer les q_i au $\frac{1}{n}$ autrement que par le produit :

$$C_n(\eta) = \left\{ \theta = \sum_{i=1}^{n} q_i X_i \;\middle|\; \sum_i \left(\frac{1}{n} - q_i\right)^2 \leq \eta \right\}$$
$$= \left\{ \theta = X\mathbb{Q} \;\middle|\; 2n||\mathbb{Q} - \mathbb{P}_n||_2^2 \leq \eta \right\}$$

$$\theta \in C_n(\eta) \Leftrightarrow \exists \mathbb{Q},\ \theta = X\mathbb{Q} \text{ et } ||\mathbb{Q} - \mathbb{P}_n||_2^2 \leq \eta$$
$$\Leftrightarrow \beta_n(\theta) = 2n \inf_{X(\mathbb{Q}-\mathbb{P}_n)=\theta-\overline{X}} ||\mathbb{Q} - \mathbb{P}_n||_2^2 \leq \eta$$

Produit

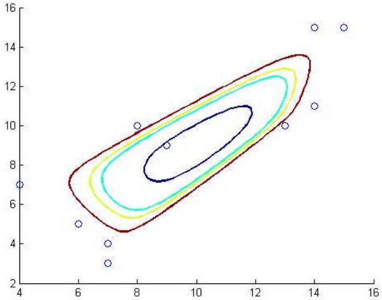
$$2\log\left(\prod_i \frac{1}{n}\right) - 2\log\left(\prod_i q_i\right)$$



Lignes de niveau

Produit

$$2\log\left(\prod_i \frac{1}{n}\right) - 2\log\left(\prod_i q_i\right)$$

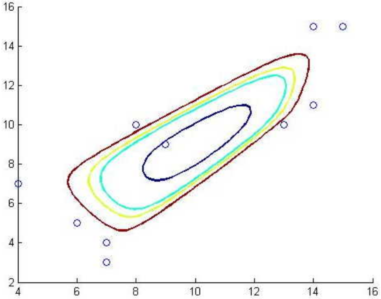


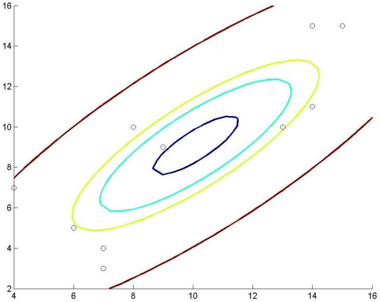
Norme Euclidienne

$$\sum_i \left(\frac{1}{n} - q_i\right)^2$$

Lignes de niveau

Produit	Norme Euclidienne
$2\log\left(\prod_i \frac{1}{n}\right) - 2\log\left(\prod_i q_i\right)$	$\sum_i \left(\frac{1}{n} - q_i\right)^2$





Lignes de niveau

Barycentres

On se place entre K et la norme euclidienne.

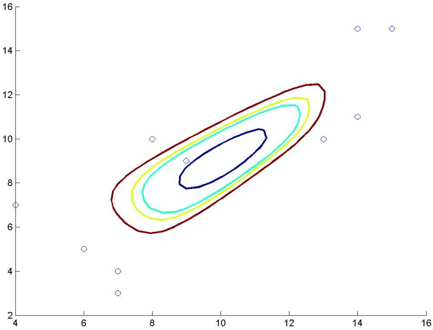
Barycentres

On se place entre K et la norme euclidienne.

$\forall \varepsilon \in [0;1],\ \forall x < 1,\ k_\varepsilon(x) = \varepsilon \cdot x^2/2 + (1-\varepsilon) \cdot [-1 - \log(1-x)]$

Barycentres

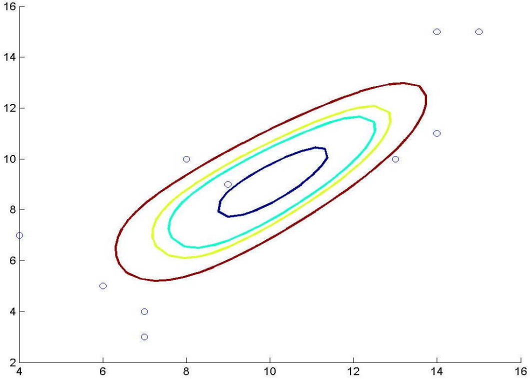
On se place entre K et la norme euclidienne.
$\forall \varepsilon \in [0;1],\ \forall x < 1,\ k_\varepsilon(x) = \varepsilon \cdot x^2/2 + (1-\varepsilon) \cdot [-1 - \log(1-x)]$



Quasi-Kullback pour $\varepsilon = 0.0$

Barycentres

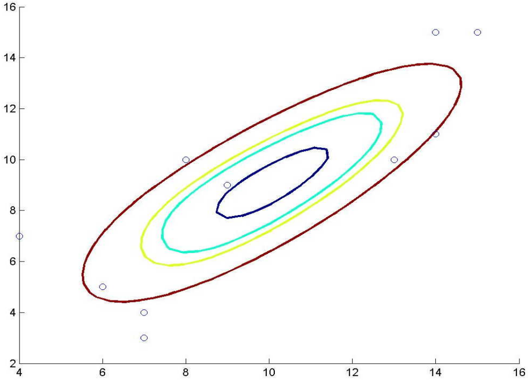
On se place entre K et la norme euclidienne.
$\forall \varepsilon \in [0;1],\ \forall x < 1,\ k_\varepsilon(x) = \varepsilon \cdot x^2/2 + (1-\varepsilon) \cdot [-1 - \log(1-x)]$



Quasi-Kullback pour $\varepsilon = 0.2$

Barycentres

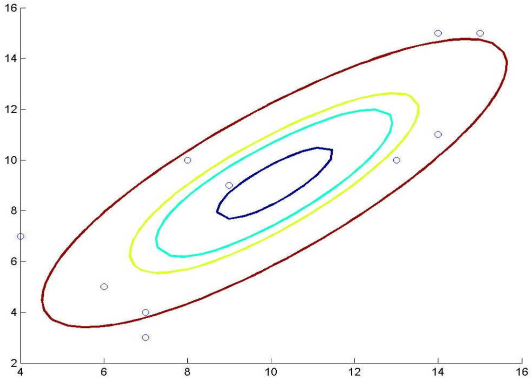
On se place entre K et la norme euclidienne.
$\forall \varepsilon \in [0;1],\ \forall x < 1,\ k_\varepsilon(x) = \varepsilon \cdot x^2/2 + (1-\varepsilon) \cdot [-1 - \log(1-x)]$



Quasi-Kullback pour $\varepsilon = 0.4$

Barycentres

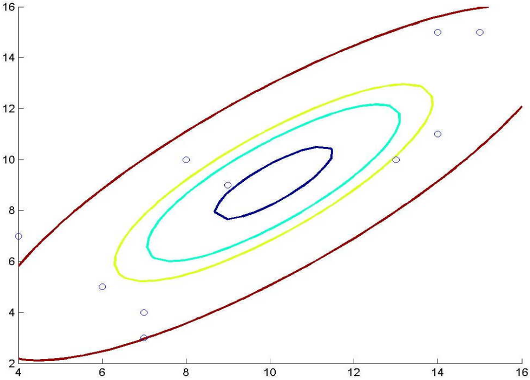
On se place entre K et la norme euclidienne.
$\forall \varepsilon \in [0;1],\ \forall x < 1,\ k_\varepsilon(x) = \varepsilon \cdot x^2/2 + (1-\varepsilon) \cdot [-1 - \log(1-x)]$



Quasi-Kullback pour $\varepsilon = 0.6$

Barycentres

On se place entre K et la norme euclidienne.
$\forall \varepsilon \in [0;1],\ \forall x < 1,\ k_\varepsilon(x) = \varepsilon \cdot x^2/2 + (1-\varepsilon) \cdot [-1 - \log(1-x)]$

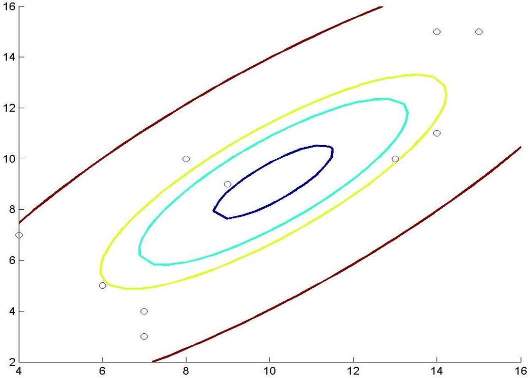


Quasi-Kullback pour $\varepsilon = 0.8$

Barycentres

On se place entre K et la norme euclidienne.

$\forall \varepsilon \in [0;1], \ \forall x < 1, \ k_\varepsilon(x) = \varepsilon \cdot x^2/2 + (1-\varepsilon) \cdot [-1 - \log(1-x)]$



Quasi-Kullback pour $\varepsilon = 1.0$

φ^*-Divergence

Définition (Conjuguée Convexe et φ^*-Divergence)

Soit φ convexe, avec $\varphi(0) = \varphi^{(1)}(0) = 0$. On construit alors :

φ^*-Divergence

Définition (Conjuguée Convexe et φ^*-Divergence)

Soit φ convexe, avec $\varphi(0) = \varphi^{(1)}(0) = 0$. On construit alors :

- la conjuguée convexe : $\varphi^*(x) = \sup_{y \in \mathbb{R}}\{xy - \varphi(y)\}$.

φ^*-Divergence

Définition (Conjuguée Convexe et φ^*-Divergence)

Soit φ convexe, avec $\varphi(0) = \varphi^{(1)}(0) = 0$. On construit alors :

- la conjuguée convexe : $\varphi^*(x) = \sup_{y\in\mathbb{R}}\{xy - \varphi(y)\}$.
- la φ^*-divergence : $I_{\varphi^*}(\mathbb{Q}, \mathbb{P}) = \frac{1}{n}\sum_i \varphi^*\left(nq_i - 1\right)$

φ^*-Divergence

Définition (Conjuguée Convexe et φ^*-Divergence)

Soit φ convexe, avec $\varphi(0) = \varphi^{(1)}(0) = 0$. On construit alors :

- la conjuguée convexe : $\varphi^*(x) = \sup_{y\in\mathbb{R}}\{xy - \varphi(y)\}$.
- la φ^*-divergence : $I_{\varphi^*}(\mathbb{Q},\mathbb{P}) = \frac{1}{n}\sum_i \varphi^*\left(nq_i - 1\right)$

Théorème (Dualité)

$$\inf_{\theta = X\mathbb{Q}} I_{\varphi^*}(\mathbb{Q},\mathbb{P}_n) = \sup_{\lambda}\left\{-\lambda'(\overline{X} - \theta) - \frac{1}{n}\sum_i \varphi\left(\lambda'(X_i - \theta)\right)\right\}$$

On peut alors remplacer la divergence K par une φ^*-divergence

On peut alors remplacer la divergence K par une φ^*-divergence

Lignes de niveau par divergence

$$C_n(\eta) = \left\{ \theta = X\mathbb{Q} \middle| \quad 2nK(\mathbb{Q}, \mathbb{P}_n) \quad \leq \eta \right\}$$

On peut alors remplacer la divergence K par une φ^*-divergence

Lignes de niveau par divergence

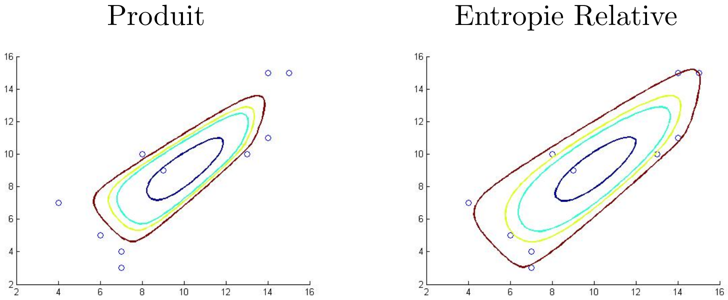
$$C_n(\eta) = \left\{ \theta = X\mathbb{Q} \;\middle|\; 2n\textcolor{red}{I_{\varphi^*}}(\mathbb{Q}, \mathbb{P}_n) \leq \eta \right\}$$

On peut choisir l'Entropie Relative comme divergence :

$$
\begin{aligned}
\phi(x) &= e^x - 1 - x \text{ sur } \mathbb{R} \\
\phi^*(x) &= (x+1)\log(x+1) - x \text{ sur }]-1;+\infty[
\end{aligned}
$$

On peut choisir l'Entropie Relative comme divergence :

$$\begin{aligned}\phi(x) &= e^x - 1 - x \text{ sur } \mathbb{R}\\ \phi^*(x) &= (x+1)\log(x+1) - x \text{ sur }]-1;+\infty[\end{aligned}$$



Lignes de niveau .

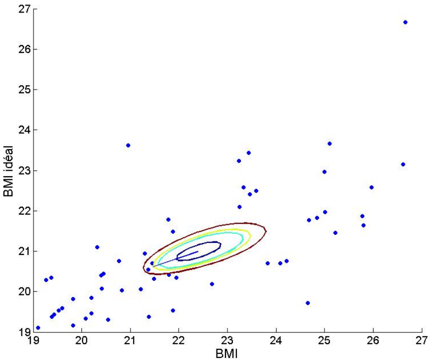
Norme sociale

On s'intéresse à l'IMC idéal : poids idéal sur taille au carré.
Il apparaît que le lien avec l'IMC réel dépend du groupe social :

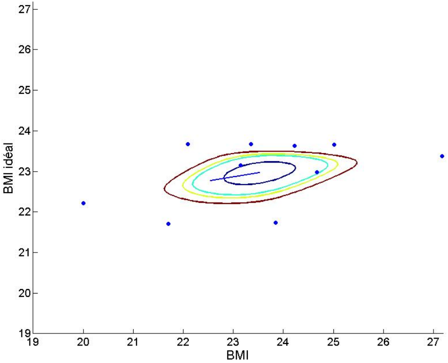
Norme sociale

On s'intéresse à l'IMC idéal : poids idéal sur taille au carré.
Il apparaît que le lien avec l'IMC réel dépend du groupe social :

Jeunes employés



Jeunes employées



Bibliographie

C. Léonard.
Convex conjugates of integral functionals.
Acta Mathematica Hungarica, 93(4) :253–280, 2001.

A. B. Owen.
Empirical Likelihood.
Chapman and Hall/CRC, Boca Raton, 2001.

R. T. Rockafellar.
Convex Analysis.
Princeton University Press, Princeton, NJ, 1970.

Divergences	$\varphi_\alpha^*(x)$	$d(\varphi_\alpha^*)$
Vraisemblance Empirique	$x - \log(1+x)$	$]-1,\infty[$
Norme euclidienne	$\frac{x^2}{2}$	$\mathbb{R}$
Entropie Relative	$(x+1)\log(x+1) - x$	$]-1,\infty]$
Hellinger	$2(\sqrt{(x+1)}-1)^2$	$]-1,\infty[$

Principales Cressie-Read