Panorama des méthodes d'estimation sur petits domaines

Pascal Ardilly
Insee (UMS)

Définition et problématique

Domaine = sous-population a (taille N_a) de U

$$Y_a = \sum_{i=1}^{N_a} Y_i \qquad \text{et} \qquad \overline{Y}_a = \frac{1}{N_a} \sum_{i=1}^{N_a} Y_i \quad ?$$

Echantillon s tiré dans U (plan complexe).

Trois catégories d'estimateurs

- **<u>Estimateurs directs</u>** : aucune information utilisée hors du domaine ;

- **<u>Estimateurs indirects avec modélisation implicite</u>** : s'appuient sur un modèle de comportement reliant le domaine au reste de la population U ;

Le « modèle » porte sur des données agrégées et il n'y a pas d'autre aléa que celui de l'échantillonnage : l'approche reste descriptive.

- **Estimateurs indirects avec modélisation explicite** : s'appuient sur un modèle de comportement sur U utilisant des variables auxiliaires explicatives et une composante stochastique classique (modèles linéaires mixtes).

Où est le problème ?

$\Rightarrow$ **Pas dans le biais !**

Estimateur direct « de base » (Horvitz-Thompson) :

$$\boxed{\hat{Y}_a = \sum_{i \in s_a} \frac{Y_i}{\Pi_i}}$$

$s_a = s \cap a$ (taille n_a)

SANS BIAIS

Echantillonnage à probabilités égales et de taille fixe :

$$\hat{Y}_a = N \cdot \frac{n_a}{n} \cdot \bar{y}_a = \hat{N}_a . \bar{y}_a$$

$\Rightarrow$ **Mais dans la variance :**

$$EQM\left(\hat{Y}_a\right) = V(\hat{Y}_a) = O\left(\frac{1}{n_a}\right)$$

PETIT domaine $\Leftrightarrow$ n_a est PETIT

Exemple du tirage aléatoire simple ,

- ***Conditionnel*** :

$$V\left[\hat{Y}_a \middle| n_a\right] = \left(N \cdot \frac{n_a}{n}\right)^2 \left(1 - \frac{n_a}{N_a}\right) \frac{S_a^2}{n_a}$$

- ***NON conditionnel*** :

$$V\left(\hat{Y}_a\right) \# N_a^2 \cdot \frac{1}{\frac{nN_a}{N}} \cdot \left[S_a^2 \left(1 - \frac{n}{N_a}\right) + \bar{Y}_a^2 \left(1 - \frac{N_a}{N}\right) \right]$$

$$\boxed{\textbf{QUE FAIRE ???}}$$

Mise en garde :

Ne pas confondre avec l'estimation de la part que le domaine représente dans la population U:

$P_a = \dfrac{N_a}{N}$ estimé par $\hat{P}_a = \dfrac{\hat{N}_a}{\hat{N}}$

et $V(\hat{P}_a) = O(\frac{1}{n})$

Estimateurs directs « par la régression »

X_i information auxiliaire dans R^p.

$$X_a = \sum_{i=1}^{N_a} X_i \quad \underline{\text{connu}}$$

On pose

$$\forall i = 1, 2, \ldots, N_a : \quad Y_i = B_a^T \cdot X_i + U_i$$

où $\quad B_a^T = \left(B_a^1, B_a^2, \ldots, B_a^p\right)$ dans R^p.

et $$Var\begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{pmatrix} = \Omega = \begin{pmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ & 0 & & \sigma_N^2 \end{pmatrix}$$

Le paramètre B_a optimum (critère des moindres carrés) est $\widetilde{B}_a$, estimé par :

$$\hat{B}_a = \left(\sum_{i \in s_a} \frac{X_i \cdot X_i^T}{\Pi_i \hat{\sigma}_i^2}\right)^{-1} \cdot \left(\sum_{i \in s_a} \frac{X_i \cdot Y_i}{\Pi_i \hat{\sigma}_i^2}\right) \quad \text{dans } \mathrm{R}^P$$

L'estimateur par la régression du total Y_a est alors défini par :

$$\boxed{\hat{Y}_{\mathrm{Re}g,a} = \hat{Y}_a + \hat{B}_a^T \cdot \left(X_a - \hat{X}_a\right)}$$

En pratique on trouve 2 cas :

$X_i \in \mathbb{R}^P$ et σ_i^2 est constant,

ou

- $X_i \in \mathbb{R}$ et σ_i^2 prop à X_i, soit $\sigma_i^2 = \sigma^2 \cdot X_i$.

<u>Cas 1</u> :

$$\hat{B}_a = \left(\sum_{i \in s_a} \frac{X_i X_i^T}{\Pi_i} \right)^{-1} \cdot \left(\sum_{i \in s_a} \frac{X_i Y_i}{\Pi_i} \right)$$

<u>Cas 2</u> :

$$\hat{B}_a = \frac{\sum_{i \in s_a} Y_i / \Pi_i}{\sum_{i \in s_a} X_i / \Pi_i} = \frac{\hat{Y}_a}{\hat{X}_a} \qquad (\hat{B}_a \in \mathbb{R})$$

$$\Rightarrow \hat{Y}_{Reg,a} = X_a \cdot \frac{\hat{Y}_a}{\hat{X}_a} \quad \text{(RATIO)}$$

Le « grand » résultat :

1/ $\hat{Y}_{Reg,a}$ sans biais si n_a assez grand ($n_a \geq 50$?)
quelle que soit la « qualité » de la liaison ;

$\Rightarrow$ Il n'y a pas de dépendance envers le « modèle ».

2/ Toujours si n_a « assez grand » :

$$\boxed{V\left(\hat{Y}_{Reg,a}\right) \approx V\left(\hat{U}_a\right) = V\left(\sum_{i \in s_a} \frac{U_i}{\Pi_i}\right)}$$

où $U_i = Y_i - \tilde{B}_a^T \cdot X_i$.

Conclusion : si n_a « assez grand » et si le caractère explicatif de X_i est fort, $\hat{Y}_{Reg,a}$ peut être compatible avec l'objectif de précision.

Estimateurs indirects avec modélisation implicite

Il y a deux classes d'estimateurs de ce type :

- Les estimateurs synthétiques
- Les estimateurs composites

I) <u>Les estimateurs synthétiques</u> :

Justification de l'estimateur synthétique : croire à un modèle descriptif du type

$$\boxed{\text{paramètre sur } a = \text{paramètre sur U}}$$

A) Estimateurs synthétiques sans info auxiliaire:

$$\boxed{\hat{Y}_{a,SYN} = N_a.\hat{\bar{Y}} = N_a \cdot \frac{\hat{Y}}{\hat{N}}}$$

où $\hat{Y} = \sum_{i \in s} \frac{Y_i}{\Pi_i}$ et $\hat{N} = \sum_{i \in s} \frac{1}{\Pi_i}$.

Cas du tirage aléatoire simple : $\hat{Y}_{a,SYN} = N_a.\bar{y}$

n grand $\Rightarrow$ $Biais = E\left(\hat{Y}_{a,SYN}\right) - Y_a \approx N_a.(\overline{Y} - \overline{Y}_a)$

<u>Modèle implicite</u> $\overline{Y} = \overline{Y}_a$

$$EQM\left(\hat{Y}_{a,SYN}\right) = N_a{}^2.[\quad \underbrace{V\left(\hat{\overline{Y}}\right)}_{\text{varie en } 1/n} \quad + \quad \underbrace{\left(\overline{Y}_a - \overline{Y}\right)^2}_{\text{ne dépend pas de n}} \quad)]$$

FAIBLE SI LE MODELE EST

(A PEU PRES) VERIFIE

<u>Nota</u> : $\hat{Y}_{a,SYN}$ est calculable même si n_a = 0 !!!

B) Estimateurs synthétiques avec info auxiliaire:

$$\hat{Y}_{Reg,a} = \hat{Y}_a + \hat{B}_a^T\left(X_a - \hat{X}_a\right)$$

$$= X_a^T\hat{B}_a + \left(\underbrace{\hat{Y}_a - \hat{B}_a^T \cdot \hat{X}_a}_{\Delta_a}\right)$$

Δ_a est « petit » devant $X_a^T\hat{B}_a$ et vaut 0 dans les cas « habituels ».

Pour stabiliser $\hat{Y}_{Reg,a}$ on remplace $\hat{B}_a$ par $\hat{B}$:

$$\boxed{\hat{Y}_{a,REGSYN} = X_a^T \hat{B}}$$

avec $$\hat{B} = \left(\sum_{i \in s} \frac{X_i \cdot X_i^T}{\hat{\sigma}_i^2 . \Pi_i} \right)^{-1} \cdot \left(\sum_{i \in s} \frac{X_i . Y_i}{\hat{\sigma}_i^2 . \Pi_i} \right)$$

estimant $$\tilde{B} = \left(\sum_{i \in U} \frac{X_i \cdot X_i^T}{\sigma_i^2} \right)^{-1} \cdot \left(\sum_{i \in U} \frac{X_i . Y_i}{\sigma_i^2} \right)$$

$$\begin{aligned} Biais &= E\left(X_a^T \hat{B}\right) - Y_a \approx X_a^T \tilde{B} - Y_a \\ &= X_a^T \left(\tilde{B} - \tilde{B}_a\right) - \left(Y_a - X_a^T \tilde{B}_a\right) \end{aligned}$$

$$= X_a^T \left(\tilde{B} - \tilde{B}_a\right)$$ dans les cas « usuels »

<u>Modèle implicite</u> $\tilde{B}_a = \tilde{B}$

$$EQM\left(\hat{Y}_{a,REGSYN}\right) \approx \left[X_a^T\left(\tilde{B} - \tilde{B}_a\right)\right]^2$$

+ fonction de $1/n$

Deux déclinaisons de l'estimateur synthétique de type régression, lorsque $X_i \in R$:

- *CAS 1* :

$$Y_i = \sum_{h=1}^{H} \lambda_h \left(X_i 1_{i \in h}\right) + U_i \;, Var\, U_i = \sigma_h^2 \cdot X_i \text{ si } i \in h$$

$$\boxed{\hat{Y}_{a,REGSYN} = \sum_{h=1}^{H} X_{ah} \cdot \frac{\hat{Y}_h}{\hat{X}_h}}$$

<u>Modèle implicite</u> pour tout h :

$$\frac{\sum_{\substack{i \in h \\ i \in a}} Y_i}{\sum_{\substack{i \in h \\ i \in a}} X_i} = \frac{\sum_{i \in h} Y_i}{\sum_{i \in h} X_i} \quad \Leftrightarrow \quad \frac{\overline{Y}_{ah}}{\overline{X}_{ah}} = \frac{\overline{Y}_h}{\overline{X}_h}$$

- *CAS 2* : $X_i = 1$

$$\boxed{\hat{Y}_{a,REGSYN} = \sum_{h=1}^{H} N_{ah} \frac{\hat{Y}_h}{\hat{N}_h}}$$

$\hat{N}_h$ estime la taille de la sous-population h.

Modèle implicite pour tout h :

$$\bar{Y}_{ah} = \bar{Y}_h$$

Cas du tirage aléatoire simple $\Rightarrow$ estimateur synthétique de type post-stratifié :

$$\hat{Y}_{a,REGSYN} = \sum_{h=1}^{H} N_{ah} \cdot \bar{y}_h$$

$$\bar{y}_h = \frac{1}{n_h} \cdot \sum_{i \in s_h} Y_i$$

$$EQM\left(\hat{Y}_{a,REGSYN}\right) = \left[\sum_{h=1}^{H} N_{ah}\left(\bar{Y}_h - \bar{Y}_{ah}\right)\right]^2 + \sum_{h=1}^{H} N_{ah}^2 \left(E\left(\frac{1}{n_h}\right) - \frac{1}{N_h}\right) S_h^2$$

PEUT-ON JUGER DU BIAIS ?

Plutôt NON : il ne semble pas possible d'estimer de manière stable les EQM des estimateurs directs. Le problème est que Y_a est inconnu !!!

Cependant, on obtient un estimateur (presque) sans biais en formant :

$$\boxed{E\hat{Q}M\left(\hat{Y}_{a,SYN}\right) \cong \left(\hat{Y}_{a,SYN} - \hat{Y}_a\right)^2 - \hat{V}\left(\hat{Y}_a\right)}$$

et $\hat{V}\left(\hat{Y}_a\right)$ fourni par un logiciel adéquat (Poulpe ?).

Suggestion : si on considère m domaines « pas trop » différents,

$$\boxed{\frac{1}{N_a^2} E\hat{Q}M\left(\hat{Y}_{a,SYN}\right) \approx \frac{1}{m} \sum_{a=1}^{m} \frac{1}{N_a^2} \left(\hat{Y}_{a,SYN} - \hat{Y}_a\right)^2 - \frac{1}{m} \sum_{a=1}^{m} \frac{1}{N_a^2} \hat{V}\left(\hat{Y}_a\right)}$$

Une dernière tentative pour supprimer le biais…

? Pourquoi pas :

$$\tilde{Y}_{a,REGSYN} = \hat{Y}_a + \hat{B}^T \left(X_a - \hat{X}_a\right)$$

Il est (presque) sans biais si n est grand . Cependant

$$V\left(\tilde{Y}_{a,REGSYN}\right) \approx V\left(\hat{Y}_a - \hat{B}^T \cdot \hat{X}_a\right) \approx V\left(\hat{\Phi}_a\right)$$
$$= O\left(\frac{1}{n_a}\right)$$

où $\Phi_i = Y_i - \tilde{B}^T X_i$

Il peut néanmoins être intéressant si n_a « tout petit » et que le caractère explicatif de X est fort.

II) Les estimateurs composites :

$\hat{Y}_a^D$: un estimateur direct

$\hat{Y}_a^{SYN}$: un estimateur synthétique

$$\boxed{\hat{Y}_{a,COMP} = \phi_a.\hat{Y}_a^D + (1-\phi_a).\hat{Y}_a^{SYN}} \quad \text{où} \quad \phi_a \in [0,1]$$

On module ainsi les poids des estimateurs *directs* (biais faible, forte variance) et *synthétique* (biais, faible variance).

A) Estimateur composite optimum :

Si on minimise l'EQM en ϕ_a, on trouve

$$\phi_a(OPTI) = \frac{1}{1+F_a} \qquad (\in [0,1])$$

avec $$F_a = \frac{EQM\left(\hat{Y}_a^D\right)}{EQM\left(\hat{Y}_a^{SYN}\right)}$$

Propriété 1 :

$$1 \geq \frac{EQM\left(\hat{Y}_{a,COMP}^{OPTI}\right)}{MIN\left(EQM\left(\hat{Y}_a^D\right), EQM\left(\hat{Y}_a^{SYN}\right)\right)} \geq \frac{1}{2}$$

Propriété 2 :

Si $Max(0{,}2\cdot\phi_a(OPTI)-1)\leq\phi_a\leq Min(1{,}2\phi_a(OPTI))$

alors
$EQM(\hat{Y}_{a,COMP})\leq Min\left(EQM(\hat{Y}_a^D),EQM(\hat{Y}_a^{SYN})\right)$

Il faut estimer $\phi_a(OPTI)$, par exemple par

$$\hat{\phi}_a(OPTI)=\frac{E\hat{Q}M(\hat{Y}_a^{SYN})}{(\hat{Y}_a^D-\hat{Y}_a^{SYN})^2}$$

On souffre de l'instabilité de cette estimation !

B) Estimateur dépendant de la taille de l'échantillon :

Idée : retrouver un estimateur direct lorsque n_a est « assez grand », c'est-à-dire lorsque

$$\hat{N}_a=\sum_{i\in s_a}\frac{1}{\Pi_i}$$

est grand. Par exemple:

$$\phi_a=\begin{cases}1 & \text{si } \hat{N}_a>\delta\cdot N_a\\ \dfrac{\hat{N}_a}{\delta\,.\,N_a} & \text{si } \hat{N}_a\leq\delta\cdot N_a\end{cases}$$

Exemple du sondage aléatoire simple avec $\delta = 1$:

- $\hat{Y}_{a,COMP} = \hat{Y}_a^D \qquad$, si $n_a \geq n \cdot \dfrac{N_a}{N}$

- $\hat{Y}_{a,COMP} = \left(\dfrac{N}{n} \dfrac{n_a}{N_a} \right) \hat{Y}_a^D + \left(1 - \dfrac{N}{n} \dfrac{n_a}{N_a} \right) \hat{Y}_a^{SYN}$, sinon

ATTENTION : il faut quand même N_a « assez grand ».

Dans la littérature, on trouve par exemple

$$\hat{Y}_{a,COMP} = \phi_a \cdot \left(\hat{B}^T \cdot X_a + \frac{N_a}{\hat{N}_a} \left[\hat{Y}_a - \hat{B}^T \hat{X}_a \right] \right) + (1 - \phi_a) \cdot \left(\hat{B}^T X_a \right)$$

avec $\phi_a = 1 \qquad si\ \hat{N}_a \geq N_a$

$$= \left(\frac{\hat{N}_a}{N_a} \right)^2 \quad si\ \hat{N}_a < N_a$$

C) Estimateur dit « de James-Stein »

<u>Contexte</u> :

- On s'intéresse à **m** domaines : a varie de 1 à m ;

- Il existe une fonction g telle que $\hat{\theta}_a = g\left(\hat{Y}_a^D\right)$ et

$$\hat{\theta}_a \to \mathcal{N}(\theta_a, \Psi_a)$$

Ψ_a = variance d'échantillonnage de $\hat{\theta}_a$, <u>connue</u>.

- On dispose de valeurs « a priori » θ_a^o supposées proches des θ_a :

  - information totalement externe, ou tirée d'une enquête passée ;

  - estimation construite à partir d'informations auxiliaires z_a (de dimension p) :

$$\theta_a^o = z_a^T \hat{\beta} \text{ où } \hat{\beta} = (Z^T Z)^{-1} Z^T \hat{\theta}$$

avec $Z^T = (z_1, z_2, \ldots, z_m)$ matrice $p \times m$.

Au minimum : $p = 1$ et $z_a = 1$

$$\Rightarrow \forall a \text{ de } 1 \text{ à } m : \quad \theta_a^o = \frac{1}{m} \sum_{a=1}^{m} \hat{\theta}_a$$

Soit le critère de qualité :

$$R\left(\theta, \tilde{\theta}\right) = \sum_{a=1}^{m} E\left(\tilde{\theta}_a - \theta_a\right)^2$$

Supposons (simplification) : $\Psi_a = \Psi$

Exemple des proportions : $\theta_a = g(P_a) = \text{Arcsin}\left(\sqrt{P_a}\right)$

$\Longrightarrow$ Estimateurs dits « de James-Stein » :

$$\boxed{\hat{\theta}_{a,JS} = \theta_a^o + \left[1 - \frac{K \cdot \Psi}{S}\right] \cdot \left(\hat{\theta}_a - \theta_a^o\right)}$$

avec $$S = \sum_{a=1}^{m} \left(\hat{\theta}_a - \theta_a^o\right)^2$$

$$K = \begin{cases} m-2 & \text{si les } \theta_a^o \text{ sont exogènes ;} \\ m-p-2 & \text{si les } \theta_a^o \text{ sont issus d'une régression} \\ & \text{sur } z_a \text{ (de dimension } p\text{).} \end{cases}$$

Cela impose une borne minimale pour m.

Cet estimateur :

- est (très) facile à calculer ;
- s'écrit aussi

$$\hat{\phi}_{JS} \cdot \hat{\theta}_a + \left(1 - \hat{\phi}_{JS}\right) \cdot \theta_a^o$$

où $$\hat{\phi}_{JS} = 1 - \frac{K \cdot \Psi}{S}$$ (ne dépend pas de a)

<u>Propriété 1</u> :

$$R\left(\theta, \hat{\theta}_{JS}\right) \leq R\left(\theta, \hat{\theta}\right) \quad \text{pour tout } \theta.$$

Propriété 2 :

$$R\left(\theta,\hat{\theta}\right)=m\cdot\Psi\,.$$

Or on montre :

$$R\left(\theta,\hat{\theta}_{JS}\right)\leq m\Psi-\frac{(m-2)^2\Psi^2}{(m-2)\Psi+\sum_{a=1}^{m}\left(\theta_a-\theta_a^o\right)^2}$$

Si $\theta_a^o\approx\theta_a \Rightarrow R\left(\theta,\hat{\theta}_{JS}\right)\approx 2\Psi$: très fort gain si m est grand (départements, ZUS).

L'estimation de James-Stein a aussi des faiblesses.

- Si la fonction g n'est pas l'identité, la dominance de $\hat{Y}_{a,JS}=g^{-1}\left(\hat{\theta}_{a,JS}\right)$ sur $\hat{Y}_a^D$ n'est pas assurée.
- Le critère de qualité est global : pas de garantie au niveau d'un domaine donné (certains $\hat{\theta}_{a,JS}$ pourront être moins bons que le $\hat{\theta}_a$).

III) Un complément intéressant pour estimer des effectifs : la méthode de préservation des structures :

A t, comment estimer les effectifs de a vérifiant les modalités d'une variable qualitative X donnée si on a :

- Les effectifs (éventuellement estimés) $N_{a,uv}$ dans a qui vérifiaient, à une date antérieure t_0, la modalité u de X et la modalité v de Z

  (origine : RP ou fichier administratif) ;

- A la date t, des estimations (fiables) $\hat{M}_{\bullet,uv}$ des effectifs croisant u et v (donc sur l'ensemble des domaines). L'actualisation de $M_{a,uv}$ est notée $N_{a,uv}$.

$$M_{\bullet,uv} = \sum_{a=1}^{m} M_{a,uv} \text{ et } M_{a,u\bullet} = \sum_{v} M_{a,uv}$$

$$\text{Minimiser} \quad \sum_{a=1}^{m} \sum_{u} \sum_{v} \left(N_{a,uv} - x_{a,uv}\right)^2 \Big/ N_{a,uv}$$

$$\text{sous contraintes} \quad \sum_{a} x_{a,uv} = \hat{M}_{\bullet,uv} \qquad \forall u \text{ et } \forall v$$

ou encore :

$$\text{Minimiser} \qquad \sum_{a=1}^{m} \sum_{u} \sum_{v} N_{a,uv} Log \frac{N_{a,uv}}{x_{a,uv}}$$

$$\text{sous contraintes} \qquad \sum_{a} x_{a,uv} = \hat{M}_{\bullet,uv}$$

La solution est :

$$x_{a,uv} = \frac{N_{a,uv}}{N_{\bullet,uv}} \times \hat{M}_{\bullet,uv}$$

Puis

$$\hat{M}_{a,u\bullet} = \sum_{v} \frac{N_{a,uv}}{N_{\bullet,uv}} \times \hat{M}_{\bullet,uv}$$

Si E($\hat{M}_{\bullet,uv}$) = $M_{\bullet,uv}$ (généralement vérifié) et si, pour tout u et pour tout v :

$$\frac{M_{a,uv}}{N_{a,uv}} \text{ ne dépend pas (peu) de } a \Leftrightarrow \frac{M_{a,uv}}{N_{a',uv}} = \frac{N_{a,uv}}{N_{a',uv}}$$

alors $\hat{M}_{a,u\bullet}$ est sans biais de $M_{a,u\bullet}$.

Cet estimateur $\hat{M}_{a,u\bullet}$ est <u>de type synthétique</u> et

$$V(\hat{M}_{a,u\bullet}) = O\left(\frac{1}{\sum_{a=1}^{m} n_a} \right)$$

$\Longrightarrow$ fort gain si m grand.

Extension possible si on connaît $\hat{M}_{a,\bullet\bullet}$

$$\text{Minimiser} \quad \sum_{a=1}^{m} \sum_{u} \sum_{v} N_{a,uv} \cdot Log \frac{N_{a,uv}}{x_{a,uv}}$$

$$\text{Sous contraintes} \begin{cases} \sum_{a} x_{a,uv} = \hat{M}_{\bullet,uv} \text{ pour tout } (u,v) \\ \sum_{u,v} x_{a,uv} = \hat{M}_{a,\bullet\bullet} \end{cases}$$

Pas de solution analytique : nécessite un algorithme itératif de type « raking ratio » (la fonction objectif n'est pas dans Calmar !).

Estimateurs indirects avec modélisation explicite

On distingue :

- 2 grandes catégories de modèles :
    - Les modèles au niveau du domaine ;
    - Les modèles au niveau des individus.

- 3 stratégies principales :
    - Les estimateurs sans biais linéaires optimaux ;
    - Les estimateurs optimaux ;
    - Les estimateurs « Bayésiens hiérarchiques » .

On s'intéresse toujours à **m** <u>domaines</u> : a varie de 1 à m : c'est seulement dans ce contexte qu'on « tirera de la force » d'une modélisation.

I) <u>Les estimateurs sans biais optimaux linéaires (SBOL)</u> :

A) Théorie générale :

$$\boxed{Y = X\beta + Zv + e}$$

Y : vecteur observé de taille n
X : matrice $n \times p$ (connue)
β : vecteur **<u>inconnu</u>** de taille $p \rightarrow$ **<u>effets FIXES</u>**
Z : matrice $n \times h$ (connue)
v : **<u>vecteur aléatoire</u>** de taille h (inobservé)
$\rightarrow$ **<u>effets ALEATOIRES</u>**
e : vecteur aléatoire de taille n (inobservé)

Modèle linéaire mixte

- $E(e) = 0$ et $E(v) = 0$
- e et v sont indépendants
- $V(v) = G(\delta)$ (on notera G)
- $V(e) = R(\delta)$ (on notera R)

où $\delta = (\delta_1, \delta_2, \ldots, \delta_q)^T$

On cherche à prédire la valeur de la variable aléatoire

$$\mu = l^T \beta + m^T v \qquad (\in R)$$

où l et m sont des vecteurs parfaitement connus. On cherchera (Y étant observé) :

$$\hat{\mu} = a^T \cdot Y + b$$

où a et b sont des vecteurs parfaitement déterministes. On va imposer :

$$E(\hat{\mu} - \mu) = 0$$

en minimisant :

$$E(\hat{\mu} - \mu)^2$$

On obtient, après calculs, l'estimateur « SBOL » :

$$a^T_{SBOL} = \left(l^T - m^T G\, Z^T V^{-1} X\right)\left(X^T V^{-1} X\right)^{-1} X^T V^{-1} + m^T G\, Z^T V^{-1}$$

$$b_{SBOL} = 0$$

avec $\qquad V = V(Y) = V(Z.v) + V(e) = ZGZ^T + R$

Le prédicteur optimum est noté $\hat{\mu}^H = a_{SBOL} Y + b_{SBOL}$.

Ecriture intéressante :

$$\boxed{\hat{\mu}^H = l^T \tilde{\beta} + m^T \left[G\, Z^T V^{-1}\left(Y - X\tilde{\beta}\right)\right]}$$

où $\qquad \tilde{\beta} = \left(X^T V^{-1} X\right)^{-1} \cdot X^T V^{-1} Y$

Terme entre crochets = prédicteur (optimum) de v.

Souvent δ est inconnu : il faut l'estimer et le remplacer par $\hat{\delta}$ $\Rightarrow$ estimateur empirique « ESBOL ».

On peut utiliser (par exemple) :

- La méthode des moments ;
- L'estimation du maximum de vraisemblance (EMV), mais il faut postuler une loi pour e et v.

- Complexité algorithmique, et difficulté à estimer l'EQM (mais c'est possible !).

- La PROC MIXED de SAS fournit l'estimateur ESBOL à partir des paramètres estimés $\left(\hat{\beta},\hat{\delta}\right)$ par EMV, ainsi que $E\hat{Q}M\left(\hat{\mu}^{H}(\hat{\beta},\hat{\delta})\right)$.

B) Application au cas d'une modélisation au niveau du domaine (modèle de Fay et Herriot) :

Pour chacun des m domaines, on dispose d'une information auxiliaire $z \in R^{p}$ au niveau « domaine ».

$$\theta_a = g\left(\overline{Y}_a\right)$$

On postule, pour tout a :

$$\theta_a = z_a^T \cdot \beta + b_a \cdot v_a$$

où $\beta \in R^{p}$ inconnu, b_a réel connu et v_a variable aléatoire (« effet aléatoire » propre au domaine) vérifiant :

$$E(v_a) = 0 \text{ et } V(v_a) = \sigma_v^2$$

Les e_a sont supposés mutuellement indépendants dans le modèle de base (c'est plus ou moins acceptable).

A noter : il n'y a pas d'hypothèse de loi de v_a.

Par ailleurs, on a observé un estimateur (direct) de θ_a, noté $\hat{\theta}_a$

$$\hat{\theta}_a = \theta_a + e_a$$

e_a = erreur d'échantillonnage, <u>supposée sans biais</u>, de variance estimée Ψ_a .

Les e_a sont supposés mutuellement indépendants dans le modèle de base. Alors

$$\boxed{\hat{\theta}_a = z_a^T \beta + b_a v_a + e_a}$$

On considérera v_a et e_a comme indépendantes.

Le paramètre σ_v^2 témoigne d'une variabilité de type « inter » domaines : $b_a \cdot v_a$ traduit la spécificité du domaine a par rapport aux autres, hors effet de z.

Au contraire, ψ_a lié à la variabilité « intra » domaines.

<u>C'est un modèle linéaire mixte,</u>

<u>qui mêle 2 natures d'aléas</u>

On s'intéresse à

$$\theta_a = z_a^T \beta + b_a v_a$$

c'est-à-dire à $\mu_a = l_a^T \beta + m_a^T \cdot v_a$

avec $l_a = z_a$ et $b_a = m_a$

L'estimateur SBOL de θ_a est :

$$\boxed{\hat{\theta}_a^H = z_a^T \tilde{\beta} + \gamma_a \left(\hat{\theta}_a - z_a^T \tilde{\beta} \right)}$$

$$\gamma_a = \frac{b_a^2 \cdot \sigma_v^2}{\Psi_a + b_a^2 \sigma_v^2} = \frac{Variance\,(b_a v_a)}{Variance\,(b_a v_a) + Variance\,(e_a)}$$

et $$\tilde{\beta} = \left[\sum_{a=1}^{m} \frac{z_a z_a^T}{\Psi_a + b_a^2 \sigma_v^2} \right]^{-1} \cdot \left[\sum_{a=1}^{m} \frac{z_a \hat{\theta}_a}{\Psi_a + b_a^2 \cdot \sigma_v^2} \right]$$

Ici $\delta = \sigma_v^2 \quad (\delta \in \mathrm{R})$.

On peut considérer l'écriture alternative de $\hat{\theta}_a^H$:

$$\hat{\theta}_a^H = \gamma_a \cdot \underbrace{\hat{\theta}_a}_{\text{Estimateur direct}} + (1 - \gamma_a) \underbrace{z_a^T \tilde{\beta}}_{\text{Estimateur synthétique}}$$

$\gamma_a \in [0,1] \quad \Rightarrow \quad \hat{\theta}_a$ est un estimateur composite de θ_a

* Si b_a est petit ou/et si σ_v^2 est petit $\Rightarrow$ l'influence de v_a est faible $\Rightarrow$ $\hat{\theta}_a^H$ est « presque » l'estimateur synthétique.

* Si Ψ_a est faible, γ_a tend vers 1 et l'estimateur direct reprend l'avantage.

Avec seulement l'aléa de sondage, cet estimateur
- est convergent $(n_a \rightarrow N_a)$
- est biaisé !

Le biais disparaît si on considère les 2 aléas.

$$EQM\left(\hat{\theta}_a^H\right) = \gamma_a \Psi_a + (1-\gamma_a)^2 z_a^T \left(\sum_{a=1}^{m} \frac{z_a z_a^T}{\Psi_a + \sigma_v^2 b_a^2} \right)^{-1} z_a$$

$$= \gamma_a \Psi_a + O(\frac{1}{m}) \approx \gamma_a \Psi_a \quad \text{si } m \text{ grand}$$

$$\Rightarrow \quad \boxed{\frac{EQM(\hat{\theta}_a^H)}{EQM(\hat{\theta}_a)} \cong \gamma_a}$$

<u>Conclusion</u> : <u>**si γ_a petit $\Rightarrow$ gain important**</u>.

Comme σ_v^2 est inconnu, il faut l'estimer et utiliser l'estimateur ESBOL : on sait faire (moments, EMV en postulant une loi,…), mais c'est assez compliqué.

Les estimateurs SBOL et ESBOL ont néanmoins le mérite d'associer « harmonieusement » les deux grandes approches de l'estimation / prédiction :

- l'approche sondage (pas de dépendance envers un modèle de comportement)
- l'approche classique par modélisation (le modèle de comportement est déterminant).

<u>en donnant priorité à celle qui semble la plus fiable</u>.

<u>NOTA</u> : On peut étendre toute cette méthodologie à des modèles plus sophistiqués, par exemple :

- *Modèles de corrélation spatiale :*

$$Cov\left(v_a, v_b\right) = \alpha \cdot e^{-\beta . d_{ab}} \qquad \left(\alpha, \beta\right) \in \mathrm{R}^2$$

Autre approche : Ω_a étant un « voisinage » de a,

$$v_a \left| \left\{v_b, b \neq a\right\} \rightarrow \mathcal{N}\left(\rho \cdot \sum_{b \in \Omega_a} v_b \, , \sigma_v^2 \right)\right.$$

- *Modèles temporels*

$$\begin{cases} \hat{\theta}_{at} = \theta_{at} + e_{at} \\ \theta_{at} = g\left(\overline{Y}_{at}\right) = Z_{at}^{T}\beta + b_a v_a + u_{at} \end{cases}$$

avec $u_{at} = \rho \cdot u_{a,t-1} + \varepsilon_{at}$

C) Application au cas d'une modélisation au niveau individuel

$$Y_{a,i} = X_{a,i}^{T} \cdot \beta + v_a + e_{a,i}$$

avec $E\left(e_{a,i}\right) = 0$ et $V\left(e_{a,i}\right) = \left(k_{a,i}\right)^2 \cdot \sigma_e^2$

En écriture matricielle, pour tout a (de 1 à m) :

$$\boxed{Y_a = X_a \beta + v_a \cdot 1_{n_a} + e_a}$$

Il faut distinguer deux cas de figure selon la valeur N_a :

CAS 1 :

C'est le cas favorable où N_a est grand : dans ce cas

$$\bar{e}_a = \frac{1}{N_a} \sum_{i=1}^{N_a} e_{a,i} \cong 0.$$

Donc le paramètre à estimer est le réel

$$\overline{Y}_a = \mu_a \approx \overline{X}_a^T \beta + v_a$$

où
$$\overline{X}_a = \frac{1}{N_a} \sum_{i=1}^{N_a} X_{a,i} \,.$$

On retrouve le modèle au niveau domaine !
L'estimateur SBOL de $\overline{Y}_a$ est

$$\hat{\mu}_a^H = \overline{X}_a^T \tilde{\beta} + \tilde{v}_a$$

Après calculs, on aboutit à :

$$\boxed{\hat{\mu}_a^H = \gamma_a \left[\bar{y}_a^\lambda + \left(\bar{X}_a - \bar{x}_a^\lambda \right)^T \tilde{\beta} \right] + (1-\gamma_a) \bar{X}_a^T \tilde{\beta}}$$

Partie (pseudo) directe   Partie synthétique

où

$$\tilde{\beta} = \left(\sum_{i \in s_a} \lambda_{ai} \cdot X_{a,i} X_{a,i}^T - \gamma_a \cdot \lambda_a \cdot \bar{x}_a^\lambda \cdot \bar{x}_a^{\lambda T} \right)^{-1} \cdot \left(\sum_{i \in s_a} \lambda_{a,i} X_{a,i} \cdot Y_{a,i} - \gamma_a \lambda_a \bar{x}_a^\lambda \cdot \bar{y}_a^\lambda \right)$$

$$\bar{x}_a^\lambda = \frac{1}{\lambda_{a\bullet}} \sum_{i \in s_a} \lambda_{a,i} \cdot X_{a,i}$$

$$\bar{y}_a^\lambda = \frac{1}{\lambda_{a\bullet}} \sum_{i \in s_a} \lambda_{a,i} \cdot Y_{a,i}$$

$$\gamma_a = \frac{\sigma_v^2}{\sigma_v^2 + \dfrac{\sigma_e^2}{\lambda_{a\bullet}}}$$

avec $\lambda_{ai} = 1 / k_{a,i}^2$ et $\lambda_{a\bullet} = \sum_{i \in s_a} \lambda_{ai}$

σ_v^2 petit $\Rightarrow$ modèle bien ajusté $\Rightarrow$ partie synthétique prime

$\lambda_{a\bullet}$ grand $\Leftrightarrow \approx n_a$ grand $\Rightarrow$ partie directe prime

Avec seulement l'aléa de sondage, si $k_{a,i} \neq 1$, cet estimateur n'a pas de bonnes propriétés (biais, non convergent) : mais les poids de sondage n'interviennent pas…

C'est bien **<u>dépendant du modèle</u>**

<u>CAS 2 :</u>

N_a est « petit ». On a

$$\overline{Y}_a = f_a \cdot \overline{y}_a + (1 - f_a)\overline{y}_a^{*}$$

où $\overline{y}_a^{*}$ = moyenne des $Y_{a,i}$ sur tous les individus a qui ne sont PAS dans s_a.

Il faut alors <u>prédire</u> $\overline{y}_a^{*}$, selon :

$$\hat{\overline{y}}_a^{*H} = \frac{1}{N_a - n_a} \sum_{i \notin s_a} (X_{a,i}^T \tilde{\beta} + \tilde{v}_a) = \overline{X}_a^{*T} \tilde{\beta} + \tilde{v}_a$$

$$\overline{X}_a^{*} = \frac{N_a \overline{X}_a - n_a \overline{x}_a}{N_a - n_a} \quad : \quad \text{valeur connue}$$

On obtient, après calculs :

$$\boxed{\hat{\overline{Y}}_a^{H} = f_a \overline{y}_a + (1 - f_a)\left[\gamma_a \left(\overline{y}_a^{\lambda} + \left(\overline{X}_a^{*} - \overline{x}_a^{\lambda}\right)^T \tilde{\beta}\right) + (1 - \gamma_a)\overline{X}_a^{*T} \cdot \tilde{\beta}\right]}$$

connu direct synthétique

Il faut in fine estimer σ_e^2 et σ_v^2 (très compliqué - mais la PROC MIXED de SAS donne les EMV) et aboutir à l'ESBOL.

II) Les estimateurs optimaux

A) Principe général et démarche d'ensemble :

On veut :

- que l'optimum soit « absolu » ;
- que la théorie s'applique aussi aux variables qualitatives ;

C'est possible avec une modélisation, mais en contrepartie il faut une **hypothèse sur les lois de e et de v.**

* Théorème central (rappel):

Pour prédire une variable aléatoire μ au moyen d'une variable aléatoire Y, le prédicteur optimum au sens de l'erreur quadratique est

$$f(Y) = E[\mu|Y]$$

Alors pour tout g, on a $E[g(Y)-\mu]^2 \geq E[f(Y)-\mu]^2$

Si $Y = X\beta + Zv + e$, si $\mu = l^T \cdot \beta + m^T \cdot v$
(β est fixe, inconnu, mais v est une variable aléatoire) et que (v,e) suit une loi de Gauss, alors on trouve

$$f(Y) = l^T \beta + m^T G Z^T V^{-1}(Y - X\beta)$$

* <u>Démarche à suivre</u>

a/ Soit

- La densité de μ : $f(\mu\,;\lambda_2)$
- La densité de Y sachant μ : $f(Y|\mu\,;\lambda_1)$

b/ Par la formule de Bayes :

$$f(\mu|Y\,;\lambda_1,\lambda_2)=\frac{f(Y|\mu\,;\lambda_1)\cdot f(\mu\,;\lambda_2)}{\int f(Y|\mu;\lambda_1)\cdot f(\mu\,;\lambda_2)d\mu}$$

c/ Puis

$$\hat{\mu}^{OPTI}=E(\mu|Y\,;\lambda_1,\lambda_2)=\int\mu\cdot f(\mu|Y\,;\lambda_1,\lambda_2)d\mu$$

C'est le prédicteur optimum théorique (incalculable en général car on ne connaît pas λ_1 ni λ_2).

d/ On estime (λ_1,λ_2) à partir de

$$f(Y\,;\lambda_1,\lambda_2)=\int f(Y|\mu\,;\lambda_1)\cdot f(\mu\,;\lambda_2)d\mu$$

par une méthode quelconque (par exemple le maximum de vraisemblance).

e/ On obtient $(\hat{\lambda}_1,\hat{\lambda}_2)$ et on termine en calculant

$$E(\mu|Y,\hat{\lambda}_1,\hat{\lambda}_2) = \hat{\mu}_E^{OPTI}$$

dit abusivement estimateur « Bayésien empirique ».

B) Application au cas des variables qualitatives : paramètre de type « proportion ».

<u>Objectif</u> : estimer des vraies **<u>proportions</u>** P_a qui traduisent l'importance d'une sous population D (proportion de chômeurs dans la ZUS a, par exemple).

Soit : $Y_{a,i}$ =1 si $(a,i) \in D$
=0 sinon

On distingue :

- Soit (cas 1) $Y_{a,i} \rightarrow \mathcal{B}(1, P_a)$
- Soit (cas 2) $Y_{a,i} \rightarrow \mathcal{B}(1, P_{a,i})$

On aura donc bien compris que les lois de Gauss en sont pas adaptées au contexte !!!

<u>Limitons nous au cas 1</u>

On suppose les $Y_{a,i}$ indépendants d'un individu à l'autre dans un domaine donné.

$$Y_a = \sum_{i \in s_a} Y_{a,i} \qquad \Rightarrow \qquad Y_a \rightarrow \mathcal{B}(n_a, P_a)$$

$\mu = P_a$ (à prédire) et $Y = Y_a$ (observé)

<u>Nota</u> : pas de poids de sondage : peu importe la méthode de tirage de s_a.

A partir de là, on peut imaginer

nombre de modèles sur P_a *!!!*

- <u>Exemple 1</u> : P_a suit une <u>**loi bêta**</u> (α, β)

$$f(P_a \; ; \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} P_a^{\alpha-1}(1-P_a)^{\beta-1}$$

Comme on a

$$f(Y_a | P_a) = \binom{n_a}{Y_a} P_a^{Y_a} \cdot (1-P_a)^{n-Y_a}$$

on déduit   (étape b /)

$$f(P_a | Y_a \; ; \alpha, \beta) = \frac{\Gamma(\alpha+\beta+n_a)}{\Gamma(\alpha+Y_a)\Gamma(n_a-Y_a+\beta)} P_a^{\alpha+Y_a-1}(1-P_a)^{n_a-Y_a+\beta-1}$$

Prédicteur optimum (étape c /) :

$$\hat{P}_a^{OPTI} = E\left(P_a \middle| Y_a \; ; \alpha, \beta\right) = \frac{Y_a + \alpha}{n_a + \alpha + \beta}$$

L'étape d / fournit la loi marginale de Y_a (« bêta ») :

$$f\left(Y_a \; ; \alpha, \beta\right) = \binom{n_a}{Y_a} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+Y_a)\cdot\Gamma(\beta+n_a-Y_a)}{\Gamma(\alpha+\beta+n_a)}$$

D'où $\hat{\alpha}$ et $\hat{\beta}$, estimateurs du maximum de vraisemblance.
Pas de solutions analytiques $\Rightarrow$ utilisation d'algorithmes convergents. On aboutit au prédicteur optimum empirique :

$$\hat{P}_{a,E}^{OPTI} = \frac{Y_a + \hat{\alpha}}{n_a + \hat{\alpha} + \hat{\beta}}$$

Il est possible aussi d'obtenir des estimateurs de α et β selon la méthode des moments. Dans ce cas, on vérifie que la solution conduit à :

$$\hat{P}_{a,E}^{OPTI} = \hat{\gamma}_a \cdot \hat{p}_a + \left(1 - \hat{\gamma}_a\right) \cdot \hat{p}$$

Direct Synthétique

où $$\hat{p} = \sum_{a=1}^{m} \frac{n_a}{n} \cdot \hat{p}_a$$

$$\hat{\gamma}_a = \frac{n_a}{n_a + \hat{\alpha} + \hat{\beta}} \quad \in [0,1].$$

<u>Nota</u> : la variance d'échantillonnage n'intervient jamais dans $\hat{\gamma}_a$, ce qui est un atout très appréciable (et qui n'avait pas lieu avec les modèles linéaires mixtes).

- <u>Exemple 2</u> : P_a suit une loi LOGIT

$$Log\,\frac{P_a}{1-P_a} = \mu + v_a$$

où v_a suit une loi $\mathcal{N}\left(0,\sigma^2\right)$

PAS d' expression analytique de $\hat{P}_a^{OPTI}$!!!

On peut néanmoins traiter la question par une méthode approchée s'appuyant sur des simulations et sur la loi des grands nombres.

C) Application au cas des risques relatifs

On cherche à prédire le paramètre (« risque relatif »)

$$\theta_a = \frac{P_a}{P}$$

P_a = proportion d'une population donnée au sein du domaine ;
P = proportion d'une population donnée au sein de la population entière U.

L'effectif Y_a est observé dans l'échantillon. On note

$$\tau_a = n_a \cdot \frac{\sum_{a=1}^{m} Y_a}{\sum_{a=1}^{m} n_a} \quad \text{(proche de } n_a.P\text{)}$$

τ_a est considéré comme fixé.

On suppose $Y_a | \theta_a \rightarrow Poisson(\tau_a.\theta_a)$

On suppose $\theta_a \rightarrow Gamma(\alpha, \beta)$, soit

$$f(\theta_a \; ; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \theta_a^{\alpha - 1} (1 - \theta_a)^{\beta - 1}$$

Alors on montre $\theta_a | Y_a \rightarrow Gamma(\alpha + Y_a, \beta + \tau_a)$. D'où

$$\hat{\theta}_{a,E}^{OPTI} = \frac{Y_a + \hat{\alpha}}{f_a + \hat{\beta}}$$

après estimation de α de β. Si on utilise des estimateurs des moments « bien choisis », on obtient :

$$\hat{\theta}_{a,E}^{OPTI} = \hat{\gamma}_a \cdot \hat{\theta}_a + (1 - \hat{\gamma}_a) \cdot \hat{\theta}$$

Direct Synthétique

où $$\hat{\theta} = \frac{\sum_{a=1}^{m} \tau_a . \hat{\theta}_a}{\sum_{a=1}^{m} \tau_a}$$

$$\hat{\gamma}_a = \frac{\tau_a}{\tau_a + \hat{\alpha}} \quad \in [0,1].$$

D) Retour au cas du modèle de Fay et Herrict

On reprend le modèle de Fay et Herriot en introduisant une hypothèse de normalité, soit :

$$\hat{\theta}_a = \theta_a + e_a \quad \text{avec } e_a \to \mathcal{N}(0, \Psi_a)$$
$$\theta_a = z_a^T \beta + b_a v_a \quad \text{avec } v_a \to \mathcal{N}(0, \sigma_v^2)$$

On vérifie

$$f\left(\theta_a \middle| \hat{\theta}_a \; ; \beta, \sigma_v^2\right) \to \mathcal{N}\left(\hat{\theta}_a^{OPTI}, \gamma_a \Psi_a\right)$$

où $$\gamma_a = \frac{b_a^2 \sigma_v^2}{b_a^2 \sigma_v^2 + \Psi_a}$$

Ψ_a est supposé connu, et

$$\hat{\theta}_a^{OPTI} = E\left[\theta_a \middle| \hat{\theta}_a ;\beta,\sigma_v^2\right] = \gamma_a \hat{\theta}_a + (1-\gamma_a) z_a^T \beta \cdot$$

Pour estimer β et σ_v^2 on utilise :

$$\hat{\theta}_a \to \mathcal{N}\left(z_a^T \beta, b_a^2 \sigma_v^2 + \Psi_a\right)$$

Par EMV, on obtient $\hat{\beta}$ et $\hat{\sigma}_v^2$. Finalement :

$$\boxed{\hat{\theta}_{a,E}^{OPTI} = \hat{\gamma}_a \hat{\theta}_a + (1-\hat{\gamma}_a) z_a^T \hat{\beta}}$$

C’est le même estimateur que l’ESBOL (donc il y a une robustesse à l’hypothèse de normalité) !

III) Les estimateurs « Bayésiens hiérarchiques » :

Approche partant de la précédente, et ajoutant une étape : on suppose que les paramètres (λ_1, λ_2) suivent une loi que l'on se fixe A PRIORI, soit $f(\lambda_1, \lambda_2)$.

Par Bayes :

$$f(\mu, \lambda_1, \lambda_2 | Y) = \frac{f(Y, \mu \mid \lambda_1, \lambda_2) \cdot f(\lambda_1, \lambda_2)}{\int f(Y, \mu | \lambda_1, \lambda_2) \cdot f(\lambda_1, \lambda_2) d\mu d\lambda}$$

$f(Y, \mu \mid \lambda_1, \lambda_2)$: voir estimateurs optimaux

Dénominateur : c'est la densité de Y : généralement incalculable analytiquement ! $\Rightarrow$ utiliser des techniques de simulation par chaînes de Markov (algorithmes de Gibbs, algorithme de Metropolis) .

Finalement, on obtient la densité « A POSTERIORI »

$$f(\mu | Y) = \int f(\mu, \lambda_1, \lambda_2 | Y) d\lambda$$

puis on termine par

$$E(\mu | Y) = \hat{\mu}^{HB}$$

dit estimateur « Bayésien hiérarchique ».

Conclusion

L'estimation sur petits domaines :

- Reste dépendante de modèles : pas de miracle ! Comme on a peu d'information locale, il faut compter sur des modèles pour aller en chercher ailleurs…
- La qualité des hypothèses portées par les modèles est (très) difficile à apprécier. C'est néanmoins plus facile avec une approche par modélisation explicite (critères de choix de modèles et critères de qualité).
- Semble mal prise en compte par les logiciels !
- On a très peu d'expérience en France sur ces questions plutôt complexes.

****