Algorithme EM. Mélanges gaussiens
Modèle linéaire Mixte.

Contexte : - Publi JRSS B. 1977. 39(1).
MLE from Incomplete data via the
EM Algorithm.

→ Proposent un algo itératif de max. de Vrais.
pour modèles à observations incomplets.

→ Applications : - mélange (Gaussien ou pas) / clustering
- HMM (Forward - Backward).
- donnés tronquées, censurés.
- estimat° de genotypes
- mod. linéaire mixte.

→ Ingrédient de base pour tte une famille d'algo.
(ex : Stat. Bayesienne, SAEM, SEM ...).

Exemple du clustering :

- On observe $(y_1 \text{---} y_n)$, on suppose ∃ K groups inconnus.

↳ les labels ne sont pas obs.

→ les données sont incomplets.

- On cherche à retrouver les labels $(z_1 — z_n)$ mais on ne dispose que des $(y_1 — y_n)$
- Données Complètes : (Y, Z)
- Si on avait (Y, Z) alors on aurait tts les infos.

$$Y_i \mid Z_i = k \sim \mathcal{N}(\mu_k, \sigma^2)$$

$$Z_i \sim \mathcal{M}(1, \pi_1 \dots \pi_K)$$

↳ loi Conditionnelle, très facile à écrire, formuler, calculer.

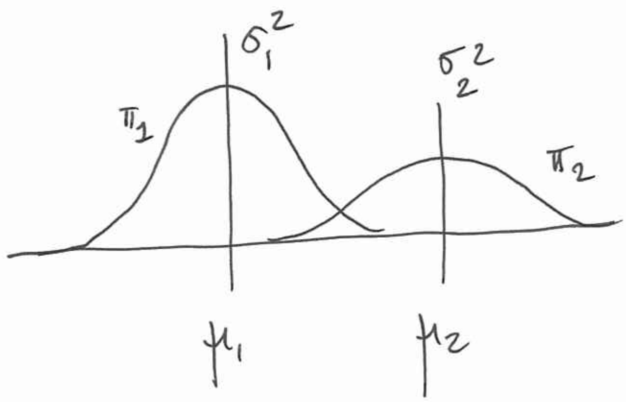
$$f(y, z; \theta) = \underbrace{f(y \mid z; \theta)}_{\downarrow \ \mathcal{N}} \ \underbrace{f(z; \theta)}_{\downarrow \ \mathcal{M}}$$

- Paramètres : $(\mu_1 \dots \mu_K)$: moyenne par Groupe.

σ^2 : dispersion

$(\pi_1 — \pi_K)$: taille des Groupes.

$$\theta : (\underline{\mu}, \sigma^2, \underline{\pi}).$$



- Estimation par Max. de Vrai.

$$\log \mathcal{L}(Y;\theta) = \sum_i \log f(y_i;\theta)$$
$$= \sum_i \log \left[\sum_{k=1}^{K} f(y_i | Z_i = k;\theta_k)\, \mathbb{P}(Z_i = k) \right].$$
$$= \sum_i \log \left(\sum_{k=1}^{K} \pi_k f(y_i;\theta_k) \right).$$

↳ Si on veut MLE alors il faut calculer.

$$\frac{\partial \log \mathcal{L}(Y;\theta)}{\partial \theta} = 0.$$

↳ Comment faire numériquem^t ?

↳ EA si on veut aussi les labels ?.

↳ Dans certains modèls, la Vrais marginale ne peut pas être calculée direct^t (ex HMM).

→ Modèle de classif. probabiliste dont les paramètres s'estiment par EM (lien avec k-means).

Exemple du Modèle linéaire mixte.

$$Y = X\beta + ZU + E \qquad V(Y) = ZGZ' + R$$

$$\beta \in \mathbb{R}^p, \quad U \in \mathbb{R}^q$$

$$E \perp U$$

$$Y \sim N(X\beta, V) \qquad U \sim N(0, G)$$

$$E \sim N(0, R).$$

→ On cherche à estimer (β, G, R) et à prédire U.
en ayant observé que Y.

- loi marginale de $Y \sim N(X\beta, V)$
- loi Conditionnelle $Y|U \sim N(X\beta + ZU, R)$

$$f(y;\theta) = \frac{1}{(2\pi)^{N/2}|V|^{1/2}} \exp\left(-\frac{1}{2}\|Y - X\beta\|^2_{V^{-1}}\right)$$

$$f(y|u;\theta) = \frac{1}{(2\pi)^{N/2}|R|^{1/2}} \exp\left(-\frac{1}{2}\|Y - X\beta - ZU\|^2_{R^{-1}}\right).$$

On prendra : $R = \sigma_\varepsilon^2 I_N$, $G = \sigma_u^2 I_q$.

et le modèle $Y_{ij} = \mu + U_i + E_{ij}$ $\quad j = 1, n_0$, $i = 1, I$.

↳ On veut estimer & prédire sans utiliser V^{-1}

Définition des vraisemblances.

- Vrais. incomplète ou marginale.

↳ c'est la densité marginale des obs.

$$\log \mathcal{L}(Y;\theta) = \sum_i \log f(y_i;\theta)$$ si les ind sont ⫫.

- Vrais. des données complètes.

↳ c'est la densité jointe des obs. & var. cachées.

$$\log \mathcal{L}(Y,Z;\theta) = \underbrace{\log \mathcal{L}(Y|Z;\theta)}_{\substack{\text{souvent facile}\\ \text{à formuler}\\ \text{calculer}}} + \log \mathcal{L}(Z;\theta).$$

Ex : <u>mélange Gaussien</u> $Y_i | Z_i = k \sim N(\mu_k, \sigma^2)$

$$\log \mathcal{L}(Y,Z;\theta) = \log \left[\prod_{i=1}^{n} f(y_i, z_i;\theta) \right]$$

$$= \log \left[\prod_i \prod_k \left[f(y_i | z_i = k;\theta) f(z_i = k) \right]^{\mathbb{1}(z_i=k)} \right].$$

$$= \log \left(\prod_{i,k} \left[\pi_k \phi(x_i, \mu_k, \sigma^2) \right]^{\mathbb{1}(z_i=k)} \right)$$

On reparamétrise : $Z_{ik} = 1$ si $i \in k$.

$$\log \mathcal{L}(Y, Z; \theta) = \sum_{ik} Z_{ik} \log \left(\pi_k \, \phi(y_i \mid \mu_k, \sigma^2) \right)$$

Ex. du modèle mixte.

$$\log \mathcal{L}(Y, U; \theta) = \log \mathcal{L}(Y \mid U; \theta) + \log \mathcal{L}(U; \theta)$$

$$-2 \log \mathcal{L}(Y \mid U; \theta) = N \log (2\pi) + \log |R| + \| Y - X\beta - ZU \|^2_{R^{-1}}$$
$$= N \log 2\pi + N \log \sigma_\varepsilon^2 + \frac{1}{\sigma_\varepsilon^2} \| Y - X\beta - ZU \|^2.$$

$$-2 \log \mathcal{L}(U; \theta) = q \log 2\pi + q \log \sigma_u^2 + \frac{1}{\sigma_u^2} \| u \|^2.$$

Idée d'EM : Maximiser une fonction de $\log \mathcal{L}(Y, U; \theta)$ pour maximiser (indirect) $\log \mathcal{L}(Y)$.

Maximisation de la vraisemb. complète.

$$\frac{\partial \log \mathcal{L}(Y, Z; \theta)}{\partial \theta} = 0.$$

↳ si on connaissait les var. cachées.

1). Mélange gaussien :

$$\frac{\partial \log \mathcal{L}(Y, Z; \theta)}{\partial \mu_k} = \sum_{i=1}^{n} Z_{ik} \frac{\partial \log f(y_i, \mu_k, \sigma^2)}{\partial \mu_k}.$$

$$= \sum_{i=1}^{n} Z_{ik} \frac{(y_i - \mu_k)}{\sigma^2} = 0$$

$$\hookrightarrow \hat{\mu}_k = \frac{\sum Z_{ik} y_i}{\sum Z_{ik}}.$$

2). Modèles mixtes

$$\frac{\partial \log \mathcal{L}(Y, U; \theta)}{\partial \beta} = \sum_{i=1}^{n} \frac{1}{\sigma_\varepsilon^2} x_i (Y_i - X_i\beta - Z_i U_i)$$

$$\tilde{Y}_i = Y_i - Z_i U_i.$$

$$\rightarrow \hat{\beta} = \sum_{i=1}^{n} ({}^t x_i x_i)^{-1} \, {}^t x_i \tilde{Y}_i$$

car $R^{-1} = 1/\sigma_\varepsilon^2 I$ facile à inverser !.

- Comment relier la maximisation de la vrais. complète (facile) à l'incomplète ?

- Comment prendre en compte le fait qu'on n'observe pas les var. cachées ?.

- Ingrédient : $\mathbb{E}_\theta(Z|Y)$.

$$\mathbb{E}_\theta(Z|Y) = \int z\, f(z|y;\theta)\, dz$$

$$\mathbb{E}(g(Z)\,|Y) = \int g(z)\, f(z|y;\theta)\, dz.$$

loi param. de la sous laquelle on intègre

↳ la vrais. est une fonction des Z et des Y.

$$\mathbb{E}_\theta\left(\log \mathcal{L}(Y, Z; \theta)\,\middle|\, Y\right)$$

$$= \int \log \mathcal{L}(Y, Z, \theta) \times f(Z|Y;\theta)\, dZ$$

On va maximiser l'esp. cond. de la vrais. complète des obs.

$$\mathbb{E}_{\theta}\left(\log \mathcal{L}(Y,Z;\theta) \,\middle|\, Y \right) = \int \log \mathcal{L}(Y,Z;\theta) \times f(Z \mid Y;\theta)\, dZ$$

paramètre sous lequel on calcule l'espérance

paramètre inconnu.

⇒ <u>Algorithme d'optimisation itératif :</u>

- si on fixe une valeur courante du paramètre $\theta = \theta^{(t)}$.

$$\mathbb{E}_{\theta^{(t)}}\left(\log \mathcal{L}(Y,Z;\theta) \,\middle|\, Y \right) = Q\left(\theta, \theta^{(t)}\right)$$

- <u>Identité de Fisher.</u>

$$\frac{\partial \log \mathcal{L}(Y;\theta)}{\partial \theta} = \mathbb{E}_{\theta}\left(\frac{\partial \log f(y,z;\theta)}{\partial \theta} \,\middle|\, Y=y \right).$$

<u>Demonstration</u> :

$$\frac{\partial \log \mathcal{L}(Y;\theta)}{\partial \theta} = \frac{\partial \mathcal{L}(Y;\theta)}{\partial \theta} \times \frac{1}{\mathcal{L}(Y;\theta)}$$

$$\mathcal{L}(Y;\theta) = \int \mathcal{L}(Y,Z;\theta)\, dZ$$

$$\frac{\partial \mathcal{L}(Y;\theta)}{\partial \theta} = \int \frac{\partial \mathcal{L}(Y;Z;\theta)}{\partial \theta}\, dZ = \int \mathcal{L}(Y,Z;\theta) \times \frac{\partial \log \mathcal{L}(Y,Z;\theta)}{\partial \theta}\, dZ$$

- $$\frac{\partial \log \mathcal{L}(Y,Z;\theta)}{\partial \theta} = \frac{\partial \mathcal{L}((Y,Z);\theta)}{\partial \theta} \times \frac{1}{\mathcal{L}(Y,Z;\theta)}$$

- $$\mathcal{L}(Y,Z;\theta) = \mathcal{L}(Y;\theta) \times \mathcal{L}(Z|Y;\theta).$$

$$\hookrightarrow \frac{\partial \log \mathcal{L}(Y;\theta)}{\partial \theta} = \frac{1}{\mathcal{L}(Y;\theta)} \int \mathcal{L}(Y;\theta) \times \mathcal{L}(Z|Y;\theta) \times \frac{\partial \log \mathcal{L}(Y,Z;\theta)}{\partial \theta}\, dZ.$$

$$= \int \mathcal{L}(Z|Y;\theta) \times \frac{\partial \log \mathcal{L}(Y,Z;\theta)}{\partial \theta}\, dZ.$$

$\hookrightarrow$ OK

• Si on fixe une valeur courante du paramètre :

$$\mathbb{E}_{\theta^{(t)}}\left(\frac{\partial}{\partial\theta} \log \mathcal{L}(Y,Z;\theta) \,\middle|\, Y\right)$$

$$= \int \underbrace{\left[\frac{\partial}{\partial\theta} \log \mathcal{L}(Y,Z;\theta)\right]}_{\text{fonct}^\circ \text{ de } \theta} \times \underbrace{\mathcal{L}(z \mid y ; \theta^{(t)})}_{\text{ce n'est plus une fonct}^\circ \text{ de } \theta \,.} dz$$

$$\Rightarrow \frac{\partial}{\partial\theta} \int \log \mathcal{L}(Y,Z;\theta) \times \mathcal{L}(Z \mid Y ; \theta^{(t)}) \, dZ \,.$$

$$= \frac{\partial}{\partial\theta}\left[\mathbb{E}_{\theta^{(t)}}\left(\log \mathcal{L}(Y,Z;\theta) \,\middle|\, Y\right)\right].$$

Exemple du mélange Gaussien :

$$\log \mathcal{L}(Y,Z;\theta) = \sum_{ik} Z_{ik} \log \pi_k + \sum_{ik} Z_{ik} \log f(y_i \mid Z_i = k ; \theta_k)$$

$$\mathbb{E}_{\theta^{(t)}}\left(\log \mathcal{L}(Y,Z;\theta) \,\middle|\, Y\right) = \sum_{ik} \mathbb{E}_{\theta^{(t)}}(Z_{ik} \mid Y_i) \log \pi_k$$

$$+ \sum_{ik} \mathbb{E}(Z_{ik} \mid Y_i) \log \text{——} .$$

On note $\tau_{ik} = \mathbb{E}_{\theta^{(t)}}(Z_{ik} | Y_i) = \mathbb{P}_{\theta^{(t)}}(Z_{ik} = 1 | Y_i)$

$$= \frac{\mathbb{P}(Z_{ik} = 1) \times \mathbb{P}(Y_i | Z_{ik} = 1)}{\sum \pi_\ell^{(t)} f(y_i ; \theta_\ell^{(t)})}$$

$$= \frac{\pi_k^{(t)} f(y_i ; \theta_k^{(t)})}{\sum_\ell \pi_\ell f_\ell}$$

$$Q(\theta, \theta^{(t)}) = \sum_{ik} \tau_{ik} \log \pi_k + \sum_{ik} \tau_{ik} \log f(y_i ; \theta_k).$$

↳ $\frac{\partial Q}{\partial \theta} = 0 \Rightarrow \widehat{\mu_k} = \frac{\sum \tau_{ik} y_i}{\sum \tau_{ik}}$

→ Estimation sous la forme d'une moyenne pondérée.

↳ <u>Algorithme itératif :</u>

1)- Calcul de l'espérance : Expectat° Step

2)- Maximisat° de $Q(\theta, \theta')$

Algorithme EM pour les modèles linéaires mixtes.

$$\log \mathcal{L}(Y,U;\theta) = \log \mathcal{L}(Y|U;\theta) + \log \mathcal{L}(U;\theta)$$

Etape E :
$$\mathbb{E}_{\theta^{(t)}}\left(\log \mathcal{L}(Y,U,\theta) \middle| Y\right) = Q(\theta,\theta^{(t)})$$
$$= Q_E(\theta,\theta^{(t)}) + Q_U(\theta,\theta^{(t)}).$$

$$Q_U(\theta,\theta^{(t)}) = -\frac{1}{2}\left[q\log 2\pi + q\log\sigma_u^2 + \mathbb{E}_{\theta^{(t)}}(U'U|Y)/\sigma_u^2\right]$$

↳ Calcul de $\mathbb{E}_{\theta^{(t)}}(U'U|Y)$

Etape M :
$$\frac{\partial Q_u}{\partial \theta} = 0 \Rightarrow q/\sigma_u^2 = \frac{1}{\sigma_u^4}\mathbb{E}(U'U|Y) = 0.$$

$$\Rightarrow \boxed{\sigma_u^2 = \frac{1}{q}\mathbb{E}(U'U|Y)}$$

↳ Reste à calculer $\mathbb{E}_\theta(U'U|Y)$.

→ On utilise l'identité.

$$\mathbb{E}(U'U|Y) = \underbrace{\mathbb{E}(U|Y)'}_{\text{BLUP}}\mathbb{E}(U|Y) + tr\,(\underbrace{V(U|Y)}))$$

$$V(U|Y) = (Z'R^{-1}Z + G^{-1})^{-1} = \sigma_\varepsilon^2(Z'Z+\lambda I)^{-1}$$

BLUP → $\hat{u}^{(t)}$

$$V(U|Y) = V(\hat{U}-U) = V(U) - V(\hat{U})$$

par les prop. du BLUP.

Or
$$V(\hat{U}) = \underbrace{G Z' V^{-1}} Z G$$
$$= \left(\left(Z'R^{-1}Z + G^{-1}\right)^{-1} Z'R^{-1}\right) \times Z G.$$
$$= \left(Z'Z + \frac{\sigma_\varepsilon^2}{\sigma_u^2} I_q\right)^{-1} Z'Z \times \sigma_u^2 .$$

Car
$$Z'R^{-1}Z + G^{-1} = \frac{1}{\sigma_\varepsilon^2} Z'Z + \frac{1}{\sigma_u^2} I_q$$
$$= \frac{1}{\sigma_\varepsilon^2}\left(Z'Z + \frac{\sigma_\varepsilon^2}{\sigma_u^2} I_q\right).$$
$$\left(Z'R^{-1}Z + G^{-1}\right)^{-1} = \sigma_\varepsilon^2 \left(Z'Z + \frac{\sigma_a^2}{\sigma_\varepsilon^2} I_q\right)^{-1}.$$

$$V(U|Y) = \sigma_u^2 I_q - \sigma_u^2 \left(Z'Z + \lambda I_q\right)^{-1} Z'Z = V(U) - V(\hat{U})$$
$$= V(U - \hat{U}).$$

On peut montrer (Searle p 298
Foulley p 135)

$$tr\left(\left(Z'Z + \lambda I\right)^{-1} Z'Z\right) = q - \lambda\, tr\left(\left(Z'Z + \lambda I_q\right)^{-1}\right)$$

↳
$$\boxed{tr\left(V(U|Y)\right) = \sigma_\varepsilon^2\, tr\left[\left(Z'Z + \lambda I_q\right)^{-1}\right]}.$$

On le trouve directement avec
$$V(U|Y) = \left(Z'R^{-1}Z + G^{-1}\right)^{-1}.$$

$$\sigma_1^2 = \frac{1}{q} \underset{\theta^{(t)}}{\mathbb{E}}\left(\|U\|^2 \mid Y\right) = \frac{1}{q}\left[\hat{U}^{(t)\prime} \hat{U}^{(t)} + tr\left(V(U|Y)\right)\right].$$

$$\boxed{\sigma_1^2 = \frac{1}{q}\left(\hat{u}^{(t)\prime} \hat{u}^{(t)} + \sigma_\varepsilon^2 \, tr\left(\left[Z'Z + \lambda I_q\right]^{-1}\right)\right)}$$

↳ Il faut maintenant Calculer $\hat{u}$.

→ Mais on est dans le cadre Gaussien donc $E(U|Y)$ est connu.

$$\mathbb{E}(U|Y) = GZ'V^{-1}(Y - X\beta)$$

mais d'après les resultats d'Henderson

$$\underset{\theta^{(t)}}{\mathbb{E}}(U|Y) = \left(Z'Z + \lambda^{(t)} I\right)^{-1} Z'\left(Y - X\beta^{(t)}\right)$$

à l'étape E, θ est fixé @ $\theta^{(t)}$

↳ $\beta^{(t)}$, $\lambda^{(t)} = \dfrac{\sigma_\varepsilon^{2\,(t)}}{\sigma_u^{2\,(t)}}$ sont fixés.

Dans le modèle $Y_{ij} = \mu + u_i + \varepsilon_{ij}$.

$i = 1, I$, $j = 1, n_0$, $u_i \sim N(0, \sigma_u^2)$
$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$.

$X\beta = \mathbb{1}_N \mu$, $Z = Id_I \otimes \mathbb{1}_{n_0}$.

→ On veut calculer $\hat{u}_i$:

$$(Z_i' Z_i + \lambda I)^{-1} = (n_0 + \lambda)^{-1} I_{n_0} = \frac{\sigma_u^2}{n_0 \sigma_u^2 + \sigma_\varepsilon^2} I_{n_0}.$$

$$Z_i' (Y_i - \mathbb{1}_{n_0} \mu) = n_0 (y_{i\bullet} - \hat{\mu}) = \sum_{j=1}^{n_0} (y_{ij} - \hat{\mu}).$$

$$\hookrightarrow \hat{u}_i = \underbrace{\frac{n_0 \sigma_u^2}{n_0 \sigma_u^2 + \sigma_\varepsilon^2}}_{\text{on retrouve bien le poids en } \mathbb{V}(Y_{i\bullet}) = \sigma_u^2 + \frac{1}{n_0} \sigma_\varepsilon^2} (y_{i\bullet} - \hat{\mu})$$

$$\sigma_u^2 = \frac{1}{I} \left(\sum_i \hat{u}_i^2 + \sigma_\varepsilon^2 \, tr (Z'Z + \lambda I)^{-1} \right).$$

$$tr \left[(Z'Z + \lambda I)^{-1} \right] = tr \begin{bmatrix} (n_0 + \lambda) & & \\ & \ddots & \\ & & n_0 + \lambda \end{bmatrix}^{-1} = \frac{I}{n_0 + \lambda}.$$

$$\boxed{\sigma_u^2 = \frac{1}{I} \sum_i \left(\hat{u}_i^2 + \frac{\sigma_\varepsilon^2}{n_0 + \lambda} \right)}$$

Estimation de σ^2_ε.

$$Q_E\left(\theta,\theta^{(r)}\right) = -2\,\mathbb{E}_\theta\left(N\log 2\pi + N\log\sigma^2_\varepsilon + \frac{1}{\sigma^2_\varepsilon}\|Y-X\beta-ZU\|^2 \,\Big|\, Y\right)$$

$\Rightarrow$ Il faut calculer $\mathbb{E}_\theta\left(\|Y-X\beta-ZU\|^2 \,|\, Y\right)$. avec la même égalité que pour $\mathbb{E}(\|U\|^2|Y)$!

$$= \|Y-X\beta-Z\underbrace{\mathbb{E}(U|Y)}_{\hat{u}}\|^2 + tr\left(Z'\mathbb{V}(U|Y)Z\right)\times\sigma^2_\varepsilon$$

$$= \underbrace{\|Y-X\beta-Z\hat{U}\|^2}_{\text{résidus du modèle}} + tr\left(Z'V(U|Y)Z\right)\times\sigma^2_\varepsilon.$$

si on prend directement $(Z'Z+\lambda I)$

↳ Dans l'étape M:

$$\sigma^2_\varepsilon = \frac{1}{N}\,\mathbb{E}_{\theta^{(r)}}\left(\|Y-X\beta-ZU\|^2 \,|\, Y\right)$$

↳ Il reste à calculer $tr\left(Z'V(U|Y)Z\right)$.

$$V(U|Y) = \sigma^2_u\left(I_q - (Z'Z+\lambda I_q)^{-1}Z'Z\right)$$

$$Tr\left(V(U|Y)\right) = \sigma^2_\varepsilon\, tr\left((Z'Z+\lambda I_q)^{-1}\right).$$

$$\sigma^2_\varepsilon\, tr\left[(Z'Z+\lambda I_q)^{-1}\right] = \sigma^2_u\times\left(q - tr\left(Z'Z\,(Z'Z+\lambda I_q)^{-1}\right)\right)$$

$$\lambda \operatorname{tr}\left((Z'Z + \lambda I_q)^{-1}\right) = q - \operatorname{tr}\left(Z'Z\,(Z'Z + \lambda I_q)^{-1}\right).$$

$$\hookrightarrow \boxed{\operatorname{tr}\left(Z\,V(U|Y)Z'\right) = q - \lambda \operatorname{tr}\left((Z'Z + \lambda I)^{-1}\right)}$$

→ Dans le cas du modele $Y_{ij} = \mu + U_i + \varepsilon_{ij}$.

$$\operatorname{tr}\left(Z'Z + \lambda I\right)^{-1} = \frac{I}{n_0 + \lambda}.$$

$$\rightarrow \boxed{\sigma_0^2 = \frac{1}{N}\left(\sum_{i=1}^{I}\sum_{j=1}^{n_0}\left(y_{ij} - \hat{\mu} - \hat{U}_i\right)^2 + \left(q - \frac{\lambda I}{n_0 + \lambda}\right) \times \sigma_\varepsilon^2\right)}$$

Résumé :

- Etape E : . Calcul de $\hat{U}$ = BLUP avec les équations d'Henderson quand $(\sigma_u^2, \sigma_\varepsilon^2, \mu)$ sont fixés.

- Etape M : . Calcul de $\sigma_u^2, \sigma_u^2, \mu$ avec les estimateurs explicits et $\hat{u}$ fixé qui a été calculé @ l'étape E précédente.

Propriétés de l'algorithme EM

1). Accroissement monotone de la vraisemblance.

↳ EM génère une suite $\theta^{(t)}$ $t \geq 0$,
et on peut mq:

$$\log \mathcal{L}(Y; \theta^{(t)}) \geq \log \mathcal{L}(Y; \theta^{(t-1)}) \quad \forall t.$$

↳ Pour le montrer il faut introduire de nouvelles quantités.

↳ l'Identité de Fisher montre : $\dfrac{\partial \log \mathcal{L}_Y}{\partial \theta} = \dfrac{\partial Q}{\partial \theta}$.

mais quel lien entre $\log \mathcal{L}_Y$ et Q ?.

→ par définition de la densité jointe

$$\underbrace{g(y;\theta)}_{\text{terme } \log \mathcal{L}(Y)} \times \underbrace{h(z \mid y;\theta)}_{\text{terme } H(\theta, \theta^{(t)})} = \underbrace{f(y, z; \theta)}_{\text{terme : } Q(\theta, \theta^{(t)})}.$$

Quand on prend l' $\mathbb{E}(\cdot \mid Y)$ $g(y;\theta)$ reste inchangée.

$$\log \mathcal{L}(Y;\theta) = \log \mathcal{L}(Y,Z;\theta) - \log \mathcal{L}(Z|Y;\theta).$$

$$\mathbb{E}_{\theta^{(t)}}\left(\text{———} \middle| Y \right) = Q(\theta,\theta^{(t)}) - H(\theta,\theta^{(t)})$$

$$H(\theta,\theta^{(t)}) = \mathbb{E}_{\theta^{(t)}}\left[\log \mathcal{L}(Z|Y;\theta) \middle| Y \right].$$

$$= \int \log h(z|y;\theta) \times h(z|y;\theta^{(t)})\, dz.$$

Accroissements de Q.

Par définition de la phase M.

$$Q(\theta^{(t+1)};\theta^{(t)}) \geqslant Q(\theta^{(t)},\theta^{(t)}).$$

Accroissements de H (Par Jensen $\mathbb{E} f(x) \leq f(x)$ f concave).

$$H(\theta^{(t+1)},\theta^{(t)}) - H(\overset{(t+1)}{\theta},\theta^{(t)}).$$

$$= \int \log\left[\frac{h(z|y;\theta^{(t+1)})}{h(z|y;\theta^{(t)})}\right] * h(z|y;\theta^{(t)})\, dz.$$

$$\leqslant \log \int \frac{h(z|y,\theta^{(t+1)})}{h(z|y,\theta^{(t)})} \times h(z|y,\theta^{(t)})\, dz = 0.$$

Autour de l'Algorithme EM.

- Initialisation ? : stratégie . stochastique
  . déterministe

- Accélération ? : . Stratégie de classement "dur"
  - algorithmes en ligne (incrementaux).

- Preuve de cv vers un max local ? global ?

- Extensions en Bayesien, / aussi SAEM

Biblio : The EM Algorithm & Extensions Wiley.
Mc Lachlan & Krishnan.