M1 : estimation et sélection de modèle

Février 2021

Modèle

- Observations

$$Y_{ij}|\psi_i = (\lambda_i, h_i) \sim \mathcal{N}(m_{ij}, s_{ij}^2 + \sigma^2)$$

où

$$m_{ij} = \frac{\sqrt{d_{ij}} - 1}{C_{2E}\lambda_i + h_i(\sqrt{d_{ij}} - 1)}$$

et

$$s_{ij}^2 = \frac{C_{2V} - C_{2E}^2}{\Delta} \frac{\lambda_i^2(\sqrt{d_{ij}} - 1)}{\left(C_{2E}\lambda_i + h_i(\sqrt{d_{ij}} - 1)\right)^3}$$

- Paramètres individuels

$$\log(\lambda_i) \sim \mathcal{N}(m_\lambda, s_\lambda^2)\ ,\ \log(h_i) \sim \mathcal{N}(m_h, s_h^2)$$

Différentes versions du modèle

- Variabilité inter-individuelle sur λ et h :

$$\log(\lambda_i) \sim \mathcal{N}(m_\lambda, s_\lambda^2)\ ,\ \log(h_i) \sim \mathcal{N}(m_h, s_h^2)$$

  - $\mathcal{M}_1^{(m)}$: $V(Y_{ij}) = \sigma^2$
  - $\mathcal{M}_2^{(m)}$: $V(Y_{ij}) = s_{ij}^2 + \sigma^2$
  - $\mathcal{M}_3^{(m)}$: $V(Y_{ij}) = s_{ij}^2$

- Absence de variabilité inter-individuelle :

$$\lambda_i \equiv \lambda\ ,\ h_i \equiv h$$

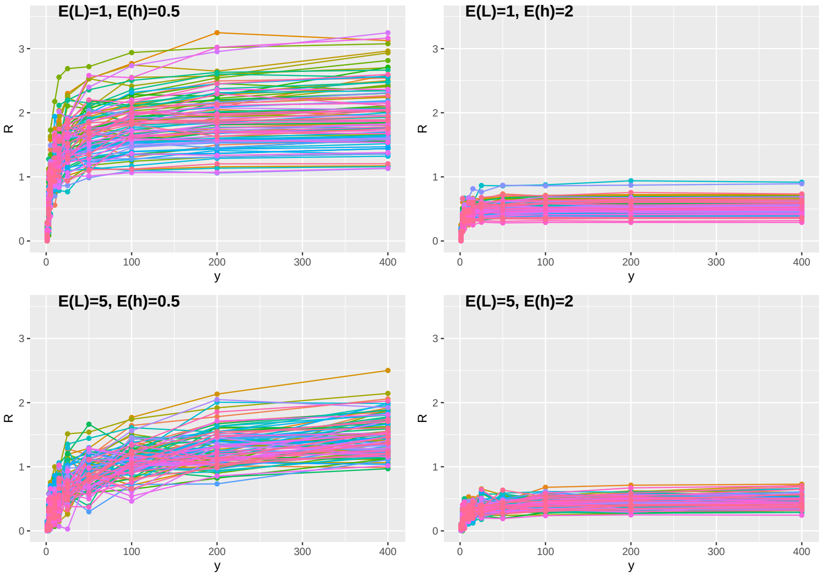
  - $\mathcal{M}_1^{(f)}$: $V(Y_{ij}) = \sigma^2$
  - $\mathcal{M}_2^{(f)}$: $V(Y_{ij}) = s_j^2 + \sigma^2$
  - $\mathcal{M}_3^{(f)}$: $V(Y_{ij}) = s_j^2$

Scénarios de simulations

- Choix d'une forme pour la variance des réponses fonctionnelles
- $E(\lambda_i) \in \{1, 5\}$, $E(h) \in \{0.5, 2\}$
- $d_{ij} \in \{1.1, 3, 5, 10, 15, 25, 50, 100, 200, 400\}$
- $N \in \{100, 200, 500\}$
- niveaux de variabilité résiduelle : $\sigma * E(h) \in \{5\%, 10\%, 25\%, 50\%\}$
- coefficients de variation de λ_i et $h_i \in \{0.25, 0.5, 0.75\}$

1. Estimation sous le vrai modèle
2. Estimation sous un modèle mal spécifié
3. Sélection de modèle

Allure des courbes, $\mathcal{M}_3^m$



Simulations sous $\mathcal{M}_2^m$, estimation sous le vrai modèle

- $E(\lambda_i) = 1,\ E(h_i) = 0.5,\ \Delta = 1,\ cv = 0.25$
- influence de σ

Table 1: residual noise to signal ratio : 0.05

	ml.m	mh.m	sl.m	sh.m	sig.m	ml.sd	mh.sd	sl.sd	sh.sd	sig.sd
N\true	-0.030	-0.723	0.246	0.246	0.100	NA	NA	NA	NA	NA
100	-0.029	-0.728	0.237	0.248	0.101	0.038	0.026	0.032	0.020	0.005
200	-0.034	-0.724	0.242	0.247	0.100	0.024	0.018	0.025	0.015	0.004
500	-0.032	-0.723	0.243	0.246	0.100	0.017	0.011	0.016	0.008	0.002

Table 2: residual noise to signal ratio : 0.50

	ml.m	mh.m	sl.m	sh.m	sig.m	ml.sd	mh.sd	sl.sd	sh.sd	sig.sd
N\true	-0.030	-0.723	0.246	0.246	1.000	NA	NA	NA	NA	NA
100	-0.037	-0.734	0.431	0.227	0.991	0.142	0.049	0.110	0.042	0.025
200	-0.035	-0.730	0.396	0.222	0.996	0.108	0.040	0.082	0.033	0.018
500	-0.028	-0.725	0.345	0.236	1.001	0.065	0.026	0.067	0.021	0.011

Simulations sous $\mathcal{M}_2^m$, estimation sous le vrai modèle

- $E(\lambda_i) = 1$, $E(h_i) = 0.5$, $\Delta = 1$, $\%\sigma = 0.10$
- influence de la variabilité inter-individuelle (Biais relatifs)

Table 3: variation ratio : 0.25

	ml.m	mh.m	sl.m	sh.m	sig.m	ml.sd	mh.sd	sl.sd	sh.sd	sig.sd
N=100	-0.002	-0.009	0.023	-0.013	-0.010	0.049	0.026	0.048	0.020	0.003
N=200	0.000	-0.001	0.007	-0.009	0.000	0.041	0.020	0.044	0.014	0.002
N=500	-0.001	0.001	-0.014	-0.004	0.001	0.022	0.013	0.028	0.009	0.001

Table 4: variation ratio : 0.50

	ml.m	mh.m	sl.m	sh.m	sig.m	ml.sd	mh.sd	sl.sd	sh.sd	sig.sd
N=100	-0.005	0.009	-0.012	-0.003	0.012	0.067	0.047	0.063	0.034	0.003
N=200	0.000	-0.003	-0.023	0.003	-0.001	0.049	0.037	0.045	0.026	0.002
N=500	0.002	0.002	-0.011	-0.008	-0.001	0.028	0.019	0.025	0.017	0.001

Table 5: variation ratio : 0.75

	ml.m	mh.m	sl.m	sh.m	sig.m	ml.sd	mh.sd	sl.sd	sh.sd	sig.sd
N=100	0.002	0.000	-0.024	-0.013	0.000	0.074	0.065	0.066	0.048	0.003
N=200	0.005	0.002	-0.012	-0.014	0.003	0.062	0.052	0.053	0.036	0.002
N=500	0.002	0.006	-0.006	-0.006	0.003	0.033	0.032	0.031	0.022	0.001

Estimation sous le vrai modèle : bilan

- plus N est grand, plus le biais est faible et la précision grande
- l'augmentation de σ entraîne
  1. une baisse de la précision pour l'estimation de m_λ, m_h, s_λ, s_{ch}
  2. une augmentation du biais pour certains paramètres pour les plus faibles valeurs de N
- l'augmentation de la variabilité inter-individuelle entraîne
  1. une augmentation du biais sur σ
  2. une baisse de la précision
- mêmes conclusions que les données soient simulées sous $\mathcal{M}_1^m$, $\mathcal{M}_2^m$, $\mathcal{M}_3^m$ et quelle que soit la valeur de Δ

Simulations sous $\mathcal{M}_2^m$, estimation sous le mauvais modèle

- influence de σ
- estimations moyennes à $N = 500$

Table 6: residual noise to signal ratio : 0.05

	mul	muh	sl	sh	sigma
true	-0.030	-0.723	0.246	0.246	0.100
M1m	-0.034	-0.724	0.273	0.244	0.169
M3m	-0.015	-0.726	0.252	0.247	NA
M1f	-0.008	-0.745	NA	NA	0.340
M2f	-0.015	-0.742	NA	NA	0.329
M3f	1.062	-1.186	NA	NA	NA

Table 7: residual noise to signal ratio : 0.50

	mul	muh	sl	sh	sigma
true	-0.030	-0.723	0.246	0.246	1.000
M1m	-0.034	-0.721	0.263	0.239	1.011
M3m	-0.005	-0.733	0.379	0.269	NA
M1f	-0.013	-0.741	NA	NA	1.053
M2f	-0.014	-0.740	NA	NA	1.045
M3f	0.631	-17.954	NA	NA	NA

Influence de la v. inter-individuelle (biais relatifs)

Table 8: variation ratio : 0.25

	mul	muh	sl	sh	sigma
M1m	-0.002	-0.006	0.344	-0.066	0.752
M3m	0.029	-0.018	0.052	0.004	NA
M1f	0.012	-0.030	NA	NA	1.225
M2f	0.289	-0.286	NA	NA	-0.214
M3f	0.488	-0.474	NA	NA	NA

Table 9: variation ratio : 0.50

	mul	muh	sl	sh	sigma
M1m	-0.004	0.001	0.055	-0.018	0.884
M3m	0.035	-0.022	0.017	-0.008	NA
M1f	3.874	-2.618	NA	NA	2.998
M2f	0.093	-1.773	NA	NA	2.156
M3f	0.727	-1.526	NA	NA	NA

Table 10: variation ratio : 0.75

	mul	muh	sl	sh	sigma
M1m	-0.010	0.014	-0.012	-0.008	1.085
M3m	0.038	-0.023	0.008	-0.011	NA
M1f	0.529	-1.124	NA	NA	4.503
M2f	1.082	-2.798	NA	NA	4.493
M3f	0.866	-4.954	NA	NA	NA

Estimation sous le mauvais modèle : influence de σ

1. Estimation sous un modèle mixte dont la variance est mal spécifiée
   - ajouter/enlever un terme mécaniste au terme résiduel ($\mathcal{M}_1^{(m)}/\mathcal{M}_2^{(m)}$) dans la variance entraîne surtout une sous/sur-estimation de σ
   - négliger la résiduelle ($\mathcal{M}_3^{(m)}$) entraîne du biais dans l'estimation des paramètres, surtout les variances des effets aléatoires. Ce biais est d'autant plus élevé que la vraie valeur de σ est grande.
2. Estimation sous un modèle à effets fixes
   - biais sur l'ensemble des paramètres
   - tendance à la sur-estimation de σ (qui absorbe la variabilité inter-individuelle non prise en compte par le modèle)
   - sur-estimation de σ d'autant plus grande que le coefficient de variation des effets aléatoires est grand

Sélection de modèle

Pénalité adaptée aux modèles à effets mixtes

$$BIC(\mathcal{M}) = -2 * \mathcal{L} + \sharp_1 \log(N) + \sharp_2 \log(n_t)$$

- $\sharp_1$ = nombre de paramètres des distributions des effets aléatoires
- $\sharp_2$ = nombre d'effets fixes
- N : nombre de perdrix
- n_t : nombre total d'observations

Vrai modèle $\mathcal{M}_1^m$: effet des deux sources de variabilité, N=500

Table 11: variation ratio of the random effects = 0.25

	M1m	M2m	M3m	M1f	M2f	M3f
s. 0.05, delta=1	0	1	18	81	0	0
s. 0.05, delta=10	0	40	1	59	0	0
s. 0.10, delta=1	0	0	2	98	0	0
s. 0.10, delta=10	0	0	0	100	0	0

Table 12: variation ratio of the random effects = 0.50

	M1m	M2m	M3m	M1f	M2f	M3f
s. 0.05, delta=1	66	15	19	0	0	0
s. 0.05, delta=10	49	50	1	0	0	0
s. 0.10, delta=1	13	38	9	40	0	0
s. 0.10, delta=10	17	52	0	31	0	0

Vrai modèle $\mathcal{M}_2^m$: effet des deux sources de variabilité, N=500

Table 13: variation ratio of the random effects = 0.25

	M1m	M2m	M3m	M1f	M2f	M3f
s. 0.05, delta=1	0	0	7	93	0	0
s. 0.05, delta=10	0	29	1	70	0	0
s. 0.10, delta=1	0	0	1	99	0	0
s. 0.10, delta=10	0	0	0	100	0	0

Table 14: variation ratio of the random effects = 0.50

	M1m	M2m	M3m	M1f	M2f	M3f
s. 0.05, delta=1	31	47	21	1	0	0
s. 0.05, delta=10	40	57	3	0	0	0
s. 0.10, delta=1	0	30	7	63	0	0
s. 0.10, delta=10	11	41	0	48	0	0

Vrai modèle $\mathcal{M}_3^m$

Table 15: Delta=1

	M1m	M2m	M3m	M1f	M2f	M3f
cv. 0.25, delta=1	0	7	34	59	0	0
cv. 0.50, delta=1	25	31	44	0	0	0
cv. 0.75, delta=1	22	22	56	0	0	0

Table 16: Delta=10

	M1m	M2m	M3m	M1f	M2f	M3f
cv. 0.25, delta=10	21	18	61	0	0	0
cv. 0.50, delta=10	23	21	56	0	0	0
cv. 0.75, delta=10	22	27	51	0	0	0

Sélection de modèle

- La variabilité inter-individuelle n'est décelable que si la variabilité résiduelle n'est pas trop grande et le coefficient de variation des effets aléatoires pas trop petit.
- Le cas échéant, forte hésitation entre $\mathcal{M}_1^m$ et $\mathcal{M}_2^m$ lorsque les données sont générées sous $\mathcal{M}_1^m$ (même pénalité pour le BIC, choix basé sur la valeur de vraisemblance seule)
- Influence de Δ
- Quelques ordres de grandeur : $E(h_i) = 0.5$, $E(\lambda_i) = 1$, $\Delta = 10$
  - $(s_{ij})_{j=1,\dots,10} = \{0.03, 0.07, 0.06, 0.06, 0.05, 0.05, 0.04, 0.03, 0.02, 0.02\}$
  - alors que $\sigma \in \{0.1, 0.2, 0.5, 1\}$