GSA avec Morris et Anova

Introduction élémentaire

École Chercheur Mexico

C. Bruchou
bruchou@avignon.inra.fr

Contexte et objectifs

- $\mathcal{G}$ modèle de simulation :

  $\Omega = [0,1]^K \to \mathbb{R}, \;\; \mathbf{x} = (x^{(1)}, .., x^{(K)}) \longmapsto \mathcal{G}(\mathbf{x}),$
- $\mathbf{x}$: vecteur des K facteurs continus du modèle,
- **Rem** : unités d'usage des facteurs par transfo. lin. de $\mathbf{x}$
- $\Omega_i = \{0, \frac{1}{Q-1}, \frac{2}{Q-1}, .., \frac{Q-1}{Q-1} = 1\}$: découpage de $[0,1]$
- grille de Ω : $\tilde{\Omega} = \otimes_{i=1}^{K} \tilde{\Omega}_i$,
- objectifs :
  - détecter et hiérarchiser les effets principaux et d'interaction des K facteurs.
  - échantillonner parcimonieusement (*si possible*) dans $\tilde{\Omega}$,
  - préserver la distribution d'échantillonnage *a priori* uniforme des facteurs,
  - analyser les variations de $\mathcal{G}$,

Plan d'échantillonnage

- principe OAT,
- trajectoire aléatoire : $\mathcal{T}_l = (P_{l,1}, ..., P_{l,K+1}) \in \tilde{\Omega}$, $l = 1, .., N$
- variation d'un facteur = pas entre deux points :
  $P_{l,j+1} = P_{l,j} + \epsilon\delta\,\vec{i}_{\sigma(j)}$,
  - $\sigma()$ permutation aléatoire sans répétition de $\{1, ..., K\}$,
  - $\vec{i}_{\alpha=\sigma(j)}$ vecteur unitaire de la base de Ω,
  - $\delta \propto 1/(Q-1)$: longueur du pas,
  - $\epsilon \pm 1$: signe aléatoire de la direction du pas.
- N trajectoires $\Rightarrow$ $N(K+1)$ simulations, N variations de $\mathcal{G}$,
- **en pratique** : Q **pair**, et $\delta = \frac{Q}{2(Q-1)}$.

Indices de Morris

- facteur $X^{(\alpha)}$:
- effet **élémentaire** : $\Delta_l^{(\alpha)}\mathcal{G} = \frac{\mathcal{G}(P_{l,j+1})-\mathcal{G}(P_{l,j})}{\epsilon\delta}$
- effet **moyen** : $\mu_\alpha = \frac{1}{N}\sum_{l=1}^{N}\Delta_l^{(\alpha)}\mathcal{G}$
- effet absolu **moyen** : $\mu^\star{}_\alpha = \frac{1}{N}\sum_{l=1}^{N}|\Delta_l^{(\alpha)}\mathcal{G}|$
- s.d. des effets élémentaires : $\sigma_\alpha = \sqrt{\frac{1}{N-1}\sum_{l=1}^{N}(\Delta_l^{(\alpha)}\mathcal{G}-\mu_\alpha)^2}$

Interprétation de μ_α

- Si $\mathcal{G}$ dérivable :
  - $\Delta_l^{(\alpha)}\mathcal{G} \simeq \dfrac{\partial \mathcal{G}}{\partial x^{(\alpha)}}$ (si δ petit)
  - $\Omega = [0,1]^{K-1} \times [0,1], dx = dx_{\sim\alpha}\, dx^{(\alpha)}$,
  - N grand, $\mu_\alpha \simeq \dfrac{1}{|\Omega|}\displaystyle\int_\Omega \dfrac{\partial \mathcal{G}}{\partial x^{(\alpha)}} dx = \int_{[0,1]^{K-1}} \int_0^1 \dfrac{\partial \mathcal{G}}{\partial x^{(\alpha)}} dx^{(\alpha)} dx_{\sim\alpha}$

    $= \displaystyle\int_{[0,1]^{K-1}} \mathcal{G}(.., x^{(\alpha)} = 1, ..) dx_{\sim\alpha} - \int_{[0,1]^{K-1}} \mathcal{G}(.., x^{(\alpha)} = 0, ..) dx_{\sim\alpha}$
  - μ_α est l'écart écart entre les moyennes de $\mathcal{G}$ aux bornes du domaine de $X^{(\alpha)}$
  - Ex 1 :$\mathcal{G}(x,y) = a_1\, x + a_2\, y \Rightarrow \mu_x = \displaystyle\int_{[0,1]^2} \dfrac{\partial \mathcal{G}}{\partial x} dxdy = a_1$
  - Ex. 2 : $\mathcal{G}(x,y) = (x-0.5)(y-0.5) \Rightarrow \mu_x = 0$.
    Pourtant, x joue un rôle !
  - Conclusion : μ_α peut prendre mal en compte l'effet d'un facteur.

Interprétation de $\mu^\star{}_\alpha$

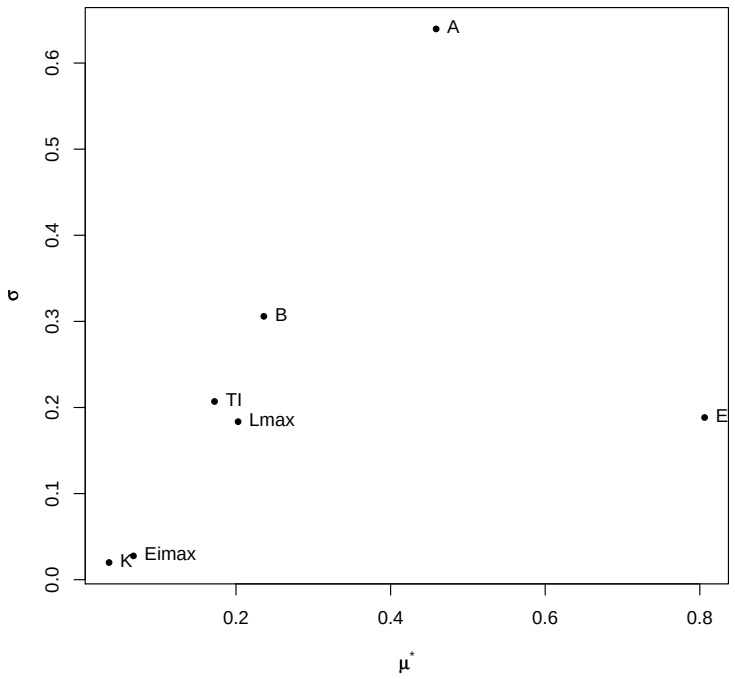
- $\Omega = \Omega^+ \cup \Omega^-$, $\Omega^+ = \{x, \frac{\partial \mathcal{G}}{\partial x^{(\alpha)}} > 0\}$, $\Omega^- = \{x, \frac{\partial \mathcal{G}}{\partial x^{(\alpha)}} \leq 0\}$
- $\mu^\star{}_\alpha \simeq \int_{[0,1]^K} |\frac{\partial \mathcal{G}}{\partial x^{(\alpha)}}| dx = \int_{\Omega^+} \frac{\partial \mathcal{G}}{\partial x^{(\alpha)}} dx - \int_{\Omega^-} \frac{\partial \mathcal{G}}{\partial x^{(\alpha)}} dx > 0$
- Ex. 2 : $\mathcal{G}(x, y) = (x - 0.5)(y - 0.5)$,
  pour le facteur facteur X on a
  $\Omega^+ = \{(x, y) \in [0, 1]^2, y > 0.5\}$, $\Omega^- = \{(x, y), y < 0.5\}$
  $\Rightarrow \mu^\star{}_x = \int_0^1 \int_{0.5}^1 (y - 0.5) dxdy + \int_0^1 \int_0^{0.5} (0.5 - y) dxdy = 0.25$
- en général, $\mu^\star$ caractérise l'effet **total** d'un facteur qui cumule l'effet individuel ou **principal** du facteur et l'effet combiné avec d'autres facteurs (**interaction**).
- Rem. : faible étendue et dérivée élevée de $\mathcal{G}$ sont compatibles.

Interprétation de σ_α

- $\sigma_\alpha^2 = \int_{[0,1]^K} \left[\frac{\partial \mathcal{G}}{\partial x^{(\alpha)}}\right]^2 dx - \mu_\alpha^2$
- Ex. 3 : $\mathcal{G}(x,y) = 4 - (x - 0.5) + 2(y - .5)^2 + 3(x - 0.5)(y - 0.5)$
  - $\mu_x = \int_0^1 \int_0^1 (3y - \frac{5}{2}) dxdy = -1$, $\mu_y = \int_0^1 \int_0^1 (4y + 3x - \frac{7}{2}) dxdy = 0$
  - $\mu_x^* = \int_{\frac{5}{6}}^1 \int_0^1 (3y - \frac{5}{2}) dxdy + \int_0^{\frac{5}{6}} \int_0^1 (\frac{5}{2} - 3y) dxdy = \frac{13}{12}$
  - $\mu^\star{}_y = \int_0^1 \int_{\frac{7}{8} - \frac{3}{4}x}^1 (4y + 3x - \frac{7}{2}) dydx + \int_0^1 \int_0^{\frac{7}{8} - \frac{3}{4}x} (\frac{7}{2} - 4y - 3x) dydx = 4.69$
  - $\sigma_x^2 = \int_0^1 \int_0^1 \frac{\partial \mathcal{G}}{\partial x}^2 dxdy - \mu_x^2 = 0.75$, $\sigma_y^2 = \int_0^1 \int_0^1 \frac{\partial \mathcal{G}}{\partial y}^2 dxdy - \mu_y^2 = 2.08$
  - interprétation des indices :
    - présence d'interaction ou de non-linéarités ($\sigma_x = 0.87$ et $\sigma_y = 1.44$), (ici, toutes les deux).
    - les dérivées patielles de G par rapport de Y se compensent ($\mu_y = 0$) mais ne sont pas nulles $\mu^\star{}_y = 4.69$)
    - **en pratique** : étude graphique des effets fournis par la fct morris() pour mettre en évidence les non-linéarités.
- Exercice : étudier l'exemple 3 avec la fct morris() de "sensitivity".

Interprétation

graphique de Morris $(\mu^{\star}{}_{\alpha} \times \sigma_{\alpha})$, $\alpha = 1, K$



- hiérarchie entre les facteurs (K, $Eimax < Lmax$, TI, B, $A < E$),
- facteurs autour de l'origine : effet négligeable (K, $Eimax$),
- à droite de l'axe des abscisses : facteurs ayant un effet total important (E),
- en haut à droite du graphique : non-linéaire ou interaction (A),
- graphique en 3d disponible : $\mu \times \mu^{\star} \times \sigma^{\star}$.

Inférence

- indépendance des $|\Delta^{(i)}\mathcal{G}|, I = 1, N$
- IC bilatéraux de probabilité 0.95 des indices $\mu^\star$ et σ_α :
  - $\mu^\star_\alpha \pm T_{N-1}(0.975)\frac{\sigma_{\mu^\star_\alpha}}{\sqrt{N}}$ avec $T_{N-1}(p)$ quantile d'ordre p de la loi de Student à $N-1$ ddl,
  - $[\sqrt{\frac{N-1}{\chi^2(N-1,0.975)}}\,\sigma_\alpha, \sqrt{\frac{N-1}{\chi^2(N-1,0.0.025)}}\,\sigma_\alpha]$ avec $\chi^2(N-1,p)$ quantile d'ordre p de la loi du χ^2 à $N-1$ ddl
- IC par simulation bootstrap : tirage avec remise des trajectoires,
- le nombre de trajectoires dépend de la précision souhaitée et du modèle de simulation (pré-étude nécessaire pour avoir un ordre de grandeur de $\sigma_{\mu^\star_\alpha}$ σ_α).

R

```
> Q=6
> delta = Q/(2*(Q-1)))
> morris.out = morris(model = wwdm.simule,
        factors= noms , r = 30,
        design = list(type = "oat", levels = Q,
        binf = 0, bsup= 1)
> plot(morris.out)
```

R

```
> IC.mu = apply(abs(x$ee),2, t.test)
> for(i in 1:7) print(IC.mu[[i]]$conf.int)
> sigma = sqrt(apply(x$ee,2,var))
> N = etude.morris$r
> for(i in 1:7) print(c(sigma[i]*(N-1)^.5/qchisq(.975,N-1)^.5,
                        sigma[i]*(N-1)^.5/qchisq(.025,N-1)^.5))
```

$\mu^\star$	Eb	Eimax	K	Lmax	A	B	TI
binf	0.73	0.065	0.033	0.19	0.38	0.15	0.15
bsup	0.80	0.08	0.042	0.28	0.64	0.27	0.2
σ							
binf	0.16	0.02	0.021	0.21	0.61	0.25	0.18
bsup	0.22	0.03	0.028	0.27	0.80	0.33	0.24

Discussion

- Quid de la parcimonie de la méthode ?
- effet de la direction du pas sur l'estimation de la variabilité :

  $\mathcal{G}(x, y) = x^2 y$

  $\mathcal{G}(\delta, 0) - \mathcal{G}(0, 0) = \mathcal{G}(0, \delta) - \mathcal{G}(0, 0) = 0 \neq \mathcal{G}(\frac{\delta}{\sqrt{2}}, \frac{\delta}{\sqrt{2}}) - \mathcal{G}(0, 0) = \frac{\delta^3}{2}$
- noeud au voisinage d'une discontinuité :

  $\mathcal{G}(x, y) = \frac{xy}{x^2+y^2}$ non-continue en $(0, 0)$,

  soit ϵ petit, conséquence de la position du noeud de la grille en $(0, \epsilon)$ ou $(\epsilon, 0)$

  $\frac{\partial \mathcal{G}}{\partial x}(\epsilon, 0) = 0$, $\frac{\partial \mathcal{G}}{\partial x}(0, \epsilon) = \frac{1}{\epsilon}$
- examen des quantiles $q_{.95}$ des effets ?
- Exercice : adapter la méthode pour traiter les effets quadratiques :
  - trajectoire (dérivée seconde partielle) :

    $(x, y), (x + \delta, y), (x + \frac{\delta}{2}, y), (x + \delta, y + \frac{\delta}{2}), (x + \delta, y + \delta)$,
  - estimation empirique de $\frac{\partial^2}{\partial x^2}$ et $\frac{\partial^2}{\partial y^2}$,
  - Comment définir une trajectoire pour estimer $\frac{\partial^2}{\partial x \partial y}$?

Rappels élémentaires

Dispositif complet et équilibré à deux facteurs A et B :

- variable analysée $G = f(A, B)$, I niveaux pour A (resp. J pour B)
- R répétitions indépendantes par combinaison des facteurs
- Tables des résultats d'une expérience agronomique :

exp.	A	B	r	G
1	1	1	1	9.5
2	1	1	2	10.5
3	1	2	1	20.2
4	1	2	2	18.8
5	1	3	1	30.3
6	1	3	2	27.7
7	2	1	1	8.4
8	2	1	2	9.6
9	2	2	1	29
10	2	2	2	31
11	2	3	1	41
12	2	3	2	39
13	3	1	1	1.5
14	3	1	2	2.5
15	3	2	1	31.5
16	3	2	2	28.5
17	3	3	1	41
18	3	3	2	39

moyenne $\bar{G}_{\dots} = 23.28$, variance $\sigma^2_G = 180.9$

Modèle statistique linéaire

- $G_{i,j,r} = \mu + A_i + B_j + AB_{i,j} + \epsilon_{i,j,r}$
  $i = 1, I,\ \ j = 1, J,\ r = 1, R$
- $\mu =$ constante,
- A_i : i^{eme} effet factoriel ; $\sum_i A_i = 0$,
- B_j : j^{eme} effet factoriel ; $\sum_j B_j = 0$,
- AB_{ij} : effet d'interaction ; $\forall j$, $\sum_i AB_{ij} = 0$ et $\forall i$, $\sum_j AB_{ij} = 0$,
- ϵ erreur *aléatoire* d'espérance nulle et de variance σ_G^2.

Estimateurs des effets

- effets principaux :
  - estimateur de A_i : $\hat{A}_i = \bar{G}_{i..} - \bar{G}$
  - estimateur de B_j : $\hat{B}_j = \bar{G}_{.j.} - \bar{G}$

	A_i	B_j
1	-3.78	-16.28
2	3.06	3.22
3	0.72	13.06

- effets d'interaction :
  - $AB_{ij} = (E(G_{ij}) - \mu) - A_i - B_j$
  - $\Rightarrow \widehat{AB_{ij}} = \bar{G}_{ij.} - \bar{G}_{i..} - \bar{G}_{.j.} + \bar{G}$

$i \setminus j$	1	2	3
1	6.78	-3.22	-3.56
2	-1.06	0.45	0.61
3	-5.72	2.78	2.94

Variabilité des effets

$$SS_T = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{r=1}^{R}(G_{ijr} - \bar{G}_{...})^2 = SS_A + SS_B + SS_{AB} + SS_\epsilon$$

$$SS_A = JR\sum_{i=1}^{I} A_i^2 = 6\sum_{i=1}^{3}(\bar{G}_{i..} - \bar{G})^2 = 144.8,$$

$$SS_B = IR\sum_{j=1}^{J} B_j^2 = 6\sum_{j=1}^{3}(\bar{G}_{.j.} - \bar{G})^2 = 2674.8$$

$$SS_{AB} = R\sum_{i=1}^{I}\sum_{j=1}^{J} AB_{ij}^2 = 2\sum_{i=1}^{3}\sum_{j=1}^{3}(\bar{G}_{ij.} - \bar{G})^2 = 239.6$$

$$SS_\epsilon = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{r=1}^{2}(G_{ijr} - \bar{G}_{ij.})^2 = 253.11 = 16.6$$

Exercice : vérifier sous R les calculs numériques de l'exemple.

Variance des effets

$N = IJR,$

- $Var(G) = MS_T = \dfrac{SS_T}{N-1}$
- estimateurs des variances des effets factoriels :
  $s_A^2 = \dfrac{\sum_{i=1}^{I} A_i^2}{I-1}$, $s_B^2 = \dfrac{\sum_{j=1}^{J} B_i^2}{J-1}$, $s_{AB}^2 = \dfrac{\sum_{i=1}^{I}\sum_{j=1}^{J} AB_{ij}^2}{(I-1)(J-1)}$
  variance résiduelle :
  $s_\epsilon^2 = \dfrac{SS_\epsilon}{IJ(R-1)}$,
- Moyennes des carrés des effets factoriels (dans la table d'anova) :
  $MS_A = \dfrac{SS_A}{I-1} = JR\, s_A^2,\ MS_B = \dfrac{SS_B}{J-1} = IR\, s_B^2,\ MS_{AB} = \dfrac{SS_{AB}}{(I-1)(J-1)} = R\, s_{AB}^2,\ MS_\epsilon = s_\epsilon^2,$
- $Var(G) = \dfrac{N}{N-1}\left\{\dfrac{(I-1)}{I}s_A^2 + \dfrac{(J-1)}{J}s_B^2 + \dfrac{(I-1)(J-1)}{IJ}s_{AB}^2 + \dfrac{R-1}{R}s_\epsilon^2\right\}$
- N grand $\Rightarrow \frac{N}{N-1} \simeq 1$,
- en supposant s_A^2 et s_B^2 peu sensibles au niveau de discrétisation des facteurs,
  $I < J \Rightarrow$ part du facteur B renforcée par le poids $\frac{(J-1)}{J}$

Indices de sensibilité d'un modèle de simulation

$\mathcal{G}(x, y)$ modèle **déterministe**, X et Y facteurs continus,
discrétisation de X et $Y \Rightarrow A, B$ avec $I = J$ niveaux,
expérience = combinaison des facteurs A et B,
IJ combinaisons possibles ($R = 1$)

- nombre de modalités proche pour les facteurs,
- Indice = contribution à SS_T (d'après Saltelli)
  - indices principaux

    facteur X : $I_x = \dfrac{SS_A}{SS_T}$

    facteur Y : $I_y = \dfrac{SS_B}{SS_T}$
  - indice totaux

    facteur X : $IT_x = \dfrac{SS_A + SS_{AB}}{SS_T}$

    facteur Y : $IT_y = \dfrac{SS_B + SS_{AB}}{SS_T}$
- N.B. : $SS_\epsilon = 0$

Analyse conditionnelle

- G modèle **stochastique** défini sur une grille $\Omega = \{1, .., I\} \otimes \{1, .., J\}$
- $(i,j) \in \Omega,\ \ G_{i,j} = G(i,j) = \mu + \alpha X(i) + \beta Y(j) + \gamma X(i) Y(j) + \epsilon(i,j)$,

  avec X, Y **v.a.i.** vérifiant $\sum_i X_i = \sum_j Y_j = 0$, α, β, γ et μ constantes, ϵ bruit de moyenne nulle et variance σ_ϵ^2
- lois uniformes : $Prob(X = X_i) = \dfrac{1}{I}$, $Prob(Y = Y_j) = \frac{1}{J} \Rightarrow \mathbb{E}(X) = \mathbb{E}(Y) = 0$
- $\mathbb{E}(G) = \mu$ et $\mathbb{V}(G) = \alpha^2\sigma_X^2 + \beta^2\sigma_Y^2 + \gamma^2\sigma_X^2\sigma_Y^2 + \sigma_\epsilon^2$
- **espérance conditionnelle** de G sachant X : à X_i est associé

  $$\mathbb{E}(G|X = X_i) = \mu + \alpha\, X_i + \beta\, \mathbb{E}(Y|X = X_i) + \gamma\, \mathbb{E}(XY|X = X_i) + \mathbb{E}(\epsilon|X = X_i) = \mu + \alpha X_i$$
- $\mathbb{E}_X(\mathbb{E}(G|X)) = \sum_i \frac{1}{I}\mathbb{E}(G|X = X_i) = \mu + \alpha\frac{1}{I}\sum_i X_i = \mu = \mathbb{E}(G)$
- $\mathbb{V}_X(\mathbb{E}(G|X)) = \sum_i \frac{1}{I}(\mu + \alpha X_i - \mu)^2 = \alpha^2\sigma_X^2$
- **variance conditionnelle** $\mathbb{V}(G|X)$ de G sachant X : à X_i est associé

  $$\mathbb{V}(G|X = X_i) = \mathbb{E}(G^2|X = X_i) - (\mathbb{E}(G|X = X_i))^2 = \beta^2\sigma_Y^2 + \gamma^2 X_i^2\sigma_Y^2 + \sigma_\epsilon^2$$
- $\mathbb{E}_X(\mathbb{V}(G|X)) = \beta^2\sigma_Y^2 + \gamma^2\sigma_X^2\sigma_Y^2 + \sigma_\epsilon^2 \quad \left(\mathbb{E}_Y(\mathbb{V}(G|Y)) = \alpha^2\sigma_X^2 + \gamma^2\sigma_X^2\sigma_Y^2 + \sigma_\epsilon^2\right)$
- $\mathbb{V}(G) = \mathbb{V}_X(\mathbb{E}(G|X)) + \mathbb{E}_X(\mathbb{V}(G|X))$

Indices de sensibilité

- $I_X = \frac{\mathbb{V}_X(\mathbb{E}(G|X))}{\mathbb{V}(G)} = \frac{\alpha^2\sigma_X^2}{\mathbb{V}(G)}$
- $IT_X = \frac{\alpha^2\sigma_X^2 + \gamma^2\sigma_X^2\sigma_Y^2}{\mathbb{V}(G)} = \frac{\mathbb{E}_Y(\mathbb{V}(G|Y)) - \sigma_\epsilon^2}{\mathbb{V}(G)} = 1 - \frac{\mathbb{V}_Y(\mathbb{E}(G|Y))}{\mathbb{V}(G)} - \frac{\sigma_\epsilon^2}{\mathbb{V}(G)}$
- $I_Y = \frac{\mathbb{V}_Y(\mathbb{E}(G|Y))}{\mathbb{V}(G)} = \frac{\alpha^2\sigma_Y^2}{\mathbb{V}(G)}$
- $IT_Y = \frac{\beta^2\sigma_Y^2 + \gamma^2\sigma_X^2\sigma_Y^2}{\mathbb{V}(G)} = \frac{\mathbb{E}_X(\mathbb{V}(G|X)) - \sigma_\epsilon^2}{\mathbb{V}(G)} = 1 - \frac{\mathbb{V}_X(\mathbb{E}(G|X))}{\mathbb{V}(G)} - \frac{\sigma_\epsilon^2}{\mathbb{V}(G)}$

Discussion

- mise en oeuvre aisée et interprétation relativement claire,
- utile pour expliquer la variabilité par une approximation linéaire
- limites :
  - si non-linéarités (augmenter le nbre de niveaux !),
  - si nombreux facteurs :
    30 facteurs, 3 niveaux, $R = 1$, $\Rightarrow 3^{30} \sim 2 \times 10^{14}$ simulations.
    $t_{exec} = 10^{-4} s$, 1000 processeurs $\Rightarrow T_{tot} = 7.8$ mois.
- Objectif : construire un plan parcimonieux

Codage d'un Plan complet

3 facteurs à 2 niveaux $(-1, +1)$, $N = 2^3$ simulations,
Matrice reliant les mesures G aux paramètres (i.e. effets) du modèle d'anova :

MU	A	B	C	AB	AC	BC	ABC	G
$+1$	-1	-1	-1	$+1$	$+1$	$+1$	-1	0.367
$+1$	-1	-1	$+1$	$+1$	-1	-1	$+1$	0.532
$+1$	-1	$+1$	-1	-1	$+1$	-1	$+1$	0.495
$+1$	-1	$+1$	$+1$	-1	-1	$+1$	-1	0.489
$+1$	$+1$	-1	-1	-1	-1	$+1$	$+1$	0.310
$+1$	$+1$	-1	$+1$	-1	$+1$	-1	-1	0.485
$+1$	$+1$	$+1$	-1	$+1$	-1	-1	-1	0.476
$+1$	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$	$+1$	0.440

colonne AB = interaction = produit terme à terme des colonnes A et B.
L'interaction d'ordre 3 est utilisée pour estimer la variance d'erreur.

Facteurs confondus

principe de construction d'un plan fractionnaire :
confondre une interaction inutile avec un nouveau facteur

MU	A	B	C	$D = AB$	$E = AC$	$F = BC$	ABC
$+$	$-$	$-$	$-$	$-$	$+$	$+$	$+$
$+$	$-$	$-$	$+$	$+$	$+$	$-$	$-$
$+$	$-$	$+$	$-$	$+$	$-$	$+$	$-$
$+$	$-$	$+$	$+$	$-$	$-$	$-$	$+$
$+$	$+$	$-$	$-$	$+$	$-$	$-$	$+$
$+$	$+$	$-$	$+$	$-$	$-$	$+$	$-$
$+$	$+$	$+$	$-$	$-$	$+$	$-$	$-$
$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$

- les effets de D et AB ne peuvent être estimés séparément ; ils sont **confondus** ou aliasés.
- 8 simulations au lieu de $2^6 = 64$
- plan 2^{6-3} = fraction $1/8$ du plan complet 2^6.

Modèle

- 3 facteurs à 2 niveaux :
  effet de U (vecteur formé de -1 et $+1$) : $e(U) = 1/N < U, G > \in \mathbb{R}$

$$\mathbb{E}(G) = \sum_{a,b,c} A^a B^b C^c e(A^a B^b C^c)$$

  $a, b, c \in \{0, 1\}$, (ex. $A^1 B^1 C^0 = AB$ vecteur colonne)
- Exemple :
  - plan complet
    $\mathbb{E}(G_{+,+,-}) = \mu + e(A) + e(B) - e(C) + e(AB) - e(AC) - e(BC) - e(ABC)$
  - plan fractionnaire
  - $\mathbb{E}(G_{+,+,-,+,-,-}) = \mu + e(A) + e(B) - e(C) + e(D) - e(E) - e(F) - e(ABC)$

Estimation des effets

- effet $e(A^a B^b C^c) = 1/N < A^a B^b C^c, G >$ (p.s. de colonnes) avec $a, b, c \in \{0, 1\}$ Ex. : $N = 8$, $e(A^1 B^1 C^0) = e(AB) = 1/8 < AB, G >$.
- Effet principal de A :
  $e(A) = \frac{1}{2}(\bar{G}_{+..} - \bar{G}_{-..}) = \frac{1}{8}[(G_{+--} + G_{+-+} + G_{++-} + G_{+++}) - (G_{---} + G_{--+} + G_{-+-} + G_{-++})] = \frac{1}{8} < A, G >$
- Effet d'interaction (plan complet)
  $e(AB) = \frac{1}{2}[e(A|B = +1) - e(A|B = -1)]$
  $e(AB) = \frac{1}{2}\{\frac{1}{4}[G_{++-} + G_{+++} - G_{-+-} - G_{-++}] - \frac{1}{4}[G_{+--} + G_{+-+} - G_{---} - G_{--+}]\} = \frac{1}{8} < AB, G >$

Résolution

- un plan de résolution $\mathcal{R}$ permet d'analyser sans confusions :
  - les interactions d'ordre $(\mathcal{R}-1)/2$ si $\mathcal{R}$ est impair,
  - les interactions d'ordre $(\mathcal{R}-2)/2$ si $\mathcal{R}$ est pair.
- Ex. : le plan 2^{6-3}, de résolution $\mathcal{R}=3$, est noté 2^{6-3}_{III}.Il permet d'estimer les effets principaux sans confusions.

Plans possibles

Relation entre nombre K de facteurs et nombre N de simulations :

- si $\mathcal{R} = III$: $N \geq 1 + K$,
- si $\mathcal{R} = IV$: $N \geq 2K$,
- si $\mathcal{R} = V$: $N \geq 1 + K + K(K-1)/2$.

Nombre *maximal* κ de facteurs pour un plan de résolution V :

- de taille $N = 2^s$,
- donnant des estimateurs de variance minimale,

s	4	5	6	7	8	9
N	16	32	64	128	256	512
κ	5	6	8	11	17	≥ 23

R

```
arguments de regular.fraction (auteur Herve Monod):
 nb facteurs, nb niveaux, r tq N = p^r unités, résolution
> plan = regular.fraction(6,2,3,3)$plan
> fich = data.frame(A=as.factor(plan[,1]),
          B=..., F=as.factor(plan[,6]),
          Y=c(.367, .532,.495, .489, .310, .485, .476, .440))
> fich.aov = aov(Y ~ A + B + C + D +E + F, fich)
> summary(fich.aov)
```

R

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	1	0.0037	0.0037	18.49	0.15
B	1	0.0053	0.0053	26.52	0.12
C	1	0.0111	0.0111	55.5	0.085
D	1	0.00016	0.00016	0.8	0.53
E	1	0.00005	0.00005	0.25	0.70
F	1	0.018	0.018	91.203	0.066
Residuals	1	0.0002	0.0002		

Discussion

- généralisation avec un nombre de niveaux supérieur à deux,
- il n'existe pas forcément un plan fractionnaire dans toutes les configurations.

Pavage de Ω

- **objectif** : prendre en compte les variations du modèle entre les noeuds,
- **proposition** : pavage régulier de Ω,
- pavés définis à partir des intervalles équiprobables des lois des incertitudes,
- tirage au hasard de R points dans chaque pavé,
- variabilité de l'échantillon à prendre en compte,
- calcul des indices de sensibilité,
- inférence à définir,
- N.B. : la stochasticité de la sortie (même si le modèle est déterministe) est induite par la loi des incertitudes.

Variante simple de Morris

- définir une trajectoire de pas non constant
- P point de l'échantillonnage de Morris = centre d'un pavé,
- bruitage uniforme des coordonnées dans un pavé,
- exemple dans $[0,1]^3$ avec $U \sim \mathcal{U}[-L/2, L/2]$ et L longueur du côté du pavé :

  $P_0 = (x_0, y_0, z_0)$, $P_1 = (x_1, y_0, z_0)$, $P_2 = (x_1, y_0, z_1)$, $P_3 = (x_1, y_1, z_1)$ centres des pavés initiaux

  - bruitage des coordonnées : $P_0 \to P_0' = (x_0' = x_0 + U, y_0' = y_0 + U, z_0' = z_0 + U)$,
  - bruitage dans la direction de x : $P_1 \to P_1' = (x_1' = x_1 + U, y_0, z_0)$, $\Delta_x = x_1' - x_0'$
  - bruitage dans la direction de z : $P_2 \to P_2' = (x_1, y_0, z_1' = z_1 + U)$, $\Delta_z = z_1' - z_0'$
  - bruitage dans la direction de y : $P_3 \to P_3' = (x_1, y_1' = y_1 + U, z_1')$, $\Delta_y = y_1' - y_0$
  - trajectoire $=(P_0', P_1', P_2', P_3')$
- la direction d'un pas est parallèle aux axes du repère,
- les indices prennent en compte les longueurs différentes des pas des trajectoires,
- *Ecrire la fonction R associée.*

Variante de Morris par balayage de Ω

- **Simplex-based screening designs for estimating metamodels** , *Reliability Engineering and System Safety* de G. Pujol (2008),
- trajectoire définie par les $K+1$ d'un simplexe de dimension K de $\Omega = [0,1]^K$ (analogue au triangle en dimension K),
- tirage aléatoire de la position et de la direction du simplexe,
- l'effet élémentaire est calculé en supposant un comportement linéaire du modèle et pas d'interaction entre les facteurs dans le voisinage du simplexe,
- calcul des coefficients de l'hyperplan passant par les sommets
- Exemple $K = 2$:
  approximation locale de $\mathcal{G}(x, y)$ par un plan d'équation $a_0 + a_1 x + a_2 y$
  $\Rightarrow$ calcul de a_0,a_0, et a_2 à l'aide de trois points

  $\Rightarrow \frac{\Delta G}{\Delta_x} \simeq a_1$ et $\frac{\Delta G}{\Delta_y} \simeq a_2$

Variante d'anova

- découpage de la gamme d'un facteur en intervalles de même amplitude,
- un niveau du facteur discrétisé correspond à un intervalle,
- tirage de R points à l'intérieur de chaque pavé,
- le modèle d'anova est représenté par une fonction en escalier (dans chaque pavé $\mathcal{G}$ est ajusté à l'aide d'une constante),
- pb : variabilités différentes entre pavés

R

- tirage de l'échantillon

```
tirage.r = function(PAV,  binf, bsup, Nbclass){
#-------------------------------------------------------------
#  TIRAGE uniforme des coordonnées d'un point
#          dans le pavé PAV de R^K
#-------------------------------------------------------------
#  binf et bsup : vecteurs des bornes inf et sup des facteurs,
#  Nbclass : nbre de classes par facteur,
#  PAV : vecteur de K numéros de classe (1 à Nbclass)
#  sortie = vecteur de K éléments
   K = length(PAV)
   bornes = matrix(NA,nrow=K, ncol=2)
   bornes[,1]= binf + (PAV-1)*(bsup-binf)/Nbclass
   bornes[,2]= binf + PAV*(bsup-binf)/Nbclass
   cc = numeric(K)
for(i in 1:K) cc[i] = runif(1,min=bornes[i,1],max=bornes[i,2])
   cc
    }
```

R

- table des simulations et anova

```
> K=2;Nbclass=3; r=2; binf=c(-1,-1); bsup=c(1,1);
> facteurs = expand.grid(X=1:Nbclass, Y=1:Nbclass, R=1:r)
# simu = matrice K x nrow(facteurs)
>simu = apply(facteurs[,1:K],1, tirage.r, binf, bsup, Nbclass)
> data = data.frame(X=as.factor(facteurs[,1],
                    Y=as.factor(facteurs[,2],
                    Z =  mon.modele(simu) )
 >aov.out= aov(Z ~X + Y + X:Y, data) ou aov(Z ~(X + Y)^2, data)
 > summary(aov.out)
```

	Df	Sum Sq	Mean Sq	F value	Pr($> F$)
A	2	144.78	72.39	39.294	3.573e-05
B	2	2674.78	1337.39	725.965	1.130e-10
AB	4	239.56	59.89	32.509	2.327e-05
Residuals	9	16.58	1.84		

Utilisation conjointe de Morris et de plans fractionnaires

Proposition *ouverte* à discussion :

- pavage régulier de Ω,
- déplacement entre pavés avec trajectoires de Morris,
- plan fractionnaire dans chaque pavé,
- chaotisation de l'ordre des facteurs dans le plan,
- estimation des moyennes, variance, effets de $\mathcal{G}$ dans les pavés,
- indices de sensibilité à définir à partir des effets des facteurs dans les pavés,
- contacter CB pour compléments.