

Extrêmes multivariés: estimation de la dépendance par mélange de Dirichlet

Anne Sabourin

Institut Camille Jordan, Lyon 1,
Laboratoire des Sciences du Climat et de l'Environnement.
Directeurs de thèse: Anne-Laure Fougères (Lyon 1), Philippe Naveau (LSCE).

Applibugs, 26 Juin 2012

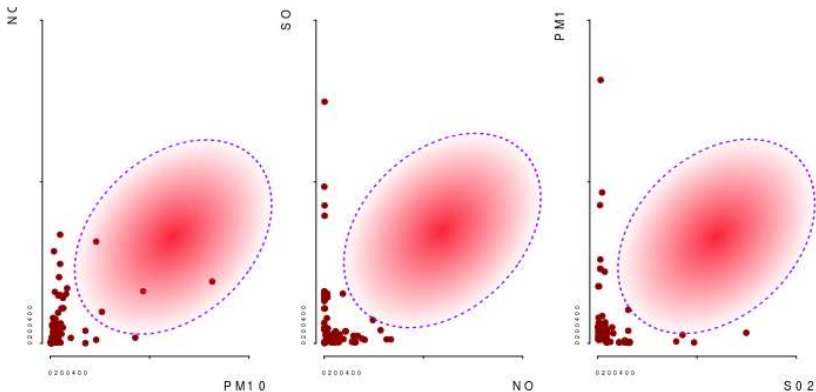
Objectif : modéliser la dépendance entre événements rares.

- **Valeurs extrêmes multivariées** : c.f De Haan, Resnick (70's, 80's).
- Inférence : problème ouvert, non paramétrique.
- **“Dirichlet Mixture model”** : semi-paramétrique. **Boldi, Davison, 2007**
 - *Problème* : Contraintes \Rightarrow Inférence bayésienne compliquée.
 - *Idée* : Changer la paramétrisation.

Qualité de l'air, Leeds, UK

Cooley *et al.*, 2010, Boldi and Davison, 2007, Heffernan and Tawn, 2004.

- Données de concentration atmosphérique pour 5 polluants : O_3 , NO_2 , PM_{10} (poussières fines), NO , SO_2 .
- Maximum journalier sur cinq saisons : 100 plus grandes observations (sur $\simeq 500$)



- 1 Extrêmes multivariés
- 2 Modèle contraint de mélange de Dirichlet : Boldi, Davison
- 3 Modèle reparamétré : sans contraintes
- 4 Résultats : dimension 3 et 5
- 5 Conclusion

- 1 Extrêmes multivariés
- 2 Modèle contraint de mélange de Dirichlet : Boldi, Davison
- 3 Modèle reparamétré : sans contraintes
- 4 Résultats : dimension 3 et 5
- 5 Conclusion

Extrêmes multivariés : décomposition polaire

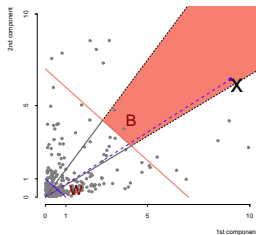
$$\mathbf{X} = (X_1, X_2) ; \quad \mathbf{P}(X_i \leq x) = e^{-\frac{1}{x}} .$$

$$R = X_1 + X_2 : \text{'rayon'} , \quad \mathbf{W} = \frac{\mathbf{X}}{R} : \text{'angle'} = \text{point du simplexe } \mathbf{S}_2 .$$

$$\mathbf{W} \in \mathbf{S}_2 = \{(w_1, w_2) : w_1 + w_2 = 1, w_i \geq 0\} .$$

- Loi jointe des extrêmes :

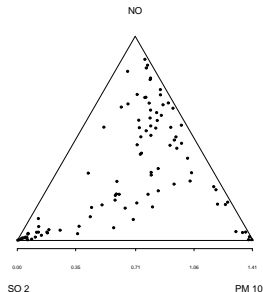
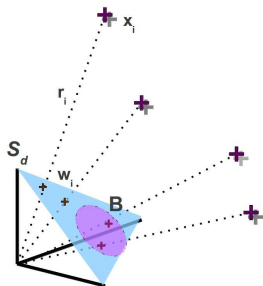
$$\mathbf{P}(\mathbf{W} \in B, R > r) ?$$



Extrêmes multivariés : décomposition polaire

- Dimension 3 : $\mathbf{X} = (X_1, X_2, X_3)$
- $\mathbf{W} = (w_1, w_2, w_3)$: point dans un triangle :

le simplexe $\mathbf{S}_3 = \left\{ \mathbf{w} = (w_1, w_2, w_3) : \sum_{i=1}^3 w_i = 1, w_i \geq 0 \right\}$.



Leeds

Extrêmes multivariés : décomposition polaire

Résultat fondamental (de Haan, Resnick) :

$$\mathbf{P}(\mathbf{W} \in B, R > r) \underset{r \rightarrow \infty}{\sim} \frac{1}{r} H(B)$$

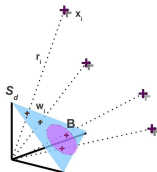
La loi limite des angles H = 'mesure spectrale' = 'mesure angulaire' détermine la loi jointe des grandes observations.

- Seule contrainte sur H :

$$\forall i \in \{1, \dots, d\}, \int_{\mathbf{S}_d} w_i dH(\mathbf{w}) = \frac{1}{d}. \quad (1)$$

Centre de gravité de H = centre du simplexe.

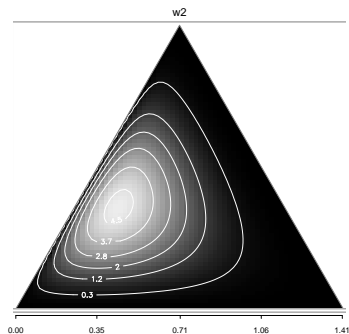
- Ensemble des H 's admissibles : famille non paramétrique.



- 1 Extrêmes multivariés
- 2 **Modèle contraint de mélange de Dirichlet : Boldi, Davison**
- 3 Modèle reparamétré : sans contraintes
- 4 Résultats : dimension 3 et 5
- 5 Conclusion

$$\forall \mathbf{w} \in \overset{\circ}{\mathbf{S}}_d, \text{diri}(\mathbf{w} \mid \boldsymbol{\mu}, \nu) = \frac{\Gamma(\nu)}{\prod_{i=1}^d \Gamma(\nu \mu_i)} \prod_{i=1}^d w_i^{\nu \mu_i - 1}.$$

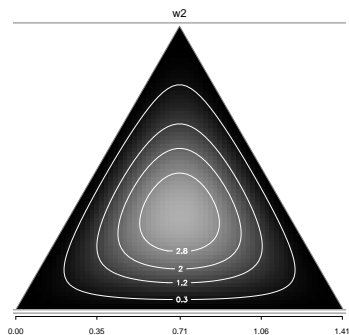
- $\boldsymbol{\mu} \in \overset{\circ}{\mathbf{S}}_d$: paramètre (vecteur) de localisation : 'centre'.
- $\nu > 0$: paramètre de concentration.



ex : $\overset{w3}{\boldsymbol{\mu}} = (0.15, 0.35, 0.5)$, $\nu = 9$.

$$\forall \mathbf{w} \in \overset{\circ}{\mathbf{S}}_d, \text{diri}(\mathbf{w} \mid \boldsymbol{\mu}, \nu) = \frac{\Gamma(\nu)}{\prod_{i=1}^d \Gamma(\nu \mu_i)} \prod_{i=1}^d w_i^{\nu \mu_i - 1}.$$

- $\boldsymbol{\mu} \in \overset{\circ}{\mathbf{S}}_d$: paramètre (vecteur) de localisation : 'centre'.
- $\nu > 0$: paramètre de concentration.



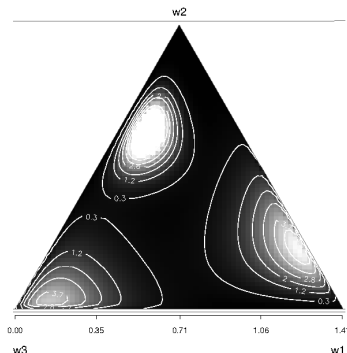
H^{w3} valide : $\boldsymbol{\mu} = (1/3, 1/3, 1/3)^{\text{w1}}$.

Mélange de k lois de Dirichlet

- $\boldsymbol{\mu} = \boldsymbol{\mu}_{\cdot, 1:k}$, $\boldsymbol{\nu} = \nu_{1:k}$, $\mathbf{p} = p_{1:k}$

$$h(\boldsymbol{\mu}, \mathbf{p}, \boldsymbol{\nu})(\mathbf{w}) = \sum_{m=1}^k p_m \text{diri}(\mathbf{w} \mid \boldsymbol{\mu}_{\cdot, m}, \nu_m)$$

- Contrainte de moments sur $(\boldsymbol{\mu}, \mathbf{p})$: $\sum_{m=1}^k p_m \boldsymbol{\mu}_{\cdot, m} = (\frac{1}{d}, \dots, \frac{1}{d})$.



Espace des paramètres : union disjointe $\Theta = \coprod_{k=1}^{\infty} \Theta_k$.

Chaque espace des paramètres Θ_k est contraint :

$$\Theta_k = \left\{ \boldsymbol{\theta} = (\boldsymbol{\mu}_{\cdot,1:k}, p_{1:k}, \nu_{1:k}) : \sum_{m=1}^k p_m \boldsymbol{\mu}_{\cdot,m} = \left(\frac{1}{d}, \dots, \frac{1}{d} \right) \right\}.$$

Inférence bayésienne :

- prior $\pi(\boldsymbol{\theta})$
- Posterior : $\pi_n(\boldsymbol{\theta}) \hat{=} \pi(\boldsymbol{\theta} | \mathbf{w}_{1:n})$ connu à une constante multiplicative près.
- π_n estimée par échantillonnage : Metropolis Hastings, sauts réversibles.

$\Theta_k = \text{espace contraint} \neq \text{pavé de } \mathbb{R}^n !!$

- Comment définit un prior π sur Θ_k ?
- Echantillonnage MCMC : Comment générer des propositions “raisonnables” ?

- Prior sur Θ_k :

$$\pi(\theta) = \pi_\nu(\nu)\pi_p(\mathbf{p})\pi_\mu(\boldsymbol{\mu} | \mathbf{p}),$$

avec

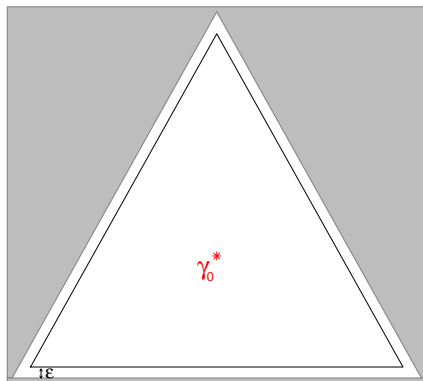
$$\pi_\mu(\boldsymbol{\mu} | \mathbf{p}) = \prod_{i \leq d, j \leq k} \pi_{i,j}(\boldsymbol{\mu}_{i,j} | \mathbf{p}, \{\boldsymbol{\mu}_{k,l}, (k,l) < (i,j)\})$$

⇒ Prior asymétrique (stick breaking) ⇒ concentration de la masse a priori dans un “coin” du simplexe.

- *propositions MCMC* : par fusion/division de paires de noyaux.
⇒ Taux d'acceptation de l'échantillonneur très faible.

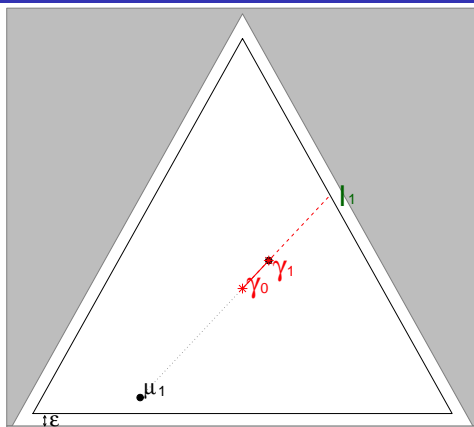
⇒ Pas d'inférence bayésienne en dimension > 3 .

- 1 Extrêmes multivariés
- 2 Modèle contraint de mélange de Dirichlet : Boldi, Davison
- 3 Modèle reparamétré : sans contraintes**
- 4 Résultats : dimension 3 et 5
- 5 Conclusion



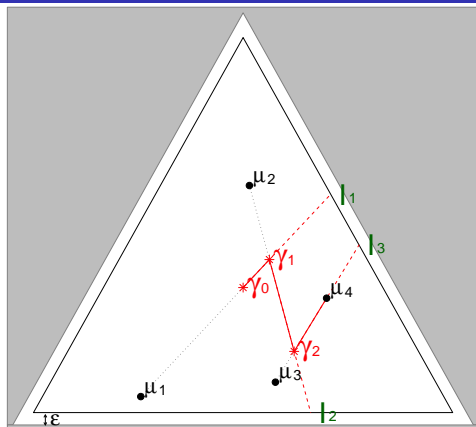
γ_m : barycentre des “noyaux après $\mu_{\cdot, m}$ ” : $\mu_{\cdot, m+1}, \dots, \mu_{\cdot, k}$.

$$\gamma_m = \sum_{j>m} \frac{p_j}{1 - \sum_1^m p_i} \mu_{\cdot, j}$$



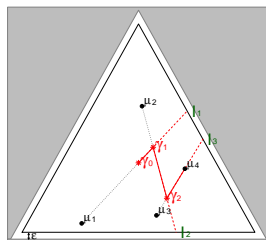
$$\mu_{.,1}, e_1 \Rightarrow \gamma_1 : \frac{\overline{\gamma_0 \gamma_1}}{\gamma_0 l_1} = e_1 ;$$

$$\Rightarrow p_1$$



$$\mu_{.,3}, e_3 \Rightarrow \gamma_3 : \frac{\overline{\gamma_2 \gamma_3}}{\gamma_2 l_3} = e_3 ; \quad \mu_{.,4} = \gamma_3.$$

$$\Rightarrow p_3, p_4$$



- On s'est donné

$$(\mu_{.,1:k-1}, \mathbf{e}_{1:k-1});$$

On en déduit

$$(\mu_{.,1:k}, \rho_{1:k}).$$

- On peut paramétrer la densité h par

$$\theta = (\mu_{.,1:k-1}, \mathbf{e}_{1:k-1}, \nu_{1:k})$$

- Espace des paramètres non contraint (union d'espaces produit) :

$$\Theta = \prod_{k=1}^{\infty} \Theta_k; \quad \Theta_k = \left\{ (\mathbf{s}_d)^{k-1} \times [0, 1]^{k-1} \times (0, \infty]^{k-1} \right\}$$

- Inférence sur k , $\theta_k = (\boldsymbol{\mu}_{1:k-1}, \mathbf{e}_{1:k-1}, \nu_{1:k})$
- Restriction (compacité) : $k \leq 15$, $\nu < \nu_{\max}$, etc ...

- Prior :

$k \sim$ Géométrique tronquée

$\boldsymbol{\mu}_{\cdot,m} | (\boldsymbol{\mu}_{\cdot,1:m-1}, \mathbf{e}_{1:m-1}) \sim$ Dirichlet

$e_m | (\boldsymbol{\mu}_{\cdot,1:m}, \mathbf{e}_{1:m-1}) \sim$ Beta

$\nu_m \sim \log N$

- *Echantillonnage du posterior* : par MCMC (sauts réversibles).

- *Principe de l'algo* :

proposer $\boldsymbol{\mu}_m^*$ dans des régions à forte vraisemblance

i.e. : près des données.

Trois types de propositions :

- *Classique* : un $\mu_{\cdot,m}$, e_m ou un ν_m est modifié. (Gibbs)

$$\mu_{\cdot,m}^* \sim \sum_{j=1}^n p(\mathbf{w}_j) \text{diri}(\cdot | \mathbf{w}_j, \nu(\mathbf{w}_j))$$

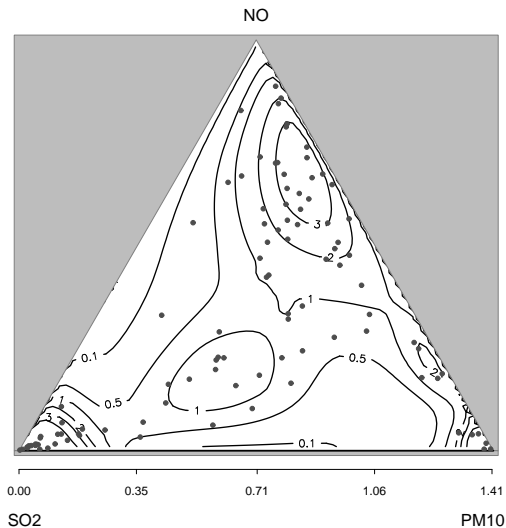
- *Trans-dimensionnelle* : Un composant $(\mu_{\cdot,k}, e_k, \nu_{k+1})$ est rajouté ou supprimé. (Green, 1995)
- *“Shuffle”* : Permutation (évite la division des derniers poids par stick-breaking).

- 1 Extrêmes multivariés
- 2 Modèle contraint de mélange de Dirichlet : Boldi, Davison
- 3 Modèle reparamétré : sans contraintes
- 4 Résultats : dimension 3 et 5
- 5 Conclusion

En dimension 3

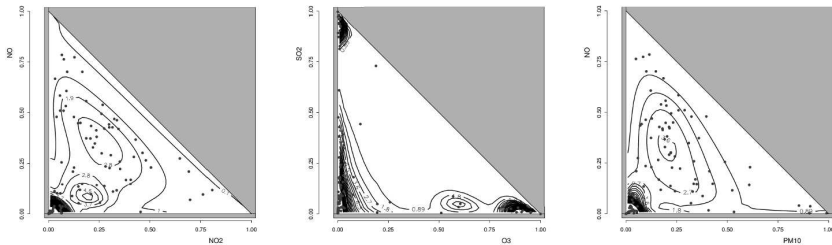
Leeds : PM_{10} , NO , SO_2 .

MCMC : $T_2 = 60 \cdot 10^3$; $T_1 = 25 \cdot 10^3$.

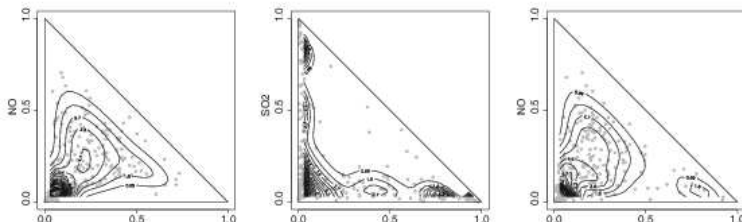


Dimension 5, Leeds : densité "projetée" par paires

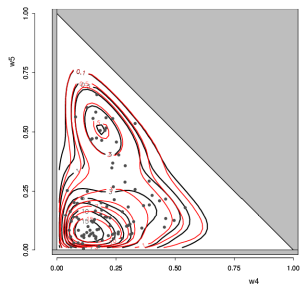
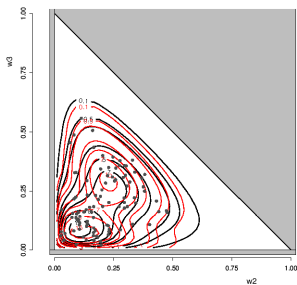
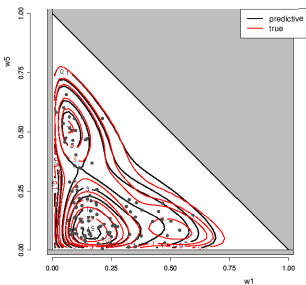
Prédictive par MCMC ($T_2 = 10^6$, $T_1 = 5 \cdot 10^5$)



Boldi, Davison, méthode non bayésienne (EM)



Dimension 5, Données simulées, mélange de 4 noyaux



$$T_2 = 150 \cdot 10^3, T_1 = 50 \cdot 10^3.$$

- 1 Extrêmes multivariés
- 2 Modèle contraint de mélange de Dirichlet : Boldi, Davison
- 3 Modèle reparamétré : sans contraintes
- 4 Résultats : dimension 3 et 5
- 5 Conclusion

- Re-paramétrisation du modèle de Boldi et Davison
⇒ utilisable en dimension 5, dans un cadre Bayésien.
- Convergence asymptotique prouvée : *posterior* + Metropolis
- Diagnostiques de convergence : intégration contre des fonctions test de Dirichlet.



M.-O. Boldi and A. C. Davison.

A mixture model for multivariate extremes.

JRSS : Series B (Statistical Methodology), 69(2) :217–229, 2007.



Coles, SG and Tawn, JA

Modeling extreme multivariate events

JR Statist. Soc. B, 53 :377–392, 1991



P.J. Green.

Reversible jump Markov chain Monte Carlo computation and Bayesian model determination.

Biometrika, 82(4) :711, 1995.



Resnick, S.I.

Heavy-Tail Phenomena : Probabilistic and Statistical Modeling.

Springer Series in Operations Research and Financial Engineering, 2007



Roberts, G.O. and Rosenthal, J.S.

Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains

The Annals of Applied Probability,16,4,2123 :2139, 2006.



S. Guillotte, F. Perron, and J. Segers.

Non-parametric bayesian inference on bivariate extremes.

JRSS : Series B (Statistical Methodology), 2011.



Einmahl, J.H.J. and Segers, J.

Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution.

The Annals of Statistics,37, 5B, 2953 :2989, 2009.

- $\varphi : \mathbf{S}_p \rightarrow \mathbf{R}^+, \mathcal{C}_b$,
- On “transforme” l'échantillon $(\boldsymbol{\theta}(t))_{t=1:T}$:

$$f(\boldsymbol{\theta}) = \langle \varphi, h_{\boldsymbol{\theta}} \rangle = \int_{\mathbf{S}_p} \varphi(\mathbf{w}) h_{\boldsymbol{\theta}}(\mathbf{w}) d\mathbf{w}$$

Dimension quelconque : Intégration contre une fonction test

- $\varphi : \mathbf{S}_p \rightarrow \mathbf{R}^+, \mathcal{C}_b$,
- On “transforme” l'échantillon $(\boldsymbol{\theta}(t))_{t=1:T}$:

$$f(\boldsymbol{\theta}) = \langle \varphi, h_{\boldsymbol{\theta}} \rangle = \int_{\mathbf{S}_p} \varphi(\mathbf{w}) h_{\boldsymbol{\theta}}(\mathbf{w}) d\mathbf{w}$$

- Si φ : Dirichlet bornée, expression explicite pour $\langle \varphi, h_{\boldsymbol{\theta}(t)} \rangle$!

Fait

Si φ_1, φ_2 sont Dirichlet, φ_1 bornée, alors $\varphi_1 \varphi_2$ est encore Dirichlet.

Proposition

- *Ergodicité :*

$$\frac{1}{T} \sum_{t=1}^T \langle \varphi, h_{\boldsymbol{\theta}(t)} \rangle \xrightarrow{T \rightarrow \infty} \mathbf{E}_{\pi_n} (\langle \varphi, h_{(\cdot)} \rangle)$$

- *Consistance de la loi a posteriori : $h_0 =$ "vraie" distribution*

$$\mathbf{E}_{\pi_n} (\langle \varphi, h_{(\cdot)} \rangle) \xrightarrow{n \rightarrow \infty} \langle \varphi, h_0 \rangle$$

Proposition

- *Ergodicité* :

$$\frac{1}{T} \sum_{t=1}^T \langle \varphi, h_{\theta(t)} \rangle \xrightarrow{T \rightarrow \infty} \mathbf{E}_{\pi_n} (\langle \varphi, h_{(\cdot)} \rangle)$$

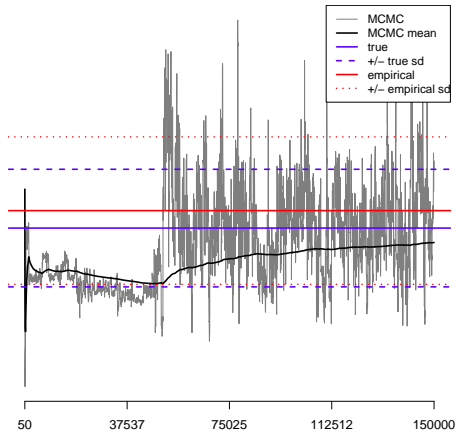
- *Consistance de la loi a posteriori* : $h_0 = \text{“vraie”}$ distribution

$$\mathbf{E}_{\pi_n} (\langle \varphi, h_{(\cdot)} \rangle) \xrightarrow{n \rightarrow \infty} \langle \varphi, h_0 \rangle$$

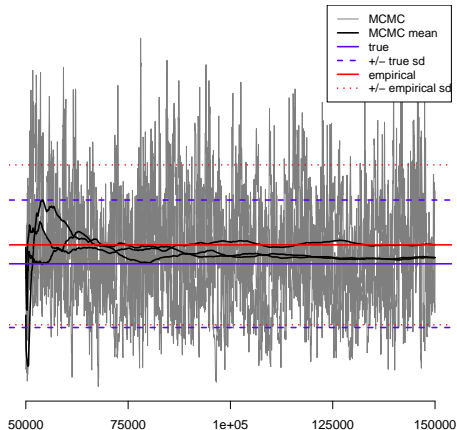
- Si $h_0 = h_{\theta_0}$ (données simulées) : Expression exacte
- Si h_0 inconnu : $\langle \varphi, h_0 \rangle = \mathbf{E}_{h_0}(\varphi)$. Estimateur empirique

$$\hat{\mathbf{E}}(\varphi) = \frac{1}{n} \sum_{j=1}^n \varphi(\mathbf{w}_j)$$

Dimension 5, données simulées (4 noyaux)



Dimension 5, données simulées (4 noyaux)



Stationnarité (Heidelberger) et Gelman (“mixing”) : OK