

Lois empiriques et quasi-empiriques

Observations x_1, \dots, x_n ; observations triées $x_{(1)}, \dots, x_{(n)}$.

Fonction de répartition empirique monte de $1/n$ à chaque observations:

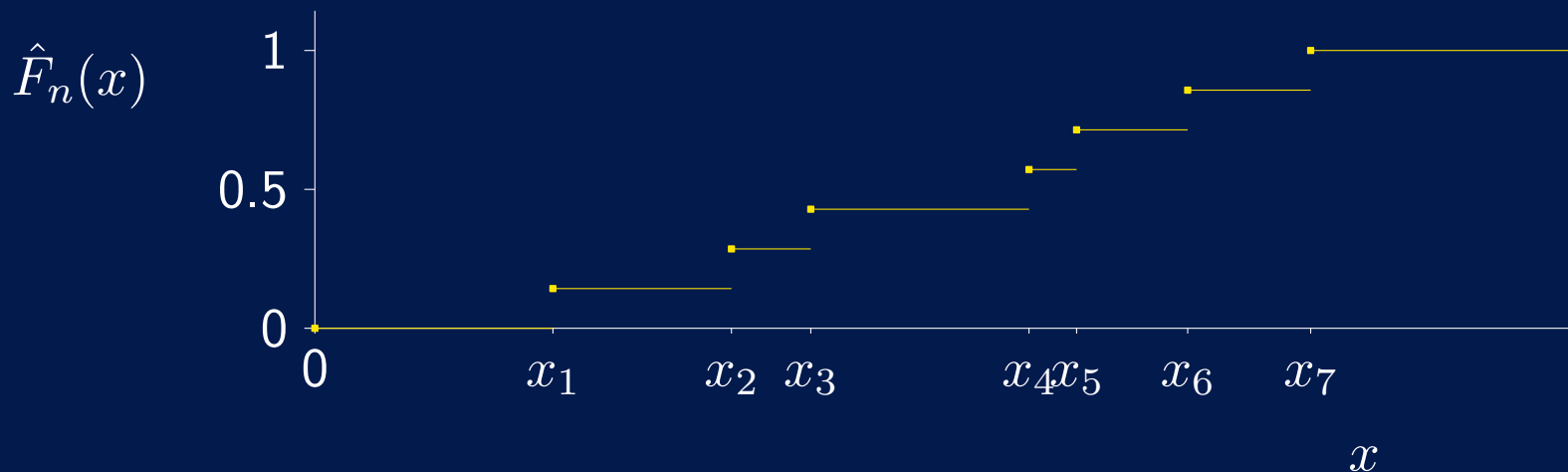
$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I[x_i \leq x].$$

Lois empiriques et quasi-empiriques

Observations x_1, \dots, x_n ; observations triées $x_{(1)}, \dots, x_{(n)}$.

Fonction de répartition empirique monte de $1/n$ à chaque observations:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I[x_i \leq x].$$

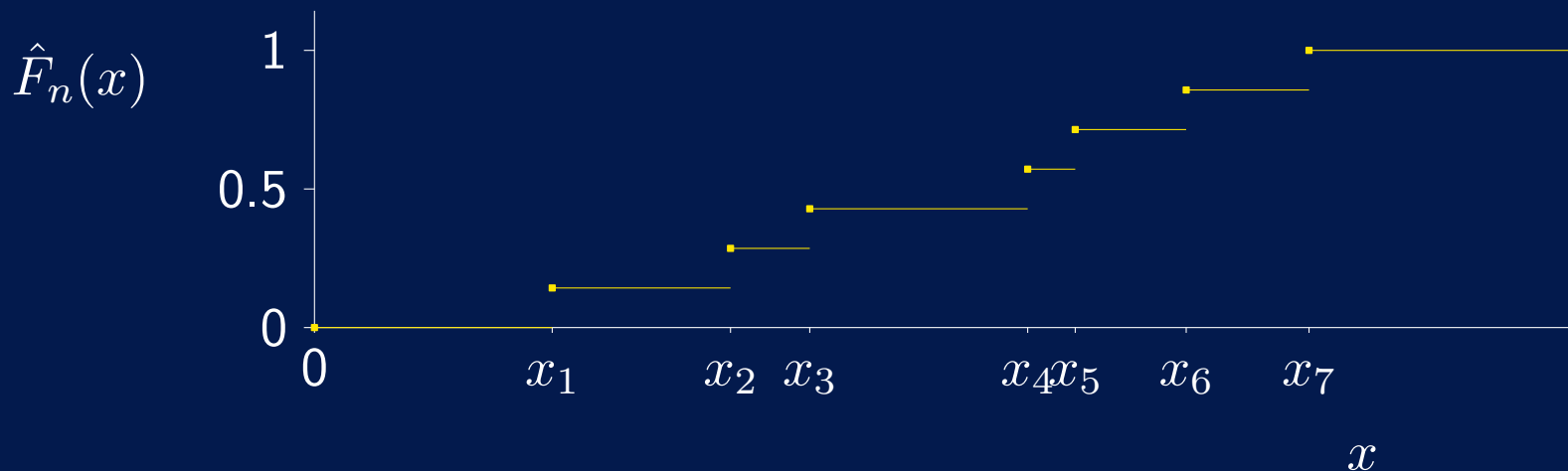


Lois empiriques et quasi-empiriques

Observations x_1, \dots, x_n ; observations triées $x_{(1)}, \dots, x_{(n)}$.

Fonction de répartition empirique monte de $1/n$ à chaque observations:

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I[x_i \leq x].$$



Générer des valeurs selon \hat{F}_n correspond à piger des valeurs au hasard dans l'échantillon, avec remplacement.

Si x_1, \dots, x_n sont les valeurs de v.a.'s i.i.d. de répartition F , alors

$$D_n = \sup_{-\infty < x < \infty} |\hat{F}_n(x) - F(x)| \rightarrow 0 \quad \text{a.p.1 lorsque } n \rightarrow \infty,$$

et D_n converge en probabilité en $O(n^{-1/2})$.

Ainsi, F_n devient très proche de F lorsque n est grand.

Si x_1, \dots, x_n sont les valeurs de v.a.'s i.i.d. de répartition F , alors

$$D_n = \sup_{-\infty < x < \infty} |\hat{F}_n(x) - F(x)| \rightarrow 0 \quad \text{a.p.1 lorsque } n \rightarrow \infty,$$

et D_n converge en probabilité en $O(n^{-1/2})$.

Ainsi, F_n devient très proche de F lorsque n est grand.

Si n est grand, on peut donc générer des valeurs directement de \hat{F}_n au lieu d'estimer une loi paramétrique.

Si x_1, \dots, x_n sont les valeurs de v.a.'s i.i.d. de répartition F , alors

$$D_n = \sup_{-\infty < x < \infty} |\hat{F}_n(x) - F(x)| \rightarrow 0 \quad \text{a.p.1 lorsque } n \rightarrow \infty,$$

et D_n converge en probabilité en $O(n^{-1/2})$.

Ainsi, F_n devient très proche de F lorsque n est grand.

Si n est grand, on peut donc générer des valeurs directement de \hat{F}_n au lieu d'estimer une loi paramétrique.

Mais: on ne pourra générer que les mêmes valeurs qui sont dans l'échantillon!

Solution potentielle: lisser \hat{F}_n .

Variante continue linéaire par morceaux:

$$\tilde{F}_n(x) = \begin{cases} 0 & \text{if } x \leq x_{(1)}, \\ \frac{i-1}{n-1} + \frac{x - x_{(i)}}{(n-1)(x_{(i+1)} - x_{(i)})} & \text{if } x_{(i)} \leq x \leq x_{(i+1)}, \\ 1 & \text{if } x \geq x_{(n)}. \end{cases}$$

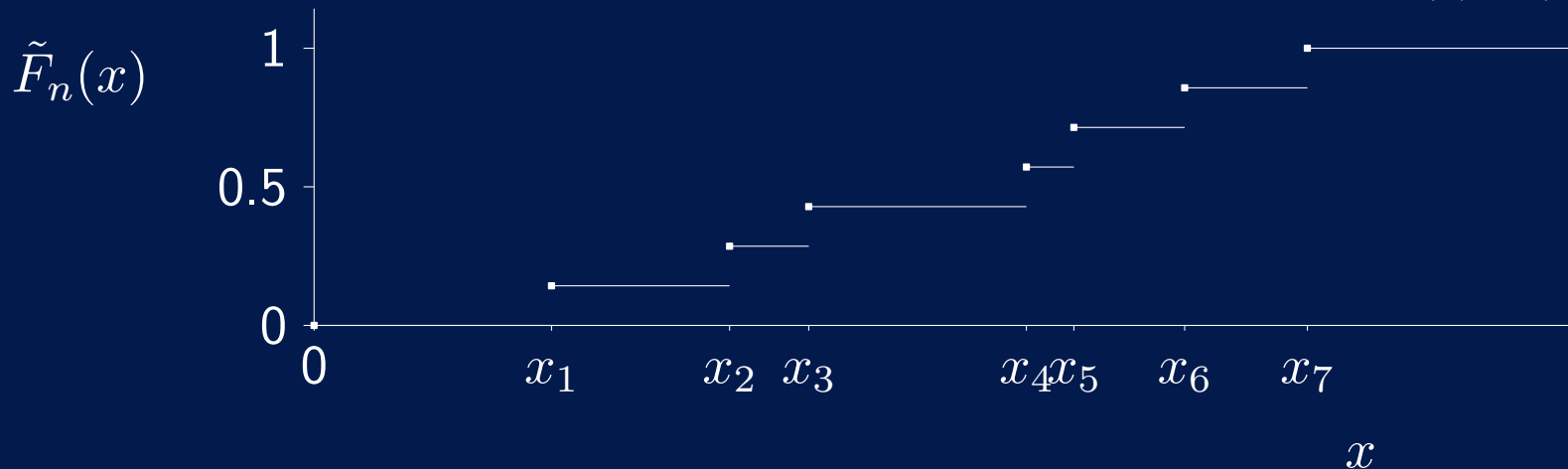
Possible de générer une valeur quelconque dans l'intervalle $(x_{(1)}, x_{(n)})$.

Solution potentielle: lisser \hat{F}_n .

Variante continue linéaire par morceaux:

$$\tilde{F}_n(x) = \begin{cases} 0 & \text{if } x \leq x_{(1)}, \\ \frac{i-1}{n-1} + \frac{x - x_{(i)}}{(n-1)(x_{(i+1)} - x_{(i)})} & \text{if } x_{(i)} \leq x \leq x_{(i+1)}, \\ 1 & \text{if } x \geq x_{(n)}. \end{cases}$$

Possible de générer une valeur quelconque dans l'intervalle $(x_{(1)}, x_{(n)})$.

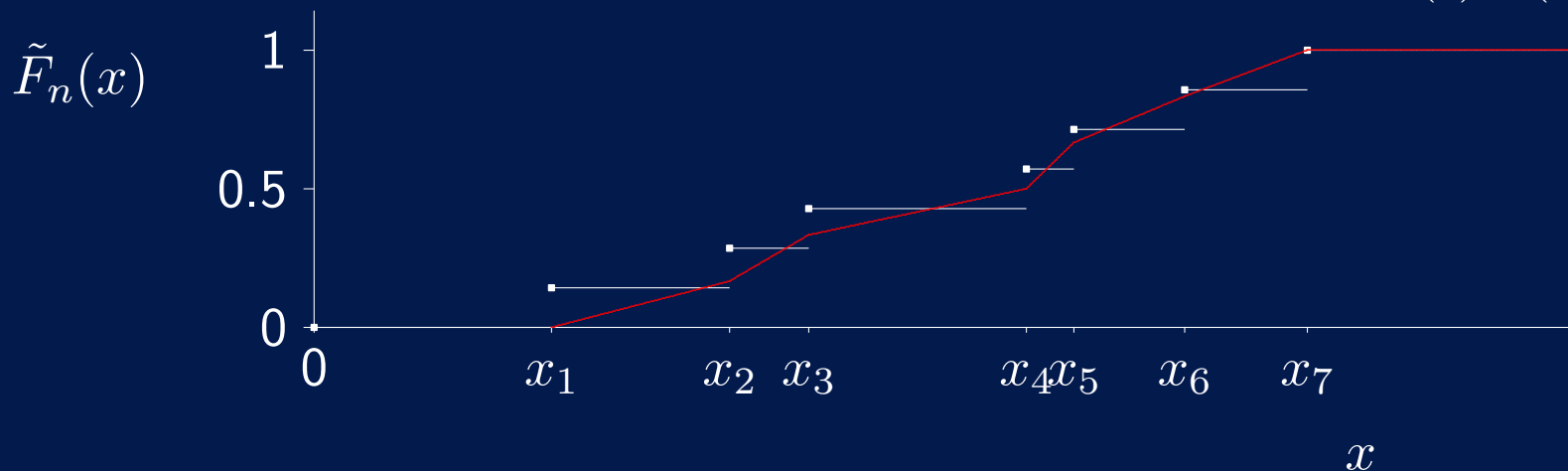


Solution potentielle: lisser \hat{F}_n .

Variante continue linéaire par morceaux:

$$\tilde{F}_n(x) = \begin{cases} 0 & \text{if } x \leq x_{(1)}, \\ \frac{i-1}{n-1} + \frac{x - x_{(i)}}{(n-1)(x_{(i+1)} - x_{(i)})} & \text{if } x_{(i)} \leq x \leq x_{(i+1)}, \\ 1 & \text{if } x \geq x_{(n)}. \end{cases}$$

Possible de générer une valeur quelconque dans l'intervalle $(x_{(1)}, x_{(n)})$.



Si on veut pouvoir générer des valeurs en dehors de $(x_{(1)}, x_{(n)})$, par exemple sur $[0, \infty)$ pour une durée de vie, on peut ajouter une queue à la distribution.

Si on veut pouvoir générer des valeurs en dehors de $(x_{(1)}, x_{(n)})$, par exemple sur $[0, \infty)$ pour une durée de vie, on peut ajouter une queue à la distribution.

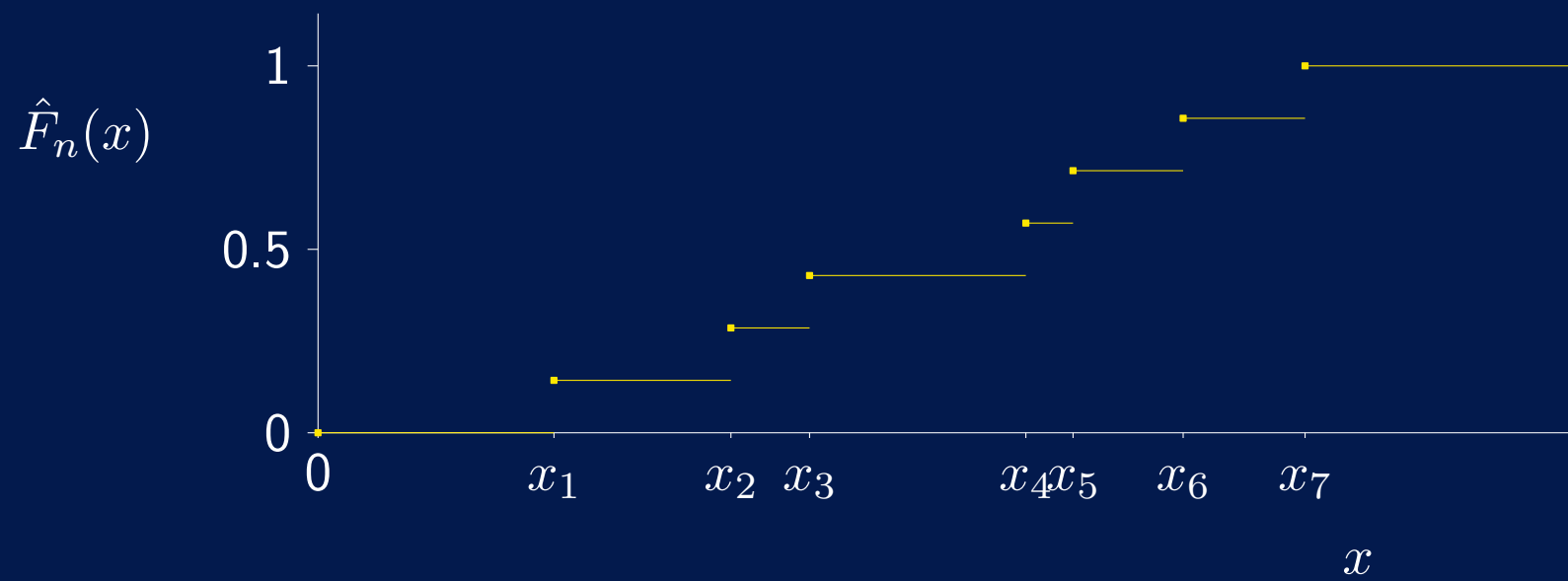
Par exemple, Bratley, Fox, and Schrage (1987) proposent une fonction linéaire par morceaux jusqu'à $x_{(n-k)}$, avec une queue exponentielle par la suite, ajustée pour obtenir la même moyenne que dans les données:

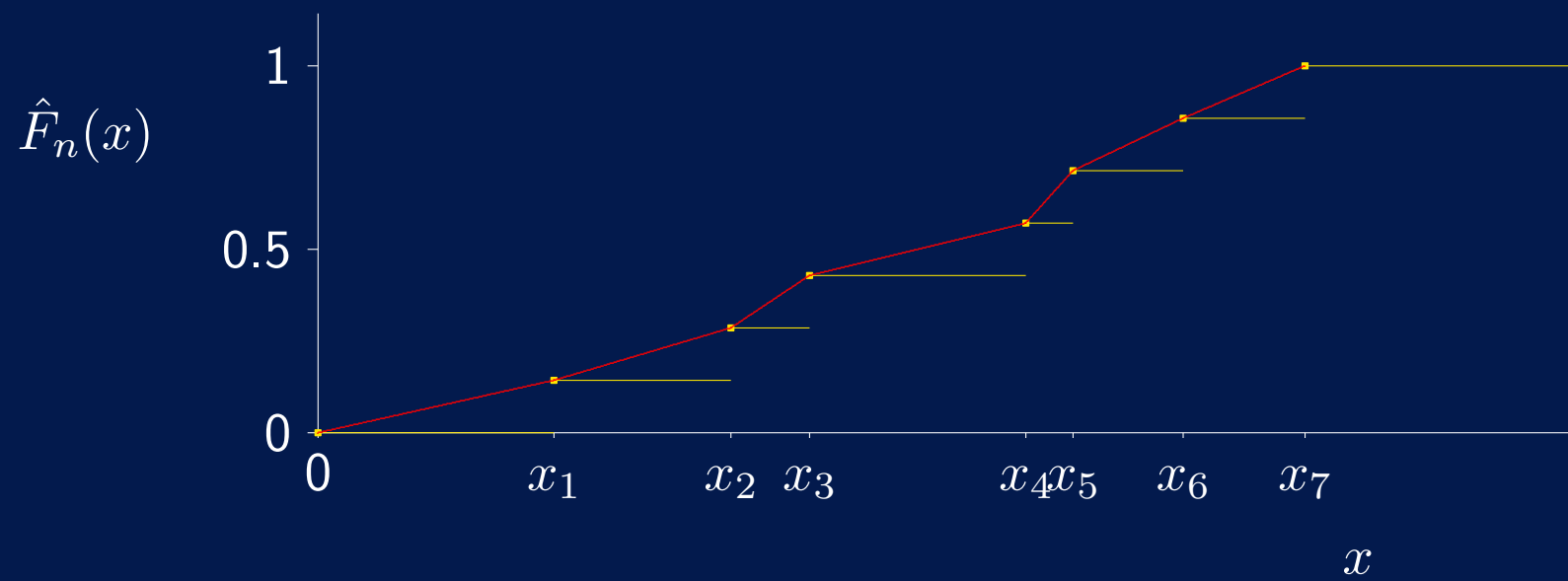
$$\check{F}_n(x) = \begin{cases} \frac{i}{n} + \frac{x - x_{(i)}}{(x_{(i+1)} - x_{(i)})n} & \text{if } x_{(i)} \leq x \leq x_{(i+1)}, 0 \leq i < n - k, \\ 1 - \frac{k}{n} \exp[-(x - x_{(n-k)})/\theta] & \text{if } x > x_{(n-k)}, \end{cases}$$

où $x_{(0)} = 0$, k est un entier de 1 à 5 (disons), et

$$\theta = \frac{1}{k} \left(\frac{x_{(n-k)}}{2} + \sum_{i=n-k+1}^n (x_{(i)} - x_{(n-k)}) \right).$$

On pourrait aussi ajuster la queue d'une autre loi (gamma, lognormale, etc.).





Estimation de densité

Les observations (données) x_1, \dots, x_n proviennent d'une densité f inconnue et on veut estimer f sans présumer de sa forme.

Estimation de densité

Les observations (données) x_1, \dots, x_n proviennent d'une densité f inconnue et on veut estimer f sans présumer de sa forme.

Peut-on estimer f par la densité associée à \tilde{F} ou \check{F} ?

Estimation de densité

Les observations (données) x_1, \dots, x_n proviennent d'une densité f inconnue et on veut estimer f sans présumer de sa forme.

Peut-on estimer f par la densité associée à \tilde{F} ou \check{F} ?

Ces densités sont constantes par morceaux (sauf pour la queue exponentielle).

La densité $\tilde{f}_n(x)$ de \tilde{F}_n est

$$\tilde{f}_n(x) = \frac{1}{(n-1)(x_{(i+1)} - x_{(i)})} \text{ pour } x_{(i)} \leq x \leq x_{(i+1)}.$$

Estimation de densité

Les observations (données) x_1, \dots, x_n proviennent d'une densité f inconnue et on veut estimer f sans présumer de sa forme.

Peut-on estimer f par la densité associée à \tilde{F} ou \check{F} ?

Ces densités sont constantes par morceaux (sauf pour la queue exponentielle).

La densité $\tilde{f}_n(x)$ de \tilde{F}_n est

$$\tilde{f}_n(x) = \frac{1}{(n-1)(x_{(i+1)} - x_{(i)})} \text{ pour } x_{(i)} \leq x \leq x_{(i+1)}.$$

Entre 2 observations proches, cette densité empirique devient très grande!
Elle est très irrégulière et ne converge pas vers f pour $n \rightarrow \infty$.

Exemple. Supposons que f est la densité $U(0, 1)$. Soit

$$d_n = \min_{1 \leq i < n} (x_{(i+1)} - x_{(i)}).$$

Alors $n(n-1)d_n \Rightarrow \text{Exponentielle}(1)$ quand $n \rightarrow \infty$.

Exemple. Supposons que f est la densité $U(0, 1)$. Soit

$$d_n = \min_{1 \leq i < n} (x_{(i+1)} - x_{(i)}).$$

Alors $n(n-1)d_n \Rightarrow$ Exponentielle(1) quand $n \rightarrow \infty$.

$$P \left[\max_{0 \leq x \leq 1} \tilde{f}_n(x) > y \right] = P \left[\frac{1}{(n-1)d_n} > y \right] = P[n(n-1)d_n < n/y] \approx 1 - e^{-n/y}.$$

Pour y fixé, converge vers 1 de manière exponentielle.
La densité \tilde{f}_n aura donc de très grands pics!

Exemple. Supposons que f est la densité $U(0, 1)$. Soit

$$d_n = \min_{1 \leq i < n} (x_{(i+1)} - x_{(i)}).$$

Alors $n(n-1)d_n \Rightarrow \text{Exponentielle}(1)$ quand $n \rightarrow \infty$.

$$P \left[\max_{0 \leq x \leq 1} \tilde{f}_n(x) > y \right] = P \left[\frac{1}{(n-1)d_n} > y \right] = P[n(n-1)d_n < n/y] \approx 1 - e^{-n/y}.$$

Pour y fixé, converge vers 1 de manière exponentielle.

La densité \tilde{f}_n aura donc de très grands pics!

Dans beaucoup de situations, une bonne approximation de F suffit, pas besoin de bien approximer f . Mais pas toujours (exercices 2.17 et 2.19).

Histogramme.

n observations sur un intervalle $[a, b]$.

On partitionne l'intervalle en m morceaux de longueur $h = (b - a)/m$.

La densité de l'histogramme, $f_{h,n}$, est constante sur chaque intervalle et proportionnelle au nombre d'observations dans l'intervalle.

Histogramme.

n observations sur un intervalle $[a, b]$.

On partitionne l'intervalle en m morceaux de longueur $h = (b - a)/m$.

La densité de l'histogramme, $f_{h,n}$, est constante sur chaque intervalle et proportionnelle au nombre d'observations dans l'intervalle.

Pour minimiser le "MISE" $E \int_a^b [f_{h,n}(x) - f(x)]^2 dx$, il faut choisir h tel que $h^3 n \int_a^b (f'(x))^2 dx \approx 6$. On a alors $m = O(n^{1/3})$ et $\text{MISE} = O(n^{-2/3})$.

Histogramme.

n observations sur un intervalle $[a, b]$.

On partitionne l'intervalle en m morceaux de longueur $h = (b - a)/m$.

La densité de l'histogramme, $f_{h,n}$, est constante sur chaque intervalle et proportionnelle au nombre d'observations dans l'intervalle.

Pour minimiser le "MISE" $E \int_a^b [f_{h,n}(x) - f(x)]^2 dx$, il faut choisir h tel que $h^3 n \int_a^b (f'(x))^2 dx \approx 6$. On a alors $m = O(n^{1/3})$ et $\text{MISE} = O(n^{-2/3})$.

Mieux: interpolation polygonale de l'histogramme.

Donne $\text{MISE} = O(n^{-4/5})$ avec $m = O(n^{1/5})$.

Estimateur de densité à noyaux.

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n k((x - x_i)/h),$$

où k est une densité appelée le noyau et $h > 0$ une constante appelée de facteur de lissage.

Estimateur de densité à noyaux.

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n k((x - x_i)/h),$$

où k est une densité appelée le noyau et $h > 0$ une constante appelée de facteur de lissage. Le changement de variable $y = (x - x_i)/h$ donne

$$\frac{1}{h} \int_{-\infty}^{\infty} k((x - x_i)/h) dx = \int_{-\infty}^{\infty} k(y) dy = 1.$$

Estimateur de densité à noyaux.

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n k((x - x_i)/h),$$

où k est une densité appelée le noyau et $h > 0$ une constante appelée de facteur de lissage. Le changement de variable $y = (x - x_i)/h$ donne

$$\frac{1}{h} \int_{-\infty}^{\infty} k((x - x_i)/h) dx = \int_{-\infty}^{\infty} k(y) dy = 1.$$

Très facile de générer des valeurs selon $f_n(x)$:

- (1) Générer I uniformément sur $\{1, \dots, n\}$;
- (2) générer D selon la densité k , indep. de I ;
- (3) retourner $X = x_I + hD$.

Estimateur de densité à noyaux.

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n k((x - x_i)/h),$$

où k est une densité appelée le noyau et $h > 0$ une constante appelée de facteur de lissage. Le changement de variable $y = (x - x_i)/h$ donne

$$\frac{1}{h} \int_{-\infty}^{\infty} k((x - x_i)/h) dx = \int_{-\infty}^{\infty} k(y) dy = 1.$$

Très facile de générer des valeurs selon $f_n(x)$:

- (1) Générer I uniformément sur $\{1, \dots, n\}$;
- (2) générer D selon la densité k , indep. de I ;
- (3) retourner $X = x_I + hD$.

La méthode se généralise aux lois multivariées (d dimensions).

Estimateur de densité à noyaux.

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n k((x - x_i)/h),$$

où k est une densité appelée le noyau et $h > 0$ une constante appelée de facteur de lissage. Le changement de variable $y = (x - x_i)/h$ donne

$$\frac{1}{h} \int_{-\infty}^{\infty} k((x - x_i)/h) dx = \int_{-\infty}^{\infty} k(y) dy = 1.$$

Très facile de générer des valeurs selon $f_n(x)$:

- (1) Générer I uniformément sur $\{1, \dots, n\}$;
- (2) générer D selon la densité k , indep. de I ;
- (3) retourner $X = x_I + hD$.

La méthode se généralise aux lois multivariées (d dimensions).

Difficulté principale: choix de k et h .

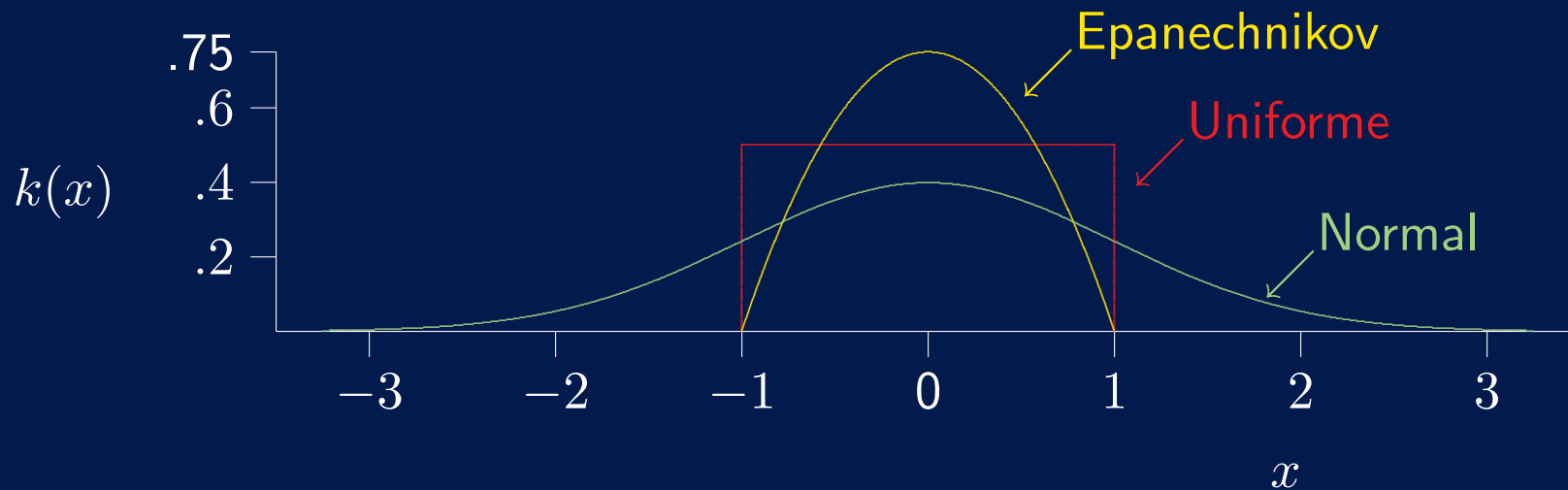
Pas de bonne méthode universelle pour cela.

Choix populaires de noyaux.

nom	loi	étendue	α_k	σ_k^2	efficacité
Epanechnikov	$2 \text{ Beta}(2, 2) - 1$	$[-1, 1]$	1.7188	1/5	1.000
triangulaire	$\text{Triangular}(-1, 1, 0)$	$[-1, 1]$	1.8882	1/6	0.986
uniforme	$U(-1, 1)$	$[-1, 1]$	1.3510	1/3	0.930
normal	$\text{Normale}(0, 1)$	$(-\infty, \infty)$	0.7764	1	0.951
logistique	$\text{Logistique}(0, 1)$	$(-\infty, \infty)$	0.4340	3.2899	0.888
Student-t(3)	$\text{Student-t}(3)$	$(-\infty, \infty)$	0.4802	3	0.674

Choix populaires de noyaux.

nom	loi	étendue	α_k	σ_k^2	efficacité
Epanechnikov	2 Beta(2, 2) - 1	$[-1, 1]$	1.7188	1/5	1.000
triangulaire	Triangular(-1, 1, 0)	$[-1, 1]$	1.8882	1/6	0.986
uniforme	$U(-1, 1)$	$[-1, 1]$	1.3510	1/3	0.930
normal	Normale(0, 1)	$(-\infty, \infty)$	0.7764	1	0.951
logistique	Logistique(0, 1)	$(-\infty, \infty)$	0.4340	3.2899	0.888
Student-t(3)	Student-t(3)	$(-\infty, \infty)$	0.4802	3	0.674



Théorie: Par exemple, si le support de f est un intervalle $[a, b]$, si f'' est continue sur $[a, b]$, et si on veut minimiser $E \int_{-\infty}^{\infty} |f_n(x) - f(x)| dx$, alors le choix optimal asymptotique, pour $n \rightarrow \infty$, est

$$k(x) = \frac{3}{4}(1 - x^2) \quad \text{pour } -1 \leq x \leq 1,$$
$$h = \left(\frac{15}{2\pi n} \right)^{1/5} \left(\frac{\int_a^b (f(x))^{1/2} dx}{\int_a^b |f''(x)| dx} \right)^{2/5}.$$

Théorie: Par exemple, si le support de f est un intervalle $[a, b]$, si f'' est continue sur $[a, b]$, et si on veut minimiser $E \int_{-\infty}^{\infty} |f_n(x) - f(x)| dx$, alors le choix optimal asymptotique, pour $n \rightarrow \infty$, est

$$k(x) = \frac{3}{4}(1 - x^2) \quad \text{pour } -1 \leq x \leq 1,$$

$$h = \left(\frac{15}{2\pi n} \right)^{1/5} \left(\frac{\int_a^b (f(x))^{1/2} dx}{\int_a^b |f''(x)| dx} \right)^{2/5}.$$

Si on veut plutôt minimiser $E \int_{-\infty}^{\infty} [f_n(x) - f(x)]^2 dx$, avec le même noyau $k(x)$, le h optimal est

$$h = \left(\frac{15}{n \int_{-\infty}^{\infty} (f''(x))^2 dx} \right)^{1/5}.$$

Théorie: Par exemple, si le support de f est un intervalle $[a, b]$, si f'' est continue sur $[a, b]$, et si on veut minimiser $E \int_{-\infty}^{\infty} |f_n(x) - f(x)| dx$, alors le choix optimal asymptotique, pour $n \rightarrow \infty$, est

$$k(x) = \frac{3}{4}(1 - x^2) \quad \text{pour } -1 \leq x \leq 1,$$

$$h = \left(\frac{15}{2\pi n} \right)^{1/5} \left(\frac{\int_a^b (f(x))^{1/2} dx}{\int_a^b |f''(x)| dx} \right)^{2/5}.$$

Si on veut plutôt minimiser $E \int_{-\infty}^{\infty} [f_n(x) - f(x)]^2 dx$, avec le même noyau $k(x)$, le h optimal est

$$h = \left(\frac{15}{n \int_{-\infty}^{\infty} (f''(x))^2 dx} \right)^{1/5}.$$

En principe, on peut estimer le h optimal à partir des données.
Mais il y a souvent beaucoup de bruit.
Puis ce h n'est optimal qu'asymptotiquement.

En principe, on peut estimer le h optimal à partir des données.

Mais il y a souvent beaucoup de bruit.

Puis ce h n'est optimal qu'asymptotiquement.

Lorsque le support n'est pas borné, on prend souvent un noyau gaussien.

En principe, on peut estimer le h optimal à partir des données.

Mais il y a souvent beaucoup de bruit.

Puis ce h n'est optimal qu'asymptotiquement.

Lorsque le support n'est pas borné, on prend souvent un noyau gaussien.

Méthode **plug-in**: (1) d'abord une première estimation de f ; (2) calculer le h "optimal" pour ce f ; (3) utiliser ce h pour estimer f (deuxième passe).

En principe, on peut estimer le h optimal à partir des données.

Mais il y a souvent beaucoup de bruit.

Puis ce h n'est optimal qu'asymptotiquement.

Lorsque le support n'est pas borné, on prend souvent un noyau gaussien.

Méthode **plug-in**: (1) d'abord une première estimation de f ; (2) calculer le h "optimal" pour ce f ; (3) utiliser ce h pour estimer f (deuxième passe).

Méthode à **double noyau**: Utiliser 2 noyaux distincts k_1 et k_2 , calculer les estimateurs de f correspondants, $f_{n,h}$ et $g_{n,h}$, puis choisir h qui minimise

$$\int_{-\infty}^{\infty} |f_{n,h}(x) - g_{n,h}(x)| dx.$$

En principe, on peut estimer le h optimal à partir des données.

Mais il y a souvent beaucoup de bruit.

Puis ce h n'est optimal qu'asymptotiquement.

Lorsque le support n'est pas borné, on prend souvent un noyau gaussien.

Méthode **plug-in**: (1) d'abord une première estimation de f ; (2) calculer le h "optimal" pour ce f ; (3) utiliser ce h pour estimer f (deuxième passe).

Méthode à **double noyau**: Utiliser 2 noyaux distincts k_1 et k_2 , calculer les estimateurs de f correspondants, $f_{n,h}$ et $g_{n,h}$, puis choisir h qui minimise

$$\int_{-\infty}^{\infty} |f_{n,h}(x) - g_{n,h}(x)| dx.$$

Une combinaison des deux donne l'une des méthodes les plus robustes en pratique.

Formule simplifiée (heuristique) de Silverman (1986):

$$h = \alpha_k h_0 \quad \text{où}$$

$$h_0 = 1.36374 \min(s_n, q/1.34) n^{-1/5}$$

$$q = \text{interquartile échantillonnal,}$$

$$s_n = \text{écart-type échantillonnal,}$$

$$\alpha_k = \left(\sigma_k^{-4} \int_{-\infty}^{\infty} k(x) dx \right)^{1/5}$$

$$\sigma_k = \text{écart-type du noyau } k.$$

Formule simplifiée (heuristique) de Silverman (1986):

$$\begin{aligned}
 h &= \alpha_k h_0 \quad \text{où} \\
 h_0 &= 1.36374 \min(s_n, q/1.34) n^{-1/5} \\
 q &= \text{interquartile échantillonnal,} \\
 s_n &= \text{écart-type échantillonnal,} \\
 \alpha_k &= \left(\sigma_k^{-4} \int_{-\infty}^{\infty} k(x) dx \right)^{1/5} \\
 \sigma_k &= \text{écart-type du noyau } k.
 \end{aligned}$$

La variance de la densité f_n est toujours plus grande que σ_n^2 .

Idée: on peut rapprocher les observations de leur moyenne \bar{x}_n afin de rendre la variance de f_n égale à σ_n^2 .

Il suffit de diviser la distance entre chaque x_i et \bar{x}_n par $\sigma_e^2 = 1 + (h\sigma_k/s_n)^2$.

(Exercice du devoir 3.)

Lois multivariées

Souvent des dépendances entre les v.a. d'un modèle.

Lois multivariées

Souvent des dépendances entre les v.a. d'un modèle.

A. Vecteurs aléatoires;

B. Séries chronologiques et autres processus stochastiques.

Lois multivariées

Souvent des dépendances entre les v.a. d'un modèle.

A. Vecteurs aléatoires;

B. Séries chronologiques et autres processus stochastiques.

Un vecteur $\mathbf{X} = (X_1, \dots, X_d)'$ a une loi multivariée de fonction de répartition F si pour tout $\mathbf{x} = (x_1, \dots, x_d)' \in \mathbb{R}^d$,

$$F(\mathbf{x}) = P[\mathbf{X} \leq \mathbf{x}] = P[X_1 \leq x_1, \dots, X_d \leq x_d].$$

Lois multivariées

Souvent des dépendances entre les v.a. d'un modèle.

A. Vecteurs aléatoires;

B. Séries chronologiques et autres processus stochastiques.

Un vecteur $\mathbf{X} = (X_1, \dots, X_d)'$ a une loi multivariée de fonction de répartition F si pour tout $\mathbf{x} = (x_1, \dots, x_d)' \in \mathbb{R}^d$,

$$F(\mathbf{x}) = P[\mathbf{X} \leq \mathbf{x}] = P[X_1 \leq x_1, \dots, X_d \leq x_d].$$

Les lois marginales: $F_j(x) = P[X_j \leq x]$.

Lois multivariées

Souvent des dépendances entre les v.a. d'un modèle.

A. Vecteurs aléatoires;

B. Séries chronologiques et autres processus stochastiques.

Un vecteur $\mathbf{X} = (X_1, \dots, X_d)'$ a une loi multivariée de fonction de répartition F si pour tout $\mathbf{x} = (x_1, \dots, x_d)' \in \mathbb{R}^d$,

$$F(\mathbf{x}) = P[\mathbf{X} \leq \mathbf{x}] = P[X_1 \leq x_1, \dots, X_d \leq x_d].$$

Les lois marginales: $F_j(x) = P[X_j \leq x]$.

X_1, \dots, X_d sont indépendantes ssi $F(x_1, \dots, x_d) = F_1(x_1) \cdots F_d(x_d)$.

Lois multivariées

Souvent des dépendances entre les v.a. d'un modèle.

A. Vecteurs aléatoires;

B. Séries chronologiques et autres processus stochastiques.

Un vecteur $\mathbf{X} = (X_1, \dots, X_d)'$ a une loi multivariée de fonction de répartition F si pour tout $\mathbf{x} = (x_1, \dots, x_d)' \in \mathbb{R}^d$,

$$F(\mathbf{x}) = P[\mathbf{X} \leq \mathbf{x}] = P[X_1 \leq x_1, \dots, X_d \leq x_d].$$

Les lois marginales: $F_j(x) = P[X_j \leq x]$.

X_1, \dots, X_d sont indépendantes ssi $F(x_1, \dots, x_d) = F_1(x_1) \cdots F_d(x_d)$.

Mesures de dépendance entre X et Y :

covariance: $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$,

coefficient de corrélation linéaire de Pearson:

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{(\text{Var}(X)\text{Var}(Y))^{1/2}} = \frac{E[XY] - E[X]E[Y]}{\sigma(X)\sigma(Y)}.$$

On a $-1 \leq \rho(X, Y) \leq 1$.

On a $-1 \leq \rho(X, Y) \leq 1$.

Matrice de covariance: $\Sigma = \text{Cov}[\mathbf{X}]$, éléments $\sigma_{ij} = \text{Cov}[X_i, X_j]$;

Matrice de corrélation \mathbf{R} , avec éléments $\rho(X_i, X_j)$.

On a $-1 \leq \rho(X, Y) \leq 1$.

Matrice de covariance: $\Sigma = \text{Cov}[\mathbf{X}]$, éléments $\sigma_{ij} = \text{Cov}[X_i, X_j]$;

Matrice de corrélation \mathbf{R} , avec éléments $\rho(X_i, X_j)$.

Contrainte de validité: Σ et \mathbf{R} doivent être définies semi-positives, i.e., $\mathbf{x}'\Sigma\mathbf{x} \geq 0$ pour tout $\mathbf{x} \in \mathbb{R}^d$,

On a $-1 \leq \rho(X, Y) \leq 1$.

Matrice de covariance: $\Sigma = \text{Cov}[\mathbf{X}]$, éléments $\sigma_{ij} = \text{Cov}[X_i, X_j]$;

Matrice de corrélation \mathbf{R} , avec éléments $\rho(X_i, X_j)$.

Contrainte de validité: Σ et \mathbf{R} doivent être définies semi-positives,

i.e., $\mathbf{x}'\Sigma\mathbf{x} \geq 0$ pour tout $\mathbf{x} \in \mathbb{R}^d$,

et la diagonale de \mathbf{R} ne peut contenir que des 1.

On a $-1 \leq \rho(X, Y) \leq 1$.

Matrice de covariance: $\Sigma = \text{Cov}[\mathbf{X}]$, éléments $\sigma_{ij} = \text{Cov}[X_i, X_j]$;

Matrice de corrélation \mathbf{R} , avec éléments $\rho(X_i, X_j)$.

Contrainte de validité: Σ et \mathbf{R} doivent être définies semi-positives,

i.e., $\mathbf{x}'\Sigma\mathbf{x} \geq 0$ pour tout $\mathbf{x} \in \mathbb{R}^d$,

et la diagonale de \mathbf{R} ne peut contenir que des 1.

Pour certains choix de marginales, il peut y avoir d'autres contraintes!

On a $-1 \leq \rho(X, Y) \leq 1$.

Matrice de covariance: $\Sigma = \text{Cov}[\mathbf{X}]$, éléments $\sigma_{ij} = \text{Cov}[X_i, X_j]$;

Matrice de corrélation \mathbf{R} , avec éléments $\rho(X_i, X_j)$.

Contrainte de validité: Σ et \mathbf{R} doivent être définies semi-positives,

i.e., $\mathbf{x}'\Sigma\mathbf{x} \geq 0$ pour tout $\mathbf{x} \in \mathbb{R}^d$,

et la diagonale de \mathbf{R} ne peut contenir que des 1.

Pour certains choix de marginales, il peut y avoir d'autres contraintes!

Plusieurs façons de généraliser certaines lois univariées.

Les F_j peuvent aussi provenir de différentes familles.

Loi normale multivariée (ou multinormale)

$\mathbf{X} = (X_1, \dots, X_d)' \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ a la densité

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma})}} \exp\left(-(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) / 2\right)$$

pour $\mathbf{x} = (x_1, \dots, x_d)' \in \mathbb{R}^d$, où $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)' \in \mathbb{R}^d$ (moyenne) et $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X})$ (matrice de covariance).

Loi normale multivariée (ou multinormale)

$\mathbf{X} = (X_1, \dots, X_d)' \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ a la densité

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma})}} \exp\left(-(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) / 2\right)$$

pour $\mathbf{x} = (x_1, \dots, x_d)' \in \mathbb{R}^d$, où $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)' \in \mathbb{R}^d$ (moyenne) et $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X})$ (matrice de covariance).

On a $E[X_j] = \mu_j$ et $\text{Cov}(X_j, X_k) = \sigma_{jk}$.

Loi normale multivariée (ou multinormale)

$\mathbf{X} = (X_1, \dots, X_d)' \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ a la densité

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma})}} \exp\left(-(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) / 2\right)$$

pour $\mathbf{x} = (x_1, \dots, x_d)' \in \mathbb{R}^d$, où $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)' \in \mathbb{R}^d$ (moyenne) et $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X})$ (matrice de covariance).

On a $E[X_j] = \mu_j$ et $\text{Cov}(X_j, X_k) = \sigma_{jk}$.

Loi normale multivariée standard: $\mathbf{X} \sim N(\mathbf{0}, \mathbf{I})$.

Loi normale multivariée (ou multinormale)

$\mathbf{X} = (X_1, \dots, X_d)' \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ a la densité

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma})}} \exp\left(-(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) / 2\right)$$

pour $\mathbf{x} = (x_1, \dots, x_d)' \in \mathbb{R}^d$, où $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)' \in \mathbb{R}^d$ (moyenne) et $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X})$ (matrice de covariance).

On a $E[X_j] = \mu_j$ et $\text{Cov}(X_j, X_k) = \sigma_{jk}$.

Loi normale multivariée standard: $\mathbf{X} \sim N(\mathbf{0}, \mathbf{I})$.

Si $\mathbf{X} \sim N(\mathbf{0}, \mathbf{I})$ et $\boldsymbol{\Sigma} = \mathbf{A}'\mathbf{A}$, alors $\mathbf{Y} = \mathbf{A}\mathbf{X} + \boldsymbol{\mu} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Facile de générer \mathbf{Y} !

Loi normale multivariée (ou multinormale)

$\mathbf{X} = (X_1, \dots, X_d)' \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ a la densité

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma})}} \exp\left(-(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) / 2\right)$$

pour $\mathbf{x} = (x_1, \dots, x_d)' \in \mathbb{R}^d$, où $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)' \in \mathbb{R}^d$ (moyenne) et $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X})$ (matrice de covariance).

On a $E[X_j] = \mu_j$ et $\text{Cov}(X_j, X_k) = \sigma_{jk}$.

Loi normale multivariée standard: $\mathbf{X} \sim N(\mathbf{0}, \mathbf{I})$.

Si $\mathbf{X} \sim N(\mathbf{0}, \mathbf{I})$ et $\boldsymbol{\Sigma} = \mathbf{A}'\mathbf{A}$, alors $\mathbf{Y} = \mathbf{A}\mathbf{X} + \boldsymbol{\mu} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Facile de générer \mathbf{Y} !

Ici, la moyenne et la covariance spécifient uniquement la distribution.

Mais ce n'est pas le cas en général pour les autres distributions.

Loi normale multivariée (ou multinormale)

$\mathbf{X} = (X_1, \dots, X_d)' \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ a la densité

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \det(\boldsymbol{\Sigma})}} \exp\left(-(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) / 2\right)$$

pour $\mathbf{x} = (x_1, \dots, x_d)' \in \mathbb{R}^d$, où $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)' \in \mathbb{R}^d$ (moyenne) et $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X})$ (matrice de covariance).

On a $E[X_j] = \mu_j$ et $\text{Cov}(X_j, X_k) = \sigma_{jk}$.

Loi normale multivariée standard: $\mathbf{X} \sim N(\mathbf{0}, \mathbf{I})$.

Si $\mathbf{X} \sim N(\mathbf{0}, \mathbf{I})$ et $\boldsymbol{\Sigma} = \mathbf{A}'\mathbf{A}$, alors $\mathbf{Y} = \mathbf{A}\mathbf{X} + \boldsymbol{\mu} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Facile de générer \mathbf{Y} !

Ici, la moyenne et la covariance spécifient uniquement la distribution.

Mais ce n'est pas le cas en général pour les autres distributions.

$\mathbf{X} = (X_1, \dots, X_d)'$ suit une loi lognormale multivariée si $\mathbf{Y} = (\ln X_1, \dots, \ln X_d)'$ est multinormal.

Bornes de Fréchet

Théorème (e.g., Lehmann 1966): Parmi les paires de v.a. (X, Y) dont les f.r. marginales sont F et G , la paire $(X, Y) = (F^{-1}(U), G^{-1}(U))$ où $U \sim U(0, 1)$, maximise $\rho[X, Y]$,

Bornes de Fréchet

Théorème (e.g., Lehmann 1966): Parmi les paires de v.a. (X, Y) dont les f.r. marginales sont F et G , la paire $(X, Y) = (F^{-1}(U), G^{-1}(U))$ où $U \sim U(0, 1)$, maximise $\rho[X, Y]$, et la paire $(X, Y) = (F^{-1}(U), G^{-1}(1 - U))$ minimise $\rho[X, Y]$.

Bornes de Fréchet

Théorème (e.g., Lehmann 1966): Parmi les paires de v.a. (X, Y) dont les f.r. marginales sont F et G , la paire $(X, Y) = (F^{-1}(U), G^{-1}(U))$ où $U \sim U(0, 1)$, maximise $\rho[X, Y]$, et la paire $(X, Y) = (F^{-1}(U), G^{-1}(1 - U))$ minimise $\rho[X, Y]$.

La loi conjointe de corrélation maximale satisfait:

$$\begin{aligned}
 P\{X \leq x, Y \leq y\} &= P[F^{-1}(U) \leq x, G^{-1}(U) \leq y] \\
 &= P[U \leq F(x), U \leq G(y)] \\
 &= P[U \leq \min(F(x), G(y))] = \min(F(x), G(y)).
 \end{aligned}$$

Bornes de Fréchet

Théorème (e.g., Lehmann 1966): Parmi les paires de v.a. (X, Y) dont les f.r. marginales sont F et G , la paire $(X, Y) = (F^{-1}(U), G^{-1}(U))$ où $U \sim U(0, 1)$, maximise $\rho[X, Y]$, et la paire $(X, Y) = (F^{-1}(U), G^{-1}(1 - U))$ minimise $\rho[X, Y]$.

La loi conjointe de corrélation maximale satisfait:

$$\begin{aligned} P\{X \leq x, Y \leq y\} &= P[F^{-1}(U) \leq x, G^{-1}(U) \leq y] \\ &= P[U \leq F(x), U \leq G(y)] \\ &= P[U \leq \min(F(x), G(y))] = \min(F(x), G(y)). \end{aligned}$$

Celle de corrélation minimale satisfait:

$$\begin{aligned} P\{X \leq x, Y \leq y\} &= P[F^{-1}(U) \leq x, G^{-1}(1 - U) \leq y] \\ &= P[U \leq F(x), 1 - U \leq G(y)] \\ &= P[1 - G(y) \leq U \leq F(x)] = \max(0, F(x) + G(y) - 1). \end{aligned}$$

Exemple. Soient $X \sim \text{Normale}(\mu_1, \sigma_1^2)$ et $Y \sim \text{Normale}(\mu_2, \sigma_2^2)$.
Soit $U \sim U(0, 1)$ et $Z = \Phi^{-1}(U)$.

Exemple. Soient $X \sim \text{Normale}(\mu_1, \sigma_1^2)$ et $Y \sim \text{Normale}(\mu_2, \sigma_2^2)$.
Soit $U \sim U(0, 1)$ et $Z = \Phi^{-1}(U)$.

Corrélation maximale: $X = \mu_1 + \sigma_1 Z$ et $Y = \mu_2 + \sigma_2 Z$.

Exemple. Soient $X \sim \text{Normale}(\mu_1, \sigma_1^2)$ et $Y \sim \text{Normale}(\mu_2, \sigma_2^2)$.
Soit $U \sim U(0, 1)$ et $Z = \Phi^{-1}(U)$.

Corrélation maximale: $X = \mu_1 + \sigma_1 Z$ et $Y = \mu_2 + \sigma_2 Z$. On a

$$\rho(X, Y) = \frac{E[(X - \mu_1)(Y - \mu_2)]}{\sigma_1 \sigma_2} = \frac{E[\sigma_1 Z \sigma_2 Z]}{\sigma_1 \sigma_2} = E[Z^2] = 1.$$

Exemple. Soient $X \sim \text{Normale}(\mu_1, \sigma_1^2)$ et $Y \sim \text{Normale}(\mu_2, \sigma_2^2)$.
Soit $U \sim U(0, 1)$ et $Z = \Phi^{-1}(U)$.

Corrélation maximale: $X = \mu_1 + \sigma_1 Z$ et $Y = \mu_2 + \sigma_2 Z$. On a

$$\rho(X, Y) = \frac{E[(X - \mu_1)(Y - \mu_2)]}{\sigma_1 \sigma_2} = \frac{E[\sigma_1 Z \sigma_2 Z]}{\sigma_1 \sigma_2} = E[Z^2] = 1.$$

Corrélation minimale: Puisque $\Phi^{-1}(1 - U) = -Z$,
c'est $X = \mu_1 + \sigma_1 Z$ et $Y = \mu_2 - \sigma_2 Z$.

Exemple. Soient $X \sim \text{Normale}(\mu_1, \sigma_1^2)$ et $Y \sim \text{Normale}(\mu_2, \sigma_2^2)$.
Soit $U \sim U(0, 1)$ et $Z = \Phi^{-1}(U)$.

Corrélation maximale: $X = \mu_1 + \sigma_1 Z$ et $Y = \mu_2 + \sigma_2 Z$. On a

$$\rho(X, Y) = \frac{E[(X - \mu_1)(Y - \mu_2)]}{\sigma_1 \sigma_2} = \frac{E[\sigma_1 Z \sigma_2 Z]}{\sigma_1 \sigma_2} = E[Z^2] = 1.$$

Corrélation minimale: Puisque $\Phi^{-1}(1 - U) = -Z$,
c'est $X = \mu_1 + \sigma_1 Z$ et $Y = \mu_2 - \sigma_2 Z$.

$$\rho(X, Y) = \frac{E[(X - \mu_1)(Y - \mu_2)]}{\sigma_1 \sigma_2} = \frac{E[\sigma_1 Z (-\sigma_2) Z]}{\sigma_1 \sigma_2} = -E[Z^2] = -1.$$

Exemple. Soient $X \sim \text{Normale}(\mu_1, \sigma_1^2)$ et $Y \sim \text{Normale}(\mu_2, \sigma_2^2)$.
Soit $U \sim U(0, 1)$ et $Z = \Phi^{-1}(U)$.

Corrélation maximale: $X = \mu_1 + \sigma_1 Z$ et $Y = \mu_2 + \sigma_2 Z$. On a

$$\rho(X, Y) = \frac{E[(X - \mu_1)(Y - \mu_2)]}{\sigma_1 \sigma_2} = \frac{E[\sigma_1 Z \sigma_2 Z]}{\sigma_1 \sigma_2} = E[Z^2] = 1.$$

Corrélation minimale: Puisque $\Phi^{-1}(1 - U) = -Z$,
c'est $X = \mu_1 + \sigma_1 Z$ et $Y = \mu_2 - \sigma_2 Z$.

$$\rho(X, Y) = \frac{E[(X - \mu_1)(Y - \mu_2)]}{\sigma_1 \sigma_2} = \frac{E[\sigma_1 Z (-\sigma_2) Z]}{\sigma_1 \sigma_2} = -E[Z^2] = -1.$$

Dans ce cas, $\rho(X, Y)$ peut prendre n'importe quelle valeur dans $[-1, 1]$.

Exemple. Soient $X \sim \text{Normale}(\mu_1, \sigma_1^2)$ et $Y \sim \text{Normale}(\mu_2, \sigma_2^2)$.
Soit $U \sim U(0, 1)$ et $Z = \Phi^{-1}(U)$.

Corrélation maximale: $X = \mu_1 + \sigma_1 Z$ et $Y = \mu_2 + \sigma_2 Z$. On a

$$\rho(X, Y) = \frac{E[(X - \mu_1)(Y - \mu_2)]}{\sigma_1 \sigma_2} = \frac{E[\sigma_1 Z \sigma_2 Z]}{\sigma_1 \sigma_2} = E[Z^2] = 1.$$

Corrélation minimale: Puisque $\Phi^{-1}(1 - U) = -Z$,
c'est $X = \mu_1 + \sigma_1 Z$ et $Y = \mu_2 - \sigma_2 Z$.

$$\rho(X, Y) = \frac{E[(X - \mu_1)(Y - \mu_2)]}{\sigma_1 \sigma_2} = \frac{E[\sigma_1 Z (-\sigma_2) Z]}{\sigma_1 \sigma_2} = -E[Z^2] = -1.$$

Dans ce cas, $\rho(X, Y)$ peut prendre n'importe quelle valeur dans $[-1, 1]$.

Ceci est vrai en général si X et Y ne diffèrent que par des paramètres de localisation et d'échelle.

Exemple. Soient $X \sim \text{Normale}(\mu_1, \sigma_1^2)$ et $Y \sim \text{Normale}(\mu_2, \sigma_2^2)$.
Soit $U \sim U(0, 1)$ et $Z = \Phi^{-1}(U)$.

Corrélation maximale: $X = \mu_1 + \sigma_1 Z$ et $Y = \mu_2 + \sigma_2 Z$. On a

$$\rho(X, Y) = \frac{E[(X - \mu_1)(Y - \mu_2)]}{\sigma_1 \sigma_2} = \frac{E[\sigma_1 Z \sigma_2 Z]}{\sigma_1 \sigma_2} = E[Z^2] = 1.$$

Corrélation minimale: Puisque $\Phi^{-1}(1 - U) = -Z$,
c'est $X = \mu_1 + \sigma_1 Z$ et $Y = \mu_2 - \sigma_2 Z$.

$$\rho(X, Y) = \frac{E[(X - \mu_1)(Y - \mu_2)]}{\sigma_1 \sigma_2} = \frac{E[\sigma_1 Z (-\sigma_2) Z]}{\sigma_1 \sigma_2} = -E[Z^2] = -1.$$

Dans ce cas, $\rho(X, Y)$ peut prendre n'importe quelle valeur dans $[-1, 1]$.

Ceci est vrai en général si X et Y ne diffèrent que par des paramètres de localisation et d'échelle.

Exemple. Si $X \sim \text{Lognormale}(0, 1)$ et $Y \sim \text{Lognormale}(0, 2)$, alors
 $-0.090 \leq \rho(X, Y) \leq 0.666$ (exercice).

Autres mesures de dépendance

Le coefficient de corrélation $\rho(X, Y)$ dépend non seulement du niveau de dépendance entre (X, Y) , mais aussi de leurs lois marginales.

Autres mesures de dépendance

Le coefficient de corrélation $\rho(X, Y)$ dépend non seulement du niveau de dépendance entre (X, Y) , mais aussi de leurs lois marginales.

Dans le cas continu, le coefficient de **corrélation des rangs** ou de **Spearman**,

$$\rho_s(X, Y) = \rho(F(X), G(Y)) = 12 E[F(X)G(Y)] - 3,$$

mesure la dépendance indépendamment des marginales.

Autres mesures de dépendance

Le coefficient de corrélation $\rho(X, Y)$ dépend non seulement du niveau de dépendance entre (X, Y) , mais aussi de leurs lois marginales.

Dans le cas continu, le coefficient de **corrélation des rangs** ou de **Spearman**,

$$\rho_s(X, Y) = \rho(F(X), G(Y)) = 12 E[F(X)G(Y)] - 3,$$

mesure la dépendance indépendamment des marginales.

Si $X = F^{-1}(U_1)$ et $Y = G^{-1}(U_2)$ pour des uniformes U_1 et U_2 , alors

$$\rho_s(X, Y) = \rho(U_1, U_2)$$

Autres mesures de dépendance

Le coefficient de corrélation $\rho(X, Y)$ dépend non seulement du niveau de dépendance entre (X, Y) , mais aussi de leurs lois marginales.

Dans le cas continu, le coefficient de corrélation des rangs ou de Spearman,

$$\rho_s(X, Y) = \rho(F(X), G(Y)) = 12 E[F(X)G(Y)] - 3,$$

mesure la dépendance indépendamment des marginales.

Si $X = F^{-1}(U_1)$ et $Y = G^{-1}(U_2)$ pour des uniformes U_1 et U_2 , alors

$$\rho_s(X, Y) = \rho(U_1, U_2)$$

Il peut aussi prendre toutes les valeurs dans $[-1, 1]$, peu importe F et G , et existe toujours.

Dépendance vs corrélation.

Les coefficients ρ et ρ_s ne mesurent que la dépendance linéaire.

Dépendance vs corrélation.

Les coefficients ρ et ρ_s ne mesurent que la dépendance **linéaire**.

Par exemple, si $X \sim N(0, 1)$ et $Y = X^2$, il est clair que Y dépend de X , mais

$$\rho(X, Y) = \frac{E[XY] - E[X]E[Y]}{\sigma(X)\sigma(Y)} = E[X^3] - E[X]E[X^2] = 0.$$

Dépendance vs corrélation.

Les coefficients ρ et ρ_s ne mesurent que la dépendance **linéaire**.

Par exemple, si $X \sim N(0, 1)$ et $Y = X^2$, il est clair que Y dépend de X , mais

$$\rho(X, Y) = \frac{E[XY] - E[X]E[Y]}{\sigma(X)\sigma(Y)} = E[X^3] - E[X]E[X^2] = 0.$$

L'indépendance implique l'absence de corrélation, mais pas l'inverse.

Fonctions de dépendance (copules)

Soit $\mathbf{X} = (X_1, \dots, X_d)$ v.a. continue, où $P[X_j \leq x] = F_j(x)$.

Alors les lois marginales de $\mathbf{U} = (U_1, \dots, U_d) = (F_1(X_1), \dots, F_d(X_d))$ sont uniformes.

Fonctions de dépendance (copules)

Soit $\mathbf{X} = (X_1, \dots, X_d)$ v.a. continue, où $P[X_j \leq x] = F_j(x)$.

Alors les lois marginales de $\mathbf{U} = (U_1, \dots, U_d) = (F_1(X_1), \dots, F_d(X_d))$ sont uniformes.

En spécifiant une forme de dépendance (ou loi conjointe) pour \mathbf{U} , on en spécifie indirectement pour \mathbf{X} . Une copule, ou fonction de dépendance pour \mathbf{U} , est une fonction de répartition conjointe C pour \mathbf{U} .

Fonctions de dépendance (copules)

Soit $\mathbf{X} = (X_1, \dots, X_d)$ v.a. continue, où $P[X_j \leq x] = F_j(x)$.

Alors les lois marginales de $\mathbf{U} = (U_1, \dots, U_d) = (F_1(X_1), \dots, F_d(X_d))$ sont uniformes.

En spécifiant une forme de dépendance (ou loi conjointe) pour \mathbf{U} , on en spécifie indirectement pour \mathbf{X} . Une copule, ou fonction de dépendance pour \mathbf{U} , est une fonction de répartition conjointe C pour \mathbf{U} .

Théorème de Sklar: La loi de \mathbf{X} est spécifiée uniquement par C et les F_j . On a

$$F(x_1, \dots, x_d) = C(u_1, \dots, u_d) = C(F_1(x_1), \dots, F_d(x_d))$$

où $u_j = F_j(x_j)$.

Fonctions de dépendance (copules)

Soit $\mathbf{X} = (X_1, \dots, X_d)$ v.a. continue, où $P[X_j \leq x] = F_j(x)$.

Alors les lois marginales de $\mathbf{U} = (U_1, \dots, U_d) = (F_1(X_1), \dots, F_d(X_d))$ sont uniformes.

En spécifiant une forme de dépendance (ou loi conjointe) pour \mathbf{U} , on en spécifie une indirectement pour \mathbf{X} . Une copule, ou fonction de dépendance pour \mathbf{U} , est une fonction de répartition conjointe C pour \mathbf{U} .

Théorème de Sklar: La loi de \mathbf{X} est spécifiée uniquement par C et les F_j . On a

$$F(x_1, \dots, x_d) = C(u_1, \dots, u_d) = C(F_1(x_1), \dots, F_d(x_d))$$

où $u_j = F_j(x_j)$. X_1, \dots, X_d sont indép. ssi $C(u_1, \dots, u_d) = u_1 \cdots u_d$.

Fonctions de dépendance (copules)

Soit $\mathbf{X} = (X_1, \dots, X_d)$ v.a. continue, où $P[X_j \leq x] = F_j(x)$.

Alors les lois marginales de $\mathbf{U} = (U_1, \dots, U_d) = (F_1(X_1), \dots, F_d(X_d))$ sont uniformes.

En spécifiant une forme de dépendance (ou loi conjointe) pour \mathbf{U} , on en spécifie une indirectement pour \mathbf{X} . Une copule, ou fonction de dépendance pour \mathbf{U} , est une fonction de répartition conjointe C pour \mathbf{U} .

Théorème de Sklar: La loi de \mathbf{X} est spécifiée uniquement par C et les F_j . On a

$$F(x_1, \dots, x_d) = C(u_1, \dots, u_d) = C(F_1(x_1), \dots, F_d(x_d))$$

où $u_j = F_j(x_j)$. X_1, \dots, X_d sont indép. ssi $C(u_1, \dots, u_d) = u_1 \cdots u_d$.

Génération: (U_1, \dots, U_d) selon C , puis $X_j = F_j^{-1}(U_j)$.

Fonctions de dépendance (copules)

Soit $\mathbf{X} = (X_1, \dots, X_d)$ v.a. continue, où $P[X_j \leq x] = F_j(x)$.

Alors les lois marginales de $\mathbf{U} = (U_1, \dots, U_d) = (F_1(X_1), \dots, F_d(X_d))$ sont uniformes.

En spécifiant une forme de dépendance (ou loi conjointe) pour \mathbf{U} , on en spécifie une indirectement pour \mathbf{X} . Une copule, ou fonction de dépendance pour \mathbf{U} , est une fonction de répartition conjointe C pour \mathbf{U} .

Théorème de Sklar: La loi de \mathbf{X} est spécifiée uniquement par C et les F_j . On a

$$F(x_1, \dots, x_d) = C(u_1, \dots, u_d) = C(F_1(x_1), \dots, F_d(x_d))$$

où $u_j = F_j(x_j)$. X_1, \dots, X_d sont indép. ssi $C(u_1, \dots, u_d) = u_1 \cdots u_d$.

Génération: (U_1, \dots, U_d) selon C , puis $X_j = F_j^{-1}(U_j)$.

Mais comment choisir C ? Trop de possibilités!

On veut quelque chose de pratique à utiliser.

Fonctions de dépendance (copules)

Soit $\mathbf{X} = (X_1, \dots, X_d)$ v.a. continue, où $P[X_j \leq x] = F_j(x)$.

Alors les lois marginales de $\mathbf{U} = (U_1, \dots, U_d) = (F_1(X_1), \dots, F_d(X_d))$ sont uniformes.

En spécifiant une forme de dépendance (ou loi conjointe) pour \mathbf{U} , on en spécifie une indirectement pour \mathbf{X} . Une copule, ou fonction de dépendance pour \mathbf{U} , est une fonction de répartition conjointe C pour \mathbf{U} .

Théorème de Sklar: La loi de \mathbf{X} est spécifiée uniquement par C et les F_j . On a

$$F(x_1, \dots, x_d) = C(u_1, \dots, u_d) = C(F_1(x_1), \dots, F_d(x_d))$$

où $u_j = F_j(x_j)$. X_1, \dots, X_d sont indép. ssi $C(u_1, \dots, u_d) = u_1 \cdots u_d$.

Génération: (U_1, \dots, U_d) selon C , puis $X_j = F_j^{-1}(U_j)$.

Mais comment choisir C ? Trop de possibilités!

On veut quelque chose de pratique à utiliser.

Certaines méthodes **restreignent** volontairement le champs de possibilités pour C .

Méthode NORTA (Normal to Anything)

On se restreint aux copules de la forme

$$C(u_1, \dots, u_d) = P[\Phi(Y_1) \leq u_1, \dots, \Phi(Y_d) \leq u_d]$$

où $\mathbf{Y} = (Y_1, \dots, Y_d) \sim N(\mathbf{0}, \Sigma^{(\mathbf{y})})$ et $\Sigma^{(\mathbf{y})} = \mathbf{R}^{(\mathbf{y})} = \mathbf{L}^{(\mathbf{y})'} \mathbf{L}^{(\mathbf{y})}$.

Méthode NORTA (Normal to Anything)

On se restreint aux copules de la forme

$$C(u_1, \dots, u_d) = P[\Phi(Y_1) \leq u_1, \dots, \Phi(Y_d) \leq u_d]$$

où $\mathbf{Y} = (Y_1, \dots, Y_d) \sim N(\mathbf{0}, \Sigma^{(y)})$ et $\Sigma^{(y)} = \mathbf{R}^{(y)} = \mathbf{L}^{(y)'} \mathbf{L}^{(y)}$.

Pour générer \mathbf{X} , générer $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I})$, puis poser $\mathbf{Y} = \mathbf{L}^{(y)} \mathbf{Z}$ et

$$\mathbf{X} = (X_1, \dots, X_d) = (F_1^{-1}(\Phi(Y_1)), \dots, F_d^{-1}(\Phi(Y_d))).$$

Méthode NORTA (Normal to Anything)

On se restreint aux copules de la forme

$$C(u_1, \dots, u_d) = P[\Phi(Y_1) \leq u_1, \dots, \Phi(Y_d) \leq u_d]$$

où $\mathbf{Y} = (Y_1, \dots, Y_d) \sim N(\mathbf{0}, \Sigma^{(y)})$ et $\Sigma^{(y)} = \mathbf{R}^{(y)} = \mathbf{L}^{(y)'} \mathbf{L}^{(y)}$.

Pour générer \mathbf{X} , générer $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I})$, puis poser $\mathbf{Y} = \mathbf{L}^{(y)} \mathbf{Z}$ et

$$\mathbf{X} = (X_1, \dots, X_d) = (F_1^{-1}(\Phi(Y_1)), \dots, F_d^{-1}(\Phi(Y_d))).$$

Il suffit donc de spécifier les marginales et les corrélations.

Beaucoup moins de degrés de liberté que si on peut spécifier C arbitrairement.

Mais plus simple à gérer.

Méthode NORTA (Normal to Anything)

On se restreint aux copules de la forme

$$C(u_1, \dots, u_d) = P[\Phi(Y_1) \leq u_1, \dots, \Phi(Y_d) \leq u_d]$$

où $\mathbf{Y} = (Y_1, \dots, Y_d) \sim N(\mathbf{0}, \Sigma^{(y)})$ et $\Sigma^{(y)} = \mathbf{R}^{(y)} = \mathbf{L}^{(y)'} \mathbf{L}^{(y)}$.

Pour générer \mathbf{X} , générer $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I})$, puis poser $\mathbf{Y} = \mathbf{L}^{(y)} \mathbf{Z}$ et

$$\mathbf{X} = (X_1, \dots, X_d) = (F_1^{-1}(\Phi(Y_1)), \dots, F_d^{-1}(\Phi(Y_d))).$$

Il suffit donc de spécifier les marginales et les corrélations.

Beaucoup moins de degrés de liberté que si on peut spécifier C arbitrairement.

Mais plus simple à gérer.

On voudra choisir $\mathbf{R}^{(y)}$ (ou $\mathbf{L}^{(y)}$) de manière à obtenir une matrice de corrélation visée $\mathbf{R}^{(x)} = \mathbf{R}$ pour \mathbf{X}

Méthode NORTA (Normal to Anything)

On se restreint aux copules de la forme

$$C(u_1, \dots, u_d) = P[\Phi(Y_1) \leq u_1, \dots, \Phi(Y_d) \leq u_d]$$

où $\mathbf{Y} = (Y_1, \dots, Y_d) \sim N(\mathbf{0}, \Sigma^{(y)})$ et $\Sigma^{(y)} = \mathbf{R}^{(y)} = \mathbf{L}^{(y)'} \mathbf{L}^{(y)}$.

Pour générer \mathbf{X} , générer $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I})$, puis poser $\mathbf{Y} = \mathbf{L}^{(y)} \mathbf{Z}$ et

$$\mathbf{X} = (X_1, \dots, X_d) = (F_1^{-1}(\Phi(Y_1)), \dots, F_d^{-1}(\Phi(Y_d))).$$

Il suffit donc de spécifier les marginales et les corrélations.

Beaucoup moins de degrés de liberté que si on peut spécifier C arbitrairement.

Mais plus simple à gérer.

On voudra choisir $\mathbf{R}^{(y)}$ (ou $\mathbf{L}^{(y)}$) de manière à obtenir une matrice de corrélation visée $\mathbf{R}^{(x)} = \mathbf{R}$ pour \mathbf{X} , ou encore $\mathbf{R}^{(u)}$ pour \mathbf{U} .

Pour simplifier la notation, supposons que les X_j ont été standardisés:
 $E[X_j] = 0$ et $\text{Var}[X_j] = 1$.

Pour simplifier la notation, supposons que les X_j ont été standardisés:
 $E[X_j] = 0$ et $\text{Var}[X_j] = 1$.

Notons $r_{ij}^{(y)}$ et r_{ij} les éléments de $\mathbf{R}^{(y)}$ et \mathbf{R} .

Pour simplifier la notation, supposons que les X_j ont été standardisés:
 $E[X_j] = 0$ et $\text{Var}[X_j] = 1$.

Notons $r_{ij}^{(y)}$ et r_{ij} les éléments de $\mathbf{R}^{(y)}$ et \mathbf{R} .

En intégrant par rapport à la densité de (Y_i, Y_j) :

$$\begin{aligned} r_{ij} &= E[X_i X_j] = \varphi_{ij}(r_{ij}^{(y)}) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_i^{-1}(\Phi(y_i)) F_j^{-1}(\Phi(y_j)) \\ &\quad \frac{1}{2\pi(1 - r_{ij}^{(y)})} \exp \left[\frac{y_i^2 + y_j^2 - 2r_{ij}^{(y)} y_i y_j}{2(1 - r_{ij}^{(y)})} \right] dy_i dy_j. \end{aligned}$$

Il faut résoudre (trouver $r_{ij}^{(y)}$) pour chacune des $d(d - 1)/2$ paires (i, j) .

Pour simplifier la notation, supposons que les X_j ont été standardisés:
 $E[X_j] = 0$ et $\text{Var}[X_j] = 1$.

Notons $r_{ij}^{(y)}$ et r_{ij} les éléments de $\mathbf{R}^{(y)}$ et \mathbf{R} .

En intégrant par rapport à la densité de (Y_i, Y_j) :

$$\begin{aligned} r_{ij} &= E[X_i X_j] = \varphi_{ij}(r_{ij}^{(y)}) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_i^{-1}(\Phi(y_i)) F_j^{-1}(\Phi(y_j)) \\ &\quad \frac{1}{2\pi(1 - r_{ij}^{(y)})} \exp \left[\frac{y_i^2 + y_j^2 - 2r_{ij}^{(y)} y_i y_j}{2(1 - r_{ij}^{(y)})} \right] dy_i dy_j. \end{aligned}$$

Il faut résoudre (trouver $r_{ij}^{(y)}$) pour chacune des $d(d-1)/2$ paires (i, j) .

On peut montrer que $\varphi_{ij} : [-1, 1] \rightarrow [-1, 1]$ est non-décroissante et continue, avec $\varphi_{ij}(0) = 0$, et va de $\underline{r}_{ij} = \varphi_{ij}(-1)$ à $\bar{r}_{ij} = \varphi_{ij}(1)$.

Pour simplifier la notation, supposons que les X_j ont été standardisés:
 $E[X_j] = 0$ et $\text{Var}[X_j] = 1$.

Notons $r_{ij}^{(y)}$ et r_{ij} les éléments de $\mathbf{R}^{(y)}$ et \mathbf{R} .

En intégrant par rapport à la densité de (Y_i, Y_j) :

$$\begin{aligned} r_{ij} &= E[X_i X_j] = \varphi_{ij}(r_{ij}^{(y)}) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_i^{-1}(\Phi(y_i)) F_j^{-1}(\Phi(y_j)) \\ &\quad \frac{1}{2\pi(1 - r_{ij}^{(y)})} \exp \left[\frac{y_i^2 + y_j^2 - 2r_{ij}^{(y)} y_i y_j}{2(1 - r_{ij}^{(y)})} \right] dy_i dy_j. \end{aligned}$$

Il faut résoudre (trouver $r_{ij}^{(y)}$) pour chacune des $d(d-1)/2$ paires (i, j) .

On peut montrer que $\varphi_{ij} : [-1, 1] \rightarrow [-1, 1]$ est non-décroissante et continue, avec $\varphi_{ij}(0) = 0$, et va de $\underline{r}_{ij} = \varphi_{ij}(-1)$ à $\bar{r}_{ij} = \varphi_{ij}(1)$.

On résoud l'équation numériquement par une méthode de recherche de racine.

Si on spécifie plutôt $\mathbf{R} = \mathbf{R}^{(u)}$ (la matrice de corrélation des rangs), l'équation à résoudre se simplifie et on obtient une solution analytique:

$$r_{i,j}^{(y)} = 2 \sin(\pi r_{i,j}^{(u)} / 6)$$

(voir exercices).

Si on spécifie plutôt $\mathbf{R} = \mathbf{R}^{(u)}$ (la matrice de corrélation des rangs), l'équation à résoudre se simplifie et on obtient une solution analytique:

$$r_{i,j}^{(y)} = 2 \sin(\pi r_{i,j}^{(u)} / 6)$$

(voir exercices).

Pour cette raison, il est nettement préférable de spécifier $\mathbf{R}^{(u)}$ au lieu de $\mathbf{R}^{(x)}$ (en estimant les corrélations de Spearman à partir des données).

Pour $d > 2$, il se peut que $\mathbf{R}^{(y)}$ trouvée ainsi ne soit pas semi-définie positive!
Dans ce cas, cela veut dire que NORTA ne peut pas donner le \mathbf{R} désiré pour les marginales choisies.

Pour $d > 2$, il se peut que $\mathbf{R}^{(y)}$ trouvée ainsi ne soit pas semi-définie positive!
Dans ce cas, cela veut dire que NORTA ne peut pas donner le \mathbf{R} désiré pour les marginales choisies.
On dit que ce \mathbf{R} n'est pas NORTA-compatible.

Pour $d > 2$, il se peut que $\mathbf{R}^{(y)}$ trouvée ainsi ne soit pas semi-définie positive!
Dans ce cas, cela veut dire que NORTA ne peut pas donner le \mathbf{R} désiré pour les marginales choisies.

On dit que ce \mathbf{R} n'est pas NORTA-compatible.

Les chances que cela se produise augmentent rapidement avec d .

Pour $d > 2$, il se peut que $\mathbf{R}^{(y)}$ trouvée ainsi ne soit pas semi-définie positive!
Dans ce cas, cela veut dire que NORTA ne peut pas donner le \mathbf{R} désiré pour les marginales choisies.

On dit que ce \mathbf{R} n'est pas NORTA-compatible.

Les chances que cela se produise augmentent rapidement avec d .

Solution: choisir une matrice de corrélations $\mathbf{L}'\mathbf{L}$, semi-définie positive, "la plus proche possible" de $\mathbf{R}^{(y)}$.

Pour $d > 2$, il se peut que $\mathbf{R}^{(y)}$ trouvée ainsi ne soit pas semi-définie positive!
Dans ce cas, cela veut dire que NORTA ne peut pas donner le \mathbf{R} désiré pour les marginales choisies.

On dit que ce \mathbf{R} n'est pas NORTA-compatible.

Les chances que cela se produise augmentent rapidement avec d .

Solution: choisir une matrice de corrélations $\mathbf{L}'\mathbf{L}$, semi-définie positive, “la plus proche possible” de $\mathbf{R}^{(y)}$.

On choisit une norme $\|\cdot\|$ sur l'espace des matrices, puis on résoud le problème d'optimisation:

Minimiser $\|\mathbf{L}'\mathbf{L} - \mathbf{R}^{(y)}\|$

sous la contrainte: \mathbf{L} est triangulaire inférieure avec des 1 sur la diagonale.