

DÉPENDANCE STOCHASTIQUE (D, N)

(26 / 04 / 2020, © Monfort, Dicostat2005, 2005-2020)

Si l'indépendance entre **événements aléatoires** ou entre **variables aléatoires** est aisée à définir, et de façon « univoque », la notion de **dépendance stochastique** peut revêtir de multiples formes (cf aussi **dépendance**).

(i) Soit $(\Omega, \mathcal{F}, P)$ un **espace probabilisé**, $(\mathcal{X}, \mathcal{B})$ et $(\mathcal{Y}, \mathcal{C})$ deux **espaces mesurables** et $\zeta = (\xi, \eta) : \Omega \mapsto \mathcal{X} \times \mathcal{Y}$ un **couple aléatoire** dont la **loi conjointe** $P^\zeta = P^{(\xi, \eta)}$ est supposée dominée par une mesure produit $\mu \otimes \nu$ (cf **loi absolument continue, famille de lois dominée**). On note resp P^ξ et P^η les **lois marginales** (ie les lois propres) de ξ et de η .

On dit que ξ et η sont **stochastiquement dépendantes**, ou **stochastiquement liées**, ssi il existe (au moins) un couple de **boréliens** non négligeables $(B, C) \in \mathcal{B} \times \mathcal{C}$ tq :

$$(1) \quad P^{(\xi, \eta)}(B \cap C) \neq P^\xi(B) \cdot P^\eta(C).$$

Si chaque mesure μ (resp ν) est une **mesure abstraite** positive dominant P^ξ (resp P^η), la **dépendance stochastique** implique l'existence d'une **partie** non vide $D \subset \mathcal{X} \times \mathcal{Y}$ tq :

$$(2) \quad f(x, y) \neq f_\xi(x) \cdot f_\eta(y), \quad \forall (x, y) \in D,$$

où f désigne la **densité** de P^ζ pr à $\mu \otimes \nu$, f_ξ (resp f_η) la densité de P^ξ (resp P^η) pr à μ (resp pr à ν).

(ii) La notion de dépendance se définit, de façon analogue, à partir des **fonctions de répartition** F de (ξ, η) , F_ξ de ξ et F_η de η .

(iii) A titre d'exemple, on considère le cas où $(\mathcal{Y}, \mathcal{C}, \nu) = (\mathcal{X}, \mathcal{B}, \mu) = (\mathbf{R}, \mathcal{B}_\mathbf{R}, \lambda)$. Etant donné deux sous-**tribu de parties** $\mathcal{R}$ et $\mathcal{S}$ de $\mathcal{F}$, on peut définir plusieurs types d'**indices de dépendance**, eg :

(a) un indice fondé sur le **coefficient de corrélation** :

$$(3) \quad \rho(\mathcal{R}, \mathcal{S}) = \sup_{(\xi, \eta) \in \mathcal{L}^2(\Omega, \mathcal{R}, P) \times \mathcal{L}^2(\Omega, \mathcal{S}, P)} |\rho_{\xi\eta}|,$$

où $\mathcal{L}^2$ désigne $\mathcal{L}^2_\mathbf{R}$ et $\rho_{\xi\eta}$ le **coefficient de corrélation linéaire** (théorique) entre ξ et η ;

(b) un indice fondé sur la notion d'**indépendance stochastique** entre **événements** :

$$(4) \quad \delta(\mathcal{R}, \mathcal{S}) = \sup_{(R, S) \in \mathcal{R} \times \mathcal{S}} |P(R \cap S) - P(R) \cdot P(S)|,$$

ou encore :

$$(5) \quad \varphi(\mathcal{R}, \mathcal{S}) = \sup_{(R, S) \in \mathcal{R} \times \mathcal{S}} |P(S / R) - P(S)|, \quad \forall R : P(R) > 0.$$

(iii) Soit $X = (X_n)_{n \in \mathbb{N}^*}$ un **processus stochastique** en **temps** discret (ie une **suite** aléatoire). On pose (cf **tribu engendrée**) :

$$(6) \quad \mathcal{S}_{np} = \sigma(X_{n+p}, X_{n+p+1}, \dots) \text{ (tribu engendrée par la suite commençant en } n+p),$$

$$\mathcal{R}_p = \sigma(X_1, \dots, X_p) \text{ (tribu engendrée par la suite finissant en } p),$$

et l'on définit, à partir des indices ρ , δ et φ précédents, les **suites** :

$$\rho_n = \sup_{p \in \mathbb{N}^*} \rho(\mathcal{R}_p, \mathcal{S}_{n+p}),$$

$$(7) \quad \delta_n = \sup_{p \in \mathbb{N}^*} \delta(\mathcal{R}_p, \mathcal{S}_{n+p}),$$

$$\varphi_n = \sup_{p \in \mathbb{N}^*} \varphi(\mathcal{R}_p, \mathcal{S}_{n+p}).$$

On dit alors que X est un :

(a) **processus rho-mélangeant** (ou ρ -mélangeant) si $\lim_n \rho_n = 0$;

(b) **processus delta-mélangeant** (ou δ -mélangeant) si $\lim_n \delta_n = 0$;

(c) **processus phi-mélangeant** (ou φ -mélangeant) si $\lim_n \varphi_n = 0$ (cf **processus phi-mélangeant**).

(iv) La dépendance est ainsi un concept très important, qui traduit, d'une certaine façon, la **complexité** d'un **phénomène** d'intérêt, ou d'une **représentation statistique** de ce dernier.

Elle se retrouve dans la plupart des branches de la **Statistique**, généralement en conjonction avec les « outils » suivants :

(a) **conditionnement** : eg notamment **relation fonctionnelle**, **régression** et **modèle de régression**, **interdépendance** et **modèle d'interdépendance** ;

(b) **processus stochastique** : eg **martingale**, **processus autorégressif**, etc ;

(c) **loi multivariée** : eg **loi multidimensionnelle**, **tableau de contingence**, etc (cf Statistique multivariée, **analyse multidimensionnelle**, **analyse des données**, **modèle d'interdépendance**, etc) ;

(d) recherche d'**hypothèses alternatives** spécifiques (ie « non omnibus ») pour un **test d'indépendance**.