

UNIVERSITE ASSANE SECK DE ZIGUINCHOR



U.F.R DES SCIENCES ET TECHNOLOGIES

DÉPARTEMENT DE MATHÉMATIQUES

Mémoire de Master

DOMAINE : Sciences et Technologies
MENTION : Mathématiques et Applications
SPÉCIALITÉ : Mathématiques Appliqués
OPTION : Probabilité et Statistique

Thème :

Temps de rupture basé sur la mesure de divergence généralisée

Présenté par : Maimouna SAGNA

Sous la direction de : Dr Clément Manga

Sous la supervision de : Professeur Alassane Diédhiou

Soutenu publiquement le 26 janvier 2021

Devant le jury ci-après :

Prénom(s) et Nom	Grade	Qualité	Établissement
Marie Salomon SAMBOU	Professeur titulaire	Président du jury	UASZ
Edouard DIOUF	Maître de conférences	Examineur	UASZ
Diène Ngom	Maître de conférences	Examineur	UASZ
Emmanuel Nicolas CABRAL	Maître assistant	Examineur	UASZ
Alassane DIEDHIOU	Professeur titulaire	Superviseur	UASZ
Clément MANGA	Maître assistant	Directeur	UASZ

Résumé

Le phénomène de rupture dans les méthodes simulées a été largement étudié ces dernières années. Il a été introduit par Aldous et Diaconis dans [1], [2] et [3] pour capturer le fait que certaines chaînes de Markov convergent brusquement vers leurs distributions stationnaires. Une question importante est de détecter le temps après lequel on peut obtenir la convergence à l'équilibre d'une chaîne de Markov intéressante. Nous donnons dans ce mémoire, une méthode pour l'évaluation de cet instant appelé temps de rupture, en utilisant des temps d'arrêt appropriés tels que ceux trouvés dans [13]. Nous donnons les conditions sous lesquelles un processus de Markov a un temps de rupture, au sens de la mesure de divergence généralisée. Pour illustration, nous évaluons l'efficacité de notre méthode sur le processus d'Ornstein-Uhlenbeck.

Dédicace

Je souhaite dédié ce mémoire à toute ma famille particulièrement à mes soeurs pour m'avoir soutenues et encouragées durant les moments difficiles.

Remerciements

Je tiens à adresser mes remerciements à l'Université Assane Seck, l'UFR des Sciences et Technologies, le Département de Mathématiques ainsi que toutes les personnes qui ont contribué à la réalisation de ce mémoire. En particulier je remercie le **Docteur Clément MANGA** qui a accepté de diriger ce mémoire. Il m'a guidé dans mon travail, m'a consacré tout le temps qu'il me fallait. Je remercie le professeur **Alassane DIEDHIOU**, qui a bien accepté de superviser les travaux. Je remercie également :

- le professeur Marie Salomon SAMBOU pour avoir accepté d'être président du jury,
- les professeurs Edouard DIOUF, Diéne NGOM et Emmanuel Nicolas CABRAL pour accepter d'être les examinateurs de ce mémoire.

Je remercie tous mes professeurs, de la Licence 1 au Master, pour leurs conseils, leurs soutiens et pour toutes les connaissances qu'ils m'ont transmises. J'adresse également mes sincères remerciements au professeur **Pape NGOM** de l'université Cheikh Anta Diop de Dakar pour l'aide et le temps qu'il m'a consacré malgré la distance et son emploi du temps chargé.

Merci à toute ma famille spécialement à mes parents **Doudou SAGNA** et **Fatou BADJI** de toujours répondre présent à mes côtés. Un grand merci à mon oncle **Bassirou SAGNA** et sa femme qui n'ont jamais cessés de répondre à mes besoins et d'oeuvrer à mon succès dans tous mes projets. Merci également à ma soeur **Fatou SAGNA** ainsi qu'à son mari pour leur aide et leur soutien.

Je ne saurais terminer sans pour autant remercier tous mes camarades étudiants, en particulier mes promotionnaires ainsi que mes aînés doctorants et docteurs du département de Mathématiques pour leur aide et leurs judicieux conseils.

Table des matières

Introduction générale	1
1 Préliminaires	2
1.1 Processus stochastiques	2
1.1.1 Filtration	2
1.1.2 Processus	2
1.1.3 Processus gaussien	3
1.1.4 Mouvement brownien	3
1.1.5 Processus de Markov	3
1.1.6 Processus d'Itô	3
1.1.7 Formule d'Itô	4
1.1.8 Equation différentielle stochastique (EDS)	4
1.2 Distribution d'une forme quadratique de vecteur gaussien	6
1.3 Mesure de divergence	7
1.3.1 Définitions et méthodes	7
1.4 Les tests d'hypothèse	9
1.4.1 Rappels de quelques notions de base de la Statistique	9
1.4.2 Les types de test d'hypothèse	10
1.4.3 Méthodologie	10
2 Résultats théoriques	12
2.1 Phénomène de rupture pour le processus d'Ornstein-Uhlenbeck (O.U)	12
2.2 Comportement asymptotique de $D_\alpha^2(\hat{\theta}_n, \theta_0)$	17
3 Résultats numériques	22
3.1 Procédure de simulations	22
3.2 Analyse des résultats de simulations	23
Conclusion générale	25

Introduction Générale

On sait depuis longtemps que pour diverses familles de processus de Markov, le phénomène de rupture a été généralement observé. Pour un grand nombre d'applications traitant des chaînes de Markov, le problème de la détection de l'instant de convergence vers la probabilité stationnaire a été résolu en étudiant le phénomène de la convergence brusque connue sous le nom de phénomène de rupture. Un processus Markovien a une rupture s'il existe un instant, appelé temps de rupture, autour duquel il y a une petite fenêtre de temps dans laquelle la convergence a lieu. Avant l'instant de rupture, le processus reste loin de sa distribution asymptotique et s'en approche rapidement de façon exponentielle après cet instant. L'évaluation de cet instant est un problème qui a donné naissance à un grand nombre de travaux.

Lorsqu'un échantillon de chaînes de Markov est simulé en parallèle, l'instant de rupture est connu explicitement (voir Barrera et al [4]). En étudiant la marche aléatoire sur l'hypercube, Diaconis et Shahshahani [8] et Diaconis et al (1990) ont observé une partie des premiers exemples du phénomène de rupture. Diaconis [6], Porod [18] et Ycart [22] ont fourni des enquêtes complètes de l'instant de rupture de la convergence brusque à l'équilibre. Récemment, Saloff-Coste [20] et Barrera et al [4] ont étendu la rupture respectivement à plusieurs marches aléatoires et à des processus convergeant exponentiellement. En 2006, Lachaud et Ycart [14] proposent une méthode qui, utilisant le temps d'atteinte nous permet de détecter la convergence au stationnaire. Cette méthode est illustrer avec le processus d'Ornstein-Uhlenbeck. Dans ce cas il a été prouvé par Lachaud [13], que le temps d'arrêt proposé est asymptotiquement équivalent au temps de rupture.

L'objet de notre travail est d'étudier le phénomène de rupture pour un échantillon de processus convergeant exponentiellement et en particulier pour un échantillon de processus d'Ornstein-Uhlenbeck (O.U) en utilisant la mesure de divergence de type Rényi. L'intérêt des processus d'O.U est que presque tous les calculs peuvent être effectués de manière explicite. Nous proposons une méthode de temps d'arrêt qui peut empiriquement détecter la convergence d'un échantillon de processus. Dans ce cas, au lieu du test de Kolmogrov-Smirnov comme dans Lachaud et Ycart [14], le test statistique de type Rényi sera présenté.

Notre travail est divisé en trois chapitres :

- Le premier chapitre sera une étude préliminaire de certaines notions qui constituent la base de notre raisonnement.
- Le second chapitre sera consacré aux résultats théoriques, d'abord sur le phénomène de rupture pour le processus d'Ornstein-Uhlenbeck et ensuite sur le comportement asymptotique de la mesure de divergence.
- Le troisième chapitre sera une illustration pratique de notre méthode en rapportant les résultats d'une expérience numérique pour la détection de la convergence.

Chapitre 1

Préliminaires

Ce chapitre contient un bref rappel de quelques notions de base de probabilités et de statistiques. Il est constitué essentiellement de définitions, de lemmes, de propriétés et de corollaires dont la plupart n'ont pas été démontrés. Tout au long de ce chapitre on utilisera le terme distribution pour désigner la loi du processus X .

1.1 Processus stochastiques

L'objet de la théorie des processus stochastiques (ou aléatoire) est l'étude des phénomènes aléatoires dépendant du temps. Soient :

- $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé, où Ω est l'espace fondamentale et \mathcal{F} une tribu sur Ω , et \mathbb{P} une probabilité sur (Ω, \mathcal{F})
- (E, \mathcal{E}) un espace mesurable appelé espace des états, avec \mathcal{E} est une tribu sur E
- \mathbb{T} , un ensemble appelé ensemble des temps (par exemple $\mathbb{T} = \mathbb{R}^+$).

1.1.1 Filtration

Définition 1.1.1 On appelle filtration toute famille croissante $(\mathcal{F}_t, t \in \mathbb{T})$ de sous-tribus de \mathcal{F} et $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ est appelé espace de probabilité filtré.

Définition 1.1.2 La tribu engendrée par une famille de variables aléatoires $(X_t, t \in [0, T])$ est la plus petite tribu contenant les ensembles $\{X_t^{-1}(A)\}$ pour tout $t \in [0, T]$ et $A \in \mathcal{B}_{\mathbb{R}}$ où $\mathcal{B}_{\mathbb{R}}$ est la tribu de Borel sur \mathbb{R} . On note cette tribu $\sigma(X_t, t \leq T)$.

1.1.2 Processus

Définition 1.1.3 Un processus stochastique à valeurs dans (E, \mathcal{E}) basé sur $(\Omega, \mathcal{F}, \mathbb{P})$ est une famille $(X_t, t \in \mathbb{T})$ de variables aléatoires (abréviation : v.a.) de $(\Omega, \mathcal{F}, \mathbb{P})$ dans (E, \mathcal{E}) .

Définition 1.1.4 Pour $\omega \in \Omega$ fixé, on appelle $(X_t(\omega))_{t \in \mathbb{T}}$ la trajectoire de ω .

On parle de processus à temps continu si $\mathbb{T} = \mathbb{R}$, à temps discret si $\mathbb{T} \subset \mathbb{Z}$ est infini et on parle de famille à paramètre si \mathbb{T} est fini.

Pour $t \in \mathbb{T}$ fixé, X_t est une variable aléatoire.

Définition 1.1.5 Un processus stochastique $(X_t, t \in \mathbb{T})$ est dit adapté (par rapport à une filtration \mathcal{F}_t) si X_t est \mathcal{F}_t -mesurable pour tout t .

On dit que le processus est à trajectoires continues (ou est continu) si les applications $t \rightarrow X_t(\omega)$ sont continues pour presque tout ω .

Un processus est dit càdlàg (continu à droite, pourvu de limite à gauche) si ses trajectoires sont continues à droite pourvues de limites à gauche. Même définition pour càglàd.

A un Processus X , on associe sa filtration naturelle $\mathcal{F}_t^X = \sigma(X_s, s \leq t)$.

1.1.3 Processus gaussien

Définition 1.1.6 On dit que la variable aléatoire X est gaussienne de moyenne m et de variance σ^2 et on note $\mathcal{N}(m, \sigma^2)$ si elle a pour densité :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

Définition 1.1.7 Un vecteur aléatoire $X = (X_1, \dots, X_d)$ est gaussien si toute combinaison linéaire de ses composantes est une variable aléatoire gaussienne c'est à dire si pour tout d -uplet $(\lambda_1, \dots, \lambda_d)$ de réels, la variable aléatoire $\lambda_1 X_1 + \dots + \lambda_d X_d$ est gaussienne.

Définition 1.1.8 Un processus stochastique $X = \{X_t, t \geq 0\}$ est gaussien si toute combinaison linéaire finie est une variable aléatoire gaussienne, c'est-à-dire si $\forall n \geq 1, \forall t_i, 1 \leq i \leq n$ le vecteur $(X_{t_1}, \dots, X_{t_n})$ est gaussien.

1.1.4 Mouvement brownien

Définition 1.1.9 On appelle mouvement brownien un processus gaussien $(B_t, t \in \mathbb{R}^+)$ presque sûrement continu, tel que $B_0 = 0$ et $B_{t+s} - B_t$ est indépendant de $\sigma(B_u, u \leq t)$ de loi $\mathcal{N}(0, s)$ pour tout $t, s \geq 0$.

1.1.5 Processus de Markov

Définition 1.1.10 Pour toute variable aléatoire réelle X positive (respectivement intégrable), et toute sous-tribu \mathcal{G} de \mathcal{F} , il existe une variable aléatoire $\mathbb{E}(X|\mathcal{G})$, unique à égalité presque sûre près, positive (respectivement intégrable), telle que :

- (i) $\mathbb{E}(X|\mathcal{G})$ est \mathcal{G} -mesurable ;
- (ii) pour tout $A \in \mathcal{G}$, $\mathbb{E}(1_A X) = \mathbb{E}(1_A \mathbb{E}(X|\mathcal{G}))$

La variable aléatoire $\mathbb{E}(X|\mathcal{G})$ s'appelle l'espérance conditionnelle de X sachant \mathcal{G} .

Définition 1.1.11 Un processus est de Markov si son comportement dans le future ne dépend du passé qu'à travers le présent. Plus précisément, on dit qu'un processus (X_t) est de Markov si pour toute fonction f borélienne bornée

$$\mathbb{E}(f(X_t)|\mathcal{F}_s^X) = \mathbb{E}(f(X_t)|X_s), \quad \forall t > s$$

Les chaînes de Markov représentent des processus de Markov à temps discret dont l'espace d'état est fini ou dénombrable.

1.1.6 Processus d'Itô

On note par Λ l'ensemble $L_{loc}^2(\Omega \times \mathbb{R}^+)$ des processus θ adaptés càglàd vérifiant $\mathbb{E}\left(\int_0^t \theta_s^2(\omega) ds\right) < \infty, \forall t \geq 0$.

Définition 1.1.12 Un processus X est un processus d'Itô si

$$X_t = x + \int_0^t b_s ds + \int_0^t \sigma_s dB_s$$

où b est un processus adapté tel que $\int_0^t |b_s| ds$ existe (au sens de Lebesgue) p.s pour tout t et $\sigma \in \Lambda$.

On utilise la notation plus concise suivante :

$$\begin{cases} dX_t = b_t dt + \sigma_t dB_t \\ X_0 = x \end{cases}$$

b est le drift ou la dérive, σ est le coefficient de diffusion.

1.1.7 Formule d'Itô

Dans ce qui suit, X est un processus d'Itô de décomposition $dX_t = b_t dt + \sigma_t dB_t$.

Théorème 1.1.1 ([12])

Soit f une fonction de \mathbb{R} dans \mathbb{R} , de classe C^2 , à dérivées bornées. Alors

$$f(X_t) = f(X_0) + \int_0^t \frac{df}{dx}(X_s) dX_s + \frac{1}{2} \int_0^t \frac{d^2 f}{dx^2}(X_s) \sigma_s^2 ds$$

ou encore

$$df(X_t) = \frac{df}{dx}(X_t) dX_t + \frac{1}{2} \frac{d^2 f}{dx^2}(X_t) \sigma_t^2 dt.$$

Pour une fonction dépendant du temps on a :

Théorème 1.1.2 ([12])

Soit f une fonction définie sur $\mathbb{R}^+ \times \mathbb{R}$ de classe C^1 par rapport à t , de classe C^2 par rapport à x , à dérivées bornées, on a :

$$f(t, X_t) = f(0, X_0) + \int_0^t \frac{\partial f}{\partial t}(s, X_s) ds + \int_0^t \frac{\partial f}{\partial x}(s, X_s) dX_s + \frac{1}{2} \int_0^t \frac{\partial^2 f}{\partial x^2}(s, X_s) \sigma_s^2 ds$$

Ce que l'on note :

$$df(t, X_t) = \left[\frac{\partial f}{\partial t}(t, X_t) + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(t, X_t) \sigma_t^2 \right] dt + \frac{\partial f}{\partial x}(t, X_t) dX_t.$$

1.1.8 Equation différentielle stochastique (EDS)

Définition 1.1.13 Une équation différentielle stochastique est une équation de la forme :

$$X_t = x + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s \quad (1.1)$$

ou sous forme condensée

$$\begin{cases} dX_t = b(t, X_t) dt + \sigma(t, X_t) dB_t \\ X_0 = x \end{cases}$$

L'inconnu est le processus X .

Définition 1.1.14 Soit b et σ deux fonctions de $\mathbb{R}^+ \times \mathbb{R}^n$ à valeurs réelles données. On se donne également un espace $(\Omega, \mathcal{F}, \mathbb{P})$ muni d'une filtration \mathcal{F}_t et d'un \mathcal{F}_t -mouvement brownien B . Une solution de (1.1) est un processus X continu, \mathcal{F}_t -adapté tel que les intégrales $\int_0^t b(s, X_s) ds$ et $\int_0^t \sigma(s, X_s) dB_s$ ont un sens et l'égalité

$$X_t = x + \int_0^t b(s, X_s) ds + \int_0^t \sigma(s, X_s) dB_s$$

est satisfaite pour tout t , \mathbb{P} -p.s.

1.1.8.1 Solution des EDS pour un modèle gaussien

Définition 1.1.15 Certains processus stochastiques qui donnent lieu à des modèles gaussiens sont de la forme :

$$dX_t = [\mu_1(t)X_t + \mu_2(t)]dt + \sigma(t)dB_t \quad (1.2)$$

où $\mu_1(\cdot)$, $\mu_2(\cdot)$, $\sigma(\cdot)$ sont des fonctions déterministes du temps et B un \mathcal{F}_t -mouvement brownien standard.

Nous avons le résultat suivant :

Théorème 1.1.3 Si $\mathbb{P}\left(\int_0^T \frac{\sigma^2(s)}{\Phi^2(s)} ds < \infty\right) = 1$, alors la solution forte de l'équation différentielle stochastique (1.2) est :

$$X_t = \Phi(t) \left[X_0 + \int_0^t \frac{\mu_2(s)}{\Phi(s)} ds + \int_0^t \frac{\sigma(s)}{\Phi(s)} dB_s \right]$$

où la fonction $\Phi(\cdot)$ est la solution de l'équation différentielle ordinaire

$$d\Phi(t) = \mu_1(t)\Phi(t)dt, \quad \Phi(0) = 1.$$

Preuve : voir [11] □

Exemple : Résolvons l'équation différentielle stochastique suivante :

$$dX_t = -\rho X_t dt + \sigma\sqrt{2\rho}dB_t, \quad X_0 = x, \quad t \geq 0. \quad (1.3)$$

Posons $Y_t = e^{\rho t} X_t$, où X_t est solution de 1.3. D'après la formule d'Itô, on a

$$\begin{aligned} Y_t &= Y_0 + \rho \int_0^t e^{\rho s} X_s ds + \int_0^t e^{\rho s} dX_s \\ &= x + \rho \int_0^t e^{\rho s} X_s ds - \rho \int_0^t e^{\rho s} X_s ds + \sigma\sqrt{2\rho} \int_0^t e^{\rho s} dB_s \\ &= x + \sigma\sqrt{2\rho} \int_0^t e^{\rho s} dB_s. \end{aligned}$$

Cela nous amène à déduire X_t , on a alors

$$X_t = xe^{-\rho t} + \sigma\sqrt{2\rho}e^{-\rho t} \int_0^t e^{\rho s} dB_s = xe^{-\rho t} + \sigma\sqrt{2\rho} \int_0^t e^{-\rho(t-s)} dB_s$$

La solution X trouvée ici est le processus d'Ornstein-Uhlenbeck. Il est gaussien d'espérance $xe^{-\rho t}$ et de variance $\sigma^2(1 - e^{-2\rho t})$. De plus X est un processus de Markov.

En effet, en écrivant

$$X_s = xe^{-\rho s} + \sigma\sqrt{2\rho} \int_0^s e^{-\rho(s-u)} dB_u$$

et

$$X_s e^{\rho(s-t)} = xe^{-\rho t} + \sigma\sqrt{2\rho} \int_0^s e^{-\rho(t-u)} dB_u$$

on en déduit pour $s \leq t$

$$X_t = X_s e^{\rho(s-t)} + \sigma\sqrt{2\rho} \int_s^t e^{-\rho(t-u)} dB_u$$

ou encore

$$X_{t+s} = X_s e^{-\rho t} + \sigma\sqrt{2\rho} \int_0^t e^{-\rho(t-u)} d\tilde{B}_u$$

où le processus \tilde{B} défini par $\tilde{B} = B_{s+u} - B_s$ est un mouvement brownien indépendant de \mathcal{F}_s (donc de X_s)

en particulier

$$\mathbb{E}(f(X_{t+s})|\mathcal{F}_s) = \mathbb{E}(f(X_s e^{-\rho t} + Y)|\mathcal{F}_s) = \mathbb{E}(f(X_{t+s})|X_s)$$

Y étant une v.a indépendante de \mathcal{F}_s , ce qui établit le caractère markovien de X .

1.2 Distribution d'une forme quadratique de vecteur gaussien

Soient A une matrice réelle symétrique d'ordre q , X un vecteur gaussien d'espérance μ et de matrice de covariance Σ , et X^T la transposée du vecteur X . On veut pouvoir déterminer la distribution de probabilité de la forme quadratique $X^T A X$ sans aucune restriction sur les paramètres du vecteur X .

Définition 1.2.1 (matrice symétrique) Soit $A = (a_{ij})_{1 \leq i, j \leq q}$ une matrice carrée réelle d'ordre q . On dit que A est symétrique si et seulement si elle est égale à sa transposée, c'est à dire si $a_{ij} = a_{ji}$ pour tout i, j compris entre 1 et q

Pour toute matrice réelle symétrique G d'ordre q on peut écrire : $G = E \Lambda_q E^T$ où $\Lambda_q = \text{diag}(\lambda_1, \dots, \lambda_q)$ est la matrice diagonales des valeurs propres de G et E la matrice orthogonales des vecteurs propres de G .

Lorsque $\text{rang}(G) = r$, G a r valeurs propres non nulles et $q - r$ valeurs propres nulles.

Définition 1.2.2 (matrice symétrique positive) Une matrice carrée symétrique A est positive si toutes ses valeurs propres sont positives ou nulles

Pour toute matrice réelles symétriques positive Q d'ordre q , on a au moins une racine carrée T d'ordre $q \times t$ telle que $T T^T = Q$. Comme $Q = E \Lambda_q E^T$, une possible racine carrée de Q est donnée par :

$$T = E \Lambda_q^{\frac{1}{2}} E^T$$

où $\Lambda_q^{\frac{1}{2}} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_q})$

Définition 1.2.3 (matrice de covariance) Si elle existe, la matrice de covariance d'un vecteur aléatoire est symétrique réelle positive. Elle est donc diagonalisable en base orthonormée.

Lemme 1.2.1 Soient L une matrice $n \times m$ et M une matrice $m \times n$. Alors LM et ML possèdent les mêmes valeurs propres non nulles avec les mêmes ordres de multiplicités.

Preuve : voir [21] p.54 □

Considérons maintenant une matrice A réelle symétrique et Σ une matrice de covariance (qui est symétrique définie positive). Soit S la racine carrée de Σ définie comme ci-dessus. Alors il suit du lemme 1.2.1 et de la symétrie de $S^T A S$ que $S^T A S$ et $A \Sigma$ ont les même valeurs propres avec les même ordres de multiplicité, leur nombre est égale au $\text{rang}(S^T A S) = \text{rang}(\Sigma A \Sigma)$.

Proposition 1.2.1 Soit $X \sim \mathcal{N}_q(0, \Sigma)$. Soit A une matrice réelle symétrique d'ordre q . soit $r = \text{rang}(\Sigma A \Sigma)$, $r \geq 1$ et soient $\lambda_1, \dots, \lambda_r$, les valeurs propres non nulles de $A \Sigma$. Alors

$$X^T A X \cong \sum_{i=1}^r \lambda_i U_i^2$$

où les U_1, \dots, U_r sont indépendants et identiquement distribuées selon la loi normale centrée réduite.

Preuve : [9] corollaire 2.1 □

1.3 Mesure de divergence

Définition 1.3.1 Soit S un ensemble convexe, une divergence sur S est une application $D : S \times S \rightarrow \mathbb{R}^+$ satisfaisant :

$$\forall (p, q) \in S \times S, \quad D(p, q) \geq 0 \text{ et } D(p, q) = 0 \text{ si et seulement si } p = q$$

Comme une distance, une divergence doit être positive, en outre elle ne satisfait pas nécessairement l'inégalité triangulaire et l'axiome de symétrie.

En théorie des probabilités, une divergence est une fonction $D(P, Q)$ qui mesure la différence entre deux distributions de probabilité P et Q .

1.3.1 Définitions et méthodes

Soit $(\mathcal{X}, \mathcal{A}, \lambda)$ un espace mesuré avec λ une mesure finie ou σ -finie. Soient μ_1, μ_2 deux mesures de probabilité sur \mathcal{X} , telles que μ_1 et μ_2 absolument continues par rapport à λ . On note les densités de Radon-Nikodym des μ_i par rapport à λ par $p_i(x)$:

$$p_i(x) = \frac{\mu_i(dx)}{\lambda(dx)}, \quad i = 1, 2$$

La divergence de Kullback Leibler est donné par la quantité :

$$K(\mu_1, \mu_2) = \int_{\mathcal{X}} p_1(x) \ln \left(\frac{p_1(x)}{p_2(x)} \right) \lambda(dx) \quad (1.4)$$

Rényi (1961) pour la première fois donna une généralisation de la divergence donnée en (1.2). Cette généralisation est la suivante :

$$R(\mu_1, \mu_2) = \frac{1}{\alpha - 1} \ln \int_{\mathcal{X}} p_1^\alpha(x) p_2^{1-\alpha}(x) \lambda(dx), \quad \alpha \in \mathbb{R}, \alpha \neq 0 \text{ et } \alpha \neq 1 \quad (1.5)$$

Csiszar (1967) généralisa la divergence de Kullback Leibler par :

$$\mathcal{C}_\phi(\mu_1, \mu_2) = \int_{\mathcal{X}} p_2(x) \phi \left(\frac{p_1(x)}{p_2(x)} \right) \lambda(dx) \quad (1.6)$$

On retrouve la divergence de Kullback Leibler lorsque $\phi(x) = x \log(x)$. Dans notre travail, on considère la version modifiée de la divergence de Rényi donnée par :

$$D_\alpha(\mu_1, \mu_2) = \left[\frac{1}{\alpha(\alpha - 1)} \ln \int_{\mathcal{X}} p_1^\alpha(x) p_2^{1-\alpha}(x) \lambda(dx) \right]^{\frac{1}{2}}, \quad \alpha \in \mathbb{R}, \alpha \neq 0 \text{ et } \alpha \neq 1 \quad (1.7)$$

ainsi que ses limites en $\alpha = 1$ et $\alpha = 0$ données par :

$$D_0^2(\mu_1, \mu_2) = \lim_{\alpha \rightarrow 0} D_\alpha^2(\mu_1, \mu_2) = \int_{\mathcal{X}} p_2(x) \ln \left(\frac{p_2(x)}{p_1(x)} \right) \lambda(dx)$$

et

$$D_1^2(\mu_1, \mu_2) = \lim_{\alpha \rightarrow 1} D_\alpha^2(\mu_1, \mu_2) = \int_{\mathcal{X}} p_1(x) \ln \left(\frac{p_1(x)}{p_2(x)} \right) \lambda(dx)$$

En effet :

$$\begin{aligned} D_\alpha^2(\mu_1, \mu_2) &= \frac{1}{\alpha(\alpha - 1)} \ln \int_{\mathcal{X}} p_1^\alpha(x) p_2^{1-\alpha}(x) \lambda(dx) \\ &= \frac{1}{\alpha(\alpha - 1)} \ln \int_{\mathcal{X}} p_2(x) \exp \left(\alpha \ln \left(\frac{p_1(x)}{p_2(x)} \right) \right) \lambda(dx) \end{aligned}$$

Or $\exp\left(\alpha \ln\left(\frac{p_1(x)}{p_2(x)}\right)\right) \simeq 1 + \alpha \ln\left(\frac{p_1(x)}{p_2(x)}\right)$ au voisinage de $\alpha = 0$ donc :

$$\begin{aligned} D_\alpha^2(\mu_1, \mu_2) &\simeq \frac{1}{\alpha(\alpha - 1)} \ln \left[\int_{\mathcal{X}} p_2(x) \lambda(dx) + \alpha \int_{\mathcal{X}} p_2(x) \ln\left(\frac{p_1(x)}{p_2(x)}\right) \lambda(dx) \right] \\ &\simeq \frac{1}{\alpha(\alpha - 1)} \ln \left[1 + \alpha \int_{\mathcal{X}} p_2(x) \ln\left(\frac{p_1(x)}{p_2(x)}\right) \lambda(dx) \right] \\ &\simeq \frac{1}{\alpha - 1} \int_{\mathcal{X}} p_2(x) \ln\left(\frac{p_1(x)}{p_2(x)}\right) \lambda(dx) \text{ au voisinage de } 0 \end{aligned}$$

D'où

$$\lim_{\alpha \rightarrow 0} D_\alpha^2(\mu_1, \mu_2) = \int_{\mathcal{X}} p_2(x) \ln\left(\frac{p_2(x)}{p_1(x)}\right) \lambda(dx)$$

De même, on voit que :

$$D_\alpha^2(\mu_1, \mu_2) = \frac{1}{\alpha(\alpha - 1)} \ln \int_{\mathcal{X}} p_1(x) \exp\left((\alpha - 1) \ln\left(\frac{p_1(x)}{p_2(x)}\right)\right) \lambda(dx)$$

Comme

$$\exp\left((\alpha - 1) \ln\left(\frac{p_1(x)}{p_2(x)}\right)\right) \simeq 1 + (\alpha - 1) \ln\left(\frac{p_1(x)}{p_2(x)}\right)$$

au voisinage de $\alpha = 1$, on montre de la même manière que pour la limite en 0 que :

$$\lim_{\alpha \rightarrow 1} D_\alpha^2(\mu_1, \mu_2) = \int_{\mathcal{X}} p_1(x) \ln\left(\frac{p_1(x)}{p_2(x)}\right) \lambda(dx)$$

Nous considérons maintenant la mesure de divergence généralisée D_α entre deux mesures produit. Soit n un entier positif. Pour $i = 1, \dots, n$, soit (E_i, \mathcal{F}_i) un espace mesurable, μ_i, ν_i , deux mesures de probabilité sur E_i . Notons par E^n le produit cartésien $E_1 \times \dots \times E_n$ muni de la σ -algèbre produit. Notons par μ^n et ν^n le produit tensoriel des μ_i et des ν_i respectivement :

$$\mu^n = \mu_1 \otimes \dots \otimes \mu_n \text{ et } \nu^n = \nu_1 \otimes \dots \otimes \nu_n$$

Si λ_i est la mesure dominante de μ_i et ν_i alors $\lambda^n = \lambda_1 \otimes \dots \otimes \lambda_n$ sera pris comme mesure dominante de μ^n et ν^n . Et on a :

Proposition 1.3.1

$$D_\alpha^2(\mu^n, \nu^n) = \sum_{i=1}^n D_\alpha^2(\mu_i, \nu_i)$$

.

Preuve : Posons $\mu_i(dx_i) = p_i(x_i) \lambda_i(dx_i)$ et $\nu_i(dx_i) = q_i(x_i) \lambda_i(dx_i)$ pour $x_i \in E_i$. Pour tout $x \in E^n$ on a :

$$\begin{aligned} \mu^n(dx) &= \mu_1(dx_1) \times \dots \times \mu_n(dx_n) \\ &= p_1(x_1) \times \dots \times p_n(x_n) \cdot \lambda_1(dx_1) \times \dots \times \lambda_n(dx_n) \\ &= P_n(x) \cdot \lambda^n(dx) \end{aligned}$$

avec

$$P_n(x) = p_1(x_1) \times \dots \times p_n(x_n) \text{ et } \lambda^n(dx) = \lambda_1(dx_1) \times \dots \times \lambda_n(dx_n)$$

De même :

$$\nu^n(dx) = Q_n(x) \cdot \lambda^n(dx) \text{ avec } Q_n(x) = q_1(x_1) \times \dots \times q_n(x_n)$$

Pour $\alpha \neq 0, 1$ on a :

$$\begin{aligned}
D_\alpha^2(\mu^n, \nu^n) &= \frac{1}{\alpha(\alpha-1)} \ln \int_{E^n} P_n^\alpha(x) Q_n^{1-\alpha}(x) \lambda^n(dx) \\
&= \frac{1}{\alpha(\alpha-1)} \ln \int_{E_1} p_1^\alpha(x_1) q_1^{1-\alpha}(x_1) \lambda_1(dx_1) \times \cdots \times \int_{E_n} p_n^\alpha(x_n) q_n^{1-\alpha}(x_n) \lambda_n(dx_n) \\
&= \sum_{i=1}^n \frac{1}{\alpha(\alpha-1)} \ln \int_{E_i} p_i^\alpha(x_i) q_i^{1-\alpha}(x_i) \lambda_i(dx_i) \\
&= \sum_{i=1}^n D_\alpha^2(\mu_i, \nu_i)
\end{aligned} \tag{1.8}$$

□

On suppose que D_α est la distance entre μ_t , la distribution d'un certain processus X au temps t et sa limite μ_∞ :

$$D_\alpha(t) = D_\alpha(\mu_t, \mu_\infty)$$

Définition 1.3.2 (convergence exponentielle) Soient $X = \{X_t, t \geq 0\}$ un processus stochastique, ν une distribution de probabilité, ρ un réel positif. On dit que le processus X converge vers ν à taux exponentiel ρ , au sens de la distance D_α si :

$$\lim_{t \rightarrow \infty} \frac{\log D_\alpha(t)}{t} = -\rho$$

1.4 Les tests d'hypothèse

Dans tous les domaines, de l'expérimentation scientifique à la vie quotidienne, on est amené à prendre des décisions sur une activité risquée au vu de résultats d'expérience ou d'observation de phénomènes dans un contexte incertain. Dans chaque cas, le problème de décision consiste à trancher, au vu d'observations, entre une hypothèse nulle, notée \mathcal{H}_0 et une autre dite alternative, notée \mathcal{H}_1 . En général, on suppose qu'une et une seule de ces deux hypothèses est vraie. Un test d'hypothèse est une procédure permet de choisir entre ces deux hypothèses.

1.4.1 Rappels de quelques notions de base de la Statistique

Soit X une variable aléatoire prenant ses valeurs sur un espace \mathcal{X} . Soit $(\mathcal{X}, \mathcal{B}_\mathcal{X}, P_\theta)_{\theta \in \Theta}$ le modèle statistique associé à la variable aléatoire X où $\mathcal{B}_\mathcal{X}$ est une σ -algèbre de Borel des sous-ensembles $A \subset \mathcal{X}$ et $\{P_\theta\}_{\theta \in \Theta}$ une famille de distribution de probabilité définies sur l'espace mesurable $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ avec Θ un ouvert de \mathbb{R}^d , $d \geq 1$. On note par f_θ la densité de P_θ si X est continue de même que si X est discret et dans ce cas $f_\theta(x)$ est définie par :

$$f_\theta(x) = P_\theta(X = x)$$

Définition 1.4.1 (Échantillon) On appelle n -échantillon de la loi P_θ , la donnée d'un vecteur $\underline{X} = (X_1, \dots, X_n)$ constitué de n v.a indépendantes et identiquement distribuées (i.i.d) de loi P_θ .

On appelle modèle d'échantillonnage, le modèle $(\mathcal{X}^n, \mathcal{B}_\mathcal{X}^{\otimes n}, P_\theta^{\otimes n})_{\theta \in \Theta}$ avec $\mathcal{B}_\mathcal{X}^{\otimes n}$ la tribu produit sur \mathcal{X}^n et $P_\theta^{\otimes n} = P_\theta \otimes \cdots \otimes P_\theta$ est la probabilité produit sur $(\mathcal{X}^n, \mathcal{B}_\mathcal{X}^{\otimes n})$ qui est la loi du vecteur \underline{X} . La densité de l'échantillon sous P_θ est alors :

$$\underline{x} = (x_1, \dots, x_n) \mapsto \prod_{i=1}^n f_\theta(x_i)$$

avec \underline{x} une observation du vecteur \underline{X} .

Définition 1.4.2 (Vraisemblance) On appelle vraisemblance de l'observation x la fonction $L(x; \cdot)$ définie sur Θ à valeur dans \mathbb{R}^+ par $\theta \mapsto L(x; \theta) = f_\theta(x)$

Dans le cas d'un modèle d'échantillonnage, la vraisemblance de l'échantillon observé

$\underline{x} = (x_1, \dots, x_n)$ s'écrit sous la forme $L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_\theta(x_i)$.

Définition 1.4.3 (Statistique) Soit $(\mathcal{X}^n, \mathcal{B}_{\mathcal{X}}^{\otimes n}, P_\theta^{\otimes n})_{\theta \in \Theta}$ un modèle d'échantillonnage, on appelle statistique la variable aléatoire $T(\underline{X}) = T(X_1, \dots, X_n)$ où T est une fonction mesurable connu de $(\mathcal{X}^n, \mathcal{B}_{\mathcal{X}}^{\otimes n}, P_\theta^{\otimes n})_{\theta \in \Theta}$ vers un espace probabilisable (F, \mathcal{F}) .

Définition 1.4.4 (Estimateur) On appelle estimateur de $g(\theta)$, toute statistique $T(\underline{X})$ de $(\mathcal{X}^n, \mathcal{B}_{\mathcal{X}}^{\otimes n})$ à valeur dans $g(\Theta)$.

Définition 1.4.5 (Estimateur du maximum de vraisemblance) On appelle estimateur du maximum de vraisemblance (EMV) la statistique $\hat{\theta}(X)$ où $\hat{\theta}$ est une application de \mathcal{X} dans Θ telle que

$$L(x, \hat{\theta}(x)) \geq L(x, \theta)$$

pour tout $\theta \in \Theta$. On note

$$\hat{\theta}(x) = \text{Argmax}_\theta L(x, \theta)$$

Dans le cas d'un modèle d'échantillonnage l'EMV est :

$$\hat{\theta}(\underline{X}) = \text{Argmax}_\theta L(\underline{X}, \theta)$$

Dans la pratique, on prend comme estimation du maximum de vraisemblance, la solution de l'équation de la vraisemblance :

$$\nabla \ln L(X, \theta) = 0$$

Remarque 1.4.1 La mesure de divergence entre les distributions de probabilité P_{θ_1} et P_{θ_2} est donnée par :

$$D_\alpha^2(P_{\theta_1}, P_{\theta_2}) = D_\alpha^2(\theta_1, \theta_2) = \frac{1}{\alpha(1-\alpha)} \ln \left(\int_{\mathcal{X}} f_{\theta_1}^\alpha(x) f_{\theta_2}^{1-\alpha}(x) d\mu(x) \right); \quad \alpha \neq 0, 1$$

1.4.2 Les types de test d'hypothèse

On distingue différentes catégories de tests :

- tests sur une hypothèse relative à la valeur particulière d'un ou de plusieurs paramètre(s) ou tests paramétriques,
- tests de conformité de deux distributions ou tests d'ajustement entre une distribution théorique et une distribution expérimentale,
- tests de comparaison de deux populations (comparaison des moyennes, des variances...),
- tests d'indépendance de deux caractères quantitatifs.

1.4.3 Méthodologie

Le principe des tests d'hypothèse est de poser une hypothèse de travail et de prédire les conséquences de cette hypothèse pour la population ou l'échantillon. On compare ces prédiction avec les observations et l'on conclut en acceptant ou en rejetant l'hypothèse de travail à partir des règles de décision objectives.

Dans un problème de décision, deux types d'erreurs sont possibles :

- **Erreur de première espèce** : décider que \mathcal{H}_1 est vrai alors que \mathcal{H}_0 est vrai.

- **Erreur de seconde espèce** : décider que \mathcal{H}_0 est vrai alors que \mathcal{H}_1 est vrai.

A toute décision correspond une probabilité de décider juste et une probabilité de se tromper :

Définition 1.4.6 *On appelle risque d'erreur de première espèce, noté α , la probabilité de rejeter \mathcal{H}_0 et d'accepter \mathcal{H}_1 alors que \mathcal{H}_0 est vrai.*

La probabilité de cet événement est le niveau de signification du test. On dit aussi que le niveau de signification est la probabilité de rejeter l'hypothèse nulle à tort. Ce risque α définit la région critique, R_c , d'aire α et de probabilité α , sous l'hypothèse \mathcal{H}_0 .

Définition 1.4.7 *On appelle risque d'erreur de seconde espèce, notée β , la probabilité de rejeter \mathcal{H}_1 et d'accepter \mathcal{H}_0 alors que \mathcal{H}_1 est vrai.*

La puissance du test est la probabilité de rejeter \mathcal{H}_0 et d'accepter \mathcal{H}_1 alors que \mathcal{H}_1 est vrai. Sa valeur est $1 - \beta$.

Différentes étapes doivent être suivies pour tester une hypothèse :

1. formuler de façon précise l'hypothèse nulle \mathcal{H}_0 et l'hypothèse alternative \mathcal{H}_1
2. fixer avant l'expérience le niveau de signification du test α et déterminer la région critique associée ;
3. préciser les conditions d'application du test : forme de la loi de probabilité de la population étudiée, taille de l'échantillon, variance connue ou inconnue... ;
4. choisir la statistique S la mieux adaptée en fonction des caractéristiques de la population étudiée et définir sa loi sous les deux hypothèses ;
5. calculer, à partir des données fournies par l'échantillon, la valeur de la statistique S_{obs} ;
6. prendre une décision concernant les hypothèses posées.

La région critique R_c correspond à l'ensemble des valeurs telles que : $\mathbb{P}(S \in R_c) = \alpha$

1.4.3.1 Règles de décision

On appelle règle de décision une règle qui permet de choisir entre \mathcal{H}_0 et \mathcal{H}_1 au vu des observations sous la contrainte que la probabilité de rejeter à tort \mathcal{H}_0 est égale à α fixé. On a deux cas de figure :

1. Sous l'hypothèse « \mathcal{H}_0 est vrai » et pour un niveau de signification α fixé
 - (a) si la valeur de la statistique S_{obs} calculée appartient à la région critique alors l'hypothèse \mathcal{H}_0 est rejetée au risque d'erreur α et l'hypothèse \mathcal{H}_1 est acceptée ;
 - (b) si la valeur de la statistique S_{obs} n'appartient pas à la région critique alors l'hypothèse \mathcal{H}_0 ne peut être rejetée.
2. La probabilité critique ou p-value (en anglais), qui est la probabilité sous \mathcal{H}_0 que les observations s'éloignent de \mathcal{H}_0 au moins autant que ce qui a été observé, c'est à dire telle que $\mathbb{P}(S \geq S_{obs}) = \alpha_{obs}$, est évaluée
 - (a) si $\alpha_{obs} \geq \alpha$, l'hypothèse \mathcal{H}_0 est acceptée car le risque d'erreur de rejeter \mathcal{H}_0 alors qu'elle est vrai est trop important ;
 - (b) si $\alpha_{obs} < \alpha$, l'hypothèse \mathcal{H}_0 est rejetée car le risque d'erreur de rejeter \mathcal{H}_0 alors qu'elle est vrai est très faible.

Chapitre 2

Résultats théoriques

Tous les résultats théoriques de notre travail sont rassemblés dans ce chapitre.

2.1 Phénomène de rupture pour le processus d'Ornstein-Uhlenbeck (O.U)

le phénomène de rupture pour des échantillons de processus convergeant exponentiellement a été étudié, et la base théorique de la méthode a été développée par Barrera et al (2006). Dans cette section, on étudie l'instant de rupture pour un n -échantillon de processus d'Ornstein-Uhlenbeck indépendants identiquement distribués. Ici nous discuterons aussi d'une façon de comprendre la divergence entre deux mesures de probabilité.

Définition 2.1.1 (Temps de rupture)

Pour chaque $n \geq 0$, soit $X^{(n)}$ un processus stochastique sur un espace $E^{(n)}$, convergeant en loi vers une certaine distribution de probabilité $\nu^{(n)}$. Soit d une distance entre distribution de probabilité, prenant ses valeurs entre 0 et M , où M est soit un réel positif soit l'infini. Pour $t \geq 0$, notons par $d^{(n)}(t)$ la distance entre la distribution de $X^{(n)}$ au temps t et $\nu^{(n)}$:

$$d^{(n)}(t) = d(\mathcal{L}X^{(n)}(t), \nu^{(n)})$$

Soit (t_n) une suite de réels positifs. La suite $(X^{(n)})$ est dit avoir une rupture au temps (t_n) au sens de la distance d si pour $c > 0$:

$$c < 1 \Rightarrow \lim_{n \rightarrow \infty} d^{(n)}(ct_n) = M$$

$$c > 1 \Rightarrow \lim_{n \rightarrow \infty} d^{(n)}(ct_n) = 0$$

On a le lemme suivant :

Lemme 2.1.1 Pour $i = 1, 2, \dots, n$, soient d_i une fonction strictement positive définie sur \mathbb{R}^+ , et ρ_i un réel strictement positif. Pour $n \geq 1$, notons par $\rho(1, n), \dots, \rho(n, n)$ les valeurs de ρ_1, \dots, ρ_n rangées dans l'ordre croissant et τ_n le réel suivant :

$$\tau_n = \max \left\{ \frac{\log i}{\rho(i, n)}, \quad i = 1, \dots, n \right\}. \quad (2.1)$$

Supposons les hypothèses suivantes :

1. Il existe une fonction positive g , décroissante et tendant vers 0 quand t tend vers l'infini et un réel positif t_0 tel que pour tout $t \geq t_0$ et tout $i \geq 1$

$$\left| \frac{\log d_i(t)}{t} + \rho_i \right| \leq g(t) \quad (2.2)$$

2.

$$\lim_{n \rightarrow \infty} \rho(1, n) \tau_n = +\infty \quad (2.3)$$

3. Pour tout réel positif c ,

$$\lim_{n \rightarrow \infty} \frac{g(c\tau_n)}{\rho(1, n)} = 0 \quad (2.4)$$

Alors pour tout entier positif k , tout réel positif c et toute suite (τ'_n) telle que $\lim_{n \rightarrow \infty} \frac{\tau'_n}{\tau_n} = 1$

$$c < 1 \Rightarrow \lim_{n \rightarrow \infty} \sum_{i=1}^n \left(d_i \left(c \frac{\tau'_n}{k} \right) \right)^k = +\infty,$$

$$c > 1 \Rightarrow \lim_{n \rightarrow \infty} \sum_{i=1}^n \left(d_i \left(c \frac{\tau'_n}{k} \right) \right)^k = 0$$

Preuve :

— Démontrons premièrement le résultat pour la suite τ_n .

Définissons $g_i(t)$ par :

$$\begin{aligned} g_i(t) &= \frac{\log d_i(t)}{t} + \rho_i \\ \Rightarrow d_i(t) &= \exp(t(g_i(t) - \rho_i)) \end{aligned}$$

Ainsi

$$\sum_{i=1}^n \left(d_i \left(\frac{c\tau_n}{k} \right) \right)^k = \sum_{i=1}^n \exp \left(-c\tau_n \rho_i + c\tau_n g_i \left(\frac{c\tau_n}{k} \right) \right)$$

Or

$$|g_i(t)| \leq g(t)$$

Donc

$$S_n \exp \left(-c\tau_n g \left(\frac{c\tau_n}{k} \right) \right) \leq \sum_{i=1}^n \left(d_i \left(\frac{c\tau_n}{k} \right) \right)^k \leq S_n \exp \left(c\tau_n g \left(\frac{c\tau_n}{k} \right) \right)$$

avec

$$S_n = \sum_{i=1}^n \exp(-c\tau_n \rho_i) = \sum_{i=1}^n \exp(-c\tau_n \rho(i, n))$$

Premier cas : $c > 1$

Pour $c > 1$, utilisons la borne supérieure de cette inégalité.

On observe que pour tout $i = 1, \dots, n$, $\exp(-c\tau_n \rho(i, n)) \leq i^{-c}$ puisque $\tau_n \geq \frac{\log i}{\rho(i, n)}$.

Pour tout $l = 1, \dots, n-1$, on peut écrire

$$S_n = \sum_{i=1}^l \exp(-c\tau_n \rho(i, n)) + \sum_{i=l+1}^n \exp(-c\tau_n \rho(i, n))$$

Ce qui implique que

$$\begin{aligned} S_n &\leq l \exp(-c\tau_n \rho(1, n)) + \sum_{l+1}^n i^{-c} \quad \text{car } \rho(1, n) \leq \rho(i, n) \quad \forall i = 1, \dots, n \\ &\leq l \exp(-c\tau_n \rho(1, n)) + \int_l^n x^{-c} dx \\ &\leq l e^{-c\tau_n \rho(1, n)} + \frac{1}{c-1} \left(l^{-(c-1)} - n^{-(c-1)} \right) \end{aligned}$$

Cette borne reste vrai pour $l = n$. Posons maintenant $l_n = \lfloor e^{\rho(1,n)\tau_n} \rfloor$ où $\lfloor \cdot \rfloor$ est la partie entière.

l_n n'est pas plus grand que n par définition de τ_n . Alors on a :

$$\begin{aligned} S_n &\leq l_n e^{-c\tau_n\rho(1,n)} + \frac{1}{c-1} \left(l_n^{-(c-1)} - n^{-(c-1)} \right) \\ &\leq l_n e^{-c\tau_n\rho(1,n)} + \frac{1}{c-1} l_n^{-(c-1)} \end{aligned}$$

Or $e^{\rho(1,n)\tau_n} - 1 \leq l_n \leq e^{\rho(1,n)\tau_n}$

$$\begin{aligned} \Rightarrow S_n &\leq e^{-(c-1)\tau_n\rho(1,n)} + \frac{1}{c-1} \left(e^{\rho(1,n)\tau_n} - 1 \right)^{-(c-1)} \\ &\leq e^{-(c-1)\tau_n\rho(1,n)} \left(1 + \frac{\left(1 - e^{-\rho(1,n)\tau_n} \right)^{-(c-1)}}{c-1} \right) \end{aligned}$$

Ce qui implique que

$$\begin{aligned} \sum_{i=1}^n \left(d_i \left(\frac{c\tau_n}{k} \right) \right)^k &\leq e^{-(c-1)\tau_n\rho(1,n)} \left(1 + \frac{\left(1 - e^{-\rho(1,n)\tau_n} \right)^{-(c-1)}}{c-1} \right) \exp \left(c\tau_n g \left(\frac{c\tau_n}{k} \right) \right) \\ &= \left(1 + \frac{\left(1 - e^{-\rho(1,n)\tau_n} \right)^{-(c-1)}}{c-1} \right) \exp \left(-\rho(1,n)\tau_n \left(c-1 + c \frac{g \left(\frac{c\tau_n}{k} \right)}{\rho(1,n)} \right) \right) \end{aligned}$$

qui tend vers 0 quand n tend vers l'infini en utilisant (2.3) et (2.4). D'où

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \left(d_i \left(\frac{c\tau_n}{k} \right) \right)^k = 0$$

Deuxième cas : $0 < c < 1$

Pour $0 < c < 1$, on utilise la borne inférieure de la sommation des d_i . Pour chaque n , on choisit i_n^* tel que $\tau_n = \frac{\log i_n^*}{\rho(i_n^*, n)} \Rightarrow i_n^* = \exp(\tau_n \rho(i_n^*, n))$

$$\begin{aligned} S_n &\geq \sum_{i=1}^{i_n^*} \exp(-c\tau_n\rho(i, n)) \\ &\geq i_n^* \exp(-c\tau_n\rho(i_n^*, n)) \\ &\geq \exp(\tau_n\rho(i_n^*, n)) \times \exp(-c\tau_n\rho(i_n^*, n)) \\ &\geq \exp((1-c)\tau_n\rho(1, n)) \end{aligned}$$

D'où

$$\sum_{i=1}^n \left(d_i \left(\frac{c\tau_n}{k} \right) \right)^k \geq \exp \left(\tau_n\rho(1, n) \left(1 - c - c \frac{g \left(\frac{c\tau_n}{k} \right)}{\rho(1, n)} \right) \right)$$

qui tend vers $+\infty$ quand n tend vers l'infini en utilisant (2.3) et (2.4).

- Démontrons le résultat maintenant pour une suite (τ'_n) équivalente à τ_n . On considère maintenant une suite (τ'_n) équivalente à τ_n . La nouvelle somme peut être bornée comme précédemment :

$$S'_n \exp \left(-c\tau'_n g \left(\frac{c\tau'_n}{k} \right) \right) \leq \sum_{i=1}^n \left(d_i \left(\frac{c\tau'_n}{k} \right) \right)^k \leq S'_n \exp \left(c\tau'_n g \left(\frac{c\tau'_n}{k} \right) \right)$$

avec

$$S'_n = \sum_{i=1}^n \exp(-c\tau'_n \rho(i, n)).$$

Pour $c > 1$:

Soit $c' < 1$ fixé tel que $cc' > 1$. Pour n assez grand $c' \leq \frac{\tau'_n}{\tau_n} \leq \frac{1}{c'}$.

Ainsi

$$S'_n \leq \sum_{i=1}^n \exp(-cc'\tau_n \rho(i, n))$$

et

$$\exp\left(c\tau'_n g\left(\frac{c\tau'_n}{k}\right)\right) \leq \exp\left(\frac{c}{c'}\tau_n g\left(\frac{cc'\tau_n}{k}\right)\right).$$

Comme g est décroissante, donc la borne supérieure de S_n peut être appliquée à S'_n en remplaçant c par cc' . On obtient :

$$\begin{aligned} \sum_{i=1}^n \left(d_i\left(\frac{c\tau'_n}{k}\right)\right)^k &\leq e^{-(cc'-1)\tau_n \rho(1, n)} \left(1 + \frac{(1 - e^{-\rho(1, n)\tau_n})^{-(cc'-1)}}{cc' - 1}\right) \exp\left(\frac{c}{c'}\tau_n g\left(\frac{cc'\tau_n}{k}\right)\right) \\ &= \left(1 + \frac{(1 - e^{-\rho(1, n)\tau_n})^{-(cc'-1)}}{cc' - 1}\right) \exp\left(-\rho(1, n)\tau_n \left(cc' - 1 - \frac{c}{c'} \frac{g\left(\frac{cc'\tau_n}{k}\right)}{\rho(1, n)}\right)\right) \end{aligned}$$

qui tend vers 0 quand n tend vers l'infini.

La preuve est similaire pour la borne inférieure. □

Soit $X^{(n)} = (X_1, \dots, X_n)$ un n -uplet de processus indépendants, notons par $D_\alpha^i(t)$ la distance à l'équilibre du i -ième processus au temps t . Supposons que ρ_1, \dots, ρ_n sont les taux de convergence exponentielle des processus X_1, \dots, X_n respectivement. On a alors le théorème suivant :

Théorème 2.1.1 *Supposons que $D_\alpha^i(t)$, définie ci-dessus, et ρ_i satisfont les hypothèses (2.2) à (2.4) du lemme précédent. Notons par $X^{(n)}$ le n -uplet de processus (X_1, \dots, X_n) Alors la suite de processus $X^{(n)}$ a une rupture au sens de la distance D_α au temps*

$$t_n = \max \left\{ \frac{\log i}{2\rho(i, n)}; \quad i = 1, \dots, n \right\} \quad (2.5)$$

Où les $\rho(1, n), \dots, \rho(n, n)$ sont les valeurs de ρ_1, \dots, ρ_n rangées dans l'ordre croissant.

Preuve : Posons $d(t) = D_\alpha^2(t) = D_\alpha^2(\mu_t^{(n)}, \mu_\infty^{(n)})$ avec $\mu_t^{(n)}$ la distribution du processus $X^{(n)}$ et $\mu_\infty^{(n)}$ sa limite.

D'après la proposition 1.3.1 du chapitre précédent

$$d(ct_n) = \sum_{i=1}^n \left(D_\alpha^i(ct_n)\right)^2$$

Comme $D_\alpha^i(t)$ et ρ_i satisfont les hypothèses du lemme 2.1.1 ci-dessus et que $t_n = \frac{\tau_n}{2}$, on peut appliquer le lemme 2.1.1 à $d(t)$. Ainsi $\sum_{i=1}^n \left(D_\alpha^i(ct_n)\right)^2$ tend vers $+\infty$ pour $0 < c < 1$ et vers 0 pour $c > 1$.

D'où $X^{(n)}$ a une rupture au temps t_n au sens de la distance D_α . □

Rappelons que le processus d'Ornstein-Uhlenbeck (O.U) est définie comme suit :

Définition 2.1.2 Soient ρ et σ deux réels positifs et x_0 un réel. Soit $(B_t)_{t \geq 0}$ un mouvement brownien standard. On appelle processus d'Ornstein-Uhlenbeck issu de x_0 de paramètres ρ et σ l'unique solution de l'équation différentielle stochastique :

$$\begin{cases} dX_t = -\rho X_t dt + \sigma \sqrt{2\rho} dB_t \\ X_0 = x_0 \end{cases} \quad (2.6)$$

La solution de cette équation différentielle stochastique peut être explicitement écrite comme :

$$X_t = x_0 e^{-\rho t} + \sigma \sqrt{2\rho} \int_0^t e^{-\rho(t-s)} dB_s, \quad t \geq 0.$$

Par conséquent le processus $(X_t)_{t \geq 0}$ est gaussien d'espérance $x_0 e^{-\rho t}$ et de variance $\sigma^2 (1 - e^{-2\rho t})$. Ainsi la distribution asymptotique μ_∞ du processus X est gaussienne centrée de variance σ^2 notée $\mathcal{N}(0, \sigma^2)$:

$$X_t \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

quand t tend vers l'infini. Cette convergence a lieu au taux exponentiel ρ .

Notons par μ_t la distribution du processus X au temps t et supposons qu'il existe un réel positif D tel que :

$$\lim_{t \rightarrow \infty} D_\alpha(\mu_t, \mu_\infty) e^{\rho t} = D \quad (2.7)$$

Remarque 2.1.1 Pour tout $\alpha \in \mathbb{R}$, $\alpha \neq 0, 1$:

$$D_\alpha(\mu_t, \mu_\infty) = D_{1-\alpha}(\mu_\infty, \mu_t)$$

Donc calculer $D_\alpha(\mu_t, \mu_\infty)$ revient à calculer $D_{1-\alpha}(\mu_\infty, \mu_t)$.

Soit $X^{(n)} = (X_1, \dots, X_n)$ un n-échantillon du processus X, i.e, un n-uplet de copies indépendantes de X. Notons par $D_\alpha^n(t)$ la mesure de divergence généralisée entre la distribution de $X^{(n)}$ au temps t la distribution à l'équilibre $\mu_\infty^{\otimes n}$. Nous avons pour un échantillon $X^{(n)}$ du processus d'O.U le théorème suivant :

Théorème 2.1.2 Supposons que l'hypothèse (2.7) est vérifiée, et soit u un réel fixé. Alors

$$\lim_{n \rightarrow +\infty} D_\alpha^n \left(u + \frac{\log n}{2\rho} \right) = D e^{-\rho u}$$

Preuve : Soit $X_t = x_0 e^{-\rho t} + \sigma \sqrt{2\rho} \int_0^t e^{-\rho(t-s)} dB_s$ le processus d'O.U,

$$X_t \sim \mathcal{N}(x_0 e^{-\rho t}, \sigma^2 (1 - e^{-2\rho t}))$$

$$\Rightarrow f_t(x) = \frac{1}{\sigma \sqrt{2\pi(1 - e^{-2\rho t})}} \exp \left(-\frac{(x - x_0 e^{-\rho t})^2}{2\sigma^2(1 - e^{-2\rho t})} \right)$$

Comme $X_\infty \sim \mathcal{N}(0, \sigma^2)$, on a $f_\infty(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left(-\frac{x^2}{2\sigma^2} \right)$. Donc pour tout $\alpha \in \mathbb{R}$, $\alpha \neq 0, 1$:

$$f_t^{1-\alpha}(x) f_\infty^\alpha(x) = \frac{(1 - e^{-2\rho t})^{\frac{\alpha}{2}}}{\sigma \sqrt{2\pi(1 - e^{-2\rho t})}} \exp \left(-\frac{(1 - \alpha)x_0^2 e^{-2\rho t}}{2\sigma^2(1 - e^{-2\rho t})} \right) \times \exp \left(-\frac{(1 - \alpha e^{-2\rho t})x^2 - 2(1 - \alpha)xx_0 e^{-\rho t}}{2\sigma^2(1 - e^{-2\rho t})} \right)$$

Pour $\alpha < e^{2\rho t}$, on a $(1 - \alpha e^{-2\rho t}) > 0$

$$f_t^{1-\alpha}(x) f_\infty^\alpha(x) = A(\alpha) \times \frac{\sqrt{1 - \alpha e^{-2\rho t}}}{\sigma \sqrt{2\pi(1 - e^{-2\rho t})}} \exp \left(-\frac{\left(x - \frac{(1-\alpha)x_0 e^{-\rho t}}{1 - \alpha e^{-2\rho t}} \right)^2}{\frac{2\sigma^2(1 - e^{-2\rho t})}{1 - \alpha e^{-2\rho t}}} \right)$$

avec

$$A(\alpha) = \frac{(1 - e^{-2\rho t})^{\frac{\alpha}{2}}}{\sqrt{1 - \alpha e^{-2\rho t}}} \exp\left(\frac{\alpha(1 - \alpha)x_0^2 e^{-2\rho t}}{2\sigma^2(1 - \alpha e^{-2\rho t})}\right)$$

Ainsi

$$D_\alpha(\mu_t, \mu_\infty) = \left[\frac{1}{\alpha(1 - \alpha)} \left(\frac{\alpha}{2} \log(1 - e^{-2\rho t}) - \frac{1}{2} \log(1 - \alpha e^{-2\rho t}) \right) + \frac{x_0^2 e^{-2\rho t}}{2\sigma^2(1 - \alpha e^{-2\rho t})} \right]^{\frac{1}{2}}$$

Or

$$\lim_{t \rightarrow \infty} \left(\frac{\alpha \log(1 - e^{-2\rho t})}{2 e^{-2\rho t}} - \frac{1 \log(1 - \alpha e^{-2\rho t})}{2 e^{-2\rho t}} \right) = 0$$

donc

$$\lim_{t \rightarrow \infty} D_\alpha(\mu_t, \mu_\infty) e^{\rho t} = \frac{|x_0|}{\sigma\sqrt{2}}$$

Puisque les X_i sont indépendants et identiquement distribués, on a

$$(D_\alpha^n(t))^2 = n D_\alpha^2(\mu_t, \mu_\infty) \Rightarrow D_\alpha^n(t) = (n D_\alpha^2(\mu_t, \mu_\infty))^{\frac{1}{2}}$$

Ce qui implique que

$$D_\alpha^n\left(u + \frac{\log n}{2\rho}\right) = \left[\frac{1}{\alpha(1 - \alpha)} \left(\frac{n\alpha}{2} \log\left(1 - \frac{e^{-2\rho u}}{n}\right) - \frac{n}{2} \log\left(1 - \alpha \frac{e^{-2\rho u}}{n}\right) \right) + \frac{x_0^2 e^{-2\rho u}}{2\sigma\left(1 - \alpha \frac{e^{-2\rho u}}{n}\right)} \right]^{\frac{1}{2}}$$

Alors on a :

$$\lim_{n \rightarrow \infty} D_\alpha^n\left(u + \frac{\log n}{2\rho}\right) = \frac{|x_0|}{\sigma\sqrt{2}} e^{-\rho u}$$

□

Ce théorème donne le comportement de l'échantillon par rapport à l'instant de rupture $\frac{\log n}{2\rho}$ en fonction du signe de u . Soit R un réel positif assez grand.

- Si $u = -R$ alors $De^{-\rho u}$ est grand : avant l'instant de rupture $\frac{\log n}{2\rho}$ l'échantillon est loin de l'équilibre.
- Si $u = R$ alors $De^{-\rho u}$ est petit : après l'instant $\frac{\log n}{2\rho}$ l'échantillon converge exponentiellement de manière rapide vers l'équilibre.

2.2 Comportement asymptotique de $D_\alpha^2(\hat{\theta}_n, \theta_0)$

L'objectif de cette section est entièrement de servir de support théorique pour les résultats numériques du chapitre suivant.

Soit $(\mathcal{X}, \mathcal{B}_\mathcal{X}, P)$ un espace statistique associé à la variable aléatoire X , où $\mathcal{B}_\mathcal{X}$ est une σ -algèbre de Borel de sous ensemble de \mathcal{X} et P une famille de distributions de probabilité

$$P = \{P_\theta : \theta \in \Theta, \text{ où } \Theta \text{ est un ouvert de } \mathbb{R}^d\}$$

sur $(\mathcal{X}, \mathcal{B}_\mathcal{X})$. On suppose que les distributions de probabilité P_θ sont absolument continues par rapport à une mesure σ -finie μ sur $(\mathcal{X}, \mathcal{B}_\mathcal{X})$, où

$$f(x, \theta) = \frac{dP_\theta}{d\mu}$$

sont les fonctions de densité qui vérifient les conditions de régularité suivantes :

1. L'ensemble $A = \{x \in \mathcal{X} / f(x, \theta) > 0\}$ est indépendant de θ et pour tout $x \in A$, $\theta \in \Theta$,

$$\frac{\partial f(x, \theta)}{\partial \theta_i}, \quad \frac{\partial^2 f(x, \theta)}{\partial \theta_i \partial \theta_j}, \quad \frac{\partial^3 f(x, \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}, \quad i, j, k = 1, \dots, d$$

existent et sont finies

2. Il existe des fonctions à valeurs réelle $F(x)$ et $H(x)$ telles que

$$\left| \frac{\partial f(x, \theta)}{\partial \theta_i} \right| < F(x), \quad \left| \frac{\partial^2 f(x, \theta)}{\partial \theta_i \partial \theta_j} \right| < F(x), \quad \left| \frac{\partial^3 f(x, \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| < H(x),$$

Où F est d'intégrale finie et $\mathbb{E}[H(X)] < L$, avec L indépendant de θ .

3. La matrice de l'information de Fisher

$$I_F(\theta) = \left[\mathbb{E} \left\{ \frac{\partial \log f(X, \theta)}{\partial \theta_i} \frac{\partial \log f(X, \theta)}{\partial \theta_j} \right\} \right]_{i,j=1,\dots,d}$$

est finie et définie positive

Ici, on choisit une importante mesure de divergence donnée par :

$$\begin{aligned} D_\alpha^2(\theta_1, \theta_2) &= \frac{1}{\alpha(1-\alpha)} \log \left(\int_{\mathcal{X}} f_{\theta_1}^\alpha(x) f_{\theta_2}^{1-\alpha}(x) d\mu(x) \right); \quad \alpha \neq 0, 1. \\ &= h(\mathcal{C}_\phi(\theta_1, \theta_2)) \end{aligned} \quad (2.8)$$

avec

$$\begin{aligned} \mathcal{C}_\phi(\theta_1, \theta_2) &= \int_{\mathcal{X}} f_{\theta_2}(x) \phi \left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)} \right) d\mu(x) \quad ; \\ h(x) &= \frac{1}{r(r-1)} \log(r(r-1)x + 1), \quad r \neq 0, 1 \\ \phi(x) &= \frac{x^r - r(x-1) - 1}{r(r-1)}, \quad r \neq 0, 1 \end{aligned}$$

Où

$$h(0) = 0, \quad h'(0) = 1; \quad \phi(1) = 0, \quad \phi'(1) = 0, \quad \phi''(1) = 1$$

Nous supposons qu'il existe une suite $\hat{\theta}_n$ fortement consistant racine de l'équation de vraisemblance telle que :

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow{L} \mathcal{N} \left(0, I_F(\theta)^{-1} \right) \quad (2.9)$$

Cette suite est l'estimateur du maximum de vraisemblance (EMV) de θ .

Supposons maintenant que la distribution de la variable aléatoire X est F_X , soit $\mathcal{P} = \{E_i\}_{i=1,\dots,M}$ une partition de l'ensemble \mathcal{X} en M ensembles. Soit $\mathbf{p} = (p_1, \dots, p_M)^T$ les probabilités des ensembles E_i , $i = 1, \dots, M$, respectivement, telles que :

$$p_i = \Pr_{F_X}(X \in E_i) = \int_{E_i} dF_X, \quad i = 1, \dots, M$$

avec

$$\sum_{i=1}^M p_i = 1$$

Soit (X_1, \dots, X_n) un n -échantillon de variables aléatoires issu de la distribution F_X . Soient $N_i = \sum_{k=1}^n I_{E_i}(X_k)$ où $I_{E_i}(X_k)$ égale 1 si $X_k \in E_i$ zéro sinon et $\hat{\mathbf{p}} = (\hat{p}_1, \dots, \hat{p}_M)^T$ avec

$\hat{p}_i = \frac{N_i}{n}$, $i = 1, \dots, M$ les fréquences relatives des ensembles, respectivement. Le vecteur $N = (N_1, \dots, N_M)$ suit par définition une loi multinomiale d'effectif n et de paramètres p_1, p_2, \dots, p_M . On pose :

$$\Theta = \left\{ \theta = (p_1, \dots, p_{M-1})^T \in \mathbb{R}^{M-1}; \sum_{i=1}^{M-1} p_i < 1, p_i > 0, i = 1, \dots, M-1 \right\}$$

Et donc on a $\hat{\theta}_n = (\hat{p}_1, \dots, \hat{p}_{M-1})^T$.

Proposition 2.2.1 *La matrice de l'information de Fisher dans un modèle multinomial est donnée par $I_F(\theta) = I_F(\mathbf{p}) = (i_{r,s})_{r,s=1,\dots,M-1}$ où*

$$i_{r,s} = \begin{cases} \frac{1}{p_r} + \frac{1}{p_M} & \text{si } r = s \\ \frac{1}{p_M} & \text{si } r \neq s \end{cases}, \quad r, s = 1, \dots, M-1$$

Preuve : A voir dans [17] □

Proposition 2.2.2 *L'inverse de la matrice de l'information de Fisher est donné par :*

$$I_F(\theta)^{-1} = \text{diag}(\theta) - \theta\theta^T$$

où $\theta = (p_1, \dots, p_{M-1})^T$

$\text{diag}(\theta)$ est une matrice diagonale dont les éléments de la diagonale sont les composantes du vecteur θ .

Preuve : A voir dans [17] □

On veut tester l'hypothèse $\mathcal{H}_0 : F_X = F_0$ contre $\mathcal{H}_1 : F_X \neq F_0$, F_0 connu.

Si \mathcal{H}_0 est vrai, alors $\Pr_{F_X}(X \in E_i) = \Pr_{F_0}(X \in E_i)$ pour tout $i = 1, \dots, M$. Posons $\mathbf{p}^0 = (p_1^0, \dots, p_M^0)^T$ avec $p_i^0 = \Pr_{F_0}(X \in E_i)$, $i = 1, \dots, M$. Cela revient à tester l'hypothèse $\mathbf{p} = \mathbf{p}^0$, et donc aussi l'hypothèse $\theta = \theta_0$, où $\theta_0 = (p_1^0, \dots, p_{M-1}^0)^T$. Comme par dessus on avait $\Theta \subset \mathbb{R}^d$, on peut donc supposer que $d = M-1$.

Théorème 2.2.1 *Supposons que le modèle et ϕ satisfont les conditions de régularité (1) à (3) et 2.9 respectivement. Soit θ_0 la vraie valeur du paramètre θ :*

(i) si $\theta \neq \theta_0$, on a :

$$\sqrt{n}[D_\alpha^2(\hat{\theta}_n, \theta_0) - D_\alpha^2(\theta, \theta_0)] \rightarrow \mathcal{N}[0, \sigma^2(\theta, \theta_0)],$$

où $\sigma^2(\theta, \theta_0) = AI_F(\theta)^{-1}A^T$ et $A = \nabla D_\alpha^2(\theta, \theta_0)$ avec $\nabla = \left(\frac{\partial}{\partial \theta_1}, \dots, \frac{\partial}{\partial \theta_d} \right)$

(ii) si $\theta = \theta_0$ alors on a :

$$2nD_\alpha^2(\hat{\theta}_n, \theta_0) \rightarrow \sum_{i=1}^r \beta_i Z_i^2$$

où les Z_i sont $\mathcal{N}[0, 1]$ et les β_i sont les valeurs propres de $M(\theta)\Sigma_\theta$ avec $r = \text{rang}(\Sigma_\theta M(\theta)\Sigma_\theta)$ avec $M(\theta) = \left(\frac{\partial^2 D_\alpha^2(\theta, \theta_0)}{\partial \theta_i \partial \theta_j} \right)_{i,j=1,\dots,d}$

Preuve :

Nous supposons que D_α a des dérivés partielles secondes continues satisfaisant les conditions de régularité (1) à (3).

(i) Si $\theta = \theta_0$;

le développement de Taylor à l'ordre 1 de $D_\alpha^2(\hat{\theta}_n, \theta_0)$ en θ donne :

$$D_\alpha^2(\hat{\theta}_n, \theta_0) = D_\alpha^2(\theta, \theta_0) + \nabla D_\alpha^2(\theta, \theta_0)(\hat{\theta}_n - \theta) + o(\|\hat{\theta}_n - \theta\|)$$

En posant $A = \nabla D_\alpha^2(\theta, \theta_0)$ on a

$$\begin{aligned} D_\alpha^2(\hat{\theta}_n, \theta_0) - D_\alpha^2(\theta, \theta_0) &= A(\hat{\theta}_n - \theta) + o(\|\hat{\theta}_n - \theta\|) \\ \sqrt{n}[D_\alpha^2(\hat{\theta}_n, \theta_0) - D_\alpha^2(\theta, \theta_0)] &= A\sqrt{n}(\hat{\theta}_n - \theta) + \sqrt{n}o(\|\hat{\theta}_n - \theta\|) \end{aligned}$$

Or $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} \mathcal{N}(0, I_F(\theta)^{-1})$, alors

$$A\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} \mathcal{N}(0, \sigma^2(\theta, \theta_0))$$

avec $\sigma^2(\theta, \theta_0) = AI_F(\theta)^{-1}A^T$

Il est clair que les variables aléatoires $\sqrt{n}[D_\alpha^2(\hat{\theta}_n, \theta_0) - D_\alpha^2(\theta, \theta_0)]$ et $A\sqrt{n}(\hat{\theta}_n - \theta)$ ont la même distribution car

$$\sqrt{n}o(\|\hat{\theta}_n - \theta\|) = \sqrt{n}o(O_P(n^{-1/2})) = o_P(1)$$

(ii) Si $\theta = \theta_0$;

le développement en série de Taylor du second ordre de $D_\alpha^2(\hat{\theta}_n, \theta_0)$ en θ donne :

$$D_\alpha^2(\hat{\theta}_n, \theta_0) = D_\alpha^2(\theta, \theta_0) + \sum_{i=1}^d \frac{\partial D_\alpha^2(\theta, \theta_0)}{\partial \theta_i} (\hat{\theta}_{ni} - \theta_i) + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2 D_\alpha^2(\theta, \theta_0)}{\partial \theta_i \partial \theta_j} (\hat{\theta}_{ni} - \theta_i)(\hat{\theta}_{nj} - \theta_j) + o(\|\hat{\theta}_n - \theta\|^2)$$

Or

$$D_\alpha^2(\theta, \theta_0) = h(\mathcal{C}_\phi(\theta, \theta_0))$$

avec

$$\mathcal{C}_\phi(\theta, \theta_0) = \sum_{k=1}^M p_k^0 \phi\left(\frac{p_k}{p_k^0}\right) = \sum_{k=1}^{d=M-1} p_k^0 \phi\left(\frac{p_k}{p_k^0}\right) + p_M^0 \phi\left(\frac{1 - \sum_{k=1}^{d=M-1} p_k}{p_M^0}\right)$$

Et

$$\frac{\partial D_\alpha^2(\theta, \theta_0)}{\partial \theta_i} = h'(\mathcal{C}_\phi(\theta, \theta_0)) \times \left(\phi'\left(\frac{p_i}{p_i^0}\right) - \phi'\left(\frac{p_M}{p_M^0}\right) \right), \quad i = 1, \dots, d = M - 1$$

Si $\theta = \theta_0$ alors $p = p^0$ et dans ce cas on a $\mathcal{C}_\phi(\theta, \theta_0) = 0$, $D_\alpha^2(\theta, \theta_0) = 0$ et $\frac{\partial D_\alpha^2(\theta, \theta_0)}{\partial \theta_i} = 0$, $i = 1, \dots, d$. On a donc

$$D_\alpha^2(\hat{\theta}_n, \theta_0) = \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2 D_\alpha^2(\theta, \theta_0)}{\partial \theta_i \partial \theta_j} (\hat{\theta}_{ni} - \theta_i)(\hat{\theta}_{nj} - \theta_j) + o(\|\hat{\theta}_n - \theta\|^2)$$

$$\Rightarrow 2nD_\alpha^2(\hat{\theta}_n, \theta_0) = n(\hat{\theta}_n - \theta)^T M(\theta)(\hat{\theta}_n - \theta) + no(\|\hat{\theta}_n - \theta\|^2)$$

avec

$$M(\theta) = \left(\frac{\partial^2 D_\alpha^2(\theta, \theta_0)}{\partial \theta_i \partial \theta_j} \right)_{i,j=1,\dots,d}$$

Ainsi

$$2nD_\alpha^2(\hat{\theta}_n, \theta_0) \text{ a la même distribution que } n(\hat{\theta}_n - \theta)^T M(\theta)(\hat{\theta}_n - \theta)$$

D'autre part

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} \mathcal{N}(0, \Sigma_\theta)$$

où

$$\Sigma_\theta = \text{diag}(\theta) - \theta\theta^T$$

Par le théorème de Slutsky on a d'une part

$$\sqrt{n}M(\theta)(\hat{\theta}_n - \theta) \rightarrow M(\theta)X$$

où

$$X \sim \mathcal{N}(0, \Sigma_\theta)$$

et d'autre part par le théorème de l'application continue

$$n(\hat{\theta}_n - \theta)^T M(\theta)(\hat{\theta}_n - \theta) = (\sqrt{n}(\hat{\theta}_n - \theta))^T (\sqrt{n}M(\theta)(\hat{\theta}_n - \theta)) \rightarrow X^T M(\theta)X$$

D'après le corollaire 1.2.1 du chapitre 1 on a $X^T M(\theta)X \simeq \sum_{i=1}^r \beta_i Z_i^2$

D'où

$$n(\hat{\theta}_n - \theta)^T M(\theta)(\hat{\theta}_n - \theta) \rightarrow \sum_{i=1}^r \beta_i Z_i^2$$

où les Z_i sont $\mathcal{N}[0, 1]$ et les β_i sont les valeurs propres de $M(\theta)\Sigma_\theta$ avec $r = \text{rang}(\Sigma_\theta M(\theta)\Sigma_\theta)$ avec $M(\theta) = \left(\frac{\partial^2 D_\alpha^2(\theta, \theta_0)}{\partial \theta_i \partial \theta_j} \right)_{i,j=1,\dots,d}$ \square

\square

Remarque 2.2.1 Sur la base du théorème 2.2.1 on a pour tout $i, j = 1, \dots, d = M - 1$:

$$M(\theta)_{i,j} = \frac{\partial^2 D_\alpha^2(\theta, \theta_0)}{\partial \theta_i \partial \theta_j} = h''(\mathcal{C}_\phi(\theta, \theta_0)) \times \frac{\partial \mathcal{C}_\phi(\theta, \theta_0)}{\partial \theta_j} \times \frac{\partial \mathcal{C}_\phi(\theta, \theta_0)}{\partial \theta_i} + h'(\mathcal{C}_\phi(\theta, \theta_0)) \times \frac{\partial^2 \mathcal{C}_\phi(\theta, \theta_0)}{\partial \theta_i \partial \theta_j}$$

Lorsque $\theta = \theta_0$ on obtient

$$M(\theta)_{i,j} = h'(0) \times \left. \frac{\partial^2 \mathcal{C}_\phi(\theta, \theta_0)}{\partial \theta_i \partial \theta_j} \right|_{\theta=\theta_0}.$$

Et

$$\left. \frac{\partial^2 \mathcal{C}_\phi(\theta, \theta_0)}{\partial \theta_i \partial \theta_j} \right|_{\theta=\theta_0} = \begin{cases} \frac{1}{p_i} + \frac{1}{p_M} & \text{si } i = j \\ \frac{1}{p_M} & \text{si } i \neq j \end{cases}$$

Par conséquent d'après la proposition 2.2.1 $M(\theta) = \Sigma_\theta^{-1}$, et donc $2nD_\alpha^2(\hat{\theta}_n, \theta_0)$ suit une loi de χ^2 à d degrés de liberté.

Ainsi donc nous obtenons une approximation de la fonction puissance $P(\theta) = \mathbb{P}(2nD_\alpha^2(\hat{\theta}_n, \theta_0) > \chi_d^2)$ de la façon suivante :

$$P(\theta) = 1 - \Phi \left[\frac{\sqrt{n}}{\sigma(\theta, \theta_0)} \left(\frac{\chi_d^2}{2n} - D_\alpha^2(\theta, \theta_0) \right) \right]$$

où $\Phi[\cdot]$ est la fonction de répartition d'une loi normale centrée réduite.

Ce résultat peut être utilisé pour obtenir une approximation de la puissance du test $D_\alpha^2(\hat{\theta}_n, \theta_0)$ de la même façon que le théorème 2.2.1.

Soient X un processus stochastique, $X^{(n)}$ un n-échantillon de ce processus, μ_t la distribution de X au temps t et μ_∞ sa limite. Les résultats précédents donnant la distribution asymptotique de la statistique de divergence

$$D_\alpha^2(\hat{\theta}_n, \theta_0) = D_\alpha^2(\mu_t(\hat{\theta}_n), \mu_\infty(\theta_0))$$

dans l'échantillonnage aléatoire peuvent être utilisés pour tester des hypothèses statistiques.

Nous voulons tester :

$$\begin{cases} \mathcal{H}_0 : & \mu_t = \mu_\infty \\ & \text{vs} \\ \mathcal{H}_1 : & \mu_t \neq \mu_\infty \end{cases} \quad (2.10)$$

Le chapitre suivant illustre ce fait.

Chapitre 3

Résultats numériques

Dans ce chapitre nous avons essayer d'illustrer les résultats théoriques du chapitre précédent.

3.1 Procédure de simulations

Soit X un processus stochastique convergeant exponentiellement vers sa distribution asymptotique. Soient (X_1, \dots, X_n) un n -échantillon de ce processus, μ_t la distribution de ce processus à l'instant t et μ_∞ sa distribution asymptotique. On note par M_n le processus moyen,

$$M_n(t) = \frac{1}{n} \sum_{i=1}^n X_i(t), \quad t \geq 0$$

Autrement dit $M_n(t)$ est la moyenne empirique de cet échantillon au temps t .

On note par T_n le temps d'atteinte de la valeur de l'espérance de la distribution asymptotique du processus échantillonné par le processus moyen M_n . Supposons que la distribution asymptotique est centrée alors on a :

$$T_n = \inf \{t \geq 0 : M_n(t) = 0\}$$

Il a été montré par Lachaud dans [13] que T_n est équivalent au temps de rupture, du moins en terme de convergence en probabilité.

Ainsi, pour la détection de l'instant de convergence du processus X à sa distribution asymptotique, nous considérons la méthode suivante pour laquelle la mise en oeuvre nécessite les calculs suivants :

1. choisissez un entier n et construisez un n -échantillon du processus ;
2. exécutez le n -échantillon et calculez le processus moyen à chaque instant ;
3. arrêtez quand le processus moyen atteint pour la première fois l'espérance de la distribution asymptotique.

Pour illustrer la méthode proposée ici, nous prenons un exemple. Le processus échantillonné est la diffusion d'Ornstein-Uhlenbeck décrite dans le chapitre précédent avec les paramètres suivants : $\rho = 1$, $\sigma = 1$, $x_0 = 10$. Dans notre étude nous considérons différents ensembles d'expériences dans lesquelles les données sont générées à partir du processus d'Ornstein-Uhlenbeck. La taille de l'échantillon varie de 100 à 2000 et pour chaque taille d'échantillon, le nombre de réplication est de 1000.

Afin de mettre en évidence le phénomène de rupture, nous considérons aussi la Méthodologie suivante basée sur l'étude de l'évolution du processus et de sa convergence. Notre procédure est la suivante :

- * nous fixons une suite de temps espacée régulièrement t_1, t_2, \dots, t_M avec $t_i = \frac{i}{70}$ et $i = 1, \dots, M = 1000$.

* l'expérience consiste à simuler un n-échantillon de processus d'Ornstein-Uhlenbeck (X_1, X_2, \dots, X_n) avec les paramètres suivants : $\rho = 1$, $\sigma = 1$, $x_0 = 10$.

A chaque temps t_i , nous testons si notre échantillon $(X_1(t_i), X_2(t_i), \dots, X_n(t_i))$ est gaussien avec $\mathcal{N}(0, \sigma^2)$.

Nous utilisons la statistique du test D_α de la qualité de l'ajustement avec $\alpha = 0.5$. La valeur p (p-value) de d_j est la probabilité d'observer une valeur de $D_\alpha(\mu_t, \mu_\infty)$ comme ou plus extrême que d_j selon la distribution normale. Ici, la valeur p du test est calculée pour chaque instant t_i . Ces valeurs p ont été produites sur la base de 1000 simulations de Monté Carlo sous les hypothèses nulles (2.10). Ici, le niveau de signification est fixé à 5%. Les moyennes de ces 1000 valeurs p pour chaque temps t_i , sont représentées dans les figures ci-dessous.

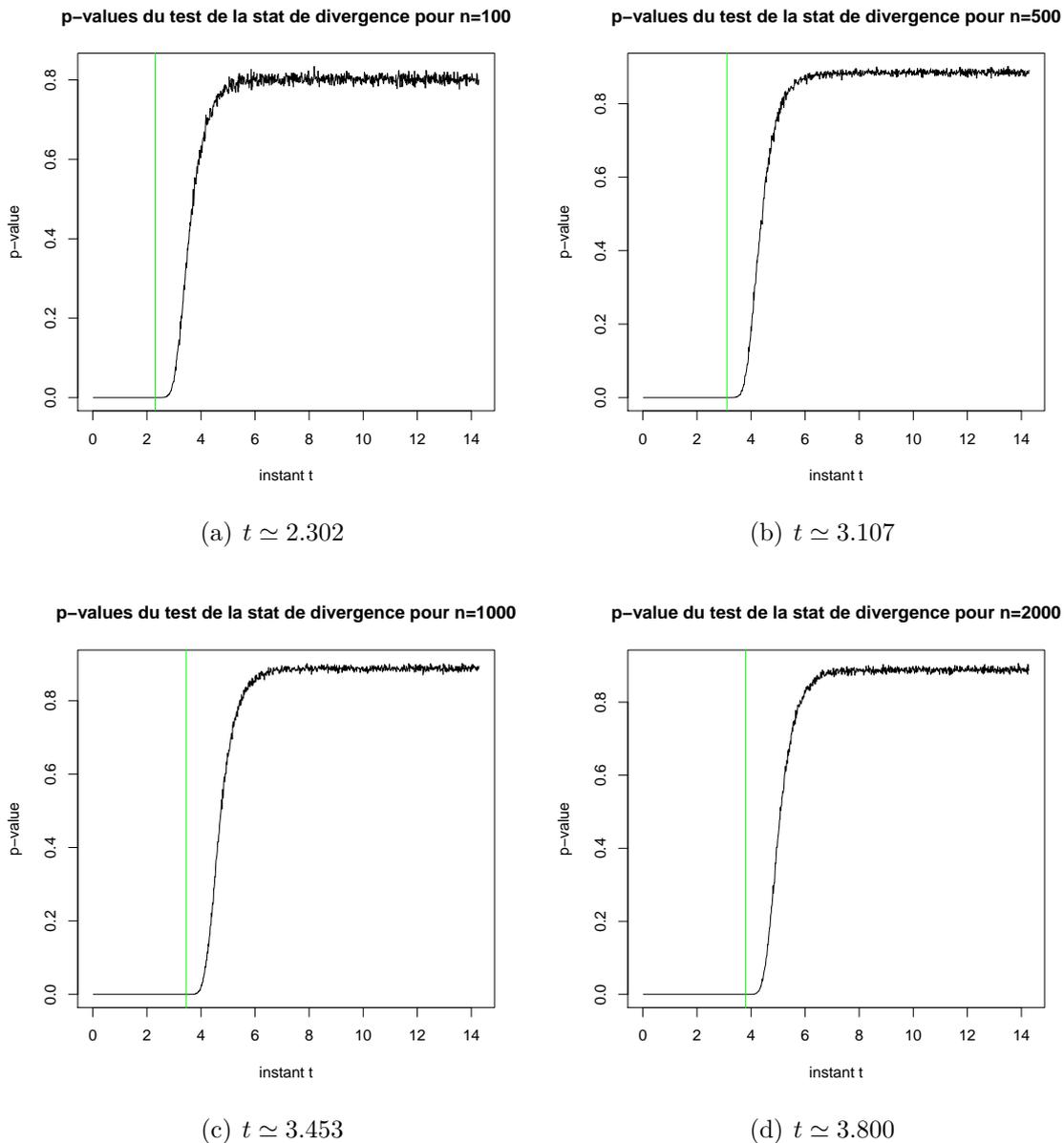


FIGURE 3.1 – Moyennes des valeurs p du test D_α pour différentes valeurs de n

3.2 Analyse des résultats de simulations

Il ressort clairement du tableau 3.1 que la convergence à la distribution asymptotique est relativement lente. Ce résultat est confirmé en figure 3.1 ci-dessus où nous présentons le tracé

n,taille d'échantillon	<i>Valeur p du test de divergence D_α</i>			
	t=02.14	t=04.28	t=05.71	t=08.57
150	0.00	0.01	0.64	0.67
300	0.00	0.07	0.73	0.76
500	0.00	0.15	0.69	0.78
600	0.00	0.20	0.72	0.73
800	0.00	0.12	0.68	0.85
1000	0.00	0.10	0.69	0.89

TABLE 3.1 – Valeurs p pour la statistique de test $D_\alpha(\mu_t, \mu_\infty)$ pour l'hypothèse que $\mu_t = \mu_\infty$

de la valeur p (p-value) pour la statistique D_α . Ces valeur p (p-value) pour la statistique D_α ont été calculées en s'appuyant sur les résultats de la seconde section du chapitre 2.

L'idée de base est simplement de représenter graphiquement la moyenne des 1000 valeurs p associées à chaque t_i , pour l'hypothèse que $\mu_t = \mu_\infty$. L'instant de rupture est matérialisé par la ligne verticale dans les figures. Afin de vérifier si l'équilibre est effectivement obtenu, l'expérience est exécutée après le temps d'atteinte.

La première figure 3.1(a) indique le cas où les valeurs p sont tirées pour $n=100$. Ici la ligne verticale correspond à l'instant de rupture $\frac{\log n}{2\rho} \simeq 2.302$.

Pour les figures 3.1(b),(c) et (d), les valeurs p sont générées à partir de $n=500, 1000, 2000$, avec temps de rupture 3.107, 3.453, 3.800 respectivement.

On a alors observé que la valeur p moyenne du test D_α est nulle avant l'instant de rupture et devient positive après cet instant. En outre l'hypothèse nulle est systématiquement rejetée avant l'instant de rupture. On attend environ le double de l'instant de rupture et on observe que le régime stationnaire est établi lorsque la valeur p moyenne se stabilise autour de 90%

Conclusion Générale

En résumé, à l'instar des notions classiques, nous avons introduit la distance D_α pour détecter la convergence d'un processus stochastique à sa distribution asymptotique. Nous avons étudié la propriété fondamentale de la convergence brusque à l'équilibre du processus d'Ornstein-Uhlenbeck (O.U). Nous nous sommes limités au cas d'O.U, mais les résultats des simulations au chapitre 3 montrent que le théorème 2.1.2 s'applique à un processus plus général. Notre étude, illustrée par des résultats numériques, a montré que la méthode proposée ici permet de détecter très bien la convergence du processus à sa distribution asymptotique, au sens de la distance D_α .

Bibliographie

- [1] D. Aldous, (1983). Random walks on finite groups and rapidly mixing Markov chains. Seminar on Probability, *XVII*, Lecture Notes in Math., vol. 986, Springer, Berlin, pp. 243 – 297. MR 770418
- [2] D. Aldous, P. Diaconis, (1986). Shuffling cards and Stopping times. *Amer. Math. Monthly*, 93, pp. 333 – 348. MR 0841111
- [3] D. Aldous, P. Diaconis, (1987). Strong uniform times and finite random walks. *Acta Appl. Math.*, 8, pp. 69 – 97. MR 0876954
- [4] J. Barrera, B. Lachaud, B. Ycart, (2006). Cutoff for n-tuples of exponentially converging processes. *Stochastic Processes and their Applications* 116, 1433 – 1446
- [5] I. Csiszar, (1967). Information-type measures of divergence of probability distributions and indirect observations. *Studia Sci. Math. Hung.* 299 – 318
- [6] P. Diaconis, (1996). The cutoff phenomenon in finite Markov chains. *Proc. Natl. Acad. Sci. USA* 93, 1659 – 1664.
- [7] P. Diaconis, R. Graham, J. Morrison, (1990). Asymptotic analysis of a random walk on a hypercube with many dimensions. *Rand. Struct. Algorithms* 1, 51 – 72
- [8] P. Diaconis, M. Shahshahani, (1987). Time to reach stationarity in Bernoulli-Laplace diffusion model. *SIAM J. Math. Anal.* 18(1), 208 – 218. MR871832.
- [9] J.J. Dik and M.C.M. de Gunst (1985). The distribution of general quadratic forms in normal variables. *Statistica Neerlandica*, 39, 14 – 26
- [10] A. Diédhiou, P. Ngom, (2009). Cutoff time based on generalized divergence measure, *Statistics and Probability Letters* 79, 1343 – 1350.
- [11] G. Gauthier. Equation différentielle stochastique (EDS) et le lemme d’Itô 80 – 646 – 08 *Calcul stochastique I*.
- [12] M. Jeanblanc, (2006). Cours de Calcul stochastique Master 2IF Every
- [13] B. Lachaud, (2005). Cutoff and hitting times for a sample of Ornstein-Uhlenbeck processes and its average. *J. Appl. Probab.* 42(4), 1069 – 1080.
- [14] B. Lachaud, B. Ycart, (2006). Convergence times for parallel Markov chains. In : *Lecture Notes in Control and Information Sciences* vol. 341. Springer, Berlin, Heidelberg, pp. 1069 – 1080

- [15] D. Morales, L. Pardo, (2000). New smooth test statistics of goodness of fit for categorized composite null hypotheses. *Sociedad de Estadística e Investigación Operativa, Test* 9,173–190.
- [16] D. Morales, M.L. Menendez, L. Pardo, M. Salicrù, (1994). On the applications of divergence type measures in testing statistical hypotheses. *J. Multivariate Anal.* 51, 372 – 391
- [17] L. Pardo (2006). *Statistical inference based on divergence measure.* Taylor and Francis Group, LLC, 18 – 179
- [18] U. Porod, (1996). The cutoff phenomenon for random reflections. *Ann. Probab.* 24(1), 74 – 96. MR1387627.
- [19] A. Rényi, (1961). On measures of entropy and information. In : *Proc. 4th Berk. Symp. Math. Stat. Probl.* vol. 1. University of California Press, pp. 547 – 561.
- [20] L. Saloff-Coste, (2004). Random walks on finite groups. In : *Kesten, H. (Ed.), Probability on Discrete Structures*, vol. 110. Springer, Berlin, pp. 263 – 346
- [21] J.H. WILKINSON, (1965). *The algebraic eigen value problem.* Clarendon Press, Oxford.
- [22] B. Ycart, (2000). Stopping tests for Markov chain Monte-Carlo methods. *Methodol. Comput. Appl. Probab.* 2(1), 23 – 36.