Optimalité du test de Wald pour des observations dépendantes

Christophe KOELL
Institut de Recherche Mathématique Avancée
Université Louis Pasteur et C.N.R.S.
7 rue René Descartes
67084 Strasbourg Cedex
France
e-mail : koell@math.u-strasbg.fr

31 janvier 1995

Résumé

Cette thèse s'insère dans le cadre de l'Analyse Séquentielle, théorie statistique reposant sur le caractère aléatoire du nombre d'observations prises en compte pour effectuer un test d'hypothèses, déterminer un intervalle de confiance pour des paramètres, etc. Nous nous intéressons plus particulièrement aux propriétés d'optimalité du test séquentiel du rapport des probabilités ou test de Wald. Ainsi nous établissons que cette procédure minimise les informations de Kullback lorsque le processus des observations est à trajectoires continues à droite et admettant une limite à gauche, puis par extension, nous montrons un résultat analogue pour des observations non nécessairement indépendantes et identiquement distribuées en temps discret. Une autre partie de ce travail est consacrée à des résultats d'optimalité asymptotique du test de Wald (on fait tendre les risques vers zéro) par rapport aux moments d'ordre r de la durée d'observation, pour r un nombre réel strictement positif, mais également par rapport à des fonctionnelles plus générales.

Abstract

This thesis is related to the topic of Sequential Analysis, a statistical theory based on the randomness of the number of observations taken into account to test hypotheses, determine confidence intervals for parameters, etc. We more precisely deal with optimality properties of the sequential probability ratio test or Wald test. Thus we establish that this procedure minimizes the Kullback informations when the observed process is right-continuous and has limits from the left, and by extension, we show that this result also holds for nonnecessarily independent and identically distributed observations in discrete time. Another part of this work is devoted to asymptotic optimality properties of the Wald test (as the error probabilities tend to zero) with respect to the r^{th} moments of observation times, where r denotes some positive real number, but also with respect to more general functionals.

N.B. Les personnes souhaitant obtenir les figures incluses dans le manuscrit peuvent me le faire savoir à l'adresse ci-dessus.

Table des matières

Chapitre 1

Introduction générale

Dans son ouvrage intitulé *Sequential Analysis* ([Wa] en abrégé dans la suite), Abraham Wald pose les fondements d'une théorie proposant une alternative, mais aussi et surtout un complément au point de vue que nous qualifierons de *classique* de Neyman et Pearson. Cette monographie est entièrement consacrée aux tests séquentiels d'hypothèses. Outre la définition du *test séquentiel du rapport des probabilités* figurent une comparaison entre le nombre moyen d'observations nécessaires pour ce test et celui déterminé par la théorie de Neyman et Pearson en fonction des risques de première et de deuxième espèce (ceci amènera A. Wald à suggérer l'optimalité des tests séquentiels par rapport aux durées moyennes d'observation), mais aussi de nombreuses applications à des tests usuels. Nous commençons par rappeler la définition du test séquentiel du rapport des probabilités ainsi que la propriété d'optimalité énoncée par Wald (cf. [Wa]).

1.1 Le test séquentiel du rapport des probabilités ou test de Wald

On considère X une variable aléatoire (v.a.) générique et une suite $(X_n)_{n\geq 1}$ de copies indépendantes de X. On désigne par $f(x;\theta)$ soit la densité de probabilité de X, soit la probabilité que X prenne la valeur x, selon que la loi de X admet une densité ou est à support discret (θ désignant un paramètre inconnu). Nous souhaitons tester une hypothèse nulle $H_0 \ : \ \theta = \theta_0$ contre une hypothèse alternative $H_1 \ : \ \theta = \theta_1$ pour θ_0 et θ_1 deux valeurs fixées.
La vraisemblance d'un n-échantillon $(X_1, \cdots, X_n)$ de la loi de X est donnée par :

$$p_{0n} = p_{0n}(x_1, \cdots, x_n) = f(x_1;\theta_0)\cdots f(x_n;\theta_0)$$

si H_0 est vraie, et par :

$$p_{1n} = p_{1n}(x_1, \cdots, x_n) = f(x_1;\theta_1)\cdots f(x_n;\theta_1)$$

si H_1 est vraie, où $(x_1, \cdots, x_n)$ désignent les réalisations de $(X_1, \cdots, X_n)$.
Le *test séquentiel du rapport des probabilités* est alors défini comme suit :

– on choisit deux constantes $0 < A \leq 1 \leq B < +\infty \quad , \quad (A < B)$

– pour chaque entier naturel n on calcule le rapport des probabilités $R_n = \dfrac{p_{1n}}{p_{0n}}$

– si $R_n \in]A, B[$, on choisit de faire une observation supplémentaire, sinon on arrête les observations et on prend une décision :

- si $R_n \leq A$, on accepte l'hypothèse H_0
- si $R_n \geq B$, on rejette l'hypothèse H_0.

En d'autres termes, le test de Wald de bornes A et B est défini par :

– le temps d'arrêt $\tau_{AB} = \inf\{n \geq 1 : R_n \notin]A, B[\}$

– la fonction de décision $d_{AB} = \begin{cases} 0 & \text{si} \quad R_{\tau_{AB}} \leq A \\ 1 & \text{si} \quad R_{\tau_{AB}} \geq B \end{cases}.$

Nous noterons $\mathbf{P}^0$ (respectivement $\mathbf{P}^1$) la loi de la suite $(X_n)_{n\geq 1}$ lorsque H_0 (respectivement H_1) est vraie ; $\mathbf{E}_0$ et $\mathbf{E}_1$ désignant alors les espérances correspondantes. Les risques de première et de deuxième espèce,

$$\alpha = \mathbf{P}^0(d_{AB} = 1) \qquad \text{et} \qquad \beta = \mathbf{P}^1(d_{AB} = 0) \ \ ,$$

sont liés aux bornes A et B par les inégalités :

$$A \geq \frac{\beta}{1-\alpha} \qquad \text{et} \qquad B \leq \frac{1-\beta}{\alpha} \ .$$

Désignant par $\Delta_{\alpha\beta}$ la classe des règles de décision $\delta = (\tau, d)$ vérifiant les conditions suivantes :

(A) les risques de première et de deuxième espèce $\mathbf{P}^0(d = 1)$ *et* $\mathbf{P}^1(d = 0)$ *sont respectivement majorés par* α *et* β*,*

(B) les nombres moyens d'observations nécessaires $\mathbf{E}_0\tau$ *et* $\mathbf{E}_1\tau$ *sont finis,*

on peut montrer que le test de Wald (τ_{AB}, d_{AB}) minimise les durées moyennes d'observation sous les deux hypothèses :

$$\forall\ (\tau, d) \in \Delta_{\alpha\beta}\ , \quad \mathbf{E}_0\tau \geq \mathbf{E}_0\tau_{AB} \qquad \text{et} \qquad \mathbf{E}_1\tau \geq \mathbf{E}_1\tau_{AB}\ .$$

Le lecteur pourra par exemple se référer aux travaux sur le sujet de Wald ([Wa]), Wald et Wolfowitz ([Wa-Wo]), Lehmann ([Le]), Shiryaev ([Sh1]).

1.2 Cas des processus en temps continu

Nous considérons une fois de plus le cas d'un test de deux hypothèses simples. A.I. Yashin a établi l'optimalité du test de Wald par rapport aux informations de Kullback lorsque les deux mesures $\mathbf{P}^0$ et $\mathbf{P}^1$ sont localement équivalentes (ce qui revient à dire que leurs restrictions $\mathbf{P}_t^0$ et $\mathbf{P}_t^1$ à chacune des sous-tribus d'une filtration $(\mathcal{F}_t)_{t\geq 0}$ sont équivalentes) mais singulières sur la tribu terminale $\mathcal{F}$. L'auteur suppose de plus que les processus de densités $\frac{d\mathbf{P}_t^1}{d\mathbf{P}_t^0}$ et $\frac{d\mathbf{P}_t^0}{d\mathbf{P}_t^1}$ sont à trajectoires presque sûrement continues. Si $\Delta_{\alpha\beta}$ désigne la classe des tests tels que les risques de première et de deuxième espèce sont respectivement majorés par α et β (il y a également une condition sur les temps d'arrêt que nous n'explicitons pas ici), alors on a le résultat suivant (cf. [Ya]) :

Théorème 1.1 *Pour tous α et β strictement positifs et tels que $\alpha + \beta < 1$, il existe dans $\Delta_{\alpha\beta}$ une règle de décision (τ_{AB}, d_{AB}) qui minimise simultanément les informations de Kullback $\mathbf{E}_0 \ln \frac{d\mathbf{P}_\tau^0}{d\mathbf{P}_\tau^1}$ et $\mathbf{E}_1 \ln \frac{d\mathbf{P}_\tau^1}{d\mathbf{P}_\tau^0}$ parmi tous les tests appartenant à $\Delta_{\alpha\beta}$. Elle est donnée par :*

- *le temps d'arrêt* $\tau_{AB} = \inf\left\{t \geq 0 : \frac{d\mathbf{P}_t^1}{d\mathbf{P}_t^0} \notin]A, B[\right\}$
- *la fonction de décision* $d_{AB} = \begin{cases} 0 & \text{si} \quad \frac{d\mathbf{P}_{\tau_{AB}}^1}{d\mathbf{P}_{\tau_{AB}}^0} \leq A \\ 1 & \text{si} \quad \frac{d\mathbf{P}_{\tau_{AB}}^1}{d\mathbf{P}_{\tau_{AB}}^0} \geq B \ , \end{cases}$.

où les bornes A et B sont liées aux risques α et β par les égalités :

$$A = \frac{\beta}{1-\alpha} \qquad \text{et} \qquad B = \frac{1-\beta}{\alpha} \ .$$

De plus,

$$\mathbf{E}_0 \ln \frac{d\mathbf{P}_{\tau_{AB}}^0}{d\mathbf{P}_{\tau_{AB}}^1} = \omega(\alpha, \beta) \qquad \text{et} \qquad \mathbf{E}_1 \ln \frac{d\mathbf{P}_{\tau_{AB}}^1}{d\mathbf{P}_{\tau_{AB}}^0} = \omega(\beta, \alpha)$$

où $\omega(x, y) = x \ln \frac{x}{1-y} + (1-x) \ln \frac{1-x}{y}$.

La démonstration de ce résultat repose sur une utilisation judicieuse de l'inégalité de Jensen conditionnelle. Il est à noter que, contrairement au cas discret, on peut fixer α et β puis définir les bornes A et B du test de Wald dont les risques sont précisément égaux à α et β (voir le contre-exemple figurant dans [Sh1], Remarque 3, p. 141).

Un résultat d'un type différent a été obtenu par Irle et Schmitz ([Ir-Sc]). Les auteurs supposent que les processus de densités sont à accroissements indépendants et stationnaires et continus en probabilité sous $\mathbf{P}^0$ et $\mathbf{P}^1$. Ils établissent alors l'optimalité du test de Wald par rapport aux durées moyennes parmi la classe des règles de décision dont les risques sont bornés par α et β respectivement.

Lorsque g désigne une fonction convexe satisfaisant à certaines conditions non restrictives, le premier de ces auteurs a établi l'optimalité du test de Wald (τ_{AB}, d_{AB}) par rapport aux fonctionnelles $\mathbf{E}_0 g\Big(\frac{d\mathbf{P}^1_\tau}{d\mathbf{P}^0_\tau}\Big)$ et $\mathbf{E}_1 g\Big(\frac{d\mathbf{P}^0_\tau}{d\mathbf{P}^1_\tau}\Big)$ pour peu que $\frac{d\mathbf{P}^1_{\tau_{AB}}}{d\mathbf{P}^0_{\tau_{AB}}}$ appartienne à $\{A, B\}$ presque sûrement (voir [Ir1]).

C'est dans ce cadre que s'insère le Chapitre 2 de la présente thèse. Nous donnons une démonstration basée sur les théories de la statitique bayésienne et de l'arrêt optimal d'un résultat d'optimalité du test de Wald par rapport aux informations de Kullback (Théorème 2.1) lorsque les processus de densités $\frac{d\mathbf{P}^1_t}{d\mathbf{P}^0_t}$ et $\frac{d\mathbf{P}^0_t}{d\mathbf{P}^1_t}$ sont à trajectoires càdlàg et si la condition

$$\frac{d\mathbf{P}^1_{\tau_{AB}}}{d\mathbf{P}^0_{\tau_{AB}}} \in \{A, B\} \quad \textit{presque sûrement}$$

est satisfaite. Le même raisonnement nous permet de déduire des résultats analogues lorsque le processus des *observations* est à accroissements indépendants ou encore une suite de variables aléatoires non nécessairement indépendantes et identiquement distribuées (Théorème 2.4).

1.3 Propriétés d'optimalité asymptotique

Il semblerait que, outre dans le cas considéré dans [Ir-Sc], l'optimalité *exacte* par rapport aux durées moyennes soit délicate à obtenir si l'on n'impose pas au processus de densité d'atteindre les bornes du test de Wald. Une façon de s'affranchir de cette condition est de faire tendre les risques vers zéro et donc les bornes du test de Wald vers zéro et l'infini respectivement. On pourra alors espérer que l'effet d'un saut de la densité au-delà de l'une des bornes sera atténué.

Divers auteurs (voir [Dr-No], [Ir2], [Ir3], [La2] par exemple) ont démontré des propriétés d'*optimalité asymptotique* du test de Wald par rapport à des fonctionnnelles diverses dont les moments d'ordre r de la durée d'observation dans le cas discret (cf. [La2]). Irle a appliqué ce résultat à des processus régénératifs, fournissant ainsi (à notre connaissance du moins) les premiers exemples d'optimalité du test de Wald pour des v.a. dépendantes : les processus de Harris et les chaînes de Markov à espace d'états fini (cf. [Ir3]).

Nous donnons au Chapitre 3 des extensions des résultats de [La2] au cas des tests de m hypothèses simples, $m \geq 2$. Nous proposons également un autre exemple de fonctionnelle minimisée par le test de Wald (dans le contexte asymptotique). Ces résultats font l'objet des Théorèmes 3.2 et 3.3, ainsi que des Corollaires 3.1, 3.2 et 3.3.

1.4 Analyse séquentielle et théorie de l'information

En guise de conclusion à ce préambule, nous souhaiterions essayer de donner une interprétation intuitive (naturelle ?) de la minimisation des informations de Kullback dans le cas de processus en temps continu, mais aussi en temps discret.

Dans son livre *Information Theory and Statistics*, Solomon Kullback rend l'hommage suivant à Abraham Wald : "*Although Wald [...] did not explicitely mention information in his treatment of sequential analysis, it should be noted that his work must be considered a major contribution to the statistical applications of information theory.*" Nous souhaitons préciser les liens entre l'analyse séquentielle (et plus précisément le test de Wald) et l'information de Kullback. Il nous paraît en effet intéressant de reconsidérer le test de Wald et ses propriétés d'optimalité sous l'éclairage fourni par la théorie de l'information, car si dans le cas de suites de v.a. indépendantes et identiquement distribuées (i.i.d.) l'optimalité est établie par rapport aux durées moyennes, elle a en revanche lieu par rapport aux informations de Kullback dans le cas où le processus des observations est en temps continu. De surcroît, on peut montrer sans peine que les informations de Kullback sont également minimisées pour des observations formant une suite de variables aléatoires i.i.d.

On considère les espaces probabilisés $(\mathcal{X}, \mathcal{A}, \mathbf{P}^i)$, $i = 0, 1$. Supposons les mesures de probabilité $\mathbf{P}^0$ et $\mathbf{P}^1$ équivalentes sur $(\mathcal{X}, \mathcal{A})$ et si λ désigne une dominante commune, on pose $f_i = \dfrac{d\mathbf{P}^i}{d\lambda}$, $i = 0, 1$, les dérivées de Radon-Nikodym.

Si $X_1, \cdots, X_n, \cdots$ est une suite de variables aléatoires réelles i.i.d., on pose $\mathcal{X} = \mathbf{R}^n$ l'espace d'épreuves du n-échantillon $(X_1, \cdots, X_n)$ et nous supposons $\mathcal{X}$ partitionné en $\mathcal{X} = E_0 \cup E_1$. Soit une procédure de test telle que si la réalisation $x = (x_1, \cdots, x_n)$ appartient à E_i on accepte l'hypothèse H_i , $i = 0, 1$. Les probabilités d'erreur sont définies par $\alpha = \mathbf{P}^0(E_1)$ et $\beta = \mathbf{P}^1(E_0)$.

L'information de Kullback de H_1 par rapport à H_0 pour le n-échantillon $(X_1, \cdots, X_n)$ est donnée par :

$$I_n(1,0) := \int_{\mathcal{X}} \ln \frac{f_1(x)}{f_0(x)} d\mathbf{P}^1(x) = \int_{\mathcal{X}} \ln \frac{P(H_1|x)}{P(H_0|x)} d\mathbf{P}^1(x) - \ln \frac{P(H_1)}{P(H_0)}$$

où $P(H_i)$ est la probabilité *a priori* de l'hypothèse H_i et $P(H_i|x)$ est la probabilité *a posteriori* de l'hypothèse H_i lorsque l'on a observé x (voir également le lien entre la propriété d'optimalité et la minimisation de risque bayésien établi au paragraphe 3 du Chapitre 2).

On définit de même l'information de Kullback de H_0 par rapport à H_1 pour le n-échantillon $(X_1, \cdots, X_n)$ par :

$$I_n(0,1) := \int_{\mathcal{X}} \ln \frac{f_0(x)}{f_1(x)} d\mathbf{P}^0(x) = \int_{\mathcal{X}} \ln \frac{P(H_0|x)}{P(H_1|x)} d\mathbf{P}^0(x) - \ln \frac{P(H_0)}{P(H_1)} \ .$$

Remarque 1.1 *Ces quantités sont positives (inégalité de Jensen).*

Remarque 1.2 *Posant* $J_n(0,1) = I_n(0,1) + I_n(1,0)$, *on a :*

$$\begin{aligned} J_n(0,1) &= \int_{\mathcal{X}} (f_1(x) - f_0(x)) \ln \frac{f_1(x)}{f_0(x)} d\lambda(x) \\ &= \int_{\mathcal{X}} \ln \frac{P(H_1|x)}{P(H_0|x)} d\mathbf{P}^1(x) - \int_{\mathcal{X}} \ln \frac{P(H_1|x)}{P(H_0|x)} d\mathbf{P}^0(x) \ . \end{aligned}$$

C'est une mesure de la divergence *entre les hypothèses H_0 et H_1 (ou bien entre les mesures $\mathbf{P}^0$ et $\mathbf{P}^1$) et constitue une mesure de la difficulté à les discerner. La divergence possède presque toutes les propriétés d'une distance ; il n'y a toutefois pas d'inégalité du triangle.*

Remarque 1.3 *Toutes ces fonctionnelles sont additives :*

$$\begin{aligned} I_n(0,1) &= I_{n-1}(0,1) + I_1(0,1) = nI_1(0,1) \\ I_n(1,0) &= I_{n-1}(1,0) + I_1(1,0) = nI_1(1,0) \\ J_n(0,1) &= J_{n-1}(0,1) + J_1(0,1) = nJ_1(0,1) \\ J_n(1,0) &= J_{n-1}(1,0) + J_1(1,0) = nJ_1(1,0) \end{aligned}$$

et la divergence est symétrique.

Montrons à présent l'optimalité du test de Wald par rapport aux informations de Kullback. Nous rappelons à cet effet l'*identité de Wald :*

Lemme 1.1 *Soient $Z_1, \cdots, Z_n, \cdots$ des variables aléatoires i.i.d. telles que $\mathbf{E}Z_1 < +\infty$. On pose $S_n = Z_1 + \cdots + Z_n$ et $\mathcal{F}_n^Z$ la σ-algèbre engendrée par $\{Z_1, \cdots, Z_n\}$. Soit T un $(\mathcal{F}_n^Z)_{n\geq 1}$-temps d'arrêt tel que $\mathbf{E}T < +\infty$. Alors :*

$$\mathbf{E}S_T = \mathbf{E}(Z_1 + \cdots + Z_T) = \mathbf{E}Z_1 \,.\, \mathbf{E}T \ .$$

Dans le cas présent, le rapport des probabilités est donné par :

$$R_n = \prod_{i=1}^{n} \frac{d\mathbf{P}^1}{d\mathbf{P}^0}(X_i) \quad ,$$

si bien que les v.a. $\ln R_n$ et $\ln \dfrac{1}{R_n}$ peuvent être représentées comme des sommes de variables aléatoires i.i.d. :

$$\begin{aligned} \ln R_n &= \sum_{i=1}^{n} \ln \frac{d\mathbf{P}^1}{d\mathbf{P}^0}(X_i) = \sum_{i=1}^{n} Z_i \\ \ln \frac{1}{R_n} &= \sum_{i=1}^{n} \ln \frac{d\mathbf{P}^0}{d\mathbf{P}^1}(X_i) = \sum_{i=1}^{n} (-Z_i) \quad . \end{aligned}$$

De plus, si $\lambda\{x : \mathbf{P}^1(x) \neq \mathbf{P}^0(x)\} > 0$, on a :

$$0 < \mathbf{E}_1(Z_1) < +\infty \qquad \text{et} \qquad 0 < \mathbf{E}_0(-Z_1) < +\infty$$

si bien que l'identité de Wald fournit alors :

$$\mathbf{E}_1 \ln \frac{d\mathbf{P}^1_\tau}{d\mathbf{P}^0_\tau} = \mathbf{E}_1 Z_1 . \mathbf{E}_1 \tau$$

et

$$\mathbf{E}_0 \ln \frac{d\mathbf{P}^0_\tau}{d\mathbf{P}^1_\tau} = \mathbf{E}_0(-Z_1) \,.\, \mathbf{E}_0 \tau$$

pour tout temps d'arrêt τ d'espérance finie sous les deux probabilités.

Remarque 1.4 *Un raisonnement analogue montre que l'on peut minimiser toute fonctionnelle $\mathbf{E}_i D_\tau$ telle que D_n puisse être représentée comme une somme de variables aléatoires i.i.d. intégrables par rapport à $\mathbf{P}^i$, $i = 0, 1$. Ceci est en fait raisonnable, car les fonctionnelles à minimiser doivent posséder les propriétés d'une fonction de coût (cf. théorie de la statistique bayésienne).*

Nous venons ainsi de montrer que minimiser les durées moyennes ou les informations de Kullback est équivalent dans le cas de variables aléatoires i.i.d. Il nous semble toutefois que ce ne soit pas systématiquement le cas. En effet, si les fonctions

$$t \longmapsto \mathbf{E}_0 \ln \frac{d\mathbf{P}^0_t}{d\mathbf{P}^1_t} \qquad \text{et} \qquad t \longmapsto \mathbf{E}_1 \ln \frac{d\mathbf{P}^1_t}{d\mathbf{P}^0_t}$$

sont croissantes et si la propriété [1] :

$$\sigma \leq \tau \quad \Longrightarrow \mathbf{E}_0 \ln \frac{d\mathbf{P}^0_\sigma}{d\mathbf{P}^1_\sigma} \leq \mathbf{E}_0 \ln \frac{d\mathbf{P}^0_\tau}{d\mathbf{P}^1_\tau} \quad \text{et} \quad \mathbf{E}_1 \ln \frac{d\mathbf{P}^1_\sigma}{d\mathbf{P}^0_\sigma} \leq \mathbf{E}_1 \ln \frac{d\mathbf{P}^1_\tau}{d\mathbf{P}^0_\tau}$$

a lieu pour tous temps d'arrêt σ et τ, cela ne signifie pas pour autant que les informations de Kullback sont des fonctions croissantes des *durées moyennes*, c'est-à-dire que minimiser ces deux quantités ne revient pas au même dans le cas général.

On voit bien quel peut être l'intérêt de minimiser les durées moyennes d'observation (minimiser le coût lié aux observations) ; mais si l'on décide de rétribuer

[1] Cette propriété ne résulte pas du théorème d'arrêt des martingales, puisque les temps d'arrêt que nous considérons ne sont pas bornés et que les processus de densités locales ne sont pas uniformément intégrables (car les mesures $\mathbf{P}^0$ et $\mathbf{P}^1$ sont singulières sur la tribu terminale). Elle provient du fait que les densités sont des surmartingales positives continues à droite, donc vérifient :

$$\frac{d\mathbf{P}^1_\sigma}{d\mathbf{P}^0_\sigma} \geq \mathbf{E}_0\Big[\frac{d\mathbf{P}^1_\tau}{d\mathbf{P}^0_\tau} \,|\, \mathcal{F}_\sigma\Big] \quad \text{et} \quad \frac{d\mathbf{P}^0_\sigma}{d\mathbf{P}^1_\sigma} \geq \mathbf{E}_1\Big[\frac{d\mathbf{P}^0_\tau}{d\mathbf{P}^1_\tau} \,|\, \mathcal{F}_\sigma\Big]$$

pour tous temps d'arrêt $\sigma \leq \tau$ (voir [Re-Yo], Théorème 3.3, p. 66). On conclut à l'aide de l'inégalité de Jensen conditionnelle et en passant aux espérances.

l'"information" apportée par une observation non pas à un prix constant mais selon sa valeur, c'est-à-dire la "quantité d'information" fournie, il devient alors raisonnable de chercher à minimiser des fonctionnelles telles que l'information de Kullback. La propriété d'optimalité se traduira simplement par le fait qu'un test permettra de conclure en utilisant le moins d'information possible (les risques étant bornés par ceux du test de Wald).

L'information de Kullback, ou plutôt la divergence, est une mesure de la "distance" entre deux hypothèses, entre deux mesures de probabilité. S'il existe un test (τ, d) tel que l'on ait :

$$\mathbf{E}_0 \ln \frac{d\mathbf{P}^0_\tau}{d\mathbf{P}^1_\tau} < \mathbf{E}_0 \ln \frac{d\mathbf{P}^0_{\tau_{AB}}}{d\mathbf{P}^1_{\tau_{AB}}} \quad \text{et} \quad \mathbf{E}_1 \ln \frac{d\mathbf{P}^1_\tau}{d\mathbf{P}^0_\tau} < \mathbf{E}_1 \ln \frac{d\mathbf{P}^1_{\tau_{AB}}}{d\mathbf{P}^0_{\tau_{AB}}}$$

pour (τ_{AB}, d_{AB}) le test de Wald, alors les hypothèses H_0 et H_1 sont moins discernables à l'instant τ qu'à l'instant τ_{AB}, au sens de la divergence. Les risques d'erreur en sont ainsi amplifiés :

$$\mathbf{P}^0(d=1) > \mathbf{P}^0(d_{AB}=1) = \alpha \qquad \text{et} \qquad \mathbf{P}^1(d=0) > \mathbf{P}^1(d_{AB}=0) = \beta\ ,$$

si bien que (τ, d) n'appartient plus à la classe des tests considérée. Nous pouvons donc interpréter τ_{AB} comme étant le premier instant auquel il est possible de discerner les hypothèses H_0 et H_1, pour des probabilités d'erreur α et β fixées.

Références bibliographiques

[Dr-No] V. P. Dragalin et A. A. Novikov. *Adaptive sequential tests for composite hypotheses.* A paraître dans Theory of Probability and its Applications.

[Ir1] A. Irle. *1984 Extended optimality of sequential probability ratio tests.* Annals of Statistics, vol. 12, no. 1, pp. 380-386.

[Ir2] A. Irle. *1990 Asymptotic optimality of general sequential probability ratio tests.* Scandinavian Journal of Statistics, vol. 17, pp. 321-332.

[Ir3] A. Irle. *1993 r-quick convergence for regenerative processes with applications to sequential analysis.* Stochastic Processes and their Applications, vol. 45, pp. 319-329.

[Ir-Sc] A. Irle et N. Schmitz. *1984 On the optimality of the SPRT for processes with continuous time parameter.* Math. Operationsforsch. u. Statist., ser. Statist., vol. 15, no. 2, pp. 91-104.

[Ku] S. Kullback. *1968 Information theory and statistics.* Dover.

[La2] T. L. Lai. *1981 Asymptotic optimality of invariant sequential probability ratio tests.* Annals of Statistics, vol. 9, no.2, pp. 318-333.

[Le] E. L. Lehmann. *1959 Testing statistical hypotheses.* Wiley, New York.

[Re-Yo] D. Revuz et M. Yor. *1991 Continuous martingales and brownian motion.* Springer Verlag.

[Sh1] A. N. Shiryaev. *1973 Statistical sequential analysis (Translations of Mathematical Monographs, vol. 38).* A.M.S., Providence.

[Sh2] A. N. Shiryaev. *1978 Optimal stopping rules.* Springer Verlag.

[Wa] A. Wald. *1947 Sequential analysis.* Wiley, New York.

[Wa-Wo] A. Wald et J. Wolfowitz. *1948 Optimum character of the sequential probability ratio test.* Annals of Mathematical Statistics, vol. 19, pp. 326-339.

[Ya] A. I. Yashin. *1983 On a problem of sequential hypothesis testing.* Theory of Probability and its Applications, vol. 28, no. 1, pp. 157-165.

Chapitre 2

Une procédure optimale pour tester la loi d'un processus càdlàg

2.1 Introduction

On se place sur l'espace $\Omega = \mathcal{D}([0,+\infty[)$ des fonctions continues à droite et admettant une limite à gauche (*càdlàg*) que l'on munit de la filtration $(\mathcal{F}_t)_{t\geq 0}$, où $\mathcal{F}_t = \sigma\{x : x_s \ , \ s \leq t\}$. Nous posons $\mathcal{F} = \bigvee_t \mathcal{F}_t$ et supposons $(\Omega, \mathcal{F})$ muni de deux mesures de probabilité $\mathbf{P}^0$ et $\mathbf{P}^1$.

Considérons la filtration $(\mathcal{G}_t)_{t\geq 0} = (\mathcal{F}^{\mathbf{Q}}_{t+})_{t\geq 0}$, qui est continue à droite et complète pour la mesure dominante $\mathbf{Q} = \frac{1}{2}(\mathbf{P}^0 + \mathbf{P}^1)$, ainsi que la tribu $\mathcal{G} = \bigvee_t \mathcal{G}_t$. Nous supposons que les mesures $\mathbf{P}^0$ et $\mathbf{P}^1$ s'étendent à $(\Omega, \mathcal{G})$ et désignons respectivement par $\mathbf{P}^0_t$ et $\mathbf{P}^1_t$ leurs restrictions à $\mathcal{G}_t$; le processus canonique étant noté X.

On observe le processus $(X_t, \mathcal{G}_t, \mathbf{P})_{t\geq 0}$, pour $\mathbf{P}$ une mesure de probabilité sur $(\Omega, \mathcal{G})$, et on souhaite savoir laquelle de $\mathbf{P}^0$ et $\mathbf{P}^1$ est la vraie loi de X. Pour ce faire, nous construisons une procédure séquentielle pour le test de :

$$H_0 : \text{la loi de } X \text{ est } \mathbf{P}^0 \quad \text{contre} \quad H_1 : \text{la loi de } X \text{ est } \mathbf{P}^1,$$

et établissons l'optimalité de cette règle de décision par rapport aux *informations de Kullback* au sein d'une certaine classe de règles de décision que nous définirons ultérieurement.

Avant de donner des conditions sous lesquelles l'optimalité a lieu, nous rappelons brièvement quelques résultats généraux liés aux *processus de densités locales* (cf. [Ka-Li-Sh2]).

1) Si $\mathbf{P}^0_t \sim \mathbf{P}^1_t$, $t \geq 0$, alors il existe des martingales locales $(Z^0_t\,,\,\mathcal{G}_t\,,\,\mathbf{P^0})_{t\geq 0}$ et $(Z^1_t\,,\,\mathcal{G}_t\,,\,\mathbf{P^1})_{t\geq 0}$ telles que $Z^0_t = \dfrac{d\mathbf{P}^1_t}{d\mathbf{P}^0_t}$ et $Z^1_t = \dfrac{d\mathbf{P}^0_t}{d\mathbf{P}^1_t}$: ce sont les processus de densités locales. On pose

$$M^i_t = \int_0^t Z^{i\oplus}_{s-}\, dZ^i_s \quad , \quad \text{où} \quad a^{\oplus} = \begin{cases} a^{-1} & \text{si} \quad a \neq 0 \\ 0 & \text{si} \quad a = 0 \end{cases} ,$$

2) les processus $M^i = (M^i_t, \mathcal{G}_t, \mathbf{P}^i)_{t\geq 0}, i = 0, 1$, sont des martingales locales,

3) les densités locales Z^0 et Z^1 sont respectivement solution des équations différentielles stochastiques (E.D.S.)

$$Z^i_t = Z^i_0 + \int_0^t Z^i_{s-}\; dM^i_s \ , \ i = 0, 1,$$

4) les limites $\lim_{t\to+\infty} Z^0_t$ et $\lim_{t\to+\infty} Z^1_t$ existent $\mathbf{Q}$-p.s. sur $[0,+\infty]$ (on les note Z^0_∞ et Z^1_∞ respectivement),

5) Pour tout $A \in \mathcal{G}$, on a les *décompositions de Lebesgue* :

$$\mathbf{P}^1(A) = \int_A Z^0_\infty \, d\mathbf{P}^0 + \mathbf{P}^1(A \cap \{Z^0_\infty = +\infty\}) \ ,$$

$$\mathbf{P}^0(A) = \int_A Z^1_\infty \, d\mathbf{P}^1 + \mathbf{P}^0(A \cap \{Z^1_\infty = +\infty\}) \ .$$

Soit M^{ic} la composante continue de M^i , $\langle M^{ic}, M^{ic}\rangle$ sa variation quadratique, μ^{M^i} la mesure associée aux sauts de M^i et ν_i le compensateur de μ^{M^i} par rapport à $\mathbf{P}^i$, $i = 0, 1$. On pose

$$B_t^i := B_t^i(M^i) := \langle M^{ic}, M^{ic}\rangle_t + \int_0^t \int_{\mathbf{R}\setminus\{0\}} \frac{x^2}{1+|x|}\, d\nu_i\ ,\ t < +\infty$$

et

$$B_\infty^i := B_\infty^i(M^i) := \lim_{t\to+\infty} B_t^i(M^i)\ ,\ i = 0, 1.$$

Nous supposons qu'à l'instant 0, les deux mesures de probabilité $\mathbf{P}^0$ et $\mathbf{P}^1$ coïncident :

I.*a*. $\mathbf{P}_0^0 = \mathbf{P}_0^1$.

Le Théorème 1 de [Ka-Li-Sh2] assure alors que si la condition

I.*b*. $\mathbf{P}_t^0 \sim \mathbf{P}_t^1\ ,\ t > 0$

est satisfaite, alors les trois assertions suivantes sont équivalentes :

II.*a*. $\mathbf{P}^1\{B_\infty^0 = +\infty\} = 1$

II.*b*. $\mathbf{P}^0\{B_\infty^1 = +\infty\} = 1$.

$\mathbf{P}^0 \perp \mathbf{P}^1$.

2.2 Enoncé du résultat principal

Définition 2.1 *Un couple $\delta = (\tau, d)$ est appelé* règle de décision *si :*

(i) τ est un $(\mathcal{G}_t)_{t\geq 0}$-temps d'arrêt

(ii) d est une variable aléatoire $\mathcal{G}_\tau$-mesurable, à valeurs dans $\{0,1\}$.

On désigne par $\alpha(\delta) = \mathbf{P}^0(d=1)$ et $\beta(\delta) = \mathbf{P}^1(d=0)$ les risques de première et de deuxième espèce respectivement.

Pour $0 < a \leq 1 \leq b < +\infty$ $(a < b)$ deux réels fixés, on définit le *test séquentiel de bornes* a et b, noté $\delta_{ab} = (\tau_{ab}, d_{ab})$, par :

$$\tau_{ab} = \inf\{t \geq 0 : Z_t^0 \notin]a,b[\} \quad , \quad d_{ab} = \begin{cases} 0 & \text{si} \quad Z^0_{\tau_{ab}} \leq a \\ 1 & \text{si} \quad Z^0_{\tau_{ab}} \geq b \end{cases} \quad .$$

Soient α et β deux nombres strictement positifs tels que $\alpha + \beta < 1$. On désigne par $\Delta_{\alpha\beta}$ la classe des règles de décision $\delta = (\tau, d)$ satisfaisant aux deux conditions suivantes :

(A) $\alpha(\delta) = \mathbf{P}^0(d=1) \leq \alpha$ et $\beta(\delta) = \mathbf{P}^1(d=0) \leq \beta$

(B) les informations de Kullback $\mathbf{E}_0 \ln Z_\tau^1$ et $\mathbf{E}_1 \ln Z_\tau^0$ sont finies et strictement positives (dans le cas de variables aléatoires indépendantes et identiquement distribuées, la condition de positivité traduit simplement le fait que la loi de X_1 sous H_0 est différente de sa loi sous H_1).

Nous énonçons à présent le résultat principal de cette partie.

Théorème 2.1 *Soit $\delta_{AB} = (\tau_{AB}, d_{AB})$ le test séquentiel de bornes A et B, avec $0 < A \leq 1 \leq B < +\infty$, et telles que $Z^0_{\tau_{AB}} \in \{A, B\}$ $\mathbf{Q}-p.s.$ On désigne respectivement par α et β les risques de première et deuxième espèce qui lui sont associés. Si les conditions I et II sont satisfaites, alors pour toute procédure $\delta = (\tau, d) \in \Delta_{\alpha\beta}$, on a les inégalités :*

$$\mathbf{E}_0 \ln Z^1_{\tau_{AB}} \leq \mathbf{E}_0 \ln Z^1_\tau \quad \text{et} \quad \mathbf{E}_1 \ln Z^0_{\tau_{AB}} \leq \mathbf{E}_1 \ln Z^0_\tau \ .$$

On a également les relations :

$$A = \frac{\beta}{1-\alpha} \ , \ B = \frac{1-\beta}{\alpha}$$

et

$$\mathbf{E}_0 \ln Z^1_{\tau_{AB}} = \omega(\alpha, \beta) \ , \ \mathbf{E}_1 \ln Z^0_{\tau_{AB}} = \omega(\beta, \alpha) \ ,$$

où

$$\omega(x,y) = x \ln \frac{x}{1-y} + (1-x) \ln \frac{1-x}{y} \ .$$

Des résultats analogues ont été obtenus dans le cas où les processus Z^i , $i = 0, 1$, sont des processus à trajectoires continues (cf. [Ya]) ou encore à accroissements indépendants et continus en probabilité (voir [Ir-Sc]).

Ce théorème s'applique par exemple lorsque la densité Z^0 vérifie la condition suivante :

On suppose qu'il existe un intervalle $[A', B'] \subset [A, B]$ *tel que les trajectoires de* Z^0 *sont* $\mathbf{Q}$*-presque sûrement continues dans* $[A, A']$ *et* $[B', B]$ *et que les sauts de* Z^0 *à l'intérieur de* $[A', B']$ *sont bornés par* $\Delta = \min\{B - B' , A' - A\}$*. Ceci entraîne alors que* Z^0 *ne pourra sortir de l'intervalle* $[A, B]$ *que de façon continue, si bien que* $Z^0_{\tau_{AB}} \in \{A, B\}$ $\mathbf{Q} - p.s.$

Avant d'aborder la démonstration de ce théorème, à la fois basée sur la théorie de la décision bayésienne et sur celle de l'arrêt optimal, nous donnons une première propriété du temps d'arrêt défini plus haut.

Lemme 2.1 (cf. [Ya]) *On suppose les conditions I et II satisfaites. Alors le temps d'arrêt* τ_{AB} *est presque sûrement fini sous les deux probabilités.*

Démonstration. Il est facile de voir que $\mathbf{P}^0 \perp \mathbf{P}^1 \iff \mathbf{P}^1\{Z^0_\infty = +\infty\} = 1 \iff \mathbf{P}^0\{Z^0_\infty = 0\} = 1 \iff \mathbf{P}^0\{Z^1_\infty = +\infty\} = 1 \iff \mathbf{P}^1\{Z^1_\infty = 0\} = 1$.

Donc $\lim_{t\to+\infty} Z^0_t = +\infty \quad \mathbf{P}^1-$p.s. et $\lim_{t\to+\infty} Z^0_t = 0 \quad \mathbf{P}^0-$p.s.

Ceci entraîne que $\mathbf{P}^1\{\tau_{AB} < +\infty\} = 1$ et $\mathbf{P}^0\{\tau_{AB} < +\infty\} = 1$.

2.3 Problème de décision bayésienne associé

La première étape de la démonstration consiste à montrer que le problème d'optimalité que nous considérons est équivalent à un problème de décision bayésienne, que nous reformulons en termes d'arrêt optimal dans une deuxième étape. Finalement, la résolution de ce dernier problème nous fournira une démonstration concise du Théorème 2.1.

Soit $(\Omega, \mathcal{G})$ l'espace mesurable et $\mathbf{P}^0, \mathbf{P}^1$ les mesures de probabilité définis précédemment. Pour tout $\pi \in [0,1]$, nous posons

$$\mathbf{P}^\pi(\cdot) = \pi \mathbf{P}^1(\cdot) + (1-\pi)\mathbf{P}^0(\cdot).$$

Soit $B \in \mathcal{G}$ tel que $\mathbf{P}^1(B) = 1$ et $\mathbf{P}^0(B) = 0$. Définissons sur $(\Omega, \mathcal{G}, \mathbf{P}^\pi)$ la variable aléatoire de Bernoulli $\theta = \mathbf{1}_B$, fonction indicatrice de l'événement B. Elle est telle que :

$$\mathbf{P}^\pi(\theta = 1) = \pi \quad , \quad \mathbf{P}^\pi(\theta = 0) = 1 - \pi.$$

Alors on a la propriété suivante :

$$\mathbf{P}^\pi(A|\theta = i) = \mathbf{P}^i(A) \ , \ i = 0, 1 \ , \ A \in \mathcal{G}.$$

Le paramètre π représente la *probabilité a priori* de θ et nous le supposons fixé dans tout ce paragraphe. Nous définissons à présent la *perte* résultant d'une décision erronée, le *coût* lié à l'observation du processus, la *fonction de risque* associée à une procédure δ, ainsi que son *risque.*

• Si la valeur du paramètre est θ et la décision prise est d, la perte subie est notée $l(\theta, d)$ et on a :

$$l(0,0) = l(1,1) = 0 \ , \ l(0,1) = w_0 > 0 \ , \ l(1,0) = w_1 > 0.$$

• Si nous observons $\{X_s, s \leq t\}$ et si la valeur du paramètre est θ, alors le coût est donné par la fonction $C_t(\theta, X)$ que nous définissons comme suit :

$$C_t(0, X) = c \ln Z_t^1 \quad , \quad C_t(1, X) = c \ln Z_t^0$$

pour c une constante positive.
• La fonction de risque est donnée par

$$R(\theta, \delta) = \mathbf{E}_\theta[l(\theta, d) + C_\tau(\theta, X)] \ ,$$

donc

$$\begin{cases} R(0,\delta) & = & \mathbf{E}_0[l(0,d) + c \ln Z_\tau^1] & = & w_0 \alpha(\delta) + c \ \mathbf{E}_0 \ln Z_\tau^1 \\ R(1,\delta) & = & \mathbf{E}_1[l(1,d) + c \ln Z_\tau^0] & = & w_1 \beta(\delta) + c \ \mathbf{E}_1 \ln Z_\tau^0 \end{cases} .$$

• Le risque associé à δ est défini par :

$$\begin{aligned} r(\pi, \delta) &= \mathbf{E}_\pi(R(\theta, \delta)) \\ &= (1-\pi)R(0,\delta) + \pi R(1,\delta) \\ &= (1-\pi)w_0\alpha(\delta) + \pi w_1 \beta(\delta) + (1-\pi)c \ \mathbf{E}_0 \ln Z_\tau^1 + \pi c \ \mathbf{E}_1 \ln Z_\tau^0. \end{aligned}$$

Définition 2.2 *On dit que δ_* est une* procédure de Bayes par rapport à la probabilité a priori π *si* $r(\pi,\delta_*) = \inf_{\delta\in\Delta} r(\pi,\delta)$, *où Δ désigne la classe de toutes les règles de décision, et $r(\pi,\delta_*)$ est appelé le* risque bayésien.

Il semble clair à présent que nous devons établir que la règle de décision δ_{AB} définie dans le Théorème 2.1 est bayésienne par rapport à toute probabilité a priori $\pi \in [0,1]$. En effet, nous avons choisi la fonction de risque afin de faire apparaître les informations de Kullback. Nous allons montrer à présent que ce problème de décision bayésienne est équivalent à un problème d'arrêt optimal. Plus précisément, on établira le lemme :

Lemme 2.2 *Soit $\mathcal{T}$ l'ensemble de tous les $(\mathcal{G}_t)_{t\geq 0}$-temps d'arrêt satisfaisant à la condition (B). Alors*

$$\inf_{\delta\in\Delta} r(\pi,\delta) = \inf_{\tau\in\mathcal{T}} \mathbf{E}_0 F(Z_\tau^0) \ ,$$

où

$$F(x) = \pi c x \ln x - (1-\pi) c \ln x + ((1-\pi)w_0) \wedge (\pi x w_1) \ .$$

Ce lemme exprime le fait que si la probabilité a priori $\pi \in [0,1]$ est fixée, alors il existe une fonction F, dépendant de π, de laquelle on peut déduire un problème d'arrêt optimal dont la résolution fournit également la solution du problème de minimisation de risque bayésien défini précédemment.
Démonstration.

$$\begin{aligned} r(\pi,\delta) &= \pi \mathbf{E}_1(w_1 \mathbf{1}_{(d=0)} + c\ln Z_\tau^0) + (1-\pi)\mathbf{E}_0(w_0\mathbf{1}_{(d=1)} + c\ln Z_\tau^1) \\ &= \mathbf{E}_0\Big\{(1-\pi)(w_0\mathbf{1}_{(d=1)} + c\ln Z_\tau^1) + \pi Z_\tau^0(w_1\mathbf{1}_{(d=0)} + c\ln Z_\tau^0)\Big\} \ . \end{aligned}$$

A $\delta = (\tau, d)$ nous associons une autre règle de décision $\tilde{\delta} = (\tau,\tilde{d})$ où

$$\tilde{d} = \begin{cases} 0 & \text{si} \quad (1-\pi)w_0 \geq \pi Z_\tau^0 w_1 \\ 1 & \text{sinon} \end{cases} \ .$$

Il est clair que

$$\begin{aligned} r(\pi,\delta) &\geq r(\pi,\tilde{\delta}) \\ &= \mathbf{E}_0\Big\{(1-\pi)c\ln Z_\tau^1 + \pi c Z_\tau^0 \ln Z_\tau^0 + ((1-\pi)w_0)\wedge(\pi Z_\tau^0 w_1)\Big\} \\ &= \mathbf{E}_0 F(Z_\tau^0) \ . \end{aligned}$$

Ainsi, $\quad r(\pi,\delta) \geq r(\pi,\tilde{\delta}) \geq \inf_{\delta\in\Delta} r(\pi,\tilde{\delta}) \quad , \quad$ pour tout $\delta\in\Delta$, donc on a :

$$\inf_{\delta\in\Delta} r(\pi,\delta) \geq \inf_{\delta\in\Delta} r(\pi,\tilde{\delta}) \ .$$

D'autre part, $\tilde{\delta}$ appartient à Δ , si bien que l'on a :

$$r(\pi,\tilde{\delta}) \geq \inf_{\delta\in\Delta} r(\pi,\delta) \quad , \quad \text{pour tout } \delta\in\Delta$$

et ceci entraîne l'inégalité dans l'autre sens :

$$\inf_{\delta\in\Delta} r(\pi,\tilde{\delta}) \geq \inf_{\delta\in\Delta} r(\pi,\delta)\ .$$

Finalement, nous obtenons l'égalité annoncée :

$$\inf_{\delta\in\Delta} r(\pi,\delta) = \inf_{\delta\in\Delta} r(\pi,\tilde{\delta}) = \inf_{\tau\in\mathcal{T}} \mathbf{E}_0 F(Z^0_\tau)\ ,$$

pour F définie ci-dessus.
Il s'agit à présent de résoudre ce problème d'arrêt optimal.
Comme on a :

$$F''(x) = (\pi c x\ln x - (1-\pi)c\ln x)'' = \pi c\frac{1}{x} + (1-\pi)c\frac{1}{x^2}\ >\ 0\ ,$$

F est une fonction convexe sur chacun des intervalles $[0,x_0[$ et $]x_0,+\infty[$, où $x_0 = \dfrac{1-\pi}{\pi}\dfrac{w_0}{w_1}$. On vérifie que $\lim_{x\downarrow 0} F(x) = +\infty$ et $\lim_{x\uparrow+\infty} F(x) = +\infty$ et que le graphe de F est de l'une des deux formes suivantes :

Nous rappelons le lemme suivant qui nous sera d'une grande utilité.

Lemme 2.3 (cf. [Mi]) *Soit G une fonction continue définie sur $[0,+\infty]$ et admettant au moins un minorant convexe. Alors il existe un plus grand minorant convexe, noté $\underline{G}$. L'ensemble $D = \{x : \underline{G}(x) < G(x)\}$ est ouvert et la fonction $\underline{G}$ est linéaire sur chacune des composantes connexes $D_i \subset D$. De plus, les dérivées de G et $\underline{G}$ coïncident en chacune des extrémités de D_i où G est dérivable.*

On applique ce lemme à F, qui est convexe par morceaux (deux composantes connexes) : elle admet donc un minorant convexe maximal $\underline{F}$, qui est linéaire entre deux points a et b définis par le système

$$\begin{cases} \underline{F}'(a) = \frac{F(b)-F(a)}{b-a} & = \quad \pi c\ln a + \pi c - (1-\pi)\frac{c}{a} + \pi w_1 \quad = \quad F'(a) \\ \underline{F}'(b) = \frac{F(b)-F(a)}{b-a} & = \quad \pi c\ln b + \pi c - (1-\pi)\frac{c}{b} \quad = \quad F'(b)\,. \end{cases}$$

Lorsque π est fixée, a et b sont des fonctions de c, w_0 et w_1. On va à présent montrer qu'il existe $c > 0$ et $0 < w < 1$ (tels que $w_0 = 1 - w$ et $w_1 = w$) pour lesquels on a $a = a(c, w) = A$ et $b = b(c, w) = B$, pour A et B les bornes du test séquentiel δ_{AB} introduit plus haut. Il s'agit de résoudre les deux équations :

$$\begin{cases} F(B) - F(A) &= F'(A)(B - A) \\ F(B) - F(A) &= F'(B)(B - A) \; . \end{cases}$$

L'égalité des dérivées en A et B se traduit par :

$$\pi c A + \pi c - (1 - \pi)\, c\, \mathbf{e}^{-A} + \pi w = \pi c B + \pi c - (1 - \pi)\, c\, \mathbf{e}^{-B}$$

ou encore :

$$w = c\,(B - A) + \frac{1 - \pi}{\pi} c\,(\mathbf{e}^{-A} - \mathbf{e}^{-B}) \; .$$

Remplaçons maintenant w par son expression dans la première équation. Après simplification, nous obtenons l'égalité suivante :

$$c \left\{ \pi(\mathbf{e}^{B} - \mathbf{e}^{A}) + 2(1 - \pi)(B - A) + \frac{(1 - \pi)^2}{\pi}(\mathbf{e}^{-A} - \mathbf{e}^{-B}) \right\} = 1 - \pi \; ,$$

d'où finalement :

$$c = \frac{1}{\frac{\pi}{1-\pi}(B - A) + 2(\ln B - \ln A) + \frac{1-\pi}{\pi}(\frac{1}{A} - \frac{1}{B})}$$

et

$$w = \frac{(\ln B - \ln A) + \frac{1-\pi}{\pi}(\frac{1}{A} - \frac{1}{B})}{\frac{\pi}{1-\pi}(B - A) + 2(\ln B - \ln A) + \frac{1-\pi}{\pi}(\frac{1}{A} - \frac{1}{B})} \; .$$

Nous allons enfin résoudre le problème d'arrêt optimal associé à π, c et w. Nous montrerons en fait que l'on a :

$$\tau_{\text{opt}} := \arg \inf_{\tau \in \mathcal{T}} \mathbf{E}_0 F(Z^0_\tau) = \tau_{AB} \; .$$

a) $\inf\limits_{\tau \in \mathcal{T}} \mathbf{E}_0 F(Z^0_\tau) \geq \inf\limits_{\tau \in \mathcal{T}} \mathbf{E}_0 \underline{F}(Z^0_\tau) \geq \inf\limits_{\tau \in \mathcal{T}} \underline{F}(\mathbf{E}_0 Z^0_\tau) = \underline{F}(1)$ par l'inégalité de Jensen.

b) D'autre part, comme $1 \in [A, B]$, la linéarité de $\underline{F}$ sur $[A, B]$ entraîne que l'on a :

$$\begin{aligned} \mathbf{E}_0 \underline{F}(Z^0_{\tau_{AB}}) &= p \mathbf{E}_0 Z^0_{\tau_{AB}} + q \quad (\text{car } Z^0_{\tau_{AB}} \in \{A, B\} \ \mathbf{Q}\text{−p.s.}) \\ &= \underline{F}(\mathbf{E}_0 Z^0_{\tau_{AB}}) \\ &= \underline{F}(1) \; . \end{aligned}$$

Or $\underline{F}(Z^0_{\tau_{AB}}) = F(Z^0_{\tau_{AB}})$ $\mathbf{Q}$-p.s., donc

$$\underline{F}(1) = \mathbf{E}_0 \underline{F}(Z^0_{\tau_{AB}}) = \mathbf{E}_0 F(Z^0_{\tau_{AB}}) \geq \inf_{\tau \in \mathcal{T}} \mathbf{E}_0 F(Z^0_\tau) \; .$$

Il s'ensuit que l'on a :

$$\inf_{\tau\in\mathcal{T}} \mathbf{E}_0 F(Z^0_\tau) \geq \underline{F}(1) = \mathbf{E}_0 F(Z^0_{\tau_{AB}}) \geq \inf_{\tau\in\mathcal{T}} \mathbf{E}_0 F(Z^0_\tau) \ ,$$

ou encore

$$\mathbf{E}_0 F(Z^0_{\tau_{AB}}) = \inf_{\tau\in\mathcal{T}} \mathbf{E}_0 F(Z^0_\tau) \ .$$

Nous venons d'établir le lemme :

Lemme 2.4 *On suppose les conditions I et II satisfaites. Soit $0 \leq \pi \leq 1$ un réel fixé. Alors le test séquentiel de bornes A et B est une procédure de Bayes par rapport à la probabilité a priori π.*

En effet, il suffit de remarquer que $\widetilde{d_{AB}} = d_{AB}$ et donc que l'on a :

$$r(\pi, \delta_{AB}) = r(\pi, \widetilde{\delta_{AB}}) = \mathbf{E}_0 F(Z^0_{\tau_{AB}}) = \inf_{\tau\in\mathcal{T}} \mathbf{E}_0 F(Z^0_\tau) = \inf_{\delta\in\Delta} r(\pi, \delta) \ .$$

2.4 Démonstration du Théorème 2.1

Soit $\delta' = (\tau', d') \in \Delta_{\alpha\beta}$ une autre procédure de test. Comme δ_{AB} réalise le risque bayésien, on a :

$$(1-\pi)(1-w)\alpha + \pi w\beta + (1-\pi)c\ \mathbf{E}_0 \ln Z^1_{\tau_{AB}} + \pi c\ \mathbf{E}_1 \ln Z^0_{\tau_{AB}}$$
$$\leq (1-\pi)(1-w)\alpha' + \pi w\beta' + (1-\pi)c\ \mathbf{E}_0 \ln Z^1_{\tau'} + \pi c\ \mathbf{E}_1 \ln Z^0_{\tau'}$$

$\Big(\alpha' = \mathbf{P}^0(d'=1)\ ,\ \beta' = \mathbf{P}^1(d'=0)\Big)$.

Comme $\alpha' \leq \alpha$ et $\beta' \leq \beta$, il vient que

$$(1-\pi)\mathbf{E}_0 \ln Z^1_{\tau_{AB}} + \pi \mathbf{E}_1 \ln Z^0_{\tau_{AB}} \leq (1-\pi)\mathbf{E}_0 \ln Z^1_{\tau'} + \pi\ \mathbf{E}_1 \ln Z^0_{\tau'}$$

pour toute probabilité a priori $\pi \in [0,1]$.
Cette dernière inégalité entraîne que $\mathbf{E}_0 \ln Z^1_{\tau_{AB}}$ et $\mathbf{E}_1 \ln Z^0_{\tau_{AB}}$ sont finies, donc que $\delta_{AB} = (\tau_{AB}, d_{AB}) \in \Delta_{\alpha\beta}$. On en conclut que

$$\mathbf{E}_0 \ln Z^1_{\tau'} \geq \mathbf{E}_0 \ln Z^1_{\tau_{AB}} \quad \text{et} \quad \mathbf{E}_1 \ln Z^0_{\tau'} \geq \mathbf{E}_1 \ln Z^0_{\tau_{AB}}$$

pour toute règle de décision $\delta' \in \Delta_{\alpha\beta}$. L'optimalité du test de Wald est donc établie.
Nous démontrons les relations liant bornes et probabilités d'erreur du test de Wald, ainsi que les égalités figurant dans le Théorème 2.1 :

$$\mathbf{E}_0 Z^0_{\tau_{AB}} = 1 = A\,\mathbf{P}^0(d=0) + B\,\mathbf{P}^0(d=1) = A(1-\alpha) + B\alpha\ ,$$

$$\mathbf{E}_1 Z^1_{\tau_{AB}} = 1 = \frac{1}{A}\,\mathbf{P}^1(d=0) + \frac{1}{B}\,\mathbf{P}^1(d=1) = \frac{1}{A}\beta + \frac{1}{B}(1-\beta)\ ,$$

d'où

$$\alpha = \frac{1-A}{B-A} \qquad ; \qquad \beta = \frac{A(B-1)}{B-A}\ ,$$

ce qui est équivalent à :

$$A = \frac{\beta}{1-\alpha} \qquad ; \qquad B = \frac{1-\beta}{\alpha}\ .$$

Calculons à présent les informations de Kullback :

$$\mathbf{E}_0 \ln Z^1_{\tau_{AB}} = -\mathbf{E}_0 \ln Z^0_{\tau_{AB}} = -\Big((1-\alpha)\ln A + \alpha \ln B\Big)$$
$$= (1-\alpha)\ln\frac{1-\alpha}{\beta} + \alpha \ln \frac{\alpha}{1-\beta} = \omega(\alpha,\beta) \quad ;$$

$$\mathbf{E}_1 \ln Z^0_{\tau_{AB}} = \beta \ln A + (1-\beta)\ln B$$
$$= \beta \ln \frac{\beta}{1-\alpha} + (1-\beta)\ln\frac{1-\beta}{\alpha} = \omega(\beta,\alpha) \quad ;$$

pour

$$\omega(x,y) = x \ln \frac{x}{1-y} + (1-x)\ln\frac{1-x}{y}\ .$$

2.5 Cas des processus à accroissements indépendants

Nous conservons les notations du paragraphe 1 et supposons de plus que X est un processus à accroissements indépendants (P.A.I.) sous $\mathbf{P}$. L'objet de ce paragraphe est de réécrire les conditions I et II du Théorème 2.1 en utilisant les décompositions de X_t par rapport aux probabilités $\mathbf{P}^0$ et $\mathbf{P}^1$.

Soit $(X_t, \mathcal{G}_t, \mathbf{P})_{t\geq 0}$ un P.A.I., μ^X la mesure associée aux sauts de X et ν le compensateur de μ^X par rapport à $\mathbf{P}$. Le compensateur ν vérifie :

$$\int_0^t \int (x^2 \wedge 1)\ \nu(dt, dx) < +\infty\ .$$

Posant

$$X'_t = X_t - X_0 - \textstyle\sum_{s\leq t} \Delta X_s.\mathbf{1}_{(|\Delta X_s|>1)} = X_t - X_0 - \displaystyle\int_0^t \int_{|x|>1} x\ d\mu^X\ ,$$

nous obtenons un P.A.I. dont les sauts sont bornés (par 1). Le processus

$$X''_t = \int_0^t \int_{|x|\leq 1} x\ d(\mu^X - \nu)$$

est une martingale locale purement discontinue et un P.A.I. Par construction, on a :

$$\Delta X''_t = \Delta X_t.\mathbf{1}_{(|\Delta X_t|\leq 1)} = \Delta X'_t\ .$$

Ainsi,

$$X'''_t = X_t - X_0 - \textstyle\sum_{s\leq t} \Delta X_s.\mathbf{1}_{(|\Delta X_s|>1)} - X''_t = X'_t - X''_t$$

est également un P.A.I., mais il est à trajectoires continues ; nous lui appliquons la formule de Lévy-Khintchine :

$$\mathbf{E}\ \mathbf{e}^{iu(X'''_t - X'''_s)} = \exp\{iu(\alpha(t) - \alpha(s)) - \frac{u^2}{2}(\sigma^2(t) - \sigma^2(s))\} \quad ,$$

pour α et σ^2 deux processus déterministes.

Soit $M = X''' - \alpha$. Ceci est un P.A.I. à trajectoires continues et tel que

$$\mathbf{E}\ \mathbf{e}^{iu(M_t - M_s)} = \exp\{-\frac{u^2}{2}(\sigma^2(t) - \sigma^2(s))\}\ ,$$

donc une martingale gaussienne et on a :

$$X_t = X_0 + \alpha(t) + M_t + \int_0^t \int_{|x|\leq 1} x\ d(\mu^X - \nu) + \int_0^t \int_{|x|>1} x\ d\mu^X \qquad \mathbf{P}-\text{p.s.}$$

En appliquant cette décomposition à $\mathbf{P}^i$, $i = 0, 1$, on obtient :

$$X_t = X_0 + \alpha_i(t) + M^i_t + \int_0^t \int_{|x|\leq 1} x\ d(\mu^X - \nu_i) + \int_0^t \int_{|x|>1} x\ d\mu^X \qquad \mathbf{P}^i-\text{p.s.}$$

où $\alpha_i(t)$ est déterministe, $(M_t^i, \mathcal{G}_t, \mathbf{P}^i)_{t\geq 0}$ est une martingale gaussienne continue et ν_i est le compensateur de μ^X par rapport à $\mathbf{P}^i$, $i = 0, 1$.

Soit $E = \mathbf{R} \setminus \{0\}$. Considérons les conditions suivantes :

I. $\mathbf{P}_0^0 = \mathbf{P}_0^1$

II.a1. $d\nu_1 = Y_0 d\nu_0$ a2. $d\nu_0 = Y_1 d\nu_1$
b. $a_t^0 = \nu_0(\{t\}, E) = 1 \iff a_t^1 = \nu_1(\{t\}, E) = 1$
c. $\langle M^0, M^0\rangle = \langle M^1, M^1\rangle$
d1. $\alpha_1(t) - \alpha_0(t) - \int_0^t \int_{|x|\leq 1} x(Y_0(s,x) - 1)\ d\nu_0 = \int_0^t \gamma_s^0\ d\langle M^0, M^0\rangle_s$
d2. $\alpha_0(t) - \alpha_1(t) - \int_0^t \int_{|x|\leq 1} x(Y_1(s,x) - 1)\ d\nu_1 = \int_0^t \gamma_s^1\ d\langle M^1, M^1\rangle_s$

III. Posons $B_t^0 = \int_0^t (\gamma_s^0)^2\ d\langle M^0, M^0\rangle_s + \int_0^t \int_E (1 - \sqrt{Y_0(s,x)})^2\ d\nu_0$

$$+ \sum_{s\leq t} \mathbf{1}_{(0<a_s^0<1)} \Big(1 - \sqrt{\frac{1 - a_s^1}{1 - a_s^0}}\Big)^2 (1 - a_s^0)$$

et

$$B_t^1 = \int_0^t (\gamma_s^1)^2\ d\langle M^1, M^1\rangle_s + \int_0^t \int_E (1 - \sqrt{Y_1(s,x)})^2\ d\nu_1$$

$$+ \sum_{s\leq t} \mathbf{1}_{(0<a_s^1<1)} \Big(1 - \sqrt{\frac{1 - a_s^0}{1 - a_s^1}}\Big)^2 (1 - a_s^1)\ .$$

a1. $B_t^0 < \infty$ si $t < +\infty$
a2. $B_t^1 < \infty$ si $t < +\infty$
b1. $B_\infty^0 = +\infty$
b2. $B_\infty^1 = +\infty$.

Puisque X est un P.A.I. sous les deux probabilités, ν_0 et ν_1 sont des mesures non aléatoires et les conditions II entraînent que Y_0 , Y_1 , B^0 et B^1 sont déterministes.

On peut d'emblée faire les remarques suivantes :

1) I, II.a1, II.b, II.c, II.d1 et III.a1 (resp. I, II.a2, II.b, II.c, II.d2 et III.a2) traduisent le fait que $\mathbf{P}_t^1 \ll \mathbf{P}_t^0$, $t \geq 0$ (resp. $\mathbf{P}_t^0 \ll \mathbf{P}_t^1$, $t \geq 0$) ; donc I, II et III.a signifient que $\mathbf{P}_t^1 \sim \mathbf{P}_t^0$, $t \geq 0$. On dira que les mesures $\mathbf{P}^0$ et $\mathbf{P}^1$ sont *mutuellement localement absolument continues*.
2) Le Théorème 15 dans [Ka-Li-Sh3] exprime les faits suivants :
- Si I, II.a1, II.b, II.c, II.d1 et III.a1 sont réalisées, alors III.b1 $\iff \mathbf{P}^1 \perp \mathbf{P}^0$;
- Si I, II.a2, II.b, II.c, II.d2 et III.a2 sont réalisées, alors III.b2 $\iff \mathbf{P}^1 \perp \mathbf{P}^0$.

Expliquons un peu les conditions II.d. Supposons, pour fixer les idées, que $\mathbf{P}_t^1 \ll \mathbf{P}_t^0$, $t \geq 0$. On rappelle la décomposition de X (par rapport à $\mathbf{P}^0$, mais aussi par rapport à $\mathbf{P}^1$) :

$$X_t = X_0 + \alpha_0(t) + M_t^0 + \int_0^t \int_{|x|\leq 1} x \; d(\mu^X - \nu_0) + \int_0^t \int_{|x|>1} x \; d\mu^X$$

On sait que $d\langle Z^{0c}, M^0\rangle \ll d\langle M^0, M^0\rangle$, où Z^{0c} désigne la partie continue de la martingale locale Z^0. Posons $\beta_t^0 = \dfrac{d\langle Z^{0c}, M^0\rangle_t}{d\langle M^0, M^0\rangle_t}$.

Alors $\widetilde{M_t^0} = M_t^0 - \int_0^t Z_{s-}^{0\oplus} d\langle Z^{0c}, M^0\rangle_s$ est une $(\mathcal{G}_t, \mathbf{P}^1)$-martingale locale (par le Théorème de Girsanov) et on a :

$$\begin{aligned} X_t &= X_0 + \alpha_0(t) + \widetilde{M_t^0} + \int_0^t Z_{s-}^{0\oplus} \; d\langle Z^{0c}, M^0\rangle_s \\ &\quad + \int_0^t \int_{|x|\leq 1} x \; d(\mu^X - \nu_0) - \int_0^t \int_{|x|\leq 1} x(Y_0(s,x) - 1) \; d\nu_0 \\ &\quad\quad + \int_0^t \int_{|x|\leq 1} x(Y_0(s,x) - 1) \; d\nu_0 + \int_0^t \int_{|x|>1} x \; d\mu^X \\ &= X_0 + \alpha_0(t) + \widetilde{M_t^0} + \int_0^t \int_{|x|\leq 1} x \; d(\mu^X - \nu_1) + \int_0^t Z_{s-}^{0\oplus} \; d\langle Z^{0c}, M^0\rangle_s \\ &\quad + \int_0^t \int_{|x|\leq 1} x(Y_0(s,x) - 1) \; d\nu_0 + \int_0^t \int_{|x|>1} x \; d\mu^X \qquad \mathbf{P}^1\text{−p.s.} \end{aligned}$$

Or la décomposition de X pour $\mathbf{P}^1$ est donnée par

$$X_t = X_0 + \alpha_1(t) + M_t^1 + \int_0^t \int_{|x|\leq 1} x \; d(\mu^X - \nu_1) + \int_0^t \int_{|x|>1} x \; d\mu^X \quad \mathbf{P}^1\text{−p.s.},$$

si bien qu'en identifiant les termes dans les deux dernières égalités, on obtient :

$$\begin{aligned} 0 &= \Big(\alpha_1(t) - \alpha_0(t) - \int_0^t \int_{|x|\leq 1} x(Y_0(s,x) - 1) d\nu_0\Big) - \int_0^t Z_{s-}^{0\oplus} d\langle Z^{0c}, M^0\rangle_s + (M_t^1 - \widetilde{M_t^0}) \\ &= A_t - A'_t + M'_t. \end{aligned}$$

Mais $M'_t \in \mathcal{M}_{loc}(\mathcal{G}_t, \mathbf{P}^1)$ est à trajectoires continues et A_t , A'_t sont à variation finie par rapport à $\mathbf{P}^1$, donc l'égalité $M'_t = A'_t - A_t$ entraîne que $M'_t = A'_t - A_t = 0$ $\mathbf{P}^1$−p.s. et donc que l'on a :

$$\begin{aligned} \alpha_1(t) - \alpha_0(t) - \int_0^t \int_{|x|\leq 1} x(Y_0(s,x) - 1) \, d\nu_0 &= \int_0^t Z_{s-}^{0\oplus} \, d\langle Z^{0c}, M^0\rangle_s \\ &= \int_0^t Z_{s-}^{0\oplus} \beta_s^0 \, d\langle M^0, M^0\rangle_s \, . \end{aligned}$$

En posant $\gamma_s^0 = Z_{s-}^{0\oplus} \; \beta_s^0$, on obtient le résultat voulu. La condition II.d2 se déduit de manière analogue du fait que $\mathbf{P}_t^0 \ll \mathbf{P}_t^1$, $t \geq 0$.

Le Théorème 2.1 reste bien entendu valable. Les conditions I et II ont simplement été réécrites en termes de décompositions de X selon les deux probabilités.

Remarquons que la condition de positivité des informations de Kullback s'écrit

$$\mathbf{E}_0 \ln \Big(\frac{d\mathbf{P}^0}{d\mathbf{P}^1}\Big|_{\sigma\{X_u, s\leq u\leq t\}} \Big) \ > \ 0 \ , \ \mathbf{E}_1 \ln \Big(\frac{d\mathbf{P}^1}{d\mathbf{P}^0}\Big|_{\sigma\{X_u, s\leq u\leq t\}} \Big) \ > \ 0 \ , \ s < t \ ,$$

dans le cas des P.A.I.

2.6 Application au cas des suites de variables aléatoires indépendantes

Il s'agit d'un corollaire immédiat de ce qui précède ; nous nous bornerons donc à préciser les notations (voir [Ja-Sh]) et à énoncer le résultat.

Notons $\mathbf{P}^0$ (resp. $\mathbf{P}^1$) la loi d'une suite de variables aléatoires (v.a.) indépendantes, à valeurs dans $\mathbf{R}^d$, si H_0 (resp. H_1) est vraie. Ceci est formalisé de la manière suivante :

$$\begin{cases} \Omega = (\mathbf{R}^d)^{\mathbf{N}^*} \\ \xi_n(\omega) \text{ est la } n^{\text{ème}} \text{ coordonnée de } \omega \in \Omega \\ \mathcal{F} = (\mathcal{B}(\mathbf{R}^d))^{\otimes \mathbf{N}^*} \ , \ \mathcal{F}_0 = \{\emptyset, \Omega\} \ , \ \mathcal{F}_n = \sigma\{\xi_1, \cdots, \xi_n\} \ , \ n \geq 1 \\ \mathbf{P}^0 = \bigotimes_{n \in \mathbf{N}^*} \rho_n^0 \ , \ \mathbf{P}^1 = \bigotimes_{n \in \mathbf{N}^*} \rho_n^1 \end{cases}$$

où ρ_n^0 et ρ_n^1 sont des mesures de probabilité sur $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$. Alors les v.a. canoniques $(\xi_n)_{n\geq 1}$ sont indépendantes sous $\mathbf{P}^0$ (resp. $\mathbf{P}^1$) et la loi de ξ_n est ρ_n^0 (resp. ρ_n^1).

Pour tout $n \geq 1$, on considère une dominante $\overline{\rho}_n$ de ρ_n^0 et ρ_n^1 et on note $\eta_n^0 = \dfrac{d\rho_n^0}{d\overline{\rho}_n}$ et $\eta_n^1 = \dfrac{d\rho_n^1}{d\overline{\rho}_n}$ les dérivées de Radon-Nikodym.

Ainsi, si on pose $\mathbf{Q} = \bigotimes_{n\in\mathbf{N}^*} \overline{\rho}_n$, on a $\mathbf{P}^0 \overset{\text{loc}}{\ll} \mathbf{Q}$ et $\mathbf{P}^1 \overset{\text{loc}}{\ll} \mathbf{Q}$, c'est-à-dire que $\forall\ n \in \mathbf{N}^*$, $\mathbf{Q}_n = \overline{\rho}_1 \otimes \cdots \otimes \overline{\rho}_n$, $\mathbf{P}_n^0 = \rho_1^0 \otimes \cdots \otimes \rho_n^0$, $\mathbf{P}_n^1 = \rho_1^1 \otimes \cdots \otimes \rho_n^1$, donc $\mathbf{P}_n^0 \ll \mathbf{Q}_n$ et $\mathbf{P}_n^1 \ll \mathbf{Q}_n$.

Soient $\gamma_n^0 = \dfrac{d\mathbf{P}_n^0}{d\mathbf{Q}_n}$ et $\gamma_n^1 = \dfrac{d\mathbf{P}_n^1}{d\mathbf{Q}_n}$. Par indépendance, on a :

$$\gamma_n^0 = \begin{cases} 1 & \text{si } n = 0 \\ \prod_{p=1}^{n} \eta_p^0(\xi_p) & \text{si } n \geq 1 \end{cases} \quad \text{et} \quad \gamma_n^1 = \begin{cases} 1 & \text{si } n = 0 \\ \prod_{p=1}^{n} \eta_p^1(\xi_p) & \text{si } n \geq 1 \end{cases} .$$

Posons $\beta_n^0 = \gamma_n^0(\gamma_{n-1}^0)^{\oplus}$ et $\beta_n^1 = \gamma_n^1(\gamma_{n-1}^1)^{\oplus}$.
Alors $\beta_n^0 = \eta_n^0(\xi_n)\mathbf{1}_{(\gamma_{n-1}^0 > 0)}$ et $\beta_n^1 = \eta_n^1(\xi_n)\mathbf{1}_{(\gamma_{n-1}^1 > 0)}$.

Nous rappelons quelques résultats concernant la *distance de Hellinger-Kakutani*.

Soient $\mathcal{P}$ et $\tilde{\mathcal{P}}$ deux probabilités sur $(\Omega, \mathcal{F})$; soit $\mathcal{Q}$ une dominante. On note $z = \dfrac{d\mathcal{P}}{d\mathcal{Q}}$ et $\tilde{z} = \dfrac{d\tilde{\mathcal{P}}}{d\mathcal{Q}}$ les dérivées de Radon-Nikodym. Alors on a :

$$\begin{aligned} \rho^2(\mathcal{P}, \tilde{\mathcal{P}}) &:= \frac{1}{2}\int_\Omega \left(\sqrt{d\mathcal{P}} - \sqrt{d\tilde{\mathcal{P}}}\right)^2 = \frac{1}{2}\int_\Omega \left(\sqrt{\frac{d\mathcal{P}}{d\mathcal{Q}}} - \sqrt{\frac{d\tilde{\mathcal{P}}}{d\mathcal{Q}}}\right)^2 d\mathcal{Q} \\ &= \frac{1}{2}\mathbf{E}_{\mathcal{Q}}[(\sqrt{z} - \sqrt{\tilde{z}})^2] . \end{aligned}$$

- $\rho(\mathcal{P},\tilde{\mathcal{P}})$ est appelée *distance de Hellinger-Kakutani* entre $\mathcal{P}$ et $\tilde{\mathcal{P}}$.
- $\rho^2(\mathcal{P},\tilde{\mathcal{P}}) = 1 - H(\mathcal{P},\tilde{\mathcal{P}})$, où $H(\mathcal{P},\tilde{\mathcal{P}})$ est l'*intégrale de Hellinger*, définie par

$$H(\mathcal{P},\tilde{\mathcal{P}}) = \mathbf{E}_{\mathcal{Q}}(\sqrt{z\tilde{z}}) = \int_\Omega \sqrt{z\tilde{z}}\; d\mathcal{Q}.$$

- Plus généralement, on définit l'*intégrale de Hellinger d'ordre* α de $\mathcal{P}$ et $\tilde{\mathcal{P}}$ par

$$H(\alpha;\mathcal{P},\tilde{\mathcal{P}}) = \mathbf{E}_{\mathcal{Q}}(z^\alpha.\tilde{z}^{1-\alpha}) \ , \quad \alpha \in]0,1[\ .$$

On a évidemment $\quad H(\mathcal{P},\tilde{\mathcal{P}}) = H(\frac{1}{2};\mathcal{P},\tilde{\mathcal{P}})$.

On rappelle le théorème :

Théorème 2.2 (Kakutani) *a)* $\mathbf{P}^1 \ll \mathbf{P}^0$ *si et seulement si les deux conditions suivantes sont réalisées :*

(i) $\sum_{n\geq 1} [1 - H(\rho_n^0,\rho_n^1)] < +\infty$

(ii) $\rho_n^1 \ll \rho_n^0 \ , \ \forall \ n \in \mathbf{N}^*$

b) $\mathbf{P}^1 \perp \mathbf{P}^0$ *si et seulement si l'une (au moins) des deux conditions suivantes est réalisée :*

(i') $\sum_{n\geq 1} [1 - H(\rho_n^0,\rho_n^1)] = +\infty$

(ii') $\exists \ n \in \mathbf{N}^*$ *tel que* $\rho_n^1 \perp \rho_n^0$.

Corollaire 2.1 *Supposons que* $\rho_n^1 \sim \rho_n^0$ *pour tout* $n \geq 1$. *Alors, soit* $\mathbf{P}^1 \sim \mathbf{P}^0$, *soit* $\mathbf{P}^1 \perp \mathbf{P}^0$.

Ce résultat est connu sous le nom d'*alternative de Kakutani*. Ces deux résultats peuvent se déduire d'un théorème de [Ka-Li-Sh1] que nous énonçons à présent.

Soient $\mathbf{P}^0$ et $\mathbf{P}^1$ deux mesures de probabilité telles que $\mathbf{P}_n^1 \ll \mathbf{P}_n^0 \ , \quad \forall \ n \geq 1$. On note $Z_n^0 = \dfrac{d\mathbf{P}_n^1}{d\mathbf{P}_n^0}$ une version (finie) de la dérivée de Radon-Nikodym de $\mathbf{P}_n^1$ par rapport à $\mathbf{P}_n^0$; $Z_0^0 = 1$). Posons $\alpha_n^0 = Z_n^0(Z_{n-1}^0)^{\oplus}$. Les lemmes 5 et 6 de [Ka-Li-Sh1] fournissent le critère suivant :

Si $\mathbf{P}_n^1 \ll \mathbf{P}_n^0 \ , \ n \geq 1$, *alors :*

$\mathbf{P}^1(Z_\infty^0 < +\infty) = 1 \Longleftrightarrow \mathbf{P}^1 \ll \mathbf{P}^0$

$\mathbf{P}^1(Z_\infty^0 = +\infty) = 1 \Longleftrightarrow \mathbf{P}^1 \perp \mathbf{P}^0$,

où $Z_\infty^0 = \lim_{n\to+\infty} Z_n^0$ *existe* $\mathbf{P}^1$-*p.s.*

On en déduit le théorème :

Théorème 2.3 (cf. [Ka-Li-Sh1]) *On suppose que* $\mathbf{P}_n^1 \ll \mathbf{P}_n^0 \ , \ n \geq 1$. *Alors :*

$$\mathbf{P}^1\Big\{\sum_{n=1}^{+\infty} [1 - \mathbf{E}_0(\sqrt{\alpha_n^0}|\mathcal{F}_{n-1})] \ < \ +\infty\Big\} = 1 \Longleftrightarrow \mathbf{P}^1 \ll \mathbf{P}^0 \ ,$$

$$\mathbf{P}^1\Big\{\sum_{n=1}^{+\infty}[1-\mathbf{E}_0(\sqrt{\alpha_n^0}|\mathcal{F}_{n-1})] \;=\; +\infty\Big\} = 1 \iff \mathbf{P}^1 \perp \mathbf{P}^0 \;.$$

Corollaire 2.2 *On suppose que les tribus $\sigma(\alpha_n^0)$ et $\mathcal{F}_{n-1}$ sont $\mathbf{P}^0$ (ou bien $\mathbf{P}^1$)-indépendantes et que $\mathbf{P}_n^1 \ll \mathbf{P}_n^0$ pour tout $n \geq 1$. Alors, ou bien $\mathbf{P}^1 \ll \mathbf{P}^0$, ou bien $\mathbf{P}^1 \perp \mathbf{P}^0$. Plus précisément,*

$$\sum_{n=1}^{+\infty}[1-\mathbf{E}_0(\sqrt{\alpha_n^0})] < +\infty \iff \mathbf{P}^1 \ll \mathbf{P}^0$$

$$\sum_{n=1}^{+\infty}[1-\mathbf{E}_0(\sqrt{\alpha_n^0})] = +\infty \iff \mathbf{P}^1 \perp \mathbf{P}^0 \;.$$

Ceci n'est rien d'autre que le Théorème de Kakutani. En effet,

$$\begin{aligned}
\alpha_n^0 = Z_n^0(Z_{n-1}^0)^{\oplus} &= \frac{d\mathbf{P}_n^1}{d\mathbf{P}_n^0}\,\frac{d\mathbf{P}_{n-1}^0}{d\mathbf{P}_{n-1}^1}\,\mathbf{1}_{(Z_{n-1}^0>0)} \\
&= \frac{d\mathbf{P}_n^1}{d\mathbf{Q}_n}\,\frac{d\mathbf{Q}_n}{d\mathbf{P}_n^0}\,\frac{d\mathbf{P}_{n-1}^0}{d\mathbf{Q}_{n-1}}\,\frac{d\mathbf{Q}_{n-1}}{d\mathbf{P}_{n-1}^1}\,\mathbf{1}_{(Z_{n-1}^0>0)} \\
&= \gamma_n^1(\gamma_n^0)^{\oplus}\gamma_{n-1}^0(\gamma_{n-1}^1)^{\oplus} \\
&= \beta_n^1(\beta_n^0)^{\oplus} \\
&= \frac{\eta_n^1(\xi_n)}{\eta_n^0(\xi_n)} \;=\; \frac{d\rho_n^1}{d\overline{\rho}_n}\,\frac{d\overline{\rho}_n}{d\rho_n^0} \;=\; \frac{d\rho_n^1}{d\rho_n^0}\;.
\end{aligned}$$

En particulier, α_n^0 est $\sigma(\xi_n)$-mesurable et on a :

$$\mathbf{E}_0(\sqrt{\alpha_n^0}) = \mathbf{E}_{\rho_n^0}(\sqrt{\alpha_n^0}) = \int_\Omega \sqrt{\frac{d\rho_n^1}{d\rho_n^0}}\;d\rho_n^0 = \int_\Omega \sqrt{d\rho_n^1\;d\rho_n^0}\;.$$

On constate également que α_n^0 est indépendante de $\mathcal{F}_{n-1}$ (puisque les ξ_n sont indépendantes). On conclut en remarquant qu'on a :

$$\begin{aligned}
H(\rho_n^0,\rho_n^1) = \mathbf{E}_{\overline{\rho}_n}(\sqrt{\eta_n^0\eta_n^1}) &= \int_\Omega \sqrt{\eta_n^0\eta_n^1}\;d\overline{\rho}_n = \int_\Omega \sqrt{\frac{d\rho_n^0}{d\overline{\rho}_n}\,\frac{d\rho_n^1}{d\overline{\rho}_n}}\;d\overline{\rho}_n \\
&= \int_\Omega \sqrt{d\rho_n^0\;d\rho_n^1} = \mathbf{E}_0(\sqrt{\alpha_n^0})\;.
\end{aligned}$$

Le corollaire suivant est une réécriture de l'alternative de Kakutani.

Supposons que $\mathbf{P}_n^1 \sim \mathbf{P}_n^0$ pour tout $n \geq 1$ et que les tribus $\sigma(\alpha_n^0)$ et $\mathcal{F}_{n-1}$ sont indépendantes par rapport à $\mathbf{P}^0$ (ou $\mathbf{P}^1$). Alors, ou bien $\mathbf{P}^1 \sim \mathbf{P}^0$, ou bien $\mathbf{P}^1 \perp \mathbf{P}^0$, selon la convergence ou la divergence des séries $\sum_{n=1}^{+\infty}[1-\mathbf{E}_0(\sqrt{\alpha_n^0})]$ et $\sum_{n=1}^{+\infty}[1-\mathbf{E}_1(\sqrt{\alpha_n^1})]$, où α_n^1 est défini de manière analogue à α_n^0.

Nous allons appliquer ce qui précède pour donner des conditions sous lesquelles le test de Wald est optimal au sens donné plus haut. On suppose réalisées :
I. $\forall\, n \geq 1$, $\mathbf{P}^1_n \sim \mathbf{P}^0_n$
II.a. $\displaystyle\sum_{n=1}^{+\infty} [1 - \mathbf{E}_0(\sqrt{\alpha^0_n})] = +\infty$ b. $\displaystyle\sum_{n=1}^{+\infty} [1 - \mathbf{E}_1(\sqrt{\alpha^1_n})] = +\infty$.

Remarque 2.1 *Les conditions II.a. et II.b. sont en fait équivalentes, puisque*

$$\mathbf{E}_0(\sqrt{\alpha^0_n}) = \mathbf{E}_1(\sqrt{\alpha^1_n}) = H(\rho^0_n, \rho^1_n) \ .$$

Ainsi, les conditions I et II ne sont autres que celles de l'alternative de Kakutani.

Remarque 2.2 *Si les $(\xi_n)_{n\geq 1}$ sont également identiquement distribuées, la condition I implique la condition II , pourvu que la loi de ξ_1 sous H_0 soit différente de sa loi sous H_1.*
En effet, si $\rho^0_n = \rho^0$ et $\rho^1_n = \rho^1$, on définit $\alpha^0 = \dfrac{d\rho^1}{d\rho^0}$ et $\alpha^1 = \dfrac{d\rho^0}{d\rho^1}$. Alors on a :

$$\mathbf{E}_0\sqrt{\alpha^0} = \mathbf{E}_{\rho^0}\sqrt{\alpha^0} = \int_\Omega \sqrt{\frac{d\rho^1}{d\rho^0}}\ d\rho^0 \leq \Big(\int_\Omega \frac{d\rho^1}{d\rho^0}\ d\rho^0\Big)^{\frac{1}{2}} = 1$$

et l'égalité a lieu si et seulement si $\dfrac{d\rho^1}{d\rho^0} = 1$, ou encore $\rho^1 = \rho^0$, ce qui est équivalent à $\mathbf{P}^1 = \mathbf{P}^0$.
Donc, si $\rho^0 \sim \rho^1$, mais $\rho^0 \neq \rho^1$, alors $\mathbf{E}_0\sqrt{\alpha^0} < 1$ et il en résulte que l'on a :
$\displaystyle\sum_{n=1}^{+\infty}(1 - \mathbf{E}_0\sqrt{\alpha^0_n}) = \sum_{n=1}^{+\infty}(1 - \mathbf{E}_0\sqrt{\alpha^0}) = +\infty$. *(II.a)*
On montre de même que $\displaystyle\sum_{n=1}^{+\infty}(1 - \mathbf{E}_1\sqrt{\alpha^1_n}) = \sum_{n=1}^{+\infty}(1 - \mathbf{E}_1\sqrt{\alpha^1}) = +\infty$. *(II.b)*

Il reste à donner un énoncé analogue à celui du Théorème 2.1 dans le cas du temps discret. La seule différence réside en le fait que τ_{AB} est un $(\mathcal{F}_n)_{n\geq 1}$-temps d'arrêt défini par :

$$\tau_{AB} = \inf\{n \geq 1\ :\ Z^0_n \notin]A, B[\ \} \ .$$

Théorème 2.4 *Soit $(\xi_n)_{n\geq 1}$ une suite de variables aléatoires indépendantes. On suppose les conditions I et II réalisées. Soit $\delta_{AB}{=}(\tau_{AB}, d_{AB})$ le test séquentiel de bornes A et B telles que $0 < A \leq 1 \leq B < +\infty$ et $Z^0_{\tau_{AB}} \in \{A, B\}$ $\mathbf{Q}$-p.s. et soient α et β les risques associés. Alors, pour toute procédure $\delta = (\tau, d) \in \Delta_{\alpha\beta}$, on a :*

$$\mathbf{E}_0 \ln Z^1_\tau \ \geq\ \mathbf{E}_0 \ln Z^1_{\tau_{AB}} \qquad \text{et} \qquad \mathbf{E}_1 \ln Z^0_\tau \ \geq\ \mathbf{E}_1 \ln Z^0_{\tau_{AB}}$$

(les égalités figurant dans le Théorème 2.1 restent valables).

Remarque 2.3 *On peut donner une extension de ce théorème au cas des suites de v.a. non nécessairement i.i.d. Les conditions ont déjà été données plus haut :*

I'. $\forall\, n \geq 1\;, \quad \mathbf{P}^1_n \sim \mathbf{P}^0_n$

II'.a. $\mathbf{P}^1\Big\{\sum_{n=1}^{+\infty}[1-\mathbf{E}_0(\sqrt{\alpha^0_n}|\mathcal{F}_{n-1})] = +\infty\Big\} = 1$

II'.b. $\mathbf{P}^0\Big\{\sum_{n=1}^{+\infty}[1-\mathbf{E}_1(\sqrt{\alpha^1_n}|\mathcal{F}_{n-1})] = +\infty\Big\} = 1$.

2.7 Application : un test séquentiel sur la dérive d'un processus de diffusion

Les notations sont celles introduites au paragraphe 1. Nous observons le processus $(V_t, \mathcal{G}_t, \mathbf{P})_{t\geq 0}$ qui représente la vitesse d'une particule et que nous supposons satisfaire à l'*équation de Langevin* :

$$dV_t = -\beta V_t dt + dW_t \quad , \; V_0 = 0 \tag{L}$$

pour β une constante positive et $(W_t, \mathcal{G}_t, \mathbf{P})_{t\geq 0}$ un mouvement brownien standard [1].

La solution V de l'équation différentielle stochastique (L) est appelée le *processus de vélocité d'Ornstein-Uhlenbeck* et V_t représente la vitesse de la particule à l'instant t. En utilisant la formule d'Itô, on montre que la solution de (L) est donnée par :

$$V_t = \mathbf{e}^{-\beta t} \int_0^t \mathbf{e}^{\beta s} \; dW_s \; .$$

Pour tout $T < +\infty$, $(V_t, \mathcal{G}_t, \mathbf{P})_{0\leq t\leq T}$ est une semimartingale gaussienne, continue en moyenne quadratique et on vérifie aisément que l'on a :

$$\int_0^T \mathbf{E}V_t^2 \; dt \; < \; +\infty. \tag{E}$$

En effet, si $\beta = 0$, on a :

$$\int_0^T \mathbf{E}V_t^2 \; dt = \int_0^T \mathbf{E}W_t^2 \; dt = \frac{T^2}{2} \; ,$$

et si $\beta > 0$, on a :

$$\begin{aligned} \int_0^T \mathbf{E}V_t^2 \; dt &= \int_0^T \mathbf{e}^{-2\beta t} \frac{1}{2\beta}(\mathbf{e}^{2\beta t} - 1) \; dt \\ &= \frac{1}{2\beta}T + \frac{1}{(2\beta)^2}(\mathbf{e}^{-2\beta T} - 1) \; . \end{aligned}$$

Cette dernière propriété nous sera utile par la suite.

---

[1]Cette équation n'est qu'une réécriture de la *loi de Newton* dans le cas d'une particule de masse m qui serait soumise à une force de frottement inversement proportionnelle à la vitesse de la particule, $-m\beta V_t$, et à une force aléatoire, $m\frac{dW_t}{dt}$, qui est formellement un processus gaussien stationnaire. La loi de Newton stipule alors que l'on a, du moins de manière formelle :

$$m\frac{dV_t}{dt} = -m\beta V_t + m\frac{dW_t}{dt} \; .$$

On passe de cette écriture à l'équation différentielle stochastique (L), qui est, quant à elle, bien définie.

Notre objectif est de montrer que la règle de Wald minimise les informations de Kullback parmi une certaine classe de règles de décision pour tester

$$H_0 : \beta = 0 \quad \text{contre} \quad H_1 : \beta = 1 \ ,$$

ou encore

$$H_0 : dV_t = dW_t \ , \ V_0 = 0 \quad \text{contre} \quad H_1 : dV_t = -V_t dt + dW_t \ , \ V_0 = 0 \ .$$

Nous munissons l'espace filtré $(\Omega, \mathcal{G}, (\mathcal{G}_t)_{t\geq 0})$ des deux mesures de probabilité $\mathbf{P}^0$ et $\mathbf{P}^1$, qui correspondent respectivement aux lois du processus V sous les hypothèses H_0 et H_1 , les espérances correspondantes étant notées $\mathbf{E}_0$ et $\mathbf{E}_1$. Il s'agit à présent de montrer que les conditions I et II du Théorème 2.1 sont réalisées.

1) Nous savons que le processus V satisfait à (E) sur tout intervalle $[0, T]$, donc le Théorème 7.15 de [Li-Sh1] entraîne que l'on a :

$$\forall \ T < +\infty \ , \quad \mathbf{P}_T^0 \ \sim \ \mathbf{P}_T^1 \ ,$$

c'est-à-dire que les mesures $\mathbf{P}^0$ et $\mathbf{P}^1$ sont localement équivalentes (condition I). De plus, on peut expliciter les processus de densités locales :

$$Z_t^0 = \frac{d\mathbf{P}_t^1}{d\mathbf{P}_t^0} = \exp\Big\{ \int_0^t -V_s dV_s - \frac{1}{2}\int_0^t V_s^2 ds \Big\}$$

et

$$Z_t^1 = \frac{d\mathbf{P}_t^0}{d\mathbf{P}_t^1} = \exp\Big\{ \int_0^t V_s dV_s + \frac{1}{2}\int_0^t V_s^2 ds \Big\} \ .$$

On sait également que les processus $(Z_t^0, \mathcal{G}_t, \mathbf{P}^0)_{t\geq 0}$ et $(Z_t^1, \mathcal{G}_t, \mathbf{P}^1)_{t\geq 0}$ sont des martingales locales à trajectoires $\mathbf{Q}$-p.s. continues et telles que $Z_t^1 = (Z_t^0)^{-1}$ $\mathbf{Q}-$p.s., où $\mathbf{Q} = \frac{1}{2}(\mathbf{P}^0 + \mathbf{P}^1)$. Ceci assure également que $Z_{\tau_{AB}}^0 \in \{A, B\}$ $\mathbf{Q}-$p.s.

2) Montrons que $\mathbf{P}^0$ et $\mathbf{P}^1$ sont singulières. Pour ce faire, nous calculons les processus $M^0, \ M^1, \ B^0$ et B^1 introduits au paragraphe 1 :

$$M_t^0 = \int_0^t (Z_s^0)^{-1} dZ_s^0 \quad , \quad M_t^1 = \int_0^t (Z_s^1)^{-1} dZ_s^1 \ ,$$

$$B_t^0 = \langle M^0, M^0 \rangle_t = \int_0^t (Z_s^0)^{-2} d\langle Z^0, Z^0 \rangle_s \ ,$$

$$B_t^1 = \langle M^1, M^1 \rangle_t = \int_0^t (Z_s^1)^{-2} d\langle Z^1, Z^1 \rangle_s \ .$$

Sous $\mathbf{P}^0$, on a :

$$Z_t^0 = \exp\Big\{ \int_0^t -V_s dW_s - \frac{1}{2}\int_0^t V_s^2 ds \Big\} \ .$$

Si l'on pose $L^0 = \int -V dW$, on constate que $(L^0_t, \mathcal{G}_t, \mathbf{P}^0)_{t\geq 0}$ est une martingale locale, et Z^0 est l'*exponentielle stochastique* de L^0, notée $\mathcal{E}(L^0)$. De plus, Z^0 est solution de l'E.D.S. :

$$dZ^0_t = Z^0_t dL^0_t = -Z^0_t V_t dW_t \ , \ Z^0_0 = 1$$

et on a :

$$\langle Z^0, Z^0 \rangle_t = \int_0^t (Z^0_s)^2 V_s^2 ds \quad , \quad B^0_t = \int_0^t V_s^2 ds \ .$$

De même, sous $\mathbf{P}^1$, on a :

$$Z^1_t = \exp\Big\{ \int_0^t V_s dW_s - \frac{1}{2} \int_0^t V_s^2 ds \Big\} = \mathcal{E}(L^1)_t \ ,$$

où $L^1 = \int V dW$. Donc Z^1 est solution de l'E.D.S. :

$$dZ^1_t = Z^1_t dL^1_t = Z^1_t V_t dW_t \ , \ Z^1_0 = 1$$

et on a :

$$\langle Z^1, Z^1 \rangle_t = \int_0^t (Z^1_s)^2 V_s^2 ds \quad , \quad B^1_t = \int_0^t V_s^2 ds \ .$$

La condition II.a. du Théorème 2.1 s'écrit alors :

$$\mathbf{P}^0 \Big\{ \int_0^{+\infty} V_s^2 ds = +\infty \Big\} = 1 \ ,$$

ou encore

$$\mathbf{P}^0 \Big\{ \int_0^{+\infty} W_s^2 ds = +\infty \Big\} = 1$$

puisque sous H_0, on a $dV_t = dW_t$, $V_0 = 0$, et cette dernière égalité résulte du fait que le mouvement brownien réel est tel que :

$$\liminf_{t\to+\infty} W_t = -\infty \text{ et } \limsup_{t\to+\infty} W_t = +\infty \quad \text{presque sûrement.}$$

On peut également vérifier de manière directe que la condition II.b. est satisfaite à l'aide du lemme suivant :

Lemme 2.5 (voir [LB]) *Soit* $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, (\widetilde{\mathcal{F}}_t)_{t\geq 0}, \widetilde{\mathbf{P}})$ *un espace probabilisé filtré. Soit* $(\widetilde{W}_t, \widetilde{\mathcal{F}}_t, \widetilde{\mathbf{P}})_{t\geq 0}$ *un mouvement brownien standard. Soit* $X = (X_t)_{t\geq 0}$ *la solution de l'E.D.S.*

$$dX_t = \alpha_t X_t dt + d\widetilde{W}_t \ , \ X_0 = 0$$

pour $\alpha = (\alpha_t)_{t\geq 0}$ *un processus continu, adapté et tel que*

$$\lim_{t\to+\infty} \alpha_t = \alpha_\infty \quad \widetilde{\mathbf{P}} - p.s.$$

où α_∞ *est une variable aléatoire* $\widetilde{\mathbf{P}}$*-p.s. finie. Alors* X *vérifie :*

$$\widetilde{\mathbf{P}} \Big\{ \int_0^{+\infty} X_s^2 ds = +\infty \Big\} = 1 \ .$$

Il suffit alors de remarquer que l'équation (L) est de la forme décrite dans ce lemme, avec $\alpha_t \equiv \alpha_\infty = -1$, $\widetilde{\mathbf{P}} = \mathbf{P}$, $\widetilde{W} = W$.

Nous définissons à présent la classe $\Delta_{\alpha\beta}$ que nous considérons dans ce problème en donnant l'expression des informations de Kullback.
Sous $\mathbf{P}^0$, on a :

$$Z_t^1 = \exp\Big\{ \int_0^t V_s dW_s + \frac{1}{2}\int_0^t V_s^2 ds \Big\} .$$

Soit $(\tau, d) \in \Delta_{\alpha\beta}$. Pour tout $n \geq 1$, on pose : $\quad \tau_n = \tau \wedge n$.
La suite de temps d'arrêts bornés $(\tau_n)_{n\geq 1}$ converge presque sûrement vers τ. Comme $\int V dW$ est une martingale locale, on a pour tout $n \geq 1$:

$$\mathbf{E}_0 \ln Z_{\tau_n}^1 = \frac{1}{2}\mathbf{E}_0 \int_0^{\tau_n} V_s^2 ds .$$

Le théorème de convergence monotone entraîne alors que l'on a :

$$\mathbf{E}_0 \ln Z_\tau^1 = \frac{1}{2}\mathbf{E}_0 \int_0^{\tau} V_s^2 ds .$$

De même, sous $\mathbf{P}^1$, on a :

$$Z_t^0 = \exp\Big\{ \int_0^t -V_s dW_s + \frac{1}{2}\int_0^t V_s^2 ds \Big\}$$

et une démarche analogue entraîne que

$$\mathbf{E}_1 \ln Z_\tau^0 = \frac{1}{2}\mathbf{E}_1 \int_0^{\tau} V_s^2 ds .$$

La classe $\Delta_{\alpha\beta}$ de règles de décision considérée est donc constituée des tests $\delta = (\tau, d)$ satisfaisant aux deux conditions suivantes :

(A) $\mathbf{P}^0(d = 1) \leq \alpha$, $\mathbf{P}^1(d = 0) \leq \beta$

(B) $\mathbf{E}_0 \int_0^\tau V_s^2 ds < +\infty$, $\mathbf{E}_1 \int_0^\tau V_s^2 ds < +\infty$.

Remarque 2.4 *Lorsque le processus observé est à trajectoires continues* (cf. [Ya]), *on peut également établir l'optimalité du test de Wald parmi la classe plus étendue des tests vérifiant (A) et*

(B') $\mathbf{P}^0\Big\{ \int_0^\tau V_s^2 ds < +\infty \Big\} = 1$, $\mathbf{P}^1\Big\{ \int_0^\tau V_s^2 ds < +\infty \Big\} = 1$.

Les résultats correspondant à ce chapitre ont donné lieu à un article publié dans la revue *Sequential Analysis* ([Ko1]) et à une communication aux 24$^{\text{èmes}}$ Journées de Statistique de Bruxelles ([Ga-Ko]).

Références bibliographiques

[Ga-Ko] L. I. Galtchouk et C. Koell. *1992 Test séquentiel de deux hypothèses simples sur le paramètre du processus de vélocité d'Ornstein-Uhlenbeck.* Communication aux 24$^{\text{émes}}$ Journées de Statistique de Bruxelles.

[Ir1] A. Irle. *1984 Extended optimality of sequential probability ratio tests.* Annals of Statistics, vol. 12, no. 1, pp. 380-386.

[Ir-Sc] A. Irle et N. Schmitz. *1984 On the optimality of the SPRT for processes with continuous time parameter.* Math. Operationsforsch. u. Statist., ser. Statist., vol. 15, no. 2, pp. 91-104.

[Ja] J. Jacod. *1979 Calcul stochastique et problèmes de martingales.* Lecture Notes in Mathematics no. 714. Springer Verlag.

[Ja-Sh] J. Jacod et A. N. Shiryaev. *1987 Limit theorems for stochastic processes.* Springer Verlag.

[Ka-Li-Sh1] Yu. M. Kabanov, R. Sh. Liptser et A. N. Shiryaev. *1977 On the question of absolute continuity and singularity of probability measures.* Mat. USSR Sbornik, vol. 33, no. 2, pp. 203-221.

[Ka-Li-Sh2] Yu. M. Kabanov, R. Sh. Liptser et A. N. Shiryaev. *1979 Absolute continuity and singularity of locally absolutely continuous probability distributions (I).* Mat. USSR Sbornik, vol. 35, pp. 631-679.

[Ka-Li-Sh3] Yu. M. Kabanov, R. Sh. Liptser et A. N. Shiryaev. *1980 Absolute continuity and singularity of locally absolutely continuous probability distributions (II).* Mat. USSR Sbornik, vol. 36, pp. 31-58.

[Ko1] C. Koell.*1994 An optimal procedure for testing hypotheses on cadlag processes.* Sequential Analysis, vol. 13, no. 3, pp. 221-236.

[LB] A. Le Breton. *1992 Adaptive control in the scalar linear-quadratic model in continuous time.* Statistics and Probability Letters, vol. 13, pp. 169-177.

[Li-Sh1] R. S. Liptser et A. N. Shiryaev. *1977 Statistics of random processes (I). General theory.* Springer Verlag.

[Li-Sh2] R. S. Liptser et A. N. Shiryaev. *1978 Statistics of random processes (II). Applications.* Springer Verlag.

[Me] P. A. Meyer. *1976 Un cours sur les intégrales stochastiques.* Séminaire Proba. XI. Lecture Notes in Mathematics no. 511, pp. 245-400. Springer Verlag.

[Mi] T. P. Miroshnichenko. *1979 Testing of two simple hypotheses in the presence of delayed observations.* Theory of Probability and its Applications, vol. 24, no. 3, pp. 467-479.

[Sh1] A. N. Shiryaev. *1973 Statistical sequential analysis (Translations of Mathematical Monographs, vol. 38).* A.M.S., Providence.

[Sh2] A. N. Shiryaev. *1978 Optimal stopping rules.* Springer Verlag.

[Wa] A. Wald. *1947 Sequential analysis.* Wiley, New York.

[Wa-Wo] A. Wald et J. Wolfowitz. *1948 Optimum character of the sequential probability ratio test.* Annals of Mathematical Statistics, vol. 19, pp. 326-339.

[Ya] A. I. Yashin. *1983 On a problem of sequential hypothesis testing.* Theory of Probability and its Applications, vol. 28, no. 1, pp. 157-165.

Chapitre 3

Optimalité asymptotique du test de Wald

3.1 Introduction

Soient $Z_1, Z_2, \cdots$ des variables aléatoires (v.a.) définies sur un espace mesurable $(\Omega, \mathcal{F})$, indépendantes et identiquement distribuées (i.i.d.), de loi $\mathbf{P}$. On veut tester l'hypothèse $H_0 \ : \ \mathbf{P} = \mathbf{P}^0$ contre l'hypothèse $H_1 \ : \ \mathbf{P} = \mathbf{P}^1$ où $\mathbf{P}^0$ et $\mathbf{P}^1$ sont deux mesures de probabilité sur $(\Omega, \mathcal{F})$. Pour ce faire, on définit p_0 et p_1 les densités respectives de $\mathbf{P}^0$ et $\mathbf{P}^1$ par rapport à une dominante $\mathbf{Q}$, ainsi que le *rapport de vraisemblance*

$$R_n = \prod_{k=1}^{n} p_1(Z_k)/p_0(Z_k)$$

correspondant au n-échantillon $(Z_1, \cdots, Z_n)$, $n \geq 1$.

Notons que si $\mathcal{F}_n$ est la tribu engendrée par les n premières observations $\{Z_1, \cdots, Z_n\}$, le processus $(R_n, \mathcal{F}_n, (\mathbf{P}^0)^{\otimes n})_{n\geq 1}$ est une surmartingale.

Pour A et B deux réels tels que $0 < A < 1 < B$, le *test séquentiel du rapport des probabilités* ou *test de Wald* de bornes A et B est défini par $\delta_{AB} = (\tau_{AB}, d_{AB})$, où

$$\tau_{AB} = \inf\{n \geq 1 : R_n \notin]A, B[\} \quad \text{et} \quad d_{AB} = \begin{cases} 0 & \text{si} \quad R_{\tau_{AB}} \leq A \\ 1 & \text{si} \quad R_{\tau_{AB}} \geq B \end{cases} .$$

Soient α et β les risques associés à δ_{AB}. Ils sont définis par

$$\alpha = \alpha(\delta_{AB}) = \mathbf{P}^0(d_{AB} = 1) \quad ; \quad \beta = \beta(\delta_{AB}) = \mathbf{P}^1(d_{AB} = 0) .$$

On peut montrer que $A \sim \dfrac{\beta}{1-\alpha}$ et $B \sim \dfrac{1-\beta}{\alpha}$ (en négligeant les sauts de R_n à l'instant τ_{AB}).

Soit $\mathcal{T}(\alpha, \beta)$ la classe des règles de décision (τ, d) satisfaisant aux deux conditions suivantes :

(A) $\mathbf{P}^0(d = 1) \leq \alpha$ et $\mathbf{P}^1(d = 0) \leq \beta$
(B) $\mathbf{E}_0\tau < +\infty$ et $\mathbf{E}_1\tau < +\infty$.

Il est bien connu (voir [Le], [Sh1], [Wa], [Wa-Wo]) que le test de Wald minimise les durées moyennes d'observation $\mathbf{E}_0\tau$ et $\mathbf{E}_1\tau$ parmi toutes les règles de décision de $\mathcal{T}(\alpha, \beta)$, à savoir :

$$\inf_{(\tau,d)\in\mathcal{T}(\alpha,\beta)} \mathbf{E}_i\tau = \mathbf{E}_i\tau_{AB} \ , \ i = 0, 1.$$

Des résultats asymptotiques ont été établis par T.L. Lai (cf. [La2]) pour α et β tendant vers 0. Nous rappelons l'un d'eux (Corollaire 1 de [La2]), ainsi que la définition :

Définition 3.1 *Soit r un nombre réel strictement positif. Un processus $(Y_t)_{t\in T}, T =$ $\mathbf{N}$ ou $\mathbf{R}_+$,* converge r-rapidement *vers a (sous la probabilité $\mathbf{P}$) si pour tout $\varepsilon > 0$, on a :*

$$\mathbf{E}L_\varepsilon^r < +\infty ,$$

où $L_\varepsilon = \sup\{n \geq 1 : |Y_n - a| > \varepsilon\}$.

Notons que cette notion est plus forte que celle de convergence presque sûre. On peut établir les propriétés suivantes (cf. [Ir3], [La1], [St]) :

1) Si $(Y_t)_{t\in T}$ converge vers a et $(Z_t)_{t\in T}$ converge vers b r-rapidement, alors :

- (i) $(Y_t+Z_t)_{t\in T}$ converge vers $a+b$,
- (ii) $(Y_tZ_t)_{t\in T}$ converge vers ab ,
- (iii) $\left(\frac{Y_t}{Z_t}\right)_{t\in T}$ converge vers $\frac{a}{b}$ si b est non nul

au sens de la convergence r-rapide ;

2) Si Z est une v.a., alors $\mathbf{E}|Z|^r<+\infty \Longleftrightarrow \frac{Z}{t}$ tend vers 0 r-rapidement ;

3) Si $(Z_n)_{n\geq 1}$ est une suite de variables aléatoires i.i.d. telle que $\mathbf{E}|Z_1|^{1+r}<+\infty$, alors $\frac{Z_n}{n}$ tend vers 0 r-rapidement ;

4) *Loi des grands nombres :*

Soit $(Y_n)_{n\geq 1}$ une suite de variables aléatoires i.i.d. On pose $S_n=Y_1+\cdots+Y_n$, $n\geq 1$. Pour tout $\alpha>\frac{1}{2}$ et tout $p>\frac{1}{\alpha}$, on a :

$$\mathbf{E}|Y_1|^p<+\infty \text{ et, lorsque } \alpha\leq 1\,,\ \mathbf{E}Y_1=0 \iff \frac{S_n}{n}\underset{n\to+\infty}{\longrightarrow}0 \quad (p\alpha-1)\text{−rapidement.}$$

En particulier, si $\alpha=1$, on a pour tout $r>0$:

$$\frac{S_n}{n}\underset{n\to+\infty}{\longrightarrow}0 \quad r\text{−rapidement} \quad \Longleftrightarrow \mathbf{E}|Y_1|^{1+r}<+\infty\ ;$$

5) *Conjecture de Strassen (loi du logarithme) :*

Soit $(Y_n)_{n\geq 1}$ une suite de variables aléatoires i.i.d. On pose $S_n=Y_1+\cdots+Y_n$, $n\geq 1$. Pour tout $r>0$, on a :

$$\limsup_{n\to+\infty}\frac{|S_n|}{\sqrt{n\ln n}}<+\infty \ \ r\text{−rapidement} \iff \mathbf{E}Y_1=0 \text{ et } \mathbf{E}\frac{|Y_1|^{2(r+1)}}{(\ln^+|Y_1|+1)^{r+1}}<+\infty\,.$$

Dans ce cas, on a :

$$\limsup_{n\to+\infty}\frac{|S_n|}{\sqrt{2n\ln n}}=\sqrt{r\mathbf{E}(Y_1^2)} \qquad (r\text{−rapidement}).$$

On considère à présent une suite $(Z_n)_{n\geq 1}$ de variables aléatoires non nécessairement i.i.d., mais vérifiant la condition suivante :

il existe $\lambda_0<0$ et $\lambda_1>0$ tels que $\frac{\ln R_n}{n}\underset{n\to+\infty}{\longrightarrow}\lambda_i$ r-rapidement sous $\mathbf{P}^i$, $i=0,1$, pour une constante positive r.

Nota Bene. La quantité R_n désigne le rapport de vraisemblance de $(Z_1,\cdots,Z_n)$, mais son expression n'est pas nécessairement celle donnée précédemment.

Remarquons que dans le cas i.i.d., la loi forte des grands nombres entraîne que l'on a :

$$\lambda_i = \mathbf{E}_i \ln \frac{p_1(Z_1)}{p_0(Z_1)} \quad , \; i = 0, 1,$$

si bien que $\lambda_0 < 0$ et $\lambda_1 > 0$.
On a le théorème :

Théorème 3.1 (Lai) *Soient $A_{\alpha\beta}$ et $B_{\alpha\beta}$ des constantes strictement positives telles que $\ln A_{\alpha\beta} \sim \ln \beta$ et $\ln B_{\alpha\beta} \sim \ln \frac{1}{\alpha}$ lorsque $\alpha + \beta \to 0$. Soit $(\tau_{\alpha\beta},\ d_{\alpha\beta})$ le test de Wald de bornes $A_{\alpha\beta}$ et $B_{\alpha\beta}$.*

S'il existe des nombres $\lambda_0 < 0$ et $\lambda_1 > 0$ tels que $\dfrac{\ln R_n}{n} \underset{n \to +\infty}{\longrightarrow} \lambda_i$ r-rapidement sous $\mathbf{P}^i$, pour r un réel positif, alors on a :

$$\mathbf{E}_i \tau_{\alpha\beta}^r < +\infty \ , \; i = 0, 1.$$

De plus, lorsque $\alpha + \beta \to 0$, on a :

$$\inf_{(\tau,d) \in \mathcal{T}(\alpha,\beta)} \mathbf{E}_0 \tau^r \quad \sim \quad \mathbf{E}_0 \tau_{\alpha\beta}^r \quad \sim \quad \frac{|\ln \beta|^r}{|\lambda_0|^r}$$

et

$$\inf_{(\tau,d) \in \mathcal{T}(\alpha,\beta)} \mathbf{E}_1 \tau^r \quad \sim \quad \mathbf{E}_1 \tau_{\alpha\beta}^r \quad \sim \quad \frac{|\ln \alpha|^r}{\lambda_1^r} \ .$$

Dans un premier temps, nous donnons une généralisation de ce résultat au cas de m hypothèses ($m \geq 2$), puis nous faisons le lien avec un travail de Dragalin et Novikov ; nous étendons ensuite la notion de convergence r-rapide et en déduisons une propriété d'optimalité asymptotique liée à cette convergence.

3.2 Une extension d'un théorème de Lai et ses applications à l'analyse séquentielle

3.2.1 Enoncé du résultat principal

Soit $\Omega = \mathbf{R}^{\mathbf{N}^*}$ l'espace des suites à valeurs réelles. Nous désignons par $X_n(\omega)$ la $n^{\text{ème}}$ coordonnée de $\omega \in \Omega$. Nous munissons Ω des σ-algèbres suivantes :

$$\mathcal{F} = (\mathcal{B}(\mathbf{R}))^{\otimes \mathbf{N}^*} \quad , \quad \mathcal{F}_n = \sigma\{X_1, \cdots, X_n\} \; ; \; n \geq 1,$$

où $\mathcal{B}(\mathbf{R})$ représente la tribu borélienne de $\mathbf{R}$. La loi de la suite $(X_n)_{n\geq 1}$ est notée $\mathbf{P}^\theta$, pour θ un paramètre inconnu appartenant à l'ensemble $\Theta = \{\theta_1, \cdots, \theta_m\}$ $(m \geq 2)$. Soit $\mathbf{Q}$ une mesure σ-finie sur $(\Omega, \mathcal{F})$ telle que $\mathbf{P}^{\theta_i} \ll \mathbf{Q}$, $1 \leq i \leq m$. Posons $\mathbf{Q}_n = \mathbf{Q}|_{\mathcal{F}_n}$ comme étant la restriction de $\mathbf{Q}$ à la σ-algèbre $\mathcal{F}_n$, $n \geq 1$. Considérons les m hypothèses

$$H_1 : \theta = \theta_1; \cdots ; H_m : \theta = \theta_m \; .$$

Soit $\mathbf{P}^i = \mathbf{P}^{\theta_i}$ la loi de $(X_n)_{n\geq 1}$ si $\theta = \theta_i$, $1 \leq i \leq m$. On suppose ces mesures toutes localement équivalentes. Soit $p_{in}(x_1, \cdots, x_n)$ la densité jointe de $(X_1, \cdots, X_n)$ par rapport à $\mathbf{Q}_n$ sous l'hypothèse H_i définie par :

$$p_{in} = \left.\frac{d\mathbf{P}^i}{d\mathbf{Q}}\right|_{\mathcal{F}_n} \; , \; n \geq 1, \; 1 \leq i \leq m.$$

Soit $\mathbf{P}$ une probabilité sur $(\Omega, \mathcal{F})$ localement équivalente aux $\mathbf{P}^i$, $1 \leq i \leq m$. On suppose que sous $\mathbf{P}$, la densité de $(X_1, \cdots, X_n)$ par rapport à $\mathbf{Q}_n$ est notée p_n , $n \geq 1$. On pose

$$R_n^{(i)} = \frac{p_n(X_1, \cdots, X_n)}{p_{in}(X_1, \cdots, X_n)} \quad , \quad R_n^{(ij)} = \frac{p_{jn}(X_1, \cdots, X_n)}{p_{in}(X_1, \cdots, X_n)} \; ; \quad 1 \leq i, j \leq m.$$

Un test des hypothèses $H_1 : \theta = \theta_1; \cdots ; H_m : \theta = \theta_m$ est défini par :
– un $(\mathcal{F}_n)_{n\geq 1}$-temps d'arrêt τ
– une fonction de décision d, $\mathcal{F}_\tau$-mesurable, à valeurs dans $\{1, \cdots, m\}$.

Pour $0 < \alpha_1, \cdots, \alpha_m < 1$, on pose $\alpha = (\alpha_1, \cdots, \alpha_m)$ et on définit la classe des tests

$$\mathcal{T}(\alpha) = \Big\{(\tau, d) \; : \; \mathbf{P}^i(d \neq i) \leq \alpha_i \; , \; 1 \leq i \leq m\Big\} \; .$$

Considérons les conditions suivantes :
(C1) il existe des constantes positives $\eta_1, \cdots, \eta_m$ dont l'une au plus est nulle et telles que

$$\frac{\ln R_n^{(i)}}{n} \underset{n\to+\infty}{\longrightarrow} \eta_i \quad \mathbf{P} - p.s. \; , \; 1 \leq i \leq m.$$

Posons $\nu_i = \dfrac{|\ln \alpha_i|}{\eta_i}$, $1 \leq i \leq m$. Soit $\nu_{(1)} \leq \cdots \leq \nu_{(m-1)} \leq \nu_{(m)}$ la suite ordonnée associée à $(\nu_1, \cdots, \nu_m)$. Nous définissons également $\alpha_{(i)}$ et $\eta_{(i)}$ par la relation :

$$\nu_{(i)} := \frac{|\ln \alpha_{(i)}|}{\eta_{(i)}} \ , \ 1 \leq i \leq m.$$

(C2) les α_i tendent vers 0 de sorte que $\nu_{(1)}, \nu_{(2)}, \cdots, \nu_{(m-1)}$ soient équivalents, i.e.

$$\frac{\nu_{(i)}}{\nu_{(j)}} \xrightarrow[\alpha_1+\cdots+\alpha_m \to 0]{} 1 \quad , \ 1 \leq i, j \leq m-1.$$

On notera cette convergence par

$$\alpha \to 0.$$

On a le théorème suivant :

Théorème 3.2 *Supposons réalisées les conditions (C1) et (C2).*

(i) Pour tout $0 < \delta < 1$, on a :

$$\inf_{(\tau,d) \in \mathcal{T}(\alpha)} \mathbf{P}\{\tau > \delta \nu_{(m-1)}\} \xrightarrow[\alpha \to 0]{} 1.$$

(ii) A $\quad 0 < \alpha_1, \cdots, \alpha_m < 1$ on associe des constantes positives $C_1, \cdots, C_m$ telles que $\ln C_i \sim |\ln \alpha_i|$ lorsque $\alpha \to 0$, $1 \leq i \leq m$.
Soit

$$\sigma_i = \inf\{n \geq 1 : R_n^{(i)} \geq C_i\}$$

le temps de rejet de l'hypothèse H_i , pour $1 \leq i \leq m$. On pose

$$\tau^* = \min_j \max_{i \neq j} \sigma_i \quad \text{et} \quad d^* = \arg\max_i \sigma_i$$

(si plusieurs indices i fournissent ce maximum, on prendra le plus petit indice comme décision).
Alors, lorsque $\alpha \to 0$, on a :

$$\frac{\tau^*}{\nu_{(m-1)}} \longrightarrow 1 \quad \mathbf{P} - p.s.$$

et, par conséquent, pour tout $0 < \delta < 1$, on a :

$$\inf_{(\tau,d) \in \mathcal{T}(\alpha)} \mathbf{P}\{\tau > \delta \tau^*\} \xrightarrow[\alpha \to 0]{} 1.$$

Remarque 3.1 *Dans le cas i.i.d., la loi des grands nombres entraîne que l'on a :*

$$\eta_i = \mathbf{E} \ln R_1^{(i)} \geq 0$$

et $\eta_i = 0$ équivaut alors à $\mathbf{P}^i = \mathbf{P}$. Cette égalité ne pouvant avoir lieu qu'au plus une fois, la condition (C1) est justifiée.

Remarque 3.2 *Le cas $m = 2$ est résolu dans [La2]. Nous en donnons une généralisation nécessitant simplement un contrôle de la convergence des α_i vers 0 à l'aide de la condition (C2). Il est à noter que pour $m = 2$, celle-ci est trivialement vérifiée.*

Remarque 3.3 *La condition (C2) implique l'existence de réels $K_{(1)}, \cdots, K_{(m-1)}$ tels que*

$$|\ln \alpha_{(j)}| \sim K_{(j)} |\ln \alpha_{(1)}| \qquad (\alpha_{(1)} \to 0)\ ,\ 1 \leq j \leq m-1.$$

En effet :

$$\frac{|\ln \alpha_{(j)}|}{|\ln \alpha_{(1)}|} = \frac{\frac{|\ln \alpha_{(j)}|}{\eta_{(j)}}.\eta_{(j)}}{\frac{|\ln \alpha_{(1)}|}{\eta_{(1)}}.\eta_{(1)}} \sim \frac{\eta_{(j)}}{\eta_{(1)}} := K_{(j)} \qquad (\alpha_{(1)} \to 0)\ ,\ 1 \leq j \leq m-1.$$

Elle n'est pas plus restrictive que la condition de Dragalin et Novikov (voir [Dr-No] et le paragraphe 3), dans la mesure où l'un des risques $(\alpha_{(m)})$ tend vers 0 sans contrainte.

Le Théorème 3.2 établit une première forme d'optimalité pour une probabilité intermédiaire $\mathbf{P}$, non nécessairement égale à l'une des $\mathbf{P}^i$, $1 \leq i \leq m$.

3.2.2 Démonstration du Théorème 3.2

(i) On pose $l_n^{(i)} = \ln R_n^{(i)}\ ,\ 1 \leq i \leq m$. Soient $\delta\ ,\ \overline{\delta}$ deux réels tels que

$$0 < \delta < 1\ ,\ \overline{\delta} > 1 \text{ et } \delta\overline{\delta} < 1.$$

Soit $k = [\delta \nu_{(m-1)}]$ la partie entière de $\delta \nu_{(m-1)}$. Considérons $(\tau, d) \in \mathcal{T}(\alpha)$. Pour $1 \leq i \leq m$, on a :

$$\begin{aligned}
\alpha_i \geq & \mathbf{P}^i\{(\tau,d) \textit{ rejette } H_i\} \\
= & \int_{\{\tau<+\infty,(\tau,d) \textit{ rejette } H_i\}} d\mathbf{P}^i \\
= & \int_{\{\tau<+\infty,(\tau,d) \textit{ rejette } H_i\}} \mathbf{e}^{-l_\tau^{(i)}} d\mathbf{P} \\
\geq & \int_{\{\tau\leq k, l_\tau^{(i)} \leq \overline{\delta}\eta_i k,(\tau,d) \textit{ rejette } H_i\}} \mathbf{e}^{-l_\tau^{(i)}} d\mathbf{P} \\
\geq & \mathbf{e}^{-\overline{\delta}\eta_i k}\ \mathbf{P}\{\tau \leq k, l_\tau^{(i)} \leq \overline{\delta}\eta_i k, (\tau,d) \textit{ rejette } H_i\}\ .
\end{aligned}$$

En vertu de la condition (C2), on a :

$$\overline{\delta}\eta_i k \leq \overline{\delta}\delta\eta_i\nu_{(m-1)} \leq \overline{\delta}\delta\eta_i\nu_i(1+o(1)) = \overline{\delta}\delta|\ln \alpha_i|(1+o(1)).$$

On en déduit :

$$\begin{aligned}\mathbf{P}\{\tau \le k, (\tau,d) \text{ rejette } H_i\} &= \mathbf{P}\{\tau \le k, (\tau,d) \text{ rejette } H_i, l_\tau^{(i)} > \overline{\delta}\eta_i k\} \\ &\quad + \mathbf{P}\{\tau \le k, (\tau,d) \text{ rejette } H_i, l_\tau^{(i)} \le \overline{\delta}\eta_i k\} \\ &\le \mathbf{P}\{\tau \le k, l_\tau^{(i)} > \overline{\delta}\eta_i k\} + \alpha_i \mathbf{e}^{\overline{\delta}\eta_i k} \\ &\le \mathbf{P}\{\max_{j\le k} l_j^{(i)} > \overline{\delta}\eta_i k\} + \alpha_i^{1-\overline{\delta}\delta(1+o(1))} \ .\end{aligned}$$

Il vient alors :

$$\begin{aligned}\mathbf{P}\{\tau \le k\} &\le \sum_{i=1}^m \mathbf{P}\{\tau \le k, (\tau,d) \text{ rejette } H_i\} \\ &\le \alpha_1^{1-\overline{\delta}\delta(1+o(1))} + \cdots + \alpha_m^{1-\overline{\delta}\delta(1+o(1))} \\ &\quad + \mathbf{P}\{\max_{j\le k} l_j^{(1)} > \overline{\delta}\eta_1 k\} + \cdots + \mathbf{P}\{\max_{j\le k} l_j^{(m)} > \overline{\delta}\eta_m k\} \ .\end{aligned}$$

Comme $\overline{\delta} > 1$, les probabilités tendent vers 0 (en effet, k tend vers l'infini lorsque les risques tendent vers zéro et la condition (C1) permet de conclure). Il en est de même pour chacun des $\alpha_i^{1-\overline{\delta}\delta(1+o(1))}$, puisque les exposants sont strictement positifs pour $\alpha \to 0$. Il est à noter que la majoration de $\mathbf{P}\{\tau \le k\}$ est indépendante de (τ, d), si bien que la borne est uniforme et il en résulte

$$\sup_{(\tau,d)\in\mathcal{T}(\alpha)} \mathbf{P}\{\tau \le k\} \xrightarrow[\alpha\to 0]{} 0 \ ,$$

ou encore

$$\inf_{(\tau,d)\in\mathcal{T}(\alpha)} \mathbf{P}\{\tau > \delta\nu_{(m-1)}\} \xrightarrow[\alpha\to 0]{} 1 \ .$$

(ii) Dans le cas de deux hypothèses, $\tau^* = \sigma_1 \wedge \sigma_2$ est le minimum de deux temps de rejet et la convergence de $\dfrac{\tau^*}{\min\{\nu_1,\nu_2\}}$ est facile à établir. Nous donnons un lemme qui assure une convergence analogue pour des temps d'arrêt qui sont des minima de temps de rejet. La deuxième assertion du théorème sera démontrée en itérant ce lemme.

Lemme 3.1 *On suppose réalisées les conditions (C1) et (C2). Soit k un entier positif non nul. Soient $0 < \alpha_{i_1}, \cdots, \alpha_{i_k} < 1$ et $C_{i_1}, \cdots, C_{i_k}$ des constantes positives telles que $\ln C_{i_j} \sim |\ln \alpha_{i_j}|$, $1 \le j \le k$. Posons*

$$T_{\alpha_{i_1},\cdots,\alpha_{i_k}} = \inf\{n \ge 1 : R_n^{(i_1)} \ge C_{i_1} \, ou \cdots ou \, R_n^{(i_k)} \ge C_{i_k}\} = \sigma_{i_1} \wedge \cdots \wedge \sigma_{i_k}.$$

Soit $(T_{\alpha_{i_1},\cdots,\alpha_{i_k}}, d)$ la procédure qui arrête les observations à l'instant $T_{\alpha_{i_1},\cdots,\alpha_{i_k}}$ et rejette H_{i_j} si et seulement si $R^{(i_j)}_{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}} \ge C_{i_j}$, $1 \le j \le k$ (nous adoptons la convention suivante : si une telle inégalité a lieu pour plusieurs indices, on ne rejettera que l'hypothèse dont l'indice est le plus petit).

Alors, lorsque $(\alpha_{i_1}, \cdots, \alpha_{i_k}) \to 0$, *on a* [1] *:*

$$\frac{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}{\min\{\frac{|\ln\alpha_{i_1}|}{\eta_{i_1}}, \cdots, \frac{|\ln\alpha_{i_k}|}{\eta_{i_k}}\}} \longrightarrow 1 \quad \mathbf{P}\text{-}p.s.$$

Pour démontrer ce lemme, il suffit de remarquer que si $\frac{|\ln\alpha_{i_j}|}{\eta_{i_j}}$ est le minimum de $\{\frac{|\ln\alpha_{i_1}|}{\eta_{i_1}}, \cdots, \frac{|\ln\alpha_{i_k}|}{\eta_{i_k}}\}$ et si son indice est minimal, cette procédure rejettera l'hypothèse H_{i_j}. En effet, les équivalences suivantes ont lieu, au moins asymptotiquement :

$$\begin{aligned}
\frac{|\ln\alpha_{i_j}|}{\eta_{i_j}} \leq \frac{|\ln\alpha_{i_l}|}{\eta_{i_l}} &\Longleftrightarrow \frac{\ln C_{i_j}}{\eta_{i_j}} \leq \frac{\ln C_{i_l}}{\eta_{i_l}} \\
&\Longleftrightarrow \frac{\ln C_{i_j}}{\ln R^{(i_j)}_{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}} \frac{\ln R^{(i_j)}_{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}}{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}} \frac{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}{\eta_{i_j}} \\
&\qquad \leq \frac{\ln C_{i_l}}{\ln R^{(i_l)}_{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}} \frac{\ln R^{(i_l)}_{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}}{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}} \frac{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}{\eta_{i_l}} \\
&\Longleftrightarrow \frac{\ln C_{i_j}}{\ln R^{(i_j)}_{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}} \leq \frac{\ln C_{i_l}}{\ln R^{(i_l)}_{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}} \ , \ 1 \leq l \leq k \ .
\end{aligned}$$

Par définition de $T_{\alpha_{i_1},\cdots,\alpha_{i_k}}$, l'un au moins de ces rapports est inférieur à 1. Ceci équivaut à dire que le plus petit d'entre eux l'est. On a alors :

$$\frac{\ln C_{i_j}}{\ln R^{(i_j)}_{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}} \leq 1 \Longleftrightarrow R^{(i_j)}_{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}} \geq C_{i_j}$$

et ceci revient à rejeter l'hypothèse H_{i_j} à l'instant $T_{\alpha_{i_1},\cdots,\alpha_{i_k}}$. Mais alors on a :

$$\begin{aligned}
\frac{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}{\frac{|\ln\alpha_{i_j}|}{\eta_{i_j}}} &= \frac{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}{\ln R^{(i_j)}_{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}} \frac{\ln R^{(i_j)}_{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}}{\frac{|\ln\alpha_{i_j}|}{\eta_{i_j}}} \\
&\sim \frac{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}{\ln R^{(i_j)}_{T_{\alpha_{i_1},\cdots,\alpha_{i_k}}}} \frac{\ln C_{i_j}}{\frac{|\ln\alpha_{i_j}|}{\eta_{i_j}}} \longrightarrow 1 \quad \mathbf{P}\text{-p.s.}
\end{aligned}$$

Le Lemme 3.1 étant démontré, nous revenons à la démonstration du Théorème 3.2.

Comme précédemment, nous désignons par $\nu_{(i)}$ le $i^{\text{ème}}$ terme de la suite ordonnée associée à $(\nu_1, \cdots, \nu_m)$ et nous définissons $\alpha_{(i)}$ et $\nu_{(i)}$ par la relation :

$$\nu_{(i)} := \frac{|\ln\alpha_{(i)}|}{\eta_{(i)}} \ , \ 1 \leq i \leq m$$

[1] Cette notation est à interpréter de la même façon que la notation $\alpha \to 0$ introduite dans la condition (C2).

sans lien aucun avec les suites ordonnées associées respectivement à $(\alpha_1, \cdots, \alpha_m)$ et $(\eta_1, \cdots, \eta_m)$. De même, les constantes $C_{(i)}$ sont liées aux $\alpha_{(i)}$ par l'équivalence entre $\ln C_{(i)}$ et $|\ln \alpha_{(i)}|$. Nous désignons enfin par $H_{(i)}$ l'hypothèse correspondant à l'indice de $\nu_{(i)}$ (si plusieurs indices k vérifient $\nu_k = \nu_{(i)}$, on ne retiendra que le plus petit d'entre eux) et par $\sigma_{(i)}$ le temps de rejet de l'hypothèse $H_{(i)}$.

Etape 1. Posons

$$T^1 = \inf\{n \geq 1 : R_n^{(1)} \geq C_1 \ ou \cdots ou \ R_n^{(m)} \geq C_m\} = \sigma_1 \wedge \cdots \wedge \sigma_m \ .$$

D'après le lemme précédent, $\frac{T^1}{\nu_{(1)}}$ converge $\mathbf{P}$-presque sûrement vers 1 lorsque $\alpha \to 0$ et nous sommes amenés à rejeter $H_{(1)}$.

Etape 2. Posons

$$T^2 = \inf\{n \geq T^1 : R_n^{((2))} \geq C_{(2)} \ ou \cdots ou \ R_n^{((m))} \geq C_{(m)}\} = \sigma_{(2)} \wedge \cdots \wedge \sigma_{(m)}.$$

Le même lemme assure alors que $\frac{T^2}{\nu_{(2)}}$ converge $\mathbf{P}$-presque sûrement vers 1 lorsque $\alpha \to 0$ et nous rejetons $H_{(2)}$.

Et ainsi de suite jusqu'à l'étape $m-1$. Posons

$$\begin{aligned} T^{m-1} &= \inf\{n \geq T^{m-2} : R_n^{((m-1))} \geq C_{(m-1)} \ ou \ R_n^{((m))} \geq C_{(m)}\} \\ &= \sigma_{(m-1)} \wedge \sigma_{(m)} = \tau^* \ . \end{aligned}$$

Il vient alors que $\frac{\tau^*}{\nu_{(m-1)}}$ converge $\mathbf{P}$-presque sûrement vers 1 lorsque $\alpha \to 0$ et l'hypothèse rejetée à ce stade est $H_{(m-1)}$.

En combinant ce dernier résultat avec l'assertion (i), pour tout $0 < \delta < 1$, nous obtenons

$$\inf_{(\tau,d)\in\mathcal{T}(\alpha)} \mathbf{P}\{\tau > \delta\tau^*\} \xrightarrow[\alpha\to 0]{} 1 \ .$$

Ceci achève la démonstration du Théorème 3.2.

3.2.3 Application à l'analyse séquentielle

Nous donnons deux corollaires du Théorème 3.2 qui établissent l'optimalité asymptotique du test séquentiel par rapport au moment d'ordre r de la durée d'observation, le premier sous la probabilité intermédiaire $\mathbf{P}$ et le second particularisant le résultat aux mesures $\mathbf{P}^j$, $1 \leq j \leq m$.

Corollaire 3.1 *On suppose réalisées les conditions (C1) et (C2). On suppose de plus que* $\frac{\ln R_n^{(i)}}{n} \xrightarrow[n\to+\infty]{} \eta_i$ *r-rapidement sous* $\mathbf{P}$ *pour* $r > 0$, $1 \leq i \leq m$.
On a alors $\mathbf{E}(\tau^*)^r < +\infty$ *et, lorsque* $\alpha \to 0$ *, on a :*

$$\inf_{(\tau,d)\in\mathcal{T}(\alpha)} \mathbf{E}\tau^r \sim \mathbf{E}(\tau^*)^r \sim (\nu_{(m-1)})^r \ .$$

Corollaire 3.2 *On suppose réalisées les conditions (C1) et (C2). On suppose de plus que* $\frac{\ln R_n^{(ij)}}{n} \underset{n\to+\infty}{\longrightarrow} \lambda_{ij}$ *r-rapidement sous* $\mathbf{P}^j$ *pour* $r > 0$, $1 \leq i, j \leq m$.
On a alors $\mathbf{E}_j(\tau^*)^r < +\infty$ *et lorsque* $\alpha \to 0$ *, on a :*

$$\inf_{(\tau,d)\in\mathcal{T}(\alpha)} \mathbf{E}_j\tau^r \sim \mathbf{E}_j(\tau^*)^r \sim (\max_{i\neq j}\left|\frac{\ln\alpha_i}{\lambda_{ij}}\right|)^r \ , \ 1 \leq j \leq m.$$

Démonstration des corollaires. Le Corollaire 3.2 s'obtient en posant $\mathbf{P} = \mathbf{P}^j$ dans le Corollaire 3.1 et en remarquant qu'alors on a :

$$\lambda_{jj} = 0 \quad , \ 1 \leq j \leq m \ .$$

Ainsi, il en résulte [2] :

$$\nu_{(m-1)} = \max_{i\neq j}\left|\frac{\ln\alpha_i}{\lambda_{ij}}\right| \ .$$

Nous donnons à présent la démonstration du Corollaire 3.1.

Nous supposons $\eta_1 > 0, \cdots, \eta_m > 0$, car si $\eta_k = 0$, cette procédure conduira à l'acceptation de H_k à l'instant

$$\tilde{\tau}^* = \max_{i\neq k}\sigma_i \ .$$

Un raisonnement analogue à celui qui suit montre que l'on a les équivalences suivantes :

$$\inf_{(\tau,d)\in\mathcal{T}(\alpha)} \mathbf{E}\tau^r \sim \mathbf{E}(\tilde{\tau}^*)^r \sim (\max_{i\neq k}\nu_i)^r \ .$$

Posons

$$L = \sup\{n \geq 1 : \max_{i=1}^{m}\left|\frac{l_n^{(i)}}{n} - \eta_i\right| > a\},$$

pour $0 < a < \min\{\eta_1, \cdots, \eta_m\}$ fixé. Si $\eta_k = 0$, on procède de même avec

$$\tilde{L} = \sup\{n \geq 1 : \max_{i\neq k}\left|\frac{l_n^{(i)}}{n} - \eta_i\right| > a\} \ .$$

Sur l'événement $\{\sigma_i - 1 > L\}$, on a :

$$\left|\frac{l_{\sigma_i-1}^{(i)}}{\sigma_i - 1} - \eta_i\right| \leq a \ , \ \text{i.e.}$$

[2] En toute rigueur, il faudrait définir

$$\nu_{i,j} := \left|\frac{\ln\alpha_i}{\lambda_{ij}}\right|$$

et utiliser la notation $\nu_{(m-1),j}$ au lieu de $\nu_{(m-1)}$.

$$\eta_i - a \leq \frac{l^{(i)}_{\sigma_i - 1}}{\sigma_i - 1} \leq \eta_i + a \Longrightarrow (\eta_i - a)(\sigma_i - 1) \leq l^{(i)}_{\sigma_i - 1} \leq \ln C_i$$
$$\Longrightarrow \sigma_i \leq 1 + \frac{\ln C_i}{\eta_i - a} \ .$$

Il en résulte alors

$$\sigma_i \leq L + 1 + \frac{\ln C_i}{\eta_i - a} \ , \ 1 \leq i \leq m,$$

$$\text{i.e.} \qquad \tau^* = \min_j \max_{i \neq j} \sigma_i \leq L + 1 + \min_j \max_{i \neq j} \left\{ \frac{\ln C_i}{\eta_i - a} \right\} \ .$$

Comme $\dfrac{l^{(i)}_n}{n}$ converge r-rapidement vers η_i sous $\mathbf{P}$, $1 \leq i \leq m$, on a $\mathbf{E}L^r < +\infty$ et par conséquent, $\mathbf{E}(\tau^*)^r < +\infty$.

D'autre part, le Théorème 3.2 assure que $\dfrac{\tau^*}{\nu_{(m-1)}}$ tend $\mathbf{P}$-p.s. vers 1 lorsque $\alpha \to 0$ et le théorème de convergence dominée entraîne que l'on a :

$$\mathbf{E}\Big(\frac{\tau^*}{\nu_{(m-1)}}\Big)^r \xrightarrow[\alpha \to 0]{} 1,$$

ou encore

$$\mathbf{E}(\tau^*)^r \sim (\nu_{(m-1)})^r \ .$$

Pour tout $0 < \delta < 1$, l'inégalité

$$\inf_{(\tau,d) \in \mathcal{T}(\alpha)} \mathbf{E}\tau^r \geq \delta^r (\nu_{(m-1)})^r (1 + o(1))$$

a lieu (elle résulte de l'assertion (ii) du Théorème 3.2). En faisant tendre δ vers 1, on obtient alors

$$\inf_{(\tau,d) \in \mathcal{T}(\alpha)} \mathbf{E}\tau^r \sim \mathbf{E}(\tau^*)^r \sim (\nu_{(m-1)})^r \ .$$

Remarque 3.4 *Les probabilités d'erreur associées à (τ^*, d^*) satisfont asymptotiquement à*

$$\mathbf{P}^i\{(\tau^*, d^*) \ rejette \ H_i\} \leq \alpha_i \ , \ 1 \leq i \leq m \ .$$

En effet :

$$\mathbf{P}^i\{(\tau^*, d^*) \ rejette \ H_i\} = \mathbf{P}^i\{d^* \neq i\} \leq \mathbf{P}^i\{\sigma_i < +\infty\}$$

car on a $\quad \mathbf{E}_i(\tau^*)^r < +\infty$ *et* $\sigma_i \leq \tau^*$. *On en déduit les inégalités suivantes :*

$$\begin{aligned} \mathbf{P}^i\{(\tau^*, d^*) \ rejette \ H_i\} &\leq \int_{\{\sigma_i < +\infty\}} d\mathbf{P}^i = \int_{\{\sigma_i < +\infty\}} \mathbf{e}^{-l^{(i)}_{\sigma_i}} \, d\mathbf{P} \\ &\leq C_i^{-1} \ \mathbf{P}\{\sigma_i < +\infty\} \\ &\leq C_i^{-1} \sim (\mathbf{e}^{|\ln \alpha_i|})^{-1} = \alpha_i \ , \ 1 \leq i \leq m \ . \end{aligned}$$

Nous venons de montrer qu'asymptotiquement (τ^*, d^*) appartient à la classe de règles de décision $\mathcal{T}(\alpha)$.

Nous donnons à présent deux exemples de tests de m hypothèses pour lesquels l'optimalité asymptotique du test de Wald a lieu (voir [La2] pour le cas $m = 2$).

Exemple 1. Soient $Z, Z_1, Z_2, \cdots$ des variables aléatoires i.i.d., de loi $\mathcal{N}(\mu;\sigma)$. On pose $\gamma = \dfrac{\mu}{\sigma}$.
On considère les hypothèses $H_1 : \gamma = \gamma_1; \cdots; H_m : \gamma = \gamma_m$. Les rapports de vraisemblance sont alors donnés par $R_n^{(ij)} = \dfrac{U_n(\gamma_j)}{U_n(\gamma_i)}$, où :

$$U_n(\gamma) = \int_0^\infty u^{-1} \exp[n\, f(u, T_n, \gamma)] du \ ;$$

$$T_n = \frac{\overline{Z_n}}{\sqrt{\frac{1}{n}\sum_{k=1}^n Z_k^2}} \ ;$$

$$f(u, y, \gamma) = -\frac{1}{2}u^2 + \gamma\, y\, u + \ln u - \frac{1}{2}\gamma^2 \ .$$

Posant

$$h(\gamma, u) = -\frac{1}{2}\gamma^2 + \frac{1}{4}(\gamma^2\, u^2 + \gamma\, u\, \sqrt{\gamma^2\, u^2 + 4}) + \ln(\gamma\, u + \sqrt{\gamma^2\, u^2 + 4})$$

et

$$\Psi_{ij}(u) = \ln h(\gamma_j, u) - h(\gamma_i, u) \ ,$$

on peut montrer qu'il existe des constantes C_{ij} telles que l'on ait :

$$|\ln R_n^{(ij)} - n\, \Psi_{ij}(T_n)| \le C_{ij} \quad , \ n \ge 1\,, \ 1 \le i, j \le m\,.$$

Comme $\mathcal{L}(Z) = \mathcal{N}(\mu;\sigma)$, $\mathbf{E}_j|Z|^s < +\infty$ pour tout $s > 0$, donc $(\overline{Z_n})_{n\ge1}$ converge vers $\mathbf{E}_j Z$ et $\Big(\sqrt{\dfrac{1}{n}\sum\limits_{k=1}^n Z_k^2}\Big)_{n\ge1}$ converge vers $\sqrt{\mathbf{E}_j Z^2}$, si bien que T_n converge vers $\dfrac{\mathbf{E}_j Z}{\sqrt{\mathbf{E}_j Z^2}}$ r-rapidement sous $\mathbf{P}^j$ pour tout $r > 0$, $1 \le j \le m$.

Mais $\mathbf{E}_j Z = \mu = \gamma_j\, \sigma$ et $\mathbf{E}_j Z^2 = \sigma^2 + \mu^2 = \sigma^2(1 + \gamma_j^2)$, donc la limite de T_n est $\dfrac{\gamma_j}{\sqrt{1+\gamma_j^2}}$, $1 \le j \le m$. L'estimation précédente montre que $\dfrac{\ln R_n^{(ij)}}{n}$ et $\Psi_{ij}(T_n)$ ont la même limite, à savoir $\Psi_{ij}\Big(\dfrac{\gamma_j}{\sqrt{1+\gamma_j^2}}\Big)$.

Exemple 2. Soit $(X_n)_{n\ge1}$ un processus autorégressif d'ordre p, gaussien et stationnaire, de niveau moyen θ inconnu. On a alors $\quad X_n = Y_n + \theta$, où

$$Y_n = \beta_1\, Y_{n-1} + \cdots + \beta_p\, Y_{n-p} + Z_n \ , \quad n > p \ ,$$

et $Z_{p+1}, Z_{p+2}, \cdots$ sont des variables aléatoires i.i.d. de loi $\mathcal{N}(0;\sigma)$, indépendantes de $Y_1, \cdots, Y_p$, dont la densité jointe est $g(y_1, \cdots, y_p)$. On suppose la densité g et les paramètres $\beta_1, \cdots, \beta_p$ connus, et que $\sigma = 1$.
La matrice du changement de variables $(Y_1, \cdots, Y_n) \longmapsto (Y_1, \cdots, Y_p, Z_{p+1}, \cdots, Z_n)$ est triangulaire inférieure et sa diagonale ne comporte que des "1". La densité jointe de $(Y_1, \cdots, Y_n)$ est alors le produit de celle de $(Y_1, \cdots, Y_p)$ par celle de $(Z_{p+1}, \cdots, Z_n)$, dans laquelle on a remplacé Z_j par $Y_j - \beta_1 Y_{j-1} - \cdots - \beta_p Y_{j-p}$. On en déduit la densité de $(X_1, \cdots, X_n)$ en substituant Y_j par $X_j - \theta$ dans celle de $(Y_1, \cdots, Y_n)$:

$$\begin{aligned}
\frac{d\mathbf{P}^\theta}{d\mathbf{P}}(X_1, \cdots, X_n) &= g(X_1 - \theta, \cdots, X_p - \theta) \frac{1}{(\sqrt{2\pi})^{n-p}} \times \\
&\times \exp\left\{ -\frac{1}{2} \sum_{k=p+1}^{n} \Big((X_k - \theta) - \beta_1 (X_{k-1} - \theta) - \cdots - \beta_p (X_{k-p} - \theta) \Big)^2 \right\} \\
&= g(X_1 - \theta, \cdots, X_p - \theta) \frac{1}{(\sqrt{2\pi})^{n-p}} \times \\
&\times \exp\left\{ -\frac{1}{2} \sum_{k=p+1}^{n} (X_k - \beta_1 X_{k-1} - \cdots - \beta_p X_{k-p})^2 - \right. \\
&\quad - \frac{1}{2}(n-p)(1 - \beta_1 - \cdots - \beta_p)^2 \theta^2 + \\
&\quad \left. + \sum_{k=p+1}^{n} (X_k - \beta_1 X_{k-1} - \cdots - \beta_p X_{k-p})\, (1 - \beta_1 - \cdots - \beta_p)\theta \right\} .
\end{aligned}$$

Donc on a :

$$\begin{aligned}
\ln R_n^{(ij)} = \ln \frac{g(X_1 - \theta_j, \cdots, X_p - \theta_j)}{g(X_1 - \theta_i, \cdots, X_p - \theta_i)} - \frac{1}{2}(n-p)(1 - \beta_1 - \cdots - \beta_p)^2 (\theta_j^2 - \theta_i^2) + \\
+ (1 - \beta_1 - \cdots - \beta_p)(\theta_j - \theta_i) \sum_{k=p+1}^{n} (X_k - \beta_1 X_{k-1} - \cdots - \beta_p X_{k-p})
\end{aligned}$$

pour $1 \le i, j \le m$. Mais sous $\mathbf{P}^j$, $1 \le j \le m$, on a :

$$\begin{aligned}
X_k - \beta_1 X_{k-1} - \cdots - \beta_p X_{k-p} &= (Y_k + \theta_j) - \beta_1 (Y_{k-1} + \theta_j) - \cdots - \beta_p (Y_{k-p} + \theta_j) \\
&= Z_k + (1 - \beta_1 - \cdots - \beta_p)\theta_j ,
\end{aligned}$$

donc également :

$$\sum_{k=p+1}^{n} (X_k - \beta_1 X_{k-1} - \cdots - \beta_p X_{k-p}) = (n-p)\,(1 - \beta_1 - \cdots - \beta_p)\theta_j + \sum_{k=p+1}^{n} Z_k \ .$$

Comme les (Z_k) sont i.i.d. de loi normale, ils admettent des moments de tous ordres, si bien que $\displaystyle\frac{1}{n} \sum_{k=p+1}^{n} Z_k \underset{n \to +\infty}{\longrightarrow} 0$ r-rapidement sous $\mathbf{P}^j$, $1 \le j \le m$, pour tout $r > 0$.

D'autre part, $\dfrac{1}{n} \ln \dfrac{g(X_1 - \theta_j, \cdots, X_p - \theta_j)}{g(X_1 - \theta_i, \cdots, X_p - \theta_i)} \underset{n \to +\infty}{\longrightarrow} 0$ r-rapidement sous $\mathbf{P}^j$, $1 \leq j \leq m$, si bien que la limite r-rapide de $\dfrac{\ln R_n^{(ij)}}{n}$ est

$$\begin{aligned}
\lambda_{ij} &= -\frac{1}{2}(1 - \beta_1 - \cdots - \beta_p)^2(\theta_j^2 - \theta_i^2) + (1 - \beta_1 - \cdots - \beta_p)^2(\theta_j - \theta_i)\theta_j \\
&= (1 - \beta_1 - \cdots - \beta_p)^2(\theta_j - \theta_i)[-\frac{1}{2}(\theta_j + \theta_i) + \theta_j] \\
&= \frac{1}{2}(1 - \beta_1 - \cdots - \beta_p)^2(\theta_j - \theta_i)^2 , \quad 1 \leq i, j \leq m.
\end{aligned}$$

Le Corollaire 3.2 permet alors d'établir l'optimalité asymptotique du test de Wald dans les deux exemples ci-dessus.

3.3 Lien avec l'article de Dragalin et Novikov

Une propriété d'optimalité asymptotique pour des tests adaptatifs a été établie par Dragalin et Novikov (cf. [Dr-No]). Ce résultat vaut pour des d'hypothèses composites lorsque la vraisemblance est d'un type particulier et lorsque les probabilités d'erreur $\alpha_1, \cdots, \alpha_m$ tendent vers zéro de la façon suivante :

$$|\ln \alpha_i| \sim K_i |\ln \alpha_1| \quad , \; \alpha_1 \to 0 \; , \; 1 \le i \le m$$

pour $K_1, \cdots, K_m$ des constantes positives.
Cette condition est similaire à la condition (C2) –voir Remarque 3.3 ci-dessus–, mais elle n'est pas moins restrictive puisque *tous* les risques doivent être contrôlés. Il est intéressant de noter que la même propriété d'optimalité asymptotique est obtenue sous deux conditions similaires mais néanmoins différentes dans le cas de m hypothèses simples. En effet, si nous considérons des variables aléatoires $Y, Y_1, Y_2, \cdots$ i.i.d. sous la probabilité $\mathbf{P}^\theta$, $\theta \in \Theta = \{\theta_1, \cdots, \theta_m\}$, et vérifiant

$$\mathbf{E}_j Y = \mu_j \quad , \quad \mathbf{E}_j Y^2 < \infty \quad , \quad Y_n \; est \; \mathcal{F}_n - mesurable \; ,$$

alors $S_n = Y_1 + \cdots + Y_n$ est un processus adapté à $(\mathcal{F}_n = \sigma\{X_1, \cdots, X_n\})_{n \ge 1}$, de carré intégrable et à accroissements indépendants et stationnaires. Comme dans [Dr-No], nous supposons que l'on a :

$$p_{in}(X_1, \cdots, X_n) = C_n \; . \; \exp\{a(\theta_i)\, S_n - b(\theta_i)\, n\} \quad , \quad 1 \le i \le m$$

où C_n est $\mathcal{F}_n$-adapté et a et b sont deux fonctions sur $\Theta = \{\theta_1, \cdots, \theta_m\}$. On a alors :

$$R_n^{(ij)} = \exp\{(a(\theta_j) - a(\theta_i))S_n - (b(\theta_j) - b(\theta_i))n\} := \exp\{a_{ji} S_n - b_{ji} n\}$$

et

$$\frac{\ln R_n^{(ij)}}{n} = a_{ji} \frac{S_n}{n} - b_{ji} \; .$$

Mais $\dfrac{\ln R_n^{(ij)}}{n}$ converge 1-rapidement sous $\mathbf{P}^j$ si et seulement si $\dfrac{S_n}{n}$ converge 1-rapidement sous $\mathbf{P}^j$ et ceci est équivalent à la finitude de $\mathbf{E}_j X_1^2$ (voir [La1]).
Le Corollaire 3.2 entraîne alors l'optimalité asymptotique du test de Wald par rapport aux durées moyennes d'observation :

$$\inf_{(\tau,d) \in \mathcal{T}(\alpha)} \mathbf{E}_j \tau \sim \mathbf{E}_j \tau^* \sim \max_{i \ne j} \left(\frac{|\ln \alpha_i|}{\lambda_{ij}} \right) \qquad ,$$

pour λ_{ij} la limite au sens de la convergence 1-rapide de $\dfrac{\ln R_n^{(ij)}}{n}$, qui n'est autre que *l'information de Kullback* pour une observation, à savoir $\mathbf{E}_j \ln R_1^{(ij)}$.

Si dans [Dr-No] nous remplaçons le rapport de vraisemblance *adaptatif* par le rapport de vraisemblance usuel, nous constatons que le temps de rejet de l'hypothèse

H_i coïncide avec celui que nous avons défini et ceci entraîne que le temps d'arrêt τ_I considéré par ces auteurs et celui que nous avons dénommé τ^* sont en fait identiques. Dragalin et Novikov ont montré que l'infimum de $\mathbf{E}_\tau$ est asymptotiquement égal à $\dfrac{|\ln \alpha_1|}{\lambda(\theta_j)}$, où $\lambda(\theta_j) := \min\limits_{i \neq j} \left(\dfrac{\lambda_{ij}}{K_i} \right)$.

Mais

$$\min_{i \neq j} \left(\frac{\lambda_{ij}}{K_i} \right) = \frac{1}{\max_{i \neq j} \left(\frac{K_i}{\lambda_{ij}} \right)} \ ,$$

si bien que l'on a :

$$\frac{|\ln \alpha_1|}{\lambda(\theta_j)} = |\ln \alpha_1| \ . \max_{i \neq j} \left(\frac{K_i}{\lambda_{ij}} \right) = \max_{i \neq j} \left(\frac{K_i |\ln \alpha_1|}{\lambda_{ij}} \right) \sim \max_{i \neq j} \left(\frac{|\ln \alpha_i|}{\lambda_{ij}} \right)$$

ce qui n'est rien d'autre que la borne inférieure donnée dans le Corollaire 3.2.

3.4 Une généralisation de la notion de convergence r-rapide : propriétés et application à l'analyse séquentielle

3.4.1

Au lieu de considérer le moment d'ordre r de la durée d'observation τ, nous nous intéressons à l'espérance d'un certain processus croissant A à l'instant τ. Le but est d'établir l'optimalité du test de Wald par rapport aux quantités $\mathbf{E}_j A_\tau$, $1 \leq j \leq m$, sous des conditions de convergence de $\dfrac{\ln R_n^{(ij)}}{n}$ vers λ_{ij} , $1 \leq i, j \leq m$, dans un sens à préciser. Nous considérons la classe de processus suivante :

Définition 3.2 *Un processus déterministe* $(A_t)_{t\geq 0}$ *nul en 0, fini pour tout* $t < +\infty$ *et croissant vers l'infini* appartient à la classe $\mathcal{A}$ *s'il satisfait aux trois conditions suivantes :*

(A1) pour toute v.a. positive T *et tous réels positifs* l *et* m, *on a :*

$$\mathbf{E}A_T < +\infty \iff \mathbf{E}A_{lT+m} < +\infty$$

(A2) pour toute suite de v.a. positives (T_γ) *et toute suite de v.a. strictement positives* (U_γ) *telles que*

$$\frac{T_\gamma}{U_\gamma} \xrightarrow[\gamma \to \gamma_0]{} 1 \ ,$$

on a :

$$\frac{A_{T_\gamma}}{A_{U_\gamma}} \xrightarrow[\gamma \to \gamma_0]{} 1$$

(A3) les quotients

$$\frac{n\, A_n - (n-1)\, A_{n-1}}{A_n} \qquad \text{et} \qquad \frac{A_n - A_{n-1}}{\frac{A_n}{n}}$$

sont bornés inférieurement et supérieurement par deux constantes positives pour n *suffisamment grand.*

Définition 3.3 *Soit* $A = (A_t)_{t\geq 0}$ *un processus appartenant à la classe* $\mathcal{A}$. *On dit qu'un processus* $(Y_t)_{t\in T}, T = \mathbf{N}$ *ou* $\mathbf{R}_+$, converge A-rapidement *vers* a *(sous la probabilité* $\mathbf{P}$*) si pour tout* $\varepsilon > 0$, *on a* $\mathbf{E}A_{L_\varepsilon} < +\infty$, *où* $L_\varepsilon = \sup\{n \geq 1 : |Y_n - a| \geq \varepsilon\}$.

Il est clair que le processus $(A_t)_{t\geq 0} = (t^r)_{t\geq 0}$, $r > 0$, appartient à la classe $\mathcal{A}$ et que dans ce cas les notions de convergence r-rapide et A-rapide coïncident. Nous allons tout d'abord établir les propriétés suivantes liées à cette notion de convergence A-rapide :

1) Si $(Y_t)_{t\in T}$ converge vers a et $(Z_t)_{t\in T}$ converge vers b A-rapidement, alors :

(*i*) $(\alpha Y_t)_{t\in T}$ converge vers $\alpha\, a$ pour tout $\alpha \in \mathbf{R}$,

(*ii*) $(Y_t + Z_t)_{t\in T}$ converge vers $a + b$,

(*iii*) $(Y_t Z_t)_{t\in T}$ converge vers ab ,

(*iv*) $\left(\frac{Y_t}{Z_t}\right)_{t\in T}$ converge vers $\frac{a}{b}$ si b est non nul

au sens de la convergence A-rapide ;

2) Si Z est une v.a., alors $\mathbf{E}A_{|Z|} < +\infty \Longleftrightarrow \frac{Z}{t}$ tend vers 0 A-rapidement.

Démonstration. On pose $L_{\varepsilon,X} = \sup\{t : |X_t - \lim\limits_{t\to+\infty} X_t| > \varepsilon\}$.

1) (*i*) Si $\alpha = 0$, c'est évident, sinon il suffit de remarquer que l'on a :

$$|\alpha\, Y_t - \alpha\, a| > \varepsilon \Longleftrightarrow |Y_t - a| > \frac{\varepsilon}{|\alpha|} \quad ,$$

ce qui entraîne : $L_{\varepsilon,\alpha Y} = L_{\frac{\varepsilon}{|\alpha|},Y}$ et le résultat s'ensuit.

(*ii*) $|(Y_t + Z_t) - (a + b)| > \varepsilon \Longrightarrow |Y_t - a| > \frac{\varepsilon}{2}$ ou $|Z_t - b| > \frac{\varepsilon}{2}$.

Donc on a :

$$L_{\varepsilon,Y+Z} \le \max\{L_{\frac{\varepsilon}{2},Y}\, ,\, L_{\frac{\varepsilon}{2},Z}\} \Longrightarrow A_{L_{\varepsilon,Y+Z}} \le \max\{A_{L_{\frac{\varepsilon}{2},Y}}\, ,\, A_{L_{\frac{\varepsilon}{2},Z}}\} \le A_{L_{\frac{\varepsilon}{2},Y}} + A_{L_{\frac{\varepsilon}{2},Z}} \ .$$

(*iii*) Montrons tout d'abord que Y_t^2 tend vers a^2 . Quitte à remplacer Y_t par $Y_t - a$, on peut supposer $a = 0$:

$$|Y_t^2| > \varepsilon \Longleftrightarrow |Y_t| > \sqrt{\varepsilon} \quad ,$$

si bien que l'on a : $L_{\varepsilon,Y^2} = L_{\sqrt{\varepsilon},Y}$, d'où la conclusion. Appliquons ce résultat :

$$Y_t\, Z_t = \frac{1}{2}[(Y_t + Z_t)^2 - Y_t^2 - Z_t^2] \underset{t\to+\infty}{\longrightarrow} \frac{1}{2}[(a+b)^2 - a^2 - b^2] = ab \quad A\text{-rapidement.}$$

(*iv*) Par l'assertion (*iii*), on a : $\frac{1}{Z_t} \underset{t\to+\infty}{\longrightarrow} \frac{1}{b}$ et $\frac{Y_t}{Z_t} \underset{t\to+\infty}{\longrightarrow} \frac{a}{b}$ A-rapidement.

2) $\frac{Z}{t} \underset{t\to+\infty}{\longrightarrow} 0$ A-rapidement $\Longleftrightarrow$ $\mathbf{E}A_{L_{\varepsilon,\frac{Z}{t}}} < +\infty$ pour tout $\varepsilon > 0$, où l'on a posé :

$$L_{\varepsilon,\frac{Z}{t}} = \sup\{t : |\frac{Z}{t}| > \varepsilon\} = \frac{|Z|}{\varepsilon} \ .$$

Il en résulte que l'on a :

$$\mathbf{E}A_{L_{\varepsilon,\frac{Z}{t}}} = \mathbf{E}A_{\frac{|Z|}{\varepsilon}} < +\infty \quad \Longleftrightarrow \quad \mathbf{E}A_{|Z|} < +\infty$$

par l'hypothèse (A1).

Le théorème suivant est une conséquence du Théorème 3.2 et plus précisément de son premier corollaire.

Théorème 3.3 *On suppose réalisées les conditions (C1) et (C2). Soit* $A \in \mathcal{A}$ *tel que* $\dfrac{\ln R_n^{(i)}}{n}$ *converge A-rapidement vers* η_i *sous* $\mathbf{P}$, $1 \leq i \leq m$.
On a alors $\mathbf{E}A_{\tau^*} < +\infty$ *et, lorsque* α *tend vers 0, on a :*

$$\inf_{(\tau,d)\in\mathcal{T}(\alpha)} \mathbf{E}A_\tau \sim \mathbf{E}A_{\tau^*} \sim A_{\nu_{(m-1)}} \ .$$

Démonstration. (les notations sont les mêmes que précédemment)
Nous supposons dans un premier temps que $\eta_1, \cdots, \eta_m$ sont strictement positifs. Soit $0 < a < \min\{\eta_1, \cdots, \eta_m\}$. On a établi plus haut (cf. Corollaire 3.1) l'inégalité

$$\tau^* \leq L + 1 + \min_j \max_{i \neq j} \left\{ \frac{\ln C_i}{\eta_i - a} \right\} \ .$$

Donc on a :

$$A_{\tau^*} \leq A_{L+1+\min_j \max_{i\neq j}\left\{\frac{\ln C_i}{\eta_i - a}\right\}} := A_* \ .$$

Comme $\mathbf{E}A_L$ est finie, la condition (A1) entraîne la finitude de $\mathbf{E}A_*$, donc celle de $\mathbf{E}A_{\tau^*}$. Par le Théorème 3.2, on a :

$$\frac{\tau^*}{\nu_{(m-1)}} \xrightarrow[\alpha\to 0]{} 1 \qquad \mathbf{P}\text{−p.s.},$$

donc

$$\frac{A_{\tau^*}}{A_{\nu_{(m-1)}}} \xrightarrow[\alpha\to 0]{} 1 \qquad \mathbf{P}\text{−p.s.}$$

par la condition (A2) et

$$\frac{A_{\tau^*}}{A_{\nu_{(m-1)}}} \leq \frac{A_*}{A_{\nu_{(m-1)}}} \leq A_*$$

pour $\alpha_1, \cdots, \alpha_m$ suffisamment petits.

En effet, les hypothèses de croissance de A vers l'infini assurent qu'il existe t_0 tel que A_{t_0} soit supérieur à 1. Les limites $\eta_1, \cdots, \eta_m$ étant fixées, on prend $\alpha_1 \cdots, \alpha_m$ suffisamment petits, de sorte que l'on ait

$$\nu_{(m-1)} \geq t_0 \ .$$

Comme A_* est intégrable, le théorème de convergence dominée entraîne

$$\mathbf{E}\Big(\frac{A_{\tau^*}}{A_{\nu_{(m-1)}}}\Big) = \frac{\mathbf{E}A_{\tau^*}}{A_{\nu_{(m-1)}}} \xrightarrow[\alpha\to 0]{} 1 \ ,$$

c'est-à-dire, lorsque $\alpha \to 0$,

$$\mathbf{E}A_{\tau^*} \sim A_{\nu_{(m-1)}} \ .$$

Par la première assertion du Théorème 3.2 et la croissance de A, on a pour tout $0 < \delta < 1$:

$$\inf_{(\tau,d)\in\mathcal{T}(\alpha)} \mathbf{P}\Big\{A_\tau > A_{\delta.\nu_{(m-1)}}\Big\} \xrightarrow[\alpha\to 0]{} 1$$

et par conséquent :

$$\inf_{(\tau,d)\in\mathcal{T}(\alpha)} \mathbf{E}A_\tau \geq A_{\delta.\nu_{(m-1)}}(1+o(1))$$

lorsque $\alpha \to 0$. Mais on sait qu'asymptotiquement (τ^*, d^*) appartient à $\mathcal{T}(\alpha)$, donc

$$\mathbf{E}A_{\tau^*} \geq \inf_{(\tau,d)\in\mathcal{T}(\alpha)} \mathbf{E}A_\tau \geq A_{\delta.\nu_{(m-1)}}(1+o(1))$$

lorsque $\alpha \to 0$, si bien qu'en faisant tendre δ vers 1, on a :

$$\inf_{(\tau,d)\in\mathcal{T}(\alpha)} \mathbf{E}A_\tau \sim \mathbf{E}A_{\tau^*} \sim A_{\nu_{(m-1)}}$$

lorsque $\alpha \to 0$.
S'il existe $k \in \{1, \cdots, m\}$ tel que $\eta_k = 0$, on procède comme dans la démonstration du Corollaire 3.1 et on aboutit au résultat analogue :

$$\inf_{(\tau,d)\in\mathcal{T}(\alpha)} \mathbf{E}A_\tau \sim \mathbf{E}A_{\tilde{\tau}^*} \sim A_{\max_{i\neq k}\nu_i} = \max_{i\neq k} A_{\nu_i}$$

lorsque $\alpha \to 0$.
Ce théorème admet de manière évidente le corollaire suivant :

Corollaire 3.3 *On suppose réalisées les conditions (C1) et (C2). Soit $A \in \mathcal{A}$ tel que $\dfrac{\ln R_n^{(ij)}}{n}$ converge A-rapidement vers λ_{ij} sous $\mathbf{P}^j$, $1 \leq i, j \leq m$.*
On a alors $\mathbf{E}_j A_{\tau^*} < +\infty$ et, lorsque α tend vers 0, on a :

$$\inf_{(\tau,d)\in\mathcal{T}(\alpha)} \mathbf{E}_j A_\tau \sim \mathbf{E}_j A_{\tau^*} \sim A_{\max_{i\neq j} \frac{|\ln \alpha_i|}{\lambda_{ij}}} = \max_{i\neq j} A_{\frac{|\ln \alpha_i|}{\lambda_{ij}}} \ , \ 1 \leq j \leq m \ .$$

3.4.2 Une condition nécessaire pour une loi des grands nombres

Lemme 3.2 *Soit $A \in \mathcal{A}$. Soient $X, X_1, X_2, \cdots$ des variables aléatoires i.i.d. telles que $\mathbf{E}X = 0$. On pose $S_n = X_1 + \cdots + X_n$, $n \geq 1$. Alors :*

$$\frac{S_n}{n} \xrightarrow[n\to+\infty]{} 0 \quad A-rapidement\ sous\ \mathbf{P} \quad \Longrightarrow \quad \mathbf{E}(|X|A_{|X|}) < +\infty \ .$$

Démonstration. Nous allons dans un premier temps donner des conditions équivalentes à la finitude de $\mathbf{E}A_{L_\varepsilon}$ et de $\mathbf{E}(|X|A_{|X|})$, où

$$L_\varepsilon = \sup\left\{n \geq 1 : \left|\frac{S_n}{n}\right| \geq \varepsilon\right\} \ .$$

$$\begin{aligned}
\mathbf{E}A_{L_\varepsilon} &= \sum_{n=0}^{+\infty} A_n \mathbf{P}\{L_\varepsilon = n\} = \sum_{n=0}^{+\infty} A_n \mathbf{P}\{L_\varepsilon \geq n\} - \sum_{n=0}^{+\infty} A_n \mathbf{P}\{L_\varepsilon \geq n+1\} \\
&= \sum_{n=1}^{+\infty} (A_n - A_{n-1}) \mathbf{P}\{L_\varepsilon \geq n\} \\
&= \sum_{n=1}^{+\infty} (A_n - A_{n-1}) \mathbf{P}\{\sup_{k\geq n} \left|\frac{S_k}{k}\right| \geq \varepsilon\}.
\end{aligned}$$

Les équivalences suivantes proviennent de la condition (A3) :

$$\mathbf{E}A_{L_\varepsilon} < +\infty \Longleftrightarrow \sum_{n=1}^{+\infty} \frac{A_n}{n} \mathbf{P}\{\sup_{k\geq n} \left|\frac{S_k}{k}\right| \geq \varepsilon\} < +\infty \qquad (a)$$

$$\begin{aligned}
\mathbf{E}(|X| A_{|X|}) < +\infty &\Longleftrightarrow \sum_{n=0}^{+\infty} n A_n \mathbf{P}\{|X| \in [n; n+1[\} < +\infty \\
&\Longleftrightarrow \sum_{n=1}^{+\infty} (nA_n - (n-1)A_{n-1}) \mathbf{P}\{|X| \geq n\} < +\infty \qquad (b) \\
&\Longleftrightarrow \sum_{n=1}^{+\infty} A_n \mathbf{P}\{|X| \geq n\} < +\infty .
\end{aligned}$$

Montrons que $\mathbf{E}(|X| A_{|X|}) < +\infty$ est une condition nécessaire à la convergence A-rapide de $\dfrac{S_n}{n}$ vers 0.

Posons $a_n = \mathbf{P}\{|X| \geq n\}$, fixons $\varepsilon = \frac{1}{2}$. On a alors :

$$+\infty > \sum_{n=1}^{+\infty} \frac{A_n}{n} \mathbf{P}\{\sup_{k\geq n} \left|\frac{S_k}{k}\right| \geq \frac{1}{2}\} \geq \sum_{n=1}^{+\infty} \frac{A_n}{n} \mathbf{P}\left\{\bigcup_{k=1}^{+\infty} (|X_{n+k}| \geq n+k)\right\} . \qquad (c)$$

En effet, supposons que l'on ait :

$$\sup_{k\geq n} \left|\frac{S_k}{k}\right| < \frac{1}{2} \quad \text{et} \quad \left|\frac{X_{n+k_0}}{n+k_0}\right| \geq 1$$

pour un certain $k_0 \geq 1$. Il vient alors :

$$\left|\frac{S_{n+k_0-1}}{n+k_0-1}\right| \geq \left|\frac{S_{n+k_0-1}}{n+k_0}\right| = \left|\frac{S_{n+k_0}}{n+k_0} - \frac{X_{n+k_0}}{n+k_0}\right| \geq \left|\left|\frac{S_{n+k_0}}{n+k_0}\right| - \left|\frac{X_{n+k_0}}{n+k_0}\right|\right| \geq \frac{1}{2}$$

ce qui est en contradiction avec l'hypothèse sur $\sup_{k\geq n} \left|\dfrac{S_k}{k}\right|$. D'où nécessairement :

$$\left(\sup_{k\geq n} \left|\frac{S_k}{k}\right| < \frac{1}{2}\right) \quad \Longrightarrow \quad \left(\left|\frac{X_{n+k_0}}{n+k_0}\right| < 1 \ , \ \forall\, k_0 \geq 1\right)$$

ce qui est équivalent à :

$$\left(\exists\, k_0 \geq 1 \text{ tel que } \left|\frac{X_{n+k_0}}{n+k_0}\right| \geq 1\right) \quad \Longrightarrow \quad \left(\sup_{k\geq n} \left|\frac{S_k}{k}\right| \geq \frac{1}{2}\right) .$$

L'inégalité (c) implique

$$\begin{aligned}
+\infty &> \sum_{n=1}^{+\infty} \frac{A_n}{n} \mathbf{P}\left\{ \bigcup_{k=1}^{+\infty} (|X_{n+k}| \geq n+k) \right\} \\
&\geq \sum_{n=1}^{+\infty} \frac{A_n}{n} \left\{ \sum_{k=1}^{+\infty} a_{n+k} - \sum_{k=1}^{+\infty} a_{n+k} . \sum_{j>k} a_{n+j} \right\} .
\end{aligned}$$

Il est clair que la convergence A-rapide de $\frac{S_n}{n}$ vers 0 entraîne la finitude de la série

$$\sum_{n=1}^{+\infty} \frac{A_n}{n} \mathbf{P}\{ \left| \frac{S_n}{n} \right| \geq \varepsilon \}$$

pour tout $\varepsilon > 0$. Le Théorème 1 de Baum et Katz (cf. [Ba-Ka]) assure la finitude de $\mathbf{E}|X|$, donc de la série de terme général a_n et la convergence de $\sum_{j=1}^{+\infty} a_{n+j}$ lorsque n tend vers l'infini, si bien que l'on a :

$$\begin{aligned}
+\infty &> \sum_{n=1}^{+\infty} \frac{A_n}{n} \left\{ \sum_{k=1}^{+\infty} a_{n+k} (1 - \sum_{j>k} a_{n+j}) \right\} \\
&\geq c \sum_{n=1}^{+\infty} \frac{A_n}{n} \sum_{k=1}^{+\infty} a_{n+k} \quad \text{pour } c \text{ une constante positive} \\
&= c \sum_{n=1}^{+\infty} a_n \sum_{k=1}^{n} \frac{A_k}{k} \\
&= c \left\{ \sum_{n=1}^{n_1} a_n \sum_{k=1}^{n} \frac{A_k}{k} + \sum_{n=n_1+1}^{+\infty} a_n \sum_{k=1}^{n_2} \frac{A_k}{k} + \sum_{n=n_1+1}^{+\infty} a_n \sum_{k=n_2+1}^{n} \frac{A_k}{k} \right\}
\end{aligned}$$

pour tous n_1 et n_2 . En particulier, la dernière de ces trois sommes est finie. Lorsque n_2 est suffisamment grand, l'hypothèse $(A3)$ entraîne qu'il existe deux constantes positives c_1 et c_2 telles que l'on ait :

$$c_1 \leq \frac{A_k - A_{k-1}}{\frac{A_k}{k}} \leq c_2, \quad k > n_2 \ ,$$

si bien que

$$\sum_{n=n_1+1}^{+\infty} a_n \sum_{k=n_2+1}^{n} \frac{A_k}{k} < +\infty \quad \Longleftrightarrow \quad \sum_{n=n_1+1}^{+\infty} a_n (A_n - A_{n_2}) < +\infty \ .$$

Comme la série de terme général a_n est convergente, on a

$$\sum_{n=1}^{+\infty} a_n A_n < +\infty \ ,$$

ce qui est équivalent à la finitude de $\mathbf{E}(|X| A_{|X|})$ par (b).

3.4.3 Une ébauche de condition suffisante

Il semblerait que la finitude de $\mathbf{E}(|X|A_{|X|})$ soit également une condition suffisante, mais nous n'avons pas encore réussi à le démontrer. Nous avons néanmoins établi que si le processus A satisfait à la condition

(A4) il existe $1 \leq l \leq 2$ tel que les suites $\left(\sum_{n=k}^{+\infty} \frac{A_n}{n^l}\right)_{k\geq 1}$ et $\left(\frac{A_k}{k^{l-1}}\right)_{k\geq 1}$ sont équivalentes,

alors on a :

$$\mathbf{E}(|X|A_{|X|}) \quad \Longrightarrow \sum_{n=1}^{+\infty} \frac{A_n}{n} \mathbf{P}\Big\{ \left|\frac{S_n}{n}\right| \geq \varepsilon \Big\} < +\infty \ , \ \forall\, \varepsilon > 0,$$

ce qui malheureusement n'est pas tout à fait la définition de la convergence A-rapide de $\frac{S_n}{n}$ vers 0.

En effet, fixons $\varepsilon = 1$ et posons $\quad X_{kn} = \begin{cases} X_k & \text{si } |X_k| < n \\ 0 & \text{sinon} \end{cases}$.

$$\begin{aligned}
\mathbf{P}\{|S_n| \geq n\} &= \mathbf{P}\{|S_n| \geq n, |X_k| \geq n \text{ pour un } k \leq n\} \\
&\qquad + \mathbf{P}\{|S_n| \geq n, |X_k| < n \text{ pour tout } k \leq n\} \\
&\leq n\,\mathbf{P}\{|X| \geq n\} + \mathbf{P}\left\{ \left| \sum_{k=1}^{n} X_{kn} \right| > n, |X_k| < n \text{ pour tout } k \leq n \right\} \\
&\leq n\,\mathbf{P}\{|X| \geq n\} + \mathbf{P}\left\{ \left| \sum_{k=1}^{n} (X_{kn} - \mathbf{E}X_{kn}) \right| > n(1 - |\mathbf{E}X_{kn}|) \right\} \ .
\end{aligned}$$

Il en résulte que l'on a :

$$\begin{aligned}
\sum_{n=1}^{+\infty} \frac{A_n}{n} \mathbf{P}\{|S_n| \geq n\} \leq \sum_{n=1}^{+\infty} & A_n \mathbf{P}\{|X| \geq n\} \\
& + \sum_{n=1}^{+\infty} \frac{A_n}{n} \mathbf{P}\left\{ \left| \sum_{k=1}^{n} (X_{kn} - \mathbf{E}X_{kn}) \right| > n(1 - |\mathbf{E}X_{kn}|) \right\}.
\end{aligned}$$

La première somme du second membre est finie en vertu de l'équivalence (b). Comme $\mathbf{E}X_{kn}$ tend vers 0 lorsque n tend vers l'infini, il suffit de montrer la finitude de

$$\sum_{n=1}^{+\infty} \frac{A_n}{n} \mathbf{P}\left\{ \left| \sum_{k=1}^{n} (X_{kn} - \mathbf{E}X_{kn}) \right| > c.n \right\}$$

pour $0 < c < 1$. Dans les inégalités suivantes, K désigne une constante positive dont la valeur est susceptible de changer de ligne en ligne :

$$\begin{aligned}
\sum_{n=1}^{+\infty} \frac{A_n}{n} \mathbf{P}\left\{\left|\sum_{k=1}^{n}(X_{kn} - \mathbf{E}X_{kn})\right| > c.n\right\} &\leq \sum_{n=1}^{+\infty} \frac{A_n}{n} \frac{1}{c^l n^l} \mathbf{E}\left|\sum_{k=1}^{n}(X_{kn} - \mathbf{E}X_{kn})\right|^l \\
&\leq K \sum_{n=1}^{+\infty} \frac{A_n}{n^{1+l}} \sum_{k=1}^{n} \mathbf{E}|X_{kn}|^l = K \sum_{n=1}^{+\infty} \frac{A_n}{n^{1+l}}.n. \int_{|x|<n} |x|^l dF(x) \\
&\leq K \sum_{n=1}^{+\infty} \frac{A_n}{n^l} \sum_{k=1}^{n} k^l \mathbf{P}\{k-1 \leq |X| < k\} \\
&= K \sum_{n=1}^{+\infty} k^l \mathbf{P}\{k-1 \leq |X| < k\}. \sum_{n=k}^{+\infty} \frac{A_n}{n^l} \\
&\leq K \sum_{n=1}^{+\infty} k^l \frac{A_k}{k^{l-1}} \mathbf{P}\{k-1 \leq |X| < k\} \\
&= K \sum_{n=1}^{+\infty} k\, A_k \mathbf{P}\{k-1 \leq |X| < k\} < \infty
\end{aligned}$$

puisque $\mathbf{E}(|X|A_{|X|})$ est finie.
Nous ne sommes pas non plus en mesure de donner une interprétation de cette notion de convergence ; nous avons simplement remarqué que $A_n = n^r$ $(0 < r < 1)$ satisfait aux conditions (A1)–(A4) avec $l = 2$, que $A_n = \ln n$ vérifie (A1), (A2), (A4) avec $l = 2$ et que l'équivalence

$$\mathbf{E}(|X|A_{|X|}) \quad \Longleftrightarrow \sum_{n=1}^{+\infty} \frac{A_n}{n} \mathbf{P}\left\{\left|\frac{S_n}{n}\right| \geq \varepsilon\right\} < +\infty\ ,\ \forall\, \varepsilon > 0$$

a lieu pour ces deux processus ; voir les Théoremes 2 et 3 de [Ba-Ka]. Il serait intéressant de voir si la classe de processus que nous considérons ne se réduit pas aux seuls polynômes.

Les résultats exposés dans ce chapitre ont donné lieu à une note à paraître aux Comptes-Rendus de l'Académie des Sciences ([Ko2]) et à un article soumis à publication à la revue *Sequential Analysis* ([Ko3]).

Références bibliographiques

[Ba-Ka] L. E. Baum et M. Katz. *1965 Convergence rates in the law of large numbers.* Trans. Amer. Math. Soc., vol. 120, pp. 108-123.

[Dr-No] V. P. Dragalin et A. A. Novikov. *Adaptive sequential tests for composite hypotheses.* A paraître dans Theory of Probability and its Applications.

[Ir2] A. Irle. *1990 Asymptotic optimality of general sequential probability ratio tests.* Scandinavian Journal of Statistics, vol. 17, pp. 321-332.

[Ir3] A. Irle. *1993 r-quick convergence for regenerative processes with applications to sequential analysis.* Stochastic Processes and their Applications, vol. 45, pp. 319-329.

[Ko2] C. Koell. *Optimalité asymptotique du test de Wald.* A paraître aux Comptes-Rendus de l'Académie des Sciences.

[Ko3] C. Koell. *Asymptotic optimality of the Wald sequential test.* Soumis à publication à Sequential Analysis.

[La1] T. L. Lai. *1976 On r-quick convergence and a conjecture of Strassen.* Annals of Probability, vol.4, no.4, pp. 612-627.

[La2] T. L. Lai. *1981 Asymptotic optimality of invariant sequential probability ratio tests.* Annals of Statistics, vol. 9, no.2, pp. 318-333.

[Le] E. L. Lehmann. *1959 Testing statistical hypotheses.* Wiley, New York.

[Sh1] A. N. Shiryaev. *1973 Statistical sequential analysis (Translations of Mathematical Monographs, vol. 38).* A.M.S., Providence.

[St] V. Strassen. *1965 Almost sure behavior of sums of independent random variables and martingales.* Proc. Fifth Berkeley Symp. Math. Statist. Prob., vol. 2, part 2, pp. 315-343.

[Wa] A. Wald. *1947 Sequential analysis.* Wiley, New York.

[Wa-Wo] A. Wald et J. Wolfowitz. *1948 Optimum character of the sequential probability ratio test.* Annals of Mathematical Statistics, vol. 19, pp. 326-339.

Bibliographie

[Ba-Ka] L. E. Baum et M. Katz. *1965 Convergence rates in the law of large numbers.* Trans. Amer. Math. Soc., vol. 120, pp. 108-123.

[Dr-No] V. P. Dragalin et A. A. Novikov. *Adaptive sequential tests for composite hypotheses.* A paraître dans Theory of Probability and its Applications.

[Dv-Ki-Wo] A. Dvoretzky, J. Kiefer et J. Wolfowitz. *1953 Sequential decision problems for processes with continuous time parameter. Testing hypotheses.* Ann. Math. Statist., vol. 24, pp. 254-264.

[Ga-Ko] L. I. Galtchouk et C. Koell. *1992 Test séquentiel de deux hypothèses simples sur le paramètre du processus de vélocité d'Ornstein-Uhlenbeck.* Communication aux 24^{émes} Journées de Statistique de Bruxelles.

[Gh-Se] B. K. Ghosh et P. K. Sen, éditeurs. *1991 Handbook of sequential analysis.* Dekker, New York.

[Ir1] A. Irle. *1984 Extended optimality of sequential probability ratio tests.* Annals of Statistics, vol. 12, no. 1, pp. 380-386.

[Ir2] A. Irle. *1990 Asymptotic optimality of general sequential probability ratio tests.* Scandinavian Journal of Statistics, vol. 17, pp. 321-332.

[Ir3] A. Irle. *1993 r-quick convergence for regenerative processes with applications to sequential analysis.* Stochastic Processes and their Applications, vol. 45, pp. 319-329.

[Ir-Sc] A. Irle et N. Schmitz. *1984 On the optimality of the SPRT for processes with continuous time parameter.* Math. Operationsforsch. u. Statist., ser. Statist., vol. 15, no. 2, pp. 91-104.

[Ja] J. Jacod. *1979 Calcul stochastique et problèmes de martingales.* Lecture Notes in Mathematics no. 714. Springer Verlag.

[Ja-Sh] J. Jacod et A. N. Shiryaev. *1987 Limit theorems for stochastic processes.* Springer Verlag.

[Ka-Li-Sh1] Yu. M. Kabanov, R. Sh. Liptser et A. N. Shiryaev. *1977 On the question of absolute continuity and singularity of probability measures.* Mat. USSR Sbornik, vol. 33, no. 2, pp. 203-221.

[Ka-Li-Sh2] Yu. M. Kabanov, R. Sh. Liptser et A. N. Shiryaev. *1979 Absolute continuity and singularity of locally absolutely continuous probability distributions (I).* Mat. USSR Sbornik, vol. 35, pp. 631-679.

[Ka-Li-Sh3] Yu. M. Kabanov, R. Sh. Liptser et A. N. Shiryaev. *1980 Absolute continuity and singularity of locally absolutely continuous probability distributions (II).* Mat. USSR Sbornik, vol. 36, pp. 31-58.

[Ko1] C. Koell.*1994 An optimal procedure for testing hypotheses on cadlag processes.* Sequential Analysis, vol. 13, no. 3, pp. 221-236.

[Ko2] C. Koell. *Optimalité asymptotique du test de Wald.* A paraître aux Comptes-Rendus de l'Académie des Sciences.

[Ko3] C. Koell. *Asymptotic optimality of the Wald sequential test.* Soumis à publication à Sequential Analysis.

[Ku] S. Kullback. *1968 Information theory and statistics.* Dover.

[La1] T. L. Lai. *1976 On r-quick convergence and a conjecture of Strassen.* Annals of Probability, vol.4, no.4, pp. 612-627.

[La2] T. L. Lai. *1981 Asymptotic optimality of invariant sequential probability ratio tests.* Annals of Statistics, vol. 9, no.2, pp. 318-333.

[LB] A. Le Breton. *1992 Adaptive control in the scalar linear-quadratic model in continuous time.* Statistics and Probability Letters, vol. 13, pp. 169-177.

[Le] E. L. Lehmann. *1959 Testing statistical hypotheses.* Wiley, New York.

[Li-Sh1] R. S. Liptser et A. N. Shiryaev. *1977 Statistics of random processes (I). General theory.* Springer Verlag.

[Li-Sh2] R. S. Liptser et A. N. Shiryaev. *1978 Statistics of random processes (II). Applications.* Springer Verlag.

[Lo] G. Lorden. *1980 Structure of sequential tests minimizing an expected sample size.* Z. Wahrscheinlichkeitstheorie verw. Gebiete, vol. 51, pp. 291-302.

[Me] P. A. Meyer. *1976 Un cours sur les intégrales stochastiques.* Séminaire Proba. XI. Lecture Notes in Mathematics no. 511, pp. 245-400. Springer Verlag.

[Mi] T. P. Miroshnichenko. *1979 Testing of two simple hypotheses in the presence of delayed observations.* Theory of Probability and its Applications, vol. 24, no. 3, pp. 467-479.

[Re-Yo] D. Revuz et M. Yor. *1991 Continuous martingales and brownian motion.* Springer Verlag.

[Sh1] A. N. Shiryaev. *1973 Statistical sequential analysis (Translations of Mathematical Monographs, vol. 38).* A.M.S., Providence.

[Sh2] A. N. Shiryaev. *1978 Optimal stopping rules.* Springer Verlag.

[St] V. Strassen. *1965 Almost sure behavior of sums of independent random variables and martingales.* Proc. Fifth Berkeley Symp. Math. Statist. Prob., vol. 2, part 2, pp. 315-343.

[Wa] A. Wald. *1947 Sequential analysis.* Wiley, New York.

[Wa-Wo] A. Wald et J. Wolfowitz. *1948 Optimum character of the sequential probability ratio test.* Annals of Mathematical Statistics, vol. 19, pp. 326-339.

[Ya] A. I. Yashin. *1983 On a problem of sequential hypothesis testing.* Theory of Probability and its Applications, vol. 28, no. 1, pp. 157-165.