Astérisque

DIDIER DACUNHA-CASTELLE

Introduction

Astérisque, tome 68 (1979), p. 3-18

<http://www.numdam.org/item?id=AST_1979__68__3_0>

INTRODUCTION

D. DACUNHA-CASTELLE.

I. LES FONDEMENTS PROBABILISTES DE LA THÉORIE DES GRANDES DÉVIATIONS.

La théorie des grandes déviations (grands écarts) concerne la loi faible des grands nombres. Si $X_1 \ldots X_n$ est une suite de variables aléatoires indépendantes, équidistribuées, d'espérance $EX = 0$, et si $S_n = X_1 + \ldots + X_n$, la loi faible des grands nombres assure de la convergence vers 0 de $\frac{S_n}{n}$ (en probabilité). Pour préciser cette convergence, on dispose d'abord du théorème de limite centrale qui étudie la convergence de $\frac{S_n}{\sqrt{n}}$. Soit $a > 0$, on obtient :

$$P\ (S_n > \sqrt{n}\,\sigma\ a) \sim 1 - \Phi\ (a)$$

où Φ est la répartition de la loi normale
et $\sigma^2 = EX^2$. Au lieu d'étudier le grossissement $\sqrt{n}$ de $\frac{S_n}{n}$, on peut étudier les grandes déviations sans grossissement, à savoir $P\ (S_n > na)$. Ces grandes déviations sont rares et le premier résultat sera que

$$P\ (S_n > na) \sim e^{-nh(a)} \qquad (1) \quad ,$$

où h est une fonction convenable.

De manière générale, on peut étudier $P\ (S_n > B_n\ a)$.

$B_n = O\ (\sqrt{n})$ donne le théorème de limite centrale et $B_n = O\ (n)$ les grandes déviations. Entre les deux $\sqrt{n} = \sigma\ (B_n)$ et $B_n = \sigma\ (n)$, on trouve la théorie des moyennes déviations, qui conduit à des probabilités de l'ordre de $\frac{1}{n^\alpha}$ pour des α convenables, $\alpha > 0$, dans certains cas.

Bien entendu, l'ensemble de ces théorèmes ne valent que sur des hypothèses convenables sur la queue des distributions des X. Très grossièrement la loi des grands nombres demande $E|X| < \infty$, le théorème limite centrale $E X^2 < \infty$ et les résultats de grandes déviations $E \exp tX < \infty$ pour des $t > 0$. Si les premiers résultats sur les grandes déviations sont dus à KHINCHINE [6], puis avant-guerre par exemple à CRAMER [3] ils ont pris sous l'influence des statisticiens le nom générique de formules de Chernoff [2]. En fait, on trouve d'abord des majorations exponentielles du type $P(S_n > na) \leqq e^{-nh(a)}$ qui dans les cas simples sont des applications directes d'une formule de Markov-Tchebychev. La minoration, même dans le cas élémentaire que nous venons d'introduire, a une démonstration intéressante en soi. Elle est une conséquence directe de la loi faible des grands nombres, appliquée à $\frac{S_n}{n}$, non pour la probabilité initiale associée à la loi dF des X, mais pour une nouvelle probabilité P_a obtenue en remplaçant dF par dF_a, avec $\frac{dF_a}{dF}(x) = \frac{e^{tx}}{\phi(t)}$, où t est choisi tel que $\int x \, dF_a(x) = a$, soit $\frac{\phi'(t)}{\phi(t)} = a$ si $\phi(t) = E \exp tX$.

Ce recentrage de la probabilité permet d'appliquer un théorème ergodique, ici la loi faible ordinaire. Cette idée sera essentielle.

Le chapitre I, élémentaire, souligne que pour qu'une suite Z_n de loi P_n satisfasse à la formule (1), il suffit que la transformée de Laplace Φ_n de P_n existe sur un intervalle contenant 0 et qu'elle ait un comportement ergodique du type $\frac{1}{n} \log \Phi_n(t) \to L(t)$. Cette convergence assure la loi faible des grands nombres pour les $dP_{n,a}$ obtenues par recentrage. On a aussi le théorème limite centrale pour les $P_{n,a}$ qui donne d'ailleurs un développement plus précis que (1).

La technique utilisée pour obtenir (1) est voisine de celle utilisée par exemple par LINNIK et FELLER [5] pour obtenir des formules de moyennes déviations mais aussi des techniques utilisées pour obtenir des développements affinant le théorème de limite centrale, comme les techniques de point de selle.

Intuitivement, si l'on interprète les X_i comme des variables à valeur

mesure (masses de Dirac) plutôt que comme des réels, on est amené à étudier, non pas $\frac{S_n}{n}$ mais $\hat{F}_n = \frac{1}{n}(\delta_{X_1} + \ldots + \delta_{X_n})$, répartition empirique. On sait que cette répartition converge (loi faible) et qu'elle satisfait au théorème de GLIVENKO-CANTELLI (convergence de $\sqrt{n} \sup_x |\hat{F}_n(x) - F(x)|$ qui est l'analogue du théorème de limite centrale. Il est donc raisonnable d'essayer d'obtenir des équivalents pour $P(\hat{F}_n \in B)$ où B est un ensemble exceptionnel ou très rare pour $\hat{F}_n$, $B \subset \mathcal{P}(\mathbb{R})$, ensemble des probabilités sur $\mathbb{R}$. La formule (1) équivaut à

$$P(\hat{F}_n \in B) \sim e^{-n\, h(a)} \qquad \text{avec}$$

$$B \{G \in \mathcal{P}(\mathbb{R}), \quad \int x\, dG > a\}$$

De manière générale, on cherchera des formules du type

$$P(\hat{F}_n(B)) \sim e^{-n\, I(B)} \qquad (2)$$

mais de telles formules ne vont exister que pour des ensembles bien particuliers; par contre, ouverts et fermés conduiront à des inégalités.

Les formules (2) ont un sens pour des variables à valeurs dans des espaces très généraux. Le cadre agréable est celui des polonais, étendu pour les nécessités au cas souslinien.

Les formules (1) ont un sens pour des variables à valeurs dans un espace vectoriel. Le cas étudié sera celui des Banach séparables. Dans tous les cas, ce qui est décisif c'est l'extension à $\mathbb{R}^k$ des formules (1) et (2). (Pour les polonais on se ramènera à ce cas, en considérant des ensembles convexes, des techniques de séparation et de projection).

Une notion apparait tout de suite comme centrale dans la théorie, c'est celle d'informations de Kullback, avec ses différentes formulations.

Rappelons que si P et Q sont deux probabilités, on définit $I(P,Q)$ par

$$I(P,Q) = E_p \log \frac{dP}{dQ}$$

$$= \int (\log \frac{dP}{dQ})\, dP$$

lorsque cette quantité a un sens, et $+\infty$ si elle n'en a pas.

I a les propriétés fondamentales bien connues d'une information. En particulier elle diminue par image. De plus elle est convexe et s.c.i. par rapport à l'argument P (et pour la convergence étroite).

Dans la démonstration élémentaire de (1), on introduit la fonction h définie par

$$h_F(a) = \sup_t \; (ta - \log \phi(t)) \qquad (3)$$

Cette fonction dite transformée de Cramer de F (qu'elle permet de retrouver), est donc la duale, au sens de Young-Orlicz de l'analyse convexe, de la fonction $\log \phi(t)$. Son lien avec l'information, connu depuis longtemps, se traduit par la formule

$$h_F(a) = \inf_{\{G,\ \int x\, d\, G(x) = a\}} I(G, F) \qquad (4)$$

Cette formule est un outil technique décisif. Elle permet d'abord d'utiliser (4) pour définir l'exposant h (B) qui sera naturel pour les extensions (2) de (1), on posera

$$h_F(B) = \inf_{G \in B} I(G, F) \qquad (5)$$

$h_F(B)$ sera donc l'exposant mesurant la rareté de l'ensemble B par rapport à un échantillonnage $\hat{F}_n$ de F. (5) a donc un sens sur un espace abstrait quelconque. La formule d'analyse convexe (3) s'étend elle au cas d'un Banach séparable E, en supposant que $\phi_F(t) = E_F \exp t \|X\| < \infty$. On peut poser alors :

$$h_F(a) = \sup_{\theta \in E^*} \; (\langle \theta, a \rangle - \log E_F \exp \langle \theta, X \rangle) \qquad (6)$$

(6) permet d'étendre (4) au cas d'un Banach.

Le chapitre III fait une étude détaillée des résultats concernant les cas Banach séparable et polonais (les principaux résultats sont dus à DONSKER et VARADHAN). On obtient essentiellement

(7) $\limsup \frac{1}{n} \log P_F \left(\frac{S_n}{n} \in B \right) \leqq - \inf h_F (B)$ si B est fermé

(8) $\limsup \frac{1}{n} \log P_F \left(\frac{S_n}{n} \in B \right) \geqq - \inf h_F (B)$ si B est ouvert (cas Banach)

et (9) $\limsup \frac{1}{n} \log P_F (\hat{F}_n \in B) \leqq - \inf_{\lambda \in B} I (\lambda, F)$ B fermé de $\mathcal{P}(E)$

(10) $\limsup \frac{1}{n} \log P_F (\hat{F}_n \in B) \geqq - \inf_{\lambda \in B} I (\lambda, F)$ B ouvert de $\mathcal{P}(E)$, (cas polonais) .

Dans la pratique statistique, il est intéressant de savoir si des ensembles B particuliers sont de bons ensembles, à savoir qu'ils réalisent à la fois (9) et (10). Un cas utile pour l'étude des tests de Kolmogorov-Smirnov est celui où l'on a

$$B = \{G, T(G) > a\}$$

où T est une fonctionnelle uniformément continue sur $\mathcal{P}(\mathbb{R})$, munie de la distance de $(G,G') = \|G-G'\|_\infty$. Lorsque F est continue, ce type de résultats a été l'objet d'efforts importants. Le passage (élémentaire et direct) du cas polonais au cas soulinien donne ce type de résultat , en conclusion du chapitre III, sans effort notable.

Le cadre probabiliste général est donc fixé par le chapitre III, avec trois notions essentielles : information de Kullback, transformée de Cramer et changement de probabilité privilégié recentrant le problème.

Les développements probabilistes sont extrêmement nombreux et touchent tous les types de processus aléatoires sans exception (processus de Markov ponctuels, gaussiens, etc...). Dans ce séminaire ne sont traités que des problèmes bien particuliers, portant sur des ensembles B très précis et choisis en vue d'applications statistiques. Cependant, les deux chapitres qui y sont consacrés présentent les techniques essentielles. Pour le reste nous renvoyons à [0]

un cours de R. AZENCOTT sur des aspects probabilistes essentiels.

Le chapitre IV, à partir d'idées de Borovkov, étudie les probabilités de certaines grandes déviations des marches aléatoires. Considérant la marche $S_n = X_1+\ldots+X_n$ introduite plus haut, on définit une suite de processus à valeurs sur $C_o(0,1)$ en posant

$$s_n\left(\frac{k}{n}\right) = \frac{S_k}{n} \quad , \qquad 1 \leqq k \leqq n \quad ; \qquad s_n(0) = 0 \quad ,$$

et $s_n(t)$ étant obtenue à partir de $s_n\left(\frac{k}{n}\right)$ par interpolation linéaire.

Les ensembles B auxquels on s'intéresse sont du type suivant. Soit $g(t)$ une fonction sur $(0,1)$, on étudie

$$B = \{f \in C_o(0,1) \quad , \quad \exists t \text{ avec } f(t) \geqq g(t)\}$$

$(S_n \in B)$ est donc l'événement : franchissement par la marche aléatoire S_n de la barrière B. Une variante est le non-franchissement d'une barrière, soit :

$$B = \{f \in C_o(0,1), \quad \forall t \ , \ f(t) < g(t)\}$$

Bien entendu, il faut que g satisfasse des conditions particulières pour que B conduise à un événement exceptionnel. Pour le franchissement, il faut qu'il existe des t tels que $\frac{g(t)}{t} > EX$, ceci puisque $\frac{s(t)}{t} \not\to \frac{S_k}{n}\frac{n}{k}$ et que $\frac{S_k}{k} \to EX$.

Si f est presque partout dérivable, on définit W par

$$W(f) = \int_o^1 h(f'(t))\,dt \qquad (11)$$

(quelquefois liée à la notion d'intégrale d'action).

W joue pour ces problèmes de suite de marches le rôle de la transformée de Cramer (de l'information). Par passage à la limite, à partir de la suite des problèmes discrets associés à chaque n, on obtient des résultats du type suivant, si $EX = 0$,

$$\underline{\lim}\ \frac{1}{n}\ \log\ P_F\ (s_n \in G)\ \geqq\ -\ W\ (\overset{\circ}{B}) \qquad (12)$$

$$\overline{\lim}\ \frac{1}{n}\ \log\ P_F\ (s_n \in G)\ \leqq\ -\ W\ (\overline{B}) \qquad (13)$$

si B est un borélien de $C\ (0,1)$, d'intérieur $\overset{\circ}{G}$ et de fermeture $\overline{G}$, et

$$W\ (B)\ =\ \inf_{\{f\in G, f(0)=0)\}}\ W\ (f)$$

Les démonstrations faites pour obtenir (12) et (13) valent pour les moyennes déviations, à condition de poser

$$s_n\ (\frac{k}{n})\ =\ \frac{S_k}{x(n)}\quad ,\quad \frac{x(n)}{n} \to 0\quad ,\quad \frac{x(n)}{\sqrt{n}} \to \infty$$

et de remplacer dans (12) et (13), $\frac{1}{n}$ par $\frac{n}{x^2\ (n)}$ et W par W_o associée à la loi normale (c'est-à-dire à $h(t) = \frac{t^2}{2}$) .

Cette première partie du chapitre IV introduit donc une technique générale de passage du discret au continu. La deuxième partie vise à calculer explicitement $W(B)$ lorsque B correspond à un franchissement (ou à un non franchissement de frontière). On trouve par exemple pour le franchissement : s'il existe t_o tel que $\frac{g\ (t_o)}{t_o} < \text{ess sup } X$

$$W\ (B)\ =\ \inf_u\ u\ h\ (\ \frac{g(u)}{u}\ )\qquad .$$

La remarque suivante semble avoir des incidences, qui ne sont pas encore suffisamment explorées, dans de nombreux problèmes tant mathématiques que directement appliqués : s'il existe un u_o unique réalisant $\inf\ u\ h\ (\ \frac{g(u)}{u}\ )$, alors on obtient une information décisive sur la façon dont s'effectue le franchissement lorsque celui-ci a lieu. A savoir : conditionnellement au franchissement de la frontière g, la fonction $\frac{g(u_o)}{u_o}\ \inf\ (t, u_o)$ est la trajectoire la plus probable.

Dans certains problèmes, il y a donc une distribution conditionnelle très "pointue" sur l'ensemble des trajectoires franchissant une frontière (cet ensemble étant de probabilité très petite). Des développements du même ordre peuvent être faits pour le même franchissement d'une frontière, ou le non échappement d'un tube.

Le chapitre VI abord l'extension des formules (1) et (2) au cas d'une chaîne de Markov, contrôlée ou non. Soit $X_1 \ldots X_n$, ... une telle chaîne, d'espace d'états (E, $\mathcal{E}$), polonais en général, et d'espace d'actions (A, $\mathcal{A}$). La chaîne est caractérisée par la transition sous contrôle π (x,a,dy) et l'on définit la fonction empirique des observations et des contrôles par

$$\hat{F}_n (B) = \sum_{j=1}^{n} 1_B (X_j, A_j, X_{j+1})$$

où A_j est le $j^{ème}$ contrôle.

Si λ est une probabilité sur $E_x A_x E$, on appellera λ sa première marginale sur E, $\lambda_1 \otimes s$ sa marginale sur $E_x A$ (s est donc une transition de E dans A), et l'on s'intéressera à l'information de λ par rapport à $\lambda_1 \otimes s \otimes \pi$, qui mesurera bien l'écart en situations initiales λ_1, s de λ à la chaîne.

On pose donc $I(\lambda) = I(\lambda, \lambda_1 \otimes s \otimes \pi)$

et $I(B) = \inf_{\lambda \in B} I(\lambda)$,

B étant un ensemble de probabilités sur $E_x A_x E$.

Le premier résultat obtenu au chapitre VI concernant les chaînes contrôlées est une majoration. On a

$$\overline{\lim_n} \frac{1}{n} \log \sup_{x,\delta} P_x^{\delta} (\hat{F}_n \in K) \leqq I(K)$$

K compact, x valeur initiale de la chaîne et P^{δ} probabilité associée à la stratégie de contrôle δ. Ce résultat uniforme par rapport à δ n'a rien de très surprenant, et vaut dans toutes les situations probabilistes connues. Mais il nécessite un certain travail technique. La situation est affinée dans le cas non contrôlé. On peut alors (plus facilement) obtenir une formule (2), en définissant un changement de probabilité convenable pour rendre la chaîne récurrente (ce qui remplace le simple centrage dans le cas des marches aléatoires, qui est lui aussi l'opération nécessaire pour rendre la marche récurrente). Une fois la chaîne récurrente, on peut appliquer un théorème ergodique et obtenir (2) et des

variantes fortes de (1). La nouvelle probabilité est construite à partir de l'information de la manière qui suit.

Soit $\alpha \in \mathcal{P}(E)$ ensemble des probabilités sur E , $\lambda \in \mathcal{P}(E \times E)$. On dira que $\lambda \in \mathcal{P}^2(\alpha)$ si les deux marginales de λ valent α . Posons alors

$$I_1(\alpha) = \inf \{ I(\lambda), \ \lambda \in \mathcal{P}^2(\alpha) \}$$

avec $I(\lambda) = F(\lambda, \alpha \otimes \pi)$.

Si α est π-invariante, $\alpha \otimes \pi \in \mathcal{P}_2(\alpha)$ est donc $I_1(\alpha) = 0$. Il est donc naturel d'étudier la mesure λ qui réalise $I_1(\alpha)$, si elle existe. Moyennant de faibles hypothèses topologiques et sur π (type transition fellerienne), on montre qu'il existe une telle λ , qui s'écrit donc $\lambda = \alpha \otimes \overline{\pi}$, où $\overline{\pi}$ est une transition de E dans E, telle que $\alpha\overline{\pi} = \alpha$ (donc α est une probabilité invariante pour la chaîne de transition $\overline{\pi}$). Mais surtout on montre que $\overline{\pi}$ est récurrente (récurrente positive au sens de Harris). Et elle définit le changement de probabilité adéquat. Si les $\pi(x,.)$ sont toutes équivalentes on obtient les formules (9) et (10) pour une chaîne de Markov. Diverses variantes peuvent être développées, y compris en direction de la transformée de Cramer. L'étude du cas des chaînes de Markov permet de mieux comprendre encore les liens entre les notions fondamentales introduites plus haut.

II. QUELQUES APPLICATIONS DE LA THÉORIE DES GRANDES DÉVIATIONS.

A) Applications à la statistique.

La première application, qui conditionne toutes les autres, concerne les tests de vraisemblance et d'abord le test de Neyman-Pearson de deux hypothèses simples P_o et P_1 .

Considérons d'abord le cas d'un échantillon ordinaire $X_1 \ldots X_n$ et faisons le test de F_o contre F_1 ($P_o = F_o^{\otimes n}$, $P_1 = F_1^{\otimes n}$). Soit f_o et f_1 les densités de F_o et F_1 par rapport à une mesure dominante, et soit

$$L(X_1, \ldots, X_n) = \sum_{j=1}^{n} \log \frac{f_1(X_j)}{f_o(X_j)}$$

le logarithme du rapport de vraisemblance $\frac{dP_1}{dP_o}$. Le test de Neyman-Pearson a une région de rejet de F_o du type $(L_n > na)$, de niveau $\alpha_n = P_o(L_n > na)$ et de probabilité de fausse alarme $\beta_n = P_1(L_n \leq na)$. Lorsque le seuil a est fixé, on voit d'après la formule (1), que l'on a sensiblement, pour a convenable,

$$\alpha_n \sim e^{-nh(a)}$$

h étant la transformée de Cramer duale de $\log \phi(t)$, où

$$\phi(t) = E_{F_o} \exp t \log \frac{f_1}{f_o}$$

$$= \int f_1^t f_o^{1-t} \, d\nu$$

$\phi(t)$ est donc directement liée à la transformée d'Hellinger de F_o et F_1. Les dérivées de ϕ en $t = 0$ et $t = 1$ valent respectivement $-I(F_o, F_1)$ et $I(F_1, F_o)$. En appliquant toujours (1), pour a fixé, on obtiendra des probabilités de fausse alarme $\beta_n \sim e^{-n\beta}$ où β se calcule en fonction de h et a. Les liens entre α, β, a sont précisés au chapitre II, application directe du chapitre I.

Le calcul est fait aussi pour le cas du test entre deux processus gaussiens stationnaires (afin de justifier une formule utilisée en théorie du signal de manière approchée). L'extension naturelle de ces résultats concerne les

tests de vraisemblance traités par BROWN, puis BIRGÉ, résultats exposés au séminaire, mais rédigés par ailleurs et à paraître.

Si Θ_o et Θ_1 sont les hypothèses à tester $\Theta_o \cup \Theta_1 = \Theta$, ensemble des paramètres, les tests de vraisemblance sont fondés sur une région de rejet du type $(T_n > na)$ où

$$T_n = \frac{\sup_{\theta \in \Theta_1} U_n (X_1 \ldots X_n, \theta)}{\sup_{\theta \in \Theta_o} U_n (X_1 \ldots X_n, \theta)}$$

où $U_n (X_1 \ldots X_n, \theta)$ est la vraisemblance sous P_θ de $X_1 \ldots X_n$.

L'optimalité, au sens de la théorie classique (par exemple de la théorie de la contiguïté) des tests de vraisemblance est difficile à obtenir (bien que récemment on ait commencé à avoir des indications). Par contre la théorie des grandes déviations permet de démontrer cette optimalité, de manière naturelle, mais pour une suite de niveaux exponentiels (ou suivant le résultat de BAHADUR exposé plus loin). En dehors des études générales, signalées plus haut, de nombreux travaux récents ont été faits, y compris dans le cadre séquentiel, où le point de vue grandes déviations est le seul opérationnel à ce jour. Ces études sont menées aux chapitres VII et VIII dans le cadre des chaînes de Markov. Les formules du type (2) permettent d'obtenir une règle d'arrêt asymptotiquement la plus économique (au sens de la moyenne).Les tests séquentiels utilisés ici sont des tests de vraisemblance et les temps d'arrêt sont construits à partir d'estimateurs du maximum de vraisemblance (ou du maximum de contraste). Les solutions asymptotiques obtenues rejoignent beaucoup de procédures empiriques utilisées dans la pratique. Notons que, des méthodes analogues peuvent être développées pour toute classe deproblèmes séquentiels portant sur des processus pour lesquels il existe une théorie de la vraisemblance (ce qui a été fait depuis ce séminaire pour le cas de différents problèmes portant par exemple sur les processus ponctuels).

Revenant aux comparaisons des tests, indiquons que nous n'avons pas développé ici le point de vue et le vocabulaire des travaux de BAHADUR [1],qui

est une forme un peu différente d'utilisation pour ces problèmes des grandes déviations.

Si T_n est une suite de tests de Θ_o contre Θ_1, de répartition G_ϑ sous P_ϑ, alors le niveau observé est défini par

$$R(T_n) = \inf_{O} \{G_\vartheta(T),\ \vartheta \in \Theta_o\}$$

et le test est d'autant meilleur que $R(T_n)$ est petit. En particulier on peut étudier les situations où $\frac{1}{n} \log [1-R(T_n)] \to -h(\vartheta)$, $h(\vartheta)$ est dit pente du test dans cette littérature. Un point de vue voisin également développé par cette même tendance est d'étudier et de comparer les tests pour lesquels il existe une suite $k_n \to \infty$ et p, $0 < p < 1$ tel que $P_\vartheta(T_n > k_n) \to p$, pour $\vartheta \in \Theta_1$, et de montrer qu'alors $\frac{1}{n} \log \alpha_n \to -h(\vartheta)$ (indépendant de p, fausse alarme fixée). Des résultats en ce sens ont été obtenus pour des classes très larges de tests, y compris dans un cadre séquentiel. Plutôt que de comparer les tests par $h(\vartheta)$, le chapitre V donne un exemple de comparaison suivant les probabilités de fausse alarme obtenues pour des niveaux exponentiels, ou suivant les seuils.

Le problème est d'étudier les ruptures de moyenne (plus généralement des ruptures de régression). Le modèle de base est

$$X_i \sim N(\theta, \sigma^2) \qquad 1 \leqq i \leqq r$$

$$X_i \sim N(\theta + d, \sigma^2) \qquad r < i \leqq N$$

(où $\sim$ veut dire suit la loi). ϑ, d, σ^2, r sont des paramètres inconnus, Θ_o correspond à $d = 0$, Θ_1 à $d \neq 0$.

Il existe des procédures classiques pour ces tests, fondées sur le comportement d'une marche aléatoire $S_r = \sum_{k-1}^{r} \hat{\epsilon}_k$, où les $\hat{\epsilon}_k$ sont une suite de variables indépendantes, de loi $N(0, \sigma^2)$, résidus récursifs construits d'après l'observation. Sous Θ_o, S_r est approximable par un mouvement brownien, en remplaçant S_r par des mesures sur $C_o(0,1)$ comme au chapitre IV (voir l'analyse faite plus haut pour les marches aléatoires), alors que sous H_1, S_r a une dérive. Les tests ont pour région de rejet de H_o la région de franchissement

d'une frontière (que l'on traverse en cas de dérive). Les calculs classiques sont rendus imprécis du fait de l'approximation marche aléatoire-mouvement brownien, qui est de l'ordre de $\frac{1}{\sqrt{n}}$, ordre sur lequel porte précisément les comparaisons usuelles. A ces tests on peut opposer le test de vraisemblance. La comparaison et l'étude des différents tests est aussi rendue compliquée par la taille de l'alternative, le caractère discret du paramètre r. Il est donc raisonnable de commencer par comparer ces tests au sens des grandes déviations, ce qui évite d'abord les problèmes d'approximation discret-continu, permet de montrer l'optimalité du test de vraisemblance et donne donc une référence. De plus certaines procédures très simples sont introduites, dont les performances au sens des grandes déviations sont assez bonnes. On a donc intérêt à en faire des études empiriques pour voir leurs performances pour des échantillons plus courts.

Une critique méthodologique est souvent faite à ces études de grandes déviations parce qu'elles portent sur des échantillons trop longs, mais surtout sur des probabilités totalement inaccesibles. Cette critique peut être fondée (sauf par exemple dans le domaine des communications, où la valeur de β est décisive). On peut y répondre cependant en doutant que les valeurs de mesure du paramètre dans le cadre classique ait un sens toujours clair et surtout par un argument de mathématique heuristique : dans bien des cas, les classements obtenus par grande déviation et par des méthodes classiques (pour des niveaux constants et éventuellement des hypothèses contigues) sont les mêmes. Les grandes déviations ont l'avantage de répondre aux problèmes de classement des tests et de recherche de solutions optimales, là où (à ce jour) les méthodes classiques sont en difficulté. La défense de ces techniques au niveau de la méthodologie nous semble donc possible.

L'exposé du chapitre IX qui termine la partie statistique est marginal en regard des autres exposés (et beaucoup plus théorique). Il comporte essentiellement l'amélioration d'un résultat connu de LE CAM sur l'estimation d'un paramètre à valeurs dans un espace métrique de dimension finie. Il met en évidence le lien

entre les majorations du type (1) et la théorie du maximum de vraisemblance sur une suite de réseaux.

Tous les résultats statistiques présentés dans le séminaire sont originaux.

B) Les applications à la littérature électronique.

Le dernier exposé est une revue des applications des grandes déviations au domaine du codage, de la transmission avec des développements sur des modèles du type ALOHA de communication par satellite. Cet exposé est un exemple de ce que peuvent apporter des idées mathématiques relativement simples à un domaine appliqué. Outre la construction de bornes et la simplification des méthodes pour les problèmes liés à la théorie de SHANNON, le changement de probabilité des grandes déviations est l'outil de la méthode de simulation dite "importance sampling". Le domaine offre de nombreuses voies de travail.

La littérature électronique n'est bien entendu pas la seule application des grandes déviations. Signalons une démonstration simple et éclairante des lois de Zipf en linguistique [8].

BIBLIOGRAPHIE:(Quelques références générales, voir après chaque exposé une bibliographie spécialisée)

[0] AZENCOTT R.
Cours de l'Ecole d'Eté de Saint-Flour, 1978
A paraître dans Lectures notes, Springer-Verlag

[1] BAHADUR (1962)
Stochastic comparison of tests
Ann. Math. Stat. 31, p. 276-295

[2] CHERNOFF H. (1952).
A measure of asymptotic efficiency for tests based on the sum of observations
Ann. Math. Stat. 23, p. 493-507

[3] CRAMER H. (1938)
Sur un nouveau théorème limite de la théorie des probabilités
Act. Sc. Ind. n° 736

[4] DOBRUSHIN (1955)
A general formulation of the fundamental theorem of Shannon
Usp. Nat. Nauk, vol. 6, p. 3-104

[5] FELLER W.
An introduction to probability theory and its applications (Wiley)

[6] KHINCHIN (1929)
Ober eine neuer grenzwertats des Wahrscheinlichertreihnung
Math. Ann. vol. 101, p. 745-752

[7] RAGHAVACHARI (1970)
On a theorem of Bahadur
Ann. Math. Stat. 41, p. 1695-1699

[8] ROUAULT A.
Démonstration simple des lois de Zipf (à paraître)

[9] SANOV (1957)
On the probability of large deviations of random variables
Mat. Sb. n° 5, 42, p. 11-44.

D. DACUNHA-CASTELLE
Mathématiques - Bât. 425
ERA CNRS 532 "Statistique Appliquée"
Université Paris-Sud
91405 ORSAY