Séminaire étudiant du $\mathbb{LSP}$

Un petit peu de grandes déviations

Djalil Chafaï <chafai@cict.fr>, 12 janvier 1999

La théorie des grandes déviations a pour objet l'étude de la vitesse de convergence vers zéro des évènements rares dans les théorèmes de convergence en loi. Le cas typique est celui des grandes déviations par rapport à la moyenne dans les théorèmes ergodiques.

Principe de Grandes Déviations (Varadhan 1966)

On dit qu'une famille de probas $(\mu_\varepsilon)_{\varepsilon>0}$ sur un espace mesurable $\mathbf{E}$ satisfait à un principe de grandes déviations (**PGD**) de vitesse décroissante $a_\varepsilon \underset{\varepsilon\to 0}{\to} 0$ et de fonction de taux $\Phi : \mathbf{E} \to [0,\infty]$ ssi
1°) $\forall\lambda \in \mathbb{R}, \Phi^{-1}((-\infty,\lambda])$ est un compact de $\mathbf{E}$ (donc Φ est s.c.i).
2°) Pour tout borélien $A \subset \mathbf{E}$ on a

$$-\inf_{\overset{\circ}{A}}\Phi \le \liminf_{\varepsilon\to 0} a_\varepsilon \log \mu_\varepsilon(A) \le \limsup_{\varepsilon\to 0} a_\varepsilon \log \mu_\varepsilon(A) \le -\inf_{\bar{A}}\Phi$$

Pour abréger, on parlera de (a_ε,Φ)-**PGD**. On parlera de **PGD** faible lorsque l'on saura seulement que Φ a des ensembles de niveau fermés (i.e. s.c.i.) et que la borne supérieure a lieu pour des A relativement compacts. On peut montrer qu'une fonction de taux est unique et qu'elle atteint on minimum sur $\mathbf{E}$, qui vaut 0.

Théorème de Cramer (1938) - Chernoff (1952)

Soit μ une proba sur $\mathbb{R}$ possédant des moments exponentiels de tous ordres. Alors $\mu^*(n\bullet)$ satisfait à un $\left(\frac{1}{n},\Lambda_\mu^*\right)$-**PGD**. Λ_μ^* étant la transformée de Cramer de μ définie par

$$\Lambda_\mu^*(\lambda) := \sup_\theta [\theta\lambda - \log \mathbf{L}_\mu(\theta)]$$

C'est la transformée de Legendre de sa «log-Laplace» Λ_μ. Elle est positive, convexe, s.c.i., tend vers $+\infty$ à l'infini, décroit sur $(-\infty,m_\mu)$, croit sur $(m_\mu,+\infty)$ et possède un unique minimum nul en m_μ.

Cramer a prouvé ce théorème dans $\mathbb{R}$ sur les demi-droites pour des probas non singulières par rapport à la mesure de Lebesgue, Chernoff leva cette condition. Donsker et Varadhan l'ont généralisé à un Banach en 1976, Bahadur et Zabell l'ont ensuite généralisé en 1979 aux e.v.t.l.c.s possédant une topologie d'e.v.t.l.c.s plus fine induisant une topologie polonaise sur un convexe fermé μ−presque-sûr. Ils ont dégagé une méthode générale dans ce cadre. Voir [1] et [4] par exemple.

Exemples de transformées de Cramer dans $\mathbb{R}$

Bernoulli $q\delta_1+(1-q)\delta_0$ avec $0<q<1$	$x\log\frac{x}{q}+(x-1)\log\frac{q-1}{x-1}$ si $0\le x\le 1$ et $+\infty$ sinon
Poisson de paramètre $\lambda>0$	$\lambda-x+x\log\frac{x}{\lambda}$ si $x>0$ et $+\infty$ sinon
Exponentielle de paramètre $\lambda>0$	$\lambda x-1-\log(\lambda x)$ si $x>0$ et $+\infty$ sinon
Gaussienne de moyenne m et de variance $\sigma^2>0$	$(x-m)^2/2\sigma^2$

Théorème de Sanov (1956) - Donsker & Varadhan (1976)

Soit μ une proba sur un espace polonais $\mathbf{E}$ (esp. top. séparable possédant une métrique complète) et $L_n : \sigma \in \mathbf{E}^n \to \frac{1}{n}\sum_{m=1}^n \delta_{\sigma_m} \in Prob(\mathbf{E})$. Alors la loi μ_n de L_n satisfait sur l'espace des probas sur $\mathbf{E}$ muni de la topologie de la convergence en loi à un **PGD** de vitesse n^{-1} dont la fonction de taux est l'entropie relative

$$\mathbf{H}(\nu|\mu) \equiv \begin{cases} \displaystyle\int_\Omega \frac{d\nu}{d\mu}\log\left(\frac{d\nu}{d\mu}\right)d\mu = \int_\Omega \log\left(\frac{d\nu}{d\mu}\right)d\nu & \text{si } \nu \ll \mu \\ +\infty & \text{sinon} \end{cases}$$

Sanov a prouvé ce théorème dans $\mathbb{R}$ mais beaucoup de gens n'y ont pas cru. La version ci-dessus est due à Donsker et Varadhan. Bahadur et Zabell l'ont retrouvé par une méthode plus générale en 1979. Voir [1]. Lorsqu'on restreint ce théorème aux lois sur un espace d'états finis, on retrouve le théorème de Cramer-Chernoff en identifiant $\mathbb{R}^p$ avec l'ensemble des mesures sur l'ensemble fini $\{1\ldots p\}$. L'entropie relative devient alors une transformée de Cramer.

Considérée comme probabilité sur $Mes(\mathbf{E})$, μ_1 a pour transformée de Cramer la fonction valant $\mathbf{H}(\bullet|\mu)$ sur $Prob(\mathbf{E})$ et $+\infty$ sur $Mes(\mathbf{E})\backslash Prob(\mathbf{E})$.

Théorème de Schilder (1962)

Considérons la mesure de Wiener μ sur l'espace des fonctions continues sur $[0,1]$ nulles en 0 à valeur dans $\mathbb{R}^d$. C'est la loi d'un brownien d-dimentionnel standard. $\mu_\varepsilon := \mu\left(\sqrt{\varepsilon}\bullet\right)$ satisfait a un **PGD** de vitesse ε dont la fonction de taux vaut $\frac{1}{2}\int_0^1 |f'(t)|^2\,dt$ si f est absolument continue et $+\infty$ sinon.

Ce théorème apparait pour la première fois dans la thèse de Shilder, qui était un élève de Donsker. Voir par exemple [4]. Ventcel et Friedlin on utilisé ce théorème pour obtenir un PGD pour la solution d'une EDS (systèmes dynamiques). Voir par exemple [4].

Tension exponentielle et PGD faibles

On dit qu'une suite de probas $(\mu_\varepsilon,\varepsilon>0)$ sur un espace mesurable $\mathbf{E}$ est exponentiellement tendue ssi pour tout $L>0$ il existe un compact K_L de $\mathbf{E}$ tel que $\varepsilon\log\mu_\varepsilon(K_L)\le -L$. Avec de la tension exponentielle un **PGD** faible devient un **PGD**. La réciproque est vraie dans certains cadres.

Principe de contraction

Soit μ_n satisfaisant à un (a_n,Φ)-**PGD** sur un espace mesurable $\mathbf{E}$ et $f:\mathbf{E}\to\mathbf{F}$ continue alors $\mu_n\circ f^{-1}$ satisfait sur l'espace $\mathbf{F}$ à un **PGD** de vitesse a_n et de fonction de taux $\inf_{f^{-1}(.)}\Phi$.

Obtention de bornes supérieures de PGD par la méthode de Cramer

Il s'agit d'une méthode générale pour obtenir la borne supérieure d'un **PGD** pour $(\mu_\varepsilon, a_\varepsilon > 0)$. Ici, on se place sur un espace de Banach $\mathbf{E}$ (on pourrait prendre plus général encore). Supposons que la fonction définie sur $\mathbf{E}'$ par

$$\Lambda(\bullet) := \lim_{\varepsilon\to 0} a_\varepsilon \left\langle \exp\left(\bullet(x)/a_\varepsilon\right)\right\rangle_{\mu_\varepsilon}$$

est bien définie et à valeurs dans $[-\infty, +\infty]$. Alors Λ est convexe sur $\mathbf{E}'$ et sa transformée de Legendre définie par

$$\Lambda^*(\bullet) := \sup_{\lambda\in\mathbf{E}'} \lambda(\bullet) - \Lambda(\lambda)$$

est convexe, s.c.i. positive. De plus, pour tout compact K de $\mathbf{E}$, on a

$$\limsup_{a_\varepsilon\to 0} a_\varepsilon \log \mu_\varepsilon(K) \leq -\inf_K \Lambda^*$$

Il ne reste donc plus qu'à démonter la tension exponentielle pour obtenir la borne supérieure d'un $(a_\varepsilon, \Lambda^*)$-**PGD**. Voir par exemple [1], [4] et [2]. Cette méthode ne permet d'obtenir que des fonctions de taux convexes, ce qui n'est pas le cas de toutes les fonctions de taux comme on peut le voir grâce au principe de contraction.

Principe de Laplace (Varadhan 1966)

Soit $(\mu_\varepsilon, \varepsilon > 0)$ vérifiant un (a_ε, Φ)-**PGD** sur un espace mesurable $\mathbf{E}$. Soit $f : \mathbf{E} \to \mathbb{R}$ continue telle que

$$\lim_{L\to+\infty} \limsup_{\varepsilon\to 0} \varepsilon \log \left\langle \exp(f/\varepsilon)\, \mathbb{1}_{f\in[L,+\infty)}\right\rangle_{\mu_\varepsilon} = -\infty$$

(c'est par exemple le cas quand f est bornée). Alors on a

$$\lim_{\varepsilon\to 0} \varepsilon \log \left\langle \exp(f/\varepsilon)\right\rangle_{\mu_\varepsilon} = \sup_{\mathbf{E}} (f - \Phi)$$

Pour le prouver, on peut montrer l'$\leq$ pour les f s.c.s. et l'$\geq$ pour les f s.c.i.. Voir par exemple [4].

Principe de Gibbs (Ellis 1985)

On se place dans le même contexte que précédement. Définissons

$$\nu_\varepsilon(\bullet) := \frac{\left\langle \exp(f/\varepsilon)\, \mathbb{1}_\bullet\right\rangle_{\mu_\varepsilon}}{\left\langle \exp(f/\varepsilon)\right\rangle_{\mu_\varepsilon}}$$

Alors $(\nu_\varepsilon, \varepsilon > 0)$ vérifie un $(\varepsilon, \Phi - f - \inf(\Phi - f))$-**PGD** sur $\mathbf{E}$. Voir [5].

Un problème d'EDP (Donsker & Varadhan 1983)

Soit $V \in C_b\left(\mathbb{R}^d\right)$ et $L = \frac{1}{2}\sum_{i,j} a_{i,j}\partial_i\partial_j + \sum_i b_i\partial_i$. Soit $(X_s, s \geq 0)$ un processus de diffusion de générateur L. On considère l'EDP suivante :

$$\begin{cases} \partial_t u &= Lu - Vu \\ u(0,\bullet) &\equiv 1 \end{cases}$$

La formule de Feynman-Kac nous donne une représentation probabiliste d'une solution :

$$u_\varepsilon(t,x) = \mathbb{E}^x\left(\exp\left(-\int_0^t V(X_s)\,ds\right)\right)$$

Considérons la mesure empirique $\Sigma_t := \frac{1}{t}\int_0^t \delta_{X_s} ds$ et la fonction $f_V(\mu) := -\langle V\rangle_\mu$. La formule de Feynman-Kac s'écrit alors

$$u_\varepsilon(t,x) = \mathbb{E}^x\left(\exp\left(t f_V(\Sigma_t)\right)\right)$$

Si $\left(\Sigma_t, t^{-1}, t > 0\right)$ vérifie un $\left(t^{-1}, \Phi(\mu)\right)$-PGD, le principe de Laplace donne alors :

$$\lim_{t\to\infty} \frac{1}{t} \log u(t,x) = \sup_{\mu\in Prob\left(\mathbb{R}^d\right)} f(\mu) - \Phi(\mu)$$

Remarque

Les **PGD** énoncés ici concernent en quelque sorte des mesures obtenues en sommant des acroissements indépendants. On peut en fait affaiblir cette hypothèse dans certains cas. Il existe par exemple des résultats analogues au théorème de Sanov pour des processus de Markov sous une hypothèse d'ergodicité. Ils contiennent le théorème de Sanov classique. Voir par exemple [4] et [9].

Références

[1] R. Azencott, *Grandes déviations et applications*, in École d'été de probabilités de Saint-Flour VIII 1978, vol. 774 of Lecture Notes in Mathematics, Springer, 1980, pp. 1–172.

[2] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Jones and Barlett, 1993.

[3] ———, *Large Deviations Techniques and Applications, Second Edition*, Springer, 1998.

[4] J.-D. Deuschel and D. W. Strook, *Large Deviations*, Academic Press, 1989.

[5] R. Ellis, *Entropy, Large Deviations and Statistical Mechanics*, Springer, 1985.

[6] D. W. Strook, *An Introduction to the Theory of Large Deviations*, Springer, 1984.

[7] Étudiants du LSP, *Les grandes déviations sans larmes*, Publications du LSP, 1995.

[8] S. R. S. Varadhan, *Large Deviations and Applications*, SIAM, 1984.

[9] ———, *Larges deviations*, in École d'été de probabilités de Saint-Flour XV 1985, vol. 1362 of Lecture Notes in Mathematics, Springer, 1988, pp. 1–50.