V) Statistiques adaptatives et inégalités exponentielles pour Processus de Poisson

Rmq $E\left(\int f\, dN\right) = \int f\, d\mu$ Var $\int f\, dN = \int f^2\, d\mu$ (si il faut écrire qd $f = \sum_I \alpha_I$

1) <u>Probability generating functional</u> (p.g.fl)

Soit h une fonction test >0, la pgfl du processus ponctuel N pris sur h est

$$E\left(\prod_{x \in N} h(x)\right).$$

2) <u>Pour le Poisson.</u> si $h = \exp(f)$.

$$E\left(\prod_{x \in N} h(x)\right) = E\left(\exp \int_{\mathcal{X}} f(t)\, dN_t\right). \quad (*)$$

Supposons f constante par morceaux $f = \sum_I \alpha_I \mathbb{1}_I$.

alors $(*)$ devient par indépendance du processus de Poisson

$$= \prod_I E\left(\exp \alpha_I \int_I \mathbb{1}_I\, dN_t\right)$$

$$= \prod_I E\left(\exp\left[\alpha_I N_I\right]\right)$$

or si $X \sim \mathcal{P}(\lambda)$ alors $E\left(e^{tX}\right) = \sum \frac{\lambda^k}{k!} e^{-\lambda} e^{tk}$

$$= \boxed{e^{(e^t - 1)\lambda}}$$

donc

λ mesure moyenne $= \lambda(x)\,dx$.

$$= \prod_I e^{\mu(I)\left(e^{\alpha_I} - 1\right)}$$

$$= e^{\int_{\mathcal{X}} \left(e^{f(x)} - 1\right) \lambda(x)\, dx}.$$

De manière générale,

<u>Thm</u> Soit f bornée mesurable,

$$E\left[e^{\int f\, dN}\right] = e^{\int (e^f - 1)\lambda\, dx}.$$

Ou

$$E\left(e^{\int f (dN - \lambda dx)}\right) = e^{\int \underbrace{(e^f - f - 1)}_{\phi(f)} \lambda\, dx}$$

3) Inégalité exponentielle (1ère).

Thm
$$\mathbb{P}\Big(\int f(x)\,(dN-\mu(dx)) \geq \xi\Big) \leq \exp\Bigg(-\frac{\int f^2\,d\mu}{\|f\|_\infty^2}\, h\Big(\frac{\xi\,\|f\|_\infty}{\int f^2\,d\nu}\Big)\Bigg).$$

avec $h(u) = (1+u)\ln(1+u) - u$.

Preuve On applique le thm précédent à λf.

Puis $\forall \lambda > 0$
$$\mathbb{P}\Big(\int f\, dN - d\mu \geq \xi\Big) = \mathbb{P}\Big(\int \lambda f\, dN - d\mu \geq \lambda\xi\Big)$$
$$= \mathbb{P}\Big(e^{\int \lambda f (dN-d\mu)} \geq e^{\lambda\xi}\Big) \leq e^{-\lambda\xi}\, \mathbb{E}\Big(e^{\int \lambda f(dN-d\mu)}\Big)$$
$$\leq e^{-\lambda\xi + \int \phi(\lambda f)\,d\mu}.$$

Or $\displaystyle \phi(\lambda f(x)) = \sum_{k=2}^{\infty} \frac{(\lambda f(x))^k}{k!} \leq \sum_{k=2}^{\infty} \frac{\lambda^k \|f\|_\infty^{k-2}}{k!}\, f^2(x)$.

donc si $v = \int f^2\,d\mu$ (Remarque c'est la variance de $\int f\,dN$).

$$\mathbb{P}\Big(\int f\, dN - d\mu \geq \xi\Big) \leq \exp\Big[-\lambda\xi + \frac{v}{\|f\|_\infty^2}\,\phi(\lambda\|f\|_\infty)\Big].$$

on dérive en λ $\quad -\xi + \frac{v}{\|f\|_\infty^2}\Big[\|f\|_\infty e^{\lambda\|f\|_\infty} - \|f\|_\infty\Big] = 0$.

ie $\quad \dfrac{v}{\|f\|_\infty}\big(e^{\lambda\|f\|_\infty} - 1\big) = \xi$

$$e^{\lambda\|f\|_\infty} = 1 + \frac{\xi\|f\|_\infty}{v}$$
$$\lambda = \frac{1}{\|f\|_\infty}\ln\Big(1 + \frac{\xi\|f\|_\infty}{v}\Big).$$

avec ce λ.

$$\leq \exp\Big[-\frac{\xi}{\|f\|_\infty}\ln\Big(1+\frac{\xi\|f\|_\infty}{v}\Big) + \frac{v}{\|f\|_\infty^2}\Big[1+\frac{\xi\|f\|_\infty}{v} - \ln\Big(1+\frac{\xi\|f\|_\infty}{v}\Big) - \cancel{1}\Big].$$

$$\Big[-\frac{v}{\|f\|_\infty^2}\Big[\Big(1+\frac{\xi v}{\|f\|_\infty}\Big)\ln\Big(1+\frac{\xi\|f\|_\infty}{v}\Big) - \frac{\xi\|f\|_\infty}{v}\Big].$$

C'est l'équivalent de la Bennett.

mais on a aussi l'équivalent de la Bernstein (version faible ici...)

Preuve rigolote :

$$P\left(e^{\lambda\int f(x)\,dN-\mu(dx)-\int(e^{\lambda f}-\lambda f-1)\,d\mu}\geq e^{x}\right)$$

$$\leq E\left(e^{\lambda\int f\,dN-\mu(dx)}\right)e^{-\int e^{\lambda f}-\lambda f-1}\,e^{-x}=e^{-x}$$

donc $P\left(\int f(x)\,dN-\mu(dx)\geq\frac{1}{\lambda}\int(e^{\lambda f}-\lambda f-1)\,d\mu+\frac{x}{\lambda}\right)\leq e^{-x}$

vrai $\forall\lambda>0$

or $(e^{\lambda f}-\lambda f-1)\leq\sum_{k=2}^{\infty}\frac{\lambda^k}{k!}\|f(x)\|_\infty^{k-2}f(x)^2$

$$\leq\frac{\lambda^2 f(x)^2}{2\left(1-\frac{\lambda\|f\|_\infty}{3}\right)}$$

$\forall\lambda>0$

donc $P\left(\int f(x)\,dN-\mu(dx)\geq\underbrace{\frac{\lambda v}{2\left(1-\frac{\lambda\|f\|_\infty}{3}\right)}+\frac{x}{\lambda}}\right)\leq e^{-x}$

on cherche λ tq soit minimal.

.... le minimum vaut $\sqrt{2vx}+\frac{\|f\|_\infty}{3}x$.

Donc $P\left(\int f(x)\,dN-d\mu\geq\sqrt{2vx}+\frac{\|f\|_\infty}{3}x\right)\leq e^{-x}$

(et Bernstein en iid)

4) Lien avec le seuillage

On veut estimer l'intensité $\lambda(x)$ du PP qu'on observe.

On suppose $\lambda(x)$ dans $L^2(\mathbb{R})$.

On se donne une BON $(\varphi_k)_{k\in\Lambda}$ (typiquement ondelette pour que ça marche bien jusqu'au bout).

donc $\lambda(x)=\sum_k\alpha_k\varphi_k$

On prend un sous ens typiquement fini de Λ : m

et sur m on estime tous les α_k

or $\alpha_k=\int\varphi_k\,\lambda(x)\,dx$ $\xrightarrow[\text{estimé sans biais par}]{}$ $\hat\alpha_k=\int\varphi_k\,dN$.

Donc on peut $\left.\begin{matrix}\text{reconstruire}\\\text{estimer}\end{matrix}\right)$ λ par $\hat\lambda_m=\sum\hat\alpha_k\varphi_k$

Mais si m mal choisi c'est une catastrophe :

en effet $\int(\hat\lambda_m-\lambda)^2\,dx=\sum_{k\in\Lambda\setminus m}\alpha_k^2+\sum_{k\in m}(\hat\alpha_k-\alpha_k)^2$

$$E(\quad)=\sum_{k\in\Lambda\setminus m}\alpha_k^2+\sum_{k\in m}\int\varphi_k^2\,d\mu.$$

Premier terme $\sum_{k\in\Lambda\setminus m} \alpha_k^2$ biais $\searrow$ si Γ grandit

Deuxième terme $\sum_{k\in m} \int \varphi_k^2\, d\mu$ $\nearrow$ si Γ grandit.

Le meilleur : $\Gamma = \{k \,/\, \int \varphi_k^2\, d\mu < \alpha_k^2\}$ ensemble "oracle" $\hookrightarrow \sum \min(\alpha_k^2, \int \varphi$

<u>Idée</u> On se fixe quand même un Λ_0 grand.

et on ne garde k que si $\hat{\alpha}_k$ est assez grand

<u>Problème</u> : Que veut dire assez grand ?

Si $\alpha_k = 0$ $\quad P\left(|\hat{\alpha}_k| \geq \sqrt{2 v_k x} + \frac{\|\varphi_k\|_\infty x}{3}\right) \leq 2e^{-x}$

où $v_k = \int \varphi_k^2\, d\mu = \int \varphi_k^2 f(x)\, dx$.

Donc si $x = \gamma \ln n$

$$P\left(\hat{\alpha}_k \geq \sqrt{2 v_k \frac{x}{n}} + \frac{\|\varphi_k\|_\infty}{3} \frac{x}{n}\right) \leq 2n^{-\gamma}$$

e^{-x}

ici garder les x

Mais v_k dépend de f

or $P\left(\hat{v}_k - v_k \geq \sqrt{2 V_k x} + \frac{\|\varphi_k\|_\infty^2}{3} x\right) \leq e^{-x}$

avec $\hat{v}_k = \int \varphi_k^2\, dN$

et $V_k = \int \varphi_k^4\, d\mu \leq v_k \|\varphi_k\|_\infty^2$.

$\rightsquigarrow$ On va garder $\hat{\alpha}_k$ si $|\hat{\alpha}_k| \geq \eta_k$.

avec $\eta_k = \sqrt{2_{(1+\varepsilon)} \hat{v}_k \frac{x}{n}} + K(\varepsilon) \frac{\|\varphi_k\|_\infty}{3} \frac{x}{n}$ Pour $\varepsilon > 0$ fi... et x bien c...

Alors on peut mq (sous des hyp sur choix base etc...) (article Rivoirard RB...

Thm

si $\hat{\alpha} = \sum_{k\in\Lambda_0} \hat{\alpha}_k \mathbb{1}_{|\hat{\alpha}_k| \geq \eta_k} \varphi_k$

Alors

$$E\left(\|\alpha - \hat{\alpha}\|^2\right) \leq \square \left[\sum_{k\notin\Lambda_0} \alpha_k^2 + \sum_{k\in\Lambda_0} \min\left(\alpha_k^2 ; v_k \ln n\right)\right] + \square$$

(et c'est la borne asymptotique i.e. négligeable devant l'autre montré si on voit un n échantillon et $n \to +\infty$

et $x = \gamma \ln n$.

5) Sélection de modèles

Principe : on se donne un contraste $\nearrow$ -log vraisemblance $\searrow$ moindres carrés

Ici on se concentre sur deuxième choix

$$\gamma(f) = -\int 2 f \, dN + \int f^2 dx .$$

Pourquoi est-ce un contraste ? $E_\lambda(\gamma(f)) = -2\int f\lambda + \int f^2$
$= \int (f-\lambda)^2 - \int \lambda^2$.

donc minimal quand $f = \lambda$.

quand $f = \sum_{k \in m} \beta_k \varphi_k$ β_k inconnu

alors quand on minimise $\gamma(f)$ pour trouver les meilleurs β_k

on trouve

$$\gamma(f) = \sum_{k \in m} \Big[-2\beta_k \underbrace{\int \varphi_k \, dN}_{\hat\alpha_k} + \beta_k^2 \Big]$$

C'est donc minimal quand $\beta_k = \hat\alpha_k$ et ça vaut $-\sum_{k \in \Gamma} \hat\alpha_k^2$.

ie $\hat\lambda_m = \underset{f \text{ écrit ru } m}{\operatorname{argmin}} \gamma(f)$

donc pour tout sous ensemble m de Λ on peut faire ça.
$=$ "modèle"

Comment choisir m ?

Ça dépend de la famille de modèles

Soit $\mathcal{M}$ une famille de modèle

$$\hat m = \underset{m \in \mathcal{M}}{\operatorname{argmin}} \gamma(\hat\lambda_m) + \text{pen } m .$$

si $\mathcal{M} = \{$ tous les sous ensembles possibles de $A_0 \}$

et $\text{pen } m = \sum_{k \in m} \eta_k^2$

alors $$- \sum_{k \in m} \hat\alpha_k^2 + \sum_{k \in m} \eta_k^2 = \sum_{k \in m} \left[\eta_k^2 - \hat\alpha_k^2 \right]$$

donc si on veut minimiser ça pour un choix de m

on a intérêt à prendre tous les k / $\hat\alpha_k^2 > \eta_k^2$ et pas les autres

ie $$\hat\lambda_{\hat m} = \sum_{k \in \Lambda_0} \hat\alpha_k \, \mathbb{1}_{|\hat\alpha_k| > \eta_k} \varphi_k$$ c'est l'estimateur par seuill[age]

(trucs + gnrx évidemment ...

Mais alors du coup à quoi ça sert ?

Avantages :	Inconvénients
. Plus général ! Et $\hat{m}$ on est pas forcé de tout écrire sur une base L^2 mais juste avec des sev (voir convexes..)	. En général il faut calculer tous les $\hat{s}_m$ $\longrightarrow$ très demandeur en calcul.
. Inégalité oracle .. possibilité de perdre le log.	. Il faut une famille généralement finie de modèles $\hat{m}$ si ça peut grossir avec n.
	. hyp du style borné, support borné etc ...

Lien avec la Concentration ? Vrai pour selec modèle cf Livre Pascal Massart — Concentration Inequalit and Model selection

Pour tout m,

$$\underbrace{-\sum_{k\in\hat{m}}\hat{\alpha}_k^2}_{\gamma(\hat{s}_{\hat{m}})} + \text{pen}\,\hat{m} \leq \underbrace{-\sum_{k\in m}\hat{\alpha}_k^2}_{\gamma(\hat{s}_m)} + \text{pen}\,m \leq \gamma(s_m) + \text{pen}\,m.$$

mais $\gamma(f) = \|f-s\|^2 - \|s\|^2 - 2\int f\,[dN - s(x)dx]$.

donc
$$\|\hat{s}_{\hat{m}} - s\|^2 \leq \|s_m - s\|^2 + \text{pen}\,m - 2\int s_m\,(dN - s(x)dx) + 2\int \hat{s}_{\hat{m}}\,(dN - s(x)dx) - \text{pen}\,\hat{m}.$$

$$\leq \|s_m - s\|^2 + \text{pen}\,m - 2\int (s_m - s_{\hat{m}})(dN - s(x)d \;+ \underbrace{2\int(\hat{s}_{\hat{m}} - s_{\hat{m}})(dN - s(x)dx)}_{\text{B}} - \text{pen}\,\hat{m}.$$

A presque de moyenne nulle

$$\text{B} = 2\sum_{k\in\hat{m}}(\hat{\alpha}_k - \alpha_k)\int \varphi_k\, dN - s(x)\,dx -$$
$$= 2\sum_{k\in\hat{m}}(\hat{\alpha}_k - \alpha_k)^2 = 2\,\chi^2(\hat{m})\ldots$$

et donc on choisit pen /

$$P(2\chi^2(m) \geq \text{pen}\,m) \quad \text{petit ... uniformément en } m.$$

Donc on cherche une inégalité exponentielle pour

$$\chi(m) = \sup_{\substack{f\in S_m\\ \|f\|=1}} \int f\,[dN - s(x)dx].$$

en effet $f = \sum \beta_k\varphi_k$ $\Rightarrow$ $\sum\beta_k(\hat{\alpha}_k - \alpha_k)$ et on maximise avec $\sum\beta_k^2 = 1$

Thm : Pour toute famille φ_a de fonctions $\|\varphi_a\|_\infty \leq b$

on considère $Z = \sup_{a \in A} \left| \int \varphi_a \, dN - \mu(dx) \right|$

et $v_0 = \sup_{a \in A} \int \varphi_a^2(x) \, \mu(dx)$.

Alors

$$P\left(Z \geq (1+\varepsilon) E(Z) + \sqrt{2 \kappa v_0 x} + K(\varepsilon) b x \right) \leq e^{-x}$$

avec $\kappa = 6$ et $K(\varepsilon) = 1.25 + 32/\varepsilon$.

Preuve : ID + Talagrand.

Cela appliqué à $\chi(m)$ on obtient

$$P\left(\chi(m) \geq (1+\varepsilon) \sqrt{\sum_{k \in m} v_k} + \sqrt{2 \kappa \|\lambda\|_\infty x} + K(\varepsilon) B x \right) \leq e^{-x}.$$

dépend de la base.

donc on voit que +/_ obligatoire que la pénalité utilise $\|\lambda\|_\infty$.
(ou une estimation).

Puis

si pen $m \geq \square \left[\sqrt{\sum_{k \in m} \hat{v}_k} + \sqrt{2 \kappa \hat{M} L_m D_m} \right]^2$.

avec $\hat{M}$ estimation de la $\|\lambda\|_\infty$.

et $\sum_{m \in \mathcal{M}} e^{-L_m D_m} \leq 1$.

alors $E \|\hat{\lambda}_{\hat{m}} - \lambda\|^2 \leq \square \inf \left[\|\lambda - \lambda_m\|^2 + \sum_{k \in m} v_k^2 + M L_m D_m \right] + \frac{1}{n}$

si famille pas trop complexe $L_m = 1$
et on n'a plus le $\ln n$ du seuillage.
(En particulier ça ne marche pas pour la famille du seuillage).

Rmq : on peut aussi faire des tests adaptatifs (ie alternative +/_ régulière et on ne sait pas sa régularité à l'avance...)