

Statistiques Quantiques

Jonas KAHN

16 octobre 2004

Les statistiques, «science de l'État» à l'origine, s'attachent au problème mathématique de remonter à la loi d'un phénomène aléatoire à partir de la distribution expérimentale d'un échantillon de ce phénomène (ou d'une transformation de celui-ci), avec autant de précision que possible, dans des sens à définir. On peut les voir aussi, peut-être moins formellement, comme le problème consistant à extraire le maximum d'information, ou les informations les plus «intéressantes», d'un jeu de données.

Dans les deux cas, cependant, on part d'un jeu de données acquis. De fait, la question du choix des données dont on peut disposer est en général peu pertinente, comme, si on peut obtenir différentes informations, il est en théorie possible de les obtenir toutes à la fois, abstraction faite des problèmes de moyens humains ou financiers. Toutefois, si on s'intéresse à la mécanique quantique, il est plus fructueux de voir les statistiques comme le problème d'extraire le maximum d'informations d'un *objet* d'étude, avant acquisition des données. En effet, comme nous allons le voir ci-après, les lois de la mécanique quantique sont telles que, si nous disposons de nombreuses *mesures* possibles, apportant chacune son lot d'information, il est impossible, *même en théorie*, de les effectuer *simultanément*.

Un des aspects les plus originaux des statistiques quantiques est donc le problème de savoir quelle mesure effectuer pour obtenir le maximum d'informations (ou l'information que nous désirons le plus) d'un objet quantique. C'est à ce genre de questions, d'abord étudiées par Helstrom et Holevo [2] [3], que s'intéresse notamment Richard Gill à Utrecht [1].

D'abord, il est néanmoins nécessaire de se demander quelle information on peut tirer d'une mesure. Mon rapport de DEA consiste en des résultats exacts dans un cas particulier.

Pour que tout ceci ait un sens un peu plus solide il convient de décrire à quels modèles peut donner lieu la physique quantique.

1 Notions de mécanique quantique

La physique s'intéresse à des systèmes, et plus précisément aux lois régissant l'évolution de ces systèmes. À un instant donné, un système sera dans un certain *état*. Tout objet mathématique contenant toute l'information quant aux *interactions* éventuelles avec d'autres systèmes est une représentation de cet état.

En mécanique classique, un jeu de variables canoniques (vitesse et position) pourra décrire cet état, en général comme un élément d'un espace vectoriel, l'espace des phases. Il est de plus théoriquement possible de mesurer simultanément chacune des coordonnées de l'état dans cet espace. Notre état peut avoir été généré de manière aléatoire, auquel cas l'état est décrit par un nuage dans l'espace des phases, mais il est toujours possibles d'échantillonner un point de cet espace des phases. On a donc accès à toutes les fonctions de cet espace des phases, qui dans le cas aléatoire, deviennent des variables aléatoires. On appelle ces fonctions les *observables* du système.

En mécanique quantique, l'une des manières les plus courantes de représenter un état est comme un *opérateur densité* :

Définition 1.1. : Opérateur densité

Un opérateur densité, habituellement noté ρ , est un opérateur sur un espace de Hilbert (complexe, séparable) \mathcal{H} qui vérifie les propriétés suivantes :

1. ρ est auto-adjoint
2. ρ est positif
3. $\text{Tr } \rho = 1$

Remarquons le cas particulier où ρ est un projecteur (forcément orthogonal comme il est auto-adjoint). On dit alors que ρ est un *état pur*. On peut dans ce cas le voir comme un élément de \mathcal{H} , celui sur lequel ρ projette. Cela correspond au cas où notre état était un point bien défini de l'espace des phases dans le cas classiques. Si ρ n'est pas un projecteur, on peut le voir comme un mélange statistique de projecteurs, donc d'états purs ; c'est-à-dire que notre état est un choix aléatoire entre ces états purs. Cela correspond au cas où notre état était décrit par un nuage de points dans le cas classique. On notera \mathcal{S} l'ensemble des opérateurs densité.

La mesure la plus générale que l'on peut effectuer sur le système peut alors se décrire ainsi :

Définition 1.2. : Mesure

Une mesure M d'un système quantique, à valeurs dans un espace mesuré $(\mathcal{X}, \mathcal{A})$ est spécifiée par une mesure de probabilité à valeurs dans les opérateurs (*OPROM* pour l'acronyme anglais), c'est-à-dire une collection d'opérateurs sur \mathcal{H} , auto-adjoints, $M(A) : A \in \mathcal{A}$, telle que :

1. $M(\mathcal{X}) = \mathbf{1}$ l'opérateur identité
2. Chaque $M(A)$ est positif
3. Pour des A_i disjoints en nombre dénombrable, $M(\bigcup_i A_i) = \sum_i M(A_i)$

On remarquera que ce sont juste les axiomes ordinaires des mesures de probabilité sur $(\mathcal{X}, \mathcal{A})$, à ceci près que la mesure prend ses valeurs dans des opérateurs auto-adjoints positifs au lieu de les prendre dans les nombres réels.

Dans la suite, nous supposons de plus que cette probabilité est dominée par une mesure σ -finie ν sur $(\mathcal{X}, \mathcal{A})$. Nous pouvons alors écrire

$$M(A) = \int_A m(x)\nu(dx)$$

Il nous reste à savoir quel résultat on obtient quand on applique la mesure M à ρ . C'est une variable aléatoire X à valeurs dans \mathcal{X} et dont la loi est donnée par la *règle de trace*.

Définition 1.3. : Règle de trace

La distribution de probabilité du résultat X est

$$\mathbb{P}[X \in A] = \text{Tr}(\rho M(A)), \quad A \in \mathcal{A} \tag{1}$$

On remarquera que (1) génère une vraie probabilité sur $(\mathcal{X}, \mathcal{A})$.

D'autre part, si M est dominée par ν , la probabilité générée est $p d\nu$ avec

$$p(x) = \text{Tr}(\rho m(x))$$

Les principes de la mécanique quantique assurent que ce sont là les seules mesures possibles de ρ , et qu'il est en théorie possible de construire un dispositif exécutant n'importe laquelle de

ces mesures. Il est notamment impossible, généralement, de combiner l'information provenant de deux mesures M et M' . En effet la mesure $(M \times M')$ qui en résulterait devrait vérifier, pour tout ρ, A et A' :

$$\text{Tr}(\rho(M \times M')(A \times A')) = \text{Tr}(\rho M(A)) \text{Tr}(\rho M'(A'))$$

ce qui n'est possible que si $M(A)$ et $M'(A')$ commutent.

Physiquement cela correspond au fait que mesurer change l'état du système et, si la mesure est assez riche, l'état du système après mesure ne dépend que du résultat de la mesure (et pas de l'état initial).

2 Comment remonter à ρ à partir du résultat de la mesure

La mesure M permet d'associer à ρ une mesure de probabilité p_ρ sur $(\mathcal{X}, \mathcal{A})$. La mesure elle-même correspondra à tirer une variable aléatoire selon p_ρ .

Comme p_ρ est linéaire en ρ , on ne peut clairement pas espérer remonter à ρ à partir d'un seul échantillon. On va donc considérer que nous avons n systèmes préparés dans le même état ρ . Effectuer la mesure M sur chacun de ces échantillons va permettre d'estimer empiriquement p_ρ , à partir de laquelle on peut, du moins pour certaines mesures, remonter à ρ . Une fois le problème posé en ces termes, on peut commencer à utiliser des méthodes traditionnelles de statistique. Ce sera l'objet du reste de ce document.

Une remarque avant d'aller plus loin : si nous considérons que nous avons n exemplaires du même état quantique, d'une part, on n'est pas forcé d'effectuer sur chacun d'eux la même mesure, et d'autre part, surtout, si on les a à disposition simultanément, on peut les considérer tous ensemble comme un même système. L'état de ce système sera alors $\rho^{\otimes n}$ un opérateur sur l'espace de Hilbert $\mathcal{H}^{\otimes n}$, soit \mathcal{H} produit tensoriel avec \mathcal{H} n fois. Les mesures dont on dispose d'après la définition 1.2 contiennent notamment les mesures produit, c'est-à-dire le fait de mesurer chacun de nos échantillons individuellement, mais ne se limitent pas à celles-ci. Il est possible de mesurer «en même temps» nos n objets, et, dans certains cas, on peut en tirer plus de renseignements que ne peut en offrir une mesure produit, quelle qu'elle soit.

Dans la suite, nous ignorerons cette remarque et nous nous concentrerons sur le cas où nous effectuons toujours la même mesure.

2.1 Deux types d'estimateur classiques

2.1.1 Estimateurs par projection

Nous allons décrire l'une des manières les plus usitées de retrouver une distribution de probabilité que l'on a échantillonnée. On attend de celle-ci qu'elle soit un élément d'un espace L^2 . Soit donc s une probabilité sur un espace \mathcal{X} , à densité par rapport à une mesure ν sur \mathcal{X} , de telle sorte que $s \in L^2(\mathcal{X}, \nu)$. On tire n variables aléatoires X_i suivant la loi s . On obtient donc une *probabilité empirique* sur \mathcal{X} , à savoir $\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$. On ne peut pas prendre, si \mathcal{X} est non dénombrable, \mathbb{P}_n comme un estimateur de s . Alors on se limite à un *modèle*, un sous-espace vectoriel V de $L^2(\mathcal{X}, \nu)$. L'estimateur naturel associé à ce modèle est la projection sur V de \mathbb{P}_n . Bien sûr, $\mathbb{P}_n \notin L^2(\mathcal{X}, \nu)$, mais comme on regarde un nombre fini de coordonnées, sa projection sur V a un sens.

Explicitons légèrement les choses. On peut écrire $s = f d\nu$. De même les vecteurs d'une base orthonormée de V s'écrivent $v_j = g_j d\nu$. Notre estimateur a pour coefficient dans la direction v_j le «produit scalaire» de la probabilité empirique et de v_j , soit $\sum_i g_j(X_i)$. L'espérance de $g_j(X)$ est alors $\int_{\mathcal{X}} g_j(x) f(x) d\nu$, qui est bien la coordonnée de s dans la direction v_j . Par la loi des

grands nombres, si on a choisi de manière raisonnable nos vecteurs, notre estimateur va donc tendre vers la projection de s sur V . Nous reviendrons par après sur le problème d'approcher s «dans son entier». Voyons d'abord comment appliquer ce genre de méthodes à ρ .

Pour commencer il nous faut remarquer que ce n'est pas ρ que l'on échantillonne (ce qui n'a d'ailleurs pas de sens), mais p_ρ . Le problème est de munir \mathcal{X} d'un produit scalaire adapté (en fait nous ne prendrons pas l'espace de la forme $L^2(\nu)$).

De par sa définition, ρ est un élément d'un espace vectoriel (quoique l'ensemble des ρ possibles soit strictement inclus dans cet espace). On peut le munir d'un produit scalaire et d'une norme associée, la norme L^2 d'opérateurs. Pour un opérateur auto-adjoint τ ,

$$\|\tau\| = \sum_i \lambda_i(\tau)^2$$

où les λ_i sont les valeurs propres de τ (qui sont réelles). En fait, du moment que l'on exprime l'opérateur dans une base orthonormée (de \mathcal{H} l'espace sous-jacent), sa norme L^2 vaut la somme des carrés du module de ses entrées matricielles.

Nous allons donc essayer de munir l'espace des fonctions sur \mathcal{X} du produit scalaire dérivant de la norme L^2 d'opérateurs. Les modèles utilisés seront logiquement les espaces vectoriels engendrés par un ensemble d'éléments matriciaux dans une base qui nous intéresse. Pour mener à bien ce programme, il nous faut légèrement expliciter comment on passe de ρ à p_ρ . Soit une base (e_j) de \mathcal{H} . Les entrées matricielles des opérateurs sur \mathcal{H} dans cette base seront notées $\rho_{j,k}$, $M_{j,k}(A)$, etc... Alors

$$\begin{aligned} p_\rho(x) &= \text{Tr}(\rho m(x)) \\ &= \sum_{j,k} \rho_{j,k} m_{k,j}(x) \end{aligned}$$

Si les $m_{k,j}$ sont des fonctions assez gentilles, alors il existe une base bi-orthogonale, c'est-à-dire des $\mu_{k,j}$ vérifiant

$$\int_{\mathcal{X}} m_{k,j}(x) \mu_{l,m}(x) \nu(dx) = \delta_{k,l} \delta_{j,m}$$

Donc la norme d'opérateurs de ρ peut se récrire :

$$\sum_{j,k} |\rho_{j,k}|^2 = \sum_{j,k} \left| \int_{\mathcal{X}} p_\rho(x) \mu_{k,j}(x) \nu(dx) \right|^2$$

Comme l'entrée $\rho_{j,k}$ est la moyenne de $\mu_{k,j}$ par rapport à p_ρ , il est logique de l'estimer par la moyenne par rapport à la probabilité empirique : $\widehat{\rho}_{j,k} = \sum_{i=1}^n \mu_{k,j}(X_i)$

et nous avons retrouvé nos projections de la probabilité empirique sur un sous-espace vectoriel (notons d'ailleurs que l'estimateur n'est pas forcément, à l'arrivée, une matrice densité).

Là encore, un modèle consistera à choisir un nombre fini d'entrées et à mettre les autres à zéro. Bien sûr cela introduit un biais, mais si on choisissait toutes les entrées, la variance L^2 de l'estimateur serait infinie. Nous reviendrons dans la sous-section 2.2 sur la manière de traiter ce problème.

D'autre part, si les fonctions $m_{j,k}$ sont mal éduquées, il peut ne pas exister de $\mu_{j,k}$ bornées, et il nous faut être plus subtil. On a un exemple de ce cas dans la dernière partie de mon rapport de DEA.

2.1.2 Estimateur du maximum de vraisemblance

Un autre estimateur extrêmement classique est l'estimateur du maximum de vraisemblance. Cet estimateur a plusieurs grandes qualités. Elles découlent de ce qu'il estime d'un coup l'objet mathématique, sans le fractionner en petits éléments (coordonnée selon un vecteur de base, par exemple) à évaluer indépendamment. On conserve donc l'information sur la structure de l'ensemble \mathcal{S} des opérateurs densité possibles. De ce fait, nous sommes aussi assurés que l'estimateur est un opérateur densité dans le sens de la définition 1.1.

L'idée de cet estimateur est la suivante : supposons ρ connu. Alors la probabilité que notre n-échantillon soit (x_1, x_2, \dots, x_n) est de densité

$$\prod_{i=1}^n p_\rho(x_i)$$

Cette fonction de $\mathcal{X}^n \rightarrow \mathbb{R}$ s'appelle la *vraisemblance* de ρ .

On peut légitimement s'attendre à ce que le ρ dont la vraisemblance est la plus forte en (x_1, x_2, \dots, x_n) soit un bon estimateur de notre état réel. Toutefois dans un certain nombre de cas, \mathcal{S} est de dimension infinie, et on ne peut donc pas optimiser sur tout l'espace.

On prend donc aussi des modèles V , des sous-variétés de \mathcal{S} (par exemple l'intersection de \mathcal{S} et de l'espace vectoriel engendré par un certain nombre d'entrées matricielles dans une base donnée), et c'est sur ces modèles que l'on maximise la vraisemblance. En pratique, comme le produit est peu agréable à manier, on prendra plutôt le logarithme de celui-ci, la *log-vraisemblance*, ce qui revient exactement au même. Notre estimateur dans V sera donc :

$$\widehat{\rho}^V = \arg \inf_{\tau \in V} \sum_{i=1}^n -\ln(p_\tau(x_i)) \quad (2)$$

Là encore, on crée un biais en se limitant à V mais on limite du même coup la variance.

Nous allons voir maintenant comment choisir V , que ce soit dans le cas des estimateurs par projection ou dans le cas de l'estimateur du maximum de vraisemblance.

2.2 Pénalisation

Les deux types d'estimateurs que nous avons vu sont contraints, dans le cas de la dimension infinie, de se restreindre à des modèles afin de garder des variances finies, mais au prix de l'introduction d'un biais. La variance (erreur d'estimation), qui décroît typiquement en $\frac{1}{\sqrt{n}}$, augmente avec la taille du modèle, tandis que le biais (erreur de modélisation) diminue. Il s'agit donc de trouver un modèle minimisant la somme des deux erreurs.

Pour cela, nous allons pour commencer mettre nos estimateurs sous la forme d'estimateurs du *minimum de contraste*. Un tel estimateur est construit en choisissant une *fonction de contraste empirique* γ_n . Pour pouvoir comparer les modèles, il ne faut pas que cette fonction dépende du modèle. Ici γ_n sera de la forme, pour τ un opérateur auto-adjoint :

$$\gamma_n(\tau) = \mathbb{P}_n[f_\tau] = \frac{1}{n} \sum_{i=1}^n f_\tau(x_i)$$

où $f_\tau : \mathcal{X} \rightarrow \mathbb{R}$ est une fonction qui dépend de τ . L'estimateur au sein d'un modèle V est alors :

$$\widehat{\rho}^V = \arg \inf_{\tau \in V} \gamma_n(\tau)$$

Comment savoir si un tel estimateur est raisonnable? Presque sûrement, (si les f_τ sont suffisamment lisses)

$$\gamma_n(\tau) \rightarrow_{n \rightarrow \infty} \gamma(\tau) = \mathbb{P}_\rho [f_\tau] = \int_{\mathcal{X}} f_\tau(x) p_\rho(x) \nu(dx)$$

Une exigence évidente est alors

$$\arg \inf_{\tau \in \mathcal{S}} \gamma(\tau) = \rho \quad (3)$$

L'estimateur de maximum de vraisemblance a déjà été mis sous la forme voulue, en (2), avec $f_\tau = -\ln(p_\tau(\cdot))$. De plus (3) est bien vérifiée :

$$\int_{\mathcal{X}} -\ln(p_\tau(x)) p_\rho(x) \nu(dx) = K(p_\tau, p_\rho) - \int_{\mathcal{X}} \ln(p_\rho(x)) p_\rho(x) \nu(dx)$$

où $K(p_\tau, p_\rho)$ est la distance de Kullback-Leibler, ou entropie relative, entre p_τ et p_ρ . Ce terme est donc strictement positif, et vaut zéro si $\tau = \rho$. D'autre part le second terme est indépendant de τ . Donc le minimum de γ est bien atteint en ρ . En fait l'estimateur du maximum de vraisemblance est l'estimateur qui minimise la distance de Kullback-Leibler entre l'estimateur et la probabilité empirique.

Il nous faut travailler à peine plus pour l'estimateur par projection. Prenons une base (e_j) sur \mathcal{H} . Alors convient

$$f_\tau(x) = \|\tau\|^2 - 2\mathcal{R} \left(\sum_{j,k} \mu_{k,j}^*(x) \tau_{j,k} \right)$$

où la norme est la norme L^2 d'opérateurs et l'étoile la conjugaison complexe. En effet la fonction de contraste empirique peut se récrire comme une somme de termes indexés par (j, k) :

$$\mathcal{R}(\tau_{j,k})(\mathcal{R}(\tau_{j,k}) - 2\frac{1}{n} \sum_{i=1}^n \mathcal{R}(\mu_{k,j}(x_i))) + \mathcal{I}(\tau_{j,k})(\mathcal{I}(\tau_{j,k}) - 2\frac{1}{n} \sum_{i=1}^n \mathcal{I}(\mu_{k,j}(x_i)))$$

Le minimum de ce terme est bien atteint en $\tau_{j,k} = \frac{1}{n} \sum_{i=1}^n \mu_{k,j}(x_i)$. On remarquera que $\gamma(\tau) = \|\rho - \tau\|^2 - \|\rho\|^2$. Intuitivement, l'estimateur par projection minimise la distance L^2 de l'estimateur à la probabilité empirique.

Pour les deux estimateurs, l'idée est maintenant de pénaliser la fonction de contraste en fonction du modèle dont fait partie τ . La pénalisation sera d'autant plus grande que le modèle est grand : elle reflète le fait qu'on s'attend à une estimation plus optimiste de la fonction de contraste empirique pour l'estimateur quand le modèle est grand. En formules pen : $\mathcal{M} \rightarrow \mathbb{R}$, où \mathcal{M} est l'ensemble des modèles. Le modèle que nous choisissons est :

$$\widehat{V} = \arg \inf_{V \in \mathcal{M}} \gamma_n(\widehat{\rho}^V) + \text{pen}(V)$$

Notre estimateur final est alors $\widehat{\rho}^{\widehat{V}}$.

Mais cela ne rime pas à grand chose si on ne sait pas comment choisir la pénalité. Pour tout $\tau \in V$,

$$\gamma_n(\widehat{\rho}^{\widehat{V}}) - \gamma_n(\tau) \leq \gamma(\widehat{\rho}^{\widehat{V}}) - \gamma(\tau) + (\gamma - \gamma_n)(\tau) + \text{pen}(V) - (\gamma - \gamma_n)(\widehat{\rho}^{\widehat{V}}) - \text{pen}(\widehat{V})$$

Nous voulons donc voir $\text{pen}(V)$ dominer $(\gamma - \gamma_n)(\tau)$, auquel cas

$$\gamma(\widehat{\rho}^{\widehat{V}}) \leq C(\gamma(\tau) + \text{pen}(V)) \quad (4)$$

Essentiellement, l'idée est de prendre

$$\text{pen}(V) = \mathbb{E} \left[\sup_{\tau \in V} |\gamma - \gamma_n|(\tau) \right] \quad (5)$$

Bien sûr, comme nous voulons que la pénalité domine l'erreur d'estimation pour tout V , il va y avoir des termes supplémentaires à la fois dans la pénalité et dans (4).

Je ne vais pas détailler beaucoup plus, comme c'est le thème principal, dans un cas particulier, de mon mémoire de DEA. Voici juste les idées directrices dans chacun des deux modèles :

Dans le cas de l'estimateur par projection, la pénalité (5) n'existe même pas, comme le sup entre parenthèses est infini : en effet, il est linéaire en τ . Mais cela implique aussi qu'on peut le récrire

$$\|\tau\| \sup_{\substack{\tau \in V \\ \|\tau\|=1}} (\gamma - \gamma_n)(\tau)$$

Si nous récrivons $\chi_n(V) = \sup_{\substack{\tau \in V \\ \|\tau\|=1}} (\gamma - \gamma_n)(\tau)$, quelques calculs permettent alors de trouver, pour tout ϵ

$$\frac{\epsilon}{2 + \epsilon} \left\| \rho - \widehat{\rho^{\hat{V}}} \right\|^2 \leq \left(1 + \frac{2}{\epsilon} \right) \|\rho - \tau\|^2 + (1 + \epsilon)(\chi_n^2(\hat{V}) + \chi_n^2(V)) - \text{pen}_n(\hat{V}) + \text{pen}_n(V)$$

Il ne nous reste plus qu'à dominer le chi-deux, qui est un supremum de processus empirique, pour lesquels il existe des outils puissants (inégalité de Talagrand).

Dans le cas de l'estimateur du maximum de vraisemblance, on peut appliquer plus directement la forme (5). On doit cependant, pour garder des fonctions bornées, travailler non pas avec $\gamma_n(\tau)$, mais avec $\gamma_n(\frac{1}{2}(\tau + \rho))$. À partir de ce moment on ne peut plus espérer dominer la distance de Kullback-Leibler, mais on domine une autre distance très proche, appelée distance de Hellinger. D'autre part la pénalité va encore consister à dominer un sup de processus empirique, mais les outils idoines sont différents de ceux de l'inégalité de Talagrand : on va contrôler la taille du modèle avec ce que l'on appelle l'entropie avec crochets.

Des théorèmes très généraux (et directement utilisables) peuvent être trouvés dans [4].

3 Choix de la mesure à effectuer

Nous n'allons qu'esquisser certaines idées dans ce sens.

Nous nous placerons dans le cas où ρ dépend d'un paramètre θ de dimension finie d . C'est notamment le cas, en physique, dans les systèmes de spin.

Nous pouvons alors effectuer l'estimation du maximum de vraisemblance sur l'ensemble de l'espace des paramètres. Les propriétés asymptotiques de ce type d'estimateur sont très bien connues (voir par exemple [5]). En effet, sous un certain nombre d'hypothèses (notamment tous les résultats de la mesure doivent pouvoir être atteints à partir de tout paramètre, et les dérivées secondes par rapport à θ de la log-vraisemblance existent) qui seront généralement vérifiées dans les cas qui nous intéressent, on sait que notre estimateur est asymptotiquement normal, de moyenne le véritable opérateur densité et de variance l'inverse de l'information de Fisher. Explicitons.

J'appelle le score en θ la dérivée de la log-vraisemblance en θ à savoir :

$$l(\theta, M) = \frac{\partial}{\partial \theta} \ln(p_\theta)$$

où nous avons abrégé p_{ρ_θ} en p_θ . C'est donc une fonction de $x \in \mathcal{X}$.

L'information de Fisher $I(\theta, M)$ est alors

$$I(\theta, M) = \mathbb{E}_\theta [l(\theta, M)l(\theta, M)^t]$$

Et on observe alors que notre estimateur $\hat{\theta}$ converge en loi, ou, pour être plus exact :

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow \mathcal{N}\left(0, \frac{1}{I(\theta, M)}\right)$$

Par ailleurs, le théorème de Cramér-Rao affirme, sous le même genre de conditions que celles qui impliquent l'existence de cette limite pour l'estimateur du maximum de vraisemblance, que la variance de tout estimateur non biaisé est supérieure à $I(\theta, M)^{-1}$. En ce sens, l'estimateur du maximum de vraisemblance est asymptotiquement optimal.

Si on applique systématiquement l'estimateur du maximum de vraisemblance, la question devient : quelle mesure M effectuer pour avoir l'information de Fisher la plus grande possible ?

Pour poser mathématiquement la question, il nous faut d'abord décider d'un coût, sanctionnant les erreurs d'estimation de θ dans les différentes directions (comme θ n'est en général pas de dimension 1). Soit alors G une matrice $d \times d$. Certains choix peuvent être plus naturels que d'autres (celui qui minimise la distance dite de Bures, par exemple). Nous cherchons M telle que $\text{Tr}(I(\theta, M)^{-1}G)$ soit minimal.

Remarquons que $I(\theta, M)$ dépend de θ . Asymptotiquement, cela n'a pas grande importance : on peut employer mettons $1/\sqrt{n}$ échantillons pour avoir une idée grossière du θ , puis choisir la M optimale pour l'estimateur $\hat{\theta}$ auquel nous sommes arrivés.

On peut déjà trouver une borne sur l'efficacité maximale d'une mesure. En effet, on peut montrer que l'information de Fisher d'une mesure est toujours inférieure, au sens des opérateurs, à l'information de Fisher quantique du modèle, que nous allons définir.

On commence par définir les matrices de score quantique. Si $\theta = (\theta_1, \theta_2, \dots, \theta_d)$, ce sont les matrices auto-adjointes λ_i pour i entre 1 et d solutions des équations implicites :

$$\frac{\partial \rho(\theta)}{\partial \theta_i} = \frac{1}{2}(\lambda_i \rho + \rho \lambda_i)$$

Là encore, les λ_i dépendent de θ .

La matrice d'information quantique a alors pour entrées :

$$(I_Q)_{i,i'} = \frac{1}{2} \text{Tr}(\rho(\lambda_i \lambda_{i'} + \lambda_{i'} \lambda_i))$$

Quand on regarde $(\lambda_i)_{1 \leq i \leq d}$ comme un vecteur à d dimensions, l'analogie avec l'information de Fisher classique devient claire.

Pour tout M ,

$$I(\theta, M) \leq I_Q(\theta)$$

Cette borne, sorte de borne de Cramér-Rao quantique, n'est souvent pas atteinte. Nous restons donc avec les questions suivantes :

1. La question qui fait titre de cette section...
2. Plus simplement, la mesure optimale peut-elle toujours être prise à valeurs dans \mathcal{X} fini ? Cela permettrait d'envisager des méthodes de recherche de la «meilleure» mesure au cas par cas.
3. Autres questions

Quelques remarques supplémentaires sur ce qui se passe quand on ne s'impose plus d'utiliser toujours la même mesure, ou même de mesurer chaque ρ individuellement (cf le début de la section 2).

Pour toute mesure produit, c'est-à-dire chaque fois que l'on mesure les ρ un à un, l'information de Fisher de la mesure de $\rho^{\otimes n}$ vaut $\sum_{i=1}^n I(\theta, M_i)$. Il n'y a donc pas d'espoir de faire mieux qu'avec une même mesure répétée par pareille méthode.

En revanche, si on s'autorise les mesures jointes (ie mesurer d'un coup $\rho^{\otimes n}$), on peut gagner. Dans le cas très simple d'un spin (θ de dimension (réelle) trois), on sait que les meilleures mesures «individuelles» vérifient $I(\theta, M) \leq \frac{1}{3}I_Q(\theta)$ tandis qu'il y a des mesures jointes sur deux spins qui atteignent $I(\theta, M) = \frac{1}{2}I_Q(\theta)$.

Pour conclure, disons que les statistiques quantiques proposent déjà à l'heure actuelle un certain nombre de problèmes non résolus, d'un certain intérêt mathématique comme, le secteur étant jeune, il est encore relativement inexploré, et d'un intérêt pratique certain (connaître la meilleure mesure à effectuer peut contribuer à construire de bons ordinateurs quantiques par exemple...).

Références

- [1] Ole E. Barndorff-Nielsen, Richard D. Gill, and Peter E. Jupp. On quantum statistical inference. *J.R.Statis.Soc.B*, 65(4) :775–816, 2003.
- [2] C.W. Helstrom. *Quantum Detection and Information Theory*. Academic Press, New York, 1976.
- [3] A.S. Holevo. *Probabilist and Statistical Aspects of Quantum Theory*. North-Holland, Amsterdam, 1982.
- [4] P. Massart. Concentration inequalities and model selection. Cours de Saint-Flour de juillet 2003, disponible sur la page web de Pascal Massart.
- [5] A.W. van der Waart. *Asymptotic Statistics*. Cambridge University Press, 1998.