PROBABILISONS UN PEU....

par *Henri ROUANET* *

En vue de cette rencontre d'Urbino autour de G.-Th. Guilbaud, j'avais d'abord songé à aller à la recherche du temps perdu, à raconter ce que furent pour moi la révélation des mathématiques discrètes à l'IHP en 1958, le grand espoir de la revue "Mathématiques et Sciences Humaines" dans les années soixante, etc. J'ai jugé plus tonique d'essayer de dire ce que le séminaire de Guilbaud m'apporte à l'heure actuelle, en évoquant deux orientations de notre groupe de recherche pour lesquelles l'influence de Guilbaud se trouve déterminante.

Première orientation : vers une conception discrète de la statistique

Pour la communauté statisticienne, les méthodes "discrètes" constituent une *spécialité*, volontiers laissée à ceux qui se complaisent dans les complications techniques. Lorsque vers 1981, avec la thèse de 3e cycle de J.-M. Bernard, nous nous sommes embarqués dans le développement des procédures bayésiennes pour les fréquences, les "gens avertis" nous ont aimablement prévenus : "the discrete case is messy". Or c'était précisément l'époque où Guilbaud, dans son séminaire, nous dévoilait la magnifique simplicité des distributions de type hypergéométrique. Ce séminaire nous a donné le courage et les moyens de poursuivre la mise en oeuvre d'une *conception discrète de l'inférence statistique.*

A l'intention de ceux qui sont habitués à l'inférence statistique traditionnelle, on peut facilement présenter cette conception en partant du *schéma d'urne* familier (les urnes mènent à tout, à condition d'en sortir!).

(1) Soit donc une urne de composition A + B + C (trois catégories suffiront pour suggérer la généralité), dont on extrait (au hasard sans remise) un échantillon ; la probabilité que cet échantillon soit de composition a + b + c est donnée par la *distribution hypergéométrique directe* (habituelle) :

$$P(a,b,c \mid A,B,C) = \frac{\binom{A}{a}\binom{B}{b}\binom{C}{c}}{\binom{A+B+C}{a+b+c}}$$

* Groupe Mathématique et Psychologie, CNRS, Université Paris V.

(2) Passons maintenant à l'inférence bayésienne, en posant (pour commencer), sur la composition de l'urne, une distribution initiale uniforme. La probabilité (inverse), conditionnellement à l'observation d'un échantillon de composition a + b + c, que l'urne soit de composition A + B + C, est donnée par une distribution de forme analogue, que nous appellerons *distribution hypergéométrique inverse* :

$$P(A,B,C \mid a,b,c) = \frac{\binom{A}{a}\binom{B}{b}\binom{C}{c}}{\binom{A+B+C+2}{a+b+c+2}}$$

(3) On peut mettre cette dernière distribution sous la forme suivante, qui donne la probabilité (prédictive) d'un échantillon à venir de composition I + J + K (en posant I = A -a, etc) :

$$P(I,J,K \mid a,b,c) = \frac{\binom{I+a}{a}\binom{J+b}{b}\binom{K+c}{c}}{\binom{I+J+K+a+b+c+2}{a+b+c+2}}$$

(4) La distribution prédictive précédente est un cas particulier de la distribution suivante (que les spécialistes ont baptisée "Dirichlet-multinomiale"), de paramètres α, β, γ (ci-dessus : $\alpha = a+1$, etc.).

$$P(I,J,K \mid \alpha, \beta, \gamma) = \frac{\binom{I+\alpha-1}{\alpha-1}\binom{J+\beta-1}{\beta-1}\binom{K+\gamma-1}{\gamma-1}}{\binom{I+J+K+\alpha+\beta+\gamma-1}{\alpha+\beta+\gamma-1}}$$

La distribution D-M possède la propriété d'être conjugée de la distribution hypergéométrique directe : si l'on prend une distribution prédictive initiale de forme D-M, et de paramètres α, β, γ, on obtient, après observation d'un échantillon de composition a + b + c, une distribution prédictive finale de forme D-M, où les paramètres α, β, γ, ont été remplacés par α + a, etc. (le cas particulier de la distribution initiale uniforme correspond donc à $\alpha = \beta = \gamma = 1$, d'où après observation de l'échantillon $\alpha = a + 1$, etc.).

Dans la ligne de ce formalisme, on peut alors poser le problème de l'inférence dans les termes suivants : sur un échantillon de taille n, on a observé une propriété $\mathcal{P}$; le but de l'inférence est tout simplement d'évaluer la probabilité (prédictive) qu'un échantillon à venir, de taille N, possède ladite propriété $\mathcal{P}$. Comme on le voit, cette conception de l'inférence est discrète à tous égards.

Cette conception n'empêchera pas d'utiliser les instruments de l'analyse, à un niveau technique, pour obtenir des solutions approchées - même si en fait, compte tenu des actuels moyens de calcul numérique, on peut penser qu'on ira très loin en restant dans ce cadre discret.

A cet égard, nous comptons beaucoup sur la *formule magique de Guilbaud,* dont nous écrirons d'abord la forme élémentaire :

$$\forall\ a,n,A,N\ :\ \sum_{I=0}^{A} \binom{I}{a}\binom{N-I}{n-a} = \sum_{i=a}^{n} \binom{A+1}{i+1}\binom{N-A}{n-a}$$

puis sa généralisation à un cléroèdre à 3 catégories (le mot χληρος, nous rappelle Guilbaud, évoque à la fois tirage au sort, donc probabilisation, et part d'héritage, donc répartition).

$$\forall\ a,b,c,A,B,C\ :\ \sum_{\substack{I\leq A\\ J\leq B}} \binom{I}{a}\binom{J}{b}\binom{K}{c} = \sum_{\substack{i\geq a\\ j\geq b}} \binom{A+1}{i+1}\binom{B+1}{j+1}\binom{C}{k}$$

Pour nous cette identité est "de Guilbaud", parce qu'il nous a appris son existence (elle ne figure pas parmi les 26 identités binomiales recensées par Knuth) ; elle est "magique", parce qu'elle préside à cette opération magique entre toutes qu'est le renversement des probabilités. Techniquement, l'identité magique devrait nous permettre, en choisissant des propriétés $\mathcal{P}'$ convenablement définies, de travailler sur de petits cléroèdres (de taille n), plutôt que sur les grands cléroèdres (de taille N). Et sur l'art de découper les cléroèdres, nous comptons bien aller consulter une fois de plus Guilbaud, pas plus tard que le plus tôt possible...

Deuxième orientation : probabilisons !

Qu'est-ce que "probabiliser", (le mot ne figure pas dans le TGL), et qui "probabilise"? Pour les statisticiens fréquentistes, la probabilité est dans les choses plutôt que dans la tête, et la statistique a pour besogne de découvrir le "bon modèle" plutôt que de peser l'incertitude ; on ne voit guère les statisticiens de cette idéologie se servir de cette "liberté de probabiliser" ardemment revendiquée par Guilbaud. Pour nous au contraire, lorsqu'on "probabilise", c'est aux fins d'évaluer la probabilité de ce qui nous intéresse à partir de ce que nous savons, et la liberté de probabiliser revêt une signification tout à fait tangible. A titre d'exemple, j'évoquerai un thème de méditation que nous proposons à nos étudiants de première année.

Le paralogisme de Larousse : dans son *Grand Dictionnaire Universel*, à l'article *Prisons*, P. Larousse intervenait dans le débat en cours il y a un siècle sur l'instauration de l'instruction publique (selon les adversaires du projet, celle-ci risquait de provoquer une augmentation de la criminalité) ; s'appuyant sur les statistiques indiquant, parmi les criminels incarcérés, une proportion énorme (82%) d'illettrés ou quasi-illettrés, P.Larousse déclare ce qui suit : "Ces chiffres sont *trop éloquents par eux-mêmes* pour avoir besoin d'un long commentaire ; ils *prouvent* que ceux qui sont hostiles à l'instruction publique favorisent le crime et que l'instruction, quoi qu'on en dise, moralise le peuple".

Le "paralogisme de Larousse", c'est clair, consiste à prétendre conclure sans tenir compte de la proportion des illettrés dans la population. Paralogisme des plus communs, dont on connaît les variantes contemporaines : "un délinquant sur quatre est un immigré, donc..." (les journaux). Les situations dans lesquelles on peut s'attendre à le retrouver relèvent d'un paradigme que nous proposons de dénommer (séminaire de sémiotique oblige) *paradigme du signe et de la chose.*

Le paradigme du signe et de la chose : soit une population d'individus qu'on peut classer selon les caractères j (le signe : chez Larousse, "illettré"), et k (la chose : chez Larousse, "criminel") :

. *on s'intéresse* aux proportions de k si j, p^jk, et de k si j' (caractère complémentaire de j), $p^{j'}k$ (dans ce qui suit nous supposerons plus particulièrement qu'on s'intéresse au rapport $p^jk / p^{j'}k$) ;

. *on connaît* la proportion de j si k, p^kj (chez Larousse, 82 %) ;

. *on ne connaît pas* la proportion marginale de j, p_j.

Il nous semble que ce paradigme présente un certain intérêt en sciences humaines, dans le domaine historique notamment, où la "répétition" chère à l'idéologie fréquentiste est absente. La probabilisation, en l'occurrence, ne peut qu'exprimer l'incertitude de qui "probabilise". La solution technique du paradigme, dans le cadre bayésien, est élémentaire. En effet, les diverses proportions en jeu sont liées par la relation :

$$\frac{p^jk}{p^{j'}k} = \frac{1 - p_j}{p_j} \times \frac{p^kj}{1-p^kj}$$

Puisque on connaît p^kj, si on s'intéresse au rapport $p^jk/p^{j'}k$ il suffit de probabiliser p_j : liberté de probabiliser! (soit dit en passant, j'aimerais savoir quelle solution les statisticiens fréquentistes pourraient proposer pour ce paradigme).

Guilbaud et la maîtrise de l'à-peu-près

Au terme de ses "Leçons d'à-peu-près" - entendons par là "leçons sur la *maîtrise* de l'à-peu-près" - Guilbaud nous invite à accéder au niveau supérieur de cette maîtrise, le *niveau probabiliste*. Le comble de la maîtrise de l'à-peu-près, ne le trouve-t-on pas dans ces admirables énoncés (dits "lois de grands nombres"), où la notion d'à-peu-près, dédoublée en les deux notions de précision et de certitude, se trouve doublement surmontée :

Quels que soient ε et α aussi petits que l'on voudra : il existe n_0 tel que $n > n_0 \Rightarrow \text{Prob (Ecart} < \varepsilon) > 1 - \alpha$.

Le séminaire de Guilbaud de ces dernières années - on ne le sait sans doute pas assez - a été consacré à ce niveau supérieur de la maîtrise de l'à-peu-près. Voici quelques-unes des têtes de chapitre de ce séminaire : de Finetti et l'échangeabilité ; Poincaré et les fonctions arbitraires ; Laplace et les marées de Brest ; J.Bernoulli et la loi forte (eh oui!) des grands nombres; sans oublier naturellement Buffon et la rumeur de Carcelle-le-Grignon...

Qu'il me soit permis, en guise de conclusion, d'appeler de mes voeux un tome 2 de "Leçons d'à-peu-près" qui reprendrait les principaux thèmes de ce séminaire. Je sais bien qu'on ne peut attendre, de la part d'une communauté statisticienne obnubilée par les illusions réductionnistes, qu'une compréhension approximative ; mais qu'importe ? Déjà, au terme de ses premières "leçons", Guilbaud nous mettait en garde contre un optimisme excessif : "Quand j'entends confrondre la convergence usuelle avec diverses lois de grands nombres ; quand j'entends oser dire "la probabilité *est* la limite d'une fréquence"... je pense que les mentalités évoluent lentement...".

Oserai-je une conclusion encore moins optimiste ? J'ai récemment lu -je vous le montrerai imprimé- la sentence : "la fréquence *est* une probabilité". Voilà bien la solution finale d'un problème qui avait tant occupé les plus grands esprits ! De tels phénomènes participent peut-être d'un déclin général de l'esprit scientifique, et sans doute vaut-il mieux en prendre son parti. Renonçant aux espérances vaines, suivons plutôt l'exemple des sages du Premier Cercle et comme eux vivons un désir sans espoir : **Che senza speme vivemo in disio.**