Abrégé

des probabilité et statistique

Marc Bourdeau[1]
École Polytechnique de Montréal

Révisé le 1er mai 2009

[1]Marc.Bourdeau@polymtl.ca

Table des matières

Avant-propos

L'appréhension, je l'ai lente et embrouillée.
Montaigne [1533 – 1592][1]

La probabilité est de l'ordre de la modélisation théorique. Les phénomènes aléatoires, c'est à dire ceux dont la mesure, le résultat, est incertain, ont parfois des modèles mathématiques ou probabilistes destinés à en dire néanmoins quelque chose de précis, nonobstant le hasard en action. Le hasard a longtemps été considéré en fait comme relevant de l'ordre du “divin”, et il a fallu attendre la quasi élimination de la divinité de la Nature pour que le hasard devienne un objet de questionnement, d'investigations, apparaisse sujet à des lois scientifiques (Bernstein, 1996 ; Senn, 2003). La probabilité est donc l'une des dernières nées des sciences naturelles. Elle remonte au XVII^{e} siècle presque deux millénaires après que naquirent les autres dans l'Antiquité.

Il y a la réalité structurelle, ou monde des lois, et la réalité naturelle. La probabilité relève de la première, la statistique de la seconde. Mais la seconde est fondée elle-même avant tout sur la probabilité. C'est grâce à la probabilité toutefois que la statistique s'avère si utile ...et réciproquement ! On modélise un phénomène aléatoire par une loi de probabilité. Des séries d'observations tirées de la nature permettent le calcul approximatif des paramètres du modèle, dont la précision fait elle-même l'objet de lois probabilistes.

Nous divisons ce travail en deux sections principales, l'une évidemment consacrée à la probabilité et l'autre à la statistique. Normalement on doit approfondir chacun de ces sujets pendant tout un cours, mais qui est souvent vite oublié, et il convient pour les cours qui suivent de mettre à disposition un résumé allant à l'essentiel de ces disciplines.

Nous devons prévenir le lecteur du caractère peu formel de ce petit texte. L'exposé nous semble suffisant pour permettre les calculs élémentaires nécessaires dans la vie courante.

Nous espérons que que ce résumé sera utile.

[1] http://fr.wikipedia.org/wiki/Montaigne.

1. La probabilité

Nothing comes of nothing.
W. Shakespeare, *King Lear* [1564 – 1616][1]

Dans ce chapitre nous verrons les concepts de variable aléatoire, discrète et continue ; de leurs principales caractéristiques ; de vecteur aléatoire ; et de fonction d'une variable aléatoire. Nous terminerons par un examen du cas d'une variable gaussienne, le plus important en pratique.

1.1 Le concept de variable aléatoire : VA

Soit une population quelconque $\mathcal{P}$ et un caractère quantitatif qu'on désire connaître de cette population. On suppose que la population est trop grande pour qu'on puisse le connaître exactement. En plus presque toute mesure est entachée d'erreur, on la dit *aléatoire*. On conçoit ainsi cette opération de mesure d'un caractère comme une fonction de $\mathcal{P}$ dans les réels, quitte éventuellement a coder numériquement les valeurs du caractères (*e.g.* la couleur des yeux, *etc.*). Cette fonction est notée par des lettres majuscules X, Y, *etc.*, ses valeurs en minuscules :

$$X \ : \mathcal{P} \longrightarrow \mathbb{R}$$

$$\omega \longmapsto X(\omega) = x \in \mathbb{R}\,.$$

Une telle fonction peut être soit continue, toutes les valeurs dans une plage de réels sont possibles, ou discrète, que quelques valeurs sont possibles.

La fonction peut être vue comme aléatoire comme si l'élément choisi, $\omega \in \mathcal{P}$ était "une boule dans une urne", et la valeur de X mesurée sur cet élément.

En fait, on note que la population, $\mathcal{P}$, ne joue plus qu'un rôle lointain, puisque c'est maintenant un charge sur le droite réelle, ou une partie de celle-ci qui définit la probabilité associée à une VA. Les sections suivantes illustreront ce concept.

1.1.1 VA discrète

Seules quelques valeurs sont possibles, par exemple le sexe ou une tranche d'âge, ou les trois qualités "excellent", "bon" ou "inacceptable" pour un produit. Soit n le nombre de valeurs possibles. On les code toujours par des nombres réels, par exemple : 2,1,0.

[1] http://en.wikipedia.org/wiki/Shakespeare.

La population étant vue comme infinie, on peut tirer autant d'observations qu'on désire, et si N, ce nombre, est assez grand, les proportions des fois où chaque valeur est atteinte donne un bonne approximations des proportions de ces valeurs dans la population. On note les "vraies" proportions :

$$\mathrm{P}[\,X = x_1],\ \cdots\ ,\ \mathrm{P}[\,X = x_n]\,,$$

et on a bien

$$\sum_1^n \mathrm{P}[\,X = x_i] = 1\,.$$

On conçoit ainsi qu'une fonction de probabilité discrète est la donnée d'un certain nombre de masses ponctuelles chargeant des points de la droite réelle. La masse totale est normalisée à l'unité. L'analogie de la masse avec une fonction de probabilité est fondamentale. On peut illustrer les probabilités discrètes par un diagramme en bâton (cf. Fig. 1.1)

Exemple 1.1 La loi binômiale. On a des boules de deux couleurs dans une urne dans les proportions p et $1-p$. On tire avec remplacement n boules dans l'urne. On pose X la variable aléatoire "nombre de boules de la première couleur choisi parmi les n". On voit aisément que bien des situations de la vie courante sont modélisées par de telles VA. Ainsi échantillonner n produits à la sortie d'une chaîne de production pour en vérifier la conformité ou non aux normes. Même si on ne remet pas la boule dans l'urne, pour utiliser une image, lorsque le nombre de boules est assez grand (encore la même image) la loi binomiale est une bonne approximation.

On écrit $X \sim \mathcal{B}(\,n\,;\,p\,)$, X suit une loi *binomiale* de paramètres n et p. On peut voir que les valeurs de X sont $k = 0, 1,\ \cdots, n$. La population est l'ensemble des $(\,a_1\,, \ldots\,, a_n\,)$ avec $a_i = 0$ ou 1, où on a noté conventionnellement 0 pour la première couleur et 1 pour la seconde. On montre que

$$\mathrm{P}[\,X = k] = \binom{n}{k}\, p^k\,(1-p)^{n-k}\,.$$

La formule dite du binôme nous donne immédiatement la raison du terme binomial pour décrire cette loi ou variable aléatoire :

$$\sum_{k=1}^{n} \left(\begin{array}{c} n \\ k \end{array}\right) p^k\,(1-p)^{n-k} = (p+1-p)^n = 1\,.$$

On peut illustrer les probabilités discrètes par un diagramme en bâton (cf. Fig. 1.1).

1.1.2 VA continue

Des exemples sont les durées de vie (utile) d'un objet ou d'un outil, ou encore le diamètre de boulons à la sortie de la production. Les valeurs des mesures ont alors des unités infiniment divisibles. Dans ce cas la probabilité d'une valeur particulière est nulle, tout ce qu'on peut mesurer c'est la probabilité d'un intervalle situé dans la plage des valeurs admissibles. En fait si on sépare celle-ci en intervalles de grandeurs égales,

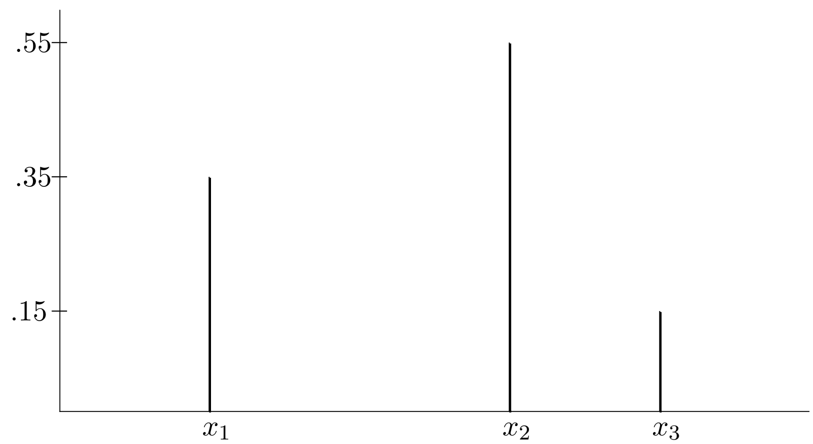

Fig. 1.1. Un diagramme en bâton pour une VA discrète de valeurs x_1, x_2, x_3

et qu'on tire un grand nombre de valeurs, chaque intervalle reçoit une proportion de $X(\omega)$ près de la vraie proportion, et on peut illustrer la VA par un histogramme, *i.e.* une fonction en escalier, donc constante par intervalles, où chaque intervalle préfixé a ces proportions comme hauteur. Il est coutumier de laisser les nombres en valeur absolue et non en proportion comme sur la Fig. 1.2, mais pour notre illustration on doit calculer en termes proportionnels. Lorsqu'on augmente le nombre d'observations on peut avoir des histogrammes avec des marches de plus en plus étroites, et on conçoit qu'à la limite on obtienne une "fonction de masse", de masse totale unitaire, une *densité de probabilité*, donc, définie sur la droite réelle. Notons-la $f(x)$. Cette fonction de $\mathbb{R}$ dans $\mathbb{R}$ est ainsi d'intégrale 1. On la considère le plus souvent définie sur tous les réels, et valant 0 hors de la plage admissible des valeurs de la VA. On supposera ici que l'intégrale de f sur tout intervalle existe, de même que ses moments de tous ordres.

On est passé des proportions aux probabilités par ce processus d'idéalisation de ce que l'empirique peut donner.

La probabilité que la variable aléatoire ait ses valeurs dans l'intervalle $[\,a\,;\,b\,]$ est donnée par la "masse" de cet intervalle :

$$\mathrm{P}[\,a \leq X \leq b\,] = \int_a^b f(x)\,dx\,.$$

On a aussi évidemment :

$$\int_{-\infty}^{-\infty} f(x)\,dx = 1 \qquad \text{et} \qquad \forall\, x\ \mathrm{P}[\,X = x] = 0\,.$$

1.1.3 Lois de probabilité

On dit qu'une variable aléatoire X suit une *loi* particulière lorsqu'on connaît sa densité. Il y a de nombreuses lois qui apparaissent dans la Nature (ou son idéalisation)

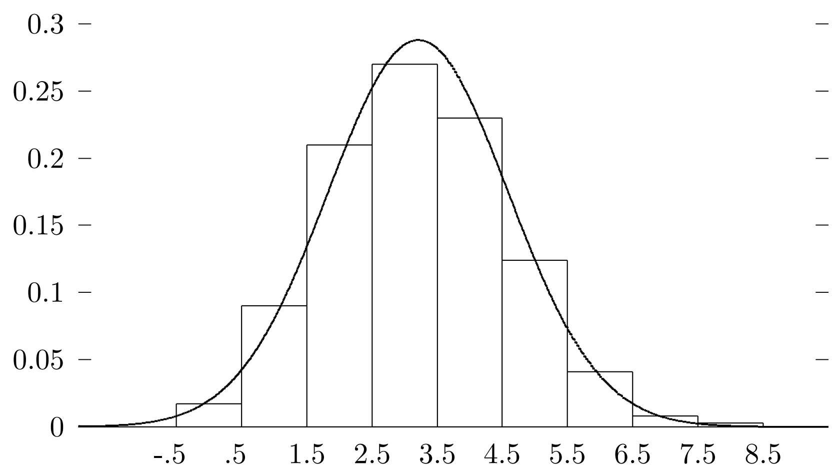

Fig. 1.2. Un exemple d'histogramme, et de son idéalisation en fonction de masse, ou densité de probabilité, lorsque le nombre de boîtes devient de plus en plus grand et celles-ci de plus en plus étroites.

et qui sont bien connues. Citons les lois discrètes binomiales, géométrique, de Poisson ; les lois continues exponentielles, de Laplace-Gauss ou gaussienne ou simplement normale (vu son importance), gamma, *etc.* Les livres de Johnson & Kotz (1969, 1970, 1972) constituent une mine de renseignements pour les lois de probabilité et leurs applications.

Les calculs de probabilité se font souvent en termes de la fonction de *répartition de probabilité* (répartition de masse), $F(x)$ qui n'est autre que la primitive de la densité, et qui donne, pour n'importe quel $x \in \mathbb{R}$ la masse cumulée jusqu'à x :

$$F(x) = \int_{-\infty}^{x} f(t)\, dt\,,$$

ce qui implique bien sûr :

$$\mathrm{P}[\, a \leq X \leq b \,] = F(b) - F(a)\,. \tag{1.1}$$

Plusieurs des lois que nous verrons par la suite ont leur fonction de répartion tabulée (il est d'ailleurs impossible d'en obtenir les valeurs exactes), et les calculs de probabilité des intervalles se font par l'équation 1.1.

1.2 Principales caractéristiques des VA

Quelle que soit la loi d'une VA, on cherche toujours à en connaître les paramètres pour la population concernée. Ces paramètres sont de deux types : les paramètres de localisation et ceux de dispersion.

Il est à noter que ces paramètres ne peuvent pas le plus souvent être connus dans leur réalité naturelle, par opposition à leur réalité théorique ou structurelle, que par échantillonnage, c'est à dire par des mesures sur un certain nombre d'éléments de la population. On reviendra sur ce point dans le chapitre suivant.

1.2.1 Les paramètres de localisation

On appelle la *moyenne* d'une VA son premier moment, ou le centre de masse de la répartition :

$$\mu = \int x f(x)\, dx \qquad \mu = \frac{1}{n} \sum_{i=1}^{n} x_i \, \mathrm{P}[\, X = x_i] \,.$$

Les grandes valeurs peuvent être vraiment excentriques mais pas très nombreuses, il n'est donc pas évident que la moyenne décrive bien la tendance centrale de la population. Le premier moment en effet est fortement tiré par les valeurs excentriques. Il suffit de prendre un exemple avec peu de points et déplacer l'une des valeurs extrêmes pour constater son influence sur la moyenne. On préfère souvent recourir à une mesure plus *robuste* pour localiser une population.

Soit p un nombre entre 0 et 1. La valeur x_p de la variable où elle cumule la proportion p de la masse totale s'appelle le p^e quantile (ou le percentile si p s'exprime en pourcentage) de la répartition. La *médiane* m est le 50^{e} percentile d'une répartition, *i.e.* la valeur de la variable telle que la moitié de la population se trouve de part et d'autre. On a évidemment :

$$F(\, x_p \,) = p \qquad F(\, m \,) = .5$$

La médiane est un bien meilleur indicateur, par exemple des revenus des populations, que la moyenne. Bien évidemment c'est à cause du fait que la médiane est totalement insensible aux valeurs excentriques des VA. Les points “aberrants”, les erreurs de mesure, n'y jouent presqu'aucun rôle.

On peut voir l'*étendue* à la fois comme un indicateur de localisation et comme un indicateur de dispersion. L'étendue est la différence entre la valeur maximum et la valeur minimum d'une VA.

$$E = x_{\mathrm{max}} - x_{\mathrm{min}} \,.$$

1.2.2 Les paramètres de dispersion

Le plus usité des paramètres de dispersion est l'écart type, σ, la racine carrée de la variance de la répartition, évidemment σ^2. L'écart type s'exprime dans les mêmes unités que la variance et est donc, de ce fait, une mesure plus intuitive.

$$\sigma^2 = \int (x-\mu)^2 f(x)\, dx \qquad \sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \, \mathrm{P}[\, X = x_i \,] \,.$$

En termes de notre métaphore mécaniste, il s'agit du deuxième moment de la répartition centré sur la moyenne. Les moments centrés d'ordre supérieur jouent parfois un rôle, notamment le troisième qui vaut 0 pour des répartitions symétriques.

1.3 Vecteurs aléatoires

D'une seule VA à plusieurs observées sur sur la même population, la généralisation est immédiate : $(\, X_1, \cdots, X_n \,)$ est un vecteur aléatoire, dont chaque coordonnée est une VA, ayant une densité dans le cas continu, *etc.* On a le concept de fonction de *répartition conjointe* définie par la probabilité suivante :

$$F(x_1, \cdots, x_n) = \mathrm{P}[\, X_1 \leq x_1, \cdots, X_n \leq x_n \,] \, .$$

La *densité conjointe* est la dérivée n^e relativement à toutes les variables. Les calculs se font maintenant sur des sous ensembles de $\mathbb{R}^n$, les concepts sont naturellement multidimensionnels.

Il est important de définir à ce stade le concept probabiliste d'indépendance qui signifie simplement que la densité conjointe d'un vecteur aléatoire dont les composants sont *indépendants* deux à deux est le produit des densités individuelles. Ainsi dans le cas d'un couple de VA, (X, Y) :

$$f(x, y) = f_1(x)\, f_2(y) \, .$$

Intuitivement parlant, le fait qu'un résultat apparaisse pour la première VA n'a aucune influence sur le résultat de la seconde. Ainsi en est-il du tir successif de pièces de monnaie, de dés. En revanche, le fait pour un individu d'avoir un haut salaire est souvent lié à son haut niveau d'éducation, les variables aléatoires X le niveau de revenu, et Y le nombre d'années d'éducation ne sont pas des variables indépendantes. On peut soupçonner de même que le fait pour une équipe de travailleurs d'être au début ou à la fin de son quart de travail peut avoir une influence sur la qualité du travail produit (quelle que soit la façon de mesurer cela).

1.4 Fonctions d'une VA

Toute variable aléatoire peut se composer avec une autre fonction H de $\mathbb{R}$ dans $\mathbb{R}$ et donne, bien entendu une fonction d'une population dans $\mathbb{R}$, une variable aléatoire donc, dite une fonction d'une VA.

$$H \circ X \equiv H(X) : \mathcal{P} \longrightarrow \mathbb{R} \, .$$

D'une façon simplifiée, on peut voir que $H(X)$ est une fonction de $\mathbb{R}$ dans les réels, avec le premier ensemble muni d'une fonction de masse unitaire. Nous ne développerons pas ici les concepts de densité de la fonction composée, comment calculer les probabilités, il nous suffira de voir certains concepts pertinents à notre objectif. Ainsi le concept d'espérance, central en théorie des probabilités.

Définition 1.2 *Soit une fonction $H(X)$ d'une variable aléatoire de densité f(x) dans le cas continu, ou sinon de probabilités $P[\, X = x_i \,]$. Alors*

$$E(\, H(X)\,) = \int H(x) f(x)\, dx \qquad E(\, H(X)\,) = \sum H(x_i)\, P[\, X = x_i \,] \, .$$

La généralisation aux fonctions d'un vecteur aléatoire est immédiate, et on montre alors sans peine que que l'*opérateur espérance* est une opérateur linéaire au sens suivant :

$$E(\sum a_i X_i\,) = \sum a_i\, E(X_i) \, . \tag{1.2}$$

Soit $H(X_1, \cdots, X_n) = \sum a_i X_i$, toute combinaison linéaire d'une suite de VA est bien une fonction d'un vecteur aléatoire. Cette fonction jouera un rôle important par la suite. Un calcul élémentaire montre alors le résultat.

Pour la variance, les choses se compliquent (voir plus loin), mais on a directement :

$$V(aX) = a^2 V(X) .$$

Les cas particuliers $H(X) = X$ et $H(X) = (X - \mu)^2$ montrent que la moyenne de X est aussi son espérance, et que sa variance est la moyenne du carré des écarts à la moyenne.

$$E(X) \equiv \mu_X \qquad V(X) = E[\,(X - \mu)^2\,] = E(X^2) - (\,E(X)\,)^2 .$$

Cette dernière équation est le résultat du développement du carré ainsi que de la linéarité de l'opérateur espérance, et elle est utilisée dans les calculs explicites de variance.

Par ailleurs la fonction du couple de VA : $(X, Y) \mapsto (X - \mu_X)(Y - \mu_Y)$ a une espérance qui s'appelle la covariance du couple de VA. La linéarité de l'opérateur espérance montre aisément :

$$E(\,(X - \mu_X)(Y - \mu_Y)\,) = E(XY) - E(X)\,E(Y) \;\equiv\; \mathrm{cov}(\,X\,,\,Y\,).$$

Cette espérance particulière porte le nom de la *covariance* entre les deux VA X et Y, $\mathrm{cov}[\,X\,,\,Y\,]$. La différence de l'espérance du produit et du produit des espérances est une marque quantitative de lien entre deux VA, de son manque d'indépendance. On montre en effet que dans le cas d'indépendance la covariance est nulle.

Un exemple fera comprendre intuitivement ce concept. Supposons que les deux variables X et Y soient discrètes et aient le même nombre de valeurs, soit n (comme dans un échantillon où on observe deux variables à la fois sur chaque "individu" ou "unité statistique", on a alors :

$$\mathrm{cov}[\,X\,,\,Y\,] = n^{-1} \sum (x_i - \mu_X)(y_i - \mu_Y)\,\mathrm{P}(\,X_i = x_i\,,\,Y_i = y_i\,) .$$

Maintenant raisonnons sur les tendances. Si une valeur supérieure à la moyenne d'une variable a tendance à impliquer le même chose pour l'autre, le produit $(x_i - \mu_X)(y_i - \mu_Y)$ est positif, de même pour le cas inverse. Pensons par exemple aux variables taille et poids des individus. Alors la somme définissant la covariance sera grande. Par ailleurs, si lorsqu'une variable est supérieure à sa moyenne n'indique nullement une tendance pour la position de l'autre relativement à la sienne, alors la somme de la covariance sera petite, les termes ayant tendance à changer de signe et à s'annuler. Pensons par exemple à la taille d'un individu et à son revenu.

On sait que deux variables indépendantes ont une covariance nulle, mais que la réciproque n'est pas vraie (sauf dans le cas où les variables sont gaussiennes). Ainsi pour deux variables indépendantes, on a l'identité : $E(XY) = E(X)\,E(Y)$.

On voit que $\mathrm{cov}(X, X) = \sigma_X^2$, et on normalise habituellement la covariance en *corrélation*.

Définition 1.3 *Soit un vecteur aléatoire* (X, Y)*, alors la corrélation entre les deux VA est :*

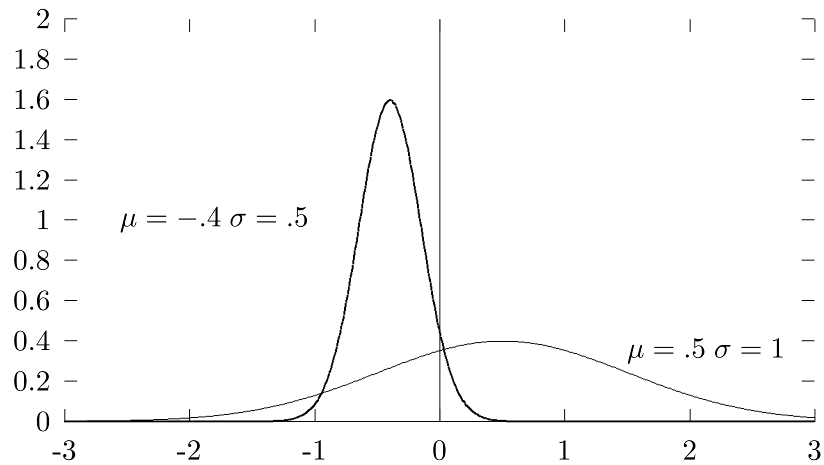

Fig. 1.3. Deux lois gaussiennes $\mathcal{N}(\mu\,;\,\sigma^2\,)$.

$$corr\,(\,X,Y\,) \equiv \rho_{XY} = \frac{cov\,(X,Y)}{\sigma_X\,\sigma_Y}\,.$$

Le coefficient de corrélation est compris entre -1 et 1, et au voisinage des extrémités ± 1, le lien entre X et Y est de plus en plus linéaire :

$$\rho_{XY} \approx \pm 1 \Longrightarrow Y \approx\, aX + b$$

avec la corrélation +1 lorsque $a > 0$ et -1 dans le cas contraire. Cette implication est probabiliste et peut souffrir des exceptions, pour certaines valeurs de X et de Y.

Le calcul direct à l'aide des définitions donne la relation suivante :

$$V(\sum_i a_i\,X_i\,) = \sum_i a_i^2 V(\,X_i\,) + 2\sum_{i\neq j} \mathrm{cov}[X_i\,,\,X_j\,]\,, \tag{1.3}$$

et cette équation se réduit à la première somme dans le cas où les variables sont indépendantes deux à deux.

1.5 Exemple important : la loi de Laplace-Gauss

Cette loi est importante et mériterait à elle seule un long développement que nous limiterons ici au strict minimum. Elle est si fréquente dans la pratique (nous verrons pourquoi) qu'on l'appelle très souvent la loi *normale* :

$$X \;\sim\; \mathcal{N}(\,\mu\,;\,\sigma^2\,)$$

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\right)\,.$$

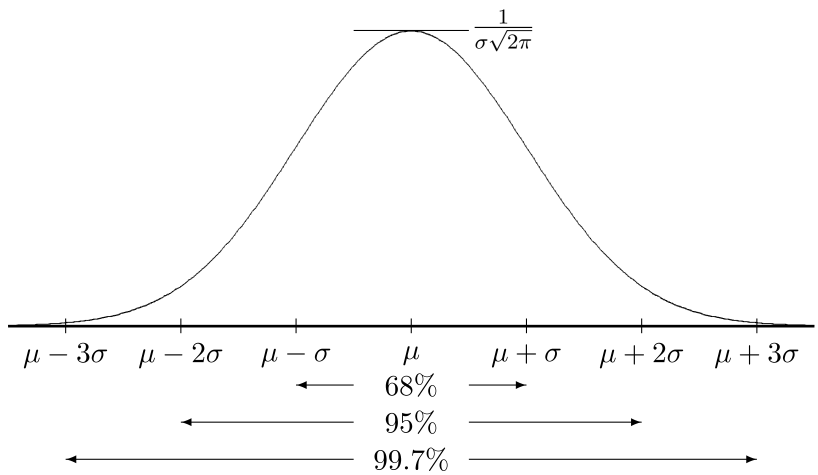

Fig. 1.4. La densité de la loi gaussienne centrée réduite avec quelques indices de concentration.Noter ce qui se passe lorsque σ diminue.

On peut centrer et réduire une loi gaussienne quelconque et nous obtenons la loi, notée conventionnellement Z, $Z \sim \mathcal{N}(0\,;1)$, qui est la seule qui soit tabulée (*cf.* Fig. 1.4). On ramène les calculs de probabilité sur toute loi gaussienne à la consultation de la table de la répartition cumulative pour la loi Z, notée Φ (pour une approximation, voir plus loin).

De toute façon, les équation 1.2 et 1.3 montrent aisément que la transformation qui consiste à *centrer* et *normer*, on dit aussi réduire, une VA donne une nouvelle VA de moyenne nulle et de variance unité :

$$Y = \frac{X - \mu}{\sigma} \Longrightarrow E(Y) = 0 \text{ et } V(Y) = 1\,.$$

Les calculs pratiques de probabilité se font en manipulant les expressions par des opérations arithmétiques simples. Ainsi, par exemple :

$$\mathrm{P}[\,a\ \leq X\ \leq\ b\,] = \mathrm{P}[\,\frac{a-\mu}{\sigma}\ \leq\ \frac{X-\mu}{\sigma}\ \leq \frac{b-\mu}{\sigma}\,] = \Phi(\frac{a-\mu}{\sigma}) - \Phi(\frac{b-\mu}{\sigma})\,,$$

et dernière égalité suppose que X est une gaussienne. Chaque valeur de la fonction cumulative de la loi gaussienne centrée réduite, Φ, est tabulée avec la précision suffisante pour la plupart des applications. La symétrie de la loi gaussienne implique d'ailleurs que seules les tables pour les valeurs positives de la variable sont nécessaires (pourquoi ?[2]).

Ce qui rend la loi gaussienne centrale sont les deux propriétés suivantes que nous énonçons sous forme de théorème.

Théorème 1.4 *Soit une suite de VA indépendantes :* $X_1, \cdots, X_n$.

[2] On doit évidemment faire un raisonnement sur le graphique 1.4, au besoin, et noter que $\Phi(-z) = 1 - \Phi(z)$.

[1] Si elles sont gaussiennes alors $\forall\, a_i \in \mathbb{R}$, $\sum a_i\, X_i$ *suit une loi gaussienne de moyenne*

$$\mu = \sum a_i\, \mu_i \quad \textit{et de variance} \quad \sum a_i^2\, \sigma_i^2$$

[2] Sous des hypothèses très souples concernant les lois des X_i*, une combinaison linéaire des* X_i *suit approximativement une loi gaussienne, d'autant plus près d'ailleurs que le nombre de termes de la somme est grand. Les moyenne et variance sont celles des équations précédentes.*

Ce dernier résultat s'appelle le *théorème limite central*, et indique que presque toute variable aléatoire qui peut être vue comme une somme de lois indépendantes est à toutes fins pratiques gaussienne, pourvu que le nombre de termes soit assez grand. Ce nombre de termes est en fait d'autant plus petit que les lois sont proches des gaussiennes, symétriques par exemple, *etc.*

Un exemple un peu surprenant est le suivant. Si on pose $X_i = 0$ ou 1, avec probabilité respective de p et $(1-p)$, donc des variables binomiales de paramètres 1 et p — on parle d'essai de Bernoulli —, on peut voir qu'une loi $X \sim \mathcal{B}(\,n\,;\,p\,)$ est une somme de n VA X_i indépendantes.

$$X = \sum_{i=1}^{n} X_i \;\sim\; \mathcal{B}(\,n\,;\,p\,) \;\approx\; \mathcal{N}(np\,;\,npq)$$

Les moyenne et variance des deux lois sont évidemment les mêmes. Dans la pratique, cette approximation gaussienne d'une loi binomiale, donc d'une loi discrète par une loi continue, doit être corrigée par la correction dite de continuité.

2. La statistique

There are lies, damn lies, and statistics.
Sir Winston Churchill [1874 – 1965][1]

Le terme de statistique dans son sens usuel de recueil de données remonte au XVII^e siècle et vient du latin "statisticum" : qui a trait à l'état. L'activité elle-même, tenir des registres de données chiffrées à des fins gouvernementales, remonte à la plus haute antiquité. Le besoin d'extrapoler à partir de données forcément incomplètes n'a pu reposer sur des méthodes scientifiques que lorsque les premiers éléments de probabilité se répandirent, notamment via l'astronomie (la théorie des erreurs). C'est sous l'impulsion d'Adolphe Quételet (1796-1874), un astronome belge devenu démographe, que la statistique fut basée sur les probabilités. Son livre " Physique sociale " publié en 1835 puis dans une seconde édition en 1869, marqua son époque et la suite des choses. Voir la monographie de Drœsbeke & Tassi (1990) pour une excellente introduction à l'histoire de la statistique.

2.1 Concept d'échantillon

Nous l'avons mentionné dans l'introduction de ce travail, l'observation de la Nature ne peut nous donner que des valeurs en nombre forcément assez limité, quelques mesures des phénomènes d'intérêt. Ainsi, même avec un bon modèle probabiliste d'un phénomène sujet à aléas, c'est la plupart du temps l'observation seule qui peut nous en donner les paramètres, et nous permettre d'agir sur celui-ci, ou plus simplement de le valider.

Le mieux qu'on puisse faire, c'est de faire ces mesures dans des conditions qui les rendent le plus possible indépendantes les unes les autres, au sens que nous avons précisé plus haut. On est souvent obligé d'ailleurs de supposer cette indépendance, sans être capable de vraiment la fonder.

Nous avons donc tiré d'une population des valeurs d'un paramètre d'intérêt. En vertu de ce qu'on vient de dire, on a, disons, n observations d'un paramètre sujet à aléas. On a n réalisations d'une variable aléatoire X. Supposons en fait qu'on ait *une* réalisation de n variables aléatoires, toutes identiques et indépendantes les unes des autres. Ce changement de perspective fait toute la différence du monde !

[1] http://en.wikipedia.org/wiki/Churchill. On ne prête qu'aux riches ce mot est aussi atribué à Mark Twain...

Définition 2.1 *Un échantillon de* taille n *d'un paramètre est une suite de* n *observations de variables identiquement distribuées et indépendantes, une observation pour chaque variable aléatoire :*

$$X_i = x_i\,, \quad X_i \;\; i.i.d.$$

On distingue communément les échantillons *exhaustifs* et *non exhaustifs*. Les premiers sont tirés d'une population finie et chaque observation la diminue d'une unité. En principe on pourrait observer toute la population. C'est comme tirer une boule à la fois d'une urne sans la remettre avant de tirer l'observation suivante. Les conditions d'observation changent de fois en fois, on ne peut alors décemment supposer l'indépendance des X_i. Dans le cas non exhaustif, la population est en pratique infinie, on remet chaque "boule" dans l'urne après usage, et on suppose l'indépendance des X_i. En fait lorsque la population est très grande et que la taille échantillonnale en est relativement petite, on a une bonne approximation de cette situation, même si on ne "sonde" jamais deux fois le même individu et qu'on se trouve en réalité dans le cas exhaustif. Naturellement, l'observation fine demande des corrections…

2.2 L'estimation

On dispose donc d'une observation d'un vecteur aléatoire $(X_1, \cdots, X_n)$, avec les X_i i.i.d. Toutes les méthodes de la statistique inférentielle reposent sur des échantillons. Le but est d'inférer à partir d'échantillons des caractéristiques de la population dans son ensemble, et de quantifier la certitude qu'on peut tenir de ces informations. La démarche statistique est inductive, et non déductive comme les sciences dites exactes. Mais y a-t-il véritablement des sciences exactes, lesquelles reposent sur l'observation forcément limitée de la Nature ?

La statistique définit sans cesse des fonctions des échantillons qui sont construites à des fins inférentielles. De telles fonctions s'appellent des *statistiques* (sens technique du mot).

Prenons par exemple, la statistique *moyenne échantillonnale* :

$$(X_1, \cdots, X_n) \longmapsto \frac{1}{n}\sum_{i=1}^{n} X_i \equiv \overline{X}\,.$$

La fonction aléatoire est observée une seule fois sur chaque échantillon de taille n d'une variable aléatoire X. La valeur observée est $\overline{x} \equiv n^{-1}\sum x_i$ (noter les minuscules). On utilise cette statistique pour approcher la moyenne de la variable.

Quel est l'avantage ? Le théorème 1.4 nous l'explique aisément ainsi que les équations 1.2 et 1.3. Si X est gaussienne, ou non pourvu que n soit assez grand, on peut affirmer :

$$\overline{X} \sim \mathcal{N}(\mu\,;\, \frac{\sigma^2}{n})\,,$$

où μ est la moyenne de X et σ^2 sa variance. La valeur centrale de $\overline{X}$ est donc μ et aura d'autant plus tendance à être près de celle-ci que n est grand puisque la variance de $\overline{X}$ diminue avec n : on passe en fait avec $\overline{X}$ d'un écart type σ à $\frac{\sigma}{\sqrt{n}}$. Voir la Fig. 2.1.

Il est ainsi avantageux d'*estimer* la valeur du paramètre inconnu μ, la moyenne de $\overline{X}$, par $\overline{x}$, la réalisation de la variable moyenne échantillonnale obtenue sur un échantilllon de X. On a en fait la définition générale suivante.

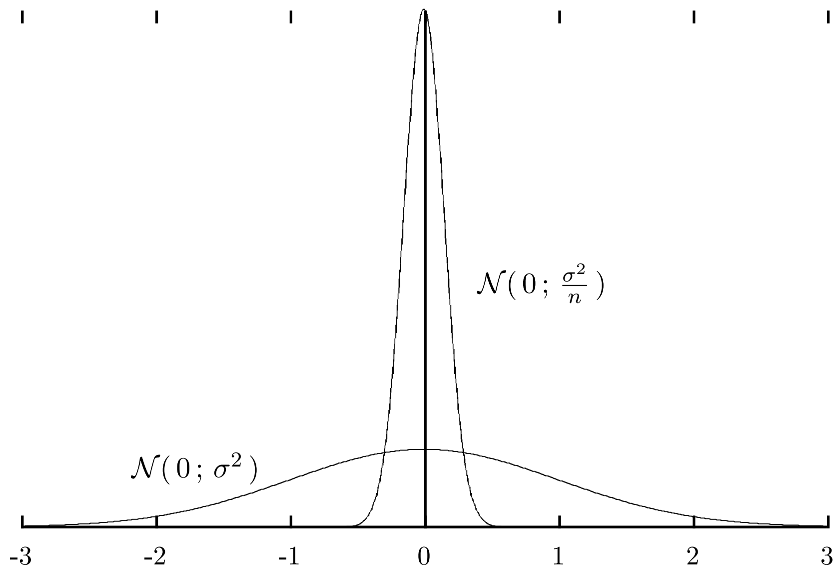

Fig. 2.1. L'effet de l'échantillonnage sur l'estimation d'u paramètre : une concentration de la masse d'autant plus marquée que n est grand.

Définition 2.2

[1] Une statistique qui vise à estimer un paramètre d'une VA s'appelle un estimateur. *On note un estimateur du paramètre θ en général par la même lettre avec un accent circonflexe : $\hat{\theta}$ On dit θ chapeau.*

[2] Un estimateur est dit sans biais si son espérance est la valeur du paramètre : $E(\hat{\theta}) = \theta$. Le biais est la différence entre les deux.

Ainsi dans l'exemple ci-haut, on voit bien que $\overline{X}$ est un estimateur sans biais de μ. Un estimateur est une VA, et en connaître la loi permet d'encadrer la valeur inconnue du paramètre à l'aide d'un intervalle dit *intervalle de confiance*. Dans notre exemple, puisque $\overline{X}$ suit une loi gaussienne la consultation de la table donne immédiatement :

$$\mathrm{P}[\,-1.96 \,\leq\, \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \,\leq\, 1.96\,] = .95\,. \tag{2.1}$$

On trouvera μ avec une probabilité de 95% dans l'intervalle $[\,\overline{X} - 1.96\,\frac{\sigma}{\sqrt{n}}\,;\,\overline{X} + 1.96\,\frac{\sigma}{\sqrt{n}}\,]$, qui est obtenu par manipulation arithmétique de l'expression à l'intérieur du signe probabilité. On écrit

$$\mu = \overline{X} \,\pm\, 1.96\,\frac{\sigma}{\sqrt{n}} \quad \text{avec probabilité } 95\%\,. \tag{2.2}$$

Tout cela est bien théorique puisqu'en pratique on n'a que rarement la valeur de σ . . . Qu'à cela ne tienne, on peut montrer qu'il existe un estimateur sans biais de σ^2 bien commode, il s'agit de S^2, dont la racine estime σ :

$$S^2 = \frac{1}{n-1}\,\sum_{i=1}^{n}(X_i - \overline{X})^2\,. \tag{2.3}$$

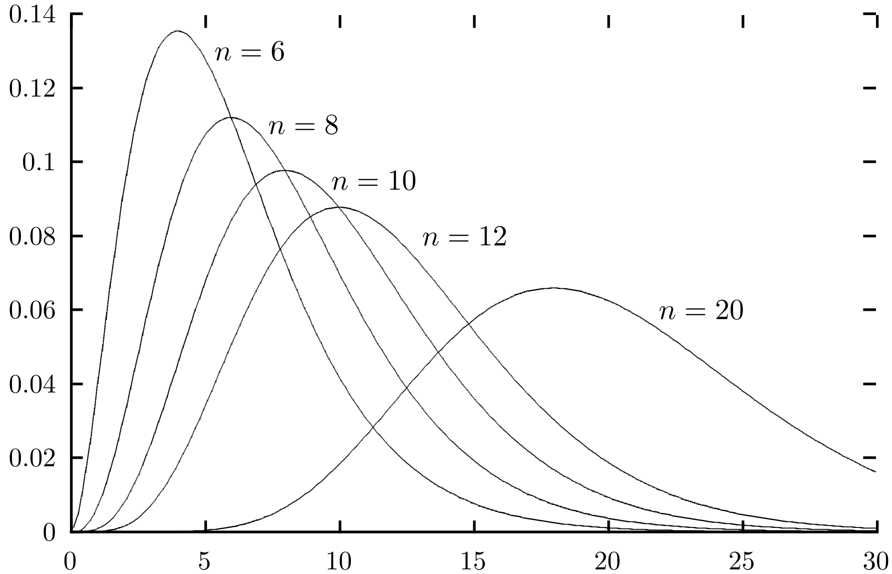

Fig. 2.2. Les lois χ^2_n sont des sommes de carrés de n gaussiennes standardisées indépendantes. On voit sur les graphiques que le théorème limite central s'applique.

On réalise S^2 une fois sur chaque échantillon $S^2 = s^2$:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2 ,$$

à noter les minuscules. Mais qu'en est-il de la loi de S^2 ? On démontre que la VA $(n-1)\, S^2 / \sigma^2$ suit une loi χ^2.

La loi χ^2 est une de ces lois qui n'existent (à toutes fins pratiques) que liées à des échantillons. Ce sont des lois dites, pour ce fait, échantillonnales. Nous en verrons deux autres ici.

La loi χ^2 est variable selon la taille échantillonnale, on parle de ses *degrés de liberté*. Notons que dans l'équation 2.2 on a n termes ou degrés de liberté dans la somme, mais le terme $\overline{X}$ ne peut être déterminé que par les autres termes. On perd ainsi 1 degré de liberté.

$$\frac{(n-1)\, S^2}{\sigma^2} \sim \chi^2_{n-1} . \tag{2.4}$$

On a à notre disposition des tables de χ^2, et à l'aide des quantiles de celles-ci, on établit aisément un intervalle de confiance pour une valeur de S^2 :

$$\mathrm{P}[\, \chi^2_{n-1,1-\frac{\alpha}{2}} \leq \frac{(n-1)\, S^2}{\sigma^2} \leq \chi^2_{n-1,\frac{\alpha}{2}} \,] = 1 - \alpha . \tag{2.5}$$

Il est grand temps d'expliquer ici le sens un peu particuliers des des quantiles qu'on trouve en général associé aux principales lois échantillonnales. Nous l'expliquerons sur la loi gaussienne centrée réduite. Ce sens particulier vient de la tradition anglaise (les anglais s'y connaissent en excentricités. . .) qui fut la première, au début du XX^e siècle à développer ces idées.

On appelle communément le *quantile* z_α ce qui est en fait le $(1-\alpha)^e$ quantile au sens énoncé plus haut. Cette petite idiosyncrasie est un des charmes de la statistique… Ainsi donc

$$\Phi(z_\alpha) = 1 - \alpha\,.$$

Après le quantile z_α , on trouve donc la masse α de la gaussienne centrée réduite.

Cette notation particulière est valable pour la loi gaussienne, on vient de l'illustrer, mais aussi pour les lois χ^2_n, les lois T_n ainsi que les lois de Fisher F_{n_1,n_2} (on dit aussi de Fisher-Snedecor) qui, elles, ont deux degrés de liberté. On verra leurs applications un peu plus loin.

Donc les mêmes principes d'écriture et de lecture de ces tables s'appliquent pour le sens des expressions : $\chi^2_{n,\alpha}$, $\mathrm{t}_{n,\alpha}$, $\mathrm{F}_{n_1,n_2,\alpha}$, et ce pour tout α entre 0 et 1.

Revenons à l'équation 2.2. Une manipulation arithmétique nous donne l'intervalle de confiance suivant (il n'est pas symétrique et ne permet pas l'écriture simplifiée de l'intervalle de confiance comme pour la moyenne échantillonnale). Avec une probabilité de $(1-\alpha)$ on trouve :

$$\sigma^2 \in \left[\, \frac{(n-1)\,S^2}{\chi^2_{n-1,\frac{\alpha}{2}}} \;\leq\; \sigma^2 \;\leq\; \frac{(n-1)\,S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} \,\right] .$$

On est en mesure maintenant de rendre notre recherche de la vraie valeur inconnue μ. On a vu qu'un échantillon de X nous donne

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \;\sim\; \mathcal{N}(0\,;\,1)\,.$$

Le problème est que σ n'est pas plus connu que μ. Maintenant à l'aide de l'estimateur S de σ on peut remplacer ce dernier dans l'équation précédente par S, mais la répartition de la statistique change légèrement d'une loi gaussienne en une loi T de Student avec $n-1$ degrés de liberté :

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \;\sim\; \mathrm{T}_{n-1}\,.$$

Les lois T_n sont évidemment tabulées en termes des degrés de liberté, et on peut y lire les quantiles d'utilité générale : $\mathrm{t}_{n,\alpha}$, avec $\alpha = .25, .1, .05$, *etc.* Il ne faut pas oublier la petite idiosyncrasie rapportée plus haut pour l'interprétation de ces quantiles (*cf.* la Fig. 2.3). La loi gaussienne centrée réduite est assez semblable aux diverses lois T_n, et pour $n \geq 60$ les deux sont pratiquement confondues. Les lois T_n sont de moyennes 0 et symétriques tout comme la loi gaussienne. Ainsi dans la Fig. 2.3, on pourrait remplacer les quantiles $\mathrm{t}_{n,\alpha}$ par z_α et avoir les mêmes interprétations.

L'équation 2.2 devient donc, lorsqu'on se place dans la situation plus réaliste d'avoir à estimer σ (noter les minuscules) :

$$\mu = \overline{x} \,\pm\, \mathrm{t}_{n-1,\frac{\alpha}{2}}\,\frac{s}{\sqrt{n}} \;\text{ avec probabilité }\; (1-\alpha)\,. \tag{2.6}$$

Cet intervalle est souvent mais pas toujours légèrement plus grand que celui obtenu lorsque σ est supposé connu. Cela se comprend puisqu'on a ici moins d'information à exploiter, σ étant estimé et non connu exactement.

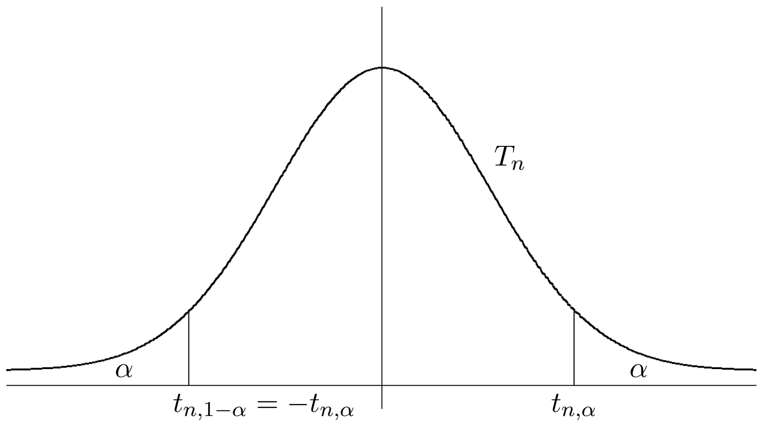

Fig. 2.3. Les lois T sont proches des gaussiennes, et à toutes fins pratiques sont identiques à celle-ci pour $n \geq 60$.

2.2.1 De l'usage du biais

Les estimateurs sans biais sont recherchés mais il ne sont pas toujours les plus commodes. Il peut arriver qu'un estimateur biaisé soit plus intéressant. C'est le concept de plus petite distance moyenne au paramètre qui est le bon critère.Plus précisément, on définit l'écart quadratique moyen d'un estimateur.

Définition 2.3 *Soit $\hat{\theta}$ un estimateur de θ. l'écart quadratique moyen de $\hat{\theta}$ est la quantité suivante :*

$$EQM(\hat{\theta}) \equiv E([\hat{\theta} - \theta]^2) = V(\hat{\theta}) + Biais^2(\theta).$$

La dernière égalité est obtenue par développement du carré dans le membre de gauche. La Fig. 2.4 montre bien qu'un biais peut être compensée par une petite variance pour rendre un estimateur biaisé plus efficace qu'un non biaisé. On peut développer plusieurs mesures correctrices pour compenser le biais d'un estimateur.

2.3 Tests d'hypothèse

Supposons qu'on ait une hypothèse concernant un paramètre, une mesure, et qu'on prenne un échantillon de cette mesure (forcément aléatoire). Si on note la VA concernée X, on peut alors tirer de $\overline{X} = \overline{x}$ un intervalle de confiance de niveau $(1 - \alpha)$ avec α petit pour μ la vraie valeur du paramètre. Comme il n'y a alors qu'une probabilité α qu'elle soit à l'extérieur de cet intervalle, on sera porté à rejeter l'hypothèse advenant le cas que sa valeur supposée y soit. L'idée est simple, et dans un premier temps apparaît la seule possible.

Mais en y réféchissant bien, on peut raisonner un peu à l'inverse, qui est d'ailleurs absolument équivalente, mais le concept est plus riche en développements. Raisonnons sur le cas d'une hypothèse concernant une moyenne : $\mu = \mu_0$. Appelons notre hypothèse, l'hypothèse nulle :

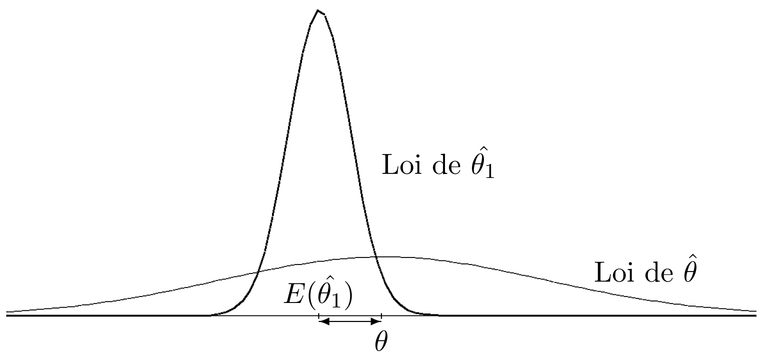

Fig. 2.4. L'estimateur $\hat{\theta}$ est sans biais, le biais de $\hat{\theta}_1$ est noté par une double flèche. l'EQM de $\hat{\theta}_1$ compense pour son biais.

$$\mathrm{H}_0 \; : \mu = \mu_0$$

Lorsqu'on prend un échantillon de la mesure, on peut alors calculer, *sous cette hypothèse*, qu'on a une probabilité de disons 95%, pour fixer les idées, que la variable aléatoire $\overline{X}$ soit située dans l'intervalle

$$[\, \mu_0 - \mathrm{t}_{n-1,.025} \, \frac{s}{\sqrt{n}} \, ; \, \mu_0 + \mathrm{t}_{n-1,.025} \, \frac{s}{\sqrt{n}} \,] \, .$$

Cet intervalle est un peu plus petit si on connaît la valeur de σ , puisqu'alors on peut remplacer s par σ , et $\mathrm{t}_{n-1,-.025}$ par $z_{.025}$.

Ainsi, il n'y a que 5% des chances, sous H_0, qu'une réalisation de $\overline{X}$ tombe hors de cet intervalle. Advenant le cas pour $\overline{x}$ la valeur obtenue de l'échantillon on rejette H_0. On a alors 5% des chances de se tromper, 5% des chances de rejeter l'hypothèse si elle est vraie.

Le concept est plus riche avons-nous dit, car apparaît naturellement ici un autre type d'erreur : si l'hypothèse dans les faits n'est pas vraie et que la vraie valeur est disons $\mathrm{H}_1 : \mu = \mu_1$ différente de la précédente, formant l'hypothèse dite alternative, alors il y a un certain risque à accepter H_0, puisqu'alors

$$\overline{X} \; \sim \mathcal{N}(\mu_1 \, ; \, \frac{\sigma}{\sqrt{n}} \,) \, .$$

Nous raisonnons ici dans le cas d'école où σ est connu pour simplifier. En réalité il faudrait, pour être plus réaliste, raisonner sur des lois T_n, ce qui change les calculs mais pas vraiment les formes des courbes qui illustrent les concepts (*cf.* Fig. 2.5).

L'intervalle d'acceptation de H_0 a une certaine probabilité de réalisation sur un échantillon, sous cette hypothèse alternative censée représenter la réalité, et accepter H_0 si la vraie valeur est celle de H_1, entraîne alors un risque, une probabilité.

La première s'appellera donc l'erreur de *première espèce*, celle-ci de *seconde espèce*. L'erreur de seconde espèce ne peut être calculée en fait qu'en postulant une autre réalité que celle donnée dans l'hypothèse nulle, c'est à dire, pour notre exemple, une valeur précisée $\mu = \mu_1$.

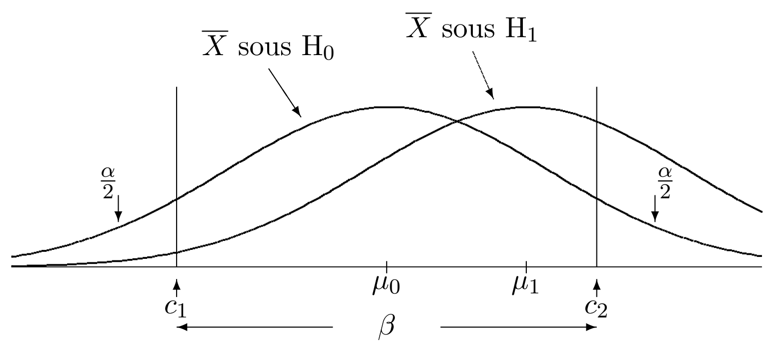

Fig. 2.5. Illustration dans le cas où σ est connu des erreurs de première espèce α et de deuxième espèce β .

Tableau 2.1. Les erreurs de décision concernant les hypothèses.

Décision	Réalité	
	H_0 vraie	H_1 vraie
H_0 acceptée	—	β erreur de seconde espèce
H_1 acceptée	α erreur de première espèce	—

Nous avons en fait le tableau 2.1 qui montre la richesse de l'introduction de la façon dite inverse de raisonner mentionnée au début du paragraphe. Cela provient de l'introduction alors évidente, parce qu'impliquant aisément un calcul de risque, des deux hypothèses : l'hypothèse nulle et son alternative. Les hypothèses viennent par paires tel que mentionné plus haut.

On voit ici la *stratégie* ou *paradigme* des tests statistiques. On a une hypothèse nulle concernant un paramètre, et la possibilité de prendre un échantillon de taille n de ce paramètre. On se fixe une erreur de première espèce de risque α. L'hypothèse nulle donne alors un intervalle $[\,c_1\,;\,c_2\,]$ construit à partir d'un estimateur du paramètre, région où on s'attend à trouver la réalisation de l'échantillon avec probabilité $(1-\alpha)$. La région complémentaire s'appelle la *région critique*. Toute autre hypothèse, une alternative à l'hypothèse nulle, concernant le paramètre, et les hypothèses viennent toujours par paires, permet un calcul alors du risque de deuxième espèce.

La quantité $(1-\alpha)$, fixée *a priori* s'appelle l'*efficacité* du test, la quantité $(1-\beta)$ s'appelle la *puissance* du test. On peut écrire directement :

$$\text{Efficacité :} \qquad 1-\alpha = \mathrm{P}[\text{ choisir } \mathrm{H}_0 \mid \mathrm{H}_0 \text{ vraie }]\,,$$

$$\text{Puissance :} \qquad 1-\beta = \mathrm{P}[\text{ choisir } \mathrm{H}_1 \mid \mathrm{H}_1 \text{ vraie }]\,.$$

Naturellement le calcul de la puissance suppose une valeur alternative précise du paramètre : β dépend explicitement d'une valeur précise de $\mu = \mu_1$. Certains théorèmes d'optimalité de la statistique permettent de construire des statistiques-tests qui, à α donné, ou efficacité donnée, minimisent en un certain sens que nous ne préciserons pas ici, le risque de seconde espèce, maximisent la puissance.

Ces idées concernant les décisions à prendre en présence de risque rejoignent bien sûr les décisions de la vie courante. On ne peut échapper aux risques : le risque d'une décision, le risque de ne pas la prendre… La statistique offre le luxe d'une optimisation. La vie courante ne fournit pas en général des tests optimaux…

Ce paradigme central en statistique a fini par s'établir au début des années trente, après de nombreux tâtonnements et controverses, notamment entre R.A. Fisher et les auteurs du paradigme Jerzy Neyman et Egon Pearson. Quelques petits exercices de réflexion préciseront les choses pour ceux qui les feront. Ce sont des extensions élémentaires de ce qui vient d'être exposé.

1. Si σ est inconnu et qu'on remplace σ par l'écart type échantillonnal, comment déterminer un test de seuil α pour une moyenne ?

2. Si on veut tester par deux échantillons, de tailles différentes ou non, une hypothèse sur deux moyennes $\mathrm{H}_0 : \mu_1 = \mu_2$, on peut se ramener aux cas déjà developpés en utilisant une différence de moyennes échantillonnales comme statistique pour tester l'hypothèse transformée mais équivalente : $\mathrm{H}_0 : \mu_1 - \mu_2 = 0$. Montrer comment les équations du théorème 1.4 permettent le même développement que pour celui d'un test sur une seule moyenne. Utiliser votre développement pour tester l'hypothèse que deux procédés de fabrication de boulons donnent les mêmes diamètres, lorsque $n_1 = n_2 = 10, \hat{X}_1 = 4.21\mathrm{mm}$ et $\hat{x_2} = 4.51\mathrm{mm}$. On admettra dans ce problème que les écarts types sont connus et égaux dans les deux cas à $\sigma = .5$ mm.

3. Y a-t-il une difficulté à développer un raisonnement analogue lorsque les variances sont inconnues ? (Se reporter éventuellement dans ce cas à un manuel complet sur les statistiques élémentaires.)

4. On a considéré ici le cas d'une hypothèse alternative dite *b*ilatère H:$\mu \neq \mu_0$. Si celle-ci est plutôt unilatère du type H_1: $\mu \geq 0$, par exemple, il est possible d'étendre sans peine ce qu'on vient de faire à ce cas.

2.4 Les lois de Fisher-Snedecor, et les tests de variance

Une application importante de la statistique est la comparaison de plusieurs groupes, divers traitements, divers procédés industriels par exemple. On utilise alors l'analyse de la variance. Celle-ci repose en grande partie sur des lois échantillonnales dites de Fisher-Snedecor, ou de Fisher plus simplement, du nom du scientifique anglais Sir Ronald Aylmer Fisher (1890-1962) qui a essentiellement fondé cette discipline.

Un quotient de deux lois χ^2 divisées chacune par leur degré de liberté, disons n_1 et n_2 suivent une autre de ces lois échantillonnales utiles, une loi F_{n_1,n_2}. En fonction de ces deux degrés de liberté, on trouve des tables des principaux quantiles utiles (ne pas oublier l'idiosyncrasie usuelle dans ce cas-ci aussi) : .1, .05, *etc.*. Ainsi donc, en se reportant à l'équation 2.4, on a

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2\,\sigma_2^2}{S_2^2\,\sigma_1^2} \;\sim\; \mathrm{F}_{n_1-1,n_2-1}\,.$$

Et donc si on a l'hypothèse nulle : $\mathrm{H}_0\ :\sigma_1^2 = \sigma_2^2$, le quotient $S_1^2\,/S_2^2$ suit une loi F_{n_1-1,n_2-1}, où n_1 et n_2 sont respectivement les tailles échantillonnales qui ont servi à calculer les réalisations de S_1^2 et S_2^2.

Remarquons que seuls les quantiles pour des α petits sont nécessaires, puisque l'inverse d'une loi F_{n_1,n_2} est une loi F_{n_2,n_1} (noter la permutation des degrés de liberté). Ainsi

$$1-\alpha = \mathrm{P}[\,F_{n_1,n_2}\;\leq\;F_{n1,n_2,\alpha}\,] = \mathrm{P}[\,\frac{1}{F_{n_1,n_2}}\;\geq\;\frac{1}{F_{n_1,n_2,\alpha}}\,]\,,$$

ce qui implique par définition pour l'inverse d'une loi F que :

$$\mathrm{P}[\,F_{n_2,n_1}\;\geq\;\frac{1}{F_{n_1,n_2,\alpha}}\,] = 1-\alpha\,.$$

Et donc par définition de $F_{n_2,n_1,1-\alpha}$:

$$F_{n_2,n_1,1-\alpha} = \frac{1}{F_{n_1,n_2,\alpha}}\,.$$

En pratique, le test bilatère sur deux variances ne se fait pas. On teste plutôt :

$$\mathrm{H}_0\ :\sigma_1^2 = \sigma_2^2 \qquad \text{versus} \qquad \mathrm{H}_1\ :\sigma_1^2\;\geq\;\sigma_2^2\,,$$

où, conventionnellement, on dénote par σ_1 la valeur exacte de la plus grande des deux variances échantillonnales.

On ne peut développer ici les applications de cette loi, à la régression notamment, en fait à presque toutes les grandes analyses statistiques en usage. Contentons-nous d'indiquer que le même paradigme de test s'applique encore.

On se donne une risque de première espèce α, puis on construit une statistique-test, on calcule sa loi, et une réalisation sur un échantillon nous indique la rareté de l'événement, et donc permet de décider en faveur ou non de l'hypothèse nulle.

Nous devons indiquer pour terminer comment lire un listage d'ordinateur où les tests sont rapportés.

2.5 Lecture des listages d'ordinateur pour les tests

Les valeurs échantillonnales des estimateurs sont rapportés dans les listages des logiciels de statistique. Ce qui demande interprétation, ce sont les conclusions des tests. Pour des raisons évidentes un ordinateur ne remplace pas l'expérimentateur et ne peut fixer à sa place un seuil α de rejet. La seule chose que tout logiciel se contente d'indiquer, c'est la rareté d'une réalisation d'une statistique-test, ou sa *probabilité de dépassement*[2].

Ainsi, par exemple, pour l'hypothèse qu'une moyenne est nulle, un logiciel se contente de donner

$$\mathrm{P}[\,|\,Z\,|\;\geq\;z_0\,] \equiv p\,,$$

[2] que le monde anglophone appelle le *p-value*.

puisque sous l'hypothèse nulle $\overline{X} \sim \mathcal{N}(0\,;1)$. Dans ce cas, z_0 est la valeur de la statistique calculée sur l'échantillon. L'expérimentateur doit éventuellement créer de nouvelles variables pour tester une valeur quelconque de μ_0 (par translation d'une variable échantillonnale donnée par exemple), pour utiliser le logiciel qui teste toujours l'hypothèse sur la moyenne $\mathrm{H}_0 : \mu = 0$. Dans la plupart des cas, surtout ceux reliés à l'analyse de la variance, le logiciel donne le résultat interprétable directement : si la probabilité de dépassement est inférieure au seuil critique fixée par l'expérimentateur, celui-ci peut ne pas accepter l'hypothèse nulle :

$$\text{on rejette } \mathrm{H}_0 \text{ si la probabilité de dépassement est} \leq \alpha\,.$$

Dans le cas unilatère on rejette H_0 si la probabilité de dépassement est inférieure à $2\,\alpha$, *i.e.* si $\frac{p}{2} \leq \alpha$.

À cet égard, il convient de mentionner que $\alpha = .05$ joue un rôle quasi magique dans beaucoup d'applications. Obtenir une probabilité de dépassement inférieure donc au .05 magique, signifie qu'on a moins qu'une chance sur vingt d'avoir un tel résultat pour la statistique échantillonnale (quelle qu'elle soit, d'ailleurs). Cela est approximativement aussi improbable que réussir 4 piles (ou faces) de file, ou qu'avoir deux fois 1 ou deux fois 2 avec une paire de dés. Cela est rare on en convient, mais ce nombre n'a rien de magique. C'est pourtant le α quasi universel en sciences humaines. Un $\alpha = .07$ suffit à ne pas rejeter une hypothèse et de .04 à la rejeter… Il convient de développer un sain scepticisme donc devant les résultats de beaucoup de travaux rapportés dans la littérature scientifique. Il est facile d' “oublier” des observations embarrassantes afin d'obtenir, en toute bonne conscience cela va sans dire, des résultats significatifs. D'autant plus facile qu'on trouve à la clé de nombreuses recherches significatives des sommes fort importantes pour la poursuite de travaux pas seulement d'un seul chercheurs mais souvent de toute une équipe. Il faudrait entrer ici dans l'aspect socio-politique de l'institution scientifique… De nombreux scandales récents ont jeté une lumière plutôt crue sur de vastes pans de la recherche scientifique (Blanc *et al.*, 1980 ; Boursin, 1978 & 1990 ; Huff, 1954 ; Wang, 1993 ; Wheeler, 1976). Remarquons toutefois que cela vient renforcer la position de la statistique puisqu'elle seule peut donner une respectabilité aux sciences appliquées et humaines : presque tous les fraudeurs ont dû trafiquer des résultats pour donner une caution à leurs fraudes, et que tout cela est maquillé en langage statistique (Senn, 2005, chapitre 1 parmi bien d'autres).

2.6 Aproximations pour les lois échantillonnales

Il peut être commode de programmer dans une calculette des approximations des lois échantillonnales les plus utiles (Zelen & Severo, 1972). Cela évite de les chercher dans des tables. Les approximations suivantes sont valables à au moins une décimale, ce qui est bien suffisant dans la plupart des applications.

- Pour la loi gaussienne :

$$\Phi(z) = \left[\,1 + exp(\,-1.5976z(\,1 + 0,044417z^2)\,)\,\right]^{-1}\,.$$

- Pour la loi gaussienne centrée réduite inverse : $z_\alpha = \mathit{\Phi}^{-1}(\,1{-}\alpha\,)$. Soit maintenant $\alpha \in [\,\frac{1}{2}\,;\,1\,]$, alors :

$$\mathit{\Phi}^{-1}(\alpha) = c - \frac{2,30753 + 0,27061\,c}{1 + 0,99229\,c + \,0,04481\,c^2}\,,$$

où

$$c = \sqrt{-2\,\log(1-\alpha)}\,.$$

- Loi T_n[3] :

$$\mathrm{t}_{n,\alpha} = \frac{z_\alpha}{1 - \frac{z_\alpha^2+1}{4\,n}}\,.$$

- Pour la loi χ^2 (approximation de Wilford-Hilferty) :

$$\chi^2_{n,\alpha} = n\left[1 - \frac{2}{9\,n} + z_\alpha\sqrt{\frac{2}{9\,n}}\right]^3\,.$$

- Pour la loi F_{n_1,n_2} :

$$\mathrm{F}_{n_1,n_2,\alpha} = \exp\left(\frac{1}{n_2} - \frac{1}{n_1} + z_\alpha\sqrt{\frac{2}{n_1} + \frac{2}{n_2}}\right)\,.$$

[3]Cette approximation est due au professeur Michel Moran de l'École Polytechnique de Montréal.

3. La vérité et la statistique

It is not that figures lie, but that liars sometimes figure.
Maurice G. Kendall [1907 – 1983][1]

On a pu écrire[2] comme le rapporte Vessereau (p.5) qu'il y a "trois formes de mensonge qui sont, par ordre de gravité, le mensonge ordinaire, le parjure, le mensonge par la statistique". L'attitude commune face à la statistique en est une en effet de scepticisme voire carrément de la plus grande méfiance. Sauf, évidemment lorsqu'on se sert des chiffres qui nous conviennent pour confirmer nos propres croyances ou simples opinions. On se vêt alors de la caution scientifique sans broncher. Quand on considère les tentatives de manipulation publique qui utilisent la caution scientifique, tout particulièrement la statistique, on est fortement souvent irrité, surtout en ayant des connaissances dans la discipline, et tenté de souscrire à ces jugements négatifs la concernant...

Plus de la moitié des étudiants universitaires québécois francophones croient en l'astrologie, à la communication par la pensée, *etc.*, plus, que toutes ces pseudo-sciences reposent sur de grandes études scientifiques (statistiques, naturellement). Jusqu'où peut-on aller dans l'ignorance[3]?

Nous faisons tous à vrai dire de la statistique (de cuisine) sans le savoir. Nous fondons une grande partie de nos jugements sur des fréquences particulières qu'on croit avoir observé... Qui n'a pas utilisé le proverbe "jamais deux sans trois"? Ce type de jugement fondé sur l'observation est le plus courant de nos vies de tous les jours. La statistique apporte ainsi souvent la force de l'évidence, vient conforter des hypothèses plus ou moins formulées...

Dans notre société largement illétrée, vulnérable donc à toutes les manipulations, on utilise bien souvent le modèle scientifique pour asservir la pensée individuelle. La massification des médias à laquelle on assiste maintenant ne vient que renforcer ces tendances. Les personnes mêmes les plus instruites, n'ont souvent lu aucun livre de réflexion sur quelque sujet que ce soit hors leur champ de spécialisation. Leur source quasi unique d'information est le journal télévisé, écouté d'un oreille distraite, vu la

[1] Important statisticien britannique, *cf.* http://en.wikipedia.org/wiki/Maurice_Kendall.

[2] À relire ce pamphlet [août 2006], il apparaît qu'il faudrait lui apporter beaucoup de nuances, même si l'ensemble reste encore, je dirais hélas, trop juste dans son ensemble.

[3] À noter que les statistiques sont utilisées la plupart du temps comme arguments de preuves... Même si l'on cherche à rouler dans la farine, même comme écran de fumée. On reconnaît par là l'importance primordiale, essentielle pour le développement de la connaissance, d'icelles dans la monde moderne... (Voir Senn, 2003). On pourrait montrer que les développements de la connaissance et ceux de la statistique sont dans une relation de symbiose (toujours Senn, 2003).

hâte obligée à faire tout chose. La classe instruite est sur le même pied que le béotien moyen dès qu'elle se trouve hors son champ spécialisé[4].

Cette section est un essai, un pamphlet en fait, où on expose les raisons qui nous font nous méfier des jugements soutenus par la statistique. Et nous oserons suggérer quelques correctifs à cette situation.

Recouvrement. Il n'y a pas aujourd'hui beaucoup de champs de connaissances où on ne doive utiliser des données quantitatives pour justifier ses assertions, opinions, idées avancées. Impossible quasiment de se faire entendre dans le moindre secteur des sciences humaines, *a fortiori* des sciences exactes ou appliquées, sans un attirail quantitatif qui échappe en grande partie à ceux qui le pratiquent. Notre société regorge de ces idiots savants qui savent presser sans discernement les bons boutons pour étayer leurs *a priori*. Et puis une grande part, sinon toutes les prises de décision individuelles se font en présence de prétendus experts dont les capacités à camoufler leur ignorance sous des oripeaux crédibles est leur plus grand talent.

La statistique est utilisée au premier chef pour en imposer, en jeter, influencer le cours des événements. On est bien loin de la science, bien loin de pourvoir trouver, même à l'université une "libre poursuite de la vérité", comme on définissait autrefois l'espace universitaire. La liberté aujourd'hui est en effet très souvent entravée par le dieu dollar, qui régente la moindre parcelle d'activité, même universitaire. L'université n'échappe plus pas plus aux impératifs commerciaux que le marchand de légumes du quartier.

De nombreux scandales sont venus récemment nous le rappeler. Le paragraphe suivant en exposera les causes institutionnelles. Nous évoquerons les autres. Nous terminerons notre essai par un examen d'une des principales applications de la statistique, les sondages. Nous parlerons de leurs effets pervers.

Nombreux scandales. Tout le monde a à l'esprit la sinistre affaire Fabrikant, du nom d'un chercheur d'origine étrangère poussé au meurtre par les moeurs locales en matières de recherche scientifique qui l'ont fait basculer dans la paranoïa. L'enquête a montré la justesse de ses plaintes ce qui a entraîné la démission d'une partie de la haute administration de l'université impliquée.

D'autre part, la recherche médicale canadienne, particulièrement la québécoise qui depuis des générations est vue comme la plus importante au Canada, est souillée par plusieurs scandales de tentatives de tromperie, dont une, particulièrement dramatique et jamais vraiment élucidée, a abouti au suicide de deux personnes. Cette fumeuse affaire a d'ailleurs commencé par une dénonciation anonyme à un journal anglophone de Montréal qui est plutôt un "journal de combat" en ce qui concerne les affaires québécoises. Quoi qu'il en soit, la fin de ce combat-là fut tragique!

Nous ne voulons pas entrer dans le fond de ces ténébreuses histoires, mais on sait par toutes sortes de révélations depuis quelques décennies, que la science médicale nord-américaine est particulièrement malade des tentatives de tricherie. Voilà une science qui s'est fortement quantifiée depuis vingt ans. Impossible maintenant de publier dans cette discipline sans utiliser l'appareil statistique; finies en grande partie donc

[4]On pourra consulter Milner (1984) à ce sujet (*De l'école,* Paris: Éditions du Seuil), de même que d'autres articles de cet auteur, notamment dans les journaux, qui débusque sans trêve les méfaits de cet 'illettrisme'. Voir aussi l'intéressante suite d'articles à ce sujet dans le journal Le Devoir [publié à Montréal, Qc] à l'été 2006.

les observations de quelques cas qui font l'histoire... Il suffit de parler au personnel statistique engagé pour combler les lacunes des chercheurs pour voir l'extension de la maladie. Les données utilisées finissent bien souvent après de longs détours par justifier (il faut absolument des $p < .05$) des rejets des hypothèses nulles. Et c'est ainsi, trop souvent hélas! qu'on fait progresser les connaissances...

Il ne faut pas croire cependant que les scandales soient récents. De tout temps on a vu des tentatives qui ont parfois nécessité des générations avant de se voir découvertes. On pourra consulter à ce sujet l'intéresssant article dans un numéro de la revue *La Recherche* (Blanc *et al.*, 1980) qui en rapporte les plus notoires. La gloire, le pouvoir sont des motivations qui ont agité les humains de tout temps. La science n'y échappe pas.

Mais l'institution (politique) de la recherche, tout particulièrement la nord-américaine, a ajouté aux motivations traditionnelles l'impératif de la survie. Les laboratoires reposent pour leur fonctionnement, de l'engagement du personnel à l'achat du moindre crayon, sur très peu d'individus pour l'approvisionnement en argent, ce nerf de la guerre. Pas d'argent pas d'équipement, pas de chercheurs. Et pas de publication, pas d'argent. On voit le cercle infernal se refermer sur les laboratoires. N'ayant pas, la plupart du temps, de structure permanente, ni même institutionnelle, il leur faut en permanence chercher les commanditaires pour assurer leur survie. Pis, c'est sur la tête de peu d'individus que reposent les responsabilités. Même à l'intérieur de la tour d'ivoire universitaire, où on a pu échapper à une certaine époque aux impératifs financiers par trop contraignants, il n'y a presque plus de reconnaissance (sauf de façade) due à l'enseignement, la mission éducative a été reléguée au dernier plan, et il faut 'battre' le voisin en méga-dollars de subventions et de contrats. L'université fonctionne comme une entreprise privée de recherche. L'argent corrompt tout ce qu'il touche. La cupidité a remplaçé la "libre recherche de la vérité", selon l'expression caduque consacrée.

Comment s'étonner que tout ce qui repose sur des données expérimentales non répétables et facilement trafiquables fasse l'objet de toutes les manipulations. Enfin n'oublions pas que les humains ont une facilité extraordinaire à trouver toutes les justifications nécessaires pour arriver à leurs fins. Avec, évidemment, une bonne conscience à la clé...

Naturellement les grands chercheurs ont une respectabilité à protéger, et peut-être une classe qui les met à l'abri des tentations. Mais peut-on demander la sainteté à tout venant, surtout aux jeunes chercheurs qui, après de très longues années d'études, cherchent à s'établir?

Nul doute en fait que les scandales qui arrivent sur la place publique ne sont que la pointe de l'iceberg . C'est l'institution même de la recherche qu'il faut modifier (Renaut, 2002). Mais, pour paraphraser Chateaubriand qui a parlé des nations pour qui cela s'applique très bien, les institutions passent longtemps sur leur lit de mort avant d'expirer[5].

[5] La situation décrite par Renaut (2002) s'applique plus particulièrement à la France, mais bien des éléments sont aisément transposables au Québec. À cet égrard, on s'interroge étonnamment peu chez nous. À croire que l'éducation n'est pas une valeur dans notre société.

Sondages: l'appel au conformisme

Le contexte

- Pour le public en général la principale utilisation des statistiques: les intervalles de confiance. Les diverses statistiques économiques sont farcies de nombres-indices dont personne ne connaît la signification exacte;
- Et les connaissances communes sont incapables d'en faire la moindre critique. On manipule les sentiments les plus bas de la population, la peur, la cupidité;
- Dans nos sociétés où même une réduction d'une fraction simple en pourcentage représente un calcul hors de portée même chez les plus instruits (30% des étudiants de Polytechnique sont incapables de réduire sans faute une expression un peu complexe comportant des fractions...).

Les sondages sont la manifestation la plus publique que connaissent les statistiques. Il deviennent omniprésents, au point qu'on a l'impresssion qu'aucune mesure gouvernementale ne se fait sans faire appel à quelque enquête pour savoir si on peut la faire accepter par avance par les populations concernées, autrement on l'abandonne sans même commencer à en parler. On parle de gouvernement par sondages, et certains se font les apôtres d'une nouvelle démocratie par le vote permanent. Voulez-vous rester au pouvoir? Eh bien, à l'intérieur de certaines limites, vous vous placez à la remorque du plus grand nombre et vous gagnez à coup sûr. La tyrannie de la majorité, car c'est bien de cela qu'il s'agit, fixe les opinions et les directions à prendre. C'est un bien grand signe de décadence que l'art de gouverner soit remis entre les mains de la masse. Le capitaine du vaisseau (gouverner vient du grec *kubernêtikè*, le terme pour gouvernail d'un bateau et qui a donné aussi cybernétique) est remplacé par le *vox populi*. Déjà Platon s'interrogeait dans *La république* sur la tyrannie du grand nombre. Aujourd'hui elle est érigée en vertu.

Mais plutôt que de voir les sondages uniquement comme un moyen pour savoir ce qui plaît au plus grand nombre pour s'en inspirer dans les directions à prendre, on peut aussi les voir comme d'un autre moyen pour asservir aux courants majoritaires les factions dissidentes. On voit les sondages construits sur les valeurs immédiates des populations sondées. On connaît bien les tendances au conformisme ("Tout le monde le fait, fais-le donc") des humains, c'est une manifestation de l'instinct grégaire. Les sondages destinés à la publication, ceux politiques notamment, sont trop souvent le fait de manipulateurs de l'opinion, deviennent des manipulations de masse. Les séquences de questions sont alors savamment étudiées par des batteries de psychologues, sociologues et autres gogues, pour en arriver à orienter l'opinion. On diffuse ensuite dans les journaux les résultats des principales questions blindés d'intervalles de confiance, et qui prennent de ce fait valeur de vérité absolue. On assortit le tout d'un paragraphe de méthodologie (taille des échantillons, *etc.*), et le tour est joué. En vérité on a actionné les principaux ressorts émotifs, les insécurités fort nombreuses et compréhensibles, en guise de mise en situation, avant d'en arriver aux questions qui vont faire l'objet de diffusions publiques.

S'il n'y avait que cela. Il est bien connu depuis longtemps que les réponses des sondés sont biaisées par leur désir de plaire, d'être conformes. Ainsi Huff (1954) rapporte un exemple saisissant de ce phénomène. Des intervievouveurs noirs ont obtenu de noirs interrogés que seulement 25% de ceux-ci croyaient que les Noirs seraient moins bien traités par les Japonais que par les Blancs américains advenant une victoire des premiers, alors que les enquêteurs blancs ont obtenu la proportion de 45% de la même population (nous sommes pendant la deuxième grande guerre). On répond pour être conforme à l'idée que l'enquêté se fait de l'opinion de l'enquêteur qui est le représentant type de la société: c'est *Big brother* qui parle par lui. Un des plus grands désirs des humains est de ne pas se différencier de la masse. On veut être conforme avant tout. La massification des médias et l'omniprésence de la télévision ont poussé au paraoxysme le phénomène. L'individualisme de la société nord-américaine dont on nous bassine tant, n'est-il pas plutôt un fait de conformisme? L'idéal pour tout commanditaire — des sondages en particulier —, est une population homogène, qu'il est d'autant plus facile de manipuler que la qualité de l'éducation est très basse. Les sentiments d'insécurité sont entretenus par les médias, la cupidité et la peur de manquer d'argent sont au cœur même de ces manipulations.

Sous prétexe d'un plus grande information concernant l'opinion publique, on ne fait que la rendre conforme aux idéaux des commanditaires[6].

En guise de conclusion: les fondements éducatifs

Le but de l'éducation est de rendre libre.

Denis Diderot [1713 – 1984][7]

Nous ne saurions trop insister: une bonne éducation, rigoureuse et exigeante, est une absolue nécessité dans le monde d'aujourd'hui avec la mondialisation à la clé — comme si cela n'avait pas toujours été le cas! —, mais aujourd'hui et dans une petite société comme la nôtre on ne peut se permettre de gaspiller nos ressources humaines.

Il n'est pas clair que que tous puissent avoir accès à une éducation supérieure. En fait, celle-ci, bien conçue et comprise, ne peut être accessible qu'à une minorité. La démocratisation de l'éducation supérieure ne peut se faire qu'au mépris de sa qualité (Jacquard *et al.*, 2003). Plus, pas d'*éducation d'élite* pour ceux qui peuvent y accéder, pas de progrès d'une société, ce qui n'est pas suffisant bien sûr. On parle aussi d'une éducation de qualité à tous les niveaux.

Le fait bien avéré aujourd'hui qu'environ 30% des écoliers sortent de l'école obligatoire québécoise (13 ans de scolarisation), après avoir usé leurs vêtements sur les bancs d'école pendant plus de 10000 heures, tout en restant essentiellement des illettrés[8] est un scandale inacceptable, un gaspillage intolérable dans les sociétés comme la nôtre.

[6]Voici bien un exemple d'affirmation où s'applique la note 3 de bas de page de cette section! En dépit de ce qui vient d'être écrit, il est néanmoins bien évident à quiconque s'intéresse à connaître la société, son évolution, que les sondages sont des outils absolument essentiels. En réalité, il n'est de connaissance *scientifique* que quantitative. Le quantitatif vient ainsi confirmer toute considération théorique, et nécessairement. Autrement, on se promène dans la boue idéologique, la manipulation, la foi religieuse — dans le mauvais sens du terme! Répétons-le: la méthode scientifique se confond avec la méthode statistique (Senn, 2003).

[7]Un des créateurs des Lumières: http://fr.wikipedia.org/wiki/Diderot.

[8]Un *analphabète* ne sait ni lire ni écrire; un *illettré* lit avec grande difficulté même les messages simples sur les contenants de médicaments, sait encore moins écrire logiquement.

Nous croyons, avec Diderot, que l'éducation est avant tout une condition de la liberté, c'est elle, presque elle seule, qui permet la pensée libre et à distance. La pensée quoi!

C'est ainsi que nous pensons qu'il convient que les études obligatoires des enfants soient exigeantes et rigoureuses, et:

- promeuvent la pensée personnelle à tous les échelons;
- donnent une mémoire aux écoliers qui va bien au-delà de leur naissance, ce qui est à la source première de leur identité collective comme personnelle[9];
- et se terminent par des cours de philosophie dont le premier but est la mise à distance, l'examen critique du monde, la présentation des grandes questions.

Nous aimerions bien aussi que passe le message suivant: la méthode scientifique et les technologies statistiques sont dans une relation de symbiose. Ces dernières sont le bras armé de la première. L'histoire du développement des connaissances depuis la Renaissance a clairement montré ce fait.

Il est possible d'introduire les premières notions de probabilité via l'analyse de données — et toute connaissance scientifique est là — très tôt dans le cursus scolaire: au secondaire sûrement, voire même au primaire.

La vie courante, elle aussi, est remplie de considérations statistiques: qu'on pense seulement aux nombres-indices, aux nombreux sondages qu'il importe de savoir comprendre, au moins en gros, de savoir les mettre à distance et souvent critiquer. Que tous les étudiants universitaires aient la possibilité de le faire paraît comme relevant du plus élémentaire bon sens.

L'introduction correcte aux technologies statistiques sont une absolue nécessité dans tout cursus universitaire, à l'exception des lettres et encore[10].

Les temps ne sont pas propices à l'honnêteté intellectuelle. Les institutions universitaires ne favorisent pas beaucoup la vie intellectuelle, on le constate aisément. Et les jeunes qui entrent enfin dans la système comprennent rapidement de quoi il retourne. C'est malheureux, mais c'est ainsi, et ce n'est pas demain la veille qu'on verra des changements à cet égard!

Quoi qu'il en soit, les technologies statistiques correctement et honnêtement utilisées sont les outils conceptuels les plus puissants. Il n'y en a pas beaucoup d'autres.

---

[9] On a assisté au Québec en cette année 2006 à cette tentative intolérable que l'histoire et son interprétation ferait éventuellement l'objet de décrets gouvernementaux... Bien des fonctionnaires ont semble-t-il obtempéré! N'eût été de la vigilance de certains journalistes (du Devoir pour ne pas le nommer), cela aurait passé comme une lettre à la poste! Quand on parle de pauvreté de l'éducation...

[10] On peut trouver cette affirmation très exagérée. Qu'on m'indique une discipline et je pourrai montrer où on y trouve assez rapidement des méthodes quantitatives! (Voir aussi Senn (2003) pour des détails.)

Références

[1] Peter L. Bernstein : *Against the gods. The remarkable story of risk.* John Wiley & Sons, New York NY, 1996.

[2] M. Blanc, G. Chapithier et A. Danchin : Les fraudes scientifiques. *La Recherche*, 11 : 858–868, 1980.

[3] Jean-Marie Bouroche et Gilbert Saporta : *L'Analyse des données.* Presses Universitaires de France, coll. Que sais-je ? n° 1854, Paris, 1980.

[4] Jean-Louis Boursin : *La forme scientifique du mensonge.* Tchou, Paris, 1976.

[5] Jean-Louis Boursin : *Les dés et les urnes.* Seuil, Paris, 1990.

[6] CERESTA : Aide-mémoire pratique des techniques statistiques. *Revue de Statistique appliquée*, 34, numéro spécial, 1986.

[7] Philippe Cibois : *L'analyse factorielle.* Presses Universitaires de France, coll. Que sais-je ? n° 2095, Paris, 1983.

[8] Jean-Jacques Droesbeke et Philippe Tassi : *Histoire de la statistique.* Presses Universitaires de France, coll. Que sais-je ? n° 2527, Paris, 1990.

[9] Darrell Huff : *How to lie with statistics.* W.W. Norton, New York, 1954.

[10] Albert Jacquard : *Les probabilités.* Presses Universitaires de France, coll. Que sais-je ? n° 571, Paris, 1974.

[11] Albert Jacquard, Pierre Manent et Alain Renaut : *Une éducation sans autorité ni sanction?* Éditions Bernard Grasset, Paris, 2003.

[12] Norman L. Johnson et Samuel Kotz : *Discrete distributions.* Houghton Mifflin, New York, 1969.

[13] Norman L. Johnson et Samuel Kotz : *Continuous univariate distributions, pt. 1.* Houghton Mifflin, New York, 1970.

[14] Norman L. Johnson et Samuel Kotz : *Continuous univariate distributions, pt. 2.* Houghton Mifflin, New York, 1970.

[15] Norman L. Johnson et Samuel Kotz : *Continuous multivariate distributions.* J. Wiley, New York, 1972.

[16] Alain Renaut : *Que faire des universités?* Bayard Éditions, Paris, 2002.

[17] Stephen Senn : *Dicing with death. Chance, risk and health.* Cambridge University Press, Cambridge UK, 2003.

[18] Michel Tenenhaus : *Méthodes statistiques en gestion.* Dunod, Paris, 1994.

[19] André Vessereau : *La statistique.* Presses Universitaires de France, coll. Que sais-je ? n° 281, Paris, 1947.

[20] Chamont Wang : *Sense and nonsense of statistical inference. Controversy, misuse, and subtlety.* Marcel Dekker Inc., New York, 1993.

[21] Michael Wheeler : *Lies, Damn Lies, and Statistics. The manipulation of public opinion in America.* Liveright, New York, 1976.

[22] Marvin Zelin et Norman C. Severo : Probability functions. *In* M. Abramowitz et Irene A. Stegun, éditeurs : *Handbook of Mathematical Functions*, chapitre 25. Dover Publications Inc., New York, 1972.