MODÉLISATION DE L'INCERTAIN - VARIABLES ALÉATOIRES

La notion de ***variable aléatoire*** (en abrégé: ***v.a***. ou va) formalise l'incertitude de situations ou une variable peut prendre différentes valeurs. Ainsi :

- le résultat du jet d'un dé ;
- le sexe d'un certain enfant à naître ;
- le rapport du prochain quinté dans l'ordre à Cagnes-sur-Mer (Alpes-Maritimes) ;
- le prénom du (de la) prochain(e) président(e) des États-Unis d'Amérique ;
- la température demain à midi au sommet de la tour Eiffel ;
- la durée d'un coup de téléphone ;
- l'indice Dow Jones le 24 octobre 2009 ;
- le PIB de la France l'année prochaine ;
- le nombre d'hectolitres de beaujolais AOC issu des vendanges 2010.

Sauf indication contraire, on s'intéressera ici à des va prenant des valeurs numériques. On distingue les va ***discrètes***, qui ne prennent que des valeurs isolées (en général, en nombre fini et petit), et les va ***continues***.

VA DISCRÈTES

Loi d'une va discrète

Une va discrète : **X**, est caractérisée par sa ***loi***, c'est à dire par la donnée des différentes valeurs possibles : $\mathbf{x_1}$, $\mathbf{x_2}$, ..., $\mathbf{x_k}$, et des probabilités associées : $\mathbf{p_1}$, $\mathbf{p_2}$, ..., $\mathbf{p_k}$.

On peut considérer que les probabilités p_i mesurent les « chances d'apparition » de ces différentes valeurs x_i.

En fait, on observe seulement des ***réalisations*** de la variable aléatoire, (la face supérieure du dé, le sexe de l'enfant, etc.); ces observations ont une valeur déterminée, et plus aucun caractère d'incertitude, mais, s'il est possible de faire un **grand nombre** d'observations suivant la même loi, les fréquences d'apparition des différentes valeurs approchent leurs probabilités.

Exemple : loi de la va qui formalise le lancer d'une pièce de monnaie :

$\mathbf{x_i}$	$\mathbf{p_i}$
Pile	0,5
Face	0,5

Et l'on sait que sur un grand nombre de lancers, environ une moitié donneront pile, et une moitié face.

Telles les fréquences, les probabilités des différentes valeurs sont comprises entre 0 et 1, et leur somme vaut 1.

Espérance, variance, écart-type

Aux caractéristiques de la distribution d'une série statistique numérique correspondent les caractéristiques d'une loi de va, définies en prenant les probabilités en place des fréquences.

On donne un tableau synoptique :

Série statistique x	**Va X**
Valeurs prises: x_i	Valeurs possibles: x_i
Fréquences: f_i	Probabilités: p_i
Moyenne: $\bar{x}$ $\bar{x} = \Sigma f_i.x_i$	Espérance: E(X) $E(X) = \Sigma p_i.x_i$
Variance: var(x) $var(x) = \Sigma f_i.(x_i - \bar{x})^2$	Variance: var(X) $var(X) = \Sigma p_i.[x_i - E(X)]^2$
Écart-type: σ(x) $\sigma(x) = [var(x)]^{0,5}$	Écart-type: σ(X) $\sigma(X) = [var(X)]^{0,5}$

Les probabilités étant les valeurs limites des fréquences pour une grande série d'essais, on voit que l'espérance d'une va est elle-même approchée par la moyenne d'une telle série, pour cette raison, l'espérance est fréquemment, et abusivement, appelée également moyenne; c'est un indicateur de ***tendance centrale*** d'une va.

De même la variance et l'écart-type sont approchés par les indicateurs de mêmes noms, calculés sur un grand nombre d'essais; ce sont des indicateurs de ***dispersion*** d'une va.

Propriétés

Les propriétés connues de la moyenne et de l'écart-type d'une série statistique se transposent à l'espérance et à l'écart-type d'une va.

L'espérance d'une somme de va est la somme de leurs espérances et plus généralement, si X et Y sont deux va et a et b deux constantes :

$$E(a.X + b.Y) = a.E(X) + b.E(Y)$$

Deux ou plusieurs va sont dites ***indépendantes*** si elles ne s'influencent pas, ainsi le jet d'un dé et le jet suivant, la position des astres le jour de sa naissance et les gains au loto d'un individu, mais non sa fortune personnelle et celle de ses ascendants.

La variance d'une somme de va **indépendantes** est la somme de leurs variances, et plus généralement, toujours sous l'hypothèse d'indépendance :

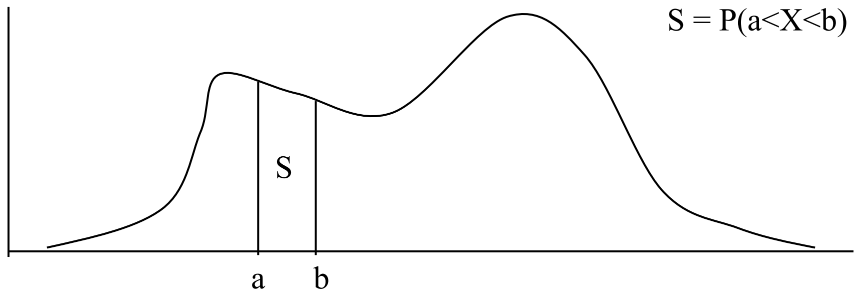
$$var(a.X + b.Y) = a^2.var(X) + b^2.var(Y)$$

VA CONTINUES

Certaines grandeurs aléatoires peuvent prendre l'infinité des valeurs d'un intervalle (la taille d'un homme, la durée d'un appel téléphonique, d'une vie, etc.), et la définition de probabilités élémentaires: p_i, n'est plus possible. Une telle va est dite continue, et sa loi est alors donnée par sa ***courbe de densité***.

La courbe de densité d'une va continue peut être considérée comme la courbe limite des diagrammes de fréquences convenablement calibrés des valeurs observées pour un nombre croissant d'essais classés en des intervalles de plus en plus étroits.

Exemple :



La signification graphique reste la même, et la surface d'une bande verticale mesure la probabilité que la va tombe dans l'intervalle.

L'espérance, la variance et l'écart-type se définissent d'une autre manière (il est nécessaire d'employer des *intégrales* en place des sommes : Σ), mais conservent la signification et les propriétés précédentes.

Il est commode de représenter les quantités aléatoires discrètes mais prenant un très grand nombre de valeurs très proches (le revenu, l'effectif d'une grande population, etc.) par des va continues.

LOIS NORMALES

Il est une famille de lois de va continues très importantes en statistique : les **lois *normales*** (ou **lois de *Gauss***), elles apparaissent en effet fréquemment et interviennent notamment dans la théorie de l'estimation.

Les lois de distribution normales, qui prennent leurs valeurs dans l'ensemble des nombres réels, dépendent de deux paramètres (l'espérance et l'écart-type) et se ramènent à la plus simple d'entre elles : la loi normale ***centrée-réduite***.

LOI NORMALE CENTRÉE-RÉDUITE : *N*(0,1)

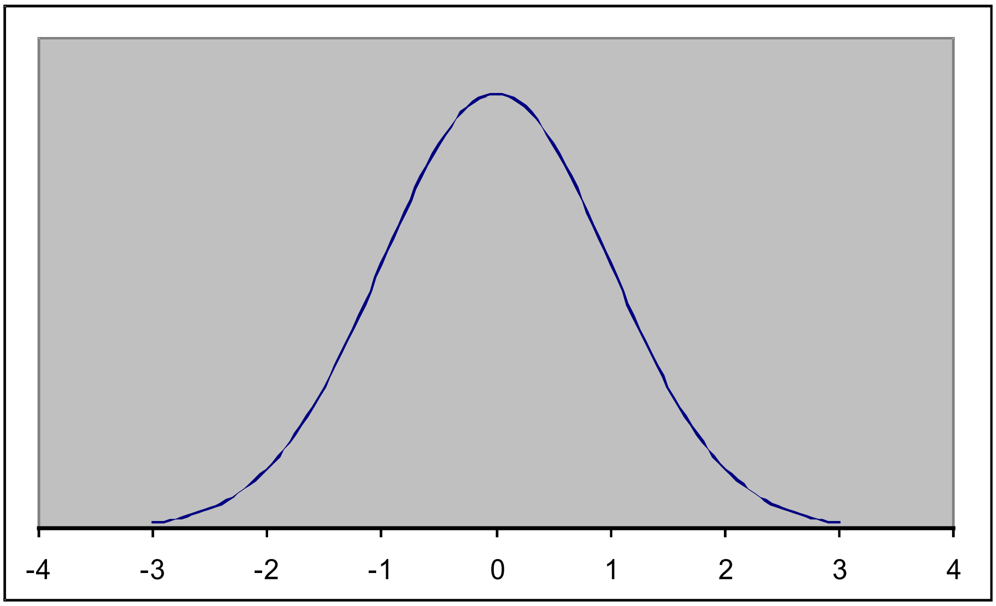
La loi normale centrée-réduite est celle d'espérance nulle et d'écart-type 1, notée : ***N*(0,1)**, elle peut prendre ses valeurs entre moins infini et plus infini, et elle est définie par sa courbe de densité symétrique, souvent appelée ***courbe en cloche*** (ou *courbe de Gauss*).

Pour la curiosité, on indique sa densité : f(x) = [1/racine(2.pi)].exp(-x^2/2).

Différentes tables permettent de calculer la probabilité qu'a une va de loi $\mathcal{N}(0,1)$ de tomber dans un intervalle donné.

A l'inverse, la table de l'***écart-réduit*** permet pour une probabilité donnée: **a** (ou ***niveau de risque***), de déterminer l'intervalle symétrique : $[-t_a, t_a]$, qui a la probabilité : **1-a**, (ou *niveau de **confiance***), de contenir la valeur observée d'une va $\mathcal{N}(0,1)$.

Exemple : pour a = 0,05 on lit $t_a = t_{0,05} = 1,96$ c'est-à-dire qu'une va $\mathcal{N}(0,1)$ a une probabilité de 0,95, ou « 95 chances sur 100 », de tomber entre -1,96 et +1,96. On dit que l'intervalle [-1,96; 1,96] est un ***intervalle de confiance*** de risque 5%, pour une $\mathcal{N}(0,1)$.



Courbe de densité de la loi de distribution $\mathcal{N}(0,1)$

On peut retenir la valeur approximative $t_{0,05} \approx 2$ et l'intervalle de confiance approché au risque 5% [-2; 2].

LOI NORMALE QUELCONQUE : $\mathcal{N}(\mu,\sigma)$

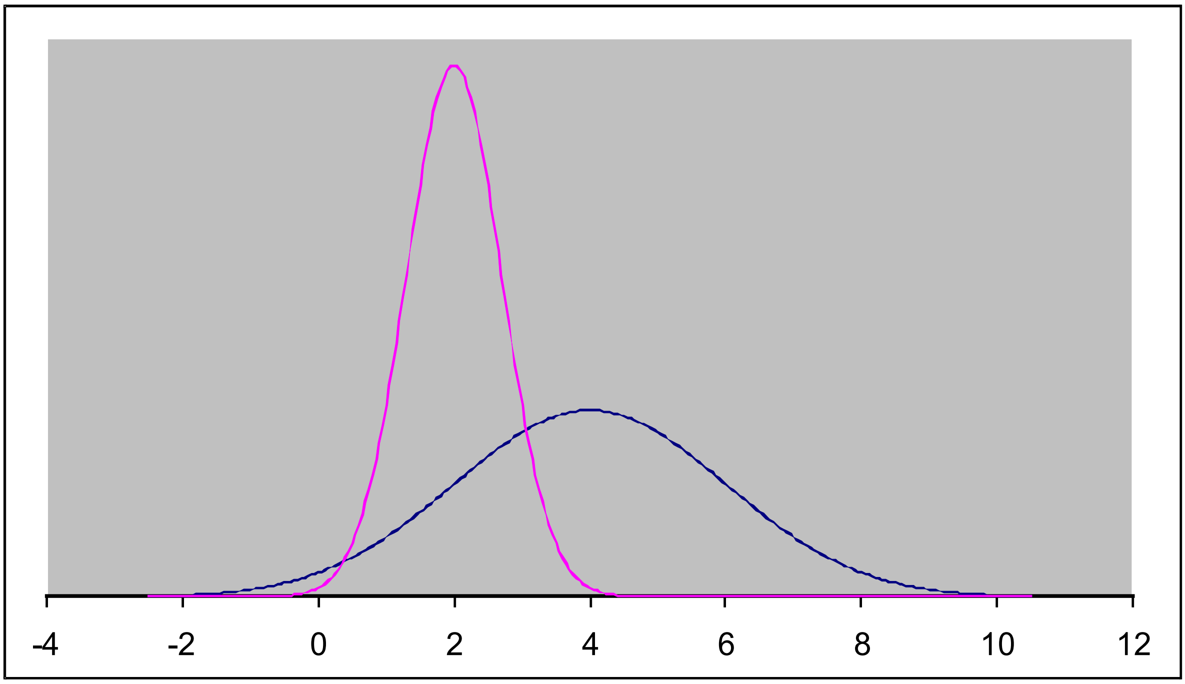
La loi normale $\mathcal{N}(\mu,\sigma)$ est celle d'espérance : μ, et d'écart-type : σ, elle se déduit de la loi normale centrée-réduite par un changement d'origine et un changement d'échelle; exactement, la va: X, suit la loi $\mathcal{N}(\mu,\sigma)$, si l'écart à μ, mesuré dans l'unité: σ, c'est-à-dire la va : $Y=(X-\mu)/\sigma$ suit la loi $\mathcal{N}(0,1)$.

La courbe en cloche de la densité est centrée sur μ et non plus sur 0, et la cloche est plus ou moins évasée suivant la dispersion : σ.

La construction d'un intervalle de confiance pour une va $\mathcal{N}(\mu,\sigma)$ se déduit du cas d'une $\mathcal{N}(0,1)$ par le changement d'origine et d'échelle convenable; au niveau de risque: a, la valeur observée d'une $\mathcal{N}(\mu,\sigma)$ appartient, c'est à dire tombera dans l'intervalle :

$$[\mu - t_a.\sigma \,;\, \mu + t_a.\sigma].$$

<u>Exemple</u> : on suppose que la durée de vie en heures d'un certain type d'appareil électronique suit une loi normale $\mathcal{N}(10000,2000)$. Un appareil donné a alors une probabilité 0,95 de durer entre $10000 - t_{0,05}.2000$ heures et $10000 + t_{0,05}.2000$ heures, c'est-à-dire de tomber en panne entre 6000 et 14000 heures de fonctionnement, en utilisant l'approximation $t_{0,05} \approx 2$.



Courbes de densité des lois $\mathcal{N}(2 ; 0{,}7)$ et $\mathcal{N}(4 ; 2)$

PROPRIÉTÉ FONDAMENTALE

Une propriété fondamentale explique l'importance des lois normales : la somme d'un grand nombre de quantités aléatoires **indépendantes** et dont aucune n'est d'échelle prépondérante, suit approximativement une distribution normale (cet énoncé « littéraire », connu sous le nom de ***théorème central-limite*** ou de ***loi forte des grands nombres***, admet naturellement des formulations mathématiquement plus rigoureuses !).

Cela explique pourquoi beaucoup de variables, qui peuvent être considérées comme la résultante additive d'un grand nombre de déterminants dont aucun n'est dominant, montrent des distributions normales (ainsi la taille des Françaises, le poids exact des paquets de 1 kg de haricots secs, les erreurs d'observations astronomiques, etc., mais non la taille des Français des deux sexes, ce dernier caractère ayant une influence prépondérante).

C'est pour cette raison que, d'une manière quelque peu abusive, on a baptisé cette loi : « normale ».

Malgré l'importance des lois normales, on gardera à l'esprit qu'il existe bien d'autres lois de distribution. Bien que celles-ci soient répandues, une grandeur aléatoire ne suit pas nécessairement une loi normale; une telle hypothèse doit en principe être confirmée (ou infirmée) par l'examen d'observations...

L'ESTIMATION - DEUX EXEMPLES - PROPRIÉTÉS

L'*estimation* est l'un des buts de la statistique; l'idée est de relier une valeur **non aléatoire, mais inconnue**, à une quantité **aléatoire, mais observable**, pour en déduire une approximation, ou ***estimation***, de la valeur inconnue, en ayant, autant que possible, une idée de sa précision et de la confiance à lui accorder.

On donne ci-après deux exemples d'estimation fondés sur la *moyenne d'échantillon*.

ÉCHANTILLON

Une certaine grandeur aléatoire: X (le poids d'une lettre, le solde d'un compte en banque, la durée de vie d'un certain équipement, d'un être humain, etc.), suit une loi d'espérance: μ, et d'écart-type: σ. Un ***échantillon de taille N*** est une série de N va : $x_1, x_2, \ldots, x_N$, **indépendantes** suivant cette même loi.

Comme pour une va unique, l'observation d'un échantillon est l'observation des réalisations des N va composant l'échantillon.

Moyenne d'échantillon

La ***moyenne d'échantillon*** est la moyenne : $\bar{x}$, de la série des va considérées : $\bar{x} = \Sigma x_i / N$, c'est naturellement encore une va, dont on observe la réalisation.

On montre que la moyenne d'échantillon a les caractéristiques suivantes :

$$E(\bar{x}) = \mu, \quad var(\bar{x}) = \sigma^2/N \quad \text{et} \quad \text{écart-type}(\bar{x}) = \sigma/\sqrt{N}$$

Enfin, d'après la propriété fondamentale, elle suit approximativement une loi normale, qui est donc la loi $\boldsymbol{N}(\mu,\sigma/\sqrt{N})$.

Cette approximation est en général satisfaisante dès que N atteint quelques dizaines, et elle ne suppose pas que la loi initiale soit normale.

ESTIMATION D'UNE ESPÉRANCE INCONNUE, INTERVALLE DE CONFIANCE

On veut estimer l'espérance inconnue : μ, d'une certaine grandeur aléatoire : X (ou, ce qui revient au même, sa moyenne sur une population très grande, considérée comme infinie). L'écart-type de la distribution, en général également inconnu, est noté s.

On prélève un échantillon de taille N suffisante, et on détermine sa moyenne $\bar{x}$ Cette moyenne est une estimation *ponctuelle* de la valeur μ, mais il faut encore pouvoir indiquer la précision et la confiance à accorder à cette estimation.

Ayant pour espérance la valeur μ qu'elle veut estimer, la moyenne d'échantillon en est un estimateur *sans biais* (cf infra). Suivant d'autre part approximativement une distribution normale $\boldsymbol{N}(\mu,\sigma/\sqrt{N})$, sa valeur observée appartient à l'intervalle :

$[\mu - t_a.\ \sigma/\sqrt{N}\ ;\ \mu + t_a.\ \sigma/\sqrt{N}\ ]$ au risque a

```
        ←---tₐ. σ/√N ---→ ←---tₐ. σ/√N ---→
--------------------------μ-----------x̄ ---------------------->
                          ←---tₐ. σ/√N ---→ ←---tₐ. σ/√N ---→
```

En fait c'est $\bar{x}$ qui est observé et connu, et μ qui ne l'est pas, mais il est équivalent de dire que l'inconnue μ appartient à l'intervalle :

$[\bar{x}\ \mathbf{-\ t_a.}\ \sigma/\sqrt{N}\ \mathbf{;}\ \bar{x}\ \mathbf{+\ t_a.}\ \sigma/\sqrt{N}\ \mathbf{]}$ au risque **a**

On montre qu'il est acceptable, pour un échantillon de taille suffisante, de remplacer la valeur inconnue de l'écart-type σ par son estimation, calculée sur l'échantillon, et on obtient ainsi une estimation de l'espérance inconnue μ par un *intervalle de confiance* de risque a encadrant la moyenne d'échantillon.

Exemple : 15000 personnes ont passé un concours. 600 premières copies prises au hasard sont déjà corrigées, on a trouvé une moyenne de 11,3 et un écart-type de 2,1. On veut estimer la moyenne générale avec un risque de 5%.
Avec $t_{0,05} \approx 2$ on obtient l'intervalle de confiance : $[11,3 - 2.2,1/\sqrt{600}\ ;\ 11,3 + 2.2,1/\sqrt{600}\ ]$, et on peut affirmer que la moyenne générale est entre 11,13 et 11,47 avec 5% de risque d'erreur.

La présence du dénominateur : $\sqrt{N}$, montre, par exemple, que pour une estimation 10 fois plus précise (à un niveau de confiance donné), il faut prendre un échantillon 100 fois plus important.

Dans le cas d'une grandeur de distribution supposée normale, la méthode reste pertinente avec un échantillon de taille réduite, mais on utilise alors pour construire l'intervalle de confiance non la loi normale, mais une loi dérivée et approchée: la *loi de **Student***, destinée à prendre en compte l'imprécision introduite par l'emploi de l'écart-type empirique en place de l'écart-type vrai mais inconnu.

Remarque importante : on retiendra impérativement que l'estimation par intervalle de confiance donne une réponse en terme de précision (par l'intervalle proposé), mais aussi de risque (par la valeur de a choisie). Ainsi un statisticien qui fait par exemple des estimations au risque 10% dans des conditions correctes doit s'attendre à obtenir une « fourchette » erronée, c'est à dire ne contenant pas la valeur estimée, environ une fois sur dix.

ESTIMATION D'UNE FRÉQUENCE INCONNUE - SONDAGE

La méthode précédente permet aussi d'estimer une fréquence dans une population très nombreuse (par exemple la proportion des électeurs qui vont voter pour un certain candidat).

Un échantillon est alors plutôt appelé ***sondage***. On note : p, la fréquence inconnue de la caractéristique étudiée dans la population totale, et : f, sa fréquence observée dans le sondage.

f est naturellement une estimation ponctuelle de p, mais on veut construire un intervalle de confiance.

Soit la va qui note par 1 ou 0, la présence ou l'absence de la propriété, elle suit une loi:

Valeurs	**Probabilités**
0	1-p
1	p

d'espérance : p, et dont on montre que l'écart-type vaut : $\sqrt{p.(1-p)}$.

Cette va sert à compter les observations dotées de la propriété, et la moyenne d'échantillon apparaît donc comme la fréquence observée f.

Pour un effectif N suffisant, f suit donc approximativement la loi N(p, $\sqrt{p.(1-p)/N}$), et on montre à nouveau que pour l'écart-type : $\sqrt{p.(1-p)}$, on peut remplacer la valeur p inconnue par son estimation ponctuelle : f.

On obtient donc l'intervalle de confiance pour la valeur de p **:**

[f - t_a. $\sqrt{f.(1-f)/N}$ **; f + t_a.** $\sqrt{f.(1-f)/N}$ **)]** au risque: **a**.

Ces approximations sont mauvaises, et cette méthode est à déconseiller, si la valeur à estimer est proche de 0 ou de 1 - ce qui se perçoit en général sur l'échantillon.

<u>Exemple</u> : 850 électeurs ont été interrogés et 441 ont déclaré vouloir voter pour le candidat X. La fréquence dans le sondage du vote pour X est : f = 441/850 ≈ 51,9%, et au risque 5% le vote en sa faveur sera dans l'intervalle :

[51,9 - 2. $\sqrt{51{,}9.48{,}1/850}$; 51,9 + 2. $\sqrt{51{,}9.48{,}1/850}$]

c'est-à-dire qu'il aura entre 48,5% et 55,3% des voix.

Dans le cas des sondages électoraux, on construit plus fréquemment des intervalles de confiance asymétriques (tests *unilatéraux*), les candidats étant surtout préoccupés d'obtenir un score estimé situé dans un intervalle totalement à droite de 50%...

QUALITÉS DES ESTIMATEURS

On appelle généralement ***estimateur*** la va observable utilisée pour une certaine estimation, et *estimation*, sa réalisation, observée sur l'échantillon considéré.

On a vu qu'un estimateur est une va observable destinée à approcher une valeur inconnue. Ses propriétés et qualités dépendent de sa loi de distribution, elle-même liée au modèle initial retenu (et supposé correct !). On présente les plus importantes.

On considère un estimateur A d'une quantité inconnue a, associé à un échantillon de taille N, ou, pour les propriétés asymptotiques, une suite A_N d'estimateurs associés à des échantillons de taille croissante (telles, par exemple, les moyennes, encore dites ***moyennes empiriques***, calculées sur les échantillons successifs).

La première qualité d'un estimateur est liée à son espérance :

- L'estimateur A de a est ***sans biais*** si son espérance E(A) est précisément égale à la quantité a estimée (il est dit *biaisé* dans le cas contraire).

Les qualités suivantes concernent la précision de l'estimateur :

- L'estimateur A de a est ***efficace*** s'il est sans biais et de variance minimale pour la taille d'échantillon considérée (s'ils sont tous deux non biaisés, le *plus efficace* de deux estimateurs est celui de plus petite variance).

S'il est vrai qu'on s'intéresse généralement d'abord à l'espérance, et au biais éventuel, d'un estimateur, et ensuite seulement à sa précision, un estimateur légèrement biaisé mais de faible variance peut être préférable à un estimateur sans biais mais beaucoup moins précis.

- Certains estimateurs biaisés pouvant ainsi être néanmoins intéressants, on définit l'*erreur quadratique moyenne* (*mean square error* ou *MSE* pour les Anglo-Saxons) comme l'espérance du carré de l'erreur d'estimation : $E[(A - a)^2]$. Un estimateur est ***optimal en moyenne quadratique*** s'il minimise cette quantité.

Les dernières propriétés sont asymptotiques et traduisent la qualité des estimateurs lorsque la taille de l'échantillon croit vers l'infini ; la convergence d'une suite d'estimateurs peut être définie de différentes manières.

- l'estimateur A_N de a est ***asymptotiquement sans biais*** si son espérance $E(A_N)$ tend vers a lorsque la taille N l'échantillon croit; on dit aussi que l'*espérance asymptotique* de cet estimateur est égale à la quantité estimée ;

- l'estimateur A_N de a est ***consistant*** s'il est asymptotiquement sans biais et si sa variance $var(A_N)$ tend vers zéro lorsque la taille N l'échantillon croit indéfiniment (il est équivalent de dire que l'erreur quadratique moyenne tend vers zéro).

La convergence d'une suite d'estimateurs A_N peut enfin être définie de plusieurs manières:

- la suite d'estimateurs A_N ***converge en probabilité*** vers la valeur a, si les valeurs « se concentrent » autour de a lorsque N croit (plus précisément et rigoureusement, si pour ε arbitrairement petit : $\lim P(|A_N - a| > \varepsilon) = 0$). On omet fréquemment le qualificatif « en probabilité » en parlant simplement d'*estimateur* ***convergent*** pour désigner cette propriété de la suite considérée ;

- la suite d'estimateurs ou de va dérivées: B_N, ***converge en loi*** vers une loi donnée s'il y a convergence en chaque valeur des fonctions de répartitions (c'est à dire si $\lim P(B_N \leq t) = F(t)$, F désignant la fonction de répartition de la loi limite retenue). On parle ainsi de ***distribution limite*** d'un estimateur ou d'une quantité dérivée (souvent normale, comme dans les exemples préliminaires), et on utilise ses propriétés dès qu'on juge l'échantillon assez grand.

Sans être équivalentes, les différentes notions précédentes entretiennent naturellement diverses relations. On montre ainsi qu'un estimateur consistant converge en probabilité vers son espérance asymptotique, tandis la convergence en loi vers une loi *dégénérée*, c'est à dire concentrée en un seul point, implique la convergence en probabilité vers cette valeur.

Exemple : la moyenne empirique $\bar{x}$ est un estimateur sans biais de l'espérance μ, sa variance : σ^2/N, tendant vers zéro lorsque la taille de l'échantillon s'accroît, l'estimation est donc consistante et par suite convergente, on montre enfin que la quantité : $\sqrt{N}.(\bar{x} - \mu)/\sigma'$, suit asymptotiquement une loi normale N(0,1), σ' désignant l'écart-type estimé usuel.

Les propriétés envisagées, ou une partie d'entre elles, ne définissent généralement pas un estimateur unique, ainsi dans l'exemple précédent, l'estimateur : $\bar{x}' = \bar{x}.[N/(N-1)]$, de l'espérance μ inconnue est asymptotiquement sans biais et possède toutes les propriétés précédentes sauf la première.

MÉTHODE OU ESTIMATION DU MAXIMUM DE VRAISEMBLANCE

Lorsque face à un problème d'estimation on ne dispose pas a priori d'un estimateur « naturel » (telle la moyenne empirique pour l'espérance), on a fréquemment recours à une technique puissante, celle du ***maximum de vraisemblance*** (*maximum-likelihood* pour les Anglo-Saxons).

Vraisemblance d'un échantillon

Soit une loi de distribution à densité f(x), la densité pour une observation x_i est $f(x_i)$ et pour un N-échantillon : $(x_1, x_2,..., x_N)$, c'est : $f(x_1).f(x_2). ... .f(x_N)$, ou $\Pi\ f(x_i)$, cette quantité est appelée ***vraisemblance*** de l'échantillon et généralement notée : $L(x_1, x_2,..., x_N)$.

Maximum de vraisemblance

Si la loi dépend d'un paramètre inconnu θ (éventuellement multidimensionnel), on note $f_\theta(x)$ la densité, et la vraisemblance de l'échantillon est $L = \Pi\ f_\theta(x_i)$.

L'***estimateur du maximum de vraisemblance*** du paramètre inconnu θ est la valeur de celui-ci qui maximise la vraisemblance de l'échantillon observé, ou en d'autres termes, celui pour lequel cet échantillon observé est le plus probable.

Fondés sur une idée intuitive, ces estimateurs présentent d'intéressantes propriétés d'optimalité. Sous des hypothèses relativement larges, parfois appelées *conditions de régularité* :

- ils sont consistants ;
- on peut estimer leur variance et leur écart-type ;
- ils sont asymptotiquement normaux.

les deux derniers points permettent la construction d'intervalles de confiance et la pratique de tests classiques pour des données de taille suffisante (une règle « au doigt mouillé » recommande d'avoir au moins une trentaine d'observations par paramètre à estimer).

Les estimateurs sont en outre optimaux en un certain sens parmi les estimateurs consistants (*efficience*).

On maximise en général le logarithme de la vraisemblance : $\ln(L) = \Sigma \ln[f_\theta(x_i)]$, ce qui est équivalent mais simplifie fréquemment les calculs.

Dans le cas discret, c'est la probabilité des observations qui est maximisée par rapport aux paramètres inconnus.

La méthode enfin n'est pas limitée aux échantillons et aux observations indépendantes ; elle peut s'appliquer dès qu'il est possible d'écrire la vraisemblance (ou la probabilité) de l'ensemble des observations en fonction des paramètres à estimer.

Exemple 1 : on suppose que 60% des Irlandais, 30% des Anglais et 5% des nationaux des autres pays sont roux. On aperçoit Place de la Concorde un car de touristes (supposés de la même origine) composés d'une forte proportion de roux ; l'estimation du maximum de vraisemblance de la nationalité des ces visiteurs est a priori : irlandaise…

Exemple 2 : espérance d'une loi normale. Soit une loi normale $\boldsymbol{N}(m, \sigma)$ dont on veut estimer l'espérance m inconnue à l'aide d'un échantillon de taille N par la méthode du maximum de vraisemblance.

$$f_m(x_i) = \frac{1}{\sigma.\sqrt{2\pi}} . \exp\left[\frac{-(x_i - m)^2}{2.\sigma^2}\right]$$

$$\ln(L) = \frac{-1}{2.\sigma^2} . \Sigma (x_i - m)^2 + (\text{terme non fonction de m})$$

$$\frac{d \ln(L)}{dm} = \frac{1}{\sigma^2} . \Sigma (x_i - m)$$

qui s'annule pour la moyenne $\bar{x}$ des x_i. On vérifie qu'il s'agit bien d'un maximum, ce qui montre que l'estimateur du maximum de vraisemblance est la moyenne observée sur l'échantillon. On remarque que la connaissance de l'écart-type s n'est pas nécessaire.

Lorsqu'il n'est pas possible, comme dans l'exemple précédent, d'obtenir une solution explicite, on fait appel à des méthodes numériques de maximisation (mises en œuvre par les logiciels économétriques en de nombreuses procédures).

----===ooOoo===---- (03.03.2009)