CLAUDE MOUCHOT

STATISTIQUE ET ECONOMETRIE

ECONOMICA

STATISTIQUE ET ECONOMETRIE

par

Claude Mouchot

Ancien Elève de l'Ecole Polytechnique
Professeur de Sciences Economiques

ECONOMICA
49, rue Héricart, 75015 Paris
1983

TABLE DES MATIERES

AVANT-PROPOS

Ce cours de Statistique et Économétrie représente un enseignement assuré pendant de nombreuses années au Département de Sciences Économiques et de Gestion de l'Université Lyon II.

Centré au départ sur la présentation des modèles linéaires de l'économétrie, il est apparu peu à peu souhaitable d'y adjoindre l'ensemble des outils nécessaires à la compréhension complète de la démarche économétrique.

C'est ainsi que nous avons été conduit à introduire successivement la présentation des tests de Neyman et Pearson, puisque l'économétrie « décide » de la validité des modèles qu'on lui propose au moyen de cette procédure ; puis l'étude des lois usuellement employées dans les tests effectivement réalisés : χ^2, Student, Fisher ; enfin cette étude elle-même a nécessité la présentation des intégrales doubles, qui permettent la construction complète de ces lois.

On l'aura compris, ce Cours forme un tout : un étudiant issu d'un DEUG de Sciences économiques pourra comprendre la démarche de l'économétrie sans impasse théorique, puisque tous les éléments de cette démarche sont exposés et démontrés.

Les connaissances requises sont en effet essentiellement :

- le calcul différentiel et intégral simple,
- le calcul matriciel élémentaire,
- les concepts fondamentaux associés à une variable aléatoire : loi de probabilité, fonction densité et fonction de répartition, espérance mathématique et variance.

En pratique, nous avons regroupé dans une première partie l'ensemble des outils qui conduisent à la compréhension de la procédure des tests. La seconde partie présente alors l'économétrie des modèles linéaires.

Cet ouvrage ne comporte pratiquement aucun exercice d'application puisque ceux-ci avaient déjà été regroupés dans les « Exercices pédagogiques de statistiques et économétrie » parus précédemment chez le même éditeur. Ces deux ouvrages se correspondent d'ailleurs tant au niveau du plan suivi que des notations employées.

D'autre part, il a paru nécessaire de reprendre intégralement certains développements des « Exercices... » qui y avaient été placés pour que cet ouvrage forme un tout, mais qui ont leur place naturelle dans le présent Cours. Il en est ainsi pour une partie du Chapitre I, pour le Chapitre II, et pour l'élaboration du test de la moyenne d'une loi normale dans le Chapitre IV.

Nous tenons enfin à signaler nos dettes envers les ouvrages suivants :

– Calot (G.), *Cours de Statistiques descriptive,* Dunod.
– Calot (G.), *Cours de Calcul des Probabilités,* Dunod.
– Fourgeaud (C.), *Cours de Statistique de 3^e^ Année de Licence ès-Sciences Économiques,* Librairie Dey.
– Fourgeaud (C.) et Fuchs (A.), *Statistique,* Dunod.
– Kane (E.J.), *Statistique économique et Économétrie,* A. Colin.

Abréviations utilisées

V.A.	variable aléatoire	f.d.	fonction densité
ddl	degrés de liberté	f.r.	fonction de répartition.
l.p.	Loi de probabilité		

Première Partie

THÉORIE DES TESTS ET TESTS USUELS

Chapitre I

INTÉGRALES DOUBLES
NOTION D'INTÉGRALES MULTIPLES

Ce chapitre a deux buts :

- permettre, grâce au calcul des intégrales doubles, l'élaboration des lois de probabilité des variables aléatoires de χ^2, de Student et de Fisher, constamment utilisées dans les tests ;
- présenter la notion d'intégrale multiple, indispensable à la construction théorique des tests de Neyman et Pearson.

1. EXEMPLE

Considérons deux V.A. X et Y, continues et indépendantes, définies sur R^+, de f.d. f(x) et g(y). On sait que l'on a :

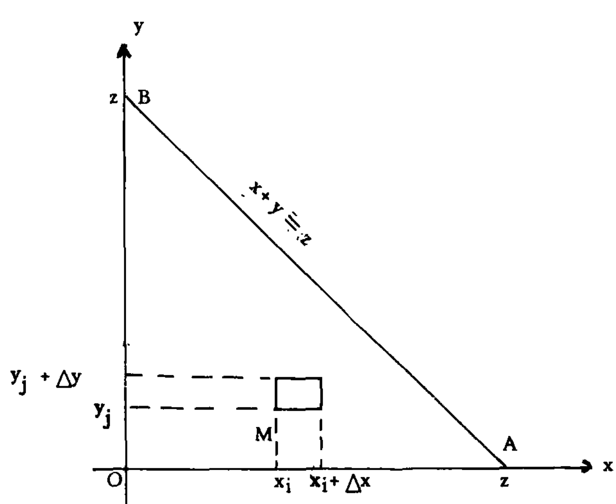
$Pr(X<x_o) = \int_o^{x_o} f(x)dx = F(x_o)$; $F(x) =$ f.r. de X.

$Pr(Y<y_o) = \int_o^{y_o} g(y)dy = G(y_o)$; $G(y) =$ f.r. de Y.

On se propose de déterminer la l.p. de la V.A. $Z = X + Y$.

X et Y étant définies sur R^+, Z sera elle aussi définie sur R^+. Il nous faut de plus déterminer sa f.r. (ou sa f.d.) sur R^+ ; soit :

$$H(z) = Pr(Z<z) = Pr(X+Y<z)$$



Géométriquement, on voit que H(z) représente la probabilité attachée au triangle OAB, domaine à l'intérieur duquel on a bien $x+y < z$.

Soit alors $M(x_i, y_j)$ un point de ce domaine ; la probabilité élémentaire attachée au rectangle figuré, de côtés Δx et Δy, est, conformément au théorème des probabilités composées :

$$f(x_i)\,\Delta x . g(y_j)\,\Delta y$$

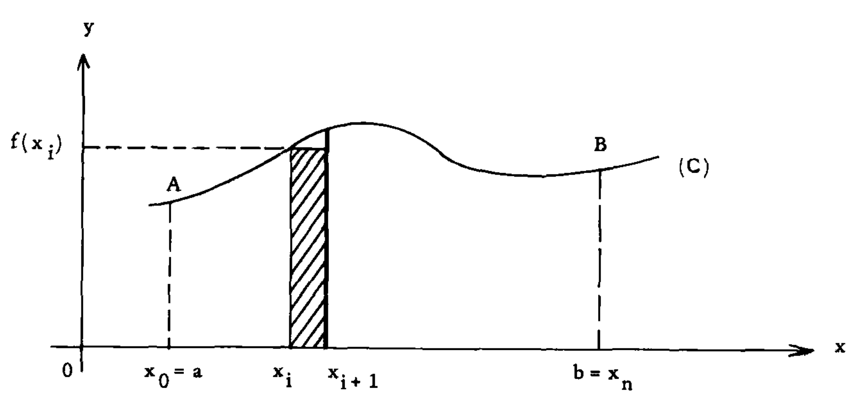
On voit ainsi que la détermination de H(z) nécessite la *sommation* (l'intégration) de toutes les probabilités élémentaires intérieures au triangle OAB conformément au Théorème des probabilités totales. C'est précisément par une intégrale double que l'on effectuera cette sommation.

2. DÉFINITION ET CALCUL D'UNE INTÉGRALE DOUBLE.

La définition d'une intégrale double peut être présentée selon une démarche en tout point analogue à celle d'une intégrale simple ; c'est pourquoi on rappellera tout d'abord la définition de celle-ci.

2.1. Intégrale définie d'une fonction f(x).

Soit une fonction $y = f(x)$, définie et continue en tout point d'un segment (a, b). Soit (C) son graphe dans un plan xOy :



Pour calculer une mesure approchée de l'aire du quadrilatère mixtiligne aABb, on peut procéder de la façon suivante : on divise le segment [a, b] en n segments égaux de longueur $\Delta x = \dfrac{b-a}{n}$ au moyen des points $x_0 = a, x_1, \ldots, x_i, \ldots, x_n = b$.

La quantité $\sum_{i=0}^{n-1} f(x_i) \,.\, \Delta x$, représentant la somme des mesures des aires des rectangles tels que ceux hachurés sur la figure, constitue alors une valeur approchée de la mesure de l'aire cherchée.

D'autre part, il est intuitivement clair que si le nombre d'intervalles n augmente indéfiniment, la somme des aires hachurées a pour limite l'aire cherchée.

On a alors les égalités suivantes :

$$\text{Aire (aABb)} = \lim_{n \to \infty} \sum_{i=0}^{n} f(x_i)\,\Delta x \underset{\substack{\text{par}\\ \text{définition}}}{=} \int_a^b f(x)\,dx.$$

On montre enfin que le *calcul* de cette intégrale définie, en général, nécessite la connaissance d'une primitive F(x) de f(x) et qu'alors on a :

$$\int_a^b f(x)dx = F(b) - F(a) = |F(x)|_a^b$$

Remarques :

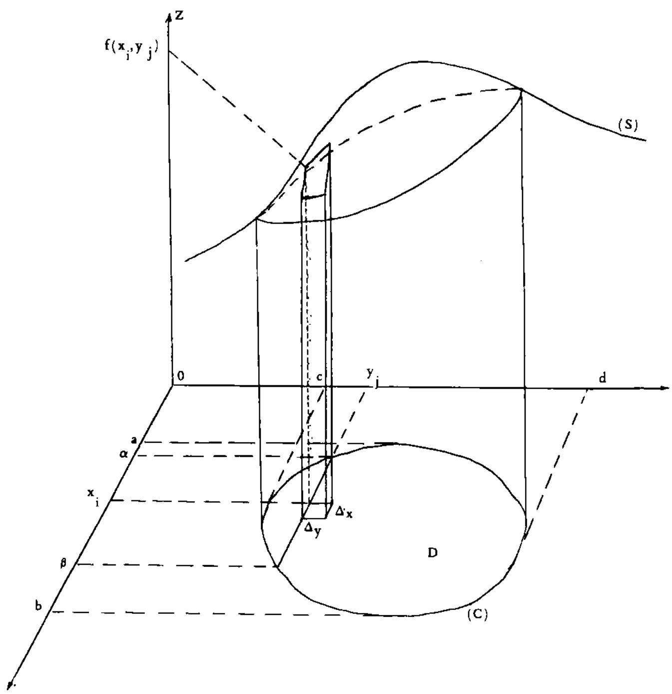
1. On démontre que l'intégrale ci-dessus définie est indépendante du découpage du segment (a, b) ; en particulier, il n'est pas nécessaire que les intervalles $x_{i+1} - x_i$ soient égaux : il suffit qu'ils tendent tous vers zéro quand n tend vers l'infini.

2. Il existe des fonctions qui n'ont pas de primitives ; ainsi $f(x) = \exp(-x^2)$: il n'existe pas de fonction F(x) telle que $F'(x) = f(x)$. Il est toutefois évident que l'intégrale définie $\int_a^b f(x)dx = \int_a^b \exp(-x^2)\,dx$ existe, puisque c'est la mesure de l'aire aABb située sous la courbe y=f(x), dite courte en cloche. Son calcul ne peut alors être effectué, en pratique, que par des méthodes numériques dont l'expression $\Sigma\, f(x_i)\, \Delta x$ est un exemple ; on obtient alors une mesure approchée, au degré d'approximation que l'on désire en choisissant la valeur de n.

2.2. Définition d'une intégrale double.

On peut suivre la même démarche que ci-dessus. Soit donc une fonction z = f(x,y) définie et continue en tout point (x_0, y_0) d'un domaine D du plan x0y, limité par une courbe (C).

Le graphe de cette fonction f est alors, dans un repère (0xyz), une surface (S).



On peut alors définir l'intégrale double de f(x,y) sur D comme la mesure du volume situé entre la surface (S), le plan x0y et le cylindre vertical de base (C).

En découpant le segment ab en n segments égaux de longueur Δx et le segment cd en m segments égaux de longueur Δy, une mesure approchée de ce volume peut s'écrire :

$$\sum_i \sum_j f(x_i, y_j)\, \Delta x\, \Delta y$$

puisque chaque terme $f(x_i, y_i)\, \Delta x\, \Delta y$ n'est autre que la mesure du volume du parallélépipède rectangle figuré.

Il est ici aussi intuitivement clair que si n et m augmentent indéfiniment, donc si Δx et Δy tendent vers zéro, la limite de l'expression ci-dessus n'est autre que la mesure du volume cherchée.

Il nous faut alors être en mesure de calculer l'expression :

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \sum_i \sum_j f(x_i, y_j)\, \Delta x\, \Delta y = V$$

Ecrivons la :

$$V = \lim_{\Delta y \to 0} \sum_j \Delta y \left[\lim_{\Delta x \to 0} \sum_i f(x_i, y_j)\, \Delta x \right]$$

Dans le crochet y_j est une constante par rapport à l'opérateur $\sum_i$. Ce crochet est alors, *par définition*, l'intégrale définie de la fonction d'une variable x, $f(x, y_j)$ sur le segment $[\alpha,\beta]$ (cf. figure).

$$A = \lim_{\Delta x \to 0} \sum_i f(x_i, y_j)\,\Delta x = \int_\alpha^\beta f(x,y_j)\,dx \qquad (y_j = \text{cte})$$

D'autre part, α et β sont fonctions de y_j, fonctions données par l'équation de la courbe (C) (cf. figure). On a donc finalement :

$$A = \int_{\alpha(y_j)}^{\beta(y_j)} f(x,y_j)dx = \Phi\ (y_j),$$

une fois l'intégration effectuée.

Reportant alors ce résultat dans V, il vient :

$$V = \lim_{\Delta y \to 0} \sum_j \Phi\ (y_j)\ \Delta y \underset{\substack{\text{par}\\ \text{définition}}}{=} \int_c^d \Phi(y)\,dy,$$

c et d étant les ordonnées extrêmes des points de (C) (cf. figure). En explicitant à nouveau $\Phi\ (y)$, il vient enfin

$$V = \int_c^d \left[\int_{\alpha(y)}^{\beta(y)} f(x,y)dx\right] dy = \int_c^d dy \int_{\alpha(y)}^{\beta(y)} f(x,y)dx =$$
$$\iint_D f(x,y)\,dx\,dy.$$

les deux dernières notations étant conventionnelles.

Nous avons intégré d'abord en x, à y fixé ; il est évident que la démarche inverse est possible et que l'on peut intégrer d'abord en y, à x fixé :

$$V = \int_a^b \left[\int_{\gamma(x)}^{\delta(x)} f(x,y)\,dy\right]dx = \int_a^b dx \int_{\gamma(x)}^{\delta(x)} f(x,y)\,dy.$$

On voit enfin

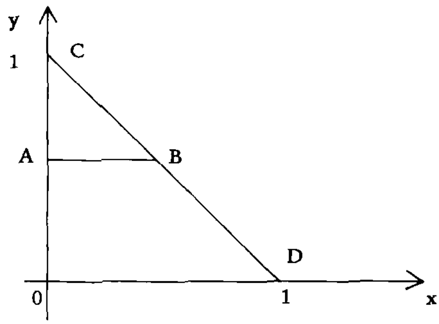
1. la raison de l'appellation «intégrale double» ; il s'agit bien du calcul successif de 2 intégrales simples.
2. l'absence *totale* de difficulté nouvelle au niveau du calcul intégral. La seule difficulté nouvelle réside dans la détermination des bornes de la première intégrale, bornes qui sont en général fonction de la seconde variable d'intégration et qui s'obtiennent très simplement à partir de l'équation de la courbe (C).

2.3. Calcul d'une intégrale double.

Exemple 1. Calculer $I = (x+y^2)dxdy$ dans le domaine :

$$\{\ x + y \leqslant 1\ ; x \geqslant 0\ ; y \geqslant 0\ \}$$

Ce domaine se représente ainsi :



Il est représenté par le triangle OCD. Nous intégrons d'abord par rapport à x, à y fixé. En se donnant alors une valeur de y, on voit sur la figure que x varie de x_A à x_B, à l'intérieur du domaine. On a évidemment $x_A = O, \forall\, y$. Quant à x_B, il est donné, en fonction de y, par l'équation de la droite CD :

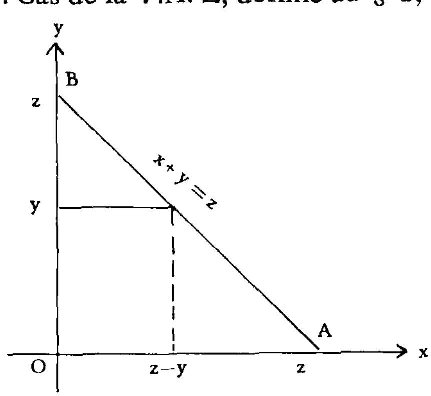
$$x + y = 1 \qquad x_B = 1 - y$$

Une fois cette intégration faite, y doit varier de 0 à 1 pour que la totalité du domaine soit balayée. D'où le calcul de I :

$$I = \iint (x+y^2)\,dx\,dy = \int_o^1 dy \int_o^{1-y} (x+y^2)\,dx = \int_o^1 \left| \frac{x^2}{2} + y^2 x \right|_o^{1-y} dy$$

$$= \int_o^1 [\frac{1}{2}(1-y)^2 + y^2(1-y)]\,dy = \int_o^1 (-y^3 + \frac{3}{2}y^2 - y + \frac{1}{2})\,dy = \frac{1}{4}$$

Exemple 2. Cas de la V.A. Z, définie au § 1, ci-dessus.



La f.r. de Z que nous cherchions à calculer est donc :

$H(z) = Pr(Z < z) = Pr(X+Y < z) = \iint f(x)\, g(y)\, dxdy$, dans le triangle OAB de la figure. Le même raisonnement que dans l'exemple 1 conduit alors à :

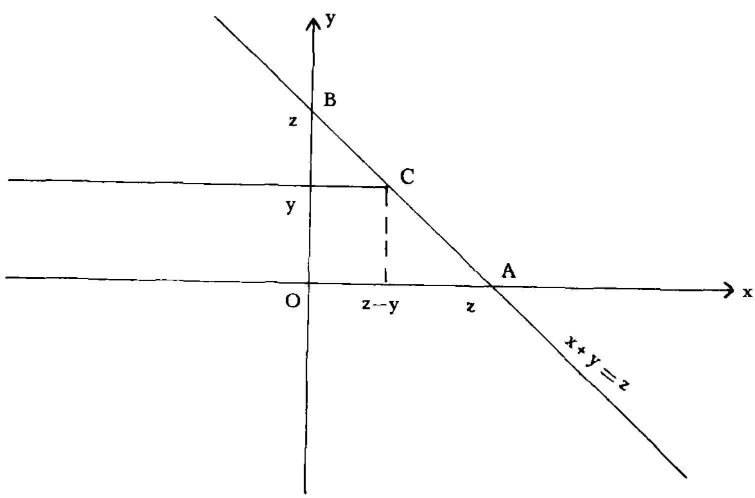
$$H(z) = \int_0^z dy \int_0^{z-y} f(x)\, g(y)\, dx = \int_0^z g(y)\, dy \int_0^{z-y} f(x)\, dx.$$

Exemple 3. Loi de propabilité de la somme de deux V.A. N(0, 1), indépendantes.

Soient X et Y les deux V.A. N(0, 1). On a donc :

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}} \; ; \; g(y) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}}$$

Mais cette fois X et Y étant définies sur R, le domaine d'intégration est le demi-plan situé sous la droite AB, d'équation : $x + y = z$.



A y fixé, x varie de $-\infty$ à $x_C = z-y$. Puis y doit varier de $-\infty$ à $+\infty$ pour balayer tout le demi-plan. On a ainsi :

$$H(z) = Pr(X+Y < z) = \int_{-\infty}^{+\infty} dy \int_{-\infty}^{z-y} \frac{1}{2\pi}\, e^{-\frac{x^2}{2}} e^{-\frac{y^2}{2}}\, dx$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\, dy \int_{-\infty}^{z-y} e^{-\frac{x^2}{2}}\, dx$$

Nous ne pouvons pas poursuivre le calcul puisqu'il n'existe pas de primitive de la fonction $\exp(-x^2/2)$ (cf. Remarque 2. § 2.1.). Mais puisque nous cherchons la l.p. de Z, il nous suffit d'avoir sa f.d. h(z), au lieu de sa f.r. H(z).

Or on sait (résultat du calcul intégral simple) que :

$$\text{si } H(z) = \int_a^z h(u)du, \text{ alors } H'(z) = \frac{dH(z)}{dz} = h(z).$$

Appliquant ce résultat à la première intégrale (en x), il vient :

$$h(z) = H'(z) = \frac{dH(z)}{dz} = \frac{dH(z)}{d(z-y)} \cdot \frac{d(z-y)}{dz} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-y^2/2} \cdot e^{-\frac{1}{2}(z-y)^2} \cdot dy$$

$$= \frac{1}{2\pi} \int e^{-\frac{1}{2}(2y^2 - 2yz + z^2)} \cdot dy$$

L'exposant peut alors s'écrire : $-\left[\left(y - \frac{z}{2}\right)^2 + \frac{z^2}{4}\right]$; d'où :

$$h(z) = \frac{1}{2\pi} e^{-z^2/4} \int_{-\infty}^{+\infty} e^{-\left(y - \frac{z}{2}\right)^2} \cdot dy$$

Posons enfin $y - \frac{z}{2} = \frac{u}{\sqrt{2}}$; $dy = \frac{du}{\sqrt{2}}$; il vient :

$$h(z) = \frac{1}{2\pi} e^{-z^2/4} \cdot \frac{1}{\sqrt{2}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} \cdot du$$

On reconnaît alors l'intégrale de la loi normale N(0, 1) et on sait que (ce résultat sera démontré plus loin) :

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du = 1 \Leftrightarrow \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du = \sqrt{2\pi}$$

D'où en définitive :

$$h(z) = \frac{1}{\sqrt{2\pi}.\sqrt{2}} \cdot e^{-\frac{z^2}{2.2}} \Leftrightarrow Z \rightsquigarrow \mathcal{N}(0, \sqrt{2}).$$

Remarque : Nous serons conduit plusieurs fois à dériver ainsi des intégrales pour parvenir à déterminer certaines l.p.