Revue de statistique appliquée

G. Rouzet

Incidence pratique de la méthode utilisée pour l'étude graphique de la loi de probabilité d'une variable aléatoire à partir d'un échantillon d'observations indépendantes

Revue de statistique appliquée, tome 14, n° 3 (1966), p. 5-24

<http://www.numdam.org/item?id=RSA_1966__14_3_5_0>

INCIDENCE PRATIQUE DE LA MÉTHODE UTILISÉE POUR L'ÉTUDE GRAPHIQUE DE LA LOI DE PROBABILITÉ D'UNE VARIABLE ALÉATOIRE A PARTIR D'UN ÉCHANTILLON D'OBSERVATIONS INDÉPENDANTES

G. ROUZET

Ingénieur à la Compagnie des Compteurs

1 - REMARQUES PRELIMINAIRES

Lorsqu'on désire étudier par voie graphique la loi de probabilité d'une variable aléatoire, à partir d'un échantillon d'observations indépendantes, il est habituel d'utiliser la notion de courbe de répartition des observations, c'est-à-dire de faire correspondre à chaque valeur observée la proportion des observations de valeurs inférieures ou égales à cette valeur. C'est cette notion que nous utiliserons dans ce qui suit. Précisons toutefois que nous ne ferons pas usage d'un tracé joignant les points représentatifs des couples (valeur, fréquence cumulée) définis ci-dessus, mais que nous nous intéresserons à l'ensemble de ces points.

Par ailleurs, deux méthodes d'investigation peuvent être envisagées :

- attendre de la représentation graphique des résultats d'échantillonnage - dans un système d'axes à échelles arithmétiques - qu'elle suggère tel modèle mathématique pour la loi de probabilité de la variable,

- faire une hypothèse quant à la loi de probabilité de la variable et juger, en fonction de la représentation graphique des résultats d'échantillonnage - dans un système d'axes utilisant éventuellement des échelles judicieusement adaptées - de la validité de l'hypothèse.

Nous laisserons de côté la première méthode, en nous bornant à noter qu'elle n'a de sens que si le nombre des observations est très grand, et nous nous intéresserons dans ce qui suit à la seconde méthode, dans son principe même.

2 - NOTATIONS

Soient :

X : la variable aléatoire étudiée. Nous supposerons cette variable continue, ce qui n'affecte en rien le caractère général des principes que nous allons expliciter.

f(x) et F(x) : respectivement, les expressions définissant la densité de probabilité et la fonction de répartition pour toute valeur x de X.

n : le nombre d'observations indépendantes effectuées.

i : le rang d'une observation dans le classement par ordre non décroissant des valeurs observées.

v_i^* : la valeur de l'observation de rang i :

$$v_1 \leqslant v_2 \leqslant \ldots \leqslant v_i \leqslant \ldots \leqslant v_n.$$

3 - ETUDE THEORIQUE DU COMPORTEMENT DES OBSERVATIONS ALEATOIRES POUR UNE LOI DE PROBABILITE DE X CONNUE

Supposons connue la loi de probabilité de X définie par f(x) ou F(x) ; proposons nous de constituer un échantillon de n observations aléatoires indépendantes et de classer ensuite ces observations dans l'ordre non décroissant des valeurs, et appelons provisoirement x_i la valeur que prendra l'observation de rang i dans ce classement.

Le symbole x_i, auquel ne correspondra une valeur numérique (v_i) que lorsque l'échantillon aura été constitué, désigne ici toute valeur possible de la variable aléatoire correspondante X_i dont la loi de probabilité (bien connue, ce qui nous dispense de l'établir) est définie par la densité de probabilité :

$$g(x_i) = \frac{[F(x_i)]^{i-1}\,[1 - F(x_i)]^{n-i}\,f(x_i)}{B(i\;;\;n - i + 1)} \qquad (1)$$

B désignant la fonction eulérienne de première espèce :

$$B(p,\; q) = \int_0^1 u^{p-1}\,(1 - u)^{q-1}\,du$$

Cette loi de probabilité pourra être caractérisée de façon concrète par son espérance mathématique $E(X_i)$ et des limites en probabilité correspondant à un seuil α donné ; soient $L_1(i)$ et $L_2(i)$ les limites en probabilité de X_i telles que :

$$\begin{cases} G(x_i = L_1) = \alpha/2 \\ G(x_i = L_2) = 1 - \alpha/2 \end{cases} \qquad (2)$$

Sur un graphique à échelles arithmétiques, on peut alors porter en abscisses les valeurs successives de i, et en ordonnées les quantités $E(X_i)$, $L_1(i)$, $L_2(i)$ correspondantes (figure 1).

(Ce n'est que pour faciliter la lecture d'un tel graphique que l'on a joint par des courbes les points qui se correspondent lorsque i varie, la notion d'interpolation étant bien évidemment dépourvue de sens).

Nous préfèrerons toutefois porter en ordonnées les proportions i/n et en abscisses les valeurs x_i caractérisant les lois de probabilité des variables X_i (figure 2). Cette représentation, équivalente, est, en effet, en correspondance directe avec la notion de courbe de répartition d'un échantillon.

Remarque

Il y a lieu d'attirer l'attention sur le fait que, sur la figure 2, le caractère aléatoire des résultats d'échantillonnage se trouve ainsi reporté sur les abscisses, alors que les ordonnées, résultant de la seule valeur de n, sont au contraire déterminées.

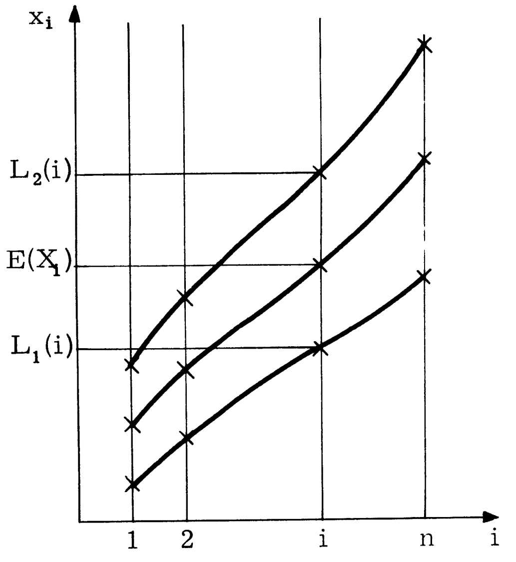


Figure 1

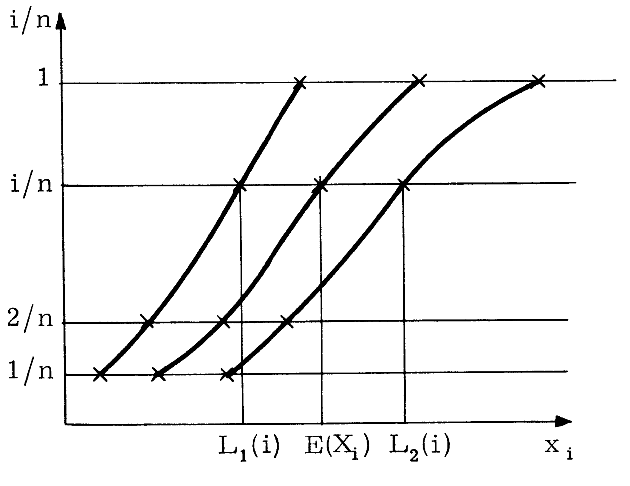


Figure 2

4 - GRAPHIQUE DES VALEURS OBSERVEES

Réalisons maintenant l'échantillonnage et observons les valeurs numériques v_i prises par les variables X_i dans l'échantillon.

Nous porterons sur un graphique les points (v_i ; i/n) de la courbe de répartition de l'échantillon (figure 3).

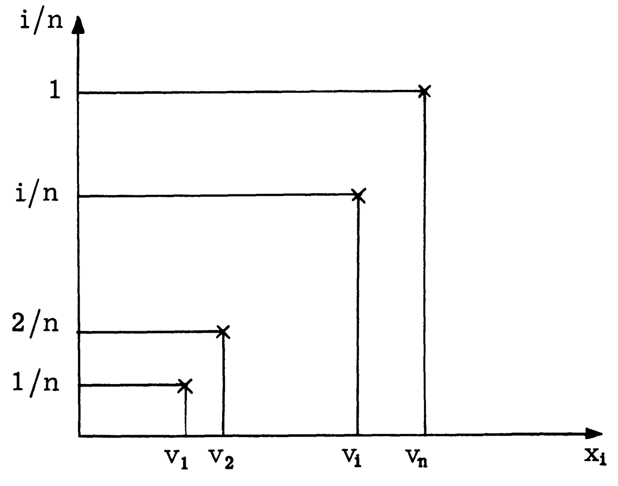


Figure 3

Remarque

Si l'on procède r fois à la constitution d'un tel échantillon et si l'on enregistre les observations sur un même graphique, on obtient n ensembles dont chacun est constitué de r points distribués sur une même ordonnée (figure 4), la valeur de cette ordonnée pouvant être définie a priori connaissant seulement n.

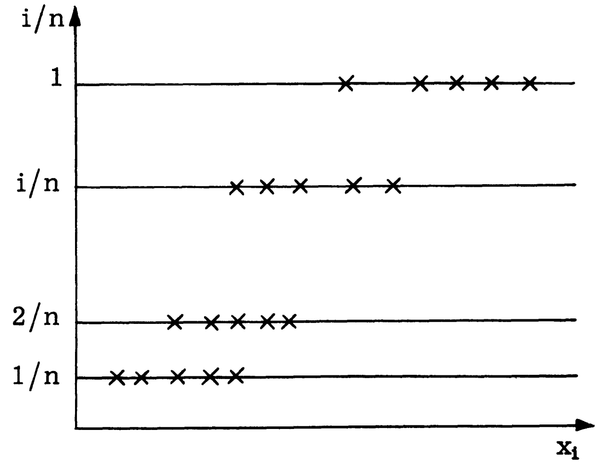


Figure 4

Nous retrouvons ici la notion mise en relief dans la remarque finale du paragraphe précédent.

Par ailleurs, pour chaque ordonnée, et indépendamment des autres, les r points obtenus correspondent à r observations indépendantes de la même variable X_i dont la loi de probabilité est définie par $g(x_i)$. Ainsi, si l'on appelle k le numéro de l'échantillon ($1 \leqslant k \leqslant r$) la moyenne empirique :

$$\overline{x_i} = \frac{1}{r} \sum_{k=1}^{k=r} x_{ki} \tag{3}$$

converge, lorsque r augmente, vers l'espérance mathématique de $g(x_i)$:

$$\lim[\overline{x_i}]_{r \to \infty} = E(X_i) \tag{4}$$

5 - JUGEMENT SUR LA VALIDITE D'UNE HYPOTHESE QUANT A LA LOI DE PROBABILITE DE X

Examinons maintenant de quelle manière ce qui vient d'être dit conduit à résoudre le problème pratique, qui consiste à juger de la validité d'une hypothèse quant à la loi de probabilité, inconnue, de X.

Dès l'hypothèse faite, l'étude mathématique du comportement des caractéristiques d'échantillonnage se fait sur la base d'une loi de X connue, et, en particulier, les notions explicitées dans les paragraphes précédents sont directement applicables.

Le critère de validité de l'hypothèse est alors un critère de compatibilité entre l'ensemble des points expérimentaux (figure 3) et le graphique préétabli à partir de f(x) et de n (figure 2). La notion de compatibilité inclut en l'occurrence deux aspects différents et complémentaires que nous allons examiner successivement.

5.1 - Comparaison de l'ensemble des points (v_i ; i/n) à l'ensemble des points ($E(X_i)$; i/n)

Si l'on porte sur un même graphique les points expérimentaux (figure 3) et les points de la "courbe" moyenne du graphique des valeurs calculées dans le cadre de l'hypothèse (figure 2), on est plus ou moins fondé à admettre la légitimité de celle-ci selon que les premiers sont plus ou moins proches, dans leur ensemble, des seconds.

Pour faciliter cette comparaison, on peut utiliser une anamorphose telle que les points "théoriques" soient alignés : si nous appelons y_i la valeur qui, sur une échelle arithmétique auxiliaire, repère la position qu'occupe l'ordonnée i/n sur l'échelle modifiée (figure 5), cette anamorphose est définie par :

$$y_i = E(X_i) \tag{5}$$

(cette relation semble restreindre à une égalité la notion de transformation linéaire de $E(X_i)$. Mais, en fait, cette dernière notion se trouve incluse dans le choix arbitraire de l'échelle arithmétique des y_i).

Notons que l'on peut alors décomposer l'appréciation de la validité de l'hypothèse en deux éléments ayant respectivement pour base :

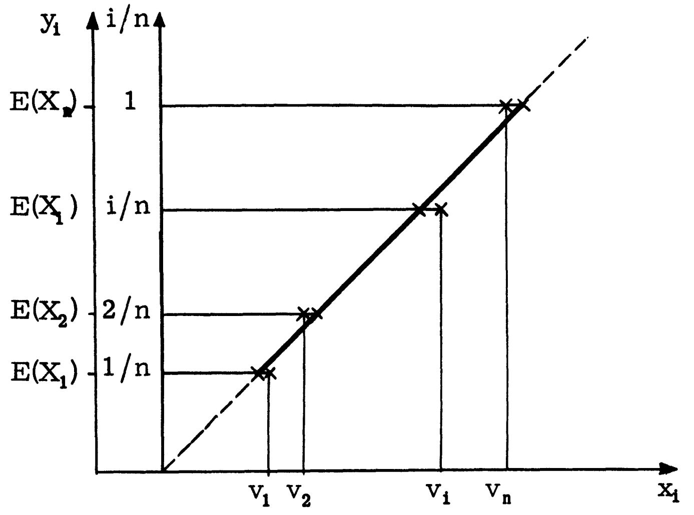


Figure 5

- l'observation d'une allure approximativement rectiligne de l'ensemble des points expérimentaux,

- l'identification de la droite moyenne correspondante, à la droite passant par l'ensemble des points $(E(X_i)\ ;\ i/n)$.

Cette remarque conduit à l'idée d'un procédé graphique susceptible de permettre de juger de la validité d'une hypothèse concernant l'appartenance de la loi de X à une certaine famille de lois de probabilité ; dans ce cas, en effet, l'expression de f(x) met en jeu des paramètres littéraux et il est souhaitable de pouvoir juger séparément de la validité du modèle paramétrique et de la validité des valeurs affectées aux paramètres littéraux. Par ailleurs, ces dernières valeurs résultant en général d'estimations qui nécessitent des calculs préalables à l'étude graphique proprement dite, on peut aussi souhaiter pouvoir mettre en oeuvre un procédé graphique d'estimation sous réserve d'une précision suffisante.

Il est effectivement des cas dans lesquels ceci peut être réalisé.

Ainsi, si les valeurs $E(X_i)$ sont proportionnelles à un paramètre "a" de valeur inconnue :

- on définira l'ordonnée arithmétique du point d'abscisse v_i par :

$$z_i = \frac{E(X_i)}{a} \tag{6}$$

- un premier examen du graphique permettra d'apprécier si l'hypothèse paraît raisonnable, en ce que l'ensemble des points est sensiblement rectiligne,

- dans l'affirmative, on tracera la droite moyenne : l'inverse de sa pente fournira l'estimation graphique de "a".

5.2 - Positions des points (v_i ; i/n) par rapport aux intervalles résultant des limites en probabilité

Si nous envisageons maintenant la superposition du graphique expérimental (figure 3) au graphique "théorique" (figure 2) en nous intéressant plus précisément sur ce dernier à l'ensemble des limites $L_1(i)$ et $L_2(i)$, le critère de compatibilité prend, pour chaque ordonnée considérée indépendamment des autres, la valeur d'un test statistique (figure 6).

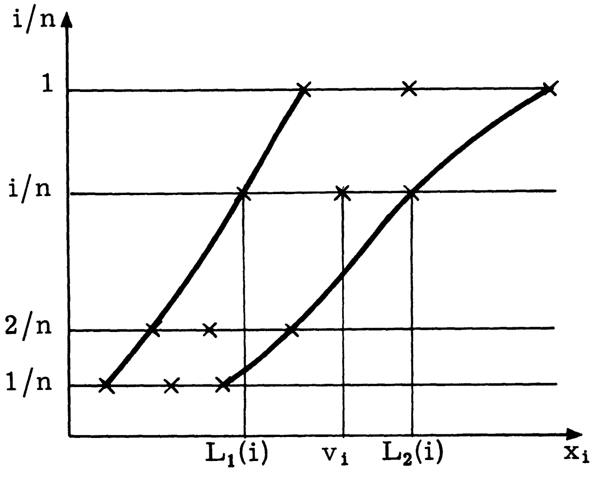


Figure 6

Pour l'ensemble des points, la même acception du mot test n'est possible que si l'on peut calculer le risque global ; notons pourtant que, dans un sens plus large, la notion de test reste dans tous les cas attachée à cette méthode graphique, en ce qu'elle a pour base la comparaison de valeurs observées à des limites précalculées.

La synthèse des possibilités exposées en 5.1 et 5.2 est aisée : il suffit d'utiliser dans tous les cas l'échelle des ordonnées définie en 5.1 (figure 5).

Le graphique auquel nous aboutissons ainsi (figure 7) constitue une disposition tout particulièrement favorable pour l'étude graphique d'une hypothèse quant à l'expression de f(x), ou dans certains cas, comme nous l'avons vu, quant à un modèle paramétrique pour f(x).

(Dans ce dernier cas, les points expérimentaux sont d'abord portés sur le graphique ; si la notion de droite moyenne s'avère avoir un sens, l'estimation des valeurs des paramètres littéraux permet ensuite le calcul des limites $L_1(i)$ et $L_2(i)$, dont les points représentatifs viennent alors compléter le graphique).

6 - EXEMPLE

Plaçons-nous dans le cas où l'hypothèse concernant la loi de X est celle d'une loi exponentielle de paramètre m.

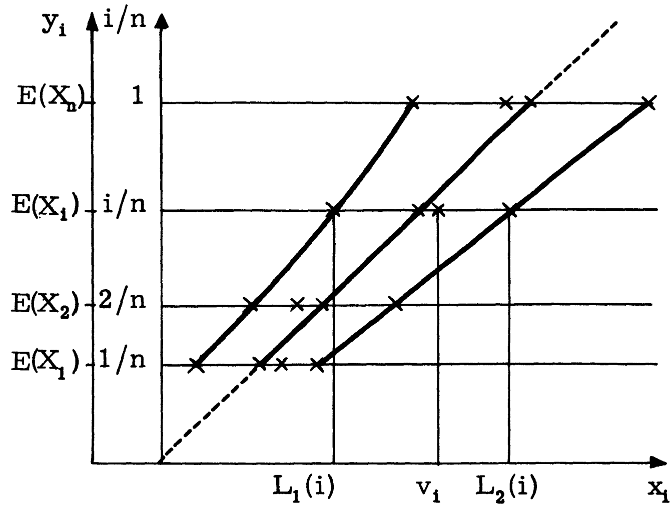


Figure 7

Nous avons :

$$\left.\begin{aligned} f(x) &= \frac{1}{m}\, e^{-\frac{x}{m}} \\ F(x) &= 1 - e^{-\frac{x}{m}} \end{aligned}\right\} \tag{7}$$

L'expression de $g(x_i)$ s'écrit alors :

$$g(x_i) = \frac{1}{m}\, \frac{\left[1 - e^{-\frac{x_i}{m}}\right]^{i-1} \left[e^{-\frac{x_i}{m}}\right]^{n-i+1}}{B(i\ ;\ n - i + 1)} \tag{8}$$

D'où, (comme le montre le calcul reporté en annexe I) :

$$E(X_i) = m \sum_{j=1}^{i=1} \frac{1}{n + 1 - j} \tag{9}$$

Par ailleurs, les limites en probabilité de X_i sont données, pour chaque rang i et indépendamment des autres, à partir de la fonction B incomplète pour laquelle nous emploierons la notation de Pearson :

$$I_x(p\ ;\ q) = \frac{\int_0^x u^{p-1}\,(1 - u)^{q-1}\,du}{B(p\ ;\ q)}$$

et en posant :

$$\left\{\begin{aligned} I_{u_1(i)}\ (i\ ;\ n - i + 1) &= \frac{\alpha}{2} \\ I_{u_2(i)}\ (i\ ;\ n - i + 1) &= 1 - \frac{\alpha}{2} \end{aligned}\right. \tag{10}$$

par :

$$\begin{cases} L_1(i) = m \ \ln \dfrac{1}{1 - u_1} \\ \\ L_2(i) = m \ \ln \dfrac{1}{1 - u_2} \end{cases} \qquad (11)$$

("ln a" désignant le logarithme naturel de a).

Les résultats exprimés par les relations (9) et (11) permettent de construire le graphique correspondant à la figure 7 ; en particulier, on portera l'ordonnée i/n à une distance de l'origine égale :

- si dans l'hypothèse m a une valeur numérique , à :

$$y_i = m \sum_{j=1}^{j=i} \frac{1}{n + 1 - j}$$

selon la relation (5)

- si dans l'hypothèse m est laissé sous forme littérale, à :

$$z_i = \sum_{j=1}^{j=i} \frac{1}{n + 1 - j}$$

puisqu'on se trouve dans le cas de la relation (6).

En utilisant les données numériques suivantes :

$$\left.\begin{array}{l} m = 1 \\ n = 9 \\ \alpha = 5\ \% \end{array}\right\} \qquad (12)$$

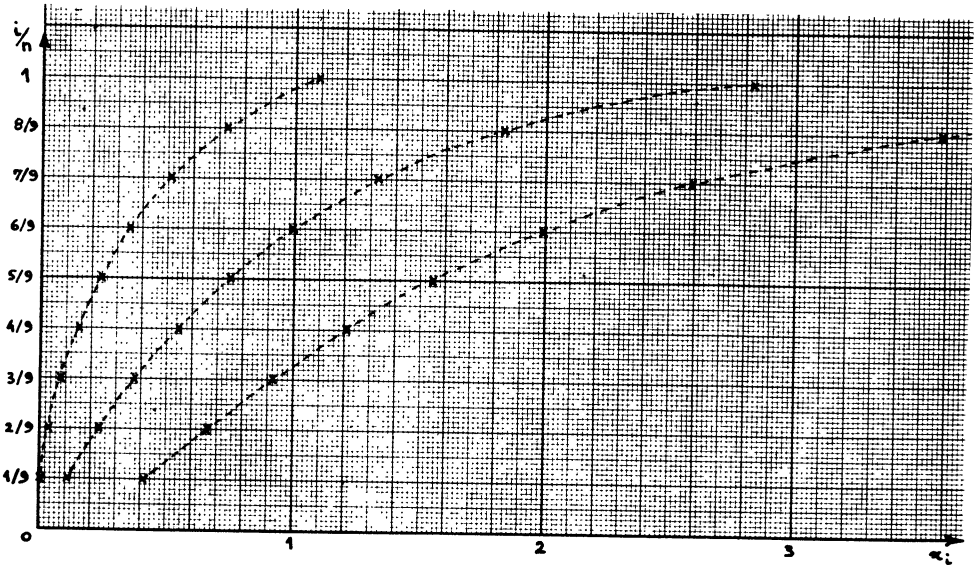
on a tracé, à titre d'illustration, les graphiques correspondant à la figure 2 (voir graphique 1) et à la figure 7 (voir graphique 2).

Les valeurs mises en jeu sont récapitulées dans le tableau ci-dessous :

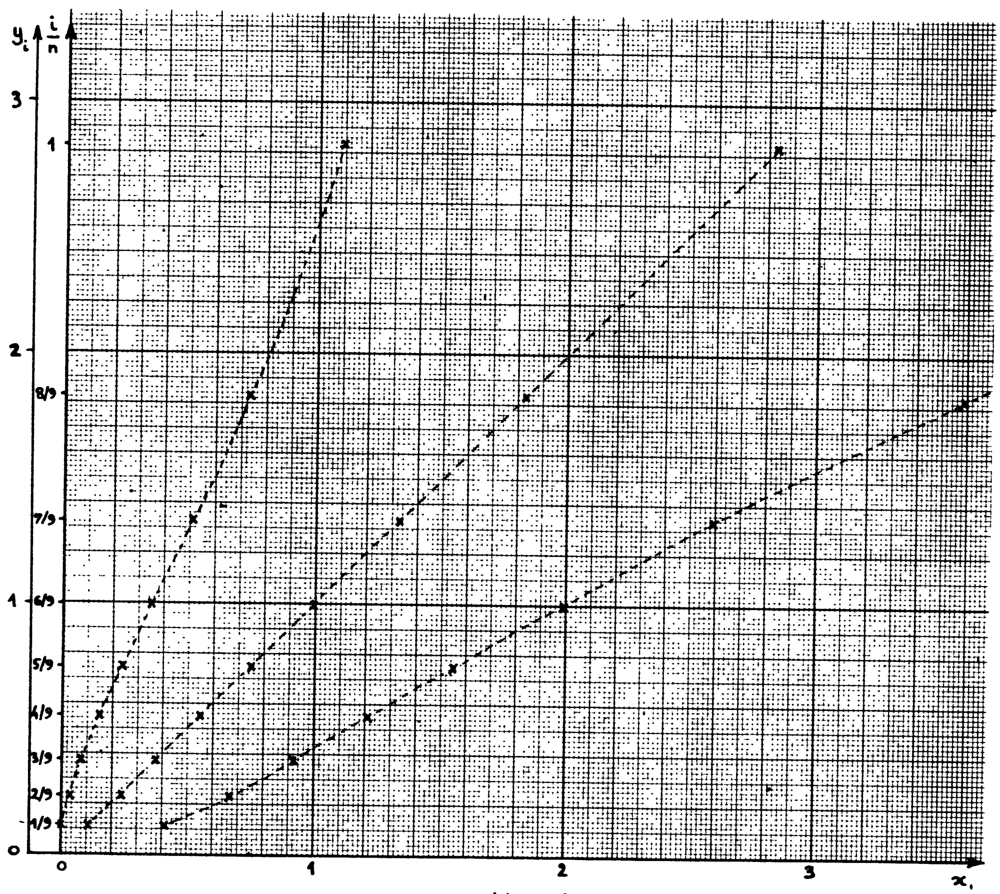
i	$\sum_{j=1}^{j=i} \frac{1}{n+1-j}$	$L_1(i)$	$L_2(i)$
1	0,111	0,00	0,41
2	0,236	0,03	0,66
3	0,379	0,08	0,92
4	0,546	0,15	1,21
5	0,748	0,24	1,55
6	0,996	0,35	1,99
7	1,329	0,51	2,59
8	1,829	0,73	3,58
9	2,829	1,09	5,8

7 - UTILISATION D'UN CHANGEMENT DE VARIABLE

Revenons à l'expression (1) qui définit la loi de X_i, pour une fonction F(x), une taille n d'échantillon et une valeur de i données.



graphique 1



graphique 2

Dans les mêmes conditions, considérons la variable aléatoire Φ_i, telle que chaque valeur φ_i de cette variable corresponde à chaque valeur x_i par la relation :

$$\varphi_i = F(x_i) \tag{13}$$

La loi de probabilité de Φ_i est alors définie par :

$$h(\varphi_i) = \frac{\varphi_i^{i-1} (1 - \varphi_i)^{n-i}}{B(i ; n - i + 1)} \tag{14}$$

On remarque que la loi de Φ_i est indépendante de l'expression de F(x). L'utilisation de ce changement de variable permet donc l'usage d'un graphique commun, quelle que soit la fonction F(x), établi pour une valeur donnée de n.

Pour réaliser ce graphique (figure 8) nous pouvons porter en abscisses les valeurs i et en ordonnées les valeurs caractérisant la loi de Φ_i, c'est-à-dire l'espérance mathématique dont la valeur est :

$$E(\Phi_i) = \frac{i}{n + 1} \tag{15}$$

et les limites en probabilité $u_1(i)$ et $u_2(i)$ au seuil α qui sont fournies par la fonction B incomplète, selon les relations (10) que nous rappelons ici :

$$\begin{cases} I_{u_1(i)} (i ; n - i + 1) = \frac{\alpha}{2} \\ I_{u_2(i)} (i ; n - i + 1) = 1 - \frac{\alpha}{2} \end{cases} \tag{10}$$

Les points représentatifs des résultats d'échantillonnage auront, par ailleurs, pour coordonnées : $[i ; F(v_i)]$

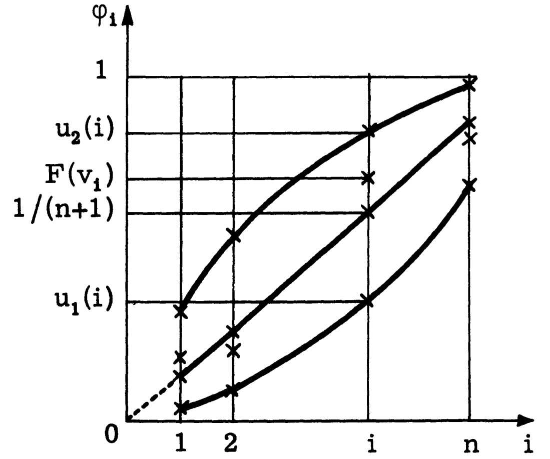


Figure 8

Remarquons que l'ensemble des points $\left(i ; \frac{i}{n+1}\right)$ se présente directement selon une droite.

Remarquons aussi qu'en abscisses, à l'échelle des valeurs φ_i peut être associée une échelle, non linéaire, des valeurs de x_i correspondantes.

Facultativement, cette dernière remarque peut nous inciter à revenir, en permutant les axes de coordonnées et en remplaçant i par i/n sur la nouvelle échelle des ordonnées, à une représentation relative à la courbe de répartition des observations (figure 9).

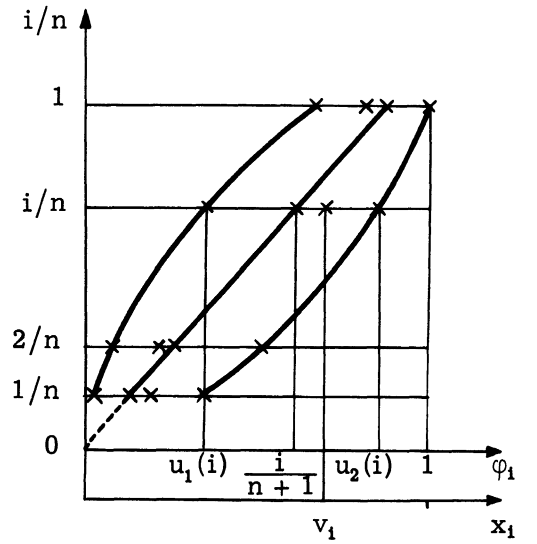


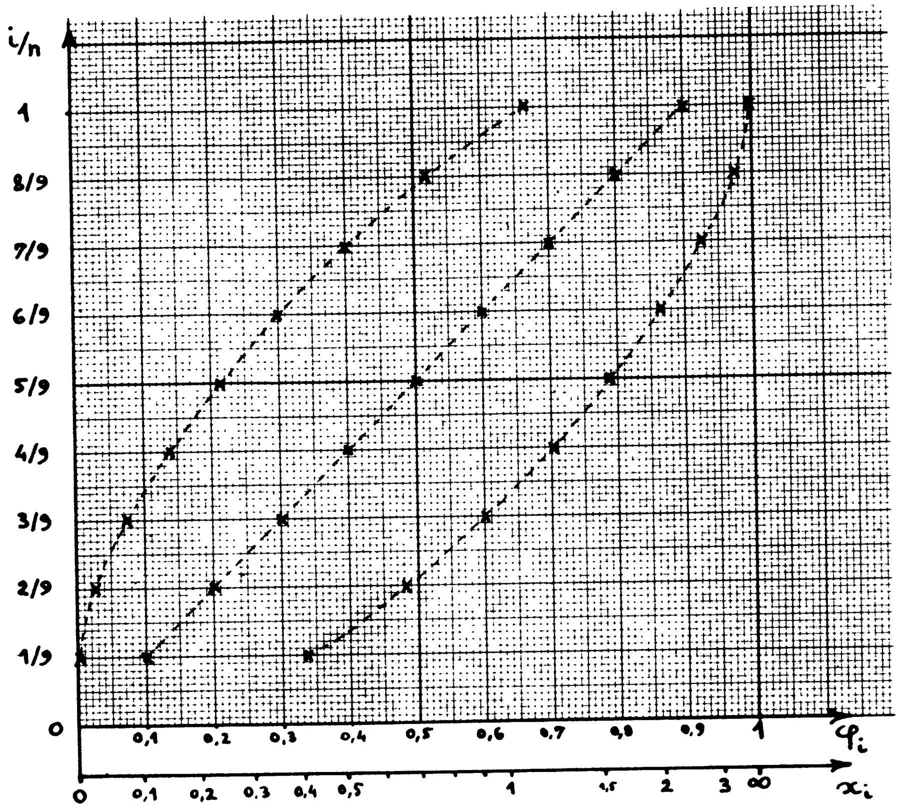
Figure 9

$$\begin{cases} I_{u_1(i)}\ (i\ ;\ n - i + 1) = \dfrac{\alpha}{2} \\ I_{u_2(i)}\ (i\ ;\ n - i + 1) = 1 - \dfrac{\alpha}{2} \end{cases}$$

Il est clair que cette méthode ne peut être mise en oeuvre que lorsque la fonction F(x) est entièrement définie, c'est-à-dire, sous l'angle pratique, lorsque l'hypothèse met en jeu une loi de probabilité déterminée, le cas d'un modèle paramétrique se trouvant exclu. (Dans ce dernier cas, on pourra seulement, par une estimation préalable des valeurs des paramètres dans le cadre de l'hypothèse, se ramener au cas d'une loi de probabilité déterminée).

Telle est la restriction associée à l'emploi de cette méthode, par rapport à la méthode de base exposée précédemment.

On trouvera ci-après (graphique 3) le graphique correspondant aux données de l'exemple précédent, précisées par les relations (7) et (12),



graphique 3

dont les valeurs caractéristiques sont récapitulées dans le tableau suivant :

i	$u_1(i)$	$\frac{i}{n+1}$	$u_2(i)$
1	0,00	0,1	0,34
2	0,03	0,2	0,48
3	0,07	0,3	0,60
4	0,14	0,4	0,70
5	0,21	0,5	0,79
6	0,30	0,6	0,86
7	0,40	0,7	0,93
8	0,52	0,8	0,97
9	0,66	0,9	1,00

8 - AUTRES METHODES UTILISEES

En dehors de quelques cas particuliers pour l'expression de f(x) - ainsi la loi exponentielle que nous utilisons à titre d'exemple - l'expression de $g(x_i)$ se prête assez mal au calcul de $E(X_i)$.

La mise en oeuvre de la méthode résultant des conclusions du paragraphe 5 implique donc en général - ainsi dans l'hypothèse de la normalité de la loi de X - l'usage de tables de valeurs numériques précalculées.

Cette difficulté n'est pas insurmontable, mais elle n'en constitue pas moins un inconvénient sur le plan pratique.

De là sans doute l'usage très courant d'autres méthodes, lesquelles prennent essentiellement pour base un ou plusieurs des trois éléments suivants :

1/ l'anamorphose telle que F(x) soit représentée par une droite sur le graphique où seront portés les points expérimentaux (droite d'Henry, dans son acception la plus générale) :

Il est clair, en effet, qu'il y a convergence, lorsque $n \longrightarrow \infty$, de l'ensemble des points (v_i ; i/n) vers la fonction de répartition F(x) de la variable.

Mais pour une taille d'échantillon finie, l'anamorphose est différente de celle définie au paragraphe 5 : ainsi, pour une même ordonnée i/n, l'abscisse ξ_i telle que :

$$F(\xi_i) = \frac{i}{n} \tag{16}$$

est différente de l'espérance mathématique $E(X_i)$, comme le montrent les calculs reportés en annexe II dans le cas de la loi exponentielle. La "courbe" moyenne autour de laquelle se distribuent les points (v_i ; i/n) est donc différente de la droite qui, dans l'anamorphose considérée, représente F(x).

2/ l'emploi de coordonnées (x_i, Q_i) telles que les points (ξ_i', Q_i) correspondant aux valeurs successives de i se placent sur une droite, ξ_i' étant défini par :

$$F(\xi_i') = \frac{i}{n+1} \tag{17}$$

Une telle disposition est justifiée dans le cas de la seconde présentation du graphique étudié au paragraphe 7 (figure 9), ce graphique se rapportant à l'étude de la variable $\varphi_i = F(x_i)$ dont les valeurs sont portées sur une échelle arithmétique.

Mais si l'on utilise une autre échelle pour x_i, correspondant à un changement de variable différent, la nouvelle variable utilisée n'admet plus pour espérance mathématique la valeur correspondant à $x_i = \xi_i'$.

En particulier, si l'on porte les valeurs x_i sur une échelle arithmétique, on a (en excluant évidemment la loi rectangulaire) :

$$\xi_i' \neq E(X_i) \tag{18}$$

puisque (en notant que $F(\xi_i') = \frac{i}{n+1} = E[F(x_i)]$ en ce qui concerne le membre de gauche) :

$$E[F(x_i)] \neq F[E(X_i)] \tag{19}$$

3/ l'étude de la loi de probabilité de la fréquence cumulée correspondant, dans un échantillon aléatoire, à une valeur de x donnée.

F(x) étant définie et n étant donné, la loi de probabilité du rang i occupé par l'observation de valeur x = A donnée se ramène en effet à une loi binômiale.

Mais il faut observer que tout se passe, sous cette optique, comme si, chacune des valeurs v_i - considérée indépendamment des autres - étant supposée donnée à l'avance, l'échantillonnage avait pour but d'en révéler le rang.

Or, les résultats obtenus de cette manière ne coïncident pas avec ceux qui découlent de la formulation du problème adoptée au paragraphe 3.

9 - INCIDENCE PRATIQUE DE LA METHODE

Pour illustrer l'influence pratique de la méthode, nous utiliserons encore l'exemple caractérisé par les données (7) et (12).

Nous avons alors :

$$F(x) = 1 - e^{-x}$$

$$E(X_i) = \sum_{j=1}^{j=i} \frac{1}{n + 1 - j}$$

Les quantités ξ_i et ξ'_i du paragraphe précédent ont par ailleurs pour expression :

$$\xi_i = \ln \frac{n}{n - i}$$

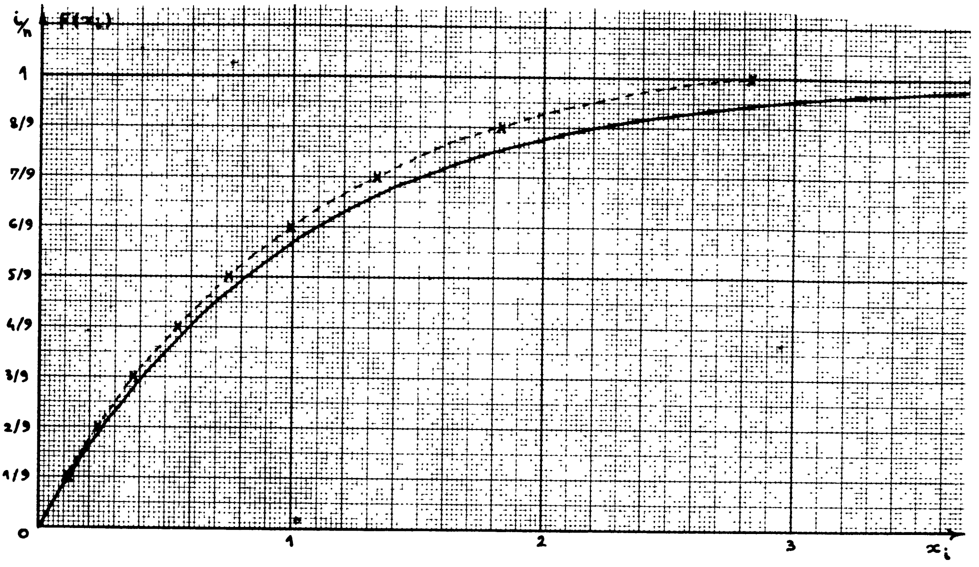
$$\xi'_i = \ln \frac{n + 1}{n + 1 - i}$$

d'où les valeurs numériques récapitulées dans le tableau ci-dessous :

i	$E(X_i)$	ξ_i	ξ'_i
1	0,111	0,118	0,105
2	0,236	0,251	0,223
3	0,379	0,405	0,357
4	0,546	0,588	0,511
5	0,748	0,811	0,693
6	0,996	1,099	0,916
7	1,329	1,504	1,204
8	1,829	2,197	1,609
9	2,829	∞	2,303

Représentons tout d'abord sur un graphique à échelles arithmétiques la fonction de répartition F(x) et l'ensemble des points $[E(X_i) ; i/n]$ qui constitue la "courbe" moyenne pour un échantillon de taille n = 9 (graphique 4).

La même comparaison se retrouve dans le tracé de la "courbe" moyenne résultant des points $[E(X_i) ; \xi_i]$ dans l'anamorphose rendant rectiligne la courbe F(x) - (graphique 5).



graphique 4

Représentons également la "courbe" moyenne résultant des points $[E(X_i)\ ;\ \xi'_i]$ dans l'anamorphose utilisant ξ'_i (graphique 6).

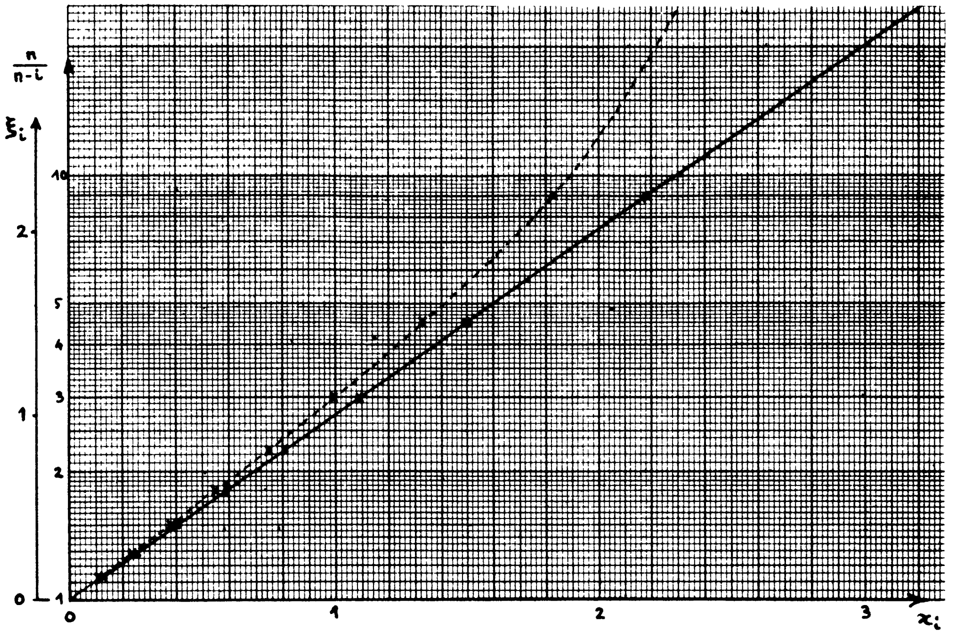
Les allures différentes des courbes, sur chacun de ces graphiques, constituent un premier aspect de l'incidence pratique de la méthode utilisée.

Par ailleurs, les écarts observés sur les graphiques 5 et 6 se traduiraient dans le cas de graphiques établis sur la base du modèle paramétrique constitué par la loi exponentielle de moyenne m inconnue, par des écarts dans les estimations de m ; en effet, si l'on avait obtenu dans un échantillonnage des valeurs de x_i coïncidant précisément avec leurs espérances mathématiques - soit $v_i = E(X_i)$ - on aurait, ayant conclu à une linéarité admissible, tracé une droite passant au mieux par les points représentatifs (points de la "courbe" moyenne) et estimé la valeur du paramètre m par l'inverse de la pente de cette droite, ce qui aurait conduit à un écart d'estimation de l'ordre de 10 % par rapport à la valeur m = 1 utilisée en fait pour le tracé de ces graphiques.

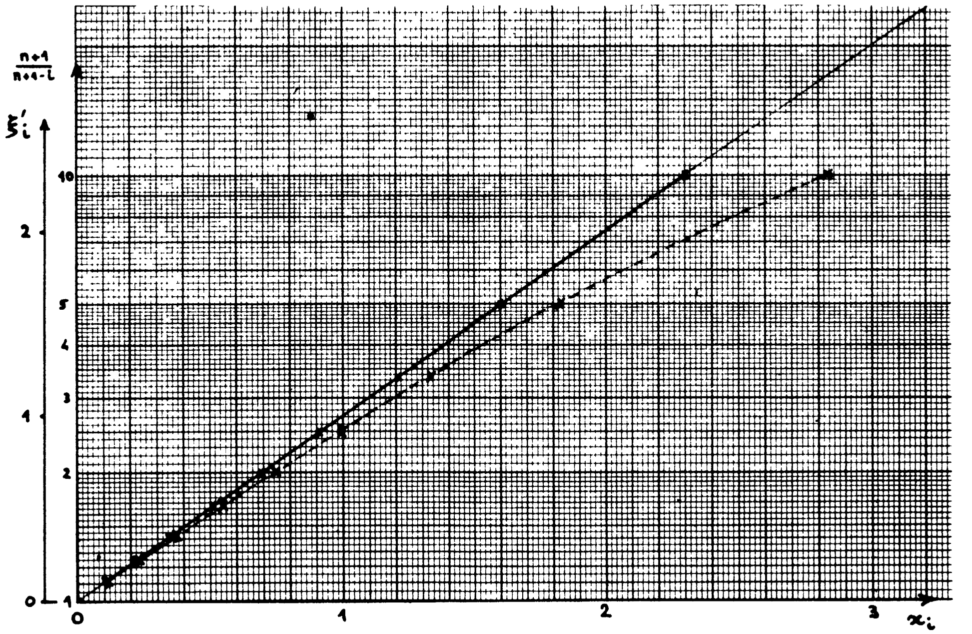
Tel est le second aspect de l'incidence pratique de la méthode utilisée.

Remarque

La taille d'échantillon n = 9 introduite dans les exemples a été choisie petite de façon à limiter le nombre des valeurs à considérer. Il est évidemment souhaitable d'être à même d'utiliser dans la pratique des échantillons plus importants. Les écarts que nous venons d'observer deviennent alors moins sensibles du fait de la convergence des différentes méthodes ; précisons pourtant que cette convergence est assez lente : ainsi le rapport de ξ'_i à $E(X_i)$, qui pour i = n = 9 a pour valeur $\frac{2,303}{2,829} = 0,814$, n'atteint encore pour i = n = 100 que la valeur $\frac{4,615}{5,187} = 0,890$.



graphique 5



graphique 6

10 - CONCLUSION

L'influence de la méthode sur les résultats de l'étude graphique peut n'être pas négligeable, en particulier en ce qui concerne l'estimation des valeurs numériques de paramètres.

Si qualitativement cette conclusion peut sembler voisine d'une évidence, son importance pratique nous a paru justifier, d'une part, que la nature de cette influence soit précisée par les considérations théoriques qui constituent l'essentiel de cet article, et, d'autre part, que l'attention soit attirée sur la nécessité de tenir compte, dans l'expression des résultats, de la méthode par laquelle ceux-ci ont été obtenus.

ANNEXE I

$$g(x_i) = \frac{1}{m} \frac{\left[1 - e^{-\frac{x_i}{m}}\right]^{i-1} \left[e^{-\frac{x_i}{m}}\right]^{n-i+1}}{B(i\ ;\ n-i+1)}$$

d'où :

$$E(X_i) = m \frac{\int_0^\infty \left[1 - e^{-\frac{x_i}{m}}\right]^{i-1} \left[e^{-\frac{x_i}{m}}\right]^{n-i+1} \frac{x_i}{m}\, d\left(\frac{x_i}{m}\right)}{B(i\ ;\ n-i+1)}$$

$$E(X_i) = m \frac{\int_0^\infty (1 - e^{-u})^{i-1} (e^{-u})^{n-i+1}\, u\, du}{B(i\ ;\ n-i+1)}$$

Or :

$$d[(1 - e^{-u})^{i-1} (e^{-u})^{n-i+1} u] = \left[\begin{array}{l} (i-1)\,(1 - e^{-u})^{i-2} (e^{-u})^{n-i+2}\, u\, du \\ -\ (n-i+1)\,(1 - e^{-u})^{i-1} (e^{-u})^{n-i+1} u\, du \\ +\ (1 - e^{-u})^{i-1} (e^{-u})^{n-i+1}\, du \end{array}\right.$$

d'où, en tenant compte de ce que :

$$[(1 - e^{-u})^{i-1} \quad (e^{-u})^{n-i+1}\, u]_0^\infty = 0$$

$$\int_0^\infty (1 - e^{-u})^{i-1} (e^{-u})^{n-i+1}\, u\, du = \left[\begin{array}{l} \dfrac{i-1}{n-i+1} \displaystyle\int_0^\infty (1 - e^{-u})^{i-2} (e^{-u})^{n-i+2}\, u\, du \\ +\ \dfrac{B(i\ ;\ n-i+1)}{n-i+1} \end{array}\right.$$

De cette relation, on tire la relation de récurrence :

$$E(X_i) = m \frac{\int_0^\infty (1 - e^{-u})^{i-2} (e^{-u})^{n-i+2}\, u\, du}{\frac{n-i+1}{i-1} B(i\ ;\ n-i+1)} + \frac{m}{n-i+1}$$

$$= m \frac{\int_0^\infty (1 - e^{-u})^{i-2} (e^{-u})^{n-i+2} u \, du}{B(i - 1 \; ; \; n - i + 2)} + \frac{m}{n - i + 1}$$

$$\boxed{E(X_i) = E(X_{i-1}) + \frac{m}{n - i + 1}}$$

Par ailleurs, la valeur de $E(X_1)$ étant :

$$E(X_1) = m \frac{\frac{1}{n^2} \Gamma(2)}{B(1 \; ; \; n)} = \frac{m}{n}$$

on a donc :

$$\boxed{E(X_i) = m \sum_{j=1}^{j=i} \frac{1}{n + 1 - j}}$$

ANNEXE II

Rappelons que :

$$\boxed{E(X_i) = m \sum_{j=1}^{j=i} \frac{1}{n + 1 - j}}$$

Par ailleurs, la quantité ξ_i telle que :

$$F(\xi_i) = 1 - e^{-\frac{\xi_i}{m}} = \frac{i}{n}$$

a pour valeur :

$$\boxed{\xi_i = m \ln \frac{n}{n - i}}$$

Considérons alors la somme donnant $E(X_i)$. Nous pouvons écrire :

$$\frac{E(X_i)}{m} = \frac{1}{n} + \frac{1}{n - 1} + \frac{1}{n - 2} + \dots + \frac{1}{n - (i - 1)}$$

$$= \left[1 . \frac{1}{n}\right] + \left[\frac{1}{1 - \frac{1}{n}} \cdot \frac{1}{n}\right] + \left[\frac{1}{1 - \frac{2}{n}} \cdot \frac{1}{n}\right] + \dots + \left[\frac{1}{1 - \frac{i - 1}{n}} \cdot \frac{1}{n}\right]$$

Cette expression correspond à la somme des aires des rectangles dessinés sur la figure 10 ci-dessous qui traduit graphiquement le fait que si u est une variable discontinue telle que $\Delta u = \frac{1}{n}$, les termes successifs de la somme précédente peuvent se mettre sous la forme $\left[\frac{1}{1 - u} \cdot \Delta u\right]$, u variant de 0 à $\frac{i - 1}{n}$.

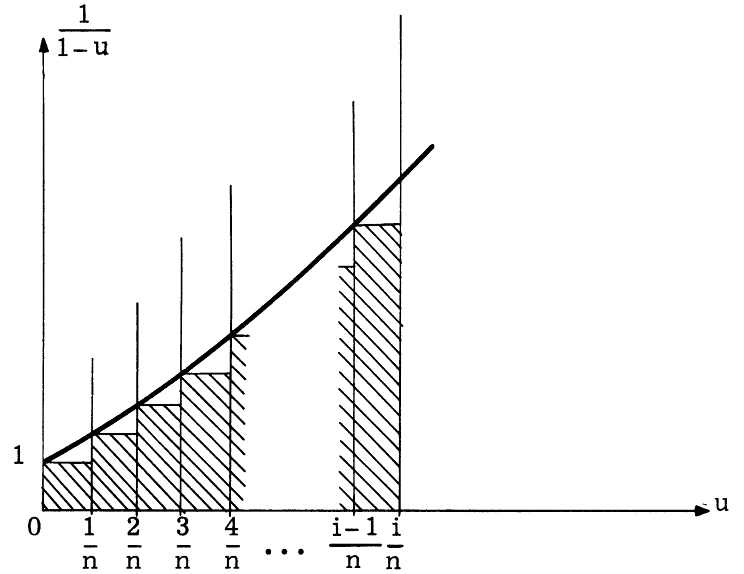


Figure 10

Lorsque n $\longrightarrow \infty$, on a :

$$\lim \left[\frac{E(X_i)}{m}\right] = \int_0^{\frac{i}{n}} \frac{1}{1-u}\, du = -\ln\left(1 - \frac{i}{n}\right) = \ln \frac{n}{n-i}$$

d'ou :

$$\boxed{\lim\, [E(X_i)]_{n\to\infty} = \xi_i}$$

Il y a donc bien convergence, lorsque la taille de l'échantillon croît indéfiniment, de l'ensemble des points d'abscisses $E(X_i)$ et d'ordonnées $\frac{i}{n}$ ("courbe" moyenne pour un échantillon de taille n) vers la courbe de répartition F(x).

Mais, pour n fini, la "courbe" moyenne est distincte de F(x). La figure montre en outre que $E(X_i)$ est alors toujours inférieur à ξ_i.