REVUE DE STATISTIQUE APPLIQUÉE

J. MOTHES

Principes fondamentaux de la méthode statistique et applications industrielles (suite)

Revue de statistique appliquée, tome 3, nº 2 (1955), p. 13-26
<http://www.numdam.org/item?id=RSA_1955__3_2_13_0>

© Société française de statistique, 1955, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (http://www.sfds.asso.fr/publicat/rsa.htm) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

PRINCIPES FONDAMENTAUX DE LA MÉTHODE STATISTIQUE ET APPLICATIONS INDUSTRIELLES (Suite)

par

J. MOTHES

Ancien Élève de l'École Polytechnique,
Ingénieur au Département Commercial du Gaz de France

CHAPITRE IV

ESTIMATION DE LA QUALITÉ D'UN PRODUIT

CHAPITRE IV

ESTIMATION DE LA QUALITÉ D'UN PRODUIT

Des indications fournies au chapitre 3, il ressort qu'une des phases fondamentales de l'analyse statistique consiste à passer d'une distribution observée à la distribution théorique susceptible de lui correspondre. Cela fait, rien ne s'oppose en effet à ce que l'on considère la distribution observée comme le résultat d'un échantillonnage opéré dans la population hypothétique infinie distribuée conformément à la loi de probabilité choisie comme modèle de référence. On est alors à même de raisonner sur un modèle mathématique bien défini et d'en tirer, en sens inverse, des conclusions d'ordre pratique.

Les distributions usuelles les plus fréquemment rencontrées en pratique présentent l'intérêt d'être entièrement définies par un petit nombre de paramètres. La distribution normale est, par exemple, **entièrement définie** par sa moyenne et son écart-type, la distribution de Poisson par sa moyenne. Pour passer d'une distribution observée à une distribution théorique, on est donc obligatoirement amené, une fois choisi le type de la distribution théorique de référence, à estimer les valeurs des paramètres qui définissent entièrement cette distribution théorique à partir des données constituant la distribution observée. Les problèmes d'estimation présentent de ce fait un caractère fondamental sur lequel il paraît inutile d'insister davantage.

On se propose de décrire, au cours des pages qui vont suivre, les principes de l'estimation statistique. Auparavant, il n'est pas sans intérêt d'analyser ce que représente le raccordement d'une distribution observée à une distribution théorique du point de vue des notions mêmes de "qualité" et "d'homogénéité" d'une fabrication.

IV. 1. - NOTIONS DE QUALITÉ ET D'HOMOGÉNÉITÉ D'UNE FABRICATION

Le raccordement d'une distribution observée à une distribution théorique est autorisé par la loi des grands nombres.

Ayant observé, par exemple, des résultats de fabrication distribués comme en IV-1, on conçoit la distribution théorique correspondante comme le système

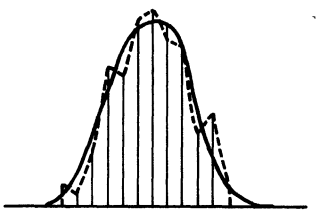

Fig. IV. 1.

limite des fréquences qui seraient observées dans l'hypothèse où le processus de fabrication, (à l'origine de celle-ci), serait capable de produire une infinité de pièces "dans des conditions de marche identiques aux conditions existantes lors de la fabrication du lot examiné". Raisonnant de la sorte, on constate que la distribution théorique de référence **résume entièrement** , du point de vue de la propriété examinée, l'influence -sur la fabrication- du jeu de l'ensemble des facteurs de production mis en oeuvre pour réalisation de cette fabrication. Cela étant, cette distribution est très étroitement liée à la notion même de "**qualité**" du produit fabriqué.

Qu'est-ce, en effet, que la qualité d'un produit donné? C'est évidemment un ensemble de conditions imposées à une ou plusieurs des propriétés du produit en question. Par conséquent, pour juger de la qualité d'une fabrication, trois opérations sont nécessaires :

a) Déterminer les propriétés à retenir comme "critères de qualité",

b) Imposer (éventuellement) des conditions à ces propriétés,

c) Comparer les résultats effectifs de la fabrication -du point de vue des propriétés retenues- aux conditions imposées à ces dernières.

Dès l'abord, il saute aux yeux que l'opération (c) implique la connaissance de la (ou des) distribution théorique de référence relative à la (ou aux) propriété examinée (dans la fabrication prise en considération).

Par ailleurs, les opérations (a) et (b) n'échappent pas aussi complètement qu'on peut le croire à première vue au domaine de la Statistique.

Si l'opération (c) est pratiquement résolue une fois connue la distribution théorique de référence de la propriété étudiée dans la fabrication examinée, encore faut-il être en mesure de définir effectivement cette distribution théorique à partir de la distribution des résultats des essais opérés en pratique. Cette distribution théorique est d'un usage d'autant plus commode qu'elle est susceptible d'être résumée par un très petit nombre de valeurs caractéristiques.

Dans ces conditions, si, au stade de l'opération (a), on a le choix entre deux critères de qualité équivalents mais dont l'un conduit, en fabrication, à une distribution théorique plus commode à manier que l'autre, on a toujours intérêt à retenir celui-là.

Au stade de l'opération (b), il est d'autre part souvent utile de se demander si les conditions imposées aux critères de qualité pris en considération sont absolument impératives. On constate souvent que ces conditions sont fixées, sans tenir compte des possibilités pratiques de fabrication, à des niveaux ne présentant, du point de vue des usagers, aucun caractère de stricte nécessité. Dans la mesure où les méthodes statistiques permettent de définir les possibilités techniques de fabrication, il peut être utile de tenir compte de leurs enseignements : c'est ce qui a été fait en Angleterre dans l'Industrie des Matières Plastiques

Toutes ces remarques sont aisées à clarifier à l'aide de quelques petits schémas.

Considérons d'abord, pour un produit donné, deux critères de qualité c_1 et c_2, équivalents, mais dont l'un est distribué, dans les fabrications réalisées en pratique, conformément à la figure IV-2 C_1 et l'autre, conformément à la figure IV-2 C_2 (distribution normale).

Fig. IV. 2. C_1.

Fig. IV. 2. C_2.

Dans l'hypothèse où, techniquement, les deux critères de qualité c_1 et c_2 sont équivalents, il est évident qu'on a intérêt (opération a) à retenir c_2 , la distribution théorique normale étant simple (elle est entièrement définie par deux caractéristiques sa moyenne et son écart-type) et bien connue.

Cette évidence devient flagrante si, au stade de l'opération (b), des conditions techniques impératives impliquent, pour que la fabrication soit "bonne", que les valeurs de c_2 tombent entre deux limites ϕ_1 et ϕ_2. Il est en effet très aisé de définir la position de la distribution théorique normale par rapport à l'intervalle ϕ_1 et ϕ_2. Du point de vue pratique, les consignes de fabrication à donner pour enregistrer un minimum de rebuts, sont par ailleurs les plus simples possibles : il suffit de régler la fabrication **au centre** de l'intervalle ϕ_1 ϕ_2 (Fig. IV-3).

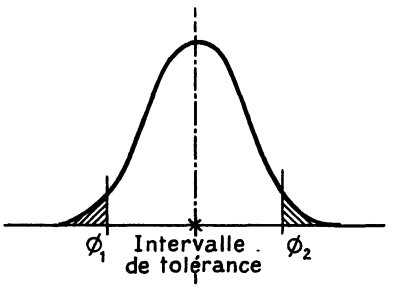

Fig. IV. 3

Avec une distribution dissymétrique comme celle de c_1 -l'aurait-on parfaitement définie- ces deux questions seraient, évidemment, beaucoup plus difficiles à traiter.

La prise en compte de la distribution théorique de référence d'une fabrication permet également une meilleure compréhension de la notion d'**homogénéité**

En pratique industrielle on a l'habitude de lier cette notion à celle de "plus ou moins grande dispersion" des résultats de fabrication. En présence de deux fabrications analogues F_1 et F_2 (Figure IV-4) et constatant que la dispersion des résultats issus de F_1 est inférieure à la dispersion des résultats issus de F_2, on dira, par exemple, que la fabrication F_1 est moins homogène que la fabrication F_2 . Or, cette conception entraîne bien des difficultés, dès qu'on veut définir la notion d'homogénéité autrement que sous forme relative. A la limite, en effet, une "fabrication homogène" (du point de vue d'une propriété donnée) ne pourrait être qu'une fabrication dont tous les individus seraient identiques, fabrication inenvisageable en pratique.

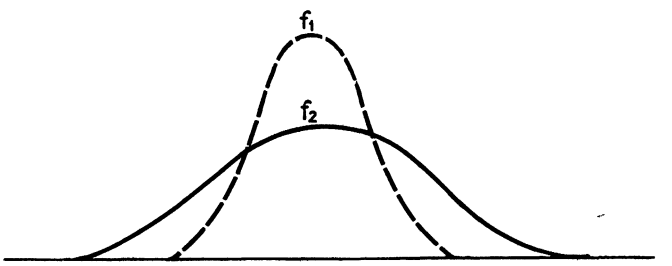

Fig. IV. 4

Il paraît beaucoup plus commode de relier la notion d'homogénéité d'une fabrication à la notion de stabilité des conditions de marche du processus de production à l'origine de la fabrication considérée. Revenant aux fabrications F_1 et F_2 , imaginons que F_1 et F_2 soient deux lots d'un même produit, fabriqués par

le même processus, en deux intervalles de temps successifs. Nous dirons que F_1 est homogène car le fait d'avoir caractérisé la distribution théorique de référence revient indirectement à caractériser des conditions de fabrication stables. De façon analogue, nous dirons également que F_2 est homogène.

La différence entre F_1 et F_2 prouve toutefois que, d'une fabrication à l'autre, il y a eu modification des conditions de fabrication. Dans ces conditions, nous dirons que le mélange de F_1 et F_2 est hétérogène.

IV. 2. - RÉSULTATS FONDAMENTAUX DE LA THÉORIE DE L'ÉCHANTILLONNAGE

Le problème de l'estimation statistique, tel que nous l'avons défini, consiste à passer des données constituant la distribution observée en pratique aux valeurs des paramètres qui définissent la distribution théorique de référence de la fabrication examinée, cela en considérant la distribution pratique comme le résultat d'un échantillonnage opéré au sein de la population (hypothétique) infinie de référence.

Pour résoudre ce problème, au moins dans les cas classiques, il est nécessaire de connaître les propriétés des échantillons extraits de populations distribuées conformément aux lois théoriques les plus couramment rencontrées en pratique, c'est-à-dire les résultats fondamentaux de la théorie de l'échantillonnage.

Tous les résultats de cette théorie reposent sur une hypothèse TRES IMPORTANTE, à savoir que les échantillons extraits de la population le sont "AU HASARD". C'est pourquoi nous reviendrons ici, en premier lieu, sur cette notion "d'Echantillonnage au hasard" déjà abordée au Chapitre II.

Echantillonner au hasard, c'est -avons-nous dit- effectuer les prélèvements de façon telle que tous les éléments de la population aient, avant chacun des tirages, autant de chances les uns que les autres d'être prélevés. Dans la pratique, cela n'équivaut absolument pas à "prélever n'importe comment" : il est, au contraire, parfois difficile de respecter rigoureusement cette règle du jeu pourtant impérative.

Dans certains cas, le problème ne présente pas de graves difficultés. Ayant, par exemple, à faire des prélèvements dans une caisse contenant de petites composantes métalliques, on peut admettre -si le contenu de la caisse est tout entier à portée de la main- qu'on respecte suffisamment le hasard dès qu'on s'astreint à prélever un peu partout après brassage, les pièces destinées à constituer l'échantillon.

Dans d'autres cas, un peu plus compliqués, on peut encore imaginer des plans d'échantillonnage susceptibles, bien que très simples, de donner satisfaction. Soit, par exemple, à procéder sur des tubes de verre de 1,5 mètre de long, à des mesures de diamètres. Il suffit de diviser une règle de 1,5m en 100 intervalles numérotés de 1 à 100 et de disposer d'un sac de billes également numérotées de 1 à 100. Ayant alors à effectuer -disons- 5 mesures sur un tube, on tirera 5 numéros dans le sac et, une fois superposé le tube à la règle graduée, on mesurera les 5 diamètres des zones du tube en regard des numéros correspondants de la règle.

On peut évidemment se demander s'il ne suffirait pas, dans ce cas, de laisser toute latitude à l'opérateur d'effectuer les mesures à sa guise, un peu partout. En fait, toutes les expériences faites à ce sujet montrent que de telles mesures, effectuées sans précaution, ne sont nullement réparties au hasard.

Dans certains cas, enfin, il faut envisager des plans d'échantillonnage beaucoup plus compliqués. De tels plans impliquent l'usage de tables dites de "nombres au hasard". La technique consiste à numéroter de 1 à N les éléments constituant la population. Ayant alors décidé de prélever un échantillon de n éléments, on lit n nombres dans la table et on prélève effectivement les n éléments portant les numéros correspondants.

Dans la Revue de Statistique Appliquée (Vol II, n° 4, page 188) est décrit un matériel de démonstration d'assemblages dont l'usage implique certaines précautions en vue de respecter les règles du hasard. A titre anecdotique son cas mérite d'être examiné.

Ce matériel a pour objet de montrer que les tolérances à adopter sur la longueur d'un ensemble de pièces mises bout à bout n'est pas la somme des tolérances fixées sur chacune des pièces.

Les élements de base de l'appareillage de démonstration sont quatre populations de pièces métalliques. Chaque population comprend des pièces de longueurs variables, la distribution du nombre de pièces en fonction des longueurs étant normale (Fig. IV-5), et le principe de la démonstration consiste à prélever au hasard une pièce dans chaque population, à mettre les quatre pièces ainsi prélevées bout à bout dans un bati spécial, cela fait à constater que la longueur totale des séries de quatre pièces ainsi constituées varie dans un intervalle beaucoup moins large que l'intervalle obtenu par addition des pièces les plus petites d'une part et les plus longues d'autre part des quatre populations.

Pour respecter la règle du jeu théorique de prélèvements au hasard de chaque pièce dans la population à laquelle elle appartient, on peut imaginer le procédé très simple consistant à mettre chaque population dans un sac puis à tirer une pièce du sac après brassage. En fait, on constate expérimentalement que cette méthode est très mauvaise. L'opérateur sent à la main la longueur des pièces métalliques et, inconsciemment, a tendance à ne saisir que les pièces ni trop courtes ni trop longues.

Pour respecter les règles du "hasard", on a été amené à passer par l'intermédiaire de quatre populations de billes, chaque bille correspondant à une pièce et portant, en guise de numéro, la longueur de la pièce dont elle est l'image. Les prélèvements portent ainsi sur des séries de quatre billes, étant entendu qu'on introduit dans le bâti les pièces qui leur correspondent.

Après ces remarques extrêmement importantes sur les conditions pratiques de l'échantillonnage au hasard, il paraît suffisant d'énumérer les principaux résultats de la théorie de l'échantillonnage susceptibles d'être utilisés dans la suite de cet ouvrage. Ces résultats sont les suivants :

a) Population distribuée de façon quelconque

Les **moyennes** m' des échantillons extraits au hasard d'une population **quelconque** de moyenne m et d'écart-type σ se distribuent suivant une loi de moyenne m et d'écart-type $\sigma / \sqrt{n}$

De plus, cette loi **tend rapidement vers la loi normale lorsque l'importance M de l'échantillon augmente.**

Graphiquement, ce résultat est schématisé sur la figure IV-6.

b) Population distribuée de façon normale

1) Les **moyennes** m' des échantillons extraits au hasard d'une population normale de moyenne m et d'écart-type σ sont elles-même distribuées **suivant une loi normale de moyenne m et d'écart type** $\sigma / \sqrt{n}$ **et cela quelle que soit l'importance de l'échantillon**

Graphiquement, le résultat est schématisé sur la figure IV-7.

2) Les distributions des **Variances** des **écarts-types** et des **étendues** des échantillons extraits au hasard d'une population normale sont parfaitement connues. D'expressions relativement compliquées, on les a "tabulées" pour les mettre sous une forme commode à utiliser en pratique. Dans la Revue de Statistique Appliquée (Vol II, n° 3, page 127) figure, par exemple, la "fonction de distribution" de la

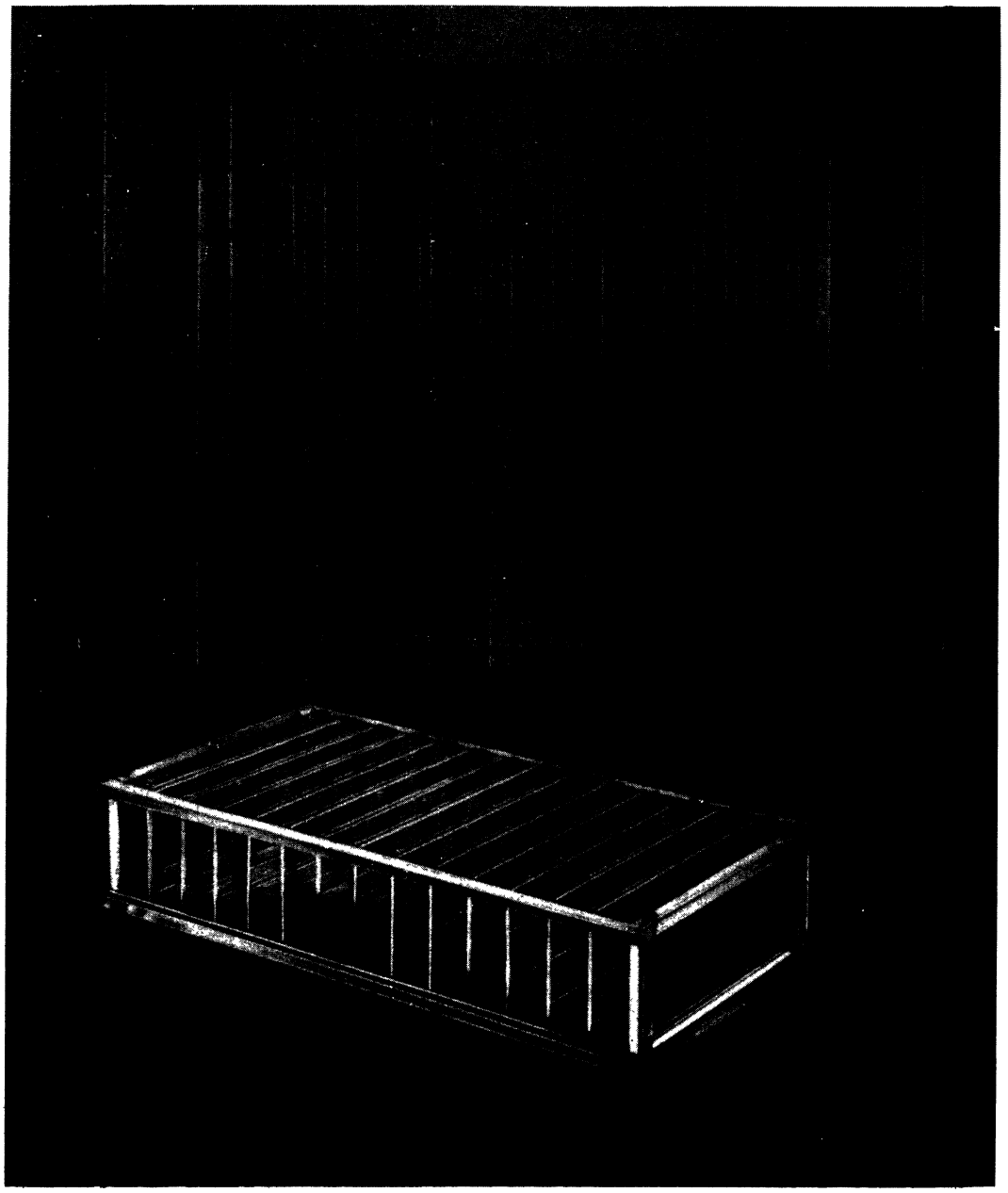

Fig. IV. 5.

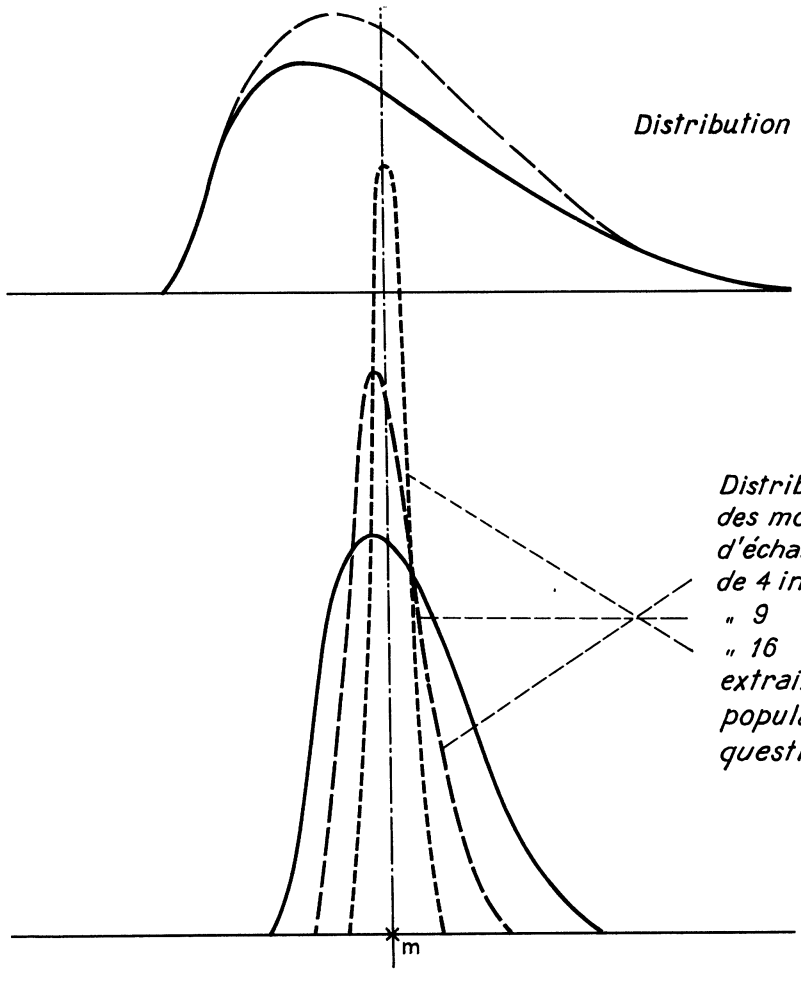

Fig. IV. 6.

Distribution de la population

Distribution des moyennes d'échantillons de 4 individus " 9 " " 16 " extraits de la population en question.

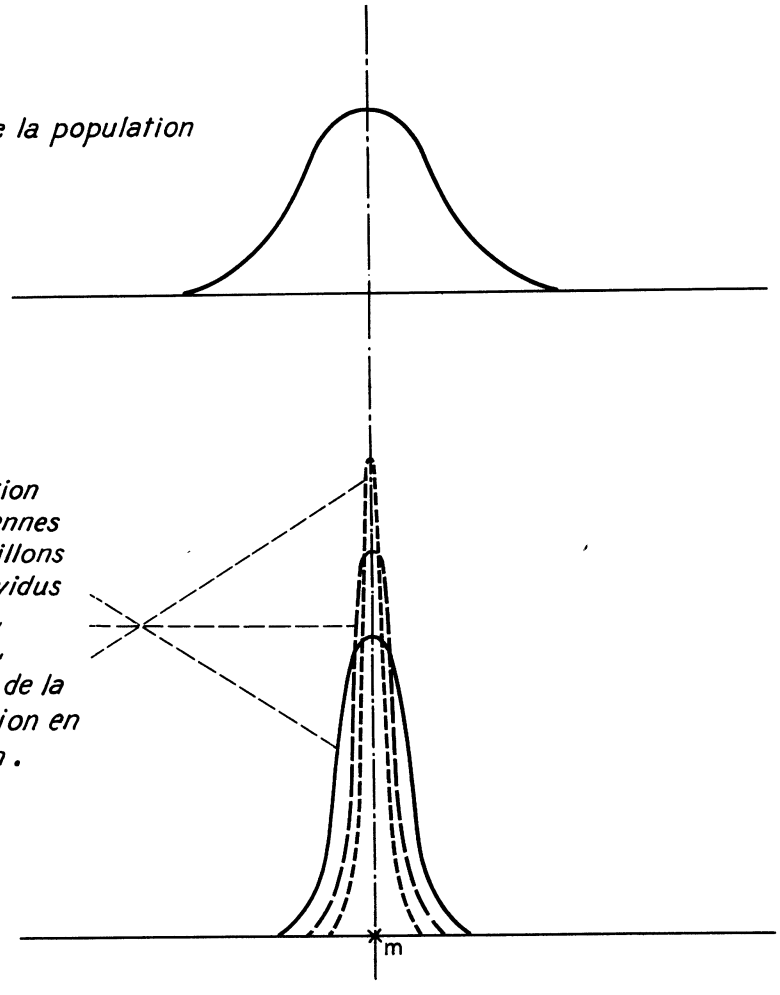

Fig. IV. 7.

loi des étendues (rapportées à l'écart-type) pour des échantillons de taille n variant de 2 à 12 unités (1).

c) Population dont les éléments sont distingués en "convenables et "défectueux"

La proportion de déchets présentée par des échantillons extraits au hasard d'une population dont les éléments sont distingués en "convenables" et "défectueux" est distribuée suivant une **loi binomiale** dont une forme limite est la **loi de Poisson** .

Les deux distributions ont été étudiées au cours du Chapitre II.

IV. 3. - LES TECHNIQUES D'ESTIMATION

Examinant, au début du chapitre III, la nature des raisonnements statistiques, on a longuement insisté sur l'introduction, dans les raisonnements, de ce qu'il est convenu d'appeler les "connaissances a priori". On a indiqué, à cette occasion, qu'il était préférable de se passer de telles connaissances lorsqu'elles ne reposaient pas sur des éléments d'appréciation objectifs.

Cela posé, on dispose de deux catégories de techniques d'estimation selon qu'on utilise ou non des données a priori. Nous désignerons les premières sous le vocable de "techniques a prioristes", les secondes sous le nom de "techniques d'estimation par estimateurs".

a) Les techniques à prioristes

Elles reposent sur un théorème célèbre -le théorème de Bayes- évoqué, sans le citer, au cours de la section III.2.

Ce théorème peut s'énoncer de la façon suivante :

"Si l'on sait qu'un évènement B s'est produit et si l'on sait que cet évènement ne peut être provoqué que par un certain nombre de causes A_1 A_2 ... A_m dont on connaît les probabilités a priori, les probabilités d'intervention des causes A_1 A_2 ... A_m , **compte-tenu de l'arrivée effective de l'évènement B**, s'écrivent :

$$Pr^B(A_1) = \frac{Pr(A_1) \cdot Pr^A(B)}{\sum_{i=1}^{m} Pr(A_i)\ Pr^{A_i}(B)} ,$$

$$Pr^B(A_2) = \frac{Pr(A_2) \cdot Pr^{A_2}(B)}{\sum_{i=1}^{m} Pr(A_i)\ Pr^{A_i}(B)} ,$$

$$\vdots$$

$$Pr^B(A_m) = \frac{Pr(A_m) \cdot Pr^{A_m}(B)}{\sum_{i=1}^{m} Pr(A_i)\ Pr^{A_i}(B)}$$

en désignant par :

- $Pr^B(A_i)$: La probabilité, ayant observé l'évènement B, que cet évènement ait été provoqué par la cause A_i :

- $Pr^{A_i}(B)$: La probabilité de l'évènement B lorsque la cause A_i se manifeste;

(1) La notion de "fonction de répartition d'une distribution" ou plus simplement de "fonction de distribution" a été fournie à la section II.1. De la fonction, on passe immédiatement à la loi proprement dite.

- $Pr(A_i)$ · La probabilité a priori de la cause A_i ;

et par $\sum_{i=1}^{m} Pr(A_i) \cdot Pr^{A_i}(B)$ la somme :

$$Pr(A_1)\ Pr^{A_1}(B) + Pr(A_2)\ Pr^{A_2}(B) + \ldots + Pr(A_m)\ Pr^{A_m}(B)\ .$$

Pour ne pas entrer dans le détail mathématique de l'application de ce thèorème qui déborderait le cadre de cet exposé sommaire, on se contentera de concrétiser les formules qui précèdent sur un exemple simple (et d'ailleurs artificiel).

Imaginons qu'à la réception d'un lot de pièces mécaniques, on reçoive une lettre du fournisseur indiquant :

1° - Qu'il a livré un second lot du même type à un autre client :

2° - Que ces deux lots, de type analogue, comprenaient respectivement 2% et 4% de déchets (d'après les vérifications à 100% opérées chez lui);

3° - Qu'il ne sait malheureusement plus sur lequel de ses deux clients a été dirigé le lot le meilleur.

et demandons-nous comment déterminer si le lot receptionné est, ou non, celui qui comprend les 2% de rebuts.

Il faut commencer par prélever un échantillon dans le lot examiné. Admettons qu'on prélève 250 pièces et qu'on détecte 8 rebuts parmi les 250 pièces c'est l'évènement (B). D'autre part, les causes possibles de cet évènement sont, soit le fait d'avoir affaire au lot à 2% de rebuts (A_1), soit le fait d'avoir affaire au lot à 4% de rebuts (A_2).

Cela posé, appliquons le théorème de Bayes.

D'après la lettre du fournisseur on a évidemment autant de chances de se trouver en présence du lot à 2% de déchets que de se trouver en présence du lot à 4% de déchets. Les **probabilités à priori** de ces deux éventualités sont donc respectivement :

$$Pr(A_1) = Pr(2\%) = 1/2$$
$$Pr(A_2) = Pr(4\%) = 1/2$$

En second lieu, dans l'hypothèse où l'on se trouve en présence du lot à 2% de rebuts, la probabilité de trouver 8 rebuts dans un échantillon de 250 pièces s'écrit d'après la loi de Poisson (cf. Table n° 5; np = 5; k = 8).

$$Pr^{A_1}(B) = Pr^{2\%}\ (8 \text{ rebuts}) = 0.0653$$

et de façon analogue si l'on se trouve en présence d'un lot à 4% de déchets (cf. table n° 5, np = 10 k = 8)

$$Pr^{A_2}(B) = Pr^{4\%}\ (8 \text{ rebuts}) = 0.1126$$

par conséquent

$$Pr^{B}(A_1) = Pr^{8\text{rebuts}}(2\%) = \frac{0,5 \times 0,0653}{0,5 \times 0,0653 + 0,5 + 0,1126} = 0,37$$

$$Pr^{B}(A_2) = Pr^{8\text{rebuts}}(4\%) = \frac{0,5 \times 0,1126}{0,5 \times 0,0653 + 0,5 \times 0,1126} = 0,63$$

On constate donc que, compte tenu des propriétés de l'échantillon extrait du lot, la probabilité d'avoir affaire au lot à 4% de rebuts est plus grande que la probabilité d'avoir affaire au lot à 2% de rebuts.

Il est donc logique de retenir 4% comme estimation de la proportion de rebuts présentées par le lot.

b) Les techniques d'estimation par estimateur

Lorsqu'on ne dispose pas de connaissances a priori dignes de confiance, il est plus rationnel d'utiliser les techniques d'estimation par estimateur.

On désigne sous le terme **d'estimateur** une fonction particulière des résultats de l'échantillon, fonction douée de certaines propriétés sur lesquelles on va revenir.

Soit un échantillon de taille n. Cet échantillon conduit à n observations

$x_1 \quad x_2 \ldots\ldots x_i \ldots\ldots x_n$

par conséquent toute fonction

$f(x_1 \quad x_2 \ldots\ldots\ldots\ldots x_n)$

de ces résultats prend une certaine valeur.

Si l'on recommence l'échantillonnage, la fonction f prend, dans chaque cas, une nouvelle valeur -du fait des fluctuations d'échantillonnage- mais, connaissant la distribution des x dans la population, on sait déterminer la distribution des valeurs possibles de f. C'est ainsi, par exemple, que la distribution de

$$f(x_1 \quad x_2 \ldots\ldots\ldots x_n) = \frac{x_1 + x_2 + \ldots\ldots\ldots x_n}{n} \text{ (Moy. arith.)}$$

pour des échantillons extraits d'une distribution normale de moyenne m et d'écart-type σ est elle-même une distribution normale de moyenne m et d'écart-type $\sigma / \sqrt{n}$ (cf. section IV-2).

Cela étant, on dit que la fonction $f(x_1 \quad x_2 \ldots\ldots\ldots x_n)$ des résultats de l'échantillon est un estimateur d'une caractéristique c donnée de la population lorsque la moyenne et l'écart-type de la distribution de f tendent respectivement vers c et o pour des échantillons d'importance n croissant indéfiniment.

De façon plus imagée, f est un estimateur de c si la distribution de f se déforme lorsque n croît, comme indiqué sur la figure IV.8 (passage de (1) en (2).

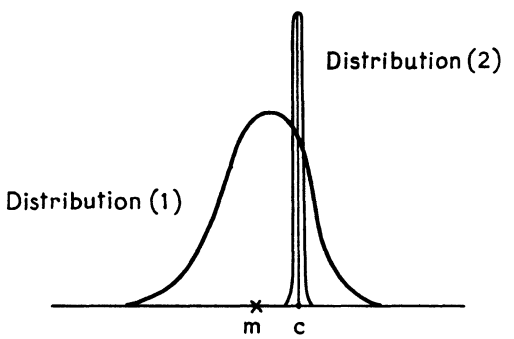

Fig. IV. 8.

Lorsque la moyenne de la distribution de f **est égale à** c quel que soit n (figure IV.9), on dit que f est un **estimateur sans biais**.

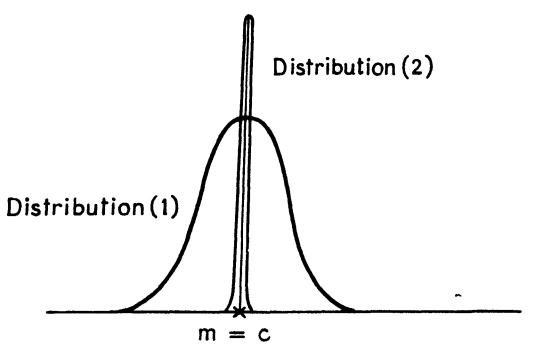

Fig. IV. 9.

Le problème de l'estimation par estimateur revient, on le voit, à trouver une fonction f répondant aux conditions précitées. Et, parmi toutes les fonctions possibles, la meilleure est évidemment celle pour laquelle le passage de (1) en (2) est le plus rapide.

A titre indicatif, on peut remarquer dès maintenant qu'un estimateur de la moyenne d'une population distribuée de façon quelconque n'est autre que la moyenne de l'échantillon prélevé dans cette population. On a en effet signalé en IV.2 que la distribution de m' tend dans ce cas, pour n croissant, vers une distribution normale

- de moyenne m
- d'écart-type $\sigma / \sqrt{n} \rightarrow 0$.

Lorsque la population initiale est normale, cet estimateur devient un estimateur **sans biais** , les moyennes m' étant distribuées normalement et ayant m pour moyenne quel que soit n.

Signalons encore que, pour une population distribuée normalement, la variance de l'échantillon n'est pas un estimateur sans biais de la variance de cette population. L'estimateur sans biais s'écrit en effet :

$$\frac{n}{n-1} \cdot \sigma'^2$$

Dans le tableau suivant figurent quelques estimations classiques. On notera que les estimations des caractéristiques considérées sont désignées, comme les caractéristiques elles-mêmes, avec, toutefois, une caractéristique en exposant.

Type de population	Caractéristique estimée.	Estimation.
Quelconque	Moyenne	$m^* = m'$
Normale	Moyenne	$m^* = m'$
"	Variance	$\sigma^{2*} = \frac{n}{n-1} \cdot \sigma'^2$
Binomiale (Eléments distingués en bons et mauvais).	Proportion de déchets	$p^* = p'$ en désignant par p' la proportion de déchets trouvée dans l'échantillon.

Bien qu'elle soit assez délicate à saisir, on ne saurait ici passer sous silence la notion **d'intervalle de confiance** d'une estimation.

Pour l'expliciter aisément, il est commode de raisonner graphiquement.

Soit n l'importance de l'échantillon et f l'estimateur de la caractéristique c examinée, f étant distribué conformément au schéma IV.10.

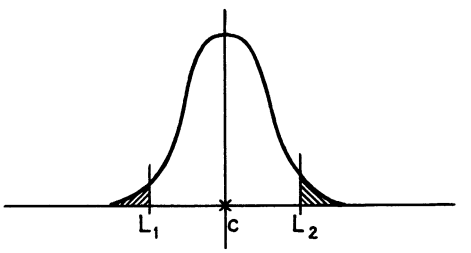

Fig. IV. 10.

Par définition de l'estimateur f, la moyenne de **l'ensemble** des valeurs possibles de f tend ou est égale) à c mais, en pratique, la valeur de f effectivement obtenue, résultat d'un seul échantillonnage, n'est pas forcément égale ou très voisine de c.

Tout ce que l'on peut dire, connaissant la distribution de f, c'est que le f obte-

nu, **pour un seuil de probabilité** $1-\alpha$ donné (section III.3), tombe entre les deux limites correspondantes :

$$\left.\begin{array}{l} L_1 = c - h_1 \\ L_2 = c + h_2 \end{array}\right\} \text{ } h_1 \text{ et } h_2 \text{ étant fournis par le calcul (1)}$$

A ce stade, nous constatons que nous avons, par définition :

$$c - h_1 \leqslant f \leqslant c + h_2$$

dans la proportion $1-\alpha$ des cas. Or, chaque fois que nous avons cela, nous avons

$$f - h_2 \leqslant c \leqslant f + h_1$$

Ainsi l'intervalle $\left[(f - h_2), (f + h_1)\right]$ **recouvre la vraie valeur c de la caractéristique examinée dans la proportion** $1-\alpha$ **des cas**. Cet intervalle est appelé **l'intervalle de confiance** (au seuil de probabilité $1-\alpha$).

On remarquera - c'est très important - qu'on ne dit pas que c a la probabilité $1-\alpha$ de se trouver dans l'intervalle de confiance. On dit que l'intervalle de confiance recouvre la valeur c dans le proportion $1-\alpha$ des cas. On ne peut logiquement, en effet, lier une probabilité à c. Sans doute la valeur c est-elle inconnue mais, elle n'en existe pas moins de façon parfaitement définie : elle n'est donc pas une variable aléatoire.

C'est exactement le contraire de ce qui se passe dans les techniques d'estimation a priori. Celles-ci considèrent, en effet, c comme une variable aléatoire et mettent d'ailleurs en jeu sa loi de probabilité (ce sont les Pr (A_i))

IV. 4. - CONCLUSIONS D'ORDRE PRATIQUE

Les résultats qui précèdent n'ont pas fait l'objet de développements mathématiques. De tels développements nous auraient, en effet, conduit assez loin. Dans le cadre de cette brochure, il nous a paru beaucoup plus important de parler de l'estimation statistique en termes -disons- philosophiques. C'est dans la mesure où l'on commence à saisir la différence de nature existant entre les techniques à prioristes et les techniques par estimateur que l'on commence à devenir statisticien.

Des indications précédentes, un certain nombre de conclusions d'ordre pratique doivent être tirées.

Il est, en premier lieu très clair que la précision d'une estimation est d'autant meilleure, c'est-à-dire que l'intervalle de confiance est d'autant réduit, que l'importance de l'échantillon est grande. L'étendue de l'intervalle de confiance est, en effet, liée à la dispersion de la distribution des f. Or, par définition, cette dispersion diminue quand n augmente. Mais, du même coup, il devient aussi très clair que, dans le cadre des prélèvements (exhaustifs ou assimilables à des prélèvements exhaustifs) qui sont à la base de la théorie, l'importance de la population initiale **n'a rien à voir à l'affaire**.

On voit encore, dans bien des entreprises décider de l'importance du prélèvement en fonction de la taille du lot examiné. Dans la plupart des cas, c'est un non-sens.

Il y a d'autre part lieu de noter que la décision d'appliquer les règles d'estimation par intervalle de confiance revient, une fois le seuil $1-\alpha$ choisi, à accepter de se tromper, en moyenne, dans la proportion α des cas. C'est en effet, par définition, dans cette proportion α des cas que l'intervalle de confiance **ne recouvre pas** la vraie valeur c cherchée.

Il devient alors évident que, la détermination du seuil de probabilité - considéré comme négligeable - n'est pas statistique mais dépend de la plus ou moins grande gravité des conséquences des erreurs possibles.

(1) Nous n'expliciterons pas, ici, les calculs.