JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

GÉRARD PILÉ

La statistique appliquée au contrôle de fabrication

Journal de la société statistique de Paris, tome 88 (1947), p. 421-436

<http://www.numdam.org/item?id=JSFS_1947__88__421_0>

© Société de statistique de Paris, 1947, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (http://publications-sfds.math.cnrs.fr/index.php/J-SFdS) 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/

VII

VARIÉTÉ

La statistique appliquée au contrôle de fabrication

Avant-propos

Le 17 mai 1939, M. Rosenfeld faisait à la Société de Statistique de Paris une communication sur l'application de diagrammes (*Control charts*) au contrôle statistique dans l'industrie.

On ne saurait trop aujourd'hui attirer de nouveau l'attention ici sur ces problèmes et les solutions sans cesse renouvelées qui y sont apportées dans les pays anglo-saxons. Il semble bien en effet qu'en Angleterre et surtout aux États-Unis, la guerre ait été cause d'un progrès décisif dans le perfectionnement des méthodes et la formation de spécialistes entraînés à les utiliser.

Lors de l'entrée en guerre des U. S. A., où progrès scientifique et technique vont de pair, il n'y avait cependant qu'un nombre relativement minime d'établissements industriels à appliquer les ressources de la statistique au contrôle de leurs fabrications et seules quelques personnes, appartenant au monde universitaire pour la plupart, étaient capables, sans études spéciales, de diffuser un enseignement à ce sujet.

L' « American Standard Association » commença en 1941 avec la publication de deux brochures intitulées : *Guide for quality Control* et *Control about method of analysing data ;* suivirent d'autres publications américaines et anglaises. Quelques cours furent organisés vers la même époque aux U. S. A., dus à des initiatives privées; ils reçurent le soutien financier de l'État. Tel fut le cas de l'Université de Stanford.

En 1941 et 1942, le Bureau de l'Artillerie de l'Armée fit des efforts sérieux pour utiliser les méthodes statistiques modernes, le fait le plus saillant ayant été l'adoption de cartes de contrôle dans la fabrication des blindages. La marine, elle, malgré une adoption plus tardive de ces méthodes, poussa leur usage beaucoup plus loin que les autres services.

Les applications connurent bientôt une plus grande extension par suite de

l'adoption par l' « Office of Production Research and development » d'un plan d'ensemble d'adaptation des méthodes statistiques de contrôle aux industries de guerre. Sous l'influence de l'E. S. M. W. P. (*Engineering-Science and Management war program*), les cours se multiplièrent. En 1945, on comptait plus de 8.000 personnes ayant suivi pendant la guerre des cours de statistique appliquée au contrôle des fabrications.

Dans toutes les branches : aviation, aciérie, optique, on a pu constater les progrès accomplis dans le rendement grâce à la statistique. Des résultats particulièrement spectaculaires ont été par exemple obtenus dans le contrôle de pièces ouvrées par des machines semi-automatiques : l'alerte étant rapidement donnée dès que se produit le moindre déréglage, le personnel se sentant surveillé, accroît son rendement et est attentif au réglage.

RAPPEL DES PRINCIPALES MÉTHODES UTILISÉES POUR LE CONTRÔLE STATISTIQUE DES FABRICATIONS

Pour déterminer si des objets fabriqués satisfont ou non à certaines conditions requises, on peut utiliser l'un des trois procédés suivants :

1° L'examen 100 %. Cette méthode fréquemment inutilisable en pratique pour des raisons faciles à concevoir, ne donne pas, contrairement à l'opinion courante, une garantie 100 % pour l'utilisateur : les gabarits subissant une usure rapide, doivent eux-mêmes être fréquemment corrigés et il est impossible d'éliminer complètement les causes d'erreur personnelle.

L'examen 100 % ne pose pas, à proprement parler, de problème statistique; un problème statistique se pose, à partir du moment où l'on doit juger de la qualité ou décider de la viabilité du lot sur la base de l'information incomplète que peut donner l'examen d'échantillons extraits du lot. A cet effet, les principales méthodes suivantes peuvent être envisagées;

2° La mise en œuvre de cartes de contrôle sur lesquelles on reporte les résultats de l'examen d'échantillons sélectionnés au cours de la fabrication. On ne reviendra pas sur ce procédé qui a déjà fait l'objet d'une étude dans ce journal, comme il a été rappelé au début;

3° Les tests sur des échantillons de la production, constitués au hasard.

Dans ce qui va suivre, on s'intéressera uniquement aux procédés rangés dans la catégorie 3°, une mention toute spéciale étant réservée aux procédés modernes dits « de score ».

A l'origine et à la base de toutes ces méthodes est celle de simple échantillonnage dont voici les principes essentiels :

On détermine un taux maximum probable de pièces défectueuses et on majore chaque lot à livrer du nombre de pièces lui correspondant (point de vue du producteur).

Pour cela, on prélève un échantillon de n pièces dont d se révèlent défectueuses. Cette opération est assimilée à un tirage non exhaustif dans une urne qui contiendrait une proportion θ de boules différenciées, par exemple, par la couleur. La probabilité d'obtenir d pièces défectueuses dans un lot de n pièces est alors donnée par :

$$P\,d = C^d_n\,\theta^d\,(1-\theta)^{n-d}$$

On sait que si n est suffisamment grand, la loi binômiale a pour limites acceptables :

Si θ n'est pas trop petit : une loi normale de moyenne np d'écart type $\sqrt{n\theta(1-\theta)}$.

Si θ est très voisin de 0 ($n\theta$ lui-même petit) : une loi de Poisson de moyenne $n\theta$ et d'écart type $\sqrt{n\theta}$.

Dans le premier cas, il y aura par exemple 95 chances sur 100 pour que d soit tel que :

$$\left|\frac{d - n\theta}{\sqrt{n\theta(1-\theta)}}\right| \leqslant 1{,}96.$$

L'égalité permet, ayant d, de déterminer la valeur maxima de θ. L'emploi de la formule précédente ou de la formule analogue associée à une probabilité d'erreur de 1 °/₀₀ sera grandement facilité par l'emploi d'abaques.

On peut aussi convenir que si la valeur maxima trouvée pour θ excède une valeur θ_0, le lot doit être refusé (point de vue de l'utilisateur). Mais quelle que soit l'application que l'on fasse de la méthode, on détermine à l'avance le nombre de pièces à examiner en fonction de la précision avec laquelle on veut estimer la proportion de pièces défectueuses. On fera souvent un premier essai sur un nombre réduit d'objets, ce n'est que si la proportion critique θ_0 est comprise par exemple dans l'intervalle $\frac{d_1}{n_1} \pm 2\sigma\left(\frac{d_1}{n_1}\right)$ que l'on fera une inspection plus complète.

Ces méthodes ne sont pas rigoureuses, on n'est pas en général placé dans les conditions de tirages non exhaustifs; en outre les lois que l'on utilise ne sont que des lois limites qui obligent pour être correctement appliquées, à faire des échantillonnages importants.

Une méthode bien plus rigoureuse, à peu près ignorée en France, est appliquée depuis de longues années déjà, à la « Bell Telephone Company » aux États-Unis, elle est connue sous le nom de méthode de Dodge et Romig (ses inventeurs). Son principe est le suivant :

Soit un lot de N pièces comprenant une proportion θ de défectueuses, un échantillon de n pièces est examiné, on en trouve d défectueuses. La probabilité d'un tel échantillon est donnée par :

$$P_{d,\, n,\, \theta,\, N} = \frac{C_{N\theta}^{d}\, C_{N-N\theta}^{n-d}}{C_N^n}$$

En effet, on a C_N^n tirages exhaustifs possibles, or il y a en tout $N\theta$ pièces défectueuses; donc $C_{N\theta}^{d}$ façons de tirer d pièces défectueuses
et $C_{N-N\theta}^{n-d}$ façons de tirer $n-d$ pièces convenables.

La fonction : $F(c, \theta, N, n) = \sum_{d=0}^{d=c} P_{d;\, n,\, \theta,\, N}$.

Donne la probabilité d'obtenir dans un échantillon de n pièces extraites du lot N un nombre de pièces défectueuses inférieur ou égal à c.

Un lot de pièces ne doit être considéré comme acceptable que si $\theta < \theta_0$, θ_0 étant une proportion fixée à l'avance. La connaissance de la fonction F permet, lorsque l'on se donne N, θ_0, n, de déterminer c de sorte qu'il n'y ait pratiquement pas de chances d'obtenir $d > c$.

Si le lot est acceptable, on améliore la qualité du lot en remplaçant les d pièces défectueuses par d pièces conformes. θ devient alors $\frac{N\theta - d}{N}$, à cette nouvelle proportion est associée la proportion P_d et l'espérance mathématique de cette proportion s'exprime par :

$$\bar{p} = \Sigma_{d=0}^{d=c} \frac{\theta N - d}{N} P_{d,\, n,\, \theta,\, N}$$

Pour N, n, c donnés, $\bar{p}$ a une valeur maximum $\bar{P}_L$.

Ces fonctions, étudiées par Dodge et Römig et calculées pour les valeurs les plus usuelles des variables, sont à la base de la méthode dont on se borne ici à indiquer le schéma.

Il faut avoir une idée préalable de la proportion de pièces défectueuses d'après l'expérience antérieure que l'on peut avoir sur la qualité des fabrications ou les données fournies par les cartes de contrôle.

On fait un premier échantillonnage sur n_1 pièces.

Si $d_1 < c_1$, le lot est accepté.

Si $d_1 > c_1$, on prélève un second échantillon de n_2 pièces parmi lesquelles on trouve d_2 défectueuses.

Si $d_1 + d_2 < c_2$, le lot est accepté.

Si $d_1 + d_2 > c_2$, on fait une inspection 100 % et on remplace les pièces défectueuses par des bonnes.

L'utilité d'un second échantillonnage est de réduire dans une large mesure l'éventualité d'une inspection 100 %; en effet les pièces défectueuses que l'on découvre dans la deuxième hypothèse étant remplacées par des conformes, le risque du consommateur s'en trouve réduit. On comprend qu'il soit alors nécessaire d'élargir l'échantillonnage et ceci d'autant plus que dans le premier échantillonnage d_1 a été trouvé voisin de c_1, pour pouvoir se prononcer avec la même chance d'erreur (1 °/₀₀ est la proportion la plus usuelle aux U. S. A. à cet égard).

La méthode de Dodge et Romig est particulièrement précise et sûre, mais elle convient mal lorsque la qualité est sujette à des fluctuations fréquentes et importantes à cause de la connaissance préalable qu'elle implique, de la qualité; elle s'applique utilement aux étapes de la fabrication pour éviter le passage de pièces défectueuses aux stades ultérieurs d'utilisation.

LES MÉTHODES DE SCORE

L'exposé qui va suivre rend compte d'une étude parue dans le numéro de septembre 1945 du Journal de l'Association américaine de Statistique. Due au professeur Abraham Wald de Colombia University, cette étude constitue un exposé théorique simplifié de la « Sequential method of sampling » que l'on peut traduire : « Méthode d'échantillonnages successifs ». Mais à cette dénomination, malgré tout assez peu explicite, nous lui préférerons celle qui a été donnée par Bornard : « Méthodes de Score ».

Ce mot suggère en effet bien mieux le caractère particulier aux méthodes en question, lesquelles présentent une analogie avec les jeux ne prenant fin que

lorsqu'une certaine marge d'avance est acquise à un parti et où : durée et nombre d'avantages ne sont pas déterminés à l'avance.

Une règle va être admise (procédé d'échantillonnage), pour conclure l'examen de chaque échantillon par l'une des trois décisions suivantes : accepter le lot (1), le refuser (2), ou recommencer un autre échantillonnage, le processus ne prenant fin que lorsque l'application du procédé conduit à prendre l'une des deux premières résolutions.

Supposons que les observations successives soient des observations indépendantes faites sur une même population et que la distribution de la variable x considérée soit connue à l'exception d'un seul paramètre Θ pouvant être la moyenne, la variance ou une autre caractéristique de la population.

Dans le cas qui nous occupe, la population est le lot et pour chaque unité du lot, la variable x est susceptible de prendre deux valeurs : 1 si l'unité est défectueuse; 0 dans le cas contraire; la caractéristique inconnue est alors la moyenne θ. On pourrait considérer la valeur x de quelque grandeur caractéristique associée aux objets du lot, comme le poids, la résistance à la pression; alors x serait une variable continue susceptible de prendre toutes les valeurs d'un certain intervalle. On supposerait, par exemple, que la distribution de x est normale.

1. *Courbe caractéristique du procédé.* — Désignons par L (θ) la probabilité pour chaque valeur de θ, que l'opération se termine avec l'adoption; L (θ) constitue ce que l'on appelle la courbe caractéristique d'opération associée au procédé (« o. c. curve »).

Précisons, sur un exemple particulièrement simple, comment peut se présenter le calcul d'une telle fonction.

Imaginons le procédé suivant : si les 25 premières unités inspectées ne présentent pas de défaut, le lot est accepté; si pour quelque valeur $m \leqslant 25$ la même unité inspectée est défectueuse, on s'en tient là et le lot est refusé. Soit θ la proportion inconnue de pièces défectueuses, la probabilité pour que le lot soit accepté s'exprime par :

$$L(\theta) = (1 - \theta)^{25}$$

2. *Nombre moyen d'observations requis par un procédé d'échantillonnage.* — Le nombre n d'observations requis par un plan déterminé d'échantillonnage n'est pas déterminé à l'avance. Si l'on répète l'opération donnée à titre d'exemple dans le paragraphe précédent, on obtiendra en général différentes valeurs de n.

Puisque par hypothèse il y a un seul paramètre inconnu θ dans la distribution de x, l'espérance mathématique de n, E (n) sera seulement fonction de θ.

Dans le cas de l'exemple précédent, le nombre moyen d'observations requis est donné par :

$$E_\theta(n) = \Sigma_{m=1}^{m=25} m\,\theta\,(1 - \theta)^{m-1} + 25\,(1 - \theta)^{25}.$$

Principes directeurs du choix d'un échantillonnage.

Considérons le cas où la préférence pour l'admission ou le rejet du lot dépend de θ de la façon suivante : il existe une valeur θ' telle que si $\theta < \theta'$ nous préférons la décision (1) et si $\theta > \theta'$ nous préférons la décision (2), si $\theta = \theta'$ notre position est indifférente.

D'autre part, si $\theta < \theta'$, notre préférence pour (1) croît lorsque décroît θ et si $\theta > \theta'$, notre préférence pour (2) croît avec θ. On a naturellement intérêt à ce que la valeur de L (θ) soit aussi voisine que possible de 1 pour $\theta < \theta'$ et de 0 pour $\theta > \theta'$. Ceci définit donc une forme idéale de la courbe caractéristique; plus on désirera s'en rapprocher et plus grand en général sera le nombre d'observations à faire.

Il sera possible en général de définir un intervalle limité par θ_0 et θ_1 de part et d'autre de θ' tel que pour toutes valeurs de θ extérieures à celui-ci on puisse considérer que l'erreur que l'on commet en ne prenant pas la décision correspondante ait une conséquence pratique.

Comme on veut réduire les chances d'erreur, on s'imposera des conditions telles que :

$$1 - L(\theta) \leqslant \alpha \qquad \text{pour } \theta \leqslant \theta_0$$

(3.1) $$\text{ou} \quad L(\theta) \leqslant \beta \qquad \text{pour } \theta \geqslant \theta_1$$

Une courbe type satisfaisant à ces conditions est donnée ci-dessous :

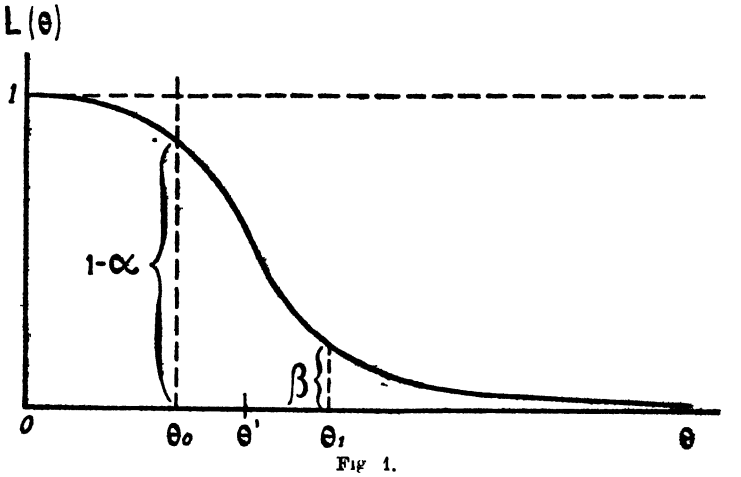

Fig. 1.

Les quatre quantités θ_0, θ_1, α et β ayant été choisies, on cherchera un procédé d'échantillonnage satisfaisant aux conditions (3-1) et tel que le nombre d'observations qu'il requiert soit aussi petit que possible, d'où l'intérêt de connaître $E_\theta(n)$ associé à chaque procédé.

4. *Généralisation.* — Généralisons au cas où la variable x est susceptible de valeurs quelconques. Pour une valeur θ nous appelons $f(x, \theta)$ la loi de distribution de x; $f(x, \theta)$ sera une densité de probabilité si x est continu; mais, pour simplifier l'exposé suivant, on va faire l'hypothèse que x est discontinu.

Pour toute valeur entière m, la probabilité d'obtenir un échantillon égal à celui observé est donnée par :

(4.1) $$P\,m = f(x_1, \theta)\, f(x_2, \theta) \ldots\ldots f(x_m, \theta).$$

Soient $P_0\,m$ et $P_1 m$ les probabilités pour θ_0 et θ_1.

A chaque stade d'observation, on calculera le rapport $\dfrac{P_1 m}{P_0 m}$.

Deux constantes A et B étant choisies $0 < B < 1 < A$.

(4.2) $$\text{Si} \quad B < \frac{P_1 m}{P_0 m} < A, \quad \text{on refait une autre observation;}$$

(4.3) Si $\frac{P_1 m}{P_0 m} \leqslant B$ on prend la décision (1)

(4.4) Si $\frac{P_1 m}{P_0 m} \geqslant A$ on prend la décision (2)

A et B sont pris de telle sorte que :

(4.5) $L(\theta_0) = 1 - \alpha$ et $L(\theta_1) = \beta$ (4.6)

Le problème de cette détermination sera discuté au paragraphe suivant. En ce qui concerne les calculs, il est préférable de prendre les logarithmes. Si on désigne par z_a la quantité $\log \frac{f(x_a, \theta_1)}{f(x_a, \theta_0)}$, la condition (4.2) écrite plus haut devient par exemple :

(4.7) $\log B < z_1 + ... z_m < \log A.$

On se contentera en général d'un calcul approché. Il résulte de l'inégalité (4.4) que, pour les échantillons conduisant à la décision (2), la probabilité d'obtenir un tel échantillon est au moins A fois plus grande lorsque $\theta = \theta_1$ que lorsque $\theta = \theta_0$. Mais la probabilité totale de tels échantillons est égale à la probabilité que la décision (2) soit prise. Cette dernière étant égale à :

$1 - \beta$ lorsque $\theta = \theta_1$

et α lorsque $\theta = \theta_0$.

on en déduit l'inégalité

(4.8) $1 - \beta \geqslant A\alpha$

ou $A \leqslant \frac{1-\beta}{\alpha}$

$\frac{1-\beta}{\alpha}$ est donc une limite supérieure de A.

On trouve de la même façon que

(4.9) $B \geqslant \frac{\beta}{1-\alpha}$

On prendra pratiquement les valeurs de A et B définies par les égalités (4.8) et (4.9).

La raison des inégalités (4.8) et (4.9) est la possibilité de l'observation de se terminer par

$$\frac{P_1 m}{P_0 m} > A \quad \text{ou} \quad \frac{P_1 m}{P_0 m} < B$$

Si $f(x, \theta_1)$ et $f(x, \theta_0)$ sont peu différentes, il est presque certain qu'une observation supplémentaire entraîne une légère modification du rapport $\frac{P_1 m}{P_0 m}$. Au stade final, ce dernier sera donc légèrement supérieur à A ou légèrement inférieur à B, lesquels seront donc très voisins de $\frac{1-\beta}{\alpha}$ et $\frac{\beta}{1-\alpha}$ respectivement.

Avantage de ce procédé. — Le procédé dont il vient d'être question a pour supériorité essentielle sur le procédé de simple échantillonnage de permettre une économie notable d'observations, économie que l'on peut calculer de la façon suivante :

On démontre (1) que, pour tout procédé d'échantillonnage satisfaisant aux conditions :

$$L(\theta_0) = 1 - \alpha$$
$$L(\theta_1) = \beta$$

(1) Une expression approchée de $E_\theta(n)$ est donnée plus loin (6,5).

on a :

$$(4.10)\quad E_{\theta_0}(n) \geqslant \frac{1}{E_0(z)}\left[(1-\alpha)\log\frac{\beta}{1-\alpha} + \frac{\beta}{1-\alpha}\log\frac{1-\beta}{\alpha}\right]$$

$$(4.11)\quad E_{\theta_1}(n) \geqslant \frac{1}{E_1(z)}\left[\beta\log\frac{\beta}{1-\alpha} + (1-\beta)\log\frac{1-\beta}{\alpha}\right]$$

où $E_0(z)$, $E_1(z)$ désignent les espérances mathématiques, pour θ, valeur vraie du paramètre, égal à θ_0 et θ_1 respectivement, de :

$$z = \log\frac{f(x, \theta_1)}{f(x, \theta_0)}$$

$E_{\theta_0}^{(n)}$ et $E_{\theta_1}^{(n)}$ sont d'ailleurs très voisins des seconds membres de ces inégalités dans le procédé des rapports successifs de probabilités. On démontre aussi que dans la pratique on peut dire que le nombre moyen d'observations est minimum pour le procédé lorsque $\theta = \theta_0$ et $\theta = \theta_1$.

Dans la méthode de simple échantillonnage, le nombre n d'observations à faire est déterminé à l'avance : si la moyenne arithmétique des observations est inférieure ou égale à une quantité choisie Δ, on accepte le lot; sinon on le refuse. Δ et n sont déterminés de telle sorte que la probabilité d'accepter le lot soit égale à :

$$\begin{cases} 1-\alpha & \text{pour } \theta = \theta_0 \text{ (1)} \\ \beta & \text{pour } \theta = \theta_1 \end{cases}$$

Appelons $n(\alpha\ \beta)$ ce nombre.

L'économie moyenne d'observations du procédé actuel sur le procédé de simple échantillonnage peut s'exprimer au moyen des expressions :

$$(4.12)\qquad 100\left[1 - \frac{E_{\theta_0}(n)}{n(\alpha, \beta)}\right] \quad \text{pour } \theta = \theta_0$$

$$(4.13)\qquad 100\left[1 - \frac{E_{\theta_1}(n)}{n(\alpha, \beta)}\right] \quad \text{pour } \theta = \theta_1$$

On a calculé $E_{\theta_0}^{(n)}$ et $E_{\theta_1}^{(n)}$ au moyen des formules approchées (4.10), (4.11) et (9.10) (l'erreur est pratiquement négligeable lorsque θ_0 et θ_1 sont voisins), dans l'hypothèse d'une distribution normale de la variable x.

Les expressions (4.12) et (4.13) sont indépendantes de θ_0 et θ_1; les nombres portés dans le tableau ci-dessous en résultent :

(1) Voir la formule indiquée dans la première partie de cette étude.

Économies moyennes d'observations, en %, permises par l'utilisation du procédé du rapport des probabilités successives (1).

α	0,01		0,02		0,03		0,04		0,05	
β	θ_0	θ_1	θ_0	θ_1	θ_0	θ_1	θ_0	θ_1	θ_0	θ_1
0,01	58	58	54	60	51	61	49	62	47	63
0,02	60	54	56	56	53	57	50	58	49	59
0,03	61	51	57	53	54	54	51	55	50	55
0,04	62	49	58	50	55	51	52	52	50	53
0,05	63	47	59	49	55	50	53	50	51	51

5. *Expression de la fonction* L (θ) *pour tout procédé utilisant le rapport* $\frac{P_1 m}{P_0 m}$.

Soit l'expression :

$$\left[\frac{f(x, \theta_1)}{f(x, \theta_0)}\right]^{h(\theta)}$$

(où $h(\theta)$ est différent de 0), telle que son espérance mathématique (pour θ : valeur vraie du paramètre) soit égale à 1 :

$$\Sigma_x f(x, \theta) \left[\frac{f(x, \theta_1)}{f(x, \theta_0)}\right]^{h(\theta)} = 1 \qquad (5.1)$$

On montre que, moyennant quelques restrictions sur $f(x, \theta)$, il existe une valeur réelle $h(\theta) \neq 0$ satisfaisant à (5.1).

Pour toute valeur de θ, la fonction :

$$\Phi(x, \theta) = f(x, \theta) \left[\frac{f(x, \theta_1)}{f(x, \theta_0)}\right]^{h(\theta)} \qquad (5.2)$$

est une fonction de distribution.

On voit, en élevant les membres des inégalités (4.2) à la puissance θ que le procédé correspondant à :

$$A, B, f(x, \theta_1), f(x, \theta_0)$$

est identique au procédé correspondant à :

$$A^{h(\theta)}, B^{h(\theta)}, \Phi(x, \theta), f(x, \theta)$$

Appliquant des résultats trouvés antérieurement à ce deuxième procédé, on obtient les égalités suivantes valables avec une bonne approximation :

$$A^{h(\theta)} = \frac{1 - \beta'}{\alpha'} \qquad (5.3)$$

$$B^{h(\theta)} = \frac{\beta'}{1 - \alpha'} \qquad (5.4)$$

où α' désigne la probabilité que soit prise la décision (2) lorsque $f(x, \theta)$ est la distribution vraie de x.

(1) On fait l'hypothèse que θ (paramètre inconnu) est la moyenne d'une distribution normale dont on connaît l'écart type.

On obtient de la sorte :

$$\alpha' = \frac{1 - B^{h(\theta)}}{A^{h(\theta)} - B^{h(\theta)}} = 1 - L(\theta)$$

(5.5) $$\text{D'où} \quad L(\theta) = \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}}$$

6. *Expression de la fonction $E_\theta^{n)}$ pour tout procédé utilisant le rapport $\frac{P_1 m}{P_0 m}$.*

Soit n le nombre d'observations requis par un tel procédé et N un entier assez grand pour que la probabilité de $n > N$ puisse être tenue pour négligeable.

On peut écrire :

(6.1) $$z_1 + \ldots\ldots + z_N = (z_1 + \ldots\ldots + z_n) + (z_{n+1} + \ldots\ldots + z_N)$$

$$\text{où} \quad z_a = \log \frac{f(x_a, \theta_1)}{f(x_a, \theta_0)}$$

Prenant l'espérance mathématique des deux membres :

(6.2) $$NE(z) = E(z_1 + \ldots\ldots + z_n) + E(z_{n+1} \ldots\ldots + z_N).$$

Puisque pour $a > n$ la variable z_a est distribuée indépendamment de n, on peut écrire :

(6.3) $$E(z_{n+1} + \ldots\ldots z_N) = E(N - n)\, E(z) = N\, E(z) - E(n)\, E(z)$$

De (6.2) et (6.3), on tire :

(6.4) $$E(n) = \frac{E(z_1 + \ldots\ldots + z_n)}{E(z)}$$

Faisons l'hypothèse que $f(x, \theta)$ soit la distribution vraie de x, alors :

$$E(n) = E_\theta(n).$$

Si l'on néglige la probabilité que le rapport $\frac{P_1 m}{P_0 m}$ excède les limites de A et B, la variable $(z_1 + \ldots\ldots + z_n)$ peut seulement prendre une des deux valeurs log A ou log B, avec les probabilités $1 - L(\theta)$ et $L(\theta)$ respectives.

d'où :

$$E(z_1 + \ldots\ldots + z_n) = L(\theta) \log B + [1 - L(\theta)] \log A$$

(6.5) $$\text{Finalement} \quad E_\theta(n) = \frac{L(\theta) \log B + [1 - L(\theta)] \log A}{E^\theta(z)}$$

7. *Applications au cas où une observation individuelle conduit à une classification dans une ou deux catégories :*

C'est par exemple le cas considéré au début de cette étude : on a un lot comprenant un grand nombre d'objets manufacturés; il s'agit de décider de son admission en fonction de la proportion d'unités défectueuses qu'il contient.

Désignons par dm le nombre d'unités défectueuses trouvées dans les m premières unités inspectées.

Alors la probabilité d'obtenir un échantillon égal à celui observé est donnée par :

(7.1) $$P_1 m = \theta_1^{dm} (1 - \theta_1)^{m - dm} \text{ pour } \theta = \theta_1$$

(7.2) et $$P_0 m = \theta_0^{dm} (1 - \theta_0)^{m - dm} \text{ pour } \theta = \theta_0$$

(7.3) Tant que $B < \left(\frac{P_1 m}{P_0 m}\right) < A$ nous faisons une observation additionnelle,

(7.4) (7.5), on s'arrête dès que $\frac{P_1 m}{P_0 m} \leqslant B$ ou $\frac{P_1 m}{P_0 m} \geqslant A$

En accord avec les résultats acquis dans le paragraphe 4.2, on considère les valeurs approchées de A et de B : $\frac{1 - \beta}{\alpha}$ et $\frac{\beta}{1 - \alpha}$.

Il est avantageux de réécrire les inégalités (7.3), (7,4), (7,5) d'une façon quelque peu différente; par exemple (7.3) devient en passant aux logarithmes :

(7.6)
$$\frac{\log \frac{\beta}{1-\alpha}}{\log \frac{\theta_1}{\theta_0} - \log \frac{1-\theta_1}{1-\theta_0}} + m \frac{\log \frac{1-\theta_0}{1-\theta_1}}{\log \frac{\theta_1}{\theta_0} - \log \frac{1-\theta_1}{1-\theta_0}}$$
$$< dm < \frac{\log \frac{1-\beta}{\alpha}}{\log \frac{\theta_1}{\theta_0} - \log \frac{1-\theta_1}{1-\theta_0}} + m \frac{\log \frac{1-\theta_0}{1-\theta_1}}{\log \frac{\theta_1}{\theta_0} - \log \frac{1-\theta_1}{1-\theta_0}}$$

On est conduit à considérer deux nombres A m et R m respectivement égaux aux nombres extrêmes de la double inégalité écrite plus haut.

A m est le nombre d'admission.

R m celui de rejet.

Si A m calculé n'est pas un nombre entier, on le remplace par l'entier immédiatement inférieur.

Si de même R m n'est pas entier, on lui substitue l'entier immédiatement supérieur.

Exemple numérique : $\theta_0 = 10$ $\alpha = 0{,}02$

$\theta_1 = 30$ $\beta = 0{,}03$

Les nombres d'admission et de rejet ont été calculés ci-dessous en même temps que sont donnés des résultats d'observations.

TABLEAU 2

n NOMBRE D'UNITÉS inspectées	Am NOMBRE d'admission	dm NOMBRE D'UNITÉS défectueuses observées	Rm NOMBRE de rejet
1	»	0	»
2	»	0	»
3	»	1	»
4	»	1	4
5	»	1	4
6	»	1	4
7	»	1	5
8	»	1	5
9	»	2	5
10	»	2	5
11	»	3	5
12	»	4	6
13	»	4	6
14	0	5	6
15	0	5	6
16	0	5	6
17	0	5	7
18	0	6	7
19	0	6	7
20	1	6	7
21	1	6	7
22	1		7
23	1		8
24	1	»	8
25	2	»	8

Courbe caractéristique du procédé.

Selon l'équation (5.5), cette courbe est donnée par l'équation :

$$L(\theta) = \frac{\left(\frac{1-\beta}{\alpha}\right)^{h(\theta)} - 1}{\left(\frac{1-\beta}{\alpha}\right)^{h(\theta)} - \left(\frac{\beta}{1-\alpha}\right)^{h(\theta)}} \tag{7.7}$$

où $h(\theta)$ est la racine de l'équation (5.1), comme dans le cas présent.

$$\begin{cases} f(x, \theta) = \theta \text{ pour } x = 1 \\ f(x, \theta) = 1 - \theta \text{ pour } x = 0 \end{cases}$$

L'équation (5.1) peut s'écrire :

$$\theta \left(\frac{\theta_1}{\theta_0}\right)^{h(\theta)} + (1-\theta)\left(\frac{1-\theta_1}{1-\theta_0}\right)^{h(\theta)} = 1 \tag{7.8}$$

On peut considérer $h(\theta)$ comme un paramètre et résoudre par rapport à θ. On obtient :

$$\theta = \frac{1 - \left(\frac{1-\theta_1}{1-\theta_0}\right)^{h}}{\left(\frac{\theta_1}{\theta_0}\right)^{h} - \left(\frac{1-\theta_1}{1-\theta_0}\right)^{h}}$$

Cette équation jointe à (7.7) permet le calcul des coordonnées de points de la courbe.

Courbe du nombre moyen d'observations du procédé.

On utilise la formule (6.5), où :

$$E_\theta(z) = \theta \log \frac{\theta_1}{\theta_0} + (1-\theta) \log \frac{1-\theta_1}{1-\theta_0}$$

On obtient ainsi :

$$E_\theta(n) = \frac{L\,\theta \log \frac{\beta}{1-\alpha} + (1 - L(\theta)) \log \frac{1-\beta}{\alpha}}{\theta \log \frac{\theta_1}{\theta_0} + (1-\theta) \log \frac{1-\theta_1}{1-\theta_0}}$$

forme type de cette courbe.

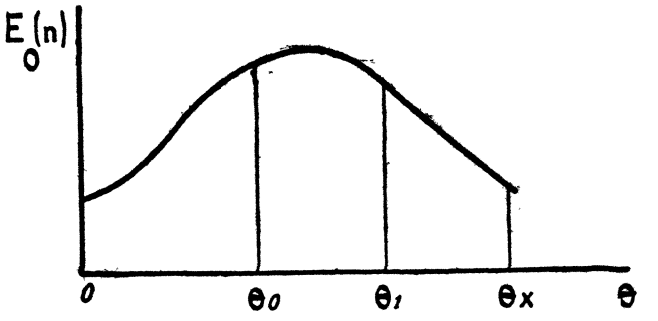

Fig. 2.

8. *Analyse en série de doubles dichotomies :*

Considérons deux distributions normales et soient p_1 et p_2 les probabilités inconnues de succès d'un simple essai, relatives à ces deux distributions.

Si $p_1 > p_2$ on préfère (1) et si $p_1 < p_2$ on préfère (2).

On va faire des observations par paires, chaque paire comprenant une observation portant sur chaque distribution. Une telle méthode s'impose par exemple, lorsqu'il s'agit de comparer l'efficacité de deux procédés de fabrication en fonction de la proportion d'unités défectueuses.

L'efficacité du procédé 1 de la fabrication peut se mesurer par le rapport des nombres d'unités non défectueuses et défectueuses :

$$k_1 = \frac{p_1}{1-p_1}$$

De façon similaire pour le procédé 2 :

$$k_2 = \frac{p_2}{1-p_2}$$

Le rapport $U = \frac{k_1}{k_2}$ peut être considéré comme une mesure raisonnable de la supériorité du procédé (2) sur le procédé (1).

Il sera possible en général de définir pour U deux valeurs U_0 et U_1 ($U_0 < U_1$) comme pour θ.

U_0 et U_1 étant choisis, les risques d'erreurs admissibles vont être concrétisés ainsi : la probabilité d'adopter le procédé (2) doit être au plus égale à une certaine valeur α pour $U \leqslant U_0$ et celle d'adopter le procédé (1) au plus égale à β pour $U \geqslant U_1$.

Afin de choisir l'un des deux procédés, on va seulement retenir les paires d'observations dissemblables. Précisons : si on désigne par 1 une unité non défectueuse et 0 une unité défectueuse, on va seulement considérer les couples (0,1) et (1,0). On justifie en effet que les observations (0,0) et (1,1) n'ont pas d'utilité.

La probabilité d'obtenir une paire 0,1 est égale à $(1 - p_1)\, p_2$; celle d'obtenir une paire 1,0 : $p_1\,(1 - p_2)$. de sorte que parmi les observations retenues, la probabilité θ d'obtenir un couple (0,1) est donnée par :

$$\theta = \frac{(1 - p_1)\, p_2}{p_1 (1 - p_2) + p_2 (1 - p_1)} \tag{8.1}$$

$$1 - \theta = \frac{p_1 (1 - p_2)}{p_1 (1 - p_2) + p_2 (1 - p_1)} \tag{8.2}$$

On peut écrire θ sous la forme :

$$\theta = \frac{U}{1 + U} \tag{8.3}$$

Le nombre m d'observations du paragraphe 7 doit être remplacé par celui de paires (0,1), (1,0) observées. Le nombre d'unités défectueuses par le nombre t_2 de paires (0,1) observées.

Les nombres d'admission et de rejet deviennent :

$$A_t = \frac{\log \frac{\beta}{1 - \alpha}}{\log U_1 - \log U_0} + t \frac{\log \frac{1 + U_1}{1 + U_0}}{\log U_1 - \log U_0}$$

$$R_t = \frac{\log \frac{1 - \beta}{1 - \alpha}}{\log U_1 - \log U_0} + t \frac{\log \frac{1 + U_1}{1 + U_0}}{\log U_1 - \log U_0}$$

Exemple numérique.

NOMBRE DE PAIRES (0,1) ou (1,0) observées	PAIRES observées	At NOMBRE d'admission	tx NOMBRE de paires (0,1) observées	Rt NOMBRE de rejet
1	(0,1)	»	1	»
2	(0,1)	»	2	»
3	(1,0)	»	2	»
4	(1,0)	»	2	»
5	(1,0)	0	2	»
6	(0,1)	1	3	»
7	(1,0)	1	3	»
8	(0,1)	2	4	»
9	(0,1)	3	5	»
10	(1,0)	3	5	»
11	(0,1)	4	6	»
12	(0,1)	5	7	»
13	(0,1)	5	8	13
14	(1,0)	6	8	14
15	(1,0)	7	8	14
16	(0,1)	7	9	15
17	(1,0)	8	9	16
18	(1,0)	9	9	16
19	»	9	»	17
20	»	10	»	18

C'est ainsi que l'échantillonnage s'arrête pour $t = 18$ avec l'adoption du procédé 1.

9. *Application dans le cas d'une distribution normale de la variable d'observation, la déviation type etant connue :*

La loi de distribution de x est donnée par :

$$f(x, \theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\theta)^2} \tag{9.1}$$

On note les valeurs successives observées de x, x_1, x_2.....

$$P_1 m = \frac{1}{(2\pi\sigma)^{\frac{m}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{a=1}^{a=m}(x_a-\theta_1)^2} \tag{9.2}$$

$$P_0 m = \frac{1}{(2\pi\sigma)^{\frac{m}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{a=1}^{a=m}(x_a-\theta_0)^2} \tag{9.3}$$

On doit avoir, pour faire une observation additionnelle :

$$B < \frac{e^{-\frac{1}{2\sigma^2}\sum_{a=1}^{a=m}(x_a-\theta)^2}}{e^{-\frac{1}{2\sigma^2}\sum_{a=1}^{a=m}(x_a-\theta_0)^2}} < A \tag{9.4}$$

en passant aux logarithmes :

$$\frac{\sigma^2}{\theta_1-\theta_0}\log\frac{\beta}{1-\alpha} + m\frac{\theta_0+\theta_1}{2} < \sum_{a=1}^{a=m} x_a < \frac{\sigma^2}{\theta_1-\theta_0}\log\frac{1-\beta}{\alpha} + m\frac{\theta_0+\theta_1}{2} \tag{9.5}$$

équation de la forme :

$$A\,m < \sum_{a=1}^{a=m} x_a < R\,m$$

Dans les applications numériques, les tables se présenteront ainsi :

m (NOMBRE d'observations)	Am (NOMBRE d'admission)	x (VALEUR observée)	Σx (VALEUR CUMULÉE de x)	Rm (NOMBRE de rejet)

Courbe caractéristique du procédé.

Si $\frac{\theta_1-\theta_0}{\sigma}$ est petit, une bonne approximation est donnée par :

$$L(\theta) = \frac{\left(\frac{1-\beta}{\alpha}\right)^{h(\theta)} - 1}{\left(\frac{1-\beta}{\alpha}\right)^{h(\theta)} - \left(\frac{\beta}{1-\alpha}\right)^{h(\theta)}} \tag{9.6}$$

où $h(\theta)$ est la racine non nulle de l'équation (5.1). Cette racine est égale à :

$$\frac{\theta_1+\theta_0-2\theta}{\theta_1-\theta_0} \tag{9.7}$$

Courbe du nombre moyen d'observations :

$$E_\theta(n) = \frac{L(\theta) \log \frac{\beta}{1-\alpha} + [1 - L(\theta)] \log \frac{1-\beta}{\alpha}}{E_\theta(z)} \tag{9.8}$$

avec :

$$z = \log \frac{f(x, \theta_1)}{f(x, \theta_0)} = \frac{1}{2\sigma^2} [2(\theta_1 - \theta_0) x + \theta_0^2 - \theta_1^2] \tag{9.9}$$

donc :

$$E_\theta(z) = \frac{1}{2\sigma^2} [\theta_0^2 - \theta_1^2 + 2(\theta_1 - \theta_0)\theta] \tag{9.10}$$

Gérard PILÉ.