REVUE DE STATISTIQUE APPLIQUÉE

E. MORICE

Les graphiques à échelles fonctionnelles du statisticien

Revue de statistique appliquée, tome 12, nº 3 (1964), p. 29-103
<http://www.numdam.org/item?id=RSA_1964__12_3_29_0>

© Société française de statistique, 1964, 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/

LES GRAPHIQUES A ÉCHELLES FONCTIONNELLES DU STATISTICIEN

par E. MORICE

Un certain nombre de papiers graphiques à échelles fonctionnelles, disponibles dans le commerce, constituent pour le statisticien des outils intéressants soit pour la solution approchée d'un problème, soit comme première approche vers une solution plus raffinée par le calcul.

Les principales échelles fonctionnelles utilisées en diverses combinaisons dans ces graphiques sont :

- l'échelle fonction logarithmique soit :

$$X = \log x \text{ ou même } X = \log_e x$$

- l'échelle fonction de répartition de la loi normale ou plutôt fonction inverse de cette loi

$$X = u \quad \text{pour} \quad x = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

- l'échelle logarithmique-logarithmique

$$X = \text{Log}(\text{Log } x)$$

Le principe général de graduation de ces échelles est le suivant : si la transformation utilisée est définie pour $X = f(x)$, chaque point coté x sur l'échelle sera situé à une distance de l'origine mesurée à l'unité graphique utilisée par :

$$X = f(x) \quad \text{ou} \quad X = f(x) + C^{te}$$

L'emploi de ces échelles fonctionnelles sur l'un des axes de coordonnées ou sur les deux permet de représenter certaines fonctions $y = f(x)$ par des droites dans le double but :

1/ de vérifier si un ensemble de couples d'observations (x, y) satisfait avec une approximation considérée comme satisfaisante à une liaison de la forme envisagée,

2/ d'estimer la valeur des coefficients numériques qui définissent la fonction $f(x)$ envisagée, c'est-à-dire les paramètres caractéristiques de cette loi.

Les papiers à échelles fonctionnelles permettront donc l'étude graphique de diverses distributions statistiques : suites ou ensembles d'obser-

vations à taux de variation constant ou à élasticité constante, loi de Pareto, lois normale et log-normale, loi binomiale, loi exponentielle, loi de Weibull.

D'autre part, un choix convenable de ces échelles permettra de construire des abaques plus satisfaisants pour leur commodité d'emploi et leur précision que ceux que l'on pourrait construire en utilisant directement les coordonnées x, y.

Les papiers semi-logarithmiques et log-log sont trop connus pour qu'il soit nécessaire de s'y arrêter longtemps.

Rappelons en cependant quelques propriétés. L'utilisation des graphiques dépend de la précision graphique des tracés et en particulier de celle de la mise en place des points : si l'on admet que cette erreur graphique est sensiblement constante, l'erreur relative pour une échelle donnée sera d'autant plus grande que la variable considérée sera plus petite, par contre dans le cas d'une échelle logarithmique cette erreur relative sera constante.

Une échelle logarithmique permettra de représenter simultanément de petites et de grandes valeurs de la variable avec la même erreur relative.

Une autre considération en justifie l'intérêt. Dans un graphique dessiné sur papier quadrillé ordinaire, la pente d'un segment AB, et par conséquent celle de la tangente à une courbe est (Fig. 1. 1) :

$$\frac{\Delta y}{\Delta x} \longrightarrow \frac{dy}{dx}$$

Fig. 1. 1

Δy et Δx étant mesurés compte tenu des unités graphiques choisies au mieux, cette pente caractérise la vitesse de variation de y par rapport à x (vitesse moyenne ou vitesse au point A(x, y).

Le graphique classique est donc un moyen de mettre en évidence l'existence et, éventuellement, le caractère constant de cette vitesse et d'en estimer la valeur. Cette vitesse caractérise pour l'ensemble étudié, ou pour un petit intervalle, la proportionnalité des variations de x et de y.

I - GRAPHIQUES A ECHELLES LOGARITHMIQUES

On peut s'intéresser à d'autres formes de liaison que celle envisagée ci-dessus, par exemple liaison entre les variations relatives de y et les variations absolues de x (taux de variation), ou liaison entre les variations relatives de y et les variations relatives de x (élasticité).

Les échelles logarithmiques répondent à ces soucis.

Un autre avantage de l'échelle logarithmique réside dans le fait que, par exemple, la longueur d'échelle nécessaire pour représenter un intervalle 1 à 10^4 ne sera que 4 fois la longueur nécessaire pour représenter l'intervalle 1 à 10, l'erreur relative de mise en place ou de lecture restant constante dans tout l'intervalle utilisé.

A - GRAPHIQUES SEMI-LOGARITHMIQUES

1/ But

Le graphique semi-logarithmique permet de faire correspondre une image graphique (longueur) à un rapport, de représenter la liaison qui existe entre la variation relative de y et la variation absolue de x (taux de variation de y par rapport à x), et éventuellement de constater l'existence d'un taux constant ou sensiblement constant et d'en estimer la valeur r dans une relation de la forme :

$$y = a(1 + r)^x.$$

2/ Définitions

Le quadrillage semi-logarithmique est un quadrillage à deux axes de coordonnées rectangulaires X O Y avec les échelles :

$X = x$ échelle linéaire

$Y = \log_{10} y$ échelle logarithmique

Dans ce système le point coté (x, y) sur les axes a des coordonnées dont les mesures à l'échelle graphiques sont :

$$X = x \qquad Y = \log y + C^{te}$$

(Ceci revient à dire que tout point coté y sur l'échelle peut être considéré comme représentant ky ou plus particulièrement 10^n y en multipliant toutes les graduations de l'axe des y par 10^n.).

Pour cette raison, l'échelle logarithmique est généralement présentée sous la forme suivante (Fig. 1. 2).

Fig. 1. 2

Les points cotés 1 représentent en réalité des puissances successives de 10, au choix de l'utilisateur, la distance entre deux de ces points successifs (ou module de l'échelle) étant mesurée, à l'échelle graphique choisie pour construire le papier, par log 10 = 1.

La pente d'un segment $A(x_1\ y_1)$, $B(x_2,\ y_2)$, représente donc :

$$\frac{Y_2 - Y_1}{X_2 - X_1} = \frac{\log y_2 - \log y_1}{x_2 - x_1} = \frac{\log y_2/y_1}{x_2 - x_1}$$

3/ Emploi du papier semi-logarithmique

3.1 - Estimation d'un rapport

Si la différence graphique des ordonnées cotées y_2 et y_1 est reportée sur l'axe des ordonnées à partir de l'origine cotée 1, on lira directement sur cet axe la valeur du rapport y_2/y_1 (ou du rapport y_1/y_2). (Fig. 1.3)

Fig. 1.3

Le procédé est évidemment sans intérêt pour un rapport isolé.

Mais il devient très commode pour estimer une valeur moyenne de plusieurs rapports y_2/y_1 devant avoir, à des variations aléatoires près, la même signification. Il suffira d'utiliser un morceau de bristol sur le bord duquel on aura marqué une origine O.

Cette origine étant placée successivement en chacun des points y_1', y_1'' ... on marquera d'un trait de crayon les points P', P'' ... correspondants aux valeurs correspondantes y_2', y_2'', ... et on déterminera le point médian M de cette série de points.

Le bristol étant placé sur l'échelle des ordonnées, le point O au point coté 1 de l'échelle logarithmique, on lira en face de M la valeur médiane estimation de la valeur moyenne des rapports y_2/y_1. (Fig. 1.4).

Fig. 1.4

L'homogénéité de l'ensemble des rapports sera manifestée par le groupement des points P, un point anormalement éloigné du groupe attirant l'attention sur une variation exceptionnelle.

On dispose ainsi d'un moyen rapide d'étude de variations périodiques, par exemple, estimation de coefficients saisonniers, les valeurs y_2 correspondant aux chiffres mensuels bruts et les valeurs y_1 étant les moyennes mobiles centrées correspondantes, calculées sur 12 mois consécutifs.

3.2 - Etude d'une tendance

Si l'on étudie la variation d'un phénomène y en fonction d'une variable x et si l'on obtient des points alignés ou des points approximativement alignés dont l'ensemble peut être, en première approximation, représenté par une droite D, on aura mis en évidence un phénomène évoluant à taux constant, la pente de cette droite étant caractéristique du taux de variation de y par rapport à x.

En effet, la relation :

$$y = a(1 + r)^x$$

est équivalente à :

$$Y = \log y = \log a + x \log(1 + r)$$

si, sur cette droite D, on prend deux points quelconques $A(x_1, y_1)$ et $B(x_2, y_2)$, on aura :

$$\log y_2 - \log y_1 = (x_2 - x_1) \log(1 + r)$$

d'où :

$$\log(1 + r) = \frac{\log y_2 - \log y_1}{x_2 - x_1} = \frac{Y_2 - Y_1}{x_2 - x_1}$$

La différence graphique d'ordonnées entre A et B devra être mesurée avec l'unité graphique de l'axe des ordonnées, c'est-à-dire la distance (module) entre deux points cotés 1, (ou 10^n et 10^{n+1}) de cette échelle, soit (Fig. 1.5)

$$\log (1 + r) = \frac{\overline{A'B'}}{\overline{MM'}} \cdot \frac{1}{x_2 - x_1}$$

$$\log(1 + r) = \frac{5}{10} \times \frac{1}{9} = 0,055$$

$$1 + r = 1,136$$

$$r = 0,136$$

ou $r \sim \frac{2,30}{9} \times \frac{5}{10} \sim 0,13$

Fig. 1.5

On obtiendra ensuite 1 + r et r à l'aide de la table logarithmique, mais on peut noter que pour un taux de variation faible, par exemple inférieur à 10 %, on peut, si la période d'étude n'est pas très longue, utiliser l'approximation :

$$\log (1 + r) = 0,43 \ \mathrm{Log}_e(1 + r) \sim 0,43 \ r$$

d'où :

$$r \sim 2,30 \log (1 + r) = 2,30 \ \frac{\log y_2 - \log y_1}{x_2 - x_1}$$

$$r \sim \frac{2,30}{x_2 - x_1} \ \frac{\overline{A'B'}}{\overline{MM'}}$$

Un rapporteur de pentes, gradué en taux, annexé au quadrillage, ou préparé sur un transparent permettrait de lire directement les taux sans calcul.

Ce rapporteur de pentes donnera immédiatement la valeur du facteur $(1 + r)^\delta$ en fonction de r et de x (noter qu'il n'est valable que pour l'échelle d'abscisses qui a servi à le construire).

3.3 - Graphique semi-logarithmique à droite directrice

Ce graphique, dans lequel l'axe des abscisses n'intervient pas, permet la représentation simultanée d'ensembles Y, Z, ... décomposés chacun en plusieurs postes de même espèce (par exemple des bilans).

$$Y = y_1 + y_2 + y_3 + \ldots\ldots + y_n$$

$$Z = z_1 + z_2 + z_3 + \ldots\ldots + z_n$$

Il permet la lecture directe des rapports y_i/Y ... z_i/Z, la comparaison de ces rapports, il montre éventuellement la liaison qui existe entre leur évolution et celle des valeurs globales Y, Z ...

L'élément fondamental du graphique est une oblique D (droite directrice) théoriquement quelconque, mais placée de manière que toutes les valeurs globales Y, Z ... lues à l'échelle logarithmique puissent être placées sur cette droite.

Pratiquement, si l'on a :

$$10^k < Y,\ Z,\ \ldots\ < 10^{k+a}$$

on tracera cette oblique en joignant le point d'ordonnée 10^k de l'échelle logarithmique côté gauche du graphique, au point d'ordonnée 10^{k+a} côté droit du graphique.

Un ensemble particulier Y par exemple sera représenté par une parallèle (Y) à l'axe des ordonnées coupant la droite directrice D au point d'ordonnée Y.

Les sous-ensembles y_1, y_2, ... de Y seront représentés par les points d'ordonnées respectives y_1, y_2, ... de la droite (Y) (Fig 1.6) .

Fig. 1.6 - Graphique à droite directrice.

On représentera de même les autres ensembles.

Un réseau de parallèles à la droite D graduées en %, déterminera sur ces droites (Y), (Z) des points tels que sur chacune d'elles, les rapports y/Y, z/Z soient égaux.

Le réseau gradué des obliques permettra donc de lire directement tous les rapports correspondant à un même poste i, y_i/Y_i, z_i/Z_i, ... et de suivre leur évolution en fonction de celle de Y, Z.

(Le choix du papier à utiliser exige que l'échelle log. des ordonnées puisse contenir la plus petite des valeurs y_i, z_i, ... et la plus grande des valeurs Y, Z).

4/ Comparaison de deux tendances

Si on a étudié simultanément deux séries d'observations y et z en fonction d'une même variable x, (par exemple deux séries chronologiques), le fait que les deux graphiques (Y) et (Z) se déduisent l'un de l'autre par une translation parallèle à l'axe des ordonnées montrera que le rapport z/y reste invariable quand x varie.

Dans le cas où les deux séries peuvent être représentées par des droites qui seront parallèles, les taux de variation de y et z par rapport à x seront les mêmes.

Plus généralement, si on obtient deux droites de tendance de pentes différentes, on aura pour un même intervalle d'étude (x_1, x_2) :

$$\frac{\log(1 + r_1)}{\log(1 + r)} = \frac{A_1'B_1'}{A'B'}$$

Pour des taux inférieurs à 10 % on aura sensiblement (Fig. 1.7).

Fig. 1.7

5/ Remarque

Une difficulté d'emploi des échelles logarithmiques provient du fait que le module graphique (distance entre deux puissances successives de 10 sur l'axe correspondant) est fixé : on trouve dans le commerce des papiers de divers modules mais la gamme en est assez restreinte.

Si y varie dans un intervalle petit, c'est-à-dire caractérisé par un faible rapport des valeurs extrêmes, les valeurs de Y = log y seront très voisines et le graphique sera d'un dessin ou d'une lecture peu précis.

On peut y remédier en utilisant du papier quadrillé ordinaire avec une échelle graduée en Y = log y, après transformation de y en Y à l'aide de la table de logarithmes : on pourra alors choisir au mieux l'unité graphique.

Etant donné l'intérêt pratique des graphiques semi-logarithmiques, on pourra très facilement construire des échelles logarithmiques de module graphique convenable pour tout ou partie du module.

L'échelle logarithmique d'un papier du commerce étant placée au bord d'une feuille de papier millimétré, il suffira de tracer une oblique de pente convenable pour graduer une nouvelle échelle que l'on pourra ensuite découper suivant cette droite (Fig. 1.8).

Fig. 1.8

B - GRAPHIQUE LOGARITHMIQUE-LOGARITHMIQUE

1/ But

Le graphique log-log permet de mettre en évidence la liaison qui existe entre les variations relatives $\frac{\Delta y}{y}$ et $\frac{\Delta x}{x}$ de deux variables, et de déterminer l'élasticité de y par rapport à x, soit :

$$\lambda = \frac{\Delta y}{y} : \frac{\Delta x}{x}$$

pour de petites variations correspondantes de x et de y.

2/ Définitions

Le quadrillage log-log est un quadrillage avec deux échelles logarithmiques en général de même module (ce que nous supposerons dans ce qui suit).

Le point coté (x y) sur les axes du graphique a des coordonnées dont les mesures à l'échelle graphique sont :

$$X = \log x \qquad Y = \log y$$

La pente d'un segment $A(x_1, y_1)$, $B(x_2, y_2)$ est alors :

$$\frac{Y_2 - Y_1}{X_2 - X_1} = \frac{\log y_2 - \log y_1}{\log x_2 - \log x_1} = \frac{\log y_2/y_1}{\log x_2/x_1}$$

Elle est donc caractéristique d'une liaison entre les variations relatives de x et de y.

3/ Emploi du graphique log-log

Les points (x, y) étant placés sur le graphique, la courbe définie par ces points mettra en évidence la liaison qui existe entre log y et log x.

Si l'ensemble des points peut être convenablement représenté par une droite D d'équation :

$$\log y = a + \alpha \log x$$

correspondant entre y et y à une relation de la forme :

$$y = k\, x^{\alpha}$$

le paramètre α positif ou négatif, caractéristique de cette liaison sera mesuré par la pente de la droite D

Si les modules des deux échelles sont les mêmes, α pourra être déterminé par un simple rapport de longueurs mesurées sur les axes (Fig. 1.9).

$$\alpha = \frac{\overline{A'B'}}{\overline{AB}}$$

Fig. 1.9

Pour de petites variations associées de x et de y, ce coefficient mesurera le rapport des variations relatives de y et de x (coefficient d'élasticité de y par rapport à x).

$$\alpha \sim \frac{y_2 - y_1}{y_1} : \frac{x_2 - x_1}{y_1}$$

/ 2c

En général, le paramètre α est petit, de l'ordre de quelques unités et l'approximation ci-dessus sera valable si $\frac{x_2 - x_1}{x_1}$ ne dépasse pas 10 %.

(De manière plus précise, on notera que la relation

$$y = k\ x^{\alpha}$$

entraîne comme conséquence :

$$\frac{dy}{y} = \alpha \frac{dx}{x} \quad).$$

Si l'ensemble des points (x y) est représenté par une courbe C, la pente de la tangente à la courbe C en un point M(xy) mesurera le coefficient d'élasticité de y par rapport à x dans une petite région autour de M.

4/ Exemple. Loi de Pareto

La loi de Pareto a été utilisée d'abord pour décrire l'inégalité de la distribution des richesses. On constate en effet dans l'étude de la distribution des revenus dans une population que le nombre N_x des revenus supérieurs à une certaine valeur x, satisfait approximativement à une loi de distribution de la forme

$$N_x = A\ x^{-\alpha},$$

α étant une constante, (en général voisine de 3/2), il en résulte que l'on a :

$$\log N_x = a - \alpha \log x$$

relation représentée par une droite sur du papier log-log.

En contrôle des fabrications, de nombreuses observations ont montré, qu'au moins d'une manière qualitative, elle était applicable à l'inégale distribution des pertes dues aux différents types de défauts, un petit nombre de défauts étant responsable d'une part très importante de la perte totale.

Une telle étude est importante dans l'élaboration d'un programme d'amélioration de la qualité : elle permet de faire porter les efforts sur les points essentiels et de négliger éventuellement des secteurs dans lesquels le coût des recherches et des corrections à apporter risquerait d'être supérieur à la diminution de perte pouvant en être tirée.

D'autres études économiques ont montré l'intérêt de ces recherches de liaisons entre variations relatives.

Dans un article, publié dans l'Onde Electrique d'Avril 1963, M. Danzin étudie "La formation des prix de revient dans l'Industrie des composants électroniques".

Cet article donne des renseignements numériques à partir de résultats d'expériences sur la variation du prix de revient d'un composant en fonction des quantités fabriquées selon trois modes de production :

- artisanale,
- semi automatique
- automatique

Etudiant ces trois courbes prix/quantité sur papier log-log, M. Borel a constaté qu'elles admettaient une enveloppe linéaire avec une précision de l'ordre de 10 %, enveloppe que l'on peut interpréter comme donnant une loi optimale du prix de revient en fonction de la quantité pour les modes de fabrication adaptés aux quantités.

M. Borel trouve alors pour le coût unitaire y en fonction de la quantité x (en millions par an) les résultats suivants :

- pour un prix de revient complet

$$y = 120\ x^{-0,21}$$

- pour un prix de revient sans matière première

$$y = 100\ x^{-0,30}$$

relations tendant à montrer que pour de petites variations, les pourcentages de variation du prix unitaire et de la quantité sont proportionnels et qu'il en est de même pour les pourcentages de variation de la dépense totale et de la qualité.

5/ Remarque

A propos des graphiques log-log, remarquons encore que des relations de la forme :

$$\frac{y}{x} = C^{te} \qquad \text{ou} \qquad xy = C^{te}$$

se traduisent par des droites parallèles aux bissectrices des angles des axes :

$$Y = X + K \qquad \text{ou} \qquad Y = -X + K$$

Si la première transformation qui remplace des droites concourantes par des droites parallèles est de peu d'intérêt (si ce n'est celui d'une précision relative constante), la seconde qui transforme des hyperboles en droites parallèles l'est évidemment davantage (Fig. 1.10).

Fig. 1.10

6/ Application des graphiques logarithmiques à la loi exponentielle

La loi exponentielle joue un rôle essentiel dans les études de fiabilité.

Si le taux de défaillance est constant, au cours de la période d'observation, la probabilité de survie au bout d'une durée t (ou de zéro défaillance pendant la durée t), c'est-à-dire la fiabilité est définie par :

$$R = \text{Prob}(T > t) = 1 - F(t) = e^{-\frac{t}{m}} = e^{-\lambda t}$$

m étant la durée de vie moyenne ou le temps moyen entre deux défaillances.

On a donc :

$$\text{Log}_e R = - t/m$$

relation qui est représentée par une droite dans un graphique semi-logarithmique dans lequel l'échelle des valeurs de R est logarithmique, l'échelle des t étant linéaire (Fig. 1. 11).

6. 1 - Emploi pour une loi exponentielle de moyenne m connue

L'échelle des abscisses étant alors graduée en t/m, le graphique fournira la solution immédiate des deux problèmes suivants :

a) Un équipement possède, d'après des informations antérieures, une vie moyenne de 300 heures entre deux défaillances. Quelle est la probabilité pour que cet équipement fonctionne sans panne pendant 60 heures ?

Pour $t/m = \frac{60}{300} = 0,2$; le graphique donne R = 0, 82.

b) Inversement, quel doit être le temps moyen entre pannes, nécessaire pour donner une probabilité 0, 90 de fonctionner sans panne pendant une durée de 60 heures ?

Pour P = 0, 90, le graphique donne t/m = 0, 10 soit $= = \frac{60}{0,10} = 600$ Heures .

Le tableau ci-après donne les valeurs de $e^{-t/m}$ pour quelques valeurs de t/m.

t/m	$e^{-t/m}$	t/m	$e^{-t/m}$
0, 01	0, 990	0, 11	0, 896
0, 02	0, 980	0, 12	0, 887
0, 03	0, 970	0, 13	0, 878
0, 04	0, 961	0, 14	0, 869
0, 05	0, 951	0, 15	0, 861
0, 06	0, 942	0, 16	0, 852
0, 07	0, 932	0, 17	0, 844
0, 08	0, 923	0, 18	0, 835
0, 09	0, 914	0, 19	0, 827
0, 10	0, 905	0, 20	0, 819

Fig. 1.11 - Loi exponentielle $P = e^{-\frac{t}{m}}$

On notera que pour $t/m < 0,10$, on peut pratiquement se contenter de l'approximation :

$$e^{-t/m} \sim 1 - t/m$$

Pour $t/m > 0,20$, les valeurs de $e^{-t/m}$ peuvent être lues avec une précision suffisante sur le graphique.

6.2 - Chercher si une loi de survie observée sur un échantillon est une loi exponentielle et, si cette hypothèse est acceptée, estimer m.

On utilisera le même papier semi-logarithmique, l'échelle des abscisses étant alors graduée en durées avec une unité graphique que l'on choisira en fonction des observations.

Soient n le nombre des éléments soumis à l'essai et $t_1 \leqslant t_2 \leqslant \ldots \leqslant t_i \leqslant \ldots \leqslant t_n$ les temps auxquels se produisent les défaillances successives.

Si la distribution est exponentielle de moyenne m, on pourra considérer le rapport $\frac{n+1-i}{n+1}$ comme une estimation correcte de $P_i = 1-F(t_i) = e^{-t_i/m}$ (1).

Il suffira donc de placer les points d'abscisses t_i et d'ordonnées $\frac{n+1-i}{n+1}$ sur le graphique et de voir s'ils sont approximativement alignés.

S'il en est ainsi, l'abscisse du point d'ordonnée 0,367 (ou 36,7 %) de la droite représentant au mieux l'ensemble des points fournira une estimation de la vie moyenne m.

$$\left(\text{Pour } m = t_i \text{ on a } R = \frac{1}{e} = 0,367\right)$$

On pourra évidemment utiliser tout aussi bien les points $\left(t_i, \frac{n+1}{n+1-i}\right)$ qui dans les mêmes conditions doivent être alignés sur une droite dont l'abscisse du point d'ordonnée e = 2,718 donnera alors l'estimation de la moyenne m (Fig. 1.12).

6.3 - Exemple

Vingt-quatre équipements ont été soumis à un essai ayant donné les résultats donnés par le tableau ci-après (t_i = époque de la $i^{ème}$ défaillance), dans lequel on a calculé les valeurs correspondantes des rapports $\frac{n+1-i}{n+1}$.

Les points étant approximativement alignés on a tracé la droite D_1 qui donne $m \sim 47$. (On a aussi tracé sur le graphique la droite D_2 correspondant aux points $\left(t_i, \frac{n+1}{n+1-i}\right)$). (Fig. 1.12).

---

(1) On montre en effet que t_i étant l'époque de la $i^{ème}$ défaillance constatée, l'espérance mathématique de $F(t_i)$, c'est-à-dire la valeur moyenne de $F(t_i)$ pour l'ensemble des échantillons de n observations provenant de la population testée est :

$$E\,[F(t_i)] = \frac{i}{n+1} \quad \text{(voir annexe)}$$

Fig. 1.12 - Etude graphique d'une distribution exponentielle.

i	t_i	$\frac{n+1-i}{n+1}$	i	t_i	$\frac{n+1-i}{n+1}$
0	0	1	13	30	0, 48
1	1	0, 96	14	35	0, 44
2	2	0, 92	15	45	0, 40
3	6	0, 88	16	46	0, 36
4	7	0, 84	17	56	0, 32
5	11	0, 80	18	60	0, 28
6	12	0, 76	19	62	0, 24
7	13	0, 72	20	75	0, 20
8	15	0, 68	21	87	0, 16
9	23	0, 64	22	95	0, 12
10	26	0, 60	23	115	0, 08
11	27	0, 56	24	152	0, 04
12	28	0, 52			

(Tableau I - Essai d'un échantillon de 24 équipements).

Ne pas oublier que cette étude suppose que la période des défaillances précoces (mortalité infantile) a été exclue des observations, ainsi que la période d'usure.

6.4 - Annexe Théorique

Considérons un élément dont la probabilité de défaillance dans l'intervalle infiniment petit t à t + dt est λ dt (λ, taux de défaillance étant une constante positive, indépendante de t : pas de vieillissement mais seulement accident aléatoire).

Soit F(t) la probabilité de défaillance entre 0 et t :

$$F(t) = Pr(T < t)$$

La probabilité de survie à l'époque t ou fiabilité sera $R(t) = 1 - F(t) = Pr\ (T > t)$.

La probabilité a priori pour que la défaillance d'un élément ait lieu entre t et t + dt est égale au produit de la probabilité de survie à l'époque t multiplié par la probabilité de défaillance dans l'intervalle (t, t + dt), soit :

$$dF(t) = [1 - f(t)]\ \lambda\, dt$$

$$\frac{dF}{1 - F} = \lambda\ dt$$

$$Log_e(1 - F) = \lambda\ t + C^{te}$$

Pour t = 0 on doit avoir 1 = F(0) = 1 d'où :

$$R(t) = 1 - F(t) = e^{-\lambda t}$$

La fonction de répartition de la distribution exponentielle des durées de vie est donc :

$$F(t) = 1 - e^{-\lambda t} \qquad \text{pour } t \geqslant 0$$
$$= 0 \qquad \text{pour } t < 0$$

Il lui correspond une fréquence instantanée de défaillance (densité de probabilité des défaillances) défini par :

$$f(t) = \frac{dF}{dt} = \lambda e^{-\lambda t} \quad \text{(Fig. 1.13)}$$

$f(t) = \frac{1}{m} e^{-\frac{t}{m}}$ $\qquad$ $F(t) = 1 - e^{-\frac{t}{m}}$ $\qquad$ $\alpha = \frac{1}{m} = c^{te}$

$R(t) = e^{-\frac{t}{m}}$

Fig. 1.13 - Distribution exponentielle.

6.4.1 - *Vie moyenne*

La durée de vie moyenne est :

$$m = E(t) = \int_0^\infty t\, f(t)\, dt = \int_0^\infty \lambda\, t\, e^{-\lambda t} dt = \frac{1}{\lambda} \int_0^\infty y\, e^{-y}\, dy \quad \text{avec } y = \lambda t$$

En intégrant par parties on a :

$$\int_0^\infty y\, e^{-y}\, dy = [-y\, e^{-y}]_0^\infty - \int_0^\infty - e^{-y}\, dy = - [e^{-y}]_0^\infty = 1$$

d'où : $\qquad m = \frac{1}{\lambda}$

On remarquera que l'on a aussi :

$$m = \int_0^\infty R(t)\, dt = \int_0^\infty e^{-\lambda t}\, dt = - \left[\frac{1}{\lambda} e^{-\lambda t}\right]_0^\infty = \frac{1}{\lambda}$$

6.4.2 - *Variance de T*

On calculerait de même :

$$V(T) = \int_0^\infty t^2\, f(t)\, dt - m^2 = \frac{1}{2} = m^2$$

Compte tenu de ces résultats, on présente souvent la distribution exponentielle de durées de vie de moyenne m sous la forme :

$F(t) = Pr(T \leqslant t) = 1 - e^{-t/m}$ $\qquad$ (Probabilité de défaillance)

$R(t) = Pr(T > t) = e^{-t/m}$ $\qquad$ (Probabilité de survie ou fiabilité)

$f(t) = \frac{1}{m} e^{-t/m}$

6.4.3 - *Estimation à partir d'un échantillon de la fonction de répartition d'une distribution exponentielle.*

Considérons n éléments mis en fonctionnement et soient $t_1 < t_2 < \ldots t_i \ldots < t_n$ les époques des défaillances.

t_i, époque de la $i^{ème}$ défaillance est une variable aléatoire, il en est de même de la variable $u = F(t_i)$, la loi de $u = F(t)$ étant une loi uniforme sur l'intervalle (0, 1).

La probabilité pour que la $i^{ème}$ avarie se produise dans un intervalle $(t, t + dt)$ est $\varphi(t)\,dt$ = Probabilité d'avoir :

$i - 1$	éléments défaillants dans l'intervalle $(0, t)$
1	élément défaillant dans $(t, t + dt)$
$n - i$	éléments non défaillants.

d'où :

$$\varphi(t) = \frac{n!}{(i-1)!\ 1!\ (n-i)!} (1 - e^{-t/m})^{i-1} \frac{1}{m} e^{-t/m} (e^{-t/m})^{n-i}$$

Si on fait le changement de variable :

$$u = F(t) = 1 - e^{-t/m}$$

$$\frac{du}{dt} = \frac{1}{m} e^{-t/m}$$

La loi de probabilité élémentaire de la variable aléatoire $u_i = F(t_i)$, définie par la correspondance $\varphi(t)dt = h(u)\,du$, est donc :

$$h(u)\,du = \frac{n!}{(i-1)!\ 1!\ (n-i)!} u^{i-1}(1-u)^{n-i}\,du$$

d'où l'espérance mathématique de $F(t_i)$

$$E[F(t_i)] = \int_0^1 u\,h(u)\,du = \frac{n!}{(i-1)!\ 1!\ (n-i)!} \int_0^1 u^i (1-u)^{n-i}\,du$$

Compte tenu de :

$$\int_0^1 u^i (1-u)^{n-i}\,du = B(i+1,\ n+1-i) = \frac{i!\ (n-i)!}{(n+1)!}$$

on a finalement :

$$E[F(t_i)] = \frac{i}{n+1}$$

$$E[1 - F(t_i)] = 1 - \frac{i}{n+1} = \frac{n+1-i}{n+1}$$

6.4.4 - *Estimation de la vie moyenne m (durée moyenne avant défaillance*

Soient n éléments soumis à un essai de durée de vie. L'expérience étant arrêtée lorsque tous les éléments sont défaillants, soient t_1, t_2, ..., t_n les époques des défaillances.

La fonction de vraisemblance est :

$$L = m^{-n} e^{-1/m} \sum t_i$$

qui sera maximum, en même temps que son logarithme :

$$\text{Log}_e L = - n \text{ Log}_e m - \frac{1}{m} \sum t_i$$

soit pour :

$$- \frac{n}{m} + \frac{1}{m^2} \sum t_i = 0,$$

d'où l'estimation :

$$m = \frac{1}{n} \sum_{i=1}^{n} t_i$$

D'autres estimations pourraient être envisagées mais on peut considérer celle-là comme la meilleure en ce sens que son espérance mathématique est égale à m et que sa variance $V(\hat{m}) = \frac{m^2}{n}$ est minimum

On remarquera toutefois que pour un échantillon de faible effectif d'éléments de haute fiabilité, cette variance est importante.

D'autre part, il faut tenir compte du fait que ce calcul correspond à l'hypothèse de défaillances aléatoires avec un taux constant, c'est-à-dire que la période de mortalité infantile a été éliminée et que, d'autre part, on n'a pas atteint la période d'usure.

Or, si l'essai est prolongé jusqu'à défaillance du dernier élément, il est vraisemblable que cette période d'usure sera atteinte ; on est donc conduit à arrêter l'essai à une durée telle que la période d'usure ne soit pas atteinte.

Epstein a montré que pour n éléments soumis à l'essai, si cet essai est arrêté à la $k^{i\text{ème}}$ défaillance, les époques des k défaillances étant t_1, t_2, ...t_i,..t_k, l'estimation de m est :

$$\hat{m} = \frac{\sum_{i=1}^{k} t_i + (n - k) t_k}{k}$$

Dans le but d'obtenir une information plus rapide, on peut envisager un test avec remplacement dans lequel :

1/ n éléments sont utilisés à l'origine de l'essai.

2/ Chaque élément défaillant est immédiatement remplacé par un nouvel élément provenant de la même population de sorte qu'il y a toujours n éléments sous test.

3/ L'essai est encore arrêté à l'époque $t = t_k$ de la $k^{ième}$ défaillance. On aura alors :

$$\hat{m} = \frac{nt_k}{k}$$

Ce test peut donner une information plus rapide mais pour un coût de réalisation plus grand (permanence d'attention de jour et de nuit pour effectuer les remplacements).

Mais l'économie de temps à précision égale est généralement faible si l'on envisage la nécessité d'observer un nombre de défaillances suffisamment grand pour espérer une estimation assez précise de $\hat{m}$.

6.4.5 - *Temps moyen entre défaillances ("mean time between failures" MTBF).*

Si l'on considère un système ou un équipement réparable, t_i, mesurera le temps de fonctionnement entre deux défaillances.

Après k avaries, n mesures de t_i seront disponibles et le temps moyen entre deux défaillances sera :

$$m = \frac{\sum t_i}{k} = \frac{T}{k}$$

en désignant par T le temps total de fonctionnement jusqu'à la $k^{ième}$ avarie.

Dans les ouvrages américains, le terme temps moyen entre défaillances (MTBF) est souvent employé aussi bien dans le cas de composants non réparables (vie moyenne ou temps moyen avant défaillance) que dans le cas d'un équipement réparable.

Dans les deux cas, il représente le même concept statistique : temps moyen au bout duquel apparaît une avarie.

Prenons comme origine t l'instant ou l'équipement est mis ou remis en service.

La probabilité pour que la défaillance suivante ait lieu entre t et t + dt est égale au produit de la probabilité de zéro défaillance entre 0 et t multipliée par la probabilité d'une défaillance entre t et t + dt, soit dans le cas d'une distribution exponentielle :

$$e^{-\lambda t} \times \lambda dt$$

L'espérance mathématique de t, ou temps moyen entre deux défaillances est donc encore :

$$\int_0^\infty \lambda t \; e^{-\lambda t} \, dt = \frac{1}{\lambda} = m$$

6.4.6 - *Probabilité des défaillances entre 0 et t.*

Soit $P_k(t)$ cette probabilité.

---

(1) Certains techniciens français utilisent l'expression "Moyenne des Temps de Bon Fonctionnement" qui correspond à l'observation américaine MTBF.

On peut écrire que la probabilité de k défaillances entre 0 et t + dt est $P_k(t + dt)$ = Pr(k entre zéro et t) x Pr(zéro entre t et t + dt)
+ Pr(k - 1 entre zéro et t) x Pr(une entre t et t + dt).

soit :

$$P_k(t + dt) = P_k(t) \quad [1 - \lambda dt] + P_{k-1}(t).\lambda dt.$$

d'où :

$$\frac{P_k(t + dt) - P_k(t)}{dt} = - \lambda P_k(t) + \lambda P_{k-1}(t)$$

$$\frac{dP_k}{dt} = - \lambda P_k + \lambda P_{k-1}$$

On obtient ainsi un système récurrent d'équations différentielles avec les conditions initiales :

$$P_0(0) = 1 \qquad P_k(0) = 0 \qquad \text{si } k \geqslant 1$$

A partir de la solution de base $P_0 = e^{-\lambda t}$ on obtient successivement

$$P_1(t) = \lambda t \; e^{-\lambda t}$$

$$P_k(t) = (\lambda t)^k \; \frac{e^{-\lambda t}}{k!}$$

c'est-à-dire une loi de Poisson de moyenne :

$$E(k) = \lambda t$$

La probabilité pour que le nombre des défaillances entre 0 et t soit inférieur à k sera :

$$P_{x<k}(t) = \sum_{x=0}^{k} e^{-\lambda t} \frac{(\lambda t)^x}{x!}$$

qui sera donnée directement par les tables ou abaques de la loi de Poisson.

6.4.7 - *Intervalle de confiance à 1 - α pour la vie moyenne m.*

A - *Essai de durée totale préfixée.*

Un essai a été poursuivi pendant une durée totale de fonctionnement θ, fixée à l'avance, au cours de laquelle on a observé k défaillances.

Pour un système isolé, réparable, la durée totale opérationnelle sera :

$$T = \sum_{i=1}^{n} t_i + t_0,$$

les temps t_i étant les durées de fonctionnement entre avaries successives, et t_0 le temps de fonctionnement entre la $k^{ième}$ défaillance et l'arrêt de l'essai.

Pour un ensemble de n éléments identiques, T sera la somme des temps opérationnels des divers éléments soumis à l'essai, au cours de la durée d'observation fixée, soit :

$$T = \sum_{i=1}^{n} t_i + (n - k)\,\theta .$$

La durée de vie moyenne ou temps moyen entre défaillance sera estimée par :

$$\hat{m} = \frac{T}{k}$$

A cette estimation, on se propose d'associer un intervalle (m_1, m_2) ayant une probabilité donnée $1 - \alpha$ de contenir la vraie valeur inconnue m dans la population d'ou provient l'échantillon testé, c'est-à-dire tel que :

$$Pr(m_1 < m < m_2) = 1 - \alpha$$

Compte tenu de la relation entre la loi de Poisson et la loi de χ^2, on sait que :

$$Pr_{x \leq k}(t) = 1 - Pr(\chi^2 < 2\,\lambda\,T) = 1 - Pr\left(\chi^2 < \frac{2T}{m}\right)$$

avec :

$$\nu = 2(k + 1) \text{ degrés de liberté.}$$

En particulier, on a :

$$Pr_{x = 0}(t) = 1 - Pr(\chi^2 < 2T/m) \text{ avec } \nu = 2 \text{ degrés de liberté.}$$

Pour une durée totale de fonctionnement T, si la vie moyenne dans la population est m_1, la probabilité d'obtenir une estimation égale ou supérieure à $\hat{m} = T/k$, c'est-à-dire la probabilité d'observer un nombre x d'avaries égal ou inférieur à k, devra satisfaire à la condition (Fig. 1. 14)

Fig. 1. 14

$$\frac{\alpha}{2} = Pr(x \leqslant k \quad si \quad m = m_1)$$

De même, si la vie moyenne est m_2, la probabilité d'une estimation égale ou inférieure à $\hat{m}$ devra satisfaire à :

$$\frac{\alpha}{2} = Pr(x \geqslant k \quad si \quad m = m_2)$$

ou encore :

$$1 - \frac{\alpha}{2} = Pr(x \leqslant k - 1 \quad si \quad m = m_2).$$

Ces conditions s'écrivent :

$$\frac{\alpha}{2} = 1 - Pr\text{-}\left(\chi^2 < \frac{2T}{m_1}\right) \quad avec \quad \nu = 2(k + 1)$$

$$1 - \frac{\alpha}{2} = 1 - Pr\left(\chi^2 < \frac{2T}{m_2}\right) \quad avec \quad \nu = 2k$$

Si l'on désigne de manière générale par $\chi^2_\beta(\nu)$ la valeur de $\chi^2(\nu)$ telle que (Fig. 1.15) :

$$\beta = Pr[\chi^2(\nu) < \chi^2_\beta(\nu)]$$

Fig. 1.15 - Loi de χ^2

l'intervalle de confiance à $1 - \alpha$ pour m est défini par :

$$\frac{2T}{m_1} = \chi^2_{1-\frac{\alpha}{2}} \ (\nu = 2k + 2)$$

$$\frac{2T}{m_2} = \chi^2_{\frac{\alpha}{2}} \ (\nu = 2k)$$

soit :

$$m_1 = \frac{2T}{\chi^2_{1-\frac{\alpha}{2}}(\nu = 2k + 2)} \quad , \qquad m_2 = \frac{2T}{\chi^2_{\frac{\alpha}{2}} \ (\nu = 2k)}$$

Compte tenu de :

$$\hat{m} = \frac{T}{k}$$

ces résultats peuvent aussi être mis sous la forme :

$$\frac{m_1}{\hat{m}} = \frac{2k}{\chi^2_{1-\frac{\alpha}{2}} \ (\nu = 2k + 2)}$$

$$\frac{m_2}{\hat{m}} = \frac{2k}{\chi^2_{\frac{\alpha}{2}} \ (\nu = 2k)}$$

Ceci permet, à partir de la table de χ^2 de construire un abaque donnant $\frac{m_1}{\hat{m}}$ et $\frac{m_2}{\hat{m}}$ en fonction de k. (voir abaque Fig. 1. 16).

(Il est commode d'utiliser pour k une échelle logarithmique).

B - *Essai arrêté dès que la k ième avarie s'est produite.*

Pour un système isolé réparable, on a :

$$T = \sum_{i=1}^{k} t_i$$

Pour un ensemble de n composantes dont les k défaillances (k < n) sont observées aux époques $t_1 \ldots t_k$, on a :

$$T = \sum_{i=1}^{k} t_i + (n - k)t_k$$

la vie moyenne étant toujours $\hat{m} = T/k$

Dans ce cas, où T est une variable aléatoire continue, Epstein a montré (1) que le rapport $\frac{2T}{m} = \frac{2k\hat{m}}{m}$ est encore distribué comme χ^2 avec 2k degrés de liberté, les deux limites de l'intervalle de confiance à $1 - \alpha$ pour m étant alors :

$$m_1 = \frac{2T}{\chi^2_{1-\frac{\alpha}{2}} \ (\nu = 2k)} = \frac{2k\hat{m}}{\chi^2_{1-\frac{\alpha}{2}} (\nu = 2k)}$$

$$m_2 = \frac{2T}{\chi^2_{\frac{\alpha}{2}} \ (\nu = 2k)} = \frac{2k\hat{m}}{\chi^2_{\frac{\alpha}{2}} \ (\nu = 2k)}$$

(On remarquera le changement de $\nu = 2k + 2$ en $\nu = 2k$ pour la détermination de la limite inférieure m_1).

On pourra comme précédemment construire l'abaque correspondant à ce problème.

---

(1) B. Epstein - Estimation from life test data.

"IRE Transactions on Reliability and Quality Control" Vol. RQC - 9 - Avril 1960.

Fig. 1.16 - Limites de confiance du rapport $\frac{m}{\hat{m}}$ (m = vie moyenne estimée d'après un test de durée opérationnelle T fixée au cours duquel on a observé k défaillance).

$$\frac{2k}{\chi^2_{1-\frac{\alpha}{2}}(\nu = 2k+2)} \leqslant \frac{m}{\hat{m}} \leqslant \frac{2k}{\chi^2_{\frac{\alpha}{2}}(\nu = 2k)}$$

6.4.8 - *Exemples.*

1/ Au cours d'un essai d'une durée opérationnelle totale fixée de T = 3.000 heures, un équipement électronique a donné lieu à k = 25 avaries. Quel est l'intervalle de confiance à 0,95 pour la durée de vie moyenne ? $\left(\frac{\alpha}{2} = \frac{1}{2}(1 - 0,95) = 0,025\right)$

La vie moyenne estimée est :

$$\hat{m} = \frac{3.000}{25} = 120 \text{ heures}$$

Les limites de l'intervalle de confiance à 0,95 sont :

$$m_1 = \frac{6.000}{\chi^2_{0,975}\ (\nu = 52)} = \frac{6.000}{73,8} = 81 \text{ heures}$$

$$m_2 = \frac{6.000}{\chi^2_{0,025}\ (\nu = 50)} = \frac{6.000}{32,4} = 185 \text{ heures}$$

L'abaque (Fig. 1.16) donne) :

$$\frac{m_1}{\hat{m}} = 0,68 \qquad \frac{m_2}{\hat{m}} = 1,53$$

2/ Un acheteur a demandé que la vie moyenne ait 95 chances sur 100 d'être supérieure à m_0= 100 heures. L'équipement ci-dessus satisfait-il à cette condition ?

La limite inférieure de confiance à ce niveau $\frac{\alpha}{2} = 0,05$ intervalle unilatéral pour m est alors définie par :

$$m_1 = \frac{2T}{\chi^2_{0,95}} = \frac{2k\hat{m}}{\lambda^2_{0,95}}, \qquad \nu = 2(k + 1)$$

soit, pour l'exemple considéré :

$$m_1 = \frac{6.000}{69,8} = 86 \text{ heures}$$

Le client demandant 100 heures à ce niveau, l'équipement testé ne répond pas aux exigences imposées.

(Pour $\frac{\alpha}{2} = 0,05$, l'abaque donne $\frac{m_1}{\hat{m}} = 0,72$, alors que l'on demande $\frac{m_1}{\hat{m}} = \frac{100}{120} = 0,83$).

3/ Un équipement réparable a eu sa troisième défaillance au bout de 600 heures de fonctionnement (temps de réparation déduits). Quel est l'intervalle de confiance à 0,95 pour la durée de vie moyenne ? Quelle est la durée de vie qui a 95 chances sur 100 d'être dépassée ?

a) On a :

$$\frac{\alpha}{2} = \frac{0,05}{2} = 0,025 \qquad T = 600 \qquad k = 3$$

$$\hat{m} = \frac{600}{3} = 200 \text{ heures}$$

$$m_1 = \frac{2T}{\chi^2_{0,975}} = \frac{1.200}{14,4} = 83 \text{ heures}$$

$$m_2 = \frac{2T}{\chi^2_{0,025}} = \frac{1.200}{1,24} = 968 \text{ heures}$$

$$\nu = 2k = 6$$

$$m_0 = \frac{2T}{\chi^2_{0,95}} = \frac{1.200}{12,6} = 95 \text{ heures}$$

4/ Reprenons les données du tableau I ci-dessus et supposons que l'essai a été arrêté immédiatement après la 15ème avarie : on ne possède que les 15 premiers temps pour estimer la vie moyenne :

$$\hat{m} = \frac{\sum_{i=1}^{15} t_i + (24 - 15)\, t_{15}}{15}$$

$$= \frac{281 + 9 \times 45}{15} = \frac{686}{15} = 46 \text{ heures}$$

Les limites de confiance à 0,95 sont :

$$m_1 = \frac{2 \times 686}{\chi^{(30)}_{0,975}} = \frac{1372}{47} = 29 \text{ heures}$$

$$m_2 = \frac{2 \times 686}{\chi^{(30)}_{0,025}} = \frac{1372}{16,8} = 81 \text{ heures}$$

5/ Supposons maintenant avec les mêmes données que l'essai a été arrêté lorsque la somme des durées de vie atteint T = 800 heures. Ceci se produit 25 heures après la 17ème avarie, la 18ème ne s'étant pas produite.

On a en effet immédiatement après la 17ème avarie :

$$\sum_{i=1}^{18} t_i = 383$$

$$T_i = 382 + 7 \times 56 = 775 \text{ heures}$$

(Immédiatement après la 18ème avarie, on aurait :

$$\sum_{i=1}^{18} t_i = 443$$

$$T_i = 443 + 360 = 803).$$

On a alors :

$$\hat{m} = \frac{800}{17} = 47 \text{ heures}$$

$$m_1 = \frac{2T}{\chi^{2\,(36)}_{0,975}} = \frac{1.600}{54,4} = 29 \text{ heures}$$

$$m_2 = \frac{2T}{\chi^{2\,(34)}_{0,025}} = \frac{1.600}{19,8} = 81 \text{ heures.}$$

6/ Un équipement doit avoir une fiabilité au moins égale à 0, 99 pour une mission d'une durée de 2 heures.

Quelle doit être la durée minimum d'un essai sans défaillance pour qu'on puisse considérer ce résultat comme acquis au niveau de confiance $1 - \alpha = 0,95$?

La vie moyenne m doit être au moins égale à celle qui correspond à l'équation :

$$0,99 = e^{-2/m} \sim 1 - \frac{2}{m}$$

soit m = 200.

D'autre part la condition :

$$0,05 = Pr(x = 0 \quad \text{si} \quad m = 200)$$

peut s'écrire, (voir page 51)

$$0,05 = 1 - Pr\left(\chi^2 < \frac{2T}{200}\right)$$

avec : $\nu = 2(k + 1)$ 2 degrés de liberté.

La valeur minimum de T sans défaillance doit donc satisfaire à :

$$\frac{2T}{200} = \chi^2_{0,95}(\nu = 2) = 5,99$$

soit : T = 600 heures.

II - PAPIER A UNE ECHELLE FONCTIONNELLE LOI NORMALE

1/ But

Ce papier permet de résoudre graphiquement les deux problèmes ci-après.

1/ voir si une distribution observée est approximativement normale

2/ Si cette hypothèse paraît acceptable, estimer sa moyenne et son écart-type.

2/ Définition

Dans ce papier (appelé quelquefois papier gausso-arithmétique), l'un des axes, par exemple ox, a une graduation linéaire habituelle dont on peut fixer arbitrairement l'unité graphique, l'axe oy étant gradué directement en fréquences cumulées de la répartition observée (fonction de répartition de la loi normale). Cet axe est généralement gradué en centièmes ou en millièmes.

Il est commode d'associer à cet axe, un axe parallèle gradué en valeurs correspondantes de la variable normale réduite u, c'est-à-dire en écarts-types.

Sur un tel graphique, la fonction de répartition de la loi normale, c'est-à-dire la courbe cumulative de la distribution sera représentée par une droite.

3/ Etude graphique d'une distribution à partir d'un échantillon

3.1 - Représentation des observations

Si, pour un échantillon de n observations, n_i est le nombre cumulé des observations inférieures à une valeur donnée x_i, il suffira de placer sur le graphique les points d'abscisses x_i et d'ordonnées $\frac{n_i}{n}$ (en % ou ‰)

Si ces points sont approximativement alignés, on tracera la droite D représentant au mieux leur ensemble (droite de Henry), et on pourra admettre que la variable x est normalement distribuée dans la population d'où provient l'échantillon. (Fig. 2.1).

3.2 - Estimation des paramètres m et σ.

- la moyenne m de la population sera estimée par l'abscisse x = m du point M d'intersection de la droite D avec l'axe des abscisses.

- l'écart-type σ sera estimé par la différence des abscisses entre le point M et le point S correspondant au point d'ordonnée y = 0,8413 de l'échelle fonctionnelle (c'est-à-dire au point d'ordonnée u = 1 de l'échelle graduée en écarts-types).

Pour diminuer l'erreur graphique, on aura intérêt à utiliser de préférence les points d'ordonnées u = 2 (ou u = 3) et à diviser par 2 (ou 3) la différence d'abscisses correspondantes.

Fig. 2.1 - Droite de Henry.

3.3 - Exemple

Etude d'un échantillon de 250 résistances.

Les observations ayant été réparties en classes d'amplitude 2/ohms, on a obtenu les résultats donnés dans le tableau ci-après :

Classes	Nombre d'observations	Nombres cumulés	% du total
104		0	0
106	2	2	0,8
108	2	4	1,6
110	4	8	3,2
112	15	23	9,2
114	14	37	14,8
116	30	67	26,8
118	33	100	40,0
120	50	150	60,0
122	28	178	71,2
124	37	215	86,0
126	15	230	92,0
128	10	240	96,0
130	7	247	98,8
132	2	249	99,6
134	1	250	100,0
	250		

On notera que, dans ce tableau, les effectifs des classes sont placés dans les intervalles de l'échelle des classes, mais que les effectifs cumulés "jusqu'à", sont placés sur les lignes des limites de classes correspondantes.

Les points représentatifs, (fig. 2. 2) étant approximativement alignés, on peut en conclure :

1/ que la distribution est très voisine d'une distribution normale,

2/ que la médiane, c'est-à-dire aussi la moyenne, peut être estimée à 119 ohms, l'écart-type étant égal à 124 - 119 = 5 ohms.

4/ Annexe théorique

Soit X une variable normale (m, σ) dont la fonction de répartition est :

$$Pr(X \leqslant x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

soit :

$$F(x) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = F(u)$$

avec :

$$u = \frac{x - m}{\sigma}$$

La correspondance entre les valeurs de u et de x définit donc une droite appelée droite de Henry.

Si, pour un échantillon de n observations, n_i est le nombre de celles qui sont inférieures à x_i, le rapport n_i/n peut être considéré comme une estimation de $F(x_i)$, c'est-à-dire de $F(u_i)$ avec $u_i = \frac{x_i - m}{\sigma}$.

En utilisant cette estimation, la table de la loi normale donnera dans cette hypothèse, la valeur correspondante de u_i, ce qui permettra de placer les points (x_i, u_i) et de voir s'ils satisfont à la relation ci-dessus.

En utilisant l'échelle fonctionnelle avec $F(u_i) = \frac{n_i}{n}$, on pourra se dispenser de passer par la table et on placera directement les points $\left(x_i, \frac{n_i}{n}\right)$

Fig. 2.2 - Loi normale - Etude graphique.

III - PAPIER A DEUX ECHELLES FONCTIONNELLES : LOI NORMALE ET LOGARITHMES (PAPIER LOG-NORMAL)

1/ But

Ce papier permet de résoudre graphiquement les problèmes suivants :

1/ Rechercher si la distribution de log x, ou de log $(x - x_0)$, est normale (x_0 étant une constante avec $x - x_0 > 0$). On dira que la distribution de x est log-normale.

2/ Si cette hypothèse paraît acceptable, estimer la moyenne et l'écart-type de log x (ou de log $(x - x_0)$).

2/ Définition

L'échelle des ordonnées est la même que dans le précédent : échelle graduée en fréquences cumulées à laquelle on associe une échelle graduée en écarts-types.

L'échelle des abscisses est une échelle logarithmique.

3/ Etude graphique d'une distribution à partir d'un échantillon.

3.1 - Représentation des observations

Comme dans le cas précédent, si $\frac{n_i}{n}$ est la fréquence cumulée des observations inférieures à $x = x_i$, on placera sur le graphique les points de coordonnées cotés x_i et n_i/n (ou 100 n_i/n si l'échelle des fréquences cumulées est graduée en centièmes).

Si ces points sont approximativement alignés, on tracera la droite D représentant au mieux leur ensemble et on pourra admettre que la variable log x est normalement distribuée dans la population d'où provient l'échantillon.

3.2 - Détermination de la moyenne et de l'écart-type de log x (Fig. 3.1).

3.2.1 - *Moyenne de log x*

L'abscisse a du point M d'intersection de la droite D avec l'échelle logarithmique donnera une estimation de la médiane de la variable X.

La valeur moyenne m de log x sera définie par m = log a.

Si on a pris la précaution d'associer à l'échelle fonctionnelle des abscisses une échelle linéaire graduée en log x, on pourra lire directement m.

3.2.2 - *Ecart-type de log x*

Il sera défini comme précédemment par :

$$\sigma = \log b - \log a,$$

Fig. 3.1.

b étant l'abscisse du point d'ordonnée 0,8413 sur l'axe des fréquences cumulées ou u = 1 sur l'axe gradué en écarts-types.

4/ Remarque

Les échelles logarithmiques des papiers que l'on trouve dans le commerce ont un module graphique déterminé, (par exemple 12,5 ou 25 cm, pour l'intervalle entre deux puissances successives de 10).

Leur emploi devient difficile si la variable X varie dans un intervalle relativement petit.

4.1 - Emploi du papier pour loi normale

On aura alors intérêt à utiliser le papier précédent à échelle fonctionnelle loi normale en ordonnée, l'axe des abscisses ayant une échelle linéaire pouvant alors être graduée en log x avec une unité graphique que l'on pourra choisir au mieux.

Ceci exigera évidemment que l'on transforme préalablement x en log x, à l'aide d'une table, mais la lecture de m et de σ se fera directement sur le graphique.

Il suffira encore de placer les points des coordonnées (log x_i, n_i/n) et de tracer, s'il y a lieu, la droite représentant l'ensemble de ces points (Fig. 3.2).

La moyenne m de log x sera l'abscisse du point M' et l'écart-type sera mesuré par la distance M'B' :

$$\sigma = b' - m.$$

Fig. 3.2

5/ Etude graphique de la distribution normale de log $(x - x_0)$

5.1 - Emploi du papier log-normal

Il arrive fréquemment qu'au lieu de log x, ce soit la variable $\log(x - x_0)$ qui soit distribuée normalement, x_0 étant une constante en général positive.

S'il en est ainsi, le papier log-normal permettra encore de s'en apercevoir et d'estimer les trois paramètres x_0, m et σ.

Comme précédemment, on placera sur le graphique log-normal les points de coordonnées cotés x_i et n_i/n (Fig. 3.3).

La possibilité d'une distribution normale de log $(x - x_0)$ - sans que cela soit une certitude - se manifestera par le fait que les points $(x_i, n_i/n)$ tendront à se placer au voisinage d'une courbe C à courbure régulière, concave vers les grandes valeurs de x (si $x_0 > 0$), et tendant à devenir rectiligne lorsque x augmente.

Etant données les propriétés de la fonction logarithmique, la transformation de x_i en $x_i - x_0$, à ordonnée invariable, pour une valeur arbitraire de x_0 aura d'autant plus d'effet sur $\log(x_i - x_0)$ que x_i sera plus petit.

On pourra, par tatonnement essayer de déterminer x_0 de la manière suivante :

On tracera à l'estime une asymptote à la courbe C et on mesurera le décalage $x_0 = x_1 - x_1'$ qui lui correspond pour une valeur de x_1 située dans la zone des petites valeurs de x.

Ce même décalage sera appliqué à tous les points du graphique dont les abscisses deviendront $x_i - x_0$, les ordonnées n_i/n restant inchangées

Fig. 3.3

On obtiendra une nouvelle courbe C'. Si elle se rapproche d'une droite, sa concavité restant dans le même sens, on fera un nouvel essai avec un nouveau décalage supérieur à x_0 (si la concavité changeait de sens, on essayerait un décalage plus petit).

Si la distribution de log $(x - x_0)$ est normale, la convergence graphique vers la droite de Henry correspondante (D) est rapide et il est rare d'avoir besoin de plus de deux ou trois tâtonnements pour l'obtenir.

5.2 - Estimation des paramètres de la loi normale de log $(x - x_0)$.

Noter que dans ce cas :

$m = \log a$	sera la moyenne de	$\log (x - x_0)$
$\sigma = \log b - \log a$	sera l'écart-type de	$\log (x - x_0)$

5.3 - Emploi du papier pour loi normale

Pour les raisons indiquées plus haut, on pourra encore utiliser un papier à seule échelle fonctionnelle (normale) en ordonnées, l'axe des abscisses ayant une échelle linéaire que l'on pourra graduer à sa convenance en log x (Fig. 3.4).

Le décalage essayé, x_0, correspondant au transfert du point d'abscisse $\log x_1$ au point d'abscisse $\log x'_1 = \log (x_1 - x_0)$, situé sur l'asymptote estimée, sera $x_0 = x_1 - x'_1$, ces deux valeurs étant fournies par la table de logarithmes à partir des logarithmes lus sur le graphique.

Fig. 3.4

Pour chaque valeur, essayée x_0, l'usage de la table sera nécessaire pour remplacer toutes les valeurs log x_i par les valeurs correspondantes log $(x_i - x_0)$.

Si la transformation conduit finalement à des points à peu près alignés sur une droite D, on lira directement sur le graphique.

$$m = \text{abscisse de } M' = \text{moyenne de } \log (x - x_0)$$

$$\sigma = b' - m = \text{écart-type de } \log (x - x_0)$$

5.4 - Autre moyen d'estimation de x_0

Si M_1 et M_3 sont deux points de la courbe C tracée à l'estime, et dont les ordonnées sont symétriques par rapport à 0,50, la transformation de x en x_0 donnera, dans le cas d'une distribution log-normale, deux points dont les abscisses mesurées en log $(x - x_0)$ devront être symétriques par rapport à log $(x_2 - x_0)$, x_2 étant l'abscisse du point d'ordonnée 0,50 de la courbe C (Fig. 3.5).

On devra avoir :

$$2 \log (x_2 - x_0) = \log (x_1 - x_0) + \log (x_3 - x_0)$$

$$(x_2 - x_0)^2 = (x_1 - x_0)(x_3 - x_0)$$

soit :

$$x_0 = \frac{x_2^2 - x_1 x_3}{2x_2 - x_1 - x_3}$$

qui fournira une estimation de x_0, basée sur la courbe C préalablement tracée à l'estime.

Fig. 3.5

5.5 - Estimation de la moyenne m et de l'écart-type σ de log x (Distribution normale de log x)

Si Q_1 et Q_3 sont les quartiles de la distribution de x, et Q_2 la médiane, la transformation en log x conservant les fractiles, on aura, en raison de la symétrie de la distribution normale de log x :

$$m = \frac{1}{2} (\log Q_1 + \log Q_3)$$

On peut obtenir une estimation de variance plus petite en calculant une moyenne de plusieurs fractiles, moyenne pondérée proportionnellement aux ordonnées correspondantes de la courbe normale, par exemple :

$$m = \frac{\log Q_1 + \log Q_3 + 1,255 \log Q_2}{3,255}$$

$$\left(1,255 = \frac{\text{ordonnée de } Q_2}{\text{ordonnée de } Q_1}\right)$$

De même, compte tenu de :

$$\log Q_3 - \log Q_1 = 2 \times 0,6745\, \sigma = 1,349\, \sigma$$

On estimera σ par :

$$\sigma = 0,7413 (\log Q_3 - \log Q_1)$$

S'il s'agit de la distribution normale de log $(x - x_0)$ et si x_0 est connu, on aura de même :

$$m_{\log(x - x_0)} = \frac{\log (Q_1 - x_0) + \log (Q_3 - x_0) + 1,255 \log (Q_2 - x_0)}{3,255}$$

$$\sigma_{\log(x - x_0)} = 0,7413 \, [\log(Q_3 - x_0) - \log (Q_1 - x_0)]$$

5.6 - Exemple - On a étudié sur un lot de 5348 transistors le rapport qui existe, en émetteur commun, entre le courant de sortie et le courant d'entrée, en signal sinusoïdal de faible amplitude.

Le graphique ci-après (fig. 3.6) montre la distribution log normale des observations (document obligeamment communiqué par la Société Industrielle Française des Transistors).

S.I.F.TRA

Fig. 3.6 - Distribution des gains (β) mesurés sur 5 348 transistors 7 700. Fabrication de janvier 1963.

6/ Annexe théorique

Si $y = x - x_0$ avec $x > x_0$, est une variable telle que $\log_e y$ (1) ait une distribution normale de moyenne m et d'écart-type , la distribution de x est dite log-normale.

La loi de :

$$u = \frac{\log_e y - m}{\sigma}$$

sera :

$$f(u)\ du = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{u^2}{2}}\ du$$

La loi de x, définie par :

$$g(x)\ dx = f(u)\ du$$

sera :

$$g(x)\ dx = \frac{1}{x - x_0}\ \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}\left[\frac{\log_e (x - x_0) - m}{\sigma}\right]^2} dx$$

6.1 - Médiane de $X - x_0$

Elle correspond à la médiane de $\log_e y$, c'est-à-dire à la moyenne m de cette variable normale.

$$\text{Log}_e \text{ Méd } (X - x_0) = \text{Méd. } \text{Log}_e(X - x_0) = \text{Moy. } \text{Log}_e(X - x_0) = m$$

d'où :

$$\text{Médiane de } (X - x_0) = e^m$$

$$\text{Médiane de } X = x_0 + e^m$$

6.2 - Moyenne de X

Elle est définie par :

$$E(X) = \int_{x_0}^{\infty} x\ g(x)\ dx = \int_{-\infty}^{\infty} (x_0 + e^{m+ou})\ \frac{1}{\sqrt{2\pi}}\ e^{-\frac{u^2}{2}}\ du$$

soit :

$$m_1 = E(X) = x_0 + e^{m + \frac{\sigma^2}{2}}$$

---

(1) Noter que dans les formules qui suivent, m et σ sont les paramètres de la distribution du logarithme népérien de $(x - x_0)$.

6.3 - Valeur modale de X

Elle est définie par le maximum de :

$$g(x) = f(u) \frac{du}{dx} = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} u'$$

c'est-à-dire par le maximum de $\log_e g(x)$, ou encore de :

$$W = -\frac{u^2}{2} + \log_e u'$$

on a :

$$\frac{dW}{dx} = -u u' + \frac{u''}{u'} = 0 \quad \text{pour}$$

$$u'' = uu'^2$$

soit :

$$\frac{-1}{\sigma(x - x_0)^2} = \frac{\log_e(x - x_0) - m}{\sigma} \cdot \frac{1}{\sigma^2(x - x_0)^2}$$

d'où :

$$\text{Mode de } x = x_0 + e^{m-\sigma^2}$$

6.4 - Variance de X

Le moment d'ordre deux de $X - x_0$ est :

$$m_2(X - x_0) = m_2(e^{m + \sigma u})$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{2(m + \sigma u)} e^{\frac{u^2}{2}} du$$

$$= e^{2(m + \sigma^2)}$$

d'où :

$$\text{Variance de } X = \text{variance de } (X - x_0) = m_2 - m_1^2$$

$$= e^{2(m + \sigma^2)} - e^{2m + \sigma^2}$$

d'où :

$$\sigma_x = e^{m + \sigma^2} \sqrt{1 - e^{-\sigma^2}}$$

On a donc en résumé, pour la variable X :

médiane	$x_0 + e^m$
mode	$x_0 + e^{m - \sigma^2}$
moyenne	$x_0 + e^{m + \frac{\sigma^2}{2}}$
variance	$e^{2(m + \sigma^2)}[1 - e^{-\sigma^2}]$

avec :

m = moyenne de $\log (x - x_0)$,

σ = écart-type de $\log_e(x - x_0)$

soit, pour l'exécution des calculs :

$m = 2,3026 \times$ moyenne de $\log_{10} (x - x_0)$

$\sigma = 2,3026 \times$ écart-type de $\log_{10}(x - x_0)$

On peut aussi écrire :

$$E(X) = x_0 + 10^{m' + 2,3026 \frac{\sigma'^2}{2}}$$

$$V(X) = 10^{2\left(m' + 2,3026 \frac{\sigma'^2}{2}\right)} - 10^{2m' + 2,3026 \sigma'^2}$$

avec :

m' = moyenne de $\log_{10}(x - x_0)$

σ' = Ecart-type de $\log_{10}(x - x_0)$.

6.5 - Estimation de x_0

De la relation :

$$u = \frac{\log_e(x - x_0) - m}{\sigma}$$

on tire :

$$\log_e (x - x_0) = u + m\sigma$$

soit :

$$\text{Log}_e(x = \text{Log}_e(x_0 + e^{m+u\sigma}) = (m + \sigma u) + \text{Log}_e \left[1 + \frac{x_0}{e^{m + \sigma u}}\right]$$

équation de la courbe cumulative dans le système $(u, \log_e x)$.

Pour x, et par conséquent u augmentant indéfiniment, $\log_e x - (m + \sigma u)$ tend vers zéro

La courbe cumulative admet pour asymptote la droite d'équation :

$$\mathrm{Log}_e\ x = m + \sigma u$$

Pour x_0 positif, ce qui est le cas le plus fréquent, et pour x tendant vers x_0, c'est-à-dire u tendant vers $-\infty$, $\log_e x - \log_e x_0$ tend vers zéro, la courbe admet l'asymptote verticale :

$$\mathrm{Log}_e\ x = \mathrm{Log}_e\ x_0 \longrightarrow x = x_0$$

Dans le cas d'une distribution tronquée pour les grandes valeurs de x, l'asymptote oblique étant mal précisée, on pourra trouver là une indication pour estimer x_0.

IV - PAPIER LOI BINOMIALE

1/ But

Le papier bonomial permet de résoudre graphiquement, ou éventuellement avec quelques calculs simples, certains problèmes d'estimation ou de comparaison qui se posent dans les problèmes de contrôle de réception ou de contrôle en cours de fabrication ou plus généralement dans les problèmes statistiques faisant intervenir la loi binomiale.

Citons, par exemple, les problèmes suivants :

- Comparaison d'une fréquence observée sur un échantillon à une fréquence théorique (par exemple pourcentage de défectueux).

- Intervalle de confiance d'un pourcentage à partir d'un échantillon.

- détermination d'un plan d'échantillonnage correspondant à des risques donnés, (α, p_1, β, p_2).

2/ Définitions

2.1 - Description du quadrillage binomial (notation anglaise : B.I.P.P. Binomial Probability Paper).

C'est un papier quadrillé à deux axes de coordonnées rectangulaires dont les deux échelles identiques sont des échelles fonctionnelles "racine carrée", c'est-à-dire telles que les points A (ou B) cotés x (ou y) sur les axes de coordonnées sont définis par les longueurs :

$$OA = \sqrt{x} \qquad OB = \sqrt{y}$$

à partir d'une unité graphique préalablement choisie (Fig. 4.1).

Fig. 4.1 - Papier loi binomiale.

2.2 - Représentation d'une population ou d'un lot

Cette représentation est en principe indépendante de l'effectif du lot.

Pour un effectif arbitraire N, un lot caractérisé par Np pièces défectueuses et $Nq = N(1 - p)$ pièces bonnes, sera représenté par une droite OP, appelée caractéristique du lot, issue de l'origine et passant par le point P de coordonnées cotées Np sur l'axe des y (par exemple) et $N(1 - p)$ sur l'axe des abscisses.

L'intersection de cette droite OP avec le quart de circonférence passant par les points (0, 100) et (100, 0) permet de lire directement les valeurs correspondantes de p et 1 - p (en %), respectivement sur les axes Oy et Ox.

2.3 - Représentation d'un prélèvement (d'un échantillon)

Un échantillon d'effectif n, qui représenterait parfaitement la population, se partagerait en deux groupes d'effectifs np et $n(1 - p)$.

Il serait représenté par le point M de la droite OP, ayant des coordonnées cotées :

$$x = n(1 - p) \qquad y = np$$

soit pour les longueurs :

$$OA = \sqrt{x} \qquad OB = \sqrt{y}$$

avec :

$$OM = \sqrt{n}$$

Un échantillon observé donnera en réalité une proportion f de défectueux, soit $y' = nf$ défectueux et $x' = n(1 - f)$ pièces bonnes.

Il sera représenté par le point M' de coordonnées cotées x' et y', (point échantillon).

Dans la mesure où la distribution normale donne une bonne approximation de la loi binomiale (pratiquement si $np > 10$), la distance du point M' à la droite OP, est, pour les divers échantillons tirés de cette population, une variable satisfaisant à la loi normale avec une moyenne nulle et un écart-type σ sensiblement constant, égal à 1/2 mesuré à l'échelle du graphique.

Une échelle graduée en σ, annexée au graphique, permettra de mesurer cette distance en écarts-types et d'apprécier la probabilité qu'un tel écart puisse être atteint ou dépassé par le simple jeu des fluctuations aléatoires d'échantillonnage.

L'échelle fonctionnelle peut évidemment être choisie de manière à donner à l'unité l'écart $\sigma = 1/2$ une longueur simple, par exemple 5 ou 10 mm.

On remarquera que si les nombres observés x', y' sont trop grands pour que le point M' soit placé sur le graphique, on pourra les remplacer par les nombres $\frac{x'}{10}$, $\frac{y'}{10}$, à condition de remplacer l'unité d'écart σ par :

$$\sigma' = \frac{\sigma}{\sqrt{10}}$$

3/ Résolution graphique de quelques problèmes

3.1 - Comparaison d'une fréquence observée f à une fréquence théorique p (Test de l'hypothèse p)

3.1.1 - *Problème*

L'écart entre la proportion observée f des pièces défectueuses dans un échantillon de n pièces et une proportion théorique p définissant un lot ou une population est-il significatif ?

3.1.2 - *Solution graphique* (Fig. 4.2).

Fig. 4.2 - Signification d'un écart.

- Tracé de la droite OP caractéristique de la population (de l'hypothèse à tester soit p = 20 %).
- Mise en place du point échantillon M' (n = 80 avec y' = 25 défectueux, x' = 55.
- Mesure de la distance du point M' à la droite OP en écarts-types σ .
- Si cette distance est inférieure par exemple à 2σ, on pourra conclure que l'écart n'est pas significatif d'un désaccord réel entre f et p. (En prenant 2σ comme limite, si la distance est supérieure à 2σ, et que ceci conduise à considérer l'écart comme significatif, on a moins de cinq chances sur 100 de se tromper, c'est-à-dire de refuser une hypothèse qui pourrait cependant être exacte).

On pourra évidemment se fixer un écart limite $t\sigma$ correspondant à tout autre risque d'erreur que l'on se sera fixé (voir annexe I).

Pour l'exemple envisagé ci-dessus, on constate que la distance du point échantillon M' à la droite OP est légèrement supérieure à 2σ : un tel écart considéré en valeur absolue à moins d'une chance sur 20 d'être dû aux seules fluctuations d'échantillonnage.

Plus généralement, la zone d'acceptation d'une hypothèse p à un risque donné, par exemple $\alpha = 0,05$ de rejeter une hypothèse vraie sera limitée par deux droites parallèles à la droite théorique OP(p, 1 - p), situées à la distance 2σ.

Quel que soit n - dans les limites où l'approximation normale de la loi binomiale est acceptable - tout point M'(x', y') situé entre ces deux parallèles conduira à l'acceptation de l'hypothèse p au risque α de la rejeter alors qu'elle est vraie.

Ce tracé fait à partir de deux droites théoriques $OP_1(p_1, 1 - p_1)$ et OP_2, fournira dans certains cas un moyen de discrimination entre l'hypothèse p_1 et l'hypothèse p_2.

3.2 - Détermination d'un intervalle de confiance

3.2.1 - *Problème*

On a observé y' pièces défectueuses dans un échantillon de n pièces soit une estimation $f = \frac{y'}{n}$ de la proportion p de pièces défectueuses dans le lot. Déterminer un intervalle (p_1, p_2) ayant une probabilité donnée, par exemple $1 - \alpha = 0,95$ de contenir la vraie valeur inconnue p.

3.2.2 - *Solution graphique*

- Placer le point échantillon M'
- Tracer la circonférence (C) de centre M' et de rayon 2σ
- Tracer les tangentes issues de l'origine à cette circonférence.
- Les ordonnées des intersections de ces tangentes avec la circonférence de centre O et de rayon 100 (à l'échelle fonctionnelle) donneront en % les limites de l'intervalle de confiance pour p (Fig. 4.3).

Fig. 4.3 - Intervalle de confiance.

3.2.3 - *Exemple numérique.*

Un échantillon de n = 300 a donné 75 pièces défectueuses, d'où le point échantillon M' (y' = 75, x' = 225).

Les ordonnées des points d'intersection avec la circonférence de centre O et de rayon 100, des tangentes issues de O à la circonférence de centre M' et de rayon 2σ sont 20 et 30.

On peut donc conclure qu'il y a 95 chances sur 100 pour que l'intervalle (20 % - 30 %) contienne la vraie valeur inconnue de p.

3.3 - Détermination d'un plan d'échantillonnage

3.3.1 - *Problème.*

Déterminer l'effectif n de l'échantillon et le nombre maximum c de pièces défectueuses qu'il peut contenir pour un plan devant satisfaire aux conditions suivantes définissant le risque α du fournisseur et le risque β de l'acheteur.

Probabilité d'acceptation $1 - \alpha = 0,95$ si la proportion de défectueux est $p_1 = 3$ % (AQL).

Probabilité d'acceptation $\beta = 0,10$ si la proportion de défectueux est $p_2 = 12$ % (LTPD).

3.3.2 - *Solution graphique* (Fig. 4.4).

- Tracé des droites caractéristiques D_1 et D_2 correspondant à p_1 et p_2.
- Tracé de la parallèle D_1' à D_1 située à $1,64\sigma$ au-dessus (voir annexe II). Tracé de la parallèle D_2' à D_2 située à $1,28\sigma$ au-dessous.
- Ces deux parallèles se coupent au point M de coordonnées y = 5, x = 61
- On utilisera l'échantillon d'effectif n = x + y - 1 = 65, la limite d'acceptation étant c = y - 1 = 4.

(La correction -1 est introduite pour des raisons de continuité : la distribution binomiale étant celle d'une variable discrète alors que la distribution normale est celle d'une variable continue ; son importance est pratiquement négligeable en ce qui concerne l'effectif de l'échantillon).

On notera que, pour cet exemple - qui, en raison de la petite valeur de p_1 pourrait paraître être au-delà des possibilités de l'approximation normale - les résultats graphiques obtenus par un dessinateur moyen sont excellents.

L'étude théorique à partir des tables de la loi binomiale donne en effet les résultats ci-après pour $p_1 = 0,03$ et $p_2 = 0,12$:

n = 65 et C = 4	$P_1 = 0,954$	$P_2 = 0,097$
n = 66 et C = 4	$= 0,951$	$= 0,091$

3.3.3 - *Remarque.*

On notera que pour p_1 et p_2 donnés, les droites D_1' et D_2', correspondant à des risques donnés α et β, partagent l'angle D_1 O D_2 en quatre régions correspondant à des risques α et β inférieurs ou supérieurs, suivant le cas, à ceux qui ont servi au tracé de D_1' et D_2' et que l'on pourrait préciser en comparant à σ les distances d'un point échantillon particulier aux droites D_1 et D_2.

Fig. 4.4 - Plan d'échantillonnage (p_1, α, p_2, β)

3.4 - Comparaison de lots d'un même ensemble

3.4.1 - *Problème.*

En cours de fabrication on a produit un certain nombre de lots de n_i pièces parmi lesquels se trouvent y_i pièces défectueuses, soit une proportion $f_i = \frac{y_i}{n_i}$. On se propose de voir si l'un de ces lots correspond à un déréglage du procédé de fabrication.

3.4.2 - *Etude graphique.*

Si le procédé est resté bien réglé, les diverses valeurs f_i sont des estimations indépendantes de la proportion p de défectueux dans la fabrication, proportion dont l'estimation d'ensemble est :

$$p = \frac{\sum y_i}{\sum n_i}$$

d'où la solution.

- Tracé de la droite D représentative de cette estimation d'ensemble.
- Tracé de deux droites parallèles D_1 et D_2 distantes de 3σ de la droite D.
- Mise en place des points échantillons M_i sur le graphique.
- Si tous ces points se placent à l'intérieur de la zone comprise entre D_1 et D_2, on pourra admettre l'homogénéité de l'ensemble.
- Un point extérieur à cette zone sera considéré comme caractéristique d'un lot correspondant à un déréglage.

On notera que dans ce cas l'élimination de ce lot modifierait la position de la droite moyenne et qu'il pourrait en résulter un changement dans le jugement relatif à un autre lot.

Le procédé envisagé n'a donc qu'une valeur indicative généralement suffisante en pratique pour attirer l'attention sur une anomalie éventuelle

4/ Annexe Théorique

La variable $f = \frac{x}{n}$ est une variable aléatoire caractérisée par :

$$E(f) = p \qquad V(f) = \frac{p(1 - p)}{n}$$

Cette variance dépend de la vraie valeur p du paramètre que l'on veut estimer ou tester à partir d'un échantillon d'effectif n tiré de la population (p, 1 - p).

On peut envisager un changement de variable $\theta = \Phi(f)$, tel que la variance de θ soit indépendante de p.

Pour de petites variations de f autour de p, on peut considérer $\Phi(f)$ comme une fonction linéaire définie par :

$$\theta = \Phi(f) = \Phi(p) + (f - p)\ \Phi'(p)$$

avec :

$$E(\theta) = \Phi(p)$$

$$V(\theta) = V(f).\ [\Phi'(p)]^2$$

$$\sigma_\theta = \Phi'(p) \sqrt{\frac{p(1 - p)}{n}}$$

La condition $\sigma_\theta = c$, c ne dépendant pas de p, donne :

$$\Phi'(p) = \frac{c \sqrt{n}}{\sqrt{p(1 - p)}}$$

Il suffira de choisir une fonction Φ satisfaisant à la condition :

$$\Phi(p) = c \sqrt{n} \int_0^1 \frac{dp}{\sqrt{p(1 - p)}}$$

Si l'on pose :

$$p = \mathrm{Sin}^2\theta$$

on obtient :

$$\Phi(p) = 2\ c \sqrt{n}\ \mathrm{Arc\ Sin} \sqrt{p}$$

En prenant $2\ c \sqrt{n} = 1$, on obtient la variable transformée :

$$\theta = \Phi(f) = \mathrm{Arc\ Sin} \sqrt{f}$$

dont la variance est, dans les conditions envisagées ci-dessus :

$$V(\theta) = \frac{1}{4} \frac{1}{p(1 - p)} \frac{p(1 - p)}{n} = \frac{1}{4n}$$

doit :

$$\sigma_\theta = \frac{1}{2 \sqrt{n}} \qquad \text{(pour } \theta \text{ mesuré en radians)}$$

Pour θ mesuré en degrés, on aurait :

$$\sigma_\theta = \sqrt{\frac{821}{n}}$$

Dans la mesure où l'approximation normale est valable pour la loi binomiale, la variable θ considérée comme fonction linéaire de f, est pratiquement une variable normale, pour de petites variations de f autour de p.

L'étude de la fonction $\theta = \mathrm{Arc\ Sin} \sqrt{f}$ montre que pour $0,1 < f < 0,9$, la linéarité de cette fonction est pratiquement assurée dans d'assez grands intervalles de variation de f.

Dans ces conditions, la population (p, 1 - p) est représentée par la droite OP telle que xOP = θ, soit :

$$\text{Sin } \theta = \sqrt{p}, \qquad \text{Cos } \theta = \sqrt{1 - p}$$

Le point M représentant un échantillon parfait a pour coordonnées np et n(1 - p), dans les échelles fonctionnelles d'ordonnées et d'abscisses, avec OM = $\sqrt{n}$.

De même, le point échantillon réel M', de coordonnées nf et n(1 - f), sera tel que :

$$xOM' = \theta' = \theta + d\theta, \qquad OM' = \sqrt{n}$$

Il en résulte que l'angle $d\theta$ et la longueur $\lambda = MM' = \sqrt{n}\, d\theta$, que l'on peut assimiler à la distance du point M' à la droite OP sont des variables aléatoires sensiblement normales avec les écarts-types.

$$\sigma_\theta = \frac{1}{2\sqrt{n}}, \qquad \sigma_\lambda = \sqrt{n}\ \sigma_\theta = \frac{1}{2}$$

La quantité $\sigma_\lambda = \frac{1}{2}$, mesurée à l'échelle graphique choisie pour établir le quadrillage est donc une caractéristique de dispersion de la distance du point échantillon M' à la droite de la population OP.

ANNEXES

Loi normale de Moyenne m et d'écart-type σ

ANNEXE I

Probabilité P pour qu'une variable normale soit comprise dans un intervalle $m - t\sigma$ à $m + t\sigma$.

100 P	t
99, 9	3, 29
99	2, 58
98	2, 33
97	2, 17
96	2, 05
95	1, 96
94	1, 88
93	1, 81
92	1, 75
91	1, 69
90	1, 64

100 P	t
89	1, 60
88	1, 55
87	1, 51
86	1, 48
85	1, 44
84	1, 40
83	1, 37
82	1, 34
81	1, 31
80	1, 28
75	1, 15

Ex. : 95 % des valeurs de la variable considérée sont comprises dans l'intervalle $m - 1,96\,\sigma$ à $m + 1,96\,\sigma$.

Dans un échantillon d'effectif n, provenant d'un lot dans lequel la proportion des pièces défectueuses est p, la proportion f de pièces défectueuses observées sera dans 95 % des cas comprise dans l'intervalle $p - 1,95\sqrt{\frac{p(1-p)}{n}}$ à $p + 1,96\sqrt{\frac{p(1-p)}{n}}$, leur nombre étant avec la même probabilité de 0, 95, compris entre $np - 1,96\sqrt{np(1-p)}$ et $np + 1,96\sqrt{np(1-p)}$.

ANNEXE II

Probabilité P pour qu'une variable normale soit inférieure à $m + t\sigma$

100 P	t
80	0, 84
85	1, 04
90	1, 28
95	1, 64
99	2, 32

Ex. : Si un lot contient une proportion $p = 0,20$ (20 %) de pièces défectueuses, et si on prélève un échantillon de $n = 100$ pièces, il y a 90 chances sur 100 pour que le nombre des défectueux observés soit inférieur à $np + 1,28 \sqrt{np(1 - p)}$ c'est-à-dire à 25, ou encore une chance sur 100 pour qu'il excède 25.

Nota - Ces approximations de la loi binomiale par la loi normale supposent que np est au moins supérieur à 10.

V - PAPIER LOI DE WEIBULL

1/ But

Parmi les nombreuses distributions théoriques utilisées dans les études de fiabilité, la loi de Weibull, dont la loi exponentielle n'est qu'un cas particulier, joue un rôle important.

Elle dépend de deux (ou même de trois) paramètres et les calculs nécessaires pour estimer ces paramètres, et par conséquent la valeur moyenne qui en est fonction, sont longs et difficiles.

Le papier à échelles fonctionnelles présenté par Allan Plait (1) permet une solution graphique de ce problème moyennant quelques calculs arithmétiques élémentaires (Fig. 5. 1).

2/ Définition et emploi

2. 1 - Réalisation d'un quadrillage loi de Weibull

Le quadrillage de ce papier comporte deux axes de coordonnées rectangulaires dont les échelles sont les suivantes :

- axe des abscisses : échelle logarithmique pour la variable observée x, par exemple le temps de fonctionnement ou le nombre de fonctionnements avant défaillance.

- axe des ordonnées : Echelle logarithmique-logarithmique pour les fréquences cumulées F(x) des défaillances correspondant aux valeurs de x, ces fréquences cumulées sont exprimées en %.

A ces deux échelles qui seront utilisées pour représenter les observations sont associées sur les axes parallèles deux échelles linéaires directement graduées :

- l'une en : $X_i = \log_e x_i$

$X_i = 0$ pour $x_i = 1$

$X_i = 1$ pour $x_i = 2,71828 = e$

- l'autre en : $Y_i = \log_e \log_e \dfrac{1}{1 - F(x_i)}$

$Y_i = 0$ pour $F(x_i) = \dfrac{e - 1}{e} = 0,6321$

$Y_i = 1$ pour $F(x_i) = \dfrac{e^e - 1}{e^e} = 0,934$

Ces échelles permettront les transformations de logarithmes népériens en nombres sans passer par l'intermédiaire d'une table.

---

(1) Allan Plait : "The Weibull Distribution with tables" - Industrial Quality Control. Novembre 1962.

Fig. 5.1 - Etude graphique de la loi de Weibull.

2.2 - Etude graphique d'un test de durée de vie sur un échantillon.

2.2.1 - *Représentation des observations.*

On placera sur le graphique les points de coordonnées x, F(x) en %.

Si ces points sont alignés, ou pratiquement, si l'on peut tracer une droite D représentant convenablement l'ensemble des points, on peut conclure à une loi de Weibull.

2.2.2 - *Estimation de la valeur moyenne m de la variable x.*

Cette estimation comporte trois stades successifs :

- Pente de la droite

Cette pente β pourra être lue directement en cherchant dans le rapporteur de pente figurant dans le coin supérieur gauche du diagramme, le rayon parallèle à la droite D.

On peut aussi calculer cette pente en prenant deux points quelconques $A(X_1, Y_1)$ et $B(X_2, Y_2)$ sur la droite, dont on lira les coordonnées dans le système (X, Y), on aura :

$$\beta = \frac{Y_2 - Y_1}{X_2 - X_1}$$

- Estimation de la moyenne m

On lira d'abord sur l'axe des x, l'abscisse η du point de la droite D dont l'ordonnée est Y = O dans le système (X Y).

La valeur moyenne cherchée m (souvent désignée par la notation MTBF - Medan Time Between Failure) est liée à η par la relation :

$$m = \eta . \Gamma\left(1 + \frac{1}{\beta}\right)$$

$\Gamma\left(1 + \frac{1}{\beta}\right)$ est une fonction de β dont on pourrait trouver la valeur dans les tables, mais plus simplement une échelle à double graduation donne la correspondance entre β et $\Gamma\left(1 + \frac{1}{\beta}\right)$: une simple lecture donne $\Gamma\left(1 + \frac{1}{\beta}\right)$ en fonction de β.

Il suffit alors de calculer le produit :

$$m = \eta . \Gamma\left(1 + \frac{1}{\beta}\right)$$

3/ Exemple

Un échantillon de 20 relais a été soumis à un test de durée de vie, chaque fois que l'un des relais cessait de fonctionner on notait le nombre x de fois qu'il avait fonctionné avant défaillance.

Chacun des relais représentant 5 % de l'effectif de l'échantillon, on a obtenu le tableau ci-après des résultats observés.

Fonctionnements avant défaillance (en 10^5)	Défaillances cumulées en %
1, 90	5
3, 34	10
3, 65	15
4, 20	20
4, 72	25
5, 89	30
6, 10	35
6, 62	40
7, 92	45
8, 40	50
8, 50	55
9, 00	60
9, 60	65
11, 02	70
11, 95	75
12, 40	80
13, 03	85
13, 42	90
18, 07	95
20, 63	100

Les points ayant été placés sur le graphique, on peut tracer la droite dont la pente, calculée à partir des deux points A(0, - 4, 45) et B(2, 23, 0) est :

$$\beta = \frac{4,45}{2,23} \sim 2$$

Pour le point de cette droite d'ordonnée Y = 0, on lit :

$$x = \eta = 9,25 \times 10^5$$

D'autre part pour $\beta = 2$, l'échelle donnant le facteur $\Gamma\left(1 + \frac{1}{\beta}\right)$ donne la valeur :

$$\Gamma\left(1 + \frac{1}{\beta}\right) = 0,8862$$

d'où la valeur moyenne du nombre de fonctionnements :

$$m = 0,8862 \times 9,25 \times 10^5 = 8,2\ 10^5$$

4/ Annexe Théorique

4. 1 - Loi de Weibull

La fonction de répartition de la loi de Weibull (fréquence cumulée des durées de vie avant défaillances) est définie de manière la plus générale par :

$$\Pr(X \leqslant x) = F(x) = 1 - \exp.\left[-\left(\frac{x-\gamma}{\eta}\right)^{\beta}\right] \quad \text{pour} \quad x \geqslant \gamma$$

$$= 0 \quad \text{pour} \quad x < \gamma$$

La fréquence élémentaire entre x et x + dx a donc pour expression :

$$f(x)dx = \frac{\beta}{\eta}\left(\frac{x-\gamma}{\eta}\right)^{\beta-1} \exp -\left(\frac{x-\gamma}{\eta}\right)^{\beta} dx,$$

la valeur moyenne de x est :

$$m = E(x) = \int_{\gamma}^{\infty} x.\ f(x)dx = \gamma + \frac{\eta}{\beta}\Gamma\left(\frac{1}{\beta}\right)$$

La fonction Γ étant définie par :

$$\Gamma(u) = \int_0^{\infty} Z^{-Z} e \quad dZ,$$

$\int_0 e^{-z} z^{u-1} dz$

et telle que :

$$u\ \Gamma(u) = \Gamma(u + 1),$$

on a encore :

$$m = \gamma + \eta\ \Gamma\left(1 + \frac{1}{\beta}\right)$$

4.2 - Etude du cas particulier $\gamma = 0$

On a alors :

$$1 - F(x) = \exp\left[-\left(\frac{x}{\eta}\right)^{\beta}\right]$$

soit encore :

$$\frac{1}{1 - F(x)} = \exp\left(\frac{x}{\eta}\right)^{\beta}$$

La loi de Weibull est alors caractérisée par deux paramètres : β paramètre de forme (pour β voisin de 3,25 la loi de Weibull est très proche de la loi normale), η paramètre d'échelle, appelé vie caractéristique.

Il en résulte :

$$\mathrm{Log}_e \frac{1}{1 - F(x)} = \left(\frac{x}{\eta}\right)^{\beta}$$

$$\mathrm{Log}_e . \mathrm{Log}_e \frac{1}{1 - F(x)} = \beta\ (\mathrm{Log}_e\ x - \mathrm{Log}_e\ \eta)$$

relation linéaire de la forme :

$$Y = a\ X + b$$

avec :

$$Y = \mathrm{Log}_e\ \mathrm{Log}_e \frac{1}{1 - F(x)} \qquad a = \beta$$

$$X = \mathrm{Log}_e\ x \qquad b = -\ \beta\ \mathrm{Log}_e\ \eta.$$

Les échelles fonctionnelles des graphiques d'Allan Plait permettent de rechercher cette droite directement à partir des résultats observés x et F(x) sans avoir besoin de faire ces transformations.

4.3 - Cas particulier - Loi exponentielle

Pour $\gamma = 0$ et $\beta = 1$, on obtient la loi exponentielle :

$$f(x) = \frac{1}{\eta}\ e^{-\frac{x}{\eta}}$$

$$m = E(x) = \eta$$

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

$$\mathrm{Log}_e\ \mathrm{Log}_e \frac{1}{1 - F(x)} = \mathrm{Log}_e\ x - \mathrm{Log}_0\ \eta$$

La détermination graphique de $x = \eta$ pour $Y = O$, c'est-à-dire $F(x) = 0,6321$ se fera comme dans le cas précédent.

On aura ainsi : $m = \eta$

4.4 - Cas général $\gamma \neq 0$

Si les points $[x, F(x)]$ se placent au voisinage d'une courbe à concavité régulière (C), devenant de plus en plus voisine d'une droite lorsque x augmente, on pourra essayer de déterminer γ par tâtonnements successifs en examinant ce que deviennent les points, à ordonnées invariables lorsqu'on remplace x_i par $x_i - \gamma$.

De même que dans le cas de la détermination du paramètre x_0 de la loi log-normale, on sera guidé pour la première valeur à essayer de γ, par le décalage qui amène un point de la courbe (C), correspondant à une petite valeur de x, sur l'asymptote estimée de la courbe.

Pour une valeur convenable de γ, si elle existe, c'est-à-dire si la distribution est du type Weibull général, on pourra comme ci-dessus estimer la vie moyenne de $x' = x - \gamma$ et en déduire celle de x.

5/ Précision à attendre du graphique

L'échelle logarithmique de la variable x peut être utilisée avec une erreur relative de 1/100 ; l'échelle log-log des fréquences cumulées, graduée en % peut l'être avec une erreur relative variant de 1/20ème à 1/50ème lorsque la fréquence cumulée varie de 0,1 % à 99 %.

Les échelles linéaires correspondantes de X et Y permettent de lire ces valeurs respectivement à 0,01 et 0,02 près.

L'échelle graphique des pentes permet de déterminer β avec une erreur variable avec β allant de 0,02 à 0,05 environ lorsque β varie de zéro à trois.

Pour β estimé à 0,02 près par exemple, l'erreur sur $\Gamma\left(1 + \frac{1}{\beta}\right)$ est très variable avec β : 0,0001 pour $\beta = 3$, 0,005 pour $\beta = 1$, elle croît rapidement pour β inférieur à 1.

6/ Tables de la loi de Weibull

L'article d'Allan Plait contient les tables ci-après :

1/ Table de la fonction de densité f(u) de la loi de Weibull pour :

$\beta = 0,10(0,10) \ldots\ldots\ldots\ldots 3(1) \ldots\ldots\ldots\ldots 10$

$u = 0,10(0,10) \ldots\ldots\ldots\ldots 2,5$

2/ Table de la fonction de répartition F(u) pour :

$\beta = 0,10(0,10) \ldots\ldots\ldots\ldots 4$

$u = 0,10(0,10) \ldots\ldots\ldots\ldots 2,5$

Ces tables sont établies dans l'hypothèse $\gamma = 0$ et pour $\eta = 1$, c'est-à-dire pour $x = \frac{u}{\eta}$.

Leur emploi implique d'abord une estimation de η à partir des données c'est-à-dire de la valeur de x pour laquelle on a Y = 0, donc F(x) = 63, 2 %.

Dans l'exemple ci-dessus, ceci donne par interpolation $x \# 9,4 \times 10^5$ qui peut être utilisée comme approximation de la vie caractéristique η.

Les valeurs de $u_i = \frac{x_i}{\eta}$ correspondant aux valeurs observées de x_i permettent, dans la table II, de trouver à chaque fois la valeur de β pour laquelle F(u) serait égale à la fréquence cumulée correspondant à $x = x_i$.

L'accord éventuel de ces valeurs de β correspondant aux observations justifie l'hypothèse d'une loi de Weibull ($\gamma = 0$) et fournit une estimation de β.

En fait, dans un échantillon au hasard provenant d'une population distribuée selon la loi de Weibull, ces valeurs de β seront plus ou moins dispersées et le choix d'une valeur optimale assez subjectif : la méthode graphique semble devoir être préférée, aussi bien du point de vue efficacité que du point de vue commodité.

7/ Méthode d'estimation par le maximum de vraisemblance

Si on pose :

$$\eta^{\beta} = \frac{1}{\lambda}$$

on peut écrire :

$$f(x) = \beta\lambda\, x^{\beta-1}\, e^{-\lambda x^{\beta}}$$

Pour un échantillon de k éléments, le logarithme de la fonction de vraisemblance sera :

$$L = \text{Log } L = k \text{ Log } \beta + k \text{ Log } \lambda + (\beta - 1) \sum_{i=1}^{k} \text{Log } x_i - \lambda \sum_{1}^{k} x_i^{\beta}$$

d'où :

$$\frac{\partial L}{\partial \beta} = \frac{k}{\beta} + \sum \text{Log } x_i - \lambda \sum x_i^{\beta} \text{ Log } x_i = 0$$

$$\frac{\partial L}{\partial \lambda} = \frac{k}{\lambda} - \sum x_i^{\beta} = 0$$

on en déduit les formules :

$$\hat{\lambda} = \frac{k}{\sum x_i^{\beta}}$$

$$\hat{\beta} = \frac{k}{\hat{\lambda} \sum x_i^{\beta} \text{ Log } x_i - \sum \text{Log } x_i}$$

La résolution de ce système, par itération, à partir d'une valeur initiale convenablement choisie de β implique des calculs assez lourds.

Ces remarques sur l'estimation des paramètres de la loi de Weibull montrent l'intérêt de la méthode graphique envisagée ci-dessus.

VI - ABAQUES DE LA LOI DE POISSON

1/ But

La loi de Poisson ne dépendant que d'un seul paramètre, il est facile de la mettre sous forme d'un abaque permettant de résoudre graphiquement les problèmes suivants.

1/ Connaissant la fréquence théorique p des défectueux dans une population (ou plus généralement des unités qui possèdent un certain caractère) quelle est la probabilité P de trouver dans un échantillon de n observations un nombre k de défectueux inférieur ou égal à c .

2/ Inversement quelle est la valeur de c qui a une probabilité donnée d'être atteinte ou dépassée dans un tel échantillon.

3/ Ayant observé C défectueux dans un échantillon de n pièces, déterminer un intervalle de confiance pour la valeur de p dans le lot.

4/ Déterminer un plan d'échantillonnage par attributs (effectif n et critère limite d'acceptation C), satisfaisant à des risques donnés du producteur et de l'acheteur.

2/ Définition

Pour des raisons de commodité de lecture cet abaque comporte deux échelles fonctionnelles (Fig. 6. 1).

a) en abscisses une échelle logarithmique pour les valeurs de m = np allant de 0, 1 à 30 (au delà, l'approximation normale est parfaitement valable).

b) une échelle d'ordonnée graduée en utilisant, pour P, l'échelle des probabilités cumulées de la loi normale.

Dans le corps de l'abaque, on trouve un réseau de courbes C = Cte de (de 1 à 50).

Pour une valeur donnée de c, l'ordonnée du point d'abscisse m = np de la courbe C donne la valeur de la probabilité P de trouver un nombre k de défectueux, inférieur ou égal à c, dans un échantillon au hasard d'effectif n tiré d'un lot important caractérisé par une proportion p de défectueux.

Plus généralement l'abaque permet de déterminer l'un des trois paramètres P, np et c connaissant les deux autres.

Le pouvoir séparateur de l'échelle logarithmique étant constant en valeur relative, on peut sur un graphique de format 21 × 27, lire m = np à 1/100 près en valeur relative. De même le pouvoir séparateur de l'échelle normale sera d'autant plus grand que la probabilité d'acceptation sera plus voisine de zéro ou de un.

Moyennant les échelles graphiques choisies, les intersections des parallèles à l'axe des abscisses, avec les courbes de l'abaque se feront sous des angles permettant une lecture facile.

Fig. 6.1 - Loi de Poisson (Probabilités cumulées)

3/ Résolution graphique de quelques problèmes

3.1 - Probabilité de trouver $k \leqslant c$ défectueux dans un échantillon de n pièces provenant d'un lot contenant par hypothèse une fraction p de défectueux (la loi de Poisson étant supposée valable ($p < 0,10$, $n > 50$).

1/ Calculer $m = np$

2/ Lire sur l'axe des ordonnées la probabilité cherchée, P, ordonnée du point d'abscisse $m = np$ de la courbe c :

$$P = Pr\ [k \leqslant c \quad \text{si} \quad m = np]$$

on remarquera que l'on a :

$$1 - P = Pr\ [k > c \quad \text{si} \quad m = np]$$

Il est donc commode d'avoir deux axes d'ordonnées, l'un gradué en P, l'autre en $P' = 1 - P$.

3.2 - Exemple

3.2.1 - Un échantillon de $n = 80$ pièces a été prélevé dans une population contenant en moyenne 5 % de défectueux. Quelle est la probabilité de trouver :

a) un nombre de défectueux inférieur ou égal à 7,

b) un nombre de défectueux supérieur à 7 (donc supérieur ou égal à $7 + 1 = 8$).

On a :

$$m = np = 80 \times 0,05 = 4$$

On lit directement, à partir de l'intersection de la courbe $c = 7$ et de l'ordonnée $m = np = 4$:

$$\text{Probabilité } [k \leqslant 7 \quad \text{si} \quad np = 4] = 0,95$$

$$\text{Probabilité } [k > 7 \quad \text{si} \quad np = 4] = 0,05$$

3.2.2 - Test de l'hypothèse $p = 0,05$ à partir d'un échantillon de $n = 80$ pièces au risque bilatéral 0,05 de rejeter une hypothèse exacte.

Le nombre de défectueux que l'on observera et qui conduira à accepter ou à rejeter l'hypothèse étant un nombre entier, ce problème ne peut avoir qu'une solution approchée.

Les points d'ordonnées 0,025 et 0,975 et d'abscisse $m = np = 4$ donnent :

$$0 < c_1 < 1 \qquad c_2 \# 8$$

Si on accepte l'hypothèse $p = 0,05$ lorsque l'on trouve :

$$1 \leqslant k \leqslant 8$$

on aura en réalité :

$$Pr \quad (k < 1 \quad si \quad np = 4) \quad = Pr(k \leqslant 0) \quad = 0,02$$

$$Pr \quad (k > 8 \quad si \quad np = 4) \quad = 1 - Pr(k \leqslant 8) = 0,025$$

soit un risque d'erreur d'environ 0,045 de rejeter l'hypothèse alors qu'elle est vraie.

3.2.3 - Un échantillon de n = 200 pièces a donné 7 défectueux, déterminer l'intervalle de confiance bilatéral centré à 0,95 pour la proportion p de défectueux dans le lot.

Les limites cherchées p_1, p_2 sont définis par :

$$Pr(k \leqslant 7 \quad si \quad m = np_2) = 0,025$$

$$Pr(k \geqslant 7 \quad si \quad m = np_1) = 0,025$$

Cette dernière condition pouvant s'écrire :

$$Pr(k \leqslant 6 \quad si \quad m = np_1) = 0,975$$

Le point d'ordonnée P = 0,025 de la courbe c = 7 a pour abscisse np = 14,4 d'où :

$$p_2 = \frac{14,4}{200} = 0,072 \qquad (7,2\ \%)$$

Le point d'ordonnée P = 0,975 de la courbe c = 6 a pour abscisse np = 2,8 d'où :

$$p_1 = \frac{2,8}{200} = 0,014 \qquad (1,4\ \%)$$

d'où l'intervalle de confiance cherché :

$$(1,4\ \% \quad à \quad 7,2\ \%)$$

Une transformation de l'abaque précédent donne directement la solution (Fig. 6.2).

Pour obtenir l'intervalle de confiance à 0,95 (0,025 à 0,975), il suffira, avec cet abaque, de lire les abscisses des points d'ordonnée 7 situés sur les courbes correspondant aux seuils de confiance 0,025 et 0,975 pour obtenir np_1 = 14,5 et np_2 = 2,8.

3.2.4 - Détermination d'un plan d'échantillonnage par bons et mauvais, satisfaisant aux conditions :

$$P_a = \text{Probabilité d'acceptation} : P_a = \begin{cases} 1 - \alpha & si\ p = p_1 \\ \beta & si\ p = p_2 \end{cases}$$

Fig. 6.2 - Loi de Poisson. Intervalles de confiance à 0,15 et 0,998.

Exemple

On s'est fixé :

$p_1 = 0,01$ $\quad P_a = 0,90$ $\quad \alpha = 0,10$

$p_2 = 0,05$ $\quad P_a = 0,10$ $\quad \beta = 0,10$

Pour différentes valeurs de c = 0, 1, 2, ... on lira sur le graphique pour P = 0, 90 et P = 0, 10 les valeurs correspondantes de np_1 et np_2 d'où on déduira les valeurs correspondantes de n_1 et n_2.

$$n_1 = \frac{(np_1)}{p_1} \qquad n_2 = \frac{(np_2)}{p_2}$$

Le plan d'échantillonnage cherché est défini par la valeur de c pour laquelle on a :

$$n_1 = n_2 = n \qquad \text{(effectif de l'échantillon)}$$

Dans le cas particulier envisagé, on obtient les résultats suivants :

Pa	p	c = 0		c = 1		c = 2		c = 3	
0, 90	$p_1 = 0,01$	$np_1 = 0,11$	$n_1 = 11$	0, 56	56	1, 16	116	1, 8	180
0, 10	$p_2 = 0,05$	$np_2 = 2,3$	$n_2 = 46$	3, 95	79	5, 4	108	6, 6	132

n et c devant être entiers, le problème ne peut pas avoir de solution rigoureuse. On a seulement :

$$108 < n < 116$$

En première approximation on pourrait prendre :

$$c = 2 \qquad n = \frac{108 + 116}{2} = 112$$

ce qui donnerait :

$$np_1 = 1,12 \qquad np_2 = 5,6$$

$$P_{a_1} = 0,89 \qquad P_{a_2} = 0,08$$

On peut remarquer que la pente des courbes $c = c^{te}$ est d'autant plus grande que np est plus élevé, ce qui conduit à faire l'interpolation plus près de 108 que de 116.

Pour : $\quad c = 2$ et $n = 110 \qquad np_1 = 1,10 \quad np_2 = 5,5$

le graphique donne :

$$P_{a_1} = 0,89 \qquad P_{a_2} = 0,09$$

3.2.5 - *Courbe d'efficacité d'un plan d'échantillonnage par bons et mauvais (Loi de Poisson).*

Pour un plan donné (n, c) le graphique donne immédiatement les probabilités d'acceptation correspondant à une qualité p caractéristique du lot, définie sur l'abaque par m = np.

L'abaque constitue donc un résumé des courbes d'efficacité de ces plans. Ainsi, pour le plan défini par :

$$n = 100 \qquad c = 2$$

la probabilité d'acceptation si p = 0,03, soit np = 3 est P = 0,42.

Les valeurs p_1 et p_2 correspondant respectivement aux risques $\alpha = 0,05$ et $\beta = 0,10$ sont définies par :

$$P = 0,95 \qquad np = 0,82 \qquad p_1 = 0,008$$

$$P = 0,10 \qquad np = 5,3 \qquad p_2 = 0,053$$

Inversement, si on se fixe à l'avance les valeurs de p_1 et p_2, on aura si n = 100 et c = 2 :

pour : $p_1 = 0,01 \qquad np_1 = 1 \qquad P = 0,92 \qquad \alpha = 0,08$

$p_2 = 0,06 \qquad np_2 = 6 \qquad P = 0,06 \qquad \beta = 0,06$

3.3 - Abaque Cavé

3.3.1 - *Détermination de n et de c pour un plan défini par* (α, p_1 β, p_2).

M. Cavé, Ingénieur en Chef des fabrications d'Armement a réalisé un abaque qui donne directement la solution du problème précédent en fonction du rapport p_2/p_1 et des risques α et β pour un certain nombre de combinaisons fréquemment utilisées de α et β.

$\alpha = \beta = 0,001$ et $0,01$

$\alpha = 0,05 \qquad \beta = 0,05$ et $0,10$

$\alpha = \beta = 0,10$

A ch cune de ces combinaisons correspond une courbe $c(\alpha, \beta)$, dont l'ordonnée du point M d'abscisse p_2/p_1 donne la valeur critique c (Fig. 6.3).

Les valeurs de np_1 et np_2 sont alors les abscisses des points N_1 et N_2 d'ordonnées c des deux courbes c_α et c_β correspondant respectivement aux valeurs α et β fixées. On en déduit ensuite :

$$n = \frac{np_1}{p_1} = \frac{np_2}{p_2}$$

Dans le cas où l'ordonnée de M n'est pas un nombre entier, on prendra pour c la valeur entière la plus voisine.

M. Cavé propose, dans ce cas, de prendre pour n la valeur n' définie en fonction des abscisses m et m des deux points N_1 et N_2 par :

$$n' = \frac{m_1 + m_2}{p_1 + p_2}$$

3.3.2 - *Courbe d'efficacité.*

Compte tenu des courbes de l'abaque correspondant aux diverses valeurs de β et de $1 - \alpha$, c'est-à-dire à diverses valeurs de la probabilité P d'acceptation, l'horizontale d'ordonnée c coupera ces courbes en des points N_i, situés sur les courbes $P = P_i$ et d'abscisses respectives np_i définies par :

$$P_i = \text{Probabilité d'acceptation si } p = p_i = \frac{(np_i)}{n}$$

Exemple : Pour $n = 105$ $c = 5$

$P = 0,30$ = Probabilité d'acceptation si $105\ p = 7$ soit :

$$p = \frac{7}{105} = 6,67\ \%$$

L'abaque permet donc de tracer par points la courbe d'efficacité du plan d'échantillonnage.

3.3.3 - *Intervalle de confiance.*

L'abaque permet aussi de déterminer - pour les valeurs α et β qui y figurent - l'intervalle de confiance à $1 - (\alpha + \beta)$ pour la proportion p de défectueux dans un lot dont on a examiné un échantillon de n pièces.

Exemple : Un échantillon de 200 pièces a donné 7 défectueux soit $f = 3,5\ \%$. Déterminer l'intervalle de confiance à 0,95 pour P, soit :

$$\alpha_1 = 0,05 \qquad \alpha_2 = 0,05$$

On doit donc avoir :

$$Pr(k \leqslant 7 \quad \text{si} \quad m = np_2) = 0,05$$

$$Pr(k \leqslant 6 \quad \text{si} \quad m = np_1) = 0,95$$

Pour $c = 7$ et $P = 0,05$, l'abaque donne $np = 13,2$ soit :

$$p_2 = \frac{13,2}{200} = 6,6\ \%$$

Pour $c = 6$ et $P = 0,95$, l'abaque donne $np = 3,3$ soit :

$$p_1 = \frac{3,3}{200} = 1,65\ \%$$

d'où l'intervalle de confiance cherché :

$$1,65\ \% < p < 6,6\ \%$$

4/ Annexe Théorique

4.1 - Loi de Poisson

La loi de Poisson, de moyenne m = np, est définie pour une variable k par la relation :

$$Pr(k \leqslant c) = \sum_{k=0}^{c} e \quad \frac{(np)^k}{k\,!} = P \qquad (1)$$

qui peut encore s'écrire :

$$pr(k > c) = \sum_{k=c+1}^{\infty} e^{-np} \frac{(np)}{k\,!} = P' = 1 - P$$

Pour toute valeur de m, c'est-à-dire pour tout couple de valeurs de n et de p, la relation (1) permet, à l'aide des tables de la loi de Poisson, de construire par points les courbes correspondant à toute valeur entière de c.

L'abaque permet donc de résoudre l'équation (1) par rapport à l'une quelconque des trois variables P, c et np, lorsque deux d'entre elles sont fixées, compte tenu du fait que c et n ne peuvent prendre que des valeurs entières.

Pour c et np donnés la solution en P de l'équation (1) est donnée immédiatement par l'abaque (Exemple 3.2.1).

4.1.1 - *Test de l'hypothèse* m = np *à un risque* α *donné de rejeter l'hypothèse alors qu'elle est vraie.*

L'hypothèse d'un lot caractérisé par une proportion p de défectueux - c'est-à-dire d'un lot pour lequel un échantillon d'effectif n donnerait en moyenne np défectueux - sera acceptée si le nombre k de défectueux donné par l'échantillon satisfait à la condition :

$$c_1 < k < c_2$$

c_1 et c_2 étant définis par :

$$Pr(k \leqslant c_1 \quad \text{si} \quad np) = \sum_{k=0}^{c_1} e^{-np} \frac{(np)^k}{k\,!} = \alpha_1$$

$$Pr(k \geqslant c_2 \quad \text{si} \quad np) = 1 - \sum_{k=0}^{c_2-1} e^{-np} \frac{(np)^k}{k\,!} = \alpha_2$$

avec $\alpha = \alpha_1 + \alpha_2$.

En fait, c_1 et c_2 devant avoir des valeurs entières, on cherchera les courbes (c_1', c_1') et (c_2', c_2'') qui encadrent les points d'abscisse np et d'ordonnées respectives $P_1 = \alpha$ et $P_2 = 1 - \alpha$.

Ayant choisi l'une des solutions approchées c_1, c_2 on pourra lire le risque réel α correspondant à cette solution (cf. Ex. 3.2.2).

4.1.2 - *Estimation de p avec intervalle de confiance à*
$$1 - \alpha = 1 - \alpha_1 - \alpha_2$$
à partir d'un échantillon de n pièces ayant donné k = c *défectueux.*

Les limites cherchées p_1 et p_2 sont définies par :

$$Pr(k \leqslant c \quad \text{si} \quad m = np_2) = \sum_{k=0}^{c} e^{-np_2} \frac{(np_2)}{k!} = \alpha_2 = P_2$$

$$Pr(k \geqslant c \quad \text{si} \quad m = np_1) = \sum_{k=c}^{\infty} e^{-np_1} \frac{(np_1)}{k!} = \alpha_1$$

cette seconde condition étant équivalente à :

$$Pr(k \leqslant c - 1 \quad \text{si} \quad m = np_1) = 1 - \alpha_1 = P_1$$

np_2 et np_1 sont les abscisses des points des courbes c et c - 1 d'ordonnées respectives P_2 et P_1.

La transformation d'abaque, déjà signalée en 3.2.3 peut évidemment être généralisée pour des valeurs quelconques de α_1 et α_2.

Remarque importante

Dans tous les problèmes relatifs à des variables discrètes, il importe de faire attention à la différence entre $k \leqslant c$ et $k < c$, cette dernière condition pouvant aussi s'écrire $k \leqslant c - 1$.

4.2 - Abaque Cavé (Fig. 6.4).

Les équations à résoudre pour définir le plan (n, c) correspondant aux probabilités d'acceptation :

$$P_a = \begin{cases} 1 - \alpha & \text{si} \quad p = p_1 \\ \beta & \phantom{\text{si}} \quad p = p_2 \end{cases}$$

sont les suivantes :

$$\sum_{k=0}^{c} e^{-m_1} \frac{(m_1)^k}{k!} = 1 - \alpha \qquad \text{avec} \qquad m_1 = np_1$$

$$\sum_{k=0}^{c} e^{-m_2} \frac{(m_2)^k}{k!} = \beta \qquad m_2 = np_2$$

Pour α et β donnés, ainsi que p_1 et p_2, chacune de ces équations définit c comme une fonction implicite de m (m_1 ou m_2), ces deux variables étant liées par la condition :

$$\frac{m_2}{m_1} = \frac{p_2}{p_1}$$

soit $\log m_2 - \log m_1 = \log p_2 - \log p_1$.

On peut donc tracer les courbes des variations de c en fonction de m (m_1 et m_2) en utilisant la transformation :

Fig. 6.4 - Emploi de l'abaque Cavé.

$p_1 = 0,01$ $\quad \alpha = 0,10$ $\quad \frac{p_2}{p_1} = 5$ $\quad c = 2$ $\quad (np_1) = 1,1$

$p_2 = 0,05$ $\quad \beta = 0,10$ $\quad (np_2) = 5,4$

$$n = \frac{1,1 + 5,4}{0,01 + 0,05} = \frac{6,5}{0,06} = 109$$

$$\sum_{k=0}^{c} e^{-m} \frac{m^k}{k!} = \frac{\int_{\sqrt{2m}}^{\infty} e^{-\frac{\chi^2}{2}} \chi^{2c+1} d\chi}{\int_0^{\infty} e^{-\frac{\chi^2}{2}} \chi^{2c+1} d\chi}$$

$$= 1 - Pr(\chi^2 < 2m), \quad \nu = 2(c + 1)$$

degrés de liberté, qui permettra de considérer c comme une variable continue.

Pour α et β donnés, on obtiendra deux courbes :

$$c_\alpha = f_\alpha(m) \qquad c_\beta = f_\beta(m)$$

Les points de ces courbes correspondant à la solution du système devront donc avoir pour coordonnées :

$$c_\alpha = c \qquad c_\beta = c$$

$$m = m_1 = np_1 \qquad m = m_2 = np_2$$

avec :

$$\log m_2 - \log m_1 = \log \frac{p_2}{p_1}$$

Si on a utilisé, pour m, une échelle logarithmique, il suffira de déterminer sur ces courbes, les deux points qui, pour une même ordonnée, ont des abscisses dont la différence est égale à $\log \frac{p_2}{p_1}$

Pour tout couple (α, β), ce même abaque établit une correspondance entre c et p_2/p_1 que l'on représente par la courbe :

$$c_{\alpha\beta} = g\left(\frac{p_2}{p_1}\right)$$

en utilisant pour $\frac{p_2}{p_1}$ une échelle logarithmique de même module que celle utilisée pour m. Un décalage convenable de cette échelle permet d'éviter la superposition des deux réseaux (c_α, c_β) et $c_{\alpha\beta}$.

On notera que, pour un couple donné α, β, c ne dépend que du rapport p_2/p_1 et non des valeurs respectives de p_1 et p_2.

M. Cavé donne dans son ouvrage (1) les valeurs de p_2/p_1 en fonction de c pour divers couples de valeurs de α et β.

---

(1) R. Cavé - Le Contrôle statistique des fabrications. Eyrolles (2e édition - 1961).