REVUE DE STATISTIQUE APPLIQUÉE

G. ROUZET

Étude graphique de l'hypothèse d'une loi exponentielle pour la durée de vie de dispositifs

Revue de statistique appliquée, tome 17, nº 4 (1969), p. 77-82

<http://www.numdam.org/item?id=RSA_1969__17_4_77_0>

ÉTUDE GRAPHIQUE DE L'HYPOTHÈSE D'UNE LOI EXPONENTIELLE POUR LA DURÉE DE VIE DE DISPOSITIFS

G. ROUZET

Chef du Contrôle général qualité de la Compagnie des Compteurs

A la suite d'essais effectués sur des dispositifs issus d'une même "population", on dispose d'un échantillon de n valeurs de durée de vie.

On désire étudier graphiquement la possibilité d'admettre, sur cette base, l'hypothèse d'une loi exponentielle de la durée de vie, dans la population.

Plusieurs méthodes peuvent être utilisées dans ce but ; en particulier, nous avons donné précédemment(1) la définition d'échelles fonctionnelles selon lesquelles les points représentatifs de l'échantillon tendent à s'aligner sur une droite si l'hypothèse est vraie, chacun d'eux devant, de plus, être compris dans l'intervalle entre deux limites en probabilité.

Dans la méthode exposée ci-après, les n points de l'échantillon se rapportent à n variables indépendantes, de sorte que des limites en probabilité peuvent être définies pour l'ensemble du nuage, ce qui facilite l'interprétation, puisque celle-ci réside dès lors en un test global unique.

Dans la mesure où elles n'ont pas été obtenues directement dans cet ordre, classons les n valeurs observées, dans l'ordre croissant. Soit x_i la valeur de l'observation de rang i dans ce classement.

On remarque que x_i s'identifie à l'instant de la $i^{ème}$ défaillance, dans le schéma - réel ou fictif - correspondant à la mise en essai simultanée des n dispositifs à l'instant $x_o = 0$.

Faisons l'hypothèse que la loi de probabilité de la durée de vie est exponentielle, de moyenne m : soit $\lambda = 1/m$ le taux de défaillance instantané.

Considérons :

- les variables aléatoires X_i, instants des n défaillances dans un échantillon aléatoire répondant au même schéma,

- les écarts $\Delta_i = X_i - X_{i-1}$.

----

(1) "Incidence pratique de la méthode utilisée pour l'étude graphique de la loi de probabilité d'une variable aléatoire à partir d'un échantillon d'observations indépendantes" - Revue de Statistique Appliquée - Vol. XIV, N° 3.

Il découle des propriétés de la loi exponentielle, que les variables aléatoires Δ_i sont indépendantes, et que leurs lois de probabilité sont exponentielles, d'expression :

$$f(\delta_i) = (n + 1 - i) \lambda \; e^{-(n+1-i)\lambda\delta_i} \tag{1}$$

Nous n'établirons pas ce résultat classique, qui se retrouve aisément en remarquant que $(n + 1 - i)$ dispositifs restent en essai pendant l'intervalle (x_{i-1}, x_i).

Posons alors :

$$Q_i = (n + 1 - i) \lambda \Delta_i = \frac{(n + 1 - i) \Delta_i}{m} \tag{2}$$

La loi de probabilité de Q_i a pour expression, quel que soit i :

$$g(q_i) = e^{-q_i} \tag{3}$$

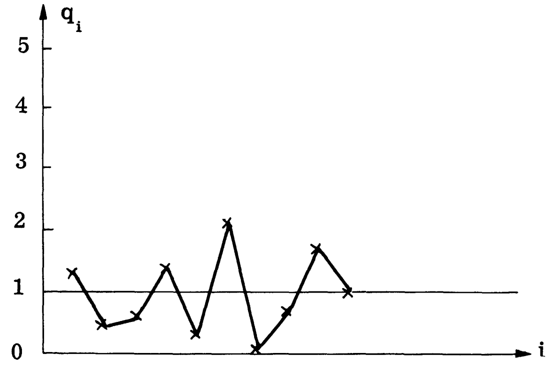
En conclusion, les variables aléatoires Q_i sont indépendantes, et chacune d'elles admet pour loi de probabilité la loi exponentielle de moyenne égale à l'unité. On a donc :

$$E(Q_i) = 1 \tag{4}$$

de sorte que, si l'hypothèse est vraie, les points de l'échantillon définis par :

$$q_i = \frac{(n + 1 - i) (x_i - x_{i-1})}{m} \tag{5}$$

doivent tendre à se répartir - selon les fluctuations d'échantillonnage - autour de la droite moyenne d'ordonnée unité, sur le graphique de coordonnées $(i \; ; \; q_i)$.



Par ailleurs, les limites en probabilité peuvent être facilement définies à partir de la fonction de répartition Q_i :

$$F(q_i) = 1 - e^{-q_i} \tag{6}$$

Toutefois, ainsi que nous l'avons annoncé initialement, c'est à l'ensemble des points que nous ferons correspondre un intervalle en probabilité.

L'intervalle bilatéral à $(1 - \alpha)$ est borné par les deux limites $L_{\alpha/2}$ et $L_{1-\alpha/2}$ que nous définirons en écrivant que :

1/ La probabilité que l'ensemble des points soit dans l'intervalle est égale à $(1 - \alpha)$, soit :

$$[(1 - e^{-L_{1-\alpha/2}}) - (1 - e^{-L_{\alpha/2}})]^n = 1 - \alpha$$

d'où :

$$e^{-L_{\alpha/2}} - e^{-L_{1-\alpha/2}} = (1 - \alpha)^{1/n} \tag{7}$$

2/ La probabilité de franchir la limite supérieure est égale à la probabilité de franchir la limite inférieure, soit :

$$(e^{-L_{1-\alpha/2}})^n = (1 - e^{-L_{\alpha/2}})^n$$

d'où

$$e^{-L_{\alpha/2}} + e^{-L_{1-\alpha/2}} = 1 \tag{8}$$

Le système formé par (7) et (8) admet pour solution les valeurs suivantes des limites de l'intervalle :

$$\left.\begin{aligned} L_{\alpha/2} &= - \ln \frac{1}{2} [1 + (1 - \alpha)^{1/n}] \\ L_{1-\alpha/2} &= - \ln \frac{1}{2} [1 - (1 - \alpha)^{1/n}] \end{aligned}\right\} \tag{9}$$

Le calcul numérique de ces limites, pour $\alpha = 5$ %, donne, pour quelques valeurs de n, les résultats suivants :

n	$L_{\alpha/2}$	$L_{1-\alpha/2}$
40	0,00064	7,35
50	0,00051	7,58
60	0,00043	7,75
100	0,00026	8,27

Comme on pouvait s'y attendre, les valeurs de $L_{\alpha/2}$ sont ridiculement faibles, de sorte que cet intervalle bilatéral offre peu d'intérêt.

On considèrera donc plutôt l'intervalle unilatéral à $(1 - \alpha)$ défini par $(0 ; L_{1-\alpha})$ avec :

$$L_{1-\alpha} = - \ln [1 - (1 - \alpha)^{1/n}] \tag{10}$$

Pour 5 %, on trouve :

n	$L_{1-\alpha}$
40	6,66
50	6,88
60	7,06
100	7,57

valeurs qui peuvent être obtenues par la relation approchée :

$$L_{0,95} \simeq 2,97 + \ln n \tag{11}$$

qui correspond à la forme limite de l'expression (10) pour n croissant, pour $\alpha = 5\ \%$

Lorsque l'hypothèse ne comporte pas de valeur numérique pour m, c'est-à-dire qu'elle est relative à la "forme" exponentielle, on pourra, comme on le fait usuellement dans les problèmes analogues, adopter pour m la valeur de la moyenne $\bar{x}$ de l'échantillon :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{i=n} x_i \tag{12}$$

ou encore :

$$x = \frac{1}{n} \sum_{i=1}^{i=n} (n + 1 - i)\ (x_i - x_{i-1}) \tag{13}$$

et considérer les quantités :

$$q'_i = \frac{(n + 1 - i)\ (x_i - x_{i-1})}{\bar{x}} \tag{14}$$

En toute rigueur, les quantités q'_i ne se comportent pas comme les quantités q_i, dans un échantillon aléatoire. Mais, si n est suffisamment grand, la méthode précédente pourra utiliser les quantités q'_i avec une approximation satisfaisante. On remarquera que, de toutes manières, l'étude graphique ne permet guère de tirer des conclusions valables que si l'échantillon n'est pas trop petit, disons si $n > 30$. Le comportement de q'_i est alors assimilable à celui de q_i, et les résultats précédents restent pratiquement valables.

RESUME

A partir de n résultats, classés dans l'ordre croissant, soit : x_1, $x_2 \ldots x_n$, on calculera les quantités :

$$q_i = \frac{(n + 1 - i)\ (x_i - x_{i-1})}{m}$$

avec $x_0 = 0$. Si l'hypothèse ne comporte pas de valeur pour la moyenne m de la loi exponentielle, on prendra pour m la valeur de :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{i=n} x_i$$

que l'on pourra vérifier par :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{i=n} (n + 1 - i)\ (x_i - x_{i-1})$$

Les quantités q_i seront portées sur un graphique de coordonnées $(i\ ;\ q_i)$.

Si l'hypothèse est vraie, les points doivent admettre pour droite moyenne la droite d'ordonnée égale à l'unité.

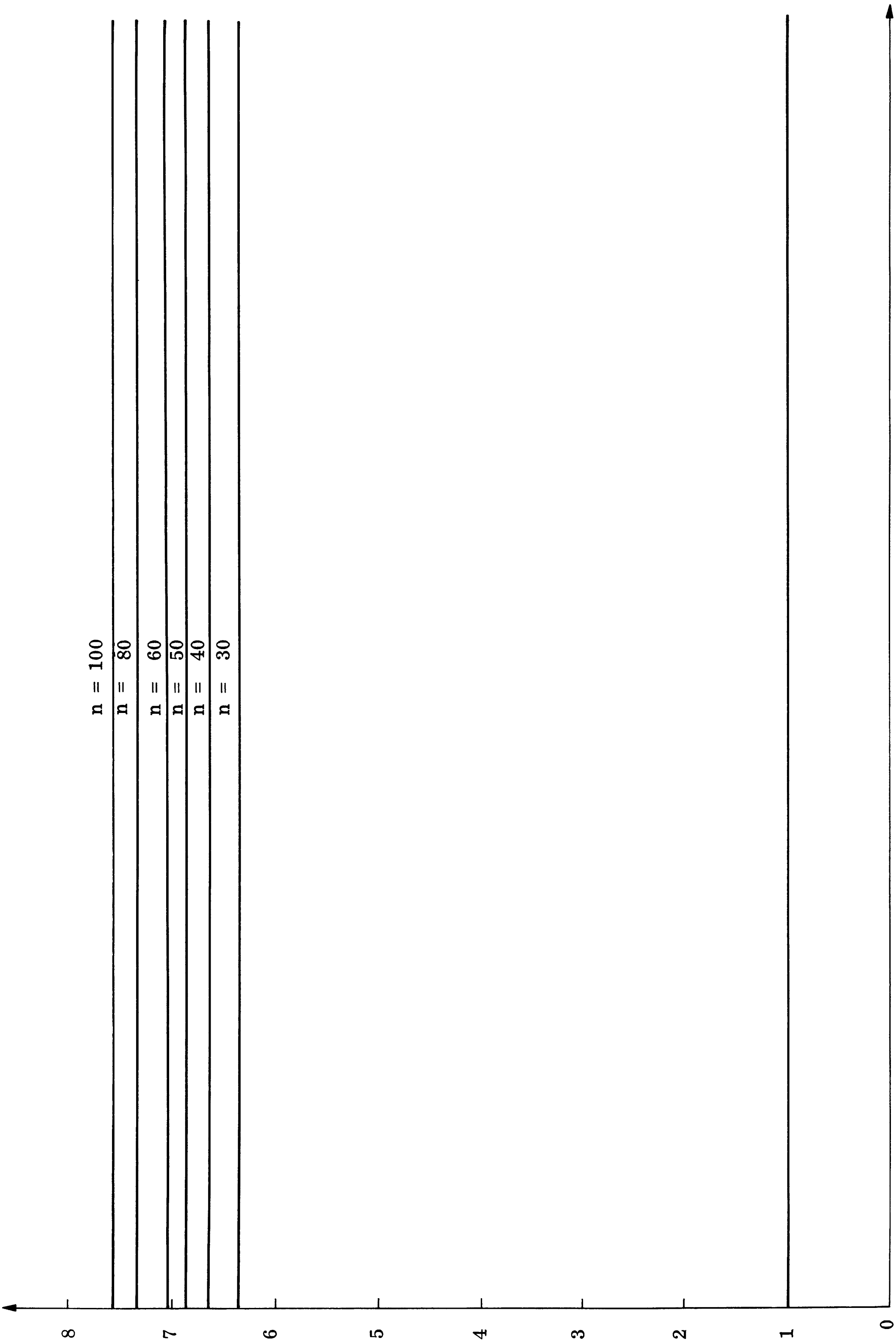
De plus, l'ensemble de ces points doit respecter la limite en probabilité à $(1 - \alpha)$ définie par :

$$L_{1-\alpha} = - \ln\ [1 - (1 - \alpha)^{1/n}]$$

On trouvera ci-après un graphique sur lequel figurent, pour quelques valeurs de n, les limites à $1 - \alpha = 95\ \%$; ce graphique est construit avec les valeurs suivantes :

n	$L_{0,95}$
30	6,37
40	6,66
50	6,88
60	7,06
80	7,35
100	7,57

$$L_{0,95} \simeq 2,97 + \ln n$$


n = 100
n = 80
n = 60
n = 50
n = 40
n = 30
0
1
2
3
4
5
6
7
8