Journal de la société statistique de Paris

E. Morice

Transformation doublement logarithmique de pourcentages cumulés

Journal de la société statistique de Paris, tome 109 (1968), p. 266-276
<http://www.numdam.org/item?id=JSFS_1968__109__266_0>

TRANSFORMATION DOUBLEMENT LOGARITHMIQUE DE POURCENTAGES CUMULÉS

Ce problème étudié par notre collègue J. Dufrénoy dans un précédent numéro (J.S.S.P., 1967, n^{os} 10-11-12) est un cas particulier d'un problème largement étudié au cours des dix dernières années et pour lequel on dispose de divers papiers à échelles fonctionnelles et de tables.

Si dans une population $\mathscr{P}$ (d'individus ou d'éléments), P (t) est la fréquence cumulée des individus dont la durée de vie T est supérieure à une valeur t, Dufrenoy envisage le cas où ln [— ln P (t)] est une fonction linéaire de t :

$$\ln [- \ln P(t)] = at + b$$

d'où l'on déduit, dans le cas le plus général :

$$Pr(T > t) = P(t) = \exp(- e^{at+b}), \qquad (1)$$
$$- \infty < t < + \infty$$

(a étant une constante positive : P (t) tend vers zéro lorsque t augmente indéfiniment).

Les paramètres a et b de la loi définie par (1), calculés par la méthode des moments, peuvent être estimés par les formules

$$\hat{a} = \frac{\pi}{s\sqrt{6}}, \qquad \hat{b} = -(0,5772 \times \widehat{at})$$

t et s étant respectivement la moyenne et l'écart-type de la variable t, estimés à partir des observations, et 0,5772, la constante d'Euler.

1° *Loi de Gompertz*

On peut remarquer que la loi de survie définie par (1) n'est autre, à un coefficient constant près, que la loi proposée par Gompertz en 1825, pour décrire une population humaine, en faisant l'hypothèse que la mort est due à l'affaiblissement progressif de l'individu, c'est-à-dire à ce qu'il nommait « force of mortality », φ, qui s'accroît dans le temps dt d'une quantité $d\varphi$ proportionnelle à φ.

Ceci définit un taux instantané de mortalité φ, satisfaisant à l'équation

$$\frac{1}{\varphi}\frac{d\varphi}{dt} = q = C^{te}$$

d'où

$$\varphi = a\, e^{qt}$$

Si V_t est le nombre des survivants à l'âge t, l'hypothèse de Gompertz conduit à un quotient moyen de mortalité à l'âge t, par unité de temps, défini par :

$$\frac{1}{dt}\frac{V_t - V_{t+dt}}{V_t},$$

d'où pour dt tendant vers zéro :

$$-\frac{V'_t}{V_t} = a\, e^{qt}$$

$$V_t = k \exp\left[-\frac{a}{q}\, e^{qt}\right]$$

soit, en utilisant les notations de Gompertz :

$$V_t = k\, g^{c^t} \tag{2}$$

avec

$$g = e^{-\frac{a}{q}}, \qquad c = e^q$$

k, g c étant trois constantes définissant le nombre de vivants à l'origine et l'allure de la courbe de survie.

La loi de Gompertz, considérée comme loi de survie d'une génération humaine, est caractérisée par un taux instantané de mortalité constamment croissant (si $q > 0$), elle ne peut évidemment s'appliquer approximativement qu'au-delà d'un certain âge t_0, à partir duquel les quotients annuels de mortalité observés augmentent avec l'âge [1].

Les divers paramètres de la loi sont liés par la condition initiale

$$V/t = t_0) = V_0 = k \exp\left[-\frac{a}{q}\, e^{qt_0}\right]$$

d'où

$$V(t) = V_0 \exp\left[-\frac{a}{q}\,(e^{qt} - e^{qt_0})\right] \tag{3}$$

soit :

$$P(t) = Pr\,[T > t] = \frac{V}{V_0} = \exp\left[\frac{a}{q}\, e^{qt_0}\right] \cdot \exp\left[-\frac{a}{q}\, e^{qt}\right], \qquad t > t_0$$

2° *Loi de Gumbel* [2]

Plus généralement, dans d'autres domaines, on peut envisager la loi de survie définie par

$$Pr\,[T > t] = P(t) = \exp\left[-\frac{a}{q}\, e^{qt}\right] \tag{4}$$

$$-\infty < t < +\infty, \quad a > 0 \quad q > 0$$

pour laquelle on aura :

$$\ln\,[-\ln P(t)] = qt + \ln\frac{a}{q}$$

1. Il est de même de la loi de Makeham, qui reprenant une hypothèse déjà aussi envisagée par Gompertz, considère un taux instantané de mortalité défini par

$$-\frac{V'}{V} = a\, e^{qt} + b,$$

la constante positive b représentant l'effet du hasard indépendant de l'âge.
On obtient alors la loi de Makeham, généralement presentée sous la forme :

$$V(t) = k\, s^x\, g^{c^x} \quad \text{avec } b = -\ln s$$

2. Rappelons que Gumbel, Professeur à Columbia University, décédé en 1966, avait, avant la guerre enseigné à la Faculté des Sciences de Lyon.

On retrouve ainsi la loi étudiée par Gumbel dans des études de résistance des métaux à la fatigue et présentée par cet auteur avec une fonction de répartition de la forme :

$$(G_1) \cdot F(t) = Pr[T < t] = 1 - P(t) = 1 - \exp[- e^{\alpha(t-u)}], \tag{5}$$
$$- \infty < t < + \infty \qquad \alpha > 0$$
$$F(- \infty) = 0 \quad F(+ \infty) = 1$$

pour laquelle on a

$$\ln[- \ln P(t)] = \alpha t - \alpha u,$$

fonction linéaire croissante de t.

Une table de la fonction

$$y = \ln[- \ln P]$$

a été calculée en 1953 par le « National Bureau of Standards » [3]. Elle peut être utilisée pour construire une échelle graphique graduée en P (t), (ou en F (t)), où les longueurs sont proportionnelles aux valeurs de la variable réduite :

$$y = \ln[- \ln P(t)] = \ln[- \ln(1 - F(t))]$$

Si t est porté en abscisses sur une échelle uniforme, les points $(t, P(t))$ correspondant aux observations devront se placer autour de la droite :

$$y = \alpha(t - u),$$

si le modèle (4) correspond au phénomène étudié (fig. 1).

La valeur modale de t (valeur correspondant au maximum de la fonction de densité)

$$f(t) = \frac{d[1 - P(t)]}{dt}$$

est égale à la valeur caractéristique u. Les deux paramètres u et $1/\alpha$ peuvent être estimés à partir des formules

$$\hat{u} = \bar{t} + \frac{0{,}5772}{\hat{\alpha}}, \hat{\alpha} = \frac{\pi}{s\sqrt{6}}$$

où C = 0,5772 est la constante d'Euler, t et s étant respectivement la moyenne et l'écart-type estimés à partir des observations.

On notera que le changement de variable

$$\alpha(t - u) = - \beta(x - \nu),$$

α et β positifs, donne pour la variable X, la fonction de répartition :

$$(G_2) \quad F(x) = Pr[X < x] = \exp[- e^{-(\beta x - \nu)}]$$
$$- \infty < x < + \infty$$
$$F(- \infty) = 0, \quad F(+ \infty) = 1$$

avec cette fois :

$$\ln[- \ln F(x)] = - \beta(x - \nu), \tag{fig. 1}$$

fonction linéaire décroissante de x, les paramètres β et v pouvant alors être estimés par :

$$\hat{\nu} = \bar{x} - \frac{0{,}5772}{\hat{\beta}}$$
$$1/\hat{\beta} = \frac{s\sqrt{6}}{\pi},$$

$\bar{x}$ et s étant respectivement la moyenne et l'écart-type de la variable x, estimés à partir des observations.

3° *Loi de Weibull*

Elle est définie par

(W). $$P(t) = Pr\,[T > t] = \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right], \quad t > 0 \tag{6}$$

β et η étant deux constantes positives :

β : paramètre de forme

η : paramètre d'échelle (vie caractéristique)

On a alors

$$\ln[-\ln P(t)] = \beta \ln t - \beta \ln \eta \tag{7}$$

La loi de Weibull est identique à la loi $P(t) = \exp[-e^y]$, moyennant le changement de variable :

$$e^y = \left(\frac{t}{\eta}\right)^{\beta} \tag{8}$$

$$y = \beta \ln t - \beta \ln \eta$$

Moyennant l'emploi d'un graphique à échelles fonctionnelles :

a) dont l'axe des ordonnées est gradué suivant les valeurs de $P(t)$, avec une échelle uniforme en $\ln[-\ln P(t)]$

b) dont l'axe des abscisses est gradué suivant les valeurs de t, avec une échelle uniforme en $\ln t$.

Les points correspondant aux valeurs observées $[t, P^*(t)]$, devront se placer au voisinage de la droite définie par l'équation (7) (*cf.* graphique ci-joint).

La droite que, dans un cas particulier, on peut éventuellement ajuster aux points observations, permet une estimation facile des paramètres β et η :

η est la valeur de t correspondant à $\ln[-\ln P(t)] = 0$, c'est-à-dire à

$$p^*(t) = t = \frac{1}{e} = 0{,}368$$

β est le coefficient angulaire de la droite d'ajustement, calculé compte tenu des modules des échelles employées.

Les valeurs typiques de la distribution théorique de la variable t sont :

Moyenne $m = E(T) = \eta\,\Gamma\left[1 + \frac{1}{\beta}\right]$

Mode $\hat{M} = \quad = \eta\left(1 - \frac{1}{\beta}\right)^{1/\beta}$, si $\beta > 1$

Médiane $M = \quad = \eta\,[\ln 2]^{1/\beta}$

Variance $\sigma^2 \doteq V(T) = \eta^2\left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right)\right]$

L'emploi du graphique de Weibull est particulièrement commode pour étudier les divers problèmes ci-dessus.

On trouve en effet dans le commerce des papiers pour graphiques à échelles fonctionnelles de Weibull [noter que dans ces papiers l'axe des ordonnées est généralement gradué non en $P(t)$, mais en $F(t) = 1 - P(t)$ c'est-à-dire en $F(t) = Pr[T < t]$], [4] et [5].

Si $t_2 < t_1 < \dots < t_n$ sont les n valeurs ordonnées des durées de vie correspondant aux n éléments d'un échantillon prélevé au hasard dans la population étudiée, la fréquence cumulée jusqu'à et y compris la valeur $t = t_i$ est une variable aléatoire, variable d'un échantillon à un autre.

Plutôt que de prendre i/n comme estimation de $F(t_i)$ dans l'ensemble continu des valeurs de t, il conviendra de prendre comme estimation $F^*(t_i)$

— soit son espérance mathématique $F^*(t_i) = \dfrac{i}{n+1}$

— soit sa valeur médiane $F^*(t_i) = \dfrac{i\,F_{0.50}}{(n-i+1) + i\,F_{0.50}}$,

$F_{0,50}$ étant le fractile d'ordre 0,50 de la loi de F pour $\nu_1 = 2i$, $\nu_2 = 2(n-i+1)$.

Deux types de papiers de Weibull couramment utilisés sont les suivants :

1° Compagnie Française des Diagrammes — Bd Inkermann à Neuilly : 1 modèle de papier avec :

— échelle des ordonnées $F(t) = 0{,}1\ \% \dots$ à $99{,}9\ \%$

— échelle des abscisses à 3 modules soit $t = 10^n \dots$ à 10^{n+3}.

Une échelle graduée en β donne immédiatement la valeur de β après tracé d'une parallèle à la droite d'ajustement. Un tableau joint au cahier de graphiques donne les valeurs des coefficients numériques $\Gamma(1 + 1/\beta)$ et $[\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta)]^{1/2}$ en fonction de β et permet de calculer la moyenne m et l'écart-type σ.

2° Technical and Engineering Aids for Management, 104 Belrose Avenue, Lowell, Massachusetts (U.S.A.) a édité 9 modèles de papiers pour graphiques de Weibull

— 3 échelles de probabilité : 0,0001 — 0,01 — 1,0 à 99,9 % avec pour chaque type des échelles logarithmiques à 3, 5 ou 7 modules.

Moyennant quelques tâtonnements rapides, ces mêmes graphiques permettent aussi d'étudier l'ajustement éventuel d'une loi de la forme :

$$F(t) = 1 - P(t) = 1 - \exp\left[-\left(\frac{t-\gamma}{\eta}\right)^\beta\right], \quad t > \gamma$$

en étudiant la loi de la variable $t' = t - \gamma$, pour des valeurs essayées du paramètre inconnu γ [fig. (2)].

Notons aussi qu'il existe des tables [6] et [7] des fonctions

$$F(x) = 1 - \exp(-x^\beta)$$

$$f(x) = \frac{dF(x)}{dx} = \beta\, x^{\beta-1} \exp(-x^\beta)$$

donnant en fonction de x et β, les valeurs de la fonction de répartition $F(x)$ et de la densité de probabilité $f(x)$ relatives à la loi de Weibull pour $\eta = 1$, $\gamma = 0$.

Remarque. — La loi de Weibull (6) correspond à la loi de Gumbel (5), moyennant la transformation définie par :

$$e^y = \left(\frac{t}{\eta}\right)^\beta$$

$$y = \beta\,(\ln t - \ln \eta)$$

ou en logarithmes décimaux

$$y = \beta'\,(\log t - \log \eta)$$

avec $\beta' = 2{,}3026\,\beta$.

Gumbel [2] donne comme estimation des paramètres η et β'

$$\log \hat{\eta} = \overline{\log t} + \frac{0{,}5772}{\hat{\beta}'},$$

$$\frac{1}{\hat{\beta}'} = \frac{\sqrt{6}}{\Pi}\, s\,(\log t),$$

$\overline{\log t}$ et $s\,(\log t)$ étant respectivement la moyenne et l'écart-type empiriques des logarithmes estimés à partir de l'échantillon observé.

Loi de Fréchet

Elle peut être présentée sous la forme

$$F(x) = Pr\,(X < x) = \exp\left[-\left(\frac{x-a}{b}\right)^{-k}\right], \qquad x > a$$

k et b étant deux paramètres positifs

$$\ln\,[-\ln F(x)] = -\,k \ln\,(x-a) + k \ln,\, b,$$

fonction linéaire décroissante de la variable auxiliaire ln (x-a), alors que pour la loi de Weibull définie par

$$F(t) = Pr\,[T < t) = 1 - \exp\left[-\left(\frac{t-\gamma}{\eta}\right)^\beta\right]$$

on avait

$$\ln\,[-\ln\,[1 - F(t)]] = \beta \ln\,(t-\gamma) - \beta \ln \eta$$

fonction croissante de $\ln\,(t-\gamma)$.

Le papier classique loi de Weibull, pourra encore être utilisé pour étudier la loi de Fréchet (fig. 2), à condition de considérer que l' échelle d'ordonnées est graduée en fréquences cumulées « au-delà », soit $P^*(x) = 1 - F^*(x)$, l'échelle fonctionnelle étant alors une échelle en $\ln\,[-\ln\,[1 - P^*(x)]]$

On notera que le changement de variable défini par

$$\left(\frac{x-a}{b}\right)^{-k} = \left(\frac{t-\gamma}{\eta}\right)^\beta, \qquad t > \gamma$$

β et η positifs, donne précisément la loi de Weibull

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

Les valeurs typiques de la loi de Fréchet sont données par :

Moyenne $m = E(X) = a + b\,\Gamma(1 - {}^1/_k), \quad k > 1$

Mode $\widehat{M} = \quad = a + b\,(1 + {}^1/_k)^{-1/k}$

Médiane $\overline{M} = \quad = a + b\,(\ln, 2)^{-1/k}$

Variance $\sigma^2 = V(X) = b^2\,[\Gamma(1 - {}^2/_k) - \Gamma^2(1 - {}^1/_k)], \quad k > 2$

Dans le cas particulier a = 0, les paramètres b et k peuvent être estimés par la méthode des moments à partir des formules :

$$\log \widehat{b} = \overline{\log.\, x} - \frac{0{,}5772}{\widehat{k'}}$$

$$\frac{1}{\widehat{k'}} = \frac{\sqrt{6}}{\eta}\, s\,(\log.\, x)$$

avec $k' = 2{,}3026\,k$, les logarithmes étant ici des logarithmes décimaux, $\overline{\log\, x}$ et $s\,(\log\, x)$ étant respectivement la moyenne et l'écart-type empiriques de $\log\, x$, estimés à partir de l'échantillon observé.

D'autres lois doublement exponentielles ont été envisagées dans les études de durée de vie (fiabilité).

Citons, par exemple, la loi définie par la fonction de répartition

$$(\Phi_1) \qquad F(x) = 1 - \exp\left[-\frac{e^x - a}{b}\right]$$

$\ln a < x < +\infty$, $b > 0$, $a > 0$, qui a été utilisée dans des études de résistance des tissus à l'abrasion, sous la forme

$$(\Phi_1) \qquad F(x) = 1 - \exp\left[-\frac{e^x - 1}{b}\right], \quad x > 0$$

On a encore

$$\ln[-\ln(1 - F(x))] = \ln u - \ln b$$

fonction linéaire croissante de la variable auxiliaire In u avec :

$$u = e^x - 1$$

Son étude graphique avec le papier loi de Weibull exigera la transformation préalable définie par $u = e^x - 1$ (il existe des tables très détaillées de la fonction e^x [9] et [10].

Le changement de variable défini par

$$\frac{e^x - 1}{b} = \frac{e^{-y} - 1}{b'}, \quad b' > 0$$

donnerait la loi définie par

$$(\Phi_2) \quad F(y) = \exp\left[-\frac{e^{-y} - 1}{b'}\right]$$

$$-\infty < y < 0,$$

d'où

$$\ln[-\ln F(y)] = \ln(e^{-y} - 1) - \ln b'$$

La loi (Φ_1) a été utilisée dans l'étude de problèmes de corrosion sous la forme plus générale, [11] :

$$F(x) = 1 - \exp\left[-\frac{e^{\lambda x} - 1}{b}\right]$$

$$0 < x < \infty, \qquad \lambda \text{ et } b \text{ positifs}$$

Le taux instantané de mortalité est alors $\varphi(x) = \frac{e^{\lambda x}}{b}$, le paramètre supplémentaire λ caractérisant, comme dans le lot de Gompertz, un taux qui s'accroît dans le temps dx d'une quantité $d\varphi$ proportionnelle à dx

$$\frac{1}{\varphi}\frac{d\varphi}{dn} = \lambda$$

La loi de Gumbel (G_1) et la loi (Φ_1) sont des cas particuliers de cette hypothèse générale, correspondant à des conditions initiales différentes :

$$\text{Loi } (G_1) : F(x) = 0 \text{ pour } x = -\infty$$
$$\text{Loi } (\Phi_1) : F(x) = 0 \text{ pour } x = 0$$

Remarques

1° Les lois de Gumbel, Weibull, Fréchet... présentées par différents auteurs sous des formes variées, résultent des travaux de Fisher et Tippett sur les formes asymptotiques des distributions des valeurs extrêmes, [8].

Moyennant un changement de variable et un changement de paramètres, elles peuvent toutes se ramener à l'une d'entre elles, par exemple à la loi de Weibull.

2° L'emploi d'un graphique comportant, d'une part, un axe des ordonnées gradué en Φ avec l'échelle fonctionnelle $\ln[-\ln(1-\Phi)]$, (avec $\Phi = F$ ou $1 - F$ suivant le cas) et, d'autre part un axe des abscisses avec, soit une échelle fonctionnelle en $\ln x$, graduée en x, soit une échelle uniforme en x, permettra d'étudier la validité de l'une quelconque de ces lois pour une population dont on a observé un échantillon et d'estimer éventuellement les paramètres de cette loi.

Pour certaines de ces lois, le taux instantané de mortalité :

$$\varphi(x) = \frac{f(x)}{1 - F(x)}$$

s'exprime simplement en fonction des paramètres :

a) Loi de Gumbel (G_1)

$$F(x) = 1 - \exp\left[-e^{\alpha(x-u)}\right], \qquad \varphi(x) = \alpha\, e^{-\alpha u}\, e^{\alpha x}$$

b) Loi de Weibull (W)

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

c) Loi (Φ)

$$F(x) = 1 - \exp\left[-\frac{e^{x}-1}{b}\right], \qquad \varphi(x) = \frac{l^x}{b}$$

Sauf pour la loi de Weibull, lorsque $0 < \beta < 1$, il s'agit de taux croissant avec x, pouvant caractériser un phénomène d'usure.

BIBLIOGRAPHIE

[1] GUMBEL. — Statistical theory of extreme values and practical applications National Bureau of Standards. *Math. Series 33.* Washington. 1954.

[2] GUMBEL. — Étude statistique de la fatigue des matériaux, *Revue de Statistique Appliquée*, 1957, n° 4, p. 51 à 86.

[3] *N.B.S.* — Probability tables for the analysis of extreme value data *Appl. Math. Series 22.* Washington. 1953.

[4] MORICE. — Les graphiques à échelles fonctionnelles du statisticien. *Revue de Statistique Appliquée*, 1964, n° 3, p. 84 à 91.

[5] MORICE. — Quelques modèles mathématiques de durée de vie. *Revue de Statistique Appliquée*, 1966, n° 1, p. 103 à 107.

[6] DOURGNON et REYROLLE. — Tables de la fonction de répartition de la loi de Weibull. *Revue de Statistique Appliquée*, 1966, vol. 14, n° 4, p. 83-116.

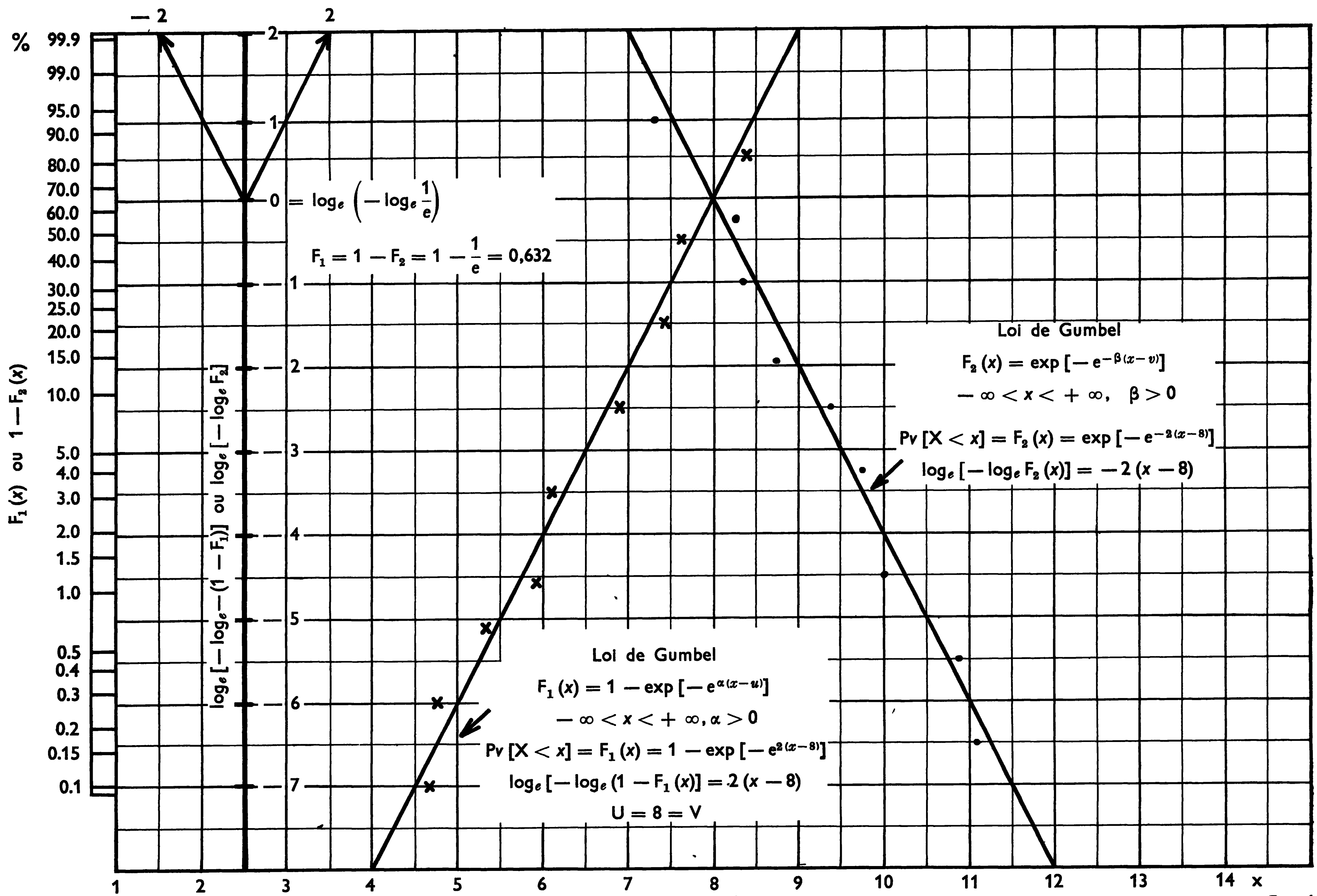
[7] Allan PLAIT. — The Weibull distribution with tables. Industrial Quality Control. Nov. 1962.

[8] FISHER and TIPPETT. — Limiting forms of the frequency distribution of the largest and smallest member of a sample. Proceedings Cambridge Philosophical Society — 24 — Part. 2, 1928, p. 180-190.

[9] M. BOLL. — Tables numériques universelles. Dunod. Cet ouvrage contient en particulier des tables des fonctions $\Gamma(x)$, $\ln[\ln x]$ et e^x.

[10] *N.B.S.* — Table of the exponential function e^x, *AMS. 14.* Nat[l] Bur. of Standards. Washington
N.B.S. — Table of the descending exponential function e^{-x} *A.M.S.* 46, Nat[l] Bur. of Standards. Washington.

[11] Lloyd and Lipow. Reliabilitz, management, methods and mathematics. Prentice-Hall, 1962, p. 140.



%
$F_1(x)$ ou $1 - F_2(x)$
$\log_e[-\log_e-(1-F_1)]$ ou $\log_e[-\log_e F_2]$
$0 = \log_e\left(-\log_e \frac{1}{e}\right)$
$F_1 = 1 - F_2 = 1 - \frac{1}{e} = 0{,}632$
Loi de Gumbel
$F_2(x) = \exp[-e^{-\beta(x-v)}]$
$-\infty < x < +\infty, \quad \beta > 0$
$Pv[X < x] = F_2(x) = \exp[-e^{-2(x-8)}]$
$\log_e[-\log_e F_2(x)] = -2(x-8)$
Loi de Gumbel
$F_1(x) = 1 - \exp[-e^{\alpha(x-u)}]$
$-\infty < x < +\infty, \alpha > 0$
$Pv[X < x] = F_1(x) = 1 - \exp[-e^{2(x-8)}]$
$\log_e[-\log_e(1-F_1(x)] = 2(x-8)$
$U = 8 = V$
x

Fig. 1

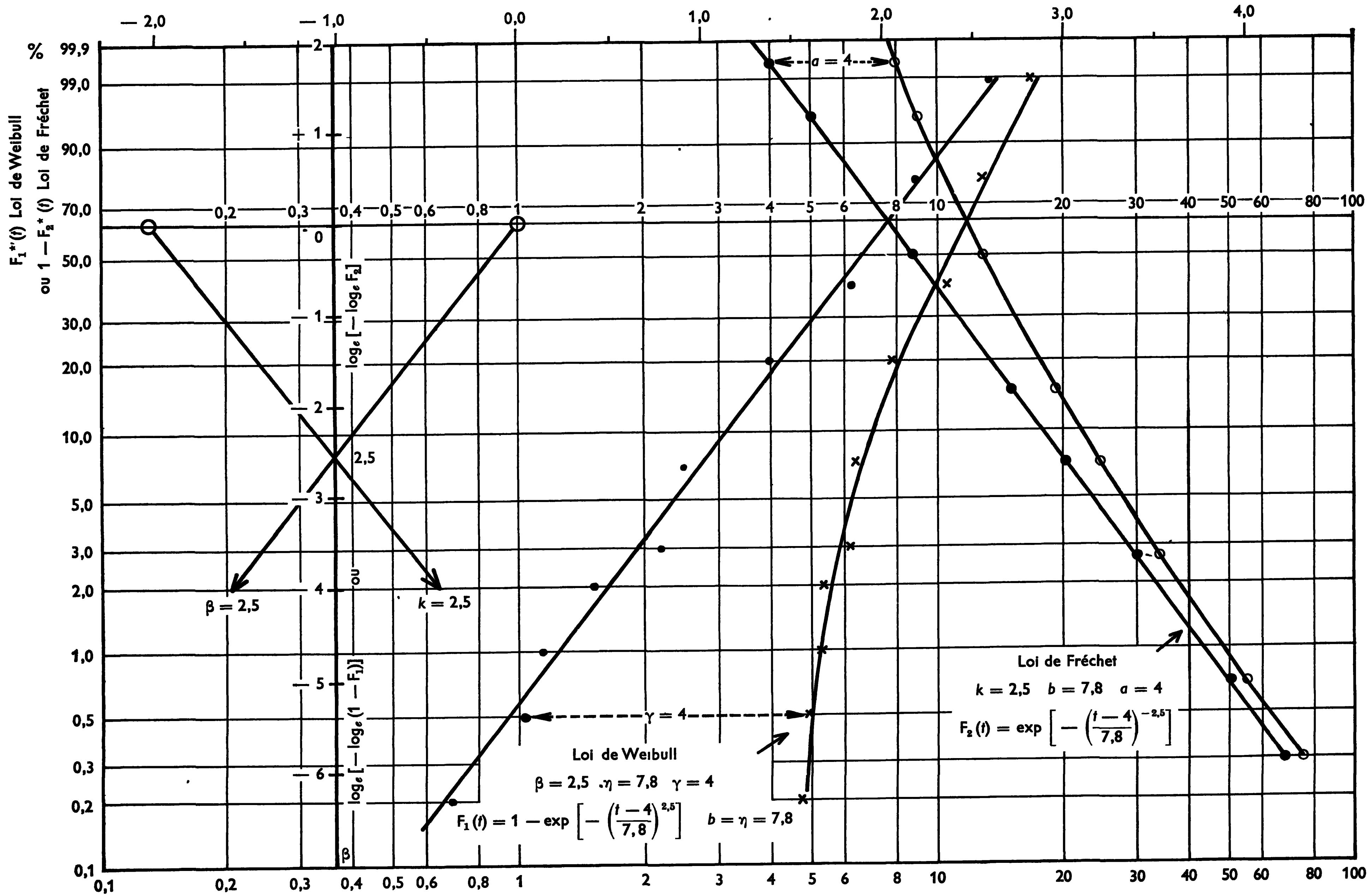

F1*'(t) Loi de Weibull
ou 1 — F2* (t) Loi de Fréchet
%
log_e [— log_e F2]
ou
log_e [— log_e (1 — F1)]
β = 2,5
k = 2,5
2,5
a = 4
γ = 4
Loi de Weibull
β = 2,5 η = 7,8 γ = 4
F1 (t) = 1 — exp [— ((t — 4)/7,8)^2,5]
b = η = 7,8
Loi de Fréchet
k = 2,5 b = 7,8 a = 4
F2 (t) = exp [— ((t — 4)/7,8)^-2,5]

Fig. 2