

ICOLD/CIGB

102.2-R-65

S. Francisco 1988

Q.63 C.....

DISTRIBUTION A DOUBLE BORNE LOGARITHMIQUE *

G. LOMBARDI

Dr Ing. Dr h.c. - Ingénieur-Conseil

Locarno (Suisse)

(*) DOUBLE LOGARITHMIC BOUNDED DISTRIBUTION

1. INTRODUCTION

Le rapport Q.63 R....., présenté au 16ème Congrès de la CIGB à San Francisco, fait usage d'une distribution bornée à double limite logarithmique, appelée 2LN. Dans la présente communication on s'étendra quelque peu sur l'aspect mathématique de cette distribution et sur la procédure à suivre en vue de son application pratique.

2. FONCTION BORNEE A DOUBLE LIMITE LOGARITHMIQUE

Partant de l'idée que toute variable physique possède des limites ou bornes supérieure et inférieure qui ne sauraient être dépassées, on a cherché une distribution qui respecte cette condition, tout en permettant une compréhension intuitive de ses particularités.

La distribution classique log-normale, qui est dérivée de la distribution normale dite de Gauss, et qui fait subir à la variable une transformation logarithmique, a servi de guide lors de la définition de cette nouvelle loi de distribution. On a ainsi cherché une fonction qui soit limi-

tée non seulement vers le bas comme la fonction log-normale, mais qui le soit d'une façon équivalente également vers le haut. La fonction est donc doublement limitée par des bornes de type logarithmique. Ainsi que nous l'avons montré dans le rapport mentionné, cette procédure équivaut tout simplement à appliquer une distribution normale non point directement à la fonction physique mais à une de ses transformées choisie d'une façon adéquate.

Dans sa forme normalisée, la distribution proposée assume l'expression suivante:

$$(1) \quad x = \frac{1}{A} \cdot (V - V_m + B)$$

$$y_1 = \log (1 + x) - \log (1 - x)$$

$$y = F \cdot y_1 - G$$

V représente la variable statistique ou physique à étudier,

V_m la valeur moyenne,

x une première transformée linéaire de la variable physique (paramètres A et B),

y_1 une deuxième transformée logarithmique, et

y une troisième transformée linéaire qui devient l'argument mathématique de la distribution (Paramètres F et G).

Alors que y est illimitée dans les deux directions x est limitée par les bornes -1 et $+1$, c'est dire que les bornes de la variable physique V sont, elles :

$V_{\max} = V_{m-B+A}$, et $V_{\min} = V_{m-B-A}$, alors que :

$V_c = V_{m-B} = (V_{\max} + V_{\min})/2$ est la valeur centrale de la variable.

Les paramètres A et B permettent donc d'accommoder, comme d'habitude, la moyenne et la dispersion quadratique.

On admet, par choix délibéré, que la variable y soit justifiable d'une distribution de probabilité normale. Le paramètre G détermine la dissymétrie de la distribution, car il permet de déplacer le maximum de la densité de probabilité à l'intérieur de l'intervalle de définition de la variable x . A son tour, le paramètre F permet de concentrer plus ou moins intensément la distribution autour de ce point, selon les nécessités du cas étudié.

La distribution suggérée porte de ce fait le nom de 2LN, c'est-à-dire 2-log-normale.

La figure 1, reprise du rapport Q63 R...., montre une façon que l'on peut dire "géométrique" de construire cette distribution. Fig.1

Il en résulte clairement une distribution des densités de probabilité doublement bornée, à un seul maximum et librement dissymétrique, dont quelques exemples d'application ont été montrés dans ledit rapport. Il va de soi que la valeur moyenne ne correspond plus à la valeur centrale (ni à la valeur médiane) et que l'écart relatif qui sépare la moyenne de la valeur centrale est une des caractéristiques de la distribution, ainsi que cela ressort de la figure 3 du rapport mentionné.

La distribution comporte de ce fait 4 paramètres libres. Outre les deux paramètres de forme F et G déjà mentionnés, les deux autres paramètres A et B s'obtiennent par la transformation habituelle de la variable physique qui permet, elle, de faire coïncider les valeurs moyennes des deux distributions, ainsi que leurs échelles, en terme de déviation quadratique moyenne, le long de l'axe de la variable.

Une fois l'ajustement obtenu, on trouve facilement par calcul les bornes V_{\max} et V_{\min} de la distribution avec lesquelles on retourne à la série expérimentale. D'une façon générale, on pourrait aussi concevoir les quatre valeurs V_{\max} , V_{\min} , F et G comme quatre paramètres indépendants à trouver pour obtenir l'ajustement optimal.

La figure 2 donne, pour un exemple quelconque, l'allure de Fig. 2 la densité de probabilité (p) et la fonction cumulative de probabilité (P) correspondante.

On conçoit aisément qu'à la limite $F \rightarrow \infty$ et $G=0$, la distribution 2LN coïncide avec la distribution normale de Gauss de laquelle elle est en fait une généralisation.

3. PROCÉDURE D'AJUSTEMENT

Pour la série expérimentale étudiée il s'agit tout simplement de déterminer les quatre paramètres libres de la fonction mathématique en satisfaisant au mieux le critère d'ajustement que l'on se sera fixé. Un premier critère pourrait être celui assez classique qui consiste à minimiser l'erreur quadratique moyenne (méthode des moindres carrés) entre la série expérimentale et la loi mathématique de distribution. Un deuxième critère consiste à faire coïncider les quatre premiers moments de la série avec ceux de la distribution, à savoir: moyenne, déviation standard, troisième et quatrième moment (c'est-à-dire rayons de premier, deuxième, troisième et quatrième or-

dre). C'est cette méthode qui sera développée au paragraphe suivant. Il va sans dire qu'il serait également concevable de choisir une méthode combinée, consistant à faire coïncider les deux premiers moments (moyenne et dispersion quadratique), et à définir les deux paramètres de forme par recours au critère des moindres carrés. D'autres combinaisons enfin sont également pensables.

Il nous a semblé que le fait de faire coïncider les rayons ou moments des quatre premiers ordres soit en définitive la méthode qui permette d'appréhender le mieux les caractéristiques intrinsèques de la série donnée. La comparaison des moments d'ordre supérieur, par exemple les cinquième et sixième, de la série expérimentale et de la fonction de distribution obtenue, permettra de juger du résultat. La qualité de l'ajustement peut également être jugée sur la base de l'écart quadratique des différences résiduelles (qui n'est donc pas toujours le minimum possible) entre les deux fonctions.

4. AJUSTEMENT PAR COÏNCIDENCE DES MOMENTS

Une série expérimentale étant donnée, il est aisé de calculer les rayons de divers ordres. Ainsi:

- le rayon ou moment de premier ordre nous donne la moyenne, c'est-à-dire la position de la variable sur son échelle,
- le deuxième rayon nous fournit la déviation standard, c'est-à-dire en fait l'échelle de la variable,
- le troisième rayon indique la dissymétrie de la distribution, et
- le quatrième moment est caractéristique de sa forme. Dans la définition de cette forme les valeurs extrêmes de la série interviennent avec davantage de poids que les valeurs centrales.

Notons qu'il est licite d'attribuer un poids différent aux diverses valeurs expérimentales.

Nous désignons les rayons de la série expérimentale par:

R_1 = moyenne

R_2 = écart quadratique moyen, soit déviation standard

R_3 = rayon de troisième ordre

R_4 = rayon de quatrième ordre

et définissons ensuite les deux rapports suivants:

$$R_{32} = \frac{R_3}{R_2} \quad \text{et} \quad R_{42} = \frac{R_4}{R_2}$$

en choisissant donc l'écart quadratique moyen R_2 comme mè-
tre de la distribution expérimentale.

De même on définit pour la distribution mathématique les
valeurs $r_1, r_2, r_3, r_4, r_{32}$ et r_{42} .

Les rapports R_{32} et R_{42} (et évidemment r_{32} et r_{42} aussi)
représentent des caractéristiques fondamentales de la
fonction étudiée et en définissent complètement l'allure.
Des tables numériques et des représentations graphiques
ont été établies, qui donnent entre autres les valeurs de
ces deux rapports en fonction des deux paramètres de forme
F et G.*

La figure 3 indique, sous forme graphique et paramétrique, Fig. 3
la valeur r_{32} en fonction des deux coefficients de forme F
et G.

La figure 4 en fait de même pour ce qui est du rapport Fig. 4
 r_{42} .

* Ces tableaux et ces graphiques peuvent être mis à
disposition des personnes intéressées.

Une fois les rapports R_{32} et R_{42} de la série expérimentale déterminés il est aisé, à l'aide de ces graphiques, de trouver les valeurs des paramètres F et G qui permettent de satisfaire aux deux conditions suivantes:

$$R_{32} = r_{32} \quad \text{et} \quad R_{42} = r_{42}.$$

La figure 5 démontre la construction suivie. De la figure Fig. 5 3 a été extraite, ou interpolée, la courbe r_{32} qui correspond à la valeur expérimentale du rapport R_{32} , et de même on a extrait de la figure 4 la fonction r_{42} qui correspond à la valeur du rapport R_{42} . Le point de croisement des deux courbes fournit directement la valeur des deux paramètres de forme F et G . La forme de la distribution étant ainsi définie par ces deux paramètres, il est alors chose aisée que de faire coïncider respectivement la valeur moyenne et la dispersion quadratique de la série donnée et de la loi de distribution.

5. EXEMPLE

A titre d'exemple on a choisi une série expérimentale portant sur 182 éléments. Il s'agit de résistances de béton.

Il va de soi que l'intérêt pratique des valeurs extrêmes est, dans le cas présent, relatif, alors que pour d'autres variables, p.ex. des débits de crue, il pourrait être prépondérant; mais cet exemple montre bien le procédé suivi.

Le calcul nous a donné immédiatement, à partir des 182 valeurs expérimentales, les valeurs suivantes:

$$R_1 = 26,262 \text{ MPa, moyenne}$$

$$R_2 = 2,935 \text{ MPa, écart quadratique moyen, ce qui donne une variance de } 2,935/26,262 = 11,2\%$$

$$R_3 = 2,078 \text{ MPa, troisième rayon}$$

$$R_4 = 3,708 \text{ MPa, quatrième rayon}$$

Les rapports R_{32} et R_{42} ressortent à

$$R_{32} = R_3/R_2 = 2,078/2,935 = 0,7080$$

$$R_{42} = R_4/R_2 = 3,708/2,935 = 1,2634$$

Le graphique qui a été reproduit à la figure 3 nous fournit la courbe dont le paramètre correspond à la valeur de R_{32} , et il en est de même du graphique de la figure 4 pour ce qui est du rapport R_{42} . Ce sont justement ces deux courbes qui ont été choisies pour dessiner la figure 5 qui livre les valeurs des paramètres de forme F et G.

$F = 1,300$, et $G = 0,66714$.

Ces valeurs peuvent s'obtenir également par calcul, cela va de soi.

Le graphique, ou le tableau correspondant, fournissent, à partir de ces deux paramètres, la valeur de

$$r_2 = 0,326931$$

ce qui nous donne immédiatement l'unité U sur l'échelle de la variable physique, soit:

$$U = \frac{R_2}{r_2} = 2,935 \text{ MPa} / 0,326931 = 8,978 \text{ MPa}.$$

Cette valeur correspond à l'intervalle compris entre la valeur centrale et les valeurs maximale et minimale respectivement.

De la même façon, on obtient la distance entre la moyenne et la valeur centrale, à partir du tableau correspondant ou par calcul, avec $x_m = - 0,2234$.

En tenant compte de l'unité d'échelle U définie ci-dessus, l'écart réel est de:

$$B = V_m - V_c = U \cdot x_m = - 8,978 \cdot 0,2234 = - 2,006 \text{ MPa}.$$

Partant, la valeur centrale est:

$$V_c = 26,262 \text{ MPa} + 2,006 \text{ MPa} = 28,268 \text{ MPa}.$$

Les valeurs extrêmes de résistance sont donc

$$V_{\max} = 28,268 + 8,978 = 37,246 \text{ MPa}$$

$$V_{\min} = 28,268 - 8,978 = 19,290 \text{ MPa.}$$

Sur la base de ces valeurs, il est aisé de dessiner la figure 6, qui représente d'une part la distribution expérimentale sous forme d'histogramme et de la comparer à la densité de probabilité mathématique (p). On remarquera la nette dissymétrie de la distribution. D'autre part la figure 7 reproduit la fonction cumulative de probabilité (P) avec les valeurs expérimentales correspondantes. Notons encore que les moments d'ordre, cinquième et sixième, expérimentaux et théoriques respectivement, nous donnent les rayons et les rapports suivants:

$$\frac{R_5}{r_5 \cdot U} = \frac{3,43105}{0,39035 \cdot 8,978} = 0,979$$

$$\frac{R_6}{r_6 \cdot U} = \frac{4,23137}{0,47882 \cdot 8,978} = 0,9843$$

La coïncidence absolue de la série expérimentale et de la formule mathématique exigerait que ces rapports soient égaux à l'unité. On constate que l'écart est très faible.

Par ailleurs, l'erreur quadratique moyenne entre la série

expérimentale et la distribution théorique est à peine de 0,188 MPa., soit 0,7% de la résistance moyenne du béton. En appliquant une distribution normale, l'erreur serait supérieure au double de cette valeur.

On notera en passant que manifestement les inévitables erreurs et les arrondis de mesure contribuent pour une partie non négligeable à former cet écart. Par exemple, 9 points se trouvent exactement au 3ème décile. Une toute faible variation, qui les ramènerait de la 4ème à la 3ème classe conduirait à la ligne pointillée de la figure 6.

Les valeurs extrêmes de résistance de 19,3 MPa et 37,2 MPa semblent physiquement acceptables et même plutôt convaincantes.

6. REMARQUES COMPLÉMENTAIRES

Remarquons tout d'abord que dans les calculs précédents la fréquence expérimentale a été obtenue par application de la formule suivante:

$$\text{Freq}(L) = 1 - (L - \alpha) / (N + 1 - 2\alpha) \quad \text{avec } \alpha = 3/8$$

utilisée habituellement pour les distributions normales et Beta, mais on pourrait envisager d'utiliser également une autre expression.

Une deuxième remarque concerne la fonction log-2log-normale (L2LN) utilisée également dans le rapport Q.63 R... déjà mentionné. Il s'agit simplement, dans ce cas, d'appliquer les méthodes de calcul ci-devant, non point directement à la variable physique mais à son logarithme. Il serait alors possible de faire intervenir un cinquième paramètre libre qui déplacerait l'origine de la fonction par rapport au logarithme et qui, de ce fait, permettrait soit de faire coïncider exactement le rayon de cinquième ordre, soit de minimaliser l'erreur quadratique moyenne entre série expérimentale et distribution mathématique tout en respectant les conditions ci-devant. Il y aurait lieu de craindre toutefois que l'introduction d'un cinquième paramètre puisse, au moins dans certaines circonstances, donner trop de flexibilité à la distribution et en quelque sorte lui couper les ailes qui lui permettent d'extrapoler vers des valeurs à faible probabilité.

Il va de soi que si les calculs ci-devant ont été présentés sous forme graphique, cela n'a été fait que pour des raisons de facilité d'exposé. Un programme de calcul électronique permet évidemment de réaliser les ajustements nécessaires avec grande facilité et une précision bien supérieure à celle que toute construction graphique concède.

7. CONCLUSIONS

Par l'exemple étudié ci-dessus, nous pensons avoir montré que l'application pratique de la distribution de 2LN ne présente aucune difficulté. Théoriquement, une impasse pourrait toutefois surgir dans les cas où la série statistique ne serait pas redevable d'une distribution de ce type.

Ce n'est qu'un usage prolongé de cette distribution qui permettra de juger si elle peut donner des résultats satisfaisants dans la pratique, puisqu'on ne saurait justifier la validité d'une loi de distribution uniquement sur la base de considérations théoriques. Au moins faut-il reconnaître qu'étant doublement bornée, cette distribution ne peut pas enfreindre certaines évidences physiques.

Locarno, octobre 1987

Dr Ing. Dr h.c. G. Lombardi

Lo/rz

RÉSUMÉ

Partant du principe que toute variable physique est doublement bornée, on établit une distribution de probabilité, appelée 2LN, qui respecte cette condition et qui présente une flexibilité suffisante mais non excessive pour s'adapter aux phénomènes physiques observés. On montre la procédure pratique utilisée pour ajuster la distribution 2LN à la série expérimentale.

Summary

Starting from the consideration that any physical variable has to be bounded on both sides, a probability distribution law, called 2LN, is proposed, which respects this condition.

This distribution offers a good flexibility but not an excessive one in order to fit with adequacy the observed physical phenomena. The practical procedure is shown, which permits to adjust the 2LN-law to the experimental row.

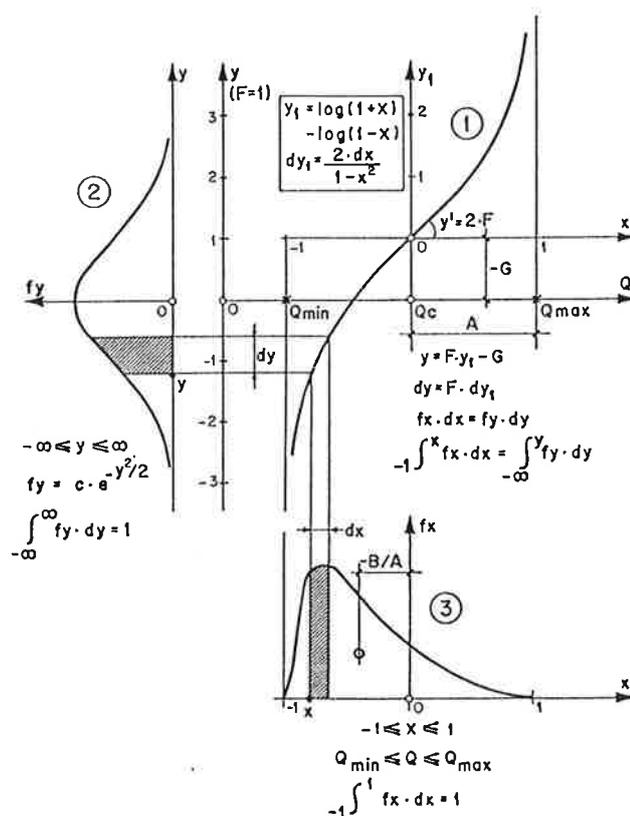


Figure 1

Distribution 2-log-normale

Q = variable physique (p.ex. débit)

Q_c = valeur centrale de Q

x = variable semi-physique = $2 \cdot (Q - Q_c) / (Q_{\max} - Q_{\min})$

y = variable mathématique = $F \cdot y_1 - G$

y_1 = variable mathématique normalisée

(1) transformation $y_1 = \log(1+x) - \log(1-x)$

(2) distribution normale de y

(3) distribution 2-log-normale de x : $2LN(x)$

Figure 1

2-log-normal distribution

Q = physical variable (e.g. flow)

Q_c = central value of Q

x = semi-physical variable = $2 \cdot (Q - Q_c) / (Q_{\max} - Q_{\min})$

y = mathematical variable = $F \cdot y_1 - G$

y_1 = normalized mathematical variable

(1) transformation $y_1 = \log(1+x) - \log(1-x)$

(2) normal distribution of y

(3) 2-log-normal distribution of x : $2LN(x)$

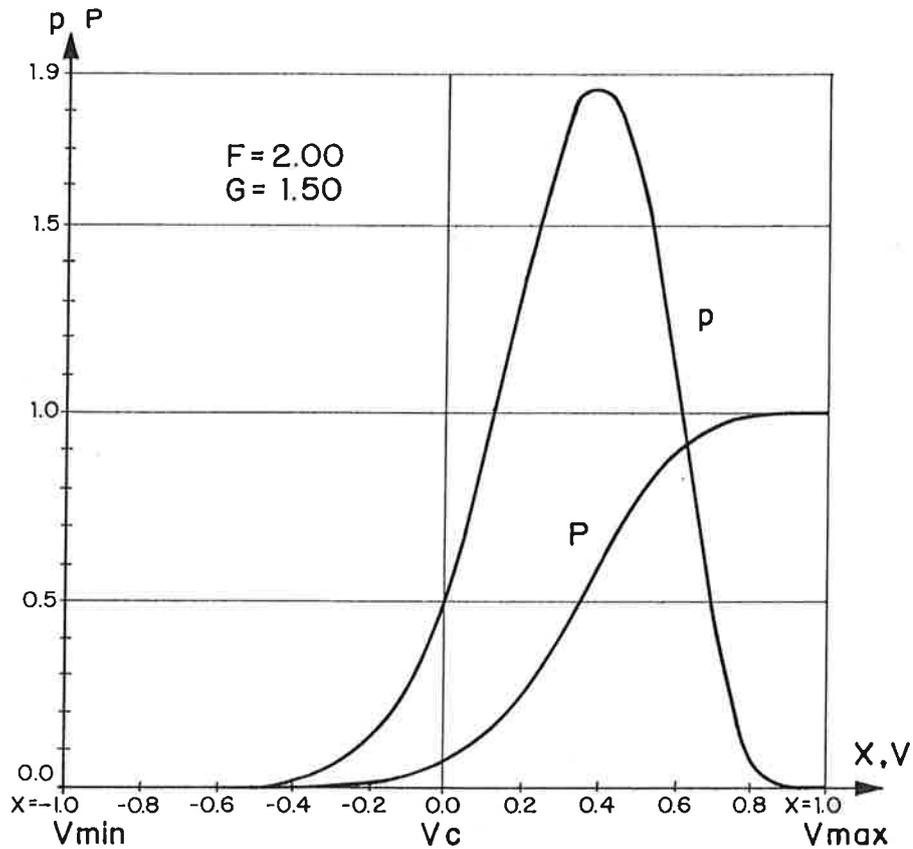


Figure 2

Exemple de fonction 2LN, avec distribution de la densité de probabilité p et valeur cumulative de la probabilité P .
($F = 2.00$, $G = 1.50$)

Figure 2

Example of a 2LN law, with distribution of the probability density p and probability of non-exceedence P .
($F = 2.00$, $G = 1.50$)

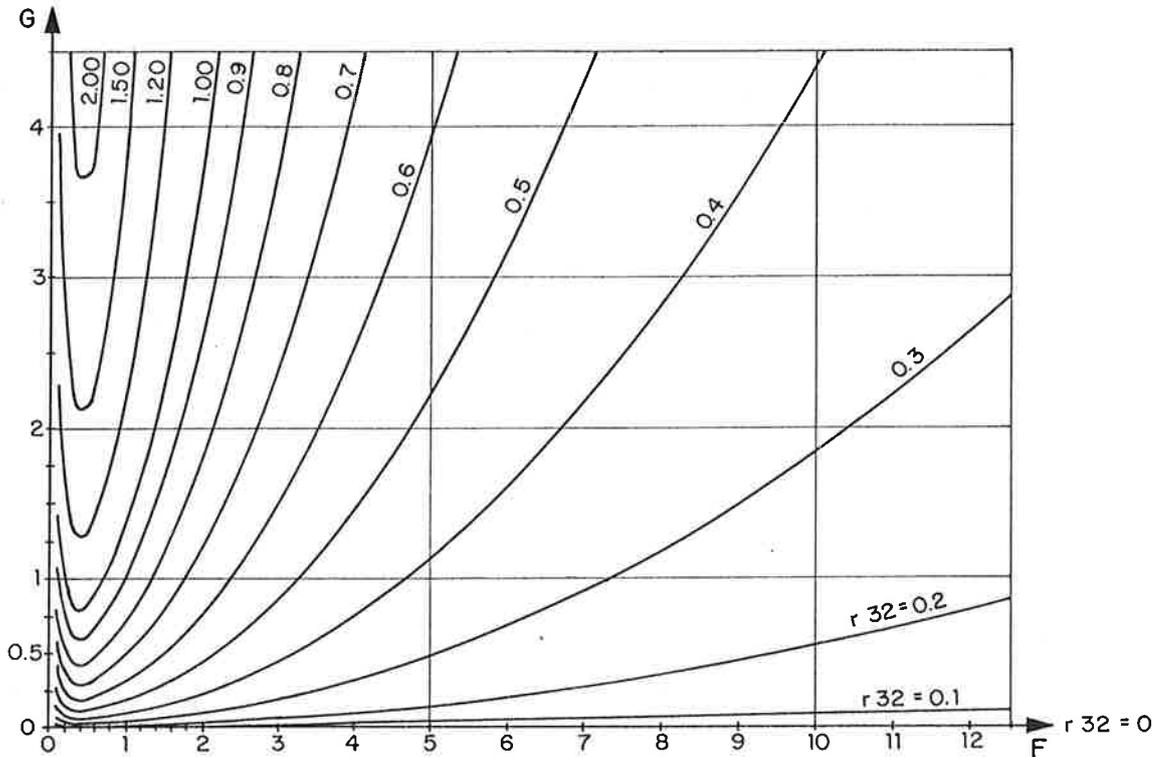


Figure 3

Graphique du rapport $r_{32} = r_3 / r_2$ en fonction des paramètres de forme F et G pour la distribution 2LN.

Figure 3

Plot of the ratio $r_{32} = r_3 / r_2$ as fonction of the two shape parameters F and G of the law 2LN.

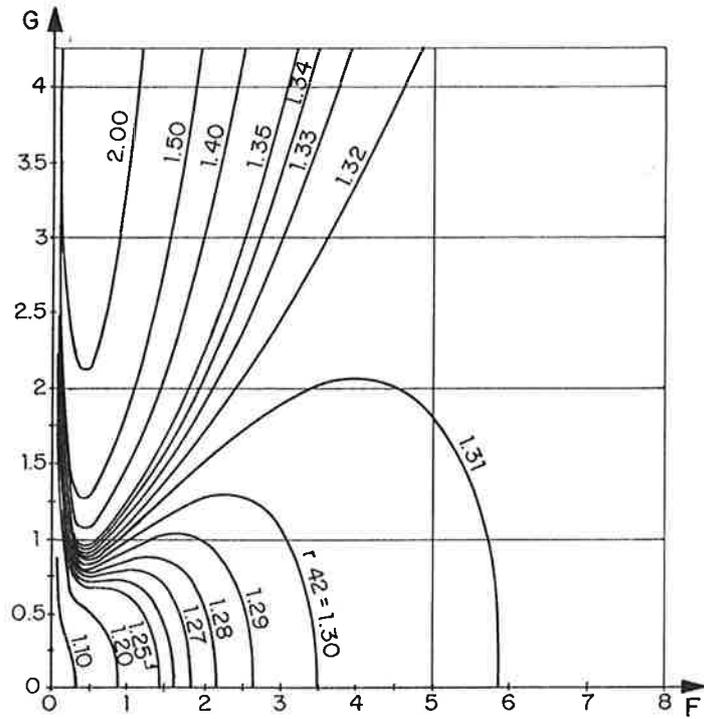


Figure 4

Graphique du rapport $r_{42} = r_4 / r_2$ en fonction des paramètres de forme F et G pour la distribution 2LN.

Figure 4

Plot of the ratio $r_{42} = r_4 / r_2$ as fonction of the two shape parameters F and G of the law 2LN.

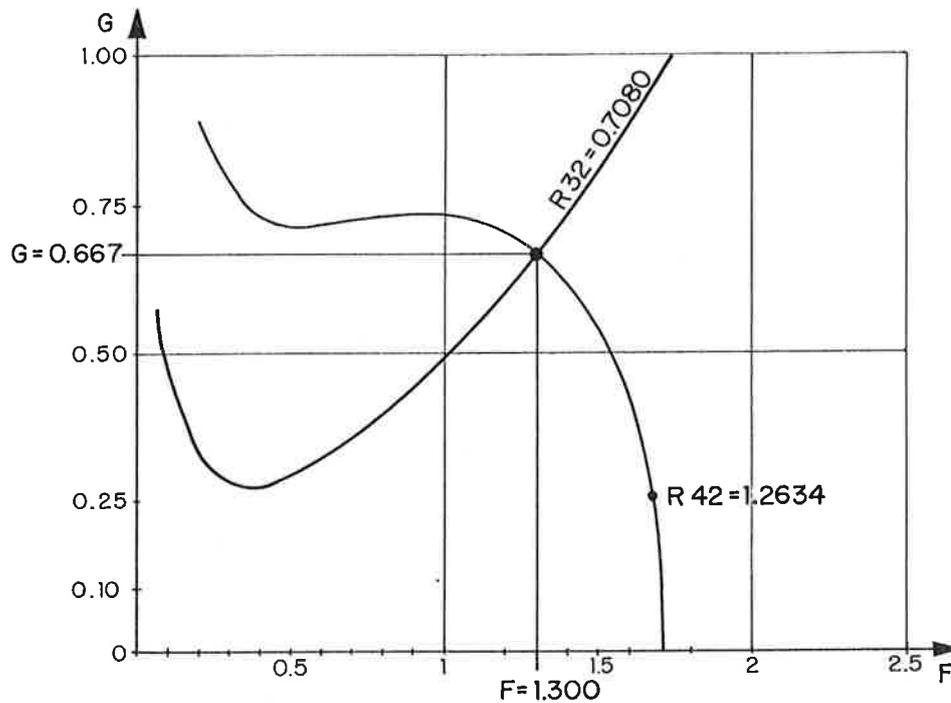


Figure 5

Détermination des valeurs des paramètres de forme $F = 1,300$ et $G = 0,667$, à partir des rapports $R_{32} = 0,7080$ et $R_{42} = 1,2634$ sur la base des courbes extraites des graphiques des figures 3 et 4.

Figure 5

Determination of the values of the shape parameters $F = 1,300$ and $G = 0,667$, from the ratios $R_{32} = 0,7080$ and $R_{42} = 1,2634$ based on the two lines out of the figures 3 and 4.

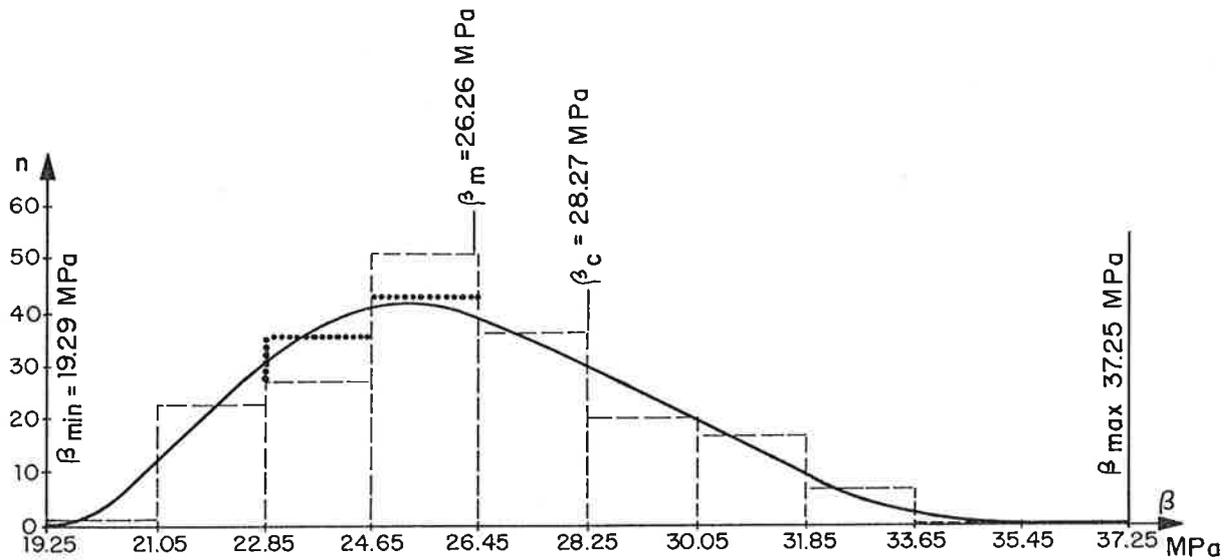


Figure 6

Exemple de distribution de résistances de béton.
Densité de probabilité et histogramme correspondant.

Figure 6

Example of a distribution of concrete strength.
Probability density and corresponding histogram.

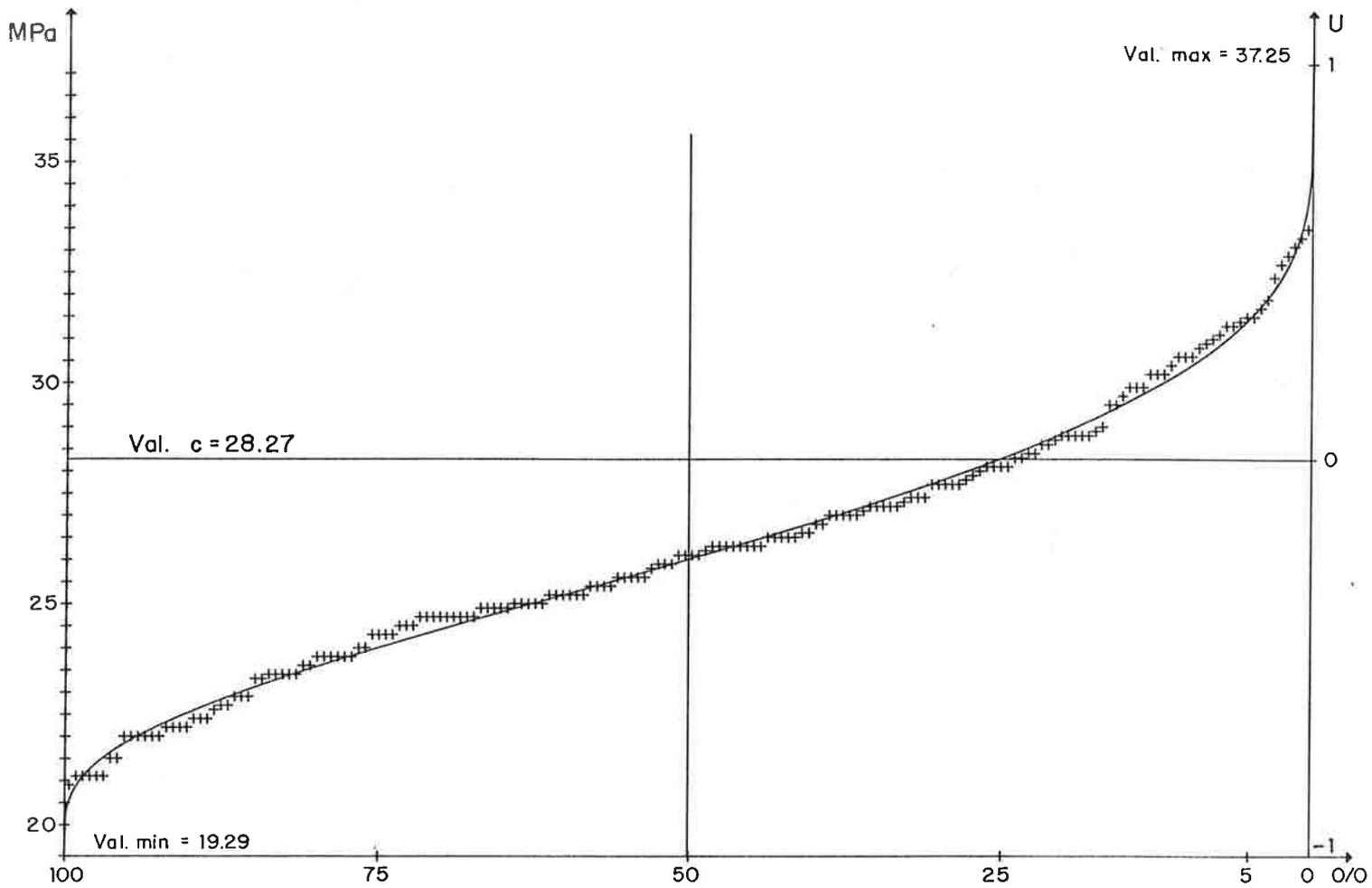


Figure 7

Exemple de la figure 6. Probabilité cumulative et série expérimentale. Echelles linéaires.

Figure 7

Same example as figure 6. Probability of exceedence and experimental row. Linear scales.

Fiche bibliographique

1. DISTRIBUTION A DOUBLE BORNE LOGARITHMIQUE	2. 1988
	3. Q63
	4. C...
	5. Français
6. Hydrologie, <u>Crue méthode statistique</u> , Résistance, Matériaux.	
7. On définit une distribution de probabilité à double borne logarithmique (2LN) et on montre la procédure à suivre pour l'ajuster à la série expérimentale.	
8.	
9. LOMBARDI G., Suisse	
10. Distribution à double borne logarithmique	
11. CIGB 16, San Francisco 1988, Vol. ..., Q63, C...	
12. A - F	

1. DOUBLE LOGARITHMIC BOUNDED DISTRIBUTION	2.1988 3.Q63 4.C ..
6. Hydrology, <u>Flood discharge statistical method</u> , Strength, <u>Materials</u> .	5.French
7. A double logarithmic bounded distribution (2LN) is proposed. The procedure to be followed to fit the experimental row is explained.	
8.	
9. LOMBARDI G., Switzerland	
10. Distribution à double borne logarithmique	
11. ICOLD 16, San Francisco 1988, Vol. .., Q63, C...	
12. A - F	