

Aktuárské vědy

E. J. Gumbel

La distribution des inondations

Aktuárské vědy, Vol. 7 (1938), No. 2, 85–93

Persistent URL: <http://dml.cz/dmlcz/144692>

Terms of use:

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://dml.cz>

f) Festschrift zum siebzigsten Geburtstag Prof. Dr. G. Rosmaniths. Prag 1935. Die Anwendung der Theorie der Elementarwahrscheinlichkeit zweier Abweichungen auf Dauerspiele.

3. Rietz-Baur, Handbuch der mathematischen Statistik. Leipzig und Berlin 1930, VIII. Einfache Korrelation.

4. a) Felix Bernstein, Variations- und Erblchkeitsstatistik. Berlin 1929.

b) Oskar Anderson, Die Korrelationsrechnung in der Konjunkturforschung, Bonn 1929.

5. Metron, Vol. XII.-Nr. 4. Roma 1936. Das Wahrscheinlichkeitsgesetz zweier wahrer einander zugeordneten Fehler und einige mit diesem zusammenhängende Betrachtungen.

6. J. L. Coolidge, Einführung in die Wahrscheinlichkeitsrechnung, Leipzig-Berlin 1927.

7. Sitzungsberichte der Kaiserl. Akademie der Wissenschaften in Wien, Mathem.-naturw. Classe Bd. CII, Abth. II. a. E. Kobald in Leoben, Über einige particuläre Lösungen der Differentialgleichung für die Wärmeleitung in einem Kreiszyylinder und deren Anwendung.

Im Übrigen verweist der Verfasser auch auf den Literaturnachweis in seinen früheren, in der Zeitschrift *Metron* erschienenen Aufsätzen, dem er noch.

R. W. M. Gibbs, *The Adjustment of Errors in practical Science*, London 1929 hinzufügen möchte.

La distribution des inondations.

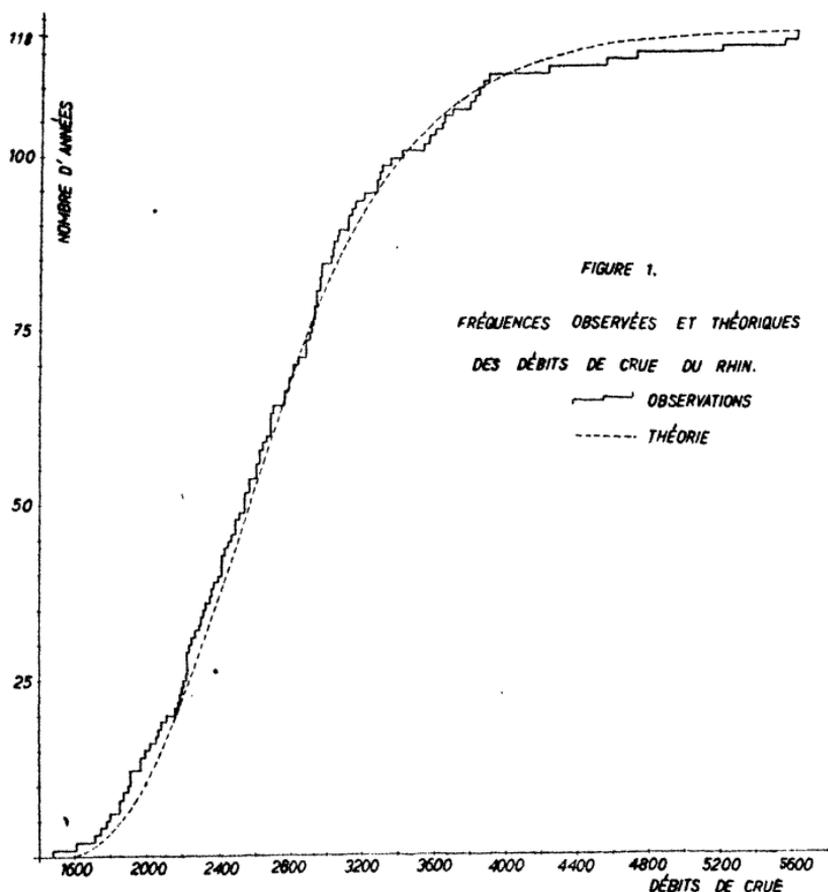
par *E. J. Gumbel*, Université de Lyon.

Dans un article précédent (8), nous avons développé la théorie de la plus grande valeur. Puis, nous l'avons appliquée au problème de l'âge limite (9) et aux distances entre les émissions radioactives (10). Mais cette théorie générale permet maintes autres applications pratiques, puisque la distribution finale a la même forme pour toutes les distributions initiales du type exponentiel. Dans ce qui suit nous allons étudier le problème des inondations.

Plusieurs auteurs, notamment Hazen (12), Gibrat (4) et Grassberger (5) ont montré dans leurs travaux systématiques que l'on peut traiter le débit d'un grand fleuve en première approximation comme une variable statistique soumise à une distribution de Galton, que l'on obtient à partir de la distribution de Gauss par une transformation logarithmique. Mais ce problème n'est pas encore tout à fait résolu. Car, même pour les grands fleuves, dont les débits montrent une certaine régularité, il existe au moins deux raisons distinctes des fluctuations des débits: au printemps la fusion de la neige et en automne la pluie. Il en résulte au moins deux valeurs dominantes des débits, ce qui complique

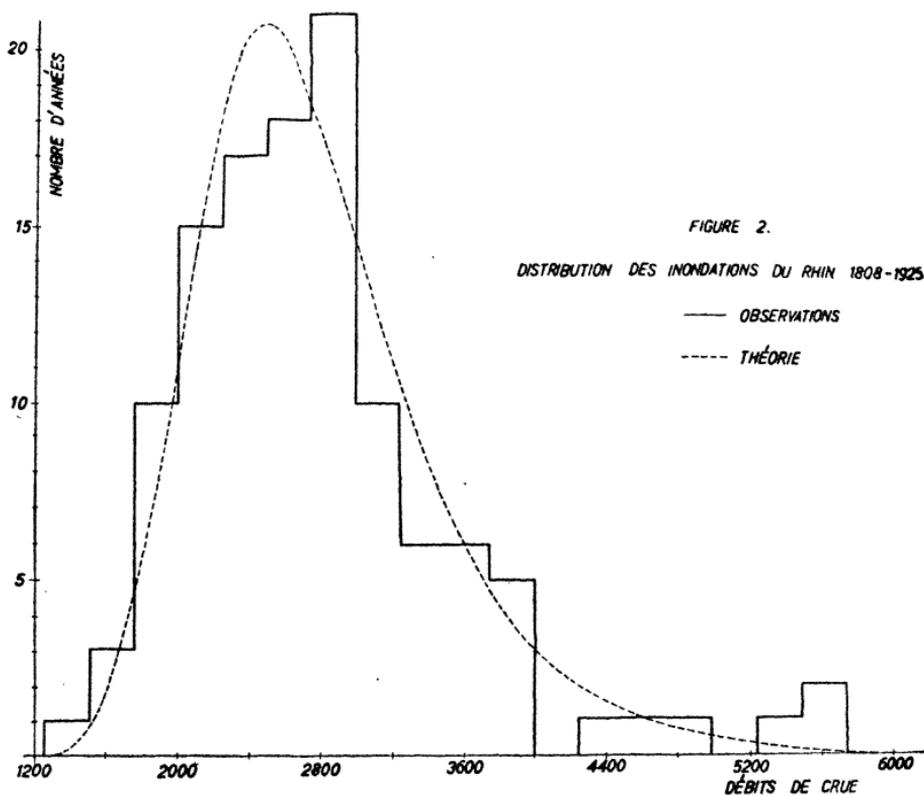
la représentation analytique. Tel est par exemple le cas pour la distribution des débits du Rhin à Bâle étudiés par M. Ghezzi (3).

La distribution des inondations a été traitée par M. A. Coutagne (1) à l'aide de formules d'interpolation. M. Grassberger (5—7) a introduit l'hypothèse que la plus grande valeur d'une distribution de Galton pourrait être soumise, elle-même à une distribution de Galton, hypothèse qu'il serait difficile de prouver.



Dans ce qui suit, nous allons représenter la distribution des inondations par celle de la plus grande valeur étant donné que les débits de crue sont certainement les plus grandes valeurs des débits. Nous nous basons sur le fait que la distribution finale de la plus grande valeur est, dans une certaine mesure, indépendante de la distribution initiale que l'on traite. Cela simplifie de beaucoup notre tâche. Car nous pouvons laisser de côté les difficultés mentionnées auparavant. La condition

nécessaire pour le passage de la distribution exacte de la plus grande valeur à sa forme finale est vérifiée. Car les débits des grands fleuves sont observés journellement, et les inondations sont connues quelquefois même pour un siècle. D'ailleurs, si la distribution initiale des débits suit la loi de Galton, la distribution finale des débits de crue est bien celle



de la plus grande valeur puisque la règle de l'Hopital vaut pour la loi de Galton.

Considérons les observations. Le travail de Ghezzi (3) contient les débits de crue pour le Rhin à Bâle et pour chaque année de 1808 à 1925, mesurés comme c'est usuel en m^3 : sec. Nous avons arrangé ces 118 valeurs d'après leur grandeur, ainsi nous obtenons la courbe en échelle tracée dans la figure 1. Pour avoir la répartition observée nous avons choisi comme unité des intervalles de $250 \frac{\text{m}^3}{\text{sec}}$ ce qui convient à la précision avec laquelle sont faites les observations. Cette répartition donnée dans le tableau I colonne 2 est reproduite dans la figure 2.

I. Répartition des débits de crue du Rhin.

Débits de crue $\times 10^2$	Nombre d'années	Fréquences classées
12,5—15	1	1
15 —17,5	3	4
17,5—20	10	14
20 —22,5	15	29
22,5—25	17	46
25 —27,5	18	64
27,5—30	21	85
30 —32,5	10	95
32,5—35	6	101
35 —37,5	6	107
37,5—40	5	112
40 —42,5	0	112
42,5—45	1	113
45 —47,5	1	114
47,5—50	1	115
50 —52,5	0	115
52,5—55	1	116
55 —57,5	2	118
	118	

Il s'agit maintenant de reproduire ces chiffres d'après la théorie de la plus grande valeur. Soit x le débit, soit n le nombre d'années observées. Alors le nombre d'années $n \mathfrak{B}(x)$ ayant un débit maximum inférieur à x , est

$$n \mathfrak{B}(x) = n e^{-e^{-y}} \quad (1)$$

où

$$y = \alpha(x - u) \quad (2)$$

est une variable réduite par l'introduction de la dominante u de la plus grande valeur et de la grandeur α , de dimension x^{-1} . Ces deux constantes α et u peuvent être déterminées, d'après les observations, à l'aide de la méthode des moments, ce qui mène aux équations

$$u = \bar{u} - \frac{0,57722}{\alpha} \quad (3)$$

et

$$\frac{1}{\alpha} = \frac{\sqrt{6}}{\pi} \sigma \quad (4)$$

où \bar{u} est la moyenne arithmétique et σ l'écart quadratique moyen de la répartition de la plus grande valeur, grandeurs qu'on tire des observations. Mais ce calcul devient fastidieux puisque le nombre $n = 118$ est assez élevé. Voilà pourquoi nous proposons une nouvelle détermi-

nation des deux constantes à l'aide de la centrale u_0 et des deux quartiles u_1 et u_2 observés.

Il s'agit donc de trouver la relation entre ces trois grandeurs, appelées valeurs de position et les deux constantes u et σ . La centrale u_0 de la distribution de la plus grande valeur est, d'après la définition de cette valeur, suivant (1) et (2) la solution de

$$e^{-e^{-\alpha(u_0-u)}} = e^{-\lg 2}$$

ce qui mène à

$$e^{-\alpha(u_0-u)} = \lg 2$$

ou

$$u_0 = u - \frac{\lg \lg 2}{\alpha}.$$

On détermine donc la constante u à l'aide de la valeur observée u_0 par :

$$u = u_0 - \frac{0,36651}{\alpha}. \quad (5)$$

Pour déterminer α , on fait usage des deux quartiles y_1 et y_2 de la variable réduite qui sont

$$y_2 = 1,24590$$

$$y_1 = -0,32663.$$

On obtient d'après (2)

$$\alpha(u_2 - u_1) = 1,57253.$$

On peut donc calculer la constante $\frac{1}{\alpha}$ d'après

$$\frac{1}{\alpha} = 0,63529 (u_2 - u_1) \quad (6)$$

et la constante u suivant (5) d'après

$$u = u_0 - 0,23307 (u_2 - u_1). \quad (7)$$

Cette détermination à l'aide des valeurs de position entraîne une certaine perte en précision quand on la compare à la détermination à l'aide des moments. Car la précision de la médiane (11) est pour la distribution de la plus grande valeur 90% de la précision de la moyenne arithmétique. Mais ce défaut est compensé de beaucoup par la facilité de la détermination.

Dans le tableau I le nombre d'observations est suffisamment grand pour qu'on puisse calculer les trois valeurs de position par la méthode ordinaire. On prend donc pour la centrale la 59^e valeur et pour le premier (second) quartile la moitié entre la 29^e et la 30^e valeur (entre la 88^e et la 89^e valeur). L'interpolation, à l'aide des fréquences classées, mène à

$$u_1 = 2250 + \frac{0,5}{17} 250 = 2257,353$$

$$u_0 = 2500 + \frac{13}{18} 250 = 2680,555.$$

$$u_2 = 3000 + \frac{3,5}{10} 250 = 3087,500.$$

D'ailleurs, si l'on emploie les données originales rangées par ordre de grandeur (figure I), on aboutit sensiblement aux mêmes valeurs. Car les différences ne sont que de 1^o/100. On peut donc se contenter de ces déterminations et l'on obtient, à l'aide de (6) et (7), les deux constantes

$$u = 2487,073 \quad (7')$$

et

$$\frac{1}{\alpha} = 527,907. \quad (6')$$

Cette détermination des constantes permet de calculer la distribution des débits de crue correspondant à la théorie de la plus grande valeur. Mais, au lieu de calculer, pour des valeurs choisies de x , la probabilité correspondante $\mathfrak{B}(x)$, nous inversons le procédé comme nous l'avons fait (9) pour la distribution du plus grand âge. Pour des valeurs choisies de $\mathfrak{B}(x)$ tirées du tableau I de notre article précédent (8), on calcule la valeur correspondante de la variable x d'après (2), (6') et (7') par

$$x = 2487,073 + 527,907y.$$

Ce procédé mène aux nombres d'années $n \mathfrak{B}(x)$ pour lesquelles le débit maximum est inférieur à x .

Pour obtenir en plus la distribution théorique comparable à la répartition observée nous calculons le nombre d'années ayant un débit maximum, contenu entre x_1 et x_2 qui sera $n(\mathfrak{B}(x_2) - \mathfrak{B}(x_1))$. Le tableau de $\mathfrak{B}(x)$ en fonction de y , donné dans notre travail précédent (8), s'échelonne d'après

$$\Delta y = \frac{1}{2}$$

c'est à dire

$$\Delta x = 263,954$$

tandis que la répartition du tableau I s'échelonne d'après $\Delta x' = 250$. Pour pouvoir comparer les nombres d'années ayant des débits maxima entre x et $x + \Delta x$ avec les valeurs observées, il faut rendre égaux les intervalles ce qui revient à la multiplication des valeurs $n \Delta \mathfrak{B}(x)$ par

$$k = \frac{250}{263,9535} \\ = 0,94714.$$

II. Distribution et fréquences théoriques des débits de crue du Rhin.

1	2	3	4	5	6
y	$y + \frac{\Delta y}{2}$	$n\mathfrak{B}(x)$	$nk\Delta \mathfrak{B}(x)$	x	$x + \frac{\Delta x}{2}$
-2,5		0,001		1167	
	-2,25		0,07		1299
-2,0		0,073		1431	
	-1,75		1,21		1563
-1,5		1,335		1695	
	-1,25		6,20		1827
-1,0		7,787		1959	
	-0,75		14,32		2091
-0,5		22,691		2223	
	-0,25		19,91		2355
0		43,410		2487	
	0,25		20,11		2619
0,5		64,340		2751	
	0,75		16,66		2883
1,0		81,680		3015	
	1,25		12,23		3147
1,5		94,401		3279	
	1,75		8,33		3411
2,0		103,064		3543	
	2,25		5,41		3675
2,5		108,700		3807	
	2,75		3,43		3939
3,0		112,269		4071	
	3,25		2,13		4203
3,5		114,490		4335	
	3,75		1,32		4467
4,0		115,858		4599	
	4,25		0,81		4731
4,5		116,696		4863	
	4,75		0,49		4995
5,0		117,208		5127	
	5,25		0,30		5259
5,5		117,519		5391	
	5,75		0,18		5523
6,0		117,709		5655	
	6,25		0,11		5786
6,5		117,823		5918	
	6,75		0,07		6050
7,0		117,893		6182	

Nous attribuons ces valeurs, $nk\Delta \mathfrak{W}(x)$ comme d'habitude, à la moitié de l'intervalle correspondant $x + \frac{\Delta x}{2}$.

Le tableau II contient dans la première colonne la variable réduite y , dans la seconde les moitiés d'intervalles $y + \frac{\Delta y}{2}$ correspondantes. La troisième colonne donne le nombre d'années $n \mathfrak{W}(x)$ ayant un débit de crue inférieur ou égal à x , grandeur reproduite dans la cinquième colonne. Les nombres $n \mathfrak{W}(x)$ sont comparables à la troisième colonne du tableau I. La quatrième colonne contient les différences $kn\Delta \mathfrak{W}(x)$, c'est à dire la distribution théorique que l'on compare à la deuxième colonne du tableau I. Les débits de crue correspondants à la première colonne du tableau I se trouvent enfin à la sixième colonne.

La figure II compare la répartition observée à la distribution théorique. L'écart entre les deux valeurs des dominantes et les irrégularités à la fin sont probablement dûs au fait que la distribution initiale des débits n'est pas simple mais se compose en réalité de plusieurs branches superposées. Mais en général, la répartition des débits de crue suit bien la distribution de la plus grande valeur. Cet accord devient évident d'après la figure I qui compare les fréquences observées, tirées du travail de Ghezzi et les valeurs théoriques correspondantes tirées du tableau II. La petite différence entre l'aire contenue dans les échelons observés et dans la courbe théorique s'explique par le fait que nous n'avons pas conservé la moyenne arithmétique mais seulement les trois valeurs de position. En général le cours des probabilités passe fort bien à travers la courbe en échelle observée. L'introduction d'hypothèses spéciales sur les débits de crue n'est point nécessaire. Ceci montre une fois de plus que le domaine d'application de la théorie de la plus grande valeur est assez vaste.

Bibliographie.

1. A. Coutagne: Etude analytique des débits de crue. Comptes Rendus du Congrès de l'Union Internationale de Géodésie et de Géophysique, Stockholm 1930. — 2. A. Coutagne: Introduction à l'étude rationnelle de l'hydrologie fluviale. Comptes Rendus du Congrès de l'Union Internationale de Géodésie et de Géophysique, Stockholm 1930. — 3. C. Ghezzi: Die Ablußverhältnisse des Rheines in Basel. Mitteilungen des Amtes für Wasserwirtschaft, Nr. 19. Berne 1926. — 4. R. Gibrat: Aménagement hydroélectrique des cours d'eau, Statistique mathématique et calcul des probabilités. Revue générale de l'électricité T. XXXII, No 15, 16, Paris 1932. — 5. H. Grassberger: Die Anwendung der Wahrscheinlichkeitsrechnung auf die Wasserführung der Gewässer. Die Wasserwirtschaft. Wien 1932, Nos 1—6. — 6. H. Grassberger: Untersuchungen über die Hochwässer des Yang-tse-kiang. Die Wasserwirtschaft, Wien 1933, No 6. — 7. H. Grassberger: Die Anwendung der Wahrscheinlichkeitsrechnung auf Hochwasserfragen. Deutsche Wasserwirtschaft, Stuttgart 1936, No 9—10. — 8. E. J. Gumbel: La plus grande valeur. Aktuárské Vědy, T. V, Nos 2, 3, 4. Prague 1935/36. — 9. E. J. Gumbel: L'âge limite.

Aktuárské Vědy, T. VI, No 1. Prague 1936. — 10. E. J. Gumbel: Les distances extrêmes entre les émissions radioactives. C. R. Ac. Sc. t. 203, p. 354, Paris 1936. — 11. E. J. Gumbel: La précision de la moyenne arithmétique et de la médiane. Aktuárské Vědy, T. VI, No 4. Prague 1937. — 12. A. Hazen: Flood flows, a study of frequencies and magnitudes, New-York 1930.

LITERATURA A ZPRÁVY.

Österreichische Sterbetafel 1930/33. (Herausgegeben vom Fachverbande der Versicherungsanstalten. Wien 1937.) Schon vor sechzig Jahren hat Lexis das Verfahren ironisiert, „dem Gesetze“ zuliebe die Beobachtungsdaten über die Sterblichkeit durch abgeänderte, nämlich ausgeglichene Daten zu ersetzen; wie würde ihm gar das in dieser Publikation befolgte (Blaschke'sche) Prinzip, daß nicht bloß die ausgeglichenen Sterbewahrscheinlichkeiten einen regelmäßigen Gang bei fortschreitendem Alter aufweisen sollen, sondern auch ihre Differenzen, was die Abänderung im allgemeinen noch vergrößert, mißfallen haben? Gleichwohl mag dem Motto „natura non facit saltum in primis differentiis, sed in secundis“ die innere Berechtigung nicht abgesprochen werden, solange die Ausgleichung keine Beobachtungsnuancen verwischt und keine übermäßige Abweichung mit sich bringt. Am besten würde dem angestrebten Zwecke gewiß eine analytische Ausgleichung dienen, aber in vorläufiger Ermanglung einer analytischen Formel für die ganze Erstreckung der Volkssterbetafel, oder selbst nur für die ganze Tafel mit Ausnahme der allerjüngsten und höchsten Alter, leistet die quasi-analytische (verallgemeinerte) King'sche Fortsetzungsmethode, wie sofort zu zeigen, entschieden besseres als die Methode des österreichischen Fachverbandes, und zwar mit viel geringerem Arbeitsaufwande.

Zunächst besitzt bei der öst. Sterbetafel 1930/33 des männlichen Geschlechtes die als Grundlage für die King'sche Methode gewählte analytische Darstellung

$$10000\bar{u}_{21+n} = 9 + 4(1,09)^n + 4,75(1,09^{-n}), \quad n \geq 0$$

daher

$$\bar{u}_{x+3} - \bar{u}_x = (1 + c + c^{-1})(\bar{u}_{x+2} - \bar{u}_{x+1}), \quad c = 1,09,$$

wo \bar{u}_x den (auf 7 Dezimalstellen zu berechnenden) ausgeglichenen Wert für $u_x = \text{colog } p_x$ und p_x die beobachtete einjährige Lebenswahrscheinlichkeit der x -Jährigen bedeutet, ungefähr denselben Annäherungsgrad und regulären Differenzengang, wie die hyperkomplizierte, mit Rücksicht auf eine nicht weniger als 20-gliedrige Umgebung von u_x bewerkstelligte „Exponential“-Ausgleichung der Publikation

$$u_x^E = b_0 u_x^{555} + b_1 u_{x\pm 1}^{555} + b_2 u_{x\pm 2}^{555} + b_3 u_{x\pm 3}^{555},$$

$$b_0 = 0,61276, \quad b_1 = 0,49571, \quad b_2 = 0,14372, \quad b_3 = -0,44581.$$

Die bekannte, wenn auch geringe Sterblichkeitssenkung, wie sie in vielen Volkssterbetafeln zu Anfang der zwanziger Lebensjahre auftritt, wird vom Verfasser der Publikation ins Gegenteil verkehrt, indem er dekretiert, diese Senkung solle nicht berücksichtigt werden, in Wirklichkeit sind voluminöse Formeln von der Art (2) nicht elastisch genug, um feine Nuancen wiederzugeben. Hingegen findet nach Formel (1) wenigstens keine Sterblichkeitssteigerung in der betreffenden Altersstrecke 21 bis 23 statt, auch hätte sich durch Wahl anderer Konstanten an Stelle derjenigen in (1) die Senkung stärker markieren lassen, aber nur bei vergrößerter Abweichung.