ON A SPECIAL DISTRIBUTION OF MAXIMUM VALUES[1]

RAYMOND SNEYERS

Institut Royal Météorologique de Belgique, Uccle, Belgium

[Manuscript received December 8, 1959]

ABSTRACT

The classical extreme-value theory does not give a good account of the distribution of maximum rainfall intensities in Belgium. Reasons are given for the use, in this case, of a probability function defined by a double exponential whose argument is a function represented by a curve with two asymptotes. The application of such a probability function, when the curve is a branch of a hyperbola, to the maximum rainfall, in 1 min., at Uccle, leads to a good fit.

1. INTRODUCTION

All problems concerned with water streaming during rainfall have generally to be solved with the knowledge of probabilities of rainfall intensities. Consequently, it is of some importance to be able to make statistical prediction of such variates with the best possible accuracy.

This paper will be more especially concerned with the monthly maximum rainfall intensities in 1 min. provided at Uccle, Belgium (Institut Royal Météorologique) by a Hellmann rain recorder.

2. EXTREME-VALUE DISTRIBUTION FUNCTIONS

The estimation of the proabilities of maximum rainfall intensities belongs obviously to extreme-value theory and therefore it should be remembered what kind of distribution functions are ordinarily best fitted to such data, functions which were introduced by Fisher and Tippett [1].

If t is the variate and if $\phi(t_o)$ gives the probability that t is less than the fixed value t_o, these distribution functions are defined by the equation:

$$\phi(t)=\exp\ [-e^{-y}]$$

where $y=ax+b$, with a and b constant and $a>0$, and with:

$$\begin{aligned}
&\text{type I: } x=t && \text{when } -\infty\leq t\leq\infty\\
&\text{type II: } x=\log t && \text{when } 0\leq t\leq\infty\\
&\text{type III: } x=\log(-t) && \text{when } -\infty\leq t\leq 0
\end{aligned}$$

There are still other forms, but they are all derived from the above types by linear transformation. Different methods of adjustment exist and all actually known have been recently reported by Gumbel [2]. The easiest one consists in adjusting by least squares $y=-\log\ (-\log F)$, where F is the observed cumulative frequency distribution, to a linear function of x.

3. MONTHLY MAXIMUM INTENSITIES IN 1 MIN., AT UCCLE

Our investigation was made on the maximum intensities observed during the period 1938–57. More precisely, the maximum rainfall intensity was determined for every month of each year of this period and the means of those monthly maximums were calculated for each month. Frequency distributions were then established with the use of class intervals having a width of one-fifth of the mean monthly maximum.

The 12 samples obtained in this way looked very similar and suggested the assumption of an identical theoretical

TABLE 1.—*Monthly maximum rainfall intensities in 1 min., at Uccle, Belgium. t is given in fifths of the mean of the monthly maximum intensity*

t	Observed and adjusted frequency distributions						
	$F(t)$	$x=\log_{10}t$ 10^{-2}	y 10^{-2}	y' 10^{-2}	$\phi(t)=\exp[-e^{-y'}]$	$z=y-y'$ 10^{-2}	z' 10^{-2}
1.5	0.0672	18	−99	−182	0.002	83	83
2.5	.2227	40	−41	−062	.156	21	24
3.5	.4076	54	11	014	.419	−3	−6
4.5	.5373	65	48	074	.621	−26	−24
5.5	.6597	74	88	122	.744	−34	−33
6.5	.7437	81	122	160	.817	−38	−37
7.5	.8403	88	175	198	.871	−23	−39
8.5	.8613	93	190	226	.901	−36	−38
9.5	.8866	98	212	253	.923	−41	−35
10.5	.9202	102	249	274	.937	−25	−31
11.5	.9412	106	280	296	.950	−16	−26
12.5	.9580	110	315	318	.959	−3	−20
13.5	.9622	113	326	334	.965	−8	−14
14.5	.9706	116	351	351	.970	0	−8
15.5	.9790	119	385	367	.975	18	−1
18.3	.9832	126	408	405	.983	3	20
19.6	.9874	129	437	421	.9853	16	31
21.4	.9916	133	478	443	.9882	35	45
24.0	.9958	138	547	470	.9909	77	67

[1] This paper was presented and discussed at a Special Climatological Seminar, U.S. Weather Bureau, on April 16, 1959.

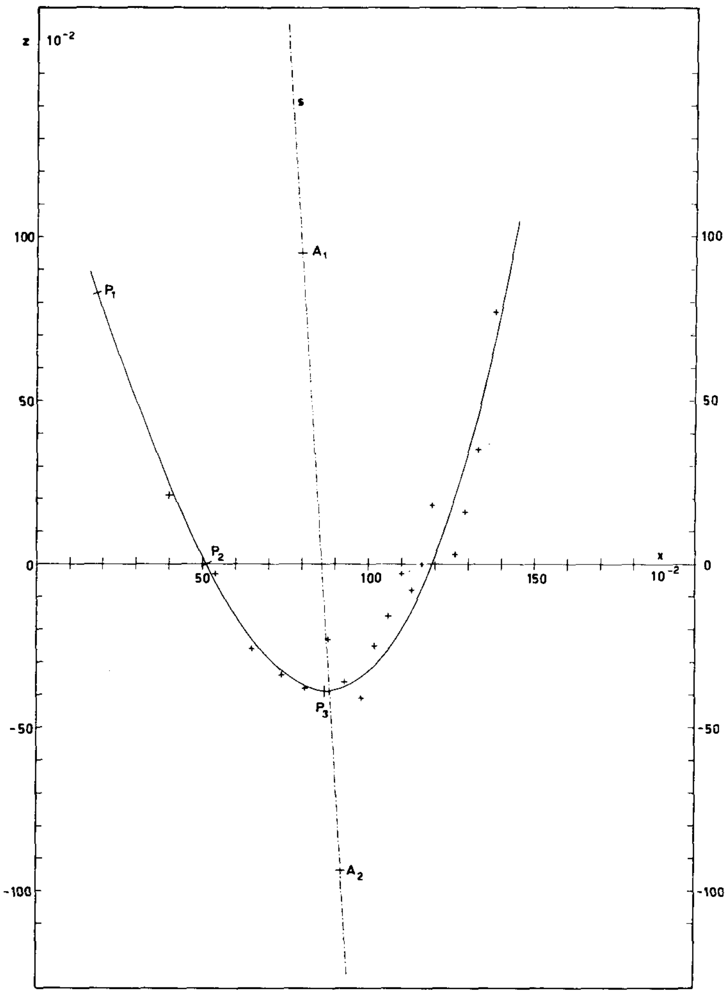

FIGURE 1.—Variation of z as a function of x. Adjusted curve and observed values.

distribution. This assumption was tested and found acceptable and therefore all data were grouped in one sample.

The cumulated frequency distribution $F(t)$ derived in that manner, t being given in fifths of the mean monthly maximum, appears in table 1. It was obtained by dividing the cumulated number of occurrences by $(n+1)$, n being the size of the total sample (here $n=237$, because no maximum rainfall intensity was measured during three of the 240 months of the considered period).

The intensities being essentially non-negative, it is clear that the adjusted function which has to be tried is a Fisher-Tippett type II distribution. Therefore we have to take $x=\log t$, or more conveniently: $x=\log_{10} t$, and to adjust $y=-\log(-\log F)$ to x.

The values of y were obtained from the Probability Tables [3] and the equation:

$$(1) \qquad y'=5.43\,x-279.36$$

was found with adjusting y to x by the ordinary method of least squares. The values of x, y, y', and $\phi(t)=\exp[-e^{-y'}]$ are also given in table 1.

The goodness of fit was tested with the Kolmogorov-Smirnov test [4] and the χ^2 test. For the first one, the differences $|F(t)-\phi(t)|$ were calculated and the largest difference compared with its critical value d. In this case the largest difference is 0.084 and, at the 0.05 level, $d=1.36/\sqrt{237}\simeq 0.088$. Since we are concerned with grouped data and an adjusted theoretical distribution, the largest difference seems to be too near its critical value to be considered as non-significant. Moreover, the χ^2 test leads to a value of χ^2 which is significant at a level smaller than 5×10^{-4}. The adjustment has thus to be rejected.

4. ADJUSTMENT OF y TO x WITH THE USE OF A QUADRATIC RELATIONSHIP

The reasons for this rejection are apparently related to a systematic variation of y with respect to y', a variation which is made evident by plotting the differences $z=y-y'$ against x (see fig. 1). In addition, this graphical representation suggests an asymptotic linear variation for small and for large values of x, variations which, in the simplest case, might be represented by a branch of a hyperbola. Such an asymptotic behavior of y had however to be expected here. It has, in fact, to be remembered that in Belgium, maximum intensities are provided by two kinds of rains, the first kind being the continuous rains falling during the passage of cyclones, and the second one being the showers accompanying certain polar air invasions as well as thunderstorms. In terms of probabilities, this means that the observed maximum is the largest between two extreme values, each of them being issued from a different population. Therefore, if $F_1(t)$ and $F_2(t)$ are the cumulative distribution functions of each population of maximums, it is clear that $\phi(t)=F_1(t)\times F_2(t)$ will be the cumulative distribution function of the largest of the two maximums. If, moreover, large values of the first population are small values of the second one, it may be expected that for the largest values of t, $\phi(t)$ will vary like $F_2(t)$, since for such values of t, $F_1(t)$ is very near to 1, and that for the smallest values of t, $\phi(t)$ will vary like $F_1(t)$, since $F_2(t)$ remains then very near to zero.

Now if we choose a branch of a hyperbola to represent such a variation of y, the adjustment has to be made with an equation of the following type:

$$(2) \qquad y=ax+b+\epsilon\sqrt{c^2(x-d)^2+e^2},\ \text{with } \epsilon=\pm 1,$$

where $a>|c|$, since the derivative dy/dx has to remain positive for any x, and where ϵ equals $+1$ or -1, according as the curvature of the curve is set toward the positive or the negative values of y.

In our case, the adjustment was performed on the differences z, and graphical estimation was preferred to a least squares procedure because the last method does not take in account the fact that all the values of z do not have the same precision. Therefore a first sketch of the

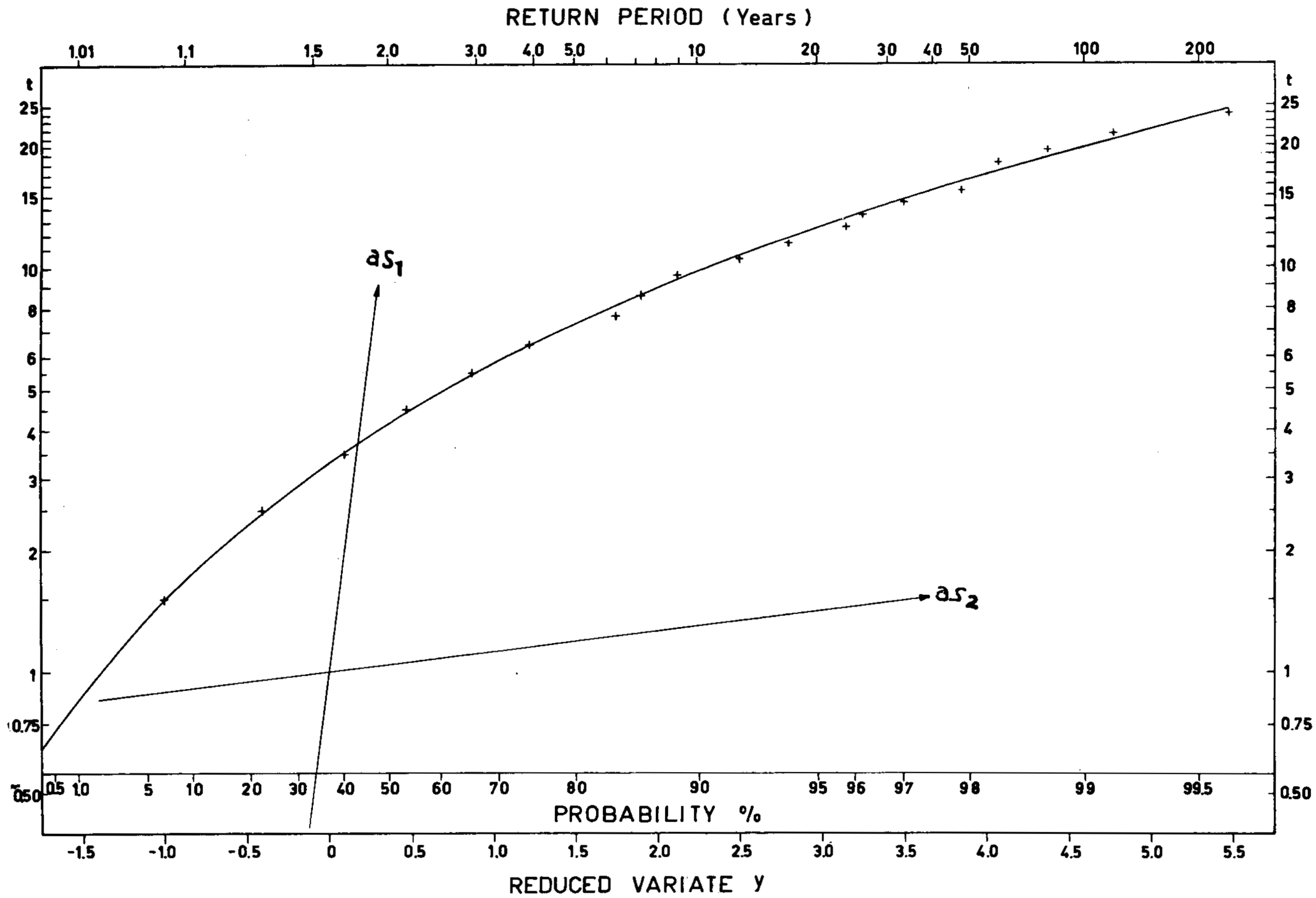

FIGURE 2.—Distribution of the monthly maximum rainfall intensities in 1 min., at Uccle. Adjusted curve and observed values. t is given in fifths of the mean monthly maximum.

curve was drawn and a graphical estimation of its axis of symmetry s was made. Using then the points $A_1(x=80, z=95)$ and $A_2(x=92, z=-94)$, the equation of this axis was found to be:

(3) $$z+15.75\,x-1355=0.$$

With this result the equation of the hyperbola may be written in the form:

(4) $$(z+15.75x-1355)^2-\alpha^2(z-0.0635x+\beta)^2=\gamma$$

where α, β, and γ are constants; note that $(15.75)^{-1}=0.0635$.

Finally, taking in account that the points $P_1(x=18, z=83)$, $P_2(x=51, z=0)$, and $P_3(x=87, z=-39)$ are on the curve, it is found that $\alpha^2=4.3031$, hence:

(5) $$\alpha=2.074;\ \beta=879.20;\ \text{and}\ \gamma=-2{,}997{,}366.$$

Whence:

(6) $$z=4.85x-1{,}555.8+\sqrt{98.6(x-141.3)^2+908{,}017}$$

or, with (1):

(7) $$y=10.28x-1835.2+\sqrt{98.6(x-141.3)^2+908{,}017}$$

Since 10.28 is larger than $\sqrt{98.6}$, the branch of hyperbola defined by (7) may be accepted.

The values of z' calculated with (6) are given in table 1 too, while both the distribution $\phi_1(t)$ defined by (7) and the observed distribution have been drawn in figure 2; this time, the comparison indicates a very good fit, which is confirmed by a value of χ^2 for which $P>0.20$.

In figure 2, the asymptotic directions as_1 and as_2 have also been drawn; they were derived from the equation:

$$y-(10.28\pm\sqrt{98.6})x=0.$$

5. FINAL REMARKS AND CONCLUSION

As was mentioned above, a relatively large sample was obtained by grouping in one sample the different samples corresponding to each month. Although the assumption permitting such a grouping was found to be acceptable, it is not uninteresting to make a last comparison in order to verify if this assumption is quite

TABLE 2.—*Differences between the theoretical distribution $\phi_1(t)$ and the observed distributfons by seasons F_{I-III}, F_{IV-VI}, F_{VII-IX}, and F_{X-XII}. t is given in fifths of the mean monthly maximum intensity.*

t	$\phi_1(t)$	ΔF_{I-III}	ΔF_{IV-VI}	ΔF_{VII-IX}	ΔF_{X-XII}
1.5	0.067	0.054	−.034	−.034	0.015
2.5	.232	.096	−.084	−.068	.014
3.5	.397	.120	−.020	−.085	.013
4.5	.545	.058	−.037	−.086	.013
5.5	.663	.027	.026	−.089	−.007
6.5	.747	−.006	.024	−.042	−.025
7.5	.816	.029	.004	.053	−.029
8.5	.859	.003	−.023	.027	−.039
9.5	.893		−.007	.009	−.040
10.5	.916	−.037	−.014	.019	.002
11.5	.935	−.021	−.017	.016	.000
12.5	.950		.001		.018
13.5	.960		.008		
14.5	.968	−.020			
15.5	.975	−.010		−.007	
18.3	.986				−.002
19.6	.989		−.005		
21.4	.992			−.008	
24.0	.995	−.012			

justified. This comparison was made by grouping the data in four samples corresponding respectively to the quarters January–March (I–III), April–June (IV–VI), July–September (VII–IX), and October–December (X–XII), and by calculating the differences of the observed frequency distributions for each of these periods from the theoretical distribution $\phi_1(t)$ defined by (7).

The results are given in table 2. They show that the fit is best for the quarter October–December, and less good for the other ones with, in particular, higher probabilities for small values of t during the period January–March and, on the contrary, lower probabilities for such values during the period April–September. A better fit is thus to be expected if each group is treated separately.

However, for high values of t, it should be noted that the fit is very good in the four cases. This last statement may be illustrated by the following feature:

It was formerly admitted that the intensity in 1 min. at Uccle might be considered as a maximum which would never occur. In reality, the probability of having in the year an intensity less than 5 mm. in 1 min., computed from the probabilities of such an intensity occurring during any month of the year, leads with the use of $\phi_1(t)$ to a probability of 0.9665. With other words, an intensity of at least 5 mm. in 1 min. has a return period of about 30 years.

Another estimation of this return period has been made by adjusting a Fisher-Tippett type I distribution to the sample of 20 yearly maximums observed during the period 1938–1957. More precisely, the procedure described by Gumbel [2] p. 226, leads to the relation:

$$x = 10.07y + 14.78 \tag{8}$$

where x is given in tenths of millimeters which, for $x=50$, gives $y=3.498$, namely a probability of 0.9702 and a return period of 33.6 years, both in very good agreement with our first estimation.

Furthermore, the secular maximum estimated in the same manner was found to be 6.17 mm. by the first method (practically in July), while $y=4.60$ in (8) leads to $x=6.11$ mm. The agreement is again excellent.

To conclude, the favorable results obtained by the very simple computations described above advocate the use of the considered double exponential each time that extreme values may come from at least two sufficiently different populations.

ACKNOWLEDGMENTS

The writer wishes to express his gratefulness to Dr. H. E. Landsberg, Director, and to Mr. H. C. S. Thom, Chief Climatologist, Office of Climatology, U.S. Weather Bureau, who kindly allowed and organized the Special Climatological Seminar at the Office of Climatology, where this paper was presented and discussed.

REFERENCES

1. R. A. Fisher and L. H. C. Tippett, "Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample," *Proceedings, Cambridge Philosophical Society*, vol. 24, 1928, pp. 180–190.
2. E. J. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, 1958.
3. U.S. National Bureau of Standards, "Probability Tables for the Analysis of Extreme-Value Data," *Applied Mathematics Series* No. 22, 1953.
4. F. J. Massey, "The Kolmogorov—Smirnov Test for Goodness of Fit," *Journal, American Statistical Association*, vol. 46, 1951, pp. 68–78.