DISASTER PREVENTION RESEARCH INSTITUTE
KYOTO UNIVERSITY
BULLETINS

Bulletin No. 66 March, 1964

Application of Extreme Value Distribution in Hydrologic Frequency Analysis

By

Mutsumi KADOYA

Contents

Application of Extreme Value Distribution in Hydrologic Frequency Analysis

By

Mutsumi KADOYA

Synopsis

It is well known that there are three types of asymptotes for the distribution of the extremes or largest values, which are expressed as follows:

$F(y) = \exp(-e^{-y})$;

$y = a(x-u)$, for the first asymtote or the Gumbel's distribution,

$y = a \log(x+b)/(u+b)$, for the second asymptote or the type A of log-extreme value distribution,

$y = a \log(g-u)/(g-x)$ for the third asymptote or the type B of log-extreme value distribution,

in which a, u, b and g are population parameters.

Several problems have remained unsolved in practical analysis of hydrologic frequency by the use of these asymptotes.

(I) In Prat I, first of all, the statistical characters of the three asymptotic distributions are discussed theoretically and it is shown that they should be applicable in limited range of the value of coefficient of skew, C_s, that is

$$C_s \left.\begin{matrix} > \\ = \\ < \end{matrix}\right\} 1.1395\cdots; \quad \text{for} \left\{\begin{matrix} \text{the second asymptote,} \\ \text{the first asymptote,} \\ \text{the third asymptote.} \end{matrix}\right.$$

Next, although methods of estimation of the parameters included in the asymptotic equations have been proposed by Gumbel and others by the help of method of moment, the results obtained by such methods seem to be not so good in fitness to hydrologic data. Then, a method of estimation based on the concept of plotting value instead of plotting position, proposed by the author is succesfully developed for the first and the second asymptotes from a view point of practical application.

(II) Generally, very large or small data are to be contained in a sample,

which is called the singular value. In estimating the population parameters of asymptotes, the rejection test of such data is essential in the sense of stochastics. Moreover, evaluation of the singular value is important in the sense of engineering.

In Part II, first, applying the concept of two-sample theory on normals the method of evaluation of a singular value is proposed. Next, on the basis of the binomial distribution, the criterion for rejection of singular data is defined.

Part I. Extreme (Largest) Value Distribution and Method of Fitting

I. Introduction

The extreme value distributions are defined generally as asymptotic forms of the distribution for the largest or smallest value in a sample. The practical methods of application of these distributions to various engineering problems have been considered by several investigators. But the introduction of statistics in this field to hydrologic forecasting seems to owe much to Gumbel, who showed the usefulness of the first asymptote of the distribution for largest value to the frequency analysis of floods in 1941[1]. Moreover, he showed that the third asymptote of the distribution for smallest value is successfully applicable to the frequency analysis of droughts in 1954[2].

In estimating the parameters included in these asymptotes, the classical method of moment and the method based on the concept of plotting position were adopted by Gumbel[1)~3)]. As methods of estimation of the parameters in such asymptotes, besides the above methods, there are useful ones proposed by Thom[4] (1954), Lieblein[5] (1954) and Jenkinson[6] (1955).

Since 1952, the statistics for the largest values of the hydrologic data have been studied by the author, who investigated (1) the statistical properties of three asymptotes for the largest value (1955)[7], (2) the method of estimation of their parameters by the classical method of moment (1955)[7], and (3) the one based on the concept of plotting position (1953, 1954)[8,9]. As a result, it was clear that the first and the second asymptotes were available to the frequency analysis of the hydrologic amount. But the results of estimation of their theoretical distribution by the classical method of moment did not prove very good in fitness to hydrologic data.

Afterwards, a reasonable method[10,11] of fitting based on the concept of plotting value, which was proposed by the author[12] instead of plotting position, was developed for these asymptotes.

In this part, the statistical properties of the extreme (largest) value distributions and the method of estimation of their parameters and so on, obtained by these studies are summarily presented.

2. Extreme (largest) value distributions and their fundamental properties

It is well known that there are three types of the asymptote for the distributions of the extreme (largest) value in a sample, which are practically expressed as follows :

$$F(y)=\exp(-e^{-y})\ ; \tag{1.1}$$

$$1\text{ st }; \ y=a(x-u), \qquad -\infty<x<\infty \tag{1.2}$$

$$2\text{ nd }; \ y=a\log(x+b)/(u+b)=k\,\mathrm{lg}\,(x+b)/(u+b), \qquad -b<x<\infty \tag{1.3}$$

$$3\text{ rd }; \ y=a\log(g-u)/(g-x)=k\,\mathrm{lg}\,(g-u)/(g-x), \qquad -\infty<x<g \tag{1.4}$$

In the above equations, $k \equiv a\log e=0.4343a$, u, b and g are population parameters, y is the reduced variate of actual extreme x and is called the reduced extreme, $F(y)$ is the asymptotic probability in which the extreme variate will not exceed a certain fixed variate, and log and lg stand for the common and natural logarithms, respectively.

In the field of hydrologic statistics in Japan, the first asymptotic distribution is usually called as the Gumbel's distribution in honour of his pioneering and fruitful work, and the second and third asymptotic distributions are called the type A and B of logarithmic extreme value distributions, respectively, since the author's proposal in 1955[7], these three asymptotes are generally called the extreme (largest) value distribution.

Since the mathematical or statistical properties of these asymptotic distributions have been studied by a number of investigators and are discussed in detail in the masterpiece "Statistics of Extremes" by Gumbel[13], it is not necessary to discuss them again here. But the following properties, among which several unpublished ones are included, should be noticed.

(1) **The first asymptote** : There are simple relations between the

population moments $\nu_i(x-u)$ and $\nu_i(y)$ about origin of order i, and between the population central moments $\mu_i(x)$ and $\mu_i(y)$ of order i, that is

$$\left.\begin{aligned} \nu_i(x-u) &= \nu_i(y)/a^i \\ \mu_i(x) \quad &= \mu_i(y)/a^i \end{aligned}\right\} \tag{1.5}$$

The population moments $\nu_i(y)$ and $\mu_i(y)$ are easily calculated by using the moment generating function $Q(t)$ and the semi-Invariant as follows[14]:

$$\begin{aligned} \lg Q(t) &= \lg \Gamma(1-t), \qquad |t|<1 \\ &= \gamma t + \sum_{r=2}^{\infty} S(r) t^r / r \end{aligned}$$

where

$$\gamma = 0.5772\cdots \quad \text{is the Euler's constant,}$$
$$S(r) = \lim_{n\to\infty} (1+2^{-r}+3^{-r}+\cdots+n^{-r}), \qquad r \geqq 2$$

Therefore, the mean m_y and m_x, the variance $\sigma_y{}^2$ and $\sigma_x{}^2$, and the other moment $\mu_i(y)$ and $\mu_i(x)$ are expressed, respectively, as follows:

$$\left.\begin{aligned} &m_y = \nu_1(y) = \gamma, && m_x = u + \gamma/a \\ &\sigma_y{}^2 = \mu_2(y) = S(2) = \pi^2/6, && \sigma_x{}^2 = S(2)/a^2 = \pi^2/6a^2 \\ &\mu_i(y) = 2S(3), && \mu_3(x) = 2S(3)/a^3 \end{aligned}\right\} \tag{1.6}$$

It should be noticed that the coefficient of skew of the extremes themselves, equal to that of the reduced extremes, becomes

$$C_s = \mu_3/\mu_2{}^{3/2} = 2S(3)/S(2)^{3/2} = 1.1395\cdots \tag{1.7}$$

(2) **The second asymptote**: The population moment $\nu_i(x+b)$ about origin of order i is expressed by

$$\nu_i(x+b) = \int_{-b}^{\infty} (x+b)^i dF(x)$$

After several calculations,

$$\nu_i(x+b) = (u+b)^i \Gamma(1-i/k) \tag{1.8}$$

The mean and the variance and the other moments are easily obtained by putting $i=1, 2, 3, \cdots$ in above equation. And the coefficient of skew C_s becomes

$$C_s = \frac{\mu_3}{\mu_2{}^{3/2}} = \frac{\Gamma(1-3/k) - 3\Gamma(1-2/k)\Gamma(1-1/k) + 2\Gamma^3(1-1/k)}{[\Gamma(1-2/k) - \Gamma^2(1-1/k)]^{3/2}} \tag{1.9}$$

It becomes clear after some examinations that

$$\frac{dC_s}{d(1/k)} > 0$$

and

$$\lim_{1/k \to 0} C_s = 2S(3)/S(2)^{3/2} = 1.1395\cdots$$

That is, the second asymptote is the extremely skewed distribution as its coefficient of skew C_s is greater than $1.1395\cdots$.

(3) **The third asymptote**: The population characters of the third asymptote can also be easily examined as well as is done for the second asymptote, and the following relations are obtained,

$$\nu_i(g-x) = (g-u)^i \Gamma(1+i/k) \tag{1.10}$$

$$C_s = -\frac{\Gamma(1+3/k) - 3\Gamma(1+2/k)\Gamma(1+1/k) + 2\Gamma^3(1+1/k)}{[\Gamma(1+2/k) - \Gamma^2(1+1/k)]^{3/2}} \tag{1.11}$$

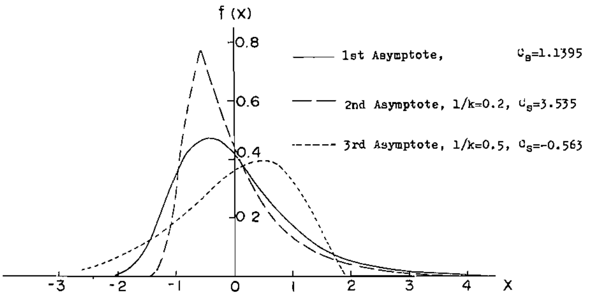

Fig. 1.1 Shapes of asymptotis distribution, provided $m_x=0$ and $\sigma_x=1$.

Moreover, it can be made clear that the coefficient of skew C_s of this asymptotic distribution is less than $1.1395\cdots$.

From the facts mentioned above, the conclusion is obtained that the three asymptotic distributions of the largest value are characterized by the coefficient of skew, or that they should be applicable in limited range of the value of coefficient of skew C_s, that is

$$C_s \begin{array}{c} > \\ = \\ > \end{array} \Bigg\} \; 1.1395\cdots; \quad \left\{ \begin{array}{l} \text{for the second asyptote,} \\ \text{for the first one,} \\ \text{for the third one.} \end{array} \right.$$

Figures 1.1 and 1.2 show this relation.

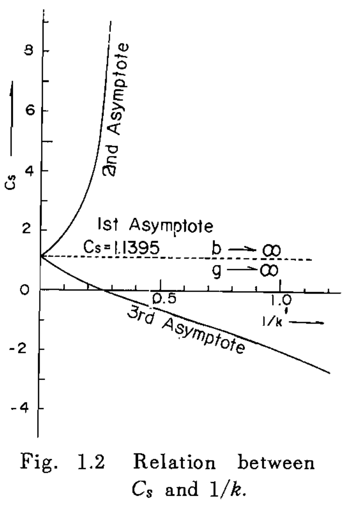

Fig. 1.2 Relation between C_s and $1/k$.

3. Estimation of population parameters by method of moment

A method of estimation of the population parameters included in these asymptotes by a method of moment is easily derived from the results in the preceding section.

(1) **The first asymptote**: Using Eq. (1.6), two parameters a and u, which are scale and location parameters of the first asymptotic distribution, respectively, can be obtained by

$$\left.\begin{aligned} a &= \sigma_y/\sigma_x \\ &= \pi/\sqrt{6}\,\sigma_x \doteqdot 1/0.7797\sigma_x \\ u &= m_x - m_y/a \\ &= m_x - \gamma/a \doteqdot m_x - 0.4500\sigma_x \end{aligned}\right\} \tag{1.12}$$

A so-called classical method of moment has been adopted by Gumbel in his earliest work[1], in which the population values σ_x and m_x in Eq. (1.12) are directly replaced by the sample values s_x and $\bar{x}$, respectively.

(2) **The second asymptote**: Three parameters $1/k$, b and u, which are skewness, location and scale parameters of the second asymptotic distribution, respectively, can lead to the following expressions after several calculations based on Eq. (1.8),

$$C_s = \frac{\Gamma(1-3/k) - 3\Gamma(1-2/k)\Gamma(1-1/k) + 2\Gamma^3(1-1/k)}{[\Gamma(1-2/k) - \Gamma^2(1-1/k)]^{3/2}} \tag{1.9$'$}$$

$$\left.\begin{aligned} &1/a = 0.4343/k \\ &b = A_1\sigma_x - m_x \equiv C_1\sigma_x - u \\ &u = m_x - B_1\sigma_x \\ &u + b = C_1\sigma_x \end{aligned}\right\} \tag{1.13}$$

where

$$\left.\begin{aligned} &A_1 \equiv \Gamma(1-1/k)/[\Gamma(1-2/k) - \Gamma^2(1-1/k)]^{1/2} \\ &B_1 \equiv [\Gamma(1-1/k) - 1]/[\Gamma(1-2/k) - \Gamma^2(1-1/k)]^{1/2} \\ &C_1 \equiv A_1 - B_1 \equiv 1/[\Gamma(1-2/k) - \Gamma^2(1-1/k)]^{1/2} \end{aligned}\right\} \tag{1.14}$$

In the above equations, it will be noticed that the values of C_s, A_1, B_1 and C_1 depend only upon the value of parameter $1/k$. In Table 1.1, those values as a function of $1/k$ are shown for the practical facilities, which are originally prepared with six decimal places[7]. If an adequate method of estimation of the population values C_s, σ_x and m_x from the sample values is found out, the parameters may by easily estimated from Eq. (1.13) by using Table 1.1.

(3) **The third asymptote**: Three parameters $1/k$, g and u, which are skewness, location and scale parameters of the third asymptotic distribution, respectively, are also obtained from Eq. (1.10), as follows:

Table 1.1 Population values of C_s, A_1, B_1 and C_1 for $1/k$ in second asymptote.

$b=A_1\sigma_x-m_x,\quad x_0=m_x-B_1\sigma_x,\quad x_0+b=C_1\sigma_x$

$1/k$	C_s	A_1	B_1	C_1
0.001	1.1455	779.13	0.450 2	778.68
2	1515	389.28	4	388.83
3	1576	259.33	6	258.88
4	1636	194.35	8	193.90
5	1697	155.37	9	154.92
0.006	1.1758	129.38	0.451 1	128.92
7	1819	110.81	3	110.36
8	1881	96.89	5	96.44
9	1943	86.06	6	85.61
10	2005	77.39	8	76.94
0.011	1.2067	70.30	0.452 0	69.85
2	2130	64.40	1	63.94
3	2193	59.40	3	58.95
4	2256	55.11	5	54.66
5	2319	51.40	6	50.95
0.016	1.2383	48.15	0.452 8	47.70
7	2447	45.28	9	44.83
8	2512	42.74	0.453 1	42.28
9	2576	40.46	2	40.00
20	2641	38.40	4	37.95
0.021	1.2706	36.54	0.453 5	36.09
2	2772	34.86	7	34.40
3	2837	33.32	8	32.86
4	2904	31.90	0.454 0	31.45
5	2970	30.60	1	30.15
0.026	1.3037	29.40	0.454 3	28.95
7	3604	28.29	4	27.84
8	3171	27.26	6	26.80
9	3239	26.30	7	25.84
30	3307	25.40	8	24.95

$1/k$	C_s	A_1	B_1	C_1
0.030	1.3307	25.40	0.454 8	24.95
1	3375	24.56	0.455 0	24.11
2	3444	23.78	1	23.32
3	3513	23.04	2	22.58
4	3582	22.34	4	21.89
5	3652	21.69	5	21.23
0.036	1.3721	21.07	0.455 6	20.61
7	3792	20.48	7	20.02
8	3862	19.92	9	19.47
9	3933	19.40	0.456 0	18.94
40	4005	18.90	1	18.44
0.041	1.4077	18.42	0.456 2	17.97
2	4149	17.97	3	17.51
3	4221	17.54	4	17.08
4	4294	17.12	6	16.67
5	4367	16.73	7	16.27
0.046	1.4441	16.35	0.456 8	15.89
7	4515	15.99	9	15.53
8	4589	15.64	0.457 0	15.19
9	4664	15.31	1	14.85
50	4739	14.99	2	14.54
0.051	1.4814	14.6867	0.457 3	14.2294
2	4890	3921	4	13.9347
3	4967	1085	5	6510
4	5043	13.8354	6	3778
5	5120	5722	7	1145
0.056	1.5198	13.3184	0.457 8	12.8606
7	5276	0735	9	6156
8	5354	12.8370	0.458 0	3790
9	5433	6085	0	1505
60	5512	3876	1	11.9295

$1/k$	C_s	A_1	B_1	C_1
0.060	1.5512	12.3876	0.458 1	11.9295
1	5592	1739	2	7157
2	5672	11.9671	3	5088
3	5753	7668	4	3084
4	5834	5728	4	1144
5	5916	3847	5	10.9262
0.066	1.5997	11.2022	0.458 6	10.7436
7	6080	0253	7	5666
8	6163	10.8534	7	3947
9	6246	6866	8	2278
70	6330	5245	9	0656
0.071	1.6415	10.3669	0.458 9	9.9080
2	6499	2137	0.459 0	7547
3	6585	0647	1	6056
4	6671	9.9197	1	4606
5	6757	7785	2	3193
0.076	1.6844	9.6411	0.459 2	9.1819
7	6932	5072	3	0479
8	7020	3766	3	8.9173
9	7108	2494	4	7900
80	7197	1254	4	6660
0.081	1.7287	9.0044	0.459 5	8.5449
2	7377	8.8863	5	4268
3	7468	7710	6	3114
4	7559	6585	6	1989
5	7651	5486	6	0890
0.086	1.7743	8.4413	0.459 7	7.9816
7	7836	3364	7	8767
8	7930	2338	7	7741
9	8024	1336	8	6738
90	8119	0355	8	5757

$1/k$	C_s	A_1	B_1	C_1
0.090	1.8119	8.0355	0.459 8	7.5757
1	8215	7.9396	8	4798
2	8311	8458	8	3860
3	8408	7539	9	2940
4	8505	6640	9	2041
5	8603	5760	9	1161
0.096	1.8702	7.4898	0.459 9	7.0299
7	8801	4054	9	6.9455
8	8901	3227	9	8628
9	9002	2416	9	7817
100	9103	1621	0.460 0	7021
0.101	1.9205	7.0842	0.460 0	6.6242
2	9308	0078	0	5478
3	9412	6.9328	0	4728
4	9516	8593	0	3993
5	9621	7872	0	3272
0.106	1.9727	6.7164	0.460 0	6.2564
7	9833	6469	0	1869
8	9941	5788	0.459 9	1189
9	2.0049	5118	9	0519
10	0158	4461	9	5.9862
0.111	2.0267	6.3815	0.459 9	5.9216
2	0378	3180	9	8581
3	0489	2557	9	7958
4	0601	1944	8	7346
5	0714	1342	8	6744
0.116	2.0828	6.0750	0.459 8	5.6152
7	0943	0168	8	5570
8	1058	5.9596	7	4999
9	1175	9033	7	4436
20	1292	8480	7	3883

$1/k$	C_s	A_1	B_1	C_1
0.120	2.1292	5.8480	0.459 7	5.3883
1	1410	7935	6	3339
2	1529	7400	6	2804
3	1649	6873	5	2278
4	1771	6354	5	1759
5	1893	5843	5	1248
0.126	2.2016	5.5341	0.459 4	5.0747
7	2140	4846	4	0252
8	2265	4359	3	4.9766
9	2391	3879	2	9287
30	2518	3406	2	8814
0.131	2.2646	5.2941	0.459 1	4.8350
2	2775	2482	1	7891
3	2905	2030	0	7440
4	3037	1585	0.458 9	6996
5	3169	1146	9	6557
0.136	2.3303	5.0714	0.458 8	4.6126
7	3438	0288	7	5701
8	3574	4.9868	6	5282
9	3711	9453	6	4867
40	3849	9045	5	4460
0.141	2.3989	4.8642	0.458 4	4.4058
2	4129	8245	3	3662
3	4271	7853	2	3271
4	4415	7467	1	2886
5	4559	7085	0	2505
0.146	2.4705	4.6709	0.457 9	4.2130
7	4852	6338	9	1759
8	5001	5971	8	1393
9	5151	5610	7	1033
50	5303	5253	6	0677

1/k	C_s	A_1	B_1	C_1
0.150	2.5303	4.5253	0.457 6	4.0677
1	5455	4901	5	0326
2	5610	4553	3	3.9980
3	5765	4210	2	9638
4	5923	3871	1	9300
5	6081	3536	0	8966
0.156	2.6142	4.3206	0.456 9	3.8637
7	6404	2879	8	8311
8	6567	2557	6	7991
9	6732	2238	5	7673
60	6899	1923	4	7359
0.161	2.7068	4.1613	0.456 3	3.7050
2	7238	1305	1	6744
3	7410	1002	0	6442
4	7583	0702	0.455 8	6144
5	7758	0406	7	5849
0.166	2.7936	4.0113	0.455 6	3.5557
7	8116	3.9823	4	5269
8	8298	9537	3	4984
9	8480	9254	1	4703
70	8665	8974	0	4424
0.171	2.8852	3.8697	0.454 8	3.4149
2	9041	8424	6	3878
3	9232	8153	5	3608
4	9426	7886	3	3343
5	9621	7622	1	3081
0.176	2.9819	3.7360	0.454 0	3.2820
7	3.0019	7101	0.453 8	2563
8	0221	6845	6	2309
9	0425	6592	4	2058
80	0632	6341	3	1808

$1/k$	C_s	A_1	B_1	C_1
0.180	3.0632	3.6341	0.453 3	3.1808
1	0842	6093	1	1562
2	1053	5848	0.452 9	1319
3	1268	5605	7	1078
4	1485	5365	5	0840
5	1704	5127	3	0604
0.186	3.1926	3.4892	0.452 1	3.0371
7	2151	4659	0.451 9	0140
8	2379	4428	7	2.9911
9	2609	4200	5	9685
90	2843	3974	3	9461
0.191	3.3079	3.3750	0.451 1	2.9239
2	3319	3529	0.450 9	9020
3	3561	3310	7	8803
4	3807	3092	4	8588
5	4056	2877	2	8375
0.196	3.4308	3.2664	0.450 0	2.8164
7	4563	2453	0.449 8	7955
8	4822	2245	5	7750
9	5085	2038	3	7545
200	5351	1833	0	7343
0.201	3.5620	3.1630	0.448 8	2.7142
2	5894	1429	6	6943
3	6171	1229	3	6746
4	6453	1032	1	6551
5	6738	0836	0.447 8	6358
0.206	3.7027	3.0642	0.447 6	2.6166
7	7321	0450	3	5977
8	7619	0260	0	5790
9	7921	0071	0.446 8	5603
10	8228	2.9884	5	5419

$1/k$	C_s	A_1	B_1	C_1
0.210	3.8228	2.9884	0.446 5	2.5419
1	8539	9699	2	5237
2	8855	9515	0	5055
3	9176	9333	0.445 7	4876
4	9503	9153	4	4699
5	9834	8974	1	4523
0.216	4.0170	2.8797	0.444 8	2.4349
7	0511	8621	5	4176
8	0858	8446	3	4003
9	1211	8273	0	3833
20	1570	8102	0.443 7	3665
0.221	4.1935	2.7932	0.443 4	2.3498
2	2305	7764	1	3333
3	2682	7596	0.442 8	3168
4	3066	7431	4	3007
5	3456	7266	1	2845
0.226	4.3853	2.7103	0.441 8	2.2685
7	4257	6941	5	2526
8	4668	6781	2	2369
9	5086	6622	0.440 8	2214
30	5512	6464	5	2059
0.231	4.5946	2.6307	0.440 2	2.1905
2	6388	6152	0.439 8	1754
3	6839	5998	5	1603
4	7298	5845	1	1454
5	7765	5693	0.438 8	1305
0.236	4.8242	2.5542	0.438 5	2.1157
7	8728	5393	1	1012
8	9223	5244	0.437 7	0867
9	9729	5097	4	0723
40	5.0245	4951	0	0581

$1/k$	C_s	A_1	B_1	C_1
0.240	5.0245	2.4951	0.43 70	2.0581
1	0772	4806	66	0440
2	1310	4662	63	0299
3	1859	4519	59	0160
4	2420	4377	55	0022
5	2993	4237	51	1.9886
0.246	5.3578	2.4097	0.43 48	1.9749
7	4176	3958	44	9614
8	4787	3820	40	9480
9	5412	3684	36	9348
50	6051	3548	32	9216
0.251	5.6706	2.3413	0.43 28	1.9085
2	7376	3279	24	8955
3	8062	3146	20	8826
4	8765	3014	15	8699
5	9485	2884	11	8573
0.256	6.0223	2.2753	0.43 07	1.8446
7	0979	2624	03	8321
8	1756	2496	0.42 98	8196
9	2552	2368	84	8074
60	3369	2242	90	7952
0.261	6.4208	2.2116	0.42 85	1.7831
2	5069	1991	81	7710
3	5954	1867	76	7591
4	6864	1744	72	7472
5	7800	1621	67	7354
0.266	6.8763	2.1499	0.42 63	1.7236
7	9754	1378	58	7120
8	7.0775	1258	53	7005
9	1827	1139	49	6890
70	2911	1020	44	6776

$1/k$	C_s	A_1	B_1	C_1
0.270	7.2911	2.1020	0.42 44	1.6776
1	4028	0903	39	6664
2	5180	0786	34	6552
3	6369	0669	29	6440
4	7598	0554	25	6329
5	8868	0439	20	6219
0.276	8.0182	2.0325	0.42 15	1.6110
7	1542	0211	10	6001
8	2950	0098	04	5894
9	4410	1.9986	0.41 99	5787
80	5924	9875	94	5681
0.281	8.749	1.9764	0.41 89	1.5575
2	912	9654	84	5470
3	9.081	9544	78	5366
4	257	9436	73	5263
5	440	9327	68	5159
0.286	9.631	1.9220	0.41 62	1.5058
7	830	9713	57	4956
8	10.037	9007	51	4856
9	254	8901	46	4755
90	481	8796	40	4656
0.291	10.717	1.8691	0.41 35	1.4556
2	965	8587	29	4458
3	11.226	8484	23	4361
4	499	8381	17	4264
5	786	8279	12	4167
0.296	12.089	1.8177	0.41 06	1.4071
7	409	8016	00	3976
8	747	7976	0.40 94	3882
9	13.104	7876	88	3788
0.300	484	7776	82	3694
0.31		1.6810	0.40 18	1.2792
0.32		5890	0.39 51	1939
0.33		5012	0.38 77	1135

$$C_s = -\frac{\Gamma(1+3/k) - 3\Gamma(1+2/k)\Gamma(1+1/k) + 2\Gamma^3(1+1/k)}{[\Gamma(1+2/k) - \Gamma^2(1+1/k)]^{3/2}} \tag{1.11$'$}$$

$$\left.\begin{aligned} &1/a \equiv 0.4343/k \\ &g = m_x + A_2\sigma_x \\ &u = m_x - B_2\sigma_x \\ &g - u = C_2\sigma_x \end{aligned}\right\} \tag{1.15}$$

where

$$\left.\begin{aligned} &A_2 \equiv \Gamma(1+1/k)/[\Gamma(1+2/k) - \Gamma^2(1+1/k)]^{1/2} \\ &B_2 \equiv [1-\Gamma(1+1/k)]/[\Gamma(1+2/k) - \Gamma^2(1+1/k)]^{1/2} \\ &C_2 \equiv A_2 + B_2 \equiv 1/[\Gamma(1+2/k) - \Gamma^2(1+1/k)]^{1/2} \end{aligned}\right\} \tag{1.16}$$

The values of C_s, A_2, B_2 and C_2 can be tabulated as a function of $1/k$ as well as that for the second asymptote. These values, however, correspond to the ones for the third asymptote of the smallest value prepared by Gumbel in his book, as follows:

For the third asymptote of largest value (the author)[7]		For the third asymptote of smallest value (Gumbel)[16]
C_s	$\equiv$	$-\beta_1(k)$
A_2	$\equiv$	$B(k) - A(k)$
B_2	$\equiv$	$A(k)$
C_2	$\equiv$	$B(k)$

Since there is no difference between the two values in essence, it will be unnecessary to show such a table in this paper.

4. Selection of applicable type of asymptote from view point of hydrologic frequency analysis

It has been shown above that the three types of asymptote for the largest value should be applicable in limited range of the value of coefficient of skew C_s. But such a discussion for the population value is not always realistic for the sample value. The reason is that, for example, the value of the coefficient of skew C'_s of the first asymptote sample, which means a sample taken from the population of the first asymptote, is not always equal to $C_s = 1.1395\cdots$ for the population, but there are various cases where it will be larger or smaller than $C_s = 1.1395\cdots$ from a view point of sample theory.

Otherwise, one of the most important problems in relation to the hydrologic largest variates is how to estimate a bigger future value. In estimation of such a value by basing on a sample of small size obtained by hydrologic observation, it may be often happen that the third asymptote is wrongly applied to the first or the second asymptote sample, or that either the first or the second asymptote is wrongly applied to the third asymptote sample. However, from the view point of the prevention of disasters, the former mistake seems to be more serious than the latter.

Under the above considerations, the following course of treatment in the hydrologic frequency analysis should be adopted.

i) If the series of plotted points of hydrologic data on the extremal probability paper is scatterd about a straight line or a curve, denoted as G or B in Fig. 1.3, the first asymptote must be applied.

ii) If the series of points is scattered about a curve, denoted as A in Fig. 1.3, the second asymptote is usefully applicable.

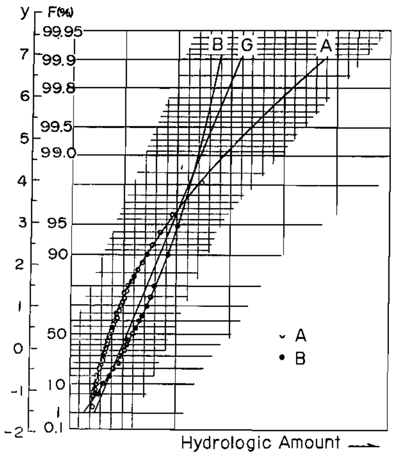

Fig. 1.3 Condition of application for three asymptotes.

iii) The third asymptote must not be applied, except for a family of data of which the plotted points are arranged about the curve B with extremely large curvature.

Therefore, a discussion of the third asymptote will be omitted in the following sections.

5. Concept of plotting value

The simplest method of estimation of the parameters of asymptotes for the distribution of the largest value is the so-called classical mothod of moment in which the population moments described in section 3 are directly replaced by the sample moments. The population moments are obtained by integration over the whole domain of variation, while the sample moments are based on the sample of limited and small size N, generally. The results obtained by the classical method of moment seem to be not so good in fitness to hydrologic data, because of the bias between two moments, although it may not be a sufficient reason. In order to eliminate this bias, an approximate method is required, by which the population moments may be evaluated as a function of sample size N.

In the field of hydrologic statistics, the population moments are often calculated by using the plotting position. However, the concept of the plotting position must be more fully considered in applying to calculation of the population moments, because it may be originally used for construction of the empirical distribution function.

Suppose that $x_1, x_2, \cdots x_N$ are a set of observations of size N, and $x_1 \leq x_2 \leq \cdots \leq x_N$. And let them be a sample of size N from a population, having the continuous *cdf* $F(x)$ of which only the type and, therefore, the value of coefficient of skew C_s, is known. Then, as is well known, the probability element $dp(x_i)dx_i$ for the i-th order statistics x_i in such a sample is given by

$$dp(x_i)dx_i = \frac{\Gamma(N+1)}{\Gamma(i)\Gamma(N-i+1)}[F(x_i)]^{i-1}[1-F(x_i)]^{N-i}f(x_i)dx_i \quad (1.17)$$

If the parameters included in $F(x)=\int_{-\infty}^{x} f(x)dx$ are not larger than three in number, and if the distribution functions of such asymmetrical types as are applicable to the hydrologic frequency analysis are supposed, the unknown parameters included in Eq. (1.17) must be two in number, since the value of coefficient of skew is known. Now, let the linear reduced variate z be defined as

$$z = a(x-b) \quad (1.18)$$

where, a and b are numerical constants.

Then, the probability element $dp(z_i)dz_i$ for the i-th order statistics z_i is given by

$$dp(z_i)dz_i = \frac{\Gamma(N+1)}{\Gamma(i)\Gamma(N-i+1)}[F(z_i)]^{i-1}[1-F(z_i)]^{N-i}f(z_i)dz_i \quad (1.19)$$

and the unknown parameter must be not included in this equation.

Since the unknown parameters a and b should be estimated as follows,

$$\sum_i [z_i - a(x_i - b)]^2 = \text{minimum} \quad (1.20)$$

under the condition

$$E(z_i) = a[E(x_i) - b], \quad (1.21)$$

the problem of estimation of the parameters is reduced to estimating the value z_i corresponding to x_i.

A schematic figure of the distribution of z_i corresponding to x_i is shown in Fig. 1.4, where x_i are regarded as the fixed variate. Since the probability element of z_i is given by Eq. (1.19) as the functions of order i and sample size N, the determination of z_i is almost equivalent to estimating the value of $\hat{z}_i$ satisfying the following condition

$$\sum_j (z_{ij}-\hat{z}_i)^2=\text{minimum}. \quad (1.22)$$

The solution of Eq. (1.22) is, clearly,

$$\hat{z}_i=E(z_i) \quad (1.23)$$

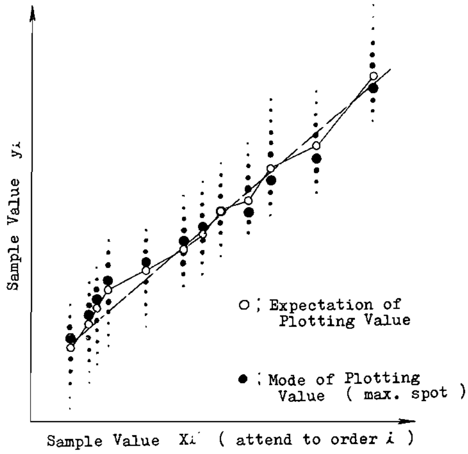

Fig. 1.4 Schematic figure for distribution of plotting value.

The line connecting each value of $E(z_i)$ is not always equivalent to the one which satisfies the condition of Eq. (1.20) in a theoretical sense, but the difference between the two lines seems to be so small that it may be ignored in a practical sense.

In this paper, the value of $E(z_i)$ obtained from this idea is named the (expected) plotting value, to distinguish it from the plotting position.

In addition, if the discussion of the plotting position is made, the distribution of the value of F_i instead of the value of x_i should be considered in Eq. (1.17).

$$dp(F_i)dF_i=\frac{\Gamma(N+1)}{\Gamma(i)\Gamma(N-i+1)}[F_i]^{i-1}[1-F_i]^{N-i}dF_i \quad (1.24)$$

This equation is originally parameter-free, differing from Eq. (1.17), and the function of i and N. The polotting position $\hat{F}_i$ must satisfy the following condition,

$$\sum_j (F_{ij}-\hat{F}_i)=\text{minimum}. \quad (1.25)$$

And its solution is, evidently,

$$\hat{F}_i=E(F_i)=i/(N+1) \quad (1.26)$$

This result differs in no way from the plotting position adopted by Thomas[18], Gumbel[2,3] and the others.

6. Plotting values for first and second asymptotes

Although the plotting position is distribution-free, the plotting value is neither distribution-free nor parameter-free, except the special type of distribution with two parameters. In this section, the plotting values for the first and the second asymptote samples will be discussed.

(1) The first asymptote :

$$F(y)=\exp(-e^{-y})$$
$$f(y)=\exp(-y-e^{-y})$$
$$y \equiv z = a(x-u)$$

Since the reduced extreme y itself is linear to the actual variate x, the value of $E(y_i)$ has only to be estimated, that is, the plotting value for the first asymptote is parameter-free. In this case, Eq. (1.19) becomes

$$dp(y_i)dy_i=\frac{\Gamma(N+1)}{\Gamma(i)\Gamma(N-i+1)}[\exp\{-(i-1)e^{-y_i}\}][1-\exp(-e^{-y_i})]^{N-i}\times$$
$$[\exp(-y_i-e^{-y_i})]dy_i$$

Therefore, $E(y_i)$ is

$$E(y_i)=\frac{\Gamma(N+1)}{\Gamma(i)\Gamma(N-i+1)}\int_{-\infty}^{\infty} y\exp(-y-ie^{-y})[1-\exp(-e^{-y})]^{N-i}dy$$
$$=\frac{\Gamma(N+1)}{\Gamma(i)\Gamma(N-i+1)}\sum_{r=0}^{N-i}(-1)^r{}_{N-i}C_r\int_{-\infty}^{\infty} y\exp\{-y-(i+r)e^{-y}\}dy$$

After several calculations, it becomes

$$E(y_i)=\frac{\Gamma(N+1)}{\Gamma(i)\Gamma(N-i+1)}\sum_{r=0}^{N-i}(-1)^r{}_{N-i}C_r\frac{1}{i+r}\{\gamma+\lg(i+r)\} \qquad (1.27)$$

where, γ is a so-called Euler's constant and C means the symbol of combination.

(2) The second asymptote :

$$F(y)=\exp(-e^{-y})\equiv\exp(-z^{-k})$$
$$y=k\lg z$$
$$z=(x+b)/(u+b)$$

The plotting value $E(z_i)$ in the case of the second asymptote is

$$E(z_i)=\frac{\Gamma(N+1)}{\Gamma(i)\Gamma(N-i+1)}k\int_0^{\infty} z^{-k}\exp(-iz^{-k})[1-\exp(-z^{-k})]^{N-i}dz$$
$$=\frac{\Gamma(N+1)}{\Gamma(i)\Gamma(N-i+1)}k\sum_{r=0}^{N-i}(-1)^r{}_{N-i}C_r\int_0^{\infty} z^{-k}\exp\{-(i+r)z^{-k}\}dz$$

Then, it is expressed as follows :

$$E(z_i)=\frac{\Gamma(N+1)}{\Gamma(i)\Gamma(N-i+1)}\Gamma(1-1/k)\sum_{r=0}^{N-i}(-1)^r{}_{N-i}C_r(i+r)^{-1+1/k} \qquad (1.28)$$

(3) A practical method of computation of the plotting values : The plotting value $E(y_i)$ or $E(z_i)$ of the i-th order statistics in the first or the second asymptote sample must be able to be calculated strictly by Eq. (1.27) or Eq. (1.28) as a functions of the order i and the sample size N. But in solving these equations for the various values of i and N, it may be seen that the smaller i is, and the larger N, the more difficult the calculation becomes. Therefore, the following method of computation may be used in a practical sense.

i) $i=N$: The plotting values for the two asymptotes are obtained by putting $i=N$ in Eqs. (1.27) and (1.28), respectively.

For the first asymptote ;

$$E(y_N)=\gamma+\lg N \tag{1.29}$$

For the second asymptote ;

$$E(z_N)=N^{1/k}\Gamma(1-1/k) \tag{1.30}$$

ii) $i=1$; The plotting values for the two asymptotes are calculated by a method of numerical integration. Several results obtained by such calculation are shown in Fig. 1.5, where the confidence limit[12] defined by Eq. (1.31) is also shown.

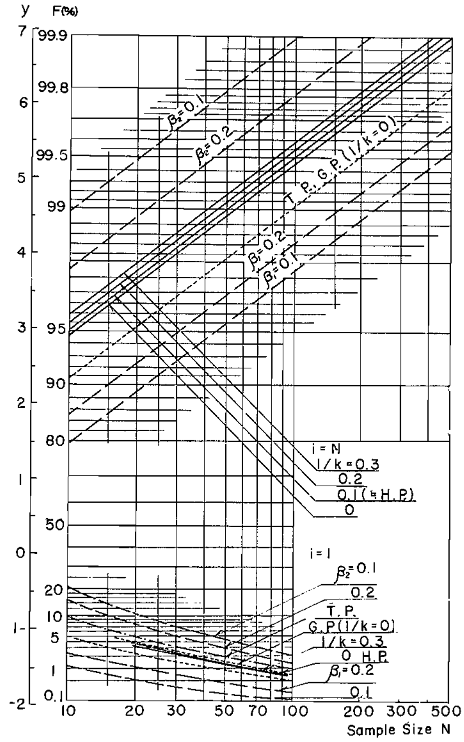

Fig. 1.5 Plotting values for $i=N$ and 1. Where T.P.=Thomas Plot$=i/(N+1)$, H.P.=Hazen Plot$=(2i-1)/2N$ and G.P. means Gumbel Plot proposed in Ref. 19).

$$\left.\begin{aligned} P(y_i\leq y_{\beta_1} \text{ or } z_i\leq z_{\beta_1})\leq\beta_1&=\frac{\Gamma(N+1)}{\Gamma(i)\Gamma(N-i+1)}\int_0^{F_{\beta_1}}F^{i-1}(1-F)^{N-i}dF_i \\ P(y_i\geq y_{\beta_2} \text{ or } z_i\geq z_{\beta_2})\leq\beta_2&=\frac{\Gamma(N+1)}{\Gamma(i)\Gamma(N-i+1)}\int_{F_{\beta_2}}^{1}F^{i-1}(1-F)^{N-i}dF_i \end{aligned}\right\} \tag{1.31}$$

iii) $1<i<N$: The plotting values for the two asymptotes are not easily obtained as long as the troublesome calculation of numerical integration is not performed. So, the values which satisfy the following equation will be

adopted as the approximations of plotting value, after a model of the method proposed by Gumbel[19] in his paper "Simplified Plotting of Statistical Observations".

$$\hat{F}_i = F_1 + \frac{i-1}{N-1}(F_N - F_1) \tag{1.32}$$

where

$$F_1 \equiv F\{E(y_1)\} \quad \text{or} \quad F\{E(z_i)\}$$
$$F_N \equiv F\{E(y_N)\} \quad \text{or} \quad F\{E(z_N)\}$$

Then, the estimates of plotting value are obtained by

$$\left.\begin{array}{ll} \text{for the first asymptote} \ ; & \hat{y}_i = -\lg(-\lg \hat{F}_i) \\ \text{for the second asymptote} \ ; & \hat{z}_i = (-\lg \hat{F}_i)^{-1/k} \end{array}\right\} \tag{1.33}$$

Therefore the plotting value $\hat{y}_i$ is expressed as a function of the sample size N and the order i, and $\hat{z}$ as a function of the skewness parameter $1/k$ and N and i.

7. Practical method of estimation of population parameters

(1) **The first asymptote**: It has already been explained that the fundamental equations for estimation of the parameters in the first asymptote are expressed as follows:

$$a = \sigma_y / \sigma_x$$
$$u = m_x - m_y / a$$

If the sample size N is very large, the classical method of moment must be valid, in which the following values are adopted as already shown by Eq. (1.12).

$$\sigma_y = \pi / \sqrt{6}$$
$$m_y = \gamma$$

But, since the number of hydrologic observations is usually 10^1 in order, it cannot but be considered that $m_y \leq \gamma$, $\sigma_y \leq \pi/\sqrt{6}$. Therefore, the population values $\bar{y}$ and s_y as a function of sample size N should be adopted instead of the values of which $m_y = \gamma$ and $\sigma_y = \pi/\sqrt{6}$, although some difficult points remain to be discussed. In this case, the parameters a and u are rewritten as follows:

$$\left.\begin{array}{l} 1/a = s_x / s_y \\ u = \bar{x} - (1/a)\bar{y} \end{array}\right\} \tag{1.34}$$

where

Table 1.2 Population values of $\bar{y}$ and s_y for N in first asymptote.
$1/a=s_x/s_y$, $u=\bar{x}-\bar{y}/a$

N	$\bar{y}$	s_y
20	0.56 92	1.1 825
2	96	895
4	99	956
6	0.57 02	1.2 009
8	05	055
30	0.57 07	1.2 095
2	09	130
4	11	162
6	13	192
8	15	219
40	0.57 17	1.2 244
2	18	266
4	20	287
6	21	306
8	23	323
50	0.57 24	1.2 338
2	25	352
4	26	366
6	27	379
8	28	391
60	0.57 29	1.2 402
2	30	413
4	31	423
6	32	433
8	33	442
70	0.57 34	1.2 451
75	36	472
80	37	490
85	39	506
90	40	520
95	41	532
100	42	542

$$\left.\begin{aligned}\bar{x}&=\frac{1}{N}\sum x, \quad \overline{x^2}=\frac{1}{N}\sum x^2\\ s^2_x&=\frac{1}{N}\sum(x-\bar{x})^2=\overline{x^2}-\bar{x}^2\end{aligned}\right\} \quad (1.35)$$

The population values $\bar{y}$ and s_y calculated by basing on the plotting value as a function of sample size N are presented in Table 1.2.

(2) **The second asymptote**: In estimation of the parameters of the second asymptote, the problem of bias between the population values arises as well as in the first one. Speaking generally, this bias seems to become large with the increase of the order of moment, because it must be caused by the difference between the domains of variate under consideration. Therefore, if the bias of moment of the highest order which is needed at least in the calculation is satisfactorily settled, the ones of the other moment of lower order may be so small that they can be ignored in a practical sense. That is, if the skewness parameter $1/k$ has only to be evaluated successfully, Eq. (1.13) must be usefully available. The theory of plotting value described in section 6 will be utilized for such a purpose.

Now, since the plotting value z is linear reduced variate of actual variate x, the sample value of the coefficient of skew $C'_s(z)$ about z must be equal to $C'_s(x)$ about x,

$$C'_s(z)=C'_s(x) \quad (1.36)$$

The calculated values of C'_s are shown in Fig. 1.6 as a function of the skewness parameter $1/k$ and the sample size N. Moreover, if the relation between the population value C_s given by Eq. (1.9) and the sample value C'_s is expressed by

$$C_s=C'_s(1+\beta_s) \quad (1.37)$$

where β_s ; the additional coefficient of skewness

the values of β_s are shown in Fig. 1.7.

Therefore, if the sample value of the coefficient of skew C'_s is calculated from a sample of size N by

$$\left.\begin{aligned}C'_s&=\frac{1}{N}\frac{\sum(x-\bar{x})^3}{s_x^3}=\frac{\overline{x^3}-3\overline{x^2}\bar{x}+2\bar{x}^3}{s_x^3}\\ s^2_x&=\overline{x^2}-\bar{x}^2, \quad \overline{x^j}=\frac{1}{N}\sum x^j\end{aligned}\right\} \quad (1.38)$$

the value of skewness parameter $1/k$ must be able to be estimated from Fig. 1.6, or from Table 1.1 by using the value of C_s presumed by Eq. (1.37)

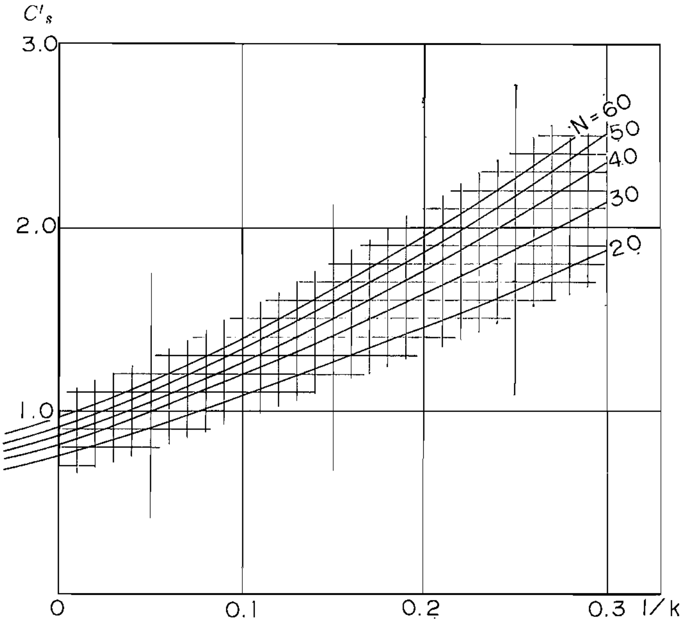

Fig. 1.6 Relation between C'_s and $1/k$ for given N.

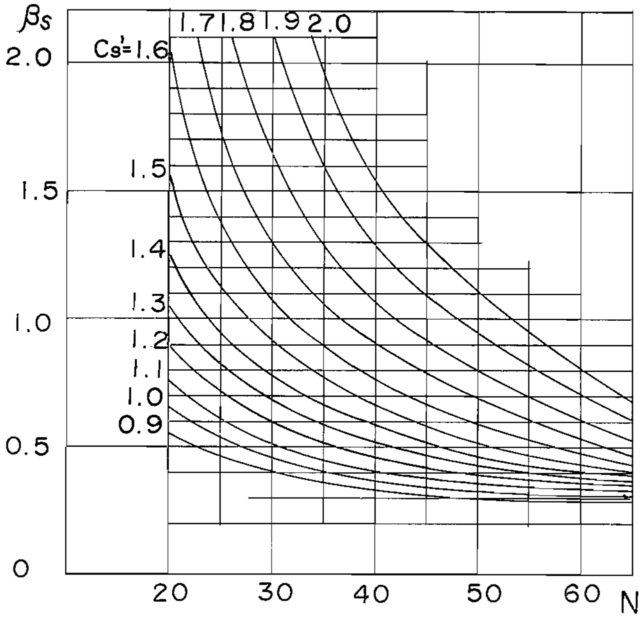

Fig. 1.7 Values of β_s provided for $C_s = C_s'(1+\beta_s)$.

and Fig. 1.7. And the other parameters u and b can be easily estimated from Eq. (1.13) and Table 1.1, using the estimates

$$\left.\begin{aligned} \sigma_x &= \sqrt{N/(N-1)}\, s_x \\ m_x &= \bar{x} \end{aligned}\right\} \tag{1.39}$$

Besides, there is a case where the location parameter b may be assumed to be zero, which happens when the series of points of $\log x$ on the extremal probability paper is scattered about a straight line. In this case, it is needless to say that the other parameters may be estimated by

$$\left.\begin{aligned} 1/a &= s_{\log x}/s_y \\ \log \hat{u} &= \overline{\log x} - (1/a)\bar{y} \end{aligned}\right\} \tag{1.40}$$

in which

$$\left.\begin{aligned} &s_y \text{ and } \bar{y}\,; \quad \text{see Table 1.2} \\ &s^2_{\log x} = \overline{(\log x)^2} - \overline{\log x}^{\,2} \\ &\overline{(\log x)^j} = \frac{1}{N}\Sigma(\log x)^j \end{aligned}\right\} \tag{1.41}$$

8. Expected value for given return period

If all parameters are reasonably estimated, the expected value of hydrologic amount for the desired return period T can be easily calculated by

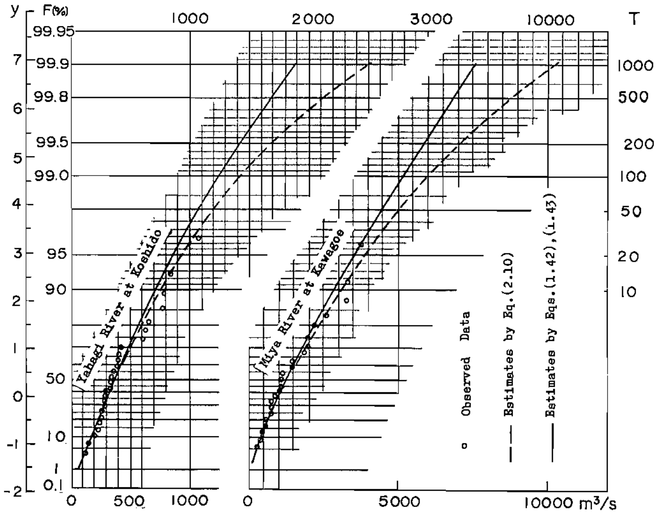

Fig. 1.8-(1) Frequency curves for annual maximum flow, Yahagi River and Miya River.

Table 1.3 Characteristic values of observed hydrologic data.

	Classification of Data		Period of Observation	Sample Size N	Sample Value of: Mean $\bar{x}$	Sample Value of: Standard Diviation s_x	Sample Value of: Coeff. of Skew C'_s	Equation for Estimates
Annual Max. Floods	Yahagi River, One-day peaks, at Koshido		1930—1956	26	m³/s 426	m³/s 234.5	1.132	$\log(x+1473)=\log 1777+0.0400y$
	Miya River, One-day peaks, at Kawagoe		1928—1950	23	1452	984	1.036	$\log(x+13417)=\log 14409+0.0239y$
	Yodo River, Momentum Peaks, at	Hazukashi	1912—1961	50	1089	669	0.944	$x=779+542y$
		Kamo			2032	1468	1.272	$\log(x+10996)=\log 12267+0.0389y$
		Hirakata			3028	1766	1.388	$\log(x+8493)=\log 10630+0.0514y$
Annual Max. Amount of Rainfall	One-Day Rainfall, at	Nara	1911—1955	36	mm/day 85	mm/day 60.5	0.165	$x=76+15.7y$
		Wakayama	1911—1955	44	113	36.5	0.852	$x=96+29.7y$
		Gobo	1911—1955	39	119	46.9	1.347	$\log(x+116)=\log 262+0.0511y$
		Tokushima	1891—1952 June-Oct.	62	153	81	1.991	$\log(x+108)=\log 225+0.0864y$
		Gojyo	1911—1955	44	100	33.2	0.677	$x=84+27.0y$
		Tsu	1901—1955	55	131	49.4	1.112	$\log(x+546)=\log 653+0.0240y$
		Ueno	1901—1953	53	98	44.9	1.784	$\log(x+67)=\log 144+0.0783y$
		Owase	1901—1955	55	297	115.4	1.397	$\log(x+466)=\log 710+0.0471y$
	4 hr-Rainfall, at	Kyoto	1896—1950	54	mm/4hr 62	mm/4hr 23.3	0.572	$x=51+18.8y$
	One-day Rainfall, at	Kyoto	1911—1950	39	mm/day 97	mm/day 26.2	0.453	$x=85+21.4y$
		Fukuchiyama	1911—1955	45	100	35.6	1.158	$\log(x+288)=\log 371+0.0298y$
		Maizuru	1911—1955	42	102	48.6	2.752	$\log(x-29)=\log 54+0.1433y$
		Hikone	1926—1955	30	90	27	1.453	$\log(x+30)=\log 106+0.068y$
		Otsu	1926—1955	30	106	32	1.149	$\log(x+169)=\log 260+0.038y$
		Minakuchi	1926—1955	30	100	45	2.272	$\log(x-18)=\log 63+0.130y$

the following relations, as is well known,

$$\left.\begin{array}{ll}\text{for the first asymptote;} & x=u+(1/a)y,\\ \text{for the second asymptote;} & \log(x+b)=\log(u+b)+(1/a)y,\end{array}\right\}\qquad(1.42)$$

where usually

$$y=-\lg[\lg T/(T-1)]\qquad(1.43)$$

Figure 1.8 shows several examples of application of the proposed method

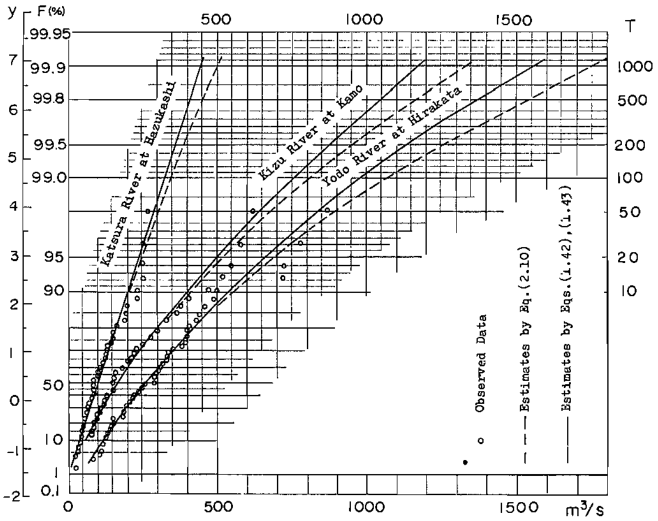

Fig. 1.8-(2) Frequency curves for annual maximum flow, Yodo River.

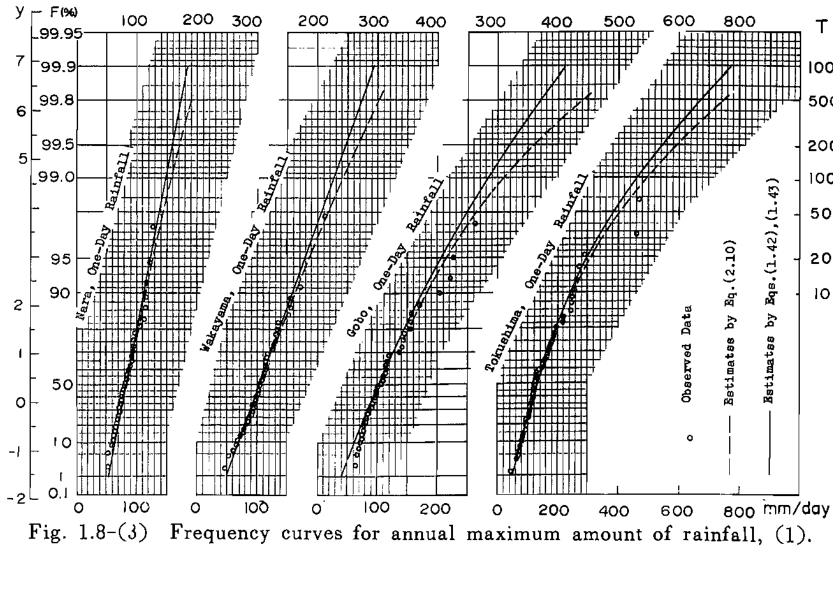

Fig. 1.8-(3) Frequency curves for annual maximum amount of rainfall, (1).

to hydrologic data in Table 1.3, where the observed data are plotted as $F=i/(N+1)$.

From a stochastic viewpoint, however, various points remain to be discussed concerning the expected value of hydrologic amont for the desired return period as expressed by Eq. (1.43). These problems will be discussed

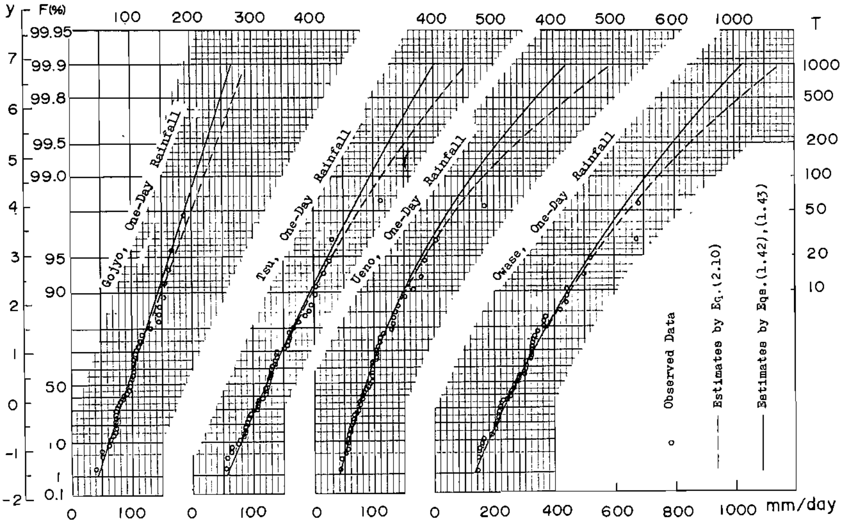

Fig. 1.8-(4) Frequency curves for annual maximum amount of rainfall, (2).

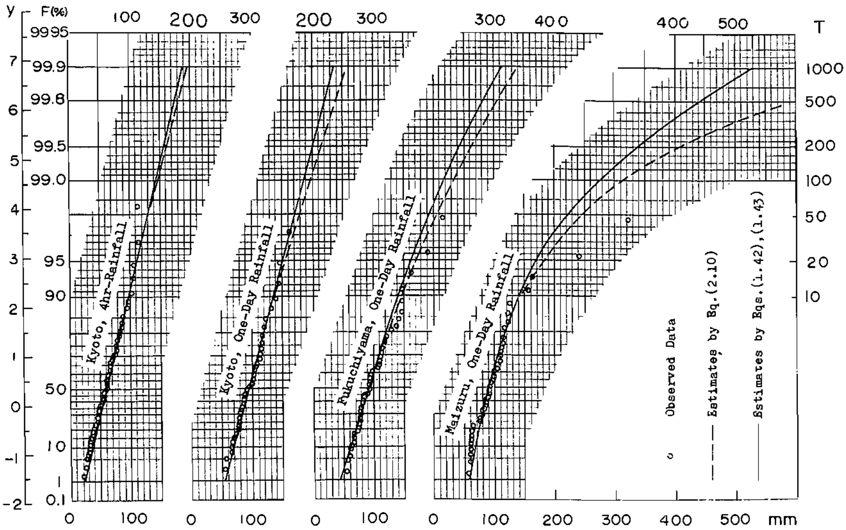

Fig. 1.8-(5) Frequency curves for annual maximum amount of rainfall, (3).

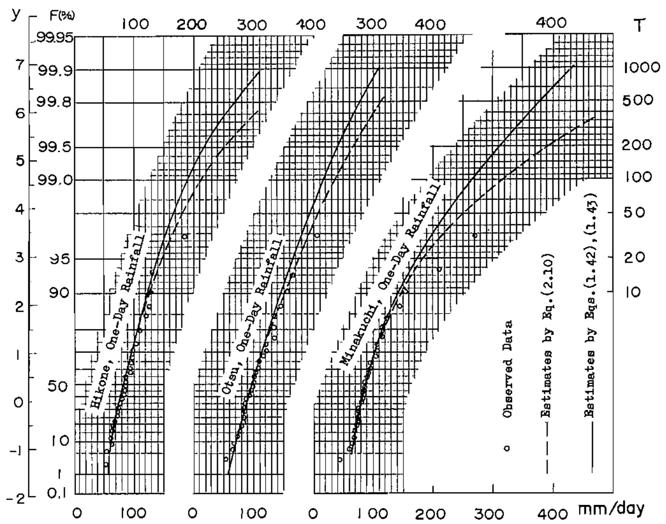

Fig. 1.8-(6) Frequency curves for annual maximum amount of rainfall, (4).

in Part II.

9. Conclusion

In this part, first, the statistical properties of three types of extreme (largest) value distribution were examined. As the result, it was disclosed that the applicable range for them must be discriminated by the population value of the coefficient of skew. Next, a practical method of estimation of the parameters included in them was successfully developed by using the concept of the plotting value.

Since these studies were already made in 1954~1956 and 1959, this publication may seem to be too late. But the author believes that this paper is still useful in the field of hydrologic frequency analysis.