VOL. 10, NO. 2 WATER RESOURCES RESEARCH APRIL 1974

Just a Moment!

J. R. Wallis

IBM Research Center, Yorktown Heights, New York 10598

N. C. Matalas and J. R. Slack

U.S. Geological Survey, Washington, D. C. 20244

The distribution functions for three statistics, the mean, standard deviation, and coefficient of skewness, for small samples from various distributions were obtained by Monte Carlo experiments. Sample sizes of 10 (10) 90 were considered, and the distributions used were the normal, Gumbel (extreme value type 1), log normal, Pearson type 3 (gamma), Weibull, and Pareto type 1 (Pearson type 4). Values of the coefficient of skewness used in generating the samples were in the range [0.0, 15.0]. Pronounced skews, biases, and constraints in the sampling properties of the statistics were observed. Examples of the available graphs of the distribution functions are presented.

Decision variables pertaining to the design of water resource systems are functions of various parameters, including those that characterize the stochastic properties of hydrologic inputs to the systems. Because hydrologic sequences are of finite lengths, only estimates of the hydrologic parameter values can be obtained, and as a result, uncertainty in the design decisions is in part attributed to hydrologic uncertainty. The particular set of hydrologic parameters to be estimated depends upon the purposes and objectives underlying the proposed development of a water resource system. In general, the set is likely to include those parameters that are defined in terms of the low-order moments, namely, the mean, standard deviation, and coefficient of skewness.

Generally, a hydrologic sequence of length n is assumed to be a sample on n identically distributed random variables, where each random variable has a finite mean μ, standard deviation σ, and coefficient of skewness γ. Moreover, the n random variables ordered in time are assumed to be a stationary stochastic process. Consequently, estimates of μ, σ, and γ, denoted by $\bar{X}$, S, and G, may be defined in terms of time averages of the n observations forming the hydrologic sequence. The sampling properties of $\bar{X}$, S, and G as functions of n depend upon the method of estimation, the marginal probability distribution function, and the type of generating mechanism of the stochastic process.

To gain some insight into the sampling properties of $\bar{X}$, S, and G, Monte Carlo experiments were carried out on the basis of the method of moments for estimating μ, σ, and γ and under the assumption that the stochastic process is purely independent, so that the n random variables are independent as well as identical. The experiments were performed on each of several distribution functions for various values of n and γ. From these experiments the probability distribution functions of $\bar{X}$, S, and G were defined empirically, and values of the mean, standard deviation, coefficient of skewness, and coefficient of kurtosis for $\bar{X}$, S, and G were calculated.

Experimental Design

Six probability distribution functions were considered: normal, three-parameter log normal, three-parameter Pearson type 3 (gamma), Gumbel (extreme value type 1), Weibull, and Pareto (Pearson type 4). For each distribution function $F(y)$, 100,000 samples of size $n = 10$ (10) 90 were generated with $\mu = 0$, $\sigma = 1$, and γ equal to the values indicated by an X in Table 1.

In general, let Y_{ij} denote the ith observation for the jth sequence of length n. For the jth sequence the mean $\bar{Y}_j$, standard deviation S_j, and coefficient of skewness G_j were defined as follows:

$$\bar{Y}_j = \sum_{i=1}^{n} Y_{ij}/n \tag{1}$$

$$S_j = \left[\sum_{i=1}^{n} Y_{ij}^2/n - \bar{Y}_j^2\right]^{1/2} \tag{2}$$

$$G_j = \left[\sum_{i=1}^{n} Y_{ij}^3/n - 3\bar{Y}_j S_j^2 - \bar{Y}_j^3\right] \Big/ S_j^3 \tag{3}$$

Let X_j, denoting any one of the above three statistics, be an observation on the random variable X having mean $\mu(X)$, standard deviation $\sigma(X)$, coefficient of skewness $\gamma(X)$, and coefficient of kurtosis $\lambda(X)$ defined as

$$\mu(X) = E(X) \tag{4}$$

$$\sigma(X) = \{E[X - E(X)]^2\}^{1/2} \tag{5}$$

$$\gamma(X) = \{E[X - E(X)]^3\}/\sigma^3(X) \tag{6}$$

$$\lambda(X) = \{E[X - E(X)]^4\}/\sigma^4(X) \tag{7}$$

From the $M = 100{,}000$ observations on X the values of $\mu(X)$, $\sigma(X)$, $\gamma(X)$, and $\lambda(X)$ were approximated by

$$\tilde{\mu}(X) = \sum_{j=1}^{M} X_j/M \tag{8}$$

$$\tilde{\sigma}(X) = \left[\sum_{j=1}^{M} X_j^2/M - \tilde{\mu}^2(X)\right]^{1/2} \tag{9}$$

$$\tilde{\gamma}(X) = \left[\sum_{j=1}^{M} X_j^3/M - 3\tilde{\mu}(X)\tilde{\sigma}^2(X) - \tilde{\mu}^3(X)\right] \Big/ \tilde{\sigma}^3(X) \tag{10}$$

$$\tilde{\lambda}(X) = \left[\sum_{j=1}^{M} X_j^4/M - 4\tilde{\mu}(X)\tilde{\sigma}^3(X)\tilde{\gamma}(X) - 6\tilde{\mu}^2(X)\tilde{\sigma}^2(X) - \tilde{\mu}^4(X)\right] \Big/ \tilde{\sigma}^4(X) \tag{11}$$

TABLE 1. Values of γ for Distribution Function F

$F(y)$	$\gamma = 0$	$\gamma = \frac{1}{4}$	$\gamma = \frac{1}{2}$	$\gamma = (\frac{1}{2})^{1/2}$	$\gamma = 1$	$\gamma = 1.14$	$\gamma = (2)^{1/2}$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$	$\gamma = 5$	$\gamma = 10$	$\gamma = 15$
Normal	X												
Log normal		X	X	X	X	X	X	X	X	X	X	X	X
Pearson		X	X	X	X	X	X	X	X	X	X		
Gumbel						X							
Weibull		X	X	X	X	X	X	X	X	X	X	X	X
Pareto									X	X	X	X	X

The distribution function $F(x) = P[X < x]$ was defined as follows. For any one of the statistics X, a range $R \equiv [R_X^-, R_X^+]$ was selected as needed, such that at least $P[R_X^- \leq x \leq R_X^+] < 10^{-4}$. The range R was divided into 400 uniform intervals

$$\left\{I_k = \left(\frac{k-1}{400} R_X^+ + \frac{401-k}{400} R_X^-, \frac{k}{400} R_X^+ + \frac{400-k}{400} R_X^-\right) \mid k = 1, \cdots, 400\right\}$$

For each I_k a count C_k was determined of the occurrences of the statistic X_{ij} falling in that interval, whereby the distribution function $F(x) = P[X < x]$ was defined as

$$F(x) = \sum_{k \mid I_k' < x} C_k/100{,}000$$

where I_k' is the lower end point of I_k.

SEQUENCE-GENERATING ALGORITHMS

For the six distributions noted above, the probability density functions $f(y)$, the distribution functions $F(y)$ (in those cases where $F(y)$ has a closed form representation), and the relations of their parameters to $\mu(Y)$, $\sigma(Y)$, $\gamma(Y)$, and $\lambda(Y)$ are as follows.

Normal

$$f(y) = \frac{1}{a(2\pi)^{1/2}} \exp\left[-\frac{1}{2}\left(\frac{y-m}{a}\right)^2\right] \qquad -\infty \leq y \leq \infty \tag{12a}$$

$$\mu(Y) = m \tag{12b}$$

$$\sigma(Y) = a \tag{12c}$$

$$\gamma(Y) = 0 \tag{12d}$$

$$\lambda(Y) = 3 \tag{12e}$$

Three-parameter log normal

$$f(y) = \frac{1}{a(2\pi)^{1/2}(y-c)} \exp\left\{-\frac{1}{2}\left[\frac{\ln(y-c) - m}{a}\right]^2\right\} \qquad y \geq c \tag{13a}$$

$$\mu(Y) = c + \exp[(a^2/2) + m] \tag{13b}$$

$$\sigma(Y) = \{\exp[2(a^2 + m)] - \exp[a^2 + 2m]\}^{1/2} \tag{13c}$$

$$\gamma(Y) = \frac{\exp[3a^2] - 3\exp[a^2] + 2}{\{\exp[a^2] - 1\}^{3/2}} \tag{13d}$$

$$\lambda(Y) = \exp[4a^2] + 2\exp[3a^2] + 3\exp[2a^2] - 3 \tag{13e}$$

Three-parameter Pearson type 3 (gamma)

$$f(y) = \frac{1}{a\Gamma(b)}\left(\frac{y-m}{a}\right)^{b-1} \exp\left[-\left(\frac{y-m}{a}\right)\right] \qquad y \geq m \tag{14a}$$

$$\mu(Y) = m + ab \tag{14b}$$

$$\sigma(Y) = a(b)^{1/2} \tag{14c}$$

$$\gamma(Y) = 2/(b)^{1/2} \tag{14d}$$

$$\lambda(Y) = 3 + 6/b \tag{14e}$$

Gumbel (extreme value type 1)

$$f(y) = \frac{1}{a}\exp\left[-\left(\frac{y-m}{a}\right)\right] \cdot \exp\left\{-\exp\left[-\left(\frac{y-m}{a}\right)\right]\right\} \qquad -\infty \leq y \leq \infty \tag{15a}$$

$$F(y) = \exp\left\{-\exp\left[-\left(\frac{y-m}{a}\right)\right]\right\} \tag{15b}$$

$$\mu(Y) = m + \kappa a \tag{15c}$$

$$\sigma(Y) = a\pi/(6)^{1/2} \tag{15d}$$

$$\gamma(Y) = -\psi''(1)/[\psi'(1)]^{3/2} \approx 1.139547 \tag{15e}$$

$$\lambda(Y) = \psi'''(1)/[\psi'(1)]^2 \approx 2.400 \tag{15f}$$

where $\kappa \approx 0.57722$ is Euler's number and $\psi(x)$ is the digamma function, i.e., $d \ln \Gamma(x)/dx$, where $\psi'(\ )$, $\psi''(\ )$, and $\psi'''(\ )$ denote the first, second, and third derivatives of $\psi(\ )$.

Weibull

$$f(y) = \frac{c}{a}\left(\frac{y-m}{a}\right)^{c-1} \exp\left[-\left(\frac{y-m}{a}\right)^c\right] \qquad y \geq m \tag{16a}$$

$$F(Y) = 1 - \exp\left[-\left(\frac{y-m}{a}\right)^c\right] \tag{16b}$$

$$\mu(Y) = m + a\Gamma(1 + 1/c) \tag{16c}$$

$$\sigma(Y) = a\{[\Gamma(1 + 2/c) - \Gamma^2(1 + 1/c)]\}^{1/2} \quad (16d)$$

$$\gamma(Y) = [\Gamma(1 + 3/c) - 3\Gamma(1 + 2/c)\Gamma(1 + 1/c) + 2\Gamma^3(1 + 1/c)]\{[\Gamma(1 + 2/c) - \Gamma^2(1 + 1/c)]\}^{-3/2} \quad (16e)$$

$$\lambda(Y) = (\Gamma(1 + 4/c) - 4\Gamma(1 + 1/c)\Gamma(1 + 3/c) + 6\Gamma^2(1 + 1/c)\Gamma(1 + 2/c) - 3\Gamma^4(1 + 1c)] \cdot [\Gamma(1 + 2/c) - \Gamma^2(1 + 1/c)]^{-2} \quad (16f)$$

Pareto type 1
(Pearson type 4)

$$f(y) = ba^b/y^{b+1} \qquad a > 0 \qquad b > 0 \qquad y \geq a \quad (17a)$$

$$F(y) = 1 - (a/y)^b \quad (17b)$$

$$\mu(Y) = ab/(b - 1) \quad (17c)$$

$$\sigma(Y) = \{[a^2b/(b - 2)] - [a^2b^2/(b - 1)^2]\}^{1/2} \quad (17d)$$

$$\gamma(Y) = 2(b + 1)(b - 2)^{1/2}/(n - 3)b^{1/2} \quad (17e)$$

$$\lambda(Y) = 3(3b^2 + b + 2)(b - 2)/b(b - 3)(b - 4) \quad (17f)$$

For these distributions, the relations of $\gamma(Y)$ to $\lambda(Y)$ are shown in Figure 1, and selected values are shown in Table 2.

To generate samples from each of the distributions, it is necessary to devise algorithms for the generation of appropriately distributed pseudorandom numbers. The generation of such numbers is at best an art. It should be noted that a number in a digital computer has a very specific form. Both the range of magnitude and the degree of precision (or 'discreteness') of the representation are circumscribed by the computer hardware and software. An analysis of a computer system's arithmetic (both the hardware and the software) and the built-in algorithms for generating pseudorandom numbers should be an integral part of the studies for any important project where decisions are dependent upon analyses utilizing such numbers.

The basic set of pseudorandom numbers are those that are uniformly distributed as these numbers are utilized in the generation of other numbers following specific distribu-

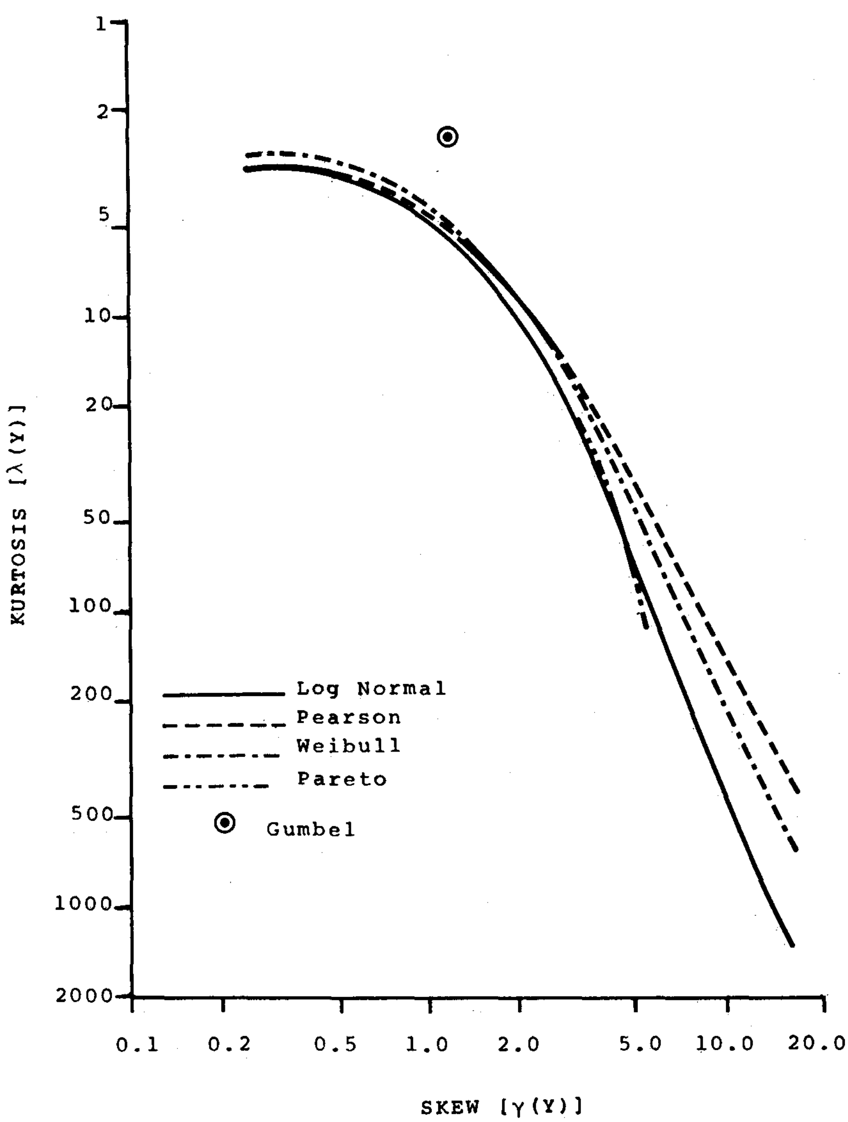

Fig. 1. Relations of skew to kurtosis.

TABLE 2. Values of Skew and Kurtosis

Skew $\gamma(Y)$	Kurtosis $\lambda(Y)$ for Indicated Distribution: Normal	Gumbel	Log Normal	Pearson	Weibull	Pareto
0.0	3.0					
0.25			3.11	3.09	2.77	
0.50			3.45	3.38	3.03	
0.71			3.90	3.75	3.40	
1.00			4.83	4.50	4.16	
1.14		2.40	5.39	4.95	4.62	
1.41			6.75	6.00	5.73	
2.00			10.86	9.00	9.00	
3.00			22.40	16.50	17.66	20.72
4.00			41.00	27.00	30.60	44.67
5.00			68.26	40.50	48.25	98.49
10.00			387.78	153.00	218.43	
15.00			1139.73	340.50	546.73	

tions. For this study of the sampling properties of the estimates of the mean, standard deviation, and coefficient of skewness, a well-tested uniform pseudorandom number generator [*Lewis et al.*, 1969] and a carefully programed Box-Muller transform [*Box and Muller*, 1958] for generating normally distributed pseudorandom numbers were used.

Let u denote a pseudorandom number uniformly distributed on the interval [0, 1], and let η denote a pseudorandom number normally distributed with zero mean and unit standard deviation. The algorithms used to generate the samples from the six distributions noted above were as follows.

Normal

$$y_i \equiv \eta_i \tag{18}$$

TABLE 3. Bias Factors for Standard Deviation

Distribution	Skew, γ	Sequence Length n: 10	20	30	40	50	60	70	80	90
Normal	0.0	1.084	1.040	1.026	1.019	1.016	1.013	1.011	1.010	1.009
Gumbel	1.14	1.108	1.053	1.035	1.027	1.021	1.018	1.015	1.013	1.012
Log normal	0.25	1.085	1.040	1.026	1.020	1.016	1.013	1.011	1.010	1.009
	0.50	1.088	1.042	1.028	1.021	1.016	1.014	1.012	1.010	1.009
	0.71	1.093	1.044	1.029	1.022	1.017	1.014	1.012	1.011	1.010
	1.00	1.101	1.048	1.032	1.024	1.019	1.016	1.014	1.012	1.011
	1.14	1.104	1.050	1.033	1.025	1.020	1.017	1.015	1.013	1.011
	1.41	1.116	1.057	1.038	1.029	1.023	1.019	1.017	1.015	1.013
	2.00	1.141	1.072	1.049	1.038	1.031	1.026	1.023	1.020	1.018
	3.00	1.186	1.100	1.071	1.056	1.047	1.041	1.036	1.032	1.029
	4.00	1.233	1.131	1.095	1.076	1.065	1.057	1.050	1.045	1.041
	5.00	1.276	1.161	1.119	1.096	1.083	1.073	1.065	1.059	1.054
	10.00	1.452	1.284	1.221	1.184	1.163	1.147	1.133	1.123	1.114
	15.00	1.581	1.378	1.299	1.254	1.226	1.206	1.188	1.175	1.164
Pareto	3.00	1.191	1.104	1.073	1.057	1.047	1.040	1.034	1.030	1.027
	4.00	1.232	1.133	1.097	1.077	1.064	1.055	1.049	1.043	1.039
	5.00	1.265	1.157	1.116	1.094	1.079	1.069	1.061	1.055	1.050
	10.00	1.354	1.224	1.173	1.145	1.125	1.111	1.101	1.092	1.085
	15.00	1.392	1.253	1.199	1.168	1.147	1.131	1.119	1.110	1.102
Pearson	0.25	1.084	1.039	1.026	1.019	1.015	1.013	1.011	1.009	1.008
	0.50	1.088	1.042	1.028	1.021	1.016	1.013	1.011	1.010	1.009
	0.71	1.091	1.044	1.029	1.021	1.017	1.014	1.012	1.011	1.009
	1.00	1.099	1.048	1.032	1.024	1.019	1.016	1.014	1.012	1.010
	1.14	1.104	1.051	1.033	1.025	1.020	1.017	1.014	1.013	1.011
	1.41	1.112	1.055	1.037	1.027	1.022	1.018	1.016	1.014	1.012
	2.00	1.138	1.070	1.047	1.036	1.029	1.025	1.021	1.018	1.016
	3.00	1.201	1.104	1.071	1.054	1.044	1.036	1.031	1.027	1.025
	4.00	1.289	1.153	1.106	1.082	1.067	1.057	1.049	1.044	1.039
	5.00	1.390	1.212	1.146	1.114	1.094	1.080	1.069	1.061	1.055
Weibull	0.25	1.080	1.037	1.024	1.018	1.014	1.012	1.010	1.009	1.008
	0.50	1.083	1.038	1.024	1.018	1.015	1.012	1.011	1.009	1.008
	0.71	1.087	1.041	1.026	1.020	1.016	1.013	1.011	1.010	1.009
	1.00	1.095	1.045	1.029	1.022	1.017	1.014	1.012	1.011	1.010
	1.14	1.100	1.047	1.031	1.023	1.018	1.015	1.013	1.011	1.010
	1.41	1.111	1.053	1.035	1.026	1.021	1.017	1.015	1.013	1.012
	2.00	1.140	1.070	1.046	1.035	1.028	1.024	1.020	1.018	1.016
	3.00	1.199	1.104	1.071	1.055	1.044	1.038	1.033	1.029	1.026
	4.00	1.263	1.144	1.100	1.078	1.064	1.055	1.048	1.043	1.039
	5.00	1.327	1.184	1.131	1.104	1.086	1.075	1.066	1.058	1.053
	10.00	1.622	1.380	1.285	1.234	1.200	1.177	1.159	1.144	1.133
	15.00	1.861	1.551	1.422	1.352	1.304	1.273	1.247	1.225	1.209

Bias ratio is $\alpha(S)$ for standard deviation. The ratio is the population value over the mean of 100,000 samples.

TABLE 4. Bias Factors for Skew

	Skew, γ	Sequence Length n: 10	20	30	40	50	60	70	80	90
Distribution										
Gumbel	1.14	2.172	1.541	1.355	1.269	1.217	1.183	1.156	1.137	1.123
Log normal	0.25	1.903	1.381	1.246	1.183	1.141	1.116	1.099	1.086	1.076
	0.50	1.960	1.413	1.267	1.198	1.156	1.129	1.109	1.096	1.085
	0.71	2.019	1.450	1.295	1.220	1.176	1.147	1.126	1.111	1.099
	1.00	2.100	1.499	1.331	1.249	1.201	1.168	1.144	1.126	1.113
	1.14	2.161	1.534	1.359	1.268	1.221	1.188	1.163	1.144	1.139
	1.41	2.251	1.595	1.404	1.309	1.252	1.218	1.189	1.168	1.151
	2.00	2.528	1.773	1.545	1.428	1.358	1.307	1.276	1.248	1.220
	3.00	3.066	2.120	1.827	1.665	1.573	1.506	1.455	1.414	1.381
	4.00	3.641	2.498	2.134	1.931	1.813	1.727	1.659	1.607	1.563
	5.00	4.234	2.888	2.453	2.209	2.064	1.959	1.876	1.811	1.757
	10.00	7.247	4.880	4.087	3.636	3.362	3.161	3.002	2.876	2.773
	15.00	10.239	6.857	5.710	5.055	4.654	4.359	4.126	3.940	3.788
Pareto	3.00	2.744	1.954	1.701	1.570	1.484	1.425	1.381	1.345	1.316
	4.00	3.464	2.429	2.089	1.910	1.791	1.708	1.646	1.595	1.553
	5.00	4.202	2.922	2.495	2.269	2.118	2.013	1.933	1.868	1.813
	10.00	7.975	5.463	4.611	4.154	3.846	3.629	3.464	3.328	3.215
	15.00	11.784	8.038	6.762	6.075	5.613	5.285	5.035	4.830	4.659
Pearson	0.25	1.868	1.359	1.232	1.169	1.129	1.103	1.088	1.078	1.066
	0.50	1.925	1.407	1.255	1.186	1.146	1.119	1.103	1.088	1.080
	0.71	1.969	1.416	1.270	1.201	1.160	1.132	1.112	1.098	1.087
	1.00	1.963	1.430	1.279	1.207	1.165	1.138	1.119	1.104	1.093
	1.14	1.972	1.441	1.291	1.216	1.174	1.145	1.125	1.109	1.096
	1.41	1.978	1.450	1.302	1.226	1.182	1.151	1.130	1.113	1.100
	2.00	2.054	1.519	1.354	1.273	1.225	1.192	1.166	1.147	1.132
	3.00	2.233	1.650	1.464	1.364	1.302	1.260	1.228	1.204	1.186
	4.00	2.466	1.811	1.590	1.473	1.398	1.345	1.307	1.276	1.252
	5.00	2.735	1.982	1.724	1.588	1.499	1.436	1.389	1.352	1.323
Weibull	0.25	1.863	1.359	1.224	1.160	1.125	1.104	1.090	1.077	1.068
	0.50	1.778	1.326	1.205	1.148	1.116	1.097	1.083	1.072	1.063
	0.71	1.777	1.333	1.213	1.155	1.124	1.104	1.090	1.078	1.070
	1.00	1.795	1.349	1.226	1.165	1.132	1.110	1.095	1.082	1.073
	1.14	1.819	1.366	1.239	1.176	1.141	1.118	1.102	1.088	1.079
	1.41	1.874	1.402	1.265	1.197	1.158	1.133	1.114	1.099	1.088
	2.00	2.057	1.521	1.357	1.274	1.224	1.191	1.166	1.147	1.132
	3.00	2.441	1.770	1.553	1.438	1.367	1.319	1.282	1.253	1.230
	4.00	2.873	2.053	1.778	1.631	1.538	1.473	1.424	1.384	1.353
	5.00	3.325	2.352	2.019	1.838	1.722	1.642	1.579	1.529	1.490
	10.00	5.661	3.904	3.279	2.934	2.708	2.547	2.422	2.320	2.240
	15.00	7.990	5.463	4.548	4.041	3.706	3.466	3.280	3.129	3.008
Bias ratio $\alpha^*(G)$ for skew	—	2.194	1.543	1.352	1.260	1.207	1.171	1.146	1.128	1.113

Bias ratio is $\alpha(G)$ for skew. The ratio is the population value over the mean of 100,000 samples.

Log normal

$$y_i = c + \exp\,[m + a\eta_i] \tag{19}$$

Gumbel

$$y_i = m + a\{-n[-\ln u_i]\} \tag{20}$$

Weibull

$$y_i = m + a[-\ln(1 - u_i)]^{1/c} \tag{21}$$

Pareto

$$y_i = a(1 - u_i)^{-1/b} \tag{22}$$

Pearson

$$y_i = m + a\left\{-\ln \prod_{k=1}^{[b]} u_{ik} - B_i \ln u_i\right\} \tag{23a}$$

where B_i is distributed as beta, $B_i \sim (b - [b], 1 - b + [b])$, and where $[b]$ denotes the greatest integer equal to or less than b. The algorithm for B_i is as follows: (1) Set $r = b - [b]$; $s = 1 - r = 1 - b + [b]$. (2) Generate $u_1, u_2 \sim U[0, 1]$. (3) Set $\zeta = u_1^{1/r}$, $\xi = u_2^{1/s}$. (4) If $\zeta + \xi > 1$, return to 2; otherwise proceed to 5. (5) Set $B_i = \zeta/(\zeta + \xi)$. Note that if b is integral, $B_i = 0$, so that

$$y_i = m + a\left\{-\ln \prod_{k=1}^{b} u_{ik}\right\} \tag{23b}$$

or if $b < 1$, then $[b] = 0$, so that

$$y_i = m + a\{-B_i \ln u_{i0}\} \tag{23c}$$

The above algorithm for Pearson type 3 generation is due to *Jöhnk* [1964] and is described by *Berman* [1971].

To generate the various sets of pseudorandom numbers, the values of the parameters appearing in the algorithms were derived by solving the equations relating $\mu(Y)$, $\sigma(Y)$, and $\gamma(Y)$ to those parameters such that $\mu(Y) = 0$, $\sigma(Y) = 1$, and $\gamma(Y)$ equal the particular values shown in Table 1.

EXPERIMENTAL RESULTS

In the appendix to this paper (the appendix is not included but is available in paper copy or microfiche form through the National Technical Information Service, U.S. Department of Commerce, Springfield, Virginia 22151) the derived distribution functions are shown for $\bar{X}$, S, and G over the experimental ranges of sequence length n, coefficient of skewness $\gamma(Y)$, and probability density function $f(y)$. The values of $\tilde{\mu}(X)$, $\tilde{\sigma}(X)$, $\tilde{\gamma}(X)$, and $\tilde{\lambda}(X)$ are presented on each figure. It should be noted that for skews greater than five, 100,000 trials do not provide sufficient resolution of the upper tail of the sample distributions of the standard deviation. In these cases the tails above the probability 99.9 are of questionable accuracy.

From these results it is seen that the distributions of $\bar{X}$, S, and G are functions of n, $\gamma(Y)$, and $f(Y)$. As might be expected, $\bar{X}$ is an unbiased estimate of $\mu(Y)$, whereas S and G are biased estimates of $\sigma(Y)$ and $\gamma(Y)$. The biases in S and G as functions of $n, \gamma(Y)$, and $f(y)$, expressed as

$$\alpha(S) = \sigma(Y)/\tilde{\sigma}(Y) \tag{24}$$

$$\alpha(G) = \gamma(Y)/\tilde{\gamma}(Y) \qquad \tilde{\gamma}(Y) > 0 \tag{25}$$

are shown in Tables 3 and 4, respectively.

For flood flow frequency studies, the *Water Resources Council* [1967] recommended that the estimate of $\gamma(Y)$ be defined as

$$G' = n^{1/2}\left(\frac{(n-1)^{1/2}}{n-2}\right)G \tag{26}$$

where G is given by (3). In a prior publication the *Interagency Committee on Water Resources* [1966] suggested that G' be multiplied by the factor $(1 + 8.5/n)$ as given by *Hazen* [1930]. Thus for the estimate G the approximate factor of bias would be given by

$$\alpha^*(G) = n^{1/2}\left(\frac{(n-1)^{1/2}}{n-2}\right)(1 + 8.5/n) \tag{27}$$

and is shown as the last line of Table 4. From these figures it is noted that

$$G^* = [\alpha^*(G)]G \tag{28}$$

is an approximately unbiased estimate of $\gamma(Y)$ for a small range of values of $\gamma(Y)$, say, $\frac{1}{2} < \gamma(Y) < 2$ for the log normal distribution. However, these figures show that $\alpha(G)$, given by (25), is very nearly distribution free for $\gamma(Y) < 2^{1/2}$.

The distributions of G suggest that G is bounded. *Kirby* [1974] has shown that bounds are given by

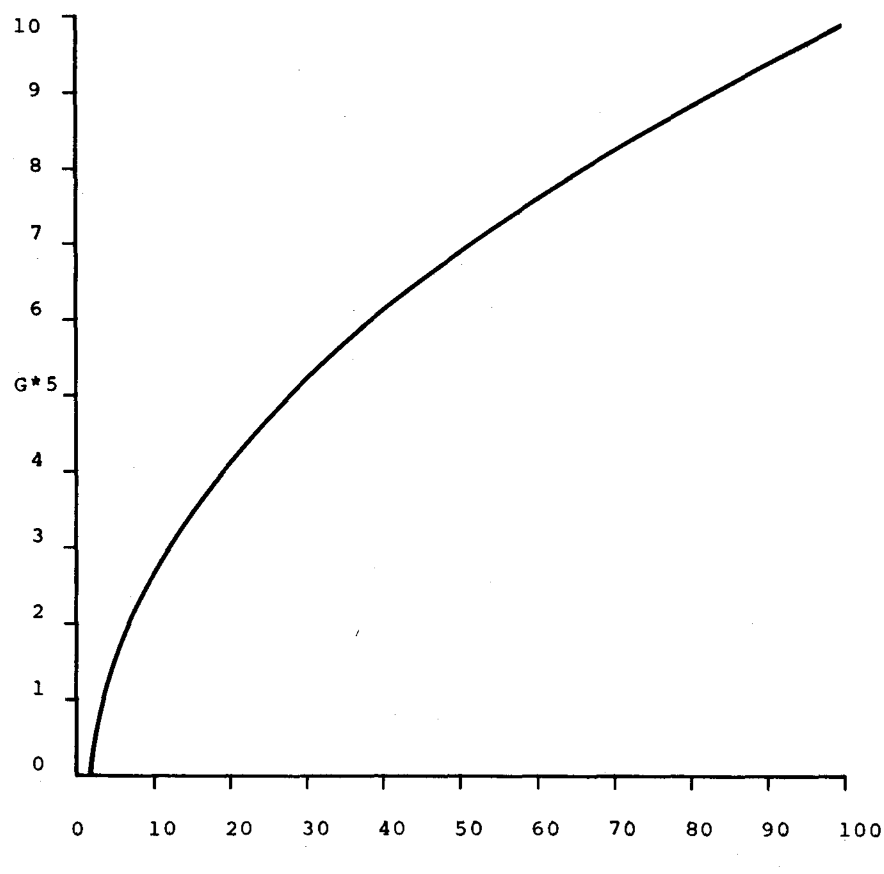

Fig. 2. Bound on coefficient skewness, with $|G| = G^* = (n-2)/(n-1)^{1/2}$.

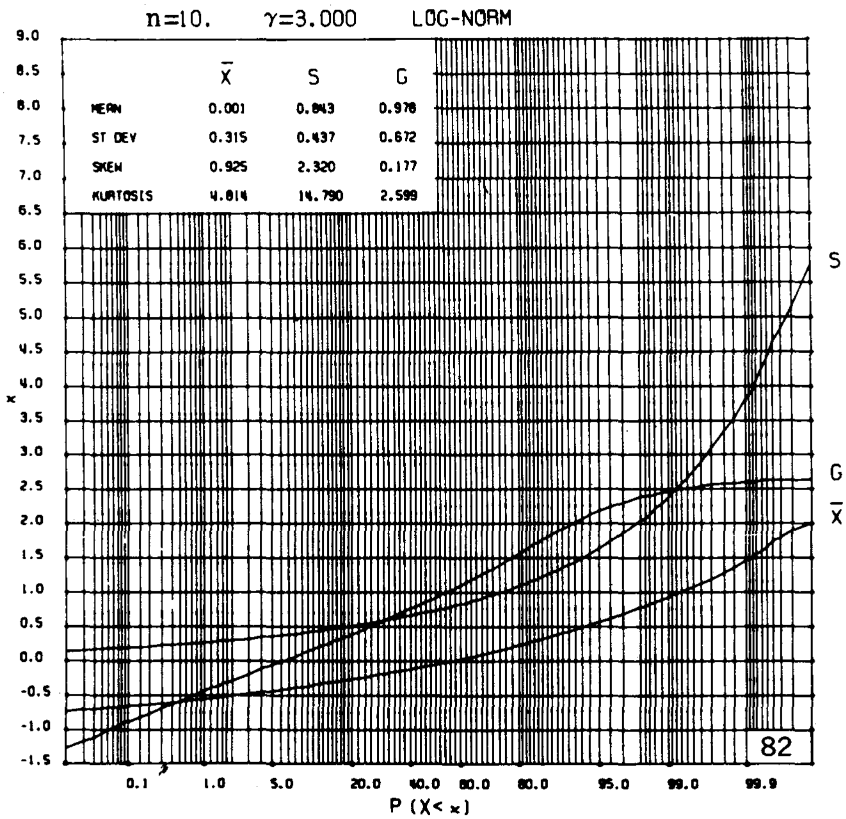

Fig. 3. Distribution functions of sample statistics for the log normal distribution.

$$|G| \leq (n - 2)/(n - 1)^{1/2} \tag{29}$$

and do not depend upon $\gamma(Y)$ or $f(y)$. Equation 29 is shown graphically in Figure 2. The bound on G is illustrated by the distribution of G for $n = 10$ and $\gamma(Y) = 3$ in Figures 3–6, which also appear in the appendix. The bounds on G will be discussed further in a subsequent paper dealing with the distribution of G for observed flood sequences.

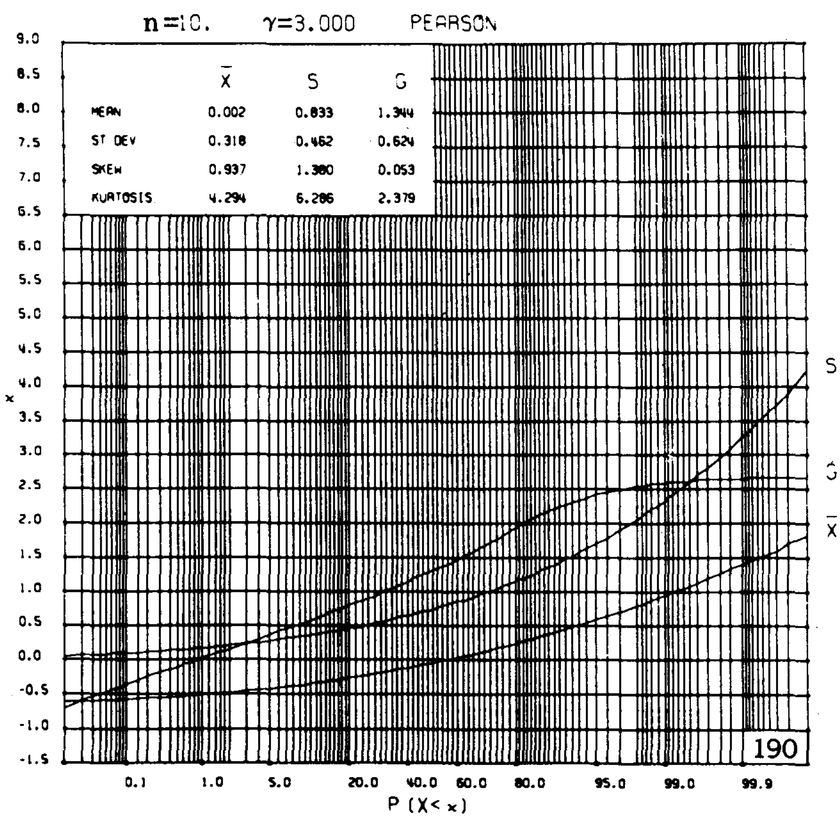

Fig. 4. Distribution functions of sample statistics for the Pearson distribution.

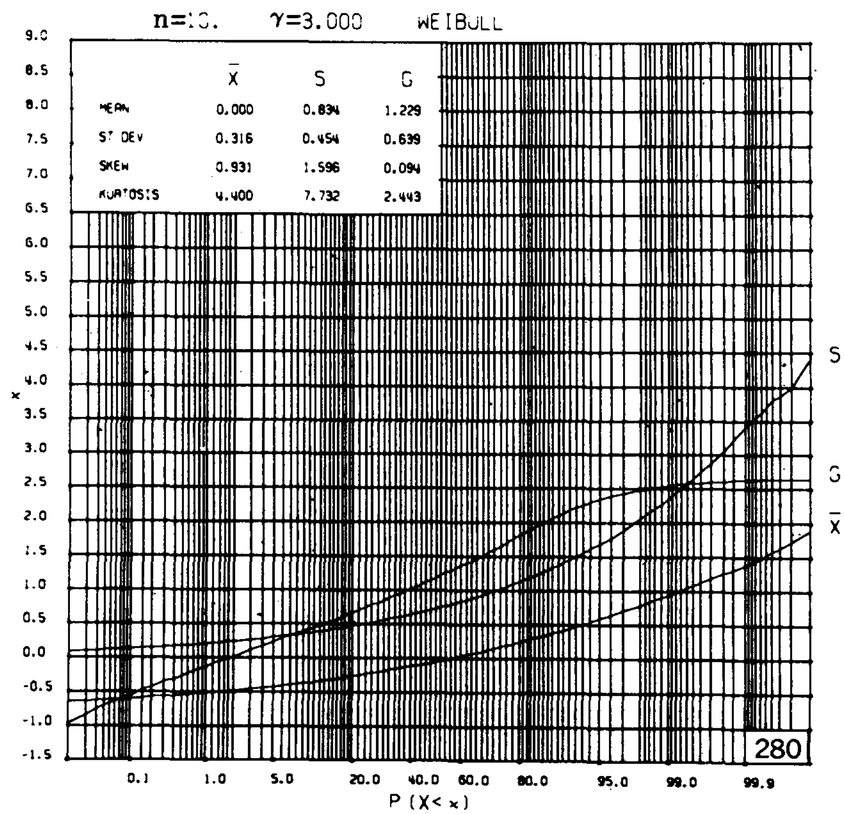

Fig. 5. Distribution functions of sample statistics for the Weibull distribution.

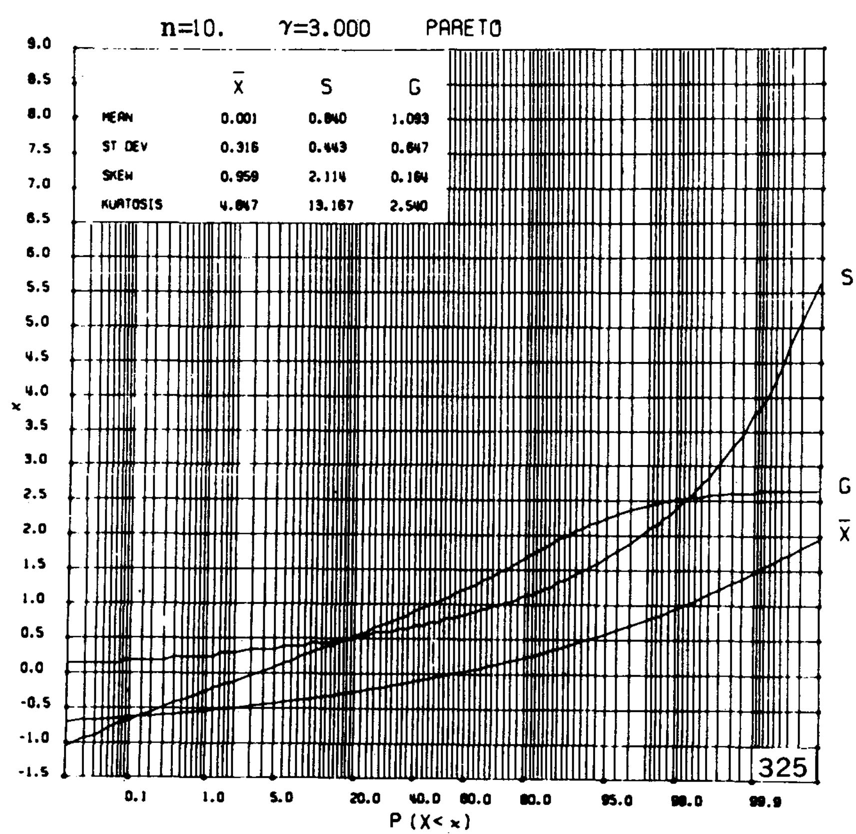

Fig. 6. Distribution functions of sample statistics for the Pareto distribution.

For certain cases some of the sampling properties of $\bar{X}$, S, and G are known theoretically. For example, given any distribution for which $\sigma(Y)$ is finite, then $\sigma(\bar{X}) = \sigma(Y)/(n)^{1/2}$. In a subsequent paper some known theoretical results will be compared with the corresponding Monte Carlo results as well as results on the goodness of fit of specific distributions to the derived distributions of $\bar{X}$, S, and G.

REFERENCES

Berman, M. B., Generating gamma distributed variates for computer simulation models, *Rep. R-641-PR,* 43 pp., Rand, Santa Monica, Calif., Feb. 1971.

Box, G. E. P., and M. E. Muller, A note on the generation of random normal deviates, *Ann. Math. Statist., 29,* 610–611, 1958.

Hazen, A., *Flood Flows,* John Wiley, New York, 1930.

Interagency Committee on Water Resources, Subcommittee on Hydrology, Methods of flow frequency analysis, *Bull. 13,* 42 pp., Washington, D. C., April 1966.

Jöhnk, M. D., Erzeugung von betaverteilten und gammaverteilten Zufallszahlen, *Metrika, 8*(1), 5–15, 1964.

Kirby, W., Algebraic boundedness of sample statistics, *Water Resour. Res.,* in press, 1974.

Lewis, P. A. W., A. S. Goodman, and J. M. Miller, A pseudorandom number generator for the System/360, *IBM Syst. J., 8*(2), 136–146, 1969.

Water Resources Council, A uniform technique for determining flood flow frequencies, *Bull. 15,* 15 pp., Washington, D. C., Dec. 1967.

(Received September 11, 1973;
revised October 29, 1973.)