CEE 621 Water Resources Planning II
Class Notes, J. R. Stedinger

Probability Weighted Moments, L-Moments,
The Generalized Extreme Value (GEV) Distribution,
and A GEV Index-Flood Procedure

Probability Weighted Moments are a different way to summarize the statistical properties of data sets. They can be defined quite generally [Greenwood *et al.*, 1979]. A definition often used in practice is that the order r PWM is

$$\beta_r = E\{ X [F(X)]^r \} \qquad (1)$$

were F(X) is the CDF for X. Instead of taking the expectation of X to some power to calculate a variance or skew, the probability weighted moments are the expectation of X times powers of F(X), its CDF. For r = 0, β_0 is just the population mean, μ_X.

To estimate β_r one can employ the order statistics $X_{(i)}$, $X_{(1)} \le \ldots \le X_{(n)}$, of a sample $\{X_i \mid i = 1, \ldots, n\}$. A natural estimator of β_r is:

$$b_r = (1/n) \sum X_{(j)} [p_j]^r \qquad (2)$$

where p_j are plotting positions assigned to each $X_{(j)}$. Hosking *et al.* [1985] suggest the APP ("A Plotting Position") formula:

$$p_i = (i-0.35)/n \qquad (3)$$

Alternatively, it is simple to show that $(r+1)\beta_r$ is the expected value of the smallest observation in a sample of size (r+1). Thus an unbiased estimator of β_r can be constructed by considering in a sample of size n every possible subsample of size (r+1) and calculating what would be the smallest observation. That calculation yields the unbiased PWM estimator:

$$b_r^{\text{unbiased}} = n^{-1} \sum_{i=1}^{n} \binom{i-1}{r} X_{(i)} / \binom{n-1}{r} \qquad (4)$$

for r = 0, 1, . . . , n-1 [Greenwood *et al.*, 1979].

L-Moments (Hosking, 1989) are an alternative, based on PWMs, to conventional descriptions of a distribution's shape, such as the coefficients of skewness and kurtosis equal to $E[(x-\mu)^r]/\sigma^r$ for r = 3,4. The drawbacks of the conventional

To obtain the software send the message
"Send lmoments from general" to the email adress:
statlib@lib.stat.cmu.edu.

CEE 621 Water Resources Planning II
Class Notes, J. R. Stedinger

product-moment sample coefficient of variation CV, and the sample coefficients of skewness and kurtosis, are that they are highly variable, and have a bias which depends upon the sample size and also the underlying distribution (Wallis, 1988; Wallis *et al.*, 1974; see Loucks *et al.*, 1981, Table 3.2 and discussion in section 3.2.3). In fact, for a sample of size n, the product-moment sample coefficient of variation cannot be larger than $(n-1)^{0.5}$, and the sample skewness cannot exceed $(n-2)/(n-1)^{0.5}$ (Kirby, 1974).

L-moments are an alternative means of describing such properties of a distribution, or of estimating coefficients of variation, skeweness, and kurtosis from samples based on probability weighted moments. Advantages of L-moments are that PWMs are linear combinations of the observations and thus do not involve squaring and cubing the observations. Also, using unbiased PWMs, unbiased analogues for the convential coefficients of variation, skewness and kurtosis result (Wallis, 1988).

L-moments are generally computed with unbiased PWMs. Let $X_{(i|n)}$ be the i^{th} largest observation in a sample of size n. Then a PWM-based estimator of scale could be based upon the expected difference between the largest and smallest observation in a sample of size 2:

$$\lambda_2 = (1/2)\, E[X_{(2|2)} - X_{(1|2)}] \qquad (5)$$

Similarly, measures of asymmetry and kurtosis are

$$\lambda_3 = (1/3)\, E[X_{(3|3)} - 2X_{(2|3)} + X_{(1|3)}]$$

$$\lambda_4 = (1/4)\, E[X_{(4|4)} - 3X_{(3|4)} + 3X_{(2|4)} - X_{(1|4)}] \qquad (6)$$

Here λ_3 is one-third the difference between the sum of the largest and smallest observation in a sample of size three, minus twice the sample median; it is positive for positively skewed distributions and negative for negatively skewed distributions. The L-kurtosis λ_4 can be written

$$\lambda_4 = (1/4)\{E[X_{(4|4)} - X_{(1|4)}] - 3\,E[X_{(3|4)} - X_{(2|4)}]\} \qquad (7)$$

and is roughly a measure of the difference between the 80-20 percentile range and 3 times the 60-40 percentile range. For highly kurtotic distributions, λ_4 will be larger; for the uniform distribution it has a value of zero.

L-moments are easily calculated in terms of PWMs using the relationships

$$\lambda_r = r^{-1} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} E[X_{r-j:r}]$$

CEE 621 Water Resources Planning II
Class Notes, J. R. Stedinger

$$\lambda_1 = \beta_0$$
$$\lambda_2 = 2\beta_1 - \beta_0$$
$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \qquad (8)$$
$$\lambda_4 = 20\,\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0$$

To obtain a dimensionless L-moment coefficient of variation (L-CV) one would employ the ratio λ_2/λ_1. The L-moment coefficient of skewness (L-Skewness) and kurtosis (L-Kurtosis) are

$$\tau_r = \lambda_r/\lambda_2 \qquad r = 3,4 \qquad (9)$$

L-moment ratios are all bounded so that $|\tau_r| < 1$. Values of λ_2, τ_3 and τ_4 are given below for several distributions. (Note, $\lambda_1 = \beta_0 = E[X]$).

Because the first r L-moments are linear combinations of the first r PWMs, fitting a distribution so as to reproduce the first r sample L-moments is equivalent to using the corresponding sample PWMs. The advantage of the use of L-moment ratios is that they are dimensionless and thus describe the fundamental characteristics of the distributions of different random variables.

Values of L-Moments for Several Distributions

Distribution	L-Moments
Uniform	$\lambda_2 = (\beta-\alpha)/6;\ \tau_3 = 0;\ \tau_4 = 0$
Exponential	$\lambda_2 = \beta/2;\ \tau_3 = 1/3;\ \tau_4 = 1/6$
Normal	$\lambda_2 = \sigma/\sqrt{\pi};\ \tau_3 = 0;\ \tau_4 = 0.1226$
Gumbel	$\lambda_2 = \alpha \ln(2);\ \tau_3 = 0.1699;\ \tau_4 = 0.1504$
GEV	$\lambda_2 = \alpha\,(1-2^{-\kappa})\,\Gamma(1+\kappa)/\kappa;\ \tau_3 = 2\,(1-3^{-\kappa})/(1-2^{-\kappa}) - 3$ and $\tau_4 = \{1 - 5\,4^{-\kappa} + 10\,3^{-\kappa} - 6\,2^{-\kappa}\}\,/\,(1-2^{-\kappa})$
Generalized Pareto	$\lambda_2 = \alpha/[(1+\kappa)(2+\kappa);\ \tau_3 = (1-\kappa)/(3+\kappa)$ and $\tau_4 = (1-\kappa)(2-\kappa)/[(3+\kappa)(4+\kappa)]$
Lognormal	See appendix
Gamma	See appendix

(From Hosking, 1989; for the Exponential distribution $F[x] = 1 - e^{-x/\beta}$; and for Generalized Pareto $x = \xi + (\alpha/\kappa)\{1 - [1-F]^{\kappa}\}$ or $F[x] = 1 - [1 - \kappa(x-\xi)/\alpha]^{1/\kappa}$.

Ref: Hosking, J. and Wallis, J. (1997). Regional Frequency Analysis, Cambridge University Press.

Distribution and inverse cdf	L moments
Uniform:	$\lambda_1 = \frac{\beta+\alpha}{2} \quad \lambda_2 = \frac{\beta-\alpha}{6}$
$x = \alpha + (\beta - \alpha)F$	$\tau_3 = \tau_4 = 0$
Exponential:*	$\lambda_1 = \xi + \frac{1}{\beta} \quad \lambda_2 = \frac{1}{2\beta}$
$x = \xi - \frac{\ln[1-F]}{\beta}$	$\tau_3 = \frac{1}{3} \quad \tau_4 = \frac{1}{6}$
Normal†	$\lambda_1 = \mu \quad \lambda_2 = \frac{\sigma}{\sqrt{\pi}}$
$x = \mu + \sigma\Phi^{-1}[F]$	$\tau_3 = 0 \quad \tau_4 = 0.1226$
Gumbel:	$\lambda_1 = \xi + 0.5772\,\alpha \quad \lambda_2 = \alpha \ln 2$
$x = \xi - \alpha \ln[-\ln F]$	$\tau_3 = 0.1699 \quad \tau_4 = 0.1504$
GEV:	$\lambda_1 = \xi + \frac{\alpha}{\kappa}(1 - \Gamma[1+\kappa]) \quad \lambda_2 = \frac{\alpha}{\kappa}(1-2^{-\kappa})\,\Gamma(1+\kappa)$
$x = \xi + \frac{\alpha}{\kappa}(1 - [-\ln F]^{\kappa})$	$\tau_3 = \left\{\frac{2(1-3^{-\kappa})}{(1-2^{-\kappa})} - 3\right\}$
	$\tau_4 = \frac{1 - 5(4^{-\kappa}) + 10(3^{-\kappa}) - 6(2^{-\kappa})}{1-2^{-\kappa}}$
Generalized Pareto:	$\lambda_1 = \xi + \frac{\alpha}{1+\kappa} \quad \lambda_2 = \frac{\alpha}{(1+\kappa)(2+\kappa)}$
$x = \xi + \frac{\alpha}{\kappa}(1 - [1-F]^{\kappa})$	$\tau_3 = \frac{1-\kappa}{3+\kappa} \quad \tau_4 = \frac{(1-\kappa)(2-\kappa)}{(3+\kappa)(4+\kappa)}$
Lognormal	See Eqs. (18.2.12), (18.2.13)
Gamma	See Eqs. (18.2.30), (18.2.31)

* Alternative parameterization consistent with that for Pareto and GEV distributions is: $x = \xi - \alpha \ln[1-F]$ yielding $\lambda_1 = \xi + \alpha$; $\lambda_2 = \alpha/2$.
† Φ^{-1} denotes the inverse of the standard normal distribution (see Sec. 18.2.1).
Note: F denotes cdf $F_X(x)$.
Source: Adapted from Ref. 72, with corrections.

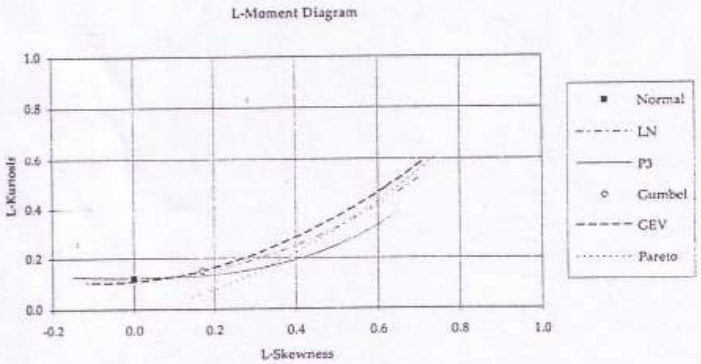


FIGURE 18.1.1 This L-moment diagram illustrates the relationship between the L-kurtosis τ_4 and the L-skewness τ_3 for the normal, lognormal (LN), Pearson type 3 (P3), Gumbel, generalized extreme value (GEV), and Pareto distributions.

TABLE 18.1.1 Definitions of Dimensionless Product-Moment and L-Moment Ratios

Name	Denoted	Definition
Product-moment ratios		
Coefficient of variation	CV_X	σ_X/μ_X
Coefficient of skewness*	γ_X	$\frac{E[X-\mu_X)^3}{\sigma_X^3}$
Coefficient of Kurtosis†		$\frac{E[X-\mu_X)^4}{\sigma_X^4}$
L-moment ratios		
L-coefficient of variation‡	L-CV, τ_2	λ_2/λ_1
L-coefficient of skewness	L-skewness, τ_3	λ_3/λ_2
L-coefficient of kurtosis	L-kurtosis, τ_4	λ_4/λ_2

* Some texts define $\beta_1 = [\gamma_X]^2$ as a measure of skewness.
† Some texts define the kurtosis as $[E[(X-\mu_X)^4]/\sigma_X^4 - 3]$; others use the term *excess kurtosis* for this difference because the normal distribution has a kurtosis of 3.
‡ Hosking[73] uses τ instead of τ_2 to represent the L-CV ratio.

CEE 621 Water Resources Planning II
Class Notes, J. R. Stedinger

A GEV Index-Flood procedure

1) At each site k compute the L-moments estimators $\lambda_1^k, \lambda_2^k, \lambda_3^k$.

2) For the region, calculate the average sample-size-weighted value of the normalized L-moments for the region λ_r^R of order r = 2 and 3 across all sites:

$$\lambda_r^R = \sum n_k (\lambda_r^k/\lambda_1^k) / (\sum n_k), \quad r = 2,3 \qquad (19)$$

where n_k is the length of record at site k, and λ_1^R is simply 1.

3) Using the λ_1^R, λ_2^R, and λ_3^R in equations (15)-(17) estimate the parameters and quantiles of the regional dimensionless GEV distribution.

4) The estimate of the 100p-percentile of the flood distribution at any site k is then

$$\hat{x}_p^k = \hat{\lambda}_1^k \hat{x}_p \qquad (20)$$

where $\hat{\lambda}_1^k$ is the sample mean for site k and $\hat{x}_p$ is the estimated index flood quantile for the region, obtained from eqn. (17).

Jin and Stedinger [1989] suggest it may be better not to weight by the sample sizes if some sites have much longer records than others, so as not to give them undue weight. However, if some sites have very short records, some weighting would be advantageous; the optimal weights depend both upon the heterogeneity of the region and the sample sizes. Jin and Stedinger illustrate the value of incorporating historical information into step 4; they employ generalized MLE's to estimate, using both systematic and historical at-site information, the scale parameter to go with the index flood distribution.

A key to the success of the index flood approach is the identification of reasonably similar sets of basins. L-moments can help in that regard. Lettenmaier *et al.*, (1987) show that modest heterocedasticity can be tolerated, but degrade the procedure's performance. There is also the question of what to do at sites which are judged to be reasonably unique.

Selected References

Greenwood, J.A., J.M. Landwehr, N.C. Matalas, and J.R. Wallis, Probability Weighted Moments: Definitions And Relation To Parameters Of Several Distributions Expressed In Inverse Form, *Water Resour. Res.*, *15*(5), 1049-54, 1979.

CEE 621 Water Resources Planning II
Class Notes, J. R. Stedinger

Hosking, J. R. M., J.R. Wallis, and E.F. Wood, Estimation of the generalized extreme-value distribution by the method of probability weighted moments, *Technometrics*, 27(3), 251-261, 1985.

Hosking, J.R.M., L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics, *Jour. of Royal Statistical Society, B*, 51(3), 1989.

Jin, M., and J.R. Stedinger, Flood Frequency Analysis with Regional and Historical Information, *Water Resour. Res.*, 25(5), 925-36, 1989.

Kirby, W., Algebric Boundness of Sample Statistics, *Water Resour. Res.*, 10(2), 220-222, 1974.

Lettenmaier, D.P., J.R. Wallis, and E.F. Wood, Effects Of Regional Heterogeneity On Flood Frequency Estimation, *Water Resour. Res.*, 23(2), 313-23, 1987.

Stedinger, J.R., Estimating a Regional Flood Frequency Distribution, *Water Resour. Res.*, 19(2), 503-10, 1983.

Wallis, J.R., N.C. Matalas, and J.R. Slack, Just a Moment!, *Water Resour. Res.*, 10(2), 211-221, 1974.

Wallis, J.R., Catastrophes, Computing and Containment: Living in our restless habitat, *Specultation in Science and Technology*, 11(4), 295-315, 1988.

June19, 1989

Chapter 4 Discrete Probability Distributions

· Bernoulli Distribution

- description

  (i) there are only two possible outcomes, called a success or failure

  (ii) the probability of occurrence of a success (or a failure) is constant

  (iii) the probability of the event occurring is independent of the time and independent of the past history of occurrences.

- probability mass function

$$f_X(x) = p^x(1-p)^{1-x} \qquad \text{for } x = 0,\ 1 \text{ and } 0 \le p \le 1$$

- population parameters

mean	= E[X]	=	p
variance	= Var[X]	=	$(1-p)p$

· Binomial Distribution

- description

  (i) a series of Bernoulli trials is made

  (ii) the trials are conducted under the same conditions

  (iii) the order in which the events in the trials occur is immaterial

- probability mass function

$$f_X(x) = f_X(x; n, p) = \binom{n}{x} p^x q^{n-x} \quad \text{for} \quad x = 0, 1, 2, \ldots, n$$

  where n = number of trials

  x = number of successes

  p = probability of success in any trial

  $q = 1 - p$ = probability of failure

- population parameters

  mean E[X] = np

  variance Var[X] = npq

  skewness coef. = $(1\text{-}2p)/(npq)^{1/2}$

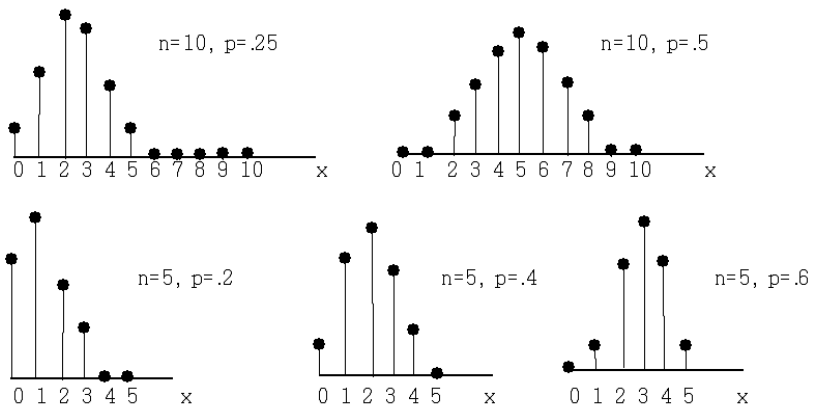
- parameter estimates

$$\hat{p} = \overline{X}/n$$

- comment

  additive property: If X~B(n1,p) and Y~B(n2,p), then Z~B(n1+n2,p) where Z = X+Y and B stands for binomial distribution (See exercise 4.17)

- distribution shape



- examples 4.4 & 4.7

· Negative Binomial Distribution

- description: a series of Bernoulli trials that are continued until exactly k successes occur when x trials are required
- probability mass function

$$f_X(x, k, p) = \binom{x-1}{k-1} p^k q^{x-k} \qquad \text{for } x = \text{k, k+1, k+2, …}$$

where x = number of trials
k = number of successes
p = probability of success in any trial
$q = 1 - p$ = probability of failure

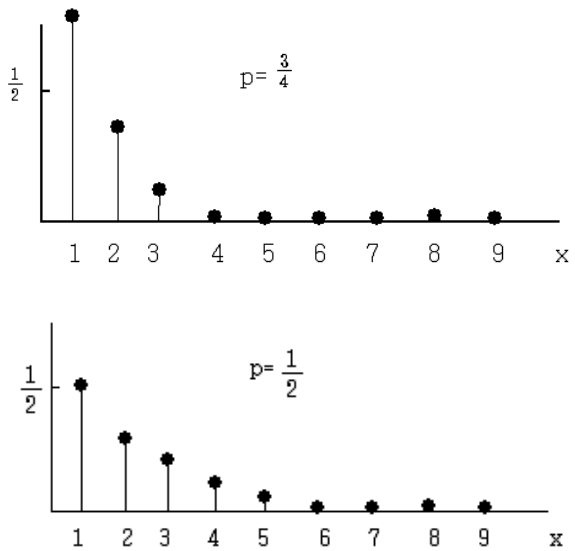
- population parameters

mean = E[X] = k/p
variance = Var[X] = $k(1-p)/p^2$

- comments

(i) for $k = 1$, the negative binomial distribution reduces to the geometric distribution (i.e. $f_X(x) = pq^{x-1}$)

- distribution shape



- example 4.11

· Geometric Distribution

- description: the probability that the first occurrence is at the x-th time
- probability mass function

$$f_X(x;p) = p(1-p)^{x-1} \qquad \text{for } x = 1, 2, 3, \ldots \text{ and } 0 \le p \le 1$$

- population parameters

mean = E[X] = $1/p$,

variance = Var(X) = $(1-p)/p^2$

- comment: the probability distribution of the length of time between occurrences
- example 4.9, 4.10

· Hypergeometric Distribution

- description

(i) drawing a random sample of size n without replacement.

(ii) from a finite population of size N with the elements of the population divided into two groups with k elements belonging to one group

- probability mass function

$$f_X(x;N,n,k) = \binom{k}{x}\binom{n-k}{n-x} / \binom{N}{n} \quad \text{for } x = 0, 1, 2, \ldots, n$$

where N= size of (finite) population,

n= size of sample,

k= number of successes in the population, and

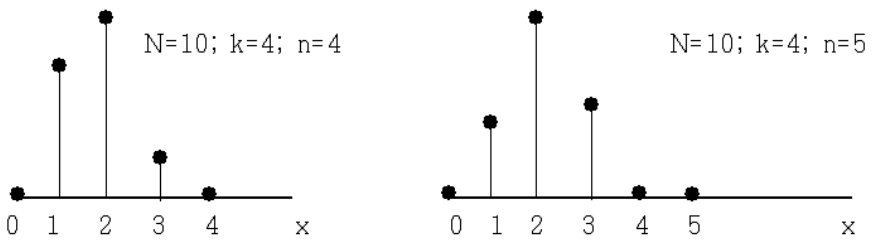
x= number of successes in the sample.

- population parameters

mean = E[X] = nk/N

variance = Var[X] = $nk(N-k)(N-n)/N^2(N-1)$

- comments: if n is "small" compared to N, the binomial distribution is reasonable approximation for the hypergeometric distribution
- distribution shape



- example 4.2

• Poisson Distribution

- description: a Bernoulli process defined over an interval of time (or space) so that p is the probability that an event may occur during the time interval

  when $\triangle t \to 0$, $p \to 0$, $np =$ constant

- probability mass function

$$f_X(x) = f_x(x;\lambda) = \frac{e^{\lambda}\lambda^x}{x!} \qquad \text{for } x = 0, 1, 2, \ldots$$

- cumulative mass function

$$F_X(x\lambda) = \sum_{i=0}^{x} \lambda^i \frac{e^{-\lambda}}{i!}$$

  where $x =$ number of occurrences of a specific event

  $\lambda =$ population parameter

- population parameters

  mean = E[X] $= \lambda$

  variance = Var[X] $= \lambda$

  skewness coef. $= \lambda^{-1/2}$

- parameter estimates

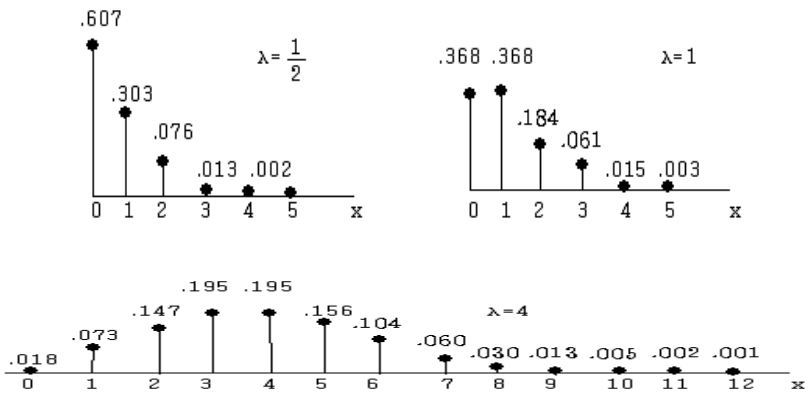
$$\hat{\lambda} = \overline{X}$$

- comments

  (i) Poisson is a convenient approximation for the binomial distribution when n is "large" and p is "small"

  (ii) the sum of two Poisson random variables with parameters λ_1 and λ_2 is a Poisson random variable with parameters $\lambda = \lambda_1 + \lambda_2$

- distribution shape



- example: May days in Denver receiving more than 0.10″ of rain from period 1878 to 1978

No of Days(x)	No of Occur	No of Total Occur Days	Binomial	Poisson	Observed
0	3	0	0.003	0.005	0.030
1	5	5	0.020	0.027	0.050
2	8	16	0.060	0.072	0.079
3	9	27	0.120	0.126	0.089
4	13	52	0.172	0.165	0.129
5	21	105	0.190	0.174	0.208
6	14	84	0.169	0.153	0.139
7	10	70	0.123	0.115	0.099
8	6	48	0.076	0.076	0.059
9	4	36	0.040	0.044	0.040
10	3	30	0.018	0.023	0.030
11	2	22	0.007	0.011	0.020
12	2	24	0.002	0.005	0.020
13	1	13	0.001	0.002	0.010
14	0	0	0.0003	0.0008	0.000
Total	101	532			

[column 4] $f(x;n,p)=\binom{n}{x}p^x(1-p)^{n-x}$

[column 5] $f(x;\lambda)=[\lambda^x e^{-\lambda}]/x!$

[column 6] = [column 2]/101

- examples 4.14 & 4.16

Chapter 5 & 6 Continuous Probability Distributions

<Basic Distributions>

Lecture 8
Advanced Hydrology
Dr. Kim, Young-Oh

• Uniform Distribution

- probability distribution function

$p_X(x) = 1/(\beta - \alpha)$ for $\alpha \le x \le \beta$

where α = population parameter, and

β = population parameter.

- population parameter

mean = E[X] = $(\alpha + \beta)/2$

variance = Var[X] = $(\beta - \alpha)^2/12$

- parameter estimates

$\hat{\beta} = x_{max}$ $\hat{\alpha} = \bar{x} - \sqrt{3}s$

$\hat{\alpha} = x_{min}$ $\hat{\beta} = \bar{x} + \sqrt{3}s$

- comments

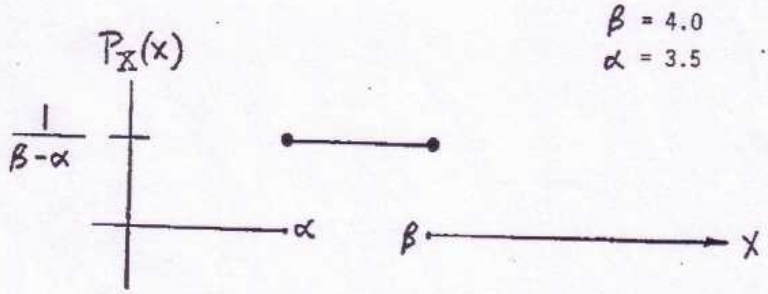
(i) see comment on Example 6.1 of the text (p.98)

- distribution shape

what is $p_X(x)$ if:

β = 4.0

α = 3.5



• Exponential Distribution

- probability distribution function

$$p_X(x) = \lambda e^{-\lambda x} = f_T(t;\lambda) \qquad \text{for } x > 0$$

where x = continuous random variable,
λ = population parameter, and
e = 2.718218...

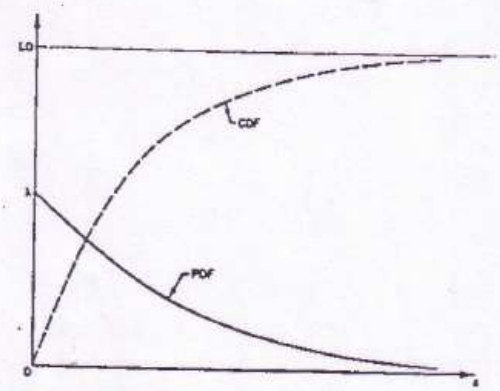
- population parameters

mean	= E[X]	$= 1/\lambda$
variance	= Var[X]	$= 1/\lambda^2$
skewness coef.		= 2

- parameter estimates

$$\lambda = 1/\bar{x}$$

- comments: special case of the gamma distribution ($\eta = 1$)
- distribution shape



PDF and CDF of the exponential distribution

From the Poisson Process, the probability of no occurrences of the r.v. X is

$$f_X(0;\lambda t) = e^{-\lambda t} = prob(T > t)$$ two adjacent

where T is the time between occurrences as the r.v.

$\therefore$ $prob(T \geq t) = 1 - e^{-\lambda t}$

(the waiting time between successive events of a Poisson process)

$\leq$

• Gamma Distribution

- description: the probability dist of the time to the nth occurrence, which is the sum of n independent r.v. $T_1 + T_2 + \ldots + T_n$ form the exponential distribution.
- probability distribution function

$$p_X(x) = \lambda^{\eta} x^{\eta-1} e^{-\lambda x} / \Gamma(\eta) = f_T(t; n, \lambda) \qquad \text{for } x \geq 0$$

wher $\Gamma(\eta) =$ gamma function,
$\eta =$ population parameter (shape),
$\lambda =$ population parameter (scale), and
$x =$ random variable.

Change $\eta \rightarrow n$

- population parameter:

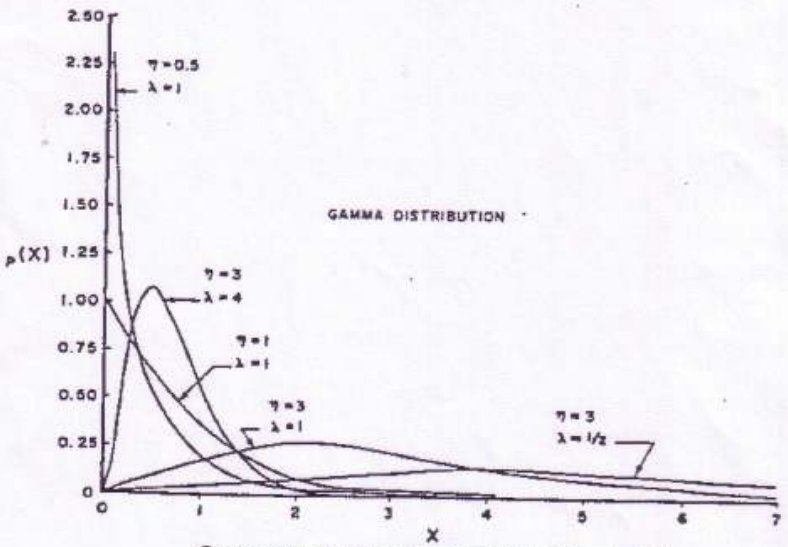
mean	= E[X]	=	η/λ
variance	= Var[X]	=	$\eta/(\lambda)^2$
skew		=	$2/(\eta)^{1/2}$

- parameter estimates (MoM, See p.102 ~103 for more.)

$$\hat{\lambda} = \bar{x}/s^2$$

$$\hat{\eta} = \bar{x}^2/s^2$$

- comments: the log-Pearson type III distribution is a 3 parameter gamma distribution
- distribution shape(s)



Gamma distribution with several values for η and λ.

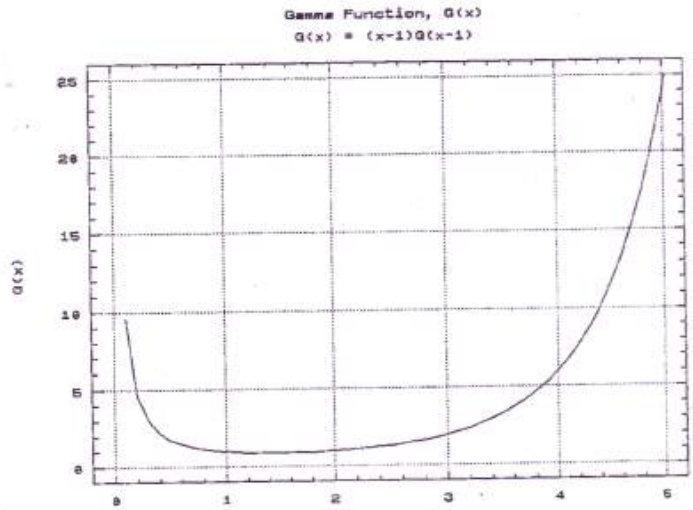
Table E.12. Gamma Function

Values of $\Gamma(n) = \int_0^\infty x^{n-1}e^{-x}\,dx$; $\Gamma(n+1) = n\Gamma(n)$

n	Γ(n)	n	Γ(n)	n	Γ(n)	n	Γ(n)
1.00	1.00000	1.25	.90640	1.50	.88623	1.75	.91906
1.01	.99433	1.26	.90440	1.51	.88659	1.76	.92137
1.02	.98884	1.27	.90250	1.52	.88704	1.77	.92376
1.03	.98355	1.28	.90072	1.53	.88757	1.78	.92623
1.04	.97844	1.29	.89904	1.54	.88818	1.79	.92877
1.05	.97350	1.30	.89747	1.55	.88887	1.80	.93138
1.06	.96874	1.31	.89600	1.56	.88964	1.81	.93408
1.07	.96415	1.32	.89464	1.57	.89049	1.82	.93685
1.08	.95973	1.33	.89338	1.58	.89143	1.83	.93969
1.09	.95546	1.34	.89222	1.59	.89243	1.84	.94261
1.10	.95135	1.35	.89115	1.60	.89352	1.85	.94561
1.11	.94739	1.36	.89018	1.61	.89468	1.86	.94869
1.12	.94359	1.37	.88931	1.62	.89592	1.87	.95184
1.13	.93993	1.38	.88854	1.63	.89724	1.88	.95507
1.14	.93642	1.39	.88785	1.64	.89864	1.89	.95838
1.15	.93304	1.40	.88726	1.65	.90012	1.90	.96177
1.16	.92980	1.41	.88676	1.66	.90167	1.91	.96523
1.17	.92670	1.42	.88636	1.67	.90330	1.92	.96878
1.18	.92373	1.43	.88604	1.68	.90500	1.93	.97240
1.19	.92088	1.44	.88580	1.69	.90678	1.94	.97610
1.20	.91817	1.45	.88565	1.70	.90864	1.95	.97988
1.21	.91558	1.46	.88560	1.71	.91057	1.96	.98374
1.22	.91311	1.47	.88563	1.72	.91258	1.97	.98768
1.23	.91075	1.48	.88575	1.73	.91466	1.98	.99171
1.24	.90852	1.49	.88595	1.74	.91683	1.99	.99581
						2.00	1.00000

* For large positive values of z, $\Gamma(z)$ approximates the asymptotic series

$$e^{-z}z^{z-\frac{1}{2}}\sqrt{2\pi}\left[1+\frac{1}{12z}+\frac{1}{288z^2}-\frac{139}{51840z^3}-\frac{571}{2488320z^4}+\cdots\right]$$



< Normal Family >

• Normal Distribution

- pdf

- cdf

- comments

(i) 2 parameter distribution

(ii) skewness coef. = 0, i.e. symmetrical about the mean

(iii) unbounded but if μ is greater than 3σ, the chances of X less than 0 are negligible in practice

(iv) reproductive properties

- standard normal distribution

- central limit theorem

(example 5.6)

• (Two Parameter) Lognormal Distribution

- random variables

X: log-normally distributed

Y = LN(X-a): normally distributed

- features

(i) population parameters

$$\mu_x = \exp(\mu_Y + \frac{\sigma_Y^2}{2})$$

$$\sigma_X^2 = \mu_X^2[\exp(\sigma_Y^2) - 1]$$

$$\gamma = 3CV_X + CV_X^3$$

(ii) bounded

• Three Parameter Lognormal Distribution(LN3)

- description

$$Y = \ln(X-\xi) \sim N(\mu_Y, \sigma^2_Y)$$

$$X = \xi + \exp(Y)$$

$$x_p = \xi + \exp(\mu_Y + \sigma_Y z_p)$$

- moments of X

$$\mu_x = \xi + \exp(\mu_Y + \frac{\sigma_Y^2}{2})$$

$$\sigma_X^2 = [\exp(2\mu_X + \sigma_Y^2)][\exp(\sigma_Y^2) - 1]$$

$$\gamma_X = 3[\exp(\sigma_Y^2) - 1]^{1/2} + [\exp(\sigma_Y^2) - 1]^{3/2}$$

- quantile lower bound estimator

(Stedinger, J. R. (1980). "Fitting Log Normal Distribution to Hydrologic Data", WRR 16(3), pp.481-490)

$$\hat{\xi} = \frac{x_1 x_n - x_{median}^2}{x_1 + x_n - 2x_{median}}$$

where x_1 and x_n are the largest and smallest observed values, respectively, and $x_{median} = x_{k+1}$ for odd sample size n=2k+1 and $x_{median} = 0.5(x_k + x_{k+1})$ for even n=2k.

((note)) When $x_1+x_n-2x_{median} < 0$, the formula provides an upper bound so that

$\ln(\xi-x) \sim$ normal ξ

WATER RESOURCES RESEARCH, VOL. 17, NO. 1, PAGES 200-208, FEBRUARY 1981

Estimating Correlations in Multivariate Streamflow Models

JERY R. STEDINGER

Department of Environmental Engineering, Cornell University, Ithaca, New York 14853

Multivariate log normal distributions are often used to model and generate multisite and multiseason streamflow sequences. A Monte Carlo study evaluated alternative estimators of the cross correlation between autocorrelated streamflow series and the lag 1 autocorrelation of a single series when some or all of the flows have a log normal distribution. Generally, smaller mean square error estimates of the correlations of flows are obtained by using the variances and covariances of the flow's logarithms to estimate a streamflow model's parameters. Sometimes it is advantageous to prewhiten two series before calculating their cross correlation and to unbias the estimated autocorrelation of the streamflow's logarithms.

Generation of seasonal streamflow sequences at a single or at several sites, for use in simulation studies of water resource systems, requires specification of the joint distribution of the flows. Physically reasonable, flexible, and mathematically tractable multivariate probability distributions are few. A common and recommended practice in hydrology is to transform the streamflows to be modeled into normally distributed random variables by taking logarithms [*Loucks et al.*, 1980] or some other Box-Cox transformation [*Hipel and McLeod*, 1978; *Box and Cox*, 1964]. Assuming that these normally distributed random variables have a joint normal distribution, a multivariate streamflow model is easily constructed. An unresolved problem is how best to estimate the covariances of the normally distributed transforms of the streamflows, the natural parameters of the multivariate normal distribution.

In the case of a single time series, such as the annual flows at a single site, the natural parameters are the lagged autocovariances of the normal transforms of the flows. When annual flows are disaggregated to seasonal flows or when flows at several sites are generated concurrently, the covariances of the normal transforms of the various flows become the natural parameters of the joint normal distribution. Two basically different approaches have been used to estimate these parameters. (1) Some individuals have recommended selecting the covariances of the transformed streamflows so as to reproduce the observed or historic correlation of the flows or record [*Burges*, 1972]. (2) Others have simply reproduced the observed covariances of the transformed flows [e.g., *Young and Pisano*, 1968]. Various authors have derived and reported relationships for preserving the observed correlation of the streamflows when the logarithms and even when some of the flows themselves are normally distributed [*Matalas*, 1967; *Burges*, 1972; *Mejia and Rodriguez-Iturbe*, 1974; *Mejia et al.*, 1974].

The first approach is often advanced with the justification that the characteristics of the actual streamflows and not the transforms are the important quantities to reproduce [*Mejia et al.*, 1974]. The second approach is often adopted because of its ease; the covariances of the normal transform of the streamflows are estimated directly rather than as nonlinear transformations of the observed correlations of the actual flows. This paper takes a rigorous look at these two approaches and evaluates their statistical efficiency. It is assumed throughout that the flows have either a two-parameter log normal or a normal distribution.

Paper number 80W1159.
0043-1397/81/080W-1159$01.00

Estimators and Estimation Criteria

Consider first the estimation of the covariance of two log normally distributed random variables x_i and y_i. These could be seasonal or annual flows at the same site or flows at different sites. To generate synthetic streamflow sequences, a multivariate model of the two stochastic processes must be selected, and its parameters estimated. The problems of model selection, though important, are not addressed here. Problems associated with fitting the marginal distributions are addressed by *Stedinger* [1980]. The question specifically addressed here is how best to estimate the parameters which determine the estimated correlation of the flows.

In particular, one should consider both the feasibility and accuracy of the estimated parameters. If the covariance matrix of the multivariate normal distribution describing the transformed flows is estimated so as to reproduce the observed correlations of the flows themselves, then this matrix may not be positive semidefinite. Hence it would not define a legitimate covariance matrix. This problem has been reported by *Hoshi et al.* [1978] and *Hoshi and Burges* [1979]. The estimated covariances of the transformed flows necessarily yield a positive definite covariance matrix if the flow records used at each site are of the same length. Hence the second estimation procedure's parameters are necessarily feasible, while those of the first estimation procedure are not necessarily so.

As was discussed by *Stedinger* [1980], criteria for evaluating the statistical performance of alternative parameter estimators should reflect the impact that misspecification of those parameters might have on the planning process, its recommendations, and the social benefits achieved. Unfortunately, it is very difficult to determine the impact of an error in one or more streamflow model parameters on the planning process and its products. As a simple surrogate for potential losses this study uses as its statistical criterion the root mean square error with which the alternative estimation procedures reproduce the true correlation of the streamflows. Thus the true correlation of the streamflows is considered to be more important or fundamental than the correlations of the transformed flows, which are the natural parameters of the standardized multivariate normal model. The root mean square error is a convenient and widely used criterion which combines the bias and variance of an estimator.

< Pearson Type III Family >

▪ Pearson Type III Distribution (P3)

- use in hydrology: distribution of flood peak
- pdf:

$$f_X(x) = |\beta|[\beta(x-\xi)]^{\alpha-1}\frac{\exp[-\beta(x-\xi)]}{\Gamma(\alpha)}$$

where ξ is a location parameter,
β is a scale parameter, and
α is a shape parameter.

- moments:

$$\mu_X = \xi + \frac{\alpha}{\beta}$$

$$\sigma^2_X = \frac{\alpha}{\beta}$$

$$\gamma_X = \frac{2}{\sqrt{\alpha}} \quad \text{for} \quad \beta > 0,\ x > \xi$$

$$\gamma_X = \frac{-2}{\sqrt{\alpha}} \quad \text{for} \quad \beta < 0,\ x < \xi$$

- parameter estimation: method of moments
- comments:

(i) The P3 becomes a large number of families of distributions including the normal, beta, and gamma distributions.

(ii) The P3 can be used with $\beta < 0$, yielding a negatively skewed distribution with an upper bound of ξ.

(iii) For $\beta > 0$ and lower bound $\xi = 0$, the P3 reduces to the gamma distribution for which $\gamma = 2CV$.

(iv) For a fixed mean and variance, as α goes to infinity and γ goes to zero, the P3 converges to the normal distribution.

(v) For $\alpha = 1$ and $\gamma = 2$, the P3 becomes the exponential distribution.

▪ Log Pearson Type III Distribution (LP3)

- description: Y = ln(X) ~ P3(α, β, ξ)
- use in hydrology: recommended for the description of floods in US by USWR Council and in Australia by their Institute of Engineers.
- moments:

$$\mu_X = e^{\xi}\left(\frac{\beta}{\beta-1}\right)^{\alpha}$$

$$\sigma^2_X = e^{2\xi}\left[\left(\frac{\beta}{\beta-2}\right)^{\alpha} - \left(\frac{\beta}{\beta-1}\right)^{2\alpha}\right]$$

$$\gamma_X = \frac{E[X^3] - 3\mu_X E[X^2] + 2\mu_X^3}{\sigma_X^3}$$

- parameter estimation method: indirect method of moments(WRC method)

- comments:

(i) The parameter ξ is a lower bound on the logarithms of the random variable if β is positive, and is an upper bound if β is negative.

(ii) It may be concluded that in general manner the methods of the WRC does not produce the desired accuracy, and that more recent methods (BOB, MM1, SAM) tend to produce better results (Bobee, B. and Ashkar, F. (1991), *The Gamma Family And Derived Distributions Applied In Hydrology*, Water Resources Publications, CO, USA, pp. 120.

< Extreme Value Family >

- description

(i) The extreme value of a set of r.v. is also random.

(ii) The pdf of extreme value r.v. in general depends on the sample size and the parent distribution
((why?))

- types of EV distributions

(i) Type I

(ii) Type II

(iii) Type III

Type	Extreme Value	Parent Distribution
Type I	largest	normal, lognormal, exponential, gamma
	smallest	normal
Type II	largest or smallest	Cauchy
Type III	largest	beta
	smallest	beta, lognormal, exponential, gamma

▪ Extreme Value Type I (Gumbel)

- pdf:

$$f_X(x) = \frac{1}{\alpha} \exp\left\{ \mp \frac{x-\beta}{\alpha} - \exp[\mp \frac{x-\beta}{\alpha}] \right\}$$

where - for max and + for min

and α and β are scale and location parameters, respectively.

-cdf:

$$F_X(x) = \exp[-\exp(-\frac{x-\beta}{\alpha})]$$

- moments:

$$\mu_X = \beta \pm 0.5772\alpha$$

$$\sigma^2_X = \frac{\pi^2\alpha^2}{6} = 1.645\alpha^2$$

$$\gamma = \pm 1.1396$$

- parameter estimation

(i) method of moments:

$$\hat{\alpha} = s_X/1.283$$

$$\hat{\beta} = \overline{X} \mp 0.45 s_X$$

(ii) MLE: Eq. (6.54) & (6.55)

- transformation: Y = (X-β)/α

$$f_Y(y) = \exp[\mp y - \exp(\mp y)]$$

$$F_Y(y) = \exp[-\exp(-y)] \quad \text{for max}$$

$$= 1 - \exp[-\exp(y)] \quad \text{for min}$$

▪ Extreme Value Type III (Weibull)

- pdf:

$$f_X(x) = \left(\frac{k}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{k-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{k}\right]$$

- cdf:

$$F_X(x) = 1 - \exp[-(x/\alpha)^k]$$

- moments

$$\mu_X = \alpha\Gamma(1+1/k)$$

$$\sigma^2_X = \alpha^2[\Gamma(1+2/k) - \Gamma^2(1+1/k)]$$

- parameter estimation methods

(i) method of moments

(ii) MLE

- comments

(i) Y = - ln(X) where X has a Weibull distribution and Y has a Gumbel distribution

(ii) For k < 1, the Weibull pdf goes to INF as x approaches 0.

(iii) For k = 0, the Weibull distribution reduces to the exponential distribution.

(iv) For k > 1, the Weibull pdf is like a P3 pdf for small x and the shape parameter (a) = k, but decays to 0 faster for large x.

▪ Generalized Extreme Value Distribution

- cdf:

$$F_X(x) = \exp\left\{-\left[1-\frac{\kappa(x-\xi)}{\alpha}\right]^{1/\kappa}\right\} \qquad \text{for} \quad \kappa \neq 0$$

where ξ, α, and κ are a location, a scale, and the important shape parameters, respectively.

- moments

Eq. (18.2.19)&(18.2.20) ((R3))

- parameter estimation method

L moments: Eq. (18.2.22a, b, c) ((R3))

- comments

(i) The GEV is similar to the Gumbel though the right-hand tail is thicker for $\kappa < 0$ and the thinner for $\kappa > 0$.

(ii) For $\kappa > 0$, the distribution has a finite upper bound at $\xi+\alpha/\kappa$ and corresponds to the EV type III distribution for maxima

(iii) For $\kappa < 0$, the distribution has a thicker right-hand tail and corresponds to the EV type II distribution for maxima