Bivariate drought analysis using entropy theory

Zengchao Hao
Department of Biological & Agricultural Engineering
Texas A & M University,
321E Scoates Hall, 2117 TAMU, College Station, Texas 77843-2117
Email: hzc07@tamu.edu

Vijay P. Singh
Department of Biological & Agricultural Engineering
Texas A & M University,
321 Scoates Hall, 2117 TAMU, College Station, Texas 77843-2117
vsingh@tamu.edu

Abstract

Drought duration and severity are two properties that are usually needed for drought analysis. To characterize the correlation between the two drought properties, a bivariate distribution is needed. A new method based on entropy theory is proposed for constructing the bivariate distribution that is capable of modeling drought duration and severity with different marginal distributions. Parameters of the joint distribution are estimated with Newton's method. Monthly streamflow data from Brazos River at Waco, Texas, are employed to illustrate the application of the proposed method to model drought duration and severity for drought analysis.

Keywords: drought analysis; joint distribution; entropy theory; principle of maximum entropy

1 Introduction

Drought analysis is important for water resources planning and management. *Yevjevich* [1967] used the run theory to define a drought as a sequence of intervals where water supply remains below water demand. This enables the characterization of drought in simple terms, such as duration and severity, using hydrological variables (e. g., streamflow). Drought duration and severity are two main characteristics that have often been used for drought analysis.

Drought duration and severity can be regarded as random variables and thus the probability distribution, either separate or joint, can be used for characterizing a drought. A traditional way to characterize the drought duration or severity is based on fitting a probability density function. Several approaches have been proposed for univariate drought analysis. The drought duration can be modeled by a geometric distribution [*Kendall and Dracup*, 1992; *Mathier et al.*, 1992] when it is treated as a discrete random variable or by an exponential distribution when it is treated as a continuous random variable [*Zelenhasi and Salvai*, 1987]. The gamma distribution is generally used to describe drought severity [*Shiau and Modarres*, 2009] .

However, the correlation between drought duration and severity cannot be characterized by univariate analysis and alternative multivariate approaches have therefore been used to model the correlation of drought variables [*González and Valdés*, 2003; *Salas et al.*, 2005; *Kim et al.*, 2006; *Shiau*, 2006; *Nadarajah*, 2007; *Nadarajah*, 2009]. Some bivariate distributions have been used for the joint distribution of drought duration and severity, such as the bivariate Pareto distribution [Nadarajah, 2009]. The joint distribution of drought duration and severity may be modeled using different distributions, in which case the general bivariate distribution does not work. To address this issue, the copula method has been applied to construct the distribution that is capable of linking two univariate distributions to form a bivariate distribution [*Shiau*, 2006; *Shiau et al.*, 2007]. The joint distribution can also be constructed from the conditional distribution and marginal distribution [*Shiau and Shen*, 2001]. Furthermore, nonparametric methods have also been proposed for bivariate drought analysis [*Kim et al.*, 2003; *Kim et al.*, 2006].

This article proposes a new method for constructing the bivariate distribution of drought duration and severity with different marginal distributions based on the principle of maximum entropy. The proposed method is applied for drought analysis based on the monthly streamflow of Brazos River at Waco, Texas.

2 Method

2.1 Principle of maximum entropy

For a continuous random variable X with probability density function (PDF) $f(x)$ defined on the interval $[a, b]$, entropy is defined as a measure of uncertainty expressed as [*Shannon*, 1948]:

$$H_1 = -\int_a^b f(x) \ln f(x)\, dx \tag{1.1}$$

For continuous random variables X and Y with a PDF $f(x, y)$ defined over the space $[a, b] \times [c, d]$, the Shannon entropy can be defined as:

$$H_2 = -\int_c^d \int_a^b f(x,y) \ln f(x,y)\, dxdy \tag{1.2}$$

The principle of maximum entropy was proposed by *Jaynes*, [1957] which states that the probability density function should be selected among all the distributions that satisfy the constraints. Generally, the constraints can be specified as:

$$\int_a^b f(x)\, dx = 1 \tag{2}$$

$$\int_a^b g_i(x) f(x)\, dx = \overline{g}_i \qquad i=0,1,2\ldots m \tag{3}$$

where m is the number of constraints; $g_i(x)$ is the function of x and $\overline{g}_i$ is the expectation of function $g_i(x)$.

The probability density function for the univariate case can be derived according to the principle of maximum entropy by maximizing the entropy defined in equation (1.1) subject to the constraints in equations (2) and (3). The maximization can be achieved using the method of Lagrange multipliers by introducing the Lagrange function L:

$$L = -\int_a^b f(x) \log f(x) dx - (\lambda_0 - 1)\int_a^b (f(x) - 1) - \sum_{i=1}^m \lambda_i \int_a^b (g_i(x) f(x) - \overline{g_i}) \tag{4}$$

where λ_i (i=0,1,2…,m) are the Lagrange multipliers. Differentiating L with respect to f and setting the derivative to zero, the maximum entropy distribution can be obtained as [*Kapur*, 1989]:

$$f(x) = \exp\left[-\lambda_0 - \lambda_1 g_1(x) - \lambda_2 g_2(x) \ldots - \lambda_m g_m(x)\right] \tag{5}$$

It has been shown that many of the commonly used distributions can be derived from entropy theory with different constraints and the maximum entropy distribution in equation (5) incorporates these distributions as special cases [*Singh*, 1998]. For example, if the first and second moments are specified as the constraints, the maximum entropy distribution is the normal distribution. It can be seen that the maximum entropy distribution is quite flexible.

2.2 Joint distribution

To derive the joint density function $f(x,y)$ of drought duration (X) and severity (Y), constraints for variables X and Y need to be specified separately and jointly. It has been shown that from the principle of maximum entropy, the exponential distribution can be obtained with the constraints in the form of mean, while the gamma distribution can be obtained with the constraints in the form of mean and logarithmic mean [*Singh*, 1998]. The separate constraints of variable X and Y can be specified accordingly as:

$$\int_a^b f(x,y) x\, dx = \overline{x} \tag{6}$$

$$\int_c^d f(x,y) y\, dy = \overline{y} \tag{7}$$

$$\int_c^d f(x,y) \ln y\, dy = \overline{\ln y} \tag{8}$$

The joint constraint can be specified to model the correlation through the product XY as:

$$\int_c^d \int_a^b xy f(x,y) dxdy = \overline{XY} \tag{9}$$

With these constraints from equations (6) and (9), the joint PDF can be obtained by maximizing the entropy in equation (1.2). Following similar steps in deriving the maximum entropy distribution in the univariate case, the joint PDF in the bivariate case can be obtained as:

$$f(x,y) = \exp(-\lambda_0 - \lambda_1 x - \lambda_2 y - \lambda_3 \ln y - \lambda_4 xy) \quad (10)$$

2.3 Marginal distribution

The marginal distribution for drought duration *X* can be obtained by integrating the joint PDF *f*(*x*, *y*) given by equation (10) over *Y as*:

$$f(x) = \int_c^d \exp(-\lambda_0 - \lambda_1 x - \lambda_2 y - \lambda_3 \ln y - \lambda_4 xy)dy \quad (11)$$

Similarly, the marginal distribution for drought severity Y can be obtained as:

$$f(y) = \int_a^b \exp(-\lambda_0 - \lambda_1 x - \lambda_2 y - \lambda_3 \ln y - \lambda_4 xy)dx \quad (12)$$

2.4 Parameter estimation

The entropy theory can be applied for parameter estimation and *Singh* [1998] gave an introduction of the entropy based method for estimating the Lagrange multipliers for the commonly used distributions. Lagrange multipliers can also be estimated by maximizing the function [*Mead and Papanicolaou*, 1984]:

$$\Gamma = \lambda_0 + \sum_{i=1}^{m} \overline{g}_i \quad (13)$$

Newton's method can be applied for maximizing the function Γ by updating $\boldsymbol{\lambda}_{(1)}$ with some initial value $\boldsymbol{\lambda}_{(0)}$ through the equation below:

$$\boldsymbol{\lambda}_{(1)} = \boldsymbol{\lambda}_{(0)} - H^{-1} \frac{\partial \Gamma}{\partial \lambda_i}, \quad i=1,2,\ldots,m \quad (14)$$

where $\boldsymbol{\lambda}=[\lambda_1,\ldots,\lambda_m]$; the gradient is expressed as:

$$\frac{\partial \Gamma}{\partial \lambda_i} = \overline{g}_i - \int_a^b f(x) g_i(x) dx, \quad i=1,2,\ldots,m \quad (15)$$

and *H* is the Hessian matrix whose elements are expressed as:

$$H_{i,j} = \int_a^b f(x) g_i(x) g_j(x) dx - \left[\int_a^b f(x) g_i(x) dx\right]\left[\int_a^b f(x) g_j(x) dx\right], i, j = 1,2\ldots,m \quad (16)$$

2.5 Comments

The steps above demonstrate how to construct the bivariate distribution of drought duration and severity with the constraints of mean and logarithmic mean. In real application, different constraints can be specified for the duration and severity separately to form different joint distributions. According to certain measures (e.g., the mean square errors of empirical and theoretical probabilities), the most suitable constraints for each variable can be used to form the joint density function of the duration and severity, which can then be used for drought analysis. The characteristic of the proposed method is that drought duration and severity can be modeled with different marginal distributions (with different constraints).

3 Results and discussion

Monthly streamflow data of Brazos River at Waco, TX (USGS 08096500) for the period from January 1941 to December 2009 was used for drought analysis. The mean streamflow of each month is used as the truncation level to define the drought event. Significant correlation exists between drought duration and severity and the bivariate density function in equation (10) was used to model them jointly.

Histograms and the fitted marginal distribution in equation (11) for drought duration and in equation (12) for drought severity are shown in Figure 1. The fitted PDFs capture the general pattern of the histograms. The empirical probability estimated from the Gringorten's plotting position formula and theoretical probability are shown in Figure 2. It can be seen that generally theoretical probabilities fitted empirical probabilities well. The Kolmogorov-Smirnov (K-S) goodness-of-fit test was used to further test whether the observed data can be modeled with the proposed model. The critical values for the duration and severity data at 5% significance level are 0.35 and 0.84. This indicates the proposed model can be applied to model the drought duration and severity data.

The return period for drought duration *D* greater or equal to a certain value *d* and for drought severity *S* greater or equal to a certain value *s* can be defined as [*Shiau*, 2003; *Shiau*, 2006]:

$$T_D = \frac{E(L)}{P_D(D \geq d)} \quad (17.1) \qquad T_S = \frac{E(L)}{P_S(S \geq s)} \quad (17.2)$$

where *E*(*L*) is the expected drought interval time that can be estimated from observed droughts; T_D , T_S are the return periods defined for drought duration and drought severity, respectively;

$P_D(D \geq d)$ and $P_S(S \geq s)$ are exceedance probability of drought duration and drought severity that can be estimated from equations (11) and (12), respectively. The univariate return periods of 2,5,10,20,50 and 100 years defined by separate drought duration and severity can then be estimated from equations (17.1) and (17.2) and are summarized in Table 1. For example, the drought duration for the 100 year return period is around 33.3 months and the drought severity for the 100 years return period is around 5.6×104 cfs months.

The joint return period of the drought duration and severity can be defined by the drought duration and severity exceeding specific values. Specifically, the joint return period T_{DS} of drought duration D and severity S can be defined as [*Shiau*, 2003; *Shiau*, 2006]:

$$T_{DS} = \frac{E(L)}{P(D \geq d, S \geq s)} \tag{18}$$

where $P(D \geq d, S \geq s)$ is the exceedance probability of drought duration and severity that can be obtained from the joint density function in equation (10). The joint return periods defined by equation (18) for different duration and severity values are shown in Figure 3.

The conditional return periods are also needed to assess the risk of water resources systems. The conditional return period $T_{D|S \geq s}$ for drought duration given drought severity exceeding a certain value can be defined as [*Shiau*, 2003; *Shiau*, 2006]:

$$T_{D|S \geq s} = \frac{T_S}{P(D \geq d, S \geq s)} \tag{19}$$

Similarly, the conditional return period $T_{S|D \geq d}$ for drought severity given drought duration exceeding a certain value can be defined as [*Shiau*, 2003; *Shiau*, 2006]:

$$T_{S|D \geq d} = \frac{T_D}{P(D \geq d, S \geq s)} \tag{20}$$

The conditional return periods are shown in Figure 4. For example, given the drought severity $s > 1 \times 10^4$ cfs month, the conditional return period of the drought duration exceeding 11 month is around 100 years.

4 Summary and Conclusion

A bivariate distribution based on entropy theory is proposed for constructing the joint distribution of drought duration and severity. The advantage of the proposed method is that it is flexible to incorporate different forms of marginal distributions of the drought duration and severity. Drought data defined by the monthly streamflow at Brazos River at Waco, Texas are used to illustrate the application of the proposed method for drought analysis. A good agreement is observed between the empirical and theoretical probabilities of the drought duration and severity. The bivariate distribution is then applied to model drought duration and severity jointly. Return periods of drought duration separately and jointly and the conditional return periods are then estimated. The results show that the proposed method is a useful tool to derive the joint distribution of drought duration and severity for drought analysis.

5 Reference

González, J. and J. Valdés (2003), Bivariate drought recurrence analysis using tree ring reconstructions, *Journal of Hydrologic Engineering*, 8(5), 247-258.

Jaynes, E. (1957), Information Theory and Statistical Mechanics, *Physical Review*, 106(4), 620-630.

Kapur, J. (1989), *Maximum-entropy models in science and engineering*, John Wiley & Sons.

Kendall, D. and J. Dracup (1992), On the generation of drought events using an alternating renewal-reward model, *Stochastic Hydrology and Hydraulics*, 6(1), 55-68.

Kim, T., J. Valdes and C. Yoo (2006), Nonparametric approach for bivariate drought characterization using Palmer drought index, *Journal of Hydrologic Engineering*, 11(2), 134-143.

Kim, T., J. Valdés and C. Yoo (2003), Nonparametric approach for estimating return periods of droughts in arid regions, *Journal of Hydrologic Engineering*, 8(5), 237-246.

Mathier, L., L. Perreault, B. Bobée and F. Ashkar (1992), The use of geometric and gamma-related distributions for frequency analysis of water deficit, *Stochastic Hydrology and Hydraulics*, 6(4), 239-254.

Mead, L. and N. Papanicolaou (1984), Maximum entropy in the problem of moments, *J. Math. Phys*, 25(8), 2404-2417.

Nadarajah, S. (2007), A bivariate gamma model for drought, *Water Resour. Res.*, 43, W08501, doi:10.1029/2006WR005641.

Nadarajah, S. (2009), A bivariate pareto model for drought, *Stochastic Environmental Research and Risk Assessment*, 23(6), 811-822.

Salas, J., C. Fu, A. Cancelliere, D. Dustin, D. Bode, A. Pineda and E. Vincent (2005), Characterizing the severity and risk of drought in the Poudre River, Colorado, *Journal of Water Resources Planning and Management*, 131(5), 383-393.

Shannon, C. E. (1948), A mathematical theory of communications, *Bell Syst. Tech. J.* , 27(7), 379-423.

Shiau, J. (2003), Return period of bivariate distributed extreme hydrological events, *Stochastic Environmental Research and Risk Assessment*, 17(1), 42-57.

Shiau, J. (2006), Fitting drought duration and severity with two-dimensional copulas, *Water resources management*, 20(5), 795-815.

Shiau, J., S. Feng and S. Nadarajah (2007), Assessment of hydrological droughts for the Yellow River, China, using copulas, *Hydrological Processes*, 21(16), 2157-2163.

Shiau, J. and H. Shen (2001), Recurrence analysis of hydrologic droughts of differing severity, *Journal of Water Resources Planning and Management*, 127(1), 30-40.

Shiau, J. T. and R. Modarres (2009), Copula based drought severity duration frequency analysis in Iran, *Meteorological Applications*, 16(4), 481-489.

Singh, V. P. (1998), *Entropy-Based Parameter Estimation in Hydrology*, Kluwer Academic Publishers.

Yevjevich, V. (1967), Objective approach to definitions and investigations of continental hydrologi droughts, *Hydrology Paper 23*, Colorado State U, Fort Collins.

Zelenhasi , E. and A. Salvai (1987), A method of streamflow drought analysis, *Water Resources Research*, 23(1), 156-168 , doi:10.1029/WR023i001p00156.

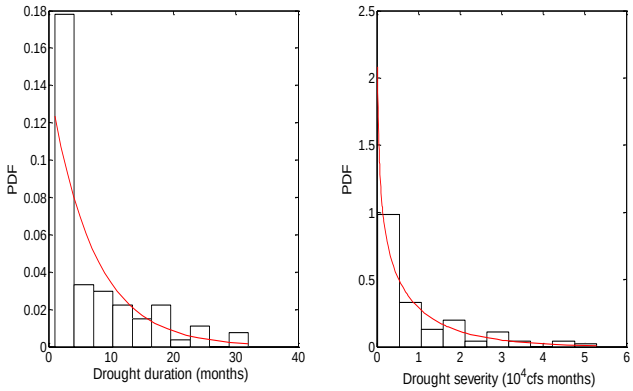

Figure 1 Comparison of the histograms and fitted PDFs

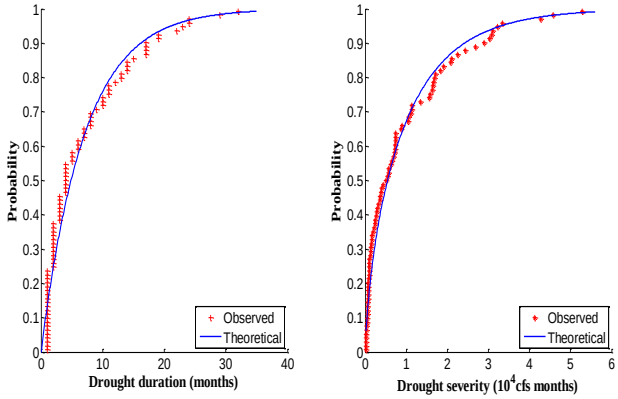

Figure 2 Comparison of the empirical and theoretical probability

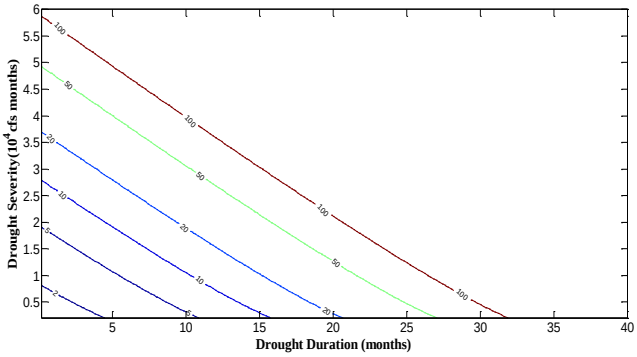

Figure 3 Bivariate drought duration and severity return periods

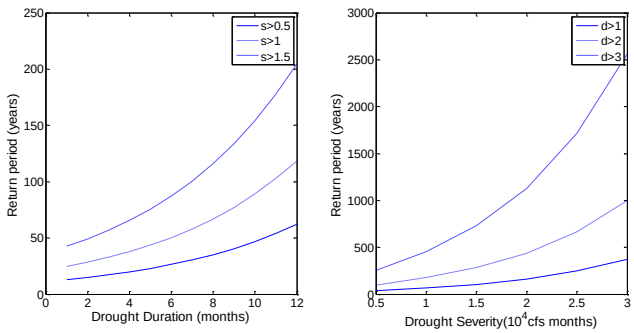

Figure 4 Conditional return periods

Table 1 Return period defined by drought duration and severity separately

Return Period (Year)	Drought duration (Month)	Drought severity (10^4cfs months)
2	5.8	0.7
5	12.3	1.7
10	17.1	2.6
20	22.0	3.5
50	28.5	4.7
100	33.3	5.6