ELSEVIER

Journal of Hydrology 169 (1995) 95–110

Journal of Hydrology

[1]

Predictions of drought length extreme order statistics using run theory

Lemuel A. Moyé*, Asha S. Kapadia

University of Texas School of Public Health, 1200 Herman Pressler, Houston, TX 77030, USA

Received 15 July 1994; revision accepted 26 October 1994

Abstract

A new model has been developed here, incorporating prospective run theory to predict events based on order statistics. The difference equations for such a system are presented and the closed form solution is obtained in terms of parameters to which the practising hydrologist has access. Predictions are provided for duration of maximum drought lengths as a function of the definition of drought, the number of years over which the prediction is made, and the probability of inadequate water resources.

1. Introduction

Workers in hydrology have long suspected that run theory with a basis in Bernoulli trials may be an appropriate foundation from which to predict sequences of hydrologic events. Yevjevich (1967) was among the first at attempting a prediction of properties of droughts using the geometric probability distribution, defining a drought of k years as k consecutive years when there are not adequate water resources. Additional work by Yevjevich (1967), and Saldariaga and Yevjevich (1970) continued the development of run theory, incorporating concepts of time series analysis in formulations to predict drought occurrence. Beginning with the examination of wet and dry periods, Şen (1976) continued this work in applying run theory to water resource predictions, evaluating run sums of annual flow series (1977). Moyé et al. (1988) developed an alternative model, extending the work of Yevjevich, equating a drought lasting k years in duration with a run of length k. This lead to the prediction of average drought lengths. The results of this model were seen to correspond closely to the findings of recorded precipitation history.

* Corresponding author.

0022-1694/95/$09.50
SSDI 0022-1694(94)02662-9

In ensuing correspondence, Şen (1989) was critical of these predictions, particularly with regard to the use of averages (e.g. average drought length). This criticism was anchored in the view that hydrologists should try to find solutions for critical values rather than averages of lengths of hydrologic phenomena. An interesting enumerative approach for maximum drought lengths was identified by Şen (1980) using a Markovian model. By integrating the notion of sequential occurrences of events with extreme order statistics (minimums and maximum event lengths), the model extends the application of formal probability modeling to hydrologic events of interest. As with earlier work, the solutions are expressed in parameters which are readily available to workers in the field. Predictions are provided for the relative frequency of extreme critical events of interest.

2. Methods

In this development, the hallmark of events of hydrologic significance is consecutive occurrences of the event. We define a drought (failure run) as a sequence of consecutive years of inadequate water resources. We may also define a bounty (success run) as a run of consecutive years of adequate water resources. We will also assume that the year-to-year availability of adequate water resources can be approximated by a sequence of Bernoulli trials in which only one of two possible outcomes (success with probability p or failure with probability q) may occur. These probabilities remain the same from year to year, and knowledge of the result of a previous year provides no information for the water resource availability of any following year. This model has been employed in modelling hydrologic phenomena (Moyé et al., 1988; Şen, 1989).

Our previous work defined the probability of exactly i runs of length k in n trials. In this paper, we define $T^*_{K,L}(n)$ as the probability all failure runs are greater than or equal to K and less than or equal to L in length where $0 \leqslant K \leqslant L$ in a sequence of n Bernoulli trials. If q is the probability of inadequate water resources, then $T^*_{K,L}(n)$ is the probability that all droughts that occur in n trials are between K and L years in length. If $T^*_{K,L}(n)$ can be formulated in terms of K, L, q, and n, quantities to which the hydrologist has access, the following probabilities for n consecutive years may be identified.

$$T^*_{K,L}(n) = P(\text{all drought lengths are between } K \text{ and } L \text{ years long})$$

$$1 - T^*_{0,L}(n) = P(\text{at least one drought greater than } L \text{ years in length})$$

Probabilities for the order statistics involving the maximum and minimum drought length are particularly noteworthy

$$P(\text{minimum drought length greater than or equal to } K \text{ years}) = T^*_{K,n}(n)$$

$$P(\text{maximum drought length less than or equal to } L \text{ years}) = T^*_{0,L}(n)$$

Access to these probabilities provides useful information for predicting extremes in drought lengths. The notion of order statistics can be further developed by observing that $T^*_{0,K}(n)$ is the probability that all droughts in n years are less than or equal to K years long, i.e. the maximum drought length is less than or equal to K years in length. $T^*_{0,K-1}(n)$ is the probability that the maximum drought length is less than or equal to $K-1$ years in length. Thus $T^*_{0,K}(n) - T^*_{0,K-1}(n)$ is the probability that the maximum drought length is exactly K years in length. The expected maximum drought length $E[M_D(n)]$ and its variance $V[M_D(n)]$ can be computed as

$$E[M_D(n)] = \sum_{K=0}^{n} K[T^*_{0,K}(n) - T^*_{0,K-1}(n)]$$

$$V[M_D(n)] = \sum_{K=0}^{n} K^2[T^*_{0,K}(n) - T^*_{0,K-1}(n)] - E^2[M_D(n)]$$

It is important to note that $E[M_D(n)]$ is not the average drought length as reported in the literature (Moyé et al. (1988)), but instead is the expected length of the worst drought in n years. For example, if in a 20 year span, the droughts which occur are of length 3 years, 2 years, and 5 years, the worst drought is the 5 year drought.

An analogous computation can be made for consecutive years of adequate rainfall. Define a bounty of length K as a sequence of K consecutive years where for each year there is adequate water resources. Then, for computations of probabilities involving bounties we define q is the probability of adequate water resources, and $T^*_{K,L}(n)$ is the probability that in n years all bounties are greater than or equal to K and less than or equal to L years in length.

We begin the general derivation of the $T^*_{K,L}(n)$ model by defining $T_{K,L}(n)$ as the probability all failure runs are either of length 0 or between K and L. The desired quantity $T^*_{K,L}(n)$ is then just $T_{K,L}(n) - q^n$. The boundary conditions for $T_{K,L}(n)$ are: $T_{K,L}(n) = 0$ for all $n < 0$, $T_{K,L}(n) = 0$ for $K > L$; $T_{K,L}(n) = 1$ for $n = 0$; $T_{K,L}(n) = p^n$ for $0 < n < K$.

Using the indicator function, we may write the recursive relationship for $T_{K,L}(n)$ for $0 < K \leqslant \min(L, n)$ as

$$\begin{aligned} T_{K,L}(n) &= p^n I_{0 \leqslant n < K} + pT_{K,L}(n-1)I_{n \geqslant K} + q^K pT_{K,L}(n-K-1)I_{n \geqslant K} \\ &\quad + q^{K+1}pT_{K,L}(n-K-2)I_{n \geqslant K} + q^{K+2}pT_{K,L}(n-K-3)I_{n \geqslant K} \\ &\quad + \dots + q^L pT_{K,L}(n-L-1)I_{n \geqslant K} + q^n I_{K \leqslant n \leqslant L} \qquad (1) \end{aligned}$$

where $I_{x \in A}$ is equal to one when $x \in A$, 0 otherwise. The motivation for Eq. (1) lies in noting the recursive relationship must include only those terms which permit failure run lengths of the required lengths. Since by its definition, $T_{K,L}(n)$ only permits failure runs of between lengths K and L, only terms with powers of q between K and L are allowed. The index functions for p^n is required since a run of length zero is permitted. In addition, a term q^n is required when $K \leqslant n \leqslant L$, since for such a small value of n a failure run length of length n meets the requirement of $T_{K,L}(n)$.

Specific cases of this recursive relationship are provided in the Appendix. Define the generating function

$$G(s) = \sum_{n=0}^{\infty} s^n T_{K,L}(n)$$

where s is a constant $0 < s < 1$. The range of the summand is $0 \leqslant n < \infty$, and the coefficient of s^n is the desired quantity $T_{K,L}(n)$. The general solution for $T_{K,L}(n)$ will be found by first collapsing the infinite number of equations for $T_{K,L}(n)$ into one equation for $G(s)$. This equation will be solved for $G(s)$, which will then be inverted, identifying the coefficients of s^n. Begin by multiplying both sides of Eq. (1) for $T_{K,L}(n)$ by s^n

$$\begin{aligned} s^n T_{K,L}(n) = {} & s^n p^n I_{0 \leqslant n < K} + ps^n T_{K,L}(n-1) I_{n \geqslant K} + q^K ps^n T_{K,L}(n-K-1) I_{n \geqslant K} \\ & + q^{K+1} ps^n T_{K,L}(n-K-2) I_{n \geqslant K} + q^{K+2} ps^n T_{K,L}(n-K-3) I_{n \geqslant K} \\ & + \ldots + q^L ps^n T_{K,L}(n-L-1) I_{n \geqslant K} + q^n s^n I_{K \leqslant n \leqslant L} \end{aligned}$$

Sum over the interval $n \geqslant 0$ to obtain

$$\begin{aligned} \sum_{n=0}^{\infty} s^n T_{K,L}(n) = {} & \sum_{n=0}^{\infty} s^n p^n I_{0 \leqslant n < K} + p \sum_{n=0}^{\infty} s^n T_{K,L}(n-1) I_{n \geqslant K} \\ & + q^K p \sum_{n=0}^{\infty} s^n T_{K,L}(n-K-1) I_{n \geqslant K} \\ & + q^{K+1} p \sum_{n=0}^{\infty} s^n T_{K,L}(n-K-2) I_{n \geqslant K} \\ & + q^{K+2} p \sum_{n=0}^{\infty} s^n T_{K,L}(n-K-3) I_{n \geqslant K} \\ & + \ldots + q^L p \sum_{n=0}^{\infty} s^n T_{K,L}(n-L-1) I_{n \geqslant K} \\ & + \sum_{n=0}^{\infty} s^n q^n I_{K \leqslant n \leqslant L} = \sum_{n=0}^{K-1} (ps)^n + p \sum_{n=K}^{\infty} s^n T_{K,L}(n-1) \\ & + \sum_{r=K}^{L} pq^r \sum_{n=K}^{\infty} s^n T_{K,L}(n-r-1) + \sum_{n=0}^{\infty} s^n q^n I_{K \leqslant n \leqslant L} \end{aligned} \quad (2)$$

The first term from the RHS of Eq. (2) may be simplified as

$$\sum_{n=0}^{K-1} (ps)^n = \frac{1-(ps)^K}{1-ps}$$

and the second term from Eq. (2) is

$$p\sum_{n=K}^{\infty} s^n T_{K,L}(n-1) = ps\sum_{n=K}^{\infty} s^{n-1} T_{K,L}(n-1) = ps\left[G(s) - \sum_{n=0}^{K-2}(ps)^n\right]$$

The rth term ($K \leqslant r \leqslant L$) in the finite sum in Eq. (2) from

$$\sum_{r=K}^{L} pq^r \sum_{n=K}^{\infty} s^n T_{K,L}(n-r-1)$$

can be written as

$$pq^r \sum_{n=K}^{\infty} s^n T_{K,L}(n-r-1) = pq^r s^{r+1} \sum_{n=K}^{\infty} s^{n-r-1} T_{K,L}(n-r-1)$$

$$= pq^r s^{r+1} \sum_{n=K-r-1}^{\infty} s^n T_{K,L}(n) = pq^r s^{r+1}\left[\sum_{n=0}^{\infty} s^n T_{K,L}(n)\right] = pq^r s^{r+1} G(s)$$

The last term from Eq. (2) is easily seen as

$$\sum_{n=K}^{\infty} q^n s^n I_{K \leqslant n \leqslant L} = \sum_{n=K}^{L} q^n s^n$$

Thus, the infinite collection of equations in terms of $T_{K,L}(n)$ may be written as one equation in terms of $G(s)$

$$G(s) = \sum_{n=0}^{K-1}(ps)^n + ps\left[G(s) - \sum_{n=0}^{K-2}(ps)^n\right] + \sum_{r=K}^{L} pq^r s^{r+1} G(s) + \sum_{n=K}^{L} q^n s^n$$

$$= \frac{\sum_{n=0}^{K-1}(ps)^n - (ps)\sum_{n=0}^{K-2}(ps)^n + \sum_{n=K}^{L}(qs)^n}{1 - ps - \sum_{r=K}^{L} pq^r s^{r+1}}$$

$$= \frac{1 - qs + (qs)^K - (qs)^{L+1}}{(1-qs)\left(1 - ps - \sum_{r=K}^{L} pq^r s^{r+1}\right)}$$

To invert $G(s)$, the result of Theorem I from the Appendix provides that if

$$H(s) = \frac{1}{1 - ps - \sum_{r=K}^{L} pq^r s^{r+1}} = \sum_{n=k}^{\infty} a_n s^n$$

Then a_n is given by

$$a_n = \sum_{m=0}^{m}\sum_{m_0=0}^{m}\sum_{m_1=0}^{m}\sum_{m_2=0}^{m}\cdots\sum_{m_{L-K+1}=0}^{m}\binom{m}{m_0 m_1 m_2 \ldots m_{L-K+1}} p^m q^{\sum_{i=0}^{L-K+1} m_i(K+i)}$$

$$\times I_{\sum_{i=0}^{L-K+1} m_i \leqslant m} I_{\sum_{i=0}^{L-K+1} m_i(K+i)=n}$$

And, according to the Corollary in the Appendix, if

$$G(s) = \frac{1 - qs + (qs)^K - (qs)^{L+1}}{(1 - qs)\left(1 - ps - \sum_{r=K}^{L} pq^r s^{r+1}\right)}$$

then the coefficient for the term s^n, $n \geqslant 0$

$$T_{K,L}(n) = c_n - qc_{n-1} + q^k c_{n-k} - q^L c_{n-L-1}$$

where

$$c_n = \sum_{j=0}^{n} a_j q^{n-j}$$

$$a_H = \sum_{m=0}^{H} \sum_{m_0=0}^{m} \sum_{m_1=0}^{m} \sum_{m_2=0}^{m} \cdots \sum_{m_{L-K+1}=0}^{m} \binom{m}{m_0 m_1 m_2 \ldots m_{L-K+1}} p^m q^{\sum_{i=0}^{L-K+1} m_i(K+i)}$$

$$\times I_{\sum_{i=0}^{L-K+1} m_i \leqslant m} I_{\sum_{i=0}^{L-K+1} m_i(K+i)=H}$$

Thus, $T^*_{K,L}(n) = T_{K,L}(n) - q^n$, and is an explicit function of n, the total number of years, q the probability of inadequate water resources in a year, and k, the length of the drought of interest.

3. Results

This model can be applied to estimating probabilities of drought/bounty lengths and statements concerning minimum and maximum expected drought/bounty lengths. The applicability of this stochastic approach in predicting droughts based on Texas state precipitation data has been established (Moyé et al., 1988). The implications of $T^*_{K,L}(n)$ for the occurrence of future hydrologic phenomena are provided, and the model allows complete freedom in choosing the drought lengths of interest. The prediction of drought length in the next 20 years as a function of the probability of inadequate water resources yield direct assessments of the relationship between the yearly probability of inadequate water resources and the probability of drought (Table 1) beginning with short drought lengths (0–5 years) and progressing

Table 1
Predictions of drought length in 20 years as a function of yearly probability of inadequate water resources (q)

Permitted drought lengths (years)	Yearly probability of inadequate water resources									
	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.99
0–5	1	0.99	0.99	0.96	0.88	0.71	0.46	0.19	0.03	0
6–15	0	0	0	0	0	0.00	0.01	0.05	0.15	0.07
16–20	0	0	0	0	0	0	0.00	0.02	0.15	0.84

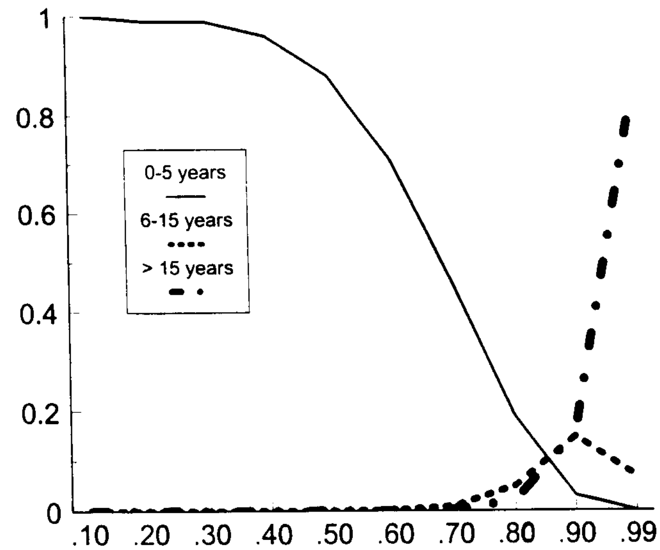

Fig. 1. Drought length as a function of q. y-axis label — probability; x-axis label — q: solid line, 0–5 years; dotted line, 6–15 years; dashed-dotted line, more than 15 years.

to medium drought lengths (6–10 years) and long drought lengths (16–20 years). For example, the probability that in the next 20 years, all drought lengths will be less than or equal to 5 years in duration is 0.99 if the probability of inadequate water resources is 0.20. Although the probability that drought lengths will be restricted to between 6 and 15 years increases as the probability of inadequate water resources increases, rising to a maximum of 0.15 for $q = 0.90$, it is notable that this probability then decreases as q increases further, falling to 0.07 for the maximum value of q examined (0.99). This somewhat paradoxical behavior is explained by the fact that 15 years is not the maximum possible drought length in 20 years (Fig. 1). As q increases from 0.10 to 0.99, the probability that the only droughts which occur are long droughts (16–20 years in length) increases slowly at first for small values of q, then increasingly dramatically as q increases to 0.99.

Table 2
Probability of maximum drought length in 20 years as a function of probability of inadequate water resources (q)

Maximum drought lengths	Yearly probability of inadequate water resources (q)						
	0.10	0.20	0.30	0.40	0.50	0.60	0.70
> 2 years	0.016	0.112	0.310	0.562	0.787	0.928	0.986
> 4 years	0	0.004	0.028	0.100	0.250	0.478	0.731
> 6 years	0	0	0.002	0.014	0.058	0.170	0.382
> 8 years	0	0	0	0.002	0.013	0.054	0.172
> 10 years	0	0	0	0	0.003	0.017	0.073

Table 2 examines the expected lengths of maximum drought lengths as a function of q. Probabilities for longer maximum drought lengths increase as q increases. For example, in a 20 year period, the probability the maximum drought length is greater than 4 years increases from 0.004 for $q = 0.20$ to 0.731 for $q = 0.70$. However, even for large values of q, the probability of having a maximum drought length of greater than 10 years is only 0.073.

Behavior of the expected maximum drought length as a function of q provides important new information for droughts over the next 20 years. For each yearly probability of inadequate water resources, Table 3 provides the expected duration of the worst drought (maximum drought length), and its standard deviation. In addition, the probability mass function is provided for the distribution of the maximum drought length. This table demonstrates that the expected maximum drought length increases as the yearly probability of inadequate water resources increases (Fig. 2). In addition, the standard deviation increases, making prediction of the length of the worst drought increasingly hazardous.

This same approach can be applied to bounty length (consecutive years of adequate water) probabilities. For fixed probability of adequate water resources, the probability the minimum bounty length exceeds K decreases as K increases (Fig. 3). In addition, as the probability of adequate water resources increases, the probability that

Table 3
Moments and probability distribution of maximum drought length M as a function of q

Parameters	q						
	0.10	0.20	0.30	0.40	0.50	0.60	0.70
$E(M)$	1.18	1.63	2.21	2.90	3.72	4.80	6.28
$SD(M)$	0.43	0.76	1.00	1.26	1.59	2.03	2.65
$P(M = 1)$	0.839	0.512	0.232	0.076	0.017	0.002	0
$P(M = 2)$	0.145	0.375	0.458	0.632	0.196	0.070	0.014
$P(M = 3)$	0.014	0.090	0.214	0.311	0.309	0.209	0.088
$P(M = 4)$	0.002	0.020	0.068	0.151	0.228	0.241	0.167
$P(M = 5)$	0	0.003	0.020	0.062	0.128	0.186	0.187
$P(M = 6)$	0	0.001	0.006	0.024	0.064	0.121	0.162
$P(M = 7)$	0	0	0.002	0.009	0.031	0.073	0.123
$P(M = 8)$	0	0	0	0.003	0.015	0.042	0.087
$P(M = 9)$	0	0	0	0.001	0.007	0.024	0.059
$P(M = 10)$	0	0	0	0	0.003	0.014	0.040
$P(M = 11)$	0	0	0	0	0.001	0.008	0.026
$P(M = 12)$	0	0	0	0	0.001	0.004	0.017
$P(M = 13)$	0	0	0	0	0	0.002	0.011
$P(M = 14)$	0	0	0	0	0	0.001	0.007
$P(M = 15)$	0	0	0	0	0	0.001	0.005
$P(M = 16)$	0	0	0	0	0	0	0.003
$P(M = 17)$	0	0	0	0	0	0	0.002
$P(M = 18)$	0	0	0	0	0	0	0.001
$P(M = 19)$	0	0	0	0	0	0	0.001
$P(M = 20)$	0	0	0	0	0	0	0.001

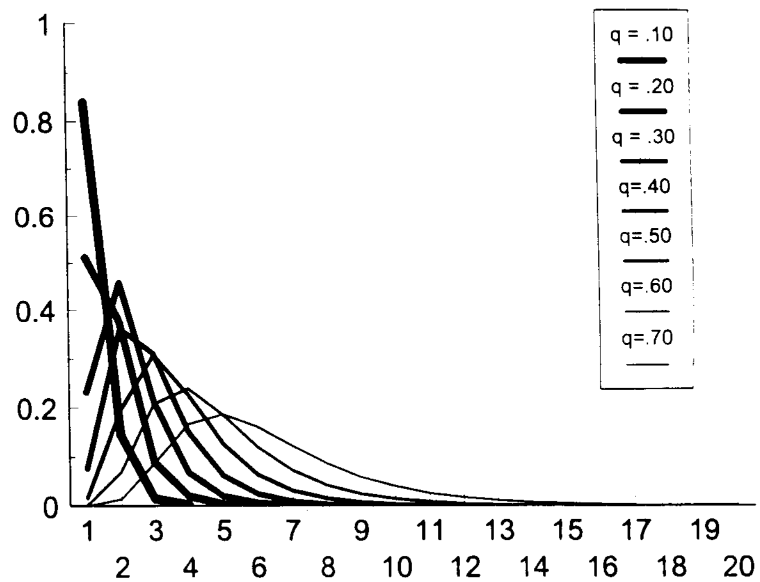

Fig. 2. Probability mass function of maximum drought length. y-axis label — probability; x-axis label — duration of maximum drought length (years).

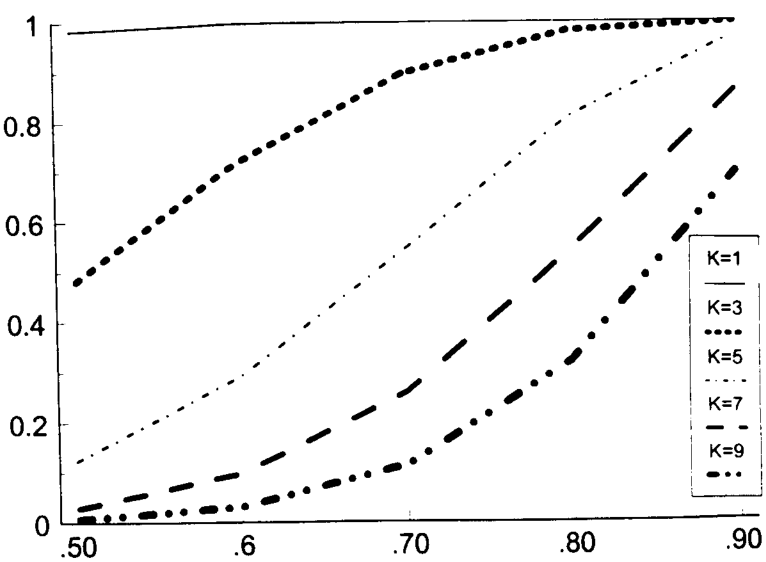

Fig. 3. Minimum bounty length probabilities. y-axis label — probability minimum bounty length exceeds K years. x-axis label — probability of adequate water resources. From top to bottom: thin line, $K = 1$; dotted line, $K = 3$; dashed-dotted line, $K = 5$; dashed line, $K = 7$; dashed-dot-dotted line, $K = 9$.

a bounty length is the minimum bounty length increases as longer bounty lengths become more likely.

4. Discussion

Despite the advances in stochastic hydrology, an important hurdle for its application remained the incomplete examination of 'critical drought length', or extreme order statistics, as mentioned by Şen (1989). The emphasis on maximum values would provide a firm basis for decisions concerning the performance of engineering structures at there most extreme risk. Initial work by David and Barton (1962) and Feller (1968) established the probability mass function for the distribution of the longest run length. Both Yevjevich (1967), and Şen, (1989) have worked in this area. We agree with this insightful perspective, and have developed a model based in prospective run theory which permits this examination of critical drought length through the use of order statistics. The model allows prediction of droughts of arbitrary length, and is easily converted to allow the computation of the probabilities of different bounty lengths. However, more importantly, we compute the extreme order statistics of drought and bounty length. Also the standard deviation of this drought length is provided, allowing for the computation of the degree of variability for the maximum drought length. Since this computation comes directly from the derived model, this computation is again a function of parameters that the hydrologist has access to, specifically q and n the number of years over which the prediction is estimated.

The work of Stern and Coe (1984) is also relevant here. There work is a conceived executed analysis of rainfall data to accurately predicting rainfall. This work begins with a Markov chain to compute the probability of rain occurring on a given day. It models the distribution of rainfall on a given day (assuming it rains on this day) as a Gamma distribution. The parameters of this distribution are assumed to vary seasonably and Fourier series are used to determine the best estimate of this seasonal variation. What is relevant for this discussion is that one of the products of this work is the probability of dry periods (i.e. consecutive days without rain), and predictions are provided for the lengths of these periods. But with this work, predictions are not provided for the extreme order statistics (minimum and maximum lengths) of the dry periods. They do not fully report the probability distribution of dry periods of 10 days, 15 days, and 20 days, the lengths their manuscript reviews. The present manuscript does not develop an estimator for daily rainfall amounts. It begins with run theory, providing not just the distribution of run lengths, but the probability distribution of the extreme order statistics.

The generating function approach is not the only possible mode of computation available to stochastic hydrologists. Direct enumeration (Şen, 1988) has been employed in the past. However, this is most easily executed if the sample sizes are small. We prefer the generating function approach since it provides a direct solution for any value of $n > 0$. In this setting, $1 - T_{K,L}(n)$ is the probability of an extremely short ($< K$) or extremely long ($> L$) drought if q is the probability of inadequate water resources. There are two ways to solve for $T^*_{K,L}(n)$. A first approach is to use the heavily nested Eq. (1), building up the sequence $T^*_{K,L}(1)$, $T^*_{K,L}(2)$, $T^*_{K,L}(3) \ldots T^*_{K,L}(n-1)$. The second way is to use the closed form solution, which does not require knowledge of $T^*_{K,L}(m)$ for $m < n$. In addition, the recursive relationship developed here allows freedom in considering the run lengths of interest through the choice of K and L such that $0 < K \leqslant L$. Thus, one can choose to focus on minimum run lengths, and easily obtain the distribution for the minimum run length from the solution provided in the methods. Alternatively, one might focus on the probability that the run lengths are neither extremely small nor extremely large.

The users of run theory in the past have not translated their run theory computations from the probability arena to the hydrology arena successfully. This is a difficulty of the parameterization of the problem. This translation is essential if the develops from run theory are to be applied to hydrology. The work provided here offers the direct translation from run theory to probabilities for drought length order statistics, the expected length of the worst drought and the expected length of the smallest bounty.

The developed model here is extremely flexible. For example, one may consider the definition of a drought as a consecutive, uninterrupted string of successes as too confining. A drought could occur when there is one year of adequate water resources, preceded and followed by consecutive years of inadequate water resources. The model should be generalized in this direction, although this generalization would not include the important case where surface or underground storage or carry over of any kind exists. A first step would be allowing as a drought a success followed by consecutive years of failure $q^{k-m-1}pq^m$. The implications of a model defining a drought as such have yet to be explored.

A criticism of this model is the use of Bernoulli trials. However, a prospective run theory model based on independent Bernoulli trials has been demonstrated to work sufficiently well in predicting rainfall in Texas (Moyé et al., 1988). Thus, although there is correlation from year to year in rainfall amounts, the correlation is sufficiently small to allow the independence assumption as a reasonable approximation. Nevertheless, the theoretical underpinnings of this model would be substantially strengthened if the model could be developed allowing correlation between events on different trials. Although work continues in this area, the present development of stochastic hydrology continue to provide useful estimates of recurrent hydrologic phenomena.

Appendix

Specific cases

We provide here two examples of the evaluation of $T_{K,L}(n)$. First consider the circumstances for the case where $n = k$, $T_{K,L}(n) = T_{K,L}(k)$. By simple enumeration, the probability of having a sequence of trials where there are no failures or, when failures occur, they occur in run lengths between length K and L is $p^k + q^k$. Using Eq (1), the result is identified as

$$T_{K,L}(K) = pT_{K,L}(K-1) + q^K pT_{K,L}[K(K+1)] + q^{K+1}pT_{K,L}[K-(K+2)] + q^{K+2}pT_{K,L}[K-(K+3)] + \dots + q^L pT_{K,L}(K-L-1) + q^k I_{K\leqslant K\leqslant L} = p^K + q^K$$

As a second example, consider $T_{K,L}(K+1) = P$ (there are no failures, or, when failure runs occur, they are between length K and L in $K+1$ trials). By enumeration this is plainly $= P$ ((no failures in $K+1$ trials) or (a failure run of length K in $K+1$ trials) or (a failure run of length $K+1$ in $K+1$ trials)) $= p^{K+1} + 2pq^k + q^{k+1}$.

Using Eq. (1), this result may also be found as

$$T_{K,L}(K+1) = pT_{K,L}(K+1-1) + q^K pT_{K,L}[K+1-(K+1)] + q^{K+1}pT_{K,L}[K+1-(K+2)] + q^{K+2}pT_{K,L}[K+1-(K+3)] + \dots + q^L pT_{K,L}(K+1-L-1) + q^k I_{K\leqslant K\leqslant L} = p(p^K + q^K) + q^k p + q^{K+1} = p^{K+1} + 2q^k p + q^{K+1}$$

and the result is verified.

Theorem:

Assertion:

If the generating function $H_k(s)$, such that, for constants k, q and p

$$H_{K,L}(s) = \sum_{n=k}^{\infty} a_n s^n \frac{1}{1 - ps - \sum_{r=K}^{L} q^r ps^{r+1}} = \frac{1}{1 - ps\left(1 + \sum_{r=K}^{L} q^r s^r\right)}$$

Then

$$a_n = \sum_{m=0}^{n}\sum_{m_0=0}^{m}\sum_{m_1=0}^{m}\dots\sum_{m_{L-K+1}=0}^{m}\binom{m}{m_0 m_1 m_2 \dots m_{L-K+1}} p^m q^{\sum_{i=0}^{L-K+1} m_i(K+i)} \times I_{\sum_{i=0}^{L-K+1} m_i \leqslant n} I_{\sum_{i=0}^{L-K+1} m_i(K+i)=n}$$

Proof:

An examination of the generating function and the inversion process for selected values of L provides guidance for the proof in the general case. Begin with $L = K+1$, $H_{K,L}(s)$ is

$$H_{K,K+1}(s) = \frac{1}{1 - ps - pq^K s^{K+1} - pq^{K+1}s^{K+2}} = \frac{1}{1 - ps(1 + q^K s^K + q^{K+1}s^{K+1})}$$

As a first step, the term $\sum_{n=0}^{\infty} a_n s^n = \sum_{n=0}^{\infty} p^n (1 + q^K s^K + q^{K+1}s^{K+1})^n s^n$. In order to collect like terms of s^n, use the multinomial theorem to write

$$(1 + q^K s^K + q^{K+1} s^{K+1})^n = \sum_{n_0=0}^{n} \sum_{n_1=0}^{n} \binom{n}{n_0 \quad n_1} (q^K s^K)^{n_0} (q^{K+1} s^{K+1})^{n_1} I_{n_0+n_1 \leqslant n}$$

where

$$\binom{n}{n_0 \quad n_1} = \frac{n!}{n_0! n_1! (n - n_0 - n_1)!}$$

Now this preliminary examination of $a_n s^n$ can be written as

$$a_n s^n = \sum_{n_0=0}^{n} \sum_{n_1=0}^{n} \binom{n}{n_0 n_1} p^n q^{Kn_0+(K+1)n_1} I_{n_0+n_1 \leqslant n} s^{n+Kn_0+(K+1)n_1}$$
$$= \sum_{n_0=0}^{n} \sum_{n_1=0}^{n} c(n, n_0, n_1) s^{n+Kn_0+(K+1)n_1}$$

The coefficients of common powers of s must now be identified by examining the powers of s generated for each n, n_0 and n_1 such that $0 \leqslant n_0$, $0 \leqslant n_1$ and $n_0 + n_1 \leqslant n$. For $n = 0$, only $c(0,0,0)$ is a coefficient of s_0; and for $n = 1$ only $c(1,0,0)$ is the coefficient of s. In general the coefficient of s^n is

$$a_n = \sum_{m=0}^{n} \sum_{m_0=0}^{m} \sum_{m_1=0}^{m} c(m, m_0, m_1) I_{m+Km_0+(K+1)m_1=n}$$
$$= \sum_{m=0}^{n} \sum_{m_0=0}^{m} \sum_{m_1=0}^{m} \binom{m}{m_0 m_1} p^m q^{Km_0+(K+1)m_1} I_{m_0+m_1 \leqslant m} I_{m+Km_0+(K+1)m_1=n}$$

When I_x is the indicator function for the condition x. This completes the examination of $L = K + 1$. We proceed in exactly the same fashion for $L = K + 2$. Begin by writing the generating function as

$$H_{K,K+2}(s) = \frac{1}{1 - ps(\sum_{r=K}^{K+2} q^r s^r)} = \frac{1}{1 - ps(1 + q^K s^K + q^{K+1} s^{K+1} + q^{K+2} s^{K+2})}$$

As before, we first write the coefficient $a_n s^n = p^n (1 + q^K s^K + q^{K+1} s^{K+1} + q^{K+2} s^{K+2})^n s^n$. In order to collect like terms of s^n, use the multinomial theorem to write

$$(1 + q^K s^K + q^{K+1} s^{K+1} + q^{K+2} s^{K+2})^n$$
$$= \sum_{n_0=0}^{n} \sum_{n_1=0}^{n} \sum_{n_2=0}^{n} \binom{n}{n_0 n_1 n_2} (q^K s^K)^{n_0} (q^{K+1} s^{K+1})^{n_1} (q^{K+2} s^{K+2})^{n_2} I_{n_0+n_1 \leqslant n}$$

Now $a_n s^n$ can be written as

$$a_n s^n = \sum_{n_0=0}^{n}\sum_{n_1=0}^{n}\sum_{n_2=0}^{n}\binom{n}{n_0 n_1 n_2} p^n q^{\sum_{i=0}^{2} n_i(K+i)} I_{\sum_{i=0}^{2} n_i \leqslant n}\, s^{n+\sum_{i=0}^{2} n_i(K+i)}$$

$$\sum_{n_0=0}^{n}\sum_{n_1=0}^{n}\sum_{n_2=0}^{n} c(n, n_0, n_1, n_2) s^{n+\sum_{i=0}^{2} n_i(K+i)}$$

As before, the coefficients of common powers of s must now be identified by examining the powers of s generated for each n, n_0 and n_1, and n_2 such that $0 \leqslant n_0$, $0 \leqslant n_1$, $0 \leqslant n_2$ and $n_0 + n_1 + n_2 \leqslant n$.

In general the coefficient of s^n is

$$a_n = \sum_{m=0}^{n}\sum_{m_0=0}^{m}\sum_{m_1=0}^{m}\sum_{m_2=0}^{m} c(m, m_0, m_1, m_2) I_{m+\sum_{i=0}^{2} m_i(K+i)=n}$$

$$= \sum_{m=0}^{n}\sum_{m_0=0}^{m}\sum_{m_1=0}^{m}\sum_{m_2=0}^{m} \binom{m}{m_0 \quad m_1 \quad m_2} p^m q^{\sum_{i=0}^{2} m_i(K+i)} I_{\sum_{i=0}^{2} m_i \leqslant m} I_{\sum_{i=0}^{2} m_i(K+i)=n}$$

This completes the examination of $L = K + 2$.

For the general case of $L > K$, begin by writing the generating function as

$$H_{K,L}(s) = \frac{1}{1 - ps - \sum_{r=K}^{L} pq^r s^{r+1}} = \frac{1}{1 - ps\left(1 + \sum_{r=K}^{L} pq^r s^r\right)}$$

Proceeding as before, we first write the coefficient $a_n = p^n(1 + q^K s^K + q^{K+1}s^{K+1} + q^{K+2}s^{K+2} + \ldots + q^L s^L)^n s^n$. In order to collect like terms of s^n, use the multinomial theorem to write

$$\left(1 + \sum_{r=K}^{L} q^r s^r\right)^n = \sum_{n_0=0}^{n}\sum_{n_1=0}^{n}\sum_{n_2=0}^{n}\cdots\sum_{n_{L-K+1}=0}^{n}\binom{n}{n_0 n_1 n_2 \ldots n_{L-K+1}}$$

$$(q^K s^K)^{n_0}(q^{K+1}s^{K+1})^{n_1}(q^{K+2}s^{K+2})\ldots(q^L s^L)^{n_{L-K+1}} I_{\sum_{i=0}^{L-K+1} n_i \leqslant n}$$

Now $a_n s^n$ can be written as

$$a_n s^n = \sum_{n_0=0}^{n}\sum_{n_1=0}^{n}\sum_{n_2=0}^{n}\cdots\sum_{n_{L-K+1}=0}^{n}\binom{n}{n_0 n_1 n_2 \ldots n_{L-K+1}} p^n q^{\sum_{i=0}^{L-K+1} n_i(K+i)}$$

$$I_{\sum_{i=0}^{L-K+1} n_i \leqslant n}\, s^{n+\sum_{i=0}^{L-K+1} n_i(K+i)}$$

$$= \sum_{n_0=0}^{n}\sum_{n_1=0}^{n}\sum_{n_2=0}^{n}\cdots\sum_{n_{L-K+1}=0}^{n} c(n, n_0, n_1, n_2, \ldots, n_{L-K+1}) s^{n+\sum_{i=0}^{L-K+1} n_i(K+i)}$$

The coefficients of common powers of s must now be identified by examining the powers of s generated for each n, n_0 and n_1 such that $0 \leqslant n_0$, $0 \leqslant n_1$, $0 \leqslant n_2, \ldots, 0 \leqslant n_{L-K+1}$ and $n_0 + n_1 + n_2 + \ldots + n_{L-K+1} \leqslant n$. Thus, in general the coefficient of s^n is

$$a_n = \sum_{m=0}^{n^-}\sum_{m_0=0}^{m}\sum_{m_1=0}^{m}\sum_{m_2=0}^{m}\cdots\sum_{m_{L-K+1}=0}^{m} c(m, m_0, m_1, m_2, \ldots, m_{L-K+1})$$

$$I_{m+\sum_{i=0}^{L-K+1} m_i(K+i)=n} = \sum_{m=0}^{n}\sum_{m_0=0}^{m}\sum_{m_1=0}^{m}\sum_{m_2=0}^{m}\cdots\sum_{m_{L-K+1}=0}^{m}\binom{m}{m_0 m_1 m_2 \ldots m_{L-K+1}}$$

$$p^m q^{\sum_{i=0}^{L-K+1} m_i(K+i)} I_{\sum_{i=0}^{L-K+1} m_i \leqslant m} I_{\sum_{i=0}^{L-K+1} m_i(K+i)=n}$$

Corollary

Assertion:

If

$$G(s) = \frac{1 - qs + (qs)^K - (qs)^{L+1}}{(1-qs)\left(1 - ps - \sum_{r=K}^{L} pq^r s^{r+1}\right)}$$

then the coefficient for the term s^n, $n \geqslant 0$

$$T_{K,L}(n) = c_n - qc_{n-1} + q^k c_{n-k} - q^L c_{n-L-1}$$

where

$$a_H = \sum_{m=0}^{H}\sum_{m_0=0}^{m}\sum_{m_1=0}^{m}\sum_{m_2=0}^{m}\cdots\sum_{m_{L-K+1}=0}^{m}\binom{m}{m_0 m_1 m_2 \ldots m_{L-K+1}} p^m q^{\sum_{i=0}^{L-K+1} m_i(K+i)}$$

$$I_{\sum_{i=0}^{L-K+1} m_i \leqslant m} I_{\sum_{i=0}^{L-K+1} m_i(K+i)=H}$$

$$c_n = \sum_{j=0}^{n} a_j q^{n-j}$$

Proof:

Begin by writing $G(s)$ as

$$G(s) = \frac{1 - qs + (qs)^K - (qs)^{L+1}}{(1-qs)\left(1 - ps - \sum_{r=K}^{L} pq^r s^{r+1}\right)}$$

The sum of generating functions are inverted to the sum of coefficients of power series.

Begin by noting the inversion of the denominator is a convolution, the nth term begin

$$\sum_{j=0}^{n} a_j q^{n-j}$$

with a_j defined from the previous theorem as

$$a_n = \sum_{m=0}^{m}\sum_{m_0=0}^{m}\sum_{m_1=0}^{m}\sum_{m_2=0}^{m}\cdots\sum_{m_{L-K+1}=0}^{m}\binom{m}{m_0 m_1 m_2 \ldots m_{L-K+1}} p^m q^{\sum_{i=0}^{L-K+1} m_i(K+i)}$$

$$I_{\sum_{i=0}^{L-K+1} m_i \leqslant m} I_{\sum_{i=0}^{L-K+1} m_i(K+i)=n}$$

The inversion of $G(s)$ is completed by examining each of the four terms of the numerator, using the power of s in that term to choose the coefficient of the power expansion of s^n for from the denominator. Thus, we identify the coefficient for the term s^n, $n \geqslant 0$

$$T_{K,L}(n) = c_n - qc_{n-1} + q^k c_{n-k} - q^L c_{n-L-1}$$

where

$$c_n = \sum_{j=0}^{n} a_j q^{n-j}$$

$$a_H = \sum_{m=0}^{H}\sum_{m_0=0}^{m}\sum_{m_1=0}^{m}\sum_{m_2=0}^{m}\cdots\sum_{m_{L-K+1}=0}^{m}\binom{m}{m_0 m_1 m_2 \ldots m_{L-K+1}} p^m q^{\sum_{i=0}^{L-K+1} m_i(K+i)}$$

$$I_{\sum_{i=0}^{L-K+1} m_i \leqslant m} I_{\sum_{i=0}^{L-K+1} m_i(K+i)=H}$$

References

David, F.N. and Barton, D.E., 1962. Combinatorial Chance. Griffin, London.

Feller, W., 1968. An Introduction to Probability Theory and its Applications, Vol 1. Wiley, New York.

Mood, A.M., 1940. The distribution theory of runs. Ann. Math Stat., 11: 367–392.

Moyé, L.A., Kapadia, A.S., Cech, I.M. and Hardy, R.J., 1988. The theory of runs with application to drought prediction. J. Hydrol., 103: 127–137.

Saldariaga, J. and Yevjevich, V., 1970. Application of Run-lengths to Hydrologic Series. Hydrology Paper, Colorado State University, Fort Collins, CO.

Şen, Z., 1976. Wet and dry periods of annual flow series. J. Hydrol. Div., ASCE, 102(HY10): 1503–1514.

Şen, Z., 1977. Run sums of annual flow series. J. Hydrol., 35: 311–324.

Şen, Z., 1980. Statistical analysis of hydrologic critical droughts. J. Hydrol. Div., ASCE, 106(HY1): 99–115.

Şen, Z., 1989. The Theory of runs with application to drought prediction — comment. J. Hydrol., 110: 383–391.

Stern, R.D. and Coe, R., 1984. A model fitting analysis of daily rainfall data. J.R. Stat. Soc., A147(Part 1): 1–34.

Yevjevich, V., 1967. An objective approach to definitions and investigations of continental hydrologic droughts. Hydrol. Papers, Colorado State University Publ., Colorado State University, Fort Collins, CO.