Journal of Hydrology (2008) **363**, 18–31

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jhydrol

Cross entropy quantile function estimation from censored samples using partial probability weighted moments

Jian Deng [a,b,*], **M.D. Pandey** [a]

[a] *Department of Civil Engineering, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada*
[b] *School of Resources and Safety Engineering, Central South University, Changsha 410083, Hunan, China*

Received 25 July 2008; accepted 29 September 2008

KEYWORDS
Quantile function;
Censored samples;
Partial probability weighted moment;
Partial minimum cross entropy principle

Summary Extreme quantile estimation using the minimum entropy principle from complete or non-censored samples has been reported in the literature. However, censored samples are often encountered in many engineering analysis such as annual extremes of river discharge and stream flow analysis. In this paper, the principle of minimum cross entropy (CrossEnt) is extended to partial minimum CrossEnt principle, which is defined on a finite interval. By interpreting the partial probability weighted moments (PPWMs) as partial moments of quantile function, the paper presents a new distribution free method for estimating the quantile function of a non-negative random variable using the principle of partial minimum CrossEnt subject to constraints on PPWMs estimated from censored data. Numerical examples are performed to assess the accuracy of extreme quantile estimates computed from censored samples.
© 2008 Elsevier B.V. All rights reserved.

Introduction

The estimation of extreme quantiles corresponding to small probabilities of exceedance (POE) is commonly required in the risk analysis of hydraulic structures. These estimates were used in the design and reassessment of sea and river defenses (Lind et al., 1989). The first step in quantile estimation involves fitting an analytical probability distribution for adequate representation of sample observations. To achieve this, the distribution type is assumed empirically from the available data, and then distribution parameters are suitably estimated using methods such as maximum likelihood method, etc. (Singh, 1998). However, the assumption of the distribution type is always arbitrary, and there is rarely any justification or verification (Lind et al., 1989). Furthermore, the bias and efficiency of quantile estimates remains sensitive to the type of assumed distribution (Pandey, 2000).

* Corresponding author. Address: School of Resources and Safety Engineering, Central South University, Changsha 410083, Hunan, China. Tel./fax: +86 731 8879612.
E-mail address: dengjian@mail.csu.edu.cn (J. Deng).

0022-1694/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2008.09.004

An alternative approach to the distribution fitting comes from the modern information theory in which entropy has been developed. Direct estimation of quantile functions using the maximum entropy principle and minimum cross entropy principle from complete or non-censored samples has been studied by Pandey (2000) and Pandey (2001). However, censored samples are often encountered in many engineering analysis. For example, in hydrology, as small floods are of little relevance to the larger ones, inclusion of data on small floods in estimating high return period floods can sometimes be only of nuisance value (Cunnane, 1987; Wang, 1990). In life testing and reaction-time experiments, observations often were terminated prior to failure or reaction of all sample specimens (Prescott and Walden, 1983; Phien and Fang, 1989; Cohen, 1991; Kroll and Stedinger, 1996; Lu et al., 1999). When a data set contains some observations within a restricted range but otherwise not measured or counted, it is called a censored sample, whose characteristics are quite different from non-censored cases (Cohen, 1991). The objective of this paper is to extend an entropy-based distribution-free method to more practical case of censored data.

Probability weighted moments (PWMs) continues to be an intensive research topic (Greenwood et al, 1979; Landwher et al., 1979; (Hosking and Wallis, 1997; Pandey, 2000; Rasmussen, 2001; Whalen et al., 2002; to name only a few). In contrast with ordinary statistical moments, the main advantage of using PWMs is that their higher order values can be accurately estimated from small samples. Also, PWMs are shown to be fairly insensitive to outliers (extreme observations) in the data, because they are linear combinations of the observed data values, unlike the ordinary moments, where the data are squared, cubed, etc. (Stedinger et al., 1993). Partial PWMs were used in distribution parameter estimation (Wang, 1990,1996,1997). As linear combination of PWM, partial and trimmed L-moments were developed to cope with censored samples (Hosking, 1995; Koulouris et al., 1998; Bayazit and Önöz, 2002; Elamir and Seheult, 2003).

In this paper, CrossEnt method is extended to deal with incomplete censored samples. Partial probability weighted moments (PPWMs) were computed from censored sample and were interpreted as partial moments of quantile functions. A new distribution-free method is presented for estimating the quantile function of a non-negative random variable using the principle of extended minimum CrossEnt subject to constraints on the PPWMs. Numerical examples were presented.

Partial probability weighted moments

Probability weighted moment

The probability weighted moment of a random variable was formally defined by Greenwood et al. (1979) as

$$M_{r,s,t} = E[X^r F^s (1-F)^t] = \int_0^1 [x(F)]^r F^s (1-F)^t \, dF \tag{1}$$

When r, s, t are real numbers (only non-negative integers are considered here). The following two forms of PWM are particularly simple and useful:

Type 1:

$$\alpha_t = M_{1,0,t} = \int_0^1 x(F)(1-F)^t \, dF \tag{2}$$

and

Type 2:

$$\beta_s = M_{1,s,0} = \int_0^1 x(F) F^s \, dF \tag{3}$$

For an ordered complete sample, $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$, the type 1 and type 2 PWMs can be unbiasedly estimated as

$$a_t = \frac{1}{n}\sum_{i=1}^{n}\left[\binom{n-1}{t} x_i\right] \Big/ \binom{n-1}{t} \tag{4}$$

$$b_s = \frac{1}{n}\sum_{i=1}^{n}\left[\binom{i-1}{s} x_i\right] \Big/ \binom{n-1}{s} \tag{5}$$

where $t, s = 0, 1, \ldots, (n-1)$ are non-negative integers.

Partial probability weighted moment

The partial probability weighted moment was defined as

$$M^P_{r,s,t} = \int_{F_0}^1 [x(F)]^r F^s (1-F)^t \, dF \tag{6}$$

where $F_0 = F(x_0)$ is a lower bound of the censored sample, x_0 being the censoring threshold. The type 1 and type 2 PPWMs becomes

Type 1:

$$\alpha^P_t = M^P_{1,0,t} = \int_{F_0}^1 x(F)(1-F)^t \, dF \tag{7}$$

and

Type 2:

$$\beta^P_s = M^P_{1,s,0} = \int_{F_0}^1 x(F) F^s \, dF \tag{8}$$

For an ordered complete sample, $x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n$, the type 1 and type 2 PPWMs can be unbiasedly estimated as Wang, (1990)

$$a^P_t = \frac{1}{n}\sum_{i=1}^{n}\left[\binom{n-i}{t} x^P_i\right] \Big/ \binom{n-1}{t} \tag{9}$$

$$b^P_s = \frac{1}{n}\sum_{i=1}^{n}\left[\binom{i-1}{s} x^P_i\right] \Big/ \binom{n-1}{s} \tag{10}$$

where $s, t = 0, 1, \ldots, (n-1)$ are non-negative integers and

$$x^P_i = \begin{cases} 0 & x_i \leqslant x_0 \\ x_i & x_i > x_0 \end{cases} \tag{11}$$

For a specific value x_0, an empirical frequency estimate of F_0 is

$$\hat{F}_0 = n_0/n \tag{12}$$

where n_0 is the number of occurrences of values which do not exceed x_0 in the sample.

When $F_0 = 0$, α^P_t and β^P_s become α_t and β_s, respectively. Their unbiased estimators a^P_t and b^P_s become a_t and b_s, respectively.

Download English Version:

https://daneshyari.com/en/article/4579126

Download Persian Version:

https://daneshyari.com/article/4579126

Daneshyari.com