UDC 519.2

A probability model for Assessments of system loads

S. A. Alawadhi

Department of Statistics and OR,
Kuwait University,
Kaldiya SAFAT 13060, Kuwait

Abstract. Let $X_1, X_2...$ be a sequence of i.i.d random variable representing successive loads on a system. A system is over loaded by X_n, off target, if the moving average exceeds a critical level. By introducing a two states Markov chain, a formula is established to measure the stationary distribution and the return period of the state of the overload. Closed forms for the return period are provided when loads follow exponential, normal or Poisson distribution. A system design policy is introduced to meet a prescribed *on target percentage.* A case study on law violations reported to police stations in some districts in Kuwait is also included.

Keywords: Markov chain, computational methods, stationary distributions.

1. Introduction

Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables, and let Y_n be the corresponding left sided moving average, as defined in the abstract. In practice, the input sequence $\{X_n\}$ may represent successive loads, excess loads, rain falls, water supply in successive periods, service time to the n^{th} arrival, etc; and the moving averages are processes indicating accumulations of certain number of immediate prior inputs. Thus by taking into account $K-1$ immediate prior inputs to the n^{th} input, the cumulative value corresponding to the n^{th} input is $\sum_{i=0}^{K-1} X_{n-i},\ n = K, K+1, K+2, \cdots$, and $Y_n = \frac{1}{K}\sum_{i=0}^{K-1} X_{n-i},\ n = K, K+1, K+2, \cdots$ is a sequence of moving averages. The process Y_n is off or on target at the commencement of the arrival of the $(n+1)^{\text{th}}$ input if $Y_n > L$ or $Y_n \leq L$ respectively. The threshold L is non-random and is considered as a parameter. Our aim in this article is to specify, or estimate, L so that the moving averages remains $(1-a)\%,\ 0 < a < 1$, of times on target.

We prove that the status, off or on target, is indeed a two state Markov chain, and derive formulas for the transition probabilities in terms of the distribution of the inputs. This allows to define a prescribed "on target significant level" for the moving averages, and then proceed to introduce a method to achieve the aim. We have examined our method for exponential or normal inputs. Interestingly in these cases L turns out to be linear in the mean of the distribution of the inputs, μ_{X_1}. Point estimation and interval estimation can be easily established using the derived linear relationships.

The methodology and results presented in this article, we believe, can be applied in Reliability, Control Theory, System Assessments, and Hydrology. Moving averages are classical tools in time series, stochastic processes and scan statistics; and are basis for many linear and nonlinear models. Moving averages, in the content presented here, had not been treated in other works, to the best of the authors' knowledge. The threshold of moving averages, considered in this article, is different from the threshold moving average which is a nonlinear model, De Gooijer (1998). Two-state Markov chains, in contents different from the one presented in this article, have been employed by different authors as underlying probability models of various hydrology events, Vogel (1987). The works Bonifacio and Salas (1999) and references therein are rich in providing applications of these types of probability techniques to hydrology data.

2. A Markov Chain

Let $X_1, X_2, \cdots$, and Y_n be as defined in the Introduction. Define

$$V_n = \begin{cases} 0, & Y_n > L \\ 1, & Y_n \leq L \end{cases} , n = K, K+1, \cdots$$

We recall that the situation $V_n = 0$ indicates that Y_n is off target by X_n, while $V_n = 1$ indicates that it is not. We prove below that $\{V_n\}$ is indeed a Markov chain and provide its transition probabilities.

Lemma 1. The process $V_n,\ n = K, K+1, \cdots$, is a Markov chain with transition probabilities.

$$P_{00} = \frac{\int_{-\infty}^{+\infty} [1 - F(KL - t)]^2 f_{T_{K-1}}(t)dt}{1 - F_{T_K}(KL)}, \quad K \geq 1, \tag{2.1}$$

$$P_{11} = \frac{\int_{-\infty}^{+\infty} [F(KL - t)]^2 f_{T_{K-1}}(t)dt}{F_{T_K}(KL)}, \quad K \geq 1, \tag{2.2}$$

where F is the distribution of X_1, and $T_K = X_1 + X_2 + ... + X_K,\ T_0 = 0$.

By using the transition probabilities, the stationary distribution of the Markov Chain $\{V_n\}$ is easily given by

$$\pi_0 = \frac{P_{10}}{P_{10} + P_{01}}, \quad \pi_1 = \frac{P_{01}}{P_{10} + P_{01}}, \tag{2.3}$$

Karlin and Taylor (1998). The return period of the state 0 and state 1 are respectively

$$m_{00} = \frac{1}{\pi_0}, \quad m_{11} = \frac{1}{\pi_1},$$

which specify the duration of successive visits to these states. Other duration are measured by

$$m_{01} = \frac{1}{1-P_{00}}, \quad m_{10} = \frac{1}{1-P_{11}}.$$

Now we are in a position to define "on target significant level".
Definition 1.1. We call the $(1-a)\%$ the "on target significant level" of the moving average process $\{Y_n\}$, where $a = \pi_0$ is the stationary probability of the state 0 of the Markov chain $\{V_n\}$.

3. Exponential And Normal Inputs

In this section we establish a relationship between the threshold L and the mean of the distribution of inputs, whenever the distribution is exponential or normal. Let us assume loads $X_1, X_2, \cdots$ are i.i.d. exponentially distributed with parameter λ, $E(X_1) = 1/\lambda$. The following theorem specifies the appropriate threshold for the moving average to possess the on target $(1-a)\%$ significant level. **Theorem 3.1.** If inputs $X_1, X_2, \cdots$ follow exponential distribution with parameter λ, then the least value L for the threshold to ensure $(1-a)\%$ on target significant level for the moving average Y_n is given by

$$L = \frac{\theta(a,K)}{K}(\frac{1}{\lambda}), \qquad (3.1).$$

where $\theta(a,K)$ is the positive solution to the equation

$$\pi_1(\theta, K) = 1 - a, \qquad (3.2)$$

and $\pi_1(\theta,K)$ is given by (2.3) with

$$P_{00} = (K-1)\frac{N(\theta, K-2)}{(K-1)! - G(\theta, K-1)}, \quad \theta = \lambda K L, \qquad (3.3)$$

and

$$P_{11} = (K-1)\frac{G(\theta,K-2) + N(\theta,K-2) - \frac{2}{K-1}e^{-\theta}\theta^{K-1}}{G(\theta,K-1)}, \quad \theta = \lambda K L, \qquad (3.4)$$

where

$$G(\theta,K) = \int_0^\theta x^K e^{-x} dx\ , \quad N(\theta,K) = \int_0^\theta (\theta - x)^K e^{-(\theta+x)} dx.$$

Remark 3.1. For $K = 7$, we solved (3.2) for the $\theta(a,K)$ with different values of $1-a$, using Mathematica 3.0, Wolfram (1991). The solutions are given in Table 1. We notice from Fig. 2 that $\pi_1(\theta, 7)$ is strictly increasing, providing a unique solution for $\theta(a, 7)$.

Table 1: Exponential Distribution; Significant Levels and Corresponding $\theta(a, 7)$ in (3.2).

$1-a$	0.9	0.8	0.7	0.6	0.5
$\theta(a,7)$	8.197	5.651	3.507	1.625	0

Normal Distribution. Suppose the inputs $X_1, X_2...$ are i.i.d normally distributed with mean μ and standard deviation σ. Interestingly, in this case also L is linear in μ. Details are given below.

Theorem 3.2. If inputs $X_1, X_2, \cdots$ follow normal distribution with mean μ and standard deviation σ, then the least value L for the threshold to ensure $(1-a)\%$ on target significant level for the moving average Y_n is given by

$$L = \mu + \eta(a, K)\sigma, \tag{3.5}$$

where $\eta(a, K)$ is the solution to the equation

$$\pi_1(\eta, K) = 1 - a, \tag{3.6}$$

and $\pi_1(\eta, K)$ is given by (2.3) with

$$P_{00} = \frac{C(\eta, K)}{1 - \Phi(\sqrt{K}\eta)}, \qquad \eta = \frac{L - \mu}{\sigma}, \tag{3.7}$$

and

$$P_{11} = \frac{B(\eta, K)}{\Phi(\sqrt{K}\eta)}, \qquad \eta = \frac{L - \mu}{\sigma} \tag{3.8}$$

where

$$C(\eta, K) = \frac{1}{\sqrt{2\pi(K-1)}} \int_{-\infty}^{+\infty} [1 - \Phi(x)]^2 \, e^{-\frac{1}{2(K-1)}(x - K\eta)^2} dx,$$

and

$$B(\eta, K) = \frac{1}{\sqrt{2\pi(K-1)}} \int_{-\infty}^{+\infty} [\Phi(x)]^2 \, e^{-\frac{1}{2(K-1)}(x - K\eta)^2} dx,$$

Remark 3.2. For $K = 4$, the (3.6) is solved for $\eta(a, K)$ with different values for $1 - a$, using Mathematica. The version of Mathematica that we used did not solve the (3.6) directly, so we had to bypass this barrier

Table 2: Normal Distribution; Significant levels and corresponding $\eta(a)$ in (3.6)

$1-a$	0.9	0.8	0.7	0.6	0.5
$\eta(a)$	0.65	0.47	0.28	0.14	0

by approximating the integrals involved in the equation by corresponding summations. The solutions are given in Table 2. The threshold L in (3.1) is also plotted in terms of the mean μ for $\sigma = 1$

Remark 3.3. The (3.1) and (3.5) can also be used estimation purposes when L is considered as an unknown parameter. It easily follows that for exponential and normal inputs, respectively

$$\hat{L} = \frac{\theta(a, K)}{K}\overline{x},$$

$$\hat{L} = \overline{x} + \eta(a, K)s.$$

Remark 3.4. Although the exponential and normal distributions were treated explicitly, the method, nevertheless, can be carried out for other distributions in order to identify or estimate the threshold parameter.

References

1. *Bonifacio F., Salas G. D.* Return Period and risk of hydrologic events // Journal of Hydrologic Engineering. — 1999. — Vol. 4, no. 4. — P. 297–315.
2. *De Gooijer J.* On threshold moving-average models // J. Time Ser. Anal. — 1998. — Vol. 19, no. 1. — P. 1–18.
3. *Karlin S., Taylor H. M.* An Introduction Stochastic Modeling. — Academic Press, New York, 1998.
4. *Vogel R. M.* Reliability indices for water supply and generation of droughts // Proc. 89 Nat. Conf. on Hydr. Engrg., ASCE, New York, 1987. — P. 50–57.
5. *Wolfram S.* Mathematica 3.0. — Adison Wesley, California, Reaward City, Second Edition, 1991.