AUSTRIAN JOURNAL OF STATISTICS
Volume 27 (1998), Numbers 1&2, 49–57

Estimation of the Parameters in the Double-periodic Model

Z. Khalil[1], M.A. El-Saidi[2], and F. Triantafillou[1]
[1]Concordia University, Montreal, Canada
[2]Ferris State University, Big Rapids, USA

Abstract: The periodic behavior of environmental conditions and its effects on waiting time is of principal interest in a number of modeling problems. Dimitrov and Khalil (1992) have used a constructive approach to introduce a new class of probability distributions which exhibits the periodic behavior of environmental conditions in time and the random occurrence of some events on each time period. Further, the authors discussed some physical and probabilistic properties of the new class. In a more recent work, Dimitrov et al. (1996) investigated a somewhat more complicated case of modeling the waiting time to occurrence of a random event governed by random environment with driving periodic or double periodic structure. In this contribution, we will make use of these important results to find estimators of the parameters in the double periodic model, where after a given number of time periods, say *m*, the conditions start to repeat, the same as from the origin. Phenomena of this type appear in a series of environmental, maintenance and financial processes. In particular, we expect that investigators working in modeling environmental evolution with periodic behavior will find our new results useful.

Keywords: Estimation, Failure Rate Function, Modeling, Periodicity, Waiting Time.

1 Introduction

Nature in general and our environment in particular, in its process of evolution, possesses many properties of periodicity. There exist events, the occurrence or non-occurrence of which depend heavily on environmental conditions. These conditions regenerate periodically. For example, the well-known cycle of tides on beaches of large oceans generates on each cycle the same probabilistic conditions for the evolution of random changes between water and continental ridges. Our own lives are controlled by the 24-hour cycle in a smaller scale, the seasonal (yearly) cycle in a larger scale, and if we look deeper into centuries we find also cycles in a larger scale. Jumping to gigantic scales, we see also cyclicity (pulsation) in the evolution of the universe.

The use of conventional probabilistic distributions for random variables without considering the periodicity of the conditions, is only an approximate way to describe the dynamics of our environment.

A class of probabilistic distributions exhibiting a periodic behavior in time and a random occurrence of some event on each period was introduced by Dimitrov and Khalil (1992). Further work on this class of distributions has been done by Dimitrov, Khalil and

El-Saidi (1996). They considered the more complicated case of modeling the waiting time to occurrence of a random event governed by random environment with driving-periodic or double-periodic structure. This means that the environmental conditions are repeatedly acting, and may be step-wise changing from one period to the next one (in some neighboring periods of length c>0). In the double-periodic case, after a given number of periods, say *m*, the conditions start to repeat as from the origin.

As an example, we consider the day-by-day changes of environmental conditions in a short term periodic environment, and the 365 day period of one year as the long term periodicity for surrounding environmental conditions. Analogously, moon periods of 28 days can be considered as a type of short term period for the bay tides, and the time of one earth orbit of the sun corresponds to a long term period.

In the present article, taking the previous results into consideration, we undertake a statistical analysis and establish estimators of the parameters in the double-periodic model. The periodic behavior of the environmental conditions and its effects on waiting times is of principle interest in a number of modeling problems in applied stochastic analysis.

2 The Double Periodic Model

2.1 Double Periodic Environment and Its Related Probability Distribution

Let X be a random variable representing the time of occurrence of an event in a double periodic random environment. We assume that :

1. the chances of an event occurring during a trial that lasts $c > 0$ are different from one trial to another,
2. the time within a period until the occurrence of the event under the condition that the event will occur in this period, has the same distribution for any other period, and
3. the time half-axis $[0, \infty)$ is subdivided into equal intervals of length c each, $c > 0$.

These assumptions completely specify the distribution of waiting time until the event occurs. Since each trial lasts time c, thus $[kc, (k+1)\,c)$; $k = 0, 1, 2, ...$ is the k^{th} interval of this subdivision during which the k^{th} trial will be performed. Then we consider a sequence of extended in time independent trials. The chance to have a success in the k^{th} trial is α_k, where $0 < \alpha_k < 1$. We say that the sequence $\{\alpha_k\}$ is periodic with period m if $m > 0$ is an integer, and the equation

$$\alpha_{k+rm} = \alpha_k \tag{1}$$

holds for any integer $r > 0$.

A periodic sequence will model the long-term periodic environment, whose period appears to be mc. Thus the global environmental conditions are considered to repeat in intervals of the form $[rmc, (r+1)mc),\ r = 0, 1, 2, \ldots$.

Now let us consider what may happen in a small interval $[kc, (k+1)\,c)$. We assume that the conditional distribution of the occurrence time of the random event during the current trial, given that it will occur on this trial and has not been observed on the preceding $(k-1)$ trials, is the same for any trial. That is the following equation:

$$P\,(X < kc + y \mid kc \le X < (k+1)\,c) = F_Y\,(y) \tag{2}$$

holds for any $k = 0, 1, 2, \ldots$, and all $y \in [0, c)$.

The random variable Y represents the time to occurrence of the considered random event within an extended in time trial, under the condition that this event has occurred, and is assumed not to depend on the number of trials. Thus $F_Y\,(y) = P\,(Y < y)$ denotes the cdf of the random variable Y. Since $Y \in [0, c)$ with probability 1, we have:

$$F_Y\,(0) = 0; \quad F_Y\,(c) = 1 \tag{3}$$

The cdf of the waiting time X to the occurrence of a random event in a sequence of extended in time independent trials in a double periodic environment has been obtained in Dimitrov et al. (1996). It has the following form:

$$F_X\,(x) = 1 - \beta^{\left[\frac{x}{mc}\right]}\left(1 - (1-\beta)\,F_{Y_m}\left(x - \left[\frac{x}{mc}\right]mc\right)\right), \quad x \ge 0 \tag{4}$$

The corresponding pdf of the random variable X, if it exists, has the form:

$$f_X\,(x) = \beta^{\left[\frac{x}{mc}\right]}\,(1-\beta)\,f_{Y_m}\left(x - \left[\frac{x}{mc}\right]mc\right), \ x \ge 0 \tag{5}$$

where

$$\beta = \prod_{i=1}^{m}(1 - \alpha_i) \tag{6}$$

represents the probability $P\,(X \ge mc)$, $\left[\frac{x}{mc}\right]$ denotes the integer part of $\frac{x}{mc}$, and the random variable Y_m is the conditional length of X given that $X < mc$, with cdf defined by:

$$F_{Y_m}\,(y) = \frac{1}{1-\beta}\left(1 - \prod_{i=1}^{\left[\frac{y}{c}\right]}(1-\alpha_i)\left(1 - \alpha_{\left[\frac{y}{c}\right]+1}F_Y\left(y - \left[\frac{y}{c}\right]c\right)\right)\right) \tag{7}$$

for all $y \in [0, c)$.

2.2 The Failure Rate Function of X

Let

$$\lambda_Z\,(t) = \frac{f_Z\,(t)}{1 - F_Z\,(t)}, \quad t \in [0, c) \tag{8}$$

be the failure rate function corresponding to the cdf $F_Z(t)$, see Barlow and Proschan (1981). We assume that the pdf $f_Z(t)$ corresponding to the cdf $F_Z(t)$ exists in order for

$\lambda_Z(t)$ to exist.

Lemma 1 *The failure rate function of the random variable* X*, assuming its pdf (5) exists, is determined by the equation:*

$$\lambda_X(x) = \frac{(1-\beta)\, f_{Y_m}\left(x - \left[\frac{x}{mc}\right] mc\right)}{1-(1-\beta) - F_{Y_m}\left(x - \left[\frac{x}{mc}\right] mc\right)} \tag{9}$$

Proof. *Can be easily obtained from* (4), (5), *and* (8).

Note that the function $\lambda_X(x)$ defined by (9) gives the chance of occurrence at time x (a realization) of the random variable X, if up to that time the realization of X had not yet been observed.

Lemma 2 *The failure rate function* $\lambda_X(x)$ *defined by* (9) *is periodic with period* mc.
Proof. *Since*

$x + rmc - \left[\frac{x+rmc}{mc}\right] mc = x + rmc - \left[\frac{x}{mc} + r\right] mc = x - \left[\frac{x}{mc}\right] mc,$
is true for any integer $r > 0$, *then it is clear that* $\lambda_X(x+rmc) = \lambda_X(x)$, $\quad x \geq 0$.

3 Estimation of the Parameters β and $F_{Y_m}(y)$

In practice, it is usually common that the periodicity parameter is known. However, a question still remains about statistical estimation of the parameters in the double periodic model; namely the probability β and the cdf $F_{Y_m}(y)$ with all its parameters and subsequently the failure rate function $\lambda_X(t)$ with all its parameters.

Theorem 1 *Let* $X_1, X_2, ..., X_n$ *be the realizations of the first event occurrence in N independent copies of the model with period* $mc > 0$. *Suppose the observations have been obtained during a time* $T = rmc$, *i.e. the phenomenon of interest has occurred* n *times and did not occur in* $N-n$ *copies. Then the estimators of the probability* β *and the cdf* $F_{Y_m}(t)$ *are given respectively by:*

$$\hat{\beta} = \frac{\sum_{i=1}^{n}\left[\frac{X_i}{mc}\right] + r(N-n)}{n + \sum_{i=1}^{n}\left[\frac{X_i}{mc}\right] + r(N-n)} \tag{10}$$

and

$$\hat{F}_{Y_m}(t) = \begin{cases} 0 & \text{for } t \leq Y_{m(1)} \\ \frac{k}{n} & \text{for } t \leq Y_{m(k+1)} - Y_{m(k)} \\ 1 & \text{for } t \leq Y_{m(n)} \end{cases} \tag{11}$$

where $Y_{m(1)} \leq Y_{m(2)} \leq ... \leq Y_{m(n)} \leq mc$ *are the order statistics corresponding to* $Y_1, Y_2, ..., Y_n$, *constructed according to the following rules:*

$$Y_{m(i)} = x_i - \left[\frac{x_i}{mc}\right] mc, \quad i = 1, 2, ..., n \tag{12}$$

Proof. *The likelihood function* $L(x_1, x_2, ..., x_n, T, N)$ *is given by*

$$L = \beta^{\sum_{i=1}^{n}\left[\frac{x_i}{mc}\right]} (1-\beta)^n \cdot \left\{\prod_{i=1}^{n} f_{Y_m}\left(x_i - \left[\frac{x_i}{mc}\right] mc\right) \beta^{\left[\frac{T}{mc}\right](N-n)}\right\}.$$

$$\left\{\beta + (1-\beta)\left(1 - F_{Y_m}\left(T - \left[\frac{T}{mc}\right] mc\right)\right)\right\} \tag{13}$$

When the observation time T *is exactly a multiple of* mc*, i.e.* $T = rmc; r = 0, 1, 2, ...,$ *then the factor* $\beta^{\left[\frac{T}{mc}\right](N-n)}$ *giving the probability of non-occurrence of the event over the* $(N-n)$ *copies, reduces to* $\beta^{r(N-n)}$*. Then the log-likelihood function* $\ell = \ln L\left(x_i, \beta, F_{Y_m}\right)$ *takes the form:*

$$\ell = \left(\sum_{i=1}^{n}\left[\frac{x_i}{mc}\right] + r\left(N-n\right)\right)\ln\beta + n\ln\left(1-\beta\right) + \sum_{i=1}^{n}\ln f_{Y_m}\left(x_i - \left[\frac{x_i}{mc}\right] mc\right) \tag{14}$$

This shows that the value of the parameter β *which maximizes the log-likelihood function in (14) does not depend on the function* $f_{Y_m}(.)$ *that also maximizes the same likelihood function. Moreover, in all terms expressing the dependence on the parameter* $F_{Y_m}(.)$*, only the order statistics* $Y_{m(i)},\ i = 1, 2, ..., n$*, calculated by (12) are present. Therefore the maximum likelihood estimator (MLE) of the second parameter* $F_{Y_m}(.)$ *will be a function of these order statistics.*

Differentiating (14) with respect to β *and setting* $\frac{\partial \ell}{\partial \beta} = 0$ *gives the MLE* $\hat{\beta}$ *specified by (10). The MLE for* $F_{Y_m}(.)$ *given by (11) follows from the above reasoning and the fact that* $Y_{m(i)}$ *defined by (12) are independent realizations of the random variable* Y_m *with cdf* $F_{Y_m}(.)$*.*

Remark 1 *Observe that for* $m = 1$ *equation (10) reduces to*

$$\hat{\beta} = \frac{\sum_{i=1}^{n}\left[\frac{X_i}{c}\right] + r\left(N-n\right)}{n + \sum_{i=1}^{n}\left[\frac{X_i}{c}\right] + r\left(N-n\right)}$$

which is the MLE for the parameter β *of the model presented by Dimitrov and Khalil (1992).*

Corollary 1 *For any integer* $k > 0$*, the MLE's for* α_k *are given by*

$$\hat{\alpha}_k = 1 - \frac{\hat{\beta}}{\prod_{i=1}^{k-1}\left(1-\hat{\alpha}_i\right)} \tag{15}$$

Proof. *Equation (6) shows that* $\beta = \prod_{i=1}^{m}\left(1-\alpha_i\right)$*. Then we have:*

$$\begin{array}{lll} \text{for} & m=1 & \hat{\alpha}_1 = 1-\hat{\beta} \\ \text{for} & m=2 & \hat{\alpha}_2 = 1-\frac{\hat{\beta}}{1-\hat{\alpha}_1} \\ \text{for} & m=3 & \hat{\alpha}_3 = 1-\frac{\hat{\beta}}{(1-\hat{\alpha}_1)(1-\hat{\alpha}_2)} \end{array}$$

Generally for $m = k$*, we obtain the expression given in (15).*

Corollary 2 *Under the conditions of Theorem 1, the estimators of the unknown parameters in* $F_{Y_m}(.)$ *include only the order statistics* $Y_{m(i)},\ i = 1, 2, ..., n$*, given by (12). These unknown parameters say* θ*, can be obtained using the conventional methods.*

Proof. *If we assume that the random variable* Y_m *has a distribution law depending on one parameter* θ*, then the dependence on* θ *in the log-likelihood function appears only in the last sum in (14); namely*

$$\ell = \ln L\left(x_i, \beta, F_{Y_m}\right) = ... + \sum_{i=1}^{n} \ln f_{Y_m}\left(Y_i, \theta\right).$$

All other terms depend only on β*. Therefore the MLE for* θ *and all other possible estimators for* θ *obtained on the basis of the order statistics* $Y_{m(i)},\ i = 1, 2, ..., n$*, given by (12) cannot have any influence on the estimators of the parameter* β*.*

3.1 Estimation of the Failure Rate Function of X

The estimator for the failure rate function

$$\lambda_X(x) = \frac{(1-\beta) f_{Y_m}\left(x - \left[\frac{x}{mc}\right] mc\right)}{1 - (1-\beta) - F_{Y_m}\left(x - \left[\frac{x}{mc}\right] mc\right)}$$

can now be easily obtained by substituting the estimators $\hat{\beta}$ of β and $\hat{F}_{Y_m}(t)$ of $F_{Y_m}(t)$ derived earlier in (10) and (11) into the above expression.

4 Distribution of $F\left(Y_{m(k)}\right)$

Theorem 2 *The random variable* $F\left(Y_{m(k)}\right)$ *has a Beta distribution with parameters* $(k, n-k+1)$.

Proof. *Consider the order statistics of the random sample* $Y_{m(1)} \leq Y_{m(2)} \leq ... \leq Y_{m(n)} \leq mc$*. Since* F_{Y_m} *is an increasing function, then*

$$F_{Y_m}(Y_{m(1)}) \leq F_{Y_m}(Y_{m(2)}) \leq ... \leq F_{Y_m}(Y_{m(n)}) \leq 1.$$

Moreover, since F_{Y_m} *is continuous, then* $F_{Y_m}\left(Y_m\right)$ *is uniformly distributed over* $[0, 1]$*. Also, since* $F\left(Y_{m(k)}\right)$ *is the* k^{th}*order statistics of a sample of size n, then its pdf, see Hogg and Craig (1995), is given by:*

$$f_k\left(F\left(Y_{m(k)}\right)\right) = \frac{n!}{(k-1)!\,(n-k)!}\left(F\left(Y_{m(k)}\right)\right)^{k-1}\left(1 - F\left(Y_{m(k)}\right)\right)^{n-k}.$$

Therefore $F\left(Y_{m(k)}\right)$ *has a Beta distribution with parameters* $(k, n-k+1)$ *and the proof is complete.*

5 A Numerical Example

To illustrate the proposed model, consider a water purifying system of a given structure. The system performs under some maintenance instructions and conditions of work as follows:

1. There is a periodic cleaning procedure for the purifying tools every day in order to ensure a certain quality of the water supply at the output.

2. Every 10 days there is another basic stress cleaning procedure that needs more specific chemicals, energy and special treatments.

3. The last operations are believed to completely restore the original features of the purifier and its production to be as new.

Suppose the water is used to wash some material. Appearance of stains on the washed materials indicates something wrong has happened to the water production of the purifier. This may be due to the incidence of random quantities of salts, input of unexpected amounts of waste into the water supply and possibly other reasons caused by the environment. Assume that the probability of detecting a stain on the washed material during the first day is 0.01, during the second day 0.02 up to the tenth day (just before the stress cleaning) is 0.1. The next day, after the basic stress cleaning of the purifier, this probability is again 0.01 and so on.

Assume that stains have been observed within a particular day, and that the time *Y* of the occurrence of the first stain measured from the beginning of that day has a triangular distribution over the interval [0,1).

The pdf and cdf of *Y* are, respectively,

$$f_Y(y) = 2y, \quad for\ y \in [0,1)$$

and

$$F_Y(y) = y^2, \quad for\ y \in [0,1)$$

Then, the failure rate function is given by

$$\lambda_Y(y) = \frac{y^2}{1-y^2}, \quad for\ y \in [0,1).$$

Let *X* the time from the beginning of work with the purifier until some stains are observed for the first time on the washed material under the above described double-periodic maintenance policy and *Y* as described above within the small period. Then the random variable *X* is generated by environmental conditions that completely correspond to the proposed model. In particular we have

$$c = 1, \quad \alpha_k = (0.01)\left(k - \left[\frac{k-1}{10}\right]10\right), \qquad m = 10$$

The following characteristics of X can be explicitly derived: $f_X(x)$, $F_X(x)$, and $\lambda_X(x)$. Graphical representations of these functions for this example are shown in Figures 1 and 2. Figure 1 shows that the failure rate varies from the 1st to 10th day, and then copies the same behavior for the next 10 days. This figure shows a typical shape and behavior for the class of probability distributions generated by a double periodic random environment. Figure 2 shows that the pdf $f_X(x)$ varies on the larger period of 10 days within any small period of 1 day, and preserving its global shape. For the next 10 days it decreases proportionally. The figure also shows that the cdf $F_X(x)$ increases fast for the first two larger periods and then slowly tends to 1, but never gets the value 1 on any finite interval. This means that the waiting time until the event of interest occurs, may have arbitrary large values with positive probability.

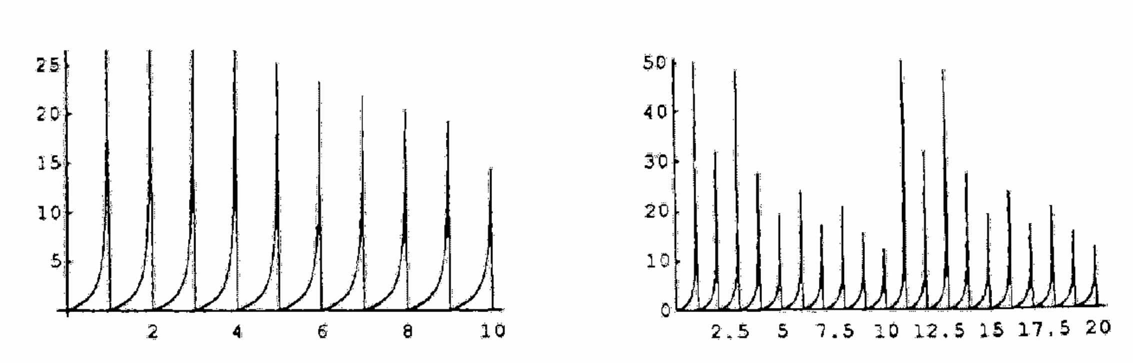

Figure 1: Failure rate $\lambda_X(x)$ for water purifier example in 10 and 20 days

Acknowledgement: We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada, Grant # 9095.

References

R. Barlow and F. Proschan. *Statistical theory of reliability and life testing*. 2nd ed., Holt Rinehart and Wilson, 1991.

B. Dimitrov and Z. Khalil. A new class of probability distributions for modelling environmental evolution with periodic behavior. *Environmetrics*, 3(4):447-464, 1992.

B. Dimitrov, Z. Khalil, and M.A. El-Saidi. On probability distributions with accumulating failure rates in periodic random environment. *Environmetrics*, 7(4):17-26, 1996.

R.V. Hogg and A.T. Craig. *Introduction to Mathematical Statistics*. 5th ed., Prentice Hall, 1995.

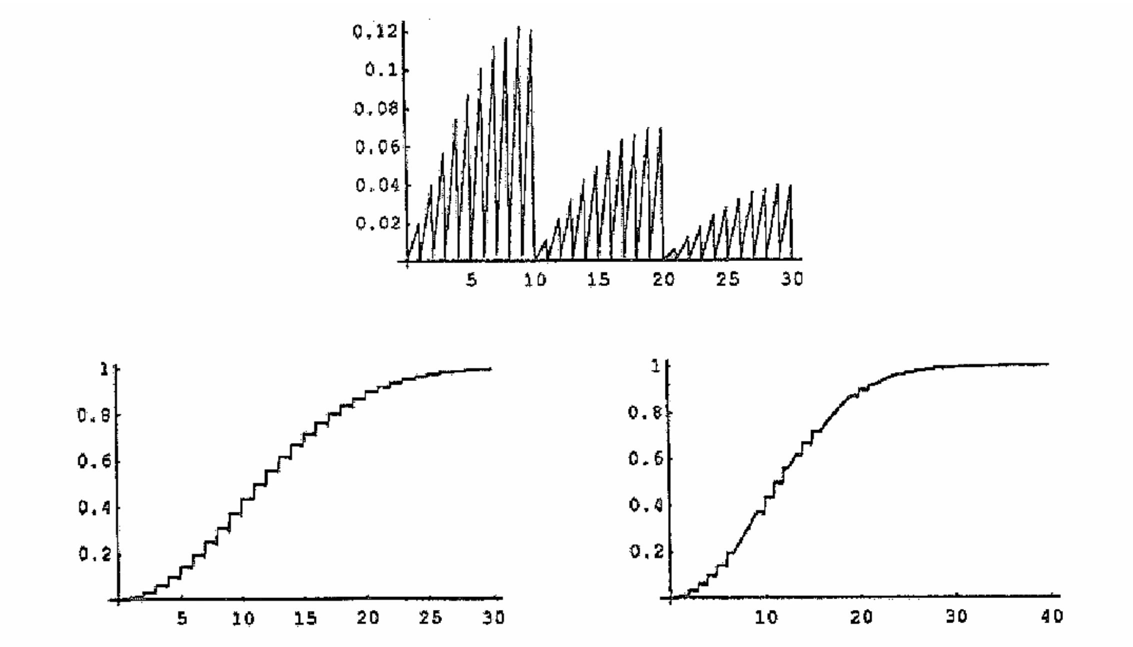

Figure 2: Probability density and cumulative distribution functions of the time to occurrence of the event of interest for the water purifier example

Authors' addresses:

Z. Khalil
F. Triantafillou
Concordia University
7141 Sherbrooke St.
Montreal, Canada H4B 1R6

M.A. El-Saidi
Ferris State University
119 South St.
Big Rapids, MI 49307
U.S.A.

Tel. (616) 592-2479
Fax (616) 592-3521
E-mail: elsaidim@pada.ferris.edu or
Mohammed_El-Saidi@ferris.edu