Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 14 (16): 1894-1897, 2014
ISSN 1812-5654 / DOI: 10.3923/jas.2014.1894.1897

Multiple Linear Regression of Maximum Queue Length Probability Function for Infected Fish in the Fish Farms

Mohammed Mohammed El Genidy
Department of Mathematics, Faculty of Science, Port Said University, Port Said, Egypt

Abstract: This study deal with the problem of obtaining some important information of the infected fish in the fish farms while there was always a difficulty in handling the mathematical formulas obtained in some of the research related to the subjects such as the maximum number of the infected fish in a fish farm using the queue system because a lot of computational procedures were required. The multiple linear regression formula for the probability function of the maximum queue length of infected fish during finite time estimated in number of days and the cumulative distribution function were obtained.

Key words: Queue system, multiple linear regression, maximum queue length, standard normal distribution, cumulative distribution function

INTRODUCTION

The maintenance of fish stock can be done through protecting it against diseases. Accordingly, the human health can be protected against infectious diseases transformed to him or her through the infected fish. Among the strategies of developing the fish resources at Egypt are fish farms constructing and exploring new areas of fish farms in areas where lands are not reclaimable. Among the common fish disease in fish farmed in Egypt are sleeping sickness, whirling, black spot and yellow grab. El Genidy (2011) has conducted study in queue system of maximal queue size with standard normal distribution for arrival times and this study was updated for providing the important relevant statistics in order to predict some related pieces of information in the future. Ekpenyong *et al.* (2008) improved the estimation based on sampling by using non linear regression method. Ismail *et al.* (2009) applied multiple linear regression method to predict the gold prices. Jin *et al.* (2006) carried out a method to estimate time traffic speed using multiple regression model. Okereke (2011a) has study the transformation on the parameter estimates of a simple linear regression model. Okereke (2011b) considered some consequences of adding a constant in the simple linear regression model. Subha and Stuwart (2010) performed multi resolution analysis in the texture classification using linear regression model. Yesmin *et al.* (2004) have done study infection in fish of swamps in Bangladesh. Naich *et al.* (2003) applied the kidney disease at fish farm on the river test. The results of this study will enable the researchers to deal with the amount of the infected fish in the farms with an economical manner and minimal costs.

DESCRIPTION OF THE PROBLEM AND ITS SOLVING

In this study, the distribution of inter-arrival times of infected fish follows the standard normal distribution. Let q_n is the number of infected fish found in the fish farm, the sequence $\{q_n\}$ forms a discrete state Markov chain and the departure distribution is exponential with rate μ. In the queue system the probability of the maximum queue length depends on the following 3 independent events:

A_1 = Is the event that an infected fish arrived at time y, $q_0 = k$, $q_1 = 0$ and $k = 0, 1, \ldots, n$ given that y varies from 0 to t

A_2 = Is the event that an infected fish at time y, $q_0 = k$, $q_1 = k$, $q_1 = k-l+1$ and $k = 0, 1 \ldots, n$ given that y varies from 0 to t

A_3 = Is the event that no infected fish arrived during the time interval [0, t] = the event that the inter-arrival time is greater than t

Then:

$$P_{k,n}(t) = \Pr(A_1) + \Pr(A_2) + \Pr(A_3)$$

It is clear that:

$$P_{k,n}(t) = \sum_{l=0}^{k} \int_0^t \Pr(x_1 = l / y).f(y).P_{k-l+1,n}(t-y).dy$$
$$+ \int_0^t \Pr(q_1 = 0 / q_0 = k, y).f(y).P_{0,n}(t-y).dy + (1 - F(t))$$

Where:

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

Then:

$$F(t) = \int_{-\infty}^{t} f(y).dy = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy$$

On the other hand:

$$\Pr(q_1 = k - l + 1 / q_0 = k, y) = \Pr(x_1 = l / y) = \frac{(\mu y)^l}{l!} e^{-\mu y};\ l = 0, 1, 2, \cdots, k$$

$$\Pr(q_1 = 0 / q_0 = k, y) = \frac{(\mu y)^{k+1}}{(k+1)!} e^{-\mu y}$$

Therefore, when k = n the probability function of the maximum queue length during time (T = t) were obtained as follows:

$$P_{n,n}(t) = \sum_{l=0}^{n} \int_{0}^{t} \frac{1}{\sqrt{2\pi}} P_{n-l+1,n}(t-y) \frac{(\mu y)^l}{l!} . e^{-(\frac{y^2}{2} + \mu y)} .dy + \int_{0}^{t} \frac{1}{\sqrt{2\pi}} P_{0,n}(t-y) . \frac{(\mu y)^{n+1}}{(n+1)!} . e^{-(\frac{y^2}{2} + \mu y)} .dy + (1 - F(t))$$

While:

$$P_{n-l+1,n}(t-y) = \frac{[\lambda(t-y)]^{n-l+1}}{(n-l+1)!} e^{-\lambda(t-y)} \text{ and } P_{0,n}(t-y) = e^{-\lambda(t-y)}$$

Hence:

$$P_{n,n}(t) = P_n(t) = \sum_{l=0}^{n} \frac{\lambda^{n-l+1} \mu^l}{\sqrt{2\pi}(n-l+1)!\, l!} \int_0^t y^l (t-y)^{n-l+1} e^{-\lambda t + \lambda y + \mu y - \frac{y^2}{2}} dy + \frac{\mu^{n+1}}{\sqrt{2\pi}(n+1)!} \int_0^t y^{n+1} e^{-\lambda t + \lambda y + \mu y - \frac{y^2}{2}} dy + (1 - F(t)) \quad (1)$$

where, n is maximum queue length, λ is arrival rate, μ is departure rate, l is the number of departure units between 2 consecutive arrival units in the system and *y* is the time between consecutive arrivals.

Suppose that, arrival rate = λ = infected fish per unit time and departure rate = μ = 2 infected fish per unit time. Then by using mathematical program on Eq. 1, consequently $P_n(t)$ was obtained as follows:

$$P_n(t) = N[\sum_{l=0}^{n} \frac{2^l}{2.5*(n+1-l)!*l!} \int_0^t y^l (t-y)^{n+1-l} \hat{E}(-t + 3y - \frac{y^2}{2}) dy + \frac{2^{n+1}}{2.5*(n+1)!} \int_0^t y^{n+1} \hat{E}(-t + 3y - \frac{y^2}{2}) dy + (1 - F(t))] \quad (2)$$

Equation 2 was converted to the formula of the multiple linear regression of the probability function of the maximum queue length Q infected fish during time T in days as a function in each random variables Q and T, where Q is a discrete random variable and T is a continuous random variable. Apply FIT command by using mathematical program on Eq. 2 as follows:

$$\text{Fit}[\{\{t_i, q_1, p_{q1}(t_i)\}, \{t_i, q_2, p_{q2}(t_i)\}, \ldots, \{t_i, q_m, p_{qm}(t_i)\}, \{t_1, q_j, p_{qj}(t_1)\}, \{t_2, q_j, p_{qj}(t_2)\}, \ldots, \{t_r, q_j, p_{qj}(t_r)\}, (t_i, q_j, p_{qj}(t_i))\}, \ldots, \{t_r, q_m, p_{qm}(t_r)\}\}, \{1, Q, T\}, \{Q, T\}]$$

where, $Q = q_1, q_2, q_3, \ldots, q_m$, $T = t_1, t_2, t_3, \ldots, t_r$, i = 1, 2,..., r and j = 1, 2,..., m. Thus, for any given data of the random variables Q and T, the formula of the multiple linear regression of the probability function Z of the maximum queue length Q during time T were obtained in the form: Z = c-aQ-bT, where a, b, c are constant and c>0, a>0, b>0.

Corollary: If the probability function Z of the maximum queue length Q during time T as a function in each random variables Q and T takes the equation Z = c-aQ-bT, where c>0, a>0, b>0 are constant, $q \to 0$ and $t \to 0$ then $z = c\text{-}aq\text{-}bt \to c$.

Lemma: If the probability function Z of the maximum queue length Q during time T as a function in each random variables Q and T takes the formula Z = c-aQ-bT, where c>0, a>0, b>0 are constant, then the cumulative distribution function F(Q, T) takes the Eq.:

$$F(Q < q, T < t) = \frac{1}{2} bt^2(-1-q) + \frac{1}{2} t(1+q)(2c - aq)$$

Proof: The function:

$$F(Q < q, T < t) = \int_{u=0}^{t} [\sum_{v=0}^{q} (c - av - bu)] du$$

was given then:

$$\sum_{v=0}^{q} (c - a*v - b*u) = c(1+q) - \frac{1}{2} aq(1+q) - b(1+q)u$$

Therefore:

$$F(Q < q, T < t) = \int_{u=0}^{t} [\sum_{v=0}^{q} (c - av - bu)] du = \frac{1}{2}(1+q)(2c - aq)t + \frac{1}{2} b(-1-q)t^2$$

APPLICATION WITH NUMERICAL RESULTS

Ekpenyong *et al.* (2008) and El Genidy (2011) taken longer time in computation processes to find the information because there was not a program to implement. The research by Ismail *et al.* (2009) and

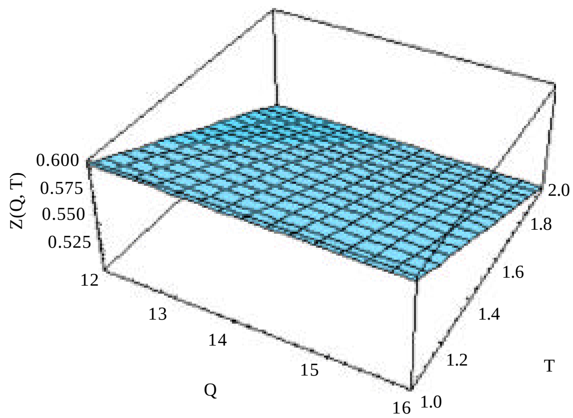

Fig. 1: Multiple linear regression of infected fish in the fish farm

Jin *et al.* (2006) deal with multiple linear regression method without finding maximum queue length probability function. Okereke (2011a, b) and Subha and Stuwart (2010) performed a study in a simple linear regression model and it could not use in multiple linear regression model. Yesmin *et al.* (2004) and Naich *et al.* (2003) needed many procedures and computations to get the information. Method of finding information on infected fish in the fish farms in this study is more easy to get the necessary information because there was a program that could be applied to the computer, as well as the probability Z(Q, T) gives the degree of confidence for each result obtained.

Suppose the following data are given:

$Q = q_1, q_2, q_3, q_4, q_5$ = 12, 13, 14, 15, 16 infected fish and $T = t_1, t_2$ = 1,2 day, respectively.

By using the FIT command on the data, the following result was obtained:

Fit[{{1, 12, 0.6025}, {1, 13, 0.6025}, {1, 14, 0.6025}, {1, 15, 0.6025}, {1, 16, 0.6025}, {2, 12, 0.5115}, {2, 13, 0.5054}, {2, 14, 0.5051}, {2, 15, 0.5047}, {2, 16, 0.5046}}, {1, T, Q}, {T, Q}]

$$Z(Q,T) = 0.70889-0.00725\,Q-0.09624\,T$$

The last formula of Z(Q,T) gives an estimation for $P_n(t)$, where $Z \cong P_n(t)$.

Figure 1 of Z (Q, T) was obtained by Plot3D command as follows:

Plot3D [Z, {Q, 12, 16}, {T, 1, 2}]

In Fig. 1, the multiple linear regression of infected fish in the fish farm was represented by the function Z(Q,T) such that the variable Q was represented by the maximum queue length of infected fish in the fish farm and the variable T was represented by the number of days. There are some values of the variables T and Q which lead to Z(Q,T)<0 or Z(Q,T)>1, that is a contradiction, therefore all these values must be excluded. The following program solved this problem such that all the required values of T and Q satisfy the condition of the probability Z(Q,T).

```
REM ProbMaxQueue_prog
INPUT MaxTime
INPUT MaxQueue
c = 0.70889
a = 0.00725
b = 0.09624
FOR t = 1 TO MaxTime
FOR q = 1 TO MaxQueue
z = c-a * q-b * t
IF z<0 OR z>1 THEN 100
PRINT z, q, t
REM "Contradiction Event"
100 NEXT q
NEXT t
END
```

Suppose t = 1, 2 days and q = 1, 2, 3, 4, 5, 6, 7 maximum number of infected fish. After the implementation of the previous program, the following result was obtained.

Table 1: Values of the probability Z with the maximum queue length q during time t

Probability (Z)	Maximum No. of infected fish (q)	No. of days (t)
0.6054	1	1
0.59815	2	1
0.5909	3	1
0.58365	4	1
0.5764	5	1
0.56915	6	1
0.5619	7	1
0.50916	1	2
0.50191	2	2
0.49466	3	2
0.48741	4	2
0.48016	5	2
0.47291	6	2
0.46566	7	2

In Table 1, some important information can be obtained such as, maximum number of infected fish q equals to one infected fish per one day with maximum probability Z = 0.6054, q = 7 infected fish per 2 days with minimum probability Z = 0.46566 and max q = 7 infected fish corresponds to Z = 0.5619 per one day and Z = 0.46566 per 2 days.

CONCLUSION

This study enables the researcher to find some of the required information such as, knowing the maximum time possible when the infected fish found in the fish farm or the maximum possible number of the infected fish that can be within the fish farm. Furthermore, the results of this study will enable the researchers to deal with the fish farms that contain infected fish in an economical manner and with minimal costs. Lastly, the idea of this study allows the researchers to apply it in various scientific fields.

REFERENCES

Ekpenyong, E.J., M.I. Okonnah and E.D. John, 2008. Polynomial (non linear) regression method for improved estimation based on sampling. J. Applied Sci., 8: 1597-1599.

El Genidy, M.M., 2011. Maximal queue size with standard normal distribution for arrival times. Asian J. Math. Stat., 4: 161-167.

Ismail, Z., A. Yahaya and A. Shabri, 2009. Forecasting gold prices using multiple linear regression method. Am. J. Applied Sci., 6: 1509-1514.

Jin, X., S.K. Hong and Q. Ma, 2006. An algorithm to estimate continuous-time traffic speed using multiple regression model. Inform. Technol. J., 5: 281-284.

Naich, M., F.M. Bilqees and N. Khatoon, 2003. Proliferative kidney disease at romsey fish farm on the river test, hampshire, U.K. Pak. J. Biol. Sci., 6: 1907-1909.

Okereke, O.E., 2011a. Effect of transformation on the parameter estimates of a simple linear regression model: A case study of division of variables by constants. Asian J. Math. Stat., 4: 174-180.

Okereke, O.E., 2011b. Some consequences of adding a constant to at least one of variables in the simple linear regression model. Asian J. Math. Stat., 4: 181-185.

Subha, A. and S.L. Stuwart, 2010. Linear regression model on multi resolution analysis for texture classification. Int. J. Comput. Appl., 2: 1-8.

Yesmin, S., M.H. Rahman, M.A. Hussain, A.R. Khan, F. Pervin and M.A. Hossain, 2004. *Aeromonas hydrophila* infection in fish of swamps in Bangladesh. Pak. J. Biol. Sci., 7: 409-411.