I.J. Engineering and Manufacturing, **2012,1, 1-7**
Published Online February 2012 in MECS (http://www.mecs-press.net)
DOI: 10.5815/ijem.2012.01.01

Available online at http://www.mecs-press.net/ijem

A Probability Model for Occurrences of Large Forest Fires

Yixun Shi

Department of Mathematics, Computer Science and Statistics, Bloomsburg University of Pennsylvania, Bloomsburg, PA 17815, USA

Abstract

Forest fires occur at various spots and various times all around the world. One of major efforts in forest fire management is to estimate the probability of occurrence of fires so that people would be better prepared to control those fires. In this paper, we establish a probability model and a numerical procedure for estimating probability of occurrences of large forest fires. A simulated numerical experiment is also presented to illustrate the application of the probability model and the numerical procedure.

Index Terms: large forest fires; probability model; numerical procedure; numerical simulation; parameter estimation

1. Introduction

Forest fires occur at various spots and various times all around the world. One of major efforts in forest fire management is to estimate the probability of occurrence of fires so that people would be better prepared to control those fires. The purpose of this paper is to establish a probability model and a numerical procedure for estimating probability of occurrences of large forest fires.

In order to precisely define “large” or “small” forest fires, we may use numerical ranks to describe the magnitudes of the fire, based on area burned, fire duration, and damage caused. For instance, we may use real numbers from the interval [1, 10] as ranks to describe the magnitudes of forest fires, 1 being the smallest and 10 the largest. Other ranking systems may also be used in practice.

A common approach to manage and control forest fires in areas of frequent occurrences is to observe the pattern or regularity of ranks in those occurrences. The purpose is to be able to predict occurrences of large forest fires, based on the probability estimation and likelihood computation, so that people can get ready to fight against the fire and reduce damages.

* Corresponding author.
E-mail address: yshi@bloomu.edu

One of the observed patterns is that in a given area where forest fires occur frequently, a large fire may occur after a series of smaller fires. Furthermore, after the occurrence of a large fire, there would be several smaller fires before the occurrence of another large one. Therefore, one assumption is that we may predict large fires by observing the occurrences of small fires. Similar assumptions are used in many studies regarding various natural disasters. See [1], [2], [3], [4] for example.

Based on this assumption, we will present a probability model for estimating the probability of the occurrence of a large forest fire in an arbitrary time period. A numerical procedure is described to implement the model, and a simulated numerical experiment is produced to illustrate the application of the model.

2. Probability Model and Numerical Procedure

Consider a given area where forest fires occur frequently. Let $x_1, x_2, \ldots x_n$ denote the magnitude rank of the n most recently recorded forest fires. Let J_i, $i = 1, 2, \ldots n - 1$, be the time interval between the i-th and $(i+1)$-th fire. Let the variable t stand for the time period of interest, which starts from the moment of occurrence of the first recorded fire and ends at an arbitrary future time moment. Finally, let X be the rank of the largest fire occurring in this time period t.

With these notations, we consider the probability $P(X < x)$ for some rank value x. Note that x can be any real number in [1, 10], and since we are concentrating on large fires, the value of x would be on the large side of the interval. Also note that $P(X < x)$ stands for the probability that no fire in the time period t would have a rank greater than or equal to x. Therefore, the probability that at least one large fire (with rank $\geq x$) would occur within the time period t equals 1 - $P(X < x)$.

Our probability model is based on a time-dependent exponential distribution involving rank as the random variable and time period as the parameter, as described in [5]. More precisely, we have the model

$$P(X < x) = [F(x)]^{t/\alpha} \tag{1}$$

with

$$F(x) = 1 - e^{-x/\theta} \quad . \tag{2}$$

Here α and θ are parameters to be estimated. In what follows, we will develop a numerical procedure to estimate these two parameters.

First, based on model (1), the likelihood function for the available data $\{x_i;\, i = 1,\ldots n\}$ and $\{J_i;\, i = 1,\ldots n\text{-}1\}$ is given by

$$L = (1/\alpha)^{n-1}[\prod_{i=1}^{n} f(x_i)][\prod_{i=1}^{n-1} J_i[F(x_{i+1})]^{(J_i/\alpha)-1} \tag{3}$$

where f stands for the derivative of F with respect to x. Therefore, the log-likelihood function is

$$\ln L = \sum_{i=1}^{n} \ln(f(x_i)) - (n-1)\ln\alpha + \sum_{i=1}^{n-1} \ln J_i + \sum_{i=1}^{n-1} ((J_i/\alpha) - 1)\ln(F(x_{i+1})) \tag{4}$$

From (4) and (2) we obtain the likelihood estimate equations

$$\frac{\partial \ln L}{\partial \alpha} = \frac{1-n}{\alpha} - \frac{1}{\alpha^2}\sum_{i=1}^{n-1} J_i \ln(1-e^{-x_{i+1}/\theta}) = 0$$
$$\frac{\partial \ln L}{\partial \theta} = \sum_{i=1}^{n}(\frac{-1}{\theta}+\frac{x_i}{\theta^2}) - \frac{1}{\theta^2}\sum_{i=1}^{n-1}(\frac{J_i}{\alpha}-1)\frac{x_{i+1}}{e^{x_{i+1}/\theta}-1} = 0 \tag{5}$$

The likelihood estimate (5) implies that the maximum likelihood estimate $\hat{\alpha}$ and $\hat{\theta}$ satisfy

$$\hat{\alpha} = \frac{-1}{n-1}\sum_{i=1}^{n-1} J_i \ln(1-e^{-x_{i+1}/\hat{\theta}})$$
$$\hat{\theta} = \bar{x} - \frac{1}{n}\sum_{i=1}^{n-1}(\frac{J_i}{\hat{\alpha}}-1)\frac{x_{i+1}}{e^{x_{i+1}/\hat{\theta}}-1} \tag{6}$$

here $\bar{x}$ stands for the mean value of $\{x_1, x_2, \dots x_n\}$.

Before presenting an iterative procedure to obtain the values of $(\hat{\alpha},\hat{\theta})$ satisfying (6), let us discuss the asymptotic distribution of $(\hat{\alpha},\hat{\theta})$ when n is large. Let us use the notation $\bar{F}(x)$ and $\bar{f}(x)$ to denote the cumulative distribution function and the density function of X, respectively. That is,

$$\bar{F}(x) = P(X < x) = [1-e^{-x/\theta}]^{t/\alpha}$$

and

$$\bar{f}(x) = \frac{t}{\alpha\theta} e^{-x/\theta}[1-e^{-x/\theta}]^{(t/\alpha-1)}.$$

We also note that the moment generating function $M(s)$ of X takes the form

$$M(s) = \frac{t}{\alpha} B(1-\theta s, \frac{t}{\alpha}),$$

where $B(.,.)$ is a Beta function.

In order to specify the asymptotic distribution of $(\hat{\alpha},\hat{\theta})$, let

$$R = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}$$

with

$$r_{11} = -E[\frac{\partial^2}{\partial \alpha^2} \ln \bar{f}],$$

$$r_{12} = r_{21} = -E[\frac{\partial^2}{\partial \alpha \partial \theta} \ln \bar{f}],$$

and

$$r_{22} = -E[\frac{\partial^2}{\partial \theta^2} \ln \bar{f}].$$

It is shown in [5] that

$$r_{11} = \frac{1}{\alpha^2},$$

$$r_{12} = r_{21} = \frac{1}{\alpha\theta} - \frac{t}{\alpha^2\theta}\sum_{k=1}^{\infty}\frac{1}{k(k+t/\alpha-1)},$$

$$r_{22} = \frac{1}{\theta^2} - \frac{2t^2}{\alpha^2\theta^2}\sum_{k=1}^{\infty}\frac{1}{k(k+t/\alpha)(k+t/\alpha-1)} + \frac{t}{\alpha\ \theta^2}\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{kj(k+j+t/\alpha-2)}.$$

It is also proved in [5] that as long as $t/\alpha > 0$, the series in r_{12} and r_{22} both converge. Now let

$$V = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

Denote the asymptotic covariance matrix of $(\hat{\alpha}, \hat{\theta})$, then $V = \frac{1}{n}R^{-1}$. Hence we have

$$\sigma_{11} = \frac{r_{22}}{n(r_{11}r_{22} - r_{12}^2)}$$

$$\sigma_{12} = \sigma_{21} = \frac{-r_{12}}{n(r_{11}r_{22} - r_{12}^2)}$$

and

$$\sigma_{22} = \frac{r_{11}}{n(r_{11}r_{22} - r_{12}^2)}.$$

The marginal distribution of $\hat{\alpha}$ and $\hat{\theta}$ are $N(\alpha, \sqrt{\sigma_{11}})$ and $N(\theta, \sqrt{\sigma_{22}})$, respectively.

Let us now present an iterative numerical procedure to obtain the values of $(\hat{\alpha}, \hat{\theta})$ satisfying (6). For convenience, let us use the notations

$$R_1(\theta) = \frac{-1}{n-1} \sum_{i=1}^{n-1} J_i \ln(1 - e^{-x_{i+1}/\theta})$$

and

$$R_2(\alpha, \theta) = \bar{x} - \frac{1}{n} \sum_{i=1}^{n-1} (\frac{J_i}{\alpha} - 1) \frac{x_{i+1}}{e^{x_{i+1}/\theta} - 1}.$$

With such notations, (6) becomes

$$\begin{aligned} \hat{\alpha} &= R_1(\hat{\theta}) \\ \hat{\theta} &= R_2(\hat{\alpha}, \hat{\theta}) \end{aligned} \tag{7}$$

Many numerical procedures can be developed to solve (7) (see [6] and [7], for example). A most straightforward numerical procedure for estimating the values of $\hat{\alpha}$ and $\hat{\theta}$ can be developed as the following Algorithm 1.

<u>Algorithm 1</u>

1. Select a user-given error tolerance ε (a small positive number), select an initial value θ_0, and let $\alpha_0 = R_1(\theta_0)$.
2. For $k = 0, 1, 2, \dots$, if the termination criterion

$$|\alpha_k - R_1(\theta_k)| + |\theta_k - R_2(\alpha_k, \theta_k)| \le \varepsilon \tag{8}$$

   is satisfied, then stop the procedure and take $\hat{\alpha} = \alpha_k$ and $\hat{\theta} = \theta_k$; otherwise let

   $\theta_{k+1} = R_2(\alpha_k, \theta_k)$ and $\alpha_{k+1} = R_1(\theta_{k+1})$,

   and continue the procedure.

One of the advantages of Algorithm 1 is that it automatically guarantees that $\alpha_k - R_1(\theta_k) = 0$ for all k. Hence the termination criterion (8) is equivalent to a simpler one $|\theta_k - R_2(\alpha_k, \theta_k)| \le \varepsilon$.

3. A Simulated Numerical Experiment

In this section we present a simulated numerical experiment to illustrate the application of the probability model (1)-(2), and the numerical procedure Algorithm 1. Suppose in a given area 10 forest fires have occurred in the past 20 years. Let us label the year when the first fire occurred as year 0. Then the year when the 10^{th} fire occurred will be labeled as year 20. Suppose the ranks $x_1, x_2, \ldots x_{10}$ and the time interval between fires $J_1, J_2, \ldots J_9$ are as given in Table I.

With this simulated data, the value of $\bar{x}$ = 3.6. Let us execute Algorithm 1 with $\varepsilon = 10^{-8}$, $\theta_0 = \bar{x} = 3.6$, and $\alpha_0 = R_1(\theta_0)$ =1.16746. After ten iterations of Algorithm 1, the termination criterion (8) is satisfied with $|\theta_{10} - R_2(\alpha_{10}, \theta_{10})| = 4.22\times 10^{-9} < 10^{-8}$, and we have obtained the maximum likelihood estimate $\hat{\alpha}$ = 0.19016 and $\hat{\theta}$ = 1.21485. Therefore, the probability model (1)-(2) based on this data is

$$P(X < x) = [1 - e^{-x/1.21485}]^{t/0.19016} \tag{9}$$

With probability model (9), we may easily estimate the probability of occurrences of large forest fires in upcoming years. For example, what is the probability that there is at least one fire of rank greater than or equal to 7 in next five years? In this case, $x = 7$ and $t = 20 + 5 = 25$. Hence the corresponding probability value equals

$$1 - P(X < 7) = 1 - [1 - e^{-7/1.21485}]^{25/0.19016} = 0.339054$$

That is, there is about 34% of chance that at least one large forest fire (of rank ≥ 7) will occur in this area within next five years.

Note that for a fixed time period t, the value of $1 - P(X < x)$ will decrease if the value of x increases. Also note that for a fixed rank value x, the value of $1 - P(X < x)$ will increase if the value of t increases. Both are consistent with the common sense expectations. In the following Table II we list a few values of $1 - P(X < x)$ with various x and t values.

4. Concluding Comments

In this paper, we have presented a probability model for estimating the probability of the occurrence of a large forest fire in an arbitrary time period. A numerical procedure is also described to implement the model. We like to point out that the numerical results presented in the previous section are based on a simulated experimental data. It is produced to illustrate the application of the model. Also note that there could be various probability models in addition to exponential distribution model that may be applied to estimate the probability of large forest fires. Furthermore, in real life applications, more and more data will become available when time goes by, and the model may be updated or improved correspondingly.

References

[1] R. Burridge and L. Knopoff, "Model and theoretical seismicity", *Bull. Seismol Soc. America*, 57, pp 341-371, 1967.

[2] W. Conover, *Practical Nonparametric Statistics*, 2nd Ed. Wiley, New York, 1980.
[3] A. Hawkes, "Point spectra of some mutually exciting point processes", *J. R. Statistics Soc. B*, 33, pp 438-443, 1971.
[4] D. Kendall, "Stochastic processes and population growth", *J.R. Statistics Soc. B*, 11, pp 230-264, 1949.
[5] R. Noubary and Y. Shi, "Seismic hazard calculation based on sizes and inter-arrival times", *Journal of Applied Statistical Science*, 10(4), pp 343-356, 2001.
[6] Y. Shi, "Some computational aspects in solving nonlinear systems", *International Journal of Applied Mathematics*, 1(5), pp 511-519,1999.
[7] Y. Shi, "A globalization procedure for solving nonlinear systems of Equations", *Numerical Algorithms*, 12, pp 272-286, 1996.

Table 1. Simulated data for the 10 fires in the past 20 years

Year of fire occurrence	Rank of the fire (x_i)	Time interval between fires (J_i)
0	4	2
2	5	1
3	3.4	2
5	3	4
9	2	3
12	4	2
14	5.1	2
16	3	1
17	3	3
20	3.5	

Table 2. Value of $1 - P(X < x)$ with various values of x and t

	$t = 25$	$t = 30$	$t = 35$	$t = 40$	$t = 45$
$x = 7$	0.339054	0.391586	0.439943	0.484456	0.525432
$x = 8$	0.166100	0.195850	0.224540	0.252205	0.278884
$x = 9$	0.076624	0.091230	0.105603	0.119750	0.113673
$x = 10$	0.034390	0.041124	0.047811	0.054452	0.061047