S11111
11111
11111
----4
CROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A
PEAKS FROM RANDOM DATA

by

Sheldon M. Ross

ORC 86-18

October 1986

This research was supported by the Air Force Office of Scientific
Research (AFOSR) USAF, under grant AFOSR-86-0153 with the University of
California. Reproduction in whole or in part is permitted for any
purpose of the United States Government.

1Department of Industrial Engineering & Operations Research, University
of California, Berkeley, California 94720.
**Title:** Peaks from Random Data

**Abstract:**

(Please see Abstract)
PEAKS FROM RANDOM DATA

Sheldon M. Ross+

Department of Industrial Engineering and Operations Research
University of California, Berkeley

1. INTRODUCTION

In an influential and controversial paper, Raup and Sepkoski [2] defined an event of mass extinction to have occurred in any time period (of roughly 6.4 million years) for which the data value for that time period (equal to the proportion of the families existing at the beginning of that period that went extinct during the period) exceeded that of its immediate neighbors. In this note we analyze the occurrence of such events when the data are randomly generated from a continuous distribution.

+This research was supported by the Air Force Office of Scientific Research (AFOSR) USAF, under grant AFOSR-86-0153 with the University of California.
2. THE PROCESS OF EVENTS UNDER RANDOM DATA

Let \( X_1, X_2, \ldots \) be a sequence of independent and identically distributed continuous random variables, and define an event to occur at time period \( j \) if \( X_{j-1} < X_j > X_{j+1} \), \( j \geq 2 \). Let

\[ I_j = \begin{cases} 1 & \text{if an event occurs at time } j \\ 0 & \text{otherwise} \end{cases} \]

and set

\[ N(n) = \sum_{j=2}^{n} I_j. \]

In words, \( N(n) \) is the number of events by time \( n \).

**Proposition 1:** With probability 1,

\[ \lim_{n \to \infty} \frac{N(n)}{n} = \frac{1}{3} \]

**Proof:** This follows from the ergodic theorem since \( \{I_j, j \geq 2\} \) is a stationary ergodic sequence and \( E[I_j] = 1/3 \). A simpler proof is obtained by noting that \( \{I_2, I_5, I_8, \ldots\}, \{I_3, I_6, I_9, \ldots\}, \text{ and } \{I_4, I_7, I_{10}, \ldots\} \) are each iid sequences whose average value converges, by the strong law of large numbers, to 1/3. QED

The mean and variance of \( N(n) \) are easily obtained:

\[ E[N(n)] = \sum_{j=2}^{n} E[I_j] = (n-1)/3 \]

\[ var[N(n)] \]
\[ \text{Var}[N(n)] = \sum_{j=2}^{n} \text{Var}(I_j) + 2 \sum_{i<j} \text{Cov}(I_i, I_j) \]

As

\[ \text{Var}(I_j) = 2/9 \]
\[ \text{Cov}(I_1, I_{i+1}) = -E[I_1]E[I_{i+1}] = -1/9 \]
\[ \text{Cov}(I_1, I_{i+2}) = P\{X_{i-1} < X_1 < X_{i+1} < X_{i+2} < X_{i+3}\} - 1/9 \]
\[ = \int \int \int \int dx_1 \ dx_2 \ dx_3 \ dx_4 \ dx_5 \ -1/9 = 1/45 \]
\[ \text{Cov}(I_1, I_{i+j}) = 0, \ j \geq 3 \]

we see from the above that, for \( n \geq 3 \),

\[ \text{Var}[N(n)] = 2(n-1)/9 -2(n-2)/9 +2(n-3)/45 \]
\[ = (2n + 4)/45 \]

It follows from the above and the central limit theorem for stationary ergodic processes that

\[ \frac{N(n) - (n-1)/3}{\sqrt{(2n+4)/45}} \quad \longrightarrow \quad N(0,1) \]
Terminology If $X_{j-1} < X_j > X_{j+1}$ call $X_{j+1}$ the end-of-peak value.

3. INTEREVENT TIMES

Let $T_i$, $i \geq 1$, denote the time between the $i^{th}$ and $(i+1)^{st}$ event - for instance, if $I_2 = I_4 = 1$ then $T_1 = 2$. Even though the $T_i$ are neither independent nor identically distributed we will show that the proportion of them that equal $j$ will, with probability 1, converge to a constant value which we shall call $p_j$.

To determine the values $p_j$, call an event a $j$-event if the event preceding it occurred exactly $j$ time periods earlier, and note that by the same argument as used in Proposition 1, the rate at which $j$-events occur, call it $r_j$, will equal the limiting probability of a $j$-event at time period $n$. Let

$$\tilde{\lambda}(k) = \sum_{j>k} r_j$$

That is, $\tilde{\lambda}(k)$ denotes the rate at which interevent times of length greater than $k$ occur. Now, an interevent time will exceed $k$ if, for some $r=0,1,...,k$, the next $r$ values after an end of peak are each less, and the next $k-r$ values are each greater, than their preceding values. Therefore, for $k \geq 2$,

$$\tilde{\lambda}(k) = P(X_2 < X_1 < X_0 < X_1 < ... < X_k) + P(X_2 < X_1 < X_0 > X_1 > ... > X_k)$$

$$+ \sum_{r=1}^{k-1} P(X_2 < X_1 < X_0 > X_1 > ... > X_r < X_{r+1} < ... < X_k)$$

$$= \int_0^1 \frac{(1-x)^k}{2} \frac{(1-x^2)}{k!} \, dx + \int_0^1 \frac{x^{k+2}}{(k+1)!} \, dx + \sum_{r=1}^{k-1} \int_0^1 \frac{(x-y)^r}{x} \frac{(1-y)^{k-r}}{(k-r)!} \, dy \, dx$$

$$= \int_0^1 \frac{(1-x)^k}{2} \frac{(1-x^2)}{k!} \, dx + \int_0^1 \frac{x^{k+2}}{(k+1)!} \, dx + \sum_{r=1}^{k-1} \int_0^1 \frac{(x-y)^r}{x} \frac{(1-y)^{k-r}}{(k-r)!} \, dy \, dx$$
where the last equality obtains from conditioning the first probability on the value of $X_0$, the second on $X_{-1}$, and the remaining on both $X_{-1}$ and $X_r$.

As the rate at which events occur is $1/3$ it follows that the proportion of events that are $j$-events is $3r(j)$. That is,

$$p_j = 3r(j).$$

A direct integration of the above yields, upon using the identity

$$\sum_{j=2}^{k} r(j) = 1 - \bar{R}(k),$$

that

$$p_2 = \frac{2}{5}$$
$$p_3 = \frac{1}{3}$$
$$p_4 = \frac{6}{35}$$
$$p_5 = \frac{1}{15}$$
$$p_6 = .02116401$$

and so on. The quantities $p_j$ also represent the limiting distribution of the $T_1$. A simple, though tedious, summing of the values $\bar{R}(k)$ yields the (already known) result that the mean of this limiting distribution is 3.

**Remark:** The above could also have been obtained by consideration of the Markov Chain of successive end-of-peak values. The limiting value of $T_1$ could then be computed by conditioning on $X$, the limiting end-of-peak value (whose density is given by $f(x)=3(1-x^2)/2, \ 0<x<1$), and then computing the probability density of the time between peaks when the initial one has an end-of-peak value $X$. 
References


END
12-86
DTIC