PERIODICA POLYTECHNICA SER. CIVIL ENG. VOL. 36, NO. 4, PP. 365–368 (1992)

GEYSERS AND TESTS[1]

Gábor J. SZÉKELY and Arjun K. GUPTA*

Department of Mathematics,
Faculty of Civil Engineering
Technical University, H–1521 Budapest
*Department of Mathematics and Statistics, BGSU, USA

Received: December 1, 1992

'Dieu! La regarderai-je?'
(E. R.: CdB Quatrième acte, scène V)

Abstract

A new test of Poissonity based on a characteristic property of Poisson distributions is proposed.

Keywords: Poisson distribution, exponential distribution, characterization of distributions.

Prelude

In 1875 a young professor of our University, Gyula König (also Rector between 1891 and 1894) gave an interesting lecture on the determination of the period ϕ of a periodic event if only very rough observations are available, say, the signs of $\cos(k\phi)$ $k = 0, 1, \ldots$. E.g. if we can only observe periodic eruptions of a geyser occurring at night $(-)$ or in the daytime $(+)$ then from this sequence of $+$ and $-$ signs we can reconstruct the value of $\phi (\mathrm{mod}\ \pi)$. If w_n denotes the changes of signs in this sequence of $+$ and $-$ of length n, then

$$\frac{\phi}{\pi} = \lim_{n\to\infty} \frac{w_n}{n}.$$

This result was published in KÖNIG (1876).

Variations

Variation 1

König's idea can be used to find a solution of the algebraic equation

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n = 0.$$

[1] Research supported by Hungarian National Foundation for Scientific Research, Grant № 1405

If we compute

$$\frac{1}{f(x)} = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \dots$$

and w_n denotes the changes of signs in the sequence $\alpha_0, \dots \alpha_n$, then one of the roots (zeros) of our equation is

$$x_0 = r(\cos\omega + i\sin\omega),$$

where

$$\frac{\omega}{\pi} = \lim_{n\to\infty} \frac{w_n}{n}$$

and

$$r = \lim_{n\to\infty} |\alpha_n|^{\frac{-1}{n}}$$

(in fact x_0 is a root with the smallest absolute value r).

Variation 2

If the period ϕ is random and the random periods $\phi_1, \phi_2, \dots$ between consecutive eruptions of a geyser are independent, identically distributed random variables with unknown distribution function F, then we might want to determine F from rough observations of

$$S_n = \phi_1 + \phi_2 + \dots + \phi_n \qquad n = 1, 2, \dots.$$

(S_n can be considered a random arithmetic progression where $\phi_1, \phi_2, \dots$ are not necessarily identical, only identically distributed. The case $\phi_1 = \phi_2 = \dots = \phi$ would correspond to König's sequence $S_k = k\phi$.) Suppose that a rough observation of S_n is the integer part of S_n denoted by $[S_n]$ (if the unit is one day, then we can only observe the dates of eruptions).

By the law of large numbers the expectation of F is

$$\lim_{n\to\infty} \frac{[S_n]}{n} = \lim_{n\to\infty} \frac{S_n}{n}$$

with probability one.

A much deeper result of BÁRTFAI (1966) and ERDÖS–RÉNYI (1970) shows that not only the expectation of F but F itself can also be reconstructed from the sequence $[S_n]$ (see also CSÖRGŐ–RÉVÉSZ (1981)).

Rondo

Try to test the hypothesis that the random number of eruptions (of a geyser) per day is a Poisson random variable. In other words if $\nu_1, \nu_2, \ldots, \nu_n$ denote the random number of eruptions on the first, second, ...,n-th day, then we want to test the hypothesis that $\nu_1, \nu_2, \ldots, \nu_n$ are (independent) Poisson random variables,

$$P(\nu_i = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad i = 1, 2, \ldots, n \ ; \ k = 0, 1, 2, \ldots .$$

(λ is the unknown expected number of eruptions per day).

Testing Poissonity in a powerful way is not an easy problem. Neither is easy to get rid of the unknown parameter λ (here λ is neither location nor scale). In fact, one can prove (see PEÑA-ROHATGI–SZÉKELY (1992)) there is no (nonconstant) function of $\nu_1, \ldots, \nu_n$ whose distribution function does not depend on λ.

Approach the problem in the following way. If $\phi_1, \phi_2, \ldots$ are independent, exponentially distributed random variables with density function $\lambda e^{-\lambda x}$ for $x \geq 0$ (and 0 otherwise) and $S_n = \sum_{i=1}^{n} \phi_i$, then in the sequence of increasing integer parts $[S_1]$,$[S_2]$,$[S_3]$,... the number of 0's, 1's, 2's, etc. is a sequence $\nu_1, \nu_2, \ldots$ of independent Poisson random variables. By Variation 2 above, this sequence uniquely determines the distribution F of ϕ's, therefore F must be exponential, ν is Poisson if and only if ϕ is exponential. Thus, testing for Poissonity of ν is equivalent to testing for exponentiality of ϕ. In the density of ϕ, however, the parameter λ is a scale parameter and, thus, the problem of testing is easier and well known (see e.g. SHAPIRO and WILK (1962), LILLIEFORS (1969), BARTHOLOMEW (1957), MORAN (1951), EPSTEIN (1960), STÖRMER (1962)).

There is only one more missing link: how can we reconstruct the sequence $\phi_1, \phi_2, \ldots$ from $\nu_1, \nu_2, \ldots$. Take independent, uniformly distributed samples of size ν_i on the interval $[i-1, i]$ $i = 1, 2, \ldots, n$. If the (increasingly) ordered joint sample is $\xi_1, \xi_2, \ldots, \xi_m$ where $m = \sum_{i=1}^{n} \nu_i$, then $\phi_i^* = \xi_{i+1} - \xi_i$, $i = 1, 2, \ldots$ is a sequence of independent, exponentially distributed random variables with parameter λ, i.e. $\{\phi_i\}$ and $\{\phi_i^*\}$ are identically distributed therefore $\{\phi_i^*\}$ can play the role of $\{\phi_i\}$ in the computations above.

Acknowledgements

The authors are thankful to G. Tusnádi for an interesting discussion on this topic.

References

BARTHOLOMEW, D. J. (1957): Testing for Departure from the Exponential Distribution. *Biometrika*, Vol. 44, pp. 253–256.

BÁRTFAI, P. (1966): Die Bestimmung der zu einem wiederkehrenden Prozess gehörenden Verteilungsfunktion aus den mit Fehlern behafteten Daten einer einzigen Realisation. *Studia Sci. Math. Hung.* Vol. 1, pp. 161–168.

CSÖRGŐ, M. – RÉVÉSZ, P. (1981): Strong Approximations in Probability and Statistics, Akadémiai Kiadó, Budapest.

EPSTEIN, B. (1960): Tests for the Validity of the Assumption that the Underlying Distribution of Life is Exponential. *Technometrics*, Vol. 2, pp. 83–101.

ERDŐS, P. – RÉNYI, A. (1970): On a New Law of Large Numbers, *Jour. Analyse Mathématique* Vol. 22, pp. 103–111.

LILLIEFORS, H. W. (1969): On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown. *J. Amer. Statist. Assoc.*, Vol. 64, pp. 387–389.

KŐNIG, Gy. (1876): On a General Solution ..., *Műegyetemi Lapok*, Vol. 1, pp. 20–28, 112–116.

MORAN, P. A. P. (1951): The Random Division of an Interval II. *J. Roy. Statist. Soc. Ser. B*, Vol. 13, pp. 147–150.

PEÑA, E. A. – ROHATGI, V. K. – SZÉKELY, G. J.(1992): On the Non-existence of Ancillary Statistics, *Probability and Statistics Letters* Vol. 15, pp. 357–360.

SHAPIRO, S. S. – WILK, M. B. (1972): An Analysis of Variance Test for the Exponential Distribution, *Technometrics*, Vol. 14, pp. 355–370.

STÖRMER, H. (1962): On a Test for the Exponential Distribution, *Metrika*, Vol. 5, pp. 128–137.