Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 126 (2016) 3790–3807

stochastic processes and their applications

www.elsevier.com/locate/spa

The Hurst phenomenon and the rescaled range statistic

David M. Mason

Department of Applied Economics and Statistics, University of Delaware, 213 Townsend Hall, Newark, DE 19716, USA

Received 17 December 2014; accepted 3 April 2015
Available online 29 April 2016

Abstract

In his 1951 study of Nile River data, H.E. Hurst introduced the rescaled range statistic-the R/S statistic. He argued via a small simulation study that if $X_i, i = 1, \ldots, n$, are i.i.d. normal then the R/S statistic should grow in the order of $\sqrt{n}$. However, Hurst found that for the Nile River data, the R/S statistic increased not in the order of $\sqrt{n}$, but in the order n^H, where H ranged between 0.75 and 0.80. He discovered that the effect also appeared in other sets of data. This is now called the *Hurst phenomenon.* We shall establish some unexpected universal asymptotic properties of the R/S statistic, which show conclusively that the Hurst phenomenon can never appear for i.i.d. data.

MSC: primary G0G10; 62M10; secondary 60F10

Keywords: Rescaled range statistic; Hurst phenomenon; Subgaussian

1. Introduction: the Hurst phenomenon

In 1951 H.E. Hurst [13] published the results of his investigations of water outflow from the great lakes of the Nile basin. Hurst wanted to determine the reservoir capacity that would be needed to develop the irrigation along the Nile to its fullest extent. The problem that he basically solved was as follows:

In the years of high runoff the Nile water is not fully utilized, but in the years of low runoff there is a shortage of water. Hence for efficient use of the water resource an optimum constant

E-mail address: davidm@udel.edu.

http://dx.doi.org/10.1016/j.spa.2016.04.008
0304-4149/

yearly outflow is required. What is that optimum constant yearly outflow and what is the storage capacity necessary to maintain it?

Here was his solution: Suppose X_i is the total yearly outflow of water from the source of the Nile in year i, $i = 1, \dots, n$. Set $\mu_n = \frac{1}{n}\sum_{i=1}^n X_i$.

Question: What is the reservoir capacity required to maintain a constant yearly outflow μ_n over the years $i = 1, \dots, n$? Let

$$S_i^* = \sum_{j=1}^{i} \left(X_j - \mu_n\right), \quad \text{for } i = 1, \dots, n,$$

$$M_n^* = \max\left(0, S_1^*, \dots, S_n^*\right) \quad \text{and} \quad m_n^* = \min\left(0, S_1^*, \dots, S_n^*\right).$$

The reservoir storage required is the *adjusted range*

$$R_n^* = M_n^* - m_n^*. \tag{1}$$

We need storage capacity M_n^* to store to a maximum overflow M_n^* over the n year mean μ_n and an additional storage m_n^* to cover the periods when there is a deficit in the outflow from the n year mean μ_n. We easily see that μ_n is the maximum constant yearly outflow that can be maintained over the n year period.

Unfortunately in planning a reservoir system we do not know *a priori* what the yearly outflows are going to be for the period in which the system will be in use. The usual assumption before Hurst was that $X_i = \mu + e_i$, where the e_i are some kind of white noise, i.e. $e_1, \dots, e_n$ are i.i.d. with $E(e_i) = 0$ and $0 < Var(e_i) = \sigma^2 < \infty$. In fact there was empirical evidence for this assumption. When Hurst depicted the measurements of the Nile River outflow in a frequency histogram, for instance, the maximum annual gage readings that were recorded at the Roda gage near Cairo for the years between 641 C.E. and 1946 C.E., he obtained a convincingly normal shaped curve around the sample mean.

Hurst made the first steps in the analysis of the random variable R_n^* under the assumption that $e_1, \dots, e_n$ are i.i.d. $N(0, 1)$. Through simulation experiments based on tossing ten sixpence coins 1000 times, cutting cards from a *probability deck* 1000 times and observing the serial numbers of bonds he found that ER_n^* grows approximately like $1.20\sqrt{n}$. (For more details see his Table 6.) Incidentally this agrees well with the following exact result of Feller [8]: Let $S^*(t) = S(t) - tS(T)/T$ for $0 \le t \le T$, $T > 0$, where $S(t)$ is a standard Brownian motion on $[0, T]$. Set

$$\mathbb{R}_T^* = \max\left(0, S^*(t), 0 \le t \le T\right) - \min\left(0, S^*(t), 0 \le t \le T\right).$$

Then $E\mathbb{R}_T^* = \sqrt{T\pi/2} \approx 1.2533\sqrt{T}$. Noting that R_n^*, defined in terms of $X_1, \dots, X_n$ i.i.d. $N\left(\mu, \sigma^2\right)$, is equal in distribution to

$$\sigma \max\left(0, S^*(i), 0 \le i \le n\right) - \sigma \min\left(0, S^*(i), 0 \le i \le n\right)$$

we have $ER_n^* \approx \sigma\sqrt{n\pi/2} \approx 1.2533\sigma\sqrt{n}$ for large n. The exact value was later shown by Solari and Anis [29] to be

$$ER_n^* = \sigma\sqrt{\frac{n}{2\pi}} \sum_{i=1}^{n-1} (i(n-i))^{-1/2} = \sigma\sqrt{n}\left(\sqrt{\pi/2} + o(1)\right).$$