Record statistics

I. INTRODUCTION

Although extreme value statistics provides relevant information about building dikes or hedging our financial bets, people are more fascinated by another type of extreme statistics, namely by records achived in sports, in lifetime, or in the strength of hurricanes.

The problem is easily formulated. Let $y_0, y_1, ..., y_{n-1}$ be an ordered set of random variables (e.g. time series of a stochastic process; e.g. the daily measured temperatures at a given location). Then y_n is an upper record if

$$y_n > \max\{y_0, y_1, ..., y_{n-1}\} . \tag{1}$$

Naturally, we can also define a lower record a $y_n < \min\{y_0, y_1, ..., y_{n-1}\}$. In general, y_0 is considered as the first record.

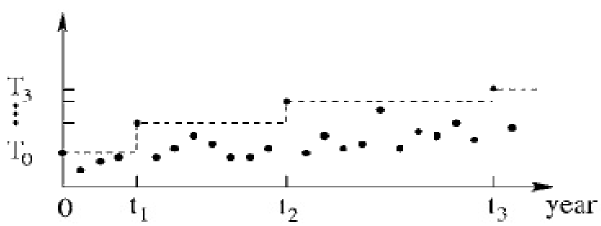

FIG. 1: xxx

One of the most investigated quantities in record statistics is the the probability of record P_n at time n. The probability P_n is the rate at which a record is produced at the given event, formally it is defined as

$$P_n = \mathrm{Prob}[y_n > \max\{y_0, y_1, ..., y_{n-1}\}] . \tag{2}$$

The other important quantity in record statistics is the record number R_n which is the number of records that occured in the time series up to time n. The average number of record number can be simply expressed through record probability

$$\langle R_n \rangle = \sum_{i=0}^{n} P_n . \tag{3}$$

II. RECORD STATISTICS FOR I.I.D. VARIABLES

There are a number of fundamental results related to record statistics for i.i.d. variables which are independent of the underlying distribution $p(y)$. Both the probability of a record P_n at n and the average record number $\langle R_n \rangle$ can be easily computed by the so called stick shuffling argument.

For i.i.d. variables the record among $n+1$ y_i values can be at any i. Thus the probability that the record is at $i = n$ is (let's remember that we start with $i = 0$)

$$P_n = \frac{1}{n+1} . \tag{4}$$

It follows then that the average record number is given by

$$\langle R_n \rangle = \sum_{i=0}^{n} P_n = \sum_{i=0}^{n} \frac{1}{i+1} \to \ln n + \gamma_E \tag{5}$$

for large n, and this is independent of the underlying parent distribution. The only assumption is the i.i.d. nature of the random variables.

Of course, there are correlations and, furthermore, usually, the time series is not stationary. E.g. one expects that if there is global warming then there is trend among the variables. These are already more complicated problems which are in the centre of attention of present day work on record statistics. Interested parties may consult the following (recent, quite well written and understandable) papers: [1–5].

[1] S. Redner and M. R. Petersen, *Role of global warming on the statistics of record-breaking temperatures*, Phys. Rev. E **74**, 061114 (2008).

[2] G. Wergen, *Records in stochastic processes - Theory and applications* arXiv:1211.6005.

[3] G. Wergen, M. Bogner, and J. Krug, *Record statistics for biased random walks, with an application to financial data,* arXiv:1103.0893.

[4] J. Krug and K. Jain, *Breaking records in the evolutionary race*, Physica A **358**, 1-9 (2005).

[5] W. I. Newman, B. D. Malamud, *Statistical properties of record-breaking temperatures*, Phys. Rev. E **82**, 099111 (2010).