TASK

Review a research article on applications of extreme value theory in sports.

SOLUTION

REVIEW

This brief document reviews paper by Einmahl, J.H.J. & Smeets, S.G.W.R. (2009), “ULTIMATE 100M WORLD RECORDS THROUGH EXTREME-VALUE THEORY”. The paper presents an interesting application of extreme value theory to sports. Overall, we believe the paper is rigorous and contributes to our understanding of the distribution of men’s and women’s world records in running 100 meters. The authors implement an objective and clean approach to data selection and apply a semi-formal statistical procedure to estimate the right tail of running speeds. The major steps of the estimation procedure are known to converge to the truth with the sample size increasing under the assumptions which are tested in the paper. The estimation is performed completely separately for men and women… For the sake of brevity we will not describe the analysis here but delegate the reader to the original source.

Still, there are a few directions in which the analysis or exposition of the results can be improved. First, information is money and ignoring information has been proven to be suboptimal by the statistical theory. Yet, the authors calculate the highest personal best time of each seasonal time list and take the lowest of these times as an upper bound on the combined list. All times higher than this upper bound are removed. The procedure is described on page 3 of the paper. It is questionable and must be substituted with a statistically efficient method. It is quite possible that the loss in the estimation accuracy is not substantial with the procedure at work. However, even if that is the case, a negligible loss in the accuracy has to be proven or illustrated.

Second, the procedure for estimating the right tail of running speeds has two subjective steps. In the first subjective step the authors identify a region of values of k for which the initial estimate of extreme value index ʹY is “stable”. Here k is the number of order statistics used in the estimation procedure. In the second subjective step the authors identify a region of values of k for which the estimate of the endpoint x* of the speed distribution is “stable”. The stability regions are identified purely visually and each time come out to be some nice round values, e.g. 50 – 80. Under such circumstances it is possible for authors to cheat (unlikely) or make small mistakes (more likely). A more formal, algorithmic procedure must be put in place here.

The third flaw of the analysis, which we find to be most important, is the following. The whole estimation procedure (with the new, innovative steps) has never been tested for convergence or accuracy. The innovative tricks that the authors apply in several different steps make perfect sense and are in line with tricks utilized in the machine learning field. We believe that, most likely, the resulting estimation procedure is consistent and relatively accurate for medium-to-large samples. It would not be too difficult to demonstrate this through a Monte Carlo simulation study. Many large and small data sets could be simulated for men (women) from

1) the empirical distribution (the sample),
2) several standard distributions which have been demonstrated to capture running speeds realistically.

Then the estimators of΄Y and x* could be calculated on each data set and their variability could be studied

1) over the sample size,
2) over data sets of the same sample size,
3) over different distributions of running speeds.

Sadly, the authors have not produced such a study. Arguably, this is the flaw of their work which is most easily corrected in follow-up paper(s).

A couple of minor issues with the analysis and exposition are the following.

- It is not clear why the authors stop the data set on June 2008. They claim themselves that August 2008 is important as the time of a world record but that measurement has not found its way into the sample. Is that because the Swedish web-site does not report statistics beyond June 2008? Quite possible but it is hard to imagine that the most recent statistics by the time of paper development (2009) was not available through other sources. Whatever the reason was it has to be clearly stated.

- Comparison of the results in the paper with those in the previous studies is non-scientific. No formal measures of discrepancy are defined. So it is not clear whether the results are substantially different in some cases. For example, the authors write: “In Denny (2008) the 9.48 <seconds> for the men agrees very much with our 9.51 <seconds> and so does the 10.39 <seconds> (calculated excluding the 10.49 world record) for women with our 10.33 <seconds>.” Well, considering that a world record is a matter of milliseconds, it is not immediately clear that the reported differences are not big. On the relative scale they are noticeable.
