Journal of the Royal Statistical Society
SERIES A Statistics in Society

J. R. Statist. Soc. A (2011)
174, *Part* 4, *pp.* 1175–1177

Correspondence

Practical and efficient estimates of one's accuracy in darts

We thank Tibshirani *et al.* (2011) for their most interesting essay. In addition to its innovative use of a personalized heat map to show the optimal strategy for throwing darts, it provides an engaging example for teaching several statistical concepts and techniques, such as fast Fourier transforms, the EM algorithm, Monte Carlo integration, importance sampling and the Metropolis–Hastings algorithm. It is a delightful blend of the applied and the theoretical, the algebraic and the graphical.

It also continues the tradition of statisticians' fascination with the imagery of marksmen (Turner and Hanley, 2010). In her chapter on metaphor and reality of target practice, Klein (1997) wrote of 'men reasoning on the likes of target practice' and described how this imagery has pervaded the thinking and work of natural philosophers and statisticians. Klein showed a frequency curve, by Yule, for 1000 shots from an artillery gun in American target practice. Pearson used it in his 1894 lectures on evolution; he decomposed the frequency curve into two chance distributions centred slightly to the right and left of the target, gave reasons why this might occur and used it to illustrate the interplay between random variation and natural selection. He also used it in Pearson (1900) in one of the illustrations of his test of goodness of fit.

Since the optimal aiming spot in darts—and thus the heat map provided by the on-line applet—depends strongly on one's accuracy, much of Tibshirani *et al.* (2011) is devoted to the challenge of estimating the (co)variance parameter(s) that describes this accuracy. All the estimators rely on the data generated by throwing n darts, aiming each time at the centre of the board, i.e. the double bull's-eye, and recording the result for each throw.

Tibshirani *et al.* (2011) noted that they would lose considerable information by not measuring the actual *locations* where the darts land but considered this to be too time consuming and error prone. Instead, they chose the individual *scores* produced by the throws (the 44 possible scores are 0–22, 24–28, 30, 32–34, 36, 38–40, 42, 45, 48, 50, 51, 54, 57 and 60). Based on $n=100$ throws by authors 1 and 2, assuming the simplest variance model (equal, uncorrelated vertical and horizontal Gaussian errors), their standard deviations were estimated to be $\hat{\sigma}=64.6$ and $\hat{\sigma}=26.9$ respectively (the applet gives $\hat{\sigma}$ to two decimal places).

We write to provide a measure of the statistical precision of these estimates of accuracy (for example, we calculate that the 95% limits to accompany the reported point estimate 64.6 derived from 100 scores are approximately 56 and 75). More importantly, we show that more precise estimates of σ can often be achieved with the same number of throws (or the same precision with fewer throws) if we use a simpler yet more informative version of the result from each throw. Here we focus on the simplest variance model.

The low information content of the scores with respect to σ is because many of them arise from throws that land at very different distances from the centre. For example, a score of 18 can arise from a throw that lands in one of four regions: double 9 (least accurate), outer single 18 (accurate), triple 6 (more accurate) or inner single 18 (most accurate). This ambiguity and loss of information are avoided if we simply record instead which of the seven '*rings*' the throw lands in: 1, the double bulls-eye; 2, the single bull's-eye; the rings formed by the 3, single bulls-eye and inner triple, 4, inner and outer triple, 5, outer triple and inner double, and 6, inner and outer double wires respectively; and 7, beyond the outer double wire (i.e. the throw misses the board), i.e. we need only to divide the dartboard into seven rings according to their distance to the centre.

To quantify how much information is conserved if the *raw location* data are reduced to

(a) 'ring' data and
(b) 'score' data,

we can measure the relative efficiency of these two latter methods of data recording. Since the log-likelihood is more symmetric in $\log(\sigma)$ than in σ, each panel in Fig. 1 shows the log-likelihood, but on a *log*-scale for σ.

 0964–1998/11/1741175

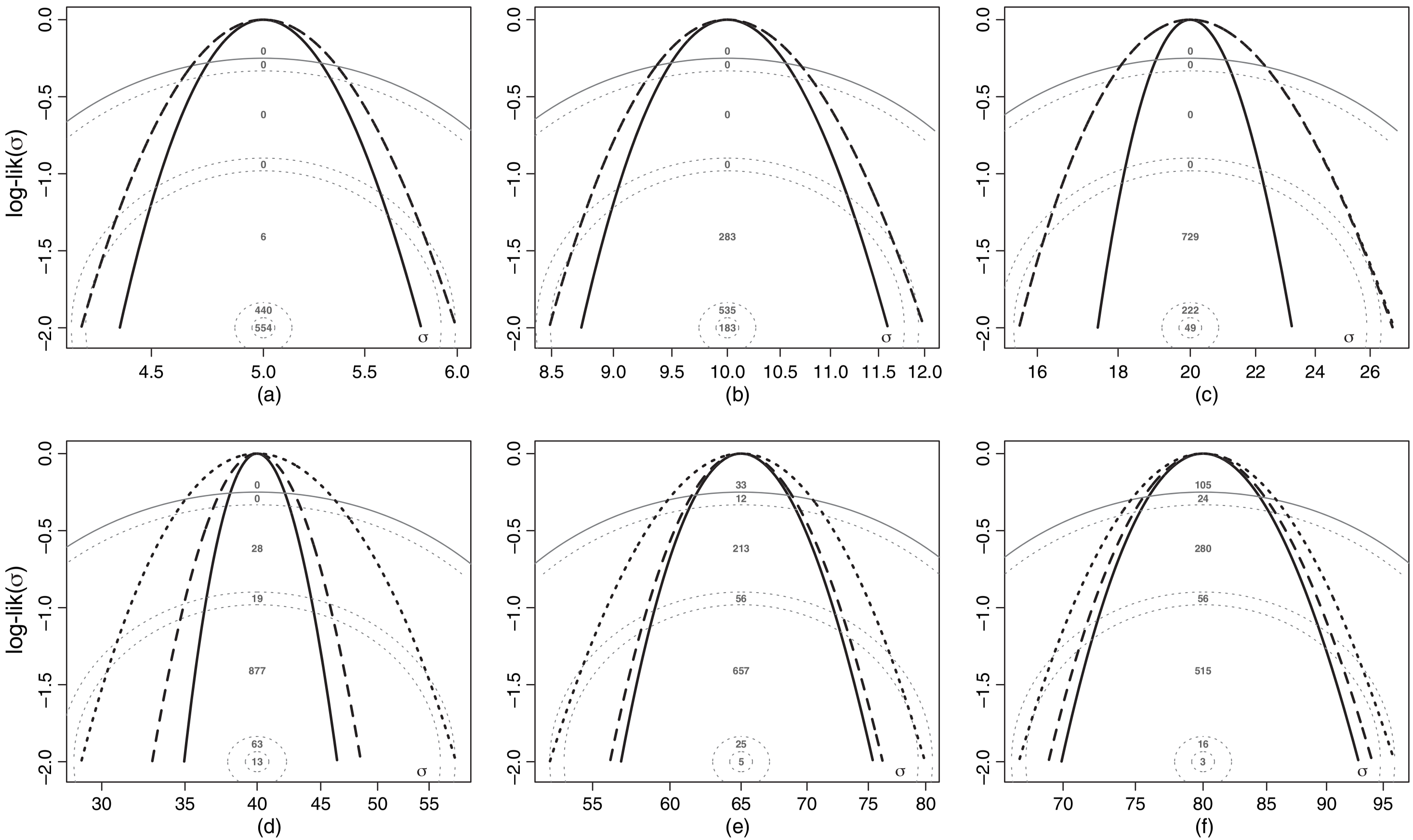

Fig. 1. Log-likelihoods, and amounts of information regarding $\log(\sigma)$, using (expected) results of $n = 50$ throws, if we record the actual locations (———, $I_{\log(\sigma)} = 4n$), or reduce them to the seven possible rings (– – –), or the 44 possible scores (· · · · · · ·) (log-likelihoods, relocated to equal 0 at $\hat{\sigma}_{\mathrm{MLE}}$, are plotted against σ, but with a log-scale for the horizontal axis, with corresponding amounts of information regarding $\log(\sigma)$; expected frequencies (scaled to sum to 1000) in the seven rings are shown in grey in the background; for low values of σ ((a)–(c)) the simpler ring data provide the same amount of information as the score data (the log-likelihoods overlap); for larger σ-values ((d)–(f)) they provide a greater amount of information): (a) locations 200, rings 128, scores 128; (b) locations 200, rings 134, scores 134; (c) locations 200, rings 54, scores 54; (d) locations 200, rings 108, scores 31; (e) locations 200, rings 171, scores 90; (f) locations 200, rings 169, scores 125

The log-likelihood function and the three amounts of (Fisher) information are based on a sample size $n = 50$, as suggested by the authors. The expected amount of information concerning $\log(\sigma)$ contained in the raw location values can be shown analytically to be $4n$, or 200 in our example. We calculated the corresponding information for the competitors by using the expected (multinomial) frequencies.

Fig.1 shows that the ring data are often much more (and never less) informative than the score data. This difference in information is greatest when the player is moderately accurate: as is seen in Fig. 1(d) and Fig. 1(e) we can obtain the same amount of information about $\log(\sigma)$ by using ring data on 26 ($= 50 \times 90/171$) throws or score data on 50 throws. This difference is least when the results from the two data recording systems overlap considerably, i.e. if most of the throws are in or close to one of the two bulls-eye regions (Figs 1(a)–1(c), where curves shown with dotted and broken lines are virtually indistinguishable), or if a large percentage of throws fall outside the board (Fig. 1(f)).

Fig.1 can be used to provide a confidence interval to accompany (for example) the reported $\hat{\sigma} = 64.6$ based on Tibshirani's 100 scores (see Fig. 1(e)). If this estimate had been based on the detailed *locations* for $n = 100$ throws, $\mathrm{SE}\{\log(\hat{\sigma})\}$ would have been approximately $(1/4n)^{1/2} = 1/400^{1/2}$, the multiplicative margin of error for a 95% confidence interval would be approximately $\exp(1.96\mathrm{SE}) = 1.1$ and so the confidence interval for σ would be approximately from $64.6/1.1$ to 64.6×1.1, or 59–71. However, since they were in fact based on scores, with an efficiency of only $90/200 = 0.45$, $\mathrm{SE}\{\log(\hat{\sigma})\}$ is approximately $\{1/(4n \times 0.45)\}^{1/2} = (1/180)^{1/2}$, the multiplicative margin of error for a 95% confidence interval is approximately 1.16 and so the limits are approximately from $64.6/1.16$ to 64.6×1.16, or 56–75. For σ-values in this range, the information content of the ring data is $2 \times 171/400$ths, or 85.5% that of the full location data.

Others may wish to explore what additional data could be used to recover more of the information about the more complex variance structures that were considered by Tibshirani *et al.* (2011).

Again, we salute Tibshirani *et al.* (2011) for their readable modern essay and for maintaining a statistical focus on marksmanship, yet using less dangerous missiles than those studied by statisticians of centuries past.

References

Klein, J. L. (1997) *Statistical Visions in Time: a History of Time Series Analysis 1662-1938*, pp. 3-11. Cambridge: Cambridge University Press.

Pearson, K. (1900) On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag.*, ser. 5, **50**, 157–175.

Tibshirani, R. J., Price, A. and Taylor, J. (2011) A statistician plays darts. *J. R. Statist. Soc.* A, **174**, 213–226.

Turner, E. L. and Hanley, J. A. (2010) Cultural imagery and statistical models of the force of mortality: Addison, Gompertz and Pearson. *J. R. Statist. Soc.* A, **173**, 483–499.

Sudipta Sadhukhan, Zhihui Liu and
James A. Hanley
McGill University
1020 Pine Avenue West
Montréal
Québec
H3A 1A2
Canada
E-mail: james.hanley@mcgill.ca

[Received February 2011. Revised May 2011]