VARIABILITY OF SCORES AND CONSISTENCY IN SPORT

Stephen R Clarke¹

Abstract

This paper discusses the concept of consistency in sport, and its relationship with the variability of the outcomes. Using examples from golf, cricket and field events, it is demonstrated that highly variable results will produce more wins, and in some cases better average scores than 'consistent' results. It is also shown that consistent behaviour can produce highly variable outcomes. This implies that if coaches or sports followers use subjective feelings about the actual scores of players or teams to judge consistency, they are likely to be severely misled. Given the importance of variability to the outcome of sporting events, the evaluation and publication of suitable measures such as variance is suggested.

1. INTRODUCTION

Mathematicians interested in sport can probably be divided into two main groups—those interested in deterministic aspects such as dynamics and mechanics, and those into stochastic aspects. I am definitely in the latter category, being particularly interested in using probabilistic models to assist with strategy or tactics. In this paper I wish to use some simple models to explore the notion of consistency in sport, and its relation to the variability of the outcomes or scores.

It is a basic tenet of sport that an essential characteristic of excellence is consistency. Sportsmen are taught to strive to improve consistency as this will improve performance, and this is often equated with low variation in outcomes. For example Pollock [1] in discussing the mean and standard deviation of scores in relation to consistency of golfers says 'it seems reasonable that a better player would have low values for both'. This paper demonstrates that this attitude is often not the case. In sports such as golf, where one player competes simultaneously against many others, highly variable scores (ie ones with a large standard deviation) will result in more wins than less variable scores with the same mean level. Furthermore, it is shown that for competitions where the final score is the best of several attempts, the variability in the individual attempts contributes positively to the final score. In judging the consistency of athletes, coaches and followers usually look at their results. This can be misleading, as it is shown that highly variable results will often arise from consistent behaviour. Indeed, in some cases, the more consistent the behaviour the more variable are the results. Given its importance it would be beneficial if statistics that measure variability were published.

¹Mathematics Department, Swinburne University, Hawthorn, Vic.
2. **Golf**

A golfer's score on each hole $X_i$ can be considered as so many shots over par (negative if under). Since par takes into account the difficulty of the hole, this should produce $X_i$'s reasonably independent and identically distributed. Thus the total $S$ for 18 holes is given by

$$ S = X_1 + X_2 + \ldots + X_{18} \quad \text{(1)} $$

and so by the central limit theorem should be normally distributed. The above normality and independence assumptions are supported by Scheid [2] who reports the net scores of all but the worst scores of 3000 Massachusetts players to be normal, with a very small correlation between scores on one hole and the next. Similarly, Wilson [3] examined the scores of many players of differing ability and found the better 40% of scores for each player approximated the normal distribution. Thus the two parameters of a normal distribution are appropriate, and we can use $\mu$ as a measure of a golfer's ability and $\sigma$ as a measure of the variability of the scores (however as I will show later, perhaps not of consistency).

### 2.1 Consistent scores mean fewer wins.

This section aims to show that of two players with the same average, the more variable will win more tournaments. In fact, a player of lesser ability can win more tournaments than a better player if they are also more variable. A typical top class tournament has over 100 players, all of varying abilities, and is played over 72 holes, the lowest total winning. We need to find the probability that a given player will have a lower score than the minimum score of all the other players, but here we make some simplifying assumptions. Also in practice, only those under the cut-off score after the first 36 holes complete the tournament. This is typically about 60 players. However, since those players cut would have little chance of winning the tournament if they were allowed to continue, we can ignore this complication. Since the score necessary to win is the best performance of many players, it will always be far better than the average performance of any individual player. Below we assume that 10 under (–10) will win the tournament, but the argument is the same for any chosen figure. Note that in some tournaments players scores are generally lower than for other tournaments. However by replacing par with the average that the course actually played, the argument remains valid. For example, in the 1992 US masters, the winner Fred Couples needed to score a total of 9.8 less than the average total of the 56 qualifiers. Given $\mu$ and $\sigma$ for a golfer, it is a simple task to find their chance of achieving a certain score for the tournament.

A player with a mean score $\mu$ and standard deviation $\sigma$ over 18 holes will have a mean $4\mu$ and standard deviation $2\sigma$ over 72 holes. Suppose a score of $w$ is required to win.

Then

$$ \Pr(\text{Score} \leq w) = \Pr [ Z \leq (w - 4\mu + 0.5) / 2\sigma ] $$

$$ = \Phi [ (w - 4\mu + 0.5) / 2\sigma ] $$

where $Z$ is the standard normal and $\Phi(z)$ is the area under the normal curve to the left of $z$. The 0.5 is needed as a correction for continuity.
Wilson [3] states that standard deviations for golfers' scores over 18 holes ranged from 2 to 6, with lower-handicap golfers having the lower values.

For example, if \( w = -10 \), \( \mu = -1 \) and \( \sigma = 2 \),

\[
\Pr( \text{Score} \leq -10 ) = \Phi\left( \frac{-10 - 4.1 + 0.5}{4} \right) \\
= \Phi(-1.375) = 0.085
\]

Thus a golfer who each round averages 1 under par with a standard deviation of 2 will win 8.5% of tournaments, assuming 10 under is needed to win.

The results of this calculation for various values of \( \mu \) and \( \sigma \) are shown in Table 1. This shows the percentage chance of winning a tournament, for values of \( \mu \) from +1 to -3, and \( \sigma \) from 1 to 4.

Table 1

<table>
<thead>
<tr>
<th>Percentage chance of score better than 10 under</th>
<th>( \sigma ) per round</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0</td>
</tr>
<tr>
<td>( \mu ) per round</td>
<td>0.0</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.0</td>
</tr>
<tr>
<td>-1.0</td>
<td>0.3</td>
</tr>
<tr>
<td>-1.5</td>
<td>4.0</td>
</tr>
<tr>
<td>-2.0</td>
<td>22.7</td>
</tr>
<tr>
<td>-2.5</td>
<td>59.9</td>
</tr>
<tr>
<td>-3.0</td>
<td>89.4</td>
</tr>
</tbody>
</table>

Note firstly that the increase in \( \sigma \) is only helpful in winning more tournaments if \( \mu \) is less than the value required to win. The last 2 rows show a decreasing chance of winning as \( \sigma \) increases. Thus for players so much better than everyone else that their average is better than the winning score (in golf there aren't any), it is in their interests for their scores to be as consistent as possible. For everyone else, the reverse is true.

A player averaging par, with \( \sigma = 2 \) will barely win 1% of tournaments. By reducing their average by 1/2 a stroke, or increasing \( \sigma \) by 1/2 a stroke, they increase their tournament wins to just on 3%. If they do both they increase their winning percentage to 6.7%.

Clearly a large variation in scores is advantageous. What are the ramifications for golfers? Clearly, given a choice, a player should choose strategies that maximise variation. For example, if faced with a choice of playing over a bunker, with a 50-50 chance of the ball getting on the green with a consequent birdie, and a 50-50 chance of finding the bunker and a certain bogie, or playing around the bunker for a certain
par, the player should choose the first alternative. Similarly, players should avoid any tendency to play the first round of a tournament 'steadily' if that means producing an average round with no disasters and no great rounds.

Note that the argument applies not just to winning tournaments. The prize-money for most tournaments is not a linear function of position, with the incremental prize-money increasing as you move up the finishing order. Thus a 10th and 12th will return more in prize-money than coming 11th twice.

If golfers put such a store in 'consistency' (even though they have it back to front), and the variance of a player's rounds is important to their chance of winning, why is it not published as a performance statistic? It would be interesting to check if top golfers do vary significantly in their round-by-round standard deviations. Rotella and Boucher [4] use regression analysis on the playing statistics of professional golfers to predict money earned. Over 13 published statistics were used, but standard deviation of scores was not one of them. However birdies divided by greens in regulation was the second most important variable behind scoring average. This statistic may be a surrogate for variance, as it goes up when the number of bogies goes up as well as when the number of birdies increases. Hale & Hale [5] find that for the leading money winners the performance statistics are not a good predictor of success - perhaps this is further evidence that some others are needed.

This argument can apply to many sports. In an earlier article on cricket Clarke [6] showed that success in Sheffield Shield is dependent on having highly variable results - viz lots of outright even if many are losses. A team can actually lose more than half their matches and still finish on top of the shield table. In discussing the triathlon, de Mestre [7] suggests that the influence of an event in determining the winner is proportional to the variance.

2.2 Consistent behaviour can mean variable results

While the above seems to go against the traditional view that consistency of performance is a desirable trait, this apparent contradiction can be resolved, as it is shown that the most variable scores are often produced by consistent behaviour or play. The consistency an athlete attempts to achieve is consistent behaviour. Thus a golfer's swing should be the same all the time, and not alter under pressure. A cricketer is required to treat every ball on its merits, to concentrate equally throughout an innings, and not allow lapses in effort. Similar golden rules exist in most sports that fundamentally require consistency of behaviour.

However, in judging the consistency or otherwise of an athlete, the performance (or set of scores) of the athlete is looked at, and consistency judged by the range of performance - the smaller the range the more consistent. Players with a large range of scores may be judged to be inconsistent, and in need of some help to improve the consistency of their behaviour. In fact the reverse is often true - a small range in scores can be indicative of inconsistent behaviour, while a large range may be generated from consistent application.
While the following is developed for golf, it applies to sports like golf, rifle shooting and archery, where a player has a series of attempts which can be classed as a success (par, birdie, bulls-eye etc) or a failure.

Let 1 represent a success, and 0 a failure, and suppose there are (a constant) n attempts. Thus in golf, for a top golfer 1 might represent a birdie, 0 a par and we have 18 attempts in one round. Suppose attempt i has a probability of success $p_i$.

Let $X_i = 1$ with probability $p_i$

$= 0$ with probability $(1 - p_i)$

Then $X_i$ (either from first principles or as a binominal variable with $n=1$, $p=p_i$ ) has mean $p_i$ and variance $p_i(1 - p_i)$

As before, let the score

$$S = X_1 + X_2 + X_3 + ... + X_n = \sum_{i=1}^{n} X_i$$

(2)

Then the mean or expected value of the score $E(S)$ is given by

$$E(S) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p_i$$

Assuming independence, the variance of the score $Var(S)$ is given by

$$Var(S) = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} p_i(1-p_i)$$

$$= \sum_{i=1}^{n} p_i - \sum_{i=1}^{n} p_i^2$$

$$= E(S) - \sum_{i=1}^{n} p_i^2$$

Thus for a given mean score $E(S)$, the variance of the score is a maximum when $\sum_{i=1}^{n} p_i^2$ is a minimum. Using calculus and the Lagrange multiplier technique, Bellman and Dreyfus [8], this is easily shown to be when all $p_i$ are equal.

Thus the maximum variation in total scores occurs when all attempts have the same chance of success.

For example, in golf, a player who always begins slowly but improves their chance of a birdie throughout the round will show less variation in scores than one who has a constant chance of success.
While this model would need to be greatly improved to form a realistic model of golf, the results of this simple model are very clear. A player whose concentration wavers, or who tries harder some holes than others, will produce scores that are less variable, and so appear more consistent than one who plays each hole the same. A coach who berates a rifle shooter for inconsistency for scores ranging between 85 and 95 may be better off concentrating on the shooter who always scores 90.

2.3 Hole-by-hole variances

Consider 2 golfers: one short but straight, a cautious putter who invariably gets par on a hole; the second a long but sometimes wayward hitter and bold putter who has a good chance of a birdie but also a good chance of a bogie. In a round both players could both get par, but one may get it by having 18 pars, while the other may get 6 pars, 6 birdies and 6 bogies. This difference could be measured by the variance or standard deviation of the X's, the hole-by-hole variance, where the first would show up with a low value and the second with a high value. Now in general \( \text{Var } S = \Sigma \text{Var } X_i + \Sigma \text{cov}(X_iX_j) \). If holes are independent, then \( \text{Var } S = 18 \text{ Var } X \). A player who plays better when up and worse when down (\( \text{cov}(X_iX_j) > 0 \)) would have a \( \text{Var } S > 18 \text{ Var } X \), while a player who loses concentration when up and tries harder when down (\( \text{cov}(X_iX_j) < 0 \)) would have \( \text{Var } S < 18 \text{ Var } X \). Thus again an insight into the psychology or mental attitude of the player might be gained by the extra statistics, with consistency being measured by the closeness of \( \text{Var } S / (18 \text{ Var } X) \) to 1.

For example, consider the following four golfers' scores on 2 rounds of 9 holes. Each has the same total score of zero over par.

<table>
<thead>
<tr>
<th>Player</th>
<th>Rnd</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>-8</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-2</td>
<td></td>
</tr>
</tbody>
</table>

The statistics for each player are shown below

<table>
<thead>
<tr>
<th>Player</th>
<th>Mean score</th>
<th>VarX</th>
<th>VarS</th>
<th>( VarS / 18\text{Var }X )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0</td>
<td>0.36</td>
<td>2</td>
<td>0.615</td>
</tr>
<tr>
<td>b</td>
<td>0</td>
<td>1.11</td>
<td>2</td>
<td>0.200</td>
</tr>
<tr>
<td>c</td>
<td>0</td>
<td>0.11</td>
<td>128</td>
<td>28.000</td>
</tr>
<tr>
<td>d</td>
<td>0</td>
<td>0.94</td>
<td>8</td>
<td>0.941</td>
</tr>
</tbody>
</table>
Note the difference in the extra statistics. Golfers b and d are shown as 'attacking' players by the high variance of $X$, golfer d is shown as consistent by a ratio near 1, c is shown as highly variable, and b & c as inconsistent.

2.4 Rating the Golf course

In this section the point of view is switched to the golf course. A lot of work has been done on rating the difficulty of a golf course, but little on the discrimination powers of individual holes or the course as a whole.

Clarke and Rice [9] have looked at rating courses. Hole-by-hole data of players' scores in the 1992 U.S. Masters' tournament obtained from the organisers included the average score on each hole, but the only indication of the variation of scores was the number of eagles, birdies and bogies. These are very difficult to interpret. Table 2 shows the mean and standard deviation of scores on each hole. It is clear that holes 2, 8, 13 and 15 (all par 5's) were easier to play than the others. However it is not only the ease or difficulty of a hole that is important. If everybody takes 1 over par on a hole, that hole is not separating players. The standard deviation gives a measure of the ability of a hole to discriminate between players. Note that hole 10, clearly the hardest hole over the 4 rounds, was on only one day in the top half of the holes in order of discrimination. In this regard holes 12 and 13 clearly stand out above the rest, between them sharing the highest and second highest standard deviation in every round. Hole 12, a par 3, is not unduly difficult, but clearly produces highly variable scores. In Parsons [10] the section of the Augusta course from holes 11 to 13 is described thus "The Masters championship has been won or lost so often between the 11th and 13th that this three hole stretch has become known as Amen corner". In Parsons 10], Jack Nicklaus describes the 12th as " the most demanding tournament hole in the world". In this case, the standard deviations reflect golfers' views of the holes. Although the standard deviation of the scores on a hole contains much more information about the importance of the hole to the result of the tournament, this statistic is never quoted. It should be:
Table 2

Descriptive Statistics for the 63 Qualifiers in the US Masters

<table>
<thead>
<tr>
<th>Hole</th>
<th>Par</th>
<th>Round</th>
<th>ONE Mean</th>
<th>Std</th>
<th>TWO Mean</th>
<th>Std</th>
<th>THREE Mean</th>
<th>Std</th>
<th>FOUR Mean</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td></td>
<td>-0.05</td>
<td>0.55</td>
<td>-0.02</td>
<td>0.68</td>
<td>0.05</td>
<td>0.55</td>
<td>0.02</td>
<td>0.63</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td></td>
<td>-0.22</td>
<td>0.73</td>
<td>-0.56</td>
<td>0.56</td>
<td>-0.21</td>
<td>0.72</td>
<td>-0.29</td>
<td>0.79</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td></td>
<td>0.00</td>
<td>0.48</td>
<td>-0.08</td>
<td>0.55</td>
<td>0.21</td>
<td>0.72</td>
<td>0.10</td>
<td>0.59</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td></td>
<td>0.11</td>
<td>0.63</td>
<td>0.16</td>
<td>0.60</td>
<td>0.22</td>
<td>0.66</td>
<td>0.14</td>
<td>0.50</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td></td>
<td>0.17</td>
<td>0.58</td>
<td>0.03</td>
<td>0.44</td>
<td>0.19</td>
<td>0.56</td>
<td>0.13</td>
<td>0.66</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td></td>
<td>0.08</td>
<td>0.52</td>
<td>-0.03</td>
<td>0.65</td>
<td>0.06</td>
<td>0.54</td>
<td>0.10</td>
<td>0.56</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td></td>
<td>0.05</td>
<td>0.58</td>
<td>0.05</td>
<td>0.61</td>
<td>-0.24</td>
<td>0.76</td>
<td>0.19</td>
<td>0.72</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td></td>
<td>-0.25</td>
<td>0.65</td>
<td>-0.29</td>
<td>0.52</td>
<td>-0.17</td>
<td>0.58</td>
<td>-0.40</td>
<td>0.58</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td></td>
<td>-0.06</td>
<td>0.47</td>
<td>0.10</td>
<td>0.64</td>
<td>0.03</td>
<td>0.59</td>
<td>0.06</td>
<td>0.54</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td></td>
<td>0.17</td>
<td>0.61</td>
<td>0.11</td>
<td>0.57</td>
<td>0.22</td>
<td>0.58</td>
<td>0.40</td>
<td>0.55</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td></td>
<td>0.11</td>
<td>0.65</td>
<td>0.11</td>
<td>0.63</td>
<td>0.03</td>
<td>0.44</td>
<td>0.22</td>
<td>0.55</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td></td>
<td>-0.03</td>
<td>0.78</td>
<td>0.21</td>
<td>0.70</td>
<td>0.17</td>
<td>0.93</td>
<td>0.41</td>
<td>1.10</td>
</tr>
<tr>
<td>13</td>
<td>5</td>
<td></td>
<td>-0.41</td>
<td>0.73</td>
<td>-0.35</td>
<td>0.77</td>
<td>-0.46</td>
<td>0.80</td>
<td>-0.44</td>
<td>0.96</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td></td>
<td>-0.21</td>
<td>0.45</td>
<td>0.11</td>
<td>0.57</td>
<td>0.14</td>
<td>0.64</td>
<td>-0.08</td>
<td>0.52</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td></td>
<td>-0.24</td>
<td>0.69</td>
<td>-0.54</td>
<td>0.64</td>
<td>-0.57</td>
<td>0.76</td>
<td>-0.65</td>
<td>0.63</td>
</tr>
<tr>
<td>16</td>
<td>3</td>
<td></td>
<td>0.00</td>
<td>0.54</td>
<td>-0.02</td>
<td>0.58</td>
<td>0.02</td>
<td>0.55</td>
<td>-0.05</td>
<td>0.68</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td></td>
<td>-0.02</td>
<td>0.58</td>
<td>-0.19</td>
<td>0.59</td>
<td>0.02</td>
<td>0.55</td>
<td>0.11</td>
<td>0.57</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td></td>
<td>0.02</td>
<td>0.49</td>
<td>-0.11</td>
<td>0.54</td>
<td>0.19</td>
<td>0.59</td>
<td>0.00</td>
<td>0.54</td>
</tr>
</tbody>
</table>

TOTAL 72  -0.78  2.50  -1.30  2.15  -0.10  2.76  -0.03  3.17

However a hole should also discriminate between golfers appropriately. The better players should do well and the poor do badly. Clarke and Rice [9] have investigated this by applying some standard psychological procedures to golf scores.

3. Field Events: Better Scores from Inconsistent Performance

In many sports the penalties for poor attempts are not great. In the long jump the athlete's score is the longest out of 3 attempts. Similar rules apply to the triple jump, javelin, discus, shot put and hammer throw. Thus unlike golf, poor attempts do not bring down the score. In such cases, variation in jumps (inconsistent jumping?) will not only increase the chance of winning the event, but will actually produce a greater score.

The statistics of extremes is quite complicated, particularly the distribution of maximum values. However Gumbel (1958) gives some graphs and formulas for the
average maximum of n standard normals from which the table below has been derived.

Table 3

Expected maximum of n normal variates with mean \( \mu \) and standard deviation \( \sigma \)

<table>
<thead>
<tr>
<th>Number of variates</th>
<th>Expected maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \mu )</td>
</tr>
<tr>
<td>3</td>
<td>( \mu + 0.85 \sigma )</td>
</tr>
<tr>
<td>6</td>
<td>( \mu + 1.25 \sigma )</td>
</tr>
<tr>
<td>100</td>
<td>( \mu + 2.50 \sigma )</td>
</tr>
</tbody>
</table>

Ladany [11] collected data for an athlete, showing the actual length of jump from take-off to landing was normally distributed with a mean 701.23 cm with a standard deviation of 20.44 cm. We will ignore problems associated with the aiming line and disallowed jumps, and take this to be the distribution of the recorded jump. Thus the expected score for this jumper is 701.23 + 0.85\times20.44 = 701.23 + 17.30 = 718.53 cm. Note that over 17 cm of his score arises through the variation in the jumps.

Note also that every increase of 10 cm in the variation results in an increase of 8.5 cm in the expected score. Thus for example, a 10 cm deficiency in the average jump will be made up for by a 12 cm increase in the variation.

In some competitions, the top competitors are allowed a further 3 jumps giving 6 jumps in all and even more advantage to the inconsistent jumper. A jumper now gains 25% more from an increase in variability than the same increase in average. Note also, that the actual maximums of the inconsistent jumper would be more variable (about their greater mean) than for the consistent jumper. Thus we also have the previous effect operating as well, giving the inconsistent jumper even more chance of winning an event.

When records are considered, we are looking at a jumper's personal best. An increase of 10 cm in the variation of jumps would add 25 cm to the best of 100 jumps. This has ramifications for selection, as some selections are effectively done on a personal best (eg Olympic qualifying times). This may disadvantage the consistent performer, who may be expected to do relatively better in the actual competition when only 3 or 6 jumps are allowed. On the other hand, the person with the large variance may produce the statistical outlier that produces a medal. In discussing Beamont's famous long jump, Brearley [12] concludes it was a 'statistical miracle'.

Do field athletes have different standard deviations? Unfortunately, once again, there is little data allowing an investigation of these effects. In most cases, only the athlete's best effort is recorded.

4. Cricket

As another example of consistent behaviour producing seemingly inconsistent scores let's look at a batsman's scores at cricket. Consider J.D Siddons, who batted about
number 6 for Victoria in the Australian Sheffield Shield competition in 1985/6. His scores for the year were

33, 17, 76, 5, 74, 7, 7, 107, 1, 45, 17, 2, 36.

Many cricket followers would say that is an inconsistent set of results, since they expect a consistent batsman to have scores with a small standard deviation, like 51, 55, 52, 53, 54. However scores like this mean that a batsman has no chance of going out until he/she reaches 50, and is almost certain to go out soon after. So in terms of probability of dismissal they are very inconsistent. A consistent batsman who has a 30% chance of making 50, should turn 30% of those 50's into centuries, and 30% of his/her centuries into 150's etc.

This assumption of a constant probability of dismissal leads to a geometric distribution (or its continuous counterpart, the negative exponential) for scores. The negative exponential distribution is common as the distribution of waiting times for random events - in this case it is the waiting time (measured by score) until a dismissal. A histogram of Siddons' scores and the geometric distribution with $p = 1/33$ are shown in Figure 1. The two are virtually identical.

Clearly Siddons' scores follow closely what theory suggests a consistent player with an average of 33 should produce. The standard deviation of Siddons' scores is 34, again agreeing to that predicted by an exponential distribution whose standard deviation is equal to the mean. Followers who judge Siddons to be inconsistent on the basis of his scores would be doing him a great injustice.

![Figure 1: Comparison of batting scores with Geometric distribution](image)

Elderton [13] and Wood [14] discuss further the meaning of consistency when applied to cricketer's scores, and the fitting of the geometric distribution. They show that not only are there theoretical reasons why the standard deviation of players scores should be equal to or greater than the mean, but that an analysis of actual scores of many players confirms this to be so. It is players with a small variation in their scores who should be the prime targets for the criticism of supporters or remedial coaching for lack of consistency in application. Wood suggested using the
Coefficient of Variation (CV) as a measure of consistency, with the closer to 100, the more consistent a batsman. Pollard [15] claims a high CV indicates a batsman has problems early, but scores runs more easily later in the innings. However, Clarke [16] showed that under a model similar to equation (2) perfectly consistent batsmen will have CV's greater than 100 and that perfectly consistent batsman, but with a different scoring profile, will have different coefficients of variation between 100 to 105. Thus it is not possible to have a single measure (closeness to 100) which indicates perfect consistency for all batsmen as claimed by Wood. The CV can be approximated by

\[ CV(S) = 100 + 50(R/m)(CV(X)^2 - 1) \]  

(3)

where R is the Run rate, m is the average score and CV(X) is the coefficient of variation of X, where X is the score each ball. This gives the coefficient of variation of scores in terms of the 3 parameters that describe a batsman's scoring profile: m is their average score and describes how many runs they get, R is the rate and describes how fast they get them, and CV(X) describes how the runs are distributed between singles, fours etc. Clarke [16] also suggests several other measures, including the mean and standard deviation of the number of balls faced, and the standard deviation of X to further describe batsmen, and gives these statistics for a 1-day series.

At the very least the publication of batsmen's standard deviation of scores would give some idea of their consistency - a standard deviation about the same level (or slightly higher) indicates consistency. Anything less does not.

5. Variability of the team's score

Just as individuals are often unjustly criticised for inconsistency, so are whole teams. It seems there is a lack of understanding in the general population as to the inherent variability in a lot of games. The variability in a cricket innings is easily investigated. The innings score is the sum of 10 partnership scores. Assuming each partnership is exponential with mean 30 (and variance 30^2) then the innings score is mean 300 and variance 9000, or standard deviation of about 95. Thus the innings could be anything from 100 to 500 - highly variable. A more realistic model assuming partnerships of different expected lengths only increases the variation. Johnston [17], Johnston et al [18] simulated one-day cricket innings using optimal batting rate policies and obtained a mean score of 215 with a standard deviation of 45. The lowest of 1000 simulations was 75, the highest 322. Clearly much of the variation we see in cricket can be explained by the highly variable nature of the game and is not necessarily due to good or bad play.

6. Conclusion

Sports followers need to be careful in assessing the attribute of consistency, and in extolling its virtues. While consistency in behaviour is to be encouraged this will not necessarily reflect itself in low variability of scores or results. However, low variability in results is often not even a desirable outcome. In some events it will decrease the actual score, while in others it will reduce the number of events won.
If coaches use subjective feelings about the actual scores of players to judge consistency, they are likely to be severely misled. Players who have small variation in results may be the prime targets for remedial coaching or the attention of psychologists, not the players with large variation in outcomes.

With the increasing use of computers, the evaluation of other statistics which measure the variability of players and teams performance is made simpler. The calculation and publication of these statistics could assist players and coaches to a better understanding of their performances, and at the least would provide followers with some interesting data. With a suitable choice of names, these could be interpreted by the general public.

REFERENCES


Second Conference on:

MATHEMATICS AND COMPUTERS IN SPORT

Held at Bond University, Queensland, Australia
11th to 13th July, 1994

*Sponsored jointly by*
The Australian Mathematical Society

and

The Australian Sports Commission

Edited by Neville de Mestre, Associate Professor of Mathematics
School of Information Technology, Bond University.