FORE OR FIVE? - THE INDEXING OF A GOLF COURSE

* I.G. Ó Muircheartaigh and J. Sheil

1. Introduction

In every golf club, for reasons which relate mainly to the rules and regulations of handicap competition, it is necessary to rank the 18 holes of the course in order of difficulty so as to provide a strokes index for each hole ranging from 1 (most difficult) to 18 (least difficult). This is normally done on a subjective basis, and is consequently a frequent topic of controversy in many golf clubs. In this note, we propose a more objective method for performing this indexing, and we apply it to a particular case (i.e. our own golf course).

The par score for a hole is the score which a top-class (scratch) player would be expected to have at that hole. An obvious measure of the difficulty of a hole is the average amount by which players exceed the par score for that hole. It was decided, however, not to use this particular measure because of the distortive effect on the mean of outlier data points; such scores could occur quite frequently at particular holes (e.g. those with psychologically intimidating features such as out of bounds close to the tee, water hazards close to the green, etc., etc.) thereby inflating the mean score disproportionately to the intrinsic difficulty of the hole. A related measure which overcomes this problem is the percentage or proportion of players who equalled or bettered par at the hole. This is the measure which we have adopted, and which is used in this paper.

Some features of this measure are worthy of note, as they motivate the analysis described below. Firstly, the percentage who better or equal par at any hole can be readily estimated from the records of golf competitions which are available in all golf clubs. In fact, the authors of this note have the distinction as applied statisticians of not only having collected their own data but also through their golf, actually contributing to it. Secondly, it is clear that this measure can be evaluated for players of different calibres (i.e. handicaps) and that different percentages (or probabilities) can be expected for the various categories of golfer: by definition, the average percentage who equal or better par at any hole will be higher for good golfers than for poor golfers. Finally, it is intuitively clear, and from (bitter!) experience glaringly obvious, that the relative difficulty of any hole is very dependent on weather conditions, and in particular, on wind speed and direction.

In this paper, we postulate that the probability of equalling or bettering par at any hole depends on the variables handicap, wind speed and wind direction via a logistic function; we estimate the parameters of this function from a large data-set of golf scores; test the goodness of fit of the model and, having accepted the model, calculate the expected value of this criterion, with respect to the variables handicap, wind speed and wind direction, for each hole and thus provide a ranking of the 18 holes.

2. The Data

The data consisted of scores for a total of 575 players, spread over five different competitions (i.e. days). An initial analysis confirmed that the probability of at least equalling par was heavily dependent on both the handicap of the player and weather conditions (i.e. wind speed and direction). Data were also available for the actual speed and direction on each day (at six-hourly intervals); the particular competitions whose scores were used in this analysis were chosen specifically on the basis of minimum variability of the six-hourly readings, and the wind speed and direction (taken as the average of the two day-time readings) were then assumed constant throughout the day.

For illustration, we provide in table 1, a typical data set - i.e. that for hole 5. This set highlights many of the points already made. In particular, on any given day the probabilities of at least equalling par are markedly different for the two handicaps classes. Furthermore, this probability, for either handicap class, varies widely from day to day. For example, in the case of competition 1, for category 1 golfers this probability was 0.20, whereas for competition 5 it was virtually doubled (to 0.39). An explanation of this variation is provided by the fact that the wind direction on day 1 was 330 (virtually directly against the hole which faces 5 degrees east of North) whereas on day 5 the wind direction was 200, almost a directly following wind.

3. Fitting the Model

We have now established that the probability of at least equalling

HOLE 5

COMPETITION	W	θ	H	# LE PAR	#GT PAR	OBSERVED PROPORTION ≤ PAR
1	8	330	1	9	35	.20
	8	330	2	1	49	.02
2	10	80	1	12	30	.29
	10	80	2	1	26	.04
3	4	210	1	37	751	.34
	4	210	2	12	115	.09
4	18	290	1	11	29	.28
5	7	200	1	53	84	.39

TABLE 1

par at a hole appears to depend on a number of factors. These are represented by the variables:

W = Wind Force

θ = Wind Direction

H = Handicap class of player (1 if player's handicap ≤ 10)
(2 if player's handicap > 10)

We postulate that there is a functional dependence of P_i, the probability of at least equalling par at hole i, on these variables, i.e. that:

(1) $P_i = g_i\ (W,\theta,H) \quad (i = 1,2,\ \ldots\ , 18)$.

We further postulate that the logistic model represents an appropriate class of functional forms to describe the relationships (1). The logistic model is given by

(2) P_i = prob {Score ≤ par $X_1, X_2, \ldots ,$}

$$= \frac{\exp\left(a_0 + \sum a_j X_j\right)}{1 + \exp\left(a_0 + \sum a_j X_j\right)}$$

where $X_1, X_2, \ldots ,$ are independent variables. The independent variables used in our analysis were

X_1 : W

X_2 : W cos $(\theta - \alpha_i)$, where α_i = direction of hole i

X_3 : H

the inclusion of each of which can be justified - *a priori*, on heuristic grounds, and *a posteriori* on the basis of their explanatory capacity.

The model (2) was then fitted to the data (using the software package BMDP) to provide estimates of the co-efficients a_0, a_1, a_2, a_3 for each hole, t-values for these co-efficients and appropriate goodness-of-fit statistics.

For illustrative purposes, we describe here the fitting of the logistic model to the data for hole 5 (presented in table 1). Similar analyses were performed for each hole.

The parameters estimated for equation (2) for hole 5 were (t-values in brackets)

$$a_o = 0.767 \quad (1.39)$$

$$a_1 = 0.015 \quad (0.43)$$

$$a_2 = -0.071 \quad (-.255)$$

$$a_3 = -1.759 \quad (-5.55)$$

The goodness of fit chisquare (1.639) has p-value 0.802, and does not lead to rejection of the model. A more intuitive presentation to highlight the adequacy of the model is to use the estimated model to predict the expected or theoretical probabilities of at least equalling par for each of the five days (i.e. W, θ combinations) and for each handicap category - i.e. a predicted probability corresponding to each row of table 1. Table 2 presents these theoretical probabilities for hole 5, together with the observed probabilities as already given in table 1. The extent of the agreement is remarkably good.

A similar analysis performed for each hole produces an estimated logistic function formulation of the functional dependence of the probability P_i on the various independent variables, which in all cases produces very good agreement between observed and predicted probabilities. In only one case does the chisquare goodness of fit statistic lead to rejection of the model at the 5% significance level, and this is just about what we would expect if the model were appropriate. We have, therefore, now established a relationship of the form $P_i = g_i(W,\theta,H)$ for each hole for any given combination of the variables W,θ,H.

To obtain an overall "average" index it is necessary to establish the joint distribution of W,θ,H. H is clearly independent of W,θ, and had for our data the very simple probability distribution

$$p(H = 1) = p(H = 2) = \tfrac{1}{2}$$

Denoting the joint distribution of W,θ by $f(W,\theta)$, the expected value of P_i as formulated by us will therefore be given by

$$\bar{P}_i = \sum_{H=1}^{2} \frac{1}{2} \iint g_i(W,\theta,H)\, f(W,\theta)\, dW\, d\theta \qquad (3)$$

HOLE 5

COMPETITION	H	OBSERVED PROPORTION ≤ PAR	PREDICTED PROBABILITY ≤ PAR
1	1	.20	.21
	2	.02	.04
2	1	.29	.26
	2	.04	.06
3	1	.34	.34
	2	.09	.08
4	1	.28	.26
5	1	.39	.40

TABLE 2

In sections 4 and 5 we describe the empirical determination of $f(W,\theta)$, the resultant derivation of $\bar{P}_i$, and the corresponding indices 1 to 18.

4. Joint Distribution of Wind Velocity and Direction

The probability distribution of wind speed is commonly assumed to be a member of the Weibull family of distributions. This is an empirically based assumption which usually gives a good fit to data from sites which exhibit a prevailing wind direction - when there is no prevailing wind direction, the Rayleigh distribution is generally found to provide a good fit. Since the Weibull model does not provide information on wind direction, it is inappropriate in our case. In a 1979 paper, MacWilliams, Newmann and Sprevak presented a simple theory for modelling the joint distribution of wind speed and direction. In a later (1980) paper, MacWilliams and Sprevak showed that the model provided a very good fit to wind data from the 14 sites in the Republic of Ireland and 5 in Northern Ireland for which hourly data were available. Our model and estimation procedure are effectively those introduced by MacWilliams et. al.

Let W = wind velocity

W_x = Component of wind velocity along the prevailing wind direction

W_y = component of velocity perpendicular to the prevailing direction

θ = Radian measure of the prevailing wind direction.

We assume that

(i) $W_x \sim N(\mu,\sigma^2)$; $W_y \sim N(0,\sigma^2)$

(ii) W_x, W_y are stochastically independent; hence the joint p.d.f. of W_x, W_y is given by

$$g(W_x, W_y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}\left[W_y^2+(W_x-\mu)^2\right]\right\}$$

By making the transformation

$W_x = W\cos\theta$, $W_y = W\sin\theta$

we obtain the joint p.d.f. of wind velocity and direction

$$f(W,\theta) = \frac{W}{2\pi\sigma^2} \exp\left(\frac{-\mu^2}{2\sigma^2}\right) \exp\left\{-\frac{1}{2\sigma^2}\left[W^2-2\mu W\cos\theta\right]\right\},$$

$$0 \leq \theta < 2\pi, \quad W \geq 0$$

The marginal distributions of wind velocity and direction are obtained by integrating $f(W,\theta)$ over θ and W respectively giving

$$h_W(\omega) = \frac{W}{\sigma^2} \exp\left(\frac{-\mu^2}{2\sigma^2}\right) \exp\left(\frac{-\omega^2}{2\sigma^2}\right) I_0\left(\frac{\mu}{\sigma^2} W\right), \quad \omega \geq 0$$

$$h_\theta(\theta) = \frac{1}{2\pi} \exp\left(\frac{-\mu^2}{2\sigma^2}\right)\left\{1+\frac{\mu}{\sigma}\sqrt{2\pi}\cos\theta \exp\left(\frac{\mu^2\cos^2\theta}{2\sigma^2}\right)\right\}\Phi\left(\frac{\mu\cos\theta}{\sigma}\right),$$

$$0 \leq \theta < 2\pi$$

where $I_0(.)$ is the modified Bessel function of the first kind and order zero and $\Phi(.)$ is the standard normal distribution function. It is worth noting that $h_w(w)$ and $h_\theta(\theta)$ became the Rayleigh density function and $^1/2\pi$ respectively when μ is zero.

5. Derivation of Expected Probabilities

Before applying $f(W,\theta)$ to compute P_i : 1, 2, ... , 18 we must first supply estimates for μ, σ^2. The data used in the estimation were recorded (4 times daily) by the Galway Meteorological Station over the 12 month period September, 1980 - August, 1981. Of the data available, only those values relating to daylight hours were used. The data consisted of wind speeds which were presented in integral units of the Beaufort scale and wind direction - specified in sectors of 10° width so that for example, a recorded direction of 9 means a wind blowing from a point 85° - 95° measured clockwise from North. The parameters specifying the function $f(W,\theta)$ were estimated by resolving the observed data into the two orthogonal directions after the prevailing direction had been obtained. The prevailing direction was taken as being the centre of the sector having the largest frequency of occurrence (sector 24).

The estimates obtained were as follows:

(i) Prevailing Direction: 240°

Hole No.	Expected Probabilities of ≤ par		Average with respect to handicap	Corresponding Indices
	Handicap Class 1	Handicap Class 2		
1	.36929	.21128	.29029	5
2	.41420	.22830	.32125	7
3	.61590	.41086	.50888	13
4	.58107	.35697	.46902	11
5	.34095	.08877	.21486	1
6	.29515	.17937	.23726	3
7	.69868	.42739	.56304	15
8	.72849	.43354	.58102	17
9	.38985	.29714	.34350	9
10	. 8524	.16725	.22625	2
11	.51290	.21006	.36148	10
12	.61996	.49026	.55511	18
13	.44748	.18975	.31862	4
14	.44264	.27441	.35853	8
15	.51062	.35130	.43096	14
16	.50353	.22195	.36274	12
17	.59382	.27348	.43365	16
18	.42472	.22587	.32530	6

TABLE 3

(ii) $\bar{W}_x = 3.215,\ S_x^2 = 58.996$

$\bar{W}_y = -0.037,\ S_y^2 = 60.611$

$\bar{W}_x$, $\bar{W}_y$, S_x, S_y represent the average wind velocities and standard deviations in the velocities along the prevailing and perpendicular directions. The values obtained were consistent with the assumptions made and 3.215 and 59.0 were then used as estimates for μ, σ^2 respectively in the model. The integrals.

$$\int\int g_i\ (W,\theta,H)f(W,\theta)dWd\theta\ :\ i = 1,\ 2,\ 3\ \ldots\ ,\ 18$$

were evaluated using the NAG library routine DOIFCF. The results together with the suggested indices are given in Table 3.

6. Summary

The dependence of the probability of at least equalling par at any given hole on handicap, wind speed and direction is shown to be adequately described by the logistic function. The expected value of this probability with respect to the empirically derived joint distribution of the independent variables is evaluated to provide a strokes index for a golf course.

References

1. MacWilliams, B., Newmann, M.M., Sprevak, D., (1979) The Probability distribution of Wind Velocity and Direction". Wind Engineering 3., 264-269.

2. MacWilliams, B., Sprevak D., (1980) "The Estimation of the Parameters of the Distribution of Wind Speed and Direction" Wind Engineering 4 227 - 237.

Department of Mathematics,
and
Department of Industrial Engineering,
University College,
Galway.