On the development of decision rules for bar quiz handicapping

David Percy and Philip Scarf

Centre for Operational Research and Applied Statistics,
Salford Business School,
University of Salford,
Salford,
Manchester
M5 4WT, UK.

email d.f.percy@salford.ac.uk, p.a.scarf@salford.ac.uk

Salford Business School Working Paper Series

Paper no. 323/06

On the development of decision rules for bar quiz handicapping

David F. Percy and Philip A. Scarf

Centre for Operational Research and Applied Statistics, Salford Business School, University of Salford, Greater Manchester, M5 4WT, U.K.

d.f.percy@salford.ac.uk

Summary

Some competitions involve teams comprising different numbers of players. For informal games, such as the popular "pub quiz", we argue that teams with fewer players are at a disadvantage. This paper investigates the properties of these games and develops several methods for allocating handicaps within such competitions, so that the competitions may be considered fair. We recommend a prior subjective handicapping rule; with this rule handicaps may be set beforehand given the judgement of the quiz setter regarding the difficulty of the questions. The paper also considers modifications of the proposed rules to cope with multiple-choice questions and progressive quizzes.

Keywords: elicitation, games, probability, quizzes, subjective.

Introduction

Formal sports and games generally involve competition between individuals or teams under equal conditions, in order to decide winners based on the best performances. However, some competitions, mostly amateur, aim to reward improved and exceptional performances relative to given expectations based on established ability. There are two common approaches in such competitions. Firstly, players may be stratified into different categories, each with their own competition. Secondly, handicapping rules may be applied so that players of different abilities can all participate in a single competition.

The sport of golf illustrates all of these procedures. The British Open Championship is a single competition open to players of all abilities with no special allowances, other than automatic qualification to the main event for leading players. Stratification is evident from the existence of separate competitions for professional, amateur and veteran players. Finally, many amateur players throughout the country compete in weekly, handicapped tournaments at their local clubs. Other settings in which stratification occurs are the coexistence of Olympic and Paralympic Games, segregation of the sexes, hierarchical league structures, and different weight categories for fighting and strength related sports such as karate and weightlifting (e.g. Cleather, 2006). Handicapping is well known in horse racing and is the subject of academic study (e.g. Bolton and Chapman, 1986; Edelman, 2003). A more subtle form of handicapping is the draft process used in North American sports whereby, at the end a season, teams that performed worst over the season are given priority in the recruitment of new talent for the coming season (Szymanski, 2003). Handicap events are common in many amateur sports. For example, most running clubs organise an annual club handicap race, in which the slowest runners, based on performance in open races, start first and the winner is the first across the line. Similarly, some badminton leagues organise formal handicapped tournaments on an annual basis, which enable teams in different divisions to compete against one another on equal terms. Indeed, individual badminton clubs often host informal handicapped tournaments for their own members.

This paper is concerned with competitions in which teams can be disadvantaged by virtue of having fewer players than their competitors. Such a situation arises in amateur golf in the Stableford competition and handicapping rules for variable-sized teams have been developed by Lewis (2005). In this paper, we are interested in this phenomenon in the context of the popular "pub quiz"; it can also arise in other impromptu activities such as "treasure hunts" and "paintball contests." While this scenario usually only arises for informal games, adaptations of the concepts in this paper may subsequently prove to be important for extension to formal games, such as the qualifying stages of international football competitions in which the size of qualification groups may vary, or the Olympic Games in which smaller countries struggle to compete against larger countries. These issues are discussed in the concluding section.

In bar quizzes competition can be intense, prizes significant, answers carefully guarded, and conflict may arise over the interpretation of ambiguous questions and answers. Therefore, there is an implicit assumption among competitors that the contest must be a fair one. However, little attention is paid to the size of teams and anecdotally the occurrence of unequally-sized teams is commonplace. The bar quiz at the recent Operational Research Society conference (OR48) in Bath, U.K., is a case in point. This quiz was won by a team of ten players; the authors of this paper were in an opposing team of seven players. Defeat for the authors' team led them to ask the following questions: Do larger teams have an advantage over smaller teams? What is the magnitude of the advantage? How should larger teams be handicapped to provide a fair contest? The fact that we ask the final question indicates that the authors believed *a priori* that larger teams do have a positive advantage. A brief analysis shortly after the competition suggested that the winning team of 10 which scored 44½ out of 60 would have finished behind the authors' team of 7 which scored 40/60 under a "fair" handicapping system. This brief analysis is developed in the next section; its basis, a simple model for a bar quiz, is also presented. A simple handicapping rule is then proposed along with an empirical method for implementing the handicap. We then discuss methods for prior handicapping (before quiz scores are known) and for combining prior knowledge with known scores. We conclude with extensions of the ideas to questions with multiple choice answers and to progressive quizzes in which the round ends with the first incorrect answer. The final section discusses the achievements, limitations and possible extensions of the modelling.

Mathematical model for a bar quiz

We suppose that a team comprising m players enters a quiz consisting of n questions. Define the Bernoulli random variables

$$X_{ij} = \begin{cases} 0\,; & \text{player } i \text{ does not know the answer to question } j \\ 1\,; & \text{player } i \text{ does know the answer to question } j \end{cases} \tag{1}$$

with probabilities of success

$$E\left(X_{ij} \middle| \theta_{ij}\right) = \theta_{ij} \tag{2}$$

for $i = 1,2,\ldots,m$ and $j = 1,2,\ldots,n$. Given these parameters, we assume that the X_{ij} are mutually independent for $i = 1,2,\ldots,m$ and $j = 1,2,\ldots,n$. This leads us to define the Bernoulli random variables

$$Y_j = \begin{cases} 0\,; & \text{team does not score a mark for question } j \\ 1\,; & \text{team does score a mark for question } j \end{cases} \tag{3}$$

with probabilities of success

$$\phi_j = E\left(Y_j \middle| \boldsymbol{\theta}_j\right) = 1 - \prod_{i=1}^{m}\left(1 - \theta_{ij}\right) \tag{4}$$

for $j = 1,2,\ldots,n$ where $\boldsymbol{\theta}_j = \left(\theta_{1j}, \theta_{2j}, \ldots, \theta_{mj}\right)'$ is a $m \times 1$ column vector of parameters for question j, as the team scores a mark for question j if any player in the team knows the correct answer. We may then represent the team's total score for the quiz by the random variable

$$Z = \sum_{j=1}^{n} Y_j \tag{5}$$

with expected value

$$E\left(Z \middle| \boldsymbol{\Theta}\right) = \sum_{j=1}^{n} E\left(Y_j \middle| \boldsymbol{\theta}_j\right) = n - \sum_{j=1}^{n}\prod_{i=1}^{m}\left(1 - \theta_{ij}\right) \tag{6}$$

from Equations (4) and (5) where $\boldsymbol{\Theta} = \left(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n\right)$ is a $m \times n$ matrix containing all of the unknown parameters.

In order to proceed, we make the simplifying assumption that $\theta_{ij} = \theta$ for $i = 1,2,\ldots,m$ and $j = 1,2,\ldots,n$. This assumption may be relaxed to allow for players with differing degrees of knowledge or ability and questions or tasks with differing degrees of difficulty. However, we are not proposing to handicap differing abilities of players in this paper, but merely to handicap larger sized teams for which the simplifying assumption is a reasonable first-order approximation. Under this assumption of common θ, it follows that $\phi_j = \phi$ for $j = 1,2,\ldots,n$ and

$$\phi = E(Y_j|\theta) = 1-(1-\theta)^m. \tag{7}$$

Furthermore, from Equations (5) and (7) we have that

$$Z|\phi \sim Bi(n,\phi) \tag{8}$$

is a binomial random variable with expected value

$$E(Z|\theta) = n\{1-(1-\theta)^m\}. \tag{9}$$

Using the method of moments to estimate the parameter θ, Equation (9) gives

$$n\{1-(1-\hat{\theta})^m\} = z \tag{10}$$

where z is the observed value of the random variable Z here. This is equivalent to the method of maximum likelihood for the binomial distribution and provides an estimate for the probability that any specified player knows the answer to any particular question. For the OR48 bar quiz with $n = 60$ questions and a winning team of size $m = 10$ players that scored $44\frac{1}{2}$ points (correct answers), Equation (10) gives

$$60\{1-(1-\hat{\theta})^{10}\} = 44\tfrac{1}{2} \Rightarrow \hat{\theta} = 0\cdot1266. \tag{11}$$

Having estimated the unknown parameter in the model, we can now use this estimate to determine the expected score for the authors' team of $m = 7$ players. Equation (9) then gives

$$E(Z|\hat{\theta}) = 60\{1-(1-\hat{\theta})^7\} = 36\cdot74, \tag{12}$$

This is the expected score under the assumption of equal ability θ, and thus the authors' team was disadvantaged to the extent of approximately eight points by virtue of having fewer players. As the actual points score of the authors' team (40) exceeds the expectation of a team of seven players, we rightly claim a moral victory. Of course, one could argue that the rules of the quiz were clear in advance and that building teams of many players is part of a winning strategy. However, a suitable system of handicapping would be useful for situations where this is not the case.

If these two teams of seven and ten players were of equal ability, the difference between the numbers of questions that the teams answer correctly is the difference D between two binomial random variables, whose distribution can be determined specifically for any given value of θ. For now, we simply calculate the mean difference

$$E(D|\theta) = 60\left\{(1-\theta)^7 - (1-\theta)^{10}\right\} \tag{13}$$

from Equation (9). Figure 1 illustrates a graph of this function, clearly demonstrating equality between the teams when all players are completely ignorant, $\theta = 0$, or perfectly informed, $\theta = 1$.

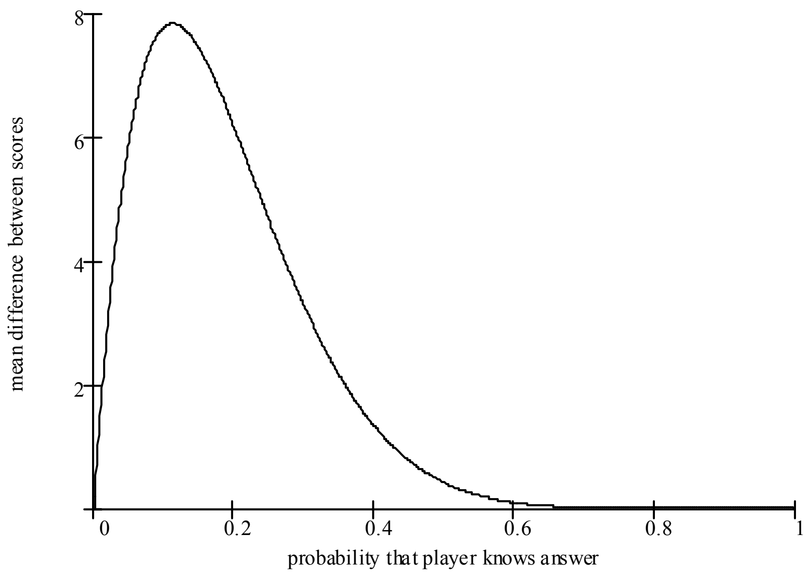

Figure 1: graph of $E(D|\theta)$ against θ for bar quiz teams of 7 and 10 players.

At $\theta = 0 \cdot 1266$, the mean difference $E(D|\theta)$ is equal to $7 \cdot 76$, which confirms our observations leading to Equation (12) and suggests that we can expect, by chance alone, the smaller team to answer eight fewer questions than the larger team. Notice in Figure 1 that the mean difference is considerable when the quiz comprises of difficult questions ($\theta \le 0.25$). The function in Equation (13) also has a unique maximum, which we now determine by calculus. Differentiating with respect to θ gives

$$\frac{d}{d\theta} E(D|\theta) = \frac{d}{d\theta} 60\left\{(1-\theta)^7 - (1-\theta)^{10}\right\} = 60\left\{10(1-\theta)^9 - 7(1-\theta)^6\right\} \tag{14}$$

and we equate this to zero, giving the minimum at $\theta = 1$ and the maximum at

$$10(1-\theta)^3 = 7 \Rightarrow \theta = 0 \cdot 1121, \tag{15}$$

which is remarkably close to our observed estimate $\hat{\theta} = 0 \cdot 1266$ and for which $E(D|\theta) = 7 \cdot 83$ or about eight correct answers again. For the contrast of an easier quiz, where each player has an equal chance of knowing the correct answer to any particular question, corresponding to $\theta = ½$, the expected difference in total scores between teams of seven and ten players is only $E(D|\theta) = 0 \cdot 41$.

Proposed handicapping system

Bar quiz organisers might wish to impose a handicapping system in advance, so that teams with differing numbers of players can compete fairly. For a quiz with n questions, Equation (9) suggests that a team of m players will score $n\left\{1-(1-\theta)^m\right\}$ marks on average. We now propose to award a handicap equal to the expected number of wrong or null answers that a team scores, defined by

$$h_{m,n}(\theta) = n - E(Z|\theta) = n(1-\theta)^m. \tag{16}$$

The corresponding handicap could be added to a team's score before or after the quiz. The advantage of waiting until after the quiz is that much information can be gained about the unknown parameter θ during the course of the quiz, allowing one to make better adjustments than if handicaps are added before the quiz begins. The disadvantage is that the necessary delay for calculations after completion of the quiz may attract accusations of unfair practice. We consider these possibilities further in a later section.

The handicapping rule above has some interesting properties. Firstly, the adjusted score of any team will be n points in expectation, with a minimum of $h_{m,n}(\theta)$ and a maximum of $n + h_{m,n}(\theta)$. Intriguingly, a team of size zero would receive a handicap of n points, which suggests that a non-existent team could win any particular quiz! However, we resolve this paradox by only allowing a team to compete if it contains a natural number of players. At the other extreme, a team with an infinite number of players would receive a handicap of 0 points, which is a natural proposition. Figure 2 displays the handicapping function for several team sizes between these two extremes with a quiz of $n = 60$ questions for illustration.

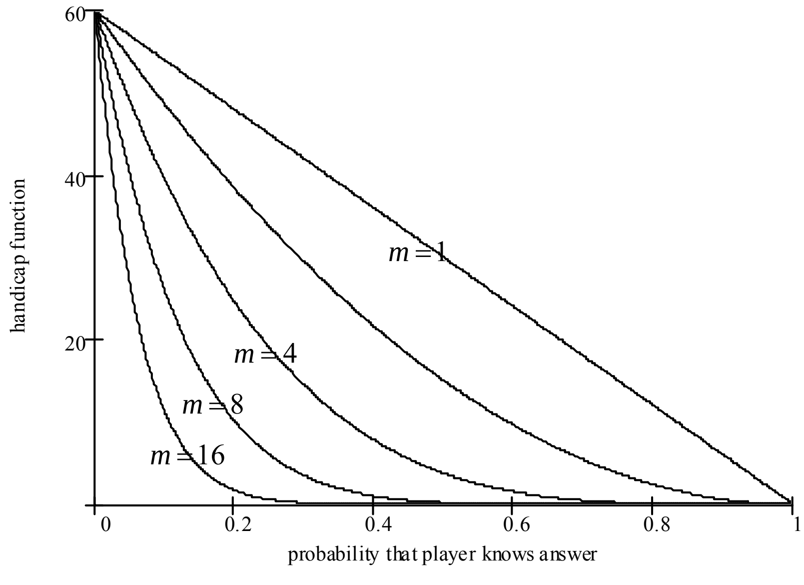

Figure 2: graph of handicap function, $h_{m,60}(\theta)$, against probability that a player knows the answer to a question, θ, for bar quiz teams of various sizes.

For the bar quiz at the Operational Research Society conference (OR48), the handicap for the authors' team of 7 players would have been

$$h_{7,60}(\theta) = 60(1-\theta)^7, \tag{17}$$

whereas the handicap for the winning team of 10 players would have been

$$h_{10,60}(\theta) = 60(1-\theta)^{10}. \tag{18}$$

Using our simple empirical estimate $\hat{\theta} = 0 \cdot 1266$, these handicaps become $h_{7,60}(\hat{\theta}) = 23 \cdot 26$ and $h_{10,60}(\hat{\theta}) = 15 \cdot 50$ respectively. The difference is the familiar eight marks that we remarked on earlier. On adding these handicaps to the actual marks scored, the adjusted scores become $40 + 23 \cdot 26 \approx 63$ for the authors' team and $44\frac{1}{2} + 15 \cdot 50 \approx 60$ for the *winning* team. Note that the latter adjusted score is necessarily $n = 60$, as we based our estimate of θ purely upon that team's performance. We consider how to improve upon this estimation procedure and develop better analytical methods in the following sections.

Note also that there would be no need for quizmasters to take calculators along with them; simple reference graphs or tables are easily constructed for routine use, such as that presented in Table 1. For our illustration, the nearest tabulated probability value to the above estimate is $\theta = 0 \cdot 15$, for which the difference in handicap allowances per question is $0 \cdot 321 - 0 \cdot 197 = 0 \cdot 124$, which becomes $7 \cdot 44$ when multiplied by $n = 60$ questions. This agrees reasonably well with the more accurate value of $7 \cdot 76$ calculated above. Clearly, the accuracy of tabulation could be improved by producing more extensive tables that consider other values of θ. Moreover, such tables could be published for particular numbers of questions n in order to avoid the need for scaling these handicaps per question.

	$h_{m,1}(\theta)$	θ 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
	1	0.950	0.900	0.850	0.800	0.750	0.700	0.650	0.600	0.550	0.500
	2	0.903	0.810	0.723	0.640	0.563	0.490	0.422	0.360	0.303	0.250
	3	0.857	0.729	0.614	0.512	0.422	0.343	0.275	0.216	0.166	0.125
	4	0.815	0.656	0.522	0.410	0.316	0.240	0.179	0.130	0.092	0.063
m	5	0.774	0.590	0.444	0.328	0.237	0.168	0.116	0.078	0.050	0.031
	6	0.735	0.531	0.377	0.262	0.178	0.118	0.075	0.047	0.028	0.016
	7	0.698	0.478	0.321	0.210	0.133	0.082	0.049	0.028	0.015	0.008
	8	0.663	0.430	0.272	0.168	0.100	0.058	0.032	0.017	0.008	0.004
	9	0.630	0.387	0.232	0.134	0.075	0.040	0.021	0.010	0.005	0.002
	10	0.599	0.349	0.197	0.107	0.056	0.028	0.013	0.006	0.003	0.001

Table 1: handicapping allowance per question $h_{m,1}(\theta)$ for various team sizes m and probability values θ.

Empirical analysis

Earlier in this paper, we estimated the unknown parameter θ by using the post-quiz score of the winning team and the method of moments on Equation (9). As the underlying assumption is that all players are equally knowledgeable, a better estimator would take into account the post-quiz scores of all the teams. Suppose that t teams enter the quiz and that team k has m_k players and correctly answers z_k questions, for $k = 1,2,\ldots,t$. As the scores achieved by each team are mutually independent, we can formulate a likelihood function as a product of binomial probability mass functions from Relation (8) as

$$L(\theta;\mathbf{m},n,\mathbf{z}) \propto \prod_{k=1}^{t} \left\{1-(1-\theta)^{m_k}\right\}^{z_k} \left\{(1-\theta)^{m_k}\right\}^{n-z_k} \tag{19}$$

where $\mathbf{m} = (m_1, m_2, \ldots, m_t)'$ and $\mathbf{z} = (z_1, z_2, \ldots, z_t)'$ are column vectors containing the numbers of players and scores for all teams. To avoid introducing further notational complexity and without loss of generalization, the remainder of this article defines the likelihood by equality rather than proportionality in Relation (19). Taking logarithms of this expression, differentiating with respect to θ and equating to zero gives the likelihood equation

$$\frac{d}{d\theta}\log L(\theta;\mathbf{m},n,\mathbf{z}) = 0 \Rightarrow \sum_{k=1}^{t} \frac{m_k z_k}{1-(1-\theta)^{m_k}} = n\sum_{k=1}^{t} m_k \tag{20}$$

whose solution $\hat{\theta}$ is the maximum likelihood estimator of θ. This needs to be evaluated numerically in practice. As a check, this agrees with Equation (10) for the special case when $t = 1$.

For the OR48 bar quiz, we can illustrate the use of Equation (20) for $t = 2$ based upon the scores of the winning team, for which $m_1 = 10$ and $z_1 = 44\frac{1}{2}$, and the authors' team, for which $m_2 = 7$ and $z_2 = 40$, and with $n = 60$ as before; the scores of other teams in the competition were unfortunately not recorded. From above, the likelihood equation is then

$$\frac{10 \times 44\frac{1}{2}}{1-(1-\theta)^{10}} + \frac{7 \times 40}{1-(1-\theta)^{7}} = 60 \times (10 + 7) \tag{21}$$

and numerical solution of this nonlinear equation using Mathcad gives the maximum likelihood estimate $\hat{\theta} = 0 \cdot 1351$ with subsequent handicaps $h_{7,60}(\hat{\theta}) = 21 \cdot 72$ and $h_{10,60}(\hat{\theta}) = 14 \cdot 05$. These handicaps have changed considerably, emphasising their sensitivity to the actual scores achieved during the quiz, although the handicap difference is little influenced. Observe that this estimate of θ is larger than the estimate based only upon the winning team's score, again reflecting the fact that the authors' team had a greater ability to answer questions correctly on the night. Indeed, this suggests an alternative but equivalent method of handicapping such quizzes. Instead of deriving an adjusted team score as recommended above, one could estimate the average number of questions that each team member scored within each team, as $n\hat{\theta}$ where $\hat{\theta}$ is the single-team estimate for θ given by

Equation (10). For the winning team this gives $60 \times 0 \cdot 1266 = 7 \cdot 6$ points, whereas for the authors' team this gives $60 \times 0 \cdot 1452 = 8 \cdot 7$ points and the conclusion is as before. The revealing truth that a team that correctly answers two thirds of the questions can comprise players who individually are only able to answer one question in eight suggests that the original handicapping proposal in Equation (16) would be received better by participants. We now return to considering the measurement of θ.

Subjective analysis

We previously explained how to calculate the maximum likelihood estimate of θ based upon all of the scores achieved during operation of the quiz. As mentioned in an earlier section, we would ideally prefer to declare the handicap adjustments proposed in Equation (16) before the quiz begins. This would allow full transparency of the handicapping procedure and avoid the time needed for calculations after marking the answers. Clearly, the simplest way of achieving this is for the quizmaster to assess the true value of θ in advance, based on his or her knowledge and experience. Although this approach is reasonable, all previous calculations assume that this assessment is the true value of θ. At best, any such calculations can only be approximations and may indeed be poor. We could perform a sensitivity analysis to measure the errors involved. However, a subjective Bayesian analysis automatically takes account of the inaccuracy attached to such an assessment for θ and so we investigate this approach now.

Instead of assessing a value for θ in advance, the quizmaster might specify a prior distribution for θ. This is easier to do than it seems. As the unknown Bernoulli parameter θ represents a probability, the natural conjugate prior distribution is the beta family. The use of natural conjugate priors is highly recommended; see Bernardo and Smith (1993) for example. The corresponding probability density function is

$$g(\theta) = \frac{1}{\mathrm{B}(a,b)} \theta^{a-1} (1-\theta)^{b-1} \; ; \quad 0 < \theta < 1 \tag{22}$$

in terms of the beta function $\mathrm{B}(a,b)$ and hyperparameters a and b that are chosen to reflect our prior knowledge accurately. Percy (2003) described and illustrated a suitable method for eliciting these hyperparameters. In the present context, the quizmaster must use his or her knowledge and experience to specify two values of θ corresponding to the tertiles, L and U, of this prior distribution. These divide the range $0<\theta<1$ into three equally likely intervals for θ, so that

$$P(0<\theta<L)=P(L<\theta<U)=P(U<\theta<1)=\frac{1}{3}. \tag{23}$$

This leads to two, simultaneous nonlinear equations to solve for the hyperparameters.

Having specified this prior distribution for θ, we revise our handicapping function in Equation (16) into a criterion that does not depend on knowing the value of θ. We achieve this by taking a prior expectation of this function, which gives

$$\begin{aligned} h_{m,n} &= E_\theta\left\{n(1-\theta)^m\right\}=\int_0^1\left\{n(1-\theta)^m\right\}\left\{\frac{1}{\mathrm{B}(a,b)}\theta^{a-1}(1-\theta)^{b-1}\right\}d\theta \\ &= \frac{n}{\mathrm{B}(a,b)}\int_0^1\theta^{a-1}(1-\theta)^{m+b-1}d\theta=\frac{n\mathrm{B}(a,m+b)}{\mathrm{B}(a,b)}=n\prod_{l=0}^{m-1}\frac{b+l}{a+b+l}\end{aligned} \tag{24}$$

for $m\in\mathbb{N}$. This handicapping system no longer depends on the unknown parameter and we can implement it before the quiz begins.

For illustration, reconsider the OR48 bar quiz. Suppose that, before the quiz, the quizmaster believed that a third of all players would be able to answer at most five of the sixty questions and that a third of all players would be able to answer at least ten of the sixty questions. With appropriate continuity correction factors, this implies that

$$L=\frac{5\tfrac{1}{2}}{60}=0\cdot0917 \quad\text{and}\quad U=\frac{9\tfrac{1}{2}}{60}=0\cdot1583. \tag{25}$$

to four decimal places. Solving Equations (23) numerically then gives the hyperparameter values

$$a = 2 \cdot 454 \quad \text{and} \quad b = 15 \cdot 62 \tag{26}$$

to four significant figures. Figure 3 displays the corresponding prior probability density function for θ, which reflects the quizmaster's expert judgements.

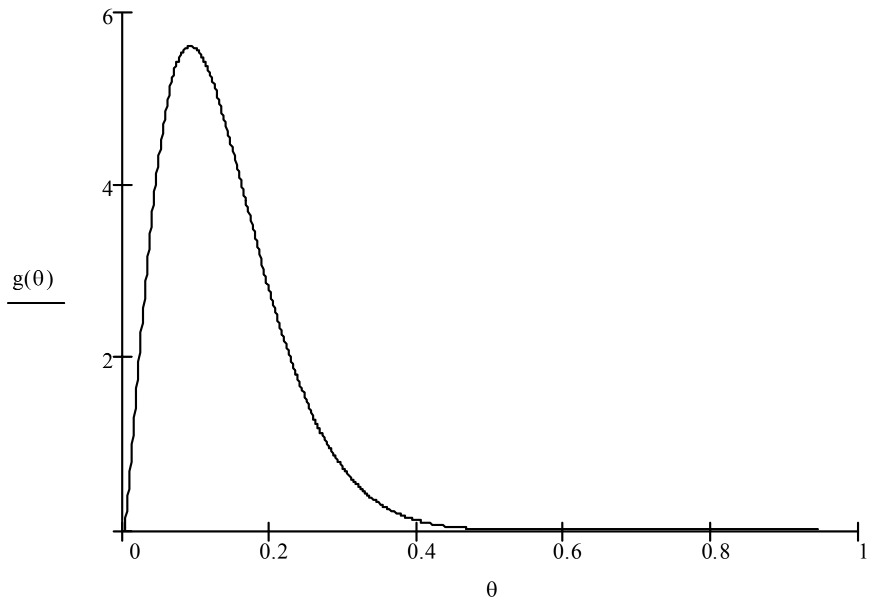

Figure 3: quizmaster's prior probability density function $g(\theta)$ for the unknown parameter θ.

We are now in a position to determine the handicapping criterion of Equation (24) for this illustration, which becomes

$$h_{m,60} = 60 \prod_{l=0}^{m-1} \frac{15 \cdot 62 + l}{18 \cdot 074 + l} \tag{27}$$

and which we illustrate in Figure 4. With this prior subjective measurement of θ, teams of seven and ten players would receive handicaps of $h_{7,60} = 25 \cdot 00$ and $h_{10,60} = 18 \cdot 58$ respectively. These compare favourably with the corresponding handicaps based only on empirical measurement of θ, which we calculated as $h_{7,60}(\hat{\theta}) = 21 \cdot 72$ and $h_{10,60}(\hat{\theta}) = 14 \cdot 05$ from Equations (17) and (18) respectively, using the estimate obtained previously.

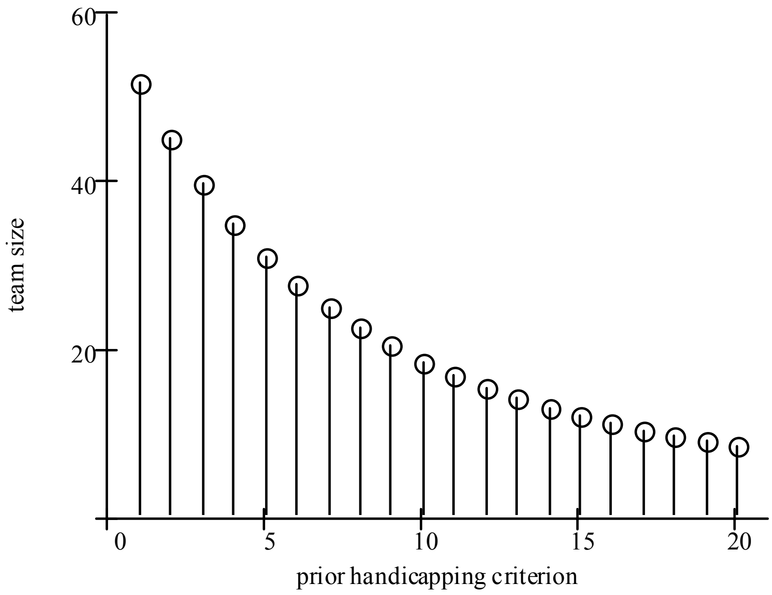

Figure 4: prior handicapping criterion $h_{m,60}$ for different team sizes m.

Two handicapping schemes have now been established; they differ only in the approach to measuring the unknown parameter θ. The first is a frequentist approach and relies upon waiting until after the quiz is complete in order to calculate a point estimate for θ, so obtaining approximate solutions. The second is a subjective Bayesian approach that we can perform before the quiz starts and which enables us to allow for the uncertainty in θ, so producing accurate solutions. For transparency, accuracy and efficiency, we prefer the second of these approaches. However, there exists a third possibility that we briefly consider now. It extends the subjective Bayesian analysis to include the quiz results. As such, this third approach relies upon waiting until after the quiz is complete like the frequentist approach, but improves upon the latter by using all available information about theta.

In passing, it is interesting to consider whether specifying a prior distribution for the unknown parameter θ is algebraically equivalent to assuming a random effects model in Equations (1) to (5), effectively relaxing the assumption that $\theta_{ij} = \theta$ for $i = 1,2,\ldots,m$ and $j = 1,2,\ldots,n$. In this case, we could suppose that $\theta_{ij} = \theta_j$ where the θ_j are independent and identically distributed with mean θ (random questions and equal players) or that the θ_{ij} are independent and identically distributed with mean θ (random questions and random players). Although such equivalence between prior distributions and random effects modelling is apparent for the general linear model, this is not the case here because the corresponding handicapping rules from Equations (16) and (6) are

$$h_{m,n}(\boldsymbol{\theta}) = n - E(Z|\boldsymbol{\theta}) = \sum_{j=1}^{n} (1-\theta_j)^m \ , \tag{28}$$

and

$$h_{m,n}(\boldsymbol{\Theta}) = n - E(Z|\boldsymbol{\Theta}) = \sum_{j=1}^{n} \prod_{i=1}^{m} (1-\theta_{ij}), \tag{29}$$

for these two cases, which do not simplify to the prior handicapping criterion $h_{m,n}$ in Equation (24) when a standard distribution (uniform, triangular or beta) is introduced for the θ_j or θ_{ij}, respectively.

Combined subjective and empirical analysis

The extension to the subjective analysis of the preceding section is simple: rather than using the prior distribution to calculate a handicapping criterion based on the expected handicapping function, we use the posterior distribution for the same purpose. We determine the posterior distribution (see Bernardo and Smith, 1993, for example) as

$$g(\theta|\mathbf{m}, n, \mathbf{z}) \propto L(\theta; \mathbf{m}, n, \mathbf{z}) g(\theta); \quad 0 < \theta < 1 \tag{30}$$

using the expressions in Relation (19) and Equation (22). We then replace the prior density in Equation (24) by the posterior density, to define the posterior handicapping criterion

$$h_{m,n}(\mathbf{m}, n, \mathbf{z}) = E_\theta \{n(1-\theta)^m | \mathbf{m}, n, \mathbf{z}\} = \int_0^1 \{n(1-\theta)^m\} g(\theta|\mathbf{m}, n, \mathbf{z}) d\theta . \tag{31}$$

Unfortunately, we cannot evaluate the constant of proportionality in Relation (30) algebraically. Moreover, we cannot integrate the function of θ in the integrand of (31) algebraically. Tierney and Kadane (1986) developed an accurate analytical approximation for this scenario, based upon truncated series expansions. However, both integrals are one-dimensional and so simple numerical quadrature is accurate and efficient for this problem.

We illustrate this best method using the OR48 scores obtained by the authors' team and the winning team, as for the empirical analysis in the previous section. With $m_1 = 10$, $z_1 = 44\frac{1}{2}$, $m_2 = 7$ $z_2 = 40$, $t = 2$, $n = 60$ and the hyperparameters a and b as defined in Equations (26), we first evaluate the constant of proportionality in Relation (30). Mathcad uses the Romberg method for numerical integration and gives the result

$$\int_0^1 L(\theta;\mathbf{m},n,\mathbf{z})g(\theta)d\theta = 4\cdot 747\times 10^{-33} \tag{32}$$

to four significant figures. Hence, the posterior probability density function for this application is given by

$$g(\theta|\mathbf{m},n,\mathbf{z}) = \frac{L(\theta;\mathbf{m},n,\mathbf{z})g(\theta)}{4\cdot 747\times 10^{-33}}; \quad 0 < \theta < 1 \tag{33}$$

from Relation (30). We can now evaluate numerically the posterior handicapping criterion in Equation (31) for different team sizes m, giving the graph in Figure 5.

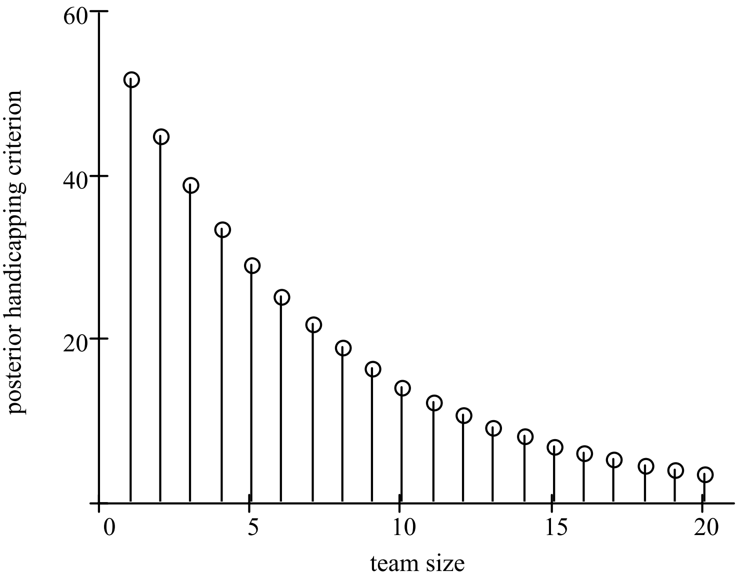

Figure 5: posterior handicapping criterion $h_{m,60}(\mathbf{m}, n, \mathbf{z})$ for different team sizes m.

With this combined subjective and empirical measurement of θ for the conference bar quiz, teams of seven and ten players would receive posterior handicaps of $h_{7,60}(\mathbf{m}, n, \mathbf{z}) = 21 \cdot 81$ and $h_{10,60}(\mathbf{m}, n, \mathbf{z}) = 14 \cdot 19$ respectively. These demonstrate the updating effect of the actual quiz scores observed upon the prior handicaps of $h_{7,60} = 25 \cdot 00$ and $h_{10,60} = 18 \cdot 58$ from the previous section. Again, the proposed handicaps are comparable to those based only on empirical measurement of θ, which we previously calculated as $h_{7,60}(\hat{\theta}) = 21 \cdot 72$ and $h_{10,60}(\hat{\theta}) = 14 \cdot 05$.

Multiple-choice questions and progressive quizzes

The handicapping rules derived earlier can be extended to cope with multiple-choice questions. Assuming all things equal, define θ to be the probability that a random player knows the answer to a random question, as in previous sections, and suppose that each question consists of $p = 2,3,4,\ldots$ possible answers. We also suppose that if any contestant in a team knows the answer, the team chooses this answer; otherwise, it chooses one of the p alternatives at random. Using the law of total probability, the probability that the team gives the correct answer to a question is

$$\phi = 1 \times \left\{1-(1-\theta)^m\right\} + \frac{1}{p} \times (1-\theta)^m = 1 + \left(\frac{1}{p} - 1\right)(1-\theta)^m \tag{34}$$

from Equation (7). We can now define a suitable handicapping function, equivalent to that in Equation (16) but adapted for multiple-choice quizzes with p possible answers to each question, by

$$h'_{m,n}(\theta) = n(1-\phi) = n\left(1 - \frac{1}{p}\right)(1-\theta)^m, \tag{35}$$

which includes an extra, multiplicative factor for the multiple-choice answers. As expected, this factor tends to one as the number of possible answers tends to infinity.

From Equations (24) and (31), it is clear that this same scaling factor applies to all three of our handicapping criteria, whereby

$$h'_{m,n}(\hat{\theta}) = \left(1 - \frac{1}{p}\right) h_{m,n}(\hat{\theta}), \tag{36}$$

$$h'_{m,n} = \left(1 - \frac{1}{p}\right) h_{m,n} \tag{37}$$

and

$$h'_{m,n}(\mathbf{m}, n, \mathbf{z}) = \left(1 - \frac{1}{p}\right) h_{m,n}(\mathbf{m}, n, \mathbf{z}). \tag{38}$$

To illustrate the effect of this scaling factor, suppose that the OR48 bar quiz with $n = 60$ consisted of questions involving $p = 3$ possible multiple-choice answers and suppose that we were interested in allocating handicaps before the quiz began. Figure 6 illustrates the prior handicapping criterion adjusted for this multiple-choice setting, for comparison with the corresponding, unadjusted criterion in Figure 4. The scaling factor in this case is ⅔ and the handicaps for our teams of seven and ten players would have been $16 \cdot 67$ and $12 \cdot 39$ respectively, which we could round off to the nearest integer for general ease of interpretation, if desired.

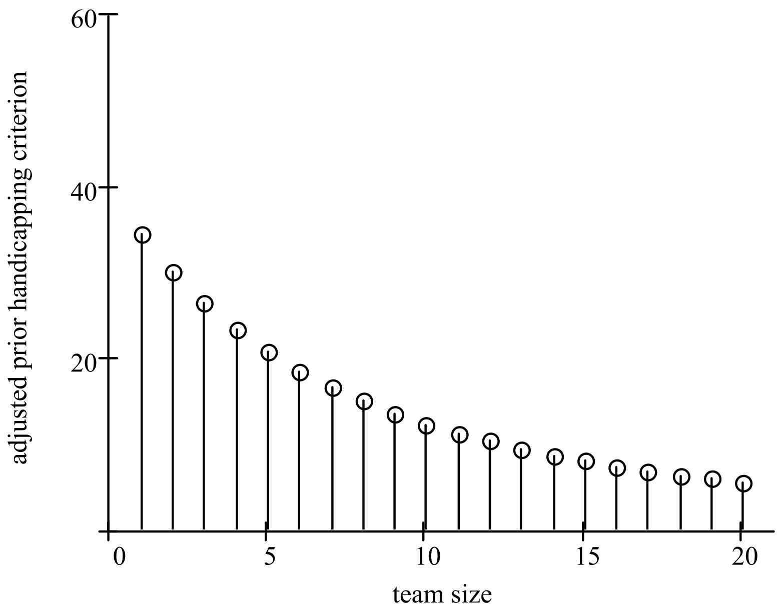

Figure 6: adjusted prior handicapping criterion $h'_{m,60}$ for different team sizes m.

We now consider how to extend the preceding analyses to develop handicapping procedures for progressive quizzes. These are characterised by the property that when a team gives a wrong answer, it is eliminated from the competition. Variations of this scenario are familiar in sporting knockout contests and in some television game shows. The problem with using the handicapping rules proposed earlier for progressive quizzes is that different teams answer different numbers of questions, so the parameter n is variable. The solution is to set $n = 1$ and reset the handicap for each round. However, a correct answer scores one point whereas a wrong answer scores zero points. With $n = 1$, all three of our handicapping criteria lie in the interval $(0,1)$. Consequently, the handicapping rule for progressive quizzes reduces to the following simple statements. In any round, eliminate the largest team that fails to answer the question correctly. If all teams answer the question correctly, eliminate the largest team. Although this is a fair handicapping rule, it seems to penalize the largest team rather harshly if all teams answer correctly. In this case, a reasonable alternative is to ask another question so that a team can only be eliminated if it answered the question incorrectly.

Conclusions and extensions

The work presented in this paper was motivated by the authors' experience at the 2006 Operational Research Society annual conference (OR48). Primarily this paper offers handicapping rules to permit fair competition for tournaments involving teams with differing numbers of players. The development of these rules has some elements in common with the handicapping procedures proposed by Lewis (2005) for team competitions in golf. The theory may also have some valuable applications in other sporting situations. For example, in tournament qualifying competitions, such as the European football championship, unequal group sizes are used; should handicapping be used here to determine qualifiers since it will be more difficult to qualify from a larger group if only a fixed number from each group qualify? Also the theory might be used to rank nations' achievements in international events, since players in national teams will be selected from pools of potential players of differing sizes. This would offer an alternative method to those that have been considered for ranking nations' achievement in the Olympic Games (e.g. Morton, 2002; Churilov and Fitman, 2006).

Under the assumptions of random questions and random players, we propose three rules for handicapping systems that apply to informal quizzes, so that teams with differing numbers of players may compete fairly against each other. The first rule is a frequentist procedure, which offers an approximation since the unknown parameter is estimated using observed quiz results. The second rule is a subjective Bayesian procedure that incorporates *a priori* information about the unknown parameter and can be determined before the quiz begins. The third of these rules, $h_{m,n}(\mathbf{m}, n, \mathbf{z})$ in Equation (31), is mathematically the most satisfactory as it includes all available knowledge about the unknown parameter θ, the probability of success for an individual player in a single question. However, the second of the rules, $h_{m,n}$ in Equation (24), is more suitable if one wishes to set handicaps before the quiz begins. It is also worth noting that $h_{m,n}$ is easy to evaluate on a portable calculator, whereas $h_{m,n}(\mathbf{m}, n, \mathbf{z})$ requires a greater level of numerical sophistication.

The handicapping rules are extended to quizzes that comprise multiple-choice questions and answers. A simple multiplicative scaling factor adjustment is all that is required. Again, the subjective Bayesian criterion, $h'_{m,n}$ in Equation (37) is straightforward to compute and is available before the quiz begins. The rules $h_{m,n}$ and $h'_{m,n}$ could be made available in tables for a range of quiz sizes, team sizes and hyperparameters for θ, perhaps accompanied graphically by a nomogram. Finally, we have demonstrated that for progressive quizzes all of the handicapping criteria for progressive quizzes amount to two simple rules.

Acknowledgements

The authors wish to thank the bar quiz organiser, Ed Rochead (DSTL), team-mates and opponents at OR48 for inspiring this research.

References

Bernardo JM, Smith AFM (1993) *Bayesian Theory*. New York: Wiley.

Bolton RN, Chapman RG (2003) Searching for positive returns at the track: a multinomial logit model for handicapping horseraces. *Management Science* 32, 1040-60.

Churilov L, Flitman A (2006) Towards fair ranking of Olympic achievements: the case of Sydney 2000. *Computers and OR* 33, 2057-2082.

Cleather DJ (2006) Adjusting powerlifting performances for differences in body mass. *J Strength and Conditioning Research* 20, 412-421.

Edelman D (203) A competitive horse-race handicapping algorithm based on analysis of covariance. In *The Economics of Gambling*, Vaughan Williams (ed.), Routledge, London, pp.106-113.

Lewis AJ (2005) Handicapping in group and extended golf competitions. *IMA J Management Math* 16, 151-160

Morton RH (2002) Who won the Sydney Olympics? an allometric approach. *J Royal Statist Soc Series D* 51, 147-155.

Percy DF (2003) Subjective reliability analysis using predictive elicitation. In *Mathematical and Statistical Methods in Reliability* (eds. Lindqvist, BH and Doksum, KA), Singapore: World Scientific, 57-72.

Szymanski S (2003) The economic design of sporting contests. *Journal of Economic Literature* 41, 1137-87.

Tierney L, Kadane JB (1986) Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, **81**, 82-86.