The Annals of Mathematical Statistics
1968, Vol. 39, No. 2, 583–592

INFERENCE FROM A KNOCKOUT TOURNAMENT[1]

By J. A. Hartigan

Princeton University

0. Summary. A parametric probability model for order data on a set of objects is developed, which is appropriate for analysis of data with tree structure; the results of knockout tournaments are an example of such data. The parameters of the model are the probabilities that the various objects will be ranked highest. An exact Bayes analysis is possible if the prior distribution of the parameters is Dirichlet. The 1965 Wimbledon tennis tournment is used as an example.

1. Introduction. We will consider the analysis of data consisting of ordering relations on a set of objects. An example of such data appears in psychology, when a subject is asked to declare his preferences among a set of objects. In paired comparisons data, the preference judgments on pairs of a set of objects are made independently over pairs; a number of judgments is made for each pair, perhaps by different subjects, or by the same subject at different times, and this sequence of judgments may be then regarded as a sequence of Bernoulli trials. Standard statistical theory is used for the estimation of probabilities of preferences; various models peculiar to paired comparisons are then available to explain and interpret these probabilities. For example Brunk (1960), David (1963), Bradley and Terry (1952).

We will be concerned with preference judgments made simultaneously over a set of objects, so that the transitivity property of ordering relations must be complied with. For example, a subject might be required to rank objects by order of preference. Complete ranking of a large number of objects is a difficult task; for this reason we wish to consider data consisting of a partial order given by a subject on a set of objects. We propose an experimental design with tree structure, in which, at each stage, the subject chooses the best out of a small group of objects (possibly two); the other objects in the group are eliminated from further comparison. Such designs used in athletic contests are known as knockout tournaments. In tree designs no testing of the transitivity property occurs; for this reason they are most desirable when the transitivity property can be reliably assumed. The outcome of a tree design is a partial order on the objects which itself has tree structure. Examples of the tree structure of the design, and the tree structure of the outcome are given in Figure 1 and Figure 2.

A partial order may be identified with that subset, of the set of all complete orders, containing all those complete orders consistent with the partial order. Thus a probability model for complete orders generates a probability model for partial orders. The probability model for complete orders used, is derived from the theory of non-parametric statistics, Savage (1956); suppose that X_1, X_2,

Received 2 August 1966; revised 29 September 1967.

[1] Research supported by Office of Naval Research Contract Nonr 1858(05).

$\cdots, X_n$ are real valued random variables with continuous distribution functions $F^{\theta_1}, F^{\theta_2}, \cdots, F^{\theta_n}$ where F is arbitrary and the θ_i are non-negative; then $P(X_1 < X_2 < \cdots < X_n) = \prod_{i=1}^{n} (\theta_i / \sum_{j=1}^{i} \theta_j)$ and the probability of other complete rankings of the X_i may be obtained similarly. A specific interpretation of the θ_i is that $\theta_i / \sum \theta_j$ is the probability that the ith object is ranked highest out of the n objects. This distribution is used in non-parametric testing of the hypothesis that some or all the X_i have identical distributions, and provides an evaluation of power if the alternative distributions have the form F^{θ_i} for some F, θ_i. As a parametric probability model for ranks, the model is consistent with the Bradley-Terry (1952) model for paired comparisons, which states that P[object i is ranked less than object j] $= \theta_j/(\theta_i + \theta_j)$, and the Luce (1963), p. 217, Axiom of choice: if i_A denotes the highest ranked object in the set A, the axiom of choice requires $P[i_A = i \mid i_A \varepsilon B] = P(i_B = i)$, where $B \subset A$. In words, the probability that i is the highest ranked object in A, given that the highest ranked object is in B, is equal to the probability that i is the highest ranked object in B.

The probability of a partial order given $\boldsymbol{\theta}$ may be formally evaluated by summing over complete orders compatible with the partial order. For partial orders with tree structure, this summation reduces to a simple closed expression, making explicit analysis of the likelihood possible. For example, if a sequence of independent tree outcomes is available, maximum likelihood could be used to estimate the θ_i. We have considered a Bayesian analysis for a single tree, in which the prior information about the θ_i is formalised as a Dirichlet variable with density $\prod \theta_i^{\alpha_i - 1}$ over the simplex $\theta_i \geqq 0$, $\sum \theta_i = 1$. This prior information is experimentally equivalent to observing that i was ranked highest in α_i out of $\sum \alpha_i$ trials. If $\sum \alpha_i$ is kept small ($\sum \alpha_i < 1$, say) the observed tree will play the main role in determining the posterior distribution of the θ_i; the means, variances, covariances, of the posterior distribution may be explicitly computed, for use in estimating θ_i.

The computations are carried through, for various priors, for the 1965 Wimbledon tennis tournament.

2. A probability model for order data. Given a set of objects $A = (1, 2, \ldots, N)$ a *partial ordering* ρ on A consists of a relation $i < j$ on some pairs (i, j) which is (a) *irreflexive* so that $i < j, j < i$ cannot occur and (b) *transitive* so that $i < j$, $j < k$ implies $i < k$. It will be sometimes convenient to denote the relation $i < j$ associated with ρ by $\rho(i) < \rho(j)$. A complete ordering or *rank* σ on A is a partial ordering such that for every pair (i, j), $i \neq j$ we have either $i < j$ or $j < i$; the rank of i in σ is $\sigma(i)$ if exactly $\sigma(i) - 1$ of the objects in A are less than i. We will use $i = j$, $\rho(i) \leqq \rho(j)$ to denote: $i < j$ or $i = j$. Let S be the set of all $N!$ ranks of A; each partial ordering ρ may be identified with a subset of S, that subset consisting of ranks σ which obey all the relations in ρ. For this reason, S is the natural sample space for analysis of a partial order.

The probability model is

$$P[\sigma \mid \theta] = \prod_{i=1}^{N} [\theta_i / \sum_{\sigma(j) \leqq \sigma(i)} \theta_j], \quad \sum_{i=1}^{N} \theta_i = 1, \ \theta_i \geqq 0. \tag{1}$$

The parameter θ_i is the probability that the ith object is ranked highest. It has been remarked that the model is consistent with the Bradley-Terry model for paired comparisons, and the Luce axiom of choice. To prove this we consider generation of the above model by associating with the family of objects a family of exponential variables $\{Y_i\}$ with means $\{1/\theta_i\}$. Then $P[Y_{\sigma(i)} > Y_{\sigma(j)}$ whenever $\sigma(i) < \sigma(j)] = P[\sigma \mid \theta]$; to prove this we develop (1) from the highest ranked object towards the lowest, using the fact that if Y is exponential with mean 1, $Y - a \mid Y > a$ is exponential with mean 1. We now see that $P[i$ is highest ranked$] = P[Y_i = \inf(Y_i, \cdots, Y_n)] = \theta_i$ as claimed above. We see that $P[i < j] = P(Y_i > Y_j) = \theta_j/(\theta_i + \theta_j)$ agreeing with the Bradley-Terry model. Finally, if i_A denotes the highest ranked object in the set A, we have $P[i_B = i] = P[Y_i = \inf_{j \varepsilon B} Y_j] = \theta_i / \sum_{j \varepsilon B} \theta_j = P[i_A = i \mid i_A \varepsilon B]$, so that the model agrees with the Luce axiom of choice.

Let $A^* = \{i_1, i_2, \cdots, i_{N^*}\}$, and consider the probability distribution induced on S^*, the set of ranks of A^*, by the mapping $\sigma \to \sigma^*$, where $\sigma^*(i_r) < \sigma^*(i_s)$ if and only if $\sigma(i_r) < \sigma(i_s)$. Then

$$(2) \qquad P[\sigma^* \mid \theta] = \prod_{k=1}^{N^*} (\theta_{i_k} / \sum_{\sigma^*(i_m) \leqq \sigma^*(i_k)} \theta_{i_m}).$$

This result is obvious in the exponential formulation. It means that the probability distribution of order within A^* is of the same form, as that of order within A, and it is determined by parameters $\{\theta_{i_k}\}$ given by the elements in A^*.

Let A^*, A^{**} be disjoint sets of objects and let σ^*, σ^{**} denote two ranks on A^*, A^{**}. Then

$$(3) \qquad P[\sigma^* \text{ on } A^*, \sigma^{**} \text{ on } A^{**} \mid \theta] = P(\sigma^* \text{ on } A^* \mid \theta) = P(\sigma^{**} \text{ on } A^{**} \mid \theta).$$

In words, orders within disjoint subsets are independently distributed. This result and a similar result for any number of disjoint subsets is obvious in the exponential formulation.

For an arbitrary partial order ρ, $P[\rho \mid \theta] = \sum_{\sigma \varepsilon \rho} P[\sigma \mid \theta]$. A considerable simplification occurs if the partial order is a *tree* τ; this is a partial order for which $\tau(i) < \tau(j), \tau(i) < \tau(k)$ implies either $\tau(j) \leqq \tau(k)$ or $\tau(k) \leqq \tau(j)$. The element j such that $i < j, j \leqq k$ for all $k > i$ is the *successor* of i denoted by τi; τi is uniquely defined unless there is no element $j, i < j$. It will be convenient to introduce an element 0, the root of the tree, and set $\tau i = 0$ if there is no $j, i < j$. The partial order τ is uniquely determined by the mapping $i \to \tau i$. We set $F_i = \{j \mid \tau j = i\}$, the *family* of i. We set $U_i = \{j \mid \tau(i) \leqq \tau(j)\}, D_i = \{j \mid \tau(j) \leqq \tau(i)\}$. A characterizing property of the tree is that $D_i \subset D_j$, $D_j \subset D_i$, or $D_i \cap D_j = \emptyset$, for every i, j. For any i, τ induces the subtree τ_i on D_i defined by $\tau_i(j) < \tau_i(k)$ if and only if $j, k \varepsilon D_i$, $\tau(j) < \tau(k)$.

THEOREM 1. *The probability of a tree is given by*

$$(4) \qquad P(\tau \mid \theta) = \prod_{i=1}^{N} (\theta_i / \sum_{j \varepsilon D_i} \theta_j).$$

To prove this we use the exponential formulation stating $P(\tau \mid \theta) = P[Y_{\tau(i)} \geqq Y_{\tau(j)}$ whenever $\tau(i) \leqq \tau(j)]$. The equation holds for $N = 1$, and we will assume it holds for 2, 3, $\cdots$, $(N - 1)$ elements. Suppose that i_0 is an element such that $\tau_{i_0} = 0$; at least one such element must exist since $i < \tau i$ for $\tau i \,\varepsilon\, A$; thus the sequence $i, \tau i, \tau(\tau i), \cdots$ consists of different elements of A, and 0. If i_0 is unique, let $\{i_1, i_2, \cdots, i_n\} = F_{i_0}$. Then $\tau = \{\sigma \mid \sigma(i_0) = N\} \cap \prod_{r=1}^{n} \tau_{i_r}$ and $P(\tau \mid \theta) = P[\{Y_{i_0} = \inf Y_i\} \cap \prod \tau_{i_0} \mid \theta]$. Now given that $Y_{i_0} = \inf Y_i$, we have that the $Y_i - Y_{i_0}$ are independent exponentials for $i \neq i_0$; thus order relations among the Y_i are independent of the knowledge $Y_{i_0} = \inf Y_i$. Therefore $P(\tau \mid \theta) = (\theta_{i_0}/\sum \theta_i) \prod_{r=1}^{n} \prod_{i \varepsilon D_{i_r}} (\theta_i / \sum_{j \varepsilon D_i} \theta_j) = \prod_{i=1}^{N} (\theta_i / \sum_{j \varepsilon D_i} \theta_j)$. The equation has been extended to N elements if i_0 is unique. If i_0 is not unique, suppose that $i_1, i_2, \cdots, i_n$ are n distinct elements satisfying $\tau i_r = 0$. Then $\tau = \prod_{r=1}^{n} \tau_{i_r}$ and $P(\tau \mid \sigma) = \prod P(\tau_{i_r} \mid \sigma) = \prod_{r=1}^{n} \prod_{i \varepsilon D_{i_r}} (\theta_i / \sum_{j \varepsilon D_i} \theta_j) = \prod_{i=1}^{N} (\theta_i / \sum_{j \varepsilon D_i} \theta_j)$. The equation holds in this case also.

3. Bayesian analysis of $\boldsymbol{\theta}$. Given τ, we wish to make inferences about $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_N)$; θ_i is the probability that the ith object will be ranked first. A convenient form of prior distribution for $\boldsymbol{\theta}$ specifies $\boldsymbol{\theta}$ to be a Dirichlet variable $\mathrm{D}[\alpha_i, i = 1 \cdots, N]$, having density proportional to $\prod_{i=1}^{N} \theta_i^{\alpha_i - 1}$ over the simplex $\theta_i \geqq 0$, $\sum_{i=1}^{N} \theta_i = 1$. This prior distribution corresponds to prior knowledge of the form: the ith object was ranked first in α_i out of $\sum \alpha_i$ competitions in the past. Thus $\alpha_i / \sum \alpha_i$ is an estimate of the probability that the ith object will be ranked first, and $\sum \alpha_i$ represents the weight associated with the estimate.

THEOREM 2. *Let* $\beta_i = \sum_{j \varepsilon D_i} \alpha_j$. *Let* $\mathbf{Z}_i = [\{W_j, j \,\varepsilon\, F_i\}, W_i^*]$ *be independent Dirichlet variables with parameters* $[\{\beta_j, j \,\varepsilon\, F_i\}, \alpha_i + 1]$, $i = 0, 1, 2, \cdots, N$. (*Let* $W_0^* = 0$). *Then the posterior distribution of* θ *given* τ *is given by* $\theta_i = W_i^* \prod_{j \varepsilon U_i} W_j$.

PROOF. We may write $\phi_i = \sum_{j \varepsilon U_i} \theta_j$, $W_i = \phi_i / \phi_{\tau i}$ so that W_i is the probability that an object in D_i is ranked first, given that an object in $D_{\tau i}$ is ranked first. In this Bayesian approach, we are regarding this probability W_i as a random variable, whose distribution changes in the light of the data τ. The prior distribution on $\boldsymbol{\theta}$ induces the distribution on $\{W_i, i = 1, \cdots, N\}$ in which $[\{W_j, j \,\varepsilon\, F_i\}, W_i^*]$ are independent Dirichlet variables with parameters $[\{\beta_j, j \,\varepsilon\, F_i\}, \alpha_i]$. (Note that $W_i^* = 1 - \sum_{j \varepsilon F_i} W_j$ is the probability that the object i is ranked first of the objects in D_i.) The prior information is equivalent, for $[\{W_j, j \,\varepsilon\, F_i\}, W_i^*]$, to performing $\beta_i = \alpha_i + \sum_{j \varepsilon F_i} \beta_j$ experiments and noting that an object in D_j, $(j \,\varepsilon\, F_i)$, was ranked first in β_j cases. In terms of the original prior information we look at those cases in which an object in D_i was first, and count the cases in which an object in D_j was first, $j \,\varepsilon\, F_i$. We now consider the effect of the present tree τ on the W variables; τ may be expressed as the intersection of the independent events: i is ranked first among the objects in D_i. Thus the information for $[\{W_j\}, j \,\varepsilon\, F_i, W_i^*]$ is now equivalent to performing $\beta_i + 1$ experiments and observing that an object in D_j, $j \,\varepsilon\, F_i$, was ranked first in β_j cases. Thus the $[\{W_j, j \,\varepsilon\, F_i\}, W_i^*]$ become independent Dirichlet variables with parameters $[\{\beta_j, j \,\varepsilon\, F_i\}, \alpha_i + 1]$, proving the theorem.

It is possible to demonstrate the theorem symbolically, working with the densities of the variables $\{W_i\}$, and following through the above verbal argument.

The verbal argument makes it possible to better understand the meaning of the prior information, and the effect of the tree τ on the parameters.

Since θ_i is a product of independent Dirichlet variables, it is relatively simple to determine its moments, and more generally any expectation of the form $E(\prod \theta_i^{k_i} \mid \tau)$. For example,

(5) $$E(\theta_i \mid \tau) = [(\alpha_i + 1)/\beta_0] \prod_{j \epsilon U_i} \beta_j/(\beta_j + 1),$$

(6) $$E(\theta_i^2 \mid \tau) = [(\alpha_i + 1)(\alpha_i + 2)/\beta_0(\beta_0 + 1)] \prod_{j \epsilon U_i} \beta_j/(\beta_j + 2),$$

(7) $$E(\theta_i \theta_j \mid \tau) = (\alpha_k + 1)E(\theta_k^2 \mid \tau)E(\theta_i \mid \tau)E(\theta_j \mid \tau)/E^2(\theta_k \mid \tau)(\alpha_k + 2),$$

where $i \neq j$, $k = \inf \{l \mid \tau(i) \leqq \tau(l), \tau(j) \leqq \tau(l)\}$.

In practice, the β_i may be computed by recurrence relations on the tree, and the above equations for the first two moments are also best exploited by recurrence relations. Empirical sampling of the posterior distribution may be performed by generating Dirichlet variables and computing θ_i from these using Theorem 2.

If the tree τ is a complete rank σ, and if $\alpha_i = \alpha$, $i = 1, \cdots, N$ we have $E(\theta_{(i)} \mid \sigma) = (\alpha + 1)/N\alpha \cdot \alpha^i \cdot \prod_{j=1}^{i} (N - j + 1)/(1 + \alpha(N - j + 1))$ where $\theta_{(i)}$ denotes the parameter associated with the object ranked ith. For $\alpha_i = 1$, $i = 1, \cdots, N$ we have $E[\theta_{(i)} \mid \sigma] = (N - i + 1)/2(N + 1)N$, which sets the expected probabilities proportional to the inverse order; these estimates reflect a very heavy prior judgment of equality, equivalent to the observation that each object was ranked first an equal number of times in N experiments. In Table 1, the posterior averages of θ given σ, $E[\theta_{(i)} \mid \sigma]$, are listed for $N = 16$, $N\alpha = 0, 1, 2,$

TABLE 1

Posterior averages of θ given ranks, for various prior weights

Rank	Prior Weight					
	0	1	2	4	8	16
1	1.000	.5313	.3750	.2500	.1667	.1176
2	*	.2571	.2446	.1974	.1471	.1103
3	*	.1200	.1556	.1535	.1287	.1029
4	*	.0538	.0963	.1174	.1115	.0956
5	*	.0230	.0578	.0880	.0956	.0882
6	*	.0094	.0335	.0646	.0809	.0809
7	*	.0036	.0186	.0461	.0674	.0735
8	*	.0013	.0098	.0319	.0551	.0662
9	*	.0004	.0049	.0213	.0441	.0588
10	*	.0001	.0023	.0135	.0343	.0515
11	*	*	.0010	.0081	.0257	.0441
12	*	*	.0004	.0045	.0184	.0368
13	*	*	.0001	.0023	.0123	.0294
14	*	*	*	.0010	.0074	.0221
15	*	*	*	.0003	.0037	.0147
16	*	*	*	.0001	.0012	.0074

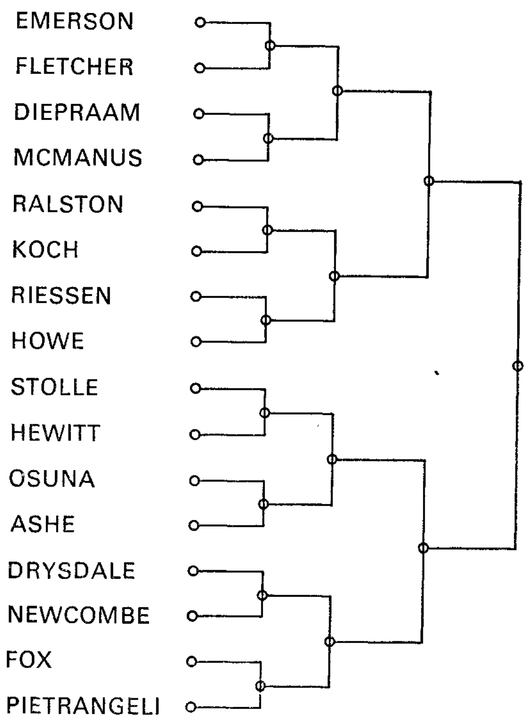

FIG. 1. Design of 1965 Wimbledon quarter finals (τ_0).

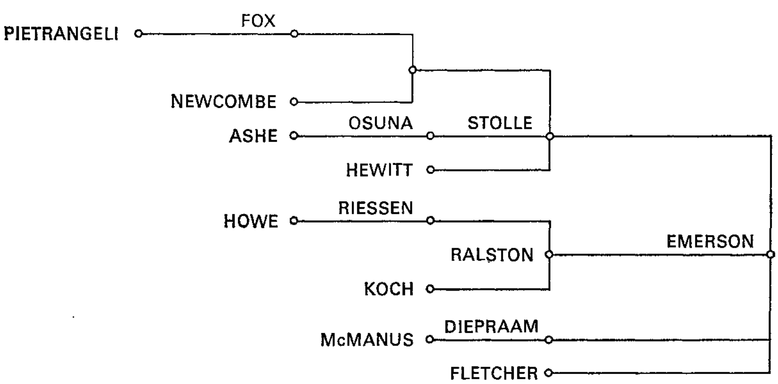

FIG. 2. Observed tree, 1965 Wimbledon quarter finals.

4, 8, 16. With $\alpha = 0$, we estimate $\theta_{(i)} = 1$, $\theta_{(i)} = 0$ for $i \neq 1$, in accordance with the single observation of which object is first. As α increases, the probability spreads out over the lower ranked objects, and at $\alpha = \infty$, $E(\theta_i \mid \sigma) = \frac{1}{16}$.

As a second example, consider the 1965 Wimbledon tennis tournament, from the quarter finals on. The tree design τ_0 is given in Figure 1. The realised tourna-

TABLE 2

Posterior Means and Variances of Players Win Probabilities Given the 1965 Wimbledon Tennis Quarter Finals

	Players	Prior Weight					
		0	1	2	4	8	16
1	Emerson	1.00	.53 (.08)	.38 (.06)	.25 (.03)	.17 (.01)	.12 (.006)
2	Stolle	*	.18 (.19)	.19 (.10)	.17 (.04)	.13 (.02)	.10 (.007)
3	Ralston	*	.11 (.11)	.13 (.06)	.13 (.03)	.11 (.01)	.09 (.006)
4	Drysdale	*	.04 (.12)	.06 (.08)	.08 (.04)	.09 (.02)	.08 (.008)
5	Diepraam	*	.06 (.06)	.08 (.04)	.08 (.02)	.08 (.01)	.08 (.005)
6	Osuna	*	.02 (.06)	.04 (.04)	.06 (.03)	.07 (.01)	.07 (.006)
7	Riessen	*	.01 (.06)	.03 (.04)	.04 (.03)	.06 (.01)	.06 (.007)
8	Fox	*	.00 (.06)	.01 (.04)	.03 (.03)	.04 (.01)	.06 (.008)
9	Fletcher	*	.03 (.03)	.04 (.02)	.05 (.01)	.06 (.01)	.06 (.005)
10	Hewitt	*	.01 (.03)	.02 (.02)	.03 (.01)	.04 (.01)	.05 (.005)
11	Koch	*	.01 (.03)	.01 (.02)	.03 (.02)	.04 (.01)	.05 (.006)
12	Newcombe	*	.00 (.03)	.01 (.02)	.02 (.02)	.03 (.01)	.04 (.006)
13	McManus	*	.00 (.03)	.01 (.02)	.02 (.02)	.03 (.01)	.04 (.006)
14	Ashe	*	.00 (.03)	.00 (.02)	.01 (.02)	.02 (.01)	.03 (.006)
15	Howe	*	.00 (.03)	.00 (.02)	.01 (.02)	.02 (.01)	.03 (.006)
16	Pietrangeli	*	.00 (.03)	.00 (.02)	.01 (.02)	.02 (.01)	.03 (.007)

ment tree is given in Figure 2. Estimation of the parameters θ_i, the probability that player i wins the tournament, is given in Table 2, with prior weights equivalent to 0, 1, 2, 4, 8, 16 tournaments; the variances of the parameters are included. The order of listings of the player is: player i is ranked over player j if he played in a later round of the tournament, or if player τi is ranked over player τj. (Here τi is the player that beat i.) In the estimates of θ_i this order is violated by Fletcher who was beaten by the final winner Emerson in the first round. With priors $\alpha_i = \frac{1}{16}$, Fletcher beats Osuna, Riessen, Fox all of whom reached the next round. These reversals disappear if the weight of prior evidence is increased.

4. Extensions, exploitation, and left-overs.

(1) *Independent matches.* It may be argued that the matches in a knockout tournament be regarded as independent events; in this case the observed tree is an intersection of independent events of the form: i is ranked highest of the objects $\{i, F_i\}$ and $P(\tau \mid \theta) = \prod_{i=1}^{N} (\theta_i/(\sum_{j\epsilon F_i} \theta_j + \theta_i)$. In our model, we do not regard the events as independent, but allow the past records of the players in the tournament to affect the probable outcomes of the matches. It will help understanding of the differences of the models if we consider players with abilities $X_1, \cdots, X_n$ which are random variables with distributions $F^{\theta_1}, F^{\theta_2}, \cdots, F^{\theta_n}$. In the dependent model we used, we imagine the $X_1, X_2, \cdots, X_n$ fixed for the duration of the tournament, which merely elicited ordering relations between the fixed but unknown X_i. In the independent model, the X_i are selected freshly and at random at each match, the outcomes of the matches are independent with the above probabilities.

The dependent model is thus appropriate if the judgments are in some sense simultaneous, so that the same X_i are used in all judgments. The independent model is appropriate if the judgments are made ignoring the results of past judgments. It seems reasonable that practice lies somewhere between these two idealizations. In many athletic contests, the independent model may be more appropriate. The advantage of the dependent or simultaneous scheme is that the transitivity requirement may be assumed; resulting in more precise estimates of θ with a given body of data.

Another model, which we shall not discuss, is proposed by Bradley (1965).

(2) *Tournament design.* Attempts are sometimes made to "seed" tournaments so that the *a priori* highly ranked players will have a better chance of winning. With the dependent model, the probability of a given player winning does not depend on the tournament design, so that the usual objective of seeding cannot be considered as a criterion for tournament design; this objective could be attempted with the independent model. A more natural criterion requires the tournament to be designed so that the results are most informative about the parameters $\theta_1, \cdots, \theta_n$. The design τ_0 of the tournament may be formalized as a family of subsets of A which are a partial order under inclusion; i.e. $\tau_0(A_1) \leqq \tau_0(A_2)$ if $A_1 \subset A_2$. The outcome τ of such an experiment states that best element in each subset in τ_0; for each τ_0, a certain family of outcome trees τ is possible, and this family is a partition of the set of all $N!$ possible ranks σ. Formally we may seek to optimize the change in entropy of θ due to τ,

$$\Delta I(\tau_0) = E[\log P(\tau \mid \theta)/P(\tau)]$$

where E denotes averaging over θ and τ, and $p(\tau)$ is the prior distribution of τ. No progress towards this objective has been made.

TABLE 3

Estimation of θ_i Based on Conditional Exponential; Means and Variances of Conditional Ranks

Players	$1/\hat{\theta}_i$	$\hat{\theta}_i/\sum\hat{\theta}_i$	$E[\sigma(i)\mid\tau)$	$\text{Var}[\sigma(i)\mid\tau]$
Emerson	1	.41	1.00	0.00
Stolle	3	.14	2.78	1.11
Ralston	5	.10	4.20	4.69
Drysdale	7	.06	5.62	4.23
Diepraam	9	.05	6.33	11.56
Osuna	11	.04	7.52	9.42
Riessen	13	.03	8.47	9.32
Fox	15	.03	9.41	7.41
Fletcher	17	.02	9.00	18.67
Hewitt	19	.02	9.89	14.85
Koch	21	.02	10.00	13.08
Newcombe	23	.02	11.31	10.30
McManus	25	.02	11.67	11.56
Ashe	27	.02	12.26	9.05
Howe	29	.01	12.73	7.75
Pietrangeli	31	.01	13.21	6.00

(3) *Exponential variables.* We know that the probability model for ranks may be obtained by letting $\{Y_i\}$ be independent exponential variables with means $\{1/\theta_i\}$, and setting $P[\sigma \mid \theta] = P\{Y_{\sigma(i)} \leqq Y_{\sigma(j)} \text{ if } \sigma(i) \geqq (j), \text{ all } i, j\}$. We then have that $Y_1, \cdots, Y_n \mid \tau, \theta$ are such that $\{Y_i - Y_{\tau i}\}$ are joint independent exponentials with means $\{1/\sum_{\tau(j) \leqq \tau(i)} \theta_j\}$; here $Y_0 = 0$. Thus the famous result, that differences of successive exponential order statistics are independent exponentials, extends to tree type partial orderings. It may be plausible to evaluate τ by its action on the exponential variables $Y_1, \cdots, Y_n$. For example, letting $\beta_k = \sum_{\tau(i) \leqq \tau(k)} \alpha_i$, where the α_i are prior estimates of θ_i, we have $E[Y_i \mid \tau, \alpha] = \sum (1/\beta_k)$. Also given θ_i we have $E[Y_i \mid \theta] = 1/\theta_i$. This suggests $\sum_{\tau(i) \leqq \tau(k)} (1/\beta_k)$ as an estimate of $1/\theta_i$. See Table 3.

It is true, and pointed out by a referee and the editor, that the exponential variables are artificial constructs, possibly useful in generating probability relationships, but probably best kept away from inferences about θ. The exponential variables may be realized in reliability problems examining the life of several components, which are connected together in a device; when the device fails, it seems plausible to suppose that examination of the components will reveal some dead, some alive, and ordering relations on the lives of the dead components. I have had no luck in imagining devices for which these ordering relations have tree structure.

(4) *Conditional ranks.* In Hartigan (1966), it was proposed to summarize the tournament τ by features of the conditional rank distribution $\sigma \mid \tau$. It was there assumed that all θ_i were equal, and that the tournament structure was of a particular type on 2^n objects. By combinatorial arguments similar to those used there (so a referee has suggested) the distribution of $\sigma(i)$ given τ may be obtained in terms of the distribution of $\sigma(\tau i)$ by

$$(8) \quad P[\sigma(i) = k + l \mid \sigma(\tau i) = k] = \binom{N_{\tau i}-1-l}{N_i-1} \Big/ \binom{N_{\tau i}-1}{N_i}, \; 1 \leqq l \leqq N_{\tau i} - N_i,$$

where N_i is the number of objects ranked below i by the tree. The mean and variance of $\sigma(i)$ may be obtained directly from (8) or from the conditional exponentials.

$$(9) \qquad E[1 - \sigma(i)/(N+1) \mid \tau] = \textstyle\prod_{\tau(j) \leqq \tau(i)} N_j/(N_j + 1),$$

$$(10) \quad E[1 - \sigma(i)/(N+1)][1 - \sigma(i)/(N+2)] \mid \tau] = \textstyle\prod_{\tau(j) \leqq \tau(i)} N_j/(N_j + 2).$$

Computation of the means and variances of conditional ranks offers a simple way to represent τ without overt commitment to a parametric approach. The above assumes that all θ_i are equal. See Table 3.

TABLE 4

Approximate Dirichlet Distributions Before and After Wimbledon

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
α Before	.006	.006	.006	.006	.006	.006	.006	.006	.006	.006	.006	.006	.006	.006	.006	.006
α After	.531	.177	.106	.035	.059	.020	.012	.004	.031	.010	.006	.002	.004	.001	.001	.000

It is interesting to note the close relation between equations (9) and (10) and equations (5) and (6) when $\alpha_i = 1, i = 1, \cdots, n$.

(5) *General partial orders.* Analytic techniques for general partial orders are not available. It is possible to use an empirical conditional rank analysis; suppose that all ranks are *a priori* equally likely. Generate ranks consistent with the partial order yielding σ with probability $p(\sigma)$. Then, for example, $E(\sigma(i))$ is estimated by $K\sigma(i)/p(\sigma)$ where K may be computed by noting that $\sum \sigma(i) = N(N+1)/2$.

(6) *Sequences of tournaments.* Suppose that $\tau_1, \tau_2, \cdots, \tau_n$ are a sequence of independent tournament results. It is not necessary that all players appear in all tournaments. We can evaluate the likelihood $\prod_{i=1}^{n} P(\tau_i \mid \theta)$, but a complete Bayesian analysis is no longer available. Maximum likelihood may be a feasible tool for n large enough (for $n = 1$ it is useless). A weak prior may be applied to each of the tournaments to obtain estimates of the θ_i for each tournament, and these may be combined to obtain an overall estimate, and compared to examine differences between tournaments. A similar analysis may be used on the conditional ranks.

An approximate Bayes analysis can be performed if we express the posterior distribution of θ_i given a tournament τ as a Dirichlet distribution; the effect on the posterior distribution of a new tournament result may then be evaluated. It seems plausible to use $D\{\alpha_i, i = 1, \cdots, n\}$, $\alpha_i/\sum \alpha_i = E[\theta_i \mid \tau]$, with $\sum \alpha_i$ equal to the weight of the original prior. Table 2 suggests that the essential effect of the observation τ is drastic changes in the estimation of θ_i and only small changes in the variability of θ_i. See Table 4 for an example.

(7) *Asymmetry.* The probability model for ranks is asymmetric in the sense that the probabilities of propositions like $j < i, k < i$ are quite different to those of propositions like $i < j, i < k$. For example θ_i is the probability that object i is ranked highest, but no simple expression is available for the probability that object i is ranked least. The axiom of choice holds for selecting the most preferred object from a set, but not for selecting the least preferred.

The independent model avoids this difficulty.

REFERENCES

BRADLEY, R. A. (1965). Another interpretation of a model for paired comparisons. *Psychometrika* **30** 315–318.

BRADLEY, R. A. and TERRY, M. E. (1952). Rank analysis of uncomplete block designs I. The method of paired comparison. *Biometrika* **39** 324–345.

BRUNK, H. D. (1960). Mathematical models for ranking from paired comparisons. *J. Amer. Statist. Assoc.* **55** 503–520.

DAVID, H. A. (1963). *The Method of Paired Comparison.* Griffin, London.

HARTIGAN, J. A. (1966). Probabilistic completion of a knockout tournament. *Ann. Math. Statist.* **37** 495–503.

LUCE, R. D., BUSH, R. B. and GALANTER, E. (1963). *Handbook of Mathematical Psychology.* Wiley, New York.

SAVAGE, I. R. (1956). Contributions to the theory of rank order statistics—the two sample case. *Ann. Math. Statist.* **27** 590–615.