The Annals of Applied Probability
1996, Vol. 6, No. 3, 862–882

OPTIMAL SELECTION PROBLEMS BASED ON EXCHANGEABLE TRIALS

BY ALEXANDER V. GNEDIN AND ULRICH KRENGEL

University of Göttingen

We consider optimal stopping problems with loss function q depending on the rank of the stopped random variable. Samuels asked whether there exists an exchangeable sequence of random variables $X_1, \dots, X_n$ without ties for which the observation of the values of the X_i's gives no advantage in comparison with the observation of just the relative ranks of the variables. We call distributions of the sequences with this property *q-noninformative* and derive necessary and sufficient conditions for this property. Extending an impossibility result of B. Hill, we show that, for any $n > 1$, there are certain losses q for which q-noninformative distributions do not exist. Special attention is given to the classical problem of minimizing the expected rank: for n even we construct explicitly universal (randomized) stopping rules which are strictly better than the rank rules for any exchangeable sequence.

1. Introduction. We consider optimal selection models, which have become known also as secretary problems. A well-known theme in this field is the comparison of decision rules based on different kinds of information regarding the alternatives available for selection. One of the most attractive features of this class of models is that simple low-information strategies often demonstrate surprisingly high performance.

We adopt the following framework. Assume a choice is to be made from a set of n alternatives examined in a random order. The alternatives are identified with their numerical values $X_1, \dots, X_n$ which are assumed to be exchangeable random variables without ties, with an n-dimensional probability distribution $\mathscr{D}$. The values of the X_i's are observed sequentially, with the object to select one of them via a stopping rule (i.e., no recall is allowed). Stopping with the kth largest value incurs the loss $q(k)$, where q is a given loss function. The performance of a stopping rule is characterized by its expected loss (also called the *risk*).

The central role in our discussion will be played by *rank rules* which, at each stage, use only the values of relative ranks of the alternatives observed so far. The class of rank rules is a finite proper subclass of the class of all stopping rules adapted to the X_i's. Under the exchangeability assumption, the risk of the rank rules does not depend on $\mathscr{D}$; thus we can speak of their performance regardless of the distribution.

Received May 1995; revised November 1995.
AMS 1991 *subject classification*. Primary 60G40.
Key words and phrases. Optimal stopping, exchangeability, rank, secretary problems.

Following Samuels (1994) we study the following questions:

1. For a given loss function q does there exist a *q-noninformative* distribution for the $\{X_i\}$ such that the minimal risk using stopping rules based on the relative ranks only is the same as that for the wider class of stopping rules adapted to the natural filtration of the sequence?
2. Is there a stopping rule (which may be randomized) with respect to the larger filtration which performs better than any rule just based on the relative ranks for *any* distribution of the sequence?

These questions are closely related to minimax considerations in the secretary problems [cf. Hill and Kennedy (1992), Gnedin (1994) and Gnedin and Krengel (1995)]. If the answer to the first question is affirmative, then the answer to the second question is negative, and the best rank rule and the relevant $\mathscr{D}$ are both minimax. On the other hand, if the second question has a positive answer and the best rank rule is minimax, then there are no minimax $\mathscr{D}$.

It is also easy to see that if there is a randomized stopping rule which beats the rank rules for all $\mathscr{D}$, then for each fixed $\mathscr{D}$ there is also a nonrandomized rule improving the rank rules. This follows from the fact that randomization does not reduce the risk in stopping problems. Therefore, if the second question is resolved positively, the answer to the first question is negative.

A restricted version of the first question, where $\mathscr{D}$ is assumed to be a mixture of product distributions with identical factors, is of some interest from the Bayesian point of view. The knowledge of $\mathscr{D}$ is (at least theoretically) equivalent to the knowledge of the "prior" distribution, that is, of the mixing measure. The decision-making process in this case involves, via the prior-to-posterior transformation, the learning of the distributions of the variables X_i, the latter being in this case conditionally iid. The affirmative answer to the first question would imply that there exist "sufficiently noninformative priors" for which learning of the underlying distribution cannot help to perform better than by using the optimal rank rule [cf. Samuels (1989, 1994)].

This restricted class of models is known in the literature as problems with "partial information" [cf. Samuels (1991)]. One extreme in this class is the "full information" problem, where $\mathscr{D}$ is a continuous product distribution and the X_i's are iid with known distribution. The generally accepted opposite to the "full information" model is the so-called "no information" problem where only the relative ranks of alternatives are observable. One of our motivations was to clarify the obscurity concerning what constitutes a "no information" problem by comparing the rank model with the logical opposite to "full information"; namely, the actual values of the X_i's are observed, although nothing is known about $\mathscr{D}$ beyond the fact that $\mathscr{D}$ is a mixture of product distributions.

The questions we study here were raised in connection with the classical best choice problem, corresponding to the special loss: $q(1) = 0$ and $q(k) = 1$

for $k > 1$ [cf. Ferguson (1989)]. In this problem, the answer to the first question is now known to be affirmative for all $n > 2$ [cf. Silverman and Nadás (1992), for $n = 3$, and Gnedin (1994), for all $n \geq 3$]. The case of two observations is special: the randomized rule "pick a normally distributed T, independent of X_1 and X_2, and then stop at X_1 if $X_1 > T$ and at X_2 otherwise" always beats the rank rules [cf. Cover (1987)]. Recently Samuels (1994) studied the general loss in the case $n = 3$. He showed that for some (but not all) losses the distributions introduced by Gnedin (1994) also provide the affirmative answer to the first question.

In this paper we give necessary and sufficient conditions for overall optimality of a rank rule in terms of some inequalities on predictive probabilities of relative ranks. We show that these inequalities can be interpreted in terms of nonnegativity of certain signed linear combinations of multidimensional marginal distributions of the order statistics vector. In the case when the optimal rank rule is not unique, some of these inequalities turn out to be equalities; that is, some marginals must be linearly dependent. We show that $(n-1)$-dimensional marginals cannot be linearly dependent. From this fact we derive that, for any n, there exist losses q for which the first question has a negative answer. For $n > 3$, there are monotone loss functions among such q. For n even, q-noninformative distributions do not exist for $q(k) \equiv k$, that is, in the minimal expected rank problem [cf. Chow, Moriguti, Robbins and Samuels (1964)]; for this case we construct explicitly a universal stopping rule whose risk is strictly smaller than the risk of any rank rule for all $\mathscr{D}$.

It is worth mentioning that the considerations regarding the predictive distribution of relative ranks are closely related to the problems of nonparametric inference studied by Hill (1968). Our result about linear independence of marginals extends his result, which says that for no exchangeable sequence without ties is the relative rank of X_n independent of $X_1, \dots, X_{n-1}$.

Finally, we give a description of all exchangeable distributions for which, in the classical best choice problem ($n > 2$), the best rank rule is overall optimal.

2. Preliminaries. Given $n > 1$, let $X_1, \dots, X_n$ be exchangeable random variables, which tie only with probability 0. Let

$$A_m := \#\{i\colon X_i \geq X_m,\ 1 \leq i \leq n\}, \qquad m = 1, \dots, n,$$

denote the *absolute rank* of the mth variable (thus smaller ranks are assigned to larger values of the X_i's). By exchangeability, all $n!$ possible values of the sequence $(A_1, \dots, A_n)$ are equally likely. We interpret $X_1, \dots, X_n$ as sequential observations and consider the class $\mathscr{X}$ of stopping rules $\tau \leq n$ satisfying the measurability condition

$$\mathscr{X}\colon (\tau = m) \text{ is measurable w.r.t. } X_1, \dots, X_m \qquad \forall\, m \leq n.$$

It is often useful, and in fact equivalent, to consider stopping rule τ as a function of n real variables with the property that the set of those $(x_1, \dots, x_n) \in \mathbb{R}^n$ satisfying $\tau(x_1, \dots, x_n) = m$ is a Borel set of the form

$B \times \mathbb{R}^{n-m}$, where $B \subset \mathbb{R}^m$. We will evaluate $\tau \in \mathscr{X}$ by means of the expected loss (called also the *risk*)

$$Eq(A_\tau) = \sum_{m=1}^{n} E\big(q(A_m)1_{(\tau=m)}\big),$$

where q: $\{1, \dots, n\} \to \mathbb{R}$ is a specified loss function measuring the loss of stopping at the observation with a given value of absolute rank. To avoid trivialities, $q(\cdot)$ will be assumed nonconstant.

Introduce the *relative ranks* as

$$R_m := \#\{i: X_i \ge X_m, 1 \le i \le m\} = \#\{i: A_i \le A_m, 1 \le i \le m\} \qquad \forall\, m \le n,$$

and the class $\mathscr{R}$ of *rank* stopping rules satisfying a weaker measurability condition

$$\mathscr{R}: (\tau = m) \text{ is measurable w.r.t. } R_1, \dots, R_m.$$

The values of these classes are

$$V_{\mathscr{X}}(\mathscr{D}) := \inf_{\tau \in \mathscr{X}} Eq(A_\tau) \quad \text{and} \quad V_{\mathscr{R}} := \inf_{\tau \in \mathscr{R}} Eq(A_\tau).$$

The existence of optimal stopping rules in problems with a finite number of observations is a standard fact of the optimal stopping theory.

The distribution of relative ranks is the same for all exchangeable sequences without ties; therefore $V_{\mathscr{R}}$ does not depend on the distribution of $(X_1, \dots, X_n)$. In contrast to this, $V_{\mathscr{X}}(\mathscr{D})$ depends heavily on the distribution of observations. The set of rank rules always constitutes a proper subclass of $\mathscr{X}$, even in the simplest case, when $(X_1, \dots, X_n)$ is a random permutation of a fixed n-tuple of different reals. Obviously, $V_{\mathscr{X}}(\mathscr{D}) \le V_{\mathscr{R}}$. Our main concern here is the existence of distributions for which these values are equal.

DEFINITION 1. Given a loss $q(\cdot)$, the probability distribution $\mathscr{D}$ of $(X_1, \dots, X_n)$ is said to be *q-noninformative* iff $V_{\mathscr{X}}(\mathscr{D}) = V_{\mathscr{R}}$.

This property means that there exists a stopping rule $\tau \in \mathscr{R}$, which is optimal in $\mathscr{X}$. Any q-noninformative distribution $\mathscr{D}_0$ is necessarily minimax in the sense that

$$V_{\mathscr{X}}(\mathscr{D}_0) = \sup_{\mathscr{D}} \inf_{\tau \in \mathscr{X}} Eq(A_\tau) = V_{\mathscr{R}}.$$

Note that $V_{\mathscr{X}}(\mathscr{D})$ is the same for $(X_1, \dots, X_n)$ and $(\phi(X_1), \dots, \phi(X_n))$, where ϕ is a strictly monotone function. The family of q-noninformative distributions is clearly convex. Given a single q-noninformative $\mathscr{D}$, we can construct new ones by combining monotone rescaling with forming convex mixtures.

If the loss function $q(\cdot)$ admits a q-noninformative distribution, the same is valid also for the losses $\alpha q(\cdot) + \beta$, with arbitrary constants β and $\alpha > 0$. The same holds also for the loss $\hat{q}(k) = q(n-k)$, because the absolute ranks of the reflected sequence $(-X_1, \dots, -X_n)$ are $n - A_1, \dots, n - A_n$.

3. Optimal rank rules. In this section we discuss the structure of optimal rank rules, with special emphasis on possible nonuniqueness. We list first some simple facts about relative and absolute ranks.

Let $(a_1, \dots, a_n)$ be a permutation of the integers $\{1, \dots, n\}$ and let $r_k := \#\{i: a_i \le a_k,\ i \le k\}$, $k = 1, \dots, n$. It is easy to see that

$$r_k = a_k - \#\{i: a_i < a_k,\ k < i \le n\}. \tag{1}$$

Since r_k is expressed in (1) through $a_k, \dots, a_n$, we can use this formula to regard $(r_m, \dots, r_n)$ as a function of $(a_m, \dots, a_n)$, $m \le n$. Keep in mind that $1 \le r_k \le k$.

PROPOSITION 2. *For any $m = 1, \dots, n$, there exists a mapping*

$$\Phi_m: (r_m, \dots, r_n) \mapsto (a_m, \dots, a_n)$$

between the set of possible values of $(r_m, \dots, r_n)$ and the set of possible values of $(a_m, \dots, a_n)$, which is inverse to the correspondence $(a_m, \dots, a_n) \mapsto (r_m, \dots, r_n)$ given by (1), *with k running from m to n.*

PROOF. Observe that both sequences may take $n!/(m-1)!$ different values, and the correspondence $(a_m, \dots, a_n) \mapsto (r_m, \dots, r_n)$ is injective.

Explicitly, a_m can be expressed through $(r_m, \dots, r_n)$ as follows. Set $z_{m,k} := \#\{i: a_i \le a_m,\ i \le k\}$, $m \le k \le n$. We have $z_{m,m} = r_m$, $z_{m,n} = a_m$ and, recursively, for $k = m, \dots, n-1$,

$$\begin{aligned} z_{m,k+1} &= z_{m,k} + 1 && \text{if } r_{k+1} \le z_{m,k}, \\ z_{m,k+1} &= z_{m,k} && \text{if } r_{k+1} > z_{m,k}. \end{aligned}$$

It is easy to see that $a_m - r_m + 1$ is equal to the number of different members of the sequence $z_{m,m}, z_{m,m+1}, \dots, z_{m,n}$, which depends only on $r_m, \dots, r_n$. Iterating this construction, we express $(a_m, \dots, a_n)$ in terms of the array $(z_{i,k};\ m \le i \le k \le n)$, which depends only on $(r_m, \dots, r_n)$. $\square$

The following facts will be used without special reference.

PROPOSITION 3. *The distribution of the ranks has the following properties*:

(i) *$R_1, \dots, R_n$ are independent, and $P(R_m = k) = m^{-1}$ for $1 \le k \le m \le n$*;

(ii) $\Phi_m(R_m, \dots, R_n) = (A_m, \dots, A_n)$;

(iii) *$(R_1, \dots, R_{m-1})$ and $(A_m, \dots, A_n)$ are independent*;

(iv) *$(R_1, \dots, R_{m-1})$ and $(X_m, \dots, X_n)$ are independent*;

(v) *$(R_m, \dots, R_n)$ is conditionally independent of $(R_1, \dots, R_{m-1})$ given the order statistics of $(X_1, \dots, X_{m-1})$.*

PROOF. The first three statements follow immediately from exchangeability and Proposition 2. Using exchangeability and rearranging $X_1, \dots, X_{m-1}$, we see that, for any fixed Borel set $B \in \mathbb{R}^{n-m+1}$, the events of the form

$$(R_1 = r_1, \dots, R_{m-1} = r_{m-1}, (X_m, \dots, X_n) \in B),$$

with different admissible combinations of the r_i's are equally likely; therefore

$$P(R_1 = r_1, \dots, R_{m-1} = r_{m-1} | X_m, \dots, X_n) = 1/(m-1)!$$

and (iv) follows. The last assertion is proved similarly. □

We will now use standard backward induction arguments to describe the structure of all optimal stopping rules in $\mathscr{R}$. Without assuming the monotonicity of q, we need to treat separately a curious family of loss functions for which all rank rules have the same risk.

For $m \leq n$, we introduce the subclass of rank stopping rules which always pass the first $m-1$ observations, $\mathscr{R}_m := \{\tau \in \mathscr{R}: \tau \geq m\}$, and set

$$v_m := \inf_{\tau \in \mathscr{R}_m} Eq(A_\tau), \qquad v_{n+1} := \infty.$$

Then we have

$$v_1 = V_{\mathscr{R}}, \qquad v_n = Eq(A_n) = n^{-1}(q(1) + \cdots + q(n)),$$

because $\mathscr{R}_1 = \mathscr{R}$ and $\mathscr{R}_n$ consists of the single constant stopping rule $\tau \equiv n$. Set further, for $r \leq m \leq n$,

$$u_m(r) := E(q(A_m)|R_m = r) = \sum_{a=r}^{n-m+r} q(a) \frac{\binom{a-1}{r-1}\binom{n-a}{m-r}}{\binom{n}{m}}. \tag{2}$$

The hypergeometric probabilities in (2) appear as the distribution of A_m conditioned on $R_m = r$.

Define, for $m = 1, \dots, n$,

$$\begin{aligned} C_m &:= \{r: u_m(r) > v_{m+1}, r \leq m\},\\ S_m &:= \{r: u_m(r) < v_{m+1}, r \leq m\},\\ I_m &:= \{r: u_m(r) = v_{m+1}, r \leq m\}. \end{aligned}$$

These three sets can be used to describe the structure of optimal rank stopping rules. Proposition 3(iii) implies that (A_m, R_m) are independent of $R_1, \dots, R_{m-1}$; the next result is a consequence of this and well-known results from optimal stopping.

PROPOSITION 4. *For any choice of $J_1 \subset I_1, \dots, J_n \subset I_n$, the stopping rule of the form*

$$\tau_m = \min\{k: R_k \in S_k \cup J_k, m \leq k \leq n\} \tag{3}$$

is optimal in $\mathscr{R}_m$.

Given that the first $m-1$ observations have been passed, it is optimal to stop at the mth observation if R_m assumes a value in S_m and to continue if

$R_m \in C_m$. For $R_m \in I_m$, it makes no difference whether one stops or continues. Obviously, $S_n = \{1, \dots, n\}$ and $C_n = I_n = \emptyset$.

Standard backward induction for optimal stopping proves also that the sequence $v_1, \dots, v_n$ is a solution to the recursion

$$v_m = m^{-1} \sum_{r=1}^{m} \min(u_m(r), v_{m+1}). \tag{4}$$

In some cases all rank rules have the same risk, and the optimal stopping problem in $\mathscr{R}$ turns out to be trivial.

DEFINITION 5. A loss function $q(\cdot)$ is said to be *singular* if constant stopping rules are optimal in $\mathscr{R}$.

This definition makes sense, because the risk of all constant rules is the same, namely, equal to $n^{-1}(q(1) + \cdots + q(n))$. More explicitly, singular losses can be characterized as solutions of the linear system

$$\frac{r}{n} q(r+1) + \frac{n-r}{n} q(r) = \frac{1}{n} \sum_{k=1}^{n} q(k), \qquad r = 1, \dots, n-1, \tag{5}$$

obtained by equating $u_{n-1}(r)$ and v_n. It is not hard to show that the algebraic rank of the system is $n - 2$. Thus the solutions form a two-dimensional linear space, including the one-dimensional subspace of constant loss functions.

PROPOSITION 6. *The loss function $q(\cdot)$ is singular iff any of the following conditions holds*:

(i) *q satisfies* (5);
(ii) *a constant rule is optimal in $\mathscr{R}_m$ for some $m < n$*;
(iii) *$I_m \cup S_m = \{1, \dots, m\}$ for some $m < n$.*

PROOF. If some constant rule is optimal in $\mathscr{R}_m$, then, by exchangeability, any constant rule in $\mathscr{R}_m$ is optimal in this class. Therefore all three conditions are equivalent and follow from the singularity of q.

Now let us prove that (i) implies that q is singular. We say that a stopping rule $\tau \in \mathscr{R}$ *ignores* the kth observation, $k < n$, if $\tau \neq k$ a.s. and for $i > k$ a weaker measurability condition holds: each event $(\tau = i)$, $i > k$, is measurable w.r.t. the variables $R_1, \dots, R_{k-1}, R'_{k+1}, \dots, R'_i$, where R'_j, $k < j \le i$, is the rank of the jth observation among the observations numbered $1, \dots, k-1, k+1, \dots, j$.

Assume (5) holds. By Proposition 3 and (4) this means exactly that the constant rules $\tau \equiv n - 1$ and $\tau \equiv n$ are optimal in $\mathscr{R}_{n-1}$, or, equivalently, $I_{n-1} = \{1, \dots, n-1\}$. Note that $\tau = n - 1$ ignores the nth observation. For $m < n - 1$, we can find, by Proposition 4, an optimal rule τ_m in $\mathscr{R}_m$ which ignores the nth observation. By symmetry, exchanging the mth and the nth observations, we see that there exists also an optimal rule τ'_m in $\mathscr{R}_m$, which ignores the mth observation. To construct this rule explicitly, represent τ_m as a function of $R_1, \dots, R_{n-2}$ and set

$$\tau'_m(r_1, \dots, r_{n-2}) = \tau_m(r_1, \dots, r_{n-2}) + 1.$$

Clearly, τ'_m is also in $\mathscr{R}_{m+1}$. We see that there is a stopping rule in $\mathscr{R}_{m+1}$, which is optimal in $\mathscr{R}_m$. Iterating this argument with $m = 1, \dots, n-2$, we prove that $\tau \equiv n$ is optimal in $\mathscr{R}$; that is, q must be singular. □

The next assertion explains the main reason why we need to distinguish the class of singular losses.

PROPOSITION 7. *If $q(\cdot)$ is not singular, then $P(\tau = n) > 0$ for any optimal stopping rule $\tau \in \mathscr{R}$.*

PROOF. If q is not singular, then $C_m \neq \varnothing\ \forall\ m < n$, as a consequence of Proposition 6. Therefore

$$\begin{aligned} P(\tau = n) &\geq P(R_1 \in C_1, \dots, R_{n-1} \in C_{n-1}) \\ &= (\#C_1 \cdot \#C_2 \cdot \cdots \cdot \#C_{n-1})/(n-1)! > 0. \end{aligned} \qquad \square$$

None of the nonconstant monotone loss functions is singular. For $n = 3$, all such q's satisfy $q(1) = q(3)$ [e.g., $q(\cdot) = (1, 0, 1)$ can be interpreted as "pay nothing if you select the second best or a dollar otherwise"].

The nonuniqueness of optimal rank rules is the case exactly when some of the indifference sets turn out to be nonempty, as stated next.

PROPOSITION 8. *An optimal rule in $\mathscr{R}_m$ is unique if and only if $I_m = \cdots = I_n = \varnothing$.*

PROOF. If there is a single optimal optimal rule $\tau \in \mathscr{R}_m$, then $q(\cdot)$ cannot be singular. Therefore, by Proposition 7,

$$P(\tau > k-1, R_k = r) = k^{-1} P(\tau > k-1) > 0, \qquad k = m, \dots, n-1.$$

It follows that the conditions

$$\begin{aligned} (\tau > k-1, R_k = r) &\Rightarrow (\tau = k) \quad \text{for } r \in S_k, \\ (\tau > k-1, R_k = r) &\Rightarrow (\tau > k) \quad \text{for } r \in C_k \end{aligned}$$

determine the optimal rank rule uniquely. The "only if" part of the statement follows from Proposition 4. □

4. Overall optimality of rank rules. To discuss optimality of the rank rules among general stopping rules, we need to proceed in a direction which is opposite to the usual setting of optimal stopping theory. Assuming overall optimality of appropriate rank rules, we will draw some inferences about the process $X_1, \dots, X_n$. Analogous to the $\mathscr{R}_m$'s introduced earlier, let $\mathscr{X}_m := \{\tau \in \mathscr{X}\colon\ \tau \geq m\}$. Our main tool is the following general result, inverting standard backward induction arguments.

PROPOSITION 9. *Assume $\tau_1 \in \mathscr{X}_1, \ldots, \tau_n \in \mathscr{X}_n$ is a sequence of stopping rules satisfying*

$$\tau_m > m \Rightarrow \tau_m = \tau_{m+1} \qquad \forall\, m < n. \tag{6}$$

The following conditions are equivalent:

(i) τ_m *is optimal in* $\mathscr{X}_m$ *for* $m = 1, \ldots, n$;
(ii) τ_m *satisfies the inequalities*

$$\tau_m = m \Rightarrow E(q(A_m)|X_1, \ldots, X_m) \le E\big(q(A_{\tau_{m+1}})|X_1, \ldots, X_m\big), \tag{7}$$

$$\tau_m > m \Rightarrow E(q(A_m)|X_1, \ldots, X_m) \ge E\big(q(A_{\tau_{m+1}})|X_1, \ldots, X_m\big) \tag{8}$$

for $m = 1, \ldots, n$.

PROOF. Assume τ_m is optimal in $\mathscr{X}_m$, then, for any $\tau \in \mathscr{X}_m$,

$$E\big(q(A_{\tau_m})|X_1, \ldots, X_m\big) \le E(q(A_\tau)|X_1, \ldots, X_m), \tag{9}$$

because otherwise τ_m could be improved by a combination of τ and τ_m. Substituting τ_{m+1} for τ and restricting the conditional expectations to the event $(\tau_m = m)$ yields (7). Substituting the constant rule $\tau \equiv m$ in (9) restricted to $(\tau_m > m)$ and using the compatibility condition (6) yields (8). Therefore (i) implies (ii).

Use induction to derive the converse implication. Suppose (7) and (8) hold and that for some m the rule τ_m is optimal. Under these assumptions (9) holds. For arbitrary $s \in \mathscr{X}_{m-1}$ and $\tau = \max(m, s)$, it follows from (7) and (8) that

$$\begin{aligned}
&E(q(A_s)|X_1, \ldots, X_{m-1})\\
&\quad = E(q(A_{m-1})|X_1, \ldots, X_{m-1})1_{(s=m-1)} + E(q(A_\tau)|X_1, \ldots, X_{m-1})1_{(s \ge m)}\\
&\quad \ge E(q(A_{m-1})|X_1, \ldots, X_{m-1})1_{(s=m-1)} + E\big(q(A_{\tau_m})|X_1, \ldots, X_{m-1}\big)1_{(s \ge m)}\\
&\quad \ge \min\Big(E(q(A_{m-1})|X_1, \ldots, X_{m-1}), E\big(q(A_{\tau_m})|X_1, \ldots, X_{m-1}\big)\Big)\\
&\quad = E\big(q(A_{\tau_{m-1}})|X_1, \ldots, X_{m-1}\big).
\end{aligned}$$

This implies that also τ_{m-1} is optimal in $\mathscr{X}_{m-1}$ and justifies the induction step from m to $m-1$. □

It is intuitively clear that exchangeability of the sequence $X_1, \ldots, X_n$ makes it possible to ignore the order in which the first $m-1$ variables have been observed, when deciding whether to stop at X_m or to proceed further.

PROPOSITION 10. *For any $m = 2, \ldots, n$, there exists an optimal stopping rule s_m in $\mathscr{X}_m$, which is independent of the relative ranks $(R_1, \ldots, R_{m-1})$.*

PROOF. We will prove first the stronger assertion that s_m as a function of $(X_1, \ldots, X_n)$ can be selected symmetric in the first $m-1$ components.

The proof is by backward induction. Assume s_{m+1} is optimal in $\mathscr{X}_{m+1}$ and invariant under permutations of $X_1, \dots, X_m$. Define s_m by setting $s_m = m$ if

$$E(q(A_m)|X_1, \dots, X_m) \le E\big(q(A_{s_{m+1}})|X_1, \dots, X_m\big),$$

and $s_m = s_{m+1}$ otherwise. Any event of the form

$$(s_{m+1} = k;\ A_m = a_m, \dots, A_n = a_n)$$

is invariant under the permutations of $X_1, \dots, X_{m-1}$. (We identify the events with their indicators, and the indicators with functions of the X_i's.) Therefore, by exchangeability, the regular conditional probability

$$P(s_{m+1} = k;\ A_m = a_m, \dots, A_n = a_n|X_1, \dots, X_m)$$

is a symmetric function of $X_1, \dots, X_{m-1}$. It follows that both conditional expectations appearing in the above inequality defining s_m are symmetric in $X_1, \dots, X_{m-1}$.

Now independence is easy. By exchangeability and invariance of s_m under permutations of $X_1, \dots, X_{m-1}$, all $(m-1)!$ probabilities $P(s_m = k;\ R_1 = r_1, \dots, R_{m-1} = r_{m-1})$ with k fixed and admissible r_i's are equal; hence the probability of the event $(s_m = k)$ conditioned on $(R_1 = r_1, \dots, R_{m-1} = r_{m-1})$ coincides with its unconditional probability, and the independence follows. The optimality of s_m follows from Proposition 9. □

Next we show that if there exist rank rules which are optimal in $\mathscr{X}$, then there are also rank rules optimal in $\mathscr{X}_m$.

PROPOSITION 11. *The distribution of* $(X_1, \dots, X_n)$ *is q-noninformative if and only if any sequence of stopping rules* $\tau_m \in \mathscr{R}_m$, $m = 1, \dots, n$, *determined through a choice of the indifference sets in* (3) *has the property that* τ_m *is optimal in* $\mathscr{X}_m$.

PROOF. For singular losses the statement is obvious, so suppose that $q(\cdot)$ is nonsingular.

Assume $\mathscr{D}$ is q-noninformative. Then τ_1 must be optimal in $\mathscr{X}$. Given $m \in \{2, \dots, n\}$, we can choose by Proposition 10 a rule $s_m \in \mathscr{X}_m$, which is optimal in this class and independent of $(R_1, \dots, R_{m-1})$. Note that τ_m also has this independence property. Decompose and estimate the expected loss using the independence and measurability of the event $(\tau_1 > m-1)$ with respect to $(R_1, \dots, R_{m-1})$ as follows:

$$\begin{aligned}
Eq(A_{\tau_1}) &= E\big(q(A_{\tau_1})1_{(\tau_1 > m-1)}\big) + E\big(q(A_{\tau_1})1_{(\tau_1 < m)}\big)\\
&= E\big(q(A_{\tau_m})1_{(\tau_1 > m-1)}\big) + E\big(q(A_{\tau_1})1_{(\tau_1 < m)}\big)\\
&= P(\tau_1 > m-1)Eq(A_{\tau_m}) + E\big(q(A_{\tau_1})1_{(\tau_1 < m)}\big)\\
&\ge P(\tau_1 > m-1)Eq(A_{s_m}) + E\big(q(A_{\tau_1})1_{(\tau_1 < m)}\big)\\
&= E\big(q(A_{s_m})1_{(\tau_1 > m-1)}\big) + E\big(q(A_{\tau_1})1_{(\tau_1 < m)}\big) = Eq(A_\tau),
\end{aligned}$$

where $\tau = \tau_1 1_{(\tau_1 < m-1)} + s_m 1_{(\tau_1 \ge m)}$. Optimality of τ_1 implies that equality holds. By Proposition 7, we have $P(\tau_1 > m-1) > 0$; therefore $Eq(A_{\tau_m}) = Eq(A_{s_m})$ and the optimality of τ_m follows. □

We are able now to derive some conditions characterizing q-noninformative distributions. Using regular conditional probabilities, define

$$\begin{aligned} &u_m(r; X_1, \dots, X_m) \\ (10) \qquad &:= \sum_{a=r}^{n-m+r} q(a) \sum_{\mathscr{S}_m(r,a)} P(R_{m+1} = r_{m+1}, \dots, \\ &\qquad\qquad R_n = r_n \mid X_1, \dots, X_m). \end{aligned}$$

Here $\mathscr{S}_m(r, a)$ is defined as the set of all integer sequences $(r_{m+1}, \dots, r_n)$ satisfying $1 \le r_i \le i$ and $\Phi_m(r, r_{m+1}, \dots, r_n) = (a_m, \dots, a_n)$ with $a_m = a$, where Φ_m is the bijection defined in Proposition 2. Further set

$$\begin{aligned} &v_m(X_1, \dots, X_m) \\ (11) \qquad &:= \sum_{a=1}^{n} q(a) \sum_{k=m+1}^{n} \sum_{\mathscr{C}_m(k,a)} P(R_{m+1} = r_{m+1}, \dots, \\ &\qquad\qquad R_n = r_n \mid X_1, \dots, X_m), \end{aligned}$$

where $\mathscr{C}_m(k, a)$ is the set of trajectories $(r_{m+1}, \dots, r_n)$ of the relative rank process satisfying

$$r_{m+1} \in C_{m+1}, \dots, r_{k-1} \in C_{k-1}, \qquad r_k \in S_k \cup I_k$$

and

$$\Phi_{m+1}(r_{m+1}, \dots, r_n) = (a_{m+1}, \dots, a_n) \quad \text{with } a_k = a.$$

By Proposition 3(v) the functions $u_m(r; \cdot)$ and $v_m(\cdot)$ are symmetric; they are uniquely determined by both q and $\mathscr{D}$. Our characterization of q-noninformative distributions is based on these functions.

THEOREM 12. *The distribution of $(X_1, \dots, X_n)$ is q-noninformative if and only if the following relations hold for all $1 \le m \le n-1$ with probability* 1:

$$(12) \qquad r \in S_m \Rightarrow u_m(r; X_1, \dots, X_m) \le v_m(X_1, \dots, X_m),$$

$$(13) \qquad r \in C_m \Rightarrow u_m(r; X_1, \dots, X_m) \ge v_m(X_1, \dots, X_m),$$

$$(14) \qquad r \in I_m \Rightarrow u_m(r; X_1, \dots, X_m) = v_m(X_1, \dots, X_m).$$

PROOF. Let $\mathscr{D}$ be q-noninformative. Choose indifference sets $J_2 \subset I_2, \dots, J_{n-1} \subset I_{n-1}$ and consider stopping rules $\tau_1, \dots, \tau_n$ given by (3). To compute the conditional risks for stopping and continuing, we need to sum over appropriate trajectories of the rank process $R_{m+1}, \dots, R_n$. On $(R_m = r)$ we have

$$E(q(A_m)|X_1, \dots, X_m) = u_m(r; X_1, \dots, X_m),$$

because the trajectories weighted with $q(a)$ are exactly those relative rank sequences which correspond to the values $R_m = r$ and $A_m = a$. The structure of the second expectation,

$$E\big(q(A_{\tau_{m+1}})|X_1, \dots, X_m\big) = v_m(X_1, \dots, X_m),$$

is more involved, because in (11) the summation is over the set of trajectories indexed by possible values of τ_{m+1} and $A_{\tau_{m+1}}$. A particular choice of indifference sets is immaterial. By Propositions 9 and 11 the overall optimality of the rank rules is valid iff the inequalities (12) and (13) hold for appropriate events $(R_m = r)$. Because the functions $u_m(r; \cdot\,)$ and $v_m(\cdot)$ are symmetric, the restriction $(R_m = r)$ can be omitted. For $r \in I_m$, we get (12) by including this value into J_m and (13) otherwise, whence (14). □

Using this criterion in particular problems is in no way easy. For a given $q(\cdot)$, just finding the summation domains $\mathscr{S}_m$ and $\mathscr{C}_m$ involves solving the optimal stopping problem in $\mathscr{R}$. The problem is less complicated for monotone losses, because the continuation sets have the form $C_m = \{l_m, \dots, n\}$, which follows from the fact that v_m decreases and $u_m(r)$ increases in m.

Taking expectations in (12), (13) and (14), we obtain some relations which are valid for any exchangeable sequence. These relations stem from the optimality of τ_1 in $\mathscr{R}$ (see Proposition 4) and the obvious equalities

$$E\big(u_m(r; X_1, \dots, X_m)\big) = u_m(r); \qquad E\big(v_m(X_1, \dots, X_m)\big) = v_m.$$

If the random variables $u_m(r; X_1, \dots, X_m) - v_m(X_1, \dots, X_m)$ were constants with probability 1, $\mathscr{D}$ would obviously be q-noninformative. It follows from the results of the next section that at least some of these differences are nonconstant for any exchangeable distribution.

5. Impossibility results. Our next goal is to characterize q-noninformative distributions by certain conditions on the cumulative distribution of the order statistics of the random vector $(X_1, \dots, X_n)$.

Rearranging terms in (12), (13) and (14), we obtain $(n-1)n/2$ relations of the form

$$\sum_{r_{m+1}, \dots, r_n} b_{m,r}(r_{m+1}, \dots, r_n) P(R_{m+1} = r_{m+1}, \dots, R_n = r_n | X_1, \dots, X_m)$$
$$\geq 0 \text{ (or } = 0).$$

The coefficients $b_{m,r}(\cdot)$ are determined solely by the loss function and not by the distribution, because the sets $\mathscr{S}_m(r, a)$ and $\mathscr{C}_m(k, a)$ depend only on q.

These relations can be written in the form

$$(15)\qquad \sum_{r_{m+1},\dots,r_n} b_{m,r}(r_{m+1},\dots,r_n) \times P(R_{m+1}=r_{m+1},\dots,R_n=r_n,(X_1,\dots,X_m)\in B)\ge 0 \ (\text{or } =0),$$

where B runs over the family of all Borel sets in $\mathbb{R}^m$. By exchangeability, it is sufficient to consider B in the smaller class, $\mathscr{B}_>$, of Borel subsets of the space $\mathbb{R}^m_>$ of strictly descending m-tuples of reals.

Let $M_1,\dots,M_n$ be the descending sequence of order statistics of $X_1,\dots,X_n$. Given $r_{m+1},\dots,r_n$, set

$$(16)\qquad (a_{m+1},\dots,a_n)=\Phi_{m+1}(r_{m+1},\dots,r_n)$$

and let $i_1,\dots,i_m$ be the increasing sequence of integers satisfying

$$(17)\qquad \{i_1,\dots,i_m\}\cup\{a_{m+1},\dots,a_n\}=\{1,\dots,n\}.$$

For $B\in\mathscr{B}_>$, the probabilities entering (15) can be written as

$$\begin{aligned} &P(R_{m+1}=r_{m+1},\dots,R_n=r_n,(X_1,\dots,X_m)\in B)\\ &\quad=P\Big(X_{m+1}=M_{a_{m+1}},\dots,X_n=M_{a_n},\\ &\qquad\qquad M_{i_1}=X_1,\dots,M_{i_m}=X_m,\big(M_{i_1},\dots,M_{i_m}\big)\in B\Big)\\ &\quad=(n!)^{-1}P\Big(\big(M_{i_1},\dots,M_{i_m}\big)\in B\Big). \end{aligned}$$

The last step is justified via exchangeability and invariance of the event $((M_{i_1},\dots,M_{i_m})\in B)$ w.r.t. permutations of the X_i's. Now (15) takes the form

$$(18)\qquad \sum_{i_1,\dots,i_m} c_{m,r}(i_1,\dots,i_m)P\Big(\big(M_{i_1},\dots,M_{i_m}\big)\in B\Big)\ge 0\ (\text{or } =0),$$

with

$$c_{m,r}(i_1,\dots,i_m):=\sum_{r_{m+1},\dots,r_n} b_{m,r}(r_{m+1},\dots,r_n),$$

the sum being extended over the $(n-m)!$ values of relative ranks $(r_{m+1},\dots,r_n)$ found from (16) and (17).

Note that exchangeability of the X_i's imposes no restrictions on the distribution of $(M_1,\dots,M_n)$, which can be arbitrary (except that we always require that there are no ties). Given such a distribution, an exchangeable sequence can be constructed via symmetrization. The problems of the existence of q-noninformative distributions is therefore reduced to the question whether certain signed linear combinations of marginal distributions of a probability measure on $\mathbb{R}^n_>$ can be nonnegative or 0. It is certainly a deep and appealing question, which linear combinations can be nonnegative or 0. Here we will give only some examples and prove that linear dependence of $(n-1)$-dimensional marginals of $(M_1,\dots,M_n)$ is impossible.

LEMMA 13. *The equality*

$$\sum_{1\le i_1<\cdots<i_{n-1}\le n} c_i P\big((M_{i_1},\dots,M_{i_{n-1}})\in B\big)=0, \tag{19}$$

where the c_i's are constant coefficients indexed by $\{i\}=\{1,\dots,n\}\setminus\{i_1,\dots,i_{n-1}\}$, holds for all Borel sets $B\subset\mathbb{R}^{n-1}_>$ if and only if $c_1=\cdots=c_n=0$. In other words, the $(n-1)$-dimensional marginal distributions of $M_1,\dots,M_n$ are linearly independent.

PROOF. Let T be a random variable with everywhere positive density (e.g., normally distributed), independent of $(M_1,\dots,M_n)$. Because $M_1,\dots,M_n$ tie only with probability 0,

$$P(M_{m+1}<T<M_m)>0,\qquad m<n. \tag{20}$$

Substitute $B=\mathbb{R}^{n-1}$ in (19) to see that $c_1+\cdots+c_n=0$. We will prove that $c_1+\cdots+c_m=0$ along with $c_{m+1}=\cdots=c_n=0$ implies $c_m=0$. The assertion will then follow by induction in m from $m=n$ to $m=2$.

Assume $c_1+\cdots+c_m=0$ and $c_{m+1}=\cdots=c_n=0$. Since (19) holds for any *fixed* Borel set, the same equality is also valid for *random* sets $[T,\infty)^{m-1}\times(-\infty,T)^{n-m}$. This follows from Fubini's theorem and because T is independent of the M_i's. We have then by (19)

$$\begin{aligned}
0&=\sum_{i=1}^m c_iP\Big((M_{i_1},\dots,M_{i_{n-1}})\in[T,\infty]^{m-1}\times(-\infty,T)^{n-m}\Big)\\
&=\sum_{i=1}^m c_iP\Big((M_{i_1},\dots,M_{i_{n-1}})\in[T,\infty)^{m-1}\times(-\infty,T)^{n-m},\,M_i>T\Big)\\
&\quad+\sum_{i=1}^m c_iP\Big((M_{i_1},\dots,M_{i_{n-1}})\in[T,\infty)^{m-1}\times(-\infty,T)^{n-m},\,M_i<T\Big)\\
&=\left(\sum_{i=1}^m c_i\right)P(M_{m+1}<T<M_m)+c_mP(M_m<T<M_{m-1}),
\end{aligned}$$

and therefore the induction hypothesis along with (20) implies $c_m=0$. □

COROLLARY 14. *Given $n>1$, there is no exchangeable sequence $X_1,\dots,X_n$ without ties such that, for some $r\le n$, the equality $P(R_n=r|X_1,\dots,X_{n-1})=n^{-1}$ holds with probability 1.*

PROOF. Because the probabilities $P(R_n=i|X_1,\dots,X_{n-1})$, $i=1,\dots,n$, sum to 1, the equality in Corollary 14 implies that they are linearly dependent. By the argument in the beginning of the section, the $(m-1)$-dimensional marginal distributions of $(M_1,\dots,M_n)$ must be linearly dependent as well, but by Lemma 13 this is impossible. □

The case $r=1$ is of special interest for extreme value theory. The event $(R_n=1)$ occurs when the nth observations is an upper record of the X_i, that is, n is a record time. The above result says that having observed some values of an exchangeable sequence we can always draw some inference regarding the future stream of records in the sense that the posterior distribution of record times turns out to be different from the unconditional distribution (see also Remark 1 in Section 8).

COROLLARY 15. *If $n > 1$ and $X_1, \dots, X_n$ are exchangeable without ties, then R_n cannot be independent of $(X_1, \dots, X_{n-1})$.*

PROOF. Indeed, by Corollary 14 the conditional distribution of R_n cannot coincide with the unconditional distribution. □

REMARK. The system of n equalities $P(R_n = r | X_1, \dots, X_{n-1}) = n^{-1}$, $r = 1, \dots, n$, was studied by Hill (1968) in connection with some problems of Bayesian nonparametric statistics. In Section 6 of his 1968 paper he proved that this system cannot be satisfied for exchangeable X_i's by showing that this is impossible for $n = 2$ and then using backward induction. Our Lemma 13 should be regarded as a stronger impossibility result. Corollary 14 says that *none* of these n equalities can hold.

Lane and Sudderth (1978) proved the existence of *finitely additive* exchangeable distributions such that R_m is independent of $(X_1, \dots, X_{m-1})$, $m = 1, 2, \dots$. These distributions have rather exotic properties like $P(M_1 - M_n < \varepsilon) = 1 \ \forall\ \varepsilon > 0$. Therefore, in the finitely additive scenario, the first question we posed in the Introduction has a positive solution, general for all losses.

We prefer to stay within the familiar framework of the countably additive probabilities and formulate next a condition when the answer to our first question is negative. Keep in mind that we do not consider constant loss functions.

THEOREM 16. *If the indifference set I_{n-1} is nonempty, then q-noninformative distributions do not exist.*

PROOF. Assume $r \in I_{n-1}$ and there is a q-noninformative distribution. By Theorem 12, (14) holds with $m = n - 1$. Obviously,

$$v_{n-1} = \sum_{i=1}^{n} q(i) P(R_n = i | X_1, \dots, X_{n-1}).$$

The absolute rank of the rth largest value among $X_1, \dots, X_{n-1}$ equals either r or $r + 1$; therefore

$$\begin{aligned} u_{n-1}(r; X_1, \dots, X_{n-1}) &= q(r) P(R_n > r | X_1, \dots, X_{n-1}) \\ &\qquad + q(r+1) P(R_n \le r | X_1, \dots, X_{n-1}). \end{aligned}$$

Now it is easy to show that (19) holds with coefficients

$$c_i = \begin{cases} q(r+1) - q(i), & \text{for } i = 1, \dots, r, \\ q(r) - q(i), & \text{for } i = r+1, \dots, n. \end{cases}$$

In view of Lemma 13, we have $c_1 = \cdots = c_n = 0$. It follows that the loss function must be constant: a contradiction. □

REMARK. The particular values of the coefficients are not important here (provided some of these values are different from 0). Only the linear dependence has been used. This kind of argument does not work for $m < n - 1$, because lower-dimensional marginal distributions of $(M_1, \dots, M_n)$ *can* be linearly dependent for $n \geq 3$, as the following construction shows.

Take two one-dimensional distribution functions F_1 and F_3 satisfying $F_1(x) < F_3(x)$ everywhere and set $F_2 = (F_1 + R_3)/2$. Then F_2 is also a distribution function, $F_1(x) < F_2(x) < F_3(x)$, and all three are linearly dependent. Consider the unit interval with Lebesgue measure as the probability space and set $M_i(\omega) = F_i^{\leftarrow}(\omega)$, $i = 1, 2, 3$, where $F_i^{\leftarrow}$ is the generalized inverse of F_i. By the construction, $M_1 > M_2 > M_3$; hence (M_1, M_2, M_3) are order statistics of an exchangeable triple (X_1, X_2, X_3) without ties. This example corresponds to the curious equality $P(R_3 = 2|X_1) = 1/3$ a.s.; that is, whatever the first observation is, the conditional probability that it is the sample median is always $1/3$. [To prove that this equality is equivalent to $F_1 + F_3 = 2F_2$, rewrite it as

$$2P(R_3 = 2|X_1 \leq x) = P(R_3 = 1|X_1 \leq x) + P(R_3 = 3|X_1 \leq x) \qquad \forall\, x$$

and substitute the formula

$$P(R_3 = i,\, X_1 \leq x) = \big(F_j(x) + F_k(x)\big)/6,$$

which holds, via exchangeability, for all $\{i, j, k\} = \{1, 2, 3\}$.]

It is worth mentioning that neither $P(R_3 = 1|X_1) = 1/3$ a.s. nor $P(R_3 = 3|X_1) = 1/3$ a.s. is possible.

A construction of distributions in $\mathbb{R}^n_>$ with linearly dependent marginals for some values of m and n can be obtained in a similar manner, namely, by first assuming appropriate marginals and then using general methods found in Vorob'ev (1962) or Kamae, Krengel and O'Brien (1977).

We give next some examples of losses which do not admit q-noninformative distributions.

EXAMPLE 1. The problem with general loss function for $n = 2$ is reduced readily to the case of the best choice problem. The class $\mathscr{R}$ consists of two constant rules. Therefore Cover's rule described in the Introduction beats the rank rules for any exchangeable (X_1, X_2).

EXAMPLE 2. For any n, there are no q-noninformative distributions for singular losses. For $n = 3$, this is the case for $(q(1), q(2), q(3)) = (1, 0, 1)$ or, what is essentially the same, for $(q(1), q(2), q(3)) = (0, 1, 0)$.

EXAMPLE 3. For general n, the loss functions which do not admit q-noninformative distributions can be found as solutions of one of the equations in system (5) for $q(1), \dots, q(n)$, with r fixed. For $n > 3$, there exists an increasing loss function with this property; we can find such a function as a solution to (5) with $r = 2$.

6. The minimal rank problem: n even. The well-known minimal rank problem found in Chow, Moriguti, Robbins and Samuels (1964) corresponds to the loss function $q(a) = a$. Here we consider the case of n even. We show that there are randomized stopping rules based on the X_i's which are strictly better than the relative rank rules for any exchangeable distribution $\mathscr{D}$. Thus we will be able to conclude that q-noninformative distributions for linear loss functions do not exist.

First of all, note that $n/2 \in I_{n-1}$ because

$$u_{n-1}(n/2) = v_n = (n+1)/2,$$

as an easy computation shows. Let τ' be an optimal rule in $\mathscr{R}$ which stops with $R_{n-1} = n/2$. That is,

$$(\tau' \geq n-1,\, R_{n-1} = n/2) \Rightarrow (\tau' = n-1),$$

Let

$$\tau'' := \begin{cases} n, & \text{if } \tau' \geq n-1,\, R_{n-1} = n/2, \\ \tau', & \text{otherwise} \end{cases}$$

be a modification of τ' which does not stop with this rank value. By Proposition 4,

$$EA_{\tau'} = EA_{\tau''}, \tag{21}$$

and τ'' is optimal as well.

The method we used in Lemma 13 extends the idea of a splitting variable due to Cover (1987) and suggests the following randomized stopping rule. Let T be an arbitrary real random variable independent of $X_1, \dots, X_n$, with everywhere positive density. Set

$$\rho := \begin{cases} n, & \text{if } \tau' \geq n-1,\, R_{n-1} = n/2 \text{ and } X_{n-1} < T, \\ \tau', & \text{otherwise.} \end{cases} \tag{22}$$

By definition, ρ coincides with τ' if $X_{n-1} > T$ and coincides with τ'' if $X_{n-1} < T$.

Let $N := \#\{i\colon i \leq n,\, X_i > T\}$ be the number of exceedances over T. Permutations of the X_i's do not change the value of N; therefore N is independent of $(A_1, \dots, A_n)$. Furthermore,

$$P(N = k) > 0, \qquad k = 0, 1, \dots, n, \tag{23}$$

because T falls with positive probability in any of the $n+1$ intervals obtained by partitioning of the real line by the order statistics $M_1, \dots, M_n$.

For the event $(N > n/2,\, R_{n-1} = n/2)$, we have $X_{n-1} > T$; therefore $N > n/2$ implies $\rho = \tau'$. Similarly, $N < n/2$ implies $\rho = \tau''$. It follows that

$$E\big(A_\rho 1_{(N > n/2)}\big) = E\big(A_{\tau'} 1_{(N > n/2)}\big). \tag{24}$$

The independence of N and the A_i's implies that N is independent of the R_i's and hence of τ''. Hence by (21) we have

$$E(A_\rho 1_{(N<n/2)}) = E(A_{\tau''} 1_{(N<n/2)}) \tag{25}$$
$$= EA_{\tau''} \cdot P(N < n/2) = E(A_{\tau'} 1_{(N<n/2)}).$$

We have also $\rho = \tau'$ for the events $(N = n/2,\ \tau' < n-1)$, $(N = n/2,\ \tau' \geq n-1,\ R_{n-1} \neq n/2)$ and $(N = n/2,\ \tau' \geq n-1,\ R_{n-1} = n/2,\ X_{n-1} > T)$.

Consider the event

$$H := (N = n/2,\ \tau' \geq n-1,\ R_{n-1} = n/2,\ X_{n-1} < T).$$

Assuming that H occurs, the number of exceedances over T among $X_1, \dots, X_{n-1}$ is $n/2 - 1$, because X_{n-1} falls below T and has relative rank $n/2$. It follows that X_n exceeds T, and thus $A_{n-1} = n/2 + 1$ holds along with $A_n \leq n/2$; that is, proceeding to X_n is better than stopping with X_{n-1}. Clearly, for this event $\rho = n$, $\tau' = n-1$ and thus ρ beats τ'.

Putting this all together along with (24) and (25), we obtain

$$\begin{aligned} EA_\rho &= E(A_\rho 1_{H^c}) + E(A_\rho 1_H) \\ &= E(A_{\tau'} 1_{H^c}) - E(A_n 1_H) \\ &< E(A_{\tau'} 1_{H^c}) + E(A_{n-1} 1_H) = EA_{\tau'}. \end{aligned}$$

The inequality is strict, because, by (23) and Proposition 7,

$$\begin{aligned} P(H) &= P(N = n/2,\ \tau' > n-2,\ R_{n-1} = n/2,\ R_n \leq n/2) \\ &= \frac{1}{2(n-1)} P(N = n/2) P(\tau' > n-2) > 0. \end{aligned}$$

The rule ρ always is strictly better than all relative rank rules. For any *given* $\mathscr{D}$, there are also nonrandomized rules in $\mathscr{X}$ which improve strictly the rank rules: take t such that the interval between the $(n/2)$th and the $(n/2+1)$th order statistics of $X_1, \dots, X_n$ covers t with positive probability and substitute this t into (22) instead of T.

REMARKS (i) The advantage of ρ over the rank rules can be arbitrarily small for some exchangeable sequences. To see this, take $\mathscr{D}$ concentrated near the diagonal $x_1 = \cdots = x_n$. Then the X_i's are clustered in a small interval, and T with probability close to 1 does not split them.

(ii) For the minimal rank problem and n odd, we have $I_{n-1} = \varnothing$, because there is no middle value for R_{n-1}; thus the method does not work. Samuels (1994) discussed the failure of similar rules for $n = 3$.

(iii) Our construction of a randomized rule which beats the rank rules is easily generalized to all increasing loss functions satisfying $I_{n-1} \neq \varnothing$.

7. Googol. Ferguson (1989) interpreted M. Gardner's Googol as a game version of the best choice problem, with the loss function $q(k) = 1_{\{k>1\}}$. In this game, Player 1 selects an exchangeable distribution for the $\{X_i\}$, and Player 2 picks a stopping rule from $\mathscr{X}$. Any q-noninformative distribution is a minimax strategy of Player 1.

It is well known that, for $n > 2$, $I_1 = \cdots I_{n-1} = \varnothing$, $S_1 = \cdots = S_{d-1} = \varnothing$ and $S_d = \cdots = S_{n-1} = \{1\}$, where d is the integer satisfying

$$\frac{1}{d} + \frac{1}{d+1} + \cdots + \frac{1}{n-1} < 1 < \frac{1}{d-1} + \frac{1}{d} + \cdots + \frac{1}{n-1},$$

$d/n \to e^{-1}$, $n \to \infty$. The inequalities (13) and (12) describing q-noninformative distributions turn into

$$P(R_{m+1} \neq 1, \dots, R_n \neq 1 | X_1, \dots, X_m) \\ \leq P(\text{exactly one of } R_d, \dots, R_n \text{ is } 1 | X_1, \dots, X_m), \qquad m = 1, \dots, d-1,$$

and

$$P(R_{m+1} \neq 1, \dots, R_n \neq 1 | X_1, \dots, X_m) \\ \geq P(\text{exactly one of } R_{m+1}, \dots, R_n \text{ is } 1 | X_1, \dots, X_m), \quad m = d, \dots, n-1,$$

respectively. The first system of inequalities is equivalent to the single inequality with $m = d - 1$, because the right-hand side there does not depend on m and the events on the left-hand side are decreasing in m. This description does not seem useful to find a q-noninformative distribution explicitly.

Tedious but straightforward calculations along the lines of Gnedin (1994) show that, for $n \geq 3$, these inequalities are satisfied if $X_1, \dots, X_n | \theta$ are iid, uniformly distributed on $[0, \theta]$, and the parameter θ has a prior distribution with density

$$f(\theta) = \frac{\varepsilon}{2}\theta^{\varepsilon-1}1_{(0,1)}(\theta) + \frac{\varepsilon}{2}\theta^{-\varepsilon-1}1_{[1,\infty)}(\theta), \tag{26}$$

where $\varepsilon > 0$ should be sufficiently small.

A more delicate analysis shows that, given N, $\varepsilon = \varepsilon(N)$ can be selected the same for all $n = 3, \dots, N$. In other words, there exists an infinite exchangeable sequence such that each subsequence of length $2 < n \leq N$ is q-noninformative. A game-theoretic meaning of this result is that Player 1 has a minimax strategy which does not depend on the particular value of n, subject to a restriction like $n < 10^{100}$. This is an improvement of the result found in Gnedin (1994), where the mixing density was dependent on n.

8. Final remarks.

REMARK 1. Samuels (1981) and Berezovskiy and Gnedin (1984) used the invariance principle to show that, in the best choice problem, the optimal rank rule is minimax. The invariance ideas can be applied also to the following version of Googol: Player 1 picks θ and samples the X_i's independently from the uniform distribution on $[0, \theta]$, while Player 2 picks a stopping rule. In this form, the game remains essentially the same if we rescale with a positive factor the values on the θ and x scales. It is therefore natural to expect that both players have minimax strategies, invariant w.r.t. these transformations.

By Corollary 14, no $\mathscr{D}$ satisfies

$$P(R_{m+1} = 1 | X_1, \dots, X_m) = (m+1)^{-1}. \tag{27}$$

However, mixing uniforms on $(0, \theta)$ with the infinite measure $\theta^{-1}\, d\theta$ yields a sequence which in a sense does satisfy (27) for all m. The *formal* posterior distribution of θ after one observation has already a probability density, and (27) can be given a proper interpretation via the Bayesian prior-to-posterior transformation. Recall that $\theta^{-1}\, d\theta$ is the invariant measure on the multiplicative group of positive reals. Thus the corresponding improper $\mathscr{D}$ can be regarded as a formal minimax strategy of Player 1. The exchangeable distribution related to the prior (26) might be interpreted as a proper approximation to the improper invariant minimax strategy.

REMARK 2. One might seek in higher dimensions for improper exchangeable distributions satisfying the condition "R_n is independent of $X_1, \dots, X_{n-1}$." For example, consideration of the group of monotone affine transformations of $\mathbb{R}$ leads, for $n = 3$, to the improper distribution with density $(\max(x_1, x_2, x_3) - \min(x_1, x_2, x_3))^{-2}$. For general n, such distributions seem to be unknown.

REMARK 3. In the best choice problem with $n > 2$, the set of all q-noninformative distributions, which are representable as mixtures of products, form a proper subset of the set of all q-noninformative distributions. The q-noninformative distributions, which are not mixtures, can be found among those described in Gnedin (1995). In this respect the mixtures do not play a special role.

REMARK 4. As $\mathscr{D}$ varies, the minimal risk $V_{\mathscr{X}}$ assumes the values between the rank stopping value $V_{\mathscr{R}}$ and $\min_{k \le n} q(k)$, the lower bound being attained, for example, in the case when $X_1, \dots, X_n$ is a random permutation of $\{1, \dots, n\}$. Let U denote the value of $V_{\mathscr{X}}(\mathscr{D})$ in the iid case. It is evident that U is none of the extremes in the range of the $V_{\mathscr{X}}$-values.

An extremal property of the iid case appears if we restrict attention to the $\mathscr{D}$'s representable as mixtures of product distributions. Clearly, U is then a lower bound. Indeed, consider the values $X_1, \dots, X_n$ coming from a mixture as a two-stage choice. A univariate continuous distribution is chosen according to some "prior," and then the X_i's are sampled independently from the resulting distribution. If the result of the first choice is known and can be used in the stopping rules, then we are exactly in the iid case, where the minimal risk does not depend on the particular distribution of the X_i's. On the other hand, we cannot do better if the result of the first choice is unknown. This argument has an interesting consequence. If $V_{\mathscr{R}}(\mathscr{D}) < U$, then $\mathscr{D}$ cannot be represented as a mixture of continuous product distributions with identical factors.

Acknowledgments. We thank the referees for their suggestions.

REFERENCES

BEREZOVSKIY, B. A. and GNEDIN, A. V. (1984). *The Best Choice Problem*. Nauka, Moscow.

CHOW, Y. S., MORIGUTI, S., ROBBINS, H. and SAMUELS, S. M. (1964). Optimal selection based on relative rank (the "secretary problem"). *Israel J. Math.* **2** 81–90.

COVER, T. M. (1987). Pick the largest number. In *Open Problems in Communication and Computation* (T. M. Cover and B. Gopinath, eds.) 152. Springer, New York.

FERGUSON, T. S. (1989). Who solved the secretary problem? *Statist. Sci.* **4** 282–289.

GNEDIN, A. V. (1994). A solution to the game of Googol. *Ann. Probab.* **22** 1588–1595.

GNEDIN, A. V. (1995). On a class of exchangeable sequences. *Statist. Probab. Lett.* **25** 351–355.

GNEDIN, A. V. and KRENGEL, U. (1995). A stochastic game of optimal stopping and order selection. *Ann. Appl. Probab.* **5** 310–321.

HILL, B. M. (1968). Posterior distribution of percentiles: Bayes theorem for sampling from a finite population. *J. Amer. Statist. Assoc.* **63** 677–691.

HILL, T. P. and KENNEDY, D. P. (1992). Sharp inequalities for optimal stopping with rewards based on ranks. *Ann. Appl. Probab.* **2** 503–517.

KAMAE, T., KRENGEL, U. and O'BRIEN, G. L. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Probab.* **5** 899–912.

LANE, D. and SUDDERTH, W. (1978). Diffuse models for sampling and predictive inference. *Ann. Statist.* **6** 1318–1336.

SAMUELS, S. M. (1981). Minimax stopping rules when the underlying distribution is uniform. *J. Amer. Statist. Assoc.* **76** 188–197.

SAMUELS, S. M. (1989). Who will solve the secretary problem? *Statist. Sci.* **4** 289–291.

SAMUELS, S. M. (1991). Secretary problems. In *Handbook of Sequential Analysis* (B. K. Ghosh and P. K. Sen, eds.). Dekker, New York.

SAMUELS, S. M. (1994). Sufficiently noninformative priors for the secretary problem; the case: $n = 3$. Preprint, Dept. Statistics, Purdue Univ.

SILVERMAN, S. and NADÁS, A. (1992). On the game of googol as *the* secretary problem. *Contemp. Math.* **125** 77–83.

VOROB'EV, N. N. (1962). Consistent families of measures and their extensions. *Theory Probab. Appl.* **7** 153–169.

UNIVERSITY OF GÖTTINGEN
INSTITUTE OF MATHEMATICAL STOCHASTICS
LOTZESTRASSE 13
37083 GÖTTINGEN
GERMANY
E-MAIL: gnedin@namu01.gwdg.de
krengel@namu01.gwdg.dr