# OPTIMAL SELECTION BASED ON RELATIVE RANKS OF A SEQUENCE WITH TIES*

by

Gregory Campbell
Purdue University

Mimeograph Series #81-40

Department of Statistics
Division of Mathematical Sciences

September, 1981

*Research supported by National Science Foundation Grant MCS78-02895.

## ABSTRACT

The optimal selection of a maximum of a sequence with the possibility of ties is considered. The object is to examine each observation in the sequence of known length n and, based only on the relative rank among predecessors, to either stop and select it as a maximum or continue without recall. Rules which maximize the probability of correctly selecting a maximum from a sequence with ties are investigated. These include rules which randomly break ties, rules which discard tied observations, and minimax rules based on the atoms of a discrete distribution function. If the sequence is random from F, a random distribution function from a Dirichlet process prior with nonatomic parameter, optimal rules are developed. The limiting behavior of these rules is studied and compared with other rules. The selection of the parameter of the Dirichlet process regulates the ties.

KEY WORDS: Best choice problem; Dirichlet process prior; Optimal stopping; Relative ranks; Secretary problem.

# OPTIMAL SELECTION BASED ON RELATIVE RANKS OF A SEQUENCE WITH TIES*

by

Gregory Campbell
Purdue University

## 1. Introduction.

Let $X_1, X_2, \ldots, X_n$ denote a sequence of independent, identically distributed random variables from distribution function F. The optimal stopping problem of interest is to observe the X's one at a time and to stop at the maximum value, with no recall permitted. In this paper it is assumed that F is not continuous, so the maximum is not necessarily unique and the optimal strategy allows for ties.

If F is continuous but unknown, and the loss function is 0-1 (0 if maximum selected, 1 if not), the optimal rule depends only on the relative ranks of the observations. This problem has been called the (best choice) "secretary problem"; see Gilbert and Mosteller (1966) for its history and solution. It will be referred to hereinafter as the continuous problem, referring to the continuous distribution function. Yang (1975), Govindarajulu (1975), and Lorenzen (1979) have discussed realistic generalizations of this problem. Numerous researchers have generalized the loss function to a function of the relative rank selected. Prior information in the form of relative ranks in a previous sample has been treated by Campbell and Samuels (1981). If the X's are observable (not just the relative ranks), prior information has been incorporated by Stewart (1978), Samuels (1981), and Campbell (1977), among others.

---

*Research supported by National Science Foundation Grant MCS78-02895.

The realistic generalization considered here is to allow ties; in particular, assume F is not continuous or, alternately, the original observations have been grouped into ordered categories. The loss function is 0-1 (best choice problem).

Section 2 treats several approaches to ties in this problem. One strategy is to randomly break ties and employ the optimal rule for the continuous problem. Another intuitive strategy is to discard all previous ties. If at $X_r$ there are m distinct values among $X_1,\ldots,X_r$, reduce n to n-(r-m) and reduce r to m and employ the optimal rule for the continuous problem. This will be referred to as the reduced rule. A third approach assumes that an upper bound k on the number of atoms of the discrete distribution function F is known. Minimax rules are obtained and the behavior of the rules as k tends to infinity is investigated.

In Section 3 F is assumed to be randomly selected by a Dirichlet process prior on the space of discrete distribution functions. The parameter of this prior due to Ferguson (1973) is assumed nonatomic; the resulting strategy depends on the parameter only through the mass C of the measure and depends also on the number of $X_1,\ldots,X_r$ tied at the relative maximum and the number of distinct values of $X_1,\ldots,X_r$. Rules are displayed for various n and C and the probability of correctly selecting a maximum computed.

The limiting behavior of the rules developed in Section 3 is investigated in Section 4. As C approaches infinity it is proved that the rules approach the reduced rules of Section 2. The usefulness of the limit as C tends to 0 is also discussed; the resultant rule is to continue if all are tied and to stop at a relative maximum otherwise.

The application of these rules based on the Dirichlet process to a random sample from an arbitrary unknown discrete distribution function is discussed. The estimation of C if it is unknown is considered. Simplified rules utilizing the Dirichlet process and reduced rules are also investigated and the limiting behavior of these rules studied.

## 2. Randomized rules, reduced rules and minimax strategies.

In the event of ties among the X's, a simple mechanism to break them is to employ a random device to order the X's tied at the relative maximum. Having resolved the ties, the optimal strategy for the continuous problem is then employed. Intuitively, ties with such a strategy improve the probability of selecting the maximum without ties in that one can select the maximum by either stopping or continuing at a tied relative maximum value.

A second strategy is to discard all previous ties at each stage. For $X_r$, let m denote the number of distinct values of $X_1,\ldots,X_r$. Then the number of tied X's is r-m (all ties are included, not just ties at the relative maximum value). The reduced strategy is then to reduce n to $n' = n-(r-m)$ and r to $r' = r-(r-m) = m$ and use the optimal strategy for the continuous problem.

The final approach of this section requires the additional information that the underlying distribution function F is discrete with a known upper bound k on the number of atoms. If a prior were placed on the atoms one could obtain the Bayes rules. However, the approach here is to develop minimax strategies.

This additional information can have a dramatic influence on the optimal rule, especially for $k < n$. For example, suppose $k = 2$. Then if $X_1 < X_2$, the optimal procedure is to stop at $X_2$. It is straightforward in this example to calculate the probability of failing to select the maximum value with the strategy of continuing if all are tied and of stopping at the first X larger than some previous X. If x and y denote the atoms with $x < y$, the value y is not selected if all y's precedes all x's. If $p_1$ is the probability for atom x and $p_2 = 1-p_1$ for atom y, the probability of not correctly selecting y is:

$$p_2p_1^{n-1} + p_2^2p_1^{n-2} + \ldots + p_2^{n-1}p_1.$$

The symmetry of this function in $p_1$ and $p_2$ implies here that it is maximized for $p_1 = \frac{1}{2}$. Thus the probability of selecting the maximum value y is bounded below by $1-(n-1)2^{-n}$.

Contrast this strategy for $k = 2$ with the randomized strategy in the case $n = 3$. The strategy above has probability of incorrect selection of $p_2p_1^2 + p_2^2p_1$. If one stops at $X_2$ if $X_1 = X_2$ the probability is $2p_1^2p_2$. The randomized rule is with probability q to stop at $X_2$ if $X_1 = X_2$. For $q = \frac{1}{2}$ this has probability of incorrect selection of $\frac{3}{2}p_1^2p_2 + \frac{1}{2}p_1p_2^2$. The least favorable configurations which maximize these three probabilities are $p_1 = .5$, $p_1 = \frac{2}{3}$, and $p_1 = .608$ with maxima $\frac{1}{4}$, $\frac{8}{27}$, and .2641, respectively. The minimax rule is non-randomized and continues at $X_2$ if $X_1 = X_2$.

For $k > 2$, suppose $n = 3$. For simplicity initially consider $k = 3$, with probabilities $p_1$, $p_2$ and $p_3$ for the ordered atoms. Let r denote the subscript of the current relative maximum X, $\ell$ the number of

$X_1,\ldots,X_r$ tied at the value of $X_r$, and m the number of distinct values in $X_1,\ldots,X_r$. Use ($\ell$, m, r) to denote this configuration. It becomes clear for n = 3, k = 3 to continue at (1, 1, 1) and stop at a relative maximum at (1, 2, 2). The strategy question is whether to continue or not at ($\ell$, m, r) = (2, 1, 2); i.e., if $X_1 = X_2$. Continuing at (2, 1, 2) results in probability of incorrect selection of

$$p_1p_2^2 + p_1p_3^2 + p_2p_3^2 + p_1^2p_2 + p_1^2p_3 + p_2^2p_3 + 3p_1p_2p_3.$$

This function is symmetric in $p_1 = p_2 = p_3$ and has maximum $\frac{1}{3}$ at $p_1 = p_2 = p_3 = \frac{1}{3}$. Stopping at (2, 1, 2) yields probability of incorrect selection of

$$2p_1^2p_2 + 2p_1^2p_3 + 2p_2^2p_3 + 3p_1p_2p_3.$$

Since this is not symmetric, $p_1 = p_2 = p_3 = \frac{1}{3}$ is not a maximum; hence the maximum exceeds $\frac{1}{3}$. Thus, the minimax strategy is to continue at (2, 1, 2). This argument extends easily to $k \geq 4$. The probability of maximal selection can be easily shown to be $\frac{1}{2} + (2k)^{-1}$. The probability of maximal selection for the continuous problem (no ties) is $\frac{1}{2}$. Thus, the probability improvement shrinks for an upper limit of k is realized, yet shrinks to 0 as k tends to infinity.

For $n \geq 4$ the rules quickly become complex. The optimal strategy at $X_r$ depends on the configuration of ties of the X's, but not on the order of arrangement of $X_1,\ldots,X_r$. In particular, it is not sufficient to record only ties at the relative maximum value.

Fix n and let k increase. It is clear that the probability of correct selection based on the minimax strategy with k or fewer atoms decreases as k increases. This is so because a least favorable $(p_1,\ldots,p_k)$ for k atoms is included with an additional atom with probability 0 in the (k+1) atoms case. As noted in the discussion of randomization to break ties, ties increase the probability of correct selection. Hence the least favorable distribution will tend to minimize the probability of ties. For $k(\geq m)$ atoms, the upper bound on the probability of no ties in n observations is $k(k-1)\ldots(k-n+1)/k^n$ (at $p_i = \frac{1}{k}$), which tends to 1 as k tends to infinity. In the case of no ties, the rule reverts to the optimal strategy for the classical continuous "secretary problem". In addition, as the probability of ties tends to zero, the strategy for such ties becomes irrelevant. A conjecture based on the simple examples of this section as well as additional investigation is that, for $k \geq n$, the minimax rules based on k or fewer atoms are reduced rules.

As a last case, if the number k of atoms is fixed and known and n tends to infinity, the probability of correct selection tends to 1. A more interesting unexplored problem is to introduce an interview cost in this situation, as Govindarajulu (1975) does in the continuous case.

## 3. *Ties and the Dirichlet process.*

In this section it is assumed F is randomly selected from the set of discrete distributions by a Dirichlet process prior of Ferguson (1973). It is assumed in this section that the measure is nonatomic with known mass C. The optimal strategy and probability of correct selection are then calculated.

The Dirichlet process was introduced by Ferguson (1973) and was employed immediately as a means of incorporating prior information into nonparametric problems (see Ferguson (1973) for a definition of the process and numerous applications). Some Bayesians have found difficulty with this approach in that the prior has frequently been inappropriate due to its concentration on the discrete distributions (see Berk and Savage (1979) for an elementary proof of this discreteness for the Dirichlet process). Unfortunately, the resulting procedures had few (in any) competitors which incorporated such general prior information. The naturally occurring ties in the Dirichlet samples have been exploited only rarely (see Campbell and Hollander (1978) for an example). Most nonparametric applications are complicated by a dual role for the mass C of the measure $\alpha$ of the Dirichlet process. First, Ferguson (1973) suggests that it is the "degree of belief" or "confidence" in the shape of the probability measure $\alpha/C$ which regulates a single observation, with the limiting case as C tends to infinity corresponding to "complete confidence". The second role of C is to regulate the ties; for example, $\frac{1}{C+1}$ is the probability two observations from a Dirichlet process with nonatomic measure are equal. Previous applications have utilized the first but generally ignored the second.

In this optimal stopping problem, the ties are of primary interest. The concentration of the Dirichlet process prior on the discrete distributions is quite appropriate. If the measure of the process is nonatomic, C can be used to regulate the ties. If the measure is otherwise unspecified the first role of C is unfulfilled and C exclusively regulates the ties. The optimal strategy is now developed for this situation.

Let $X_1,\ldots,X_r$ denote the observed X's at the rth stage, of which m are distinct. Let $z_1 <\ldots< z_m$ denote the ordered values of these X's and $k_i$ the number of the rX's tied at $z_i$, for $i = 1,\ldots,m$. Define $z_0 = -\infty$ and $z_{m+1} = \infty$. The next theorem calculates the probability distribution for $X_{r+1}$.

<u>Theorem 3.1</u>. If $X_1,\ldots,X_r$, $X_{r+1}$ is a sample from a Dirichlet process with nonatomic measure with mass C then for the above notation based on $X_1,\ldots,X_r$

a) $P\{X_{r+1} = z_i\} = k_i/(C+r)$. $\quad i = 1,\dots,m$

b) $P\{X_{r+1} \in (z_i, z_{i+1})\} = C/[(m+1)(C+r)]$. $\quad i = 0, 1,\dots,m$.

Proof. The proof of a) proceeds directly from application of the theorem of Ferguson (1973) that a Dirichlet process with parameter $\alpha$ conditional on a sample $X_1,\dots,X_r$ is again a Dirichlet process with updated measure $\alpha + \sum_{i=1}^{r} \delta_{X_i}$ where $\delta_z$ denotes the measure concentrating unit mass at the point $z$.

For b) assume without loss of generality that the measure $\alpha$ is concentrated uniformly on the interval $[0, C]$, such that $\alpha([0, y]) = y$ for $y \in [0, C]$. Then the distribution of $(z_1,\dots,z_m)$ is uniform over the region $0 = z_0 \le z_1 \le z_2 \le \dots \le z_m \le z_{m+1} = C$. Therefore,

$$P\{X_{r+1} \in (z_0, z_1) | X_{r+1} \ne z_i \text{ for } i = 1,\dots,m\}$$

$$= \int_{z_1<\dots<z_m} \cdots \int z_1 \; m! \; dz_1 \dots dz_m / C^{m+1} = \frac{1}{m+1}.$$

But by a) $P\{X_{r+1} \ne z_i \text{ for } i = 1,\dots,m\} = C/(C+r)$. Thus $P\{X_{r+1} \in (0, z_1)\} = C/[(m+1)(C+r)]$. For $i = 1,\dots,m$

$$P\{X_{r+1} \in (z_i, z_{i+1}) | X_{r+1} \ne z_i \text{ for } i = 1,\dots,m\}$$

$$= \int_{z_1} \cdots \int_{z_m} (z_{i+1} - z_i) \; m! \; dz_1 \dots dz_m / C^{m+1}.$$

A straightforward change of variables gives $\frac{1}{m+1}$ as for $i = 0$ and the theorem is proved. □

Application of this theorem to the optimal stopping problem at the rth stage, with $X_r$ a relative maximum, gives

$$P\{X_{r+1} \geq X_r\} = [(k_m + C)/(m+1)]/(C+r).$$

It is important to note that this probability depends on C, r, m, and $k_m$, but not explicitly on $k_1,\ldots,k_{m-1}$. For convenience denote the number $k_m$ of ties at the relative maximum by $\ell$. The probability $p_s$ of correct selection by stopping at $X_r$ and the probability $p_c$ of correct selection by continuing at $X_r$ with the optimal strategy are now derived iteratively.

Theorem 3.2. If $X_1,\ldots,X_n$ is from a Dirichlet process with nonatomic parameter with mass C the probability of correctly selecting the maximum value for $X_1,\ldots,X_n$ by stopping at a relative maximum $X_r$ is given by:

$$p_s(\ell, m, r; C) = \frac{r-\ell}{C+r} p_s(\ell, m, r+1; C) + \frac{\ell}{C+r} p_s(\ell+1, m, r+1; C) + \frac{mC/(m+1)}{C+r} p_s(\ell, m+1, r+1; C). \tag{3.1}$$

Proof. It is necessary to calculate $P\{X_{r+1},\ldots,X_n \leq z_m\}$. If $X_{r+1} \leq z_m$ then either $X_{r+1} = z_m$, $X_{r+1} = z_i$ $i < m$ or $X_{r+1} \in (z_i, z_{i+1})$ for $i \leq m - 1$. Since $\ell$ denotes the number of ties at the relative maximum, the probability $X_{r+1} = z_i$ for $i = 1,\ldots,m - 1$ is $(r-\ell)/(C+r)$. If $X_{r+1}$ is tied with a previous X that is not a relative maximum, the probability $X_r$ is a relative maximum is the probability $X_{r+2},\ldots, X_n$ are also $\leq z_m$ which is $p_s(\ell, m, r+1; C)$. If $X_{r+1} = z_m$, which occurs with probability $\ell/(C+r)$, the probability $X_{r+2},\ldots,X_n$ are $\leq z_m$ is $p_s(\ell+1, m, r+1; C)$. If $X_{r+1} \in (z_i, z_{i+1})$ for some $i = 0,\ldots,m-1$, which occurs with probability $mC/[(m+1)\ (C+r)]$, the conditional probability $X_r$ is a maximum is $p_s(\ell, m+1, r+1; C)$. □

It is possible to calculate $p_s$ at stage r by backward iteration from stage n-1 together with the initial condition

$$p_s(\ell, m, n-1; C) = [(n-1) + mC/(m+1)]/(C + n - 1).$$

In that the initial condition does not depend on $\ell$, neither does $p_s$ for the rth stage. Thus, for convenience the notation $p_s(m,r;C)$ will be used for this probability.

The function $p_s(m, r; C)$ is increasing in m. This is proved by backward induction on r. It is trivial for $r = n - 1$. Further, if it is true for $r = s + 1$ then

$$\begin{aligned} p_s(m+1, s; C) - p_s(m,s;C) &= \frac{s}{C+s} [p_s(m+1, s+1; C) - p_s(m, s+1; C)] \\ &+ \frac{mC/(m+1)}{C+s} [p_s(m+2, s+1; C) - p_s(m+1, s+1; C)] \\ &+ C/[(C+s)(m+1)(m+2)]\, p_s(m+2; s+1; C). \end{aligned}$$

Applying the induction hypothesis for $r = s + 1$ proves monotonicity of $p_s(m, s; C)$ in m for $r = s$. Similar arguments prove $p_s(m, r; C)$ is increasing in C and $p_s(m+i, r+i; C)$ is increasing in i for $i < n - r$.

Theorem 3.3. If $X_1,\ldots,X_n$ is from a Dirichlet process with nonatomic parameter with mass C the probability of correctly selecting the maximum value for $X_1,\ldots,X_n$ by continuing at $X_r$ with the optimal strategy is given by:

$$\begin{aligned} p_c(\ell, m, r; C) &= \frac{r-\ell}{C+r} p_c(\ell, m, r+1; C) + \frac{mC/(m+1)}{C+r} p_c(\ell, m+1, r+1; C) \\ &+ \frac{\ell}{C+r} p_M(\ell+1, m, r+1; C) + \frac{C/(m+1)}{C+r} p_M(1, m+1, r+1; C) \end{aligned} \tag{3.2}$$

where

$$p_M(\ell, m, r; C) = \max(p_s(m, r; C), p_c(\ell, m, r; C)).$$

Proof. There are four cases for $X_{r+1}$. If $X_{r+1} = z_i$ for some $i < m$, which occurs with probability $(r-\ell)/(C+r)$, one must continue with correct selection probability $p_c(\ell, m, r+1; C)$. If $X_{r+1} \in (z_i, z_{i+1})$ for some $i < m$, which has probability $mC/[(m+1)(C+r)]$, again one must continue with probability $p_c(\ell, m+1, r+1; C)$. If $X_{r+1} = z_m$, with probability $\ell/(C+r)$, it is unclear whether to stop or continue so use the better strategy of the two with probability $p_M(\ell+1, m, r+1; C)$. Lastly if $X_{r+1} > z_m$ with probability $C/[(m+1)(C+r)]$, the maximum conditional probability is $p_M(1, m+1, r+1; C)$. □

The initial condition for the backward iteration on r is obtained from Theorem 3.1:

$$p_c(\ell, m, n-1; C) = [\ell+(C/(m+1))]/(C+r).$$

For any C and n one can calculate $p_s$ and $p_c$ for any $(\ell, m, r)$ configuration and hence determine whether to stop or continue for any $X_r$.

The expression $p_c(\ell, m, r; C)$ is increasing in $\ell$. Again a backward induction argument is used. For $r = n - 1$, it is trivially true. If it is true for $r = s + 1$ then

$$p_c(\ell+1, m, s; C) - p_c(\ell, m, s; C) = \frac{r-\ell-1}{C+r}[p_c(\ell+1, m, s+1; C) -$$

$$p_c(\ell, m, s+1; C)] + \frac{mC/(m+1)}{C+r}[p_c(\ell+1, m+1, s+1; C) -$$

$$p_c(\ell, m+1, s+1; C)] + (\ell+1)\, p_M(\ell+1, m, s+1; C) - \ell$$

$$p_M(\ell, m, s+1; C) - p_c(\ell, m, s+1; C).$$

By the induction hypothesis the first two terms on the right are positive. Further, since $p_c(\ell, m, s+1; C) \leq p_M(\ell, m, s+1; C)$ the remaining terms on the right are $\geq (\ell+1)[p_M(\ell+1, m, s+1; C) - p_M(\ell, m, s+1; C)]$ which by induction is positive. Thus $p_c(\ell+1, m, s; C) \geq p_c(\ell, m, s; C)$; the induction is complete.

The ramification of this result is as follows. Since $p_s$ does not depend on $\ell$ and $p_c$ increases in $\ell$, if there is a configuration $(\ell, m, r)$ for which $p_c > p_s$ then the optimal strategy is to continue for all $(\ell', m, r)$ configuration, where $\ell \leq \ell' \leq r - m + 1$.

The value of C can influence the decision to stop or to continue. For example, for $n = 5$,

$$p_s(1, 2, 2; C) = (24 + \frac{52}{3}C + \frac{9}{2}C^2 + \frac{2}{5}C^3)/[(C+2)(C+3)(C+4)];$$

$$p_c(1, 2, 2; C) = (18 + 16C + \frac{9}{2}C^2 + \frac{13}{30}C^3)/[(C+2)(C+3)(C+4)].$$

Therefore at $X_2(>X_1)$ the optimal rule is to stop if $C < C_0$ and to continue if $C > C_0$ where $C_0 \approx 7.93$. The complete rule for $n = 5$ is to continue also at (1, 1, 1), (2, 1, 2), (3, 1, 3) and to stop otherwise. At (4, 1, 4) it doesn't matter whether one stops or continues.

Optimal rules are presented for $n \leq 8$ and for various C in Table 3.1. In particular, the $(\ell, m, r)$ configurations at which one continues are listed. The configuration $(r, 1, r)$ appears in all cases, in that the optimal strategy is always to continue at $(r, 1, r)$ if $r \leq n - 2$, with indifference at $r = n - 1$. It is interesting to note that the strategies for $C = 100$ are just the reduced rules of Section 2.

Table 3.2 contains the probabilities of correct selection for the optimal strategies; it is just a table of $p_C(1, 1, 1; C)$ for various values of n and C. For comparison the optimal probability for the continuous case as in Gilbert and Mosteller (1966) is presented in the last column of the table.

| n | C = 0.1 | C = 1.0 | C = 10 | C = 100 |
|---|---|---|---|---|
| 3 | (r,1,r) | (r,1,r) | (r,1,r) | (r,1,r) |
| 4 | (r,1,r) | (r,1,r) | (r,1,r) | (r,1,r) |
| 5 | (r,1,r) | (r,1,r) | (r,1,r)(1,2,2) | (r,1,r)(1,2,2) |
| 6 | (r,1,r) | (r,1,r) | (r,1,r)(2,2,3)(1,2,2) | (r,1,r)(1,2,3)<br>(2,2,3)(1,2,2) |
| 7 | (r,1,r) | (r,1,r)(2,2,3)(3,2,4) | (r,1,r)(2,2,4)(3,2,4)<br>(1,2,3)(2,2,3)(1,2,2) | (r,1,r)(1,2,4)<br>(2,2,4)(3,2,4)<br>(1,2,3)(2,2,3)(1,2,2) |
| 8 | (r,1,r) | (r,1,r)(4,2,5)(3,2,4)<br>(2,2,3)(1,2,3)(2,1,2) | (r,1,r)(3,2,5)(4,2,5)<br>(1,2,4)(2,2,4)(3,2,4)<br>(1,2,3)(1,3,3)(1,2,2) | (r,1,r)(1,2,5)(2,2,5)<br>(3,2,5)(4,2,5)(1,2,4)<br>(2,2,4)(3,2,4)(1,2,3)<br>(1,3,3)(2,2,3)(1,2,2) |

TABLE 3.1

($\ell$,m,r) CONFIGURATIONS AT WHICH THE
OPTIMAL STRATEGY FOR DIRICHLET SAMPLES IS
TO CONTINUE ($1 \leq r \leq n-2$)

| n \ C | 0.1 | 1.0 | 10 | 100 | 1000 | continuous case |
|---|---|---|---|---|---|---|
| 2 | .9545 | .7500 | .5455 | .50495 | .5005 | .5000 |
| 3 | .9545 | .7500 | .5455 | .50495 | .5005 | .5000 |
| 4 | .9615 | .7691 | .5241 | .4657 | .4591 | .4583 |
| 5 | .9682 | .7802 | .5004 | .4411 | .4341 | .4333 |
| 6 | .9733 | .7825 | .5050 | .4371 | .4287 | .4279 |
| 7 | .9769 | .7805 | .5048 | .4254 | .4154 | .4143 |
| 8 | .9793 | .7897 | .5007 | .4213 | .4110 | .4098 |
| 9 | .9809 | .7948 | .5066 | .4191 | .4073 | .4060 |
| 10 | .9819 | .7982 | .5097 | .4135 | .4002 | .3987 |
| 12 | .9835 | .8047 | .5156 | .4124 | .3973 | .3955 |
| 15 | .9849 | .8132 | .5260 | .4100 | .3916 | .3894 |
| 20 | .9858 | .8242 | .5435 | .4102 | .3870 | .3842 |

TABLE 3.2

OPTIMAL PROBABILITIES OF CORRECT SELECTION
FOR DIRICHLET PROCESS SAMPLES

## 4. Asymptotic rules and discussion.

The limiting behavior of the strategies of Section 3 are examined.

First, consider the limit as C tends to infinity. Then equation (3.1) becomes:

$$\lim_{C\to\infty} p_S(m, r; C) = \frac{m}{m+1} \lim_{C\to\infty} p_S(\ell, m+1, r+1; C).$$

Repeated application and the known limit for the initial stage r = n - 1 yields:

$$\lim_{C\to\infty} p_S(m, r; C) = \frac{m}{m+1} \frac{m+1}{m+2} \cdots \frac{m+n-r-1}{m+n-r} = \frac{m}{m+n-r} . \qquad (4.1)$$

Note that this is merely the probability of correct selection by stopping for the continuous version of the problem where ties are completely ignored. This is the appropriate probability for the reduced strategy of Section 2; namely, discard the r-m tied values, reducing n to n' = n - (r-m) and reducing r to m, and obtain the analogous probability for the classic untied strategy.

Similarly, equation (3.2) becomes:

$$\lim_{C\to\infty} p_C(\ell, m, r; C) = \frac{m}{m+1} \lim_{C\to\infty} p_C(\ell, m+1, r+1; C) + \frac{1}{m+1} \lim_{C\to\infty} p_M(1, m+1, r+1; C).$$

For $\ell$=1, $\lim_{C\to\infty} p_C(1, m+i, r+i; C)$ is decreasing in i for $0 \le i \le n - r - 1$. This is so because

$$\lim_{C\to\infty} p_C(1, m, r; C) - \lim_{C\to\infty} p_C(1, m+1, r+1; C) =$$

$$\frac{1}{m+1} \lim_{C\to\infty} p_M(1, m+1, r+1; C) - \frac{1}{m+1} \lim_{C\to\infty} p_C(1, m+1, r+1; c) \ge 0.$$

On the other hand $p_S(m+i, r+i; C)$ is increasing in i for $0 \le i \le n - r - 1$. Thus there exists a unique s depending on n and r such that the optimal rule is to stop at (1, s, r) and hence also for (1, s+i, r+i) for $1 \leq i \leq n - r - 1$.

Therefore for $m \geq s$ the repeated application of (4.1) and (4.2) yields

$$\lim_{C\to\infty} p_c(\ell, m, r; C) = \frac{1}{m+1}\,\frac{m+1}{n-(r-m)} + \frac{m}{m+1}\,\frac{1}{m+2}\,\frac{m+2}{n-(r-m)}$$

$$+\ldots+ \frac{m}{m+1}\,\frac{m+1}{m+2}\cdots\frac{m+n-r-2}{m+n-r-1}\,\frac{1}{m+n-r}\,\frac{m+n-r}{n-(r-m)}$$

$$= \frac{1}{n-(r-m)}\left[\frac{m}{m} + \frac{m}{m+1} +\ldots+ \frac{m}{m+n-r-1}\right].$$

But this is just equation (2a-2) of Gilbert and Mosteller (p. 39) for the optimal probability of continuing in the continuous case for a sample of size $n' = n-(r-m)$ at stage $r' = m$; i.e., it is the probability for the reduced strategy.

This then proves that as C tends to infinity the rules from the Dirichlet process approach of the reduced rules of Section 2. Further, the optimal probability of selecting a maximum therefore converges as C tends to infinity to the optimal probability for the continuous distribution function case. The convergence rate can be observed in Table 3.2.

The behavior as C tends to 0 is also quite interesting. Sethuraman and Tiwari (1981) point out in general that the limit as C tends to zero in the Dirichlet process does not correspond to the non-informative prior. As C tends to zero, all observations are tied at $X_1$, a random value. In this optimal stopping application, the only role of C is to regulate ties, so this limit should be interpretable in this application.

Taking the limit as C tends to 0, (3.1) becomes

$$\lim_{C\to 0} p_s(m, r; C) = \frac{r}{r} \lim_{C\to 0} p_s(m, r+1; C).$$

Repeated application of this along with the limiting initial condition implies

$$\lim_{C\to 0} p_s(m, r; C) = 1.$$

Equation (3.2) becomes

$$\lim p_c(\ell, m, r; C) = \frac{r-\ell}{r} \lim_{C\to 0} p_c(\ell, m, r+1; C) + \frac{\ell}{r} \lim_{C\to 0} p_M(\ell+1, m, r+1; C)$$

$$= \frac{r-\ell}{r} \lim_{C\to 0} p_C(\ell, m, r+1; C) + \frac{\ell}{r} .$$

Since $\lim_{C\to 0} p_C(\ell, m, n-1; C) = \ell/(n-1)$, this becomes

$$\lim_{C\to 0} p_c(\ell, m, r; C) = \frac{\ell}{r} [1 + \frac{r-\ell}{r+1} + \frac{(r-\ell)(r-\ell+1)}{(r+1)\ (r+2)} +\ldots+ \frac{(r-\ell)\ldots(n-\ell-2)}{(r+1)\ldots\ (n-1)}].$$

Note that $\lim_{C\to 0} p_c(\ell, m, r; C) < 1$ unless $\lim_{C\to 0} p_c(\ell, m, r+i; C) = 1$ for all $1 \leq i \leq n - r - 1$. But $\lim p_c(\ell, m, n-1; C) = \frac{\ell}{n-1}$ which is $< 1$ unless $m = 1$.

The conclusion is that the limiting strategy as $C \to 0$ is to stop at a relative maximum at $X_r$ with configuration $(\ell, m, r)$ if $\ell < r$. If $\ell = r$ a more detailed analysis is necessary:

$$\lim_{C\to 0} [p_c(r, 1, r; C) - p_s(1, r; C)] =$$

$$\lim_{C\to 0} \{\frac{r}{c+r} [p_M(r+1, 1, r+1; C) - p_s(1, r+1; C)] + \frac{C/2}{C+r} [p_c(r, 2, r+1; C)$$

$$- p_s(2, r+1; C)] + \frac{C/2}{C+r} p_M(1, r+1, r+1; C)\} .$$

Since $p_M(1, r+1, r+1; C) \geq p_s(2, r+1; C)$ because $p_s(m, r)$ is increasing in m, it follows that $\lim_{C\to 0} p_C(r, 1, r; C) \geq \lim_{C\to 0} p_s(r, 1, r; C)$; i.e., the optimal strategy for $r \leq n - 2$ is to continue at $(r, 1, r)$.

An inspection of Table 3.1 reveals that for $n \leq 8$ these limiting rules are achieved at $C = 0.1$.

In that the optimal strategy based on the Dirichlet process is admittedly complex, depending at the rth stage on $\ell$ and m, a simplified strategy is proposed. Let D denote the number of distinct values in the sequence of fixed length n from a Dirichlet process with nonatomic parameter with mass C. If ties are completely ignored, the search for the optimal rule is then similar to the approach of Gianinni-Pettitt (1979), who found an optimal rule for a sequence of random length N, where N has a discrete uniform distribution. For simplicity here, the reduced rules of Section 2 are utilized if the number of distinct observations were known. The strategy is to continue at $X_r$ with a relative maximum of m distinct observations if

$$P\{D \geq d_c | m,r\} > P\{D \leq d_s | m,r\}$$

where the reduced rule continues at $X_r$ with m distinct observations if $D \geq d_c$, stops if $D \leq d_s$, is indifferent if possibly $d_s < D < d_c$; the strategy stops otherwise. This strategy does not depend on $\ell$, the number tied at the relative maximum.

Consider the simplified rule for n = 8, C = 10. To easily obtain the reduced rules, the optimal strategy for the continuous problem for $n \leq 8$ is displayed; for a sequence of length d, this rule stops at the first relative maximum at stage s or later (from Gilbert and Mosteller (1966):

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| d | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| s | 1 | 1,2 | 2 | 2 | 3 | 3 | 3 | 4. |

Note indifference in stopping or continuing for d = 2. For a relative maximum at (m,r) = (1,6), if D = 3 the reduced rule, discarding the first five observations, continues; if D = 2 the reduced rule is indifferent;

if D = 1 the reduced rule stops. Now $P\{D = 3|(1,6)\} =$ $P[X_6, X_7 \notin \{X_1, X_2, X_3, X_4, X_5\}, X_6 \neq X_7] = (\frac{C}{C+6})(\frac{C}{C+7}) = 0.3676$ and $P\{D = 1|(1,6)\} = (\frac{6}{C+6})(\frac{7}{C+7}) = 0.1544$. Since 0.3674 > .1544, the simplified rule continues at (1,6). At (m,r) = (2,4), $d_c = 5$ and $P\{D \geq 5|(2,4)\} = \frac{10^3}{14 \cdot 15 \cdot 16 \cdot 17} [10+4+5+6+7] = 0.5602$, so the rule continues at (2,4). At (2,3), $d_c = 5$ and $P\{D \geq 5\ (2,3)\} = 0.6114$, so again the rule continues. For (m,r) = (2,5), $d_s = 4$ and $P\{D \leq 4\ (2,5)\} = 0.7549$; the rule is to stop. The complete simplified rule is to continue if r < 3, and also at (2,3) (2,4) and (1,r) (for $r \leq 6$) and to stop otherwise. The optimal Dirichlet rule for n = 8, C = 10 from Table 3.1 continues when the simplified rule does and also for $(\ell,m,r) = (3,2,5)$ and (4,2,5). The optimal rule is approximated by the simplified rule.

The limiting behavior of these simplified rules are easily obtained. As C tends to infinity, $P\{D = n-r+m|(m,r)\}$ tends to 1. The rule is then just the reduced rule at the mth stage for a sample of size n-r+m. As C tends to zero, $P\{D = m|(m,r)\}$ tends to 1 in which case one stops. These limiting cases are as in the limiting cases for the optimal Dirichlet rules.

If the parameter C is unknown for the Dirichlet process, it can be estimated based on a previous sample. If n' denotes the size of the previous sample and D' = d' is the observed number of distinct observations out of n', then the maximum likelihood approach of Ewens (1972) can be employed: the estimate is the unique solution for C of the implicit equation:

$$d' = C(\frac{1}{C} + \frac{1}{C+1} + \ldots + \frac{1}{C+n'-1}).$$

For large n' the consistent estimator D'/log n' of Korwar and Hollander (1973) is more easily computed.

An important question concerns the use of the Dirichlet process to generate rules for the maximal selection from a sequence with ties if the sequence is <u>not</u> from a Dirichlet process. If the sequence arises as a random sample from a discrete distribution, the use of Dirichlet process is particularly appropriate in that the prior from the Dirichlet process concentrates its mass on discrete distributions. The selection of a nonatomic measure as the parameter for the Dirichlet process is necessitated by the fact that only relative ranks of the sequence are observable. In addition, the resulting rules from the nonatomic measure depend only on the mass C of this measure. The limiting behavior of the Dirichlet rules as C tends to 0 and tends to infinity strongly reinforces the use of C as a mechanism to model the ties in the sequence. Therefore, prior information concerning the incidence of ties would be incorporated into the procedure by the choice of C. In particular, the distribution of D for the Dirichlet model is given by Antoniak (1974):

$$P\{D = d\} = S_n^{(d)} C^d / C^{[n]} \qquad d = 1,\ldots,n,$$

where $C^{[n]} = C(C+1)\ldots(C+n-1)$ and $S_n^{(d)}$ is the absolute value of the Stirling number of the first kind. Since the distribution of D for the random sample from the discrete distribution depends on the probabilities at the atoms, C could be chosen to approximate the prior belief of one's distribution of D.

# REFERENCES


Antoniak, C.E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. Ann. Statist. **2**, 1152-1174.

Berk, R. H. and Savage, I. R. (1979). Dirichlet processes produce discrete measures: an elementary proof. Contributions to Statistics - Jaroslav Hájek Memorial Volume, J. Jurečkova (editor), Dordrecht, Holland. 25-32.

Campbell, G. (1977). The maximum of a sequence with prior information. Purdue University Department of Statistics Mimeograph Series #485 Revised.

Campbell, G. and Hollander, M. (1978). Rank order estimation with the Dirichlet prior. Ann. Statist. **6**, 142-153.

Campbell, G. and Samuels, S. M. (1981). Choosing the best of the current crop. Adv. in Appl. Probab. **13** (in press).

Ewens, W. J. (1972). The sampling theory of selectively neutral alleles. Theor. Pop. Biol. **3**, 87-112.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. Ann. Stat. **1**, 209-230.

Gianini-Pettitt, J. (1979). Optimal selection based on relative ranks with a random number of individuals. Adv. in Appl. Probab. **11**, 720-736.

Gilbert, J. P. and Mosteller, F. (1966). Recognizing the maximum of a sequence. J. Am. Statist. Assoc. **61**, 35-73.

Govindarajulu, Z. (1975). The secretary problem: optimal selection with interview cost. Proceedings of the Symposium on Statistics and Related Topics.

Korwar, R. M. and Hollander, M. (1973). Contributions to the theory of Dirichlet processes. Ann. Probab. **1**, 705-711.

Lorenzen, T. J. (1979). Generalizing the secretary problem. Adv. Appl. Probab. **11**, 384-396.

Samuels, S. M. (1981). Minimax stopping rules when the underlying distribution is uniform. J. Am. Statist. Assoc. **76**, 188-197.

Sethuraman, J. and Tiwari, R. C. (198~~1~~ 1982). Convergence of Dirichlet measures and the interpretation of their parameter. Third Purdue Symposium on Decision Theory and Related Topics ~~(to appear)~~. vol. 2 pp 305-315 ed J.O. Berger S.S. Gupta

Stewart, T. J. (1978). Optimal selection from a random sequence with learning of the underlying distribution. J. Am. Statist. Assoc. **73**, 775-780.

Yang, W. C. K. (1974). Recognizing the maximum of a random sequence based on relative rank with backward solicitation. J. Appl. Probab. **11**, 504-512.

# OPTIMAL SELECTION BASED ON RELATIVE RANKS OF A SEQUENCE WITH TIES

by

GREGORY CAMPBELL*, Purdue University and National Institutes of Health

Abstract

The optimal selection of a maximum of a sequence with the possibility of ties is considered. The object is to examine each observation in the sequence of known length n and, based only on the relative rank among predecessors, to either stop and select it as a maximum or continue without recall. Rules which maximize the probability of correctly selecting a maximum from a sequence with ties are investigated. These include rules which randomly break ties, rules which discard tied observations, and minimax rules based on the atoms of a discrete distribution function. If the sequence is random from F, a random distribution function from a Dirichlet process prior with nonatomic parameter, optimal rules are developed. The limiting behavior of these rules is studied and compared with other rules. The selection of the parameter of the Dirichlet process regulates the ties.

BEST CHOICE PROBLEM; DIRICHLET PROCESS PRIOR; OPTIMAL STOPPING; RELATIVE RANKS; SECRETARY PROBLEM

*Present address: Division of Computer Research and Technology, National Institutes of Health, Bethesda MD 20~~014~~205 USA. Research supported in part by National Science Foundation Grant MCS78-02895.

## 1. Introduction

Let $X_1, X_2, \ldots, X_n$ denote a sequence of independent, identically distributed random variables from distribution function F. The optimal stopping problem of interest is to observe the X's one at a time and to stop at the maximum value, with no recall permitted. In this paper it is assumed that F is not continuous, so the maximum is not necessarily unique and the optimal strategy allows for ties.

If F is continuous but unknown, and the loss function is 0-1 (0 if maximum selected, 1 if not), the optimal rule depends only on the relative ranks of the observations. This problem has been called the (best choice) "secretary problem"; see Gilbert and Mosteller (1966) for its history and solution. It will be referred to hereinafter as the continuous problem, referring to the continuous distribution function F. Yang (1974), Govindarajulu (1975), Lorenzen (1979), Petrucelli (1981), Rasmussen (1975), Smith (1975) and Smith and Deely (1975) have discussed realistic generalizations of this problem. The assumption of n fixed and known has been relaxed by Presman and Sonin (1972), Rasmussen and Robbins (1975), Gianini-Pettitt (1979), and Irle (1980). Prior information in the form of relative ranks in a previous sample has been treated by Campbell and Samuels (1981). If the X's are observable (not just the relative ranks), prior information has been incorporated by Stewart (1978), Samuels (1981), and Campbell (1983), among others.

The realistic generalization considered here is to allow ties; in particular, assume F is not continuous or, alternately, the original observations have been grouped into ordered categories. The loss function is 0-1 (best choice problem).

Section 2 treats several approaches to ties in this problem. One strategy is to randomly break ties and employ the optimal rule for the continuous problem. Another intuitive strategy is to discard all previous ties. If at $X_r$ there are m distinct values among $X_1, \ldots, X_r$, reduce n to $n-(r-m)$ and reduce r to m and employ the optimal rule for the continuous problem. This will be referred to as the reduced rule. A third approach assumes that an upper bound k on the number of atoms of the discrete distribution function F is known. Minimax rules are obtained and the behavior of the rules as k tends to infinity is investigated.

In Section 3 F is assumed to be randomly selected by a Dirichlet process prior on the space of discrete distribution functions. The parameter of this prior due to Ferguson (1973) is assumed nonatomic; the resulting strategy depends on the parameter only through the mass C of the measure and depends also on the number of $X_1,\ldots,X_r$ tied at the relative maximum and the number of distinct values of $X_1,\ldots,X_r$. Rules are displayed for various n and C and the probability of correctly selecting a maximum is computed.

The limiting behavior of the rules developed in Section 3 is investigated in Section 4. As C approaches infinity it is proved that the rules approach the reduced rules of Section 2. The usefulness of the limit as C tends to 0 is also discussed; the resultant rule is to continue if all are tied and to stop at a relative maximum otherwise. The application of these rules based on the Dirichlet process to a random sample from an arbitrary unknown discrete distribution function is discussed. The estimation of C if it is unknown is considered. Simplified rules utilizing the Dirichlet process and reduced rules are also investigated and the limiting behavior of these rules studied.

## 2. Randomized rules, reduced rules and minimax strategies

In the event of ties among the X's, a simple mechanism to break them is to employ a random device to order the X's tied at the relative maximum. Having resolved the ties, the optimal strategy for the continuous problem is then employed. Intuitively, ties with such a strategy improve the probability of selecting the maximum without ties in that one can select the maximum by either stopping or continuing at a tied maximum value.

A second strategy is to discard all previous ties at each stage. At stage r, let m denote the number of distinct values of $X_1,\ldots,X_r$ . Then the number of tied X's is r−m (all ties are included, not just ties at the relative maximum value). The reduced strategy is then to reduce $n$ to $n'=n-(r-m)$ and $r$ to $r'=r-(r-m)=m$ and use the optimal strategy for the continuous problem.

The final approach of this section requires the additional information that the underlying distribution function $F$ is discrete with a known upper bound $k$ on the number of atoms. If a prior were placed on the atoms one could obtain the Bayes rules. However, the approach here is to develop minimax strategies.

This additional information can have a dramatic influence on the optimal rule, especially for $k<n$. For example, suppose $k=2$. Then if $X_1<X_2$, the optimal procedure is to stop at $X_2$. It is straightforward in this example to calculate the probability of failing to select the maximum value with the strategy of continuing if all are tied and of stopping at the first $X$ larger than some previous $X$. If $x$ and $y$ denote the atoms with $x<y$, the value $y$ is not selected if all $y'$s precede any of the $x'$s. If $p_1$ is the probability for atom $x$ and $p_2=1-p_1$ for atom $y$, the probability of not correctly selecting $y$ is:

$$p_2p_1^{n-1}+p_2^2p_1^{n-2}+\dots+p_2^{n-1}p_1.$$

The symmetry of this function in $p_1$ and $p_2$ implies here that it is maximized for $p_1=1/2$. Thus the probability of selecting the maximum value $y$ is bounded below by $1-(n-1)2^{-n}$.

Contrast this strategy for $k=2$ with the randomized strategy in the case $n=3$. The strategy of stopping at $X_2$ if $X_1<X_2$ and continuing otherwise has probability of incorrect selection of $p_2p_1^2+p_2^2p_1$. If one stops at $X_2$ if $X_1\leq X_2$ the probability is $2p_1^2p_2$. These are maximized at $p=1/2$ and $p=2/3$ with probabilities of incorrect selection of $1/4$ and $8/27$, respectively. The randomized rule is to stop at $X_2$ with probability $q$ if $X_1=X_2$. The probability of incorrect selection is

$$p_1^2p_2+qp_1^2p_2+(1-q)p_1p_2^2 = p_1^2p_2+p_1p_2^2+qp_1p_2(2p_1-1).$$

For $p_1 \geq 1/2$ and any $q \in [0,1]$, this incorrect selection probability is $\geq p_1^2 p_2 + p_1 p_2^2$ and hence has maximum $\geq 1/4$. Therefore, the minimax rule is non-randomized and continues at $X_2$ if $X_1 = X_2$.

For $k>2$, suppose $n=3$. For simplicity initially consider $k=3$, with probabilities $p_1$, $p_2$ and $p_3$ for ordered atoms. At stage r, with $X_r$ a relative maximum, let $l$ denote the number of $X_1,\ldots,X_r$ tied at the value of $X_r$, and m the number of distinct values in $X_1,\ldots,X_r$. Let $(l,m,r)$ denote this configuration. The optimal strategy for $n=3$, $k=3$ is to continue at $(1,1,1)$ and to stop at a relative maximum at $(1,2,2)$. The question is whether to continue or not at $(l,m,r)=(2,1,2)$; i.e., if $X_1=X_2$. The strategy which continues at $(2,1,2)$ results in an overall probability of incorrect selection of

$$p_1 p_2^2 + p_1 p_3^2 + p_2 p_3^2 + p_1^2 p_2 + p_1^2 p_3 + p_2^2 p_3 + 3 p_1 p_2 p_3.$$

This function is symmetric in $p_1 = p_2 = p_3$ and has maximum $1/3$ at $p_1 = p_2 = p_3 = 1/3$. Stopping at $(2,1,2)$ yields an overall probability of incorrect selection of

$$2p_1^2 p_2 + 2p_1^2 p_3 + 2p_2^2 p_3 + 3p_1 p_2 p_3.$$

Since this is not symmetric, $p_1 = p_2 = p_3 = 1/3$ is not a maximum; hence the maximum exceeds $1/3$. Thus, the minimax strategy is to continue at $(2,1,2)$. This argument extends easily to $k \geq 4$. The probability of maximal selection can be easily shown to be $1/2 + (2k)^{-1}$. The probability of maximal selection for the continuous problem (no ties) is $1/2$. Thus, the probability improvement for a small number of atoms is substantial, yet shrinks to 0 as k tends to infinity.

For $n \geq 4$ the rules quickly become complex. The optimal strategy at $X_r$ depends on the configuration of ties of the X's, but not on the order of arrangement of $X_1,\ldots,X_r$ provided $X_r$ is still a relative maximum. In particular, it is not sufficient to record only ties at the relative maximum value.

Fix n and let k increase. It is clear that the probability of correct selection based on the minimax strategy with k or fewer atoms decreases as k increases. This is so because a least favorable $(p_1,\ldots,p_k)$ for k atoms is included with an additional atom with probability 0 in the $(k+1)$ atoms case. As noted in the discussion of randomization to break ties, ties increase the probability of correct selection. Hence the least favorable distribution will tend to minimize the probability of ties. For $k(\geq n)$ atoms, the upper bound on the probability of no ties in n observations is $k(k-1)\ldots(k-n+1)/k^n$ (at $p_i=1/k$), which tends to 1 as k tends to infinity. In the case of no ties, the rule reverts to the optimal strategy for the classical continuous "secretary problem". In addition, as the probability of ties tends to zero, the strategy for such ties becomes irrelevant. A conjecture based on the simple examples of this section as well as additional investigation is that, for $k\geq n$, the minimax rules based on k or fewer atoms are reduced rules.

As a last case, if the number k of atoms is fixed and known and n tends to infinity, the probability of correct selection tends to 1. A more interesting unexplored problem is to introduce an interview cost in this situation, as Govindarajulu (1975) does in the continuous case.

## 3. Ties and the Dirichlet process

In this section prior distribution function information in the form of a Dirichlet process on the real line is used to model the ties. The Dirichlet process of Ferguson (1973) induces a prior distribution on the set of all distribution functions; in fact, with probability 1 it concentrates on the set of all discrete distribution functions (see Berk and Savage (1979) for an elementary proof of this discreteness). This promotes ties and suggests that the Dirichlet process may prove quite useful in treating the above selection problem with ties. The single parameter of the Dirichlet process is a measure $\alpha$ on the real line. If C denotes the mass of the measure $\alpha$, the shape $\alpha(.)/C$ corresponds to the prior distribution and the mass C to the "degree of belief" (Ferguson) in this shape. If the measure is atomic at a point x (so $\alpha(\{x\}) > 0$), then with probability 1, the discrete distribution function randomly selected by the Dirichlet process jumps at x. In the above selection model, it is assumed in this section that $X_1, X_2, \ldots, X_n$ are independent identically distributed from a distribution function randomly selected by the process. Ties can arise in two ways: first, at the atoms (if any) of the measure and secondly, at nonatomic points if not all the mass of the measure is concentrated at atoms. This second source of ties is due to contagion which arises from the discreteness of the selected distribution function even if the measure is non-atomic. The purpose of investigating the Dirichlet process in selection problems is not because this contagion exhibited by Dirichlet processes arises so frequently in nature but primarily as a device to incorporate prior information into the problem. The Dirichlet process is particularly suited to a selection problem with ties in that ties occur naturally in such a model.

Consider the best choice selection problem in which the actual values $X_1, \ldots, X_n$ from the Dirichlet process are observable (not just the relative ranks) and the measure $\alpha$ is known and non-atomic. This has been treated by Campbell (1983); optimal rates have been obtained and the probability of correct selection calculated. If $\alpha$ is known and not non-atomic and the X's observable the same backward iterative approach of Campbell can be easily adapted to find for the particular measure the optimal strategy and probability of correct selection. In both cases (non-atomic or not), as C tends to infinity for constant shape, one approaches the optimal strategy for the case of independent, identically distributed observations from a known distribution function; in the non-atomic case this is just the strategy of the "full information game" of Gilbert and Mosteller (1966). For small values of C, the rule allows for the ties which can occur even in the non-atomic case.

The purpose of this section is to use the Dirichlet process for the selection problem in which only the relative ranks of the X's are observable. If atoms occur in the measure of the process, the number and location of the atoms (although unknown) play a major role in the optimal strategy. Hence only non-atomic measures are considered below. Since only relative ranks are observable the shape of the non-atomic measure will be shown to not be important and hence can be assumed to be unknown. The mass C of the measure will be required since in the absence of atoms, it alone controls the probability of ties. Rarely have the naturally occurring ties in the Dirichlet process been fully exploited; see Campbell and Hollander (1978) for one example. Furthermore, in applications of the Dirichlet process usually C is required to play dual roles of reflecting the confidence or "degree of belief" in the shape of the measure as well as regulating the ties. The role of confidence is generally stressed in applications and the aspect of tie regulation usually ignored. Here, since the shape of the measure is unknown, it is only the latter which is important. Large values of C reduce the probability of ties; small values inflate it; for example, for a Dirichlet process with non-atomic measure of mass C, the probability that two observations from the process are tied is $1/(C+1)$. This suggests the use of the rules developed below for the selection problem from an unknown distribution function (not from a Dirichlet process), where C is used to reflect the prior belief in the probability of ties; this is discussed in the next section.

Let $X_1,\ldots,X_n$ denote a sample of size n from a Dirichlet process with non-atomic measure of known mass C. For $r<m$, let $X_1,\ldots X_r$ denote the observed X's at the rth stage, of which m are distinct. Let $z_1<\ldots<z_m$ denote the ordered values of these X's and $k_i$ the number of the r X's tied at $z_i$, for $i=1,\ldots,m$. Define $z_0=-\infty$ and $z_{m+1}=\infty$. The next theorem calculates the probability distribution for $X_{r+1}$.

__Theorem 3.1__. If $X_1,\ldots,X_r,X_{r+1}$ is a sample from a Dirichlet process with non-atomic measure with mass C then for the above notation, conditional on $X_1,\ldots,X_r$

a) $P\{X_{r+1}=z_i\} = k_i/(C+r)$. $\qquad i=1,\ldots,m$

b) $P\{X_{r+1}\epsilon(z_i,z_{i+1})\} = C/[(m+1)(C+r)]$. $\qquad i=0,1,\ldots,m$.

__Proof__. The proof of a) proceeds directly from application of the theorem of Ferguson (1973) that a Dirichlet process with parameter $\alpha$, conditional on a sample $X_1,\ldots,X_r$, is again a Dirichlet process, with updated measure $\alpha + \sum_{i=1}^{r} \delta_{X_i}$ where $\delta_z$ denotes the measure concentrating unit mass at the point z.

For b) assume without loss of generality that the measure is concentrated uniformly on the interval $[0,C]$ such that $\alpha([0,y]) = y$ for $y \in [0,C]$. Then the distribution of $(z_1,\ldots,z_m)$ is uniform over the region $0=z_0 \le z_1 \le z_2 \le \ldots \le z_m \le z_{m+1}=C$. Therefore,

$$P\{X_{r+1} \in (z_0,z_1) | X_{r+1} \neq z_i \text{ for } i = 1,\ldots,m\}$$

$$= \int_{z_1 < \ldots < z_m} \ldots \int z_1 \, m! \, dz_1 \ldots dz_m / C^{m+1} = 1/m+1.$$

But by a) $P\{X_{r+1} \neq z_i \text{ for } i=1,\ldots,m\} = C/(C+r)$. Thus $P\{X_{r+1} \in (z_0,z_1)\} = C/[(m+1)(C+r)]$. For $i=1,\ldots,m$

$$P\{X_{r+1} \in (z_i,z_{i+1}) | X_{r+1} \neq z_i \text{ for } i=1,\ldots,m\}$$

$$= \int_{z_1} \ldots \int_{z_m} (z_{i+1}-z_i) \, m! \, dz_1 \ldots dz_m / C^{m+1}.$$

A straightforward change of variables gives $1/(m+1)$ as for $i=0$ and the theorem is proved.

Application of this theorem to the optimal stopping problem at the rth stage, with $X_r$ a relative maximum, gives

$$P\{X_{r+1} \ge X_r\} = [k_m + C/(m+1)]/(C+r).$$

It is important to note that this probability depends on C, r, m, and $k_m$, but not expicitly on $k_1,\ldots,k_{m-1}$. For convenience denote the number $k_m$ of ties at the relative maximum by $l$. The probability $p_s$ of correct selection by stopping at $X_r$ and the probability $p_c$ of correct selection by continuing at $X_r$ with the optimal strategy are now derived iteratively.

<u>Theorem 3.2</u>. If $X_1,\ldots,X_n$ is from a Dirichlet process with non-atomic parameter with mass C the probability of correctly selecting the maximum value for $X_1,\ldots,X_n$ by stopping at a relative maximum $X_r$ is given by:

$$\begin{aligned} p_S(l,m,r;C) &= \frac{r-l}{C+r}\, p_S(l,m,r+1;C) + \frac{l}{C+r}\, p_S(l+1,m,r+1;C) \\ &\quad + \frac{mC/(m+1)}{C+r}\, p_S(l,m+1,r+1;C). \end{aligned} \tag{3.1}$$

<u>Proof</u>. It is necessary to calculate $P\{X_{r+1},\ldots,X_n \le z_m\}$. If $X_{r+1} \le z_m$ then either $X_{r+1}=z_m$ or $X_{r+1}=z_i$, for $i<m$, or $X_{r+1}\epsilon(z_i,z_{i+1})$ for $i\le m-1$. Since $l$ denotes the number of ties at the relative maximum, the probability $X_{r+1}=z_i$ for $i=1,\ldots m-1$ is $(r-l)/(C+r)$. If $X_{r+1}$ is tied with a previous X that is not a relative maximum, the probability $X_r$ is a relative maximum is the probability $X_{r+2},\ldots,X_n$ are also $\le z_m$ which is $p_S(l,m,r+1;C)$. If $X_{r+1}=z_m$, which occurs with probability $l/(C+r)$, the probability $X_{r+2},\ldots,X_n$ are $\le z_m$ is $p_S(l+1,m,r+1;C)$. If $X_{r+1}\epsilon(z_i,z_{i+1})$ for some $i=0,\ldots,m-1$, which occurs with probability $mC/[(m+1)(C+r)]$, the conditional probability $X_r$ is a maximum is $p_S(l,m+1,r+1;C)$. The proof is complete.

It is possible to calculate $p_S$ at stage r by backward iteration from stage n-1 together with the initial condition

$$p_S(l,m,n-1;C) = [(n-1)+mC/(m+1)]/(C+n-1).$$

In that the initial condition does not depend on $l$, neither does $p_S$ for the $r^{th}$ stage. Thus, for convenience the notation $p_S(m,r;C)$

will be used for this probability.

The function $p_S(m,r;C)$ is increasing in m. This is proved by backward induction on r. It is trivial for $r=n-1$. Further, if it is true for $r=s+1$ then

$$p_S(m+1,s;C)-p_S(m,s;C) = \frac{s}{C+s}[p_S(m+1,s+1;C)-p_S(m,s+1;C)]$$

$$+\frac{mC/(m+1)}{C+s}[p_S(m+2,s+1;C)-p_S(m+1,s+1;C)]$$

$$+\frac{C}{[(C+s)(m+1)(m+2)]}p_S(m+2;s+1;C).$$

Applying the induction hypothesis for $r=s+1$ proves monotonicity of $p_S(m,s;C)$ in m for $r=s$. Similar arguments prove $p_S(m,r;C)$ is decreasing in C and $p_S(m+i,r+i;C)$ is increasing in i for $i<n-r$.

Theorem 3.3. If $X_1,\ldots X_n$ is from a Dirichlet process with non-atomic parameter with mass C the probability of correctly selecting the maximum value for $X_1,\ldots,X_n$ by continuing at $X_r$ with the optimal strategy is given by:

$$p_c(l,m,r;C) = \frac{r-l}{C+r}p_c(l,m,r+1;C)+\frac{mC/(m+1)}{C+r}p_c(l,m+1,r+1;C)$$

(3.2)

$$+\frac{l}{C+r}p_M(l+1,m,r+1;C)+\frac{C/(m+1)}{C+r}p_M(l,m+1,r+1;C)$$

where

$$p_M(l,m,r;C) = \max[p_S(m,r;C),p_c(l,m,r;C)].$$

Proof. There are four cases for $X_{r+1}$. If $X_{r+1}=z_i$ for some $i<m$, which occurs with probability $(r-l)/(C+r)$, one must continue with correct selection probability $p_c(l,m,r+1;C)$. If $X_{r+1}\epsilon(z_i,z_{i+1})$ for some $i<m$,

which has probability $mC/[(m+1)(C+r)]$, again one must continue, with probability $p_c(l,m+1,r+1;C)$. If $X_{r+1}=z_m$, with probability $l/(C+r)$, it is unclear whether to stop or continue to use the better strategy of the two with probability $p_M(l+1,m,r+1,C)$. Lastly, if $X_{r+1}>z_m$ with probability $C/[(m+1)(C+r)]$, the maximum conditional probability is $p_M(l,m+1,r+1;C)$.

The initial condition for the backward iteration on r is obtained from Theorem 3.1:

$$p_c(l,m,n-1;C) = \{l+[C/(m+1)]\}/(C+r).$$

For any C and n one can calculate $p_s$ and $p_c$ for any $(l,m,r)$ configuration and hence determine whether to stop or continue for any $X_r$.

The expression $p_c(l,m,r;C)$ is increasing in $l$. Again a backward induction argument is used. For $r=n-1$, it is trivially true. If it is true for $r=s+1$ then

$$\begin{aligned} p_c(l+1,m,s;C) - p_c(l,m,s;C) &= \frac{s-l-1}{C+s}\,[p_c(l+1,m,s+1;C) \\ &\quad - p_c(l,m,s+1;C)] + \frac{mC/(m+1)}{C+s}\,[p_c(l+1,m+1,s+1;C) \\ &\quad - p_c(l,m+1,s+1;C)] + \frac{(l+1)}{C+s}\,p_M(l+2,m,s+1;C) \\ &\quad - \frac{l}{C+s}\,p_M(l+1,m,s+1;C) - \frac{1}{C+s}\,p_c(l,m,s+1;C). \end{aligned}$$

By the induction hypothesis the first two terms on the right are positive. Further, since $p_c(l,m,s+1;C)\le p_M(l,m,s+1;C)$ the remaining terms on the right are $\ge \frac{l}{C+s}\,[p_M(l+2,m,s+1;C)-p_M(l+1,m,s+1;C)]$ +

$\frac{1}{C+s}$ $[p_M(l+2,m,s+1;C)-p_M(l,m,s+1;C)]$ which by induction is positive. Thus $p_c(l+1,m,s;C)\geq p_c(l,m,s;C)$; the induction is complete.

The ramification of this result is as follows. Since $p_s$ does not depend on $l$ and $p_c$ increases in $l$, if there is a configuration $(l,m,r)$ for which $p_c>p_s$ then the optimal strategy is to continue for all $(l',m,r)$ configurations, where $l\leq l'\leq r-m+1$.

The value of C can influence the decision to stop or to continue. For example, for n = 5,

$$p_s(1,2,2;C) = (24 + \frac{52}{3}C + \frac{9}{2}C^2 + \frac{2}{5}C^3)/[(C+2)(C+3)(C+4)];$$

$$p_c(1,2,2;C) = (18 + 16C + \frac{9}{2}C^2 + \frac{13}{30}C^3)/[(C+2)(C+3)(C+4)].$$

Therefore at $X_2$ $(>X_1)$ the optimal rule is to stop if $C<C_0$ and to continue if $C>C_0$ where $C_0=7.93$. The complete rule for n=5 is to continue also at (1,1,1), (2,1,2), (3,1,3) and to stop otherwise. At (4,1,4) it doesn't matter whether one stops or continues.

Optimal rules are presented for $n\leq 8$ and for various C in Table 3.1. In particular, the $(l,m,r)$ configurations at which one continues are listed. The configuration $(r,1,r)$ appears in all cases, in that the optimal strategy is always to continue at $(r,1,r)$ if $r\leq n-2$, with indifference at $r=n-1$. It is interesting to note that the strategies for C=100 are just the reduced rules of Section 2.

Table 3.2 contains the probabilities of correct selection for the optimal strategies; it is just a table of $p_c(1,1,1;C)$ for various values of n and C. For comparison the optimal probability for the continuous case as in Gilbert and Mosteller (1966) is presented in the last column of the table.

| n | C = 0.1 | C = 1.0 | C = 10 | C = 100 |
| --- | --- | --- | --- | --- |
| 3 | (r,1,r) | (r,1,r) | (r,1,r) | (r,1,r) |
| 4 | (r,1,r) | (r,1,r) | (r,1,r) | (r,1,r) |
| 5 | (r,1,r) | (r,1,r) | (r,1,r)(1,2,2) | (r,1,r)(1,2,2) |
| 6 | (r,1,r) | (r,1,r) | (r,1,r)(2,2,3)(1,2,2) | (r,1,r)(1,2,3)<br>(2,2,3)(1,2,2) |
| 7 | (r,1,r) | (r,1,r)(2,2,3)(3,2,4) | (r,1,r)(2,2,4)(3,2,4)<br>(1,2,3)(2,2,3)(1,2,2) | (r,1,r)(1,2,4)<br>(2,2,4)(3,2,4)<br>(1,2,3)(2,2,3)(1,2,2) |
| 8 | (r,1,r) | (r,1,r)(4,2,5)(3,2,4)<br>(2,2,3)(1,2,3)(2,1,2) | (r,1,r)(3,2,5)(4,2,5)<br>(1,2,4)(2,2,4)(3,2,4)<br>(1,2,3)(1,3,3)(1,2,2) | (r,1,r)(1,2,5)(2,2,5)<br>(3,2,5)(4,2,5)(1,2,4)<br>(2,2,4)(3,2,4)(1,2,3)<br>(1,3,3)(2,2,3)(1,2,2) |

TABLE 3.1

($\ell$,m,r) CONFIGURATIONS AT WHICH THE
OPTIMAL STRATEGY FOR DIRICHLET SAMPLES IS
TO CONTINUE ($1 \le r \le n-2$)

| n \ C | 0.1 | 1.0 | 10 | 100 | 1000 | continuous case |
|---|---|---|---|---|---|---|
| 2 | .9545 | .7500 | .5455 | .50495 | .5005 | .5000 |
| 3 | .9545 | .7500 | .5455 | .50495 | .5005 | .5000 |
| 4 | .9615 | .7691 | .5241 | .4657 | .4591 | .4583 |
| 5 | .9682 | .7802 | .5004 | .4411 | .4341 | .4333 |
| 6 | .9733 | .7825 | .5050 | .4371 | .4287 | .4279 |
| 7 | .9769 | .7805 | .5048 | .4254 | .4154 | .4143 |
| 8 | .9793 | .7897 | .5007 | .4213 | .4110 | .4098 |
| 9 | .9809 | .7948 | .5066 | .4191 | .4073 | .4060 |
| 10 | .9819 | .7982 | .5097 | .4135 | .4002 | .3987 |
| 12 | .9835 | .8047 | .5156 | .4124 | .3973 | .3955 |
| 15 | .9849 | .8132 | .5260 | .4100 | .3916 | .3894 |
| 20 | .9858 | .8242 | .5435 | .4102 | .3870 | .3842 |

TABLE 3.2

OPTIMAL PROBABILITIES OF CORRECT SELECTION
FOR DIRICHLET PROCESS SAMPLES

## 4. Asymptotic rules and discussion

The limiting behavior of the strategies of Section 3 is examined.

First consider the limit as C tends to infinity. Then equation (3.1) becomes:

$$\lim_{C\to\infty} p_S(m,r;C) = \frac{m}{m+1} \lim_{C\to\infty} p_S(m+1,r+1;C).$$

Repeated application and the known limit for stage $r=n-1$ yields:

$$\lim_{C\to\infty} p_S(m,r;C) = \frac{m}{m+1}\,\frac{m+1}{m+2}\cdots\frac{m+n-r-1}{m+n-r} = \frac{m}{m+n-r}. \tag{4.1}$$

Note that this is merely the probability of correct selection by stopping for the continuous version of the problem where ties are completely ignored. This is the appropriate probability for the reduced strategy of Section 2; namely, discard the $r-m$ tied values, reducing $n$ to $n'=n-(r-m)$ and reducing $r$ to $m$, and obtain the analogous probability for the classic untied strategy.

Similarly, equation (3.2) becomes:

$$\lim_{C\to\infty} p_C(l,m,r;C) = \frac{m}{m+1} \lim_{C\to\infty} p_C(l,m+1,r+1;C)$$
$$+ \frac{1}{m+1} \lim_{C\to\infty} p_M(l,m+1,r+1;C). \tag{4.2}$$

For $l=1$, $\lim_{C\to\infty} p_C(1,m+i,r+i;C)$ is decreasing in $i$ for $0\le i\le n-r-1$.

This is so because

$$\lim_{C\to\infty} p_C(1,m,r;C) - \lim_{C\to\infty} p_C(1,m+1,r+1;C) =$$

$$\frac{1}{m+1}\lim_{C\to\infty} p_M(1,m+1,r+1;C) - \frac{1}{m+1}\lim_{C\to\infty} p_C(1,m+1,r+1;C) \geq 0.$$

On the other hand $p_S(m+i,r+i;C)$ is increasing in i for $0\leq i\leq n-r-1$. Thus in the limit as C tends to infinity there exists a unique s depending on n and r such that the optimal rule is to stop at $(1,s,r)$ (for the unique minimum s) and hence also for $(1,s+i,r+i)$ for $1\leq i\leq n-r-1$.

Therefore for $m\geq s$ the repeated application of (4.1) and (4.2) yields

$$\lim_{C\to\infty} p_C(1,m,r;C) = \frac{1}{m+1}\,\frac{m+1}{n-(r-m)} + \frac{m}{m+1}\,\frac{1}{m+2}\,\frac{m+2}{n-(r-m)}$$

$$+ \cdots + \frac{m}{m+1}\,\frac{m+1}{m+2}\cdots\frac{m+n-r-2}{m+n-r-1}\,\frac{1}{m+n-r}\,\frac{m+n-r}{n-(r-m)}$$

$$= \frac{1}{n-(r-m)}\left[\frac{m}{m} + \frac{m}{m+1} + \cdots + \frac{m}{m+n-r-1}\right].$$

But this is just equation (2a-2) of Gilbert and Mosteller (p. 39) for the optimal probability of continuing in the continuous case for a sample of size $n'=n-(r-m)$ at stage $r'=m$; i.e., it is the probability for the reduced strategy.

This then proves that as C tends to infinity the rules from the Dirichlet process approach those of the reduced rules of Section 2. Further, the optimal probability of selecting a maximum therefore converges as C tends to infinity to the optimal probability for the continuous distribution function case. The convergence rate can be observed in Table 3.2.

The behavior as C tends to zero is also quite interesting. Sethuraman and Tiwari (1982) point out in general that the limit as C tends to zero in the Dirichlet process does not correspond to the non-informative prior. As C tends to zero, all observations are tied at $X_1$, a random value. In this optimal stopping application, the only role of C is to regulate ties, so this limit should be interpretable in this application.

Taking the limit as C tends to 0, (3.1) becomes

$$\lim_{C\to 0} p_S(m,r;C) = \frac{r}{r} \lim_{C\to 0} p_S(m,r+1;C).$$

Repeated application of this along with the limiting initial condition implies

$$\lim_{C\to 0} p_S(m,r;C) = 1.$$

Equation (3.2) becomes

$$\lim_{C\to 0} p_c(l,m,r;C) = \frac{r-l}{r} \lim_{C\to 0} p_c(l,m,r+1;C) + \frac{l}{r} \lim_{C\to 0} p_M(l+1,m,r+1;C)$$

$$= \frac{r-l}{r} \lim_{C\to 0} p_C(l,m,r+1;C) + \frac{l}{r}.$$

Since $\lim_{C\to 0} p_C(l,m,n-1;C) = l/(n-1)$, this becomes

$$\lim_{C\to 0} p_c(l,m,r;C) = \frac{l}{r} [1 + \frac{r-l}{r+1} + \frac{(r-l)(r-l+1)}{(r+1)(r+2)}$$

$$+ \ldots + \frac{(r-l)\ldots(n-l-2)}{(r+1)\ldots(n-1)}].$$

Note that $\lim_{C\to 0} p_c(l,m,r;C)<1$ unless $\lim_{C\to 0} p_c(l,m,r+i;C)=1$ for all $1\leq i\leq n-r-1$. But $\lim_{C\to 0} p_c(l,m,n-1;C)=\frac{l}{n-1}$ which is $<1$ unless $m=1$.

The conclusion is that the limiting strategy as C tends to zero is to stop at a relative maximum at $X_r$ with configuration $(l,m,r)$ if $l<r$. If $l=r$ a more detailed analysis is necessary:

$$\lim_{C\to 0} [p_C(r,1,r;C)-p_S(1,r;C)] = \lim_{C\to 0} \{\frac{r}{C+r}[p_M(r+1,1,r+1;C)$$

$$- p_S(1,r+1;C)] + \frac{C/2}{C+r} [p_C(r,2,r+1;C)$$

$$- p_S(2,r+1;C)] + \frac{C/2}{C+r} p_M(1,2,r+1;C)\}.$$

Since $p_M(1,2,r+1;C) \geq p_S(2,r+1;C)$ because $p_S(m,r;C)$ is increasing in m, it follows that $\lim_{C\to 0} p_C(r,1,r;C) \geq \lim_{C\to 0} p_S(1,r;C)$; i.e., the optimal strategy for $r \leq n-2$ is to continue at $(r,1,r)$.

An inspection of Table 3.1 reveals that for $n \leq 8$ these limiting rules are achieved at $C=0.1$.

In that the optimal strategy based on the Dirichlet process is admittedly complex, depending at the rth stage on l and m, a simplified strategy is proposed. Let D denote the number of distinct values in the sequence of fixed length n from a Dirichlet process with non-atomic parameter with mass C. If ties are completely ignored, the search for the optimal rule is then similar to the approach of Gianini-Pettitt (1979), who found an optimal rule for a sequence of random length N, where N has a discrete uniform distribution. For simplicity here, the reduced rules of Section 2 are utilized if the number of distinct observations are known. The strategy is to continue with a relative maximum at stage r with m distinct observations if

$$P\{D \geq d_C | m,r\} > P\{D \leq d_S | m,r\}$$

where the reduced rule continues at $X_r$ with m distinct observations if $D \geq d_C$, stops if $D \leq d_S$, is indifferent if $d_S < D < d_C$. This strategy does not depend on l, the number tied at the relative maximum.

Consider the simplified rule for $n=8$, $C=10$. To easily obtain the reduced rules, the optimal strategy for the continuous problem for $n \leq 8$

is displayed; for a sequence of length d, this rule stops at the first relative maximum at stage s or later (from Gilbert and Mosteller (1966)):

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| d | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| s | 1 | 1,2 | 2 | 2 | 3 | 3 | 3 | 4. |

Note indifference in stopping or continuing for d=2. For a relative maximum at $(m,r)=(1,6)$, if D=3 the reduced rule, discarding the first five observations, continues; if D=2 the reduced rule is indifferent; if D=1 the reduced rule stops. Now $P\{D=3|(1,6)\} = P[X_7,X_8 \notin \{X_1,X_2,X_3,X_4,X_5,X_6\}, X_7 \neq X_8] = \frac{C}{C+6}\,\frac{C}{C+7} = 0.3676$ and $P\{D=1|(1,6)\} = \frac{6}{C+6}\,\frac{7}{C+7} = 0.1544$. Since $0.3674>0.1544$, the simplified rule continues at (1,6). At $(m,r)=(2,4)$, $d_s=4$ and $d_c=5$. Now $P\{D\geq 5|(2,4)\} = \frac{10^3}{14\cdot 15\cdot 16\cdot 17}[10+4+5+6+7] = 0.5602$, so the rule continues at (2,4). At (2,3), $d_s=4$, $d_c=5$ and $P\{D\geq 5|(2,3)\}=0.6114$, so the rule continues. For $(m,r)=(2,5)$, $d_s=4$ and $P\{D\leq 4|(2,5)\}=0.7549$; the rule is to stop. The complete simplified rule is to continue if $r<3$, and also at (2,3), (2,4) and (1,r) (for $r\leq 6$) and to stop otherwise. The optimal Dirichlet rule for n=8, C=10 from Table 3.1 continues when the simplified rule does and also for $(l,m,r)=(3,2,5)$ and $(4,2,5)$. The optimal rule is approximated by the simplified rule.

The limiting behavior of these simplified rules is easily obtained. As C tends to infinity, $P\{D=n-r+m|(m,r)\}$ tends to 1. The rule is then just the reduced rule at the mth stage for a sample of size $n-r+m$. As C tends to zero, $P\{D=m|(m,r)\}$ tends to 1 in which case one stops. These limiting cases are as in the limiting cases for the optimal Dirichlet rules.

If the parameter C is unknown for the Dirichlet process, it can be estimated based on a previous sample. If $n'$ denotes the size of the previous sample and $D'=d'$ is the observed number of distinct observations out of $n'$, then the maximum likelihood approach of Ewens

(1972) can be employed: the estimate is the unique solution for $C$ of the implicit equation:

$$d' = C\left(\frac{1}{C}+\frac{1}{C+1}+\dots+\frac{1}{C+n'-1}\right).$$

For large $n'$ the consistent estimator $D'/\log n'$ of Korwar and Hollander (1973) is more easily computed.

An important question concerns the use of the Dirichlet process to generate rules for the maximal selection from a sequence with ties if the sequence is <u>not</u> from a Dirichlet process. If the sequence arises as a random sample from a discrete distribution, the use of Dirichlet process is particularly appropriate in that the prior from the Dirichlet process concentrates its mass on discrete distributions. The selection of a non-atomic measure as the parameter for the Dirichlet process is necessitated by the fact that only relative ranks of the sequence are observable. In addition, the resulting rules from the non-atomic measure depend only on the mass $C$ of this measure. The limiting behavior of the Dirichlet rules as $C$ tends to 0 and tends to infinity strongly reinforces the use of $C$ as a mechanism to model the ties in the sequence. Therefore, prior information concerning the incidence of ties would be incorporated into the procedure by the choice of $C$. In particular, the distribution of $D$ for the Dirichlet model is given by Antoniak (1974):

$$P\{D=d\} = S_n^{(d)} C^d / C^{[n]} \qquad d=1,\dots,n,$$

where $C^{[n]} = C(C+1)\dots(C+n-1)$ and $S_n^{(d)}$ is the absolute value of the Stirling number of the first kind. Since the distribution of $D$ for the random sample from the discrete distribution depends on the probabilities at the atoms, $C$ could be chosen to approximate the prior belief of one's distribution of $D$.

## Acknowledgement

The author gratefully acknowledges the helpful comments and references of a referee.

## References

[1] ANTONIAK, C.E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. Ann. Statist. 2, 1152–1174.

[2] BERK, R. H. and SAVAGE, I. R. (1979). Dirichlet processes produce discrete measures: an elementary proof. Contributions to Statistics-Jaroslav Hájek Memorial Volume, J. Jureckova (editor), Dordrecht, Holland. 25–32.

[3] CAMPBELL, G. (1983). The maximum of a sequence with prior information. Commun. Statist. C (to appear).

[4] CAMPBELL, G. and HOLLANDER, M. (1978). Rank order estimation with the Dirichlet prior. Ann. Statist. 6, 142–153.

[5] CAMPBELL, G. and SAMUELS, S. M. (1981). Choosing the best of the current crop. Adv. Appl. Prob. 13, 510–532.

[6] EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. Theoret. Popn Biol. 3, 87–112.

[7] FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. Ann. Stat. 1 209–230.

[8] GIANINI-PETTITT, J. (1979). Optimal selection based on relative ranks with a random number of individuals. Adv. Appl. Prob. 11, 720–736.

[9] GILBERT, J. P. and MOSTELLER, F. (1966). Recognizing the maximum of a sequence. J. Am. Statist. Assoc. 61, 35–73.

[10] GOVINDARAJULU, Z. (1975). The secretary problem: optimal selection with interview cost. Proceedings of the Symposium on

Statistics and Related Topics.

[11] IRLE, A. (1980). On the best choice problem with random population size. Z. Opns Res. 24, 177–190.

[12] KORWAR, R. M. and HOLLANDER, M. (1973). Contributions to the theory of Dirichlet processes. Ann. Prob. 1, 705–711.

[13] LORENZEN, T. J. (1979). Generalizing the secretary problem. Adv. Appl. Prob. 11, 384–396.

[14] PETRUCCELLI, J. D. (1981). Best choice problems involving uncertainty of selection and recall of observations. J. Appl. Prob. 18, 415–425.

[15] PRESMAN, E. L. and SONIN, I. M. (1972). The best choice problem for a random number of objects. Theory Prob. Appl. XVIII, 657–668.

[16] RASMUSSEN, W. T. (1975). A generalized choice problem. J. Optimization Theory Appl. 15, 311–325.

[17] RASMUSSEN, W.T. and ROBBINS, H. (1975). The candidate problem with unknown population size. J. Appl. Prob. 12, 692–701.

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

[19] SETHURAMAN, J. and TIWARI, R. C. (1982). Convergence of Dirichlet measures and the interpretation of their parameter. Third Purdue Symposium on Decision Theory and Related Topics.

[20] SMITH, M. H. (1975). A secretary problem with uncertain employment. J. Appl. Prob. 12, 620–624.

[21] SMITH, M. H. and DEELY, J.J. (1975). A secretary problem with finite memory. J. Am. Statist. Assoc. 70, 357–361.

[22] STEWART, T. J.(1978). Optimal selection from a random sequence with learning of the underlying distribution. J. Am. Statist. Assoc. 73, 775–780.

[23] YANG, W. C. K. (1974). Recognizing the maximum of a random

sequence based on relative rank with backward solicitation. *J. Appl. Prob.* **11**, 504-512.