A85

The Maximum of a Sequence with Prior Information

by

Gregory Campbell
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #485

March, 1977

The Maximum of a Sequence with Prior Information

by

Gregory Campbell

ABSTRACT

Let $X_1,\ldots,X_n$ be a random sample of size n which are observed sequentially. The problem is to select the maximum observation in the sequence while observing the random variables one by one without recall. Gilbert and Mosteller [J. Amer. Statist. Assoc. 61, 1966, 35-73] have treated this problem in the situation of no prior information concerning the random variables (the dowery or secretary problem) and in the situation that the random variables are i.i.d. with known distribution F (the full-information game). The problem is treated here in the intermediate case of partial prior information by means of the Dirichlet process prior introduced by Ferguson [Ann. Statist. 1, 1973, 209-230]. Optimal sequential decision rules are developed and compared, using the probabilities of correct selection, to the optimal rules for the secretary problem and the complete information game.

The Maximum of a Sequence with Prior Information

by

Gregory Campbell

Purdue University

1. Introduction. In the dowery problem, there are n slips of paper with one number on each, corresponding to the monetary values of the doweries of n girls. The problem is to select the slips of paper one at a time, stopping at that slip with the largest number, where one is prohibited from going back to a previous selection. Thus, at each stage, the choice is to either select that dowery and stop or to continue and select one of the remaining. In the event that there is no information concerning the distribution of the numbers, this is also called the secretary problem. If the numbers on the slips of paper are i.i.d. from a known distribution function, the problem is referred to as the total- or full-information game. Gilbert and Mosteller [6], among others, have treated both these problem in terms of optimal decision rules and the probabilities of winning. However, often there is neither total distributional ignorance nor certainty concerning the distribution of the dowery numbers. The purpose of this paper is the treatment of this problem under partial prior information concerning the distribution function. An allied problem is rank estimation with prior

information, which has been treated in Campbell and Hollander [3].

The key to the partial prior information problem is Ferguson's [5] Dirichlet process prior. The notation and preliminaries concerning the Dirichlet process are introduced in Section 2. Section 3 is devoted to the development of optimal rules for a game based on the Dirichlet process. In Section 4 the probabilities of winning for this Dirichlet game are contrasted with the probabilities for the secretary problem and for the full-information game. The question of performance of the Dirichlet rules in non-random situations is addressed in Section 5 under the realistic assumption that the underlying distribution function is unknown yet estimated. In the event of less than perfect estimation, the Dirichlet rules will be seen to exhibit an element of robustness not to be found in the full-information rules employed with the estimated distribution function.

2. Preliminaries. This section defines the Dirichlet process on the real line $\mathcal{R}$ with Boel σ-field $\mathcal{B}$ and its properties which will be of use in the sequel.

DEFINITION 2.1. (Ferguson) Let $(\mathcal{R},\mathcal{B})$ denote the real line with Boel σ-field $\mathcal{B}$ and let α denote a non-null, finite measure on $(\mathcal{R},\mathcal{B})$. Then P is a Dirichlet process on $(\mathcal{R},\mathcal{B})$ with parameter α if, for any $k = 1,2,\ldots,$ and every measurable partition $(B_1,\ldots,B_k)$ of $\mathcal{R}$, the vector $(P(B_1),\ldots,P(B_k))$ has a Dirichlet distribution with parameter $(\alpha(B_1),\ldots,\alpha(B_k))$.

DEFINITION 2.2 (Ferguson). The real-valued random variables $X_1,\ldots X_n$

constitute a sample of size n from a Dirichlet process P on $(\mathcal{R},\mathcal{B})$ with parameter α if, for every m = 1,2,..., and measurable sets $A_1,\ldots A_m$, $C_1,\ldots,C_n$,

$$Pr\{X_1 \in C_1,\ldots,X_n \in C_n | P(A_1),\ldots,P(A_m),P(C_1),\ldots,P(C_n)\} = \prod_{i=1}^{n} P(C_i) \text{ a.s.}$$

where Pr denotes probability.

THEOREM 2.3. (Ferguson). If X is a sample of size one from P, then $Pr\{X \in A\} = \alpha(A)/\alpha(\mathcal{R})$ for every $A \in \mathcal{B}$.

THEOREM 2.4. (Ferguson(1973)). Let P be a Dirichlet process on $(\mathcal{R},\mathcal{B})$ with parameter α and let $X_1\ldots,X_n$ be a sample of size n from P. Then the conditional distribution of P given $X_1,\ldots X_n$ is a Dirichlet process with updated parameter $\alpha + \sum_{i=1}^{n} \delta_{X_1}$, where for each $A \in \mathcal{B}$, $\delta_z(A) = 1$ if $z \in A$ and $\delta_z(A) = 0$ if $z \notin A$.

THEOREM 2.5. (Campbell and Hollander [1]). If $X_1,\ldots,X_r$ is a sample of size r from a Dirichlet process on $(\mathcal{R},\mathcal{B})$ with parameter α and if $A \in \mathcal{B}^n$, the n-dimensional Boel σ-field for $\mathcal{R}^n$, then for $r < n$,

$$Pr\{(X_1,\ldots,X_n) \in A | X_1,\ldots,X_r\}$$

$$= \int Pr\{(X_1\ldots,X_n) \in A | X_1,\ldots X_r, F\} dQ_\alpha^*(F),$$

where Q_α^* denotes the probability measure induced on the space of distribution functions by the Dirichlet process with updated parameter $\alpha^* = \alpha + \sum_{i=1}^{r} \delta_{X_i}$.

For $X_1\ldots,X_n$ a sample of size n, let K,L, and M denote the random variables for the number of $X_1\ldots,X_n$ that are less than, equal to, and greater than X_1, respectively. For a sample from a Dirichlet process, the following theorem gives the distribution of the triple (K,L,M), where the notation $y^{[k]}$ denotes the k^{th} ascending factorial $y(y-1)\ldots(y-k+1)$ for k a positive integer and $y^{[0]} \equiv 1$.

THEOREM 2.6. If $X_1,\ldots,X_n$ is a sample of size n from a Dirichlet process on $(\mathcal{R},\mathcal{B})$ with parameter α, then

$$\Pr\{(K,L,M)\} = (k,\ell,m) = \binom{n-1}{k,\ell-1,m-1} \Psi_{k,\ell,m}(\alpha)/\alpha(\mathcal{R})^{[n]}, \qquad (2.1)$$

where

$$\Psi_{k,\ell,m}(\alpha) = \int \alpha(-\infty,x)^{[k]}(\alpha(\{x\}) + 1)^{[\ell-1]}\alpha(x,\infty)^{[m]}d\alpha(x).$$

The proof of Theorem 2.6 is contained in the proof of Theorem 3.6 of Campbell and Hollander [3]. It is clear that expression $\Psi_{k,\ell,m}(\alpha)$ in Theorem 2.7 depends on the measure α. However, if α is a non-atomic measure (so $\alpha(\{x\}) = 0$ for all x), it is easily verified that

$$\Psi_{k,\ell,m}(\alpha) = (\ell-1)!\Phi_{k,m}(\alpha(\mathcal{R})),$$

where $$\Phi_{k,m}(c) = \int_0^c x^{[k]}(c-x)^{[m]}dx.$$

For k and m small, the closed form expression of $\Phi_{k,m}(c)$ is easily derived; for example, $\Phi_{1,1}(c) = \frac{c^3}{6}$, $\Phi_{3,1}(c) = \Phi_{1,3}(c) = c^5/20 + c^4/4 + c^3/3$. Thus, if α is nonatomic,

$$\Pr\{(K,L,M) = (k,\ell,m)\} = \frac{(n-1)!}{k!\ m!}\Phi_{k,m}(\alpha(\mathcal{R})/\alpha(\mathcal{R}))^{[n]} \qquad (2.1')$$

Note that the distribution of (K,L,M) in such a case depends on the measure α only through $\alpha(\mathcal{R})$ and does not depend on L, the number of ties at X_1.

3. The optimal strategy for the Dirichlet game. In this section the strategy for the following game is developed. Let $X_1,\ldots,X_n$ be a sample of size n from a Dirichlet process on $(\mathcal{R},\mathcal{B})$ with parameter α, where α is known and, for convenience, nonatomic. The problem is to observe the X's one at a time and to stop at the maximum X, where recall

is not permitted. This is called the Dirichlet game. Note that it is the dowery problem for which the numbers on the slips of paper are from a Dirichlet process.

It suffices to consider the problem for α in the following form: without loss of generality, assume $\alpha([0,x]) = \alpha(\mathcal{R})\cdot x$ for $x\varepsilon\ [0,1]$ with $\alpha(\mathcal{R}\setminus[0,1]) = 0$. If α is of the form $\alpha([0,x]) = \alpha(R)\cdot F_0(x)$ for F_0 a strictly increasing distribution function on $(-\infty,\infty)$, the sequential optimal rules are given by F_0^{-1} of the optimal rules for $\alpha[0,x]) = \alpha(\mathcal{R})\cdot x$.

A second general commentconcerns the nature of the optimal rules for this game. If the loss function of the game is 0 if one fails to select the maximum and 1 if one succeeds in stopping at the maximum, the optimal rule must stop only at observations which are at least as large as those preceding it; such observations are called candidates. Unlike the secretary problem or the full information game in which F is continuous, ties can occur in the Dirichlet game, even if α is non-atomic. This stems from the fact that the Dirichlet prior places all its mass on discrete distribution functions (see Ferguson [5], Blackwell [2], Berk and Savage [1]). Thus, the maximum of the sequence may not be unique. It is therefore necessary to distinguish the different candidates. A candidate of order 1 (or a primary candidate) is a candidate which strictly exceeds all previous observations. A (secondary) candidate of order ℓ ($\ell > 1$) is a candidate which is as large as all previous observations, but which is tied with (ℓ-1) previous observations at the time of its selection.

Let $p_s(y;k,\ell)$ denote the probability that a candidate x of order ℓ at the $(k+\ell)^{\underline{th}}$ selection is the maximum of the sequence $X_1,\ldots X_n$ from a Dirichlet process, where y is such that $\alpha(\infty,x) = y$. Thus $p_s(y;k,\ell)$ denotes the probability of winning the Dirichlet game by stopping at a candidate of order ℓ at the $(k+\ell)^{\underline{th}}$ draw. It is straightforward to compute this probability.

THEOREM 3.1. If $X_1,\ldots,X_n$ is a sample of size n from a Dirichlet process on $(\mathcal{R},\mathcal{B})$ with parameter α, then

$$p_s(y;k,\ell) = (y + k + \ell)^{[j]}/(\alpha(\mathcal{R}) + k + \ell)^{[j]}, \tag{3.1}$$

where $j = n-k-\ell$ and $X_{k+\ell} = x$ with $\alpha(\infty,x) = y$.

PROOF. For distribution function F given, the conditional probability that $X_{k+\ell}$ is a maximum given it is a candidate or order ℓ is $(F(x))^j$, where $j = n-k-\ell$ is number of unobserved X's remaining. By Theorem 2.4 the random distribution function F given $X_1,\ldots,X_{k+\ell}$ has a Dirichlet process with parameter $\alpha'' = \alpha + \sum_{i=1}^{k+\ell} \delta_{x_i}$. Thus, by Theorem 2.5,

$$p_s(y;k,\ell) = \int (F(x))^j dQ_\alpha''(F) = \alpha''(\infty,x)^{[j]}/\alpha''(\mathcal{R})^{[j]}.$$

It is interesting to note that this probability depends on $X_1,\ldots X_{k+\ell}$ only through the value x of the $\ell^{\underline{th}}$ order candidate, the number $(\ell-1)$ of previous ties at x, and the number k of observations less than x; it does not depend on the value of the other k preceding observations.

Let $p_c(y;k,\ell)$ denote the probability of winning the Dirichlet game by deciding at stage $(k+\ell)$ for which there is a candidate x of order ℓ such that $y = \alpha(\infty,x)$ to not select the candidate x but to continue observing X's and to stop later using the optimal strategy. To the end of finding an expression for $p_c(y;k,\ell)$, the following lemma is most helpful.

LEMMA 3.2. Let $X_1,\ldots,X_n$ be a sample of size n from a Dirichlet process on $(\mathbb{R},\mathcal{B})$ with parameter β which is nonatomic except possibly at the point x. The probability that, of i observations that are $\geq$ x, the first selected such that it is $\geq$ x is as large as the other (i-1) is given by:

$$p(i,\beta(\{x\}),\beta(x,\infty)) = \left\{\sum_{q=0}^{i-1}\binom{i-1}{q}(\beta(\{x\}+1)^{[i-1-q]}\Phi_{q,0}(\beta(x,\infty)) + (\beta\{x\})^{[i]}\right\}/\beta(x,\infty)^{[i]}.$$

PROOF. Let z represent the value of the first random variable $\geq$ x. Then two cases are possible since β can be nonatomic at x:

Case 1 z = x. Then z is a maximum only if all i values at equal to x. Given F, the probability of this event is $(F(x)-F(x^-))^i$ which when integrated using the Dirichlet process yields $\beta(\{x\})^{[i]}/\beta[x,\infty)^{[i]}$.

Case 2 z > x. The probability that z is a maximum for the i values at least as large as x, conditional on F and Z, is $(F(z)-F(x^-))^{i-1}$ which integrates to $\beta'[x,z]^{[i-1]}/\beta'[x,\infty)^{[i-1]}$, where $\beta' = \beta + \delta_z$. Use that $(a+b)^{[k]} = \sum_{j=0}^{k}\binom{k}{j}a^{[j]}b^{[k-j]}$ to rewrite this as $\sum_{q=0}^{i-1}\binom{i-1}{q}\beta(x,z)^{[q]}\cdot$ $(\beta(\{x\})+1)^{[i-1-q]}/\beta'[x,\infty)^{[i-1]}$.

From Theorem 2.3 the distribution of z (z > x) is known, so integrating over all z > x,

$$\int_x^\infty \sum_{q=0}^{i-1}\binom{i-1}{q}\beta(x,z)^{[q]}(\beta(\{x\})+1)^{[i-1-q]}d\beta(z)\Big/\beta'[x,\infty)^{[i-1]}\beta(x,\infty)$$

$$= \sum_{q=0}^{i-1}\binom{i-1}{q}\Phi_{q,0}(\beta(x,\infty))(\beta(\{x\})+1)^{[i-1-q]}\Big/\beta'[x,\infty)^{[i-1)}\beta(x,\infty).$$

Multiplying this probability by the probability that $z > x$; namely, $\frac{\beta(x,\infty)}{\beta[x,\infty)}$ and adding the probability from Case 1, the lemma is proved. ||

It is easy to intuit that $p(i,\beta(\{x\}),\beta(x,\infty))$ must be at least as large as $\frac{1}{i}$, since if the observations were from some non-random distribution the probability would $\frac{1}{i}$ if the distribution function were continuous and greater than $\frac{1}{i}$ if not continuous. The role that the mass of the measure plays in the Dirichlet process is illustrated by the following limiting cases for $p(i,\ell,c)$ for ℓ a constant, where c tends to either 0 or ∞.

PROPOSITION 3.3 As $c \to \infty$, $p(i,\ell,c) \to 1$, and as $c \to \infty$, $p(i,\ell,c) \to \frac{1}{i}$.

PROOF.
$$\lim_{c \to \infty} p(i,\ell,c) = \lim_{c \to \infty} \{\sum_{q=0}^{i-1} \binom{i-1}{q} (\ell + 1)^{[i-1-q]} \Phi_{q,0}(c) + \ell^{[i]}\}.$$
$$1/(c + \ell)^{[i]} = \lim_{c \to \infty} \{\Phi_{i-1,0}(c) + \mathcal{O}(c^{i-1})\}/(c^{[i]} + \mathcal{O}(c^{i-1}))$$
$$= \lim_{c \to \infty} \frac{c^i}{i}/c^i = \frac{1}{i}.$$

Also,
$$\lim_{c \to 0} p(i,\ell,c) = \lim_{c \to 0} \{\sum_{q=0}^{i-1} \binom{i-1}{q} (\ell + 1)^{[i-1-q]} \Phi_{q,0}(c) + \ell^{[i]}\}.$$
$$1/(c + \ell)^{[i]} = \frac{\ell^{[i]}}{\ell^{[i]}} = 1. \quad ||$$

The quantity $p(i,\ell,c)$ is useful in calculating the probability of winning by continuing, $p_c(y;k,\ell)$.

THEOREM 3.4. If $X_1,\ldots X_n$ is a sample of size n from a Dirichlet process on $(\mathfrak{R},\mathfrak{B})$ with parameter α, then the probability of winning at the $(k + \ell)^{th}$ stage with a candidate x of order ℓ by continuing with the optimal strategy is:

$$p_c(y;k,\ell) = \sum_{i=1}^{j} \binom{j}{i} \alpha'(-\infty,x)^{[j-i]} \alpha'[x,\infty)^{[i]} p(i,\ell,\alpha(x,\infty))/\alpha'(\mathbb{R})^{[j]}, \qquad (3.2)$$

where $j = n-k-\ell$, $\alpha' = \alpha + \sum_{i=1}^{k+\ell} \delta_{X_i}$, and $y = \alpha(-\infty,x)$.

PROOF. The conditional probability that of the j remaining observations exactly i are greater than or equal to x given F is $\binom{j}{i}F(x^-)^{j-i}(1-F(x^-))^i$. Integrating F over the updated Dirichlet prior by Theorem 2.5, the probability that exactly i of the j remaining values are greater than or equal to x is $\binom{j}{i}\alpha'(-\infty,x)^{[j-i]}\alpha'[x,\infty)^{[i]}/\alpha'(\mathbb{R})^{[j]}$. It is reasonable to expect that the optimal strategy be such that the indifference values of x at each stage for which one stops for candidates larger or continues for candidates smaller must be decreasing. Thus, if the candidate falls exactly on such a point, if one continues, the optimal strategy is to select the first observation at least as large as x of the i such observations. The probability that this occurs is $p(i,\ell,\alpha(x,\infty))$. Therefore, by multiplying $p(i,\ell,\alpha(x,\infty))$ by the probability that exactly i future x's are at least as large as x and by summing i from 1 to j, the proof is completed. ||

In order to obtain the optimum rules for the Dirichlet game, it is a matter of equating the probabilities $p_s(y;k,\ell)$ and $p_c(y;k,\ell)$ for each pair k and ℓ and solving for y. Let $y_{k,\ell}$ denote the solution for the pair k and ℓ. The decision rule is then to stop at candidate x if x is such that $\alpha(-\infty,x) > y_{k,\ell}$ and to continue if $\alpha(-\infty,x) < y_{k,\ell}$.

Consider the simple case in which n = 2. At the first stage $(k = 0, \ell = 1)$,

$$p_s(y;0,1) = (y + 1)/(\alpha(\mathcal{R}) + 1);$$

$$p_c(y;0,1) = (\alpha(\mathcal{R})-y + 1)p(1,1,\alpha(\mathcal{R})-y)/(\alpha(\mathcal{R}) + 1).$$

But $p(1,1,x) = 1$ for all x by definition since there is only one value $\geq x$ to choose. Solving for y yields $y = \alpha(\mathcal{R})/2$. Then if x_0 is the value such that $\alpha(-\infty,x_0) = \alpha(\mathcal{R})/2$, the rule chooses $X_1 = x$ if $x > x_0$ and selects X_2 if $x < x_0$.

For each pair k and ℓ, the solution $y_{k\ell}$ (if one exists) to the equation $p_c(y;k,\ell) = p_s(y;k,\ell)$ is unique. In order to see this, the monotonicity of p_s and p_c must be noted. That p_s is strictly increasing in y follows directly from equation (3.1). The non-increasing nature of p_c is most easily demonstrated by appealing to the definition of p_c -- the probability of winning by continuing with the best strategy if a candidate x is such that $\alpha(-\infty,x) = y$. As y increases, so must x, and hence the probability of winning is non-decreasing. The strict monotonicity of p then assures that there can by at most one intersection of functions p_s and p_c, for k and ℓ fixed.

It is possible that no solution exists if one probability uniformly dominates the other. In such instances, by convention y will be chosen to be 0 if $p_s > p_c$ (one would then wish to stop) and to be $\alpha(\mathcal{R})$ if $p_s < p_c$ (one would wish to continue).

Let d_i denote the solution $y_{k,\ell}$ when $\ell = 1(i = k + 1)$. Then Table 3.1 gives the optimum decision numbers d_i for the Dirichlet game for $n = 5,10$, and 20 and $\alpha(\mathcal{R}) = 1.0$, 10.0, 100.0, and 10,000. Table 3.2 provides the Dirichlet decision numbers $y_{k,\ell}$ in the presence of ties for

TABLE 3.1

OPTIMUM DECISION NUMBERS d_i FOR THE DIRICHLET GAME (ℓ = 1)

n	i^{th} draw	$\alpha(R)$ = 1.0	$\alpha(R)$ = 10.0	$\alpha(R)$ = 100.0	full information
5	1	.80981	.82111	.82418	.82459
	2	.53054	.75016	.77323	.77584
	3	.04939	.62529	.68342	.68990
	4	.00000	.35000	.48500	.50000
10	1	.90498	.91258	.91558	.91604
	2	.79427	.89329	.90487	.90627
	3	.65820	.86899	.89133	.89391
	4	.48380	.83738	.87370	.87781
	5	.24962	.79456	.84975	.85595
	6	.00000	.73320	.81541	.82459
	7	.00000	.63794	.76203	.77584
	8	.00000	.47017	.66791	.68990
	9	.00000	.10000	.46000	.50000
20	1	.95198	.95604	.95845	.95892
	2	.90130	.94934	.95580	.95671
	3	.84640	.94190	.95286	.95424
	4	.78601	.93360	.94955	.95148
	5	.71880	.92426	.94583	.94836
	6	.64318	.91367	.94160	.94483
	7	.55713	.90154	.93674	.94077
	8	.45807	.88751	.93111	.93606
	9	.34252	.87109	.92451	.93054
	10	.20573	.85160	.91666	.92398
	11	.04099	.82807	.90718	.91604
	12	.00000	.79909	.89549	.90627
	13	.00000	.76251	.88072	.89391
	14	.00000	.71488	.86147	.87781
	15	.00000	.65026	.83535	.85595
	16	.00000	.55761	.79787	.82459
	17	.00000	.41367	.73961	.77584
	18	.00000	.16000	.63690	.68990
	19	.00000	.00000	.41000	.50000

TABLE 3.2

OPTIMUM DIRICHLET RULE FOR N = 5, ALLOWING FOR TIES

k	ℓ	$\alpha(\mathcal{R}) = 10.0$
0	1	.8211
0	2	.7883
0	3	.7083
0	4	.5000
1	1	.7502
1	2	.6674
1	3	.4500
2	1	.6253
2	2	.4000
3	1	.3500

$n = 5$ and $\alpha(\mathcal{R}) = 10.0$.

Let $\alpha(-\infty,x) = \alpha(\mathcal{R})F(x)$ for distribution function $F(x)$. The next theorem establishes the limits of the rules $\frac{d_i}{\alpha(\mathcal{R})}$ as $\alpha(\mathcal{R})$ tends to infinity.

THEOREM 3.5. If $X_1 \ldots, X_n$ is a sample of size n from a Dirichlet process with parameter α with decision numbers d_i for the Dirichlet game, then, as $\alpha(R) \to \infty$ such that $\frac{\alpha(-\infty,x)}{\alpha(R)} \to F(x)$, a distribution function, $\frac{d_i}{\alpha(\mathcal{R})}$ approaches the optimum numbers for the full-information game with distribution $F(x)$.

PROOF. It is first shown that as $\alpha(\mathcal{R}) \to \infty$, $p_c(y;k,\ell) \to z^j$ for $j = n-k-\ell$ and $z = \lim_{\alpha(\mathcal{R}) \to \infty} \frac{y}{\alpha(\mathcal{R})}$. The j terms in the product $\frac{y + k + i}{\alpha(\mathcal{R}) + k + i}$ $(i = 1, \ldots, j)$ converge to z as $\alpha(\mathcal{R}) \to \infty$, so $p_c(y;k,\ell) \to z^j$. By a similar termwise argument, $\lim_{\alpha(\mathcal{R}) \to \infty} p_c(y;k,\ell) = \sum_{i=1}^{j} \binom{j}{i} z^{j-i}(1-z)^i \lim_{\alpha(\mathcal{R}) \to \infty} p(i,\ell,\alpha(\mathcal{R})-y)$. By Proposition 3.3, $\lim_{\alpha(\mathcal{R}) \to \infty} p(i,\ell,\alpha(\mathcal{R})-y) = \frac{1}{i}$. Since all these functions are continuous in z and $\alpha(\mathcal{R})$, the solution of $p_s(y;k,\ell) = p_c(y;k,\ell)$ must converge as $\alpha(\mathcal{R}) \to \infty$ such that $y/\alpha(\mathcal{R}) \to z$

to the solution of

$$z^j = \sum_{i=1}^{j} \binom{j}{i} z^{j-i}(1-z)^i/i,$$

which is the equation in Gilbert and Mosteller [6] from which the full-information rules are derived. ||

It is natural to inquire as to the behavior of the Dirichlet rules as $\alpha(\mathfrak{R})$ tends to zero. It seems natural to expect that the limiting rules are just those for the secretary problem. Unfortunately, this is not the case. Let $z = \lim_{\alpha(\mathfrak{R}) \to 0} y/\alpha(\mathfrak{R})$. Two cases are of interest:

First, suppose $k = 0$ and $\ell > 0$. Equating $p_s(y;0,\ell)$ and $p_c(y;0,\ell)$, we have:

$$\frac{(y+\ell)^{[j]}}{(\alpha(\mathfrak{R})+\ell)^{[j]}} = \sum_{i=1}^{j} \binom{j}{i} \frac{y^{[j-i]}(\alpha(\mathfrak{R})-y+\ell)^{[i]}}{(\alpha(\mathfrak{R})+\ell)^{[j]}}$$

$$\frac{\sum_{q=0}^{i-1}\binom{i-1}{q}(\ell+1)^{[i-q-1]}\Phi_{q,0}(\alpha(\mathfrak{R})-y) + \ell^{[i]}}{(\alpha(\mathfrak{R})-y+\ell)^{[i]}}$$

Multiplying both sides by the non-zero $(\alpha(\mathfrak{R})+\ell)^{[j]}$ and simplifying,

$$(y+\ell)^{[j]} = \sum_{i=1}^{j}\binom{j}{i} y^{[j-i]}\ell^{[i]} + \sum_{i=1}^{j}\binom{j}{i} y^{[j-i]}\sum_{q=0}^{i-1}\binom{i-1}{q}(\ell+1)^{[i-q-1]}\Phi_{q,0}(\alpha(\mathfrak{R})-y).$$

Using the fact that $(a+b)^{[j]} = \sum_{k=0}^{j}\binom{j}{k} a^{[j-k]}b^{[k]}$ in the case where $a = y$ and $b = \ell$, the above equation can be reduced to:

$$y^{[j]} = \sum_{i=1}^{j} y^{[j-1]}\sum_{q=0}^{i-1}\binom{i-1}{q}(\ell+1)^{[i-q-1]}\Phi_{q,0}(\alpha(\mathfrak{R})-y).$$

Now, as $\alpha(\mathfrak{R}) \to 0$, rewrite the left-hand side as $(j-1)!y + R_1$, where R_1 is such that $\lim_{\alpha(\mathfrak{R}) \to 0} R_1/\alpha(\mathfrak{R}) = 0$. Also, the right-hand side can be rewritten as $\Phi_{0,0}(\alpha(\mathfrak{R})-y)(\ell + 1)^{[i-1]} + R_2$, where $\lim_{\alpha(\mathfrak{R}) \to 0} R_2/\alpha(\mathfrak{R}) = 0$.

Since $\Phi_{0,0}(c) = c$ and $\lim_{\alpha(\mathfrak{R}) \to 0} y/\alpha(\mathfrak{R}) = z$, the solution for z is given by:

$$(j-1)!z = (1-z)(\ell + 1)^{[j-1]};$$

i.e., $z = (\ell + 1)/\{(j-1)! + (\ell + 1)^{[j-1]}\}$. For $\ell = 1$, $z = \frac{n-1}{n}$; for $\ell = 2$, $z = (n-1)(n-2)/((n-1)(n-2) + 2)$.

Consider the second case of $k > 0$, $\ell \geq 1$. It suffices as before to solve $(y + k)^{[j]} = \sum_{i=1}^{j} \binom{j}{i} (y + k)^{[j-i]} \sum_{q=0}^{i-1} \binom{i-1}{q} (\ell + 1)^{[i-q-1]} \Phi_{q,0}(\alpha(\mathfrak{R})-y)$.

Note that as $\alpha(\mathfrak{R}) \to 0$, $y \to 0$, so that the left-hand side approaches $k^{[j]}$ whereas the right-hand side goes to 0. By convention, then $z = 0$ for $k > 0$.

The optimal rules $d_i^* = \lim_{\alpha(\mathfrak{R}) \to 0} \frac{d_i}{\alpha(\mathfrak{R})}$ in the Dirichlet game for the limiting case of $\alpha(\mathfrak{R}) \to 0$ and for $\ell = 1$ is $d_1^* = \frac{n-1}{n}$, $d_2 = 0, \ldots$, $d_n^* = 0$. Contrast this with the optimal rule for the secretary problem which is of the form $d_1 = 1$, $d_2 = 1$, ..., $d_{s'} = 1$, $d_{s'+1} = 0, \ldots, d_n = 0$, where s' is the optimal integer depending on n, with $\frac{s'}{n} \to e^{-1}$ as $n \to \infty$. For extremely small $\alpha(\mathfrak{R})$, it is clear that the Dirichlet game is a poor model for the non-random distribution function game with the distribution function unknown. This is not the first situation in which the Dirichlet model so fails; see Ramsey [7] and Campbell and Hollander [3] for other instances.

4. The probability of winning. The probability of winning for any non-random strategy that has non-increasing decision numbers d_i is given in Gilbert and Mosteller [6]. Here, the adjective non-random refers to the distribution function, be it known or unknown. In the non-random game with less than perfect distributional information, one might consider employing a Dirichlet model and using the Dirichlet decision strategy. In this section the Dirichlet strategy is compared to the optimal strategies for the full-information game and for the secretary problem by means of the probabilities of winning. Also considered in this section is the probability of winning if the sample is actually from a Dirichlet process and the Dirichlet strategy employed.

Assume that one has correctly assessed the true distribution function $F(x)$ for the non-random game, but elects to use the Dirichlet strategy with measure α given by $\alpha(-\infty,x) = \alpha(\mathcal{R})F(x)$ where the positive value $\alpha(\mathcal{R})$ is used to reflect the degree of belief (see Ferguson [5]) or the confidence in the assessment $F(x)$. (Larger values of $\alpha(\mathcal{R})$ correspond to more confidence.). Table 4.1 compares the full-information strategy and the secretary problem strategy with family of the Dirichlet strategies indexed by $\alpha(\mathcal{R})$ by means of the probabilities of winning. It can clearly be seen that, as $\alpha(\mathcal{R})$ increases, the probability of winning with the Dirichlet strategy approaches that of the full-information game. It is further interesting that for $n = 20$ and $\alpha(\mathcal{R}) = 1.0$ the probability of winning for the Dirichlet game is exceeded by that of the secretary problem. This is a reflection of the poor modeling of the

TABLE 4.1

PROBABILITIES OF WINNING THE NONRANDOM GAME

n	Secretary problem	Dirichlet Rules $\alpha(\mathbb{R}) = 1.0$	$\alpha(\mathbb{R}) = 10.0$	$\alpha(\mathbb{R}) = 100.0$	Full information
2	.5000	.7500	.7500	.7500	.7500
3	.5000	.6424	.6831	.6843	.6843
4	.4583	.5956	.6526	.6554	.6554
5	.4333	.5608	.6349	.6391	.6392
10	.3987	.4532	.5944	.6085	.6087
15	.3894	.3938	.5726	.5984	.5990
20	.3842	.3535	.5554	.5932	.5942

Dirichlet process with small $\alpha(\mathbb{R})$ (relative to n) for the situation of the unknown distribution function. In short, for a truly nonrandom situation, if the amount of prior information concerning the distribution is small, it is actually better to use no information rather than to incorporate what little information there is into a Dirichlet model. Ferguson (1973) suggests a prior sample size interpretation for $\alpha(\mathbb{R})$ which may help in terms of how to use the prior information.

Consider the truly Dirichlet game in which $X_1,\ldots,X_n$ is a sample of size n from a Dirichlet process with parameter α. It is quite difficult in general to find an expression for the probability of winning this game by employing the optimal Dirichlet strategy, as detailed in Section 3. The strategy for this game depends on the number of ties at the candidates and the order of the observations, so that the sequential decision procedure is a triangular array $d_{k,\ell}$ with $1 \le k + \ell \le n-1$ and $\ell \ge 1$. In the very simple case of $n = 2$, a compact expression for the probability of winning the Dirichlet game is possible.

PROPOSITION 4.1. For a sample of size 2 from a Dirichlet process with nonatomic measure α, the probability of selecting the maximum of the sequence with the optimal Dirichlet strategy is $\frac{1}{4}(3 + (\alpha(\mathcal{R}) + 1)^{-1})$.

PROOF. From Section 3, it is known that $d_{0,1} = \frac{1}{2}\alpha(\mathcal{R})$. Let X_1 denote the $i^{\underline{th}}$ observation and let p_i denote the probability of winning by stopping after the $i^{\underline{th}}$ draw.

$$p_i = P\{X_1 \ge X_2,\ X_1 \ge \tfrac{1}{2}\alpha(\mathcal{R})\} = \int_{\frac{1}{2}\alpha(\mathcal{R})}^{\alpha(\mathcal{R})} \frac{\alpha(-\infty, x] + 1}{\alpha(\mathcal{R}) + 1} \frac{d\alpha(x)}{\alpha(\mathcal{R})}$$

$$= \frac{1}{2} - \frac{1}{8}(\alpha(\mathcal{R})/(\alpha(\mathcal{R}) + 1))$$

In a similar manner,

$$p_2 = P\{X_1 \le X_2, X_1 \le \tfrac{1}{2}\alpha(\mathcal{R})\} = \int_0^{\frac{1}{2}\alpha(\mathcal{R})} \frac{\alpha[x,\infty) + 1}{\alpha(\mathcal{R}) + 1} \frac{d\alpha(x)}{d\alpha(\mathcal{R})}$$

$$= \frac{1}{2} - \frac{1}{8}(\alpha(\mathcal{R})/(\alpha(\mathcal{R}) + 1)).$$

Thus, the probability of winning is $p_1 + p_2 = \frac{1}{4}(3 + (\alpha(\mathcal{R}) + 1)^{-1})$. ||

Note that as $\alpha(\mathcal{R})$ tends to infinity, the probability of winning approaches $\frac{3}{4}$, which is the nonrandom full information winning probability. More interestingly, as $\alpha(\mathcal{R})$ tends to zero, the Dirichlet probability of winning approaches 1. This is due to the proliferation of ties which the Dirichlet process with small $\alpha(\mathcal{R})$ encourages.

5. The Dirichlet strategy for the nonrandom game. This section focuses on the nonrandom game with less than perfect information concerning the distribution function. There are basically three

alternatives in such a situation. (i) One can ignore all prior information concerning the distribution and use the secretary problem's strategy. (ii) One could estimate the distribution function and then employ the full-information rule with the estimated rather than known distribution function. (iii) One could use the prior information to estimate the shape and mass of the measure α from a Dirichlet process and then employ the Dirichlet model to obtain a sequential decision procedure. These three approaches can be compared via the probability of winning for the non-random model as in Section 4 if the distribution function estimate is correct. The difference is that here the estimate F and the true distribution function F_0 can differ considerably. From Section 4, it is clear that if $\alpha(\mathcal{R})$ is small relative to n that procedure (i) (the secretary problem's strategy) is the only one to be considered. If $\alpha(\mathcal{R})$ is relatively large, one of (ii) or (iii) should be employed. In order to compare (ii) and (iii), it will be assumed that the estimate of F of F_0 in (ii) is also the shape of the Dirichlet measure in (iii), i.e., $\alpha(-\infty,x) = \alpha(\mathcal{R})F(x)$. If the estimate F agrees everywhere with the true distribution F_0, table 4.1 demonstrates that the full information strategy is to be preferred. The question addressed in this section concerns the comparison of the sequential decision rules in the presence of less than perfect estimation of the distribution function.

In the study of the robustness of these procedures, it is helpful to interpret the parameter $\alpha(\mathcal{R})$ for the Dirichlet procedure as the size of a previous sample on which the estimate F of F_0 is based. Without loss of generality the true distribution function F_0 is the uniform distribution [0,1]. The metric employed to reflect the distance of the estimate F from F_0 is $D = \sup_x |F(x)-F_0(x)|$. The following are the inverses of 8 smoothed but incorrect estimates for $F_0(x) = x$:

		D
1.	$F_1^+(x) = 2x-x^2$	.25
2.	$F_2^+(x) = x^3 + 3x^2(1-x)$	.0963
3.	$F_3^+(x) = x^5 + 5x^4(1-x) + 10x^3(1-x)^2$	.16
4.	$F_4^+(x) = \sqrt{x}$	.25
5.	$F_5^+(x) = x + k(.05) \quad k = 1,\ldots,9$	(0.05)k
6.	$F_6^+(x) = (10/(10-k))(x-\frac{1}{2}) + \frac{1}{2} \quad k = 1,\ldots,9$	(0.05)k
7.	$F_6^-(x) = ((10-k)/10)(x-\frac{1}{2}) + \frac{1}{2} \quad k = 1,\ldots,9$	(0.05)k
7.	$F_7^+(x) = \begin{cases} 2^{k-1}x^k & 0 \le x < \frac{1}{2} \\ 1-2^{k-1}(1-x)^k & \frac{1}{2} \le x \le 1 \end{cases}$ k = 1.5; k = 2.0	.0741; .1250
8.	$F_8^-(x) = \begin{cases} \frac{1}{2} - 2^{k-1}(\frac{1}{2} - x)^k & 0 \le x < \frac{1}{2} \\ \frac{1}{2} + 2^{k-1}(x - \frac{1}{2})^k & \frac{1}{2} \le x \le 1 \end{cases}$ k = 1.5,2.0,	Same as F_7^+

Define $F_i^-(x) = 2x-F_i^+(x)$, for i = 1,2,3,5,7 and define $F_8^+(x) = 2x-F_8^-(x)$. The reason that pairs F^+ and F^- are considered is that the pairs F^+ and F^- are the same vertical distance but in opposite directions from the line $y = x$ <u>at the same point</u>. The superscripts "+" and "-" are used to denote whether the distribution function is above or below the line $F_0(x) = x$ near $x = 1$.

Tables 5.1 and 5.2 present the probability of winning for n = 5 and n = 10, respectively, when the true distribution function is

TABLE 5.1

COMPARISON OF RULES FOR DOWERY PROBLEM FOR N = 5

WITH INCORRECT DISTRIBUTION FUNCTION

	F_0	F_1^+	F_1^-	F_2^+	F_2^-	F_3^+	F_3^-	F_4^+
$\alpha(\Re) = 5$	.6267	.5437	.5323	.6128	.5944	.5813	.5672	.6149
100	.6391	.4809	.5651	.6039	.6124	.5520	.5857	.5835
Full information	.6392	.4749	.5762	.6019	.6133	.5485	.5869	.5672

F_5^+									
k	.05	.10	.15	.20	.25	.30	.35	.40	.45
$\alpha(\Re) = 5$	.6298	.6142	.5740	.5224	.4806	.4436	.4205	.3840	.3482
$\alpha(\Re) = 100$	.6299	.5959	.5301	.4505	.3915	.3415	.3092	.2862	.2552
Full information	.6288	.5933	.5253	.4435	.3843	.3311	.3030	.2778	.2441

F_5^-									
k	.05	.10	.15	.20	.25	.30	.35	.40	.45
$\alpha(\Re) = 5$	.6098	.5832	.5504	.5141	.4766	.4394	.4039	.3707	.3404
$\alpha(\Re) = 100$	.6299	.6074	.5761	.5393	.5000	.4603	.4217	.3855	.3523
Full information	.6307	.6087	.5775	.5408	.5015	.4617	.4229	.3865	.3531

F_6^+									
k	1	2	3	4	5	6	7	8	9
$\alpha(\Re) = 5$	.6275	.6191	.5892	.5414	.5060	.4658	.4664	.4562	.3500
$\alpha(\Re) = 100$	.6356	.6184	.5708	.4972	.4414	.3895	.3529	.3526	.3517
Full information	.6352	.6172	.5679	.4917	.4362	.3765	.3531	.3531	.3531

F_6^-									
k	1	2	3	4	5	6	7	8	9
$\alpha(\Re) = 5$	.6212	.6120	.5996	.5848	.5679	.5496	.5302	.5099	.4892
$\alpha(\Re) = 100$	.6359	.6276	.6153	.5994	.5809	.5604	.5384	.5154	.4919
Full information	.6363	.6282	.6159	.6001	.5815	.5609	.5388	.5157	.4920

F_7^+			F_8^+	
k	1.5	2	1.5	2.0
$\alpha(\Re) = 5$	.6220	.5991	.6248	.5991
$\alpha(\Re) = 100$	.6192	.5781	.6215	.5781
Full information	.6179	.5758	.6203	.5758

F_7^-			F_8^-	
k	1.5	2.0	1.5	2.0
$\alpha(\Re) = 5$	.6041	.5787	.6049	.5787
$\alpha(\Re) = 100$	.6223	.5985	.6240	.5985
Full information	.6231	.5996	.6249	.5996

TABLE 5.2

COMPARISON OF RULES FOR DOWERY PROBLEM FOR N = 10

WITH INCORRECT DISTRIBUTION FUNCTION

	F_0	F_1^+	F_1^-	F_2^+	F_2^-	F_3^+	F_3^-	F_4^+
$\alpha(\Re)$ = 10	.5944	.4587	.5881	.5558	.5228	.4854	.4921	.5873
$\alpha(\Re)$ = 100	.6084	.3766	.5286	.5134	.5553	.4112	.5294	.5554
Full information	.6087	.3628	.5340	.5040	.5594	.3974	.5344	.5487

F_5^+									
k	.05	.10	.15	.20	.25	.30	.35	.40	.45
$\alpha(\Re)$ = 10	.5895	.5110	.4379	.3804	.3401	.3083	.2735	.2554	.2412
$\alpha(\Re)$ = 100	.5735	.4424	.3430	.2856	.2406	.2124	.1813	.1732	.1586
Full information	.5683	.4287	.3294	.2706	.2295	.1919	.1752	.1622	.1392

F_5^-									
k	.05	.10	.15	.20	.25	.30	.35	.40	.45
$\alpha(\Re)$ = 10	.5548	.4970	.4356	.3783	.3283	.2861	.2514	.2230	.1997
$\alpha(\Re)$ = 100	.5807	.5249	.4608	.3991	.3444	.2982	.2602	.2293	.2043
Full information	.5834	.5283	.4641	.4018	.3466	.2998	.2614	.2302	.2049

F_6^+									
k	1	2	3	4	5	6	7	8	9
$\alpha(\Re)$ = 10	.5912	.5202	.4515	.3955	.3597	.3302	.2970	.2653	.2653
$\alpha(\Re)$ = 100	.5834	.4621	.3635	.3085	.2607	.2413	.1889	.1889	.1889
Full information	.5793	.4495	.3514	.2944	.2534	.2109	.1889	.1889	.1889

F_6^-									
k	1	2	3	4	5	6	7	8	9
$\alpha(\Re)$ = 10	.5687	.5288	.4828	.4358	.3908	.3496	.3129	.2808	.2529
$\alpha(\Re)$ = 100	.5905	.5517	.5033	.4525	.4034	.3585	.3185	.2839	.2542
Full information	.5924	.5542	.5058	.4545	.4050	.3596	.3192	.2843	.2544

	F_7^+		F_8^+	
k	1.5	2.0	1.5	2.0
$\alpha(\Re)$ = 10	.5833	.5301	.5956	.5301
$\alpha(\Re)$ = 100	.5588	.4733	.5792	.4733
Full information	.5524	.4619	.5743	.4619

	F_7^-		F_8^-	
k	1.5	2.0	1.5	2.0
$\alpha(\Re)$ = 10	.5432	.5084	.5540	.5084
$\alpha(\Re)$ = 100	.5736	.5441	.5844	.5441
Full information	.5771	.5487	.5878	.5487

$F_0(x) = x$ but the estimated distribution functions as given in the table are used for the full-information strategy and the Dirichlet rules ($\alpha(\mathfrak{R}) = n$ and $\alpha(\mathfrak{R}) = 100$). The following observations are pertinent in the light of these tables:

1. In both tables all the entries are bounded above by the probability of winning with the full-information strategy and true distribution F_0; namely .6292 for $n = 5$ and .6087 for $n = 10$.

2. As $\alpha(\mathfrak{R})$ increases (the full-information probabilities can be considered as the limit as $\alpha(\mathfrak{R})$ tends to infinity), the probability of winning for the class F^- of distribution functions $(F^-)^{-1}$ which overestimate $F_0(x)$ near 1 is strictly increasing and for the class of functions $(F^+)^{-1}$ which underestimate $F_0(x)$ near 1, it is generally decreasing. Thus, in the presence of the incorrect estimate $(F^+)^{-1}$ for the distribution function, the Dirichlet rules outperform the full-information rule. The reason that this is so is as follows:

The Dirichlet decision numbers d_i' for true distribution $F_0(x)$ are less than the full information indifference numbers d_i for each i due to the ties inherent in the Dirichlet process. Further, most of the non-zero decision numbers are in the interval $[\frac{1}{2},1]$. If $F_0(x)$ is overestimated by the distribution function $(F^-)^{-1}$, the attendant rule is to stop if candidate x is greater than $F^-(d_i)$ (for full information strategy) or $F^-(d_i')$ (for Dirichlet strategy). But $F^-(x) < x$ for x near 1, so F^- deflates the Dirichlet decision number even more. Then as $\alpha(\mathfrak{R})$ increases, the Dirichlet decision numbers with incorrect distribution function increase to $F^-(d_i)$. Thus, the probability of winning is strictly increasing. However, if $F_0(x)$ is underestimated by the distribution function

$(F^+)^{-1}$ near 1, the Dirichlet rule becomes to stop if candidate $x > F^+(d_i')$. But although $d_i' < d_i$, $F^+(d_i') > d_i$, so that the effects of using the Dirichlet rule and underestimating $F_0(x)$ cancel each other out. The Dirichlet rule in this situation outperforms the full-information rule with incorrect distribution function, and, as $\alpha(\mathfrak{R}) \to \infty$, $d_i \to d_i$, causing the probability of winning to generally decrease.

3. For $\alpha(\mathfrak{R}) = n$ and $\alpha(\mathfrak{R}) = 100$, the probability of winning using the Dirichlet strategy for the class F^- of overestimated distribution functions $(F^-)^{-1}$ deviates little from the value for the full information strategy. From Tables 5.1 and 5.2 this deviation rarely exceeds .03 for $n = 5$ and .045 for $n = 10$. In contrast for the F^+ class of functions, the deviations frequently exceed .100 for both $n = 5$ and $n = 10$. In that the Dirichlet rules outperform the full-information rules for the class F^+, this suggests a robustness for the Dirichlet rule in the face of incorrect estimation of the distribution function which the full-information strategy cannot match.

4. If all prior information is ignored, the strategy of the secretary problem is optimal, with probability of winning .4333 for $n = 5$ and .3987 for $n = 10$. It is clear that there are entries in Tables 5.1 and 5.2 respectively, that are smaller than these probabilities of winning. This discrepancy is due to the poor quality of the estimated distribution function. In such situations it is better to have discarded what erroneous prior information one might have and to proceed with the secretary problem's strategy. A more detailed examination of the tables reveals that

such small probabilities occur in Table 5.1 when the measure D is $\geq .25$ for $\alpha(\mathcal{R}) = 100$ and for the full-information strategy and is $\geq .35$ for the Dirichlet rule with $\alpha(\mathcal{R}) = 5$. Suppose that $\alpha(\mathcal{R}) = m$ represents a previous sample of size m on which the prior information and hence the estimated distribution function is based. If one were to use the empirical distribution function $\hat{F}_m$ as the estimate, the distribution of $D = \sup_x |\hat{F}_m(x) - F_0(x)|$ is that of the well-known Kolmogorov-Smirnov statistic. For m = 100, the upper 80% grantile of D is .107 and for m = 5 the upper 80% quantile is .447. (Conover [4]). If a smoothed version of the empirical distribution function were employed as the shape of the parameter α (in order to assure that α is nonatomic), and if $\alpha(\mathcal{R})$ were selected to represent the previous sample size, it is clear that large values of D are not too likely to occur, and, further, that the relative largeness of D depends on $\alpha(\mathcal{R})$. For example, if $\alpha(\mathcal{R}) = 100$, it is unlikely that D be as large as .25 if α is selected on the basis of a previous sample. This self-correcting mechanism for the Dirichlet rules alleviates the problem of when to ignore the prior information by using the secretary problem's strategy, provided that $\alpha(\mathcal{R})$ is sufficiently large. Note that the full-information rule has no self-regulating ability concerning the quality of the prior information.

In conclusion, optimium decision rules have been developed for the Dirichlet game of recognizing the maximum of a sequence. These rules behave quite well for the non-random dowery problem, in some sense intermediate to the case of no prior information and total prior information. Furthermore, these Dirichlet estimates exhibit a property of robustness which the full-information rule with estimated distribution function cannot match. This suggests the Dirihclet strategy in all situations in which some prior information is available, especially when the prior information is quite reliable.

REFERENCES

[1] Berk, R.H. and Savage, I.R., "Dirichlet processes produce discrete measures: An elementary proof," unpublished manuscript, 1973.

[2] Blackwell, D., "Discreteness of Ferguson selections," Annals of Statistics, 1 (March, 1973), 356-358.

[3] Campbell, G. and Hollander, M., "Rank Order Estimation with the Dirichlet Prior," Florida State University Technical Report No. M360 (December, 1975).

[4] Conover, W.J., Practical Nonparametric Statistics, New York: John Wiley and Sons, Inc., 1971.

[5] Ferguson, T.S., "A Bayesian analysis of some nonparametric problems," Annals of Statistics, 1 (March, 1973), 209-230.

[6] Gilbert, J.P. and Mosteller, F., "Recognizing the maximum of a sequence," Journal of the American Statistical Association, 61 (March, 1966), 35-73.

[7] Ramsey, F.L., "A Bayesian approach to bioassay," Biometrics, 28 (September, 1972), 841-858.