INFERENCE ABOUT PERMUTATION PARAMETER IN LARGE SAMPLES*

by

Andrew L. Rukhin
Purdue University, W. Lafayette, IN., USA

Mimeograph Series #81-12

Department of Statistics
Division of Mathematical Sciences
May 1981

*This work was supported by National Science Foundation grant MCS 7802300.

INFERENCE ABOUT PERMUTATION PARAMETER IN LARGE SAMPLES*

by

Andrew L. Rukhin
Purdue University, W. Lafayette, IN., USA

*This work was supported by National Science Foundation Grant MCS 7802300.

ABSTRACT

The estimation problem of a permutation parameter on the basis of a random sample of increasing size is considered. Necessary and sufficient condition for the existence of an estimator asymptotically fully efficient for two different distributions families is derived. We also study the application of this result to cyclic groups of order two and three.

AMS 1970 subject classifications. Primary 62F10, Secondary 62F12, 62A05, 62B10, 94B70.

Key words and phrases: Permutation parameter, asymptotic efficiency, adaptive estimators, maximum likelihood estimator, equivariant estimator, deciphering.

I. Introduction

Let S be a finite set consisting of m elements, say $S = \{1,\ldots,m\}$, and let P be a probability measure on S, $P(k) = p_k$, $k = 1,\ldots,m$, $p_1 + \ldots + p_m = 1$. We assume that a finite group G of transformations acts on the set S. Let $y_1,\ldots,y_n$ be a sequence of independent random variables taking values in S and with common distribution P. Suppose that the statistician observes a transformed sample $x_1,\ldots,x_n$, $x_j = gy_j$, $j = 1,\ldots,n$, where g is unknown transformation from G. In this paper the statistical estimation problem of the permutation parameter g on the basis of random sample $x = (x_1,\ldots,x_n)$ is considered. Let $C = \{c: c \in G, P_c = P\}$ where P_g denotes the shift of measure P by g, i.e., $P_g(k) = P(g^{-1}k)$, $k = 1,\ldots,m$. The parametric space Θ can be identified in this problem with the cosets space G/C, $\Theta = \{\theta_0,\theta_1,\ldots,\theta_\ell\}$, $\theta_0 = C$. Thus an estimator $\delta(x)$ taking values in Θ is desired, or more generally a randomized procedure, which is a probability distribution on Θ, has to be constructed.

This problem has been studied by Maksimov (1975) who constructed estimators which are equivariant with respect to every automorphism of G. The author (Rukhin (1975)) considered the situation when G is a cyclic group of a prime order. In more general situation (see Rukhin (1977)) the asymptotical efficiency $e = -\lim n^{-1} \log P_\theta\{\delta(x) \neq \theta\}$ was evaluated for the class of minimum contrast estimators. It was shown that (unlike the real location parameter problem) there exist probability distributions P and Q such that the maximum likelihood estimator based on Q has the same asymptotical efficiency as the maximum likelihood estimator based on P when the observations have the latter distribution.

The estimation of a finite permutation parameter is motivated by deciphering problem of cryptoanalysis. In this application the sequence of y's represents the plain sequence (message), which must be transmitted, and x's correspond to the cipher sequence. Exactly this estimation problem is discussed informally by Sinkov (1966), Sec. 1.5, 1.7 for monoalphabetic ciphers in the case of cyclic group and Galois field. In a version of the same problem one observes two sequences $x_1,\ldots,x_n$ and $x'_1,\ldots,x'_k$ each belonging to a transformation parameter family with C=G and unknown parameters g and g' respectively. The shift $f^{-1}g'$ between these two samples must be determined. This setting arises in the situation when two (or more) alphabets are used in the enciphering (see Sinkov (1966), Sec. 3.6) and is reduced to our original formulation by treating $x_j^{-n}x'_j$, j=1,...,n as the observed sequence. Efficient deciphering methods have been developed in cryptoanalysis to estimate the "key" g when the language of the message is known (i.e., when the distribution P is given). Assume however that the communication can be written in one of, say, two different languages which use the same alphabet. In other terms there are two probability distributions P and Q over S, and x's have a distribution which is a member of one of the two permutation parameter families.

In this paper we obtain a necessary and sufficient condition on two measures P and Q for the existence of one estimator fully asymptotically efficient under both P and Q which can be considered as adaptive deciphering scheme. It turns out that the natural overall maximum likelihood deciphering scheme may fail to be adaptive while adaptive procedures exist. The case of a cyclic group of order two or three is considered in detail.

II. Asymptotically Efficient Equivariant Estimators

The vector $v = (v_1,\ldots,v_m)$ of relative frequencies of different values from S forms a sufficient statistic for the permutation parameter g. This fact follows immediately from the factorization theorem since

$$\prod_1^n p_{g^{-1}x_j} = \prod_1^m p_k^{v_{gk}} .$$

Therefore every estimator δ can be taken to be a function of v only, $\delta = \delta(v)$.

In this paper we assume that the risk function of an estimator δ is just the probability of incorrect decision $P_\theta(\delta(v)\neq\theta)$. In other terms zero-one loss function is used. A non-randomized estimator is equivariant if $\delta(gv) = g\delta(v)$, where $gv = (v_{g1},\ldots,v_{gm})$, and the extension of this definition to randomized procedures is clear. It is well known that equivariant estimators have constant risk, and that the best equivariant estimator, which minimizes this risk, coincides with the maximum likelihood estimator and is minimax. If the latter is not defined uniquely, i.e., when ties occur, we shall use for concreteness sake the symmetric version

of randomized procedure (see Rukhin (1977)). However the probability of this event is negligible in the asymptotic analysis.

We associate with an equivariant estimator δ the subset $N = N_\delta = \{v: \delta(v)=\theta_0\}$ of the unit simplex Σ in the m-dimensional Euclidean space. The maximum likelihood estimator is randomized on the boundary of the set N (which is an open convex polyhedron) and its consistency just means that $(p_1,\ldots,p_m)$ is an inner point of N. The group G acts on Σ, $gv = (v_{g1},\ldots,v_{gm})$, and if $\theta = g\theta_0$ (such g always exists), then $\{v: \delta(v)=\theta\} = gN$. The maximum likelihood estimator is nonrandomized on the union of all sets gN, $g \in G$, the closure of which coincides with Σ.

It follows from Krafft and Puri (1974) or Rukhin (1977) that the risk of the maximum likelihood estimator $\hat{\delta}$ tends to zero exponentially fast when the sample size n increases. More precisely

$$e_1 = -\lim n^{-1} \log P_\theta\{\hat{\delta}(v)\neq\theta\} = -\max_{\theta\neq\theta_0} \min_t \log\Big(\sum_k p_{\theta k}^{1-t} p_k^t\Big).$$

Therefore for any other estimator δ

$$-\lim n^{-1} \log \max_\theta P_\theta\{\delta(v)\neq\theta\} \leq e_1.$$

If Q is another distribution with probabilities q_k, then it is of interest to find out if there exists an equivariant estimator δ_a which is asymptotically efficient both under $\{P_g\}$ and $\{Q_g\}$, i.e., for which

$$-\lim n^{-1} \log P_\theta\{\delta_a\neq\theta\} = e_1$$

and

$$-\lim n^{-1} \log Q_\theta\{\delta_a\neq\theta\} = e_2,$$

where

$$e_2 = -\max_{\theta\neq\theta_0} \min_t \log\Big(\sum_k g_{\theta k}^{1-t} g_k^t\Big).$$

We call such δ_a adaptive estimator. The next result gives a necessary condition and a sufficient condition for the existence of adaptive procedures. Before formulating it we introduce the following notation. Let, for real d

$$e_{11}(d) = -\max_{\theta\neq\theta_0} \inf_{s,t\geq 0} [\log(\sum_k p_{\theta k}^{s+t} p_k^{1-s} q_k^{-t}) - td],$$

$$e_{21}(d) = -\max_{\theta\neq\theta_0} \inf_{s,t\geq 0} [\log(\sum_k q_{\theta k}^{s+t} p_k^{1-s} q_k^{-t}) + sd],$$

$$e_{22}(d) = -\max_{\theta\neq\theta_0} \inf_{s,t\geq 0} [\log(\sum_k p_{\theta k}^{s+t} p_k^{-s} q_k^{1-t}) + sd],$$

$$e_{12}(d) = -\max_{\theta\neq\theta_0} \inf_{s,t\geq 0} [\log(\sum_k p_{\theta k}^{s+t} p_k^{-s} q_k^{1-t}) - td].$$

Clearly all these functions are continuous on the intervals where they are finite; e_{11}, e_{12} are monotonically increasing and e_{21}, e_{22} are monotonically decreasing. Also $e_{11}(d) \geq e_1$ and $e_{22}(d) \geq e_2$ for all d.

Theorem 1. Assume that P and Q are two distributions over S with positive probabilities p_k and q_k, k = 1,...,m. If an adaptive estimator δ_a exists then for all real d

$$\min\{e_1, e_2 - d\} \leq \min\{e_{11}(d), e_{21}(d), e_{22}(d) - d, e_{12}(d) - d\}. \tag{1}$$

If for some real d

$$\begin{aligned} e_1 &= \min\{e_{11}(d), e_{21}(d)\}, \\ e_2 &= \min\{e_{22}(d), e_{12}(d)\}, \end{aligned} \tag{2}$$

then an adaptive estimator exists.

Proof. Assume that an adaptive estimator δ_a exists. Define $\tilde{\delta}$ as the maximum likelihood esimator based on the log-likelihood function

$$n \max\{\sum_k v_{g^{-1}k} \log p_k, \Sigma v_{g^{-1}k} \log q_k + d\}, \text{ i.e.,}$$

$$\max\{\sum_k v_k \log p_{\tilde{\delta}k}, \Sigma v_k \log q_{\tilde{\delta}k} + d\} = \max_g\{\sum_k v_k \log p_{gk}, \Sigma v_k \log q_{gk} + d\}.$$

Then $\tilde{\delta}$ is the best equivariant estimator for the distribution

$$\Gamma_n(v) = C_n \max\{\prod_k p_k^{nv_k}, e^{nd} \prod_k g_k^{nv_k}\} ,$$

and

$$C_n \max\{P\{\tilde{\delta}(v)\neq\theta_0\}, e^{nd}Q\{\tilde{\delta}(v)\neq\theta_0\}\} \leq \sum_{v:\tilde{\delta}(v)\neq\theta_0} \Gamma_n(v) \leq \sum_{v:\delta(v)\neq\theta_0} \Gamma_n(v)$$

$$\leq C_n[P\{\delta(v)\neq\theta_0\} + e^{nd}Q\{\delta(v)\neq\theta_0\}].$$

Therefore

$$-\lim n^{-1} \log P\{\tilde{\delta}(v)\neq\theta_0\}$$

$$\leq \min\{-\lim n^{-1} \log P\{\delta(v)\neq\theta_0\}, -\lim n^{-1} \log Q\{\delta(v)\neq\theta_0\}-d\}$$

$$= \min\{e_1, e_2-d\}$$

and analogously

$$-d - \lim n^{-1} \log Q\{\tilde{\delta}(v)\neq\theta_0\} \leq \min\{e_1, e_2-d\}.$$

We deduce the necessary condition of Theorem 1 by evaluating $\lim n^{-1} \log P\{\tilde{\delta}(v)\neq\theta_0\}$ and $\lim n^{-1} \log Q\{\tilde{\delta}(v)\neq\theta_0\}$.

Let A_g denote the following event,

$$A_g = \{\max\{\Sigma v_k \log p_k, \Sigma v_k \log q_k+d\} < \max\{\Sigma v_k \log p_{gk}, \Sigma v_k \log q_{gk}+d\}\};$$

then for any fixed $g \in \Theta$, $g \neq \theta_0$

$$P(A_g) \leq P\{\tilde{\delta}(v)\neq\theta_0\} \leq P\{\bigcup_{g\neq\theta_0} A_g\} \leq \ell \max_{g\neq\theta_0} P\{A_g\}.$$

Thus

$$n^{-1} \log P\{\tilde{\delta}(v)\neq\theta_0\} \sim n^{-1} \log \max_{g\neq\theta_0} P\{A_g\} .$$

Also

$$\max\Big\{P\{\max\{\Sigma v_k \log p_k, \Sigma v_k \log q_k + d\} \le \Sigma v_k \log p_{gk}\},$$

$$P\{\max\{\Sigma v_k \log p_k, \Sigma v_k \log q_k + d\} \le \Sigma v_k \log q_{gk} + d\}\Big\} \le P(A_g)$$

$$\le P\{\max\{\Sigma v_k \log p_k, \Sigma v_k \log q_k + d\} \le v_k \log p_{gk}\}$$

$$+ P\{\max\{\Sigma v_k \log p_k, \Sigma v_k \log q_k + d\} \le v_k \log q_{gk} + d\}.$$

The asymptotical behavior of the latter probabilities can be obtained by means of the multivariate version of Chernoff's Theorem (see Steinebach (1978), Groeneboom et al (1979)) or Sanov's Theorem (Bahadur (1971) p.18 formula (5.35)). According to these theorems

$$\lim n^{-1} \log P\{\max\{\Sigma v_k \log p_k, \Sigma v_k \log q_k + d\} \le \Sigma v_k \log p_{gk}\}$$

$$= \lim n^{-1} \log P\{\Sigma v_k \log (p_{gk}/p_k) \ge 0, \Sigma v_k (\log(p_{gk}/q_k) - d) \ge 0\}$$

$$= \inf_{s,t \ge 0} [\log(\sum_k p_{gk}^{s+t} p_k^{1-s} q_k^{-t}) - td]$$

and

$$\lim n^{-1} \log P\{\max\{\Sigma v_k \log p_k, \Sigma v_k \log q_k + d\} \le \Sigma v_k \log q_{gk} + d\}$$

$$= \lim n^{-1} \log P\{\Sigma v_k \log (q_{gk}/q_k) \ge 0, \Sigma v_k (\log(q_{gk}/p_k) + d) \ge 0\}$$

$$= \inf_{s,t \ge 0} [\log(\sum_k q_{gk}^{s+t} p_k^{1-s} q_k^{-t}) + sd].$$

Therefore

$$-n^{-1} \log P\{\tilde{\delta}(v) \ne \theta_0\} \sim \min\{e_{11}(d), e_{21}(d)\}. \tag{3}$$

Similar argument establishes the formula

$$-n^{-1} \log Q\{\tilde{\delta}(v) \ne \theta_0\} \sim \min\{e_{22}(d), e_{12}(d)\}. \tag{4}$$

Thus if an adaptive estimator exists, (1) holds. Also if (2) is true, then because of (3) and (4) the estimator $\tilde{\delta}$ is adaptive.

Remark. It follows from the proof of Theorem 1 that for all d

$$e_1 \geq \min\{e_{11}(d), e_{21}(d)\}, e_2 \geq \min\{e_{22}(d), e_{12}(d)\}. \tag{5}$$

Define

$$d_{21} = \sup\{d: e_{21}(d) \geq e_1\},$$

$$d_{12} = \inf\{d: e_{12}(d) \geq e_2\}.$$

Since $e_{12}(d) \to +\infty$ as $d \to +\infty$ and $e_{12}(d)$ is continuous from the right, $e_{12}(d_{12}) = e_2$. If $e_{21}(d_{21}-) < \infty$, then also $e_{21}(d_{21}) = e_1$.

Theorem 2. Under assumptions of Theorem 1 adaptive estimator exists if and only if

$$d_{12} \leq d_{21}. \tag{6}$$

Proof. From the definition of numbers d_{12} and d_{21} it follows that $e_{21}(d) \geq e_1$ for $d \leq d_{21}$ and $e_{12}(d) \geq e_2$ for $d \geq d_{12}$. Therefore of $d_{12} \leq d_{21}$ then for all d from the closed interval $[d_{12}, d_{21}]$

$$\min\{e_{11}(d), e_{21}(d)\} \geq e_1,$$

i.e., because of (5)

$$e_1 = \min\{e_{11}(d), e_{21}(d)\}$$

and similarly

$$e_2 = \min\{e_{22}(d), e_{12}(d)\}.$$

Theorem 1 implies the existence of an adaptive estimator.

Assume now that (6) does not hold. Then for d belonging to the open interval (d_{21}, d_{12}), $e_{21}(d) < e_1$ and $e_{12}(d) < e_2$. For such d

$$\min\{e_1, e_2-d\} > \min\{e_{21}(d), e_{12}(d)-d\}$$

$$\geq \min\{e_{11}(d), e_{21}(d), e_{22}(d)-d, e_{12}(d)-d\},$$

and because of (1) adaptive estimator cannot exist.

Corollary 1. If condition (6) holds then for all $d, d_{12} \leq d \leq d_{21}$ the estimator δ_a,

$$\{v: \delta_a(v)=\theta_0\} = \{v: \max\{\Pi p_k^{v_k}, e^d \Pi q_k^{v_k}\} = \max_g \max\{\Pi_k p_k^{v_{gk}}, e^d \Pi q_k^{v_{gk}}\}\}$$

is adaptive.

This fact easily follows from the proof of Theorem 1. It is also easy to see that if the interval $[d_{12}, d_{21}]$ is nonempty and does not contain zero, then the usual maximum likelihood procedure which corresponds to $d = 0$ is not adaptive, while adaptive procedures do exist.

Corollary 2. Adaptive estimator exits if

$$e_2 < e_{12}(d_{21})$$

or if

$$e_1 < e_{21}(d_{12}).$$

If an adaptive estimator exists, then

$$e_2 \leq e_{12}(d_{21}).$$

If an adaptive estimator exists and $e_{21}(d_{12}-) < \infty$ then

$$e_1 \leq e_{21}(d_{12}).$$

Proof. If $e_2 < e_{12}(d_{21})$, then for all d greater than $d_{12}, e_{12}(d) > e_2$, which implies that $d_{12} \leq d_{21}$. Analogously the inequality $e_1 < e_{21}(d_{12})$ implies $d_{12} \leq d_{21}$, so that in these cases adaptive estimator exists.

Assume now that an adaptive procedure exists, i.e., that $d_{12} \leq d_{21}$. If $d_{12} < d_{21}$ then $e_{21}(d_{12}) \geq e_{21}(d_{21})$ and $e_{12}(d_{12}) \leq e_{12}(d_{21})$. Since $e_{12}(d_{12}) \leq e_2$ it follows that $e_2 \leq e_{12}(d_{21})$. Because of the condition $e_{21}(d_{12}-) < \infty$, the function e_{21} is continuous at d_{12}, so that $e_{21}(d_{21}) = e_1$ and $e_1 \leq e_{21}(d_{12})$.

If $d_{12} = d_{21}$, then $e_2 \leq e_{12}(d_{12}) = e_{12}(d_{21})$, and also $e_1 = e_{21}(d_{21}) = e_{21}(d_{12})$. It follows from this argument that if e_{21} (or e_{12}) is strictly monotone function then adaptive estimator exists if and only if $e_2 \leq e_{21}(d_{12})$ (or $e_2 \leq e_{12}(d_{21})$).

Corollary 3. Adaptive estimator exists in each of the following cases:

(i) $e_{21}(0) \geq e_1, e_{12}(0) \geq e_2$;

(ii) $d_{21} \geq e_2$;

(iii) $d_{12} \leq -e_1$.

Proof. In the case (i) $d_{12} \leq 0 \leq d_{21}$, so that adaptive estimator exists. One proves (ii) and (iii) by noticing that $e_{21}(d) \geq -d$ and $e_{12}(d) \geq d$, so that $e_2 \geq d_{12}$ and $e_1 \geq -d_{21}$.

III. Adaptive Estimators of Permutation Parameter for Cyclic Groups

In this section we illustrate Theorem 2 by considering adaptive estimators of permutation parameter for cyclic groups of order two and three.

If $G = S = \{0,1\}$ with the group operation being addition modulo two, and P and Q are two probability distributions over S, with positive probabilities, $p_0 \neq p_1$, $q_0 \neq q_1$, then

$$e_1 = -\min_{0 \leq t \leq 1} \log(p_0^{1-t}p_1^t + p_1^{1-t}p_0^t) = -\log(2\sqrt{p_0 p_1}),$$

$$e_2 = -\log(2\sqrt{q_0 q_1}).$$

We assume that $q_1 > q_0$, i.e. $q_1 > 1/2$. Then

$$e_{21}(d) = -\inf_{s,t\geq 0}[\log(q_1^{s+t}p_0^{1-s}q_0^{-t}+q_0^{s+t}p_1^{1-s}q_1^{-t})+sd]$$

$$= -\min\{\inf_{s:q_1^s p_0^{1-s}\geq q_0^s p_1^{1-s}}[\log(q_1^s p_0^{1-s}+q_0^s p_1^{1-s})+sd],$$

$$\inf_{s:q_1^s p_0^{1-s}<q_0^s p_1^{1-s}}[\log(2(q_0q_1)^{s/2}(p_0p_1)^{(1-s)/2})+sd]\}.$$

If $p_1 > p_0$, then $p_0q_0 < p_1q_1$ and $\{s: q_1^s p_0^{1-s}\geq q_0^s p_1^{1-s}\} = [s_0,\infty)$, where the number s_0 is found from the formula $q_1^{s_0}p_0^{1-s_0} = q_0^{s_0}p_1^{1-s_0}$, $s_0 < 1$. Therefore $e_{21}(d) = \infty$ if $\log(q_0q_1/(p_0p_1)) + 2d < 0$, which implies that

$$2d_{21} \geq \log(q_0q_1/(p_0p_1)).$$

Also if $\log(q_0q_1/(p_0p_1)) + 2d \geq 0$,

$$e_{21}(d) = -\min_{0\leq s\leq s_0}[\log(q_1^s p_0^{1-s}+q_0^s p_1^{1-s})+sd];$$

and

$$e_{21}(\log(p_0p_1/(q_0q_1))/2)$$

$$= -\log(2(q_0q_1)^{s_0/2}(p_0p_1)^{(1-s_0)/2}) - s_0\log(p_0p_1/(q_0q_1))/2 = e_1.$$

One concludes that if $p_1 > p_0$, then

$$d_{21} = \log(p_0p_1/(q_0q_1))/2.$$

Analogously

$$e_{12}(d) = -\min\{\inf_{0\leq t\leq t_0}[\log(p_1^t q_0^{1-t}+p_0^t q_1^{1-t})-td],$$

$$\inf_{t>t_0}[\log(2(p_0p_1)^{t/2}(q_0q_1)^{(1-t)/2})-td]\},$$

where t_0 is defined by the formula $p_1^{t_0} q_0^{1-t_0} = p_0^{t_0} q_1^{1-t_0}$.

Thus

$$e_{12}(d) = - \min_{0 \le t \le t_0} [\log(p_1^t q_0^{1-t} + p_0^t q_1^{1-t}) - td],$$

if $\log(p_0 p_1/(q_0 q_1)) - 2d \ge 0$; and $e_{12}(d) = +\infty$, otherwise. Therefore

$$d_{12} \le \log(p_0 p_1/(q_0 q_1))/2 = d_{21}.$$

According to Theorem 2 adaptive estimator exists when $p_1 > p_0$, $q_1 > q_0$. It is easy to see that in this situation the estimator δ_1, $\{v: \delta_1(v)=0\}=\{v_0=\min_k v_k\}$ is adaptive. In the case when $p_1 < p_0$, $q_1 < q_0$ (which can be treated quite similarly to the previous one) the adaptive estimator δ_0 also exists. This estimator has the form $\{v: \delta_0(v)=0\}=\{v_0=\max_k v_k\}$.

We show now that if $q_1 > q_0$ and $p_1 < p_0$ then there is no adaptive procedure. If $p_0 p_1 > q_0 q_1$, which means $q_0 p_0 \le q_1 p_1$ or $q_1 > p_0$ then

$$d_{12} = \log(p_0 p_1/(q_0 q_1))/2.$$

Also

$$e_{21}(d) = -\min_{s>0}[\log(q_1^s p_0^{1-s} + q_0^s p_1^{1-s}) + sd] = 0,$$

if $d + p_0 \log(q_1/p_0) + p_1 \log(q_0/p_1) \ge 0$.

But

$$\log(p_0 p_1/(q_0 q_1))/2 + p_0 \log(q_1/p_0) + p_1 \log(q_0/p_1) \ge 0.$$

Indeed for fixed $p_0 > 1/2$, $p_1 = 1 - p_0$, the function

$$H(r) = \log(p_0 p_1/(r(1-r)))/2 + p_0 \log(r/p_0) + p_1 \log((1-r)/p_1)$$

is increasing for r, $0 < r < 1$, and $H(p_0) = 0$. Therefore $H(r) > 0$ for $r > p_0$, so that

$$d_{12} + p_0 \log(q_1/p_0) + p_1 \log(q_0/p_1) \geq 0,$$

and $e_{21}(d_{12}) = 0 < e_1$.

Thus because of Corollary 2 adaptive estimators cannot exist.

If $p_0p_1 \leq q_0q_1$, then similarly

$$d_{21} = \log((p_0p_1)/(q_0q_1))/2$$

and

$$e_{12}(d_{21}) = 0 < e_2,$$

so that again necessary condition for the existence of adaptive procedures is violated.

We formulate these results as

Theorem 3. Adaptive estimator of the permutation parameter from a cyclic group of order two exists for two probability distributions P and Q if and only if either $p_0 > p_1$, $q_0 > q_1$ or $p_0 < p_1$, $q_0 < q_1$.

Now let us consider the case of cyclic group of order three, $G = S = \{0,1,2\}$. Assume that both distributions P and Q are symmetric, i.e. $p_1 = p_2 > 0$, $q_1 = q_2 > 0$ and non-uniform, $p_1 \neq 1/3$, $q_1 \neq 1/3$. As in the previous problem adaptive estimator exists if $p_0 > p_1$, $q_0 > q_1$ or if $p_0 < p_1$, $q_0 < q_1$. Indeed estimators $\delta_0(v)$ and $\delta_1(v)$ are adaptive in the first and in the second case respectively. This fact can be obtained with the help of Theorem 2, but also easily follows by noticing that δ_0 (δ_1) is the maximum likelihood estimator of the permutation parameter for any symmetric distribution P such that $p_0 > p_1$ ($p_0 < p_1$). Figures 1 and 2 show the form of the corresponding subset N of the unit simplex. It is clear from the form of these sets that, for instance, δ_0 is consistent (although not asymptotically fully efficient) for all (not necessarily symmetric) distributions P such that $p_0 > \max_{k\neq 0} p_k$.

One has

$$e_1 = -\min_{0\le t\le 1} \log(p_0^{1-t}p_1^t+p_1^{1-t}p_0^t+p_1) = -\log(p_1+2\sqrt{p_0p_1}),$$

$$e_2 = -\log(q_1+2\sqrt{q_0q_1}).$$

The adaptiveness of δ_0 and δ_1 constitutes a part of the following result.

Theorem 4. Adaptive estimator of the permutation parameter from a cyclic group of order three exists for two symmetric probability distributions P and Q if and only if one of the following conditions holds:

(i) $p_0 > p_1$, $q_0 > q_1$

(ii) $p_0 < p_1$, $q_0 < q_1$

(iii) $p_0 > p_1$, $q_0 < q_1$,

$$\min_{0\le s\le 1} \log(\tilde{p}_0^{1-s}\tilde{q}_1^s+\tilde{p}_1^{1-s}\tilde{q}_0^s+\tilde{p}_1^{1-s}\tilde{q}_1^s] \le 0, \tag{7}$$

where $\tilde{p}_i = p_i\exp\{e_1\}$, $\tilde{q}_i = q_i\exp\{e_2\}$, $i = 0,1$,

(iv) $p_0 < p_1$, $q_0 > q_1$ and (7) is satisfied.

Proof. It suffices to consider the case (iii). Assume that $p_1q_1 > p_0q_0$. (The case $p_1q_1 \le p_0q_0$ is treated analogously.)

We have

$$e_{21}(d) = -\inf_{s\ge 0}[\log(p_0^{1-s}q_1^s+p_1^{1-s}q_0^s+p_1^{1-s}q_1^s)+sd]$$

so that $e_{21}(d) = 0$ if $-K_2 + d \ge 0$, where

$$K_2 = p_1 \log (p_1/q_0) + p_0 \log (p_0/q_1) + p_1 \log (p_1/q_1).$$

Also e_{21} is strictly decreasing for $d < K_2$, and hence

$$d_{21} < K_2.$$

If

$$d \geq -[q_1 \log (q_1/p_0)+q_0 \log (q_0/p_1)+q_1 \log (q_1/p_1)] = -K_1,$$

which means that the derivative at s = 1 of a (convex) function $\log(p_0^{1-s}q_1^s+p_1^{1-s}q_0^s+p_1^{1-s}q_1^s) + sd$ is nonnegative, then

$$e_{21}(d) = - \min_{0\leq s\leq 1} [\log(p_0^{1-s}q_1^s+p_1^{1-s}q_0^s+p_1^{1-s}q_1^s +sd].$$

and

$$e_{21}(-K_1) = K_1.$$

An easy calculation shows that

$$e_1 = - \min_{1/2>r>1/3} [r \log(r/p_0)+(1-2r)\log((1-2r)/p_1)+r \log (r/p_1)].$$

Since $1/3 < q_1 < 1/2$,

$$e_1 \leq K_1.$$

It follows that

$$e_1 = e_{21}(d_{21}) < e_{21}(-K_1)$$

and

$$-K_1 < d_{21} < K_2. \tag{8}$$

Now let us turn to the function e_{21},

$$e_{12}(d) = -\min\{ \inf_{0\leq t\leq t_1} [\log(p_1^t q_1^{1-t}+p_1^t q_0^{1-t}+p_0^t q_1^{1-t})-td],$$

$$\inf_{t>t_1} [\log(p_1^t q_1^{1-t}+2(p_0p_1)^{t/2}(q_0q_1)^{(1-t)/2})-td]\},$$

where t_1 is defined by the formula $p_0^{t_1}q_1^{1-t_1} = p_1^{t_1}q_0^{1-t_1}$, $t_1 > 1$.

Thus

$$e_{12}(d) = - \min_{0\leq t\leq t_1} [\log(p_1^t q_1^{1-t}+p_1^t q_0^{1-t}+p_0^t q_1^{1-t})-td]$$

if

$$d < d_1 = [\sum_i p_i^{t_1} q_{i+1}^{1-t_1} \log(p_i/q_{i+1})]/[\sum_i p_i^{t_1} q_{i+1}^{1-t_1}].$$

Indeed for $d < d_1$

$$\inf_{t>t_1} [\log(p_1^t q_1^{1-t} + 2(p_0 p_1)^{t/2}(q_0 q_1)^{(1-t)/2}) - td]$$

$$= \log(p_1^{t_1} q_1^{1-t_1} + 2(p_0 p_1)^{t_1/2}(q_0 q_1)^{(1-t_1)/2}) - t_1 d$$

$$= \log(p_1^{t_1} q_1^{1-t_1} + p_1^{t_1} q_0^{1-t_1} + p_0^{t_1} q_1^{1-t_1}) - t_1 d$$

$$\geq \inf_{0\leq t\leq t_1} [\log(p_1^t q_1^{1-t} + p_1^t q_0^{1-t} + p_0^t q_1^{1-t}) - td].$$

It is easy to see that the point of minimum $t = t(d)$ of the function $\log(p_1^{1-t} q_1^t + p_1^t q_0^{1-t} + p_0^t q_1^{1-t}) - td$ is an increasing function of d. Since $t(K_2) = 1$ and $t(d_1) = t_0 > 1$ one concludes that $K_2 \leq d_1$.

Because of (8)

$$e_1 = e_{21}(d_{21}) = -\min_{0\leq s\leq 1} [\log(p_0^{1-s} q_1^s + p_1^{1-s} q_0^s + p_1^{1-s} q_1^s) + sd_{21}]$$

$$= -\min_{0\leq t\leq 1} [\log(p_0^t q_1^{1-t} + p_1^t q_0^{1-t} + p_1^t q_1^{1-t}) + (1-t)d_{21}]$$

$$= -d_{21} + e_{12}(d_{21}).$$

We know that an adaptive estimator exists if and only if

$$e_{12}(d_{21}) \geq e_2,$$

which means that

$$e_2 - e_1 \leq d_{21}. \tag{9}$$

It is easy to check that

$$e_2 = \min_{0<r<1/3} [r \log(r/q_0) + (1-2r)\log((1-2r)/q_1) + r \log(r/q_1)]$$

which implies the inequality

$$e_2 \leq K_2,$$

so that

$$-K_1 \leq -e_1 < e_2 - e_1 < e_2 \leq K_2.$$

Because of this relation, (9) is equivalent to the following inequality

$$e_1 = e_{21}(d_{21}) \leq e_{21}(e_2-e_1)$$

that is,

$$-\min_{0\leq s\leq 1}[\log(p_0^{1-s}q_1^s+p_1^{1-s}q_0^s+p_1^{1-s}q_1^s)+s(e_2-e_1)] \geq e_1.$$

The latter formula is clearly equivalent to (7), which proves Theorem 4.

The quantity in (7) was evaluated numerically for different distributions P and Q. In Figure 3 the subset of the square

$$\{(p_1,q_1),0<p_1<1/2,0<q_1<1/2\},$$

where (7) holds, is shaded. One can interpret Figure 3 by stating that an adaptive procedure exists unless distributions P and Q are quite dissimilar.

In the case when (7) is satisfied an adaptive estimator can be obtained as the maximum likelihood estimator from the proof of Theorem 1 with $d = e_2 - e_1$. In Figure 4 and Figure 5 the corresponding subsets N of the unit simplex are shown for $p_1 = 0.1$, $q_1 = 0.35$, $d = -0.41$ and $p_1 = 0.3$, $q_1 = 0.49$, $d = 0.37$.

Since there is no adpative procedure for any pair of symmetric distributions it is of interest to find a reasonable estimator of the permutation parameter, which can be used with no information about p_0. The natural choice seems to be the overall maximum likelihood estimator $\hat{\delta}_0$ with respect to all symmetric distributions.

This estimator is defined in the following way

$$N_0 = \{v: \hat{\delta}_0(v)=0\}$$

$$= \{v: \max_{0\le p_0\le 1} p_0^{v_0}[(1-p_0)/2]^{1-v_0} = \max_k \max_{0\le p_0\le 1} p_0^{v_k}[(1-p_0)/2]^{1-v_k}\}$$

$$= \{v: v_0 \log v_0+(1-v_0)\log((1-v_0)/2) = \max_k [v_k \log v_k+(1-v_k)\log((1-v_k)/2)\}.$$

The form of this set N_0 is given in Figure 6.

Table 1

The Asymptotic Efficiency e_0 of the Overall Maximum Likelihood Estimator and the Asymptotic Efficiency e of the Maximum Likelihood Estimator

p	e_0	e	p	e_0	e
0.001	0.147	0.608	0.51	0.017	0.049
0.002	0.141	0.576	0.52	0.019	0.055
0.005	0.132	0.515	0.53	0.021	0.061
0.01	0.117	0.453	0.54	0.022	0.067
0.02	0.103	0.374	0.55	0.024	0.074
0.03	0.090	0.320	0.56	0.027	0.081
0.04	0.081	0.278	0.57	0.030	0.089
0.05	0.071	0.244	0.58	0.033	0.097
0.06	0.064	0.215	0.59	0.036	0.105
0.07	0.058	0.191	0.60	0.039	0.113
0.08	0.052	0.170	0.61	0.042	0.129
0.09	0.047	0.151	0.62	0.045	0.132
0.10	0.041	0.134	0.63	0.048	0.142
0.11	0.038	0.119	0.64	0.051	0.152
0.12	0.033	0.106	0.65	0.054	0.163
0.13	0.029	0.094	0.66	0.058	0.174
0.14	0.025	0.083	0.67	0.062	0.186
0.15	0.022	0.073	0.68	0.066	0.199
0.16	0.020	0.064	0.69	0.070	0.212
0.17	0.018	0.055	0.70	0.074	0.266
0.18	0.016	0.048	0.71	0.078	0.240
0.19	0.014	0.041	0.72	0.082	0.255
0.20	0.011	0.035	0.73	0.087	0.271
0.21	0.009	0.030	0.74	0.093	0.287
0.22	0.008	0.024	0.75	0.099	0.305
0.23	0.007	0.020	0.76	0.105	0.323
0.24	0.006	0.016	0.77	0.111	0.342
0.25	0.005	0.013	0.78	0.117	0.363
0.26	0.004	0.010	0.79	0.123	0.384
0.27	0.003	0.007	0.80	0.130	0.407
0.28	0.002	0.005	0.81	0.137	0.431
0.29	0.001	0.003	0.82	0.145	0.457
0.30	0.001	0.002	0.83	0.153	0.484
0.31	0.000	0.001	0.84	0.162	0.514
0.32	0.000	0.000	0.85	0.171	0.544
0.35	0.000	0.000	0.86	0.180	0.578
0.36	0.000	0.001	0.87	0.189	0.615
0.37	0.001	0.002	0.88	0.199	0.654
0.38	0.001	0.004	0.89	0.209	0.698
0.39	0.002	0.005	0.90	0.222	0.746
0.40	0.003	0.007	0.91	0.239	0.799
0.41	0.003	0.009	0.92	0.260	0.857
0.42	0.004	0.012	0.93	0.282	0.927
0.43	0.005	0.015	0.94	0.308	1.006
0.44	0.006	0.018	0.95	0.342	1.099
0.45	0.007	0.022	0.96	0.384	1.214
0.46	0.008	0.022	0.97	0.442	1.362
0.47	0.009	0.026	0.98	0.526	1.570
0.48	0.001	0.034	0.99	0.675	1.926
0.49	0.013	0.039	0.995	0.829	2.280
0.50	0.015	0.044	0.998	1.035	2.746
			0.999	1.192	3.097

In Table 1 the asymptotic efficiency of the estimator $\hat{\delta}_0$,

$$e_0 = -\lim n^{-1} \log P_\theta\{\hat{\delta}_0(v) \neq \theta\},$$

and the asymptotic efficiency $e = -\log(p_1 + 2\sqrt{p_0 p_1})$ of the maximum likelihood estimator δ are evaluated as functions of $p = p_0$, $0 < p < 1$. The relatively worst behavior of $\hat{\delta}_0$ happens when $p_0 \to 1/3$. In this case both estimators have vanishing efficiencies, and $e_0 \sim e/3$. If $p_0 \to 0$, then $e \to \log 2$ and $e_0 \sim 0.773e$; if $p_0 \to 1$, then $e_0 \sim 0.454e$, $e \to \infty$.

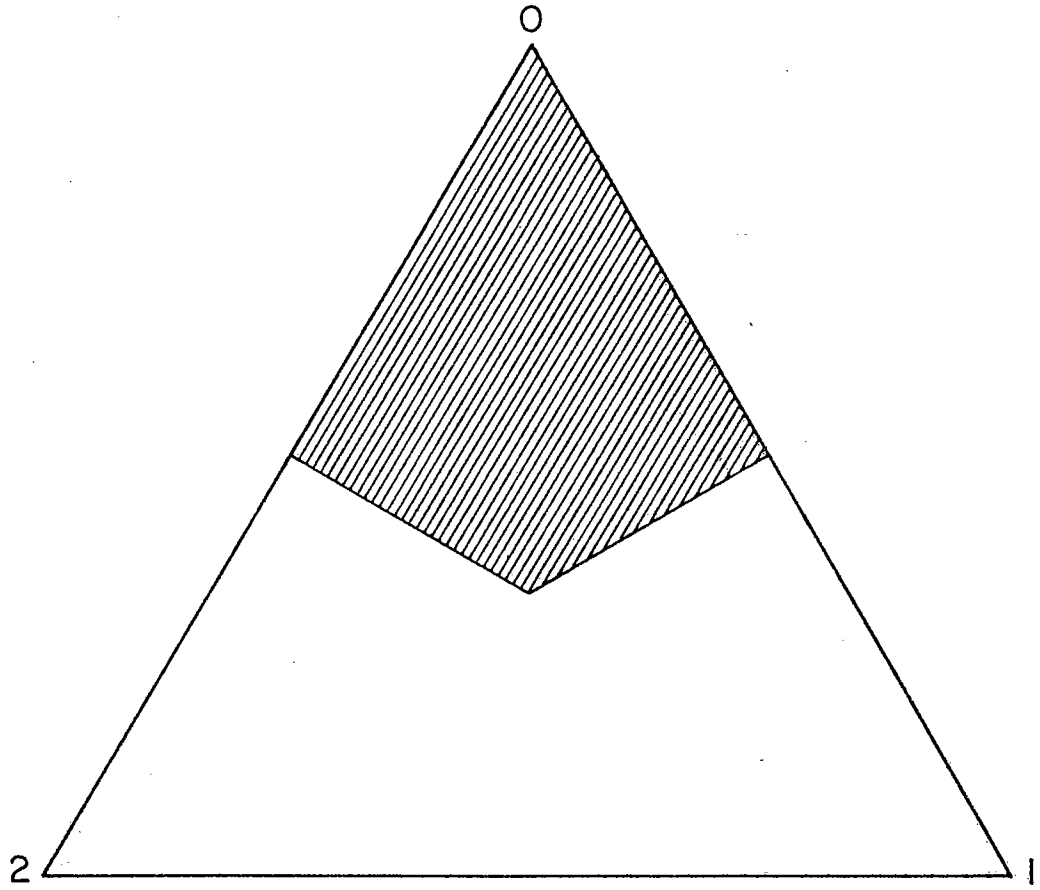

Figure 1. The Shaded Area Represents the Set N Corresponging to the Estimator δ_0.

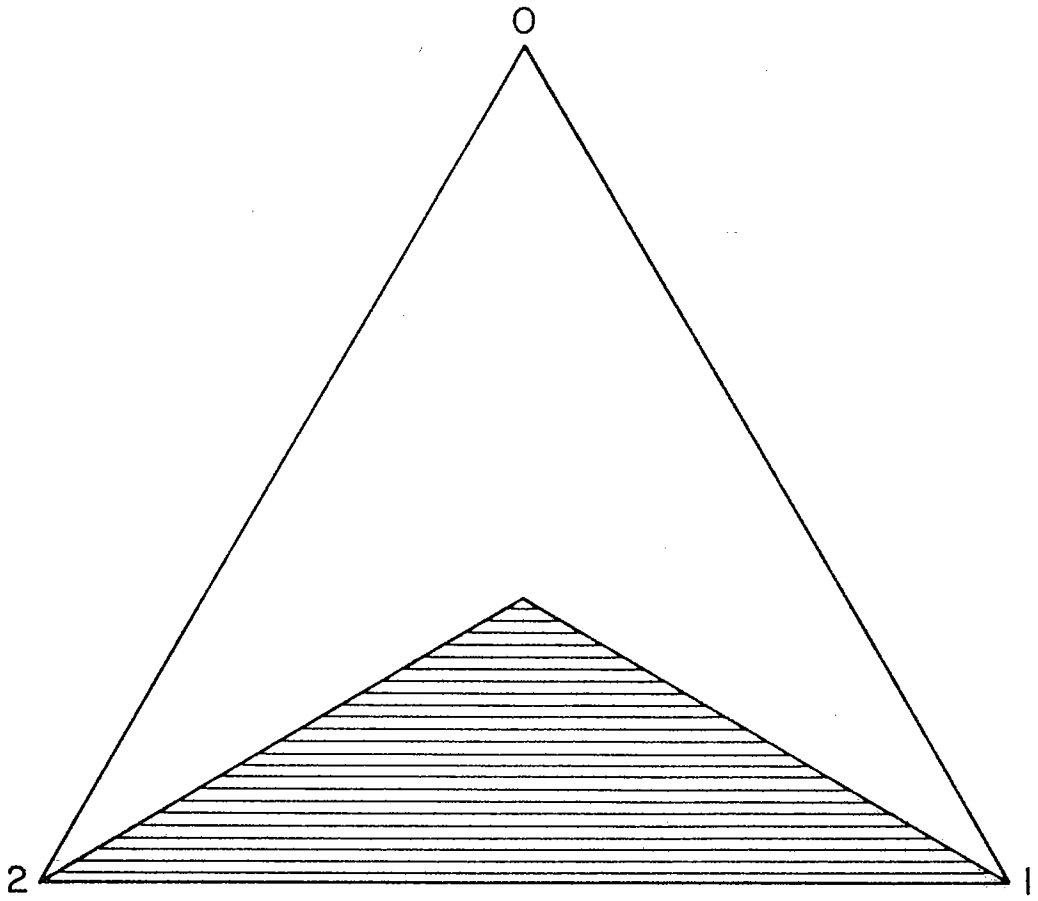

Figure 2. The Shaded Area Represents the Set N Corresponding to the Estimator δ_1.

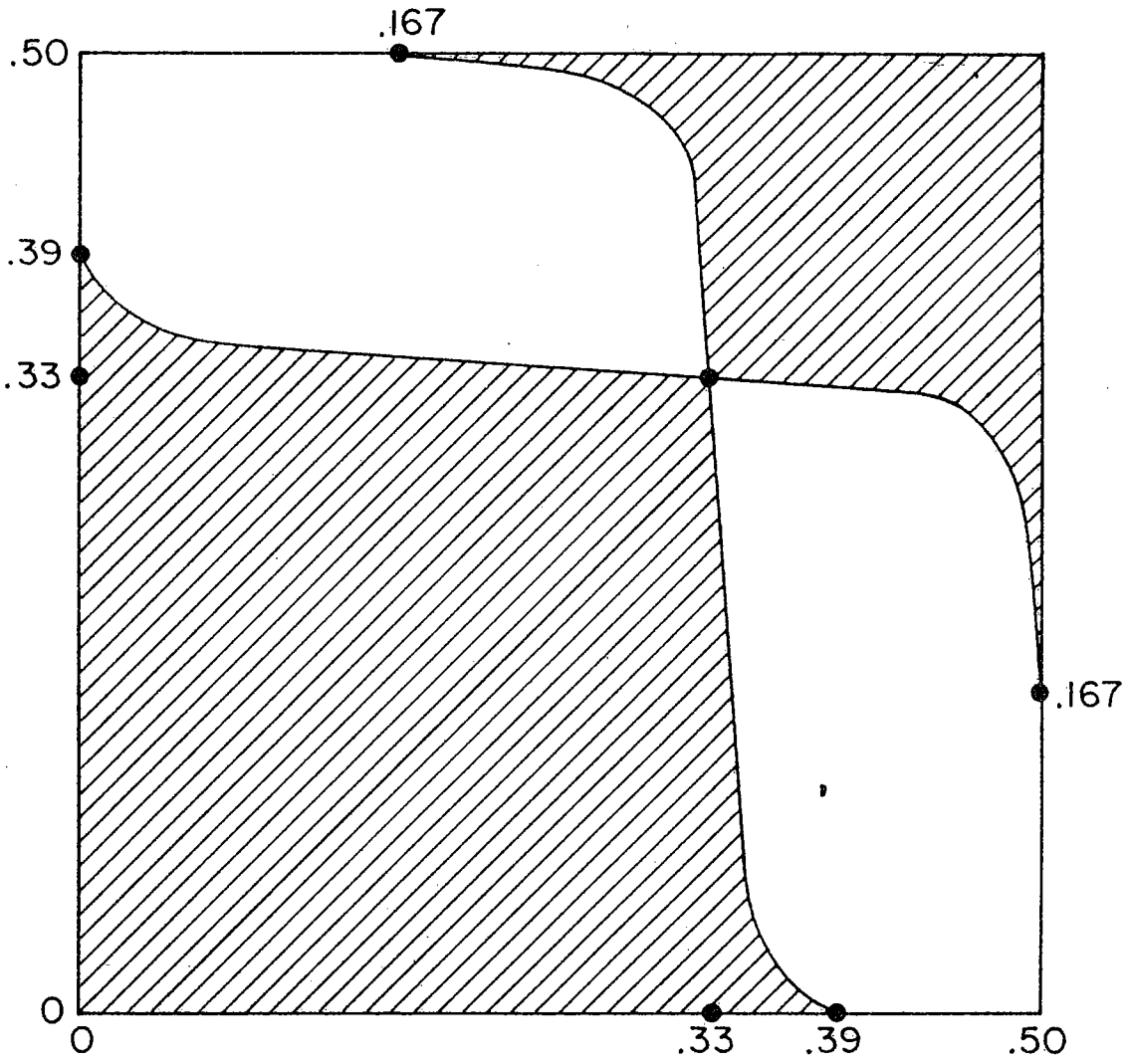

Figure 3. The Shaded Area Represents the Set of Probabilities (p_1, q_1) for which an Adaptive Estimator Exists.

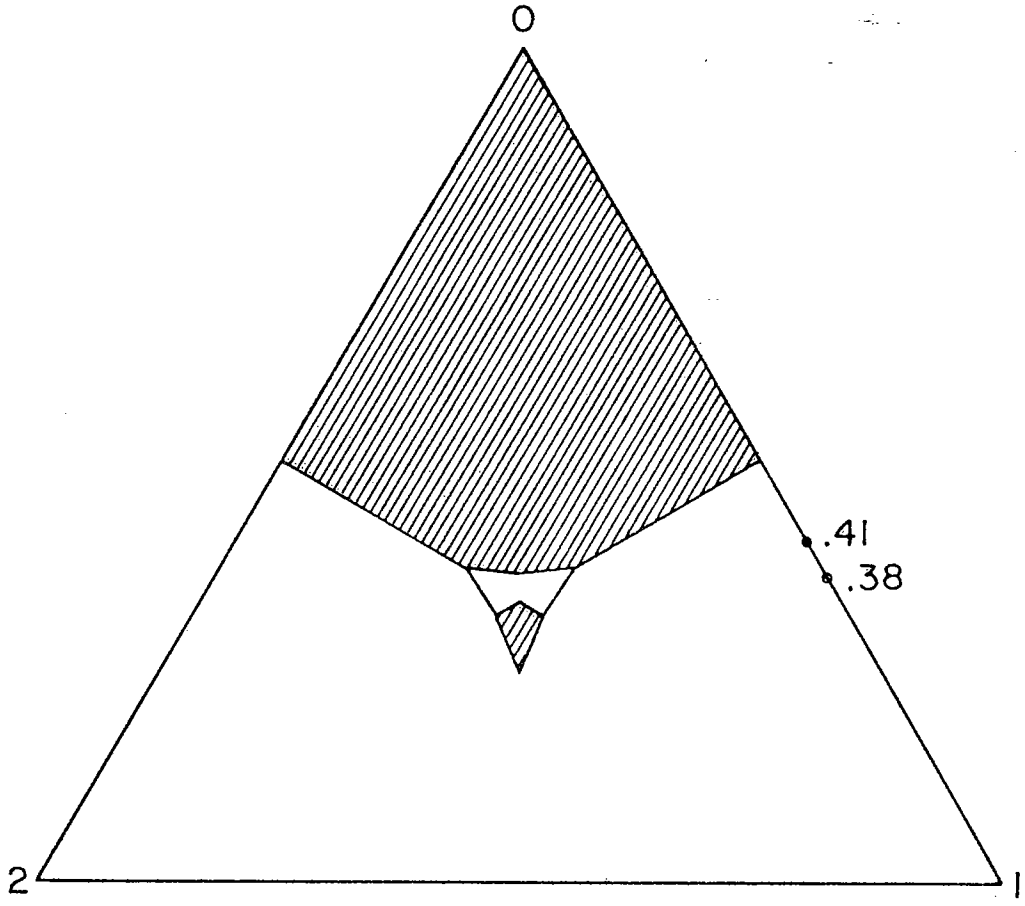

Figure 4. The Shaded Area Represents the Set N Corresponding to the Adaptive Estimator δ_a when $p_1 = 0.1$, $q_1 = 0.35$, $d = -0.41$.

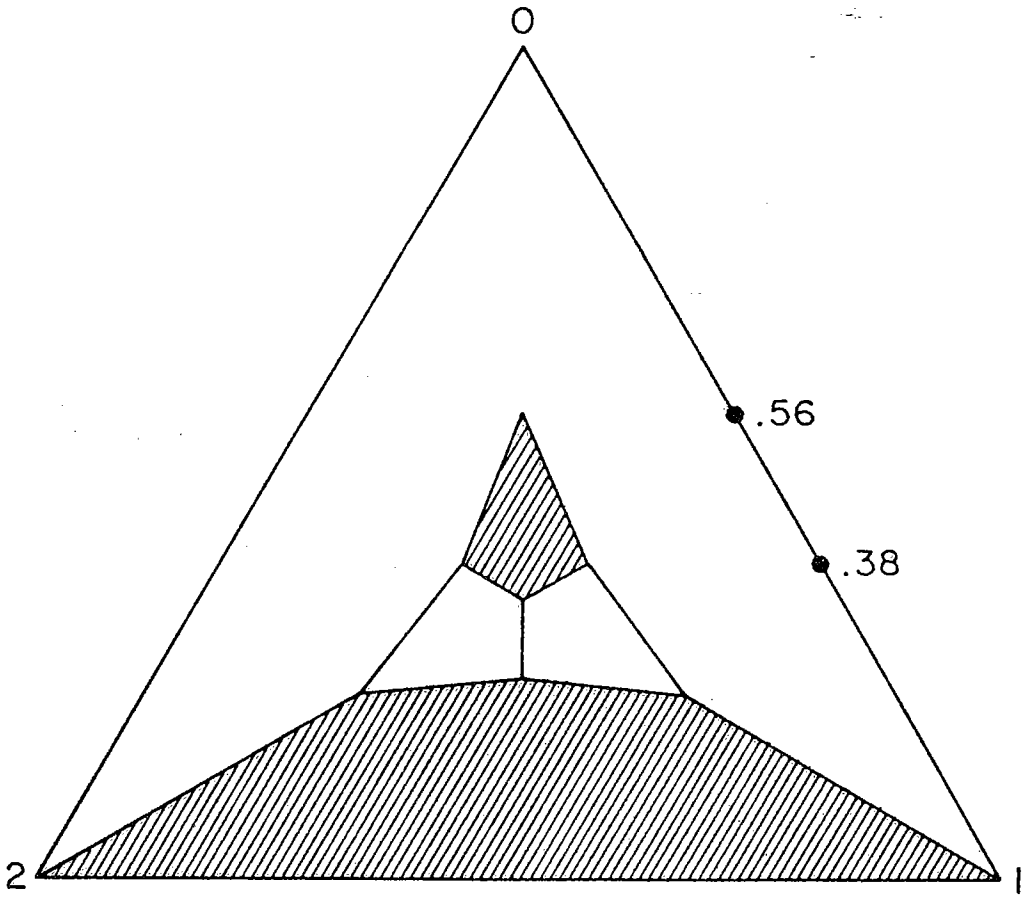

Figure 5. The Shaded Area Represents the Set N Corresponding to the Adaptive Estimator δ_a when $p_1 = 0.3$, $q_1 = 0.49$, $d = 0.37$.

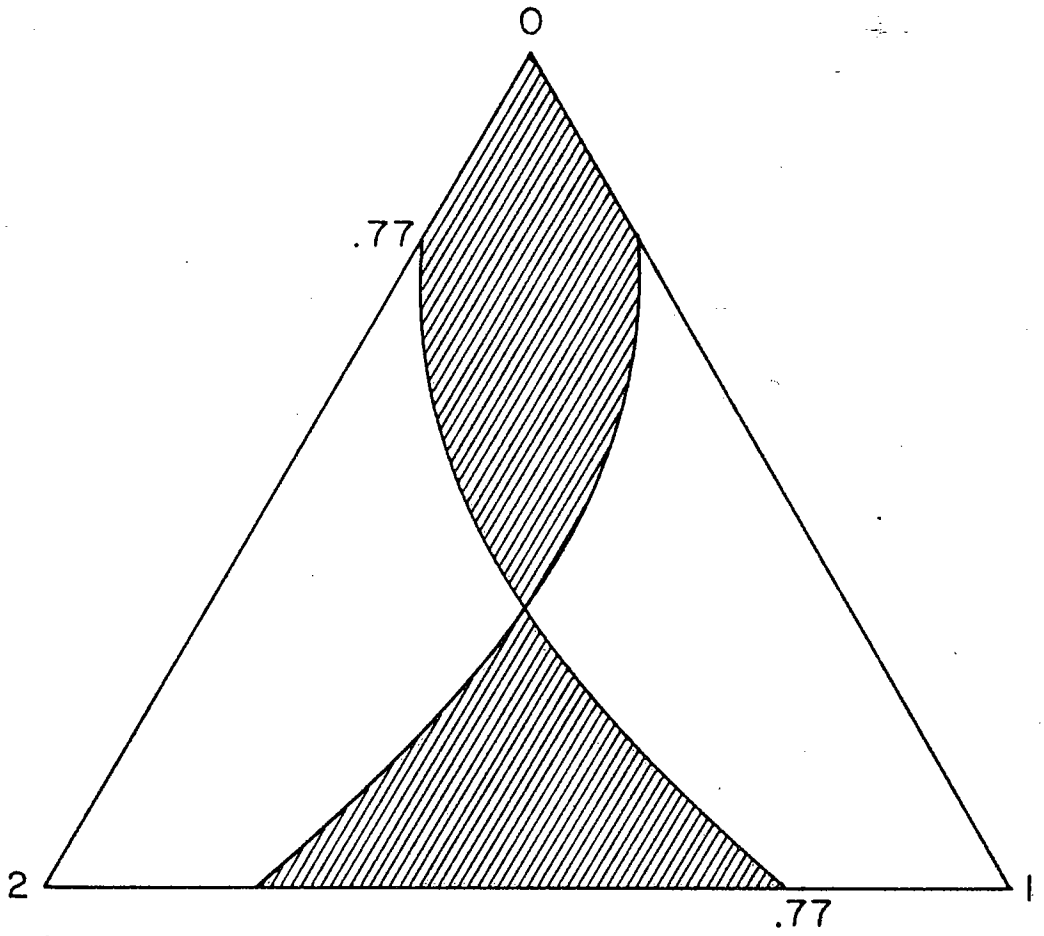

Figure 6. The Shaded Area Represents the Set N Corresponding to the Overall Maximum Likelihood Estimator $\hat{\delta}_0$.

REFERENCES

Bahadur, R. R. (1971). Some Limit Theorems in Statistics. Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia.

Groeneboom, P., Oosterhoff, J. and Ruymgaart, F. H. (1979). Large deviation theorems for empirical probability measures, Ann. Probability, 7, 553-586.

Krafft, O. and Puri, M. L. (1974). The asymptotic behavior of the minimax risk for multiple decision problems. Sankhyā, 36, 1-12.

Maksimov, V. M. (1975). Invariant estimates of the shift for distributions on finite groups. Theory Prob. Applications 20, 69-82.

Rukhin, A. L. (1975). Statistical inference about distributions on a cyclic group. (in Russian) Math. Operationsforsch. u. Statist. 6, 225-238.

Rukhin, A. L. (1977). On the estimation of a transformation parameter of a distribution given on a fintie group. Trans. Seventh Prague Conference Inform. Theory Statist. Dec. Functions Random Processes, Vol. B, 439-448.

Sinkov, A. (1966). Elementary Cryptanalysis. An Mathematical Approach. The Matheamtical Association of American, 22, Washington.

Steinebach, J. (1978). Convergence rates of large deviation probabilities in the multidimensional case. Ann. Probability 6, 751-759.

REFERENCES

Bahadur, R. R. (1971). Some Limit Theorems in Statistics. Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, Philadelphia.

Groeneboom, P., Oosterhoff, J. and Ruymgaart, F. H. (1979). Large deviation theorems for empirical probability measures, Ann. Probability, 7, 553-586.

Krafft, O. and Puri, M. L. (1974). The asymptotic behavior of the minimax risk for multiple decision problems. Sankhyā, 36, 1-12.

Maksimov, V. M. (1975). Invariant estimates of the shift for distributions on finite groups. Theory Prob. Applications 20, 69-82.

Rukhin, A. L. (1975). Statistical inference about distributions on a cyclic group. (in Russian) Math. Operationsforsch. u. Statist. 6, 225-238.

Rukhin, A. L. (1977). On the estimation of a transformation parameter of a distribution given on a fintie group. Trans. Seventh Prague Conference Inform. Theory Statist. Dec. Functions Random Processes, Vol. A, 439-448.

Steinebach, J. (1978). Convergence rates of large deviation probabilities in the multidimensional case. Ann. Probability 6, 751-759.