Journal of National Taiwan Normal University: Mathematics, Science & Technology
1999, 44(1), 1-9

The Maximum Likelihood Estimates of Multinomial Parameters Subject to Stochastic Order with Equality

Shao-Tung Chang
Department of Mathematics
National Taiwan Normal University

Hsiao-Chuan Lu
Department of Applied Mathematics
Chinese Culture University

Abstract

The problems of maximum likelihood estimates of multinomial parameters subject to stochastic ordering have been widely discussed. However, in some applications, the multinomial parameters **p** and **q** may not only satisfy the stochastically ordered constraints but also the equality of some of these parameters. We follow Barlow and Brunk's method (1972) and obtain the Fenchel duality projection-type maximum likelihood estimates of multinomial parameters under this modified hypothesis for both one-sample and two-sample problems. The consistency of the estimates is proved and an example is also presented as an illustration.

Keywords: Multinomial parameters, Maximum likelihood estimate, Stochastic ordering, Least square projection

Introduction

Let $\mathbf{p}=(p_1, p_2, ..., p_k)$ and $\mathbf{q}=(q_1, q_2, ..., q_k)$ be two collections of multinomial parameters and assume that $\forall i=1, 2, ..., k$, $p_i>0$, $q_i>0$, $\Sigma_{i=1}^k p_i=1$ and $\Sigma_{i=1}^k q_i=1$. The distribution associated with **p** is said to be stochastically smaller than the distribution associated with **q** if

$$\sum_{j=1}^{i} p_j \geqslant \sum_{j=1}^{i} q_j,\ i=1, 2, ..., k\text{-}1,$$

which we denote symbolically by $\mathbf{p} \gg \mathbf{q}$. The problems of maximum likelihood estimates of multinomial parameters subject to stochastic ordering have been widely discussed. Brunk et al. (1966) obtained the two-sample maximum likelihood estimates of stochastically ordered continuous distributions. Robertson and Wright (1974) extended Brunk et al.'s result and found two-sample maximum likelihood estimates of multivariate stochastically ordered distributions. Barlow and Brunk (1972) applied Fenchel duality and developed projection-type maximum likelihood estimates of multinomial distributions subject to the stochastically ordered constraints. They considered both one-sample and two-sample estimation problems. Sampson and Whitaker (1989) extended Barlow and Brunk's estimation to multivariate distributions under stochastic ordering. They considered both one-sample and two-sample estimation problems also. Dykstra (1982) considered the problem of finding maximum likelihood estimates of stochastically ordered survival functions for both one-sample and two-sample cases. Feltz and Dykstra (1985) found the maximum likelihood estimators of N survival functions that satisfy linear stochastic ordering. Dykstra and Feltz(1989) extended the above result and discussed estimation of survival functions under an arbitrary partial stochastic ordering.

However, in some applications, the distributions associated with **p** and **q** may not only satisfy the stochastically ordered constraints but also the equality of some of their parameters. For instance, drug A may have the same effect in relieving the headache as drug B in the first time period after taking. But the effect of drug A cannot compete with drug B after the first time period. Hence, to estimate the effect of drug A and B, it is required to add additional equality restrictions to the stochastic ordering. In this paper, we generalize this type of restrictions as follows:

$$\mathbf{H} : \mathbf{p} \gg \mathbf{q} \text{ and } p_j = q_j,\ \forall j = \eta_1, \eta_2, ..., \eta_r,$$

where $\eta_1, \eta_2, ..., \eta_r$ are the integers between 1 and k. In Section 2, we solve the maximum likelihood estimation problem under the modified hypothesis, **H**, by following Barlow and Brunk's method (1972). We discussed both one-sample and two-sample problems and obtain the Fenchel duality projection-type maximum likelihood estimates. In addition, the consistency of the estimates is discussed in Section 3. For illustration, we give a real example in Section 4.

Maximum likelihood estimates

We consider one-sample problem first. Assume **q** is known and a random sample of size m from the population associated with **p** is taken. Let $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, ..., \hat{p}_k)$ be the vector of relative frequencies; i.e., $m\hat{\mathbf{p}}$ has a multinomial distribution with parameters m and **p**. The following theorem supplies the maximum likelihood estimates of **p** subject to the restriction of **H**. For simplicity, define $\mathbf{A}^r = \{\eta_1, \eta_2, ..., \eta_r\}$ hereafter. Let $\mathbf{g} = (g_1, g_2, ..., g_k)$ be any fixed vector, $\mathbf{w} = (w_1, w_2, ..., w_k)$ be a positive weight vector, and **C** be a closed convex set in $\mathbf{R}^k$. The notation $\mathbf{P}_w(\mathbf{g}|\mathbf{C})$ represents the least squares projection of **g** onto the collection **C** with weight **w**; i.e., it is the solution to the following problem

$$\text{minimize } \sum_i^k [g_i - f_i]^2 w_i \text{ subject to } \mathbf{f} \in \mathbf{C},$$

where $\mathbf{f} = (f_1, f_2, ..., f_k)$. For x, y $\in R^k$, we also use $\mathbf{x \cdot y}$ and $\mathbf{x/y}$ to prepresent $(x_1y_1, x_2y_2, ..., x_ky_k)$ and $(x_1/y_1, x_2/y_2, ..., x_k/y_k)$ respectively.

Theorem 2.1 Suppose **q** is known and $\hat{p}_i > 0$, $\forall$ i= 1, 2, ..., k; k>1. Then, the maximum likelihood estimate (m.l.e.) of **p** subject to **H** is given by

$$\bar{\mathbf{p}} = \hat{\mathbf{p}} \cdot P_{\hat{p}}(\mathbf{q}/\hat{\mathbf{p}}|\mathbf{C}_{(r)}),$$

where

$$\mathbf{C}_{(r)} = \{(x_1, x_2, ..., x_k) \in \mathbf{R}^k : x_i \geq x_j, \text{ when } 1 \leq i < j \leq k \text{ and } i, j \not\in \mathbf{A}^r\}.$$

Proof. The m.l.e. of **p** is $\bar{\mathbf{p}}$ which is the vector that maximize $\sum_{i=1}^k (\hat{p}_i \ln p_i)$, subject to **H**

$\Longleftrightarrow$ The vector that minimize $-\sum_{i=1}^k (\hat{p}_i \ln p_i)$, subject to **H**

$\Longleftrightarrow$ The vector that

$$\text{minimize } -\sum_{i=1}^{k-r} (\hat{p}_{u_i} \ln p_{u_i}), + \Big[-\sum_{i \in A^r} (\hat{p}_i \ln p_i)\Big],$$

subject to **H**, where $1 \leq u_1 \leq u_2 \leq ... \leq u_{k-r} \leq k$ and $u_1, u_2, ..., u_{k-r} \not\in \mathbf{A}^r$. Let $\bar{p}^r = (\bar{p}_{u_1}, \bar{p}_{u_2}, ..., \bar{p}_{u_{k-r}})$.

Under **H**,

$$\Big[-\sum_{i \in A^r} (\hat{p}_i \ln p_i)\Big] = \Big[-\sum_{i \in A^r} (\hat{p}_i \ln q_i)\Big]$$

is known. The above minimization problem is equivalent to

$$\text{minimize } -\sum_{i=1}^{k-r} (\hat{p}_{u_i} \ln p_{u_i}).$$

This is equivalent to

$$\text{minimize } -\sum_{i=1}^{k-r} \Big(\hat{p}_{u_i} \ln \frac{p_{u_i}}{\hat{p}_{u_i}}\Big). \tag{1}$$

we define

$$\mathbf{w}^r = (w_1^r, w_2^r, ..., w_{k-r}^r) = (\hat{p}_{u_1}, \hat{p}_{u_2}, ..., \hat{p}_{u_{k-r}}) \hat{\mathbf{p}}^r,$$

$$\mathbf{q}^r = (q_{u_1}, q_{u_2}, ..., q_{u_r}),$$

$$\mathbf{p}^r = (p_{u_1}, p_{u_2}, ..., p_{u_r}),$$

$$\mathbf{s}^r = (s_1^r, s_2^r, ..., s_{k-r}^r) = \Big(\frac{p_{u_1}}{\hat{p}_{u_1}}, \frac{p_{u_2}}{\hat{p}_{u_2}}, ..., \frac{p_{u_{k-r}}}{\hat{p}_{u_{k-r}}}\Big) = \frac{\mathbf{p}^r}{\hat{\mathbf{p}}^r},$$

$$\mathbf{g}^r = (g_1^r, g_2^r, ..., g_{k-r}^r) = \Big(\frac{q_{u_1}}{\hat{p}_{u_1}}, \frac{q_{u_2}}{\hat{p}_{u_2}}, ..., \frac{q_{u_{k-r}}}{\hat{p}_{u_{k-r}}}\Big) = \frac{\mathbf{q}^r}{\hat{\mathbf{p}}^r},$$

and $\Phi(y) = -\ln(y)$. Thus, we can rewrite (1) as

$$\text{minimize } \sum_{i=1}^{k-r} (w_i^r \Phi(s_i^r)) \tag{2}$$

and the corresponding restriction **H** as

$$\sum_{i=1}^{j} w_i^r (g_i^r - s_i^r) \leq 0,\ 1 \leq j < k\text{-}r, \text{ and} \sum_{i=1}^{k-r} w_i^r (g_i^r - s_i^r) = 0.$$

The Fenchel dual, $\mathbf{C}^{r*w}$, of

$$\mathbf{C}^r = \{(x_{u_1}, x_{u_2}, ..., x_{u_{k-r}}) \in \mathbf{R}^{k-r} : x_{u_i} \geq x_{u_j}, \text{ when } 1 \leq i < j \leq k\text{-}r\}$$

is

$$\mathbf{C}^{r*w} = \{u^r; (u^r, v^r)_w^r \leq 0,\ \forall\ v^r \in C^r\}$$
$$= \left\{u^r;\ \sum_{i=1}^{j} u_i^r w_i^r \leq 0,\ \forall\ 1 \leq j < k-r,\ \sum_{i=1}^{k-r} u_i^r w_i^r = 0\right\}.$$

(cf. Robertson et al., 1988, page 50). Thus (2) becomes

minimize $\sum_{i=1}^{k-r} (w_i \Phi(s_i^r))$ subject to $\mathbf{g}^r - \mathbf{s}^r \in \mathbf{C}^{r*w}$.

Refer to the corollary on page 48 of Robertson et al., 1988, we have the m.l.e. of $\mathbf{s}^r$ and $\mathbf{p}^r$ as

$$\bar{\mathbf{s}}^r = P_{\hat{p}^r}(\mathbf{g}^r \mid \mathbf{C}^r),$$
$$\bar{\mathbf{p}}^r = \hat{\mathbf{p}}^r \cdot \bar{\mathbf{s}}^r = \hat{\mathbf{p}}^r \cdot \mathbf{P}_{\hat{p}^r}(\mathbf{q}^r/\hat{\mathbf{p}}^r \mid \mathbf{C}^r),$$

and

$$\bar{p}_j = \bar{q}_j,\ j \in \mathbf{A}^r.$$

So,

$$\begin{cases} \bar{p}_{u_i} = \hat{p}_{u_i} \cdot \mathbf{P}_{\hat{p}^r}(\mathbf{q}^r/\hat{\mathbf{p}}^r \mid \mathbf{C}^r)_i, & \text{where } i = 1, \dots, k-r \\ \bar{p}_i = q_i & i \in \mathbf{A}^r \end{cases} \quad (3)$$

The remaining work is to show that (3) is the same as $\bar{\mathbf{p}} = \hat{\mathbf{p}} \cdot \mathbf{P}_{\hat{p}^r}(\mathbf{q}/\hat{\mathbf{p}} \mid \mathbf{C}_{(r)}))$. We see that when $\mathbf{q}/\hat{\mathbf{p}}$ project onto $\mathbf{C}_{(r)}$, the η_1th, η_2th, ..., η_rth terms will be singletons (i.e., these terms will not amalgamate with any other terms through the least square projection). Refer to Theorem 1.4.2, Theorem 1.4.4, and Theorem 1.4.5 of Robertson et al., 1988, we have

$$\bar{p}_{\eta_i} = \hat{p}_{\eta_i} \cdot q_{\eta_i}/\hat{p}_{\eta_i} = q_{\eta_i},\ \forall\ i = 1, \dots, r.$$

and the other terms are derived by multiplying $\hat{\mathbf{p}}^r$ with the projection of $\mathbf{q}^r/\hat{\mathbf{p}}^r$ onto $\mathbf{C}^r$. Thus (3) is the same as

$$\bar{\mathbf{p}} = \hat{\mathbf{p}} \cdot \mathbf{P}_{\hat{p}}(\mathbf{q}/\hat{\mathbf{p}} \mid \mathbf{C}_{(r)})).$$

This completes the proof.

In two-sample problem, assume both $\mathbf{p}$ and $\mathbf{q}$ are unknown. In addition to $\hat{\mathbf{p}}$, let $\hat{\mathbf{q}}$ denote the vector of relative frequencies of a sample of size n from the $\mathbf{q}$ population and assume that $\hat{\mathbf{p}}$ and $\hat{\mathbf{q}}$ are independent. Let $S = m + n$ and we denote the m.l.e. of $(\mathbf{p}, \mathbf{q})$ under $\mathbf{H}$ as $(\bar{\mathbf{p}}, \bar{\mathbf{q}})$. To derive the m.l.e. of $\mathbf{p}$ and $\mathbf{q}$ under $\mathbf{H}$, we need the following five lemmas.

Lemma 1 Suppose that $\hat{p}_i, \hat{q}_i > 0,\ \forall\ i = 1, \dots, k$ and $\sum_{i=1}^{k} p_i = \sum_{i=1}^{k} q_i = 1$. If there exist $\eta_1, \eta_2, \dots, \eta_r$ such that

$$p_i = q_i,\ \forall\ i = \eta_1, \eta_2, \dots, \eta_r, \quad (4)$$

where $r \in N$, r is given, $1 \leq r \leq k-2$, and $\eta_1, \eta_2, \dots, \eta_r$ are integers between 1 and k. If $(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$ maximize the likelihood function subject to the above restriction (4), then

$$\tilde{p}_i = \tilde{q}_i = \frac{m\hat{p}_i + n\hat{q}_i}{S};\ \forall\ i = \eta_1, \eta_2, \dots, \eta_r.$$

Proof: Recall that $\mathbf{A}^r = \{\eta_1, \eta_2, \dots, \eta_r\}$. If $\mathbf{A}^r = \{1, 2, \dots, k\}$, the result holds trivially. Hence, assume $\mathbf{A}^r \neq \{\eta_1, \eta_2, \dots, \eta_r\}$ and let u_{k-r} be the largest index in $\{1, 2, \dots, k\} \setminus \mathbf{A}^r$, so the log-likelihood function

$$\mathbf{L}(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{k} (m\hat{p}_i \ln p_i + n\hat{q}_i \ln q_i)$$
$$= \sum_{i \notin \mathbf{A}^r \cup \{u_{k-r}\}} (m\hat{p}_i \ln p_i + n\hat{q}_i \ln q_i) + \sum_{i \in \mathbf{A}^r} (m\hat{p}_i + n\hat{q}_i) \ln p_i + (m\hat{p}_{u_{k-r}} \ln \hat{p}_{u_{k-r}} + n\hat{q}_{u_{k-r}} \ln \hat{q}_{u_{k-r}}.$$

So, $\mathbf{L}$ attains its maximum iff

$$\begin{cases} \dfrac{\partial L}{\partial p_i} = 0 & i \notin \mathbf{A}^r \cup \{u_{k-r}\} \\ \dfrac{\partial L}{\partial q_i} = 0 & i \notin \mathbf{A}^r \cup \{u_{k-r}\} \\ \dfrac{\partial L}{\partial p_i} = 0 & i \in \mathbf{A}^r. \end{cases}$$

Since $(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$ maximize the likelihood function, $(\tilde{\mathbf{p}}, \tilde{\mathbf{q}})$ satisfies the following equivalent equations

$$\begin{cases} \dfrac{m\hat{p}_i}{\tilde{p}_i} = \dfrac{m\hat{p}_{u_{k-r}}}{\tilde{p}_{u_{k-r}}} & i \notin \mathbf{A}^r \\ \dfrac{m\hat{q}_i}{\tilde{q}_i} = \dfrac{n\hat{q}_{u_{k-r}}}{\tilde{q}_{u_{k-r}}} & i \notin \mathbf{A}^r \\ \dfrac{m\hat{p}_i + n\hat{q}_i}{\tilde{p}_i} = \dfrac{m\hat{p}_{u_{k-r}}}{\tilde{p}_{u_{k-r}}} + \dfrac{n\hat{q}_{u_{k-r}}}{\tilde{q}_{u_{k-r}}} & i \in \mathbf{A}^r. \end{cases}$$

This implies

$$\frac{\sum_{i \notin A^r} m\hat{p}_i}{\sum_{i \notin A^r} \tilde{p}_i} = \frac{m\hat{p}_{u_{k-r}}}{\tilde{p}_{u_{k-r}}},$$

$$\frac{\sum_{i \notin A^r} n\hat{q}_i}{\sum_{i \notin A^r} \tilde{q}_i} = \frac{n\hat{q}_{u_{k-r}}}{\tilde{q}_{u_{k-r}}},$$

and

$$\frac{\sum_{i \notin A^r} m\hat{p}_i + n\hat{q}_i}{\sum_{i \notin A^r} \tilde{p}_i} = \frac{m\hat{p}_{u_{k-r}}}{\tilde{p}_{u_{k-r}}} + \frac{n\hat{q}_{u_{k-r}}}{\tilde{q}_{u_{k-r}}}.$$

Thus, we have

$$\frac{\sum_{i \notin A^r} m\hat{p}_i + n\hat{q}_i}{\sum_{i \notin A^r} \tilde{p}_i} = \frac{\sum_{i \notin A^r} m\hat{p}_i}{\sum_{i \notin A^r} \tilde{p}_i} + \frac{\sum_{i \notin A^r} n\hat{q}_i}{\sum_{i \notin A^r} \tilde{q}_i}.$$

Since $\tilde{p}_i = \tilde{q}_i,\ \forall\ i \in \mathbf{A}^r$ and $\sum_{i=1}^{k} \tilde{p}_i = \sum_{i=1}^{k} \tilde{q}_i$, $\sum_{i \notin A^r} \tilde{p}_i = \sum_{i \notin A^r} \tilde{q}_i$. Hence, the above equation can be rewritten as

$$\frac{\sum_{i \in A^r} m\hat{p}_i + n\hat{q}_i}{\sum_{i \in A^r} \tilde{p}_i} = \frac{\sum_{i \notin A^r} m\hat{p}_i + n\hat{q}_i}{\sum_{i \notin A^r} \tilde{p}_i}$$
$$= \frac{\sum_{i \in A^r} m\hat{p}_i + n\hat{q}_i + \sum_{i \notin A^r} m\hat{p}_i + n\hat{q}_i}{\sum_{i \in A^r} \tilde{p}_i + \sum_{i \notin A^r} \tilde{p}_i}$$
$$= S.$$

So,

$$\sum_{i\in A^r}\tilde{p}_i = \frac{\sum_{i\in A^r} m\hat{p}_i + n\hat{q}_i}{S}.$$

On the other hand,

$$\frac{m\hat{p}_i + n\hat{q}_i}{\tilde{p}_i} = \frac{m\hat{p}_{u_{k-r}}}{\tilde{p}_{u_{k-r}}} + \frac{n\hat{q}_{u_{k-r}}}{\tilde{q}_{u_{k-r}}}\ \forall\ i\in \mathbf{A}^r.$$

Thus,

$$\frac{\tilde{p}_i}{\tilde{p}_j} = \frac{m\hat{p}_i + n\hat{q}_i}{m\hat{p}_j + n\hat{q}_j},\ \forall\ i, j\in \mathbf{A}^r.$$

So. $\tilde{p}_i = \frac{m\hat{p}_i + n\hat{q}_i}{S}$, $\forall\ i\in \mathbf{A}^r$. This completes the proof.

Lemma 2 Suppose that $\hat{p}_i, \hat{q}_i > 0$, $\forall\ i=1,2,...,k$ and that both $\mathbf{p}$ and $\mathbf{q}$ are unknown. Let $(\bar{\mathbf{p}},\bar{\mathbf{q}})$ be the maximum likelihood estimator of $(\mathbf{p},\mathbf{q})$ under the condition of $\mathbf{H}$ and $(\tilde{\mathbf{p}},\tilde{\mathbf{q}})$ be the m.l.e. of $(\mathbf{p},\mathbf{q})$ under the condition that $p_i = q_i$, $\forall\ i\in \mathbf{A}^r$ respectively. If $\sum^j \tilde{p}_i > \sum^j \tilde{q}_i$, $\forall\ 1\leqslant j<k$, the $(\bar{\mathbf{p}},\bar{\mathbf{q}}) = (\tilde{\mathbf{p}},\tilde{\mathbf{q}})$.

Proof: Suppose that $(\bar{\mathbf{p}},\bar{\mathbf{q}}) \neq (\tilde{\mathbf{p}},\tilde{\mathbf{q}})$. Consider the function h, defined on [0,1] by $h(\lambda) = \mathbf{L}(\lambda(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) + (1-\lambda)(\bar{\mathbf{p}},\bar{\mathbf{q}}))$. Since h is concave on [0,1],

$$(h''(\lambda) = \sum_{i=1}^{k} \frac{-m\hat{p}_i(\tilde{p}_i - \bar{p}_i)^2}{[\lambda\tilde{p}_i + (1-\lambda)\bar{p}_i]^2} + \frac{-n\hat{p}_i(\tilde{q}_i - \bar{q}_i)^2}{[\lambda\tilde{q}_i + (1-\lambda)\bar{q}_i]^2} < 0)$$

and $(\bar{\mathbf{p}},\bar{\mathbf{q}}) \neq (\tilde{\mathbf{p}},\tilde{\mathbf{q}})$, $\mathbf{L}$ is increasing along the direction from $(\bar{\mathbf{p}},\bar{\mathbf{q}})$ to $(\tilde{\mathbf{p}},\tilde{\mathbf{q}})$. On the other hand, $(\tilde{\mathbf{p}},\tilde{\mathbf{q}})$ is the m.l.e. of $(\mathbf{p},\mathbf{q})$ subject to less constraints, so $h(1) = \mathbf{L}(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) \geqslant \mathbf{L}(\bar{\mathbf{p}},\bar{\mathbf{q}}) = h(0)$. Because, h increase strictly at $\lambda = 0$, there exists a $\lambda_0 \in (0,1)$ such that $h(\lambda_0) > h(0) = \mathbf{L}(\bar{\mathbf{p}},\bar{\mathbf{q}})$ and

$$\sum_{i=1}^{j}(\lambda_0\tilde{p}_i + (1-\lambda_0)\bar{p}_i) \geqslant \sum_{i=1}^{j}(\lambda_0\tilde{q}_i + (1-\lambda_0)\bar{q}_i),$$

for $j=1,...,k-1$. Furthermore, $\lambda_0\tilde{p}_i + (1-\lambda_0)\bar{p}_i = \lambda_0\tilde{q}_i + (1-\lambda_0)\bar{q}_i$, $\forall\ i\in \mathbf{A}^r$. Hence, $\lambda_0(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) + (1-\lambda_0)(\bar{\mathbf{p}},\bar{\mathbf{q}})$ satisfies $\mathbf{H}$ condition and

$$\mathbf{L}(\lambda_0(\tilde{\mathbf{p}},\tilde{\mathbf{q}}) + (1-\lambda_0)(\bar{\mathbf{p}},\bar{\mathbf{q}})) > \mathbf{L}(\bar{\mathbf{p}},\bar{\mathbf{q}}).$$

This result contradicts with the assumption that $(\bar{\mathbf{p}},\bar{\mathbf{q}})$ is the m.l.e. of $(\mathbf{p},\mathbf{q})$ under condition $\mathbf{H}$, so $(\bar{\mathbf{p}},\bar{\mathbf{q}}) = (\tilde{\mathbf{p}},\tilde{\mathbf{q}})$. This completes the proof.

The following lemma is an extension of Lemma 1, but it is proved in a different way.

Lemma 3 Suppose that $\hat{p}_i, \hat{q}_i > 0$, $\forall\ i=1,...,k$ and $\sum_{i=1}^{k} p_i = \sum_{i=1}^{k} q_i = 1$. There exist $\mathbf{A}^r = \{\eta_1, \eta_2, ..., \eta_r\}$ and $\mathbf{A}^s = \{j_1, j_2, ..., j_s\}$, $j_1 < ... < j_s$ such that

$$p_i = q_i,\ \forall\ i\in \mathbf{A}^r \text{ and } \sum_{i=1}^{j_t} p_i = \sum_{i=1}^{j_t} q_i,\ \forall\ j_t\in \mathbf{A}^s. \quad (5)$$

Note that $j_s = k$. If $(\tilde{\mathbf{p}},\tilde{\mathbf{q}})$ maximize the likelihood function subject to the above restriction (5), then $\tilde{p}_i = \tilde{q}_i = \frac{m\hat{p}_i + n\hat{q}_i}{S}$, $\forall i\in \mathbf{A}^r$ and $\sum_{i=1}^{j_t}\tilde{p}_i = \sum_{i=1}^{j_t}\tilde{q}_i$, $\forall i\in \mathbf{A}^s$.

Proof: Let $\mathbf{B}_1 = \{1,...,j_1\}$, $\mathbf{B}_2 = \{j_1+1,...,j_2\}$,...,$\mathbf{B}_s = \{j_{s-1}+1,...,j_s\}$. We see that

$$\sum_{i=1}^{j_t} p_i = \sum_{i=1}^{j_t} q_i,\ \forall\ j_t\in \mathbf{A}^s$$

is equivalent to

$$\sum_{i\in B_t} p_i\ \ \sum_{i\in B_t} q_i,\ \forall\ t=1,2,...,s.$$

We prove the result by using the Lagrange Multipliers. Let

$$\begin{aligned} F(\mathbf{p},\mathbf{q},\lambda,\boldsymbol{\delta},\boldsymbol{\beta}) = {} & \sum_{i=1}^{k} m\hat{p}_i \ln p_i + \sum_{i=1}^{k} n\hat{q}_i \ln q_i + \lambda(\sum_{i=1}^{k} p_i - 1) \\ & + \delta_{\eta_1}(p_{\eta_1} - q_{\eta_1}) + \cdots + \delta_{\eta_r}(p_{\eta_r} - q_{\eta_r}) \\ & + \beta_1(\sum_{i\in B_1} p_i - \sum_{i\in B_1} q_i) + \cdots + \beta_s(\sum_{i\in B_s} p_i - \sum_{i\in B_s} q_i). \end{aligned}$$

Taking the partial derivative of F over $\mathbf{p}$, $\mathbf{q}$, λ, $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ respectively and setting them equal to zero, we have

$$\begin{cases} \begin{cases} \frac{\partial F}{\partial p_i} = \frac{m\hat{p}_i}{p_i} + \lambda + \delta_i + \beta_t = 0 \\ \frac{\partial F}{\partial q_i} = \frac{n\hat{q}_i}{q_i} - \delta_i - \beta_t = 0 \end{cases} & \text{if } i\in \mathbf{B}_t \cap \mathbf{A}^r \\ \begin{cases} \frac{\partial F}{\partial p_i} = \frac{m\hat{p}_i}{p_i} + \lambda + \beta_t = 0 \\ \frac{\partial F}{\partial q_i} = \frac{n\hat{q}_i}{q_i} - \beta_t = 0 \end{cases} & \text{if } i\in \mathbf{B}_t \setminus \mathbf{A}^r \\ \sum_{i=1}^{k} p_i - 1 = 0 & \\ p_i - q_i = 0, & \forall\ i\in \mathbf{A}^r \\ \sum_{i\in B_t} p_i - \sum_{i\in B_t} q_i = 0, & \forall\ i \leqslant t \leqslant s. \end{cases}$$

If $(\tilde{\mathbf{p}},\tilde{\mathbf{q}})$ solve the above equations, after simplification, we have

$$\frac{m\sum_{i\in A^r}\hat{p}_i + n\sum_{i\in A^r}\hat{q}_i}{\sum_{i\in A^r}\tilde{p}_i} = -\lambda,$$

$$\frac{m\sum_{i\in B_t\setminus A^r}\hat{p}_i}{\sum_{i\in B_t\setminus A^r}\tilde{p}_i} = -\lambda - \beta_t,\ \forall\ 1\leqslant t\leqslant s,$$

and

$$\frac{m\sum_{i\in B_t\setminus A^r}\hat{q}_i}{\sum_{i\in B_t\setminus A^r}\tilde{q}_i} = -\beta_t,\quad \forall\ 1\leqslant t\leqslant s.$$

Since $\sum_{i=1}^{j_t} \tilde{p}_i = \sum_{i=1}^{j_t} \tilde{q}_i$, $\sum_{i\in B_t\setminus A^r} \tilde{p}_i = \sum_{i\in B_t\setminus A^r} \tilde{q}_i$. We have $-\lambda = m+n = S$.

Moreover

$$\tilde{p}_i = \tilde{q}_i = \frac{m\hat{p}_i + n\hat{q}_i}{S},\ \forall\ i \in \mathbf{A}^r$$

and

$$\sum_{i\in B_t\setminus A^r} \tilde{p}_i = \frac{m \sum_{i\in B_t\setminus A^r} \hat{p}_i + n \sum_{i\in B_t\setminus A^r} \hat{q}_i}{S} = \sum_{i\in B_t\setminus A^r} \tilde{q}_i,\ \forall\ 1 \leqslant t \leqslant s.$$

Hence

$$\sum_{i\in B_t} \tilde{p}_i = \sum_{i\in B_t} \tilde{q}_i,\ \forall\ j_t \in \mathbf{A}^s$$

and

$$\sum_{i=1}^{j_t} \tilde{p}_i = \sum_{i=1}^{j_t} \tilde{q}_i,\ \forall\ j_t \in \mathbf{A}^s.$$

This completes the proof.

In Lemma 4, $\mathbf{A}^r$ and $\mathbf{A}^s$ are defined the same way as in Lemma 3.

Lemma 4 Suppose that $\hat{p}_i, \hat{q}_i > 0,\ \forall\ i=1,2,\ldots,k$ and that both **p** and **q** are unknown. Let $(\bar{\mathbf{p}},\bar{\mathbf{q}})$ be the maximum likelihoodestimator of $(\mathbf{p},\mathbf{q})$ under the condition of **H** and $(\tilde{\mathbf{p}},\tilde{\mathbf{q}})$ be the m.l.e. of $(\mathbf{p},\mathbf{q})$ under the condition that

$$p_i = q_i,\ \forall\ i \in \mathbf{A}^r \text{ and } \sum_{i=1}^{j_t} p_i = \sum_{i=1}^{j_t} q_i,\ \forall\ j_t \in \mathbf{A}^s.$$

If $\sum_i^j \bar{p}_i > \sum_i^j \bar{q}_i,\ \forall\ j \notin \mathbf{A}^s$, then $(\bar{\mathbf{p}},\bar{\mathbf{q}}) = (\tilde{\mathbf{p}},\tilde{\mathbf{q}})$.

Proof: The proof is similar to the proof of Lemma 2.

Lemma 5 Suppose that $\hat{p}_i, \hat{q}_i > 0,\ \forall\ i=1,2,\ldots,k$ and that both **p and q** are unknown. Let $(\bar{\mathbf{p}},\bar{\mathbf{q}})$, be the maximum likelihood estimator of $(\mathbf{p},\mathbf{q})$ under the condition of **H**. Then,

$$\sum_{i=1}^{j} \bar{p}_i \geqslant \frac{\sum_{i=1}^{j} m\hat{p}_i + n\hat{q}_i}{S} \geqslant \sum_{i=1}^{j} \bar{q}_i,\ \forall\ j=1,2,\ldots,k,$$

and $\bar{p}_i = \bar{q}_i = \dfrac{m\hat{p}_i + n\hat{q}_i}{S},\ \forall\ i = \eta_1, \eta_2, \ldots, \eta_r$.

Proof: Depending on the number of inequalities,

$$\sum_{i=1}^{j} \bar{p}_i \geqslant \sum_{i=1}^{j} \bar{q}_i$$

that hold strictly , there are two possible cases as follows.

(i) $\sum_i^j \bar{p}_i > \sum_i^j \bar{q}_i,\ \forall\ 1 \leqslant j < k$.

(ii) There exists $W \subset \{1,\ldots,k-1\}$ such that $\sum_i^j \bar{p}_i > \sum_i^j \bar{q}_i$, $\forall\ j \in W$ and $\sum_i^j \bar{p}_i = \sum_i^j \bar{q}_i,\ \forall\ j \in \{1,2,\ldots,k\}\setminus W$.

In case (i), by Lemma 1 and Lemma 2,

$$\bar{p}_i = \bar{q}_i = \frac{m\hat{p}_i + n\hat{q}_i}{S},\ \forall\ i \in \mathbf{A}^r. \qquad (6)$$

In case (ii), by Lemma 3 and Lemma 4, equation (6) holds too.

Let $\mathbf{L}_0 = \sum_{i\notin A^r} (m\hat{p}_i \ln p_i + n\hat{q}_i \ln q_i)$, then the log-likelihood

$$\mathbf{L}(\mathbf{p},\mathbf{q}) = \mathbf{L}_0 + \sum_{i\in A^r} (m\hat{p}_i \ln p_i + n\hat{q}_i \ln q_i).$$

Thus, $(\bar{\mathbf{p}}^r,\bar{\mathbf{q}}^r)$ (Here, $\bar{\mathbf{p}}^r = (\bar{p}_{u_1},\ldots,\bar{p}_{u_{k-r}})$ and $\bar{\mathbf{q}}^r = (\bar{q}_{u_1},\ldots,\bar{q}_{u_{k-r}})$ must maximize $\mathbf{L}_0$ under the condition

$$\mathbf{H}_0 : \sum_{i=u_1}^{u_j} p_i \geqslant \sum_{i=u_1}^{u_j} q_i, \text{ for } j=1,\ldots,k\text{-}r.$$

Hence, equivalently, $(\bar{\mathbf{p}}^r,\bar{\mathbf{q}}^r)$ maximize

$$\sum_{i\notin A^r} \left\{ m\left(\sum_{i\notin A^r} \hat{p}_i\right) \cdot \frac{\hat{p}_i}{\sum_{i\notin A^r} \hat{p}_i} \cdot \ln\left(\frac{p_i}{\sum_{i\notin A^r} p_i}\right) + n\left(\sum_{i\notin A^r} \hat{q}_i\right) \cdot \frac{\hat{q}_i}{\sum_{i\notin A^r} \hat{q}_i} \cdot \ln\left(\frac{q_i}{\sum_{i\notin A^r} q_i}\right) \right\}$$

under the condition that

$$\sum_{i=u_1}^{u_j} \frac{p_i}{\left(\sum_{i=u_1}^{u_{k-r}} p_i\right)} \geqslant \sum_{i=u_1}^{u_j} \frac{q_i}{\left(\sum_{i=u_1}^{u_{k-r}} q_i\right)},\ j=1,\ldots,k\text{-}r\text{-}1,$$

and

$$\sum_{i=u_1}^{u_{k-r}} \frac{p_i}{\left(\sum_{i=u_1}^{u_{k-r}} p_i\right)} = \sum_{i=u_1}^{u_{k-r}} \frac{q_i}{\left(\sum_{i=u_1}^{u_{k-r}} q_i\right)} = 1.$$

Then, by lemma on page 251 of Robertson et al.1988,

$$\sum_{i=u_1}^{u_j} \frac{\bar{p}_i}{\sum_{i=u_1}^{u_{k-r}} \bar{p}_i} \geqslant \sum_{i=u_1}^{u_j} \frac{m\left(\sum_{i=u_1}^{u_{k-r}} \hat{p}_i\right) \cdot \frac{\hat{p}_i}{\left(\sum_{i=u_1}^{u_{k-r}} \hat{p}_i\right)} + n\left(\sum_{i=u_1}^{u_{k-r}} \hat{q}_i\right) \cdot \frac{\hat{q}_i}{\left(\sum_{i=u_1}^{u_{k-r}} \hat{q}_i\right)}}{m\sum_{i=u_1}^{u_{k-r}} \hat{p}_i + n\sum_{i=u_1}^{u_{k-r}} \hat{q}_i} \geqslant \sum_{i=u_1}^{u_j} \frac{\bar{q}_i}{\sum_{i=u_1}^{u_{k-r}} \bar{q}_i}$$

for $j=1,\ldots,k$-r. On the other hand, it is easy to see that

$$\sum_{i=u_1}^{u_{k-r}} \bar{p}_i = \sum_{i=u_1}^{u_{k-r}} \bar{q}_i = \sum_{i=u_1}^{u_{k-r}} \frac{m\hat{p}_i + n\hat{q}_i}{S}.$$

Hence,

$$\sum_{i=u_1}^{u_j} \bar{p}_i \geq \frac{\sum_{i=u_1}^{u_j} m\hat{p}_i + n\hat{q}_i}{S} \geq \sum_{i=u_1}^{u_j} \bar{q}_i, \text{ for } j=1,2,\dots,k\text{-}r. \quad (7)$$

Combining the results of (6) and (7), we have

$$\sum_{i=1}^{j} \bar{p}_i \geq \frac{\sum_{i=1}^{j} m\hat{p}_i + n\hat{q}_i}{S} \geq \sum_{i=1}^{j} \bar{q}_i, \text{ for } j=1,2,\dots,k.$$

This completes the proof.

Theorem 2.2 If $\hat{p}_i, \hat{q}_i > 0$, $i=1,2,\dots,k$ and $k>1$, then the maximum likelihood estimate of $(\mathbf{p},\mathbf{q})$ under $\mathbf{H}$ is

$$(\bar{\mathbf{p}},\bar{\mathbf{q}}) = \mathbf{w}\mathbf{P}_w(\mathbf{h}|\mathbf{B}_{(r)})$$

where

$$\mathbf{w} = (m\hat{p}_1,\dots,m\hat{p}_k,n\hat{q}_1,\dots,n\hat{q}_k)$$

$$\mathbf{h} = \left(\frac{m\hat{p}_1+n\hat{q}_1}{Sm\hat{p}_1},\dots,\frac{m\hat{p}_k+n\hat{q}_k}{Sm\hat{p}_k},\frac{m\hat{p}_1+n\hat{q}_1}{Sn\hat{q}_1},\dots,\frac{m\hat{p}_k+n\hat{q}_k}{Sn\hat{q}_k}\right)$$

and

$$\mathbf{B}_{(r)} = \{(x_1,x_2,\dots,x_{2k},) \in \mathbf{R}^{2k}: x_i \geq x_j, \text{ when } 1 \leq i<j \leq k \text{ and } i,j \notin \mathbf{A}^r;\ x_i \leq x_j, \text{ when } k+1 \leq i<j \leq 2k \text{ and } i,j \notin \mathbf{A}^{2r}\}.$$

Proof: By Lemma 5, we have $\bar{p}_i = \bar{q}_i \frac{m\hat{p}_i + n\hat{q}_i}{S}$, $\forall i \in \mathbf{A}^r$. Hence, our maximum likelihood estimation problem become maximizing

$$\mathbf{L}(\mathbf{p},\mathbf{q}) = \sum_{i\notin A^r} (m\hat{p}_i \ln p_i + n\hat{q}_i \ln q_i) + \sum_{i\in A^r} (m\hat{p}_i \ln p_i + n\hat{q}_i \ln q_i)$$

under $\mathbf{H}$ condition and $\sum_{i\in A^r} (m\hat{p}_i \ln p_i + n\hat{q}_i \ln q_i)$ is a given number. Let

$$\mathbf{t}^{2r} = \left(\frac{p_{u_1}}{m\hat{p}_{u_1}},\dots,\frac{p_{u_{k-r}}}{m\hat{p}_{u_{k-r}}},\frac{q_{u_1}}{n\hat{p}_{u_1}},\dots,\frac{q_{u_{k-r}}}{n\hat{q}_{u_{k-r}}}\right)$$

$$\mathbf{w}^{2r} = (m\hat{p}_{u_1},\dots, m\hat{p}_{u_{k-r}}, n\hat{q}_{u_1},\dots, n\hat{q}_{u_{k-r}})$$

$$\mathbf{h}^{2r} = \left(\frac{m\hat{p}_{u_1}+n\hat{q}_{u_1}}{Sm\hat{p}_{u_1}},\dots,\frac{m\hat{p}_{u_{k-r}}+n\hat{q}_{u_{k-r}}}{Sm\hat{p}_{u_{k-r}}},\frac{m\hat{p}_{u_1}+n\hat{q}_{u_1}}{Sn\hat{q}_{u_1}},\dots,\frac{m\hat{p}_{u_{k-r}}+n\hat{q}_{u_{k-r}}}{Sn\hat{q}_{u_{k-r}}}\right)$$

and $\Phi(y) = -\ln(y)$. Then, equivalently, $(\bar{\mathbf{p}}^r,\bar{\mathbf{q}}^r)$ solves the following optimization problem:

$$\text{minimize } \sum_{i=1}^{2r} w_i^{2r}\Phi(t_i^{2r})$$

subject to

$$\sum_{i=1}^{j} w_i^{2r}(h_i^{2r} - t_i^{2r}) \leq 0,\ j=1,\dots,k\text{-}r\text{-}1.$$

$$\sum_{i=k-r+1}^{k-r+j} w_i^{2r}(h_i^{2r} - t_i^{2r}) \geq 0,\ j=1,\dots,k\text{-}r\text{-}1,$$

and

$$\sum_{i=1}^{k-r} w_i^{2r}(h_i^{2r} - t_i^{2r}) = 0 = \sum_{i=k-r+1}^{2k-2r} w_i^{2r}(h_i^{2r} - t_i^{2r}).$$

These restrictions are equivalent to $\mathbf{h}^{2r}\mathbf{t}^{2r} \in \mathbf{B}^{2r^*w}$, where $\mathbf{B}^{2r^*w}$ is the Fenchel dual of $\mathbf{B}^{2r}$ and

$$\mathbf{B}^{2r} = \{(x_1,\dots,x_{k-r}, x_{k-r+1},\dots,x_{2k-2r}) \in R^{2k-2r}: x_1 \geq x_2 \geq \dots \geq x_{k-r}; x_{k-r+1} \leq x_{k-r+2} \leq \dots \leq x_{2k-2r}\}.$$

By the theorem of Robertson et al., 1988 on page 50, we have

$$(\bar{\mathbf{p}}^r,\bar{\mathbf{q}}^r) = \mathbf{w}^{2r} \cdot \mathbf{P}_{w^{2r}}(\mathbf{h}^{2r}|\mathbf{B}^{2r}).$$

Combining with the result that

$$\bar{p}_i = \bar{q}_i = \frac{m\hat{p}_i + n\hat{q}_i}{S},\ \forall\ i \in \mathbf{A}^{2r}$$

we have

$$(\bar{\mathbf{p}},\bar{\mathbf{q}}) = \mathbf{w}P_w \cdot (\mathbf{h}|\mathbf{B}_{(r)}).$$

(see the reason we gave in Theorem 2.1.) Separating $\mathbf{B}_{(r)}$ into two parts with the first k coordinates in the first part and the others in the second part, we can compute $\bar{\mathbf{p}}$ and $\bar{\mathbf{q}}$ independently as follows

$$\bar{\mathbf{p}} = (w_1,w_2,\dots,w_k)\cdot P_{(w1,w2,\dots,wk)}((h_1,h_2,\dots,h_k)|\mathbf{C}_{(r)})),$$

and

$$\bar{\mathbf{q}} = (w_{k+1},w_{k+2},\dots,w_{2k}) \cdot(-P_{(wk+1,wk+2,\dots,w2k)}(-(h_{k+1},h_{k+2},\dots,h_{2k})|\mathbf{C}_{(r)}))).$$

This completes the proof.

The Asymptotic Properties

As we know, the unconstrained maximum likelihood estimators, $\hat{\mathbf{p}}$ and $\hat{\mathbf{q}}$ converge almost surely to $\mathbf{p}$ and $\mathbf{q}$ for both one- and two-sample cases respectively. Applying this fact, we prove that the constrained maximum likelihood estimators we derived in the previous section also preserve the desired consistency property.

Theorem 3.1 As $m \to \infty$, $\bar{\mathbf{p}}$ converges almost surely to $\mathbf{p}$ provided $\mathbf{p} >> \mathbf{q}$ and $p_i = q_i$, for $i \in \mathbf{A}^r$.
Proof: By the strong law of large numbers, $\hat{\mathbf{p}} \to \mathbf{p}$ a.s. as $m \to \infty$. Furthermore, $\mathbf{P}_w(\mathbf{X}|\mathbf{C}_r)$ is continuous in both $\mathbf{w}$ and $\mathbf{X}$ so that $\bar{\mathbf{p}} \to \mathbf{p} \cdot \mathbf{P}_p(\mathbf{q}/\mathbf{p}|\mathbf{C}_{(r)})$ a.s. Using the maximum upper sets algorithm (cf. Robertson et. al. 1988 on page 25) to compute $\mathbf{p}$ ·

$\mathbf{P}_p(\mathbf{q}/\mathbf{p}|\mathbf{C}_{(r)})$, those terms with indices belonging to $\mathbf{A}^r$ form a level set so we only need to consider the other k-r terms. Refer to the equation on page 18 of Robertson et al., 1988, we have

$$AV((u_1,u_2,\ldots,u_j)) \overset{def}{=} \frac{\sum_{i=1}^{j}\left(\frac{q_{u_i}}{p_{u_i}}\right)\cdot p_{u_i}}{\sum_{i=1}^{j} p_{u_i}} = \frac{\sum_{i=1}^{j} q_{u_i}}{\sum_{i=1}^{j} p_{u_i}} \leqslant 1,$$

for $1 \leqslant j<k-r$ and the equality holds for $j=k-r$. Hence, the maximum upper set is $\{u_1,u_2,\ldots,u_{k-r}\}$ and $\mathbf{P}_p(\mathbf{q}/\mathbf{p}|\mathbf{C}_{(r)})=1$. Since $\bar{\mathbf{p}}\rightarrow\mathbf{p}\cdot\mathbf{P}_p(\mathbf{q}/\mathbf{p}|\mathbf{C}_{(r)})$, thus $\bar{p}_{u_j}\rightarrow p_{u_j}$, $j=1,\ldots,k-r$. In addition, to satisfy the assumption that $p_j=q_j$ for $j\in\mathbf{A}^r$, thus $\bar{p}_j=p_j=q_j$, for $j\in\mathbf{A}^r$. Hence, $\bar{\mathbf{p}}\rightarrow\mathbf{p}$. **This completes the proof.**

Theorem 3.2 Suppose that $\mathbf{p}>>\mathbf{q}$ and $p_i=q_i$, for $i\in A^r$, then

$$P\left\{\lim_{m,n\rightarrow\infty}(\bar{\mathbf{p}},\bar{\mathbf{q}})=(\mathbf{p},\mathbf{q})\right\}=1.$$

Proof: Since $\mathbf{P}_w(\mathbf{g}+\mathbf{d}_k|\mathbf{C}_{(r)})=\mathbf{P}_w(\mathbf{g}|\mathbf{C}_{(r)})+\mathbf{d}_k$, for any k dimensional constant vector $\mathbf{d}_k=(d,d,\ldots,d)$, it follows from Theorem 2.2 that

$$\bar{\mathbf{p}}-\hat{\mathbf{p}}=\hat{\mathbf{p}}\mathbf{P}_{\hat{p}}\left(\frac{n(\hat{\mathbf{q}}-\hat{\mathbf{p}})}{S\hat{\mathbf{p}}}|\mathbf{C}_{(r)}\right)$$

and

$$\bar{\mathbf{q}}-\hat{\mathbf{q}}=-\hat{\mathbf{q}}\mathbf{P}_{\hat{q}}\left(\frac{m(\hat{\mathbf{p}}-\hat{\mathbf{q}})}{S\hat{\mathbf{q}}}|\mathbf{C}_{(r)}\right).$$

By the strong law of large numbers, $\mathbf{P}\{\lim_{m,n\rightarrow\infty}(\hat{\mathbf{p}},\hat{\mathbf{q}})=(\mathbf{p},\mathbf{q})\}=1$. Since $(n/S)\hat{\mathbf{p}}$ and $(m/S)\hat{\mathbf{q}}$ are bounded by 0 and 1 (The bound still holds as m and $n\rightarrow\infty$ simultaneously) and $\mathbf{P}_w(\mathbf{X}|\mathbf{C})$ is continuous in $\mathbf{X}$ and $\mathbf{w}$, we only need to show that

$$\mathbf{P}_p\left(\frac{\mathbf{q}-\mathbf{p}}{\mathbf{p}}|\mathbf{C}_{(r)}\right)=\mathbf{P}_q\left(\frac{\mathbf{q}-\mathbf{p}}{\mathbf{q}}|\mathbf{C}_{(r)}\right)=0 \quad (8)$$

or equivalently

$$P_p\left(\frac{q}{p}|C_{(r)}\right)=P_q\left(-\frac{p}{q}|C_{(r)}\right)=\mathbf{e}_k, \quad (8)$$

where $\mathbf{e}_k$ is the k-dimensional vector of ones. By the assumption,

$AV((u_1,u_2,\ldots,u_j))\leqslant 1$, for $1\leqslant j<k-r$

and the equality holds for $j=k-r$. Then by the smae argument as that of Theorem 3.1,

$$\mathbf{P}_p\left(\frac{\mathbf{q}}{\mathbf{p}}|\mathbf{C}_{(r)}\right)=\mathbf{e}_k$$

follows easily and

$$\mathbf{P}_q\left(\frac{\mathbf{p}}{\mathbf{q}}|\mathbf{C}_{(r)}\right)=-\mathbf{e}_k$$

can also be established similarly. This completes the proof.

An Example

It is believed that political ideology are different by gender. Table 1, taken from the 1991 US General Social Survey (cf. Agresti (1996) page 203), shows the political ideology of Democratic affiliation in five levels (very liberal, slightly liberal, moderate, slightly conservative, and very conservative), stratified by gender. For simplicity, we denote this five level of political ideology as 1, 2, ..., 5. Let p_i (q_i) be the proportion of male (female) with political ideology level i, $i=1, 2, \ldots, 5$. If we know that male tends to be more liberal and female tends to be more conservative, a natural constraint is $\Sigma_{i=1}^{j} p_i \geqslant \Sigma_{i=1}^{j} q_i$ for $j=1, \ldots, 5$.

Table 1 The political ideology of Democratic affiliation stratified by gender.

	Political Ideology of democratic Party				
Gender	Very Liberal	Slightly Liberal	Moderate	Slightly Conservative	Very Conservtive
Male	36	34	53	18	23
Female	44	47	118	23	32

If we believe that about the same proportion of two genders tends to be moderate in the political ideology. We might consider the constraint

$$H: \sum_{i=1}^{j} p_i \geq \sum_{i=1}^{j} q_i \text{ for } j=1,\dots,5 \text{ and } p_3 = q_3.$$

In Table 2, $\hat{\mathbf{p}}$ ($\hat{\mathbf{q}}$) is the unrestricted m.l.e. of **p**(**q**); $\mathbf{p}^s$ ($\mathbf{q}^s$) is the m.l.e. of **p**(**q**) subject to the regular stochastic order **p**>>**q**; $\mathbf{p}^H$($\mathbf{q}^H$) is the m.l.e. of **p**(**q**) subject to the constraint **H**; fit^1 is the estimated frequency subject to usual stochastic order; fit^2 is the estimated frequency subject to the constraint **H**.

Table 2 Estimates for the political ideology data.

		fit^1		fit^2							
m	n	m	n	m	n	$\hat{p}$	$\hat{q}$	p^s	q^s	p^H	q^H
36	44	37.23	43.11	31.94	47.77	.22	.17	.23	.16	.19	.18
34	47	35.17	46.05	30.16	51.03	.21	.18	.21	.17	.18	.19
53	118	54.82	115.62	65.52	105.48	.32	.45	.33	.44	.40	.40
18	23	16.15	24.76	15.97	24.97	.11	.09	.10	.09	.10	.09
23	32	20.64	34.45	20.41	34.75	.14	.12	.13	.13	.12	.13

References

Agresti, A. (1996). *An introduction to categorical data analysis*, John Wiley & Sons, Inc.

Barlow, R. E. and Brunk, H. D. (1972). The isotonic regression problem and its dual. *J. Amer. Statist. Assoc., 67,* 140-147.

Brunk, H. D., Granck, W. E., Hanson, D. L., and Hogg, R. V. (1966). Maximum likelihood estimation of the distributions of two stochastically ordered random variables. *J. Amer. Statist. Assoc., 61,* 1067-1080.

Dykstra, R. L. (1982). Maximum likelihood estimation of the survival functions of stochastically ordered random variables. *J. Amer. Statist. Assoc., 77,* 621-628.

Dykstra, R. L. and Feltz, C. J. (1989). Nonparametric maximum likelihood estimation of survival functions with a general stochastic ordering and its dual. *Biometrika, 76,* 331-341.

Feltz, C. J. and Dykstra, R. L. (1985). Maximum likelihood estimation of the survival functions of N stochastically ordered random variables. *J. Amer. Statist. Assoc. 80,* 1012-1019.

Robertson, T and Wright, F. T. (1974). On the maximum likelihood estimation of stochastically ordered random variates. *Ann. Statist. 2,* 528-534.

Robertson, T and Wright, F. T. and Dykstra, R. L. (1988). *Order restricted statistical inference,* John Wiley & Sons, New York.

Sampson, A. R. and Whitaker, L. R. (1989). Estimation of multivariate distributions under stochastic ordering. *J. Amer. Statist. Assoc. 84,* 541-548.

收稿日期：民國87年9 月4 日
修正日期：民國87年12月10日
接受日期：民國87年12月11日

有等式隨機序的最大概似估計

張少同 呂小娟

國立台灣師範大學數學系 中國文化大學應用數學系

長久以來，有關隨機序的估計問題一直被廣泛的討論與應用，然而在許多實際的應用上，我們發現兩組多項式分配的參數 p 和 q 除了隨機序以外，可能還有其他的關係，例如相等、相差一個常數、或是成倍數關係。我們對參數了解愈多，愈能夠得到更精確的估計，因而在文章中，我們仿傚 Barlow 及 Brunk (1972) 的方法推得在隨機序與部份等式條件下，多項式參數的最大概似估計。我們依單樣本和雙樣本分別討論並提供應用實例，此外我們也探討估計的一致性問題。

關鍵詞：多項式參數、最大概似估計、隨機序、最小平方投影