The Annals of Statistics
1979, Vol. 7, No. 2, 282–283

ESTIMATION OF MULTINOMIAL PROBABILITIES[1]

BY KHURSHEED ALAM

Clemson University

This paper deals with the estimation of the parameters (cell probabilities) of a multinomial distribution. For the binomial distribution it is known that the maximum likelihood estimator (MLE) is admissible with respect to the squared error loss. It is shown that the MLE is admissible also in the case of the multinomial distribution.

1. Main results. Let $x = (x_1, \ldots, x_K)$ be distributed according to a multinomial distribution $M(\mathbf{x}, \mathbf{p}, n)$ with K cells, where $\mathbf{p} = (p_1, \ldots, p_K)$, $0 \leqslant p_i \leqslant 1$ $(i = 1, \ldots K)$, $\Sigma_{i=1}^K p_i = 1$ and $\Sigma_{i=1}^K x_i = n$. For estimating $\mathbf{p}$, let the loss be given by

$$L(\boldsymbol{\delta}, \mathbf{p}) = n\Sigma_{i=1}^K(\delta_i - p_i)^2$$

where $\delta_i = \delta_i(\mathbf{x})$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_K)$ is any estimator. The maximum likelihood estimator (MLE) is given by $\boldsymbol{\delta}° = (x_1/n, \ldots, x_K/n)$. Its risk is given by

$$\begin{aligned} R(\boldsymbol{\delta}°, \mathbf{p}) &= EL(\boldsymbol{\delta}°, \mathbf{p}) \\ &= 1 - \Sigma_{i=1}^K p_i^2. \end{aligned} \tag{1.1}$$

For $K = 2$, that is, in the case of the binomial distribution, it is well known that the MLE is admissible with respect to the given loss. A proof can be found in Johnson (1971). A proof for $K \geqslant 2$ is also given in that paper. We offer a different and somewhat simpler proof of the extended result.

Let

$$g(\mathbf{x}, \mathbf{p}, n) = \frac{n!}{x_1! \ldots x_K!} p_1^{x_1} \ldots p_K^{x_K}$$

denote the multinomial probability function, S denote the set of values of $\mathbf{x}$ for which $x_i \geqslant 1$, $i = 1, \ldots, K$ and T denote the complement of S. Let

$$C(\boldsymbol{\delta}, \mathbf{p}) = \Sigma_{i=1}^K \Sigma_S(\delta_i(\mathbf{x}) - p_i)^2 g(\mathbf{x}, \mathbf{p}, n)$$

$$D(\boldsymbol{\delta}, \mathbf{p}) = \Sigma_{i=1}^K \Sigma_T(\delta_i(\mathbf{x}) - p_i)^2 g(\mathbf{x}, \mathbf{p}, n)$$

where Σ_S and Σ_T denote the summation over the sets S and T, respectively. The admissibility of $\boldsymbol{\delta}°$ follows from the following theorem.

THEOREM 1.1. *If* $R(\boldsymbol{\delta}, \mathbf{p}) \leqslant R(\boldsymbol{\delta}°, \mathbf{p})$, $\forall \mathbf{p}$ *then* $\boldsymbol{\delta}(\mathbf{x}) = \boldsymbol{\delta}°(\mathbf{x})$, $\forall \mathbf{x}$.

Received October 1977; revised January 1978.

[1]The author's work was supported by the Office of Naval Research under Contract N00014-75-C-0451.

AMS 1970 *subject classification.* 62C15.

Key words and phrases. Multinomial distribution, maximum likelihood, empirical Bayes, admissible estimator.

PROOF. Suppose that

$$R(\boldsymbol{\delta}, \mathbf{p}) \leqslant R(\boldsymbol{\delta}^\circ, \mathbf{p}) = 1 - \Sigma_{i=1}^K p_i^2, \qquad \forall \mathbf{p}. \tag{1.2}$$

Let B denote the boundary of the parameter space. That is, if $\mathbf{p} \in B$ then $p_i = 0$ for some value of i. We have

$$R(\boldsymbol{\delta}, \mathbf{p}) = C(\boldsymbol{\delta}, \mathbf{p}) + D(\boldsymbol{\delta}, \mathbf{p}). \tag{1.3}$$

Clearly, $C(\boldsymbol{\delta}, \mathbf{p}) = 0$ for $\mathbf{p} \in B$.

First let $K = 2$. Then $R(\boldsymbol{\delta}^\circ, \mathbf{p}) = 0$ for $\mathbf{p} \in B$ and

$$D(\boldsymbol{\delta}, p) = 0 \qquad \text{for} \quad \mathbf{p} \in B. \tag{1.4}$$

It follows from (1.4) that $\boldsymbol{\delta}(\mathbf{x}) = \boldsymbol{\delta}^\circ(\mathbf{x})$ for each $\mathbf{x} \in T$. Hence, $D(\boldsymbol{\delta}, \mathbf{p}) = D(\boldsymbol{\delta}^\circ, \mathbf{p})$ and so

$$C(\boldsymbol{\delta}, \mathbf{p}) \leqslant C(\boldsymbol{\delta}^\circ, \mathbf{p}), \qquad \forall \mathbf{p}. \tag{1.5}$$

Multiplying both sides of (1.5) by $(p_1 p_2)^{-1}$ and integrating with respect to p_1 over the interval [0, 1] we have

$$\textstyle\int_0^1 C(\boldsymbol{\delta}, \mathbf{p})(p_1 p_2)^{-1} dp_1 \leqslant \int_0^1 C(\boldsymbol{\delta}^\circ, \mathbf{p})(p_1 p_2)^{-1} dp_1. \tag{1.6}$$

The integral on the left side is uniquely minimized by $\boldsymbol{\delta}(\mathbf{x}) = \boldsymbol{\delta}^\circ(\mathbf{x})$ for each $\mathbf{x} \in S$. It is already shown that $\boldsymbol{\delta}(\mathbf{x}) = \boldsymbol{\delta}^\circ(\mathbf{x})$ for each $\mathbf{x} \in T$. Therefore, $\boldsymbol{\delta}(\mathbf{x}) = \boldsymbol{\delta}^\circ(\mathbf{x})$, $\forall \mathbf{x}$ and so the theorem holds for $K = 2$.

Suppose that the theorem is true for $K = K^*$, say. Now let $K = K^* + 1$. Consider the inequality (1.2) and let $p_i = 0$ in turn for $i = 1, \ldots, K$. It follows from the assumption that the theorem is true for $K = K^*$ that $\boldsymbol{\delta}(\mathbf{x}) = \boldsymbol{\delta}^\circ(\mathbf{x})$ for each $\mathbf{x} \in T$. Hence $D(\boldsymbol{\delta}, \mathbf{p}) = D(\boldsymbol{\delta}^\circ, \mathbf{p})$ and so the inequality (1.5) holds. Multiplying both sides of the inequality by $(p_1 \ldots p_K)^{-1}$ and integrating with respect to $p_1, \ldots, p_{K^*}$ over the parameter space, we have

$$\textstyle\int C(\boldsymbol{\delta}, \mathbf{p})(p_1 \ldots p_K)^{-1} dp_1 \ldots dp_{K^*} \leqslant \int C(\boldsymbol{\delta}^\circ, \mathbf{p})(p_1 \ldots p_K)^{-1} dp_1 \ldots dp_{K^*}. \tag{1.7}$$

Since $g(\mathbf{x}, \mathbf{p}, n)(p_1 \ldots p_K)^{-1}$ is finite for $\mathbf{x} \in S$, the integral on the right side of (1.7) is finite and the integral on the left side is uniquely minimized by $\boldsymbol{\delta}(\mathbf{x}) = \boldsymbol{\delta}^\circ(\mathbf{x})$ for $\mathbf{x} \in S$. It has been shown above that $\boldsymbol{\delta}(\mathbf{x}) = \boldsymbol{\delta}^\circ(\mathbf{x})$ for $\mathbf{x} \in T$. Therefore, $\boldsymbol{\delta}(\mathbf{x}) = \boldsymbol{\delta}^\circ(\mathbf{x})$, $\forall \mathbf{x}$. By the induction principle the theorem is true for all $K \geqslant 2$.

Acknowledgment. Thanks are due to the referee for his useful comments on the paper.

REFERENCES

[1] JOHNSON, B. MCK. (1971). On the admissible estimators for certain fixed sample binomial problems. *Ann. Math. Statist.* **42** 1579-1587.

DEPARTMENT OF MATHEMATICAL SCIENCES
CLEMSON UNIVERSITY
CLEMSON, SOUTH CAROLINA 29631