Comm. Korean Math. Soc. **11** (1996), No. 4, pp. 1187–1192

UNBIASED ESTIMATORS IN THE MULTINOMIAL CASE

Choon-Il Park

Abstract. It is known that an unbiased estimator of $f(\mathbf{p})$ for binomial $B(n,p)$ exists if and only if f is a polynomial of degree at most n, in which case the unbiased estimator of a real-valued function $f(\mathbf{p}), \mathbf{p} = (p_0, p_1, \cdots, p_r)$ is unique. In general, this estimator has the serious fault of not being range preserving; that is, its value may fall outside the range of$f(\mathbf{p})$. In this article, a condition on a real-valued function f is derived that is necessary for the unbiased estimator to be range preserving that this is sufficient when n is large enough.

1. Introduction

An estimator is said to be range preserving if its values are confined to the range of what it is to estimate. The property of being range preserving is an essential property of an estimator, a sine qua non. Other properties, such as unbiasedness, may be desirable in some situations, but an unbiased estimator that is not range preserving should be ruled out as an estimator.

In 1977, authors [4] have explicitly insisted that an estimator should be range preserving. Occasionally it has been noticed that a proposed estimator dose not have this property, and comments on this regrettable fact have been made. Examples of unbiased estimators that fail to preserve the range have been given by several authors (see, e.g., Halmos [1] and Lehmann [3], for further related work, see Hoeffding [2] and the references cited there).

For clarity, the result will first be stated and proved for the binomial case. The extension to the multinomial case, which is straightforward

Received May 11, 1996. Revised September 8, 1996.
1991 AMS Subject Classification: 62B05.
Key words and phrases: Unbiased estimator, range preserving.

except for the more complicated notation, will be dealt with in section 3.

Let the random variable X have the binomial (n,p) distribution, $0 \leq p \leq 1$, and consider the unbiased estimation, based on an observation of X, of a real-valued funtion $f(p)$. It is known that an unbiased estimator of $f(p)$ exists if and only if f is a polynomial of degree at most n, say

$$f(p) = a_0 + a_1 p + \cdots + a_m p^m \tag{1.1}$$

where $m \leq n$. Because

$$p^k = \sum_{x=0}^{n} \frac{x^{(k)}}{n^{(k)}} \binom{n}{x} p^x (1-p)^{n-x}, \; k = 0, 1 \cdots, n,$$

where $n^{(k)} = n(n-1)\cdots(n-k+1)$, an unbiased estimator $t_n(X)$ of $f(p)$ in (1.1) is given by

$$t_n(x) = \sum_{k=0}^{m} a_k \frac{x^{(k)}}{n^{(k)}}, \; x = 0, 1, \cdots, n, \; n \geq m. \tag{1.2}$$

By the completeness of the binomial family (Lehmann 1959), this is the only unbiased estimator of $f(p)$.

In general, the values of $t_n(x)$ can fall outside the range of $f(p)$. Thus the unbiased estimator of $p(1-p)$ is $x(n-x)/n(n-1)$, and its maximum with respect to x exceeds $1/4 = \max p(1-p)$.

2. Unbiased estimators

DEFINITION 2.1. Let $\mathcal{P}$ be a family of probability distributions on a measurable space $(\mathcal{X}, \mathcal{A})$, $\mathcal{P} = \mathcal{P}_n$ is the family of the binomial (n,p) distributions P_p, $0 \leq p \leq 1$, with n fixed. Let $f(p) = \theta(P_p)$. The prior range of $f(p)$ is $\Theta = \{f(p) : 0 \leq p \leq 1\}$. The posterior range Θ_x is given by

$$\Theta_x = \{f(p) : 0 < p < 1\} \text{ if } x = 1, 2, \cdots, n-1,$$
$$\Theta_0 = \{f(p) : 0 \leq p < 1\}, \; \Theta_n = \{f(p) : 0 < p \leq 1\}.$$

THEOREM 2.1. *Let f be a nonconstant polynomial. In order for a range-preserving unbiased estimator of $f(p)$, $0 \le p \le 1$, to exist, f must attain its extreme values only at 0 and 1 and f' must be nonzero there, that is, either*

1. $f(0) < f(p) < f(1)$ *for* $0 < p < 1$; $f'(0) > 0$, $f'(1) > 0$,

or

2. $f(1) < f(p) < f(0)$ *for* $0 < p < 1$; $f'(0) < 0$, $f'(1) < 0$.

Thus the polynomial $f(p) = (p-c)^3$, $0 < c < 1$, satisfies condition 1, but for $c > 1/2$, $t_3(1) = c^2(1-c) > (1-c)^3 = f(1)$.

PROOF. Suppose that $f(0) < f(p) < f(1)$ for $0 < p < 1$. It is sufficient to show that the unbiased estimator $t_n(x)$ of $f(p)$ is range-preserving only if $f'(0) > 0$ and $f'(1) > 0$. According to the definitions, for $t_n(x)$ to be range-preserving it is necessary that $t_n(1) > f(0)$ and $t_n(n-1) < f(1)$. By (1.2) and (1.1),

$$t_n(1) = a_0 + a_1 n^{-1} = f(0) + f'(0)n^{-1},$$

$$t_n(n-1) = \sum_0^m a_k - \sum_0^m k a_k n^{-1} = f(1) - f'(1)n^{-1},$$

and the stated conditions follow. For case 2 the completion of the proof is similar.

THEOREM 2.2. *Let f be a polynomial that satisfies condition 1 or 2 of Theorem 1. Then there exists a number $N(f) \ge deg(f)$ such that the unbiased estimator $t_n(x)$ of $f(p)$ is range preserving for $n \ge N(f)$.*

PROOF. Suppose that $f(p)$ satisfies condition 1 of Theorem 1. Case 2 can be reduced to case 1 by a simple change of notation. We must show that there is an integer $N(f) \ge \deg(f)$ such that for $n \ge N(f)$, the unbiased estimator $t_n(x)$ is range preserving; that is,

$$\begin{aligned} &f(0) < t_n(x) < f(1), \quad x = 1, 2, \cdots, n-1, \\ &f(0) \le t_n(0) < f(1), \quad f(0) < t_n(n) < f(1). \end{aligned} \tag{2.1}$$

Introduce the function $s_n(p)$, $0 \le p \le 1$, as

$$s_n(p) = \sum_{k=0}^{m} a_k \prod_{j=0}^{k-1} \frac{p - j/n}{1 - j/n}.$$

Then it follows from (1.1) and (1.2) that

$$s_n(x/n) = t_n(x), \; s_n(0) = f(0), \; s_n(1) = f(1).$$

Thus in particular, s_n and s_n' converge uniformly on [0,1] to f and f', respectively. This ensures that for large enough n, s_n is range preserving, that is, $f(0) < s_n(p) < f(1)$ for $0 < p < 1$, and thus the inequalities (2.1) are satisfied.

3. Multinomial case

Let $\mathbf{X}$ have the multinomial $(n, \mathbf{p})$ distribution:

$$\Pr\{\mathbf{X} = x\} = \frac{n!}{\mathbf{x}!}\mathbf{p}^{\mathbf{x}}, \; \mathbf{x} \in \mathcal{X}, \; \mathbf{p} \in \Sigma_r,$$

$$\Sigma_r = \{\mathbf{p} = (p_0, \cdots, p_r) : p_0 \ge 0, \cdots, p_r \ge 0, \sum_0^r p_i = 1\},$$

where the notation $\mathbf{x}! = \prod_0^r x_i!$, $\mathbf{p}^{\mathbf{x}} = \prod_0^r p_i^{x_i}$ is used.

Consider the unbiased estimation, based on an observation of $\mathbf{X}$, of the value of a real-valued function $f(\mathbf{p})$, $\mathbf{p} \in \Sigma_r$. As in the binomial case, an unbiased estimator of $f(\mathbf{p})$ exists if and only if f is a polynomial of degree at most n, say

$$f(\mathbf{p}) = \sum_m a_{\mathbf{j}} \mathbf{p}^{\mathbf{j}}, \tag{3.1}$$

where $m \le n$ and the sum $\sum_m$ is extended over the vector indexes $\mathbf{j} = (j_0, j_1, \cdots, j_r)$ with nonnegative integer-valued coordinates whose sum is $|\mathbf{j}| = \sum_0^r j_j \le m = \deg(f)$. As a result of the identity

$$\mathbf{p}^{\mathbf{j}} = \sum_{\mathbf{x} \in \mathcal{X}} \frac{\mathbf{x}^{(\mathbf{j})}}{n^{(|\mathbf{j}|)}} \frac{n!}{\mathbf{x}!} \mathbf{p}^{x},$$

where $\mathbf{x}^{(\mathbf{j})} = \prod_0^r x_i^{(j_i)}$, an unbiased estimator $t_n(\mathbf{X})$ of $f(\mathbf{p})$ in (3.1) is given for $n \geq \deg(f)$ by

$$t_n(\mathbf{x}) = \sum_m a_{\mathbf{j}} \mathbf{x}^{(\mathbf{j})} / n^{(|\mathbf{j}|)}. \tag{3.2}$$

By the completeness of the multinomial family (Lehmann 1951, p.132) this is the only unbiased estimator of $f(\mathbf{p})$.

THEOREM 3.1. *Let the random vector* $\mathbf{X}$ *have a multinomial* $(n, \mathbf{p})$ *distribution,* $\mathbf{p} \in \Sigma_r$, *and let* $f(\mathbf{p})$, $\mathbf{p} \in \Sigma_r$, *be a nonconstant polynomial,* $\deg(f) \leq n$. *For the unbiased estimator* $t_n(\mathbf{x})$ *of* $f(\mathbf{p})$ *to be range preserving, it is necessary that:*

1. if F *is an* s*-dimensional open face of* $\Sigma_r, 1 \leq s \leq r$ *(its interior when* $s = r$*), and if* $f(\mathbf{p})$ *is not constant on* F, *then* $f(\mathbf{p})$ *attains its extreme values only on the boundary of* F*; and*

2. if F *is an* s*-dimensional closed face of* $\Sigma_r, 0 \leq s \leq r-1$ *(a vertex when* $s = 0$*), on which* $f(\mathbf{p})$ *is constant and attains its minimum (maximum) over* Σ_r, *then the derivatives of* $f(\mathbf{p})$ *in directions orthogonal to* F *are strictly positive (negative).*

PROOF. The necessity of condition 1 follows by the argument given in Theorem 2.1. The necessity of condition 2 is shown by focusing on a direction orthogonal to a face with the stated properties and using a one-dimensional argument similar to that used in Theorem 2.1.

THEOREM 3.2. *Let* f *be a polynomial satisfying the conditions of Theorem 3.1. Then there exists a number* $N(f)$ *such that the unbiased estimator* $t_n(\mathbf{x})$ *of* $f(\mathbf{p})$ *is range-preserving for* $n \geq N(f)$.

PROOF. As in the proof of Theorem 2.2, the function $s_n(\mathbf{p})$, $\mathbf{p} \in \Sigma_r$, is introduced by

$$s_n(\mathbf{p}) = \sum_m a_{\mathbf{j}} \frac{\prod_0^{j_0}(p_0 - k_0/n) \cdots \prod_0^{j_r}(p_r - k_r/n)}{\prod_0^{|\mathbf{j}|}(1 - k/n)} \tag{3.3}$$

It is easily seen from (3.1) that s_n equals f at all vertices of Σ_r and at every s-dimensional face of Σ_r, $1 \leq s \leq r-1$, where f is constant. It is also clear from (3.1) and (3.4) that s_n converges to f on Σ_r, and the same is true for all partial derivatives. Finally, in view of (3.3), $t_n(\mathbf{x}) = s_n(\mathbf{x}/n)$. The proof is then completed as in the binomial case.

References

1. Halmos, Paul R., *The Theory of Unbiased Estimation*, Annals of Mathematical Statistics **17** (1946), 34-43.
2. Hoeffding, Wassily, *"Unbiased Range-Preserving Estimators", in A Festschrift for Erich L. Lehmann, eds. P. J. Bickel, K. Doksum, and J. L. Hodges, Jr., Belmont, Calif.*, Wadsworth (1983), 249-260.
3. Lehmann, E. L., *A General Concept of Unbiasedness*, Annals of Mathematical Statistics **22** (1951), 587-592.
4. Peter J. Bickel and Kjell A. Doksum, *Mathematical Ststistics*, Holden-Day, Inc. (1977), 120-123.
5. V. K. Rohatgi, *An Introduction to Probability Theory and Mathematical Statistics*, New-York and London (1972),350-359.

Department of Applied Mathematics
Korea Maritime University
Pusan 606-791, Korea