MINIMUM VARIANCE UNBIASED ESTIMATION OF COIN TOSSING PROBLEMS

Abdul wahid Brohi
Azizullah Memon
Jawaid Hussain Khawaja

Abstract

The use of series expansion in estimation problems for distributions involving more than one parameter is illustarted in this paper. Unbiased minimum variance estimates are derived for the probabilities p_1 , p_2 of obtaining heads with two coins 1, 2 in two " Markov" binomial sampling problems.

1. INTRODUCTION

We consider the problem of biased minimum variance (UMV) estimation in binomial sampling with two coin 1, 2 with probabilities p_1 , p_2 of obtaining heads respectively ($p_1 \neq p_2$). Such problems have been considered by Naryana (1970), Naryana Sathe and Chroneyko (1980) and the use of series expansions in UMV estimation problems has been discussed by DeGroot (1969), Guttman (1968). In Section 2, we consider an example of estimation in an inverse binomial sampling problem, which though artificial, illustrates these principles and generalises previous results. The generating function (g.f) for this problem is given explicity in section 3 following methods given by Feller (1975). The rest of the paper deals with the application of similar combinatorial methods and the use of g.f.'s in related problems. (Narayana and ladouceur)

2. UMV ESTIMATION IN MARKOV INVERSE BINOMIAL SAMPLING

Consider the game N'r played with two coins 1,2 with probabilities p_1 , p_2 for head

$$(0 < p_1 < 1; p_1 \neq p_2 , p_1 \neq q_2 = 1 - p_2^2)$$

according to the following rules:

i) The first trial is made with coin 1.

ii) The n-th trial is made with coin 1 or 2 according as the (n-1)-st trial is tail or head (n > 1).

iii) We stop the sequence of trials at that trial where the total number of heads equals r for the first time r being a positive integer.

In what follows x and o stand for head and tail respectively.

The game N'r can end at the (n+r)-th trial n= 0, 1, 2, ...) of a sequence of no's and r x's, the last trial being, of course x. Consider the sequences ending at the $(n_0 + r)$-th trial which (a) either begin with an x and have K-1 changes of types ox or (b) begin with a) and have K changes of type ox.

The probability associated with such sequences is $p_1^{k} p_2^{r-k} q_1^{no-k+1} q^{2k-1}$ and therefore (n,k), by and appropriate choice of k for each n, can be considered a minimal sufficient statistic for the distribution. For given n,k varies between 1 and min (r, n+1). The number of sequences of type (a) i.e. stating with x and having k -1 changes ox are in one to one correspondence with different possible arguments of r-2x's and no's with k-1 runs of o's. The number of (a) sequences is thus

$$\begin{bmatrix} n - k \\ k - 2 \end{bmatrix} \qquad \begin{bmatrix} r - 1 \\ k - 1 \end{bmatrix} \tag{2.1}$$

Similarly the number of sequences (b) starting with exactly v o's (v=1,.....,n - k+ 1):

$$\begin{bmatrix} n - 1 - v \\ k - 2 \end{bmatrix} \qquad \begin{bmatrix} r - 1 \\ k - 1 \end{bmatrix} \tag{2.2}$$

As the number of sequences in (2.1) is the same as substituting v = o in (2.2), the total number of sequences with probability

$p_1^{k}\, p_2^{r-k}\, q_1^{n-k+1}\, q_2^{k-1}$, is

$$N'r(k,n) = \sum_{r=0}^{n-k+1} \begin{bmatrix} n-1-v \\ k-2 \end{bmatrix} \begin{bmatrix} n \\ k-1 \end{bmatrix}$$

$$= \begin{bmatrix} r-1 \\ k-1 \end{bmatrix} \begin{bmatrix} n \\ k-1 \end{bmatrix} \qquad (2.3)$$

Hance $\displaystyle\sum_{n=0}^{\infty} \sum_{k=1}^{\min(r,n+1)} N_r'(k,n)\, p_1^{k}\, p_2^{r-k} q_1^{n-k+1} q_2^{k-1}$

$$= \sum_{n=0}^{\infty} \sum_{k=1}^{\min(r,n+1)} N_r'(K, n)\, \theta_1^{k}\, \theta_2^{n}\, m(\theta_1, \theta_2) = 1 \qquad (2.4)$$

Where $\theta 1 = \dfrac{p_1 q_2}{p_2 q_1}$, $\theta_2 = q_1$ and $m(\theta_1, \theta_2)$

$$= \frac{p_2^{r} q_1}{q_2} = \frac{(1 - \theta_2 + \theta_1\theta_2)^{r-1}}{\theta_1 (1 - \theta_2)^{r-2}}$$

The form (2.4) indicates that the statistic (k, n) is complete. Therefore the unique UMV estimates of p_1 , p_2 are

$$\hat{p}_1 = \frac{\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}}{\begin{bmatrix} n \\ k-1 \end{bmatrix}} = \frac{k-1}{n}; \quad \hat{p}_2 = \frac{\begin{bmatrix} r-2 \\ k-1 \end{bmatrix}}{\begin{bmatrix} r-1 \\ k-1 \end{bmatrix}} = \frac{r-1}{r-k} \qquad (2.5)$$

3. THE GENERATING FUNCTION OF $N_r'(s)$

Consider the game N_r which is played with the same rules as N_r' except that at the first trial coin 2 is used instead of coin 1 (of (i), Section 2). Let N_r', $N_r(s)$ denote the g.f of these games i.e.

$$N_r'(s) = \sum_{n=0}^{\infty} \sum_{k=1}^{\min(r,n+1)} N_r'(k,n)\, x \; p_1^{k} p_2^{r-k} q_1^{n-k+1} q_2^{k-s^{n+2}} \qquad (3.1)$$

A method similar to section 2 gives

$$Nr(s) \sum_{n=0}^{\infty} \sum_{k=0}^{\min(r,n+1)} \begin{bmatrix} n \\ k-1 \end{bmatrix} \begin{bmatrix} r \\ k \end{bmatrix} p_1^{k} p_2^{r-k} q_1^{n-k+1} q_2^{k-s^{n+2}} \qquad (3.2)$$

wher we make the convention that the coefficient

$$\begin{bmatrix} 0 \\ -1 \end{bmatrix} = 1, \begin{bmatrix} n \\ -1 \end{bmatrix} = 0 \text{ for all positive n.}$$

Setting $N_0(s) = N'_0(s) = 1$ we have the relarions

$$\begin{cases} N_r(s) = p_2^{\,s} N_{r-1}(s) + q_2^{\,s} N'_r(s) \\ N'_r(s) = p_1^{\,s} N_{r-1}(s) + q_1^{\,s} N_r(s) \end{cases} \qquad (3.3)$$

$$N'_1(s) = \frac{p_1^{S}}{1.-q_1^{S}} \text{ and } N_1(s) = p_2^{S} + \frac{p_1 q_2^{S^2}}{1-q_1^{S}} \qquad (3.4)$$

Using (3.3) and (3.4) we see easily

$$N_r(s) = [N_1(s)]^r \text{ and } N'_r(s) = N'_1(s)[N_1(s)]^{r-1} \qquad (3.5)$$

a relation, which we expect to old from the definitions of the games. Writing equations (3.1), (3.2) as power series in s, the relations (3.3) are valid. Inversely (3.4), (3.5) yield in turn (3.1),(3.2) and the UMV unbiased estimates for p_1, p_2. Differentiation of (3.5) using (3.4) gives the means and variance of the number of trials required to terminate N_r, N'_r.

$$E(N_r) = \frac{r(p_1+q_2)}{p_1}, \quad E(N'r) = (r-1)\frac{p_1+q_2}{p_1} + \frac{1}{p_1}$$

4. GENERATING FUNCTIONS FOR RELATED PROBLEMS

Consider the games G_r , G'_r and their g.f's $G_r(s)$, $G'_r(s)$ defined analogously to N_r , N'_r except that the stopping rule specifies that the total number of heads exceeds the total number of tails by r for the first time Narayana (1970); Narayana, Sathe and Chorneyko (1980). In order that the games G_r, G'_r termainate with probability I it is necessary and sufficient that $p_1 + p_2 \geq q_1 + q_2$. We assume this condition holds. As

$$G_r(s) = p_2^{\ s}\, G_{r-1}(s) + q_2^{\ s}\, G'_{r+1}(s)$$
$$G'_r(s) = p_1^{\ s}\, G_{r-1}(s) + q_1^{\ s}\, G'_{r+1}(s).$$

Where $G_0(s) = G'_0(s) = 1$, we see as before

$$G_r(s) = [G_1(s)]^r,\ G_2'(s) = G'_1(s)\,[G_1(s)]^{r-1} \tag{4.1}$$

$G_1(s)$, $G'_1(s)$ are in fact the negative roots of the equations

$$\begin{cases} q_1^{\ s} x^2 - x\{1 + p_2 q_1^{\ s^2} - p_1 q_2^{\ s^2}\} + p_2^{\ s} = 0 \\ q_1^{\ s} x^2 - x\{1 + p_1 q_2^{\ s^2} - p_2 q_1^{\ s^2}\} + p_1^{\ s} = 0 \end{cases} \tag{4.2}$$

Differentiating (4.2) we have means and variances of the number of trials required to ferminate G_r, G'_r. These results generalise Narayana (1967). The expected values are

$$E(G_r) = r\,(p_1 + q_2)\,(p_1 - q_2)$$

$$E(G'_r) = (r-1)\,\frac{(p_1+q_2)}{(p_1-q_2)} + \frac{(p_2+q_1)}{(p_1-q_2)}.$$

Their connection with Feller (1975) and the replacement of the two coins in G_r by a single coin with probabilities $\frac{p_1}{p_1+q_2}$, $\frac{q_2}{p_1+q_2}$ for the heads and tails is clear. The problem of obtaining unbiased estimates for P_1, P_2 in G_r , G'_r has been treated by Narayana, Sathe and Chorneyko (1980) and these results in turn imply the identity

$$\sqrt{1+u^2+v^2-2u-2v-2uv} = 1-u-v-2x$$

$$\sum_{s=1}^{\infty}\sum_{t=1}^{\infty} \frac{\left[{s+t-1 \atop s}\right]\left[{s+t-1 \atop t}\right]}{s+t-1} u^s v^t \quad (4.3)$$

We are investigating and independent proof of (4.3) by analytic methods. We finally remarks that (4.3) and the use of g.f's simplifies similar problems in unbiased estimation of (Narayana and ladoceur))

REFERENCES:

DEGROOT: M.H. (1969): Unbiased sequential estimation for populations. Ann. Math. Stat., 30, 80.

FILLER, W, (1975): As introduction to the Theory of Probability and Its Applications, Vol.I, John wiley and Sons, Inc. (New Yourk).

GUTTMAN, I. (1968): A note on a series solution of a problem in estimation. Biometrika, 45, 365.

NARAYANA,T.V. (1967): A problem in the theory of Probability. J. Indian Soc. Agric. Stat., 6, 139.

NARAYANA (1955): Surles treillis formes parles partitions d'unentier; leurs applications a la des probabilities, Cont. 240, 1188.

NARAYANA, T.V., MOHANTY, S.C. and LADOCEUR, J.D.: A combinatorinal problem and its application to proabability theory-II (To appear in J. Indian Soc. Agric. Stat.).

NARAYANA, T.V. SATHE, Y.S. and CHORNEYKO. i. (1980): Sufficient partitions for a class of coin-tossing problems Biometrische Zeitschrift, Band 2, Heft 4.