Efficient Estimator of Parameters of a Multivariate Geometric Distribution

U. J. Dixit[1], * and S. Annapurna[2]

[1]*Department of Statistics, University of Mumbai, Mumbai, India*
[2]*Department of Statistics, St. Xaviers College (Autonomous), Mumbai, India*

ABSTRACT

The maximum likelihood estimator (MLE) and uniformly minimum variance unbiased estimator (UMVUE) for the parameters of a multivariate geometric distribution (MGD) have been derived. A modification of the MLE estimator (modified MLE) has been derived in which case the bias is reduced. The mean square error (MSE) of the modified MLE is less than the MSE of the MLE. Variances of the parameters and the corresponding generalized variance (GV) has been obtained. It has been shown that the MLE and modified MLE are consistent estimators. A comparison of the GVs of modified MLE and UMVUE has shown that the modified MLE is more efficient than the UMVUE. In the final section its application has been discussed with an example of actual data.

1. INTRODUCTION

It is appropriate and convenient to measure lifetime of devices such as on/off switches, bulbs, engines of an airplane on a discrete scale. Discrete random variables also help study lifetimes such as the incubation period of diseases like AIDS, the remission time of cancers as well as time to failure of engineering systems (see [1]). The discrete multivariate distributions are useful to measure lifetime data. The (MGD) has been vital in studying reliability analysis. Various models of the bivariate geometric distribution (BGD) have been proposed to study lifetime devices. Downtown [2] has described a model for developing a BGD. This arises in a shock model with two components. Downtown [2] describes this model asfollows. Suppose that the number of shocks suffered by each component before failure can be represented by a population in which proportions p_1 and p_2 affected the first and second components respectively, without failure and a proportion 1-p_1-p_2 of the shocks lead to failure of both the components. Hence X is number of shocks to component 1 prior to the first failure and Y is number of shocks to component 2 prior to the first failure. The joint probability function of (X, Y) is given by

$$\pi_{X,Y}(t_1, t_2) = \left(1 - p_1 - p_2\right)\left(1 - p_1 t_1 - p_2 t_2\right)^{-1} \qquad (1)$$

The corresponding joint probability mass function of (X, Y) is

$$P\left(X = x, \ Y = y\right) = \begin{cases} \begin{pmatrix} x + y \\ x \end{pmatrix} p_1^x p_2^y p_3 & ; \ x = 0, \ 1, \ 2, \dots, \ y = 0, \ 1, \ 2, \dots, \\ & \quad 0 < p_1 < 1; \ 0 < p_2 < 1; \ p_3 = 1 - p_1 - p_2 \\ 0 & ; \ \text{otherwise} \end{cases} \qquad (2)$$

Hare Krishna and Pundir [3] have obtained the MLE and Bayes estimators of the parameters for this BGD. Dixit and Annapurna [4] have further obtained the UMVUE estimators and have compared the MLE and UMVUE based on the (MSEs). Phatak and Sreehari [5] introduced a version of bivariate geometric distribution as a stochastic model for giving the distribution of good and marginally good items that are produced by a production unit. Marshall and Olkin [6] constructed a BGD based on the sequence of Bernoulli random variables in which X was defined as the number of trials required for the *rth* occurence of an event A and Y was the number of trials required for the *sth* occurence of an event B. If we consider Bernoulli trials then (X, Y) will have 4 possible values (0,0), (0,1), (1,0) and (1,1).

*Corresponding author. Email: ullllhasdixit@yahoo.com.in

We define $P\left(X = i, Y = j\right) = p_{ij}$; $i, j = 0, 1$.

Let $p_{00} + p_{01} = p_{0+}, p_{10} + p_{11} = p_{1+}, p_{00} + p_{10} = p_{+0}, p_{01} + p_{11} = p_{+1}$

Marginally X and Y have negative binomial distribution. X follows NB (r, p_{1+}) and Y follows NB (s, p_{+1}).

The marginal probability functions are of the usual form but the joint probability function of (X, Y) is quite complicated. Reference may be made to Marshall and Olkin ([6] Eq. 7.2).

On the same lines, Gultekin and Bairamov [7] constructed a trivariate geometric distribution and the corresponding multivariate extension. Srivastava and Bagchi [8] introduced the multivariate version of a geometric distribution and obtained certaincharacterizations. Vasudeva and Srilakshminarayana [9] established some properties of the MGD and also obtained a characterization assuming it to follow the power series distribution. Sreehari and Vasudeva [10] have given characterization of the MGD based on conditional distributions. Esary and Daniel [11] studied properties of MGDs that were generated by a cumulative damage process. In this paper we look at another form of the MGD and estimate its parameters by a new approach that reduces the bias.

Consider a system which comprises of k components namely $C_1, C_2,...,C_k$. The system is so designed that at any given time not more than one component can function. The system functions when any one component functions. The system initially functions because C_1 functions. When C_1 stops functioning C_2 starts functioning in the next trial keeping the system functioning. Thus system continues to function in this manner till C_k functions. Let probability that component C_i fails be p_i, $i = 1,2,3, ..., k$. Let X_i denote the trial at which component C_i fails, $i = 1,2, ..., k$.

The joint probability mass function of $(X_1, X_2, \cdots, X_k)$ is given as

$$P(X_1 = x_1, X_2 = x_2, ... X_k = x_k) = \begin{cases} (1-p_1)^{x_1-1}p_1(1-p_2)^{x_2-x_1-1}p_2 ... (1-p_k)^{x_k-x_{k-1}-1}p_k; \\ 1 \le x_1 < x_2 < ... < x_k; 0 < p_i < 1; i = 1, ..., k \\ 0 \; ; \quad \text{otherwise} \end{cases} \tag{3}$$

The probability generating function (pgf) is given as

$$P_{X_1 X_2 ... X_k}(t_1, t_2, ..., t_k) = \frac{p_1 p_2 ... p_k t_1 t_2^2 ... t_k^k}{\left[1 - \left(1-p_1\right) t_1 t_2 ... t_k\right]\left[1 - \left(1-p_2\right) t_2 t_3 ... t_k\right] ... \left[1 - \left(1-p_k\right) t_k\right]} \tag{4}$$

In this paper we obtain UMVUE and MLE of the parameters in Eq. (1) and their functions.

2. UNIFORM MINIMUM VARIANCE UNBIASED ESTIMATOR (UMVUE)

Here we obtain the UMVUE of the parameters as well as of the functions of the parameters. We consider the trivariate case which has three parameters, p_1, p_2 and p_3. Here p_i denotes the corresponding probability of the ith component in the system failing.

Consider the case where $k = 3$. Eqs. (1) and (2) become

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \begin{cases} (1-p_1)^{x_1-1}p_1(1-p_2)^{x_2-x_1-1}p_2(1-p_3)^{x_3-x_2-1}p_3 ; \\ 1 \le x_1 < x_2 < x_3; 0 < p_i < 1; i = 1, 2, 3 \\ 0; \quad \text{otherwise} \end{cases} \tag{5}$$

$$P_{X_1 X_2 X_3}(t_1, t_2, t_3) = \frac{p_1 p_2 p_3 t_1 t_2^2 t_3^3}{\left[1 - \left(1-p_1\right) t_1 t_2 t_3\right]\left[1 - \left(1-p_2\right) t_2 t_3\right]\left[1 - \left(1-p_3\right) t_3\right]} \tag{6}$$

The pgf of $S_1 = \sum_{i=1}^{n} X_{1_i}$, $S_2 = \sum_{i=1}^{n} X_{2_i}$ and $S_3 = \sum_{i=1}^{n} X_{3_i}$ is

$$P_{S_1 S_2 S_3}(t_1, t_2, t_3) = \left(\frac{p_1 p_2 p_3 t_1 t_2^2 t_3^3}{\left[1 - \left(1-p_1\right) t_1 t_2 t_3\right]\left[1 - \left(1-p_2\right) t_2 t_3\right]\left[1 - \left(1-p_3\right) t_3\right]}\right)^n \tag{7}$$

Hence the pmf of S_1, S_2 and S_3 is co-efficient of $t_1^{s_1} t_2^{s_2} t_3^{s_3}$ in Eq. (7) and is given as

$$P(s_1, s_2, s_3)$$
$$= \binom{s_3 - s_2 - 1}{s_3 - s_2 - n}\binom{s_2 - s_1 - 1}{s_2 - s_1 - n}\binom{s_2 - 1}{s_2 - n} p_1^n p_2^n p_3^n (1-p_1)^{s_1-n}(1-p_1)^{s_2-s_1-n}(1-p_1)^{s_3-s_2-n};$$
$$n \le s_1 < s_2 < s_3 < \infty; 0 < p_1 < p_2 < p_3 < 1; 0 < p_3 \tag{8}$$

Theorem 2.1. The UMVUE of $p_1^{a_1} p_2^{a_2} p_3^{a_3} (1 - p_1)^{b_1} (1 - p_2)^{b_2} (1 - p_3)^{b_3}$ is

$$\frac{\binom{s_1 - b_1 - a_1 - 1}{s_1 - b_1 - n} \binom{s_2 - s_1 - b_2 - a_2 - 1}{s_2 - s_1 - b_2 - n} \binom{s_3 - s_2 - b_3 - a_3 - 1}{s_3 - s_2 - b_3 - n}}{\binom{s_1 - 1}{s_1 - n} \binom{s_2 - s_1 - 1}{s_2 - s_1 - n} \binom{s_3 - s_2 - 1}{s_3 - s_2 - n}} \tag{9}$$

Proof. The trivariate joint distribution of $[(X_{11} X_{21} X_{31}), (X_{12} X_{22} X_{32}), ..., (X_{1n} X_{2n} X_{3n})]$ belongs to the exponential family and $(S_1, S_2 \text{ and } S_3)$ is sufficient and complete for Eq. (1). Hence by using Rao-Blackwell theorem we can obtain the UMVUE of

$$p_1^{a_1} p_2^{a_2} p_3^{a_3} (1 - p_1)^{b_1} (1 - p_2)^{b_2} (1 - p_3)^{b_3}.$$

Let $\phi(s_1, s_2, s_3)$ be the UMVUE of $p_1^{a_1} p_2^{a_2} p_3^{a_3} (1 - p_1)^{b_1} (1 - p_2)^{b_2} (1 - p_3)^{b_3}$

$$\begin{aligned} E(\phi(s_1, s_2, s_3)) &= \sum_{s_1=n}^{\infty} \sum_{s_1=s_2+n}^{\infty} \sum_{s_3=s_2+n}^{\infty} \phi(s_1, s_2, s_3) \binom{s_3 - s_2 - 1}{s_3 - s_2 - n} \binom{s_2 - s_1 - 1}{s_2 - s_1 - n} \binom{s_1 - 1}{s_1 - n} \\ &\quad \times p_1^n p_2^n p_3^n (1 - p_1)^{s_1 - n} (1 - p_2)^{s_2 - s_1 - n} (1 - p_3)^{s_3 - s_2 - n}; \\ &= p_1^{a_1} p_2^{a_2} p_3^{a_3} (1 - p_1)^{b_1} (1 - p_2)^{b_2} (1 - p_3)^{b_3} \end{aligned} \tag{10}$$

Hence

$$\begin{aligned} & \sum_{s_1=n}^{\infty} \sum_{s_2=s_1+n}^{\infty} \sum_{s_3=s_2+n}^{\infty} \phi(s_1, s_2, s_3) \binom{s_3 - s_2 - 1}{s_3 - s_2 - n} \binom{s_2 - s_1 - 1}{s_2 - s_1 - n} \binom{s_1 - 1}{s_1 - n} \\ & \times p_1^{n-a_1} p_2^{n-a_2} p_3^{n-a_3} (1 - p_1)^{s_1 - n - b_1} (1 - p_2)^{s_2 - s_1 - n - b_2} (1 - p_3)^{s_3 - s_2 - n - b_3} = 1; \end{aligned} \tag{11}$$

Therefore

$$\phi(s_1, s_2, s_3) = \frac{\binom{s_1 - b_1 - a_1 - 1}{s_1 - b_1 - n} \binom{s_2 - s_1 - b_2 - a_2 - 1}{s_2 - s_1 - b_2 - n} \binom{s_3 - s_2 - b_3 - a_3 - 1}{s_3 - s_2 - b_3 - n}}{\binom{s_1 - 1}{s_1 - n} \binom{s_2 - s_1 - 1}{s_2 - s_1 - n} \binom{s_3 - s_2 - 1}{s_3 - s_2 - n}} \tag{12}$$

$\square$

Particular Cases

1. $a_1 = a_2 = a_3 = 1$ and $b_1 = b_2 = b_3 = 0$

$$\begin{aligned} \widehat{p_1 p_2 p_3} &= \frac{\binom{s_1 - 2}{s_1 - n} \binom{s_2 - s_1 - 2}{s_2 - s_1 - n} \binom{s_3 - s_2 - 2}{s_3 - s_2 - n}}{\binom{s_1 - 1}{s_1 - n} \binom{s_2 - s_1 - 1}{s_2 - s_1 - n} \binom{s_3 - s_2 - 1}{s_3 - s_2 - n}} \\ &= \frac{(n - 1)^3}{(s_1 - 1)(s_2 - s_1 - 1)(s_3 - s_2 - 1)} \end{aligned} \tag{13}$$

2. $a_1 = 1$ and $a_2 = a_3 = b_1 = b_2 = b_3 = 0$

$$\hat{p_1} = \frac{\binom{s_1 - 2}{s_1 - n}}{\binom{s_1 - 1}{s_1 - n}} = \frac{n - 1}{s_1 - 1} \tag{14}$$

3. $a_2 = 1$ and $a_1 = a_3 = b_1 = b_2 = b_3 = 0$

$$\hat{p_2} = \frac{\binom{s_2 - s_1 - 2}{s_2 - s_1 - n}}{\binom{s_2 - s_1 - 1}{s_2 - s_1 - n}} = \frac{n - 1}{s_2 - s_1 - 1} \tag{15}$$

4. $a_3 = 1$ and $a_1 = a_2 = b_1 = b_2 = b_3 = 0$

$$\widehat{p_3} = \frac{\binom{s_3 - s_2 - 2}{s_3 - s_2 - n}}{\binom{s_3 - s_2 - 1}{s_3 - s_2 - n}} = \frac{n-1}{s_3 - s_2 - 1} \tag{16}$$

5. $a_1 = b_1 = 1$ and $a_2 = a_3 = b_2 = b_3 = 0$

$$\widehat{p_1 (1 - p_1)} = \frac{\binom{s_1 - 3}{s_1 - n - 1}}{\binom{s_1 - 1}{s_1 - n}} = \frac{(n-1)(s_1 - n)}{(s_1 - 1)(s_1 - 2)} \tag{17}$$

Similarly it is possible to obtain UMVUE for various combinations of $p_1^{a_1} p_2^{a_2} p_3^{a_3} (1-p_1)^{b_1} (1-p_2)^{b_2} (1-p_3)^{b_3}$ for different values of a_1, a_2, a_3, b_1, b_2 and b_3.

Theorem 2.2. The UMVUE in the multivariate case of

$$\prod_{i=1}^{k} p_i^{a_i} (1 - p_i)^{b_i} = \prod_{i=1}^{k} \frac{\binom{s_i - s_{i-1} - b_i - a_i - 1}{s_i - s_{i-1} - b_i - n}}{\binom{s_i - s_{i-1} - 1}{s_i - s_{i-1} - n}}; \ s_i = \sum_{j=1}^{n} X_{i_j}, \ s_0 = 0 \tag{18}$$

Proof is similar to proof of Theorem 2.1

3. MAXIMUM LIKELIHOOD ESTIMATOR (MLE)

In the earlier section we have obtained an estimator based on the criteria of unbiasedness and minimum variance. We now look at another very popular principle used namely method of maximum likelihood to obtain the estimators of the functions of the parameters. These shall be compared with the corresponding estimators obtained by UMVUE in order to study their efficiency. The likelihood function based on n systems put on test strictly under the same conditions will be

$$L = \prod_{i=1}^{n} (1 - p_1)^{x_{1_i} - 1} p_1 (1 - p_2)^{x_{2_i} - x_{1_i} - 1} p_2 \ldots (1 - p_k)^{x_{k_i} - x_{k_{i-1}} - 1} p_k \\ = (1 - p_1)^{s_1 - n} p_1^n (1 - p_2)^{s_2 - s_1 - n} p_2^n \ldots (1 - p_k)^{s_k - s_{k-1} - n} p_k^n \tag{19}$$

Taking log and differentiating w.r.t. p_i, $i = 1, 2 \ldots, k$, the MLEs are obtained as

The MLE of p_i; $i = 1, 2, \ldots, k$ is

$$\widehat{p_i} = \frac{n}{s_i - s_{i-1}}; s_0 = 0 \tag{20}$$

By invariance property, MLE of $\prod_{i=1}^{k} p_i^{a_i} (1 - p_i)^{b_i}$ is

$$\prod_{i=1}^{k} p_i^{a_i} \widehat{(1 - p_i)}^{b_i} = \prod_{i=1}^{k} \left(\frac{n}{s_i - s_{i-1}} \right)^{a_i} \left(\frac{s_i - s_{i-1} - n}{s_i - s_{i-1}} \right)^{b_i} \tag{21}$$

4. MODIFIED MAXIMUM LIKELIHOOD ESTIMATOR (MODIFIED MLE)

In the earlier two sections we have applied two procedures to obtain the estimators and either of them could be good. We now try to improve on the MLE by reducing the bias and thus derive a modified estimator namely modified MLE. We have further shown that this modified MLE is better than the UMVUE. Hence to derive a modification that reduces the bias and the MSE of the MLE of p_i, $i = 1,2,3 \ldots, k$, we apply the Taylor Series two-parameter expansion to

$$\widehat{p_i} = \phi (s_{i-1}, s_i) = \frac{n}{s_i - s_{i-1}}; \quad s_0 = 0, \quad i = 1, 2, \ldots, k. \tag{22}$$

$$\phi\left(s_{i-1}, s_i\right) = \phi\left(\mu_{s_{i-1}}, \mu_{s_i}\right) + \left(s_{i-1} - \mu_{s_{i\grave{a}1}}\right)\frac{d\phi}{ds_{i-1}}|\left(s_j = \mu_{s_j}\right) + \left(s_i - \mu_{s_i}\right)\frac{d\phi}{ds_i}|\left(s_j = \mu_{s_j}\right)$$

$$+ \frac{\left(\left(s_{i-1} - \mu_{s_{i-1}}\right)\right)^2}{2!}\frac{d^2\phi}{ds_{i-1}^2}|\left(s_j = \mu_{s_j}\right) + \frac{\left(s_i - \mu_{s_i}\right)^2}{2!}\frac{d^2\phi}{ds_i^2}|s_j = \mu_{s_j}$$

$$+ \frac{2\left(s_{i-1} - \mu_{s_{i-1}}\right)\left(s_i - \mu_{s_i}\right)}{2!}\frac{d^2\phi}{ds_{i-1}ds_i}|\left(s_j = \mu_{s_j}\right) + \dots \tag{23}$$

where $j = 1, 2, \dots, k$ and

$$\mu_{s_i} = E(S_i) = \sum_{i=1}^{n}\frac{n}{p_i}, i = 1, 2, \cdots, k. \tag{24}$$

On substition of Eqs. (21) in (22) we obtain

$$\widehat{p}_i = p_i + \left(s_{i-1} - \sum_{j=1}^{i-1}\frac{n}{p_j}\right)\frac{p_i^2}{n} + \left(s_i - \sum_{j=1}^{i}\frac{n}{p_j}\right)\frac{-p_i^2}{n} + \frac{\left(s_{i-1} - \sum_{j=1}^{i-1}\frac{n}{p_j}\right)^2}{2!}\frac{2p_i^3}{n^2}$$

$$+ \frac{\left(s_i - \sum_{j=1}^{i}\frac{n}{p_j}\right)^2}{2!}\frac{2p_j^3}{n^2} + \frac{\left(s_i - \sum_{j=1}^{i}\frac{n}{p_j}\right)\left(s_{i-1} - \sum_{j=1}^{i-1}\frac{n}{p_j}\right)}{2!}\frac{-2p_i^3}{n^2} + O\left(\frac{1}{n}\right). \tag{25}$$

On taking expectaion of Eq. (24) we obtain

$$E\left(\widehat{p}_i\right) = p_i\left(1 + \frac{1}{n}\right) - \frac{p_i^2}{n} + O\left(\frac{1}{n}\right) \tag{26}$$

We observe that

$\widehat{p}_i$ is an asymptotically unbiased estimator of p_i

Consider a linear function of the MLE of p_i given as

$\widetilde{p}_i = \alpha\widehat{p}_i + \beta$, α and β are constants

Hence $E(\widetilde{p}_i) = \alpha.E(\widehat{p}_i) + \beta$

$$= \alpha\left[p_i\left(1 + \frac{1}{n}\right) - \frac{p_i^2}{n} + O\left(\frac{1}{n}\right)\right] + \beta \tag{27}$$

If the coefficient of p_i is set equal to 1 and the constant term set equal to zero, we obtain aproximate equality between $E(\widetilde{p}_i)$ and p_i.

This gives us $\alpha = \dfrac{n}{n+1}$ and $\beta = 0$.

Therefore we obtain a modified MLE of p_i

$$\widetilde{p}_i = \frac{n^2}{(n+1)\left(s_i - s_{i-1}\right)}; \qquad i = 1, 2, 3, \cdots, k, \qquad s_0 = 0. \tag{28}$$

Since

$$\widetilde{p}_i = \frac{n}{n+1}\widehat{p}_i \tag{29}$$

$$E(\widetilde{p}_i - p_i)^2 = (\frac{n}{n+1})^2 \left(E(\widehat{p}_i - p_i)^2 + \frac{p_i^2}{n^2} - \frac{2p_i E\left(\widehat{p}_i - p_i\right)}{n} + O\left(\frac{1}{n}\right) \right)$$

$$E(\widetilde{p}_i - p_i)^2 = \left(\frac{n}{n+1}\right)^2 \left(E(\widehat{p}_i - p_i)^2 - \frac{p_i^2}{n^2} \right)$$

Thus

$$E\left(\widetilde{p}_i - p_i\right)^2 < (E(\widehat{p}_i - p_i)^2 \tag{30}$$

Thus the MSE of modified MLE of p_i namely $\widetilde{p}_i$ is less than the of the corresponding MLE of p_i i.e. $\widehat{p}_i$, $i = 1,2, ..., k$.

5. CONSISTENCY OF MODIFIED MLE

If we collect a large number of observations, then we can obtain a lot of information about the unknown parameter. We can thus construct an estimator T(X) with a small MSE and we can call it a consistent estimator if $\lim_{n \to \infty} \text{MSE}\left(T\left(X\right)\right) = 0$.

Theorem 5.1. $\widehat{p}_1, \widehat{p}_2, ..., \widehat{p}_k$ are consistent estimators where $\widehat{p}_i$ is the MLE of p_i, $i = 1,2, ..., k$.

where

$$\widehat{p}_i = \frac{n}{s_i - s_{i-1}} ; \quad s_0 = 0 \tag{31}$$

Proof.   From Eq. (26) it is clear that $E\left(\widehat{p}_i\right)$ tends to p_i as n tends to ∞, $i = 1,2, ..., k$. We shall prove by method of induction that $V\left(\widehat{p}_i\right)$ tends to 0 as n tends to ∞.

Let $k = 1$

$$\widehat{p}_1 = \frac{n}{s_1} \tag{32}$$

Applying the Taylor series expansion

$$\widehat{p}_1 = p_1 + \left(s_1 - \frac{n}{p_1} \right) \frac{-p_1^2}{n} + \frac{\left(s_1 - \frac{n}{p_1} \right)^2}{2!} \frac{2p_1^3}{n^2} + O\left(\frac{1}{n}\right) \tag{33}$$

On taking expectaion of $(\widehat{p}_1 - p_1)^2$ we obtain

$$E\left((\widehat{p}_1 - p_1)^2\right) = E\left(s_1 - \frac{n}{p_1} \right)^2 \frac{p_1^4}{n^2} + E\left(s_1 - \frac{n}{p_1} \right)^4 \frac{p_1^6}{n^4} + O\left(\frac{1}{n}\right) \tag{34}$$

$$E\left((\widehat{p}_1 - p_1)^2\right) = \frac{n\left(1 - p_1\right)}{p_1^2} \frac{p_1^4}{n^2} + \frac{n\left(1 - p_1\right)\left[p_1^2 - 3p_1\left(n + 2\right) + 3\left(n + 2\right)\right]}{p_1^4} \frac{p_1^6}{n^4} + O\left(\frac{1}{n}\right) \tag{35}$$

Hence $V\left(\widehat{p}_1\right) = E((\widehat{p}_1 - p_1)^2)$ tends to 0 as n tends to ∞.

Thus $\widehat{p}_1$ is a consistent estimator of p_1

Consider $k = 2$

Here

$$\widehat{p}_2 = \frac{n}{s_2 - s_1} \tag{36}$$

Applying Taylor Series two parameter expansion we obtain

$$
\begin{aligned}
E\left(\left(\widehat{p_2} - p_2\right)^2\right) &= E\left(s_1 - \frac{n}{p_1}\right)^2 \frac{p_2^4}{n^2} + E\left(s_2 - \frac{n}{p_1} - \frac{n}{p_2}\right)^2 \frac{p_2^4}{n^2} + E\left(s_1 - \frac{n}{p_1}\right)^4 \frac{p_2^6}{n^4} \\
&\quad + E\left(s_2 - \frac{n}{p_1} - \frac{n}{p_2}\right)^4 \frac{p_2^6}{n^4} + E\left[\left(s_1 - \frac{n}{p_1}\right)^2 \left(s_2 - \frac{n}{p_1} - \frac{n}{p_2}\right)^2\right] \frac{4p_2^6}{n^4} + O\left(\frac{1}{n}\right)
\end{aligned}
\tag{37}
$$

$$
\begin{aligned}
E\left(\left(\widehat{p_2} - p_2\right)^2\right) &= \left(\frac{n}{p_1^2} - \frac{n}{p_1}\right) \frac{p_2^4}{n^2} + \left(\frac{n}{p_1^2} + \frac{n}{p_2^2} - \frac{n}{p_1} - \frac{n}{p_2}\right) \frac{p_2^4}{n^2} \\
&\quad + \frac{n\left(1 - p_1\right)\left[p_1^2 - 3p_1\left(n + 2\right) + 3\left(n + 2\right)\right]}{p_1^4} \frac{p_1^6}{n^4} + O\left(\frac{1}{n}\right)
\end{aligned}
\tag{38}
$$

Hence $V\left(\widehat{p_2}\right) = E(\left(\widehat{p_2} - p_2\right)^2)$ tends to 0 as n tends to ∞. Thus $\widehat{p_2}$ is a consistent estimator of p_2.

Assume $\hat{p}_{i-1}$ is a consistent estimator of p_{i-1}.

To prove that $\widehat{p}_i$ is a consistent estimator of p_i.

We need to prove that $V\left(\widehat{p}_i\right)$ tends to 0 as n tends to ∞, $i = 2, 3, ..., k$.

$$
\widehat{p}_i = \hat{p}_{i-1} + \frac{n\left[2s_{i-1} - s_i - s_{i-2}\right]}{\left[s_i - s_{i-1}\right]\left[s_{i-1} - s_{i-2}\right]}
\tag{39}
$$

To obtain $E(\left(\widehat{p}_i - \hat{p}_{i-1}\right)^2)$ we apply the Taylor series expansion to $\left(\frac{n\left[2s_{i-1} - s_i - s_{i-2}\right]}{\left[s_i - s_{i-1}\right]\left[s_{i-1} - s_{i-2}\right]}\right)^2$ and then take the expectation. Hence

$$
E\left(\frac{n\left[2s_{i-1} - s_i - s_{i-2}\right]}{\left[s_i - s_{i-1}\right]\left[s_{i-1} - s_{i-2}\right]}\right)^2 = \left(p_i - p_{i-1}\right)^2 + O\left(\frac{1}{n}\right)
$$

$$
E\left(\left(\widehat{p}_i - \hat{p}_{i-1}\right)^2\right) = \left(p_i - p_{i-1}\right)^2 + O\left(\frac{1}{n}\right)
$$

Thus as n tends to ∞, $E(\left(\widehat{p}_i - \hat{p}_{i-1}\right)^2)$ tends to $\left(p_i - p_{i-1}\right)^2$

We now consider

$$
V\left(\widehat{p}_i - \hat{p}_{i-1}\right) = E(\widehat{p}_i - \hat{p}_{i-1})^2 - (E\left(\widehat{p}_i - \hat{p}_{i-1}\right))^2
$$

$$
V\left(\widehat{p}_i - \hat{p}_{i-1}\right) = E(\widehat{p}_i - \hat{p}_{i-1})^2 - (E\left(\widehat{p}_i\right) - E\left(\hat{p}_{i-1}\right))^2
$$

Thus from Eq. (39) and since $\widehat{p}_i$ is asymptotically unbiased from Eq. (26) we can conclude that $V\left(\widehat{p}_i - \hat{p}_{i-1}\right)$ tends to 0 as n tends to ∞.

But

$$
V\left(\widehat{p}_i - \hat{p}_{i-1}\right) = V\left(\widehat{p}_i\right) + V\left(\hat{p}_{i-1}\right) - 2Cov\left(\hat{p}_{i-1}, \hat{p}_i\right)
\tag{40}
$$

The $Cov\left(\hat{p}_{i-1}, \hat{p}_i\right)$ can also be shown as tending to 0 as n tends to ∞ by applying the Taylor series expansion.

Thus, since $V\left(\widehat{p}_i - \hat{p}_{i-1}\right)$, $Cov\left(\hat{p}_{i-1}, \hat{p}_i\right)$ and $V\left(\hat{p}_{i-1}\right)$ all tend to 0 as n tends to ∞, we conclude that $V\left(\widehat{p}_i\right)$ also tends to 0 as n tends to ∞.

Hence $\widehat{p}_i$, the MLE of p_i, is a consistent estimator of p_i, $i = 1, 2, \cdots, k$.

Note: The MSE of modified MLE is less than MSE of MLE from Eq. (29).

From Eq. (28) we can conclude that the modified MLE, $\tilde{p}_i$ is a consistent estimator of p_i, $i = 1, 2, \cdots, k$.   □

6. CONCLUSION AND COMPARISION OF ESTIMATORS

We have observed in the earlier section that an improvement over the MLE is the modified MLE. We now have two estimators the UMVUE and the modified MLE. Our objective is to compare both the estimators with respect to efficiency. We make a comparative study of the two based on the determinant of the variance covariance matrix also called as the generalised variance (GV). We consider the trivariate case namely $k = 3$. The variances and covariances of the UMVUE and modified MLE of the parameters can be obtained as below. Consider the case when $k = 3$.

Variance of UMVUE of p_i, $i = 1, 2, 3$ where $s_0 = 0$ is

$$\sum_{s_1=n}^{\infty} \sum_{s_2=s_1+n}^{\infty} \sum_{s_3=s_2+n}^{\infty} \left[\left(\frac{\binom{s_i - s_{i-1} - 2}{s_i - n}}{\binom{s_i - s_{i-1} - 1}{s_i - s_{i-1} - n}} \right)^2 P(s_1 s_2 s_3) \right] - \left(p_i \right)^2 \tag{41}$$

Covariance of the UMVUEs of p_i and p_j, $i, j = 1, 2, 3$ and $s_0 = 0$ is

$$\sum_{s_1=n}^{\infty} \sum_{s_2=s_1+n}^{\infty} \sum_{s_3=s_2+n}^{\infty} \left[\left(\frac{\binom{s_i - s_{i-1} - 2}{s_i - n}}{\binom{s_i - s_{i-1} - 1}{s_i - s_{i-1} - n}} \right) \left(\frac{\binom{s_j - s_{j-1} - 2}{sj - n}}{\binom{s_j - s_{j-1} - 1}{s_j - s_{j-1} - n}} \right) P(s_1 s_2 s_3) \right] - \left(p_i.p_j \right) ; \tag{42}$$

Variance of modified MLE of p_i when $k = 3$ is

$$\sum_{s_1=n}^{\infty} \sum_{s_2=s_1+n}^{\infty} \sum_{s_3=s_2+n}^{\infty} \left(\frac{n^2}{(n+1)(s_i - s_{i-1})} \right)^2 P(s_1 s_2 s_3)$$
$$- \left(\sum_{s_1=n}^{\infty} \sum_{s_2=s_1+n}^{\infty} \sum_{s_3=s_2+n}^{\infty} \left(\frac{n^2}{(n+1)(s_i - s_{i-1})} \right)^2 P(s_1 s_2 s_3) \right)^2 \tag{43}$$

where $s_0 = 0; \quad i = 1, 2, 3$

Covariance of modified MLE of p_i and p_j, $i, j = 1, 2, 3$ and $s_0 = 0$ is

$$\sum_{s_1=n}^{\infty} \sum_{s_2=s_1+n}^{\infty} \sum_{s_3=s_2+n}^{\infty} \left(\frac{n^2}{(n+1)(s_i - s_{i-1})} \right) \left(\frac{n^2}{(n+1)(s_j - s_{j-1})} \right)^2 P(s_1 s_2 s_3)$$
$$- \left(\sum_{s_1=n}^{\infty} \sum_{s_2=s_1+n}^{\infty} \sum_{s_3=s_2+n}^{\infty} \left(\frac{n^2}{(n+1)(s_i - s_{i-1})} \right) \left(\frac{n^2}{(n+1)(s_j - s_{j-1})} \right)^2 P(s_1 s_2 s_3,) \right)^2$$

The determinant of the variance covariance matrix is calculated and are compared for the two estimators in the graphs below for a range of values of the parameters p_1, p_2 and p_3.

It can be observed from the graphs in Figs. 1 and 2 that the generalised variance of the modified MLE is less than the corresponding GV of the UMVUE for numerical values of p_1, p_2 and p_3 ranging from 0.1 to 0.9. Thus we obtain a new and better estimator called modified MLE which is consistent. It is an improvement over the MLE and is also more efficient than the UMVUE.

7. AN EXAMPLE FOR *K* = 3

A game of cricket has been considered. When a batsman is out he is replaced by another batsman in the next ball and the game continues. When the replaced batsman is declared out another replacement is sent forth and the game continues. Consider the 2016 season of Cricket's Indian Premium League' IPL 2016'. A total of 17 matches were played by the winning team, Sunrisers Hyderabad, of which 15 were suitable for our study. We have recorded the following details.

Let

X_i denote the ball at which the first player becomes out in the *ith* match.

Y_i denote the ball at which the player who replaces the first batsman becomes out in the *ith* match.

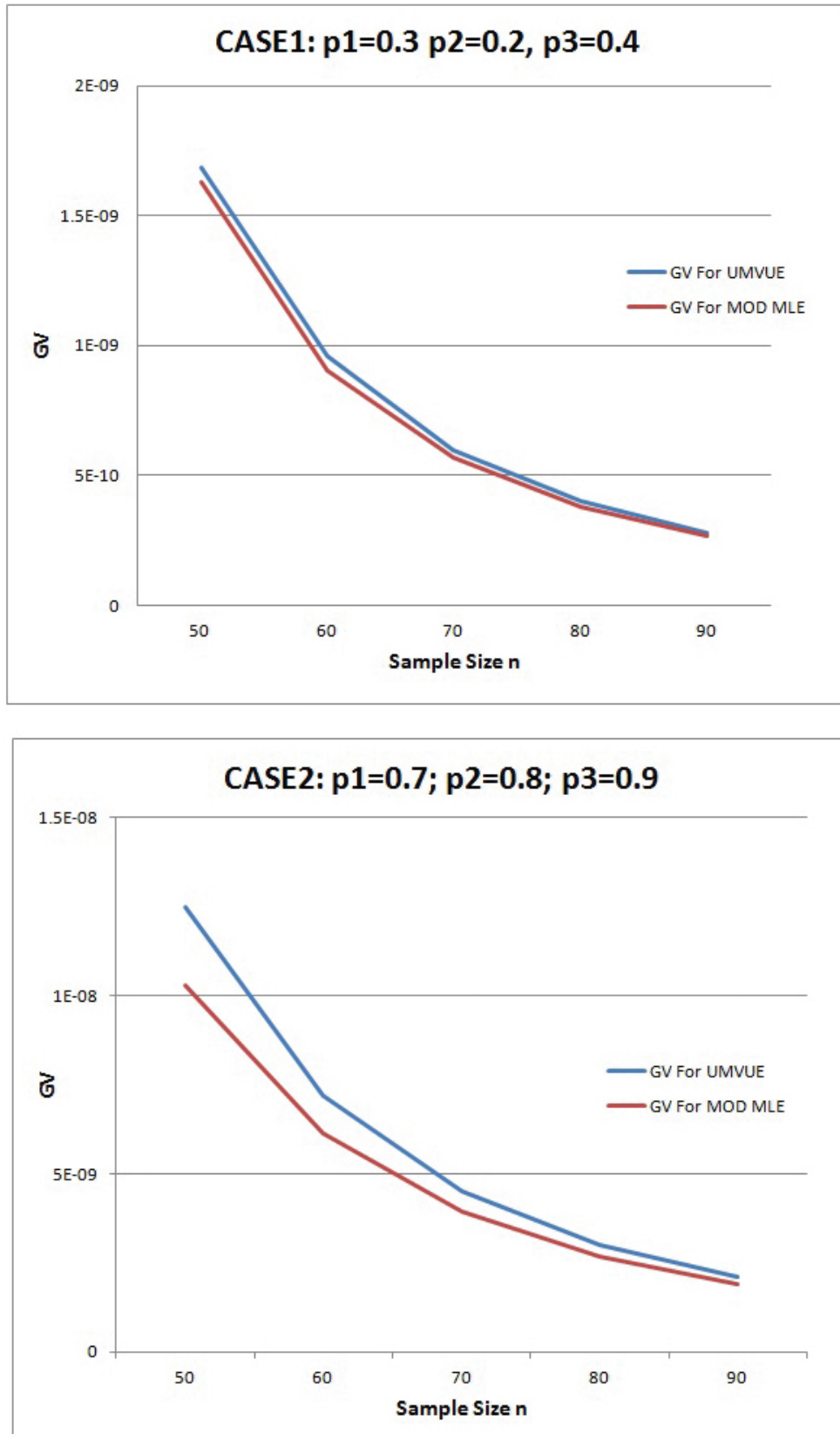Z_i denote the ball at which the player who replaces the second batsman becomes out in the *ith* match, $i = 1, 2, ..., 15$.

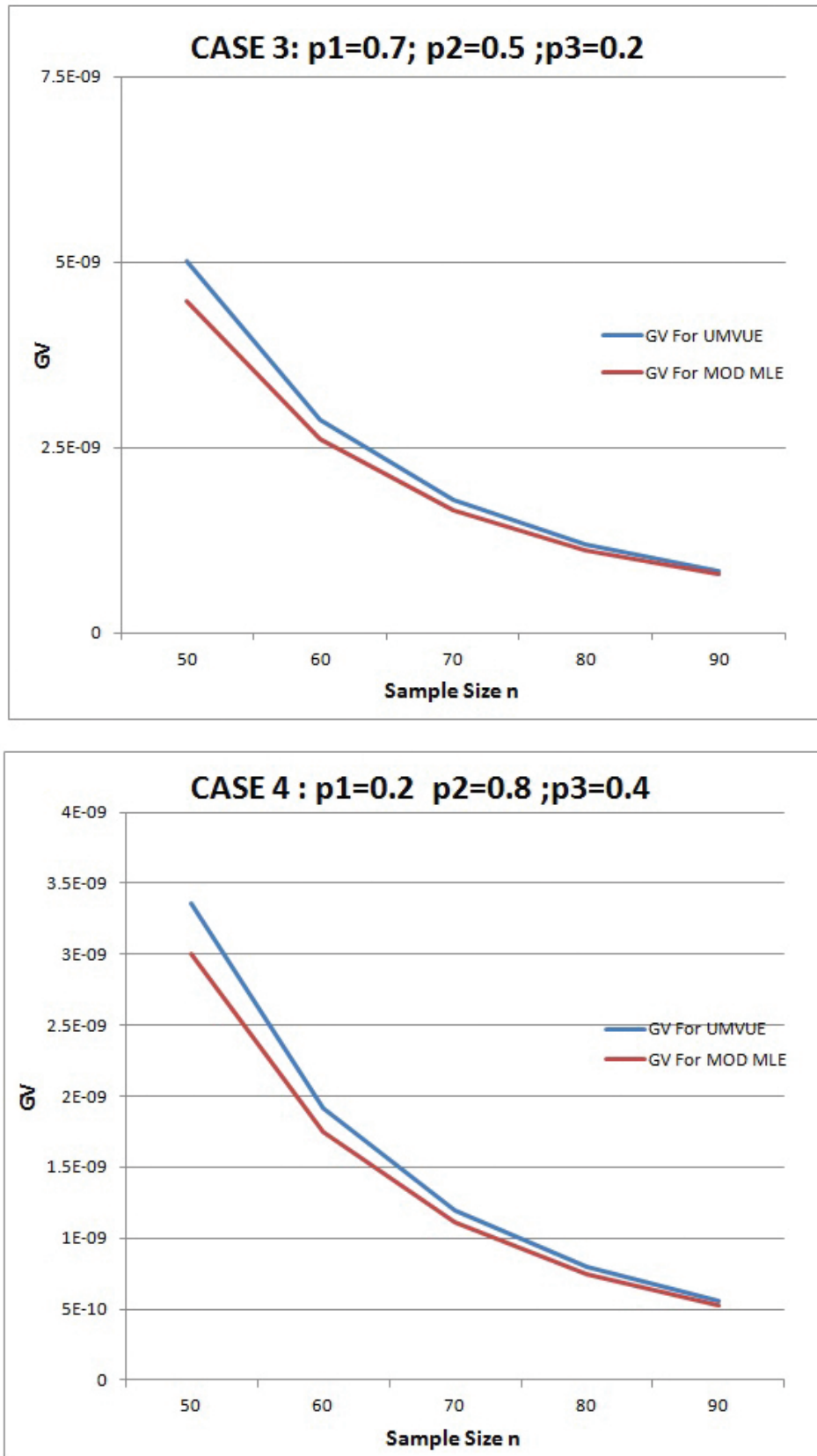**Figure 1** | The generalised variance of the modified MLE

Figure 2 | The generalised variance of the modified MLE

Table 1 | **IPL 2016: Team Hyderabad Sunrisers**

Match No i	Xi	Yi	Zi
1	25	29	35
2	7	16	27
3	4	26	38
4	31	32	52
5	4	18	20
6	11	49	52
7	17	26	42
8	33	38	61
9	14	51	56
10	30	54	56
11	8	9	20
12	16	41	50
13	10	31	61
14	4	10	23
15	25	30	53

Table 2 | **Values of MLE, UMVUE and modified MLE for parameters**

	p_1	p_2	p_3
UMVUE	0.05882	0.063636	0.075676
MLE	0.06276	0.06787	0.0806
Modified MLE	0.05884	0.063631	0.075605

Hence $p_1 = $ P(First player is out), $p_2 = $ P (Second player is out) and $p_3 = $ P (Third player is out).

Thus for $n = 15$, we obtain $s_1 = 239$, $s_2 = 460$ and $s_3 = 646$. The MLE, UMVUE and modified MLE for the following parameters are as shown in Table 2.

From the example of IPL2016, it is observed that the UMVUE, MLE and modified MLE estimates calculated for p_1, p_2 and p_3 are very close to each other.

ACKNOWLEDGEMENT

REFERENCES

1. J. Li, Modeling with Bivariate Geometric Distribution. A dissertation submitted to the Faculty of New Jersey Institute of Technology (NJIT) and Rutgers, The State University of New Jersey-Newark. Department of Mathematical Sciences (NJIT). Department of Mathematics and Computer Science, Rutgers, 2010.
2. F. Downtown, JRSS. B, 32 (1970), 408–417.
3. H. Krishna, P.S. Pundir, Commun. Statist. Theor. Meth. 38 (2009), 1079–1093.
4. U.J. Dixit, S. Annapurna, J. Stat. Theory Appl. 14 (2015), 324–349.
5. A.G. Phatak, M. Sreehari, J. Indian Stat. Assoc. 19 (1981), 141–146.
6. A.W. Marshall, I. Olkin, J. Amer. Statist. Assoc. 80 (1985), 332–338.
7. O.E. Gultekin, I. Bairamov, EGE Univ. J. Fac. Sci. 37(1) (2013), 1–18.
8. R.C. Srivastava, K.S.N. Bagchi, J. Indian Stat. Assoc. 23 (1985), 27–33.
9. R. Vasudeva, G. Srilakshminarayana, Some Properties of Trivariate and Multivariate Geometric Distribution. Paper presented at the International Conference on Multivariate Statistical Methods in 21st Century. Indian Statistical Institute, Kolkata, India, 2006.
10. M. Sreehari, R. Vasudeva, Metrika. 75 (2012), 271–286.
11. J.D. Esary, A.W. Marshall,, Calhoun. 3 (1973).
12. A.W. Marshall, I. Olkin, J. Am. Stat. Assoc. 62 (1967), 30–44.