# ESTIMATION OF PROCESS AVERAGE IN ATTRIBUTE SAMPLING PLANS

P.D. Ghangurde
Household Surveys Development Division

Exact formulae for bias and mean square error of an estimator of process average in single sampling with rectification for finite lots are obtained. Efficiency of the estimator as compared to an unbiased estimator based on the first sample is obtained for a number of values of lot size, sample size, acceptance number and process average used in sampling plans in quality control of data processing.

## 1. INTRODUCTION

In single sampling with rectification plans a sample of fixed size is drawn from a lot. If the number of defectives in the sample is less than or equal to c, an acceptance number, the lot is accepted; otherwise the lot is completely verified and rectified. The sample size and acceptance number are determined to minimize average amount of inspection for a given lot size and proportion of defectives [3]. These plans are appropriate in situations where inspection is nondestructive and rectification is not costly.

Maximum likelihood estimators (m.l.e.) of process average for single and double attribute sampling plans have been given in the literature assuming constant process average and large lot sizes (see e.g. [1] and [4]). Formulae for asymptotic variance of the m.l.e. are also obtained under these assumptions. However, there are many industrial processes in which process average can change even when the process is in control [5] and sampling plans have to be altered according to changes in the process average.

The estimator proposed in this paper is appropriate in situations in which fraction defective and lot size could vary considerably from lot to lot. Exact expressions for bias and variance of the estimator can be easily derived.

## 2. ESTIMATOR OF FRACTION DEFECTIVE

Consider a lot of size N and proportion of defectives P from which a random sample of size $n_1$ is drawn without replacement. If the number of defectives, $x_1 \leq c$, the acceptance number, the lot is accepted. If $x_1 > c$ the lot is completely verified and hence P can be determined without any sampling error. An estimator of P can be defined as

$$e_1 = \frac{x_1}{n_1} \alpha + P(1 - \alpha), \tag{2.1}$$

where $\alpha$ is a random variable defined by

$$\alpha = \begin{cases} 1 & \text{if } x_1 \leq c, \\ 0 & \text{if } x_1 > c. \end{cases} \tag{2.2}$$

The estimator $e_1$ is of the same form as m.l.e. in double sampling scheme discussed in section 4 (where N is assumed infinite) when $n_2$, the second sample size, tends to infinity.

## 3. BIAS AND MEAN SQUARE ERROR

$$E(e_1) = \frac{1}{n_1} E(x_1 | x_1 \leq c)\phi + (1 - \phi)P,$$

where

$$\phi = P[x_1 \leq c]. \tag{3.1}$$

Hence,

$$\begin{aligned} \text{Bias } (e_1) &= E(e_1) - P \\ &= \phi\left[ \frac{E(x_1 | x_1 \leq c)}{n_1} - P\right] . \end{aligned} \tag{3.2}$$

Let $L(> 1)$ lots with known sizes, $N_i$, be inspected by a single sampling with rectification plan $(n_1, c)$. The process average can be defined as

$$\bar{P} = \sum_{i=1}^{L} \pi_i P_i ,$$

where

$\pi_i$ = known proportion of ith lot size in the total items in L lots,

$P_i$ = fraction defective of ith lot,

and $\sum_{i=1}^{L} \pi_i = 1.$

The estimator $e_1$ is defined as

$$e_1 = \sum_{i=1}^{L} \pi_i \left[ \frac{x_{1i}\ \alpha_i + X_{1i}(1 - \alpha_i)}{n_1\ \alpha_i + N_i(1 - \alpha_i)} \right], \tag{3.3}$$

where

$$\alpha_i = \begin{cases} 1 & \text{if } x_{1i} \leq c, \\ 0 & \text{otherwise} \end{cases}$$

$x_{1i}$ = defectives in sample of size $n_1$ from ith lot,

$X_{1i}$ = total number of defectives in ith lot,

$N_i$ = size of ith lot.

Hence,

$$\text{Bias } (e_1) = \sum_{i=1}^{L} \pi_i\ \phi_i \left[ \frac{E(x_{1i} | x_{1i} \leq c)}{n_1} - P_i \right] \tag{3.4}$$

where

$$\phi_i = P[x_{1i} \leq c].$$

In order to study behavior of Bias $(e_1)$ for single lot we consider Poisson approximation.

Let $\lambda = n_1 P$ then $\phi = e^{-\lambda} \sum_{x_1=0}^{c} \lambda^{x_1}/x_1!$ and

$$E(x_1|x_1 \leq c) = \frac{e^{-\lambda}}{\phi} \sum_{x_1=1}^{c} \lambda^{x_1}/(x_1-1)! \ .$$

The bias expression in (3.2) takes the form Bias $(e_1) = - e^{-\lambda} \lambda^{c+1}/n_1 c!$. Relative Bias $(e_1)$ can be defined as Bias $(e_1)/P$ and is given by Relative Bias $(e_1) = - e^{-\lambda} \lambda^c/c!$. Thus Bias $(e_1)$ is negative and for a given optimum sampling plan $(n_1, c)$ obtained for certain process average P the absolute value of Relative Bias $(e_1)$ is a monotonic increasing function of P for $P < \frac{c}{n_1}$ and a monotonic decreasing function of P for $P > \frac{c}{n_1}$. Though a plan $(n_1, c)$ is determined to give minimum inspection for certain process average P, the fraction defective of a lot could be different from P and hence study of behavior of bias as a function of P for a given plan is of practical importance.

Table 1 gives numerical values of Bias $(e_1)$ and Relative Bias $(e_1)$ obtained by using binomial probabilities in formula (3.2) for single lot for various values of P, N and plans $(n_1, c)$ used in quality control of data capture by keypunch and key-edit. The values of $n_1$ and c are the optimum ones giving minimum average inspection for given lot size N and fraction defective P in single sampling with rectification assuring 3% AOQL [3]. The binomial probabilities tabulated in [6] are used in the calculations. It can be seen that absolute value of Relative Bias $(e_1)$ decreases as $n_1$ and c are increased for a given P.

The tabulated values of Bias $(e_1)$ for various plans show the extent of bias of $e_1$ for various plans and values of P. The bias of the estimator in (3.3) can be estimated from rejected lots by

$$\hat{\text{Bias}}\ (e_1) = \sum_{i=1}^{L} \pi_i (1 - \alpha_i)(P_i - \frac{x_{1i}}{n_{1i}}). \tag{3.5}$$

For single lot

$$\begin{aligned} V(e_1) &= E[e_1 - E(e_1)]^2 \\ &= E[\frac{x_1}{n_1}\ \alpha + P(1 - \alpha) - \frac{1}{n_1}\ E(x_1 | x_1 \leq c)\phi - P(1 - \phi)]^2 \\ &= E[\frac{1}{n_1}\ (x_1\alpha - E(x_1|x_1 \leq c)\phi) - P(\alpha - \phi)]^2 \\ &= \frac{1}{n_1^2}\ E[x_1\alpha - E(x_1|x_1 \leq c)\phi]^2 + P^2E(\alpha - \phi)^2 \\ &\quad - \frac{2P}{n_1} . E[(x_1\alpha - E(x_1|x_1 \leq c)\phi)(\alpha - \phi)] \\ &= \frac{\phi}{n_1^2}\ [E(x_1^2|x_1 \leq c) - E^2(x_1|x_1 \leq c)\phi] + P^2\phi(1 - \phi) \\ &\quad - \frac{2P}{n_1} . E(x_1|x_1 \leq c)\phi(1 - \phi) \qquad (3.6) \\ &= \frac{\phi}{n_1^2}\ [E(x_1^2|x_1 \leq c) - E^2(x_1|x_1 \leq c)\phi] - \phi(1 - \phi)P^2 \\ &\quad - 2P(1 - \phi)[\text{Bias}\ (e_1)] \end{aligned}$$

The mean square error of $e_1$ is given by

$$\text{MSE}\ (e_1) = V(e_1) + [\text{Bias}\ (e_1)]^2 \tag{3.7}$$

This is the expression used in numerical efficiency comparison of $e_1$ with $e_1'$ (see Table 1) which is based on first sample only and given by $e_1' = x_1/n_1$. Its variance for finite lot sizes is given by

$$V(e_1') = \frac{N-n_1}{N-1}\ .\ \frac{P(1-P)}{n_1}\ . \tag{3.8}$$

Though finite population correction is used in $V(e_1')$, the MSE $(e_1)$ is obtained by using binomial probabilities from Tables in [6]. Since $N/n_1 > 10$ for most entries in Table 1, the binomial approximation to hypergeometric probabilities is expected to be very close.

## 4. ESTIMATION IN DOUBLE SAMPLING

Consider a lot of size N from which a random sample of size $n_1$ is drawn without replacement. If the number of defectives, $x_1 \leq c$, the acceptance number, the lot is accepted. If $x_1 > c$ a second sample of size $n_2$ is drawn and the number of defectives, $x_2$, is observed. In practice, in double sampling, if $c < x_1 \leq c_1$, where $c_1$ is another acceptance number, a second sample of size $n_2$ is drawn. The lot is accepted if $x_1 + x_2 \leq c_1$ and rejected if $x_1 + x_2 > c_1$. The above double sampling scheme is considered for simplicity, since the purpose is to obtain the results of single sampling with rectification for large lots as a limiting case of double sampling when $n_2 \to \infty$.

The likelihood of the sample, assuming large N, is given by

$$L(n_1, n_2, x_1, x_2, P) = \begin{cases} \binom{n_1}{x_1} P^{x_1} (1-P)^{n_1-x_1}, & \text{if } x_1 \leq c \\ \binom{n_1}{x_1}\binom{n_2}{x_2} P^{x_1+x_2} (1-P)^{n_1+n_2-x_1-x_2} & \text{if } x_1 > c \end{cases}$$

Let $\alpha$ be a random variable defined as in (2.2). Hence

$$L(n_1, n_2, x_1, x_2, P) = \left[\binom{n_1}{x_1} P^{x_1} (1-P)^{n_1-x_1}\right] \left[\binom{n_2}{x_2} P^{x_2} (1-P)^{n_2-x_2}\right]^{(1-\alpha)} .$$

Differentiating L with respect to P the m.l.e. is obtained as

$$e_2 = \frac{x_1 + x_2(1 - \alpha)}{n_1 + n_2(1 - \alpha)} \qquad (4.1)$$

$$E(e_2) = E[\frac{x_1}{n_1} \,|\alpha=1]\phi + E[\frac{x_1 + x_2}{n_1 + n_2} \,|\alpha=0](1 - \phi)$$

$$= \frac{1}{n_1} \; E(x_1|x_1 \leq c)\phi + \frac{1}{(n_1+n_2)} \; [E(x_1|x_1 > c)(1 - \phi) + n_2P(1 - \phi)]$$

$$= \frac{1}{n_1(n_1+n_2)} \; [n_1^2 P + n_2 \; E(x_1|x_1 \leq c)\phi + n_1 \; n_2 \; P(1 - \phi)]$$

since $\qquad n_1P = E(x_1|x_1 \leq c)\phi + E(x_1|x_1 > c) \; (1 - \phi).$

Hence,

$$\text{Bias}(e_2) = E(e_2) - P$$

$$= (\frac{n_2}{n_1 + n_2})\phi \; [\frac{E(x_1|x_1 \leq c)}{n_1} - P]. \qquad (4.2)$$

As $n_2 \to \infty$ (4.2) takes the same form as (3.2).

For L lots assuming the same plan $(n_1, c)$ with the same notations as before

$$e_2 = \sum_{i=1}^{L} \pi_i \; [\frac{x_{1i} + x_{2i}(1 - \alpha_i)}{n_1 + n_2(1 - \alpha_i)}] \qquad (4.3)$$

with obvious meanings for $x_{1i}$, $x_{2i}$, $\alpha_i$, $i = 1, 2, \ldots, L$. Assuming fraction defective for each of L lots to be P, the m.l.e. of P for double sampling can be obtained as

$$e_3 = \frac{\sum_{i=1}^{L} [x_{1i} + x_{2i}(1 - \alpha_i)]}{\sum_{i=1}^{L} [n_1 + n_2(1 - \alpha_i)]} \; . \qquad (4.4)$$

This is the estimate of P generally used in the literature. If the process average of individual lots is known to be different (4.3) is more appropriate than (4.4). The estimator (4.3) takes the form (3.3) for single sampling with rectification. The expression for Bias $(e_3)$ for general L is complicated. For simplicity we consider the case L = 2 and for single sampling with rectification obtain for large lot sizes

$$\text{Bias } (e_3) = \phi_1 \phi_2 \left[ \frac{E(x_{11} | x_{11} \leq c) + E(x_{12} | x_{12} \leq c)}{2n_1} - \pi_1 P_2 - \pi_2 P_1 \right]$$

$$+ (P_2 - P_1)(\phi_1 \pi_1 - \phi_2 \pi_2). \qquad (4.5)$$

For L = 2 and single sampling with rectification Bias $(e_2)$ is obtained as

$$\text{Bias } (e_2) = \pi_1 \phi_1 \left[ \frac{E(x_{11} | x_{11} \leq c)}{n_1} - P_1 \right] + \pi_2 \phi_2 \left[ \frac{E(x_{12} | x_{12} \leq c)}{n_1} - P \right].$$

When $P_1 = P_2$, $\pi_1 = \pi_2 = 1/2$ absolute value of Bias $(e_3)$ is less than that of Bias $(e_2)$. Since for given $(n_1, c)$, $\phi$ decreases as P increases the contribution of second term in (4.5) is positive when $P_1 \neq P_2$ and $\pi_1 = \pi_2 = 1/2$. It seems that absolute value of Bias $(e_3)$ would be lesser than that of Bias $(e_2)$. However, no conclusions can be drawn for the general case of L > 2.

We now obtain variance of $e_2$ for single lot,

$$V(e_2) = V\, E(e_2 | \alpha) + E\, V(e_2 | \alpha)$$

After some algebra and reduction we obtain

$$V(e_2) = \frac{\phi}{n_1^2} [E(x_1^2|x_1 \leq c) - E^2(x_1|x_1 \leq c)\phi]$$

$$+ \frac{(1 - \phi)}{(n_1 + n_2)^2} [E(x_1^2|x_1 > c) - E^2(x_1|x_1 > c)(1 - \phi)]$$

$$+ \frac{(1 - \phi)\ n_2P[1 - P + n_2P\ \phi]}{(n_1 + n_2)^2}$$

$$+ \frac{2n_2P\ \phi(1 - \phi)\ E(x_1|x_1 > c)}{(n_1 + n_2)^2}$$

$$- \frac{2\ E(x_1|x_1 \leq c)\ n_2P\ \phi(1 - \phi)}{n_1(n_1 + n_2)}$$

$$- \frac{2\ E(x_1|x_1 \leq c)\ E(x_1|x_1 > c)\phi(1 - \phi)}{n_1(n_1 + n_2)} \tag{4.6}$$

As $n_2 \to \infty$ $V(e_2)$ takes the limiting form

$$V(e_2) = \frac{\phi}{n_1^2} [E(x_1^2|x_1 \leq c) - E^2(x_1|x_1 \leq c)\phi]$$

$$+ \phi(1 - \phi)P^2 - \frac{2PE(x_1|x_1 \leq c)\phi(1 - \phi)}{n_1}$$

which is the same as (3.6)

## 5. ACKNOWLEDGMENTS

The author would like to acknowledge the referee for some helpful comments.

Table 1:

c = 1

| P | B | n | $\phi$ | Bias($e_1$) | Relative Bias($e_1$) | MSE($e_1$) | V($e_1'$) | Eff($e_1$) |
|---|---|---|---|---|---|---|---|---|
| .004 | 900 | 26 | .995122 | -.000362 | -.090466 | .000126 | .000149 | 118.089 |
| .010 | 775 | 26 | .972277 | -.001945 | -.194455 | .000241 | .000368 | 153.003 |
| .012 | 600 | 26 | .961318 | -.002662 | -.221843 | .000267 | .000437 | 163.808 |
| .014 | 496 | 26 | .948979 | -.003444 | -.246032 | .000289 | .000504 | 174.459 |
| .016 | 424 | 26 | .935418 | -.004276 | -.267262 | .000309 | .000570 | 184.658 |
| .018 | 361 | 25 | .926016 | -.005028 | -.279357 | .000347 | .000660 | 190.400 |
| .020 | 310 | 25 | .911355 | -.005911 | -.295575 | .000365 | .000723 | 198.360 |
| .022 | 277 | 25 | .895892 | -.006811 | -.309575 | .000382 | .000786 | 205.699 |
| .024 | 250 | 25 | .879735 | -.007717 | -.321527 | .000400 | .000847 | 211.900 |
| .026 | 230 | 24 | .871832 | -.008483 | -.326255 | .000443 | .000949 | 214.370 |
| .028 | 230 | 24 | .855512 | -.009384 | -.335127 | .000462 | .001020 | 220.862 |

c = 2

| P | N | n | $\phi$ | Bias($e_1$) | Relative Bias($e_1$) | MSE($e_1$) | V($e_1'$) | Eff($e_1$) |
|---|---|---|---|---|---|---|---|---|
| .010 | 900 | 42 | .991416 | -.000549 | -.054856 | .000200 | .000225 | 112.412 |
| .012 | 900 | 42 | .985999 | -.000874 | -.072854 | .000227 | .000269 | 118.879 |
| .014 | 900 | 42 | .979009 | -.001280 | -.091442 | .000249 | .000314 | 126.140 |
| .016 | 900 | 42 | .970414 | -.001762 | -.110118 | .000267 | .000358 | 134.104 |
| .018 | 900 | 42 | .960216 | -.002313 | -.128475 | .000282 | .000402 | 142.662 |
| .020 | 750 | 42 | .948450 | -.002924 | -.146190 | .000294 | .000441 | 150.227 |
| .022 | 637 | 42 | .935178 | -.003586 | -.163010 | .000304 | .000479 | 157.810 |
| .024 | 563 | 42 | .920478 | -.004290 | -.178743 | .000312 | .000517 | 165.526 |
| .026 | 500 | 41 | .909603 | -.004907 | -.188730 | .000333 | .000568 | 170.374 |
| .028 | 449 | 41 | .893128 | -.005657 | -.202020 | .000341 | .000605 | 177.172 |
| .030 | 409 | 41 | .875552 | -.006420 | -.214010 | .000349 | .000640 | 183.453 |

RESUME

On définit les formules exactes pour calculer le biais et l'erreur quadratique moyenne d'un estimateur de moyenne du processus dans un échantillonnage unique avec correction pour les lots finis. L'efficacité de l'estimateur par rapport à un estimateur non biaisé fondé sur le premier échantillon s'obtient pour un certain nombre de valeurs de la taille des lots de l'échantillon, du nombre d'acceptation et de la moyenne du processus utilisées dans les schémas d'échantillonnage servant au contrôle qualitatif du traitement des données.

REFERENCES

[1] Cohen, A.C. (1970), "Curtailed Attribute Sampling", Technometrics, 12, 295-298.

[2] Duncan, A.J. (1965), "Quality Control and Industrial Statistics", R.D. Irwin.

[3] Dodge and Romig, (1959), "Sampling Inspection Tables, Single and Double Sampling", Second Edition, Page 60, John Wiley and Sons, Inc., New York.

[4] Phatak, A.G. and Bhatt, M.M., (1967), "Estimation of Fraction Defective in Curtailed Sampling Plans", Technometrics, 9, 219-288.

[5] Wetherill, G.B., (1969), "Sampling Inspection and Quality Control", Page 60, Methuen and Co. Ltd., London.

[6] Weintraub, S., (1963), "Tables of the Cumulative Binomial Probability Distribution for Small Values of P", Free Press of Glencoe, New York.