Inspection Errors in Link Sampling

by

Samuel Kotz
University of Maryland
College Park

Norman L. Johnson
University of North Carolina
Chapel Hill

and

Robert N. Rodriguez
SAS Institute, Cary, North Carolina

December 1983

ABSTRACT

Results of an investigation into the sensitivity to two types of inspection error of link sampling procedures described by Harishchandra and Srivenkataramana are reported. Relevant compound distributions are derived. Some comparisons with results obtained in a similar investigation for standard double sampling are also given.

Key Words and Phrases: Quality control; Compound distributions; Binomial distribution; Hypergeometric distribution; Multivariate hypergeometric distribution; Two-stage sampling.

1. Introduction

In two stage acceptance sampling, it is necessary to inspect further items from a lot when the evidence provided by the first sample is inconclusive. Commonly the second sample is twice as large as the first, though this is by no means universal. In order to save sampling costs, sometimes the additional evidence is provided from the results of standard ('first stage') inspection on the immediately preceding and following lots. There is, of course, an implicit assumption that these neighboring lots are of about the same quality as the lot under examination!

In this paper we will derive some formulas for the properties of this kind of acceptance sampling procedure, for inspection by attributes when inspection is not perfect, there being a probability, p, of (correctly) declaring a defective item to be defective, and a probability, p', of (incorrectly) declaring a nondefective item to be defective.

2. Link Sampling

A procedure described by Harishchandra and Srivenkataramana (1982) (referred to as HS in the sequel) is as follows:

Routine sampling takes random samples of size n (without replacement) from lots of size N. Denoting the number of items classified as defective (whether correctly or not) in the i-th lot by Z_i;

if $Z_i \leq a_1$ the lot is accepted

if $Z_i > a_2$ the lot is rejected

if $a_1 < Z_i \leq a_2$, the quantity $Z_{i-1}+Z_i+Z_{i+1}$ - that is, the total number of items found defective in the random samples of size n from the (i-1)-th, i-th, and (i+1)-th lots - is calculated, and if $Z_{i-1}+Z_i+Z_{i+1} \leq a_2'$, the lot is accepted

if $Z_{i-1}+Z_i+Z_{i+1} > a_2'$, the lot is rejected.

The numbers a_1, a_2 and a_2' are integers chosen to give some desired probabilities of acceptance for specified numbers of defectives among the N items in the i-th lot. Commonly, but not necessarily, $a_2 = a_2'$.

3. Distributions

The variables Z_{i-1}, Z_i, Z_{i+1} are mutually independent. Generally

$$Z_j \frown (\text{Bin}(Y_j,p) * \text{Bin}(n-Y_j,p')) \underset{Y_j}{\wedge} \text{Hypg}(n,D_j,N) \qquad (1)$$

where $\frown$ means "is distributed as" ;

$*$ denotes "convolution"'

$\underset{Y}{\wedge}$ is a "mixing" or "compounding" symbol indicating the Y has the distribution following the symbol;

D_j is the number of defective items in the j-th lot;

Y_j is the number of (really) defective items in the sample of size n from the j-th lot;

Bin(g,h) denotes the binomial distribution

$$\Pr[X=x] = \binom{g}{x} h^x (1-h)^{g-x} \quad (0 < h < 1;\ x=0,1,\ldots,g);$$

and Hypg (g,h,k) denotes the hypergeometric distribution

$$\Pr[X=x] = \binom{h}{x}\binom{k-h}{g-x}/\binom{k}{g} \quad (\max(0,g-k+h) \leq x \leq \min(g,h))$$

4. Calculation of Acceptance Probabilities

From (1) we can obtain the explicit expression

$$\Pr[Z_j = z] = \frac{1}{\binom{N}{n}} \sum_y \binom{D_j}{y}\binom{N-D_j}{n-y} \sum_{u=0}^{y} \binom{y}{u}\binom{n-y}{z-u} p^u p'^{z-u} (1-p)^{y-u} (1-p')^{n-y-z+n} \tag{2}$$

$$= P(z|D_j), \text{ say.} \qquad (0 \le z \le n)$$

The probability of acceptance of the i-th lot is

$$\Pr[Z_i \le a_1] + \Pr[(a_1 < Z_i \le a_2) \quad (Z_{i-1}+Z_i+Z_{i+1} \le a_2')]$$

$$= \sum_{z_i \le a_1} P(z_i|D_i) + \sum_{a_1 < z_i \le a_2} P(z_i|D_i) \sum\sum_{z_{i-1}+z_{i+1} \le a_2'-z_i} P(z_{i-1}|D_{i-1}) P(z_{i+1}|D_{i+1}) \tag{3}$$

For the link sampling method to be useful, it is necessary that D_{i-1}, D_i and D_{i+1} do not differ too greatly. In the case of binomial sampling (corresponding to $N \to \infty$, the case considered in HS), _if_ the lot proportion defective is the _same_ in all three lots, the acceptance probabilities for link sampling are the _same_ as they would be for regular two-stage sampling (with the same values of a_1, a_2 and a_2') with the second sample (of size 2n) being chosen, when needed, from the i-th lot. A similar result is not valid when N is finite (as it is in this paper), even when $D_{i-1}=D_i=D_{i+1}$ (=D, say) because the convolution

$$\text{Hypg}\ (n,D,N) * \text{Hypg}\ (n,D,N) \tag{4}$$

is not the same as the distribution

$$\text{Hypg.}\ (2n,D,N)$$

(or, indeed, as Hypg (2n,2D,2N).)

5. Partial Link Sampling Procedures

It may sometimes be a drawback, in the link sampling procedure, that it is necessary to wait for the results of inspection of the (i+1)-th lot before reaching a decision on the i-th lot, if the sample from the i-th lot is inconclusive. One way around this difficulty would be to replace $(Z_{i-1}+Z_i+Z_{i+1})$ by $(Z_{i-1}+Z_i)$, which can be calculated immediately. It may be felt that this is straining the assumption of (roughly) constant D values unduly. In order to meet this difficulty HS propose the use of a partial linking sampling procedure in which a second sample of size n (not 2n) is taken from the i-th lot and used in place of the sample from the (i+1)-th lot is reaching a decision. This means that $(Z_{i-1}+Z_i+Z_{i+1})$ is replaced by $(Z_{i-1}+Z_i+Z_i')$, where Z_i' denotes the number of items found to be defective (whether correctly or not) is the second sample from the i-th lot. The acceptance probability is

$$\Pr[Z_i \leq a_1] + \Pr[(a_1 < Z_i \leq a_2) \cap (Z_{i-1}+Z_i+Z_i' \leq a_2')] \tag{5}$$

Now, of course, Z_i and Z_i' are not independent, though Z_{i-1} is independent of (Z_i,Z_i'). The joint distribution of Z_i and Z_i', obtained from equation (4) of Kotz and Johnson (1983) by putting $n_1 = n_2 = n$, is

$$\begin{pmatrix} Z_i \\ Z_i' \end{pmatrix} \sim \begin{pmatrix} \text{Bin}(Y_1,p) * \text{Bin}(n-Y_1,p') \\ \text{Bin}(Y_2,p) * \text{Bin}(n-Y_2,p') \end{pmatrix} \underset{Y_1,Y_2}{\wedge} \text{Mult. Hypg}(D;n,n;N) \tag{6}$$

Here Mult. Hypg (D;n,n;N) is a multivariate hypergeometric distribution, with

$$\Pr[(Y_1=y_1) \cap (Y_2=y_2)] = \binom{n}{y_1}\binom{n}{y_2}\binom{N-2n}{D-y_1-y_2} / \binom{N}{D} \tag{7}$$

$$(0 \leq y_1,y_2 \leq n;\ D-N+2n \leq y_1+y_2 \leq D)$$

$$\left\{\binom{N}{n,n,N-2n} = \frac{N!}{[(n)!]^2(N-2n)!}\right\}$$

The expected number of items inspected in the i-th lot is

$$n\{1+\Pr[a_1 < Z_i \leq a_2]\} = \tfrac{1}{2}[n+n\{1+2\Pr[a_1 < Z_i \leq a_2]\}]$$

while with regular two-stage sampling (with second sample size 2n) it is,

$n\{1+2\Pr[a_1 < Z_i \le a_2]\}$

[HS also describe another modification in which Z_i' replaces Z_{i-1} rather than Z_{i+1}. The analysis is exactly similar to that set out above].

6. Tables

Table 1 gives acceptance probabilities for both link sampling and partial link sampling for $a_1 = 1$, $a_2 = a_2' = 5$ (the values used in HS); and N = 100, 200; n = 20, 50; p = 1, 0.9. 0.75; p' = 0, 0.1 and four sets of values of proportions defective

D_{i-1}/N	D_i/N	D_{i+1}/N
0.05	0.05	0.05
0.1	0.1	0.1
0.05	0.1	0.15
0.15	0.1	0.15

These tables were computed by RNR. The multivariate hypergeometric distribution in (7) can be computed by calculating its logarithm, using the fact that $\ell n(k!) = \ell n\Gamma(k+1)$ and calling a log-gamma subroutine repeatedly. However, this can result in a severe loss of significant digits, even when the log-gamma function is evaluated in double precision.

A better alternative (used in the construction of Table 1) is to re-express (7) as the product of two univariate hypergeometric probabilities:

$$\Pr[(Y_1=y_1) \cap (Y_2=y_2)] = \left[\frac{\binom{D}{y_1+y_2}\binom{N-D}{2n-(y_1+y_2)}}{\binom{N}{2n}}\right]\left[\frac{\binom{y_1+y_2}{y_1}\binom{2n-(y_1+y_2)}{n-y_1}}{\binom{2n}{n}}\right] \tag{8}$$

The equivalence of (7) and (8) is not obvious, but it is easily established algebraically or probabilistically. The advantage of computing (8) rather than (7) is that subroutines for the univariate hypergeometric distribution are available in a number of high-level computing languages.

The values in the tables exhibit the very marked influence of p' on the acceptance probabilities.

Table 2 contains acceptance probabilities for standard double sampling, with the same parameter values ($a_1 = 1$, $a_2 = a_2' = 5$) as in Table 1. The size of the first sample is n = 20 and the second sample (taken from the <u>same</u> (i-th) lot) is n" = 2n = 40. The probability of acceptance is

$$\Pr[Z_i \le a_1] + \Pr[(a_1 < Z_i \le a_2') \cap (Z_i + Z_i'' \le a_2)] \qquad (9)$$

where Z_i'' denotes the number of item judged to be defective in the second sample (size 2n") from the i-th lot. The joint distribution of Y_i and Y_i'', the actual numbers defective in the two samples is the multivariate hypergeometric

$$\Pr[(Y_i = y) \cap Y_i'' = y'')] = \binom{n}{y}\binom{n''}{y''}\binom{N-n'-n''}{D-y-y''}\Big/\binom{N}{D} \qquad (10)$$

$$= \left\{\binom{n+n''}{y+y''}\binom{N-n'-n''}{D-y-y''}\Big/\binom{N}{D}\right\}\left\{\binom{n}{y}\binom{n''}{y''}\Big/\binom{n+n''}{y+y''}\right\} \qquad (10)'$$

[Given Y_i and Y_i'', Z_i and Z_i'' are distributed independently as in (2)

The figures in Table 2 should be compared with those in Table 1 for n = 20 and $D_{i-1}/N = D_i/N = D_{i+1}/N$ (= 0.05, 0.1). As is to be expected the partial link sampling acceptance probabilities fall between the values for link sampling and standard double sampling. It appears that they are closer to the link sampling values than to the standard values. The differences decrease as the lot size increases (and would be zero for infinite lot size).

Acknowledgement

Samuel Kotz's research was supported by the U.S. Office of Naval Research under Contract N0014-81-K-0301.

REFERENCES

Harishchandra, K. and Srivenkataramana, T. (1982). "Link sampling for attributes, Commun. Statist. Theor. Meth., 11, 185-1868.

Kotz, S. and Johnson, N.L. (1984). "Effects of false and incomplete identification of defective items on the reliability of acceptance sampling" (to appear in Operations Research.)

Table 1: Probabilities of Acceptance for

Link Sampling (LS) and Partial Link Sampling (PLS)

Acceptance numbers $a_1 = 1$; $a_2 = a_2' = 5$

N	n	D_{i-1}/N	D_i/N	D_{i+1}/N	p	p'	Probability of Acceptance LS	Probability of Acceptance PLS	Probability of Acceptance On 1st Sample (LS and PLS)
100	20	0.05	0.05	0.05	1	0	.9572	.9752	.7395
					1	0.1	.2312	.2302	.1818
					0.9	0	.9733	.9854	.7802
					0.9	0.1	.2598	.2591	.2006
					0.75	0	.9887	.9943	.8379
					0.75	0.1	.3067	.3065	.2308
		0.1	0.1	0.1	1	0	.5366	.5464	.3630
					1	0.1	.0808	.0797	.0735
					0.9	0	.6385	.6535	.4297
					0.9	0.1	.1043	.1031	.0922
					0.75	0	.7837	.8012	.5396
					0.75	0.1	.1504	.1494	.1269
		0.05	0.1	0.15	1	0	.5348	.6866	.3630
					1	0.1	.0806	.0880	.0735
					0.9	0	.6373	.7788	.4297
					0.9	0.1	.1040	.1144	.0922
					0.75	0	.7835	.8879	.5396
					0.75	0.1	.1502	.1652	.1269
		0.15	0.1	0.15	1	0	.4037	.4547	.3630
					1	0.1	.0746	.0760	.0735
					0.9	0	.4939	.5585	.4297
					0.9	0.1	.0948	.0973	.0922
					0.75	0	.6474	.7202	.5396
					0.75	0.1	.1343	.1396	.1269
		0.05	0.05	0.05	1	0	.9503	.9589	.7372
					1	0.1	.2376	.2371	.1868
					0.9	0	.9680	.9741	.7763
					0.9	0.1	.2649	.2645	.2046
					0.75	0	.9857	.9889	.8327
					0.75	0.1	.3102	.3101	.2336
		0.1	0.1	0.1	1	0	.5504	.5550	.3782
					1	0.1	.0872	.0867	.0789
					0.9	0	.6456	.6522	.4412
					0.9	0.1	.1102	.1097	.0972
					0.75	0	.7833	.7910	.5457
					0.75	0.1	.1553	.1548	.1308

N	n	D_{i-1}/N	D_i/N	D_{i+1}/N	p	p'	LS	PLS	Probability of Acceptance On 1st Sample (LS and PLS)
		0.05	0.1	0.15	1	0	.5486	.6782	.3782
					1	0.1	.0870	.0959	.0789
					0.9	0	.6444	.7637	.4412
					0.9	0.1	.1099	.1216	.0972
					0.75	0	.7830	.8717	.5457
					0.75	0.1	.1551	.1709	.1308
		0.15	0.1	0.15	1	0	.4225	.4713	.3782
					1	0.1	.0804	.0822	.0789
					0.9	0	.5081	.5665	.4412
					0.9	0.1	.1002	.1032	.0972
					0.75	0	.6538	.7168	.5457
					0.75	0.1	.1388	.1447	.1308
100	50	0.05	0.05	0.05	1	0	.2517	.2041	.1811
					1	0.1	.0018	.0018	.0018
					0.9	0	.3701	.3543	.2512
					0.9	0.1	.0029	.0029	.0029
					0.75	0	.5804	.6163	.3786
					0.75	0.1	.0052	.0052	.0052
		0.1	0.1	0.1	1	0	.0079	.0078	.0078
					1	0.1	.0001	.0001	.0001
					0.9	0	.0195	.0191	.0191
					0.9	0.1	.0002	.0002	.0002
					0.75	0	.0631	.0594	.0581
					0.75	0.1	.0007	.0007	.0007
		0.05	0.1	0.15	1	0	.0079	.0078	.0078
					1	0.1	.0001	.0001	.0001
					0.9	0	.0195	.0192	.0191
					0.9	0.1	.0002	.0002	.0002
					0.75	0	.0629	.0679	.0581
					0.75	0.1	.0007	.0007	.0007
		0.15	0.1	0.15	1	0	.0078	.0078	.0078
					1	0.1	.0001	.0001	.0001
					0.9	0	.0191	.0191	.0191
					0.9	0.1	.0002	.0002	.0002
					0.75	0	.0582	.0582	.0581
					0.75	0.1	.0007	.0007	.0007

N	n	D_{i-1}/N	D_i/N	D_{i+1}/N	p	p'	Probability of Acceptance		
							LS	PLS	On 1^{st} Sample (LS and PLS)
200	50	0.05	0.05	0.05	1	0	.3250	.3198	.2368
					1	0.1	.0028	.0028	.0028
					0.9	0	.4273	.4290	.2991
					0.9	0.1	.0039	.0039	.0039
					0.75	0	.6051	.6180	.4108
					0.75	0.1	.0062	.0062	.0062
		0.1	0.1	0.1	1	0	.0197	.0195	.0194
					1	0.1	.0002	.0002	.0002
					0.9	0	.0368	.0361	.0354
					0.9	0.1	.0004	.0004	.0004
					0.75	0	.0895	.0872	.0807
					0.75	0.1	.0010	.0010	.0010
		0.05	0.1	0.15	1	0	.0197	.0207	.0194
					1	0.1	.0002	.0002	.0002
					0.9	0	.0367	.0408	.0354
					0.9	0.1	.0004	.0004	.0004
					0.75	0	.0892	.1119	.0807
					0.75	0.1	.0010	.0010	.0010
		0.15	0.1	0.15	1	0	.0194	.0194	.0194
					1	0.1	.0002	.0002	.0002
					0.9	0	.0355	.0355	.0354
					0.9	0.1	.0004	.0004	.0004
					0.75	0	.0810	.0819	.0807
					0.75	0.1	.0010	.0010	.0010

Table 2: Acceptance Probabilities With Regular Double Sampling

In all cases $n = 20$, $n'' = 40$; $a_1 = 1$, $a_2 = a_2' = 5$

N	p	p'	Di/N = 0.05	Di/N = 0.1
100	1	0	1.0000	0.5305
	1	0.1	0.2241	0.0769
	0.9	0	1.0000	0.6565
	0.9	0.1	0.2542	0.0997
	0.75	0	1.0000	0.8202
	0.75	0.1	0.3033	0.1458