MODIFIED DORFMAN-STERRETT SCREENING (GROUP TESTING) PROCEDURES AND THE EFFECTS OF FAULTY INSPECTION

Qishen Huang
University of Maryland
College Park

Norman L. Johnson
University of North Carolina
at Chapel Hill

Samuel Kotz
University of Maryland
College Park

Abstract

Continuing previous work on effects of errors in inspection on group sampling schemes, a modification of Dorfman-Sterrett schemes is studied. The modification consists of reversion to group sampling when a specified number of decisions of nonconformance have occurred in the course of inspection of individual items.

KEY WORDS: Average sample number; inspection errors; quality control; screening

1. INTRODUCTION

Dorfman (1943) screening procedures are schemes designed to reduce the expected amount of testing needed to identify nonconforming (NC) items in a lot. They can be applied when it is possible to apply a single test to a group of n items, that will indicate whether there is at least one NC item present; for example, the items might be bottles of a liquid product; samples from a bottles can be mixed and the mixture tested for presence of a contaminant. The most prominent application has been in blood testing, which has been extensively covered in the work of M. Sobel (Sobel (1968) and the references thereon).

If there is no indication of presence of a NC item all n items in the group are declared to be conforming (C). If an NC item is indicated, each item in the group is tested individually. The total number of tests needed is either 1 or (n+1), as compared with n tests if plain individual testing is used. If the proportion (ω) of NC items in the lot is small, the probability that only 1 test is needed is high, with consequent reduction in the expected number of tests. Assuming that lot size is large enough for a binomial distribution to be used for the number (Y) of NC items in the group of n items, the reduction would be from n tests to $1+n\{1-(1-\omega)^n\} \approx 1+n^2\omega$, provided there are no inspection errors.

Sterrett (1957) introduced a modification, aimed at effecting further reduction in the expected number of tests. This modification calls for reversion to group testing whenever an item is declared NC on individual test. Thus, if the M-th item so tested is the first one declared NC, and $M \leq n-2$, Dorfman screening is applied to the remaining (as yet untested) (n-M) items. In the original proposal, there was no limit to the number of times this could occur - whenever an individual test gives a NC decision and there are at least two items not yet tested individually. Practical considerations may limit the number of reversions allowed; if up to k reversions are permitted, we have a k-stage Dorfman-Sterrett procedure.

A further modification, which will be studied in the present paper, requires that reversion to Dorfman screening is not undertaken until the g-th NC declaration ($g \geq 2$) occurs. This modification might be expected to be more effective than Sterrett's when ω is somewhat higher, (though still quite low) because g NC items have been removed, making it less likely that any remain among the items again subjected to Dorfman screening.

Effects of inspection errors on Dorfman screening procedures have been described by Johnson, Kotz and Rodriguez (1988); and effects on 1- and 2-stage Dorfman-Sterrett procedures have been described by the same authors (1987 - henceforth referred to as JKR). In the present paper we study effects of inspection errors on modified Dorfman-Sterrett procedures, restructing our analysis to 1-stage procedures.

2. NOTATION

As in earlier studies, we wish to evaluate three measures of performance:

PC(C): probability of correct classification of C items;

PC(NC): probability of correct classification of NC items;

E: expected total number of tests (group and individual).

These will depend on ω and n. To indicate dependence on n we add a subscript. As in JKR, we will present tables of values, $PC(C)_{n|y}$, $PC(NC)_{n|y}$ and $E_{n|y}$ of these three quantities conditional on fixed values of the number (Y=y) of NC items among the n items in the group. To obtain overall values, these values must be averaged with respect to the appropriate distributions of Y. If lot sizes are large (effectively infinite) these are

For $PC(C)_{n|y}$: binomial with parameters $(n-1,\omega)$.

For $PC(NC)_{n|y}$: 1 + [binomial with parameters $(n-1,\omega)$].

For $E_{n|y}$: binomial with parameters (n,ω).

(For finite lot sizes, appropriate hypergeometric distributions are relevant.)

As in JKR, also, we first consider situations conditional not only on Y=y, but also on the order (M=m) in the individual test sequence at which the second NC declaration occurs, and the number (T=t) of truly NC items among these first M items tested.

To evaluate $PC(C)_{n|y}$ and $PC(NC)_{n|y}$ we will use the relations

$$PC(C)_{n|y} = E(C)_{n|y}/(n-y); \quad PC(NC)_{n|y} = E(NC)_{n|y}/y$$

where $E(C)_{n|y}$, $E(NC)_{n|y}$ denote the expected numbers of C, NC items, respectively, that are identified correctly. To indicate conditioning on M=m, T=t and J=j in addition to Y=y, the set of subscripts $n|m,t,y,j$ will be used.

A group containing at least one NC item will be called a <u>NC group</u>; if a group is not NC, it is a <u>C group</u>.

Inspection quality will be defined in terms of

$p_0(p_0')$: probability that a NC(C) group is declared NC (assumed to be the same whatever the value of $n \geq 2$).

$p(p')$: probability that a NC(C) item is declared NC when tested individually.

The event $(M=m) \cap (T=t)$ can be split into subevents according to the number of truly C items among those declared NC. We will denote this number by J. (The order in which these events occur is not relevant.)

We define

$P(m,t,j|n,y)$ = probability that M=m, T=t and J=j, given Y=y.

Also

$$P(m,t|n,y) = \sum_j P(m,t,j|n,y) = \text{probability that M=m and T=t, given Y=y.}$$

(Limits for j are given in (1), below.)

It is assumed that the result of any test (group or individual) is independent of the result(s) of any other test(s).

3. ANALYSIS

For $y > 0$

$P(m,t,j|n,y)$ = Pr[group test gives NC result]

× Pr[t NC's and (m−t) C's in first n items tested) ∩ ((g−j) NC's and j C's declared NC)]

× Pr[g-th item declared NC is m-th item tested, given g NC decisions in first m test]

$$= p_0 \times \binom{n}{y}^{-1} \binom{n-m}{y-t}\binom{m}{t}\binom{t}{g-j}\binom{m-t}{j} p^{g-j}(1-p)^{t-g+j} p'^{j}(1-p')^{m-t-j}$$

$$\times (gm^{-1}) \quad (\max(0,y-n+m) \leq t \leq \min(y,m);$$

$$\max(0,g-t) \leq j \leq \min(g,m-t)) \qquad (1)$$

If $y=0$, p_0 is replaced by p_0' and only $j=g$ is relevant. If $y=n$, only $j=0$ is relevant.

For $2 \leq m \leq n-2$

$$E(C)_{n|m,t,y,j} = m-t-j + (n-y-m+t)\{1-h(t,y)p'\}$$

$$E(NC)_{n|m,t,y,j} = g-j + (y-t)p_0 p$$

$$E_{n|m,t,y} = m+2 + (n-m)h(t,y)$$

where $h(t,y) = \begin{cases} p_0 & \text{if } t < y \\ p_0' & \text{if } t = y \end{cases}$.

(Note that $E_{n|m,t,y,j}$ does not depend on j.)

Also

$$E(C)_{n|n-1,t,y,j} = n-y-j-(t+1-y)p'$$

$$E(C)_{n|n,y,y,j} = n-y-j$$

$$E(NC)_{n|n-1,t,y,j} = g-j + (y-t)p$$

$$E(NC)_{n|n,y,y,j} = g-j$$

and $$E_{n|n-1,t,y} = E_{n|n,y,y} = n+1$$

In calculating $E(C)_{n|y}$, $E(NC)_{n|y}$ and $E_{n|y}$ we have to take into account

(a) cases when fewer than g NC declarations occur and also (b) cases when group-testing leads to immediate declaration that all items in the group are C.

We have (for limits for t and j see (1)):

$$E(C)_{n|y} = \sum_{m=g}^{n} \sum_{t} \sum_{j} P(m,t,j|n,y)E(C)_{n|m,t,y,j} +$$

$$+ p_0 \sum_{h=0}^{g-1} \sum_{j=0}^{h} (n-y-j)\binom{y}{h-j}\binom{n-y}{j} p^{h-j}(1-p)^{y-h+j} p'^{j}(1-p')^{n-y-j}$$

$$+(n-y)\{1-h(n-y,n)\}$$

$$E(NC)_{n|y} = \sum_{m=g}^{n} \sum_{t} \sum_{j} P(m,t,j|n,y)E(NC)_{n|m,t,y,j} +$$

$$+ p_0 \sum_{h=0}^{g-1} \sum_{j=0}^{h} (h-j)\binom{y}{h-j}\binom{n-y}{j}p^{n-j}(1-p)^{y-h+j}p'^{j}(1-p')^{n-y-j}$$

For $y > 0$

$$E_{n|y} = \sum_{m=g}^{n} \sum_{t} P(m,t|n,y)E_{n|m,t,y}$$

$$+ np_0 \sum_{h=0}^{g-1} \sum_{j=0}^{h} \binom{y}{h-j}\binom{n-2-y}{j}p^{h-j}(1-p)^{y-h+j}p'^{j}(1-p')^{n-2-y-j} + 1-p_0$$

For $y=0$

$$E_{n|0} = 1 + p_0'\left[\sum_{m=g}^{n-2} \begin{bmatrix} m-1 \\ g-1 \end{bmatrix} p'^{g}(1-p')^{m-g}\{m+1+(n-m)p_0'\} + n\sum_{j=0}^{g-1}\binom{n-2}{j}p'^{j}(1-p')^{n-2-j}\right]$$

4. TABLES

Tables are presented here for the case g=2 only. Further tables are in preparation for a later paper, in which the question of choice of value of g will be addressed, among other topics.

Values of $E_{6|y}$, $PC(NC)_{6|y} = E(NC)_{6|y}/y$ and $PC(C)_{6|y} = E(C)_{6|y}/(6-y)$ for all relevant values of y are given in Tables 1,2 and 3 respectively. They are comparable with values in Tables 1-6 of JKR according to the scheme.

Table	Tables in JKR
1 $(E_{6|y})$	1 $({}_1E_{6|y})$; 4 $({}_2E_{6|y})$
2 $(PC(NC)_{6|y})$	2 $({}_1PC(NC)_{6|y})$; 5 $({}_2PC(NC)_{6|y})$
3 $(PC(C)_{6|y})$	3 $({}_1PC(C)_{6|y}$; 6 $({}_2PC(NC)_{6|y})$

In the JKR tables, the prefix (1 or 2) refers to the number of stages in the (nonmodified) Dorfman-Sterrett procedure.

For example, with $p_0=p=0.75$ and $p'_0=p'=0.10$, we find

y	$E_{6\|y}$	${}_1E_{6\|y}$	${}_2E_{6\|y}$	$PC(NC)_{6\|y}$	${}_1PC(NC)_{6\|y}$	${}_2PC(NC)_{6\|y}$	$PC(C)_{6\|y}$	${}_1PC(C)_{6\|y}$	${}_2PC(C)_6$
0	1.59	1.52	1.52	-	-	-	.987	.992	.994
1	5.36	4.78	4.56	.612	.533	.531	.924	.949	.951
2	5.30	5.44	5.31	.570	.500	.491	.938	.938	.947
3	5.56	5.47	5.54	.538	.479	.456	.935	.938	.945
4	5.63	5.44	5.51	.514	.467	.429	.934	.939	.946
5	5.62	5.40	5.43	.497	.459	.409	.935	.940	.947
6	5.69	5.37	5.36	.485	.433	.395	-	-	-

The expected numbers of tests are somewhat greater for the modified procedure, but the probabilities of correct decision for NC items (PC(NC)) are substantially greater. There is a relatively slight decrease in the probability of correct decision for C items, especially by comparison with the 2-stage unmodified Dorfman Sterrett procedure. (However, comparison with 2-stage procedures is perhaps unwarranted.)

The somewhat irregular progression of values of $E_{6|y}$ and $PC(C)_{6|y}$ with respect to y should be noted. Note also that $PC(C) < {}_1PC(C) < {}_2PC(C)$, but $PC(NC) > {}_1PC(NC) > {}_2PC(NC)$. These features are associated with increasing use of group tests, leading to increased chances of C decision without individual testing.

Tables 1A, 2A and 3A give values for n=8, but only for $y \leq 6$. These exhibit similar patterns to those for n=6.

5. A SPECIAL CASE

For the case of lots containing a large (effectively infinite) number of items, it is possible to evaluate PC(NC) directly, in an elegant, elementary fashion, without finding $PC(NC)_{n|y}$ first. This approach exploits the fact that, given a particular NC item, $\mathscr{A}$, say in a group of n items, the probability of proceeding to individual inspection as a result of a 'NC group' decision is p_0, whatever the status of the other (n-1) items. Hence, we can evaluate performance on individual testing by regarding the number of NC decisions on the other (n-1) items as a binomial variable with parameters (n-1) and $\omega = \omega p+(1-\omega)p'$ (the probability that a randomly chosen item will receive a NC decision on individual test).

The probability that there would be h NC decisions among the other (n-1) items, and a (correct) NC decision for $\mathscr{A}$, supposing each item is tested individually is

$$p\binom{n-1}{h}\,\omega^h(1-\omega)^{n-1-h} = P_h, \text{ say.}$$

Given this event, the conditional probability that $\mathscr{A}$ will actually be declared NC is

$$\begin{cases} 1 & \text{if } h < g \\ \dfrac{g}{g+1} + \dfrac{g}{n(n-1)} + \left\{1 - \dfrac{g}{g+1} - \dfrac{g}{n(n-1)}\right\}p_0 & \text{if } h=g \\ \dfrac{g}{h+1} + \left(1 - \dfrac{g}{h+1}\right)p_0 & \text{if } h > g \end{cases}$$

Note that the correct decision is reached without reversion to group testing if (a) $\mathscr{A}$ is among the first g NC decisions or (b) with h=g, it is the last item tested and the immediately preceeding item is NC (necessarily the g-th NC). The probability of (b) is

$$\binom{n-2}{g-1} \Big/ \left\{n\binom{n-1}{g}\right\} = \frac{g}{n(n-1)} .$$

Hence

$$PC(NC) = p_0 \left[\sum_{h=0}^{g-1} P_h + \left\{ \frac{g}{g+1} + \frac{g}{n(n-1)} + \left(1 - \frac{g}{g+1} - \frac{g}{n(n-1)}\right) p_0 \right\} P_g + \sum_{h=g+1}^{n-1} \left\{ \frac{g}{h+1} + \left(1 - \frac{g}{h+1}\right) p_0 \right\} P_h \right] .$$

This method cannot be applied to evaluate PC(C), because, given a particular C item, the probability of proceeding to individual testing does depend on the status of the other (n-1) items in the set. Given that individual testing starts, therefore, these (n-1) items cannot be regarded as randomly chosen from the lot.

ACKNOWLEDGMENTS

The authors acknowledge stimulating discussions with Mr. Luo Zhen and Mr. Wang Xi Bing of the Department of Mathematics, Nankai University, Tianjin, China.

REFERENCES

Dorfman, R. (1943) "The detection of defective members of large populations," *Ann. Math. Statist.*, **14**, 436-440.

Johnson, N.L., Kotz, S. and Rodriguez, R.N. (1987) "Dorfman-Sterrett (group testing) schemes and the effects of faulty inspection," *Inst. Statist. Univ. North Carolina*, Mimeo Series No. 1722.

Johnson, N.L., Kotz, S. and Rodriguez, R.N. (1988) "Statistical effects of imperfect inspection sampling, III", *J. Qual. Technol.*, **20**, (to appear).

Sobel, M. (1968) "Binomial and hypergeometric group testing" *Stud. Scient. Math. Hungarica*, **3**, 19-42.

Sterrett, A. (1957) "On the detection of defective members of large populations," *Ann. Math. Statist.*, **28**, 1033-1036.

TABLE 1 VALUES OF $E_{6|y}$

y	$p_0=p=$	0.75			0.90			0.95		
	$p_0'=p'=$	0.25	0.10	0.05	0.25	0.10	0.05	0.25	0.10	0.05
0		2.43	1.59	1.30	2.43	1.59	1.30	2.43	1.59	1.30
1		5.29	5.36	5.42	6.15	6.21	6.29	6.44	6.49	6.57
2		5.41	5.30	5.26	6.42	6.14	6.03	6.77	6.41	6.28
3		5.59	5.56	5.55	6.88	6.83	6.80	7.38	7.32	7.29
4		5.62	5.63	5.63	6.97	6.96	6.96	7.47	7.47	7.47
5		5.62	5.62	5.63	6.97	6.97	6.98	7.47	7.48	7.48
6		5.59	5.59	5.59	6.96	6.96	6.96	7.46	7.46	7.46

TABLE 2 VALUES OF $PC(NC)_{6|y}$

y	$p_0=p=$	0.75			0.90			0.95		
	$p_0'=p'=$	0.25	0.10	0.05	0.25	0.10	0.05	0.25	0.10	0.05
1		.513	.612	.672	.711	.778	.828	.785	.836	.883
2		.530	.570	.588	.784	.802	.810	.887	.896	.900
3		.520	.538	.546	.779	.787	.790	.884	.889	.890
4		.506	.514	.517	.770	.773	.774	.879	.880	.881
5		.494	.497	.498	.764	.765	.765	.876	.876	.876
6		.485	.485	.485	.759	.759	.759	.873	.873	.873

TABLE 3 VALUES OF $PC(C)_{6|y}$

y	$p_0=p=$	0.75			0.90			0.95		
	$p_0'=p'=$	0.25	0.10	0.05	0.25	0.10	0.05	0.25	0.10	0.05
0		.906	.987	.997	.906	.987	.997	.906	.987	.997
1		.827	.924	.961	.779	.902	.949	.760	.894	.945
2		.844	.938	.969	.810	.928	.965	.798	.924	.963
3		.837	.935	.967	.790	.916	.958	.771	.908	.954
4		.837	.934	.967	.788	.915	.958	.770	.908	.954
5		.838	.935	.968	.789	.916	.958	.770	.908	.954

TABLE 1A VALUES OF $E_{8|y}$

y	$p_0=p=$	0.75			0.90			0.95		
	$p_0'=p'=$	0.25	0.10	0.05	0.25	0.10	0.05	0.25	0.10	0.05
0		2.43	1.71	1.39	2.43	1.71	1.39	2.43	1.72	1.39
1		5.62	5.86	6.29	6.75	6.76	7.26	7.16	7.06	7.58
2		5.82	5.20	5.00	7.19	5.79	5.16	7.70	5.98	5.15
3		6.27	5.82	5.62	8.26	7.76	7.56	9.07	8.64	8.39
4		6.54	6.23	6.24	8.62	8.55	8.50	9.40	9.40	9.38
5		6.68	6.62	6.59	8.70	8.74	8.73	9.44	9.50	9.50
6		6.73	6.73	6.73	8.69	8.73	8.74	9.42	9.48	9.48

TABLE 2A VALUES OF $PC(NC)_{8|y}$

y	$p_0=p=$	0.75			0.90			0.95		
	$p_0'=p'=$	0.25	0.10	0.05	0.25	0.10	0.05	0.25	0.10	0.05
1		.509	.612	.672	.711	.778	.828	.785	.836	.883
2		.515	.558	.580	.777	.796	.806	.883	.893	.898
3		.505	.529	.539	.772	.782	.787	.881	.886	.888
4		.495	.493	.495	.765	.771	.773	.877	.875	.880
5		.486	.493	.495	.760	.763	.764	.874	.875	.876
6		.479	.482	.484	.757	.758	.759	.872	.873	.873

TABLE 3A VALUES OF $PC(C)_{8|y}$

y	$p_0=p=$	0.75			0.90			0.95		
	$p_0'=p'=$	0.25	0.10	0.05	0.25	0.10	0.05	0.25	0.10	0.05
0		.867	.982	.996	.867	.982	.996	.867	.982	.996
1		.820	.915	.954	.762	.885	.937	.739	.874	.931
2		.842	.936	.968	.805	.926	.964	.793	.924	.963
3		.839	.935	.968	.790	.916	.958	.771	.909	.954
4		.839	.935	.967	.789	.915	.958	.770	.908	.954
5		.841	.936	.968	.790	.916	.958	.771	.908	.954
6		.842	.937	.968	.791	.916	.958	.771	.908	.954