SEQUENTIAL TESTS FOR THE COEFFICIENT OF CORRELATION EXACT WALD REGIONS, OPERATING CHARACTERISTIC AND AVERAGE SAMPLE NUMBER

Don B. Campbell, Vidya Taneja, Western Illinois University
Leo A. Aroian, Union College and University

ABSTRACT

It is shown how exact Wald regions for the sequential probability ratio test for the coefficient of correlation $\rho = \rho_0$ versus $\rho = \rho_1$ may be found, and also how to determine the operating characteristic function, OC, and the average sample number ASN, by Monte Carlo techniques. A two decision example, and a three decision example $\rho = \rho_0$ versus $\rho = \rho_1$ and $\rho = \rho_2$ are included.

1. Introduction

Let $\{x_{1i}, x_{2i}\}$ be pairs of observations given by a normal bivariate distribution with unknown parameters μ_1, μ_2, σ_1, σ_2, and ρ. We determine exact regions for the Wald sequential test: H_0, $\rho = \rho_0$ versus H_1, $\rho = \rho_1$, $\rho_1 > \rho_0$. The values of the operating characteristic function, OC, and the average sample number ASN, are approxomated by Monte Carlo methods. We expect to obtain exact results in the future. So far as the authors know, no exact Wald regions for the coefficient of correlation have been determined. A FORTRAN program is available so any desired regions may be constructed very quickly.

2. Description of the Test

Let

$$\bar{x}_{1n} = \sum_{i=1}^{n} x_{1i}/n, \quad \bar{x}_{2n} = \sum_{i=1}^{n} x_{2i}/n,$$

$$s_{1n}^2 = \sum_{i=1}^{n} (x_{1i} - x_{1n})^2/n,$$

$$s_{2n}^2 = \sum_{i=1}^{n} (x_{2i} - \bar{x}_{2n})^2/n,$$

$$r_n = \sum_{i=1}^{n} (x_{1i} - \bar{x}_{1n})(x_{2i} - \bar{x}_{2n})/n s_{1n} s_{2n},$$

define the two sample means, the sample variances, and the sample coefficient of correlation respectively after observation n is taken. We test the hypothesis H_0: $\rho = \rho_0$, versus H_1: $\rho = \rho_1$, $\rho_1 > \rho_0$. The Wald sequential probability ratio test limits are given by $r_n(u)$, the upper limit for r_n, and $r_n(\ell)$ the lower limit for r_n. As soon as $r_n \le r_n(\ell)$ accept $\rho = \rho_0$, and as soon as $r_n \ge r_n(u)$ accept $\rho = \rho_1$. The values of $r_n(u)$ and $r_n(\ell)$ are determined as follows.

First $Z_n(r_n)$ must be found:

$$\begin{aligned} Z_2(r_2) &= \ln(\pi - 2\sin^{-1}\rho_1) - \ln(\pi - 2\sin^{-1}\rho_0), \quad \text{if } r_2 = -1, \\ &= \ln(\pi + 2\sin^{-1}\rho_1) - \ln(\pi + 2\sin^{-1}\rho_0), \quad \text{if } r_2 = 1. \end{aligned} \tag{2.1}$$

$$\begin{aligned} Z_n(r_n) &= .5(n-1)(\ln(1-\rho_1^2) - \ln(1-\rho_0^2)) \\ &- (n-1.5)(\ln(1-\rho_1 r_n) - \ln(1-\rho_0 r_n)) \\ &+ \ln F(.5,.5,n-.5;\ .5(1+\rho_1 r_n)) \\ &- \ln F(.5,.5,n-.5;\ .5(1+\rho_0 r_n)), \text{if } n_2 > 2. \end{aligned}$$

Note that $F(v,v';v'';z) =$

$$\sum_{j=0}^{\infty} \frac{\Gamma(v+j)\Gamma(v'+j)\Gamma(v'')}{\Gamma(v)\Gamma(v')\Gamma(v''+j)} (z^j/j!)$$ is the

hypergeometric function. Let $b=\ln(\beta/(1-\alpha))$ and $a=\ln(1-\beta)/\alpha$. If $n = 2$ and $r_2 = -1$, and $Z_2(-1) \le b$, accept $\rho = \rho_0$; if $n = 2$ and $r_2 = 1$, and $Z_2(1) \ge a$, accept $\rho = \rho_1$. If $n \geq 3$, r_n is computed from $Z_n(r_n) = b$ and $Z_n(r_n) = a$ and $\rho = \rho_0$ or $\rho = \rho_1$ is accepted depending on whether $r_n \le r_n(\ell)$, or $r_n \ge r_n(u)$ where $r_n(\ell)$ and $r_n(u)$ are solutions of $Z_n(r_n(\ell)) = b$, and $Z_n(r_n(u)) = a$.

This test and all the preceding results are due to B.K. Ghosh (1970). As n becomes large, the following approximations to $r_n(\ell)$ and $r_n(u)$ are useful:

$$r_n(u) \text{ or } r_n(\ell) = (W-1)/(W\rho_1 - \rho_0), \tag{2.2}$$

$$W = \{(1-\rho_0^2)/(1-\rho_1^2)\}^{\frac{1}{2}} \exp(\omega/n) \tag{2.3}$$

$$W = \{(1-\rho_0^2)/(1-\rho_1^2)\}^{\frac{1}{2}+\frac{1}{4}n} \exp(\omega/n-1.5) \tag{2.4}$$

Note ω is a dummy variable to be replaced by b or a. Formula (2.3) is correct to $0(n^{-1})$ and formula (2.4) is correct to $0(n^{-2})$. If b or a is used in the exponentials in formulas (2.3) or (2.4) for ω, and if the resulting W's are substituted in formula (2.2), then $r_n(\ell)$ or $r_n(u)$ is determined. If $n \to \infty$ we obtain

$$r_\infty = \{((1-\rho_0^2)(1-\rho_1^2))^{\frac{1}{2}} - 1\} / \{(\rho_1((1-\rho_0^2)/(1-\rho_1^2))^{\frac{1}{2}} - \rho_0\} \tag{2.5}$$

Equations (2.2), (2.3) and (2.4) are reformulations of those of Ghosh (1970), page 324, and are somewhat simpler to calculate. If $\rho_1 < \rho_0$, a simple interchange of ρ_0 and ρ_1 is used in (2.1) with corresponding changes in α and β.

3. Calculation of the Regions

$Z_n(r_n(\ell))=b$ and $Z_n(r_n(u))=a$ are solved by trial and error and repeated linear interpolation. The solutions are nearly correct to four decimal places throughout, but occasionally the fourth decimal may be in error by as much as ± 2. A computer program is available from Don Campbell. The programming and calculation of the tables were efficiently handled by Sheri Butler. The approximate formulas for $r_n(\ell)$ and $r_n(u)$ (2.2) and (2.4) may be used to extent the tables.

4. Monte Carlo

In all cases 1000 values of the coefficient of correlations were calculated at the beginning of each run and these were continued until they went into the rejection region or acceptance region or were truncated at the truncation point, where they were placed in the acceptance or rejection region by the use of r_∞ given by (2.5). If $r_\infty \leq r(u)$ then H_1 was chosen, otherwise H_o. We generate two unit normal variates (Y_1, Y_2) with correlation coefficient as follows. First generate two independent unit normal variates U_1 and U_2 by the Box-Muller formulas

$$U_1 = (-2 \log R_1)^{\frac{1}{2}} \sin 2\pi R_2$$

$$U_2 = (-2 \log R_1)^{\frac{1}{2}} \cos 2\pi R_2$$

where R_1 and R_2 are random (0,1) variates. Then

$$Y_1 = U_1 \text{ and } Y_2 = \rho U_1 + (1-\rho^2)^{\frac{1}{2}} U_2,$$

as given by Wold (1948).

The direct method was used throughout these simulations. At each trial, the number of acceptances for H_o, H_1 and the number continuing into the next trial are found. Thus, at each exit point the number of items for each value of ρ has been determined and the distribution at this point may be found using the values of $\rho = -1$, $\rho_o-\Delta$, ρ_o, $\rho_o+\Delta$, $\rho_o+2\Delta$, $\rho_o+3\Delta$, ρ_1, $\rho_1+\Delta$, and 1, where $\Delta = .25(\rho_1-\rho_o)$. From this distribution an estimate of ρ may be made and also approximate confidence limits may be found provided the Monte Carlo trials are sufficiently extensive.

The direct method not only provides the OC and ASN but gives the DSN, decisive sample number distribution, and the conditional distribution at each point.

5. Example, $\alpha = \beta = .10$, $\rho_o = 0$, $\rho_1 = .25$

As an example, we choose $\alpha = \beta = .10$, $\rho_o = 0$, $\rho_1 = .25$. We give the region, the point of truncation m_o, the OC and the ASN, the conditional distribution at one point, the estimate for ρ after a sequential decision has been reached, and the sample size N for the corresponding fixed size test with the same α and β. Corresponding results for a variety of other cases are given in another paper, Taneja et al (1977). The region is given in Table 1. The truncation point is $m_o = 124$, the fixed size sample is N=104. The OC and ASN are given in Table 2. Usually the truncation point m_o was chosen as 1.2 times the fixed size sample N.

The actual value of α is $\alpha^1 = .111$ instead of .10 while the actual value of β is $\beta^1 = .105$ instead of the planned .10. The Monte Carlo trials, 1000, are too few to estimate α if the trials were continued to infinity. The greatest value of the ASN is 75.81, so the fixed size sample test is 73% efficient and only 55% efficient at $\rho_o = 0$ and $\rho_1 = .25$. This shows the real savings in observations needed to reach a decision.

Replications of the Monte Carlo trials show that the OC may be off as much as one unit in the second decimal place, and the ASN by one unit in the second digit at $\rho = \rho_o$ and $\rho = \rho_1$. Elsewhere the errors are somewhat larger.

Suppose the test terminates with acceptance at observation 25, what is the estimate of ρ using the mean value of the conditional distribution and approximate confidence limits for ρ based on this result? From the Monte Carlo trials at decision point 25 (not given here) we have the results:

ρ = -1,	-.0625,	0,	.0625	.125
0	10	13	8	7
0	.244	.317	.195	.171
(continued)				
ρ = .1875	.25,	.3125,	1	
1	2	0	0	
.024	.049	0	0	

where the first line are the values of ρ, the second line the number of times out of the 1000 trials that the test was terminated at 25, and the last line the estimated probabilities.

The mean value of this estimated distribution is the estimated value of ρ, .035, based on the 41 exits at this point.

6. Theory of the Three Decision Test

For three decision test we choose H_o $\rho = \rho_o$ versus a two-sided alternative H_1, $\rho = \rho_1$, $\rho_1 > \rho_o-\Delta$, and $\rho = \rho_2$, $\rho_2 < \rho_o + \Delta$.

The operating characteristic function is given piecewise:

$OC_o(\rho)$ = probability of accepting H_o, that is, $\rho = \rho_o$,

$OC_1(\rho)$ = probability of accepting $\rho = \rho_1$,

$OC_2(\rho)$ = probability of accepting $\rho = \rho_2$ and

$OC_1 + OC_o + OC_2 = 1.$

We choose α as follows:

$$OC_o\ (\rho_o | \rho = \rho_o) = 1-2\alpha,$$

$$OC_1\ (\rho_1 | \rho = \rho_1) = 1-\alpha,\ OC_2(\rho_2 | \rho = \rho_2) = 1-\alpha.$$

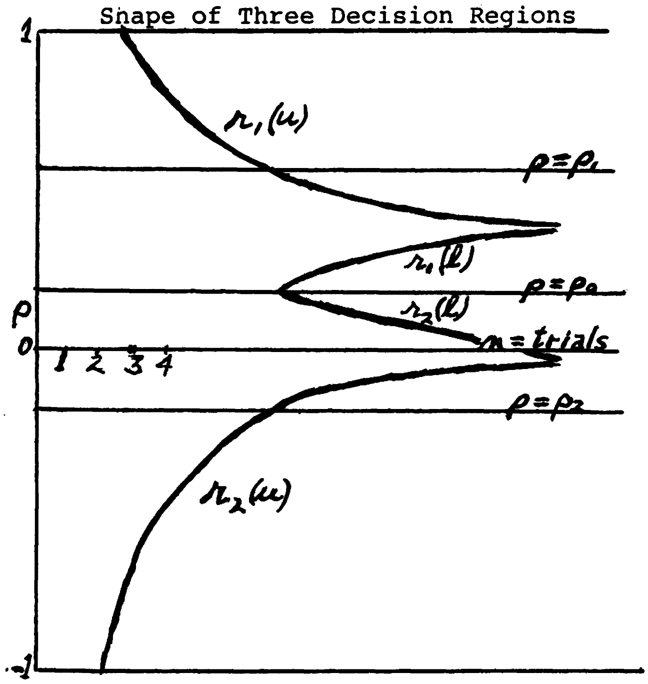
Thus, if α is chosen as .10, the probability of rejecting $\rho = \rho_o$ when $\rho = \rho_o$ is .20, while if either $\rho = \rho_1$ or ρ_2 the probability of rejecting $\rho = \rho_1$ or ρ_2 is .10 when $\rho = \rho_1$ or $\rho = \rho_2$ is true.

The region for the three decision test is determined by combining two two decision regions. First a two decision region is found for $\rho = \rho_o$ versus $\rho = \rho_1$ with $\alpha = \beta$, $\rho_1 = \rho_o + \Delta$, $\Delta > 0$, since $\rho_1 > \rho_o$ in formula (2.4). This gives us $r_1(u)$ and $r_1(\ell)$ boundaries. Next, a two decision region is found with $\rho_1^1 = -(\rho_o - \Delta)$, and $\rho_o^1 = -\rho_o$, since $\rho_1^1 > \rho_o^1$ in formulas (2.1)-(2.5), replacing ρ_1 and ρ_o by ρ_1^1 and ρ_o^1. This gives the values of $r_2(u)$ and $r_2(\ell)$. The two regions are combined. The values of $r_i(u)$ are unchanged, but the values of $r_i(\ell)$ are deleted until they intersect on the line $\rho = \rho_o$.

Intuitively we may expect the two regions to be symmetric about the line $\rho = \rho_o$. This happens if $\rho_o = 0$, but not otherwise since the distribution of r is only symmetric if $\rho = 0$. If we choose any two decision plan with $\alpha = \beta$ then the three decision regions will have approximately $OC(\rho = \rho_o) = 1-2\alpha$, $OC(\rho=\rho_1) = 1-\alpha = OC(\rho=\rho_2)$. If we choose a two decision plan with $\alpha = .5\beta$, then approximately $OC(\rho=\rho_o) = OC(\rho=\rho_1) = OC(\rho=\rho_2) = (1-\alpha)$. We may, of course, combine the two two-decision regions in such a way $OC(\rho=\rho_o) = 1-\alpha_1$, $OC(\rho=\rho_1) = 1-\alpha_2$ and $OC(\rho=\rho_2) = 1-\alpha_3$. The three decision regions always have the following shape, Figure 1:

FIGURE 1

Shape of Three Decision Regions



These three decision regions are essentially generalized **Barnard regions.** They are also similar to **Wald-Sobel** regions except no decisions are possible until one of the boundaries is reached.

8. Example of a Three Decision Test

We take the two decision test with $\alpha = \beta = .10$, $\rho_o = 0$ and $\rho_1 = .25$. We rotate the region about $\rho = 0$ to obtain the region below $\rho = 0$. The values of $r_i(u)$ are now plus and minus, the values of the middle section $r_i(\ell)$ start at trial 70 as $\pm.0017$, trial 71 $\pm.0030$, trial 72 $\pm.0053$, . . ., and at trial 124 $\pm.0574$. Thus, if the value of r exceeds $r_1(u)$ make the decision $\rho = \rho_1 = .25$, if r lies between $\pm r_i(\ell)$ conclude $\rho = \rho_o = 0$, or if $r < r_2(u)$ conclude $\rho = \rho_2 = -.25$.

The results for the OC and ASN are given in Table 3. The actual values of the OC are very close to the planned values .90, .80, .90 at $\rho = -.25$, 0, .25, namely .901, .797, and .890. The OC_{-1} and OC_1 are symmetric to each other about zero. The actual values mirror this and serve as a check on the Monte Carlo trials. The fixed size test takes 104 observations, so the efficiency of it varies from 11% to 73% as compared to the sequential test.

TABLE 1.
Values of $r_n(\ell)$ and $r_n(u)$
$\rho_o = 0$, $\rho_1 = .25$, $\alpha = \beta = .10$.

n	$r_n(\ell)$	$r_n(u)$
1 - 10	(No decision possible)	
11	-.8707	.9307
12	-.7658	.8620
13	-.6807	.8043
14	-.6104	.7548
15	-.5514	.7120
16	-.5011	.6748
17	-.4577	.6420
18	-.4199	.6129
19	-.3867	.5868
20	-.3573	.5634
21	-.3310	.5424
22	-.3075	.5232
23	-.2863	.5056
24	-.2671	.4897
25	-.2495	.4749
26	-.2334	.4614
27	-.2187	.4488
28	-.2051	.4372
29	-.1926	.4264
30	-.1809	.4162
31	-.1701	.4068
32	-.1600	.3981
33	-.1506	.3897
34	-.1417	.3819
35	-.1335	.3745
36	-.1257	.3676
37	-.1184	.3610
38	-.1114	.3549
39	-.1049	.3489
40	-.0987	.3433

n	$r_n(\ell)$	$r_n(u)$
41	-.0928	.3381
42	-.0872	.3330
43	-.0819	.3282
44	-.0769	.3235
45	-.0721	.3190
46	-.0675	.3148
47	-.0632	.3109
48	-.0590	.3070
49	-.0549	.3033
50	-.0511	.2998
52	-.0439	.2931
54	-.0373	.2869
56	-.0311	.2811
58	-.0254	.2758
60	-.0201	.2707
62	-.0106	.2661
64	-.0106	.2617
66	-.0062	.2577
68	-.0021	.2538
70	.0017	.2502
72	.0053	.2467
74	.0087	.2434
76	.0119	.2403
78	.0150	.2375
80	.0178	.2346
82	.0206	.2320
84	.0233	.2295
86	.0257	.2271
88	.0281	.2249
90	.0304	.2227
92	.0325	.2206
94	.0346	.2186
96	.0365	.2166
98	.0385	.2148
100	.0403	.2130
102	.0420	.2114
104	.0437	.2097
106	.0453	.2082
108	.0468	.2067
110	.0484	.2052

Table 1 (Continued)

n	$r_n(\ell)$	$r_n(u)$
112	.0498	.2038
114	.0512	.2025
116	.0525	.2011
118	.0538	.1999
120	.0551	.1987
122	.0563	.1975
124	.0574	.1963

REFERENCES

Aroian, L.A. Applications of the direct method in sequential analysis, Technometrics, 1976, 18, 301-306.

Box, G.E.P. and Muller, M.E., A note on the generation of random normal deviates, Annals of Mathematical Statistics, 1958, 29, 610-611.

Ghosh, B.K., *Sequential tests of statistical hypotheses*, Reading, Mass.: Addison-Wesley, 1970.

Taneja, V., Campbell, D. and Aroian, L.A., Tables of the regions, operating characteristic function, and average sample number for Wald sequential tests of the coefficient of correlation, submitted to Psychometrika, 1977.

Wold, H.O., *Random normal deviates.* Tracts for Computers, No. 25, Cambridge, England: Cambridge University Press, 1948.

TABLE 2.

OC and ASN, $\rho_o = 0$, $\rho_1 = .25$, $\alpha = \beta = .10$

ρ	-1	-.0625	0	.0625	.125	.1875	.25	.3125	1
OC	1	.957	.889	.749	.487	.272	.105	.036	0
ASN	11	48.49	57.10	68.78	75.81	69.30	57.27	45.70	11

TABLE 3.

OC and ASN, Three Decision Test, $\rho_2 = -.25$, $\rho_o = .0$, $\rho_1 = .25$

ρ	-1.	-.3125	-.25	-1875	-.125	-.0625	0	.0625	.125	.1875	.25	.3125	1.
OC_{-1}	1.	.967	.901	.722	.475	.247	.092	.035	.007	.003	.003	.002	0
OC_o	0	.033	.097	.271	.506	.718	.797	.724	.527	.277	.107	.027	0
OC_1	0	.000	.002	.007	.019	.035	.111	.241	.466	.720	.890	.971	1
ASN	9	46.50	57.30	69.44	76.27	74.56	75.76	74.81	76.50	69.06	56.08	45.94	9.