American Journal of Mathematical and Management Sciences, 33:228–238, 2014

ISSN: 0196-6324 print / 2325-8454 online
DOI: 10.1080/01966324.2014.928244

PERCENTAGE POINTS FOR TESTING HOMOGENEITY OF SEVERAL BIVARIATE GAUSSIAN POPULATIONS

DAYA K. NAGAR[1] and ARJUN K. GUPTA[2]
[1]Instituto de Matemáticas, Universidad de Antioquia, Calle 67, No. 53-108, Medellín, Colombia
[2]Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403-0221, USA

SYNOPTIC ABSTRACT

In this article, the exact distribution and exact percentage points for testing equality of q *bivariate Gaussian populations are obtained. The distribution has been derived using the inverse Mellin transformation and the residue theorem. The percentage points have been computed for* q = *2(1)5.*

Key Words and Phrases: inverse Mellin transformation, bivariate normal distribution, percentage points, residue theorem.

1. Introduction

Let $\boldsymbol{X}_{i1}, \ldots, \boldsymbol{X}_{iN_i}$ be a random sample from a bivariate normal population with mean vector $\boldsymbol{\mu}_i$ and covariance matrix Σ_i, $i = 1, \ldots, q$. Consider testing the null hypothesis $H : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_q = \boldsymbol{\mu}; \Sigma_1 = \cdots = \Sigma_q = \Sigma$, against general alternative K which says that H is not true. In H the common mean vector $\boldsymbol{\mu}$ and the common covariance matrix Σ are unknown. Let $\bar{\boldsymbol{X}}_i$ and A_i be the mean vector and covariance matrix formed from the ith population; that is $N_i\bar{\boldsymbol{X}}_i = \sum_{j=1}^{N_i} \boldsymbol{X}_{ij}$ and $A_i = \sum_{j=1}^{N_i} (\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}}_i)(\boldsymbol{X}_{ij} - \bar{\boldsymbol{X}}_i)'$, $i = 1, \ldots, q$, and put $N_0\bar{\boldsymbol{X}} = \sum_{i=1}^{q} N_i\bar{\boldsymbol{X}}_i$, $A = A_1 + \cdots + A_q$, $B = \sum_{i=1}^{q} N_i(\bar{\boldsymbol{X}}_i - \bar{\boldsymbol{X}})(\bar{\boldsymbol{X}}_i - \bar{\boldsymbol{X}})'$ and $N_0 = N_1 + \cdots + N_q$. The likelihood ratio test criterion Λ for testing H can be expressed as

$$\Lambda = \frac{N_0^{N_0} \prod_{i=1}^{q} \det(A_i)^{N_i/2}}{\prod_{i=1}^{q} N_i^{N_i} \det(A + B)^{N_0/2}}.$$

Address correspondence to Prof. Daya Nagar, Institute de Matamáticas, Universidad de Antioquia, Calle 67, No. 53-108, Medellín, Colombia. E-mail: dayaknagar@yahoo.com

This test statistic was first derived by Pearson and Wilks (1933). They also derived null moments of Λ and the distribution of Λ in special cases. It has been shown (e.g., see Perlman, 1980) that the test given by Λ is unbiased. The hth null moment of the test statistic Λ is given by

$$E(\Lambda^h) = \frac{N_0^{N_0 h}}{\prod_{i=1}^q N_i^{N_i h}} \frac{\Gamma(N_0 - 2)}{\Gamma[N_0(1+h) - 2]} \prod_{i=1}^q \frac{\Gamma[N_i(1+h) - 2]}{\Gamma(N_i - 2)}, \tag{1}$$

where $\mathrm{Re}(N_i h) > -(N_i - 2)$, $i = 1, \ldots, q$ and $\mathrm{Re}(\cdot)$ denotes the real part of $(\cdot)$. The asymptotic expansion of a constant multiple of $-2\ln\Lambda$ is available in Muirhead (1982).

In this article, we give the exact percentage points of $V_1 = \Lambda^{1/N}$ for $N_1 = \cdots = N_q = N$. The exact distribution of V_1 has been derived using the inverse Mellin transform and the residue theorem. We compute percentage points using the distributional results given in this article. The distribution and percentage points of the likelihood ratio statistic for testing homogeneity of several univariate Gaussian populations have been obtained by Nagar and Gupta (2004).

2. The Null Distribution

Substituting $N_1 = \cdots = N_q = N$, in (1) and using the Gauss–Legendre multiplication formula for gamma function, the hth moment of $V_1 = \Lambda^{1/N}$ is simplified as

$$E(V_1^h) = \frac{\Gamma^q(N - 2 + h)}{\Gamma^q(N - 2)} \prod_{k=0}^{q-1} \frac{\Gamma[N + (k-2)/q]}{\Gamma[N + h + (k-2)/q]}.$$

Now, using the inverse Mellin transform and the above moment expression, the density of V_1 is obtained as

$$f(v_1) = K(N, q)(2\pi\iota)^{-1} \int_C \frac{\Gamma^q(N - 2 + h)}{\prod_{k=0}^{q-1} \Gamma[N + h + (k-2)/q]} v_1^{-1-h}\, dh, \tag{2}$$

where $0 < v_1 < 1$, $\iota = \sqrt{-1}$, C is a suitable contour and

$$K(N, q) = \frac{\prod_{k=0}^{q-1} \Gamma[N + (k-2)/q]}{\Gamma^q (N-2)}.$$

Substituting $N - 2 + h = t$ and simplifying, the density (2) is restated as

$$f(v_1) = K(N, q)(2\pi\iota)^{-1} v_1^{N-3} \times \int_{C_1} \frac{\Gamma^{q-1}(t)}{t(t+1) \prod_{k=0(\neq 2)}^{q-1} \Gamma[t + 2 + (k-2)/q]} v_1^{-t}\, dt, \quad (3)$$

where $0 < v_1 < 1$ and C_1 is the changed contour. The poles of the integrand in (3) are given by $t = -j$, $j = 0, 1, 2, \ldots$, and each pole is of order $q-1$ except $t = 0, -1$, which are of order q. Hence by the residue theorem

$$f(v_1) = K(N, q) v_1^{N-3} \sum_{j=0}^{\infty} R_j, \; 0 < v_1 < 1, \quad (4)$$

where R_j is the residue at $t = -j$. From the calculus of residues, the residue at $t = -j$ for $j \geq 2$ is derived as

$$R_j = \frac{1}{(q-2)!} \lim_{t \to -j} \frac{\partial^{q-2}}{\partial t^{q-2}} \left[A_j v_1^{-t} \right], \quad (5)$$

where

$$A_j = \frac{(t+j)^{q-1} \Gamma^{q-1}(t)}{t(t+1) \prod_{k=0(\neq 2)}^{q-1} \Gamma[t + 2 + (k-2)/q]} = \frac{\Gamma^{q-1}(t+j+1)}{t(t+1) \prod_{\ell=0}^{j-1} (t+\ell)^{q-1} \prod_{k=0(\neq 2)}^{q-1} \Gamma[t + 2 + (k-2)/q]}.$$

By using the Leibnitz rule for successive differentiation of product of two functions, we can write the expression (5) as

$$R_j = \frac{1}{(q-2)!} \lim_{t \to -j} v_1^{-t} \sum_{u=0}^{q-2} \binom{q-2}{u} A_j^{(u)} (-\ln v_1)^{q-2-u}. \qquad (6)$$

Also,

$$\begin{aligned} A_j^{(u)} &= \frac{\partial^u}{\partial t^u} A_j = \frac{\partial^{u-1}}{\partial t^{u-1}} \left(\frac{\partial}{\partial t} A_j \right) = \frac{\partial^{u-1}}{\partial t^{u-1}} \left(A_j \frac{\partial}{\partial t} \ln A_j \right) \\ &= \frac{\partial^{u-1}}{\partial t^{u-1}} \left(A_j B_j \right), \qquad (7) \end{aligned}$$

where

$$\begin{aligned} B_j &= \frac{\partial}{\partial t} \ln A_j \qquad (8) \\ &= (q-1)\psi(t+j+1) - \frac{1}{t} - \frac{1}{t+1} - (q-1) \sum_{\ell=0}^{j-1} \frac{1}{t+\ell} \\ &\quad - \sum_{k=0(\neq 2)}^{q-1} \psi\left(t + 2 + \frac{k-2}{q} \right). \qquad (9) \end{aligned}$$

Consequently, all the derivatives of A_j can be obtained from the following recursive relation, which has been derived from (7) by using the Leibnitz rule. We have

$$A_j^{(u)} = \sum_{m=0}^{u-1} \binom{u-1}{m} A_j^{(u-1-m)} B_j^{(m)},$$

with

$$\begin{aligned} B_j^{(m)} &= \frac{\partial^m}{\partial t^m} B_j = \frac{\partial^{m+1}}{\partial t^{m+1}} \ln A_j \qquad (10) \\ &= (-1)^{m+1} m! \Bigg[(q-1)\zeta(m+1, t+j+1) + \frac{1}{t^{m+1}} + \frac{1}{(t+1)^{m+1}} \\ &\quad + (q-1) \sum_{\ell=0}^{j-1} \frac{1}{(t+\ell)^{m+1}} - \sum_{k=0(\neq 2)}^{q-1} \zeta\left(m+1, t+2+\frac{k-2}{q} \right) \Bigg]. \end{aligned}$$

(11)

In the expressions (9) and (11), $\psi(\cdot)$ and $\zeta(\cdot,\cdot)$ are well known psi- and Hurwitz zeta functions respectively (see Abramowitz & Stegun, 1992; Apostol, 2010; Askey & Roy, 2010). From (6) it is now possible to write

$$R_j = \frac{1}{(q-2)!} v_1^j \sum_{u=0}^{q-2} \binom{q-2}{u} A_{j0}^{(u)} (-\ln v_1)^{q-2-u}, \tag{12}$$

where

$$\begin{aligned} A_{j0}^{(u)} &= A_j^{(u)} \quad \text{at} \quad t=-j \\ &= \sum_{m=0}^{u-1} \binom{u-1}{m} A_{j0}^{(u-1-m)} B_{j0}^{(m)}, \end{aligned} \tag{13}$$

with

$$\begin{aligned} A_{j0}^{(0)} &= A_j^{(0)} \quad \text{at} \quad t=-j \\ &= \frac{(-1)^{(q-1)j}}{j(j-1)(j!)^{q-1}\prod_{k=0(\neq 2)}^{q-1}\Gamma[2-j+(k-2)/q]}, \end{aligned}$$

$$\begin{aligned} B_{j0}^{(0)} &= B_j^{(0)} \quad \text{at} \quad t=-j \\ &= (q-1)\psi(j+1) + \frac{1}{j} + \frac{1}{j-1} - \sum_{k=0(\neq 2)}^{q-1} \psi\left(2-j+\frac{k-2}{q}\right), \end{aligned}$$

and

$$\begin{aligned} B_{j0}^{(m)} &= B_j^{(m)} \quad \text{at} \quad t=-j \\ &= (-1)^{m+1} m! \Bigg[(q-1)\zeta(m+1,1) + \frac{1}{(-j)^{m+1}} + \frac{1}{(-j+1)^{m+1}} \\ &\quad + (q-1)\sum_{\ell=0}^{j-1} \frac{1}{(\ell-j)^{m+1}} - \sum_{k=0(\neq 2)}^{q-1} \zeta\left(m+1, 2-j+\frac{k-2}{q}\right) \Bigg]. \end{aligned}$$

Similarly, the residue at $t = -j$, $j = 0, 1$ is obtained as

$$R_j = \frac{1}{(q-1)!} v_1^j \sum_{u=0}^{q-1} \binom{q-1}{u} A_{j0}^{(u)} (-\ln v_1)^{q-1-u}, \tag{14}$$

where

$$A_{j0}^{(0)} = \frac{(-1)^{qj}}{\prod_{k=0(\neq 2)}^{q-1} \Gamma[2 - j + (k-2)/q]},$$

$$B_{j0}^{(0)} = (q-1)\psi(1) + qj + j - 1 - \sum_{k=0(\neq 2)}^{q-1} \psi\left(2 - j + \frac{k-2}{q}\right),$$

and

$$B_{j0}^{(m)} = (-1)^{m+1} m! \Bigg[(q-1)\zeta(m+1, 1) + q(-j)^{1+m} + 1 - j - \sum_{k=0(\neq 2)}^{q-1} \zeta\left(m+1, 2 - j + \frac{k-2}{q}\right)\Bigg].$$

Substituting (12) and (14) in (4), we get the density of V_1 as

$$f(v_1) = K(N, q) v_1^{N-3} \Bigg[\frac{1}{(q-1)!} \sum_{j=0,1} v_1^j \sum_{u=0}^{q-1} \binom{q-1}{u} A_{j0}^{(u)} (-\ln v_1)^{q-1-u} + \frac{1}{(q-2)!} \sum_{j=2}^{\infty} v_1^j \sum_{u=0}^{q-2} \binom{q-2}{u} \times A_{j0}^{(u)} (-\ln v_1)^{q-2-u} \Bigg], \tag{15}$$

where $0 < v_1 < 1$.

3. Computation

The computation of the percentage points has been carried out by using $F(v, q) = \int_0^v f(t)\,dt$ where $f(t)$ is given by (15). The computation is carried out by using the series representation given in (15). First, $F(v, q)$ is computed for various values of v. It is checked for monotonicity and for conditions $F(v, q) \to 0$ as $v \to 0$ and $F(v, q) \to 1$ as $v \to 1$. Then v is computed for various values of q, N and $F(v, q) = \alpha$. These are given in Tables 1–4. The tables are given for values of q from 2 to 5. We have used

TABLE 1 Table of the percentage points of V_1 for $q = 2$

N	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.1$
3	0.000901093	0.00263465	0.00606903	0.0143968
4	0.0276701	0.0475534	0.072533	0.112431
5	0.088825	0.127788	0.169741	0.228003
6	0.160494	0.211137	0.26157	0.326837
7	0.229695	0.286293	0.340054	0.406719
8	0.292211	0.351278	0.405628	0.471122
9	0.347361	0.406885	0.460419	0.523616
10	0.395671	0.454515	0.506535	0.567002
11	0.437995	0.495534	0.545721	0.603357
12	0.4752	0.531104	0.579343	0.634207
13	0.508058	0.562174	0.60846	0.660684
14	0.537227	0.589506	0.633891	0.68364
15	0.563258	0.613711	0.656279	0.703722
16	0.586606	0.63528	0.676126	0.721432
17	0.60765	0.654611	0.693836	0.737162
18	0.626704	0.672028	0.709731	0.751223
19	0.644028	0.687796	0.724073	0.763866
20	0.659844	0.702136	0.737076	0.775294
21	0.674336	0.71523	0.748919	0.785672
22	0.68766	0.727232	0.759748	0.795138
23	0.699951	0.738273	0.769687	0.803806
24	0.711321	0.748461	0.778841	0.811774
25	0.72187	0.757891	0.787299	0.819122
26	0.731682	0.766644	0.795136	0.825919
27	0.740831	0.77479	0.802419	0.832226
28	0.749381	0.78239	0.809204	0.838093
29	0.75739	0.789496	0.815541	0.843564
30	0.764905	0.796155	0.821471	0.848679

TABLE 2 Table of the percentage points of V_1 for $q = 3$

N	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.1$
3	0.0000590665	0.000197137	0.000511359	0.0013989
4	0.00578465	0.0107401	0.0175603	0.0295888
5	0.029086	0.0442891	0.0618996	0.0883674
6	0.0671487	0.0924337	0.119244	0.156375
7	0.112297	0.145368	0.178588	0.222349
8	0.15917	0.197672	0.234957	0.282439
9	0.204874	0.246932	0.286595	0.335886
10	0.248053	0.292288	0.333174	0.383054
11	0.288184	0.333613	0.374949	0.42465
12	0.325174	0.371109	0.412378	0.461423
13	0.35914	0.4051	0.445963	0.494061
14	0.390291	0.435944	0.47618	0.523162
15	0.418869	0.463986	0.503457	0.549231
16	0.445117	0.489546	0.528167	0.572694
17	0.469266	0.512907	0.550633	0.593904
18	0.49153	0.534319	0.571129	0.613161
19	0.5121	0.554001	0.589892	0.630713
20	0.531147	0.572144	0.607125	0.646772
21	0.548823	0.588912	0.623002	0.661515
22	0.565262	0.604451	0.637671	0.675096
23	0.580584	0.618885	0.651263	0.687643
24	0.594894	0.632326	0.663888	0.699269
25	0.608285	0.64487	0.675646	0.710071
26	0.620839	0.656601	0.686619	0.720131
27	0.632632	0.667594	0.696884	0.729523
28	0.643727	0.677916	0.706505	0.73831
29	0.654184	0.687625	0.715541	0.746549
30	0.664055	0.696773	0.724042	0.754288

Mathematica 7.0 to carry out these computations. To compute v for a given value of $\alpha = F(v, q)$, we have used `FindRoot` which searches for a numerical solution to the given equation, using Newton's method or a variant of the secant method. An eight-place accuracy is kept throughout. It may be noted here that for $q \geq 6$, the exact density of V can also be derived by applying the technique used in this article. But, the series expansions so obtained are extremely lengthy and therefore it is difficult to compute the percentage points. Hence, the tables are given for $q = 2, 3, 4$ and $q = 5$.

TABLE 3 Table of the percentage points of V_1 for $q = 4$

N	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.1$
3	4.97728×10^{-6}	0.0000185597	0.0000531041	0.000163148
4	0.00140554	0.00277975	0.00480586	0.00864993
5	0.010613	0.0169148	0.024611	0.0368412
6	0.0305724	0.0436293	0.0580986	0.0790881
7	0.0588413	0.0784817	0.0989837	0.127105
8	0.0919404	0.11714	0.142398	0.17575
9	0.12714	0.156709	0.185485	0.222451
10	0.162646	0.195501	0.226769	0.266105
11	0.197379	0.232625	0.265583	0.306369
12	0.23074	0.267661	0.301699	0.343269
13	0.262425	0.300465	0.335128	0.377001
14	0.292317	0.331047	0.365995	0.407829
15	0.320404	0.359496	0.394478	0.436028
16	0.346738	0.385942	0.420772	0.461867
17	0.371404	0.410529	0.445073	0.485593
18	0.394503	0.433405	0.467565	0.50743
19	0.416142	0.454713	0.488418	0.527575
20	0.436428	0.474587	0.507788	0.546205
21	0.455462	0.493151	0.525815	0.563474
22	0.473342	0.510517	0.542624	0.57952
23	0.490158	0.526789	0.558327	0.594461
24	0.505992	0.54206	0.573023	0.608405
25	0.520921	0.556415	0.586804	0.621444
26	0.535014	0.569927	0.599747	0.633661
27	0.548336	0.582667	0.611924	0.64513
28	0.560943	0.594696	0.6234	0.655915
29	0.572891	0.606069	0.634231	0.666075
30	0.584225	0.616837	0.644468	0.675661

4. Unequal Sample Sizes

When the sample sizes are unequal the problem is a bit complicated because of the presence of parameters $N_1, \ldots, N_q$ but still one can use the same technique considered in this article. Writing the gamma functions with the help of the Gauss–Legendre multiplication formula for gamma function, the expression in (1) can be rewritten as

$$E(\Lambda^h) = C\frac{\prod_{i=1}^{q}\prod_{k=0}^{N_i-1}\Gamma\left[h+1+(k-2)/N_i\right]}{\prod_{k=0}^{N_0-1}\Gamma\left[h+1+(k-2)/N_0\right]}, \qquad (16)$$

TABLE 4 Table of the percentage points of V_1 for $q = 5$

N	$\alpha = 0.01$	$\alpha = 0.025$	$\alpha = 0.05$	$\alpha = 0.1$
3	4.73662×10^{-7}	1.94749×10^{-6}	6.07185×10^{-6}	0.000020635
4	0.000366893	0.000766726	0.00139126	0.0026503
5	0.00407826	0.0067619	0.010185	0.0158772
6	0.0144969	0.0213412	0.0292016	0.0410354
7	0.0318815	0.0436296	0.0562758	0.0741842
8	0.0546414	0.0711613	0.0881766	0.11129
9	0.0808794	0.101616	0.122301	0.149562
10	0.109009	0.133276	0.156899	0.187326
11	0.137871	0.164993	0.190895	0.22366
12	0.166675	0.196049	0.223671	0.258106
13	0.194906	0.226018	0.254903	0.290482
14	0.222249	0.25467	0.284451	0.320766
15	0.24852	0.281901	0.312284	0.349019
16	0.273628	0.307684	0.338438	0.375348
17	0.297541	0.33204	0.362983	0.399882
18	0.320264	0.355021	0.386009	0.422754
19	0.341826	0.376693	0.407614	0.444096
20	0.362273	0.39713	0.427895	0.464033
21	0.381657	0.416408	0.44695	0.482683
22	0.400035	0.434605	0.464871	0.500153
23	0.417465	0.451792	0.481742	0.516541
24	0.434003	0.468041	0.497644	0.531939
25	0.449705	0.483416	0.512651	0.546426
26	0.464623	0.49798	0.526829	0.560077
27	0.478808	0.511788	0.540242	0.572958
28	0.492306	0.524893	0.552945	0.585129
29	0.505161	0.537345	0.56499	0.596646
30	0.517415	0.549186	0.576425	0.607557

where

$$C = \frac{\prod_{k=0}^{N_0-1} \Gamma\left[1 + (k-2)/N_0\right]}{\prod_{i=1}^{q} \prod_{k=0}^{N_i-1} \Gamma\left[1 + (k-2)/N_i\right]}. \tag{17}$$

Then the density of Λ is available as the sum of the residues at the poles of the gamma product in the numerator of the right-hand side of (16). Depending on the values of $N_1, \ldots, N_q$, some of the poles may be of higher orders, some of the factors may be canceled with the denominator factors, and soforth. Once $N_1, \ldots, N_q$ are known, the orders of the poles can be determined and the density can be written with the help of the residue theorem. Thus, the exact percentage points can be computed by using this result for the exact density. If q is large and if a number of the sample sizes are different, then it is more convenient to look for approximations. The well known Box's approximation is a good choice, because it gives an expansion of the cumulative distribution function of $-2 \ln \Lambda$ in series involving cumulative distribution functions of chi-square variables. The problem created by different sample sizes is completely eliminated here. This is usually good except for small values of q. Hence, the Box's approximation is recommended for $q > 5$ and $N_j > 5$ when the sample sizes are different.

Funding

Supported by the Sistema Universitario de Investigacion, Universidad de Antioquia under the project no. IN10133CE.

References

Abramowitz, M., & Stegun, I. A. (1992). *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* (Reprint of the 1972 ed.). New York, NY: Dover Publications, Inc.

Apostol, T. M. (2010). Hurwitz zeta function. *NIST Handbook of Mathematical Functions.* Edited by Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark, U. S. Department of Commerce, National Institute of Standards and Technology. Cambridge: Cambridge University Press.

Askey, R. A., & Roy, R. (2010). Polygamma functions. *NIST Handbook of Mathematical Functions.* Edited by Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert and Charles W. Clark, U. S. Department of Commerce, National Institute of Standards and Technology, Washington, DC; Cambridge: Cambridge University Press.

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory.* New York, NY: John Wiley & Sons.

Nagar, D. K., & Gupta, A. K. (2004). Percentage points for testing homogeneity of several univariate Gaussian populations. *Applied Mathematics and Computations, 156,* 551–561.

Pearson, E. S., & Wilks, S. S. (1933). Methods of statistical analysis appropriate for *k* samples of two variables. *Biometrika, 25*(3–4), 353–337.

Perlman, M. D. (1980). Unbiasedness of the likelihood ratio tests for testing equality of several covariance matrices and equality of several multivariate normal populations. *Annals of Statistics, 8,* 247–263.