Distributional inference

Albers, Casper Johannes

Publication date:
2003

Appendices

Appendix A
Proof of Theorem 4.1

The bivariate normality follows from general U-statistic theory, and the central limit theorem. Theorem 4.1 is essentially equal to the following three lemmata.

Lemma A.1 *Under* $H_0 : f = \psi$ *we have* $\mathrm{Var}\ \varepsilon = \frac{1}{n}\sigma^2 = \frac{1}{12n}$.

Lemma A.2 *Under* $H_0 : f = \psi$ *we have*

$$\mathrm{Var}\ \delta = \frac{1}{45}\frac{n+3}{n(n-1)} \to \tau^2 = \frac{1}{45}.$$

Lemma A.3 *Under* $H_0 : f = \psi$ *we have* $\mathrm{Cov}\ (\varepsilon, \delta) = 0$.

Proof of Lemma A.1
If $Y_i \sim$U$(0,1)$ i.i.d. then $\mathrm{Var}\ (Y_i - \frac{1}{2}) = \frac{1}{12}$ and $\mathrm{Var}\ (\frac{1}{n}\sum(Y_i - \frac{1}{2})) = \frac{1}{12n}$. □

Proof of Lemma A.2
This result can be found in NAIR (1936)[1] Alternatively, it can be derived along the lines of HOEFFDING (1948, p. 308) and FRASER (1959, p.230), which will be repeated here. The variance of Gini's mean difference can be written as

$$\mathrm{Var}\ \ g = \frac{2}{n(n-1)}[2\zeta_1(n-2) + \zeta_2]$$

with[2]

$$\begin{aligned}
\Delta &= \int_0^1 \int_0^1 |y_1 - y_2|\ \mathrm{d}y_1\ \mathrm{d}y_2 = \frac{1}{3} \\
\zeta_1 &= \int_0^1 [\int_0^1 |y_1 - y_2|\ \mathrm{d}y_2]^2\ \mathrm{d}y_1 - \Delta^2 \\
&= \int_0^1 (y_1^2 - y_1 + \frac{1}{2})^2\ \mathrm{d}y_1 - \Delta^2 = \frac{7}{60} - \frac{1}{9} = \frac{1}{180} \\
\zeta_2 &= \int_0^1 \int_0^1 (y_1 - y_2)^2\ \mathrm{d}y_1\ \mathrm{d}y_2 - \Delta^2 \\
&= 2\mathrm{Var}\ y_i - \left(\tfrac{1}{3}\right)^2 = \tfrac{1}{18}.
\end{aligned}$$

[1] LOMNICKI (1952) showed that there was an error in Nair's formulations, though in the uniform case the results were true (cf. KENDALL AND STUART (1958, p. 241–242)).

[2] Note that the formula for ζ_1 in FRASER (1959, p.230) is wrong.

So,

$$\begin{aligned}\text{Var } g &= \frac{2}{n(n-1)}\left(\frac{2(n-2)}{180}+\frac{1}{18}\right)=\frac{1}{45}\frac{n+3}{n(n-1)}\\ \tau^2 &= \lim_{n\to\infty} n \text{ Var } g = \frac{1}{45}.\end{aligned}$$ $\square$

Proof of Lemma A.3
Approach 1
Of course, Cov (ε, δ) = Cov $(\bar{u}, g)$. That Cov $(\bar{u}, g) = 0$ is easily seen with this symmetry-argument: define $y_i^* = 1 - y_i$. Then, of course, $\bar{u^*} = 1 - \bar{u}$ is also transformed, yet $g^* = g$ remains the same. By definition

$$\text{Cov } (\bar{u}, g) = \text{Cov } (1-\bar{u}^*, g^*) = -\text{Cov } (\bar{u}^*, g^*).$$

Therefore, the covariance is equal to zero.

Approach 2
The covariance between $\bar{u}$ and g can be written as a function of the covariances of the order statistics,

$$\begin{aligned}\text{Cov } (\bar{u}, g) &= \text{Cov } \left(\sum \frac{u_{[i]}}{n}, \sum \frac{2(2i-n-1)}{n(n-1)} u_{[i]}\right)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{2(2j-n-1)}{n^2(n-1)} \text{Cov } (u_{[i]}, u_{[j]})\\ &= \frac{2(2j-n-1)}{n^2(n-1)} \sum_{i=1}^{n}\sum_{j=1}^{n} (2j-n-1) \text{Cov } (u_{[i]}, u_{[j]})\end{aligned}$$

To compute the Cov $(u_{[i]}, u_{[j]})$, we need the joint distribution function of $U_{[i]}$ and $U_{[j]}$, which is, according to WILKS (1962, eq. 7.7.6), for $i < j$

$$\begin{aligned}f_{U_{[i]},U_{[j]}}(u,v) &= c_1 u^{i-1}(v-u)^{j-i-1}(1-v)^{n-j}\\ \text{with } c_1 &= \tfrac{\Gamma(n+1)}{\Gamma(i)\Gamma(j-i)\Gamma(n+1-j)}\end{aligned}$$

and $0 < u < v < 1$, $1 \leq i < j \leq n$. This joint probability density function is that of the ordered bivariate Dirichlet$(i, j-i, n+1-j)$ distribution. (This result can also be found using KENDALL AND STUART, 1958, p. 325, who provides the formula for the joint distribution for general F.) Hence

$$\begin{aligned}\text{Cov } (u_{[i]}, u_{[j]}) &= \mathbf{E}\ u_{[i]}u_{[j]} - \mathbf{E}\ u_{[i]}\mathbf{E}\ u_{[j]}\\ &= \mathbf{E}\ u_{[i]}u_{[j]} - \frac{ij}{(n+1)^2}.\end{aligned}$$

The expectation of the product being

$$\begin{aligned}
\mathbf{E}\, U_{[i]}U_{[j]} &= \int_0^1 \int_0^v uv f_{U_{[i]},U_{[j]}}(u,v)\,\mathrm{d}u\,\mathrm{d}v \\
&= c_1 \int_0^1 v(1-v)^{n-j}\left[\int_0^v u^i(v-u)^{j-i-1}\,\mathrm{d}u\right]\mathrm{d}v \\
&= c_1 \int_0^1 v(1-v)^{n-j}\left[v^s \mathrm{Beta}(r+1,s-r)\right]\mathrm{d}v \\
&= c_1c_2 \int_0^1 v^{j+1}(1-v)^{n-j}\,\mathrm{d}v \\
&= \tfrac{\Gamma(n+1)}{\Gamma(i)\Gamma(j-i)\Gamma(n-j+1)} \tfrac{\Gamma(i+1)\Gamma(j-i)}{\Gamma(j+1)} \tfrac{\Gamma(j+2)\Gamma(n-j+1)}{\Gamma(n+3)} \\
&= \tfrac{i(j+1)}{(n+1)(n+2)}.
\end{aligned}$$

That $c_2 = \mathrm{Beta}(r+1, s-r)$ can be seen through

$$\int_0^v u^i(v-u)^{j-i-1}\,\mathrm{d}u = v^j \int_0^1 w^i(1-w)^{j-i-1}\,\mathrm{d}w = \mathrm{B}(i+1, j-i)v^j.$$

Using $\mathbf{E}\, U_{[i]}U_{[j]}$, we can calculate the covariance

$$\mathrm{Cov}\,(U_{[i]}U_{[j]}) = \begin{cases} \tfrac{i(j+1)}{(n+1)(n+2)} - \tfrac{ij}{(n+1)^2} = \tfrac{i(n+1-j)}{(n+1)^2(n+2)} & \text{if } i < j \\ \tfrac{i(n+1-i)}{(n+1)^2(n+2)} & \text{if } i = j \\ \tfrac{j(n+1-i)}{(n+1)^2(n+2)} & \text{if } i > j \end{cases}$$

Note that the formula for $i > j$ immediately follows from that for $i < j$ by swapping the i and j. The formula for $i = j$ is easily derived by having that Cov $(U_{[i]}U_{[j]}) =$ Var $(U_{[i]})$ and $U_{[i]} \sim \mathrm{Beta}(i, n+1-i)$. Note that this formula coincides with the formula for $i < j$ as well as the formula for $i > j$. The final step is to use all these covariances to derive that Cov $(\bar{u}, g) = 0$, and, hence, $\rho = 0$. Cov $(\bar{u}, g) =$

$$\begin{aligned}
&= \frac{2}{n^2(n-1)} \sum_{j=1}^n \sum_{i=1}^n (2j-n-1)\mathrm{Cov}\,(u_{[i]}, u_{[j]}) \\
&= \tfrac{2}{(n-1)n^2(n+1)^2(n+2)} \left(\sum_{j=1}^n \sum_{i=1}^{j-1} (2j-n-1)i(n+1-j) + \right. \\
&\quad \left. + \sum_{j=1}^n (2j-n-1)j(n+1-j) + \sum_{j=1}^n \sum_{i=j+1}^n (2j-n-1)j(n+1-i) \right) \\
&= \tfrac{2}{(n-1)n^2(n+1)^2(n+2)} \left(\tfrac{(n-1)n(n+1)(n+2)(n+3)}{120} + 0 - \tfrac{(n-1)n(n+1)(n+2)(n+3)}{120} \right) \\
&= 0 \qquad \square
\end{aligned}$$

Appendix B
Critical values for the $||f_n^{(m)} - \psi||_1$-test statistic

The Tables B.1, B.2, and B.3 are each constructed on the basis of the following simulation experiment. For $m = 2$, 3, and 4, and for each value of n, a sample of size n was drawn from the standard uniform distribution providing an outcome $t_n^{(m)}$ of the test statistic $T_n^{(m)}$. This process is repeated 100 000 times. (Simulation studies showed that this number of replications suffices to obtain reliable critical values.) The percentiles are taken from the empirical distribution of $T_n^{(m)}$. (Note: these values are computed using the true $t_n^{(m)}$, not the approximation mentioned in Section 4.4 for $m = 2$.)

n	.90	.75	.50	.25	.10	.05	.025	.01
5	.0499	.0838	.1422	.2311	.3241	.3801	.4302	.4865
6	.0439	.0746	.1283	.2093	.2949	.3482	.3954	.4466
7	.0394	.0675	.1177	.1934	.2717	.3215	.3641	.4133
8	.0365	.0627	.1102	.1807	.2539	.3007	.3436	.3906
9	.0336	.0582	.1029	.1700	.2401	.2842	.3224	.3664
10	.0320	.0552	.0979	.1613	.2276	.2684	.3055	.3473
12	.0286	.0494	.0889	.1468	.2075	.2463	.2802	.3207
14	.0259	.0455	.0818	.1357	.1914	.2268	.2574	.2950
16	.0241	.0424	.0765	.1274	.1798	.2125	.2415	.2769
18	.0225	.0395	.0718	.1198	.1700	.2013	.2286	.2620
20	.0212	.0374	.0685	.1132	.1604	.1899	.2164	.2486
22	.0201	.0354	.0651	.1085	.1531	.1816	.2065	.2379
24	.0191	.0339	.0623	.1037	.1471	.1741	.1990	.2283
26	.0185	.0326	.0596	.0995	.1406	.1667	.1908	.2178
28	.0176	.0312	.0576	.0962	.1361	.1613	.1838	.2118
30	.0171	.0303	.0555	.0922	.1310	.1562	.1788	.2050
35	.0155	.0278	.0512	.0849	.1209	.1434	.1633	.1890
40	.0147	.0262	.0482	.0798	.1133	.1345	.1538	.1759
45	.0137	.0244	.0454	.0754	.1066	.1266	.1452	.1677
50	.0130	.0232	.0428	.0715	.1014	.1211	.1379	.1580
60	.0118	.0211	.0391	.0651	.0925	.1101	.1259	.1437
70	.0109	.0195	.0362	.0607	.0856	.1015	.1160	.1331
80	.0103	.0183	.0339	.0564	.0802	.0954	.1084	.1248
90	.0096	.0172	.0319	.0532	.0756	.0898	.1024	.1175
100	.0090	.0162	.0302	.0506	.0718	.0852	.0971	.1115
120	.0082	.0148	.0276	.0459	.0653	.0777	.0889	.1024
140	.0077	.0138	.0257	.0425	.0606	.0722	.0823	.0940
160	.0072	.0129	.0238	.0399	.0567	.0675	.0770	.0881
180	.0067	.0121	.0225	.0378	.0535	.0637	.0729	.0843
200	.0064	.0115	.0215	.0355	.0504	.0600	.0688	.0788
250	.0057	.0103	.0190	.0319	.0454	.0538	.0614	.0708
300	.0052	.0093	.0174	.0291	.0414	.0493	.0562	.0646
400	.0045	.0081	.0151	.0254	.0360	.0429	.0490	.0560
500	.0040	.0072	.0135	.0226	.0321	.0382	.0436	.0498

Table B.1: Critical values for $m = 2$

n	.90	.75	.50	.25	.10	.05	.025	.01
5	.0784	.1267	.1957	.2899	.3976	.4656	.5253	.5930
6	.0684	.1119	.1747	.2611	.3595	.4228	.4776	.5393
7	.0618	.1007	.1590	.2411	.3331	.3915	.4422	.5001
8	.0565	.0924	.1469	.2246	.3101	.3641	.4131	.4730
9	.0517	.0860	.1376	.2114	.2932	.3456	.3918	.4444
10	.0491	.0811	.1302	.1992	.2767	.3265	.3712	.4234
12	.0440	.0728	.1176	.1817	.2526	.2980	.3398	.3879
14	.0397	.0665	.1081	.1679	.2336	.2760	.3137	.3592
16	.0366	.0616	.1003	.1567	.2188	.2584	.2938	.3360
18	.0344	.0576	.0947	.1477	.2061	.2440	.2779	.3180
20	.0327	.0546	.0892	.1397	.1952	.2310	.2625	.3015
22	.0310	.0519	.0851	.1327	.1863	.2204	.2512	.2869
24	.0293	.0494	.0812	.1272	.1777	.2103	.2403	.2735
26	.0284	.0474	.0778	.1223	.1714	.2027	.2302	.2645
28	.0270	.0454	.0747	.1182	.1642	.1949	.2228	.2554
30	.0259	.0437	.0723	.1137	.1590	.1880	.2147	.2465
35	.0240	.0403	.0668	.1051	.1475	.1748	.1992	.2287
40	.0223	.0375	.0623	.0985	.1380	.1640	.1870	.2144
45	.0210	.0353	.0585	.0923	.1287	.1532	.1749	.2009
50	.0198	.0336	.0557	.0877	.1230	.1457	.1657	.1889
60	.0181	.0306	.0508	.0799	.1119	.1329	.1516	.1733
70	.0165	.0282	.0471	.0743	.1043	.1237	.1405	.1609
80	.0155	.0263	.0440	.0694	.0972	.1155	.1318	.1519
90	.0147	.0248	.0414	.0655	.0917	.1085	.1232	.1411
100	.0138	.0232	.0390	.0619	.0868	.1032	.1172	.1343
120	.0126	.0214	.0357	.0565	.0794	.0940	.1073	.1234
140	.0117	.0197	.0330	.0523	.0735	.0870	.0992	.1142
160	.0108	.0185	.0310	.0488	.0687	.0814	.0924	.1058
180	.0102	.0173	.0291	.0461	.0649	.0767	.0871	.0996
200	.0097	.0165	.0276	.0439	.0616	.0730	.0832	.0954
250	.0087	.0148	.0248	.0394	.0552	.0652	.0744	.0854
300	.0079	.0133	.0225	.0357	.0502	.0594	.0677	.0774
400	.0068	.0116	.0195	.0310	.0435	.0516	.0587	.0673
500	.0061	.0104	.0174	.0278	.0389	.0461	.0525	.0600

Table B.2: Critical values for $m = 3$