The Annals of Statistics
1973, Vol. 1, No. 6, 1185–1188

THE ASYMPTOTIC EXPANSION OF THE DISTRIBUTION OF THE GOODNESS OF FIT STATISTIC V_{NN}[1]

BY URS R. MAAG

Université de Montréal

The complete asymptotic expansion in powers of $N^{-\frac{1}{2}}$ for the distribution of the two-sample statistic V_{NN} is derived under the null hypothesis. The proof is based on methods developed by Kemperman (1959).

1. Introduction and summary. Consider two independent random samples of sizes M and N from the same unknown but continuous distribution $F(x)$. Let $F_M(x)$ and $G_N(x)$ be the corresponding empirical distribution functions. Kuiper (1960) defined

$$V_{MN} = \sup_{-\infty<x<\infty} (F_M(x) - G_N(x)) - \inf_{-\infty<x<\infty} (F_M(x) - G_N(x)) \tag{1}$$

as a two-sample statistic suitable for tests of homogeneity for distributions on a circle. Tables and results on the distribution theory of V_{MN} are given by Kuiper (1960), Maag and Stephens (1968) and Steck (1969). The following formula for the exact distribution of V_{NN} was derived in Maag and Stephens (1968):

$$\begin{aligned}\Pr(V_{NN} < c/N) \\ = 2^{2N+1}\binom{2N}{N}^{-1}\{\textstyle\sum_{k=1}^{[\frac{1}{2}c]} (\cos k\pi/(c+1))^{2N} - \sum_{k=1}^{[\frac{1}{2}(c-1)]} (\cos k\pi/c)^{2N}\}\end{aligned} \tag{2}$$

where c can take the values $2, 3, \cdots, N+1$ and $[x]$ stands for the greatest integer not exceeding x.

2. Results. The following notation will be used: Let

$$A_{\nu-1} = \frac{2^{2\nu}(2^{2\nu}-1)}{2\nu(2\nu)!} B_\nu \qquad (\nu = 1, 2, \cdots)$$

where $B_\nu > 0$ denotes the νth Bernoulli number ($B_1 = \frac{1}{6}, B_2 = \frac{1}{30}, B_3 = \frac{1}{42}$, etc.). Let

$$A_{sh} = \sum A_1^{\nu_1} \cdots A_s^{\nu_s}(\nu_1! \cdots \nu_s!)^{-1}$$

where the summation is extended over all the sets $(\nu_1, \cdots, \nu_s)$ of nonnegative integers which satisfy $\nu_1 + \cdots + \nu_s = h$ and $\nu_1 + 2\nu_2 + \cdots + s\nu_s = s$, and finally let

$$g_r(x) = \textstyle\sum_{k=1}^{\infty} (2xk^2)^r e^{-k^2x} \qquad (r = 0, 1, 2, \cdots). \tag{3}$$

Received June 1971; revised March 1973.

[1] Research supported in part by the National Research Council of Canada and carried out while the author was on leave at the Forschungsinstitut für Mathematik of the Swiss Federal Institute of Technology Zürich.

AMS 1970 *subject classifications.* Primary 41A60, 62E20; Secondary 62G10, 62G30.

Key words and phrases. Asymptotic expansion, goodness of fit test, distribution-free statistic, two-sample statistic.

THEOREM. *For any positive integer m we have*

$$\Pr(V_{NN} < c/N) = 2^{2N+1}\binom{2N}{N}^{-1}\Big\{\sum_{s=0}^{m-1}(2N)^{-s}\sum_{h=0}^{s}(-1)^h A_{sh} \times\left[g_{s+h}\left(\frac{N\pi^2}{(c+1)^2}\right) - g_{s+h}\left(\frac{N\pi^2}{c^2}\right)\right] + O(N^{-m+\frac{1}{2}})\Big\} \tag{4}$$

where the remainder holds uniformly in c. A more precise estimate of the remainder in terms of N and c is $O(cN^{-m-\frac{1}{2}})$.

PROOF. Since the proof follows closely the work of Kemperman (1959), only the principal steps are stated and his notation is used wherever possible. Define

$$S_c = \sum_{k=1}^{[\frac{1}{2}(c-1)]}(\cos k\pi/c)^{2N}, \qquad \beta = N\pi^2/c^2, \qquad \sigma = -2/N \quad \text{and} \quad \tau = 1/N,$$

and let

$$\varphi(w) = (-\log\cos w^{\frac{1}{2}} - w/2)w^{-2}.$$

Kemperman's (1959) Lemma 1 says that, for the appropriate branch of $-\log\cos w^{\frac{1}{2}}$ and w real, positive and less than $\pi^2/4$, $\varphi(w) \geqq 0$, and that the Taylor expansion

$$e^{u\varphi(w)} = \sum_{s=0}^{\infty}\sum_{h=0}^{s} A_{sh}\, u^h w^{s-h}$$

holds for arbitrary u and $|w| < \pi^2/4$.

Observe that

$$S_c = \sum_{k=1}^{[\frac{1}{2}(c-1)]} e^{-\beta k^2}\exp[\sigma(\beta k^2)^2\varphi(\tau\beta k^2)]$$

and that every term of S_c has $\exp[-\beta k^2]$ as an upper bound since $\sigma < 0$ and $\varphi(\tau\beta k^2) \geqq 0$.

Writing $S_c = S_c' + S_c''$ where S_c' contains the terms with $1 \leqq k \leqq \lambda$ and S_c'' the terms with $\lambda < k \leqq [\frac{1}{2}(c-1)]$ permits us to obtain the following bounds:

(i) For an arbitrarily chosen positive number a, let $\lambda = a^{\frac{1}{2}}c\pi^{-1}N^{-\frac{1}{4}}$. Then

$$S_c'' \leqq (\tfrac{1}{2}c)e^{-aN^{\frac{1}{2}}} \leqq (N/2)e^{a-N^{\frac{1}{2}}}.$$

(ii) To obtain an approximation for S_c' we shall apply Kemperman's (1959) Lemma 4 with $x = 0$, $p = 2$, $q = 1$ and $s = 0$. Let

$$T_m = \sum_{s=0}^{m-1}\sum_{h=0}^{s} A_{sh}\,\sigma^h\tau^{s-h}\sum_{k=1}^{\infty} e^{-\beta k^2}(\beta k^2)^{s+h}. \tag{5}$$

Choose $u_1 \geqq 2a^2$, w_1 real and such that $0 < w_1 < \pi^2/4$, and finally choose $N \geqq \max[a^{-2}, a^2 w_1^{-2}]$. These choices of λ, u_1, w_1 and N now guarantee that the conditions of Kemperman's lemma hold, i.e.

$$\beta\lambda^2 \geqq 1, \qquad |\sigma|(\beta\lambda^2)^2 \leqq u_1 \qquad \text{and} \qquad |\tau|(\beta\lambda^2) \leqq w_1.$$

Under these conditions

$$|S_c' - T_m| \leqq K\beta^{-\frac{1}{2}}[e^{-\beta\lambda^2}(\beta\lambda^2)^{\frac{1}{2}} + e^{-\beta}|\sigma|^m(1 + \beta^{2m+\frac{1}{2}}) + e^{-\beta}|\tau|^m(1 + \beta^{m+\frac{1}{2}})].$$

Since $e^{-\beta}(1 + \beta^{2m+\frac{1}{2}})$ and $e^{-\beta}(1 + \beta^{m+\frac{1}{2}})$ are bounded there exists a constant K' such that

$$|S_c - T_m| \leqq K'[Ne^{-aN^{\frac{1}{2}}} + cN^{-m-\frac{1}{2}}] \tag{6}$$

for any choice of $c \leqq N$ and $N \geqq N_0$. Rewriting T_m in terms of N, c and g_r completes the proof.

LEMMA. *For any positive integer* q

$$(7)\qquad g_{s+h}\left(\frac{N\pi^2}{(c+1)^2}\right) - g_{s+h}\left(\frac{N\pi^2}{c^2}\right) = \sum_{j=1}^{q-1}\left(\frac{c+\frac{1}{2}}{(c+1)^2}\right)^j \frac{1}{j!}\sum_{\nu=0}^{j} a(j,\nu,s+h)g_{s+h+\nu}\left(\frac{N\pi^2}{c^2}\right) + O\left\{\left(\frac{(c+\frac{1}{2})N}{(c+1)^2c^2}\right)^q\right\}.$$

The coefficients $a(j, \nu, s+h)$ *are independent of* c *and* N; *they are given by the formula*

$$(8)\qquad a(j,\nu,r) = \binom{j}{\nu}(-1)^{j-\nu}2^{j-\nu}r_{(j-\nu)}$$

where $r_{(p)}$ *stands for* $r(r-1)\cdots(r-p+1)$ *with* $r_{(0)} = 1$. *They can also be calculated recursively from*

$$(9)\qquad a(j,\nu,r) = a(j-1,\nu-1,r) - 2(r+\nu-j+1)a(j-1,\nu,r)$$

with the initial values $a(0,0,r) = 1$ *and* $a(j,\nu,r) = 0$ *if* $\nu < 0$ *or if* $\nu > j$. *The first coefficients are* $a(1,0,r) = -2r$, $a(2,0,r) = 4r(r-1)$, $a(2,1,r) = -4r$ *and* $a(j,j,r) = 1$.

PROOF. The functions $g_r(x)$ and their derivatives are bounded for positive x with the Taylor series

$$g_r(x') - g_r(x) = \sum_{j=1}^{q-1}\frac{(x'-x)^j}{j!}\left(\frac{d}{dx}\right)^j g_r(x) + O\{(x'-x)^q\}.$$

Furthermore, they have the property that $(2x)^j(d/dx)^j g_r(x)$ can be written as a linear combination of $g_r(x), g_{r+1}(x), \cdots, g_{r+j}(x)$. Here $x' = N\pi^2(c+1)^{-2}$, $x = N\pi^2c^{-2}$ and thus $x' - x = (-2N\pi^2c^{-2})(c+\frac{1}{2})(c+1)^{-2}$, i.e., $x'-x$ contains $2x$ as a factor.

For large N only values of c of the form $c = xN^{\frac{1}{2}}$ yield probabilities which are sufficiently different from zero and one. In order to obtain a useable formula for the distribution of $N^{\frac{1}{2}}V_{NN}$ one applies to the Theorem

(i) the Lemma with $q = 2m - 2c$,

(ii) the expansion

$$\left[\frac{c+\frac{1}{2}}{(c+1)^2}\right]^j = c^{-j}\left[1 + \sum_{t=1}^{\infty} c^{-t}\sum_{i=0}^{\min(t,j)} 2^{-i}\binom{j}{i}\binom{-2j}{t-i}\right]$$

with $c = xN^{\frac{1}{2}}$, and

(iii) the expansion

$$2^{2N+1}\binom{2N}{N}^{-1} = 2(\pi N)^{\frac{1}{2}}[1 + (8N)^{-1} + (128N^2)^{-1} + \cdots].$$

For example, the case $m = 1$ yields

$$\begin{aligned}\Pr(N^{\frac{1}{2}}V_{NN} < x) &= 2(\pi N)^{\frac{1}{2}}[1 + O(N^{-1})]\{A_{00}[x^{-1}N^{-\frac{1}{2}}(1 + O(N^{-\frac{1}{2}}))][a(1, 0, 0)g_0(\pi^2x^{-2}) \\ &\qquad + a(1, 1, 0)g_1(\pi^2x^{-2}) + O(N^{-1})] + O(N^{-1})\} \\ &= 4\pi^{\frac{5}{2}}x^{-3}\textstyle\sum_{k=1}^{\infty} k^2 \exp[-k^2\pi^2/x^2] + O(N^{-\frac{1}{2}})\end{aligned}$$

which is the limiting distribution. For $m = 2$ the result agrees with formula (11) in Maag and Stephens (1968). The remainder which becomes $O(N^{-m+\frac{1}{2}})$ holds uniformly in x for bounded values of x.

3. Remark. The method of Kuiper (1960, formula (4.3)), which led to the correct limiting term but no term in $N^{-\frac{1}{2}}$ and an incorrect one in N^{-1}, can be adjusted to yield the correct term in $N^{-\frac{1}{2}}$ and the order of the remainder. Kuiper approximated

$$\Pr(N^{\frac{1}{2}}V_{NN} < c) = \textstyle\sum_{a^*=1}^{c^*}[P_N(a, c - a + N^{-\frac{1}{2}}) - P_N(a, c - a)]$$

where $P_N(a, b) = \Pr(-b < N^{\frac{1}{2}}(F_M(x) - G_N(x)) < a$ for all $x)$, $a = a^*N^{-\frac{1}{2}}$ and $c = c^*N^{-\frac{1}{2}}$, by

$$\int_0^c \frac{\partial}{\partial y} P_N(a, y)|_{y=c-a}\, da$$

but neglected the error terms of order $N^{-\frac{1}{2}}$ which this approximation introduces.

Acknowledgment. The author wishes to thank Professor J. H. B. Kemperman for suggesting the approach taken here. He is also grateful to a referee for his detailed and helpful report.

REFERENCES

[1] KEMPERMAN, J. H. B. (1959). Asymptotic expansions for the Smirnov test and for the range of cumulative sums. *Ann. Math. Statist.* **30** 448–462.

[2] KUIPER, N. H. (1960). Tests concerning random points on a circle. *Nederl. Akad. Wetensch. Proc. Ser. A* **63** 38–47.

[3] MAAG, U. R. and STEPHENS, M. A. (1969). The V_{NM} two-sample test. *Ann. Math. Statist.* **39** 923–935.

[4] STECK, G. P. (1969). The Smirnov two sample tests as rank tests. *Ann. Math. Statist.* **40** 1449–1466.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MONTREAL
P. O. BOX 6128
MONTREAL 101, CANADA