PREAMBLE

True eloquence consists in saying all that is necessary,
and nothing but what is necessary.

La Rouchefoucauld

Any scholar who, like myself, has experienced pure mathematics will recognize in this paper the illumination of formal mathematics that so often is missing in applied statistics–definitions, theorems, proofs, lemmas, and analytics. If you have never encountered these constructions before in an advanced mathematics context, then you may find this paper difficult reading; but, it certainly is eloquent! The purpose of this paper is twofold, namely to derive (1) those formulae needed to compute the exact distributions of the Moran and Geary spatial autocorrelation indices under an assumption of normality, and (2) expressions for their asymptotic mean and variance under an assumption of non-normality. Mardia's reaction to the formalism employed here is that the results are theoretically impressive, but not yet helpful to practitioners. Perhaps Sen has been a bit too concise. On the other hand, Mardia suggests that this paper should generate further research of an applied nature.

The Editor

Distribution of Spatial Correlation Statistics

Ashish Sen*

School of Urban Planning and Policy and the Urban Transportation Center, University of Illinois at Chicago, Chicago, IL 60680, U. S. A.

Overview: This paper contains a number of results on the distribution of Moran and Geary statistics. Two key ones are (1) formulæ to compute the exact distributions of these statistics when observations are normally distributed, and (2) expressions for their asymptotic mean and variance, when the observations are not normal. The results are sufficiently general that they may be applied to a wide range of situations. However, in order to be somewhat specific, the presentation assumes that the spatial correlation statistics are being applied to linear least squares residuals.

1. Introduction

Two statistics that may be applied on regression residuals $e_1, \cdots, e_n$ to test for the existence of spatial correlation are of the form

$$c \sum_{ij} w_{ij} e_i e_j / s^2 \tag{1}$$

and

$$c \sum w_{ij} (e_i - e_j)^2 / s^2 \tag{2}$$
$$ij$$

where c is a suitable constant, $s^2 = (n-k-1)^{-1} \sum_{i=1}^{n} e_i^2$ is the usual unbiased estimate of the variance of the regression error term when there are k independent variables, and w_{ij} is some measure of inverse distance. For example, we could have $w_{ij} = 1$ when the ith and jth observations are from contiguous zones and $w_{ij} = 0$ otherwise; or $w_{ij} = d_{ij}^{-2}$ where d_{ij} is the distance between the locations where observations i and j were taken. The expressions (1) and (2) are obvious generalizations of statistics previously given by Moran and Geary, respectively (see Cliff and Ord, 1981).

In Section 3 we present exact distributions for these statistics under the hypothesis of no spatial correlation and under the assumption of normality of regression errors. While the formulæ are somewhat complicated and depend on the matrix of w_{ij}'s, they can be programmed so that for any given situation, relevant portions of tables can be obtained from a computer. They also can be used to obtain exact tail probabilities to help identify suitable approximate methods for obtaining such probabilities. In this context, it should be mentioned that computations using these exact formulæ are certainly less time consuming than Monte Carlo methods.

If the errors are not normal, the only recourse available is to invoke large sample theory and use the fact that under certain mild conditions (1) and (2) are asymptotically normal.

* I would like to thank Prof. Tony Smith, Regional Science Department, University of Pennsylvania, Prof. Muni Srivastava, Statistics Department, University of Toronto, and the discussant for many valuable comments and suggestions on an earlier draft of this paper. I would also like to express my gratitude to Ms. Marilyn Engwall for the diagrams. (School of Urban Planning and Policy and the Urban Transportation Center, University of Illinois at Chicago.)

However, we still need means and variances of the statistics, and the asymptotic expressions for these (in, say Sen, 1976) are a bit too crude. In Section 4 we present formulæ for such means and variances. It may be mentioned in passing that when the observations are non-normal, and sample sizes small, the analyst may wish to use rank-test equivalents of (1) and (2) — see Sen and Soot (1977) — or obtain critical points of (1) and (2) by permutation methods (see Cliff and Ord, 1981).

The next section is devoted to some preliminaries and to notation. Although they are known, for the sake of completeness, we also obtain means and variances of the statistics under normality.

2. Preliminaries

Let $\Omega_1, \ldots, \Omega_n$ be n regions and assume that for each Ω_i we have an observation y_i on a (dependent) variable and values $x_{i1}, \ldots, x_{ik}$ of k independent variables. Then a linear regression model is written in the form

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{3}$$

where $\mathbf{y} = (y_1, \ldots, y_n)'$,

$$X = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1k} \\ 1 & x_{21} & \ldots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{pmatrix}$$

$\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$ is the vector of the error terms, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)'$ is the vector of parameters and a prime denotes matrix transpose. The least squares estimate $\mathbf{b}$ of $\boldsymbol{\beta}$ is (assuming X is non-singular)

$$(b_0, \ldots, b_k)' = \mathbf{b} = (X'X)^{-1}X'\mathbf{y} \tag{4}$$

and the residuals are, therefore,

$$(e_1, \ldots, e_n)' = \mathbf{e} = \mathbf{y} - X\mathbf{b} = \mathbf{y} - X(X'X)^{-1}X'\mathbf{y} = M\mathbf{y} \tag{5}$$

where $M = I - H$ and $H = X(X'X)^{-1}X'$. It follows that $\mathbf{e}$ also could have been written as $\mathbf{e} = MX\boldsymbol{\beta} + M\boldsymbol{\epsilon} = M\boldsymbol{\epsilon}$ and, moreover, it may be verified that M and H are idempotent (*i. e.*, $M^2 = M$ and $H^2 = H$).

In order to be assured that the linear combination $\boldsymbol{\ell}'\mathbf{b}$ of b_i's is a best (*i. e.*, minimum variance) unbiased linear estimate of $\boldsymbol{\ell}'\boldsymbol{\beta}$, three conditions — called Gauss–Markov conditions — must be met. These are

$$E(\epsilon_i) = 0 \tag{6}$$

$$E(\epsilon_i^2) = \sigma^2 (\text{a constant}) \tag{7}$$

$$E(\epsilon_i \epsilon_j) = 0 \tag{8}$$

for all i and j. In matrix notation, the Gauss–Markov conditions become

$$E(\boldsymbol{\epsilon}) = \mathbf{0}, \quad E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \sigma^2 I.$$

If, in addition, $\boldsymbol{\epsilon}$ is normally distributed we write $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I)$, *i. e.*, $\boldsymbol{\epsilon}$ is normally distributed with mean $\mathbf{0}$ and covariance matrix $\sigma^2 I$. From (3) it follows that if $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I)$, then $\mathbf{y} \sim N(X\boldsymbol{\beta}, \sigma^2 I)$.

When (8) holds we shall also say (since the observations are taken over space) that the ϵ_i's are spatially uncorrelated. On the other hand, if (8) does not hold and particularly if $E(\epsilon_i \epsilon_j) = \rho_{ij}$, where $|\rho_{ij}|$ declines with the spatial separation between Ω_i and Ω_j, we say the ϵ_i's are spatially correlated (see Cliff and Ord, 1981, for more on this subject). As is well known, the presence of spatial correlation does not bias the estimate $\mathbf{b}$, but the covariance matrix of $\mathbf{b}$ and any quantity that depends on it (the t, the F and, indeed, most statistics used for tests) are seriously affected.

The seriousness can be seen as follows. If $\boldsymbol{\epsilon}$ has covariance matrix $\sigma^2\Omega$, then the variance of $\boldsymbol{\ell}'\mathbf{b}$ is of the form $\sigma^2 \mathbf{c}'\Omega\mathbf{c}$ with $\mathbf{c}' = \boldsymbol{\ell}'(X'X)^{-1}X'$. This contains $n(n-1)$ terms involving non–diagonal elements of Ω. Therefore, even if each such element is small, their combined effect *can be* considerable. Even worse is the fact that, when we use ordinary least squares, computer packages typically compute estimates of variance under the assumption that Gauss–Markov Conditions hold, *i. e.*, $\Omega = I$. Therefore, unaccounted for non–diagonal elements can substantially affect any inferences we reach.

As mentioned in Section 1, the statistics (1) and (2) may be used to test for the existence of spatial correlation. To simplify matters we shall write both (1) and (2) in the form

$$c\mathbf{e}'Z\mathbf{e}/s^2 \tag{9}$$

which is obviously appropriate since numerators of both are quadratic forms.

Since $\mathbf{e} = M\boldsymbol{\epsilon}$, it follows that $\mathbf{e}'Z\mathbf{e} = \boldsymbol{\epsilon}'M'ZM\boldsymbol{\epsilon} = \boldsymbol{\epsilon}'B\boldsymbol{\epsilon}$ where $B = M'ZM$. Since the matrix B is symmetric there exists an orthogonal matrix Γ such that $B = \Gamma D_\lambda \Gamma'$ where $D_\lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ and λ_i's are the eigenvalues of B. Therefore, writing $\Gamma\boldsymbol{\epsilon} = \mathbf{u}$, we have

$$\mathbf{e}'Z\mathbf{e} = \boldsymbol{\epsilon}'B\boldsymbol{\epsilon} = \boldsymbol{\epsilon}'\Gamma D_\lambda \Gamma'\boldsymbol{\epsilon} = \mathbf{u}'D_\lambda\mathbf{u} = \sum_{i=1}^{n} \lambda_i u_i^2. \tag{10}$$

Now consider the denominator of (9):

$$s^2 = p^{-1}\mathbf{e}'\mathbf{e}$$

where $p = n - k - 1$. Since $\mathbf{e} = M\boldsymbol{\epsilon}$ and M is idempotent we may write s^2 as

$$p^{-1}\boldsymbol{\epsilon}'M'M\boldsymbol{\epsilon} = p^{-1}\boldsymbol{\epsilon}'M\boldsymbol{\epsilon}.$$

Obviously $BM = MB$ and hence the same matrix Γ that diagonalized B also diagonalizes M (see Bellman, 1960, p. 56). Hence

$$B = \Gamma D_\lambda \Gamma', \text{ and } M = \Gamma D_\xi \Gamma' \tag{11}$$

where $D_\xi = \text{diag}(\xi_1, \ldots, \xi_n)$. Since M is idempotent, its eigenvalues ξ_i are either one or zero and since the rank of M is $p = n - k - 1$, exactly p of the eigenvalues are ones. Therefore, we may write

$$c\mathbf{e}'Z\mathbf{e}/s^2 = c\boldsymbol{\epsilon}'B\boldsymbol{\epsilon}/p^{-1}\boldsymbol{\epsilon}'M\boldsymbol{\epsilon} = c\sum_{i=1}^{n}\lambda_i u_i^2 / p^{-1}\sum_{i=1}^{p} u_i^2 = cP, \tag{12}$$

where $P = U/V$, $U = \sum_{i=1}^{n} \lambda_i u_i^2$ and $V = p^{-1} \sum_{i=1}^{p} u_i^2$.

It is a property of normal distributions (see Rao, 1973; Srivastava and Khatri, 1979) that if $\boldsymbol{\epsilon} \sim N(0, \sigma^2 I)$, then $\mathbf{u} \sim N(0, \sigma^2 I)$. From (11), and the fact that the ξ_i's are either one or zero (and therefore $D_\xi^2 = D_\xi$), we get

$$\boldsymbol{\epsilon}' M \boldsymbol{\epsilon} = \boldsymbol{\epsilon}' \Gamma D_\xi \Gamma' \boldsymbol{\epsilon} = \mathbf{u}' D_\xi \mathbf{u} = \mathbf{u}' D_\xi' D_\xi \mathbf{u}.$$

Also, because D_ξ is diagonal with elements either 0 or 1, p of the components of $D_\xi \mathbf{u} = \mathbf{u}^\star$ are identical to those in $\mathbf{u}$ and the remainder are zeros. Now

$$\begin{aligned} \boldsymbol{\epsilon}' B \boldsymbol{\epsilon} = \boldsymbol{\epsilon}' M Z M \boldsymbol{\epsilon} &= \boldsymbol{\epsilon}' \Gamma D_\xi \Gamma' Z \Gamma D_\xi \Gamma' \boldsymbol{\epsilon} = \mathbf{u}' D_\xi \Gamma' Z \Gamma D_\xi \mathbf{u} \\ &= \mathbf{u}^\star (\Gamma' Z \Gamma) \mathbf{u}^\star = (\mathbf{u}^\star)' Z^\star \mathbf{u}^\star \end{aligned}$$

where $Z^\star = \Gamma' Z \Gamma$, showing that the same p u_i's are in the numerator of (12) as are in its denominator. Because of this and because $\mathbf{u}$ is normal, it follows that P and V are independent (Cliff and Ord, 1981, p.43, Theorem 1, Pitman, 1937; the original result is due to R. A. Fisher). An important consequence of this is that

$$E(P^s) E(V^s) = E(U^s) \tag{13}$$

Theorem 2.1.

When $\boldsymbol{\epsilon} \sim N(0, \sigma^2 I)$, *i. e.*, when $\mathbf{y} \sim N(X\beta, \sigma^2 I)$, the mean and variance of (9) are

$$E(\mathbf{e}' Z \mathbf{e} / s^2) = \mathrm{tr}[B] \tag{14}$$

and

$$\mathrm{var}(\mathbf{e}' Z \mathbf{e} / s^2) = 2(n-k+1)^{-1} \{(n-k-1)\mathrm{tr}[B^2] - (\mathrm{tr}[B])^2\} \tag{15}$$

where var($\cdot$) stands for the 'the variance of', and $B = M'ZM$.

As mentioned earlier, this result is known, e.g., see Ripley (1981, p. 100) or Brandsma and Ketellapper (1979).

Proof of Theorem 2.1:

From (10)

$$E(\mathbf{e}' Z \mathbf{e}) = E(\sum_{i=1}^{n} \lambda_i u_i^2) = \sigma^2 \sum_{i=1}^{n} \lambda_i = \sigma^2 \mathrm{tr}[D_\lambda] = \sigma^2 \mathrm{tr}[B] \tag{16}$$

since $\mathrm{tr}[B] = \mathrm{tr}[\Gamma' D_\lambda \Gamma] = \mathrm{tr}[D_\lambda \Gamma \Gamma'] = \mathrm{tr}[D_\lambda]$. Also, since u_i's are independent normal with mean 0 and variance σ^2,

$$\begin{aligned} E(\mathbf{e}' Z \mathbf{e})^2 &= E(\sum_{i=1}^{n} \lambda_i u_i^2)^2 \\ &= E[\sum_{i=1}^{n} \lambda_i^2 u_i^4 + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \lambda_i \lambda_j u_i^2 u_j^2] \\ &= \mu_4 \sum_{i=1}^{n} \lambda_i^2 + \sigma^4 \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \lambda_i \lambda_j \end{aligned} \tag{17}$$

where $\mu_4 = E(u_i^4) = 3\sigma^4$. Hence

$$\begin{aligned}\text{var}(\mathbf{e}'Z\mathbf{e}) &= E(\mathbf{e}'Z\mathbf{e})^2 - [E(\mathbf{e}'Z\mathbf{e})]^2 \\ &= \mu_4 \sum_{i=1}^{n} \lambda_i^2 + \sigma^4 \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \lambda_i \lambda_j - \sigma^4 (\sum_{i=1}^{n} \lambda_i)^2,\end{aligned}$$

and since

$$(\sum_{i=1}^{n} \lambda_i)^2 = \sum_{i=1}^{n} \lambda_i^2 + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \lambda_i \lambda_i$$

it follows that

$$\text{var}(\mathbf{e}'Z\mathbf{e}) = (\mu_4 - \sigma^4) \sum_{i=1}^{n} \lambda_i^2 = 2\sigma^4 \sum_{i=1}^{n} \lambda_i^2 = 2\sigma^4 \text{tr}[B^2]. \tag{18}$$

From (18) and (16), we have, alternatively

$$E(\mathbf{e}'Z\mathbf{e})^2 = 2\sigma^4 \text{tr}[B^2] + \sigma^4 \text{tr}[B])^2. \tag{19}$$

Now replacing B by M in (15) and (19) we get

$$E(s^2) = p^{-1} E(\boldsymbol{\epsilon}' M \boldsymbol{\epsilon}) = p^{-1} \sigma^2 \text{tr}[M] = \sigma^2 \tag{20}$$

and

$$E(s^2)^2 = p^{-2} E(\boldsymbol{\epsilon}' M \boldsymbol{\epsilon})^2 = p^{-2} \{2\sigma^4 \text{tr}[M] + \sigma^4 (\text{tr}[M])^2\} = \sigma^4 (2p^{-1} + 1). \tag{21}$$

It is now simple to verify that (14) follows from (13), (16) and (20). To verify (15), notice that from (13), (19) and (21)

$$E(P^2) = (2\text{tr}[B^2] + (\text{tr}[B])^2)/(1 + 2p^{-1}).$$

Therefore, from (14)

$$\begin{aligned}\text{var}(P) &= E(P^2) - (E(P))^2 \\ &= 2(1 + 2p^{-1})^{-1} \text{tr}[B^2] + (\text{tr}[B])^2 ((1 + 2p^{-1})^{-1} - 1) \\ &= 2(p + 2)^{-1} \{p\text{tr}[B^2] - (\text{tr}[B])^2\}.\end{aligned}$$

This proves the theorem.

3. Exact Distribution of P under Normality

Since P is a ratio of quadratic forms of normal variables, there are several methods available for computing its distribution function (cdf) under the hypothesis of no spatial correlation, although perhaps not as many as one would expect. In this section we describe versions of the two key ones. Sections 3.1 and 3.2 deal with the case where the λ_i's are distinct, while Section 3.3 is concerned with the λ_i's having common values. Section 3.1 is devoted to the proof of a theorem on the distribution of U, which then is used in Section 3.2 to provide formulæ for the cdf of P. Notice that without loss of generality we can set $\sigma = 1$ and assume that $\mathbf{u} \sim N(0, I)$.

3.1. Distribution of U when λ_i's are distinct

Theorem 3.1

Without loss of generality we can ignore zero valued λ_i's. Let $\mu_1 > \ldots > \mu_{n_1}$ be the negative valued λ_i's (if any) and $\nu_1 < \ldots < \nu_{n_2}$ be the positive valued λ's with $n = n_1 + n_2$. Define

$$\int_{(+)} = \sum_{k=1}^{\lfloor (n_2+1)/2 \rfloor} (-1)^{k+1} \int_{\nu_{2k-1}^{-1}}^{\nu_{2k}^{-1}}$$

and

$$\int_{(-)} = \sum_{k=1}^{\lfloor (n_1+1)/2 \rfloor} (-1)^{k+1} \int_{\mu_{2k-1}^{-1}}^{\mu_{2k}^{-1}}$$

where $\lfloor \psi \rfloor$ is the integer part of ψ, $\mu_{n_1+1} = -\infty$ and $\nu_{n_2+1} = \infty$. Further define

$$D(\lambda) = \left[- \prod_{1=1}^{n} (1 - \lambda\lambda_i) \right]^{-1/2}.$$

Then if the $\mathbf{u} \sim N(0, I)$, the cdf $F(z)$ of U is

$$F(z) = \begin{cases} 1 - \pi^{-1} \int_{(+)} \lambda^{-1} D(\lambda) \exp(-\lambda z/2)\, d\lambda & \text{for } z \geq 0 \\ \pi^{-1} \int_{(-)} \lambda^{-1} D(\lambda) \exp(-\lambda z/2)\, d\lambda & \text{for } z < 0. \end{cases} \tag{22}$$

Notice that $\int_{(+)}$ and $\int(-)$ depend on n_1, n_2 and the λ_i's. However, no confusion need occur if we note that whenever these symbols arise, we simply act as if they were exactly equivalent to their definitions above. This theorem is similar to one in Smirnov (1937) and also may be obtained from Plackett (1960, pp. 20-22). However, Smirnov made a mistake in signs which carried over into several subsequent papers (some of which did not refer to Smirnov!). The mistake is pointed out in the proof of the theorem which is given after the following lemma.

Lemma 3.1

The absolute value of each of the integrals that comprise either $\int_{(+)} \lambda^{-1} D(\lambda)\, d\lambda$ or $\int(-)\lambda^{-1} D(\lambda)\, d\lambda$ is less than a constant $M > 0$.

Proof:

Let a be a number between ν_{2k-1}^{-1} and ν_{2k}^{-1}. For $\nu_{2k-1}^{-1} \leq \lambda \leq a$,

$$|\lambda^{-1}| \prod_{\substack{i=1 \\ i \neq 2k-1}}^{n} |1 - \lambda\lambda_i|^{-1/2} < M_1$$

for some $M_1 > 0$. Hence

$$\int_{\nu_{2k-1}^{-1}}^{a} \lambda^{-1} D(\lambda)\, d\lambda < M_1 \int_{\nu_{2k-1}^{-1}}^{a} |1 - \lambda_{2k-1}\lambda|^{-1/2}\, d\lambda$$
$$= 2M_1 \lambda_{2k-1}^{-1/2} (a - \lambda_{2k-1}^{-1})^{1/2} < M_2(\text{say}).$$

Similarly it can be shown that if $2k \leq n_2$, $\int_a^{\nu_{2k}^{-1}} \lambda^{-1} D(\lambda) d\lambda$ is finite. When $2k = n_2 + 1$, *i. e.*, when $\nu_{2k} = \infty$, let $a - \nu_{n_2}^{-1} = c$ and

$$\int_a^{\infty} \lambda^{-1} D(\lambda) d\lambda < \left(\prod_{i=1}^{n} \lambda_i^{-1/2} \right) \int_c^{\infty} x^{-(n/2+1)} dx$$

which also is finite. Similarly, each of the components of $\int_{(-)} \lambda^{-1} D(\lambda) d\lambda$ may be shown to be less than some number M.

Proof of Theorem 3.1

Since each $u_i \sim N(0,1)$, it follows that u_i^2 has a chi–square distribution with 1 degree of freedom. Hence (Rao, 1973, p.167), its characteristic function is $(1 - 2\imath t)^{-1/2}$ where $\imath = \sqrt{-1}$. Therefore, the characteristic function of $\lambda_i u_i^2$ is $(1 - 2\imath t\lambda_i)^{-1/2}$ (Rao, 1973, p. 100). Since the u_i's are independent, the characteristic function of $U = \sum_{i=1}^{n} \lambda_i u_i^2$ is (Rao, 1973, p. 104)

$$\Phi_U(t) = \prod_{i=1}^{n} (1 - 2\imath t\lambda_i)^{-1/2}. \tag{23}$$

The distribution function $F(z)$ of U can be found from (23). From the general inversion theorem for characteristic functions (Wilks, 1962, p. 252), we have

$$\begin{aligned} F(z) - F(0) &= (2\pi)^{-1} \int_{-\infty}^{\infty} \Phi_U(t)[1 - \exp(\imath tz)](\imath t)^{-1} \, dt \\ &= (2\pi)^{-1} \int_{-\imath\infty}^{\imath\infty} \lambda^{-1} D(\lambda)[1 - \exp(-\lambda z/2)] \, d\lambda \end{aligned} \tag{24}$$

on making the substitution $\lambda = 2\imath t$. Hence for two different values z and z',

$$F(z) - F(z') = (2\pi)^{-1} \int_{-\imath\infty}^{\imath\infty} \lambda^{-1} D(\lambda)[\exp(-\lambda z'/2) - \exp(-\lambda z/2)] \, d\lambda. \tag{25}$$

Consider first the case where z and z' are non–negative. Call the integrand in (25) g and consider the non–negative half plane of λ, *i. e.*, the half plane containing complex values of λ with non–negative real parts. On this half plane $[D(\lambda)]^{-1}$ can take zero values only on the positive real half axis. Therefore, g is analytic over the entire half plane from which the slits S_k (see Exhibits 1.a and 1.b) have been removed and by Cauchy's Theorem (see, for example, Nehari, 1961 or Hille, 1959),

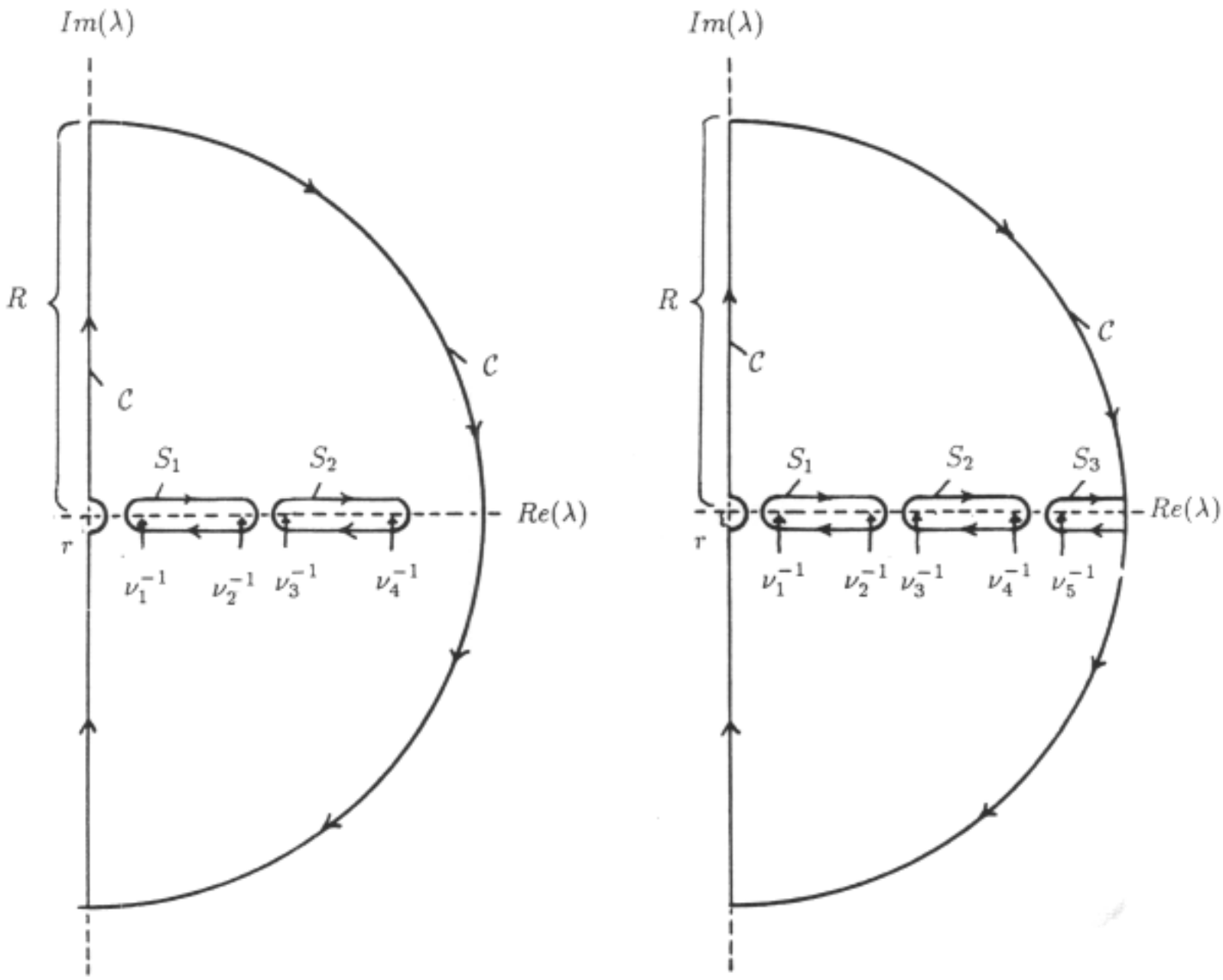
$$\int_C g + \sum_k \int_{\partial S_k} g = 0 \tag{26}$$

where the integral $\int_C$ is taken over the contour C illustrated in Exhibits 1.a and 1.b, and $\int_{\partial S_k}$ is the integral over the boundary ∂S_k of the slit S_k. When $R \to \infty$, the integral of g goes to zero over the part of C that is semicircular with radius R. Hence,

$$\lim_{R \to \infty} \int_C g = \int_{-\imath\infty}^{\imath\infty} g.$$

Exhibit 1. a, left; b, right.

Contour C and slits S_k.



Therefore, from (26),

$$\int_{-i\infty}^{i\infty} g = -\sum_k \int_{\partial S_k} g. \tag{27}$$

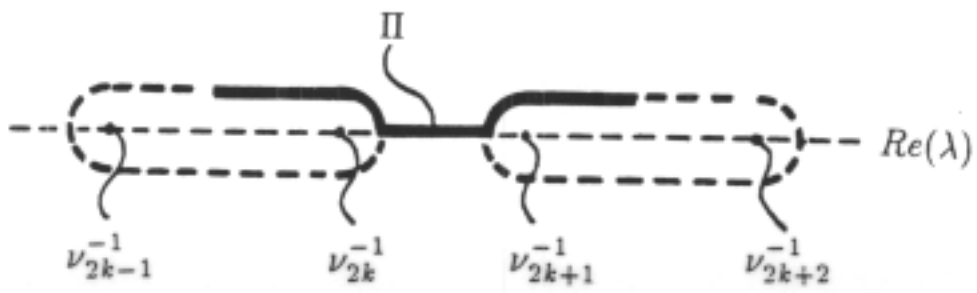
For any branch of $D(\lambda)$, as λ moves from the lower boundary of any of the slits S_k around the point ν_{2k-1}^{-1} to the upper boundary, $\arg(1-\nu_{2k-1}\lambda)$ changes by 2π while all other $\arg(1-\nu_i\lambda)$'s remain the same. Hence $\arg[D(\lambda)]$ changes by π, and consequently $D(\lambda)$ and g change sign. Moreover, if λ follows the path Π (see Exhibit 2) from the upper boundary of S_k around the points ν_{2k}^{-1} and ν_{2k+1} to the upper boundary of S_{k+1}, both $\arg(1-\nu_{2k}\lambda)$ and $\arg(1-\nu_{2k+1}\lambda)$ change by π, while other $\arg(1-\nu_i\lambda)$'s remain the same. Hence, again $d(\lambda)$ changes sign (this is the discussion Smirnov missed. I found the mistake when I used Smirnov's formula and obtained values of a distribution function which were greater than one). The value of g over the semicircular ends of the narrow slits S_k of width $2r$ is of the order $r^{-1/2}$. Therefore, the integral of g over one of these semicircles goes to 0

as $r \to 0$. Consequently, as $r \to 0$,

$$F(z) - F(z') = \pm\pi^{-1} \int_{(+)} \lambda^{-1} D(\lambda)[\exp(-\lambda z'/2) - \exp(-\lambda z/2)]\, d\lambda. \tag{28}$$

Exhibit 2.

Path π.



As $z' \to \infty$, $F(z') \to 1$ and $\int_{(+)} \lambda^{-1} D(\lambda)\exp(-\lambda z'/2) \to 0$. Therefore

$$F(z) = 1 \pm \pi^{-1} \int_{(+)} \lambda^{-1} D(\lambda)\exp(-\lambda z/2)\, d\lambda. \tag{29}$$

We choose that branch of $D(\lambda)$ which replaces '$\pm$' by '$-$' because, as we now show, $\int_{(+)} \lambda^{-1} D(\lambda)\exp(-\lambda z/2)\, d\lambda > 0$.

For $n_2 \leq 2$, $\int_{(+)} g$ is obviously positive. If $n_2 \geq 3$, then by Lemma 3.1

$$\int_{(+)} g > \exp[-z/(2\nu_2)] \int_{\nu_1^{-1}}^{\nu_2^{-1}} \lambda^{-1} D(\lambda)\, d\lambda - (n_2/2) M \exp[(-z/(2\nu_3)]$$

which is positive when z is large enough. Since $\int_{(+)} g$ is obviously continuous, if it were negative for some value of z, it would be zero for some other value, $z'' < \infty$. Then for either branch of $D(\lambda)$, $F(z'')$ would be 1, which is impossible.

This proves the part of (22) for $z \geq 0$. The proof for $z \leq 0$ is quite similar; only now we would consider a mirror image (in the imaginary axis) of the contour C and slits $S_1', S_2', \ldots$ enclosing pairs of points μ_i^{-1}. This proves the theorem.

If the reader is at all disturbed by the prospect of discontinuity of $F(z)$ at $z = 0$, he needs to notice the following: consider a region in the λ plane, enclosed by a large circular contour of radius R, from which slits S_k and S_k' have been removed as has been a small circular hole of radius r around $\lambda = 0$ (to account for the λ^{-1} in g). By the residue theorem (see Nehari, 1961, or Hille, 1959) the integral around the small hole is $2\pi i \lim_{\lambda \to 0} \lambda g = 1$. Notice also that as we move from the upper boundary of S_1' around the upper part of the circumference of the small circular hole to the upper boundary of S_k, $\arg(g)$ changes first

by $\pi/2$, then by π and finally by $\pi/2$ for a total of 2π. Therefore, $\int_{\partial S_1'}$ and $\int_{\partial S_1}$ have the same sign. Hence it may be shown that

$$\int_{(-)} \lambda^{-1} D(\lambda)\, d\lambda + \int (+)\lambda^{-1} D(\lambda)\, d\lambda - 1 = 0 \tag{30}$$

which establishes continuity.

3.2. Distribution of P when λ_i's are Distinct

In order to get the distribution of P from Theorem 3.1, we can proceed in at least two different ways. One is the approach taken by Imhoff (1961). Let $G(z)$ be the distribution of P. Then

$$G(z) = \Pr(P \le z) = \Pr(U \le zV) = \Pr(U - zV \le 0) = \Pr(\sum_{i=1}^{n} [\lambda_i - z] u_i^2 \le 0)$$

Thus, for each z, $G(z)$ is the same as $F(0)$ where $F(z)$ is as in Theorem 3.1 with the difference that λ_i's are replaced by $[\lambda_i - z]$. One computational advantage of this approach is that $\exp[-z\lambda]$, the repeated computation of which can be quite time consuming, becomes 1. Several alternative expressions and approximations also have been found for such $F(0)$'s — see Koertz and Abrahamse (1969, Ch. 5), Imhoff (1961), L'Esperance *et al.* (1976), White (1978).

Another approach is more classical. We present it as a theorem:

Theorem 3.2

Define

$$\psi(z, \lambda) = \begin{cases} 1 - \lambda z & \text{when } 0 \le \lambda z \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then, in the same notation as in Theorem 3.1, the distribution $G(z)$ of P is

$$G(z) = \begin{cases} 1 - \pi^{-1} \int_{(+)} \lambda^{-1} D(\lambda) [\psi(z, \lambda/p)]^{\nu - 1}\, d\lambda & \text{for } z \ge 0 \\ \pi^{-1} \int_{(-)} \lambda^{-1} D(\lambda) [\psi(z, \lambda/p)]^{\nu - 1}\, d\lambda & \text{for } z < 0, \end{cases} \tag{31}$$

where $2\nu = p$ and ν is not necessarily an integer.

Before using (31) in numerical computations, it is desirable to plot the function $\psi(z, \lambda/p)$ for some representative values of the arguments. This would help in determining time saving limits of integration. In this context, it might also be mentioned that since numerical integration is equivalent to computing the dot–product of (rather long) vectors, modern compilers can perform the operation rather efficiently.

Notice that as $p \to \infty$, (31) becomes (22), as indeed should have been expected. Koopmans (1942, see also von Neumann, 1942, and Plackett, 1960) obtained the density function corresponding to (31) in terms of a complex integral from which (31) can be derived in a manner analogous to the proof of Theorem 3.1. Mulholland (1970) also has obtained a result similar to Theorem 3.2.

Proof of Theorem 3.2:

The density function (p.d.f.) of U can easily be obtained from (22) by differentiation with respect to z. It is

$$f(z) = \begin{cases} +(2\pi)^{-1} \int_{(+)} D(\lambda)\exp\,(-\lambda z)\, d\lambda & \text{for } z \geq 0 \\ -(2\pi)^{-1} \int_{(-)} D(\lambda)\exp(-\lambda z)\, d\lambda. & \text{for } z < 0. \end{cases}$$

Therefore, the sth moment of U is

$$E(U^s) = (2\pi)^{-1} \left[\int_0^\infty \int_{(+)} D(\lambda) z^s \exp(-\lambda z)\, d\lambda dz \right.$$
$$\left. - \int_{-\infty}^0 \int_{(-)} D(\lambda) z^s \exp(-\lambda z)\, d\lambda dz. \right. \tag{32}$$

The Gamma function $\Gamma(s+1)$ is defined as $\int_0^\infty x^s \exp(-x)\, dx$. Therefore,

$$\int_0^\infty z^s \exp(-\lambda z/2) dx = (2/\lambda)^{s+1} \Gamma(s+1).$$

The integrability of the relevant functions in (32) is easily established. Hence we can use Fubini's theorem to exchange the order of integration. Therefore, the first integral within square brackets in (32) can be written as

$$\Gamma(s+1) \int_{(+)} D(\lambda)(2/\lambda)^{s+1}\, d\lambda.$$

After carrying out a similar exercise with the second integral we get

$$E(U^s) = (2\pi)^{-1}\Gamma(s+1) \cdot \left[\int_{(+)} D(\lambda)(2/\lambda)^{s+1}\, d\lambda + (-1)^s \int_{(--)} D(\lambda)(2/\lambda)^{s+1}\, d\lambda \right].$$

where $\int_{(--)}$ is obtained from $\int_{(-)}$ by replacing each ν_i^{-1} by $-\nu_i^{-1}$.

Since pV has a chi–square distribution with p degrees of freedom, it can be shown that

$$E[(pV)^s] = 2^s \Gamma(s+\nu)/\Gamma(\nu).$$

Hence, by (13), it follows that

$$E[(pP)^s] = \pi^{-1}[\Gamma(s+1)\Gamma(\nu)/\Gamma(\nu+s)]$$
$$\cdot \left[\int_{(+)} D(\lambda)\lambda^{-s-1}\, d\lambda + (-1)^s \int_{(-)} D(\lambda)\lambda^{-s-1}\, d\lambda \right]. \tag{33}$$

Now consider the function

$$h(z,\lambda) = \pi^{-1}(\nu-1)D(\lambda)[\psi(z,\lambda)]^{\nu-2}.$$

Since it can easily be verified that $z^r h(z,\lambda)$ is integrable over the appropriate region, we can use Fubini's theorem to invert order of integration and get

$$\begin{aligned}\int_0^\infty z^s \int_{(+)} h(z,\lambda)\,d\lambda dz &= \pi^{-1}(\nu-1)\int_{(+)} D(\lambda)\int_0^{\lambda^{-1}} z^s(1-\lambda z)^{\nu-2}\,dz d\lambda \\ &= \pi^{-1}(\nu-1)\int_{(+)} D(\lambda)\lambda^{-s-1}\int_0^1 (\lambda z)^s(1-\lambda z)^{\nu-1}\,d(\lambda z)d\lambda \\ &= \pi^{-1}[\gamma(s+1)\Gamma(\nu)/\Gamma(\nu+s)]\int_{(+)} D(\lambda)\lambda^{-s-1}\,d\lambda, \qquad (34)\end{aligned}$$

on noting that $\int_0^1 x^{a-1}(1-x)^{b-1}\,dx$ is the beta–function $\Gamma(a)\Gamma(b)/\Gamma(a+b)$ (see Wilks, 1962, p. 174). Therefore, (34) is the first term on the right side of (33). A similar result can be derived for the second term. Hence,

$$E[(pP)^s] = \int_0^\infty z^s \int_{(+)} h(z,\lambda)\,d\lambda dz - \int_{-\infty}^0 z^s \int_{(-)} h(z,\lambda)\,d\lambda dz. \qquad (35)$$

Therefore, by the uniqueness theorem for moments of bounded random variables, it follows that the p.d.f. of pP is $\int_{(+)} h(z,\lambda)\,d\lambda$ when $z \geq 0$, and when $z < 0$ it is $\int_{(-)} h(z,\lambda)\,d\lambda$. A straightforward change of variable yields the p.d.f. of P as

$$\begin{cases} +(2\pi\nu)^{-1}(\nu-1)\int_{(+)} D(\lambda)[\psi(z,\lambda/p)]^{\nu-2}\,d\lambda \text{ for } z \geq 0 \\ -(2\pi\nu)^{-1}(\nu-1)\int_{(-)} D(\lambda)[\psi(z,\lambda/p)]^{\nu-2}\,d\lambda \text{ for } z < 0. \end{cases} \qquad (36)$$

To complete the theorem and obtain the distribution function all we need do is integrate the density function (36) from 0 to z. When $z < 0$, it is easy to see that this integral [which is simply the integral of the lower expression in (36)] is the lower expression in (31). When $z \geq 0$, integration of the upper expression in (36) from 0 to z, followed by the use of (30), yields the upper part of (31) as the required distribution function.

3.3. Distribution of P when Values of λ_i's are Repeated

In most practical applications the λ_i's are distinct, particularly when the regions considered are irregular — as census tracts, states and Zip–code zones usually are. If a pair (or two) of the λ_i's become alike, there is perhaps not much lost by adding a small number to one and subtracting it from the other in order to make them distinct. However, this recourse is not too satisfying if a large number of pairs of λ_i's are the same. This can happen when observations are taken over a regular lattice or over regions bounded by a uniform grid (*e.g.*, quarter sections). In this section we first consider the case when

$$U = \sum_{\ell=1}^{2}\sum_{i=1}^{m} \lambda_i u_{i\ell}^2, \qquad (37)$$

λ_i's are distinct and $n = 2m$ is necessarily even. Then we shall consider the case when some λ_i's are singletons while others come in pairs. Finally, we shall briefly examine the general case when λ_i's may be repeated arbitrarily often.

As before, let $\mu_1, \ldots, \mu_{m_1}$ be the negative λ_i's and let $\nu_1, \ldots, \nu_{m_2}$ be the positive λ_i's. Let

$$B_{k_{(+)}} = \prod_{i=1}^{m_1} (1 - \mu_i/\nu_k)^{-1} \prod_{\substack{i=1 \\ i \neq k}}^{n_2} (1 - \nu_i/\nu_k)^{-1}$$

$$B_{k_{(-)}} = \prod_{\substack{i=1 \\ i \neq k}}^{m_1} (1 - \mu_i/\mu_k)^{-1} \prod_{\substack{i=1 \\ i \neq k}}^{n_2} (1 - \nu_i/\mu_k)^{-1}.$$

Then, for $z \geq 0$, considering the same semicircular contour C that we did in the proof of Theorem 3.1, and for $z < 0$, considering the mirror image of C in the imaginary axis, and applying the residue theorem, it may be shown that the distribution function of U is

$$F(z) = \begin{cases} 1 - \sum_k^{n_2} B_{k_{(+)}} \exp(-z\nu_k/2) & \text{for } z \geq 0 \\ \sum_k^{n_1} B_{k(-)} \exp(-z\mu_k/2) & \text{for } z < 0. \end{cases} \tag{38}$$

The result is very well known, having been given by several authors (see Box, 1954).

Using (38) and following steps similar to those in the proof of Theorem 3.2, it may be shown that the distribution of P, when U has the form (37), is

$$F(z) = \begin{cases} 1 - \sum_k^{n_2} B_{k(+)} (\psi[z, (p\nu_k)^{-1}])^{\nu - 1} & \text{for } z \geq 0 \\ \sum_k^{n_1} B_{k(-)} (\psi[z, (p\mu_k)^{-1}])^{\nu - 1} & \text{for } z < 0 \end{cases} \tag{39}$$

where ψ is the same function we defined in Theorem 3.2. Obviously, since no integration is involved, (38) is easier to use than Theorem 3.1.

When some λ_i's are singletons while others come in pairs, we can write $U = U_1 + U_2$, where

$$U_1 = \sum_{i=1}^{q} \lambda_{i(1)} u_i^2 \tag{40}$$

with distinct $\lambda_{i(1)}$'s and U_2 is as in (37). The p.d.f. f_1 of U_1 can be found from Theorem 3.1, and f_2, the p.d.f. of U_2, can be found from (38). The p.d.f. $f(z)$ of U can then be found by convolution:

$$f(z) = \int_{-\infty}^{\infty} f_1(z - y) f_2(y)\, dy.$$

Since z enters both $f_1(z)$ and $f_2(z)$ only as an argument of the exponential function, such a convolution is easy to obtain analytically and is roughly of the same form as $F(z)$ in Theorem 3.1. The distribution of the corresponding P can be found using the same means as in the proof of Theorem 3.2, and is numerically no more difficult to compute than (31).

It is unlikely that, in practice, we would encounter values of λ_i repeated more than three times, and, therefore, the discussion above should cover most practical situations. However, for completeness, we briefly outline a treatment for the most general case where $U = U_1 + U_2$, with U_1 as in (40) and U_2 containing coefficients the values of which are repeated an even number of times. The distribution of U_2 can be obtained using the residue theorem. It will have the same general form as (38), but B_k's will now be polynomials and not constants.

The distribution of U again can be found analytically by convolution, although now it will contain Γ–functions (because B_k are polynomials). That of P can be obtained using the same method as in the proof of Theorem 3.2, but because of the Γ–functions in the convolution, it would contain β–functions.

4. Asymptotic Mean and Variance

If we do not assume that $\boldsymbol{\epsilon}$ is normal, then in general we cannot say much about $\mathbf{u}$. Consequently we must proceed differently and indeed the exact mean and variance of (9) are not available. However, we can obtain the mean and variance of the numerator of (9) and obtain asymptotic results by letting the denominator $s^2 \to \sigma^2$. These results will be sharper than the ones in Sen (1976).

To obtain the mean of the numerator $c\mathbf{e}'Z\mathbf{e}$ we can proceed directly. As before we can write $\mathbf{e}'Z\mathbf{e} = \boldsymbol{\epsilon}'MZM\boldsymbol{\epsilon} = \boldsymbol{\epsilon}'B\boldsymbol{\epsilon}$, and, since we are interested in the distribution under the hypothesis, $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{\epsilon}) = \sigma^2 I$. Hence

$$\begin{aligned} E(\boldsymbol{\epsilon}'B\boldsymbol{\epsilon}) &= E(\text{tr}[\boldsymbol{\epsilon}'B\boldsymbol{\epsilon}]) = E(\text{tr}[B\boldsymbol{\epsilon}\boldsymbol{\epsilon}']) = \text{tr}[E(B\boldsymbol{\epsilon}\boldsymbol{\epsilon}')] \\ &= \text{tr}[BE(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')] = \text{tr}[B(\sigma^2 I)] = \sigma^2 \text{tr}[B] \end{aligned} \tag{41}$$

as before. Hence $E(c\mathbf{e}'Z\mathbf{e}) = c\sigma^2\text{tr}(B)$ and $E(P) = \text{tr}(B)$.

In order to obtain the variance, we first compute

$$E(\boldsymbol{\epsilon}'B\boldsymbol{\epsilon})^2 = E(\boldsymbol{\epsilon}'B\boldsymbol{\epsilon}\boldsymbol{\epsilon}'B\boldsymbol{\epsilon}) = E(\text{tr}[\boldsymbol{\epsilon}B\boldsymbol{\epsilon}\boldsymbol{\epsilon}'B\boldsymbol{\epsilon}]) = E(\text{tr}[B\boldsymbol{\epsilon}\boldsymbol{\epsilon}'B\boldsymbol{\epsilon}\boldsymbol{\epsilon}']) = \text{tr}[BC] \tag{42}$$

where

$$C = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}'B\boldsymbol{\epsilon}\boldsymbol{\epsilon}'). \tag{43}$$

To compute (43) we note that

$$E(\boldsymbol{\epsilon}_i^4) = \mu_4 \quad \text{and} \quad E(\boldsymbol{\epsilon}_i^2\boldsymbol{\epsilon}_j^2) = \sigma^4 \text{ when } i \neq j, \tag{44}$$

and all other expectations of the products of four $\boldsymbol{\epsilon}_i$'s are zero, *i. e.*,

$$\begin{aligned} &E(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}, 3_j) = E(\boldsymbol{\epsilon}_i)E(\boldsymbol{\epsilon}_j^3) = 0 \text{ for } i \neq j, \\ &E(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_j^2\boldsymbol{\epsilon}_k) = 0 \text{ for } i \neq j, i \neq k, \\ &E(\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_j\boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_{\boldsymbol{\ell}}) = 0 \text{ for } i \neq j, k, \boldsymbol{\ell}. \end{aligned} \tag{45}$$

Now set $\boldsymbol{\epsilon}\boldsymbol{\epsilon}'B = A = (a_{i\boldsymbol{\ell}})$ (i.e, A is the matrix with elements $a_{i\boldsymbol{\ell}}$). Obviously $a_{i\boldsymbol{\ell}} = \boldsymbol{\epsilon}_i \sum_{k=1}^n \boldsymbol{\epsilon}_k b_{k\boldsymbol{\ell}}$. Since the $(\boldsymbol{\ell}, j)$th element of $\boldsymbol{\epsilon}\boldsymbol{\epsilon}'$ is $\boldsymbol{\epsilon}_{\boldsymbol{\ell}}\boldsymbol{\epsilon}_j$, if we set $C = (c_{ij})$,

$$c_{ij} = \sum_{\boldsymbol{\ell}} c_{\boldsymbol{\ell}}(i,j) \text{ where } c_{\boldsymbol{\ell}}(i,j) = E[\boldsymbol{\epsilon}_i(\sum_{k=1}^n \boldsymbol{\epsilon}_k b_{k\boldsymbol{\ell}})\boldsymbol{\epsilon}_{\boldsymbol{\ell}}\boldsymbol{\epsilon}_j].$$

From (45) we see that the terms of $c_{\boldsymbol{\ell}}(i,j)$ are zeros unless

$$i = j \text{ and } k = \boldsymbol{\ell} \tag{46}$$

$$i = \boldsymbol{\ell} \text{ and } k = j \tag{47}$$

$$i = k \text{ and } \boldsymbol{\ell} = j. \tag{48}$$

First consider $i = j$. Then the only non-zero term in $c_\ell(i,i)$ occurs when $k = \ell$. Therefore, if $i \neq \ell$, $c_\ell(i,i) = b_{\ell\ell}\sigma^4$, and if $i = \ell$, $c_i(i,i) = b_{ii}\mu_4$. It follows that

$$c_{ii} = \sum_{\ell=1}^{n} b_{\ell\ell}\sigma^4 + b_{ii}(\mu_4 - \sigma^4) = \sigma^4 \mathrm{tr}[B] + b_{ii}(\mu_4 - \sigma^4).$$

If $i \neq j$, then $c_\ell(i,j) \neq 0$ only if either $i = \ell$ or $j = \ell$. In the former case $c_i(i,j) = \sigma^4 b_{ji}$ and in the latter case $c_j(i,j) = \sigma^4 b_{ij}$. Since B is symmetric, this yields

$$c_{ij} = 2\sigma^4 b_{ij}.$$

Thus

$$\begin{aligned} C = \sigma^4 &\begin{pmatrix} \mathrm{tr}[B] & 2b_{12} & \dots & 2b_{1n} \\ 2b_{21} & \mathrm{tr}[B] & \dots & 2b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 2b_{n1} & 2b_{n2} & \dots & \mathrm{tr}[B] \end{pmatrix} \\ &+ (\mu_4 - \sigma^4)\,\mathrm{diag}(b_{11}, b_{22}, \dots, b_{\mathrm{nn}}) \\ &= 2\sigma^4 B + (\mu_4 - 3\sigma^4)\,\mathrm{diag}(b_{11}, b_{22}, \dots, b_{\mathrm{nn}}) + \sigma^4(\mathrm{tr}[B])I_n. \end{aligned}$$

From (41) and (42)

$$\mathrm{var}(\mathbf{e}'Z\mathbf{e}) = \mathrm{tr}[BC] - \sigma^4(\mathrm{tr}[B])^2 = 2\sigma^4\mathrm{tr}[B^2] + (\mu_4 - 3\sigma^4)\sum_{i=1}^{n} b_{ii}^2. \tag{49}$$

It has been shown (Sen, 1976) that $\mathbf{e}'Z\mathbf{e} = \boldsymbol{\epsilon}'B\boldsymbol{\epsilon}$ is asymptotically normal when $b_{ii} \to 0$ and some other mild conditions mentioned in that paper are met. In most practical situations where a small set of observations are not unduly influential these conditions will always be met. Then, of course, $(\mu_4 - 3\sigma^4) \to 0$. Since $s^2 \to \sigma^2$ almost surely, the asymptotic normality of $e'Ze/s^2$ follows from Slutsky's Theorem (see Rao, 1971, p. 122). Hence the asymptotic variance of P is $2\mathrm{tr}[B^2]$.

5. References

Bellman, R. A. (1962) *Introduction to Matrix Analysis*, McGraw–Hill, New York.

Box, G. E. P. (1954) Some Theorems on Quadratic Forms Applied in the Study of Variance Problems, I. Effect of Inequality of Variance in the one–way Classification, *Annals of Mathematical Statistics*, **26** 464-477.

Brandsma, A. S. and R. H. Ketellapper (1979) Further Evidence on Alternative Procedures for Testing Spatial Correlation Amongst Regression Disturbances, in *Exploratory and Explanatory Statistical Analysis of Spatial Data*, C.P.A. Bartels and R.H. Ketellaper, Eds. Martinus Nijhoff, Boston, 113-136.

Cliff, A. D. and J. K. Ord. (1981) *Spatial Processes*, Pion, London.

Hille, E. (1959) *Analytic Function Theory, Vol. 1*, Ginn and Co., Boston.

Imhoff, J. P. (1961) Computing the Distribution of Quadratic Forms in Normal Variables, *Biometrika* **48** 419-426.

Koertz, J. and A. P. J. Abrahamse (1969) *On the Theory and Applications of the General Linear Model*, Rotterdam University Press.

Koopmans, T. C. (1942) Serial Correlation and Normal Quadratic Forms in Normal Variables, *Annals of Mathematical Statistics*, **13** 14-33.

L'Esperance, W. L., D. Chall and D. Taylor (1976) An Algorithm for Determining the Distribution Function of the Durbin-Watson Statistic, *Econometrica*, **44** 1325-1346.

Mullholland, H. P. (1970) On Singularities of Sampling Distributions, and Particular for Ratios of Quadratic Forms, *Biometrika*, **57** 155-174.

Nehari, Z. (1961) *Complex Analysis*, Allyn and Bacon, Boston.

Pitman, E. J. G. (1937) The 'Closest' Estimates of Statistical Parameters, *Proceedings, Cambridge Philosophical Society*, **33** 212.

Plackett, R. L. (1960) *Principles of Regression Analysis*, Clarendon Press, Oxford.

Rao, C. R. (1973) *Linear Statistical Inference and its Applications*, Wiley.

Ripley, B. D. (1981) *Spatial Statistics*, Wiley.

Sen, A. (1976) Large Sample–Size Distribution of Statistics Used in Testing for Spatial Correlation, *Geographical Analysis*, **9** 175-184.

Sen, A. and S. Soot. (1977) Rank Tests for Spatial Correlation, *Environment and Planning A*, **9** 897-903.

Smirnov, N. V. (1937) O Raspredelenie ω^2 kriteriia Mizesa, *Matematicheskii Sbornik, 2* **44** 973-993. A French summary is given in Smirnoff, N. V. (1936) Sur la Distribution de ω^2 (Criterium de M. R. v. Mises), *Com. Rend. Acad. Sci. (Paris)* **202** 449-452.

Srivastava, M. S. and C. G. Khatri (1979) *An Introduction to Multivariate Statistics*, North–Holland.

Von Neumann, J. (1941) Distribution of the Ratio of the Mean Square Successive Difference to the Variance, *Annals of Mathematical Statistics*, **12** 367-395.

White, K. J. (1978) A General Computer Program for Econometric Models — SHAZAM, *Econometrica*, **46** 151–159.

Wilks, S. S. (1962) *Mathematical Statistics*, Wiley.

DISCUSSION

"Distribution of spatial correlation statistics"

by Ashish Sen

The paper looks into the difficult problem of obtaining the exact distribution of spatial correlation statistics P for normal residuals under the null hypothesis of spatial independence. Also, the first two moments of P are obtained together with the first two asymptotic moments for non–normal variables. Theoretically the results are impressive, but I believe that substantial further work is required to make them applicable in practice. My main comments are as follows:

(i) How will the actual formulæ for the distribution function help with numerical examples? Cannot the same work be achieved in practice by simulating the percentage points under the null hypothesis? [I am not clear of the precise overlap of Theorem 3.2 with Mulholland (1970)].

(ii) The mean and variance of the statistic P under the normality assumption are given in Section 2. How will these help in approximating the distribution function? Should one take higher moments and use a Beta distribution, for example? The work of Jones (1987) and the references therein are relevant to obtain the moments.

(iii) The asymptotic variance in Section 4 also must depend on $\mathrm{cov}(\mathbf{e}'\mathbf{Ze}, s^2)$. This value can be obtained in the same way as $\mathrm{var}(\mathbf{e}'\mathbf{Ze})$ in Section 4. Using the asymptotic variance for the ratio through Taylor series expansion one can obtain an improved approximation to the variance of P for random samples from non–normal variables. The use of the permutation approach of Box and Watson (1962) and Mardia (1970) will be a step forward. Note that for $b_{ii} \to 0$ asymptotically, the variance of P under the normal and non–normal cases is equivalent. Thus, at least for the first two moments, the effect of non–normality seems to be minimal! Does this result require the assumption that the kurtosis is negligible? A few simulation studies should prove useful.

Now coming to a point of detail, I might mention that Theorem 2.1 follows trivially from a well–known result (see Mardia, Kent and Bibby, 1979, p. 95, Exercise 3.4.21) that if $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$, then

$$\mathrm{var}(\mathbf{X}'\mathbf{AX}) = 2\mathrm{tr}(\mathbf{A\Sigma})^2,$$
$$\mathrm{cov}(\mathbf{X}'\mathbf{AX}, \mathbf{X}'\mathbf{BX}) = 2\mathrm{tr}(\mathbf{A\Sigma B\Sigma}).$$

Here $\mathbf{e} \sim N(0, \mathbf{M})$ where $\mathbf{M}$ is idempotent so that

$$\mathrm{cov}(\mathbf{e}'\mathbf{Ze}, \mathbf{e}'\mathbf{e}) = 2\sigma^4\mathrm{tr}(\mathbf{ZM}) = 2\sigma^4\mathrm{tr}(\mathbf{M}'\mathbf{ZM}) = 2\sigma^4\mathrm{tr}(\mathbf{B})$$
$$\mathrm{var}(\mathbf{e}'\mathbf{Ze}) = 2\mathrm{tr}(\mathbf{ZM})^2 = 2\sigma^4\mathrm{tr}(\mathbf{B})^2,$$
$$\mathrm{var}(\mathbf{e}'\mathbf{e}) = 2\sigma^4 p.$$

Further, $P = \mathbf{e}'\mathbf{Ze}/\mathbf{e}'\mathbf{e}$ and $V = \mathbf{e}'\mathbf{Ze}$, are independent under normality of $\mathbf{e}$ because P is a scale free quantity. Hence $\mathrm{var}(P)$ can be written down.

In conclusion, I am sure that Dr. Sen's paper will generate further research work with the aim of some specific recommendations for the practitioners.

References

Box, G. E. S., and G. S. Watson. (1962) Robustness to nonnormality of regression tests. *Biometrika*, **49**, 93-106 (correction, *Biometrika*, **52**, 669).

Jones, M. C. (1987) On moments of ratios of quadratic forms in normal variables. *Statistics & Probability Letters*, **6**, 129-136.

Mardia, K. V. (1979) The effect of nonnormality on some multivariate tests and robustness to nonnormality in the linear model. *Biometrika*, **58**, 105-121.

Mardia, K. V., J. T. Kent, and J. M. Bibby. (1979) *Multivariate Analysis*. New York: Academic Press.

Kanti V. Mardia, University of Leeds