Transformations which preserve exchangeability and application to permutation tests.

Daniel Commenges

Transformations which preserve exchangeability and application to permutation tests

Daniel Commenges
Biostatistics Team INSERM E03 38
ISPED, 146 rue Léo Saignat, Bordeaux, 33076, France
Tel: (33) 5 57 57 11 82; Fax (33) 5 56 24 00 81

January 21, 2003

Nonparametric Statistics 2001, in press

Abstract:

Exchangeability of observations is a key condition for applying permutation tests. We characterize the linear transformations which preserve exchangeability, distinguishing second-moment exchangeability and global exchangeability; we also examine non-linear transformations. When exchangeability does not hold one may try to find a transformation which achieves approximate exchangeability; then an approximate permutation test can be done. More specifically, consider a statistic $T = \phi(Y)$; it may be possible to find V such that $\tilde{Y} = V(Y)$ is exchangeable and to write $T = \tilde{\phi}(\tilde{Y})$. In other cases we may be content that $\tilde{Y}$ has an exchangeable variance matrix, which we denote second-moment exchangeability. When seeking transformations towards exchangeability we show the privileged role of residuals. We show that exact permutation tests can be constructed for the normal linear model. Finally we suggest approximate permutation tests based on second-moment exchangeability . In the case of an intraclass correlation model, the transformation is simple to implement. We also give permutational moments of linear and quadratic forms and show how this can be used through Cornish-Fisher expansions.

Keywords: Exchangeability, Permutation test, Homogeneity test, Residuals, Score Test, Transformation.

1 Introduction

Exchangeability is a key requirement for building a permutation test. A vector Y has an exchangeable distribution if PY has the same distribution as Y, for any permutation matrix P. If we consider a test statistic $T = \phi(Y)$ a permutation test is obtained, if Y is exchangeable, by conditioning on the order statistics $Y_{(o)} = \{Y_{(1)}, \ldots, Y_{(n)}\}$ (Cox and Hinkley 1974; Kalbfleisch 1978); that is, the conditional distribution of T is $Pr[T = \phi(PY_{(o)})] = 1/n!$, for any permutation matrix P. The assumption of exchangeability, although a little less stringent than the assumption of identically independently distributed observations, is still quite restrictive, and does not hold for instance in regression problems. Consider for instance the linear regression problem (studied in section 3.3): $Y = Z\beta + \varepsilon$ where Z is a matrix of explanatory variables, β a vector of coefficients and ε an exchangeable vector of errors with zero expectation. The exchangeability property will not in general be retained by Y in such a model (because the Y_i do not have the same expectation) so permutation tests cannot be applied in a simple way.

The possibility of applying permutation tests to situations where exchangeability is not given *a priori* has received only limited attention (Romano 1990; Schmoyer 1994) and it is clear that there are situations (for instance when data have a non-exchangeable correlation matrix) where permutation tests will not be robust.

This paper is an attempt to extend some theoretical results about tranformations which preserve exchangeability and to explore a way of extending permutations tests to complex situations. This way is to find a transformation V such that $\tilde{Y} = V(Y)$ is exactly or approximately exchangeable and to rely on the permutation distribution of T (expressed as a function of $\tilde{Y}$) induced by the permutations of $\tilde{Y}$. Since exchangeability is a very stringent assumption, we have been led to examine a weaker assumption, that of second moment exchangeability (PY has same first and second moments as Y). We have considered the tests of regression coefficients and the tests of homogeneity as described by Commenges and Jacqmin-Gadda (1997). We have not obtained by this way a particularly appealing new test; however, along the way, we have obtained several theoretical results which may be useful for future reflexion on this topic.

The organization of the paper is as follows. As a prerequisite, in section 2, we first consider linear transformations which preserve exchangeability, a topic already studied by Dean and Verducci (1990) and Dean and Wolfe (1990); here we distinguish between second-moment and global exchangeability; non-linear transformations are also considered. Then in section 3 we consider linear transformations which achieve exchangeability: the use of residuals is highlighted and we give exact permutation tests for the normal linear model and an example in a matched pairs design. Section 4 presents a more empirical approach to the general case, in particular in the case of an intraclass correlation model; we examine both linear and quadratic statistics. Finally, a last section (which is not tightly linked to the rest of the paper) presents some technical indications, difficult to find elsewhere, for performing a permutation test using approximations based on moments.

2 Characterisation of transformations which preserve exchangeability

2.1 Exchangeable matrices

A random $n \times 1$ vector Y has an exchangeable distribution if and only if any permutation of Y has the same distribution. This implies that the variance matrix of PY, where P is a permutation matrix, must be the same as that of Y. If E is this matrix we must have

$$PEP^T = E$$

for any matrix $P \in \mathcal{P}_n$, the set of $n \times n$ permutation matrices. Such a matrix is called an exchangeable matrix and can be written

$$E_{(a,b)} = a_1 I + b_1 \bar{\mathbf{1}} = a(I - \bar{\mathbf{1}}) + b\bar{\mathbf{1}}$$

where $\bar{\mathbf{1}}$ is a $n \times n$ matrix which has all its elements equal to $1/n$, and I is the $n \times n$ identity matrix. The second representation is interesting because

$\bar{\mathbf{1}}$ and $I - \bar{\mathbf{1}}$ are idempotent and orthogonal. They form an orthonormal basis in the space $\mathcal{E}_n$ of $n \times n$ exchangeable matrices and we have

$$E_{(a,b)} + E_{(a',b')} = E_{(a+a',b+b')}$$

$$E_{(a,b)}E_{(a',b')} = E_{(aa',bb')}$$

$$E_{(a,b)}^{-1} = E_{(a^{-1},b^{-1})}$$

$$E_{(a,b)}^{1/2} = E_{(a^{1/2},b^{1/2})}, \text{ with } E_{(a,b)} = E_{(a,b)}^{1/2}E_{(a,b)}^{1/2}.$$

Finally, the rank of $E_{(a,b)}$ is 0 if $a = b = 0$, 1 if $a = 0$, $b \neq 0$, $n - 1$ if $a \neq 0$, $b = 0$ and n if $a \neq 0$, $b \neq 0$.

Notations: in some cases the dimension of a matrix will be indicated as a subscript, in other cases where the dimension is obvious it will be omitted: for instance the $n \times n$ matrix which has all its elements equal to $1/n$ will be denoted $\bar{\mathbf{1}}$ or $\bar{\mathbf{1}}_n$.

2.2 Linear transformations which preserve second-moment exchangeability

Let $\mathcal{K}_{\mathcal{E}2}$ the space of linear transformations from $\mathcal{R}^n$ to $\mathcal{R}^m$, with $m \leq n$, which preserve second-moment exchangeability. The following theorem characterizes these transformations.

Theorem 1 *A linear transformation from $\mathcal{R}^n$ to $\mathcal{R}^m$ preserves second moment exchangeability (that is belongs to $\mathcal{K}_{\mathcal{E}2}$ if it is represented by a matrix which can be written as*

$$K_{m,n} = E_m Q_{m,n},$$

where, $E_m \in \mathcal{E}_m$ and $Q_{m,n}$ is such that $Q_{m,n}Q_{m,n}^T = I_m$; the row vectors q_i of Q form thus an orthonormal basis; in addition they must satisfy the condition $\sum_j q_{ij} = \omega$, where $\omega \in \mathcal{R}$.

Proof :

It is clear that if $E \in \mathcal{E}_m$, for any $E_1 \in \mathcal{E}_m$ we have $EE_1E^T \in \mathcal{E}_m$ and it is easily verified that if $E_2 \in \mathcal{E}_n$, $QE_2Q^T \in \mathcal{E}_m$. Thus any composition of these two transformations, compatible with the dimensions, preserves second-moment exchangeability , which proves that matrices of the form stated in the Theorem belong to $\mathcal{K}_{\mathcal{E}2}$.

Note that if Y has second moment $E_1 \in \mathcal{E}_n$, if K preserves second moment exchangeability we must have $KE_1K^T \in \mathcal{E}_m$. That all matrices preserving exchangeability have the form of the Theorem can be seen by $E_1 = I$. This implies that we have $KK^T = E$, for some $E \in \mathcal{E}_m$. We can write $KK^T = [E^{1/2}][E^{1/2}]^T$. From proposition 1.31 of Eaton (1983), this implies that $K^T = \psi E$, where ψ is an isometric operator from $\mathcal{R}^m$ to $\mathcal{R}^n$. Thus $K = E\psi^T = EQ$; here $Q = \psi^T$ which implies that $QQ^T = I$. There is an additional condition when we apply K to the matrix $\bar{\mathbf{1}}_n$ we must have $K\bar{\mathbf{1}}_n = \omega\bar{\mathbf{1}}_m$, where $\omega \in \mathcal{R}$.

Now it can be seen that ψ^T can be represented by the composition of an isometric operator from $\mathcal{R}^n$ to $\mathcal{R}^n$ and the sub-vector operator: $Q = I_{m,n}U_{n,n}$, where $I_{m,n} = [I_{m,m}|O_{m,n-m}]$ is the "subvector operator" and $U_{n,n}$ is an isometric transformation ($U_{n,n}U_{n,n}^T = I$). Also, geometrically, QY can be interpreted as the projection of the original vector Y on a subspace of dimension m, represented in a orthonormal basis of this subspace whose vectors are the rows of Q.

In particular, transformations from $\mathcal{R}^n$ to $\mathcal{R}^n$ can be written

$$K_{(\alpha,\beta)} = \alpha U + \beta \bar{\mathbf{1}},$$

where U is a $n \times n$ isometric transformation. An example of such matrices are Cholesky decomposition of exchangeable matrices.

It is clear that transformations given in Theorem 1 preserve exchangeability of the first moment too. Thus, for normal distributions, if $K \in \mathcal{K}_{\mathcal{E}2}$, it preserves (global) exchangeability. From this we deduce an interesting theoretical consequence: there are in general several valid permutation distributions for a given statistic. This is not surprising if we consider restricted permutations, for instance permutation on subvectors. More surprising is the fact that, if Y has a normal exchangeable distribution, there exist regular transformations, which preserve exchangeability but not the distribution of the statistics depending on Y, for instance the Cholesky decomposition of an exchangeable variance matrix: consider $E \in \mathcal{K}_{\mathcal{E}2}$ regular; $E = VV^T$; V preserves second-moment exchangeability; if we have a statistic $T = \phi(Y)$, it can be written $T = \tilde{\phi}(\tilde{Y})$, with $\tilde{\phi} = \phi \circ V^{-1}$. If Y has a normal exchangeable distribution, it is also the case for $\tilde{Y}$ and then valid inference for T can be done either by permuting Y or $\tilde{Y}$. It can be verified that the two inferences are really different.

To make things more concrete consider a linear statistic (of the type discussed in 3.2): $T = z^T Y$. If Y is exchangeable and normal under the null hypothesis to be tested, we can use a permutation distribution (that is we use a probablity conditional on $Y_{(o)}$); in this situation $\tilde{Y} = VY$, with $VV^T = E$ and E an exchangeable matrix, is also exchangeable; so, if we rewrite the same statistic as $T = z^T V^{-1} V Y = \tilde{z}^T \tilde{Y}$ we can also use a permutation distribution based on the permutation of $\tilde{Y}$ (that is using a probability conditional on $\tilde{Y}_{(o)}$). These two distributions are different: we illustrate it with a small example with $n = 3$. Take $z^T = (0.26, 0.40, 0.56)$ and $Y^T = (-1.38, 0.119, -1.09)$ (this value of Y was indeed obtained as a realization of three independent normal pseudo-random variables); the value of T is -0.92. The six possible values obtained by permutation of Y are: -1.17, -1.13, -1.00, -0.92, -0.76, -0.73. If the p-value is defined as $P(T \geq -0.92)$, this is equal to 0.5. Take V obtained by a Cholesky decomposition of $I + 2\bar{\mathbf{1}}$. We obtain $\tilde{z}^T = (-0.075, 0.20, 0.49)$ and $\tilde{Y}^T = (-1.78, -0.57, -1.90)$. Then the six possible values obtained by permutation of $\tilde{Y}$ are: -1.25, -1.21, -0.92, -0.85, -0.52, -0.49); this leads to a p-value of 0.66, different from the first one. Thus it is clear that for a given value of Y we can find different p-values for different transformations. We may conjecture that in most situations, the tests obtained would have exactly (or approximately) the same power.

2.3 Linear transformations which preserve global exchangeability

The above results are in apparent contradiction with the result of Dean and Verducci (1990) which stated that linear transformations from $\mathcal{R}^n$ to $\mathcal{R}^n$ preserving exchangeability were of the form:

$$K_{\alpha,\beta} = \alpha P + \beta \bar{\mathbf{1}},$$

where $P \in \mathcal{P}_n$. The difference is that these matrices preserve exchangeability whatever the distribution of Y; we have shown that the set of transformations which preserve exchangeability when Y is normal is larger.

Dean and Verducci (1990) in their corollary 1.1 have characterized the linear transformations which preserve exchangeability. We reproduce here their important result:

Theorem 2 *A linear transformation K from $\mathcal{R}^n$ to $\mathcal{R}^m$ preserves exchangeability if and only if K can be represented in the form $K = [B_1, \ldots, B_t]P$ where $P \in \mathcal{P}_n$, and for each $i = 1, \ldots, t$, B_i is an $m \times n_i$ matrix ($\sum n_i = n$) satisfying the following conditions: if the first column of B_i contains the distinct elements $d_1, \ldots, d_k$ with multiplicities $p_1, \ldots, p_k$ (so that ($\sum p_j = m$), then the n_i columns of B_i consist of the $n_i = m!/\prod p_j!$ distinct permutations of the first column.*

Particular cases are obtained by starting with the form EQ given in Theorem 1 and choosing for the rows of Q permutations of the vector $(1, 0, \ldots, 0)$. If $m = n$ this gives all the transformations but for some $m < n$ the class given by Theorem 2 is larger; that is if we accept a loss of dimension, we have a greater choice for these transformations and this makes it possible to construct useful permutation tests as we will see in a small example in section 3.

It is quite clear that $\mathcal{K}_{\mathcal{E}}$, the set of transformations preserving exchangeability, is smaller than $\mathcal{K}_{\mathcal{E}2}$. For instance the triangular matrix obtained by Cholesky decomposition of an exchangeable matrix is in $\mathcal{K}_{\mathcal{E}2}$ but not in $\mathcal{K}_{\mathcal{E}}$.

2.4 Linear transformations which preserve the permutation distribution of statistics

Consider a statistic $T = \phi(Y)$. If Y is exchangeable we can apply the permutation distribution of T; T takes the values $\phi(P_j Y_{(o)})$ with equal probabilities for the $n!$ possible permutation matrices P_j. Consider linear transformations preserving exchangeability $\tilde{Y} = KY$. We can write T as a function of $\tilde{Y}$: $T = \phi(Y) = \tilde{\phi}(\tilde{Y})$. The question is to find the transformations for which the distributions of T based on the permutation of Y and of $\tilde{Y}$ are the same. It is clear that K must be $n \times n$ since there exist some T which take $n!$ different values when Y is permuted. From Theorem 2 it can be seen that the $n \times n$ transformations must be of the form

$$K = EP = \alpha P + \beta \bar{\mathbf{1}},$$

where $P \in \mathcal{P}_n$ and $E \in \mathcal{R}_n$. Conversely these transformations preserve the distribution of any statistic. To see that, we have first to define the transformed form; it is

$$\tilde{\phi}_{\bar{Y}}(u) = \phi[\alpha^{-1} P^T (u - \beta \bar{Y})].$$

It is easily verified (remembering that $P^T P = I$) that $\tilde{\phi}_{\bar{Y}}(\tilde{Y}) = \phi(Y)$. Similarly, using the fact that the permutation matrices permute, we obtain that for any $P_1 \in \mathcal{P}_n$, $\tilde{\phi}_{\bar{Y}}(P_1 \tilde{Y}) = \phi(P_1 Y)$. This shows that the values of T obtained by permutation of Y or $\tilde{Y}$ are the same, and occuring with the same probability $1/n!$.

Note that these transformations may be of rank n or $n-1$. Note also that $\tilde{\phi}_{\bar{Y}}(.)$ generally depends on $\bar{Y}$. It is interesting here to define "clean" forms. A form ϕ is clean if $\phi \circ \bar{\mathbf{1}} = 0$ or equivalently $\phi \circ (I - \bar{\mathbf{1}}) = \phi$. This property simplifies the study of both conventional (de Jong, 1990) and permutational (see section 4.2) distributions of quadratic forms in particular. If ϕ is clean then it can be seen that $\tilde{\phi}_{\bar{Y}}(u) = \phi(\alpha^{-1} P^T u)$, not depending on $\bar{Y}$. If ϕ is not clean we can consider the clean version of ϕ: $\phi'(.) = \phi \circ (I - \bar{\mathbf{1}})(.)$. Consider a linear form $\phi(Y) = L^T Y$; we define the clean version $T' = \phi'(Y) = L'^T Y$, with $L' = (I - \bar{\mathbf{1}})L$. Then $T' = T + L^T \bar{Y}$; since the second term is constant under permutation of Y, the statistics T and T' lead to the same permutation test.

Finally if we denote $\mathcal{K}_d$ the set of linear transformations which preserve the distribution we have the following inclusions: $\mathcal{E}_n \subset \mathcal{K}_d \subset \mathcal{K}_{\mathcal{E}} \subset \mathcal{K}_{\mathcal{E}2}$.

Dean and Verducci showed that a necessary and sufficient condition for a transformation K of $\mathcal{R}^n$ to $\mathcal{R}^m$, with $m \leq n$, to preserve exchangeability was that for each permutation matrix $P_1 \in \mathcal{P}_m$, one can find a permutation matrix $P_2 \in \mathcal{P}_n$ such that

$$P_1 K = K P_2.$$

2.5 General transformations which preserve exchangeability

The Dean and Verducci condition can be extended to non-linear transformations. A transformation G(.) preserves exchangeability of Y whatever its distribution if and only if for each $P_1 \in \mathcal{P}_m$ there exist $P_2 \in \mathcal{P}_n$ such that

$$P_1 G(Y) = G(P_2 Y).$$

The sufficiency of this condition is simple to prove: $P_1 G(Y) = G(P_2 Y)$, $P_2 Y$ has the same distribution as Y, thus $G(P_2 Y)$ has the same distribution as $G(Y)$. Thus for all $P_1 \in \mathcal{P}_m$, $P_1 G(Y)$ has the same distribution as $G(Y)$ which is the definition of exchangeability. If we want $G(.)$ to preserve exchangeability whatever the distribution of Y, the condition is necessary and this can be proved by choosing a specific distribution, with the same reasoning as in Dean and Verducci.

Let $\mathcal{G}_{\mathcal{E}}$ the set of transformations which preserve exchangeability.

Lemma 1 *$\mathcal{G}_{\mathcal{E}}$ is stable by addition and composition*

Let G_1, G_2 belong to $\mathcal{G}_{\mathcal{E},n,m}$, the subset of transformations from $\mathcal{R}^n$ to $\mathcal{R}^m$ which preserve exchangeability. Applying the general Dean-Verducci condition, for each P_1 there must be a P_2 such that $P_1(G_1 + G_2)(Y) = (G_1 + G_2)(P_2 Y)$. This follows from the linearity of permutation operators and the fact that the Dean–Verducci condition holds separately for G_1 and G_2. We omit the proof for composition which is also very easy.

Identical componentwise non-linear transformations (for example taking the logarithm of all the components) are an example of a non-linear transformation which preserves exchangeability. The transformation which consists in taking the ranks of the observations satisfies the condition and thus preserves exchangeability.

These results can still be extended to the case where observation i is itself a vector. As an application consider a regression problem where $p+1$-uplets $X_i = (Y_i, Z_i)$ are observed, where Y_i is a "dependent" variable and Z_i is a $1 \times p$ vector of explanatory variables. It can be assumed, similarly as in Wei, Lin and Weissfeld (1989), that $X = (X_1, \ldots, X_n)$ is exchangeable; see an application in section 3.5.

3 Transformations towards exchangeability

3.1 General idea

If we start with non-exchangeable observations Y our aim is to find a transformation towards exchangeability. Consider a transformation $\tilde{Y} = V(Y)$ such that $\tilde{Y}$ is exchangeable. If we can write as before $T = \phi(Y) = \tilde{\phi}(\tilde{Y})$, inference can be based on the distribution of T generated by the permutation of $\tilde{Y}$. Except in special cases (see below the normal linear model) it will not be possible to construct an exactly exchangeable $\tilde{Y}$; we shall be content to achieve first and second-moment (in short, second-moment) exchangeability. In section 3.2 we show that residuals in generalized linear models are important because they are "closer" to exchangeability than the original observations, and also because score statistics are simple forms of them. In sections 3.3 and 3.4 we turn to more particular results for the linear model. In section 3.5 we consider an assumption of joint exchangeability of the couple (response, covariate) in regression problems and show its possible use in some special cases. Sections 3.6 and 3.7 consider the more complex case of correlated data.

3.2 First moment exchangeability: residuals and score tests

The first obstacle to exchangeability in regression problems is that the Y_i do not have the same expectation. Assume for the moment that the parameters of the model are known and let us first consider component-wise affine transformations $g_i(Y_i) = a_i Y_i + b_i$ which lead to first moment exchangeability. We must have $\mathrm{E}(a_i Y_i + b_i) = \lambda$ for some λ, hence $b_i = \lambda - a_i \mathrm{E}(Y_i)$ so that $g_i(Y_i) = a_i[Y_i - \mathrm{E}(Y_i)] + \lambda$. This is a linear form of residuals plus a constant; for linear or quadratic forms, the tests do not depend on λ so that if we restrict to component-wise affine transformations we are led to ordinary residuals. Other transformations may lead to other types of residuals; in 3.3 we shall consider general linear transformations. Also, the Cox-Snell residuals (Cox and Snell, 1968) could be used but imply more complex tranformations; this will not be developed here.

An interesting feature is that in generalized linear models (McCullagh and Nelder, 1989), score tests quite generally lead to statistics which are linear or quadratic functions of these ordinary residuals. Assume that under the null hypthesis Y_i are independently distributed with probability density function in the exponential family defined as follows:

$$f(Y_i;\theta_i,\phi) = \exp\left\{\phi^{-1}\left[\theta_i Y_i - g(\theta_i)\right] + C(Y_i,\phi)\right\} \tag{1}$$

with $E(Y_i) = g'(\theta_i) = \mu_i$ and $\theta_i = Z_i\beta$ where $Z_i = (z_i^1,\ldots,z_i^p)$, β is a $p\times 1$ vector of coefficients; here ϕ denotes, as is conventional, the dispersion parameter (and not the function of the data used previouly); the alternative hypothesis specifies that an additional variable z_i^{p+1} may be part of the model so that: $\theta_i = Z_i\beta + z_i^{p+1}\beta_{p+1}$. We will denote by z_{p+1} the vector of the z_i^{p+1}.

Lemma 2 : *Assume the generalized linear model (1), the score statistic for $\beta_{p+1} = 0$ is $T = \phi^{-1} z_{p+1}^T R$ where R is the vector of residuals $R_i = Y_i - \mu_i$ computed under the null hypothesis.*

Finally, consider the model (1) where $\theta_i = Z_i\beta + \alpha_i$, where the α_i are random effects with zero expectation, and the vector of α_i has (known) correlation matrix W and variance matrix ωW. Here we assume that under the alternative the Y_i are independently distributed conditional on the α_i. In this model, homogeneity (or independence of the Y_i) obtains if $\omega = 0$. Then, we have:

Lemma 3 : *The score statistic for $\omega = 0$ is*

$$T = \phi^{-2} R^T W R.$$

The proof of lemma 2 is easy and lemma 3 has been given in Commenges and Jacqmin-Gadda (1997).

Note that for testing purpose we can eliminate ϕ from the statistics and consider the simpler statistics $z^T R$ and $R^T W R$ which lead to the same tests as the previous ones whatever the value of ϕ.

It is tempting to use a test based on permutation of the residuals R since they have first moment exchangeability; in general however they fail to have more.

3.3 Second-moment exchangeability: the linear model

In this section (and the following) we focus on the linear model which is a special case of (1). This model can be written in the simpler way:

$$Y = Z\beta + \varepsilon, \tag{2}$$

where ε is exchangeable with expectation zero; we will assume that the variance of ε is $\sigma^2 I$; Z is the $n \times p$ matrix with rows Z_i. If β were known, the residuals would obviously be exchangeable since $R = Y - Z\beta = \varepsilon$.

In practice β is unknown and we most often use the least-square estimator $\hat{\beta} = (Z^T Z)^{-1} Z^T Y$. The estimated residuals are then $R = Y - Z\hat{\beta} = (I - H)Y$, where $H = Z(Z^T Z)^{-1} Z^T$ is the so-called "hat matrix" and Z is the $n \times p$ matrix of explanatory variables. It can be seen that R is not exchangeable by examining the variance matrix which is $\sigma^2(I - H)$; this, in general does not have the form of an exchangeable matrix and this in turn can be seen by looking at the ranks: it is known that the rank of $I - H$ is $n - p$, while we know from section 2.1 that an exchangeable matrix has rank n, $n - 1$, 1 or 0: the latter two cases are degenerate; the only interesting case is $n - 1$ which occurs if there is only one parameter; we can see that we must in addition have $Z = \bar{\mathbf{1}}$, that is the model must have only an intercept. Most usual non-parametric tests can be derived through the argument derived here, when under the null hypothesis the model contains only an intercept.

In models containing explanatory variables, We could rely on results of Randles (1984) based on the asymptotic independence of the residuals. We can also try to find a transformation of the residuals which gets us closer to exchangeability. Remember that R belongs to the space $\mathcal{Z}^{\perp}$, that is a subspace orthogonal to the subspace $\mathcal{Z}$ spanned by the columns of Z. Thus, if p is the number of explanatory variables, the rank of the variance matrix of R is generally $n-p$; using the constraints already mentioned on the rank of an exchangeable matrix, a rank $n-p$ is compatible with a $(n-p+1)\times(n-p+1)$ or $(n-p)\times(n-p)$ exchangeable matrix; thus a linear transformation towards exchangeability has necessarily to reduce the dimension of R (except if $p=1$).

Theorem 3 : *In the linear model, the linear transformations from $\mathcal{R}^n$ to $\mathcal{R}^m$ of R which achieve second-moment exchangeability can be written in the form*

$$G = EQ + M$$

where E is a $m\times m$ exchangeable matrix, and Q is a $m\times n$ matrix whose rows form an orthonormal basis of F, and F is any subspace included in $\mathcal{Z}^{\perp}$; in addition the elements q_{ij} of Q must be such that $\sum_j q_{ij}=\omega$, $\omega\in\mathcal{R}$; M is any matrix whose rows are in $\mathcal{Z}$.

Let $\tilde{R}=GR$. Since $R=(I-H)Y=(I-H)\varepsilon$, we have: $\tilde{R}=G(I-H)\varepsilon$.

Sufficient condition: if $G=EQ+M$, then $G(I-H)=GQ$ which has the form given by Theorem 1, thus $G(I-H)\in\mathcal{K}_{\mathcal{E}2}$, thus since ε is exchangeable, so is $\tilde{R}$. Necessary condition: the transformations G which lead to second-moment exchangeability must be such that $G(I-H) \in \mathcal{K}_{\mathcal{E}2}$; from Theorem 1, we must have $G(I-H) = E_m Q_{m,n}$. Now, any matrix G can be written $G = G_1 + M$ with $G_1 = G(I-H)$ and $M = GH$. If $G(I-H)$ preserves second-moment exchangeability we have $G_1 = EQ$, and M is any matrix whose rows are in $\mathcal{Z}$. Since the rows of G_1 must be in $\mathcal{Z}^{\perp}$, it must be so of Q, which finishes the proof. Note that if the rank of G is m then E must be of rank m too. The rank of G may also be $m-1$ in which case $E = a(I - \bar{\mathbf{1}})$. Then the rows of Q can be chosen in $\mathcal{Z}^{\perp}$ or in the space orthogonal to all the vectors z_i except the vector 1.

The rank of the variance of $\tilde{R}$ must be lower or equal to $n-p$; in order to keep the maximum power to the permutation test, we have to keep the maximum rank. Then, the simplest choice is to take $E = I_{n-p}$, in which case $\mathrm{var}(\tilde{R}) = \sigma^2 I_{n-p}$.

An orthonormal basis of $\mathcal{Z}^{\perp}$ can easily be constructed, for instance by the Gram-Schmidt procedure, but is not unique. Theil (1965, 1968) has given an algorithm for finding uncorrelated residuals, the so-called BLUS residuals. Theil's residuals can be computed in O(n) operations so that the transformation does not increase the complexity of the computation of the permutational variance for linear or quadratic forms.

To keep the same value of the statistic T we have to compute the inner product (z,R) as a function of $\tilde{R}$; remembering that R is $\mathcal{Z}^{\perp}$, we can compute the scalar product of the projected vector represented in the basis of $\mathcal{Z}^{\perp}$ given by the rows of Q. This leads to $T = \tilde{z}^T \tilde{R}$ with $\tilde{R} = QR$ and $\tilde{z} = Qz$. By the same reasoning, a quadratic form of the residuals $T = R^T W R$ can be written $T = \tilde{R}^T \tilde{W} \tilde{R}$ with $\tilde{W} = QWQ^T$. Thus this approach makes it possible to make inference for a linear or quadratic form of the residuals, using transformed residuals which have second-moment exchangeability.

Corrolary 1 *In the normal linear model, it is possible to find exact permutation tests for linear and quadratic forms of residuals R.*

This comes from the fact that second-moment exchangeability implies exchangeability if Y has a normal distribution. Note that we have not only one but an infinity of different permutation tests obtained by choosing different bases of $\mathcal{Z}^{\perp}$ (see the example at the end of section 2.2).

This result can easily be extended to the case where the error has a known covariance matrix:

$$Y = Z\beta + \varepsilon$$

with $\mathrm{var}(\varepsilon) = \Sigma = V^{-2}$. Then premultiplying by V we obtain the same problem as above, that is $Y' = Z'\beta+\varepsilon'$ with $Y' = VY$, $Z' = VZ$ and $\varepsilon' = V\varepsilon$. The score statistic for testing "$\beta_{p+1} = 0$" is $T = z'^{T}_{p+1}R'$ with $z'_{p+1} = Vz_{p+1}$ and R' the residual in the transformed problem. This transformed residual is still in $\mathcal{Z}^{\perp}$ so that the solution for finding uncorrelated residuals is still to find a matrix Q whose rows are an orthonomal basis of $\mathcal{Z}^{\perp}$. If the distribution of ε is normal then we obtain exact permutation tests.

In the linear model with continuous explanatory variables, it is generally impossible to find an exact distribution-free permutation test because we cannot find a matrix $K = G(I - H)$ which has its rows othogonal to a vector of explanatory variable z and must be of the form given in Theorem 2. However it is possible to find exact permutation tests in some simple cases. The simplest case where it works, as already noted, is the case without explanatory variables, in which case the residuals are directly exchangeable.

In section 3.4 we shall see it also works in a matched pairs design; in more complex designs we might still use transformations of theorem 3 which produce residuals with second moment exchangeability, in fact Theil residuals, and use the adapted test statistic as above; the permutation test will not be stricly valid.

3.4 Example of permutation tests in the linear model

We give an example, a matched pairs design, where an exact permutation test, for a regression parameter and for homogeneity, can be found in the presence of an explanatory variable; this is to illustrate how the general results of section 3.3 lead, in a particular case, to the conventional Wilcoxon test for matched pairs. Consider model (2) in which $n = 2p$ and Z_i has a 1 in the $[(i+1)/2]$ position and zero elsewhere. The residuals are $R_i = Y_i - \bar{Y}_{i.}$ where $\bar{Y}_i = \bar{Y}_{i+1} = \frac{Y_{i+1}+Y_i}{2}, i = 1, 3, \dots, n-1$. Consider the $1 \times n$ vectors $q_i = (0, 0, \dots, 1/\sqrt{2}, -1\sqrt{2}, \dots, 0)$ where the $1/\sqrt{2}$ is in position $2i-1$, $i = 1, \dots, p$. These vectors are orthogonal to the vectors of Z and form an orthonormal basis of $\mathcal{Z}^{\perp}$ and $\sum_j q_{ij} = 0$. Thus, from Theorem 3, the matrix Q which has i^{th} row q_i $(i = 1, \dots, p)$ preserves second-moment exchangeability. The transformed residual vector $\tilde{R} = QR$ is a $p \times 1$ vector with elements $\frac{1}{\sqrt{2}}(Y_{i+1} - Y_i), i = 1, 3, \dots, n-1$. It is easily seen that Q also satisfies the Dean and Verducci condition; it can be verified that Q has the form given by Theorem 2: $Q = [B_1, B_2]P$ with B_1 formed by all the p distinct permutations of the vector $1/\sqrt{2}, 0, \dots, 0$ and B_2 formed by all the p distinct permutations of the vector $-1/\sqrt{2}, 0, \dots, 0$; the permutation P puts the columns in the right order. Thus, Q preserves global exchangeability.

Now consider the model

$$Y = Z\beta + z_{p+1}\beta_{p+1} + \varepsilon,$$

where z_{p+1} is the vector of values of another explanatory variable. The score test for "$\beta_{p+1} = 0$" is by Lemma 2 $T = z_{p+1}^T R = \tilde{z}_{p+1}^T \tilde{R}$ where $\tilde{z}_{p+1} = Qz_{p+1}$. An exact permutation test is then possible for T. If the values of z_{p+1} are binary then only pairs who present the two different values for this variable are kept. If the further (non-linear) exchangeability-preserving transformation $\check{R}_i = \text{rank}(\tilde{R}_i)$ is applied, then the permutation test obtained is the Wilcoxon test for paired observations.

Consider now an homogeneity problem

$$Y = Z\beta + \alpha + \varepsilon_{ij},$$

where α is a vector of random effects with expectation zero and known variance matrix ωW. From Lemma 3, the score test for "$\omega = 0$" is $T = R^T W R = \tilde{R}^T \tilde{W} \tilde{R}$, where $\tilde{W} = QWQ^T$. Then, as before, a valid permutation test is based on the permutation of $\tilde{R}$.

Finally note the contrast between this model and a model in which we would assume only exchangeability between ε_i and ε_{i+1} for all odd i . Then only partial exchangeability can be used with the consequence that the number of permutations is 2^p which is much less than $p!$ as soon as p is larger than 4. We may expect that the permutation test using the global exchangeability assumption for ε is more powerful than the one using the partial exchangeability assumption (which is weaker).

3.5 Exchangeability of (Y, Z)

If we consider the $p+1$-uplets $X_i = (Y_i, Z_i)$ as exchangeable (as in section 2.5), the residuals R are exchangeable (here we consider Z as a random variable). It can be shown by application of lemma 1 that $\hat{\mu}(Y, Z)$ preserves exchangeability and finally that the residual operator $R(X)$ preserves exchangeability; the proof uses also the fact that $\hat{\beta}$ is invariant by permutation that is, $\hat{\beta}(Y) = \hat{\beta}(PY)$ for any permutation matrix P. Thus the residuals are exchangeable in that probability space. This result can be used for deriving permutation tests in particular regression problems. Consider for instance the score test statistic of section 3.2: $T = z_{p+1}^T R$. When using this statistic we condition on the values z_{p+1}; for using permutation test we need exchangeability of $X = (Y, Z)$ conditional on z_{p+1}. This does not hold in general: for instance if Z and z_{p+1} (considered as a random variable) are not independent, the distribution of Z conditionally on z_{p+1} depends on z_{p+1}. It may hold in particular cases, such as in a randomized clinical trial where z_{p+1} represents randomly assigned treatment values (thus making z_{p+1} and Z independent). In that case of course a randomisation test would also be valid: this test would condition on X (hence on R also) and be based on the known-by-design distribution of z_{p+1}. The two tests would coincide only if the distribution of z_{p+1} was obtained by permutation of a given vector.

For homogeneity tests it is more likely that the matrix W of the quadratic form does not depend on Z. For instance, for testing homogeneity of the rates of a disease in a region, the quadratic statistic $T = R^T W R$ can be used where W has entries which depend on the distance d_{ij} between two observations, for instance $w_{ij} = d_{ij}^{-1}$. Jacqmin-Gadda and Commenges (1997) have shown by simulations that the permutation test of residuals had good properties in some situations. If the explanatory variables Z do not depend on the geographical position the permutation homogenity test will be valid. However this assumption may be questionable. On the other hand, in the simulations of the Jacqmin and Commenges paper, Z was generated independently of the geographical position; thus this simulation was hardly worth to do since the theory ensures that the test is valid in that case.

3.6 Second-moment exchangeability: general case

When the data are correlated and the model is not linear normal, exact permutation tests are not available. We may still try to construct approximately uncorrelated residuals, thus achieving second-moment exchangeability and hope that the permutation tests will be robust to departure from global exchangeability.

An approach which has been tried in pratice is the following one. Assume that we observe Y_i, $i = 1, \ldots, n$, which have marginally distributions belonging to the exponential family (1). However we suspect that the Y_i are not independent. For testing a regression parameter, we may still use the same statistic $T = z_{p+1}^T R$, but we cannot directly apply the permutation distribution. If the observations belong to groups (such as familial or geographical groups), we expect the variance matrix of the residuals to be approximately a block diagonal matrix Σ with exchangeable blocks (a BDEB matrix). We estimate Σ and compute $\tilde{R} = \hat{\Sigma}^{-1/2} R$. Then $T = \tilde{z}_{p+1}^T \tilde{R}$ with $\tilde{z}_{p+1} = \hat{\Sigma}^{1/2} z_{p+1}$. For a BDEB matrix relatively simple formulas for computing the transformed residuals and the transformed form (both for the linear and the quadratic case) are given in section 3.7.

An example is given in Commenges and Abel (1996) who treat the problem of genetic linkage using a generalized homogeneity test of the form $T = R^T W R$. In that case, R is the vector of residuals under the hypothesis that the marker is not linked to the disease susceptibility gene, and the elements w_{ij} of W are the number of alleles shared by subjects i and j. However it often happens that even under the null hypothesis, the residuals are correlated within sibships because of genetic or environmental factors. A simulation study has shown that the permutation test based on transformed residuals leads to better type-one errors than the uncorrected test.

The homogeneity test can also be applied in geographical epidemiology. The problem is to test whether the distribution of the disease exhibits a specific spatial pattern. The quadratic statistic $T = R^T W R$ can be used where W has entries which depend on the distance d_{ij} between two observations, for instance $w_{ij} = d_{ij}^{-1}$. Jacqmin-Gadda and Commenges (1997) have shown that the uncorrected permutation test had good properties in some situations. The results of this paper could be used to improve the approximation and to extend the application of this test to the case where the data are correlated even under the null hypothesis.

3.7 The case of the intraclass correlation model

Suppose we observe Y with exchangeable first moment but not second moment. The simplest model with non-exchangeable correlation is the intraclass correlation model: observations belonging to the same group k are correlated with coefficient ρ_k, observations belonging to different groups are uncorrelated. In that case the variance matrix of Y can be written

$$\Sigma = D^{1/2} C D^{1/2},$$

where D is any diagonal matrix, C is a block diagonal matrix with exchangeable blocks (BDEB); the blocks C_k are the exchangeable correlation matrices of group k:

$$C_k = (1 - \rho_k)I_k + n_k\rho_k\bar{\mathbf{1}}_k$$

A square-root decomposition of Σ^{-1} obtains with $\Sigma^{-1/2} = C^{-1/2}D^{-1/2}$ and $C^{-1/2}$ a BDEB matrix with blocks

$$C_k^{-1/2} = a'(\rho_k)I_k + b'_k(\rho_k)\bar{\mathbf{1}}_k,$$

with $a'(\rho_k) = (1 - \rho_k)^{-1/2}$ and $b'_k(\rho_k) = [(n_k - 1)\rho_k + 1]^{-1/2} - (1 - \rho_k)^{-1/2}$. If $D^{-1/2}Y$ is partially exchangeable (that is the subvectors corresponding to the groups have exchangeable distributions), this transformation preserves this property while achieving second-moment exchangeability for the whole vector. Thus $\tilde{Y} = \Sigma^{-1/2}Y$ can easily be computed as

$$\tilde{Y}_i = a'(\rho_{g(i)})Y'_i + b'_{g(i)}(\rho_{g(i)})\bar{Y}'_{g(i)} \tag{3}$$

where $Y'_i = \frac{Y_i}{d_i}$ and $\bar{Y}'_{g(i)}$ is the mean of Y'_i over the group of i, $g(i)$. The transformed form $\tilde{\phi}$ is

$$\tilde{\phi} = \phi \circ D^{1/2}C^{1/2}.$$

If for example ϕ is a linear form we have $\phi(Y) = L^TY$ and $\tilde{\phi} = L^TD^{1/2}C^{1/2}\tilde{Y} = \tilde{L}^TY$. $\tilde{L}$ is obtained by a formula similar to 3

$$\tilde{L}_i = a(\rho_{g(i)})L'_i + b_{g(i)}(\rho_{g(i)})\bar{L}'_{g(i)},$$

with $a(\rho_k) = (1 - \rho_k)^{1/2}$ and $b_k(\rho_k) = [(n_k - 1)\rho_k + 1]^{1/2} - (1 - \rho_k)^{1/2}$.

If ϕ is a quadratic form

$$\phi(Y) = Y^TWY = \tilde{Y}^T\tilde{W}\tilde{Y},$$

with $\tilde{W} = C^{1/2}D^{1/2}WC^{1/2}D^{1/2}$. Using the fact that $C^{1/2}$ is a BDEB matrix with blocks

$$C_k^{1/2} = a(\rho_k)I_k + b_k(\rho_k)\bar{\mathbf{1}}_k$$

$\tilde{W}$ can easily be computed. For the sake of simplicity, we give the formulas for the common intraclass correlation model where $\rho_k = \rho$:

$$\tilde{w}_{ij} = a^2 w_{ij} + ab_{g(j)}\bar{w}_{i;g(j)} + ab_{g(i)}\bar{w}_{g(i);j} + b_{g(i)}b_{g(j)}\bar{w}_{;g(i);g(j)}$$

where $a = \sqrt{(1-\rho)}$, $b_k = \sqrt{(n_k-1)\rho+1} - \sqrt{(1-\rho)}$, $a' = 1/a$, $b'_k = 1/\sqrt{(n_k-1)\rho+1} - 1/\sqrt{(1-\rho)}$ and $\bar{w}_{i;g(j)}$ is the mean of the weights for sujects i and subjects belonging to the same group as j and $\bar{w}_{;g(i);g(j)}$ is the mean of the weights relating subjects of the same groups of i to subjects of the same groups as j; n_k is the size of group k.

4 Doing permutation tests in practice

4.1 Generalities

The direct way for computing the permutation distribution of $T = \phi(Y)$ is to compute all the values of the statistic T for all possible permutations of Y. This is generally impossible for moderately large n. One possibilitity is to use a random sample of permutations. However it is faster to approximate the distribution. A normal approximation may be sufficient; for linear forms, the first four moments are easy to compute, so that the accuracy of the approximation can be improved using saddle-point approximation (Robinson, 1982) in simple cases and Edgeworth approximation more generally. We give here the first and second moments of quadratic forms and the first four moments of linear forms.

4.2 Moments of linear and quadratic forms

Consider a symmetric quadratic form $T = Y^T W Y$. It is useful to decompose T in terms of a so-called "clean" quadratic form T' specified by a matrix W' whose entries are

$$w'_{ij} = w_{ij} - \bar{w}_{.j} - \bar{w}_{i.} + \bar{w}_{..},$$

where $\bar{w}_{.j}$ is the mean over index i of w_{ij} and $\bar{w}_{..}$ is the grand mean. For such a matrix the sum of each line and row is zero. In matrix form the clean version of W is $W' = (I - \bar{\mathbf{1}})W(I - \bar{\mathbf{1}})$. We have that

$$T = T' + 2L' + H^T \bar{Y}$$

where $L' = (H - \bar{H})^T Y$, $H = W\bar{Y}$, $\bar{Y}$ is the vector whose components are the mean of Y, that is $\bar{Y} = \bar{\mathbf{1}}Y$. Then we have that $\mathrm{E}(T) = \mathrm{E}(T') + H^T\bar{Y}$ (because $\mathrm{E}(L') = 0$) and $\mathrm{var}(T) = \mathrm{var}(T') + 4\mathrm{var}(L') + 4\mathrm{cov}(T', L')$. We can compute using similar arguments as in Mantel (1967)

$$\mathrm{E}(T') = \frac{n}{n-1} m_2 \mathrm{Tr}(W')$$

$$\begin{aligned}\mathrm{var}(T') &= \frac{n}{n^{(3)}} m_2^2 [2(n^2 - 3n + 3)\mathrm{Tr}(W'^2) + \frac{n^2 - 3}{n-1}(\mathrm{Tr}(W'))^2 - 3n(n-1)\mathrm{Tr}(D^2)] \\ &+ \frac{nm_4}{n^{(3)}}[-2(n-1)\mathrm{Tr}(W'^2) - (n-1)(\mathrm{Tr}(W'))^2 + n(n+1)\mathrm{Tr}(D^2)]\end{aligned}$$

$$\mathrm{var}(L') = \frac{1}{(n-1)} \sum_{i=1}^{n} (y_i - \bar{y})^2 \sum_{i=1}^{n} (h_i - \bar{h})^2$$

$$\mathrm{cov}(T', L') = \bar{Y} \sum_{i=1}^{n} w_{ii} w_{.i} m_3$$

where D is a diagonal matrix with same diagonal as W', $m_j = \sum_{i=1}^{n} y_i^j / n$; $n^{(3)} = (n-1)(n-2)(n-3)$. This is valid for $n > 3$

Cox and Hinkley (1974) gave the first three moments of linear forms (however with an error in the third moment). From the formula of the variance of a quadratic form, we deduce, the formula for the fourth moment of a linear form. By the transformation $I - \bar{\mathbf{1}}$ we can restrict our attention to clean linear forms $T = L^T Y$. We have $\mathrm{E}(T^4) = \mathrm{var}(Q) + E(Q)^2$, where $Q = T^2 = Y^T L L^T Y$. We apply the formulas above with $W = W' = L^T L$. For this matrix, $\mathrm{Tr}(W'^2) = [\mathrm{Tr}(W')]^2 = (\sum L_i^2)^2 = n^2 \nu_2^2$ and $\mathrm{Tr} D^2 = \sum L_i^4 = n\nu_4$. We obtain for a clean linear form $\mu_1 = 0$, $\mu_2 = \frac{n^2 m_2 \nu_2}{n-1}$ and

$$\mu_4 = \frac{3n^3}{n^{(3)}}(n^2 - 3n + 3) m_2^2 \nu_2^2 - 3 \frac{n^3}{(n-2)(n-3)} (m_2^2 \nu_4 + m_4 \nu_2^2) + \frac{n^3 (n+1)}{n^{(3)}} m_4 \nu_4;$$

We have also $\mu_3 = \frac{n^3}{(n-1)(n-2)} m_3 \nu_3$.

4.3 Cornish-Fischer expansion

The Cornish-Fischer expansion (Kendall and Stuart, 1977; McCune and Gray, 1981) allows to compute the p-percentile of the normal distribution u_p as a function of the p-percentile of a distribution F, x_p, and the first cumulants of F, while the inverse Cornish-Fischer expansion allows to compute the x_p as a function of u_p and the first cumulants of F. Thus the Cornish-Fisher expansion allows to compute a corrected p-value while the inverse Cornish-Fischer expansion allows to compute a corrected percentile for a given size of a test. For instance the formula given by McCune and Gray (1981) for the inverse Cornish-Fisher expansion up to the fourth cumulant (order n^{-1}), valid for a standardized statistic, and corrected for a typographical error for the second term, is

$$x_p = u_p + \frac{1}{6}(u_p^2 - 1)k_3 + \frac{1}{24}(u_p^3 - 3u_p)k_4 - \frac{1}{36}(2u_p^3 - 5u_p)k_3^2$$

The cumulants k_j of the standardized statistic $T/\sqrt{\mu_2}$ are $k_3 = \mu_3/\mu_2^{3/2}$ and $k_4 = \mu_4/\mu_2^2$. For instance this formula can be used for computing the 95%-percentile of the permutation distribution of a standardized statistic by putting $u_p = 1.645$.

Acknowledgements: Daniel Commenges is Biostatistician, Director of Research, Biostatistics Team INSERM E03 38, ISPED, 146 rue Léo Saignat, Bordeaux, 33076, France.

I thank Hélène Jacqmin-Gadda and Frank Gaüzère for helpful discussion.

References

Commenges, D. and Abel, L. (1996), “Improving the robustness of the Weighted Pairwise Correlation test for linkage analysis”. *Genetic Epidemiology* 13, 559-573.

Commenges, D. and Jacqmin–Gadda, H. (1997), “Generalized score test of homogeneity based on correlated random effect models”. *J Roy Stat Soc B* 59, 157-171.

Cox, D. R. and Hinkley, D. V. (1974), “Theoretical Statistics”. London: Chapman and Hall.

Cox, D. R. and Snell, E. J. (1968), “A general definition of residuals”. *J Roy Stat Soc B* 30, 248-75.

de Finetti, B. (1990), ”Theory of Probability”. Wiley and Sons: Chichester.

de Jong, P. (1987), "A central limit theorem for generalized quadratic forms". *Prob. Th. Rel. Fields*, 75, 261-277.

Dean, A. M. and Verducci, J. S. (1990), "Linear transformations that preserve majorization, Schur concavity, and exchangeability". *Linear algebra and its applications*, 127, 121-138.

Dean, A. M. and Wolfe, D. A. (1990), "A note on exchangeability of random variables". *Statistica Neerlandica*, 44, 23-27.

Eaton, M. L. (1983), "Multivariate statistics: a vector space approach". Wiley and Sons.

Jacqmin–Gadda, H., Commenges, D, Nejjari, C. and Dartigues, JF. (1997), "Test of geographical correlation with adjustment for explanatory variables: application to dyspnoea in the elderly". *Statistics in Medicine*, 16, 1283-1297.

Good, P. (1992), "Permutations tests". Berlin: Springer-Verlag.

Mantel, N. (1967), " The detection of disease clustering and a generalized regression approach". *Cancer Research*, 27 Part 1, 209-220.

McCullagh, P., and Nelder, J.A. (1989) *Generalized Linear Models* (2nd ed.), New York: Chapman & Hall.

McCune, E. D. and Gray, H. L. (1981), "Cornish-Fisher and Edgeworth expansions" in *Encyclopedia of Statistical Sciences*, eds. S. Kotz and L. Johnson, New-York: John Wiley, 188-193.

Randles, R. H. (1984), "On tests applied to residuals". *Journal of the American Statistical Association*, 79, 349-354.

Robinson, J. (1982), Saddlepoint approximation s for permutation tests and confidence intervals. *J Roy Statist Soc B*, 91-101.

Romano, J.P. (1990), "On the behaviour of randomization tests without a group invariance assumption". *Journal of the American Statistical Association*, 85, 686-692.

Schmoyer, R. L. (1994), "Permutation tests for correlation in regression errors". *Journal of the American Statistical Association*, 89, 1507-1516.

Theil, H. (1965), "The analysis of disturbances in regression analysis". *Journal of the American Statistical Association*, 60, 1067-1079.

Theil, H. (1968), "A simplification of the BLUS procedure for analyzing regression disturbances". *Journal of the American Statistical Association*, 63, 242-251.

Wei L. J. , Lin D. Y. and Weissfeld L. (1989), "Regression analysis of multivariate incomplete failure time data by modeling marginal distributions". *Journal of the American Statistical Association*, 84, 1065-1073.

Welsh, W. J. (1990), "Construction of permutation tests". *Journal of the American Statistical Association*, 85, 693-698.