RANDOM EFFECTS MODEL: NONPARAMETRIC CASE

Z. Govindarajulu AND Jayant V. Deshpandé

(Received Nov. 16, 1970)

1. Introduction and summary

The random effects model for the analysis of variance has been dealt with adequately in Scheffé ([8], see Ch. 7). For one factor experiments the model consists of the following. We have random variables \(X_{ij}, j=1, 2, \cdots, n_i\) and \(i=1, 2, \cdots, c\), such that

\[X_{ij} = \mu + Y_i + \varepsilon_{ij},\]

where \(\{Y_i\}\) and \(\{\varepsilon_{ij}\}\) are completely independent random variables. Assuming that the variances exist, \(\nu(X_{ij}) = \nu(Y_i) + \nu(\varepsilon_{ij})\). The null hypothesis usually to be tested, then, is

\[H_0: \nu(X_{ij}) = \nu(\varepsilon_{ij}) \quad \text{or equivalently} \quad \nu(Y_i) = 0 \quad \text{for all } i,
\]

or the more general

\[H'_0: \nu(Y_i) \leq \theta \nu(\varepsilon_{ij}) \quad \text{for all } i, j,
\]

where \(\theta(\geq 0)\) is a preassigned constant. The general idea is to test the hypothesis that \(Y_i\) have very little, if any, dispersion as compared to \(\varepsilon_{ij}\).

In the classical analysis, developed by Scheffé and others, \(Y_i\) and \(\varepsilon_{ij}\) are further assumed to have normal distributions \(N(0, \sigma_i^2)\) and \(N(0, \sigma^2)\) respectively. Greenberg [4] has considered a more general model wherein \(Y_i\) are still assumed to have normal distribution but the \(\varepsilon_{ij}\) may have an arbitrary continuous distribution function \(F\) with density function \(f\) and variance \(\sigma^2\). Thus she develops partially distribution-free tests for the hypotheses quoted above and for hypotheses involving more general nested designs.

In this paper we try to develop completely distribution-free tests for the above. In particular, we are able to obtain the locally most powerful tests for the hypotheses described above and also for some related hypotheses. In Section 2 we present the main results of this paper, viz., the derivation of these locally most powerful tests. In
Section 3, using well-known results, we establish the asymptotic normality of the test statistics obtained in Section 2. In Section 4 we discuss the limitations that distribution-free or rank tests have for testing such hypotheses. A list of references follows.

2. Locally most powerful rank tests

(a) Let us consider the following model. We have random variables \( X_{ij}, j=1, 2, \ldots, n_i; i=1, 2, \ldots, c \), such that \( X_{ij} = Y_i + \varepsilon_{ij} \), where \( Y_i \) and \( \varepsilon_{ij} \) are completely independent random variables: \( \varepsilon_{ij} \) are identically distributed with distribution \( F \); and \( Y_i \) have distribution \( G_i \) respectively. We are interested in testing the null hypothesis

\[
H_0 : G_i(x) = \begin{cases} 
0 & \text{if } x < 0 \\
1 & \text{if } x \geq 0, \ \forall i.
\end{cases}
\]

We would either like to accept the null hypothesis or reject it in favour of the alternative which says that at least one \( G_i \) is nontrivial. In order to derive the locally most powerful (LMP) rank test, we consider the alternative

\[
H_\Delta : X_{ij} = \Delta Y_i + \varepsilon_{ij}, \quad \text{for small } \Delta,
\]

where not all \( \mathcal{E}(Y_i) \) are the same.

Let \( W_1 < W_2 < \cdots < W_N, \ (N=\sum_{i=1}^{c} n_i) \), denote the combined ordered sample and let \( Z=(Z_1, Z_2, \ldots, Z_N) \) denote the \( c \)-sample rank order, i.e., \( Z_i = j \) if \( W_i = X_{jk} \) for some \( k=1, 2, \ldots, n_j \). Let \( z \) be a possible realization of the \( \prod_{i=1}^{c} (n_i!) \) possible rank orders.

It is assumed that \( X_{ij} \) has conditional distribution \( F_i \) with density \( f_i \) for given \( Y_i = y_i \) \((i=1, 2, \ldots, c)\), then one may write

\[
P[Z=z|Y_j=y_j, \ \forall j] = \left( \prod_{i=1}^{c} (n_i!) \right) \left( \prod_{i=1}^{N} \left[ \sum_{j=1}^{c} f_j^{z_{j,i_1}}(w_i) \right] \right) \prod_{i=1}^{N} dw_i
\]

where \( f_j(w_i) = f(w_i - y_j) \) and \( \delta_{j,i_1}=1 \) if \( z_i = j \) and zero otherwise. Note that \( \sum_{j=1}^{c} \delta_{j,i_1}=1 \).

Now, it may be seen that

\[
P[Z=z|H_\Delta] - P[Z=z|H_0] = \left( \prod_{j=1}^{c} \left[ \sum_{i=1}^{n_j} P(Z=z|H_\Delta; \ Y_j=y_j, \ j=1, 2, \ldots, c) \right] \right) - P(Z=z|H_0) \prod_{j=1}^{c} dG_j(y_j).
\]
Consider

\[ P[Z=z|H_1; Y_j = y_j, j=1, 2, \ldots, c] - P[Z=z|H_0] = \prod_{i=1}^{c} (n_i!) \left[ \prod_{j=1}^{c} \left( \prod_{i=1}^{N} \left( \int_{-\infty}^{w_i} f^{i,j,1}(w_i - \Delta y_i) \right) \right)^{N} \int_{1=1}^{N} dw_i \right]. \]

If we now assume the regularity conditions stated below in Theorem 2.1, it is seen that using the dominated convergence theorem of Lebesgue, one may write

\[ \lim_{d \to 0} \frac{1}{d} \{ P[Z=z|H_1; Y_j = y_j, j=1, 2, \ldots, c] - P[Z=z|H_0] \} = \prod_{i=1}^{c} (n_i!) \left[ \prod_{j=1}^{c} \left( \prod_{i=1}^{N} \left( \int_{-\infty}^{w_i} f^{i,j,1}(w_i - \Delta y_i) \right) \right) \right] \lim_{d \to 0} \frac{1}{d} \{ \int_{1=1}^{N} \left( \int_{j=1}^{N} f^{i,j,1}(w_i) \right)^{N} \int_{1=1}^{N} dw_i \}. \]

And hence

\[ \lim_{d \to 0} \frac{1}{d} \{ P[Z=z|H_1] - P[Z=z|H_0] \} = \left( \begin{array}{cc} N & n_1 \end{array} \right)^{-1} \left( \begin{array}{c} n_1 \end{array} \right) \sum_{j=1}^{n_1} \mathcal{E}(Y_j) \sum_{j=1}^{n_1} \mathcal{E} \left( \frac{f'(W_j)}{f(W_j)} \right) \]

where \( \left( n_1, \ldots, n_c \right) = N! / n_1! \cdots n_c! \) and s_1, s_2, \ldots, s_{n_1} are the ranks of the l_th sample in the combined ordered sample. The derivations of the above results are not dissimilar to those of Hájek [6] in the case of the bivariate independence problem, and hence are omitted here. We now have

**Theorem 2.1.** If (i) \( F \) has a density \( f \) which is absolutely continuous on finite intervals, (ii) \( f'(\cdot) \) is continuous almost everywhere, (iii) \( G \) is such that \( \mathcal{E}|Y_j| < \infty \) for each \( j \), and \( \mathcal{E}(Y_j) \) are not all equal; and (iv) \( \int_{-\infty}^{\infty} |f'(t)| dt < \infty \), then the test with the critical region

\[ \sum_{i=1}^{c} \mathcal{E}(Y_i) \sum_{j=1}^{n_1} \mathcal{E} \left[ \frac{-f'(W_j)}{f(W_j)} \right] \geq K \]

is the LMP rank test for \( H_0 \) against \( H_1 \) where \( K = K(\alpha) \) and \( \alpha \) denotes the level of significance.

(b) As a special case of this theorem we may consider the LMP
test for the corresponding two sample problem, viz., \( c=2 \). The test statistic is

\[
\mathcal{E}(Y_i) \sum_{k=1}^{n_k} \mathcal{E} \left[ \frac{-f'(W_{k})}{f(W_{k})} \right] + \mathcal{E}(Y_i) \sum_{k=1}^{n_k} \mathcal{E} \left[ \frac{-f'(W_{rk})}{f(W_{rk})} \right]
\]

where \( s_k \) and \( r_k \) are the ranks of \( X_{ik} \) and \( X_{ik} \) respectively in the combined ordered sample for \( k=1, 2, \ldots, n_i, \ i=1, 2 \).

(c) Let us now consider the following one-sample situation:

\[
H_0: X_i = \varepsilon_i
\]

and

\[
H: X_i = Y_i + \varepsilon_i, \quad i=1, 2, \ldots, n.
\]

Here the \( \varepsilon_i \) are assumed to be independent and identically distributed with distribution function \( F \). The \( Y_i \) and \( \varepsilon_i \) are assumed to be completely independent. In order to obtain the LMP test, we consider the probability of the rank orders under

\[
H_0: X_i = \Delta Y_i + \varepsilon_i \quad \text{and let } \Delta \to 0.
\]

Under the same conditions (quoted in Th. 2.1) on \( F \) and \( G \) we obtain that the LMP test with size \( \alpha \) has to use

\[
\sum_{k=1}^{n} \mathcal{E} \left[ \frac{-f'(X_{(k)})}{f(X_{(k)})} \right] \mathcal{E}(Y_k) \leq C
\]

as the critical region. Here \( X_{(k)} \) is the \( k \)th order statistic from the random sample \( X_1, \ldots, X_n \). This is really the special case of the result in Theorem 2.1 with \( n_1 = n_2 = \cdots = n_c = 1 \) and \( c = n \).

(d) We might also consider the following local alternatives which yield different LMP rank tests. If \( H_0: Y_i + \varepsilon_{ij} \) and

\[
H_d: X_{ij} = Y_i (1 + \Delta) + \varepsilon_{ij}, \quad j=1, 2, \ldots, n_i \text{ and } i=1, 2, \ldots, c,
\]

then the LMP rank test is of the form: Reject \( H_0 \) if

\[
\sum_{i=1}^{c} \mathcal{E} \left[ Y_i \sum_{j=1}^{n_i} \mathcal{E} \left( \frac{-f'(W_{ij} - y_i)}{f(W_{ij} - y_i)} \right) \right] < K.
\]

(e) If \( H_0: X_{ij} = \varepsilon_{ij} \) and

\[
H_1: X_{ij} = \varepsilon_{ij} (1 + \Delta Y_i), \quad j=1, 2, \ldots, n_i, \ i=1, 2, \ldots, c,
\]

then the LMP rank test is of the form: Reject \( H_0 \) if

\[
\sum_{i=1}^{c} \mathcal{E}(Y_i) \sum_{j=1}^{n_i} \mathcal{E} \left( -W_{ij} \frac{f'(W_{ij})}{f(W_{ij})} \right) > K.
\]
3. Asymptotic normality of the test statistics

If we assume the normal density for \( f \), the conditions for Theorem 2.1 are met and the statistics obtained are variations of the normal scores statistics with the additional \( \mathcal{E}(Y) \) factor coming in. Under the null hypothesis, when all the observations are independent, the asymptotic normality, after standardization, of these statistics follows from the results of Chernoff and Savage [2] and Govindarajulu, LeCam and Raghavachari [3], etc. In the one sample case, 2(c), treated above the observations are completely independent under the null as well as the alternative hypotheses, and consequently the asymptotic normality holds under all alternatives.

Even if we do not assume the normal density function for \( f \), the asymptotic normality of the statistics of the kind

\[
S = \sum_{i=1}^{N} a_i b(W_{a_i}),
\]

where \( b(W_{a_i}) \) may be of the form \( \mathcal{E}[-(f'(W_{a_i})/f(W_{a_i}))] \), under various conditions has been investigated by several authors including Hájek [5].

4. Remarks

The interesting fact to be noticed here is that although in the normal theory these tests reduce to tests of dispersions and one uses tests based on estimates of different variances, in the nonparametric set up we again come up with tests based on normal scores and the expectations of the random effects. Obviously, if all the random effects had the same expectations, none of our test statistics would be valid. This seems to indicate that although we have introduced random effects, rank tests seem to be able to distinguish only between their locations. This observation led us to introduce the model considered in 2(e). We do not know if it is a realistic model for any practical situation. In this model the random factor \( Z \) is a scale factor (one may very easily consider the corresponding several sample situation), rather than a location factor as in the usual models. The LMP rank test obtained with normal alternatives for this problem is similar to that obtained by Capon [1] for the problem of scales in the two-sample fixed effects situation. All this reinforces the thesis proposed by Moses [7] and others that no rank test can hope to be a satisfactory test against dispersion alternatives without some sort of strong assumption being placed on the class of admissible distributions.

This paper covers only one factor experiments. The derivations and the LMP tests become rather complicated with more factors. It
is believed, however, that there will not be any serious difficulties involved.

University of Kentucky
Case Western Reserve University

References


