Testing Independence with Additional Information

by

Francois Perron
Université de Montréal and Purdue University

Technical Report # 90–24C

Department of Statistics
Purdue University

May, 1990

△

TESTING INDEPENDENCE WITH ADDITIONAL INFORMATION

by

Francois Perron

Université de Montréal and Purdue University

ABSTRACT

Summary. Let $X = (X^j : j = 1, \ldots, n)$ be n row vectors of dimension p independently and identically distributed multinormal. For each j, X^j is partitioned as $X^j = (X_1^j, X_2^j, X_3^j)$ where p_i is the dimension of X_i^j with $p_1 = 1, p_1 + p_2 + p_3 = p$. In addition, consider $Y_i^j, i = 1, 2, j = 1, \ldots, n_i$ where the vectors Y_i^j are independent and distributed as X_i^1. We treat here the problem of testing independence between X_1^1 and X_3^1 knowing that X_1^1 and X_2^1 are uncorrelated. A locally best invariant test is proposed for this problem.

Key words: Invariance, group, orbits, maximal invariant.

Résumé. À partir de vecteurs aléatoires lignes indépendants entre eux $(X^j : j = 1, \ldots, n)$ distribués selon une loi multinormale de dimension p on cherche à tester s'il y a indépendance ou non entre la première composante et les p_3 dernières composantes des vecteurs X^j sachant que la première composante est indépendante des p_2 composantes suivantes. Pour cela, on partitionne tous les vecteurs X^j en sous-vecteurs $(X^j = X_1^j, X_2^j, X_3^j)$ de dimensions $p_1 = 1, p_2, p_3$ respectivement. Dans ce problème on suppose qu'il existe également des vecteurs $(Y_i^j : i = 1, 2, j = 1, \ldots, n_i)$ dont les distributions en loi sont les mêmes que celles des marginales X_i^1 de X^1. Ces vecteurs forment l'information additionnelle. Dans cet article, on trouve un test localement le plus puissant pour ce problème.

AMS 1985 subject classifications: Primary 62H15; secondary 62A05, 62F03.

Supported by the National Sciences and Engineering Research Council of Canada.

1. Introduction

Consider a random vector in $\mathbb{R}^p$ having a multivariate normal distribution with mean 0 and covariance Σ (Σ positive definite). Decompose this vector into three subvectors of dimensions $p_1 = 1, p_2$ and p_3. Assuming that the first subvector is independent of the second subvector we wish to test whether the first subvector is independent of the third subvector or not. Partitioning Σ in an appropriate way, the problem corresponds to testing $H_0 : \Sigma_{13} = 0$ versus $H_1 : \Sigma_{13} \neq 0$ with

$$\Sigma = \begin{pmatrix} \Sigma_{11} & 0 & \Sigma_{13} \\ 0 & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix}$$

where Σ_{ij} is a $p_i \times p_j$ matrix $i, j = 1, 2, 3$.

In order to test this hypothesis a sample consisting of three random matrices (X, Y_1, Y_2) of dimensions $n \times p$, $n_1 \times p_1$ and $n_2 \times p_2$ respectively is observed. It is assumed that X, Y_1 and Y_2 are mutually independent and distributed as $N(0, I_n \otimes \Sigma), N(0, I_{n_1} \otimes \Sigma_{11})$ and $N(0, I_{n_2} \otimes \Sigma_{22})$ respectively. In other words, the row vectors of X are independent and identically distributed $N_p(0, \Sigma)$ and, similarly, the row vectors of Y_i are independent and identically distributed $N(0, \Sigma_{ii}), i = 1, 2$. The matrix X is partitioned as $X = (X_1 X_2 X_3)$ where X_i is $n \times p_i, i = 1, 2, 3$. Let $S_{ij} = X_i^t X_j$ and $W_k = Y_k^t Y_k, k = 1, 2, i, j = 1, 2, 3$. The matrices S, W_1, W_2 are independent Wishart and they form a sufficient statistic. The couple (W_1, W_2) is what we call the additional information.

In practice, additional information appears in many circumstances. It occurs in sample surveys when two types of questionnaires (one partial and the other one complete) are distributed. It may also occur when we combine data obtained from different experiments.

The likelihood ratio test for testing H_0 versus H_1 is given by rejecting H_0 if and only if R is large where

$$R = \frac{S_{13.2}S_{33.2}^{-1}S_{31.2}/p_3}{S_{11.2}/(n-p_2)}$$

and $S_{ij.k} = S_{ij} - S_{ik}S_{kk}^{-1}S_{kj}$. This test does not take into account the additional information. When no additional information is available the locally best invariant test (Giri 1979) is given by rejecting H_0 if and only if φ_1 is large where

$$\varphi_1 = (R-1)\left(\frac{S_{11.2}/(n-p_2)}{S_{11}/n}\right).$$

The purpose of this article is to find the locally best invariant test when additional information is available. It is expected that this test will make use of (W_1, W_2). For instance, consider the following related problem where $p_1 \geq 1, p_2 = 0$ and W_3 is $Wp_3(\Sigma_{33}, n_3)$. In this case the locally best invariant test (Eaton and Kariya 1983) consists in rejecting H_0 if and only if φ_2 is large where

$$\varphi_2 = \frac{(n+n_1)(n+n_3)}{p_1p_3} tr((S_{11}+W_1)^{-1}S_{13}(S_{33}+W_3)^{-1}S_{31}) - \sum_{i=1,3} \frac{(n+n_i)}{p_i} tr((S_{ii}+W_i)^{-1}S_{ii}).$$

Notice how the statistic φ_2 makes use of (W_1, W_3).

The invariance context is introduced in Section 2 and the locally bets invariant test is given. Computations are reported in Section 3. Throughout this text the multiplicative group of $p \times p$ non-singular matrices is denoted by $G_\ell(p)$. The subgroups of orthogonal matrices and lower triangular matrices with positive elements on the diagonal are respectively denoted by $O(p)$ and $T(p)$.

2. Invariance context

Let S, W_1, W_2 be independently distributed Wishart with parameters $(p, \Sigma, n), (p_1, \Sigma_{11}, n_1)$ and (p_2, Σ_{22}, n_2) respectively. Let P_Σ be the probability measure associated with (S, W_1, W_2) when Σ is the parameter and let G be the group of transformations given by

$$G = \left\{ g = \begin{pmatrix} g_{11} & 0 & 0 \\ 0 & g_{22} & 0 \\ 0 & g_{32} & g_{33} \end{pmatrix} : g \in G_\ell(p), g_{ii} \in G_\ell(p_i) \text{ for } i = 1, 2, 3 \right\}.$$

Corresponding to $g \in G$ the transformations on the sufficient statistic (S, W_1, W_2) and the parameter Σ are given by

$$(S, W_1, W_2) \longrightarrow \tilde{g}(S, W_1, W_2) = (gSg^t, g_{11}^2 W_1, g_{22} W_2 g_{22}^t),$$

$$\Sigma \longrightarrow g \Sigma g^t$$

and a maximal invariant in the parameter space is given by ρ where $\rho(\Sigma) = \Sigma_{13} \Sigma_{33.2}^{-1} \Sigma_{31} \Sigma_{11}^{-1}$.

The power function of any invariant test is constant on each orbit of the parameter space so there is no loss of generality in working on a class of representatives instead of working on the original parameter space. Let

$$A(\delta) = \begin{pmatrix} 1 & 0 & \delta D^t \\ 0 & I & 0 \\ \delta D & 0 & I \end{pmatrix}^{-1} \tag{2.1}$$

with $D^t = (0, \ldots, 0, 1)$. The set $\{A(\delta) : \delta \in [0, 1)\}$ consists in a class of representatives for the orbits of the parameter space and $\rho(A(\delta)) = \delta^2$. The last inequality indicates a bijection between the class of representatives and the range of ρ.

Define P_δ^I as the restriction of $P_{A(\delta)}$ over the class of all measurable sets which are invariant. According to Theorem 1 Section 3

$$\frac{dP_\delta^I}{dP_0^I}(S, W_1, W_2) = (1 - \delta^2)^{n/2} [1 + \frac{(n + n_1)}{2} + \frac{(n - p_2)}{2} (R - 1) \frac{S_{11.2}/(n - p_2)}{(S_{11} + W_1)/(n + n_1)} + 0(\delta^3)].$$

Therefore, the locally best invariant test is given by rejecting H_0 if and only if φ_3 is large where

$$\varphi_3 = (R-1)\frac{S_{11.2}/(n-p_2)}{(S_{11}+W_1)/(n+n_1)}.$$

The statistic φ_3 is the product of two factors. The first factor is equivalent to the likelihood ratio statistic. It essentially measures the multiple correlation between X_1 and X_3 after removing the effect of X_2. The second factor is a ratio of two estimators of Σ_{11}. The additional information is used to get an improved estimator in the denominator. (Giri's test has $n_1 = 0$ and $W_1 = 0$.) The second factor provides a measure of orthogonality between X_1 and the columns of X_2. The fact that this test is locally most powerful suggests that, as X_1 becomes more nearly orthogonal to the columns of X_2, the first factor becomes more effective in detecting near-zero correlation.

A comment on the fact that the test does not involve Y_2 : Giri's test uses X_2 only through the projection matrix $X_2(X_2^t X_2)^{-1}X_2^t$, which contains no information about Σ_{22}. Thus it is not surprising that additional information about Σ_{22} is ignored.

3. Computations

In this section, we derive expression (2.2). Using Wijsman's representation theorem (ref. Wijsman 1967)

$$\frac{dP_\delta^I}{dP_0^I}(S, W_1, W_2) = \frac{r(\delta, S, W_1, W_2)}{r(0, S, W_1, W_2)}$$

with

$$r(\delta, S, W_1, W_2) = \int_G f(\tilde{g}(S, W_1, W_2)|A(\delta))\lambda(dg)$$

where λ is a left invariant measure on G and $f(\cdot|\Sigma)$ is the density function of (S, W_1, W_2)

with respect to an invariant measure μ when the parameter is Σ. The measures λ and μ are unique up to a multiplicative factor. Let

$$\lambda(dg) = |g_{33}g_{33}^t|^{-p_2/2} \prod_{i=1}^{3} \lambda_{p_i}(dg_{ii})$$

and

$$\mu(d(S, W_1, W_2)) = \mu_p(dS)\mu_{p_1}(dW_1)\mu_{p_2}(dW_2)$$

where $\lambda_q(dh) = |hh^t|^{-q/2}dh$ is a left invariant measure on the space of all $q \times q$ matrices and $\mu_q(dW) = |W|^{-(q+1)/2} \prod_{1 \leq i \leq j \leq q} dw_{ij}$ is an invariant measure on the space of all $q \times q$ positive definite matrices. The joint density of (S, W_1, W_2) with respect to the measure μ is given by

$$f(S, W_1, W_2|\Sigma) = K|\Sigma^{-1}S|^{n/2} \prod_{i=1}^{2} |\Sigma_{ii}W_i|^{n_i/2}$$

$$\exp -\frac{1}{2}\{tr(\Sigma^{-1}S) + \sum_{i=1}^{2} tr(\Sigma_{ii}^{-1}W_i)\}$$

where K is a normalization constant independent of Σ.

Theorem 1.

$$\frac{dP_\delta^I}{dP_0^I}(S, W_1, W_2) = (1-\delta^2)^{n/2}\left[1 + \frac{(n+n_1)}{2} + \frac{(n-p_2)}{2}\{R-1\} \frac{S_{11.2}/(n-p_2)}{(S_{11}+W_1)/(n+n_1)} + 0(\delta^3)\right] .$$

Proof. First, begin with

$$f(\tilde{g}(S, W_1, W_2|A(\delta)) = K(1-\delta^2)^{n/2}|S|^{n/2}W_1^{n_1/2}|W_2|^{n_2/2}$$

$$(g_{11}^2)^{(n+n_1)/2}|g_{22}g_{22}^t|^{(n+n_2)/2}|g_{33}g_{33}^t|^{n/2}$$

$$\exp -\frac{1}{2}\{(S_{11} + (1-\delta^2)W_1)g_{11}^2 + tr(g_{22}(S_{22}+W_2)g_{22}^t) + tr(g_{33}S_{33}g_{33}^t)$$

$$+ tr(g_{32}S_{22}g_{32}^t) + 2tr(g_{32}S_{23}g_{33}^t) + 2\delta g_{11}D^t g_{32}S_{21} + 2\delta g_{11}D^t g_{33}S_{31}\}.$$

Introducing the variable transformation

$$h_{32} = [g_{32} + g_{33}S_{32}S_{22}^{-1} + \delta D g_{11} S_{12} S_{22}^{-1}] S_{22}^{1/2}$$

$$h_{ii} = g_{ii} V_i, V_1 = (S_{11} + W_1)^{1/2}, V_2 = (S_{22} + W_2)^{1/2}, V_3 = S_{33.2}^{1/2}$$

$$\lambda(dg) = |S_{22}|^{-p_3/2} \lambda(dh).$$

We get

$$r(\delta, S, W_1, W_2) = K(1-\delta^2)^{n/2} |S|^{n/2} |S_{22}|^{-p_3/2} V_1^{-(n+n_1)} |V_2|^{-(n+n_2)}$$

$$|V_3|^{-n} W_1^{n_1/2} |W_2|^{n_2/2} \int_G (h_{11}^2)^{(n+n_1)/2} |h_{22}h_{22}^t|^{(n+n_2)/2} |h_{33}h_{33}^t|^{(n-p_2)/2}$$

$$\exp -\frac{1}{2}\{h_{11}^2 + tr(h_{22}h_{22}^t) + tr(h_{33}h_{33}^t) + tr(h_{32}h_{32}^t)$$

$$+ 2\delta D^t h_{33} V_3^{-1} S_{31.2} V_1^{-1} h_{11} - \delta^2 h_{11}^2 S_{11} V_1^{-2}(R_1 + Z)\} \lambda(dh)$$

with $Z = S_{11}^{-1} W_1$ and $R_1 = S_{12} S_{22}^{-1} S_{21} / S_{11}$.

Integrating over h_{22}, h_{32} and developing the exponential for δ close to zero we find

$$r(\delta, S, W_1, W_2) = (1-\delta^2)^{n/2} b(S, W_1, W_2)$$

$$\{\int_{G_\ell(p_1)} \int_{G_\ell(p_3)} [1 - D^t h_{33} V_3^{-1} S_{31.2} V_1^{-1} h_{11}\delta + \frac{1}{2}((D^t h_{33} V_3^{-1} S_{31.2})^2 +$$

$$S_{11}(R_1 + Z))\delta^2 V_1^{-2} h_{11}^2](h_{11}^2)^{(n+n_1)/2} |h_{33}h_{33}^t|^{(n-p_2)/2} \exp -\frac{1}{2}\{h_{11}^2 + tr(h_{33}h_{33}^t)\}$$

$$\lambda_{p_1}(dh_{11}) \lambda_{p_3}(dh_{33}) + Q(\delta, S, W_1, W_2)\}$$

where $Q(\delta, S, W_1, W_2) = O(\delta^3)$ uniformly in (S, W_1, W_2) (can be proven).

Next, decompose $G_\ell(p)$ as a product space $G_\ell(p) = T(p) \times O(p)$ where $g = to$, $g \in G_\ell(p), t \in T(p)$ and $o \in O(p)$. According to this decomposition the left invariant measure λ_p is decomposed as $\lambda_p(dg) = \tau_p(dt) \times \nu_p(do)$ where τ_p is a left invariant measure on $T(p)$ and ν_p is a left invariant probability measure on $O(p)$. Moreover, corresponding to ν_p the following identities (James 1954).

$$\int_{O(p)} tr(AO)\nu_p(dO) = 0, \int_{O(p)} tr^2(AO)\nu_p(dO) = tr(AA^t)/p, A(p \times p)$$

are used to obtain

$$r(\delta,S,W_1,W_2)=(1-\delta^2)^{n/2}b(S,W_1,W_2)\{\int_{T(p_1)}\int_{T(p_3)}$$

$$[1+\frac{1}{2}(D^tT_{33}T_{33}^tDR_2/p_3+Z+R_1)(1+Z)^{-1}T_{11}^2\delta^2]T_{11}^{n+n_1}|T_{33}|^{n-p_2}$$

$$\exp-\frac{1}{2}\{T_{11}^2+tr(T_{33}T_{33}^t)\}\tau_{p_1}(dT_{11})\tau_{p_3}(dT_{33})+O(\delta^3)\}.$$

Finally, recognizing the Bartlett decomposition of a Wishart distribution (ref. Giri 1977) we obtain

$$r(\delta,S,W_1,W_2)/r(0,S,W_1,W_2)=(1-\delta^2)^{n/2}$$

$$\{1+\frac{1}{2}E\{D^tU_1DR_2/p_3+Z+R_1\}U_2(1+Z)^{-1}\delta^2+O(\delta^3)\}$$

$$=(1-\delta^2)^{n/2}\{1+\frac{(n+n_1)}{2(1+Z)}\{\frac{(n-p_2)}{p_3}R_2+Z+R_1\}\delta^2+O(\delta^3)\}$$

$$=(1-\delta^2)^{n/2}\{1+\frac{(n+n_1)}{2}+\frac{(n-p_2)}{2}(R-1)\frac{S_{11.2}/(n-p_2)}{(S_{11}+W_1)/(n+n_1)}+O(\delta^3)\}$$

where U_1 is $W_{p_3}(n-p_2,I)$ and U_2 is $\chi^2_{n+n_1}$ with $R_2=S_{13.2}S_{33.2}^{-1}S_{31.2}/S_{11}$ and $R=(n-p_2)R_2/p_3(1-R_1)$. Q.E.D.

Acknowledgement: I am grateful to the referees for the valuable suggestions concerning the intuitive interpretation of the statistic φ_3.

REFERENCES

Eaton, M.L. and Kariya, T. (1983). Multivariate test with incomplete data. *Ann. Statist.*, **11**, 653–665.

Giri, N. (1977). *Multivariate Statistical Inference.* Academic Press, New York.

Giri, N. (1979). Locally minimax test for multiple correlations. *Canadian Journal of Statistics*, **7**, 53–60.

James, A.T. (1954). Normal multivariate analysis and the orthogonal group. *Ann. Math. Statist.*, **25**, 40–75.

Wijsman, R.A. (1966). Cross-sections of orbits and their application to densities of maximal invariants. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.* Vol. 1, University of California Press, 389–400.