A LIKELIHOOD RATIO TEST FOR MTP_2 WITHIN BINARY VARIABLES

F. Bartolucci

Dipartimento di Scienze Statistiche
Università di Perugia, Italy

Definition

- Let

$$\mathbf{X} = (X_1, \ldots, X_J)' \in \mathcal{X}$$

be a random vector of binary variables and $p(\mathbf{X})$ denote its joint probability distribution.

- If, for any pair of vectors $\mathbf{x}_1$ and $\mathbf{x}_2 \in \mathcal{X}$,

$$p[\min(\mathbf{x}_1, \mathbf{x}_2)]p[\max(\mathbf{x}_1, \mathbf{x}_2)] \geq p(\mathbf{x}_1)p(\mathbf{x}_2)$$

then the random vector $\mathbf{X}$ is *Multivariate Totally Positive* (MTP_2).

- MTP_2 defines a stochastic ordering amounting to a strong form of positive dependence since implies *Association* (A) and *Strongly Positive Orthant Dependence* (SPOD) (e. g. Holland and Rosenbaum, 1986).

- $J = 2$,

	$X_2 = 0$	$X_2 = 1$
$X_1 = 0$	$p(00)$	$p(01)$
$X_1 = 1$	$p(10)$	$p(11)$

$$\rho(X_1, X_2) = \log \frac{p(00)p(11)}{p(01)p(10)} \geq 0$$

- $J = 3$,

	$X_3 = 0$		$X_3 = 1$	
	$X_2 = 0$	$X_2 = 1$	$X_2 = 0$	$X_2 = 1$
$X_1 = 0$	$p(000)$	$p(010)$	$p(001)$	$p(011)$
$X_1 = 1$	$p(100)$	$p(110)$	$p(101)$	$p(111)$

$$\rho(X_1, X_2 | X_3 = 0) \geq 0 \qquad \rho(X_1, X_2 | X_3 = 1) \geq 0$$

$$\rho(X_1, X_3 | X_2 = 0) \geq 0 \qquad \rho(X_1, X_3 | X_2 = 1) \geq 0$$

$$\rho(X_2, X_3 | X_1 = 0) \geq 0 \qquad \rho(X_2, X_3 | X_1 = 1) \geq 0$$

- Let $\mathcal{P} = \{j_1, j_2\}$ be any pair of indices belonging to $\mathcal{J} = \{1, 2, \ldots, J\}$ and $\mathcal{U} \subseteq \bar{\mathcal{P}}$.

- $\rho(\mathcal{P}, \mathcal{U})$ be the log-odds ratio in the 2×2 subtable corresponding to the conditional distribution of (X_{j_1}, X_{j_2}) given $X_j = 1, \forall j \in \mathcal{U}$ and $X_j = 0, \forall j \notin \mathcal{P} \cup \mathcal{U}$.

- $J = 3$, $\mathcal{P} = \{1, 2\}, \mathcal{U} = \emptyset$,

$$\rho(\mathcal{P}, \mathcal{U}) = \rho(X_1, X_2 | X_3 = 0).$$

- $J = 4$, $\mathcal{P} = \{2, 3\}, \mathcal{U} = \{1\}$,

$$\rho(\mathcal{P}, \mathcal{U}) = \rho(X_2, X_3 | X_1 = 1, X_4 = 0).$$

• In general (for any J), MTP_2 holds if and only if (Karlin and Rinott, 1980) for any $\mathcal{P} \subset \mathcal{J}$ and any $\mathcal{U} \subseteq \bar{\mathcal{P}}$

$$\rho(\mathcal{P}, \mathcal{U}) \geq 0.$$

• In the whole we have a condition that concerns

$$\binom{J}{2} 2^{J-2}$$

log-odds ratios.

• We propose a procedure to test if MTP_2 holds for a certain data set.

• This condition is relevant in many fields: *statistical mechanics*, *computer storage*, *Item Response Theory* (IRT) models. IRT models are latent variable models used for the analysis of the results of a test assigned to a group of subjects. X_j is the response of an examinee to the jth item of the test:

– $X_j = 1$ the response is correct

– $X_j = 0$ otherwise

• These models are usually based on the non-parametric assumptions of *Local independence* (LI) and *Unidimensionality* (U) and *Monotonicity* (M).

• LI, U and M imply MTP_2: violation of MTP_2 for a data set implies that no IRT model for the data set may exist.

- The procedure to test MTP_2 is based on the likelihood ratio statistic between a saturated log-linear model and the same model whose parameters are constrained to take into account MTP_2.

- The saturated model log-linear model is defined as

$$\log(\mathbf{p}) = \mathbf{Z}\boldsymbol{\beta}$$

where

– $\mathbf{p}$ is the vector of all the joint probabilities apart from the first that is redundant;

– $\mathbf{Z}$ is an invertible matrix obtained by deleting the first row and the first column from

$$\underbrace{\mathbf{E} \otimes \cdots \otimes \mathbf{E}}_{J \text{ times}}, \text{ with } \mathbf{E} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

– $\boldsymbol{\beta}$ is the parameter vector.

• The vector of all the conditional log-odds ratios ($\rho(\mathcal{P}, \mathcal{U})$) may be obtained as

$$\boldsymbol{\rho} = \mathbf{R}\boldsymbol{\beta}$$

where $\boldsymbol{R}$ is an appropriate matrix that may be obtained trough a series of Kronecker products between matrices $\mathbf{E}$ and vectors $\mathbf{e}^{'} = (\,0 \quad 1\,)$.

• Since MTP$_2$ holds if and only if $\boldsymbol{\rho} \geq \mathbf{0}$, MTP$_2$ is also equivalent to the requirement that the $\boldsymbol{\beta}$ belongs to the convex cone

$$\mathcal{C} = \{\boldsymbol{\beta} : \mathbf{R}\boldsymbol{\beta} \geq 0\}.$$

Maximum Likelihood Estimation

• Under the multinomial sampling scheme, let $\boldsymbol{y}$ denote the vector of the observed frequencies apart from the first $(y(\mathbf{0}))$ which is redundant.

• To estimate $\boldsymbol{\beta}$ under MTP_2 we have to maximize the log-likelihood

$$L(\boldsymbol{\beta};\mathbf{y}) = \mathbf{y}'\log(\mathbf{p}) + (n - \mathbf{1}'\boldsymbol{y})\log[1 - \mathbf{1}'\log(\mathbf{p})] + \text{constant}$$

under the constraint $\boldsymbol{\beta} \in \mathcal{C}$.

- To maximize $L(\boldsymbol{\beta}; \mathbf{y})$ under $\boldsymbol{\beta} \in \mathcal{C}$, an iterative algorithm based on reweighted least squared is proposed.

- At step any step the operation

$$\max_{\boldsymbol{\beta} \in \mathcal{C}} Q(\boldsymbol{\beta}, \boldsymbol{\beta}_0),$$

where $\boldsymbol{\beta}_0$ is the estimate at previous step, is performed. $Q(\boldsymbol{\beta}, \boldsymbol{\beta}_0)$ is the second order Taylor expansion of L in $\boldsymbol{\beta}_0$.

- The starting value is given by the unrestricted estimate

$$\mathbf{Z}^{-1} \log[\mathbf{y}/y(\mathbf{0})].$$

- This algorithm converges to the maximum of the L under $\boldsymbol{\beta} \in \mathcal{C}$ (concavity of L).

Hypothesis testing

- H_0 hypothesis of independence ($\mathbf{R}\boldsymbol{\beta} = \mathbf{0}$).
- H_P hypothesis that MTP_2 holds ($\boldsymbol{\beta} \in \mathcal{C}$).
- H_U hypothesis that $\boldsymbol{\beta}$ is unrestricted.

- Let $L_h(\mathbf{y})$ be the maximum likelihood obtained under the hypothesis h ($h = 0, P, U$).

- To test H_0 versus H_P/H_0 and H_P versus H_U/H_P a decomposition of the G^2 statistic for testing independence is used:

$$G^2 = 2[L_U(\mathbf{y}) - L_0(\mathbf{y})] = T_{PU} + T_{0P}$$

where:

– $T_{0P} = 2[L_P(\mathbf{y}) - L_0(\mathbf{y})]$ is a measure of the discrepancy against H_0 in the direction of H_P;

– $T_{PU} = 2[L_U(\mathbf{y}) - L_P(\mathbf{y})]$ is a measure of the discrepancy against H_P in the direction of H_U.

Asymptotic distribution of T_{0P} and T_{PU}

- Under H_0, when n increases while J remains constant T_{PU} converges in distribution to

$$Q_{\bar{\mathcal{C}}} \;\sim\; \bar{\chi}^2(\bar{\mathcal{C}}, \mathbf{\Sigma})$$

and T_{0P} converges in distribution to

$$Q_{\mathcal{C}} \;\sim\; \bar{\chi}^2(\mathcal{C}, \mathbf{\Sigma})$$

where $\mathbf{\Sigma}$ is the asymptotic variance of $\hat{\boldsymbol{\beta}}$ (unconstrained estimator of $\boldsymbol{\beta}$).

- In general for a cone $\mathcal{S}$ and a covariance matrix $\mathbf{V}$, $\bar{\chi}^2(\mathcal{S}, \mathbf{V})$ is the distribution of

$$Q_{\mathcal{S}} = \hat{\mathbf{v}}'\mathbf{V}^{-1}\hat{\mathbf{v}}$$

where $\hat{\mathbf{v}}$ is the orthogonal projection of $\mathbf{v} \sim N(\mathbf{0}, \boldsymbol{V})$ in the $\mathbf{V}^{-1}$ metric. $\hat{\mathbf{v}}$ solves the problem

$$\min_{\hat{\mathbf{v}} \in \mathcal{S}} (\mathbf{v} - \hat{\mathbf{v}})'\mathbf{V}^{-1}(\mathbf{v} - \hat{\mathbf{v}}).$$

• $\bar{\chi}^2(\mathcal{S}, \mathbf{V})$ is a mixture of χ^2 distributions with appropriate weights which depend on $\mathcal{S}$ and $\mathbf{V}$.

• In practice, once compute the value of T_{PU} to test for MTP_2 it is possible to compute a *local* p-value as

$$\lim_{n \to \infty} P(T_{PU} > t_{PU}) = \sum_0^t w_j(\bar{\mathcal{C}}, \hat{\mathbf{\Sigma}}_0) Pr(\chi^2_j > t_{PU})$$

where:

– $t = s^J - J - 1$;

– $\hat{\mathbf{\Sigma}}_0$ is the estimate of $\mathbf{\Sigma}$ under H_0;

– weights are estimated, with the required precision, by a Monte Carlo Simulation.

• This p-value depends on the local estimate of $\mathbf{\Sigma}$.

• It has been proven that for any $c > 0$,

$$P(\bar{\chi}^2(\mathcal{O}_t) \geq c) \leq \lim_{n\to\infty} Pr(T_{PU} \geq c)$$
$$\leq P(\chi^2_{t-u} + \bar{\chi}^2(\mathcal{O}_u) \geq c),$$

where

– $u = J(J-1)/2$;

– $\mathcal{O}_t$ and $\mathcal{O}_u$ are the positive orthants in $\mathcal{R}^t$ and $\mathcal{R}^u$, respectively;

– the covariance matrix in the $\bar{\chi}^2$ distributions is the identity matrix.

• It is possible to compute an interval for the p-value which does not depend on the local estimate.

• The weights of the extreme distribution may be computed without simulation since correspond to probabilities of appropriate binomial distributions.

An Application

- We analyzed a data set concerning the responses of $n = 150$ students to a test made-up of $J = 4$ items used within an assessment for a basic course in Statistics at Perugia University.

$$\mathbf{y}' = (\,0 \quad 1 \quad 1 \quad 4 \quad 24 \quad 0 \quad 3 \quad 0 \quad 0 \quad 4 \quad 10 \quad 0 \quad 3 \quad 10 \quad 90\,)\,.$$

- The value of T_{PU} equals 12.0603: the p-value is bounded between 0.0599 and 0.1564 with a local estimate equal to 0.1114.

- MTP_2 cannot be rejected and we cannot state that IRT models are not adequate to analyze these data.

REFERENCES

DARDANONI, V. AND FORCINA, A. (1998). A unified approach to likelihood inference on stochastic orderings in a nonparametric context. *J. Amer. Statist. Assoc.* **93**: 1112-1123.

HOLLAND, P. W. AND ROSENBAUM, P. R. (1986). Conditional association and unidimensionality in monotone latent variable models. *Ann. Statist.* **14**: 1523-1543.

JUNKER, B. W. AND ELLIS, J. L. (1997). A characterization of monotone unidimensional latent variable models. *Ann. Statist.* **25**: 1327-1343.

KARLIN, S. AND RINOTT, Y. (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Mult. An.* **10**: 467-498.

LAWSON, L. L. AND HANSON, R. J. (1995). *Solving Least Squares Problems.* Philadelphia: SIAM ed.

LINDSAY, B., CLOGG, C. AND GREGO, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis, *J. Amer. Statist. Assoc.* **86**: 86-107.

ROSENBAUM, P. R. (1984). Testing the conditional independence and monotonicity assumption of item response theory. *Psychometrika* **49**: 425-435.

SHAPIRO, A. (1988). Towards a Unified Theory of Inequality Constrained Testing in Multivariate Analysis. *International Statistical Review* **56**: 49-62.

WOLAK, F. A. (1989). Testing inequality constraints in linear econometric models. *J. Economtrics* **41**: 205-235.