Friedman's Test with Missing Observations

Edyta Mrwka

Seminarium Pracowni Metod Przetwarzania Informacji Niepewnej

September 26, 2005

Outline

- Friedman's test -classical approach
- IF-sets – basic definitions
- IF-sets in modelling rankings
- Generalization of Friedman's test
- Conclusion

Friedman's test -classical approach

Let:

- $X = \{x_1, \ldots, x_n\}$ denote a finite universe of discourse.
- Elements (objects) $x_1, \ldots, x_n$ are ordered according to preferences of k observers $A_1, \ldots, A_k$.
- R_{ij}, $i = 1, 2, \ldots, k$, $j = 1, 2, \ldots, n$, denote the ranked observations so that R_{ij} is the rank given by ith observer to the jth object.
- $R = (R_1, ..., R_n)$ denote observed column totals, e.g.:

$$R_j = \sum_{i=1}^{k} R_{ij}, \qquad j = 1, ..., n.$$

- $\bar{R}$ denote the average column total.

Our data could be presented in the form of a two-way layout (or matrix) M with k rows and n columns:

$$M = \begin{bmatrix} R_{11} & R_{12} & \dots & R_{1n} \\ R_{21} & R_{22} & \dots & R_{2n} \\ \dots & \dots & \dots & \dots \\ R_{k1} & R_{k2} & \dots & R_{kn} \end{bmatrix}. \qquad (1)$$

Comment:

One may easily see that each row in (1) is a permutation of numbers $1, 2, \dots, n$. If, e.g., x_j has the same preference relative to all other objects

$$x_1, \dots, x_{j-1}, x_{j+1}, \dots x_n$$

in the opinion of each of the k observers, then all ranks in the jth column will be identical.

Hypothesis Testing Problem:

Suppose we are interested in testing the null hypothesis that the k variates are independent or - in other words - that there is no association between rankings given by k observers.

Test Statistic:

The sum of squares of deviations between actually observed column total and average column total for perfect agreement is given by:

$$S(R) = \sum_{j=1}^{n} [R_j - \bar{R}]^2 = \sum_{j=1}^{n} \left[R_j - \frac{k(n+1)}{2} \right]^2 \qquad (2)$$

maybe used to test the null hypothesis H that the rankings are independent.

Critical Region:

A linear function of statistic (2) defined as

$$Q = \frac{12S}{kn\left(n+1\right)} \tag{3}$$

can be used to define the rejection region for our hypothesis testing problem. It was proved that statistic (3) approaches the chi-square distribution with $n-1$ degrees of freedom as k increases. Therefore we reject the null hypothesis H if

$$Q \geq \chi^2_{n-1,\alpha}.$$

A test based on Q is called *Friedman's test*.

Problems with ranking specyfication:

The Friedman's test could be used provided all elements are univocally classified by all observers. However, it may happen that one (or more) observer cannot rank all the elements under study. In other situations someone may have problems with specifying his or her preferences.

Solution:

- Remove all objects which are not ordered by all of the observers and not to include them into considerations (this approach involves always a loss of information.
- Generalize the classical Friedman's test to make it possible to infer about possible association between rankings with missing information or non-comparable outputs.

IF-sets – basic definitions

Let X denote a universe of discourse. Then a fuzzy set C in X is defined as a set of ordered pairs

$$C = \{\langle x, \mu_C(x)\rangle : x \in X\},$$

where $\mu_C : X \to [0, 1]$ is the membership function of C and $\mu_C(x)$ is the grade of belongingness of x into C. Thus automatically the grade of nonbelongingness of x into C is equal to $1 - \mu_C(x)$.

Comment:

In real life the linguistic negation not always identifies with logical negation. This situation is very common in natural language processing, computing with words, etc. Therefore Atanassov suggested a generalization of classical fuzzy set, called an intuitionistic fuzzy set.

An IF-set C in X is given by a set of ordered triples

$$C = \{\langle x, \mu_C(x), \nu_C(x)\rangle : x \in X\},$$

where $\mu_C, \nu_C : X \to [0,1]$ are functions such that

$$0 \le \mu_C(x) + \nu_C(x) \le 1 \qquad \forall x \in X.$$

For each element $x \in X$ we can compute, so called, the IF-index of x in C defined as follows

$$\pi_C(x) = 1 - \mu_C(x) - \nu_C(x),$$

which quantifies the amount of indeterminacy associated with x_i in C.
A distance between two IFS of the universe of discourse $X = \{x_1, \ldots, x_n\}$ it is a function $d : IFS(X) \times IFS(X) \to R^+ \cup \{0\}$ defined as follows:

$$d(B,C) = \sum_{j=1}^{n} \left[(\mu_B(x_j) - \mu_C(x_j))^2 + (\nu_B(x_j) - \nu_C(x_j))^2 \right].$$

IF-sets in modelling rankings

Let

$$A_i = \{\langle x_j, \mu_{A_i}(x_j), \nu_{A_i}(x_j)\rangle : x_j \in X\}$$

denote an intuitionistic fuzzy subset of the universe of discourse $X = \{x_1, \ldots, x_n\}$, where membership function $\mu_{A_i}(x_j)$ indicates the degree to which x_j is the most preferred element by ith observer, while nonmembership function $\nu_{A_i}(x_j)$ shows the degree to which x_j is the less preferred element by ith observer.

Queastion:

How to determine these membership and nonmembership functions?

For each observer one can always specify two functions $w_{A_i}, b_{A_i} : X \to \{0, 1, \ldots, n-1\}$ defined as follows:

- Let $w_{A_i}(x_j)$ denote the number of elements $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$ surely worse than x_j.
- Let $b_{A_i}(x_j)$ be equal to the number of elements $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$ surely better than x_j in the ordering corresponding to the preferences expressed by observer A_i.

Using functions $w_{A_i}(x_j)$ and $b_{A_i}(x_j)$ we may determine the requested membership and nonmembership functions as follows

$$\mu_{A_i}(x_j) = \frac{w_{A_i}(x_j)}{n-1}, \quad \nu_{A_i}(x_j) = \frac{b_{A_i}(x_j)}{n-1}.$$

In such a way we get k well defined IF-sets which describe nicely orderings corresponding to k observers. It is seen that:

- $\pi_{A_i}(x_j) = 0$ for each $x_j \in X$ if and only if all elements are ranked by ith observer and there are no ties.
- If there exist such element $x_j \in X$ that $\pi_{A_i}(x_j) > 0$ then it means that there are ties or non-comparable elements in the ordering made by ith observer.
- $\pi_{A_i}(x_j) = 1$ if and only if element $x_j \in X$ is non-comparable with other element or all elements $x_1, \ldots, x_n$ have obtained the same rank in the ordering made by ith observer.

Generalization of Friedman's test

According to (2) the test statistic for testing independence might be expressed in a following way

$$S(R) = d(R, \overline{R^*}),$$

where $d(R, \overline{R^*})$ denotes a distance between the observed column totals $R = (R_1, \ldots, R_n)$ and the average column totals $\overline{R^*}$ obtained for perfect agreement between rankings. It can be shown that $\overline{R^*} = (\overline{R_1^*}, \ldots, \overline{R_n^*})$ and $\overline{R_j^*} = \frac{k(n+1)}{2}$ for each $j = 1, \ldots, n$.

We will consider appropriate IF-sets $A_1, \ldots, A_k$ for modelling ill-defined rankings. Thus instead of R will also consider an IF-set A, defined as follows

$$A = \{\langle x_j, \mu_A(x_j), \nu_A(x_j)\rangle : x_j \in X\},$$

where the membership and nonmembership functions μ_A and ν_A, respectively, are given by

$$\mu_A(x_j) = \frac{1}{k}\sum_{i=1}^{k} \mu_{A_i}(x_j),$$

$$\nu_A(x_j) = \frac{1}{k}\sum_{i=1}^{k} \nu_{A_i}(x_j).$$

If there is a perfect agreement within the group of observers and all objects are ranked without ties, then the resulting IF-set is of a form

$$A^* = \{\langle x_j, \mu_{A^*}(x_j), \nu_{A^*}(x_j)\rangle : x_j \in X\}$$

such that the membership function is given by

$$\mu_{A^*}(x_{j_1}) = 1, \quad \mu_{A^*}(x_{j_2}) = \frac{n-2}{n-1}, \quad \mu_{A^*}(x_{j_3}) = \frac{n-3}{n-1},$$
$$\ldots \quad \mu_{A^*}(x_{j_{n-1}}) = \frac{1}{n-1}, \quad \mu_{A^*}(x_{j_n}) = 0.$$

where $x_{j_1}, \ldots, x_{j_n}$ is a permutation of elements $x_1, \ldots, x_n$ and the nonmembership function is

$$\nu_{A^*}(x_j) = 1 - \mu_{A^*}(x_j)$$

for each $j = 1, \ldots, n$.

Therefore, for perfect agreement between rankings, instead of the average column totals $\overline{R^*}$ we obtain an IF-set $\overline{A^*} = \{\langle x_j, \mu_{\overline{A^*}}(x_j), \nu_{\overline{A^*}}(x_j)\rangle : x_j \in X\}$ such that

$$\begin{aligned} \mu_{\overline{A^*}}(x_1) &= \ldots = \mu_{\overline{A^*}}(x_n) = \frac{1}{2}, \\ \nu_{\overline{A^*}}(x_1) &= \ldots = \nu_{\overline{A^*}}(x_n) = \frac{1}{2}. \end{aligned}$$

For actual observed rankings, modelled by IF-sets $A_1, \ldots, A_k$, test statistic is a distance between IF-set A and $\overline{A^*}$:

$$S_{IF}(A) = d(A, \overline{A^*}) = \sum_{j=1}^{n} \left[\left(\mu_A(x_j) - \frac{1}{2} \right)^2 + \left(\nu_A(x_j) - \frac{1}{2} \right)^2 \right].$$

Lemma

For any $A \in IFS$ there exist two intuitionistic fuzzy sets A^+ and A^- such that:

$$S_{IF}\left(A^-\right) \leq S_{IF}\left(A\right) \leq S_{IF}\left(A^+\right)$$

where

$$\begin{aligned} A^- &= \{\langle x_j, \mu_A(x_j), \nu_A(x_j) + \pi_A(x_j)\rangle : x_j \in X\}, \\ A^+ &= \{\langle x_j, \mu_A(x_j) + \pi_A(x_j), \nu_A(x_j)\rangle : x_j \in X\}. \end{aligned}$$

Hence our test statistic $S_{IF}(A)$ based on ill-defined data is bounded by two other statistics $S_{IF}^{-}(A) = S_{IF}(A^{-})$ and $S_{IF}^{+}(A) = S_{IF}(A^{+})$ corresponding to situations with perfect rankings. Indeed, for each $x_j \in X$

$$\begin{aligned} \mu_{A^-}(x_j) &= 1 - \nu_{A^-}(x_j) \Rightarrow \pi_{A^-}(x_j) = 0 \\ \mu_{A^+}(x_j) &= 1 - \nu_{A^+}(x_j) \Rightarrow \pi_{A^+}(x_j) = 0 \end{aligned}$$

which means that A^{-} and A^{+} describe situations when all elements are univocally classified.

Therefore, there exist two systems of rankings (in a classical sense) R^- and R^+ and one-to-one mapping transforming A^- and A^+ onto R^- and R^+. Next, it could be shown that

$$
\begin{aligned}
S_{IF}^{-}(A) &= \frac{2}{k^2 (n-1)^2} S\left(R^-\right) \\
S_{IF}^{+}(A) &= \frac{2}{k^2 (n-1)^2} S\left(R^+\right).
\end{aligned}
$$

Since $S_{IF}^{-}(A)$ and $S_{IF}^{+}(A)$ are linear functions of S, we can easily find their distributions necessary for the desired test.

If k is large enough then

$$\begin{aligned} T_1 &= \frac{6k}{n(n-1)(n^2-1)} S_{IF}^{-} \\ T_2 &= \frac{6k}{n(n-1)(n^2-1)} S_{IF}^{+}. \end{aligned}$$

are chi-square distributed with $k-1$ degrees of freedom.

Comment:

In hypothesis testing we reject the null hypothesis H if test statistic belongs to critical region or accept H otherwise. In our problem with missing data we get two test statistics T_1 and T_2. We have to utilize both statistics for decision making.

The hypothesis H should be rejected on the significance level α if

$$T_1(A) \geq \chi^2_{n-1,\alpha}$$

while there are no reasons for rejecting H (i.e. we accept H) if

$$T_2(A) < \chi^2_{n-1,\alpha}.$$

These two situations are quite obvious. However, it may happen that

$$T_1(A) < \chi^2_{n-1,\alpha} \leq T_2(A).$$

In such a case we are not completely convinced neither to reject nor to accept H.

Instead of a binary decision we could indicate a degree of conviction that one should accept or reject H.

$$Ness(\text{reject } H) = \begin{cases} 1 & \text{if} \quad T_1(A) \geq \chi^2_{n-1,\alpha} \\ \frac{T_2(A)-\chi^2_{n-1,\alpha}}{T_2(A)-T_1(A)} & \text{if} \quad T_1(A) < \chi^2_{n-1,\alpha} \leq T_2(A) \\ 0 & \text{if} \quad T_2(A) < \chi^2_{n-1,\alpha} \end{cases}$$

describing the degree of necessity for rejecting H. Simultaneously we get another measure

$$Poss(\text{accept } H) = 1 - Ness(\text{reject } H)$$

describing the degree of possibility for accepting H.

Conclusion

- We have proposed how to generalize the well-known Friedman's test to situations in which not all elements could be ordered.
- We have discussed Friedman's test as a nonparametric tool for testing independence of k variates.
- Our generalized version of Friedman's test also works for two-way analysis of variance by ranks with missing data.