A Test for Dominance

Atsu S.S. Dorvlo

Department of Mathematics and Statistics, College of Science, Sultan Qaboos University, P.O. Box: 36, PC 123, Al-Khod, Muscat, Sultanate of Oman. Email: atsu@squ.edu.om.

ABSTRACT: In some selection situations a simple majority is not enough to indicate that one item dominates the others. In this paper we develop a simple procedure for testing the dominance of one item with respect to others in a multi option situation. Expressions are given for computing the relevant p-values. The procedure is illustrated.

Keywords: Dominance; Modal class; Categorical variable; Hypothesis testing.

اختبار السيطرة

اتسو دورفلو

ملخص: إن الأغلبية البسيطة في بعض الحالات المختارة غير كافية لتدل على أن عنصرا واحداً هو العامل المسيطر من بين العوامل الأخرى. لقد طورنا في هذا البحث طريقة بسيطة لاختيار العامل المسيطر لعنصر واحد بالمقارنة مع عناصر أخرى عند وجود خيارات متعددة. لقد تم وضع الصيغة المناسبة لحسابات القيم p ووضحت الطريقة.

مفتاح الكلمات: العوامل المسيطرة، الحزم المشروطة ، المتغير الحاسم ، اختبار الفرضيات.

1. Introduction

Suppose that a categorical variable X can take on m possible values arbitrarily numbered 1 to m. Let $p_i = \text{P}(X \,\varepsilon\,$ category i), i $=,\ldots,$m. Are all these categories equally likely? Is there a statistically dominant category? These two questions can be put formally as (i) are all the items equally likely versus (vrs) does at least one item occur more often. This hypothesis is that the probability of selecting each item is the same against the probability of selecting at least one item which is different from the others. That is $H_\circ : p_i = \dfrac{1}{m} \,\forall\, i$, vrs. $H_A : p_i \neq \dfrac{1}{m}$ for at least one i. Secondly (ii) there is no dominant category vrs there is a dominant category. Both hypotheses can be stated more generally as (iii) $H_\circ : p_i = p_{\circ_i}$, $i = 1,\ldots,m$ where $p_{\circ_i}$ are specified probabilities vrs at least one of them is different from the specified $p_{\circ_i}$.

The rejection of the null hypothesis under (i) does not indicate which item is most preferred. Also rejecting this hypothesis does not necessarily mean that a particular item is dominant. However, accepting the null hypothesis under (i) will imply that the null hypothesis under (ii) is true. The reverse is not necessarily true. The hypothesis (i) or (iii) can be tested using either the Pearson chi-square statistic or the likelihood ratio statistic [1]. However the statement of dominance does not necessarily imply any particular probabilities under the null hypothesis.

Statistical dominance of G over F implies that $G(x) \leq F(x) \forall x \in \mathbb{R}$ with strict inequality for some x. There are numerous Kolmogorov-Smirnov type tests for the dominance of G over F with many applications in financial mathematics [2-4]. Tse and Zhang [5] compared several tests for stochastic dominance using Monte Carlo simulation. This formulation of dominance implies a form of simple majority.

Instances where it is desired to know which item is most preferred occur frequently. In an opinion poll of say six politicians, which politician is the dominant one? In a restaurant that serves four types of deserts, which desert is the most patronized? When a group of seniors from a high school are asked to select a university from a list, which university will be the most selected? In an 'eye-movement thought process experiment', is there a dominant pattern for a particular thought process[6,7]? These examples and others like them often occur in practice. Frequently a simple majority is enough to declare one option the favorite. The notion of dominance, however, involves a larger majority than a simple majority of the item where the simple majority item is not necessarily the dominant item. In this paper we propose a test that will indicate whether the modal item is dominant when there are m items/choices and the (selectors)

sample size is n. To establish if a particular item is dominant we propose computing the probabilities associated with selecting an item i by j selectors out of n.

2. Distribution of an item X_i

Suppose X_i is the number of subjects who have selected the ith item. This implies that $\sum_{i=1}^{m} X_i = n$. The sample space for any X_i is {0, 1, 2,..., n}. The null hypothesis of there being no dominant item implies that no particular item is favoured, hence without loss of generality we shall discuss the distribution of the number of subjects, X_1, selecting item 1. The number of combinations such that $X_1 = j$ and the sum of the rest of the X_is is equal to (n-j) is given by the binomial coefficient $\binom{m-2+n-j}{m-2}$, for m, the given number of items to be selected from, and n, the number of subjects doing the selection. Then

$$P_m^n(X_1 = j) = \frac{1}{K}\binom{m-2+n-j}{m-2}, \quad j = 0,1,2,.....,n \tag{1}$$

where $K = \sum_{j=0}^{n}\binom{m-2+n-j}{m-2} = \binom{m+n-1}{m-1}$.

After some simplification we have

$$P_m^n(X_1 = j) = \begin{cases} \dfrac{m-1}{m+n-1}, & for\ j = 0 \\ \dfrac{m-1}{m+n-1}\prod_{t=1}^{j}\dfrac{n+1-t}{m+n-1-t}, & for\ j = 1,2,...,n. \end{cases} \tag{2}$$

Table 1 gives an example of the frequency of X_1, for various values of m from 3 to 10 and n = 10. Table 2 gives the densities for X_1. It is easy to prove that for fixed $m \geq 2$, $P_m^n(X_1 = i) < P_m^n(X_1 = j)$ for all i>j. Also for a fixed i, $P_{m+k}^n(X_1 = i) > P_m^n(X_1 = i)$ for positive integer k. These are illustrated in Tables 1 and 2.

Table 1. Frequency of X_1 for given m and n=10.

X_1	3	4	5	6	7	8	9	10
0	11	66	286	1001	3003	8008	19448	43758
1	10	55	220	715	2002	5005	11440	24310
2	9	45	165	495	1287	3003	6435	12870
3	8	36	120	330	792	1716	3432	6435
4	7	28	84	210	462	924	1716	3003
5	6	21	56	126	252	462	792	1287
6	5	15	35	70	126	210	330	495
7	4	10	20	35	56	84	120	165
8	3	6	10	15	21	28	36	45
9	2	3	4	5	6	7	8	9
10	1	1	1	1	1	1	1	1

Table 2. Distribution of X_i for a given m and n = 10.

X_i	3	4	5	6	7	8	9	10
0	0.1667	0.2308	0.2857	0.3333	0.3750	0.4118	0.4444	0.4737
1	0.1515	0.1923	0.2198	0.2381	0.2500	0.2574	0.2614	0.2632
2	0.1364	0.1573	0.1648	0.1648	0.1607	0.1544	0.1471	0.1393
3	0.1212	0.1259	0.1199	0.1099	0.0989	0.0882	0.0784	0.0697
4	0.1061	0.0979	0.0839	0.0699	0.0577	0.0475	0.0392	0.0325
5	0.0909	0.0734	0.0559	0.0420	0.0315	0.0238	0.0181	0.0139
6	0.0758	0.0524	0.0350	0.0233	0.0157	0.0108	0.0075	0.0054
7	0.0606	0.0350	0.0200	0.0117	0.0070	0.0043	0.0027	0.0018
8	0.0455	0.0210	0.0100	0.0050	0.0026	0.0014	0.0008	0.0005
9	0.0303	0.0105	0.0040	0.0017	0.0007	0.0004	0.0002	0.0001
10	0.0152	0.0035	0.0010	0.0003	0.0001	0.0001	0.0000	0.0000

3. The joint distribution of two items X_1 and X_2

Without loss of generality we consider the joint distribution of $X_1 = i$ and $X_2 = j$. The sample space for the (X_1, X_2) combination is $\big((0,0),(0,1),...,(0,n),(1,0),...,(1,n-1),...,(n,0)\big)$. Let K be the total combinations needed to fill the m items/choices. Hence if $X_1 = i$ and $X_2 = j$ such that $i + j \leq n$, for $\sum X_i = n$ then the number of combinations for this set is $\binom{m+n-3-i-j}{m-3}$. Therefore the joint distribution of $X_1 = i$ and $X_2 = j$ is

$$P_m^n\left(X_1 = i, X_2 = j\right) = \frac{1}{K}\binom{m+n-3-i-j}{m-3}, \ i+j \leq n \tag{3}$$

where $K = \sum\limits_{0 \leq i+j \leq n} \binom{m+n-3-i-j}{m-3}$ and the summation is over all possible i and j values such that the sum over all the items is n. This sum reduces to

$$K = \sum\limits_{t=0}^{n}(t+1)\binom{m+n-3-t}{m-3} = \binom{m+n-1}{n} = \binom{m+n-1}{m-1}. \tag{4}$$

The probability in (3) can therefore be written in the computationally friendly form

$$P_m^n\left(X_1 = i, X_2 = j\right) = \begin{cases} \dfrac{(m-1)(m-2)}{(m+n-1)(m+n-2)}, & for \quad i=0, j=0 \\[3mm] \left(\dfrac{(m-1)(m-2)}{(m+n-1)(m+n-2)}\right)\prod\limits_{t=1}^{k}\dfrac{n+1-t}{m+n-2-t}, & for \quad 1 \leq i+j \leq n. \end{cases} \tag{5}$$

The joint distribution is dependent on only the sum of units selecting these two items. Therefore if the interest is in the number of subjects, $k \leq n$, selecting any two items then the distribution is

$$P_m^n\left(X_1 + X_2 = k\right) = \begin{cases} \dfrac{(m-1)(m-2)}{(m+n-1)(m+n-2)}, & for \quad k=0 \\[3mm] \dfrac{(k+1)(m-1)(m-2)}{(m+n-1)(m+n-2)}\prod\limits_{t=1}^{k}\dfrac{n+1-t}{m+n-2-t}, & for \quad k=1,2,...,n. \end{cases} \tag{6}$$

4. The conditional distribution of X_k given X_l

From the distribution of X_k and the joint distribution of X_k and X_l we can derive the conditional distribution of say, $X_k \mid X_l$. Without loss of generality, the joint distribution of $X_1 \mid X_2$ is given below.

$$P_m^n\left(X_1 = i \mid X_2 = j\right) = \begin{cases} \dfrac{(m-2)}{(m+n-2)}, & for \quad i=0, j=0 \\[3mm] \dfrac{(m-2)}{(m+n-2-j)}, & for \quad i=0, j \neq 0, j \leq n \\[3mm] \dfrac{(m-2)}{(m+n-2)}\prod\limits_{t=1}^{i}\dfrac{n+1-t}{m+n-2-t}, & for \quad i \neq 0, j=0, i \leq n \\[3mm] \dfrac{(m-2)}{(m+n-2-j)}\prod\limits_{t=j+1}^{i+j}\dfrac{n+1-t}{m+n-2-t}, & for \quad i \neq 0, j \neq 0, i+j \leq n \end{cases} \tag{7}$$

5. A Dominance Test

The acceptance of $H_\circ : p_i = \dfrac{1}{m} \forall i$ implies that there is no dominant item. However, the acceptance of the hypothesis of there being no dominant item does not necessarily imply $H_\circ : p_i = \dfrac{1}{m} \forall i$. Hence the first step in testing for dominance may be to perform the chi-square test for $H_\circ : p_i = \dfrac{1}{m} \forall$ i vrs $H_\circ : p_i \geq \dfrac{1}{m}$ for at least one i. There may therefore be a dominant choice if the null hypothesis is rejected.

In preference testing, one expects that the dominant item be the most selected. The item X_i is said to be the modal item if the number of subjects that selected it is more than the number of subjects that selected any other item. We say that item X_i is the modal item if $X_i > X_j$, for $i \neq j$ for all j. In considering hypothesis (ii), it is intuitively clear that X_i, the modal item, will be the dominant item the closer X_i is to n. We can therefore base a test on the statistic X_i the modal item. Reject that there is no dominant item if X_i is large. In fact, when the modal item is small, that means that the subjects are divided among the items equally except for random fluctuations. Therefore the modal item is the dominant item if X_i is greater than some $y_\circ$ for which $P(X_i > y_\circ) = \alpha$. From eq (2) the p-value can be computed from

$$P(X_i \geq y) = \frac{m-1}{m+n-1} \sum_{j=y}^{n} \prod_{t=1}^{j} \frac{n+1-t}{m+n-1-t},$$ where y is some observed modal value. When this p-value is less than some α, then that particular item can be said to dominate the others.

6. Discussion

The proposed test checks if the modal item is the dominant one. However, suppose the interest is in determining if a particular item is dominant. The above test can still be used in one of two ways.
1. Compute the p-value associated with the realization of this item. If it is less than α of choice, then the item is dominant.
2. Compute the p-value associated with the modal item. If this is less than α, then the modal item is dominant. Hence the item of choice cannot be said to be dominant, assuming there is only one dominant item.

7. An Illustration

Twenty-five council members were asked to decide which of four amendments (M1, M2, M3 and M4) to a bill will be most favorable to their constituents. For example, suppose 14 voted for M1, 6 for M2, and 3 and 2 for M3 and M4 respectively. A chi-square test showed that not all the amendments are equally appealing to the councilors (χ =14.2, p-value = 0.003). However no amendment is dominant (p-value = 0.111). For an amendment to be dominant, at $\alpha = 0.05$, it should have the vote of at least 18 out of the 25 council members. Table 3 gives the critical values for some m and n at $\alpha = 0.05$.

For m and n large, it is possible that this test will show that two or more items can be dominant. For example, suppose that m = 8 and n = 30, the critical value for dominance at $\alpha = 0.05$ is 12. It is therefore possible that two items can have this number or more. However, the probability of this happening will be very small. The joint probability of two items having large values is generally small (eq (5) and (6)). In this example the probability of two items getting 12 or more at the same time is only 0.0023. Also it can be shown that the conditional probability of a large item, X_i given that X_j is large, is small (eq (7)).

An interesting application of the above formulation is to be able to compute the p-value corresponding to a particular modal value and thereby know ones position of strength.

8. Acknowledgement

9. Appendix

The following identities were useful in proving the distributions above.

$$\sum_{t=0}^{a-b}(t+1)\binom{a-t}{b}=\binom{a+2}{a-b}$$

$$\sum_{t=0}^{a-b}\binom{a-t}{b}=\binom{a+1}{b+1}=\binom{a+1}{a-b}$$

$$\sum_{t=0}^{a-b}t\binom{a-t}{b}=\frac{a-b}{a+2}\binom{a+2}{a-b}$$

Table 3. Critical values for dominance at $\alpha=0.05$ for some n and m.

n/m	3	4	5	6	7	8	9	10
3	-	3	3	3	3	3	3	2
4	-	4	4	3	3	3	3	3
5	5	5	4	4	4	3	3	3
6	6	5	5	4	4	4	4	3
7	7	6	5	5	4	4	4	4
8	8	7	6	5	5	5	4	4
9	9	7	6	6	5	5	5	4
10	9	8	7	6	6	5	5	5
11	10	9	8	7	6	6	5	5
12	11	9	8	7	6	6	6	5
13	12	10	9	8	7	6	6	5
14	12	11	9	8	7	7	6	6
15	13	11	10	9	8	7	6	6
16	14	12	10	9	8	7	7	6
17	15	12	11	9	8	8	7	7
18	16	13	11	10	9	8	7	7
19	16	14	12	10	9	8	8	7
20	17	14	12	11	10	9	8	7
21	18	15	13	11	10	9	8	8
22	19	16	13	12	10	9	9	8
23	19	16	14	12	11	10	9	8
24	20	17	14	13	11	10	9	9
25	21	18	15	13	12	11	10	9
26	22	18	15	14	12	11	10	9
27	23	19	16	14	12	11	10	9
28	23	19	17	14	13	12	11	10
29	24	20	17	15	13	12	11	10
30	25	21	18	15	14	12	11	10
31	26	21	18	16	14	13	12	11
32	26	22	19	16	14	13	12	11
33	27	23	19	17	15	13	12	11
34	28	23	20	17	15	14	12	11
35	29	24	20	18	16	14	13	12
36	30	24	21	18	16	14	13	12
37	30	25	21	18	16	15	13	12
38	31	26	22	19	17	15	14	13
39	32	26	22	19	17	15	14	13
40	33	27	23	20	18	16	14	13

References

1. Snedecor, G.W. and Cochran W.G. Statistical Methods, Iowa State University Press, 1989.
2. Davidson, R. and Duclos, J. Statistical inference for stochastic dominance and the measurement of poverty and inequality. *Econome*., 2000, **68**, 1453-1464.
3. Larsen, G.A. and Resnick, B.G. Bootstrapping a distance test for stochastic dominance analysis. *Rev. Qual. Fin. Account*., 2006, **3(1)**, 61-69.
4. Merino, F. Firms' productivity and internationalization: a statistical dominance test. *Appl. Econ. Lett*., 2004,

11(13), 851-854.

5. Tse, Y. and Zhang, X. A Monte Carlo Investigation of Some Tests for Stochastic Dominance. Mimeo. Department of Econometrics and Business Statistics, Monash University, Victoria, Australia., 2003.

6. Burke, D.T., Meleger, A., Schneider, J.C., Snyder, J., Dorvlo, A.S.S. and Al-Adawi, S. Eye-movements and ongoing task processing. *Perceptual and Motor Skills*., 2003, **96(3 Pt. 2),** 1330-1338.

7. Zhou, Y.H., Gao, J.B., White, K.D., Merk, I and Yao, K. Perceptual dominance time distributions in multistable visual perception. *Biol. Cybernet*., 2004, **90(4**), 256-263.