

An on-line exact small-sample test of significance for extended matching tasks

RAPHAEL GILLETT

University of Leicester, Leicester, England

An extension of the traditional matching paradigm that enables researchers to test a variety of new experimental hypotheses is outlined. An on-line computer program that provides an exact small-sample test of hypotheses in the extended matching paradigm is described. The program, which has an intuitive graphical interface, may be accessed and executed via the Internet by using an ordinary browser.

An extension of the traditional matching paradigm that enables researchers to test a variety of new experimental hypotheses was introduced by Gillett (1985a, 1985b, 1985c). In the present article, an on-line computer program that provides an exact small-sample test of hypotheses in the extended matching paradigm is described. The program, which has an intuitive graphical interface, may be accessed and executed via the Internet by using an ordinary browser.

In the traditional matching task, two collections each containing n objects are paired in a one-to-one arrangement on two different occasions, yielding two separate pairing configurations. For example, the first pairing configuration might represent the standard, or correct, pairings between objects, and the second configuration might denote a subject's estimate of the correct pairings. Interest lies in the number of matches between the two pairing configurations.

Figure 1 illustrates the simple matching task. A subject was asked to pair a collection of photographs of human faces displaying a range of emotions with a collection of printed labels bearing the names of different emotional states. The subject's performance might be assessed by the number of matches between the known (correct) pairings and the subject's own pairings. The null probability distribution for the number of matches in the traditional matching task has been described by Feller (1968, p. 107).

The traditional matching test was extended to cover a broader class of research designs by Gillett (1985b). In the extended version, one of the one-to-one pairing configurations (either the standard or the one generated by the subject) is replaced by a set of postulated pairings between the objects in the two collections. The set of postulated pairings need not be one-to-one. The aim is to determine the frequency of occurrence of postulated pairings in the one-to-one pairing configuration. When a postulated pairing is observed to occur in the one-to-one pairing configuration,

the event is termed a match. The number of matches is used as a statistic to test the research hypothesis.

The extended matching paradigm is useful in catering for uncertainty and in giving credit for partial knowledge. It provides researchers with a more sensitive test of matching performance. For example, an uncertain subject who thinks that a particular photograph could be categorized as either of two different emotional states might be allowed to make both pairings. In this way, it is possible to give the subject credit for partial knowledge. A subject would not be able to artificially enhance performance by displaying exaggerated uncertainty, because an increase in the number of postulated pairings that is not based on knowledge would produce a corresponding increase in the number of matches required to reach a given level of significance.

In some research designs, the postulated pairings are known in advance and the one-to-one pairing configuration constitutes the empirical data; in other designs, the one-to-one pairing configuration is known in advance and the postulated pairings represent the empirical data.

A comparison of Figure 1 and Figure 2 illustrates the difference in structure between the traditional and extended matching paradigms by using the example of a subject attempting to pair facial photographs with emotional states. The same subject performs both the traditional matching task and the extended matching task. In the traditional matching task, each photograph can be paired with only one emotional state, and vice versa. In the extended matching paradigm, more than one photograph can be paired with an emotional state, and more than one emotional state can be paired with a photograph.

The subject's responses in Figures 1 and 2 are quite similar, as might be expected. The only difference between them is that, in Figure 2, the subject was able to express a degree of uncertainty experienced with photographs A and B.

In both Figures 1 and 2, the known (correct) pairing configuration occupies the negative diagonal that runs from top-left to bottom-right. The number of hits or matches j achieved by a subject is equal to the number of crosses in the negative diagonal. Our subject scored $j = 3$ hits on the traditional matching task in Figure 1 and $j = 4$ hits on the extended matching task in Figure 2.

This article was written during a period of study leave granted by the University of Leicester. Correspondence concerning this article should be addressed to R. Gillett, School of Psychology, University of Leicester, Leicester LE1 7RH, U. K. (e-mail: rtg@le.ac.uk).

| | | Emotional States | | | | | |
|-----------------------|---|------------------|---|---|---|---|---|
| | | a | b | c | d | e | f |
| Facial Photographs | A | | X | | | | |
| | B | | | | | X | |
| | C | | | X | | | |
| | D | | | | X | | |
| | E | X | | | | | |
| | F | | | | | | X |

Figure 1. Traditional matching paradigm in which a subject may pair each photograph with only one emotional state.

Note that it is only a matter of convention that objects in the two collections are arranged so that known or correct pairings occupy the negative diagonal. The computer program does not require that data should follow this convention. The analysis is unaffected by interchanging any two rows or any two columns. In many applications, such as rater agreement (Gillett, 1985c), the concept of a "correct" pairing configuration does not apply.

The method of analysis for the data in Figures 1 and 2 is illustrated in the next section, which describes the input to, and output from, the on-line computer program.

ON-LINE COMPUTER PROGRAM: XMATCH

The program XMATCH may be accessed via the Internet using a reasonably recent version of any browser, such as Netscape Communicator 4.7 or Internet Explorer 5.0, at <http://www.le.ac.uk/pc/rtg/xmatch.html>.

The program runs in real-time on the Internet, accepting data from, and returning results to, a number of users. Because combinatorial programs like the present one are

processor-intensive, there are restrictions on the number of simultaneous users and on the duration of a computation.

The algorithm is based on a technique of enumerating permutations called *rook methodology*, described by Kaplansky and Riordan (1946) and Riordan (1958, Chapter 7). In the game of chess, a rook can only take another piece if it lies in the same row or column as the other piece. A permutation may be visualized as an arrangement of n rooks on a $n \times n$ chessboard so that no rook can take any other rook. For example, the six crossed cells in Figure 1 may be viewed as six nontaking rooks on a 6×6 chessboard. Likewise, the correct pairing configuration in Figure 1 may be represented by six nontaking rooks that occupy the negative diagonal. Hence, the enumeration of permutations in the matching test is equivalent to the enumeration of arrangements of n nontaking rooks on an $n \times n$ chessboard.

The rook approach provides important statistical and combinatorial benefits. It supplies a framework for constructing exact, small-sample statistical tests in the matching paradigm and yields closed-form expressions for the expectation, variance, and other moments of the extended matching statistic (Gillett, 1985b). Furthermore, rook

| | | Emotional States | | | | | |
|-----------------------|---|------------------|---|---|---|---|---|
| | | a | b | c | d | e | f |
| Facial Photographs | A | | X | | | | X |
| | B | | X | | X | X | |
| | C | | | X | | | |
| | D | | | | X | | |
| | E | X | | | | | |
| | F | | | | | | X |

Figure 2. Extended matching paradigm in which a subject may pair each photograph with more than one emotional state.

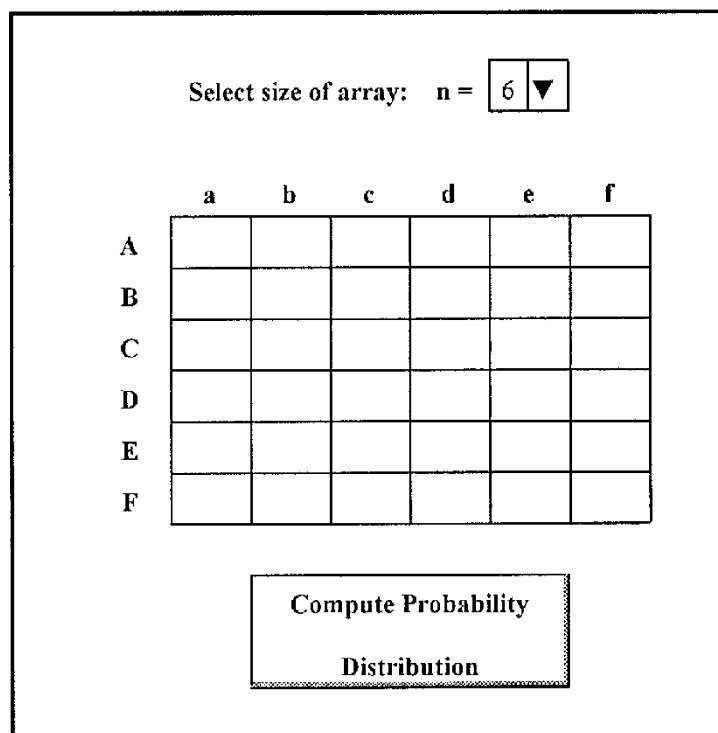


Figure 3. Data-input screen for on-line Internet computer program XMATCH.

methodology consumes considerably less computational time than would be required by full exhaustive enumeration of all permutations.

The languages employed to write the computer program and its graphical user interface include Java, JavaScript, XML, and HTML.

Data Input

The opening data-input screen displays a 9 × 9 grid array. The dimension of the array may be changed by selecting an appropriate size of *n* from the pop-down list above the array. A choice of array sizes between *n* = 4 and *n* = 25 is always available. Figure 3 provides an illustration of the data-input screen for an array of size *n* = 6.

The data that a user enters into the input array of XMATCH are the crossed cells, or postulated pairings. To enter a crossed cell into the input array, the user should click the mouse on the desired cell. The cell will immediately darken. Dark cells in the display represent crossed cells. To correct a mistake, the user can simply click on a cell a second time and it will revert to its original shade.

Once all the crossed cells have been entered, the probability distribution of the number of hits, or matches, expected under the null hypothesis can be calculated. The user should click on the large button at the bottom of the input screen labeled “Compute Probability Distribution,” and after a delay, which varies according to the amount of computation required, the probability distribution will be displayed.

The null hypothesis is that the subject pairs objects in the two collections in a random manner. The distribution of the number of matches under the null hypothesis is derived from the assumption that all permutations, or nontaking rook configurations, are equally likely. Thus, the null probability of exactly *j* matches is equal to the proportion of all nontaking rook configurations that occupy exactly *j* crossed cells and *n* − *j* noncrossed cells.

The expected number of matches under the null hypothesis is equal to the number of crossed cells divided by *n* (Gillett, 1985b). In the special case of the traditional matching task, where the number of crossed cells equals *n*, the expected number of matches is precisely one.

Analysis Output

The output of the analysis, the probability distribution of the number of hits, or matches, expected under the null hypothesis, is displayed in a separate window. Thus, if a

Table 1
Probability Distribution for Data in Figure 1

| <i>k</i> | <i>p</i> (exactly <i>k</i> hits) |
|----------|----------------------------------|
| 0 | .36806 |
| 1 | .36667 |
| 2 | .18750 |
| 3 | .05556 |
| 4 | .02083 |
| 5 | .00000 |
| 6 | .00139 |

Table 2
Probability Distribution for Data in Figure 2

| k | $p(\text{exactly } k \text{ hits})$ |
|-----|-------------------------------------|
| 0 | .18194 |
| 1 | .35833 |
| 2 | .29028 |
| 3 | .12500 |
| 4 | .03750 |
| 5 | .00556 |
| 6 | .00139 |

glance at the output were to indicate that a mistake had been made at the data-input stage, the user can simply switch windows to return to the data-input screen, correct the mistake, and rerun the analysis.

When the crossed cells of Figure 1 are entered into the data-input grid, the probability distribution that is obtained is displayed in Table 1. Likewise, the crossed cells in Figure 2 yield the probability distribution in Table 2.

Interpreting the Results

It will be recalled that our subject scored $j = 3$ hits on the traditional matching task in Figure 1 and $j = 4$ hits on the extended matching task in Figure 2. From Table 1, the probability of obtaining $j \geq 3$ hits in Figure 1 is equal to .07778. Therefore, at a significance level of $\alpha = .05$, the null hypothesis of the traditional matching test cannot be rejected. On the other hand, from Table 2, the probability of obtaining $j \geq 4$ hits in Figure 2 is equal to .04445. Hence, it can be concluded that the subject's performance in the extended matching task is significantly better than would be expected under the null hypothesis.

The difference between the two tasks resides in the response options that are available to a subject who is uncertain. The traditional matching task obliges the subject to guess when he/she is uncertain, whereas the extended matching task enables the subject to communicate the precise nature of their uncertainty so that partial knowledge may be taken into consideration. Although appropriate credit is awarded for partial awareness, a subject gains no advantage by overstating the degree of uncertainty actually experienced. By disregarding partial knowledge, the traditional matching task provides a less sensitive, more hit-and-miss test, especially when n is small. This helps to explain why the extended matching test reaches statistical significance, whereas the traditional test does not.

FURTHER TESTS WITHIN THE EXTENDED MATCHING PARADIGM

The extended matching paradigm constitutes a flexible analytic framework that researchers may utilize, adapt, and develop in order to build statistical tests for a wide variety of purposes.

The application of the extended matching paradigm to nominal-scale rater agreement is discussed by Gillett (1985c). An advantage of the approach is that it is able to cater for rater uncertainty. In the article, the test procedure is outlined, and examples that may be re-analyzed using XMATCH are provided.

Small-sample contingency table analysis is another area of application. Gillett (1985b) discussed the application of the extended matching paradigm to construct a test of a composite hypothesis in the area of gender discrimination in promotion practices. Suppose that a number of employees are being considered for promotion and that the posts available are of high, medium, and low responsibility. It is desired to test the composite hypothesis that posts of high responsibility will be assigned predominantly to men *and/or* posts of low responsibility will be allocated mostly to women. The employees are represented by the rows of the data-input grid and the posts by the columns. It is helpful to group the rows representing men together and likewise the rows for women. Similarly, columns representing posts of a given level of responsibility are grouped together. Crosses are placed in every cell linking a man with a high-responsibility post and also in every cell linking a woman with a low-responsibility post. Thus, the data-input grid should contain two rectangular areas of crossed cells. The composite hypothesis is tested by comparing the combined number of men actually assigned to high-responsibility posts and women actually assigned to low-responsibility posts with the probability of obtaining that number or greater, as supplied by the probability distribution yielded by the computer program.

In summary, XMATCH provides a useful and flexible analytic tool for researchers who wish to perform one of the tests described in this article or to construct custom-built, exact, small-sample tests in the extended matching paradigm.

REFERENCES

- FELLER, W. (1968). *An introduction to probability theory and its applications: Vol. 1* (3rd ed.). New York: Wiley.
- GILLETT, R. (1985a). Allowing for infeasible pairings in the matching paradigm. *Psychometrika*, **50**, 265-274.
- GILLETT, R. (1985b). The matching paradigm: An exact test procedure. *Psychological Bulletin*, **97**, 106-118.
- GILLETT, R. (1985c). Nominal scale response agreement and rater uncertainty. *British Journal of Mathematical & Statistical Psychology*, **38**, 58-66.
- KAPLANSKY, I., & RIORDAN J. (1946). The problem of the rooks and its applications. *Duke Mathematical Journal*, **13**, 259-268.
- RIORDAN, J. (1958). *An introduction to combinatorial analysis*. New York: Wiley.