

Statistical Models for the Assessment of Verbal and Other ESP Responses

W. G. ROLL AND D. S. BURDICK¹

ABSTRACT: In the development of methods for assessing verbal material produced by mediums and sensitives, important advances have been made by Saltmarsh, Soal, and Fisher; by Pratt; and by Pratt and Birge. The Pratt-Birge method is limited by the fact that it cannot yield a high level of significance in tests with small groups of sitters or "target persons." In the search for a more sensitive method of evaluation, a statistical model has been found which can yield lower P values in such tests, which fulfills all necessary statistical requirements, and which is simpler to calculate than the Pratt-Birge method. These factors should make the new method a useful addition to the parapsychologist's statistical repertoire. We also present a method which can be used for combining results of ESP tests with targets having mixed probabilities.

INTRODUCTION

Parapsychologists are becoming increasingly preoccupied with gifted subjects, including so-called mediums and sensitives. Research with such individuals seems to reveal new aspects of the ESP process, as in studies of object association or psychometry, and it is one of the main hopes for an empirical solution of the question whether human personality survives death.

The ostensible ESP responses produced by mediums and sensitives are usually verbal utterances and, occasionally, drawings or behavior by which the subject attempts to imitate some target person (TP) or situation. This type of material is much more difficult to appraise statistically than, say, ESP card-guessing trials. The latter are planned to provide a constant and mathematically defined probability of a hit or a miss on any trial and also to enable the experimenter to combine a series of such hits and misses and arrive at an overall probability value. It is obviously more difficult to determine how unlikely it is that such statements as "Your grandmother's name is Joanna" and "You have an uncle who is a sea captain" are true by chance coincidence and to combine hits and misses of this type.

An important step in the solution of these problems was taken by H. F. Saltmarsh and S. G. Soal (6) in the appraisal of a series of statements by the medium, Mrs. Warren Elliott. Saltmarsh and Soal divided the medium's statements about the TP into categories

¹ We are grateful to Drs. T. N. E. Greville, J. A. Greenwood, and J. G. Pratt for reading this paper and providing helpful comments.

according to how applicable these statements seemed to be to the general population. Thus a quite specific response as "You got young girl in house, when trouble comes she collapses" (6, p. 270) received a probability figure of 1/10 while "used to read a lot" (6, p. 271), which seems to be true of many people, was assigned the larger p of 1/2. After this Saltmarsh and Soal combined the probabilities associated with all the responses, using a method developed by R. A. Fisher, to arrive at a probability figure representing the total result.

The determination of probabilities by subjective judgments represents an uncertain and undesirable element in a statistical analysis. J. G. Pratt (3) replaced this by a procedure where the probability of each statement was determined by the group of target persons participating in the tests. In these experiments, where Mrs. Eileen J. Garrett was the subject, there were fifteen TPs. In a given test, one of these occupied a room adjacent to Mrs. Garrett while she produced a series of statements supposedly pertaining to him. After the completion of all sessions, each TP annotated all the statements according to whether or not they applied to his circumstances, but without knowing which statements had actually been made by Mrs. Garrett when he was in the next room. On the basis of these annotations, it could be determined which statements were true for one, two, three, four, etc., of the TPs and the statements were assigned probability values accordingly. Using Fisher's method for combining probabilities, Pratt then arrived at a total result in the same way as had Saltmarsh and Soal.

This method was an improvement over the Saltmarsh-Soal approach because it substituted an objective for a subjective determination of the probabilities of the individual responses. However, a difficulty remained. This was due to the possible interdependence between the subject's responses. For instance, the subject might say after or prior to the statement about the uncle being a sea captain that "I hear the sound of engines" and "someone is looking at a chart." It is obvious that these responses would be likely to apply to the captain of a ship and that they would be checked as "correct" by a TP to whom the statement about the sea captain applied.

Again J. G. Pratt, this time in collaboration with W. R. Birge (4), responded to the problem. The Pratt-Birge method allows for any degree of interdependence between the subject's responses to any target person. In an experiment to be assessed by this method, each member of a group of TPs attends an experimental session with the subject under conditions where the latter has no sensory contact with the TP and where the TP, in turn, has no way of knowing what the subject says about him. At the completion of the

sessions, the experimenter divides the subject's responses to all TPs into separate items for annotation and sends copies of the full transcript to the TPs. Each TP then annotates the material, and the number of correct responses each TP recognizes as applying to his own circumstances or those of his living or dead relatives and friends is entered in a table and evaluated by analysis of variance (an earlier paper in this *Journal* (5) gives additional detail about these procedures). Table 1 shows the results of an imaginary

Table 1
PRATT-BIRGE ASSESSMENT OF FREE RESPONSES
(TARGET PERSON PERMUTATION MODEL)

	A	B	C	D	E	F	Total
A	3	2	2	1	1	2	11
B	0	3	0	0	2	1	6
C	2	3	4	1	2	2	14
D	2	1	2	2	2	2	10
E	0	3	1	0	3	2	9
F	2	1	1	1	1	2	8
Total:	9	13	10	5	11	11	59

Hits (entries on diagonal) = 17

Expected Hits: $\frac{1}{6} \times 59 = 9.83$

Deviation: $17 - 9.83 = 7.17$

$$V_a = \frac{3481 + (36 \times 131) - (6 \times 617) - (6 \times 619)^*}{36 \times (6 - 1)} = \frac{781}{180}$$

$SD = \sqrt{V_a} = 2.08$

$CR = \frac{Dev}{SD} = \frac{7.17}{2.08} = 3.4$

* In this formula, 6 represents the number of TPs; 36 is the square of this value; 3481 is the square of the total of all entries (59^2); 131 is the sum of the squares of the entries ($3^2 + 2^2 + \dots + 2^2$); 617 is the sum of the squares of the column totals ($9^2 + 13^2 + \dots + 11^2$); and 619 is the sum of the squares of the row totals ($11^2 + 6^2 + \dots + 8^2$).

experiment assessed by the Pratt-Birge method. The capital letters stand for the names of six TPs and the figures in the six-by-six matrix for the check marks entered by each of these TPs against the statements made by the subject in his own session and in each of the five other sessions. Thus A found three correct items in the session devoted to him, two in the session for B, two for C, etc. The hits are represented in the descending left-to-right diagonal

across the table. In this case the deviation amounts to 7.17 and the CR to 3.4.

A CR of 3.4 normally corresponds to a probability of .00034. However, Pratt and Birge warn us against using P values obtained by their method when the number of TPs is small. The warning is given because their method utilizes an approximation of the P value that becomes reliably close to the true value only for larger numbers of TPs.

This example illustrates a statistic which is calculated from the data as distinguished from a statistical model which allows us to take the further step and say how unlikely it is that the result given by the statistic could arise from chance coincidence. In the next section we shall compare the statistics and underlying models of the Pratt-Birge method and other methods. A new model will be introduced which is less limited than the Pratt-Birge method in its capacity to produce significant P values in experiments with small numbers of TPs.

THE RELATION BETWEEN A STATISTIC AND A MODEL

A "statistic" can be defined as a quantity computed from a set of data. Quantities such as hit totals, standard deviations, and critical ratios are statistics. We may also speak of "criterion" or "test statistic" upon which a particular statistical inference is to be based. When we attempt to ask a question of the data such as "Does ESP appear to be present?" it is standard practice to base the answer on the observed value of some test statistic such as the total number of hits. For all the methods discussed in this section, the criterion statistic is the total number of hits.

However, in order to make statistical inferences from data, it is not enough just to compute the value of a criterion statistic. One must also consider whether the value which is observed "could reasonably have arisen by chance." Implicit in the phrase "could reasonably have arisen by chance" is the existence of a mathematical specification or assumption involving probability which we can use as a model for the process which generated our data. In our mathematical model the test statistic, or any statistic we might obtain from the data, will be represented by a random variable, i.e., a variable whose value depends on chance.

For our study of models, we need a little more information about the experimental data than Table 1 provides. In Table 2, the same imaginary test is described, the capital letters again referring to six TPs. Among other things, this table shows that there was a total of twenty-five responses produced by the subject in the course of his six sessions with these TPs.

Table 2

ASSESSMENT OF FREE RESPONSES FOR TARGET PERSONS
PERMUTATION MODEL WITH INDEPENDENT ITEMS

Test No.	Target	Response	A	B	C	D	E	F	N-n	n(N-n)	Totals
1	A	1	✓						5	5	
		2	✓						5	5	
		3		✓		✓		✓	3	9	
		4	✓		✓				4	8	
		5		✓	✓		✓	✓	2	8	
2	B	6		✓				✓	3	9	
		7									
		8		✓			✓		4	8	
		9		✓					5	5	
3	C	10	✓	✓	✓			✓	2	8	
		11									
		12	✓		✓				4	8	
		13		✓	✓	✓		✓	3	9	
		14									
		15		✓	✓	✓	✓	✓	1	5	
4	D	16	✓		✓	✓	✓	✓	1	5	
		17	✓	✓	✓		✓	✓	1	5	
		18				✓			5	5	
5	E	19					✓		5	5	
		20		✓			✓		4	8	
		21	✓	✓			✓		3	9	
		22		✓			✓	✓	3	9	
		23									
6	F	24	✓	✓	✓	✓	✓	✓	0		
		25	✓					✓	4	8	
Totals			9	13	10	5	11	11			59
Total Hits			9	13	10	5	11	11			59
Total n(N-n)			26	34	21	10	27	23	141	141	

$$\text{Expected Hits: } \frac{1}{6} \times 59 = 9.83$$

$$\text{Va.} = \frac{1}{N^2} \sum n(N-n) = \frac{1}{6^2} \times 141$$

$$\text{Deviation: } 17 - 9.83 = 7.17$$

$$\text{SD} = \sqrt{\text{Va}} = \frac{1}{6} \sqrt{141} = 1.98$$

$$\text{CR} = \frac{\text{Dev}}{\text{SD}} = \frac{7.17}{1.98} = 3.6; \quad P = .00016$$

$$\text{Corrected Dev.} = 7.17 - .50 = 6.67;$$

$$\text{Adjusted CR} = \frac{6.67}{1.98} = 3.37$$

$$\text{Adjusted P} = .00038$$

Let us denote the total number of hits in our twenty-five item example experiment by H . Then H will be a random variable in our mathematical model. With our model the question "Could seventeen hits have reasonably occurred by chance?" can be given the mathematically precise formulation: "What is the probability that H would be seventeen or greater?" The answer to this question is called the P value for the observation $H = 17$. In order to be capable of answering this question with a P value, our model must include either a direct specification of the probability distribution of H or else enough information to permit this distribution to be deduced.

There are usually many different mathematical specifications (often called assumptions) which might be used to model a particular experiment. Different mathematical models will normally lead to different distributions for the criterion H and therefore to different P values. The basic difference between the Pratt-Birge method and the method suggested in this paper is a difference in the underlying model. As mentioned above, the criterion statistic H is the same for both methods.

The construction of a mathematical model or the choice between alternative models will depend most strongly on how well the assumptions correspond to reality. In other words, the extent to which the statements and rules which define a mathematical model correspond to the real world phenomena represented by the model is by far the most important criterion for evaluating the model. Also important is mathematical tractability—the ability of the model to yield answers to the questions we ask of it. Difficulties in its mathematical analysis can severely limit the usefulness of an otherwise realistic model.

In preparation for a detailed study of possible models for our example experiment, we introduce some more notation. As before, we denote the total number of hits by H . The possible values for H are the non-negative integers between 0 and 25. In the example the observed value of H is 17. Let X_i denote the number of hits on the i th item of the total of 25 items. The possible values for X_i are 0 and 1. In the example as shown in Table 2, $X_1 = 1$, $X_2 = 1$, $X_3 = 0$, $X_4 = 1$, $X_5 = 0$, $X_{25} = 1$. The relation between H and the X_i 's is $H = X_1 + X_2 + \dots + X_{25}$.

Mathematically speaking, the X_i 's are random variables as is their sum H . If we specify the probability distributions of the X_i 's (or, more properly, their joint distribution), we should in principle be able to derive the probability distribution for H and to answer the question: " $\text{Pr}(H \geq 17) = ?$ " We suggest this approach, rather than a direct specification of the probability distribution of H , because assumptions about the distribution of X_i 's tend to be easier to assess for realism than direct assumptions about the distribution of H .

There are several models which might conceivably be suggested for this experiment. We begin with one which is mathematically simple and which is familiar to most parapsychologists. In this model each X_i has the same probability of being one (scoring a hit) and the X_i are stochastically independent, i.e., the occurrence of a 0 or 1 for any X_i has no effect on the probability of any other X_i being 0 or 1. Under these conditions the random variable H will have a binomial distribution and the P value for $H = 17$ may be computed as

$$P = \text{Pr}(H = 17) + \text{Pr}(H = 18) + \dots + \text{Pr}(H = 25) = \sum_{h=17}^{25} \binom{25}{h} p^h (1-p)^{25-h}$$

provided p is specified. The letter, h , refers to an arbitrary possible value for the random variable H . It is natural to call this model the binomial model because it results in a binomial distribution for H .

The binomial model may be criticized on several points, but it will serve well to illustrate the use of critical ratios. There is an important mathematical theorem, the Central Limit Theorem, which states that under certain rather broad conditions the distribution of a random variable (in our case, H) which is the sum of a large number of other random variables (in our case, the X_i 's) may be approximated by the normal distribution with the same mean and variance. Without going into a lengthy discussion of the precise conditions under which the Central Limit Theorem holds (such a discussion can be found in (2), Chapter VI), we can say

roughly that the random variables in the sum must be not too highly dependent and their distributions must be not too different.

In the binomial model the X_i 's are stochastically independent and their distributions are identical. We may therefore wish to avoid the rather tedious computation of a P value from the exact distribution of H and compute instead an approximate P value using the normal approximation to the distribution of H. This is accomplished by computing a critical ratio and checking it against a table of the cumulative distribution function for the standard normal distribution.

The fitting of a normal approximation to a random variable H requires the mean and variance of H. Although the formulas for the mean and variance of a binomial random variable are widely known, we prefer to obtain them here as a consequence of general formulas for the mean and variance of a sum of random variables. If we write $E(H)$ for the mean (or expectancy) of H and $E(X_i)$ for the mean of X_i , our general formula for the mean can be written $E(H) = E(X_1) + \dots + E(X_{25})$. In other words, the mean of a sum is the sum of the means. In our binomial model the mean of each X_i can be shown by straightforward methods to be p so that $E(H) = p + p + \dots + p = 25p$.

The formula for the variance of a sum is a bit more elaborate. The variance of a sum is the sum of the variances plus twice the sum of all the covariances. In symbols, $\text{Var}(H) = \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_{25}) + 2[\text{Cov}(X_1, X_2) + \text{Cov}(X_1, X_3) + \dots + \text{Cov}(X_1, X_{25}) + \dots + \text{Cov}(X_{24}, X_{25})]$.

In the case where the X_i 's are uncorrelated, all the covariances are zero and the formula for the variance simplifies to: "The variance of a sum of uncorrelated random variables is the sum of the variances." In our binomial model the X_i 's are independent and therefore uncorrelated. Furthermore, each X_i has variance $p(1-p)$ and so $\text{Var}(H) = p(1-p) + \dots + p(1-p) = 25p(1-p)$. Thus, $(17 - 25p)/\sqrt{25p(1-p)}$

would be the appropriate critical ratio to use in the calculation of an approximate P value for the binomial model.

Of course, the binomial model is open to criticism for failure to correspond to the real world experiment it represents. The binomial model "assumes" the same probability of a hit on each item, but items certainly vary in their applicability to the general population. The assumption of independence from item to item is also questionable.

As a first step in improving the binomial model, let us permit item-to-item variation in the hit probabilities while retaining the

assumption of independence between items. This model could be called the binomial model with mixed probabilities. If we write $\text{Pr}(X_i = 1) = p_i$ so that p_i is the probability of a hit on the i^{th} item, we will have by our general formulas:

$$E(H) = p_1 + p_2 + \dots + p_{25}$$

$$\text{Var}(H) = p_1(1-p_1) + \dots + p_{25}(1-p_{25})$$

Unless many of the p_i are extreme (very near 0 or 1), we can expect the normal approximation to the distribution of H suggested by the Central Limit Theorem to give reasonably good approximate P values. An expression giving exact probabilities for H can be obtained, but the computation is even more tedious than for the binomial model.

A difficulty, which we have glossed over until now is the requirement that the p_i be known (or p in the binomial model). Whether we are computing exact probabilities or critical ratios, a knowledge of the p_i is necessary. Such knowledge is frequently hard to come by. We could, of course, use the data in Table 2 to obtain estimates

$\hat{p}_i$ of the true p_i . Thus, $\hat{p}_1 = \frac{1}{6}$, $\hat{p}_2 = \frac{1}{6}$, $\hat{p}_3 = \frac{3}{6}$ etc., could be

used in the formulas in place of the true p_i . Unfortunately, the formulas were derived under the assumption that the p_i 's were fixed and not subject to random variation. The $\hat{p}_i$'s are computed from the data and are therefore subject to chance fluctuations in repeated experiments. This difficulty can be overcome by a reformulation of the model, as we shall soon see.

The distribution for the random variable H, which is used to calculate a P value, is based upon a null hypothesis, which in turn is based on an assumption that no psi phenomena occur. It is sometimes possible to make a translation of an assumption of no psi phenomena into a suitable mathematical model. For example, if there is no relationship between an item and the person for whom it was intended, then it ought not to matter much if we relabel the persons who checked the items so that the target person for any particular item might be changed. In fact, we might consider making a random relabeling or permutation of target persons. Permuting target persons at random induces a random behavior in H which we can study. This device is employed in the Pratt-Birge method.

Our new model will not be fully specified until we decide on the nature of the permutation. We must decide whether (a) to make an independent random relabeling for each item; (b) to make one random relabeling serve for the entire set of data; or (c) to have one randomly selected permutation serve for all the items in

a particular test record, but to make separate, independent random selections for different records. We call the three models resulting from these three possibilities: (a) the TP permutation model with independent items; (b) the TP permutation model; and (c) the TP permutation model with independent records.

(a) Let us consider the TP permutation model with independent items. It is easily verified that $\Pr(X_i = 1) = \hat{p}_i$, where $\hat{p}_i$ is obtained as before, and that the X_i 's are independent. The test statistic H , therefore, has the same distribution that it did in the binomial model with mixed probabilities if the p_i 's are replaced by the $\hat{p}_i$'s. In the corresponding permutation model, however, the $\hat{p}_i$'s are the same for each permutation and so are not subject to random variation. As a result, the TP permutation model with independent items permits us to analyze our data by methods appropriate to the binomial model with mixed probabilities using $\hat{p}_i$ instead of p_i . This analysis has been done in Table 2.

(b) The assumption of item-to-item independence is a serious deficiency in the models we have considered so far. The TP permutation model does not presume item independence. This is the model upon which the Pratt-Birge method is based. Let us consider the probability distribution of H for this model. There are $6! = 720$ possible permutations of six people. For each permutation a value of H may be calculated. Different permutations can lead to the same value of H , but it is clear from a consideration of Table 1 that no rearrangement of the columns will lead to a value of H that is greater than or equal to 17. Therefore, our P value =

$\Pr(H \geq 17) = \Pr(H = 17) = \frac{1}{720} = .0014$. There are three permutations which yield an H of 16. These may be obtained from Table 1 by exchanging column A with column F, or column B with column E, or column D with column F. Thus, $\Pr(H \geq 16) =$

$$\Pr(H = 16) + \Pr(H = 17) = \frac{3}{720} + \frac{1}{720} = \frac{1}{180} = .0056.$$

For less extreme values for H , hand computation of exact probabilities may be impractical. It is not so easy to obtain the number of permutations for which $H = 13$, for example. It is easier to calculate the mean and variance of H in this model and use a critical ratio to obtain an approximate P value. This is the Pratt-Birge method. In our example it yields a critical ratio of 3.4 and an approximate P value of .00034 compared with the exact P value of .0014. A better approximation is obtained if H is corrected for continuity (see 1, Chapter 7) in the formula for the critical ratio. To correct for continuity, we simply subtract 1/2 point from the

observed number of hits. In this case the corrected H is $(17 - .5) = 16.5$. This adjustment yields a critical ratio of 3.2 and an approximate P of .00069.

(c) Consideration of the TP permutation model with independent records will be facilitated if we introduce some further notation. Let us write Y_i for the number of hits on the i th test record. In our example $Y_1 = X_1 + X_2 + X_3 + X_4 + X_5 = 1 + 1 + 0 + 1 + 0 = 3$, $Y_2 = X_1 + X_2 + X_3 + X_4 = 1 + 0 + 1 + 1 = 3$, etc. Thus, $H = X_1 + \dots + X_{25} = Y_1 + \dots + Y_6$.

Now, the name "target person permutation model with independent records" is a slight misnomer. The effect of selecting TP permutations independently for each record is to make the Y_i 's independent. As mentioned before, item-to-item independence for checks made by the same TP is a highly questionable assumption, but different Y_i 's arise from checks made by different TPs. Independent Y_i 's would occur in any model which assumed that the pattern of checks by one TP would have no effect on the probabilities of checks of another TP. It is not hard to argue that such an assumption would correspond well to the real world experiment we are attempting to model.

If a random permutation of TPs is made for a particular record, the only thing that matters is who will be the new TP. Equivalently, we may observe that the six values in row 1 of Table 1 represent

the six possibilities for Y_1 , each occurring with probability $\frac{1}{6}$.

Allowing for duplications, we have $\Pr(Y_1 = 3) = \frac{1}{6}$, $\Pr(Y_1 = 2) = \frac{3}{6}$, and $\Pr(Y_1 = 1) = \frac{2}{6}$.

The Y_1 of 3, which was actually observed, is the largest possible value for Y_1 in this model. Similarly, the observed values of Y_2 , Y_3 , Y_4 , Y_5 , and Y_6 are maximal. Therefore, $H = Y_1 + \dots + Y_6$ can be no greater than 17 and our P value is given by:

$$\Pr(H \geq 17) = \Pr(H = 17) = \Pr(Y_1 = 3, Y_2 = 3, Y_3 = 4, Y_4 = 2, Y_5 = 3, \text{ and } Y_6 = 2) = [\text{by independence of the } Y_i] \\ \Pr(Y_1 = 3) \times \Pr(Y_2 = 3) \times \dots \times \Pr(Y_6 = 2) = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} \times \\ \frac{5}{6} \times \frac{2}{6} \times \frac{2}{6} = \frac{5}{11,664} = .00043.$$

For this model also the computation of exact probabilities is much harder at less extreme values for H , and we might prefer to use a normal approximation. The formulas for the mean and

variance of H in this model can be evaluated using the information in Table 1. In general let us write m_{ij} for the number of checks made by the j^{th} TP on the i^{th} record, and let n be the number of TPs (or records). H , corrected for continuity, as before is $(17 - .5) = 16.5$.

$$E[H] = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n m_{ij} = \frac{59}{6} = 9.83^*$$

$$\begin{aligned} \text{Var}[H] &= \frac{1}{n^2} \left[n \sum_{i=1}^n \sum_{j=1}^n m_{ij}^2 - \sum_{i=1}^n \left(\sum_{j=1}^n m_{ij} \right)^2 \right] \\ &= \frac{1}{36} [6(3^2 + 2^2 + \dots + 2^2) - (11^2 + 6^2 + \dots + 8^2)] \\ &= \frac{1}{36} [6(131) - 619] = \frac{167}{36} = 4.64 \end{aligned}$$

$$\begin{aligned} \text{CR} &= (H - E[H]) / \sqrt{\text{Var}[H]} \\ &= (16.5 - 9.83) / 2.15 = 3.1 \end{aligned}$$

$$P = .00097$$

In this method we do not use the sum of the squares of the column totals (617) nor the square of the total of all entries (3481), both of which figures enter into the Pratt-Birge calculation (see Table 1).

Our method has a limitation which the Pratt-Birge does not have. Because the latter method allows for interdependence between the items (X_i 's), the subject need not be ignorant of the TPs and their characteristics. For instance, it would not matter if he knew that one of them is a sea captain. Of course, he should be kept ignorant of when, in the course of the tests, the sea captain is the TP. However, when our method is to be used, information about any one or more members of the group of TPs must be kept from the subject. If he knew that one TP is a sea captain and others not, this would erode the independence between the test records (Y 's) assumed by our model.

In both methods when the results from different tests have to be combined, the individual variances from each test are added together, as are the deviations. The standard deviation is the square root of the sum of the variances and the critical ratio is the sum of deviations divided by the standard deviation.

In summary, we can make the following remarks in comparing the Pratt-Birge method, which uses the TP permutation model, with the method suggested here, which uses the TP permutation

* This figure is obtained in the same way as in Tables 1 and 2.

model with independent records. First, the new method can only be employed if the subject has no normally acquired information about any of the TPs. Secondly, the new method uses the same way of calculating the number of hits and the expected number of hits as does the Pratt-Birge method. Thirdly, the new method can yield a smaller exact P value than is possible with the Pratt-Birge method. With six TPs the smallest possible P value using the Pratt-Birge method is $1/6! = .0014$, but the new method could have a P as low as $1/6^6 = .00002$. Of course, the new method will attain its minimum exact P less often than the Pratt-Birge method will attain its minimum P . However, it is likely that the new method will usually result in smaller exact P values than the Pratt-Birge method. In our imaginary experiment, the exact P value of .00043 for the new method is smaller than the exact P of .0014 for the Pratt-Birge method even though the Pratt-Birge P is at its minimum and the new method P is not.

Fourthly, there is a difference between the exact P values and the approximate P values computed from CRs for the two methods. Although general conclusions cannot be made at this point, the results of our example experiment are of interest. The approximate P value (derived from the CR) given by the Pratt-Birge method is .00069, while the exact P was less significant at .0014. Also, in the new method the approximate P differed from the exact P by a factor of about 2. But in this case the approximate P value erred on the safe side, giving the less significant figure of .00097 as against the exact P of .00043. Thus, for this example at least, use of the CR approximation with the new method would not lead to an overstatement of the significance of the result.

Fifthly, the formula for the variance used in the CR approximation is easier to compute with the new than with the Pratt-Birge method.

Finally, we repeat for emphasis that independence of the Y_i 's, which the new method requires, is a consequence of the reasonable assumption of independence between different TPs. We conclude therefore that the new method should be a useful addition to the parapsychologist's statistical repertoire.

APPLICATION TO ESP GUESSING EXPERIMENTS

Most ESP tests, including card-calling trials, consist of long series of guesses of targets with the same probability values, such as the five common ESP symbols where the probability of a hit is always one fifth. The method described here makes it easy to conduct and assess tests with targets of mixed probabilities. Such tests might be designed to study the preferential effect, to inhibit

guessing habits, or for a number of other purposes. The results of tests with targets having different probabilities can be assessed together as in the imaginary experiment described in the present section (Table 3), or the tests with the same targets can be assessed separately and compared to the results obtained with the other targets.

Table 3

BINOMIAL MODEL WITH MIXED PROBABILITIES

Trials		Target Situation										Hits	P
1	7	9	5	3	8	10	2	6*	1	4	1	1/10	
1	8	3	5	7*	4	6	10	2	1	9	1	1/10	
3	6	7	2	3	8	1	5*	9	4	10	1	1/10	
4	L	△	△	△	△	△	△	△	△	△	1	1/5	
5	+	+	+	+	+	+	+	+	+	+	1	1/5	
6	Green*	Green*	Green*	Green*	Green*	Green*	Green*	Green*	Green*	Green*	1	1/2	
7	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	1	1/2	
8	White	White	White	White	White	White	White	White	White	White	1	1/2	
9	Green*	Green*	Green*	Green*	Green*	Green*	Green*	Green*	Green*	Green*	1	1/2	
10	White	White	White	White	White	White	White	White	White	White	1	1/2	

Totals

6 32/10 3.2

$$\text{Expected Hits: } p_1 + p_2 + \dots + p_{10} = \frac{1}{10} + \frac{1}{10} + \dots + \frac{1}{2} = 3.2$$

$$\text{Deviation: } 6 - 3.2 = 2.80$$

$$\text{Va.: } \frac{1}{10} \left(1 - \frac{1}{10}\right) + \frac{1}{10} \left(1 - \frac{1}{10}\right) + \dots + \frac{1}{2} \left(1 - \frac{1}{2}\right) = 1.84$$

$$\text{SD} = \sqrt{\text{Va}} = \sqrt{1.84} = 1.36$$

$$\text{CR} = \frac{2.80}{1.36} = 2.0 \quad P = .02275$$

$$\text{Adjusted Deviation: } 2.80 - .50 = 2.30$$

$$\text{Adjusted CR} = \frac{2.30}{1.36} = 1.69 \quad \text{Adjusted P} = .04551$$

* The targets are in italics and the responses are indicated by asterisks.

Table 3 shows the results of an imaginary experiment consisting of ten trials. Of these, trials 1-3 are associated with probabilities of 1/10; trials 4-5, of 1/5; and trials 6-10, of 1/2. The first three trials might consist in guessing the numbers 1-10, the next two might be guessing the five ESP cards, and the last five might consist in calling two colors, say white and green. The numbers, symbols, and words in the center of the table illustrate the target range. The target for each response is underlined and the guess is

represented by an asterisk. There are six hits. Our model is the binomial model with mixed probabilities.

For convenience, we use a slightly different expression of the formula shown in the previous section of this paper. However, the calculations remain simple. The variance is: $\text{Va.} = p_1(1 - p_1) + p_2(1 - p_2) + \dots + p_{10}(1 - p_{10})$, where p is the probability of a hit on any of the items 1-10. The variance thus works out at:

$$\text{Va.} = \frac{1}{10} \left(1 - \frac{1}{10}\right) + \frac{1}{10} \left(1 - \frac{1}{10}\right) + \dots + \frac{1}{2} \left(1 - \frac{1}{2}\right) = 1.84.$$

The expected number of hits is the sum of the probabilities of hits on each of the items:

$$\begin{aligned} \text{Expected Hits: } p_1 + p_2 + \dots + p_{10} \\ = \frac{1}{10} + \frac{1}{10} + \dots + \frac{1}{2} = 3.2 \end{aligned}$$

Consequently the deviation is: $6 - 3.2 = 2.80$.

The standard deviation is the square root of the variance:

$$\text{SD} = \sqrt{\text{Va}} = \sqrt{1.84} = 1.36;$$

and the critical ratio is:

$$\text{CR} = \frac{2.80}{1.36} = 2.0 \text{ which has a probability}$$

of $P = .023$. The adjusted deviation is: $2.80 - .50 = 2.30$, the

adjusted critical ratio is: $\text{CR} = \frac{2.30}{1.36} = 1.69$, and the adjusted probability is: $P = .046$.

The assumption of independence between the items which underlies the binomial method with mixed probabilities is the same which underlies Fisher's method (6) we referred to in the Introduction. The difference between the two methods is the result of a transformation of the X_i 's in the Fisher method so that instead of being 1 or 0 for a hit or a miss they are, respectively, $-(1 - p_i) \log p_i$ and $+p_i \log p_i$. In other words, Fisher's method provides a weighting of the items with different values for different probabilities. This type of transformation is likely to add to the sensitivity of the assessment. Other transformations than Fisher's can be used in the models discussed in this paper, the choice mainly depending on how extreme the probabilities are. This subject will be taken up in a later paper.

REFERENCES

1. FELLER, W. *An Introduction to Probability Theory and its Applications*. New York: John Wiley and Sons, 1957.
2. LOEVE, M. *Probability Theory*. New York: Van Nostrand, 1960.
3. PRATT, J. G. "Towards a Method of Evaluating Mediumistic Material." *Bulletin XXIII*, Boston Society for Psychic Research, March, 1936.
4. PRATT, J. G., AND BIRGE, W. R. "Appraising Verbal Test Material in Parapsychology." *Journal of Parapsychology*, Vol. 12, December, 1948, 236-256.
5. ROLL, W. G. "Designs for Tests with Free Response Material." *Journal A.S.P.R.*, Vol. 56, October, 1962, 184-195.
6. SALTMARSH, H. F., AND SOAL, S. G. "A Method of Estimating the Supernormal Content of Mediumistic Communications." *Proc. S.P.R.*, Vol. 39, 1930-31, 266-271.

Psychical Research Foundation, Inc.
 Box 6116, College Station
 Durham, North Carolina 27708

Department of Mathematics
Duke University
 Durham, North Carolina 27706