Chapter 7. The Metastudy: A review

The empirical foundation for adopting the logistic-J distribution as a universal descriptor of abundances in natural communities rests on the metastudy described briefly in Chapter 1. In that metastudy, 125 randomly selected biosurveys were treated as random samples of their respective communities (as defined by the respective authors) and fitted to a logistic-J distribution having the same mean abundance. The resulting scores were then compiled into a new "meta-histogram" that portrayed the collective shape of the scores.

The concept of a perturbation is central to the application of the theory underlying the chi square distribution, as explained in Section 4.3. The chi square scores of a random collection P of perturbations of a given theoretical distribution will have the chi square distribution. It is shown in Section 7.4 how the random samples that make up the metastudy data can be treated as perturbations, each of a different logistic-J distribution, but with a collective behaviour that is close enough to draw the conclusion that the field data do, in fact, follow the logistic-J distribution.

A parallel to the general idea behind the experiment about to be reported can be found in the Central Limit theorem (Hays and Winkler, 1971). If one takes m samples from m different distributions, each with mean μ, then the averages of those samples have a normal distribution having the same mean. In other words, if you distribute the resulting m mean values into the appropriate range of numbers, the histogram that emerges has a normal shape, albeit possibly ragged. However, the more samples one takes, the more closely the histogram fits the normal curve in question. This sort of experiment is often performed by statistics students, thanks to software that simulates the process. (Regan, 2011)

The following sections provide a description of the experimental protocol, a summary of the results, and an explanation of the theory behind the protocol.

7.1 The test protocol

Data for the metastudy originated in some 125 papers, as published in a wide variety of journals. (See Metastudy References in the Bibliography.) Each paper was selected at random from the University of Western Ontario library by graduate students who had been given a method for selecting journal volumes at random and, within those volumes, a paper at random. The only nonrandom element was an occasional intervention by the author to include a specific group, hitherto missing from the study by chance. It can be said, however that selection of papers within such groups followed the same randomizing procedures. In all cases, however, papers that listed fewer than 30 species were excluded from the study, as were papers that addressed only "common" species within a group. The papers were collected under these criteria prior to any analysis of their contents. None were rejected at any time beyond the collection stage.

From each paper a species-abundance table (field histogram) was extracted and the abundances were compared, via the chi square test, with the predictions made by both the logistic-J distribution and the log-series distribution, that being the theoretical proposal closest to the logistic-J in overall shape. The parameters for either distribution were selected as follows:

> For the logistic-J distribution, the Fa-μ method of curve-fitting was used. In this method, the mean abundance μ of the field histogram, along with the number of species Fa in the minimum abundance category were used to derive the corresponding values of ε and Δ in order to obtain a candidate source distribution. (See the transfer equations in Appendix A.1) While some of the resulting fits were better than average (less than the number of degrees of freedom), other fits were worse. Such a result was not only consistent with chi square theory, but demanded by it in this context. The 125 scores were then compiled into a "meta-table", as shown in Appendix C.1.
>
> For the log-series distribution (see Appendix A.5), there is only one way to select a candidate source distribution with the same mean, the log-series being a one-parameter distribution. The parameter value c and constant α, not being independent, were both derived from from μ, the mean abundance. Thus, when it comes to comparisons, both distributions were on an equal footing, both being fitted to curves based on the mean abundance and adjusted with their respective degrees of freedom. Again, some of the scores were sub-optimal, while others were higher, as was the case with the logistic-J distribution. Indeed it was fully expected that some of the log-series fits would be better than their logistic-J counterparts and this happened from time to time. Again, 125 chi square scores were recorded, this time for the log-series. (See Appendix C.1)

All chi square scores for both sets of tests were then converted into the equivalent scores at 10 degrees of freedom. (This corresponds to a common p-value, for those familiar with the chi square test.) The purpose of this normalization process was to make the results of the 250 tests comparable within the data for each distributions, as well as between distributions. A paired interval test was employed to detect any significant difference between the two sets of scores.

Since any fit of a logistic-J distribution to empirical data will result in a specific value for the parameter Δ, that value automatically becomes a prediction of the corresponding empirical parameter, Δ'. The comparisons were made in the form of a percentage ratio $\Delta'/\Delta \times 100$ as described in Section 5.1. We note in passing that to detect any difference between the two sets of test scores by eye alone would be completely impossible. Averages, standard deviations, and other basic statistics were indispensable to the study.

7.2 Experimental results of the metastudy

The results of the metastudy support the claim that the collective score for the logistic-J distribution was optimal, or as near to optimal as one could reasonably expect, given the variability in the scores themselves. The score of 10.0 is the expected score for a family of perturbations under the null hypothesis at 10 degrees of freedom. The average chi square (normalized) score of the metastudy was 10.43 which is reasonably close, given the high variance (sd 4.64} of the scores. The degree of closeness to the chi square distribution itself was calculated in another way. If we compute the overall score differently, taking the percentage ratio of each individual chi square score with the corresponding number of degrees of freedom, we end with an average ratio of 0.998, slightly below the expected 1.000.

For those who may not yet have grasped the full significance of optimality in this context, it could be pointed out that any researcher who employed the protocol described above, ending with a score of 7.3, say, would immediately fall under suspicion of falsifying the results, whether consciously or unconsciously.

The overall test score for the log-series distribution was 13.56 (s.d. 9.40), well above the logistic-J score, but within a standard confidence interval of it. To tease out systematic score differences for the two theoretical distributions, a paired interval test was applied. In this test the statistic is formed as the sum of individual squared differences of pairs of scores, one pair for each set of biosurvey data. The compiled sum of these differences resulted in the interval [0.64, 4.89]. In other words, the mean difference between the paired scores lay within this interval with a probability of 99.9 %. The statistical interpretation can also be phrased as follows: "We would err with a probability of 0.001 in claiming that the paired differences are bounded away from zero." In other words, there is a residuum of difference that cannot be accounted for as random. The better score of the logistic-J distribution in this context is apparently the result of it being the better descriptor of species-abundance field histograms, in general.

Although the (slightly) higher than expected average chi square score of 10.43 and the (slightly) higher than expected Δ-prediction of 101.6 are about as close as can be expected, one is still free to speculate that the two results are related in the following sense: there was a slight tendency for those histograms that fit the source distribution less well also to have maximum abundances that had ratios of Δ'/Δ higher than 100.0. A reverse finding would of course be more surprising.

The tests carried out as part of the metastudy were very time-consuming, involving hundreds of hours to complete. There seemed little point in including tests of the lognormal and other distributions for the following reasons. First, of all the extant proposals for theoretical species abundance distributions, the log-series is closest in general shape to the logistic-J, both being essentially hyperbolic functions. Although the log-series is a reasonable approximation to the logistic-J for some purposes, the insertion of a convergent series (as suggested by R. A. Fisher to Corbett & Williams 1943) amounts to what physicists would call a "fudge factor," meant to force the area under the log-series curve to have a finite value. Inaccuracies introduced by this factor, although slight, appear to show up in the significantly different overall test scores for the two distributions.

As for the lognormal distribution, the situation is basically hopeless. When used with the correct "veil curve" (Dewdney 1998) and the standard abundance axes, the lognormal distribution would yield average normalized scores one or two orders of magnitude higher than the results given above, putting the lognormal out of the running. One reason for the poor fits would be the fact that the theoretical lognormal curve would tend to have a low value in the lowest abundance category, while the histogram with which it is compared would have a rather high one. Moreover, at slightly higher abundances, the situation is reversed: where the typical field histogram has declined to fewer species, the lognormal is peaking. As I have pointed out in Section 4.2, the use of logarithmic axes in such a comparison would prove nothing, since both the logistic-J and log-series distributions produce unimodal distributions when subjected to this treatment.

In short, to do well in this kind of metastudy, any candidate theoretical distribution would have to be exceedingly close in general shape to the logistic-J. A humble example may serve to make the point. If one tested the uniform distribution U[2] x 100 against the outcomes of 100 tosses of an unbiased coin, a total of 50 heads would be the expected outcome, with scores from 45 to 55 not uncommon. A proposed theoretical distribution for the outcomes of the coin-tossing experiment that was different from U[2] x 100 would clearly have to be close to the uniform distribution to score as well or even (by accident) slightly better.

7.3 The distribution of chi square scores

In Section 2.5 I described Chi square theory in terms of multiple random perturbations of a source distribution, namely the theoretical distribution under test. The theory asserts that the perturbations, when tested collectively against the source distribution yield chi square scores that follow the chi square distribution. Since Pearson's theorem applies to any distribution whatever, we can illustrate it with the simplest of distributions, the Uniform Random, symbolized in this case by U[5] x 50, shown in Figure 7.1.

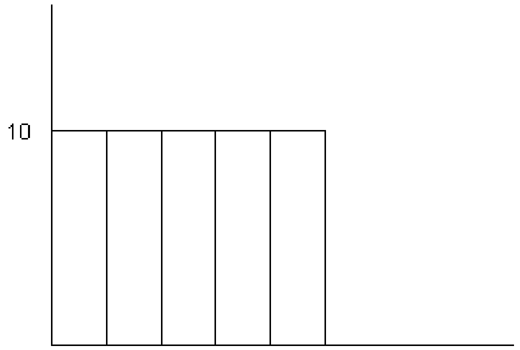

Figure 7.1 A uniform source distribution

The distribution shown has five categories and a total of 50 items (interpreted as species, if one wishes) distributed uniformly among them. Using U as the source distribution, we may generate as many perturbations as we like by using a computer program that mimics the variations within columns in a precise way. Figure 7.2 shows three examples of perturbations thus generated. These were selected from among the first dozen examples that were generated by computer, seeming pathologies being relatively common. They illustrate the point that visual inspection of samples can be misleading. While Figure 7.2a looks straightforward enough, Figures 7.2b and 7.2c appear to have originated in unimodal and triangular distributions, respectively.

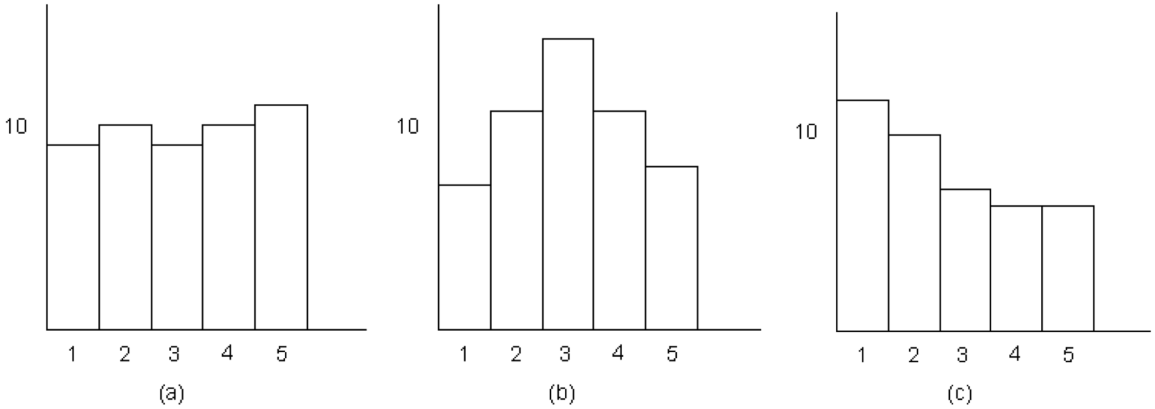

Figure 7.2. Three random perturbations of the uniform distribution

If we now generate 100 perturbations of the theoretical curve U[5] x 50 and compute a chi square score for each one in comparison with U[5] x 50, we may compile the scores into a new histogram in which the height of each bar is the number of scores that fell into that category. Such an experiment was performed using the computer program Uniform, which also computed the chi square score for each perturbation. The experiment was run for 100 trials. The resulting histogram appears in Figure 7.3 with the expected chi square distribution superposed on it.

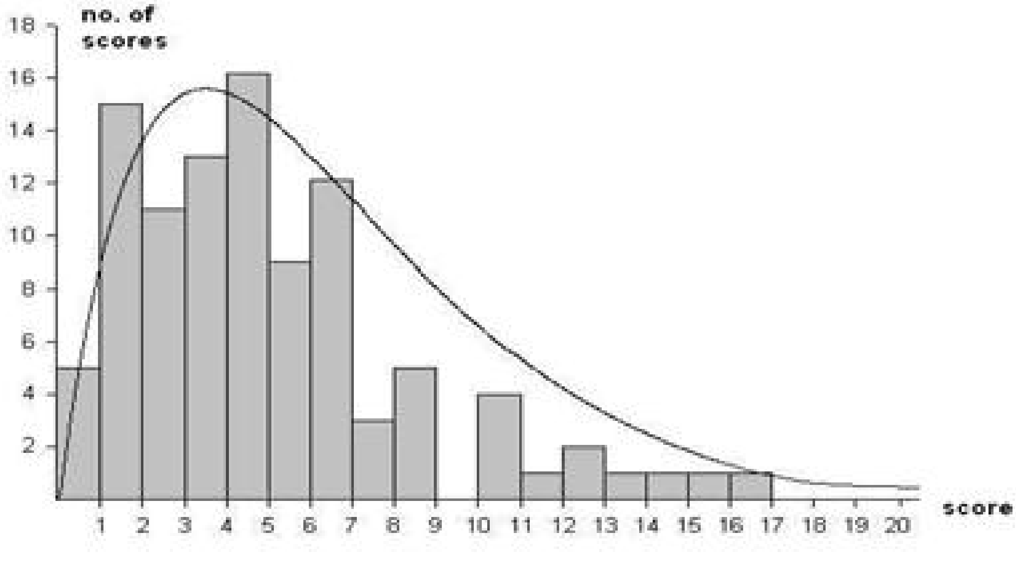

Figure 7.3. Histogram of chi square scores for uniform distribution

As expected, the fit is fairly good for a compilation of chi square scores, the more important result being the average score of 4.958, rather close to the expected 5.00 for an average fit under the null hypothesis. The theoretical chi square values, rendered as a curve, may be compared with the experimental values, rendered as bars. Naturally, there is some variation in the latter, owing to the usual statistical fluctuations.

7.3.1 Free and bound perturbations

In what follows it will be necessary to distinguish between perturbations that are "free" and ones that are "bound." The following definition incorporates the notion of degrees of freedom into the definition of a perturbation.

In Section 1.6.1 a perturbation of a source distribution F was defined for each abundance category k as the sum of F(k) and a random variable V(k) having a normal distribution. with mean zero and standard deviation σ(k). A perturbation will be called *free* if the variations in its individual functional values F(k) are independent, otherwise it will be called a *bound perturbation*.

It may well happen that a perturbation that is bound over m columns is free over fewer than m columns. To make the notion precise, the *degree* of a perturbation is the maximum number of columns over which it is free. For example, if the source distribution has m columns and σ(k) = 0 for a given column k, there is no freedom for the column to vary at all. In such a case, the perturbation will be bound, but free of degree m-1, provided that the remaining columns are free to vary independently. (In general, the standard deviations will be appropriate to the expected variation.) Another, more subtle example involves a global restriction on the variations within columns. One way to generate perturbations of this distribution is to allocate the 50 species in F among the various abundance categories at random. Such a perturbation is not free because the number of species that end up in a given category is a strict function of the number of species in the other categories. The restriction that the frequencies so assigned must sum to 50 means that the columns are not independent, so the perturbation is bound and not free. In the face of such a restriction on perturbations, there is a loss of one or more degrees of freedom. Such a perturbation would have degree 4. Thus if we had generated the perturbations in the previous experiment by this method, the average chi square score would have been approximately 4.0, instead of approximately 5.0.

7.4 Applying chi square theory to multiple-source histograms

In Section 7.3, it was shown how chi square theory works in the standard case of multiple perturbations of a single source. In this section we apply it to single perturbations of multiple sources. The space of all possible perturbations and sources is represented in Figure 7.4. Here, the horizontal plane represents the parameter space of the logistic-J distribution, with each grid point representing a pair of parameter values. The parameter epsilon is quantified by the cross-page axis, while the parameter delta follows the other horizontal axis into the page. If all such distributions had, say 10 degrees of freedom, every possible logistic-J distribution with 10 degrees of freedom would be represented by a point in this two-dimensional parameter space.

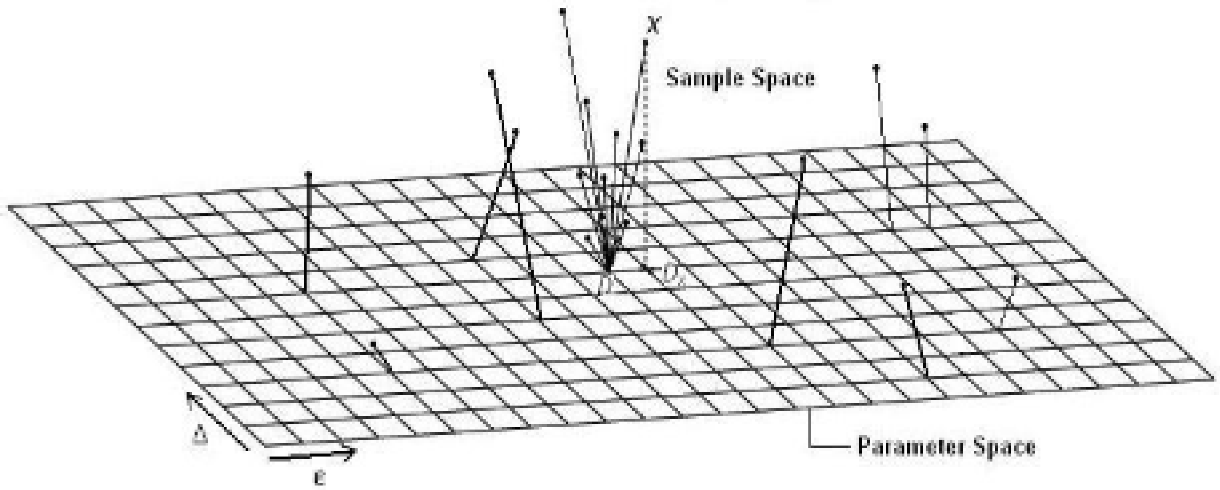

Figure 7.4 Parameter space and sample space

In the vertical direction, above the plane, we may locate all possible field histograms as follows. Given any field histogram X, we may carry out a best fit procedure (see Section 4.3) to determine a unique theoretical version of the logistic-J distribution yielding the minimum chi square score with respect to the histogram. One may then locate the histogram X in question directly above the best-fit theoretical distribution Dx at a height equal to the score. Points high above the parameter plane do not fit any of the theoretical versions of the logistic-J distribution very well, while points lower down do so. In this representational scheme the best fit procedure plays no other role than to enable us to assign a position to the perturbation in sample space.

If one now selects a specific logistic-J distribution in parameter space and takes multiple free perturbations of it, each perturbation can be represented by a single point located somewhere above the parameter plane, as described above. A straight line segment joining the source distribution to that point helps one to visualize the relationship between the source and the perturbation. The result is a cluster of line segments, as shown in Figure 7.4.

Each of the remaining (isolated) lines is identical to one of the lines in the cluster; having the same length and orientation. The two sets of lines, one clustered, the other dispersed, illustrate the application of chi square theory to the many tests comprising the metastudy. For the moment, we will pretend that each isolated source is a logistic-J distribution having 10 degrees of freedom and the point at the other end of each such segment represents a field histogram being fitted to this source. The point made by this figure is that to carry out many tests on different field histograms is equivalent to perturbing a single logistic-J distribution many times. Since the scores associated with the clustered source follow the chi square distribution, so do the scores associated with the isolated points. The only question that remains is how close the average score will be to the number of degrees of freedom of the source, namely 10 in the case at hand.

One more condition must be satisfied before the argument gets home free. The degree of the perturbation represented by the field histogram must equal the number of degrees of freedom of the logistic-J distribution acting as its source. If the source has ten degrees of freedom, it must have 12 abundance categories and so must the field histogram, as explained in Section 4.3.

In the metastudy, source distributions were selected by the F_a-μ method, as applied to the field histogram. Here, values for F_a (height of the initial peak) and μ (the mean abundance) become inputs for the transfer equations and the outputs are values for the parameters ε and Δ. The transfer equations (See Appendix A.4.) guarantee that the logistic-J distribution with these particular parameter values also have initial peak size F_a and mean μ.

Figure 7.5 displays an example of the fitting procedure. The dashed lines represent bar heights of the field histogram, while the bars represent the source distribution arrived at via the transfer equations. As one can see, the solid bars show an even progression to smaller and smaller values as abundance increases, whereas the dashed bar heights show a similar, but rather more ragged progression. We note that the dashed and solid lines coincide in the first column of minimum abundance. (a = 1)

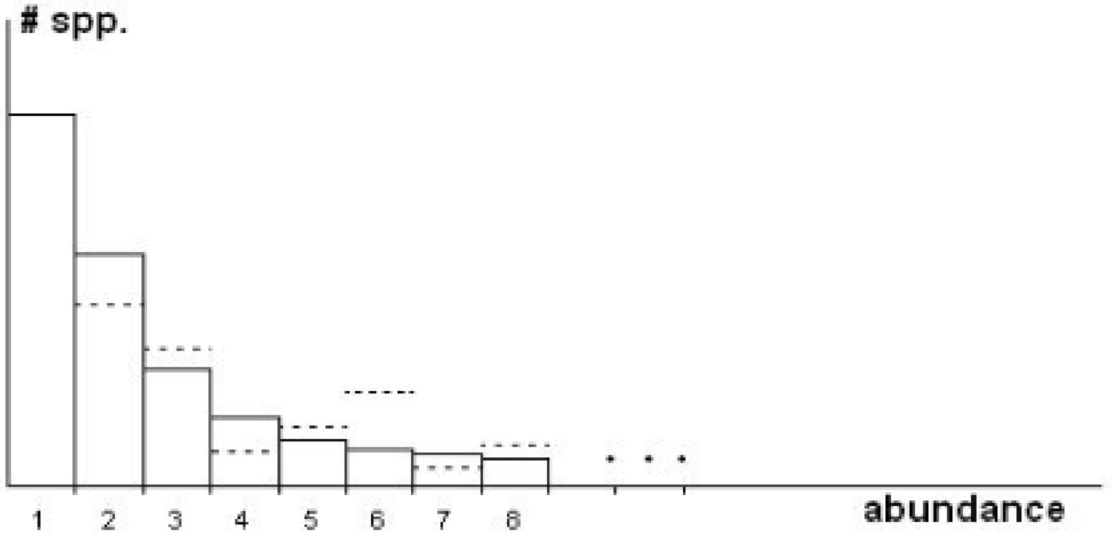

Figure 7.5a Field histogram compared with source distribution

It is immediately evident that the field histogram cannot be regarded as a free perturbation, since the value of the perturbation at abundance category 1 is completely restricted and not free to vary. This cancels out one abundance category, in effect. Now if the first category is deleted, the result appears to be a free perturbation on one less category, as shown in Figure 7.5b.

However, even the reduced histogram is not free, since the number of species is now (R - F_a) and no matter how the remaining bar heights may vary, they are constrained to add up to this quantity. This means that the height of any one bar can be expressed as a function of the remaining bar heights and they are not independent.

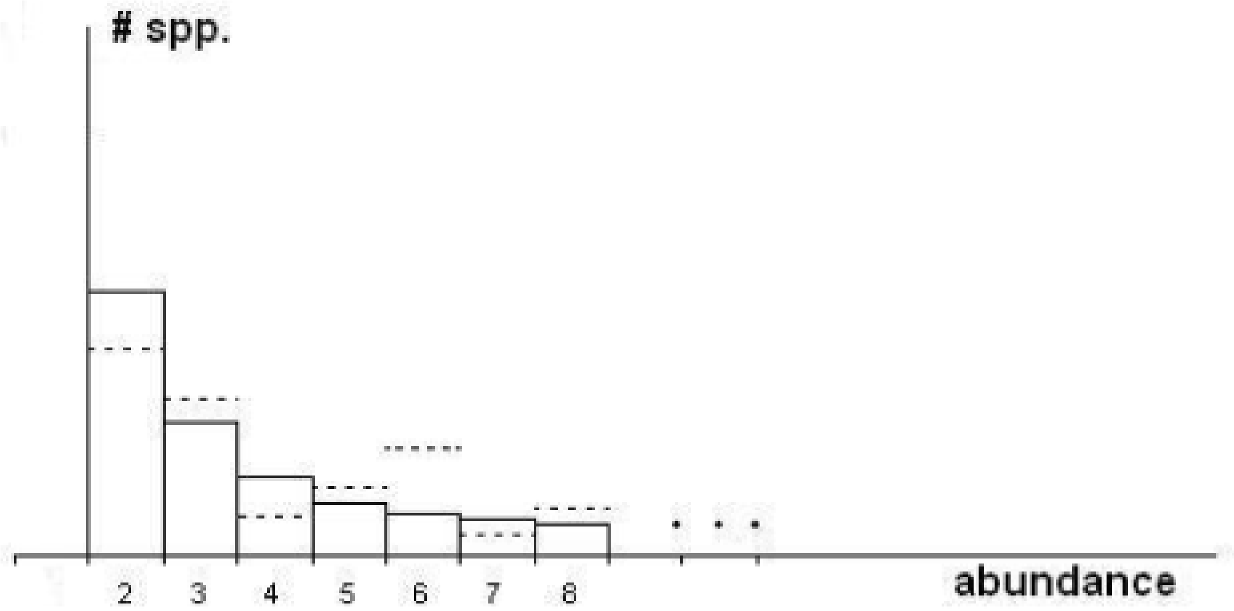

Figure 7.5b Reduced field histogram

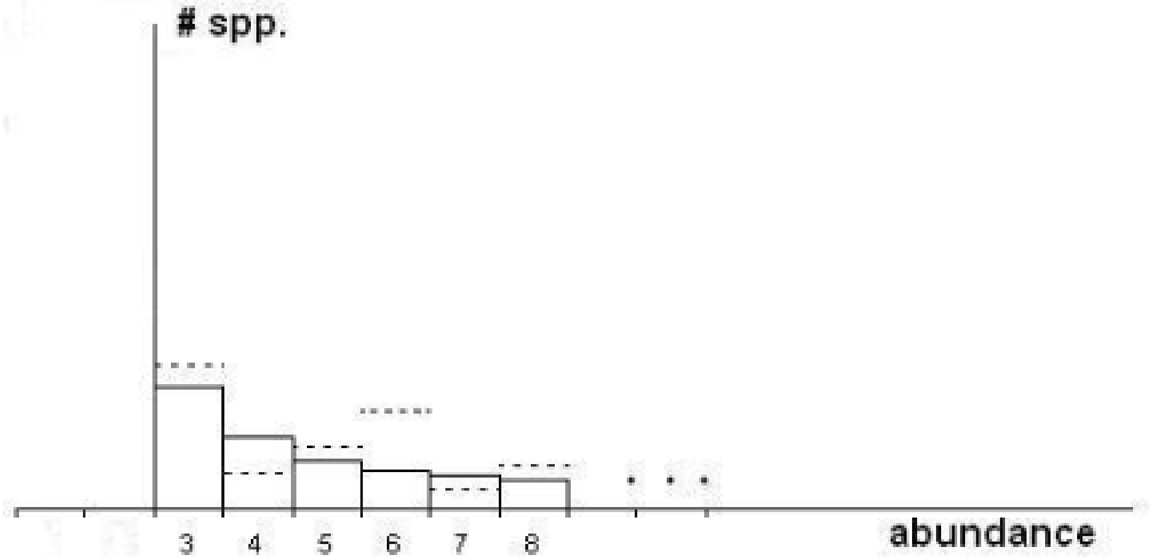

Figure 7.5c A free perturbation with eight degrees of freedom

One must therefore remove one more column, as in Figure 7.5c, before the perturbation is actually free. The process has resulted in a perturbation of degree 10. Since the degree matches the number of degrees of freedom, the optimal average score for the tests is 10.0.

The foregoing arguments lean heavily on the assumption that the results of a chi square test with some other number of degrees of freedom can be made equivalent to the results of a test on 10 degrees of freedom. I will illustrate the equivalence operation with a simple example wherein the chi square table plays a key role.

In Table 7.1 I have excerpted a small portion of two rows from a standard chi square table of critical values. In the process I have reduced the decimal factions from five decimal digits to three for the sake of simplicity. As explained in Section 4.2 the q-values represent the probability that one is incorrect in rejecting a chi square score that is higher than the corresponding critical value. Suppose, for example, that one has conducted a chi square test with six degrees of freedom on a particular field histogram, obtaining a score of 4.721 in the process. Note that this score exceeds the critical value of 3.455, yet falls short of 5.348.

Because it exceeds the lower critical value, one could say that the score is too large to be “accepted” as having the logistic-J distribution at the level of p = 0.750. In other words, one could reject the null hypothesis (that the histogram has the logistic-J distribution) with a probability of 0.750 of being wrong. The corresponding probability p of being “right” is 0.250, since the complementary probability has the relationship p = 1 - q with q. If the score exceeded 5.348, on the other hand, the probability of rejecting it incorrectly falls to 0.500. As I pointed out in Section 4.2, one must have a rather higher score in order to reject it at the highest level, where the probability of being wrong has fallen to q = 0.050. At this level the corresponding p-value is 0.950 and one says that one has 95 % confidence that the rejection was correct.

Q	0.750	0.500
6 df	3.455	5.348
10 df	6.737	9.342

Table 7.1. critical values for 6 and 10 degrees of freedom

The foregoing explanation serves only to refresh the reader’s memory about the chi square table and how it is normally used. Here, I shall use the table to provide an index of equivalence, so to speak. What score at 10 degrees of freedom corresponds to the score of 4.721 at 6 degrees of freedom?

The simplest approach is somewhat approximate. Linear interpolation treats the probability curve between the two critical values as a straight line, whereas the probability function is somewhat curved over the domain spanned by the two critical values. The discovery of an equivalent score at 10 degrees by this method involves a simple proportion. If a and b represent the adjacent critical values at 6 degrees of freedom, while c and d represent the corresponding critical values at 10 degrees of freedom, we seek a value x such that x occupies the same position in relation to the interval [a, b] as 4.721 does in relation to the interval [c, d]. One has then,

$$x = 9.479$$

The chi square score at 10 degrees of freedom corresponding to 4.721 at 6 degrees of freedom is, under this scheme, 9.479. This score occupies the same position in the interval [6.737, 9.342] as 4.721 does in the interval [3.455, 5.348] and the two scores may be said to be equivalent under linear interpolation. The continuous versions of the curve formed by critical values as a function of p-value is very slightly curved upward over the domain exceeding $p = 0.500$ and very slightly downward below that value. The interpolated values are therefore slightly too high in the first case and slightly too low in the second case. Such errors would have an impact on the final average by introducing consistent inaccuracies over the domains in question. Under the null hypothesis, this would drive scores up slightly.

Note: We can explain the lower scores achieved by histograms with fewer degrees of freedom by the increased tendency of a single column (say) to match the mean more frequently than sevral columns to do the same. There is also the question of weighting to deal with. Should low-df histograms be weighted the same as high-df histograms?