

Count Data Models in Biometry and Randomness Patterns in Birds Extra-Pair Paternity

Tiago André Marques¹, Dinis Duarte Pestana^{2*}, Sílvia Filipe Velosa³

¹Universidade de Lisboa, CEAUL, Centro de Estatística e Aplicações da Universidade de Lisboa.
Campo Grande, 1749-016 Lisboa, Portugal, +351-21-7500000 tiagomarques@fc.ul.pt

²Universidade de Lisboa, DEIO and CEAUL, Centro de Estatística e Aplicações da Universidade de
Lisboa, Campo Grande, 1749-016 Lisboa, Portugal, +351-21-7500040 dinis.pestana@fc.ul.pt

³Universidade da Madeira, DMCE and CEAUL, Centro de Estatística e Aplicações da Universidade
de Lisboa, Campus da Penteada, 9000-390 Funchal, Portugal, +351-291-705177
svelosa@math.uma.pt

SUMMARY

The number of extra pair nestlings in a brood is the basic information to investigate extra-pair fertilization in socially monogamous birds, an interesting pattern of behaviour that has been observed in some species. Under unconstrained randomness, Poisson streams of events are expected. But other patterns of randomness may arise, suggesting new research questions. Starting from a coordinated approach to count models, we discuss Zipf-Mandelbrot self-organizing scaling laws, which are typical of phenomena shaped as a result of conflicting interests, and some extensions of Mandelbrot's model. While the traditional count models (Poisson, binomial, negative binomial or hypergeometric) seem inappropriate, the logarithmic, truncated logarithmic, Zipf-Mandelbrot and discrete lognormal models consistently provide the best fit to the available data, indicating that probably some females are more prone than others to have extra pair nestlings. This suggests a delicate balance: the number of extra pair nestlings in the progeny is the result of conflicting behaviours, the search for genetic diversity and the need to ensure male cooperation in raising the brood.

* Corresponding author: Dinis Pestana, fax +351-21-7500081

Key words: Count data, extra-pair fertilization, goodness-of-fit, Poisson related and Zipf – Mandelbrot related models, randomness.

This book concentrates, not on how to do an analysis, but on how to choose the right sort of analysis, and how to make sense of the answers. [...]

Statistical methods, apparently quite unrelated to each other, are in fact different aspects of the same central theory. [...] Statistical methods constitute a tool, often useful but sometimes abused.

[...] Biologists now use computers to proliferate statistical analyses, which have become an integral part of their work. But [...] sometimes an analysis is applied to data which contradict the assumptions of that analysis!

N. Gilbert, *Biometrical Interpretation* (Preface)

1. Introduction

The reason why some females of socially monogamous bird species seek extra-pair fertilizations is not clear, and it is to some extent important to assert whether the number of extra-pair young within and among broods follows some pattern of randomness which may support one of the possible explanations put forward by other researchers (Petrie and Kempenars, 1998).

The Poisson model may be looked at as a yardstick, compared to which other models are overdispersed or underdispersed. Moreover, the fact that the differential entropy $I(p) = \int_{-\infty}^{\infty} f_x(x) \ln[f_x(x)] dx$ attains its maximum over all f_x which are positive only for $x \geq 0$ and have finite expectation for the exponential density $f_x(x) = \frac{1}{\delta} e^{-\frac{x}{\delta}} I_{(0,\infty)}$ may be interpreted as meaning that the exponential distribution is the most uncertain among all distributions of nonnegative random variables with finite expectation. Therefore, for all $t > 0$, the Poisson process has the greatest λt -dimensional entropy in the interval $(0, t)$ among all homogeneous point processes with the same given intensity $\lambda = \frac{1}{\delta} > 0$ (Rényi, 1964). For that reason, Poisson streams of events are generally interpreted as representing unconstrained randomness, but this does not preclude the data from exhibiting other patterns of randomness.

Recently, Neuhäuser *et al.* (2001), using data about the number of extra-pair nestlings in broods of yellow warblers (*Dendroica petechia*), collected by Yezerinac *et al.* (1995), of hooded warblers (*Wilsonia citrina*), published by Stutchbury *et al.* (1994), and collared flycatchers (*Ficedula albicollis*), from Sheldon and Ellegren (1999), cf. columns 1–3 in Tables 1–3, challenged previous Poisson fits, and carried out a detailed investigation into the departure from a specific randomness hypothesis, namely an exchangeability hypothesis, expressed by a multivariate hypergeometric model, a form of mild dependence, with stationary probabilities. But in fact the

observed value of the G^2 goodness-of-fit statistic and the corresponding p -values show that their multivariate hypergeometric model is in general worse than the Poisson model. On the other hand, this complex multi-hypergeometric model and the cumbersome computations associated with it are irrelevant: as we shall explain in Section 4, to evaluate expected values only univariate hypergeometric margins are needed, and hence their exceedingly complex algorithm is unjustified.

The purpose of modelling is to achieve a generality that does not exist in actual observational results, in other words, to abstract from the data a general model that encompasses all possible sampling results, to extract knowledge from information. The built in dependence assumption that comes from sampling results rather than a sensible rationale is far from convincing, and the poor fit exhibited in Neuhäuser *et al.* (2001) ought to be expected, as will be further discussed in Section 4. But one of the merits of Neuhäuser *et al.* (2001) is to show that there is evident departure from the kind of randomness expressed by the hypergeometric model. We may therefore suspect that in those species where extra-mating has been studied, there are individual differences from female to female, as regards extra-mating strategies. We cannot conclude, however, whether this means different attitudes towards this kind of sexual behaviour, or whether different females have diverse extra-mating opportunities. We also ignore the distribution of successful extra-mating.

Models that are more closely related to the Poisson — even a crude Poisson model itself, with infinite support — are more appropriate than the hypergeometric model. The Poisson model may, as a side effect, incorporate the randomness derived from extra-mating opportunities, and on the other hand, as the Bernoulli filtered Poisson process is still Poisson, it may also incorporate extra-mating success.

The fact that both the Poisson and the multivariate hypergeometric provide exceedingly bad fit, in the generality of cases, does not mean that randomness has to be rejected, since other patterns of randomness may apply. We examine several alternative count models that stem from a basic Poissonian assumption in Section 2, and in Section 3 we investigate their goodness-of-fit to the number of extra-pair nestlings in broods of size s . Although much care must be taken in drawing conclusions based on such small samples, we shall observe that Zipf's law or the more general Zipf-Mandelbrot and discrete lognormal models account fairly well for the observed data.

The presentation of discrete models in Section 2 focuses on enhancing relations between different count models and how the unconstrained randomness of the Poisson model is progressively reduced by conditioning to find other well-known models such as the binomial and the hypergeometric, or by allowing for individual variability in negative binomial models. The logarithmic model, which can be derived as the limit of zero-truncated negative binomial distributions, may then be considerably generalized — in the sense that the parameter space is much wider — by right truncation, and the resulting Zipf or extended Zipf-Mandelbrot scaling laws exhibit an interesting "manicheist" pattern of randomness, that is appropriate to account for the equilibrium between conflicting tensions observed in many social

phenomena. A likely pattern for these females' behaviour, which seems to combine the eagerness for genetic diversity in the future with the present need to ensure male cooperation in raising the brood, which is supported by an excellent fit.

Most of the models we use are from the Katz family (Katz, 1965), whose probability mass functions exhibit nice recursive relations (Panjer, 1981). We also work out a discrete lognormal distribution that seems appropriate for skewed data; moreover, it generalizes Zipf–Mandelbrot laws for self-organizing phenomena that maintain a similar structure at various scales: scale has a bearing in structural organization and, on the other hand, the limit of discrete lognormal distributions when the location parameter goes to $-\infty$ is a Zipf–Mandelbrot distribution. The fit is in most cases almost as good as with the logarithmic or truncated-logarithmic distributions, and for very small samples it is even better. In the concluding Section 4, we comment on some points in the philosophy and practice of model fitting, stating some disagreements with Neuhäuser *et al.* (2001).

2. Count models and random patterns — a coordinate presentation of useful models

The classical presentation of count models is to use $X \sim \text{Hypergeometric}(N, n, p)$ for the number of successes in random sampling without replacement from a finite population (which implies a mild form of dependence, *exchangeability*), to use $Y \sim \text{Binomial}(n, p)$ as suitable for the number of successes in Bernoulli trials, i.e. when sampling with replacement (independent trials), and to adopt $W \sim \text{Poisson}(\lambda)$ as the limit of a sequence of binomial random variables under a “stable in average” condition, $E\left(X_n\right) = np_n \xrightarrow[n \rightarrow \infty]{} \lambda > 0$. Multivariate extensions with these margins are easily derived along similar paths.

This classical presentation has the advantage of immediate interpretation in terms of the most common sampling strategies. Moreover, when $n \ll N$ it is almost irrelevant to sample with or without replacement, and in that case $X \sim \text{Hypergeometric}(N, n, p)$ may be approximated by the simpler (two parameters instead of three) $Y \sim \text{Binomial}(n, p)$. On the other hand, for large values of n and $p \approx 0$, $W \sim \text{Poisson}(np)$ is a good approximation for $Y \sim \text{Binomial}(n, p)$, a further simplification, since we have to deal with only one parameter.

For some population studies, more sophisticated counting models are needed. In what follows, we present a coordinate description of the most important count models. Starting from the Poisson model, appropriate for unconstrained randomness, we obtain the binomial model by conditioning a Poisson summand on the observed sum, and the hypergeometric model by further conditioning a binomial summand on the observed sum, and comment on the constrained randomness brought in by

increased information (smaller variance). On the other hand, there may be room for individual variability, and mixing leads to negative binomial models.

At that stage, the concepts of underdispersion and overdispersion deserve some comments, as well as the recursive expressions used by Katz (1965) to organize discrete families of distributions, later explored by Panjer (1981) to obtain simple expressions for the density of randomly stopped sums. Binomial, Poisson and negative binomial are the non-degenerate solutions of Panjer's functional equation. If we relax the recursive relation, we obtain two other nondegenerate solutions, the Engen and the logarithmic models; the latter is appropriate when data show some tendency for clustering.

All these models can be further specialized by truncating the right tail. An important point is that the parameter space may be then considerably enlarged. The truncated logarithmic model, for instance, no longer has the restriction $\theta \in (0,1)$, and for $\theta = 1$ we have Zipf's model $P(X = k) \propto \frac{1}{k}$, $k = 1, \dots, N$, which accounts for "least effort", namely for equilibrium in accommodating conflicting needs.

Mandelbrot (1983) considerably generalized the useful part of Zipf's ideas by introducing location (λ) and shape (δ) parameters, $P(X = k) \propto \frac{1}{(k-\lambda)^{1+\delta}}$, $k = 1, \dots, N$.

Observe that a scale parameter is irrelevant in these "scaling laws", since it would be absorbed into the multiplicative normalizing constant. They are therefore useful for self-organizing phenomena, those that in some sense are scale invariant.

We introduce a non-trivial extension of Mandelbrot's class of models, by letting the shape $\theta(k)$ depend on k . The discrete lognormal random variable, with

probability mass function $p_k \propto \frac{1}{k} \exp\left(-\frac{(\ln k)^2}{2}\right) = \frac{1}{k^{1+\ln k}}$, $k = 1, 2, \dots$, is an important

model from that extended class. More generally, the discrete lognormal family, with

probability mass function $p_k \propto \frac{1}{k} \exp\left(-\frac{1}{2} \left(\frac{\ln k - \mu}{\sigma}\right)^2\right)$, $k = 1, 2, \dots$, $\mu \in R$, $\sigma > 0$, seems a

likely model for many discrete skewed data.

Alongside the description of models, we comment on parameter estimation and give appropriate references.

2a. Poisson Randomness and Count Data Models — Poisson and truncated Poisson, binomial and hypergeometric.

First order approximations are successfully used in all branches of mathematical modelling, and this is a natural explanation of the wide use of the Poisson model.

In fact, apart from stationarity and independence (in disjoint observation windows) of the increments, a third axiom rules out coincidental observations, and postulates local linearity, in the sense that the probability of a single observation is approximately proportional to the size of an infinitesimal observation window of area dA :

$$P(X_{A+dA} - X_A = 1) = \lambda dA + o(dA).$$

These three basic ideas are enough to establish that the number of occurrences in a region of size A is

$$X_A = \begin{cases} k & k = 0, 1, \dots \\ p_k = e^{-\lambda A} \frac{(\lambda A)^k}{k!} \end{cases}.$$

$E(X_A) = \lambda A$, so that the local linearity of the probability of a single occurrence in an infinitesimal window is mirrored by a global linearity for the expected values. In what follows we shall assume that the global area has unit size $A = 1$.

Poisson randomness is therefore appropriate whenever we feel that expected values may be considered "stable on average" (Recall that the Poisson arises as the limiting form of a sequence $X_n \sim \text{Binomial}(n, p_n)$ when the expected value $E(X_n) = n p_n \approx \lambda$). Examples are the number of sugar cane plants germinating per 10 m² plot, the number of nests per 100 m² plot in a pine wood and the number of butterflies caught per hour in a specific field. This is, indeed, the simpler mathematical translation of our faith in the regularity (and therefore predictability) of phenomena. It is natural to estimate λ by \bar{x} , which is in fact the maximum likelihood estimate of the rate λ .

There are, moreover, many other mathematical advantages of the Poisson model, for instance general Poisson random variables are the building blocks of infinitely divisible random variables, those that may be decomposed as sums of infinitesimal independent random summands. This is, of course, a strong modelling asset, since many observed phenomena are the result of infinitely many contributing effects. We shall pursue the matter no further here, but draw the reader's attention to the guidance that statistical knowledge may provide on the choice of models, that on one hand they must be useful — i.e. mathematically tractable — and on the other hand they must be appropriate, in the sense that they have built in properties that reflect known properties of the phenomena at hand.

A crude Poisson model may be useful, even though it has infinite support, and the phenomenon we wish to model is clearly finite. In order to use a chi-square goodness-of-fit test, care must be taken that $\sum_k O_k = \sum_k e_k$, and a standard practice is to consider the N -th class as a composite aggregate class, corresponding to $k \geq N$.

Another possibility is to truncate the right tail of the Poisson model, considering that for physical reasons $k > N$ is unobservable. The truncated model has probabilities

$$P_k = \frac{P(X=k)}{P(X \leq N)} = \frac{1}{\sum_{j=1}^N \frac{\lambda^j}{j!}} \frac{\lambda^k}{k!}, \quad k=0,1,\dots,N.$$

In this case, in order to estimate λ (maximum likelihood) from the data, we must solve

$$\sum_{j=0}^N (\bar{x} - j) \frac{\lambda^j}{j!} = 0,$$

which is straightforward using Cohen's (1961) tables. Moore (1954) suggested the estimator

$$\tilde{\lambda} = \sum_j \frac{X_j}{m},$$

where m stands for the number of observed values that are less than $N-1$, which we shall use since it is easier to compute and is an unbiased estimator of λ .

The success of the Poisson model derives from striking "conservative" mathematical properties: if $X \sim \text{Poisson}(\lambda)$, $Y \sim \text{Poisson}(\mu)$ and X and Y are independent, then $X+Y \sim \text{Poisson}(\lambda+\mu)$, and thus incorporation of new information can be achieved, in many situations, through a simple modification of the parameter, the structure of the model not being modified. A similar conservative result holds for binomial filtering (or thinning): if $X \sim \text{Poisson}(\lambda)$ goes through a binomial filter with probability of success p , the resulting model is $X_j \sim \text{Poisson}(p\lambda)$.

From the result on the addition of independent Poisson random variables, by conditioning one of the summands on the observed value of the sum we obtain a binomial (more information yields a model with smaller variance): if $X_k \sim \text{Poisson}(\lambda_k)$, for $k=1,\dots,r$, are independent,

$$X_k \mid \sum_{j=1}^r X_j = s \sim \text{Binomial}(s, \pi_k), \quad \pi_k = \frac{\lambda_k}{\sum_{j=1}^r \lambda_j}.$$

More generally, if $\mathbf{X} = (X_1, \dots, X_N)$ is multi-Poisson, the conditional distribution of \mathbf{X} given that $X_1 + \dots + X_N = n$ is multinomial; this is the basis for the analysis of count data with general chi-square statistics comparing observed counts with the corresponding expected values, using either the classical chi-square statistic X_N^2 , the likelihood ratio statistic G^2 , or general power divergence statistics (Cressie and Read, 1984).

A similar result holds for independent binomial summands: if $X_i \sim \text{Binomial}(n_i, p)$, and $N = \sum_{j=1}^r n_j$, then $X_1 + \dots + X_r \sim \text{Binomial}(N, p)$ and $X_i \mid \sum_{j=1}^r X_j = s \sim \text{Hypergeometric}(N, s, \pi_i)$, where $\pi_i = \frac{n_i}{N}$. As this last relation may have

been one (unstated) reason for Neuhäuser *et al.*'s (2001) choice of hypergeometric randomness, we also investigate a binomial fit, since we feel more inclined towards an unconditional Poisson model than towards an unconditional binomial model. We use the maximum likelihood estimate of the binomial parameter p , the individual probability of success in the sequence of Bernoulli trials, which in this case coincides with the minimum chi-square and the method of moments estimates.

The above results show that the binomial and the hypergeometric (and their multivariate extensions), usually derived as count models associated with simple random sampling, with or without replacement, respectively, may also arise in connection with Poisson randomness. In the following sections, we shall show that other important discrete distributions are closely related to the Poisson law.

As we observed above, binomial filtering (or *thinning*) of the Poisson distribution results in a filtered Poisson distribution. For illustration purposes, consider the following situation. Let us assume that the number of times a female encounters another male (not her social mate) may be modelled as $X \sim \text{Poisson}(\lambda)$. Assume that the probability that copulation ensues from each encounter is α , and that the probability that a copulation will eventually lead to a (extra-pair) nestling is β , independently of what happens in any other occurrence. Thus the number of extra-pair nestlings is modelled by the filtered $X_f \sim \text{Poisson}(\lambda\alpha\beta)$.

Assume also that the number of times the female copulates with her social mate can be modelled as $Y \sim \text{Poisson}(\mu)$, and that the probability that a copulation will eventually lead to a nestling is π , so that the number of non extra-pair nestlings is $Y_f \sim \text{Poisson}(\mu\pi)$.

Accepting independence between X and Y (and, for honesty sake, this might not be true in the real world), then, if we draw a sample of broods and consider a specific brood size of s (that is, conditioning on the fact that $X_f + Y_f = s$), the model that describes the number of extra-pair nestlings will be $X_f \sim \text{Binomial}\left(s, \frac{\lambda\alpha\beta}{\lambda\alpha\beta + \mu\pi}\right)$. Thus it seems worthwhile to investigate a binomial fit in this case study (with disappointing results, as we shall see).

A final observation on Poisson, binomial and hypergeometric models: we derived the binomial distribution by conditioning a Poisson summand on the observed value of the sum, and the hypergeometric law by conditioning a binomial summand on the observed value of the sum (assuming independence of summands, and equal parameter p in the latter case). Comparing the variances of $X \sim \text{Poisson}(\lambda)$, $Y \sim \text{Binomial}(n, p)$ and $W \sim \text{Hypergeometric}(N, n, p)$, with equal means, $np = \lambda$,

$$\text{var}(W) = np(1-p) \frac{N-n}{N-1} < \text{var}(Y) = np(1-p) < \text{var}(X) = \lambda = np.$$

The Poisson model is mathematically simpler — one single parameter, and the family of Poisson random variables is closed under summation and binomial filtering —, but on the other hand the hypergeometric model has more information.

As we have seen, the Poisson model accounts for unconstrained randomness. The binomial model, as an alternative to the usual presentation via counting successes in sampling with replacement, has been obtained as the posterior distribution of a Poisson summand conditioned on the value of the sum — extra information that diminishes randomness, bringing in more information, and hence a smaller variance. The hypergeometric model, aside from being the model for counting successes in sampling without replacement, has been shown to be the model for a binomial summand conditioned on the observed sum, and increased information once again decreases dispersion and further constrains randomness.

If the sampling fraction $\frac{n}{N}$ is low, i.e. $n \ll N$, $\text{var}(W) \approx \text{var}(Y)$: the probability of including the same element more than once in the sample, when sampling with replacement, is very low, and sampling with or without replacement is an idle question, since both sampling strategies give the same amount of information. On the other hand, if $p \approx 0$, $\text{var}(Y) \approx \text{var}(X)$, a clear indication that the Poisson law is a good approximation of the binomial when the variance is close to the expectation.

2b. Negative binomial as a gamma mixture of Poisson distributions; truncated negative binomial.

The geometric model — and, more generally, the negative binomial model of which the geometric is not but a special case — are interesting alternatives to the Poisson model, when we need to account for individual variability. In the case at hand, this is individual variability in the females' patterns of sexual behaviour.

Let us suppose that the appropriate model for each female is $X \sim \text{Poisson}(\lambda)$

$$X = \begin{cases} k & k = 0, 1, \dots \\ p_k = e^{-\lambda} \frac{\lambda^k}{k!} \end{cases}$$

($\text{var}(X) = E(X) = \lambda$). As regards the whole population, we may model λ as a random variable $\Lambda \sim \text{Exponential}(\delta)$, i. e. with distribution function

$$F_{\Lambda}(\lambda) = \left[1 - e^{-\frac{\lambda}{\delta}} \right] I_{(0, \infty)}(\lambda), \quad \delta > 0.$$

In this hierarchical model,

$$E(X_{ii}) = E[E(X | \Lambda)] = E(\Lambda) = \delta$$

and

$$\text{var}(X_{ii}) = E[\text{var}(X | \Lambda)] + \text{var}[E(X | \Lambda)] = E(\Lambda) + \text{var}(\Lambda) = \delta + \delta^2.$$

Thus we obtain a model with $\text{var}(X_{ii}) > \text{var}(X)$, which clearly accounts for higher diversity.

The mixture distribution can easily be derived: $f_{\Lambda}(\lambda) = \frac{1}{\delta} e^{-\frac{\lambda}{\delta}} I_{(0, \infty)}(\lambda)$, hence

$$\begin{aligned} P(X_{ii} = k) &= \int_0^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} \frac{1}{\delta} e^{-\frac{\lambda}{\delta}} d\lambda = \frac{1}{\delta k!} \int_0^{+\infty} \lambda^k e^{-(1+\frac{1}{\delta})\lambda} d\lambda = \\ &= \frac{1}{\delta k!} \int_0^{+\infty} \left(\frac{y}{1+\frac{1}{\delta}}\right)^k e^{-y} \frac{dy}{1+\frac{1}{\delta}} = \frac{1}{1+\delta} \left(\frac{\delta}{1+\delta}\right)^k, \quad k = 0, 1, \dots \end{aligned}$$

and thus $X_{ii} \sim \text{Geometric}\left(\frac{1}{1+\delta}\right)$. For that reason, some authors consider the geometric model (and more generally $Y \sim \text{NegativeBinomial}\left(\nu, \frac{1}{1+\delta}\right)$ with index $\nu > 0$, i.e.

$P_k = \binom{k+\nu-1}{\nu-1} \left(\frac{1}{1+\delta}\right)^{\nu} \left(\frac{\delta}{1+\delta}\right)^k$, $k = 0, 1, \dots$, which may be derived using similar arguments with the Gamma with index ν as mixing distribution) as a "more dispersed" Poisson.

In the general case, when both parameters are unknown, the easiest estimation method is the method of moments, equating sample and population means and variances, respectively. As we are working with the negative binomial located at 0, $E(Y) = \frac{\nu(1-p)}{p}$ and $\text{var}(Y) = \frac{\nu(1-p)}{p^2}$, therefore

$$\tilde{p} = \frac{\bar{x}}{s} \quad \text{and} \quad \tilde{\nu} = \frac{\bar{x}^2}{s^2 - \bar{x}}$$

As an alternative, the *Mean-and-Zero-Frequency* method equates the observed and expected number of zero values, and the sample mean and population mean. Thus

$$f_0 = (p^*)^{\nu^*} \Rightarrow \nu^* = \frac{\ln f_0}{\ln p^*}$$

and

$$\tilde{v} \cdot \frac{1-p^*}{p^*} = \bar{x},$$

from which we get $\frac{\ln f_0}{\ln p^*} \frac{1-p^*}{p^*} = \bar{x} \iff \frac{1-p^*}{p^* \ln p^*} = \frac{\bar{x}}{\ln f_0} \iff \frac{p^*}{\ln(1+p^*)} = -\frac{\bar{x}}{\ln f_0}$, where

$P^* = \frac{1-p^*}{p^*}$. Piegorsch (1990) recommends minimum chi-square estimation, on the grounds that it is slightly less biased than maximum likelihood or method of moments estimation, provided that the sample size is greater than 20. For a discussion of the relative merits of minimum chi-square and maximum likelihood criteria, cf. Berkson (1978) and the ensuing discussion.

Although the negative binomial distribution is infinitely divisible, and hence a non-integer index may be used, we shall limit our investigation to the classical case, and use an integer approximation for \tilde{v} . Whenever \tilde{v} is too small, we use the shifted logarithmic model (Fisher *et al.*, 1943, derived the logarithmic model as the limit of zero truncated negative binomial distributions).

As in the Poisson case, the support of the negative binomial is infinite (as happens for all infinitely divisible laws), and we may derive finite support models using truncation. There are no closed forms for parameter estimation in the case of truncated models. For instance, for the geometric model truncated to the right of s ,

with $p_k = \frac{p(1-p)^k}{1-(1-p)^{s+1}}$, $k=0, \dots, s$, the maximum likelihood estimate \hat{p} satisfies the equation

$$\frac{n}{p} - \sum_{j=1}^n \frac{x_j}{1-p} - \frac{n(s+1)(1-p)^s}{1-(1-p)^{s+1}} = 0,$$

which, however, is easily handled in the example that we work out.

2c. Katz family of discrete random variables.

If $W \sim \text{Poisson}(\lambda)$, the recursive expression $p_{k+1} = \frac{\lambda}{k+1} p_k = \left(\alpha + \frac{\beta}{k+1}\right) p_k$ holds for $k=0, 1, \dots$, with $\alpha=0$ and $\beta=\lambda$.

If $Y \sim \text{Binomial}(n, p)$, we may write $p_{k+1} = \frac{n-k}{k+1} \frac{p}{1-p} p_k = \left(\alpha + \frac{\beta}{k+1}\right) p_k$, with $\alpha = \frac{p}{p-1}$ and $\beta = \frac{(n+1)p}{1-p}$, for $k=0, \dots, n-1$.

On the other hand, if $X \sim \text{NegativeBinomial}(\nu, p)$, i.e. $P(X = k) = \binom{k+\nu-1}{\nu-1} p^\nu (1-p)^k$, for $k = 0, 1, \dots$ we have $p_{k+1} = \frac{(\nu+k)(1-p)}{k+1} p_k = \left(\alpha + \frac{\beta}{k+1}\right) p_k$, with $\alpha = 1-p$ and $\beta = (\nu-1)(1-p)$.

It is interesting to observe that the models mentioned above are the only ones whose probability mass function satisfies the recursive relation

$$\frac{p_{n+1}}{p_n} = \alpha + \frac{\beta}{n+1} \quad n = 0, 1, \dots$$

(Panjer's recursive expression, which had been used by Katz (1965) in the equivalent form $\frac{p_{n+1}}{p_n} = \frac{\alpha + \beta \frac{n}{n+1}}{n+1}$, $n = 0, 1, \dots$, to classify families of discrete distributions).

In fact, multiplying $p_{n+1} = \alpha + \frac{\beta}{n+1} p_n$ by s^{n+1} and adding for $n \geq 0$, we obtain the differential equation

$$\frac{P'(s)}{P(s)} = \frac{\alpha + \beta}{1 - \alpha s},$$

for the probability generating function $P_{X_{\alpha, \beta}}(s) = \sum_n p_n s^n$, whose absolutely monotone solutions are the probability generating functions

1. $P_{X_{0,0}}(s) = 1$, case $\alpha = \beta = 0$, i. e. $X_{0,0} = 0$, the degenerate random variable with unit mass at 0.
2. $P_{X_{0,\beta}}(s) = e^{\beta(s-1)}$ if $\alpha = 0$ (and necessarily $\beta > 0$), and therefore $X_{0,\beta} \sim \text{Poisson}(\beta)$.
3. $P_{X_{\alpha,\beta}}(s) = \left(\frac{1-\alpha}{1-\alpha s}\right)^{\frac{\alpha+\beta}{\alpha}}$; $X_{\alpha,\beta} \sim \text{NegativeBinomial}\left(\frac{\alpha+\beta}{\alpha}, 1-\alpha\right)$ if $\alpha \in (0, 1)$ and $\alpha + \beta > 0$.
4. $P_{X_{\alpha,\beta}}(s) = \left(1 - \frac{\alpha}{\alpha-1} + \frac{\alpha}{\alpha-1} s\right)^{-\left(1+\frac{\beta}{\alpha}\right)}$; if $\alpha < 0$, $\frac{\alpha}{\alpha-1} = p > 0$ and we recognize the probability generating function of $X_{\alpha,\beta} \sim \text{Binomial}\left(-1 - \frac{\beta}{\alpha}, \frac{\alpha}{\alpha-1}\right)$ with $-1 - \frac{\beta}{\alpha} \in \mathbb{N}^+$.

Cf. Rolski *et al.* (1999) for an alternative proof.

An important point is that if $W \sim \text{Poisson}(\lambda)$, $\frac{\text{var}(W)}{E(W)} = \frac{\lambda}{\lambda} = 1$, a good reason to use the Poisson distribution as a yardstick in what regards dispersion.

On the other hand, if $Y \sim \text{Binomial}(n, p)$, $\frac{\text{var}(Y)}{E(Y)} = \frac{np(1-p)}{np} < 1$, Y is underdispersed.

We have seen that the geometric random variable, allowing for individual diversity, is more dispersed than the Poisson, and more generally if $X \sim \text{NegativeBinomial}(v, p)$, $\frac{\text{var}(X)}{E(X)} = \frac{1}{p} > 1$, X is *overdispersed*.

Panjer's family may be considerably extended by using weights,

$$\frac{p_{n+1}}{p_n} = \alpha + \frac{\beta}{(n+1)\gamma_n} \quad n = 0, 1, \dots$$

(Velosa, 2003), or if we relax the defining condition, by allowing

$$\frac{p_{n+1}}{p_n} = \alpha + \frac{\beta}{n+1} \quad n = k, k+1, \dots,$$

i.e. the relation holds only for $n \geq k > 0$ (Hess *et al.*, 2002). The case $k = 1$ has been studied by Sundt and Jewell (1981) and by Willmot (1987). The results are worth recording:

If $p_0 = 0$ and $\frac{p_{n+1}}{p_n} = \alpha + \frac{\beta}{n+1}$, $n = 1, 2, \dots$, multiplying by s^{n+1} and summing for $n \geq 1$, we obtain the differential equation

$$(1 - \alpha s)P'(s) - (\alpha + \beta)P(s) = p_1,$$

whose non-degenerate absolutely monotone solutions are:

$$1. P_{x_{\alpha, \beta}} = \frac{1 - (1 - \alpha s)^{\frac{\alpha + \beta}{\alpha}}}{1 - (1 - \alpha s)^{\frac{\alpha + \beta}{\alpha}}}, \text{ the probability generating function of}$$

(a) the zero-truncated negative binomial distribution if $\alpha \in (0, 1)$ and $\beta > -\alpha$;

(b) the zero truncated Engen generalized negative binomial distribution, if $\alpha \in (0, 1]$ and $\beta \in (-2\alpha, -\alpha)$;

(c) *Logarithmic*(α) if $-\beta = \alpha \in (0, 1)$; the logarithmic random variable may be obtained either as the limit of zero truncated Engen's extended negative binomial random variables, or of zero truncated negative binomial random variables, with index (shape) parameter going to 0. Each of these models has a prestigious history of applications to population studies, since they were introduced by Fisher *et al.* (1943) and by Engen (1974).

$$2. P_{x_{\alpha, \beta}} = \frac{e^{-\beta}}{1 - e^{-\beta}} (e^{\beta s} - 1) \text{ (the zero truncated Poisson, if } \alpha = 0 \text{ and } \beta > 0 \text{);}$$

$$3. P_{x_{\alpha, \beta}} = \frac{(1 - \alpha s)^{\binom{1 + \beta}{\alpha}} - 1}{(1 - \alpha)^{\binom{1 + \beta}{\alpha}} - 1} \text{ (the zero truncated binomial, if } \alpha < 0 \text{ and } -\frac{\beta}{\alpha} \in \mathbb{N} \text{).}$$

More generally, the probability mass function of the random variable X satisfies $\frac{p_{n+1}}{p_n} = \alpha + \frac{\beta}{n+1}$, $n = 1, 2, \dots$, if the corresponding probability generating function $\tilde{P}_X(s)$ can be written

$$\tilde{P}_X(s) = \tau + (1 - \tau)P_{X_{\alpha, \beta}}(s),$$

where $P_{X_{\alpha, \beta}}$ is one of the probability generating functions in the preceding enumeration, and $\tau \in \left[\frac{P(0)}{P(0)-1}, 1 \right)$. Hence, the ratio of observed frequencies $\frac{o_{k+1}}{o_k} \approx \frac{p_{k+1}}{p_k}$ may be of some guidance on the choice of an appropriate model, by fitting $\frac{(k+1)o_{k+1}}{o_k} \approx \hat{\alpha}k + (\hat{\alpha} + \hat{\beta})$. Considerations on dispersion may give further insight on which model to prefer.

2d. Clustering and the logarithmic model; truncated logarithmic and Zipf-Mandelbrot models.

If none of the nestlings is extra-pair we cannot presume that the female did not mate with other males aside from her social partner. It may be that she has a very promiscuous behaviour, but extra-pair mating turns out to be infertile. On the other hand, the availability of extra-pair males may also have a bearing on the number of extra-pair nestlings. Promiscuous mating is certainly an opportunistic behaviour, perhaps a delicate equilibrium between compulsion to bring in genetic diversity to the progeny, and the need to maintain a monogamic social organization that favours the raising of nestlings.

The fact that we cannot exclude promiscuous mating behaviour when there are no extra-pair young in the nest is to some extent compensated by clustering when we do find "bastard" progeny. This is a common situation in biology. For instance, some female insects lay eggs on leaves appropriate for feeding their maggots. Of course the fact that we do not find any eggs on a particular leaf does not mean that none of these females landed on that particular leaf. On the other hand, we often find several eggs on the same leaf, and "blindness" in what regards 0 is compensated by some clustering. We now describe an appropriate model for this kind of phenomena: for any $\theta \in (0, 1)$, we have $\ln(1 - \theta) = -\sum_{k=1}^{\infty} \frac{\theta^k}{k}$. Therefore $p_k = -\frac{1}{\ln(1 - \theta)} \frac{\theta^k}{k}$, $k = 1, 2, \dots$, is the probability mass function of a random variable $W_\theta \sim \text{Logarithmic}(\theta)$, with support $k = 1, 2, \dots$, since it sums to 1:

$$W_\theta = \begin{cases} k & k = 1, 2, \dots \\ p_k = -\frac{1}{\ln(1-\theta)} \frac{\theta^k}{k} & (0 < \theta < 1) \end{cases}$$

As $\frac{\text{var}(W_\theta)}{E(W_\theta)} = \frac{1 + \frac{\theta}{\ln(1-\theta)}}{1-\theta}$, for $\theta = 1 - \frac{1}{e}$ we have $E(W_\theta) = \text{var}(W_\theta)$ as in the Poisson model.

For $\theta \in (1 - \frac{1}{e}, 1)$ the random variable W_θ is overdispersed, and for $\theta \in (0, 1 - \frac{1}{e})$ we observe that W_θ is an interesting infinite support underdispersed model.

Fisher *et al.*'s (1943) derivation of the logarithmic random variable shows that it is the weak limit as $\nu \rightarrow 0$ of a sequence of zero-truncated negative binomial random variables with index ν , cf. Johnson *et al.* (1992, p. 286), and hence it is also related to the Poisson randomness model. They have shown that if in a batch the number of species represented by exactly one individual is n_1 then, denoting $\alpha = \frac{n_1}{\theta}$ the *index of diversity*, the number of species represented by k individuals is approximated by $\frac{\alpha \theta^k}{k}$, $k = 2, 3, \dots$. The logarithmic distribution is considered a good fit for count data whenever there is underlying clustering — number of bacteria per colony, number of inhabitants per house, or number of animals per litter, for instance.

The maximum likelihood estimate of θ is the solution $\hat{\theta}$ of

$$-\frac{\theta}{(1-\theta)\ln(1-\theta)} = \bar{x} = \frac{1}{n} \sum_{k=1}^n x_k,$$

which is easily computed using, for instance, the Newton–Raphson method.

By truncating to the right of s , we obtain the truncated logarithmic model

$$W^* = \begin{cases} k & k = 1, \dots, s \\ \frac{1}{\sum_{j=1}^s \frac{\theta^j}{j}} \frac{\theta^k}{k} & \end{cases}$$

The maximum likelihood estimator of θ is the solution of

$$\frac{\theta(1-\theta^s)}{(1-\theta) \sum_{j=1}^s \frac{\theta^j}{j}} = \bar{x},$$

which may be computed from Patil and Wani's (1965) tables. But for our purposes it is simpler to equate sample and population moments, obtaining the explicit estimate

$$\tilde{\theta} = \frac{m'_3 - (s+2)m'_2 + (s+1)m'_1}{m'_3 - sm'_2},$$

where $m'_k = \frac{1}{n} \sum_{j=1}^n x_j^k$ denotes the k -th sample moment.

We shall, naturally, shift these random variables to 0, $X = W - 1$ and $X^* = W^* - 1$; this is equivalent to taking $x_k + 1$ instead of x_k as observations.

An important remark: in this truncated case it is no longer necessary to consider that the parameter space is $\Theta = (0,1)$. In fact, $\{p_k\}_{k=1}^s$, $p_k = \frac{\theta^k}{k \sum_{j=1}^s \frac{\theta^j}{j}}$, is a probability

mass function for any $\theta > 0$, as in Table 1 ($s=4,5$), Table 2 ($s=2,3,5$) and Table 3 ($s=4,5$). In particular, if $\theta=1$ we get Zipf's law $\{p_k = \frac{c}{k}\}_{k=1}^s$, where

$\frac{1}{c} = \sum_{k=1}^s \frac{1}{k} \approx \ln s + \gamma$ ($\gamma \approx 0.577$ is Euler's constant), which Zipf claimed was tied to the 'principle of least effort' in the sense that it would model phenomena shaped by conflicting interests (Zipf's primary examples are verbal communication — the need to be socially understood constraining the use of rich personal vocabulary — and the size of cities, seen as the result of the attraction/repulsion feelings that large human settlements exert upon different individuals).

More general "scaling laws" or "power laws" $\left\{p_k = \frac{c}{(k-\lambda)^{1+\rho}}\right\}_{k=1}^s$, where

$\frac{1}{c} = \sum_{k=1}^s \frac{1}{(k-\lambda)^{1+\rho}}$, with location parameter λ and shape parameter $\rho > 0$, have been

used by Mandelbrot (1983), namely to model self-organizing phenomena. Gell-Mann (1994, p. 91–94) is an excellent naive introduction to the matter, with enlightening comments on the role of parameters. When $\rho > 0$, s may be ∞ ; in the simpler case

$\lambda = 0$, $\left\{p_k = \frac{1}{\zeta(1+\rho)} \frac{1}{k^{1+\rho}}\right\}_{k \geq 1}$, where $\zeta(\cdot)$ is Riemann's zeta function. Some authors

(Seal, 1952; Adamic, 2001) regard this *zeta distribution* as a *discrete Pareto* law. The maximum likelihood estimator $\hat{\rho}$ of the shape parameter ρ satisfies

$$\frac{\zeta'(\rho+1)}{\zeta(\rho+1)} = -\frac{1}{n} \sum_{k=1}^n \ln X_k$$

(Seal, 1952). Alternatively, equating the population and sample means we obtain the estimator $\tilde{\rho}$, that satisfies $\frac{\zeta(\rho)}{\zeta(\rho+1)} = \bar{X}$, which may be easily solved using the tables provided by Moore (1956).

Power laws play an important role in modelling dynamic phenomena where scale effects contribute to self-organization. With the sole exception of the hypergeometric distribution, all the other models related to Poisson randomness that we have described are generalized power laws, cf. Johnson *et al.* (1992, p. 81). We believe this is one more reason to prefer any of the other models (moreover, in the context of initial Poisson randomness, the hypergeometric arises from a double addition/conditioning scheme that seems farfetched in this context of modelling the number of extra-pair young in each nest). Pérez-Abreu (1991) has shown that under very mild assumptions the Poisson arises as the limit of power series distributions, further justifying for the central role we have chosen for the Poisson model:

2e. Discrete lognormal model.

Discrete probability mass functions of the form $\left\{ p_k = \frac{1}{C(\theta)} \frac{1}{k^{1+\theta(k)}} \right\}_{k \geq 1}$ provide an even wider choice of models, among which Zipf–Mandelbrot’s, corresponding to $\theta(k) = \rho$, is just the easiest to deal with. The generalized case, on the other hand, may provide appropriate models where location may influence scale effects.

A particular choice $\theta(k) = \ln(\sqrt{k})$ leads to $p_k \propto \frac{1}{k^{1+\ln\sqrt{k}}}$, $k = 1, 2, \dots$, i.e. to the *discrete lognormal* random variable

$$X_{0,1} = \begin{cases} k & k = 1, 2, \dots \\ p_k = \frac{1}{C(0,1)k} \exp\left(-\frac{(\ln k)^2}{2}\right) \end{cases}$$

where $C(0,1) = \sum_{k=1}^{\infty} \frac{1}{k} \exp\left(-\frac{(\ln k)^2}{2}\right)$ is the normalizing constant.

More generally, the discrete lognormal law with parameters μ and σ

$$X_{\mu,\sigma} = \begin{cases} k & k = 1, 2, \dots \\ p_k = \frac{1}{C(\mu,\sigma)k} \exp\left[-\frac{1}{2}\left(\frac{\ln k - \mu}{\sigma}\right)^2\right] \end{cases}$$

with $C(\mu, \sigma)$ the appropriate normalizing constant, seems a likely model for discrete skew populational data. The parameters μ and σ may be estimated by numerical maximization of the likelihood function.

Further observe that when $\mu \rightarrow -\infty$, the discrete lognormal law may be considered a Zipf-Mandelbrot law with shape parameter $\rho = \frac{\mu}{\sigma}$. In fact, if $\ln k \ll |\mu|$,

$$p_i = \frac{1}{C(\mu, \sigma)k} \exp \left[-\frac{1}{2} \left(\frac{\ln k - \mu}{\sigma} \right)^2 \right]$$

$$\propto \frac{1}{k} \exp \left[-\frac{\ln k (\ln k - 2\mu)}{2\sigma^2} \right] \approx \frac{1}{C(\mu, \sigma)} \frac{1}{k^{\frac{1-\mu}{\sigma}}}$$

3. Goodness-of-fit

For the data provided in Tables 1–6 — Tables 1 and 2 for yellow warblers (*Dendroica petechia*) studied by Yezerinac *et al.* (1995), Tables 3 and 4 for hooded warblers (*Wilsonia citrina*) provided by Stutchbury *et al.* (1994), Tables 5 and 6 for collared flycatchers (*Ficedula albicollis*) studied by Sheldon and Ellegren (1999) —, and using G^2 as a comparison criterion, we evaluate the goodness-of-fit for the relevant models described above: the more traditional hypergeometric, Poisson, truncated Poisson, binomial, negative binomial (in general geometric), truncated geometric, in the odd-numbered tables; and the less common logarithmic, truncated logarithmic, truncated Zipf-Mandelbrot and discrete lognormal models in the even-numbered tables. We present asymptotic theory, since this way our results may be checked easily using a spreadsheet, since they do not depart substantially from exact results using exact algorithms such as those incorporated in StaXact.

For comparison purposes, our Tables 1 to 6 exhibit some overlap — i.e., for the hypergeometric and Poisson models — with Tables 2–4 in Neuhäuser *et al.* (2001); our goal is to show that the simple univariate hypergeometric model achieves exactly the same results as their multi-hypergeometric model. We omitted brood size $s=1$ (hooded warblers) and $s=2$ (collared flycatchers), and in fact should have refrained from analyzing other cases: statistical analysis with exceedingly small sample sizes cannot be recommended. But this is a case study to assess the potential interest of these models, not a statistical data analysis aimed at drawing conclusions from the data. Brood size $s=4$ (collared flycatchers), and $s=5$ (hooded warblers) are maintained in our analysis, as paradigmatic cases, important for the discussion in Section 4 and comparison with Neuhäuser *et al.*'s (2001) views.

For the Poisson, logarithmic and geometric models, which have infinite support, we give results both using the last class as an “aggregate” class, and using the

corresponding truncated models. We use truncated negative binomial models only for estimated index $\nu = 1$ (geometric), since parameter estimation for the general truncated negative binomial is unreliable. Anyway, only in the case of yellow warblers, $s = 4$ (Table 1) do we have the unexpected index estimate $\nu^* \approx 10$; in all the other cases (we would have $\nu^* \approx 5$ for hooded warblers, $s = 2$ in Table 2; but for this case, since we need to estimate two parameters, the number of degrees of freedom would be 0), the estimate is either $\nu^* \approx 0$ — in which case the logarithmic model seems appropriate — or $\nu^* \approx 1$.

As we have seen in Section 2, the logarithmic random variable is the limit of zero-truncated negative binomial random variables with index going to zero. Hence, whenever the index estimate of the negative binomial model is near zero we proceed directly to a logarithmic model fit.

Parameter estimation has been done either by maximum likelihood estimation or the simpler — as a rule, unbiased — estimation described in the appropriate subsection of Section 2; we emphasize simple methods, and this is the reason why we described them beforehand, providing a fair choice to the reader. Computer intensive methods, bootstrap estimates, and other sophisticated methods may be preferable in many situations — but in all cases the researcher must ponder whether the sample size calls for such investment. In addition to the G^2 observed value, degrees of freedom and p -values, we present parameter estimates (though the index of the negative binomial may be fractional — $\frac{pe^u}{1-(1-p)e^u}$ is an infinitely divisible characteristic

function, and thus $\left(\frac{pe^u}{1-(1-p)e^u}\right)^v$ is a characteristic function for any $\nu > 0$ —, we shall use the integer part of ν^* as the binomial index), and expected frequencies, for immediate visual comparison with the observed values.

There is no need, in the hypergeometric case, for explicit evaluation of the parameter estimate $\hat{p} = \frac{r}{N}$; but this does not preclude the fact that it is implicitly used in the computation of expected values under the null hypothesis, and that the appropriate number of degrees of freedom is $(s+1) - 2 = s - 1$.

As regards the hypergeometric, the computation is done with the univariate hypergeometric marginal model, which provides exactly the same expected values, with much less computational effort than in Neuhäuser *et al.* (2001). In fact, only expected values of the univariate model are needed, and it seems farfetched to compute univariate moments using the complicated algorithm for the multi-hypergeometric model given in the appendix to Neuhäuser *et al.* (2001); this matter will be further discussed in Section 4.

The hypergeometric and the binomial model give poor fits, even worse than the crude Poisson model criticized by Neuhäuser *et al.* (2001), or the negative binomial that could account for some individual variability. Thus the classical count models seem useless in the present context.

On the other hand, the logarithmic and truncated logarithmic (power laws, in the strict sense) provide in general excellent fits, the more general Zipf–Mandelbrot scaling laws and the skewed discrete lognormal distribution seem interesting candidates to elicit the randomness exhibited by these data. This clear pattern, exhibited by the comparison of Tables 1 and 2, Tables 3 and 4, and Tables 5 and 6, has been confirmed using StaXact exact algorithms to compute observed values and p -values.

It therefore seems plausible to consider the working hypothesis that eagerness for genetic diversity and the need to have male cooperation in raising the brood are two polarities that generate an elaborate equilibrium in these females mating behaviour.

4. Concluding Remarks

The purpose of the present paper, aside from contributing to the understanding of some female birds' remarkable sexual behaviour, is twofold:

1. To stress that the discrete models presented in elementary and intermediate Probability and Statistics courses (hypergeometric, binomial and negative binomial, and Poisson) are insufficient to model the rich randomness patterns that arise when counting biostatistical phenomena. Johnson *et al.*'s (1992) presentation, on the other hand, which in many cases uses Gauss hypergeometric generating functions, is too specialized for most practical users. Our choice would be to invest in a coordinate approach to count data models as outlined in Section 2, stressing patterns of randomness. One point we would like to emphasize once again: truncation (and indeed other techniques aimed at restricting the support) may be compensated by an interesting stretching of the parameter space (in our examples, although the parameter space for the *Logarithmic*(θ) is $\Theta = (0,1)$, for the right truncated logarithmic — power laws — the parameter space $\Theta = (0, \infty)$ is much wider).
2. To underline the importance of clearly understanding statistical concepts, in order to avoid misconceptions and mishandling of statistical algorithms. This is worrying mostly because errors tend to have a fertile progeny following their publication. For that reason, we make explicit some points detailed in Marques and Pestana (2002):
 - Neuhäuser *et al.* (2001) claim that the Poisson is not a plausible model, since it attributes positive probabilities to unobservable values, and since the percentage of extra-pair nestlings is high, this not a rare event. The first objection — pointing out that impossible configurations have positive probabilities in a model — is a statistical misconception (truncation of the right tail solves part of the problem). Sharing their views about modelling would prevent us from using almost all useful (i.e., mathematically tractable) statistical models, most of which have infinite support, while all

our observations are finite. "*A little inaccuracy saves tons of explanations*", wrote Saki; in our view, this is the role of models: a good model is general — since there is no Science of particular facts — free from concrete details, although its properties reflect the striking patterns of the phenomena that it represents; and this inbuilt simplification makes it mathematically tractable.

Regarding rare events, notice that the same model can be generated by very diverse mechanisms, and that the point is whether or not it provides a good fit to the available data. The fact that the Poisson distribution can be derived as the limit of binomial distributions under a condition of stability in the mean that has an appealing interpretation as a limit law of rare events cannot be taken as a crude statement that it applies only to rare events; and it should be noted that "rarity" concerns the binomial parent distribution, not the limiting Poisson law: $np_n \xrightarrow{n \rightarrow \infty} \lambda \in (0, \infty)$, and hence the mean value $E(X) = \lambda$ can be very large!

- Although the question addressed in Neuhäuser *et al.* (2001), as they clearly stated, was not "*to model observed distributions, but to provide a pattern of randomness with which observed patterns can be compared*", their multi-hypergeometric distribution seems quite unrealistic. It comes as no surprise that Neuhäuser *et al.* (2001) obtained very low p -values using G^2 goodness-of-fit statistics for their hypergeometric fitting. In fact they were, as a rule, lower than using the Poisson that they criticize, as can be seen in their Tables 2–4 and in Tables 1–3 below.

It should be noted that there is a systematic error in their computations: in order to use the chi-square goodness-of-fit statistic care must be taken that, under the validity of H_0 , $\sum_{k=1}^N p_k = 1$. This is an important point, namely in

deriving the asymptotic chi-square statistic — the classes considered must be a partition of the support of the distribution, $\sum_{k=0}^N e_k = \sum_{k=0}^N o_k$ is the reason why

the number of degrees of freedom in the asymptotic chi-square is always $N-1$ (minus the number of estimated parameters, if any). This accounts for the discrepant values between their Tables 2–4 and our Tables 1–3. The relative error in their Table 4, $s = 4$, is quite high, 12%. We point this out because it is a frequent error that needs to be eradicated, a mishandling that can be misleading.

- The computation of expected frequencies is highly simplified by using a marginal hypergeometric model, instead of their complex algorithm for computing expected values using the multivariate distribution.

The reader can verify that all calculations in Tables 2, 3 and 4 in Neuhäuser *et al.* (2001) are greatly simplified with $e_k = k p_k$, the p_k being the **univariate** hypergeometric probabilities

$$p_k = \frac{\binom{r}{k} \binom{n-r}{s-k}}{\binom{n}{s}} = \frac{\binom{s}{k} \binom{n-s}{r-k}}{\binom{n}{r}},$$

where s is the brood size, n is the number of nests with brood size s , $n = s n_s$ is the total number of young in broods of size s , and r the total number of extra-pair young (we use the notations s , n and r of Neuhäuser *et al.* (2001) for an easy comparison).

In fact, the appropriate model for the number of individuals from each of r classes C_1, \dots, C_r in a sample of size n taken randomly, without replacement, from a population of size $N = K_1 + \dots + K_r$ such that K_j individuals in the population belong to class C_j , $j = 1, \dots, r$, is

$\mathbf{X} = (X_1, \dots, X_r) \sim \text{Multihypergeometric}(N; n; p_1, \dots, p_r)$, where $p_j = \frac{K_j}{N}$:

$$X = (X_1, \dots, X_r) = \begin{cases} (k_1, \dots, k_r) & k_1 + \dots + k_r = n \\ P_{(k_1, \dots, k_r)} = \frac{\binom{K_1}{k_1} \dots \binom{K_r}{k_r}}{\binom{N}{n}} \end{cases}$$

But it is obvious that any univariate marginal random variable X_j counting the number of elements classified in class C_j has *Hypergeometric*(N, n, p_j) distribution: we need only to collect all the other classes in a residual

$\bar{C}_j = \bigcup_{k \neq j} C_k$. Therefore we get

$$P(X_j = x_j) = \frac{\binom{K_j}{x_j} \binom{N-K_j}{n-x_j}}{\binom{N}{n}}, \text{ and } E_{(x_1, \dots, x_r)}(X_j^v) = E_{x_k}(X_j^v).$$

Thus a much easier univariate hypergeometric model achieves the same goals as the multi-hypergeometric put forward by Neuhäuser *et al.* (2001).

Finally, we shall comment on the most important issue, i.e. what can be grasped from quantitative data as regards the understanding of mating behaviour. As has been established by Neuhäuser *et al.* (2001), the multivariate hypergeometric model and the Poisson law must be rejected, since the associated p -values are always

exceedingly small. Our results confirm this, and further establish that there is departure from the kind of randomness defined by the classical (Poisson, binomial, negative binomial and hypergeometric) count models, which always overestimate frequencies for the intermediate cases.

But other patterns of randomness may be an excellent fit with interesting behavioral interpretations — they are indeed an excellent fit, although as a general remark we would like to point out that it is easier, of course, to obtain a better fit with a model depending on more parameters, and that Akaike's criterion ought to be used to evaluate the relative benefit from using a more complicated model.

The observed frequencies seem indicate that the most common situation is either zero or at most one extra-pair, or, more seldom, many extra-pairs. This seems to indicate that only a small proportion of these females have a very promiscuous behaviour (or successful promiscuous behaviour), hypothesis that deserves further investigation.

Also, we have to bear in mind that we are modelling the number of extra-pair young as a step in trying to ascertain why females seek extra-pair fertilization. We are therefore looking in the present for questions which arose in the past, at an evolutionary level, which are far from straightforward. On the other hand, even random search by some females of extra pair fertilizations might lead to more elaborate random models for the number of extra-pair nestlings if other self-control or self-organizing mechanisms intervene in the process (such as regulation of the probability of success of extra pair fertilizations, or the probability that a female finds an available male, based on territorial constraints and nearest neighbour models). More sophisticated models, such as Zipf's equilibrium, therefore seem appropriate for seeking a much better explanation.

Nevertheless, the importance of modelling the observed number of extra pair nestlings should not be overlooked, as close fits to specific models may lead to interesting working hypotheses. The different models presented in this work are some of the possible alternatives when dealing with count data. For the available data, which is very scarce, there is clear indication that the logarithmic or the truncated logarithmic model often provide the best fit. This agrees with the explanation that different females have different behaviour, with some of them displaying more "eagerness" for promiscuous mating than others. Power laws, Zipf's law and Pareto models are closely related (Adamic, 2001, Mandelbrot, 1983), and seem to account for sophisticated forms of equilibrium reached by natural populations. The more general Zipf-Mandelbrot scaling models, and the discrete lognormal model whose structure changes with scale, provide good fits and deserve further investigation with larger samples. Scale regulation seems, indeed, to be the governing force behind most dynamic self-organizing phenomena.

Acknowledgements: This work has been partially supported by FCT/POCTI/FEDER, VEXTRA Project. T. A. Marques' research has been sponsored by FCT, SFRH/BM/2068/2000, FSE, III Quadro Comunitário de Apoio.

REFERENCES

- Adamic L.A. (2001). Zipf, Power-Laws and Pareto — a ranking tutorial.
<http://ginger.hpl.hp.com/shl/papers/ranking/>
- Berkson J. (1980). Minimum chi-square, not maximum likelihood! (with discussion), *Annals of Statistics*, 457–487.
- Cohen A. C. (1961). Estimating the Poisson parameter from samples that are truncated to the right, *Technometrics*, 433–438.
- Cressie N., Read T. R. C. (1984). Multinomial Goodness-of-fit tests, *Journal of the Royal Statistical Society B* 46, 440–464.
- Engen S. (1974). On species frequency models, *Biometrika*, 263–270.
- Fisher R. A., Corbet A. S., Williams C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population, *Journal of Animal Ecology*, 42–58.
- Gell-Mann M. (1994). *The Quark and the Jaguar*, Freeman, New York.
- Gilbert, N. (1989). *Biometrical Interpretation. Making Sense of Statistics in Biology*, Oxford University Press, Oxford.
- Hess K. Th., Lewald A., Schmidt K. D. (2002). An extension of Panjer's recursion, *ASTIN Bulletin* 32, 283–297.
- Johnson N. L., Kotz S., Kemp A. W. (1992). *Univariate Discrete Distributions*, Wiley, New York.
- Katz L. (1965). Unified treatment of a broad class of discrete probability distributions, *Classical and Contagious Discrete Distributions*, G. P. Patil (ed.), Pergamon Press, Oxford, 175–182.
- Mandelbrot B. (1983). *The Fractal Geometry of Nature*, Freeman, New York.
- Marques T. A., Pestana D. D. (2002). *Extra-Pair Fertilizations and the Distribution of Extra-Pair Young Within and Among Broods: A Case Study on Count Data Models in Biometry*, CEAUL, Lisboa.
- Moore P. G. (1954). A note on truncated Poisson distribution, *Biometrics* 10, 402–406.
- Moore P. G. (1956). The geometric, logarithmic and discrete Pareto forms of series, *Journal of the Institute of Actuaries* 82, 130–136.
- Neuhäuser M., Forstmeier W., Bretz F. (2001). The distribution of extra-pair young within and among broods — a technique to calculate deviations from randomness, *Journal of Avian Biology* 32, 358–363.
- Panjer H. H. (1981). Recursive evaluation of a family of compound distributions, *ASTIN Bulletin*, 22–26.
- Patil G. P., Wani J. K. (1965). Maximum likelihood estimation for the complete and truncated logarithmic series distribution, in G. P. Patil (ed.): *Classical and Contagious Discrete Distributions*, 398–409, Pergamon Press, Oxford.
- Pérez-Abreu V. (1991). Poisson approximation to power series distributions, *American Statistician* 45, 42–45.
- Petrie M., Kempenars B. (1998). Extra-pair paternity in birds: explaining variation between species and populations, *Trends Ecol. Evol.* 13, 52–58.
- Piegorsch W.W. (1990). Maximum likelihood estimation for the negative binomial dispersion parameter, *Biometrics* 46, 863–867.

- Rényi A. (1964). On an extremal property of the Poisson process, *Annals of the Institute of Statistical Mathematics* 16, 129–133.
- Rolski T., Schmidli H., Schmidt V., Teugels J. (1999). *Stochastic Processes for Insurance and Finance*, Wiley, New York.
- Seal H. L. (1952). The maximum likelihood fitting of the discrete Pareto law, *Bulletin of the Assoc. Institute of Actuaries* 78, 115–121.
- Sheldon B. C., Ellegren H. (1999). Sexual selection resulting from extra-pair paternity in collared flycatchers, *Animal Behavior* 57, 285–298.
- Stutchbury B. J., Rhymer J. M., Morton E. S. (1994). Extrapair paternity in hooded warblers, *Behavioral Ecology* 5, 384–392.
- Sundt B. Jewell W.S. (1981). Further results on recursive evaluation of compound distributions. *Astin Bulletin* 12, 27-39.
- Velosa S. F. (2003). *Discrete infinitely divisible laws*, CEAUL, Lisboa.
- Velosa S. F., (2003). New classes of discrete infinitely divisible laws in *Literacy and Statistics*, P. Brito, A. Figueiredo, F. Sousa, P. Teles e F. Rosado, eds., Sociedade Portuguesa de Estatística, Porto, 673–684.
- Willmot G. E. (1987). Sundt and Jewell's family of discrete distributions, *ASTIN Bulletin* 18, 17–29.
- Yezerinac S. M., Weatherhead P. J., Boag P. T. (1995). Extra-pair paternity and the opportunity for sexual selection in a socially monogamous bird (*Dendroica petechia*), *Behavioral Ecology and Sociobiology* 37, 179–188.
- Zar J. H. (1999). *Biostatistical Analysis*, 4th ed., Prentice Hall, Upper Saddle River.

Table 1. Extra pair nestlings in broods of size s , yellow warblers (*Dendroica petechia*), data from Yezerinac *et al.* (1995) - traditional count models. Observed and expected frequencies, G^2 observed value, degrees of freedom (d.f.), and corresponding p -value. For the untruncated models with infinite support, the last residual class represents $P(X \geq s)$.

	k	o_k	<i>Hyperg.</i>	<i>Pois.</i>	<i>Tr.Pois.</i>	<i>Bin.</i>	<i>Neg.Bin.</i>	<i>Tr.Neg.Bin.</i>
<i>est.par.</i>				0.429	0.500	0.143	1; 0.750	0.680
$s=3$	0	10	8.707	9.120	8.506	8.816	10.500	9.618
	1	2	4.610	3.909	4.253	4.408	2.625	3.080
	2	2	0.659	0.838	1.063	0.735	0.656	0.986
	3	0	0.024	0.134	0.177	0.041	0.219	0.316
G^2			3.872	2.643	2.745	3.364	2.394	1.880
d.f.			2	2	2	2	2	1
p -value			0.144	0.267	0.254	0.186	0.122	0.170
<i>est.par.</i>				1.730	2.462	0.432	10; 0.850	
$s=4$	0	10	3.720	6.561	3.522	3.839	7.247	
	1	7	11.757	11.349	8.668	11.701	10.902	
	2	9	11.350	9.816	10.669	13.373	9.021	
	3	5	6.748	5.659	8.754	6.793	5.428	
	4	6	1.225	3.614	5.387	1.294	4.403	
G^2			21.220	4.944	10.511	20.170	3.090	
d.f.			3	3	3	3	2	
p -value			0	0.176	0.015	0	0.213	
<i>est.par.</i>			~	1.538	1.935	0.308	1; 0.458	0.299
$s=5$	0	17	6.060	8.374	5.712	6.202	17.855	13.216
	1	8	13.878	12.883	11.055	13.783	9.681	9.271
	2	3	12.406	9.910	10.698	12.252	5.249	6.504
	3	3	5.410	5.082	6.902	5.445	2.846	4.563
	4	3	1.151	1.955	3.340	10.210	1.543	3.201
	5	5	0.095	0.797	1.293	0.108	1.827	2.246
G^2			59.537	27.051	32.162	57.398	6.299	6.659
d.f.			4	4	4	4	3	3
p -value			0	0	0	0	0.098	0.084

Table 2. Extra pair nestlings in broods of size s , yellow warblers (*Dendroica petechia*), data from Yezerinac *et al.* (1995) - power, generalized Zipf-Mandelbrot and discrete lognormal models.

Observed and expected frequencies, G^2 observed value, degrees of freedom (d.f.), and corresponding p -value. Maximum p -value, indicating the best fit, in boldface. For the untruncated models with infinite support, the last residual class represents $P(X \geq s)$.

	k	o_k	<i>Log.</i>	<i>Tr.Log.</i>	<i>Mand.</i>	<i>D.Logn.</i>
<i>est. par.</i>			0.491	0.563	-3.32; 5.02	0.07; 0.62
$s=3$	0	10	10.175	9.782	9.795	9.788
	1	2	2.500	2.751	2.801	2.972
	2	2	0.819	1.032	0.993	0.829
	3	0	0.505	0.435	0.410	0.411
G^2			2.330	1.813	1.664	2.387
d.f.			2	2	1	1
p -value			0.312	0.404	0.197	0.122
<i>est. par.</i>			0.827	1.251	-4.57; 0.00	0.86; 0.59
$s=4$	0	10	17.424	11.829	9.697	8.519
	1	7	7.209	7.401	8.211	11.840
	2	9	3.977	6.174	7.135	7.570
	3	5	2.468	5.795	6.303	4.139
	4	6	5.922	5.801	5.644	4.933
G^2			10.402	1.573	0.970	2.916
d.f.			3	3	2	2
p -value			0.015	0.665	0.616	0.233
<i>est. par.</i>			0.807	1.140	0.10; 0.00	0.55; 0.84
$s=5$	0	17	19.148	12.866	16.589	15.829
	1	8	7.722	7.332	7.858	9.665
	2	3	4.152	5.571	5.148	5.282
	3	3	2.512	4.763	3.828	2.987
	4	3	1.621	4.343	3.047	1.771
	5	5	3.846	4.125	2.530	3.467
G^2			1.956	4.085	3.499	2.890
d.f.			4	4	3	3
p -value			0.744	0.395	0.321	0.409

Table 3. Extra pair nestlings in broods of size s , hooded warblers (*Wilsonie citrina*), data from Stutchbury et al. (1994) - traditional count models. Observed and expected frequencies, G^2 observed value, degrees of freedom (d.f.), and corresponding p -value. For the untruncated models with infinite support, the last residual class represents $P(X \geq s)$.

	k	o_k	Hyperg.	Pois.	Tr.Pois.	Bin.	Neg.Bin.	Tr.Neg.Bin.
<i>est.par.</i>				0.692	1.200	0.346		
$s=2$	0	15	11	13.011	8.904	11.115		
	1	4	12	9.008	10.685	11.769		
	2	7	3	3.982	6.411	3.115		
G^2			12.378	5.673	9.016	11.692		
d.f.			1	1	1	1		
p -value			0	0.017	0.003	0.001		
<i>est.par.</i>				0.841	1.194	0.280	1; 0.566	0.437
$s=3$	0	28	16.255	18.978	13.797	16.402	24.904	21.388
	1	3	19.402	15.959	16.467	19.165	10.808	12.033
	2	5	7.430	6.710	9.827	7.464	4.691	6.770
	3	8	0.913	2.353	3.910	0.969	3.598	3.809
G^2			50.026	28.388	34.119	48.589	12.300	15.594
d.f.			2	2	2	2	1	1
p -value			0	0	0	0	0	0
<i>est.par.</i>				0.944	1.172	0.236	1; 0.473	0.455
$s=4$	0	23	12.098	14.000	11.225	12.258	17.027	17.217
	1	2	15.377	13.222	13.161	15.155	8.974	9.378
	2	4	7.048	6.244	7.715	7.027	4.729	5.108
	3	4	1.379	1.966	3.015	1.448	2.492	2.782
	4	3	0.097	0.568	0.884	0.112	2.777	1.515
G^2			45.957	27.390	29.800	44.203	10.734	12.187
d.f.			3	3	3	3	2	2
p -value			0	0	0	0	0.005	0
<i>est.par.</i>				1.667	2.500	0.333	$(\hat{\nu} \approx 0)$	
$s=5$	0	2	0.252	0.567	0.257	0.395		
	1	0	1.049	0.944	0.643	0.988		
	2	0	1.199	0.787	0.803	0.988		
	3	0	0.450	0.437	0.669	0.494		
	4	0	0.050	0.182	0.418	0.123		
	5	1	0.001	0.083	0.209	0.012		
G^2			22.107	10.032	11.335	15.276		
d.f.			4	4	4	4		
p -value			0	0.040	0.023	0.004		

Table 4. Extra pair nestlings in broods of size s , hooded warblers (*Wilsonia citrina*), data from Stutchbury *et al.* (1994) - power, generalized Zipf-Mandelbrot and discrete lognormal models.

Observed and expected frequencies, G^2 observed value, degrees of freedom (d.f.), and corresponding p -value. Maximum p -value, indicating the best fit, in boldface. For the untruncated models with infinite support, the last residual class represents $P(X \geq s)$.

	k	o_k	<i>Log.</i>	<i>Tr.Log.</i>	<i>Mand.</i>	<i>D.Logn.</i>
<i>est. par.</i>			0.622	1.261		
$s=2$	0	15	16.616	12.035		
	1	4	5.171	7.587		
	2	7	4.213	6.378		
G^2			1.985	2.789		
d.f.			1	1		
p -value			0.159	0.095		
<i>est. par.</i>			0.673	1.137	0.54; 0.00	-0.03; 0.91
$s=3$	0	28	26.487	18.585	26.911	25.924
	1	3	8.915	10.568	8.479	9.457
	2	5	4.001	8.012	5.032	4.007
	3	8	4.597	6.834	3.578	4.612
G^2			7.668	13.200	9.051	7.310
d.f.			2	2	1	1
p -value			0.022	0.001	0.003	0.007
<i>est. par.</i>			0.702	0.968	0.63; 0.00	-0.24; 1.05
$s=4$	0	23	20.882	16.368	22.213	21.124
	1	2	7.327	7.924	5.999	7.304
	2	4	3.428	5.115	3.468	3.207
	3	4	1.804	3.715	2.439	1.634
	4	3	2.560	2.877	1.881	2.731
G^2			7.809	9.014	4.441	7.669
d.f.			3	3	2	2
p -value			0.050	0.029	0.109	0.022
<i>est. par.</i>			0.821	3.000	0.69; 0.00	-2.47; 2.01
$s=5$	0	2	1.432	0.044	1.881	1.667
	1	0	0.588	0.065	0.445	0.518
	2	0	0.322	0.131	0.252	0.247
	3	0	0.198	0.294	0.176	0.142
	4	0	0.130	0.705	0.135	0.912
	5	1	0.331	1.762	0.110	0.335
G^2			3.549	14.179	8.232	2.381
d.f.			4	4	3	3
p -value			0.470	0.007	0.041	0.497

Table 5. Extra pair nestlings in broods of size s , collared flycatchers (*Ficedula albicollis*), data from Sheldon and Ellegren (1999) - traditional count models. Observed and expected frequencies, G^2 observed value, degrees of freedom (d.f.), and corresponding p -value. For the untruncated models with infinite support, the last residual class represents $P(X \geq s)$.

	k	o_k	<i>Hyperg.</i>	<i>Pois.</i>	<i>Tr.Pois.</i>	<i>Bin.</i>	<i>Neg.Bin.</i>	<i>Tr.Neg.Bin.</i>
<i>est.par.</i>				2.500	5.000	0.625		
$s=4$	0	0		0.164	0.031	0.040		
	1	0	0.143	0.410	0.153	0.264		
	2	1	0.857	0.513	0.382	0.659		
	3	1	0.857	0.428	0.637	0.732		
	4	0		0.485	0.797	0.305		
G^2			0.617	3.034	2.823	1.456		
d.f.			3	3	3	3		
p -value			0.893	0.386	0.420	0.692		
<i>est.par.</i>				1.529	2.000	0.306	1; 0.393	0.301
$s=5$	0	9	2.595	3.683	2.339	2.739	6.684	5.795
	1	2	6.133	5.633	4.679	6.035	4.056	4.050
	2	0	5.476	4.308	4.679	5.319	2.461	2.830
	3	2	2.306	2.196	3.119	2.344	1.493	1.978
	4	2	0.457	0.840	1.560	0.516	0.906	1.382
	5	2	0.034	0.340	0.624	0.046	1.399	0.966
G^2			39.528	22.126	24.729	36.906	8.291	9.538
d.f.			4	4	4	4	3	3
p -value			0	0	0	0	0.04	0.02
<i>est.par.</i>				0.804	0.822	0.134	1;0425	0.546
$s=6$	0	31	19.230	20.579	20.215	19.395	19.533	25.204
	1	3	18.244	16.553	16.622	18.015	11.239	11.45
	2	6	6.987	6.657	6.833	6.973	6.466	5.201
	3	4	1.382	1.785	1.873	1.439	3.721	2.363
	4	1	0.149	0.359	0.385	0.167	2.141	1.073
	5	0	0.008	0.058	0.063	0.010	1.232	0.488
	6	1	0	0.009	0.009	0	0.669	0.221
G^2			46.464	31.893	32.149	44.73	17.846	13.596
d.f.			5	5	5	5	4	4
p -value			0	0	0	0	0.001	0.009
<i>est.par.</i>				0.250	0.250	0.036	($\hat{v} \approx 0$)	
$s=7$	0	11	9.212	9.346	9.346	9.303		
	1	0	2.579	2.336	2.336	2.412		

	2	0	0.204	0.292	0.292	0.268
	3	1	0.004	0.024	0.024	0.017
	4	0		0.002	0.002	0.001
	5	0		0	0	0
	6	0		0	0	0
	7	0		0	0	0
G^2			14.75	11.017	11.017	11.89
d.f.			6	6	6	6
p -value			0.022	0.088	0.088	0.064

Table 6. Extra pair nestlings in broods of size s , collared flycatchers (*Ficedula albicollis*), data from Sheldon and Ellegren (1999) - power, generalized Zipf-Mandelbrot and discrete lognormal models.

Observed and expected frequencies, G^2 observed value, degrees of freedom (d.f.), and corresponding p -value. Maximum p -value, indicating the best fit, in boldface. For the untruncated models with infinite support, the last residual class represents $P(X \geq s)$.

	k	o_k	<i>Log.</i>	<i>Tr.Log.</i>	<i>Mand.</i>	<i>D.Logn.</i>
<i>est. par.</i>			0.882	1.235	-500; 0.00	1.25; 0.08
$s=4$	0	0	0.825	0.653	0.402	0
	1	0	0.364	0.403	0.401	0
	2	1	0.214	0.332	0.900	0.974
	3	1	0.142	0.308	0.904	1.026
	4	0	0.455	0.304	0.398	0.001
G^2			6.993	4.562	3.005	0.0015
d.f.			3	3	2	2
p -value			0.072	0.207	0.223	0.999
<i>est. par.</i>			0.806	1.215	0.40; 0.00	0.19; 1.07
$s=5$	0	9	8.364	4.966	8.458	8.005
	1	2	3.368	3.016	3.172	3.641
	2	0	1.809	2.442	1.952	1.890
	3	2	1.093	2.225	1.410	1.088
	4	2	0.704	2.162	1.103	0.675
	5	2	1.662	2.188	0.906	1.701
G^2			6.568	7.961	4.716	6.173
d.f.			4	4	3	3
p -value			0.160	0.093	0.194	0.103
<i>est. par.</i>			0.662	0.724	0.77; 0.00	-0.58; 1.11
$s=6$	0	31	28.088	26.419	30.736	29.231

	1	3	9.291	9.569	5.747	8.678
	2	6	4.098	4.622	3.170	3.360
	3	4	2.033	2.511	2.189	1.744
	4	1	1.076	1.455	1.671	0.958
	5	0	0.593	0.879	1.352	0.570
	6	1	0.820	0.546	1.135	1.259
G^2			9.572	10.274	6.978	9.036
d.f.			5	5	4	4
<i>p</i> -value			0.088	0.068	0.13?	0.060
<i>est. par.</i>			0.350	0.426	0.96; 0.01	-29.99; 3.61
<i>s</i> =7	0	11	9.749	9.214	10.935	10.47
	1	0	1.707	1.960	0.407	1.043
	2	0	0.398	0.556	0.206	0.266
	3	1	0.105	0.177	0.138	0.100
	4	0	0.029	0.060	0.103	0.047
	5	0	0.009	0.021	0.083	0.025
	6	0	0.003	0.008	0.069	0.015
	7	0	0.001	0.003	0.059	0.035
G^2			7.172	7.357	6.323	9.548
d.f.			6	6	5	5
<i>p</i> -value			0.305	0.289	0.276	0.089