A class of distributions applicable to the description of the number of nematodes parasitizing birds

I. KOPOCIŃSKA† AND B. KOPOCIŃSKI

Mathematical Institute, University of Wrocław, pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland

AND

A. OKULEWICZ

Institute of Microbiology, University of Wrocław, ul. Przybyszewskiego 63/77, 51-148 Wrocław, Poland

[Received on 20 November 2000; revised on 20 November 2001, 13 January 2003, 25 September 2003; accepted on 20 October 2003]

We define the class of bivariate binomial distributions as Poisson distributions transformed by a binomial dilution and mixing with a gamma distribution. They are used to study the number of nematodes, considering the sex ratio, parasitizing birds.

Keywords: negative binomial distribution; binomial dilution; mixed distribution; nematode; parasite distributions.

1. Introduction

The starting point of the paper is the estimation of the number of intestinal nematodes capillaria of the species B. o. (Baruscapillaria ovopunctata) and P. e. (Pterothominx (Avesonchotheca) exilis) parasitizing the song thrush (Turdus philomelos). Elementary analysis of the data indicates that the mean number of nematodes and the proportion of sexes of nematodes parasitizing the birds vary, the shape of the distributions of the number of nematodes is diverse, Poisson or negative binomial, and the numbers of the nematodes of both sexes are correlated. The aim of the paper is to mathematically model these phenomena under simple assumptions.

We introduce three probabilistic mechanisms from which our probability distributions arise: a mixing of Poisson distributions, a binomial dilution of population and a mixing of the dilution parameter. The propositions and theorems presented in the mathematical sections of the paper show which of the remarks presented in the Introduction may be explained by a given mechanism. Finally, it turns out that our bivariate data are compound in such a way that all mechanisms must necessarily be used for their description.

Data analyses and mathematical models considered in the biological literature often lead to bivariate negative binomial distributions. In the paper of Lonc et al. (1997) the frequency distributions of the total number of parasitic nematodes in a few species of birds are considered and negative binomial distributions are found in many cases. It is

†Email: ibk@hera.math.uni.wroc.pl
demonstrated by Kopocińska et al. (1997) that the distribution of the total number of nematodes in the blackbird was a mixture of negative binomial distributions, and that among the number of nematodes there were extreme outliers. Bivariate negative binomial distributions, with the same aggregation parameter of marginal distributions were found in the case of mallophagans (Mallophaga) Goniocotes chryzocephalus and Zlotorzyckiella colchici in the pheasant (Phasianus colchicus L.) by Kopociński (1999) and Kopociński et al. (1998).


In all the above bivariate distributions the marginal distributions are negative binomial with the same parameter of aggregation. This does not create an obstacle in many applications, e.g. insurance, but limits the range of applications in parasitology, where the variation of aggregation describes the variability of individuals within sexes and species.

Tables 1–2 show data collected in the Department of General Parasitology, Institute of Microbiology, University of Wrocław. In the analysis we used 50 individuals infected with one of the two nematoda species mentioned. Non-infected birds and those with extremely large numbers of nematodes (over 20), which are treated as outliers were omitted from the analysis. The purpose of the paper is the analysis of mathematical models. When commenting on the data, we mark the text with □ □. We present most of our propositions without proofs, but by presenting the generating function of the distributions, we suggest a method of the proof. For proofs of some theorems see also Johnson et al. (1997).

### Table 1 Frequency distribution of the number of nematodes B. o. parasitizing song thrush (T. philomelos)

<table>
<thead>
<tr>
<th>$Y_1$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>*</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$Y_1$, the number of male nematodes; $Y_2$, the number of female nematodes; *, missing data.
TABLE 2 Frequency distribution of the number of nematodes P. e. parasitizing song thrush (T. philomelos)

<table>
<thead>
<tr>
<th>$Y_1$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_2$</td>
<td>*</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

$Y_1$, the number of male nematodes; $Y_2$, the number of female nematodes; *, missing data.

TABLE 3 The description of the distribution of the number of nematodes and its parameters

<table>
<thead>
<tr>
<th></th>
<th>Poisson distributions</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>$\lambda$</td>
<td>$df$</td>
<td>$\chi^2$</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>$Y_1$</td>
<td>2.340</td>
<td>4</td>
<td>0.29</td>
</tr>
<tr>
<td>2</td>
<td>$Y_1$</td>
<td>1.740</td>
<td>3</td>
<td>1.49</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Negative binomial distributions</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>$p$</td>
<td>$a$</td>
<td>$df$</td>
</tr>
<tr>
<td>1</td>
<td>$Y_1$</td>
<td>0.3**</td>
<td>5.55</td>
</tr>
<tr>
<td>2</td>
<td>$Y_1$</td>
<td>0.2**</td>
<td>7.13</td>
</tr>
<tr>
<td>1</td>
<td>$Y_2$</td>
<td>0.683</td>
<td>2.64</td>
</tr>
<tr>
<td>2</td>
<td>$Y_2$</td>
<td>0.606</td>
<td>2.25</td>
</tr>
</tbody>
</table>

$\lambda$, the expected value in a Poisson distribution; $p$, $a$, the parameters in a negative distribution: $ap/(1-p)$, the expected value; $a$, the aggregation parameter. ***, the heuristically assumed value. The chi-square statistics are calculated for $\min(Y_j, df + 3)$, $j = 1, 2$.

1.1 Description of the empirical distributions

When describing the distribution of the number of nematodes of each sex in a given bird species the Poisson and negative binomial distributions were considered.

Tables 1–2 contain the data used, $Y_1$ denotes the number of male nematodes, $Y_2$—the number of female nematodes and further $R = Y_1 + Y_2$ denotes the total number of nematodes.

Table 3 shows the results of fitting the frequency distributions: the type of distribution and parameters. The goodness of fit is illustrated by the $\chi^2$-statistic and $df$ denotes the degrees of freedom. It follows that the distributions of the number of nematodes and the parameters are diverse. In both cases the male nematodes have a Poisson distribution, the female nematodes often have a negative binomial distribution.

The proportion of sexes of the nematodes is unequal, it appears to depend on the total number of nematodes in the host. Linear regression of $Y_1$ with respect to $R$ gives a positive constant term, suggesting that the proportion of males decreases in $R$. The variances of the
number of males conditioned on \( R \) are sub-binomial. The numbers of nematodes of both sexes are positively correlated: 0.72, 0.57 (in order of the data tables).

1.2 Proportions of the sexes

Let \( p = \frac{E(Y_1)}{E(R)} \) denote the proportion of males in the population of nematodes; \( p_j = \frac{E(Y_1|R = j)}{j}, j \geq 1 \), denotes the proportion of males under the condition \( R = j \).

Let us assume that \( p \) is known and consider the hypothesis \( H_0^{(1)}: p_j = p, j \geq 1 \), and the alternative hypothesis \( H_1^{(1)} \) that \( H_0^{(1)} \) is not satisfied. Let \( n_{ij} \) denote the number of observations \((X = i, R = j)\). The test statistic is

\[
\chi^2_1 = \frac{1}{p(1-p)} \sum_j (\hat{p}_j - p)^2 j n_j,
\]

where

\[
\hat{p}_j = \frac{1}{jn_j} \sum_{i=1}^j i n_{ij}, \quad n_j = \sum_{i=0}^j n_{ij}.
\]

Assuming \( H_0^{(1)} \) is true the random variable \((\hat{p}_j - p)/in_j/p(1-p)\) is asymptotically normal \( N(0, 1) \) and thus \( \chi^2_1 \) has a chi-square distribution (asymptotically) and the number of degrees of freedom \( df \) is equal to the number of non-empty classes in the frequency distribution of \( R \). In practice, when \( p \) is unknown, we use its estimator instead:

\[
\hat{p} = \frac{1}{n} \sum_j \sum_{i=1}^j i n_{ij}, \quad n = \sum_j n_j.
\]

Recall that the standardized statistic \( \chi^2 \) is asymptotically normal

\[
U = \sqrt{2\chi^2 - \sqrt{2n - 1}} \sim N(0, 1).
\]

\( \square \) For Data 1 we have \( p = 0.291, df = 16, \chi^2 = 7.82 \). For Data 2 we have \( p = 0.335, df = 14, \chi^2 = 9.91 \). At a significance level of 0.05 we do not reject \( H_0^{(1)} \). \( \square \)

1.3 Mathematical models

For the easy identification of standard probability distributions and their parameters we recall the standard definitions.

**Definition 1**

(i) The random variable \( II = II(\lambda) \) has Poisson distribution with parameter \( \lambda > 0 \), if \( P(II = n) = (n!)^{-1} \lambda^n e^{-\lambda}, n \geq 0 \). We have \( E(II) = \lambda \) and \( \text{Var}(II) = \lambda \).

(ii) The random variable \( \Gamma = \Gamma(v, \alpha) \) has gamma distribution with parameters \( v > 0, \alpha > 0 \), if its density is of the form \( f_{\Gamma}(x) = (\Gamma(\alpha))^{-1} x^{v-1} \alpha^v e^{-\alpha x}, x \geq 0 \). We have \( E(\Gamma) = \alpha v^{-1} \) and \( \text{Var}(\Gamma) = \alpha v^{-2} \).
(iii) The random variable $B = B(p, N)$ has binomial distribution with parameters $0 \leq p \leq 1$, $N \geq 0$, if $P(B = n) = b(n; p, N) = \binom{N}{n}p^n(1 - p)^{N-n}$, $0 \leq n \leq N$. We have $E(B) = np$ and $\text{Var}(B) = np(1 - p)$.

(iv) The random variable $\mathcal{N}_B = \mathcal{N}_B(p, a)$ has negative binomial distribution with parameters $0 < p < 1$, $a > 0$, if $P(\mathcal{N}_B = n) = \mathcal{N}_b(n; p, a) = (n!)^{-1}a(a + 1)\cdots(a + n - 1)p^n(1 - p)^a$, $n \geq 0$. The parameter $a$ is called the aggregation parameter. We have $E(\mathcal{N}_B) = ap(1 - p)^{-1}$ and $\text{Var}(\mathcal{N}_B) = ap(1 - p)^{-2}$.

(v) We define the generating function of the distribution of a integer valued random variable $X \phi_X(u)$ to be $E(u^X)$. We define the Laplace transform of the distribution of a non-negative random variable $X \phi^*_X(s)$ to be $E(e^{-sX})$, $s > s^*$, for some $s^*$. We have: $\phi^*_X(u) = e^{-\lambda(u-1)}$, $\phi^*_Y(s) = (\frac{s}{s+\lambda})^\alpha$, $\phi_G(u) = (pu + 1 - p)^N$ and $\phi_{\mathcal{N}_B}(u) = (\frac{1-p}{1-pu})^a$. Transforms of multivariate distributions are denoted and defined analogously.

(vi) Let $X = X(w)$ be a random variable with probability distribution function $F_{X,w}$ with parameter $w$. Let $W$ be a random variable with distribution $G$ on a domain $D$ and let $X$, $W$ be mutually independent. Then $Y = X(W)$ is a random variable with mixed distribution function $F_Y(x) = \int_D F_{X,w}(x)G(dw)$.

□ The mixing operation in biological sciences is interpreted in the following way: $X(w)$ represents a random feature of an individual distributed according to some parameter $w$; $W$ represents the random variability of the parameter in a population of individuals. Hence $X(W)$ is the random feature of an individual from the population.

We assume that the number of nematodes parasitizing the birds is a result of the environment as well as disease immunity. The diversity of environment mixes the parameters of the distribution of the parasite invasion while the diversity of the immunity of birds mixes the parameters of the dilution.

We estimate the model parameters by the least squares method based, however, on an incomplete data set (a number of observations in one cell in the data set is missing). For the estimation of the bivariate distribution the marginal parameters are partially used. □

2. Binomial dilution

Below we consider models leading to a mixed binomial distribution.

2.1 In one dimension

Let $X$ be an integer valued random variable and let $\delta_j$, $j \geq 1$, be a sequence of Bernoulli trials with $P(\delta_j = 1) = p$. The random variable

$$Y = B(p, X) = \sum_{j=1}^X \delta_j$$

(1)

is the binomial dilution of the random variable $X$. It has mixed binomial distribution.

**Proposition 1** Let $X$, $Y$ be integer valued random variables. Suppose that the random variable $Y|(X = n)$ has binomial distribution with parameters $p$, $n$. The random variable $Y$
has mixed binomial distribution $P(Y = k) = \mathcal{E}(b(k; p, X))$, $k \geq 0$. Moreover, we have
\[ \phi_{\Phi(p, X)}(u) = \phi_X(pu + 1 - p), E(Y) = pE(X), \text{Var}(Y) = p^2\text{Var}(X) + p(1 - p)E(X). \]

In the sequel let $B_i(a, X) = \sum_{j=1}^{X} \delta_j^{(i)}$, where $a = E(\delta_j^{(i)}), \delta_j^{(i)}, j \geq 1, i = 1, 2$, are mutually independent.

**Proposition 2** The binomial dilution of random variable preserves some type of distribution, namely:

- binomial: $B_1(p, B_2(a, N)) \overset{d}{=} B(ap, N)$;
- Poisson: $B(p, \Pi(\lambda)) \overset{d}{=} \Pi(p\lambda)$;
- negative binomial: $B(p, N_B(v, a)) \overset{d}{=} NB_2\left(\frac{vp}{1-v(1-p)}, a\right)$.

**Proposition 3** If $Y_1 = B(p, X), Y_2 = X - Y_1$, then

\[ \phi_{Y_1,Y_2}(u_1, u_2) = \phi_X(pu_1 + (1 - p)u_2), \]
\[ \text{Cov}(Y_1, Y_2) = p(1 - p)(\text{Var}(X) - E(X)). \]

**Proposition 4** If $Y_1 = B_1(p_1, X), Y_2 = B_2(p_2, X)$, then

\[ \phi_{Y_1,Y_2}(u_1, u_2) = \phi_X((p_1u_1 + 1 - p_1)(p_2u_2 + 1 - p_2)), \]
\[ \text{Cov}(Y_1, Y_2) = p_1p_2\text{Var}(X). \]

**Proposition 5** If $X_i, i = 1, 2$, are mutually independent, then $B_i(p, X_i), i = 1, 2$, are mutually independent and $B(p, X_1 + X_2) = B_1(p, X_1) + B_2(p, X_2)$.

### 2.2 Mixed dilution

Let $P$ be a random variable on $[0,1]$ and $X$ be an integer random variable. Then $Y = B(P, X)$ defined as $Y|\{P = p\} = B(p, X)$ is a doubly mixed random variable. If $\Pi(\lambda)$ is a Poisson random variable then $B(\Pi(\lambda), P) \overset{d}{=} B(P, \Pi(\lambda))$.

**Proposition 6** If $P, X$ are mutually independent, then

\[ \phi_Y(u) = E_P(\phi_X(Pu + 1 - P)), \]
\[ E(Y) = E(P)E(X), \]
\[ \text{Var}(Y) = E(P^2)E(X(X - 1)) + E(Y)(1 - E(Y)). \]

**Proposition 7** If $P, \Pi(\lambda)$ are mutually independent and $\Pi(\lambda)$ is a Poisson random variable, then $\text{Var}(B(\Pi(\lambda), P)) = \lambda^2\text{Var}(P) + \lambda E(P)$.

The binomial dilution with an essentially random $P$ (e.g. when $\text{Var}(P) > 0$) does not preserve the Poisson distribution because

\[ \frac{\text{Var}(Y)}{E(Y)} = \lambda \frac{\text{Var}(P)}{E(P)} + 1. \]
2.3 In two dimensions

For \( i = 1, 2 \) let \( P_i \) be a random variable on \([0,1]\), \( X_i \) an integer random variable and let \( Y_i = B_i(P_i, X_i) \).

**Proposition 8** If \((X_1, X_2), (P_1, P_2)\) are mutually independent random vectors, then

\[
\text{Cov}(Y_1, Y_2) = \text{Cov}(P_1, P_2) \text{Cov}(X_1, X_2) + \text{Cov}(P_1, P_2) E(X_1)E(X_2)
\]

\[
+ \text{Cov}(X_1, X_2)E(P_1)E(P_2).
\]

Note that if \((X_1, X_2), P_1, P_2\) are mutually independent random elements, then

\[
\text{Cov}(Y_1, Y_2) = \text{Cov}(X_1, X_2)E(P_1)E(P_2).
\]

If \( P_1 = p_1 \) is deterministic, then

\[
\text{Cov}(Y_1, Y_2) = p_1 E(P_2) \text{Cov}(X_1, X_2).
\]

\(\Box\) We interpret the binomial dilution (1) as choosing a nematode out of organism by the Bernoulli trials: a bird is infected by a number \( X \) of parasites, but only a reduced part \( Y \) remains in the host, where \( p \) is the probability that the parasite will remain in the host. The random variables \( Y_1, Y_2 \) in Proposition 3 may be interpreted as the division of the offspring number \( X \) between the two sexes. These random variables are negatively correlated when \( X \) is constant, or independent when \( X \) has a Poisson distribution. When \( X \) has a negative binomial distribution with parameters \( p, \alpha \), then \( \text{Cov}(Y_1, Y_2) = \alpha(1 - p)^2/p^2 \) is positive. Proposition 4 describes the case when the number of individuals of both sexes is strictly equal and chances of survival unequal. The shape of the distribution of \( X \) describes the variability of birds in the population. The shape of the distribution of \( P \) describes the variability of parasite immunity to protective action by the bird. \( \Box \)

3. Negative binomial distributions

The negative binomial distributions in the probabilistic model are derived as mixed Poisson distributions.

**Proposition 9** If \( \Pi(\lambda) \) is a Poisson random variable with parameter \( \lambda \), \( \Gamma = \Gamma(v, \alpha) \) is a random variable with gamma distribution and \( \Pi(\lambda), \Gamma(v, \alpha) \) are mutually independent, then \( \Pi(\Gamma) \) has a negative binomial distribution with parameters \( (v + 1)^{-1}, \alpha \);

\[
\Pi(\Gamma(v, \alpha)) = \text{NB}
\]

After generalizing the negative binomial distribution to the multivariate distribution, we consider multivariate Poisson distributions (see Johnson et al., 1997) and analogously define multivariate gamma distributions.

**Definition 2**

(i) Let \( \Pi^{(i)} = \Pi^{(i)}(\lambda^{(i)}) \), \( i = 1, 2, 3 \), be mutually independent Poisson random variables with parameters \( \lambda^{(i)} \), respectively. Then \( \Pi_1 = \Pi^{(1)} + \Pi^{(2)}, \Pi_2 = \Pi^{(2)} + \Pi^{(3)} \) have a bivariate Poisson distribution with parameters \( \lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)} \).
42  I. KOPOCIŃSKA ET AL.

(ii) Let $I_i^{(i)} = I_i^{(i)}(v, \alpha_i)$, $i = 1, 2, 3$, be mutually independent random variables with a gamma distribution with parameters $v, \alpha_i$, respectively. Then $I_1' = I_1^{(1)} + I_2^{(2)}$, $I_2' = I_1^{(2)} + I_2^{(3)}$ have a bivariate gamma distribution with parameters $v, \alpha^{(1)}, \alpha^{(2)}$, $\alpha^{(3)}$.

Both the bivariate Poisson random variables and the bivariate gamma random variables are positively correlated.

**Proposition 10** Let $X_1 = II_1(\alpha) + II_2(\beta)$, $X_2 = II_2(\beta) + II_3(\gamma)$ be bivariate Poisson random variables. Then

$$P(X_1 = m|X_1 + X_2 = n) = \sum_{i=0}^{\min(m,n-n)} \frac{\beta^i \alpha^{m-i} \gamma^{n-m-i}}{i!(m-i)!(n-m-i)!} \left( \sum_{i=0}^{n/2} \frac{\beta^i (\alpha + \gamma)^{n-2i}}{i!(n-2i)!} \right)^{-1}, \quad 0 \leq m \leq n. \quad (2)$$

**Proposition 11** From (2) we obtain

$$E(X_1|X_1 + X_2 = n) = \frac{\alpha n}{\alpha + \gamma} + \frac{\gamma - \alpha}{\gamma + \alpha} \left( \sum_{i=0}^{n/2} \frac{\beta^i (\alpha + \gamma)^{n-2i}}{i!(n-2i)!} \right)^{-1} \left( \sum_{i=0}^{n/2} \frac{\beta^i (\alpha + \gamma)^{n-2i}}{i!(n-2i)!} \right)^{-1}.$$

For $\alpha = \gamma$ we have $E(X_1|X_1 + X_2 = n) = n/2$, and for $\alpha < \gamma$ this expected value decreases in $n$.

In Proposition 4 one random variable is dilated in two ways. We return to the problem where a Poisson random variable is dilated in two ways and mixed.

**Definition 3** Let $II(\lambda)$ be a Poisson random variable, $P_1, P_2$ are random variables with a joint distribution on $[0,1] \times [0,1]$ and the random elements $II(\lambda)$, $(P_1, P_2)$ are mutually independent. Then the random variables $Y_i = II(\lambda, P_i)$, $i = 1, 2$, are the bivariate dilution of $II(\lambda)$.

**Proposition 12** Let $Y_1, Y_2$ be the bivariate dilution of a Poisson random variable $II(\lambda)$. Then

$$\phi_{Y_1,Y_2}(u_1, u_2) = E_{P_1, P_2}(\exp(-\lambda(1 - (P_1u_1 + 1 - P_1)(P_2u_2 + 1 - P_2)))).$$

$$P(Y_1 = m, Y_2 = n) = E_{P_1, P_2} \left( \sum_{k=\max(m,n)}^{\infty} \binom{k}{m} P_1^m (1 - P_1)^{k-m} \times \binom{k}{n} P_2^n (1 - P_2)^{k-n} \right), \quad m \geq 0, n \geq 0,$$

$$\text{Cov}(Y_1, Y_2) = \lambda(\lambda + 1)\text{Cov}(P_1, P_2) + \lambda E(P_1)E(P_2).$$

**Definition 4** If $II_i^{(i)}(\lambda_i), i = 1, 2, 3$, are mutually independent random variables with Poisson distributions and $A_j = I_i^{(j)}(v, \alpha_j)$, $j = 1, 2, 3$, are mutually independent random variables with gamma distributions, then the random variables

$$Y_1 = II^{(1)}(A_1) + II^{(2)}(A_2), \quad Y_2 = II^{(2)}(A_2) + II^{(3)}(A_3) \quad (3)$$

have a bivariate negative binomial distribution of type I with parameters $v, \alpha_1, \alpha_2, \alpha_3$. 


PROPOSITION 13 For the random variables (3) we have
\[
\phi_{Y_1,Y_2}(u_1, u_2) = \left( \frac{v}{v+1-u_1} \right)^{\alpha_1} \left( \frac{v}{v+1-u_1 u_2} \right)^{\alpha_2} \left( \frac{v}{v+1-u_2} \right)^{\alpha_3},
\]
(4)
\[
\text{Cov}(Y_1, Y_2) = \text{Var}(NB((v+1)^{-1}, \alpha_2)).
\]

From (4) we obtain the following proposition.

PROPOSITION 14 If \( Y_1, Y_2 \) are bivariate negative binomial random variables of type I with parameters \( v, \alpha_1, \alpha_2, \alpha_3 \), then \( Y_1 \) is \( NB(p, \alpha_1 + \alpha_2) \), \( Y_2 \) is \( NB(p, \alpha_2 + \alpha_3) \), where \( p = (v+1)^{-1} \). Hence
\[
\text{E}(Y_1) \text{Var}(Y_2) = \text{Var}(Y_1) \text{E}(Y_2).
\]

DEFINITION 5 If \( II^{(i)}(\lambda_i) \), \( i = 1, 2 \), are mutually independent random variables with Poisson distributions and \( A_i = \Gamma^{(i)}(1, \alpha_j) \), \( j = 1, 2, 3 \), are mutually independent random variables with gamma distributions, then
\[
Y_1 = II^{(1)}(\lambda_1(A_1 + A_2)), \quad Y_2 = II^{(2)}(\lambda_2(A_2 + A_3))
\]
(5)
have a bivariate negative binomial distribution of type II with parameters \( \lambda_1, \lambda_2, \alpha_1, \alpha_2, \alpha_3 \).

PROPOSITION 15 For the random variables (5) we have
\[
\phi_{Y_1,Y_2}(u_1, u_2) = \left( \frac{1}{1 + \lambda_1(1-u_1)} \right)^{\alpha_1} \left( \frac{1}{1 + \lambda_1(1-u_1) + \lambda_2(1-u_2)} \right)^{\alpha_2} \times \left( \frac{1}{1 + \lambda_2(1-u_2)} \right)^{\alpha_3},
\]
(6)
\[
\text{Cov}(Y_1, Y_2) = \lambda_1 \lambda_2 \text{Var}(A_2).
\]

From (5) we obtain the following proposition.

PROPOSITION 16 If \( Y_1, Y_2 \) are bivariate negative binomial of type II with parameters \( \lambda_1, \lambda_2, \alpha_1, \alpha_2, \alpha_3 \), then \( Y_1 \) is \( NB(p_1, \alpha_1 + \alpha_2) \), \( Y_2 \) is \( NB(p_2, \alpha_2 + \alpha_3) \),
\[
\text{Corr}(Y_1, Y_2) = \alpha_2 \frac{p_1 p_2}{(\alpha_1 + \alpha_2)(\alpha_2 + \alpha_3)},
\]
(7)
where \( p_i = \lambda_i(1 + \lambda_i)^{-1}, i = 1, 2 \).

Remarks. We have \( \text{Corr}(Y_1, Y_2) \leq \sqrt{p_1 p_2} \); the inequality is strict, if \( \alpha_1 = \alpha_3 = 0 \). If \( Y_1 \) is Poisson, then \( p_1 = 0 \) and \( \text{Cov}(Y_1, Y_2) = 0 \).

In Definition 4 it is assumed that the random variables \( Y_1, Y_2 \), conditioned by three mutually independent random variables with gamma distributions have a bivariate Poisson distribution. In Definition 5 it is assumed that \( Y_1, Y_2 \) conditioned by random variables with bivariate gamma distributions are independent Poisson random variables. In both these types of bivariate negative binomial distributions the marginal distributions are negative binomial (with some aggregation) and the correlation coefficient is non-negative,
4. Limit theorems

When we generate the negative binomial distribution starting from a Poisson random variable and mixing it with a gamma distribution, the domain of the mixture is finite. The negative binomial variable and mixing it with a gamma distribution, the domain of the mixture is infinite. When we generate the negative binomial distribution starting from a Poisson random variable, we can obtain the asymptotically negative binomial distribution with parameters $v$ and $\alpha$.

**Theorem 1** Let $\Pi(\lambda)$ have Poisson distribution with parameter $\lambda$ and $P = I_{[0,1]}(\Gamma(\lambda v, \alpha))$ have gamma distribution with parameters $\lambda v, \alpha$, truncated to $[0,1]$, and $Y = B(P, \Pi(\lambda))$. If $P, \Pi$ are mutually independent and $\lambda \to +\infty$, then $Y$ has an asymptotically negative binomial distribution with parameters $(v+1)^{-1}, \alpha$.

**Proof of Theorem 1.** Let $F_{\Gamma(\lambda v, \alpha)}$ denote the probability distribution function of $\Gamma(\lambda v, \alpha)$ and $f_{\Gamma(\lambda v, \alpha)}$ denote the density of the distribution. The generating function of $Y$ is of the form

$$\phi_Y(u) = E(u^{B(P, \Pi(\lambda))}) = E_P(E_{\Pi(\lambda)}(P^{1-u}P^{\Pi(\lambda)})) = E_P(e^{-\lambda P(1-u)})$$

$$= \int_0^1 e^{-\lambda x(1-u)} f_{\Gamma(\lambda v, \alpha)}(x) dx / F_{\Gamma(\lambda v, \alpha)}(1)$$

$$= \int_0^\lambda \frac{x^\alpha - 1}{\Gamma(\alpha)} e^{-x(1+u-v)} dx / F_{\Gamma(\lambda v, \alpha)}(\lambda).$$

If $\lambda \to \infty$, then it converges to the generating function $\left( \frac{v}{\Gamma(\lambda v, \alpha)} \right)^\alpha$ of $NB((v+1)^{-1}, \alpha)$.

**Theorem 2** Let $\Pi^{(i)}(\lambda_1^{(i)}, \lambda_2^{(i)})$, $i = 1, 2, 3$, be mutually independent Poisson random variables with parameters $\lambda_1^{(1)}, \lambda_2^{(1)}, \lambda_2^{(3)}$, respectively and $\Pi_1 = \Pi^{(1)}(\lambda_1^{(1)}) + \Pi^{(2)}(\lambda_1^{(2)})$, $\Pi_2 = \Pi^{(2)}(\lambda_1^{(2)}) + \Pi^{(3)}(\lambda_2^{(3)})$ have a bivariate Poisson distribution. Let $\Gamma_i = \Gamma_i(\lambda v, \alpha_i)$, $i = 1, 2, 3$, be mutually independent random variables with gamma distributions, respectively. If $\Pi_1, \Pi_2$ have a bivariate gamma distribution with parameters $\lambda_1, \alpha_1, \alpha_2, \alpha_3$, truncated to $[0,1] \times [0,1]$ and $Y_i = B(P_i, \Pi_i)$, $i = 1, 2$. If $(\Pi_1, \Pi_2)$, $(P_1, P_2)$ are mutually independent and $\lambda \to +\infty$, then $Y_1, Y_2$ have an asymptotically bivariate negative binomial of type $II$ with parameters $\lambda_1 = \lambda^{(1)} + \lambda^{(2)}, \lambda_2 = \lambda^{(2)} + \lambda^{(3)}, \alpha_1, \alpha_2, \alpha_3$.

The proof of Theorem 2 is similar to the proof of Theorem 1.

5. Fitting the bivariate distribution to the data

Bivariate negative binomial distributions of type I or type II may be described in terms of a negative binomial distribution.

**Theorem 3** The bivariate negative binomial distribution (3) is of the form

$$P(Y_1 = i, Y_2 = j) = \sum_{k=0}^{\min(i, j)} N(b(i-k; p, \alpha_1) \cdot b(k; p, \alpha_2) \cdot N(b(j-k; p, \alpha_3), i, j \geq 0),$$

where $p = (v+1)^{-1}$. 

Proof of Theorem 3. Expanding (4) as a series we obtain
\[
\phi_{Y_1, Y_2}(u_1, u_2) = \left( \sum_{n=0}^{\infty} Np(n; p, \alpha_1)u_1^n \right) \times \left( \sum_{n=0}^{\infty} Np(n; p, \alpha_2)u_2^n \right) \left( \sum_{n=0}^{\infty} Np(n; p, \alpha_3)u_1^n \right).
\]

After changing the order of summation, we obtain the bivariate series
\[
\phi_{Y_1, Y_2}(u_1, u_2) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} P(Y_1 = i, Y_2 = j)u_1^i u_2^j.
\]

THEOREM 4 The bivariate negative binomial distribution (6) is of the form
\[
P(Y_1 = i, Y_2 = j) = \sum_{k=0}^{i} \sum_{l=0}^{j} \binom{k + l}{k} p_{21}^{k} p_{22}^{l} \times N_{b}(i - k; p_1, \alpha_1) N_{b}(k + l; p_2, \alpha_2) N_{b}(j - l; p_3, \alpha_3), \quad i, j \geq 0,
\]
where
\[
p_1 = \frac{\lambda_1}{1 + \lambda_1}, \quad p_2 = \frac{\lambda_1 + \lambda_2}{1 + \lambda_1 + \lambda_2}, \quad p_3 = \frac{\lambda_2}{1 + \lambda_2}, \quad p_{21} = \frac{\lambda_1}{\lambda_1 + \lambda_2}, \quad p_{22} = 1 - p_{21}.
\]

Proof of Theorem 4. Expanding (6) as a series we obtain
\[
\phi_{Y_1, Y_2}(u_1, u_2) = \left( \sum_{n=0}^{\infty} Np(n; p_1, \alpha_1)u_1^n \right) \times \left( \sum_{n=0}^{\infty} Np(n; p_2, \alpha_2)(p_{21}u_1 + p_{22}u_2)^n \right) \left( \sum_{n=0}^{\infty} Np(n; p_3, \alpha_3)u_2^n \right).
\]

After expanding \( (p_{21}u_1 + p_{22}u_2)^n \) and changing the order of summation, we obtain Theorem 4. □

Consider the problem of the parameter estimation for the bivariate negative binomial distribution of type II. The parameters of the marginal distributions may be estimated by the least squares method. The parameters \( p_1, p_2, \alpha_1, \alpha_2, \alpha_3 \) of the bivariate distribution may be estimated using the previously estimated parameters \( p_1, \alpha^{(1)} = \alpha_1 + \alpha_2 \) and \( p_2, \alpha^{(2)} = \alpha_2 + \alpha_3 \) of the marginal distribution and correlation (7). For \( \alpha_1, \alpha_2, \alpha_3 \) we have estimators
\[
\hat{\alpha}_2 = \text{Corr}(Y_1, Y_2) \sqrt{\alpha^{(1)} \alpha^{(2)} p_1^{-1} p_2^{-1}}, \quad \hat{\alpha}_1 = \alpha^{(1)} - \hat{\alpha}_2, \quad \hat{\alpha}_3 = \alpha^{(2)} - \hat{\alpha}_2.
\]
For both sets of data $Y_1$ is Poisson (see Table 3) and $\text{Cov}(Y_1, Y_2) > 0$, hence the bivariate negative binomial distribution (type I or type II) cannot be fitted. If for Data 1 we substitute the Poisson distribution for $Y_1$ by the negative binomial $NB(0.3, 5.55)$, then we can fit a bivariate distribution. The expected frequencies and parameters are presented in Table 4. If for Data 2 we substitute the Poisson distribution for $Y_1$ by the negative binomial distribution $NB(0.2, 7.13)$, then we can fit a bivariate distribution. The expected frequencies and parameters are presented in Table 5. In both cases $\alpha_2 > \alpha(2)$, hence we take $\alpha_2 = \alpha(2)$, $\hat{\alpha}_1 = \alpha(1) - \hat{\alpha}_2$, $\hat{\alpha}_3 = 0$. □

6. Applications

Now we return to the discussion of the data from Tables 1–2. On the basis of the mathematical models introduced we attempt to explain the observations previously presented.

The missing number in the (0, 0) cell of our data set causes some difficulties in the parameter estimation of the models. In the case of the marginal distributions we estimate the parameters as well as the number in the (0) cell. In the case of the bivariate distribution a few experiments of the parameter estimation using 0, 1, 2, . . . for the number in the (0, 0) cell were performed, but the problem of the most proper missing number is not considered, and we assumed this number to be equal to zero.

For both sets of data the best fitted distribution for $Y_1$ in the class of negative binomial distributions is a Poisson distribution. But then a bivariate negative binomial distribution with positive correlation cannot be taken into account in the model. Note that marginal distributions do not determine the bivariate distribution and the statistical tests do not reject
TABLE 5  The approximation of the frequency distribution of the number of nematodes in Data 2

<table>
<thead>
<tr>
<th>( Y_1 ) ( Y_2 )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>( \geq 10 )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0-0.68</td>
<td>10.19</td>
</tr>
<tr>
<td>0**</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.74</td>
<td>14.53</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.90</td>
<td>11.81</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.93</td>
<td>7.19</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0.09</td>
<td>8.0</td>
</tr>
<tr>
<td>( \geq 4 )</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.03</td>
<td>3.64</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>10</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>5.13</td>
<td>50.00</td>
</tr>
</tbody>
</table>

** the heuristically assumed value, integer numbers—the observed values, real numbers—the expected values. Parameters: \( \alpha_1 = 4.88, \alpha_2 = 2.25, \alpha_3 = 0, p_1 = 0.2, p_2 = 0.606, \chi^2 = 68.97, df = 49, U = 1.90. **

Grouping the neighbouring cells into eight subclasses indicated by frames, we obtain \( \chi^2 = 5.72, df = 2. **

any other parameters of marginal distributions. Therefore for the parameter estimation of the bivariate distribution we modify the parameters of the marginal distributions with an element of heuristic influence.

In terms of our models the difference in the sex ratio of nematodes may result either from the difference in the infection or from the difference in the dilution parameters. The differences both in the sizes of parasites (males are smaller than females) as well as in the ages of birds (birds acquire the immunity against successive infections caused by the same species of parasites) may be of some importance. The univariate Poisson distribution of the number of male nematodes may be justified using the Poisson Theorem, similarly to that employed in many other applied probability models. The negative binomial distribution for the number of female nematodes may be justified by the large variability of the dilution parameter. In biological terms it may be caused by the variable immunity of birds infected by female nematodes. The decreasing proportion of males in distributions conditioned by a fixed total number of nematodes is not statistically significant. It may result from the dependence of the numbers of male and female nematodes during infection. Thus, the assumption of a bivariate Poisson distribution for the number of nematodes is sufficient (Proposition 11).

The positive correlation of the number of nematodes of both sexes follows independently from a few assumptions. Firstly, a deterministic binomial dilution is sufficient (Proposition 3), but Poisson marginal distributions are not obtained. Secondly, the positive correlation of the parameters of the random dilution, even assuming its independence from the extent of infection, with nematodes is sufficient, but then the Poisson marginal distribution is also not obtained (Definition 2). Thirdly, the positive correlation of the number of nematodes during infection or positive correlation of the probabilities of dilution is sufficient (Proposition 8). Note that for our data a Poisson
marginal distribution may be substituted by a negative binomial distribution with statistical tolerance.

REFERENCES


