A PARAMETRIC MODEL FOR HETEROGENEITY IN PAIRED POISSON COUNTS

Constantinos Goutis and
Rex F. Galbraith

95-61

WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 95-61
Statistics and Econometrics Series 25
December 1995

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid 126
28903 Getafe (Spain)
Fax (341) 624-9849

A PARAMETRIC MODEL FOR HETEROGENEITY IN PAIRED POISSON COUNTS

Constantinos Goutis and Rex F. Galbraith*

Abstract

We present a model for data in the form of match pairs of counts. Our work is motivated by a problem in fission track analysis, where the determination of a crystal age is based on the ratio of counts of spontaneous and induced tracks. It is often reasonable to assume that the counts follow a Poisson distribution but, typically, they are overdispersed and there exists a positive correlation between the numbers of spontaneous and induced tracks at the same crystal. We propose a model that allows for both overdispersion and correlation by assuming that the mean densities follow a bivariate Wishart distribution. Our model is quite general, having the usual negative binomial or Poisson models as special cases. We propose a maximum likelihood estimation method based on a stochastic implementation of the EM algorithm and we derive the asymptotic standard errors of the parameter estimates. We illustrate the method by a data set of fission tracks counts in matched areas of zircon crystals.

Key words and phrases : EM algorithm, fission track analysis, maximum likelihood estimation, overdispersion, Wishart distribution.

* Goutis, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, tel. (341) 624-9852, e-mail: *costas@est-econ.uc3m.es*; Galbraith, Department of Statistical Science, University College London, United Kingdom, e-mail: *rex@stats.ucl.ac.uk*.

1 Introduction

Data in the form of matched pairs of counts arise in a variety of applications, such as numbers of grasshopper progeny with and without a B-chromosome (Shaw *et al.*, 1985), treatment and control counts in autoradiography (Blackett and Parry, 1977), numbers of red and white corpuscles in blood samples, numbers of accidents experienced by individuals in successive time periods, and many others. A natural model for such data is that the counts have Poisson distributions with means that may vary both within and between pairs.

In such applications a parameter of interest is often the ratio of the Poisson means, or some function of this ratio, and there is usually another incidental parameter associated with each pair of counts. This situation has been considered recently by Davison (1992) and Morton (1991) to which we refer further below. In this article we consider a parametric model for overdispersion in paired Poisson counts, in the context of modelling "mixed" ages in fission track analysis, though the model is more gererally applicable.

The rest of the paper is organized as follows: Section 2 gives some illustrative data and context. Section 3 briefly discusses extra-binomial models and Section 4 presents a parametric extra-Poisson model. In Section 5 we consider maximum likelihood fitting of the parametric model and we give some concluding remarks in the last section.

2 Data and context

The modelling of "mixed" fission track ages can provide estimates of times and temperatures that are of interest in the oil exploration industry and in various geological applications (see Hurford, 1991, for a recent review).

Table 1 shows a typical set of data, which are numbers of spontaneous and induced fission tracks counted in matched areas of crystal and mica for 24 zircon crystals. Spontaneous tracks form over geological time by spontaneous fission of trace ^{238}U. Induced tracks are created artificially by placing the sample in a nuclear reactor and bombarding it with thermal neutrons, a measured proportion of which collide with trace ^{235}U atoms, thereby causing them to fission. This indirectly measures the amount of trace uranium in the crystal.

Galbraith and Laslett (1993) considered statistical models for such data. It is supposed that the numbers of spontaneous and induced tracks (Y_1, Y_2) counted over matched areas A for a single crystal, have conditionally independent Poisson distributions with means $A\rho_1$ and $A\rho_2$ respectively. In this context the Poisson model is particularly convincing (Galbraith *et al.*, 1990). The spontaneous track density ρ_1 depends on the age of the crystal, the amount of trace ^{238}U it contains, and the mean length of spontaneous tracks. The induced track density ρ_2 depends on the amount of trace ^{235}U and on the mean length of induced tracks; ρ_2 also depends on the thermal neutron dose, which is measured independently. To a close approximation, the ratio ρ_1/ρ_2 is given by

$$\frac{\rho_1}{\rho_2} = \frac{2\lambda_f}{\Phi\sigma_f I}\, t\, \frac{l_1}{l_2} \tag{1}$$

which depends on

Table 1.
Numbers of spontaneous and induced fission tracks counted in matched areas for 27 zircon crystals:

crystal	Y_1	Y_2	area	crystal	Y_1	Y_2	area
1	24	459	80	15	2	70	49
2	8	52	30	16	3	94	28
3	136	310	30	17	23	128	60
4	56	257	70	18	153	264	70
5	3	57	70	19	90	143	32
6	6	332	80	20	31	49	16
7	73	98	14	21	38	120	40
8	131	226	50	22	51	46	25
9	9	173	80	23	38	85	12
10	6	28	12	24	127	45	20
11	141	229	70	25	5	24	30
12	11	74	36	26	24	56	20
13	12	61	18	27	10	31	18
14	10	28	40				

- the crystal's age t, which is a parameter of interest,
- the ratio l_1/l_2 of mean lengths of spontaneous and induced tracks, which reflects the amount of heat the crystal has experienced and is also of interest,
- the ^{235}U:^{238}U isotopic ratio I, which is usually assumed to be fixed at 0.00725, but which conceivably may vary on a microscopic scale, and
- the thermal neutron dose Φ and constants λ_f and σ_f that are independently calibrated.

Typically the amounts of trace uranium and areas vary substantially between crystals (as they do in Table 1) and hence Y_1 and Y_2 will be highly correlated when considering their variation between crystals.

In a sample of crystals the ratios ρ_1/ρ_2 will vary if the crystals have different ages. They may also vary due to the effect of heat (particularly for the mineral apatite), even if all crystals have the same age, because the spontaneous tracks will shorten, possibly by different amounts for different crystals, so that l_1/l_2 varies. Thus it is of interest to develop models that allow for variation between crystals of ρ_1, ρ_2 and of ρ_1/ρ_2.

3 Extra-binomial models

A standard way to model such data is to use the binomial distribution, considering the total $Y_1 + Y_2$ as the "number of trials" and Y_1 as the "number of successes"; that is, to condition on the sum of two Poisson random variables. If θ is the probability of success, then

$$\frac{\theta}{1-\theta} = \frac{\rho_1}{\rho_2}.$$

Then, to model variation between crystals, one can assume a distribution for θ. This approach was taken by Galbraith and Laslett (1993) where θ was assumed to have a logistic normal distribution, a model discussed by Williams (1982), Anderson (1988) and Goutis (1993) among others.

Although this approach simplifies the computations, particularly by avoiding reference to the many nuisance parameters, it is not clear that such a model faithfully represents the reality. With respect to this point, Morton (1991) considered "extra-binomial" models derived from "extra-Poisson" variables conditional on their total. That is, he allowed the Poisson means to vary randomly between pairs, but then he analysed Y_1 conditional on $Y_1 + Y_2$.

In the absence of extra-binomial variation (i.e. when ρ_1/ρ_2 is constant), likelihood based inferences are identical whether one considers Poisson data or binomial data, as $Y_1 + Y_2$ is in some sense ancillary. This situation is analogous to that for contingency tables, where there has been a long-standing discussion on whether or not one should condition on marginal totals, with good arguments on both sides. But if the random variation in ρ_1 and ρ_2 induces variation in the ratio ρ_1/ρ_2, then the statistic $Y_1 + Y_2$ is no longer ancillary. Both the conditional distribution of Y_1 given $Y_1 + Y_2$ and the marginal distribution of $Y_1 + Y_2$ depend on all parameters, so that conditioning on $Y_1 + Y_2$ leads to a loss of information. This argues against the approach of Morton (1991).

Davison (1992) developed a test for treatment effect heterogeneity (i.e. variation of ρ_1/ρ_2) for paired Poisson counts, and briefly discussed a mixture model for inference in the presence of overdispersion. In §4 we propose a different model that seems more straightforward for assessing dispersion.

4 A parametric model

We model the data directly as overdispersed Poisson counts, by assuming a joint distribution for (ρ_1, ρ_2) that allows for the correlation between them, due to their being from the same crystal.

Traditional univariate models for extra-Poisson variation assume that the Poisson mean comes from some distribution, common choices being the gamma and the lognormal. From an empirical point of view both these distributions often yield similar results. For a lognormal mixing distribution, the resulting distribution of the count does not have a tractable form; however this mixing distribution is a natural generalisation of log-linear models, where fixed and random effects are added on the same scale. For a gamma mixing distribution, the count is negative binomial. This choice is also satisfying in that the parameters (ρ_1, ρ_2) are mixed on their natural scale of mean density per unit area.

To model paired data we use a multivariate generalisation of the gamma distribution, namely the Wishart. We introduce an auxiliary random variable ρ_c, to model the covariance, and let the 2×2 matrix random variable

$$R = \begin{pmatrix} \rho_1 & \rho_c \\ \rho_c & \rho_2 \end{pmatrix}$$

have a Wishart distribution $W_2(\nu, M/\nu)$ with degrees of freedom ν and mean matrix M. This has probability density function

$$f(R) = \frac{|R|^{\frac{\nu-3}{2}}\, \nu^{\nu}\, \exp\{-\frac{\nu}{2}\,\mathrm{tr}(M^{-1}R)\}}{2^{\nu}\sqrt{\pi}\,|M|^{\frac{\nu}{2}}\,\Gamma(\frac{\nu}{2})\,\Gamma(\frac{\nu-1}{2})} \tag{2}$$

defined for R positive definite. In (2), M is a positive definite matrix defined as

$$M = \begin{pmatrix} \mu_1 & \mu_c \\ \mu_c & \mu_2 \end{pmatrix} \tag{3}$$

where $E(\rho_1) = \mu_1$, $E(\rho_2) = \mu_2$ and $E(\rho_c) = \mu_c$. The parameter μ_c determines the correlation between ρ_1 and ρ_2 as

$$\mathrm{corr}\,(\rho_1, \rho_2) = \frac{\mu_c^2}{\mu_1 \mu_2} \tag{4}$$

which is necessarily positive. The fourth parameter ν (the degrees of freedom) is also positive and describes the variation of ρ_1, ρ_2 and ρ_c about their means; in particular each of them has coefficient of variation equal to $\sqrt{2/\nu}$.

Note that the main parameter of interest is μ_1/μ_2 since we are interested in the fission track age, while $\mu_1 + \mu_2$ is of no direct interest. The squared coefficient of variation of the ratio ρ_1/ρ_2 is approximately

$$\begin{aligned} \mathrm{cv}^2(\rho_1/\rho_2) &\approx \mathrm{cv}^2(\rho_1) + \mathrm{cv}^2(\rho_2) - 2\,\mathrm{corr}(\rho_1, \rho_2)\,\mathrm{cv}(\rho_1)\,\mathrm{cv}(\rho_2) \\ &= \frac{4}{\nu}\left(1 - \frac{\mu_c^2}{\mu_1 \mu_2}\right). \end{aligned} \tag{5}$$

The square root of (5) is also of interest because it quantifies the relative variation of ρ_1/ρ_2 between crystals, and hence it quantifies the variation of fission track ages t, assuming the other factors in (1) are constant.

In this model, the marginal distributions of Y_1 and Y_2 are negative binomial with the same index ν. Their bivariate distribution has mean $(A\mu_1, A\mu_2)$ and covariance matrix given by

$$\begin{pmatrix} A\mu_1 + \frac{2}{\nu}A^2\mu_1^2 & \frac{2}{\nu}A^2\mu_c^2 \\ \frac{2}{\nu}A^2\mu_c^2 & A\mu_2 + \frac{2}{\nu}A^2\mu_2^2 \end{pmatrix}.$$

This is a more general bivariate negative binomial distribution than that defined in Johnson and Kotz (1969, page 292). The latter has just three parameters and corresponds to the special case $\mu_c^2 = \mu_1 \mu_2$ discussed below. In our context, the fourth parameter (essentially μ_c) allows for the important possibility that the ratio ρ_1/ρ_2 might vary between pairs.

Summarising, for data Y_{1i}, Y_{2i}, $i = 1, 2, \ldots, n$ we have

$$Y_{1i} \sim \text{Poisson}(A_i \rho_{1i}), \qquad Y_{2i} \sim \text{Poisson}(A_i \rho_{2i})$$

independently given R_i, and

$$R_i = \begin{pmatrix} \rho_{1i} & \rho_{ci} \\ \rho_{ci} & \rho_{2i} \end{pmatrix} \sim \mathrm{W}_2\left(\nu, \frac{M}{\nu}\right)$$

where M is given by (3).

4.1 Submodels

It is interesting to see how the model degenerates as the parameters take values on the boundary of the parameter space.

When $\mu_c^2 = \mu_1\mu_2$ the matrix M is singular and $\text{corr}(\rho_1, \rho_2) = 1$. This implies that the ratio $\lambda = \rho_2/\rho_1$ is the same for all crystals, corresponding to the absence of extra-binomial variation, though not of extra-Poisson variation. Marginally, ρ_1 and ρ_2 have gamma distributions with the same index $\nu/2$ amd means μ_1 and $\lambda\mu_1$ respectively (indeed $\rho_2 = \lambda\rho_1$ where λ is fixed). Johnson and Kotz (1969, page 292) defined a negative multinomial distribution as a k-variate version of this case.

When $\mu_c = 0$ then ρ_1 and ρ_2 have independent gamma distributions with the same index $\nu/2$ amd means μ_1 and μ_2. Hence the counts for the same pair (Y_1, Y_2) have independent negative binomial distributions with the same index $\nu/2$. This model was discussed, though not in the context of paired data, by Bliss and Owen (1958). It is equivalent to the model considered by Morton (1991, §2) who then went on to condition on $Y_1 + Y_2$.

When $\nu \to \infty$ the variances of ρ_1, ρ_2 and ρ_c are zero, corresponding to the absence of extra-Poisson variation. In this case Y_1 and Y_2 have independent Poisson distributions with means $A\mu_1$ and $A\mu_2$.

When $\nu \to 0$ the variances of ρ_1 and ρ_2 become infinite. Then estimation of ρ_1/ρ_2 is impossible as there is too much noise in the data. There is also a technical problem with the Wishart distribution because the density (2) does not exist for $\nu \leq 1$. Hence small estimates of ν should be a warning that the data are uninformative.

5 Estimation of parameters

We develop a computational procedure to compute maximum likelihood estimates and their asymptotic precisions for the full model and for the various sub-models, using a stochastic implementation of the EM algorithm (Dempster, Laird and Rubin, 1977).

5.1 Log likelihood

For the four parameter model the log-likelihood, apart from an additive constant, is

$$L(\nu, M) = \sum_{i=1}^{n} \log \left\{ \int \rho_{1i}^{Y_{1i}}\, \mathrm{e}^{-A_i\rho_{1i}}\, \rho_{2i}^{Y_{2i}}\, \mathrm{e}^{-A_i\rho_{2i}}\, f(R_i)\, dR_i \right\} \tag{6}$$

where $f(R)$ is given by (2) and the integral dR_i is taken over the domain $\rho_{1i} \geq 0$, $\rho_{2i} \geq 0$, $\rho_{ci}^2 \leq \rho_{1i}\rho_{2i}$.

Direct maximisation of L is daunting, so we turn to other computational techniques, in particular to the EM algorithm, which is often suitable for mixture models. For the M-step we suppose that we have observed $R_i, i = 1, 2, \ldots, n$, so we find M and ν to maximise

$$\sum_{i=1}^{n} \log f(R_i)$$

thereby avoiding the integrals in (6). For the E-step we "estimate" R_i from the data and current parameter values, using appropriate conditional expectations. Starting with trial values of M and ν and iterating between the E-step and M-step eventually gives convergence to the maximum likelihood estimates of M and ν. One can also derive the information matrix from the EM algorithm using the method of Louis (1982) or its stochastic implementation (Wei and Tanner, 1990), for which only the gradient and second derivatives of the complete likelihood are needed. Computational details are given in §5.3.

For the special case $\mu_c^2 = \mu_1\mu_2$ there are just three parameters λ, μ_1 and ν, where $\lambda = \mu_1/\mu_2 = \rho_{2i}/\rho_{1i}$. The model simplifies considerably, though the log-likelihood does not have the form (6) because M^{-1} does not exist. The log-likelihood now equals

$$L(\nu, \mu_1, \lambda) = \sum_{i=1}^{n} \left[Y_{2i} \log \lambda - \frac{\nu}{2} \log \mu_1 + \frac{\nu}{2} \log \frac{\nu}{2} - \log \Gamma\left(\frac{\nu}{2}\right) + \log\left(\int_0^\infty h_i(\rho_{1i})\, d\rho_{1i}\right)\right] \qquad (7)$$

where

$$h_i(\rho_{1i}) = \exp\left\{ -A_i\rho_{1i} + Y_{+i} \log \rho_{1i} - \lambda A_i \rho_{1i} - \frac{\nu \rho_{1i}}{2\mu_1} + \left(\frac{\nu}{2} - 1\right) \log \rho_{1i} \right\}.$$

The EM algorithm described in §5.3 can easily be modified to maximise (7) by considering ρ_{1i} as the missing data. Alternatively the integral in (7) can be expressed is a multiple of a Gamma function:

$$\int_0^\infty h_i(\rho_{1i})\, d\rho_{1i} = \left(\lambda A_i + A_i + \frac{\nu}{2\mu_1}\right)^{-(Y_{+i} + \frac{\nu}{2})} \Gamma\left(Y_{+i} + \frac{\nu}{2}\right)$$

so that (7) becomes a familiar negative binomial log likelihood:

$$\begin{aligned} L(\nu, \mu_1, \lambda) &= Y_{2+} \log \lambda - n\frac{\nu}{2} \log \mu_1 + n\frac{\nu}{2} \log \frac{\nu}{2} - n \log \Gamma\left(\frac{\nu}{2}\right) \\ &\quad - \sum_{i=1}^{n} \left[\left(Y_{+i} + \frac{\nu}{2}\right) \log\left(\lambda A_i + A_i + \frac{\nu}{2\mu_1}\right) - \log \Gamma\left(Y_{+i} + \frac{\nu}{2}\right)\right], \end{aligned} \qquad (8)$$

which can be maximised numerically by standard methods.

When $\mu_c = 0$ the log-likelihood is again a standard negative binomial form:

$$\begin{aligned} L(\nu, \mu_1, \mu_2) &= -n\frac{\nu}{2} \log(\mu_1\mu_2) + n\nu \log \frac{\nu}{2} - 2n \log \Gamma\left(\frac{\nu}{2}\right) \\ &\quad - \sum_{i=1}^{n} \sum_{j=1}^{2} \left[\left(Y_{ji} + \frac{\nu}{2}\right) \log\left(A_i + \frac{\nu}{2\mu_j}\right) - \log \Gamma\left(Y_{ji} + \frac{\nu}{2}\right)\right]. \end{aligned} \qquad (9)$$

Finally for the case $\nu \to \infty$, where Y_{1i}, Y_{2i} have independent Poisson distributions there are just two parameters μ_1 and μ_2 and the corresponding log-likelihood is:

$$L(\mu_1, \mu_2) = Y_{1+} \log \mu_1 + Y_{2+} \log \mu_2 - A_+ (\mu_1 + \mu_2) . \tag{10}$$

5.2 Example

We now present the results for the data in Table 1. Before looking at the numerical answers, we examine a graphical display of the data in Figure 1. The figure shows a radial plot (see Galbraith 1990 for the general principles of radial plots and various applications). We show a scatterplot of the standardised estimates of age, using crude estimates of the ratios ρ_{1i}/ρ_{2i}, against a measure of their precision. In this plot we have used the Anscombe (1948) modified angular transformation

$$z_i = \arctan \sqrt{\frac{Y_{1i} + \frac{3}{8}}{Y_{2i} + \frac{3}{8}}}, \qquad \sigma_i = \frac{1}{2\sqrt{Y_{1i} + Y_{2i} + \frac{1}{2}}} \tag{11}$$

in order to produce a precision nearly independent of the value of the ratio ρ_{1i}/ρ_{2i} (and hence of age). The y-axis represents values of the centered $(z_i - z_0)/\sigma_i$ where z_0 is a pooled estimate of the ratio. The x-axis indicates both the precision $1/\sigma$ and the total number of tracks $Y_1 + Y_2$. The values of age are represented by slopes of lines from (0, 0) through the points and can be read off the circular axis. The age scale is graduated in equal divisions of $\arctan \sqrt{\rho_1/\rho_2}$ rather than ρ_1/ρ_2.

The plot indicates a clear overdispersion. Most points have a y coordinate outside the "two-sigma" band and the age estimates using raw z_i's vary from about 1 to about 100 Ma. They vary continuously, over this range, suggesting that a continuous mixing density for ρ_1 and ρ_2 is indeed appropriate. It should be noted that even data points with high precisions yield varying estimates of age. Most of the high precision pairs, however seem to correspond to small values of ρ_1/ρ_2, with the exception of one outlier. We are not aware of any specific information regarding the latter point, which yields an estimate of age of about 100 Ma.

The numerical results for these data are presented in Table 2. We fitted the general four parameter model as well as the special cases $\mathrm{corr}(\rho_1, \rho_2) = 1$, $\mathrm{corr}(\rho_1, \rho_2) = 0$ and $\nu = +\infty$. Table 2 gives the maximum likelihood estimates of various parameters, including μ_1/μ_2 and the age which are of particular interest. We have also derived the asymptotic standard errors, either directly (see §5.3) or by the delta method, and they are also given in Table 2.

The first five rows of Table 2 give the maximum likelihood estimates of the main parameters and their asymptotic standard errors for each model, along with the maximised value of the log likelihood. The full model is clearly superior to the submodel with $\mu_c = 0$; the asymptotic likelihood ratio test statistic is $2 \times (742.933 - 738.724) = 8.42$ to be compared with $\chi^2(1)$. The submodel with $\mu_c^2 = \mu_1\mu_2$ is much worse-fitting by this criterion, and the Poisson model ($\nu \to \infty$) has an appalling fit. This implies that, for

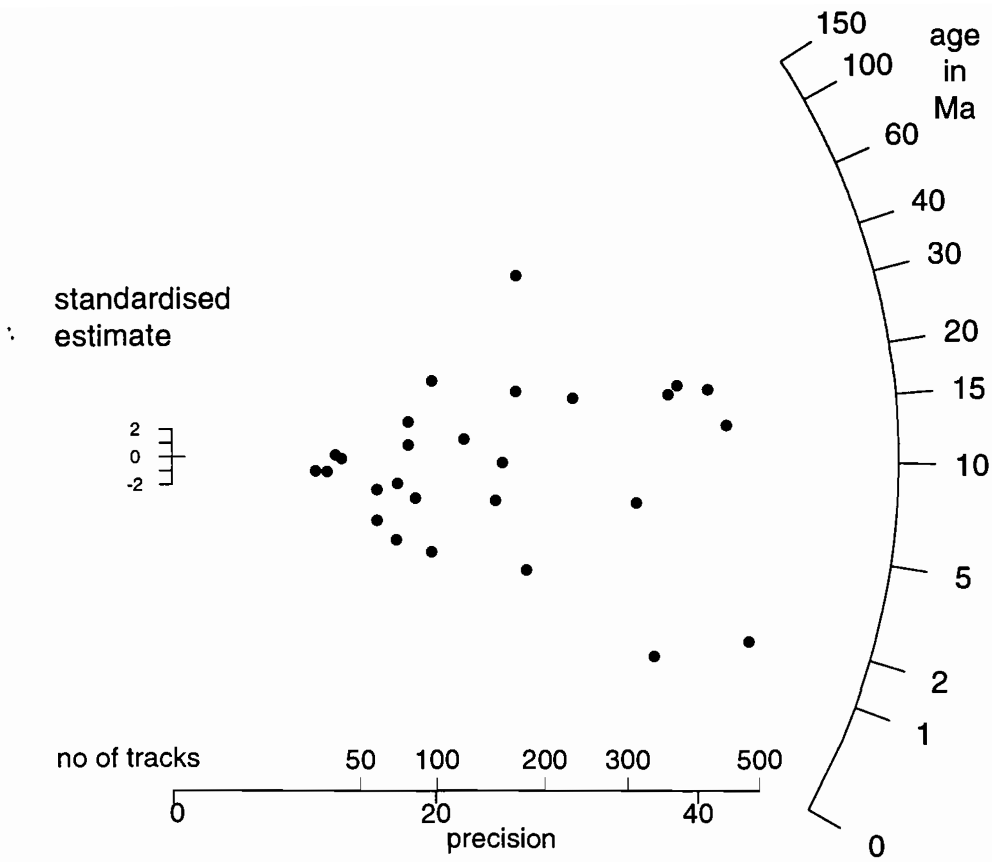

Figure 1: Radial plot of the fission track data in Table 1.

Table 2. Maximum likelihood estimates.

parameter	Full model est.	s.e.	Sub-models $\mu_c = 0$ est.	s.e.	$\mu_c^2 = \mu_1\mu_2$ est.	s.e.	$\nu \to \infty$ est.	s.e.
μ_1	1.46	0.27	1.46	0.26	1.23	0.16	1.11	0.03
μ_2	3.32	0.59	3.31	0.58	3.56	0.46	3.22	0.05
μ_c	1.58	0.39	0	–	2.09	0.27	–	–
ν	2.35	0.42	2.42	0.41	4.45	1.15	∞	–
$L_{\max}$	742.933		738.724		395.219		−497.135	
$\sqrt{2/\nu}$	0.92	0.08	0.91	0.07	0.67	0.03	0	–
$\mathrm{corr}(\rho_1, \rho_2)$	0.51	0.15	0	–	1	–	–	–
$\mathrm{cv}(\rho_1/\rho_2)$	0.91	0.93	1.29	0.48	0	–	0	–
μ_1/μ_2	0.440	0.079	0.439	0.110	0.345	0.011	0.345	0.011
age (Ma)	13.25	2.38	13.22	3.31	10.39	0.34	10.39	0.34

these data, a model should account for extra-Poisson variation, correlation, as well as variation of the ratio of the track densities between crystals.

The next three rows of Table 2 give the values of various functions of the parameters to aid interpretation. The value of the coefficient of variation $\sqrt{2/\nu} = 0.92$ for the full model represents a very substantial variability of the true track densities between crystals. For the submodel with $\mu_c = 0$ this quantity is similar, but for $\mu_c^2 = \mu_1\mu_2$ it is reduced somewhat to 0.67. For the Poisson model this variation is forced to be zero, which accounts for the very poor fit of this model.

The value of $\mathrm{corr}(\rho_1, \rho_2) = 0.51$ for the full model, calculated using equation (4), is distinctly between 0 and 1, the values for the first two submodels. Also the coefficient of variation of ρ_1/ρ_2, calculated from (5) is substantial for the first two models (91% and 129%), indicating that the ratio of densities varies between crystals, whereas for the last two models this quantity is zero by assumption.

The last two rows of Table 2 give the estimate of μ_1/μ_2 and the corresponding age estimate for each model. The estimates are the same for the last two models (independent Poisson and a single negative binomial) although their estimates of μ_1 and μ_2 differ. In both cases, though, the standard errors are grossly under-estimated. This is typical in the cases where overdispersion is erroneously omitted. The full model gives a more sensible estimate and standard error.

5.3 Computational details

The "missing data" for our implementation of the EM algorithm will be the matrices R_i, $i = 1, 2, \ldots, n$. Note that given R_i, the data Y_{1i}, Y_{2i} are independent of the parameters M, ν so the M-step involves only these matrices. The E-step is not straightforward, so we use the stochastic implementation of Wei and Tanner (1989). More specifically:

M-step. We maximise

$$\frac{\nu}{2}\sum_i \log|R_i| - \frac{\nu}{2}\mathrm{tr}\left(M^{-1}\sum_i R_i\right) + n\nu\log\frac{\nu}{2} - \frac{n\nu}{2}\log|M| - n\log\Gamma\left(\frac{\nu}{2}\right) - n\log\Gamma\left(\frac{\nu-1}{2}\right)$$

with respect to M and ν. Maximising first with respect to M gives

$$\hat{M} = \frac{1}{n}\sum_i R_i\,.$$

Hence the profile log likelihood of ν, apart from multiplication by n, is

$$\frac{\nu}{2}\left(\frac{1}{n}\sum_i \log|R_i| - \log\left|\frac{1}{n}\sum_i R_i\right|\right) + \nu\log\frac{\nu}{2} - \nu - \log\Gamma\left(\frac{\nu}{2}\right) - \log\Gamma\left(\frac{\nu-1}{2}\right).$$

This has a unique maximum and indeed is sharply peaked. It is an easy computational problem to find $\hat{\nu}$ by direct numerical maximisation of this.

E-step. The method needs $E[\sum_i \log|R_i|]$ and $E[\mathrm{tr}(M^{-1}\sum_i R_i)]$ where the expectation is taken over the distribution of R_i given the current parameter estimates $(\hat{\nu}, \hat{M})$ and all the data. The R_i's are conditionally independent and the conditional distribution of each R_i depends only on Y_{1i}, Y_{2i} and the parameters M and ν. It is hard to compute these expectations because the conditional density function has the form

$$f(R|Y_1, Y_2, \nu, M) \propto \rho_1^{Y_1}\rho_2^{Y_2}|R|^{\frac{\nu-3}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}\left[\left(\nu M^{-1} + 2AI\right)R\right]\right\}$$

where we omit the subscript i. Nevertheless, it is easy to simulate from this distribution using the acceptance-rejection method.

To simulate from $f(R|Y_1, Y_2, \nu, M)$ we reparameterise (ρ_1, ρ_2, ρ_c) to $(\rho_1, \rho_2, \tau\sqrt{\rho_1\rho_2})$. Then the density is proportional to

$$\rho_1^{Y_1+\frac{\nu}{2}-1}\rho_2^{Y_2+\frac{\nu}{2}-1}\left(1-\tau^2\right)^{\frac{\nu-3}{2}}\exp\left\{-\frac{1}{2}\left(\kappa_{11}\rho_1 + \kappa_{22}\rho_2 + 2\kappa_{12}\tau\sqrt{\rho_1\rho_2}\right)\right\}$$

where $\kappa_{11}, \kappa_{12}, \kappa_{22}$ are the entries of the symmetric matrix $K = \nu M^{-1} + 2AI$. Then we simulate:

$$\begin{aligned}
\rho_1 &\sim \mathrm{Gamma}\left(Y_1 + \frac{\nu}{2}, A + \frac{\nu(\mu_2 - \delta)}{2|M|}\right) \\
\rho_2 &\sim \mathrm{Gamma}\left(Y_2 + \frac{\nu}{2}, A + \frac{\nu(\mu_1 - \delta)}{2|M|}\right) \\
\tau &\sim f(\tau) \propto \left(1-\tau^2\right)^{\frac{\nu-3}{2}}
\end{aligned}$$

independently, and accept the simulated values of ρ_1, ρ_2, τ (and hence the simulated R) with probability

$$\exp\left\{-\frac{\nu}{2|M|}\left[\delta(\rho_1+\rho_2)-2\mu_c\tau\sqrt{\rho_1\rho_2}\right]\right\}. \tag{12}$$

The constant δ is chosen so that (12) is always less than 1; note that it equals 1 when $\rho_1 = \rho_2 = 0$. It turns out that (12) is a decreasing function of δ, and that (12) will be less than 1 if $\delta > |\mu_c|$. In our experience the acceptance probability (12) can become small, particularly if the marginal distribution of τ is concentrated near 1 or -1.

The observed information matrix can be derived using the method of Louis (1982) or Wei and Tanner (1990). Denote by L_i the contribution to the complete data log likelihood from pair i, i.e. let

$$L_i = \frac{\nu}{2}\log|R_i| - \frac{\nu}{2}\operatorname{tr}\left(M^{-1}R_i\right) + \nu\log\frac{\nu}{2} - \frac{\nu}{2}\log|M| - \log\Gamma\left(\frac{\nu}{2}\right) - \log\Gamma\left(\frac{\nu-1}{2}\right)$$

and write

$$\begin{aligned}
H_{1i} = \frac{\partial\operatorname{tr}(M^{-1}R_i)}{\partial\mu_1} &= \frac{\rho_{2i}}{|M|} - \frac{\mu_2}{|M|}\operatorname{tr}\left(M^{-1}R_i\right)\\
H_{2i} = \frac{\partial\operatorname{tr}(M^{-1}R_i)}{\partial\mu_2} &= \frac{\rho_{1i}}{|M|} - \frac{\mu_1}{|M|}\operatorname{tr}\left(M^{-1}R_i\right)\\
H_{ci} = \frac{\partial\operatorname{tr}(M^{-1}R_i)}{\partial\mu_c} &= -\frac{2\rho_{ci}}{|M|} + \frac{2\mu_c}{|M|}\operatorname{tr}\left(M^{-1}R_i\right).
\end{aligned}$$

Then straightforward algebra yields:

$$\begin{aligned}
\frac{\partial L_i}{\partial\nu} &= \frac{1}{2}\left\{\log\frac{|R_i|}{|M|} + 2\log\frac{\nu}{2} + 2 - \operatorname{tr}\left(M^{-1}R_i\right) - \psi\left(\frac{\nu}{2}\right) - \psi\left(\frac{\nu-1}{2}\right)\right\}\\
\frac{\partial L_i}{\partial\mu_1} &= -\frac{\nu}{2}H_{1i} - \frac{\nu}{2}\frac{\mu_2}{|M|}\\
\frac{\partial L_i}{\partial\mu_2} &= -\frac{\nu}{2}H_{2i} - \frac{\nu}{2}\frac{\mu_1}{|M|}\\
\frac{\partial L_i}{\partial\mu_c} &= -\frac{\nu}{2}H_{ci} + \frac{\nu\mu_c}{|M|}
\end{aligned}$$

where $\psi(x)$ is the derivative of $\log\Gamma(x)$. The second derivatives of L_i are:

$$\begin{aligned}
\frac{\partial^2 L_i}{\partial\nu^2} &= \frac{1}{\nu} - \frac{1}{4}\psi'\left(\frac{\nu}{2}\right) - \frac{1}{4}\psi'\left(\frac{\nu-1}{2}\right)\\
\frac{\partial^2 L_i}{\partial\nu\partial\mu_1} &= -\frac{1}{\nu}\frac{\partial L_i}{\partial\mu_1}\\
\frac{\partial^2 L_i}{\partial\nu\partial\mu_2} &= -\frac{1}{\nu}\frac{\partial L_i}{\partial\mu_2}\\
\frac{\partial^2 L_i}{\partial\nu\partial\mu_c} &= -\frac{1}{\nu}\frac{\partial L_i}{\partial\mu_c}
\end{aligned}$$

$$\frac{\partial^2 L_i}{\partial \mu_1^2} = -\frac{\nu\mu_2}{|M|}\left(H_{1i} + \frac{\mu_2}{2|M|}\right)$$

$$\frac{\partial^2 L_i}{\partial \mu_2^2} = -\frac{\nu\mu_1}{|M|}\left(H_{2i} + \frac{\mu_1}{2|M|}\right)$$

$$\frac{\partial^2 L_i}{\partial \mu_1 \partial \mu_2} = \frac{\nu(\mu_1 H_{1i} + \mu_2 H_{2i})}{2|M|} + \frac{\nu\mu_1\mu_2}{2|M|^2} - \frac{\nu[1 - \mathrm{tr}(M^{-1}R_i)]}{2|M|}$$

$$\frac{\partial^2 L_i}{\partial \mu_1 \partial \mu_c} = \frac{\nu(\mu_2 H_{ci} + \mu_c H_{1i})}{2|M|} - \frac{\nu\mu_2\mu_c}{|M|^2}$$

$$\frac{\partial^2 L_i}{\partial \mu_2 \partial \mu_c} = \frac{\nu(\mu_1 H_{ci} + \mu_c H_{2i})}{2|M|} - \frac{\nu\mu_1\mu_c}{|M|^2}$$

$$\frac{\partial^2 L_i}{\partial \mu_c^2} = -\frac{2\nu\mu_c}{|M|}\left(H_{ci} - \frac{\mu_c}{|M|}\right) + \frac{\nu[1 - \mathrm{tr}(M^{-1}R_i)]}{|M|}$$

The Hessian of the log likelihood can be found by applying Wei and Tanner (1989, formula[1] 3.3) of using the simulation method described above.

6 Discussion

The model that we have put forward here is appropriate for paired Poisson data where overdispersion is present and it allows for correlation of the Poisson counts in the same pair. The main interest in the motivating example lies on the estimation of the ratio of the means of the counts but our model is more generally applicable. Various special cases are of interest. In particular, we can model pure overdispersion or pure correlation of the data in a straightforward manner. This immediately leads to an easy computation of the corresponding likelihood ratio statistics which allow for testing for positive correlation or overdispersion or both. Score tests seem more difficult to costruct.

Parameter estimation in this model is feasible, though not straightforward. The EM algorithm is a natural candidate method for maximum likelihood estimation, since we can view our model as a mixture of distributions. The stochastic version of EM algorithm allows the computation of intractable expectations. Other numerical maximisation methods such as Newton-Raphson that have been proposed for negative binomial likelihoods might also be used but we suspect that their implementation will be problematic since the likelihood does not have a closed form. It may be worthwhile developing approximate methods for computing maximum likelihood estimates.

The Wishart distribution that we used as a mixing distribution is simply one of the various existing choices. The bivariate lognormal could also have been used. In our context, it is rather inappropriate, since we would like the means of the Poisson counts to be additive in the crystal area. A bivariate lognormal distribution on the track densities would not yield expected counts proportional to the area. Other candidate mixing distributions include the various generalisations of the gamma in higher dimensions. We suspect that there is not much to differentiate between them, except that they are lesser known and analytically less tractable than the Wishart distribution.

[1] This formula should have + instead of $\times$ in the second line.

7 References

ANDERSON, D.A. (1988) Some models for overdispersed binomial data. *Austral. J. Statist,* **30**, 125-148.

ANSCOMBE, F.J. (1948) The transformation of Poisson, binomial, and negative-binomial data. *Biometrika,* **35**, 246-254.

BLACKETT, N.M. & PARRY, D.M. (1977) A simplified method of 'hypothetical grain' analysis of electron microscope autoradiographs. *J. Histchem. Cytochem.,* **25**, 206-214.

BLISS, C.I. & OWEN, A.R.G. (1958) Negative binomial distributions with a common *k*. *Biometrika,* **45**, 37-58.

DAVISON, A.C. (1992) Treatment effect heterogeneity in paired data. *Biometrika,* **79**, 463-474.

DEMPSTER, A.P., LAIRD, N.M. & RUBIN, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B,* **39**, 1-22.

GALBRAITH, R.F. (1994) Some applications of radial plots. *J. Amer. Statist. Assoc.,* **89**, 1232-1242.

GALBRAITH, R.F. & LASLETT, G.M. (1993) Statistical models for mixed fission track ages. *Nuclear Tracks and Radiation Measurements,* **21**, 459-470.

GALBRAITH, R.F., LASLETT, G.M., GREEN, P.F. & DUDDY, I.R. (1990) Apatite fission track analysis: geological thermal history analysis based on a three dimensional random process of linear radiation damage. *Phil. Trans. R. Soc. Lond.* A, **332**, 419-438.

GOUTIS, C. (1993) Recovering extra-binomial variation. *J. Statist. Comput. Simul.,* **45**, 233-242.

HURFORD, A.J. (1991) Uplift and cooling pathways derived from fission track analysis and mica dating: a review. *Geologisches Rundschau,* **80**, 349-368.

JOHNSON, N.L.& KOTZ, S. (1969) *Distributions in Statistics: Discrete Distributions.* Wiley, NY.

LOUIS, T.A. (1982) Finding the observed information matrix when using the EM algorithm. *J. R. Statist. Soc.* **B**, 226-233.

MORTON, R. (1991) Analysis of extra-multinomial data derived from extra-Poisson variables conditional on their total. *Biometrika,* **78**, 1-6.

SHAW, M.W., HEWITT, G.M. & ANDERSON, D.A. (1985) Polymorphism in the rates of meiotic drive acting on the B-chromosome of *Myrmeleotettix maculatus. Heredity,* **55**, 61-68.

WEI, G.C.G. & TANNER, M.A. (1990) A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J. Amer. Statist. Assoc.,* **85**, 699-704.

WILLIAMS, D.A. (1982) Extra-binomial variation in logistic linear models. *Appl. Statist.,* **31**, 144-148.