CENTRE FOR **STOCHASTIC GEOMETRY**
AND ADVANCED **BIOIMAGING**

Ege Rubak, Jesper Møller and Peter McCullagh

Statistical Inference for a Class of Multivariate Negative Binomial Distributions

No. 08, August 2010

Statistical Inference for a Class of Multivariate Negative Binomial Distributions

EGE RUBAK[1,*] JESPER MØLLER[1,**] and PETER MCCULLAGH[2],

[1] *Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7G, DK-9220 Aalborg Ø, Denmark E-mail:* [*] rubak@math.aau.dk; [**] jm@math.aau.dk.

[2] *Department of Statistics, University of Chicago, 5734 University Avenue, Chicago 60637, U.S.A. E-mail:* pmcc@galton.uchicago.edu.

This paper considers statistical inference procedures for a class of models for positively correlated count variables called α-permanental random fields, and which can be viewed as a family of multivariate negative binomial distributions. Their appealing probabilistic properties have earlier been studied in the literature, while this is the first statistical paper on α-permanental random fields. The focus is on maximum likelihood estimation, maximum quasi-likelihood estimation and on maximum composite likelihood estimation based on uni- and bivariate distributions. Furthermore, new results for α-permanents and for a bivariate α-permanental random field are presented.

Keywords: α-permanent, α-permanental random field, composite likelihood, doubly stochastic construction, maximum likelihood, quasi-likelihood.

1. Introduction

Møller and Rubak (2010) provided a review of a class of models for positively correlated count variables $\mathbf{N} = (N_1, \ldots, N_m)$, which possess a number of appealing properties. This model class was referred to as α-permanental random fields, since it is a special case of the class of general α-permanental point processes which have been the subject of much research interest in recent years, see Macchi (1971, 1975), Shirai and Takahashi (2003a,b), Georgii and Yoo (2005), and McCullagh and Møller (2006). As each count variable N_i follows a negative binomial distribution, an α-permanental random field may be referred to as a multivariate negative binomial distribution. The probabilistic properties of these multivariate distributions have been studied in detail in Griffiths and Milne (1987), Vere-Jones (1997), and Møller and Rubak (2010), but to the best of our knowledge no statistical inference based on the models have been conducted. In this paper we develop statistical inference procedures using the full likelihood, quasi-likelihood or composite likelihoods.

Section 2 introduces the notation and provides the necessary background material. Section 3 describes the inferential procedures, and Section 4 illustrates their use for analyzing two different data sets. Technicalities are deferred to Appendix A, which, among other things, establishes a new result concerning the joint density of any two count variables (N_i, N_j).

2. The α-permanental random field

This section contains a very brief introduction to the necessary background material about the α-permanental random field. We mainly follow the notation and terminology of Møller and Rubak (2010), and further details can be found therein.

We start by recalling the definition of the α-permanent of an $n \times n$ matrix A with entries $A_{i,j}$,

$$\mathrm{per}_\alpha(A) = \sum_{\sigma \in \mathcal{S}_n} \alpha^{c(\sigma)} A_{1,\sigma(1)} A_{2,\sigma(2)} \cdots A_{n,\sigma(n)},$$

where $\mathcal{S}_n$ is the set of all permutations of $1, \ldots, n$, and $c(\sigma)$ denotes the number of cycles in σ. In a more general setup, it may be convenient to work with the related α-determinant $\det_\alpha(A) = \alpha^n \mathrm{per}_{1/\alpha}(A)$ as in Møller and Rubak (2010), but it is not necessary here. In general the α-permanent is very expensive computationally, and apart from a few special cases it can only be approximated (see Appendix A for details).

The distribution of an α-permanental random field $\mathbf{N} = (N_1, \ldots, N_m)$ is specified by a positive real parameter α and a real $m \times m$ matrix C, and we write $\mathbf{N} \sim \mathrm{per}(\alpha, C)$. Throughout this paper we assume that the matrix

$$\tilde{C} = \alpha C (I + \alpha C)^{-1} \tag{1}$$

exists. As discussed below, further restrictions need to be satisfied by (α, C) or by $(\alpha, \tilde{C})$

to ensure the existence of the distribution $\mathrm{per}(\alpha, C)$. Then, for

$$\mathbf{n} = (n_1, \ldots, n_m) \in \{0, 1, \ldots\}^m, \quad n_\star = \sum_{i=1}^{m} n_i,$$

the probability function is given by

$$p(\mathbf{n}) = \frac{|I - \tilde{C}|^{1/\alpha}}{\prod_{i=1}^{m} n_i!} \mathrm{per}_{1/\alpha}(\tilde{C}[\mathbf{n}]), \tag{2}$$

where $\tilde{C}[\mathbf{n}]$ is the $n_\star \times n_\star$ block matrix obtained from $\tilde{C}$ by repeating the i'th index n_i times (cf. Section A.2 for further details). Marginally each N_i follows a negative binomial distribution with mean $\mathrm{E}N_i = C_{i,i}$ and variance $\mathrm{Var}N_i = C_{i,i} + \alpha C_{i,i}^2$. Furthermore, $\mathrm{Cov}(N_i, N_j) = \alpha C_{i,j} C_{j,i} \geq 0$ for $i \neq j$, so all correlations are non-negative. The parameter α influences both the amount of over-dispersion and the strength of correlation between variables. In particular these decrease as α tends to zero and the limiting distribution is Poisson with independent components regardless of the matrix C. No combination of parameters (α, C) exists such that the components of $\mathbf{N}$ are Poisson variables with positive correlation. However, over-dispersion without correlation is possible, in which case the components are independent negative binomial variables. In other words, the α-permanental model is such that, if there is correlation among the counts, over-dispersion will also be present. The over-dispersion factor for each N_i is $1 + \alpha \mathrm{E}(N_i)$.

In this paper we mainly consider the case where the following doubly stochastic construction applies: First, let $\mathbf{X} = (X_1, \ldots, X_m)$ follow a certain multivariate gamma distribution denoted $\Gamma_m(\alpha, C)$, where Proposition 4.5 in Vere-Jones (1997) gives a sufficient and necessary condition for the existence of this multivariate gamma distribution, but the following sufficient condition (C1) is simpler to use:

(C1) C is a covariance matrix and $\alpha \in \left(0, \frac{2}{m-1}\right] \cup \left\{\frac{2}{m-2}, \frac{2}{m-3}, \ldots, 1, 2\right\}$.

Under (C1), $\mathbf{X}$ is distributed as the diagonal of a Wishart matrix with $2/\alpha$ degrees of freedom and mean C, so marginally X_i is gamma distributed with $\mathrm{E}X_i = C_{i,i}$ and $\mathrm{Cov}(X_i, X_j) = \alpha C_{i,j} C_{j,i}$ (Møller and Rubak, 2010, Section 4.1). Second, conditionally on $\mathbf{X}$, let the N_i's be independent Poisson random variables with $\mathrm{E}(N_i \mid X_i) = X_i$.

Under the doubly stochastic scheme, for $k = 1, 2, \ldots$ and given an observation of

$\mathbf{N} = \mathbf{n}$, the Bayes estimate of the k'th moment of the unknown mean X_i is

$$\begin{aligned}
\mathrm{E}(X_i^k \mid \mathbf{n}) &= \frac{1}{p(\mathbf{n})} \int_{\mathbb{R}_+^m} x_i^k p(\mathbf{n} \mid \mathbf{x}) p(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\
&= (n_i + 1) \cdots (n_i + k) \frac{1}{p(\mathbf{n})} \int_{\mathbb{R}_+^m} \frac{x_i^{n_i+k}}{(n_i+k)!} e^{-x_i} \prod_{j \neq i} \Big[\frac{x_j^{n_j}}{n_j!} e^{-x_j} \Big] p(\mathbf{x}) \, \mathrm{d}\mathbf{x} \\
&= (n_i + 1) \cdots (n_i + k) \frac{p(\mathbf{n}_i^k)}{p(\mathbf{n})} \\
&= \frac{\mathrm{per}_{1/\alpha}(\tilde{C}[\mathbf{n}_i^k])}{\mathrm{per}_{1/\alpha}(\tilde{C}[\mathbf{n}])}, \qquad (3)
\end{aligned}$$

where $\mathbf{n} = (n_1, \ldots, n_m)$ and $\mathbf{n}_i^k = (n_1, \ldots, n_{i-1}, n_i + k, n_{i+1}, \ldots, n_m)$. Furthermore, if D is a diagonal matrix with diagonal entries $D_{i,i} = \sqrt{a_i}$ where $a_i \geq 0$, then

$$(X_1, \ldots, X_m) \sim \Gamma_m(\alpha, C) \;\Rightarrow\; (a_1 X_1, \ldots, a_m X_m) \sim \Gamma_m(\alpha, DCD). \qquad (4)$$

As noted in Vere-Jones (1997) the doubly stochastic construction is not necessary for the existence of the α-permanental random field: there are (α, C) such that $\mathrm{per}(\alpha, C)$ exists, but a corresponding gamma random field $\Gamma_m(\alpha, C)$ does not exist. Another sufficient condition for the existence of $\mathrm{per}(\alpha, C)$ is

(C2) $\tilde{C}$ has non-negative entries and all eigenvalues have modulus less than 1

(Vere-Jones (1997); Møller and Rubak (2010)). When (C1) is satisfied, simulation of first $\mathbf{X}$ and second $\mathbf{N}$ is easily done by the doubly stochastic construction. If (C2) but not (C1) is satisfied, a Poisson randomization can be used for simulation (Møller and Rubak, 2010, Section 4.2).

3. Inference

3.1. Full likelihood

Given a realization $\mathbf{n}$ of an α-permanental random field with a parametric model for the matrix $C = C_\psi$, where ψ is a real d-dimensional parameter, note that $\tilde{C} = \tilde{C}_\theta$ depends on $\theta = (\alpha, \psi)$, cf. (1). In principle, we can evaluate the log-likelihood

$$\ell(\alpha, \psi; \mathbf{n}) = \frac{1}{\alpha} \log |I - \tilde{C}_\theta| + \log \mathrm{per}_{1/\alpha}(\tilde{C}_\theta[\mathbf{n}]) \qquad (5)$$

on a grid of (α, ψ) in order to obtain the maximum likelihood estimate (MLE) $(\hat{\alpha}, \hat{\psi})$ (provided it exists). Further, for each grid point (α, ψ), we have access to the log-likelihood ratio $\lambda(\alpha, \psi) = 2(\ell(\hat{\alpha}, \hat{\psi}) - \ell(\alpha, \psi))$, which may be compared with quantiles of the χ^2_{d+1} distribution to find approximate confidence regions.

However, as mentioned previously and discussed in Appendix A, exact calculation of the α-permanent is usually not tractable, and in fact even approximate calculation may be computationally expensive. Furthermore, the grid evaluation requires some knowledge of the range of (α, ψ) values to include in the grid. Therefore we study composite likelihoods which both serve as a computationally simple method for inference in its own right and can be used for initializing the grid evaluation of the full likelihood.

3.2. Composite likelihood

Composite likelihoods have been extensively studied in many connections, see e.g. Lindsay (1988) and Cox and Reid (2004). Here we outline how composite likelihood methods can be used for the α-permanental random field model, using either the univariate or the bivariate distributions.

Given an observation $\mathbf{n}$ and a parametric model as in Section 3.1, we define the *first-order composite log-likelihood* by

$$\ell^1(\theta) = \sum_{i=1}^{m} \log p_i(n_i \,|\, \theta), \tag{6}$$

where p_i is the marginal probability function for N_i. It corresponds to the log-likelihood for m independent negative binomial random variables. In this case likelihood inference can be done using an iterative Newton-Raphson procedure and efficient software implementations are readily available (Venables and Ripley, 2002, Section 7.4). Depending on the parametric model for C, some parameters may be unidentifiable using this procedure, since only the diagonal elements of C enter in the first-order composite log-likelihood, as exemplified in Section 4.1. Due to the computational simplicity of this composite log-likelihood, it is well suited for initialization of the parameters in more complicated methods.

In a similar manner as above, we define the *pairwise composite log-likelihood* by

$$\ell^2(\theta) = \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \log p_{i,j}(n_i, n_j \,|\, \theta), \tag{7}$$

where $p_{i,j}$ denotes the bivariate probability function for (N_i, N_j). These bivariate distributions have been thought to be quite complicated, cf. the discussion in Griffiths and Milne (1987), and previously it was not possible to use these bivariate distributions in practice. However, in Appendix A.2.3 we give a computationally simple formula for calculation of the relevant α-permanent. The resulting bivariate probability function is

$$p_{i,j}(n_i, n_j) = \frac{\left(\frac{a_i}{b}\right)^{n_i} \left(\frac{a_j}{b}\right)^{n_j} \Gamma(\frac{1}{\alpha} + n_i)\Gamma(\frac{1}{\alpha} + n_j)}{b^{\frac{1}{\alpha}} n_i! n_j! \Gamma(\frac{1}{\alpha})\Gamma(\frac{1}{\alpha})} \sum_{k=0}^{n_i \wedge n_j} \binom{n_i}{k} \binom{n_j}{k} \frac{k!\Gamma(\frac{1}{\alpha})}{\Gamma(\frac{1}{\alpha} + k)} c^{2k}, \tag{8}$$

where

$$a_i = \alpha^2(C_{i,i}C_{j,j} - C_{i,j}^2) + \alpha C_{i,i}, \quad a_j = \alpha^2(C_{i,i}C_{j,j} - C_{i,j}^2) + \alpha C_{j,j},$$
$$b = \alpha^2(C_{i,i}C_{j,j} - C_{i,j}^2) + \alpha(C_{i,i} + C_{j,j}) + 1, \quad c = \frac{\alpha C_{i,j}}{\sqrt{a_i a_j}}.$$

This makes it practically feasible to implement the pairwise composite log-likelihood for statistical inference.

In many applications there is a distance function or neighbourhood structure attached to the domain, or index set, of the field. For example, when modeling spatial regions some regions will share a boundary and will be called neighbours. In this way there will also be a natural notion of higher order neighbours, such that regions not sharing a boundary but with a common neighbour are second order neighbours etc. The part of the pairwise composite log-likelihood (7) corresponding to contributions from k'th order neighbours is denoted

$$\ell_k^2(\theta) = \sum_{(i,j)\in\mathcal{P}(k)} \log p_{i,j}(n_i, n_j \,|\, \theta),$$

where $\mathcal{P}(k)$ denotes the set of distinct pairs (i, j) that are k'th order neighbours. It may then be interesting to use the *k'th order pairwise composite log-likelihood*

$$\ell_{<k}^2(\theta) = \sum_{l=1}^{k} \ell_l^2(\theta).$$

Note that the pairwise composite log-likelihood defined in (7) corresponds to including neighbours of all orders and we may write $\ell^2(\theta) = \ell_{<\infty}^2(\theta)$.

3.3. Quasi-likelihood

As an alternative to composite likelihood inference based on low dimensional marginal distributions as above we may consider inference based on low order moments. For an α-permanental random field the factorial moments are given by α-permanents and are especially tractable for low orders (see Vere-Jones, 1997; Møller and Rubak, 2010).

The quasi-likelihood as introduced by Wedderburn (1974) has been widely used in the literature and has a well developed asymptotic theory (cf. McCullagh, 1983). In the following we detail how to apply quasi-likelihood methods for α-permanental random fields, and only briefly recall the necessary general results.

As an initial step we construct a vector

$$\mathbf{Y} = (N_i, N_i(N_i - 1), N_{ij})_{\{i=1,\ldots,m\,;\, i<j\le m\}},$$

and denote the length of $\mathbf{Y}$ by n. We do not necessarily include products of all pairs of counts $N_i N_j$ with $j > i$; we may only consider a subset based on neighbour relations. Note, that the mean $\mu = \mu(\theta) = \mathrm{E}_\theta(\mathbf{Y})$ and the covariance matrix $\Sigma = \Sigma_\theta = \mathrm{Cov}_\theta(\mathbf{Y})$

can be expressed in terms of factorial moments of order at most 4, which are easily evaluated analytically.

Let $D = D_\theta$ be the $n \times d$ derivative matrix with entries $D_{ir} = \partial\mu_i/\partial\theta_r$. Then the quasi-likelihood estimating function for θ is

$$\mathbf{U}(\theta; \mathbf{Y}) = D_\theta^\top \Sigma_\theta^{-1}(\mathbf{Y} - \mu(\theta)),$$

which has zero expectation and covariance matrix $V = V_\theta = \mathrm{Cov}(\mathbf{U}) = D^\top \Sigma^{-1} D$. The quasi-likelihood estimator $\hat{\theta} \equiv \hat{\theta}(\mathbf{Y})$ is the root of the vector equation $\mathbf{U}(\hat{\theta}) = \mathbf{0}$, which can be found iteratively. Using a modified Newton-Raphson scheme the current estimate $\hat{\theta}^{(i)}$ is updated to

$$\hat{\theta}^{(i+1)} = \hat{\theta}^{(i)} + V_{\hat{\theta}^{(i)}}^{-1} \mathbf{U}(\hat{\theta}^{(i)}; \mathbf{Y}).$$

The iterative procedure is stopped once the estimate has converged within a specified tolerance. Under regularity conditions the quasi-likelihood estimator is asymptotically Gaussian with covariance matrix V^{-1}, which is calculated in each step of the iterative procedure. This allows us to attach an asymptotic variance V_θ^{-1} to the quasi-likelihood estimate.

4. Examples

4.1. One dimensional example

Figure 1 shows counts of clover leaves in 200 squares of size 5×5 cm along a 10 m transect line as detailed in Augustin et al. (2006). This data can be viewed as a realization of a one-dimensional random field consisting of 200 sites on the real line, with positive association expected between the counts due to the multiplicity of leaves per plant and the clustering of plants in patches. We model the leaf counts $\mathbf{N} = (N_1, \ldots, N_{200})$ as $\mathbf{N} \sim \mathrm{per}(\alpha, C)$, where $C_{i,j} = \kappa\rho^{|i-j|}$ with $0 \le \rho \le 1$ and $\kappa \ge 0$. Then condition (C2) is satisfied (Møller and Rubak, 2010, Proposition 1). Furthermore, by arguments similar to those used in the proof of Proposition 1 in Møller and Rubak (2010), it can be shown that for all $\alpha > 0$, C satisfies a regularity condition (Vere-Jones, 1997, Proposition 4.5) implying the existence of $\mathbf{X} \sim \Gamma_m(\alpha, C)$ so that $\mathrm{per}(\alpha, C)$ has a doubly stochastic construction, cf. Section 2. Therefore, it makes sense to calculate the Bayes estimate $\mathrm{E}_\theta(\mathbf{X} \,|\, \mathbf{n})$ of the conditional intensity for all positive α. The Bayes estimate for the model using the MLE as found below is superimposed as a line in Figure 1.

As an initial step in the parameter estimation we use the first-order composite log-likelihood. Notice, since $\ell^1(\alpha, \kappa, \rho)$ is independent of ρ, it is not possible to estimate this parameter using ℓ^1. Using the iterative Newton-Raphson procedure of Venables and Ripley (2002), the estimate of $\log(\kappa)$ is 0.247 ± 0.257 and the estimate of $1/\alpha$ is 0.396 ± 0.141, where both estimates are quoted plus/minus two standard errors. The point estimates correspond to $\hat{\kappa} = 1.28$ and $\hat{\alpha} = 2.5$. For a grid of parameter values evaluation of the full log-likelihood yielded the MLE $(\hat{\alpha}, \hat{\kappa}, \hat{\rho}) = (2.3, 1.28, 0.860)$. A three dimensional approximate 95% confidence region can be found by calculating the

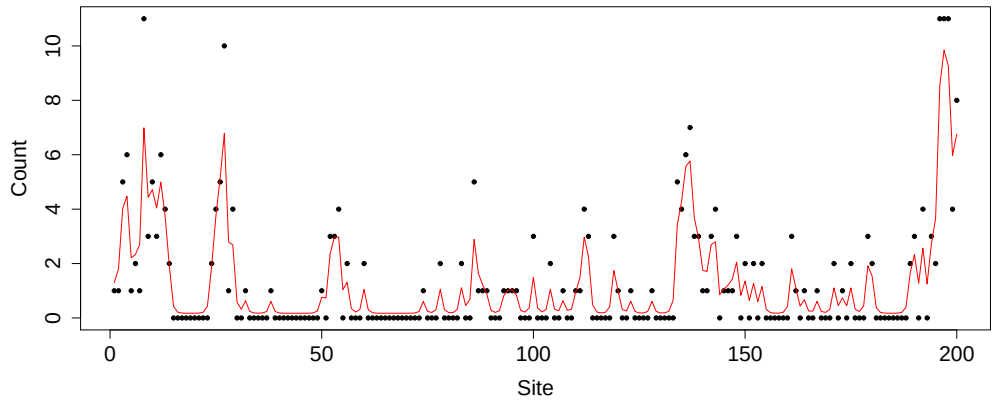


Figure 1: Counts of clover leaves in 200 square patches with Bayes estimate of the random mean field superimposed as a line.

likelihood ratio $\lambda(\alpha,\kappa,\rho) = 2(\ell(\hat{\alpha},\hat{\kappa},\hat{\rho}) - \ell(\alpha,\kappa,\rho))$ for all points of the parameter grid and compare with the 95th percentile of the χ^2_3-distribution. Marginal confidence intervals are $(1.4, 4.4)$ for α, $(0.7, 4.7)$ for κ, and $(0.8, 0.95)$ for ρ. To visualize the confidence region in two dimensions Figure 2a shows a contour plot of $\lambda(\alpha,\hat{\kappa},\rho)$ as a function of (α,ρ) with κ fixed at the MLE $\hat{\kappa} = 1.28$. The contours are based on the 50th, 95th, and 99th percentile of the χ^2_3-distribution. Figures 2b-2d are similar contour plots based on ℓ^2_1, ℓ^2_9 and $\ell^2_{<\infty}$ with κ fixed at $\hat{\kappa} = 1.28$. In these plots the contours are no longer related to any confidence regions. It is clear that ℓ^2_1 in Figure 2b determines ρ quite well, and the higher order neighbour pairs do not contain much information about ρ. A plot of the empirical autocorrelation function (not shown) reveals that it is negative for neighbours of order 9, which explains the shape of the contour plot in Figure 2c, where the maximum is at $\rho = 0$. The pairwise composite log-likelihood with neighbours of all orders is a sum of many composite log-likelihoods, where ρ is poorly determined for the majority of them, which causes the shape of the contour plot in Figure 2d. However, the point estimates of the parameters other than ρ do not change much when inference is based on $\ell^2_{<k}$ for growing k. Based on $\ell^2_{<2}$ the estimates are $(\hat{\alpha},\hat{\kappa},\hat{\rho}) = (2.5, 1.28, 0.860)$ whereas $\ell^2_{<\infty}$ yields the point estimates $(\hat{\alpha},\hat{\kappa},\hat{\rho}) = (2.5, 1.28, 0.855)$.

Using the modified Newton-Raphson scheme described in Section 3.3 the quasi-likelihood estimates (with corresponding two standard errors) are found to be $\hat{\alpha} = 2.2 \pm 1.5$, $\hat{\kappa} = 1.35 \pm 0.7$ and $\hat{\rho} = 0.85 \pm 0.16$ when only first order neighbours are used. The quasi-likelihood estimates only change slightly when higher order neighbours are used, and they are not quoted here.

The full likelihood calculations have been carried out using the Monte Carlo (MC) importance sampling algorithm of Kou and McCullagh (2009), which provides an estimate of both the α-permanent in (5) and the standard error of this estimate. We used

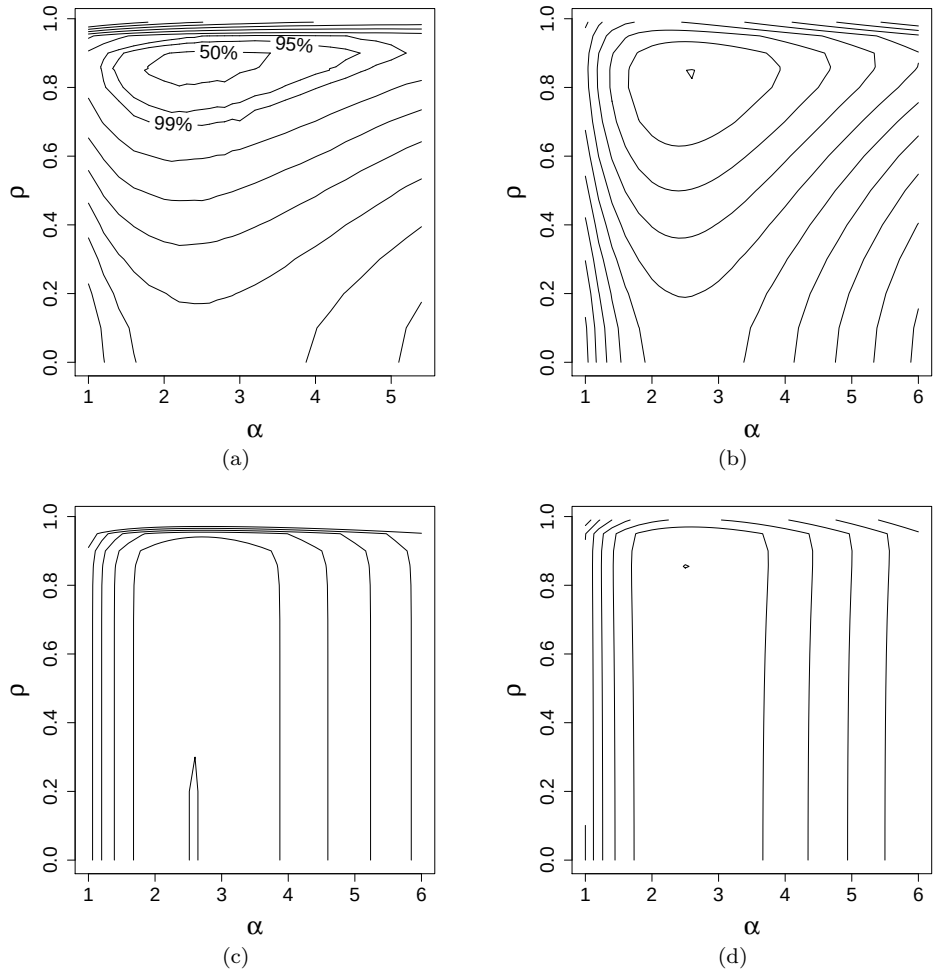


Figure 2: Contour plot of (a) the full log-likelihood, $\ell(\theta)$, compared with contour plots of the pairwise composite log-likelihood with (b) first order neighbours only, $\ell_1^2(\theta)$; (c) ninth order neighbours only, $\ell_9^2(\theta)$; (d) neighbours of all orders, $\ell_{<\infty}^2(\theta)$. For all the plots κ is fixed at the MLE $\hat{\kappa} = 1.28$.

10^5 samples, giving an average relative error (ratio of the standard error and the estimate) of 0.077. As noted in Kou and McCullagh (2009), their algorithm is especially well suited for estimating ratios of α-permanents as required in the Bayes estimate (3). For the calculation used for obtaining Figure 1, 10^4 MC samples were sufficient.

It is possible to perform model validation based on simulation using the Poisson randomization (Møller and Rubak, 2010, Section 4.2). We simulated 100 realizations from the model using the MLE, and checked some properties of the data against the simulated realizations. A characteristic feature for the data is the large number of zeros overall and the apparent clustering of the zeros. For example, the average total number of zeros in the simulated realizations was 111 with the first and third quartile at 103 respectively 119, while data has 114 zeros. The largest cluster of zeros in data is 13 where the simulated realizations have an average of 12 with the first and third quartile at 10 respectively 15. Besides the simulation based validation we also checked empirical first and second order moments of data with the theoretical moments of the fitted model, and they also revealed a very good fit.

In conclusion any of the proposed estimation methods provide good point estimates, but in particular the composite likelihood based approach including neighbours of all orders has a big information loss about the correlation parameter ρ. When it is computationally feasible, as it was the case here, using the full likelihood is preferred.

4.2. Disease mapping

Choo and Walker (2008) presented a so-called multivariate Poisson-Gamma (MPG) model to investigate the spatial variations of cases $\mathbf{n} = (n_1, \ldots, n_m)$ of testis cancer in the $m = 19$ municipalities of the county of Frederiksborg in Denmark, where corresponding expected values $\mathbf{e} = (e_1, \ldots, e_m)$ based on the population and age structures are treated as covariates. For illustrative purposes, we present another approach using α-permanental random fields and leading to the perhaps surprising conclusion that there is little evidence in these data of either over-dispersion or spatial correlation.

The parameters of interest are the incidence ratios γ_i, $i = 1, \ldots, m$, which indicate whether municipality i has an over-representation of testis cancer ($\gamma_i > 1$) or not ($0 < \gamma_i \leq 1$). Specifically, conditional on $\Gamma = (\gamma_1, \ldots, \gamma_m)$, we assume the data is a realization of independently Poisson distributed counts N_i with $\mathrm{E}(N_i \,|\, \Gamma) = \gamma_i e_i$, $i = 1, \ldots, m$. The raw estimates are given by $\hat{\gamma}_i = n_i/e_i$, which agree with the MLE if Γ is a deterministic parameter vector. However, typically Γ would be treated as a random field with spatial dependence, cf. Choo and Walker (2008) and the references therein.

Before proceeding any further, some general remarks about modelling of this type of spatial epidemiological data are needed. In principle, each count N_i can be viewed as the aggregation over an area A_i of an underlying point process specifying the domestic location of each individual diagnosed with the disease. It would be natural to specify a Cox point process model for this underlying data process, where the random intensity at location x, $\gamma(x)$ has mean $e(x)$, which is the known age-adjusted population density at x. Then, conditional on γ the counts N_i are independent Poisson variables with mean

$\int_{A_i} \gamma(x)\,\mathrm{d}x$. The distributional properties of this integral are usually intractable, and it is a well-known unsolved problem in the literature to specify a point process model where inference based on aggregated count data is tractable (see Richardson, 2003; Møller, 2003). A common approach, which we follow here, is simply to specify a model directly in terms of the aggregated data without considering a consistent underlying point process model. However, an important point to be derived from this discussion is that the model should respect geographic integrity, namely that the marginal distribution for a subset of the aggregated data should belong to the same class.

We assume $\Gamma \sim \Gamma_m(\alpha, R)$ where R is a correlation matrix. This ensures that $\mathrm{E}(N_i) = \mathrm{E}(\gamma_i e_i) = e_i$, as one may naturally require. Let $C = DRD$ with D diagonal and $D_{i,i} = \sqrt{e_i}$. We consider a doubly stochastic construction as in Section 2 with $\mathbf{X} = D\Gamma D \sim \Gamma_m(\alpha, C)$ and $\mathbf{N} \sim \mathrm{per}(\alpha, C)$, cf. (4). Moreover, assuming that $\alpha \in \left(0, \frac{2}{m}\right] \cup \left\{\frac{2}{m-1}, \frac{2}{m-2}, \ldots, 2\right\}$, condition (C1) is satisfied, and so the model is well defined.

The final stage of the model is to specify the off diagonal entries of R which determine the correlation structure of the model. A natural approach is to use a neighbourhood relation when specifying R, and we assume that

$$R_{i,j} = \begin{cases} \rho & \text{if } i \sim j \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where for the present data, $i \sim j$ indicates that municipalities i and j share a border. Care must be taken to ensure R is indeed semi-definite; we realized empirically that R is only semi-definite if $0 \leq \rho \leq \rho_c$, where $\rho_c \leq 1$ is a critical value depending on the neighbourhood structure. The critical value can be approximated before any inference is carried out, e.g. by using a spectral decomposition, which for the data at hand revealed $\rho_c \approx 0.416$.

In the special case $\rho = 0$, the model reduces to m independent negative binomial random variables, and so the full log-likelihood is equivalent to the first-order composite log-likelihood. For this model α is the only parameter, and it is straightforward to find the Bayes estimates

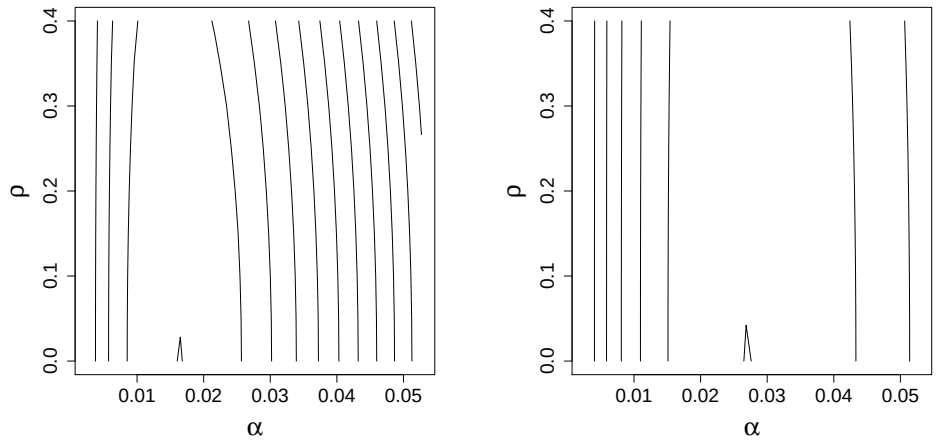
$$\mathrm{E}(\gamma_i \,|\, \mathbf{n}) = \frac{1 + \hat{\alpha} n_i}{1 + \hat{\alpha} e_i} \quad i = 1, \ldots, m.$$

The MLE of $1/\alpha$ is 36.2 ± 69.2 leading to the point estimate $\hat{\alpha} = 0.0277$. The large value of twice the standard error indicates that a negative binomial model is not necessary and a likelihood ratio test against the simpler Poisson null model is performed. The negative binomial model has $-2\ell(\hat{\alpha}) = 107.66$ whereas the Poisson model (corresponding to $\alpha = 0$) has $-2\ell(0) = 105.44$, and the likelihood ratio test yields a p-value of about 14%. Similarly, the standard Pearson χ^2 test for over-dispersion yields the test statistic of $\sum (n_i - e_i)^2 / e_i = 25.5$ on 18 degrees of freedom, for a p-value of about 11%. In other words, there is little evidence of either over-dispersion or spatial correlation.

If ρ is included as a parameter in the model, either full, quasi- or pairwise composite likelihood inference can be used. However, in this example the MC importance sampling algorithm of Kou and McCullagh (2009) used to estimate the α-permanent performs

poorly; even for a very large number of MC samples (10^8) the standard error of the estimate is relatively large. On the other hand, both quasi- and pairwise composite likelihood inference is fast and does not require any approximation (apart from the inherent surrogate nature of these methods).

For the quasi-likelihood iterative scheme ρ quickly approaches zero at which point the covariance matrix V becomes singular, so no standard errors can be given. However, α stabilizes at 0.027 ± 0.064 making it clear that $\alpha = 0$ is well within two standard errors of the estimate. Figure 3a shows a contour plot of the pairwise composite log-



(a) (b)

Figure 3: Contour plots based on pairwise composite log-likelihood using (a) first order neighbours only (b) neighbours of all orders.

likelihood based on first order neighbours only, whereas the contour plot in Figure 3b is based on neighbours of all orders. Notice that in both cases the correlation parameter ρ is poorly determined and the maximal composite likelihood value is attained at $\rho = 0$ confirming the findings of the quasi-likelihood method. Furthermore, it appears that Figure 3b contains less information about ρ than Figure 3a. This is explained by the fact that ρ only enters in bivariate distributions of directly neighbouring sites, and all the terms of $\ell^2_{<\infty}$ not appearing in ℓ^2_1 are independent of ρ. The estimate of α is respectively 0.0165 and 0.0268 when using ℓ^2_1 and $\ell^2_{<\infty}$. Thus, it seems preferable to use $\ell^2_{<\infty}$ to estimate α since it yields an estimate close to the MLE for $\rho = 0$.

For this dataset the main interest is in estimating the incidence ratios γ_i, which is done by calculating the Bayes estimates $\mathrm{E}_{\hat{\theta}}(\gamma_i \mid \mathbf{n})$ under the fitted model. Table 1 lists these estimates for each model as well as the estimates for the MPG model in Choo and Walker (2008). The model with $\rho = \rho_c$ is included for illustrative purposes and for both this model and the independent negative binomial model with $\rho = 0$ the value of α is fixed at 0.0277. The table reveals that estimates based on the MPG model are

Table 1. Bayes estimates of the incidence ratios for the two α-permanental models with $\rho = \rho_c$ and $\rho = 0$ compared with raw Poisson estimates and MPG estimates of Choo and Walker (2008).

	n_i	e_i	raw	$\rho = 0$	$\rho = \rho_c$	MPG
Allerød	18	17.61	1.02	1.01	0.97	1.01
Birkerød	17	18.20	0.93	0.98	1.01	1.00
Farum	14	13.65	1.03	1.01	1.02	0.99
Fredensborg-Humlebæk	14	14.29	0.98	0.99	0.97	0.93
Frederikssund	21	13.17	1.59	1.16	1.17	1.14
Frederiksværk	14	14.63	0.96	0.99	0.99	0.98
Græsted-Gilleleje	13	12.38	1.05	1.01	0.98	0.93
Helsinge	8	13.66	0.59	0.89	0.89	0.86
Helsingør	31	47.18	0.66	0.81	0.81	0.73
Hillerød	28	27.23	1.03	1.01	1.00	0.98
Hundested	8	6.44	1.24	1.04	1.22	1.03
Hørsholm	28	17.04	1.64	1.21	1.03	1.23
Jægerspris	4	6.05	0.66	0.95	0.98	0.97
Karlebo	12	13.78	0.87	0.96	1.01	0.99
Skibby	6	4.57	1.31	1.03	1.10	1.09
Skævinge	6	4.28	1.40	1.04	0.98	1.02
Slangerup	3	6.44	0.47	0.92	1.05	0.95
Stenløse	13	10.47	1.24	1.05	0.95	1.05
Ølstykke	14	10.93	1.28	1.06	1.06	1.11

close to estimates based on the the independent negative binomial model lending further support to the findings that a complex model is unnecessary for this particular dataset. In conclusion, it appears that it suffices to use the model with no spatial dependence between incidence ratios, which was not touched upon by Choo and Walker (2008).

To calculate the Bayes estimates for the model with $\rho = \rho_c$, ratios of α-permanents are again needed, but this poses no significant problem, since the MC importance sampling algorithm estimates these well even though the individual α-permanents are difficult to estimate.

5. Discussion

For the dataset of counts of clover leaves in Section 4.1 the α-permanental random field model with an exponential covariance matrix provides a good fit. Estimation based on both the full, quasi- and pairwise composite likelihood gives similar point estimates, but the shape of the pairwise composite likelihood is sensitive to the choice of neighbourhood order included in the model. This adds the disadvantage of having to choose the neighbourhood order when using the pairwise composite likelihood, while the quasi-likelihood appears to be less sensitive to this choice. In the analysis of this dataset, it is noticeable that the Bayes estimate of the random mean field in Figure 1 is spiky, which may be caused by the choice of covariance model. An immediate advantage of using the exponential covariance model is that the α-permanental model is well defined for all values of $\alpha \geq 0$. For a general covariance model the largest generally admissible value of α is 2. However, it may be possible to find covariance models allowing for $\alpha > 2$ as it was the

case for the exponential covariance model. Alternatively, it may be possible to obtain a good fit with α fixed at 2 using an alternative covariance model of e.g. polynomial type, which would be expected to yield a smoother Bayes estimate of the random mean field.

The dataset of testis cancer cases in Section 4.2 illustrates a simple yet important fact: There is little point in using a complicated model with over-dispersion and spatial dependence if the data shows evidence of neither. However, the example still allows us to illustrate the potential use of the α-permanental model for disease mapping.

Acknowledgments

Supported by the NSF, grant no. DMS-0906592, by the Danish Natural Science Research Council, grants 272-06-0442 and 09-072331('Point process modelling and statistical inference'), by the Danish Agency for Science, Technology and Innovation, grant 645-06-0528, 'International PhD student', and by Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from the Villum Foundation.

References

Augustin, N. H., J. McNicol, and C. A. Marriott (2006). Using the truncated auto-Poisson model for spatially correlated counts of vegetation. *J. Agric. Biol. Environ. Stat. 11*, 1–23.

Choo, L. and S. G. Walker (2008). A new approach to investigating spatial variations of disease. *J. Roy. Statist. Soc. Ser. A 171*, 395–405.

Cox, D. R. and N. Reid (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika 91*, 729–737.

Georgii, H.-O. and H. J. Yoo (2005). Conditional intensity and Gibbsianness of determinantal point processes. *J. Stat. Phys. 118*, 617–666.

Griffiths, R. C. and R. K. Milne (1987). A class of infinitely divisible multivariate negative binomial distributions. *J. Multivariate Anal. 22*, 13–23.

Hammersley, J. M. and D. C. Handscomb (1964). *Monte Carlo Methods.* London: Methuen.

Kou, S. C. and P. McCullagh (2009). Approximating the α-permanent. *Biometrika 96*, 635–644.

Lindsay, B. G. (1988). Composite likelihood methods. In N. U. Prabhu (Ed.), *Statistical Inference from Stochastic Processes*, Providence, pp. 221–239. American Mathematical Society.

Macchi, O. (1971). Stochastic point processes and multicoincidences. *IEEE Trans. Inform. Theory 17*, 2–7.

Macchi, O. (1975). The coincidence approach to stochastic point processes. *Adv. in Appl. Probab. 7*, 83–122.

Maybee, J. and J. Quirk (1969). Qualitative problems in matrix theory. *SIAM Rev. 11*, 30–51.

McCullagh, P. (1983). Quasi-likelihood functions. *Ann. Statist. 11*, 59–67.

McCullagh, P. and J. Møller (2006). The permanental process. *Adv. in Appl. Probab. 38*, 873–888.

Minc, H. (1978). *Permanents*. Reading, MA: Addison-Wesley.

Møller, J. (2003). A comparison of spatial point process models in epidemiological applications. In P. Green, N. Hjort, and S. Richardson (Eds.), *Highly Structured Stochastic Systems*, pp. 264–268. Oxford: Oxford University Press.

Møller, J. and E. Rubak (2010). A model for positively correlated count variables. *Int. Stat. Rev. 78*(1), 65–80.

Richardson, S. (2003). Spatial models in epidemiological applications. In P. Green, N. Hjort, and S. Richardson (Eds.), *Highly Structured Stochastic Systems*, pp. 237–259. Oxford: Oxford University Press.

Shirai, T. and Y. Takahashi (2003a). Random point fields associated with certain Fredholm determinants I: fermion, Poisson and boson point processes. *J. Func. Anal. 205*, 414–463.

Shirai, T. and Y. Takahashi (2003b). Random point fields associated with certain Fredholm determinants II: fermion shifts and their ergodic and Gibbs properties. *Ann. Probab. 31*, 1533–1564.

Sweet, R. A. (1969). A recursive relation for the determinant of a pentadiagonal matrix. *Communications of the ACM 12*, 330–332.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S*. New York: Springer.

Vere-Jones, D. (1997). Alpha-permanents and their applications to multivariate gamma, negative binomial and ordinary binomial distributions. *New Zealand J. Math. 26*, 125–149.

Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika 61*, 439–447.

Appendix A: Evaluating α-permanents

In this appendix we both present some general results for α-permanents (Appendix A.1) and some results on simple patterned matrices (Appendices A.2-A.3) as well as illustrate how an existing algorithm for approximating α-permanents in some cases may be improved (Appendix A.4).

A.1. Preliminary results

Here we give a few general results for α-permanents, which we will need later.

A.1.1. Expansion by sums of cyclic products

For any positive integer n let $I_n = (1, \ldots, n)$, and let I_0 denote the "empty subsequence". Given a positive integer $m \leq n$, let $I = (i_1, \ldots, i_m)$ be an ordered subsequence of I_n meaning that $1 \leq i_1 < \cdots < i_m \leq n$, and let $I^c = (j_1, \ldots, j_{n-m})$ denote the complementary subsequence so that $\{i_1, \ldots, i_m\}$ and $\{j_1, \ldots, j_{n-m}\}$ are disjoint with union $\{1, \ldots, n\}$. For any such I we let $\mathcal{I}(r, I)$ denote the class of ordered subsequences of I of length $r \geq 0$ using the convention $\mathcal{I}(0, I) = \{I_0\}$.

For any $n \times n$ matrix A, we define A_I as the $m \times m$ submatrix of A with (k, l)'th entry A_{i_k, i_l}. Furthermore, we let $\mathrm{cyp}(A_I)$ denote the sum of cyclic products of length $|I| = m$ formed from A_I. Thus, $\mathrm{cyp}(A_I)$ is a sum over $(m-1)!$ terms, and if e.g. $m = 3$ we have

$$\mathrm{cyp}(A_I) = A_{i_1,i_2} A_{i_2,i_3} A_{i_3,i_1} + A_{i_1,i_3} A_{i_3,i_2} A_{i_2,i_1}.$$

Maybee and Quirk (1969) provides the following formula for calculating the determinant of a $n \times n$ matrix A.

Theorem 1. *For $n > 1$ and any fixed $I \in \mathcal{I}(n-1, \{1, \ldots, n\})$,*

$$|A| = A_{I^c, I^c} |A_I| + \sum_{r=0}^{n-2} (-1)^{n-1-r} \sum_{J \in \mathcal{I}(r, I)} |A_J| \mathit{cyp}(A_{J^c}),$$

where we define $|A_\emptyset| = 1$.

This result extends straightforwardly to α-permanents.

Corollary 1. *For all $\alpha \in \mathbb{R}$, $n > 1$ and any fixed $I \in \mathcal{I}(n-1, \{1, \ldots, n\})$,*

$$\mathrm{per}_\alpha(A) = \alpha A_{I^c, I^c} \mathrm{per}_\alpha(A_I) + \sum_{r=0}^{n-2} \sum_{J \in \mathcal{I}(r, I)} \alpha \mathrm{per}_\alpha(A_J) \mathit{cyp}(A_{J^c}). \tag{10}$$

where we define $\mathrm{per}_\alpha(A_\emptyset) = 1$.

Proof. From Theorem 1 we know that the right hand side of (10) has all the $n!$ terms of the form $A_{1,\sigma(1)} \cdots A_{n,\sigma(n)}$ and we only need to verify each term is weighted correctly. The first term on the right hand side of (10) is

$$\alpha A_{I^c,I^c} \mathrm{per}_\alpha(A_I) = \alpha A_{I^c,I^c} \sum_{\sigma \in \mathcal{S}_{n-1}} \alpha^{c(\sigma)} \prod_{i=1}^{n-1} (A_I)_{i,\sigma(i)},$$

and since A_{I^c,I^c} introduces a new cycle to all terms the weighting with $\alpha^{c(\sigma)+1}$ is correct. The rest of the terms are

$$\sum_{r=0}^{n-2} \sum_{J \in \mathcal{I}(r,I)} \alpha \mathrm{per}_\alpha(A_J) \mathrm{cyp}(A_{J^c}) = \sum_{r=0}^{n-2} \sum_{J \in \mathcal{I}(r,I)} \alpha \sum_{\sigma \in \mathcal{S}_r} \alpha^{c(\sigma)} \prod_{i=1}^{r} (A_J)_{i,\sigma(i)} \mathrm{cyp}(A_{J^c})$$

and since again exactly one new cycle is introduced by the cyclic product the weight is correct. □

A.1.2. Expansion by cofactors

Let A be a $n \times n$ matrix. By isolating a given element $A_{r,s}$ of $\mathrm{per}_\alpha(A)$ it is obvious that the coefficient of $A_{r,s}$ depends only on the elements of the reduced matrix of order $n-1$ with row r and column s deleted. However, the coefficient is in general not the α-permanent of the reduced matrix, and Vere-Jones (1997) remarks that no simple cofactor expansion of $\mathrm{per}_\alpha(A)$ is known. However, in the following we give a cofactor expansion by a slight modification of the recipe for determinants. The (r,s) minor is a square matrix $A^{r,s}$ of order $n-1$ obtained from A in two steps as follows: First switch rows r and s; then delete column s and row s (row r from the original matrix). The switching of rows is not a part of the standard definition of a minor, but it is needed for $\alpha \neq \pm 1$ to maintain the cycle structure, and is done prior to deletion to avoid ambiguity about labelling. If $r = s$, the first step is nugatory; otherwise if $r \neq s$ the symmetrically opposed component $A_{s,r}$ occurs on the diagonal of $A^{r,s}$, and every other element on the diagonal of the minor also occurs on the diagonal of A. The components of the cofactor matrix $\mathrm{cof}_\alpha(A)$ are defined as

$$\mathrm{cof}_\alpha(A)_{r,s} = \begin{cases} \alpha \mathrm{per}_\alpha(A^{r,s}) & r = s \\ \mathrm{per}_\alpha(A^{r,s}) & \text{otherwise.} \end{cases}$$

On row r, the cofactor expansion of the α-permanent is

$$\begin{aligned} \mathrm{per}_\alpha(A) &= \alpha A_{r,r} \mathrm{per}_\alpha(A^{r,r}) + \sum_{s \neq r} A_{r,s} \mathrm{per}_\alpha(A^{r,s}) \\ &= \sum_{s=1}^{n} A_{r,s} \ \mathrm{cof}_\alpha(A)_{r,s}. \end{aligned} \tag{11}$$

Although the definition of a minor has been modified to suit the general case, for $\alpha = -1$ this reduces to the standard cofactor expansion of a determinant.

A.2. Block matrices

Evaluating multivariate negative binomial probabilities involves the α-permanent of a block matrix, which are studied in this appendix. For any $m \times m$ matrix A and non-negative integers $\mathbf{n} = (n_1, \ldots, n_m)$, let $n_\star = \sum_{i=1}^m n_i$ and define the block matrix $A[\mathbf{n}]$ as the $n_\star \times n_\star$ matrix obtained from A by repeating index i n_i times. For example, if $m = 4$ and $\mathbf{n} = (2, 0, 1, 3)$

$$A[\mathbf{n}] = A[(2,0,1,3)] = \begin{bmatrix} A_{1,1} & A_{1,1} & A_{1,3} & A_{1,4} & A_{1,4} & A_{1,4} \\ A_{1,1} & A_{1,1} & A_{1,3} & A_{1,4} & A_{1,4} & A_{1,4} \\ A_{3,1} & A_{3,1} & A_{3,3} & A_{3,4} & A_{3,4} & A_{3,4} \\ A_{4,1} & A_{4,1} & A_{4,3} & A_{4,4} & A_{4,4} & A_{4,4} \\ A_{4,1} & A_{4,1} & A_{4,3} & A_{4,4} & A_{4,4} & A_{4,4} \\ A_{4,1} & A_{4,1} & A_{4,3} & A_{4,4} & A_{4,4} & A_{4,4} \end{bmatrix}.$$

We call A the generating matrix, $\mathbf{n}$ the block sizes, and $A[\mathbf{n}]$ a m-dimensional block matrix.

A.2.1. One-dimensional block matrices

When A is a 1×1 matrix with element a and the block size is n we have $A[n] = a\mathbf{1}_n$, where $\mathbf{1}_n$ is the $n \times n$ matrix whose elements are all one. This matrix has α-permanent

$$\mathrm{per}_\alpha(\mathbf{1}_n) = \alpha^{\uparrow n} = \alpha(\alpha+1)\cdots(\alpha+n-1),$$

called the ascending factorial function. Furthermore, $\mathrm{per}_\alpha(A[n]) = a^n \alpha^{\uparrow n}$.

A.2.2. Block-diagonal matrices

For a general block-diagonal matrix it is easy to verify that the α-permanent is the product of the α-permanent of the blocks. The special block diagonal matrix with constant blocks can be written as $A[\mathbf{n}]$, where the generator A is a diagonal matrix with diagonal $(a_1, \ldots, a_m)$, and in this case we have

$$\mathrm{per}_\alpha(A[\mathbf{n}]) = \prod_{i=1}^m \mathrm{per}_\alpha(a_i \mathbf{1}_{n_i}) = \prod_{i=1}^m a_i^{n_i} \alpha^{\uparrow n_i}.$$

A.2.3. Two-dimensional block matrix

For two-dimensional block matrices we have the following result allowing efficient calculation of the α-permanent.

Proposition 1. *Let A be a 2×2 matrix and define*

$$\rho = \frac{A_{1,2}A_{2,1}}{A_{1,1}A_{2,2}}.$$

Then

$$\mathrm{per}_\alpha(A[n_1,n_2]) = A_{1,1}^{n_1}A_{2,2}^{n_2}\alpha^{\uparrow n_1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_1\wedge n_2}\frac{n_1^{\downarrow j}\,n_2^{\downarrow j}\,\rho^j}{j!\;\alpha^{\uparrow j}}, \tag{12}$$

where we define $\alpha^{\uparrow n_1}\alpha^{\uparrow n_2}/\alpha^{\uparrow j} = 0$ *when both the numerator and denominator is zero,* $n^{\downarrow j} = n(n-1)\cdots(n-j+1)$*,* $n^{\downarrow 0} = 1$ *and* $\alpha^{\uparrow 0} = 1$*. Thus,* $n^{\downarrow n} = 1^{\uparrow n} = n!$ *and* $\mathrm{per}_\alpha(A[0,0]) = 1$*.*

Proof. As a preliminary, we note the following property of the ascending factorial function:

$$\sum_{r=0}^{n}\alpha^{\uparrow(k+r)}/r! = \frac{\alpha^{\uparrow(n+k+1)}}{n!\,(\alpha+k)} \tag{13}$$

for non-negative integer k such that $\alpha + k \neq 0$. If $k = 0$ the sum is $(\alpha+1)^{\uparrow n}/n!$, which is readily established by induction on n. The result for general k then follows from $\alpha^{\uparrow(k+r)} = \alpha^{\uparrow k}(\alpha+k)^{\uparrow r}$.

Any 2×2 matrix A can be factorized as

$$A = DRD = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}\begin{bmatrix} 1 & \rho_1 \\ \rho_2 & 1 \end{bmatrix}\begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix},$$

where the (possibly complex) numbers d_1, d_2, ρ_1, ρ_2 satisfy $d_1^2 = A_{1,1}$, $d_2^2 = A_{2,2}$, $\rho_1 = A_{1,2}/(d_1d_2)$, and $\rho_2 = A_{2,1}/(d_1d_2)$. Then it can be verified that

$$\mathrm{per}_\alpha(A[n_1,n_2]) = A_{1,1}^{n_1}A_{2,2}^{n_2}\mathrm{per}_\alpha(R[n_1,n_2]), \tag{14}$$

and therefore to prove (12) it is sufficient to show that

$$\mathrm{per}_\alpha(R[n_1,n_2]) = \alpha^{\uparrow n_1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_1\wedge n_2}\frac{n_1^{\downarrow j}\,n_2^{\downarrow j}\,\rho^j}{j!\;\alpha^{\uparrow j}} \tag{15}$$

with $\rho = \rho_1\rho_2$.

For $n_2 > 0$, let $S[n_1,n_2]$ be the matrix obtained from $R[n_1,n_2]$ by replacing the first row by the last row. Then S is square but asymmetric, and the cofactor expansion gives the bivariate permanental recurrence relation

$$\begin{aligned}\mathrm{per}_\alpha(R[n_1+1,n_2]) &= (\alpha+n_1)\mathrm{per}_\alpha(R[n_1,n_2]) + n_2\rho_{12}\mathrm{per}_\alpha(S[n_1+1,n_2-1]),\\ \mathrm{per}_\alpha(S[n_1+1,n_2]) &= (\alpha+n_1)\rho_{21}\mathrm{per}_\alpha(R[n_1,n_2]) + n_2\mathrm{per}_\alpha(S[n_1+1,n_2-1]).\end{aligned}$$

For successively smaller values of n_2, repeated substitution of the second expression into the first eliminates $\mathrm{per}_\alpha(S[...])$, giving the linear recurrence relations

$$\mathrm{per}_\alpha(R[n_1+1,n_2]) = (\alpha+n_1)\mathrm{per}_\alpha(R[n_1,n_2]) + \rho(\alpha+n_1)\sum_{i=1}^{n_2}n_2^{\downarrow i}\mathrm{per}_\alpha(R[n_1,n_2-i]), \tag{16}$$

one equation for each $n_1, n_2 \geq 0$. These equations are linearly independent of full rank, and have a unique solution for any given boundary value $\mathrm{per}_\alpha(R[0,0])$. It follows immediately that $\mathrm{per}_\alpha(R[n,0]) = \alpha^{\uparrow n}\mathrm{per}_\alpha(R[0,0])$, so the desired boundary value is one.

On the assumption that $n_2 \leq n_1$, we now show that (15) is a solution of the system of linear equations (16). We start by noticing from (16) that $\alpha^{\uparrow n_1}$ is a common factor in all terms of $\mathrm{per}_\alpha(R[n_1,n_2])$ implying $\mathrm{per}_\alpha(R[n_1,n_2]) = 0$ when α is a non-positive integer bigger than $-n_2$. This proves the claim for these values of α and in what follows we consider all other values of α. After substituting (15), the right side of (16) becomes

$$\begin{aligned} & \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2}\frac{n_1^{\downarrow j}\,n_2^{\downarrow j}}{j!}\frac{\rho^j}{\alpha^{\uparrow j}} + \rho\alpha^{\uparrow n_1+1}\sum_{i=1}^{n_2} n_2^{\downarrow i}\alpha^{\uparrow n_2-i}\sum_{j=0}^{n_2-i}\frac{n_1^{\downarrow j}\,(n_2-i)^{\downarrow j}}{j!}\frac{\rho^j}{\alpha^{\uparrow j}} \\ = \quad & \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2}\frac{n_1^{\downarrow j}\,n_2^{\downarrow j}}{j!}\frac{\rho^j}{\alpha^{\uparrow j}} + \rho\alpha^{\uparrow n_1+1}\sum_{j=0}^{n_2-1}\binom{n_1}{j}\frac{\rho^j}{\alpha^{\uparrow j}}\sum_{i=1}^{n_2-j} n_2^{\downarrow i}\alpha^{\uparrow n_2-i}(n_2-i)^{\downarrow j}. \end{aligned}$$

On account of the ascending factorial identity (13), the final sum reduces to

$$\sum_{i=1}^{n_2-j} n_2^{\downarrow i}\alpha^{\uparrow n_2-i}(n_2-i)^{\downarrow j} = \alpha^{\uparrow n_2}n_2!/((\alpha+j)(n_2-j-1)!) = \alpha^{\uparrow n_2}n_2^{\downarrow j+1}/(\alpha+j),$$

which simplifies the preceding expression to

$$\begin{aligned} & \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2}\binom{n_1}{j}\frac{n_2^{\downarrow j}\,\rho^j}{\alpha^{\uparrow j}} + \rho\alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2-1}\binom{n_1}{j}\frac{n_2^{\downarrow j+1}\,\rho^j}{(\alpha+j)\alpha^{\uparrow j}} \\ = \quad & \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2}\binom{n_1}{j}\frac{n_2^{\downarrow j}\,\rho^j}{\alpha^{\uparrow j}} + \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2-1}\binom{n_1}{j}\frac{n_2^{\downarrow j+1}\,\rho^{j+1}}{\alpha^{\uparrow j+1}} \\ = \quad & \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2}\binom{n_1}{j}\frac{n_2^{\downarrow j}\,\rho^j}{\alpha^{\uparrow j}} + \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=1}^{n_2}\binom{n_1}{j-1}\frac{n_2^{\downarrow j}\,\rho^j}{\alpha^{\uparrow j}} \\ = \quad & \alpha^{\uparrow n_1+1}\alpha^{\uparrow n_2}\sum_{j=0}^{n_2}\binom{n_1+1}{j}\frac{n_2^{\downarrow j}\,\rho^j}{\alpha^{\uparrow j}}, \end{aligned}$$

showing that (15) satisfies the permanental recurrence equations (16). Since the recurrence equations are linear, any solution satisfying the desired boundary condition $\mathrm{per}_\alpha(A[0,0]) = 1$ is necessarily unique. □

Proposition 1 can be combined with the result on block-diagonal matrices (Section A.2) to calculate the α-permanent of a block-diagonal matrix where each block possibly is a two-dimensional block matrix.

A.2.4. The ordinary permanent ($\alpha = 1$)

For a general m-dimensional block matrix we have the following result in the special case $\alpha = 1$, for which the α-permanent reduces to the ordinary permanent (Minc, 1978).

Proposition 2. *Let $A[\mathbf{n}]$ be a m-dimensional block matrix with block sizes $\mathbf{n} = (n_1, \ldots, n_m)$. Further, let $T_{\mathbf{n}}$ denote the set of all two way tables $k = \{k_{ij}\}_{i,j=1,\ldots,m}$, $k_{ij} \in \mathbb{N}_0$ with row and column totals $n_1, \ldots, n_m$. Then*

$$\mathrm{per}_1(A[\mathbf{n}]) = \sum_{T_{\mathbf{n}}} \prod_{i=1}^{m} (n_i!)^2 \prod_{i,j=1}^{m} \frac{A_{i,j}^{k_{ij}}}{k_{ij}!}$$

Proof. By definition,

$$\mathrm{per}_1(A[\mathbf{n}]) = \sum_{\sigma \in \mathcal{S}_{n_\star}} A[\mathbf{n}]_{1,\sigma(1)} \cdots A[\mathbf{n}]_{n_\star,\sigma(n_\star)}.$$

In each term of the sum the i'th row index must occur exactly n_i times and the j'th column index must occur exactly n_j times. This makes it clear that each term in the sum is of the form

$$\prod_{i=1}^{m} \prod_{j=1}^{m} A_{i,j}^{k_{ij}}, \quad \text{where} \quad \sum_{j=1}^{m} k_{ij} = n_i \quad \text{and} \quad \sum_{i=1}^{m} k_{ij} = n_j. \tag{17}$$

The question is how many times each term of this form occurs in the sum over all permutations. First k_{11} $A_{1,1}$'s must be chosen from $A[\mathbf{n}]$, which can be done in

$$\frac{n_1 n_1 \cdot (n_1-1)(n_1-1) \cdots (n_1-k_{11}+1)(n_1-k_{11}+1)}{k_{11}!}$$

ways. When these are chosen we must choose k_{12} $A_{1,2}$'s, which can be done in

$$\frac{(n_1-k_{11})n_2 \cdot (n_1-k_{11}-1)(n_2-1) \cdots (n_1-k_{11}-k_{12}+1)(n_2-k_{12}+1)}{k_{12}!}$$

ways. We can continue in this fashion and finally find the number of ways to choose the k_{1m} $A_{1,m}$'s. Then we can start a new row and find that the k_{21} $A_{2,1}$'s can be chosen in

$$\frac{n_2(n_1-k_{11}) \cdot (n_2-1)(n_1-k_{11}-1) \cdots (n_2-k_{21}+1)(n_1-k_{11}-k_{21}+1)}{k_{21}!}$$

ways. Continuing in this fashion we see that for $i, j = 1, \ldots, m$ the number of ways to choose the k_{ij} A_{ij}'s is

$$\frac{(n_i-k_{i1}-\cdots-k_{i,j-1})(n_j-k_{1j}-\cdots-k_{i-1,j}) \cdots (n_i-k_{i1}-\cdots-k_{ij}+1)(n_j-k_{1j}-\cdots-k_{ij}+1)}{k_{ij}!}. \tag{18}$$

To find the coefficient for the term in (17) we need to take the product over $i, j = 1, \ldots, m$ of (18). The product of the numerators simplifies considerably and we end up with

$$\frac{\prod_{i=1}^{m} (n_i!)^2}{\prod_{i=1}^{m} \prod_{j=1}^{m} k_{ij}!},$$

and so the result follows. □

In the two-dimensional case Proposition 2 extends to $\alpha > 0$ as detailed in Section A.2.3. It is plausible that this is also the case in the m-dimensional case. More precisely, if we as in the proof of Proposition 1 consider a $m \times m$ matrix R with unit diagonal we conjecture

$$\mathrm{per}_\alpha(R[\mathbf{n}]) = \prod_{i=1}^{m} (n_i! \alpha^{\uparrow n_i}) \sum_{T_{\mathbf{n}}} \prod_{i,j=1}^{m} \frac{R_{i,j}^{k_{ij}}}{k_{ij}!} P(k, \alpha)$$

where P is a (hopefully simple) rational function in k and α necessarily satisfying $P(k, 1) = 1$. However, we have not yet been able to establish the correct form for P even though some promising patterns have been found in low dimensional examples.

A.3. Penta-diagonal matrices

This section generalizes the efficient algorithm of Sweet (1969) for computing the determinant of a penta-diagonal matrix to the more general case of the α-permanent. The development follows the same lines as Sweet (1969).

Let A be a $n \times n$ penta-diagonal matrix (i.e. $A_{i,j} = 0$ for $|i - j| > 2$) and let $n > 3$. By applying Corollary 1 with $I = I_{n-1} = (1, \ldots, n-1)$ we have

$$\mathrm{per}_\alpha(A) = \alpha A_{n,n} \mathrm{per}_\alpha(A_{I_{n-1}}) + \sum_{r=0}^{n-2} \sum_{J \in \mathcal{I}(r, I_{n-1})} \alpha \mathrm{per}_\alpha(A_J) \mathrm{cyp}(A_{J^c}).$$

Note that for $J \in \mathcal{I}(r, I_{n-1})$ the subsequence J^c always contains n. If $|J^c| \geq 3$ (i.e. $r \leq n-3$) the only subsequences J^c giving rise to non-zero cyclic products are of the form $J^c = I_r^c = (r+1, \ldots, n)$. This can be seen by considering a subsequence of the form $\tilde{J} = (r, \ldots, r+j-1, r+j+1, \ldots, n)$. In order to make a cyclic product $A_{i_1,i_2} \cdots A_{i_r,i_1}$ non-zero using $\tilde{J}$ as index set, we have to choose the element $A_{r+j-1,r+j+1}$, but then we have no way of connecting the upper and lower end of the index set without having a difference of more than two in the indices leading to one of the elements being zero.

When $r = n-2$ such that $|J^c| = 2$ the only non-zero two-cycles clearly arise when $J^c = (n-1, n)$ and $J^c = (n-2, n)$. Consequently we have

$$\begin{aligned}
&\sum_{r=0}^{n-2} \sum_{J \in \mathcal{I}(r, I_{n-1})} \alpha \mathrm{per}_\alpha(A_J) \mathrm{cyp}(A_{J^c}) \\
&\quad = \alpha A_{n-1,n} A_{n,n-1} \mathrm{per}_\alpha(A_{I_{n-2}}) + \alpha A_{n-2,n} A_{n,n-2} \mathrm{per}_\alpha(A_{(1,\ldots,n-3,n-1)}) \\
&\qquad + \sum_{r=0}^{n-3} \alpha \mathrm{per}_\alpha(A_{I_r}) \mathrm{cyp}(A_{I_r^c}).
\end{aligned}$$

If we consider $\mathrm{per}_\alpha(A_{(1,\ldots,n-3,n-1)})$ similar arguments as above yield

$$\mathrm{per}_\alpha(A_{(1,\ldots,n-3,n-1)}) = \alpha A_{n-1,n-1} \mathrm{per}_\alpha(A_{I_{n-3}}) + \alpha A_{n-3,n-1} A_{n-1,n-3} \mathrm{per}_\alpha(A_{I_{n-4}}).$$

Finally we need to analyze the sum of cyclic products $\text{cyp}(A_{I_r^c})$ when $0 \leq r \leq n-3$. In this case when $n-r$ is even the only two non-zero terms in the sum are

$$\text{cyp}(r,n) = A_{r+1,r+2}A_{r+2,r+4}\cdots A_{n-2,n}A_{n,n-1}A_{n-1,n-3}\cdots A_{r+5,r+3}A_{r+3,r+1},$$

$$\text{cyp}^t(r,n) = A_{r+1,r+3}A_{r+3,r+5}\cdots A_{n-3,n-1}A_{n-1,n}A_{n,n-2}\cdots A_{r+4,r+2}A_{r+2,r+1},$$

and when $n-r$ is odd

$$\text{cyp}(r,n) = A_{r+1,r+2}A_{r+2,r+4}\cdots A_{n-3,n-1}A_{n-1,n}A_{n,n-2}\cdots A_{r+5,r+3}A_{r+3,r+1},$$

$$\text{cyp}^t(r,n) = A_{r+1,r+3}A_{r+3,r+5}\cdots A_{n-2,n}A_{n,n-1}A_{n-1,n-3}\cdots A_{r+4,r+2}A_{r+2,r+1}.$$

To ease the notation we let

$$\begin{aligned} a_i &= A_{i,i}, \quad i = 1,\ldots,n \\ b_i &= A_{i,i+1}A_{i+1,i}, \quad i = 1,\ldots,n-1 \\ \beta_i &= A_{i,i+2}A_{i+2,i}, \quad i = 1,\ldots,n-2 \\ p_i^\alpha &= \text{per}_\alpha(A_{I_i}), \quad i = 0,\ldots,n, \end{aligned}$$

when stating the formula in the following corollary.

Corollary 2. *Let A be a $n \times n$ penta-diagonal matrix with $n > 3$. Then, using the notation from above,*

$$\begin{aligned} p_n^\alpha = {} & \alpha a_n p_{n-1}^\alpha + \alpha b_{n-1}p_{n-2}^\alpha + \alpha^2\beta_{n-2}a_{n-1}p_{n-3}^\alpha + \alpha^2\beta_{n-2}\beta_{n-3}p_{n-4}^\alpha \\ & + \sum_{r=0}^{n-3} \alpha p_r^\alpha (cyp(r,n) + cyp^t(r,n)). \end{aligned}$$

This gives an easy way to recursively calculate the α-permanent of a penta-diagonal matrix, and if we also assume that $b_i \neq 0$, $i = 1,\ldots,n-1$, we can simplify the calculations further. This follows the exact same lines as for the regular determinant in Sweet (1969), and we leave out the details in the following. The key idea is that the cyclic products of length greater than three can be written in terms of shorter cyclic products. Using the notation

$$\begin{aligned} c_i &= A_{i,i+1}A_{i+1,i+2}A_{i+2,i}, \\ c_i^t &= A_{i,i+2}A_{i+2,i+1}A_{i+1,i}, \end{aligned}$$

for $i = 1,\ldots,n-2$ we have that

$$\text{cyp}(r,n) = \begin{cases} \frac{c_{r+1}c_{r+2}^t\cdots c_{n-3}c_{n-2}^t}{b_2b_3\cdots b_{n-2}} & \text{for } n-r \text{ even} \\ \frac{c_{r+1}c_{r+2}^t\cdots c_{n-4}c_{n-3}^tc_{n-2}}{b_2b_3\cdots b_{n-2}} & \text{for } n-r \text{ odd.} \end{cases}$$

The recursive algorithm for calculating the α-permanent is then:
Set $p_{-1} = 0$, $p_0 = 1$, $p_1 = \alpha a_{1,1}$, $p_2 = \alpha^2 a_{1,1}a_{2,2} + \alpha b_1$, $\epsilon_{-1} = e_{-1} = 0$, and compute

$$\begin{aligned}
\rho_{k-2} &= a_{k-1}p_{k-3} + \beta_{k-3}p_{k-4},\\
\epsilon_{k-3} &= p_{k-3} + \frac{c_{k-3}^t}{b_{k-2}}e_{k-4},\\
e_{k-3} &= p_{k-3} + \frac{c_{k-3}}{b_{k-2}}\epsilon_{k-4},\\
p_k &= \alpha a_k p_{k-1} + \alpha b_{k-1}p_{k-2} + \alpha^2\beta_{k-2}\rho_{k-2} + \alpha c_{k-2}\epsilon_{k-3} + \alpha c_{k-2}^t e_{k-3}.
\end{aligned}$$

A.4. Approximating the α-permanent

As mentioned previously, the exact calculation of the α-permanent is in general computationally intractable apart from the special cases treated in the previous sections, but it can be approximated using the importance sampling scheme of Kou and McCullagh (2009). Using this method approximation of the α-permanent in e.g. the log-likelihood (5) is feasible for datasets with a moderate total number of counts $n_\star$ (of the order a couple of hundreds). In the following we will discuss how the introduction of a control variate (see Hammersley and Handscomb (1964)) potentially can improve the performance of the algorithm.

To approximate $\mathrm{per}_\alpha(A)$ for a given $n \times n$ matrix A the algorithm uses permutations $\sigma_1, \ldots, \sigma_N \in \mathcal{S}_n$ independently sampled from a certain probability distribution $f(\sigma)$ on $\mathcal{S}_n$ as detailed in Kou and McCullagh (2009). The unbiased estimate is then

$$X = g(\sigma_1, \ldots, \sigma_N; A, \alpha) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{f(\sigma_i)}\alpha^{n-c(\sigma_i)}A_{1,\sigma_i(1)}A_{2,\sigma_i(2)}\cdots A_{n,\sigma_i(n)}.$$

Now let A' approximate A in some sense and have a form such that $\mathrm{per}_\alpha(A')$ can be calculated efficiently (e.g. a block-diagonal or penta-diagonal approximation as detailed in Sections A.2-A.3). Then we use the same set of permutations to form the zero mean random variable $Y = g(\sigma_1, \ldots, \sigma_N; A', \alpha) - \mathrm{per}_\alpha(A')$, and introduce the control variate corrected unbiased estimate of $\mathrm{per}_\alpha(A)$ as $Z = X - \beta Y$. Notice that

$$\sigma_Z^2 = \sigma_X^2 + \beta^2\sigma_Y^2 - 2\beta\rho\sigma_X\sigma_Y,$$

where $\sigma_X^2 = \mathrm{Var}(X)$, $\sigma_Y^2 = \mathrm{Var}(Y)$, $\sigma_Z^2 = \mathrm{Var}(Z)$ and $\rho = \mathrm{Corr}(X, Y)$. Hereby, the optimal value of β minimizing the variance of Z is

$$\hat{\beta} = \rho\frac{\sigma_X}{\sigma_Y}, \tag{19}$$

which changes the variation in the estimate of $\mathrm{per}_\alpha(A)$ by a factor $\sigma_Z^2/\sigma_X^2 = 1 - \rho^2$. In Hammersley and Handscomb (1964) the suboptimal fixed value of $\beta = 1$ is used, but we prefer the optimal value (19), which only requires the additional calculation of an estimate of ρ. We exemplify the use of control variates in what follows below.

A.4.1. Example using control variates

Consider a multivariate negative binomial distribution of dimension $m = 10$ parametrized by $\alpha = 1$ and C with entries $C_{i,j} = \kappa\rho^{|i-j|}$, where $\kappa = 2$ and $\rho = 0.5$. The probability of observing any given outcome $\mathbf{n}$ is given by (2), which depends on the α-permanent of the block matrix $\tilde{C}[\mathbf{n}]$. We have approximated this α-permanent for different outcomes $\mathbf{n}$ using either a penta-diagonal control variate or a block-diagonal control variate. The penta-diagonal matrix is obtained simply by truncating $\tilde{C}[\mathbf{n}]$ to be penta-diagonal (i.e. all entries not on the diagonal or the two first super- or sub-diagonals are set to zero). The block-diagonal matrix is obtained by only retaining the five two-dimensional block matrices of sizes $n_{2i-1} + n_{2i}$, $i = 1, \ldots, 5$ along the diagonal of $\tilde{C}[\mathbf{n}]$ and setting all other entries to zero. Table 2 shows the estimated probability for three different outcomes plus/minus two standard errors. Results are shown for both types of control variates as well as with no control variate using 500 MC samples.

Table 2. Comparison of control variates

		(1,1,1,1,1,1,1,1,1,1)	(1,3,1,3,1,3,1,3,1,3)	(3,3,3,3,3,3,3,3,3,3)
none	$\beta = 0$	$37.31 \pm 1.34 \times 10^{-8}$	$38.85 \pm 3.87 \times 10^{-10}$	$13.88 \pm 2.21 \times 10^{-11}$
block	$\beta = 1$	$37.70 \pm 1.04 \times 10^{-8}$	$40.19 \pm 2.40 \times 10^{-10}$	$14.48 \pm 1.94 \times 10^{-11}$
	$\beta = \hat{\beta}$	$37.55 \pm 0.75 \times 10^{-8}$	$39.93 \pm 2.28 \times 10^{-10}$	$14.39 \pm 1.93 \times 10^{-11}$
penta	$\beta = 1$	$38.16 \pm 0.02 \times 10^{-8}$	$40.81 \pm 2.45 \times 10^{-10}$	$13.29 \pm 2.00 \times 10^{-11}$
	$\beta = \hat{\beta}$	$38.16 \pm 0.02 \times 10^{-8}$	$40.41 \pm 2.32 \times 10^{-10}$	$13.37 \pm 1.99 \times 10^{-11}$