AMERICAN UNIVERSITY

WASHINGTON, DC

Erlang Renewal Models for Genetic Recombination

John P. Nolan

June 2000

Technical Report No. 2000-5

https://doi.org/10.17606/p2a8-cr57

Erlang Renewal Models for Genetic Recombination

John P. Nolan*
American University

Revised 20 June 2000,printed September 26, 2000

Abstract

Closed form expressions for multilocus probabilities are given for the crossover process when it is a renewal process with the distance between crossovers modeled by a Erlang distribution. Closed form expressions are also given for the multilocus probabilities for the chiasma process on the four strand bundle under the same model of recombination for single gamete and for tetrad data. These expressions yield explicit formulas for the map functions, coincidence functions and distributions of the identity-by-descent process for a class of models that incorporate interference. The alternating renewal models used may be of interest in other fields, e.g. telecommunication networks and queues, where they can be used to model the busy/non-busy state of a system with buffers.

Abbreviated title: Erlang models for recombination

AMS 1991 subject calssifications: Primary 92D10; secondary 60K05.

Key words and phrases: Erlang renewal processes, genetic recombination, multilocus probabilities, crossover interference, generalized hyperbolic functions, alternating renewal processes.

1 INTRODUCTION

Multilocus probabilities are the basic quantities that are used to build genetic maps and to compute linkage scores. Suppose there are $n+1$ markers $\mathcal{M}_1, \ldots, \mathcal{M}_{n+1}$ along a chromosome. For each of the inter-marker intervals, let

$$i_j = \begin{cases} 1 & \text{if a recombination has occurred between } \mathcal{M}_j \text{ and } \mathcal{M}_{j+1} \\ 0 & \text{otherwise.} \end{cases}$$

*The author was supported in part by a Visiting Investigator grant at National Human Genome Research Institute, NIH

A sequence $(i_1, \ldots, i_n)$ of 0's and 1's is called a recombination pattern and the multilocus probabilities are:

$$p(i_1, \ldots, i_n) = \text{ probability of the recombination pattern } (i_1, \ldots, i_n). \quad (1)$$

These multilocus probabilities depend on inter-marker distances $d_1, \ldots, d_n$, where d_j= distance between markers $\mathcal{M}_j$ and $\mathcal{M}_{j+1}$. (Throughout this paper, distances will be expressed in genetic units (Morgans), not physical units.) Just as important, the multilocus probabilities depend on the model used to describe the way crossovers occur. The standard model is a Poisson process which is used because it was a reasonable first approximation that is mathematically tractable. It's use in genetics was introduced by Haldane (1919), who knew it assumes no crossover interference.

It is widely accepted that there is positive crossover interference - a crossover at a point apparently inhibits crossovers at nearby points, see Kwiatkowski *et al.*(1993) and Harushima *et al.*(1998). Various models have been proposed to represent that interference. Many authors have attempted to express multilocus probabilities in terms of map functions, e. g. Geiringer (1944), Schnell (1961), Karlin and Liberman (1979), Liberman and Karlin (1984), Risch and Lange (1983), Weeks *et al.* (1993), Karlin and Liberman (1994). However, as Zhao and Speed (1996) point out, such efforts cannot accurately describe general multilocus probabilities because different models can yield the same map function. The root of the problem is that a map function can only describe what happens among three loci and a multilocus probability depends on more information. The "adjacent interval" coincidence coefficient (Sturtevant (1915), Muller (1916)) provides some information, and the "nonadjacent interval" coincidence coefficient of Foss *et al.* (1993) appears to provide more, but neither can fully characterize interference in general. The approach here is to specify a model for recombination and derive multilocus probabilities, and then compute map functions and coincidence functions from them.

Renewal process models for genetic recombination have a long history in genetics, see Fisher *et al.* (1947), Owens (1949), Bailey (1961), Cobbs (1978), Stam (1979), and Lange (1997). In these models, the distance between crossovers is modeled by a random variable with some distribution. Once a crossover has occurred, the distance until the next crossover is an independent random variable with the same distribution. The choice of that distribution completely determines the properties of the crossover process, and hence the multilocus probabilities (1). The mathematical complexity of these models comes from the fact that, except when an exponential distribution is used, the process is non-Markov.

Recently there has been renewed interest in using an Erlang renewal process to model distances between crossovers. Foss *et al.* (1993) suggested such models on biological grounds. McPeek and Speed (1995) fit various data sets to different models of interference, and found that the Erlang models do as good a job

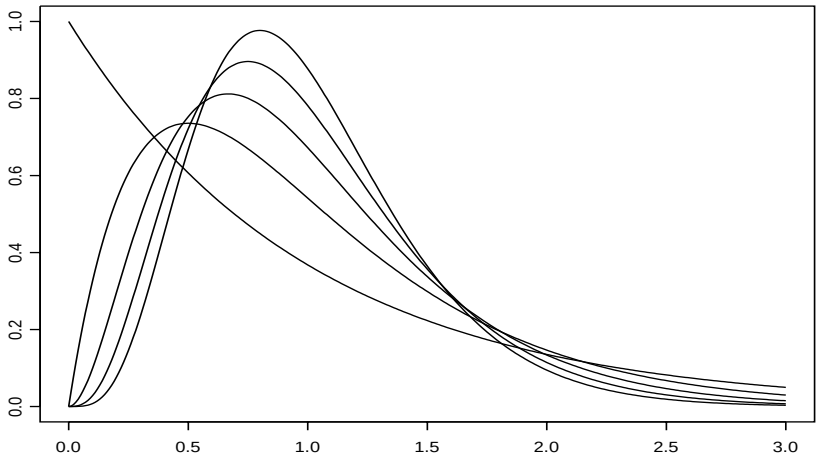

Figure 1: Densities of a Erlang$(m, \lambda m)$ random variable, with $m = 1, 2, 3, 4, 5$ and $\lambda = 1.0$. For small positive x, the topmost curve is $m = 1$, next lower is $m = 2$, etc.

fitting the data as any of the others. An Erlang distribution is described by a shape parameter m, a positive integer, and a scale parameter $\lambda > 0$, with density $f(x) = x^{m-1}e^{-\lambda x}/(m-1)!$, $x > 0$, see Figure 1. Erlang distributions include the exponential distribution as a special case ($m = 1$), and are in turn a subclass of the gamma distributions (as are the chi-squared distributions). Previous authors have called the models analyzed below gamma models or chi-squared models, but since the integer value of the shape parameter is essential, we prefer to use the term Erlang models.

For our purposes it is convenient to parameterize the Erlang distributions as Erlang$(m, \lambda m)$, where m is a positive integer and λ is a positive number. In Zhao, Speed and McPeek (1995) and Lin and Speed (1996), values of $m = 4$ for Drosophila, $m = 2$ for Neurospora, and $m = 3$ for humans give the best fit. (For notational simplicity, our m is their $m + 1$; in the notation of Foss, *et al.* (1993), we are using a $Cx(Co)^{m-1}$ model.) Figure 2 of Harushima, *et al.*(1998) shows a histogram of distances between 555 recombinations for a rice data set. It is poorly described by the Haldane model ($m = 1$), but well described by an Erlang distribution with $m = 2$. The main point of biological interest in using an Erlang distribution is that as m increases, the distance to the next crossover gets more concentrated around the mean, which has the same value, $1/\lambda$, for all Erlang$(m, \lambda m)$ distributions. This implies that it is less likely to see two

crossovers close to each other as m increases. There is an opposing shift in the probabilities for large distances, but that doesn't appear to be significant until the genetic length of a chromosome exceeds $2/\lambda$ Morgans.

The main results of this paper give closed form expressions for multilocus probabilities when the inter-event distribution is Erlang$(m, \lambda m)$. This work is an extension of the models of Owen (1949), Bailey (1961), Cobbs (1978), and Stam (1979). These results in Section 2 are based on ideas in Zhao, Speed and McPeek (1995), where infinite series expressions for multilocus probabilities are given for the chiasma model on the four strand bundle. We show that the the matrix functions they consider answer the multilocus probability question for the crossover process as well, and we give a closed form expression for both the crossover process and the four strand chiasma process. The next section derives map functions and coincidence functions for Erlang models of recombination, filling in some gaps in the work of Cobbs (1978) and Foss *et al.* (1993). The description of the identity by descent process and the effect on genome wide thresholds are contained in the following section. Section 5 reviews our findings and makes some general comments about the plausibility of renewal models for recombination. The proofs are concentrated in Section 6.

In mathematical terms, the crossover process is an alternating renewal process, with state alternating between 0 and 1 and the multilocus probabilities are essentially the finite dimensional distributions of the process. While we focus on the genetic application of these results, the results may be useful in other areas. They may describe busy or non-busy status in a queueing system that buffers exponential arrivals. Examples include a shuttle bus that waits for m passengers before leaving, a computer system that buffers m bytes before initiating an input or output operation, and communication networks that relay packets through m nodes.

2 ERLANG RENEWAL MODELS

2.1 The crossover process

The renewal crossover process is a model for recombination in diploid individuals that involves two strands - maternal and paternal haploids. Crossovers occur between these two strands according to a renewal process, leading to the exchange of genetic material. These models do not appear to take into account the fact that eukaryotic meiosis involves four strands. However, this is not true - Section 5 shows that the chiasma model considered below is a crossover process, albeit with a non-Erlang inter-event distribution.

The formula for multilocus probabilities for a crossover renewal process with

Erlang$(m, \lambda m)$ inter-event distribution is given by:

$$p(i_1, \ldots, i_n) = \frac{1}{m}\mathbf{1}\left(\prod_{j=1}^{n} M_{i_j}^{cross}(\lambda m d_j)\right)\mathbf{1}^T. \tag{2}$$

where $\mathbf{1} = (1, \ldots, 1)$ is a row vector of m 1's, and the $m \times m$ matrix functions $M_0^{cross}(u)$ and $M_1^{cross}(u)$ are given by

$$M_0^{cross}(u) = e^{-u}\begin{bmatrix} f_{0,2m}(u) & f_{1,2m}(u) & f_{2,2m}(u) & \cdots & f_{m-1,2m}(u) \\ f_{2m-1,2m}(u) & f_{0,2m}(u) & f_{1,2m}(u) & \cdots & f_{m-2,2m}(u) \\ f_{2m-2,2m}(u) & f_{2m-1,2m}(u) & f_{0,2m}(u) & \cdots & f_{m-3,2m}(u) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ f_{m+2,2m}(u) & f_{m+3,2m}(u) & f_{m+4,2m}(u) & \cdots & f_{1,2m}(u) \\ f_{m+1,2m}(u) & f_{m+2,2m}(u) & f_{m+3,2m}(u) & \cdots & f_{0,2m}(u) \end{bmatrix}, \tag{3}$$

and

$$M_1^{cross}(u) = e^{-u}\begin{bmatrix} f_{m,2m}(u) & f_{m+1,2m}(u) & f_{m+2,2m}(u) & \cdots & f_{2m-1,2m}(u) \\ f_{m-1,2m}(u) & f_{m,2m}(u) & f_{m+1,2m}(u) & \cdots & f_{2m-2,2m}(u) \\ f_{m-2,2m}(u) & f_{m-1,2m}(u) & f_{m,2m}(u) & \cdots & f_{2m-3,2m}(u) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ f_{2,2m}(u) & f_{3,2m}(u) & f_{4,2m}(u) & \cdots & f_{m+1,2m}(u) \\ f_{1,2m}(u) & f_{2,2m}(u) & f_{3,2m}(u) & \cdots & f_{m,2m}(u) \end{bmatrix}, \tag{4}$$

and the *generalized hyperbolic functions* $f_{r,q}$ are given by

$$f_{r,q}(u) = \frac{1}{q}\sum_{j=0}^{q-1} e^{a_j u}\cos(b_j u - (2\pi r j/q)), \tag{5}$$

where q is a positive integer, $r = 0, \ldots, q-1$, $a_j = a_j(q) = \cos(2\pi j/q)$ and $b_j = b_j(q) = \sin(2\pi j/q)$.

These matrices are given in Bailey (1961, pg. 203), in infinite series form. The derivation is given in Section 6, where a transition matrix interpretation is given for the above matrices. Note that when $m = 1$, (2) simplifies to the Haldane no interference model.

2.2 The chiasma process

In a complete model of gamete formation in diploid eukaryotes, each haploid replicates itself, and a four stranded bundle is formed. In a renewal chiasma process, crossovers occur among these four strands according to a renewal process, and the bundle pulls apart to form four gametes. The crucial difference between this model and the crossover process is that a crossover among sister chromatids

does not result in a genetically observable exchange of material, although it does interfere with the location of nearby chiasma. Karlin and Libermann (1984) and Speed (1999) discuss a mathematical model for this, based on work of Mather (1936, 1937) and others. This approach allows one to model more concretely what goes on in the biological process of recombination. First we will focus on a multilocus probabilities for a single gamete produced by an individual; then tetrad multilocus probabilites will be derived. In what follows, we assume no chromatid interference (NCI), that is, which chromatids crossover at a given point are not dependent on which chromatids crossover at other points.

If the distance between successive crossovers are Erlang$(m, \lambda m)$, then with the NCI model, there are an average of $\lambda/2$ genetically observable crossovers in a distance of one Morgan. Hence to keep this model comparable to the crossover process, which has an average of λ crossovers per Morgan, an Erlang$(m, 2\lambda m)$ inter-event distribution should be used. For a gamete formed by a renewal chiasma process with Erlang$(m, 2\lambda m)$ inter-event distribution, the multilocus probabilities are given by

$$p(i_1, \ldots, i_n) = \frac{1}{m}\mathbf{1}\left(\prod_{j=1}^{n} M_{i_j}^{NCI}(2\lambda m d_j)\right)\mathbf{1}^T, \tag{6}$$

where

$$M_0^{NCI}(u) = \frac{1}{2}(D_\infty(u) + D_0(u)) \quad \text{and} \quad M_1^{NCI}(u) = \frac{1}{2}(D_\infty(u) - D_0(u)).$$

The $m \times m$ matrix functions D_∞ and D_0 are

$$D_\infty(u) = e^{-u}\begin{bmatrix} f_{0,m}(u) & f_{1,m}(u) & f_{2,m}(u) & \cdots & f_{m-1,m}(u) \\ f_{m-1,m}(u) & f_{0,m}(u) & f_{1,m}(u) & \cdots & f_{m-2,m}(u) \\ f_{m-2,m}(u) & f_{m-1,m}(u) & f_{0,m}(u) & \cdots & f_{m-3,m}(u) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ f_{2,m}(u) & f_{3,m}(u) & f_{4,m}(u) & \cdots & f_{1,m}(u) \\ f_{1,m}(u) & f_{2,m}(u) & f_{3,m}(u) & \cdots & f_{0,m}(u) \end{bmatrix}, \tag{7}$$

and

$$D_0(u) = e^{-u}\begin{bmatrix} 1 & u & u^2/2 & u^3/3! & \cdots & u^{m-1}/(m-1)! \\ 0 & 1 & u & u^2/2 & \cdots & u^{m-2}/(m-2)! \\ 0 & 0 & 1 & u & \cdots & u^{m-3}/(m-3)! \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{8}$$

The matrices M_0^{NCI} and M_1^{NCI} are called N and R respectively in Zhao *et al.* (1995), where they are given as infinite series. Lin and Speed (1996) have implemented a numerical approximation to these matrices.

2.3 Tetrad case

For tetrad data, there are three possible tetrad patterns between each pair of markers. There are now 3^n recombination patterns for $n+1$ markers, each can be represented by a pattern $(i_1, \ldots, i_n)$, where each $i_j \in \{0, 1, 2\}$. The same notation can be used as before, with the multilocus probabilities for tetrads, assuming NCI, given by

$$p(i_1, \ldots, i_n) = \frac{1}{m}\mathbf{1}\left(\prod_{j=1}^{n} M_{i_j}^{tetrad}(2\lambda m d_j)\right)\mathbf{1}^T,$$

where

$$\begin{aligned}
M_0^{tetrad}(u) &= \frac{e^{-u}}{2}\left(D_\infty(u) - D_0(u) - M_1^{tetrad}(u)\right), \\
M_1^{tetrad}(u) &= e^{-u}\begin{bmatrix} h_{m,m}(u) & h_{m+1,m}(u) & h_{m+2,m}(u) & \cdots & h_{2m-1,m}(u) \\ h_{m-1,m}(u) & h_{m,m}(u) & h_{m+1,m}(u) & \cdots & h_{2m-2,m}(u) \\ h_{m-2,m}(u) & h_{m-1,m}(u) & h_{m,m}(u) & \cdots & h_{2m-3,m}(u) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ h_{2,m}(u) & h_{3,m}(u) & h_{4,m}(u) & \cdots & h_{m+1,m}(u) \\ h_{1,m}(u) & h_{2,m}(u) & h_{3,m}(u) & \cdots & h_{m,m}(u) \end{bmatrix} \\
M_2^{tetrad}(u) &= \frac{e^{-u}}{2}\left(D_\infty(u) + D_0(u) - M_1^{tetrad}(u)\right).
\end{aligned}$$

The functions $h_{r,m}$, $r = 0, 1, \ldots, 2m-1$ are given by

$$h_{r,m}(u) = \begin{cases} \frac{2}{3}f_{r,m}(u) + (-1)^r c_{m,r,m} e^{-\alpha u} \\ \qquad +2\sum_{j=1}^{m-1} c_{j,r,m} e^{\alpha a_j' u}\cos(\alpha b_j' u - rj\pi/m) & r < m \\ \frac{2}{3}f_{r-m,m}(u) - 2\frac{u^{r-m}}{(r-m)!} + (-1)^r c_{m,r,m} e^{-\alpha u} \\ \qquad +2\sum_{j=1}^{m-1} c_{j,r,m} e^{\alpha a_j' u}\cos(\alpha b_j' u - rj\pi/m) & r \geq m, \end{cases} \tag{9}$$

where $\alpha = (1/2)^{1/m}$, $a_j' = a_j(2m) = \cos(\pi j/m)$, $b_j' = b_j(2m) = \sin(\pi j/m)$ and

$$c_{j,r,m} = \begin{cases} 0 & j \text{ even} \\ -2/(3m\alpha^r) & j \text{ odd.} \end{cases}$$

The matrices M_0^{tetrad}, M_1^{tetrad} and M_2^{tetrad} correspond to the matrices P, T and N respectively in Zhao *et al.* (1995).

3 MAP FUNCTIONS AND COINCIDENCE FUNCTIONS

As we noted in the introduction, the map function gives only partial information about multilocus probabilites. Still, it is of interest to know what the map function is for the Erlang renewal models, and we will use it below to describe the IBD process. If d is the genetic distance between two loci, then for a crossover renewal process with Erlang$(m, \lambda m)$ inter-event distribution, the recombination fraction between them is

$$\begin{aligned} r_m^{cross}(d) &= \frac{1}{m}\mathbf{1}M_1^{cross}(\lambda md)\mathbf{1}^T = e^{-\lambda dm} f_{m,2m}(\lambda md) \\ &\quad + \frac{e^{-\lambda dm}}{m}\sum_{j=1}^{m-1}(m-j)\left[f_{m+j,2m}(\lambda md) + f_{m-j,2m}(\lambda md)\right]. \end{aligned}$$

For a chiasma renewal process with NCI and Erlang$(m, 2\lambda m)$ inter-event distribution, the recombination fraction is,

$$\begin{aligned} r_m^{NCI}(d) &= \frac{1}{m}\mathbf{1}M_1^{NCI}(2\lambda md)\mathbf{1}^T \\ &= \frac{e^{-2\lambda dm}}{2}\sum_{j=0}^{m-1}\left(f_{j,m}(2\lambda md) - \left(\frac{m-j}{m}\right)\frac{(2\lambda md)^j}{j!}\right) \\ &= \frac{1}{2}\left(1 - e^{-2\lambda dm}\frac{1}{m}\sum_{j=0}^{m-1}(m-j)\frac{(2\lambda md)^j}{j!}\right), \end{aligned} \tag{10}$$

where the last equality uses (14) in Section 6. This last result is equation (30) of Cobbs (1978), and equation (7) of Foss *et al.* (1993).

Figure 2 shows a graph of the map function for various values of m. As expected, the functions start from the no interference model (Haldane distance, $m = 1$) and get closer and closer to the complete interference model $\theta = d$. Note that recombination fractions for the crossover models exceed the level $r = 1/2$ when $m > 1$. In fact, for the crossover model, the recombination fraction oscillates around $1/2$ as $d \to \infty$. In contrast, under the chiasma model with NCI, Mather's formula shows that the recombination fraction cannot exceed $1/2$. In our case, this is obvious from (10) because the sum in the last term is always less than $e^{2\lambda md}$. While not shown here, the graphs of the other commonly used map functions (Kosambi, Binomial with $N = 2$, Sturt with any $L > 0.79$) generally lie between the crossover $m = 1$ and the crossover $m = 2$ curves shown. (One can get above the $m = 2$ curve by taking L small enough in the Sturt map function, or by making N large enough in Binomial map function.) It is straightforward to show that both recombination fractions are of order $\lambda d + o(d)$ as $d \to 0$: use the fact that $f_{r,q}(u) = u^r/r! + o(u^r)$ as $u \to 0$ (see equation (12)) and some algebra.

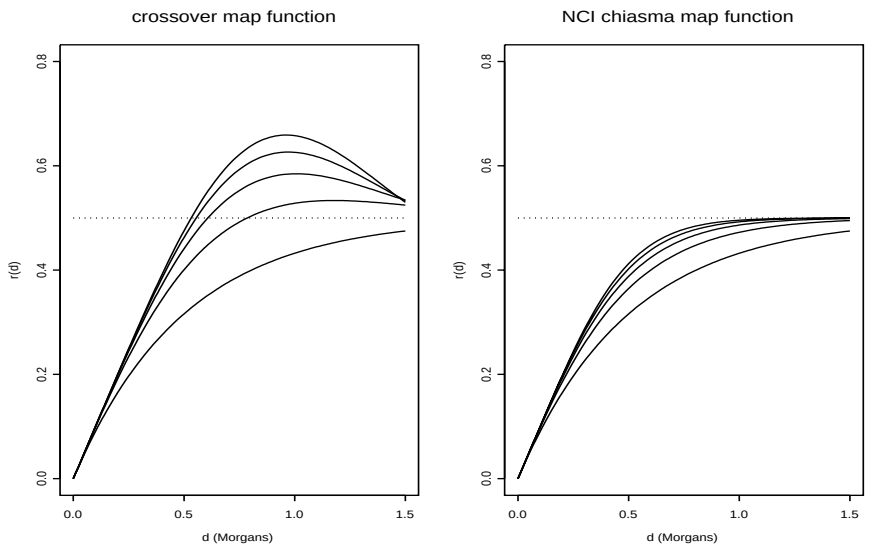

Figure 2: Map functions for the crossover process (left) and chiasma process (right) with Erlang interference models for various values of m and $\lambda = 1.0$. The lowest curve is $m = 1$, where the crossover and chiasma model coincide. Above that curve, m increases from 2 to 5.

In describing the coincidence functions, it is convenient to allow the symbol "*" in a multilocus probability to denote either a 0 or a 1. Thus $p(1,*) = p(1,0) + p(1,1)$, etc. The classical "adjacent interval" coincidence coefficient is defined by taking three markers, separated by inter-marker distances d_1 and d_2:

$$C_3(d_1, d_2) = \frac{p(1,1)}{p(1,*)p(*,1)} = \frac{r(d_1) + r(d_2) - r(d_1 + d_2)}{2r(d_1)r(d_2)}.$$

The discussion on page 307 of Lange, Zhao and Speed (1997) shows all Erlang renewal models have positive interference. The "nonadjacent interval" coincidence coefficient is defined by taking four markers, with intermarker distances d_1, d_2 and d_3:

$$C_4(d_1, d_2, d_3) = \frac{p(1,*,1)}{p(1,*,*)p(*,*,1)} = \frac{p(1,*,1)}{r(d_1)r(d_3)}.$$

With these definitions, S_3 and S_4 of Foss *et al.* (1993) are given by

$$\begin{aligned} S_3(d) &= C_3(d,d) = [r(d) - (1/2)r(2d)]/[r(d)]^2 \\ S_4(d) &= \lim_{d_1 \downarrow 0} \lim_{d_3 \downarrow 0} C_4(d_1, d, d_3). \end{aligned}$$

These equations are general, they depend only on valid multilocus probabilities. When Erlang interference is assumed, S_3 can be computed using the formulas for map functions above. For S_4, Section 6 shows that

$$\begin{aligned} S_4^{cross}(d) &= me^{-\lambda m d} f_{m-1,m}(\lambda m d) \\ S_4^{NCI}(d) &= me^{-2\lambda m d} f_{m-1,m}(2\lambda m d) = S_4^{cross}(2d). \end{aligned}$$

Figure 3 shows plots of S_3 and S_4 for both models. We note that the last equation above is a closed form expression for their equation (8) of Foss *et al.* (1993).

4 IBD PROCESS

The primary goal in genetic linkage studies is to localize disease genes by determining where affected relative pairs have segments of their chromosome identical by descent (IBD), i.e. inherited from the same ancestor. The IBD process $X(t)$ is a model for these shared segments. For simplicity, consider two half-sibs and compare the chromosomes that they inherited from their common parent. Let t denote position along the chromosome and define

$$X(t) = \begin{cases} 1 & \text{if the DNA at } t \text{ came from the same grandparent} \\ 0 & \text{otherwise.} \end{cases}$$

The places where $X(t)$ changes value are precisely the points where a crossover has occurred.

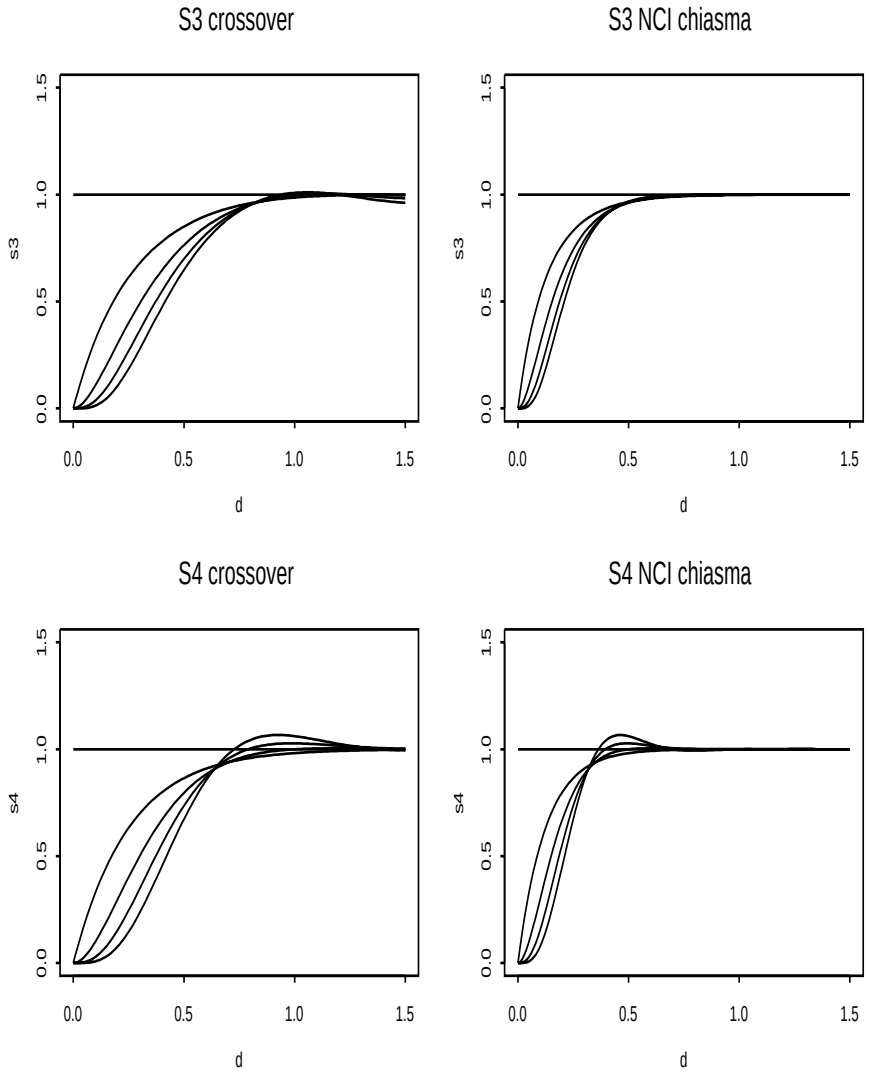

Figure 3: Coincidence functions S_3 and S_4 for the crossover process (left) and chiasma process (right) with Erlang interference models for various values of m and $\lambda = 1.0$. The horizontal lines at height 1 correspond to $m = 1$ (no interference). Values of m increase from $m = 2$ to $m = 5$ from left to right.

If we have the multilocus probabilities (1), then the multilocus IBD probabilities are given by:

$$\begin{aligned} & P(X(t_1) = i_1, X(t_2) = i_2, \ldots, X(t_{n+1}) = i_{n+1}) \\ &= P(X(t_1) = i_1, X(t_2) - X(t_1) = |i_2 - i_1|, \ldots, X(t_{n+1}) - X(t_n) = |i_{n+1} - i_n|) \\ &= \frac{1}{2} p(|i_2 - i_1|, |i_3 - i_2|, \ldots, |i_{n+1} - i_n|). \end{aligned}$$

Note that this equality always holds, regardless of what model is used (crossover, chiasma, NCI or chromatid interference, etc.). The awkward looking absolute value signs are explained by the fact that $|i_{j+1} - i_j| = 1$ or 0, depending on whether or not a recombination has or has not occurred in the j^{th} interval. For the crossover and chiasma models described above, the value of λ used in evaluating the multilocus probability in this formula depends on the type of relative pair considered.

4.1 Thresholds for dense markers

Lander and Kruglyak (1995) used the no interference model to derive appropriate thresholds for an infinitely dense scan of the genome. Using their approach on a genome wide scan is questionable because it is based on a null hypothesis of no contributing gene, and no one would undertake a full genome scan unless there was strong evidence of a genetic factor. However, their approach does make sense for a limited region, say a single chromosome. We show that the thresholds don't change when the Erlang renewal processes described above are used instead of the Haldane model.

The basic IBD process is a stationary 0-1 valued process with mean and covariance

$$\begin{aligned} E\, X(t) &= P(X(t) = 1) = 1/2 \\ \mathrm{Cov}(X(t+d), X(t)) &= P(X(t) = 1, X(t+d) = 1) - P(X(t) = 1) P(X(t+d) = 1) \\ &= \frac{1}{2}(1 - r(d)) - \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} - \frac{1}{2} r(d), \end{aligned} \tag{11}$$

where $r(d)$ is the recombination fraction. Given a sample of n relative pairs, sum over all pairs and normalize to get $Z(t) = 2\sqrt{n} \sum (X_j(t) - \frac{1}{2})$. When n is large, this is approximately a (stationary) Gaussian process. When X is based on the no interference model, the large sample limit is an Ornstein-Uhlenbeck process; when $m > 1$ the crossover and chiasma models considered above do not have an Orstein-Uhlenbeck process as the limit.

The main technical result used in deriving the thresholds is a large deviation result, e.g. Theorem 12.2.9 of Leadbetter, Lindgren and Rootzen (1983). That result shows that the threshold for a dense set of markers depends on the rate

at which $\mathrm{Cov}(Z(t), Z(t+d)) \to 1$ as $d \to 0$. Using (11),

$$\begin{aligned}\mathrm{Cov}(Z(0), Z(d)) &= 4n\,\mathrm{Cov}((1/n)\sum_i (X_i(0) - \frac{1}{2}), (1/n)\sum_i (X_i(d) - \frac{1}{2})) \\ &= 4\mathrm{Cov}(X_i(0), X_i(d)) = 1 - 2r(d).\end{aligned}$$

As we noted above, for any m, the Erlang renewal processes have $r(d) = \lambda d + o(d)$ as $d \to 0$, so $\mathrm{Cov}(Z(t), Z(t+h)) = 1 - 2\lambda d + o(d)$ as $d \to 0$. In fact, any plausible model of recombination will have the map function approximately linear near the origin, so it will yield the same thresholds as the no interference model for a dense map.

When a finite set of markers are used, there is a change in the threshold. The reason is that with positive interference, nearby markers are less dependent, and the multiple comparison problem is heightened. Quantifying this difference depends on being able to accurately compute cumulative probabilities for multivariate normal distributions with dependence, a difficult computational problem for $n > 4$ markers.

5 DISCUSSION

We have used Erlang renewal processes to model both the crossover process and the chiasma process with NCI. Closed form expressions are given for multilocus probabilities in both cases, completing the work of Owens (1949), Bailey (1961), Cobbs (1978), Stam (1979) and Zhao *et al.* (1995). These formulas lead to expressions for map functions, coincidence functions, IBD probabilities as well as closed form expressions for tetrad multilocus probabilities.

The fact that crossover models with $m > 1$ yield recombination fractions above $1/2$ may be desirable in certain cases. This can happen in prokaryotes, so these models may be directly applicable there. In fact, the observance of recombination fractions above $1/2$ (Falconer (1947) and Wright (1947)) in mouse data was seen as a deficiency of the Haldane, Kosambi, etc. map functions. The second cited source is a careful study involving 453 offspring in a balanced block design. Convinced that $r > 1/2$ was possible, Fisher *et al.* (1947), Owen (1949), and Bailey (1961) specifically tried to develop models that had this property. We do not know whether such fractions have been seen in other data sets or whether other factors, e.g. differential viability of the organisms, may have caused the value of $r > 1/2$.

There is a mathematical explanation for $r > 1/2$ in terms of the underlying renewal process. When $m > 1$, the Erlang densities are concentrated around the mean of $1/\lambda$, which means a recombination is most likely to occur approximately $1/\lambda$ Morgans away from the first crossover. Equivalently, for the associated IBD process, (11) shows that the covariance becomes negative when $r > 1/2$, so that the process is most likely to be in opposite states at that distance. This is not

restricted to Erlang models; any renewal process model for the crossover process whose inter-event distribution has a strong enough peak will have $r > 1/2$.

Crossover models are all that are strictly necessary in mammalian genetics (excluding oocyte mapping), because we only observe the single gamete that was used at conception. For example, the renewal chiasma model with NCI described above is a crossover process with inter-event distribution a geometric mixture of Erlangs. The equivalent crossover interevent distribution has density

$$\begin{aligned} h(u) &= \sum_{j=1}^{\infty} (1/2)^j \exp(-2\lambda m u)(2\lambda m)(2\lambda m u)^{jm-1}/(jm-1)! \\ &= 2^{1-1/m}\lambda m \exp(-2\lambda m u) \sum_{j=1}^{\infty} (2^{(1-1/m)}\lambda m u)^{jm-1}/(jm-1)! \\ &= 2^{1-1/m}\lambda m \exp(-2\lambda m u) f_{m-1,m}(2^{(1-1/m)}\lambda m u), \end{aligned}$$

where the last term uses (12). In words, if we thin an Erlang point process, we get a different process with interevent distances given by the expression above.

It is an open question whether or not a renewal process is an appropriate model for recombination. First we address some technical issues, then make a general comment.

One criticism of renewal processes is that they are not generally "multilocus feasible" in the sense of Liberman and Karlin (1984). On this issue, we agree with Speed (1999), where it is pointed out that Liberman and Karlin define what might be called "nonadjacent interval multilocus feasibility". While mathematically elegant, their definition puts conditions on recombinations in intervals separated by an arbitrary distance, which does not agree with the basic intuition of interference being a local phenomenon. Zhao and Speed (1996) show that most of the common map functions can arise from renewal processes, even though some are not "nonadjacent interval multilocus feasible."

Another criticism of the use of renewal processes is that multiple chiasma apparently can occur simultaneously, making a serial renewal process inappropriate. As Bailey (1961, pg. 178) points out, we do not necessarily need a serial explanation for using Erlang inter-event distributions - they may just describe what's going on in the spatial point process (ignoring the temporal dimension). Molecular interactions may act spatially, not temporarly, inhibiting nearby crossovers. The counting model of Foss *et al.* (1993) assumes intermediates (C's) being distributed according to a Poisson point process, and then some of these convert to crossovers. They focus on a fixed number ($m-1$ in our notation) of non-crossover events (Co's) between crossovers (Cx's), but also mention a variable number of Co's. Lange, Speed and Zhao (1997) and Lange (1997) analyze this "random-skip" process and give infinite series for multilocus probabilities and derived quantities for that model.

In the end, experimentation will have to resolve whether Erlang (or any) renewal process realistically models recombination. A more relevant question

right now is whether these models do a better job than the commonly used no interference model. The results of Foss *et al.* (1993), e. g. Figure 4, and McPeek and Speed (1995) indicate that that they do. A maximum likelihood fit to Figure 2 of Harushima *et al.*(1998) shows that an Erlang distribution with $m = 2, \lambda = 1/2$ fits the data well. If so, these models may help detect disease or trait loci and build genetic maps.

6 PROOFS

Mathematical proofs of the results described above are given in this section. We follow the argument and notation of Zhao, Speed and McPeek (1995). For $k = 1, 2, 3, \ldots$, define the sequence of $m \times m$ matrix functions $D_k(u)$ with $(i,j)^{th}$ entry $e^{-u}u^{mk+j-i}/(mk+j-i)!$. The matrix $D_0(u)$ was defined in (8) and should have zeros below the main diagonal (there is a misprint in Zhao *et al.*). The $D_k(\cdot)$ matrices have an interpretation as transition matrices that is implicit in Zhao *et al.* (1993). Let $N(t)$ be a renewal counting process with inter-event distribution Erlang(m, λ), and let $N^*(t)$ be a Poisson process, i.e. a renewal process with inter-event distribution Exponential(λ)=Erlang$(1, \lambda)$. The key idea in what follows is the observation that $N(t) \stackrel{d}{=} \lfloor N^*(t)/m \rfloor =$ the integer part of $N^*(t)/m$, i. e. the Erlang renewal process $N(t)$ increases by 1 every time m events have occurred for the Poisson process $N^*(t)$. The Markov nature of $N^*(t)$ then allows an analysis of $N(t)$.

The phrase "$N(t)$ is in phase i" will be used as shorthand for $N^*(t) = i \,(\mathrm{mod}\, m)$. In the terminology of Foss, *et al.* (1993), this means that i Co events have occurred since the last Cx event. Then the $(i,j)^{th}$ entry of $D_k(u)$ is $P(N(u) - N(0) = k, N(u)$ is in phase $j | N(0)$ is in phase $i)$. For $k > 1$, this probability is equal to $P(N^*(u) - N^*(0) = (m-i) + m(k-1) + j) = P(N^*(u) - N^*(0) = mk + j - i) = e^{-u}u^{mk+j-1}/(mk+j-i)!$. A similar argument gives D_0. If $\mathbf{p}_0$ is the distribution of the phase of $N(0)$, then $\mathbf{p}_0 D_k(u)$ is the distribution of the phase of $N(u)$ given that k renewal events occurred in $[0, u]$. This is the intuitive content of Lemma 1 of Zhao *et al.* In particular, for an Erlang renewal process with $N(0)$ having distribution $\mathbf{p}_0$,

$$P(N(t_1) = n_1, N(t_2) = n_2, \ldots, N(t_k) = n_k) = \\ \mathbf{p}_0 D_{n_1}(\lambda t_1) D_{n_2 - n_1}(\lambda(t_2 - t_1)) \cdots D_{n_k - n_{k-1}}(\lambda(t_k - t_{k-1})) \mathbf{1}^T.$$

The memoryless property of $N^*(t)$ is what makes the multiplication of matrices give the correct probabilities for $N(t)$. The choice $\mathbf{p}_0 = (1/m)\mathbf{1} = (1/m, \ldots, 1/m)$ used in the formulas for multilocus probabilities represents the equiprobable initial distribution for the phase of $N(0)$ in the stationary case.

For the crossover process, a recombination is seen precisely when there is an odd number of crossovers. This leads to the formulas: $M_0^{cross}(u) = \sum_{k=0}^{\infty} D_{2k}(u)$ and $M_1^{cross}(u) = \sum_{k=0}^{\infty} D_{2k+1}(u)$, i.e. M_0 takes into account

all possibilities with an even number of crossovers, whereas M_1 takes into account all possibilities with an odd number of crossovers. Like the D_k matrices above, the entries of these matrices have a transition matrix interpretation. For example, the i, j^{th} entry of $M_0(u)$ is the probability of starting in phase i, having an even number of crossovers in distance u, and ending in phase j.

Using these sums and the definitions of D_k, some algebra shows that M_0 and M_1 have the form claimed in (3) and (4) respectively, where

$$f_{r,q}(u) = \sum_{k=0}^{\infty} u^{qk+r}/(qk+r)! \qquad r = 0, \ldots, q-1. \tag{12}$$

Bailey (1961, pg. 202), calls these the q^{th} order segmental functions. A closed form expression for these follows from the next lemma.

Lemma 1 *$f_{r,q}(u)$ defined by (12) can be written as (5).*

Proof Differentiating $f_{r,q}$ with respect to u repeatedly shows that $f_{r,q}^{(q)}(u) = f_{r,q}(u)$. The initial conditions for this q^{th} order differential equation are $g_{r,q}^{(i)}(0) = 1$ if $i = r$; and $= 0$ otherwise. After solving this equation, the author discovered that these differential equations are known in the mathematical literature. The solutions are

$$f_{r,q}(u) = \frac{1}{q} \sum_{j=0}^{q-1} \omega^{-jr} e^{\omega^j u},$$

where $\omega = \exp(2\pi i/q)$ is a q^{th} root of unity. This is given in Erdélyi *et al.* (1955, pg. 212), where $f_{r,q}$ are called generalized hyperbolic functions of order q. A recent survey of these functions is given in Muldoon and Ungar (1996). To eliminate the complex terms in this expression, set $\Delta = 2\pi/q$, then the constants from (5) are $a_j = \cos(j\Delta)$ and $b_j = \sin(j\Delta)$. Since $\omega = \exp(i\Delta)$, $\omega^j = a_j + ib_j$ and

$$\begin{aligned} \omega^{-jr} \exp(\omega^j u) &= \exp(i(-jr\Delta) + (a_j + ib_j)u) \\ &= e^{a_j u}(\cos(b_j u - rj\Delta) + i \sin(b_j u - rj\Delta)) \end{aligned}$$

Thus

$$f_{r,q}(u) = \frac{1}{q} \sum_{j=0}^{q-1} e^{a_j u} \cos(b_j u - rj\Delta) + i \frac{1}{q} \sum_{j=0}^{q-1} e^{a_j u} \sin(b_j u - rj\Delta). \tag{13}$$

It remains to be shown that the imaginary term above is zero. When q is odd, say $q = 2m+1$, then the $j = 0$ term is zero because $b_0 = 0$. For $j = 1, \ldots, m$, $a_{q-j} = a_j$, $b_{q-j} = -b_j$, and $\sin(b_{q-j}u - r(q-j)\Delta) = \sin(-b_j u - 2\pi r + rj\Delta)$ $= -\sin(b_j u - rj\Delta)$. Hence the imaginary term is zero. When q is even, say

$q = 2m$, then the $j = 0$ and $j = m$ terms in the sum are zero, and the j^{th} and $(q-j)^{th}$ terms will cancel as above. □

If we sum (12) over $r = 0, 1, \ldots, q-1$, all powers appear in the series, leading to

$$\sum_{r=0}^{q-1} f_{r,q}(u) = \sum_{j=0}^{\infty} u^j/j! = e^u. \tag{14}$$

We next consider the chiasma process on the four strand bundle. Theorem 1 of Zhao *et al.* (1995) gives the solution of (1) for the single gamete case: $M_i^{NCI} = (1-i)D_0 + (1/2)\sum_{k=1}^{\infty} D_k = (1/2)(\sum_{k=0}^{\infty} D_k + (-1)^{i+1} D_0)$. Straightforward algebra shows that $D_\infty(u) = \sum_{k=0}^{\infty} D_k(u)$ has the form claimed in (7). For the tetrad case, a more involved argument is needed.

Lemma 2 *The matrices M_0^{tetrad}, M_1^{tetrad} and M_2^{tetrad} are given by (9).*

Proof For the tetrad case, Zhao *et al.* (1995) give the following series representations for M_0, M_1 and M_2:

$$M_0 = D_0 + \sum_{k=2}^{\infty} p_0^{(k)} D_k, \qquad M_1 = D_1 + \sum_{k=2}^{\infty} p_1^{(k)} D_k, \qquad M_2 = \sum_{k=2}^{\infty} p_2^{(k)} D_k,$$

where $p_0^{(k)} = p_2^{(k)} = \frac{1}{3}(\frac{1}{2} + (-\frac{1}{2})^k)$, $p_1^{(k)} = \frac{2}{3}(1 - (-\frac{1}{2})^k)$. Note that $M_0 = D_0 + M_2$ and since $p_0^{(k)} + p_1^{(k)} + p_2^{(k)} = 1$, $M_0 + M_1 + M_2 = D_0 + D_1 + \sum_{k=2}^{\infty}[p_0^{(k)} + p_1^{(k)} + p_2^{(k)}]D_k = D_\infty$. Therefore $M_0 + M_1 + M_2 = (M_2 + D_0) + M_1 + M_2 = D_\infty$, so $M_2 = \frac{1}{2}[D_\infty - D_0 - M_1]$ and $M_0 = \frac{1}{2}[D_\infty + D_0 - M_1]$.

It remains to show that M_1 has form (9). Noting that $p_1^{(1)} = 1$ is consistent with the definition of $p_1^{(k)}$, we have $M_1 = \sum_{k=1}^{\infty} p_1^{(k)} D_k$ has the claimed form, where $h_{r,m}(u) = \sum_{k=1}^{\infty} p_1^{(k)} u^{mk+r}/(mk+r)!$, $r = 0, \ldots, 2m-1$. Differentiating $h_{r,m}$ $2m$ times gives

$$h_{r,m}^{(2m)}(u) = \begin{cases} \frac{1}{2}\frac{u^r}{r!} + \sum_{k=1}^{\infty} p_1^{(k+2)} u^{mk+r}/(mk+r)! & r < m \\ \frac{u^{r-m}}{(r-m)!} + \frac{1}{2}\frac{u^r}{r!} + \sum_{k=1}^{\infty} p_1^{(k+2)} u^{mk+r}/(mk+r)! & r \geq m \end{cases}.$$

Luckily, $p_1^{(k+2)} = \frac{1}{2} + \frac{1}{4} p_1^{(k)}$, so using (12)

$$\begin{aligned} \sum_{k=1}^{\infty} p_1^{(k+2)} u^{mk+r}/(mk+r)! &= \frac{1}{2}\sum_{k=1}^{\infty} u^{mk+r}/(mk+r)! + \frac{1}{4}\sum_{k=1}^{\infty} p_1^{(k)} u^{mk+r}/(mk+r)! \\ &= \begin{cases} \frac{1}{2}\left[f_{r,m}(u) - \frac{u^r}{r!}\right] + \frac{1}{4} h_{r,m}(u) & r < m \\ \frac{1}{2}\left[f_{r-m,m}(u) - \frac{u^r}{r!} - \frac{u^{r-m}}{(r-m)!}\right] + \frac{1}{4} h_{r,m}(u) & r \geq m. \end{cases} \end{aligned}$$

Substituting this in the above equation shows that $h_{r,m}$ satisfies the $(2m)^{th}$ order differential equation

$$h_{r,m}^{(2m)}(u) \quad = \quad \begin{cases} \frac{1}{2} f_{r,m}(u) + \frac{1}{4} h_{r,m}(u) & r < m \\ \frac{1}{2} f_{r-m,m}(u) + \frac{1}{2} \frac{u^{r-m}}{(r-m)!} + \frac{1}{4} h_{r,m}(u) & r \geq m. \end{cases}$$

The initial conditions are $h_{r,m}^{(j)}(0) = 1$ if $j = m + r$ and $= 0$ otherwise. The general solution to these equations is

$$h_{r,m}(u) = \begin{cases} \sum_{j=0}^{2m-1} \gamma_{j,r,m} e^{\alpha \omega^j u} + \frac{2}{3} f_{r,m}(u) & r < m \\ \sum_{j=0}^{2m-1} \gamma_{j,r,m} e^{\alpha \omega^j u} + \frac{2}{3} f_{r-m,m}(u) - 2\frac{u^{r-m}}{(r-m)!} & r \geq m, \end{cases}$$

where $\omega = \omega(2m) = \exp(i\pi/m)$: the summation gives the solution of the homogeneous equation and the remaining terms give a particular solution. Laborious calculations with the initial conditions show that $\gamma_{j,r,m} = c_{j,r,m}\omega^{-jr}$. More algebra shows that the above simplifies to (9). □

Next we derive formulas for the coincidence function S_4. For the crossover process, $S_4^{cross}(d) = (1/m)\mathbf{v}(M_0^{cross}(\lambda d) + M_1^{cross}(\lambda d))\mathbf{v}^T$, where $\mathbf{v} = \mathbf{1}(\lim_{d \downarrow 0} M_1(\lambda d)/r^{cross}(d))$. Now $r(d) = \lambda d + o(d)$ and $f_{r,q}(\lambda d) = (\lambda d)^r + o(d^r)$ as $r \downarrow 0$, so the limiting matrix of $M_1(\lambda d)/r(d)$ is all zero, except for the lower left element which is the constant m. Hence $\mathbf{v} = (0, \ldots, 0, m)$, with only one non-zero entry. Now $(M_0^{cross}(\lambda d) + M_1^{cross}(\lambda d))$ has $(m,m)^{th}$ entry $f_{m-1,2m}(\lambda d) + f_{2m+1,2m}(\lambda d) = f_{m-1,m}(\lambda d)$, where the last identity is obtained by adding two series of form (12). Hence $S_4^{cross}(d) = (1/m)m^2 \exp(-\lambda d)\ f_{m-1,m}(\lambda d)$.

The argument is similar for the NCI chiasma model: $S_4^{NCI}(d) = (1/m)\mathbf{v}$ $(M_0^{NCI}(2\lambda d) + M_1^{NCI}(2\lambda d))\mathbf{v}^T = (1/m)\mathbf{v}D_\infty(2\lambda d)\mathbf{v}^T$, where $\mathbf{v}$ is the same as above. The $(m,m)^{th}$ entry of D_∞ is $\exp(-2\lambda d)\ f_{m-1,m}(2\lambda d)$, giving the formula for S_4^{NCI}.

We close with the comment that the functions and matrices used above are rich in mathematical structure. The matrix $D_\infty(u)$ is called the 1-hyperbolic matrix in Muldoon and Ungar (1996). It is a circulant matrix, is related the the fast Fourier transform, always has determinant 1, and $D_\infty(u)D_\infty(v) = D_\infty(u+v)$. The matrices $M_0^{cross}(u)$ and $M_1^{cross}(u)$ are blocks of $D_\infty(u; 2m)$, i.e.

$$D_\infty(u; 2m) = \begin{bmatrix} M_0^{cross}(u; m) & M_1^{cross}(u; m) \\ M_1^{cross}(u; m) & M_0^{cross}(u; m) \end{bmatrix}.$$

This structure may be useful in compressing formulas or speeding up computations of multilocus probabilities.

The computational effort needed to evaluate Erlang multilocus probabilities need not be an obstacle to using them in a genetic linkage study. One can precompute many of the terms needed in the formulas and the remaining computations are small compared to the total computation time used in linkage programs.

The author was supported during this work as a Visiting Investigator at the National Human Genome Research Institute, National Institutes of Health. The author would like to thank Soumitra Ghosh for discussions on the genetics and Richard Holzsager for discussions on the mathematics in this paper.

References

[1] Bailey, N. T. J. (1961). *Introduction to the Mathematical Theory of Genetic Linkage*. Oxford University Press, London.

[2] Cobbs, G. (1978). Renewal process approach to the theory of genetic linkage. *Genetics* **89**:563-581.

[3] Erdélyi, A. Magnus, W., Oberhettinger, F. and Tricomi, F. G. (1955). *Higher Transcendental Functions, Vol. 3* McGraw-Hill, NY.

[4] Falconer, D. S. (1947). Linkage of Rex with Shaker-2 in the house mouse. *Heredity* **1**:133-135.

[5] Fisher, R. A., Lyon, M. F. and Owen, A. R. G. (1947). The sex chromosome in the house mouse. *Heredity* **1**:355-365.

[6] Foss, E., Lande, F., Stahl, F. W., and Steinberg, C. M. (1993). Chiasma interference as a function of genetic distance. *Genetics* **133**:681-691. Erratum *Genetics* **134**:997.

[7] Foss, E. and Stahl, F. W. (1995). A test of a counting model for chiasma interference. *Genetics* **139**:1201-1209.

[8] Geiringer, H. (1944). On the probability theory of linkage in Mendelian heredity. *Ann. Math. Stat.* **15**:25-57.

[9] Haldane, J. B. S. (1919). The combination of linkage values and the calculation of distances between the loci of linked factors. *J. Genet.* **8**:299-309.

[10] Harushima, Y., *et al.*(1998). A high density rice genetic linkage map with 2275 markers using a single F_2 generation. *Genetics* **148**:479-494.

[11] Karlin, S. and Liberman, U. (1979). A natural class of multilocus recombination processes and the related measures of crossover interference. *Adv. Appl. Probl.* **11**:479-501.

[12] Karlin, S. and Liberman, U. (1983). Measuring interference in the chiasma renewal formation process. *Adv. Appl. Probl.* **15**:471-478.

[13] Karlin, S. and Liberman, U. (1994). Theoretical recombination processes incorporating interference effects. *Theor. Popul. Biol.* **46**:(198-231.

[14] Lander, E. and Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**:241-247.

[15] Lange, K. (1997) *Mathematical and Statistical Methods for Genetic Analysis* Springer-Verlag, NY.

[16] Lange, K., Zhao, H. and Speed, T. (1997). The Poisson-skip model of crossing-over. *Ann. Applied Prob.* **7**, 299-313.

[17] Leadbetter, M. R., Lindgren, G, and Rootzen, H. (1983). Extremes and Related Properties of Random Sequences and Processes. Springer Verlag, NY.

[18] Liberman, U. and Karlin, S. (1984). Theoretical models of genetic map functions. *Theor. Popul. Biol.* **25**:331-346.

[19] Lin, S., Speed, T. (1996). Incorporating crossover interference into pedigree analysis using the chi-square model. *Hum. Heredity*, in press.

[20] Mather, K. (1936). Reduction and equational separation of the chromsomes in bivalents and multivalents. *J. Genet.* **30**:207-235.

[21] Mather, K. (1937). The determination of position in crossing over. II. The chromosome length-chiasma frequency relation. *Cytologia, Jub. Vol.*, 514-526.

[22] McPeek, M. S. and Speed, T. P. (1995). Modeling interference in genetic recombination. *Genetics* **139**:1031-1044.

[23] Muldoon, M. E. and Ungar, A. A. (1996). Beyond Sin and Cos. *Mathematics Magazine* **69**:3-14.

[24] Muller, H. J. (1916). The mechanism of crossing over. *Am. Nat.* **50**:193-207.

[25] Ott, J. (1991). *Analysis of Human Genetic Linkage, Revised Ed.* The Johns Hopkins University Press, Baltimore, MD.

[26] Owen, A. R. G. (1949). The theory of genetical recombination. I. Long-chromosome arms. *Proc. Roy. Soc. London, B* **136**:67-94.

[27] Risch, N. and Lange, K. (1983). Statistical analysis of multilocus recombination. *Biometrics* **39**:949-963.

[28] Schnell, F. W. (1961). Some general formulations of linkage effects in inbreeding. *Genetics* **46**:947-957.

[29] Speed, T. P. (1999). What is a map function? *IMA Volume on Molecular Biology*, ed. by T. Speed and Waterman, Springer Verlag, NY.

[30] Stam, P. (1979). Interference in genetic crossing over and chromosome mapping. *Genetics* **92**:573-594.

[31] Sturtevant, A. H. (1915). The behavior of the chromosomes as studied through linkage. *Z. Indukt. Abstammungs. Vererbungsl.* **13**:234-287.

[32] Weeks, D. E., Lathrop, G. M., Ott, J. (1993). Multipoint mapping under genetic interference. *Hum. Hered.* **43**:86-97.

[33] Wright, M. E. (1947). Two sex linkages in the house mouse with unusual recombination values. *Heredity* **1**:349-354.

[34] Zhao, H., Speed, T. P., McPeek, M. S. (1995). Statistical analysis of crossover interference using the chi-square model. *Genetics* **139**:1045-1056.

[35] Zhao, H. and Speed, T. P. (1996). On genetic map functions. *Genetics* **142**:1369-1377.

John P. Nolan
Department of Mathematics and Statistics
American University
4400 Massachusetts Avenue, NW
Washington, DC 20016
E-mail: jpnolan@american.edu