

RESEARCH

Open Access



Erlang renewal models for genetic recombination

John P. Nolan

Correspondence:
jpnolan@american.edu
Department of Mathematics and
Statistics, American University,
Washington, DC, USA

Abstract

Erlang renewal models, also called chi-squared models, provide a tractable model for genetic recombination that exhibits positive interference. Closed form expressions for multilocus probabilities are derived for the crossover process when it is a renewal process with the distance between crossovers modeled by an Erlang distribution. These expressions yield explicit formulas for the map functions, coincidence functions and distributions of the identity-by-descent process, giving exact results for a class of models that better model observed biological data.

Keywords: Erlang renewal processes, Genetic recombination, Multilocus probabilities, Crossover interference, Generalized hyperbolic functions, Alternating renewal processes

Mathematics Subject Classification (2000): 60C05, 60E15, 92D10

1 Introduction

During the process of meiosis, germ cells are produced from the genetic material an organism has inherited from its parents. The top of Fig. 1 shows a simple case where genetic material from each parent is unchanged. However, with some regularity, the strands of genetic material from the parents cross-over and recombine as in the bottom of Fig. 1, with a mixture of genes from different parents in the final gametes. This mixing of genetic material allows children to have different traits than their parents. Recombination can be used to locate disease genes and other traits on the chromosomes.

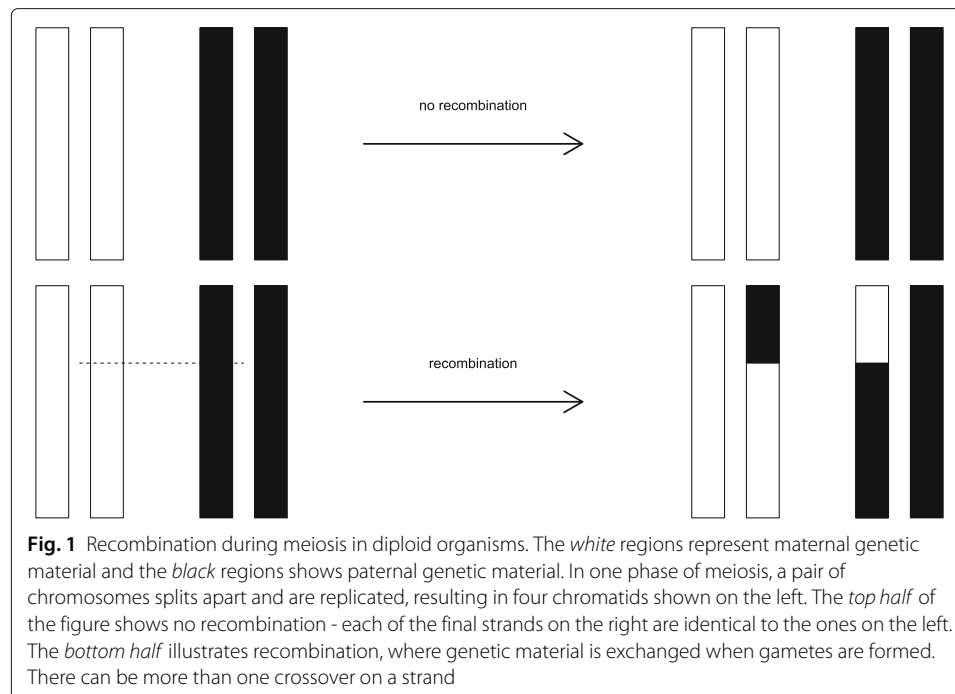
Multilocus probabilities are the basic quantities that are used to build genetic maps and to compute linkage scores. Suppose there are $n + 1$ markers $\mathcal{M}_1, \dots, \mathcal{M}_{n+1}$ along a chromosome. For each of the inter-marker intervals, let

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

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

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

These multilocus probabilities depend on inter-marker distances d_1, \dots, d_n , where $d_j =$ distance between markers \mathcal{M}_j and \mathcal{M}_{j+1} . (Throughout this paper, distances will be expressed in Morgans, e.g. genetic units, 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 and it is mathematically tractable. Its 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), Harushima et al. (1998), and Broman and Weber (2000). Various models have been proposed to represent that interference. A map function is a relation $r = r(d)$ that expresses the recombination fraction r between two locations on a chromosome to the genetic distance d between them. Several authors have attempted to express multilocus probabilities in terms of map functions, e. g. Geiringer (1944); Karlin and Liberman (1994, 1979); Liberman and Karlin (1984); Risch and Lange (1983); Schnell (1961); Weeks et al. (1993). 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 loci and a multilocus probability requires more information. The “adjacent interval” coincidence coefficient (Muller (1916); Sturtevant (1915)) 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 directly from that model. Expressions for map functions and coincidence functions follow from the multilocus probabilities.

The main results of this paper are closed form expressions for multilocus probabilities when the inter-event distribution is Erlang. These results in Section 2 are based on Zhao et al. (1995), where infinite series expressions for multilocus probabilities are given for the chiasma model on the four strand bundle. We show that the (infinite series) matrix

functions they consider can also be used to specify the multilocus probabilities for the crossover process as well, and we give closed form expressions for both the crossover process and the four strand chiasma process. The next section derives closed form expressions for 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.

Since we hope that these results will be used by geneticists, we have attempted to make them more accessible by stating them without the standard lemma/theorem format and separating the proofs to Section 6. The interesting mathematical ideas are that one can explicitly write down the finite dimensional distributions of a counting renewal process with Erlang inter-arrival distances. The crossover process is then an alternating renewal process that toggles between two states (recombination/no recombination), and its finite dimensional distributions are given in closed form. The result is a new class of probability distributions that are based on a novel use of generalized hyperbolic functions. These results can be used in other applications where a system alternates between an on-state and an off-state; some examples are mentioned at the end of the last section.

2 Erlang renewal models

Renewal process models for genetic recombination have a long history in genetics, see Bailey (1961); Cobbs (1978); Fisher et al. (1947); Lange (1997); Owen (1949) and Stam (1979). 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 (Section 1). The mathematical complexity of these models comes from the fact that, except when an exponential distribution is used, the process is non-Markov.

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. McPeck and Speed (1995) fit various data sets to different models of interference, and found that the Erlang models do as good a job 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 $f(x) = \lambda^m x^{m-1} e^{-\lambda x} / (m-1)!$, $x > 0$, see Fig. 2. The name chi-squared model is used by the above paper, by Zhao et al. (1995), and by Armstrong et al. (2006), though they always assume an even number of degrees of freedom. Since it is essential that the shape parameter is an integer for the work below and Erlang laws have a long history in queueing theory, it seems appropriate to call these Erlang models. 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).

For our purposes it is convenient to parameterize the Erlang distributions in the form $\text{Erlang}(m, \lambda m)$, where m is a positive integer and λ is a positive number. In Lin and Speed (1996) and Zhao et al. (1995), 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).

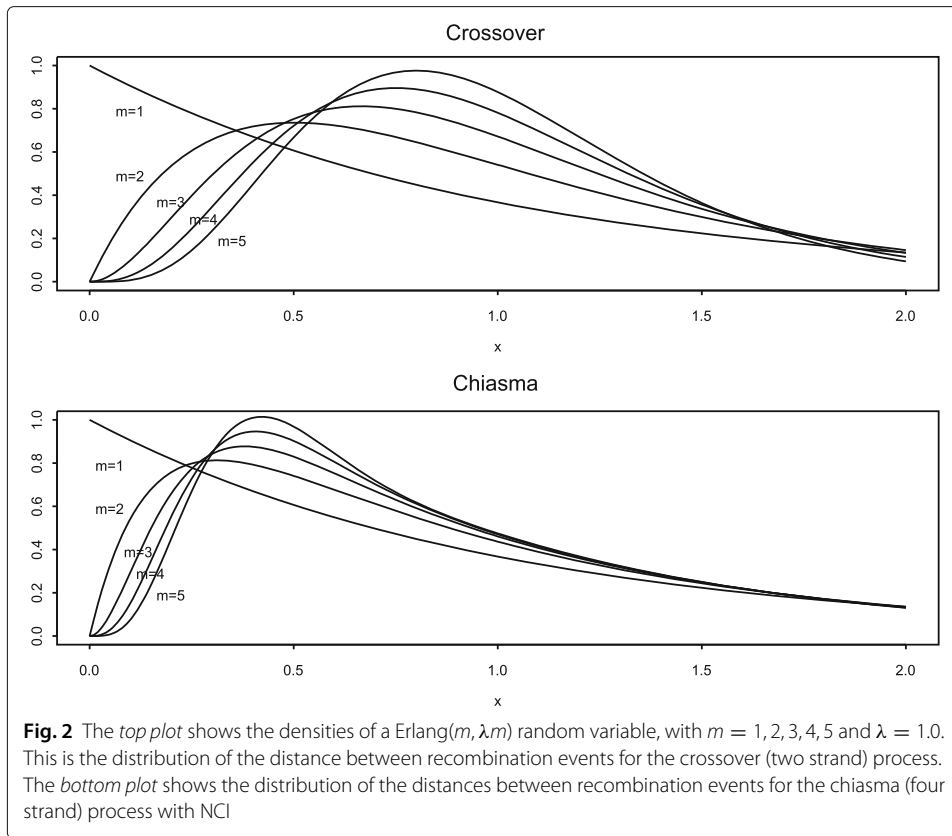


Figure 2 of Harushima, et al. (1998) shows a plot of distances between 555 recombinations for a rice data set. It is poorly described by the Haldane model ($m = 1$), but well described by an Erlang distribution with $m = 2$. Likewise Figure 2 of Broman and Weber (2000) shows clear evidence of positive interference, with close recombinations visibly depleted. Broman et al. (2002) discusses the fit to mouse data.

The main point of biological interest in using an Erlang distribution is that as m increases, it is less likely to see two crossovers close to each other. 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. 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.

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 diploid meiosis involves four strands. Lemma 2 shows that a chiasma model for all four strands is equivalent to a crossover model with a different inter-event distribution. In particular, Eq. (9) below gives an formula for the inter-event distribution for a crossover model that yields the same multilocus probabilities as the Erlang chiasma model considered below.

The formula for multilocus probabilities for a crossover renewal process with an Erlang($m, \lambda m$) inter-event distribution is given by:

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

where $\mathbf{1} = (1, \dots, 1)$ 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, $r = 0, \dots, q - 1$, $a_j = a_j(q) = \cos(2\pi j/q)$ 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 the biological process of meiosis in diploid organisms, 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 probabilities 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

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

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.

The effective distance between genetically observable crossovers is shown to be

$$h(u) = \frac{\exp(-2\lambda mu)}{2^{1-1/m}\lambda m} f_{m-1,m} \left(2^{(1-1/m)}\lambda mu \right), \tag{9}$$

in Section 6. Graphs of this distribution are shown in the bottom plot of Fig. 2.

2.3 The tetrad case

Simple organisms like yeast produce tetrads, where all four strands remain together, instead of separating into four distinct gametes as discussed above. For tetrad data, there are three possible tetrad patterns between each pair of markers and therefore there are now 3^n recombination patterns for $n + 1$ markers, each represented by a pattern (i_1, \dots, i_n) , where $i_j \in \{0, 1, 2\}$. The same notation can be used as before, with the multilocus probabilities for tetrads, assuming NCI, given by

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

where

$$M_0^{tetrad}(u) = \frac{e^{-u}}{2} \left(D_\infty(u) - D_0(u) - M_1^{tetrad}(u) \right),$$

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

$$M_2^{tetrad}(u) = \frac{e^{-u}}{2} \left(D_\infty(u) + D_0(u) - M_1^{tetrad}(u) \right).$$

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

$$h_{r,m}(u) = \begin{cases} \frac{2}{3}f_{r,m}(u) + (-1)^r c_{m,r,m} e^{-\alpha u} & \\ \quad + 2 \sum_{j=1}^{m-1} c_{j,r,m} e^{\alpha a_j' u} \cos(\alpha b_j' u - rj\pi/m) & r < m \\ \frac{2}{3}f_{r-m,m}(u) - 2 \frac{u^{r-m}}{(r-m)!} + (-1)^r c_{m,r,m} e^{-\alpha u} & \\ \quad + 2 \sum_{j=1}^{m-1} c_{j,r,m} e^{\alpha a_j' u} \cos(\alpha b_j' u - rj\pi/m) & r \geq m, \end{cases} \tag{10}$$

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 probabilities. 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 identity-by-descent 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

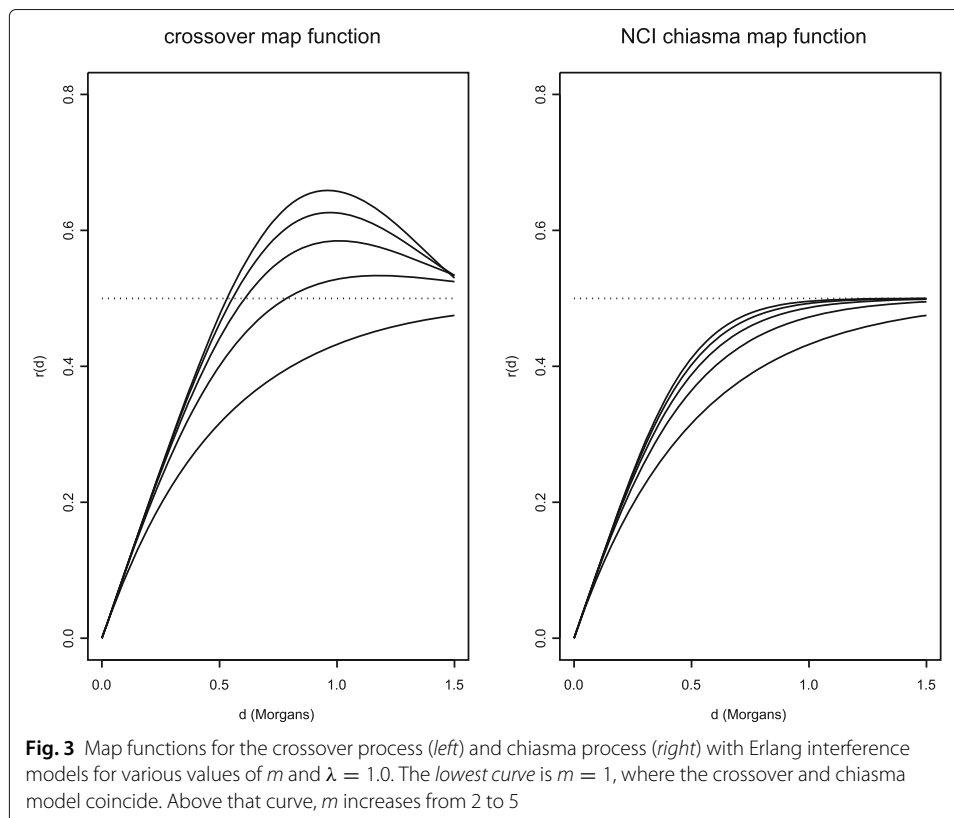
$$r_m^{cross}(d) = \frac{1}{m} \mathbf{1} M_1^{cross}(\lambda md) \mathbf{1}^T = e^{-\lambda dm} f_{m,2m}(\lambda md) + \frac{e^{-\lambda dm}}{m} \sum_{j=1}^{m-1} (m-j) [f_{m+j,2m}(\lambda md) + f_{m-j,2m}(\lambda md)].$$

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

$$\begin{aligned}
 r_m^{NCI}(d) &= \frac{1}{m} \mathbf{1} M_1^{NCI}(2\lambda md) \mathbf{1}^T \\
 &= \frac{e^{-2\lambda dm}}{2} \sum_{j=0}^{m-1} \left(f_{j,m}(2\lambda md) - \binom{m-j}{m} \frac{(2\lambda md)^j}{j!} \right) \\
 &= \frac{1}{2} \left(1 - e^{-2\lambda dm} \frac{1}{m} \sum_{j=0}^{m-1} (m-j) \frac{(2\lambda md)^j}{j!} \right), \tag{11}
 \end{aligned}$$

where the last equality uses (15) in Section 6. This last result is Eq. (30) of Cobbs (1978), and Eq. (7) of Foss et al. (1993).

Figure 3 shows graphs of the map functions 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 $d \rightarrow \infty$. In contrast, under the chiasma model with NCI, Mather’s formula shows that the recombination fraction cannot exceed $1/2$. In our case, this follows from (11) because the sum in the last term is always less than $e^{2\lambda md}$, and thus the term in parentheses is less than 1. 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.) Furthermore both recombination fractions are approximately linear for small distances. This is because (13) shows $f_{r,q}(u) = u^r/r! + o(u^r)$ as $u \rightarrow 0$ and therefore the dominant terms in $r_m^{cross}(\cdot)$ and $r_m^{NCI}(\cdot)$ are of order $\lambda d + o(d)$ as $d \rightarrow 0$.

In describing the coincidence functions, it is convenient to allow the symbol “*” in a multilocus probability to denote either a 0 or a 1. So $p(1, *) = p(1, 0) + p(1, 1)$, 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 et al. (1997) shows all Erlang renewal models have positive interference. The “nonadjacent interval” coincidence coefficient is defined by taking four markers, with inter-marker 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

$$S_3(d) = C_3(d, d) = [r(d) - (1/2)r(2d)] / [r(d)]^2$$

$$S_4(d) = \lim_{d_1 \downarrow 0} \lim_{d_3 \downarrow 0} C_4(d_1, d, d_3).$$

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

$$S_4^{cross}(d) = me^{-\lambda md} f_{m-1,m}(\lambda md)$$

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

Figure 4 shows plots of S_3 and S_4 for both models. We note that the last equation above is a closed form expression for Eq. (8) of Foss et al. (1993).

4 Identity-by-descent process

One goal of genetic linkage studies is to localize genes for disease (or other traits) 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

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

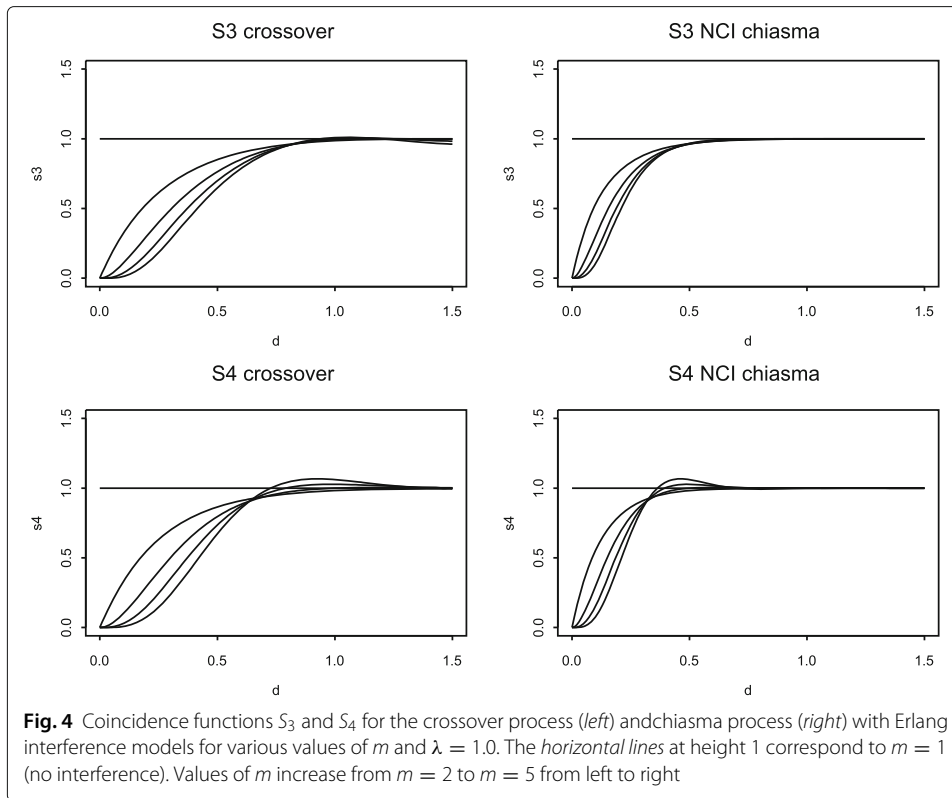
The places where $X(t)$ changes value are precisely the points where a crossover has occurred.

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

$$P(X(t_1) = i_1, X(t_2) = i_2, \dots, X(t_{n+1}) = i_{n+1})$$

$$= P(X(t_1) = i_1, X(t_2) - X(t_1) = |i_2 - i_1|, \dots, X(t_{n+1}) - X(t_n) = |i_{n+1} - i_n|)$$

$$= \frac{1}{2} p(|i_2 - i_1|, |i_3 - i_2|, \dots, |i_{n+1} - i_n|).$$



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.

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. 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

$$\begin{aligned}
 EX(t) &= P(X(t) = 1) = 1/2 \\
 \text{Cov}(X(t+d), X(t)) &= P(X(t) = 1, X(t+d) = 1) - P(X(t) = 1)P(X(t+d) = 1) \\
 &= \frac{1}{2}(1 - r(d)) - \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} - \frac{1}{2}r(d), \tag{12}
 \end{aligned}$$

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; but when $m > 1$ the crossover and chiasma models considered above do not have an Ornstein-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 et al. (1983). That result shows that the threshold for

a dense set of markers depends on the rate at which $\text{Cov}(Z(t), Z(t + d)) \rightarrow 1$ as $d \rightarrow 0$. Using (12),

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

As we noted above, all Erlang renewal processes have $r(d) = \lambda d + o(d)$ (for any m) as $d \rightarrow 0$, so $\text{Cov}(Z(t), Z(t + d)) = 1 - 2\lambda d + o(d)$ as $d \rightarrow 0$. 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 precisely depends on being able to accurately compute cumulative probabilities for multivariate normal distributions with dependence given by (12).

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 Bailey (1961); Cobbs (1978); Owen (1949); 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 apparently can happen in prokaryotes, so these models may be directly applicable there. In fact, the observance of recombination fractions above 1/2 in mouse data, e.g. Falconer (1947) and Wright (1947), 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, Bailey (1961); Fisher et al. (1947) and Owen (1949) 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 observed values of $r > 1/2$ in those older studies.

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, (12) 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 given by (9). In general, any chiasma renewal process is

equivalent to a crossover renewal process with inter-event distribution that is a geometric mixture of the chiasma inter-event distribution.

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 temporarily, inhibiting nearby crossovers. The counting model of Foss et al. (foss1) 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 et al. (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 is whether these models, which incorporate positive interference, do a better job than the commonly used Haldane model. The results of Foss et al. (1993), e. g. Figure 4, Copenhaver et al. (2002); Housworth and Stahl (2003) and McPeck 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 rice data well. Better models for recombination, even if not exactly correct, may help detect disease or trait loci and build genetic maps.

6 Mathematical proofs

We will work with three processes:

$N^*(t)$ = Poisson counting process with Exponential (λ) inter-arrival distances

$N(t)$ = counting process with Erlang ($m, \lambda m$) inter-arrival distances

$X(t)$ = alternating renewal process with Erlang ($m, \lambda m$) inter-arrival distances

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 $N^*(t) = i \pmod{m}$. In the terminology of Foss, et al. (1993), this means that i Co events have occurred since the last Cx event. Another key idea is that $X(t) \stackrel{d}{=} N(t) \pmod{2}$, i.e. the alternating renewal process switches state every time $N(t)$ increases.

We follow the reasoning of Zhao et al. (1995). The matrix $D_0(u)$ was defined in (8) (the zeros below the main diagonal correct a misprint there). For $k = 1, 2, 3, \dots$, 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 $D_k(\cdot)$ matrices have an interpretation as transition matrices: for $k > 0$, the $(i, j)^{th}$ entry of $D_k(u)$ is

$$\begin{aligned} P(N(u) - N(0) = k, N(u) \text{ is in phase } j | N(0) \text{ is in phase } i) \\ &= P(N^*(u) - N^*(0) = (m - i) + m(k - 1) + j) \\ &= P(N^*(u) - N^*(0) = mk + j - i) \\ &= e^{-u}u^{mk+j-1}/(mk + j - i)! \end{aligned}$$

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 $[0, u]$. In particular, for an Erlang renewal process with $N(0)$ having initial distribution \mathbf{p}_0 ,

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

This gives a closed form expression for the finite dimensional distributions of a counting process with Erlang inter-arrival distances. The memoryless property of $N^*(t)$ is what makes the multiplication of matrices give the correct probabilities for $N(t)$. The choice $\mathbf{p}_0 = (1/m)\mathbf{1} = (1/m, \dots, 1/m)$ 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:

$$M_0^{cross}(u) = \sum_{k=0}^{\infty} D_{2k}(u) \quad \text{and} \quad M_1^{cross}(u) = \sum_{k=0}^{\infty} D_{2k+1}(u),$$

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

$$f_{r,q}(u) = \sum_{k=0}^{\infty} u^{qk+r}/(qk+r)! \quad r = 0, \dots, q-1. \tag{13}$$

A closed form expression for these follows from the next lemma.

Lemma 1 $f_{r,q}(u)$ defined by (13) can be written as (5).

Proof Differentiating $f_{r,q}$ with respect to u q times 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, we 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 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

$$f_{r,q}(u) = \frac{1}{q} \sum_{j=0}^{q-1} e^{a_j u} \cos(b_j u - rj\Delta) + i \frac{1}{q} \sum_{j=0}^{q-1} e^{a_j u} \sin(b_j u - rj\Delta). \tag{14}$$

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 $j = 1, \dots, m$, $a_{q-j} = a_j$, $b_{q-j} = -b_j$, 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 $(q-j)^{\text{th}}$ terms will cancel as above. □

If we sum (13) over $r = 0, 1, \dots, q-1$, all powers appear in the series, leading to

$$\sum_{r=0}^{q-1} f_{r,q}(u) = \sum_{j=0}^{\infty} u^j / j! = e^u. \tag{15}$$

We next consider the chiasma process on the four strand bundle. Theorem 1 of Zhao et al. (1995) gives an infinite series expression for (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).

Next the derivation of (9) is given. A p -thinning of a point process is when successive points are retained with probability p or eliminated with probability $1-p$, with independent decisions being made at each point. The effective inter-arrival distribution when an Erlang point process is thinned is given in the following lemma. As mentioned above, the chiasma model is obtained by a $(1/2)$ -thinning of a crossover model with rate $\mu = 2\lambda m$, which yields (9).

Lemma 2 *A p -thinning of an Erlang(m, μ) point process is a point process with inter-arrival density*

$$h(x) = \frac{p e^{-\mu x}}{\mu (1-p)^{1+1/m}} f_{m-1,m}((1-p)^{1/m} \mu x).$$

Proof Set $q = 1 - p$ and $\Delta = \mu q^{1/m}$, then the density is

$$\begin{aligned} h(x) &= \sum_{j=1}^{\infty} q^{j-1} p \frac{\mu^j x^{jm-1} e^{-\mu x}}{(jm-1)!} = \frac{pe^{-\mu x}}{q} \sum_{j=1}^{\infty} q^j \frac{\mu^j x^{jm-1}}{(jm-1)!} \\ &= \frac{pe^{-\mu x}}{q\Delta} \sum_{j=1}^{\infty} \frac{(\Delta x)^{jm-1}}{(jm-1)!} = \frac{pe^{-\mu x}}{q\Delta} f_{m-1,m}(\Delta x), \end{aligned}$$

where the last equality uses (13). □

The derivation of multilocus probabilities in the tetrad case requires a more involved argument.

Lemma 3 *The matrices M_0^{tetrad} , M_1^{tetrad} and M_2^{tetrad} are given by (10).*

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

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

where $p_0^{(k)} = p_2^{(k)} = \frac{1}{3} \left(\frac{1}{2} + \left(-\frac{1}{2}\right)^k \right)$, $p_1^{(k)} = \frac{2}{3} \left(1 - \left(-\frac{1}{2}\right)^k \right)$. 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 (10). 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 $h_{r,m}(u) = \sum_{k=1}^{\infty} p_1^{(k)} u^{mk+r} / (mk+r)!$, $r = 0, \dots, 2m-1$. 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 (13)

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

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

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

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 (10). \square

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 $\mathbf{v} = (0, \dots, 0, m)$, 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 (13). 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 a few miscellaneous comments. 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 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.

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 focused on the genetic application of these results, the results 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. 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.

When $m = 1$, i.e. the inter-event distances are exponential, the alternating renewal process is called a random telegraph process and is used in the study of defects in semiconductors, see Simoen and Claeys (2016). Using $m > 1$ gives a generalized random telegraph process; such processes may be useful in semiconductor models.

R programs for computations of map distances, multilocus probabilities and coincidence functions for Erlang renewal models are available from the author.

Acknowledgements

The author would like to thank Soumitra Ghosh for discussions on the genetics in this paper, Richard Holzsager for discussions on the mathematics in this paper, and the anonymous referees whose suggestions improved the paper.

Competing interests

The author declares that he has no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 6 March 2017 Accepted: 6 July 2017

Published online: 15 October 2017

References

- Armstrong, NJ, McPeck, MS, Speed, TP: Incorporating interference into linkage analysis for experimental crosses. *Biostatistics*. **7**, 374–386 (2006)
- Bailey, NTJ: Introduction to the mathematical theory of genetic linkage. Oxford University Press, London (1961)
- Broman, KW, Weber, JL: Characterization of human crossover interference. *Amer. J. of Human Genetics*. **66**, 1911–1926 (2000)
- Broman, K, Rowe, L, Churchill, G, Paigen, K: Crossover Interference in the Mouse. *Genetics*. **160**, 1123–1131 (2002)
- Cobbs, G: Renewal process approach to the theory of genetic linkage. *Genetics*. **89**, 563–581 (1978)
- Copenhaver, GP, Housworth, EA, Stahl, FW: Crossover Interference in Arabidopsis. *Genetics*. **160**, 1631–1639 (2002)
- Erdélyi, A, Magnus, W, Oberhettinger, F, Tricomi, FG: Higher transcendental functions, Vol. 3. McGraw-Hill, NY (1955)
- Falconer, DS: Linkage of Rex with Shaker-2 in the house mouse. *Heredity*. **1**, 133–135 (1947)
- Fisher, RA, Lyon, MF, Owen, ARG: The sex chromosome in the house mouse. *Heredity*. **1**, 355–365 (1947)
- Foss, E, Lande, F, Stahl, FW, Steinberg, CM: Chiasma interference as a function of genetic distance. *Genetics*. **133**, 681–691 (1993). Erratum *Genetics* **134**:997
- Foss, E, Stahl, FW: A test of a counting model for chiasma interference. *Genetics*. **139**, 1201–1209 (1995)
- Geiringer, H: On the probability theory of linkage in Mendelian heredity. *Ann. Math. Stat.* **15**, 25–57 (1944)
- Haldane, JBS: The combination of linkage values and the calculation of distances between the loci of linked factors. *J. Genet.* **8**, 299–309 (1919)
- Harushima, Y, et al: A high density rice genetic linkage map with 2275 markers using a single F_2 generation. *Genetics*. **148**, 479–494 (1998)
- Housworth, EA, Stahl, FW: Crossover Interference in Humans. *Am. J. Human Genet.* **73**, 188–197 (2003)
- Karlin, S, Liberman, U: A natural class of multilocus recombination processes and the related measures of crossover interference. *Adv. Appl. Probl.* **11**, 479–501 (1979)
- Karlin, S, Liberman, U: Measuring interference in the chiasma renewal formation process. *Adv. Appl. Probl.* **15**, 471–478 (1983)
- Karlin, S, Liberman, U: Theoretical recombination processes incorporating interference effects. *Theor. Popul. Biol.* **46**, 198–231 (1994)
- Kwiatkowski, DJ, Dib, C, Slaugenhaupt, SA, Povey, S, Gusella, JF, Haines, JL: An index marker map of chromosome 9 provides strong evidence for positive interference. *Am. J. Hum. Genet.* **53**, 1279–1288 (1993)
- Lander, E, Kruglyak, L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.* **11**, 241–247 (1995)
- Lange, K: Mathematical and statistical methods for genetic analysis. Springer-Verlag, NY (1997)
- Lange, K, Zhao, H, Speed, T: The Poisson-skip model of crossing-over. *Ann. Applied Prob.* **7**, 299–313 (1997)
- Leadbetter, MR, Lindgren, G, Rootzen, H: Extremes and related properties of random sequences and processes. Springer Verlag, NY (1983)
- Liberman, U, Karlin, S: Theoretical models of genetic map functions. *Theor. Popul. Biol.* **25**, 331–346 (1984)
- Lin, S, Speed, T: Incorporating crossover interference into pedigree analysis using the χ^2 model. *Hum. Heredity*. **46**, 315–322 (1996)
- Mather, K: Reduction and equational separation of the chromosomes in bivalents and multivalents. *J. Genet.* **30**, 207–235 (1936)
- Mather, K: The determination of position in crossing over. II. The chromosome length- chiasma frequency relation. *Cytologia*. **FujiiJubilai(1)**, 514–526 (1937). doi:10.1508/cytologia.FujiiJubilai.514
- McPeck, MS, Speed, TP: Modeling interference in genetic recombination. *Genetics*. **139**, 1031–1044 (1995)
- Muldoon, ME, Ungar, AA: Beyond Sin and Cos. *Math. Mag.* **69**, 3–14 (1996)
- Muller, HJ: The mechanism of crossing over. *Am. Nat.* **50**, 193–207 (1916)
- Owen, ARG: The theory of genetical recombination. I. Long-chromosome arms. *Proc. Roy. Soc. London B*. **136**, 67–94 (1949)
- Risch, N, Lange, K: Statistical analysis of multilocus recombination. *Biometrics*. **39**, 949–963 (1983)
- Schnell, FW: Some general formulations of linkage effects in inbreeding. *Genetics*. **46**, 947–957 (1961)
- Simoen, E, Claeys, C: Random telegraph signals in semiconductor devices. IOP Publishing, Bristol (2016)
- Speed, TP: What is a map function? IMA Volume on Molecular Biology (Speed, T, Waterman, eds.) Springer Verlag, NY (1999)
- Stam, P: Interference in genetic crossing over and chromosome mapping. *Genetics*. **92**, 573–594 (1979)
- Sturtevant, AH: The behavior of the chromosomes as studied through linkage. *Z. Indukt. Abstammungs. Vererbungsleh.* **13**, 234–287 (1915)
- Weeks, DE, Lathrop, GM, Ott, J: Multipoint mapping under genetic interference. *Hum. Hered.* **43**, 86–97 (1993)
- Wright, ME: Two sex linkages in the house mouse with unusual recombination values. *Heredity*. **1**, 349–354 (1947)
- Zhao, H, McPeck, MS, Speed, TP: Statistical analysis of crossover interference using the chi-square model. *Genetics*. **139**, 1045–1056 (1995)
- Zhao, H, Speed, TP: On genetic map functions. *Genetics*. **142**, 1369–1377 (1996)