

THE POISSON-SKIP MODEL OF CROSSING-OVER¹

BY KENNETH LANGE, HONGYU ZHAO AND TERENCE P. SPEED

*University of Michigan, Yale University and
University of California, Berkeley*

The Poisson-skip model introduced in this paper generalizes the chi-square model of crossover interference. Both models are constructed from the random points of a Poisson process occurring along a meiotic bundle of four chromatids. The points of the Poisson process are divided into χ points and o points, with χ points corresponding to crossovers. In the chi-square model, a fixed number of o points intervene between every adjacent pair of χ points; in the Poisson-skip model, a random number of o points intervene. Both of these renewal models permit reasonably straightforward calculation of gamete and tetrad probabilities for multiple linked markers. We illustrate the data analysis possibilities of the Poisson-skip model by fitting it to classical recombination data on *Drosophila*, the mouse, and *Neurospora*. We also describe conditions on the discrete skip distribution that guarantee positive interference.

1. Introduction. The phenomenon of genetic interference first observed by Sturtevant (1915) and Muller (1916) in *Drosophila* has proved to be almost universal in higher organisms. Although many mathematical models have been proposed in the literature to describe crossing-over, the biological mechanisms of interference remain obscure. The recent review by McPeck and Speed (1995) highlights the renewal process models pioneered by Fisher, Lyon and Owen (1947) and subsequently elaborated by Owen (1950), Carter and Robertson (1952), Payne (1956), Cobbs (1978), Stam (1979), Karlin and Liberman (1983), Foss, Lande, Stahl and Steinberg (1993), and Zhao, Speed and McPeck (1995a). Competing point process models include the count-location model [Karlin and Liberman (1979), Risch and Lange (1979), Goldgar and Fain (1988)] and the King and Mortimer (1990) polymerization model.

To fit a specific model to recombination data involving several linked loci, one must calculate the probabilities of the various gamete types. These probabilities are nontrivial, and most mathematical work on renewal process models has focused on the chi-square model, where it is assumed that the interevent (or interarrival) distribution is chi-square with an even number of degrees of freedom $2m$. The chi-square model has been interpreted by Foss, Lande, Stahl and Steinberg (1993) as a Poisson process involving two types of random points. Changing their notation slightly, we designate these points as χ and o points. Chiasmata, which are the cytological sites of crossing-over,

Received April 1996; revised November 1996.

¹Research supported in part by USPHS Grants GM-53275 (Lange) and HG-01093 (Zhao and Speed).

AMS 1991 subject classifications. Primary 92D10; secondary 60K05.

Key words and phrases. Genetic recombination, interference, Poisson process, Markov chain, renewal process, reliability theory.

correspond to χ points; o points merely serve to separate χ points. Between each successive pair of χ points exactly $m - 1$ o random points intervene. The level of interference associated with the chi-square model increases with its order m .

The chi-square model fits data from a variety of organisms well [Zhao, Speed and McPeck (1995a)]. Its major statistical drawback is that the order of the model is confined to the positive integers. This discreteness of the underlying parameter space is a nuisance. In the current paper, we extend the chi-square renewal model to the more general Poisson-skip renewal model for which this objection no longer holds. [In his appendix to the paper by Foss, Lande, Stahl and Steinberg (1993), Lande partially anticipates this extension.] The Poisson-skip model still permits relatively painless calculation of multi-locus gamete probabilities. We demonstrate the versatility of the Poisson-skip model by fitting it to classical recombination data on *Drosophila*, the mouse, and *Neurospora*. In contrast to the chi-square model, the Poisson-skip model can entail negative as well as positive interference. Thus, it is of considerable interest to formulate sufficient conditions for positive interference. Using arguments from reliability theory, we establish such conditions.

2. The Poisson-skip model. During meiosis, two homologous chromosomes duplicate to form two pairs of two sister chromatids each. These two pairs then align in a synapsed bundle of four chromatids. Each crossover along the chromatid bundle involves two nonsister chromatids. If the random pair of chromatids participating in a given crossover depends on the pairs chosen at other crossovers, then the crossing-over process exhibits chromatid interference. Because there is little consistent evidence for this phenomenon [Zhao, McPeck and Speed (1995b)], we assume its absence in the sequel. Convincing evidence does exist for chiasma interference (Zhao, Speed and McPeck (1995a)). This second form of interference appears to involve the suppression (positive interference) or induction (negative interference) of additional chiasmata in the vicinity of a chiasma already formed. Positive interference is the rule.

In the genetic crosses of the kind considered here, we will assume that relevant parents are fully informative for linkage in the sense that it is possible to classify the recombination pattern of each of their contributed gametes. If we consider $n + 1$ ordered marker loci along a chromosome, then such a parent should be heterozygous at each participating locus and should possess two distinguishable haplotypes $(M_1, M_2, \dots, M_{n+1})$ and $(m_1, m_2, \dots, m_{n+1})$. The recombination pattern displayed by one of the parent's transmitted gametes is summarized by a sequence (i_1, i_2, \dots, i_n) of indicators, where $i_j = 1$ corresponds to recombination and $i_j = 0$ to nonrecombination between loci j and $j + 1$. For instance, with three loci, the two equivalent gametes M_1, m_2, m_3 and m_1, M_2, M_3 are represented by the sequence $(1, 0)$.

Although the chi-square model was initially pursued because of its mathematical tractability, it has recently been revived by Foss, Lande, Stahl and Steinberg (1993) on biological grounds. They suggest that certain gene conver-

sion events occur along the chromatid bundle according to a Poisson process. In their interpretation, some of these crossover intermediates, the χ points, resolve into chiasmata; the o crossover intermediates do not. Under the chi-square model, if a random point is a χ point, then the next $m - 1$ random points are o points. These are followed by a χ point, then $m - 1$ o points, and so forth. The resulting chiasma process is stationary provided each random point is assigned probability $1/m$ of being a χ point. The chi-square model makes the specific prediction that tetrads with close double crossovers will be enriched for o points. Unfortunately, Foss and Stahl (1995) find no evidence for such structures in yeast.

In spite of this empirical failure, the close fit of the chi-square model to recombination data argues that it is worth refining as a statistical tool. The Poisson-skip model now introduced generalizes the chi-square model in a useful way. Both the chi-square and Poisson-skip models are generated by a Poisson process with constant intensity λ and by a skip distribution s_n on the positive integers. Random Poisson points are divided into o points and χ points. The o points are “skipped” to reach χ points or chiasmata. In the case of the chi-square model, a fixed number $m - 1$ of o points are always skipped to reach the next χ point. In the more general Poisson-skip model, one skips $n - 1$ o points with probability s_n . The choice of how many o points to skip is made independently at each χ point. This probabilistic mechanism creates a renewal process with interevent distribution

$$(1) \quad F(x) = \sum_{n=1}^{\infty} s_n \sum_{m=n}^{\infty} \frac{(\lambda x)^m}{m!} e^{-\lambda x}.$$

The Poisson tail probability $\sum_{m=n}^{\infty} e^{-\lambda x} (\lambda x)^m / m!$ appearing in this formula is the probability that the n th random point to the right of the current χ point lies within a distance x of the current χ point. If $\omega = \sum_{n=1}^{\infty} n s_n$ denotes the mean number of random points until the next χ point, then Wald’s formula implies that $F(x)$ has mean ω / λ . Hence, the density of the equilibrium distribution (Ross, 1983) to the next χ point is

$$\begin{aligned} \frac{\lambda}{\omega} [1 - F(x)] &= \frac{\lambda}{\omega} \sum_{n=1}^{\infty} s_n \left[1 - \sum_{m=n}^{\infty} \frac{(\lambda x)^m}{m!} e^{-\lambda x} \right] \\ &= \frac{\lambda}{\omega} \sum_{n=1}^{\infty} s_n \sum_{m=0}^{n-1} \frac{(\lambda x)^m}{m!} e^{-\lambda x}. \end{aligned}$$

Mather’s (1938) formula

$$(2) \quad \theta = \frac{1}{2} [1 - \Pr(N_{[a, a+x]} = 0)]$$

gives the recombination fraction of a chromosome interval $[a, a + x]$ in terms of the random number $N_{[a, a+x]}$ of chiasmata occurring on the corresponding bundle interval. If we require the Poisson-skip model to be stationary, then

formula (2) reduces to

$$\begin{aligned} \theta &= \frac{1}{2} \left\{ 1 - \frac{\lambda}{\omega} \int_x^\infty [1 - F(y)] dy \right\} \\ &= \frac{1}{2} \left[1 - \frac{\lambda}{\omega} \int_x^\infty \sum_{n=1}^\infty s_n \sum_{m=0}^{n-1} \frac{(\lambda y)^m}{m!} e^{-\lambda y} dy \right]. \end{aligned}$$

This can be simplified by successive integrations by parts. Indeed,

$$\int_x^\infty \frac{(\lambda y)^m}{m!} e^{-\lambda y} dy = \frac{1}{\lambda} \sum_{k=0}^m \frac{(\lambda x)^k}{k!} e^{-\lambda x}$$

implies that

$$\begin{aligned} \theta &= \frac{1}{2} \left[1 - \frac{1}{\omega} \sum_{n=1}^\infty s_n \sum_{m=0}^{n-1} \sum_{k=0}^m \frac{(\lambda x)^k}{k!} e^{-\lambda x} \right] \\ &= \frac{1}{2} \left[1 - \frac{e^{-\lambda x}}{\omega} \sum_{n=1}^\infty s_n \sum_{k=0}^{n-1} (n-k) \frac{(\lambda x)^k}{k!} \right]. \end{aligned}$$

3. Gamete probabilities. We now turn to the problem of calculating gamete probabilities under the Poisson-skip model. It is helpful to consider two associated Markov chains as we move from left to right along the chromatid bundle. The first of these has state space $\{0, 1, 2, \dots\}$. When this chain is in state 0, the most recently encountered point was a χ point. When it is in state $i > 0$, it must pass exactly $i - 1$ o points before encountering the next χ point to the right. If the chain is currently in state 0, then we adopt the perspective that it decides how many o points to skip simultaneously with moving to the next random point. Thus, the chain exits state 0 to state $n - 1$, $n > 1$, with transition rate λs_n . If the chain decides to skip no o points, then it remains in state 0. If the chain is currently in state $n > 0$, then it encounters the next random point and falls back to state $n - 1$ with transition rate λ . Finally, if at most $r - 1$ o points can be skipped, then the motion of the chain on the reduced state space $\{0, 1, 2, \dots, r - 1\}$ is summarized by the infinitesimal generator

$$\Gamma = \begin{matrix} & \begin{matrix} 0 & 1 & \cdots & r-2 & r-1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ \vdots \\ r-1 \end{matrix} & \begin{pmatrix} -\lambda(1 - s_1) & \lambda s_2 & \cdots & \lambda s_{r-1} & \lambda s_r \\ \lambda & -\lambda & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & -\lambda \end{pmatrix} \end{matrix}.$$

It is straightforward to check that the chain has equilibrium distribution π with components $\pi_m = \omega^{-1} \sum_{n>m} s_n$. Indeed, this follows directly from the

balance conditions

$$-\frac{1}{\omega} \sum_{n>0} s_n \lambda (1 - s_1) + \frac{1}{\omega} \sum_{n>1} s_n \lambda = 0$$

$$\frac{1}{\omega} \sum_{n>0} s_n \lambda s_{m+1} - \frac{1}{\omega} \sum_{n>m} s_n \lambda + \frac{1}{\omega} \sum_{n>m+1} s_n \lambda = 0$$

and the constraint $\sum_{n>0} s_n = 1$.

The second Markov chain is identical to the first except for an absorbing state 0_{abs} . When the Poisson-skip process hits a new χ point, the second Markov chain exits to state 0_{abs} . Thus, the chain moves to state 0_{abs} from state 0 with transition rate λs_1 . In state 1 it moves to state 0_{abs} instead of state 0. Otherwise, the second chain behaves exactly as the first chain. If at most $r - 1$ o points can be skipped, then the second chain has infinitesimal generator

$$\Delta = \begin{matrix} & \begin{matrix} 0 & 1 & \cdots & r-2 & r-1 & 0_{\text{abs}} \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ \vdots \\ r-1 \\ 0_{\text{abs}} \end{matrix} & \begin{pmatrix} -\lambda & \lambda s_2 & \cdots & \lambda s_{r-1} & \lambda s_r & \lambda s_1 \\ 0 & -\lambda & \cdots & 0 & 0 & \lambda \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & -\lambda & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix} \end{matrix}.$$

We can capture the dynamics of both Markov chains by computing the matrix exponentials $P(t) = \exp(t\Gamma)$ and $Q(t) = \exp(t\Delta)$. The entry $p_{ij}(t)$ in row i and column j of $P(t)$ gives the probability that the Poisson-skip process moves from state i of the first chain at time 0 to state j of the same chain at time t . The corresponding entry $q_{ij}(t)$ of $Q(t)$ provides the probability that the process moves from state i of the first chain to state j of the first chain without encountering a χ point during the interim.

It is possible to find explicit expressions for these probabilities by simple path counting arguments. In fact,

$$(3) \quad q_{0j}(t) = \begin{cases} e^{-\lambda t}, & j = 0, \\ \sum_{k>j} s_k \frac{(\lambda t)^{k-j}}{(k-j)!} e^{-\lambda t}, & j > 0 \end{cases}$$

for $i = 0$, and

$$(4) \quad q_{ij}(t) = \begin{cases} 0, & j > i \text{ or } j = 0, \\ \frac{(\lambda t)^{i-j}}{(i-j)!} e^{-\lambda t}, & 0 < j \leq i \end{cases}$$

for $i > 0$. The expression for $q_{00}(t)$ in (3) is based on the observation that the Poisson-skip process cannot leave state 0 and return to it without encountering a χ point. The process stays in state 0 with probability $e^{-\lambda t}$. On the other hand, the process can leave state 0 and wind up in state $j > 0$ if the k th point to its right is the next χ point, and if it encounters $k - j$ o points during the time interval $[0, t]$. Conditioning on the value of k gives the expression in (3) for

$q_{0j}(t)$ when $j > 0$. Similar reasoning produces the expressions in (4) for $q_{ij}(t)$ when $i > 0$.

The entries $p_{ij}(t)$ of $P(t)$ are harder to calculate, but useful formulas can be derived by considering the discrete renewal process corresponding to the skip distribution $\{s_k\}_{k \geq 1}$. One can characterize this process by defining u_n to be the probability that the n th point to the right of the current χ point is a χ point. The u_n satisfy the initial condition $u_0 = 1$ and the classical recurrence relation

$$(5) \quad u_n = s_1 u_{n-1} + s_2 u_{n-2} + \dots + s_{n-1} u_1 + s_n u_0.$$

The recurrence (5) is derived by conditioning on the position of the last χ point before point n .

Armed with the probabilities u_n , we can now express

$$(6) \quad p_{ij}(t) = \mathbf{1}_{\{0 < j \leq i\}} \frac{(\lambda t)^{i-j}}{(i-j)!} e^{-\lambda t} + \mathbf{1}_{\{j=0\}} \sum_{n=0}^{\infty} u_n \frac{(\lambda t)^{i+n}}{(i+n)!} e^{-\lambda t} + \mathbf{1}_{\{j>0\}} \sum_{n=0}^{\infty} u_n \sum_{k>j} s_k \frac{(\lambda t)^{i+n+k-j}}{(i+n+k-j)!} e^{-\lambda t}.$$

Indeed, the first term $e^{-\lambda t} (\lambda t)^{i-j} / (i-j)!$ of (6) expresses the probability of encountering $i - j$ o points during $[0, t]$; this is relevant when there is a direct path from state i to j that does not pass through state 0. The term $u_n e^{-\lambda t} (\lambda t)^{i+n} / (i+n)!$ is the probability of passing through the $i - 1$ current o points to the right, hitting the next χ point, and returning to a χ point after encountering n further points. Finally, the term

$$u_n s_k \frac{(\lambda t)^{i+n+k-j}}{(i+n+k-j)!} e^{-\lambda t}$$

is the probability of passing through the $i - 1$ current o points to the right, hitting the next χ point, returning to a χ point after encountering n further points, and then passing through $k - j$ remaining o points enroute to a χ point k further points to the right.

The finite-time transition probabilities $p_{ij}(t)$ and $q_{ij}(t)$ permit calculation of the probability of an arbitrary gamete recombination pattern (i_1, i_2, \dots, i_n) over $n + 1$ loci. Let t_j be the distance between loci j and $j + 1$. For our purposes, it is crucial to keep track of the state X_j of the first Markov chain at locus j as j increases from 1 to $n + 1$. The X_j are unobservable. What is observable are the indicator random variables $Y_j = i_{j-1}$ flagging for each j whether a recombination event has occurred between loci $j - 1$ and j . One strategy to compute $\Pr(Y_2 = i_1, \dots, Y_{n+1} = i_n)$ is to invoke Baum's forward algorithm from the theory of hidden Markov chains [Baum (1972) and Devijver (1985)]. This suggests computing recursively the joint probabilities

$$f_j(x_j) = \Pr(Y_2 = i_1, \dots, Y_j = i_{j-1}, X_j = x_j)$$

beginning with $f_1(x_1) = \pi_{x_1}$ for a stationary process. At locus $n + 1$ we recover the gamete probability $\Pr(Y_2 = i_1, \dots, Y_{n+1} = i_n)$ from the identity

$$\Pr(Y_2 = i_1, \dots, Y_{n+1} = i_n) = \sum_{x_{n+1}} f_{n+1}(x_{n+1}).$$

In view of Mather’s formula (2), if $i_j = 1$, then

$$f_{j+1}(x_{j+1}) = \sum_{x_j} f_j(x_j) \frac{1}{2} [p_{x_j, x_{j+1}}(t_j) - q_{x_j, x_{j+1}}(t_j)]$$

because $[p_{x_j, x_{j+1}}(t_j) - q_{x_j, x_{j+1}}(t_j)]/2$ is the probability that the chain moves from state $X_j = x_j$ at locus j to state $X_{j+1} = x_{j+1}$ at locus $j + 1$ and that the chosen gamete is recombinant on the interval between the loci. On the other hand, if $i_j = 0$, then

$$f_{j+1}(x_{j+1}) = \sum_{x_j} f_j(x_j) \frac{1}{2} [p_{x_j, x_{j+1}}(t_j) + q_{x_j, x_{j+1}}(t_j)]$$

because

$$q_{x_j, x_{j+1}}(t_j) + \frac{1}{2} [p_{x_j, x_{j+1}}(t_j) - q_{x_j, x_{j+1}}(t_j)] = \frac{1}{2} [p_{x_j, x_{j+1}}(t_j) + q_{x_j, x_{j+1}}(t_j)]$$

is the probability that the chain moves from state $X_j = x_j$ at locus j to state $X_{j+1} = x_{j+1}$ at locus $j + 1$ and that the chosen gamete is nonrecombinant on the interval between the loci.

If there is an upper bound $r - 1$ on the number of o points that can be skipped, then the result of the forward algorithm can be summarized by defining the $r \times r$ matrices

$$R_j = \frac{1}{2} [P(t_j) - \tilde{Q}(t_j)]$$

$$N_j = \frac{1}{2} [P(t_j) + \tilde{Q}(t_j)]$$

appropriate to recombination and nonrecombination, respectively, between loci j and $j + 1$. Here $\tilde{Q}(t)$ is the matrix $Q(t)$ omitting its last row and last column. The matrices R_j and N_j provide the explicit expression

$$\Pr(Y_2 = i_1, \dots, Y_{n+1} = i_n) = \pi R_1^{i_1} N_1^{1-i_1} \dots R_n^{i_n} N_n^{1-i_n} \mathbf{1},$$

where π is the equilibrium distribution and $\mathbf{1}$ is a column vector of 1’s. If the process is nonstationary, then we merely replace the equilibrium distribution π in this product by an appropriate initial distribution for the state of the chain at the first locus.

4. Fitting the model to data. In this section, we describe the outcomes of fitting a special form of the Poisson-skip model to data from a variety of organisms. This mixture form has its skip distribution determined by $s_m = \alpha$ and $s_{m+1} = 1 - \alpha$ for $\alpha \in [0, 1]$ and its equilibrium distribution by $\pi_i = \omega^{-1}$ for $0 \leq i \leq m - 1$ and $\pi_m = \omega^{-1}(1 - \alpha)$, where $\omega = m\alpha + (m + 1)(1 - \alpha)$. The case $\alpha = 1$ corresponds to a chi-square model of order m , and the case $\alpha = 0$

to a chi-square model of order $m + 1$. For data on $n + 1$ linked markers, the stationary mixture model involves n scaled genetic distances t_1, \dots, t_n , an interference parameter m , and a mixture parameter α . The intensity parameter λ is arbitrarily set to 1 because only the products λt_i are identifiable. The $n + 2$ parameters of the model can be estimated by the method of maximum likelihood. Optimization methods such as the Nelder–Mead simplex algorithm are particularly attractive in this context because they do not require derivatives of the loglikelihood [Press, Teukolsky, Vetterling and Flannery (1992)].

Table 1 presents the results of fitting the stationary chi-square and mixture models to nine data sets. Data sets five through nine involve *Neurospora*. This bread mold occupies a special place in genetics because one can recover all four products of a meiosis in an ensemble called an ascus. We have extended our calculation of gamete probabilities to the calculation of tetrad probabilities for data of this sort. For the sake of brevity, we omit the theoretical development here. Interested readers can find a partial treatment in [Zhao, McPeck and Speed (1995b)] based on Mather’s early work [Mather (1935)].

The first data set listed in Table 1 was collected by Weinstein (1935), the second by Morgan, Bridges and Schultz (1935), the third by Blank, Campbell, Calabro and Eustachio (1988), the fourth by Todd, Aitman, Cornall, Ghosh and Hall (1991), the fifth by Perkins (1962), and the sixth through ninth by Strickland (1961). For the Morgan, Bridges and Schultz (1935) *Drosophila* data, our findings are in close agreement with the results of McPeck and Speed (1995), who estimate a shape parameter of 4.94 for a gamma interevent distribution. Weeks, Ott and Lathrop (1994) previously found positive interference in several models fitted to the mouse chromosome 1 data of Blank, Campbell, Calabro and Eustachio (1988) and the mouse chromosome 12 data of Todd, Aitman, Cornall, Ghosh and Hall (1991). The four Linkage group I *Neurospora*

TABLE 1
Data analysis with the Poisson-skip model

Organism	N	Best chi-square	Best mixture	$\Delta \ln L$
<i>Drosophila</i> ¹	28,239	$s_5 = 1$	$(s_4, s_5) = (0.40, 0.60)$	0.36
<i>Drosophila</i> ²	16,136	$s_5 = 1$	$(s_4, s_5) = (0.06, 0.94)$	0.53
mouse ³	317	$s_7 = 1$	$(s_6, s_7) = (0.33, 0.67)$	0.02
mouse ⁴	288	$s_5 = 1$	$(s_4, s_5) = (0.30, 0.70)$	0.02
<i>Neurospora</i> ⁵	1,252	$s_3 = 1$	$(s_3, s_4) = (0.58, 0.42)$	0.70
<i>Neurospora</i> ⁶	1,802	$s_3 = 1$	$(s_3, s_4) = (0.44, 0.56)$	0.30
<i>Neurospora</i> ⁶	1,968	$s_3 = 1$	$(s_2, s_3) = (0.16, 0.84)$	1.84
<i>Neurospora</i> ⁶	2,239	$s_3 = 1$	$(s_2, s_3) = (0.15, 0.85)$	2.09
<i>Neurospora</i> ⁶	2,810	$s_3 = 1$	$(s_3, s_4) = (0.34, 0.66)$	1.32

¹Weinstein (1935).

²Morgan, Bridges and Schultz (1935).

³Blank, Campbell, Calabro and Eustachio (1988).

⁴Todd, Aitman, Cornall, Ghosh, et al. (1991).

⁵Perkins (1962).

⁶Strickland (1961).

markers typed by Strickland (1961) are a subset of the six markers typed by Perkins (1962). Table 1 lists the organism, the number of offspring N typed, the best fitting chi-square model, the best fitting mixture model, and the increase in loglikelihood (base e) in passing from the discrete chi-square model to the continuous mixture model.

The improvements in loglikelihoods under the mixture model are modest for the data sets displayed in Table 1. In this regard, it is helpful to recall the major improvements of the chi-square model over competing models such as the count-location model. Furthermore, the mixture model is by far the simplest Poisson-skip model. More complex skip distributions might yield more impressive results.

5. Interference. We now turn to the question of when a Poisson-skip model entails positive chiasma interference. If the two adjacent intervals $[a, b]$ and $[b, c]$ on the chromatid bundle contain $N_{[a, b]}$ and $N_{[b, c]}$ chiasmata respectively, then the coincidence coefficient

$$\begin{aligned}
 C &= \frac{(1/4) \Pr(N_{[a, b]} > 0, N_{[b, c]} > 0)}{(1/2) \Pr(N_{[a, b]} > 0)(1/2) \Pr(N_{[b, c]} > 0)} \\
 &= \frac{(1/4)[1 - \Pr(N_{[a, b]} = 0) - \Pr(N_{[b, c]} = 0) + \Pr(N_{[a, b]} + N_{[b, c]} = 0)]}{(1/2)[1 - \Pr(N_{[a, b]} = 0)](1/2)[1 - \Pr(N_{[b, c]} = 0)]}
 \end{aligned}$$

is the traditional measure of chiasma interference in the absence of chromatid interference. Chiasma interference is positive when $C > 1$ and negative when $C < 1$. Haldane's Poisson process model gives $C = 1$ for all adjacent, and indeed all disjoint, intervals.

When crossing-over is governed by a renewal process with interevent distribution F and density f , Karlin and Liberman (1983) show that positive interference occurs when F has increasing failure rate (IFR) and negative interference occurs when F has decreasing failure rate (DFR). Recall that F has IFR (DFR) if $\bar{F}(t+x)/\bar{F}(t)$ is decreasing (increasing) in t for every x , where $\bar{F}(x) = 1 - F(x)$ is the survivorship function. In practice, one verifies the IFR (DFR) property by showing that the hazard rate $f(x)/\bar{F}(x)$ is increasing (decreasing) in x . The gamma density $f(x) = \beta(\beta x)^{\alpha-1}\Gamma(\alpha)^{-1}e^{-\beta x}$ has IFR for $\alpha > 1$ and DFR for $0 < \alpha < 1$. Hence, the chi-square model always exhibits positive interference. The Weibull distribution $F(x) = 1 - e^{-(\lambda x)^\alpha}$ follows the same interference pattern as the gamma relative to its parameter α . If the crossing-over process is stationary, then these sufficient conditions for positive and negative interference can be replaced by the weaker conditions that F has decreasing mean residual lifetime (DMRL) and increasing mean residual lifetime (IMRL), respectively. If the random variable X has distribution F , then F has DMRL (IMRL) when the conditional expectation $E(X - t | X > t)$ is decreasing (increasing) in t .

To apply these results, it is helpful to introduce a constraint that we will impose on the discrete skip distribution of the Poisson-skip model to achieve

positive interference. A nonnegative sequence $\{c_n\}_{n=-\infty}^{\infty}$ is said to be log concave whenever $c_n^2 \geq c_{n-1}c_{n+1}$ for all n . To avoid certain pathologies entailed by this standard definition, we will add the stipulation that $c_n > 0$ if and only if $n \in I$ for some interval I , which can be finite, semiinfinite, or infinite. For instance, any sequence concentrated on a single integer or two adjacent integers is log concave. The geometric density $c_n = 1_{\{n \geq 0\}}p(1-p)^n$ provides another example.

If for the sake of convenience in the Poisson-skip model we take $\lambda = 1$, then each interevent distance for the Poisson-skip model can be represented as a random sum $S = \sum_{i=1}^N X_i$, where the X_i are independent, exponentially distributed random variables of unit mean and the number of summands $N \geq 1$ is independent of all X_i . If the count variable N has the skip density $\{s_n\}_{n \geq 1}$, then S has the gamma mixture density $\sum_{n=0}^{\infty} w_n e^{-x} x^n / n!$ with $w_{n-1} = s_n$.

THEOREM 1. *In the Poisson-skip model, let \mathcal{C} be the set of discrete densities $\{w_n\}_{n \geq 0}$ for $N-1$ such that $S = \sum_{i=1}^N X_i$ has the IFR property and consequently entails positive or no interference. Then we have the following:*

- (a) \mathcal{C} is closed under weak limits;
- (b) \mathcal{C} is closed under convolution;
- (c) \mathcal{C} contains all discrete log concave densities;
- (d) \mathcal{C} contains all densities concentrated on a single integer or two adjacent integers, all binomial, geometric, Poisson, uniform, and integer-order negative binomial densities, and the shifts of any of these densities by a positive integer.

To avoid interrupting our discussion, we defer the proof to the Appendix.

For a counterexample to the IFR property, consider the mixture density $pe^{-x} + qx^2e^{-x}/2$ with $q = 1 - p$. Its hazard function

$$\frac{p + q(x^2/2)}{p + q(1 + x + (x^2/2))}$$

has derivative

$$\frac{q(qx + q(x^2/2) - p)}{[p + q(1 + x + (x^2/2))]^2},$$

which for small x is negative. The Poisson-skip interevent density

$$\begin{aligned} \frac{p-r}{1-r}e^{-x} + \frac{q}{1-r}re^{-rx} &= pe^{-x} + \frac{qr}{1-r}[e^{(1-r)x} - 1]e^{-x} \\ (7) \qquad \qquad \qquad &= pe^{-x} + q \sum_{n=1}^{\infty} (1-r)^{n-1} re^{-x} \frac{x^n}{n!}. \end{aligned}$$

with $0 < r < p$ provides a more striking counterexample. Because exponential distributions enjoy the DFR property, and the collection of DFR distributions is closed under admixture, the density (7) has the DFR property and consequently entails negative interference.

6. Discussion. Haldane's Poisson model is widely used in gene mapping. Although Speed, McPeck and Evans (1992) show that Haldane's model is asymptotically consistent for ordering multiple linked loci, there is an inevitable loss in statistical efficiency if it is invoked in the presence of chiasma interference [Goldstein, Zhao and Speed (1995)]. Thus, better models of crossing-over are worth pursuing. The stationary renewal models are appealing candidates. For instance, Zhao and Speed (1996) prove that most of the map functions geneticists employ in practice arise from stationary renewal processes. As an analytically tractable subset of renewal models, Poisson-skip models have even more to recommend them.

First, any interevent distribution on $[0, \infty)$ can be arbitrarily well approximated by a Poisson-skip interevent distribution [Neuts (1981)]. Second, most of the analytic advantages of the chi-square model carry over to the Poisson-skip model. For instance, one can extend our calculation of gamete probabilities to the calculation of tetrad probabilities for *Neurospora*. Our data analysis in Section 4 relies on this extension. Third, the Poisson-skip model allows for recombination hot spots since inhomogeneity of the underlying Poisson process relative to physical distance does not contradict its homogeneity relative to genetic distance. Fourth and finally, the Poisson-skip model fits recombination data well. Even a simple mixture of two chi-square distributions performs well on the fruit fly, mouse, and *Neurospora* data examined here.

Among the many crossover models proposed in the literature, only the model of Goldgar and Fain (1988) and a version of the polymerization model of King and Mortimer (1990) fit *Drosophila* data as well as the chi-square model [McPeck and Speed (1995)]. Under each of these three models, it is possible to estimate the joint distributions of the crossovers along a chromosome conditional on there being one, two, and three crossovers on the chromosome. For the data sets we have explored, the estimated distributions are always similar across models. Thus, in spite of incorporating very different mechanisms, all three models appear capable of capturing some of the salient features of the recombination process. It is our hope that the analysis of new, more extensive data sets will better discriminate among the Poisson-skip and these competing models.

There is ample room for improvement and generalization of the Poisson-skip model. For example, one could construct a version of the model that includes chromatid as well as chiasma interference [Zhao (1995)]. Alternatively, one could extend the model to include the phase distributions used with great success in queuing theory [Neuts (1981)]. A phase-type model involves a Markov chain with a single absorbing state and a given initial distribution. Visits to the absorbing state constitute a renewal process. Compared to the Poisson-skip model, phase-type models present greater difficulties in calculating gamete and tetrad probabilities. Furthermore, because the representation of a phase-type distribution may not be unique, some of its underlying parameters may not be identifiable. Finally, whichever of these generalizations is pursued, massive amounts of data will be required to discriminate between it and the Poisson-skip mixture model.

APPENDIX

Before embarking on the proof of Theorem 1, let us consider a less obvious example of a log concave sequence, namely the sequence of partial sums $e_n(x) = \sum_{i=0}^n x^i/i!$ of the exponential function for $x \geq 0$. By convention, we take $e_n(x) = 0$ for $n < 0$. The inequality $e_n(x)^2 \geq e_{n-1}(x)e_{n+1}(x)$ can be restated as

$$\begin{aligned} e_{n-1}(x)^2 + \frac{2x^n}{n!}e_{n-1}(x) + \left(\frac{x^n}{n!}\right)^2 \\ \geq e_{n-1}(x)^2 + \frac{x^n}{n!}e_{n-1}(x) + \frac{x^{n+1}}{(n+1)!}e_{n-1}(x). \end{aligned}$$

To prove this inequality, it suffices to show that

$$\begin{aligned} \sum_{k=0}^n \frac{x^{n+k}}{n!k!} &= \frac{x^n}{n!}e_{n-1}(x) + \left(\frac{x^n}{n!}\right)^2 \\ &= \frac{x^n}{n!}e_n(x) \\ &\geq \frac{x^{n+1}}{(n+1)!}e_{n-1}(x) \\ &= \sum_{k=1}^n \frac{x^{n+k}}{(n+1)!(k-1)!}. \end{aligned}$$

When we examine the coefficient of x^{k+n} on both sides of this last inequality, it is apparent that

$$\frac{1}{n!k!} \geq \frac{1}{(n+1)!(k-1)!}$$

holds because $n+1 > k$. This proves the log concavity of $e_n(x)$ with strict inequality in the defining inequality when $x > 0$.

If we let $m \leq n$ and multiply the inequalities

$$\begin{aligned} c_m^2 &\geq c_{m-1}c_{m+1} \\ c_{m+1}^2 &\geq c_m c_{m+2} \\ &\vdots \\ c_n^2 &\geq c_{n-1}c_{n+1}, \end{aligned}$$

then the inequality

$$\prod_{k=m}^n c_k^2 \geq c_{m-1}c_m \left(\prod_{k=m+1}^{n-1} c_k^2 \right) c_n c_{n+1}$$

results. When $m, n \in I$, we can cancel balancing positive factors on both sides and infer that

$$(8) \quad c_m c_n \geq c_{m-1} c_{n+1}.$$

Inequality (8) remains true when either $m \notin I$ or $n \notin I$ because our added condition on log concave sequences forces both sides to be 0.

PROOF OF THEOREM 1. It is well known that the set of IFR distributions is closed under weak limits and convolutions (Ross 1983). Part (a) is now proved by observing that if $\lim_{k \rightarrow \infty} N_k = N_\infty$ in distribution, then $\lim_{k \rightarrow \infty} \sum_{i=1}^{N_k} X_i = \sum_{i=1}^{N_\infty} X_i$ in distribution. Part (b) follows from the representation

$$\sum_{i=1}^{N_1+N_2} X_i = \sum_{i=1}^{N_1} X_i + \sum_{j=1}^{N_2} X_{N_1+j}$$

for N_1 and N_2 independent. The assertions in part (d) are immediate consequences of parts (a), (b) and (c).

To prove part (c), first note that the survivorship function of the sum S amounts to

$$\sum_{n=0}^{\infty} w_n \int_x^{\infty} \frac{t^n}{n!} e^{-t} dt = \sum_{n=0}^{\infty} w_n e_n(x) e^{-x}$$

in our notation for the partial sums of the exponential function. This representation follows from integration by parts or from standard waiting time arguments for a Poisson process. The derivative of the hazard function of S therefore is

$$\begin{aligned} & \frac{d}{dx} \frac{\sum_{n=0}^{\infty} w_n (x^n/n!) e^{-x}}{\sum_{n=0}^{\infty} w_n e_n(x) e^{-x}} \\ &= \frac{d}{dx} \frac{\sum_{n=0}^{\infty} w_n [e_n(x) - e_{n-1}(x)]}{\sum_{n=0}^{\infty} w_n e_n(x)} \\ &= \frac{\sum_{k=0}^{\infty} w_k [e_{k-1}(x) - e_{k-2}(x)] \sum_{l=0}^{\infty} w_l e_l(x)}{[\sum_{n=0}^{\infty} w_n e_n(x)]^2} \\ (9) \quad & - \frac{\sum_{k=0}^{\infty} w_k [e_k(x) - e_{k-1}(x)] \sum_{l=0}^{\infty} w_l e_{l-1}(x)}{[\sum_{n=0}^{\infty} w_n e_n(x)]^2} \\ &= \frac{[\sum_{k=0}^{\infty} w_k e_{k-1}(x)]^2 - [\sum_{k=0}^{\infty} w_k e_{k-2}(x)] [\sum_{l=0}^{\infty} w_l e_l(x)]}{[\sum_{n=0}^{\infty} w_n e_n(x)]^2} \\ &= \frac{(1/2) \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} [w_k w_l - w_{k-1} w_{l+1}] [e_{k-1}(x) e_{l-1}(x) - e_{k-2}(x) e_l(x)]}{[\sum_{n=0}^{\infty} w_n e_n(x)]^2}, \end{aligned}$$

where the sequences $w_n = e_n(x) = 0$ for $n < 0$. In light of inequality (8), the two factors $w_k w_l - w_{k-1} w_{l+1}$ and $e_{k-1}(x) e_{l-1}(x) - e_{k-2}(x) e_l(x)$ in the last line of (9) are simultaneously nonnegative or nonpositive. Hence, the derivative of the hazard function is nonnegative. \square

REFERENCES

- BARLOW, R. E. and PROSCHAN, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston, New York.
- BAUM, L. E. (1972). An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities* **3** 1–8.
- BLANK, R. D., CAMPBELL, G. R., CALABRO, A. and EUSTACHIO, P. D. (1988). A linkage map of mouse chromosome 12: localization of *Igh* and effects of sex and interference on recombination. *Genetics* **120** 1073–1083.
- CARTER, T. C. and ROBERTSON, A. (1952). A mathematical treatment of genetical recombination using a four-strand model. *Proc. Roy. Soc. London Ser. B* **139** 410–426.
- COBBS, G. (1978). Renewal process approach to the theory of genetic linkage: case of no chromatid interference. *Genetics* **89** 563–581.
- DEVLJVER, P. A. (1985). Baum's forward-backward algorithm revisited. *Pattern Recognition Lett.* **3** 369–373.
- FISHER, R. A., LYON, M. F. and OWEN, A. R. G. (1947). The sex chromosome in the house mouse. *Heredity* **1** 335–365.
- FOSS, E., LANDE, R., STAHL, F. W. and STEINBERG, C. M. (1993). Chiasma interference as a function of genetic distance. *Genetics* **133** 681–691.
- FOSS, E. and STAHL, F. W. (1995). A test of a counting model for chiasma interference. *Genetics* **139** 1201–1209.
- GOLDGAR, D. E. and FAIN, P. R. (1988). Models of multilocus recombination: nonrandomness in chiasma number and crossover positions. *Amer. J. Hum. Genet.* **43** 38–43.
- GOLDSTEIN, D. R., ZHAO, H. and SPEED, T. P. (1995). Relative efficiencies of several statistical models of recombination for exclusion mapping and gene ordering. *Genomics* **27** 265–273.
- KARLIN, S. and LIBERMAN, U. (1979). A natural class of multilocus recombination processes and related measure of crossover interference. *Adv. in Appl. Probab.* **11** 479–501.
- KARLIN, S. and LIBERMAN, U. (1983). Measuring interference in the chiasma renewal formation process. *Adv. in Appl. Probab.* **15** 471–487.
- KING, J. S. and MORTIMER, R. K. (1990). A polymerization model of chiasma interference and corresponding computer simulation. *Genetics* **126** 1127–1138.
- MATHER, K. (1935). Reduction and equational separation of the chromosomes in bivalents and multivalents. *J. Genet.* **30** 53–78.
- MATHER, K. (1938). Crossing over. *Biol. Rev.* **13** 252–292.
- MCPEEK, M. S. and SPEED, T. P. (1995). Modeling interference in genetic recombination. *Genetics* **139** 1031–1044.
- MORGAN, T. H., BRIDGES, C. B. and SCHULTZ, J. (1935). Report of investigations on the constitution of the germinal material in relation to heredity. *Carnegie Institute of Washington* **34** 284–291.
- MULLER, H. J. (1916). The mechanism of crossing over. *Amer. Nat.* **50** 193–207.
- NEUTS, M. F. (1981). *Matrix-Geometric Solutions in Stochastic Models*. Johns Hopkins Univ. Press.
- OWEN, A. R. G. (1950). The theory of genetical recombination. *Adv. Genet.* **3** 117–157.
- PAYNE, L. C. (1956). The theory of genetical recombination: a general formulation for a certain class of intercept length distributions appropriate to the discussion of multiple linkage. *Proc. Roy. Soc. London Ser. B* **144** 528–544.
- PERKINS, D. D. (1962). Crossing-over and interference in a multiply marked chromosome arm of *Neurospora*. *Genetics* **47** 1253–1274.
- PRESS, W. H., TEUKOLSKY, S. A., VETTERLING, W. T. and FLANNERY, B. P. (1992). *Numerical Recipes in Fortran: The Art of Scientific Computing*, 2nd ed. Cambridge Univ. Press.
- RISCH, N. and LANGE, K. (1979). An alternative model of recombination and interference. *Ann. Hum. Genet.* **43** 61–70.
- ROSS, S. M. (1983). *Stochastic Processes*. Wiley, New York.
- SPEED, T. P., MCPEEK, M. S. and EVANS, S. N. (1992). Robustness of the no-interference model for ordering genetic markers. *Proc. Nat. Acad. Sci. U.S.A.* **89** 3103–3106.

- STAM, P. (1979). Interference in genetic crossing over and chromosome mapping. *Genetics* **92** 573–594.
- STRICKLAND, W. N. (1961). Tetrad analysis of short chromosome regions of *Neurospora crassa*. *Genetics* **46** 1125–1141.
- STURTEVANT, A. H. (1915). The behavior of the chromosomes as studied through linkage. *Z. Indukt. Abstammungs. Vererbungsl.* **13** 234–287.
- TODD, J. A., AITMAN, T. J., CORNALL, R. J., GHOSH, S., HALL J. R. et al. (1991). Genetic analysis of autoimmune type 1 diabetes mellitus in mice. *Nature* **351** 542–547.
- WEEKS, D. E., OTT, J. and LATHROP, G. M. (1994). Detection of genetic interference: simulation studies and mouse data. *Genetics* **136** 1217–1226.
- WEINSTEIN, A. (1935). The theory of multiple strand crossing over. *Genetics* **21** 155–199.
- ZHAO, H. (1995). *Statistical Analysis of Genetical Interference*. Ph.D. dissertation, Univ. California, Berkeley.
- ZHAO, H. and SPEED, T. P. (1996). On genetic map functions. *Genetics* **142** 1369–1377.
- ZHAO, H., SPEED, T. P. and MCPEEK, M. S. (1995a). Statistical analysis of crossover interference using the chi-square model. *Genetics* **139** 1045–1056.
- ZHAO, H., MCPEEK, M. S. and SPEED, T. P. (1995b). Statistical analysis of chromatid interference. *Genetics* **139** 1057–1065.

KENNETH LANGE
DEPARTMENTS OF BIOSTATISTICS
AND MATHEMATICS
UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109-2029
E-MAIL: klange@umich.edu

H. ZHAO
DEPARTMENT OF EPIDEMIOLOGY
AND PUBLIC HEALTH
YALE UNIVERSITY SCHOOL OF MEDICINE
NEW HAVEN, CONNECTICUT 06520-8034
E-MAIL: hz27@email.med.yale.edu

T. P. SPEED
DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720-3860
E-MAIL: terry@stat.berkeley.edu