The Relationship Between Count-Location and Stationary Renewal Models for the Chiasma Process

Sharon Browning

Department of Statistics, Texas A&M University, College Station, Texas 77843

Manuscript received September 15, 1999
Accepted for publication April 12, 2000

ABSTRACT

It is often convenient to define models for the process of chiasma formation at meiosis as stationary renewal models. However, count-location models are also useful, particularly to capture the biological requirement of at least one chiasma per chromosome. The Sturt model and truncated Poisson model are both count-location models with this feature. We show that the truncated Poisson model can also be expressed as a stationary renewal model, while the Sturt model cannot. More generally, we show that there is only one family of count-location models for the chiasma process that can also be expressed as stationary renewal models. The models in this family can exhibit either positive or negative interference.

BECAUSE of its mathematical simplicity, Haldane's Poisson process model (Haldane 1919) for the occurrence of crossovers along chromosomes at meiosis is generally assumed when calculating probabilities on multiple linked genetic loci. However, there is increasing interest in investigating alternative models. Several authors (*e.g.*, Foss *et al.* 1993; McPeek and Speed 1995; Zhao *et al.* 1995b) have compared the fit of various models to meiosis and tetrad data. Browning (1999) presents a method for comparing the fit of models to identity-by-state data from pairs of related individuals. The approach in Browning (1999) can also be used to incorporate any model for the crossover process into relationship analysis based on identity-by-state data from two individuals of uncertain relationship.

Regardless of the crossing-over model, the genetic distance in Morgans between two loci on a chromosome is defined to be the expected number of crossovers between them in a single meiosis; hence the rate of crossing over along a chromosome will be 1/M.

A map function $M(d)$ gives the probability of recombination (an odd number of crossovers) in an interval of genetic length d. Until recently, it was common to define map functions rather than to model the crossover point process directly. However, as Fisher (1947) pointed out, map functions do not uniquely determine multilocus recombination probabilities (the probabilities of patterns of recombination and nonrecombination between multiple loci) for more than three loci. The issue was complicated by the approach of Liberman and Karlin (1984) for extending map functions to multilocus recombination probabilities, which leads to inconsistencies for many of the map functions in use. A much more satisfactory approach is presented in Zhao and Speed (1996). Zhao and Speed (1996) show that stationary renewal processes give rise to most of the map functions found in the literature and characterize the class of map functions that may be achieved by stationary renewal process models.

Rather than define models for the crossover process directly, it is best to start with a model for the underlying chiasma process, since the formation of chiasmata is the physical event underlying crossing over. At meiosis, each chromosome duplicates to form two sister chromatids and the two pairs of homologous chromatids line up into a bundle of four. On this bundle, chiasmata occur, and each chiasma causes two nonsister chromatids to cross over. Figure 1 illustrates the process. Any one of the four resulting chromatids may be the one transmitted to the offspring. It is common to assume no chromatid interference (NCI), so that each chromatid is involved in a crossover with probability one-half at each chiasma independent of the outcome at other chiasmata on the bundle. This assumption seems to fit the available data (see Zhao *et al.* 1995a) and is convenient. A chiasma process model with a model for chromatid interference or the assumption of NCI defines a corresponding crossover process model.

The two main classes of chiasma process models that have been studied are stationary renewal models and count-location models (see Lange 1997). Count-location models (Karlin and Liberman 1978) define a distribution for the number, N, of chiasmata and assume that the locations of the chiasmata given N are independent. In this article, we characterize the relationship between count-location and stationary renewal models, showing that there is only one family of count-location models that can also be expressed as stationary renewal models. In particular, we examine the Sturt and trun-

Address for correspondence: Department of Statistics, 447 Blocker Bldg., Texas A&M University, College Station, TX 77843-3143.
E-mail: browning@stat.tamu.edu

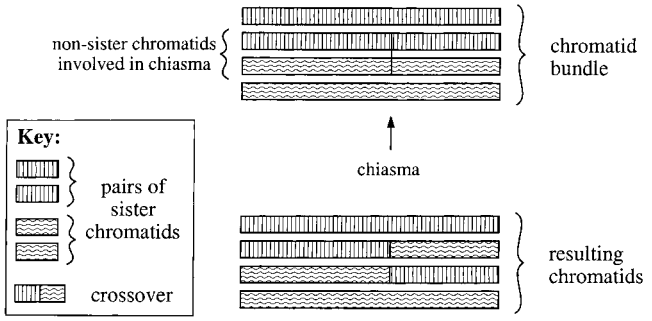

Figure 1.—The process of chiasma formation.

cated Poisson count-location models, showing that the truncated Poisson model is a stationary renewal model while the Sturt model is not.

Sturt and truncated Poisson models: The formation of at least one chiasma per chromatid bundle seems to be essential for correct disjunction of chromatids at meiosis. As a result, several chiasma process models have been proposed that ensure at least one chiasma per chromatid bundle. The Sturt (1976) and truncated Poisson (Lange 1997) models are both count-location models, with the locations of chiasmata distributed uniformly and independently along the chromosome given N, the number of chiasmata.

For a chromosome of genetic length L M, the Sturt model has the number N of chiasmata distributed as 1 + Poisson($2L - 1$)—that is, the model places one chiasma at a uniformly chosen random location on the chromatid bundle and then superimposes a Poisson chiasma process with mean $2L - 1$ so that the overall number of chiasmata has mean $2L$ as required. (Note that a genetic length of L M implies an average of L crossovers or $2L$ chiasmata per meiosis). The probability distribution of N under this model is, for $n \geq 1$,

$$P(N = n) = \frac{(2L - 1)^{n-1} e^{-(2L-1)}}{(n - 1)!}.$$

The truncated Poisson model has N, the number of chiasmata, following a Poisson distribution conditional on $N \geq 1$. The rate α of the distribution must be chosen such that the expected value of $N = 2L$, hence α solves $\alpha L/(1 - e^{-\alpha L}) = 2L$ and thus $\alpha/2 = 1 - e^{-\alpha L}$. For the truncated Poisson model, the probability distribution of N is, for $n \geq 1$,

$$P(N = n) = \frac{(\alpha L)^n e^{-\alpha L}}{n!(1 - e^{-\alpha L})}.$$

Note that under these two models, all chromosomes must have $L \geq 0.5$. The requirement of at least one chiasma per chromatid bundle guarantees that at least half of the resulting chromatids show at least one crossover. Thus the genetic length in morgans (the expected number of crossovers per meiosis) of a chromosome is at least 0.5.

RESULTS

In this section we show that the truncated Poisson model can be expressed as a stationary renewal model while the Sturt model cannot. Moreover, we show that there is only one family of count-location chiasma process models that can be expressed as stationary renewal models, and we characterize interference for this family.

Theorem 1. *Count-location chiasma process models with probabilities of the following form can be expressed as stationary renewal models:*

$$p_n = (1 - p_0) \frac{\beta^n L^n e^{-\beta L}}{n!(1 - e^{-\beta L})} \quad \text{for } n \geq 1, \tag{1}$$

where $p_n = P(N = n)$ *and* $\beta > 0$ *solves*

$$(1 - p_0)\beta/2 = 1 - e^{-\beta L}. \tag{2}$$

Conversely, count-location models can be expressed as renewal models only if they have probabilities of this form.

In this family of models, the probability, p_0, of no chiasmata can take any value between zero and one. But by definition of genetic length (genetic length, in morgans, equals expected number of crossovers per meiosis, which equals half the expected number of chiasmata on the chromatid bundle), $L = E(N)/2$, so L and p_0 must satisfy $L \geq (1 - p_0)/2$ since $E(N) \geq P(N \geq 1) = 1 - p_0$.

The models are mixtures of scaled truncated Poisson and point mass at zero. Let X_1 be the genetic distance to the first (leftmost) chiasma (if there is no chiasma on the chromosome, we consider $X_1 > L$), and let X_2 be the genetic distance between the first chiasma and the second ($X_2 > L - X_1$ if there is only one chiasma). Write $f_{X_1}(x)$ for the probability density of X_1, and $F_{X_2}(x) = P(X_2 \leq x)$ for the cumulative distribution function of X_2. On $0 < x < L$, $f_{X_1}(x) = 2e^{-\beta x}$ for these models, and $F_{X_2}(x) = 1 - e^{-\beta x}$. Thus $f_{X_1}(x) = 2(1 - F_{X_2}(x))$, as expected for a stationary renewal process of rate 2. The renewal distribution on $x \geq L$ is in general not uniquely determined by the count-location model.

Proof of Theorem 1 can be found in the appendix.

Corollary 1. *It follows immediately that the truncated Poisson model can be expressed as a stationary renewal model and that the truncated Poisson model is the only count-location with $P(N = 0) = 0$ that can be expressed as a stationary renewal model. In the* appendix *we show that the renewal distribution is uniquely determined on $x \geq L$ for this model and has $F_{X_2}(x) = 1$ for $x \geq L$.*

Note that $\int_0^L f_{X_1}(x)\,dx = 1$ so that f_{X_1} is a proper density with support on $(0,L)$, which ensures at least one chiasma on the chromosome. Also, the distribution of X_2 has mass of $e^{-\alpha L}$ at L. This mass does not affect the process on $[0,L]$ since we observe X_2 only in the interval $[X_1, L]$, with $X_1 > 0$.

Corollary 2. *The Sturt model cannot be expressed as a stationary renewal model.*

Characterization of interference: A traditional measure of chiasma interference is the coincidence coeffi-

cient, C, which is, for two disjoint intervals on a chromosome, the ratio of the probability of recombination in both intervals to the product of the marginal probabilities of recombination in each of the intervals. Equivalently,

$$C(I_1, I_2) = \frac{\frac{1}{4}P(N_{I_1} > 0 \text{ and } N_{I_2} > 0)}{\frac{1}{2}P(N_{I_1} > 0)\frac{1}{2}P(N_{I_2} > 0)}$$

(Risch and Lange 1979), where N_{I_1} and N_{I_2} are the numbers of chiasmata in the two intervals. If $C < 1$, we say that chiasma interference is positive, while if $C > 1$, we say that chiasma interference is negative. Risch and Lange (1979) give a partial characterization of interference for count-location models.

Theorem 2. *For the family of models given in Theorem 1, $C = \beta/2$ for any pair of disjoint intervals on a chromosome. Hence C increases monotonically as a function of p_0, with $C < 1$ for $p_0 < e^{-2L}$, and $C > 1$ for $p_0 > e^{-2L}$.*

Proof of Theorem 2 can be found in the appendix.

COMPARISON WITH THE RESULTS OF ZHAO AND SPEED

Zhao and Speed (1996) investigate properties of genetic mapping functions corresponding to stationary renewal chiasma processes. According to Theorem 2 of their article, the truncated Poisson model does not correspond to a stationary renewal model, which contradicts Corollary 1 of this article.

We reproduce Zhao and Speed's theorem and then resolve the discrepancy.

> For a map function M defined on $[0, L]$, where $L < \infty$, we say that M satisfies condition (B) if
> $M(0) = 0$ (B1), $M'(d) \geq 0$,
> for all d (B2), $M'(0) = 1$ (B3), $M''(d) \leq 0$,
> for all d (B4), $M'(L) = 0$ (B5), $M(L) = \frac{1}{2}$ (B6).
> We say that M satisfies condition (B)′ if it satifies (B1)–(B4) and
>
> $$M'(L) > 0 \text{ (B5)}', \quad M(L) < \tfrac{1}{2} \text{ (B6)}'.$$
>
> Theorem [Zhao and Speed (1996)]. Let M be the map function for a stationary renewal chiasma process satisfying NCI on a chromosome arm of finite length. Then M satisfies (B) or (B)′ for any L. Conversely, suppose that a function M from $[0, L]$ into $[0, \frac{1}{2}]$ satisfies (B) or (B)′. Then there is a stationary renewal chiasma process whose map function is M and whose renewal density is $-M''$ when $d \leq L$.

There is a difficulty with this theorem as M is only defined on $[0, L]$ and thus $M'(L)$ is not actually defined. Even if M can be extended beyond L (as will be the case when there is a stationary renewal chiasma process whose map function is M), the extension may not be differentiable at L. In particular, if M satisfies (B), then, assuming NCI, the only possible extension is $M(L) = \frac{1}{2}$ for $d \geq L$, and thus if $\lim_{d \nearrow L} M'(d) \neq 0$, then M is not differentiable at L.

If M corresponds to a stationary renewal model, this nondifferentiability at L arises when the interevent distribution has mass at L, as is the case for the renewal representation of the truncated Poisson model. The map function for the truncated Poisson model is

$$M(d) = \frac{1}{2}\left(1 - \sum_{n=0}^{\infty} p_n(1 - d/L)^n\right) = \frac{1}{2}\left(1 - \frac{e^{-\alpha d} - e^{-\alpha L}}{1 - e^{-\alpha L}}\right)$$

(see Lange 1997). It has $M(L) = \frac{1}{2}$ and $\lim_{d \nearrow L} M'(d) = \frac{1}{2}\alpha e^{-\alpha L}/(1 - e^{-\alpha L}) = e^{-\alpha L} > 0$, and hence M does not satisfy (B) or (B)′.

Thus, if one is willing to consider only stationary renewal processes for which the interevent distribution is absolutely continuous on $[0,\infty)$ [and thus has a density on $[0, \infty)$, as the statement of the theorem implicitly assumes], the theorem is correct, although it would be best to replace $M'(L)$ by $\lim_{d \nearrow L} M'(d)$. However, while it is, for biological plausibility and computational simplicity, reasonable to require the interevent distribution to have a density on $[0,L)$, it seems unnecessary to require the interevent distribution to be absolutely continuous on $[L,\infty)$ as this part of the distribution has no relevance or practical application for chromosomes of length L.

DISCUSSION

In some cases, the crossover process model has a simple form and is as easy to work with as the chiasma process model, while in other cases, it is simplest to work directly with the chiasma process. In Monte Carlo methods that sample possible realizations of either the chiasma process or the crossover process underlying the observed data (such as in Browning 1999), Monte Carlo error is reduced by sampling the crossover process rather than the chiasma process. Assuming NCI, the crossover process corresponding to a count-location chiasma process is also a count-location process. Let M be the number of crossovers on a chromosome. For the truncated Poisson,

$$P(M = 0) = \frac{e^{-\alpha L/2}}{1 + e^{-\alpha L/2}}$$

and

$$P(M = m) = \frac{e^{-\alpha L/2}(\alpha L/2)^m}{m!(1 - e^{-\alpha L})} \quad \text{for } m \geq 1.$$

Derivation of this result, along with corresponding probabilities for the Sturt model, can be found in Browning (1999). By thinning, the truncated Poisson crossover process is a stationary renewal process and has interevent density

$$g(x) = \frac{\alpha}{2}e^{-\alpha x/2} \quad \text{for } x < L.$$

The truncated Poisson and Sturt models, along with

other count-location models that require at least one chiasma per chromosome, are useful models because of the biological requirement that they satisfy. As renewal models have received more attention in the recent statistical genetics literature than count-location models, and thus methods for working with chiasma process models may be described only for renewal models, it is helpful to be able to express the truncated Poisson model as a renewal model and to be aware that other models, such as the Sturt model, cannot be expressed in this way.

Count-location models do have a severe drawback, in that they cannot incorporate chiasma interference in any meaningful way since the locations of the chiasmata are assumed to be independent given the number of chiasmata. An ideal model would incorporate the requirement of at least one chiasma, along with a pattern of chiasma interference that fits the available data, such as is found in the Kosambi map function (Kosambi 1944) and in the chi-square model (for example, see Zhao *et al.* 1995b). Goldgar and Fain (1988) present one approach to constructing such a model. The probability distribution for the number of crossovers is modeled as for a count-location model, but the locations of the crossovers are not assumed to be independent but to follow a distribution that allows for crossover interference. The model does seem to fit data well, but has several drawbacks, which are discussed in Zhao *et al.* (1995b).

I am grateful to Elizabeth Thompson for many helpful discussions and to Terry Speed, Hongyu Zhao, and the referees for their comments. This work was supported in part by the Burroughs Wellcome Fund through a Program in Mathematics and Molecular Biology training fellowship.

LITERATURE CITED

Browning, S., 1999 *Monte Carlo Likelihood Calculation for Identity by Descent Data.* Ph.D. thesis, University of Washington, Seattle.

Fisher, R. A., 1947 The theory of linkage in polysomic inheritance. Philos. Trans. R. Soc. Lond. Ser. B **233:** 55–87.

Foss, E., R. Lande, F. W. Stahl and C. M. Steinberg, 1993 Chiasma interference as a function of genetic distance. Genetics **133:** 681–691.

Goldgar, D. E., and P. R. Fain, 1988 Models of multilocus recombination: nonrandomness is chiasma number and crossover positions. Am. J. Hum. Genet. **43:** 38–45.

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.

Karlin, S., and U. Liberman, 1978 Classifications and comparisons of multilocus recombination distributions. Proc. Natl. Acad. Sci. USA **75:** 6332–6336.

Kosambi, D. D., 1944 The estimation of map distances from recombination values. Ann. Eugen. **12:** 172–175.

Lange, K., 1997 Models of recombination, chapter 12, pp. 206–227 in *Mathematical and Statistical Methods for Genetic Analysis.* Springer, New York.

Liberman, U., and S. Karlin, 1984 Theoretical models of genetic map functions. Theor. Popul. Biol. **25:** 331–346.

McPeek, M. S., and T. P. Speed, 1995 Modeling interference in genetic recombination. Genetics **139:** 1031–1044.

Risch, N., and K. Lange, 1979 An alternative model of recombination and interference. Ann. Hum. Genet. **43:** 61–70.

Sturt, E., 1976 A mapping function for human chromosomes. Ann. Hum. Genet. **40:** 147–163.

Zhao, H., and T. P. Speed, 1996 On genetic map functions. Genetics **142:** 1369–1377.

Zhao, H., M. S. McPeek and T. P. Speed, 1995a Statistical analysis of chromatid interference. Genetics **139:** 1057–1065.

Zhao, H., T. P. Speed and M. S. McPeek, 1995b Statistical analysis of crossover interference using the chi-square model. Genetics **139:** 1045–1056.

Communicating editor: S. Tavaré

APPENDIX

Chiasma process probabilities for count-location models: Let p_n, X_1, and X_2 be as in Theorem 1. For count-location models, $P(X_1 \le x|N = n) = 1 - (1 - x/L)^n$ (for $x < L$), so

$$\begin{aligned} P(X_1 \le x) &= \sum_{n=0}^{\infty} P(N = n)P(X_1 \le x|N = n) \\ &= \sum_{n=0}^{\infty} p_n\left[1 - \left(1 - \frac{x}{L}\right)^n\right] \\ &= 1 - \sum_{n=0}^{\infty} p_n\left(1 - \frac{x}{L}\right)^n \end{aligned}$$

and the density for X_1 is

$$f_{X_1}(x) = \sum_{n=1}^{\infty} \frac{np_n}{L}\left(1 - \frac{x}{L}\right)^{n-1}. \tag{3}$$

Let $f_{X_1|N}(x|n)$ be the probability density of X_1 at x given that $N = n$. Then

$$\begin{aligned} P(N = n|X_1 = x) &= f_{X_1|N}(x|n)P(N = n)/f_{X_1}(x) \\ &= \frac{n}{L}\left(1 - \frac{x}{L}\right)^{n-1} p_n \Big/ \sum_{m=1}^{\infty} \frac{mp_m}{L}\left(1 - \frac{x}{L}\right)^{m-1} \end{aligned}$$

and for $x < L - y$,

$$\begin{aligned} P(X_2 \le x|X_1 = y) &= \sum_{n=0}^{\infty} P(N = n|X_1 = y)P(X_2 \le x|X_1 = y, N = n) \\ &= \sum_{n=1}^{\infty}\left(\frac{np_n(1 - y/L)^{n-1}}{\sum_{m=1}^{\infty} mp_m(1 - y/L)^{m-1}}\right)\left(1 - \left(1 - \frac{x}{L - y}\right)^{n-1}\right) \\ &= 1 - \frac{\sum_{n=1}^{\infty} np_n((L - y)/L)^{n-1}((L - y - x)/(L - y))^{n-1}}{\sum_{n=1}^{\infty} np_n(1 - y/L)^{n-1}} \\ &= 1 - \frac{\sum_{n=1}^{\infty} np_n(1 - (x + y)/L)^{n-1}}{\sum_{n=1}^{\infty} np_n(1 - y/L)^{n-1}}. \end{aligned} \tag{4}$$

For stationary renewal models, X_2 is independent of X_1. Hence if the count-location model can be expressed as a stationary renewal model, then the final expression in Equation 4 must simplify so that it does not depend on y.

Proof of Theorem 1. From Equation 3, for count-location models of the form in Equation 1, the density of X_1 on $0 < x < L$ is

$$f_{X_1}(x) = \sum_{n=1}^{\infty}(1-p_0)\frac{ne^{-\beta L}(\beta L)^n}{Ln!(1-e^{-\beta L})}\left(1-\frac{x}{L}\right)^{n-1}$$
$$= (1-p_0)\frac{\beta e^{-\beta L}}{1-e^{-\beta L}}\sum_{n=1}^{\infty}\frac{\beta^{n-1}(L-x)^{n-1}}{(n-1)!}$$
$$= (1-p_0)\frac{\beta e^{-\beta L}}{1-e^{-\beta L}}e^{\beta(L-x)}$$
$$= 2e^{-\beta x} \quad \text{since } (1-p_0)\beta/2 = 1-e^{-\beta L}.$$

From Equation 4,

$$P(X_2 \le x | X_1 = y) = 1 - \frac{\Sigma_{n=1}^{\infty}(\beta L)^n(1-(x+y)/L)^{n-1}/(n-1)!}{\Sigma_{n=1}^{\infty}(\beta L)^n(1-y/L)^{n-1}/(n-1)!}$$
$$= 1-\exp\left\{\beta L\left(1-\frac{x+y}{L}\right)-\beta L\left(1-\frac{y}{L}\right)\right\}$$
$$= 1-e^{-\beta x}.$$

Since $P(X_2 \le x|X_1 = y)$ does not depend on y, we have $F_{X_2}(x) = 1-e^{-\beta x}$.

We can check that the probability distribution of the distance X_{i+1} between the ith and $(i+1)$st chiasmata $(i \ge 2)$ is the same as that for X_2, and does not depend on $\{X_j : j \le i\}$, by induction. Suppose $X_1, X_2, \ldots, X_i$ are independent, with probability densities $f_{Xj}(x) = \beta e^{-\beta x}$ for $2 \le j \le i$ and $f_{X_1}(x) = 2e^{-\beta x}$. Then

$$f(X_1 = x_1, \ldots, X_i = x_i | N = n)$$
$$= f(X_1 = x_1|N=n)\, f(X_2 = x_2|N=n, X_1 = x_1)$$
$$\ldots f(X_i = x_i|N=n, X_1 = x_1, \ldots, X_{i-1} = x_{i-1})$$
$$= \frac{n}{L}\left(1-\frac{x_1}{L}\right)^{n-1}\frac{n-1}{L-x_1}\left(1-\frac{x_2}{L-x_1}\right)^{n-2}$$
$$\ldots\frac{n-i+1}{L-\Sigma_{j=1}^{i-1}x_j}\left(1-\frac{x_i}{L-\Sigma_{j=1}^{i-1}x_j}\right)^{n-i}$$
$$= \frac{n!(L-\Sigma_{j=1}^{i}x_j)^{n-i}}{(n-i)!L^n}$$

and, writing X_i for $\{X_j : j \le i\}$,

$$P(X_{i+1} \le x_{i+1}|X_1 = x_1, \ldots, X_i = x_i)$$
$$= \sum_{n=i}^{\infty}P(X_{i+1} \le x_{i+1}|N=n, X_i = x_i)P(N=n|X_i = x_i)$$
$$= \sum_{n=i}^{\infty}P(X_{i+1} \le x_{i+1}|N=n, X_i = x_i)$$
$$\times\frac{f(X_i = x_i|N=n)P(N=n)}{f(X_i = x_i)}$$
$$= \sum_{n=i}^{\infty}\left(1-\left(1-\frac{x_{i+1}}{L-\Sigma_{j=1}^{i}x_j}\right)^{n-i}\right)\frac{n!(L-\Sigma_{j=1}^{i}x_j)^{n-i}}{(n-i)!L^n}$$
$$\times(1-p_0)\frac{(\beta L)^n e^{-\beta L}}{n!(1-e^{-\beta L})}\Big/(2\beta^{i-1}e^{-\beta\Sigma_{j=1}^{i}x_i})$$
$$= e^{-\beta(L-\Sigma_{j=1}^{i}x_j)}\sum_{n=i}^{\infty}\left(\frac{\beta^{n-i}(L-\Sigma_{j=1}^{i}x_j)^{n-i}}{(n-i)!}-\frac{\beta^{n-i}(L-\Sigma_{j=1}^{i+1}x_j)^{n-i}}{(n-i)!}\right)$$
$$= e^{-\beta(L-\Sigma_{j=1}^{i}x_j)}(e^{\beta(L-\Sigma_{j=1}^{i}x_j)}-e^{\beta(L-\Sigma_{j=1}^{i+1}x_j)})$$
$$= 1-e^{-\beta x_{i+1}}$$

so that X_{i+1} is independent of $\{X_j : j \le i\}$, and the density of X_{i+1} is $f_{X_{i+1}}(x) = \beta e^{-\beta x}$ as required. Hence the stationary renewal model with interevent distribution $F_{X_2}(x) = 1-e^{-\beta x}$ for $x < L$ is a representation of the count-location model in Equation 1.

We now show that count-location models can be expressed as renewal models only if they have probabilities following the form in Equation 1. We do so by showing that for any given value of p_1, there can be at most one count-location model that can be expressed as a renewal model—and hence for any given value of p_0 the same is true.

For a stationary renewal chiasma model,

$$f_{X_1}(x) = 2(1-F_{X_2}(x)). \tag{5}$$

Also, for a stationary renewal process, X_2 is independent of X_1, so that $F_{X_2}(x) = P(X_2 \le x|X_1 = y)$ (and this expression does not depend on y). Hence for a count-location chiasma model that can be expressed as a stationary model, Equations 3, 4, and 5 imply that

$$\sum_{n=1}^{\infty}\frac{np_n}{L}\left(1-\frac{x}{L}\right)^{n-1} = \frac{2\Sigma_{n=1}^{\infty}np_n(1-(x+y)/L)^{n-1}}{\Sigma_{n=1}^{\infty}np_n(1-y/L)^{n-1}}$$

and hence

$$\left(\sum_{n=1}^{\infty}np_n\left(1-\frac{x}{L}\right)^{n-1}\right)\left(\sum_{n=1}^{\infty}np_n\left(1-\frac{y}{L}\right)^{n-1}\right) = 2L\sum_{n=1}^{\infty}np_n\left(1-\frac{x+y}{L}\right)^{n-1}. \tag{6}$$

Write

$$\phi(s) = \sum_{n=1}^{\infty}np_n(1-s)^{n-1}. \tag{7}$$

Then Equation 6 can be written

$$\phi\left(\frac{x}{L}\right)\phi\left(\frac{y}{L}\right) = 2L\phi\left(\frac{x+y}{L}\right)$$

and this holds for all $\{(x,y) : x \ge 0, y \ge 0, x+y < L\}$. Hence,

$$\phi(s)\phi(t) = 2L\phi(s+t) \tag{8}$$

for all $\{(s,t) : s \ge 0, t \ge 0, s+t < 1\}$. Now $\phi(0) = E(N)$ and hence $\phi(0) = 2L$ by the definition of genetic distance. Also $\lim_{s\nearrow 1}\phi(s) = p_1$.

Suppose p_1 is set at some value, while the $\{p_n, n \ne 1\}$ can take any values. Then we can show that ϕ is uniquely determined by Equation 8. To show this by induction, suppose that for some n the value of $\phi(i/2^n)$ is determined for $i = 0, 1, \ldots, 2^n$ (for fixed L and p_1 this holds for $n = 0$). Then trivially $\phi(0/2^{n+1}) = \phi(0/2^n)$, which is determined. From Equation 8, $\phi(1/2^{n+1})\phi(1/2^{n+1}) = 2L\phi(1/2^n)$; thus $\phi(1/2^{n+1}) = \sqrt{2L\phi(1/2^n)}$ is determined. Further, applying Equation 8 multiple times, [$\phi(1/$

$2^{n+1})]^i = (2L)^{i-1}\phi(i/2^{n+1})$ and hence $\phi(i/2^{n+1}) = (1/2L)^{i-1}[\phi(1/2^{n+1})]^i$ is determined for $i = 2, 3, \ldots, 2^{n+1}$. Thus by induction, $\phi(i/2^n)$ is determined for all n and $i = 0, 1, \ldots, 2^n$. By continuity of ϕ (see the definition of ϕ in Equation 7), $\phi(s)$ is determined for all $0 \leq s < 1$.

Note that $p_n = \lim_{s\nearrow 1}\phi^{(n-1)}(s)/n!$, where $\phi^{(i)}(s)$ is the ith derivative of ϕ at s, for $n \geq 1$ and $p_0 = 1 - \Sigma_{n=1}^{\infty}p_n$. Thus for a given value of p_1, Equation 5 uniquely determines p_n for all $n \neq 1$. Equivalently, given a value of p_0, Equation 5 uniquely determines p_n for all $n \neq 0$.

Hence we have demonstrated that, for a given value of p_0, there is only one count-location model that can be expressed as a renewal model. In fact, for a given value of p_0, the unique count-location model that can be expressed as a renewal model has the form given in Equation 1.

Proof of extension in Corollary 1: We show that for the truncated Poisson model, F_{X_2} has a unique extension to $[0, \infty)$. First, we require $\lim_{x\to\infty}F_{X_2}(x) = 1$. We have $\lim_{x\nearrow L}F_{X_2}(x) = 1 - e^{-\alpha L}$, thus we have probability $e^{-\alpha L}$ that X_2 takes the value L or greater. Second, we require $E(X_2) = \frac{1}{2}$ since chiasmata occur at rate 2/M. Let $1_{\{X_2<L\}}$ equal one if $X_2 < L$ and zero otherwise.

$$\begin{aligned}E(X_2 1_{\{X_2<L\}}) &= \int_0^L x f_{X_2}(x)dx\\ &= \int_0^L x\alpha e^{-\alpha x}dx\\ &= \frac{1 - e^{-\alpha L}}{\alpha} - Le^{-\alpha L}\\ &= \frac{1}{2} - Le^{-\alpha L}.\end{aligned}$$

Thus $E(X_2 1_{\{X_2\geq L\}}) = E(X_2) - E(X_2 1_{\{X_2<L\}})$ should equal $Le^{-\alpha L}$. The two conditions are met by giving the distribution of X_2 mass $e^{-\alpha L}$ at L.

Proof of Theorem 2. Let I be an interval, or union of two disjoint intervals, with total genetic length d on a chromosome of length L. Let N_I be the number of chiasmata in I. Then, for the models in Equation 1,

$$\begin{aligned}P(N_I = 0) &= \sum_{n=0}^{\infty}p_n\left(1 - \frac{d}{L}\right)^n\\ &= p_0 + \sum_{n=1}^{\infty}(1 - p_0)\frac{\beta^n L^n(1 - d/L)^n e^{-\beta L}}{n!(1 - e^{-\beta L})}\\ &= p_0 - \frac{(1 - p_0)e^{-\beta L}}{1 - e^{-\beta L}} + \sum_{n=0}^{\infty}(1 - p_0)\frac{\beta^n L^n(1 - d/L)^n e^{-\beta L}}{n!(1 - e^{-\beta L})}\\ &= 1 - \frac{1 - p_0}{1 - e^{-\beta L}} + \frac{(1 - p_0)e^{-\beta d}}{1 - e^{-\beta L}}\\ &= 1 - \frac{(1 - p_0)(1 - e^{-\beta d})}{1 - e^{-\beta L}}\\ &= 1 - (2/\beta)(1 - e^{-\beta d}),\end{aligned}$$

substituting Equation 1 for p_n in the second line and applying condition (2) in the last line.

Hence, for two disjoint intervals with genetic lengths d_1 and d_2,

$$\begin{aligned}C &= \frac{P(N_{I_1} > 0 \text{ and } N_{I_2} > 0)}{P(N_{I_1} > 0)P(N_{I_2} > 0)}\\ &= \frac{1 - P(N_{I_1} = 0) - P(N_{I_2} = 0) + P(N_{I_1\cup I_2} = 0)}{[1 - P(N_{I_1} = 0)][1 - P(N_{I_2} = 0)]}\\ &= \frac{(2/\beta)(1 + e^{-\beta(d_1+d_2)} - e^{-\beta d_1} - e^{-\beta d_2})}{(2/\beta)^2(1 - e^{-\beta d_1})(1 - e^{-\beta d_2})}\\ &= \beta/2.\end{aligned}$$

Now, condition (2) gives us $p_0 = 1 - 2(1 - e^{-\beta L})/\beta$. We can check that p_0 increases monotonically with β (for $\beta > 0$) and hence that C increases monotonically with p_0. Let $g(\beta) = 1 - 2(1 - e^{-\beta L})/\beta$. The derivative of g is $g'(\beta) = 2(e^{\beta L} - (1 + \beta L))/(\beta^2 e^{\beta L}) > 0$ for $\beta > 0$, since $e^{\beta L} = \Sigma_{i=0}^{\infty}(\beta L)^i/i! > 1 + \beta L$, and hence $p_0 = g(\beta)$ is an increasing function of β for $\beta > 0$. The no interference model has $p_0 = e^{-2L}$. Thus $C < 1$ for $p_0 < e^{-2L}$ ($C \to 0$ as $p_0 \to 1 - 2L$) and $C > 1$ for $p_0 > e^{-2L}$ ($C \to \infty$ as $p_0 \to 1$).