Published in final edited form as:
Commun Stat Theory Methods. 2014 ; 43(1): 44–71. doi:10.1080/03610926.2012.655876.

NON-HOMOGENEOUS POISSON PROCESS MODEL FOR GENETIC CROSSOVER INTERFERENCE

Szu-Yun Leu[1] and **Pranab K. Sen**[2]

[1]Department of Pediatrics, Institute for Clinical and Translational Science, 1115 Hewitt Hall, Zot 1385, University of California, Irvine, Irvine, California 92697, sleu@uci.edu

[2]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Abstract

The genetic crossover interference is usually modeled with a stationary renewal process to construct the genetic map. We propose two non-homogeneous, also dependent, Poisson process models applied to the known physical map. The crossover process is assumed to start from an origin and to occur sequentially along the chromosome. The increment rate depends on the position of the markers and the number of crossover events occurring between the origin and the markers. We show how to obtain parameter estimates for the process and use simulation studies and real *Drosophila* data to examine the performance of the proposed models.

Keywords

Non-homogeneous; Poisson; Crossover; Interference

1. INTRODUCTION

Reproduction is the basis of heredity. In order for offspring to preserve genetic information from both parents and still keep the same number of chromosomes, each parent will only pass, on average, half of their chromosomes to the offspring. The process for the diploid chromosome number reducing to the haploid state is called meiosis, which occurs during gamete formation in animals. During the meiosis process, each cell first replicates its chromosomes and the two sister chromatids joined together by the centromere with the ends of the chromosome named telomeres. The homologous chromosomes then line up and pair together to form tight bundles of four chromatids. At some locations along the chromosomes, one chromatid from each of the two homologous chromosomes contacts each other and crosses over for physical materials to separate, exchange, and recombine. The locations of crossover or recombination are called chiasmata (sing.=chiasma). The replicated cell is then followed by two separate cell divisions and thus each diploid cell results in four haploid cells. Eventually, every haploid reproductive cell may contain different combinations of genes from the parents. (Snustad, 1992)

The actual crossover events cannot be seen, but there are two observable outcomes, the formation of chiasmata and the recombination between genes on the opposite sides of the crossover point. Chiasmata can be observed cytologically, but there are technical difficulties. The recombination events can only be observed at the next generation when the

Correspondence to: Szu-Yun Leu.

genes on the recombinant chromosomes are expressed. If two genes on two loci of the same chromosome are from different paternal origin, we can confirm that an odd number of crossovers have occurred between the two loci and that they are recombinants.

Interference is a phenomenon where the formation of crossover events in adjacent regions are non-independent. The actual mechanism is not yet fully understood but it is a cause of the inconsistency between genetic distance and physical distance. Genetic distance of 1 centiMorgan (cM) or 0.01 Morgan (M) is about the distance with average 1 percent recombinants. Physical distance is the distance of nucleotide bases, in base pairs (bp), kilobases pairs (kb, 1000 bp), or megabases pairs (Mb, 1 million bp). There are two types of interference: crossover interference, in which the number and location of crossovers in a given region depend on the numbers and locations of crossovers in disjoint regions; and chromatid interference, in which each pair of non-sister chromatids is not equally likely to be involved in a crossover and depends on which were involved in other crossovers (McPeek and Speed, 1995). Crossover interference has been observed in almost all organisms, while no consistent evidence of chromatid interference has been found (Zhao et al., 1995b). Since inferences about chromatid interference require data from all four products of meiosis (tetrad data), which are not available in most species, the assumption of no chromatid interference (NCI) is usually made (McPeek and Speed, 1995; Zhao et al., 1995a; Zhao and Speed, 1996). Thus NCI is also assumed throughout this paper.

By assuming NCI with the four-strand crossover process, Mather's (1938) formula below can be applied to construct the relationship between the unobservable four-strand crossover probability and the observable single-strand recombination probability. Let $N(a, b)$ be the random number of chiasmata occurring in interval $[a, b]$ of the four-strand chromatid bundle, and $R_h(a, b)$, $h = 1, 2, 3, 4$, represent the index binary variable of whether or not a and b are recombinants for the h-th chromatid, Mather showed that

$$Pr\{R_h(a,b)=1\}=\frac{1}{2}Pr\{N(a,b)>0\}=\frac{1}{2}[1-Pr\{N(a,b)=0\}].$$

The formula can be extended for multiple markers with disjoint genomic regions I_j, $j = 1, 2, \ldots, k$.

$$Pr\{E_1 \text{and} E_2 \text{and} \cdots E_k\}=\frac{1}{2^k}Pr\{F_1 \text{and} F_2 \text{and} \cdots F_k\}, \quad (1)$$

where E_j is the event of recombination in I_j (requiring an odd number of crossover events) for a single-strand meiotic product, and F_j is the event of at least one crossover occurs within I_j for the four-strand chromatid bundle. The boundary for the probability of recombination occurring in all k intervals is $\frac{1}{2^k}$. When chromatid interference is allowed, the boundary may go up to 1. (Karlin and Liberman, 1983; Risch and Lange, 1983)

The unobservable crossover process has usually been modeled with a stationary renewal process so the genetic distance between markers can be estimated. Available models include homogeneous Poisson model (Haldane, 1919), chi-square model (Fisher et al. 1947; Owen, 1949; Zhao et al. 1995a) and gamma model (Broman and Weber, 2000). The homogeneous Poisson model does not allow for interference, while the chi-square model and the gamma model allow for interference. However, it has always been known that the genetic distance and the actual physical distance between markers do not match. Now, with molecular biology advances, the physical positions of many genes along the same chromosome in some organisms have been confirmed. It is more possible for us to study the crossover

formation process directly using the physical map; and a counting process is a good start for this purpose.

In a Poisson process, there are generally three elements, the distance or time t, the number of events n, and the increment rate λ. When λ does not depend on t or n, the process has homogeneous and independent increments. When λ depends on t, the process becomes non-homogeneous, and when λ depends on n, the process becomes dependent. The λ in our proposed models actually depends on both t and n, so the crossover process depends not only on the physical position of the markers, but also on the number of crossover events that have occurred.

In this paper, we propose two non-homogeneous and dependent Poisson process models with different rates of positive interference which limits the number of chiasmata. The crossover events are assumed to start from a point of origin, which can be the centromere or one of the telomeres, and consecutive events occur sequentially along the chromosome. The increment rate $\lambda(t, n)$ increases with the position of marker t, but decreases with the number of crossover events n occurring between the marker and the origin. Our models can be reduced to homogeneous models but are always dependent. Since the physical position starts from the telomere of the p arm (the short arm), for consistency and simplicity, we set the origin point to be the same telomere throughout this paper.

2. METHODS

2.1. MODEL SETUP

Suppose, including the origin point, there are $l + 1$ markers along the chromosome and these markers construct l intervals starting from the origin, labeled as $I_1, I_2, \ldots, I_l$. Let $t_j, j = 1, 2, \ldots, l$, be the physical distance (in bp) for interval I_j, and $n_j = N(t_j)$ be the number of crossovers occurred within I_j for the four-strand chromatid bundle. Then, $T_j = \sum_{r=1}^{j} t_r$ represents the physical position of the j-th marker on the chromosome, and $N(T_j) = \sum_{r=1}^{j} n_r$ represents the number of crossovers occurred before the j-th marker for the four-strand chromatid bundle. The position of the marker at the origin point T_0 is set to be 0, and also $N(T_0)$ is set to be 0. Note that T_l and $N(T_l)$ are both considered finite.

Let $\boldsymbol{q}_l = q(x_1, x_2, \ldots, x_l)$ be the joint probability of the unobservable crossover pattern $(x_1, x_2, \ldots, x_l)$ for the four-strand chromatid bundle and $\boldsymbol{p}_l = p(y_1, y_2, \ldots, y_l)$ be the joint probability of the observable recombination pattern $(y_1, y_2, \ldots, y_l)$ for a single-strand meiotic product in marker intervals $(I_1, I_2, \ldots, I_l)$, and for $j = 1, 2, \ldots, l$,

$$x_j = \begin{cases} 1, & \text{if at least one crossover occurs in} I_j, \\ 0, & \text{if no crossover occurs in} I_j, \end{cases}$$

$$y_j = \begin{cases} 1, & \text{if recombination(odd number of crossover)occurs in} I_j, \\ 0, & \text{if no recombination(even number of crossover)occurs in} I_j. \end{cases}$$

Then, for the $l + 1$ markers, the crossover probability $\boldsymbol{q}_l$ is defined as

$$\begin{aligned}\boldsymbol{q}_l=&[Pr\{N(t_1)=0\}]^{1-x_1}[Pr\{N(t_1)>0\}]^{x_1}\\&\times\prod_{j=2}^{l}[Pr\{N(t_j)=0|N(T_{j-1})=0\}]^{(1-x_j)\prod_{r=1}^{j-1}(1-x_r)}\\&\times[Pr\{N(t_j)=0|N(T_{j-1})>0\}]^{(1-x_j)\left[1-\prod_{r=1}^{j-1}(1-x_r)\right]}\\&\times[Pr\{N(t_j)>0|N(T_{j-1})=0\}]^{x_j\prod_{r=1}^{j-1}(1-x_r)}\\&\times[Pr\{N(t_j)>0|N(T_{j-1})>0\}]^{x_j\left[1-\prod_{r=1}^{j-1}(1-x_r)\right]}\\=&\sum_{\substack{n_i=1\\\forall i=1,\ldots,l:x_i=1}}^{\infty}\left\{\prod_{j=1}^{l}[Pr\{N(t_j)=0|N(T_{j-1})=0\}]^{(1-x_j)c_j}\times[Pr\{N(t_j)=0|N(T_{j-1})=d_j\}]^{(1-x_j)(1-c_j)}\times[Pr\{N(t_j)=n_j|N(\right.\end{aligned}\tag{2}$$

where $x_0 = 0$, $c_j=\prod_{r=0}^{j-1}(1-x_r)$, and $d_j=\sum_{r=0}^{j-1}x_r n_r$; and the marginal and conditional probabilities are derived from the non-homogeneous Poisson process models as proposed in Section 2.1.1 and 2.1.2.

By assuming NCI, the extended version of Mather's formula in equation (1) can be reformulated (Speed, 1996) and the recombination probability becomes

$$\boldsymbol{p}_l=\sum_{\substack{(x_1,\ldots,x_l):\\\forall x_r\geq y_r,r=1,\ldots,l}}\left(\frac{1}{2}\right)^{\sum_{r=1}^{l}x_r}\boldsymbol{q}_{l\cdot}\tag{3}$$

Taking 0, T_1 and T_2 as an example, the relationship between the recombination probability and the crossover probability can be expressed as

$$\begin{aligned}p_{0,T_1,T_2}(0,0)&=\frac{q_{0,T_1,T_2}(1,1)}{4}+\frac{q_{0,T_1,T_2}(1,0)+q_{0,T_1,T_2}(0,1)}{2}+q_{0,T_1,T_2}(0,0)=\frac{1+q_{0,T_1,T_2}(0)+q_{0,T_1,T_2}(1,0)}{4}+\frac{q_{0,T_1,T_2}(0,0)}{2}\\p_{0,T_1,T_2}(0,1)&=\frac{q_{0,T_1,T_2}(1,1)}{4}+\frac{q_{0,T_1,T_2}(0,1)}{2}=\frac{1+q_{0,T_1,T_2}(0)-q_{0,T_1,T_2}(1,0)}{4}-\frac{q_{0,T_1,T_2}(0,0)}{2}\\p_{0,T_1,T_2}(1,0)&=\frac{q_{0,T_1,T_2}(1,1)}{4}+\frac{q_{0,T_1,T_2}(1,0)}{2}=\frac{1-q_{0,T_1,T_2}(0)+q_{0,T_1,T_2}(1,0)}{4}\\p_{0,T_1,T_2}(1,1)&=\frac{q_{0,T_1,T_2}(1,1)}{4}=\frac{1-q_{0,T_1,T_2}(0)-q_{0,T_1,T_2}(1,0)}{4}\\p_{0,T_1,T_2}(0,.)&=\frac{q_{0,T_1,T_2}(1,.)}{2}+q_{0,T_1,T_2}(0,.)=\frac{1+q_{0,T_1,T_2}(0,.)}{2}\\p_{0,T_1,T_2}(1,.)&=\frac{q_{0,T_1,T_2}(1,.)}{2}=\frac{1-q_{0,T_1,T_2}(0,.)}{2}\\p_{0,T_1,T_2}(.,0)&=p_{0,T_1,T_2}(0,0)+p_{0,T_1,T_2}(1,0)=\frac{1+q_{0,T_1,T_2}(0,0)+q_{0,T_1,T_2}(1,0)}{2}\\p_{0,T_1,T_2}(.,1)&=p_{0,T_1,T_2}(0,1)+p_{0,T_1,T_2}(1,1)=\frac{1-q_{0,T_1,T_2}(0,0)-q_{0,T_1,T_2}(1,0)}{2}.\end{aligned}$$

2.1.1. MODEL I—The increment rate $\lambda_1(t, n)$, as defined below, of the first proposed model is a linear increasing function of t and a monotone decreasing function of n.

$$\lambda_1(t,n)=\frac{2\alpha t+\beta}{n+\mu},\quad \alpha,\beta\geq 0,\mu>0,\alpha+\beta>0,$$
$$\int\lambda_1(t,n)dt=\Lambda_1(t,n)=\frac{\alpha t^2+\beta t}{n+\mu},$$

Various increment rates are shown in Figure 1(a) and it can be seen that $\frac{\beta}{\mu}$ and $\frac{2\alpha}{\mu}$ are, respectively, the intercept and the slope of $\lambda_1(t, n)$ with respect to t when $n = 0$. With smaller μ, the reduction of $\lambda_1(t, n)$ from $n = 0$ to $n = 1$ is greater. The lower bound of $\lambda_1(t, n)$ is 0 and the upper bound is $\frac{2\alpha L+\beta}{\mu}$, where L represents the length of the chromosome in bp. When $\alpha = 0$, this model reduces to a homogeneous Poisson process model with $\lambda=\frac{\beta}{n+\mu}$.

By working through the Poisson process theory, the probability components of $\boldsymbol{q}_l$ in equation (2) are found to be

$$Pr\{N(t_j)=0|N(T_{j-1})=0\}=\exp\left[-\frac{\alpha A_{2j}+\beta A_{1j}}{\mu}\right]=G_{j0}(0)$$
$$Pr\{N(t_j)=0|N(T_{j-1})=d_j\}=\exp\left[-\frac{\alpha A_{2j}+\beta A_{1j}}{d_j+\mu}\right]=G_{j1}(0)$$
$$Pr\{N(t_j)=n_j|N(T_{j-1})=0\}=\sum_{k_j=0}^{n_j}\frac{(k_j+\mu)^{n_j-1}(n_j+\mu)}{(-1)^{n_j-k_j}k_j!(n_j-k_j)!}\exp\left[-\frac{\alpha A_{2j}+\beta A_{1j}}{k_j+\mu}\right]=\sum_{k_j=0}^{n_j}G_{j0}(n_j)$$
$$Pr\{N(t_j)=n_j|N(T_{j-1})=d_j\}=\sum_{k_j=0}^{n_j}\frac{(d_j+k_j+\mu)^{n_j-1}(d_j+n_j+\mu)}{(-1)^{n_j-k_j}k_j!(n_j-k_j)!}\exp\left[-\frac{\alpha A_{2j}+\beta A_{1j}}{d_j+k_j+\mu}\right]=\sum_{k_j=0}^{n_j}G_{j1}(n_j)$$

where $j = 1, 2, \ldots, l$, $A_{1j} = T_j - T_{j-1}$, and $A_{2j}=T_j^2 - T_{j-1}^2$. We can further obtain the first derivatives of $\boldsymbol{q}_l$ with respect to (α, β, μ) as below and the second derivatives as shown in the Supplement.

$$\frac{\partial \boldsymbol{q}_l}{\partial\alpha}=\sum_{\substack{n_i=1\\ \forall i=1,\ldots,l:x_i=1}}^{\infty}\sum_{h=1}^{l}\left\{-A_{2h}\times\left[\frac{G_{h0}(0)}{\mu}\right]^{(1-x_h)c_h}\left[\frac{G_{h1}(0)}{d_h+\mu}\right]^{(1-x_h)(1-c_h)}\left[\sum_{k_j=0}^{n_j}\frac{G_{h0}(n_h)}{k_h+\mu}\right]^{x_hc_h}\left[\sum_{k_j=0}^{n_j}\frac{G_{h1}(n_h)}{d_h+k_h+\mu}\right]^{x_h(1-c_h)}\times\prod_{\substack{j=1\\ j\neq h}}^{l}[\right.$$

$$\frac{\partial \boldsymbol{q}_l}{\partial\beta}=\sum_{\substack{n_i=1\\ \forall i=1,\ldots,l:x_i=1}}^{\infty}\sum_{h=1}^{l}\left\{-A_{1h}\times\left[\frac{G_{h0}(0)}{\mu}\right]^{(1-x_h)c_h}\left[\frac{G_{h1}(0)}{d_h+\mu}\right]^{(1-x_h)(1-c_h)}\left[\sum_{k_j=0}^{n_j}\frac{G_{h0}(n_h)}{k_h+\mu}\right]^{x_hc_h}\left[\sum_{k_j=0}^{n_j}\frac{G_{h1}(n_h)}{d_h+k_h+\mu}\right]^{x_h(1-c_h)}\times\right.$$

$$\frac{\partial \boldsymbol{q}_l}{\partial\mu}=\sum_{\substack{n_i=1\\ \forall i=1,\ldots,l:x_i=1}}^{\infty}\sum_{h=1}^{l}\left\{\left[\frac{\alpha A_{2j}+\beta A_{1j}}{\mu^2}G_{j0}(0)\right]^{(1-x_j)c_j}\left[\frac{\alpha A_{2j}+\beta A_{1j}}{(d_j+\mu)^2}G_{j1}(0)\right]^{(1-x_j)(1-c_j)}\times\left[\sum_{k_j=0}^{n_j}\left(\frac{1}{n_j+\mu}+\frac{n_j-1}{k_j+\mu}+\frac{\alpha A_{2j}}{(k_j-\right.\right.\right.$$

With $\boldsymbol{q}_l$ defined, $\boldsymbol{p}_l$ can then be obtained. In Figure 1(b), we display the expected recombination probability between various pairs of markers (T_1 and T_2) from the corresponding increment rates shown in Figure 1(a). We can see that the expected recombination probability between two markers increases with α and β but decreases with μ.

The rate also changes almost linearly with T_1 for the same distance between T_1 and T_2.

When the ratio $\frac{\beta}{\mu}$ is the same, as the two panels on the left and the two panels on the right, the distributions of the expected recombination probability of paired markers look similar. The distributions are more distinguishable when α gets larger (bottom panel vs. top panel). Also, on the left two panels, when β = 0, some of the recombination probabilities are too close to distinguish, while on the right panels they spread out.

2.1.2. MODEL II—For the second proposed model, the increment rate $\lambda_2(t, n)$ decreases with n at the same rate as in Model I but increases with t as a concave function when t is small and a convex function when t gets larger.

$$\begin{aligned}\lambda_2(t,n)&=\tfrac{\beta}{n+\mu}\left(1+\tfrac{\alpha}{t}\right)e^{-\frac{\alpha}{t}},\\ \Lambda_2(t,n)&=\tfrac{\beta t}{n+\mu}e^{-\frac{\alpha}{t}},\end{aligned}\quad \alpha\geq 0,\beta,\mu>0$$

Figure 2(a) shows that $\lambda_2(t, n)$ is upper bounded by $\frac{\beta}{\mu}$ and larger α slows down the rate of $\lambda_2(t, n)$ increasing with t. The effect of μ is the same as in Model I and this model reduces to the same homogeneous Poisson process model as in Model I when α = 0.

Following the Poisson process theory, the probabilities in $\boldsymbol{q}_l$ for the second model are found to be

$$Pr\{N(t_j)=0|N(T_{j-1})=0\}=\exp[-\frac{\beta B_{1j}}{\mu}]=H_{j0}(0)$$

$$Pr\{N(t_j)=0|N(T_{j-1})=d_j\}=\exp[-\frac{\beta B_{1j}}{d_j+\mu}]=H_{j1}(0)$$

$$Pr\{N(t_j)=n_j|N(T_{j-1})=0\}=\sum_{k_j=0}^{n_j}\frac{(k_j+\mu)^{n_j-1}(n_j+\mu)}{(-1)^{n_j-k_j}k_j!(n_j-k_j)!}\exp\left[-\frac{\beta B_{1j}}{k_j+\mu}\right]=\sum_{k_j=0}^{n_j}H_{j0}(n_j)$$

$$Pr\{N(t_j)=n_j|N(T_{j-1})=d_j\}=\sum_{k_j=0}^{n_j}\frac{(d_j+k_j+\mu)^{n_j-1}(d_j+n_j+\mu)}{(-1)^{n_j-k_j}k_j!(n_j-k_j)!}\exp\left[-\frac{\beta B_{1j}}{d_j+k_j+\mu}\right]=\sum_{k_j=0}^{n_j}H_{j1}(n_j)$$

where $j = 1, 2, \ldots, l$ and $B_{1j}=T_je^{-\frac{\alpha}{T_j}}-T_{j-1}e^{-\frac{\alpha}{T_{j-1}}}$. The first derivatives of $\boldsymbol{q}_l$ with respect to (α, β, μ) are shown below and the second derivatives can be found in the Supplement.

$$\frac{\partial \boldsymbol{q}_l}{\partial \alpha}=\sum_{\substack{n_i=1\\ \forall i=1,\ldots,l:x_i=1}}^{\infty}\sum_{h=1}^{l}\left\{\beta\left(e^{-\frac{\alpha}{T_h}}-e^{-\frac{\alpha}{T_{h-1}}}\right)\times\left[\frac{H_{j0}(0)}{\mu}\right]^{(1-x_j)c_j}\left[\frac{H_{j1}(0)}{d_j+\mu}\right]^{(1-x_j)(1-c_j)}\left[\sum_{k_j=0}^{n_j}\frac{H_{j0}(n_j)}{k_j+\mu}\right]^{x_jc_j}\left[\sum_{k_j=0}^{n_j}\frac{H_{j1}(n_j)}{d_j+k_j+\mu}\right]^{x_j}\right.$$

$$\frac{\partial \boldsymbol{q}_l}{\partial \beta}=\sum_{\substack{n_i=1\\ \forall i=1,\ldots,l:x_i=1}}^{\infty}\sum_{h=1}^{l}\left\{-\left(T_je^{-\frac{\alpha}{T_j}}-T_{j-1}e^{-\frac{\alpha}{T_{j-1}}}\right)\times\left[\frac{H_{j0}(0)}{\mu}\right]^{(1-x_j)c_j}\left[\frac{H_{j1}(0)}{d_j+\mu}\right]^{(1-x_j)(1-c_j)}\left[\sum_{k_j=0}^{n_j}\frac{H_{j0}(n_j)}{k_j+\mu}\right]^{x_jc_j}\left[\sum_{k_j=0}^{n_j}\frac{H_{j1}(n_j)}{d_j+k_j+\mu}\right]^{x_j}\right.$$

$$\frac{\partial \boldsymbol{q}_l}{\partial \mu}=\sum_{\substack{n_i=1\\ \forall i=1,\ldots,l:x_i=1}}^{\infty}\sum_{h=1}^{l}\left\{\left[\frac{\beta B_{1j}}{\mu^2}H_{j0}(0)\right]^{(1-x_j)c_j}\left[\frac{\beta B_{1j}}{(d_j+\mu)^2}H_{j1}(0)\right]^{(1-x_j)(1-c_j)}\times\left[\sum_{k_j=0}^{n_j}\left(\frac{1}{n_j+\mu}+\frac{n_j-1}{k_j+\mu}+\frac{\beta B_{1j}}{(k_j+\mu)^2}\right)H_{j0}(n\right.\right.$$

Figure 2(b) demonstrates the expected recombination probability between various pairs of markers (T_1 and T_2) from the corresponding increment rates in Figure 2(a). The figure shows that the expected recombination probability increases with β and decreases with α and μ; and it changes nonlinearly with T_1 for the same distance between T_1 and T_2 and the nonlinearity seems to be affected by α and $\frac{\beta}{\mu}$. Also like in Model I, the distributions of the expected recombination probabilities are very similar if they have the same α and the same $\frac{\beta}{\mu}$ (not shown).

2.2. PARAMETER ESTIMATION

2.2.1. MAXIMUM LIKELIHOOD ESTIMATOR—Let $(y_{i1}, \ldots, y_{il})$, $i = 1, \ldots, m$, be m i.i.d. random variables with p.d.f. $\boldsymbol{p}_l$ as defined in equation (3) and parameter vector $\boldsymbol{\theta} = (\alpha, \beta, \mu)'$. Then the likelihood function and the log-likelihood function are simply

$$L_m(\boldsymbol{\theta};\boldsymbol{y})=\prod_{i=1}^{m}\boldsymbol{p}_{il}\text{and}\ \ \ln L_m(\boldsymbol{\theta};\boldsymbol{y})=\sum_{i=1}^{m}\ln\boldsymbol{p}_{il}.$$

The score function $\boldsymbol{U}_m(\boldsymbol{\theta})$ and the information matrix $\boldsymbol{V}_m(\boldsymbol{\theta})$ are further defined as

$$\boldsymbol{U}_m(\boldsymbol{\theta})=\frac{\partial}{\partial\boldsymbol{\theta}}\ln L_m(\boldsymbol{\theta};\boldsymbol{y})=\sum_{i=1}^{m}\frac{\partial\ln\boldsymbol{p}_{il}}{\partial\boldsymbol{\theta}}=\sum_{i=1}^{m}\frac{1}{\boldsymbol{p}_{il}}\frac{\partial\boldsymbol{p}_{il}}{\partial\boldsymbol{\theta}}$$

$$\boldsymbol{V}_m(\boldsymbol{\theta})=\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\ln L_m(\boldsymbol{\theta};\boldsymbol{y})=\sum_{i=1}^{m}\frac{\partial^2\ln\boldsymbol{p}_{il}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}=\sum_{i=1}^{m}\left[\frac{1}{\boldsymbol{p}_{il}}\frac{\partial\boldsymbol{p}_{il}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}-\frac{1}{\boldsymbol{p}_{il}^2}\frac{\partial\boldsymbol{p}_{il}}{\partial\boldsymbol{\theta}}\frac{\partial\boldsymbol{p}_{il}}{\partial\boldsymbol{\theta}'}\right]$$

where

$$\frac{\partial\boldsymbol{p}_{il}}{\partial\boldsymbol{\theta}}=\sum_{\substack{(x_1,\ldots,x_l):\\ \forall x_r\geq y_r,r=1,\ldots,l}}\left(\frac{1}{2}\right)^{\sum_{r=1}^{l}x_r}\frac{\partial\boldsymbol{q}_{il}}{\partial\boldsymbol{\theta}}\ \ \frac{\partial^2\boldsymbol{p}_{il}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}=\sum_{\substack{(x_1,\ldots,x_l):\\ \forall x_r\geq y_r,r=1,\ldots,l}}\left(\frac{1}{2}\right)^{\sum_{r=1}^{l}x_r}\frac{\partial^2\boldsymbol{q}_{il}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}.$$

Let $\hat{\theta}_m = (\hat{\alpha}, \hat{\beta}, \hat{\mu})'$ be the MLE of θ. An explicit form of the MLE cannot be found by setting the score function $\boldsymbol{U}_m(\theta)|_{\theta=\hat{\theta}_m} = 0$. Hence, an iteration method needs to be applied to solve for MLE. Suppose $\boldsymbol{\theta}_m^{(0)}$ is an initial guess of θ that is close to the true MLE, and $\boldsymbol{\theta}_m^*$ lies between $\hat{\theta}_m$ and $\boldsymbol{\theta}_m^{(0)}$. According to the following Taylor expansion,

$$\boldsymbol{U}_m(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\hat{\theta}_m} = \boldsymbol{U}_m(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m^{(0)}} + \left(\hat{\theta}_m - \boldsymbol{\theta}_m^{(0)}\right)' \boldsymbol{V}_m(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m^*} = 0,$$

it is found that

$$\hat{\theta}_m = \boldsymbol{\theta}_m^{(0)} + \left(\boldsymbol{V}_m(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m^*}\right)^{-1} \boldsymbol{U}_m(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m^{(0)}}.$$

Based on the above equation, the Newton-Raphson method was further applied to the following repeated iteration,

$$\boldsymbol{\theta}_m^{(i)} = \boldsymbol{\theta}_m^{(i-1)} + \left(\boldsymbol{V}_m(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m^{(i-1)}}\right)^{-1} \boldsymbol{U}_m(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_m^{(i-1)}},$$

where $\boldsymbol{\theta}_m^{(i)}$ is the estimator of θ at the *i*-th iteration. The MLE can then be found when the difference between $\boldsymbol{\theta}_m^{(i)}$ and $\boldsymbol{\theta}_m^{(i-1)}$ is almost 0. Other methods such as EM algorithm and Markov chain Monte Carlo (MCMC) may also be considered for obtaining the MLE.

2.2.2. STARTING VALUE—The starting values $\boldsymbol{\theta}_m^{(0)}$ can be obtained using method of moments estimator (MME). Let $R_{j_1,j_2,\ldots,j_s}$, where $\{j_1, j_2, \ldots, j_s\}$ is a subset of $\{0, 1, \ldots, l\}$ and $j_1 < j_2 < \ldots < j_s$, be the observed joint recombination rate between the j_1-th, j_2-th, …, and j_s-th markers. By equating some of the observed recombination rates to the corresponding recombination probabilities, the starting values for θ_m can be found.

We use the origin, the first and the last known markers as an illustration. Let R_{0,j_1} and R_{0,j_s} represent the observed recombination rates between T_0 and T_{j1} and between T_0 and T_{j_s}, respectively. As detailed in the Supplement, the starting value of μ, say $\mu^{(0)}$, can be obtained through iteration. Then the starting value of α and β for Model I can be found as

$$\alpha^{(0)} = \frac{\mu^{(0)}\ln(1-2R_{0,j_1})T_{j_s} - \mu^{(0)}\ln(1-2R_{0,j_s})T_{j_1}}{T_{j_s}^2 T_{j_1} - T_{j_1}^2 T_{j_s}} \text{ and } \beta^{(0)} = -\alpha^{(0)} T_{j_1} - \frac{\mu^{(0)}\ln(1-2R_{0,j_1})}{T_{j_1}},$$

and for Model II as

$$\alpha^{(0)} = \frac{T_{j_1} T_{j_s}}{T_{j_1} - T_{j_s}} \ln\left[\frac{T_{j_s}\ln(1-2R_{0,j_1})}{T_{j_1}\ln(1-2R_{0,j_s})}\right] \text{ and } \beta^{(0)} = -\frac{\mu^{(0)}\ln(1-2R_{0,j_1})}{T_{j_1}} e^{\frac{\alpha^{(0)}}{T_{j_1}}}.$$

Note that if $R_{0,j1} = 0$ in Model II, it will be necessary to use a different marker.

2.2.3. MARKER INFORMATION UNKNOWN AT THE ORIGIN—In reality, marker information is usually unknown at the origin point, so R_{0,j_1} and R_{0,j_s} are unobservable. Unfortunately, the crossover probabilities based on markers without the origin are too complicated to provide a formula for a starting value of θ_m. Thus, an estimate for R_{0,j_1} and R_{0,j_s} may be necessary. We use the following summary information from the observed recombination rate of all pairs of markers to obtain the estimates.

$$R_{min}=\min\left(\frac{R_{j_1,j_2}}{T_{j_2}-T_{j_1}},\frac{R_{j_1,j_3}}{T_{j_3}-T_{j_1}},\ldots,\frac{R_{j_{s-1},j_s}}{T_{j_s}-Tj_{s-1}}\right)\ R_{mean}=\text{mean}\left(\frac{R_{j_1,j_2}}{T_{j_2}-T_{j_1}},\frac{R_{j_1,j_3}}{T_{j_3}-T_{j_1}},\ldots,\frac{R_{j_{s-1},j_s}}{T_{j_s}-Tj_{s-1}}\right)$$

Since R_{0,j_1} should be relatively small and R_{0,j_s} should be no more than 0.5, the following estimates are employed assuming L is the length of the chromosome.

$$\hat{\mathrm{R}}_{0,j_1}=\min(R_{min}T_{j_1},\frac{0.5}{L}T_{j_1})\ \hat{\mathrm{R}}_{0,j_s}=\begin{cases}R_{mean}T_{j_s}, & \text{if}R_{mean}T_{j_s}<0.5,\\ \frac{0.5}{L}T_{j_s}, & \text{if}R_{mean}T_{j_s}\geq 0.5.\end{cases}$$

We also suggest a small arbitrary value, say 0.1, for $\mu^{(0)}$, instead of running iteration using $\hat{R_{0,j_1}}$ and R_{0,j_s}, and let the Newton-Raphson method play the primary role for finding the MLE of the parameters.

3. SIMULATION

In the simulation study, we assume that the crossover process occurs at a hypothetical chromosome of 10 Mb length. There are a total of 10 markers located at evenly distributed physical positions 0 Mb (the origin), 1 Mb, 2 Mb, …, and 9 Mb. For each proposed model, we present the results of four simulations with different parameters. For Model I, the parameters are $\alpha = 0.0025$, $\beta = 0$, $\mu = 0.2$ [same as Figure 1(a)(5)] for Simulation (1a); $\alpha = 0.0025$, $\beta = 0.01$, $\mu = 0.5$ for Simulation (1b); $\alpha = 0.0004$, $\beta = 0.01$, $\mu = 0.2$ for Simulation (1c); and $\alpha = 0.0004$, $\beta = 0.05$, $\mu = 0.5$ [same as Figure 1(a)(4)] for Simulation (1d). For Model II, the parameters are $\alpha = 6.5$, $\beta = 0.1$, $\mu = 0.5$ [same as Figure 2(a)(8)] for Simulation (2a); $\alpha = 6.5$, $\beta = 0.5$, $\mu = 1$ for Simulation (2b); $\alpha = 2$, $\beta = 0.05$, $\mu = 0.5$ [same as Figure 2(a)(2)] for Simulation (2c); and $\alpha = 2$, $\beta = 0.1$, $\mu = 0.4$ for Simulation (2d).

For each proposed model and assumed parameters, the crossover probabilities were first calculated based on the marker location and the crossover events which occurred previously. Next, 10,000 random samples with different crossover patterns were generated according to the above crossover probabilities and then the recombination rates were obtained for all recombination patterns among the 10 markers. Furthermore, the recombination information with the origin was removed from the dataset to represent the typical situation where the information at the origin point is unknown. The process was repeated and 100 replicated datasets were created.

For each dataset, 12 subsets with four different markers were further produced. The four markers are considered as the observed markers and are used for parameter estimation. The 12 subsets were chosen to represent different allocation of existing markers: the first four subsets [1236], [1239], [2356] and [1456] have markers located mainly at the first half of the chromosome; the last four subsets [4569], [4578], [1789] and [5789] have the markers mainly located at the latter half of the chromosome; and the remaining subsets [1347], [2468], [1289] and [3679] are considered to have markers more evenly distributed along the entire chromosome.

Figure 3 and Figure 5 display the parameter estimation results from Model I and Model II, respectively. The mean estimate and the corresponding 95% confidence interval (CI) of the three parameters from the 100 replicates are presented for each subset of each simulation. The figures also include the ratio $\frac{\beta}{\mu}$, which is $\lambda_1(t, n)$ when $t = 0$ and $n = 0$ and the upper bound of $\lambda_2(t, n)$. Most of the figures are plotted within 2 units scale of the base unit of the true parameter, 1 unit above and 1 unit below. For example, 0.001 is considered as the base unit for parameter 0.0025 and the scale is from 0.0015 to 0.0035, and 0.0001 is considered as the base unit for parameter 0.0004 and the scale is from 0.0003 to 0.0005. One parameter in Figure 3 and two parameters in Figure 5 are shifted with half unit and five parameters in Figure 3 have scales over 2 units.

From the estimated parameters, we can obtain the estimated recombination rate between any two markers. The mean difference and the corresponding 95% CI between the estimated and the observed recombination rate are demonstrated in Figure 4 for Model I and Figure 6 for Model II. Since each subset contains four markers, there are six pairs of markers for each subset and 72 CIs, in total, for each simulation. All figures are centered at 0 and scaled from −0.002 to 0.002.

For Model I, Simulation (1c) has the best estimation, where only one CI does not cover the true parameter, one replicate for subset [1235] could not converge, and seven CIs, all for subsets [1789] and [5789], in Figure 4 do not cover 0. In Simulation (1d), the performance of the parameter estimation also does well with only three CIs not covering the true parameter and four CIs not covering $\frac{\beta}{\mu}$; but 25% of the recombination rate estimates are biased. Interestingly, although the α, μ and $\frac{\beta}{\mu}$ are biased for subset [5789], the recombination rate estimates are unbiased. This shows some inconsistencies between the parameter estimation and the recombination rate estimation. In Simulation (1b), most of the parameter estimates tend to overestimate with over 50% of the CIs not covering the true parameter. However, all of the CIs cover $\frac{\beta}{\mu}$ and there are only three biased estimates for recombination rate. For Simulation (1a), although the estimates for α are all unbiased, they are all biased for β because the true value 0 is at the boundary and five of the CIs do not cover the true μ. The recombination rate estimation is also the worst with no more than 20% of the CIs covering 0.

For Model II, most of the parameter estimates and the recombination rate estimates are unbiased. The exceptions include β and μ from subsets with markers mainly located at the first half in Simulation (2a), two βs and five μs in Simulation (2b), four $\frac{\beta}{\mu}$s in Simulation (2c), all parameters for subset [5789] and two others in Simulation (2d), and a total of 13 recombination rates in Figure 6. There are only two replicates, one from subset [1456] in Simulation (2b) and one from subset [5789] in Simulation (2d), which could not converge. Also, the variation of the estimates for α tends to be larger when the markers are mainly located at the latter part of the chromosome.

4. APPLICATION

There are two large datasets of *Drosophila* X-chromosome, from Morgan et al. (1935) and Weinstein (1936), that have been used in other papers (McPeek and Speed, 1995; Zhao et al., 1995a; Risch and Lange, 1985). Morgan et al.'s (1935) dataset (M-data) contained the counts of recombination events of 16,136 *Drosophila* offsprings on nine markers based on

the phenotype of the flies, while Weinstein's (1936) dataset (W-data) had 28,239 events on seven markers and six of them were the same as in the M-data. The X-chromosome of *Drosophila* has a total length of 21.78 Mb of DNA with one arm and recombination occurs only in females. Table 1 summarizes the six markers with their cytogenetic position, genetic position and physical position from the telomere. The number of recombination events and rates among the six markers from both M-data and W-data are listed in Table 2. From the table, we find that the non-recombination rate is lower in the W-data than in the M-data (0.45 vs. 0.58), the recombination rates between pairs of markers are all higher in the Wdata except between *ec* and *cv*, and the recombination rate between *v* and *f* in the W-data is almost 3 times the rate in the M-data (0.22 vs. 0.08).

For each dataset, the two proposed models were first applied to all six markers. Then one marker was removed and the models were fit to each of the six different combinations of five markers. Results for parameter estimation are shown in Table 3. The β in Model I is estimated to be close to 0 from both datasets except when one of the first two markers, *sc* or *ec*, is removed. We also observe that the estimates of α and μ in Model I and β and μ in Model II with five markers are similar to those with six markers except when the last marker *f* in W-data and when one of the last two markers, *v* or *f*, in M-data is removed. The estimate of α in Model II is inconsistent in M-data but is similar in W-data except when *sc* is excluded.

Furthermore, for each of the seven different combinations of markers, we obtained the estimated recombination rate between any two markers using the estimated parameters. Figure 7 illustrates the difference between the estimated and observed recombination rate using box-plots and showing that the estimation for W-data (range −0.026, 0.019) is much better than for M-data (range −0.029, 0.098). For M-data, the recombination rates tend to be over-estimated for most combinations of markers except for models not involving either the first marker *sc* or the last marker *f*. For most combinations of markers except [*ec*,*cv*,*ct*,*v*,*f*], the differences obtained from Model II have a smaller range and the median is closer to 0 compare to Model I. For W-data, the differences from Model I are slightly tighter than from Model II and the rates seem to be slightly under-estimated from Model II.

5. DISCUSSION

The crossover process has long been studied for building the genetic map since the physical map was not well constructed. It is usually modeled with a stationary renewal process so that different pairs of markers with the same recombination rate are considered to have the same genetic distance between the two markers in each pair. However, this is not the case for physical distance, where different pairs of markers with the same physical distance between the two markers may not have the same recombination rate. Now, since the physical map on nucleotide bases is available for many species, we have more information to model the crossover process based directly on the physical positions of genetic markers.

In order to incorporate crossover interference, we propose two non-homogeneous Poisson process models with positive interference. The assumptions include (1) the crossover process is for the four-strand chromatid bundle with no chromatid interference; (2) the crossover process starts from an origin point and continues sequentially along the chromosome; and (3) the crossover interference depends on the location of the markers, and the number of crossover events that have occurred previously.

The increment rates of the two proposed models are $\lambda_1(t,n)=\frac{2\alpha t+\beta}{n+\mu}$ and $\lambda_2(t,n)=\frac{\beta}{n+\mu}(1+\frac{\alpha}{t})e^{-\frac{\alpha}{t}}$, and both models decrease with n, the number of events occurred previously, at the same rate. $\lambda_1(t, n)$ is a simple linear increasing function of distance t and is bounded by 0 and $\frac{2\alpha L+\beta}{\mu}$. Since L is the length of the chromosome, $\lambda_1(t, n)$ can be quite large. Alternatively, $\lambda_2(t, n)$ is also an increasing function of t but has more flexibility and is upper bounded by $\frac{\beta}{\mu}$.

In order to better understand these two models, we present in Figure 1 and Figure 2 various increment rates and provide the expected recombination probability between various pairs of markers. In Model I, while the increment rate increases with t linearly, the expected recombination probability also changes with T_1 relatively linearly when the distance between T_1 and T_2 is the same. In Model II, the changes are both nonlinear and the expected recombination probability between two markers with the same distance can be much higher when T_1 is located in the middle of the chromosome compare to its location at the two ends of the chromosome. This flexibility makes Model II more preferred.

From the simulation study, we learned that the performance of estimation depends on the values of the true parameters. Take graph (3) in Figure 2 as an example, since the increment rate increases quickly from 0 to 0.4 in a short distance and the expected recombination probability between markers located close to the origin is quite high, it takes much longer time for the model to converge and also leads to a higher rate of convergence failure. Another example is Simulation (1a) for Model I, since the true β is 0, the parameter estimation is obviously biased and also results in a biased estimation for the recombination rate. The performance also depends on the subsets involved in the model. Depending on the values of the parameters, when observable markers are confined to certain part of the chromosome, the estimation tends to have larger variation and be more biased.

Since some values of the parameters may not be appropriate for the crossover process, the values chosen for the simulation study are modified from the parameters estimated from the *Drosophila* data in order for the simulations to be more realistic. Based on the selected simulations, the estimation from Model II seems to perform better than that of Model I, and this may be due to the flexibility of Model II and also the choice of true parameters. The performance of the recombination rate estimation seems to be more related to $\frac{\beta}{\mu}$ although there are several exceptions such as subset [5789] in Model I Simulation (1d). Some of the biased estimates may be due to the loss of information at the origin or that only recombination events instead of crossover events are observed. When the recombination information between the existing markers and the origin is restored for Model I (results not shown), most of the estimates become unbiased in Simulations (1c) and (1d) but only the variation is reduced in Simulations (1a) and (1b).

When working with the two *Drosophila* X-chromosome data, we observed that the recombination rates are quite different between the two datasets. The rates between pairs of markers are all higher in the W-data except for one pair (difference −0.003), and the differences involving marker *f* are much larger (range 0.075, 0.14) than the differences not involving marker *f* (range −0.003, 0.03). Figure 7 shows that both of our models fit the W-data better than the M-data, and the performance improves in the M-data when marker *f* is

removed. Neither model shows obvious advantages over the other one although Model II seems to fit the M-data a little better then Model I.

In McPeek and Speed (1995), they summarized several crossover process models for genetic distance and used the Monte Carlo methods to fit the recombination data from the first five markers [*sc,ec,cv,ct,v*] from the M-data. Figure 8 shows the difference between the estimated and the observed recombination rate between all pairs markers from the six models they presented. All the models in their paper tend to underestimate the recombination rate, however, the estimation from the gamma model, the count-location model and the King-Mortimer II model is very good with the difference ranging from −0.006 to 0.0015. The difference from our two models for these five markers ranges more similarly to the other three models but is not as biased.

The following issues may limit the performance of our proposed models. First, only the recombination events not the actual crossover events are observed; second, recombination information is only obtained from some observable markers; and third, recombination information related to the origin point is missing. Furthermore, the estimates in our models are nonlinear and hence as in many cases of MLE, the unbiasedness of the estimating equations may not transmit to the unbiasedness of the derived estimators. If the order of bias of the estimators is $O(\frac{1}{n})$, one could use the jackknife method to reduce the bias to the order $O(\frac{1}{n^2})$. However, in the present context that needs further appraisal.

In the actual data application, our proposed models are compatible with some of the renewal process models. The relatively larger difference between the estimated and the observed recombination rate suggests that additional parameters may be needed for the proposed models or different increment rates should be considered. We also recognize that the underlying assumptions of our proposed crossover process may not be accurate. A thorough discussion of various crossover processes can be found in Karlin and Liberman (1994).

We consider this paper a starting point to construct alternative modeling for the crossover interference based on the physical map. Since the positions of markers are known, the model is not burdened with estimating the genetic distance and hence can focus on modeling the recombination events. With the model parameters estimated, it is also possible to predict the location of new unknown markers, which is not shown in this paper. The two models we propose are non-stationary and thus provide more flexibility. However, because the proposed models do require information about the physical positions of some markers, they are not for the purpose of constructing a genetic map but rather an additional and potentially useful tool for modeling the crossover process and interference. We intend to undertake more studies in developing such models based on the physical map and hope to have provided useful information in this area.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors are grateful to both the reviewers and the editor for their helpful comments and criticisms.

REFERENCE

Broman KW, Weber JL. Characterization of human crossover interference. American Journal of Human Genetics. 2000; 66:1911–1926. [PubMed: 10801387]

Fisher RA, Lyon MF, Owen ARG. The sex chromosome in the house mouse. Heredity. 1947; 1:355–365.

Haldane JBS. The combination of linkage values and the calculation of distances between the loci of linked factors. Journal of Genetics. 1919; 8:299–309.

Karlin S, Liberman U. Measuring interference in the chiasma renewal formation process. Advances in Applied Probability. 1983; 14:471–487.

Karlin S, Liberman U. Theoretical recombination processes incorporating interference effects. Theoretical Population Biology. 1994; 46:198–231. [PubMed: 7974280]

Mather K. Crossing Over. Biological reviews of the Cambridge Philosophical Society. 1938; 13:252–292.

McPeek MS, Speed TP. Modeling interference in genetic recombination. Genetics. 1995; 139:1031–1044. [PubMed: 7713406]

Morgan TH, Bridges CB, Schultz J. Constitution of the germinal material in relation to heredity. Carnegie Institution Washington Yearbook. 1935; 34:284–291.

Owen ARG. The theory of genetical recombination. Proceedings of the Royal Society of London. Series B, Biological sciences. 1949; 136:67–94.

Risch N, Lange K. Statistical analysis of multilocus recombination. Biometrics. 1983; 39:949–963. [PubMed: 6231060]

Snustad, DP.; Simmons, MJ.; Jenkins, JB. Principles of Genetics. John Wiley & Sons, Inc.; 1992.

Speed, TP. What Is A Genetic Map Function?. In: Speed, TP.; Waterman, MA., editors. Genetic Mapping and DNA Sequencing. Springer-Verlag; 1996.

Weinstein A. The theory of multiple-strand crossing over. Genetics. 1936; 21:155–199. [PubMed: 17246790]

Zhao H, Speed TP, McPeek MS. Statistical analysis of crossover interference using the chi-square model. Genetics. 1995a; 139:1045–1056. [PubMed: 7713407]

Zhao H, McPeek MS, Speed TP. Statistical analysis of chromatid interference. Genetics. 1995b; 139:1057–1065. [PubMed: 7713408]

Zhao H, Speed TP. On genetic map functions. Genetics. 1996; 142:1369–1377. [PubMed: 8846913]

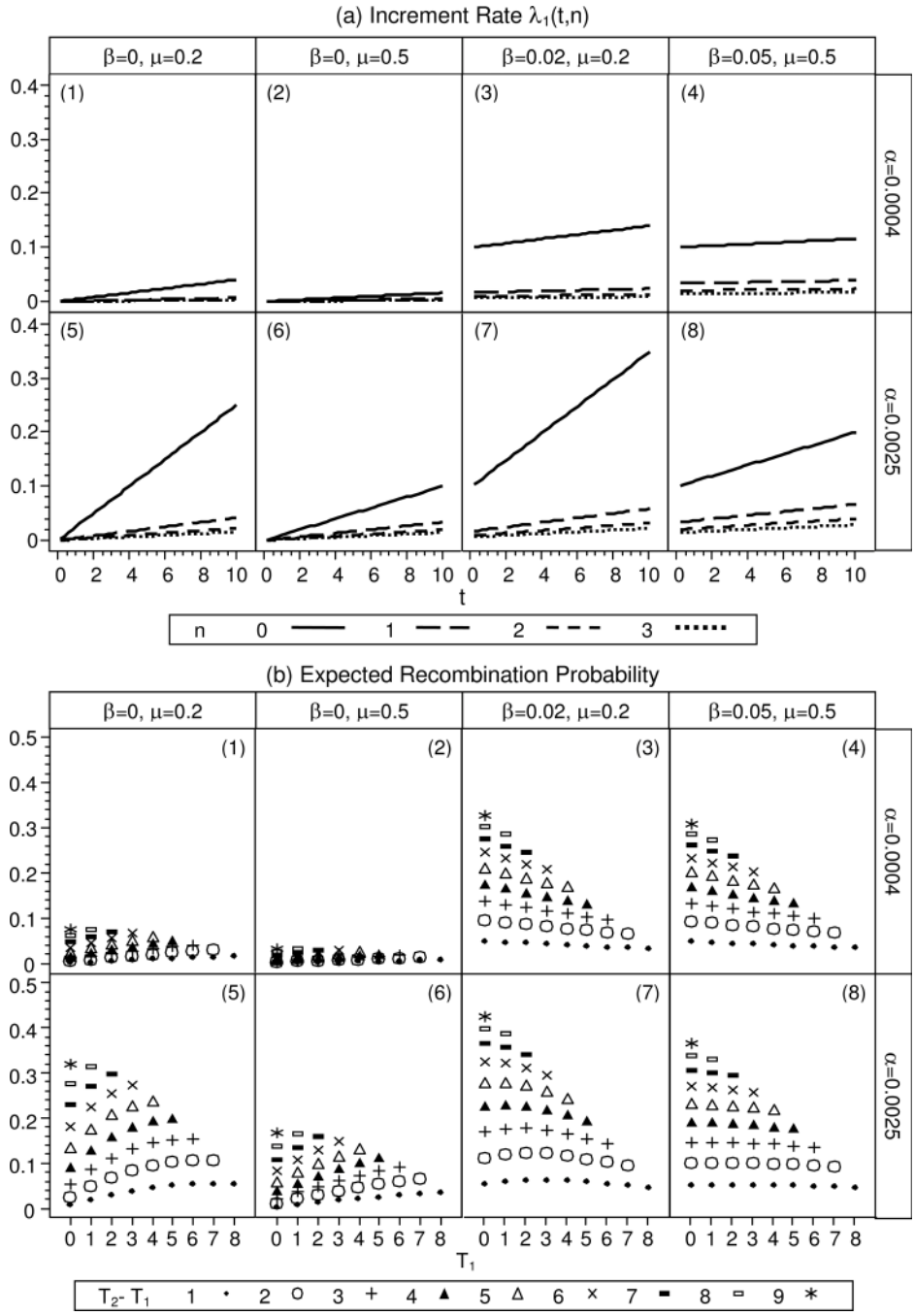

Figure 1.
Model I — Various Increment Rates and Expected Recombination Probabilities

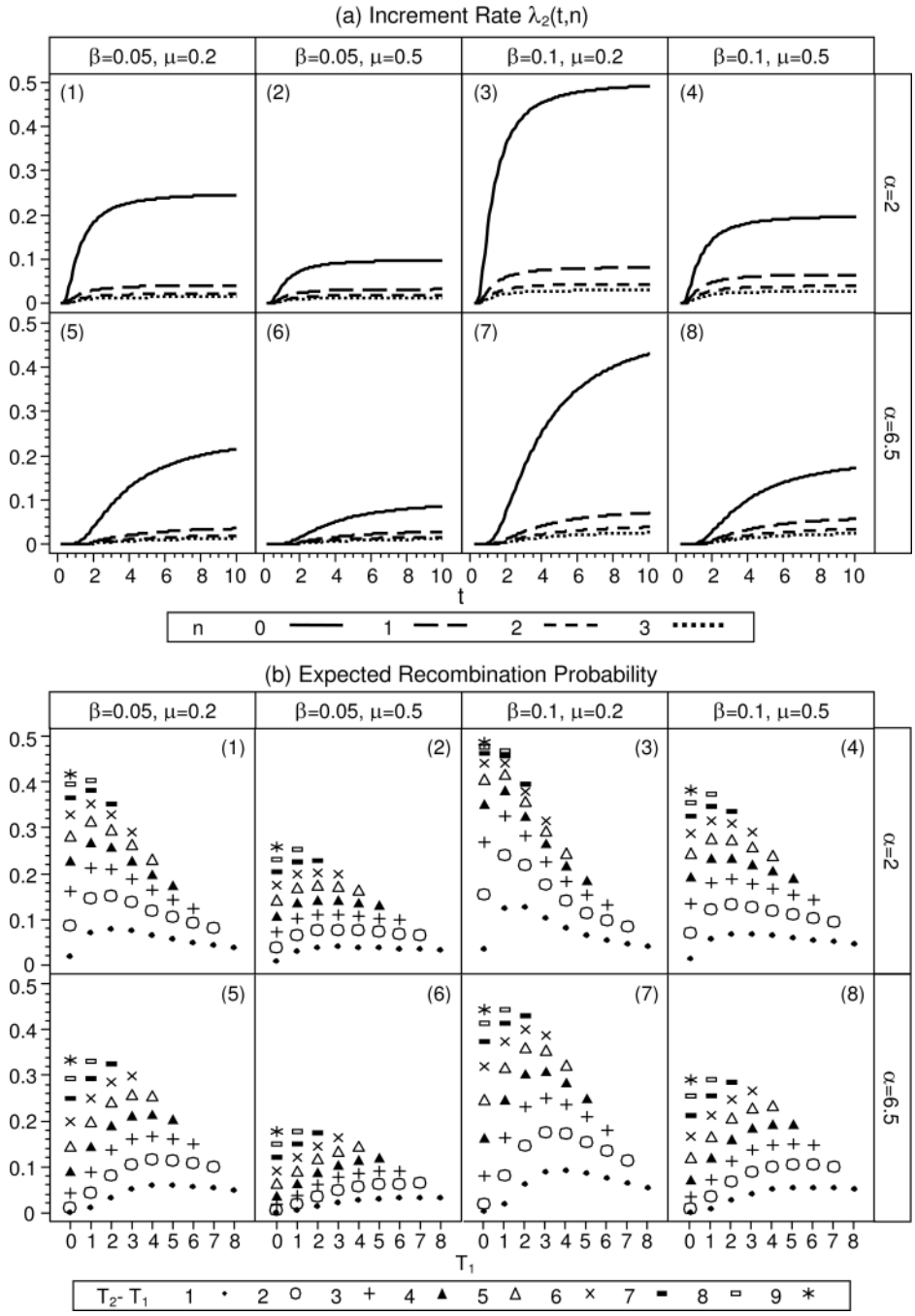

Figure 2.
Model II — Various Increment Rates and Expected Recombination Probabilities

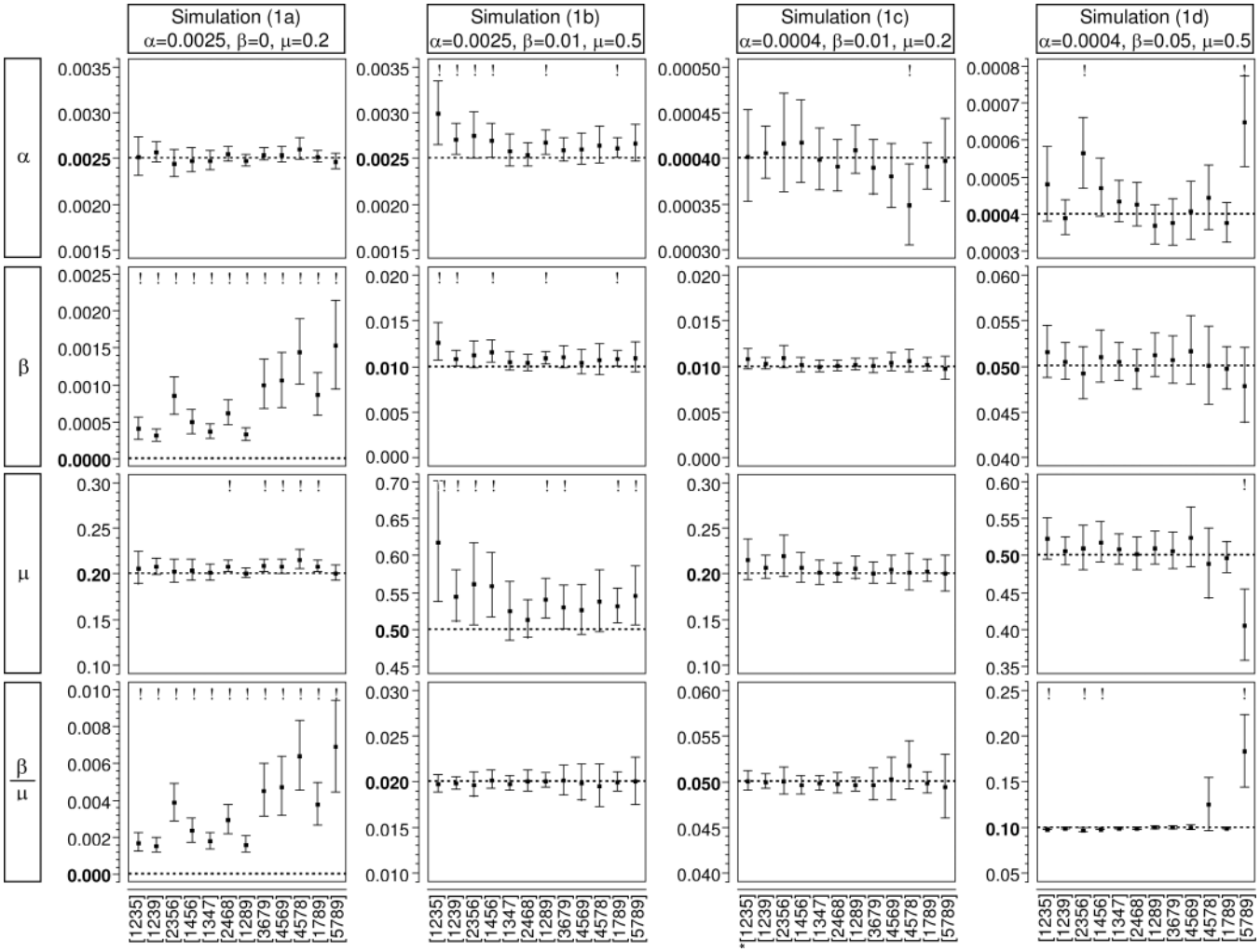

Figure 3.

Model I — Parameter Estimation for Simulated Data

Dash line represents the true parameter, dot is the mean and the bars are the upper and lower 95% CI.

!: The 95% CI does not cover the true parameter. *: Could not converge in one replicate.

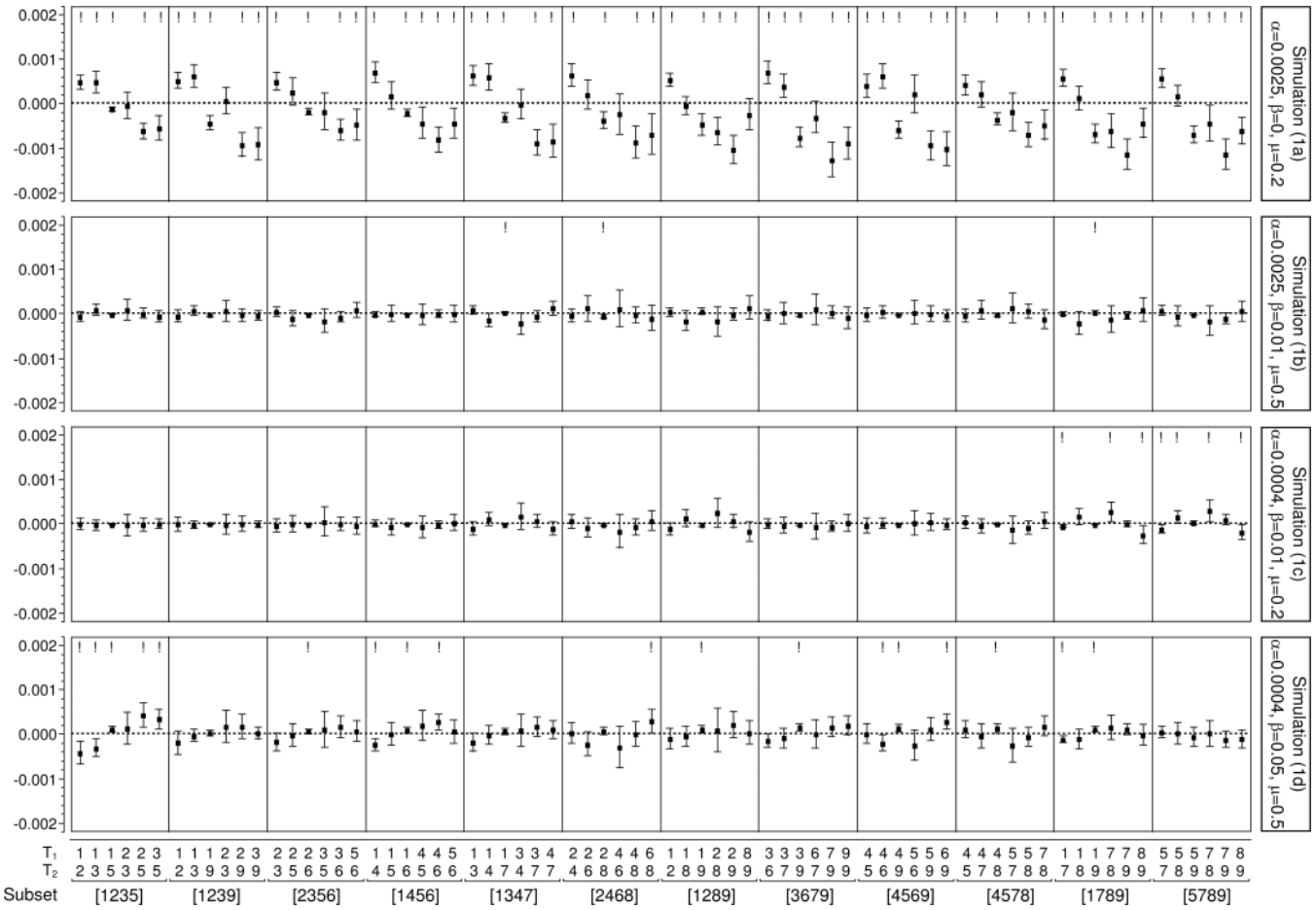

Figure 4.

Model I — Difference between Estimated and Observed Recombination Rate between Two Markers (T_1, T_2)

Dash line represents 0, dot is the mean and the bars are the upper and lower 95% CI.

!: The 95% CI does not cover 0.

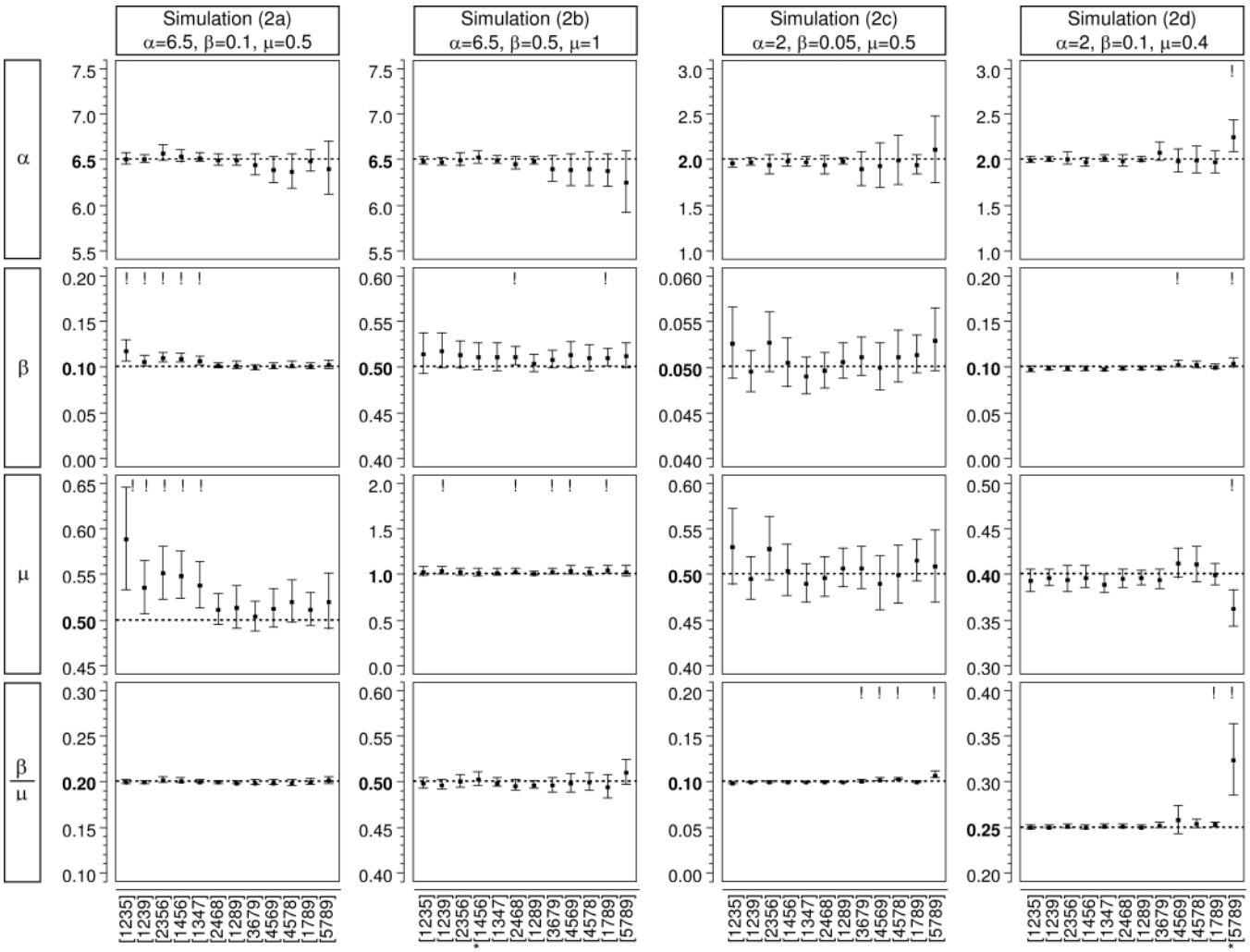

Figure 5.

Model II — Parameter Estimation for Simulated Data

Dash line represents the true parameter, dot is the mean and the bars are the upper and lower 95% CI.

!: The 95% CI does not cover the true parameter. *: Could not converge in one replicate.

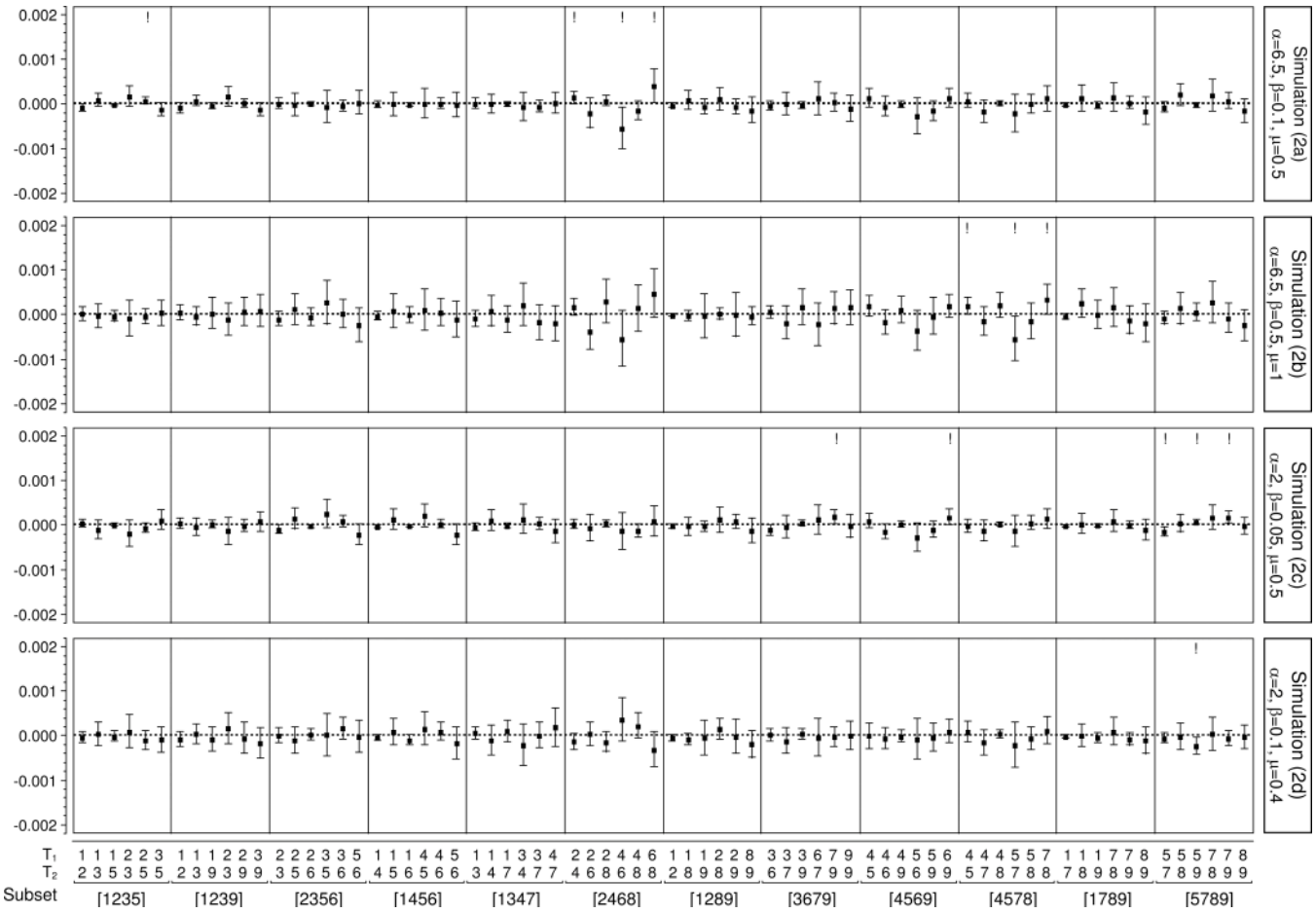

Figure 6.

Model II — Difference between Estimated and Observed Recombination Rate between Two Markers (T_1, T_2)

Dash line represents 0, dot is the mean and the bars are the upper and lower 95% CI.

!: The 95% CI does not cover 0.

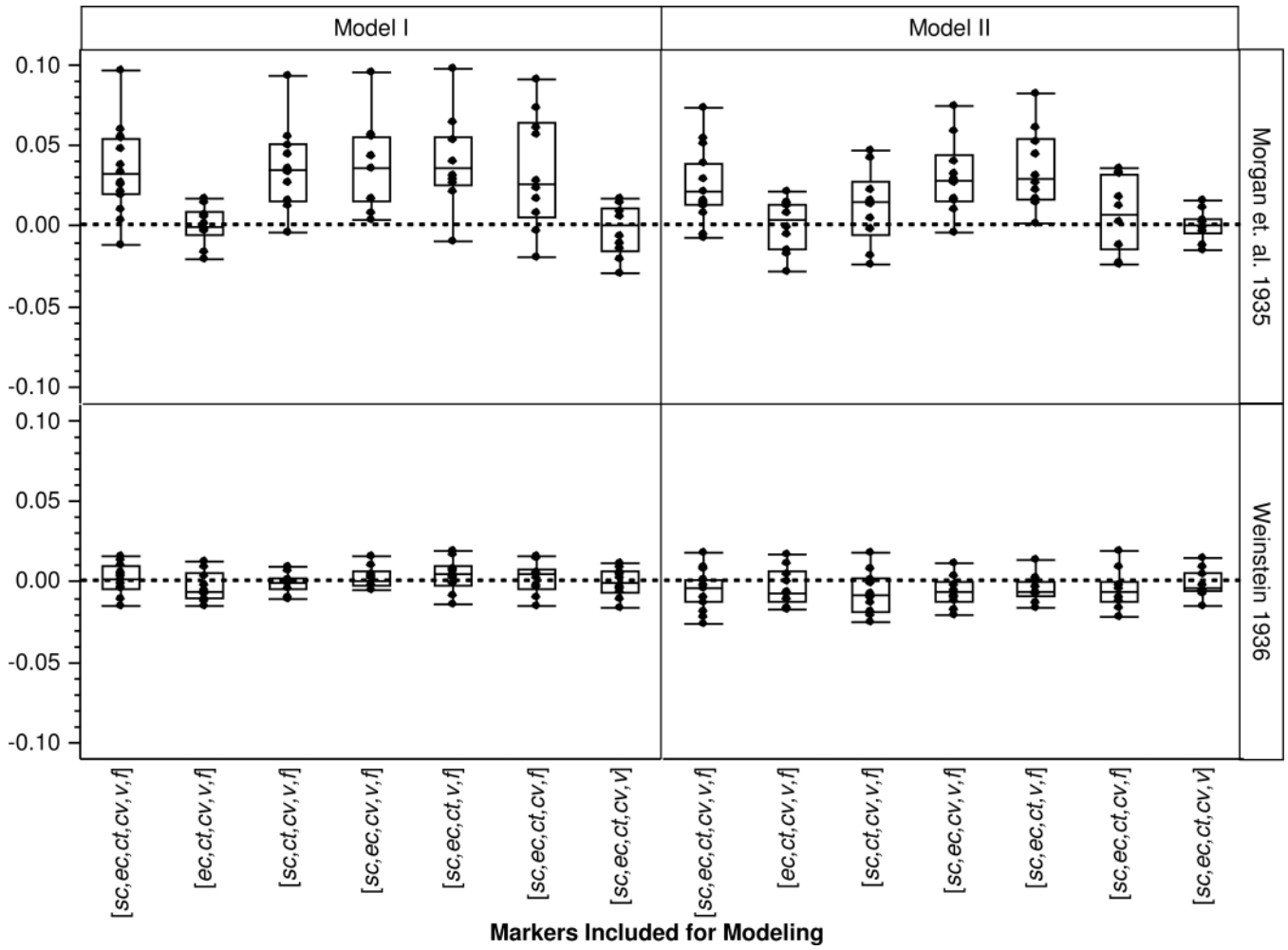

Figure 7.
Difference between Estimated and Observed Recombination Rate between Two Markers for Morgan et. al. (1935) and Weinstein (1936)

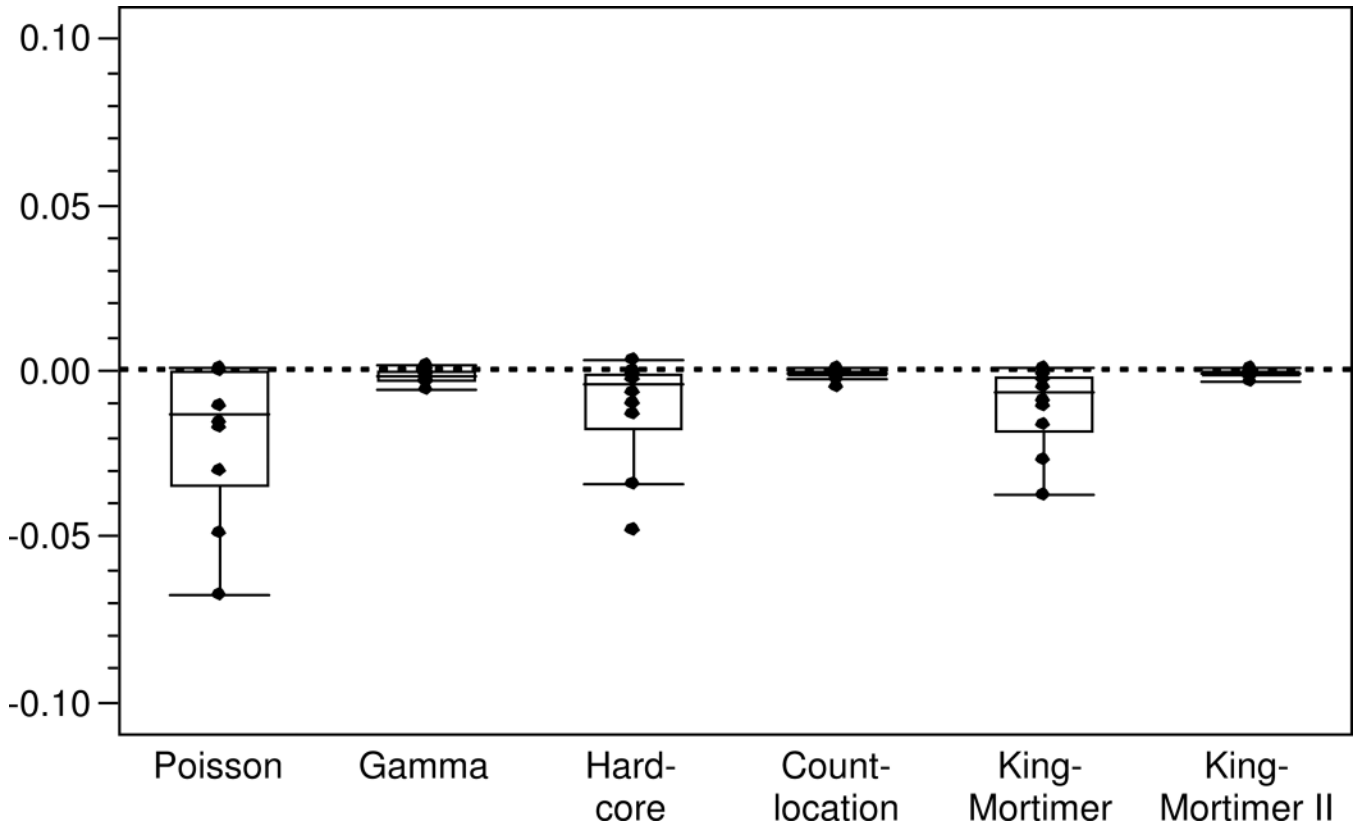

Figure 8.
Difference between Estimated and Observed Recombination Rate between Two Markers in Morgan et al. (1935) from Various Models for Genetic Maps Presented in McPeek and Speed (1995)
* Five markers *sc, ec, cv, ct, v* are included for modeling.

Table 1

Summary of Six Markers on *Drosophila* X-chromosome Used in Both Morgan et al. (1935) and Weinstein (1936)

Marker (Abbreviation)	Cytogenetic Position	Genetic Position (cM)	Physical Position (mb)
scute (sc)	1A8	0.0	0.29
echinus (ec)	3F3	5.5	3.73
crossveinless (cv)	5A13	13.7	5.59
cut (ct)	7B4-6	20.0	7.54
vermilion (v)	9F11	33.0	10.82
forked (f)	15F4-7	56.7	17.16

Source: All information is obtained from Flybase website at University of Indiana, http://flybase.bio.indiana.edu/ (Version FB2007_03, released November 1,

Table 2

Recombination Data from Morgan et al. (1935) and Weinstein (1936)

Recombination Pattern					Morgan et al. (1935)		Weinstein (1936)	
sc-ec	*ec-cv*	*cv-ct*	*ct-v*	*v-f*	Event	Rate	Event	Rate
0	0	0	0	0	9369	0.58	12835	0.45
0	0	0	0	1	1062	0.066	4423	0.16
0	0	0	1	0	1952	0.12	3379	0.12
0	0	1	0	0	1160	0.072	1981	0.070
0	1	0	0	0	1486	0.092	2019	0.072
1	0	0	0	0	713	0.044	1408	0.050
0	0	0	1	1	42	0.0026	438	0.016
0	0	1	0	1	61	0.0038	360	0.013
0	0	1	1	0	25	0.0015	42	0.0015
0	1	0	0	1	93	0.0058	579	0.021
0	1	0	1	0	53	0.0033	136	0.0048
0	1	1	0	0	4	2.5e-4	11	3.9e-4
1	0	0	0	1	58	0.0036	404	0.014
1	0	0	1	0	45	0.0028	142	0.0050
1	0	1	0	0	6	3.7e-4	16	5.7e-4
1	1	0	0	0	4	2.5e-4	9	3.2e-4
0	0	1	1	1	0	0	5	1.8e-4
0	1	0	1	1	0	0	12	4.2e-4
0	1	1	0	1	0	0	0	0
0	1	1	1	0	0	0	1	3.5e-5
1	0	0	1	1	1	6.2e-5	25	8.9e-4
1	0	1	0	1	1	6.2e-5	6	2.1e-4
1	0	1	1	0	0	0	0	0
1	1	0	0	1	0	0	2	7.1e-5
1	1	0	1	0	0	0	1	3.5e-5
1	1	1	0	0	1	6.2e-5	3	1.1e-4
0	1	1	1	1	0	0	0	0

Commun Stat Theory Methods. Author manuscript; available in PMC 2015 January 01.

Recombination Pattern					Morgan et al. (1935)		Weinstein (1936)	
sc-ec	*ec-cv*	*cv-ct*	*ct-v*	*v-f*	Event	Rate	Event	Rate
1	0	1	1	1	0	0	0	0
1	1	0	1	1	0	0	0	0
1	1	1	0	1	0	0	1	3.5e-5
1	1	1	1	0	0	0	1	3.5e-5
1	1	1	1	1	0	0	0	0
Total					16136		28239	

Table 3

Parameter Estimation for Morgan et al. (1935) and Weinstein (1936)

Markers Included for Modeling	α (se)	β (se)	μ (se)
Morgan et al. (1935) – Model I			
[*sc*,*ec*,*cv*,*ct*,*v*,*f*]	4.3e-4 (2.2e-5)	1e-8	0.030 (0.0016)
[*ec*,*cv*,*ct*,*v*,*f*]	4.6e-4 (2.8e-5)	0.0014 (2.0e-4)	0.031 (0.0019)
[*sc*,*cv*,*ct*,*v*,*f*]	4.1e-4 (2.2e-5)	5.3e-4 (1.8e-4)	0.033 (0.0023)
[*sc*,*ec*,*ct*,*v*,*f*]	4.4e-4 (2.3e-5)	1e-8	0.031 (0.0017)
[*sc*,*ec*,*cv*,*v*,*f*]	4.3e-4 (2.3e-5)	1e-8	0.029 (0.0016)
[*sc*,*ec*,*cv*,*ct*,*f*]	5.8e-4 (3.3e-5)	1e-8	0.047 (0.0029)
[*sc*,*ec*,*cv*,*ct*,*v*]	0.0010 (9.4e-5)	1e-8	0.097 (0.0095)
Morgan et al. (1935) – Model II			
[*sc*,*ec*,*cv*,*ct*,*v*,*f*]	6.99 (0.24)	0.013 (7.1e-4)	0.055 (0.0041)
[*ec*,*cv*,*ct*,*v*,*f*]	4.10 (0.22)	0.013 (8.5e-4)	0.056 (0.0039)
[*sc*,*cv*,*ct*,*v*,*f*]	4.61 (0.25)	0.014 (8.1e-4)	0.086 (0.0070)
[*sc*,*ec*,*ct*,*v*,*f*]	7.33 (0.25)	0.014 (7.4e-4)	0.055 (0.0042)
[*sc*,*ec*,*cv*,*v*,*f*]	7.40 (0.25)	0.013 (7.2e-4)	0.047 (0.0036)
[*sc*,*ec*,*cv*,*ct*,*f*]	5.38 (0.18)	0.020 (0.0013)	0.13 (0.010)
[*sc*,*ec*,*cv*,*ct*,*v*]	7.15 (0.19)	0.023 (0.0021)	0.11 (0.011)
Weinstein (1936) – Model I			
[*sc*,*ec*,*cv*,*ct*,*v*,*f*]	0.0025 (6.4e-5)	1e-8	0.21 (0.0060)
[*ec*,*cv*,*ct*,*v*,*f*]	0.0023 (2.1e-5)	0.10 (0.0062)	0.15 (0.0085)
[*sc*,*cv*,*ct*,*v*,*f*]	0.0025 (6.6e-5)	0.0048 (7.8e-4)	0.25 (0.011)
[*sc*,*ec*,*ct*,*v*,*f*]	0.0026 (6.7e-5)	1e-8	0.22 (0.0064)
[*sc*,*ec*,*cv*,*v*,*f*]	0.0027 (7.0e-5)	1e-8	0.22 (0.0065)
[*sc*,*ec*,*cv*,*ct*,*f*]	0.0028 (9.3e-5)	1e-8	0.23 (0.0086)
[*sc*,*ec*,*cv*,*ct*,*v*]	0.0017 (9.3e-5)	1e-8	0.14 (0.0082)
Weinstein (1936) – Model II			
[*sc*,*ec*,*cv*,*ct*,*v*,*f*]	6.38 (0.14)	0.088 (0.0027)	0.41 (0.018)
[*ec*,*cv*,*ct*,*v*,*f*]	7.88 (0.28)	0.085 (0.0030)	0.38 (0.020)
[*sc*,*cv*,*ct*,*v*,*f*]	6.22 (0.19)	0.091 (0.0029)	0.43 (0.021)
[*sc*,*ec*,*ct*,*v*,*f*]	6.49 (0.14)	0.094 (0.0030)	0.43 (0.019)
[*sc*,*ec*,*cv*,*v*,*f*]	6.45 (0.14)	0.096 (0.0031)	0.43 (0.020)
[*sc*,*ec*,*cv*,*ct*,*f*]	6.31 (0.14)	0.097 (0.0039)	0.45 (0.024)
[*sc*,*ec*,*cv*,*ct*,*v*]	6.42 (0.13)	0.038 (0.0022)	0.17 (0.011)

se: estimated standard error.