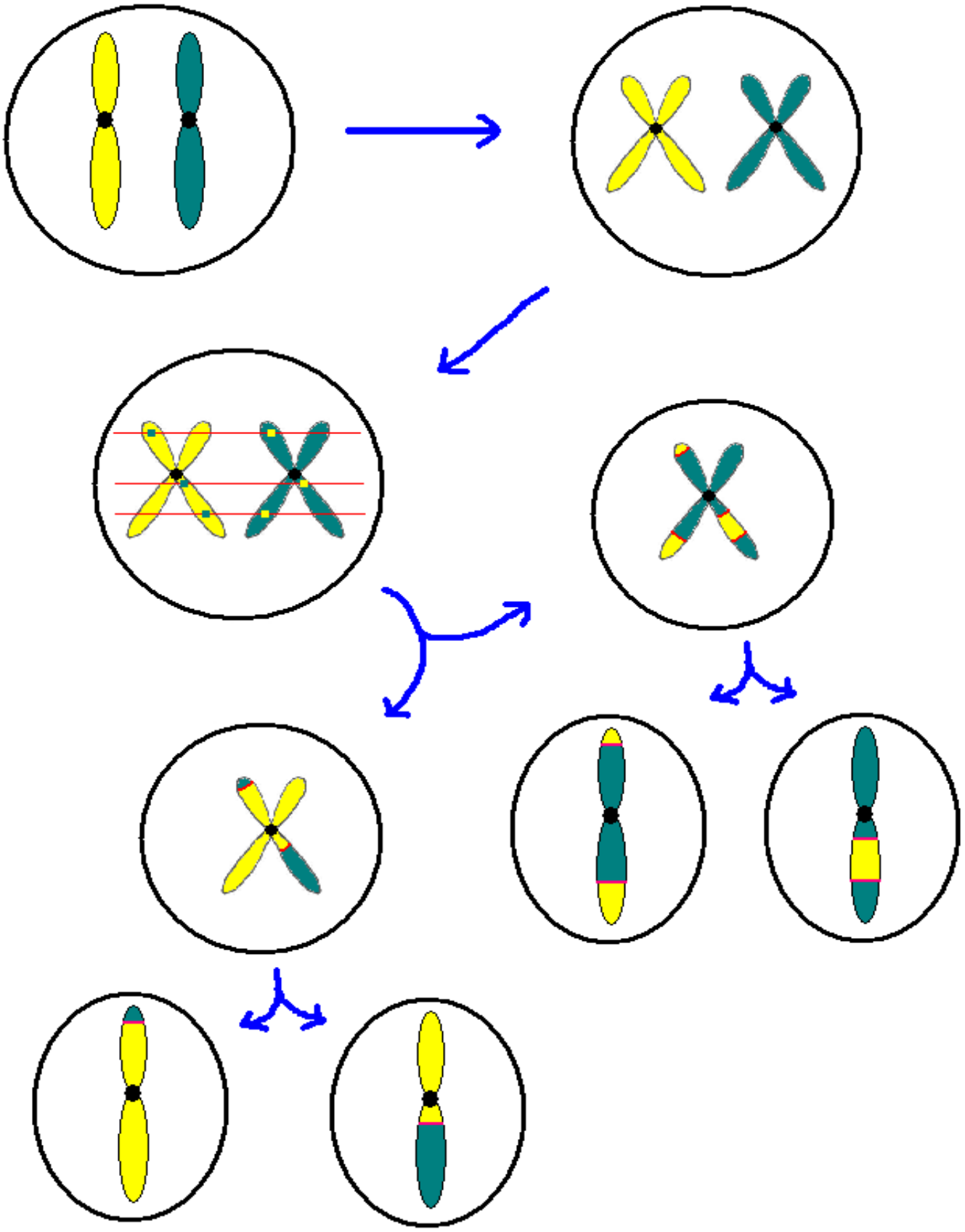
Human crossover interference

Karl W Broman
Department of Biostatistics
Johns Hopkins University

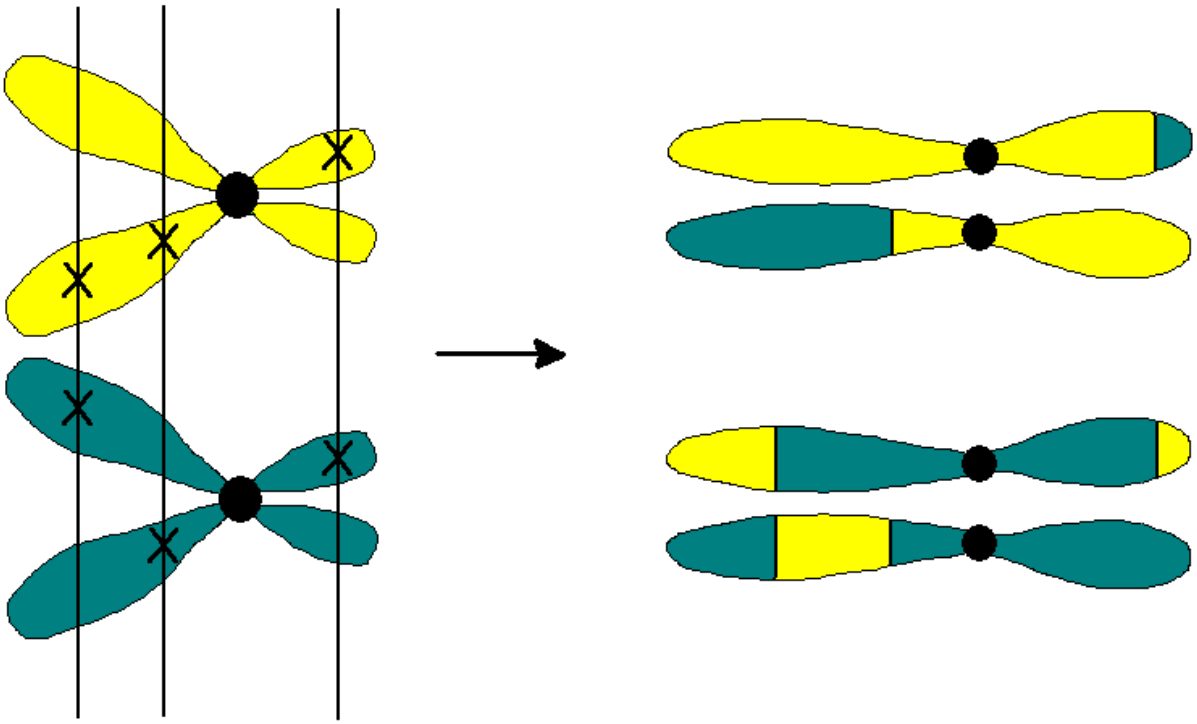
Joint work with James L. Weber,
Marshfield Medical Research Foundation

http://kbroman.homepage.com

Meiosis



Interference



- Strand choice

  → Chromatid interference

- Spacing

  → Chiasma (crossover) interference

Why study interference?

- Estimate the probability of a double crossover in a small interval
- Obtain a model of meiosis for simulation and analysis
- Compare human meiosis to that of other organisms

Goals

- Demonstrate the presence of interference in human meiosis
- Obtain an empirical map function
- Find a good model

Model organisms

- Lots of meioses
- A few linked markers
- Look at frequency of rare multiple recombination events

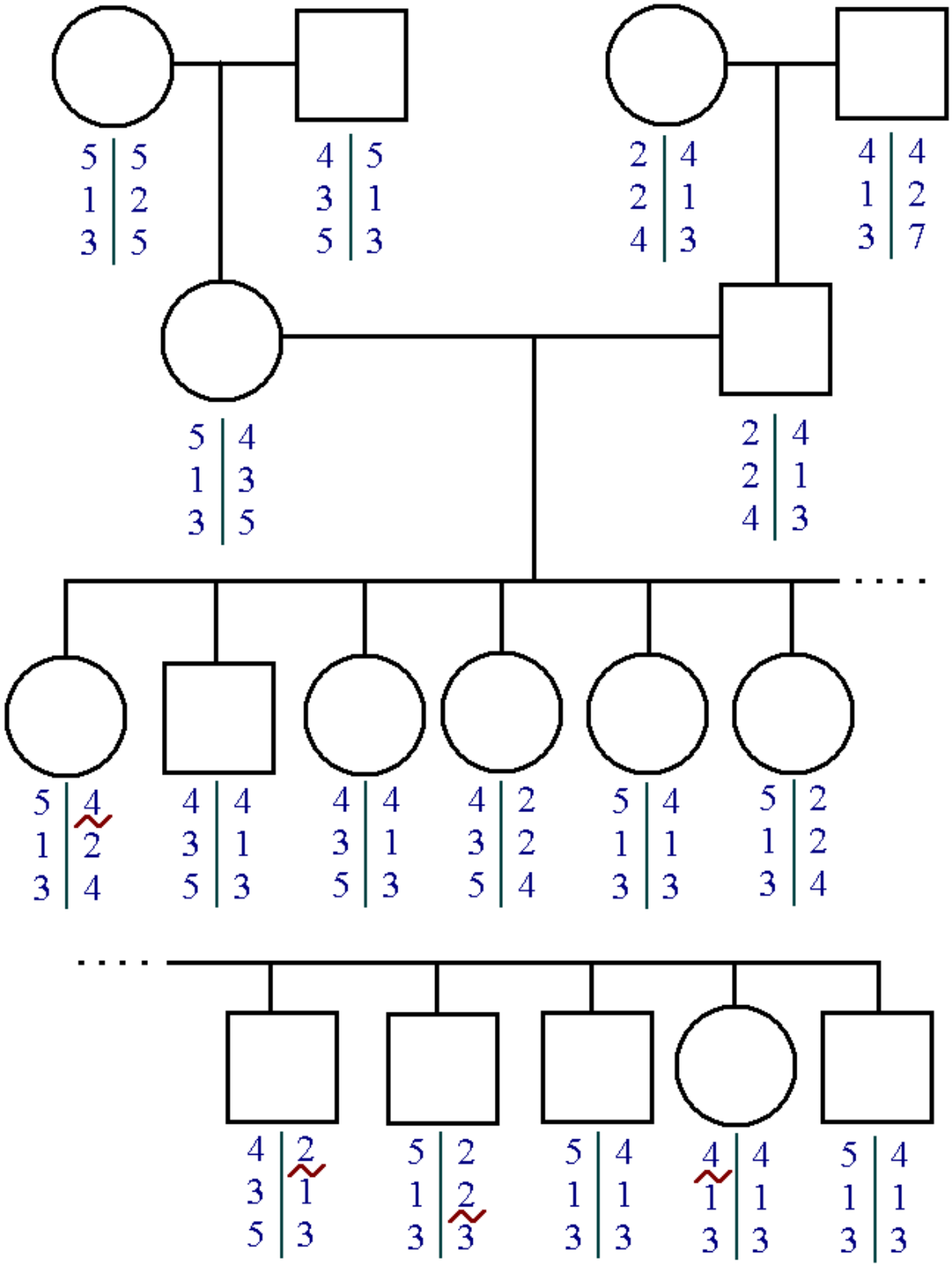
***Drosophila* data (Morgan et al 1935)**

Event	Count	Event	Count
0000	10,431	1001	46
1000	771	0101	53
0100	1,579	0011	25
0010	1,221	1110	1
0001	1,994	1101	1
1100	4	1011	1
1010	7	0111	1
0110	4	1111	1

Human data

- research.marshfieldclinic.org/genetics
- 8 CEPH families
  - three generations
  - 11 to 15 progeny
  - 92 meioses
- ~8,000 STRP markers
  - 90 ± 7 % typed
- Average spacing
  - female: 0.6 ± 1.2 cM
  - male: 0.4 ± 1.0 cM
  - sex-ave: 0.5 ± 0.9 cM
- Data cleaning
  - Removed 764/964,425 (~0.08%) genotypes resulting in tight double recombinants

CEPH pedigree

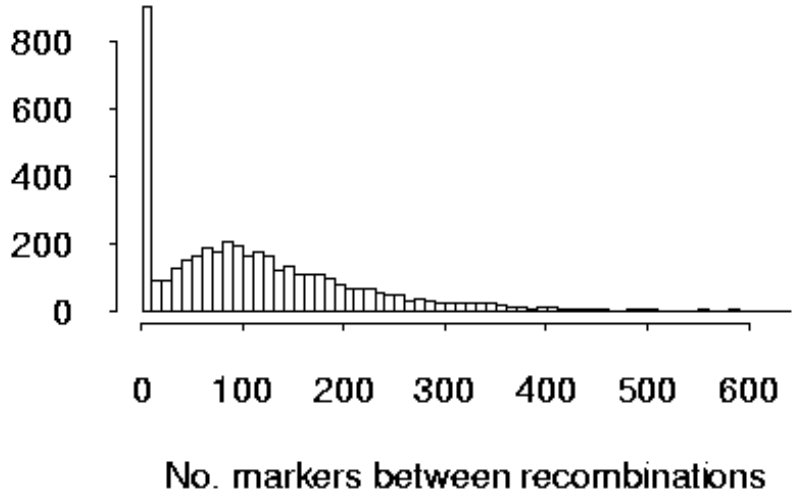


CRI-MAP *chrompic* output

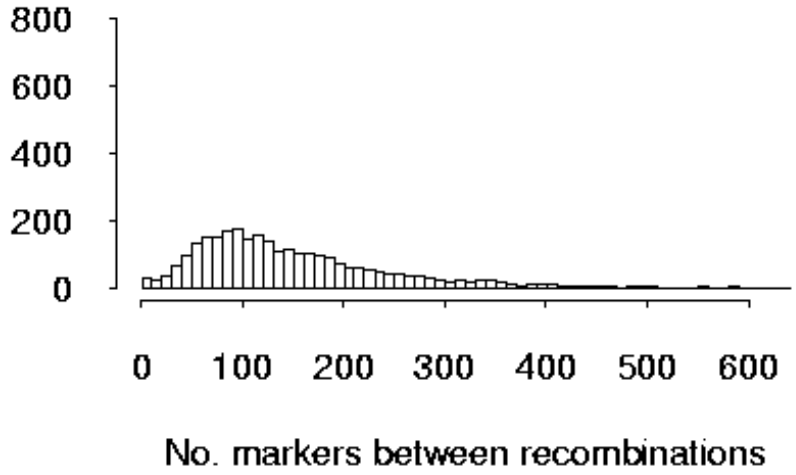
CEPH individual 1331–11
maternal chromosome 10

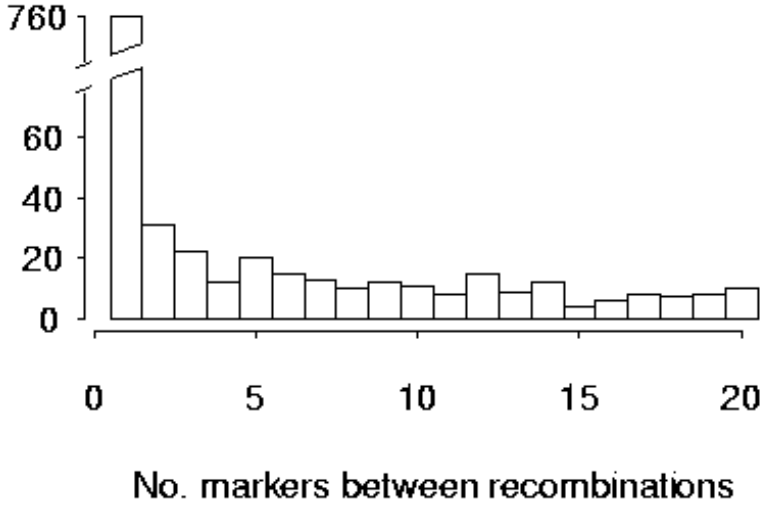
```
1111111--- 11-111-11- --11--111i
1-11---11- 111111--i- 11-1111-11
--11111-11 1111-11111 11--111111
111-111i-i 1111111111 10000-0-00
0--o00-000 0000-00000 0000--0000
0000--0000 0000-00000 o00-0--0--
--0-11-11- -111ii1i-1 ---1-i-1-i
1111-i--11 11111-11i1 -11i-11111
-1----i111 1i1111-111 -11i1-111-
11-111111i 111-i111i- 1111111-i-
1111111-1i 1i-111i11- 1i--1-11-1
111-1i-1-1 1-1----1-1 1i-1ii1i11
1i--1--1i- 11i11--111 11--1i111i
1i1i-11111 i-0---0000 00000-000o
o0-00o
```

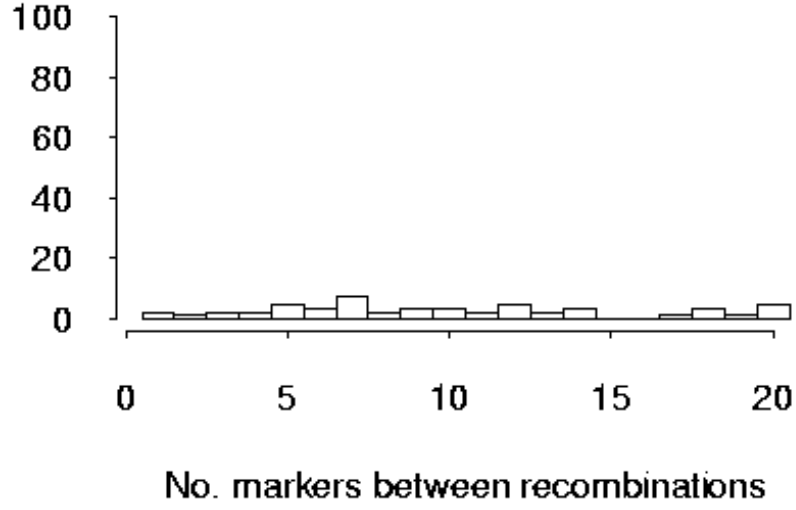
Raw Data



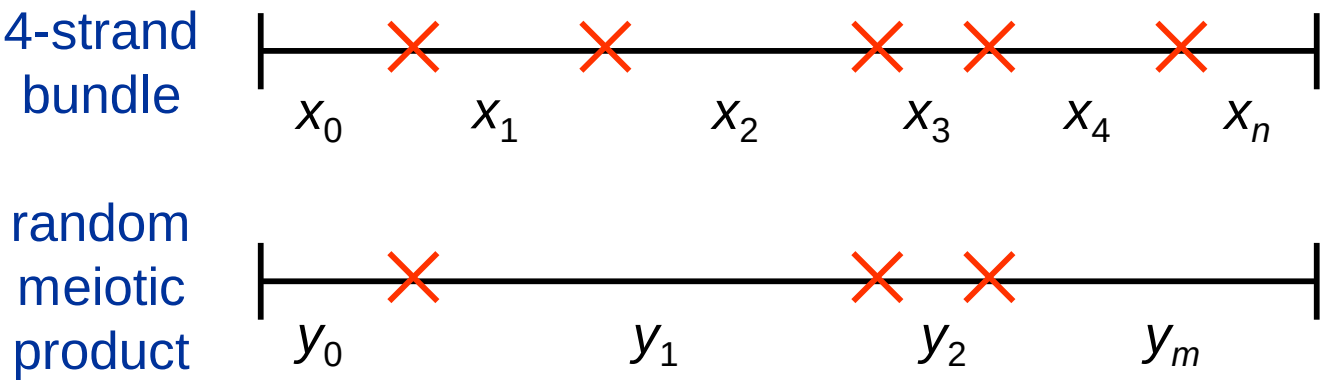
Clean Data







Models



- Count-location model

  $n \sim (p_0, p_1, p_2, \ldots)$

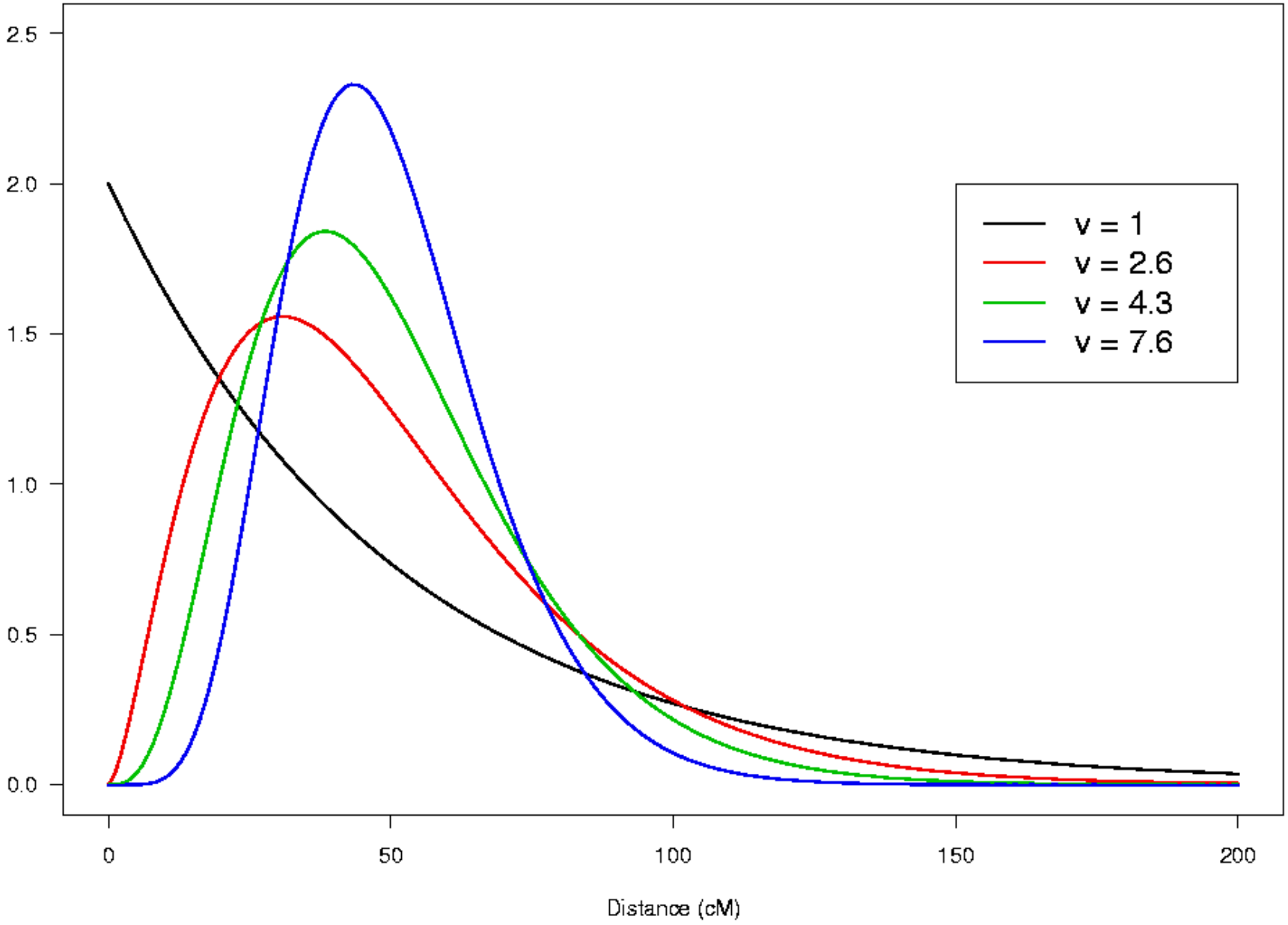
  locations | n ~ iid uniform

- Gamma model

  x_i's ~ stationary gamma renewal process (shape = v, rate = 2v)

  y_i's ~ mixtures of gammas

Dist'n of distance between chiasmata



Model fitting

- Count-location model

m_i = # crossovers
n_i = underlying # chiasmata

$n_i \sim (p_0, p_1, p_2, \ldots)$
$m_i \mid n_i \sim \text{binomial}(n_i, 1/2)$

MLEs via a version of the EM algorithm

Model fitting

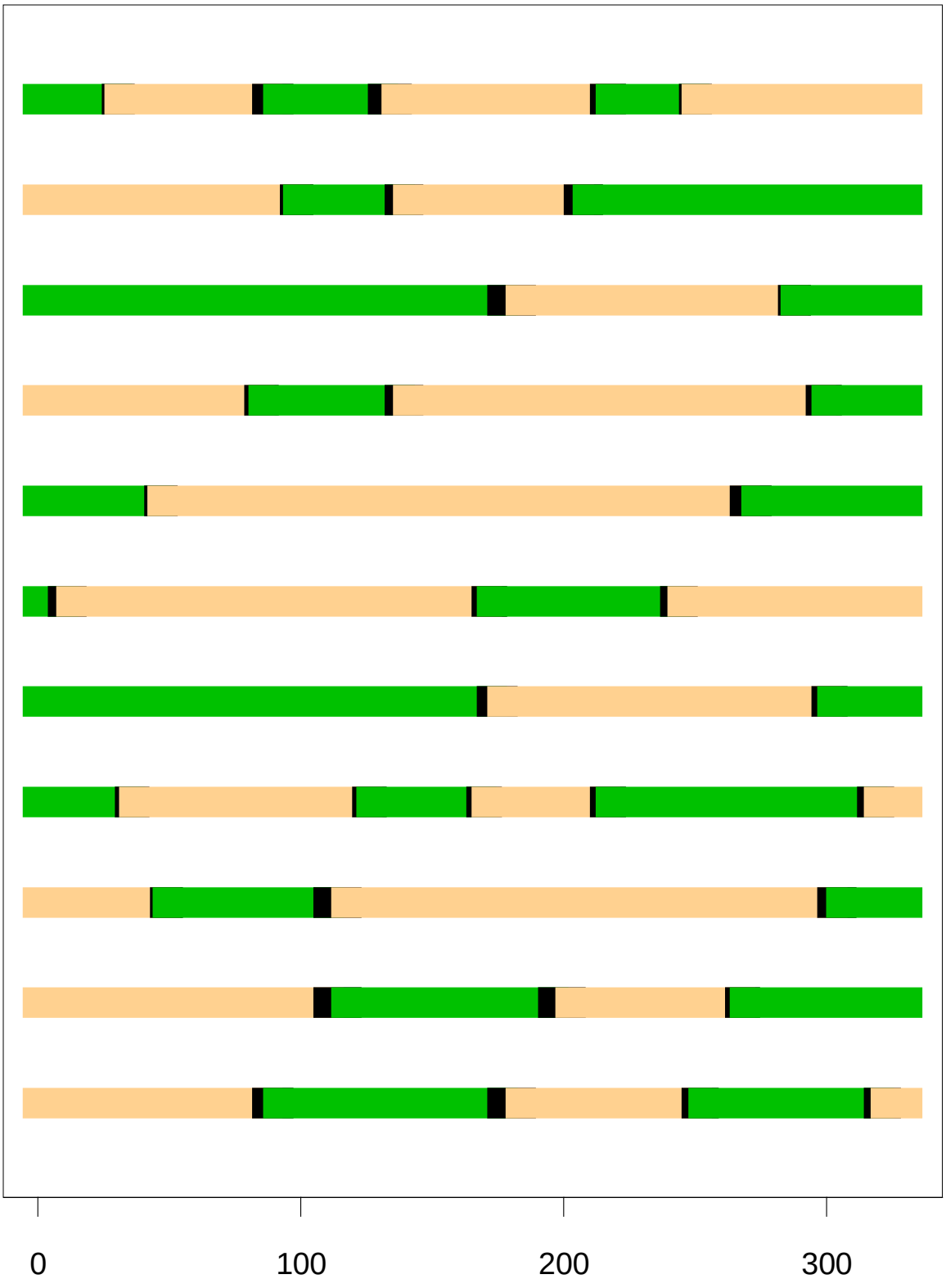
- Gamma model

$x_1, x_2, \ldots \sim f(v, 2v)$
$x_0 \sim g = 2[1 - F(v, 2v)]$
x_i's independent

$y_1, y_2, \ldots \sim \Sigma\, (½)^k\, f(kv, 2v)$
$y_0 \sim ½\, g + \Sigma\, (½)^{(k+1)}\, g * f(kv, 2v)$
y_i's independent

- MLE of v using y_i's
- g calculated numerically
- Convolutions calculated numerically
- Maximization performed using a quasi-Newton method

Another view of the data



Distributions of # XOs / chr

Maternal chromosome 1

	0	1	2	3	4	5	> 5	X^2
Obs.	2	7	12	24	23	14	10	
Pois.	3	9	17	20	17	12	14	**9.2**
C-L	2	7	14	22	23	16	9	0.8
Gamma	1	5	14	23	23	16	10	1.2

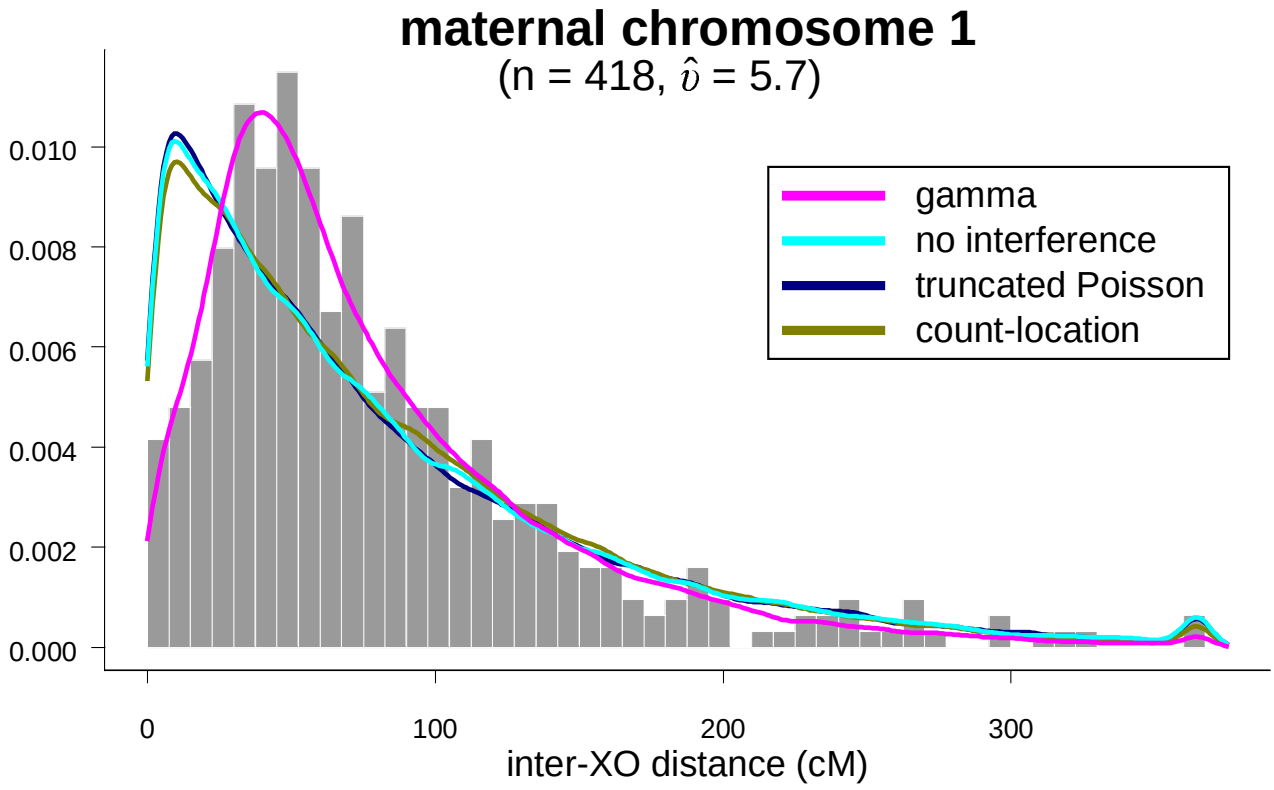
Maternal chromosome 4

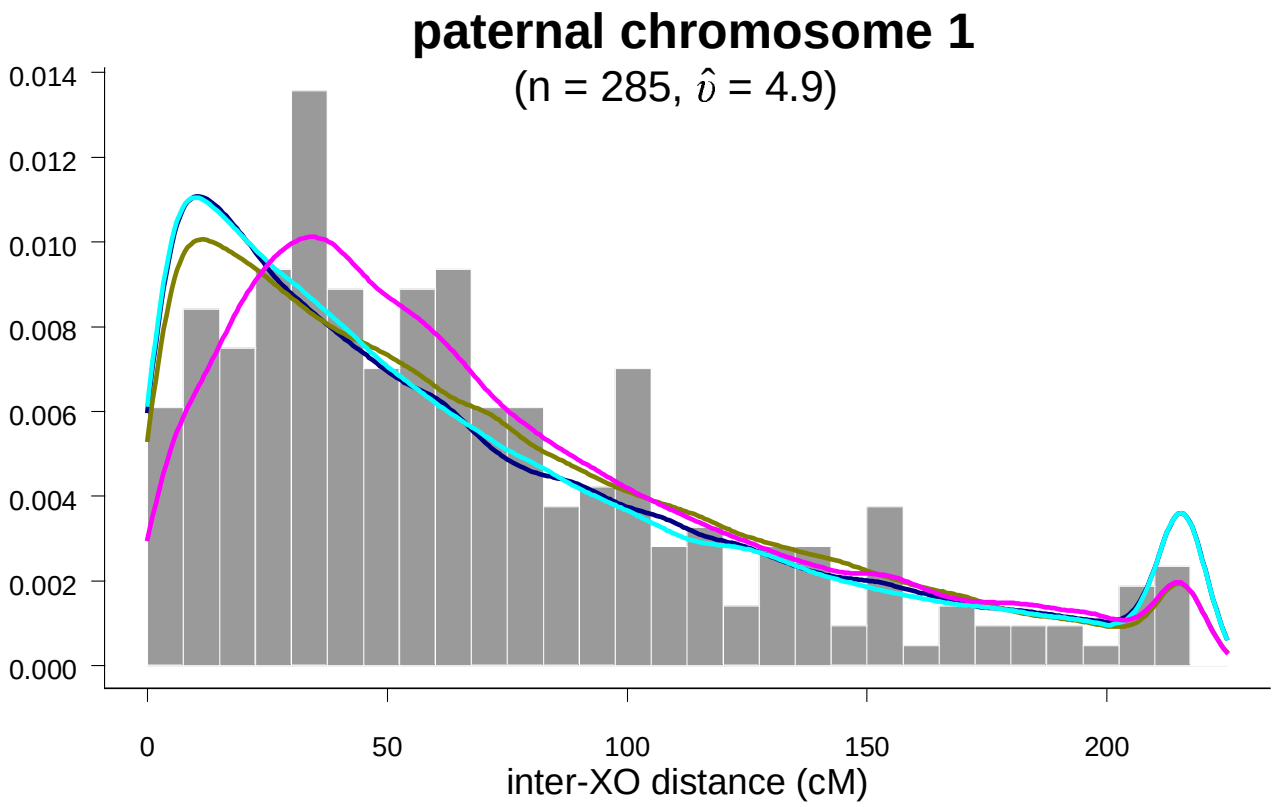
	0	1	2	3	4	5	> 5	X^2
Obs.	1	16	**36**	15	15	9	0	
Pois.	7	18	23	20	13	7	4	**14.4**
C-L	4	16	26	25	15	6	1	**12.8**
Gamma	4	15	26	24	15	6	1	**7.1**

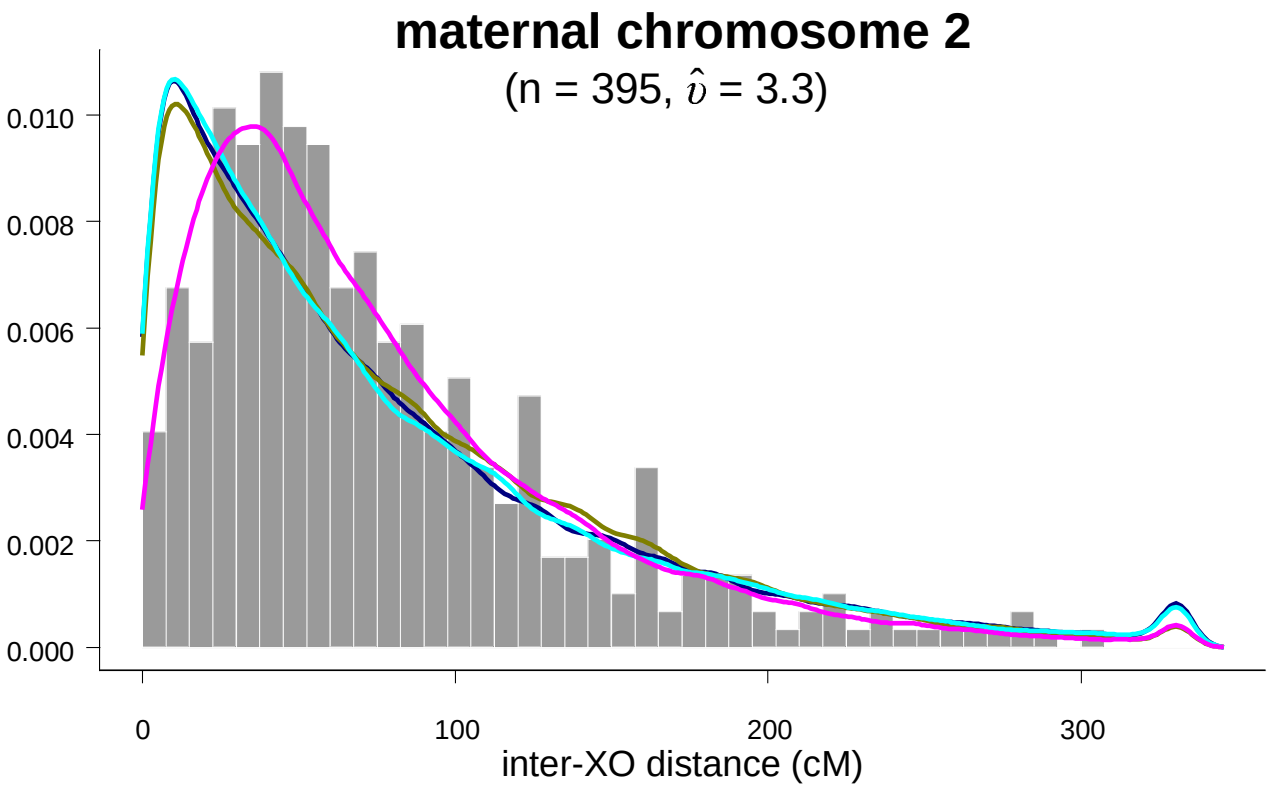
Evidence for interference:

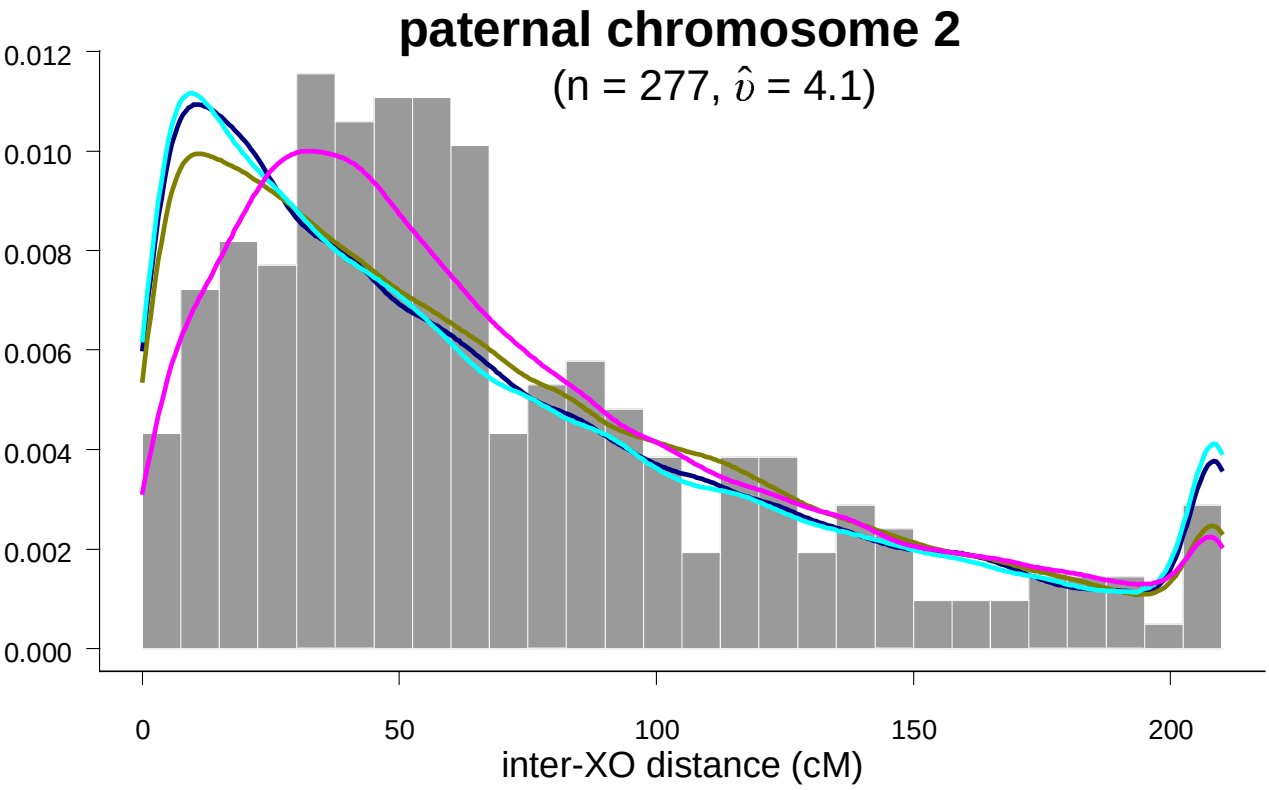
maternal 3, 9, 12, 14, 15, 17

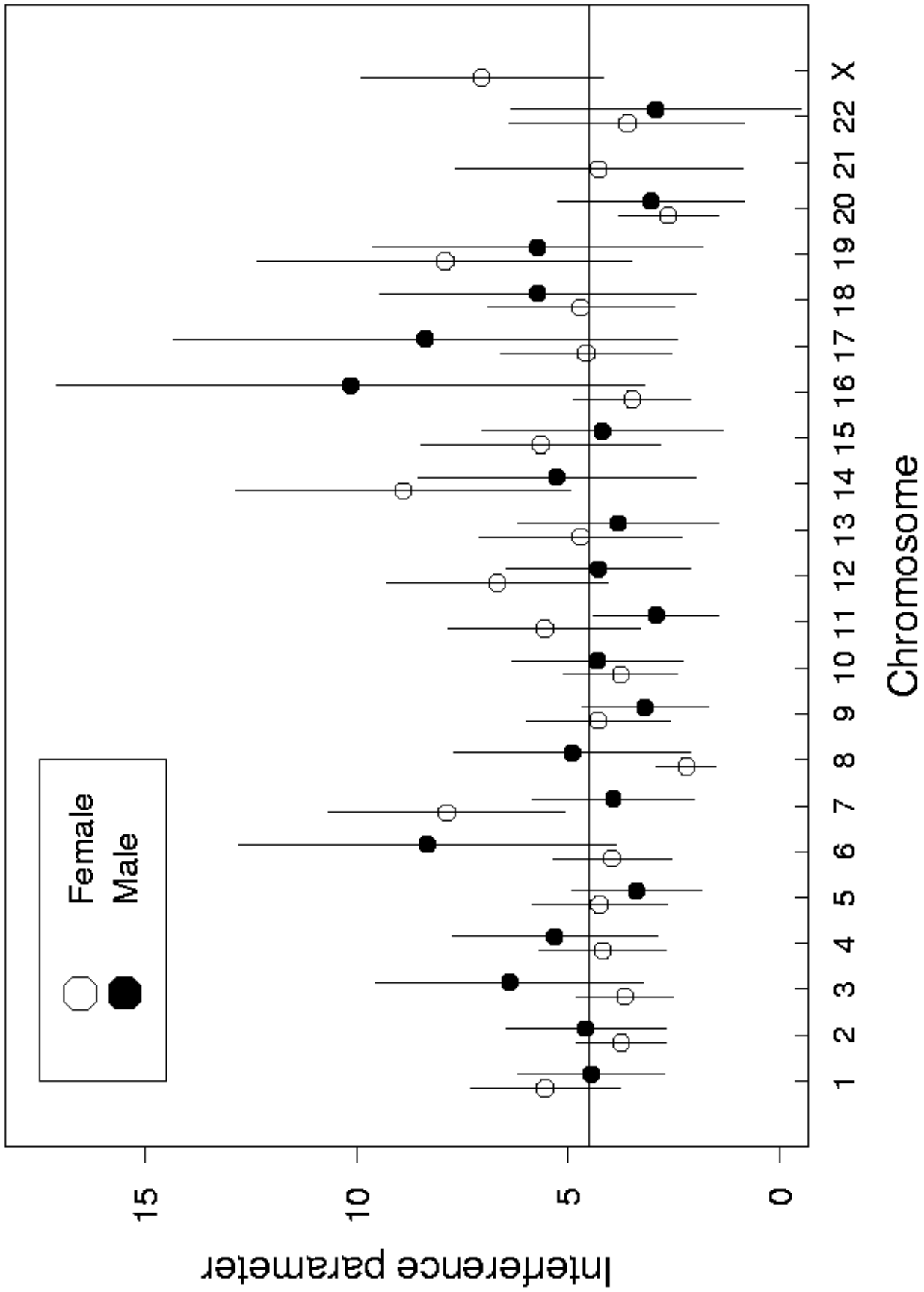
paternal 1, 4, 5, 9, 14


maternal chromosome 1
(n = 418, $\hat{\upsilon}$ = 5.7)
gamma
no interference
truncated Poisson
count-location
0.010
0.008
0.006
0.004
0.002
0.000
0
100
200
300
inter-XO distance (cM)


paternal chromosome 1
(n = 285, $\hat{\upsilon}$ = 4.9)
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
50
100
150
200
inter-XO distance (cM)

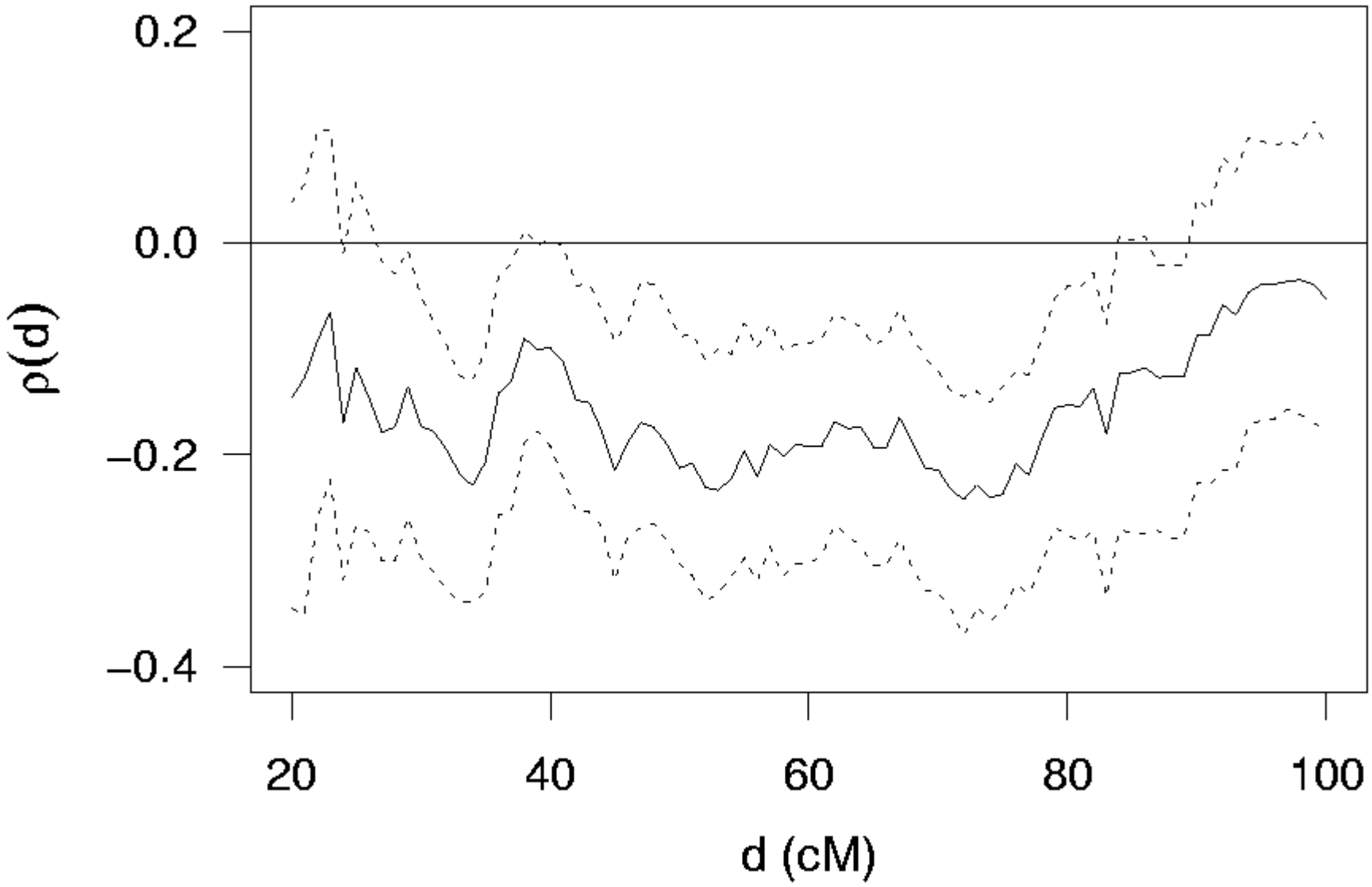

maternal chromosome 2
(n = 395, $\hat{\upsilon}$ = 3.3)
0.010
0.008
0.006
0.004
0.002
0.000
0
100
200
300
inter-XO distance (cM)

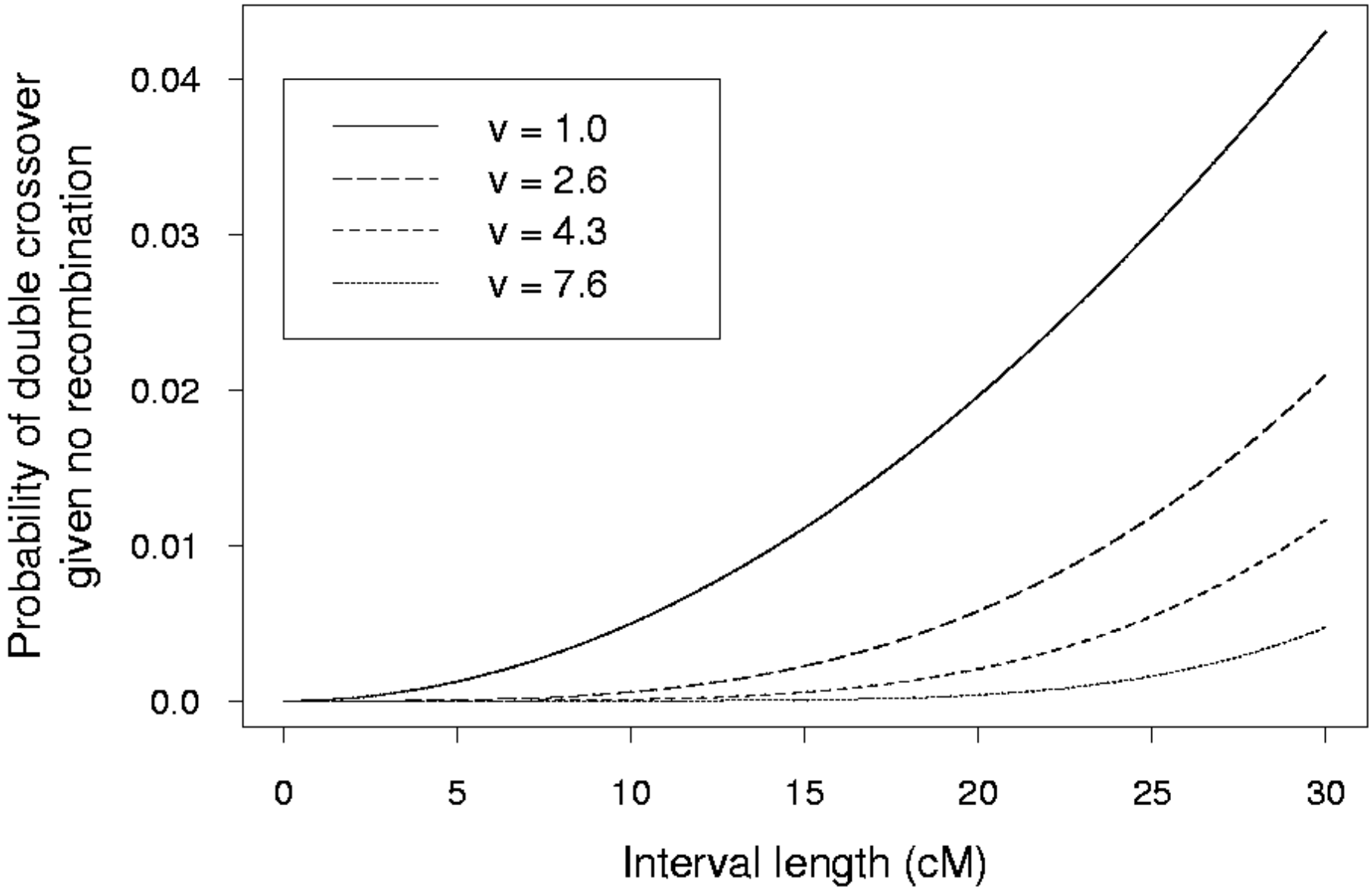

paternal chromosome 2
(n = 277, $\hat{\upsilon}$ = 4.1)
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
50
100
150
200
inter-XO distance (cM)


Female
Male
Interference parameter
0
5
10
15
Chromosome
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X

Fathers
Mothers
Family
102
884
1331
1332
1347
1362
1413
1416
Interference parameter
8
7
6
5
4
3

$$\rho(d) = \mathrm{corr}(x_p, x_q | \max\{x_p, x_q\} \leqslant d, \min\{L_p, L_q\} \geqslant d)$$




Probability of double crossover given no recombination
0.0
0.01
0.02
0.03
0.04
v = 1.0
v = 2.6
v = 4.3
v = 7.6
0
5
10
15
20
25
30
Interval length (cM)

Discussion

- Approximations
  - Correct marker order
  - Correct genetic distances
  - All crossovers observed
  - Interval censoring unimportant
  - No individual variation in recombination
  - Interference constant across chromosome
- Conclusions
  - Gamma model fits well
  - Count-location model fits poorly
  - Gamma parameter, $\nu \approx 4.3$ (stronger than Kosambi, $\nu \approx 2.6$)
  - Little variation between chr or sexes
  - Possible variation between mothers
  - Interference across the centromere