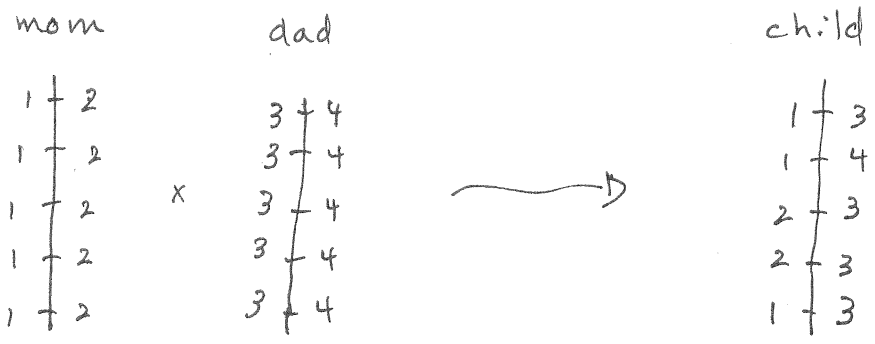
Phase-known data on 2 parents + child



Recombination (across an interval)

- change in grandparental origin of DNA

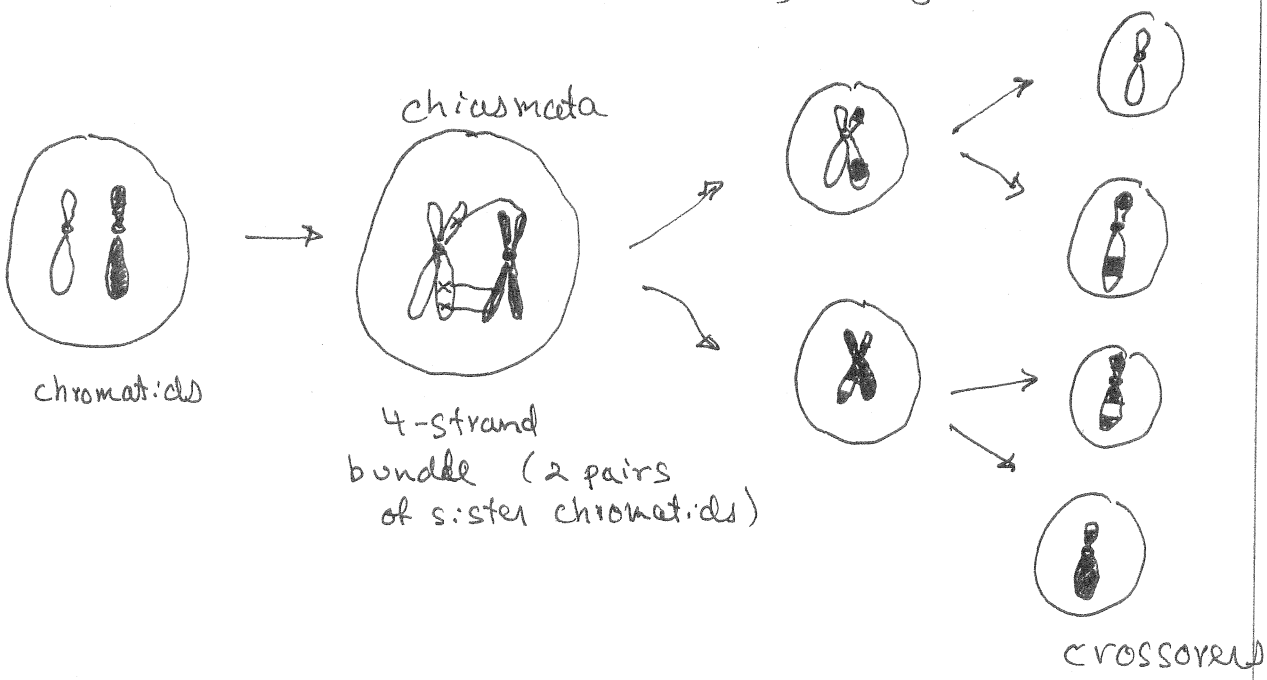
Recombination fraction

- Probability of recombination

Note: to observe recombinations, we need:

- phase information (i.e. haplotypes)
- parent heterozygous at both loci

MEIOSIS : cell division resulting in gametes



CHIASMATA : exchanges between non-sister chromatids (each involves one pair)

CROSSOVERS : points of exchange on a product of meiosis

Note : obligate chiasma on 4-str. bundle for proper chromosome segregation (disjunction)

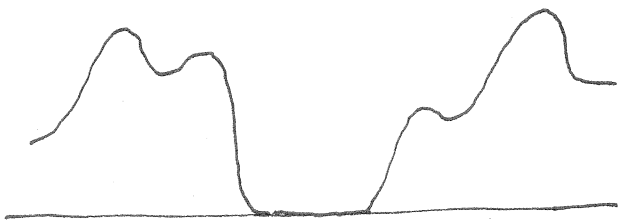
(betw. 2 pts)

GENETIC DISTANCE : average no. crossovers per meiotic product (in the int'l)

d cM = ave. of d crossovers per 100 meiotic products

(ad chiasmata per 100 meioses)

[Generally unobservable]



intensity function (expected no. XOs per unit dist.)

phys. distance

$$\lambda(x) = \lim_{\Delta \to 0} \frac{\Pr\{N(x, x+\Delta) > 0\}}{\Delta}$$

Genetic distance : rescale to have constant intensity

Genetic ⟷ Physical distance

Recombination rate varies by...

A. Organism

human ~ 3000 cM — i.e. 30 M

mouse & rat ~ 1500 cM — 15 M

B. Sex

human female ~ 44 M

human male ~ 27 M

ratio ~ 1.6

C. chromosome

female : male ratio varies by chr.

c15, c19 : ~1.2

c8 : ~2.0

X chr : fem ~180 cM

mal

pseudoautomal regions

pter ~ 20cM;

qter ~ 5cM? (or 0?)

D. position on chr

- telomere shows more recomb'n (esp in males)
- centromere shows very little recomb'n

E Individuals

- sperm typing: differences betw. regions
- total no. XOs: females vary, males don't

Recomb. nation Two pts. have recombined if the gr. pa'l origin of DNA is different

Genetic distance Two pts separated by d cM if ave no. crossovers, per 100 meiotic products, is d

MAP FUNCTIONS

- crossovers generally not observable

OBSERVE: recombination across interval $\equiv$ odd no. XOs in interval

Rec. frac $=$ Pr(recombination) $=$ Pr(odd no. XOs)

Map function : connect rec. frac. with genet. dist.

$$r = M(d) = Pr\{ \text{odd no. XOs} \mid \text{ave no. XOs} = d \}$$

[we generally use $d = M^{-1}(r)$]

Examples : Morgan $M(d) = d$ (generally true for small d)

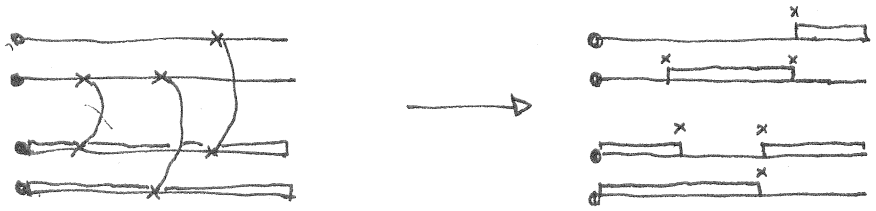
(here, d in Morgans) Haldane $M(d) = \frac{1}{2}\{1 - e^{-2d}\}$

Kosambi $M(d) = \frac{1}{2}\tanh(2d)$

$$= \frac{1}{2}\left\{\frac{e^{4d}-1}{e^{4d}+1}\right\}$$

$\Longrightarrow$ leads us to seek MODELS of the recombination process

INTERFERENCE



chiasmata on 4-str. bundle

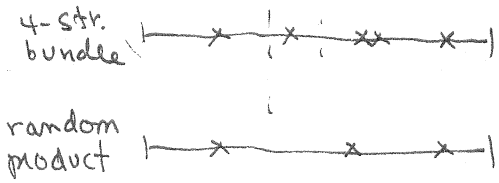
crossovers on meiotic products

A. CHROMATID INTERF. : choice of strands
- little evidence (need tetrad data)

B. CROSSOVER (chiasma) INTERF. : positions / spacing
- strong evidence for POSITIVE interference
(X's & XOs more evenly spaced than "random")

NO CHROMATID INTERFERENCE (NCI)

each chiasma involved random choice of non-sister chromatids, equally likely among 4 poss. pairs, independent from chiasma to chiasma



- left with pt. process model of chiasma locations on 4-str. bundle
- Implied XO process on random meiotic product
- Many ways to define (construct) pt. process models; we won't be too rigorous.

MATHER'S FORMULA

Assuming NCI, the map function must be

$$r = M(d) = \tfrac{1}{2}\{1 - Pr(\text{no chiasma in interval})\}$$

on 4-str. bundle

r = rec. frac. for interval

d = genetic length of interval (in Morgans)
= ave no. XOs per meiotic product

→ constant confusion betw. chiasmata & XOs
– be careful of factors of 2.

Proof Let $r_i = Pr(\text{recombination} \mid i \text{ chiasmata in int'l})$

$r_0 = 0$

$r_1 = 1/2$

$r_n = \frac{1}{2} r_{n-1} + \frac{1}{2}(1 - r_{n-1})$ for $n > 0$ ‖ ← keypt.

$= \frac{1}{2}$

$$\Rightarrow r = \sum_i r_i p_i \quad \text{where } p_i = Pr(i \text{ chiasmata in interval})$$

$$= \tfrac{1}{2} \sum_{i>0} p_i$$

$$= \tfrac{1}{2}(1 - p_0)$$

POINT PROCESS MODELS

A. POISSON PROCESS ("NO INTERFERENCE")

chiasma process : rate 2/M (0.02/cM)

XO process : rate 1/M (0.01/cM)

i. For any disjoint subintervals

$I_1, I_2, \ldots, I_n$ of lengths $d_1, \ldots, d_n$ (in M)

N_{I_k} = # chiasmata in I_k

$\rightsquigarrow (N_{I_1}, \ldots, N_{I_n})$ ~ independent

$N_{I_k} \sim$ Poisson $(2d_k)$

ii. distances betw. chiasmata are i.i.d exponential (rate = 1/2M)

Map function : Haldane $M(d) = \frac{1}{2}(1 - e^{-2d})$

"memoryless" property $Pr(X_i > s+t \mid X_i > s) = Pr(X_i > t)$

Note: under NCI, the locations of XOs on a random meiotic product will also follow a Poisson process.

B. COUNT-LOCATION MODEL

n = no. chiasmata on 4-strand bundle

$n \sim (p_0, p_1, p_2, \dots)$ $\qquad L = \sum_i p_i / 2$

Given n, locations of chiasmata are iid

(w.l.o.g., uniform)

assuming NCI

↘ crossover process: (also a C-L model)

m = no. XOs

$m|n \sim \text{Binomial}(n, \frac{1}{2})$

$$P_r(m=i) = \sum_{n=i}^{\infty} \binom{n}{i} \left(\tfrac{1}{2}\right)^n p_n$$

special cases

1. obligate chiasma C-L model

$$p_0 = 0$$

2. Poisson (i.e., NI)

$$p_n = \frac{e^{-\lambda}\lambda^n}{n!} \qquad \text{where } \lambda = 2L$$

3. truncated Poisson

$$p_0 = 0$$

$$p_n = \frac{e^{-\lambda}\lambda^n}{n!\,(1-e^{-\lambda})} \qquad \text{for } n > 0$$

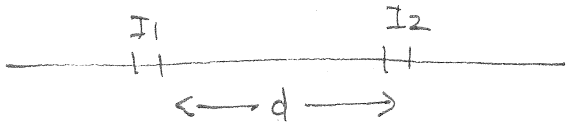
C-L model (continued)

Jung Ott's split of crossover interference

- numerical interference
  - dist'n of no. XOs differs from Poisson (or truncated Poisson)
- positional interference

[This is meaningful only in the context of the C-L model]

coincidence : classical measure of interference



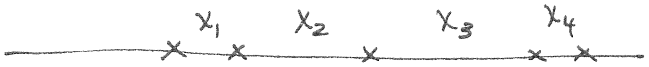
$$S_4 = \frac{\Pr(\text{rec. in } I_1 \text{ and } I_2)}{\Pr(\text{rec. in } I_1)\ \Pr(\text{rec. in } I_2)}$$

$\left\{ \begin{array}{l} = \\ < \\ > \end{array} \right.$ | implies | no interf. / pos'ive interf. / neg've interf.

For the C-L model, S_4 is independent of the distance betw. the intervals → BAD!

C. Stationary Gamma Renewal Process

Renewal process



inter-arrival distances (X_i) are iid with density f.

f = exponential $\rightsquigarrow$ Poisson process

Gamma distribution (shape $= \nu$, rate $= \lambda$; $\nu, \lambda > 0$)

$$f(x; \nu, \lambda) = \frac{\lambda^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\lambda x} \quad \text{for } x > 0$$

where $\Gamma(\nu) = \int_0^\infty e^{-x} x^{\nu-1} dx$

$$\text{mean} = \nu/\lambda \qquad \text{var} = \nu/\lambda^2$$

$$\text{CV} = \frac{\text{SD}}{\text{mean}} = 1/\sqrt{\nu}$$

Special cases

$\nu = 1$: exponential (λ)

$\nu = k/2$, $\lambda = 1/2$: $\chi^2(k)$

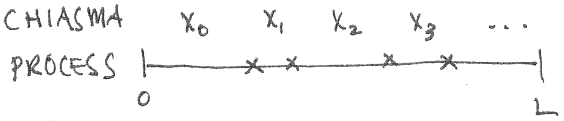
$\nu = m+1$, $\lambda = \frac{m+1}{2}$: scaled version of $\chi^2(2m+1)$

GAMMA MODEL

Locations of chiasmata on 4-str. bundle according to stationary gamma renewal process

(indep. increments gamma (shape $= \nu$, rate $= 2\nu$) so that mean $= 1/2$ (in morgans))

$\nu \begin{cases} = \\ > \\ < \end{cases} 1$ — no interf. / pos're interf. / neg're interf.



X_i indep.

$X_1, X_2, \ldots$ iid $f(x; \nu, 2\nu)$

$X_0 \sim g = 2(1-F)$ (cdf of f)

(assuming NCI

again, stationary renewal process

$Y_1, Y_2, \ldots$ iid $f^* = \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k f_k$

where $f_k = f(\cdot; k\nu, 2\nu)$

$Y_0 \sim g^* = 1 - F^*$

FACT : $X_1, X_2, \ldots, X_n$ independent w/

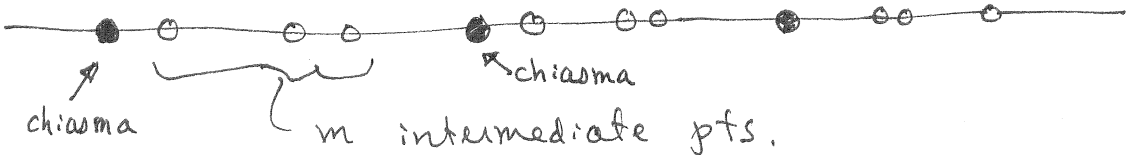
$X_i \sim$ Gamma (ν_i, λ)

Then $\sum_{i=1}^{n} X_i \sim$ Gamma $(\Sigma \nu_i, \lambda)$

χ^2 model

Special case of gamma model where

$\nu = m+1$ for a nonneg.'ve integer, m.



- locations according to a Poisson process
- every $(m+1)$ st point is a chiasma

Poisson-skip model : take m to be random

Map functions and renewal processes

[stat'y renewal process for chiasma locations
+ NCI

renewal density = f

map function = M

$$\boxed{f = -M''}$$

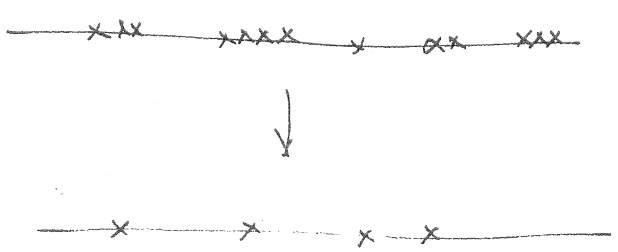
gamma model $\nu = 2.6$ $\approx$ Kosambi

$\nu = 7.2$ $\approx$ Carter-Falconer

Humans : $\hat{\nu} = 4.3$

Mouse : $\hat{\nu} = 11.3$

Superposition of point processes

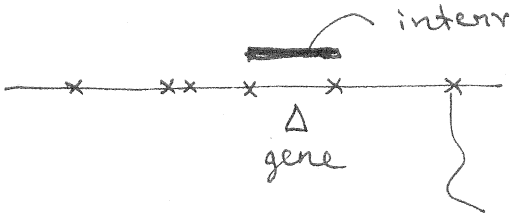


throw down (overlay) multiple pt. processes

rescale to get constant intensity

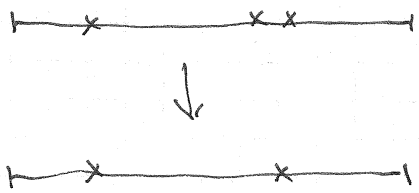
In the limit : a Poisson process

APPLICATION Fine mapping of mendelian diseases



points generally from different point processes (i.e., different people)

- Want a model for the XO process
- X on 4-str. bundle; NCI: thin X process indep'ly w/ prob 1/2



Poisson process model:

X acc. to Poisson process w/ rate $\frac{1}{2}$ per Morgan

$\Rightarrow$ XO also Poisson process w/ rate 1 per Morgan

I. Disjoint intervals $I_1, I_2, \ldots, I_k$

N_i = # X in interval I_i

N_i indep., $N_i \sim$ Poisson $(2|I_i|)$

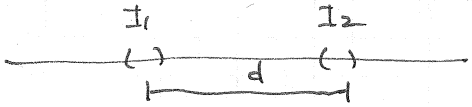
II. distances betw. X are iid exponential $(\frac{1}{2})$

Count-loc'n model

n = no. X on 4-str. bundle $\sim (p_0, p_1, p_2, \ldots)$

locations $X_1, X_2, \ldots$

$X_1, \ldots, X_n \mid n \sim$ iid Uniform $(0, L)$



$$S_4 = \frac{\Pr(X \text{ in both } I_1 \text{ and } I_2)}{\Pr(X \text{ in } I_1) \times \Pr(X \text{ in } I_2)}$$

Under C-L model, S_4 indep't of d (distance betw. intervals)

Count-location model (cont.)

XO process also a C-L model (if we assume NCI)

m = no. XO on random meiotic product

$m/n \sim$ Binomial $(n, 1/2)$

$m \sim (q_0, q_1, \dots)$

where $q_i = \sum_{n=i}^{\infty} p_i \binom{n}{i} \left(\frac{1}{2}\right)^n$

Gamma model : fits data well

Assume χ locations acc'g to gamma renewal model (inter-χ distances iid gamma)

- constrain ave inter-χ distance to be $\frac{1}{2}$ Morgan
- shape parameter ν controls interference

$\nu = 1$ no interference

$\nu > 1$ pos've interference

Assuming NCI, XO locns acc'g to renewal model with inter-XO distance dist'n being a mixture of gammas

Special case: "χ^2 model"

$\nu = m + 1$ where m is a non-neg've integer

[Simpler to work with]