A MODEL FOR ADDITIVE CROSS-OVER UNITS IN GENETIC LINKAGE ANALYSIS

BU-164-M D. S. Robson March, 1964

Abstract

If the number of cross-overs occurring between two loci is a Poisson random variable with mean \( \lambda \) then the probability of an odd number of cross-overs is

\[ p = \frac{1-e^{-2\lambda}}{2}. \]

Cross-over units measured on the scale of

\[ \lambda = -\frac{\log(1-2p)}{2} \]

rather than on the p-scale would therefore be additive.

Biometrics Unit, Plant Breeding Department, Cornell University
A parent carrying the gene arrangement \( \frac{AB}{ab} \) on a particular chromosome will transmit \( AB \) and \( ab \) gametes with equal probability, say \((1-p)/2\), while gametes of type \( Ab \) and \( aB \) will each be transmitted with probability \( p/2 \).

The fraction \( p \) therefore denotes the probability that in the process of gamete production an odd number of chromosomal cross-over will occur between these two loci. It is an empirical fact that as the distance between loci increases, the probability \( p \) approaches \( 1/2 \), in which case the two loci effectively segregate independently.

The fraction \( p \) is employed in linkage analysis as a measure of the "cross-over distance" between loci; thus, if gametes in the repulsion phase (Ab or aB) are produced in \( 100p_{AB} \) percent of the cases then the A and B loci are said to lie \( 100p_{AB} \) cross-over units apart. If a third locus appears with the configuration \( \frac{ABC}{abc} \), however, then \( p_{Ac} \neq p_{AB} + p_{Bc} \), and cross-over units so defined are not additive.

If the number of cross-overs occurring between two loci is a Poisson random variable with mean \( \lambda \) then \( p \), the probability of an odd number of cross-overs, is

\[
p = \frac{1-e^{-2\lambda}}{2}.
\]

This result may be obtained by noting from the relation

\[
p = e^{-\lambda} \left( \lambda + \frac{\lambda^3}{3!} + \frac{\lambda^5}{5!} + \frac{\lambda^7}{7!} + \cdots \right)
\]
that as a function of \( \lambda \), \( p \) satisfies the differential equation

\[
\frac{dp}{d\lambda} = e^{-\lambda} \left( 1 + \frac{\lambda^2}{2!} + \frac{\lambda^4}{4!} + \cdots \right) - e^{-\lambda} \left( \lambda + \frac{\lambda^3}{3!} + \frac{\lambda^5}{5!} + \cdots \right)
\]

or

\[
\frac{dp}{d\lambda} = 1 - 2p
\]

The boundary condition that \( p = 0 \) when \( \lambda = 0 \) then determines the solution indicated above. As \( \lambda \) increases, \( p \) rapidly approaches the limiting value of \( 1/2 \); for an average of \( \lambda = 1 \) cross-over between two loci the value of \( p \) is already .432, and if \( \lambda = 2 \) then \( p = .491 \).

If, as would be implied by the Poisson model, cross-overs between the A- and B-locus occur independently of those between the B- and C-locus in the configuration \( \frac{ABC}{abc} \) then the corresponding Poisson parameters \( \lambda_{AB} \) and \( \lambda_{BC} \) are additive; that is, \( \lambda_{AC} = \lambda_{AB} + \lambda_{BC} \). For chromosomal mapping purposes, therefore, the transformation from \( p \) to

\[
\lambda = \log_2 \frac{1}{\sqrt{1-2p}} = -\frac{1}{2} \log_2 (1-2p)
\]

would provide a much more convenient scale of "cross-over distance" and would have the further advantage of being simply interpreted as the mean number of cross-overs.

As is characteristic of classical linkage analysis, statistical estimation of \( \lambda \) (or \( p \)) would approach the impossible as \( p \) approaches 1/2. Since the estimate of \( p \) is a binomial variable, however, then for large samples the
estimate of $\lambda$ would be approximately log-normally distributed with mean $\lambda$ and variance $16p(1-p)/n(1-2p)^2$. For closely linked loci this variance would become simply $16p/n$.

A check on the validity of this model can be readily made by checking the additivity of $\lambda$ from available data. Sample size, however, must be truly large in order to detect departures from this model. Three examples (1) involving triple heterozygotes have been checked.

<table>
<thead>
<tr>
<th>Gamete numbers</th>
<th>ABC or abc</th>
<th>Abc or aBC</th>
<th>ABC or aBC</th>
<th>AbC or aBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>5246</td>
<td>229</td>
<td>1227</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3344</td>
<td>245</td>
<td>1504</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>4287</td>
<td>721</td>
<td>2403</td>
<td>235</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates of $\lambda$</th>
<th>$\hat{\lambda}<em>{AB} \pm 2\hat{\sigma}</em>{\lambda_{AB}}$</th>
<th>$\hat{\lambda}<em>{BC} \pm 2\hat{\sigma}</em>{\lambda_{BC}}$</th>
<th>$\hat{\lambda}<em>{AC} \pm 2\hat{\sigma}</em>{\lambda_{AC}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.053 ± .025</td>
<td>.221 ± .071</td>
<td>.275 ± .072</td>
<td></td>
</tr>
<tr>
<td>.069 ± .031</td>
<td>.477 ± .133</td>
<td>.562 ± .162</td>
<td></td>
</tr>
<tr>
<td>.144 ± .040</td>
<td>.586 ± .140</td>
<td>.849 ± .246</td>
<td></td>
</tr>
</tbody>
</table>

none of which contradict the model $\lambda_{AB} + \lambda_{BC} = \lambda_{AC}$.

Reference