There are two characterizations of the Poisson “parameter”: the rate $\lambda$ or the mean $\mu$.

The probability density function $P(X = k) = e^{-\mu} \frac{\mu^k}{k!}$ requires $\mu$.

In 4.2, $\lambda$ is a rate, with dimensions inverse time, so it is necessary to compute the mean $\mu = \lambda \cdot t$ where $t$ is the amount of time; this is analogous to computing $\mu = \lambda \cdot d$ for distance and crossovers.

The scenario described in 4.2 is possible for any value of $\lambda > 0$, and the problem does not give enough information to solve for a specific value of $\lambda$, so leave $\lambda$ as a variable.

In 5.2, give an interpretation.

Use Matlab or R to make the table in H-202(a) and the graph in H-203(d). While these can be done by hand or calculator, you should get experience using Matlab or R on easier problems like these now, since some problems on later homeworks will be infeasible to do by hand.

**Problem H-201.** This problem concerns Haldane’s formulas $d = -\frac{1}{2} \ln(1 - 2r)$ and $r = \frac{1}{2} (1 - \exp(-2d))$ when $d$ is in Morgans and $r$ is expressed as a number between 0 and 1/2; conversions are necessary when $d$ is in centi-Morgans or $r$ is expressed as a percent.

(a) What is the maximum possible recombination rate for genes on the same chromosome? How does this compare to the constant recombination rate of 1/2 for genes on separate chromosomes?

(b) A portion of mouse chromosome 1 is shown on the last page of this assignment, using coordinates in cM (centi-Morgans). What is the distance and the recombination rate (to at least 3 significant figures) between $D1Hun27$ and $Cops5$?

(c) Compute the distance between two genes if the recombination rate is 35%. Compute the recombination rate between two genes if their distance is 125 cM. Comment on how these computations compare to the similarity or dissimilarity between recombination rate and distance in part (b).

(d) We derived Haldane’s formula for recombination in class based on certain independence assumptions. Unfortunately, recombination rates are not additive. The formula for $d$ in terms of $r$ was developed to transform recombination rates into a different scale that is additive. Assuming genes $A$, $B$, and $C$ are in that order, show that $r_{AC} = r_{AB} + r_{BC} - 2r_{AB}r_{BC}$ can be rewritten as $1 - 2r_{AC} = (1 - 2r_{AB})(1 - 2r_{BC})$. Then show how to convert this to $d_{AC} = d_{AB} + d_{BC}$.

**Problem H-202.** Assume crossovers are a Poisson process with rate $\lambda = 1 \text{ M}^{-1} = .01 \text{ cM}^{-1}$, giving mean $\mu = \lambda d$.

(a) Two genes are on the same chromosome, 240 cM apart. Make a table of the probabilities of $k$ crossovers occurring in-between them for $k = -1, 0, 1, 2, 3, 4, 5, 6$. Also give a column with the cumulative distribution function at these same values (i.e., probability that $\#$ crossovers is $\leq k$).

(b) Again, two genes are on the same chromosome, 240 cM apart. What is the expected number of crossovers in-between them, and what’s the standard deviation of that? What is the probability that the number of crossovers occurring will exactly equal the expected number?
Problem H-203. During meiosis, two genes on the same chromosome are recombined if an odd number of crossovers occurs in-between them. Suppose that two genes are located a distance \(d\) apart, where \(d\) is measured in Morgans.

(a) Show that the probability that the number of crossovers is odd equals \(e^{-d} \sum_{k=0}^{\infty} \frac{d^{2k+1}}{(2k+1)!} \).

(b) Show that \(\sum_{k=0}^{\infty} \frac{x^{2k+1}}{(2k+1)!} = \frac{e^x - e^{-x}}{2}\) (which is called \(\sinh(x)\)) by using the Taylor series

\[
e^x = \sum_{m=0}^{\infty} \frac{x^m}{m!}
\]

(c) Combine (a) and (b) to derive the formula given in problem H-201: when the distance \(d\) between \(A\) and \(B\) is measured in Morgans, the recombination rate between \(A\) and \(B\) is \(r = \frac{1}{2}(1 - \exp(-2d))\).

(d) When \(d\) is “small,” the distance \(d\) in Morgans is approximately the same as the recombination rate as a fraction. Now we will see why.

(i) Expand \(r\) as a Taylor series in \(d\) about \(d = 0\) to at least three terms.

(ii) The first nonzero term in the Taylor series is \(d\). Plot \(r_1 = d\) and \(r = (1 - \exp(-2d))/2\) as functions of \(d\) on the same graph.

(iii) Examine the plots and give approximate values of \(d\) where \(r_1\) and \(r\) are almost equal.

Problem H-204. Suppose that we are able to count the number of crossovers between two genes \(A\) and \(B\) occurring in the formation of independent gametes. Over the sample examined, we obtain the following results. Estimate how far apart \(A\) and \(B\) are in cM.

<table>
<thead>
<tr>
<th>(k)</th>
<th># gametes with (k) crossovers between (A) and (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>&gt; 4</td>
<td>0</td>
</tr>
</tbody>
</table>

Problem H-205. We want to sequence a BAC of size 250 kbp with reads of constant size 200 bp. (a) Use the Lander-Waterman estimates in Chapter 5.1.1–5.1.2 to estimate the number of reads required for 90\% of the genome to be sequenced. For this scenario, also estimate (b) the mean number of reads per contig; (c) the mean contig size; and (d) the mean number of contigs.
This map was obtained from the Mouse Genome Database (MGD), Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, Maine. Jan. 19, 2004.
http://www.informatics.jax.org/searches/linkmap.cgi