

GENOME 560, Spring 2012

Problem Set #2

(Due May 17th 9:00am)

1. [20 points] Inter-species Cross

Let's say that we are interested in functionally characterizing offsprings between inter-species cross. The Saccharomyces Genome Resequencing Project (SGRP) has completely resequenced 37 strains of the closely related species *S.cerevisiae* (25 natural isolate, 7 laboratory strains, and 5 clinical isolates) and 27 strains of the closely related species *S.paradoxus* (22 natural isolates and 5 laboratory strains).

- If we want to perform all pair-wise crosses between the SGRP *S.cerevisiae* and *S.paradoxus* strains, how many matings should we do?
- Let's assume that none of the crosses performed above worked so instead, we decide to perform all pair-wise matings within each species. How many matings should we now do for *S.cerevisiae* and *S.paradocus*?
- More bad news. In setting up the cross in part (b), we somehow managed to switch the labels on the plates for the *S.cerevisiae* strains and we have no idea which one is which. What we're really interested in are cross between the natural and clinical isolates. Rather than doing the obvious and reordering the strains, we decided to gamble and randomly pick two *S.cerevisiae* strains to cross. What is the probability that we guess correctly and mate a natural and clinical isolate?

2. [10 points] Hypergeometric Distribution

Say that in a gene expression study, we identified 100 differentially expressed genes. We performed a Gene Ontology analysis and find that 15 out of the 100 are annotated as "sensory perception of sound". Our microarray contains 10,000 genes in total, of which 1,000 are annotated as "sensory perception of sound". What is the probability that we would observe 15 or more genes out of 100 annotated with this term by chance?

3. [20 points] Wright-Fisher Model

In last Thursday's lecture, we discussed the Wright-Fisher model of binomial sampling alleles from one generation to the next (slide 23 in lecture 2). Assume a population size of $2N = 100$. In the current generation, there are $i = 10$ copies of the A allele ($p_A = 0.1$). The number of A alleles in the next generation is j . Calculate the following probabilities:

- (a) $P(j = 10)$
- (b) $P(j = 0)$
- (c) $P(j \geq 10)$
- (d) $P(5 \leq j \leq 15)$

4. **[15 points] Bayes' Rule**

Only 1 in 1,000 adults is afflicted with a rare disease for which a diagnostic test has been developed. The test is such that, when an individual actually has a disease, a positive result will occur 99% of the time, while an individual without the disease will show a positive test result only 2% of the time. If a randomly selected individual is tested and the result is positive, what is the probability that the individual has the disease?

5. **[30 points] Expression QTLs**

There has been considerable recent interest in mapping loci that influence inter-individual variation in gene expression levels. In these experiments, gene expression levels are treated as a quantitative trait and linkage analysis is performed to find positions in the genome that contribute to variation in transcript abundance.

Let's say that you have performed a linkage analysis experiment to map gene expression quantitative trait loci (eQTL). In total, you have detected significant linkage for 1,013 expression traits. Next, you want to test for the presence of linkage "hotspots", which are regions in the genome showing linkage to multiple gene expression traits. To detect "hotspots", you divide the genome into 579 bins of equal size. Define the random variable X_i to be the number of linkages observed in the i^{th} bin.

- (a) What distribution does X follow? Briefly explain your choice.
- (b) Assuming that eQTLs are randomly distributed, what is the probability that a bin contains no eQTLs.
- (c) What is the probability that a bin contains 40 or more eQTLs?
- (d) Find the number of eQTL in a bin x , such that $P(X \geq x) = 0.05$.

6. **[5 points] Easy Question**

Among the set of all families with two children, a family is selected at random and found to have a girl. What is the probability that the other child is a girl? Briefly explain your result.