Chapter 5 – Inferring Population History and Demography

Inferring Demography Using Summary Statistics

Coalescence Simulations and Confidence Intervals

Estimating Evolutionary Trees

Gene Trees Versus Species Trees

Interpreting Estimated Trees from Population Genetic Data

Likelihood and the Felsenstein Equation

MCMC and Bayesian Methods

The Effect of Recombination

Population Assignment, Clustering, and Admixture

Sequence Divergence

❖ simple genetic distance, d = the proportion of sites that differ between two aligned, homologous sequences

❖ given a constant mutation/substitution rate, d should provide a measure of time since divergence
  ✧ but this is greatly complicated by **multiple hits** (homoplasy)

❖ given that there are **not** an infinite number of sites in a sequence, how is d expected to change with time?

consider two recently diverged sequences…

ACGTACGTACGTACGTACGTACGTACGT
ACGTACGTACGTACGTACGTACGTACGT

consider two recently diverged sequences…

ACGTACGTACGTACGTACGTACGTACGT
ACGTACGTACGTACGTAAGTACGTACGT

what is the chance that the next substitution obscures the first?

consider two recently diverged sequences...
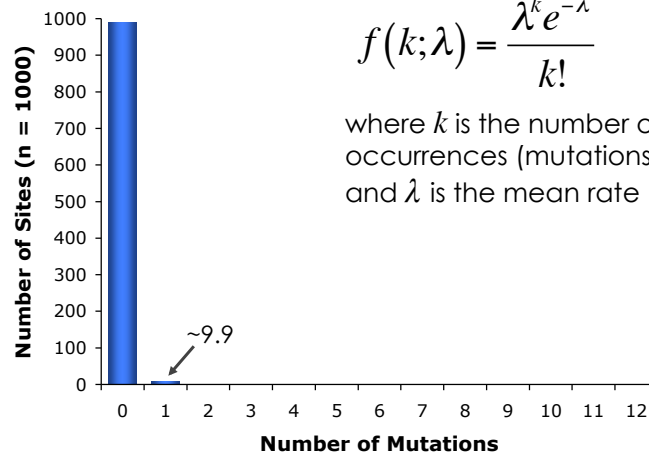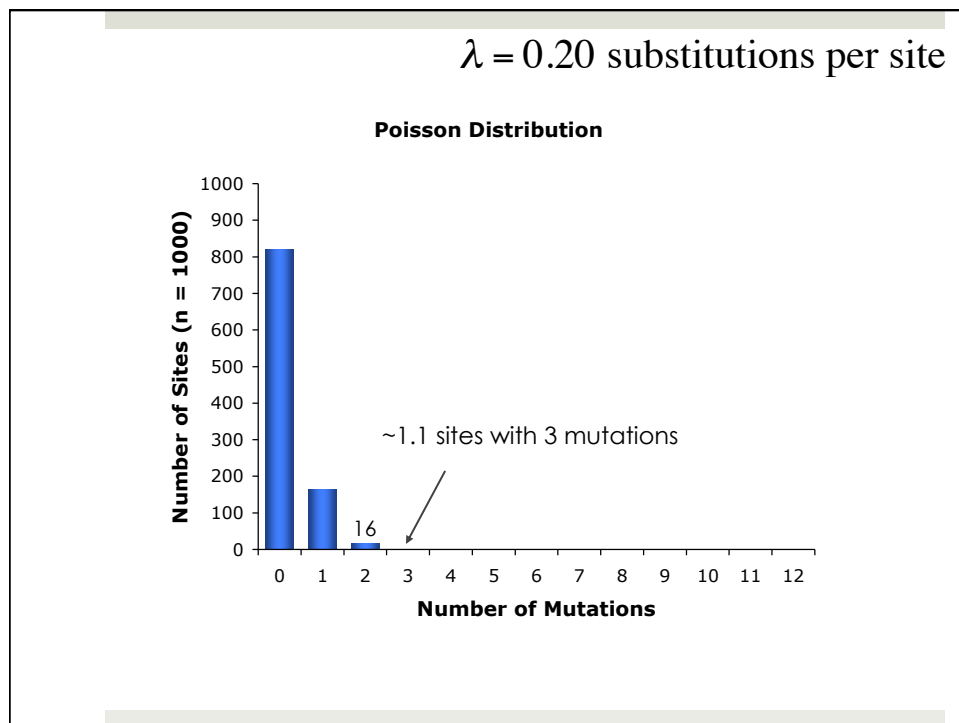
ACGTACGC**C**ACGTACGTACGTACGTACGT
ACGTACGTACGTACGTA**A**GTACGTACGT

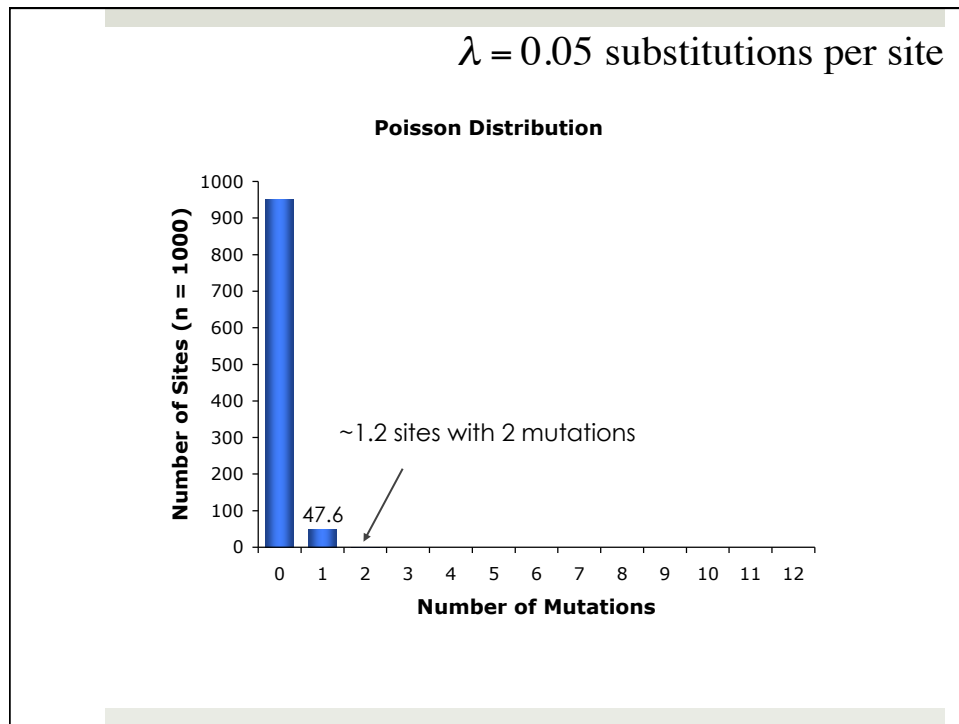now, what is the chance that the next substitution obscures one of the first two?
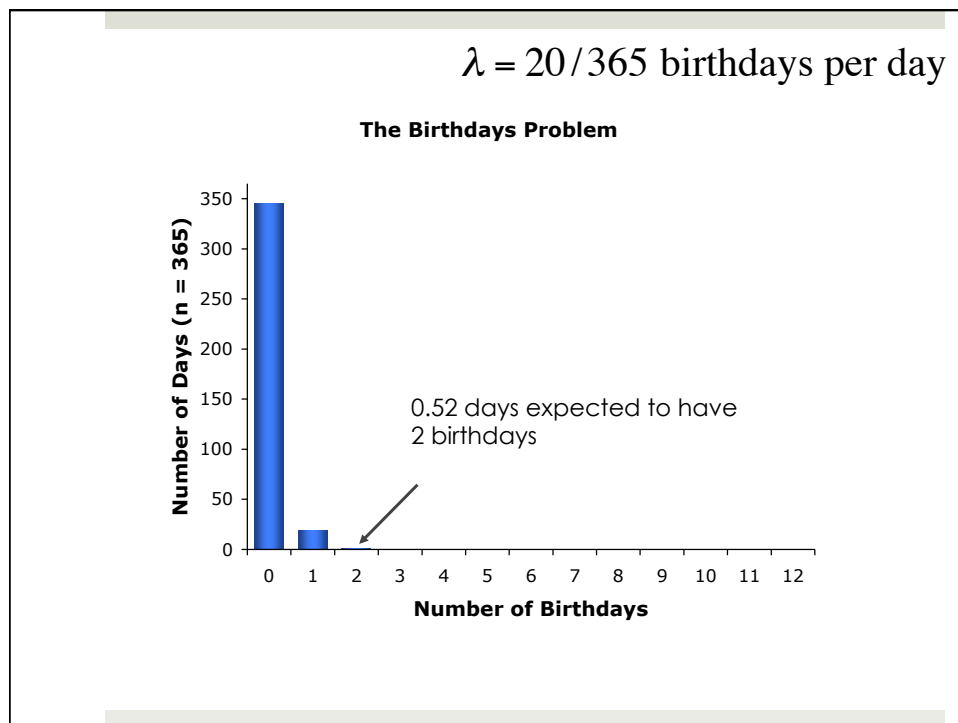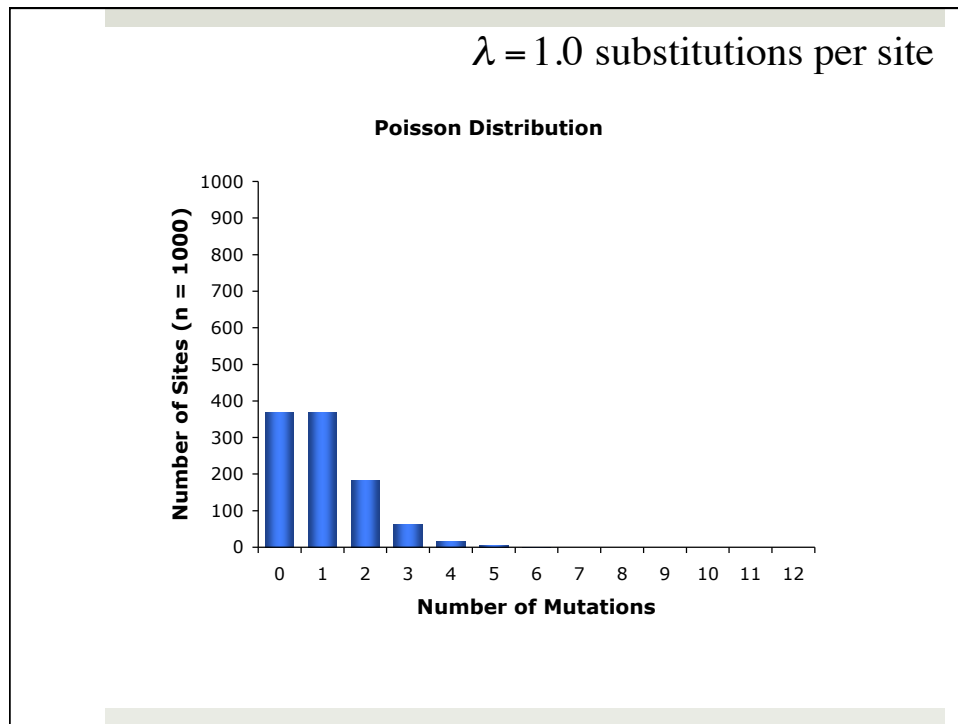

$\lambda = 0.01$ substitutions per site

Poisson Distribution
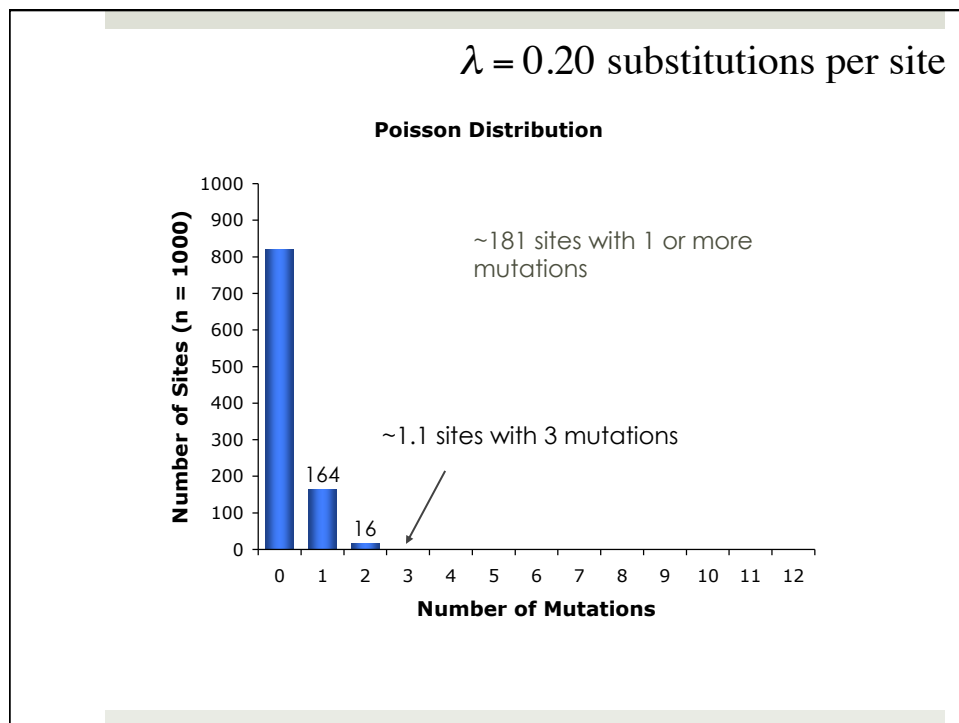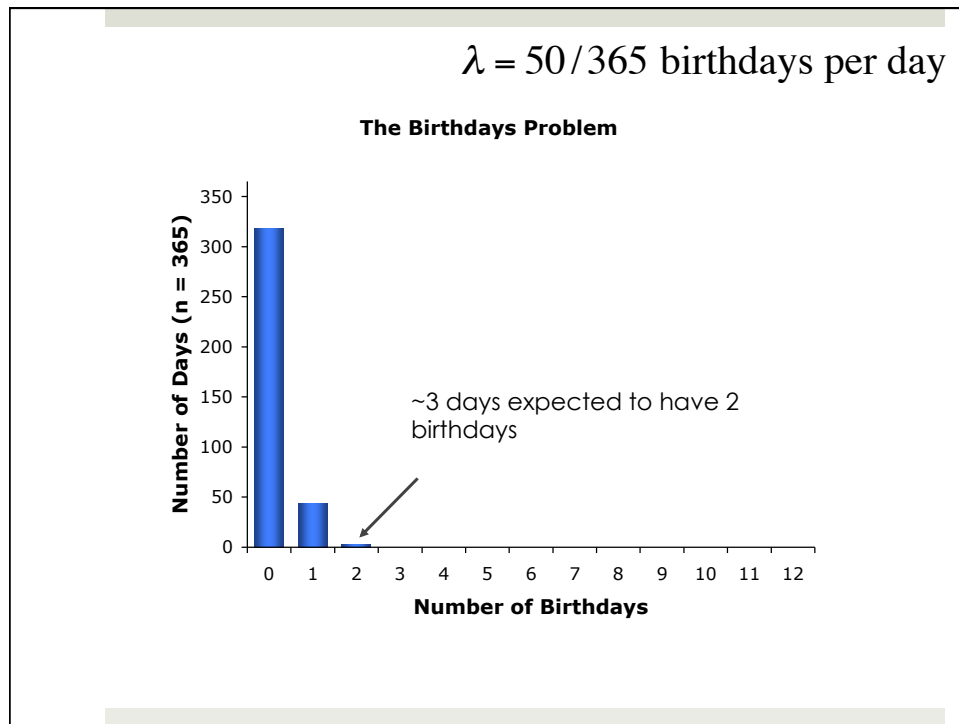
$$f(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

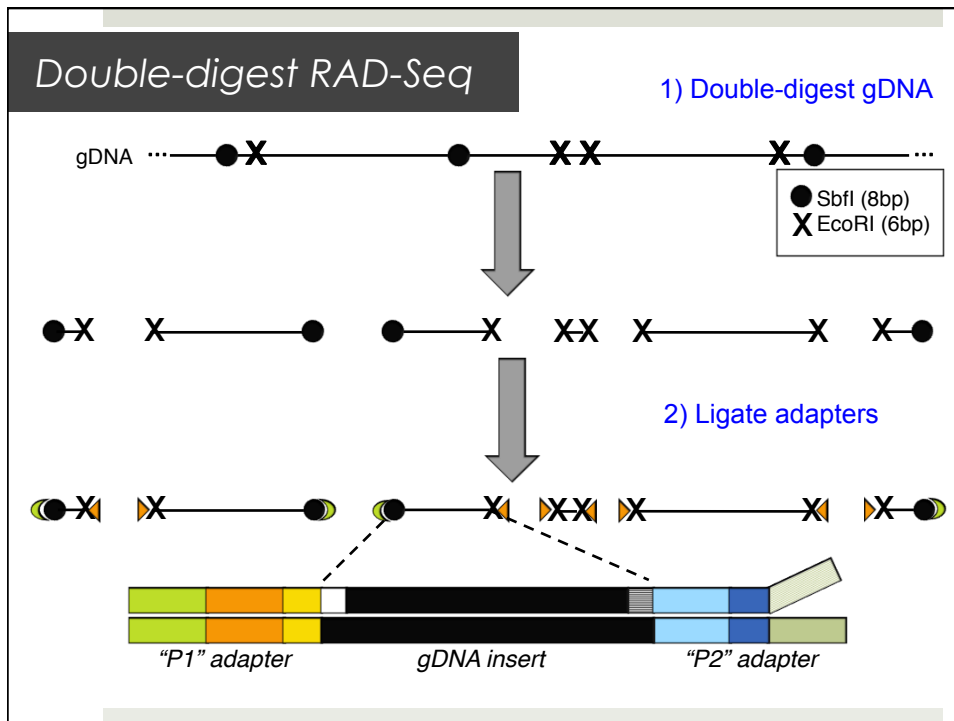where k is the number of occurrences (mutations) and λ is the mean rate

~9.9

Number of Sites (n = 1000)

Number of Mutations

$\lambda = 0.05$ substitutions per site

Poisson Distribution

~1.2 sites with 2 mutations

47.6

Number of Sites (n = 1000)

Number of Mutations

$\lambda = 0.20$ substitutions per site

Poisson Distribution

~1.1 sites with 3 mutations

16

Number of Sites (n = 1000)

Number of Mutations

$\lambda = 1.0$ substitutions per site

Poisson Distribution

Number of Sites (n = 1000) vs. Number of Mutations

$\lambda = 20/365$ birthdays per day

The Birthdays Problem

Number of Days (n = 365) vs. Number of Birthdays

0.52 days expected to have 2 birthdays

$\lambda = 50/365$ birthdays per day

The Birthdays Problem

~3 days expected to have 2 birthdays

$\lambda = 0.20$ substitutions per site

Poisson Distribution

~181 sites with 1 or more mutations

~1.1 sites with 3 mutations

Divergence of DNA Sequences

❖ even if mutation occurs by a random Poisson process...

  ♦ divergence (genetic distance) depends on changes in both sequences, not just one
  ♦ mutations yield one of four different nucleotides (A, C, G, T)
  ♦ parallel and reverse mutations may result in sequences being identical at a particular position

Double-digest RAD-Seq

1) Double-digest gDNA

● SbfI (8bp)
X EcoRI (6bp)

2) Ligate adapters

"P1" adapter   gDNA insert   "P2" adapter

ddRAD-Seq output