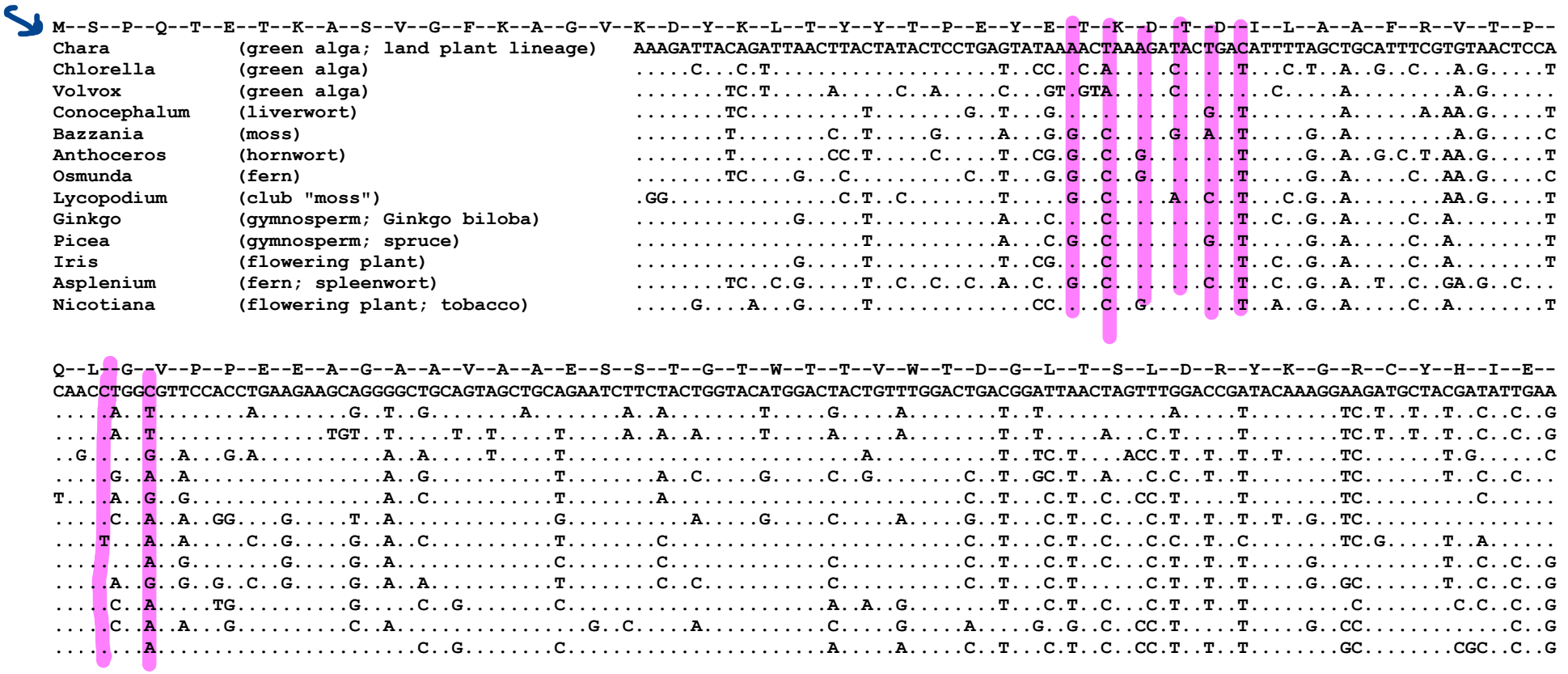
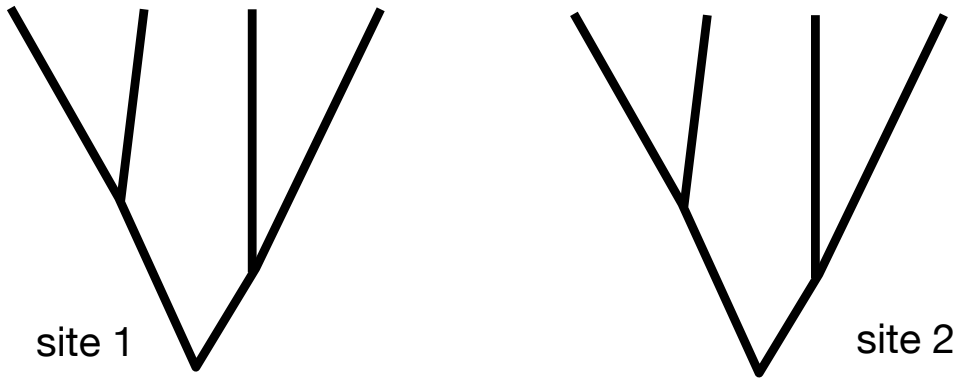
Green Plant rbc*L*

First 88 amino acids (translation is for *Zea mays*)



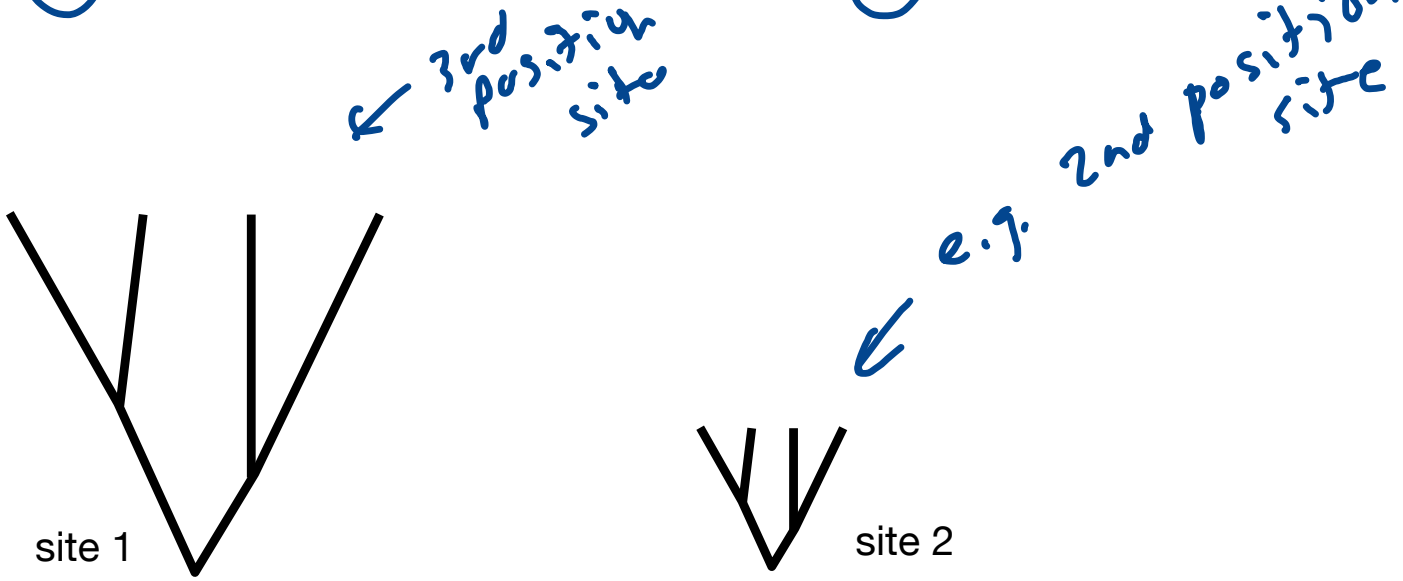
Rate homogeneity



All edges same for every site

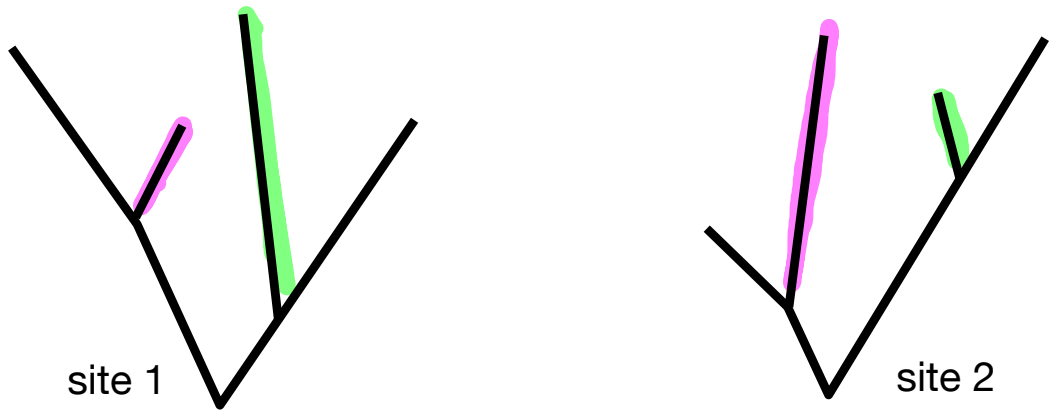
Among-site rate variation (ASRV)

(this is modeled by site-specific rates, discrete gamma (G), or proportion of invariable sites (I) models)



Rate varies across sites but edge lengths proportional (tree scales up or down across sites)

Heterotachy

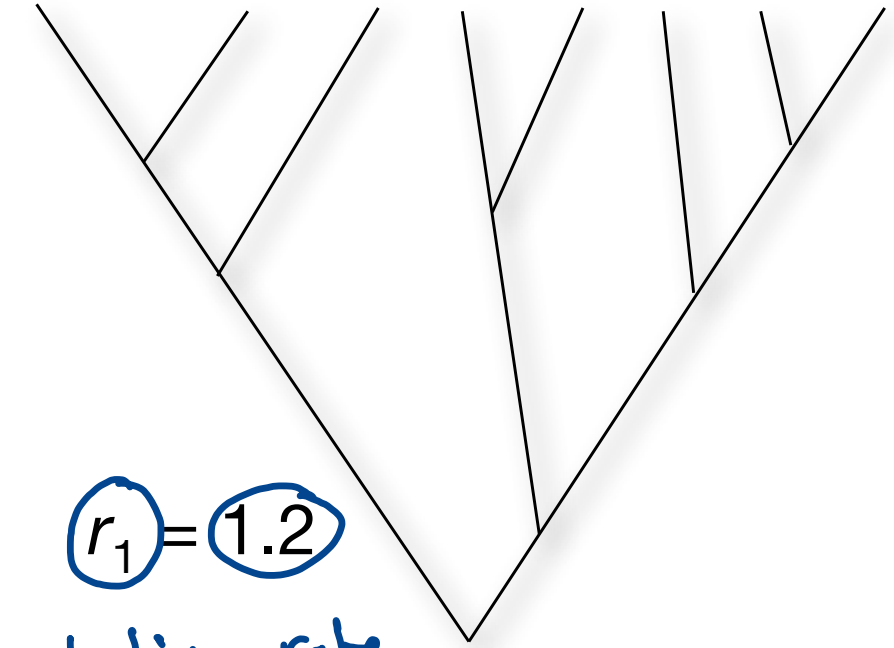


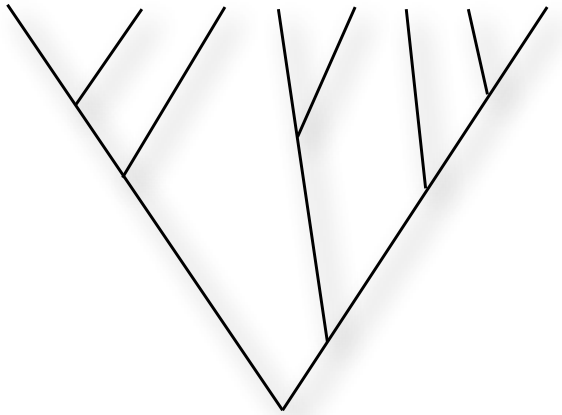
Rates vary but not simple scaling
(rate for a particular edge slow for
some sites, fast for others)

Site-specific rates

1000 sites r_1 1000 sites r_2

$$L = \left(\Pr(D_1|r_1) \cdots \Pr(D_{1000}|r_1)\right)\left(\Pr(D_{1001}|r_2) \cdots \Pr(D_{2000}|r_2)\right)$$





Site-specific rates

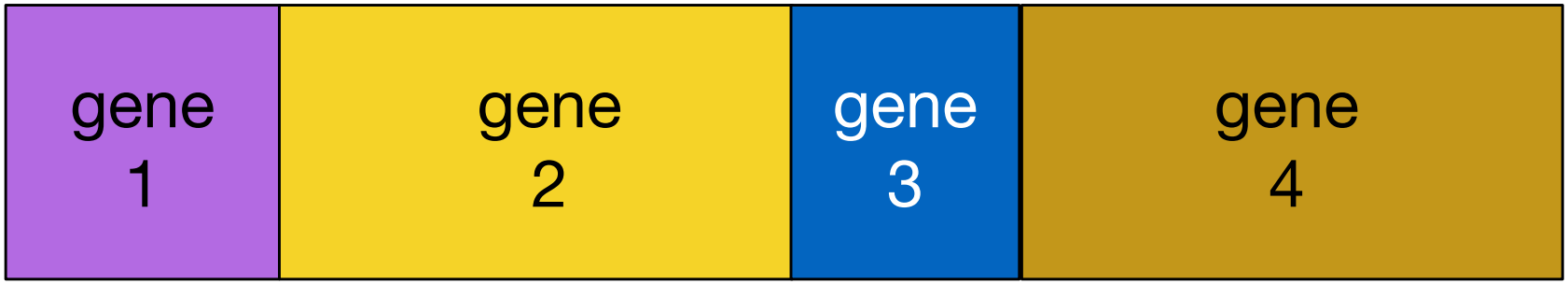
$$p_{ij} = \frac{1}{4} - \frac{1}{4} e^{-\frac{4 v r_1}{3}}$$

$$p_{ii} = \frac{1}{4} + \frac{3}{4} e^{-\frac{4 v r_1}{3}}$$

use relative rate r_1

Partitioned models

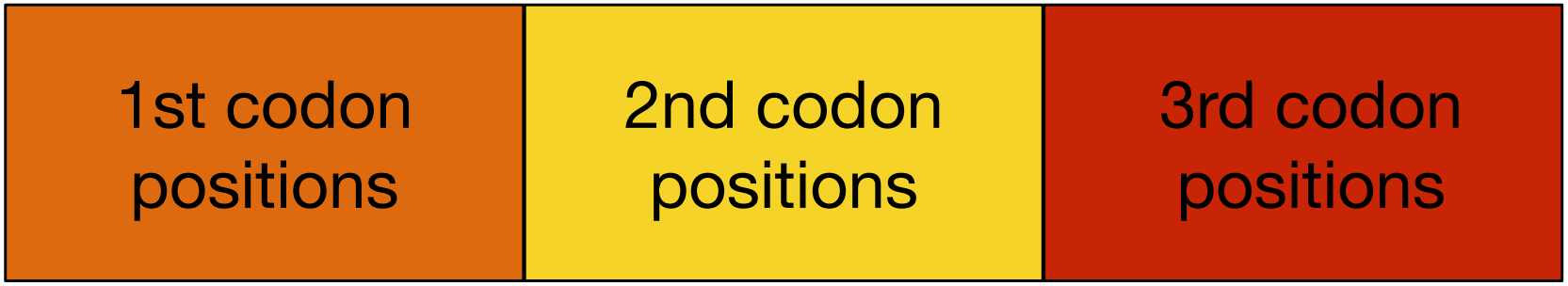
$p_1 r_1 + p_2 r_2 + p_3 r_3 + p_4 r_4 = 1$



partitioned by gene, 4 subsets

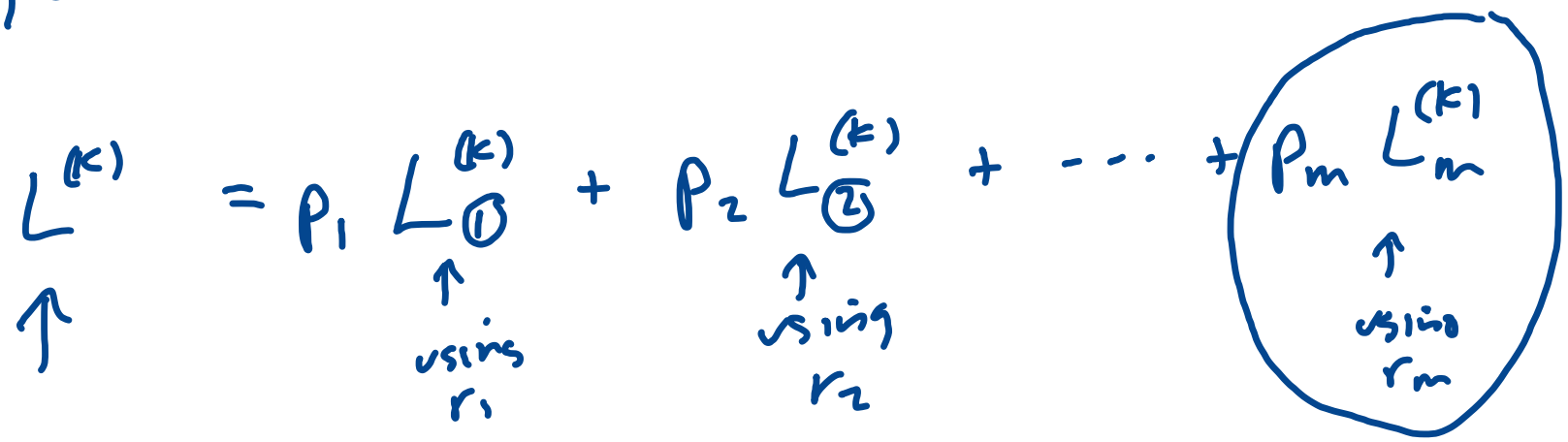
$p_1 = p_2 = p_3 = \frac{1}{3}$

$p_1 r_1 + p_2 r_2 + p_3 r_3 = 1$



partitioned by codon position, 3 subsets

G, I models Mixture Models



free rates/probs

Invariable sites (I) model

$$L^{(k)} = p_{invar} L_1^{(k)} + (1 - p_{invar}) L_2^{(k)}$$

p_{invar}: proportion of invariable sites

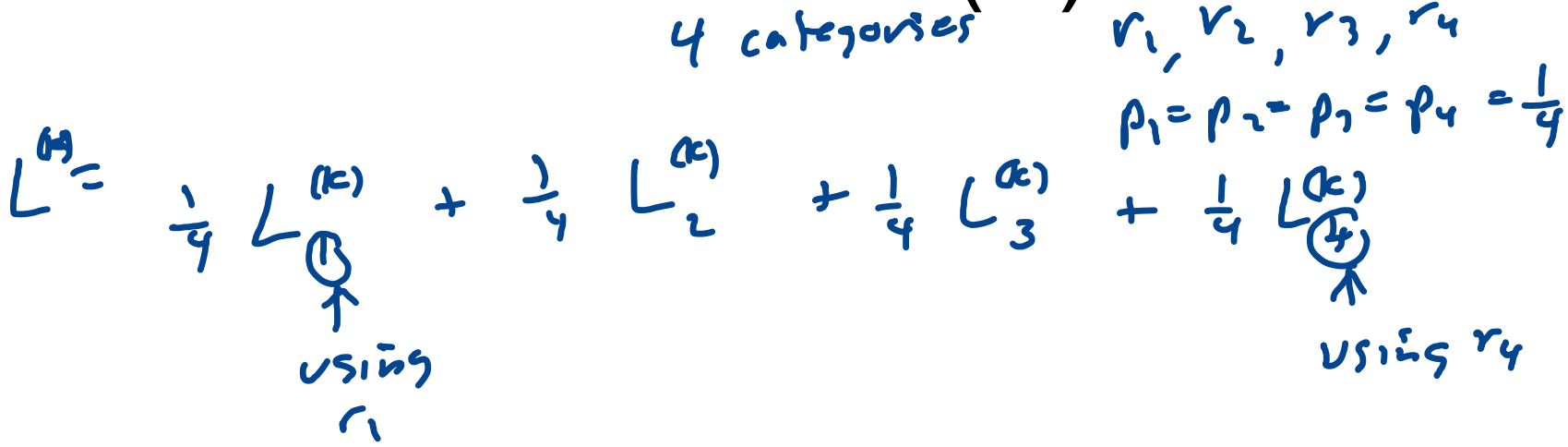
$r_1 = 0$ $\qquad$ $r_2 =$

$$p_{invar} r_1 + (1 - p_{invar}) r_2 = 1$$

$$(p_{invar})(0) + (1 - p_{invar}) r_2 = 1$$

$$r_2 = \frac{1}{1 - p_{invar}}$$

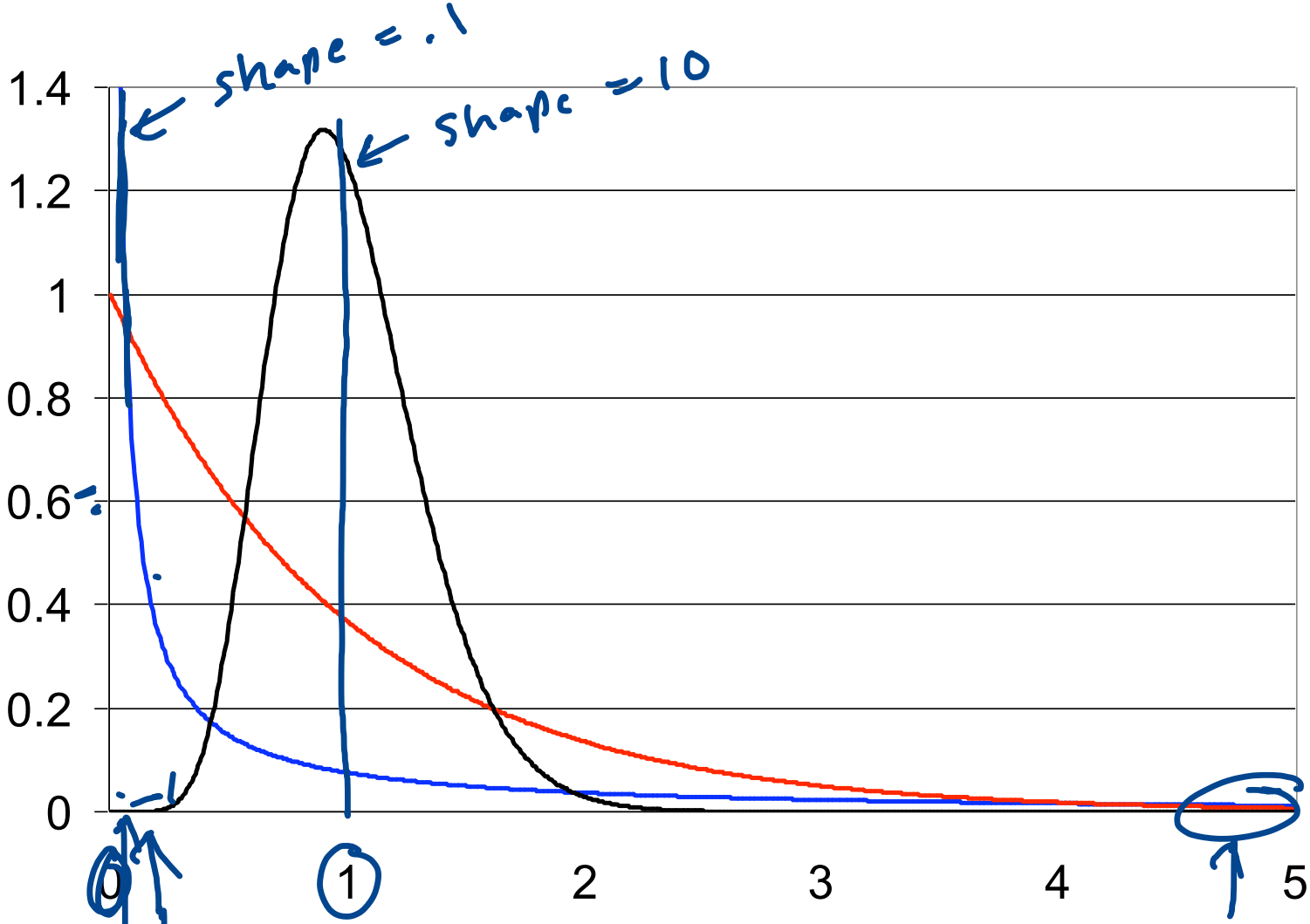
Discrete Gamma (G) model



mixture model formula

Yang (1993, 1994)

Gamma distributions



STOPPED HERE 2024-02-08

mean = 1 axis labels shape = 1/variance

Relative rates in 4-category case