CS221 Quiz Solutions

Release v10

Christopher Lee

October 27, 2017

Contents

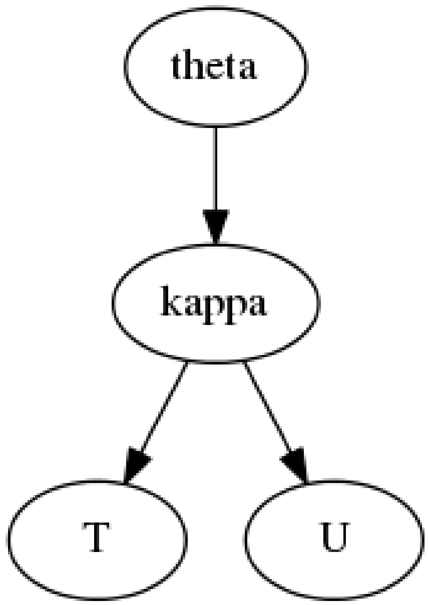
1. **Genetic Testing of Twins**

   (a) • θ: population allele frequency;

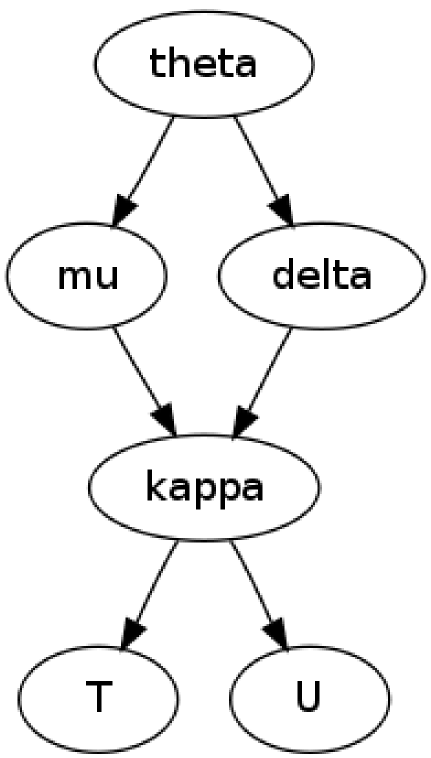
   • $\mu, \delta, \kappa, \lambda$: hidden genetic state (copy number of the SNP) in the mom, dad and twins. All of these are necessary in the fraternal twins model.

   • In the identical twins model, we don't **have to** include the parents (although we could), because including the parents does not change the likelihood of T,U. Hence we prefer to omit them.

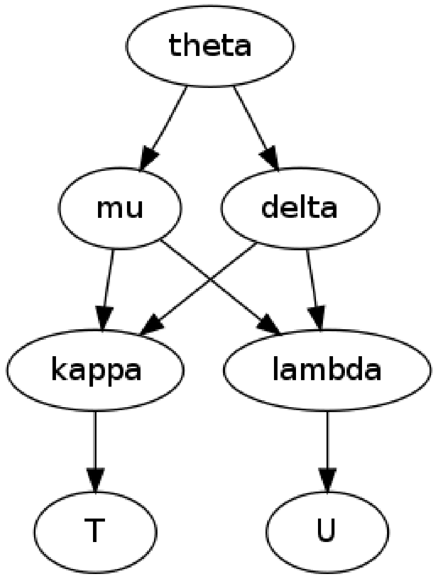
   Identical (without parents, preferred):



Identical (with parents, more complex than strictly necessary):



Fraternal:



(b) The fraternal twins model involves a term $p(\kappa|\mu, \delta)$ which is $O(N^3)$. We cannot get rid of this complexity, because these hidden variables are the key connection between the observations T,U.

The identical twins model is a bit more subtle. We could of course include the parents (as we did in the fraternal twins model) making the complexity math:*O(N^3)*, but this is not strictly necessary, because only κ actually connects T,U. Using the simplified model the complexity is only math:*O(N)*.

(c) We have *n* independent markers, whose joint probability will just be the product of their individual probabilities:

$$p(T_1, U_1, T_2, U_2, ...T_n, U_n|\theta_1, \theta_2, ...\theta_n) = \prod_{i=1}^{n} p(T_i, U_i|\theta_i)$$

Finally, we wish to obtain a posterior odds ratio for the two models:

$$\frac{p(identical|obs, \theta_1, ...\theta_n)}{p(fraternal|obs, \theta_1, ...\theta_n)} = \frac{p(obs|identical, \theta_1, ...\theta_n)p(identical)}{p(obs|fraternal, \theta_1, ...\theta_n)p(fraternal)}$$

$$= \frac{p(identical) \prod_{i=1}^{n} p(T_i, U_i|identical, \theta_i)}{p(fraternal) \prod_{i=1}^{n} p(T_i, U_i|fraternal, \theta_i)}$$

where the priors are $p(identical) = 0.001$ and $p(identical) = 0.999$.