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.
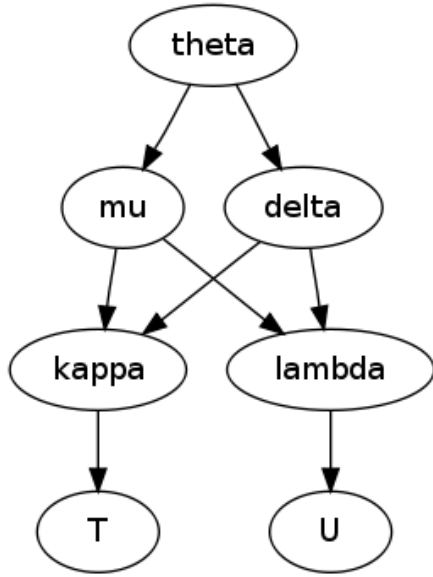
   Identical (without parents, preferred):
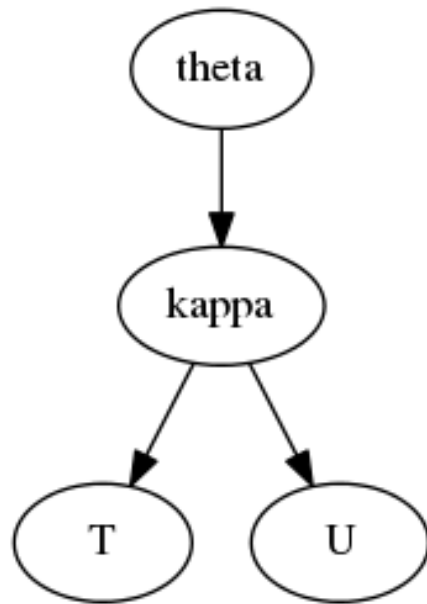
Identical (with parents, more complex than strictly necessary):
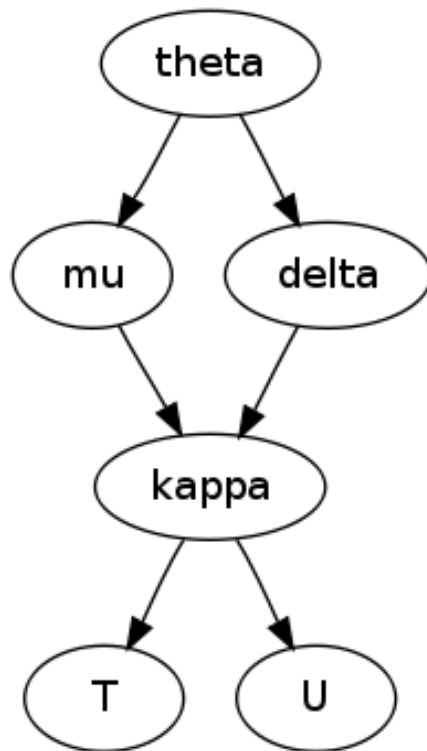
Fraternal:

(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)

$$= \frac{\sum_{\mu,\delta} p(T = 0.5|\kappa = 1)p(U = 0.5|\kappa = 1)p(\kappa = 1|\mu, \delta)p(\mu, \delta|\theta)p(identical)}{\sum_{\mu,\delta} p(T = 0.5|\kappa = 1)p(U = 0.5|\lambda = 1)p(\kappa = 1|\mu, \delta)p(\lambda = 1|\mu, \delta)p(\mu, \delta|\theta)p(fraternal)}$$

The first two terms in numerator and denominator cancel, giving us:

$$= \frac{p(identical) \sum_{\mu,\delta} p(\kappa = 1|\mu, \delta)p(\mu|\theta)p(\delta|\theta)}{p(fraternal) \sum_{\mu,\delta} p(\kappa = 1|\mu, \delta)p(\lambda = 1|\mu, \delta)p(\mu|\theta)p(\delta|\theta)}$$

(d) The copy number for a twin (κ) should be the result of two independent binomial draws (one chromosome copy from each parent). For example, the conditional probability $p(\kappa = 0 + 1|\mu = 0, \delta = 1) = p(0|\mu = 0)p(1|\delta = 1) = 1(1/2) = 1/2$.