CS121 / 221 Quiz 2 Solutions

Release v10

Christopher Lee

November 07, 2017

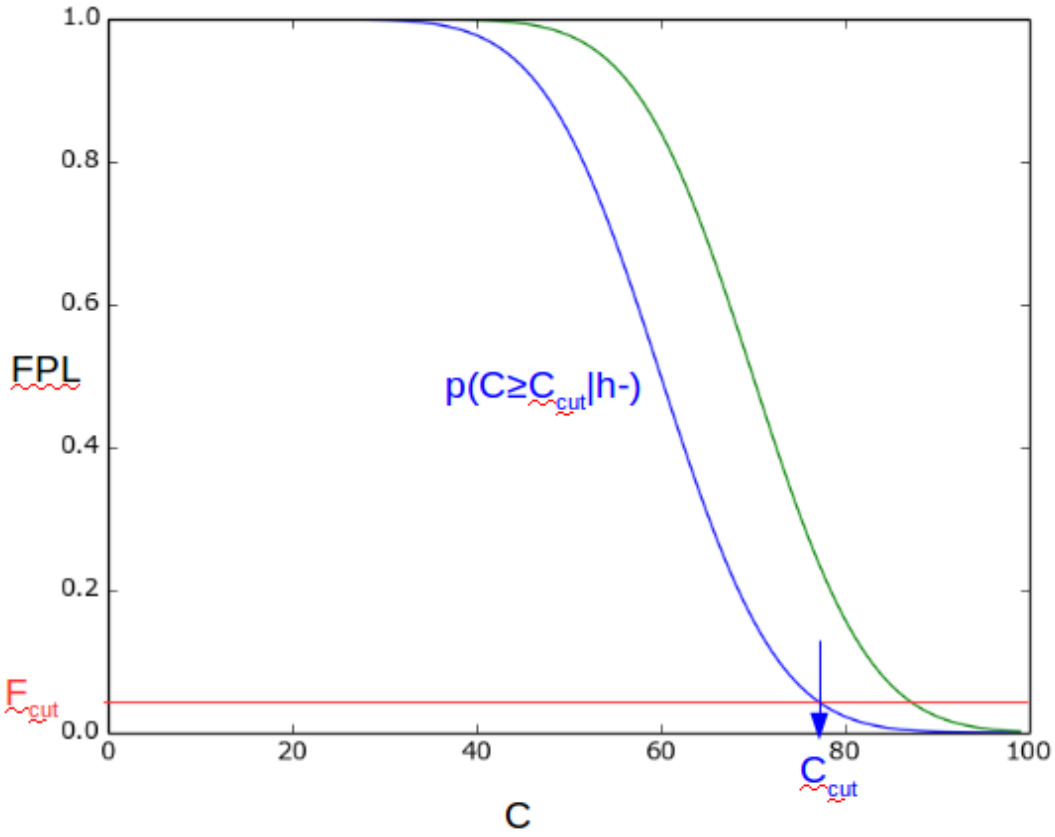
Contents

1. **X-linked vs. Autosomal Inheritance Hypothesis Test**

   (a) Under the X-linked recessive model, disease symptoms should be much less frequent among women (who must get two copies of the disease mutation, with probability θ^2) than among men (who only need to get a single copy, with probability θ). By contrast, under the autosomal recessive model, the fraction in men should be equal to that in women (because in both cases they must get two copies of the disease mutation).

   (b) Say that our total sample contains N people, of whom M are men, and D total people have disease symptoms. Then we could use the count of men who have disease symptoms (call it *C*) as our scoring function, because it will be much larger under the X-linked recessive model than under the autosomal recessive model.

   (c) Using *C* as our scoring function, we will choose an appropriate cutoff value c_{cut}, and then define t^+ as $C \geq c_{cut}$ (otherwise t^-). We then must choose that cutoff value such that $p(C \geq c_{cut}|h^-) \leq F_{cut}$.



We could use a hypergeometric likelihood model to represent the autosomal recessive hypothesis (i.e. that the disease rates are equal in men and women), where

$$p(C|N, M, D, h^-) = \frac{\binom{D}{C}\binom{N-D}{M-C}}{\binom{N}{M}}$$

so for a given c_{cut}, the false positive likelihood becomes:

$$p(t^+|h^-) = \sum_{C=c_{cut}}^{M} \frac{\binom{D}{C}\binom{N-D}{M-C}}{\binom{N}{M}}$$

Alternatively, we could use a binomial likelihood model, where ϕ is the fraction of men expected to get disease:

$$p(C|M, \phi, h^-) = \binom{M}{C}\phi^C(1-\phi)^{M-C}$$

We could set the value of ϕ either from $\phi = \theta^2$ or estimated from the sample of women as $\phi = (D - C)/(N - M)$. Then the false positive likelihood becomes:

$$p(t^+|h^-) = \sum_{C=c_{cut}}^{M} \binom{M}{C}\phi^C(1-\phi)^{M-C}$$

(d) We want to calculate the posterior odds ratio for the X-linked vs. autosomal recessive models:

$$\frac{p(h^+|t^+)}{p(h^-|t^+)} = \frac{p(t^+|h^+)p(h^+)}{p(t^+|h^-)p(h^-)}$$

Clearly we need to know the prior $p(h^+)$. We could estimate this from the fraction of the genome that is on chromosome X, simplistically as $p(h^+) = 1/23$ (because there is one X chromosome and 22 autosomal chromosomes).

We also need to know $p(t^+|h^+)$. If the *C* distributions for the two models are well-separated, then it would be reasonable to assume $p(t^+|h^+) \approx 1$.