JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Statistics for laboratory scientists

Solutions for the homework problems for lecture 6

1.
   a. Suppose that A and B are *mutually exclusive*.

      i. Pr(A or B) = Pr(A) + Pr(B)

      ii. Pr(A and B) = 0, since they can't both happen.

   b. Suppose that A and B are *independent*.

      i. Pr(A and B) = Pr(A) × Pr(B)

      ii. Pr(A or B) = Pr(A) + Pr(B) - Pr(A) × Pr(B)

   c. Suppose that A and B are *both* mutually exclusive and independent.

      Since A and B are independent, Pr(A and B) = Pr(A) × Pr(B). But since A and B are mutually exclusive, Pr(A and B) = 0. Thus Pr(A) × Pr(B) = 0. And so either Pr(A) = 0 or Pr(B) = 0 or both. In other words, either A or B (or both) *cannot happen*!

      The point: generally when we talk about independent events, they are *not* mutually exclusive, and vice versa.

2. Recall that we pick a woman at random from the population. Let U = {She is unaffected}, C = {She is a carrier}, A = {her brother is affected and her parents are both unaffected}, B = {her first child is affected}, D = {at least one of her five children is affected}. Note that by "carrier," I mean "has genotype +d."

   Calculate the following.

   a. Pr(C | U) = $2p(1-p) / [2p(1-p) + (1-p)^2]$ = ... = $2p / (1 + p)$ = approx 2.0%.

   b. Pr(C | U and A) = 2/3 = approx 67%, since her parents must both be carriers.

c. Pr(B | U and A) = Pr(C | U and A) × Pr(B | C and U and A) = (2/3) × [(1/2) × *p*] = p/3 = approx 3/1000.

   Note that Pr(B | C and U and A) = Pr(B | C). We're assuming here that the woman is mating at random (*with respect to genotype at this gene*). Thus the allele contributed by the father has probability *p* of being the disease allele.

d. Let E = {none of the five children is affected}, F_0 = {father has genotype ++}, F_1 = {father has genotype +d}, and F_2 = {father has genotype ++}.

   Note that Pr(E | C) = Pr(F_0) × Pr(E | C and F_0) + Pr(F_1) × Pr(E | C and F_1) + Pr(F_2) × Pr(E | C and F_2)
   = $(1-p)^2 \times 1 + 2\,p\,(1-p) \times (3/4)^5 + p^2 \times (1/2)^5$ = approx 98.5%.

   Thus, Pr(D | U and A) = Pr(C | U and A) × Pr(D | C and U and A)
   = Pr(C | U and A) × [1 - Pr(E | C)]
   = approx 1%.

3.

a. Pr(G_1) = Pr(A) × Pr(G_1 | A) + Pr(B) × Pr(G_1 | B) = (1/3) × (1/2) + (2/3) × (1/5) = 3/10 = 30%

b. Pr(G_2 | A and G_1) = 4/9

c. Pr(G_2 | A) = 1/2

d. Pr(G_2 | B and G_1) = 1/9

e. Pr(G_2 | G_1) = Pr(G_2 and G_1) / Pr(G_1)

   Note that Pr(G_2 and G_1) = Pr(A) × Pr(G_1 and G_2 | A) + Pr(B) × Pr(G_1 and G_2 | B) = (1/3) × (5/10) × (4/9) + (2/3) × (2/10) × (1/9) = 4/45 = approx 8.9%

   We found that Pr(G_1) = 3/10 (part a).

Thus, $Pr(G_2 | G_1)$ = (4/45) / (3/10) = 8/27 = approx 30%

f. Pr(exactly one green ball) = $Pr(G_1$ and not $G_2)$ + $Pr(G_2$ and not $G_1)$ = 2 × $Pr(G_1$ and not $G_2)$ = 2 × [Pr(A) × $Pr(G_1$ and not $G_2)$ + Pr(B) × $Pr(G_1$ and not $G_2)$] = 2 × [(1/3) × (5/10) × (5/9) + (2/3) × (2/10) × (8/9)] = 114/270 = approx 42%.

g. Note that Pr(at least one green ball | A) = 1 - Pr(no green balls | A) = 1 - (1/2) × (4/9) = 7/9 = approx 78%. Pr(at least one green ball | B) = 1 - (4/5) × (7/9) = 17/45 = approx 38%.

   Thus Pr(at least one green ball) = (1/3) × (7/9) + (2/3) × (17/45) = approx 51%.

   And so Pr(A | at least one green ball) = Pr (A) × Pr(at least one green ball | A) / Pr(at least one green ball) = (1/3) × (7/9) / [(1/3) × (7/9) + (2/3) × (17/45)] = approx 51%.

4. Let F = {the champion is the colt's father} and C = {the colt carries the Y marker}.

   a. Pr(F) = 1/2.

   b. Pr(F | C) = Pr(F) Pr(C|F) / [Pr(F) Pr(C|F) + Pr(not F) Pr(C | not F)] = (1/2) × (1) / [(1/2) × 1 + (1/2) × 0.02] = approx 98%.

   c. Given that the mare was exposed to a total of 1000 stallions, we have Pr(F) = 1/1000. Thus Pr(F | C) = Pr(F) Pr(C|F) / [Pr(F) Pr (C|F) + Pr(not F) Pr(C | not F)] = (1/1000) × (1) / [(1/1000) × 1 + (999/1000) × 0.02] = approx 4.8%.

[3rd term syllabus | 4rd term syllabus | R for Windows]

Last modified: Wed Feb 22 09:44:30 EST 2006