STAT 886 HW 1

David Nguyen

February 8, 2021

1.3

Probability calculation for genetics (from Lindley, 1965): suppose that in each individual of a large population there is a pair of genes, each of which can be either x or X, that controls eye color: those with xx have blue eyes, while heterozygotes (those with Xx or xX) and those with XX have brown eyes.

The proportion of blue-eyed individuals is p^2 and of heterozygotes is $2p(1-p)$, where 0 < p < 1. Each parent transmits one of its own genes to the child; if a parent is a heterozygote, the probability that it transmits the gene of type X is 1/2 .

Assuming random mating, show that among brown-eyed children of brown-eyed parents, the expected proportion of heterozygotes is 2p/(1+2p).

Probabilities of heterozygote child given both parent's have brown eyes

- p(xX | XX, XX) = 0
- p(xX | xX, XX) = 1/2
- p(xX | xX, xX) = 1/2

Prior probabilities of parent chromosomal combinations

- p(XX, XX) = $(1-p)^2(1-p)^2 = (1-p)^4$
- p(xX, XX) = $(1-p)^2 2p(1-p) + 2p(1-p)(1-p)^2 = 4p(1-p)^3$
- p(xX, xX) = $2p(1-p)2p(1-p) = 4p^2(1-p)^2$

probability a child has brown eyes given their parent's

- p(brown | XX,XX) = 1
- p(brown | xX, XX) = 1
- p(brown | xX, xX) = 3/4

$$
\begin{aligned}
p(\text{xX, child has brown eyes, parents have brown eyes}) &= \frac{p(\text{xX, parents have brown eyes})}{p(\text{child has brown eyes, parents have brown eyes})} \\
&= \frac{0 \cdot (1-p)^4 + 1/2 \cdot 4p(1-p)^3 + 1/2 \cdot 4p^2(1-p)^2}{1 \cdot (1-p)^4 + 1 \cdot 4p(1-p)^3 + 3/4 \cdot 4p^2(1-p)^2} \\
&= \frac{2p(1-p)^3 + 2p^2(1-p)^2}{(1-p)^4 + 4p(1-p)^3 + 3p^2(1-p)^2} \\
&= \frac{2p(1-P) + 2p^2}{(1-p)^2 + 4p(1-p) + xp^2} \\
&= \frac{2p - 2p^2 + 2p}{p^2 - 2p + 1 + 4p - 4p^2 + 3p^2} \\
&= \frac{2p}{1+2p}
\end{aligned}
$$

Suppose Judy, a brown-eyed child of brown-eyed parents, marries a heterozygote, and they have n children, all brown-eyed. Find the posterior probability that Judy is a heterozygote.

From the previous problem we know $P(\text{Judy is xX}|\text{her and her parents eyes are brown} = \frac{2p}{1+2p}$ which implies that $P(\text{Judy is XX}|\text{her and her parents eyes are brown} = \frac{1}{1+2p}$, The probability that Judy and her heterozygous partner produce n children with brown eyes is $\left(\frac{3}{4}\right)^n$ given that she is a heterozygote and 1^n given that she is a homozygote (XX).

$$
P(\text{Judy is xX}|\text{Judy, her parents, and n children have brown eyes}) = \frac{\frac{2p}{1+2p}\left(\frac{3}{4}\right)^n}{\frac{2p}{1+2p}\left(\frac{3}{4}\right)^n + \frac{1}{1+2p}1^n}
$$

1.4

Probability assignment: we will use the football dataset to estimate some conditional probabilities about professional football games. There were twelve games with point spreads of 8 points; the outcomes in those games were: -7, -5, -3, -3, 1, 6, 7, 13, 15, 16, 20, and 21, with positive values indicating wins by the favorite and negative values indicating wins by the underdog. Consider the following conditional probabilities:

- Pr(favorite wins | point spread = 8),
- Pr(favorite wins by at least 8 | point spread = 8),
- Pr(favorite wins by at least 8 | point spread = 8 and favorite wins).

1.4a

Estimate each of these using the relative frequencies of games with a point spread of 8.

$$
\begin{aligned}
Pr(\text{favorite wins} \mid \text{point spread} = 8) &= \frac{8}{12} \\
Pr(\text{favorite wins by at least 8} \mid \text{point spread} = 8) &= \frac{5}{12} \\
Pr(\text{favorite wins by at least 8} \mid \text{point spread} = 8 \text{ and favorite wins}) &= \frac{5}{8}
\end{aligned}
$$

1.4b

Estimate each using the normal approximation for the distribution of (outcome − point spread).

The sample standard deviation of deviations (score difference - point spread) is 10.1 which we will round to 10. Even though the sample mean of deviations is -1.25 we will assume, following BDA3, that a mean of zero is appropriate for approximating the deviations such that $d \sim N(0, 10^2)$.

- Pr(favorite wins | point spread = 8) = 0.7881446
- Pr(favorite wins by at least 8 | point spread = 8) = 0.5
- Pr(favorite wins by at least 8 | point spread = 8 and favorite wins) = 0.6344014

1.6

Conditional probability: approximately 1/125 of all births are fraternal twins and 1/300 of births are identical twins. Elvis Presley had a twin brother (who died at birth). What is the probability that Elvis was an identical twin? (You may approximate the probability of a boy or girl birth as 1/2.)

Note that:

- pr(identical twin brothers) = (1/300)(1/2) = 1/600
- pr(fraternal twin brothers) = $(1/125)(1/2)^2$ = 1/500
- pr(twin brothers) = 1/600 + 1/500 = 11/3000

So,

$$\begin{aligned} pr(\text{Elvis' brother was identical}) &= \frac{pr(\text{identical twin brothers})}{pr(\text{twin brothers})} \\ &= \frac{1/600}{11/3000} \\ &= \frac{5}{11} \end{aligned}$$

1.7

Conditional probability: the following problem is loosely based on the television game show Let's Make a Deal. At the end of the show, a contestant is asked to choose one of three large boxes, where one box contains a fabulous prize and the other two boxes contain lesser prizes. After the contestant chooses a box, Monty Hall, the host of the show, opens one of the two boxes containing smaller prizes. (In order to keep the conclusion suspenseful, Monty does not open the box selected by the contestant.) Monty offers the contestant the opportunity to switch from the chosen box to the remaining unopened box. Should the contestant switch or stay with the original choice? Calculate the probability that the contestant wins under each strategy. This is an exercise in being clear about the information that should be conditioned on when constructing a probability judgment. See Selvin (1975) and Morgan et al. (1991) for further discussion of this problem.

Denote that the door we initially choose as door "A" and the door that Monty Hall opens as door "B". We want to know if switching our choice to the remaining door "C" gives us a different probability of winning that sticking with our initial choice of door "A". Let $\theta = i$ denote where the greatest prize is, where $i = \{A, B, C\}$

- P(θ = A) = 1/3

- P(opens B | θ = A) = 1/2
- P(opens B | θ = C) = 1 (since he cannot open the door we picked, or the door with the prize)
- p(A)p(opens B | θ = A) = (1/3) (1/2) = 1/6
- p(C)p(opens B | θ = C) = (1/3) 1 = 1/3
- p(opens B) = 1/6 + 1/3 = 1/2

$$
\begin{aligned}
p(\theta = A|\text{opens } B) &= \frac{p(A)p(\text{opens B}|\theta = A)}{P(\text{opens B})} \\
&= \frac{1/6}{1/2} \\
&= 1/3 \\
p(\theta = C|\text{opens } B) &= \frac{p(C)p(\text{opens B}|\theta = C)}{P(\text{opens B})} \\
&= \frac{1/3}{1/2} \\
&= 2/3
\end{aligned}
$$

Therefore, the probability of winning is actually twice as large if we switch our choice ((pr(win) = 2/3)) than if we didn't (pr(win) = 1/3).

2.2

Predictive distributions: consider two coins, C_1 and C_2, with the following characteristics: Pr(heads$|C_1$) = 0.6 and Pr(heads$|C_2$) = 0.4. Choose one of the coins at random and imagine spinning it repeatedly. Given that the first two spins from the chosen coin are tails, what is the expectation of the number of additional spins until a head shows up?

- pr(C_1) = pr(C_2) = 0.5
- pr(TT | C_1) = 0.6^2
- pr(TT | C_2) = 0.4^2

We can use this information to compute the posterior probability that we have selected C_1 or C_2:

$$
\begin{aligned}
P(C_1|TT) &= \frac{pr(C_1)pr(TT|C_1)}{pr(C_1)pr(TT|C_1) + pr(C_2)pr(TT|C_2)} \\
&= \frac{0.5 \times 0.6^2}{0.5 \times 0.6^2 + 0.5 \times 0.4^2} \\
&= 0.692 \\
P(C_2|TT) &= 1 - p(C_1|TT) = 1 - 0.692 = 0.308
\end{aligned}
$$

Note that the number of additional spins until the first head follows a geometric distribution which has mean $\frac{1}{p}$ where p is the probability of success. Since the geometric distribution is "memoryless" we do not need to account for the two tails we have already observed when computing the expected number of additional spins.

$$
\begin{aligned}
E(Spins|TT) &= 0.692\left(\frac{1}{0.6}\right) + 0.308\left(\frac{1}{0.4}\right) \\
&= 1.92
\end{aligned}
$$

The expected number of additional spins to get heads is 1.92.