Section 4.6 Bayes' Theorem

Math 141

Main ideas

Bayes' Theorem: if the sample space $S = E \cup E'$ (everything is either in E or not), then

$$\Pr(E|F) = \frac{Pr(E \cap F)}{\Pr(F)} = \frac{Pr(E \cap F)}{Pr(E \cap F) + Pr(E' \cap F)} = \frac{Pr(E) \cdot \Pr(F|E)}{Pr(E) \cdot \Pr(F|E) + Pr(E') \cdot \Pr(F|E')}$$

So in order to find $\Pr(E|F)$ we use $\Pr(F|E)$, plus some other values.

Where the entire sample space can be divided into mutually exclusive (non-overlapping) categories

$$S = E_1 \cup E_2 \cup \cdots \cup E_n$$

then

$$\Pr(F) = Pr(F \cap E_1) + Pr(F \cap E_2) + \cdots + Pr(F \cap E_n)$$

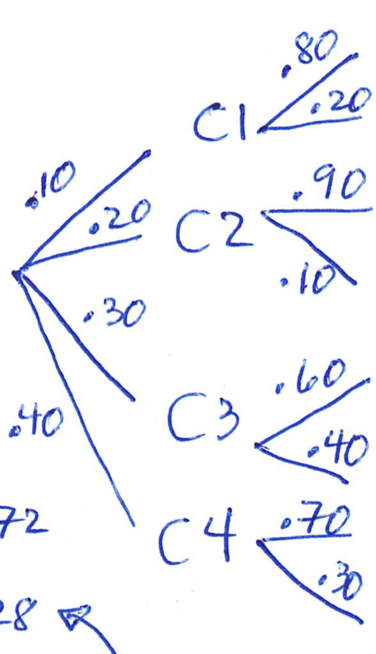
and

$$\Pr(E_i|F) = \frac{Pr(E_i \cap F)}{\Pr(F)} = \frac{Pr(E_i) \cdot \Pr(F|E_i)}{Pr(E_1) \cdot \Pr(F|E_1) + \cdots + Pr(E_i) \cdot \Pr(F|E_i) + \cdots + Pr(E_n) \cdot \Pr(F|E_n)}$$

Problems

1. Classes and grades.

Class	Fraction of all students	Fraction of group with an A
C1 (freshman)	.10	.80
C2 (sophomore)	.20	.90
C3 (junior)	.30	.60
C4 (senior)	.40	.70



$Pr(A) = (.10)(.80) + \cdots + (.40)(.70) = .72$

$Pr(A') = (.10)(.20) + \cdots + (.40)(.30) = .28$

$= 1 - .72$

Notice that $.60 < Pr(A) < .90$ and $.10 < Pr(A') < .40$.

$Pr(C2\,|\,A) = \dfrac{Pr(C2 \text{ and } A)}{Pr(A)} = \dfrac{.18}{.72} = .25$

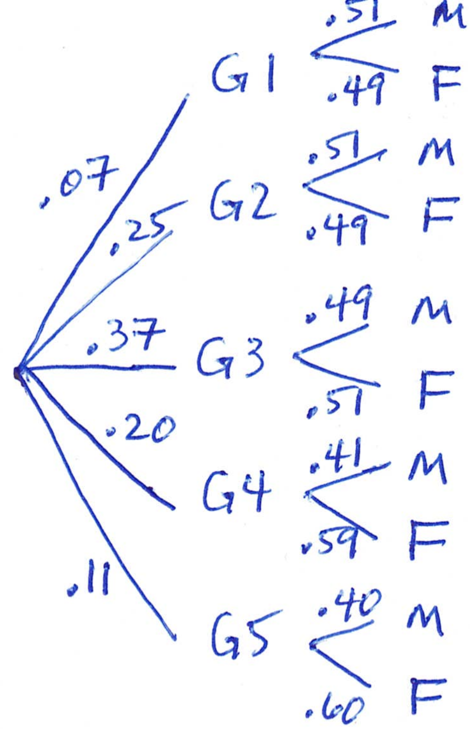
Why does it makes sense that $Pr(C2|A) > Pr(C2)$?

$Pr(C3|A') = \dfrac{Pr(C3 \text{ and } A')}{Pr(A')} = \dfrac{(.30)(.40)}{.28} \approx .43$

Why does it make sense that $Pr(C3|A') > Pr(C3)$

2. Age and gender.

Group	Fraction of population	Fraction of this group that is male
G1 (0 – 5)	.07	.51
G2 (5 – 19)	.25	.51
G3 (20 – 44)	.37	.49
G4 (45 – 64)	.20	.41
G5 (65 –)	.11	.40



Prediction: $.40 < Pr(M) < .51$

$Pr(M) = (.07)(.51) + \cdots + (.11)(.40) = .4705$

$Pr(G1|M) = \dfrac{(.07)(.51)}{.4705} = .076$

$Pr(G2|M) = \dfrac{(.25)(.51)}{.4705} = .271$

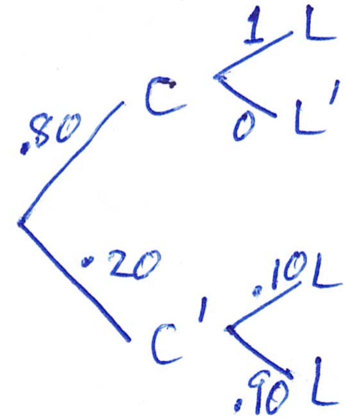
$Pr(G3|M) = \dfrac{(.37)(.49)}{.4705} = .385$

$Pr(G4|M) = \dfrac{(.20)(.41)}{.4705} = .174$

$Pr(G5|M) = \dfrac{(.11)(.40)}{.4705} = .094$

	Probability of being in group	
Group	If no info on gender	If person is male
G1 (0 – 5)	.07	↗ .076
G2 (5 – 19)	.25	↗ .271
G3 (20 – 44)	.37	↗ .385
G4 (45 – 64)	.20	↘ .174
G5 (65 –)	.11	↘ .094

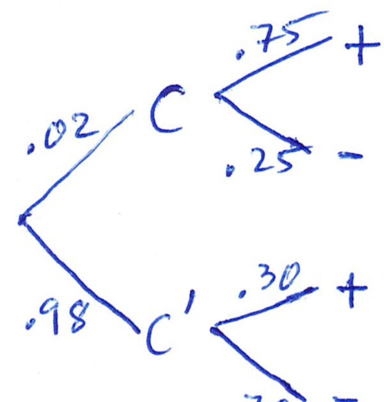
3. Approximately 10% of the population is left-handed. A person is on trial for a particular crime. The prosecution has proven with approximately 80% certainty that the defendant committed the crime (without using information about whether the defendant is left- or right-handed). In addition, the prosecution has proven that the person who did commit the crime is left-handed. The defendant is left-handed. With the additional information that crime was committed by a left-handed person and that the defendant is left-handed, how likely is it he actually committed the crime?



$$\Pr(C \mid L) = \frac{\Pr(C \text{ and } L)}{\Pr(L)}$$

$$= \frac{(.80)(1)}{(.80)(1) + (.20)(.10)} \approx .9756$$

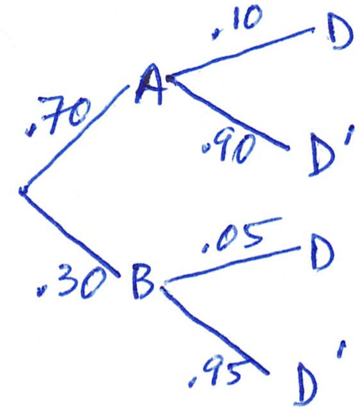
4. According to a NY Times article, about 2% of women aged 40 to 49 years old develop breast cancer during that decade of her life. But the mammogram used for women in that age group has a high rate of false positives and false negatives. The false positive rate is .30 and the false negative rate is .25. If a woman in her 40s has a positive mammogram test result, what is the probability that she actually has breast cancer? (More on medical testing next class.)



$$\Pr(C \mid +) = \frac{\Pr(C \text{ and } +)}{\Pr(+)}$$

$$= \frac{(.02)(.75)}{(.02)(.75) + (.98)(.30)} \approx .0485$$

5. 10% percent of the pens made by Apex are defective. Only 5% made by its competitor, B-ink, are defective. Since Apex pens are cheaper than B-ink pens, an office orders 70% of its stock from Apex and 30% from B-ink. A pen is chosen at random and found to be defective. What is the probability that it was produced by Apex?



$$\Pr(A \mid D) = \frac{\Pr(A \text{ and } D)}{\Pr(D)}$$

$$= \frac{(.70)(.10)}{(.70)(.10) + (.30)(.05)} \approx .82$$