Define events using probability notation

Express all the percents on the previous slide using probability notation.

- Let M denote the age of the mother and B denote whether the birth was intended. Then we can define the events on the previous slides as:
  - P(M = teen) = 0.09
  - P(M = young adult) = 0.24
  - P(M = older adult) = 0.67
  - P(B = intended|M = teen) = 0.23
  - P(B = intended|M = young adult) = 0.5
  - P(B = intended|M = older adult) = 0.75

In class (Sept 26) we talked about how to estimate the relevant marginal probability P(B=unintended) from the information provided. Below we go over the specific calculations using the rules of probability directly. The slides cover the frequencies and tree methods, but they just are examples of intuitive approaches to using these probability rules.

Getting marginal probabilities of single variable from joint probabilities (probabilities of two variables)

Let's say I have two discrete random variables, X and Y. I'm given the joint probabilities: *P(X=x and Y=y)*, for all values of $X \in \{1,2,3,4,5\}$ and $Y \in \{1,2,3,4\}$, that is, these are the possible values of *X* and *Y*. These are so-called *joint* probabilities (that is probability of more than one variable assuming values), so if the context is sampling from a large target population, then *P(X=x and Y=x)* represent the *marginal* proportion of observations with the specific values (e.g., P(*X=2* and *Y=3)*= proportion of individuals in targeted population with both *X=2* and *Y=3)* . In the example above, it would be the set of the probabilities of the 6 possible groups defined by a combination of age of the mother (teen, young adult, older) and intention of the pregnancy (intended, unintended).

To calculate the marginal probability of a single variable, e.g., P(X=x) from the joint probabilities, P(X=x and Y=y) is derived by summing up the joint probabilities where X=x:

$$P(X=x)=\sum_{y=1}^{4} P(X=x, Y=y) \quad \text{(Equation 1)}$$

However, this is not the most convenient form for calculating the *P(B=intended*) in the above example, because we were not provided the joint marginal joint probabilities (e.g., *P(B=b,M=m)*). However, we can use the *generalized multiplication rule* on slide 19 in Lecture (Day) 12:

$$P(A \text{ and } B) = P(A|B)P(B) = P(B|A)P(A)$$

Note, this is for when the two random variables, A and B, are binary, so can be represented as A vs. A^C and B vs. B^C. For our *X,Y* example the joint probability for X and Y can be written as:

$$P(X=x \text{ and } Y=y)=P(X=x|Y=y)P(Y=y)=P(Y=y|X=x)P(X=x).$$

If one only has the marginal probabilities of *P(Y=y)* for every y and the conditional probabilities, *P(X=x|Y=y)*, then one can re-write Equation 1 as:

$$P(X=x)=\sum_{y=1}^{4} P(X=x \mid Y=y)P(Y=y) \quad \text{(Equation 2)}$$

Looking at the example above, this means that we can write the *marginal* probability of unintended pregnancies (nothing that *P(B=Unintended|M=m)=1-P(B=Intended|M=m)*) as:

$$P(B=Unintended)=\sum_{m=\{teen,younger,older\}} P(B=Unintended \mid M=m)P(M=m)$$
$$=(1-0.23)*0.09+(1-0.50)*0.24+(1-0.75)*0.67=0.36,$$

or approximately 36% (Lec 13, page 8 gives the answer at one more decimal place, or 35.7%).

Diagnostic Testing Example

We went over how to calculate the positive predictive in class using the tree and frequencies approach. However, one can also apply the rules discussed above

We have two variables, cancer status (C), and test results (T), both binary. We were given the following:

- P(T=positive|C=cancer) = 0.9 (sensitivity)
- P(T=negative|C=not cancer) = 0.9 (specificity)
- P(C=cancer) = 0.01 (and thus P(C=not cancer)=0.99).

Our goal is to estimate P(C=cancer|T=positive) or the positive predictive value (PPV).

We know from the *generalized multiplication rule* we have:

$$P(C = cancer \mid T = positive) = \frac{P(C = cancer \text{ and } T = positive)}{P(T = positive)}.$$

We also know that we can use the same rule to represent the numerator by another conditional probability times a marginal, in this case: P(C=cancer and T=positive) = P(T=positive|C=cancer)P(C=cancer), so plugging this in we get Bayes Theorem representation or:

$$P(C = cancer \mid T = positive) = \frac{P(T = positive \mid C = cancer)P(C = cancer)}{P(T = positive)}$$

We have the info above to calculate the numerator, we just need to figure out how to calculate the denominator. Using how we did this above (Equation (1)).

$$P(T = positive) = P(T = positive \text{ and } C = cancer) + P(T = positive \text{ and } C = \text{not cancer})$$

Finally, we have not been given these directly but can again use the generalized multiplication rule to re-formulate as in Equation 2:

$$P(T = positive) = P(T = positive \mid C = cancer)P(C = cancer) + P(T = positive \mid C = \text{not cancer})P(C = \text{not cancer})$$

We are almost there! We just need to notice that:
P(T=positive|C= not cancer) = 1-P(T=negative | C=not cancer) = 1-specificity.

So, writing it all out in terms of what we were given, we get:

$$P(C = cancer \mid T = positive) = \frac{P(T = positive \mid C = cancer)P(C = cancer)}{P(T = positive \mid C = cancer)P(C = cancer) + P(T = positive \mid C = \text{not cancer})P(C = \text{not cancer})}$$

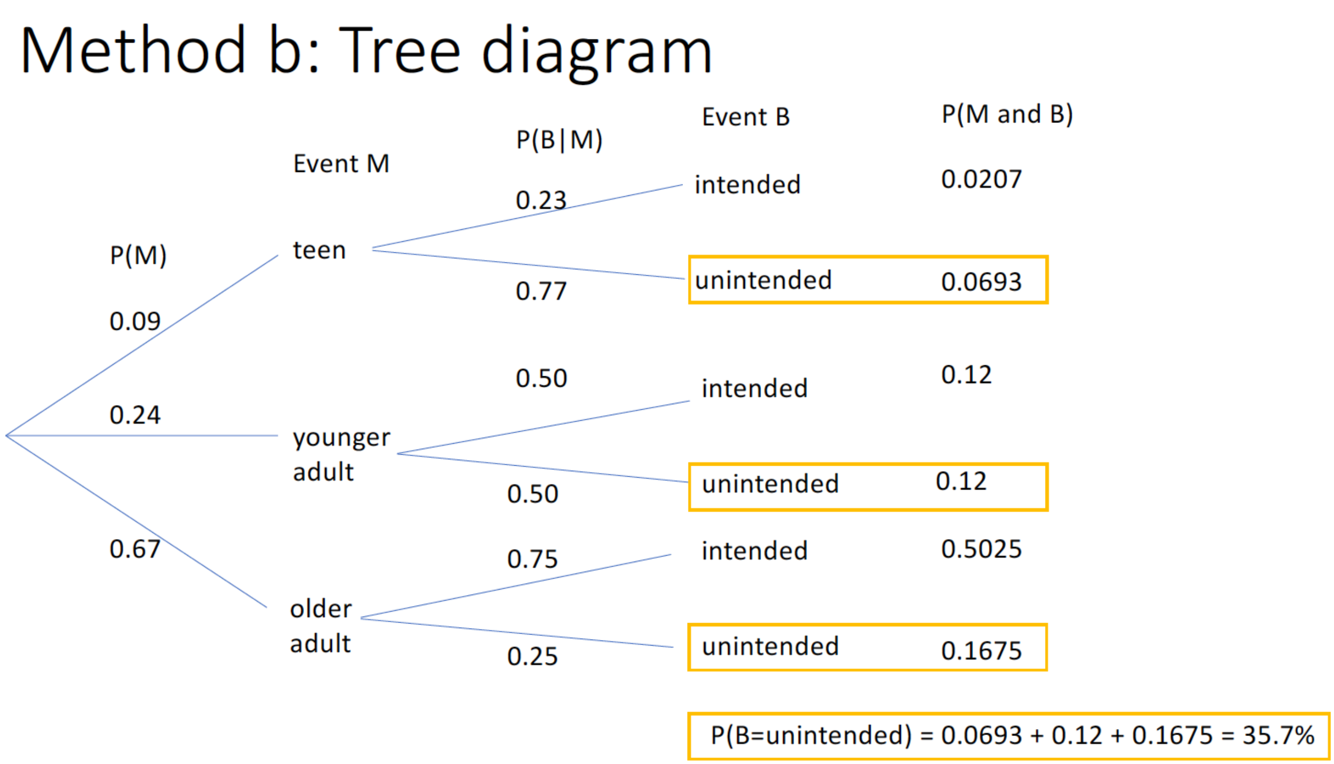
Now, let's plug and chug:

$$P(C = cancer \mid T = positive) = \frac{0.90 * 0.01}{0.90 * 0.01 + 0.10 * 0.99} = 0.083$$

or 8.3%.

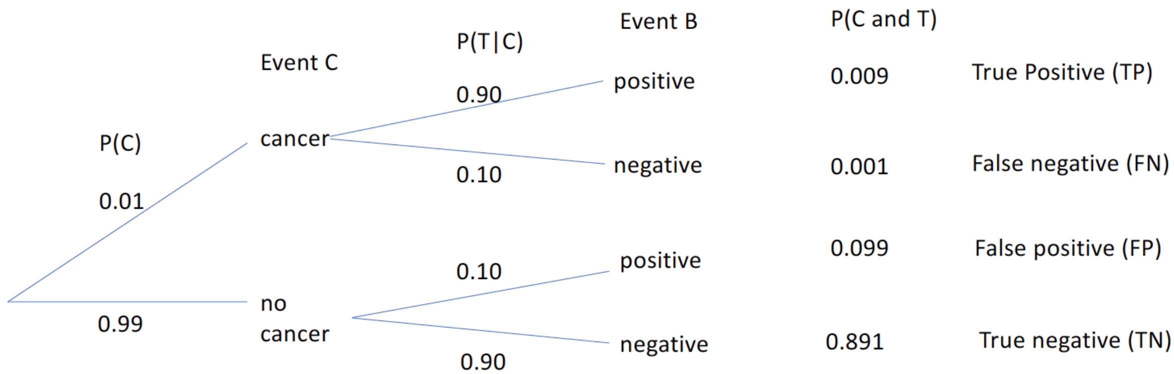
Below are some annotated slides that connect the tree methods to the basic probability rules.

<u>Slide 11 in Lecture 13</u>



$$P(B = Unintended) = \sum_{m=\{teen, younger, older\}} P(B = Unintended \text{ and } M = m)$$

Method b: Tree diagram



P(C=cancer|T=positive) = P(cancer & test positive)/P(test positive)
= P(cancer & test positive)/[P(test positive & cancer) + P(test positive & no cancer)]
= P(true positive)/[P(true positive) + P(false positive)]
= 0.009/(0.009 + 0.099) = 8.3%

$$P(C = cancer\,|\,T = positive) = \frac{P(C = cancer \text{ and } T = positive)}{P(T = positive)} = 0.083,$$

with $P(C = cancer \text{ and } T = positive) = 0.009$, and

$P(T = positive) = P(T = positive \text{ and } C = cancer) + P(T = positive \text{ and } C = \text{not cancer}) = 0.009 + 0.099 = 0.108$