

Stat 400, section 2.4 Conditional Probability and Bayes' Theorem

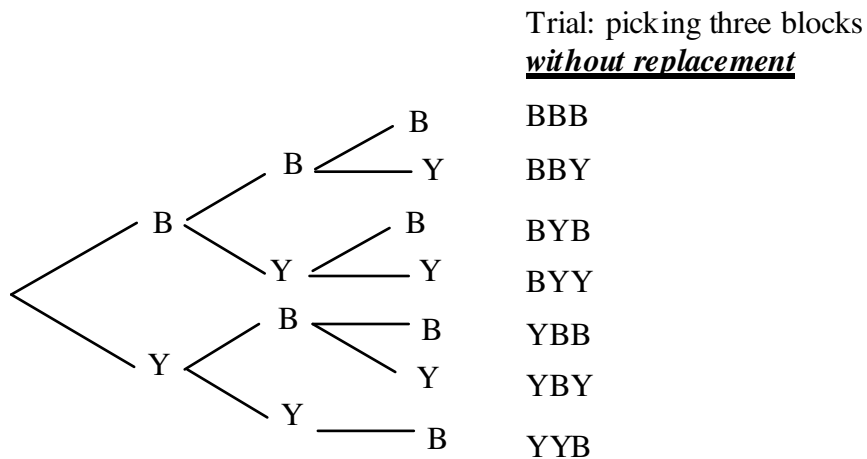
notes by Tim Pilachowski

In a **conditional probability** an outcome or event E is dependent upon another outcome or event F .

2.2 Example E revisited: Suppose that a box contains 3 blue blocks and 2 yellow blocks.

As long as we put each block back into the box after we pick, i.e. if we picked with replacement the probabilities do not change. $P(B) = \frac{3}{5}$ and $P(Y) = \frac{2}{5}$ whether it's the first pick, the second, the third or the hundredth pick.

The tree diagram for randomly picking three blocks without replacement, with associated probabilities, would look like this:



Let B_1 = blue on the first pick, B_2 = blue on the second pick, B_3 = blue on the third pick, Y_1 = yellow on the first pick, Y_2 = yellow on the second pick, Y_3 = yellow on the third pick. You can fill in the probabilities on the tree diagram above as we calculate them in class. Those marked with * below will be left for you to find on your own.

$$P(\text{picking blue second given a blue was picked first}) = P(B_2 | B_1) =$$

$$P(Y_2 | B_1) =$$

$$P(B_2 | Y_1) =$$

$$P(Y_2 | Y_1) =$$

$$*P(\text{picking blue third given a blue was picked first and second}) = P(B_3 | B_1 \cap B_2) =$$

$$P(Y_3 | B_1 \cap B_2) =$$

$$P(B_3 | B_1 \cap Y_2) =$$

$$* P(Y_3 | B_1 \cap Y_2) =$$

$$P(B_3 | Y_1 \cap B_2) =$$

$$* P(Y_3 | Y_1 \cap B_2) =$$

$$* P(B_3 | Y_1 \cap Y_2) =$$

$$* P(Y_3 | Y_1 \cap Y_2) =$$

In an “intersection”, i.e. an “and” situation, moving left-to-right on the tree diagram, multiply probabilities.

$$*P(\text{blue first and blue second and blue third}) = P(B_1 \cap B_2 \cap B_3) = P(B_1) * P(B_2 | B_1) * P(B_3 | B_1 \cap B_2) =$$

$$P(B_1 \cap B_2 \cap Y_3) =$$

$$P(B_1 \cap Y_2 \cap B_3) =$$

$$* P(B_1 \cap Y_2 \cap Y_3) =$$

$$P(Y_1 \cap B_2 \cap B_3) =$$

$$* P(Y_1 \cap B_2 \cap Y_3) =$$

$$* P(Y_1 \cap Y_2 \cap B_3) =$$

In a “union”, i.e. an “or” situation, moving up-and-down on the tree diagram, add probabilities.

$$*P(3 \text{ blue blocks}) = P(B_1 \cap B_2 \cap B_3) =$$

$$P(2 \text{ blue blocks and 1 yellow block}) =$$

$$*P(1 \text{ blue block and 3 yellow blocks}) =$$

$$*P(3 \text{ yellow blocks}) =$$

Note that the sum of the conditional probabilities in each “column” of the tree diagram add up to 1, i.e. the two possibilities for the first pick are the sample space for “pick one block”, the four possibilities for the (first intersection second) pick are the sample space for the “pick two blocks”, and the seven possibilities for the (first intersection second intersection third) pick are the sample space for “pick three blocks”.

$$P(B_1) + P(Y_1) = \quad * P(B_1 \cap B_2) + P(B_1 \cap Y_2) + P(Y_1 \cap B_2) + P(Y_1 \cap Y_2) =$$

$$P(B_1 \cap B_2 \cap B_3) + P(B_1 \cap B_2 \cap Y_3) + P(B_1 \cap Y_2 \cap B_3) + P(B_1 \cap Y_2 \cap Y_3) \\ + P(Y_1 \cap B_2 \cap B_3) + P(Y_1 \cap B_2 \cap Y_3) + P(Y_1 \cap Y_2 \cap B_3) + P(Y_1 \cap Y_2 \cap Y_3) =$$

For Example E revisited, we were easily able to count how many blocks were left. In other situations, we won’t be able to count so easily. So we have a formal definition and formula. Earlier, we used the multiplication principle, $P(E) * P(F | E) = P(E \cap F)$. With a little algebraic manipulation, we get

$$P(F | E) = \frac{P(E \cap F)}{P(E)} = \frac{\frac{N(E \cap F)}{N}}{\frac{N(E)}{N}} = \frac{N(E \cap F)}{N(E)}.$$

2.2 Example H revisited: In 2009, households were surveyed about health insurance coverage.

	Age 18-64	Age < 18	Totals
Public Plan	10286	10771	
Private Plan	47000	15914	
Uninsured	14144	1885	
Totals			100000

Health Insurance Coverage: Early Release of Estimates From the National Health Interview Survey, January–March 2010 by Robin A. Cohen, Ph.D., and Michael E. Martinez, M.P.H., M.H.S.A., Division of Health Interview Statistics, National Center for Health Statistics. (caveat: The report from which Example H was developed did not provide the numbers in the table above. Rather they were calculated from the given data, using an entirely fictional sample size of 100,000.)

Let F = individual is under 18; $P(F) =$

Let E = individual has health insurance; $P(E) =$

$E \cap F$ = individual is under 18 **and** has health insurance; $P(E \cap F) =$

$E \setminus F =$

$P(E \setminus F) =$

From the Stat 400 page you can link to a supplement, [conditional probability and Bayes' Theorem](#), which has an example worked out, then two scenarios with questions designed to help you in all of the concepts and procedures we've covered in chapter 2. Example 1 is a completely made-up scenario. Example 2 uses information from the Gallup organization and Example 3 uses information from the U.S. Bureau of Labor Statistics.

Example G: The report from which Example F was developed did not provide the numbers given in the table. Rather they were calculated from the summary data, using an entirely fictional sample size of 100,000. Example G presents information drawn directly from the same report, but for a different time period.

Percentage of persons who lacked health insurance coverage (Table 1), number of persons who lacked health insurance coverage (Table 2), and percentage of persons with public health plan or private health insurance coverage (Table 3), at the time of interview, by age group: United States, Jan–March 2010.

	Under 65 years	18-64 years	Under 18 years
Uninsured %	17.5%	21.5%	7.4%
Number uninsured (millions)	46.4	40.9	5.5
Public %	21.2%	14.4%	38.4%
Private %	62.7%	65.5%	55.5%

Health Insurance Coverage: Early Release of Estimates From the National Health Interview Survey, January–March 2010 by Robin A. Cohen, Ph.D., and Michael E. Martinez, M.P.H., M.H.S.A., Division of Health Interview Statistics, National Center for Health Statistics.

IMPORTANT: A footnote states, “A small number of persons were covered by both public and private plans and were included in both categories.” As a result, note that the percents in each column add up to greater than 100%.

Let U = persons in the U.S. under 65 years of age, and B = “uninsured” given “under 18”, C = “having public health insurance” given “under 18”, D = “having private health insurance” given “under 18”.

$P(\text{under 18 years}) =$

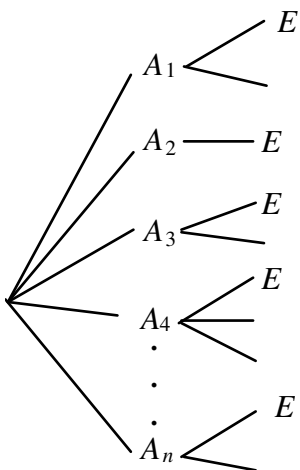
So, $P(C \cap D) =$

So, $P(C \cup D) =$

Let S = persons in the U.S. under 65 years of age . Let F = age under 18 years, so F' = age 18-64 years. Let E = lacking health insurance coverage, so E' = having public or private health coverage.

From the table: $P(\text{individual is uninsured given that she/he is under 18}) = P(E | F) =$

What if we wanted to know $P(\text{individual is under 18 given that she/he is uninsured}) = P(F | E)?$

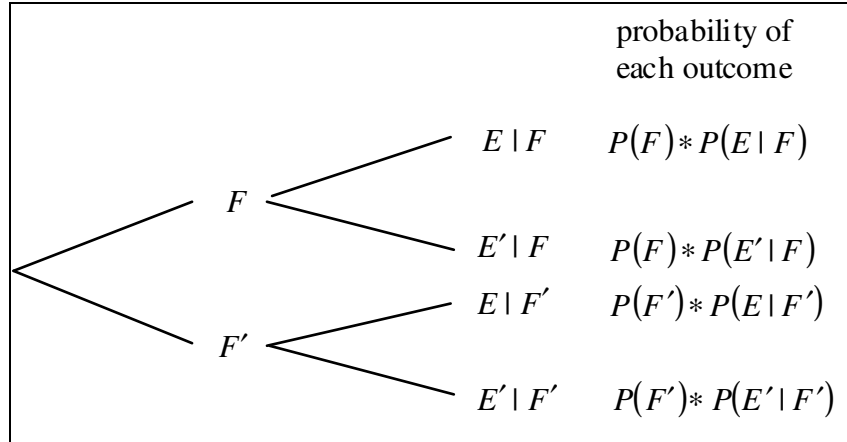


Theory: Bayes' Theorem

The second set of branches in a tree diagram represent conditional probabilities: $P(E | A_1)$, $P(E | A_2)$, $P(E | A_3)$, etc..

$$\begin{aligned}
 P(A_i | E) &= \frac{P(E \cap A_i)}{P(E)} \\
 &= \frac{P(E \cap A_i)}{P(E \cap A_1) + P(E \cap A_2) + \dots + P(E \cap A_n)} \\
 &= \frac{P(A_i) * P(E | A_i)}{P(A_1) * P(E | A_1) + P(A_2) * P(E | A_2) + \dots + P(A_n) * P(E | A_n)}
 \end{aligned}$$

Back to Example G: Let S = persons in the U.S. under 65 years of age . Let F = age under 18 years, so F' = age 18-64 years. Let E = lacking health insurance coverage, so E' = having public or private health coverage. (Fill in the probabilities as we find them.)



From the table, $P(E | F) =$ _____ , so $P(E' | F) =$ _____

From the table, $P(E | F') =$ _____ , so $P(E' | F') =$ _____

IMPORTANT: Complements were used, rather than adding percents from the table, because of the reporting anomaly noted above.

We need both $P(F)$ and $P(F')$, neither of which is given directly in the tables. However,

$$P(E | F) = \frac{N(E \cap F)}{N(F)} \Rightarrow$$

$$P(E) = \frac{N(E)}{N} \Rightarrow$$

$$P(F) =$$

$$P(F') =$$

$$\text{So, (finally)! } P(F | E) =$$

interpretation of $P(E | F)$:

interpretation of $P(F | E)$: