Calculus 131, section 12.3 Conditional Probability

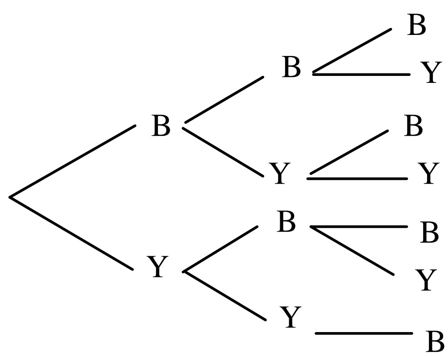
notes by Tim Pilachowski

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

12.2 Example D 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:



Trial: picking three blocks ***without replacement***

BBB
BBY
BYB
BYY
YBB
YBY
YYB

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(picking blue second given a blue was picked first) $= P(B_2 \mid B_1) =$

$P(Y_2 \mid B_1) =$ $\qquad$ $P(B_2 \mid Y_1) =$ $\qquad$ $P(Y_2 \mid Y_1) =$

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

$P(Y_3 \mid B_1 \cap B_2) =$ $\qquad$ $P(B_3 \mid B_1 \cap Y_2) =$ $\qquad$ * $P(Y_3 \mid B_1 \cap Y_2) =$

$P(B_3 \mid Y_1 \cap B_2) =$ $\qquad$ * $P(Y_3 \mid Y_1 \cap B_2) =$ $\qquad$ * $P(B_3 \mid Y_1 \cap Y_2) =$

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

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

*P(blue first and blue second and blue third) = $P(B_1 \cap B_2 \cap B_3) = P(B_1) * P(B_2 \mid B_1) * P(B_3 \mid B_1 \cap B_2) =$

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

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

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

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

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

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

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

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

P(2 blue blocks and 1 yellow block) =

*P(1 blue block and 3 yellow blocks) =

*P(3 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".

$P(B_1) + P(Y_1) =$

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

$$\begin{aligned} &P(B_1 \cap B_2 \cap B_3) + P(B_1 \cap B_2 \cap Y_3) + P(B_1 \cap Y_2 \cap B_3) + P(B_1 \cap Y_2 \cap Y_3) \\ &\quad + P(Y_1 \cap B_2 \cap B_3) + P(Y_1 \cap B_2 \cap Y_3) + P(Y_1 \cap Y_2 \cap B_3) + P(Y_1 \cap Y_2 \cap Y_3) = \end{aligned}$$

For Example D, 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. $P(E) * P(F \mid E) = P(E \cap F)$. With a little algebraic manipulation, we get

$$P(F \mid E) = \frac{P(E \cap F)}{P(E)} = \frac{\frac{n(E \cap F)}{n(S)}}{\frac{n(E)}{n(S)}} = \frac{n(E \cap F)}{n(E)}.$$

Definition: If two events E and F are **independent** then $P(E \mid F) = P(E)$ and $P(F \mid E) = P(F)$.
Using the multiplication principle, $P(E \cap F) = P(E) * P(F \mid E)$.
Then, if the two events are independent, $P(E \cap F) = P(E) * P(F)$.
Either of these can be used to prove or disprove independence of events.

12.2 Example F 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 F 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; $P(E \cap F) =$

$E \mid F =$

$P(E \mid F) =$

Are events E and F independent? How do we know?

12.2 Examples A, B and C revisited. Tossing two coins, tossing two dice, and picking two cards ***with replacement*** from a standard deck of 52 are examples of independent events. The result of the first toss/pick doesn't affect the result of the second. In statistical analysis, establishing whether or not events are independent is a primary part of showing correlation and/or causality.

On your own, show that $P(E \mid F) = P(E)$ for each of the following:

A: E = second coin is tails, F = first coin is tails

B: E = second die is even, F = first die is odd

C: E = picking an Ace, F = picking a Spade

Text practice exercises also ask questions about these, and other, similar scenarios.

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(under 18 years) =

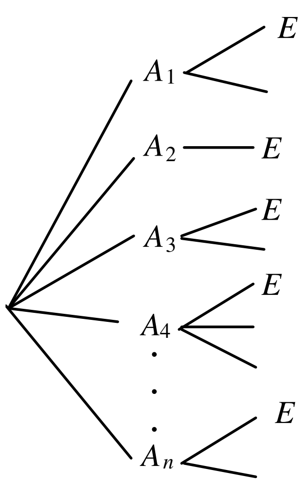
So, $P(C \cap D) =$

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(individual is uninsured given that she/he is under 18) = $P(E \mid F)$ =

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



Theory: Bayes' Theorem

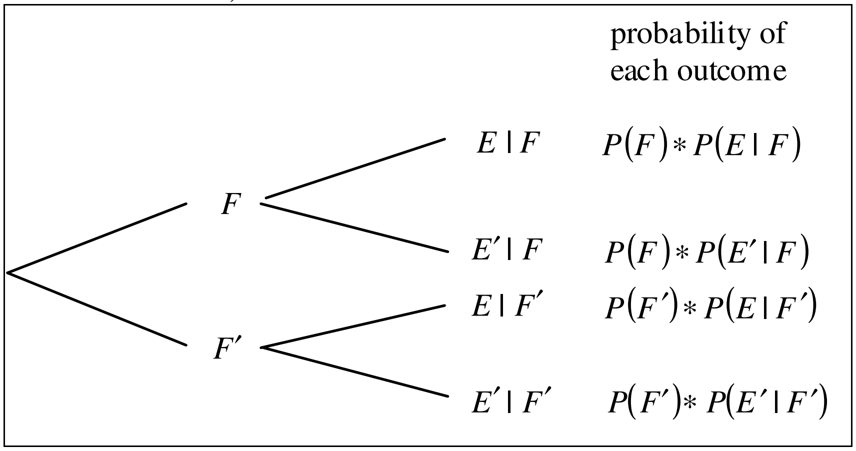
The second set of branches in a tree diagram represent conditional probabilities: $P(E \mid A_1)$, $P(E \mid A_2)$, $P(E \mid A_3)$, etc..

$$P(A_i \mid E) = \frac{P(E \cap A_i)}{P(E)}$$

$$= \frac{P(E \cap A_i)}{P(E \cap A_1) + P(E \cap A_2) + \ \dots \ + P(E \cap A_n)}$$

$$= \frac{P(A_i) * P(E \mid A_i)}{P(A_1) * P(E \mid A_1) + P(A_2) * P(E \mid A_2) + \ \dots \ + P(A_n) * P(E \mid A_n)}$$

Back to Example F: 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, $P(E \mid F) =$, so $P(E' \mid F) =$

From the table, $P(E \mid F') =$, so $P(E' \mid F') =$

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,

$$P(E \mid F) = \frac{n(E \cap F)}{n(F)} \quad \Rightarrow$$

$$P(E) = \frac{n(E)}{n(S)} \quad \Rightarrow$$

$P(F) =$

$P(F') =$

So, (finally)! $P(F \mid E) =$

interpretation of $P(E \mid F)$:

interpretation of $P(F \mid E)$: