Classroom Voting Questions: Statistics

General Probability Rules & Conditional Probability

1. In a certain semester, 500 students enrolled in both Calculus I and Physics I. Of these students, 82 got an A in calculus, 73 got an A in physics, and 42 got an A in both courses. Which of the following probabilities is the smallest? The probability that a randomly chosen student

   (a) got an A in at least one of the two courses.

   (b) got less than an A in at least one of the two courses.

   (c) got an A in both of the two courses.

   (d) got an A in calculus but not in physics.

   (e) got an A in physics but not calculus.

   Answer: (e). 42 students got an A in both courses. However, only 73 students got an A in physics, so 73 - 42 = 31 students got an A in physics but did not get an A in calculus. Similarly 82 - 42 = 40 students got an A in calculus but not physics.

   by Derek Bruff

   STT.04.05.010

   CC HZ MA336 S10: 0/3/72/0/**25**
   AS DH 3321 010 F14: 0/0/4/96/**0** time 4:00 ,
   CC KC MA315 F15: 0/0/44/6/**50** time 2:30
   CC KC MA315 F18: 0/3/42/3/**52**
   CC KC MA207 S19: 8/8/46/0/**38**

2. Three cards are placed in a hat–one card is blue on both sides, one card is red on both sides, and one card has one side blue and one side red. A card is drawn at random from the hat and you see that one side is blue. What is the probability that the other side is also blue?

   (a) 1/3

   (b) 1/2

   (c) 2/3

Answer: (b). There are two cards with a blue side. One of these cards is blue on the other side and one of these cards is red on the other side.

by Derek Bruff: Grinstead and Snell, Examples 4.7, 4.8, and 4.9

STT.04.05.020

AS DH 3321 010 S14: 30/**65**/5 time 3:10
AS DH 3321 010 F14: 0/**90**/10 time 3:10
AS DH 3321 010 S15: 12/**77**/12 time 3:10
AS DH 1333 020 F15: 29/**71**/0 time 4:00
AS DH 3321 010 F15: 60/**30**/10 time 3:00
CC KC MA315 F15: 12/**88**/0 time 1:00
AS DH 3321 010 F16: 48/**52**/0 time 3:30
AS DH 1342 010 F17: 21/**67**/13 time 3:30
AS DH 1342 020 F18: 50/**36**/14 time 3:40
AS DH 1342 040 S19: 7/**79**/14 time 2:30
AS DH 1342 030 F19: 76/**24**/0 time 3:50

3. Consider tossing a fair coin, that is, one that comes up heads half of the time and tails half of the time. Let A be the event "the first toss is a head," B be the event "the second toss is a tails," C be the event "the two outcomes are the same," D be the event "two heads turn up." Which of the following pairs of events is not independent?

   (a) A and B

   (b) A and C

   (c) A and D

   Answer: (c). Use the definition of independence ($P(A|B) = P(A)$ must be true) to see why.

   by Derek Bruff

   STT.04.05.030
   AS DH 3321 010 S14: 17/33/**50** time 3:30 ,
   AS DH 3321 010 F14: 0/17/**83** time 4:10 ,
   AS DH 3321 010 S15: 0/4/**96** time 3:20 ,
   AS DH 3321 010 F15: 4/60/**36** time 3:30 ,
   CC KC MA315 F15: 17/17/**66** time 2:00
   AS DH 3321 010 F16: 10/16/**74** time 4:00

4. Suppose A is the event that it rains today and B is the event that I brought my umbrella into work today. What is wrong with the following argument? "These events are independent because bringing an umbrella to work doesn't affect whether or not it rains today."

(a) These events are not independent, because one's decision of bringing an umbrella is dependent on the likelihood of rain. (However, rain is definitely not dependent on one carrying an umbrella although Murphy's Law might prove the opposite.)

(b) Although bringing an umbrella to work doesn't cause it to rain, given that you've brought your umbrella to work, the probability that it's a rainy day is higher than the chance of rain on any random day.

(c) These events are independent because the probability of bringing an umbrella to work doesn't affect the probability of the event its rains today and vice versa.

(d) It is false because the fact that it is raining today means that it was probably predicted to rain. If you checked that prediction then you would be more likely to bring in an umbrella making the events linked.

Answer: (b). Here's another example: Suppose you're running an ice cream stand. Let A be the event that it is a hot day and let B be the event that you sell more ice cream than usual. It seems clear that these events aren't independent since it's more likely that B will occur if A occurs–you'll sell more ice cream than usual on hot days. Mathematically, we know that if $P(B) \neq P(B|A)$, then $P(A) \neq P(A|B)$, but what does this mean in terms of the events? It doesn't mean that selling more ice cream than usual *causes* it to be a hot day. It does, however, mean that when you just look at days on which you sell more ice cream, the proportion of hot days is higher than the proportion of hot days to all days. That is, given that you're selling more ice cream than usual, the probability that it's a hot day is greater than the probability of any given day being hot. This is a great example, since causation creates the dependence in one direction, but not in the other direction.

by Derek Bruff

STT.04.05.040
AS DH 3321 010 S14: 22/**33**/11/33 time 3:00 ,
CC KC MA315 F15: 47/**10**/30/13 time 2:30

5. Assume that two events A and B are independent events. Which of the following statements is *false*?

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

   (b) $P(B|A) = [P(A|B) * P(B)]/P(A|B)$

   (c) A and B are mutually exclusive events.

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

   Answer: (c). (A) This is one definition of independent events.

   (B) If A and B are independent, then $P(B|A) = P(B)$, to which the right hand side of the equation simplifies.

(C)* correct Although the words independence and exclusive appear to be semantically-related in most peoples minds, the fact is that any mutually exclusive events cannot be independent. A simple example will serve to prove the point: Two mutually exclusive outcomes (Male and Female). Does knowing the one outcome (i.e. the sex of the person) tell you anything about the other outcome? Of course it does since all of the information in the occurrence of the outcome Female (yes or no) is contained in the outcome Male (if Male, then person is NOT female with probability 1). For those really clever students, I am excluding those people with biologically complex sexual markers for simplicity.

(D) As in (b), $P(A|B) = P(A)$ and $P(B|A) = P(B)$. In addition, $P(A) * P(B) = P(A \text{ and } B)$.

by Murphy, McKnight, Richman, and Terry

STT.04.05.050

CC HZ MA336 S10: 0/19/**68**/14
AS DH 3321 010 S14: 0/44/**56**/0 time 3:10 ,
AS DH 3321 010 F14: 0/61/**13**/26 time 3:30 ,
AS DH 3321 010 F15: 9/48/**0**/43 time 3:10 ,
AS DH 3321 010 F16: 0/65/**20**/15 time 3:10

6. Through accounting procedures, it is known that about 10% of the employees in a store are stealing. The managers would like to fire the thieves, but their only tool in distinguishing them from the honest employees is a lie detector test that is only 90% accurate. That is, if an employee is a thief, he or she will fail the test with probability 0.9, and if an employee is not a thief, he or she will pass the test with probability 0.9. If an employee fails the test, what is the probability that he or she is a thief?

   (a) 90%
   (b) 75%
   (c) 66 2/3%
   (d) 50%

   Answer: (d). Construct a tree diagram or a contingency table to see why.

   by Derek Bruff

   STT.04.05.060

   CC HZ MA336 S10: 31/47/9/**9**
   AS DH 3321 010 S14: 0/21/0/**79** time 5:30 ,
   AS DH 3321 010 F14: 5/32/63/**0** time 4:30 ,
   AS DH 3321 010 S15: 7/17/3/**72** time 6:00 ,
   AS DH 3321 010 F15: 14/32/14/**41** time 6:00 ,
   AS DH 3321 010 F16: 42/21/17/**21** time 8:00

CC KC MA207 S19: 53/27/20/**0**
CC KC MA315 S19: 25/0/0/**75**
CC KC MA315 S20: 22/12/0/**66**

7. A recent article in the Oklahoma Daily suggested that marijuana is a gateway drug for harder drug use. Suppose we have the following "facts". When asked, 90% of current "hard drug" users admit previously using marijuana; 40% of the general population admit using marijuana at some point during their lives; and 20% of the general population admit to using "hard drugs" at some point in their life. Given these three facts, what is the conditional probability of "hard drug" use given prior marijuana usage?

   (a) 0.16
   (b) 0.20
   (c) 0.25
   (d) 0.45
   (e) 0.90

   Answer: (d). Note: The next question is a different version of this question.

   This is a standard Bayes problem, in which the information presented is in the form of a retrospective probability: That is, when selecting on hard drug use, it is found that 90% have also previously used marijuana. In considering the diagnosticity of this probability, it is sometimes useful to consider that 100% of selected hard drug uses have also previously used water.

   In determining the correct answer, one can use several forms of the Bayes equation, but perhaps a table is more instructive. A hypothetical 2 by 2 table can be constructed, with the numbers filled in to be consistent with the "facts." For example, if we assume 100 people in the sample, then 40 would have used marijuana, and 20 would have used "hard drugs." 90% of the hard drug users admit to using marijuana, giving 18 who admit to both. The other numbers can now be filled in using arithmetic.

   Finally, the answer should now be more apparent. When selecting marijuana use, what is the probability of future hard drug use? There are 40 marijjuana users, and 18 of those have also used hard drugs. This gives $p = 18/40 = 0.45$, a number surely surprising and much less convincing than the $p = 0.90$ probability used to support the gateway theory. The number obtained when selecting on marijuana use is often called the prospective probability, and is more useful in assessing the causal efficacy on the risk factor being studied than the retrospective probability given in the question stem.

   by Murphy, McKnight, Richman, and Terry

   STT.04.05.070

   CC HZ MA336 S10: 3/3/13/**81**/0
   AS DH 3321 010 S14: 31/15/0/**23**/31 time 5:00 ,

AS DH 3321 010 F14: 17/17/17/**42**/8 time 7:00 ,
AS DH 3321 010 S15: 0/4/0/**81**/15 time 6:00 ,
AS DH 3321 010 F15: 6/0/65/**24**/6 time 6:00 ,
AS DH 3321 010 F16: 5/30/0/**60**/5 time 8:00
CC KC MA315 F18: 9/6/3/**50**/31
CC KC MA315 S20: 0/0/10/**85**/5

8. A recent article in the Oklahoma Daily suggested that marijuana is a gateway drug for harder drug use. The following fact which we will take as accurate - was used to support their argument: 9 out of 10 of "hard drug" users have previously used marijuana. Additionally, the newspaper also reported that 4 out of every 10 persons in the general population have admitted using marijuana and that 2 out of 10 persons in the general population have admitted partaking of "harder" drugs.

   You now find out that one of your children has used marijuana. What is the probability of your child subsequently using some "hard drug" based on the information presented above?

   (a) 0.16
   (b) 0.20
   (c) 0.25
   (d) 0.45
   (e) 0.90

   Answer: (d). Solution: This is the same answer as before, but with a different context embedded within the stem. This context has two features that distinguish it from the previous question: 1) the information is presented in a frequency format, which has been suggested (Gigerenzer, 19xx) as providing a more natural understanding of the data; and 2) a real situation that often arises is included as part of the problem - should you be concerned about your child?

   by Murphy, McKnight, Richman, and Terry

   STT.04.05.080
   AS DH 3321 010 S14: 0/0/0/**100**/0 time 2:00 ,
   AS DH 3321 010 F14: 18/5/14/**64**/0 time 1:30 ,
   AS DH 3321 010 F15: 0/5/0/**74**/21 time 1:00 ,

9. A cab was involved in a hit and run accident at night. Only two cab companies, the Transporter and the Rock, operate in the city. You are given the following data:

   (a) 85% of the cabs in the city are Transporters and 15% are Rocks.

(b) A witness identified the cab as a Rock. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two cabs 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was indeed a Rock?

(a) 0.75
(b) 0.41
(c) 0.27
(d) 0.63
(e) 0.80

Answer: (b). Another Bayes problem made famous by a Nobel Prize-winning Psychologist (Daniel Kahneman) and his lifelong collaborator (Amos Tversky). In this example, an eyewitness is fairly accurate (80%) so most people are inclined to believe that the Cab involved in the accident was most likely a Rock. However, most people also forget to properly account for the base rate of cabs in the population; in this case study, most cabs are not Rocks, but Transporters. This suggests that the prior probability of the cab being a Transporter is much higher than that of the Rock. Is the evidence, the identification of the cab from a witness known to be 80% accurate, enough to overcome the strong prior probability of the cab being a Transporter.

Again, we use the Table approach, with a hypothetical table completed to be consistent with the facts given in the question stem.

Note that in this problem, we are given the following pieces of information: 85% of the cabs are Transporters (85 out of 100) and the eyewitness is 80% accurate in identification (in the same conditions, the eyewitness would correctly identify 68 out of 85 Transporter cabs and 12 out of 15 Rock cabs, should the experiment be repeated 100 times with 85% of the experimental trials containing Transporter Cabs).

Now that we have 1 set of marginal totals (base rate of cabs), and 2 cells (68 and 12) reflecting the reported accuracy rate, we can complete our hypothetical table. The problem can now be phrased as follows: if the eyewitness claims the cab was a Rock, what is the probability of a Rock actually being the offending cab? This posterior probability can be calculated as 12/29, or an astounding low .413. As Kahneman and Tversky put it, Something about this result unsettles the average human being. What did they mean by this?

by Murphy, McKnight, Richman, and Terry

STT.04.05.090

CC HZ MA336 S10: 3/**7**/13/13/60
AS DH 3321 010 S14: 0/**38**/0/38/23 time 5:10 ,
AS DH 3321 010 F14: 0/**62**/4/14/19 time 6:20 ,
AS DH 3321 010 S15: 0/**76**/0/7/17 time 5:30 ,

AS DH 3321 010 F15: 0/**75**/0/25/0 time 5:20 ,
AS DH 3321 010 F16: 11/**17**/33/6/33 time 6:30
CC KC MA315 F18: 0/**90**/0/0/6
CC KC MA207 S19: 12/**42**/27/4/15
CC KC MA315 S19: 10/**50**/15/10/15

10. The following contingency table/two-way table classifies the members of a certain government into political party (Liberal or Conservative) and whether they support or oppose the spending bill that is currently up for adoption.

	Support	**Oppose**	**Total**
Liberal	47	11	58
Conservative	14	35	49
Total	61	46	107

Imagine randomly selecting one member of the government. Let L, C, S, and O denote the events of selecting a liberal, a conservative, a bill supporter, and a bill opposer, respectively. Find $P(C \,\&\, S)$.

(a) $\frac{14}{61}$
(b) $\frac{14}{49}$
(c) $\frac{14}{107}$
(d) None of the above

Answer: (c).

(a) This $P(C|S)$.
(b) This is $P(S|C)$.
(c) Correct.
(d) Incorrect. (c) is the correct answer. (See above.)

by David A. Huckaby

STT.04.05.100
AS DH 3321 010 F16: 3/3/**94**/0 time 2:00

11. The following contingency table/two-way table classifies the members of a certain government into political party (Liberal or Conservative) and whether they support or oppose the spending bill that is currently up for adoption.

	Support	Oppose	Total
Liberal	47	11	58
Conservative	14	35	49
Total	61	46	107

Imagine randomly selecting one member of the government. Let L, C, S, and O denote the events of selecting a liberal, a conservative, a bill supporter, and a bill opposer, respectively. Find $P(S|L)$.

(a) $\frac{47}{61}$

(b) $\frac{47}{58}$

(c) $\frac{47}{107}$

(d) None of the above

Answer: (b).

(a) This $P(L|S)$.

(b) Correct.

(c) This is $P(L \,\&\, S)$.

(d) Incorrect. (b) is the correct answer. (See above.)

by David A. Huckaby

STT.04.05.110
AS DH 3321 010 F16: 3/**97**/0/0 time 1:50

12. A sample of sports fans from Canada and the Unites States were asked whether they would prefer to attend a professional basketball game or a professional (ice) hockey game. The following table gives a joint probability distribution for the sample.

	Basketball	Hockey
Canada	0.026	0.312
United States	0.512	0.150

Imagine randomly selecting one member of this sample. Let C, S, B, and H denote the events of selecting someone from Canada, someone from the United States, someone who chose basketball, and someone who chose hockey, respectively. Find $P(C|H)$.

(a) 0.312

(b) 0.675

(c) 0.923
(d) None of the above

Answer: (b).

(a) This simply $P(C \,\&\, S)$, which is provided in the given distribution.
(b) Correct: $P(C|H) = \dfrac{P(C \,\&\, S)}{P(H)} = \dfrac{0.312}{0.312 + 0.15} = 0.675$
(c) This is $P(H|C) = \dfrac{P(C \,\&\, S)}{P(C)} = \dfrac{0.312}{0.312 + 0.026} = 0.923$
(d) Incorrect. (b) is the correct answer. (See above.)

[Follow-up question: Just by eyeballing the numbers in the given distribution, can you explain why an answer of just over two-thirds makes sense?]

by David A. Huckaby

STT.04.05.120
AS DH 3321 010 F16: 33/**64**/3/0 time 3:00

13. A woman is the victim of a homicide, and her husband is on trial for her murder. It is known that prior to her murder, her husband had verbally threatened to kill her. During the trial, the defense attorney tells the court, "Only 1% of all men who threaten to kill their wives actually go on to kill them." There are, of course, many other pieces of evidence presented in the trial, but let's focus on this statement made by the defense attorney. TRUE or FALSE: The attorney's statement is a significant piece of evidence in favor of the man's innocence.

(a) True, and I am very confident.
(b) True, and I am not very confident.
(c) False, and I am not very confident.
(d) False, and I am very confident.

Answer: (d). The attorney's statement is that the probability that a man has killed his wife given that he threatened to kill her is 0.01. This probability is far from being the most relevant: The woman in question has been killed! Much more relevant is the probability that a man has killed his wife given that he threatened to kill her *and* she has been killed by someone. Without knowing this probability, one has a feeling it is *much* higher than 1%. [Follow-up question: Can you provide some intuition on why this more relevant probability is much higher? Answer: Say the general murder rate of women, which varies by location, is 0.001 in this couple's area. So there is 1 chance in 1000 the woman will be murdered. But if her husband threatens to kill her, there is a 1 in 100 chance that he will kill her. So her chances of being murdered have increased dramatically, between 10-fold and 11-fold. (Why? If the original 1 in 1000 probability was all due to her husband, the chances have gone up 10-fold, from 1 in 1000 to 1 in 100. If, at the other extreme, the original 1 in 1000 probability was all due to people other than her husband, the chances have gone up 11-fold, from 1 in 1000 to 11 in 1000. The truth is somewhere in between, close to the latter one would hope. So let's assume that's the case, that is, that the husband's contribution to the original 1 in 1000 probability is very small. This assumption only helps the defendant.) She is now found murdered. It's roughly 10 times more likely that she was murdered by her husband than that she was murdered by someone else. Indeed, there was an 11 out of 1000 chance she was going to be murdered. Of that 11 out of 1000, roughly 10 out of 1000 was due to her husband murdering her, while only roughly 1 out of 1000 was due to the possibility of someone else murdering her. So based on this information, the probability that her husband murdered her is roughly $\frac{10}{11} = 0.91$. (Note that when we say "roughly" we are using the assumption that the possibility of her husband's killing her accounted for a very small part of the original 1 in 1000 chance of her being murdered.) Drawing a figure to illustrate these ideas would be helpful if this result is not yet intuitive. (Can you find any hidden assumptions we are making? One assumption is that the (roughly 1 in 1000) probability that someone other than her husband murders the wife doesn't increase when her husband threatens to kill her. One might instead argue that if he threatens to kill her the probability is higher that she is an unlikeable person and that the probability that she will be murdered by someone other than her husband has therefore increased just like the probability that she will be murdered by her husband has increased. In the absence of supporting data, however, we have the feeling that the "problem" is much more with the threatening husband than with the wife's alleged unlikeability, and that even if we grant an increase in ill-will toward her from others, we sense that it would be dwarfed by the increase in her husband's ill-will. If possible, we could try to obtain data and do a study to shore up our intuition on this.)

More a follow-up project than a follow-up question: Can you make the above reasoning more formal? One possible approach: Let T be the event a man threatens to kill his wife, K the event he kills her, and B the event that she has been killed. The attorney gave $P(K|T) = 0.01$. The more relevant probability is $P(K|T\,\&\,B) = \frac{P(K\,\&\,T\,\&\,B)}{P(T\,\&\,B)} = \frac{P(K\,\&\,T)}{P(T\,\&\,B)}$. The last step follows because K is a subset of B. This can be further rewritten as $\frac{P(T)P(K|T)}{P(T)P(B|T)} = \frac{P(K|T)}{P(B|T)} = \frac{0.01}{P(B|T)}$. Can we reasonably estimate $P(B|T)$? One technique worth trying is to rewrite the event B as $B = K \cup (B - K)$. Then $P(B|T) = P(K|T) + P(B - K|T) = 0.01 + P(B - K|T)$. So altogether we have $P(K|T\,\&\,B) = \frac{0.01}{0.01 + P(B-K|T)}$. What can you say about $P(B - K|T)$? (Suggestions to consider: Can you invoke independence and say that $P(B - K|T) = P(B - K)$? Can you say that $P(B - K|T) < P(B - K)$? Note that if $P(B - K|T) \le P(B - K)$, then $P(K|T\,\&\,B) = \frac{0.01}{0.01 + P(B-K|T)} \geq$

$\dfrac{0.01}{0.01 + P(B - K)}$. In this connection, note that $P(B - K) < P(P)$, so that the fraction would then be bounded below by $\dfrac{0.01}{0.01 + P(B)}$. Or is it perhaps true that $P(B-K|T) > P(B-K)$, but that equality almost holds?) Let's say that you can somehow show that $P(K|T \,\&\, B) = \dfrac{0.01}{0.01 + P(B - K|T)}$ is commensurate with or bounded below by $\dfrac{0.01}{0.01 + P(B)}$. If $P(B)$ is much smaller than 0.01, then this fraction—and therefore $P(K|T \,\&\, B)$—will be large, that is, much closer to 1 than one might think. (You can see that achieving this result depends on showing that the quantity in the place of $P(B - K|T)$ is small compared to 0.01. We have made suggestions on possibly relating it to the general murder rate $P(B)$, which would presumably be known. Perhaps you can argue more directly that $P(B - K|T)$ is small. Or perhaps you have data or could gather data and perform a study.) If we assume that $P(B) = 0.001$, as we did in our informal argument above, then we have that $P(K|T \,\&\, B)$ is roughly $\dfrac{0.01}{0.01 + P(B)} = \dfrac{0.01}{0.01 + 0.001} = \dfrac{0.01}{0.011} = \dfrac{10}{11} = 0.91$. This is very different than the 0.01 cited by the defense attorney. Indeed: Assuming our argument here is valid (Can you make it so by successfully following through on one of the suggestions given above?), by giving the jury this 1% figure, a figure that is so much higher than the general murder rate, the defense attorney was actually presenting evidence for the man's guilt, not his innocence. It would be a good exercise to create a Venn diagram drawn to scale in order to illustrate the ideas in this problem. Were you to present your ideas in court, a jury would probably have a much better chance of understanding a Venn diagram argument than this cascade of symbols we have just tumbled down.]

by David A. Huckaby

STT.04.05.130
AS DH 3321 010 F16: 0/7/28/**66** time 3:00