MAT 3701: Selected solutions to Problem Set 3.

February 14, 2013

1. First, I must apologize for an incorrect (meaning non-standard) definition. *Mutually independent* is generally taken to mean that for the intersection of *any* subcollection of the events, the probability of the intersection is equal to the product of the individual probabilities. So in the case of three events, A, B, and C, they are mutually independent if they are pairwise independent *and* $P(A \cap B \cap C) = P(A)P(B)P(C)$.

   (a) $P(A) = \frac{1}{2}$, since A contains two of the four equally likely outcomes. Similarly, $P(B) = P(C) = \frac{1}{2}$. $A \cap B = A \cap C = B \cap C = A \cap B \cap C = \{1\}$, so $P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$. Thus, a quick calculation shows the events to be pairwise independent ($\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$), but not mutually independent ($\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \neq \frac{1}{4}$).

   (b) Using the standard definition, it is of course impossible to find three events that are mutually independent but not pairwise independent (since the former property includes the latter). However, it is possible to find three events A, B, and C that are not pairwise independent, but such that $P(A \cap B \cap C) = P(A)P(B)P(C)$. If you didn't find an example, keep trying! (Hints: Remember that mutually *exclusive* events with positive probability are definitely *not* independent, and remember that the empty set is an event, and $P(\emptyset) = 0$.

   (c) Observe that the probability of picking a point from a subset of S is its area (since the area of the whole square is one). Event A is the rectangle $[0, \frac{1}{2}] \times [0, 1]$; its area is $\frac{1}{2}$. Event B is the rectangle $[0, 1] \times [0, \frac{1}{4}]$; its area is $\frac{1}{4}$. Event $A \cap B$ consists of the rectangle $[0, \frac{1}{2}] \times [0, \frac{1}{4}]$, which has area $\frac{1}{8}$. So by definition, A and B are independent. This result should be intuitively clear: for each x-coordinate, the same interval of y-coordinates lies in the sample space, so restricting the x-coordinates won't make a difference to the probability of an event defined by the y-coordinates.

2. It is somewhat easier to first consider the probability that a given atom survives. Since the half-life is 8 days, it has a 50% chance of surviving for 8 days. If it does survive for 8 days, it has once again a 50% chance of surviving the next 8. For the atom to survive 16 days, both events must occur. Thus its chances of surviving the full 16 days is $(50\%)(50\%) = 25\%$. Therefore, the probability that it decays within 16 days is $100\% - 25\% = 75\%$.

   Since we know the probability an atom will survive for 8 days is 50%, we can compute the probability it will survive for 4 days by breaking this 8-day period into two equal parts. Let p be the probability the atom survives the first 4 days. If it does, its probability of surviving the second 4 days is the same. So its probability of surviving for 8 days is p^2. Of course, we know this probability is $\frac{1}{2}$, so $p^2 = \frac{1}{2}$. Therefore, $p = \frac{1}{\sqrt{2}} = 2^{-\frac{1}{2}}$. (For this calculation, it was obviously easier to use fractional rather than percentage form.) Finally, the probability the atom will decay with 4 days time is $1 - p = 1 - 2^{-\frac{1}{2}}$.

We can do a similar analysis for any fraction of the half-life by dividing this time period into the appropriate number of equal parts. In general, it is clear now that if we use the half-life of the substance as our unit of time, then the probability the atom will survive t half-lives is 2^{-t}, and thus the probability it will decay within t half-lives is $1 - 2^{-t}$. (If we use days as the unit of time, then t half-lives $u = 8t$ days, so $t = \frac{u}{8}$, and the formula for the probability the atom will decay in u days is $1 = 2^{-\frac{u}{8}}$.)

Remark. Extending this analysis to irrational values of t requires us to use limits and the fact that the probability of the union of a sequence of nested events, each included in the next, is the limit of their probabilities, as proven in the previous problem set.

Remark. It is not necessary to use the half-life as the unit of time; we can use any amount of time for which the probability of survival is known. The base of the exponential expression will then be different. For example, if the probability an atom of a radioactive substance survives 3 days is known to be $\frac{3}{11}$, the the probability it will survive $3t$ days is $(\frac{3}{11})^t$. (In natural exponential form this would be e^{-kt}, where $k = \ln \frac{11}{3}$.)

To compute the probability an atom will decay between times t_1 and t_2, consider the following three events: Let E_1 be the event that the atom decays by time t_1 (including t_1). Let E_2 be the event that it decays after time t_1 but by time t_2; that is, E_2 is the event whose probability we want to find. Note that it does not matter which, if any, of the endpoints t_1 and t_2 are included in event E_2, since the probability the atom decays *exactly* at a particular instant must be 0. (Why? Can you prove this?) Let E_3 be the event that the atom decays sometime by time t_2.

Events E_1 and E_2 are *mutually exclusive.* In other words, we take our sample space to consist of the non-negative real numbers specifying the times at which the atom could decay, and subsets E_1 and E_2 of the sample space are disjoint. In short, these events cannot occur simultaneously. Therefore the probability of their union is the sum of their probabilities, by the countable additivity of probability measure.

Furthermore, the union of events E_1 and E_2 is E_3. Therefore, the sum of their probabilities is the probability of E_3.

Since the probabilities of E_1 and E_3 are known, we can compute $P(E_2) = P(E_3) - P(E_1)$. I leave the calculation to you.

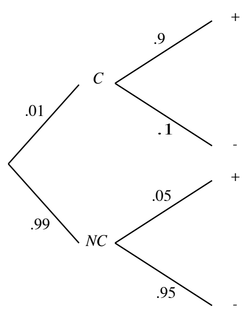
Remark. Let T be the time (that is, instant) at which the atom actually decays. (It is traditional in probability theory to use capital letters for variables that depend on the outcome of an indeterminate experiment. As we now know, they are called *random variables.*) Observe that $P(E_1)$ is the probability that $t \leq t_1$, and $P(E_3)$ is the probability that $t \leq t_2$. Thus, our computation may be denoted $P(t_1 < T \leq t_2) = P(T \leq t_2) - P(T \leq t_1)$. In general, the probability of any interval can be computed if the probabilities of all intervals of the form $T \leq t$ are known. This standard method proves to be so useful that the probabilities of events of the from $T \leq t$ are given a special name. Let $F(t) = P(T \leq t)$. As we now know, the function F is called the *cummulative distribution function of T.*

3. The similarity to the previous problem should be obvious! This is an example of the same mathematical model applying to various real situations.

4. Here is the easy way. Since each ball is equally likely to be removed from the box, and we know nothing about the three balls removed first, each ball is equally likely to be the *fourth* ball removed from the box! Hence the probability that it is white is just $\frac{4}{10} = .4$.

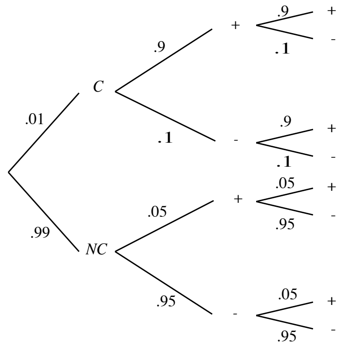
5. None of these questions can be answered without knowing further information, such as the prevalence of cancer in the population to which the person being tested belongs. For a dramatic illustration of this, suppose you belong to a population in which everyone has cancer (say, you are a patient at the Sloan-Kettering Cancer Center). In other words, you already *know* you have cancer. Then there can be no false positives; the probability you have cancer given a positive test result is 100% (and, for that matter, the probability you have cancer given a negative test result is 100%, too!). On the other hand, suppose you belong to a population that is completely free from cancer. (Obviously, such a population exists only in theory.) Then every positive will be a false positive; even if you test positively, the probability you have cancer is 0. I invite you to try populations with different cancer rates between these extremes and see that the probability of actually having cancer given a positive test result varies with the cancer rate.

6. Now we have enough information to construct the sample space:



The probability that a person chosen at random has cancer and gets a positive result is $(.01)(.9) = .009$. (Notice that this is small, because the cancer rate is pretty low.) The probability that a person does not have cancer and gets a positive result is $(.99)(.05) = .0495$. (Notice that this is larger, because even though most people without cancer test negative, there is a high percentage of people without cancer.) The probability of simply getting a positive result is the sum of these two mutually exclusive events: .0504. So the probability that someone who tests positively actually has cancer is the measure of portion of this event consisting of those who have cancer and get a positive result: $\frac{.009}{.0504} \simeq 18\%$. Obviously, the test is good enough that if you test positively your chances of having cancer are higher than if the test result weren't known. The test adds diagnostic value. However, the fact that you test positive does not mean you probably have cancer (if you live in this fairly healthy population).

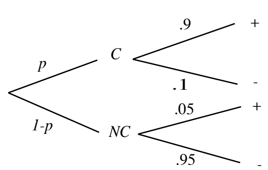
Diagnostic value is added by doing the test twice, because the probability of a double false positive is lower than the probability of a single false positive. We construct the sample space as follows:



We see that the probability of two positive tests is (.01)(.9)(.9) + (.99)(.05)(.05), whereas the measure of the portion of this event consisting of those who have cancer is (.01)(.9)(.9). So the probability a person has cancer given two positive results is $\frac{(.01)(.9)(.9)}{(.01)(.9)(.9)+(.99)(.05)(.05)} \simeq 77\%$. Two positive tests indicates that you probably have cancer, although there is still a pretty big - about 23% - chance that you don't.

Using our first sample space, we find that the probability that someone who tests negatively actually has cancer (false negative) is $\frac{(.01)(.1)}{(.01)(.1)+(.99)(.95)} \simeq .1\%$. So a negative test result is quite reassuring, reducing one's chance of having cancer by a factor of 10.

7. Construct the sample space as follows, using an unknown cancer rate, p:



$p(.9) + (1 - p)(.05) = .1$. Solve for p.

8. We want the failure rate to be less than 1%. All n components fail with probability $(.05)^n$, so we need $(.05)^n < .01$. The smallest integer n for which this is true is 2. Two components are needed.

9. The probability that the other drawer contains a gold coin is $\frac{2}{3}$, because each drawer is equally likely to be picked, and two of the three drawers with gold coins are in the first chest. Your chances are not 50-50 because, if you have picked a gold coin, you are more likely to be in the first chest.

   A diagram of the sample space confirms and vividly illustrates this answer:

G	*G*	*S*
G	*S*	*S*

10. The contestant should always switch, because the chance of winning if you switch is $\frac{2}{3}$, whereas if you don't switch your chance of winning is clearly only $\frac{1}{3}$! To see this, suppose you picked one of the goats on the first try. Then the host has no choice but to reveal the other goat, so if you switch you will certainly win. In two out of the three equally likely scenarios, you have picked a goat and the remaining curtain hides the prize, whereas in only one of the three have you picked the prize and the remaining curtain hides a goat. In summary, in two out of the three equally likely scenarios, switching ensures you will win, whereas in only one of the three - the one in which you were right the first time - will switching cause you to lose.

11. Let $A'_n = A_n \setminus A_{n-1}$. Here are some facts in basic set theory:

    (a) $\bigcup_{i=1}^{n} A'_i = A_n$.

    (b) Hence, $\bigcup_{i=1}^{\infty} A'_i = \bigcup_{n=1}^{\infty} A_n$.

    (c) $A'_i \bigcap A'_j = \emptyset$ if $i \neq j$.

    You should be able to prove all of these completely and rigorously. If you have difficulty, come see me (or your Foundations instructor!).

    Using the facts above, we have:

    - $\rho(\bigcup_{n=1}^{\infty} A_n) = \rho(\bigcup_{i=1}^{\infty} A'_i)$, because the two input events are equal (fact (b)), and probability measure is a function.
    - $\rho(\bigcup_{i=1}^{\infty} A'_i) = \sum_{i=1}^{\infty} \rho(A'_i)$, by the countable additivity of probability measure, using fact (c).
    - $\sum_{i=1}^{\infty} \rho(A'_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \rho(A'_i)$, by definition of the sum of an infinite series. Note that this series must converge, because its terms are all positive and it is bounded above by 1.

- $\lim_{n\to\infty} \sum_{i=1}^{n} \rho(A'_i) = \lim_{n\to\infty} \rho(\bigcup_{i=1}^{n} A'_n) = \lim_{n\to\infty} \rho(A_n)$, by the countable additivity of probability measure and fact (a).

12. Use Problem 2 and the fact that $\bigcap_{n=1}^{\infty} A_n = (\bigcup_{n=1}^{\infty} A_n^c)^c$.

13. Since a point is being chosen at random via an (ideal) spinner, rotating an event must give you an event with the same probability. (The spinner does not favor any direction, so rotation cannot make a difference.) This is the key observation for doing this problem.

    If we take set A and rotate it by $n\alpha\pi$, we get a new set, call it A_n, disjoint from A (except for $A_0 = A$). This is because A contains exactly one point from each equivalence class, and A_n contains a *different* point from each equivalence class, namely the rotation of the point in A by $n\alpha\pi$. (Since α is irrational, the rotation of a point by $n\alpha\pi$ can never be the same point; if it were, that would mean $n\alpha\pi = 2\pi m$, for some integer m, which means $\alpha = \frac{2m}{n}$, which is rational.) Similarly, any two distinct rotations of A, A_n and A_m, $n \neq m$, are disjoint, because they each contain a different element from each equivalence class.

    Furthermore, $\bigcup_{i=-\infty}^{\infty} A_n = S$, where S is the entire circle, since these rotations cover every element in every equivalence class.

    Suppose $P(A)$ is defined. Then $P(A_n) = P(A)$ for each n (by the rotation principle), and $\sum_{i=-\infty}^{\infty} P(A_n) = P(S) = 1$, since the sets A_n are pairwise disjoint.

    Case 1: $P(A) = 0$. Then $0 + 0 + 0 + \cdots = 1$, which contradicts the fact that $0 + 0 + 0 + \cdots = 0$. (Since each partial sum is 0, the limit is 0.)

    Case 2: $P(A) \neq 0$. Then the partial sums of $P(A) + P(A) + P(A) + \cdots$ increase without bound, so $\sum_{i=-\infty}^{\infty} P(A_n) = 1$ again leads to a contradiction.

    Since every possible case leads to a contradiction, we must conclude that our initial supposition that $P(A)$ is defined is false.