1 Some Review questions on the basics of set theory and probability theory

September 10, 2010

1. Make sure you can define and explain the following terms: random variable, discrete random variable, continuous random variable, sample space, probability, events, event space, expected value, and ....

2. In both words and functional notation, what is a probability function? What kind of function is it?

3. I have a student who is doing research on diabetes. She tells me the following: "The probability that someone has diabetes multiplied by the probability that they are skinny given that they have diabetes equals the probability they are skinny multiplied by the probability that they have diabetes given they are skinny" Is she right? Yes or No, and explain. She also said, " The probability that a person is skinny is equal to the probability that a person is skinny conditional on them having diabetes." Yes or No, and explain.

   answer: The question is whether $P[skinny\,|diabetes]P[diabetes] = P[diabetes\,|skinny]P[skinny]$. Or in terms of A and B where A is skinny, whether it is correct that $P[A\,|B]P[B] = P[B\,|A]P[A]$. This equality follows directly from the definition of conditional probability, which is

   $$P[A\,|B] = \frac{P[AB]}{P[B]}$$

   Since there is nothing special about A and B, it follows that $P[B\,|A] = \frac{P[BA]}{P[A]}$ =. We also know that $AB = BA$, so $P[B\,|A] = \frac{P[AB]}{P[A]}$.So, solving $P[B\,|A] = \frac{P[AB]}{P[A]}$ and $P[A\,|B] = \frac{P[AB]}{P[B]}$,each for $P[AB]$, one gets $P[B\,|A]P[A] = P[AB]$ and $P[A\,|B]P[B] = P[AB]$. Since they are equal. $P[A\,|B]P[B] = P[B\,|A]P[A]$. If B is diabetes and A is skinny, this is $P[skinny\,|diabetes]P[diabetes] = P[diabetes\,|skinny]P[skinny]$. Said another say, $P[AB] = P[BA]$

   second part: In general, $P[A\,|B]$ does not equal $P[A]$. For example, the the probability that one has HIV is not equal to the probaility that one has HIV if one is an intraveneous drug user. But they can be the equal. Since $P[A\,|B] = \frac{P[AB]}{P[B]}$ they will be equal only if $\frac{P[AB]}{P[B]} = P[A]$. This requires that $P[AB] = P[A]P[B]$, in words, they are independent, which is usually not the case.

4. Consider two events: $C \equiv \{$the individual is a clown$\}$ and $E \equiv \{$the individual is an economist$\}$. Convince me that the probability that one is

either a clown or an economist is probably not the probablity of being a clown plus the probability of being an economist.

answer: You could convince the reader of Theorem 19 on page 24 of the third edition of MGB, which implies that $P[C \cup E] = P[C] + P[E] - P[CE]$. So, the above statement is only correct if there are no individuals who are both clowns and economists. My existence, since I am both, proves otherwise.

A formal proof: $C \cup E = A \cup \overline{C}E$ - everything that is in either C or E is everthing that is in C plus everything that is in E that was not in C. In addition, $C \cap \overline{C}E = \phi$ - these two sets have no intersection. Since these two sets are mutually exclusive, by Theorem 16 in MGB, $P[C \cup E] = P[C] + P[\overline{C}E]$. So, to complete the proof, we need to show that $P[\overline{C}E] + P[E] - P[CE]$, which is Theorem 16 in MGB.

Or, do a proof by contradiction. E.g. give the reader an example where $P[C \cup E]$ does not equal $P[C] + P[E]$. For example, draw a Venn diagram where economists are a subset of clowns and explain, in words, or whatever, why that demonstrates what you need to demonstrate.

5. Make up a data generating process with a random component that would determine whether an individual hits their spouse on Sunday night ($h_{it} = 1$ if individual i hits spouse on Sunday t, and zero otherwise). Police reports show that whether a spouse gets hit is a function of whether the Denver Broncos (the local football team - American footbal) won or lost, so make the outcome a function of that ($B_t = 1$ if lost on Sunday t, and zero otherwise).

   a possible answer: Imagine two urns: the Broncos' Won urn, and the Broncos' lost urn. Each urn contains 100 balls. In the first urn, 10 of the balls say "hit spouse". In the second urn, 25 of the balls say "hit spouse". When the football game is over, in each household each spouse draws a ball - from the Win urn if the Broncos won and from the Lost urn if they lost. In each household, there are $0, 1$, or 2 spouse hits.

   the following is an abbrev. of Maximillan's answer: Assume $h_{it} = B_t \varepsilon_t$ where ε_t can only take the values zero or one. Assume the probability of $\varepsilon = 1$ is p and the probability of $\varepsilon = 0$ is $(1 - p)$; that is ε is a draw from a Bernoulli distribution. Think of ε as a draw from an urn where p proportion of the balls say "hit" on them. So, if the Broncos win, $B_t = 0$, so $(0)(1 \text{ or } 0) = h_{it} = 0$, but if Broncos lose $B_t = 1$ and the probability of a spouse being hit (drawing a hit ball) is p.

   my second answer: I am going to start by specifying a classical OLS Model:

   $$D_{it} = \beta_t B_t + \beta_g G_i + \beta_r R_i + \beta_b Bud_{it} + \varepsilon_{it}$$

   where each ε_{it} is an independent draw from a normal distribution with mean zero and variance σ^2, $G_i = 1$ if the individual is a male, and zero otherwise, Bud_{it} is the number of beer individual i drank on Sunday t, and $R_i = \frac{w_i}{w_{is}}$ where w_i is the weight of individual i and w_{is} is the weight of their spouse. I am going to define D_{it} as individual $i's$ desire to hit their spouse on Sunday t. Not that D_{it} is normally distributed with mean, $\beta_t B_t + \beta_g G_i + \beta_r R_i + \beta_b Bud_{it}$ and variance σ^2. (Note that it would be innapropriate to assume $h_{it} = \beta_t B_t + \beta_g G_i + \beta_r R_i + \beta_b Bud_{it} + \varepsilon_{it}$ because that would say that h_{it} can take any value between plus and minus infinity - it can only take two values)

We don't observe D_{it}. So, let's specify a rule for converting desire into behavior. Assume $h_i = 1$ if $D_{it} > \alpha$, and zero otherwise. That's my model - it is a threshold model. The unknown parameters in this model are β_t, β_g, β_r, β_b, σ^2 and α. If we were econometricans, we would specify some estimators for these six parameters, collect some data, and do some estimating.

Do you have expectations as to the signs on the parameters? Do you think you could figure out the probability that $h_{it} = 1$. Let $m_{it} \equiv \beta_t B_t + \beta_g G_i + \beta_r R_i + \beta_b Bud_{it}$. So,

$$\begin{aligned}\Pr[h_{it} &= 1] = \Pr[m_{it} + \varepsilon > \alpha] \\ &= \Pr[\varepsilon > \alpha - m_{it}] \\ &= \Pr[\varepsilon < m_{it} - \alpha]\end{aligned}$$

so

$$\Pr[h_{it} = 1] = \int_{-\infty}^{m_{it}-\alpha} \frac{1}{\sigma\sqrt{2\pi}} e^{-(\varepsilon - m_{it})^2/2\sigma^2} d\varepsilon$$

if I did that right.

additional thoughts: It never crossed my mind to make B a random variable. But one could? That said, the problem is more difficult to answer if you assume two random variables: B and ε. One could have no epsilon and assume B is the single random variable and have a simple model: Broncos win no one gets hit and if they lose every spouse gets hit.

a number of students sort of had answers along the following lines: Assume that if the Broncos win, no one gets hit and if they lose everyone gets hit, but whether they win or lose is a random variable. Let p be the probability that they win. OK so far. Then assume $\ln(p_t/(1-p_t)) = \alpha + \beta B_t + \varepsilon_t$ where ε_t is a draw from a normal distribution (a logistic equation??). In this case $\ln(p_t/(1-p_t))$ is normally distributed with mean $\alpha + \beta B_t$. $\ln(p_t/(1-p_t)) = \alpha + \beta B_t + \varepsilon_t$ cannot be correct because on has made the probability of the Broncos winning a function of whether they won or lost. One could make p the probability of hitting one's spouse - I would be more comfortable with that. In which case, hitting is a rv but not whether the Broncos won or lost. Remember it is Sunday night and the game is over. But I am still confused a bit. What is the purpose of the ε_t? What if one assumed $\ln(p_t/(1-p_t)) = \alpha + \beta B_t$ where p is the probability

7818 Quiz 1 - Take-Home Part

Produced by James Meldrum
for Prof Edward Morey, ECON 7818

Problem Statement:

Make up a data generation process that would explain whether an individual hits their spouse on Sunday night ($h_{it} = 1$ if individual i hits spouse on Sunday t, and zero otherwise). Police reports show that whether a spouse gets hit is a function of whether the Denver Broncos (the local football team - American footbal) won or lost, so make the outcome a function of that ($B_t = 1$ if lost on Sunday t, and zero otherwise).

My Answer:

In this problem, the initial random variable determines whether the Denver Broncos won or lost on any given Sunday, denoted t. I use the built-in *Mathematica* function "RandomInteger" to generate the outcome of Broncos game. "RandomInteger" accesses a "pseudorandom" number generator, which performs a set algorithm upon a "seed" number derived from the time of day, and outputs an integer that is close enough to random for most practical purposes. Leaving the argument empty (i.e. following the command with empty brackets "[]") tells *Mathematica* to output either a 0 or a 1, with probability 1/2 in either case.

```
Bt = RandomInteger[];
```

Now I am ready to declare the function of whether a spouse gets hit on Sunday t. This function will have the independent variable of B_t (the outcome of the Broncos game), and the dependent variable will be h_{it} (the random variable assigned a 1 if individual i hits spouse on Sunday t and a zero otherwise. I will assume that the individual is a Denver Broncos fan and thus hits her spouse if the Broncos lost on Sunday, i.e. if $B_t = 1$. Because the independent variable can only take on the values of 0 and 1, the function describing whether or not the individual hits her spouse could logically be a very simple function, which passes the value of B_t directly to h_{it}, i.e. $h_{it}(B_t) = B_t$. In this instance, I have not surpressed the output, so that the function will automatically display its result.

```
hit = Bt
```

```
0
```

However, let me now assume that the individual is a bit more complicated than that. Her actions are not strictly deterministic based on the outcome of the game, but rather are influenced by a large variety of factors (amount of alcohol consumed during the game, success of her fantasy football players, the weather, etc.). Let me assume that the influence that these many contributing factors exerts upon her decision to hit her spouse can be represented by a randomly generated uniformly distributed error term, ϵ. I generate this error term with the *Mathematica* function "RandomReal," which outputs a pseudorandom real number between 0 and 1 if I leave the argument empty. The pseudorandom number generator works analogously to that which produces RandomInteger, only the output is now as a real number uniformly distributed across a range, rather than as integer values.

of hitting one's spouse? One would have a random data generating process where $p = \frac{e^{\alpha+B\beta}}{e^{\alpha+B\beta}+1}$. This can be viewed as an urn problem: Where $\frac{e^{\alpha+\beta}}{e^{\alpha+\beta}+1}$ is the proportion of balls in the urn if the Broncos lost and $\frac{e^{\alpha}}{e^{\alpha}+1}$ is the proportion if they won$\frac{e^{\alpha}}{e^{\alpha}+1}$

another possible answer: Whether the spouse gets hit is determined by a random draw from the unit uniform distribution. If the draw is $\leq m$ there is a hit, where $m = .05$ if the Broncos win and .10 if they lose.

Jame's mathematica answer:

6. What is the binomial theorem?

7. Prove that $P(A|B)P(B) = P(B|A)P(A)$

8. What is $\binom{n}{k}$ in both words and functional notation? Why do we care?

9. Events W and Z are independent if ?

10. Consider the experiment of drawing four cards from an ordinary deck. How many possible outcomes are there; that is, how many combinations are there of four cards?

11. So, one year there were 28 students in 7818, give or take. Assuming order in line matters, how many different ways are there to line up the class?

Or said another way, if we sampled the whole class without replacement and the first person sampled was put in position one, the second person in person two, etc., how many different samples are there (how many elements in the sample ? What if order did not matter.

What if instead of taking sample of 28 without replacement we took sample of three without replacement. How many possible samples are there if order matters? If order doesn't matter.

What are the general formulas involved in the above calculations. Use m to denote the population size and k the sample size.

answer: The answer to the first question is $28! = 304\,888\,344\,611\,713\,860\,501\,504\,000\,000$, a big number. In explanation, there are 28 possibilities for first position, 27 for second position, etc. If order doesn't matter the answer is one.

If the samples are of size three. If order matters there are $(28)(27)(26) = 19,656$. 28 different people can be in position 1, since there is not replacement there are only 27 candidates for position 2, and 26 for position 3. But many of these $19,656$ samples will consist of the same three people. How many different ways are there to order the same three people (Natasha, Bullwinkle and Rocky). Six, I think, $3! = 6$ (R,B,N; RNB, B,R,N; B,N,R; N,R,B; N,B,R). So, if order does not matter, it seems the answer is $\frac{(28)(27)(26)}{3!} = 3,276$

What are the general formulas involved in the above calculations. Let m equal the number in the population and let k equal the sample size. Consider the formula

$$(m)_k = m(m-1)(m-2)(m-k+1) \equiv \frac{m!}{(m-k)!}$$

This is the formula we used for the sample size of 28 and the sample size of 3 when order matters. To convert this to the case where order does not matter, we divide by $k!$ to get

$$\frac{(m)_k}{k!} = \frac{m(m-1)(m-2)(m-k+1)}{k!} \equiv \frac{m!}{(m-k)!k!}$$

This expression is typically abbreviated $\binom{m}{k}$ and called the *binomial coefficient* or the *combinatorial symbol*. And, is read as "the combination of m things taking k at a time" (MGB page 529). Note that the "taking" is without replacement and order does not matter. If order matter the answer is $(m)_k$.

What if the sampling was with replacement? In that case one could take a sample of any size, including sizes greater than 28. Consider a sample of 28 people with replacement: 28 possibilities on every draw, so the answer seems to be $28^{28} = 33\,145\,523\,113\,253\,374\,862\,572\,728\,253\,364\,605\,812\,736$ if order matters. If order does not matter one has to divide by $28!$, $\frac{28^{28}}{28!} =$

$\frac{411\,417\,561\,653\,470\,839\,904\,139\,214\,848}{3784\,415\,134\,680\,984\,375}$: $1.087\,1 \times 10^{11}$ (without replacement the answer was 1). Whether one samples with or without replacement makes a big difference.

For samples of size 3 with replacement, if order matters, the answer is $28^3 = 21,952$ and if order matters, $\frac{28^3}{3!} = \frac{10\,976}{3} = 3658.7$ (without replacement the answer was $3,276$).

12. The experiment is that a coin is flipped twice. How many outcomes are in the sample space and what are they? Now define and enumerate the event space. If the coin is fair, what is the probability associated with each of your events. Is there a formula you can use to figure out the size of the event space? If yes explain the formula.

    answer: There are $(2)(2) = 2^2 = 4$ possible outcomes to this experiment: (HH), (TT), (HT), (TH). There are $2^4 = 16$ possible events. I will try and list them

    the first four are elementary events (the four possible outcomes):

    HH ($prob = \frac{1}{4}$)

    TT ($prob = \frac{1}{4}$)

    HT ($prob = \frac{1}{4}$)

    TH ($prob = \frac{1}{4}$)

    The rest are nothing happens ($prob = 0$)

    HH or TT or HT or TH (something happens) ($prob = 1$)

    HT or TH (one of each) ($prob = \frac{1}{2}$)

    HH or HT (H on the first flip) ($prob = \frac{1}{2}$)

    TT or TH (T on the first flip) ($prob = \frac{1}{2}$)

    HH or TH (H on the second flip) ($prob = \frac{1}{2}$)

    TT or HT (T on the second flip) ($prob = \frac{1}{2}$)

    HH or TT (the outcomes of the first and second flip are the same) ($prob = \frac{1}{2}$)

    HH or TT or HT (how do you describe this one) ($prob = \frac{3}{4}$)

    HH or TT or TH (how do you describe this one) ($prob = \frac{3}{4}$)

    HH or TH or HT (at least one H) ($prob = \frac{3}{4}$)

    TH or TT or HT (at least one T) ($prob = \frac{3}{4}$)

    I hope I got that right. The event space is all sixteen of these events.

    Note that each event is a subset of the sample space. Note that each of these events are something you might want to bet on before the coin is flipped for the first time. Need to add an explanation of the formula 2^w where w is the number of elements in the sample space (see the question below).

Stuff to note about some student answers: It is correct to say that "an event is a subset of the sample space". It is incorrect to say that "event space is a subset of the sample space." The opposite is the case: sample space is a subset of event space.

If the experiment is to draw n observations (e.g. if the experiment is to flip two coins, $n = 2$). Everyone is then described in terms of n things: the outcome on the first draw, the second draw, ... the n^{th} draw. So, for example if two coins are flipped in sequence, an outcome is described in terms of two things: what happened on the first flip and wha happened on the second flip. For example, if the experiment is a coin is flipped once there are 2 outcomes, if the experiment is two flips there are 4 outcomes, and if the experiment is 100 flips there are $2^{100} = 1267\,650\,600\,228\,229\,401\,496\,703\,205\,376$ outcomes. Think of it another way, if one takes a sample with 20 observations, and realization of the rv can take only two values, there are 2^{20} possible samples.

13. If a sample space has w elements, how many subsets are there of these w elements. Explain your answer that would make a wary reader believe you. Give me a reason for wanting to know this.

    answer: If there are w elements in the sample space, there are subsets with w elements, $(w-1)$ elements, $(w-2)$ elements, ..., 1 element. So we need to figure out how many elements there are in each size of subset, and then add up these numbers. Note that order does not matter for subsets: two sets are, by definition, the same sets if they have the same elements.

    How many subsets with w elements? One

    How many subsets with 1 element? w

    Now things get a bit more complicated. How many subset with k elements? The answer is $\binom{w}{k} \equiv \frac{w!}{(w-k)!k!}$, the combinatorial symbol - this is explained in another question.

    But note that $\binom{w}{w} \equiv \frac{w!}{(w-w)!w!} = 1$ because 0! is defined as 1, and $\binom{w}{1} = w$.[1] This confirms our two answers above.

    Further note that $\binom{w}{0} \equiv \frac{w!}{(w-0)!0!} = 1$ so there is is one subset that has no elements, the empty set.

    So, to finish we just need to add the number of subsets of each size, which is $\sum_{k=0}^{w} \binom{w}{k}$. It is possible to show that this equals 2^w. Wow.

14. An urn contains six red and four black balls. Two balls are drawn without replacement. What is the probability that the second ball is red if it is known that the first is red?

[1] Note that $\binom{w}{1} = \binom{w}{w-1}$.

15. Assume an urn contains $M = 3$ balls numbered $1, 2$ and 3. Further assume that balls 1 and 2 are purple. Ball 3 is white. You randomly draw, without replacement, two balls from the urn. What is the probability that you draw only one purple ball? Explain, in word, how you determined your answer.

    answer: The sample space is $\Omega = \{(1,2),(1,3),(2,1),(2,3),(3,1),(,3,2)\}$: there are 6 possible outcomes to the experiment. Four of these include one, and only one, purple ball: $(1,3),(2,3),(3,1),(3,2)$. So the probability of drawing one purple ball is $4/6 = 2/3$. The formula is $\frac{outcomes\ that\ contain\ one\ purple}{total\ number\ of\ outcomes}$. The denominator is $(3)_2 = (3)(2) = 6$. Or more generally $(M)_n$ where M is the number of balls in the urn and n is the sample size.

    Another, more "difficult" way to answer the quiz question. There are two ways to get one purple ball: draw a purple first and a white second (P, W), or draw a white first and a purple second (W, P). The probability of drawing a purple first is $\frac{2}{3}$ and the probability of drawing a white first is $\frac{1}{3}$. If one draws a purple first the probability of drawing a white second is $\frac{1}{2}$ and, if one draws a white first, the probability of drawing a purple second is 1. So the probability of drawing $(P, W) = (\frac{2}{3})(\frac{1}{2}) = \frac{1}{3}$, and the probability of drawing $(W, P) = (\frac{1}{3})(1) = \frac{1}{3}$. So the probability of getting one purple and one white is the sum 2/3.

    How to get the right answer for the wrong reason (incorrectly assuming all events from an experiment are equally likely). Correctly identify three events: (two purple), (purple, white) and (white purple). Then assume, but don't demonstrate, that all three of these events events are equally likely. In this case, all three of these events are equally likely but that wil typically not be true. What is always true is that if the draws are random, each outcome is equally likely. It turns out that in this example each event can be produced by the same number of outcomes, but that won't generally be the case.

    Be careful with notation. Note that Pr(purple then white) does not equal $\Pr[P \cap W]$; in the first order matter, in the second it does not matter.

    So, another way to approach this problem (this is equivelant to how I approached the problem). Conclude that events with one and only one purple ball require that one of the balls is white. So determine the probability that one of the balls is white. $\Pr[W] = \frac{2}{3}$ because four of the six equally likely outcomes has a white ball. Some of you wanted to look at $\Pr[P\,|W] = \frac{\Pr[PW]}{\Pr[W]} = \frac{\frac{2}{3}}{\frac{2}{3}} = 1$ instead of $\Pr[W]$. I am not sure why this conditional probability would be of interest - it just says that if one of the draws is a white ball the other one has to be a purple ball.

    So, another way to get confused. If you forget that sampling is without replacement, you might start thinking about the Binomial distribution. With the binomial distribution the probabilities don't change with the draws, here the probability of drawing a color depends on what color was drawn previously.

More generally (and not required for your answer) We determined the number of outcomes with only one purple ball by counting but is there a general formula? MGB (page 29 thrid edition) say yes. They say it is $\binom{n}{k}(K)_k(M-K)_{n-k}$ where K is the number of purple balls in the urn, n is the sample size, and k is the number of purple balls in the sample . In our case, $M=3$, $K=2$, $n=2$ and $k=1$. So let's see if the formula works for our simple case. $\binom{2}{1}=2$ equals the number of subsets of size 1 from a sample with two elements. $(K)_k = K(K-1)K(K-2)...(K-k+1)$; so, $(2)_1 = 2(1) = 2$. $(M-K)_{n-k} = (M-K)((M-K)-1)((M-K)-2)...((M-K)-(n-k)+1)$; so $(1)_1 = 1$. So in this case $\binom{n}{k}(K)_k(M-K)_{n-k} = \binom{2}{1}(2)_1(1)_1 = (2)(2)(1) = 4$, which is the right answer - wow. I don't really understand their explanation of why this formula makes sense. Do you?

But, let's assume it is correct. In which case the probability of getting k purple balls if one randomly draws, without replacement, a sample of n balls from an urn with M balls, K of which are purple, is

$$\Pr[k] = \frac{\binom{n}{k}(K)_k(M-K)_{n-k}}{(M)_n}$$

which some fool has shown is equal to

$$= \frac{\binom{K}{k}\binom{M-K}{n-k}}{\binom{M}{n}}$$

So, let's try it out. Assume $M=10$, $n=5$, $K=3$, and $k=2$. Plugging in $\frac{\binom{K}{k}\binom{M-K}{n-k}}{\binom{M}{n}} = \frac{\binom{3}{2}\binom{7}{3}}{\binom{10}{5}} = \frac{5}{12}$, saying that the probability of drawing a sample with two purple balls is $\frac{5}{12}$. MGB say you should know this formula if you play poker for money. I don't (I hate games), so I don't.

16. Assume that event space has the following properties: (1) $\Omega \in \mathcal{A}$. That is, the event that one of the outcomes occurs is an event. Note that $\Pr[\Omega]=1$. (2) if $A \in \mathcal{A}$ then $\bar{A} \in \mathcal{A}$ where $\bar{A}$ is the compliment of A. That is, if A is an event, then not A is an event. And (3) if A_1 and $A_2 \in A$ then $A_1 \cup A_2 \in \mathcal{A}$. That is, either event happening is an event

    These three axioms/assumptions supposedly imply the following:

    $\emptyset \in \mathcal{A}$ This follows from the first two Axioms. Why

    if A_1 and $A_2 \in \mathcal{A}$, then $A_1 \cap A_2 \in \mathcal{A}$

    if $A_1, A_2, ...A_n \in \mathcal{A}$ then $\bigcup_{i=1}^{n} A_i$ and $\bigcap_{i=1}^{n} A_i \in \mathcal{A}$

    Can you convince your fellow students that these three theorems follow logically from the three axioms?

17. What is random variable? and what role do they play in statistics?

    answer: Before we define a rv we should define a *variable*. A simple answer, too simplistic, would be that a variable is something that varys. But what is that something? A variable is something that can take at least two states and each state can be represented by a different number. For example, gender is a variable because it can take two states (male or female) and each state can be represented with a number (for example, 1 for females and 0 for males, or any other two numbers). Height is a variable because it can take an infinite number of states (heights) and each state is represented by a unique number.

    A rv can be defined in two ways. I find both definitions add insights. (1) A rv is a variable that has some distribution. That is, it has some density function or probability density function. (2) A rv is a function that associates a real number with each event in event space. Because it associates different numbers with events, it varies.

    Simply put, RVs characterize events in terms of the value of a real number. They are the link between probability theory and distribution theory. Since events are uncertain, the specific value a rv will take is also uncertain, but it will have some distribution, which can be described by a density function or probability density function. Associating numbers with events makes it much easier to study events and their probabilities.

    Historically, student answers have been weak on what is a *variable*. A complete answer needs to define both *variable* and then explain what it means for a variable to be a random variable. Imagine I asked you define *white elephant* and all you did was explained what the adjective *white* implies - never defining *elephant*. Keep in mind that the we use numbers to distinguish between the different states that a variable can realize, but it is important to distinguish between the different states and the numbers used to represent the different states.

    Zach asked how a something can be a variable, but not a rv. Good question. A variable that is not a rv is something that can realize two or more states, but one cannot determine, in theory, the probabilities associated with different states. Said another way, a rv is a variable whose varying cannot be described with a density function. Examples become tricky. Start simply by assuming that x is not a rv, in which case $y = 5x$ is not a rv, but $y = 5x + \varepsilon$, where ε is a random draw from some distribution is a rv, because it is a function of a rv. The issue is whether the process that determines which state the variable of interest will realize has a random component. If not, it is not a random variable. Often variables are simply assumed not to be random variables, even thought an argument could be made that they are a function of random variables. Can God be deterministic but not always do the same thing, does his behavior follow deterministic rules? Might there be some variables that are deterministic but random variables from our perspective?

18. I am always confused about when two events are, or are not, independent. Make up a two page or less handout, with examples and intuition, that will help me to understand these concepts. Include in your handout a definition of independence. Why should I care whether statistical events are or are not independent. Thanks for the help.

19. How would you determine whether being a male and being homosexual are, or are not, independent? Are maleness (being a male) and femaleness independent or dependent events?

20. Consider two events, A and B. In general, what is $P(A \cup B)$ and how does it relate to $P(A)$ and $P(B)$? Why? What if A and B are independent? Mutually exclusive. As part of your answer define all of your terms.

    answer: $P(A \cup B)$ is the the probability of either events A, B, or both occurring. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, where $P(A \cap B)$ is the probability that events A and B simultaneously occur. If one did not subtract $P(A \cap B)$ from $P(A) + P(B)$, one would be double counting in terms of the likelihood of A or B.

    If A and B are mutually exclusive, $A \cap B = \emptyset \Longrightarrow P(A \cap B) = 0$. In which case, $P(A \cup B) = P(A) + P(B)$.

    Mutually exclusive and independent are, in general, not compatible. If A and B are independent, $P(A \cup B) = P(A) + P(B) - P(A)P(B)$ because independence implies $P(A \cap B) = P(A)P(B)$.

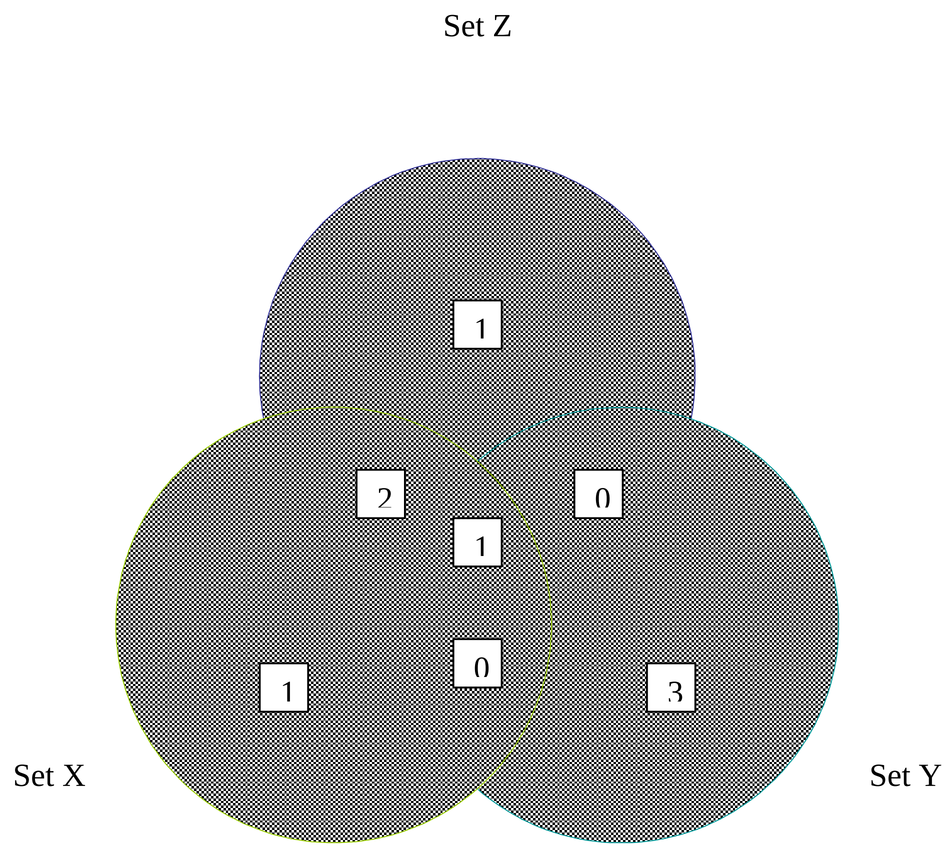
    Note that if $P(A) > 0$, $P(B) > 0$ and A and B are independent, A and B cannot be mutually exclusive.

    If A and B are independent, $P(A \cup B) = P(A) + P(B) - P(A)P(B)$. Which simplifies to $P(A \cup B) = P(A) + P(B)$ iff $P(A) = 0$, or $P(B) = 0$.

21. Consider three overlapping sets: X, Y and Z. Draw a Venn diagram representing these three sets and put a number in each segment of each set (where the number in each segment represents the number of elements in that segment. Draw your figure and choose numbers such that $\Pr(X \cap Y \cap Z) = \Pr(X)\Pr(Y)\Pr(Z)$, but $\Pr(X \cap Y) \neq \Pr(X)\Pr(Y)$ and $\Pr(X \cap Z) \neq \Pr(X)\Pr(Z)$. Explain to your reader why your figure meets the stated requirements.

    answer: there are some other questions like this one at the end of chapter 2 is Amemiya. Consider the following example provided by a number of students:

    This example was chosen because it is simple: there are 8 possible outcomes. In the Venn diagram above, the numbers represent the number of outcomes in each subset, for example, $(Z \cap X)/Y$ has 2 elements. Let's see if this example has the specified properties. There are 4 elements in X, in Y, and in Z, so $\Pr[X] = \Pr[Y] = \Pr[Z] = \frac{4}{8} = \frac{1}{2}$. $(X \cap Y \cap Z)$ has one element, so $\Pr[X \cap Y \cap Z] = \frac{1}{8}$. The intersection


Set Z
1
2
0
1
0
1
3
Set X
Set Y

$(X \cap Z)$ has 3 elements, so $\Pr[X \cap Z] = \frac{3}{8}$. And, $(X \cap Y)$ has 1 element, so $\Pr[X \cap Y] = \frac{1}{8}$. $\Pr[X]\Pr[Y] = (\frac{1}{2})(\frac{1}{2}) = \frac{1}{4} = \Pr[X]\Pr[Z]$, and $\Pr[X]\Pr[Y]\Pr[Z] = (\frac{1}{2})(\frac{1}{2})(\frac{1}{2}) = \frac{1}{8}$

So, the three conditions are met: $\Pr[X \cap Y \cap Z] = \frac{1}{8} = \Pr[X]\Pr[Y]\Pr[Z]$, $\Pr[X \cap Y] = \frac{1}{8} \neq \frac{1}{4} = \Pr[X]\Pr[Y]$, and $\Pr[X \cap Z] = \frac{3}{8} \neq \frac{1}{4} = \Pr[X]\Pr[Z]$. This is all that is needed to answer the question as asked.

That said, two additional questions come to mind. Can one work out the necessary conditions to fulfill these conditions? And, what does the example express in terms of the concept of independence. In terms of independence, X and Y are not independent, X and Z are not independent. Are X, Y and Z independent? No they are not even though $\Pr[X \cap Y \cap Z] = \Pr[X]\Pr[Y]\Pr[Z]$, which is a necessary but not a sufficient condition for the three sets to be independent. Independence of the three sets requires $\Pr[X \cap Y \cap Z] = \Pr[X]\Pr[Y]\Pr[Z]$, $\Pr[X \cap Y] = \Pr[X]\Pr[Y]$, $\Pr[X \cap Z] = \Pr[X]\Pr[Z]$ and $\Pr[Y \cap Z] = \Pr[Y]\Pr[Z]$.

Here is another example that meets the three properties. It was developed by Yiqing in class. Note that in this example, $\Pr[X] = \Pr[Y] \neq \Pr[Z]$

22. Consider three overlapping sets: X, Y and Z. Draw a Venn diagram representing these three sets and put a lower-case letter in each segment of each set (where the letter in each segment represents the number of elements in that segment. Assume that $\Pr[X \cap Y \cap Z] = \Pr[X]\Pr[Y]\Pr[Z]$ and that $\Pr[Y \cap Z] = \Pr[Y]\Pr[Z]$. Derive some conditions on $a, b, c, ..., g$ for these two conditions to hold. Let $m = a + b + c + ... + g$. Begin by figuring out some general conditions. $\Pr[X] = \frac{b+c+f+g}{m}$, $\Pr[Y] = \frac{d+e+f+g}{m}$, $\Pr[Z] = \frac{a+b+c+d}{m}$, so

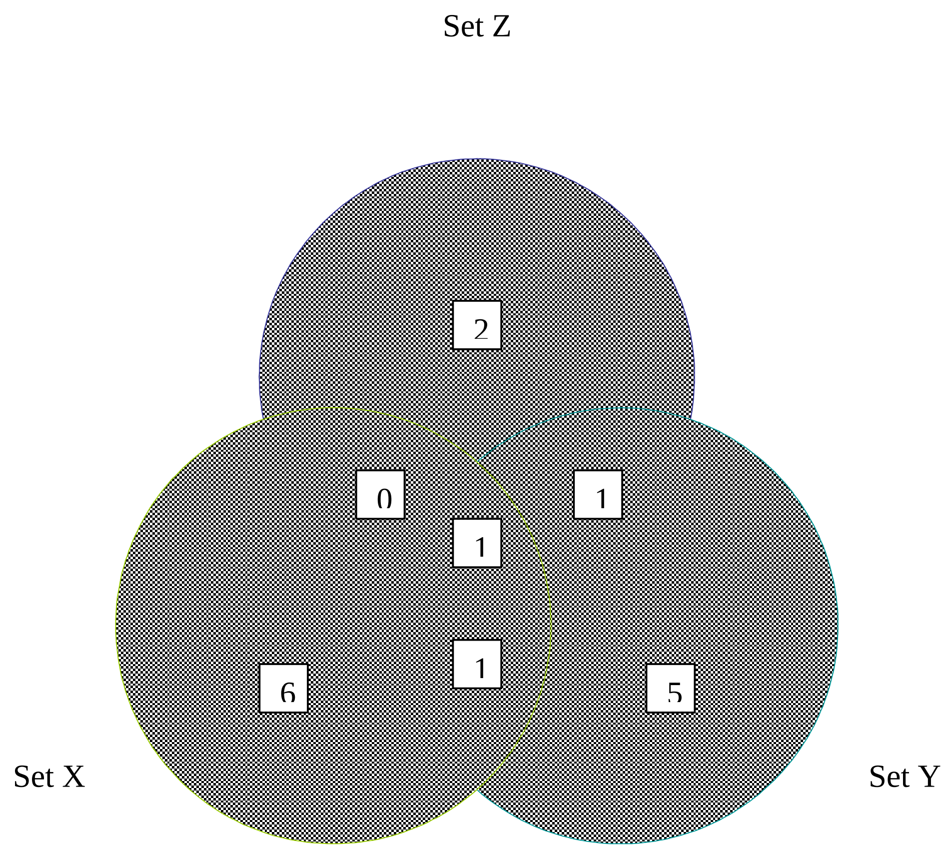
$$\begin{aligned} \Pr[X]\Pr[Y]\Pr[Z] &= (\frac{b+c+f+g}{m})(\frac{d+e+f+g}{m})(\frac{a+b+c+d}{m}) \\ &= \frac{1}{(a+b+c+d+e+f+g)^3}(d+f+g+e)\big(ab+ac+2bc+bd+af+cd+ag+ \end{aligned}$$

Wow!

In addition $\Pr[Y \cap Z] = \frac{c+d}{m}$. So, the second condition requires that $\Pr[Y \cap Z] = \Pr[Y]\Pr[Z]$ requires that $\frac{c+d}{m} = (\frac{d+e+f+g}{m})(\frac{a+b+c+d}{m})$. This implies that $c + d = (d + e + f + g)(a + b + c + d)$.

In addition $\Pr[X \cap Y \cap Z] = \frac{c}{m}$, so the first condition requires that

$$\begin{aligned} &\frac{c}{(a+b+c+d+e+f+g)} \\ = &\frac{1}{(a+b+c+d+e+f+g)^3}(d+f+g+e)\big(ab+ac+2bc+bd+af+cd+ag+bf+bg+cf+c \end{aligned}$$


Set Z
2
0
1
1
1
6
5
Set X
Set Y

which implies

$$c = \frac{1}{(a+b+c+d+e+f+g)^2}(d+f+g+e)\left(ab+ac+2bc+bd+af+cd+ag+bf+bg+cf+cg+\right.$$

So, it would seem, if I did this correctly, that the two condition require that

$c = \frac{1}{(a+b+c+d+e+f+g)^2}(d+f+g+e)\left(ab+ac+2bc+bd+af+cd+ag+bf+bg+cf+cg+df+dg+\right.$

and $c + d = (d + e + f + g)(a + b + c + d)$. Whatever.

answer:

23. (from MGB) If events M and N are mutually exclusive (disjoint), $\Pr(M) = .5$ and $\Pr[M \cup N] = .6$, what is $\Pr[N]$?

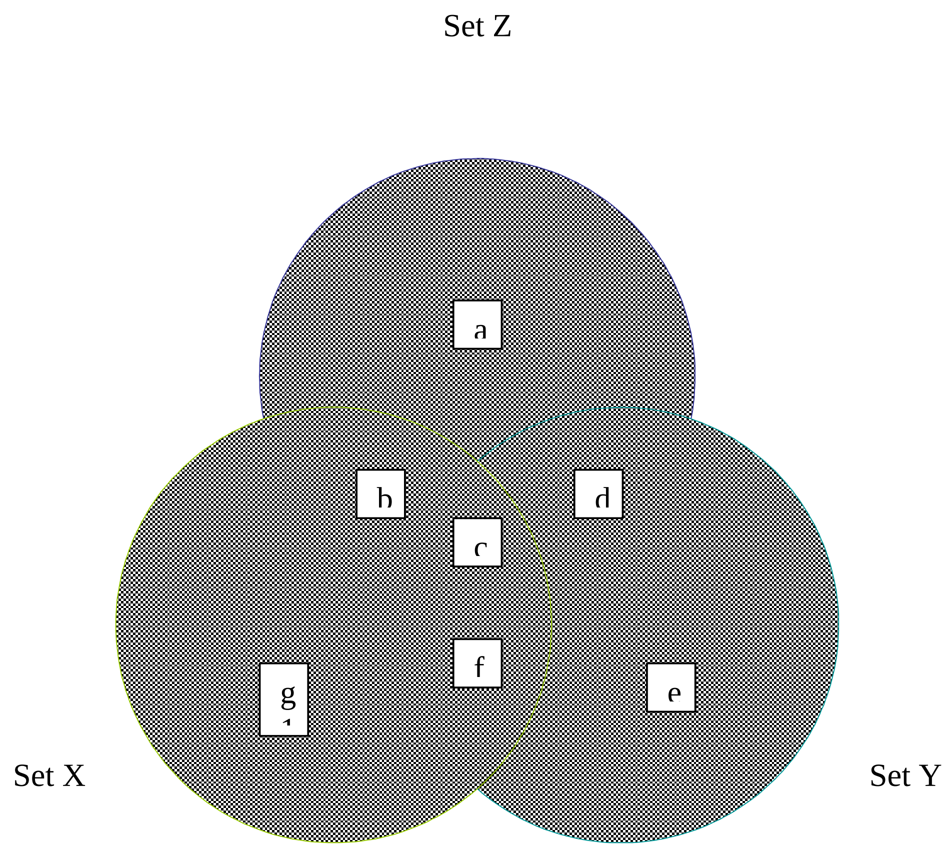
24. Prove that $(B - C) \cap (S - C) = (S \cap B) - C$.

25. (from MGB) In an assembly-line operation, 1/3 of the items produced are defective. Consider three items I picked at random and tested. What is the probability that exactly one of them will be defective, that at least one of them will be defective, that all of them will be defective.

26. In 7818 we study sample spaces, outcomes, events, probability and probability functions. Explain to the reader what is meant by the "probability of an event".

    answer: Simply put, the probability of an event is the number associated with that event by the probability function. A probability function is a function with certain properties that associates a number with every possible event. This is a sufficient explanation, but not very enlightening.

    Explaining further: The probability function imposes certain properties on the number it attaches to an event. The number must be between zero and one, inclusive. In addition, if one "summed" the numbers associated with all possible events, the sum would equal one. The number associated with no event occurring is zero, and the number associated with events A or B occurring cannot be less than the number associated with either A or B alone. What does a "probability" indicate. Considering its properties, it indicates the "likelihood" of an event, in that the "probability" of nothing happening is zero (zero indicates the event cannot happen), the "probability" of something happening is one (one indicates the event is certain), and the "probability" of events A and B happening can't be less than the "probability" of either alone. Sometimes it is appropriate to define the probability of an event in the following way. If an experiment were repeated an infinite number of time, the probability of the event would equal the proportion of times the event occurred. This would be a particular specification (a frequency specification) of the probability function.


Set Z
a
b
d
c
f
g
e
Set X
Set Y

One could then estimate the probability of an event by the proportion of times it occurs in n experiments.

The classical definition of the probability of an event and the frequency definition of the probability of an event imply two specific probability functions.

Note that when answering questions: A giraffe is a giraffe and a probability is a probability but each term does not define itself. That is, to say "a giraffe is a giraffe" or that a "giraffe exudes giraffeness" are both true statements, but they are not helpful when it comes to determining what is and is not a giraffe.

27. Who was Bayes? What is his theorem? Why do we care?

28. So tomorrow you might go skiing and you have J sites to choose from. Let $P[S_j]$ be the probability that you ski at area j. You can also not ski; denote the probability of not skiing as $P[NS] > 0$. Let $P[S]$ denote the probability of skiing somewhere. There are $J+1$ alternatives in your choice set. Starting with Bayes formula, convince the reader that the probability that you will ski site j conditional on your skiing must be greater than the simple probability that you will ski site j.

    answer: In this notation, Bayes formula says $\Pr[S_j\,|S\,] = \frac{\Pr[S|S_j\,]\Pr[S_j]}{\Pr[S]}$. But $\Pr[S\,|S_j\,] = 1$, so $\Pr[S_j\,|S\,] = \frac{\Pr[S_j]}{\Pr[S]} = \frac{\Pr[S_j]}{1-P[NS]}$. Since $1 > P[NS] > 0$, $1 > 1-P[NS] > 0$, so we are dividing by a positive fraction so $\Pr[S_j\,|S\,] = \frac{\Pr[S_j]}{1-P[NS]} > \Pr[S_j]$.

29. Imagine that 30% of the dogs in Boulder are vegetarians. Further assume that 60% of these vegetarians eat a raw diet (only raw food), but only 10% of the other dogs in Boulder eat a raw diet. What is the probability that a Boulder dog who eats a raw diet is not a vegetarian, $\Pr[NV\,|R]$. Explain how you solved this problem, including the tool(s) you used. If appropriate, use a Venn diagram to represent this situation.

    answer: One can use Bayes formula to answer this question. Bayes formula, in this context, is

    $$\Pr[NV\,|R] = \frac{\Pr[R\,|NV\,]\Pr[NV]}{\Pr[R]}$$

    where NV denotes non-vegetarian and R is raw diet. From the information provided, 70% of Boulder dog are not vegetarians ($\Pr[NV] = .7$) and only 10% of the non-vegetarians eat a raw diet ($\Pr[R\,|NV\,] = .1$). So,

    $$\Pr[NV\,|R] = \frac{(.1)(.7)}{\Pr[R]}$$

    So, what is $\Pr[R]$, the probability that a Boulder dog eats a raw diet? It is, based on the *theorem of total probabilities*, $\Pr[R] = \Pr[R\,|NV\,]\Pr[NV]+$

$\Pr[R\,|V\,]\Pr[V] = (.1)(.7) + (.6)(.3) = 0.25$. So

$$\Pr[NV\,|R] = \frac{(.1)(.7)}{.25} = .28$$

which is the probability that one eats meat (is a NV), given that one eats a raw diet. As an aside, it is not easy to observe whether a dog is a true vegetarian and/or only eats raw food - dogs eat behind our backs - my dog can grab and swallow items off the sidewalk in the flash of a second.

To be complete, one might derive Bayes theorem. Start with the definition of conditional probability, which in this context is $\Pr[NV\,|R] = \frac{\Pr[NV,R]}{\Pr[R]}$. Combing this with $\Pr[NV,R] = \Pr[R\,|NV\,]\Pr[NV]$, and rearranging one gets Bayes theorem.

30. Edna Gomer and Wilbur Guber, after years of research, isolated a disease that kills 1/1000 Americans every year. For their trouble, it has been named the *Guber-Gomer disease*. (Edna and Wilbur died in each other's arms after injecting each other with the disease - dying in loving bliss.) If one contracts the disease, one will be dead within the year. Recently a very accurate test has been developed to see if one has the disease – it is 99% accurate: everyone with the disease will test positive for the disease, but 1% of those who take the test will test positive even though they do not have disease (1/100 are false positives). You get the test and the results are positive. Should you worry? Should you assume you are a goner? What is the probability of you dying in the next 12 months from GG. Make sure to completely explain how your determined this probability. What did having the test do to your probability of having GG?

    answer: Before giving up hope, apply Bayes Theorem. The probability of dying from GG is $\Pr[GG] = .001$. The probability of a positive test result if one has GG, $\Pr[P\,|GG]$, is 1 where P is a positive test result. And, the probability of a positive test result, $\Pr[P]$, is $\frac{11}{1000} = 0.011$(11 people out of 1000 will test positive). So, applying Bayes theorem

$$\begin{aligned}\Pr[GG\,|P\,] &= \frac{\Pr[P\,|GG\,]\Pr[GG]}{\Pr[P]} \\ &= \frac{1(.001)}{.011} = .0909\end{aligned}$$

    about 9%, so you have less than a 10% chance of dying from GG this year, not a 99% chance of dying. Things are not as bad as many people in your position would conclude. That said, things could be better: on the basis of the test your chances of dying this year of GG increased 90-fold.

    Or you could have looked at

$$\begin{aligned}\Pr[notGG\,|P\,] &= \frac{\Pr[P\,|notGG\,]\Pr[notGG]}{\Pr[P]} \\ &= \frac{.01(.999)}{.011} = .90818\end{aligned}$$

which is the probability that you don't have GG given that you tested positive. The test kind of sucks.

31. Consider two events, A and B. In general, what is $P(A \cup B)$ and what would it be if A and B were independent?

32. MGB and I have made a big deal of the distinction between *outcomes* and *sample space*, on the one hand, and *events* and *event space* on the other. Explain to the reader how outcomes and sample space differ from events and event space. Why is it important to separately define sample space and event space - make sure the reader understand why it is important to know this stuff (motivate the material). Make sure to define all your terms. Examples are always good for making points but remember that examples are not definitions. You answer will obviously require some notation. In addition it will also require words and, maybe, some figures. I will try to evaluate your "essay" trying to maintain the fiction? that I was clueless when I started reading your essay.

    answer: *outcomes* are defined in the context of experiments (or samples). Every possible outcome of an experiment is called an outcome, sometimes it is called a *sample point*, in the sense that each possible sample is an outcome. The realization of one outcome precludes all the others - outcomes are mutually exclusive. The set of all possible outcomes is called the sample space, typically denoted with the letter Ω and an outcome is denoted ω.

    An event, A, is defined as a subset of the sample space including nothing happening and something happening, so all outcomes are events (simple/*elemental* events) but not all events are outcomes. Let $\mathcal{A}$ denote the set of all events, *event space*. If one groups or classifies outcomes one is defining events.

    I tend to think of events as things that one might gamble on. You win if the event occurs and lose if it does not.

    So, in terms of notation $\omega \in \Omega$, $A \in \mathcal{A}$, $\omega \in \mathcal{A}$, , $\varnothing \in \mathcal{A}$ and $\Omega \in \mathcal{A}$. Also note that typically $\Omega \subset \mathcal{A}$.

    It is correct to say that an event is a subset of the sample space.

    Note that outcomes are mutually exclusive, but many events are not mutually exclusive.

    Given an event is a subset of the sample space, it is tempting, but wrong, to conclude that event space is a subset of sample space. This is wrong; it is backwards. Sample space is a subset of event space: every outcome of an experiment is an event (a simple or elemental event) but most events are not outcomes. Many of you drew examples of sets that proved this, even though you asserted the opposite.

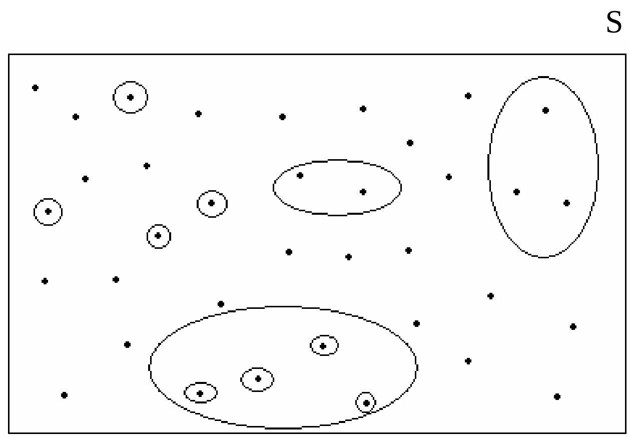
Given that sample space and outcomes are understood, why is it important to then introduce the concept of events and events space? Simply put, most of the things that we want to predict the probability of are not outcomes, that is, they are not members of the sample space, rather they are subsets of the sample space. We want to predict the probability that some specified subset of the sample space will be realized. For example, if our interest is in the probability of being dealt a full house (three of a kind plus two of a kind), we are interested in the probability of drawing an outcome from the subset of sample space that implies a full house. Being dealt a full house is an event, a complicated event - there are many different outcomes of the experiment (dealing 5 cards) that can generate this outcome. We need a name for a subset of sample space, a generic name for what we are calculating the probability of, *event* fits the bill. If we didn't call these things events, we would call them something else, *happenings*, for example. We then need a name for the set of all events.

To calculate the probability of an event, one typically needs to know the number of outcomes in the sample space, the probability associated with each (often these are equal to $\frac{1}{N}$ where N is the number of elements in the sample space) and knowledge of which outcomes imply the event. Note that the domain of the probability function, $\Pr[A]$ is event space, $\mathcal{A}$.

The following, from a student, is a graphical example of a finite sample space, indentifying all the outcomes, some of the events, and the sample space. S is the entire sample space, each dot is an outcome, every circle is an example of an event (note that an event is one or more outcomes), and that many events are not drawn.

In closing, I quote from a student: An event will always be a subset of the sample space, but for sufficiently large sample spaces, not all subsets will be events. This means that the class of all subsets of the sample space will not necessarily correspond to the event space. However, if the sample space consists of only a finite (not infinite) number of points, then the event space will be the class of all subsets of the sample space.

Primary interest is not in the events themselves, but in the likelihood that an event does or does not occur. While the sample space is fairly basic and generally easy to define for a given experiment, it is the event space that is essential in defining probability. When designing or observing experiments, it is important to understand the interplay of all these terms / designations. Certain properties and theorems apply to events, event space, sample space, and outcomes. Being able to understand the context of those theorems allows us to be more effective and efficient in communicating with other economists, econometricians or researchers in general. It provides rules to the communications, so that all



terms do not have to be defined every time they are used. By using common definitions, it eliminates confusion later and provides structure to the communication. If these terms were not able to be distinguished from each other, we would not know what we were attempting measure unless a clear definition was specified each time we wished to use them.

33. from MGB). If $\Pr[A] = 1/3$ and $\Pr[B] = 1/4$, can A and B be mutually exclusive. Yes or No and explain.

34. Imagine that every day I do one of two things: drink a gallon of Coke, or drink a gallon of Pepsi, my stomach can only handle one of these experiences each day. Let $\Pr_c$ denote the probability that it is a gallon of Coke, and $\Pr_P$ the probability that it is a Pepsi. Over a ten day period you observe me drinking two Pepsi, and eight Coke. What is the probability of observing this sample of my beverage consumption. Is the probability $(\Pr_C)^8(\Pr_P)^2 = (\Pr_C)^8(1 - \Pr_C)^2$? Yes or no, and explain why in as much detail as you can. Now generalize assuming c is the number of Coke I consume in a n day period.

    answer: The answer to the first part is No. $(\Pr_C)^8(1 - \Pr_C)^2$ is the probability of drinking 8 Cokes and 2 Pepsi in a particular sequence/order (the probability is the same for each possible ordering of 2 Pepsi and 8 Coke). So, to figure out the probability of drinking Coke 8 of the 10 days, we need to figure out how many different ways there are to drink 8 Coke and 2 Pepsi in 10 days. Or said another way, how many different ways are there to drink Coke 8 out of ten days. The answer is provided by the binomial coefficient $\binom{10}{8} = \frac{10!}{(10-8)!8!} = 45$; there are 45 different ways to get

8 Cokes in 10 trials (there are 45 subsets of the sample space that have the property 8 Cokes and 2 Pepsi). So, the correct answer is for the probability of observing me drinking Coke 8 times out of 10 is $\binom{10}{8}(\Pr_C)^8(1-\Pr_C)^2$. For example if $\Pr_C = .23$, $f(10,8) = \binom{10}{8}(.23)^8(.77)^2 = 2.0894 \times 10^{-4}$, not very likely. In contrast the probability of Coke on the first eight days and Pepsi on the last two is $(.23)^8(1-.23)^2 = 4.6431 \times 10^{-6}$ which is also the probability of Pepsi on the first two days and Coke on the last eight.

Now, generalizing the problem, $f(n,c) = \binom{n}{c}(\Pr_C)^c(1-\Pr_C)^{n-c}$ if $n = o, 1, 2, ..., n$, and zero otherwise. This is the binomial distribution.

Note that $f(n,p) = \binom{n}{p}(\Pr_P)^p(1-\Pr_P)^{n-p}$ which if $\Pr_C = .23$ and p is 2, is $\binom{10}{2}(.77)^2(.23)^8 = 2.0894 \times 10^{-4}$, which, as it must, equals the probability of drinking Coke 8 out of the 10 times.

35. Can you prove that, in general, $\binom{n}{c}(\Pr_C)^c(1-\Pr_C)^{n-c} = \binom{n}{p}(\Pr_P)^p(1-\Pr_P)^{n-p}$, where $\Pr_P = 1 - \Pr_C$ and $c + p = n$?

    answer: start by proving/showing that $\binom{n}{c} = \binom{n}{p}$ if $p = n - c \Longrightarrow \binom{n}{p} = \binom{n}{n-c} = \frac{n!}{(n-c)!(n-(n-c)!} = \frac{n!}{(n-c)!c!} = \binom{n}{c}$. So, all that is left is to show that $(\Pr_C)^c(1-\Pr_C)^{n-c} = (1-\Pr_P)^c(1-(1-\Pr_P))^{n-c} = (1-\Pr_P)^c(\Pr_P)^{n-c} = (1-\Pr_P)^{n-p}(\Pr_P)^p$

36. Imagine that every day I do one of three things: drink a gallon of Coke, drink a gallon of Pepsi, or eat a gallon of gelato. Let $\Pr_c$ denote the probability that it is a gallon of Coke, ... Over a ten day period you observe me drinking two Pepsi, five Cokes and eating three gallons of bacio gelato. What is the probability of observing this sample of my beverage and gelato consumption?

    answer: the probability of observing two Pepsi, five Coke and three gallons of gelato in a particular order is $(\Pr_C)^5(\Pr_P)^2(\Pr_G)^3 = (\Pr_C)^5(\Pr_P)^2(1-\Pr_C-\Pr_P)^3$ but there are many different possible sequences/orders. How many? The Binomial coefficient can help us here. $\binom{n}{c} = \frac{n!}{(n-c)!c!}$ applies when there are two possible outcomes, Coke and *not* Coke and can be viewed as the number of outcomes that have c Cokes and $(n-c)$ not cokes. So, we could write the Binomial coefficient $\binom{n}{c} = \frac{n!}{(n-c)!c!} = \frac{n!}{p!!c!}\frac{n!}{(notc)!c!}$. When there are three alternatives, rather than two, then $notc$ can be either Pepsi or gelato. So, I am guessing that $\frac{n!}{c!p!g!}$ is the number of different ways one can get c Cokes, p Pepsi and g gelati in n draws. If so, the the probability of observing two Pepsi, five Coke and three gallons of gelato is $\frac{10!}{5!2!3!}(\Pr_C)^5(\Pr_P)^2(1-\Pr_C-\Pr_P)^3$. Note that $\frac{10!}{5!2!3!} = 2520$ is the number of different of possible outcomes that have five Coke, two Pepsi and two gelati.

    If, for example, $\Pr_C = .37$, $\Pr_P = .3$ and $\Pr_G = .33$, then $\frac{10!}{5!2!3!}(.37)^5(.3)^2(.33)^3 = 5.6519 \times 10^{-2}$

    More generally, $f(n,c,p) = \frac{n!}{c!p!(n-c-p)!}(\Pr_C)^c(\Pr_P)^p(1-\Pr_C-\Pr_P)^{n-c-p}$. This is the multinomial distribution with three alternatives.

37. The "laws"/ "theorems" listed below can be derived from the basic definitions and properties of sets.

    Communitve law: $A \cup B = B \cup A$ and $A \cap B = B \cap A$

    Associate law: $A \cup (B \cup C) = (A \cup B) \cup C$ and $A \cap (B \cap C) = (A \cap B) \cap C$

    Distributive law: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

    $(\overline{\overline{A}}) = A$: the compliment of the compliment of A is A.

    $A \cap \Omega = A$, $A \cup \Omega = \Omega$, $A \cap \emptyset = \emptyset$, and $A \cup \emptyset = A$

    $A \cap \overline{A} = \emptyset$, $A \cup \overline{A} = \Omega$, $A \cap A = A$ and $A \cup A = A$

    $\overline{A \cup B} = \overline{A} \cap \overline{B}$ and $\overline{A \cap B} = \overline{A} \cup \overline{B}$ called *De Morgan's laws*

    $A/B = A \cap \overline{B}$

    $A = AB \cup A\overline{B}$ and $AB \cap A\overline{B} = \emptyset$

    If $A \subset B$, then $A \cap B = A$, and $A \cup B = B$

    Be prepared to convince someone that each of these theorems follow from the basic properties of sets. Venn diagrams might help you to be convincing, so would formal proofs, if your audience understands such things.

38. Assume that everytime one has a kid there is a 50% chance that it is a female. Comment on the following: Fred and Mabel have two kids (two boys, or two girls, or a boy and a girl) and the probability that they are both boys is 1/3.

39. Consider a continuous random variable with density $f_X(x)$. Describe in word $f_X(4)$. Now consider a discrete random variable with probability density function $f_X(x)$ and describe in words $f_X(4)$.

    answer: For the discrete rv X, $f_X(4)$ is the probability that X takes the value 4; it with be a non-negative number less than one. On the other hand, if X is a continuous random variable, $f_X(4)$ is simply the height of the density function at $X = 4$. It is a non-negative number than is not bounded from above by one (it can be greater than one). It is not the probability that $X = 4$ (the probability that X is exactly 4 is zero).

40. Consider the population of individuals who have entered the graduate program in Economics at CU. Further assume that the distribution of individuals in this population in terms of remaing in the program at the end of their first year, second year, ... is not a function of the year in which they started the program.

    In Fall 2007, Pat took a random sample of 15 students from the incoming class. Call this sample s_1.By the fall of 2008 4 of these students had flunked out of the program; by fall 2009, 7 had flunked out of the program.

    In Fall 2008, Pat took a random sample of 25 students from the incoming class. Call this sample, s_2. By the fall of 2009, 9 of these students had flunked out of the program.

What is your best estimate of the probability of surviving in the program at least two years without flunking out?

answer: Consider some different estimates of potentially relevant probabilities. Based on the first sample only, the estimated probability of surviving at least one year is $\widehat{\Pr}_{s_1}(1) = 11/15 = 0.733\,33$. Based on the first sample only, the estimated probability of surviving at least two years is $\widehat{\Pr}_{s_1}(2) = 8/15 = 0.533\,33$.

Using the first sample only, the estimated conditional probability of surviving a second year given that one has survived the first year is $\widehat{\Pr}_{s_1}(2\,|1\,) = 8/11 = 0.727\,27$. Note that one can estimate this conditional probability only with the first sample.

Using the second sample, the estimated probability of surviving at least one year is $\widehat{\Pr}_{s_2}(1) = 16/25 = 0.64$. Note that $\widehat{\Pr}_{s_2}(1) \neq \widehat{\Pr}_{s_1}(1)$. There is not a $\widehat{\Pr}_{s_2}(2)$ nor a $\widehat{\Pr}_{s_1}(2\,|1\,)$.

What is the best estimate of Pr(1)? Note that we have data for the first year from both samples. They are both random samples from the sample population, so can be combined. So $\widehat{\Pr}_{s_1+s_2}(1) = \frac{11+16}{15+25} = 0.675$. This is a better estimate of Pr(1) than is $\widehat{\Pr}_{s_1}(1)$ or $\widehat{\Pr}_{s_2}(1)$ because it uses all of the data - it is more efficient.

Remember we are looking for the best estimate of Pr(2). We know that

$$\Pr(2) = \Pr(2\,|1\,)\Pr(1)$$

We have only one estimate of $\Pr(2\,|1\,)$, but three estimates of Pr(1), the best of which is $\widehat{\Pr}_{s_1+s_2}(1) = 0.675$. So our best estimate of Pr(2) is $\widehat{\Pr}(2) = \widehat{\Pr}_{s_1}(2\,|1\,)\widehat{\Pr}_{s_1+s_2}(1) = (0.727\,27)(0.675) = 0.490\,91$ which is a better estimate than $\widehat{\Pr}_{s_1}(2) = 0.533\,33$

We used the data for the 2008 class even though they were only observed for one year, to get a better estimate of the probability of lasting at least two years - wow.

41. Everyday for the last ten years you have recored the temperature at noon. You want to "model" and "predict what the temperature will be tomorrow at noon. That is, you want to "estimate" tomorrow's temperature - tomorrow's temperature is a random variable. God tells you that tomorrow's termperature is not a function of any explanatory variables. **Propose an estimator** for tomorrow's temperature, explain to me why you chose this estimator, and tell me how you would use it to estimate tomorrow's weather.

    answer: You problem is to estimate the density function for tomorrow's temperature, $f(t)$; $f(t)$ is unknown. I would propose the following estimate and estimator for $f(t)$: the estimate, $\widehat{f}(t)$, is a plot of your **data** (your sample) with temperatures on the horizontal axis, and the proportion of days at each temperature on the vertical axis, the estimator using the proportions from your data to estimate $f(t)$. Note that your data set is being plugged into the estimator to estimate the population density function for tomorrow's weather. Every different sample of past temperatures will generate a different estimate of $f(t)$.

As far as predicting tomorrow's temperature: Your predictions will be of the following form, $\Pr[a \leq t \leq b] = \int_a^b \widehat{f}(t)dt$: the probability that the temperature tomorrow will be between a and b. This is an interval estimate. The average temperature over the last ten years would be a point estimate.

I chose the estimator I did for $f(t)$ using the following logic. I knew I had to choose an $\widehat{f}(T)$, an estimator for $f(T)$, with $\widehat{f}(t) \geq 0$ $\forall t$ and $\int_{-\infty}^{+\infty}$ $\widehat{f}(t)dt = 1$, otherwise, $\widehat{f}(t)$ would not be an estimated density function.

My first inclination was to specify a specific function form for $f(t)$, for example, the normal distribution. If I had followed that path I would have then used my data to estimate the parameters of my chosen distribution (the mean and variance if I had chosen the normal).

But I rejected this idea. Maybe temperature is not Guber distributed, or normal, or whatever I arbitrarily might assume for the distribution. Why not look at the data and see what the distribution of temperatures looks like? That is, use the estimated distribution, with no functional form imposed, as an estimate of the popultation distribution.

What I proposed is a non-parameteric estimator - no function form was imposed on $f(t)$.

comments on answers:

They were difficut to grade. Most have serious problems but what is wrong varys. Some of you, many of you, were compelled to assume that tomorrow's temperature depends on recent events, even though you were expliticly told it did not. Note that I have changed the words of the question to make clearer, I hope, that you not do do.

Your data is consistent with the temperature being normally distributed? No, but maybe as an approximation. That said, if you assumed temperature was normally distributed you were at least talking about a density function. Some of you never mentioned or specified a density function. The problem was to "specifiy the density function for tomorrow's weather". Did you do that? Did the word *density* ever appear in your answer?

Did you write down a deterministic model, a model where t_i is not a r.v.? If it is not a r.v., there is no density function, and nothing to estimate. Sometimes I had no way of determining whether you model included any random variables. If you write down a model of t_i as a function of past temperatures, what exactly is random? You need to be explicit.

Some of you wanted to estimate tomorrow's termperature based on peoples' expectations of tomorrow's temperature. Are you assuming expectations have nothing to do with recent events?

Some of you said that would would use the average temperature for the last 10 years as your estimate. Fair enough. I called this a "good start." You actually have a lot more information that than the 10-year average. You have the observed sampling distribution. I less liked estimators that put more weight on more recent temperatures.

42. Consider three players, a,b, and c, playing a in a tennis tournament consisting of games. Only two players can play at once. Assume no game ends in a tie. The tournament starts with a playing b in the first game, the winner of the game then plays c. That is, the loser of a game is replaced in the next game by the individual who was not playing in that game. The tournament proceeds until someone wins two games in a row. Identify the sample space for this tournament. Try to explain your characterization of the sample space. Then discuss whether the sample space is finite or infinite and, if infinite, countable or not countable.

    answer: The sample space consists of all possible tournaments: a sequence of games that ends with the same person winning the last two games, and a tournament that never ends. Following Hakon's suggestion let a_c denote a beat c, and a_b denote a beat b. The sample space includes two sequences (omitting the subscripts is fine as well).

    The sequence of sample points when a wins the first game: $a_b a_c$, $a_b c_a c_b$, $a_b c_a b_c b_a$, $a_b c_a b_c a_b a_c$, $a_b c_a b_c a_b c_a c_b$, $a_b c_a b_c a_b c_a b_c b_a$, $a_b c_a b_c a_b c_a b_c a_b a_c$, $a_b c_a b_c a_b c_a b_c a_b c_a c_b$, $a_b c_a b_c a_b c_a b_c a_b c_a b_c b_a$, $a_b c_a b_c a_b c_a b_c a_b c_a b_c a_b a_c$

    The sequence of sample points when b wins the first game: $b_a b_a$, $b_a c_b c_a$, $b_a c_b a_c a_b$, $b_a c_b a_c b_a b_c$, $b_a c_b a_c b_a c_b c_a$, $b_a c_b a_c b_a c_b a_c a_b$, $b_a c_b a_c b_a c_b a_c b_a b_c$

    The next sample point in each sequence is created by dropping the tournament ending last game from the previous sample point, and adding two won games by the player who was sitting out, first against the winner of the second-to-last game in the previous sample point, then, in the last game, beating the other guy.

    Another way to represent the sequence where a wins first is, with tournament ending games in bold,

$1st$ game	$2nd$	$3rd$	$4th$	$5th$	$6th$	$7th$	$8th$	
a_b	a_ba_c							
	a_bc_a	$a_bc_ac_b$						
		$a_bc_ab_c$	$a_bc_ab_cb_a$					
			$a_bc_ab_ca_b$	$a_bc_ab_ca_ba_c$				
				$a_bc_ab_ca_bc_a$	$a_bc_ab_ca_bc_ac_b$			
					$a_bc_ab_ca_bc_ab_c$	$a_bc_ab_ca_bc_ab_cb_a$		
						$a_bc_ab_ca_bc_ab_ca_b$	$a_bc_ab_ca_bc_ab_ca_ba_c$	
							$a_bc_ab_ca_bc_ab_ca_bc_a$	a_bc
								a_bc

The sample points are bolded, not the unbolded ones. Some of the student answers suggest that all of the above sequences are sample points, which is not correct.

The tournament can, in theory, last forever, so the sample space has an infinite number of points. But the number is countable: tournaments with two games, three games, four games, ...

Priti,in one part of her answer, defined the sample space as $\Omega = \{(a,i),(b,i),(c,i)\}$ $i = 1,2,...n$, where first element in the pair is who won the tournament, and i is the number of games it took to win. This is an interesting start but a bit flawed: it does not allow the tournament (sample pt.) where the tournament continues forever, c cannot win when i is a number other than $3,6,9,...$, etc. It is also does not show what the sequence has to be for, for example a to win in i games.

Zack had an interesting notation. He defined a^c to mean a beats b while c sits out. For the a wins first sequence: he defined $a^c c^b b^a$ as a set, then $\frac{a^c c^b b^a}{j} \equiv \frac{a_1^c c_1^b b_1^a a_2^c c_2^b b_2^a \ldots a_j^c c_j^b b_j^a}{j}$ as a sequence of j sets. Then, a tournament (sample point) won by a in the $j+1$ set is denoted $\frac{a^c c^b b^a a_{j+1}^c a_{j+2}^b}{j}$. A tournament (sample point) won by b in the $j+1$ set is denoted $\frac{a^c c^b b^a b_{j+1}^c}{j}$. And, a tournament (sample point) won by c in the $j+1$ set is denoted $\frac{a^c c^b b^a c_{j+1}^a c_{j+1}^b}{j}$. Another way of characterizing sample points.

Note that the probability of winning a game (relative abilities) is completely immaterial to the question, and the next question as well.

43. Now imagine the tournament in the previous question ends after four games. What is the probability that the tournament was won by player c? Explain your answer.

answer: The following is the possible sequence of winners in a tournament that has a maximum of four games (remember for the tournament to win the same player has to win the last two games): a_ba_c, $a_bc_ac_b$, $a_bc_ab_cb_a$, b_ab_a, $b_ac_bc_a$, and $b_ac_ba_ca_b$. It is impossible for player c to win a tournament that ends after four games, so **the probability is zero**; if the tournament ends after four games, either a or b must win the tournament: player c could be the winner after three games; in fact, if the tournament ends after three games, player c must be the winner. If player c wins the tournament it must be in game 3, or 6, or 9, or or ? Alternatively, player a can win in game 2, or 4, or 5, or 7, or 8. Player b can also win the tournament in these numbers of games. It is kind of cool to figure this stuff out.

Hakon formalized this to **if a wins the first game**,

a can only win the tournament after $3k-1$ games with k being an integer greater than or equal to 1.

b can only win the tournament after $3k+1$ games.

c can only win the tournament after $3k$ games.

if b wins the first game,

b can only win the tournament after $3k-1$ games.

a can only win the tournament after $3k+1$ games.

c can only win the tournament after $3k$ games.

You would be amazed, I was, at the number of students who proved, with a diagram, that c could not win the tournament in the 4th game of the tournament, but then told me the probability of c winning in the $4th$ game was positive.

44. So, you observe a sample that consists of two individuals, Wanda and her husband Alfred. How many observations are in this sample? Then you observe a sample that consists of three coins heads up and one coin tails up. How many observations are in this sample? What is the point of these two questions?

    answer: In the first case the answer is either one or two: one if one is sampling from the population of couples, two if one is sampling from the population of individuals. In the coin example, the sample size is one, two or four. If the population consists of sets of four coin tosses, the sample size is one. If the population consists of sets of one coin toss, the sample size is four. If the population consists of sets of two coin tosses, the sample size is two.

    What is the point? Until the population is defined, sample size is indeterminate.

    You remember my emphatic point that. "One cannot talk about sampling until the population of interest has been defined."

    I was surprised, a lot, by some of the answers to this question. An observation and an event are not the same thing: an event is a property that a sample does ot does not have have (either the observations weighs more than a ton, or it does not, where the event is *weighs more than a ton)*. The observation occurs whether it is or is not an occurance of a particular event. An observation is a draw: when one samples, one takes draws, the number of observations is the number of draws from the population in the sample.

If the population is couples and the sample is the couple Wanda and Alfred, there is one observation and order is immaterial: there is no order. If the population is sets of of four coin tosses, and one observes 3 heads and on tail, this is one observation, independent of the order is which the heads and tell occured if the tosses were sequential. Order is also immaterial is all the four coins were tossed at once - no order.

45. A fair coin is flipped an infinite number of times. What is the probability of a sequence of flips with 992 tails in a row?

    answer: ONE. In an infinite number of flips every finite sequence will occur with probability one; 992 tails in a row is a finite sequence, so is head-tail-head a million times, or heads a billion times in a row.

    Hakon adds the following insights: If a fair coin is tossed 992 times the probability of all tails is $(.5)^{992}$ very low. Then consider T flips, where $T > 993$. Then the probability of 992 tails in a row is greater, and this probabilityis increasing in T. And, as T approaches infinifity, it approaches 1. And with an infinite number of flips it is one.

46. **now on prob review sheet.** So assume that there are 32 students in 7818. You might not have noticed, you can tell if you look closely, but 12 of the students are zombies (the "walking dead"). You randomly sample 9 students from class, what is the probability that $k, k = 0, 1, 2, ..., 9$ of those in your sample are zombies. Show all of your work, and explain, in words what you are doing and why? Then report the general formula for $\Pr[k]$ for a population of M with K zombies and a random sample size of n.

    answer: Note that by the way the question is specified, the order in which the k zombies are drawn is matterial.

    Assuming sampling **without replacement**, and Ignoring order, there are $\binom{M}{n} = \binom{32}{9} = 2.8049 \times 10^7$ equally likely samples. If a sample of 9 has k zombies, it must have $9-k$ non-zombies. $\binom{K}{k} = \binom{12}{k}$ is, ignoring order, the number of different ways to choose k zombies out of 12 zombies. $\binom{M-K}{n-k} = \binom{32-12}{9-k}$ is, ignoring order, the number of ways to choose $9-k$ non-zombies out of $32-12 = 20$ non-zombies. So, since any choice of k zombies can be combined with any choice of $9-k$ non-zombies, $\binom{K}{k}\binom{M-K}{n-k} = \binom{12}{k}\binom{32-12}{9-k}$ is the number of equally likely samples with k zombies and $9-k$ non-zombies. So, the proportion samples with the correct property is

    $\Pr[k] = \frac{\binom{K}{k}\binom{M-K}{n-k}}{\binom{M}{n}} = \frac{\binom{12}{k}\binom{32-12}{9-k}}{\binom{32}{9}} = \frac{1}{28\,048\,800}\binom{12}{k}\binom{20}{9-k}$, $k = 01, 2, ..., 9$.

    As required, $\sum_{k=0}^{9} \frac{1}{28\,048\,800}\binom{12}{k}\binom{20}{9-k} = 1.0$

    For example, if $k = 3$, $\Pr[3] = \frac{1}{28\,048\,800}\binom{12}{3}\binom{20}{9-3} = 0.304\,01$

One can write $\Pr[k] = \frac{\binom{K}{k}\binom{M-K}{n-k}}{\binom{M}{n}} = \frac{C_K^k C_{M-K}^{n-k}}{C_M^n} = \frac{\binom{n}{k}(K)_k(M-K)_{n-k}}{(M)_n}$

The probability density function $\Pr[k] = \frac{\binom{K}{k}\binom{M-K}{n-k}}{\binom{M}{n}}$ is called the *hypergeometric* distribution. The corresponding CDF is $\Pr[k \le x] = \sum_{k=0}^{x} \frac{\binom{K}{k}\binom{M-K}{n-k}}{\binom{M}{n}}$.

Calculating the different probabilities

$\text{HypergeomDen}(0; 32, 12, 9) = 5.9881 \times 10^{-3}$

$\text{HypergeomDen}(1; 32, 12, 9) = 5.3893 \times 10^{-2}$

$\text{HypergeomDen}(2; 32, 12, 9) = 0.18241$

$\text{HypergeomDen}(3; 32, 12, 9) = 0.30401$

$\text{HypergeomDen}(4; 32, 12, 9) = 0.27361$

$\text{HypergeomDen}(5; 32, 12, 9) = 0.13681$

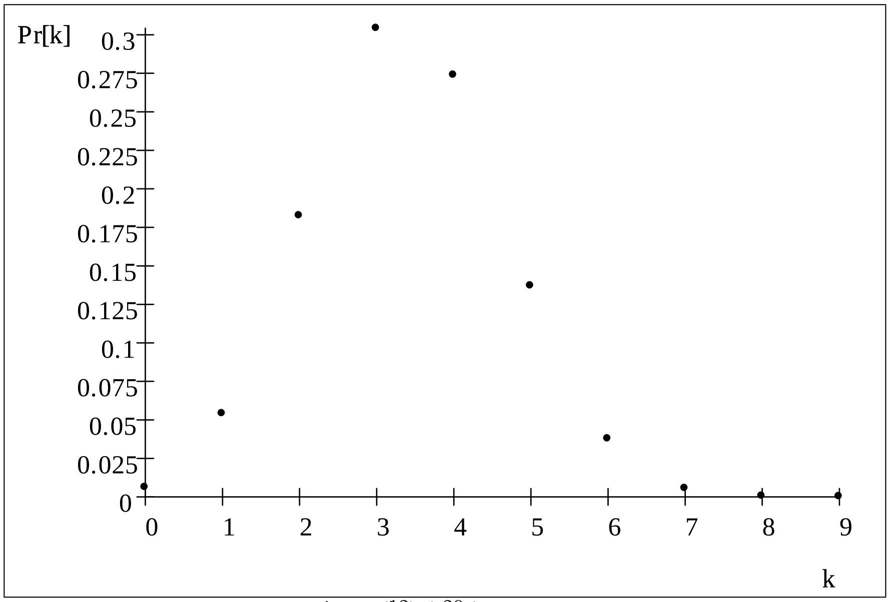
$\text{HypergeomDen}(6; 32, 12, 9) = 3.7555 \times 10^{-2}$

$\text{HypergeomDen}(7; 32, 12, 9) = 5.3649 \times 10^{-3}$

$\text{HypergeomDen}(8; 32, 12, 9) = 3.5296 \times 10^{-4}$

$\text{HypergeomDen}(9; 32, 12, 9) = 7.8435 \times 10^{-6}$

Graphing this discrete probability distribution



$\Pr[k] = \frac{1}{28\,048\,800}\binom{12}{k}\binom{20}{9-k}$, $k = 0, 1, 2, ..., 9$

Note that $\Pr[k] = \frac{\binom{12}{k}\binom{20}{9-k}}{\binom{32}{9}} = \frac{\binom{9}{k}\frac{12!}{(12-k)!}\frac{20!}{(20-(k-1))!}}{\frac{32!}{(32-9)!}} = \frac{\binom{9}{k}(12)_k(20)_{(k-1)}}{(32)_9}$, so these are different ways of writing $\Pr[k]$.

When I first made up the question, I did not consider sampling with replacement - sampling without replacement seemed more likely, but many of you did and answered the question both ways. Consider the answer if one samples without replacement.

When there is sampling with replacement on each of the n trials one either draws a zombie, or not, so there are two possible outcomes. Since there is sampling with replacement, the trials are independent. So the $\Pr[k]$, the probability of k sucesses (zombies) in n trials, is a binomial, $\Pr[k] = \binom{n}{k} p^k (1-p)^{n-k}$ where p is the probability of zombie on a trial $= 12/32 = \frac{3}{8}$, so $\Pr[k] = \binom{n}{k}(\frac{3}{8})^k(\frac{5}{8})^{n-k}$. Another way to write this is $\Pr[k] = \frac{\binom{n}{k} K^k (M-K)^{n-k}}{M^n} = \frac{\binom{n}{k} 12^k (20)^{n-k}}{32^n}$

$\Pr[0] = \binom{9}{0}(\frac{3}{8})^0(\frac{5}{8})^9 = 1.455\,2 \times 10^{-2}$ which is $\frac{\binom{9}{0k} 12^0 (20)^{9-0}}{32^9} = 1.455\,2 \times 10^{-2}$

$\Pr[1] = \binom{9}{1}(\frac{3}{8})^1(\frac{5}{8})^{9-1} = 0.078\,58$

$\Pr[2] = \binom{9}{2}(\frac{3}{8})^2(\frac{5}{8})^{9-2} = 0.188\,59$

When I wrote the question I did not realize it was on page 29 of MGB.

As some of you correctly noted, and as I asked on the inclass final, MGB would say you cannot randomly sample from this finite popultion if one samples without replacement.

47. Consider some of the ways to characterize an individual: Let $B = 1$ if the individual went to graduate school in business or is in the process of going, and zero otherwise. Let $M = 1$ if the individual meditates at least two hours a day; let $C = 1$ if the individual is a capitalist pig, and let $R = 1$ if the individual is a member of the religious right.

I am interested in the probability that an individual is a meditating capitalist pig given that he went to business school, $\Pr[MC\,|B]$. We know that the probability that one went to business school conditional on being a meditating capitalist pig is .4

Part 1: Determine $\Pr[MC\,|B]$ and what happens to $\Pr[MC\,|B]$ if $\Pr[MC]$ increases.

answer to part 1:

By definition of the conditional probability $\Pr[MC\,|B] = \frac{\Pr[BMC]}{\Pr[B]}$

We know by Bayes' Theorem, that $\Pr[MC\,|B] = \frac{\Pr[B|MC]\Pr[MC]}{\Pr[B]}$

So $\Pr[MC\,|B] = \frac{\Pr[B|MC]\Pr[MC]}{\Pr[B]} = \frac{.4\Pr[MC]}{\Pr[B]}$

What happens to $\Pr[MC\,|B]$ when $\Pr[MC]$ increases? This is tricky. Playing around (note that we can no longer assume $\Pr[B\,|MC]$ is a constant equal to .4 or that $\Pr[B]$ stays the same:

$$
\begin{aligned}
\frac{\partial(\Pr[MC\,|B\,])}{\partial(\Pr\,[MC])} &= \frac{\partial(\frac{\Pr[B|MC\,]\Pr[MC]}{\Pr[B]})}{\partial(\Pr\,[MC])} \\
&= \frac{\partial(\Pr[B\,|MC\,]\Pr\,[MC]\,(\Pr[B])^{-1})}{\partial(\Pr\,[MC])} \\
&= \frac{(\Pr[B])^{-1})\partial(\Pr[B\,|MC\,]\Pr\,[MC])}{\partial(\Pr\,[MC])} \\
&\quad -\Pr[B\,|MC\,]\Pr\,[MC]\,(\Pr[B])^{-2})\frac{\partial\Pr[B]}{\partial(\Pr\,[MC])} \\
&= (\Pr[B])^{-1})\left[\Pr[B\,|MC\,] + \Pr\,[MC]\,\frac{\partial(\Pr[B\,|MC\,])}{\partial(\Pr\,[MC])}\right] \\
&\quad -\Pr[B\,|MC\,]\Pr\,[MC]\,(\Pr[B])^{-2})\frac{\partial\Pr[B]}{\partial(\Pr\,[MC])}
\end{aligned}
$$

But before $\Pr\,[MC]$ changes, $\Pr[B\,|MC\,] = .4$, so plugging in this information

$$
\begin{aligned}
&\frac{\partial(\Pr[MC\,|B\,])}{\partial(\Pr\,[MC])} \\
&= (\Pr[B])^{-1})\left[.4 + \Pr\,[MC]\,\frac{\partial(\Pr[B\,|MC\,])}{\partial(\Pr\,[MC])}\right] - .4\Pr\,[MC]\,(\Pr[B])^{-2})\frac{\partial\Pr[B]}{\partial(\Pr\,[MC])}
\end{aligned}
$$

Maybe I even did that correctly - without any independence assumptions everything can depend on everything else. So, the answer, in general, is who knows what happens to $\Pr[MC\,|B\,]$ when $\Pr\,[MC]$ increases - it looks like it can go up or down. If one were willing to add the restrictions that $\frac{\partial(\Pr[B|MC\,])}{\partial(\Pr[MC])} = \frac{\partial\Pr[B]}{\partial(\Pr[MC])} = 0$, $\frac{\partial(\Pr[MC|B\,])}{\partial(\Pr[MC])}$ simplifies to $\frac{.4}{\Pr[B]} > 0$. This question is more complicated than I originally thought.

Part 2:

Now add the assumption that whether one is a capitalist pig is independent of whether one meditates, and the assumption that the probability one meditates equals the probability one went to business school. In this case what more can you tell me about the probability that one is a meditating capitalist pig given that one has been to business school.

answer to part 2: Without these additional assumption we know, from above, that $\Pr[MC\,|B\,] = \frac{.4\Pr[MC]}{\Pr[B]}$. Independence of M and C implies that $\Pr\,[MC] = \Pr[M]\Pr[C]$, so with this independence restriction $\Pr[MC\,|B\,] = \frac{.4\Pr[M]\Pr[C]}{\Pr[B]}$. If we add to this the restriction that $\Pr[M] = \Pr[B]$, we get $\Pr[MC\,|B\,] = .4\Pr[C]$. In words, the probability that one is a meditating capitalist pig, conditional on going to business school is .4 multiplied by the probability that one is a capitalist pig. However, what about the $\Pr[B\,|MC\,] = .4$ part? Remember the restriction

$\Pr[M] = \Pr[B]$; I am getting confused. **We need to think about this questions some more.**

48. Some people, quite a few, are named Shirley- they suffer from *Shirleyness*, others, quite a few, are named Bob - they suffer from *Bobness*. Bobness and Shirleyness are disjoint. Bobness and Shirleyness are independent.

    Comment in an informative and intuitive manner.

    answer: this statement is incorrect. Given that $\Pr[B] > 0$, $\Pr[S] > 0$, and $B \cap S = \varphi$, B and S cannot be independent. The intuition is straightforward: independence is defined as $\Pr[B \cap S] = \Pr[B] \Pr[S]$. But this can't be because $\Pr[B] \Pr[S] > 0$ but $\Pr[B \cap S] = 0$