Stat 110 Strategic Practice 1, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1 Naive Definition of Probability

1. For each part, decide whether the blank should be filled in with $=$, $<$, or $>$, and give a short but clear explanation.

(a) (probability that the total after rolling 4 fair dice is 21) ____ (probability that the total after rolling 4 fair dice is 22)

(b) (probability that a random 2 letter word is a palindrome[1]) ____ (probability that a random 3 letter word is a palindrome)

2. A random 5 card poker hand is dealt from a standard deck of cards. Find the probability of each of the following (in terms of binomial coefficients).

(a) A flush (all 5 cards being of the same suit; do not count a royal flush, which is a flush with an Ace, King, Queen, Jack, and 10)

(b) Two pair (e.g., two 3's, two 7's, and an Ace)

3. (a) How many paths are there from the point $(0, 0)$ to the point $(110, 111)$ in the plane such that each step either consists of going one unit up or one unit to the right?

(b) How many paths are there from $(0, 0)$ to $(210, 211)$, where each step consists of going one unit up or one unit to the right, and the path has to go through $(110, 111)$?

4. A *norepeatword* is a sequence of at least one (and possibly all) of the usual 26 letters a,b,c,...,z, with repetitions not allowed. For example, "course" is a norepeatword, but "statistics" is not. Order matters, e.g., "course" is not the same as "source".

A norepeatword is chosen randomly, with all norepeatwords equally likely. Show that the probability that it uses all 26 letters is very close to $1/e$.

[1] A *palindrome* is an expression such as "A man, a plan, a canal: Panama" that reads the same backwards as forwards (ignoring spaces and punctuation). Assume for this problem that all words of the specified length are equally likely, that there are no spaces or punctuation, and that the alphabet consists of the lowercase letters a,b,...,z.

2 Story Proofs

5. Give a story proof that $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

6. Give a story proof that

$$\frac{(2n)!}{2^n \cdot n!} = (2n-1)(2n-3)\cdots 3 \cdot 1.$$

7. Show that for all positive integers n and k with $n \geq k$,

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k},$$

doing this in two ways: (a) algebraically and (b) with a "story", giving an interpretation for why both sides count the same thing.

Hint for the "story" proof: imagine $n+1$ people, with one of them pre-designated as "president".

Stat 110 Strategic Practice 1 Solutions, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1 Naive Definition of Probability

1. For each part, decide whether the blank should be filled in with $=$, $<$, or $>$, and give a short but clear explanation.

(a) (probability that the total after rolling 4 fair dice is 21) $\underline{\;>\;}$ (probability that the total after rolling 4 fair dice is 22)

Explanation: All *ordered* outcomes are equally likely here. So for example with two dice, obtaining a total of 9 is more likely than obtaining a total of 10 since there are two ways to get a 5 and a 4, and only one way to get two 5's. To get a 21, the outcome must be a permutation of $(6, 6, 6, 3)$ (4 possibilities), $(6, 5, 5, 5)$ (4 possibilities), or $(6, 6, 5, 4)$ ($4!/2 = 12$ possibilities). To get a 22, the outcome must be a permutation of $(6, 6, 6, 4)$ (4 possibilities), or $(6, 6, 5, 5)$ ($4!/2^2 = 6$ possibilities). So getting a 21 is more likely; in fact, it is exactly twice as likely as getting a 22.

(b) (probability that a random 2 letter word is a palindrome[1]) $\underline{\;=\;}$ (probability that a random 3 letter word is a palindrome)

Explanation: The probabilities are equal, since for both 2-letter and 3-letter words, being a palindrome means that the first and last letter are the same.

2. A random 5 card poker hand is dealt from a standard deck of cards. Find the probability of each of the following (in terms of binomial coefficients).

(a) A flush (all 5 cards being of the same suit; exclude the possibility of a royal flush, which is a flush with an Ace, King, Queen, Jack, and 10)

A flush can occur in any of the 4 suits (imagine the tree, and for concreteness suppose the suit is Hearts); there are $\binom{13}{5}$ ways to choose the cards in that suit, except for one way to have a royal flush in that suit. So the probability is

$$\frac{4\left(\binom{13}{5} - 1\right)}{\binom{52}{5}}.$$

[1] A *palindrome* is an expression such as "A man, a plan, a canal: Panama" that reads the same backwards as forwards (ignoring spaces and punctuation). Assume for this problem that all words of the specified length are equally likely, that there are no spaces or punctuation, and that the alphabet consists of the lowercase letters a,b,...,z.

(b) Two pair (e.g., two 3's, two 7's, and an Ace)

Choose the two ranks of the pairs, which specific cards to have for those 4 cards, and then choose the extraneous card (which can be any of the $52-8$ cards not of the two chosen ranks). This gives that the probability of getting two pairs is

$$\frac{\binom{13}{2} \cdot \binom{4}{2}^2 \cdot 44}{\binom{52}{5}}.$$

3. (a) How many paths are there from the point $(0, 0)$ to the point $(110, 111)$ in the plane such that each step goes either one unit up or one unit to the right?

Encode a path as a sequence of U's and R's, like $URURURUUUR \dots UR$, where U and R stand for "up" and "right" respectively. The sequence must consist of 110 R's and 111 U's, and to determine the sequence we just need to specify where the R's are located. So there are $\binom{221}{110}$ possible paths.

(b) How many paths are there from $(0, 0)$ to $(210, 211)$, where each step goes either one unit up or one unit to the right, and the path has to go through $(110, 111)$?

There are $\binom{221}{110}$ paths to $(110, 111)$, as above. From there, we need 100 R's and 100 U's to get to $(210, 211)$, so by the multiplication rule the number of possible paths is $\binom{221}{110} \cdot \binom{200}{100}$.

4. A *norepeatword* is a sequence of at least one (and possibly all) of the usual 26 letters a,b,c,...,z, with repetitions not allowed. For example, "course" is a norepeatword, but "statistics" is not. Order matters, e.g., "course" is not the same as "source".

A norepeatword is chosen randomly, with all norepeatwords equally likely. Show that the probability that it uses all 26 letters is very close to $1/e$.

Solution: The number of norepeatwords having all 26 letters is the number of ordered arrangements of 26 letters: 26!. To construct a norepeatword with $k \leq 26$ letters, we first select k letters from the alphabet ($\binom{26}{k}$ selections) and then arrange them into a word ($k!$ arrangements). Hence there are $\binom{26}{k} k!$ norepeatwords with k letters, with k ranging from 1 to 26. With all norepeatwords equally likely, we have

$$\begin{aligned} P(\text{norepeatword having all 26 letters}) &= \frac{\#\text{ norepeatwords having all 26 letters}}{\#\text{ norepeatwords}} \\ &= \frac{26!}{\sum_{k=1}^{26} \binom{26}{k} k!} = \frac{26!}{\sum_{k=1}^{26} \frac{26!}{k!(26-k)!} k!} \\ &= \frac{1}{\frac{1}{25!} + \frac{1}{24!} + \ldots + \frac{1}{1!} + 1}. \end{aligned}$$

The denominator is the first 26 terms in the Taylor series $e^x = 1 + x + x^2/2! + \ldots$, evaluated at $x = 1$. Thus the probability is approximately $1/e$ (this is an *extremely* good approximation since the series for e converges very quickly; the approximation for e differs from the truth by less than 10^{-26}).

2 Story Proofs

5. Give a story proof that $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

Consider picking a subset of n people. There are $\binom{n}{k}$ choices with size k, on the one hand, and on the other hand there are 2^n subsets by the multiplication rule.

6. Give a story proof that

$$\frac{(2n)!}{2^n \cdot n!} = (2n-1)(2n-3)\cdots 3 \cdot 1.$$

Take $2n$ people, and count how many ways there are to form n partnerships. We can do this by lining up the people in a row and then saying the first two are a pair, the next two are a pair, etc. This overcounts by a factor of $2^n \cdot n!$ since the order of pairs doesn't matter, nor does the order within each pair. Alternatively, count the number of possibilities by noting that there are $2n - 1$ choices for the partner of person 1, then $2n - 3$ choices for person 2 (or person 3, if person 2 was already paired to person 1), etc.

7. Show that for all positive integers n and k with $n \geq k$,

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k},$$

doing this in two ways: (a) algebraically and (b) with a "story", giving an interpretation for why both sides count the same thing.

Hint for the "story" proof: imagine $n+1$ people, with one of them pre-designated as "president".

Solution: For the algebraic proof, start with the definition of the binomial coefficients in the left-hand side, and do some algebraic manipulation as follows:

$$\begin{aligned}\binom{n}{k} + \binom{n}{k-1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)!} \\ &= \frac{(n-k+1)n! + (k)n!}{k!(n-k+1)!} \\ &= \frac{n!(n+1)}{k!(n-k+1)!} \\ &= \binom{n+1}{k}.\end{aligned}$$

For the "story" method (which proves that the two sides are equal by giving an interpretation where they both count the same thing in two different ways), consider $n+1$ people, with one of them pre-designated as "president". The right-hand side is the number of ways to choose k out of these $n+1$ people, with order not mattering. The left-hand side counts the same thing in a different way, by considering two cases: the president is or isn't in the chosen group.

The number of groups of size k which include the president is $\binom{n}{k-1}$, since once we fix the president as a member of the group, we only need to choose another $k-1$ members out of the remaining n people. Similarly, there are $\binom{n}{k}$ groups of size k that don't include the president. Thus, the two sides of the equation are equal.

Stat 110 Homework 1, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1. A certain family has 6 children, consisting of 3 boys and 3 girls. Assuming that all birth orders are equally likely, what is the probability that the 3 eldest children are the 3 girls?

2. (a) How many ways are there to split a dozen people into 3 teams, where one team has 2 people, and the other two teams have 5 people each?

(b) How many ways are there to split a dozen people into 3 teams, where each team has 4 people?

3. A college has 10 (non-overlapping) time slots for its courses, and blithely assigns courses to time slots randomly and independently. A student randomly chooses 3 of the courses to enroll in (for the PTP, to avoid getting fined). What is the probability that there is a conflict in the student's schedule?

4. A city with 6 districts has 6 robberies in a particular week. Assume the robberies are located randomly, with all possibilities for which robbery occurred where equally likely. What is the probability that some district had more than 1 robbery?

5. Elk dwell in a certain forest. There are N elk, of which a simple random sample of size n are captured and tagged ("simple random sample" means that all $\binom{N}{n}$ sets of n elk are equally likely). The captured elk are returned to the population, and then a new sample is drawn, this time with size m. This is an important method that is widely-used in ecology, known as *capture-recapture*.

What is the probability that exactly k of the m elk in the new sample were previously tagged? (Assume that an elk that was captured before doesn't become more or less likely to be captured again.)

6. A jar contains r red balls and g green balls, where r and g are fixed positive integers. A ball is drawn from the jar randomly (with all possibilities equally likely), and then a second ball is drawn randomly.

(a) Explain intuitively why the probability of the second ball being green is the same as the probability of the first ball being green.

(b) Define notation for the sample space of the problem, and use this to compute the probabilities from (a) and show that they are the same.

(c) Suppose that there are 16 balls in total, and that the probability that the two balls are the same color is the same as the probability that they are different colors. What are r and g (list all possibilities)?

7. (a) Show using a story proof that

$$\binom{k}{k} + \binom{k+1}{k} + \binom{k+2}{k} + \cdots + \binom{n}{k} = \binom{n+1}{k+1},$$

where n and k are positive integers with $n \geq k$.
Hint: imagine arranging a group of people by age, and then think about the oldest person in a chosen subgroup.

(b) Suppose that a large pack of Haribo gummi bears can have anywhere between 30 and 50 gummi bears. There are 5 delicious flavors: pineapple (clear), raspberry (red), orange (orange), strawberry (green, mysteriously), and lemon (yellow). There are 0 non-delicious flavors. How many possibilities are there for the composition of such a pack of gummi bears? You can leave your answer in terms of a couple binomial coefficients, but not a sum of lots of binomial coefficients.

Stat 110 Homework 1 Solutions, Fall 2011

Prof. Joe Blitzstein (Department of Statistics, Harvard University)

1. A certain family has 6 children, consisting of 3 boys and 3 girls. Assuming that all birth orders are equally likely, what is the probability that the 3 eldest children are the 3 girls?

Label the girls as $1, 2, 3$ and the boys as $4, 5, 6$. Think of the birth order is a permutation of $1, 2, 3, 4, 5, 6$, e.g., we can interpret 314265 as meaning that child 3 was born first, then child 1, etc. The number of possible permutations of the birth orders is 6!. Now we need to count how many of these have all of $1, 2, 3$ appear before all of $4, 5, 6$. This means that the sequence must be a permutation of $1, 2, 3$ followed by a permutation of $4, 5, 6$. So with all birth orders equally likely, we have

$$P(\text{the 3 girls are the 3 eldest children}) = \frac{(3!)^2}{6!} = 0.05.$$

Alternatively, we can use the fact that there are $\binom{6}{3}$ ways to choose where the girls appear in the birth order (without taking into account the ordering of the girls amongst themselves). These are all equally likely. Of these possibilities, there is only 1 where the 3 girls are the 3 eldest children. So again the probability is $\frac{1}{\binom{6}{3}} = 0.05$.

2. (a) How many ways are there to split a dozen people into 3 teams, where one team has 2 people, and the other two teams have 5 people each?

Pick any 2 of the 12 people to make the 2 person team, and then any 5 of the remaining 10 for the first team of 5, and then the remaining 5 are on the other team of 5; this overcounts by a factor of 2 though, since there is no designated "first" team of 5. So the number of possibilities is $\binom{12}{2}\binom{10}{5}/2 = 8316$. Alternatively, politely ask the 12 people to line up, and then let the first 2 be the team of 2, the next 5 be a team of 5, and then last 5 be a team of 5. There are 12! ways for them to line up, but it does not matter which order they line up in *within* each group, nor does the order of the 2 teams of 5 matter, so the number of possibilities is $\frac{12!}{2!5!5!\cdot 2} = 8316$.

(b) How many ways are there to split a dozen people into 3 teams, where each team has 4 people?

By either of the approaches above, there are $\frac{12!}{4!4!4!}$ ways to divide the people into a Team A, a Team B, and a Team C, if we care about which team is which (this is called a *multinomial coefficient*). Since here it doesn't matter which team is which, this over counts by a factor of 3!, so the number of possibilities is $\frac{12!}{4!4!4!3!} = 5775$.

3. A college has 10 (non-overlapping) time slots for its courses, and blithely assigns courses to time slots randomly and independently. A student randomly chooses 3 of the courses to enroll in (for the PTP, to avoid getting fined). What is the probability that there is a conflict in the student's schedule?

The probability of no conflict is $\frac{10 \cdot 9 \cdot 8}{10^3} = 0.72$. So the probability of there being at least one scheduling conflict is 0.28.

4. A city with 6 districts has 6 robberies in a particular week. Assume the robberies are located randomly, with all possibilities for which robbery occurred where equally likely. What is the probability that some district had more than 1 robbery?

There are 6^6 possible configurations for which robbery occurred where. There are 6! configurations where each district had exactly 1 of the 6, so the probability of the complement of the desired event is $6!/6^6$. So the probability of some district having more than 1 robbery is

$$1 - 6!/6^6 \approx 0.9846.$$

Note that this also says that if a fair die is rolled 6 times, there's over a 98% chance that some value is repeated!

5. Elk dwell in a certain forest. There are N elk, of which a simple random sample of size n are captured and tagged ("simple random sample" means that all $\binom{N}{n}$ sets of n elk are equally likely). The captured elk are returned to the population, and then a new sample is drawn, this time with size m. This is an important method that is widely-used in ecology, known as *capture-recapture*.

What is the probability that exactly k of the m elk in the new sample were previously tagged? (Assume that an elk that was captured before doesn't become more or less likely to be captured again.)

We can use the naive definition here since we're assuming all samples of size m are equally likely. To have exactly k be tagged elk, we need to choose k of the n tagged elk, and then $m - k$ from the $N - n$ untagged elk. So the probability is

$$\frac{\binom{n}{k} \cdot \binom{N-n}{m-k}}{\binom{N}{m}},$$

for k such that $0 \leq k \leq n$ and $0 \leq m - k \leq N - n$, and the probability is 0 for all other values of k (for example, if $k > n$ the probability is 0 since then there aren't even k tagged elk in the entire population!). This is known as a *Hypergeometric* probability; we will encounter these probabilities again later in the course.

6. A jar contains r red balls and g green balls, where r and g are fixed positive integers. A ball is drawn from the jar randomly (with all possibilities equally likely), and then a second ball is drawn randomly.

(a) Explain intuitively why the probability of the second ball being green is the same as the probability of the first ball being green.

This is true by *symmetry*. The first ball is equally likely to be any of the $g + r$ balls, so the probability of it being green is $g/(g + r)$. But the second ball is also equally likely to be any of the $g+r$ balls (there aren't certain balls that enjoy being chosen second and others that have an aversion to being chosen second); once we know whether the first ball is green we have information that affects our uncertainty about the second ball, but before we have this information, the second ball is equally likely to be any of the balls.

Alternatively, intuitively it shouldn't matter if we pick one ball at a time, or take one ball with the left hand and one with the right hand at the same time. By symmetry, the probabilities for the ball drawn with the left hand should be the same as those for the ball drawn with the right hand.

(b) Define notation for the sample space of the problem, and use this to compute the probabilities from (a) and show that they are the same.

Label the balls as $1, 2, \ldots, g+r$, such that $1, 2, \ldots, g$ are green and $g+1, \ldots, g+r$ are red. The sample space can be taken to be the set of all pairs (a, b) with $a, b \in \{1, \ldots, g + r\}$ and $a \neq b$ (there are other possible ways to define the sample space, but it is important to specify all possible outcomes using clear notation, and it make sense to be as richly detailed as possible in the specification of possible outcomes, to avoid losing information). Each of these pairs is equally likely, so we can use the naive definition of probability. Let G_i be the event that the ith ball drawn is green.

The denominator is $(g + r)(g + r - 1)$ by the multiplication rule. For G_1, the numerator is $g(g + r - 1)$, again by the multiplication rule. For G_2, the numerator is also $g(g + r - 1)$, since in counting favorable cases, there are g possibilities for the second ball, and for each of those there are $g + r - 1$ favorable possibilities for the first ball (note that the multiplication rule doesn't require the experiments to be listed in chronological order!); alternatively, there are $g(g - 1) + rg = g(g + r - 1)$ favorable possibilities for the second ball being green, as seen by considering 2 cases (first ball green and first ball red). Thus,

$$P(G_i) = \frac{g(g + r - 1)}{(g + r)(g + r - 1)} = \frac{g}{g + r},$$

for $i \in \{1, 2\}$, which concurs with (a).

(c) Suppose that there are 16 balls in total, and that the probability that the two balls are the same color is the same as the probability that they are different colors. What are r and g (list all possibilities)?

Let A be the event of getting one ball of each color. In set notation, we can write $A = (G_1 \cap G_2^c) \cup (G_1^c \cap G_2)$. We are given that $P(A) = P(A^c)$, so $P(A) = 1/2$. Then

$$P(A) = \frac{2gr}{(g+r)(g+r-1)} = \frac{1}{2},$$

giving the quadratic equation

$$g^2 + r^2 - 2gr - g - r = 0,$$

i.e.,

$$(g-r)^2 = g + r.$$

But $g + r = 16$, so $g - r$ is 4 or -4. Thus, either $g = 10, r = 6$, or $g = 6, r = 10$.

7. (a) Show using a story proof that

$$\binom{k}{k} + \binom{k+1}{k} + \binom{k+2}{k} + \cdots + \binom{n}{k} = \binom{n+1}{k+1},$$

where n and k are positive integers with $n \geq k$.
Hint: imagine arranging a group of people by age, and then think about the oldest person in a chosen subgroup.

Consider choosing $k+1$ people out of a group of $n+1$ people. Call the oldest person in the subgroup "Aemon." If Aemon is also the oldest person in the full group, then there are $\binom{n}{k}$ choices for the rest of the subgroup. If Aemon is the second oldest in the full group, then there are $\binom{n-1}{k}$ choices since the oldest person in the full group can't be chosen. In general, if there are j people in the full group who are younger than Aemon, then there are $\binom{j}{k}$ possible choices for the rest of the subgroup. Thus,

$$\sum_{j=k}^{n} \binom{j}{k} = \binom{n+1}{k+1}.$$

(This is sometimes called the *hockey-stick identity*, since if shown visually using Pascal's triangle it resembles a hockey-stick.)

(b) Suppose that a large pack of Haribo gummi bears can have anywhere between 30 and 50 gummi bears. There are 5 delicious flavors: pineapple (clear), raspberry (red), orange (orange), strawberry (green, mysteriously), and lemon (yellow). There are 0 non-delicious flavors. How many possibilities are there for the composition of such a pack of gummi bears? You can leave your answer in terms of a couple binomial coefficients, but not a sum of lots of binomial coefficients.

For a pack of i gummi bears, there are $\binom{5+i-1}{i} = \binom{i+4}{i} = \binom{i+4}{4}$ possibilities since the situation is equivalent to getting a sample of size i from the $n = 5$ flavors (with replacement, and with order not mattering). So the total number of possibilities is

$$\sum_{i=30}^{50} \binom{i+4}{4} = \sum_{j=34}^{54} \binom{j}{4}.$$

Applying the previous part, we can simplify this by writing

$$\sum_{j=34}^{54} \binom{j}{4} = \sum_{j=4}^{54} \binom{j}{4} - \sum_{j=4}^{33} \binom{j}{4} = \binom{55}{5} - \binom{34}{5}.$$

(This works out to 3200505 possibilities!)