Math 493. On sampling without replacement.

We discuss here problems in which \( k \) balls are drawn without replacement from an urn which contains \( r \) red and \( b \) black balls. Problems in which there are more than two types of objects can be attacked analogously.

Write \( n = r + b \). The sample space we have most often used to model the experiment of drawing the \( k \) balls is the space of permutations of \( n \) objects taken \( k \) at a time:

\[
S = \{(x_1, \ldots, x_k) : \text{each } x_i \in \{1, \ldots, n\} \text{ and } x_i \neq x_j \text{ if } i \neq j\}.
\]

Then \( |S| = n(n-1) \cdots (n-k+1) \). According to our basic definitions, the probability of an event \( A \) equals \( \frac{|A|}{|S|} \). Thus, calculation of \( \Pr(A) \) according to definitions entails counting the number of objects in \( A \).

Happily, it turns out that there exist fast ways for solving some of the most frequently occurring problems. Three illustrations are provided below.

1. Let \( i_1 < i_2 < \cdots < i_m \), where \( 1 \leq m \leq k \). What is the probability of some specified color on each of the \( i_j \text{'th} \) draws?

Answer: If the total numbers of reds and blacks specified are \( x \) and \( y = m - x \), respectively, then the probability is

\[
\frac{r(r-1) \cdots (r-x+1)b(b-1) \cdots (b-y+1)}{n(n-1) \cdots (n-m+1)}.
\]

For example, if we specify red on second, blue on fifth, red on seventh, the numerator is \( r(r-1)b \).

To see why (1) is true, consider first the case when red balls are specified for draws \( 1, \ldots, x \) and black for draws \( x+1, \ldots, m \). If \( A \) denotes the set of points in \( S \) which satisfy the conditions, then it's easy to count that \( |A| \) is the numerator in (1) multiplied by the number of permutations of \( n-m \) things taken \( k-m \) at a time. After cancellations, \( |A|/|S| \) reduces to (1). Of course, the usual thought process is not to count, but to think sequentially: \( \Pr(A) \) equals

\[
\frac{r\ r\ r-1\ \cdots\ b\ y+1}{n\ n-1\ \cdots\ n-m+1}.
\]
For the general case of problem 1, (1) asserts that that the probability for any specified order is the same as that for any other specified order, as long as the number \( m \) of specifications and the total number \( n \) of \( r \)'s and \( b \)'s drawn does not change. Also, one can leave unspecified some draws in between draws which are specified. One way to justify this result is to let \( B \) denote the event corresponding to the prescribed order. Then, using the symmetries in \( S \), one can establish a bijective (fancy name for 1-1) correspondence between \( B \) and the event \( A \) introduced above.

2. With \( x \) as in 1, what is the probability that, in all of the \( k \) draws, exactly \( x \) of the balls chosen are red? Answer:

\[
\binom{r}{x} \binom{b}{k-x} \binom{n}{k}.
\]

(2)

Question 2 differs from question 1 mainly in that we don't care in question 2 what order the balls were drawn in, but only how many of each color turned up over all. One way to establish (2) is to take \( m = k \) in question 1, and to let \( C \) denote the subset of \( S \) consisting of all \( k \) tuples with \( x \) "red" entries and \( k-x \) "black" entries. Then \( |C| = \binom{k}{x} |A| \). (2) follows from multiplying (1) by \( \binom{k}{x} \), then simplifying.

A simpler way to obtain (2) is to imagine the balls being laid out in a row. Since order is not relevant in this problem, it seems reasonable to recast the sample space as being, instead of \( S \), the set \( S_1 \) of all (unordered) \( k \)-element subsets of \( \{1, \ldots, n\} \). Then questions about total numbers of red and black balls drawn are like problems about composition of \( k \)-person committees from a group with \( r \) boys and \( b \) girls, and lead to familiar answers like (2). Of course, one should understand why the \( S \) and \( S_1 \) models give the same answers for such questions.

3. Problems with conditional probabilities. Here's an example. If red was chosen on the first and second draws and black on the third, what's the probability that red will be chosen on the fourth draw? Clearly, the answer should be

\[
\frac{r-2}{n-3}.
\]

(3)
To answer this question using the definition of conditional probability, one must express the question in the form: find $Pr(D|E)$, where $D$ and $E$ are appropriate subsets of $S$. Then $Pr(D|E) = \frac{|DE|}{|E|}$. After simplification, one can verify that this second answer indeed agrees with (3).

What's the probability that red appeared on the 5'th draw, if we know that red appeared on draws 2 and 6, and black appeared on draw 10? The answer is still given by (3). This may be less intuitively obvious, but it can be proved from the definitions using symmetry properties of $S$. Some of our conditional "Urn problems" can be quickly dispatched using this line of thinking. Others, though, seem to require casting in the form: find $Pr(A|B)$, then computation of $|AB|$ and $|B|$. 
Urn problems. Suppose that an urn contains 5 green, 7 blue, and 8 white balls, and that
8 of them are drawn at random without replacement. Find the following probabilities.

1. The first ball is black.
2. The fifth ball is black.
3. The first and fifth balls are black.
4. The first and fifth balls are black, the second and third are white, and the eighth is
green.
5. At least one ball is black.
6. Exactly three balls are black.
7. Exactly 3 are black, 2 are white, and 3 are green.
8. Fewer than 3 are black.
9. Exactly 3 are black, given that exactly 2 are white.

10. The eighth is black, given that the first was white.
11. The eighth is black, given that the first and fifth were black, the second, fourth, sixth
and seventh white, and the third green.
12. The third is black, given that at least one one of the first two was black.
13. Suppose that three of the whites are pained pink. Find the probability that all four
of the colors appear on the first four draws.
14. With the situation of 13, find the probability that all four colors appear on the first
five draws.
15. With the situation of 13, find the probability that all four colors appear on the first
four draws, if we know that black did appear.
Urn answers.

1. $\frac{7}{20}$; 2. $\frac{7}{20}$; 3. $\frac{7}{20} \div \frac{8}{19}$; 4. $\frac{7}{20} \div \frac{8}{19} \div \frac{7}{18}$; 5. $\frac{13 \cdot 12 \cdot 11 \cdots 6}{20 \cdot 19 \cdot 18 \cdots 13}$; 6. $\frac{\binom{7}{3} \binom{13}{5}}{\binom{20}{8}}$; 7. $\frac{\binom{7}{3} \binom{5}{2}}{\binom{8}{3}}$; 8. $\frac{\binom{13}{5} \binom{7}{1}}{\binom{20}{6}}$; 9. $\frac{\binom{7}{3} \binom{5}{2}}{\binom{12}{6}}$; 10. $\frac{7}{20}$; 11. $\frac{5}{13}$; 12. $\frac{a}{b}$, where $a = \frac{1}{20 \cdot 19 \cdot 18} (\frac{7 \cdot 13 \cdot 6}{6 \cdot 7})$, $b = \frac{1}{20 \cdot 19} (\frac{7 \cdot 13}{13 \cdot 6})$; 13. $\frac{7 \cdot 5 \cdot 3 \cdot 5}{\binom{20}{4}}$; 14. $\frac{a}{\binom{20}{8}}$, where $a = \binom{5}{2} \cdot 7 \cdot 3 \cdot 5 \div 5 \cdot \binom{7}{2} \cdot 3 \cdot 5 \div 5 \cdot 7 \cdot \binom{5}{2} \cdot 5 \div 5 \cdot 7 \cdot 3 \cdot \binom{5}{2}$; 15. $\frac{7 \cdot 5 \cdot 3 \cdot 5 \cdot 4!}{(20 \cdot 19 \cdot 18 \cdot 17) - (13 \cdot 12 \cdot 11 \cdot 10)}$. 