MATHEMATICS 191, FALL 2004
MATHEMATICAL PROBABILITY
Outline #2 (Sets and Probability)

Last modified: September 17, 2004

References:

- PRP(Probability and Random Processes, by Grimmett and Stirzaker), Sections 1.1 to 1.3 and 1.7.
- EP(Elementary Probability, by Stirzaker), Chapters 0, 1, and 3 (up through section 3.4).
- 1000Ex(One Thousand Exercises in Probability, by Grimmett and Stirzaker), page 136.
- Web pages on poker and bridge from www.durangobill.com (attached)

1. (EP section 1.1, PRP section 1.2)Give a couple of "real-world" examples of "experiments," each with an associated sample space Ω. For each, specify an "individual outcome" in Ω to which you might assign a probability and an "event" (subset of Ω) that is not just an individual outcome. On what basis might you assign these probabilities?

2. (EP section 1.2) Derive formulas for the intersection and difference of sets A and B in terms only of union and complement. Illustrate the formulas using a Venn diagram.

3. (PRP, section 1.2) State the definition of a σ-field (EP uses the synonym "event space") and explain why the closure requirements in the definition are sufficient to prove closure under difference and intersection. Give a couple of trivial examples of σ-fields. Then give a simple example to show that the union of two σ-fields is not necessarily a σ-field (1000Ex, solution 3c on p. 142).

4. (PRP, section 1.3, EP section 1.3) Define "probability measure"(EP uses the synonyms "probability distribution" and "probability function"). Define "probability space", and illustrate the definition by describing the probability space associated with rolling a single die.

5. (PRP, section 1.3, EP section 1.3)Prove that for a probability measure $\mathbb{P}$,
   - $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$ if $A \subseteq B$.
   - $\mathbb{P}(A) \leq \mathbb{P}(B)$ if $A \subseteq B$.

6. (PRP, section 1.3, EP section 1.4)Prove that for a probability measure $\mathbb{P}$,
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$
   and illustrate this theorem with a diagram (EP figure 1.2) where ω is the set of points in a square region of the blackboard.

7. By induction on the preceding result, prove the "generalized inclusion-exclusion formula" (Lemma 4d on page 6 of PRP, formula 8 on p.35 of EP, proved on p. 136 of 1000Ex, problem 4.)

8. Prove that for a probability measure $\mathbb{P}$, $\mathbb{P}(A \cup B) <= \mathbb{P}(A) + \mathbb{P}(B)$. Then by induction on this result, prove "Boole's inequality"
$$\mathbb{P}(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} \mathbb{P}(A_i).$$
   (proved on p. 143 of 1000Ex, problem 11.)

9. (PRP, section 1.3, EP section 1.5) State and prove Lemma 5 on page 7 of PRP, and explain why this lemma would be inappropriate if the definition of a probability measure involved only finite unions and inadequate if the definition allowed uncountable unions.

10. (PRP, problem 16 on p.22, solved on p. 144 of 1000Ex, and PRP, p. 320, theorem 10a) Consider an infinite sequence of events $A_1, A_2, \ldots$. For example, A_n might be the event "the total amount spent by the U.S. government for years 2000 through 2000+n exceeds tax revenues by more than 5 percent."

    Let B be the event that infinitely many of the events A_n occur. Argue that
$$B = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$
    and use Boole's inequality to show that
$$P(B) = 0$$
    if
$$\sum_{n=1}^{\infty} P(A_n)$$
    converges. This important result is called the "first Borel-Cantelli lemma."

11. (PRP, section 1.7, EP, section 1.8) Using the principles listed at the start of section 1.7 in PRP, explain how to answer the following questions:

    (a) What is the probability of getting exactly two heads when a fair coin for which $\mathbb{P}(h) = \frac{1}{2}$ is tossed three times?

    (b) What is the probability of getting a total of either 7 or 11 when tossing two "unloaded" dice?

    In the following items, use these principles of counting, taken for granted in PRP but nicely explained in EP, sections 3.1 through 3.4.

    - Principle 1: Multiply for sequential counting. If we are forming a set of n-tuples where there are k_1 choices for the first element x_1, k_2 for x_2, and so on, the number of n-tuples $x_1 x_2 ... x_n$ in the set is $k_1 k_2 ... k_n$
    - Principle 2: Divide to correct systematic overcounting. If M different n-tuples all correspond to the same element in our sample space, count the n-tuples and then divide by the number of times M that each was overcounted.
    - Principle 3: Divide and Conquer. To count the number of elements in a union of disjoint subsets, count each subset and sum the results.
    - Principle 4: Subtract off special cases. To count the number of elements in a difference $A \setminus B$ where $B \subset A$, count each set and take the difference.

    Classic example of principles 1 and 2: counting the number of k-element subsets of an n-element set.

    First count the k-tuples whose elements are all different:

    $$n(n-1)(n-2)...(n-k+1).$$

    Then divide by $k!$ to correct for overcounting. The result is the familiar "combinations" formula

    $$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

12. Calculate the probability, when a pair of cards are dealt from a well-shuffled single deck of cards, that at least one of them is a spade. Let A be the event that the first card dealt is a spade (and the second can be anything). Let B be the event that the second card dealt is a spade (and the first can be anything). Show that you get the same answer, $\frac{15}{34}$, by each of the following approaches. State in words, and illustrate with a Venn diagram, the reasoning behind each of the four approaches.

(a) $P(A \cup B) = 1 - P(A^c \cap B^c)$
(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(c) $P(A \cup B) = P(A) + P(A^c \cap B)$
(d) $P(A \cup B) = P(A \cap B^c) + P(A^c \cap B) + P(A \cap B)$

For the following topics, the easiest way to get numerical answers is to use Mathematica. Any Harvard undergraduate can download a copy for free. For details, go to http://www.fas.harvard.edu/cgi-bin/software/download.pl and follow instructions.

13. Count the number of ways to get each of the following types of 5-card poker hands, using a deck of 52 cards with 4 cards of each of 13 ranks.

    - 4 of a kind(four cards of one rank, the fifth of a different rank)
    - a full house(three cards of one rank, two of another)
    - 3 of a kind(three cards of one rank, two others of different ranks)

    Solutions:

    - How many distinct 5-card poker hands can be dealt from a 52-card deck?
      Answer: Select 5 cards sequentially: the number of ways to do this is $52 \times 51 \times 50 \times 49 \times 48$
      But this generates each distinct hand $5! = 120$ times.
      So the number of distinct hands is $\binom{52}{5} = \frac{52!}{47!5!} = 2,548,960$
    - How many distinct ways are there to get 4 of a kind, and what is the probability?
      13 choices for the rank of the 4, 48 choices for the other card.
      $$\mathbb{P} = \frac{13 \cdot 48}{2,548,960} = 0.000240096$$
    - A full house (3 of one rank, 2 of another)
      13 choices for the first rank (with 3 cards), 4 for the suit that is missing.
      12 choices for the second rank (with 2 cards), $\binom{4}{2} = 6$ for the pair of suits that have the second rank.
      $$\mathbb{P} = \frac{13 \cdot 4 \cdot 12 \cdot 6}{2,548,960} = 0.0014405762$$
    - Three of a kind (3 of one rank, the other two do not match)
      13 choices for the first rank (with 3 cards), 4 for the suit that is missing.
      $\binom{12}{2} = 66$ choices for the pair of ranks of the other two cards; 16

choices for the suit of each. So the number is $13 \cdot 4 \cdot 66 \cdot 16 = 54,912$ and

$$\mathbb{P} = \frac{54,912}{2,548,960} = 0.0211284514$$

14. Bridge problems

    Count the number of bridge hands with 6 spades, 4 hearts, 2 diamonds, and 1 club.

- Count the number of bridge hands with 6-4-2-1 suit distribution (6 cards in the longest suit, 4 in the second-longest, 2 in the third-longest)
- Count the number of bridge hands with 4-4-3-2 or 4-3-3-3 suit distribution, and show that the former has a higher probability.

Solution:

- How many ways are there to select 6 cards from the 13 spades?

$$binom136 = \frac{13!}{7!\,6!}$$

- How many distinct hands have 6 spades, 4 hearts, 2 diamonds, 1 club?
  Apply the same analysis to each suit in turn:

$$\binom{13}{6}\binom{13}{4}\binom{13}{2}\binom{13}{1} = \frac{13!}{7!\,6!}\frac{13!}{9!\,4!}\frac{13!}{11!\,2!}\frac{13!}{12!\,1!}$$

- How many distinct hands have a 6-4-2-1 distribution?
  Multiply the preceding number by the number of ways to choose the 6-card suit, the 4-card suit, etc, which is 4!=24.
- How many distinct hands have a 4-4-3-2 distribution?
  There are 4 ways to select the 2-card suit followed by 3 ways to select the 3-card suit, so we will multiply by 12.

$$\mathbb{P}_{4432} = 12\binom{13}{4}\binom{13}{4}\binom{13}{3}\binom{13}{2} = 12\frac{13!}{9!\,4!}\frac{13!}{9!\,4!}\frac{13!}{10!\,3!}\frac{13!}{11!\,2!}$$

- How many distinct hands have a 4-3-3-3 distribution?
  Now there are 4 ways to select the 3-card suit, so we will multiply only by 4.

$$\mathbb{P}_{4333} = 4\binom{13}{4}\binom{13}{3}\binom{13}{3}\binom{13}{3} = 4\frac{13!}{9!\,4!}\frac{13!}{10!\,3!}\frac{13!}{10!\,3!}\frac{13!}{10!\,3!}$$

The ratio is

$$\frac{\mathbb{P}_{4432}}{\mathbb{P}_{4333}} = \frac{12\,10!\,3!\,10!\,3!}{4\,9!\,4!\,11!\,2!} = \frac{45}{22}$$

15. Dividing up a set with two type of objects

Suppose we have a set S with $2n$ elements, m of one type and $2n-m$ of the other. As a concrete example, suppose that a mother has $2n=6$ Dunkin' Munchkins, of which $m=2$ are chocolate and the remaining $2n-m=4$ are plain. She chooses $n=3$ of them at random and puts them in a bag for one child. The remaining three go into a bag for the other child.

In general we divide up S at random into set S_1 and set S_2, each with n elements. The number of different ways of doing this is equal to the number of ways of selecting the n elements of S_1, namely

$$N = \binom{2n}{n} = \frac{(2n)!}{n!n!}$$

We are interested in the probability that k of the "special objects" (chocolate Munchkins) end up in set S_1. Assume that $k \leq m$, $k \leq n$, and $k \geq m-n$. The number of ways of selecting k special objects and $n-k$ non-special objects for set 1 is

$$M = \binom{m}{k}\binom{2n-m}{n-k} = \frac{m!}{k!(m-k)!}\frac{(2n-m)!}{(n-k)!(n+k-m)!}$$

If we assume that each way of selecting the elements of S_1 is equally likely, the ratio $\frac{M}{N}$ gives the probability that set S_1 contains precisely m special objects.

As a practical matter, it is much easier to repeat this reasoning in special cases than to memorize these formulas!