MATH 210, PROBLEM SET 6

DUE IN LECTURE ON WEDNESDAY, OCT. 27

1. σ-algebras and Borel sets

Let S be a set. Recall that a σ-algebra B for S is a set of subsets of S such that:

i. S is in B.
ii. If E is in B then the complement $E^c = S - E = \{s \in S : s \notin E\}$ is in B.
iii. If $\{E_i\}_{i=1}^{\infty}$ is in B then $\cup_{i=1}^{\infty} E_i$ is in B.

1. With the assumptions of (iii) above, show that
$$\cap_{i=1}^{\infty} E_i = (\cup_{i=1}^{\infty} E_i^c)^c$$
Explain why this implies that $\cap_{i=1}^{\infty} E_i$ must be in B.

2. When S is the closed interval $[0, 1]$ on the real line, we defined the σ-algebra $B([0, 1])$ of Borel sets to be the intersection of all σ-algebras of subsets of S which contain every closed interval $[a, b]$ for which $0 \le a \le b \le 1$. Explain why the intervals $[0, a)$, $(b, 1]$ and (a, b) must also be in $B([0, 1])$.

3. A subset E of $S = [0, 1]$ is open if for each $x \in E$, there is an open interval $(x-\epsilon, x+\epsilon)$ on the real line for some $\epsilon > 0$ such that $(x - \epsilon, x + \epsilon) \cap [0, 1]$ is contained in E. Show that if $0 < a < b < 1$ then $[a, b]$ is not open. Is $[0, a)$ open in $S = [0, 1]$? Note: This shows that there are Borel sets which are not open!

2. Multinomial coefficients

The multinomial Theorem says that if $t \ge 1$ and $n = n_1 + \cdots + n_t$ for some non-negative integers n_i, the number c of ways of picking disjoint unordered subsets T_i of $\{1, \ldots, n\}$ such that $\#T_i = n_i$ is

$$c = \frac{n!}{n_1! \cdots n_t!}. \tag{2.1}$$

Note that $\{1, \ldots, n\}$ must be the disjoint union of the T_i since $n = n_1 + \cdots + n_t$. This exercise is about proving the multinomial Theorem by induction on n.

4. Show that if $n = 1$ then exactly one of the T_i is not empty, and then $T_i = \{1\}$. Explain how this implies the theorem when $n = 1$. (Recall that $0! = 1$ by definition.)

5. Suppose now that $n > 1$ and that the theorem is true for all integers smaller than n. The integer 1 in $\{1, \ldots, n\}$ must lie in exactly one of the T_i for which $n_i = \#T_i > 0$. Having picked such a T_i, the set $\{2, \ldots, n\}$ is the disjoint union of $T_i - \{1\}$ together with the T_j associated to $j \ne i$. Use induction to show that the number of ways of writing $\{2, \ldots, n\}$ as the disjoint union of $T_i - \{1\}$ with the other T_j is
$$\frac{(n-1)!}{n_1! n_2! \cdots (n_i - 1)! (n_{i+1})! \cdots n_t!}$$

Explain why this leads to the formula

$$
\begin{aligned}
c &= \sum_{1\le i\le t, n_i>0} \frac{(n-1)!}{n_1!n_2!\cdots(n_i-1)!(n_{i+1})!\cdots n_t!} \\
&= \sum_{1\le i\le t, n_i>0} \left(\frac{n_i}{n}\right)\cdot\left(\frac{n!}{n_1!n_2!\cdots n_i!(n_{i+1})!\cdots n_t!}\right) \qquad (2.2)
\end{aligned}
$$

6. Use $n = n_1 + \cdots + n_t$ to complete the proof of (2.1) using problem # 5.

3. Conditional probabilities

7. The video for October 22 describes a legal case in California. After having a look at this, calculate the conditional probability that there will be at least three couples matching the description of the perpetrators, given that there is a least one couple that matches this description. As in the video, you can use the parameters $p = \frac{1}{12{,}000{,}000}$ for the probability that a randomly chosen couple matches the description given by eyewitnesses, and you can use $m = 8,000,000$ for the number of couples in Los Angeles. First find an exact expression for the above conditional probabilitiy, and then evaluate this using Maple or Wolfram Alpha.

   Warning: Maple may stall or crash. I would suggest first trying Wolfram Alpha. To get around length limitations on the input, use scientific notation, e.g. write $1/(12 * 10^6)$ for p.

8. The analysis in the video for October 22 applies directly to the use of DNA evidence in criminal trials. For example, during the trial of O. J. Simpson, the prosecution argued that the probability that a randomly chosen person would have the same DNA results as O. J. were

   $$p = \frac{1}{170,000,000}.$$

   Suppose the population of Los Angeles during the trial was

   $$m = 10,000,000.$$

   Given that at least one person (O.J.) has this DNA pattern, what is the conditional probability that more than one person has the DNA pattern? The result is surprising. There have been some statistical analyses of the O. J. trial evidence, but I have not seen this particular one. (The DNA evidence was thrown out due to misconduct by the police department.) Quite a few of the $10,000,000$ people living in L. A. could not have conceivably committed the double homicide of which O. J. was accused. For instance, most people would not have been physically strong enough to have committed the crime. If one drops m to $m = 1,000,000$ how does this change the outcome of your calculations?

Comment: Problems 7 and 8 are a wonderful example of Poisson distributions, which we will discuss in detail later. When p is very small and m is very large, the relevant parameter will turn out to be $\delta = mp$, which is the expected number of people in the population who will have the special characteristic in question. We will show later that the conditional probability that there will be more than one person with the special characteristics, given that at least one person has these characteristics, is estimated closely by

$$1 - \frac{\delta}{e^{\delta} - 1}.$$

You might try comparing this to the exact answers you find in problem # 8 - this is certainly optional, though!

Extra Credit (due anytime during the semester)

1. With the notation of problem # 3 above, show that every open subset of $[0, 1]$ is a Borel set, i.e. is an element of $B([0, 1])$. Hint: Use that every open interval contains a rational number.