10-751 Assignment 1

Due: Tuesday September 17

Problem 1. If the probability that student A will fail a certain computer science exam is 0.5, the probability that student B will fail the exam is 0.2, and the probability that both A and B will fail the exam is 0.1, what is the probability that at least one of these two students will fail the exam? (DeGroot 1.5.1.)

Solution. Let F_A be the event that A fails and F_B the event that B fails. Then $\Pr(F_A \cup F_B) = 0.5 + 0.2 - 0.1 = 0.6$.

Problem 2. Prove that for every two events A and B, the expression

$$\Pr(A) + \Pr(B) - 2\Pr(AB)$$

is the probability that exactly one of the two events will occur. (DeGroot 1.5.9.)

Solution.

$$\begin{aligned} \Pr(AB^c) + \Pr(A^c B) &= (\Pr(A) - \Pr(AB)) + (\Pr(B) - \Pr(AB)) \\ &= \Pr(A) + \Pr(B) - 2\Pr(AB) \end{aligned}$$

Problem 3. Consider an experiment in which a fair coin is tossed once and a balanced die is rolled once.

1. Describe the sample space for this experiment.
2. What is a probability that a head will be obtained on the coin and an odd number will be obtained on the die? (DeGroot 1.6.8.)

Solution. 1. The sample space is $\{(H,1),(T,1),(H,2),(T,2),\ldots,(H,6),(T,6)\}$. 2. $\frac{1}{4}$

Problem 4. Suppose you want to find someone with the same birthday as yours. What is the smallest number of people whose birthdays you need to ask about in order to have at least a 50% chance of finding someone with the same birthday? (This summer, I met identical twins having the same birthday as me!)

Solution. One reasonable guess might be $\lceil \frac{365}{2} \rceil = 183$. However, this is an underestimate. When asking n people, the probability that none of them have your birthday is (roughly) $\left(\frac{364}{365}\right)^n$, so the probability that at least one of them has your birthday is $1 - \left(\frac{364}{365}\right)^n$. For this to be at least $\frac{1}{2}$, we take logarithms and find that n should be at least 253.

Problem 5. This problem will involve some computing using Matlab. (To get acquainted with the basics of Matlab, please read David Touretzky's tutorial, or one of the other tutorials on the Web.)

Write a Matlab program to generate n pairs of random integers, between 2 and 1000. Then, find the percentage that have greatest common divisor (gcd) $1, 2, 3, \ldots$, and make a histogram of the results for different values of n. Can you come up with any analytical explanation for the results you see? The Matlab functions `gcd`, `histc`, and `bar` might be useful for this problem.

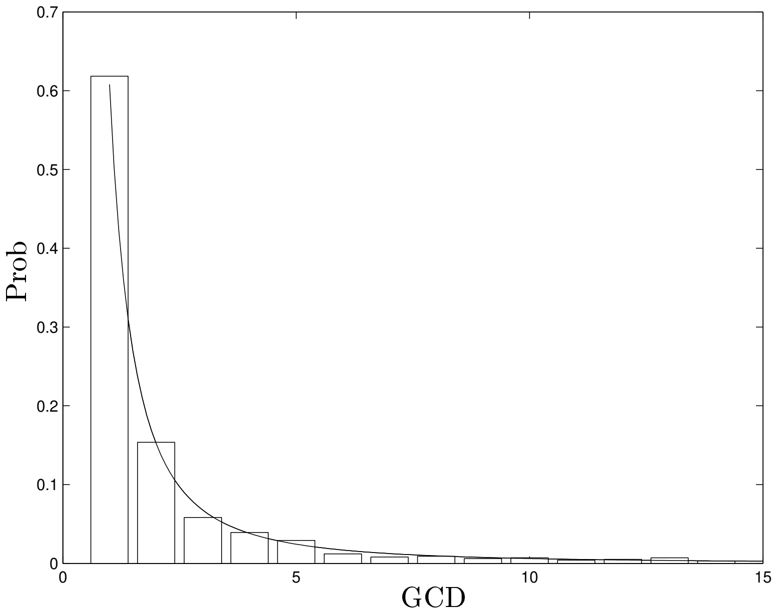
Pass in plots of your results, your Matlab program, and your analysis of the probability that explains the curve.

Solution. Here is a rough analysis to suggest the qualitative behavior seen in the data. Let g_1 denote the probability that the gcd of two random integers is 1. Intuitively, the probability that an integer m divides a random integer $n \gg m$ is about $\frac{1}{m}$. Thus, since $\gcd(n_1, n_2) = m$ if m divides n_1, m divides n_2, and $\gcd(n_1/m, n_2/m) = 1$, the probability that the gcd is m is $g_1 \frac{1}{m^2}$. This allows us to come up with an estimate for g_1. Using $g_1 \sum_m \frac{1}{m^2} = 1$ we get that $g_1 = \frac{6}{\pi^2} \approx 0.6079$.

Making all of this precise is a bit involved. Note that a better estimate for g_1 comes from restricting to the primes:

$$g_1 = \prod_{\text{prime } p} \left(1 - \frac{1}{p^2}\right)$$

Here is a histogram from generating random integers and computing their gcd:



And here is the Matlab code that generated the data:

```
% Generate two pairs of vectors, each
% 1000 random numbers between 1 & 1000
n = floor(1000 * rand(1000,2)+1);

% Compute the gcd of the pairs
g = gcd(n(:,1), n(:,2));

% Form a histogram of the result
c = histc(g, 1:100);

% Finally, make a bar plot of the empirical dist.
colormap([1 1 1])
bar(1:14, c(1:14) ./ sum(c))
xlabel('gcd'); ylabel('prob');
```