MATH 318 Assignment 3: Due Friday, Jan 26 at start of class

I. Problems to be handed in:

1. A fair dice is rolled four times. Let Y denote the number of distinct rolls. Find the probability mass function of Y and its expectation.
   Hint: Y can take values $1, 2, 3$ and 4.

2. Smith (the brown eyed guy from Assignment 2) collects coupons. There are m different types of coupons. He receives a new coupon every day, and that coupon is equally likely to be any one of the m types. He is interested in how long it would take on average to have a complete set of all m coupons.
   (a) Let X be the number of coupons he collects until obtaining a complete set. Explain why X can be written as a sum of m independent Geometric random variables (say what their parameters are).
   (b) Compute the expected value of X. (Use the fact that the expectation of a sum of random variables is the sum of the expectations.)

3. This problem gives a mathematical explanation of the similarity between the geometric and exponential random variables observed in Assignment 2, Problem 5. Let Y be a geometric random variable with parameter p, so that Y represents the trial number of the first success in a sequence of independent Bernoulli trials. Suppose the trials occur at times $\delta, 2\delta, 3\delta, \ldots$, and that δ and p are both very small with $\delta/p = 1/\lambda$, with $\lambda > 0$ fixed. At time t, about t/δ trials have taken place.
   (a) Compute $P(Y > m)$, which represents the probability that no success has been observed by time $t = m\delta$.
   (b) Show that the probability that no success has been observed by time t converges to $e^{-\lambda t}$, in the limit $\delta \to 0$ and $p \to 0$ with $\delta/p = 1/\lambda$ fixed. Conclude that the time of the first success is approximately an exponential random variable with parameter λ.

4. A binary message either—0 or 1—must be transmitted by wire from location A to location B. However, data sent over the wire are subject to channel noise disturbance, so to reduce the possibility of error, the value 2 is sent if the message is 1 and the value -1 is sent if the message is 0. If x is the value sent at A ($x = -1$ or $x = 2$), then the value received at B is $R = x + N$, where N represents the noise. Assume that N is a normal random variable with mean $\mu = 0$ and variance $\sigma^2 = 0.25$. Assume that a message sent is equally likely to be 0 or 1. When the message is received at B the receiver decodes it according to the following rule:
   If $R \geq 0.5$, then 1 is concluded.
   If $R < 0.5$, then 0 is concluded.
   The message concluded at B is 0. What is the probability that the message was incorrectly transmitted? (Recall that if $X \sim N(\mu, \sigma^2)$ then the random variable $Z = \frac{X-\mu}{\sigma}$ is standard normal, $Z \sim N(0, 1)$, and use symmetry and the Table in Ross.)

5. An airline books passengers for a flight on an airplane with 500 seats. From experience, the airline knows that each passenger has probability $p = 0.06$ of missing the flight. As such, the airline takes a risk and books 525 passengers for the flight.
   Using Octave:
   (a) Compute the true probability that the flight in question is overbooked. One helpful Octave command here is `binocdf`: calling `binocdf(x,N,p)` returns the CDF of a binomial random variable with parameters N and p evaluated at x. That is, if B is $\text{Bin}(N, p)$, then `binocdf(x,N,p)` returns $P(B \leq x)$.
   (b) Use the Poisson approximation to compute an approximation to this true probability. Here, you may find the `poisscdf` command helpful: `poisscdf(x,lambda)` returns the CDF of a Poisson random variable with parameter λ evaluated at x.
   (c) Simulate the number of no-shows for a booked flight 40000 times. Compute the proportion of times which the airline overbooks the flight. Here, you may want to use the Octave command `binornd`: calling `binornd(N,p,R,C)` generates an $R \times C$ matrix of random numbers distributed binomially with parameters N and p.
   (d) After you have simulated the no-shows 40000 times, define

$$O_n = \text{number of overbookings in the first } n \text{ simulated bookings;}$$

   then O_n/n is the running proportion of overbookings in the first n simulated bookings. Plot this running proportion against n. What happens to it as n gets large?
   Note: When calculating O_n, do not simulate n new bookings for every n. Simulate 40000 bookings first, and then for every n, calculate O_n based on the first n of these bookings.
   Print out and submit: your code; the answers to parts (a), (b), and (c); and your plot for part (d).

II. Recommended problems: These provide additional practice but are not to be handed in.

A. Suppose that X is a normal random variable with parameters μ and σ^2. (a) Compute the probability that $|X - \mu|$ exceeds $k\sigma$ for $k = 1, 2, 3$. (b) For what value of k is $P(|X - \mu| \leq k\sigma)$ equal to 0.95? $(0.3174, 0.0456, 0.0026, k = 1.96)$

B. Three sections of a class contain 20, 30 and 50 students respectively. (a) If a section is chosen at random, what is its expected size? (b) If a student is chosen at random, what is his/her expected section size?
   $(33\frac{1}{3}, 38)$.

C. Chapter 2: #16*, 23*, 27*, 31, 38*, 39 (31/6), 40 (93/16), 49*.
   Chapter 5: #3.

Quote: *The chance at any instant whether an atom disintegrates or not in any particular second is fixed. It has nothing to do with any external or internal consideration we know of, and in particular is not increased by the fact that the atom has already survived any period of past time .... All that can be said is that the immediate cause of atomic disintegration appears to be due to chance.*
Frederick Soddy, writing of his work on radioactivity (and appreciating the memoryless property of the exponential random variable), as quoted in *Uncle Tungsten* by Oliver Sacks.