Probability and stats mop-up unit

- Jointly distributed random variables (spring-break eve)
- Understanding covariance and correlation (last week Monday)
- Convergence of random variables (last week Wednesday)
- The weak law of large numbers (last week Friday)
- The central limit theorem (Monday)
- Review for test (**Today**)
- Test 2 (Friday)
- Statistical inference and confidence intervals (next week Monday)

Today:

- What we have covered so far
- What to expect on the test
- Solutions to previous exercises

(Discrete) random variables. Don't forget the basic definition of a random variable, even though that was from before Test 1. Know the definitions of expected value and variance and be able to use the definition of expected value to prove simple things like Theorems 1-4 on the Feb 18 slides/handout, and be able to find the expected value from the description of a random variable, as in the exercises on the Feb 18 handout.

Discrete distributions. Be able to identify the Bernoulli, binomial, and geometric distributions. Know what they are used for, be able to demonstrate their expected values, and use them to compute probabilities.

Continuous random variables. Know the elements of continuous random variables, such as PDFs, how to calculate the probability of an interval, and the definition of expected value for continuous random variables. Be able to identify the uniform, Gaussian (normal), exponential, and Pareto distributions, both by their formula and by the graphs of their PDFs. Be able to apply the uniform distribution to data to compute a probability of an interval.

Jointly distributed random variables. Know the definitions and formulas of joint CDF and PDF, marginal PDF, conditional PDF, and and independence. Know the definition and formula for covariance and correlation coefficient. Know the results about covariance from Mar 16.

Sequences of random variables. Know Markov's and Chebyshev's inequalities. Know what sequences of random variables and sample means are. Be able to identify the claims of the Weak Law of Large Numbers and the Central Limit Theorem.

Exercises. "Expected value and variance" (Feb 18)

1. Let X be a random variable with PMF given in the table below. Find $E[X]$, $E\left[\frac{1}{X}\right]$, $E\left[X^2 - 1\right]$, and $Var(X)$.

x	1	2	3	4
$p(x)$	.4	.3	.2	.1

   (Wackerly et al, *Mathematical Statistics with Applications*, 2008)

2. A sample of 3 items is selected at random from a box containing 20 items of which 4 are defective. Find the expected number of defective items in the sample. (Ross, *A First Course in Probability*, 1997)

3. Prove this theorem:

Theorem

For any $b \in \mathbb{R}$,

$$E[X + b] = \sum_{x \in \chi} (x + b)\ p(x) = E[X] + b$$

Exercises. "Discrete distributions" (Feb 20 & 23)

1. On a multiple-choice test with 3 possible answer for each of the 5 questions, what is the probability that a student would get 4 or more correct answers by just guessing?
2. A communications channel transmits the digits 0 and 1. Due to static, the digit transmitted is incorrectly received with probability .2. Suppose that we want to transmit an important message consisting of 1 bit. To reduce the chance of error, we transmit 00000 instead of 0 and 11111 instead of 1. The reciever decodes this based on the majority of bits—that is, if most recieved bits are 0, then the message is interpreted as 0, and similarly for 1. What is the probability that the message will be wrong when decoded? Assume that for each bit sent, the probabilities of error are independent.
3. A newsie (i.e., an old-time street vendor of newspapers) purchases copies for 10 cents and sells them at 15 cents. The daily demand for newspapers in his territory is binomially distributed with $n = 10$ and $p = \frac{1}{3}$. How many papers should the newsie purchase in order to maximize profit? (Hint: Notice that there are two random variables here—the number of papers sold and the profit for number of papers bought.)

(Ross, *A First Course in Probability*, 1997)

Exercise. "Uniform distribution" (Feb 27)

1. Trains headed for destination A arrive at the train station at 15-minute intervals starting at 7 AM, whereas trains headed for destination B arrive at 15-minute intervals starting at 7:05 AM.
   a. If a certain passenger arrives at the station at a time uniformly distributed between 7 and 8 AM and then gets on the first train that arrives, what proportion of time does he or she go to destination A?
   b. What if the passenger arrives at a time uniformly distributed between 7:10 and 8:10 AM?

Exercises. "Gaussian and Exponential" (Mar 2)

1. A company that manufactures and bottles apple juice uses a machine that automatically fills 16-ounce bottles. There is some variation, however, in the amounts of liquid dispensed into the bottles that are filled. The amount dispensed has been observed to be approximately normally distributed with mean 16 ounces and standard deviation 1 ounce. Determine the proportion of bottles that have more than 17 ounces dispensed into them.
2. The weekly amount of money spent on maintenance and repairs by a company was observed, over a long period of time, to be approximately normally distributed with mean \$400 and standard deviation \$20. If \$450 is budgeted for next week, what is the probability that the actual costs will exceed the budgeted amount?

(Wackerly et al, *Mathematical Statistics with Applications*, 2008)