University of Illinois **Spring 2010**

ECE 313: Problem Set 4
Counting Random Variables, Maximum-Likelihood Estimation

Due: Wednesday February 17 at 4 p.m.
Reading: *Ross* Chapter 4.
Noncredit Exercises: Chap. 4: Problems 32, 38-39, 40-43, 47-52;
Theoretical Exercises 10, 11, 13-20, Self-test problems 11-14

1. **[The Binomial Random Variable I]**
Consider a football game in where the offense has 5 offensive linemen lined up at the line of scrimmage. An offensive lineman will move before the ball is snapped - resulting in a false start - independently of any other lineman with probability 10^{-3}.

   (a) Specify the probability of a false start, $p_{\mathbf{fs}}$ on any offensive play

   (b) Suppose a typical football game has 100 offensive snaps in a game. Specify the pmf for the random variable $\mathbf{X}$ denoting the number of false starts in the game.

   (c) A false start results in loss of 5 yards on any play. What is the expected amount of total yards in penalty incurred during a game?

2. **[Geometric Random Variables]**
During a bad economy, a graduating ECE student goes to career fair booths in the technology sector (e.g. Google, Apple, Qualcomm, Texas Instruments, Motorola, etc) - and his/her likelihood of receiving an interview request after a career fair booth visit depends on how well he/she did in ECE 313. Specifically, an A in 313 results in a probability $p_{\mathbf{A}} = 0.95$ of an interview, whereas a C in 313 results in a probability of $p_{\mathbf{B}} = 0.15$ of an interview.

   (a) Give the pmf for the random variable $\mathbf{Y}$ that denotes the number of career fair booth visits a student must make before his/her first interview invitation.

   (b) On average, how many booth visits must an A student make before getting an interview, as compared to a C student?

   (c) Suppose that during a typical career fair, there are a total of 5 booth visits that can be made in the technology sector. Find the minimum value of p for which a student can expect to get an interview in his/her senior year. What does this mean, on average, about the C student in 313?

   (d) Find the probability that an A student in 313 will not get an interview during a typical career fair. Similarly, find the probability that a C student in 313 will get an interview during a typical career fair

   For full credit, correct numerical answers must be given. Feel free to use a computer (e.g., MATLAB, Mathematica) for this problem.

3. **[The Binomial Random Variable II]**
Suppose that eight bits are used to encode a standard ASCII character; the eighth bit is a checksum bit set to the XOR of the seven other bits. Suppose that these bits are transmitted on an optical fiber, and that the probability of an error in any bit transmission is $p = 0.001$. The receiver detects an error whenever the XOR of the bits of a received 8-bit word is non-zero.

(a) Specify the pmf for the random variable $\mathbf{X}$ that denotes the number of error bits in an eight-bit word.

(b) Find p_{ue}, the probability that a received eight-bit word will contain undetected errors.

(c) Suppose n eight-bit words are transmitted. Determine the pmf for the random variable $\mathbf{Y}$ that denotes the number received words that contain undetected errors.

(d) Compute $E[\mathbf{Y}]$ for $n = 1, 10, 100, 1000, 10000$.

For full credit, correct numerical answers must be given. Feel free to use a computer (e.g., MATLAB, Mathematica) for this problem.

4. **[Poisson Random Variables]**
Consider for another time a communication system that transmits bits with independent probability of error $p = 0.001$, with each group of eight bits corresponding to a seven-bit ASCII code with checksum. Suppose that the number of undetected errors in a transmission of n eight-bit words is modeled as a Poisson random variable $\mathbf{X}$, with parameter $\lambda = np_{\text{ue}}$.

(a) Compute $E[\mathbf{X}]$ for $n = 1, 10, 100, 1000, 10000$.

(b) Compute the probability that a transmission will contain at least one undetected error using both the binomial and the Poisson pmf for $n = 1, 10, 100, 1000, 10000$.

For full credit, correct numerical answers must be given. Feel free to use a computer (e.g., MATLAB, Mathematica) for this problem.

5. **[The Binomial Random Variable III]**
Bo Schembechler is considered to Michigan's (if not all of college football's) greatest football coach ever. But things started out shaky: during his rocky first season as Head Coach in 1969, his 140 player roster dwindled to 75 due to his intense workouts. To encourage perseverance and hard work, Bo wrote a now-legendary motivational phrase on a sign in the locker room, stating "Those Who Stay Will Be CHAMPIONS". (It is indeed legendary: while Bo was coach, every single Michigan player that stayed for all four years left Michigan with at least one Big Ten Championship). Since Bo had 234 wins and 65 losses, let us assume that on any given game, the probability of a Michigan win, independently of other games, is given by $p_{\mathbf{win}} = 1 - 65/234 = 0.72$. Suppose that Michigan played in 9 Big Ten games each season, and they won the Big Ten Championship (and thus went to the Rose Bowl) if they won at least 6 of those games.

(a) What is the expected number of Big Ten Michigan wins in a season under Bo?

(b) What is the probability that under Bo, Michigan went to the Rose Bowl in any season?

(c) Suppose the Michigan athletic department loses $\$100,000$ for each Big Ten game they lose, and they profit $\$50,000$ for each Big Ten game they win. Also, if they go to the Rose Bowl, they profit another $\$1,000,000$ from online Bowl merchandise sales from alumni all across the world. What is the expected amount of money the Michigan Athletic department makes (or loses) each year under Bo?

(d) Suppose a given player is tough enough to endure Bo's workouts and stays at Michigan for all four years. What is the probability that the phrase "Those Who Stay Will Be CHAMPIONS" rings true?

For full credit, correct numerical answers must be given. Feel free to use a computer (e.g., MATLAB, Mathematica) for this problem.

6. **[Maximum-Likelihood Estimation and Confidence Intervals]**
An urn contains 10 red balls and an unknown number x of blue balls. The experiment consists of drawing one ball at random from the urn and noting its color. Consider 100 independent trials of this experiment. Thus, the ball drawn is replaced, and the urn shaken well before the next ball is drawn. It is observed that 25 of the drawings resulted in a red ball and 75 in a blue ball.

   (a) What is the maximum likelihood estimate $\hat{x}$ of the number of blue balls x in the urn?

   (b) Let $\hat{p} = \dfrac{10}{10+\hat{x}}$. With what level of confidence can we say that $p = \dfrac{10}{10+x}$ lies in the interval $[\hat{p} - 0.1, \hat{p} + 0.1]$?

   (c) Find a confidence interval with confidence level 0.96 for p.

7. **[Maximum-Likelihood Estimation]**
There are N multiple-choice questions (with 5 possible answers each) on a certain exam. A student knows the answers to K questions and answers them correctly. On the remaining $N - K$ questions, the student guesses randomly among the 5 choices. The examiner knows N, and can observe the values of $\mathcal{C}$, the number of *c*orrect answers, and $\mathbb{W} = N - \mathcal{C}$, the number of *w*rong answers on the answer sheet. Note that $\mathcal{C}$ can have values K, $K+1$, $\ldots$, N. What the examiner is really interested in, though, is *estimating* the value of K.

   (a) Explain why it is reasonable to model $\mathbb{W}$ as a binomial random variable with parameters $(N - K, 0.8)$. What assumptions are you making?

   (b) Suppose that n answers are incorrect, that is, $\mathbb{W} = n$ and $\mathcal{C} = N - n$. What is the *likelihood* of this observation? Hint: your answer will depend on N, n and the unknown parameter K that the examiner is interested in estimating.

   (c) Having observed that $\mathbb{W} = n$, the examiner is sure that K cannot exceed $N - n$, i.e., K can have value $0, 1, 2, \ldots, N - n$ only. Use the method of Proposition 6.1 of Chapter 4 in Ross to show that the likelihood you found in part (b) is maximized at $\hat{K} = \lfloor N - 1.25n + 1 \rfloor$.

   (d) Since $\mathcal{C} = N - n$, a *guessing penalty* is applied by subtracting $\lfloor 0.25n \rfloor$ from $\mathcal{C}$ to get an estimate of K. For $N = 100$ and $K = 90$, compare the *examiner's estimate* $\tilde{K} = N - n - \lfloor 0.25n \rfloor$ and the maximum likelihood estimate $\hat{K}$ for each possible value that n can take on, *viz.* $n = 0, 1, \ldots 10$. Notice that lucky guesses cause the examiner to overestimate K while the unlucky student who blows all ten problems has to suffer the further indignity of having the score reduced to something smaller than K.