University of Illinois **Fall 2011**

ECE 313: Problem Set 4
Confidence intervals, ML parameter estimation, Bernoulli processes, Poisson Distribution

Due: Wednesday September 21 at 4 p.m.
Reading: *313 Course Notes* Sections 2.5–2.9

1. **[Confidence Interval]**
   A communication system designer needs to simulate a communication link. The link is being designed for binary transmission, i.e., bits $b_k \in \{0,1\}$ are transmitted, and bits $\hat{b}_k \in \{0,1\}$ are received in the k^{th} bit-period. Noise and other sources of channel impairments result in a *bit-errors*, i.e., the event $\hat{b}_k \neq b_k$. The bit errors are assumed to occur independently from one bit-period to the next with a probability of error or *bit error-rate* (BER) of p_e. The communication link is being designed to meet the specifications set by an international standards committee such as ITU, IEEE, or ETSI. The standards document specifies that $p_e \leq 10^{-4}$. The designer estimates p_e by running n bits through her simulation model, counts the number of errors E by comparing the transmitted bit b_k with recovered bit $\hat{b}_k$, and obtains a BER estimate $\hat{p}_e = \frac{E}{n}$, where E is the error count, i.e., the total number of bits in error in a stream of n bits. The designer wants to impress her manager and wishes to report that her design meets the BER specifications with a ±10% tolerance around the 10^{-4} specified value.

   (a) Reason that E is a binomial random variable with parameters (n, p_e).

   (b) How many bits will she need to simulate the communication link over in order to achieve a confidence level of 99%?

   (c) The designer has a desk-top with a dual core CPU running at a clock frequency of $4GHz$. The complexity of her simulation program and the associated compiler are such that each bit period takes 40 clock cycles (using both cores) to simulate. How much time in minutes will it take her to simulate all n bits, where n is the answer to part (a).

   (d) Realizing the simulation time in part (c) is too long, she settles upon a strategy to run simulations with fewer bits initially until the design has matured and then run a long one. She decides to give 10 minutes per run. How many bits can she simulate in this time and what is her level of confidence in her design assuming that the tolerance remains fixed at ±10%?

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 off-campus interview invitation 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 $p = 0.95$ of obtaining an invitation, whereas a C in 313 results in a probability of $p = 0.15$ of an invitation.

   (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 invitation.

   (b) On average, how many booth visits must an A student make before getting an off-campus interview invitation? How about 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 off-campus interview invitation. What does this mean, on average, for a C student in 313?

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

3. **[Maximum-likelihood Estimation]**
   Let $\mathbb{X}$ denote a discrete random variable that takes on integer values $1, 2, \ldots, n$. The value of n is unknown, and we wish to find its maximum-likelihood estimate $\hat{n}_{\text{ML}}$ from the observation that $\mathbb{X}$ had value 10 on a trial of the experiment.

   (a) Explain why $\hat{n}_{\text{ML}}$ must be 10 or more.

   (b) Suppose that $\mathbb{X}$ has the *increasing-ramp* pmf $p_{\mathbb{X}}(k) = \begin{cases} \frac{2k}{n(n+1)}, & 1 \leq k \leq n, \\ 0, & \text{otherwise.} \end{cases}$

   What is $\hat{n}_{\text{ML}}$ in this case?

   (c) Suppose that $\mathbb{X}$ has the *decreasing-ramp* pmf $p_{\mathbb{X}}(k) = \begin{cases} \frac{2(n+1-k)}{n(n+1)}, & 1 \leq k \leq n, \\ 0, & \text{otherwise.} \end{cases}$

   i. Compute the value of $p_{\mathbb{X}}(10)$ for $n = 10, 11, 12, \ldots$ and find the maximum-likelihood estimate $\hat{n}_{\text{ML}}$ numerically.

   ii. Now suppose that $\mathbb{X}$ has value i. Find $\hat{n}_{\text{ML}}$ as a function of i and verify that when $i = 10$, your function gives the same value for $\hat{n}_{\text{ML}}$ as you found in part (c)(i).

4. **[Poisson and Binomial Distributions]**
   Suppose that 105 passengers hold reservations for a 100-passenger flight from Chicago to Champaign. The number of passengers who show up at the gate can be modelled as a binomial random variable $\mathbb{X}$ with parameters $(105, 0.9)$.

   (a) On average, how many passengers show up at the gate?

   (b) If $\mathbb{X} \leq 100$, everyone who shows up gets to go. Find the value of $P\{\mathbb{X} \leq 100\}$.

   (c) Explain why the number of *no-shows* can be modelled as a binomial random variable $\mathbb{Y}$ with parameters $(105, 0.1)$.

   (d) Notice that the probability that everyone who shows up gets to go can also be expressed as $P\{\mathbb{Y} \geq 5\}$. Use the *Poisson approximation* to compute $P\{\mathbb{Y} \geq 5\}$ and compare your answer to the "more exact" answer that you found in part (b).

5. **[Inferring true performance on multiple-choice examinations]**
   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 on the answer sheet, 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. Show that the likelihood you found in part (b) is maximized at $\hat{K}_{\text{ML}} = \lfloor N - 1.25n + 1 \rfloor$.

   (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 *guessing-penalty estimate* $\hat{K}_{\text{GP}} = N - n - \lfloor 0.25n \rfloor$ and the maximum likelihood estimate $\hat{K}_{\text{ML}}$ 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.