University of Illinois **Spring 2012**

ECE 313: Problem Set 4
Geometric, Poisson, and Negative Binomial Distributions, Bernoulli Process, ML estimation, and Markov and Chebychev Inequalities

Due: Wednesday, February 15 at 4 p.m.
Reading: *ECE 313 Course Notes,* Sections 2.5–2.9.

1. **[Buffering in Video Streaming]**
Suppose that you are watching a video clip consisting of thirty frames per second from YouTube. Each frame consists of 1000 packets. When a frame is about to be played out, if five or more packets of the frame are lost (i.e., if you have received 995 or fewer packets of the frame), you will experience buffering during that frame. Each packet is lost with the probability θ, independently of other packets.

   (a) Suppose that θ is unknown. If you have observed that 20 packets of a frame are lost, what is the maximum likelihood estimate $\hat{\theta}_{ML}(20)$ of probability θ?

   (b) Suppose $\theta = 0.001$. Using the Poisson approximation, find the probability that you experience buffering during a given frame.

2. **[Insurance Company]**
An insurance company sold life insurance policies to 100,000 people for a premium of \$500 each. Assume that the probability of death of each insured person during the contract term is 0.001 and the deaths occur independently. In the case of death, the company pays \$200,000 to a designated beneficiary. Let p_L denote the probability that the company loses money for the contract term.

   (a) Find the upper bound on p_L yielded by Markov's inequality.

   (b) Find the upper bound on p_L yielded by Chebychev's inequality.

3. **[Hypergeometric Distribution]**
Suppose there is a box containing m red balls and $n-m$ blue balls. When a sample of k balls is taken randomly from among the n balls in the box, let X represent the number of red balls in the sample.

   (a) Find the probability mass function $p_X(x)$ and mean $E[X]$ of X if the k balls are taken one at a time with replacement.

   (b) Suppose there are at least k blue balls and at least k red balls in the box, i.e., $m \geq k$ and $n-m \geq k$. We now assume that the k balls are taken without replacement. Find the probability mass function $p_X(x)$ and mean $E[X]$ of X. The distribution of X is known to be *the hypergeometric distribution.* To find $E[X]$, you may need the following Chu-Vandermonde identity: for $s \leq \min\{w, h-w\}$,

$$\sum_{y=0}^{s} \binom{w}{y}\binom{h-w}{s-y} = \binom{h}{s}.$$

4. **[Unreliable Wireless Communications and Opportunistic Scheduling]**
Suppose that a smart phone attempts to transmit a packet to a base station using wireless communications as shown in Fig. 1. Due to unreliability of the wireless channel, the base station successfully receives the packet with probability $q \in (0,1)$, independently of outcomes of prior transmissions. The smart phone knows whether or not the packet is successfully received by the base station right after the transmission. If the packet is not received, then the smart phone immediately retransmits.

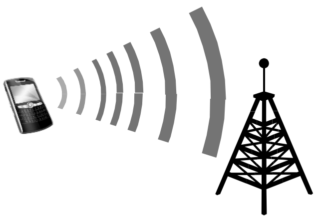

Figure 1: Packet transmission from a smart phone to a cell tower

   (a) Let X be the number of transmission attempts until a packet is successfully delivered to the base station. Find the pmf, mean, and variance of X.
   (b) Suppose the smart phone has r packets to send, one at a time. Let Y be the number of transmission attempts until all r packets are successfully transmitted. Find the pmf, mean, and variance of Y.
   (c) Now assume that there are M base stations around the smart phone as shown in Fig. 2. When the smart phone transmits a packet, each base station successfully receives the packet with probability q, independently of the other base stations and prior transmissions. We say that *the smart phone successfully delivers a packet* if at least one of the M base stations successfully receives the packet. Let Q be the number of transmission attempts until a packet is successfully delivered. Find the pmf, mean, and variance of Q.
   (d) Suppose the smart phone has r packets to send, one at a time. For each packet, the smart phone keeps transmitting the packet until one of the M base stations successfully receives the packet. Let Z be the number of transmission attempts until r packets are successfully delivered. Find the pmf, mean, and variance of Z. Further, show that when $q \approx 0$, $E[Y]$ in (b) can be approximated by $M \cdot E[Z]$, i.e., when the channel condition is pretty bad, the expected delay to deliver r packets as in (d) is M times smaller that the expected delay as in (b). (Hint: Use L'Hopital's rule: If $\lim_{t\to c} f(t) = \lim_{t\to c} g(t) = \infty$ and $f'(t)$ and $g'(t)$ exist near $t = c$, then
$$\lim_{t\to c}\frac{f(t)}{g(t)} = \lim_{t\to c}\frac{f'(t)}{g'(t)}.$$

5. **[Matlab Simulation: Unreliable Wireless Communications and Opportunistic Scheduling]**
Verify your analysis in the previous problem using a Matlab simulation. In this simulation, we will generate the samples of Bernoulli processes, geometric random variables, and negative binomial random variables. The hints for generating these samples are described as below:

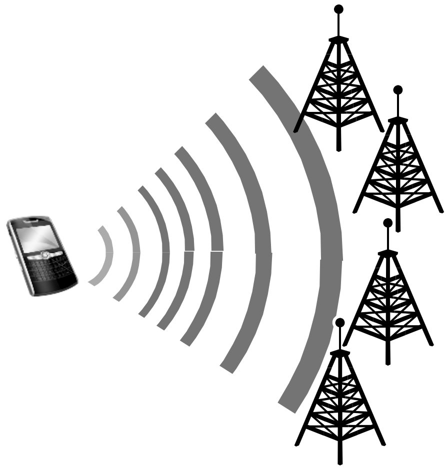

Figure 2: Opportunistic Scheduling: Packet transmission from a smart phone to multiple cell towers

- A length T segment of a Bernoulli process with parameter q can be obtained using `B=( rand(1, T)<q)` for some large enough `T`.

- The indices in B where the corresponding Bernoulli random variable is one can be obtained using `S=find(B==1)`. Here `length(S)` will be around qT.

- The intervals between ones can be computed using `S(2:length(S))-S(1:(length(S)-1))`. The result is a row vector of length `length(S)-1` filled with geometric random variables with parameter q.

- The indices when every r-th one occurs in the Bernoulli process can be obtained by `C=S(r:r:length(S))`. The length will be around qT/r.

- The intervals between these timeslots can be obtained using `C(2:length(C))-C(1:(length(C)-1))`. The result is a row vector filled with negative binomial random variables with parameters r and q.

(a) For `q=0.3` and `T=floor(2000/q)`, generate the samples of X in 4(a). Display the distributions of X using `hist()`.

(b) For `r=300`, `q=0.3`, and `T=floor(2000*r/q)`, generate the samples of Y in 4(b). Display the distributions of Y in 4(b) using `hist()`.