MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Department of Electrical Engineering & Computer Science
**6.041/6.431: Probabilistic Systems Analysis**
(Fall 2010)

**Problem Set 10**
**Due December 2, 2010 (in recitation)**

1. **A financial parable.** An investment bank is managing $1 billion, which it invests in various financial instruments ("assets") related to the housing market (e.g., the infamous "mortgage backed securities"). Because the bank is investing with borrowed money, its actual assets are only $50 million (5%). Accordingly, if the bank loses more than 5%, it becomes insolvent. (Which means that it will have to be bailed out, and the bankers may need to forgo any huge bonuses for a few months.)

   (a) The bank considers investing in a single asset, whose gain (over a 1-year period, and measured in percentage points) is modeled as a normal random variable $R$, with mean 7 and standard deviation 10. (That is, the asset is expected to yield a 7% profit.) What is the probability that the bank will become insolvent? Would you accept this level of risk?

   (b) In order to safeguard its position, the bank decides to diversify its investments. It considers investing $50 million in each of twenty different assets, with the $i$th one having a gain $R_i$, which is again normal with mean 7 and standard deviation 10; the bank's gain will be $(R_1 + \cdots + R_{20})/20$. These twenty assets are chosen to reflect the housing sectors at different states or even countries, and the bank's rocket scientists choose to model the $R_i$ as independent random variables. According to this model, what is the probability that the bank becomes insolvent?

   (c) Based on the calculations in part (b), the bank goes ahead with the diversified investment strategy. It turns out that a global economic phenomenon can affect the housing sectors in different states and countries simultaneously, and therefore the gains $R_i$ are in fact positively correlated. Suppose that for every $i$ and $j$ where $i \neq j$, the correlation coefficient $\rho(R_i, R_j)$ is equal to $1/2$. What is the probability that the bank becomes insolvent? You can assume that $(R_1 + \cdots + R_{20})/20$ is normal.

2. The adult population of Nowhereville consists of 300 males and 196 females. Each male (respectively, female) has a probability of 0.4 (respectively, 0.5) of casting a vote in the local elections, independently of everyone else. Find a good numerical approximation for the probability that more males than females cast a vote.

3. Let $S_n$ be the number of successes in $n$ independent Bernoulli trials, where the probability of success in each trial is $p = \frac{1}{2}$. Provide a numerical value for the limit as $n$ tends to infinity for each of the following three expressions:

   (a) $\mathbf{P}(\frac{n}{2} - 10 \leq S_n \leq \frac{n}{2} + 10)$
   (b) $\mathbf{P}(\frac{n}{2} - \frac{n}{10} \leq S_n \leq \frac{n}{2} + \frac{n}{10})$
   (c) $\mathbf{P}(\frac{n}{2} - \frac{\sqrt{n}}{2} \leq S_n \leq \frac{n}{2} + \frac{\sqrt{n}}{2})$

4. Alice has two coins. The probability of heads for the first coin is $1/3$; the probability of heads for the second coin is $2/3$. Other than this difference in their bias, the coins are indistinguishable through any measurement known to man. Alice chooses one of the coins randomly and sends it to Bob. Let $p$ be the probability that Alice chose the first coin. Bob tries to guess which of the two coins he received by flipping it 3 times in a row and observing the outcome. Assume that all coin flips are independent. Let $Y$ be the number of heads Bob observed.

(a) Given that Bob observed $k$ heads, what is the probability that he received the first coin?

(b) Find values of $k$ for which the probability that Alice sent the first coin increases after Bob observes $k$ heads out of 3 tosses. In other words, for what values of $k$ is the probability that Alice sent the first coin given that Bob observed $k$ heads greater than $p$? If we increase $p$, how does your answer change (goes up, goes down, or stays unchanged)?

(c) Help Bob develop the rule for deciding which coin he received based on the number of heads $k$ he observed in 3 tosses if his goal is to minimize the probability of error.

(d) For this part, assume $p = 2/3$.

    i. Find the probability that Bob will guess the coin correctly using the rule above.

    ii. How does this compare to the probability of guessing correctly if Bob tried to guess which coin he received before flipping it?

(e) If we increase $p$, how does that affect the decision rule?

(f) Find the values of $p$ for which Bob will never guess he received the first coin, regardless of the outcome of the tosses.

(g) Find the values of $p$ for which Bob will always guess he received the first coin, regardless of the outcome of the tosses.

5. Consider a Bernoulli process $X_1$, $X_2$, $X_3$, ... with unknown probability of success $q$. As usual, define the $k$th inter-arrival time $T_k$ as

$$T_1 = Y_1, \qquad T_k = Y_k - Y_{k-1}, \quad k = 2, 3, \ldots$$

where $Y_k$ is the time of the $k$th success. This problem explores estimation of $q$ from observed inter-arrival times $\{t_1, t_2, t_3, \ldots\}$.

You may find the following integral useful: For any non-negative integers $k$ and $m$,

$$\int_0^1 q^k (1-q)^m dq = \frac{k! \, m!}{(k + m + 1)!}$$

Assume $q$ is sampled from the random variable $Q$ which is uniformly distributed over $[0, 1]$.

(a) Compute the PMF of $T_1$, $p_{T_1}(t_1)$

(b) Compute the least squares estimate (LSE) of $Q$ from the first recording $T_1 = t_1$.

(c) Compute the maximum a posteriori (MAP) estimate of $Q$ given the $k$ recordings, $T_1 = t_1, \ldots, T_k = t_k$.

For this part only assume $q$ is sampled from the random variable $Q$ which is now uniformly distributed over $[0.5, 1]$

(d) Find the linear least squares estimate (LLSE) of the second inter-arrival time $(T_2)$, from the observed first arrival time $(T_1 = t_1)$.

6. The joint PDF of $X$ and $Y$ is defined as follows:

$$f_{X,Y}(x, y) = \begin{cases} cxy & \text{if } 0 < x \leq 1, \ 0 < y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the normalization constant $c$.

(b) Compute the conditional expectation estimator of $X$ based on the observed value $Y = y$.

(c) Is this estimate different from what you would have guessed before you saw the value $Y = y$? Explain.

(d) Repeat (b) and (c) for the MAP estimator.