STA 247 — Assignment #2, Due in class on December 4

Late assignments will be accepted only with a valid medical or other excuse.

This assignment is to be done by each student individually. You may discuss it in general terms with other students, but the work you hand in should be your own. Handing in work that is not your own is a serious academic offense. Fabricating results, such as handing in fake output that was not actually produced by your program, is also an academic offense.

Part I

1) Let X be a random variable with range $\{1, 2, 3\}$ and let Y be a random variable with range $\{0, 1\}$. The joint probability mass function of X and Y is defined as follows:

$$p_{XY}(1,0) = 0.2, \quad p_{XY}(2,0) = 0.3, \quad p_{XY}(3,0) = 0.1$$

$$p_{XY}(1,1) = 0.1, \quad p_{XY}(2,1) = 0.1, \quad p_{XY}(3,1) = 0.2$$

   Do all of the following, obtaining, where appropriate, an actual numerical answer (decimal number of simple fraction). Show for each part how you obtained your answer.

   a) Find the marginal distributions of X and Y.
   b) Compute $P(X \leq 2Y)$.
   c) Determine whether or not the random variables X and Y are independent.
   d) Compute $E[X]$ and $E[Y]$.
   e) Compute $E[X^2 + Y]$ and $E[XY - 2]$.
   f) Compute $P(X = 1|Y = 0)$ and $P(Y < 1|X > 1)$.
   g) Compute $VAR(X)$ by using its marginal distribution, and then by using the conditional variance formula (conditioning on Y) — that is, use $VAR(X) = E[VAR(X|Y)] + VAR(E[X|Y])$.
   h) Compute $COV(X, Y)$.

2) I have a fair 6-sided die. I throw it many times and I stop when every number from 1 to 6 has appeared at least once. Let X be the total number of rolls needed for this to happen. Define another colection of random variables Y_i with $i = 1, 2, ...6$ as follows: Y_1 is the number of rolls until the first number apprears (obviously, $Y_1 = 1$.) Y_2 is the number of additional rolls until another number (different from the first number) appears. After this second number appears, Y_3 is the number of rolls until a third number appears (different from the first two). In similar fashion, define Y_4, Y_5, and Y_6.

   (a) Describe the distribution of each of Y_2, Y_3, Y_4, Y_5, and Y_6 — ie, what family the distribution is in, and what the parameters of the distribution within that family are. Calculate the expectation and the variance of each of $Y_1, \ldots, Y_6$.
   (b) Explain why $X = \sum_{i=1}^{6} Y_i$, and use this fact to calculate $E[X]$.
   (c) Explain why the random variables $Y_1, \ldots, Y_6$ are independent of one another, and use this fact to calculate $VAR(X)$.

Part II

Bits sent through a communications channel are sometimes received with the wrong value. For some channels, such errors often occur in bursts, with several errors occurring in a row. We can model such error behaviour using a Markov chain. Let $E(i)$ be the random variable having the value 1 if an error occurred in bit i, and 0 otherwise. Suppose that these errors have the Markov property, so that

$$P(E(i) = e_i \,|\, E(i-1) = e_{i-1}) \;=\; P(E(i) = e_i \,|\, E(i-1) = e_{i-1},\, E(i-2) = e_{i-2}, \ldots, E(0) = e_0)$$

We can then specify the error behaviour of the channel by the one-step transition probabilities for this Markov chain. Suppose that these transition probabilities are the same at all times (ie, the Markov chain is homogeneous). The one-step transition probabilities will then be determined by just two numbers, P_{01} and P_{11}, defined by

$$P_{01} = P(E(i) = 1 \,|\, E(i-1) = 0), \quad P_{11} = P(E(i) = 1 \,|\, E(i-1) = 1)$$

a) Find $P(E(i+3) = 1 \,|\, E(i) = 1)$ exactly, assuming that $P_{01} = 0.01$ and $P_{11} = 0.4$.

b) Find the steady-state probabilities for this Markov chain. In other words, find the limit of $P(E(i) = 1)$ as i becomes very large. Express this probability as a function of P_{01} and P_{11}, and also find its numerical value when $P_{01} = 0.01$ and $P_{11} = 0.4$.

c) Suppose that data is sent through this channel in blocks of 120 bits, using an error-correcting code that is capable of fixing up to four errors in a block (but no more). Write an R function that simulates the transmission of 1000 such blocks through this channel, with the values P_{00} and P_{01} being arguments of this function. The function should assume that the bit before the first bit in the first block simulated was not an error. The function should return a list containing the following three values:

`failure_rate`	The fraction of blocks with errors that couldn't be fixed
`mean_errors`	The average number of errors in a block
`var_errors`	The variance of the number of errors in a block

Hand in a listing of your function, the output of four runs of this function with $P_{01} = 0.01$ and $P_{11} = 0.4$, and the output of four runs of this function with P_{01} and P_{11} both set to the steady-state probability for $E(i) = 1$ when $P_{01} = 0.01$ and $P_{11} = 0.4$ that you found in part (b).

d) Discuss your results. In particular, try to explain why the results (failure rates, mean number of errors, and variance of the number of errors) for the two setting of P_{01} and P_{11} are or are not different, and try to explain why the means and variances should be as they are. Can any of these results found approximately using computer simulation be easily found exactly?