ESD.86

Models, Data, Inference for Socio-Technical Systems

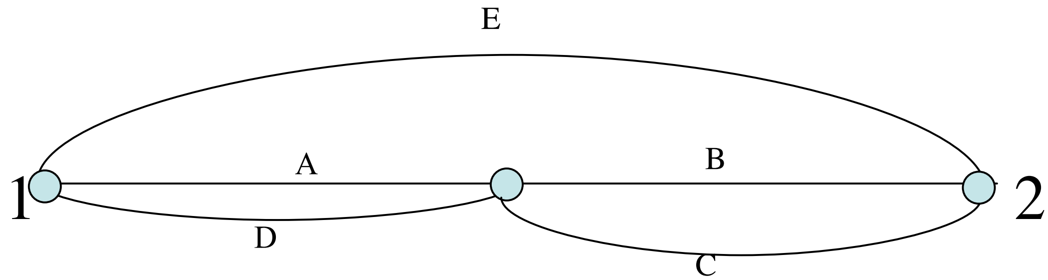
Massachusetts Institute of Technology

Engineering Systems Division

Problem Set #2

Issued: Tuesday February 20, 2007. Due: Wednesday February 28, 2007.

1. Network Robustness. In Lecture #2 we showed how to find the probability of Transmission from one point to another on a communications network having faulty links. In particular, we dealt with this network:



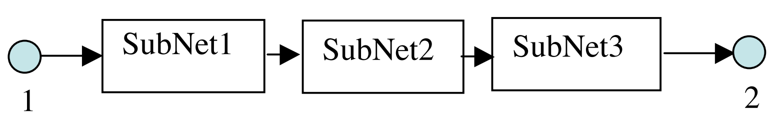
And we found that the probability of successful transmission T_{12} from node 1 to node 2 is

$$P\{T_{12}\} = p_E + (1-p_E)\{[p_A+(1-p_A)\,p_D]\ [p_B+(1-p_B)\,p_C]\}.$$

When expanded, we see the 'Venn diagram' adding and subtracting of overlapping points in the diagram:

$$\begin{aligned} P\{T_{12}\} = {} & p_E + p_A p_B + p_B p_D + p_A p_C + p_C p_D \\ & - p_A p_B p_D - p_A p_C p_D - p_A p_B p_C - p_B p_C p_D \\ & + p_A p_B p_C p_D \\ & - p_A p_B p_E - p_B p_D p_E - p_A p_C p_E - p_C p_D p_E \\ & + p_A p_C p_D p_E + p_A p_B p_D p_E + p_A p_B p_C p_E + p_B p_C p_D p_E \\ & - p_A p_B p_C p_D p_E \end{aligned}$$

(a) Considering that each link can either be functioning correctly or broken, write out the sample space for this network.
(b) For a general faulty network having N links, what is the size of the sample space?
(c) Consider a network structured like this:



Suppose that each of the three SubNets has the 5-link network topology as shown in the in the figure from lecture and copied above. But the individual link transmission probabilities are dependent on the SubNet containing them. For instance p_A would have different values, depending on whether we are in SubNet 1, 2 or 3. Determine the probability of successful transmission from node 1 to node 2 in this more complex transmission network.

(d) **"…ility".** Suppose you were given a budget to add one redundant link to one of the three subnets in the problem above. Your only option is to add a duplicate link in parallel to one of the existing links, and the probability of successful transmission of the new link will be identical to that of the link it is directly parallel to. We assume that the 2 links will operate independently. You want to do this in order to maximize the increase in probability of successful transmission from node 1 to node 2. Explain carefully how you would frame and formulate this problem.

2. Binomial Distribution and Baseball. Suppose the Boston Red Sox this year, in their 162-game season, can be modeled with a simple probability model. In particular, we suppose that we can model the outcome (Win or Loss) of each game as being determined by an independent Bernoulli trial. We assume (optimistically) that P{Win} = 0.60 and P{Loss} = 0.40. Then the total number of wins (N_w) over the course of the 162-game season can be written as the sum of 162 independent indicator random variables,

$$N_w = \sum_{i=1}^{162} X_i$$

where

$$X_i = \begin{cases} 1 & \text{if game } i \text{ is a Win} \\ 0 & \text{if game } i \text{ is a Loss} \end{cases}.$$

(a) Show that the discrete or z-transform of X_i is equal to $[0.4 + 0.6z]$.
(b) From what you know about the transform of the sum of independent random variables, argue that the discrete or z-transform of N_w is equal to $[0.4 + 0.6z]^{162}$.
(c) From what you know about the definition of the z-transform, show that you can 'invert' the transform to find P_k = P{Red Sox win precisely k games this season}. In fact, show that we have the Binomial Distribution,

$$P_k = \binom{162}{k}(0.6)^k(0.4)^{162-k}, k = 0,1,2,...,162.$$

3. Interviewing Movie Goers. You post a questionnaire on the web, asking moviegoers questions about the films that they go to the theater and see. One of the items on the questionnaire is this: *"The last time you went to the theater to see a film, estimate the fraction of seats in the theater that were occupied by fellow moviegoers."* Let's assume that each answer is precisely correct. That is, each individual's ability to estimate the percentage of seats occupied is perfect; there is no estimation error. And let's assume that the average of the answers you get back, averaged over many responses (say 1,000+), is 55%. Now, I am going to give you a piece of data: For the entire USA, the theater industry has computed this fraction:

f = (# of seats sold during 1 week)/(total # of seats offered for sale during 1 week)

In a typical week, the theater industry has found that $f = 0.05$, far from 0.55. Construct a quantitative argument, based on probabilistic reasoning and people's behavior, explaining the apparent huge discrepancy.

4. Jogging. It's a nice day in January in the Boston area. The temperature is a balmy 70 degrees, and people are jogging. Suppose the jogging path of interest is an infinitely long straight East-West path and that all joggers move at the same speed. Joggers can enter the jogging path at any point. Suppose any given jogger is equally likely to first jog East, then West, or to first jog West and then East. Each jogger will jog some jogger-specific maximum distance in the first direction chosen, stop, turn around and then complete the jog by retracing steps and exiting at the point at which the jog was started. That is, joggers are assumed to enter and exit the jogging path at the same location. And, each is assumed to jog a finite distance, but distances will differ. Now, at a random time and random place, you enter the jogging path and start your jog.

(a) At some (random) time during your jog, you pass another jogger moving towards you, moving in the direction opposite to yours. Show that with probability at least 0.75 you will not pass that jogger again on today's jog.
(b) Jogger J enters the path at the same time as you, 2 units of distance away from you and jogs 3 units of distance before turning around. Suppose you decided to jog X units of distance before turning around, where X is exponentially distributed with parameter 1. Find the probability of meeting jogger J and the probability of meeting her twice. *NB: Remember that any given jogger is equally likely to first jog East, then West, or the reverse.*

[You have just successfully completed a written doctoral exam question in applied probability, administered last month to MIT Operations Research doctoral candidates.]

5. **Returning to the Broken Stick Live Experiment.** In the experiment we did in class we used the **Four Steps to Happiness** to derive that the probability is 0.25 that one can form a triangle with the three pieces of yardstick broken at two random points. In this problem let's think of the problem in a different way.
Suppose we think of the lengths of the three pieces of stick as L, M and R, corresponding to L = length of left most piece, M = length of middle piece and R = length of right-most piece. Each is a non-negative random variable over [0, 1 The sample space remains a unit-area square in the positive quadrant, with axes labeled x_1 and x_2 as the places on the

stick of the random marks. Each (x_1, x_2) pair gives rise to experimental values of L, M and R.

(a) Argue from basic principles that
$P\{L \geq 1/2\} = 1/4,\ P\{M \geq 1/2\} = 1/4,\ P\{R \geq 1/2\} = 1/4.$

(b) Let $P\{\Delta\} \equiv P$(a triangle can be formed from the 3 pieces}. Argue that
$P\{\Delta\} = 1 - P\{[L \geq 1/2] \cup [M \geq 1/2] \cup [R \geq 1/2]\}$

(c) We know from basic probability that P{A ∪ B} = P{A} + P{B} – P{A ∩ B}. Show that the analogous result for three events, A, B and C is
P{A ∪ B ∪ C}= P{A}+P{B}+P{C}-P{B ∩ C}-P{A ∩ B}-P{A ∩ C)}+P{A ∩ B ∩ C}.

(d) If you apply the result of part (c) to your result in part (b), do you get the correct answer for $P\{\Delta\}$? Can you identify in the (x_1, x_2) sample space the three separate events $P\{L \geq 1/2\},\ P\{M \geq 1/2\},\ P\{R \geq 1/2\}$? Can you see at once that the respective events are disjoint and that $P\{L \geq 1/2\} = 1/4,\ P\{M \geq 1/2\} = 1/4,\ P\{R \geq 1/2\} = 1/4$?

(e) Extend the logic to marking the stick at 3 random points, all uniformly distributed over [0,1] and mutually independent. What is the probability that we can form a quadrilateral with the four pieces we obtain when we cut the yardstick at the 3 marked places?

(f) Extend the logic to marking the stick at (n-1) random points, all uniformly distributed over [0,1] and mutually independent (n = 3, 4, 5, ...). What is the probability that we can form an “n-gon” with the n pieces we obtain when we cut the yardstick at the (n-1) marked places? (An n-gon is a polygon with n sides.)