Massachusetts Institute of Technology
Department of Electrical Engineering & Computer Science
6.041/6.431: Probabilistic Systems Analysis
(Fall 2011)

Problem Set 7
Due November 16, 2011

1. Fred is giving out samples of dog food. He makes calls door to door, but he leaves a sample (one can) only on those calls for which the door is answered *and* a dog is in residence. On any call the probability of the door being answered is 3/4, and the probability that any household has a dog is 2/3. Assume that the events "Door answered" and "A dog lives here" are independent and also that the outcomes of all calls are independent.

   (a) Determine the probability that Fred gives away his first sample on his third call.

   (b) Given that he has given away exactly four samples on his first eight calls, determine the conditional probability that Fred will give away his fifth sample on his eleventh call.

   (c) Determine the probability that he gives away his second sample on his fifth call.

   (d) Given that he did not give away his second sample on his second call, determine the conditional probability that he will leave his second sample on his fifth call.

   (e) We will say that Fred "needs a new supply" immediately *after* the call on which he gives away his last can. If he starts out with two cans, determine the probability that he completes at least five calls before he needs a new supply.

   (f) If he starts out with exactly m cans, determine the expected value and variance of D_m, the number of homes with dogs which he passes up (because of no answer) before he needs a new supply.

2. An elementary experiment is independently performed N times, where N is a Poisson random variable with mean λ. Let $\{a_1, a_2, \ldots, a_k\}$ be the set of sample outcomes of the elementary experiment, and let p_i, $1 \le i \le k$, denote the probability of the outcome a_i.

   (a) Let N_i denote the number of elementary experiments performed for which the outcome is a_i. Find the PMF of N_i for $1 \le i \le k$.

   (b) Find the PMF of $N_1 + N_2$.

   (c) Find the conditional PMF of N_1 given that $N = n$.

   (d) Find the conditional PMF of $N_1 + N_2$ given that $N = n$.

   (e) Find the conditional PMF of N given that $N_1 = n_1$.

3. All ships travel at the same speed through a wide canal. Eastbound ship arrivals at the canal are a Poisson process with an average arrival rate λ_E ships per day. Westbound ships arrive as an independent Poisson process with average arrival rate λ_W per day. An indicator at a point in the canal is always pointing in the direction of travel of the most recent ship to pass it. Each ship takes t days to traverse the length of the canal.

   (a) Given that the pointer is pointing west:

      i. What is the probability that the next ship to pass it will be westbound?

      ii. What is the PDF for the remaining time until the pointer changes direction?

   (b) What is the probability that an eastbound ship will see no westbound ships during its eastward journey through the canal?

(c) We begin observing at an arbitrary time. Let V be the time we have to continue observing until we see the seventh eastbound ship. Determine the PDF for V.

4. (a) Shuttles depart from New York to Boston every hour on the hour. Passengers arrive according to a Poisson process of rate λ per hour. Find the expected number of passengers on a shuttle. (Ignore issues of limited seating.)

   (b) Now suppose that the shuttles are no longer operating on a deterministic schedule, but rather their interdeparture times are independent and exponentially distributed with rate μ per hour. Find the PMF for the number of shuttles arriving in one hour.

   (c) Let us define an "event" to be either the arrival of a passenger, or the departure of a shuttle (or both simultaneously). With the same assumptions as in (b) above, find the expected number of "events" that occur in one hour.

   (d) If a passenger arrives at the gate, and sees 2λ people waiting, find his/her expected time to wait until the next shuttle.

   (e) Find the PMF for the number of people on a shuttle.

5. Let T_1, T_2 (respectively, S) be exponential random variables with parameter λ (respectively, μ). We assume that all three of these random variables are independent. Derive an expression for the expected value of $\min\{T_1 + T_2, S\}$. *Hint:* Refer to Chapter 6, Problem 19 in the text.

6. Consider a Poisson process of rate λ. Let random variable N be the number of arrivals in $(0, t)$ and M be the number of arrivals in $(0, t + s)$, where $t > 0$ and $s > 0$.

   (a) Find the joint PMF of N and M, $p_{N,M}(n, m)$.

   (b) Find $\mathbf{E}[NM]$.

G1†. The interval between consecutive mammographic (X-ray) screenings for different groups of women in the population significantly affects the breast cancer survival rate. The breast cancer death rate can be reduced by finding the tumors at smaller sizes, and the probability of finding a tumor while it is small increases with the frequency of screening.

In this problem you will develop the simplest model for the probability a tumor arises and metastasizes prior to detection by mammography (generally a lethal event), as a function of the inter-screening interval T. Assume that tumors arise as a homogeneous Poisson process with a rate λ in each woman in the population independently. The model defines a tumor size in terms of the diameter $d(t)$ of the tumor, which is assumed to be spherical, and first notes the existence of a tumor when it reaches a minimal diameter $d_{\min}$. It is known that many kinds of tumors grow exponentially with a deterministic rate:

$$d(t) = d_{\min} e^{(t-t_b)/t_g}, t \geq t_b, \tag{1}$$

where $d(t)$ is the tumor diameter at time t, the tumor has reached the minimal diameter $d_{\min}$ at its "birth" time t_b, and t_g is the growth time constant. (We choose $d_{\min} = 1$mm, rather than the diameter of 1 cell, as the "birth diameter" to avoid modeling the more rapid tumor growth that occurs when $d < 1$mm, since these diameters are too small to present any significant opportunity for either detection or metastasis.)

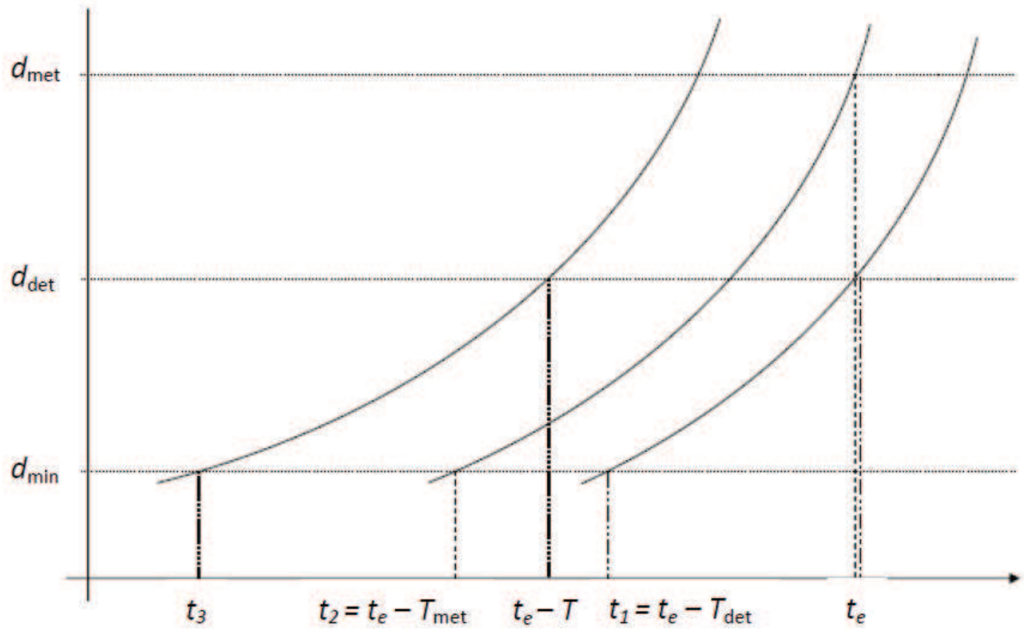

Figure 1: Growth curves for tumors born at times t_1, t_2 and t_3: A tumor "born" later than t_1 will not be detected in an exam at time t_e. A tumor born between t_1 and t_2 will first be detected at t_e and will not yet have metastasized. A tumor born between t_2 and t_3 will first be detected at t_e and will already have metastasized. A tumor born before t_3 will first be detected at an earlier exam before t_e. The length of the interexam period T may be larger or smaller than shown here. In particular, T could be quite large or T could be smaller than T_{det}. It is important to understand the consequences of both alternatives.

When a tumor is present, the probability of detection on a single mammographic exam, P_{det}, grows monotonically with d. In the simple model we use here, $P_{\text{det}}(d)$ is a step function, *i.e.*,

$$P_{\text{det}}(d) = \begin{cases} 0, & d < d_{\text{det}}, \\ 1, & d \geq d_{\text{det}}, \end{cases} \tag{2}$$

where d_{det} is the "smallest detectable tumor diameter". Assume that a tumor that is detected is always successfully removed immediately following detection.

(a) Find T_{det}, the time required for a tumor to grow from diameter d_{min} to diameter d_{det}, as a function of d_{min}, d_{det}, and t_g. Assume that $d_{\text{min}} < d_{\text{det}}$.

Similarly, the literature shows that the probability a tumor has already metastasized by the time it is detected is a monotone increasing function of the tumor diameter d at detection. In the simple model we use here, the probability that a tumor has already metastasized, P_{met}, is also a step function of the tumor diameter d, *i.e.*,

$$P_{\text{met}}(d) = \begin{cases} 0, & d < d_{\text{met}}, \\ 1, & d \geq d_{\text{met}}. \end{cases} \tag{3}$$

(b) Find T_{met}, the time required for a tumor to grow from diameter d_{min} to diameter d_{met}, as a function of d_{min}, d_{det}, and t_g. Assume that $d_{\text{min}} < d_{\text{met}}$.

(c) A typical breast tumor grows at a rate such that the tumor **volume** doubles every 130 days, at least for tumor diameters $\geq$ 1mm. Find the time constant t_g in Equation (1) for the growth rate of the tumor **diameter**.

For the remainder of this problem, assume that $d_{\text{min}} < d_{\text{det}} < d_{\text{met}}$. For part (d), assume that a woman goes in for a mammogram every T years, and that the exam in question occurs at a time t_e. Assume steady state, i.e., assume the patient has been going for mammograms at a period of once every T years for a long time in the past and will continue going every T years for a long time in the future. Assume that any tumor detected at an exam prior to t_e has been completely removed.

(d) Find the probability, as a function of T, that one or more tumors come into existence in one interexam period T and eventually metastasize.

(e) Find a simple approximation to your exact answer to part (d) for the realistic case $\lambda(T + T_{\text{det}} + T_{\text{met}}) \ll 1$. Use this approximation to give a simple verbal explanation for what these results mean.

(f) Now substitute in the following numerical values, which will let you determine the behavior of this model (which is approximately correct) for cancer mammography for a typical woman.

- For the majority of women during much of their lives, λ is approximately 1/250 tumors per year.
- The doubling time of tumor **volume** is typically about 130 days.
- Current mammogram technology gives about a 50% probability of detecting a tumor of diameter 0.7cm, a reasonable value for d_{det}.
- A breast tumor has a 50% probability of metastasizing by the time it has a diameter of 5cm, a reasonble value for d_{met}.
- As already mentioned, d_{min} is 1mm.

Find the probability that over a 30-year period, a woman ever develops breast cancer that will eventually metastasize, if for her entire adult life she has a mammogram

i. every 3 years;
ii. every 5 years.

As a final comment, this crude model gives surprisingly good answers despite its many flaws, which include

- ignoring the genetic differences between women which make certain women much more likely to have breast cancer than the average,
- ignoring the change in λ with age (including the fact that younger women are much less likely to develop a tumor, but the ones that do develop are often faster growing),
- the grading of P_{det} and P_{met} with tumor size,
- the varying abilities of different machines and doctors to spot a small tumor,
- the difficulties posed by biopsies that do not find a tumor, and
- the radiation exposure over a lifetime of mammograms.