UC Berkeley
Department of Electrical Engineering and Computer Sciences

EE126: PROBABILITY AND RANDOM PROCESSES

Midterm 2
Spring 2011

Solution to Problem 1. ...

(a) Let us define

$$T = \Big\{(x, y) \in \mathbb{R}_+^2 :\ x + \frac{y}{2} \le 1\Big\},$$
$$C_t = \{(x, y) \in \mathbb{R}_+^2 :\ x + y \le t\}$$

We know that (X, Y) is uniformly distributed in T. The CDF of $X + Y$ is given by

$$\mathbb{P}(X + Y \le t) = \mathbb{P}\big((X, Y) \in C_t\big) = \frac{\text{area}(C_t \cap T)}{\text{area}(T)} = \text{area}(C_t \cap T).$$

If $t < 0$, then $C_t \cap T = \emptyset$. If $t \in (0, 1)$, then $C_t \subset T$ and since C_t is a triangle with side lengths t, we get

$$\text{area}(C_t \cap T) = \text{area}(C_t) = \frac{1}{2}t^2.$$

If $t \in (1, 2)$, C_t and T partially intersect. The intersection of the hypotenuses is given by solving

$$\begin{cases} x + y/2 = 1, \\ x + y = t \end{cases} \implies x = -t + 2, \quad y = 2(t - 1).$$

The area of $C_t \cap T$ is that of T minus a triangle bounded by the two lines above and y-axis,

$$\text{area}(C_t \cap T) = 1 - \frac{1}{2}(2 - t)(-t + 2) = 1 - \frac{1}{2}(2 - t)^2.$$

Hence the CDF of $Z := X + Y$ is

$$\mathbb{P}(Z \le t) = \begin{cases} 0, & t \in (-\infty, 0) \\ \frac{1}{2}t^2, & t \in [0, 1) \\ 1 - \frac{1}{2}(2 - t)^2, & t \in [1, 2) \\ 1, & t \in [2, \infty) \end{cases}$$

. The PDF is

$$f_Z(t) = \frac{d}{dt}\mathbb{P}(Z \le t) = \begin{cases} 0 & t \in (-\infty, 0) \cup (2, \infty) \\ t & t \in (0, 1) \\ 2 - t & t \in (1, 2) \end{cases}$$

To compute $\mathbb{E}[X \mid Y = 0.25]$, first note that by inspection of the figure, conditioned on $Y = y$, X is distributed as $\text{Unif}(0, 1 - y/2)$. Hence, conditioned on $Y = y$, the expected value of X is $\frac{1}{2}(1 - y/2)$. Thus,

$$\mathbb{E}[X \mid Y = 0.25] = \frac{1}{2}\Big(1 - \frac{0.25}{2}\Big) = \frac{7}{16}.$$

(b) First assume that $F_X(\cdot)$ is (strictly) increasing, hence it has a proper inverse $F_X^{-1}(\cdot)$ which is also increasing. Then, for $y \in [0, 1]$

$$\mathbb{P}\big(Y \le y\big) = \mathbb{P}\big(F_X(X) \le y\big) = \mathbb{P}\big(X \le F_X^{-1}(y)\big) = F_X(F_X^{-1}(y)) = y.$$

Note also that since $Y \in [0, 1]$ due to the properties of the CDF $F_X(\cdot)$, we have that $\mathbb{P}(Y \le y)$ is 0 if $y < 0$ and is 1 if $y > 1$. Hence the CDF of Y is that of a $\text{Unif}(0, 1)$. That is, $Y \sim \text{Unif}(0, 1)$ or

$$f_Y(y) = 1\{y \in (0, 1)\}.$$

(c) Let the position of the three cars be X_1, X_2 and X_3. We know that $X_i \overset{iid}{\sim} \text{Unif}(0, 1)$ for $i = 1, 2, 3$. One approach to the problem is to consider their order statistics

$$X_{(1)} \le X_{(2)} \le X_{(3)}.$$

That is, $X_{(1)}$ is the minimum of $\{X_1, X_2, X_3\}$ and so on. We know that the joint PDF of $(X_{(1)}, X_{(2)}, X_{(3)})$ is

$$f(x_1, x_2, x_3) := 3!\, 1\{0 < x_1 < x_2 < x_3 < 1\}.$$

The event of interest is $\big\{X_{(3)} \ge X_{(2)} + d,\ X_{(2)} \ge X_{(1)} + d\big\}$, whose probability is

$$I := \int_{\substack{x_3 \ge x_2 + d, \\ x_2 \ge x_1 + d}} f(x_1, x_2, x_3) dx_1 dx_2 dx_3 = 3! \int_0^1 dx_3 \int_0^{x_3 - d} dx_2 \int_0^{x_2 - d} dx_1$$

Note that the inner integral evaluates to $x_2 - d$ if $x_2 > d$ and zero otherwise[1].

[1] We are using the notation $\int_a^b$ to denote $\int_{[a,b]}$. That is, our integrals are not oriented and $\int_a^b f = 0$ if $a > b$.

One way to write this is $\max\{x_2 - d, 0\} =: (x_2 - d)_+$. Hence,

$$
\begin{aligned}
I &= 3! \int_0^1 dx_3 \int_0^{x_3 - d} (x_2 - d)_+ \, dx_2 \\
&= 3! \int_0^1 dx_3 \int_d^{x_3 - d} (x_2 - d) \, dx_2 \\
&\overset{(a)}{=} 3! \int_0^1 dx_3 \int_0^{x_3 - 2d} u \, du \\
&= 3! \int_0^1 \frac{1}{2} (x_3 - 2d)_+^2 dx_3 \\
&= 3! \int_{2d}^1 \frac{1}{2} (x_3 - 2d)^2 dx_3 \\
&\overset{(b)}{=} 3! \int_0^{1-2d} \frac{1}{2} v^2 dv = 3! \frac{1}{3 \cdot 2} (1 - 2d)^3 = (1 - 2d)^3.
\end{aligned}
$$

(d) We have $X \sim \text{Geom}(p)$ where $\mathbb{E}(X) = \frac{1}{p} = 5$ and $Y \sim \text{Unif}(0, a)$ where $a = 1$, and the two are independent. Let $r(t)$ denote the CDF of a $\text{Unif}(0, 1)$ random variable. That is,

$$
r(t) = \begin{cases} 0, & t \in (-\infty, 0) \\ t, & t \in [0, 1] \\ 1, & t \in (1, \infty) \end{cases}
$$

Since Y/a is $\text{Unif}(0, 1)$, the CDF of Y is $r(t/a)$. Let us compute the CDF of $T = X + Y$. First, consider

$$
\mathbb{P}(T \le t \mid X = k) = \mathbb{P}(Y \le t - k \mid X = k) = \mathbb{P}(Y \le t - k) = r\Big(\frac{t - k}{a}\Big)
$$

$$
\mathbb{P}(T \le t) = \sum_{k=1}^{\infty} \mathbb{P}(T \le t \mid X = k) \, \mathbb{P}(X = k) = \sum_{k=1}^{\infty} r\Big(\frac{t - k}{a}\Big)(1 - p)^{k-1} p.
$$

The PDF is obtained by differentiating with respect to t. Note that we can write $r'(t) = 1\{t \in (0, 1)\}$, hence

$$
f_T(t) = \sum_{k=1}^{\infty} \frac{1}{a} r'\Big(\frac{t - k}{a}\Big)(1 - p)^{k-1} p = \frac{p}{a} \sum_{k=1}^{\infty} (1 - p)^{k-1} \, 1\big\{t - k \in (0, a)\big\}.
$$

Plugging in the numerical values,

$$
f_T(t) = \frac{1}{5} \sum_{k=1}^{\infty} \Big(\frac{1}{5}\Big)^{k-1} 1\big\{t - k \in (0, 1)\big\}.
$$

Simplified form. It is possible to simplify this a bit. Fix some $t \in [1, \infty)$, then $t - k \in (0, a)$ iff $k \in (t - a, t)$. Since k is an integer, this is equivalent to

$k \in (\lceil t-a \rceil, \lfloor t \rfloor)$. Hence, (using the notation $a \vee b = \max\{a,b\}$)

$$f_T(t) = \frac{p}{a} \sum_{k=\lceil t-a \rceil \vee 1}^{\lfloor t \rfloor} (1-p)^{k-1} = \frac{p}{a} \frac{(1-p)^{(\lceil t-a \rceil \vee 1)-1} - (1-p)^{\lfloor t \rfloor}}{1-(1-p)}$$
$$= \frac{1}{a} \Big\{ (1-p)^{(\lceil t-a \rceil \vee 1)-1} - (1-p)^{\lfloor t \rfloor} \Big\}.$$

For our numerical example, we get

$$f_T(t) = \left\{ \left(\frac{4}{5}\right)^{(\lceil t-1 \rceil \vee 1)-1} - \left(\frac{4}{5}\right)^{\lfloor t \rfloor} \right\}, \quad t \in [1, \infty).$$

For this special case $a = 1$, the answer actually simplifies further (see Figure 1),

$$f_T(t) = \left(\frac{4}{5}\right)^{\lfloor t \rfloor - 1} \left(\frac{1}{5}\right), \quad t \in [1, \infty).$$

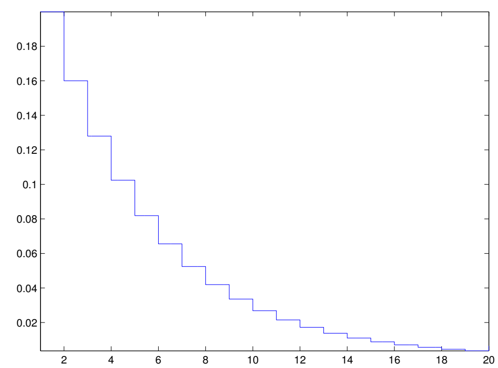

Figure 1: PDF of problem 1(d)

(e) Let X_i, $i = 1, \dots, n$ be the number of pennies in each of the $n = 100$ rolls the bank receives. We have

$$X_i \overset{iid}{\sim} \begin{cases} 49 & \text{w.p. } 0.3 \\ 50 & \text{w.p. } 0.6 \\ 51 & \text{w.p. } 0.1 \end{cases}$$

with

$$\mathbb{E}X_1 = 49(0.3) + 50(0.6) + 51(0.1) = 49.8,$$
$$\text{var}(X_1) = (0.8)^2 0.3 + (0.2)^2 0.6 + (1.2)^2 0.1 = 0.36.$$

Hence the standard deviation of X_i is $\sigma = \sqrt{\text{var}(X_i)} = 0.6$. Let $S_n := \sum_{i=1}^n X_i$. The problem asks for the probability of the following event

$$A := \Big\{ 50n - S_n \geq 25 \Big\}$$

We assume that by CLT, $Z := \frac{S_n - n(\mathbb{E}X_1)}{\sqrt{n}\sigma}$ is roughly distributed as $N(0,1)$. Note that $S_n = 10(0.6)Z + 49.8n$. Hence

$$\mathbb{P}(A) = \mathbb{P}(0.2n - 6Z \geq 25) = \mathbb{P}(Z \leq -\frac{5}{6}) \approx \Phi(-\frac{5}{6}) = 1 - \Phi(\frac{5}{6}) = 0.2023.$$

Solution to Problem 2. ...
The pair (X, Y) is uniformly distributed in the square $[0, 60]^2$ with area 3600. Figure 2 shows the region where their meeting is feasible and the desired sub-regions, for parts (a) and (b).
The feasible region can be described in equations as follows:

- For $X < 30$ and $X < Y$, they meet iff $Y \leq \min\{X + 10, 30\}$.
- For $X > 30$ and $X < Y$, they meet iff $Y \leq X + 15$.
- For $X > Y$, they meet iff $X \leq Y + 15$.

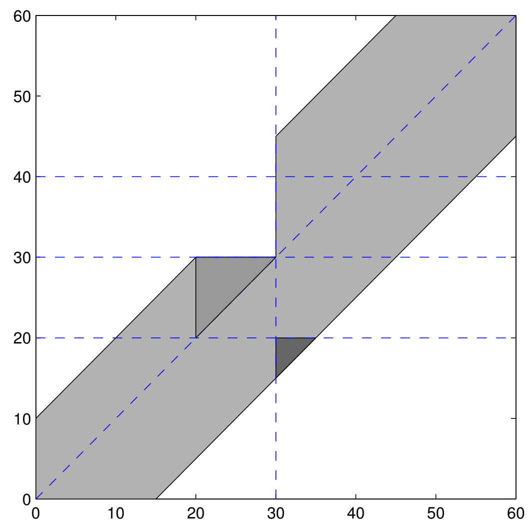

Figure 2: Feasible region for problem 2

(a) Based on the figure, we have to compute the area of a triangle with side lengths equal 10 (see the figure.) Hence, the desired probability is

$$\frac{\frac{1}{2}10^2}{60^2} = \frac{1}{72}.$$

(b) Here ,the region is a triangle of side lengths 5. Hence,

$$\frac{\frac{1}{2}5^2}{60^2} = \frac{1}{288}.$$

(c) Here, the region is the part of the feasible region to the left of the line $x = 30$. We can compute the area as that of the square $[0, 30]^2$ minus that of two (isosceles) triangles of side lengths 15 and 20, respectively.

$$\frac{30^2 - \frac{1}{2}20^2 - \frac{1}{2}15^2}{60^2} = \frac{47}{288}.$$

Solution to Problem 3. ...

(a) Let $N(t)$ be the total number of votes in the first t hours. Similarly, let $N_A(t)$ and $N_B(t)$ be the vote counts of candidates A and B in the first t hours, respectively.

Conditioned on $N(10) = 300$, the 300 votes are uniformly distributed in $[0, 10]$. The chance the any of these votes lies in $[0, 4]$ is then $\frac{4}{10}$. Since the votes are independently tagged A and B with probability $\frac{1}{2}$, the chance that any of those votes lies in $[0, 4]$ and belongs to A is $\frac{1}{2}\frac{4}{10} = \frac{1}{5}$.

Hence, conditioned on $N(10) = 300$, $N_A(4)$ is binomial with parameters 300 and $\frac{1}{5}$, that is

$$\mathbb{P}\big(N_A(4) = n \mid N(10) = 300\big) = \binom{300}{n}\Big(\frac{1}{5}\Big)^n\Big(\frac{4}{5}\Big)^{300-n},$$

for $n = 0, 1, \dots, 300$.

(b) Let X_i be the label of the i-th vote, either A or B. From the problem statement, the sequence X_i, $i = 1, 2, \dots$ is a Bernoulli process with parameter $p = \frac{1}{2}$, independent of the underlying Poisson process. That is, the label of the i-th vote is independent of the time it is cast.

Hence, the distribution of the number of votes B has received just before A receives their first vote is (shifted) geometric with parameter $p = \frac{1}{2}$. If we let M denote this number, we have

$$\mathbb{P}(M = k) = (1-p)^k p = \Big(\frac{1}{2}\Big)^{k+1}, \quad k = 0, 1, 2, \dots.$$

(c) Let us call the interval between two reversals, a reversal interval, denoted as RI. Consider the following sequence of votes

B	B	B	B	A	A	A	B	A	A
1	2	3	4	5	6	7	8	9	10

Reversals occur at positions 5 and 8. This is a general pattern for a RI ending in "A B". (There is corresponding pattern for a RI ending in "B A", obtained by interchanging A and B. Due to symmetry, it is enough only to consider those ending in "A B", for example.)

It is tempting to say that the length of the RI is that of an inter-arrival time of the voting process of B, which is a Poisson process with rate $\frac{1}{2}\lambda$. Hence, the expected length of the RI of the first type is $\frac{2}{\lambda}$. But this is not quite the case. (There is probability a way to fix this argument, but I don't see it right now!)

Let us argue as follows. Let t_i, $i = 1, 2, \dots$ be i.i.d. Expo(λ). Consider a RI ending in "A B". The length of this portion is distributed as t_1. Moving backwards, we could get either A before "A B", in which case we continue, or we could get B at which case we stop. As long as we get A's we keep adding t_i to get the length of RI. It not hard to see that the number of $\{t_i\}$ that we add up is distributed as $\widetilde{M} := 1 + M$, where M is as in part (b). Hence, the expected length of RI is

$$\mathbb{E}\Big(\sum_{i=1}^{\widetilde{M}} t_i\Big) = \mathbb{E}(\widetilde{M})\mathbb{E}(t_1) = \frac{1}{p}\frac{1}{\lambda} = \frac{2}{\lambda} = \frac{1}{15}.$$

(d) Again, it is tempting to argue that this should be exponential with rate $\lambda/2$ based on the inter-arrival argument. Let us, however, use our formula $R := \sum_{i=1}^{\widetilde{M}} t_i$. Conditioned on $\widetilde{M} = n$, R is an Erlang RV with parameters n and λ. We have, for $t > 0$,

$$\begin{aligned} f_R(t) = \sum_{k=1}^{\infty} f_{R|\widetilde{M}=k}(t)\mathbb{P}(\widetilde{M} = k) &= \sum_{k=1}^{\infty} \Big(\frac{1}{2}\Big)^k \frac{\lambda^k t^{k-1} e^{-\lambda t}}{(k-1)!} \\ &= \frac{\lambda}{2} e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\frac{\lambda}{2}t)^k}{k!} = \frac{\lambda}{2} e^{-\lambda t} e^{\frac{\lambda}{2}t} = \frac{\lambda}{2} e^{-\frac{\lambda}{2}t} \end{aligned}$$

which shows that R is indeed $\text{Expo}(\lambda/2) = \text{Expo}(15)$.

Solution to Problem 4. (Bonus problem)

(a) Let L be the length of the chord and let D be its distance from the origin. We have the following relation

$$D^2 + \Big(\frac{L}{2}\Big)^2 = r^2.$$

We know that D is $\text{Unif}(0, r)$. We have

$$\begin{aligned} \mathbb{P}(L > r) = \mathbb{P}\Big(\big(\frac{L}{2}\big)^2 > \big(\frac{r}{2}\big)^2\Big) &= \mathbb{P}\Big(r^2 - D^2 > \big(\frac{r}{2}\big)^2\Big) \\ &= \mathbb{P}\Big(D < \frac{\sqrt{3}}{2} r\Big) = \frac{\sqrt{3}}{2}. \end{aligned}$$

(b) Unless the midpoint of a chord lies at the center of the circle, the chord is uniquely determined by its midpoint. Let (X, Y) be the location of the midpoint which we assume is uniformly distributed in the circle. With the notations of part (a), we have $D = \sqrt{X^2 + Y^2}$. Hence,

$$\begin{aligned} \mathbb{P}(L > r) &= \mathbb{P}\Big(D^2 < \frac{3}{4} r\Big) \\ &= \mathbb{P}\Big(X^2 + Y^2 < \frac{3}{4} r^2\Big) \\ &= \int_{x^2+y^2<\frac{3}{4}r^2} \frac{1}{\pi r^2}\, dx\, dy = \frac{1}{\pi r^2} \int_0^{\frac{\sqrt{3}}{2}r} \int_0^{2\pi} \rho\, d\rho\, d\theta = \frac{2}{r^2} \Big[\frac{1}{2}\rho^2\Big|_0^{\frac{\sqrt{3}}{2}r} = \frac{3}{4}. \end{aligned}$$