

# Assignment 1

Benjamin Jakubowski

September 16, 2015

## 1. STRANDED

### A. PROBABILITY THREE FRIENDS COME

In this problem, each friend comes and picks you up with a probability of 0.1. If you assume each friend decides independently from the others, the number of friends that come pick you up can be modeled as  $X \sim \text{Binomial}(.1)$ . Thus, the probability that three of your five friends come is

$$P(X = 3) = \binom{5}{3} .1^3 .9^2 = .0081$$

### B. LOWER BOUND ON PROBABILITY NO FRIENDS COME

If your friends do not act independently, we can place a lower bound of 0.5 on the probability that no friends come and you are left stranded (i.e.  $P(X = 0) \geq 0.5$ ). To see this, we can model all your friends' decisions using a single uniform random variable  $U(0, 1)$ . Each friend has a decision rule based on  $U$ . WOLOG, for friend 1 this rule is

$$F_1 = \begin{cases} \text{Picks you up} & : \text{if } U \in [0, 0.1] \\ \text{Doesn't pick you up} & : \text{if } U \notin [0, 0.1] \end{cases}$$

Now consider friend 2. To minimize the probability you are stranded (or equivalently maximize the probability you are picked up),

$$(F_2 = \text{Pick up}) \subseteq (F_1 = \text{Doesn't pick you up})$$

Thus, let

$$F_1 = \begin{cases} \text{Picks you up} & : \text{if } U \in [0.1, 0.2] \\ \text{Doesn't pick you up} & : \text{if } U \notin [0.1, 0.2] \end{cases}$$

This argument can be extended to the next three friends (constructing a decision rule such that  $(F_i = \text{Pick up}) \subseteq (F_k = \text{Doesn't pick you up})$  for all  $k \in 1, \dots, i - 1$ ). This results in you being picked up with a probability of 0.5. Thus,  $P(X = 0) \geq 0.5$ .

## 2. ARMY CAMP

### A. PROBABILITY JOE DISEASED GIVEN GROUP TEST POSITIVE

First, notation: let  $Joe_+$ ,  $Test_+$ , and  $Group_+$  represent that Joe is diseased, the test is positive, and the group sample contains the disease.  $Joe_-$ ,  $Test_-$ , and  $Group_-$  represent the inverses. Then the probability Joe is diseased given the group test is positive is given by:

$$\begin{aligned}
 P(Joe_+|Test_+) &= \frac{P(Joe_+ \cap Test_+)}{P(Test_+)} \\
 &= \frac{P(Joe_+)P(Test_+|Joe_+)}{P(Test_+|Joe_+) + P(Test_+|Joe_- \cap Group_+) + P(Test_+|Joe_- \cap Group_-)} \\
 &= \frac{.2 * .9}{.2 * .9 + .8 * (1 - .8^9) * .9 + .1 * .8^9 * .8} \\
 &= .2211
 \end{aligned}$$

(Note- implicit in this expression is the fact that if Joe is positive, the group sample is also positive- thus  $P(Test_+|Joe_+) = P(Test_+|Joe_+ \cap Group_+)$ .)

### B. CONDITIONAL INDEPENDENCE OF $Test_+$ AND $Joe_+$ GIVEN FRIDGE IS BROKEN

First, some new notation: let  $F_b$  and  $F_w$  represent the events the fridge is broken and the fridge is working, respectively. If  $F_b$  (which happens with probability 0.4), then the test always reads positive. Therefore, given the event  $F_b$ ,  $T_+$  and  $Joe_+$  are independent. To see this, note

$$P(Joe_+|Test_+ \cap Fridge_b) = P(Joe_+|Fridge_b)$$

since, given the fridge is broken, the event of a positive test  $T_+$  gives no additional information about Joe's state (since the test result is now completely determined by the broken fridge).

### C. PROBABILITY FRIDGE IS BROKEN GIVEN $Test_+$

The probability the fridge is broken given the test is positive is given by:

$$\begin{aligned}
 P(Fridge_b|Test_+) &= \frac{P(Fridge_b \cap Test_+)}{P(Test_+)} \\
 &= \frac{P(Fridge_b)}{P(Test_+|Fridge_b) * P(Fridge_b) + P(Test_+|Fridge_w) * P(Fridge_w)} \\
 &= \frac{.4}{1 * .4 + .6 * .2211} \\
 &= .4502
 \end{aligned}$$

### 3. OLD CAR

#### A. ANALYSIS OF DISTRIBUTION OF FIRST TIME CAR BREAKS

First, let  $X$  be the random variable modeling the number of drives until the car breaks down. Then  $X \sim \text{Geometric}(.25)$ . The probability the car breaks down after  $k'$  drives is

$$P(X = k') = .75^{k'-1} * .25$$

The probability that the car breaks down after  $k + k'$  drives, given it has already gone  $k$  drives, is<sup>1</sup>:

$$\begin{aligned} P(X = k + k' | X > k) &= \frac{P(X = k + k' \cap X > k)}{P(X > k)} \\ &= \frac{P(X = k + k')}{P(X > k)} \\ &= \frac{P(X = k + k')}{1 - P(X \leq k)} \\ &= \frac{.75^{k+k'-1} * .25}{1 - (1 - .75^k)} \\ &= \frac{.75^{k+k'-1} * .25}{.75^k} \\ &= .75^{k'-1} * .25 = P(X = k') \end{aligned}$$

Since  $P(X = k') = P(X = k + k' | X > k)$ , this implies the distribution of the first time the car breaks is memoryless.

#### B. PROBABILITY CAR BREAKS DOWN $n$ TIMES IN FIRST $k$ DRIVES

The probability the car breaks down  $n$  times in the first  $k$  drives is a binomial random variable  $B(n, .25)$

$$P(B = k) = \binom{k}{n} .25^n .75^{k-n}$$

#### C. PROBABILITY THE $n$ TH TIME THE CAR BREAKS DOWN IS AFTER $k$ DRIVES

The probability the  $n$ th time the car breaks down is after  $k$  drives is equivalent to the probability the car breaks down  $(n - 1)$  or fewer times in the first  $k$  drives:

$$P(B < n) = \sum_{i=0}^{n-1} \binom{k}{i} .25^i .75^{k-i}$$

Note this is simply the CDF of the binomial distribution.

---

<sup>1</sup>Note the denominator was evaluated using the CDF reported for a Geometric RV on Wikipedia.

#### D. EVALUATING A NEW PMF

Now, let  $X$  be the number of drives until the car breaks down. Then

$$P(X = k') = 1 - 2^{-k'}$$

Now consider  $P(X = k + k' | x > k)$

$$\begin{aligned} P(X = k + k' | X > k) &= \frac{P(X = k + k')}{P(X > k)} \\ &= \frac{P(X = k + k')}{1 - P(X = k)} \\ &= \frac{1 - 2^{-(k+k')}}{1 - (1 - 2^{-k})} \\ &= \frac{1 - 2^{-(k+k')}}{2^{-k}} \\ &= (1 - 2^{-(k+k')}) * 2^k \\ &= P(X = k + k') * 2^k. \end{aligned}$$

Thus, given this probability model, if you've already driven  $k$  times, you're  $2^k$  times more likely to end up having an accident.

This model seems to be (in some sense) more appropriate than the memoryless geometric distribution from part a. It is more appropriate if we assume that damage to your 2000 Ford Taurus accumulates with each drive. A further improvement to the model might be changing the geometric ratio from 2 to a smaller value (such that damage isn't modeled to accumulate so quickly).

### 4. RADIOACTIVE DECAY

#### A. PMF OF NUMBER OF HOURS TILL POSITIVE READING

The true decay time is an exponential random variable  $X \sim \text{exponential}(\lambda)$  with

$$f_X(x) = \lambda e^{-\lambda x}$$

Discretizing for hourly measurements, for  $y \in \mathbb{N}$  our new random variable  $Y$  has the pmf

$$\begin{aligned} f_Y(y) &= \int_{y-1}^y \lambda e^{-\lambda x} dx \\ &= -e^{-\lambda x} \Big|_{y-1}^y = -e^{-\lambda y} - (-e^{-\lambda(y-1)}) = e^{-\lambda(y-1)} - e^{-\lambda y} \end{aligned}$$

## B. PDF FOR ERROR BETWEEN READING AND TRUE DECAY TIME

Let  $y_i$  be the reading. Then the true decay time  $T$  is  $y_i - 1 \leq T \leq y_i$ .

The PDF and CDF for  $T$  (expressed in terms of the unconditional PDF and CDF of  $X$ ) is

$$f_T(t) = \begin{cases} \frac{f_X(x)}{F_X(y_i) - F_X(y_i - 1)} & : x \in [y_i - 1, y_i] \\ 0 & : x \notin [y_i - 1, y_i] \end{cases}$$

and

$$F_T(t) = \begin{cases} 0 & : x < y_i - 1 \\ \frac{F_X(x)}{F_X(y_i) - F_X(y_i - 1)} & : x \in [y_i - 1, y_i] \\ 1 & : x \geq y_i \end{cases}$$

Now, let  $E$  be the random variable describing the error between our on-the-hour measurement,  $y_i$ , and the true time of decay,  $T$ . Then  $E = y_i - T$ . Thus

$$\begin{aligned} F_E(e) &= P(E \leq e) = P(y_i - T \leq e) = P(T \geq y_i - e) = 1 - P(T \leq y_i - e) \\ &= 1 - F_T(y_i - e) = 1 - \frac{F_X(y_i - e)}{F_X(y_i) - F_X(y_i - 1)} \end{aligned}$$

Finally,

$$\begin{aligned} f_E(e) &= \frac{d}{de} F_E(e) \\ &= \frac{d}{de} \left( 1 - \frac{F_X(y_i - e)}{F_X(y_i) - F_X(y_i - 1)} \right) \\ &= \frac{f_X(y_i - e)}{F_X(y_i) - F_X(y_i - 1)} \end{aligned}$$

## 5. GENERATING RANDOM VARIABLES

### A. CDF OF $W = F_Y(Y)$

Assume  $Y$ , the available random variable, is continuous. Let  $W = F_Y(Y)$ . Then the CDF of  $W$  is

$$\begin{aligned} F_W(w) &= P(W \leq w) \\ &= P(F_Y(Y) \leq w) \\ &= P(Y \leq F_Y^{-1}(w)) \\ &= F_Y[F_Y^{-1}(w)] = w \end{aligned}$$

Thus,  $F_W(w) = w$ . This CDF implies that  $W$  is  $U \sim (0, 1)$ .

### B. CDF OF $F_X^{-1}(W)$

Let  $Z = F_X^{-1}(W)$ . Then the CDF of  $Z$  is

$$\begin{aligned}
 F_Z(z) &= P(Z \leq z) \\
 &= P(F_X^{-1}(W) \leq z) \\
 &= P(W \leq F_X(z)) \\
 &= F_W(F_X(z)) \\
 &= F_X(z)
 \end{aligned}$$

That's why our friend is telling us to do this- it gives us the desired CDF  $F_X$ .

### C. $F_Y$ NON-INVERTIBLE

Even if  $F_Y$  is non-invertible, this method will allow you to determine the CDF  $F_X$ .

Let  $R_Y$  be the support of  $Y$ . Assume that for a fixed constant  $w$ , all the values of  $y$  such that  $F_Y(y) = w$  belong to a closed interval  $[a(w), b(w)]$ .

Then take  $R_Y \setminus [a(w), b(w))$ . Now the only supported value of  $y$  such that  $F_Y(y) = w$  is  $y = b$ .

Repeating this process for every such constant  $w$  (noting the number of such constants must be a countable) returns a set  $S$  over which  $F_Y$  is invertible. Thus, restricting ourselves to this set allows us to construct  $F_Y^{-1} : S \rightarrow [0, 1]$ , thus allowing us to construct a  $U \sim (0, 1)$  random variable (and ultimately our desired CDF) even if  $F_Y$  is non-invertible.