

# Friends Problem

Declan Stacy

## Introduction

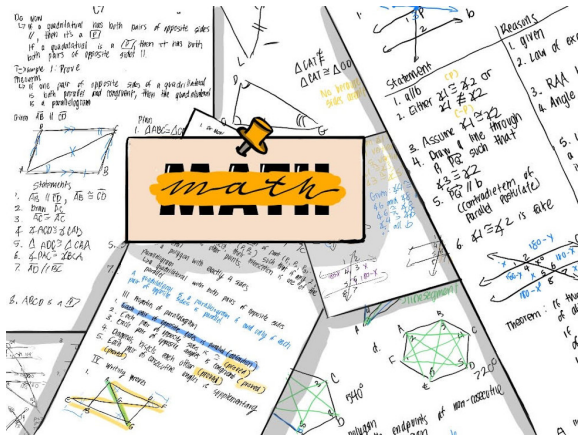


Figure 1: Image by Joey Chen

**Problem:** Two friends want to meet up at some time between 1 and 2 pm, but neither friend remembers the time they agreed to meet. So, both come at a random time chosen uniformly between 1 and 2. Both friends will wait 15 minutes for the other before giving up and going home. What is the probability that they successfully meet?

This is a problem that you may have heard of before and already know how to solve. In this article, I will present a solution and generalize the problem through the use of random variables, probability distributions, and expectation.

## Solution

One approach is to split the problem into cases. The first case is when both people arrive between 1:45 and 2. Since both friends will wait 15 minutes, they are guaranteed to meet if they both come within the same 15 minute interval. Thus, this case contributes  $\frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$  to the probability of them meeting (they both have a  $\frac{1}{4}$  chance of arriving between 1:45 and 2).

The second case is when the first person is the first to arrive and arrives before 1:45. In this case, the second person must arrive within 15 minutes of the first, which happens with probability  $\frac{1}{4}$ . Since there is a  $\frac{3}{4}$  chance of the first person arriving before 1:45, this case contributes  $\frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16}$  to the probability of them meeting.

Similarly, the second person could be the first to arrive and arrive before 1:45, so the first person would have to arrive within 15 minutes of the second, which also happens with probability  $\frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16}$ .

$$\text{Thus, the answer is } \frac{1}{16} + \frac{3}{16} + \frac{3}{16} = \boxed{\frac{7}{16}}.$$

This approach would work if the problem had more than two friends too, as long as they all are willing to wait for the same period of time. For example, if there were ten friends and each were willing to wait ten minutes, you could use the above approach to find the probability of the friends meeting. (As an exercise, try solving this problem for  $n$  friends that will all wait  $t$  hours.)

However, what if one friend was particularly impatient and would only wait for five minutes? What if another had no life and would be fine waiting 45 minutes? This problem would require more cases.

## The First Generalization

Consider  $n$  friends and the set  $\{t_1, t_2, \dots, t_n\}$ , where  $\forall i \in [n], t_i \in (0, 1)$ . Each friend  $i$  arrives at a time  $a_i$  chosen uniformly at random between 0 and 1 hours, and departs  $t_i$  hours later. Let  $M$  be the event that  $\exists T \in [0, 1)$  such that every friend is present at time  $T$  (present means the friend has arrived but not yet departed). What is  $P(M)$ ?

## Solution to the First Generalization

-1

### Intuition

Assuming  $M$  occurred, there is some earliest time  $T$  for which every friend is present. Thus,  $T$  is the time of the final arrival (otherwise, it would not be the earliest time that satisfies the condition). Knowing a friend is present at time  $T$  gives us information about when they could have arrived. They must have arrived sometime before  $T$ , but also, not too far before  $T$ , as otherwise they would have departed before  $T$ . Thus, we can easily calculate the probability that a particular friend is present at time  $T$ . Then we can do casework on which friend was the last to arrive to calculate the probability that the final arrival is at time  $T$  and all friends are present at time  $T$ . By summing this over all  $T \in [0, 1)$ , we will obtain  $P(M)$ .

However, we can't take a sum over all  $T \in [0, 1)$ , since  $[0, 1)$  is uncountable. So, we will make this argument rigorous by splitting the interval  $[0, 1)$  into  $m$  intervals of equal length and compute the sum over all intervals of the probability that  $T$  lies within the interval, taking the limit as  $m \rightarrow \infty$ . This will give us an integral from 0 to 1 that will evaluate to  $P(M)$ .

### Initial Claim

To solve this problem, we will partition the interval  $[0, 1)$  into  $m$  disjoint intervals,  $I_1, I_2, \dots, I_m$  of equal length. For example, if  $m = 4$ , the intervals would be  $[0, .25)$ ,  $[\cdot 25, .5)$ ,  $[\cdot 5, .75)$ , and  $[\cdot 75, 1)$ . For every interval  $I_i$ , let  $E_i$  be the event that exactly one friend arrived during  $I_i$  and all other friends were present during the entire interval  $I_i$ . I claim that

$$P(M) = \lim_{m \rightarrow \infty} \sum_{i=1}^m P(E_i).$$

## Showing $E_i$ 's Disjoint

Assume  $E_j$  and  $E_k$  both occur for  $0 < j < k \leq m$ .  $E_j$  occurring implies that all friends have arrived within or before the time interval  $I_j$ . However,  $E_k$  occurring implies that one friend arrived during interval  $I_k$ , a time interval after  $I_j$ . This is a contradiction, so all the events  $E_i$  are disjoint.

Let  $E = \bigcup_{i=1}^m E_m$ .

By the disjoint property this gives

$$P(E) = P\left(\bigcup_{i=1}^m E_m\right) = \sum_{i=1}^m P(E_i).$$

$$\text{Thus, } P(M) = \lim_{m \rightarrow \infty} \sum_{i=1}^m P(E_i) \iff P(M) = \lim_{m \rightarrow \infty} P(E).$$

## Bounding $P(E^C \cap M)$

$\forall i \in [m], E_i \subset M$  because  $E_i$  occurring implies that all the friends are present for some time in  $I_i \subset [0, 1)$ . Thus,  $\bigcup_{i=1}^m E_i = E \subset M$ , so  $P(M) = P(E) + P(M \cap E^C)$ .

Assuming  $M$  occurs, let  $T$  be the earliest time for which every friend is present. Since  $\bigcup_{i=1}^m I_m = [0, 1)$  and  $T \in [0, 1)$ ,  $\exists i$  such that  $T \in I_i$ . Since  $T$  is the earliest time when every friend is present, there must be an arrival at time  $T$ , so there must be an arrival during  $I_i$ .

If  $T \in I_i$ ,  $E_i \subset E$  will not occur if and only if more than one friend arrived during  $I_i$ , or there was a departure during  $I_i$ . (If you are having trouble understanding why this is, go back to the definition of  $E_i$ . There needs to be *exactly* one arrival during  $I_i$  and all other friends must be present for *the entire* interval  $I_i$ , so there must be no departures during  $I_i$ .)

Let  $B$  be the event  $\exists a \in [m]$  such that more than one friend arrived during  $I_a$ , and let  $C$  be the event  $\exists a \in [m]$  such that a friend arrived and a friend left during  $I_a$ . Let  $A = B \cup C$ .

$E^C \cap M \subset A$  because if  $E$  did not occur and  $M$  did occur,  $A$  must have occurred. So,  $P(E^C \cap M) \leq P(A)$ .

## Showing $P(A) = 0$

Since we know  $t_k$ , the difference between the arrival and departure times of friend  $k$ , knowing friend  $k$  left during an interval  $I_a = [\frac{a-1}{m}, \frac{a}{m})$  is equivalent to knowing they arrived during  $I_{\text{arrival}} := [\max(0, \frac{a-1}{m} - t_k), \frac{a}{m} - t_k)$ . Since the interval  $I_a$  has length greater than or equal to the length of  $I_{\text{arrival}}$ , we have that

$$\begin{aligned} P(\text{friend } k \text{ leaves during interval } I_a) &= P(\text{friend } k \text{ arrives during interval } I_{\text{arrival}}) \\ &\leq P(\text{friend } k \text{ arrives during interval } I_a) \end{aligned}$$

(Since arrival times are uniformly distributed between 0 and 1, the probability of an arrival during a certain interval contained in  $[0, 1)$  is simply the length of that interval.)

Event  $C$  requires at least one departure and at least one arrival during a certain interval, while event  $B$  requires at least two arrivals during a certain interval. Since the probability of a departure in a certain interval is less than or equal to the probability of an arrival,  $P(C) \leq P(B)$ . Also,  $P(A) = P(B \cup C) \leq P(B) + P(C)$ . Thus,  $P(A) \leq P(B) + P(C) \leq 2P(B)$ .

The probability that the  $k$ th friend arrives during a certain interval  $I_a$  is  $\frac{1}{m}$  as all intervals  $I_i$  are of length  $\frac{1}{m}$ . Since the arrival times of all friends are independent,

$$\begin{aligned} & P(\text{more than one friend arrived during } I_a) \\ &= 1 - P(\text{one or less friend arrived during } I_a) \\ &= 1 - P(\text{one friend arrived during } I_a) - P(\text{zero friends arrived during } I_a) \\ &= 1 - n \cdot \frac{1}{m} \cdot \left(\frac{m-1}{m}\right)^{n-1} - \left(\frac{m-1}{m}\right)^n \end{aligned}$$

(If this did not make sense, think about how you would calculate the probability of getting more than 1 heads as a result of  $n$  independent coin flips where each coin has a  $\frac{1}{m}$  chance of flipping heads.)

Then we can bound  $\lim_{m \rightarrow \infty} P(B)$ :

$$\begin{aligned} \lim_{m \rightarrow \infty} P(B) &= \lim_{m \rightarrow \infty} P(\exists j \in [m] \text{ such that more than one friend arrived during } I_j) \\ &\leq \lim_{m \rightarrow \infty} \sum_{j=1}^m P(\text{more than one friend arrived during } I_j) \\ &= \lim_{m \rightarrow \infty} \sum_{j=1}^m \left(1 - \frac{n}{m} \cdot \left(\frac{m-1}{m}\right)^{n-1} - \left(\frac{m-1}{m}\right)^n\right) \\ &= \lim_{m \rightarrow \infty} m \left(1 - \frac{n}{m} \cdot \left(\frac{m-1}{m}\right)^{n-1} - \left(\frac{m-1}{m}\right)^n\right) \\ &= \lim_{m \rightarrow \infty} \frac{m^n - n(m-1)^{n-1} - (m-1)^n}{m^{n-1}} \\ &= \lim_{m \rightarrow \infty} \frac{m^n - nm^{n-1} - m^n + nm^{n-1} + O(m^{n-2})}{m^{n-1}} = 0 \end{aligned}$$

(In the last line,  $-n(m-1)^{n-1}$  and  $-(m-1)^n$  are expanded.  $O(m^{n-2})$  is big-O notation, and is used to represent the sum of the terms in the expansions which have degree at most  $n-2$ . We say  $f(m) = O(g(m))$  if there exists a positive real number  $k$  such that  $|f(m)| \leq kg(m)$  for all  $m$  greater than or equal to some real number. So, as  $m \rightarrow \infty$ ,  $f(m)$  will behave like a constant multiple of  $g(m)$ . Thus,  $\lim_{m \rightarrow \infty} \frac{O(m^{n-2})}{m^{n-1}} \leq \lim_{m \rightarrow \infty} \frac{km^{n-2}}{m^{n-1}} = 0$ .)

Thus,  $\lim_{m \rightarrow \infty} P(M \cap E^C) \leq \lim_{m \rightarrow \infty} P(A) \leq \lim_{m \rightarrow \infty} 2P(B) = 0$ . Since probabilities are non-negative,  $\lim_{m \rightarrow \infty} P(M \cap E^C) = 0$ . So,  $P(M) = \lim_{m \rightarrow \infty} (P(E) + P(M \cap E^C)) = \lim_{m \rightarrow \infty} P(E)$ , proving the equivalent statement  $P(M) = \lim_{m \rightarrow \infty} \sum_{i=1}^m P(E_i)$ .

## Finding $E_i$

For  $E_i$  to occur, there must be exactly 1 friend that arrives during  $I_i$ . Let friend  $k \in [n]$  be the friend who arrives during  $E_i$ .

The second condition for  $E_i$  occurring is that all other friends are present for the entire interval  $I_i$ . This is equivalent to saying  $\forall j \neq k \in [n]$ , friend  $j$  arrived before  $\min(I_i)$ , but within  $t_j$  of  $\sup(I_i)$  (otherwise they would have left before the end of  $I_i$ ).

Since  $I_i = [\frac{i-1}{m}, \frac{i}{m})$ , this means that  $a_j < \frac{i-1}{m}$  and  $a_j + t_j \geq \frac{i}{m}$ , or equivalently,  $a_j \in [\max(0, \frac{i}{m} - t_j), \frac{i-1}{m})$  ( $a_j$  is the arrival time of friend  $j$ , so  $a_j$  can't be less than 0). This occurs with probability  $\frac{i-1}{m} - \max(0, \frac{i}{m} - t_j)$ .

Note that the arrival times of friends are independent of each other, so the probability of the intersection of multiple events in the form  $a_j \in I$  ( $I$  being an interval) occurring is equal to the product of the probabilities of those events occurring. Thus,  $P(a_k \in I_i \text{ and all other friends are present for the entire interval } I_i) = P(a_k \in I_i) \prod_{j \neq k \in [n]} P(\text{friend } j \text{ is present for the entire interval } I_i)$ .

Also, we can treat events in the form " $a_k \in I_i$  and all other friends are present for the entire interval  $I_i$ " as disjoint for different values of  $k$ . (Two of those events occurring would imply both friend  $k_1$  and friend  $k_2 \neq k_1$  are present for the entire interval  $I_i$  (which implies  $a_{k_1}, a_{k_2} \leq \min(I_i)$ ), but also that they both arrived during  $I_i$  ( $a_{k_1}, a_{k_2} \in I_i$ ). This could only be possible if they both arrive at  $\min(I_i)$ , but that has probability 0.)

Thus, we can sum  $P(a_k \in I_i \text{ and all other friends are present for the entire interval } I_i)$  over all possible  $k$  to get  $P(E_i)$ :

$$\begin{aligned}
& P(E_i) \\
&= \sum_{k=1}^n \frac{1}{m} \prod_{j \neq k \in [n]} \left( \frac{i-1}{m} - \max(0, \frac{i}{m} - t_j) \right) \\
&= \frac{1}{m} \sum_{k=1}^n \prod_{j \neq k \in [n]} \left( -\max(\frac{1-i}{m}, \frac{1}{m} - t_j) \right) \\
&= \frac{1}{m} \sum_{k=1}^n \prod_{j \neq k \in [n]} \min(\frac{i-1}{m}, t_j - \frac{1}{m}) \\
&= \frac{1}{m} \sum_{k=1}^n \prod_{j \neq k \in [n]} \left( \min(\frac{i}{m}, t_j) - \frac{1}{m} \right)
\end{aligned}$$

## Summing $P(E_i)$

$$\begin{aligned}
P(M) &= \lim_{m \rightarrow \infty} \sum_{i=1}^m P(E_i) \\
&= \lim_{m \rightarrow \infty} \sum_{i=1}^m \frac{1}{m} \sum_{k=1}^n \prod_{j \neq k \in [n]} \left( \min\left(\frac{i}{m}, t_j\right) - \frac{1}{m} \right) \\
&= \lim_{m \rightarrow \infty} \sum_{i=1}^m \frac{1}{m} \sum_{k=1}^n \prod_{j \neq k \in [n]} \min\left(\frac{i}{m}, t_j\right) + \lim_{m \rightarrow \infty} \sum_{i=1}^m \frac{1}{m} \sum_{k=1}^n O\left(\frac{1}{m}\right) \\
&= \int_0^1 \sum_{k=1}^n \prod_{j \neq k \in [n]} \min(x, t_j) dx + 0 \\
&= \boxed{\int_0^1 \left( \prod_{j \in n} \min(x, t_j) \right) \sum_{j \in n} \frac{1}{\min(x, t_j)} dx}
\end{aligned}$$

The final step uses the fact that the sum of  $n$  positive numbers taken  $n - 1$  at a time is equal to the product of all  $n$  numbers multiplied by the sum of the reciprocals of the  $n$  numbers. It is not necessary, but it does make calculations easier (instead of computing  $n$  products with  $n - 1$  numbers each and then adding those  $n$  numbers, you take the sum of  $n$  numbers and the product of  $n$  numbers and then multiply them).

## Summary

This integral confirms our original intuition, that the answer is essentially the sum over all possible times between 0 and 1 of the probability that one friend arrived exactly at that time, and that all the other friends had arrived before that time and are still present. If we were to compute this integral for a specific choice of  $\{t_1, t_2, \dots, t_n\}$  where  $t_1 \leq t_2 \leq \dots \leq t_n$ , the easiest way would be to express the integral from 0 to 1 as a sum of  $n + 1$  integrals, the first from 0 to  $t_1$ , the second from  $t_1$  to  $t_2$ , etc., with the last integral being from  $t_n$  to 1. This would allow us to get rid of the  $\min(x, t_j)$  terms and replace them with either  $x$  or  $t_j$ , as we would know based on the bounds of each integral if  $x \leq t_j$ . This would turn all of the integrands into polynomials, which are easy to integrate.

As mentioned in section 4.0, we did casework on the interval during which the final arrival occurred. It is almost the opposite of the solution to the simpler problem, where we did casework on the interval during which the first arrival occurred.

When I first tried to generalize the problem, I did casework on the interval during which the first arrival occurred. I encourage you to try this on your own, as it does lead to a solution (although it is recursive and arguably more of a pain to write out). However, I much prefer the approach presented in this paper because it easier to generalize further (as you will see in the next couple of sections), easier to code, and the resulting integral is relatively fast for a computer to compute (using the method described above).

# Introduction to Random Variables

## Motivation

In real life, your friends do not choose the time they show up for a meeting uniformly at random from 0 to 1. Chances are, you have some friends who tend to arrive early, and others who tend to arrive late. Maybe one friend is twice as likely to arrive between .1 and .2 than they are to arrive between .2 and .3. In order to further generalize this problem, we will need to find a way of representing how likely each friend is to arrive within a certain interval of time.

## Definition of Random Variable

For this paper, we will define a random variable  $X$  as a function from events to real numbers such that  $P(\{\omega \in \Omega | X(\omega) \leq x\})$  exists for  $x \in \mathbb{R}$ , where  $\Omega$  is the domain of  $X$ , also called the sample space.

(Note: There are random variables that do not map to the real numbers, but these will not be discussed. There are also a couple of other restrictions that are needed to make the definition of a random variable more rigorous, but this is also outside the scope of this paper.)

For example, let's say we roll a fair, standard 6-sided die, and we let  $X$  be a random variable representing the outcome of one roll. In this case,  $\Omega = \{\text{rolled a 1, rolled a 2, rolled a 3, rolled a 4, rolled a 5, rolled a 6}\}$ ,  $X(\text{rolled a 1}) = 1$ ,  $X(\text{rolled a 2}) = 2$ , etc.

## Cumulative Distribution Function (CDF)

$F_X(x) := P(\{\omega \in \Omega | X(\omega) \leq x\})$  is called the cumulative distribution function (CDF) of the random variable  $X$ . For the previous example with rolling a die, the CDF of  $X$  would be

$$F_X(x) = \begin{cases} 0 & x < 1 \\ \frac{\lfloor x \rfloor}{6} & 1 \leq x < 6 \\ 1 & 6 \leq x \end{cases}$$

## Discrete and Continuous Random Variables

In the previous example, there were only 6 possible outcomes for  $X$ . Since the number of outcomes is countable, we say that  $X$  is a discrete random variable. For a discrete random variable with outcomes  $x_1, x_2, \dots$ , there is a probability mass function (PMF),  $p_X(x_i) := P(X = x_i)$ . For rolling a die,  $p_X(1) = \frac{1}{6}$ ,  $p_X(2) = \frac{1}{6}$ , etc.

If  $X$  is not discrete, then we instead define the probability density function (PDF),  $f_X(x)$ . Note that  $f_X(x) = k$  does NOT mean that  $P(X = x) = k$ . Instead,  $f_X(x) = k$  iff for arbitrarily small  $\delta$ ,  $P(\{X \in (x, x + \delta)\}) \approx \delta k$ . The PDF does NOT represent a probability, and it can take on values greater than 1 (unlike a PMF).

If there exists a function  $f_X(x)$  that satisfies  $F_X(y) = \int_{-\infty}^y f_X(x)dx$  for all  $y$ , then  $X$  is a continuous random variable. Note that this definition implies that the CDF of a continuous random variable is continuous, and that if  $F_X(x)$  is differentiable at  $x$ , then  $\frac{d}{dx}F_X(x) = f_X(x)$ .

## Adding Independent Continuous Random Variables

Let's say  $Z = X + Y$ , where  $X$  and  $Y$  are independent continuous random variables (independent meaning  $\forall x, y \in \mathbb{R}, P(X \leq x \cap Y \leq y) = F_X(x)F_Y(y)$ ). Then  $Z$  is also a continuous random variable with  $f_Z(y) = \int_{-\infty}^{\infty} f_X(x)f_Y(y-x)dx$ . (You may recognize this as the convolution of  $f_X$  and  $f_Y$  evaluated at  $y$ ,  $(f * g)(y)$ .) This implies  $F_Z(y) = \int_{-\infty}^{\infty} F_X(x)f_Y(y-x)dx = \int_{-\infty}^{\infty} f_X(x)F_Y(y-x)dx$ . We will not actually use these rules in the paper, but it is useful to have an intuitive understanding of why they are true, so consider the discrete case.

Imagine  $X$  and  $Y$  were independent discrete random variables and could only take on integer values between  $-10$  and  $10$ . Then the probability that  $X + Y = 5$  would be the sum of the probabilities that  $X = i$  and  $Y = 5 - i$  for all possible  $i$ . Since  $X$  and  $Y$  are independent,  $P(X = i \cap Y = 5 - i) = P(X = i)P(Y = 5 - i)$ . Thus,  $p_{X+Y}(5) = \sum_{i=-10}^{10} p_X(i)p_Y(5-i)$ . For the continuous case, you would have PDFs instead of PMFs and an integral as opposed to a sum.

Note: This should also remind you of multiplying polynomials, which may motivate the representation of discrete random variables as polynomials. For our example of a die roll, you could represent the random variable  $X$  as  $\frac{1}{6}(x^1 + x^2 + x^3 + x^4 + x^5 + x^6)$ , or, more generally,  $\sum_i p_X(i)x^i$ , where the coefficients are the PMFs. Then  $P_{X+Y}(i)$  is the coefficient of the  $x^i$  term of the product of the polynomials representing  $X$  and  $Y$ . For example,  $p_{X+X}(5)$  is the coefficient of the  $x^5$  term of  $(\frac{1}{6}(x^1 + x^2 + x^3 + x^4 + x^5 + x^6))^2$ , which is  $\frac{4}{36}$ , meaning that the probability of rolling a sum of five on two fair, standard 6-sided dice is  $\frac{1}{9}$ . This representation will not be used in this paper, but if you would like to learn more about using polynomials to encode information, check out the Generating Functions article in the previous edition of Math & CS Research.

## Expected Value

One important piece of information that is useful when describing a random variable  $X$  is its expectation (expected value),  $E[X]$ . For a discrete random variable  $X$ ,  $E[X] := \sum_x xp_X(x)$ . In other words, it is the average over all possible outcomes of  $X$ , weighted by the PMF. For the previous example of a die roll,  $E[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{7}{2}$ .

The definition for a continuous random variable  $X$  is similar:  $E[X] := \int_{-\infty}^{\infty} xf_X(x)dx$ .

The expected value of a function of a random variable is exactly what you would expect; for a discrete random variable,  $E[g(X)] = \sum_x g(x)p_X(x)$ ; for a continuous random variable,  $E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$  (assuming  $g$  is a real-valued function defined over  $\mathbb{R}$ ).

Other useful properties about expected value (which will not be used or proven in this paper but are important to know) are:

$E[X + Y] = E[X] + E[Y]$  for random variables  $X$  and  $Y$  (linearity of expectation)

$E[kX] = kE[X]$  for a random variable  $X$  and constant  $k$ .

$E[XY] = E[X]E[Y]$  if  $X$  and  $Y$  are independent random variables.

## Law of Iterated Expectation

The Law of Iterated Expectation (also called Law of Total Expectation) states that for random variables  $X$  and  $Y$  defined over the same sample space,  $E[E[X|Y]] = E[X]$ .

Note that  $E[X|Y]$ , the "expected value of  $X$  given  $Y$ ," is a function of the random variable  $Y$ . Thus, we can write  $E[E[X|Y]] = \sum_y E[X|Y = y]p_Y(y)$  when  $Y$  is discrete, and  $E[E[X|Y]] = \int_{-\infty}^{\infty} E[X|Y = y]f_Y(y)dy$  when  $Y$  is continuous. (Note: When  $Y$  is a continuous random variable



and we condition on the event  $Y = y$ , it really refers to  $Y \in [y, y + \delta]$  where  $\delta \rightarrow 0$ , but for the sake of notation  $Y = y$  is used.) This is essentially a formalization of casework: instead of calculating  $E[X]$  directly, you go through every possible  $y$  that could be the outcome of  $Y$  and take the weighted average of  $E[X|Y = y]$  over all of those cases, with  $P(Y = y)$  as the weights.

I will not go in depth about what " $X$  given  $Y = y$ " means, but the basic idea is that if you know the outcome of  $Y$ , that tells you which possible events  $\omega$  could have occurred, which affects  $E[X]$ . In the case of a die roll, let's say you had another random variable  $Y$ , defined over the same sample space as  $X$  (from the previous examples), that is 1 if the die roll is greater than 3 and 0 otherwise. If we know  $Y = 1$ , then  $X$  must be either 4, 5, and 6, all with equal probability. (In this way, you can imagine conditioning on  $Y$  as a transformation of the sample space.) Thus,  $E[X|Y = 1] = 5$ . Similarly,  $E[X|Y = 0] = 2$ . Since  $P(Y = 1) = P(Y = 0) = \frac{1}{2}$ ,  $E[E[X|Y]] = 5 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{7}{2} = E[X]$ , as expected.

(If you would like to learn more about conditional probability or get practice with solving conditional probability problems, be sure to read the "Conditional Probability" article in the previous edition of Math & CS Research.)

## Generalization with Random Variables

Consider  $n$  friends. The arrival time for each friend  $i \in [n]$  can be represented by a continuous random variable  $X_i$  with  $F_{X_i}(1) = 1$  and  $\forall x < 0, F_{X_i}(x) = 0$ , and the amount of hours they wait before departing can be represented by a continuous random variable  $T_i$  with  $F_{T_i}(1) = 1$  and  $\forall x \leq 0, F_{T_i}(x) = 0$ . (Also, all random variables  $X_i$  and  $T_i$  are independent of each other.) Let  $M$  be the event that  $\exists T \in [0, 1]$  such that every friend is present at time  $T$  (present means the friend has arrived but not yet departed). What is  $P(M)$ ?

### Intuition

When all  $X_i$  were uniform random variables on  $[0, 1]$  (uniform random variables model the process of picking a number uniformly at random from an interval) and  $t_i$  were constants, the answer was

$$\int_0^1 \left( \prod_{j \in [n]} \min(x, t_j) \right) \sum_{j \in [n]} \frac{1}{\min(x, t_j)} dx$$

Recall that  $\min(x, t_j)$  was the probability that friend  $j$  arrived before and within  $t_j$  hours of  $x$ . In more mathematical terms,  $\min(x, t_j) = P(\{\omega \in \Omega | x - t_j < X_j(\omega) \leq x\}) = F_{X_j}(x) - F_{X_j}(x - t_j)$ . (In this case,  $\Omega$  is the set of all events "friend  $j$  arrives at time  $t$ " for  $t \in [0, 1]$ .) This is consistent with the previous answer because all  $X_j$  were uniform random variables on  $[0, 1]$ , so

$$F_{X_j}(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & 1 < x \end{cases}$$

meaning  $\forall x \in [0, 1], F_{X_j}(x) - F_{X_j}(x - t_j) = \min(x, t_j)$ .

Before, we were able to calculate each friends' departure time by adding  $t_j$  to their arrival time. Now,  $T_j$  is a random variable. This motivates us to define random variables  $D_j := X_j + T_j$ , representing the departure time of friend  $j$ .

One interpretation of the  $F_{X_j}(x - t_j)$  term is the probability that friend  $j$  departed before  $x$ . So, it would make sense if our new answer had  $F_{D_j}(x) = F_{X_j+T_j}(x)$  instead. (We can also rewrite  $F_{X_j}(x - t_j)$  as  $F_{X_j+t_j}(x)$ , making this connection even easier to see.)

Furthermore, 1 was the ratio of the probability of friend  $j$  arriving within a small interval  $(x, x + dx)$  to  $dx$ , or  $f_{X_j}(x)$ . (This is also consistent with the previous answer, as the probability of a uniform random variable on  $[0, 1]$  taking on a value within an interval  $I \subset [0, 1]$  is equal to the length of that interval.)

So, it would make sense if the answer was

$$\int_0^1 \prod_{j \in n} (F_{X_j}(x) - F_{X_j+T_j}(x)) \sum_{j \in n} \frac{f_{X_j}(x)}{F_{X_j}(x) - F_{X_j+T_j}(x)} dx$$

## Solution

Let  $\mathbb{I}_M$  be a random variable with  $\mathbb{I}_M(\omega) = 1$  if  $\omega \in M$  and  $\mathbb{I}_M(\omega) = 0$  otherwise. (This is called an indicator random variable. Essentially, it is 1 if  $M$  occurred, and 0 if  $M$  didn't.) Then  $E[\mathbb{I}_M] = 1 \cdot P(M) + 0 \cdot (1 - P(M)) = P(M)$ . (In general, for any indicator random variable  $\mathbb{I}_E$  of an event  $E$ ,  $P(E) = E[\mathbb{I}_E]$ .)

Let  $U$  be a random variable with  $U(\omega) = T$  if  $\omega \in M$ , with  $T$  being the earliest time for which every friend is present, and  $U(\omega) = -1$  otherwise.

By the Law of Iterated Expectation,  $E[\mathbb{I}_M] = E[E[\mathbb{I}_M|U]]$

$$\begin{aligned} E[E[\mathbb{I}_M|U]] &= \int_{-\infty}^{\infty} E[\mathbb{I}_M|U = x] f_U(x) dx \\ &= \int_0^1 f_U(x) dx \end{aligned}$$

since  $\mathbb{I}_M = 1$  iff  $U \in [0, 1]$  and  $\mathbb{I}_M = 0$  otherwise.

For  $x \in [0, 1]$ ,  $U \in [x, x + \delta)$  iff a friend  $i \in [n]$  arrived at a time  $a_i \in [x, x + \delta)$  and  $\forall j \neq i \in [n]$ , friend  $j$  is present at time  $a_i$ . If all other friends are present at the arrival time of friend  $i$  ( $a_i$ ), then friend  $i$  was the last arrival. Since the probability that two friends arrive at exactly the same time is 0, there can only be one last arrival, so the events  $\{a_i \in [x, x + \delta) \cap \forall j \neq i \in [n] \text{ friend } j \text{ is present at } a_i\}$  can be treated as disjoint for different values of  $i$ . Thus,  $P(\{U \in [x, x + \delta)\}) = \sum_{i=1}^n P(\{a_i \in [x, x + \delta) \cap \forall j \neq i \in [n] \text{ friend } j \text{ is present at } a_i\})$ .

Furthermore, since the arrival and departure times of different friends are independent, the events  $\{\text{friend } j \text{ is present at } a_i\}$  are all independent (friend  $j$  being present at a certain time depends only on friend  $j$ 's arrival and departure times), and are also independent from  $\{a_i \in [x, x + \delta)\}$ . So,  $P(\{a_i \in [x, x + \delta) \cap \forall j \neq i \in [n] \text{ friend } j \text{ is present at } a_i\}) = P(\{a_i \in [x, x + \delta)\}) \prod_{j \neq i \in [n]} P(\{\text{friend } j \text{ is present at } a_i\})$ .

The probability that friend  $i$  arrived at a time  $a_i \in [x, x + \delta)$ ,  $P(X_i \in [x, x + \delta))$ , is  $f_{X_i}(x)\delta$  for small  $\delta$  (by the definition of a PDF).

To find  $P(\{\text{friend } j \text{ is present at } a_i\})$ , remember that the definition of friend  $j$  being present is that friend  $j$  has arrived but not yet left. For the sake of notation, let  $A_j$  be the event that friend  $j$  arrived before  $a_i$  ( $\{X_j \leq a_i\}$ ), and let  $B_j$  be the event that friend  $j$  left before  $a_i$  ( $\{D_j \leq a_i\}$ ). Then  $P(\text{friend } j \text{ is present at } a_i) = P(A_j \cap B_j^C)$ . Since  $B_j \subset A_j$  (a friend can't leave before they arrive),  $P(A_j) = P(A_j \cap B_j^C) + P(B_j)$ . So,  $P(A_j \cap B_j^C) = P(A_j) - P(B_j) = P(\{X_j \leq a_i\}) - P(\{D_j \leq a_i\})$ .

Writing  $a_i$  as  $x + k\delta$  for some  $k \in [0, 1)$  and substituting,

$$\begin{aligned}
f_U(x) &= \lim_{\delta \rightarrow 0} \frac{P(\{U \in [x, x + \delta)\})}{\delta} \\
&= \lim_{\delta \rightarrow 0} \frac{\sum_{i=1}^n P(\{a_i \in [x, x + \delta) \cap \forall j \neq i \in [n] \text{ friend } j \text{ is present at } a_i\})}{\delta} \\
&= \lim_{\delta \rightarrow 0} \frac{\sum_{i=1}^n f_{X_i}(x) \delta \prod_{j \neq i \in [n]} P(\{X_j \leq x + k\delta\}) - P(\{D_j \leq x + k\delta\})}{\delta} \\
&= \lim_{\delta \rightarrow 0} \sum_{i=1}^n f_{X_i}(x) \prod_{j \neq i \in [n]} F_{X_j}(x + k\delta) - F_{D_j}(x + k\delta) \\
&= \sum_{i=1}^n f_{X_i}(x) \lim_{\delta \rightarrow 0} \prod_{j \neq i \in [n]} F_{X_j}(x + k\delta) - F_{D_j}(x + k\delta) \\
&= \sum_{i=1}^n f_{X_i}(x) \prod_{j \neq i \in [n]} \lim_{\delta \rightarrow 0} (F_{X_j}(x + k\delta) - F_{D_j}(x + k\delta)) \\
&= \sum_{i=1}^n f_{X_i}(x) \prod_{j \neq i \in [n]} \lim_{\delta \rightarrow 0} F_{X_j}(x + k\delta) - \lim_{\delta \rightarrow 0} F_{D_j}(x + k\delta)
\end{aligned}$$

We were able to bring the limit inside the sum and products because both  $\lim_{\delta \rightarrow 0} F_{X_j}(x + k\delta)$  and  $\lim_{\delta \rightarrow 0} F_{D_j}(x + k\delta)$  exist, so  $\lim_{\delta \rightarrow 0} F_{X_j}(x + k\delta) - F_{D_j}(x + k\delta)$  exists, etc. In fact, since  $F_{X_j}$  and  $F_{D_j}$  are continuous,  $\lim_{\delta \rightarrow 0} F_{X_j}(x + k\delta) = F_{X_j}(x)$  and  $\lim_{\delta \rightarrow 0} F_{D_j}(x + k\delta) = F_{D_j}(x)$ . Thus,

$$\begin{aligned}
f_U(x) &= \sum_{i=1}^n f_{X_i}(x) \prod_{j \neq i \in [n]} F_{X_j}(x) - F_{D_j}(x) \\
&= \prod_{j \in n} (F_{X_j}(x) - F_{D_j}(x)) \sum_{j \in n} \frac{f_{X_j}(x)}{F_{X_j}(x) - F_{D_j}(x)}
\end{aligned}$$

Note: Again, the last step is not necessary, but does make calculations easier. Just be careful if  $\exists i \in [n]$  such that  $F_{X_i}(x) - F_{D_i}(x) = 0$ . In that case, use the expression in the second to last line instead to avoid dividing by 0 (most terms in the sum will be 0 anyway.)

Remembering that  $D_j := X_j + T_j$  and substituting,

$$P(M) = \int_0^1 f_U(x) dx = \boxed{\int_0^1 \prod_{j \in n} (F_{X_j}(x) - F_{X_j+T_j}(x)) \sum_{j \in n} \frac{f_{X_j}(x)}{F_{X_j}(x) - F_{X_j+T_j}(x)} dx}$$

as expected.

## Conclusion

As we further generalized the problem, the answer got more complicated and harder to compute. The answer to the first generalization was an integral that could be expressed as a sum of integrals with polynomial integrands. This is very nice for computation. However, the answer to the second generalization may be very hard to calculate depending on the given functions  $F_{X_i}$  and  $F_{T_i}$ . In some cases, it may be more practical to estimate these integrals with Riemann sums, in which case you could treat  $X_j$  and  $T_j$  as discrete random variables since you would only need to know the value of their CDFs and PDFs over a finite sequence of values. This should raise the question: how much faster and more accurate is it to calculate these integrals, as opposed to running simulations and averaging the results?

Although I have not delved deep into the answer to this question, I would guess that the answer would be this:

For all cases, if the anti-derivative of the integrand can be computed, it is best to do the integral, as this will yield an exact answer. (This will always be the case for the first generalization.)

If not, then see if  $\int_0^y f_{X_j}(x)f_{T_j}(y-x)dx$  can be easily expressed in terms of  $y$  for  $y \in [0, 1]$ , which will give you  $f_{X_j+T_j}(y)$ . If this is the case, then estimating the integral with a Riemann sum of  $m$  intervals will take about as long as simulating the situation  $m$  times, but will be more accurate.

If the answer is still no, then a Riemann sum with  $m$  intervals will take about as long as simulating the situation  $m \log(m)$  times.

These are purely hypotheses and I have not tested these claims, but for  $n$  friends, I would imagine running the simulation  $m$  times would take  $O(mn)$  time, approximating the integral with a Riemann sum with  $m$  intervals when you have  $f_{X_j+T_j}(x)$  in terms of  $x$  would take  $O(mn)$  time, and approximating the integral with a Riemann sum with  $m$  intervals when  $f_{X_j+T_j}(x)$  is not easily expressed in terms of  $x$  would take  $O(m \log(m)n)$  time (since you would need to multiply polynomials of degree  $m$  to find  $f_{X_j+T_j}(x)$  for  $x = \frac{1}{m}, \frac{2}{m}, \dots, \frac{m}{m}$ , which can be done in  $O(m \log(m))$  time.)

I would encourage you to think about how you could construct algorithms for both simulating the situation and estimating the integral to test my hypotheses and possibly prove or disprove them. (To clarify, these are purely conjectures and I have not laid out the details of any such algorithms. These are just my initial guesses.)

In conclusion, even if the result from the second generalization may not always be easy to compute, it was still a good exercise in utilizing properties of random variables, probability distributions, and expectation.

## References

Robert Gallager. 6.262 Discrete Stochastic Processes. Spring 2011. Massachusetts Institute of Technology: MIT OpenCourseWare, [https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-262-discrete-stochastic-processes-spring-2011/course-notes/MIT6\\_262S11\\_chap01.pdf](https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-262-discrete-stochastic-processes-spring-2011/course-notes/MIT6_262S11_chap01.pdf). License: Creative Commons BY-NC-SA.