As you noticed, limits, integrals, and derivatives show up in Lecture 8 and beyond. Some of you have not studied calculus before so you may have a hard time figuring out what is going on. I will not teach you calculus. Instead I will try to give you an idea of what limits, integrals, and derivatives are by examples from probability. If you are familiar with calculus already, you may still find these notes useful as a review of some topics from class.

1 Taking accuracy to the limit

Let's start with the example of raindrops. Suppose that rain is falling at a rate of 1 drop per second. This doesn't mean you will get *exactly* one raindrop in every second. If you observe what happens within, say, a minute of time, divided into 60 seconds, you'd expect a fair number of these seconds to have exactly 1 raindrop, but some of them will have 2, some of them 0, maybe some of them even 3 or more.

We can represent the actual number of raindrops X in any given second by a random variable. Now we want to come up with a reasonable probability model that describes the probability mass function of X. Here is one extremely simple model. Either a raindrop falls within this second (event A_1), or it doesn't (event A_1^c). In this model we get one raindrop in the second with probability $p = \mathrm{P}(A_1)$, and zero raindrops with probability $1 - p$. The expected number of raindrops in the second is p. Since we are already told that rain drops at the rate of 1 per second, we should set $p = 1$. This is not a terribly interesting model: We get the PMF $\mathrm{P}(X = 0) = 0, \mathrm{P}(X = 1) = 1$. It tells us that with certainty we will get exactly one drop in a given second.

Now we'll refine the model like this. Let's divide up the second into two half-seconds and take A_1 be the event that there is a drop in the first half-second interval and A_2 be the event that there is a drop in the second one. We will disregard the possibility that there are two or more drops within the same interval. Since rain drops at a rate of 1 per second, we would expect to get 1/2 of a raindrop in each of our two intervals, so we need to set $\mathrm{P}(A_1) = \mathrm{P}(A_2) = 1/2$.

Now we want to calculate the PMF of the number of raindrops X in this refined probability model. X can take the values 0, 1, and 2. For example, X takes value 0 when neither A_1 nor A_2 happens, so $\mathrm{P}(X = 0) = \mathrm{P}(A_1^c A_2^c)$. To proceed further, we must specify how the events A_1 and A_2 relate to one another. Our model of rainfall makes the assumption that these events are *independent* and then we can write

$$\mathrm{P}(X = 0) = \mathrm{P}(A_1^c A_2^c) = \mathrm{P}(A_1^c)\,\mathrm{P}(A_2^c) = (1 - \mathrm{P}(A_1))(1 - \mathrm{P}(A_2)) = (1 - 1/2)^2 = 0.25.$$

This model seems more accurate than the first one: Now there is some probability that there are zero raindrops within the second. However the assumption that there is at most one raindrop in every half-second interval seems dubious. So let's refine our model further: Instead of splitting the second into two half-second intervals, let's divide it up into 10 intervals of duration 0.1 seconds each. Because these intervals are so much shorter, the assumption that each one of them can handle at most one raindrop makes much more sense. To be consistent with our assumption of one rain drop per second, each one of these intervals should have a probability of 0.1 of getting its raindrop. Assuming these events are independent of one another, we get that the probability of no raindrops within the second is

$$\mathrm{P}(X = 0) = (1 - 1/10)^{10} \approx 0.3487.$$

There is nothing to make us stop at 10 intervals: If we refine our model to 50 intervals of duration 0.02 seconds, we would get

$$P(X = 0) = (1 - 1/50)^{50} \approx 0.3642$$

and if we take 100 intervals of duration 0.01 seconds, we would get

$$P(X = 0) = (1 - 1/100)^{100} \approx 0.3660$$

As we refine our model, the value $P(X = 0)$ does not change that much. If we make a model with n intervals, we would find out that $P(X = 0) = (1 - 1/n)^n$ approaches the value $1/e \approx 0.3679$ as n becomes very large. This is the *limit* of $(1 - 1/n)^n$ as n tends to infinity. In calculus notation we write

$$\lim_{n\to\infty}(1 - 1/n)^n = 1/e \approx 0.3679.$$

There is nothing special about the probability of getting *zero* raindrops. We can repeat the same reasoning for one raindrop: The probability of getting exactly one raindrop in a second in a model with n intervals is $n \cdot (1/n)(1 - 1/n)^{n-1}$. As n becomes large, this also approaches the value $1/e$:

$$\lim_{n\to\infty} n(1/n)(1 - 1/n)^{n-1} = 1/e.$$

For 2 raindrops in a second, the n interval model gives a probability of $\binom{n}{2}(1/n)^2(1 - 1/n)^{n-2}$. As n gets large, this approaches the value $0.1839... = 1/2e$:

$$\lim_{n\to\infty} \binom{n}{2}(1/n)^2(1 - 1/n)^{n-2} = 1/(2e).$$

Continuing in this way, we would find out that for any *fixed* number k, the probability of getting exactly k raindrops in a model with n intervals is $\binom{n}{k}(1/n)^k(1 - 1/n)^{n-k}$. As n becomes large, this gets closer to the value $1/(k!e)$:

$$\lim_{n\to\infty} \binom{n}{k}(1/n)^k(1 - 1/n)^{n-k} = 1/(k!e) \quad \text{for any fixed } k = 0, 1, 2, ...$$

You can see it happening in the following table:

	$P(X = 0)$	$P(X = 1)$	$P(X = 2)$	$P(X = 3)$
1 interval model	0	1	0	0
2 interval model	0.25	0.5	0.25	0
10 interval model	0.3487...	0.3874...	0.1937...	0.0574...
50 interval model	0.3642...	0.3716...	0.1858...	0.0607...
100 interval model	0.3660...	0.3697...	0.1848...	0.0609...
1000 interval model	0.3676...	0.3681...	0.1840...	0.0613...
⋮	⋮	⋮	⋮	⋮
approaches	0.3678... ($1/e$)	0.3678... ($1/e$)	0.1839... ($1/2e$)	0.0613... ($1/6e$)

Instead of worrying whether to work with a 20 interval, 50 interval, or 1000 interval model for raindrops, let's just take the last line of this table as our probability mass function for the number of raindrops within any given second. This is the Poisson(1) random variable.

2 Cumulative distribution functions

Your friend says she'll come to your house between 12 noon and 1pm; but other than that you have no idea when she will come. This looks like a great scenario for a probability model.

To determine when your friend has arrived you look at a clock at the time she enters the door. You have a digital clock, but for some reason the last digit is broken: it always shows zero. So if your friend arrives at 12:41, you will see `12:40` on the clock. If she arrives at 12:47, you will also see `12:40`. If you record what you see on the clock as the arriving time of your friend, there are 6 possibilities: `12:00`, `12:10`, `12:20`, `12:30`, `12:40`, `12:50`. We take these six clock readings as our sample space, and we assume equally likely outcomes.

Let T be the minute part of the registered arrival time. Then T has the following PMF:

t	0	10	20	30	40	50
$\mathrm{P}(T=t)$	1/6	1/6	1/6	1/6	1/6	1/6

You want to know the probability you friend has arrived *at or before* 12:33. This is the value $\mathrm{P}(T \leq 33)$ and you can calculate it from by looking at the PMF:

$$\begin{aligned}
\mathrm{P}(T \leq 33) &= \mathrm{P}(T = 0 \text{ or } T = 10 \text{ or } T = 20 \text{ or } T = 30) \\
&= \mathrm{P}(T = 0) + \mathrm{P}(T = 10) + \mathrm{P}(T = 20) + \mathrm{P}(T = 30) \\
&= 4 \cdot 1/6 \\
&= 4/6 \approx 0.6666.
\end{aligned}$$

This estimate is not very precise. To get a better one we need a better clock. You fix the last digit of the clock and you can now accurately record the minute of the arrival time. We use this clock to make a more accurate probability model. Now there are 60 equally likely readings of the clock, from `12:00` to `12:59`, and the corresponding random variable T has PMF

t	0	1	2	...	58	59
$\mathrm{P}(T=t)$	1/60	1/60	1/60	...	1/60	1/60

Then

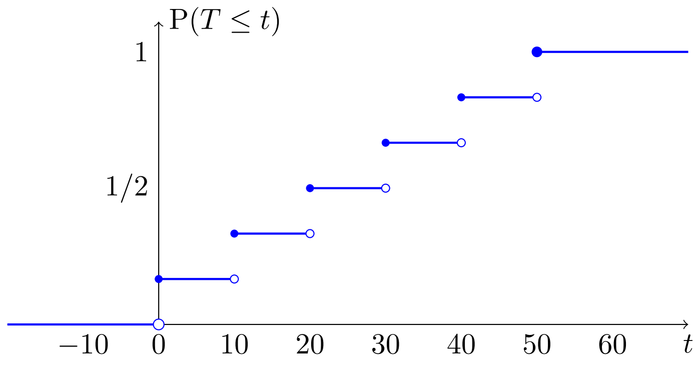
$$\mathrm{P}(T \leq 33) = \mathrm{P}(T = 0) + \mathrm{P}(T = 1) + \cdots + \mathrm{P}(T = 33) = 34 \cdot 1/60 = 34/60 \approx 0.5666.$$

Now we use an even more precise clock that displays not only the minute but also the second of the arrival time. We get an even more accurate probability model, in which the random variable T takes values $0, \frac{1}{60}, \frac{2}{60}$, up to $59\frac{59}{60}$ equally likely (all with probability 1/3600). For this random variable we get that

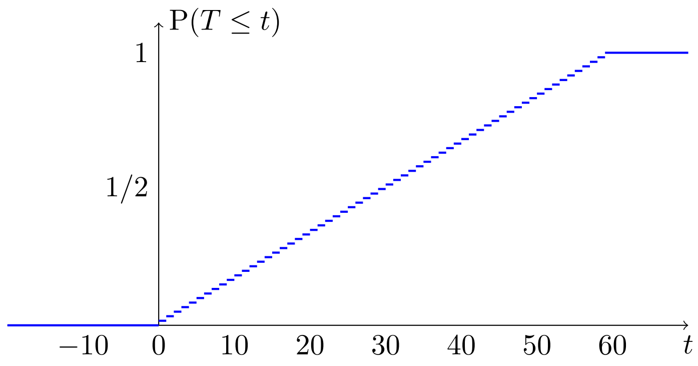
$$\mathrm{P}(T \leq 33) = \mathrm{P}(T = 0) + \mathrm{P}(T = \tfrac{1}{60}) + \cdots + \mathrm{P}(T = 33) = 1981/3600 \approx 0.5503.$$

We can continue refining our probability models by adding more and more precision to the clock. What we would find is that the more precise we make the clock, the value $\mathrm{P}(T \leq 33)$ gets closer and closer to $33/60 = 0.55$.

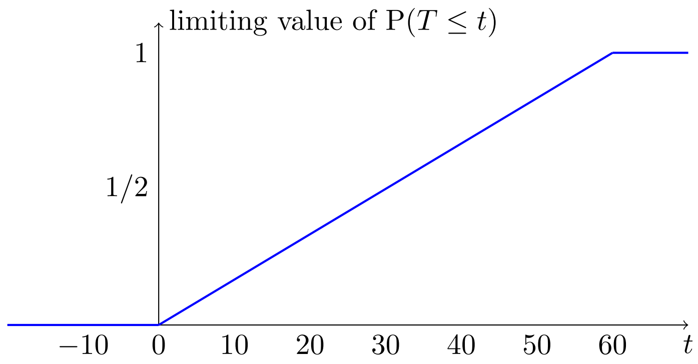
There is nothing special about the value 33. We can do the same for any value t like $-3, 7, 29.773$. Let's again start with the broken clock and ask what is $\mathrm{P}(T \leq t)$. This is the *cumulative distribution function* of t:



In the refined probability model with the more precise clock that shows minutes correctly, we get this:



In the limit as the clock gets more and more precise, for every t, $\mathrm{P}(T \leq t)$ *converges* to the value $t/60$ (for t between 0 and 60:



It would be nice not to worry about precision at all, so we can base our probability model on the last diagram. So we *define* our random variable T (the arrival time of your friend) by the CDF indicated in the last diagram:

$$F(t) = \mathrm{P}(T \leq t) = \begin{cases} 0, & \text{if } t \leq 0 \\ t/60, & \text{if } t \text{ is between 0 and 60} \\ 1, & \text{if } t > 60. \end{cases}$$

This is the Uniform$(0, 60)$ random variable. We can now calculate the probability of any event of the form $T \leq t$, for example

$$\mathrm{P}(T \leq 41.5) = 41.5/60 \approx 0.692.$$

Using the axioms of probability, we can also calculate for example:

$$\mathrm{P}(T > 30) = 1 - \mathrm{P}(T \le 30) = 1 - 30/60 = 1/2$$
$$\mathrm{P}(15 < T \le 30) = \mathrm{P}(T \le 30 \text{ but not } T \le 15) = \mathrm{P}(T \le 30) - \mathrm{P}(T \le 15) = 1/4.$$
$$\mathrm{P}(5 < T \le 10 \text{ or } 15 < T \le 20) = \mathrm{P}(T \le 20) - \mathrm{P}(T \le 15) + \mathrm{P}(T \le 10) - \mathrm{P}(T \le 5) = 1/6.$$

This is a bit different from the way we calculate probabilities for discrete random variables. There we used the formula

$$\mathrm{P}(15 \le T < 30) = \sum\nolimits_{t \text{ between 15 and 30}} \mathrm{P}(T = t).$$

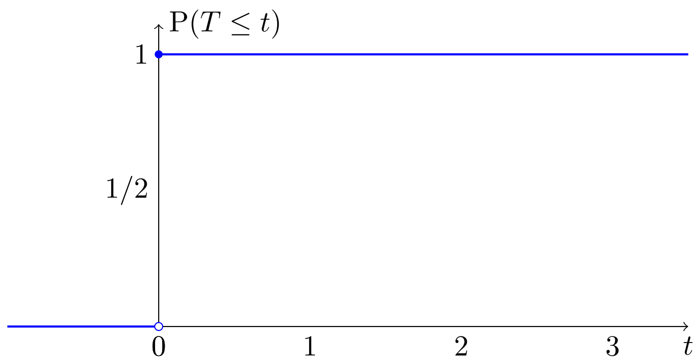
Such a formula doesn't make sense for continuous random variables because $\mathrm{P}(T = t) = 0$ for every t, so we are adding up an infinite number of zeros! To get a sensible answer, we must rely on the axioms of probability.

3 Rainfall again

Let's look again at rain falling at the rate of 1 raindrop per second and ask the following question. It has been raining for some time (at this rate) and you hit the street. What is the time T at which the *first* raindrop hits you? To answer this we'll have to combine the reasoning from the previous two sections.

We start with a very simple model: In each second s of time there are two possibilities: Either a single raindrop arrives (event A_s) or it doesn't arrive (event A_s^c). Moreover, we cannot register the arrival time of the raindrop exactly: For instance if a raindrop arrives at second 0.23764, our crude clock will register that it has arrived at second 0. If it arrives at second 2.4376, the clock will register second 2.

Let p be the probability of each of the events A_s. To get one raindrop per second, we better set $p = 1$: If p were any smaller we would get less than one drop per second on average. But then we can say with certainty that the first raindrop fell within the first second, and the clock will show a reading of 0. So we get that $T = 0$ with probability 1. The CDF of T is not terribly interesting but let's draw it anyway:



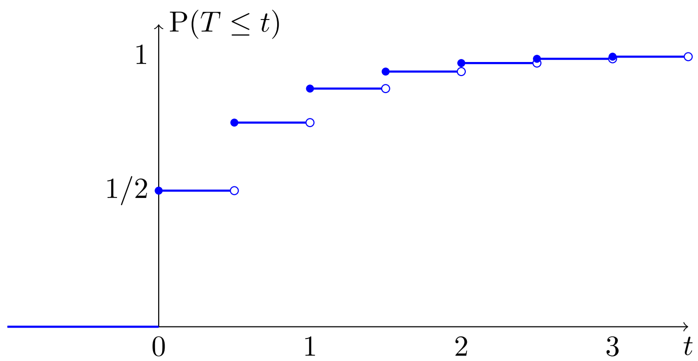
Now let's refine our probability model. For every half-second interval $[0, 1/2), [1/2, 1), [1, 3/2), \ldots$ we have an event A_s that a single raindrop fell within this interval. To have an average of 1 raindrop per second, we set $p = \mathrm{P}(A_s) = 1/2$. We assume that events $A_{[0,1/2)}, A_{[1/2,1)}, \ldots$ are independent. In addition to working with shorter intervals, our clock also becomes more precise: Now it registers the time T of the first raindrop to within a half-second. So T will take values 0, 1/2, 1, 3/2, and so on.

Let's calculate the PMF of T. The event $T = 0$ happens if a raindrop falls in the first interval, so $\mathrm{P}(T = 0) = 1/2$. The event $T = 1$ happens if a raindrop falls in the second interval but not in the first interval, so $\mathrm{P}(T = 1/2) = 1/2 \cdot (1 - 1/2) = 1/4$. Continuing this way, we get $\mathrm{P}(T = 1) = 1/2 \cdot (1 - 1/2)^2 = 1/8$, $\mathrm{P}(T = 3/2) = 1/16$, and so on.

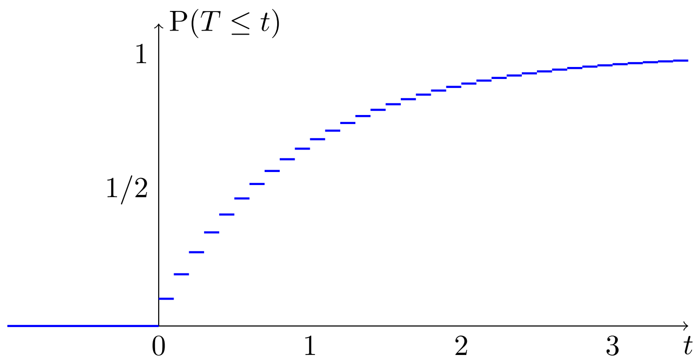
But what we really want is the CDF of T, not the PMF For any specific value of t we can calculate it as usual:

$$\begin{aligned} \mathrm{P}(T \leq 0.3) &= \mathrm{P}(T = 0) = 1/2 \\ \mathrm{P}(T \leq 0.7) &= \mathrm{P}(T = 0) + \mathrm{P}(T = 1/2) = 3/4 \\ \mathrm{P}(T \leq 0.9) &= \mathrm{P}(T = 0) + \mathrm{P}(T = 1/2) = 3/4 \\ \mathrm{P}(T \leq 1.246) &= \mathrm{P}(T = 0) + \mathrm{P}(T = 1/2) + \mathrm{P}(T = 1) = 7/8 \end{aligned}$$

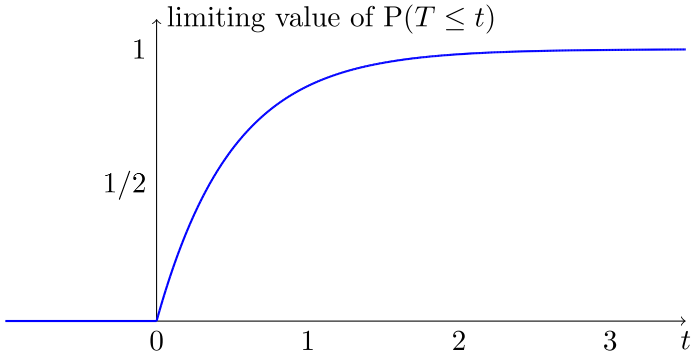
and so on. If we look at what happens with $\mathrm{P}(T \leq t)$ for different values of t, we get the following picture for the CDF of T:



Without doing the calculation, here is the CDF for a probability model with 10 intervals per second and a clock with precision 1/10 of a second:



As we make the model more and more precise by shortening the intervals and improving the precision of the clock, the values $\mathrm{P}(T \leq t)$ become closer and closer to this:



This is the function

$$F(t) = \begin{cases} 1 - e^{-t}, & \text{if } t \geq 0 \\ 0, & \text{if } t < 0 \end{cases}$$

We define the Exponential(1) random variable T to be the random variable with CDF $F(t)$. This is the most accurate probability model for the time of the first raindrop. So if we want to know, for instance, what is the probability that the first raindrop fell after 2 seconds but within 3 seconds from now, we can calculate it from the CDF like this:

$$\mathrm{P}(2 < T \leq 3) = \mathrm{P}(T \leq 3) - \mathrm{P}(T \leq 2) = (1 - e^{-3}) - (1 - e^{-2}) = e^{-2} - e^{-3} \approx 0.0855.$$

4 The probability density function

If we know the CDF of a random variable, we can calculate the probability of any event that depends on its value like in the above examples. What about the expectation? For a discrete random variable X, we had the formula

$$\mathrm{E}[X] = \sum_x x\,\mathrm{P}(X = x)$$

but this formula doesn't make sense in the continuous case because $\mathrm{P}(X = x)$ is always zero.

To make sense of this formula for a continuous random variable, we'll go in the reverse direction from what we did just now: We'll come up with a discrete model for this continuous random variable and see what happens in the limit as the model becomes more and more precise.

Let's look at the Exponential(1) random variable T. This is the arrival time (in seconds) of the first raindrop. Now suppose the clock that records the arrival time of the raindrop is only precise up to the second. (The raindrop itself can arrive at any time; we just don't have a precise record of this time.) If the first raindrop arrives at time 0.523, the clock will read 0. If the raindrop arrives at time 2.889, the clock will read 2. So the clock time C takes only discrete values 0, 1, 2, 3, and so on.

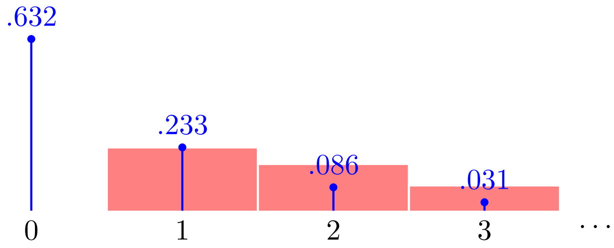
We can calculate the PMF of C: The event $C = 0$ happens when the raindrop arrives within the first second

$$\mathrm{P}(C = 0) = \mathrm{P}(T < 1) = 1 - e^{-1} \approx 0.632$$

and $C = 1$ if the raindrop arrives at some point in the interval $[1, 2)$:

$$\mathrm{P}(C = 1) = \mathrm{P}(1 \leq T < 2) = \mathrm{P}(T < 2) - \mathrm{P}(T < 1) = (1 - e^{-1}) - (1 - e^{-2}) = e^{-1} - e^{-2} \approx 0.233$$

similarly we get that for any value c of the clock, $\mathrm{P}(C = c) = e^{-c} - e^{-(c+1)}$. The *blue* bars in the following picture represent this PMF:

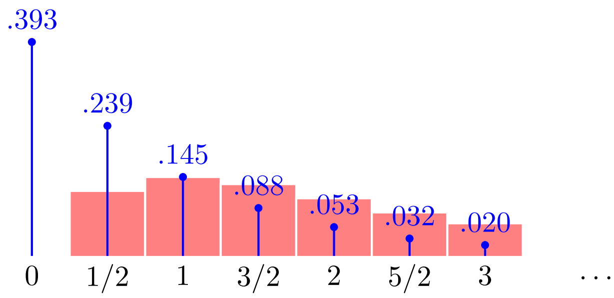


Because C is a discrete random variable, we can calculate its expected value using the formula $\mathrm{E}[C] = \sum c \cdot \mathrm{P}(C = c)$. The *red* bar at position c in the above chart has width 1 and height $c\,\mathrm{P}(C = c)$ (c times the height of the blue bar). So $\mathrm{E}[C]$ equals the area covered by the red bars.

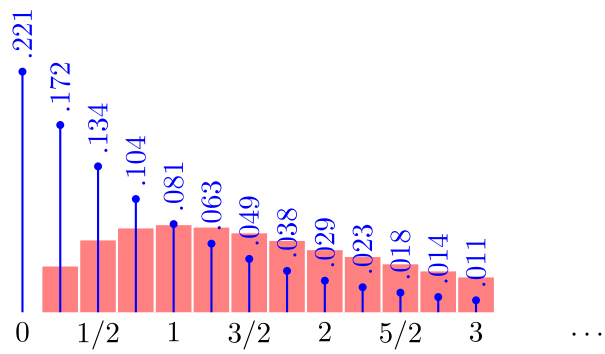
Let's now repeat the same reasoning, but with a more precise clock which records the arrival time of the first raindrop up to a half-second. Now C will take values 0, 1/2, 1, 3/2, and so on. We can calculate the PMF of C

$$\mathrm{P}(C = c) = \mathrm{P}(c \le T < c + 1/2) = (1 - e^{-(c+1/2)}) - (1 - e^{-c}) = e^{-c} - e^{-(c+1/2)}$$

and plot the PMF $\mathrm{P}(C = c)$ in blue, together with the values $c \cdot \mathrm{P}(C = c)$ in red.



Now the expected value of C equals *twice* the area behind the red bars, because each red bar has width 1/2. If the clock is precise up to 1/4 of a second, we get the following picture for $\mathrm{P}(C = c)$ (in blue) and $c\,\mathrm{P}(C = c)$ in red. The expectation of C is now 4 times the area behind the red bars.



As we improve the accuracy of the clock we find the following. For a clock that is accurate up to $1/n$ seconds, the height of each bar is on the order of $1/n$. But to calculate the expectation of C we need to multiply the area of the red bars by a factor of n. Instead of multiplying the height of the bars, let's scale the whole diagram upward by a factor of n. Then the height of the leftmost bar becomes

$$\text{height of blue bar } 0 = n \cdot \mathrm{P}(C = 0) = n \cdot \mathrm{P}(0 \le T < 1/n) = \frac{F(1/n) - F(0)}{1/n}$$

where $F(t) = 1 - e^{-t}$ (for $t \geq 0$) is the CDF of T. Similarly

$$\text{height of blue bar } c = n \cdot \mathrm{P}(C = c) = n \cdot \mathrm{P}(c \leq T < c + 1/n) = \frac{F(c + 1/n) - F(c)}{1/n}$$

As n gets larger, the (scaled) height of blue bar c approaches the *derivative* of the CDF F at c. We write this as

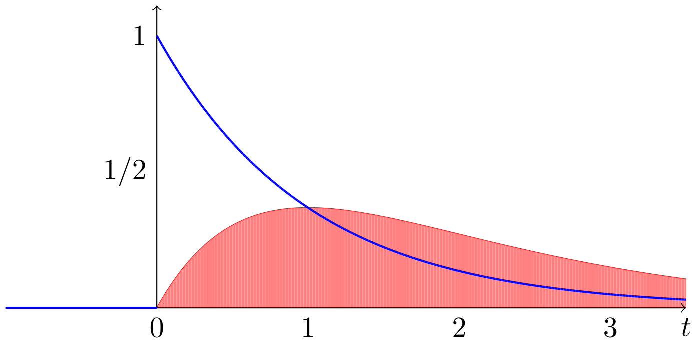
$$\lim_{n\to\infty} \frac{F(c + 1/n) - F(c)}{1/n} = \frac{dF(c)}{dc}.$$

This is the *probability density function* $f(c)$.

As the clock gets more and more precise, the difference between the time measured by the clock and the actual arrival time of the raindrop gets smaller and smaller, so it makes sense that the expected value of C should approach the expected value of T. Recall that the expected value of C is the area under the (scaled) red bars. As precision increases, the tips of the red bars connect up to form the curve $c \cdot f(c)$, so we can obtain E[T] as the area under this curve. This is the *integral* of the function $c \cdot f(c)$, which we write as

$$E[T] = \int_{-\infty}^{\infty} c \cdot f(c) dc.$$

Here is what the scaled bar charts look like in the limit as n gets large.



There are rules in calculus for calculating derivatives. For example, when $F(t) = 1 - e^{-t}$ as in this example, then $dF(c)/dc = e^{-c}$ (for $c > 0$), so we get that the Exponential(1) random variable has PDF

$$f(c) = \begin{cases} e^{-c}, & \text{when } c > 0 \\ 0, & \text{when } c < 0. \end{cases}$$

Similarly, there are rules for calculating integrals which tell us that the area under the red curve ce^{-c} is exactly 1, so $E[T] = 1$.

Exercise: Draw similar charts for the Uniform$(0, 60)$ random variable.

The value of the PDF $f(c)$ at c does not represent the probability of any particular event. This value can sometimes be greater than 1 (although it is not in the example above). So we cannot use it to reason about probabilities of events directly. It tells us, however, something about relative probabilities. For example you can see from the diagram above that for the exponential random variable, $f(0) = 1$ is substantially bigger than $f(1) = 1/e$. What this tells us is that the exponential random variable is more likely to have a value close to zero than a value close to 1. Say we want to compare the probabilities of the events that $0 \leq T < 0.01$ and $1 \leq T < 1.01$. The PDF suggests that $\mathrm{P}(0 \leq T < 0.01)/\,\mathrm{P}(1 \leq T < 1.01) \approx f(0)/f(1) = 1/(1/e) = e$.