Lecture 7: Continuous Random Variables

21 September 2005

1 Our First Continuous Random Variable

The back of the lecture hall is roughly 10 meters across. Suppose it were exactly 10 meters, and consider throwing paper airplanes from the front of the room to the back, and recording how far they land from the left-hand side of the room. This gives us a *continuous* random variable, X, a real number in the interval $[0, 10]$. Suppose further that the person throwing paper airplanes has really terrible aim, and every part of the back wall is as likely as evry other — that X is **uniformly distributed**. All of the random variables we've looked at previously have had discrete sample spaces, whether finite (like the Bernoulli and binomial) or infinite (like the Poisson). X has a continuous sample space, but most things work very similarly.

2 The Cummulative Distribution Function

For instance, we can ask about the probability of events. What is the probability that $0 \leq X \leq 10$, for instance? Clearly, 1. What is the probability that $0 \leq X \leq 5$? Well, since every point is equally likely, and the interval $[0, 5]$ is half of the total interval, $\Pr(0 \leq X \leq 5) = 5/10 = 0.5$. Going through the same reasoning, the probability that X belongs to any interval is just proportional to the length of the interval: $\Pr(a \leq X \leq b) = \frac{b-a}{10}$, if $a \geq 0$ and $b \leq 10$.

Let's plot $\Pr(0 \leq X \leq x)$ versus x.

Now, I only plotted this over the interval $[0, 10]$, but we can extend this plot pretty easily, by looking at $\Pr(X \leq x) = \Pr(-\infty < X \leq x)$. If $x < 0$, this probability is exactly 0, because we know X is non-negative. If $x > 10$, this probability is exactly 1, because we know X is less than ten. And if x is between 0 and 10, then $\Pr(X \leq x) = \Pr(0 \leq X \leq x)$. So then we get this plot

What we have plotted here is the **cummulative distribution function** (CDF) of X. Formally, the CDF of any continuous random variable X is $F(x) = \Pr(X \leq x)$, where x is any real number. When X is uniform over $[0, 10]$, then we have the following CDF:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x}{10} & 0 \leq x \leq 10 \\ 1 & x > 10 \end{cases}$$
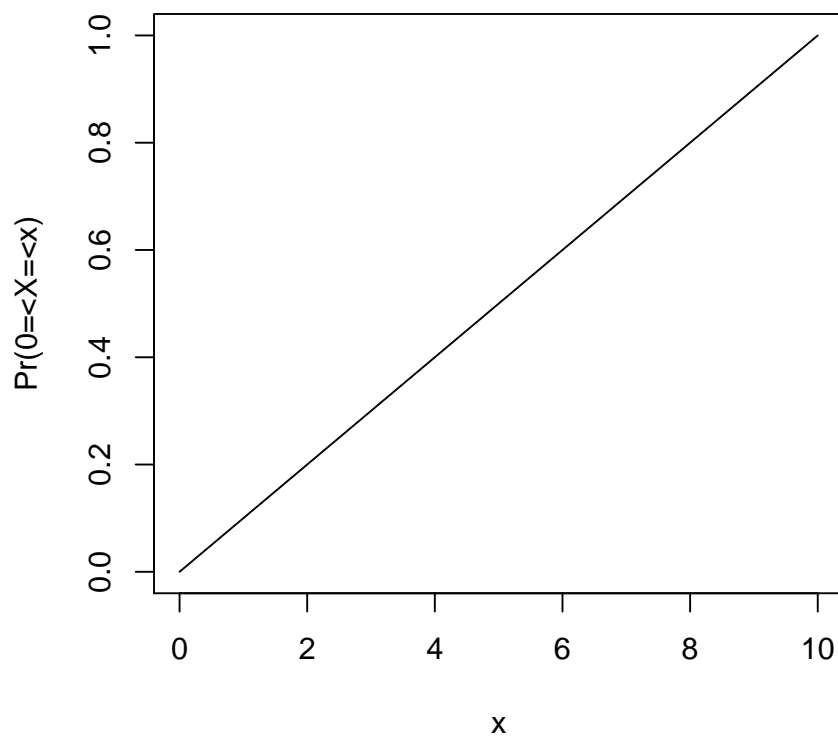
1

Figure 1: Probability that X, uniformly distributed over $[0, 10]$, lies in the interval $[0, x]$.
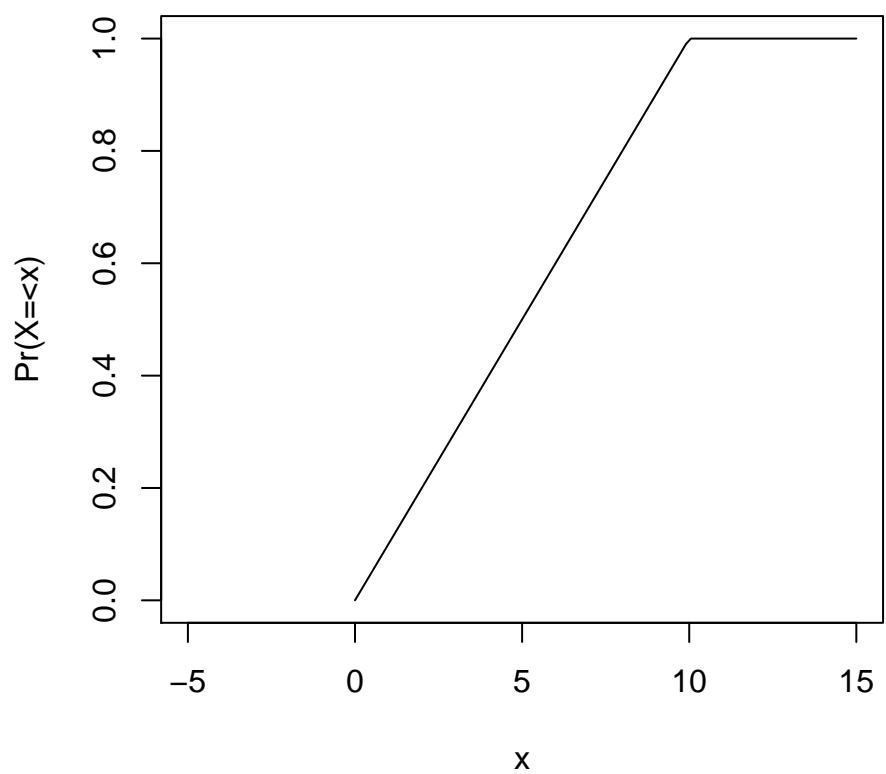
Figure 2: Probability that X, uniformly distributed over $[0, 10]$, is less than or equal to x.

For any CDF, we have the following properties:

1. The CDF is non-negative: $F(x) \geq 0$. Probabilities are never negative.

2. The CDF goes to zero on the far left: $\lim_{x \to -\infty} F(x) = 0$. X is never less than $-\infty$.

3. The CDF goes to one on the far right: $\lim_{x \to \infty} F(x) = 1$. X is never more than ∞.

4. The CDF is non-decreasing: $F(b) \geq F(a)$ if $b \geq a$. If $b \geq a$, then the event $X \leq a$ is a sub-set of the event $X \leq b$, and sub-sets never have higher probabilities. (This was a problem in HW2.)

Any function which satisfies these four properties can be used as the CDF of some random variable or other.

The reason we care about the CDF is that if we know it, we can use it to get the probability of any event we care about. For instance, suppose I tell you the CDF of X, and ask for $\Pr(a \leq X \leq b)$. We can find that from the CDF. Start with $\Pr(X \leq b) = F(b)$. By total probability,

$$\Pr(X \leq b) = \Pr((X \leq b) \cap (X \leq a)) + \Pr((X \leq b) \cap (X \leq a)')$$

Now notice that, since $a < b$, $(X \leq b) \cap (X \leq a) = (X \leq a)$. Also, $(X \leq b) \cap (X \leq a)' = (a \leq X \leq b)$. So

$$
\begin{aligned}
\Pr(X \leq b) &= \Pr(X \leq a) + \Pr(a \leq X \leq b) \\
F(b) &= F(a) + \Pr(a \leq X \leq b) \\
\Pr(a \leq X \leq b) &= F(b) - F(a)
\end{aligned}
$$

For instance, with our uniform-over-$[0, 10]$ variable X, $\Pr(5 \leq X \leq 6) = F(6) - F(5) = 0.6 - 0.5 = 0.1$, as we'd already reasoned, and $\Pr(5 \leq X \leq 12) = F(12) - F(5) = 1 - 0.5 = 0.5$.

Say we want to know $\Pr((0 \leq X \leq 2) \cup (8 \leq X \leq 10))$. Notice that the two intervals are disjoint events, because they don't overlap. The probability of disjoint events adds, and we get the sum of the probabilities of the two intervals, which we know how to do. If the intervals overlap, say in $\Pr((2 \leq X \leq 6) \cup (4 \leq X \leq 8))$, then we use the rule for unions:

$$
\begin{aligned}
\Pr((2 \leq X \leq 6) \cup (4 \leq X \leq 8)) &= \Pr((2 \leq X \leq 6)) + \Pr((4 \leq X \leq 8)) \\
&\quad - \Pr((2 \leq X \leq 6) \cap (4 \leq X \leq 8))
\end{aligned}
$$

Notice that the intersection is just another interval: $(2 \leq X \leq 6) \cap (4 \leq X \leq 8) = (4 \leq X \leq 6)$. This is always true: two intervals are either disjoint, or their intersection is another interval. Using this, it turns out that pretty much any event for real-valued variables which you can describe can be written as a union of dis-joint intervals.[1] Since we can find the probability of any interval from the CDF, we can use the CDF to find the probability of arbitrary events.

---

[1] The technicalities behind the "pretty much" form the subject of measure theory in mathematics. We won't go there.

3 The Probability Density Function

Let's try to figure out what the probability of $X = 5$ is, in our uniform example. We know how to calculate the probability of intervals, so let's try to get it as a limit of intervals around 5.

$$
\begin{aligned}
\Pr\left(4 \leq X \leq 6\right) &= F(6) - F(4) = 0.2 \\
\Pr\left(4.5 \leq X \leq 5.5\right) &= F(5.5) - F(4.5) = 0.1 \\
\Pr\left(4.95 \leq X \leq 5.05\right) &= F(5.05) - F(4.95) = 0.01 \\
\Pr\left(4.995 \leq X \leq 5.005\right) &= F(5.005) - F(4.995) = 0.001
\end{aligned}
$$

Well, you can see where this is going. As we take smaller and smaller intervals around the point $X = 5$, we get a smaller and smaller probability, and clearly in the limit that probability will be exactly 0. (This matches what we'd get from just plugging away with the CDF: $F(5) - F(5) = 0$.) What does this mean? Remember that probabilities are long-run frequencies: $\Pr\left(X = 5\right)$ is the fraction of the time we expect to get the value of *exactly* 5, in infinitely many repetitions of our paper-airplane-throwing experiment. But with a really continuous random variable, we never expect to *repeat* any particular value — we could come close, but there are uncountably many alternatives, all just as likely as $X = 5$, so hits on that point are vanishingly rare.

Talking about the probabilities of particular points isn't very useful, then, for continuous variables, because the probability of any one point is zero. Instead, we talk about the *density* of probability. In physics and chemistry, similarly, when we deal with continuous substances, we use the density rather than the mass at particular points. Mass density is defined as mass per unit volume; let's define probability density the same way. More specifically, let's look the probability of a small interval around 5, say $[5 - \epsilon, 5 + \epsilon]$, and divide that by the length of the interval, so we have probability per unit length.

$$
\frac{\Pr\left(5 - \epsilon \leq X \leq 5 + \epsilon\right)}{2\epsilon} = \frac{F(5 + \epsilon) - F(5 - \epsilon)}{2\epsilon} = \frac{5 + \epsilon - (5 - \epsilon)}{10}\frac{1}{2\epsilon} = \frac{2\epsilon}{(2\epsilon)(10)} = \frac{1}{10}
$$

Clearly, there was nothing special about the point 5 here; we could have used any point in the interval $[0, 10]$, and we'd have gotten the same answer.

Let's formally defined the **probability density function** (pdf) of a random variable X, with cummulative distribution function $F(x)$, as the derivative of the CDF

$$
f(x) = \frac{dF}{dx}
$$

To make this concrete, let's calculate the pdf for our paper-airplane example.

$$
f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{10} & 0 \leq x < 10 \\ 0 & x > 10 \end{cases}
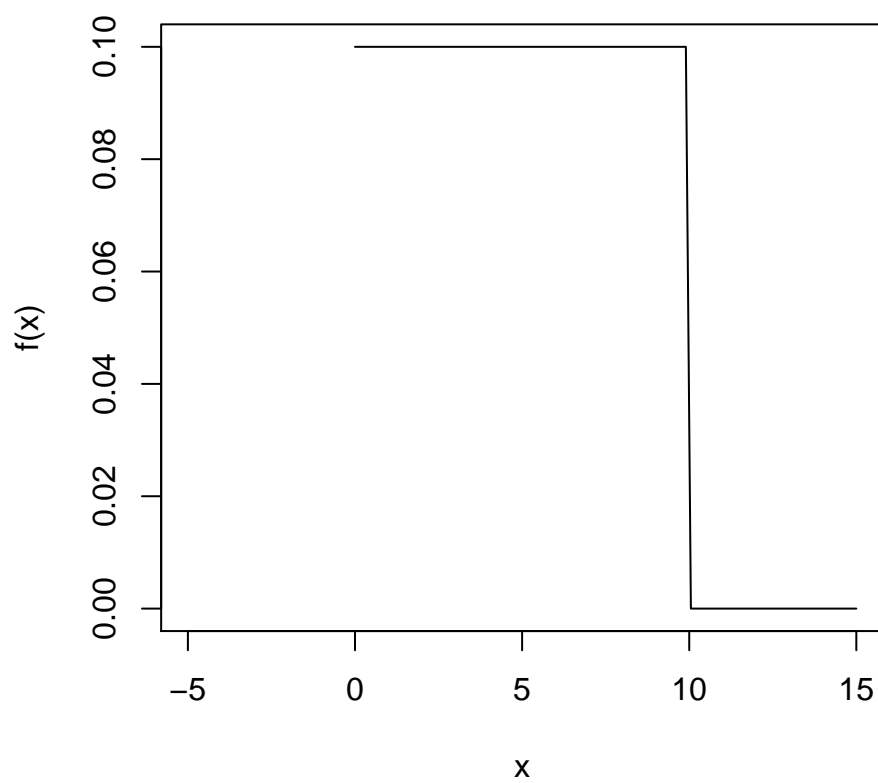$$

Figure 3: Probability density function of a random variable uniformly distributed over $[0, 10]$.

From the fundamental theorem of calculus,

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

So, for any interval $[a, b]$,

$$\Pr(a \leq X \leq b) = \int_{a}^{b} f(x)dx$$

That is, the probability of a region is the area under the curve of the density in that region. For small ϵ,

$$\Pr(x \leq X \leq x + \epsilon) \approx \epsilon f(x)$$

Notice that I write the CDF with an upper-case F, and the pdf with a lower-case f — the density, which is about small regions, gets the small letter.

With discrete variables, we used the probability mass function $p(x)$ to keep track of the probability of individual points. With continuous variables, we'll use the pdf $f(x)$ similarly, to keep track of probability densities. We'll just have to be careful of the fact that it's a probability *density* and not a probability — we'll have to do integrals, and not just sums.

From the definition, and the properties of CDFs, we can deduce some properties of pdfs:

- pdfs are non-negative: $f(x) \geq 0$. CDFs are non-decreasing, so their derivatives are non-negative.

- pdfs go to zero at the far left and the far right: $\lim_{x\to-\infty} f(x) = \lim_{x\to\infty} f(x) = 0$. Because $F(x)$ approaches fixed limits at $\pm\infty$, its derivative has to go to zero.

- pdfs integrate to one: $\int_{-\infty}^{\infty} f(x)dx = 1$. The integral is $F(\infty)$, which has to be 1.

Any function which satisfies these properties can be used as a pdf.

Suppose we have a function g which satisfies the first two properties of a pdf, but $\int_{-\infty}^{\infty} g(x)dx = C \neq 1$. Then we can **normalize** g to get a pdf:

$$f(x) = \frac{g(x)}{C} = \frac{g(x)}{\int_{-\infty}^{\infty} g(x)dx}$$

is a pdf, because it satisfies all three properties.

To see that in action, let's say $g(x) = e^{-\lambda x}$, if $x \geq 0$, and is zer otherwise. Then

$$\int_{-\infty}^{\infty} g(x)dx = \int_{0}^{\infty} e^{-\lambda x}dx = \frac{-1}{\lambda}\left[e^{-\lambda x}\right]_{0}^{\infty} = \frac{1}{\lambda}$$

so

$$f(x) = \frac{g(x)}{1/\lambda} = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$
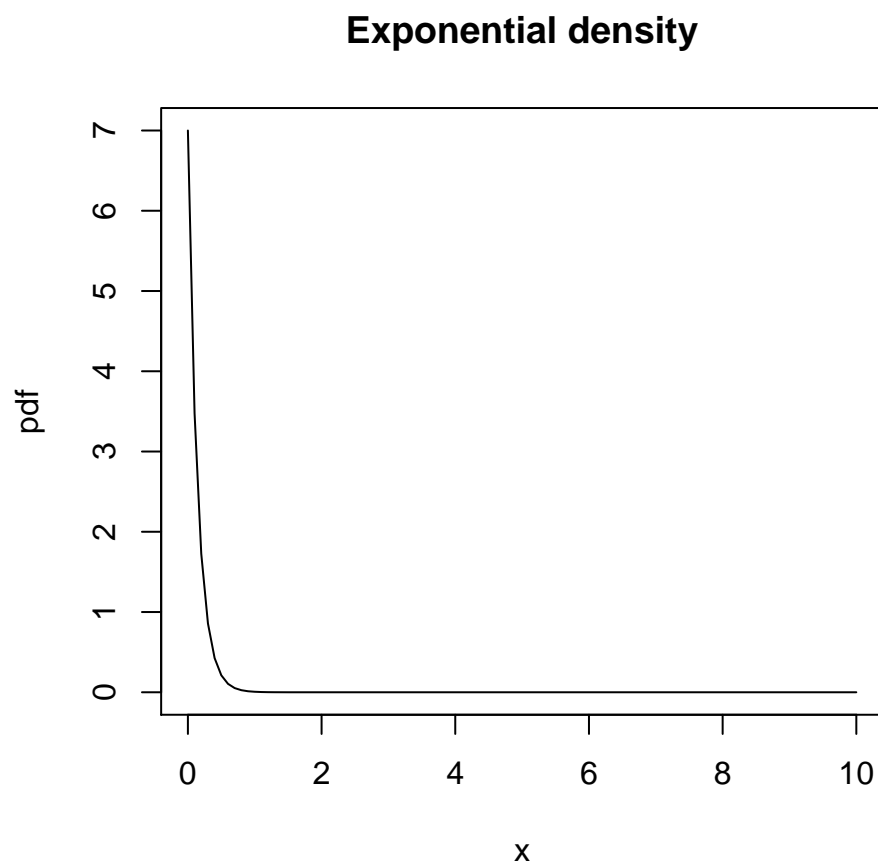
7

Exponential density

Figure 4: Exponential density with $\lambda = 7$.

is a pdf. This is called the **exponential density**, and we'll be using it a lot.
(See figure 4.)

What is the corresponding CDF?

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(y)dy \\
&= \begin{cases} \int_0^x \lambda e^{-\lambda y} dy & x \geq 0 \\ 0 & x < 0 \end{cases} \\
&= \begin{cases} \lambda \frac{-1}{\lambda} [e^{-\lambda y}]_0^x & x \geq 0 \\ 0 & x < 0 \end{cases} \\
&= \begin{cases} -[e^{-\lambda x} - 1] & x \geq 0 \\ 0 & x < 0 \end{cases} \\
&= \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}
\end{aligned}
$$

(See Figure 5.)

Let's check that this is a valid cummulative distribution function:

1. *Non-negative*: If $x < 0$, then $F(x) = 0$, which is non-negative. If $x \geq 0$,
   then $e^{-\lambda x} \leq 1$, so $F(x) = 1 - e^{-\lambda x} \geq 0$.

2. *Goes to 0 on the left*: $F(x) = 0$ if $x < 0$, so

$$\lim_{x \to -\infty} F(x) = 0$$

3. *Goes to 1 on the right*:

$$\lim_{x \to \infty} F(x) = \lim_{x \to \infty} 1 - e^{-\lambda x} = 1 - \lim_{x \to \infty} e^{-\lambda x} = 1$$

4. *Non-decreasing*: If $0 < a < b$, then

$$F(b) - F(a) = 1 - e^{-\lambda b} - (1 - e^{-\lambda a}) = e^{-\lambda a} - e^{-\lambda b} > 0$$

   If $a \leq 0 < b$, then $F(b) > 0$, $F(a) = 0$, and clearly $F(b) > F(a)$.

So the F we obtain by integration makes for a good cummulative distribution
function.

The next figure summarizes the relationships between cummulative distribu-
tion functions, probability density functions, and un-normalized functions like
g.

Expectation

I said that for most purposes the probability density function $f(x)$ of a continu-
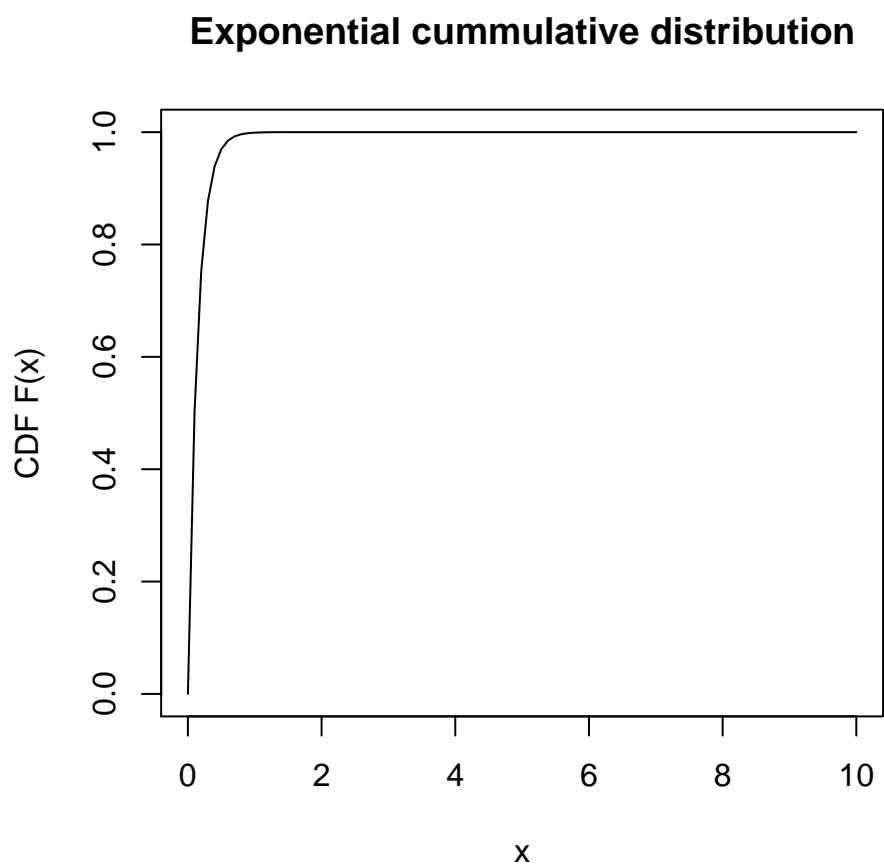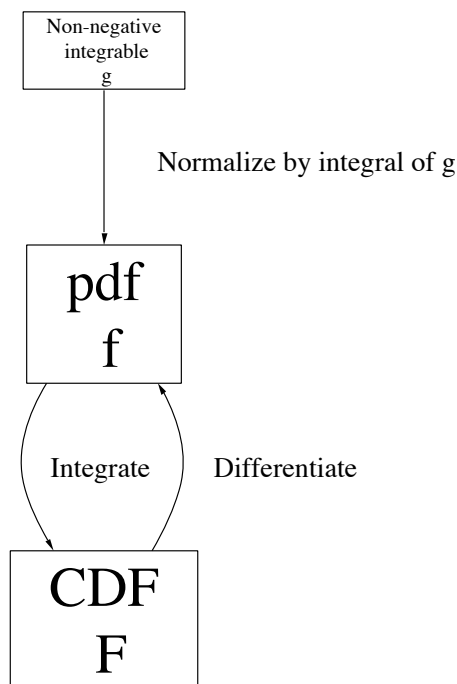ous variable works like the probability mass function $p(x)$ of a discrete variable,

Exponential cummulative distribution

Figure 5: Cummulative distribution function of an exponential random variable with $\lambda = 7$.

```
┌─────────────┐
│ Non-negative │
│  integrable  │
│      g       │
└─────────────┘
```

Normalize by integral of g

```
┌─────────┐
│   pdf   │
│    f    │
└─────────┘
```

Integrate   Differentiate

```
┌─────────┐
│   CDF   │
│    F    │
└─────────┘
```

and one place that's true is when it comes to defining expectations. Remember that for discrete variables

$$\mathbf{E}\left[X\right] \equiv \sum_x x p(x)$$

For a continuous variable, we just substitute $f(x)$ for $p(x)$ and an integral for a sum:

$$\mathbf{E}\left[X\right] \equiv \int_{-\infty}^{\infty} x f(x) dx$$

All of the rules which we learned for discrete expectations still hold for continuous expectations.

Let's see how this works for the uniform-over-$[0, 10]$ example.

$$\mathbf{E}\left[X\right] = \int_{-\infty}^{\infty} x f(x) dx = \int_{0}^{10} x \frac{1}{10} dx = \frac{1}{10}\frac{1}{2}\left[x^2\right]_0^{10} = \frac{1}{10}\frac{1}{2}(100 - 0) = 5$$

Notice that 5 is the mid-point of the interval $[0, 10]$. Suppose we had a uniform distribution over another interval, say (to be imaginative) $[a, b]$. What would the expectation be? First, find the CDF $F(x)$, from the same kind of reasoning we used on the interval $[0, 10]$: the probability of an interval is its length, divided by the total length. Then, find the pdf, $f(x) = dF/dx$; finally, get the expectation,

11

by integrating $xf(x)$.

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & b < x \end{cases}$$

$$f(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \le x \le b \\ 0 & x > b \end{cases}$$

$$\mathbf{E}\left[X\right] = \int_a^b x \frac{1}{b-a} dx$$

$$= \frac{1}{b-a} \frac{1}{2} \left[x^2\right]_a^b$$

$$= \frac{1}{b-a} \frac{b^2 - a^2}{2}$$

$$= \frac{1}{b-a} \frac{(b-a)(b+a)}{2}$$

$$= \frac{b+a}{2}$$

In words, the expectation of a uniform distribution is always the mid-point of its range.

4   Uniform Random Variables in General

We've already seen most of the important results for uniform random variables, but we might as well collect them all in one place. The uniform distribution on the interval $[a, b]$, written $U(a, b)$ or $Uni(a, b)$, has two parameters, namely the end-points of the distribution. The CDF is $F(x) = fracx - ab - a$ inside the interval, and the pdf $f(x) = \frac{1}{b-a}$. The expectation $\mathbf{E}\left[X\right] = \frac{b+a}{2}$, the mid-point of the interval, and the variance $\mathrm{Var}\left(X\right) = \frac{(b-a)^2}{12}$.

Notice that if $X \sim U(a, b)$, then $cX + d \sim U(ca + d, cb + d)$.

4.1   Turning Non-Uniform Random Variables into Uniform Ones

Suppose X is a non-uniform random variable with CDF F. Then $F(X)$ is again a random variable, but it's always uniform on the interval $[0, 1]$ — because $F(X) \le 0.5$ exactly half the time, $F(X) \le 0.25$ exactly a quarter of the time, and so on. This is one way to test whether you've guessed the right distribution for a non-uniform variable: if you think the CDF is F, then calculate $F(x_i)$ for all the data points you have, and the result should be very close to uniform on the unit interval. (We will come back to this idea later.)

4.2 Turning Uniform Random Variables into Non-Uniform Ones

Computer random number generators typically produce random numbers uniformly distributed between 0 and 1. If we want to simulate a non-uniform random variable, and we know its distribution function, then we can reverse the trick in the last section. If $Y \sim U(0,1)$, then $F^{-1}(Y)$ has cummulative distribution function F. We can think of this as follows: our random number generator gives us a number y between 0 and 1. We then slide along the domain of the CDF F, starting at the left and working to the right, until we come to a number x where $F(x) = y$. Because F is non-decreasing, we'll never have to back-track. This x is $F^{-1}(y)$, by definition. Because, again, F is non-decreasing, we know that $F^{-1}(a) \leq F^{-1}(b)$ if and only if $a \leq b$. If we pick $a = y$ and $b = F(x)$, then this gives us that $F^{-1}(y) \leq x$ if and only if $y \leq F(x)$. So if $0 \leq x \leq 1$,

$$\Pr\left(F^{-1}(Y) < x\right) = \Pr\left(Y < F(x)\right) = F(x)$$

because, remember, Y is uniformly distributed between 0 and 1.

Using this trick assumes that we can compute F^{-1}, the inverse of the CDF. This isn't always available in closed form.

5 Exponential Random Variables

Exponential random variables generally arise from decay processes. The idea is that things — radioactive atoms, unstable molecules, etc. — have a certain probability per unit time of decaying, and we ask how long we have to wait before seeing the decay. (The model applies to any kind of event with a constant rate over time.) They also arise in statistical mechanics and physical chemistry, where the probability that a molecule has energy E is proportional to $e^{-E/k_B T}$, where T is the absolute temperature and k_B is Boltzmann's constant.

There's a constant probability of decay per unit time, call it (with malice aforethought) λ. So let's ask what the probability is that the decay happened by time t, i.e., $\Pr\left(0 \leq T \leq t\right)$. We can imagine dividing that interval up into n equal parts. The probability of decay in each time-interval then is $\lambda t/n$. The probability of *no* decay is

$$\approx \left(1 - \frac{\lambda t}{n}\right)^n$$

because we can't have had a decay event in any of the n sub-intervals. So

$$\Pr\left(0 \leq T \leq t\right) \approx 1 - \left(1 - \frac{\lambda t}{n}\right)^n$$

Since time is continuous, we should really take the limit as $n \to \infty$:

$$\Pr\left(0 \leq T \leq t\right) = \lim_{n \to \infty} 1 - \left(1 - \frac{\lambda t}{n}\right)^n$$

$$= 1 - \lim_{n \to \infty} \left(1 - \frac{\lambda t}{n} \right)^n$$

$$= 1 - e^{-\lambda t}$$

which is the CDF of the exponential distribution.

The expectation of an exponential variable is $\mathbf{E}[X] = 1/\lambda$; its variance, $\mathbf{E}[X] = 1/\lambda^2$. I leave showing both of these facts as exercises; there are several ways to do it, including integration by parts.

6 Discrete vs. Continuous Variables

Discrete	Continuous
p.m.f. $p(x) \equiv \Pr(X = x)$	p.d.f. $f(x) \equiv \frac{dF}{dx} \approx \frac{\Pr(x \leq X \leq x+\epsilon)}{\epsilon}$
$\sum (\ldots) p(x)$	$\int (\ldots) f(x) dx$
CDF $F(x) \equiv \Pr(X \leq x) = \sum_{y=-\infty}^{y=x} p(y)$	CDF $F(x) \equiv \Pr(X \leq x) = \int_{-\infty}^{x} f(y) dy$
$\mathbf{E}[X] \equiv \sum_x x p(x)$	$\mathbf{E}[X] \equiv \int_{-\infty}^{\infty} x f(x) dx$