Chapter 4. Continuous Random Variables

4.1: Continuous Random Variables Basics

Slides (Google Drive) Alex Tsun Video (YouTube)

Up to this point, we have only been talking about *discrete* random variables - ones that only take values in a countable (finite or countably infinite) set like the integers or a subset. What if we wanted to model quantities that were continuous - that could take on *uncountably infinitely* many values? If you haven't studied or seen cardinality (or types of infinities) before, you can think of this as being intervals of the real line, which take decimal values. Our tools from the previous chapter were not suitable to modelling these situations, and so we need a new type of random variable.

Definition 4.1.1: Continuous Random Variables

A **continuous random variable** is a random variable that takes values from an uncountably infinite set, such as the set of real numbers or an interval. For e.g., height (5.6312435 feet, 6.1123 feet, etc.), weight (121.33567 lbs, 153.4642 lbs, etc.) and time (2.5644 seconds, 9321.23403 seconds, etc.) are continuous random variables that take on values in a continuum.

Why do we need continuous random variables?

Suppose we want a random number in the interval [0, 10], with each possibility being "equally likely".

- What is $\mathbb{P}(X = 3.141592)$ for such a random variable X? That is, if I chose a random decimal number (with infinite precision/decimal places), what is the probability you guess it exactly right (matching infinitely many decimal places)? The probability is actually 0, it's not even a tiny positive number!

- What is $\mathbb{P}(5 \leq X \leq 8)$ for such a random variable X? That is, what if you were allowed to guess a range instead of a single number? As you might expect, $\frac{\text{size of the required interval}}{\text{size of the total interval}} = \frac{3}{10}$ since the random number is uniformly distributed.

Suppose we want to study the set of possible heights (in feet) a person can have, supposing that the range of possible heights is the interval $[1, 8]$.

- What is the probability that someone has a height of 5.2311333 feet? This is again 0, since you have to be exactly precise!

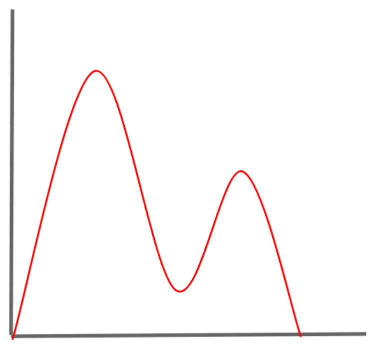
- What is the probability that someone has a height between 5 and 6 feet? This is non-zero, since we are studying an interval. It isn't necessarily $\frac{6-5}{8-1} = \frac{1}{7}$ though since heights aren't necessarily uniformly distributed! More people will have heights in the interval $[4, 6]$ feet than say $[1, 3]$ feet.

Notice, that since these values can have infinite precision, the probability that a variable has a specific value is 0, in contrast to discrete random variables.

4.1.1 Probability Density Functions (PDFs)

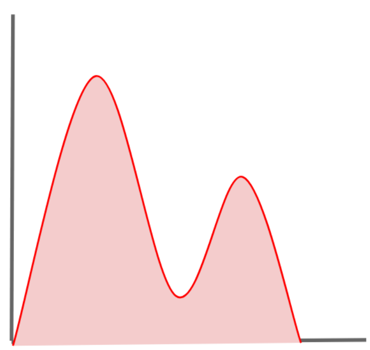
Every continuous random variable has a probability density function (PDF), *instead* of a probability mass function (PMF), that defines the relative likelihood that a random variable X has a particular value. Why do we need this new construct? We already said that $\mathbb{P}(X = a) = 0$ for any value of a, and so a "PMF" for a continuous random variable would equal 0 for any input and be useless. It wouldn't satisfy the constraint that the sum of the probabilities is 1 (assuming we could even sum over uncountably many values; we can't). Instead, we have the idea of a probability density function where the x-axis has values in the random variable's range (usually an interval), and the y-axis has the probability *density* (not mass), which is explained below.

A PDF may look something like this:

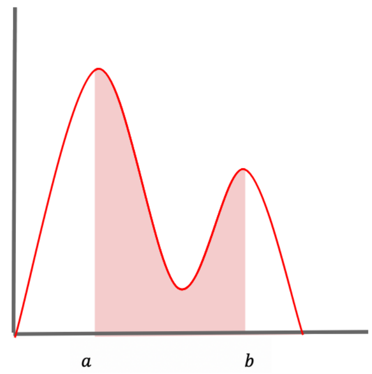


The probability density function f_X has some characteristic properties (denoted with f_X to distinguish from PMFs p_X). Notice again I will use different dummy variables inside the function like $f_X(z)$ or $f_X(t)$ to ensure you get the idea that the density is f_X (subscript indicates for rv X) and the dummy variable can be anything.

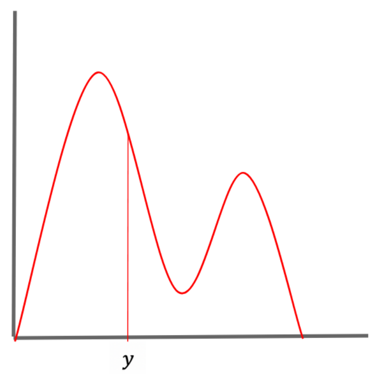
- $f_X(z) \geq 0$ for all $z \in \mathbb{R}$; i.e., it is always non-negative, just like a probability mass function.
- $\int_{-\infty}^{\infty} f_X(t)dt = 1$; i.e., the area under the entire curve is equal to 1, just like the sum of all the probabilities of a discrete random variable equals 1.



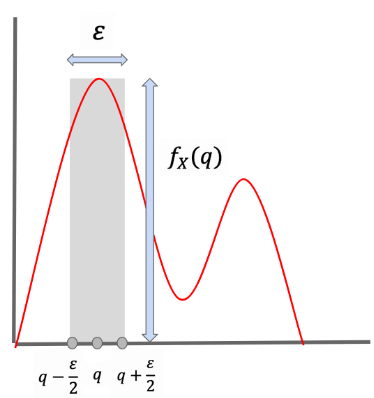
- $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(w)dw$; i.e., the probability that X lies in the interval a to b is the area under the curve from a to b. This is key - **integrating** f_X gives us **probabilities**.



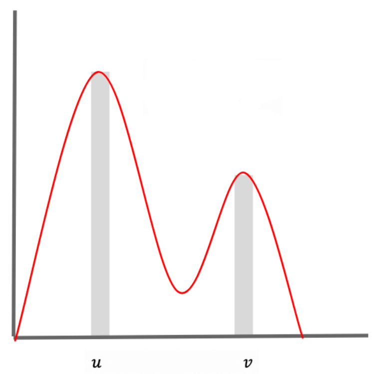
- $\mathbb{P}(X = y) = \mathbb{P}(y \le X \le y) = \int_y^y f_X(w)dw = 0$. The probability of being a particular value is 0, and NOT equal to the density $f_X(y)$ which is nonzero. This is particularly confusing at first.



- $\mathbb{P}(X \approx q) = \mathbb{P}\left(q - \frac{\varepsilon}{2} \le X \le q + \frac{\varepsilon}{2}\right) \approx \varepsilon f_X(q)$; i.e., with a small epsilon value, we can obtain a good rectangle approximation of the area under the curve. The width of the rectangle is ε (from the difference between $q + \frac{\varepsilon}{2}$ and $q - \frac{\varepsilon}{2}$). The height of the rectangle is $f_X(q)$, the value of the probability density function f_X at q. So, the area of the rectangle is $\varepsilon f_X(q)$. This is similar to the idea of Riemann integration.



- $\frac{\mathbb{P}(X \approx u)}{\mathbb{P}(X \approx v)} \approx \frac{\varepsilon f_X(u)}{\varepsilon f_X(v)} = \frac{f_X(u)}{f_X(v)}$; i.e., the PDF tells us ratios of probabilities of being "near" a point. From the previous point, we know the probabilities of X being approximately u and v, and through algebra, we see their ratios. Since the density is twice as high at u as it is at v, it means we are twice as likely to get a point "near" u as we are to get one "near" v.



Definition 4.1.2: Probability Density Function (PDF)

Let X be a continuous random variable (one whose range is typically an interval or union of intervals). The probability density function (PDF) of X is the function $f_X : \mathbb{R} \to \mathbb{R}$, such that the following properties hold:

- $f_X(z) \geq 0$ for all $z \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f_X(t)\, dt = 1$
- $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(w)\, dw$
- $\mathbb{P}(X = y) = 0$ for any $y \in \mathbb{R}$
- The probability that X is close to q is proportional to its density $f_X(q)$;

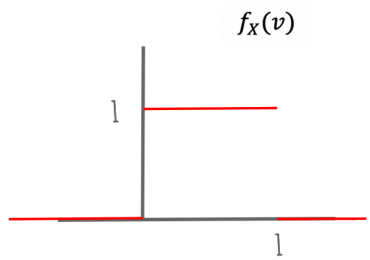
$$\mathbb{P}(X \approx q) = \mathbb{P}\left(q - \frac{\varepsilon}{2} \leq X \leq q + \frac{\varepsilon}{2}\right) \approx \varepsilon f_X(q)$$

- Ratios of probabilities of being "near points" are maintained;

$$\frac{\mathbb{P}(X \approx u)}{\mathbb{P}(X \approx v)} \approx \frac{\varepsilon f_X(u)}{\varepsilon f_X(v)} = \frac{f_X(u)}{f_X(v)}$$

4.1.2 Cumulative Distribution Functions (CDFs)

Here is the density function of a "uniform" random variable on the interval $[0, 1]$:



$$f_X(v) = \begin{cases} 1, & 0 \leq v \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

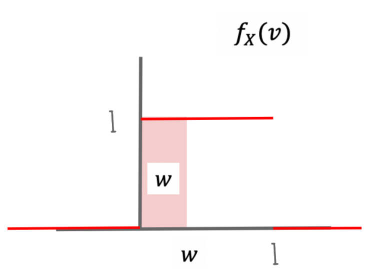
We know this is valid, because the area under the curve is the area of a square with side lengths 1, which is $1 \cdot 1 = 1$.

We define the cumulative distribution function (CDF) of X to be $F_X(w) = \mathbb{P}(X \leq w)$. That is, the all the area to the left of w in the density function. Note we also have CDFs for discrete random variables, they are defined exactly the same way (the probability of being less than or equal to a certain value)! They just don't usually have a nice closed form like they do for continuous RVs. Note for continuous random variables, the CDF at w is just the cumulative area to the left of w, which can be found by an integral (the dummy variable of integration should be different than the input variable w)

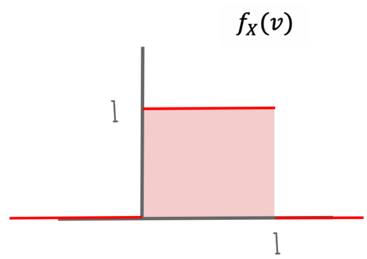
$$F_X(w) = \mathbb{P}(X \leq w) = \int_{-\infty}^{w} f_X(y)dy$$

Let's try to compute the CDF of this uniform random variable on $[0, 1]$. There are three cases to consider here.

- If $w < 0$, $F_X(w) = 0$ since $\Omega_X = [0, 1]$. For example, if $w = -1$, then $F_X(w) = \mathbb{P}(X \leq -1) = 0$ since there is no chance that $X \leq -1$. Formally, there is also no area to the left of $w = -1$ as you can see from the PDF above, so the integral evaluates to 0!
- If $0 \leq w \leq 1$, the area up to w is a rectangle of height 1 and width w (see below), so $F_X(w) = w$. That is, $\mathbb{P}(X \leq w) = w$. For example, if $w = 0.5$, then the probability $X \leq 0.5$ is actually just 0.5 since X is just equally likely to be anywhere in $\Omega_X = [0, 1]$! Note here we didn't do an integral since there are nice shapes, and we sometimes don't have to! We just looked at the area to the left of w.



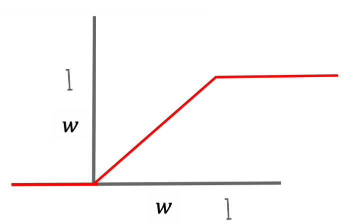
- If $w > 1$, all the area is up to the left of w, so $F_X(w) = 1$. Again, since $\Omega_X = [0, 1]$ and suppose $w = 2$, then $F_X(w) = \mathbb{P}(X \leq 2) = 1$ since X is always between 0 and 1 (X must be less than or equal to 2). Formally, the cumulative area to the left of $w = 2$ is 1 (just the area of the square)!



We can put these conclusions together to show:

$$F_X(w) = \begin{cases} 0 & \text{if } w < 0 \\ w & \text{if } 0 \leq w \leq 1 \\ 1 & \text{if } w > 0 \end{cases}$$

On a graph, $F_X(w)$ looks like this:



The cumulative distribution function has some characteristic properties:

- $F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^{t} f_X(w)dw$ for all $t \in \mathbb{R}$ - i.e. the probability that $X \leq t$ is the area to the left of t of the density curve.
- You have a function that is defined to be the integral up to a point of a function, so by the Fundamental Theorem of Calculus, the derivative of the CDF is actually the PDF - i.e. $\frac{d}{du}F_X(u) = f_X(u)$. **This is probably the most important observation** that explains the relationship between PDF and CDF.
- The probability that X is between a and b is the probability that $X \leq b$ minus the probability that $X \leq a$; i.e., $\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a)$. For those with an eagle eye, you might have noticed I lied a little - it should be $\mathbb{P}(a < X \leq b)$. But since $\mathbb{P}(X = a) = 0$, it doesn't matter!
- F_X is always monotone increasing because we are integrating a non-negative function ($f_X \geq 0$). That is, if $c \leq d$, then $F_X(c) \leq F_X(d)$. For example, if $c = 2$ and $d = 5$, then $\mathbb{P}(X \leq 2) \leq \mathbb{P}(X \leq 5)$ because $X \leq 2$ implies that $X \leq 5$ automatically.
- As $v \to -\infty$, the CDF at v is the probability that X is less than negative infinity which is 0; so the left-hand limit is 0, i.e. $\lim_{v \to -\infty} F_X(v) = \mathbb{P}(X \leq -\infty) = 0$.
- With similar logic to the previous point, $\lim_{v \to +\infty} F_X(v) = \mathbb{P}(X \leq +\infty) = 1$.

Definition 4.1.3: Cumulative Distribution Function (CDF)

Let X be a continuous random variable (one whose range is typically an interval or union of intervals). The cumulative distribution function (CDF) of X is the function $F_X : \mathbb{R} \to \mathbb{R}$ such that:

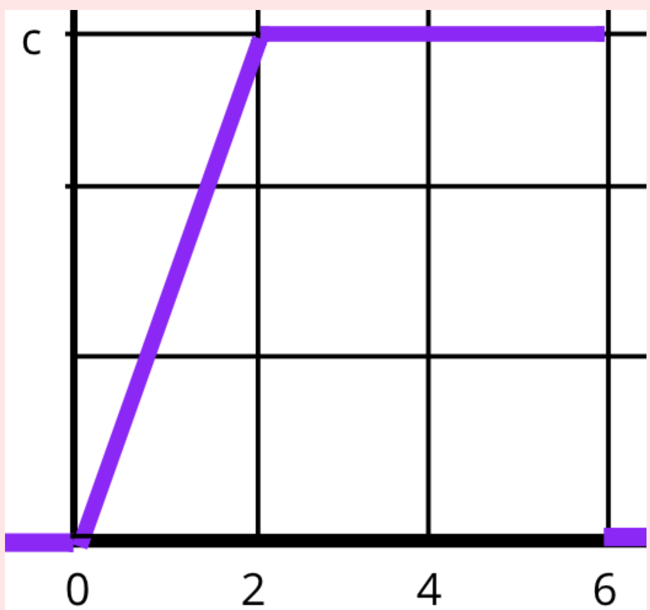
- $F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^{t} f_X(w)\, dw$ for all $t \in \mathbb{R}$
- $\frac{d}{du} F_X(u) = f_X(u)$
- $\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a)$
- F_X is monotone increasing, since $f_X \geq 0$. That is, $F_X(c) \leq F_X(d)$ for $c \leq d$.
- $\lim_{v \to -\infty} F_X(v) = \mathbb{P}(X \leq -\infty) = 0$
- $\lim_{v \to +\infty} F_X(v) = \mathbb{P}(X \leq +\infty) = 1$

Example(s)

Suppose the number of hours that a package gets delivered past noon is modelled by the following PDF:

$$f_X(x) = \begin{cases} x/10 & 0 \leq x \leq 2 \\ c & 2 < x \leq 6 \\ 0 & \text{otherwise} \end{cases}$$

Here is a graph of the PDF as described above:



1. What is the range Ω_X?
2. What is the value of c that makes f_X a valid density function?
3. Find the cumulative distribution function (CDF) of X, $F_X(x)$, and make sure to define it piecewise for any real number x.

4. What is the probability that the delivery arrives between 2pm and 6pm?
5. What is the expected time that the package arrives at?

Solution

1. The range is all values where the density is nonzero; in our case, that is $\Omega_X = [0, 6]$ (or $(0, 6)$), but we don't care about single points or endpoints because the probability of being exactly that value is 0.

2. Formally, we need the density function to integrate to 1; that is,

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

But, the density function is split into three parts, we can split our integral into three. However, anywhere the density is zero, we will get an integral of zero, so we'll only set up the two integrals that are nontrivial:

$$\int_0^2 x/10dx + \int_2^6 cdx = 1$$

Solving this equation for c would definitely work. But let's try to use geometry instead, as we do know how to compute the area of a triangle and rectangle. So the left integral is the area of the triangle with base from 0 to 2 and height c, so that area is $2c/2 = c$ (the area of a triangle is $b \cdot h/2$). The area of the rectangle with base from 2 to 6 is $4c$. We need the total area of $c + 4c = 1$, so $c = 1/5$.

3. Our CDF needs four cases: when $x < 0$, when $0 \leq x \leq 2$, when $2 < x \leq 6$, and when $x > 6$.

(a) The outer cases are usually the easiest ones: if $x < 0$, then $F_X(x) = \mathbb{P}(X \leq x) = 0$ since X cannot be less than zero.

(b) If $x > 6$, then $F_X(x) = \mathbb{P}(X \leq x) = 1$ since X is guaranteed to be at most 6.

(c) For $0 \leq x \leq 2$, we need the cumulative area to the left of x, which happens to be a triangle with base x and height $x/10$, so the area is $x^2/20$. Alternatively, evaluate the integral

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_0^x t/10dt = t^2/20$$

(d) For $2 < x \leq 6$, we have the entire triangle of area $2 \cdot 1/5 \cdot 0.5 = 1/5$, but also a rectangle of base $x - 2$ and height $1/5$, for a total area of $1/5 + 1/5(x - 2) = x/5 - 1/5$. Alternatively, the integral would be

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_0^2 t/10dt + \int_2^x 1/5dt = x/5 - 1/5$$

Again, I skipped all the integral evaluation steps as they are purely computational, but feel free to verify!

Finally, putting this together gives

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x^2/20 & 0 \leq x \leq 2 \\ x/5 - 1/5 & 2 < x \leq 6 \\ 1 & x > 6 \end{cases}$$

4. Using the formula, we find the area between 2 and 6 to get $\mathbb{P}(2 \leq X \leq 6) = \int_2^6 f_X(t)dt = \int_2^6 1/5dt = 4/5$. Alternatively, we can just see the area from 2 to 6 is just a rectangle with base 4 and height 1/5, so the probability is just 4/5.

   We could also use the CDF we so painstakingly computed.

   $$\mathbb{P}(2 \leq X \leq 6) = F_X(6) - F_X(2) = (6/5 - 1/5) - (2^2/20) = 1 - 1/5 = 4/5$$

   This is just the area to the left of 6, minus the area to the left of 2, which gives us the area between 2 and 6.

5. We'll use the formula for expectation of a continuous RV, but split into three integrals again due to the piecewise definition of our density. However, the integral outside the range $[0, 6]$ will evaluate to zero, so we won't include it.

   $$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf_X(x)dx = \int_0^2 xf_X(x)dx + \int_2^6 xf_X(x)dx = \int_0^2 x \cdot (x/10)dx + \int_2^6 x \cdot (1/5)dx$$

   We won't do the computation because it's not important, but hopefully you get the idea of how similar this is to the discrete version!

□

4.1.3 From Discrete to Continuous

Here is a nice summary chart of how similar the formulae for continuous RVs and discrete RVs are! Note that to compute the expected value of a discrete random variable, we took a weighted sum of each value multiplied by its probability. For continuous random variables though, we take an integral of each value multiplied by its density function! We'll see some examples below.

	Discrete	**Continuous**
PMF/PDF	$p_X(x) = \mathbb{P}(X = x)$	$f_X(x) \neq \mathbb{P}(X = x) = 0$
CDF	$F_X(x) = \sum_{t \leq x} p_X(t)$	$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$
Normalization	$\sum_x p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x)dx = 1$
Expectation/LOTUS	$\mathbb{E}[g(X)] = \sum_x g(x)p_X(x)$	$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx$

4.1.4 Exercises

1. Suppose X is continuous with density

   $$f_X(x) = \begin{cases} cx^2 & 0 \leq x \leq 9 \\ 0 & \text{otherwise} \end{cases}$$

   Write an expression for the value of c that makes X a valid pdf, and set up expressions (integrals) for its mean and variance. Also, find the cdf of X, F_X.

Solution: We need the total area under the curve to be 1, so

$$1 = \int_{-\infty}^{\infty} f_X(y)dy = \int_0^9 cy^2 dy = c\left[\frac{1}{3}y^3\right]_0^9 = c\frac{729}{3} = 243c$$

Hence, $c = \frac{1}{243}$. The expected value is the weighted average of each point weighted by its density, so

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} z f_X(z)dz = \int_0^9 z\frac{1}{243}z^2 dz = \frac{1}{243}\int_0^9 z^3 dz$$

Similarly, by LOTUS,

$$\mathbb{E}\left[X^2\right] = \int_{-\infty}^{\infty} z^2 f_X(z)dz = \int_0^9 z^2\frac{1}{243}z^2 dz = \frac{1}{243}\int_0^9 z^4 dz$$

Finally, we can set

$$\mathsf{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$

For the CDF, we know that

$$F_X(t) = \mathbb{P}(X \le t) = \int_{-\infty}^{t} f_X(y)dy$$

We actually have three cases, similar to the example earlier. If $t < 0$, $F_X(t) = 0$ since there's no way to get a negative number (the range is $\Omega_X = [0, 9]$). If $t > 9$, $F_X(t) = 1$ since we are guaranteed to get a number less than t. And for $0 \le t \le 9$, we just do a normal integral to get that

$$F_X(t) = \mathbb{P}(X \le t) = \int_{-\infty}^{t} f_X(s)ds = \int_{-\infty}^{0} f_X(s)ds + \int_0^t f_X(s)ds = 0 + \int_0^t cs^2 ds = \frac{c}{3}t^3$$

Putting this together gives:

$$F_X(t) = \begin{cases} 0 & t < 0 \\ \frac{c}{3}t^3 & 0 \le t \le 9 \\ 1 & t > 9 \end{cases}$$

2. Suppose X is continuous with pdf

$$f_X(x) = \begin{cases} \frac{c}{x^2} & 1 \le x \le \infty \\ 0 & \text{otherwise} \end{cases}$$

Write an expression for the value of c that makes X a valid pdf, and set up expressions (integrals) for its mean and variance. Also, find the cdf of X, F_X.

Solution: We need the total area under the curve to be 1, so

$$1 = \int_{-\infty}^{\infty} f_X(y)dy = \int_1^{\infty} \frac{c}{y^2}dy = -c\left[\frac{1}{y}\right]_1^{\infty} = -c(0-1) = c$$

Hence, $c = 1$. The expected value is the weighted average of each point weighted by its density, so

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} z f_X(z)dz = \int_1^{\infty} z \cdot \frac{1}{z^2}dz = \int_1^{\infty} \frac{1}{z}dz = [\ln(z)]_1^{\infty} = \infty$$

Actually, the mean and variance are undefined (since they are infinite)! If the integral for $\mathbb{E}[X]$ did not converge, then the integral for $\mathbb{E}\left[X^2\right]$ had no chance either (try it)! For the cdf, we know that

$$F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^{t} f_X(y)dy$$

We actually have two cases. If $t < 1$, $F_X(t) = 0$ since there's no way to get a number less than 1 (the range is $\Omega_X = [1, \infty)$). For $t > 1$, we just do a normal integral to get that

$$F_X(t) = \mathbb{P}(X \leq t) = \int_{-\infty}^{t} f_X(s)ds = \int_{-\infty}^{1} f_X(s)ds + \int_{1}^{t} f_X(s)ds = \int_{1}^{t} \frac{1}{s^2}ds = -\left[\frac{1}{s}\right]_1^t = -\left(\frac{1}{t} - 1\right) = 1 - \frac{1}{t}$$

Putting this together gives:

$$F_X(t) = \begin{cases} 0 & t < 1 \\ 1 - \frac{1}{t} & t \geq 1 \end{cases}$$