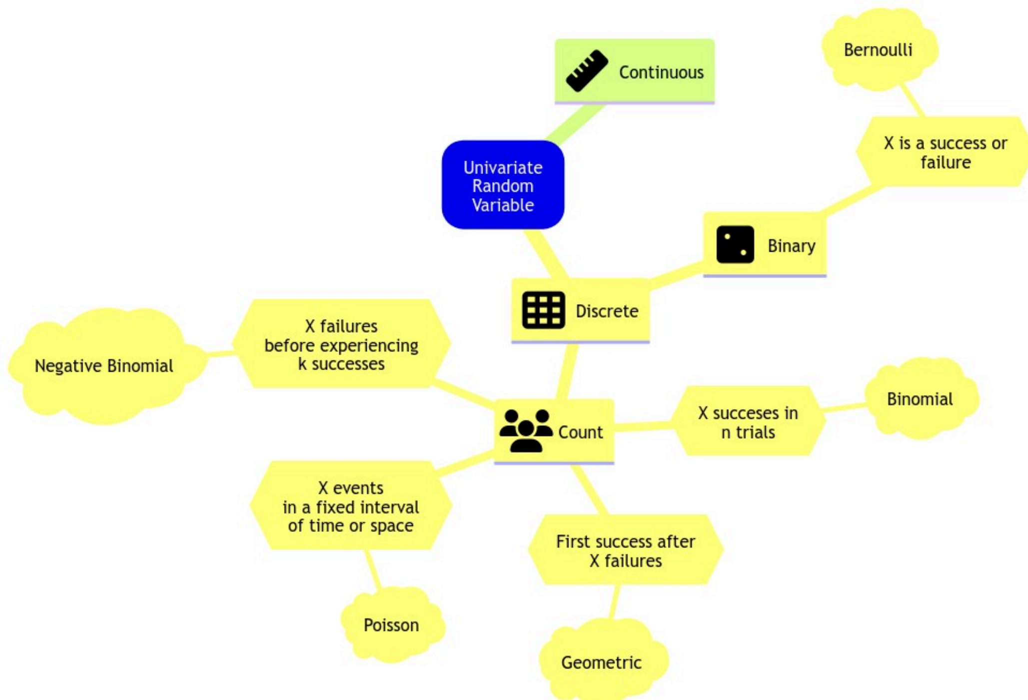
Continuous Distribution

Lecture 5

Please, sign in on iClicker

Roadmap

- So far, everything we have done has been in the context of discrete random variables.



Outline

1. Continuous Random Variables
2. Probability Density
3. Distribution Properties
4. Representing Distributions besides PDF or PMF
5. Exponential Distribution

1. Continuous Random Variables

> What is the current water level of the Bow River at Banff, Alberta?

> What about the current temperature in Vancouver, BC?

> What is the stock price for Nvidia today?

All the previous questions pertain to continuous cases!

- A continuous random variable has uncountably infinite amount of outcomes.
  - For example, temperature can be 20, 20.0001, 20.1415926

However, we can never measure anything perfectly on a continuous scale

- In practice, our measurements are limited by the precision of our tools.
- As a rule of thumb, we consider a variable continuous if the difference between neighboring values is negligible or not significant.

A First Example

- You record your total monthly expenses each month.
- You end up with 20 months worth of data.

```
 [1] "$1903.68" "$3269.61" "$6594.05" "$1693.94" "$2863.71" "$3185.01"
 [7] "$4247.04" "$2644.27" "$8040.42" "$2781.11" "$3673.23" "$4870.13"
[13] "$2449.53" "$1772.53" "$7267.11" "$938.67"  "$4625.33" "$3034.81"
[19] "$4946.4"  "$3700.16"
```

- A difference of $\$0.01$ is not a big deal, thus we may treat this as a continuous random variable:

$$X = \text{Total monthly expenses.}$$

A Second Example

- If you play a game where you can win and lose pennies (1 cent coin).

- Here are your 10 net winnings:

```
[1]  0.01 -0.01  0.02  0.01  0.04  0.02 -0.03 -0.01  0.05  0.04
```

- Since a difference of $\$0.01$ is a big deal, it is best to treat this as discrete.

2. Probability Density

- A continuous random variable has a probability density function.
- To understand this concept, we'll start with a physical analogy of density.

A Note on Integrals:

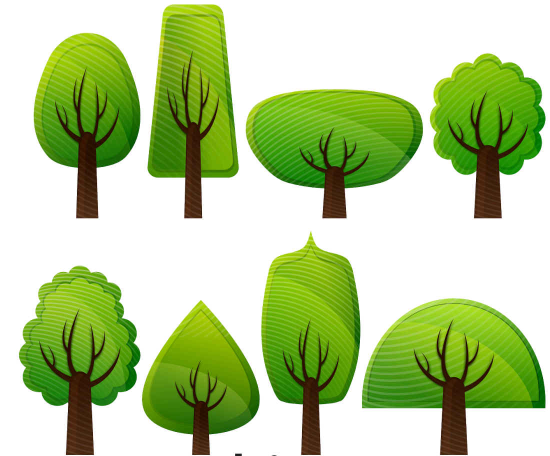
- Working with probability density functions involves integrating some functions $\int_a^b f(x) \mathrm{d}x$.
- In this course, we will only ask you to integrate basic functions.

> **Heads-up:** The scope of these upcoming lectures is to understand probability concepts for continuous random variables rather than Calculus.

Tree Density Example

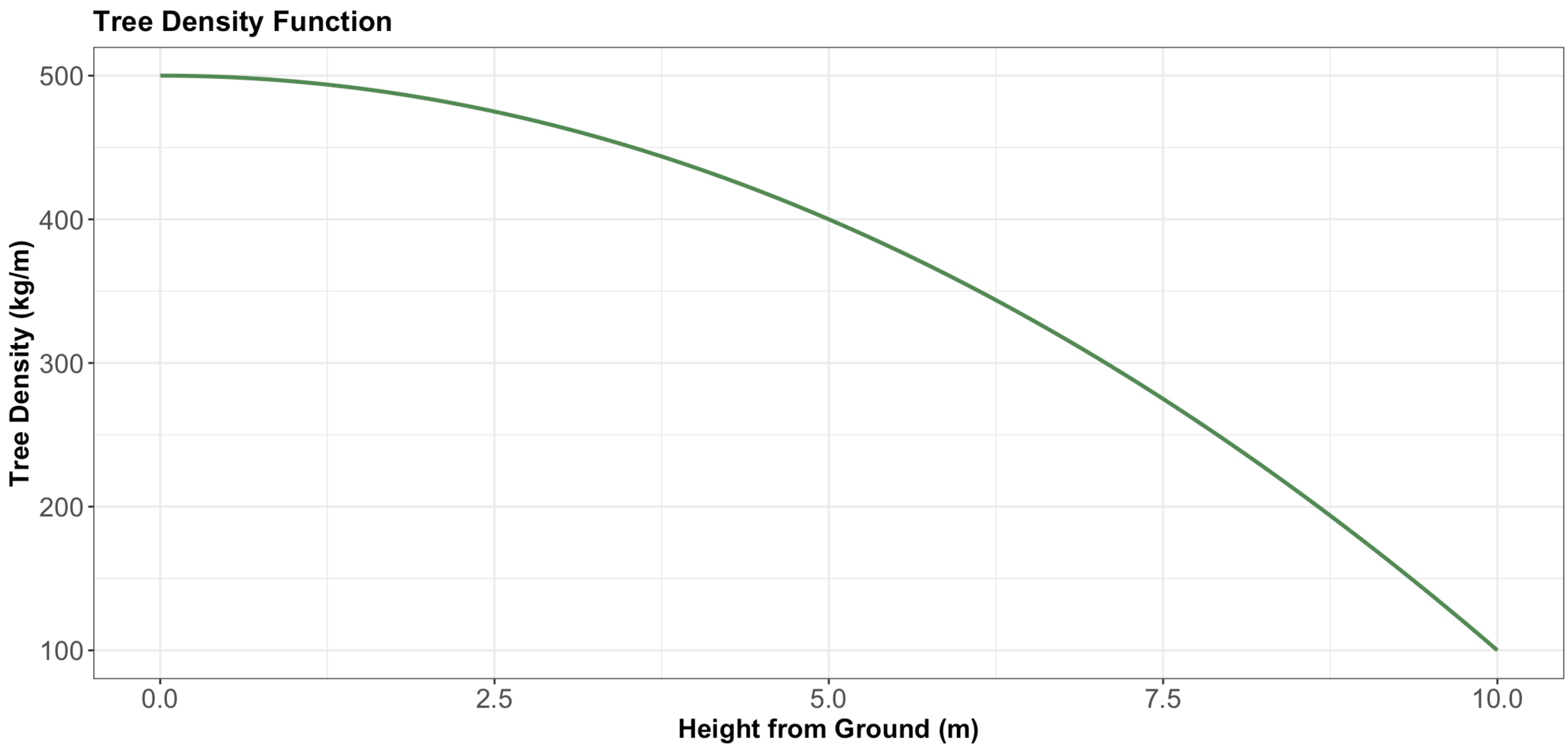
- Imagine a tree, every meter of the tree weighs differently!!!
- Some parts of the tree might be dense and heavy (like the base), while other parts might be lighter and less dense.

Tree Density Example



- Density describes how much weight is concentrated in a given space.
- The density of the tree, measured in kg/m, is a function that varies as you move up the tree.
- You can think of it as the weight of an infinitely thin slice of tree, divided by the thickness of that slice.

Visualization of a density function



What is the total weight of the tree?

- We sum the weights of all infinitely thin slices.
- This is done by taking the integral of the density function $f(x)$:

$$\int_0^{10} f(x) \mathrm{d}x.$$

UBC Master of Data Science

How about the weight between two heights?

- We can get the weight of the tree in an interval between two heights a and b

$$\int_a^b f(x)\mathrm{d}x$$

How much does the tree weigh at a height equal to 5 m?

- The tree only has a density at a height equal to 5 m!
- Since the weight between heights a and b is $\int_a^b f(x) \mathrm{d}x$.
- The weight at $x = 5 \text{ m}$ is $\int_5^5 f(x) \mathrm{d}x = 0$.
- **The tree weight at 5 m is zero!**

From Density to Probability Density

- Physical density tells us how mass is distributed over space
- Probability density tells us how **probability** is distributed over a range of values for a random variable.

Example

- Let us consider a continuous random variable:

$$X = \text{Height of a person in meters.}$$

- We have a probability density $f_X(x)$ (in “probability per meter”) of heights.

Computing Probabilities

- We can compute the probability that a randomly selected person is between 1.5 and 1.6 m:

$$P(1.5 \leq X \leq 1.6) = \int_{1.5}^{1.6} f_X(x) \mathrm{d}x.$$

2.1. A Note on Units

- Technically, in the height case, the density $f_X(x)$ has units of $1/\mathrm{m}$ or, equivalently, m^{-1}.
- For instance, what is the probability that a person's height is between $x = 1\ \mathrm{m}$ and $x = 2\ \mathrm{m}$?

$$P(1 \leq X \leq 2) = \int_1^2 f_X(x)\mathrm{d}x,$$

- Recall, $\mathrm{d}x$ is an infinitely tiny interval of x, measured in metres.

Probability Density Function

Probability of a Particular Value is Zero

- Let us answer the following question:

> What is the probability that a person is 1.5 meters tall?

- That is $1.5000000000 \ldots \text{ m}$.
- Just like the weight of a tree at a height equal to 5 m, this is 0.
- But we can ask for the probability of a range of heights!

Connection to PMF

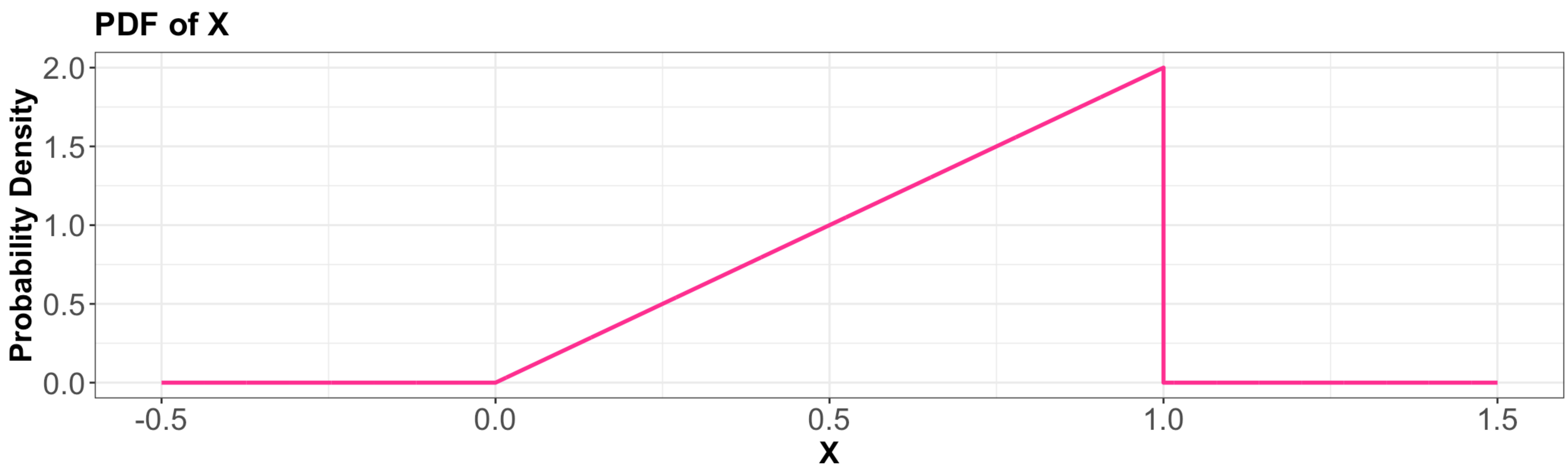
- **PMF:** $p_X(x) = P(X = x)$
- **PDF:** $f_X(x)$

- Similarities
  - Both must satisfy normalization conditions.
- Differences:
  - For PMF, we can compute the probability of specific values.
  - For PDF, we compute the probability by integrating the PDF over an interval.

2.3. Example: Low Purity Octane

Let X be a random variable with the following PDF:

$$f_X(x) = \begin{cases} 2x & \text{for} \quad 0 \leq x \leq 1 \\ 0 & \text{elsewhere.} \end{cases}$$



In-Class Question

What is the probability of $X = 0.25$? That is,

$$P(X = 0.25).$$

$$f_X(x) = \begin{cases} 2x & \text{for} \quad 0 \leq x \leq 1 \\ 0 & \text{elsewhere.} \end{cases}$$

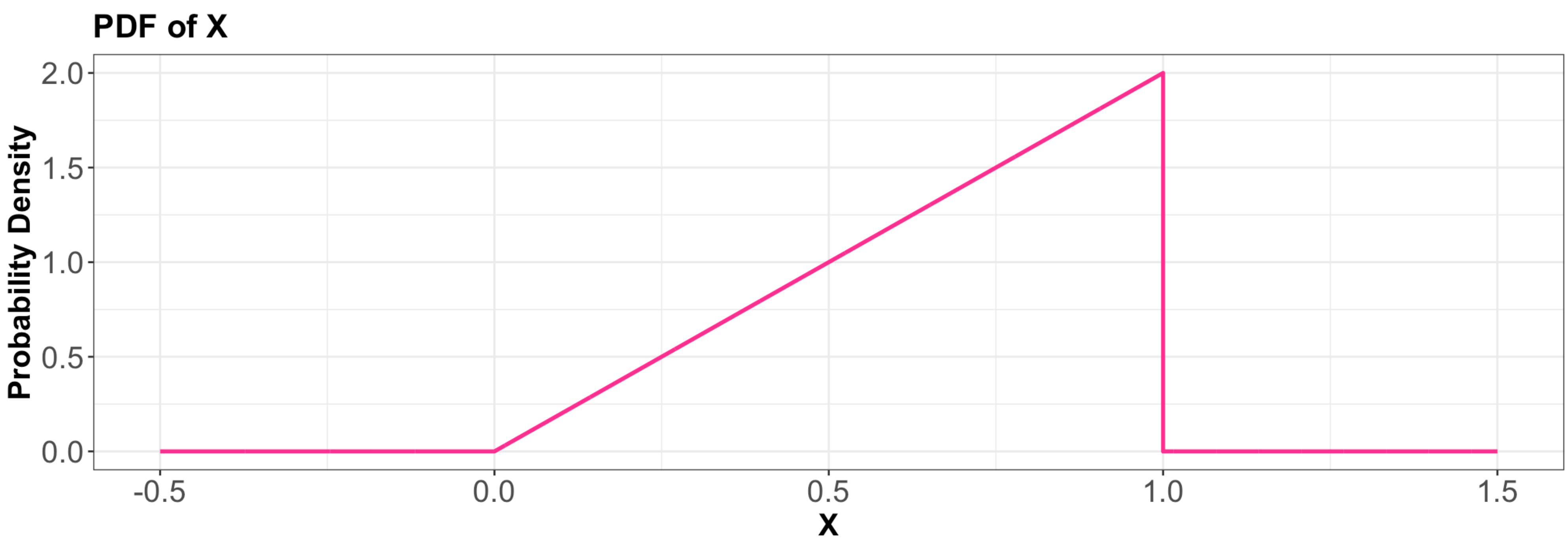
Answer

- For a single value, we would only have the corresponding density value but a probability equal to zero.
- Thus,

$$P(X = 0.25) = 0.$$

In-Class Question

The PDF evaluates to be > 1 in some places on the vertical axis. Does this mean that this is not a valid density? Why?

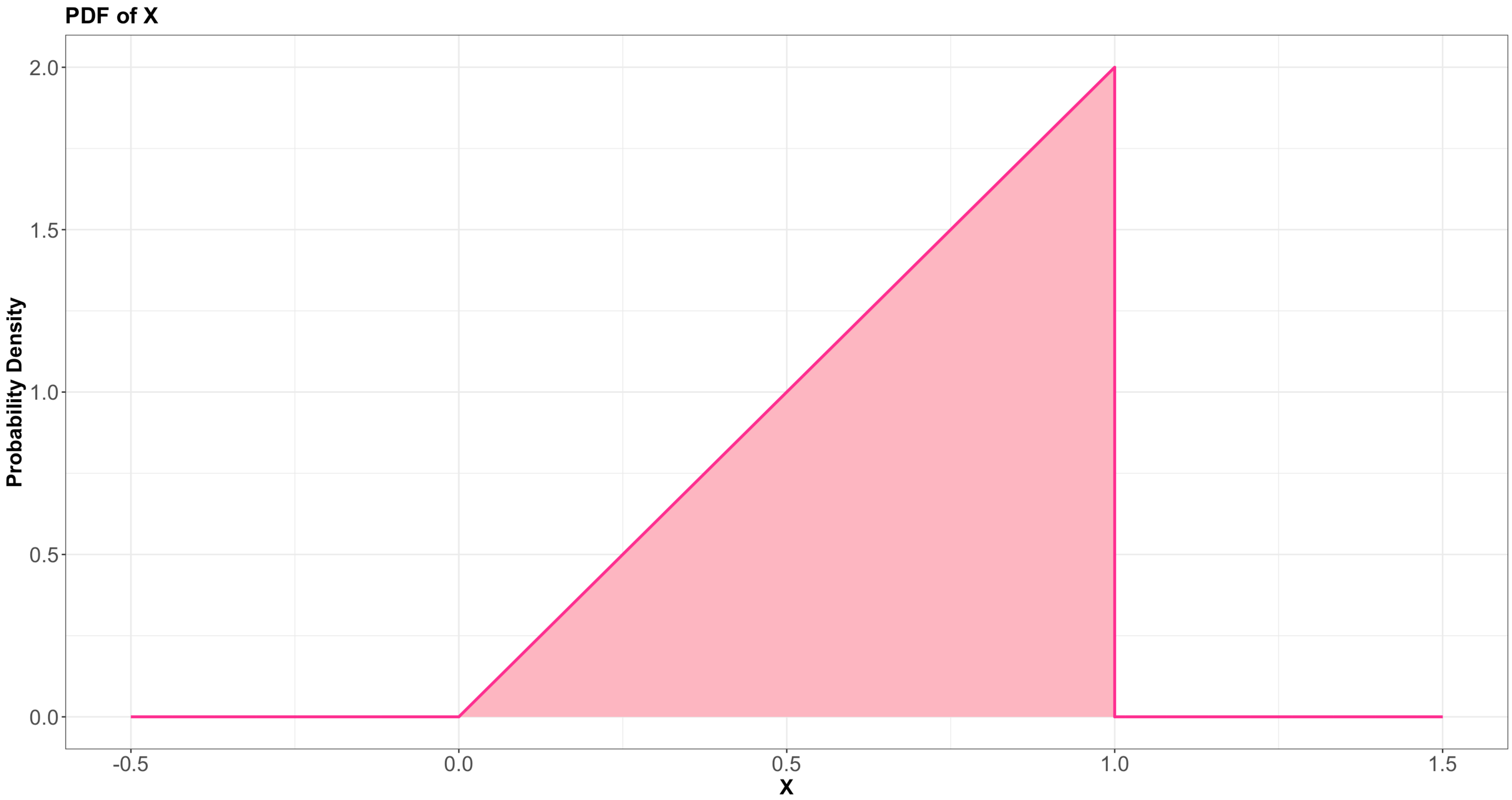


Answer

- No, it is valid since the area under the curve is 1:

$$
\begin{aligned}
P(0 \leq X \leq 1) &= \int_0^1 2x \mathrm{d}x \\
&= \frac{2x^2}{2} \Big|_0^1 \\
&= 1 - 0 \\
&= 1.
\end{aligned}
$$

Graphically...



iClicker Question

Using the PDF, what is the probability of $X < 0.5$?

$$f_X(x) = \begin{cases} 2x & \text{for} \quad 0 \leq x \leq 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Select the correct option:

A. 0.75

B. 0.1

C. 0.25

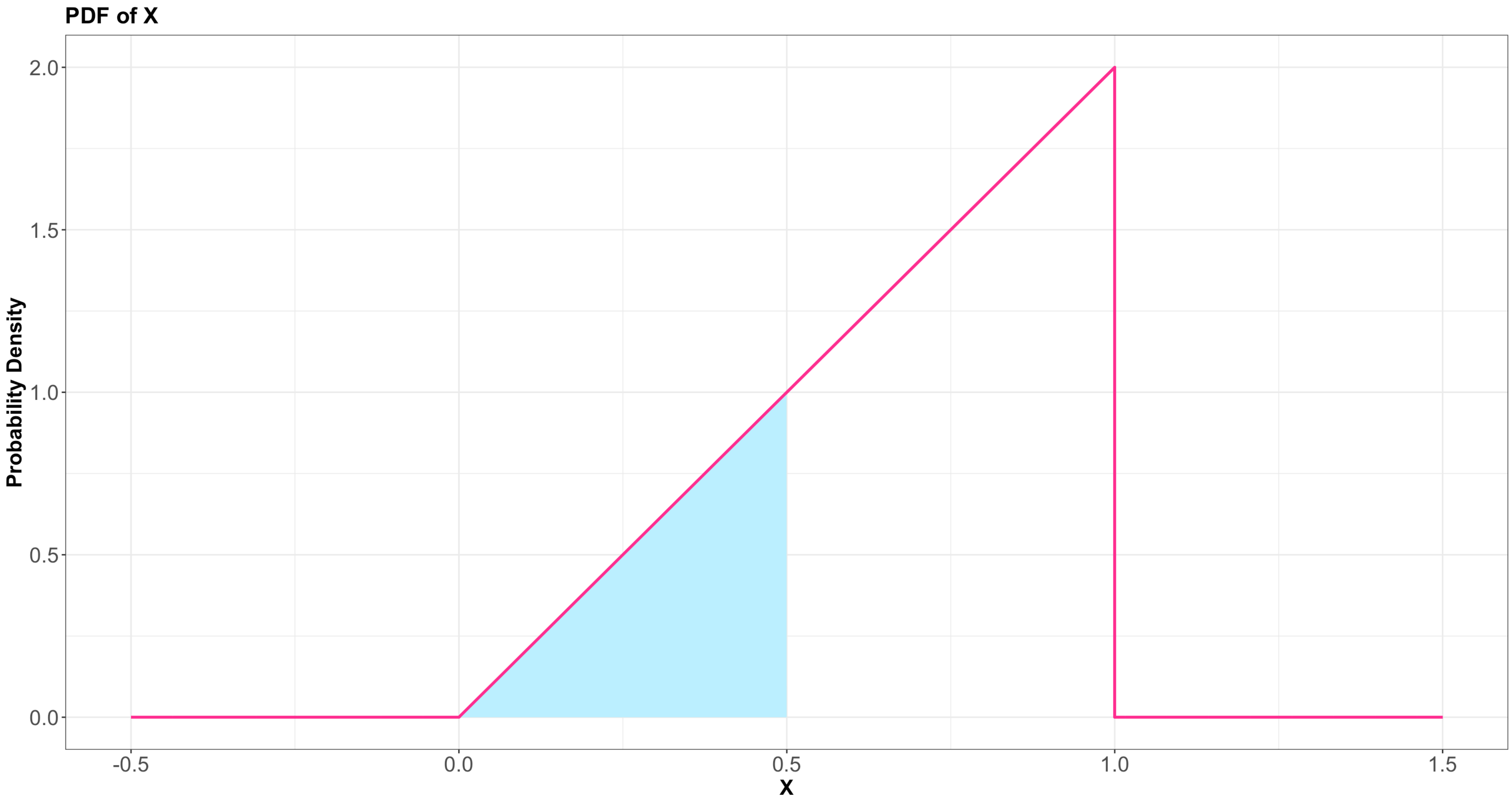
D. 0.5

Answer

- We integrate the PDF as follows:

$$
\begin{aligned}
P(X < 0.5) &= \int_0^{0.5} 2x \mathrm{d}x \\
&= \frac{2x^2}{2} \Big|_0^{0.5} \\
&= 0.5^2 - 0^2 \\
&= 0.25.
\end{aligned}
$$

Graphically...



iClicker Question

Given that $P(X < 0.5) = 0.25$, what is the probability of $X \leq 0.5$?

Select the correct option:

A. 0.25

B. 0.249

C. 0.749

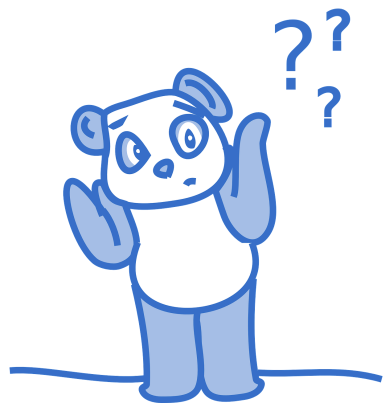
D. 0.75

Answer

- It is also 0.25.
- In continuous random variables, there is no difference between $\leq$ and $<$, or $\geq$ and $>$.

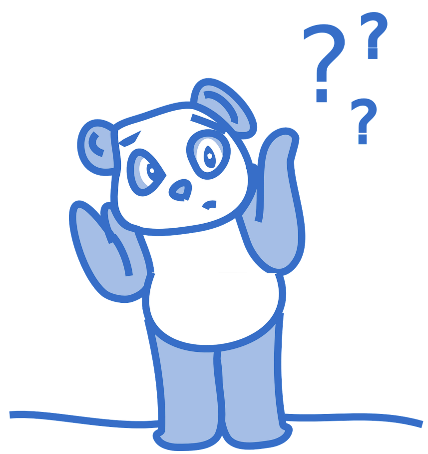
3. Distribution Properties

- With continuous random variables, it becomes easier to expand how we describe a distribution in terms of central tendency and uncertainty measures.



3.1. Mean, Variance, Mode, and Entropy

- The central tendency and uncertainty measures still apply in the continuous cases, but with slight variations.



Mode

- The mode is the outcome having the highest density. That is, for a continuous random variable X with PDF $f_X(x)$:

$$\text{Mode}(X) = \arg\max_x f_X(x).$$

- Mode is not used as much for continuous variables.

Entropy

- The entropy can be defined by replacing the sum in the discrete case with an integral:

$$H(X) = -\int_x f_X(x) \log[f_X(x)] \mathrm{d}x.$$

- It is also called differential entropy.
- Entropy of a continuous random variable may be infinitely large, positive, or even negative (why?).

Mean and Variance

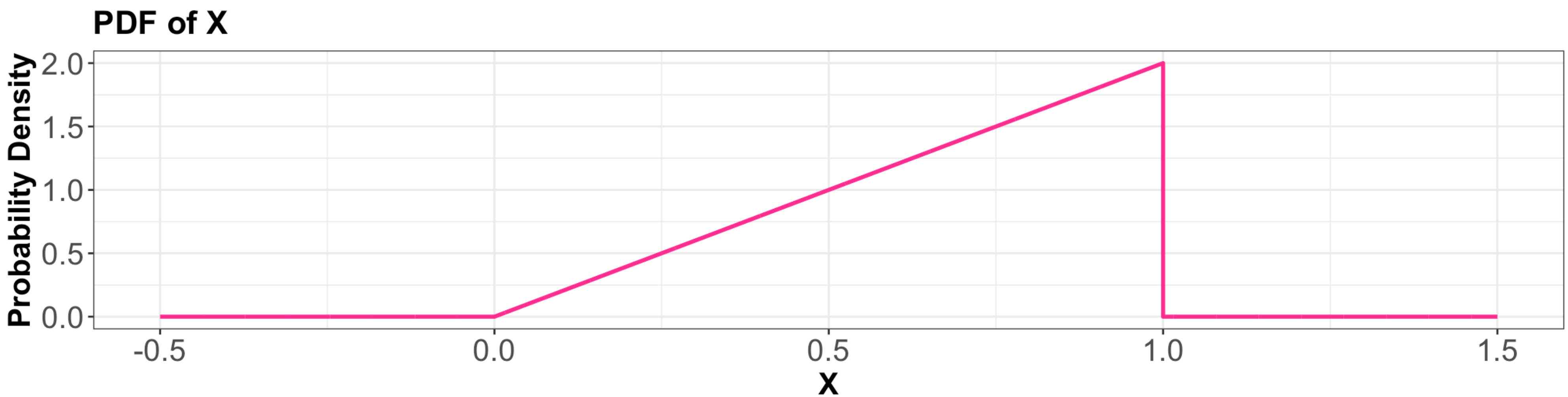
- The mean is a central tendency measure:

$$\mathbb{E}(X) = \int_x x \, f_X(x) \mathrm{d}x.$$

- The variance is an uncertainty measure. Let μ_X be $\mathbb{E}(X)$, this measure is defined as:

$$\begin{aligned}\mathrm{Var}(X) = \mathbb{E}[(X - \mu_X)^2] &= \int_x (x - \mu_X)^2 \, f_X(x) \mathrm{d}x \\ &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2\end{aligned}$$

Going back to the previous example



- The mode is $\text{Mode}(X) = \arg\max_x f_X(x) = 1.$
- The entropy works out to be

$$H(X) = -\int_0^1 \underbrace{(2x)\log(2x)}_{f_X(x)\,\log f_X(x)} \mathrm{d}x \approx -0.1931.$$

More common measures...

- The mean is (which is very different from the mode),

$$\mathbb{E}(X) = \int_0^1 x \underbrace{(2x)}_{f_X(x)} \mathrm{d}x = \frac{2}{3} \approx 0.66.$$

- The variance ends up being

$$\mathrm{Var}(X) = \int_0^1 \left(x - \frac{2}{3}\right)^2 \underbrace{(2x)}_{f_X(x)} \mathrm{d}x = \frac{1}{18} \approx 0.056.$$

3.2. Median

- It is another central tendency measure.
- The median $\mathrm{M}(X)$ is the outcome for which there is a 50-50 chance of seeing a greater or lesser value:

$$P[X \leq \mathrm{M}(X)] = 0.5.$$

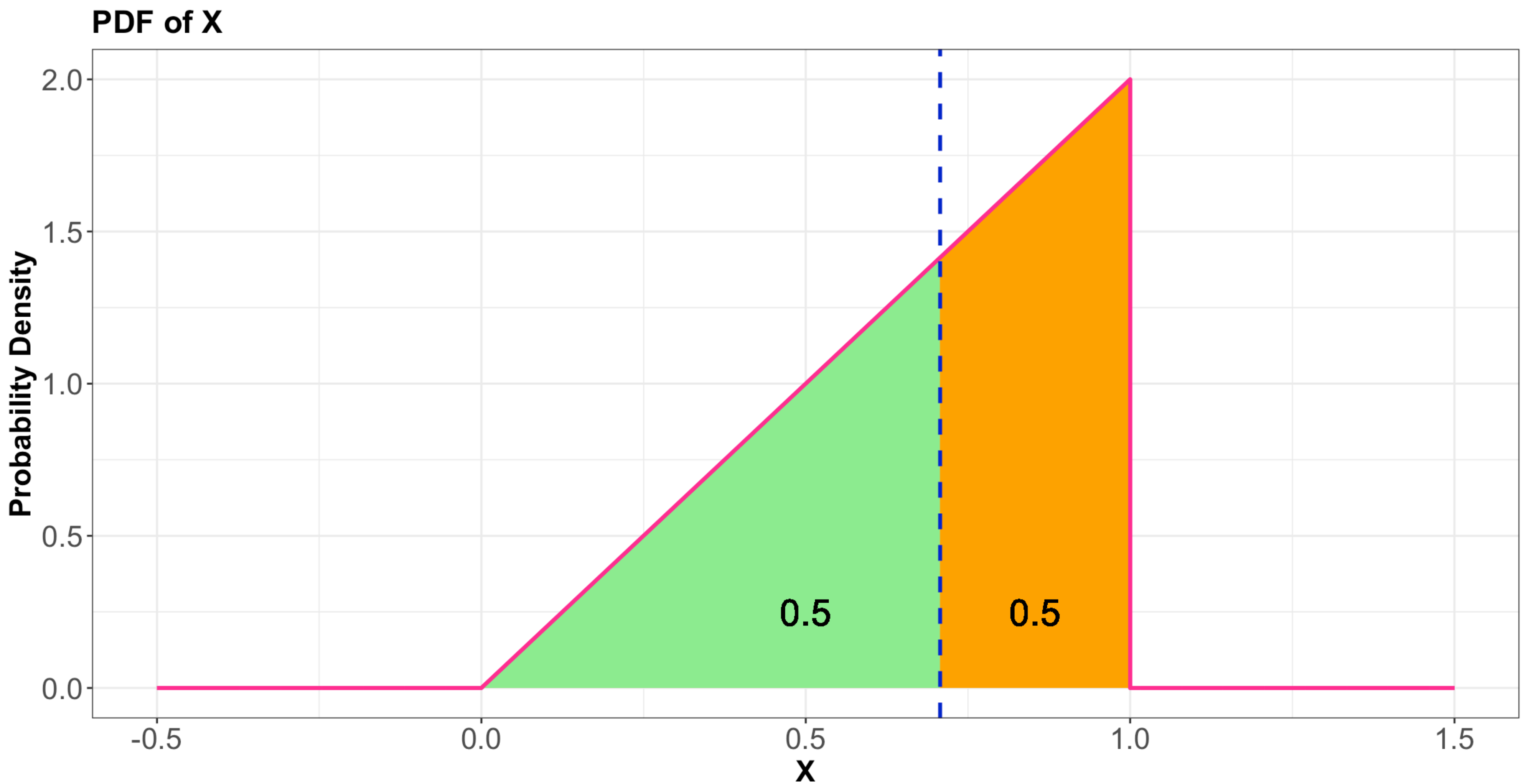
Going back to the previous example

$$
\begin{aligned}
P[X \leq \mathrm{M}(X)] = \int_0^{\mathrm{M}(X)} f_X(x) \mathrm{d}x &= \int_0^{\mathrm{M}(X)} 2x \mathrm{d}x \\
&= x^2 \Big|_0^{\mathrm{M}(X)} \\
&= [\mathrm{M}(X)]^2 = 0.5.
\end{aligned}
$$

- Thus, we have:

$$
\mathrm{M}(X) = \sqrt{0.5} = 0.7071.
$$

Graphically...



3.3. Quantiles

- More general than a median is a quantile.
- The definition of a p-quantile $Q(p)$ is the outcome with a probability p of getting a smaller outcome:

$$P[X \leq Q(p)] = p.$$

- The median is a special case, and is the 0.5-quantile.

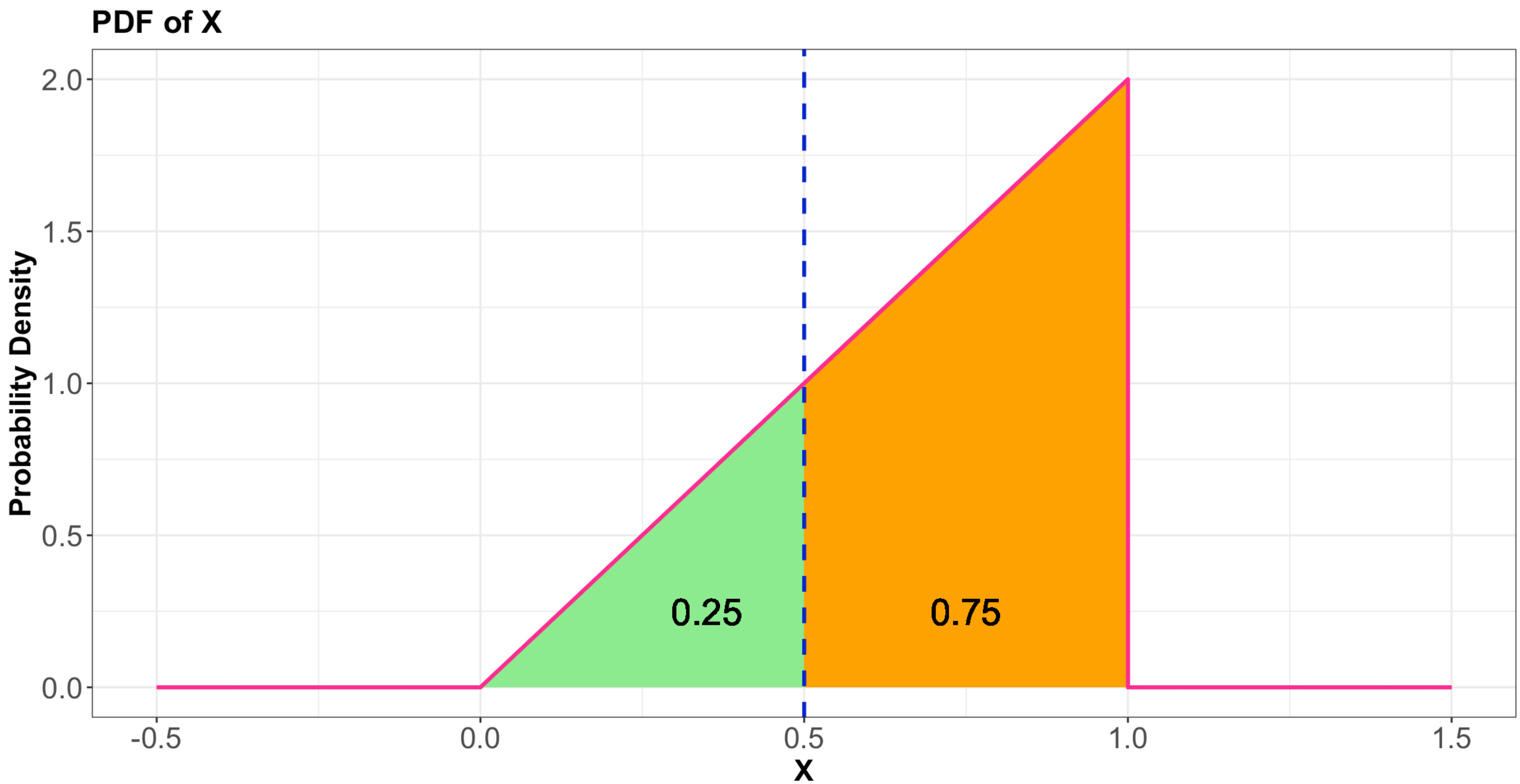
Going back to the previous example for the 0.25-quantile!

$$
\begin{aligned}
P[X \leq Q(0.25)] &= \int_0^{Q(0.25)} f_X(x) \mathrm{d}x = \int_0^{Q(0.25)} 2x \mathrm{d}x \\
&= x^2 \Big|_0^{Q(0.25)} \\
&= [Q(0.25)]^2 = 0.25.
\end{aligned}
$$

- Hence, we have:

$$
Q(0.25) = \sqrt{0.25} = 0.5.
$$

Graphically...



3.4 Prediction Intervals using Quantiles

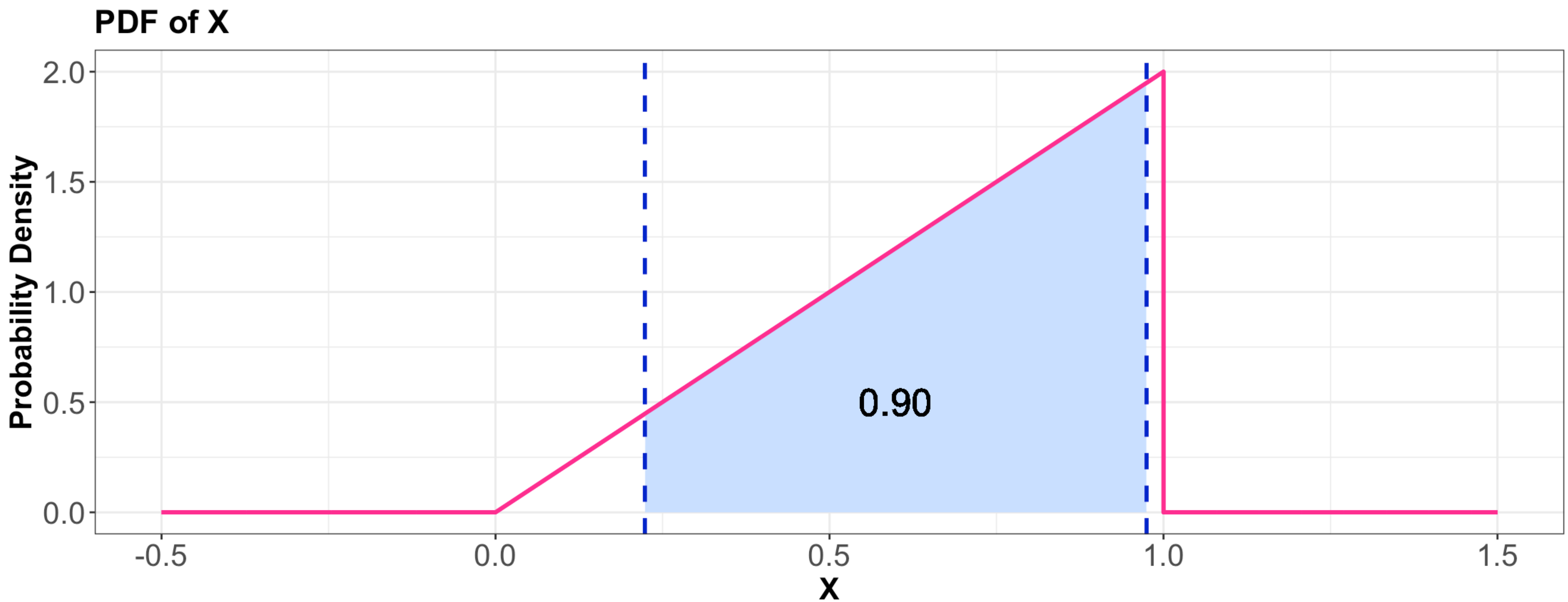
- A prediction interval gives a range that will contain a random outcome with a pre-specified probability p.
- We can use $\frac{1-p}{2}$-quantile as the lower limit and $\frac{1+p}{2}$-quantile as the upper limit to construct the interval:

$$P(Q(\frac{1-p}{2}) < X < Q(\frac{1+p}{2})) = p$$

UBC Master of Data Science

Going back to the previous example

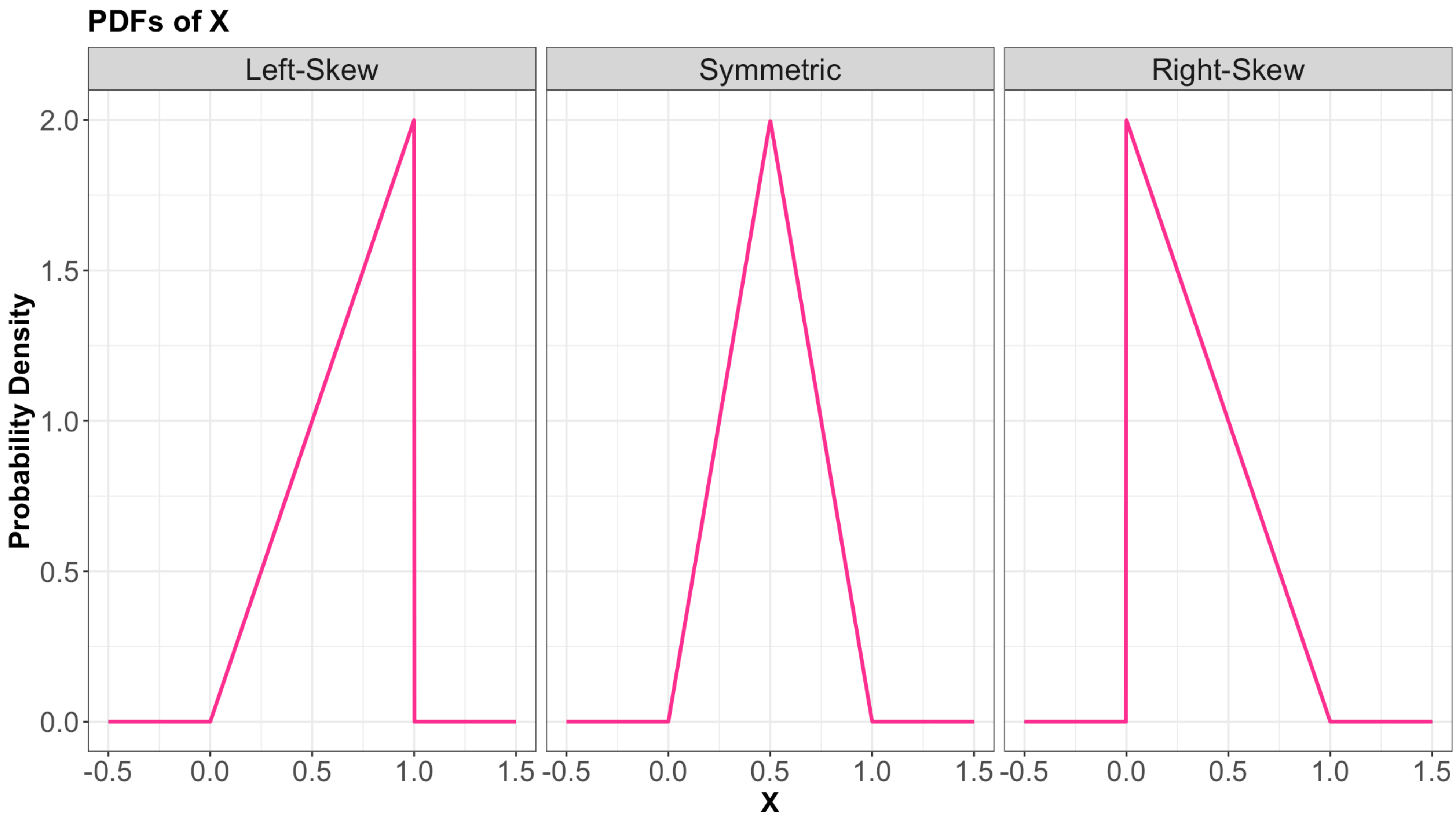
- We can construct a 90% prediction interval for X using the 0.05 and 0.95-quantiles: $[0.2236, 0.9746]$



3.5. Skewness

- Skewness measures how “asymmetric” a distribution is, as well as the direction of the skew:
  - If the density is symmetric about a point, then the skewness is 0.
  - If the density is more “spread-out” towards the right/positive values, then the distribution is said to be right-skewed (positive skewness).
  - If the density is more “spread-out” towards the left/negative values, then the distribution is said to be left-skewed (negative skewness).

Going back to the previous example



Definition of Skewness

- If X is a continuous random variable, with $f_X(x)$ as a PDF, then

$$\text{Skewness}(X) = \mathbb{E}\left[\left(\frac{X - \mu_X}{\sigma_X}\right)^3\right] = \int_x \left(\frac{x - \mu_X}{\sigma_X}\right)^3 \cdot f_X($$

4. Representing Distributions

- There are more ways we can represent a distribution besides the PMF or PDF.
- Keep in mind that all of these representations capture full information about a distribution.

4.1. Cumulative Distribution Function

- The cumulative distribution function (CDF) is defined as

$$F_X(x) = P(X \leq x).$$

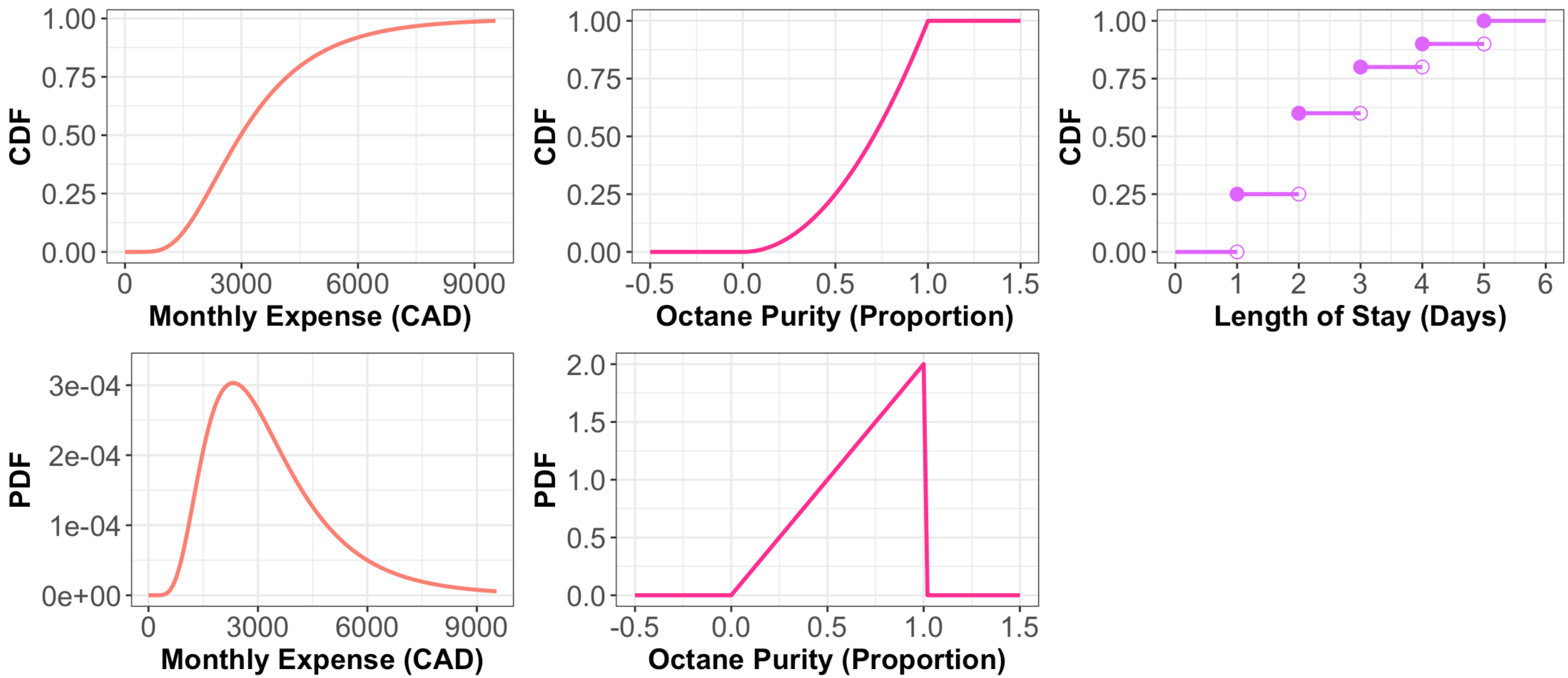
- We can calculate this using a density $f(\cdot)$ by

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t.$$

- Every numeric random variable has a corresponding CDF!

Example of CDFs

- The CDF is still defined for discrete random variables but has a jump-discontinuity at the discrete values!



More about CDFs

- For a continuous variable:
  - PDFs are derivatives of the CDFs and CDFs are integrals of PDFs.
  - When the CDF is flat, that means the PDF is zero.
- For a discrete variable, the CDF is a step function.

What Makes a Valid CDF?

1. Must never decrease.
2. It must never evalute to be < 0 or > 1.
3. $F_X(x) \to 0$ as $x \to -\infty$.
4. $F_X(x) \to 1$ as $x \to \infty$.

Calculate probability that X is between a and b using CDFs

- Recall that the probability of X being between a and b is

$$P(a \leq X \leq b) = \int_a^b f_X(x)\mathrm{d}x.$$

- Connecting the dots via CDFs:

$$\begin{aligned} P(a \leq X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F_X(b) - F_X(a) \end{aligned}$$

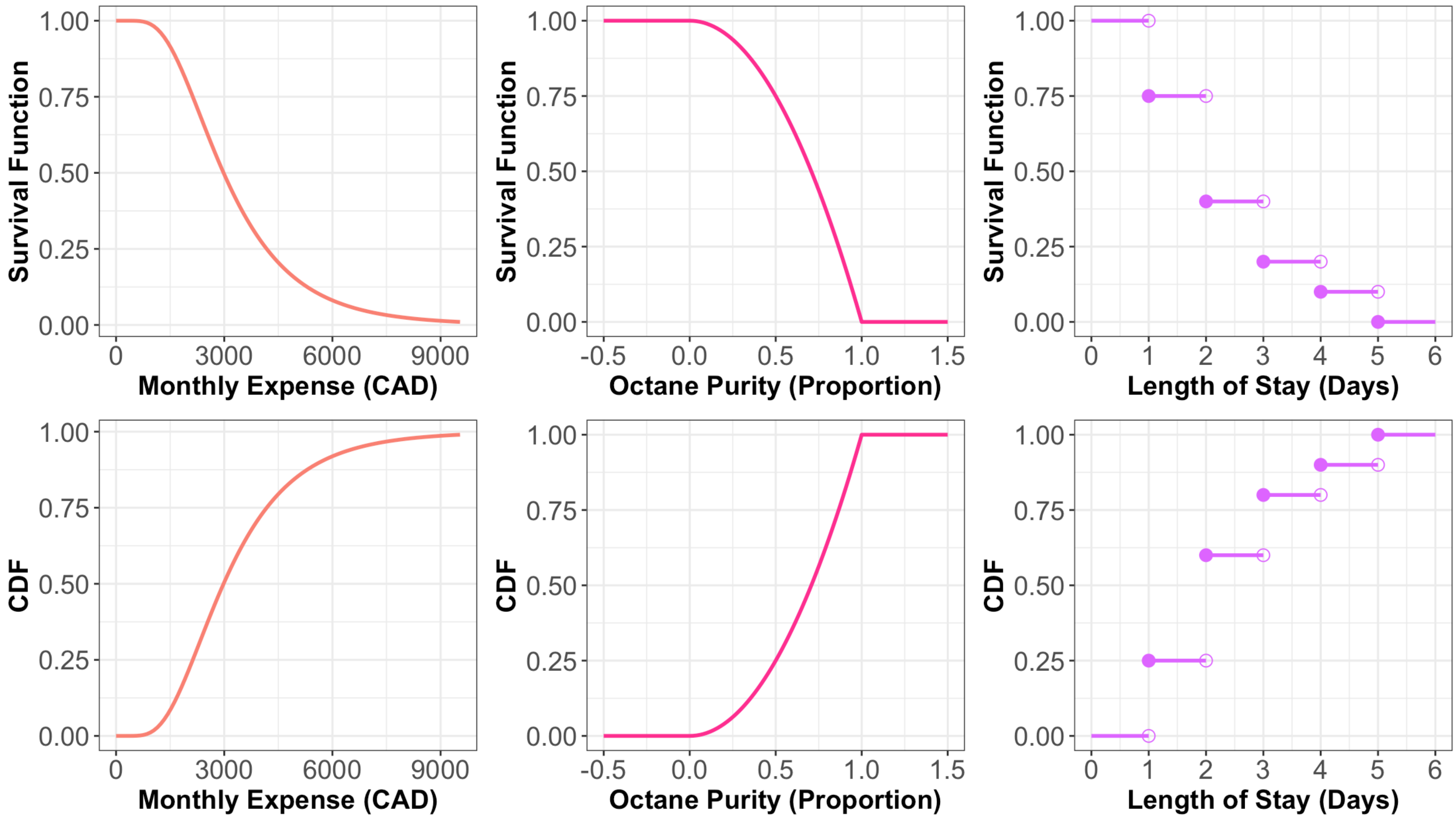
4.2. Survival Function

- It is the CDF "*flipped upside down*":

$$S_X(x) = P(X > x) = 1 - F_X(x).$$

- The name comes from Survival Analysis, where X is interpreted as a "*time of death*", so that the survival function is the probability of surviving beyond x.

Comparing Different Survival Functions



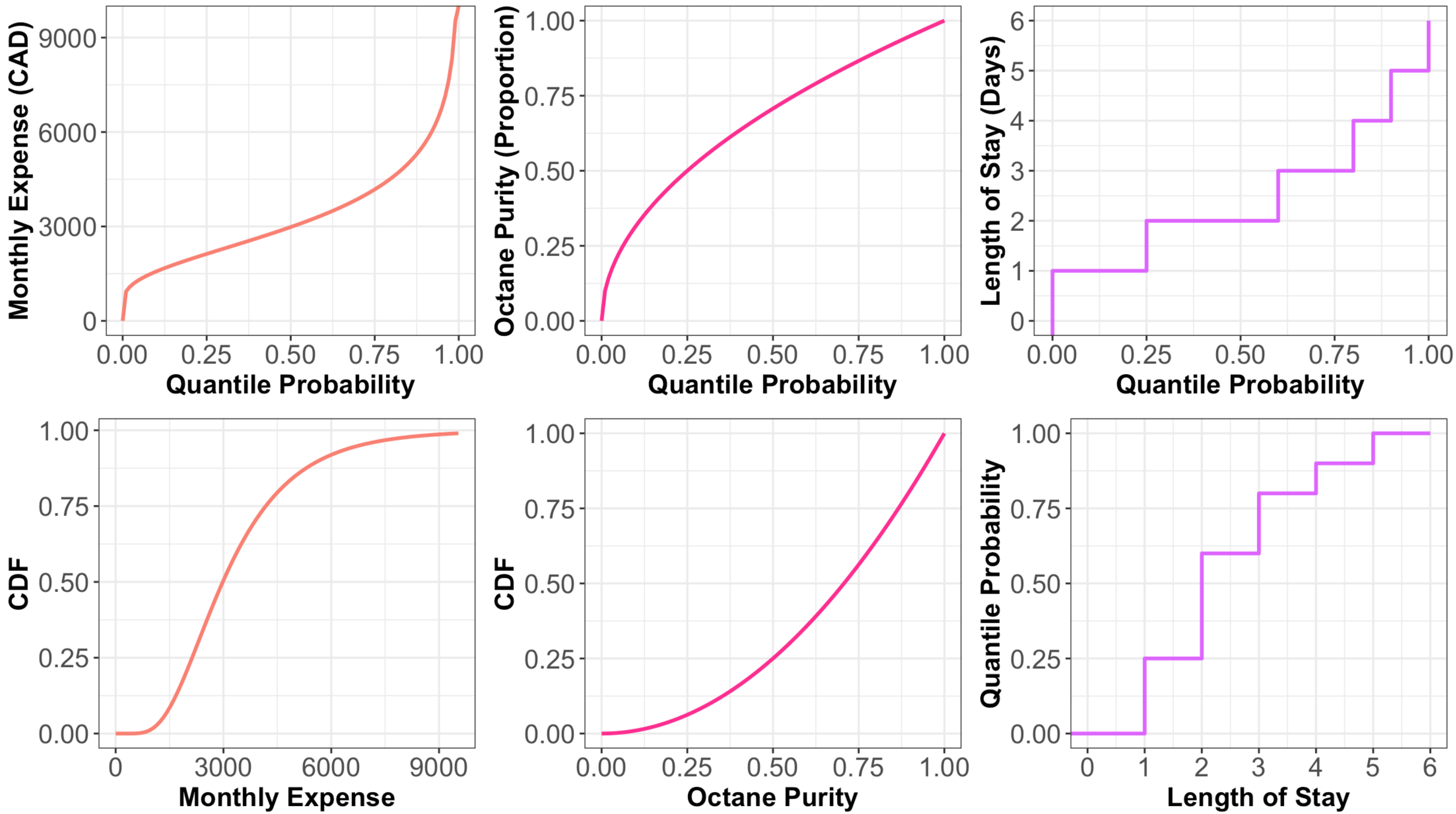
4.3. Quantile Function

- The quantile function $Q(\cdot)$ takes a probability p and maps it to the p-quantile.
- It turns out that this is the inverse of the CDF:

$$Q(p) = F^{-1}(p).$$

- Note that this function does not exist outside of $0 \leq p \leq 1$!

Comparing Different Quantile Functions



Other Ways of Representing a Distribution

- Moment generating function
- Hazard function (DSCI 562)

All these representations contain all the information about the distribution; they are different “views” of the same distribution.

5. Exponential distribution

- Just like for discrete distributions, there are also parametric families of continuous distributions.

- The Exponential distribution is the distance between events in a Poisson process (a process where events occur independently at a average rate).
- It is often interpreted as wait time.

Definition

- The family is characterized by a single parameter, usually either the mean wait time or average rate:
- The average rate $\lambda > 0$ at which events happen:

$$X \sim \text{Exponential}(\lambda).$$

- The mean wait time $\beta > 0$ between two events:

$$X \sim \text{Exponential}(\beta).$$

- $\beta = 1/\lambda.$

Using R to calcualte pdf, cdf, and quantile

- The functions are of the form `<x><dist>`, where `<dist>` is an abbreviation of a distribution family, and `<x>` is one of the values:
  - d: density function $f_X(x)$.
  - p: cumulative distribution function $F_X(x)$.
  - q: quantile function (inverse CDF).
  - r: random number generator

UBC Master of Data Science

R function for Exponential distribution

- For the Exponential distribution, we have the following R functions:
  - `dexp(x, rate)`
  - `pexp(quantile, rate)`
  - `qexp(probability, rate)`
  - `rexp(sample_size, rate)`.

Today's Learning Objectives

By the end of this lecture, you will be able to…

- Differentiate between continuous and discrete random variables.
- Interpret probability density functions and calculate probabilities from them.
- Calculate and interpret probabilistic quantities (mean, quantiles, prediction intervals, etc.) for a continuous random variable.

And…

- Explain whether a function is a valid probability density function, cumulative distribution function, quantile function, and survival function.
- Calculate quantiles from a cumulative distribution function, survival function, or quantile function.