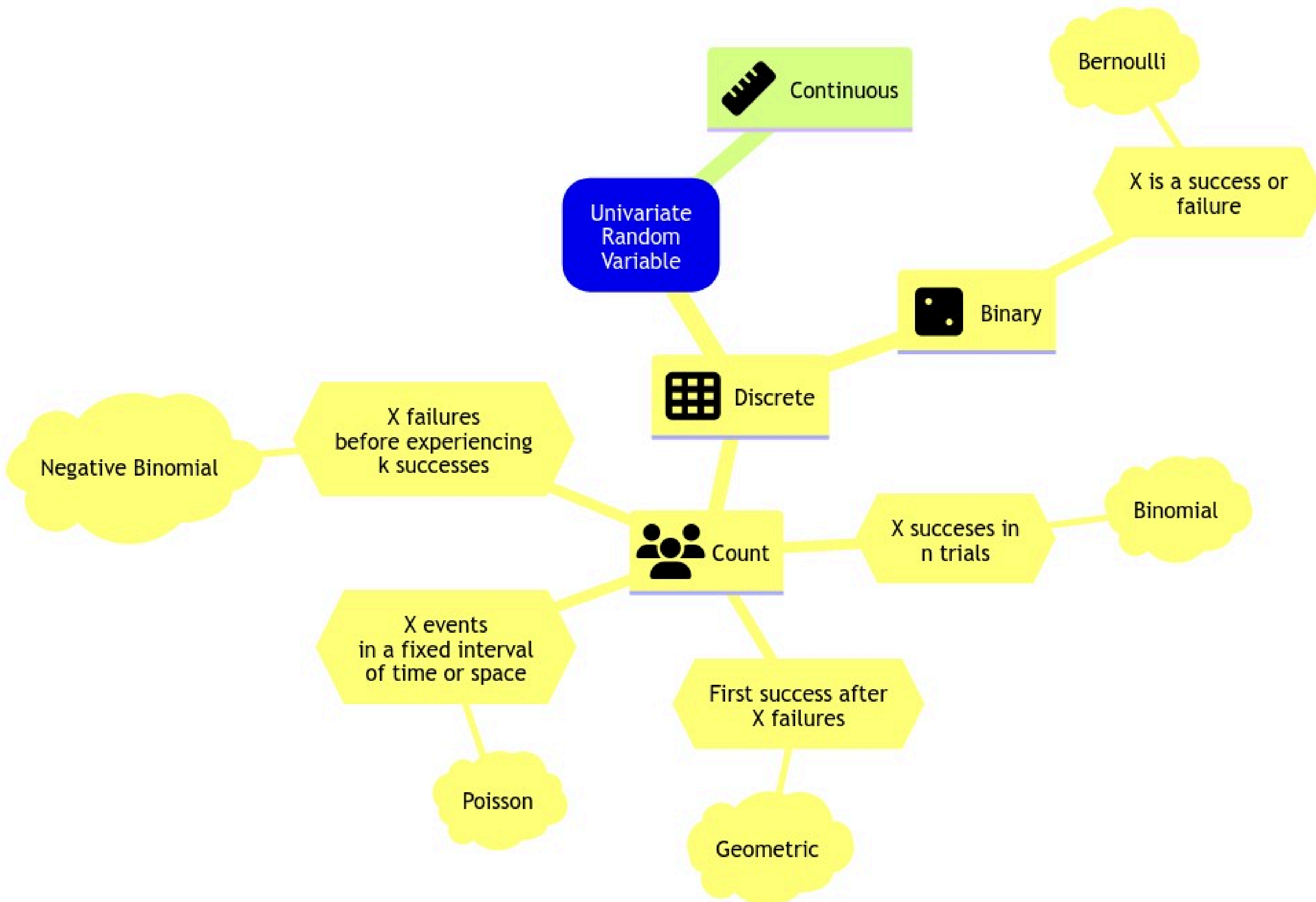
Continuous Distribution Families

Lecture 6

Please, sign in on iClicker

Time to check different continuous families!

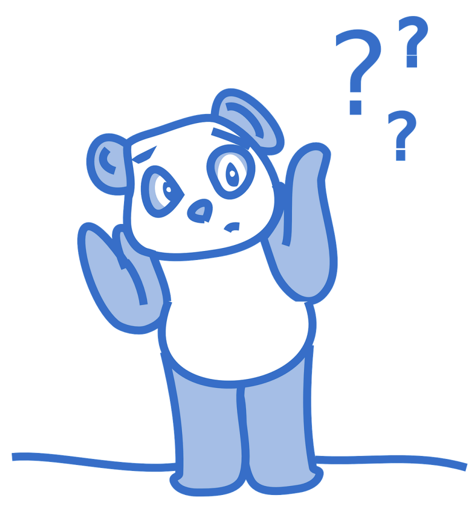


Outline

1. Common Continuous Distribution Families
2. Continuous Joint Distributions
3. Continuous Conditional Distributions

1. Common Continuous Distribution Families

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



1.1. Uniform

- A Uniform distribution has equal density in between two points a and b (for $a < b$).
- There are two parameters: one for each end-point.
- Commonly, a reference to a Uniform distribution usually implies continuous uniform.

PDF, Mean, and Variance

- It is denoted as

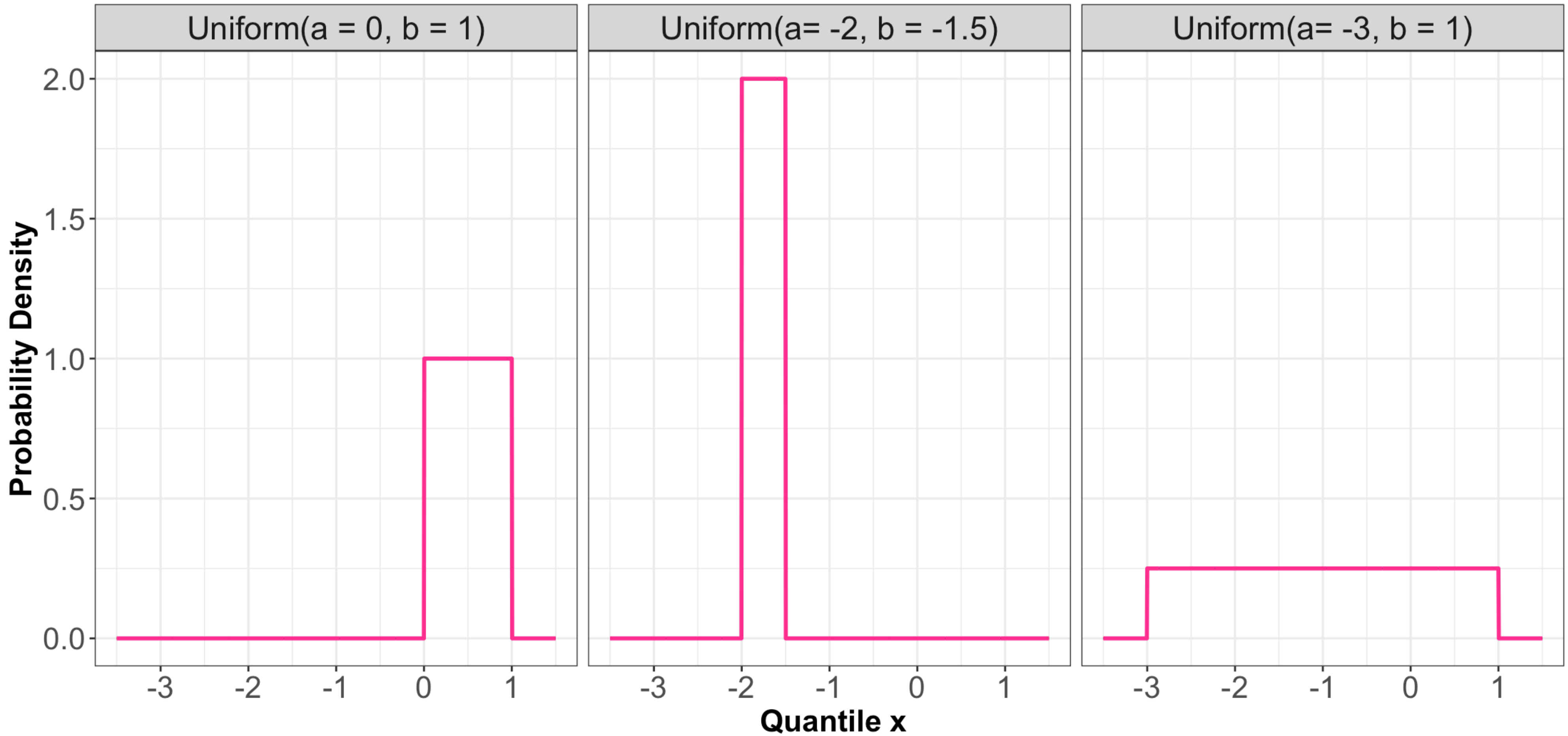
$$X \sim \text{Uniform}(a, b).$$

- The PDF is given by

$$f_X(x) = \frac{1}{b - a} \qquad \text{for} \quad a \leq x \leq b.$$

- The mean is $\mathbb{E}(X) = \frac{a+b}{2}$.
- The variance is $\text{Var}(X) = \frac{(b-a)^2}{12}$.

Some members of the Uniform family…

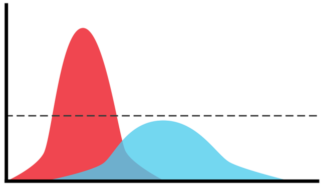


Discrete Uniform

- A discrete uniform distribution has equal probability for all integers between two integers a and b (for $a < b$).
- The PMF is

$$p_X(x) = \frac{1}{b - a + 1} \quad \text{for } x = a, a+1, \ldots, b.$$

1.2. Gaussian or Normal



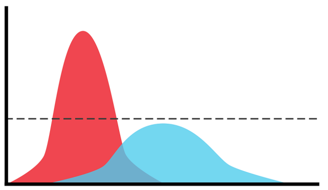
- Probably the most important distribution.
- It has a density that follows a "bell-shaped" curve.
- It is parameterized by its mean

$$-\infty < \mu < \infty$$

and variance

$$\sigma^2 > 0.$$

PDF, Mean, and Variance



- It is denoted as

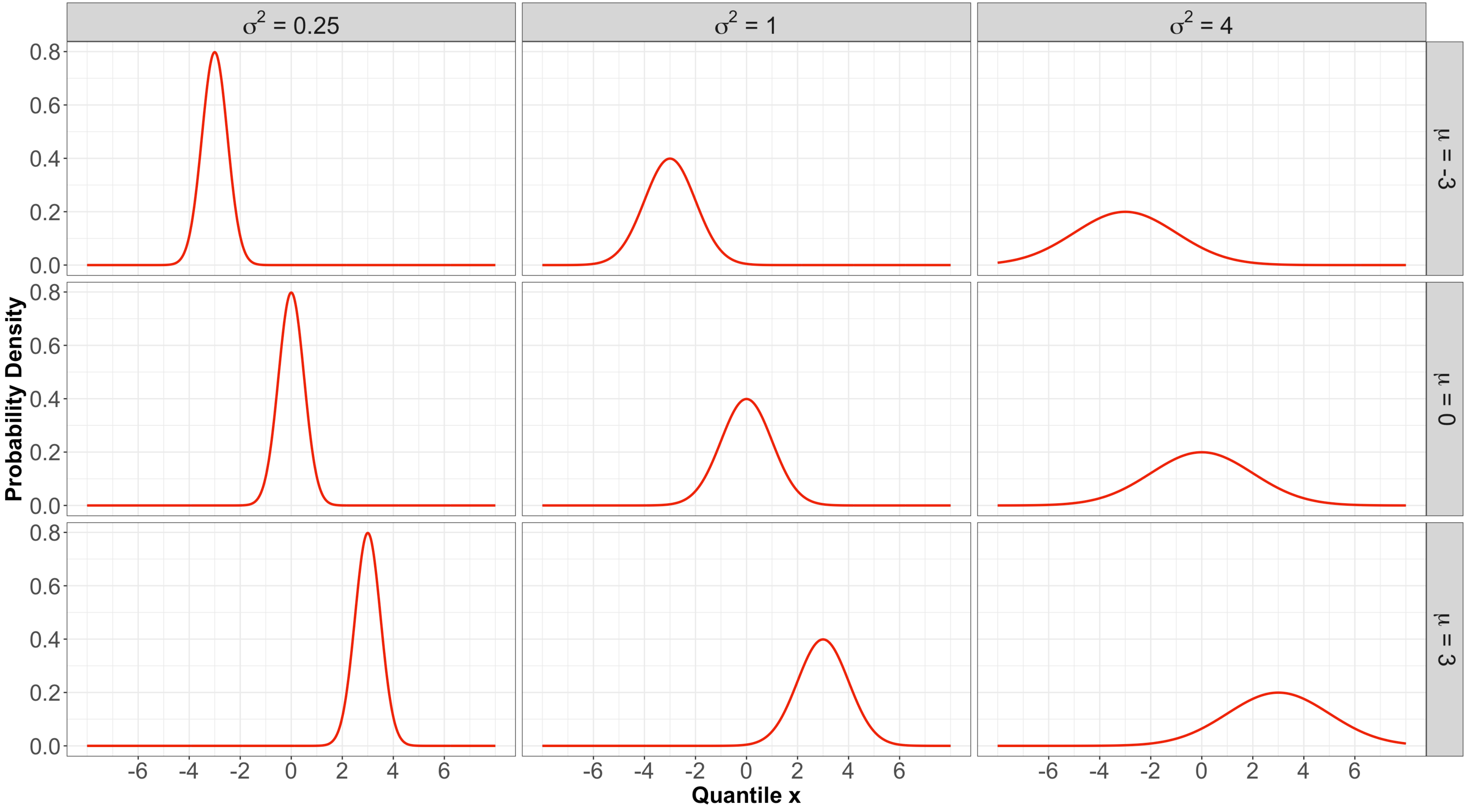
$$X \sim \mathcal{N}\left(\mu, \sigma^2\right).$$

- The PDF is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad \text{for } -\infty < x < \infty.$$

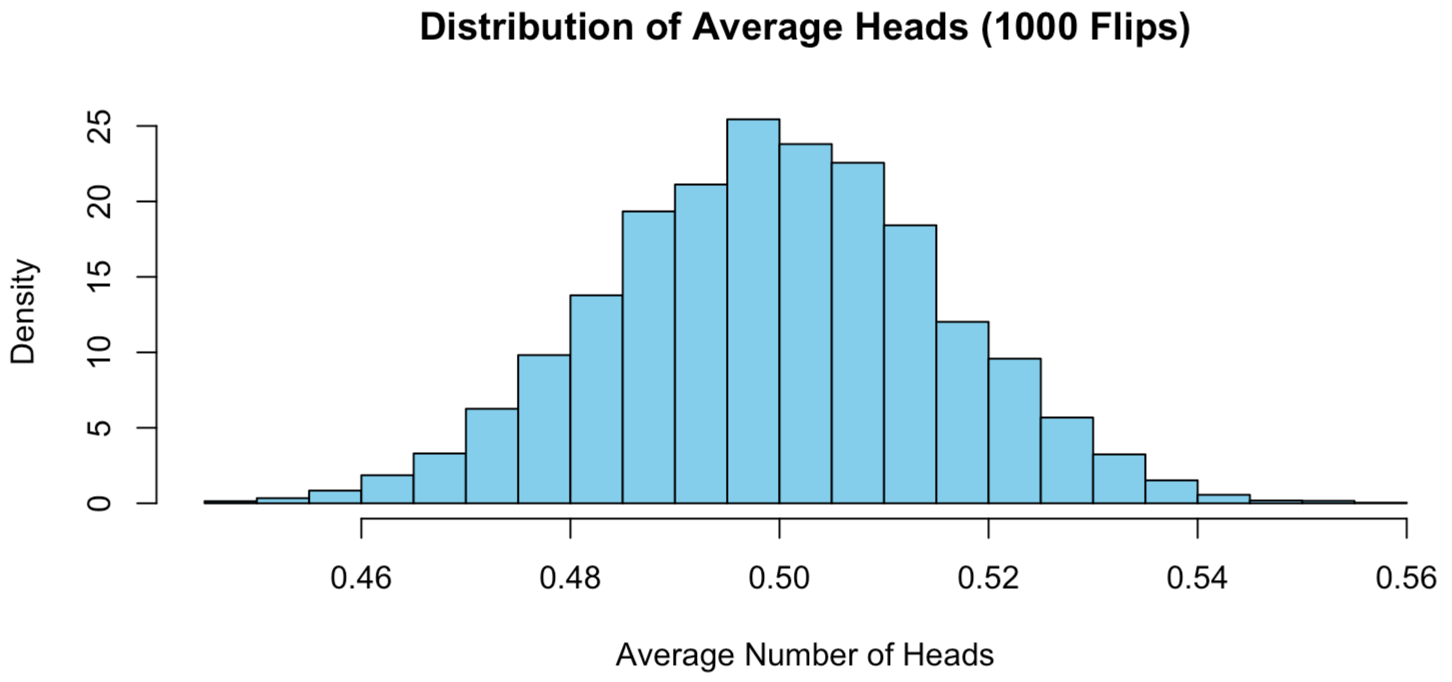
- The mean is $\mathbb{E}(X) = \mu$.
- The variance is $\text{Var}(X) = \sigma^2$.

Some members of the Gaussian or Normal family…



Why is Gaussian distribution important?

- It can be used to approximate the distribution of the average of independently and identically distributed (i.i.d.) random variables.



1.3. Log-Normal

- A random variable X is a Log-Normal distribution if $\log(X)$ follows a normal distribution.
- This family is often parameterized by the mean

$$-\infty < \mu < \infty$$

and variance

$$\sigma^2 > 0$$

of $\log(X)$.

PDF, Mean, and Variance

- It is denoted as

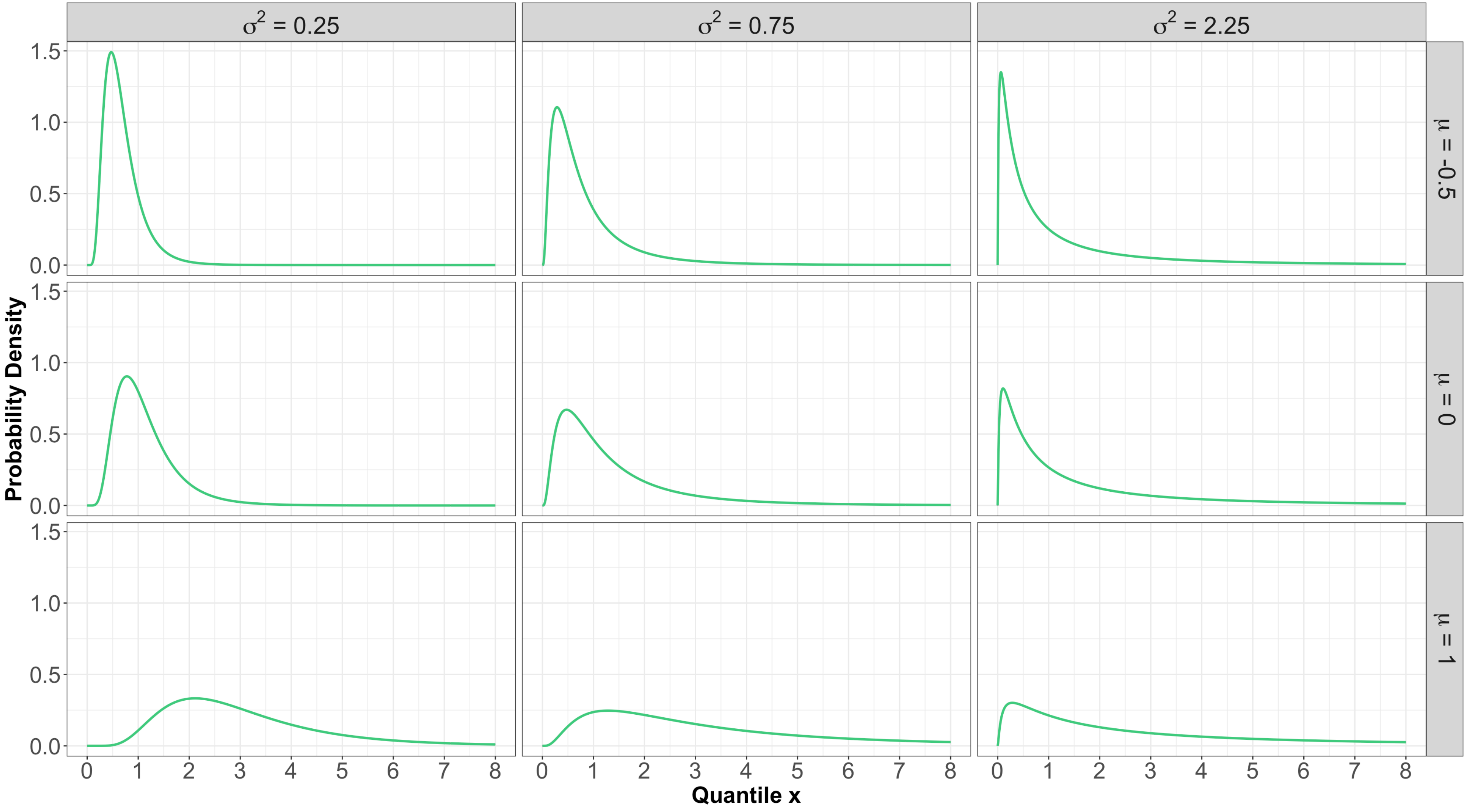
$$X \sim \text{Log-Normal}\left(\mu, \sigma^2\right).$$

- The PDF is

$$f_X(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left[\log(x) - \mu\right]^2}{2\sigma^2}\right\} \quad \text{for } x \geq 0.$$

- The mean is $\mathbb{E}(X) = \exp\left[\mu + \left(\sigma^2/2\right)\right]$.
- The variance is
$\text{Var}(X) = \exp\left[2\left(\mu + \sigma^2\right)\right] - \exp\left(2\mu + \sigma^2\right)$.

Some members of the Log-Normal family…

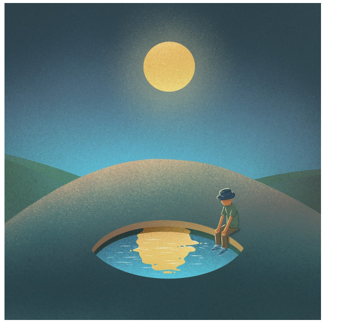


Properties of Log-Normal

- Non-negative support.
- Skewed to the right with heavy tails.

- Suitable for
  - Stock prices
  - Income

1.4. Exponential



- The Exponential distribution is for positive random variables.
- It is often interpreted as wait time for some event to happen.
- The distribution is characterized by a single parameter, usually either the mean wait time $\beta > 0$, or the average rate $\lambda > 0$ at which events happen.

Definition

- The Exponential family is denoted as

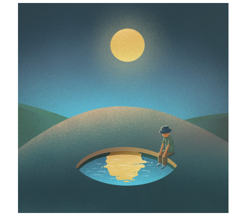
$$X \sim \text{Exponential}(\beta),$$

or

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

- where $\beta = 1/\lambda$.

PDFs



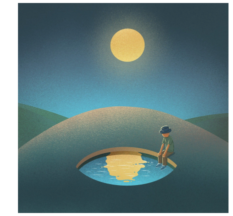
- The PDF can be parameterized as

$$f_X(x) = \frac{1}{\beta} \exp(-x/\beta) \qquad \text{for} \quad x \geq 0$$

or

$$f_X(x) = \lambda \exp(-\lambda x) \qquad \text{for} \quad x \geq 0.$$

Mean



- Using a β parameterization, the mean is:

$$\mathbb{E}(X) = \beta.$$

- On the other hand, using a λ parameterization, the mean is:

$$\mathbb{E}(X) = 1/\lambda.$$

Variance

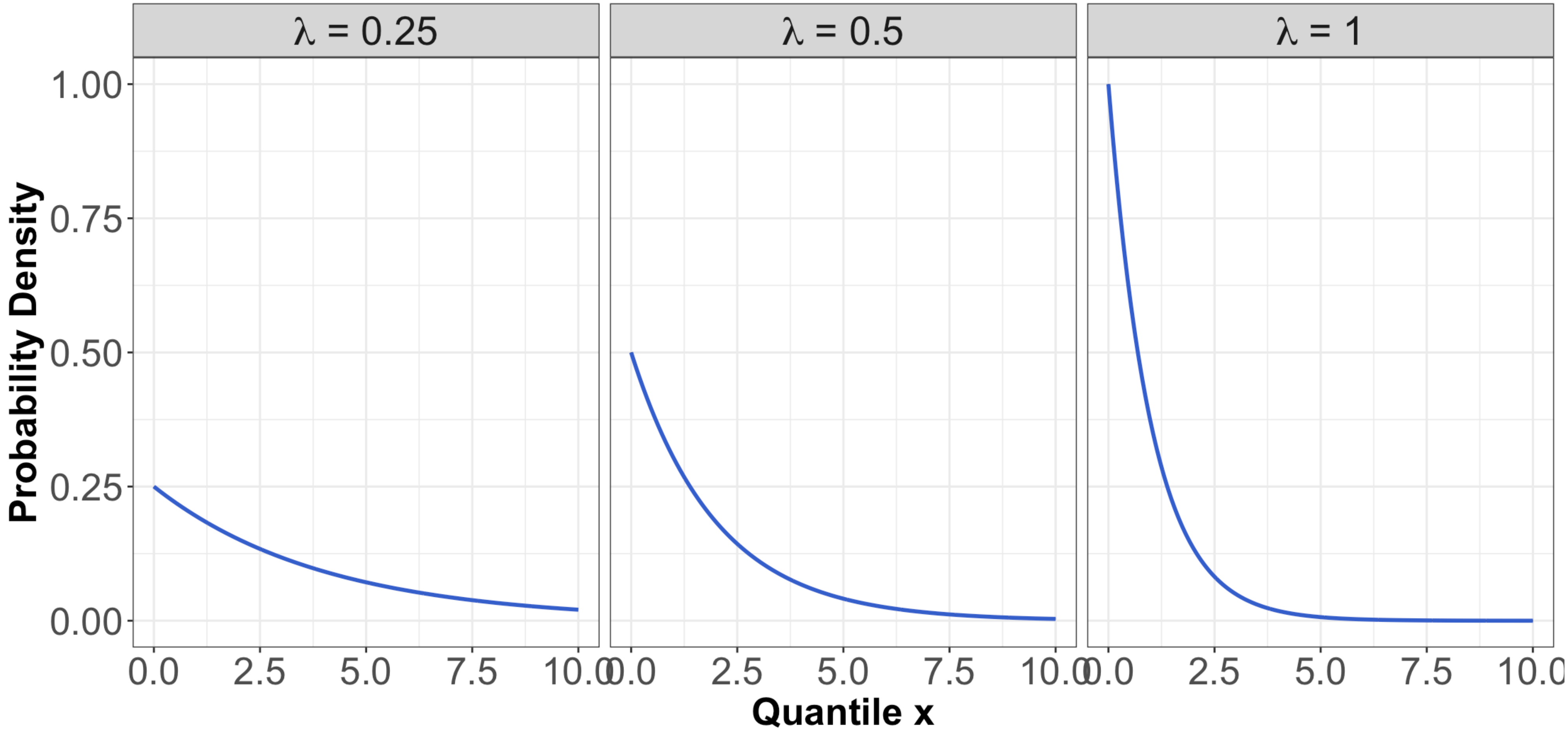
- Using a β parameterization, the variance is:

$$\mathrm{Var}(X) = \beta^2.$$

- On the other hand, using a λ parameterization, the variance is:

$$\mathrm{Var}(X) = 1/\lambda^2.$$

Some examples of Exponential distributions



Memoryless Property

- The probability of an event occurring in the next time period does not depend on how much time has already elapsed.
- Suppose $X \sim \mathrm{Exponential}(\lambda)$, then

$$P(X > s + t | X > t) = P(X > s).$$

1.5. Beta

- The Beta family of distributions is defined for random variables taking values between 0 and 1.
- Hence, it is useful for modelling the distribution of proportions or probabilities.
- It is characterized by two positive shape parameters, $\alpha > 0$ and $\beta > 0$.

PDF

- It is denoted as

$$X \sim \text{Beta}(\alpha, \beta).$$

- The PDF is given by

$$f_X(x) = \underbrace{\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}}_{\text{normalization constant}} x^{\alpha - 1}(1 - x)^{\beta - 1} \quad \text{for} \quad 0 \leq x \leq$$

where $\Gamma(\cdot)$ is the Gamma function.

Mean and Variance

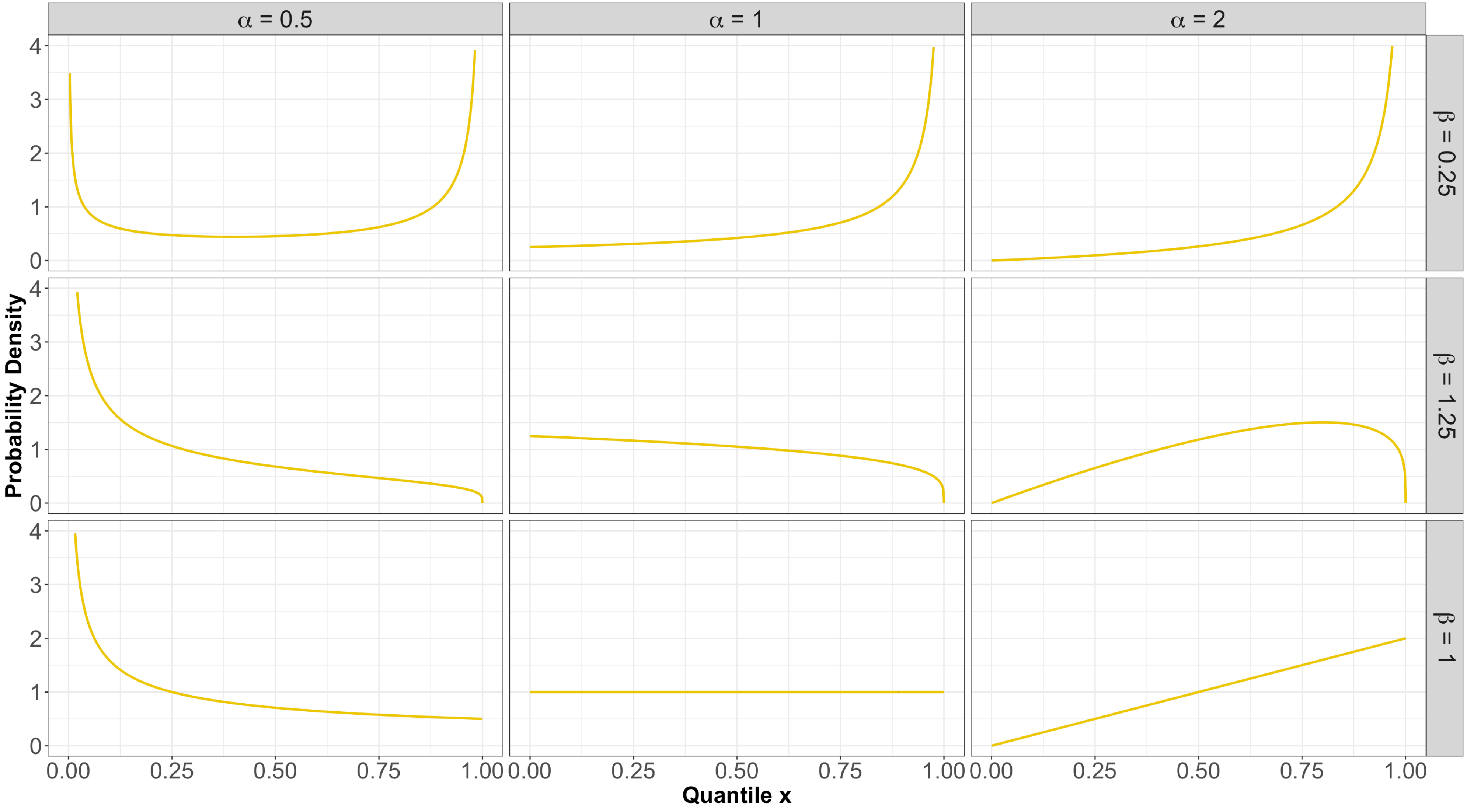
- The mean is

$$\mathbb{E}(X) = \frac{\alpha}{\alpha + \beta}.$$

- The variance is

$$\text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Some members of the Beta family…



1.6. Weibull

- A generalization of the Exponential distribution, which allows for an event to be more likely the longer you wait.
- Because of this flexibility and interpretation, this family is used heavily in survival analysis when modelling time until an event.
- This family is characterized by two parameters, a scale parameter $\lambda > 0$ and a shape parameter $k > 0$.

PDF

- It is denoted as

$$X \sim \mathrm{Weibull}(\lambda, k).$$

- The PDF is

$$f_X(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp^{-(x/\lambda)^k} \qquad \text{for} \quad x \geq 0.$$

- When $k = 1$, the Weibull distribution reduces to the Exponential distribution.

Mean and Variance

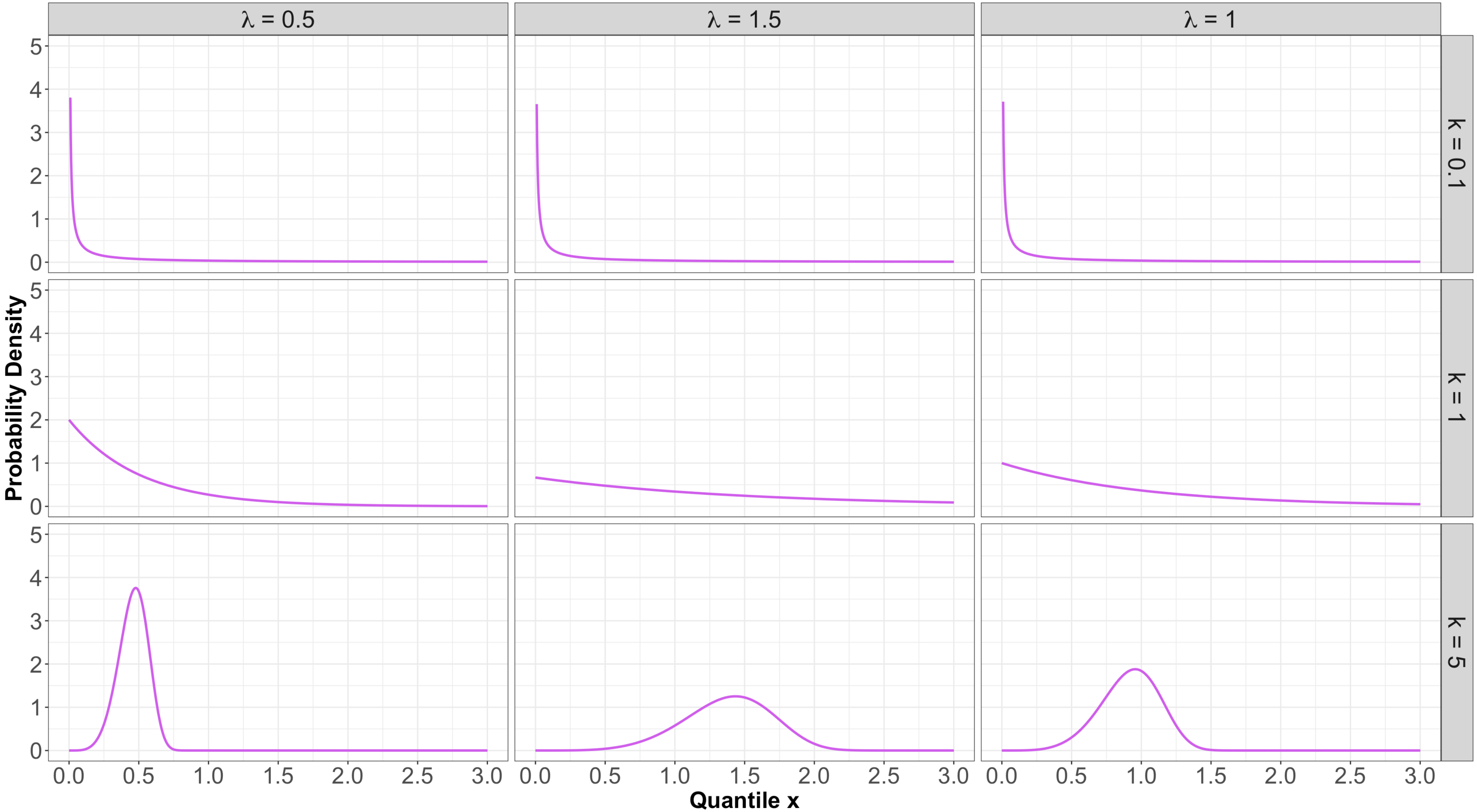
- The mean is

$$\mathbb{E}(X) = \lambda^{1/k}\Gamma\left(1 + \frac{1}{k}\right).$$

- The variance is

$$\mathrm{Var}(X) = \lambda^{2/k}\left[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)\right].$$

Some members of the Weibull family...



1.7. Gamma

- Another useful two-parameter family with support on non-negative numbers.
- One common parameterization is with a shape parameter $k > 0$ and a scale parameter $\theta > 0$.

PDF, Mean, and Variance

- It is denoted as

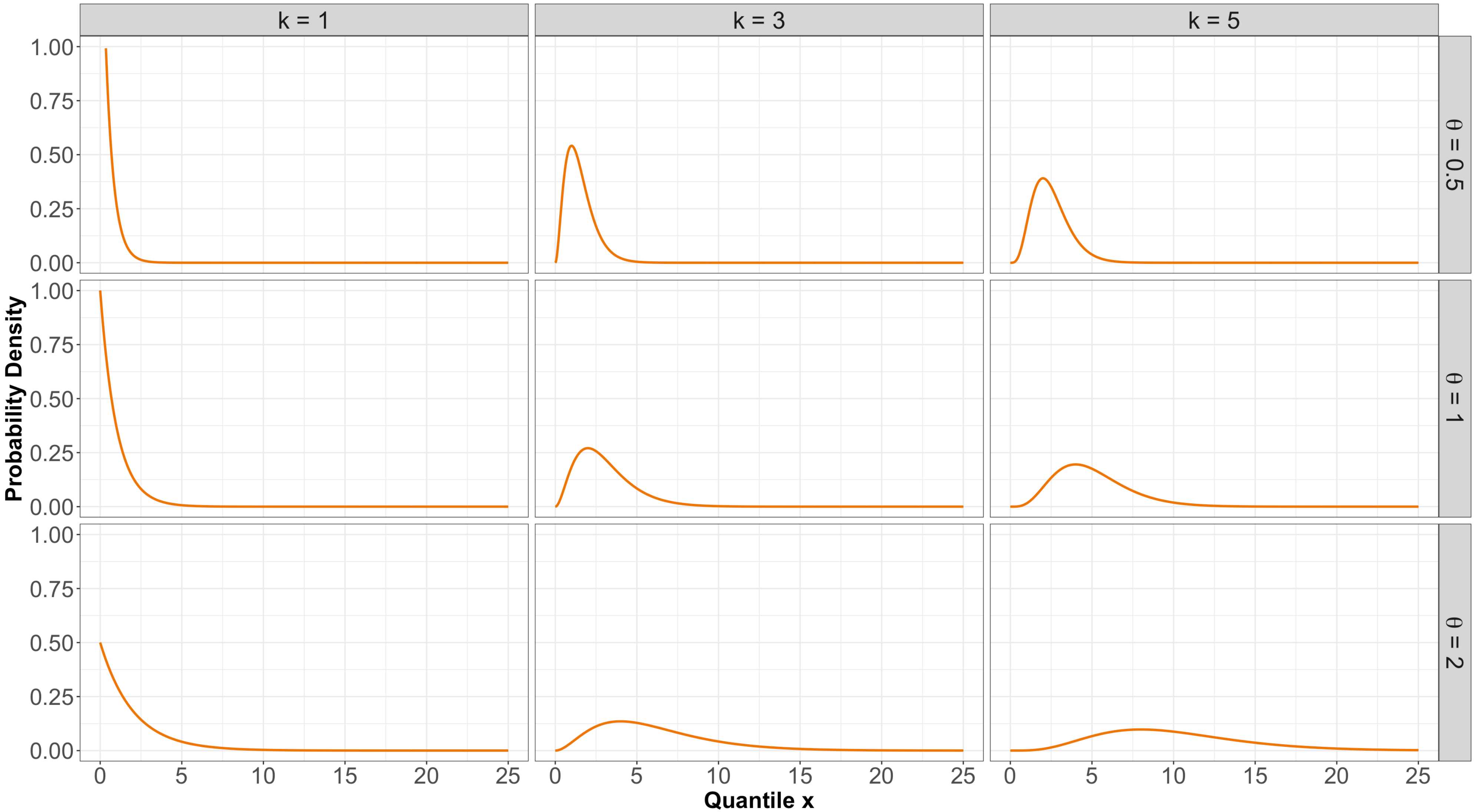
$$X \sim \text{Gamma}(k, \theta).$$

- The PDF is given by

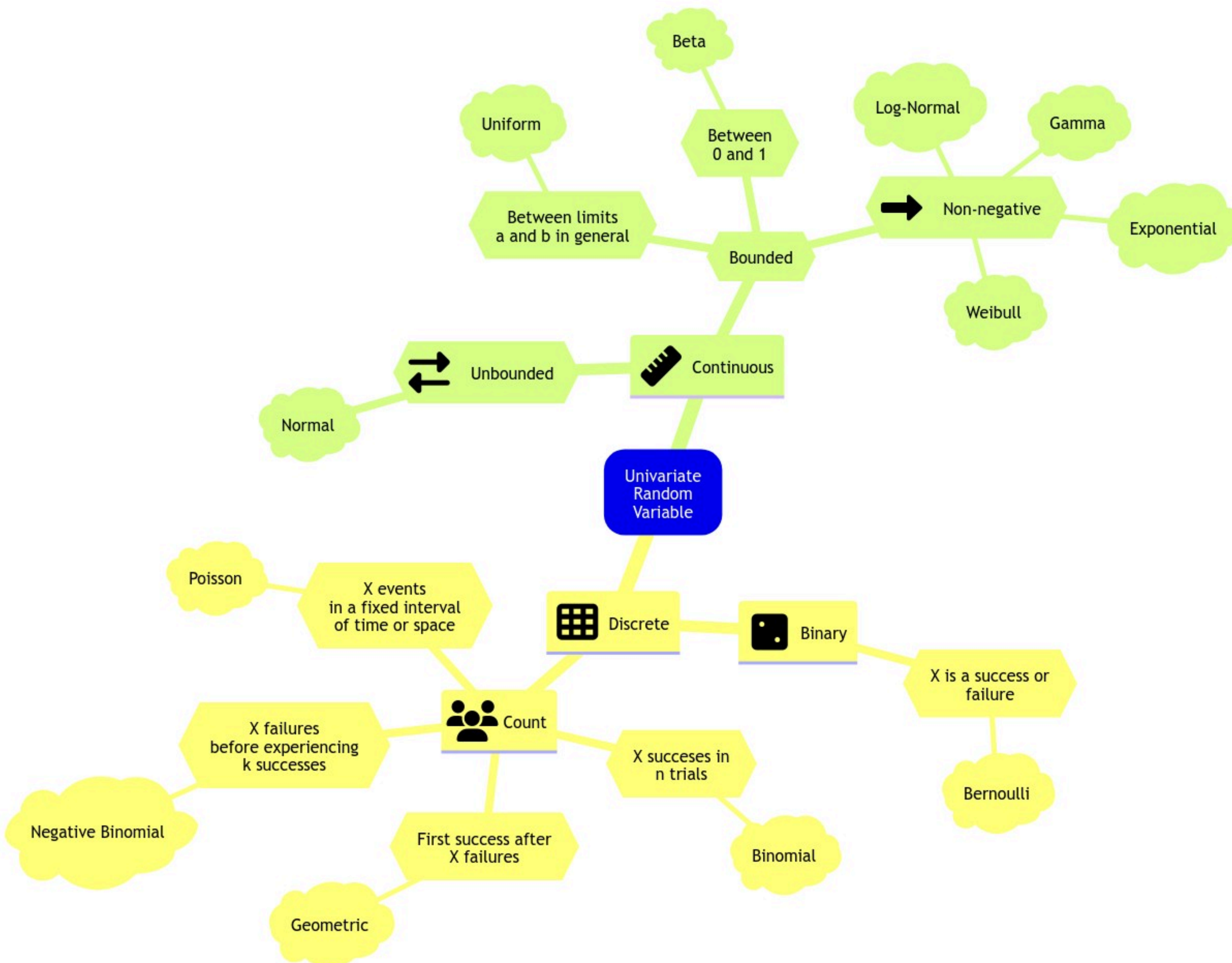
$$f_X(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} \exp(-x/\theta) \qquad \text{for} \quad x \geq 0,$$

- The mean is $\mathbb{E}(X) = k\theta$.
- The variance is $\text{Var}(X) = k\theta^2$.

Some members of the Gamma family...



Summarizing!



1.8 Relevant R Functions

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

Abbreviations for <dist>

- `unif`: Uniform (continuous).
- `norm`: Normal (continuous).
- `lnorm`: Log-Normal (continuous).
- `exp`: Exponential (continuous).
- `geom`: Geometric (discrete).
- `pois`: Poisson (discrete).
- `binom`: Binomial (discrete).
- etc.

CDF or Survival function

- We can specify the argument `lower.tail = TRUE` indicates that we want cumulative probabilities beginning on the left-hand side of the PDF.
- If we don't specify `lower.tail`, the default is `TRUE`.

iClicker Question

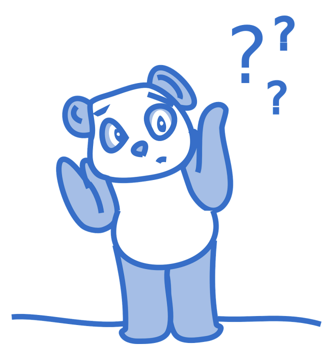
What R function do we need to obtain the density corresponding for $X \sim \mathcal{N}(\mu = 2, \sigma^2 = 4)$ at point $x = 3$? Select the correct option:

A. `pnorm(q = 3, mean = 2, sd = 2)`

B. `dnorm(x = 3, mean = 2, sd = 4)`

C. `pnorm(q = 3, mean = 2, sd = 4)`

D. `dnorm(x = 3, mean = 2, sd = 2)`



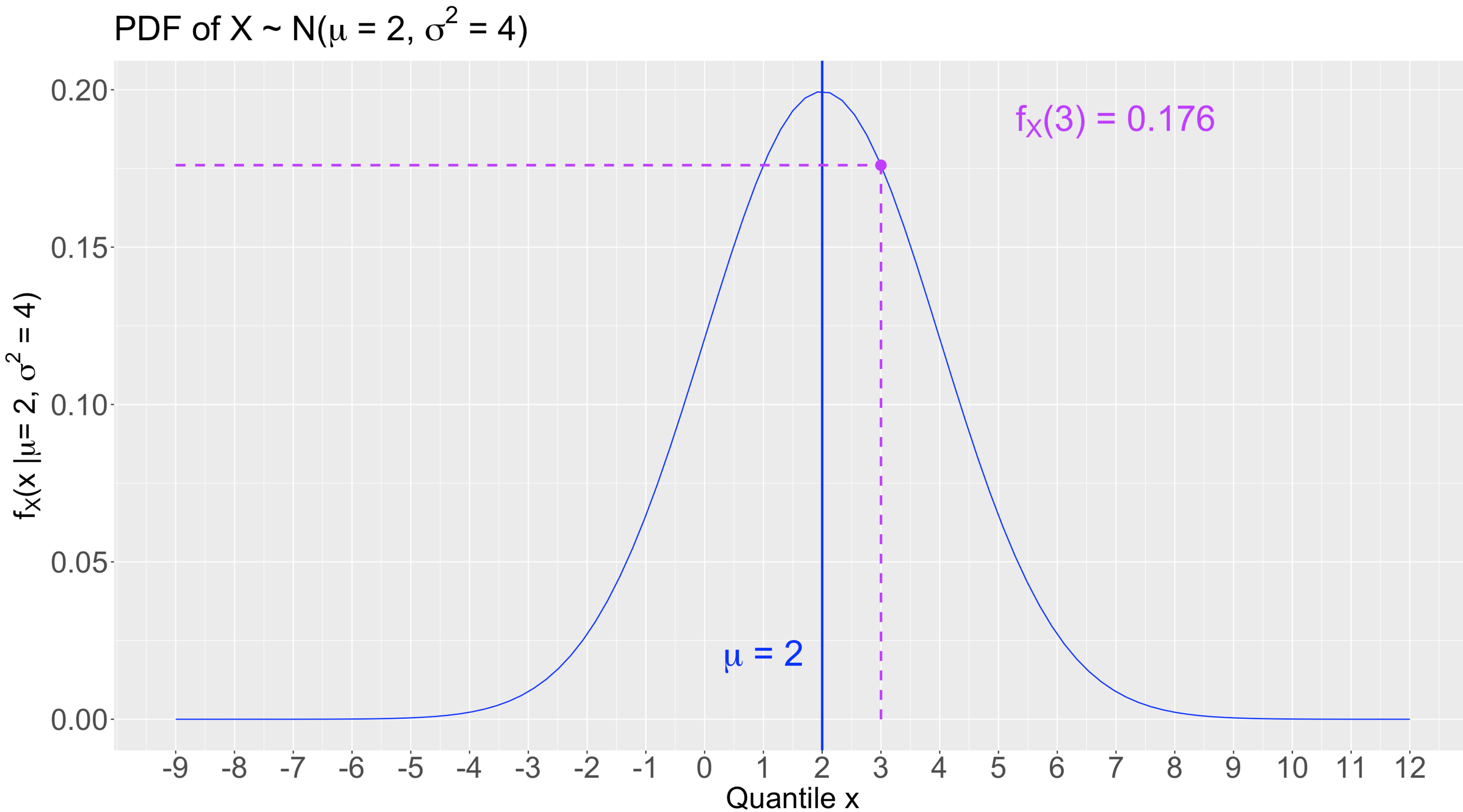
Answer

- We need the density, not the probability.
- Thus, we use `dnorm()` with parameters `mean = 2` (i.e., $\mu = 2$) and `sd = 2` (i.e., $\sigma^2 = 4$).

```
round(dnorm(x = 3, mean = 2, sd = 2), 3)
```

```
[1] 0.176
```

Graphically...



iClicker Question

What R function do we need for the CDF for $X \sim \text{Uniform}(a = 0, b = 2)$ at $x = 0.25, 0.5, 0.75$?

A. `qunif(p = c(0.25, 0.5, 0.75), min = 0, max = 2, lower.tail = TRUE)`

B. `punif(q = c(0.25, 0.5, 0.75), min = 0, max = 2, lower.tail = TRUE)`

C. `dunif(x = c(0.25, 0.5, 0.75), min = 0, max = 2)`

D. `punif(q = c(0.25, 0.5, 0.75), min = 0, max = 2, lower.tail = FALSE)`

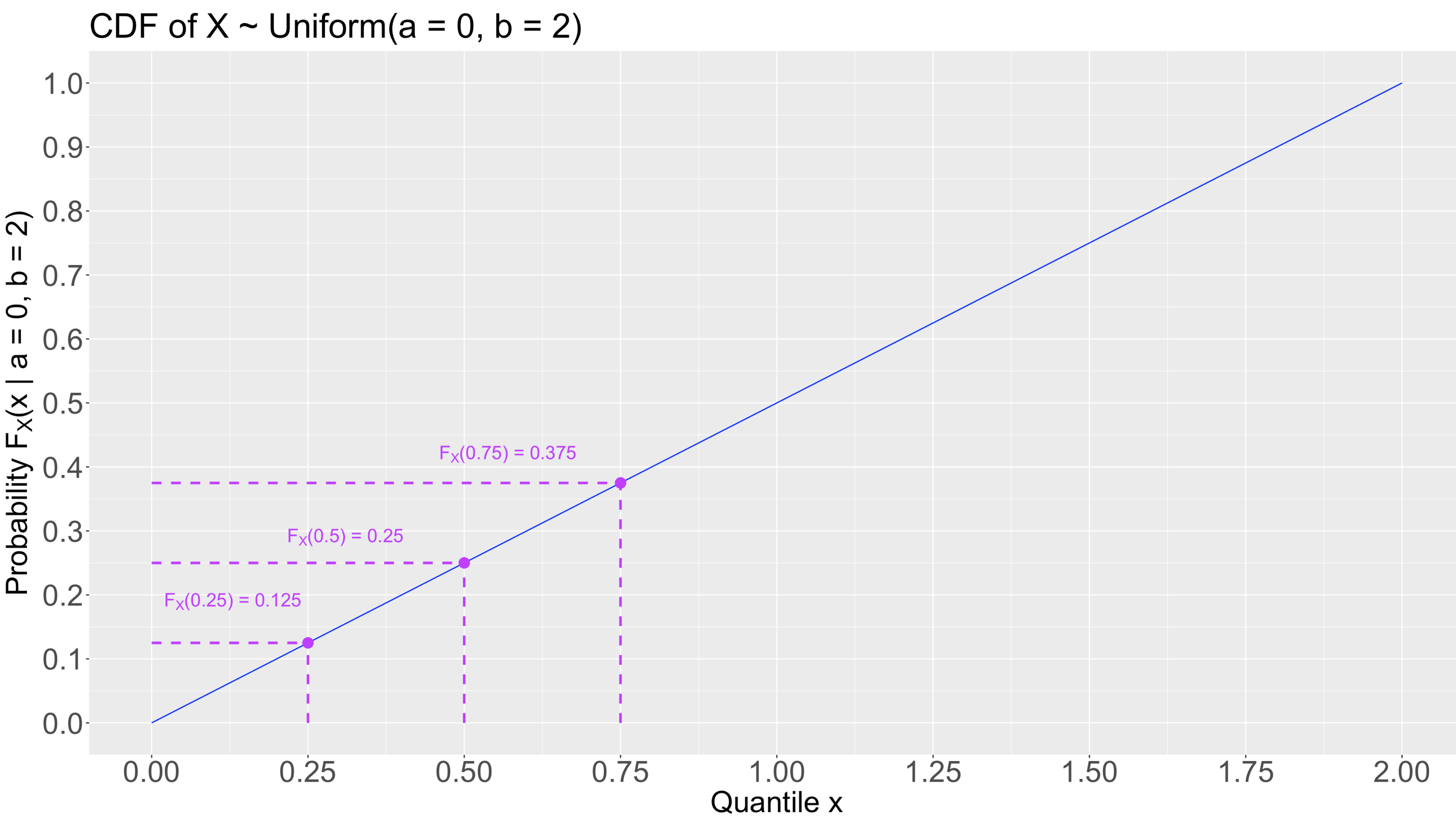
Answer

- We need the values of the CDF at different points on the x-axis, i.e. three different cumulative probabilities.

```r
punif(q = c(0.25, 0.5, 0.75), min = 0, max = 2, lower.tail = TRUE)
```

```
[1] 0.125 0.250 0.375
```

Graphically...



iClicker Question

What R function do we need to obtain the median of $X \sim \text{Uniform}(a = 0, b = 2)$?

A. `qunif(p = 0.5, min = 0, max = 2, lower.tail = TRUE)`

B. `punif(q = 0.5, min = 0, max = 2, lower.tail = TRUE)`

C. `dunif(x = 0.5, min = 0, max = 2)`

D. `punif(q = 0.5, min = 0, max = 2, lower.tail = FALSE)`

Answer

- We need the quantile x of the PDF at which we have 50-50 chance.
- Therefore, with $X \sim \text{Uniform}(a = 0, b = 2)$, we need to use `qunif()`.

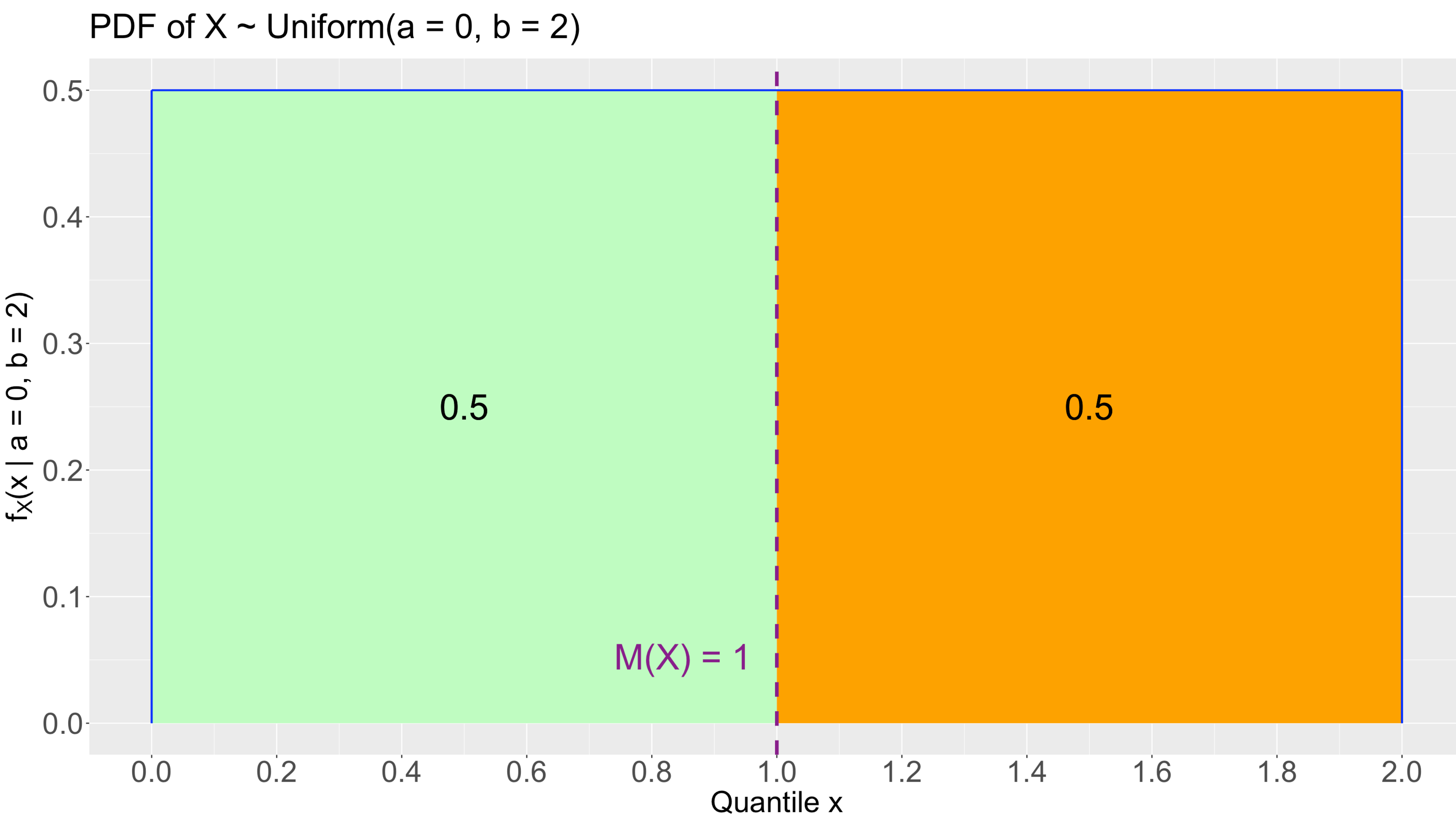
```
qunif(p = 0.5, min = 0, max = 2, lower.tail = TRUE)
```

```
[1] 1
```

```
qunif(p = 0.5, min = 0, max = 2, lower.tail = FALSE)
```

```
[1] 1
```

Graphically...



UBC Master of Data Science

iClicker Question

What R function do we need to generate a random sample of size 10 from the distribution $\mathcal{N}(\mu = 0, \sigma^2 = 25)$?

A. `runif(n = 10, min = 0, max = 25)`

B. `rnorm(n = 10, mean = 0, sd = 25)`

C. `rnorm(n = 10, mean = 0, sd = 5)`

D. `runif(n = 10, min = 0, max = 5)`

Answer

- `rnorm()` is the correct function.
- `n = 10` indicates 10 random numbers.
- The R syntax can be found below, with its output rounded to three decimal places: a vector containing ten random numbers.

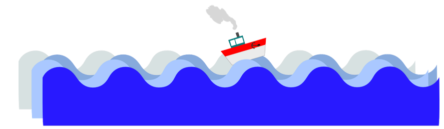
```
set.seed(551) # To ensure reproducibility (a must in simulations!)
round(rnorm(n = 10, mean = 0, sd = 5), 3)
```

```
[1]  4.205  1.188  3.049 -7.458 -2.280  3.930 -5.429 -6.474 -1.001  6.683
```

2. Continuous Joint Distributions

- In the discrete case we have seen joint distributions, conditional distributions, marginal distributions, etc.
- All these concepts are carried over to the continuous world.
- Let us start with two continuous variables (i.e., a bivariate case).

2.1. Continuous Multivariate Distributions

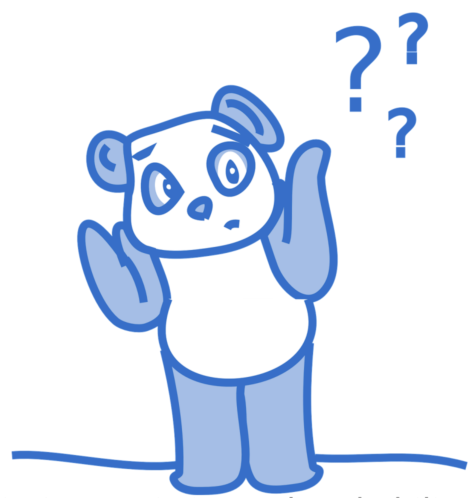


- Recall the joint probability mass function (PMF) can be represented as a table or a data frame:

	Gangs = 1	**Gangs = 2**	**Gangs = 3**	**Gangs = 4**
LOS = 1	0.00170	0.04253	0.12471	0.08106
LOS = 2	0.02664	0.16981	0.13598	0.01757
LOS = 3	0.05109	0.11563	0.03203	0.00125
LOS = 4	0.04653	0.04744	0.00593	0.00010
LOS = 5	0.07404	0.02459	0.00135	0.00002

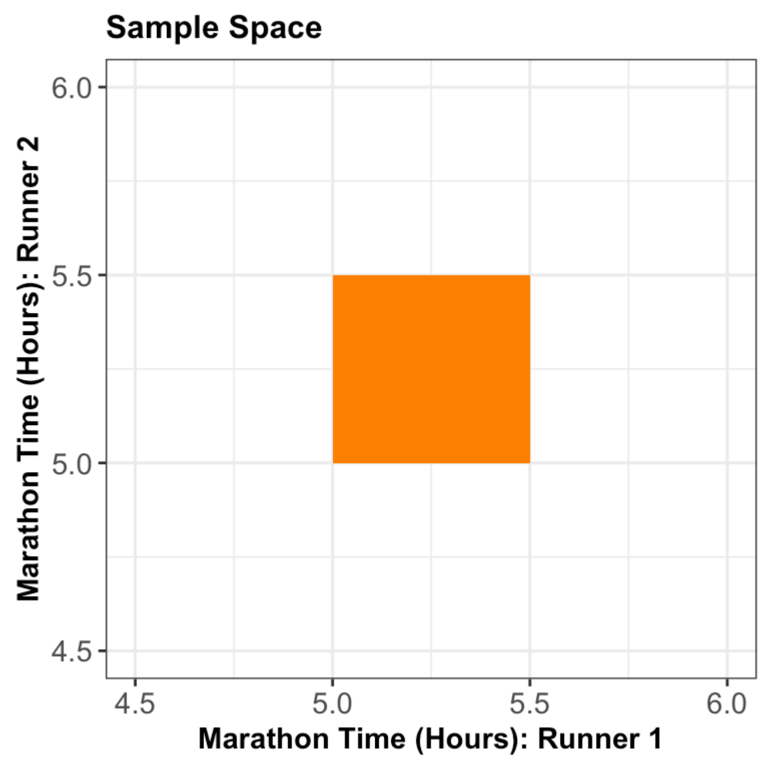
How can we set up a joint PDF?

- Suppose you have two continuous random variables that have a joint PDF.
- For this continuous case, instead of rows and columns, we have an x and y-axis for our two random variables, defining a region of possible values.



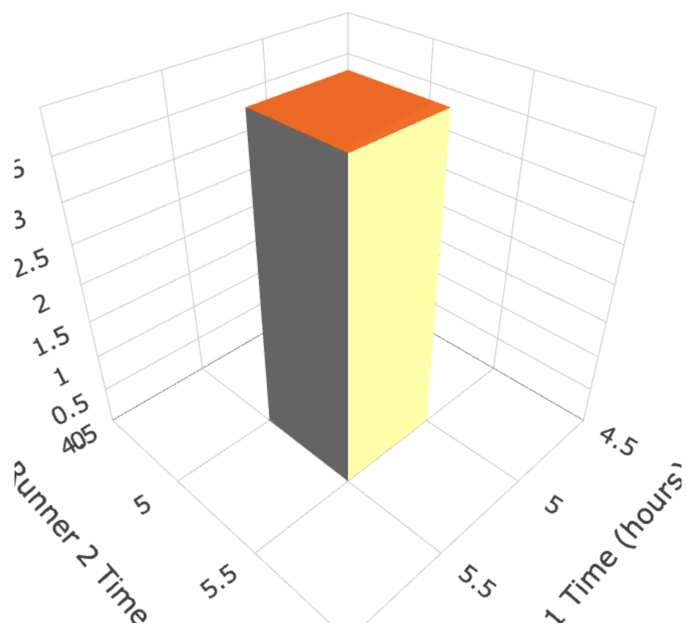
The two-runner example!

- Suppose two marathon runners can only finish a marathon between 5.0 and 5.5 hours each, and their end times are totally random!



A bivariate density function in 3D!

- The density function is a surface overtop of this square.
- The bivariate density function takes two variables and calculates a single density value.



Joint PDF

- Let X and Y be two random variables. Their joint PDF evaluated at the points x and y is usually denoted

$$f_{X,Y}(x, y).$$

- Example: if both X and Y follow $\mathrm{Uniform}(0, 1)$, then

$$f_{X,Y}(x, y) = \begin{cases} 1, & \text{for } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Normalization condition

- The total volume under the density function must equal 1.
- Formally, this may be written as

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, \mathrm{d}x \, \mathrm{d}y = 1.$$

2.2. Calculating Probabilities from Joint PDF

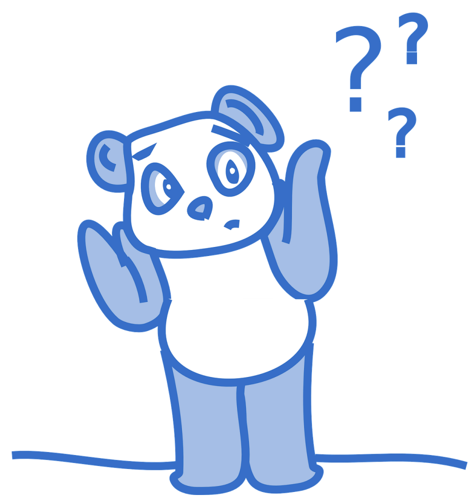
- Let X and Y be two random variables with joint PDF $f_{X,Y}(x, y)$.
- We can calculate probabilities by integrating the joint PDF:

$$P(a_1 < X < a_2, b_1 < Y < b_2) = \int_{a1}^{a_2} \int_{b_1}^{b_2} f_{X,Y}(x, y)\,\mathrm{d}y\,\mathrm{d}x$$

- If the joint PDF is constant over a region, it may be easier to calculate probabilities directly using geometric methods (e.g., finding the area of the region).

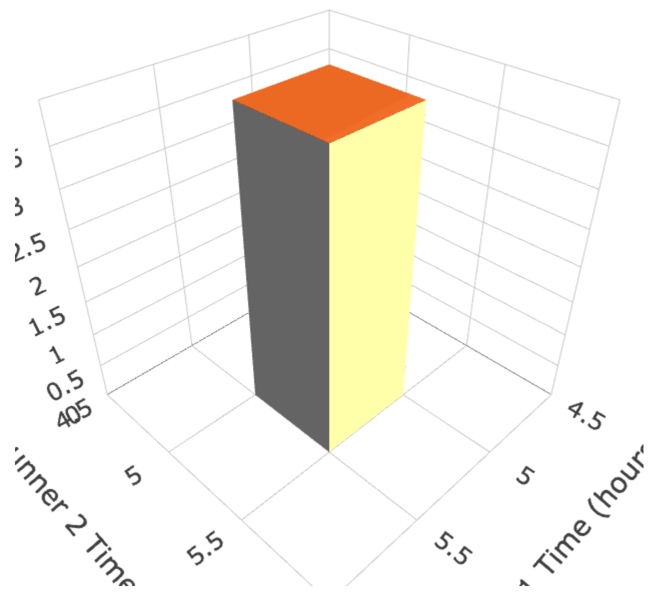
In-Class Question

- In the two-runner example, if the density is constant the entire sample space, what is the height of this surface?
- That is, what does the density evaluate to?
- What does it evaluate to outside of the sample space?



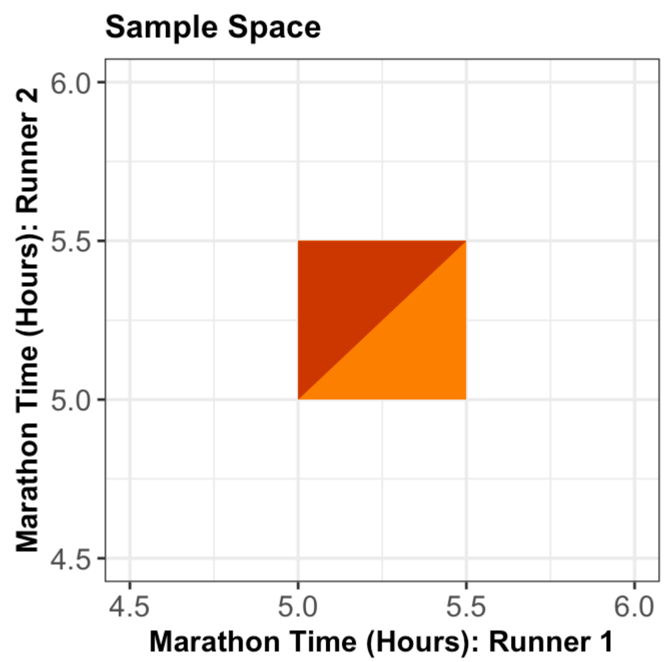
Answer

- The height of the surface on the z-axis is equal to 4, which ensures that the volume under the density function is equal to 1.
- The probability of any point outside of this prism is zero.



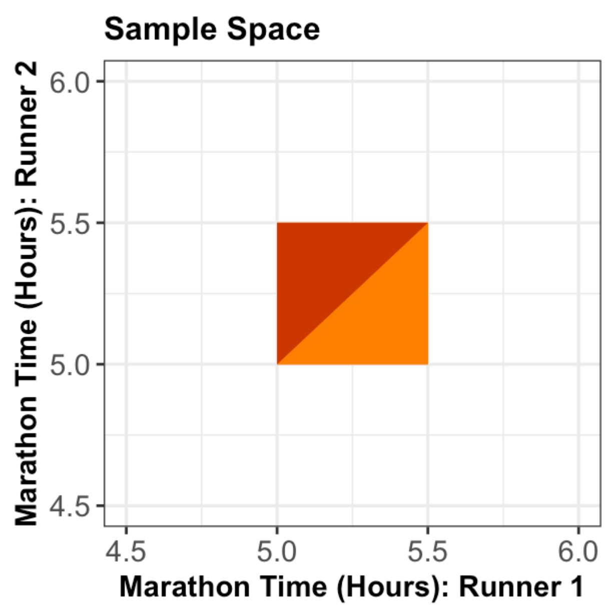
In-Class Question

- Let X and Y be the marathon times of Runner 1 and Runner 2 in hours, respectively.
- What is the probability that Runner 1 will finish the marathon before Runner 2, i.e., $P(X < Y)$?



Answer

- Identify the surface region in the sample's outcome space.



- Calculate the volume of the space overtop of this region, $P(X < Y) = \frac{1}{2}$.

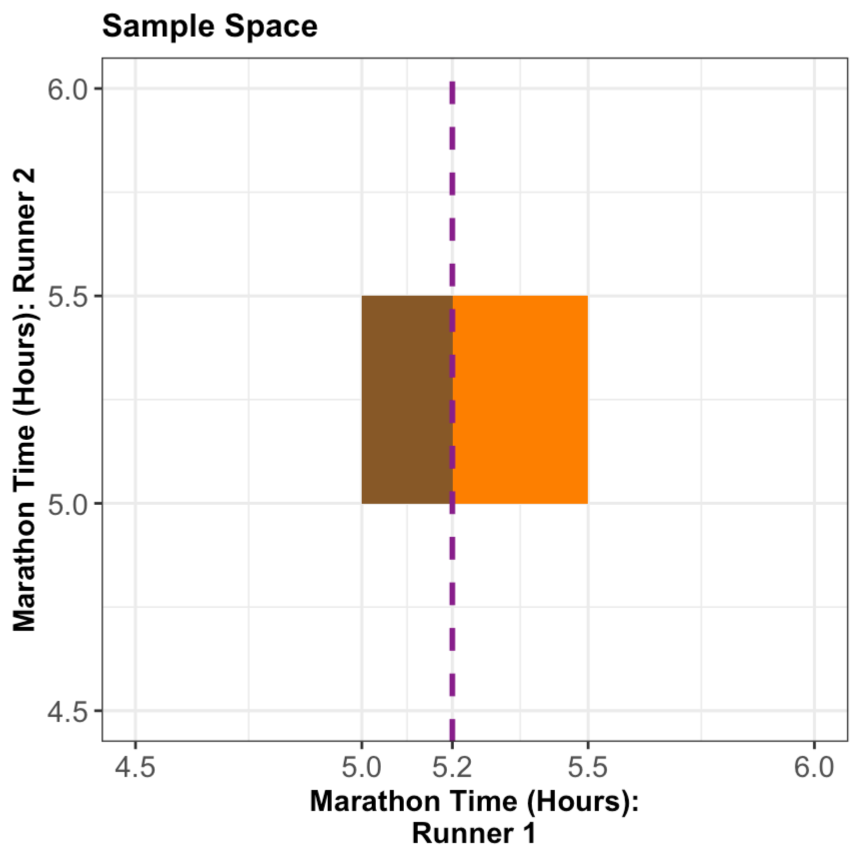
Alternative Answer

- We can also use the joint PDF to calculate the probabilities:

$$
\begin{aligned}
\int_{5}^{5.5} \int_{x}^{5.5} 4 \, \mathrm{d}y \, \mathrm{d}x &= \int_{5}^{5.5} 4 \cdot (5.5 - x) \, \mathrm{d}x \\
&= \int_{5}^{5.5} 4 \cdot 5.5 \, \mathrm{d}x - \int_{5}^{5.5} 4x \, \mathrm{d}x \\
&= 11 - 2x^2 \Big|_{5}^{5.5} \\
&= 11 - (2 \cdot (5.5)^2 - 2 \cdot (5.0)^2) = 0.5
\end{aligned}
$$

In-Class Question

- What is the probability that Runner 1 finishes in 5.2 hours or less, i.e., $P(X \leq 5.2)$?



Answer

- The question only asks about Runner 1.
- We do not take into account any specific time limits for Runner 2's time Y.
- Identify the surface region in the sample's outcome space.
- Calculate the volume of the space overtop of this region:

$$P(X \leq 5.2) = \underbrace{[(5.2 - 5.0) \times (5.5 - 5.0)]}_{\text{Brown Area}} \times \underbrace{4}_{\text{Prism's Height}} = 0.4.$$

- We are actually computing a marginal probability for X.

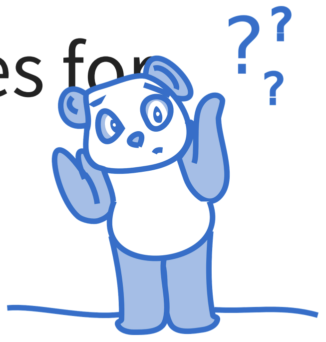
Alternative Answer

- Again we can also use the joint PDF to calculate the probabilities:

$$\int_{5}^{5.2} \int_{5}^{5.5} 4 \, \mathrm{d}y \, \mathrm{d}x = \int_{5}^{5.2} 0.5 \cdot 4 \, \mathrm{d}x$$
$$= 0.2 \cdot 0.5 \cdot 4 = 0.4$$

3. Continuous Conditional Distributions

- Recall the basic formula for conditional probabilities for events A and B:



$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

3.1. When $P(A) = 0$

- Suppose the month is half-way over, and you find that you have spent $\$2500$ so far!
- What is the distribution of this month's total expenditures now, given this information?

Applying the previous Law of Conditional Probability...

- Let

$$X = \text{Monthly Expenses in CAD.}$$

- Assume

$$X \sim \text{Log-Normal}(\mu = 8, \sigma^2 = 0.5).$$

- Using the conditional probability formula would give:

$$P(X = x \mid X \geq 2500) = \frac{P(X = x)}{P(X \geq 2500)} \qquad \text{(NO!)}$$

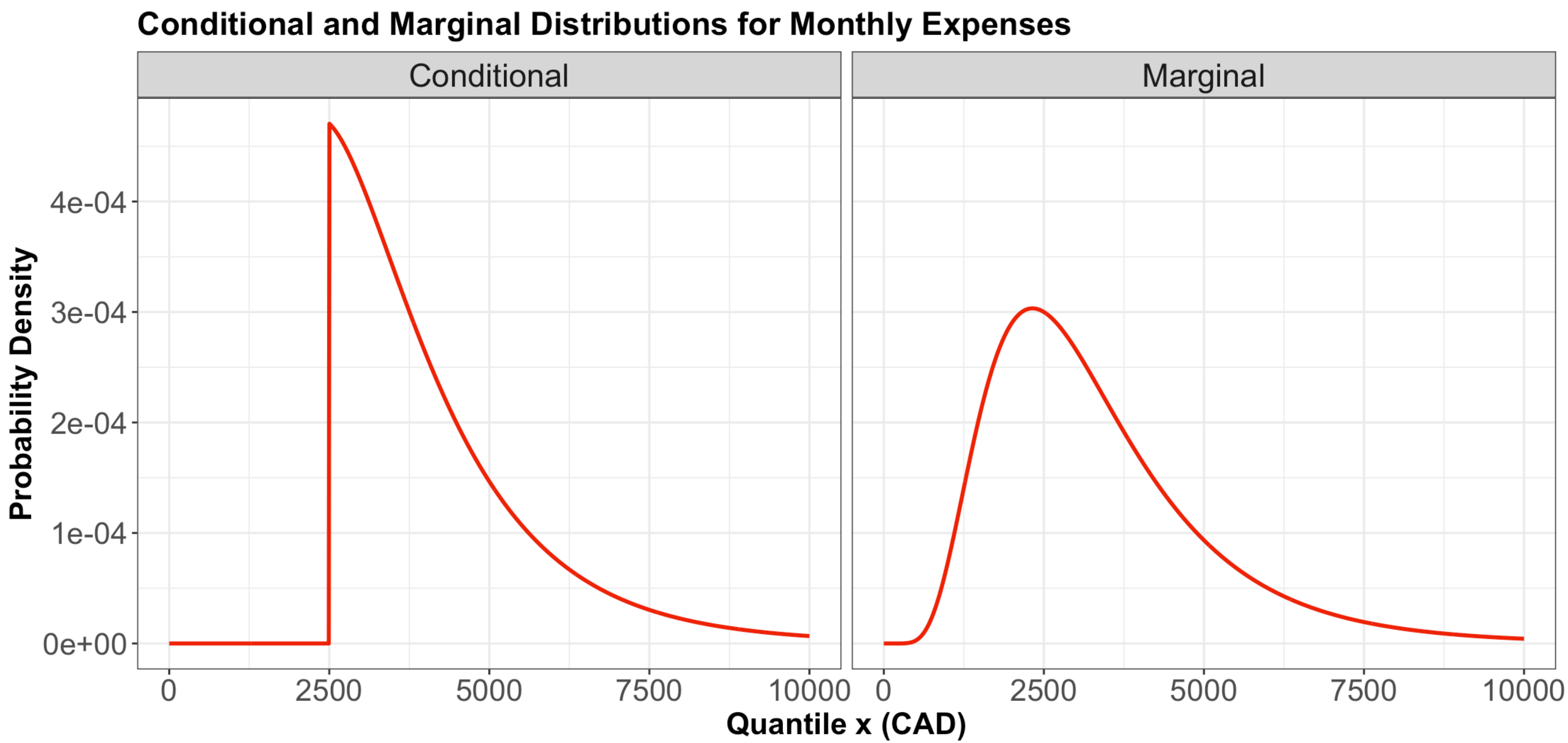
Instead...

- In general, we replace probabilities with densities.
- In this case, what we actually have is:

$$f_{X|X\geq 2500}(x) = \begin{cases} \frac{f_X(x)}{P(X\geq 2500)} & \text{for} \quad x \geq 2500 \\ 0 & \text{for} \quad x < 2500. \end{cases}$$

- Isn't it strange to divide a density by a probability?
  - The division simply re-normalizes the density function to ensure that the total probability in the region $X \geq 2500$ sums to 1.

Comparing Densities



3.2. When $P(B) = 0$

- To describe this situation, let us use the marathon runners' example again:

> **If Runner 1 ended up finishing in 5.2 hours**, what is the distribution of Runner 2's time?

- Let X be the time for Runner 1, and Y for Runner 2, what we are asking for is

$$f_{Y|X=5.2}(y).$$

However...

- Note that $P(X = 5.2) = 0$.

- Again we can replace probabilities with densities.

$$f_{Y|X=5.2}(y) = \frac{f_{X,Y}(5.2, y)}{f_X(5.2)}.$$

Conditional Density and Marginal Density

- This formula is true in general:

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

  where

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}y.$$

- We can also write $f_{Y|X}(y|x)$.

Today's Learning Goals

By the end of this lecture, we will be able to

- Identify and apply common continuous distribution families.
- Identify what makes a function a bivariate probability density function.
- Calculate probabilities from bivariate probability density functions.
- Compute conditional distributions for continuous random variables.