Lecture 6: Common Distribution Families and Conditioning

Contents

Learning Goals

By the end of this lecture, you should be able to:

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

> **Important**
>
> Let us make a **note on the Greek alphabet**. Throughout MDS, we will use diverse Greek letters in various statistical topics. Usually, these letters represent distributional parameters. That said, it is not necessary to memorize these letters right away. With practice over time, you will get familiar with this alphabet. You can find the whole alphabet in Greek Alphabet.

1. Common Continuous Distribution Families

Just like for discrete distributions, there are also parametric families of continuous distributions. Recall the chart of univariate distributions, where you will find key information, such as their corresponding probability density functions (PDFs). Furthermore, the chart illustrates how these distributions are related via random variable transformations.

1.1. Uniform

Process

A Uniform distribution has equal density in between two points a and b (for $a < b$), and is usually denoted by

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

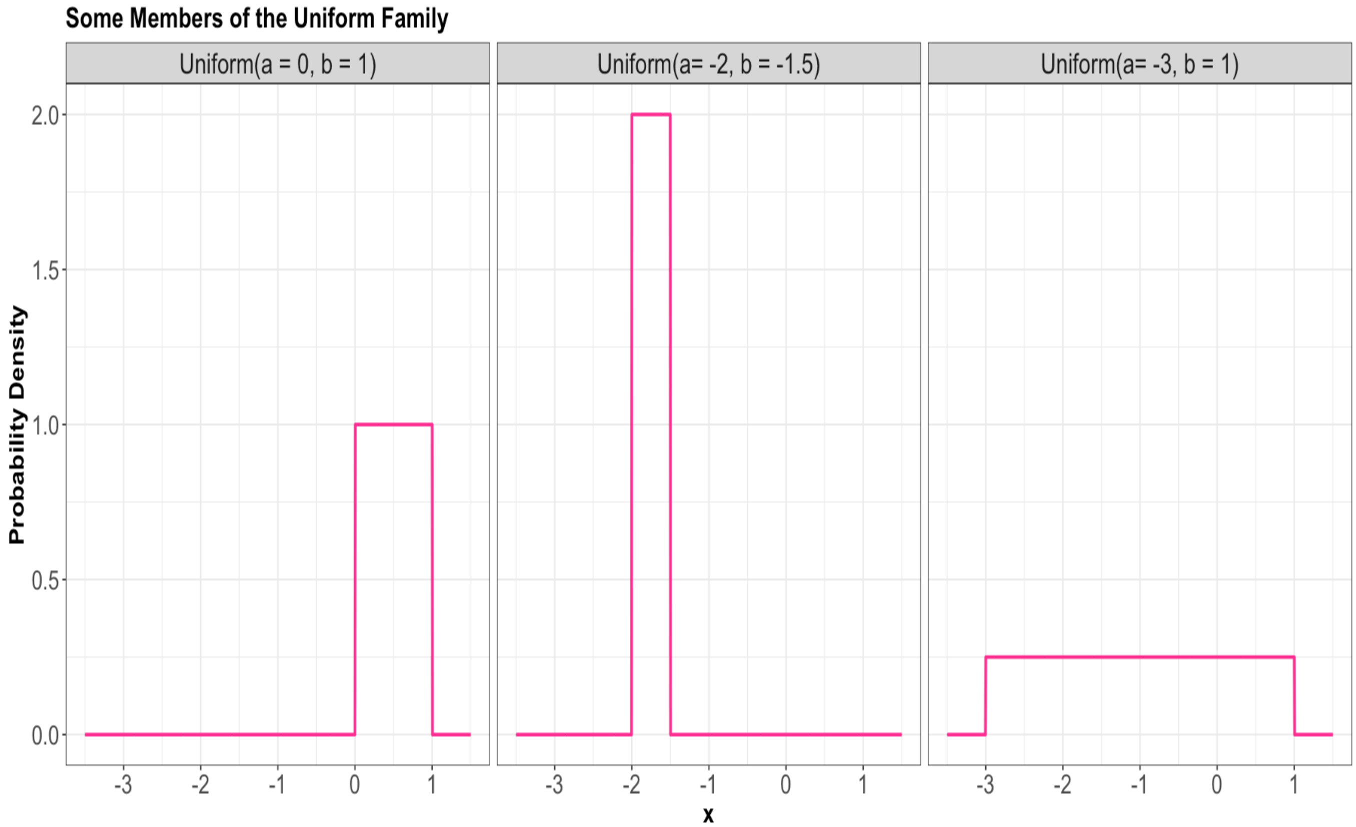
That means that there are **two parameters**: **one for each end-point**. A reference to a "*Uniform distribution*" usually implies **continuous uniform**, as opposed to discrete uniform.

PDF

The density is

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

Here are some densities from members of this family:



Mean

The mean of a Uniform random variable is defined as:

$$\mathbb{E}(X) = \frac{a+b}{2}.$$

Variance

The variance of a Uniform random variable is defined as:

$$\mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

1.2. Gaussian or Normal

Process

Probably the most famous family of distributions. It has a density that follows a **"bell-shaped"** curve. It is parameterized by its mean $-\infty < \mu < \infty$ and variance $\sigma^2 > 0$. A Normal distribution is usually denoted as

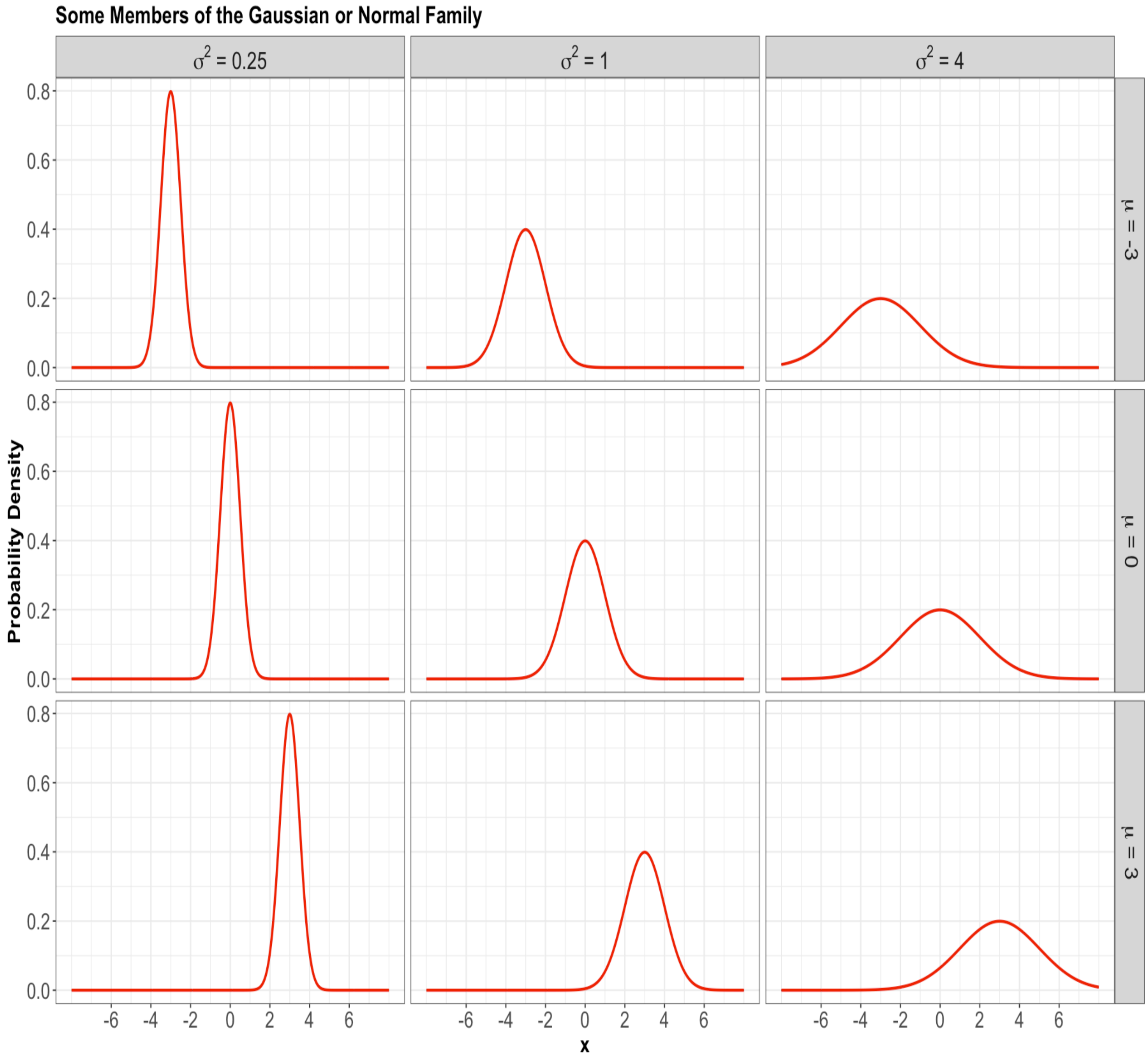
$$X \sim \mathcal{N}\left(\mu, \sigma^2\right).$$

PDF

The density is

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

Here are some densities from members of this family:



Mean

The mean of a Normal random variable is defined as:

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

Variance

The variance of a Normal random variable is defined as:

$$\operatorname{Var}(X) = \sigma^2.$$

1.3. Log-Normal

Process

A random variable X is a Log-Normal distribution if the transformation $\log(X)$ is Normal. This family is often parameterized by the mean $-\infty < \mu < \infty$ and variance $\sigma^2 > 0$ of $\log X$. The Log-Normal family is denoted as

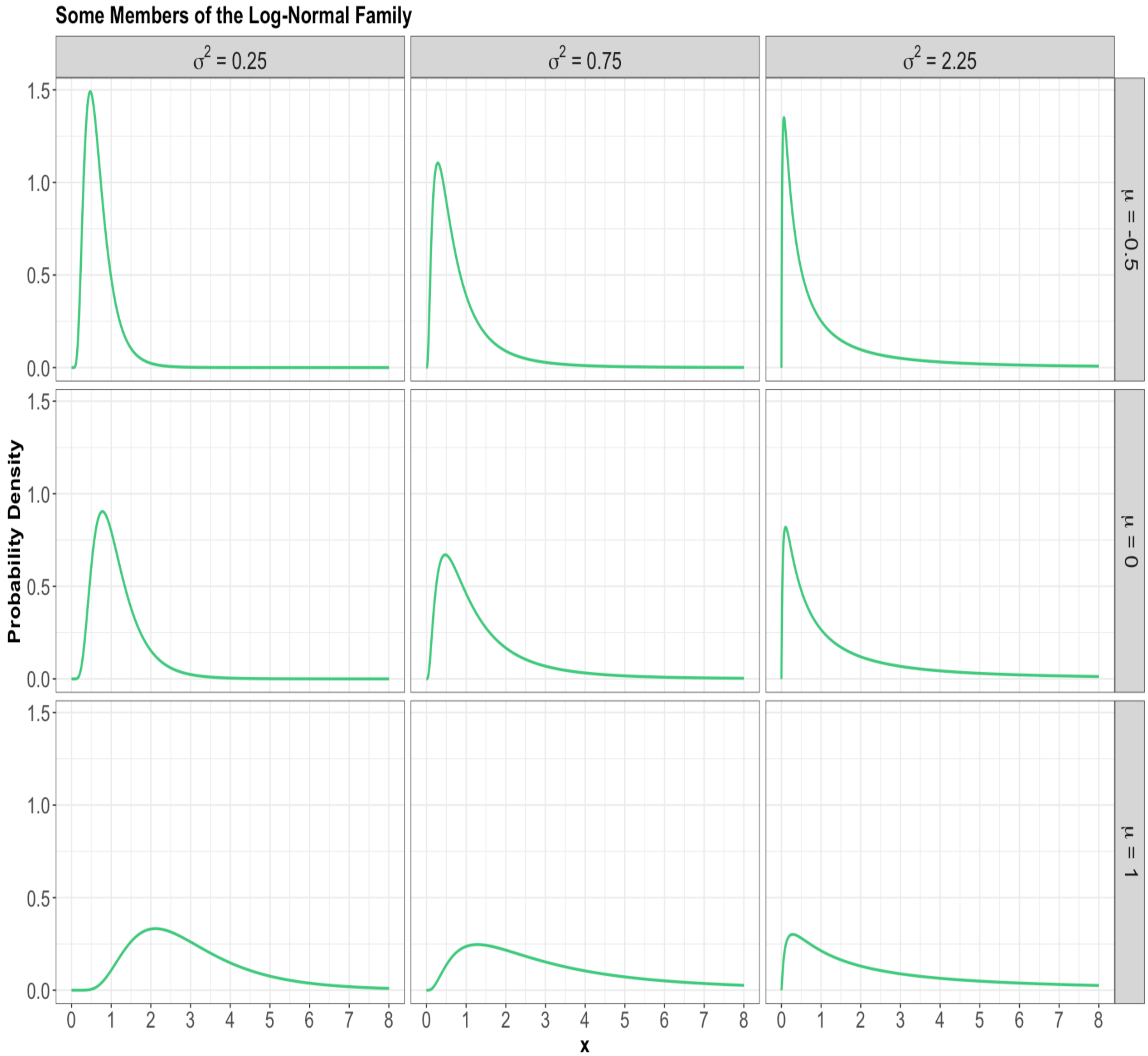
$$X \sim \text{Log-Normal}\left(\mu, \sigma^2\right).$$

PDF

The density is

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

Here are some densities from members of this family:



Mean

The mean of a Log-Normal random variable is defined as:

$$\mathbb{E}(X) = \exp\left[\mu + \left(\sigma^2/2\right)\right].$$

Variance

The variance of a Log-Normal random variable is defined as:

$$\operatorname{Var}(X) = \exp\left[2\left(\mu + \sigma^2\right)\right] - \exp\left(2\mu + \sigma^2\right).$$

1.4. Exponential

Process

The Exponential family is for positive random variables, often interpreted as **wait time** for some event to happen. Characterized by a **memoryless property**, where after waiting for a certain period of time, the remaining wait time has the same distribution.

The family is characterized by a single parameter, usually either the **mean wait time** $\beta > 0$, or its reciprocal, the **average rate** $\lambda > 0$ at which events happen.

The Exponential family is denoted as

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

or

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

PDF

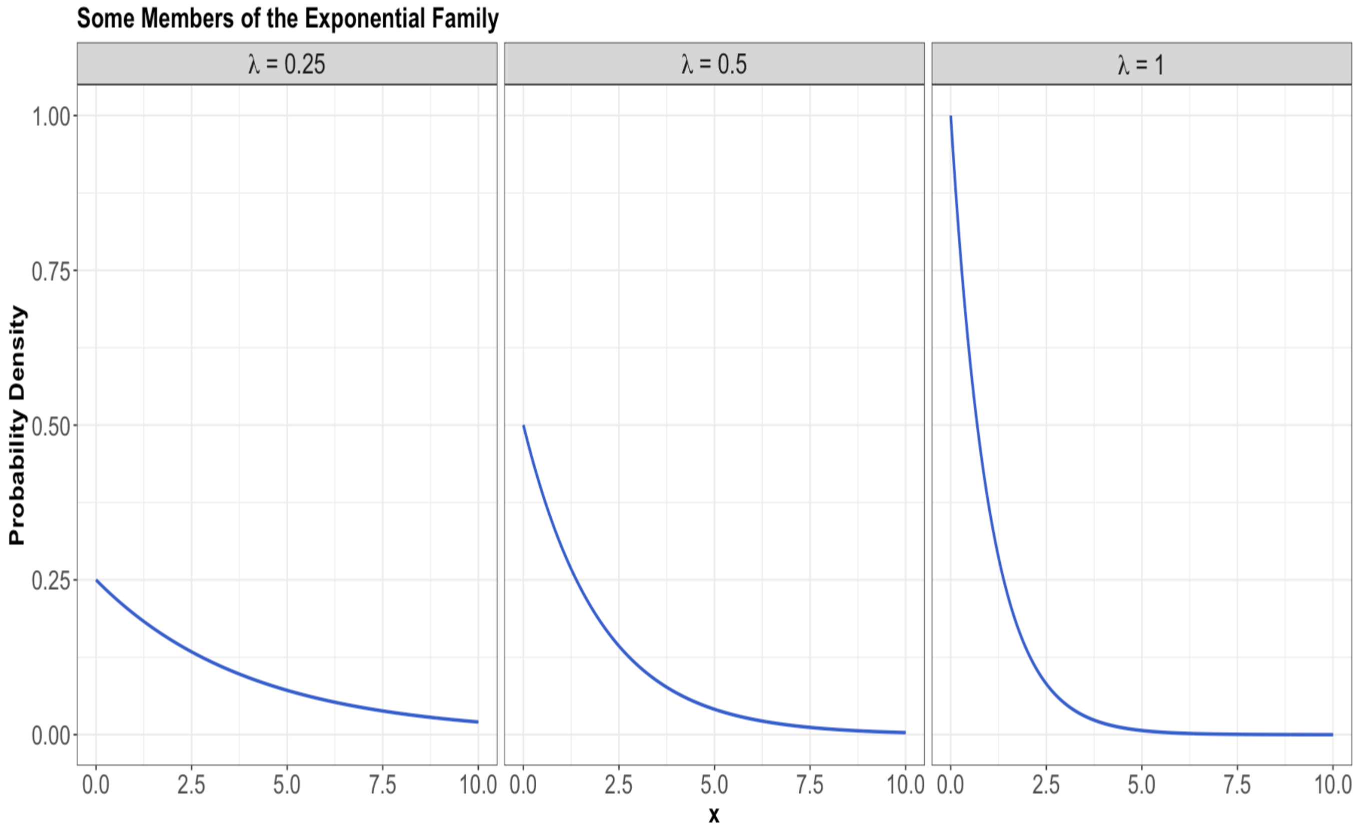
The density can be parameterized as

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

or

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

The densities from this family all decay starting at $x = 0$ for rate λ:



Mean

Using a β parameterization, the mean of an Exponential random variable is defined as:

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

On the other hand, **using a λ parameterization**, the mean of an Exponential random variable is defined as:

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

Variance

Using a β parameterization, the variance of an Exponential random variable is defined as:

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

On the other hand, **using a λ parameterization**, the variance of an Exponential random variable is defined as:

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

1.5. Beta

Process

The Beta family of distributions is defined for random variables taking values between 0 and 1, so is useful for modelling the distribution of proportions. This family is quite flexible, and has the Uniform distribution as a special case. It is characterized by two positive shape parameters, $\alpha > 0$ and $\beta > 0$.

The Beta family is denoted as

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

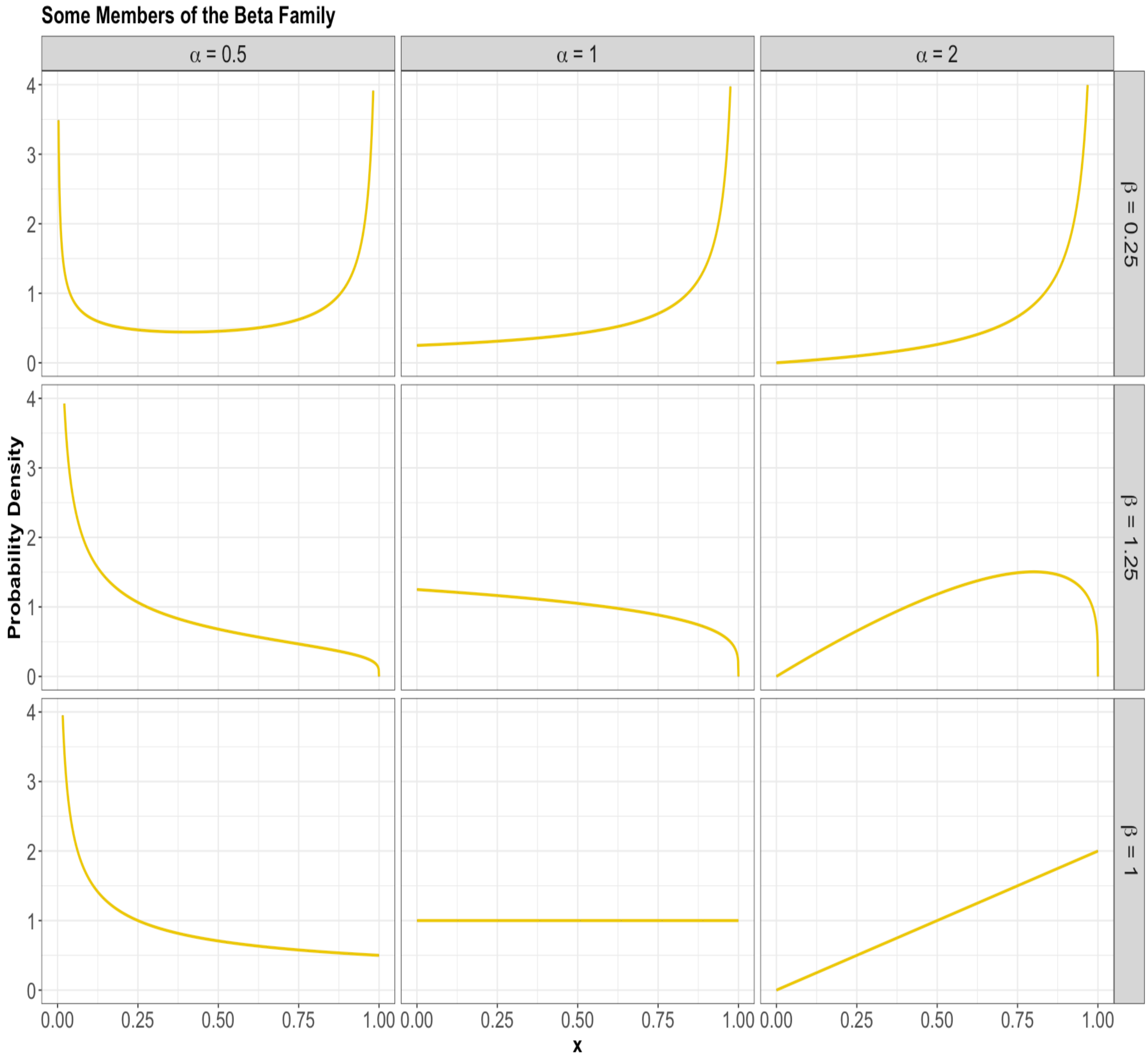
PDF

The density is parameterized as

$$f_X(x \mid \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha - 1}(1 - x)^{\beta - 1} \qquad \text{for} \quad 0 \leq x \leq 1,$$

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

Here are some examples of densities:



Mean

The mean of a Beta random variable is defined as:

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

Variance

The variance of a Beta random variable is defined as:

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

1.6. Weibull

Process

A generalization of the Exponential family, 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** $k > 0$ (where $k = 1$ results in the Exponential family).

The Weibull family is denoted as

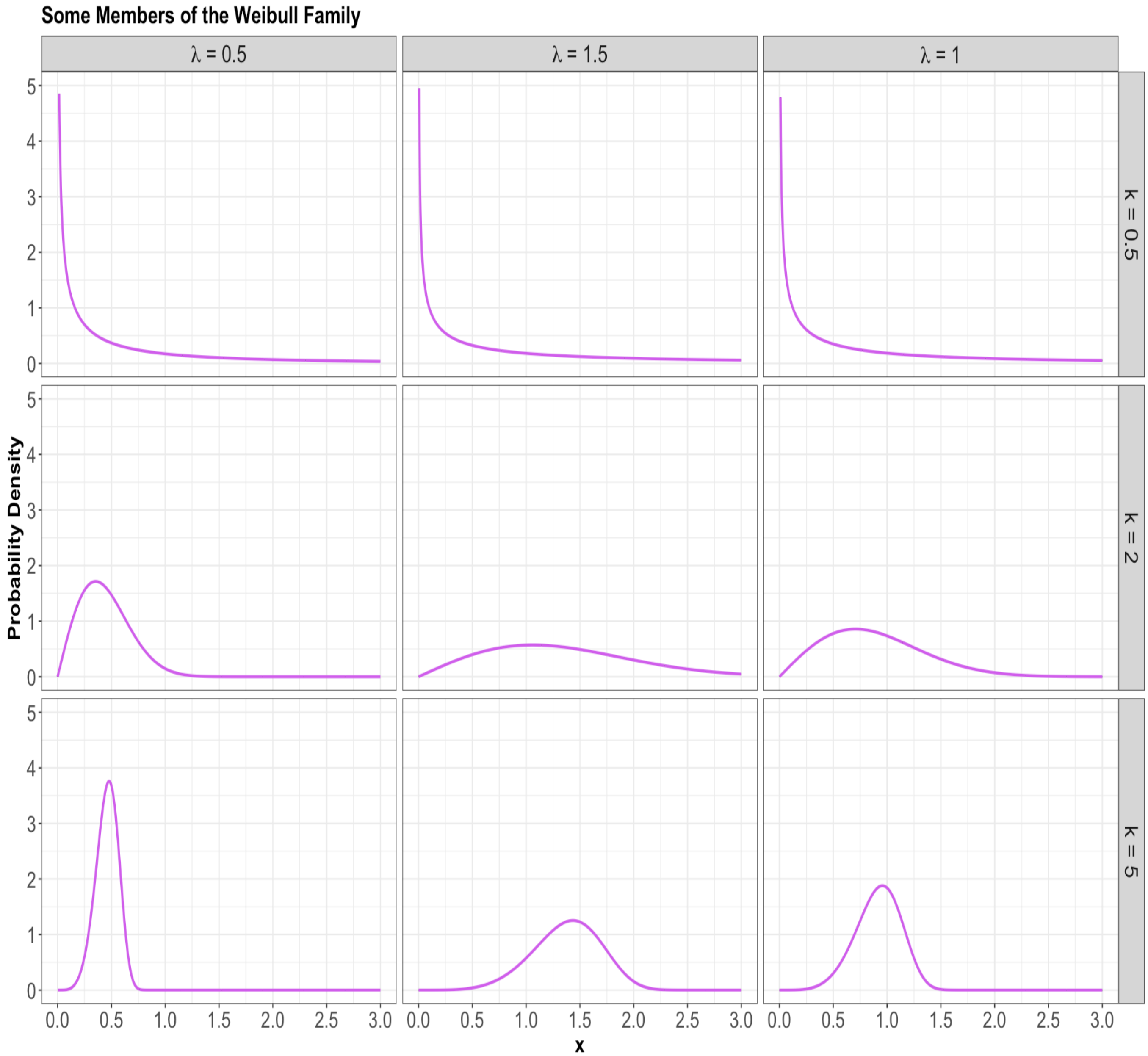
$$X \sim \text{Weibull}(\lambda, k).$$

PDF

The density is parameterized as

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

Here are some examples of densities:



Mean

The mean of a Weibull random variable is defined as:

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

Variance

The variance of a Weibull random variable is defined as:

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

1.7. Gamma

Process

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 $\theta > 0$.

The Gamma family can be denoted as

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

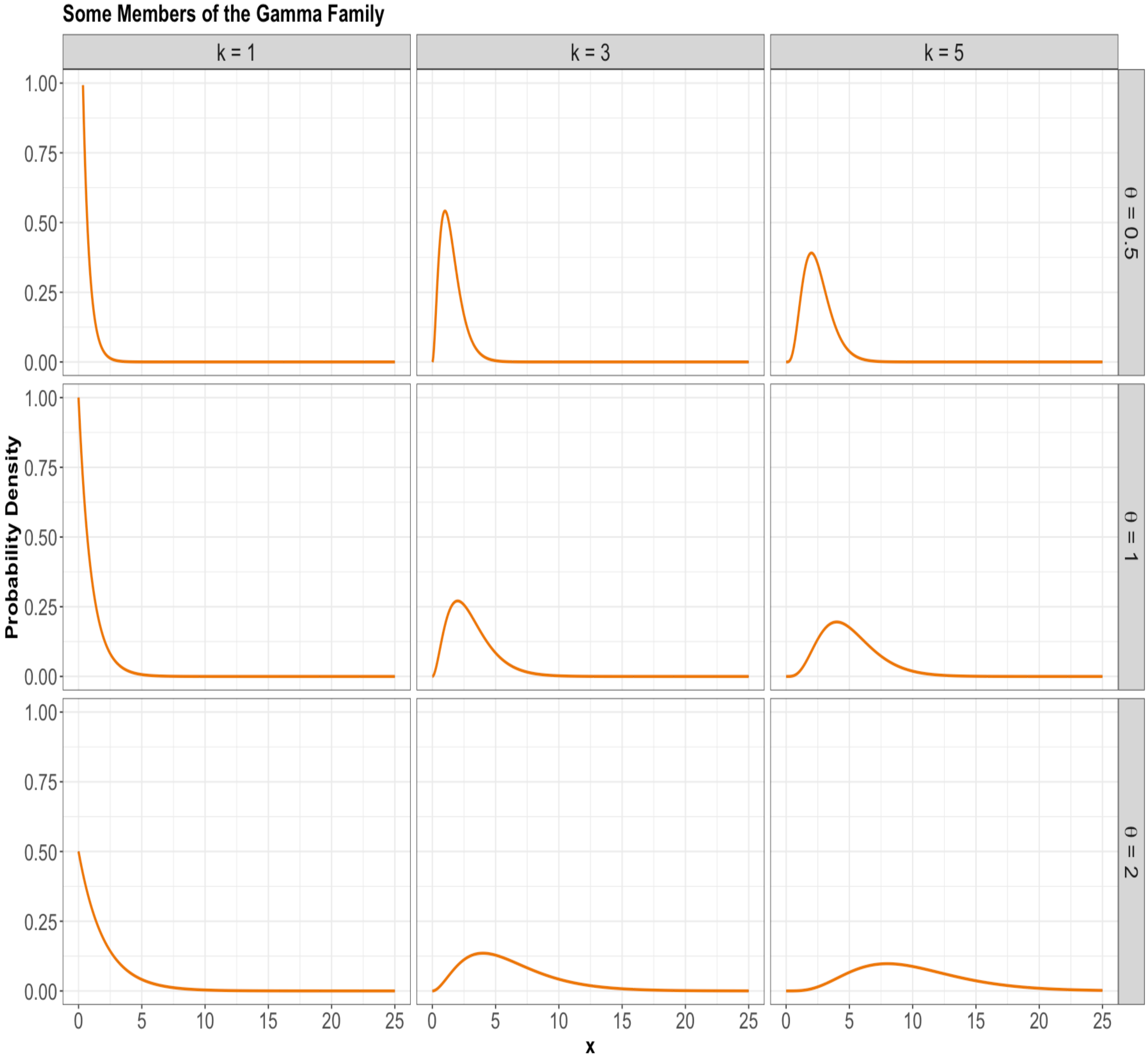
PDF

The density is parameterized as

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

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

Here are some densities:



Mean

The mean of a Gamma random variable is defined as:

$$\mathbb{E}(X) = k\theta.$$

Variance

The variance of a Gamma random variable is defined as:

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

1.8. Relevant `R` Functions

`R` has functions for many distribution families. We have seen a few already in the case of discrete families, but here is a more complete overview.

The functions are of the form `<x><dist>`, where `<dist>` is an abbreviation of a distribution family, and `<x>` is one of `d`, `p`, `q`, or `r`, depending on exactly what about the distribution you would like to calculate.

The possible prefixes for `<x>` indicate the following:

- `d`: density function – we call this $f_X(x)$.
- `p`: cumulative distribution function (CDF) – we call this $F_X(x)$.
- `q`: quantile function (inverse CDF).
- `r`: random number generator.

Here are some abbreviations for `<dist>`:

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

Let us check the following examples:

- For the Uniform family, we have the following `R` functions:
  - `dunif()`, `punif()`, `qunif()`, and `runif()`.
- For the Gaussian or Normal family, we have the following `R` functions:
  - `dnorm()`, `pnorm()`, `qnorm()`, and `rnorm()`.

Important

For any `R` function, you can get syntax help via the `R` Studio console. For instance, if we want help with the function `dunif()`, we type `?dunif` in the console. More ways to find help can be found here.

Now, let us proceed with some in-class questions via **iClicker**.

Exercise 28

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)`

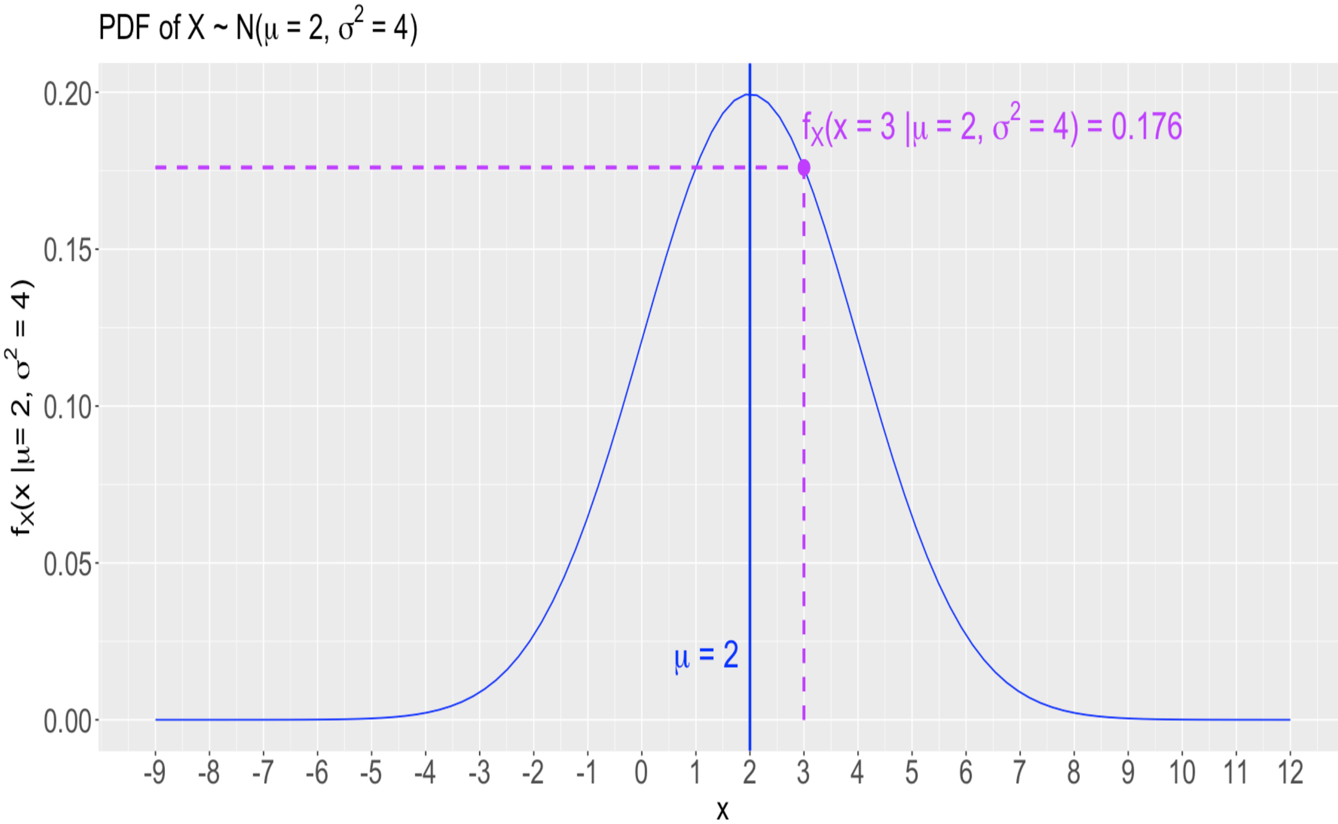
Solution to Exercise 28

We need the **density**, not the probability. Thus, we use `dnorm()` with parameters `mean = 2` (i.e., $\mu = 2$) and `sd = 2` (i.e., $\sigma^2 = 4$ **since the function takes the standard deviation** $\sigma = 2$ **as an argument**).

The `R` syntax can be found below, with its output rounded to three decimal places and the graphical representation of this density point in purple. Moreover, the mean $\mu = 2$ is represented as a vertical solid blue line.

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

0.176



Exercise 29

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

Select the correct option:

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)`

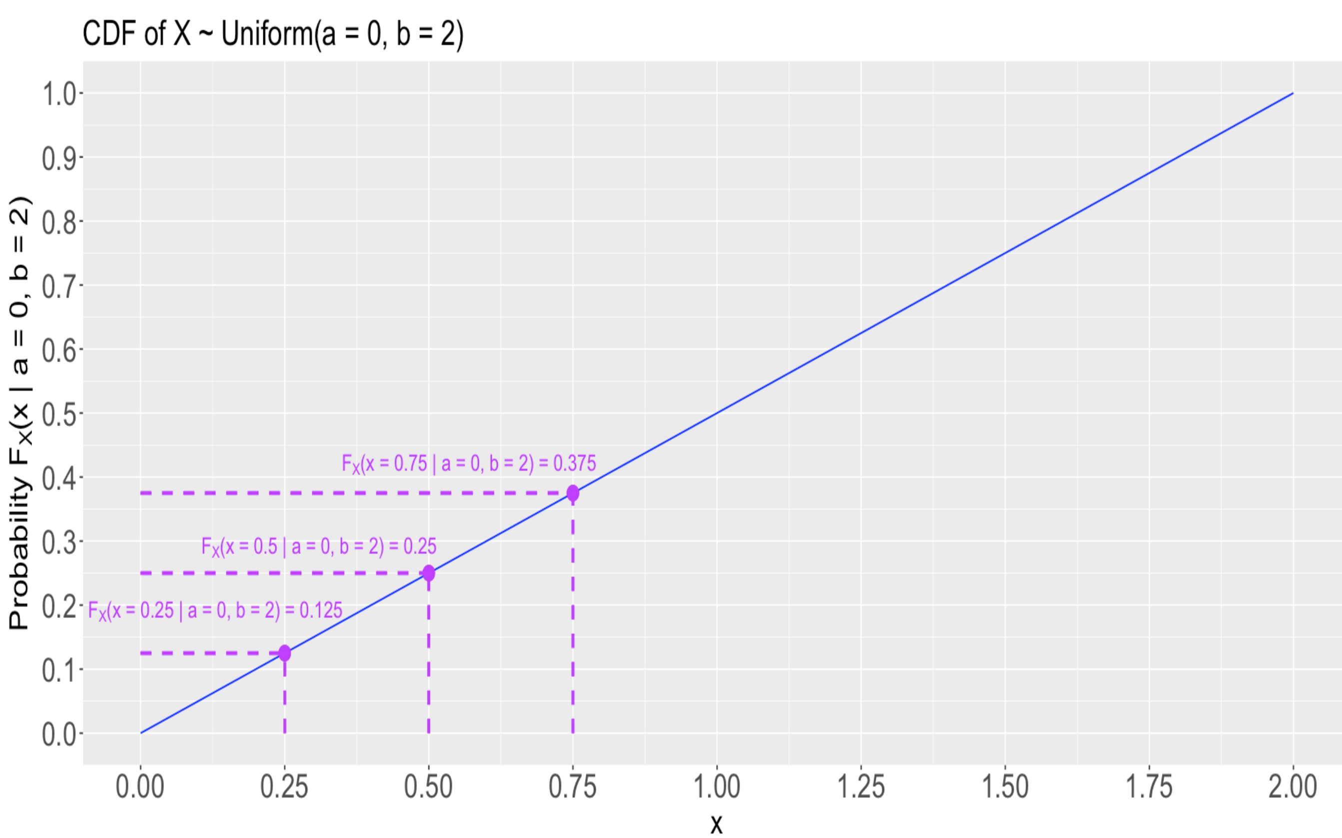
Solution to Exercise 29

We need the values of the CDF at different points on the x-axis, i.e. three different cumulative **probabilities**. Therefore, with $X \sim \text{Uniform}(a = 0, b = 2)$, we need to use `punif()` with arguments `min = 0` and `max = 2` for the quantiles in vector `q = c(0.25, 0.5, 0.75)`. Moreover, the argument `lower.tail = TRUE` indicates that we want cumulative probabilities **beginning on the left-hand side of the PDF**. Note we do not need to specify `lower.tail = TRUE` in practice since this is a default argument.

The `R` syntax can be found below with the graphical representation of the CDF $F_X(x \mid a = 0, b = 2)$ (**not the PDF!**). The three probabilities are also indicated in purple points.

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

0.125 · 0.25 · 0.375



Exercise 30

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

Select the correct option:

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)`

Solution to Exercise 30

We need the **quantile** x of the PDF at which we have 50-50 chance (i.e., 0.5 probability on its left-hand side and 0.5 probability on its right-hand side). Therefore, with $X \sim \text{Uniform}(a = 0, b = 2)$, we need to use `qunif()` with arguments `min = 0` and `max = 2` for the probability `p = 0.5`. Moreover, the argument `lower.tail = TRUE` indicates we want the cumulative probability **beginning on the left-hand side of the PDF**. Note we do not need to specify `lower.tail = TRUE` in practice since this is a default argument. Even `lower.tail = FALSE` (i.e., the cumulative probability **beginning on the right-hand side of the PDF**) yields the same quantile since we are talking about the median.

The `R` syntax can be found below with the graphical representation of the PDF $f_X(x \mid a = 0, b = 2)$. The median $M(X) = 1$ as a quantile is also indicated as a dashed vertical purple line with the shaded probabilities (i.e., areas under the PDF).

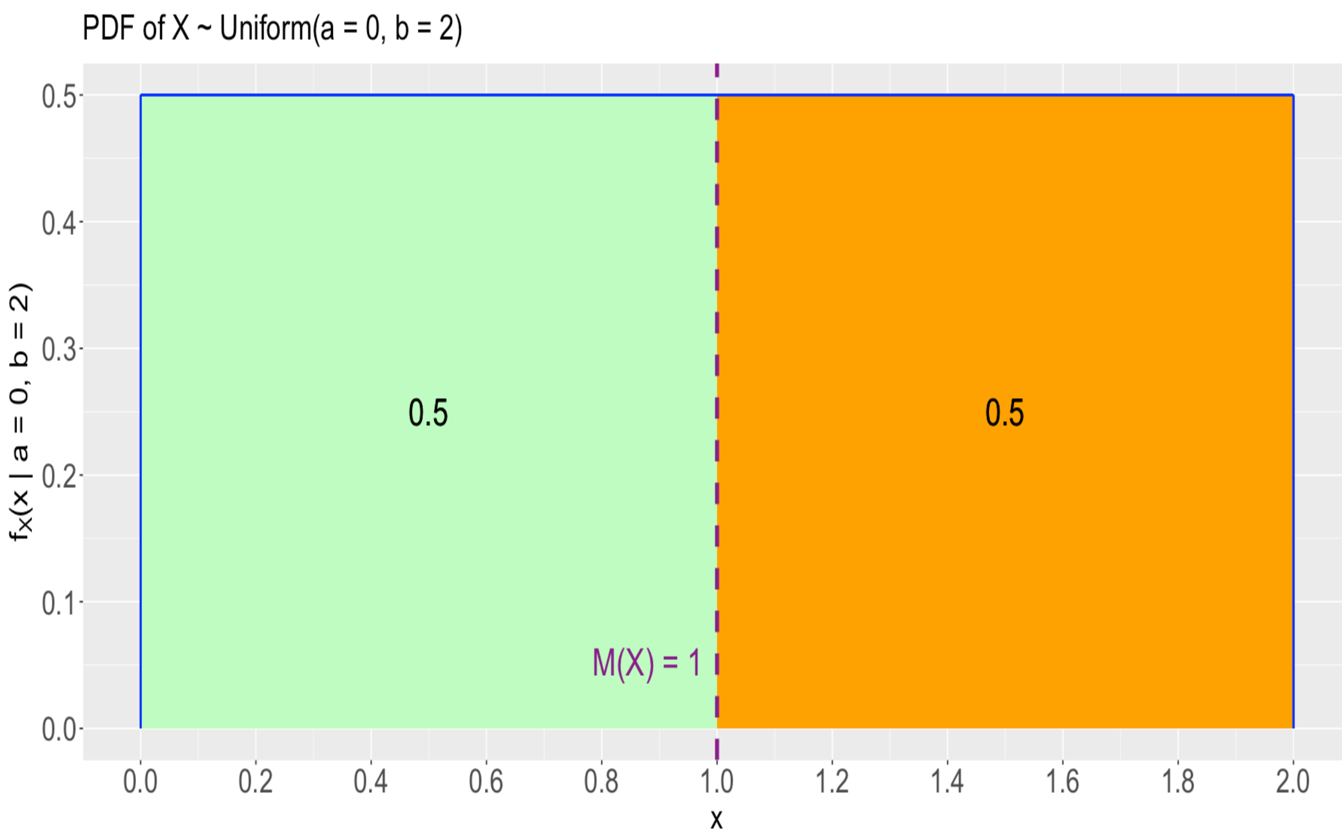
```
qunif(p = 0.5, min = 0, max = 2, lower.tail = TRUE)
```

1

Or equivalently, since this is the median:

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

1



Exercise 31

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

Select the correct option:

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)`

Solution to Exercise 31

Since we need to generate random numbers from a Normal distribution, `rnorm()` is the correct function and not `runif()` (which is meant for random numbers coming from a Uniform distribution). Thus, we need to indicate `n = 10` since we require 10 random numbers with `mean = 0` (i.e., $\mu = 0$) and `sd = 5` (i.e., $\sigma^2 = 25$ **since the function takes the standard deviation $\sigma = 5$ as an argument**).

The `R` syntax can be found below, with its output rounded to three decimal places: a vector containing ten random numbers. As a side note, every time we generate random numbers, it is necessary to set a seed so we can ensure reproducibility when running simulations (more details on this matter will be provided in `lecture8`).

```
set.seed(551)
round(rnorm(n = 10, mean = 0, sd = 5), 3)
```

4.205 · 1.188 · 3.049 · -7.458 · -2.28 · 3.93 · -5.429 · -6.474 · -1.001 · 6.683

2. Continuous Multivariate Distributions

In the discrete case, we already saw joint distributions, univariate conditional distributions, multivariate conditional distributions, marginal distributions, etc. All these concepts are carried over to continuous distributions. Let us start with two continuous random variables (i.e., a **bivariate** case).

Important

Let us make a **note on depictions of distributions**. There is such thing as a multivariate CDF. It comes in handy in Copula Theory (a more advanced field in Statistics). But otherwise, it is not as useful as a **multivariate density**, so we will not cover it in this course. Moreover, there is no such thing as a multivariate quantile function.

2.1. Multivariate Probability Density Functions

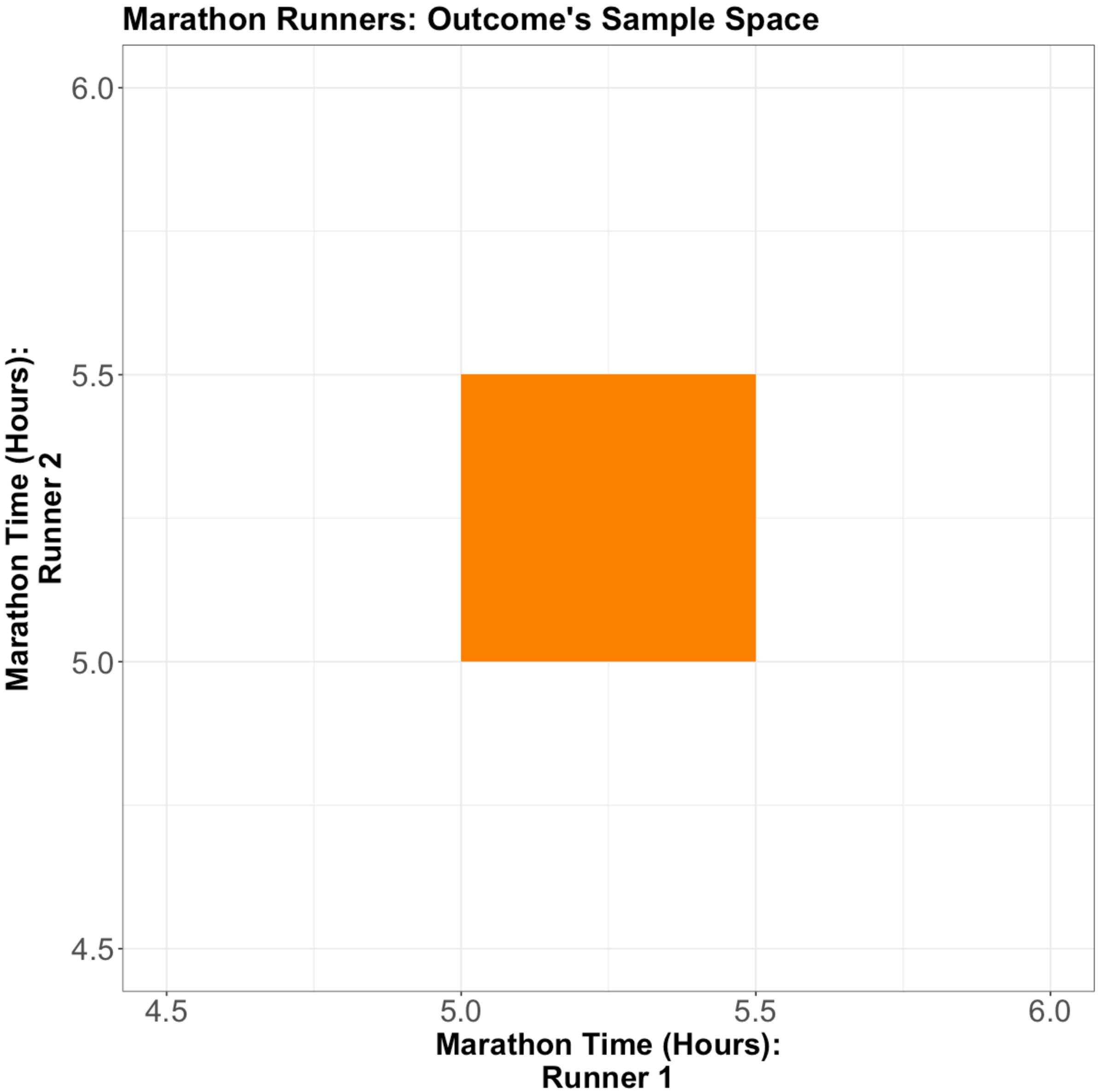
Recall the joint **probability mass function** (PMF) from [Lecture 3: Joint Probability](#) between Gang demand and length of stay (LOS):

A matrix: 5 × 4 of type dbl

	Gang = 1	Gang = 2	Gang = 3	Gang = 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

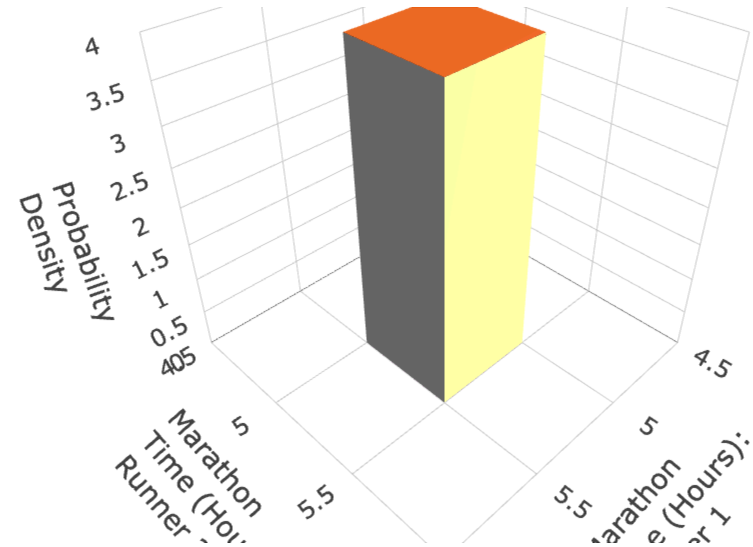
Each entry in the table corresponds to the probability of that unique row (LOS value) and column (Gang value). These probabilities add up to 1.

For the **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. For 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. In that case, the possible values are indicated by the **orange square** in the following plot:



Each point in the square is like an entry in the joint probability mass function (PMF) table in the discrete case, except now, instead of holding a probability, it holds a **density**. Then, the **density function** is a **surface** overtop of this square (or, generally, the outcome's sample space). It is a function that takes two variables (marathon time for Runner 1 and Runner 2) and calculates a single density value from those two points. This function is called a **bivariate density function**.

We can extend the above plot to a 3D setting as below. Note that the orange distribution surface corresponds to the projection of the joint distribution between the two runners on the z-axis.



2.2. Calculating Probabilities

Recall the univariate continuous case; we calculated probabilities as the **area under the density curve**. Now, let us check the specifics of the bivariate density function.

Definition of the Bivariate Probability Density Function

Let X and Y be two random variables. For these two random variables X and Y, their joint PDF evaluated at the points x and y is usually denoted

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

Therefore, we have a **density surface** and can calculate probabilities as the **volume under the density surface**. This means that the total volume under the density function must equal 1. Formally, this may be written as

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

Note what this implies about the units of $f_{X,Y}(x, y)$. For example, if x is measured in metres and y is measure in seconds, then the units of $f_{X,Y}(x, y)$ are $\mathrm{m}^{-1}\mathrm{s}^{-1}$.

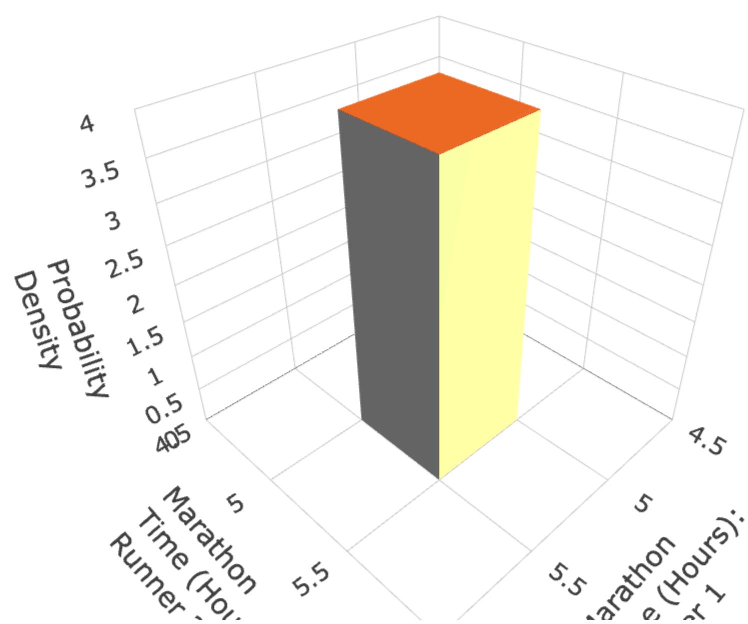
Now, let us answer the following questions:

Exercise 32

If the density is equal/flat across 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?

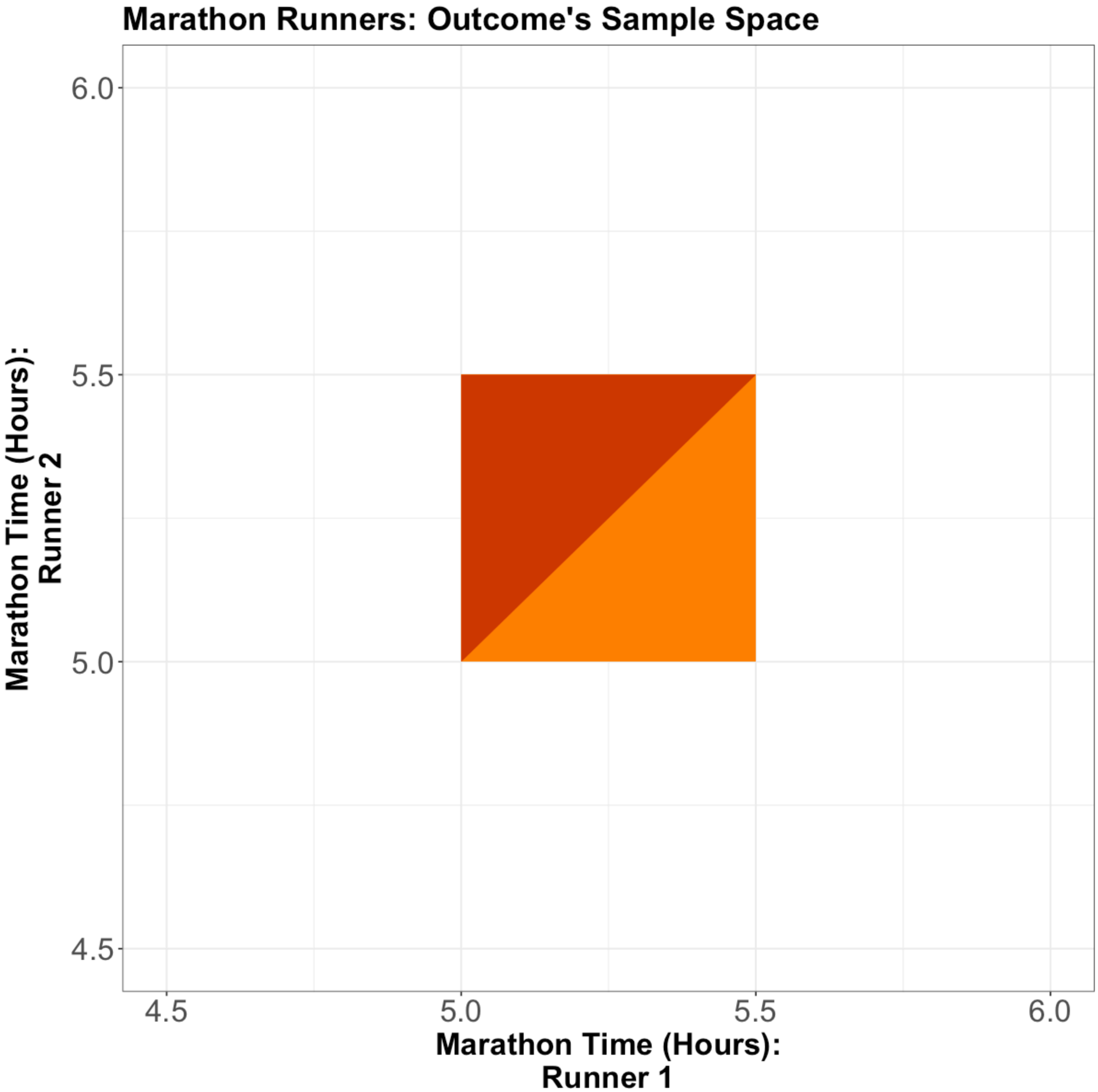
Solution to Exercise 32

By rechecking the below 3D plot, the height of the surface on the z-axis is equal to 4, which ensures that **the prism's volume is equal to 1 (i.e., the probability of the sample $P(S) = 1$)**. That said, if we are covering the whole sample space with the volume of this prism, then the probability of any point outside of this prism will be zero.



Exercise 33

Let X be the marathon time of Runner 1 in hours, and Y be the marathon time of Runner 2 in hours. What is the probability that Runner 1 will finish the marathon before Runner 2, i.e., $P(X < Y)$? The below plot might be helpful to visualize this.


Marathon Runners: Outcome's Sample Space
6.0
5.5
5.0
4.5
Marathon Time (Hours):
Runner 2
4.5
5.0
5.5
6.0
Marathon Time (Hours):
Runner 1

Solution to Exercise 33

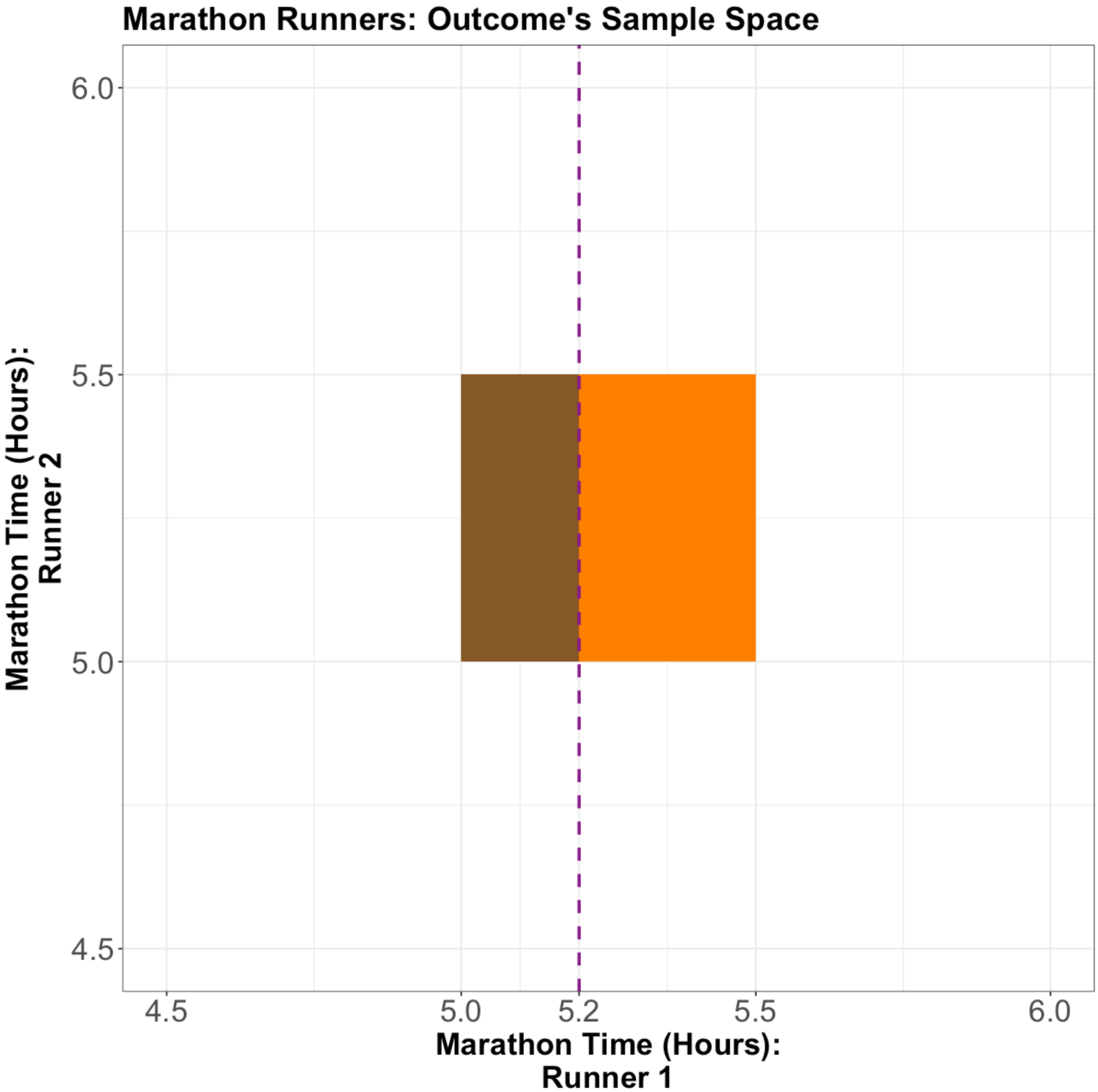
First, identify the **surface region** in the sample's outcome space that corresponds to this event to calculate this. This is plotted above as the top's darker orange triangle. Across this darker orange triangle, we can see that Runner 2's time Y (on the y- axis) is always larger than Runner 1's time X (on the x- axis).

Then, calculate the volume of the space overtop of this region and below the density surface. We already know our whole prism's volume is equal to 1, and this darker orange part would correspond to half of the volume **resulting in**

$$P(X < Y) = \frac{1}{2}.$$

Exercise 34

Now, what is the probability that Runner 1 finishes in 5.2 hours or less, i.e., $P(X \leq 5.2)$? The below plot might be helpful to visualize this.


Marathon Runners: Outcome's Sample Space
Marathon Time (Hours): Runner 2
Marathon Time (Hours): Runner 1
6.0
5.5
5.0
4.5
4.5
5.0
5.2
5.5
6.0

Solution to Exercise 34

Again, we have to identify the **surface region** in the sample's outcome space that corresponds to this event to calculate this. The question only asks about Runner 1; thus, our upper time limit to compute this probability is at $x = 5.2$ hours (denoted by the vertical dashed purple line).

Note we do not take into account any specific time limits for Runner 2's time Y (**as long as we cover their whole corresponding time range**). Therefore, we are computing a **marginal probability** for X corresponding to Runner 1. That said, we can compute the following probability as a "nested" volume:

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

3. Conditional Distributions (Revisited)

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

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{34}$$

Nonetheless, this is only true if $P(B) \neq 0$, and it is not useful if $P(A) = 0$ – two situations we face in the continuous world

3.1. When $P(A) = 0$

To describe this situation, let us use a univariate continuous example: the monthly expenses.

Suppose the month is halfway over, and you only have $\$2500$ worth of expenses so far! Given this information, what is the distribution of this month's total expenditures now?

If we use the Law of Conditional Probability (34), we would get a formula that is not useful. Let

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

Moreover, assume

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

Using Equation (34), the conditional probability **would** be given by

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

This is not good because the outcome x would have a probability of 0 in a continuous case, which makes no sense!

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

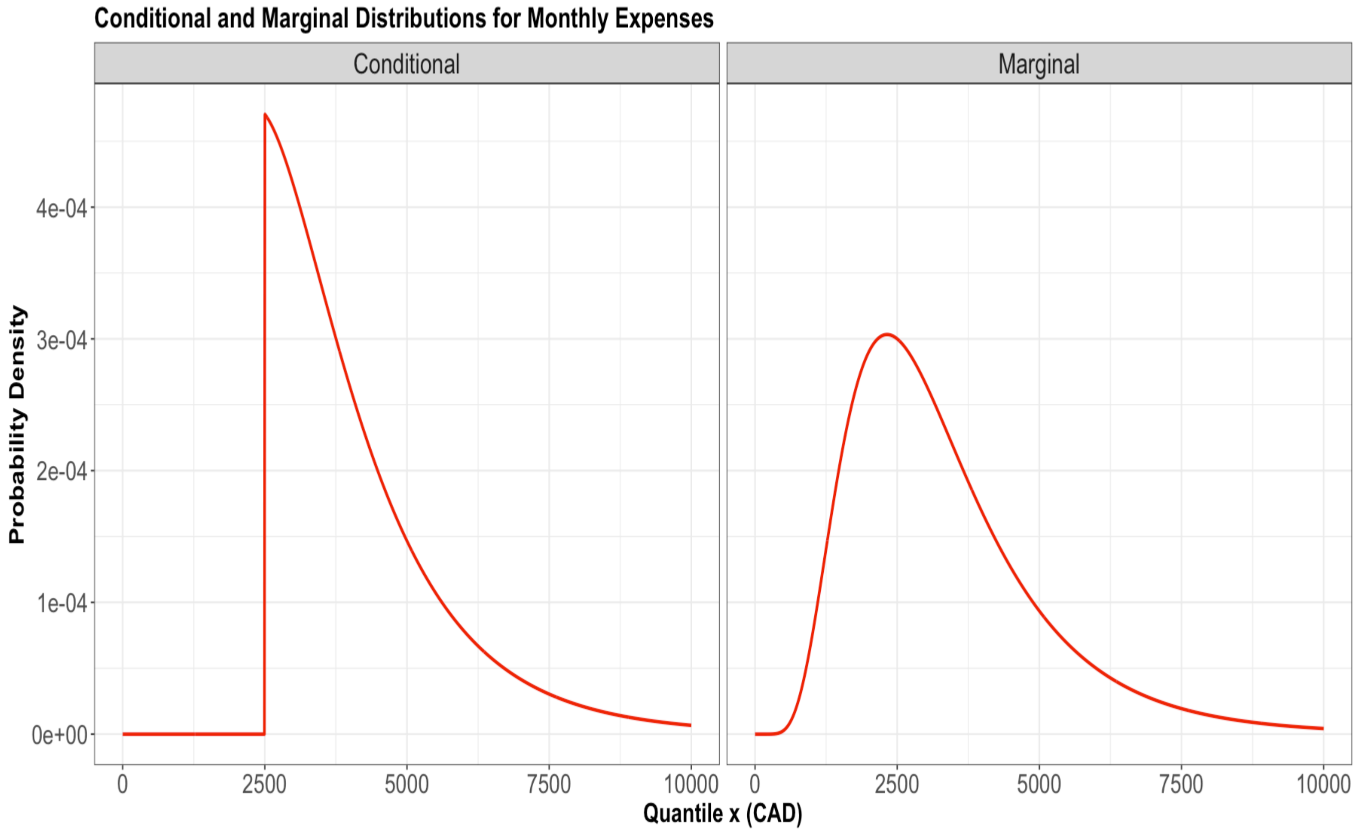
$$f_{X \mid X \geq 2500}(x) = \frac{f_X(x)}{P(X \geq 2500)} \qquad \text{for} \quad x \geq 2500,$$

and

$$f_{X \mid X \geq 2500}(x) = 0 \qquad \text{for} \quad x < 2500.$$

The formula of the resulting conditional PDF is just the original PDF but confined to $x \geq 2500$, and re-normalized to have area 1.

Below, we plot the marginal distribution $f_X(x)$ and the conditional distribution $f_{X \mid X \geq 2500}(x)$. Notice the conditional distribution is just a segment of the marginal, and then re-normalized to have an area under the curve equal to 1.



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, we already pointed out that

$$P(X = 5.2) = 0.$$

Therefore, we need to find a proper workaround. In this case, the stopwatch used to calculate the run time has rounded the true run time to 5.2 hours, even though, in reality, it would have been something like $5.2133843789373\ldots$ hours.

As seen earlier, plugging in the formula for conditional probabilities will not work. But, as in the case of $P(A) = 0$, we can generally replace probabilities with densities.

We end up with

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

This formula is true in general

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