2. Probability spaces

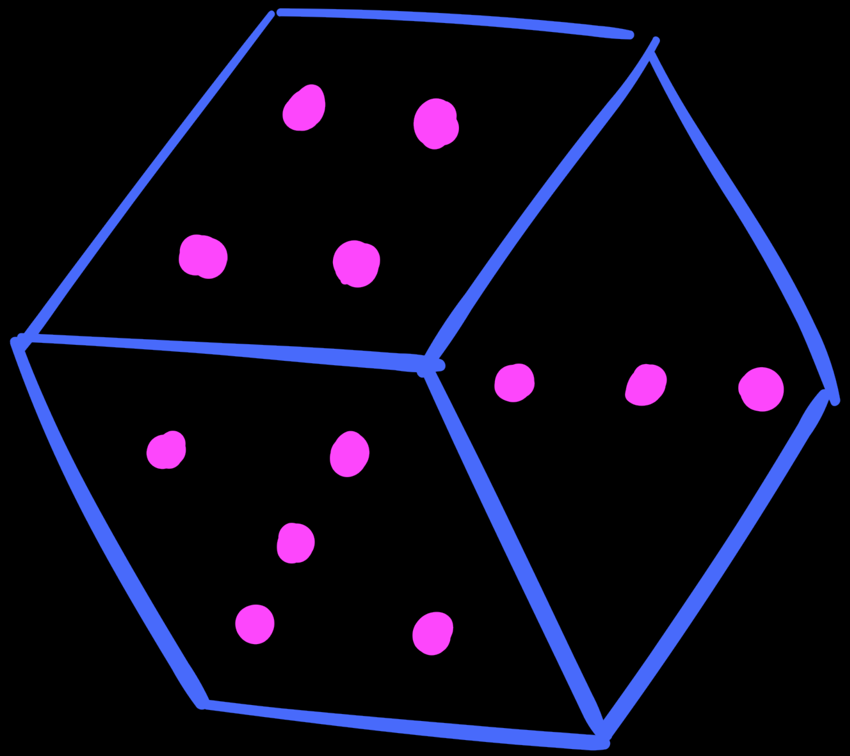
2.1. What is probability?

Definition 2.1

The interpretation that conceptualizes probabilities as measurements of *degrees of belief* is called the *subjective* (or *personal*) *interpretation* of probability.

an example you'll get sick of:

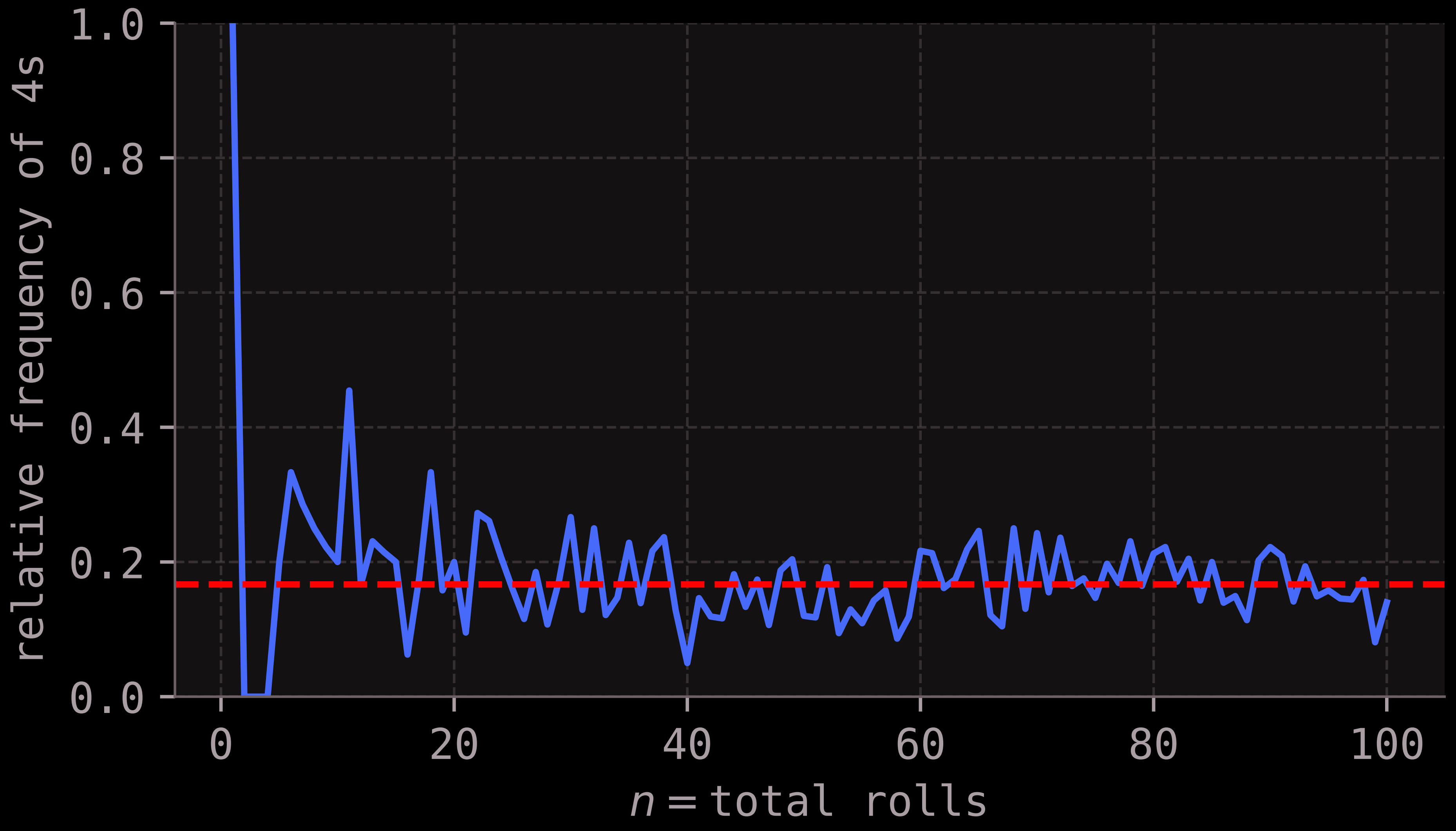
rolling a die

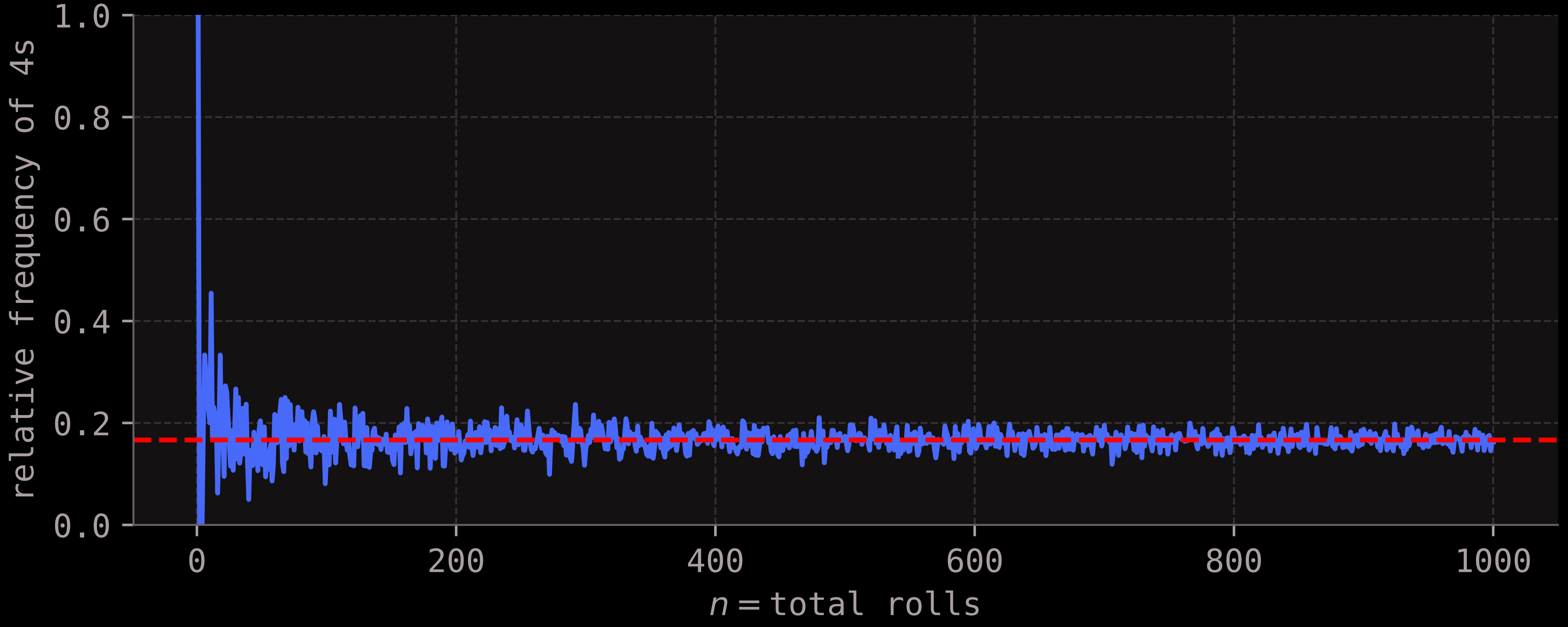


> “Well, since the die is fair and symmetric, there’s an equal chance of rolling any number. And since there are six possible numbers that we could roll, the probability of rolling any one particular number is one in six.”

Definition 2.2

The interpretation of probability whose characteristc qualities are appeals to “symmetry” and decompositions of events into “equally likely outcomes” is called the *classical interpretation* of probability.


1.0
0.8
0.6
0.4
0.2
0.0
relative frequency of 4s
0
20
40
60
80
100
n = total rolls



relative frequency of 4s
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
$n =$ total rolls

Definition 2.3

The interpretation that conceptualizes probabilities as long-run relative frequencies is called the *frequentist interpretation* of probability.

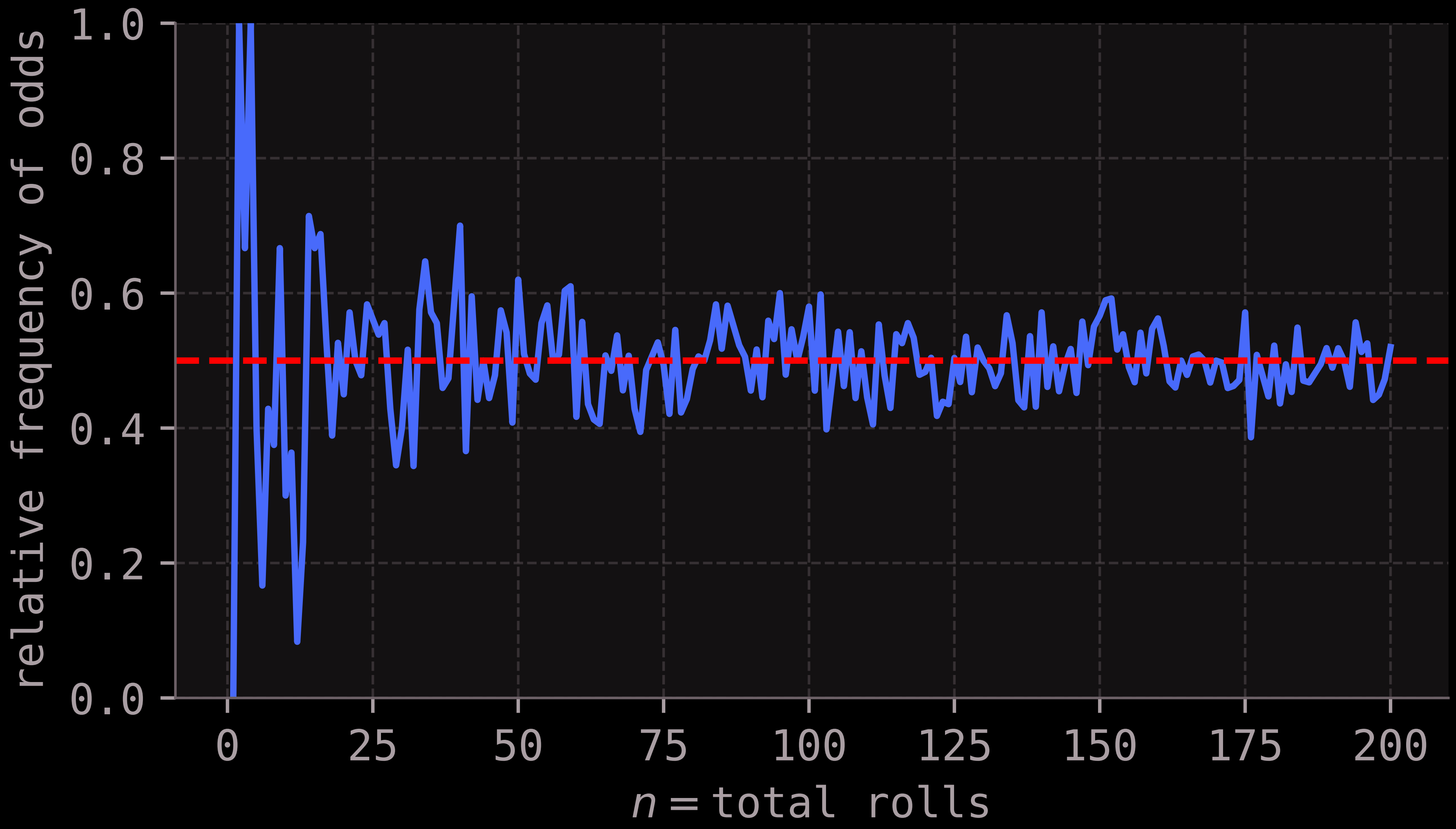
2.2. A first look at the axiomatic framework

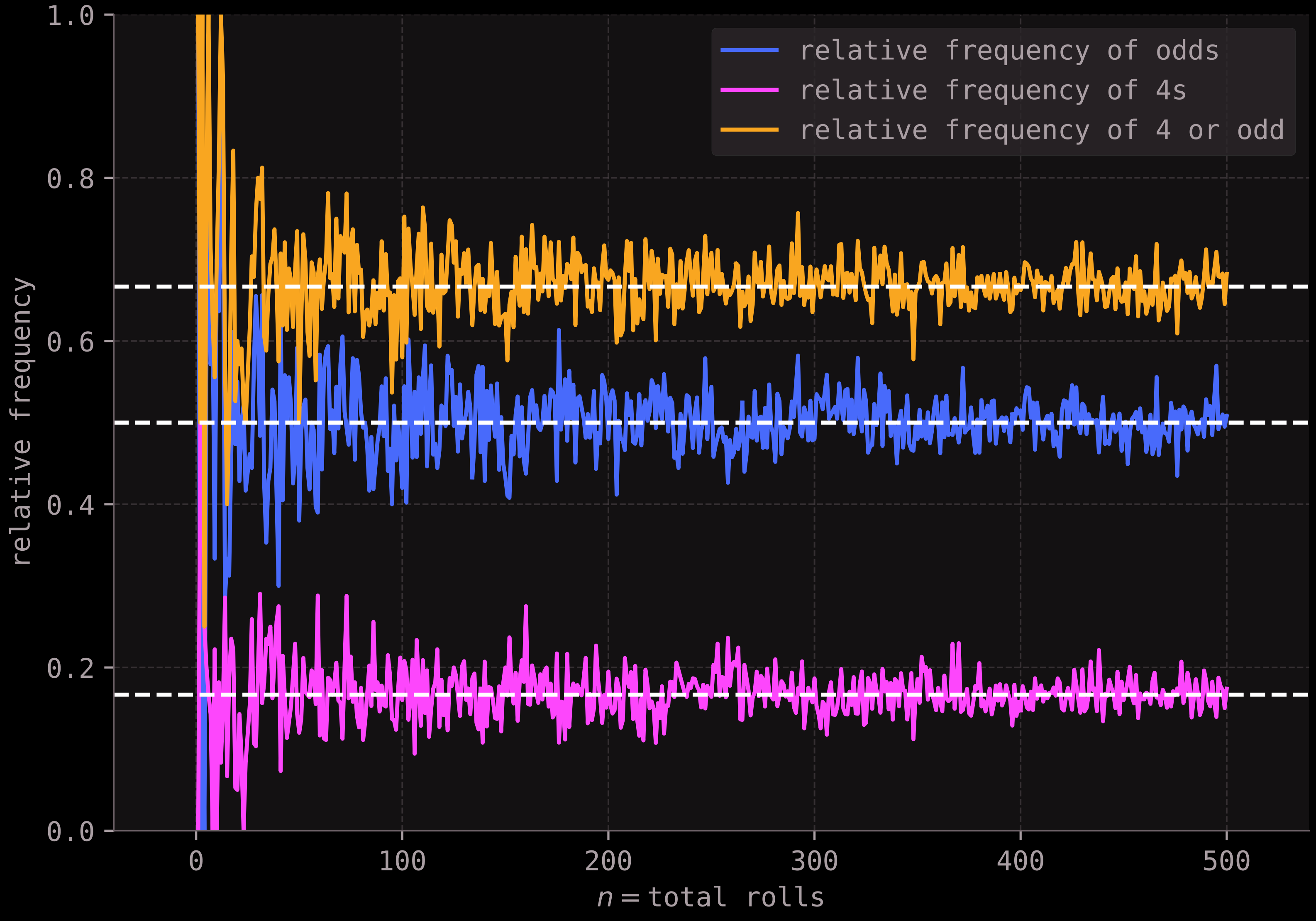
Axiom 2.1 (Probability)

Probabilities are represented by real numbers between 0 and 1, inclusive.

Axiom 2.2 (Probability)

The probability that *some* outcome occurs is 1.


relative frequency of odds
1.0
0.8
0.6
0.4
0.2
0.0
0
25
50
75
100
125
150
175
200
$n =$ total rolls


relative frequency of odds
relative frequency of 4s
relative frequency of 4 or odd
1.0
0.8
0.6
0.4
0.2
0.0
relative frequency
0
100
200
300
400
500
n = total rolls

Axiom 2.3 (Probability)

The probability of one or the other of two *disjoint* events occurring is the sum of the individual probabilities.

2.4. Probability spaces

Q: What *is* probability?

Q: What *is* probability?

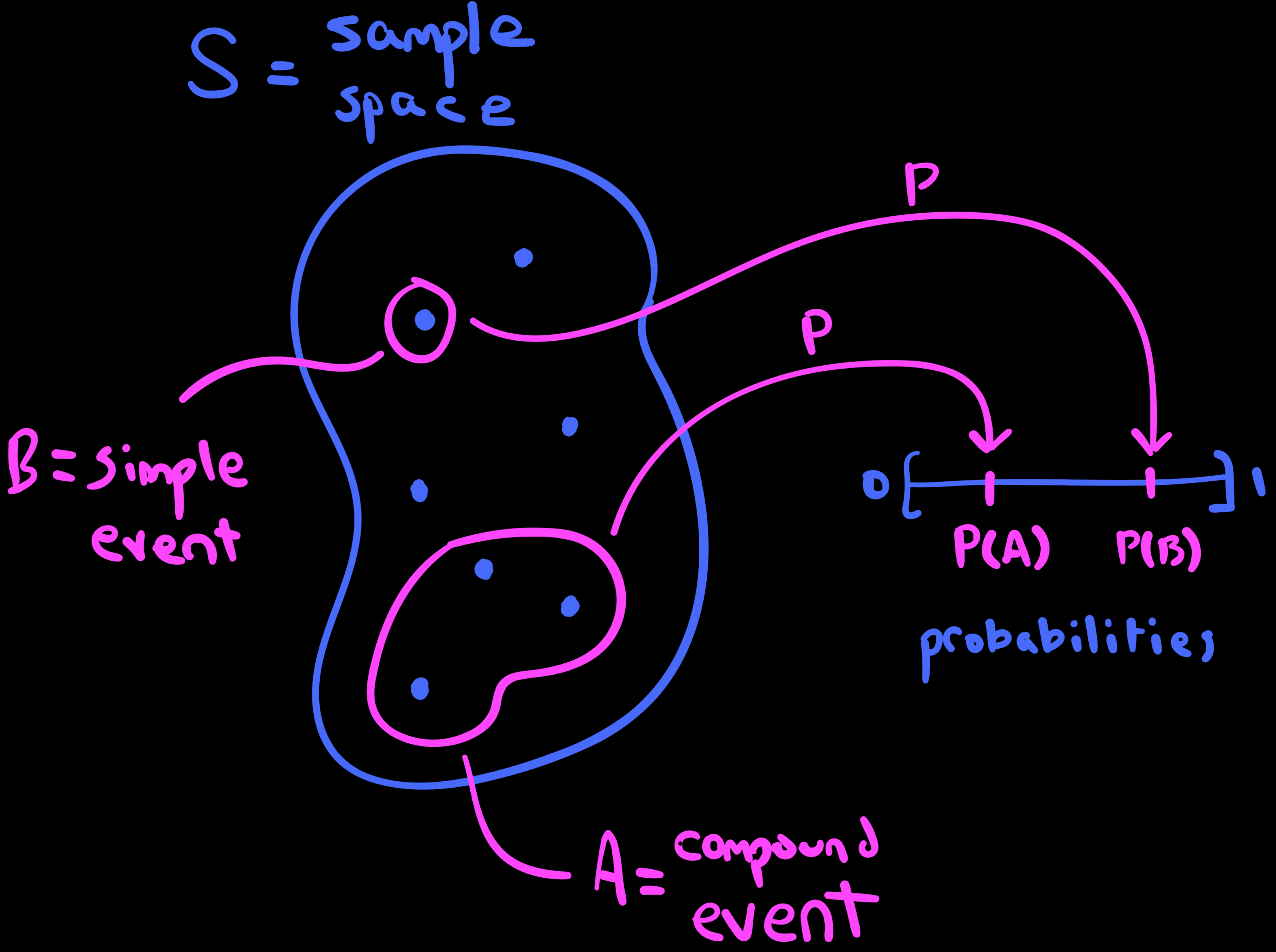
A: I don't know. But I **do** know that probabilities, no matter what they are, should follow a simple set of three rules, or axioms:

1. Probabilities are represented by real numbers between 0 and 1, inclusive.
2. The probability that *some* outcome occurs is 1.
3. The probability of one or the other of two *disjoint* events occurring is the sum of the individual probabilities.

Definition 2.6

A *probability space* consists of three things:

1. A set S called the *sample space*.
    - A sample space S often consists of all possible outcomes of a process or experiment, or it is the population under study (as defined back in Definition 1.1).
    - The elements of S are called *sample points* or *outcomes*.
2. A collection of subsets of S, called *events*.
    - So, an event in a sample space is nothing but a subset of the sample space.
    - An event containing just one sample point is called a *simple event*. All other events are called *compound events*.
3. A *probability measure* P.
    - Briefly, a probability measure is a function that assigns probabilities to events. (The precise definition is given in Definition 2.7 below.)



S = sample space
P
P
B = simple event
0
1
P(A)
P(B)
probabilities
A = compound event

Problem Prompt

Do problem 3 on the worksheet.

Problem Prompt

Do problems 4-6 on the worksheet.

Theorem 2.1 (Properties of Events)

Let S be a sample space.

1. If A is an event, then so too is its complement $S \setminus A$.
2. The entire sample space itself is always an event, and so is the empty set $\emptyset$.
3. If A and B are events, then so too is the union $A \cup B$ and intersection $A \cap B$.
4. In fact, if $A_1, A_2, A_3, \ldots$ is an infinite sequence of events, then the infinite union

   $$A_1 \cup A_2 \cup A_3 \cup \cdots$$

   and the infinite intersection

   $$A_1 \cap A_2 \cap A_3 \cap \cdots$$

   are also events.

Definition 2.7

Let S be a sample space. A *probability measure* P (also called a *probability distribution*) is a function that to each event A in S assigns a number $P(A)$, called the *probability of* A, subject to the following axioms:

1. $P(A) \geq 0$ for all events A.
2. $P(S) = 1$.
3. If $A_1, A_2, A_3, \ldots$ is a sequence of pairwise disjoint events in S, then

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = \sum_{i=1}^{\infty} P(A_i).$$

Theorem 2.2 (Properties of Probability Measures)

Let P be a probability measure on a sample space S.

1. If A is an event, then $P(A) \leq 1$.
2. If A and B are events and $A \subset B$, then $P(A) \leq P(B)$.
3. If A is an event, then $P(A^c) = 1 - P(A)$.
4. If A and B are disjoint events, then $P(A \cup B) = P(A) + P(B)$.
5. We have $P(\emptyset) = 0$.

$$P(\underset{\substack{\uparrow\\ \text{input}}}{\text{event}}) = \underset{\substack{\uparrow\\ \text{output}}}{\text{probability}}$$

2.6. Discrete and uniform probability measures

Definition 2.8

Let S be any set and $p : S \to \mathbb{R}$ a function.

1. The *support* of p is the set of all points $s \in S$ where $p(s)$ is nonzero, i.e., it is the set

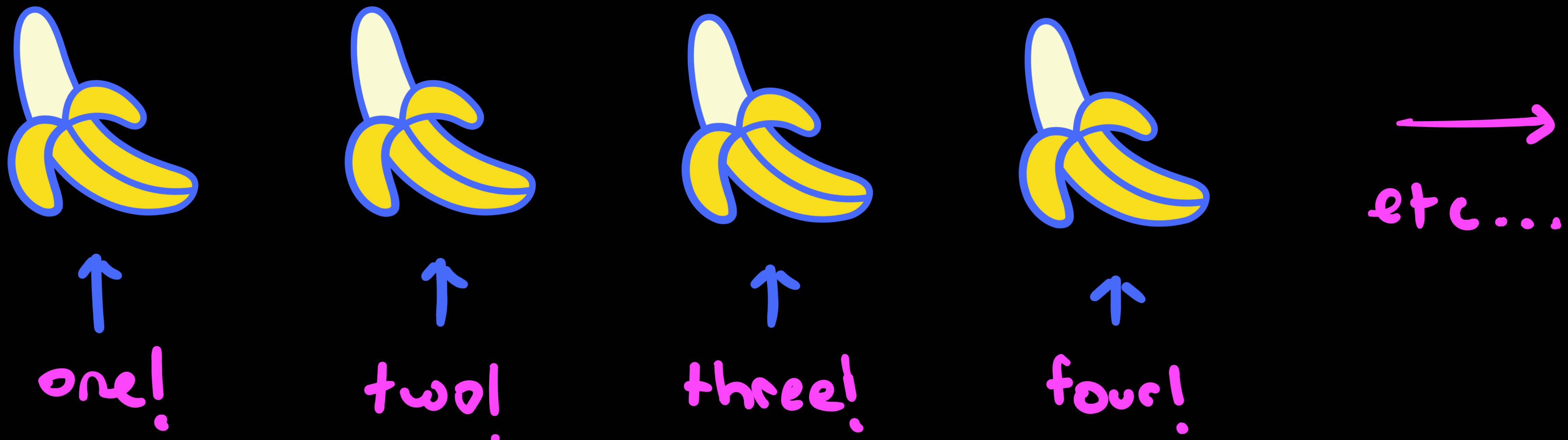
$$\{s \in S : p(s) \neq 0\}. \tag{2.1}$$

2. The function p is said to have *discrete* support if its support (2.1) is either finite or countably infinite.

Definition 2.9 (sort of)

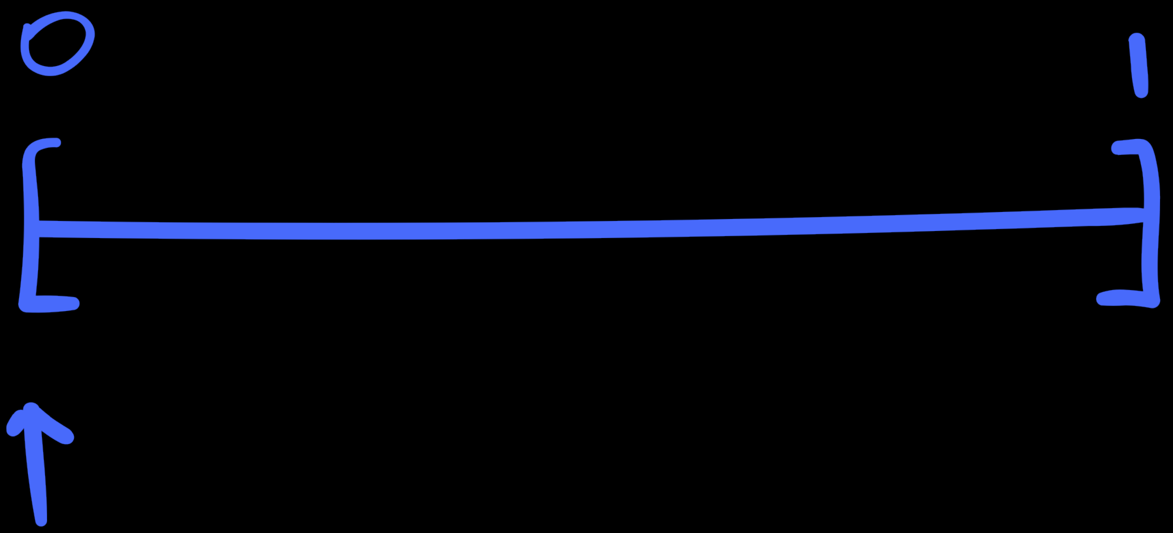
A set A is *countably infinite* if, given an infinite amount of time, I could count the elements of A one at a time, counting one element per second.

Countably infinite



uncountably infinite

all numbers between 0 and 1



one! but where is "two"?!

Definition 2.10

Let P be a probability measure on a sample space S. We shall say P is *discrete* if every subset $A \subset S$ is an event and there is a function $p : S \to \mathbb{R}$ with discrete support such that

$$P(A) = \sum_{s \in A} p(s), \tag{2.2}$$

for all events A. In this case, the function p is called the *probability mass function* (PMF) of the probability measure P (or sometimes just the *probability function*), and S is called a *discrete probability space* (when equipped with P).

Problem Prompt

Do problems 7 and 8 on the worksheet.

Theorem 2.3 (Properties of Probability Mass Functions)

Let $p(s)$ be the probability mass function of a discrete probability measure P. Then:

1. $p(s) \geq 0$ for all $s \in S$, and
2. $\sum_{s \in S} p(s) = 1$.

Theorem 2.4 (Discrete Probability Construction Lemma)

Let S be a set and $p : S \to \mathbb{R}$ a function with discrete support. If

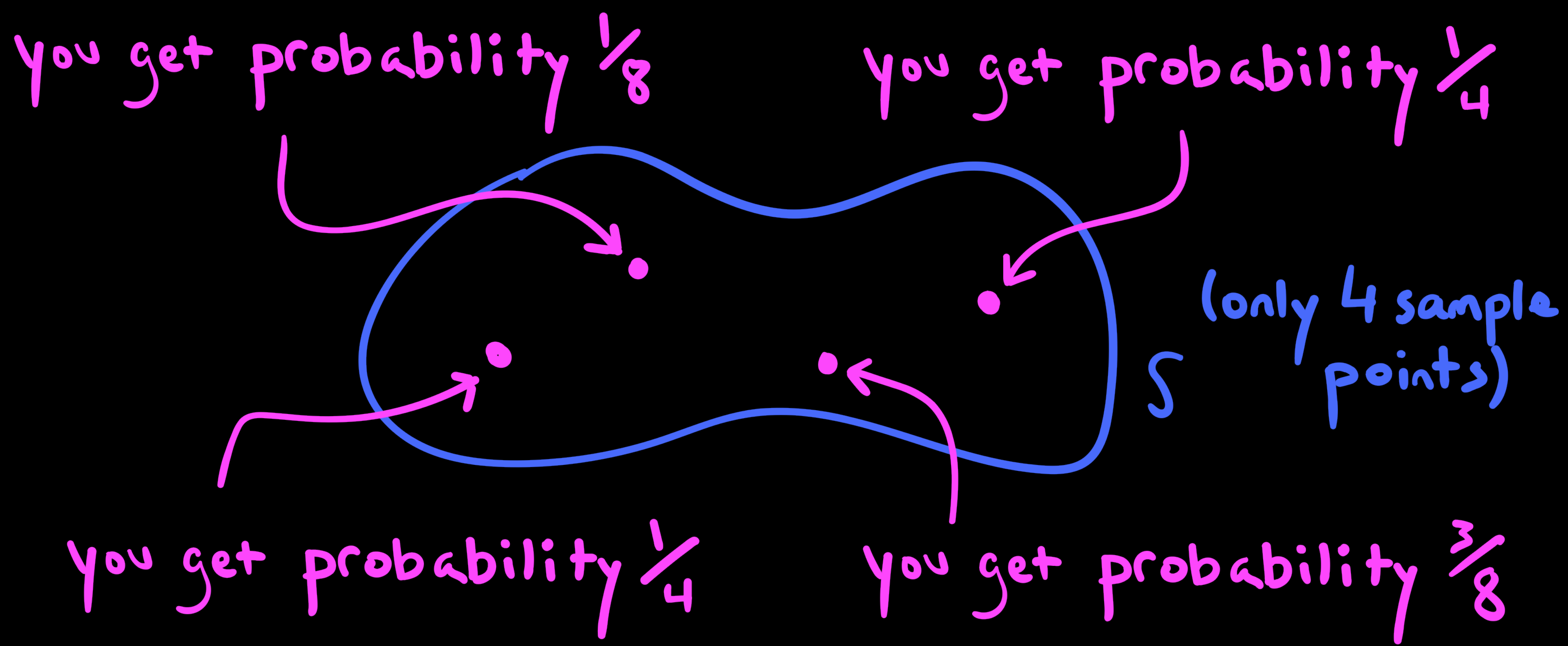
1. $p(s) \geq 0$ for all $s \in S$, and
2. $\sum_{s \in S} p(s) = 1$,

then there is a unique discrete probability measure P on S such that

$$P(A) = \sum_{s \in A} p(s)$$

for all $A \subset S$.

Discrete probability space



$$\frac{1}{8} + \frac{1}{4} + \frac{3}{8} + \frac{1}{4} = 1 \checkmark$$

Definition 2.11

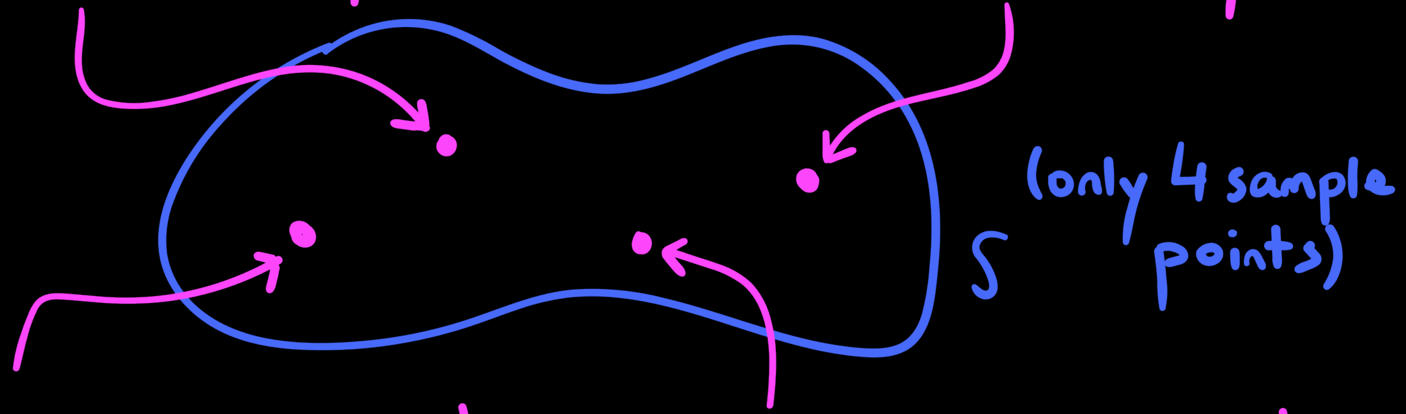
Let P be a discrete probability measure on a sample space S with probability mass function $p(s)$. Then P is called a *uniform probability measure* if the support of $p(s)$ has finite cardinality $n > 0$, and if

$$p(s) = \frac{1}{n}$$

for each s in the support of p.

uniform probability space

you get probability $\frac{1}{4}$ you get probability $\frac{1}{4}$



you get probability $\frac{1}{4}$ you get probability $\frac{1}{4}$

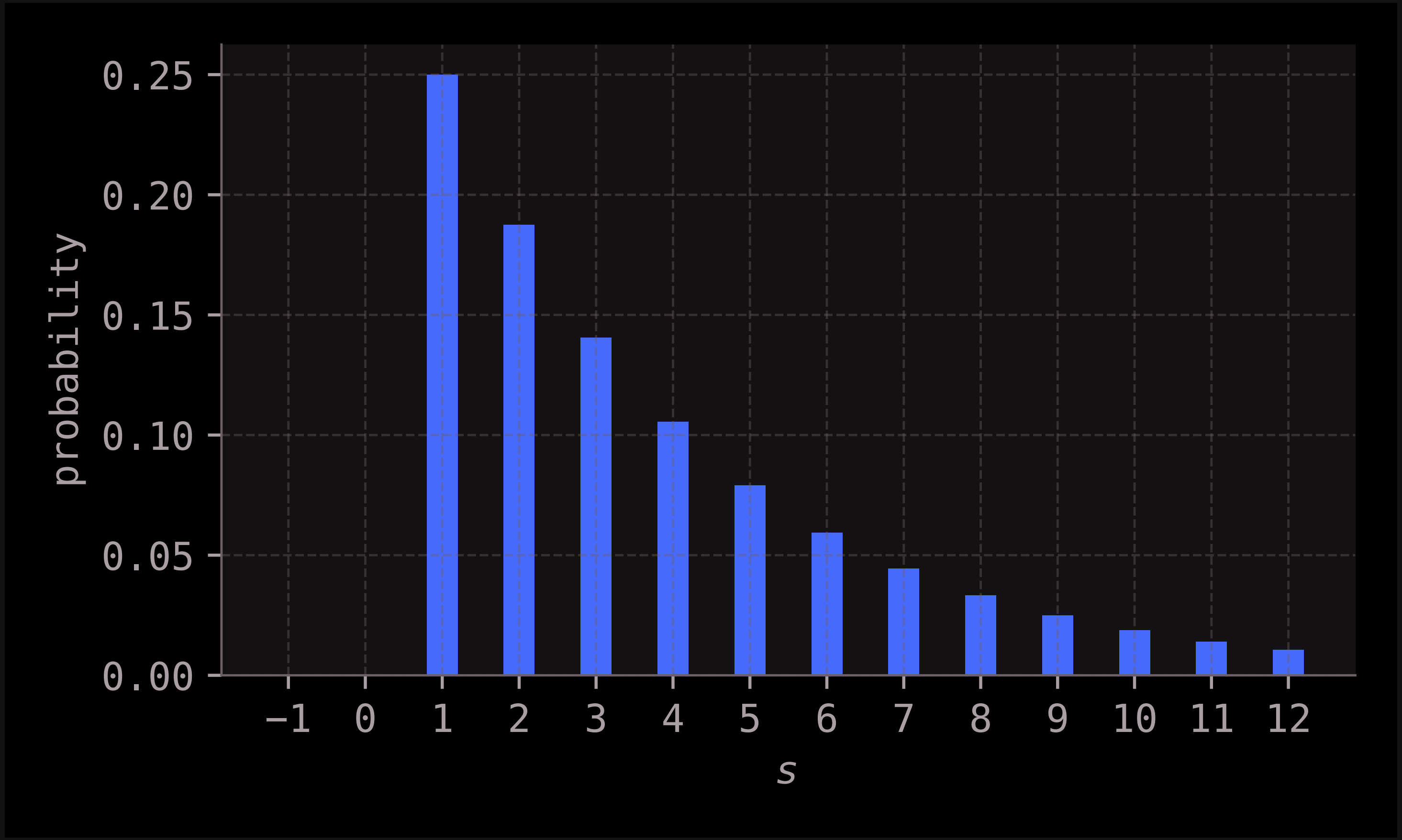
$$\frac{1}{4}+\frac{1}{4}+\frac{1}{4}+\frac{1}{4}=1 \checkmark$$

all probabilities <u>equal</u>

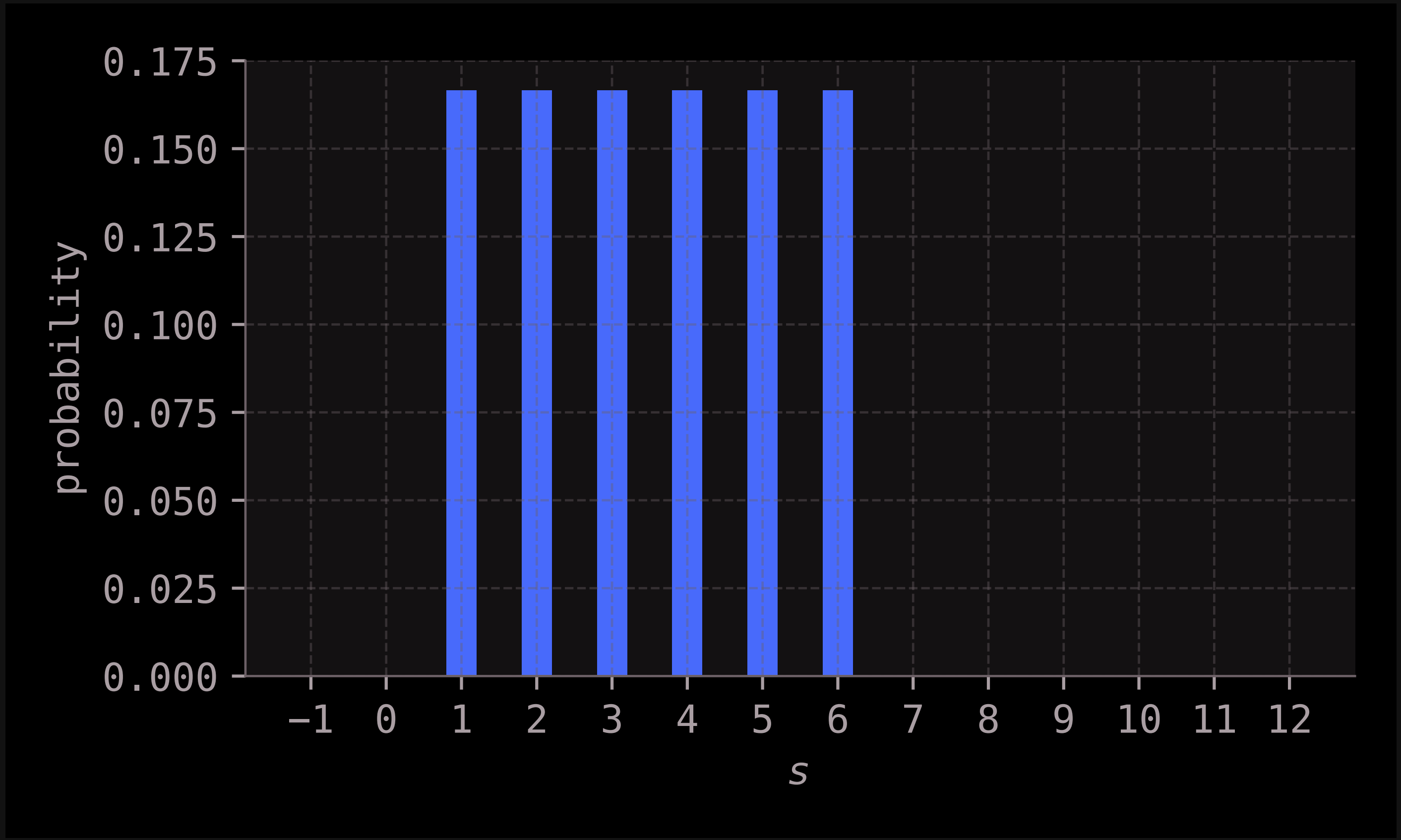
Problem Prompt

Do problems 9-11 on the worksheet.

2.7. Probability histograms



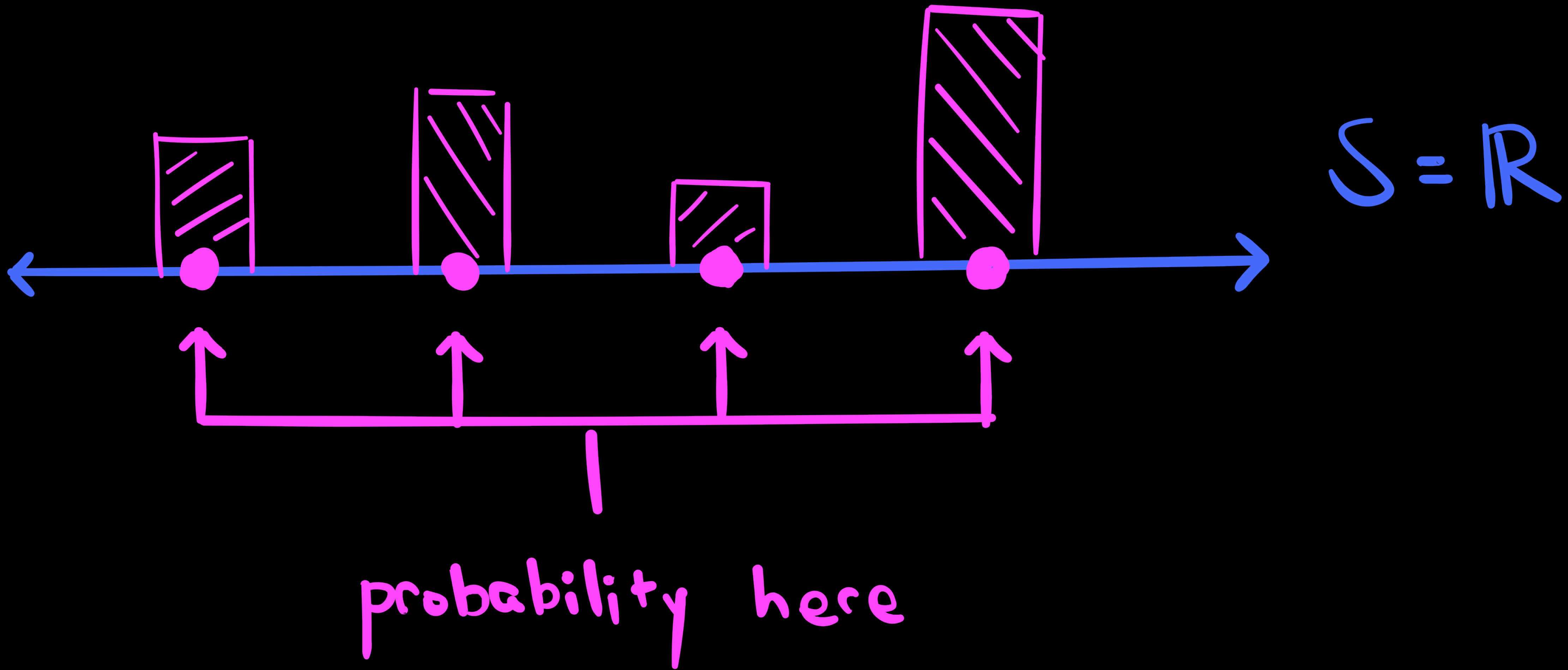
$$p(s) = \begin{cases} (0.25)(0.75)^{s-1} & : s = 1, 2, \ldots, \\ 0 & : \text{otherwise.} \end{cases}$$



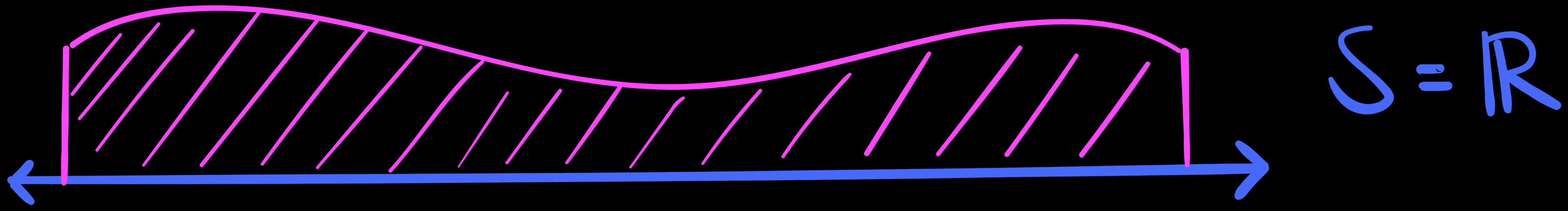
$$\text{support} = \{1, 2, 3, 4, 5, 6\} \quad \text{and} \quad p(s) = \frac{1}{6}.$$

2.8. Continuous probability measures

Discrete probability measure



continuous probability measure



probability everywhere!

Definition 2.12

Let P be a probability measure on $\mathbb{R}$. We shall say P is *continuous* if there is a function $f : \mathbb{R} \to \mathbb{R}$ such that

$$P(A) = \int_A f(s)\, \mathrm{d}s \tag{2.4}$$

for all events $A \subset \mathbb{R}$. In this case, the function $f(s)$ is called the *probability density function* (PDF) of the probability measure P, and $\mathbb{R}$ is called a *continuous probability space* (when equipped with P).

Problem Prompt

Do problems 12 and 13 on the worksheet.

Theorem 2.5 (Properties of Probability Density Functions (univariate version))

Let $f(s)$ be the probability density function of a continuous probability measure P on $\mathbb{R}$. Then:

1. $f(s) \geq 0$ for all $s \in \mathbb{R}$, and
2. $\int_{\mathbb{R}} f(s) \, \mathrm{d}s = 1$.

Theorem 2.6 (Continuous Probability Construction Lemma (univariate version))

Let $f : \mathbb{R} \to \mathbb{R}$ be a function such that

1. $f(s) \geq 0$ for all $s \in \mathbb{R}$, and
2. $\int_{\mathbb{R}} f(s)\,\mathrm{d}s = 1$.

Then there is a unique continuous probability measure P on $\mathbb{R}$ such that

$$P(A) = \int_A f(s)\,\mathrm{d}s$$

for all events $A \subseteq \mathbb{R}$.

Problem Prompt

Do problem 14 on the worksheet.

2.10. Distribution and quantile functions

Definition 2.13

A *distribution function* of a probability measure P on $\mathbb{R}$ is the function $F : \mathbb{R} \to \mathbb{R}$ such that

$$F(s) = P\big((-\infty, s]\big).$$

In particular:

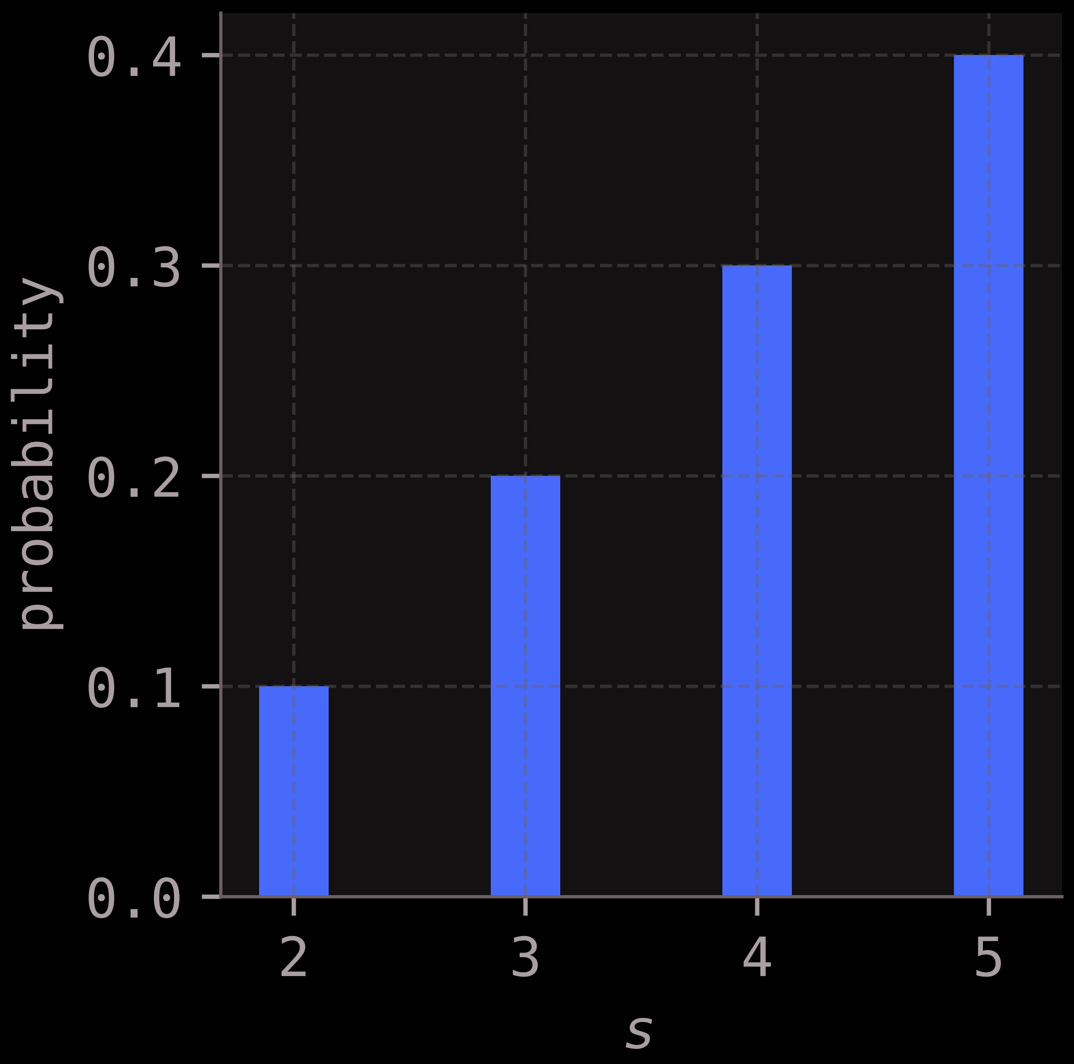
1. If P is discrete with probability mass function $p(s)$, then

$$F(s) = \sum_{s^\star \leq s} p(s^\star), \tag{2.5}$$

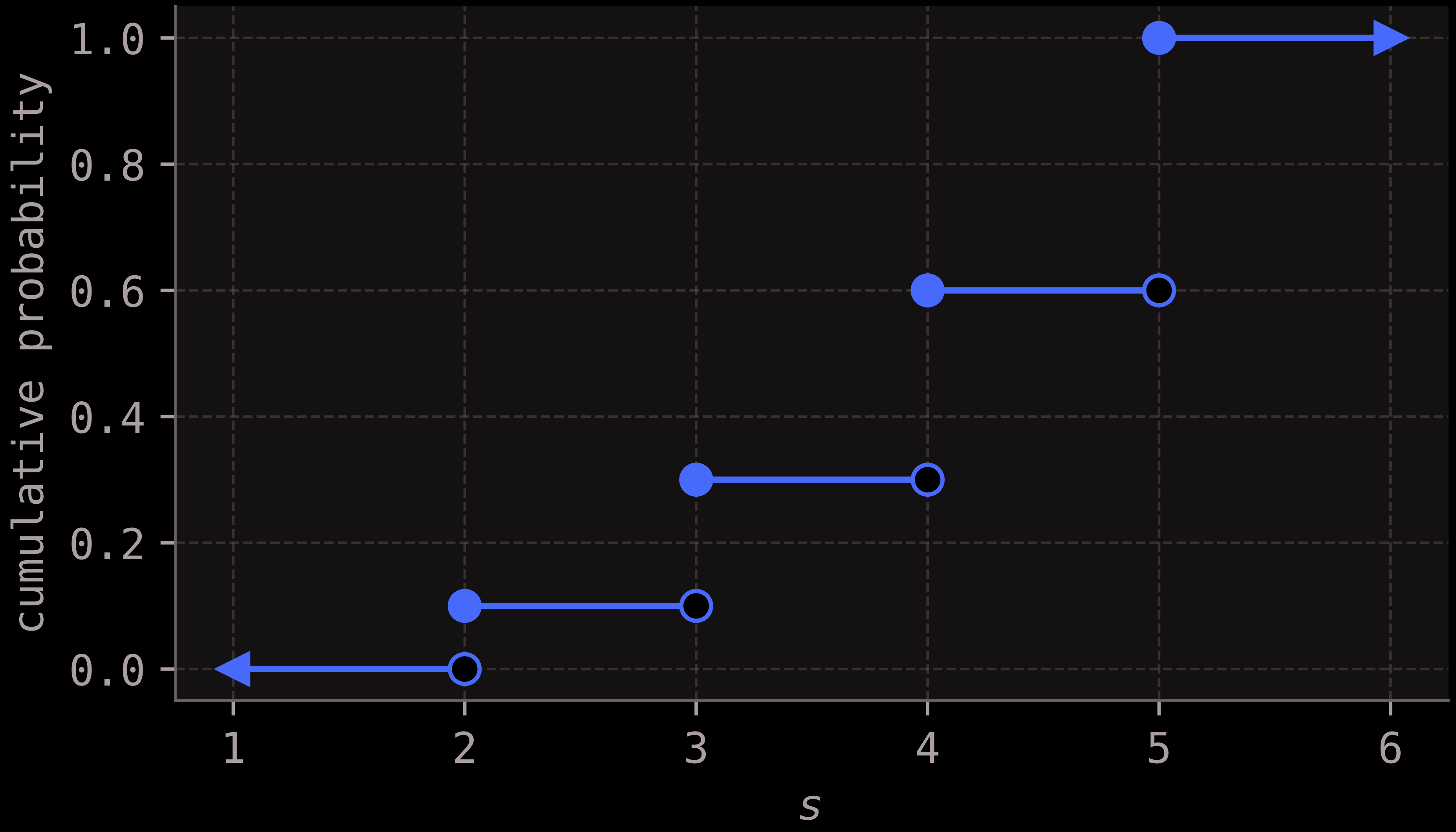
   where the sum ranges over all $s^\star \in \mathbb{R}$ with $s^\star \leq s$.
2. If P is continuous with probability density function $f(s)$, then

$$F(s) = \int_{-\infty}^{s} f(s^\star)\, \mathrm{d}s^\star.$$

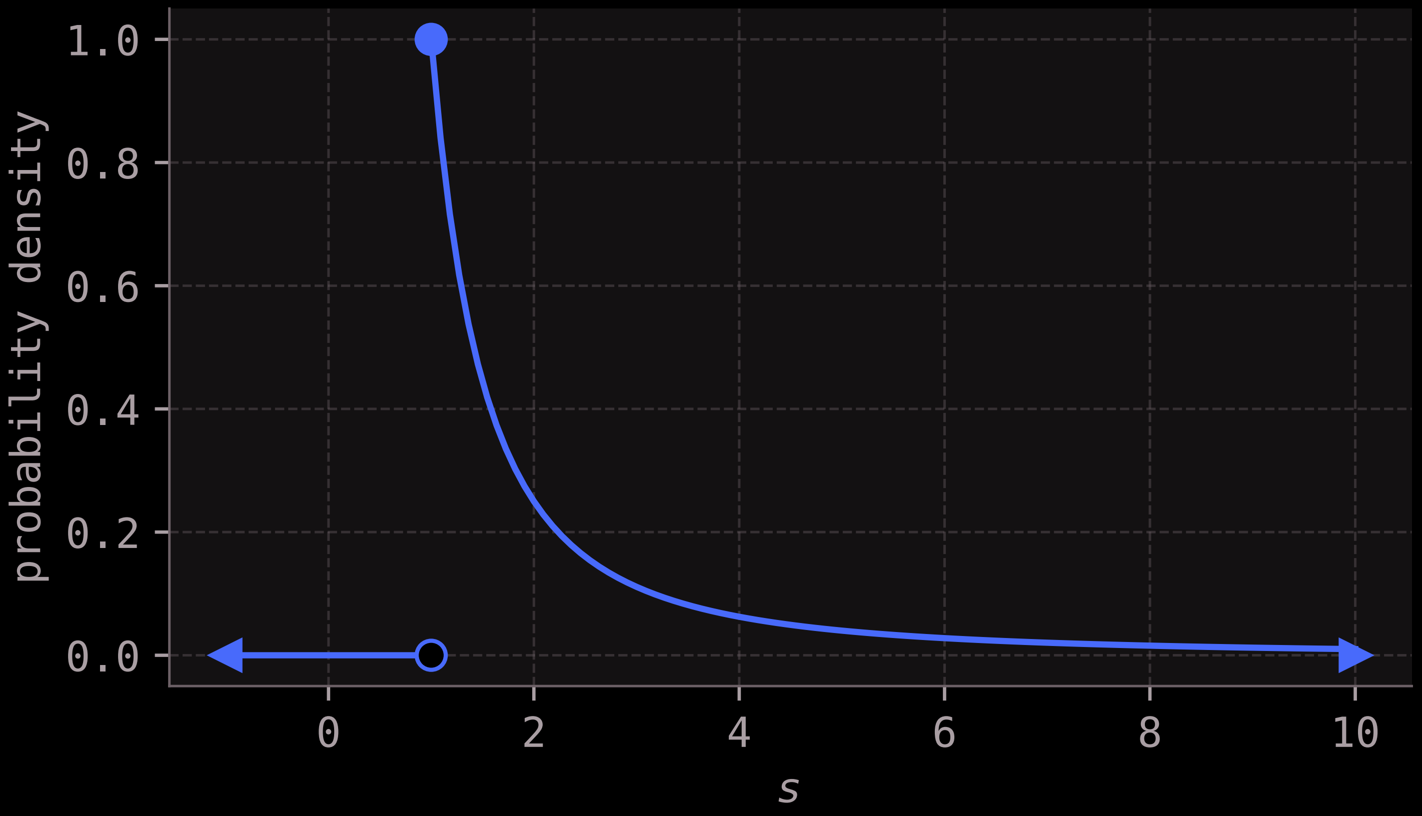
Distribution functions are also frequently called *cumulative distribution functions* (CDFs).



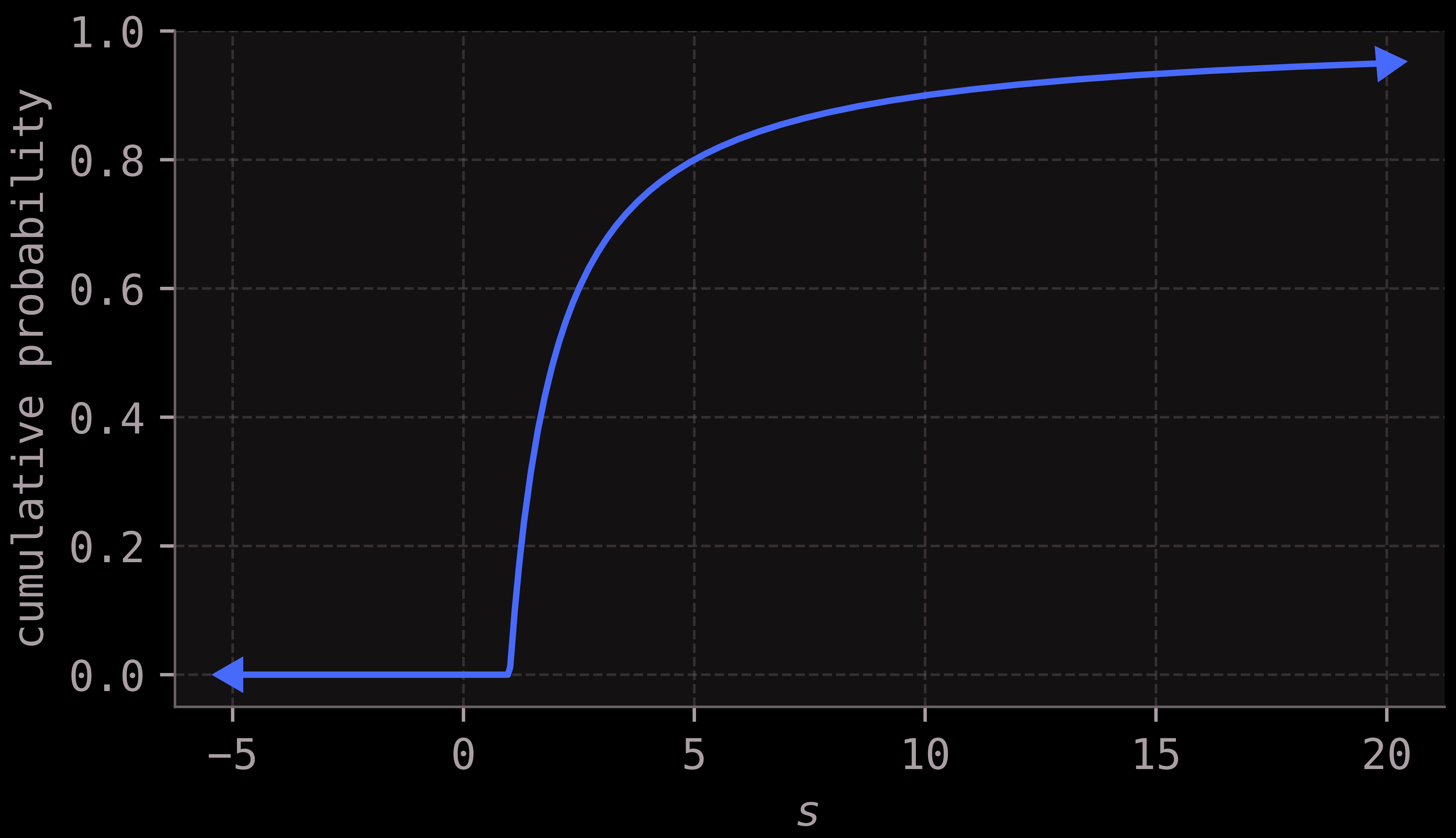
0.4
0.3
0.2
0.1
0.0
probability
2
3
4
5
s



cumulative probability
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
s



$$f(s) = \begin{cases} \frac{1}{s^2} & : s \geq 1, \\ 0 & : \text{otherwise.} \end{cases}$$



1.0
0.8
0.6
0.4
0.2
0.0
cumulative probability
−5
0
5
10
15
20
s

Theorem 2.7 (Properties of Distribution Functions)

Let $F(s)$ be the distribution function of a probability measure P on $\mathbb{R}$. Then:

1. We have

$$\lim_{s\to\infty} F(s) = 1 \quad \text{and} \quad \lim_{s\to-\infty} F(s) = 0.$$

2. $F(s)$ is a non-decreasing function, in the sense that

$$s \le t \quad \Rightarrow \quad F(s) \le F(t).$$

3. $F(s)$ is *right-continuous* at every $s \in \mathbb{R}$, in the sense that

$$F(s) = \lim_{s^\star \to s^+} F(s^\star).$$

4. The probability measure P is discrete if and only if $F(s)$ is a step function.
5. The probability measure P is continuous if and only if $F(s)$ is continuous.

Problem Prompt

Time for some practice with distribution functions. Do problems 15 and 16 on the worksheet.

Theorem 2.8 (The Fundamental Theorem of Calculus (probability version))

Let $F(s)$ be the distribution function of a probability measure P on $\mathbb{R}$. If $F(s)$ is continuous, then:

1. The measure P is continuous.
2. Wherever the derivative $F'(s)$ exists, we have

$$F'(s) = f(s),$$

   where $f(s)$ is the density function of P.

Problem Prompt

Do problem 17 on the worksheet.

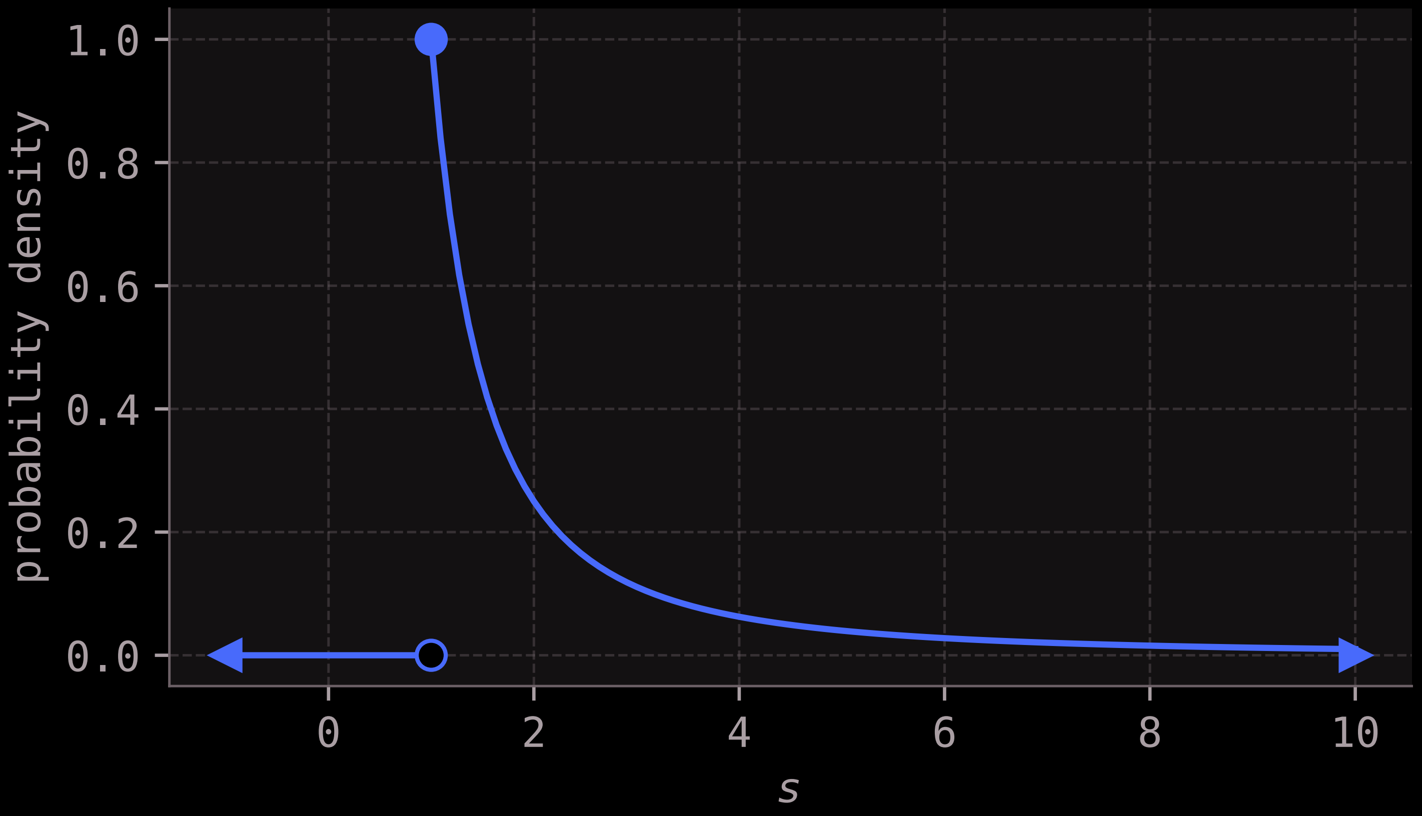
Definition 2.14

Let P be a probability measure on $\mathbb{R}$ with distribution function $F : \mathbb{R} \to [0, 1]$. The *quantile function* $Q : (0, 1) \to \mathbb{R}$ is defined so that

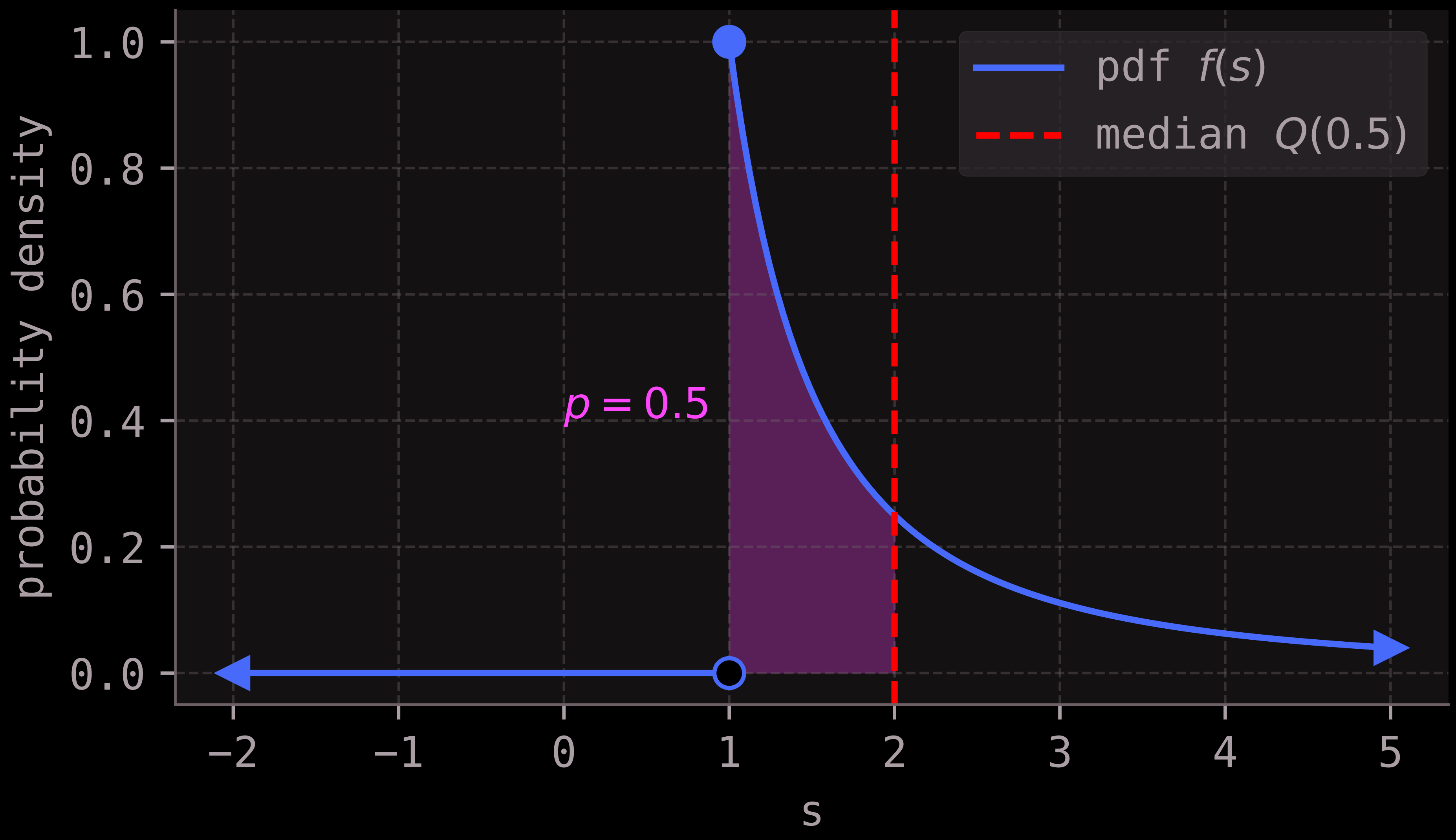
$$Q(p) = \inf\{s \in \mathbb{R} : p \leq F(s)\}.$$

In other words, the value $s = Q(p)$ is the smallest $s \in \mathbb{R}$ such that $p \leq F(s)$.

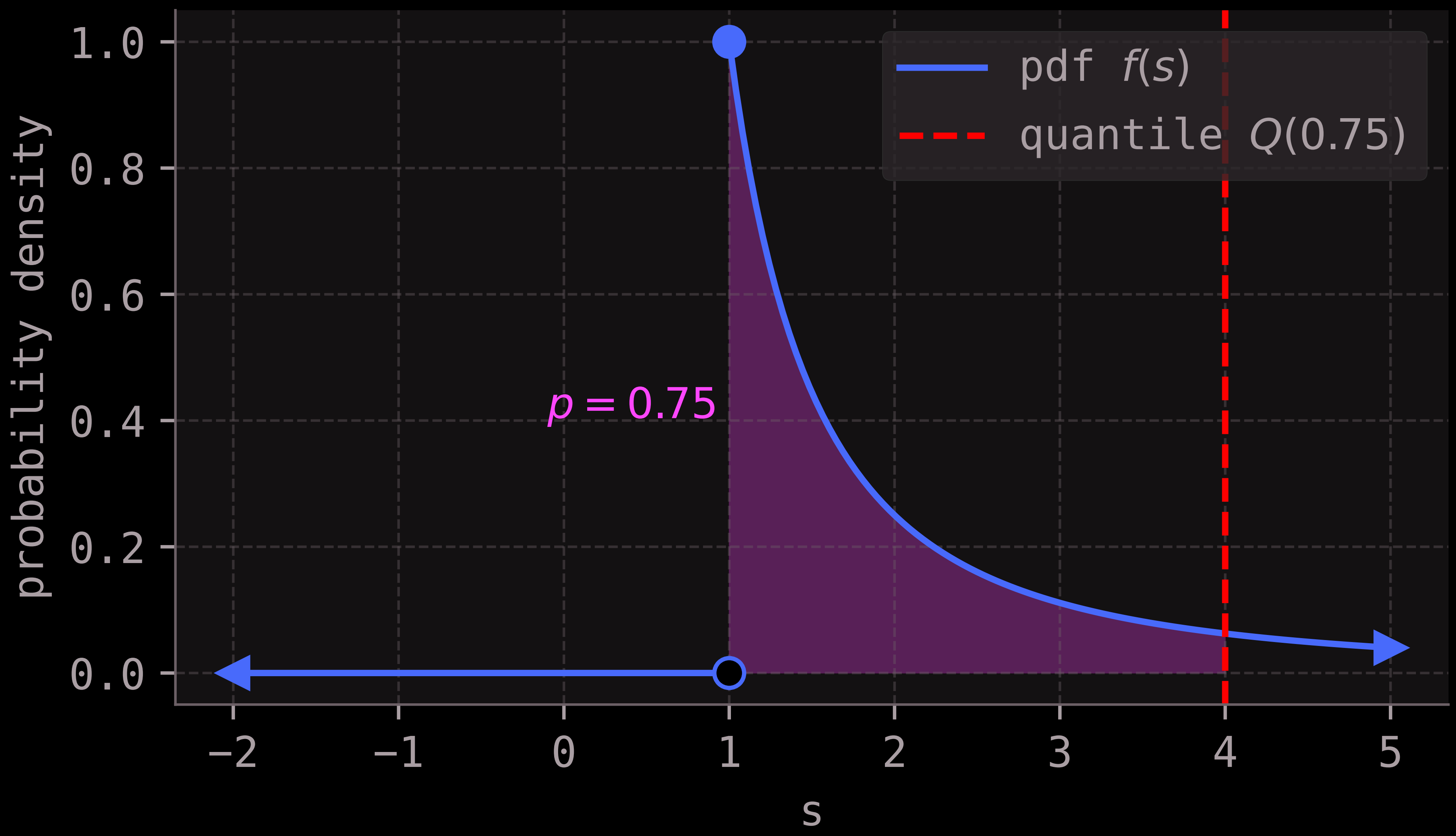
1. The value $Q(p)$ is called the *p-th quantile*.
2. The quantile $Q(0.5)$ is called the *median* of the probability measure P.



$$f(s) = \begin{cases} \dfrac{1}{s^2} & : s \geq 1, \\ 0 & : \text{otherwise.} \end{cases}$$



pdf $f(s)$
median $Q(0.5)$
$p = 0.5$
probability density
s
1.0
0.8
0.6
0.4
0.2
0.0
−2
−1
0
1
2
3
4
5



1.0
0.8
0.6
0.4
0.2
0.0
probability density
−2
−1
0
1
2
3
4
5
s
pdf $f(s)$
quantile $Q(0.75)$
$p = 0.75$

Problem Prompt

Do problems 18 and 19 on the worksheet.

2.11. Bivariate continuous probability measures

Definition 2.15

Let P be a probability measure on $\mathbb{R}^2$. We shall say P is *continuous* if there is a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that

$$P(C) = \iint_C f(s,t)\,\mathrm{d}s\mathrm{d}t$$

for all events $C \subset \mathbb{R}^2$. In this case, the function f is called the *probability density function* (PDF) of the probability measure P, and $\mathbb{R}^2$ is called a *continuous probability space* (when equipped with P).

Problem Prompt

Do problems 20 and 21 on the worksheet.

Theorem 2.9 (Properties of Probability Density Functions (bivariate version))

Let $f(s,t)$ be the probability density function of a continuous probability measure P on $\mathbb{R}^2$. Then:

1. $f(s,t) \geq 0$ for all $s, t \in \mathbb{R}$, and
2. $\iint_{\mathbb{R}^2} f(s,t)\, \mathrm{d}s\mathrm{d}t = 1$.

Theorem 2.10 (Continuous Probability Construction Lemma (bivariate version))

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function such that

1. $f(s,t) \geq 0$ for all $s, t \in \mathbb{R}$, and
2. $\iint_{\mathbb{R}^2} f(s,t)\, \mathrm{d}s\mathrm{d}t = 1$.

Then there is a unique continuous probability measure P on $\mathbb{R}^2$ such that

$$P(C) = \iint_C f(s,t)\, \mathrm{d}s\mathrm{d}t$$

for all events $C \subset \mathbb{R}^2$.

Problem Prompt

Do problem 22 on the worksheet.