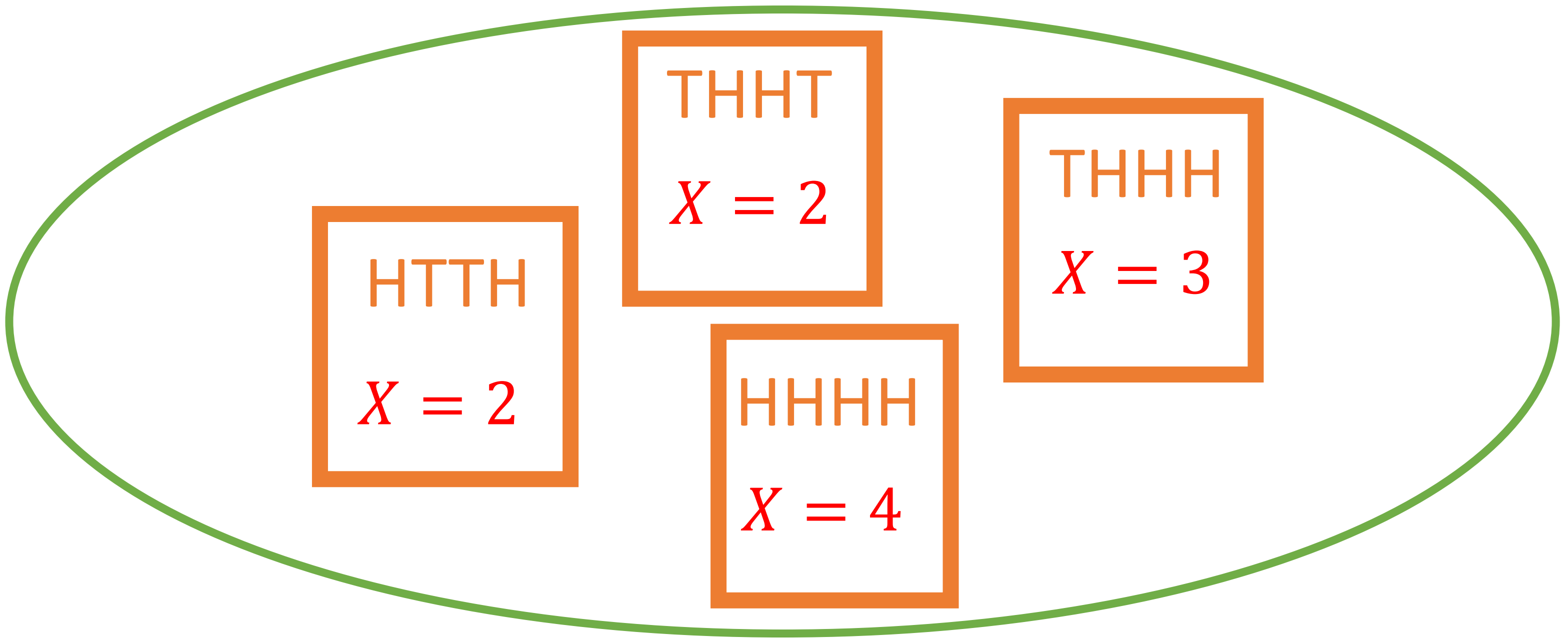
Lecture 15: Random Variables



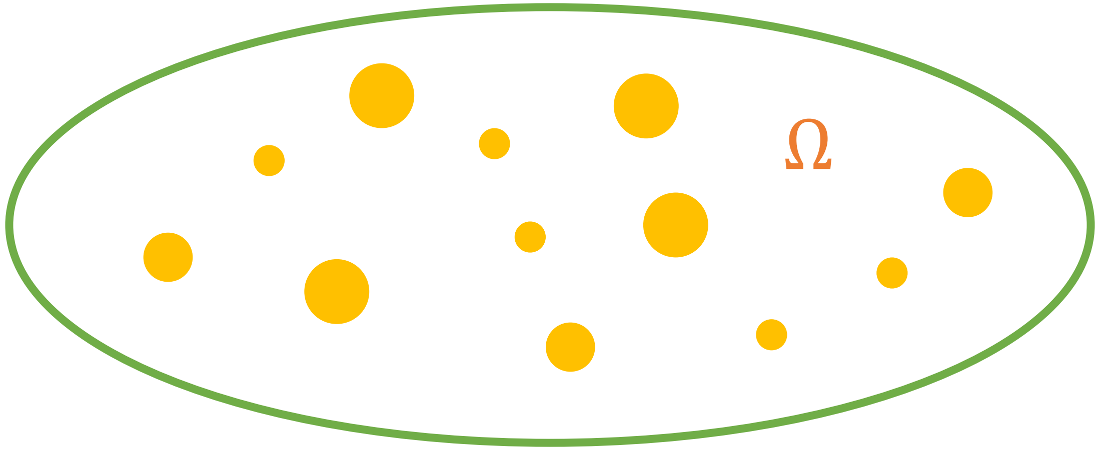
Recap: The theory of Probability

Probability Space

Sample space Ω = the set of all possible outcomes

Probability measure $\mathbb{P}: \Omega \rightarrow [0,1]$. The probability of each outcome.

$\sum_{\omega \in \Omega} \mathbb{P}[\omega] = 1$



Recap: The theory of Probability

Common pitfall

N possibilities $\neq$ 1/N probability

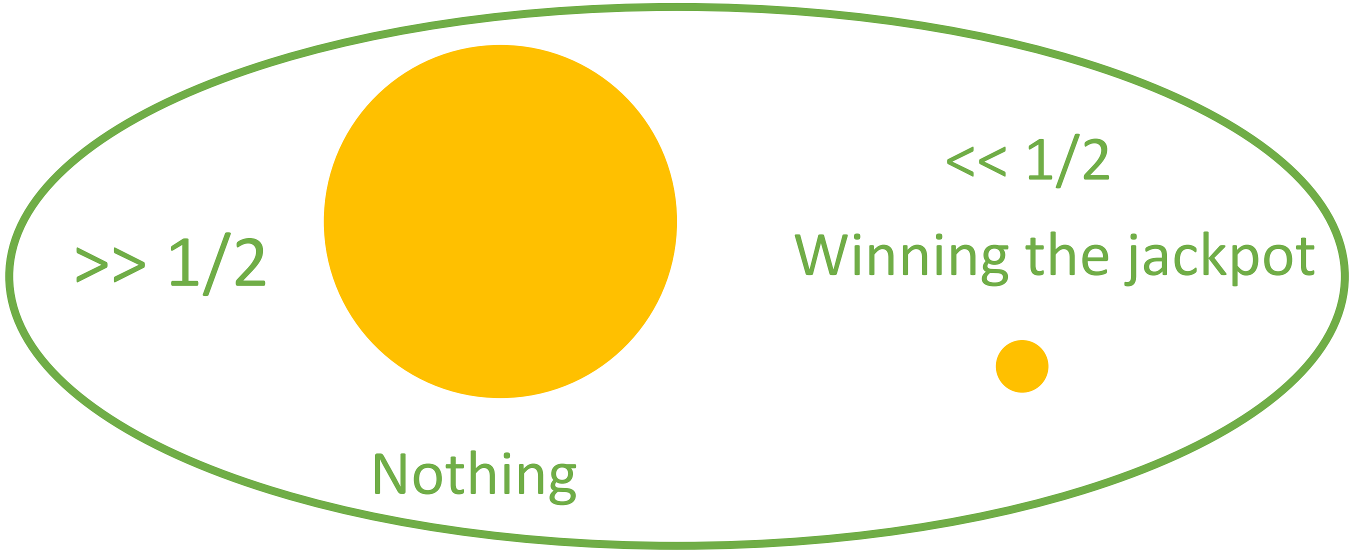
For example, lottery has 2 possibilities

Recap: The theory of Probability

Common pitfall

N possibilities $\neq$ 1/N probability

For example, lottery has 2 possibilities

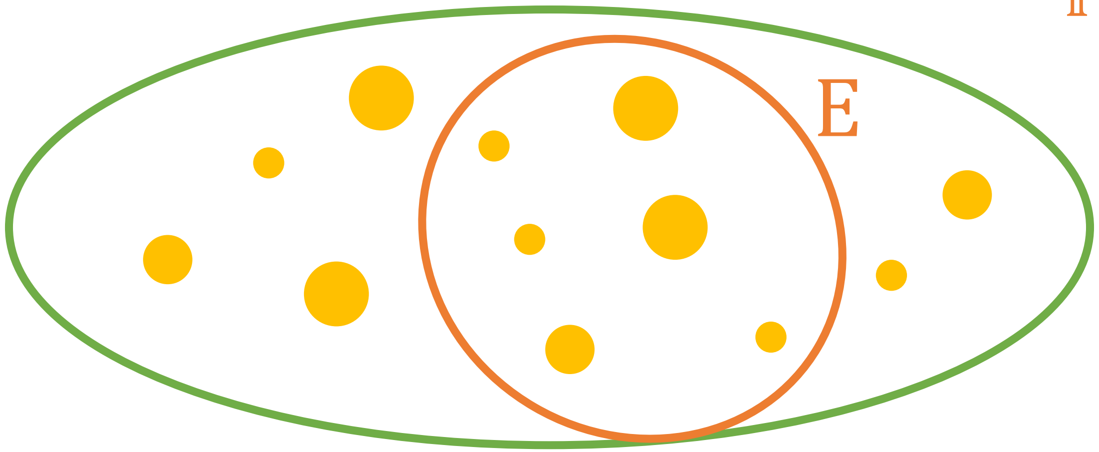


Recap: Event

Event

An event E is a subset of outcomes.

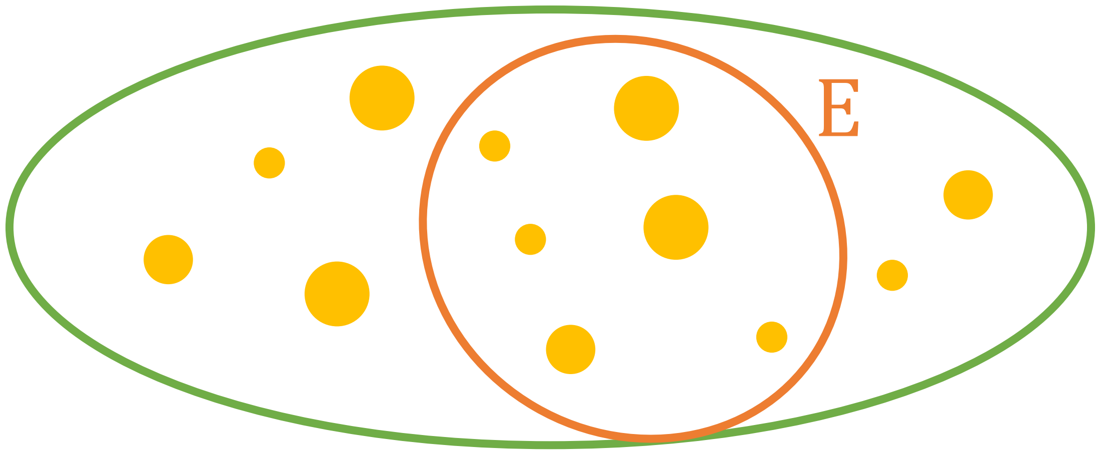
$$\mathbb{P}[E] = \sum_{\omega \in \mathrm{E}} \mathbb{P}[\omega]$$



Recap: Conditional Probability

Conditional Probability

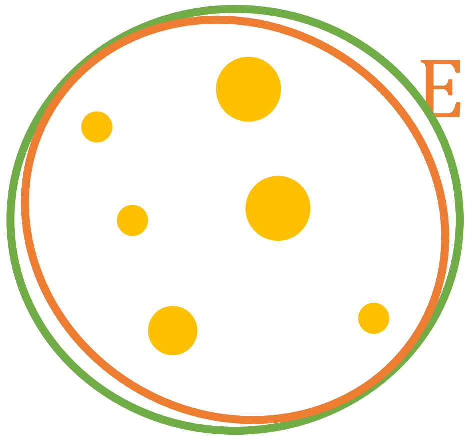
Conditioning on an event E is shrinking the probability space to E.



Recap: Conditional Probability

Conditional Probability

Conditioning on an event E is shrinking the probability space to E.

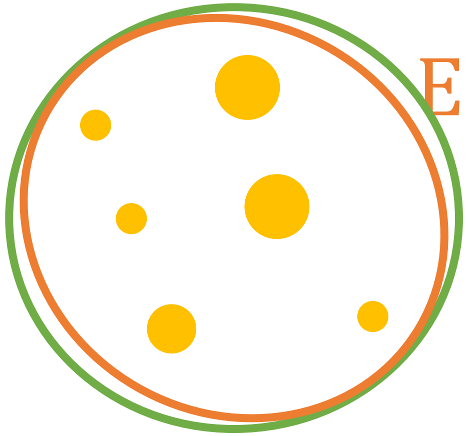


Recap: Conditional Probability

Conditional Probability

Conditioning on an event E is shrinking the probability space to E.

For every outcome $\omega \in E,\quad \mathbb{P}[\omega|\mathrm{E}] = \frac{\mathbb{P}[\omega]}{\mathbb{P}[E]}$.



So that

$$\sum_{\omega \in E} \mathbb{P}[\omega|\mathrm{E}] = \frac{\sum_{\omega \in \mathrm{E}} \mathbb{P}[\omega]}{\mathbb{P}[E]}$$

$$= 1$$

Today's Plan

Random Variables.

Definition.

Joint random variables.

Conditional random variables.

Bayesian Inference for Random Variables.

Prior/Posterior distribution

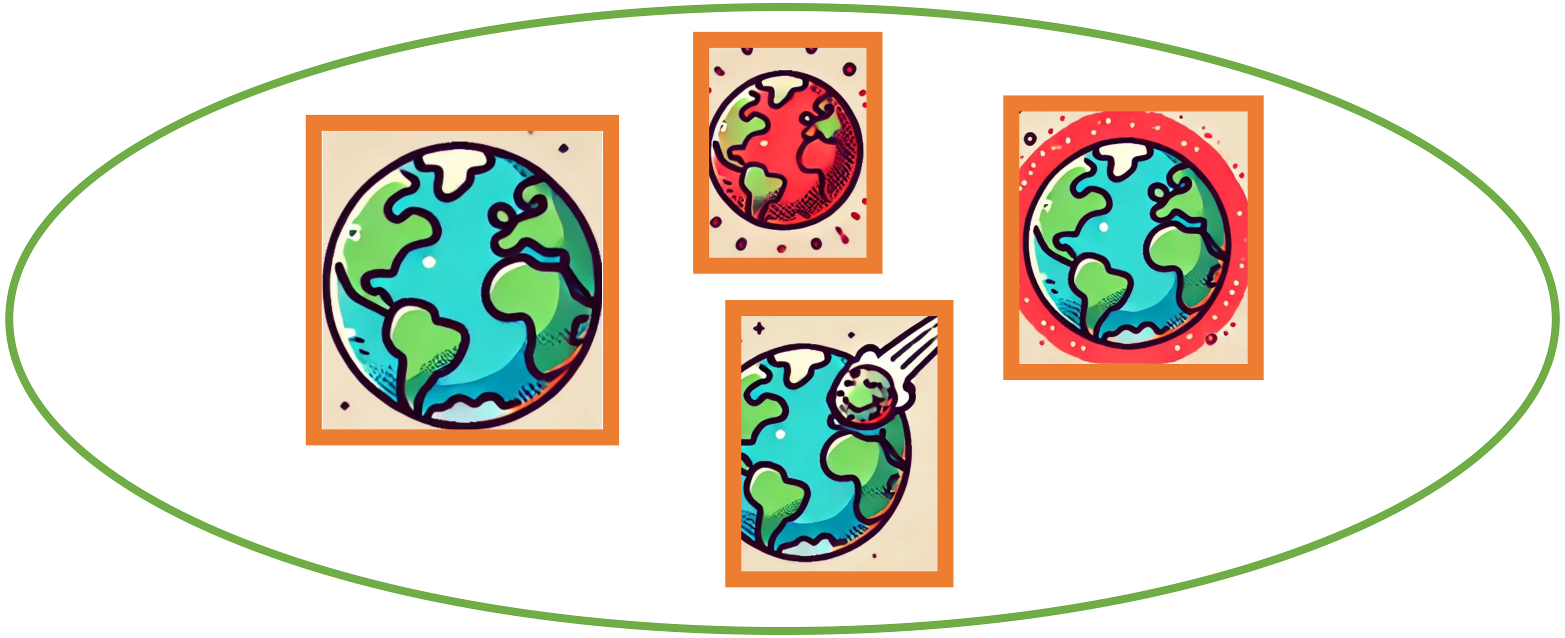
Example: Estimate the parameter of a coin. (maybe tmr)

Random Variables (Intuition)

Probability Space

Here is a way to think about probability space:

Every outcome is a state that the world could be in.

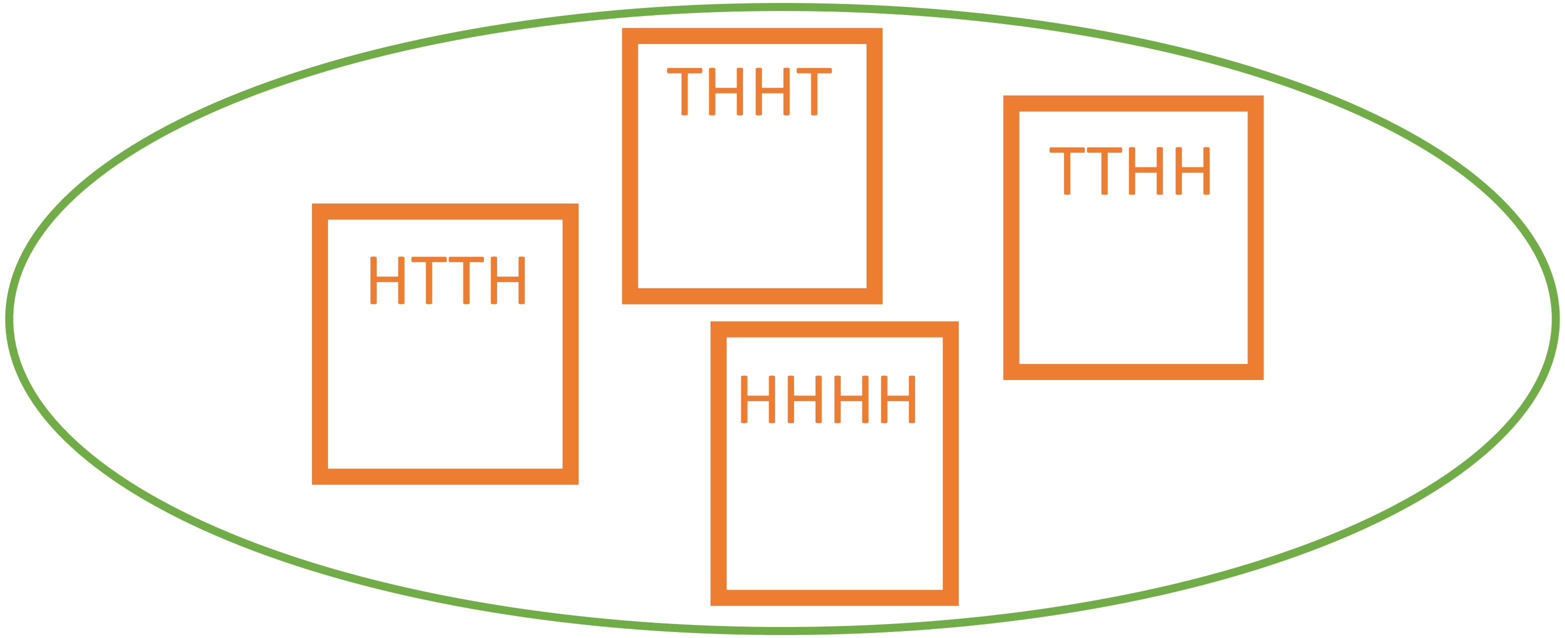


Random Variables (Intuition)

Probability Space

Here is a way to think about probability space:

Every outcome is a state that the world could be in.

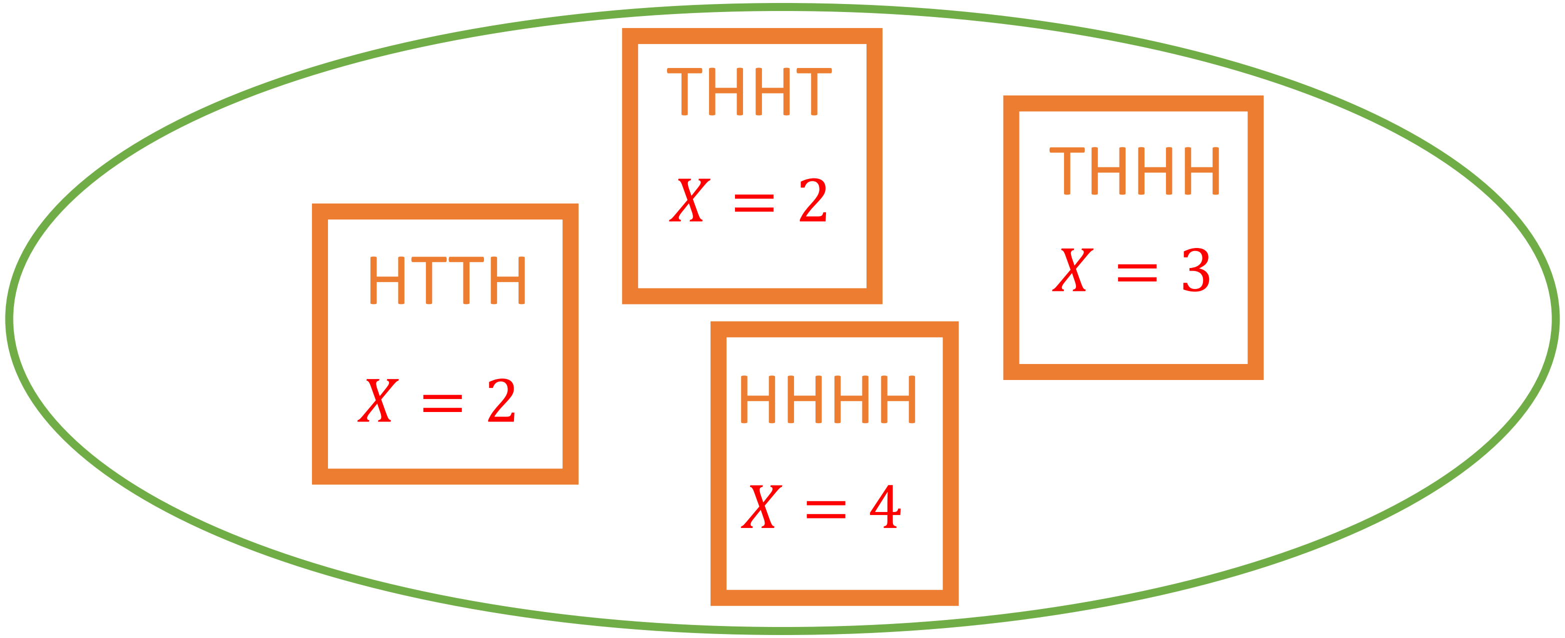


Random Variables (Intuition)

Random Variable

Consider a quantity X, say number of heads.

It has different values in different outcomes.

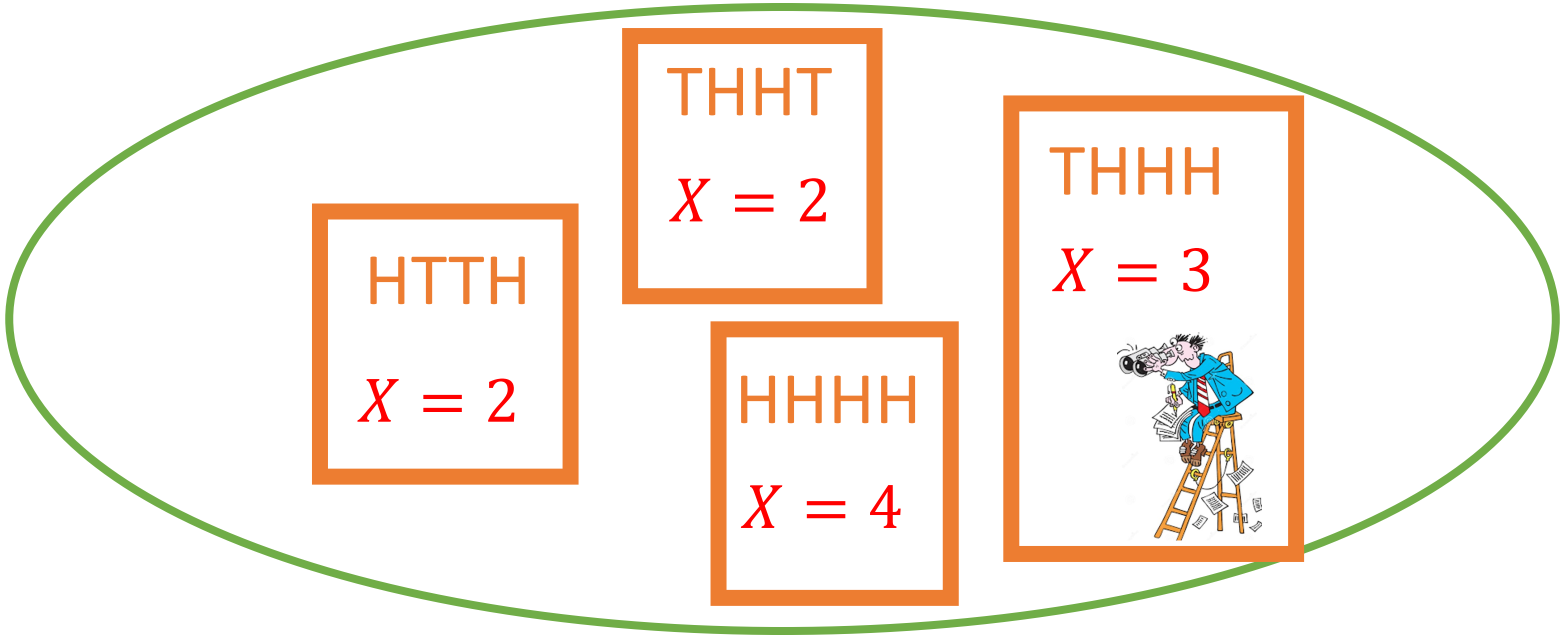


Random Variables (Intuition)

Random Variable

Consider a quantity X, say number of heads.

For an observer inside a world (after we toss the coins), X is a value.

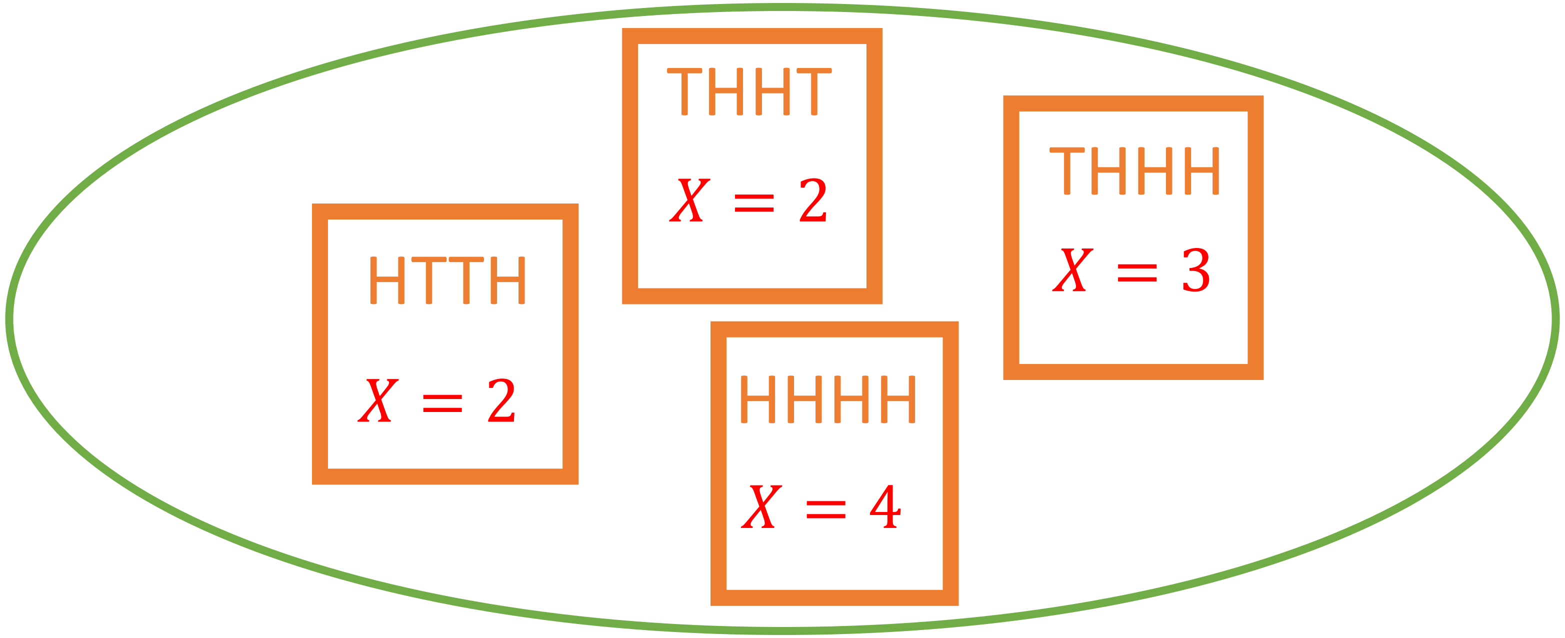


Random Variables (Intuition)

Random Variable

Consider a quantity X, say number of heads.

For an observer outside (before we toss the coins), X is a variable.

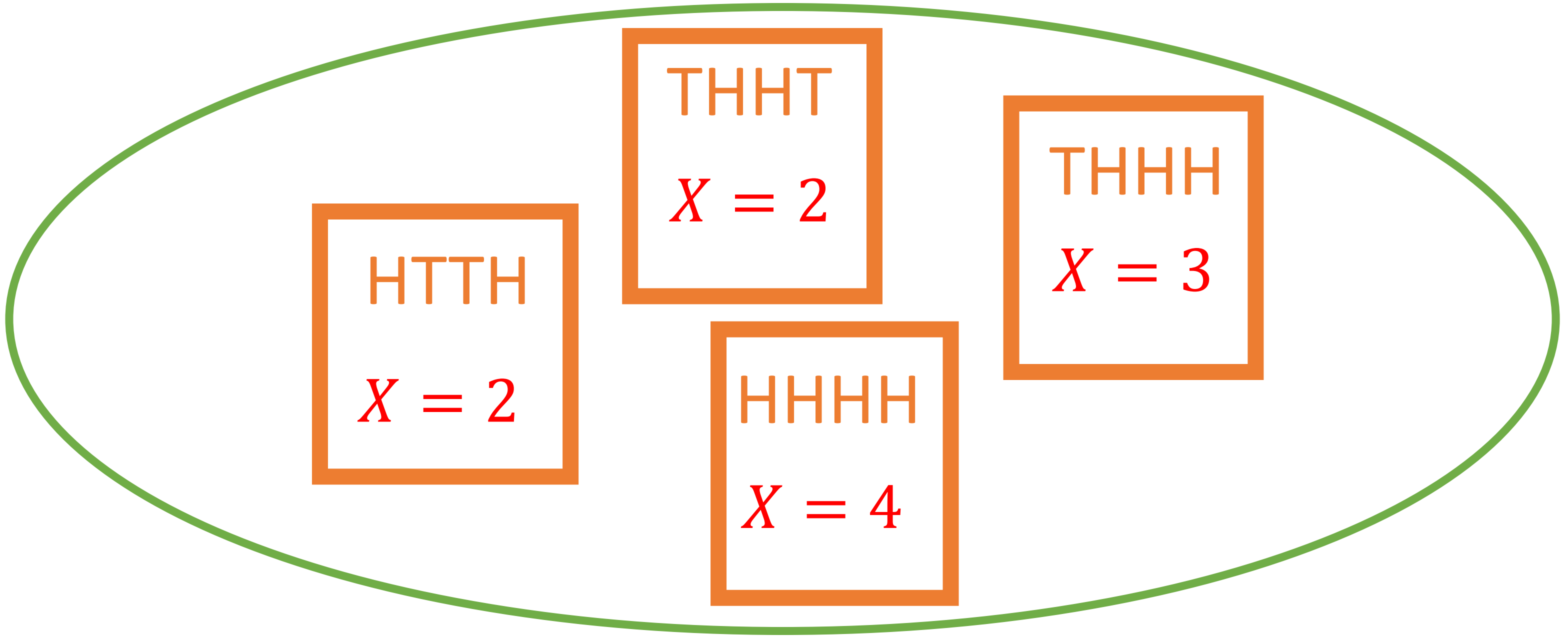


Random Variables (Formal Definition)

Definition (Random Variable)

A random variable X is a function $X: \Omega \to \mathbb{R}$.

For every outcome ω, it has a value $X(\omega)$.

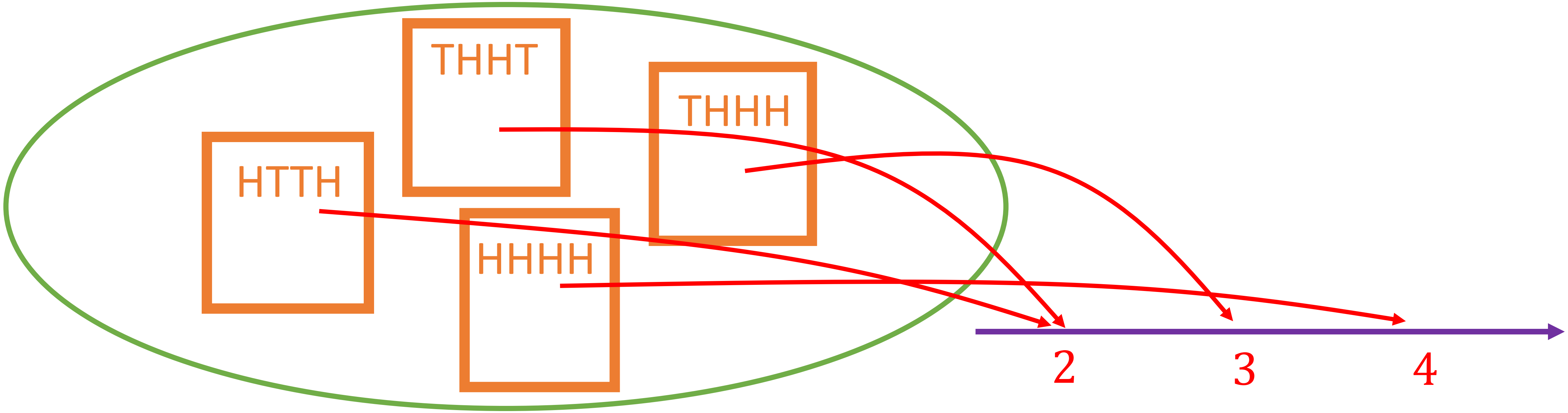


Random Variables (Formal Definition)

Definition (Random Variable)

A random variable X is a function $X: \Omega \to \mathbb{R}$.

For every outcome ω, it has a value $X(\omega)$.



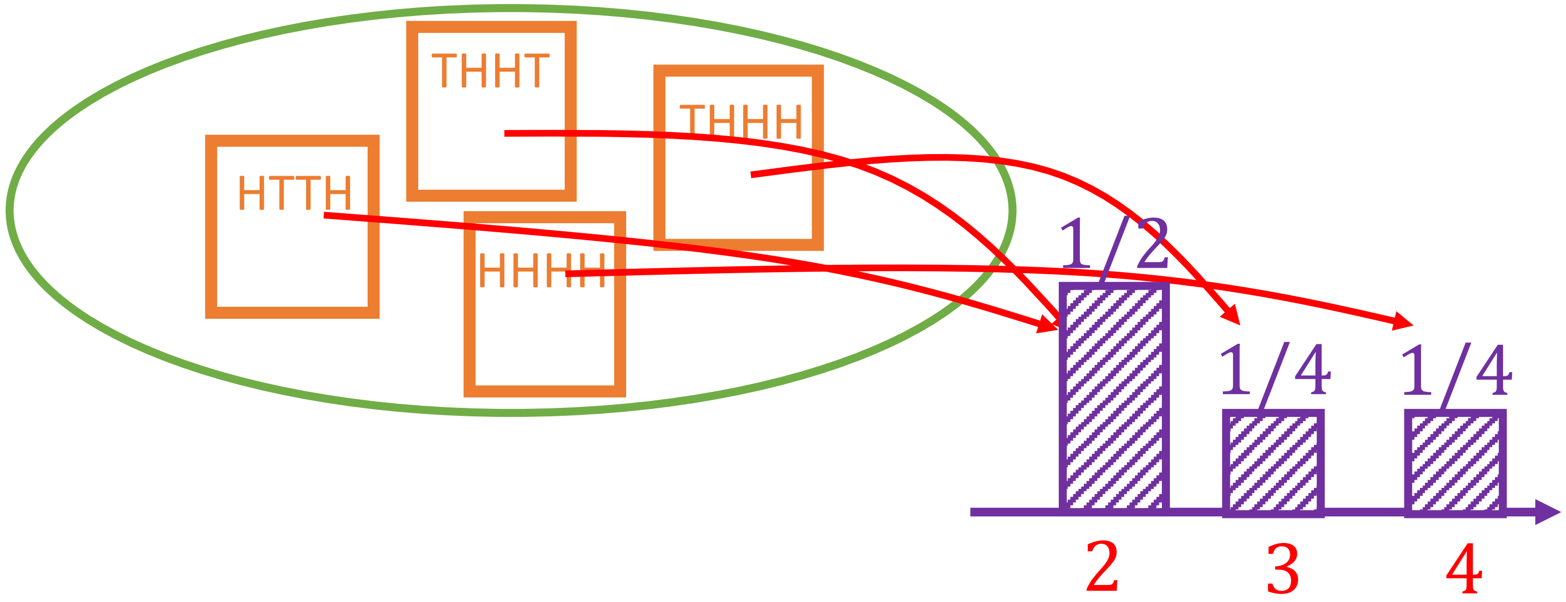
Distribution.

Definition (Distribution)

A Distribution D of random variable X is a tuple of:

- its support: All possible values of X.

- for each possible value a, the probability $\mathbb{P}[X = a]$.



Distribution.

Check.

Sum over possible value a, $\sum_a \mathbb{P}[X = a] = 1$.

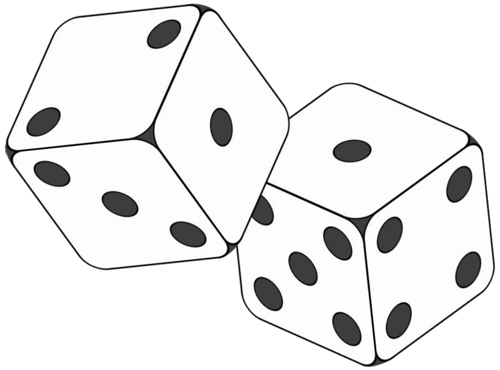
Proof.

$$\sum_a \mathbb{P}[X = a] = \sum_a \sum_{\omega : X(\omega) = a} \mathbb{P}[\omega].$$

$$= \sum_\omega \sum_{a : X(\omega) = a} \mathbb{P}[\omega]$$

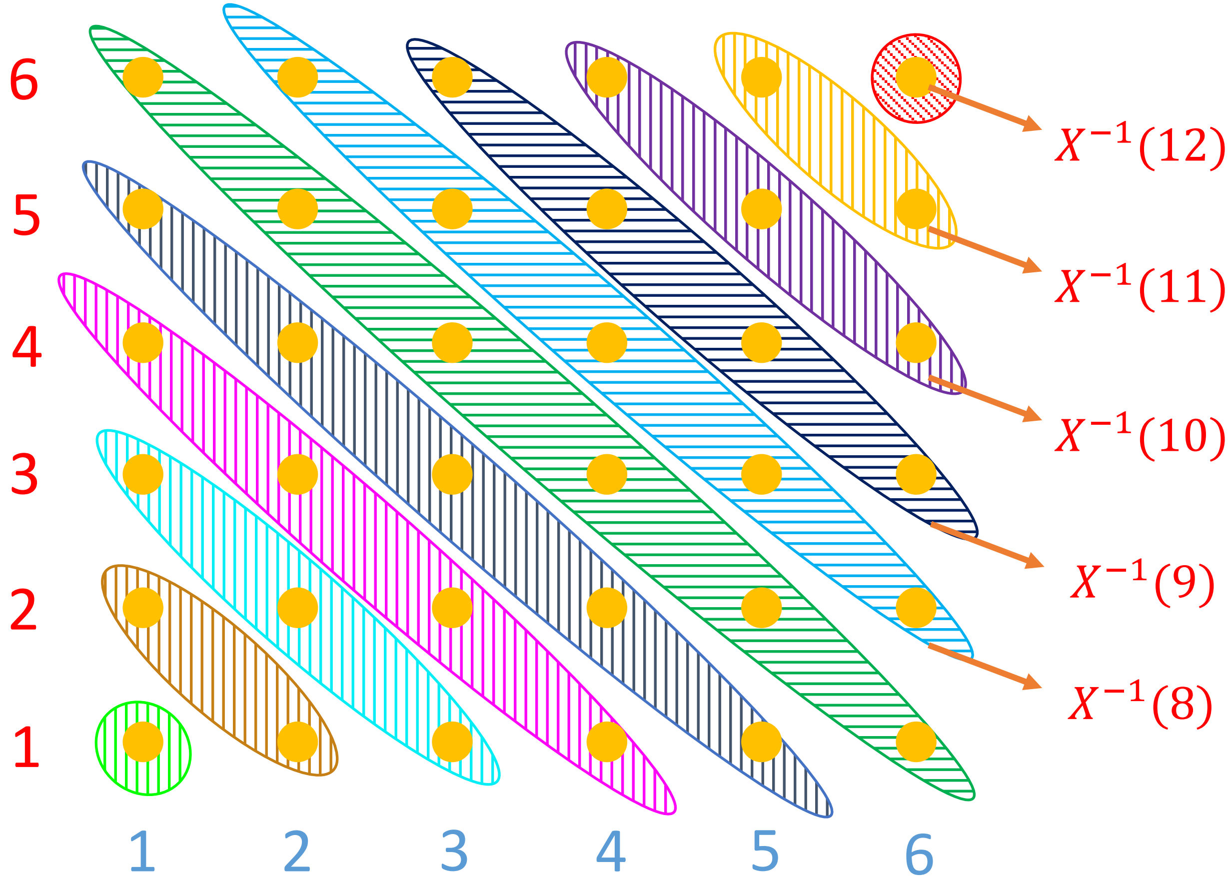
$$= \sum_\omega \mathbb{P}[\omega] = 1$$

Random Variables (Examples)

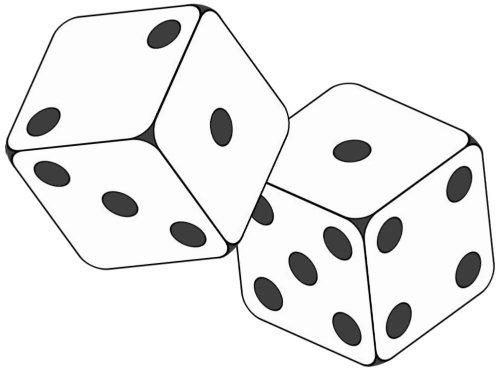


Example 1 (Rolling two dice).

Let X be the sum of two dice.

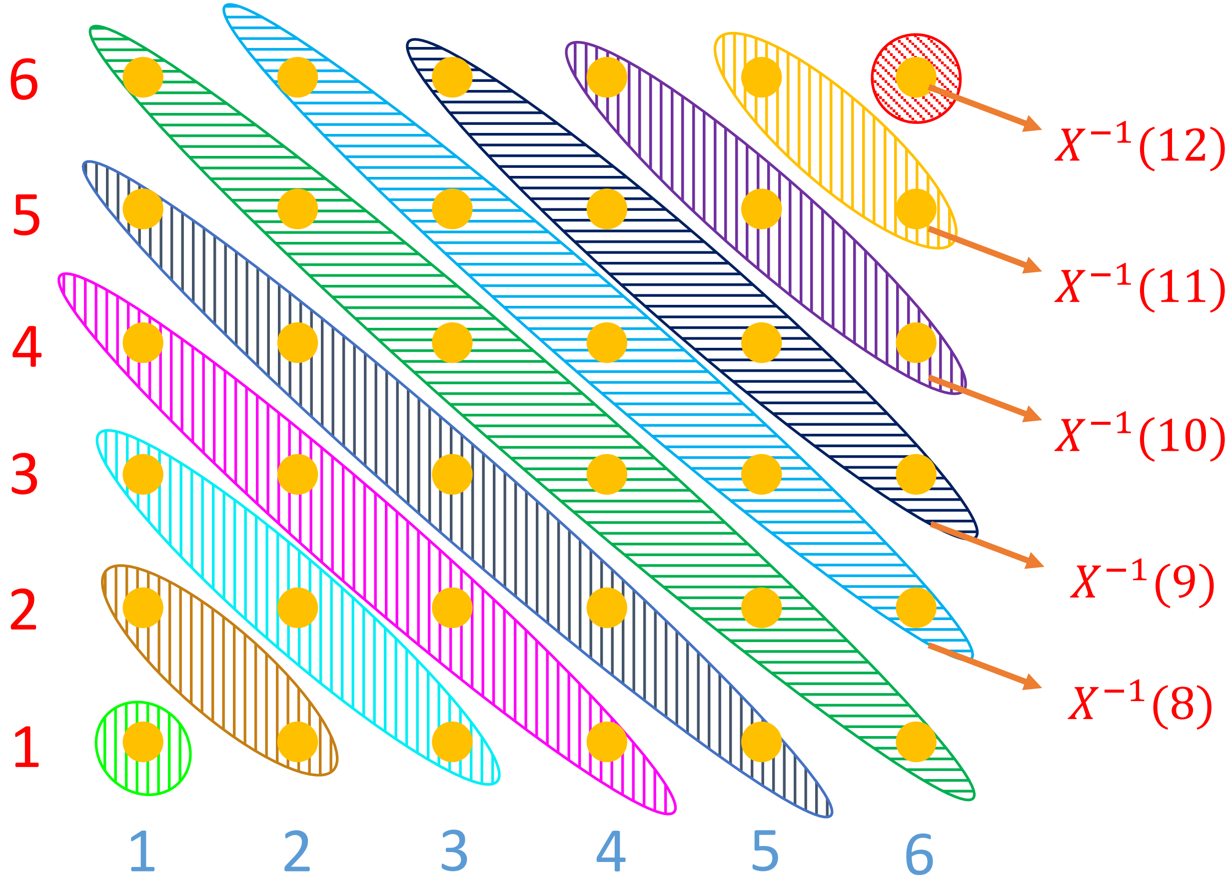


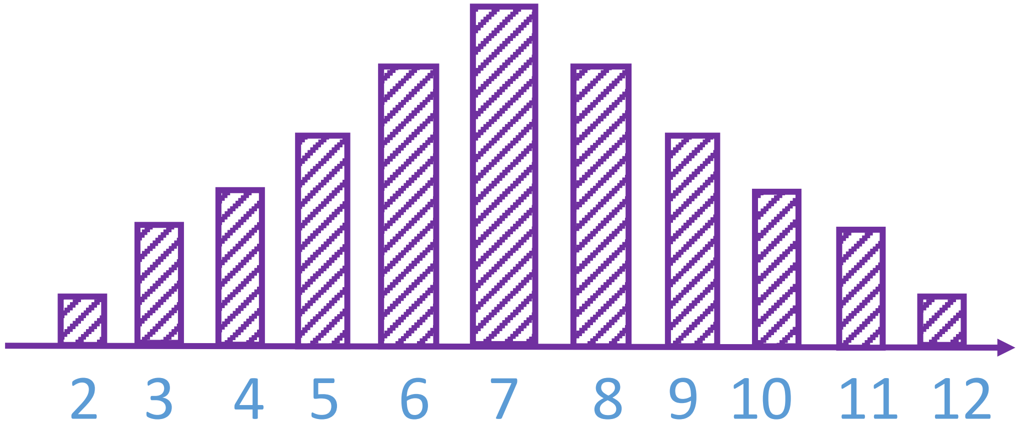
Random Variables (Examples)



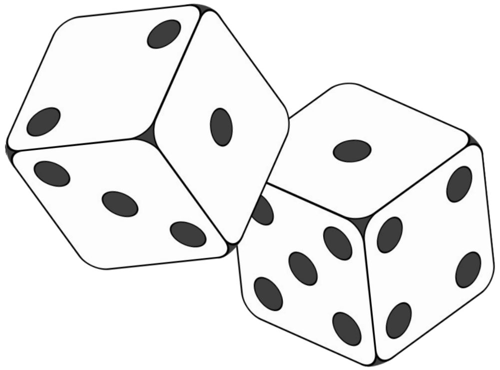
Example 1 (Rolling two dice).

Let X be the sum of two dice.



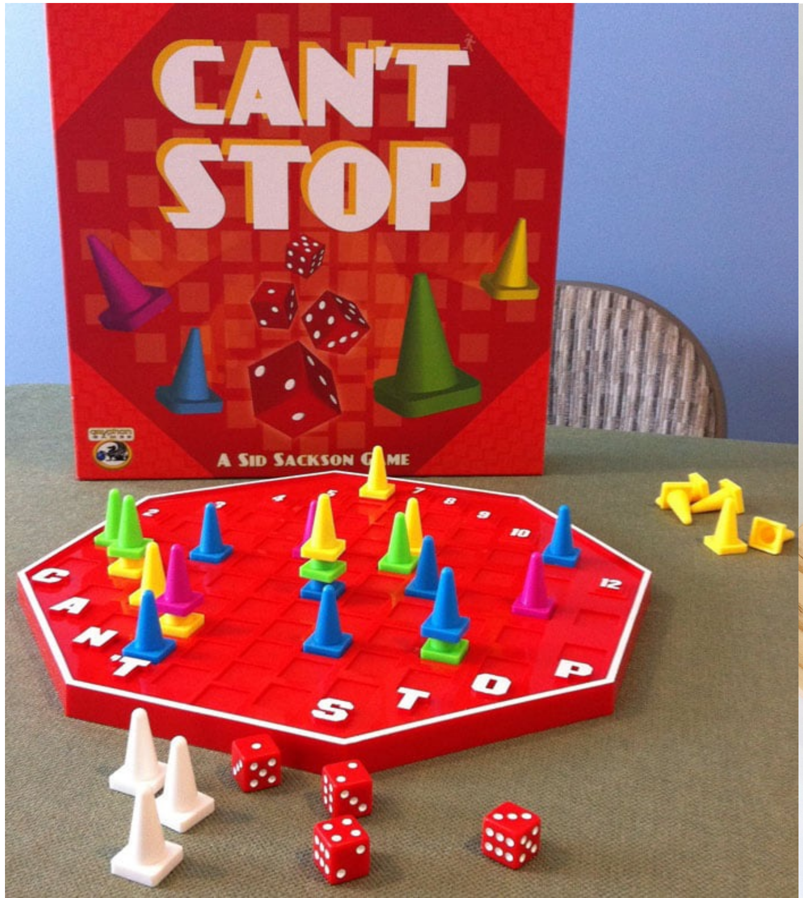


Random Variables (Examples)



Example 1 (Rolling two dice).

Let X be the sum of two dice.



Random Variables (Examples)

Example 2 (Toss 100 coins).

Let X be the number of heads.

Because all outcomes are equally likely (uniform distribution),

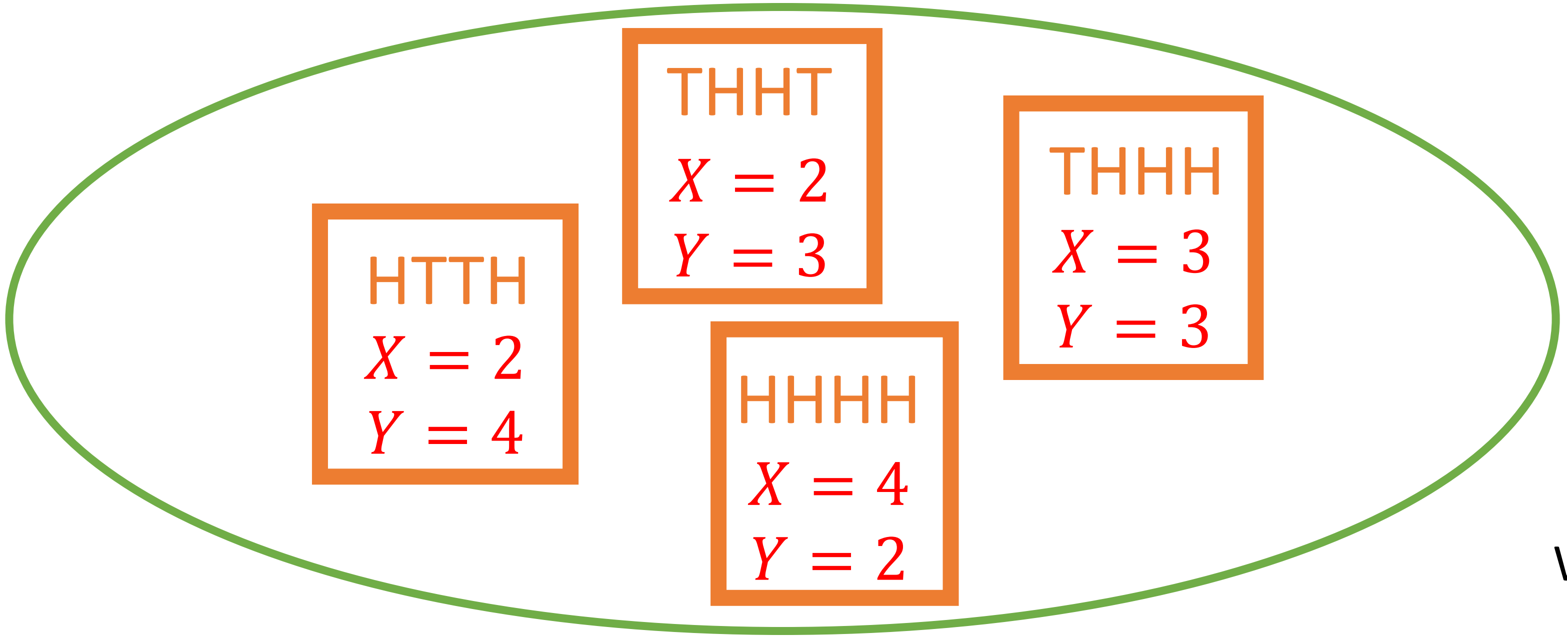
$$\mathbb{P}[X = a] = \frac{\text{Outcomes with } a \text{ heads}}{\text{Total number of outcomes}} = \frac{\binom{100}{a}}{2^{100}}.$$

Joint Random Variable (Definition)

Definition (Joint Random Variable)

For two random variable X, Y that are functions $X, Y: \Omega \to \mathbb{R}$.

For every outcome ω, it has a value $X(\omega)$ and a value $Y(\omega)$.



What is X? What is Y?

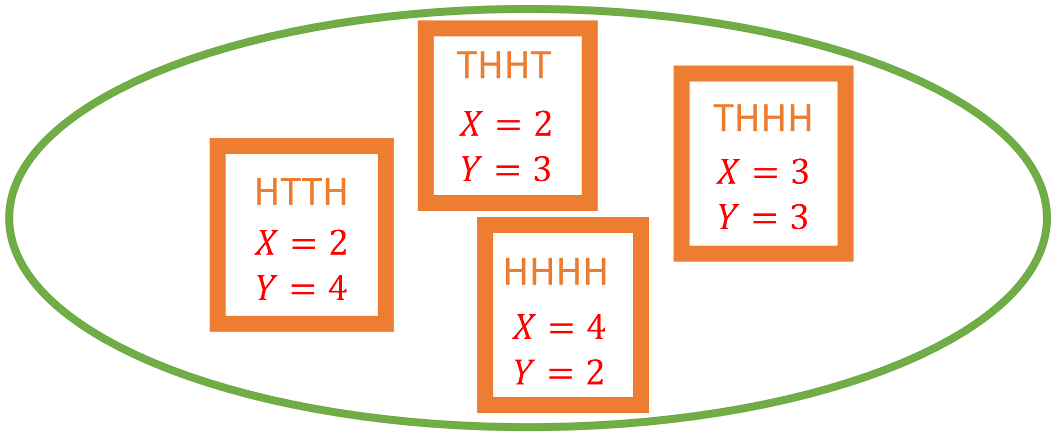
Joint Distribution (Definition)

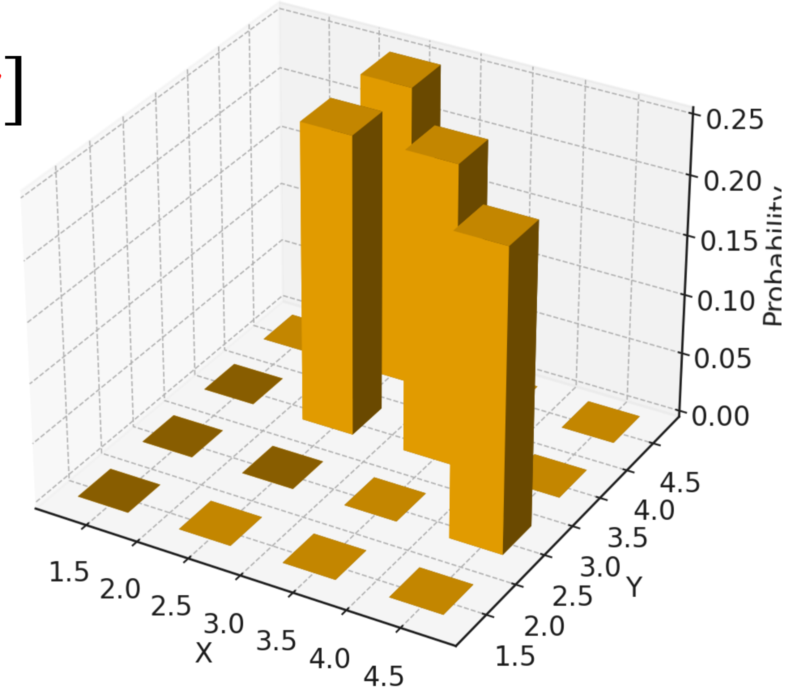
Definition (Joint Distribution)

The joint distribution of X, Y has:

- support over pairs of possible values (x, y).
- For each (x, y), a probability

$$\mathbb{P}[X = x \wedge Y = y]$$



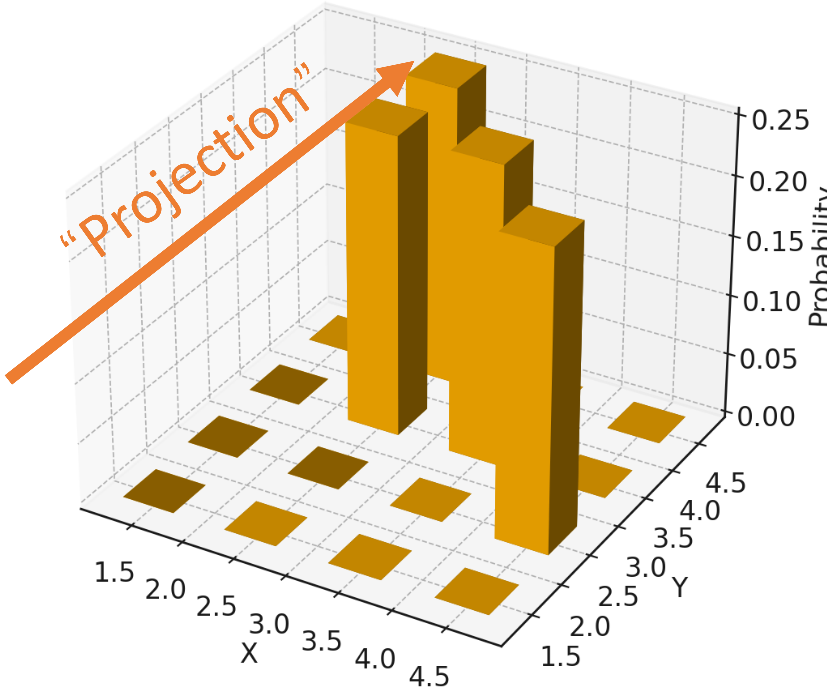


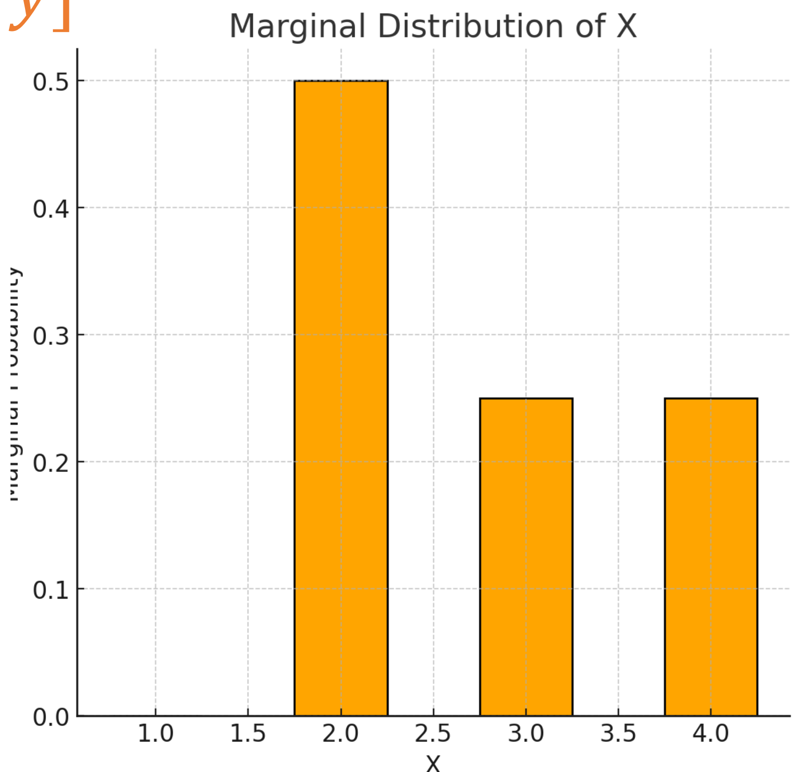
Marginal Distribution (Definition)

Definition (Marginal Distribution)

Given the joint distribution of X, Y, we can calculate the distribution of X, called the X-marginal distribution.

$$\mathbb{P}[X = x] = \sum_y \mathbb{P}[X = x, Y = y]$$





Independence (Definition)

Equivalent Definition 1:

We say two jointly distributed random variables, X, Y are independent if
$\mathbb{P}[X = x \mid Y = y] = \mathbb{P}[X = x]$.

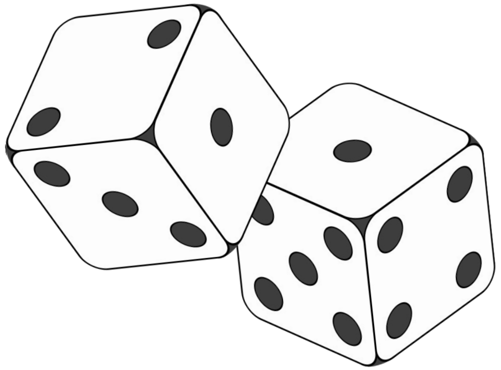
"independence ⇔ conditioning does not change distribution."

Equivalent Definition 2:

We say two jointly distributed random variables, X, Y are independent if
$\mathbb{P}[X = x \wedge Y = y] = \mathbb{P}[X = x]\mathbb{P}[Y = y]$.

"independence ⇔ Joint distribution = product of the marginal distributions."

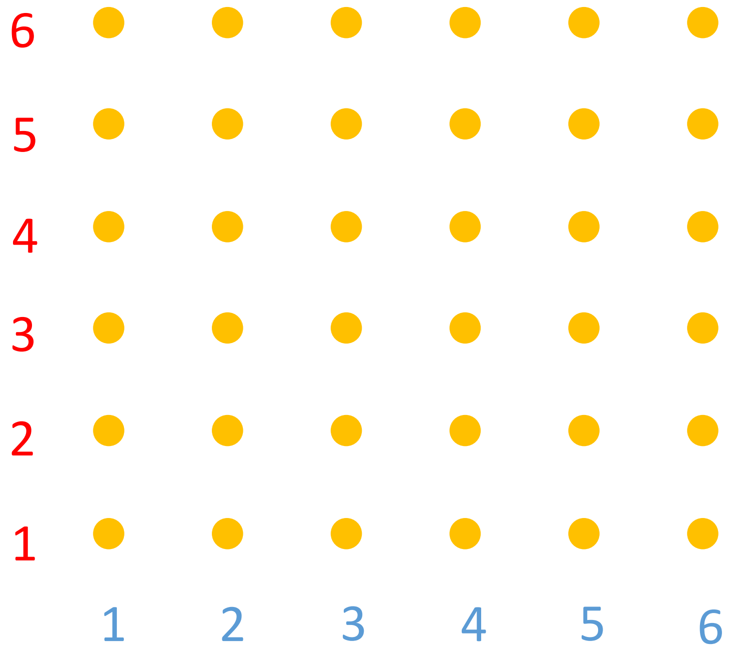
Joint Distribution (Example)



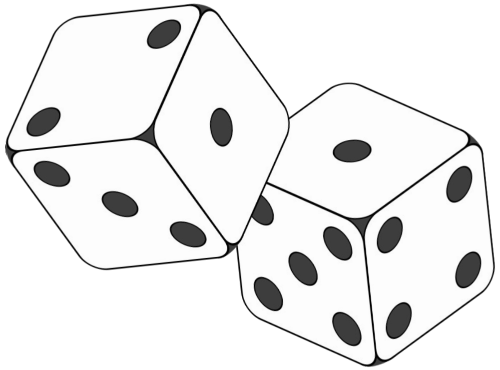
Example 1 (Rolling two dice).

Let X_1 be the first die and X_2 be the second die.

We have the following probability space.



Joint Distribution (Example)



Example 1 (Rolling two dice).

Let X_1 be the first die and X_2 be the second die.

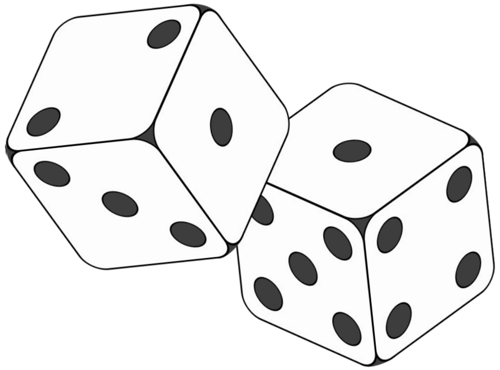
The joint distribution is just uniform.

$$\mathbb{P}[X_1 = a, X_2 = b] = \frac{1}{36}$$

We can calculate the marginal distribution.

$$\mathbb{P}[X_1 = a] = \frac{1}{6}, \mathbb{P}[X_2 = b] = \frac{1}{6}$$

Joint Distribution (Example)



Example 1 (Rolling two dice).

Let X_1 be the first die and X_2 be the second die.

The joint distribution is just uniform.

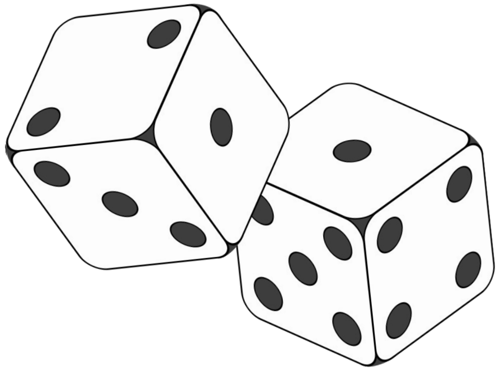
$$\mathbb{P}[X_1 = a, X_2 = b] = \frac{1}{36}$$

We can calculate the marginal distribution.

$$\mathbb{P}[X_1 = a] = \frac{1}{6}, \mathbb{P}[X_2 = b] = \frac{1}{6}$$

X_1 and X_2 are independent because the joint distribution is the product of two marginal distributions.

Joint Distribution (Example)



Example 2 (Rolling two dice).

Let S be the sum of two dice and

X_2 be the second die.

They are NOT independent.

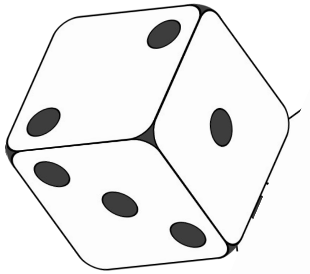
$\mathbb{P}[X_2 = 1 \mid S = 2] = 1$

$\mathbb{P}[X_2 = 1 \mid S = 12] = 0$

Here we use the first equivalent definition:

Conditioning on the value of S changes the distribution of X_2

Joint Distribution (Example)



Example 3 (Rolling one die).

Let $X_{(2)}$ be the die mod 2 and

$X_{(3)}$ be the die mod 3.

Are they independent?

Two equivalent way to roll the die:

1. Generate a random number $X = \{0,1,2,\ldots,5\}$.
2. Generate random $X_{(2)} = \{0,1\}$ and $X_{(3)} = \{0,1,2\}$.

Merge via CRT to get X.

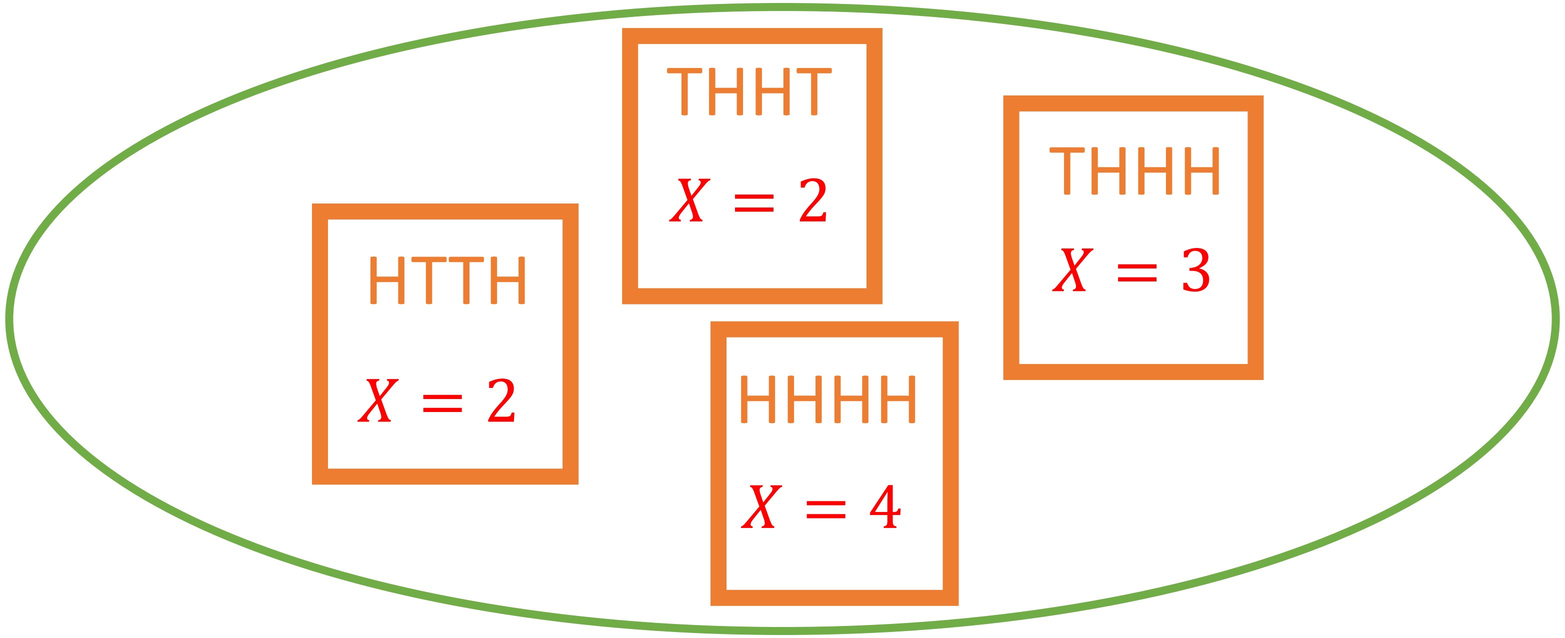
These two processes are completely equivalent.

Conditional Random Variables (Intuition)

Conditioning

Say the observer know extra information:

The number of head is even.

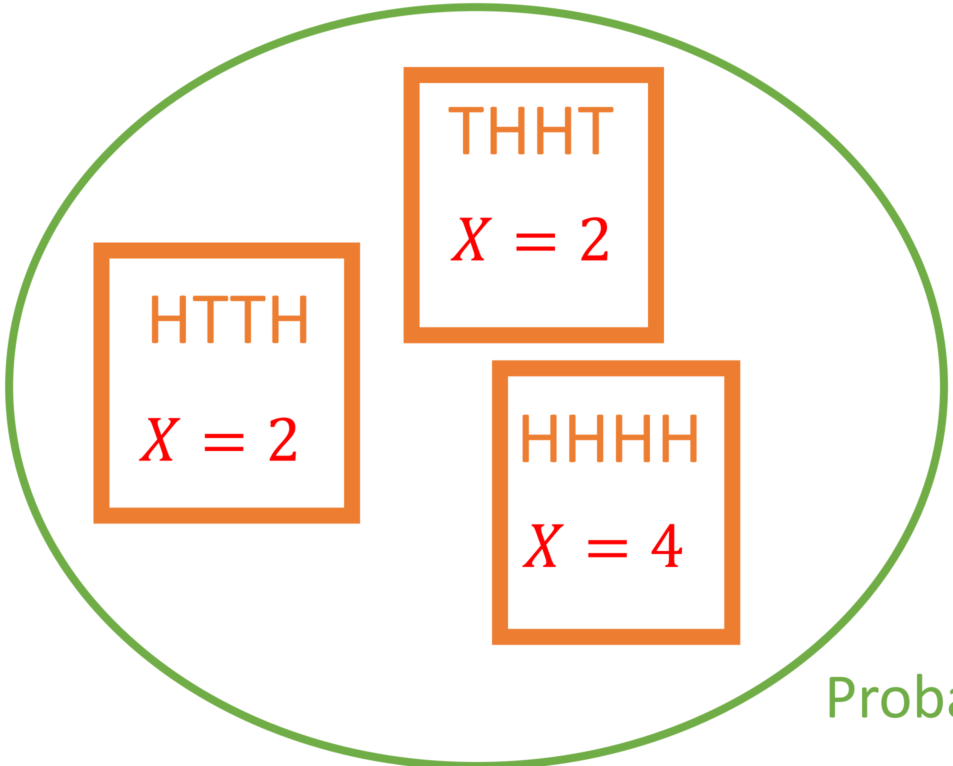


Conditional Random Variables (Intuition)

Conditioning

Say the observer know extra information:

The number of head is even.



Probability space shrinks.

Conditional Random Variables (Formal Definition)

Definition (Conditional Random Variable)

A random variable X is a function $X: \Omega \to \mathbb{R}$.

After conditioning on a event E, we get $X|E: E \to \mathbb{R}$

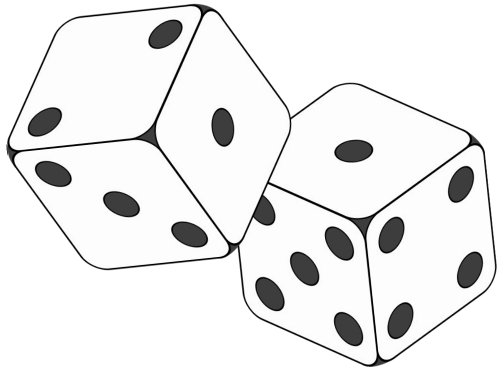
which is just the function X restrict to E.

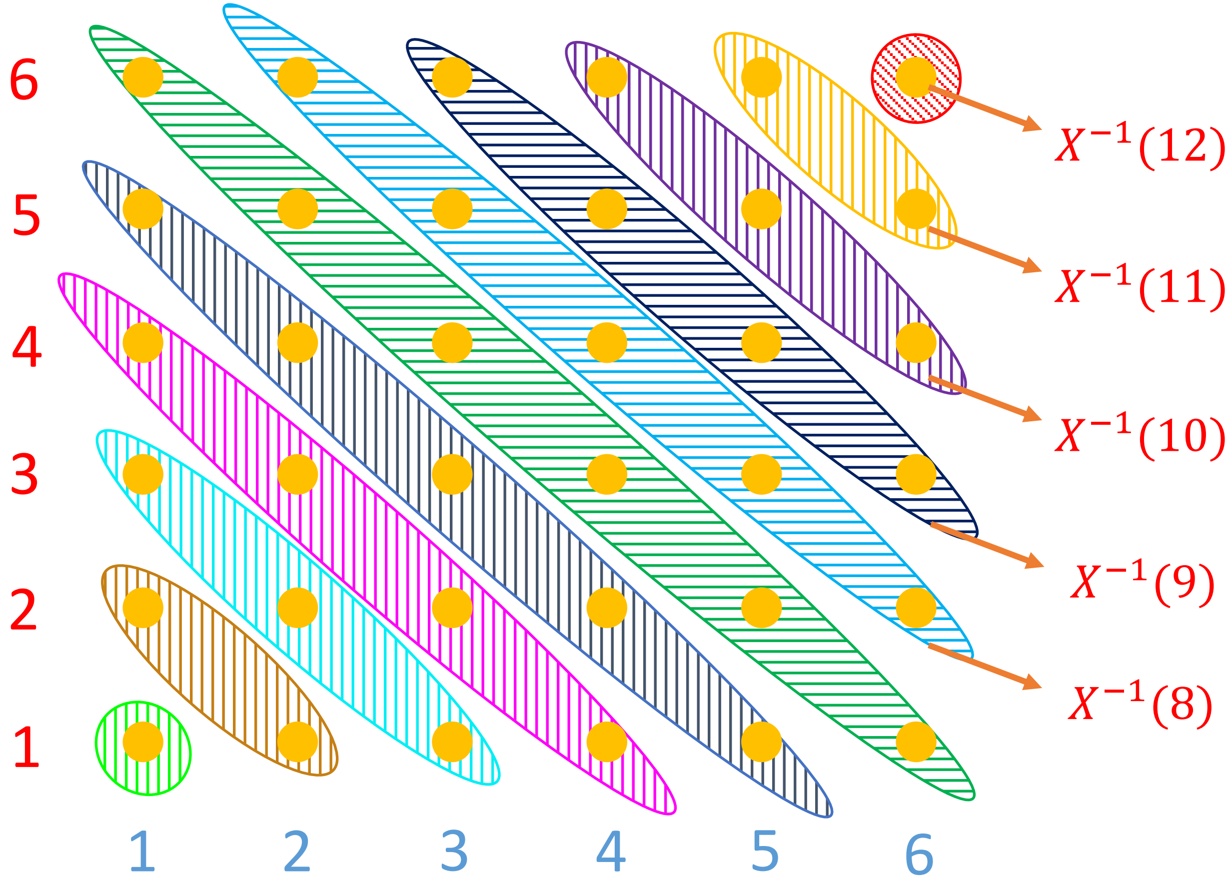
$$\mathbb{P}[X|E = a] = \frac{\mathbb{P}[X = a \wedge E]}{\mathbb{P}[E]}$$

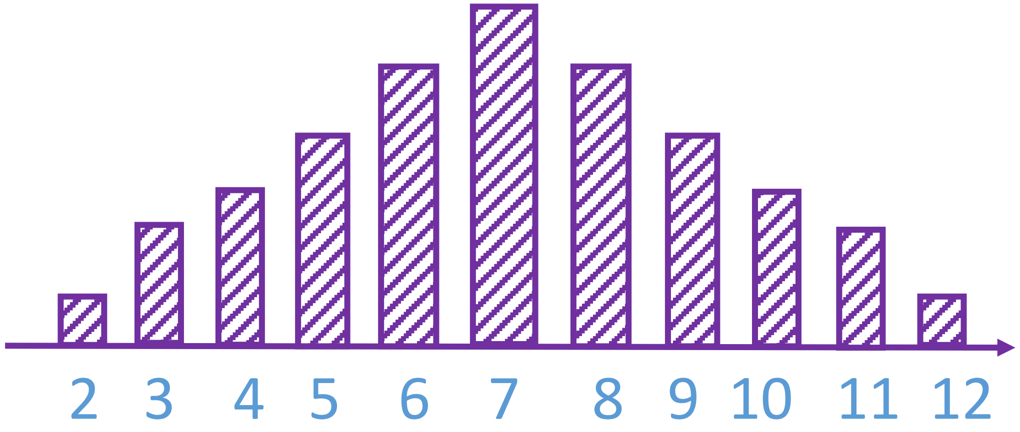
Conditional Random Variables (Examples)

Example 1 (Rolling two dice).

Let X be the sum of two dice. E be X is prime.



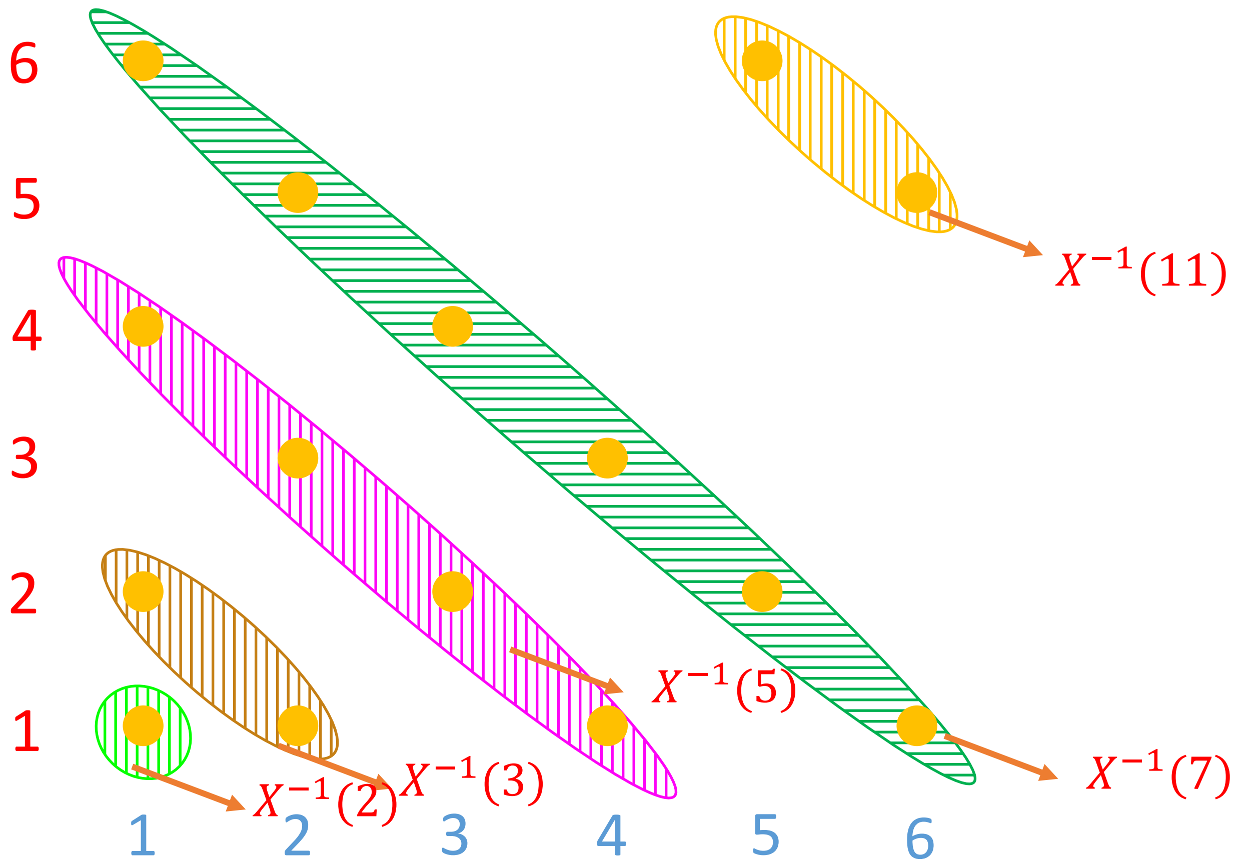


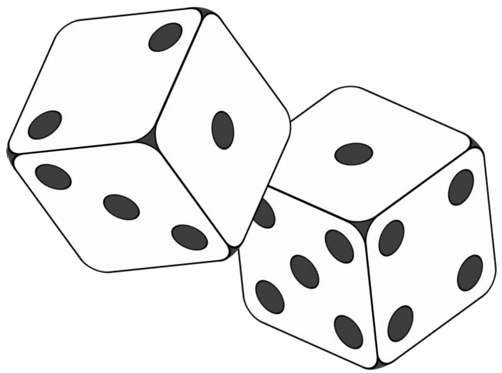


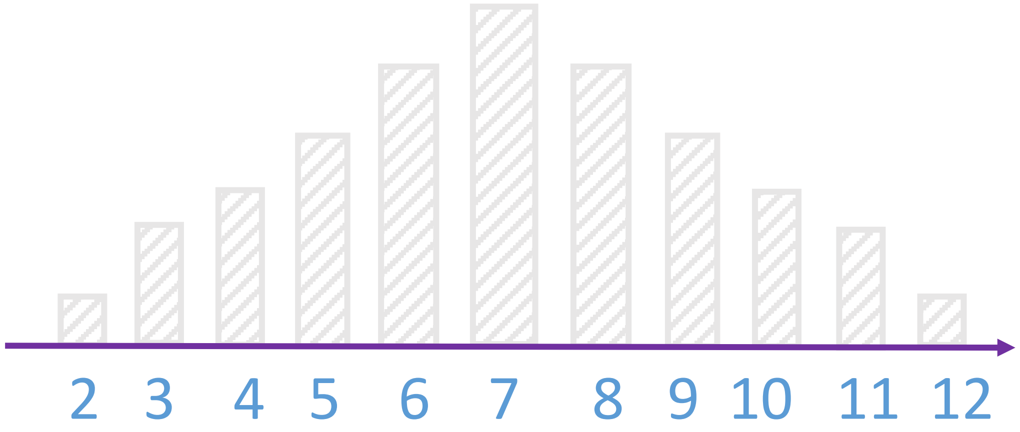
Conditional Random Variables (Examples)

Example 1 (Rolling two dice).

Let X be the sum of two dice. E be X is prime.



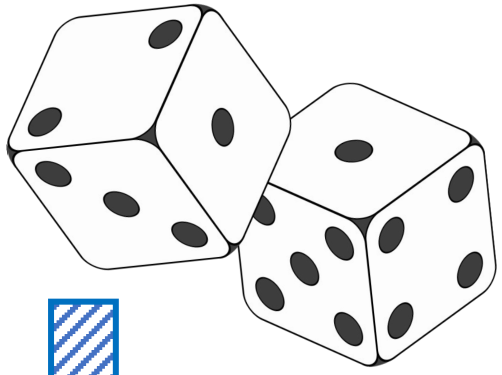


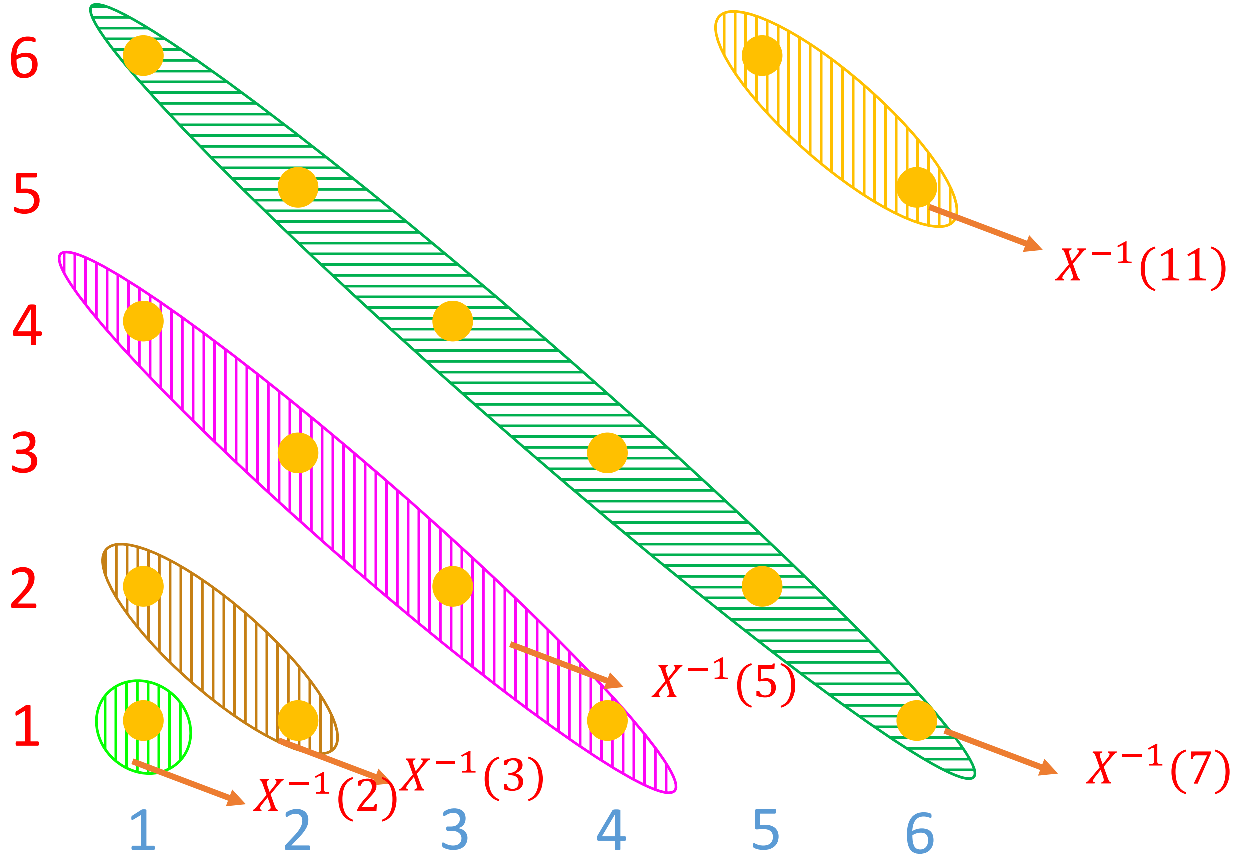


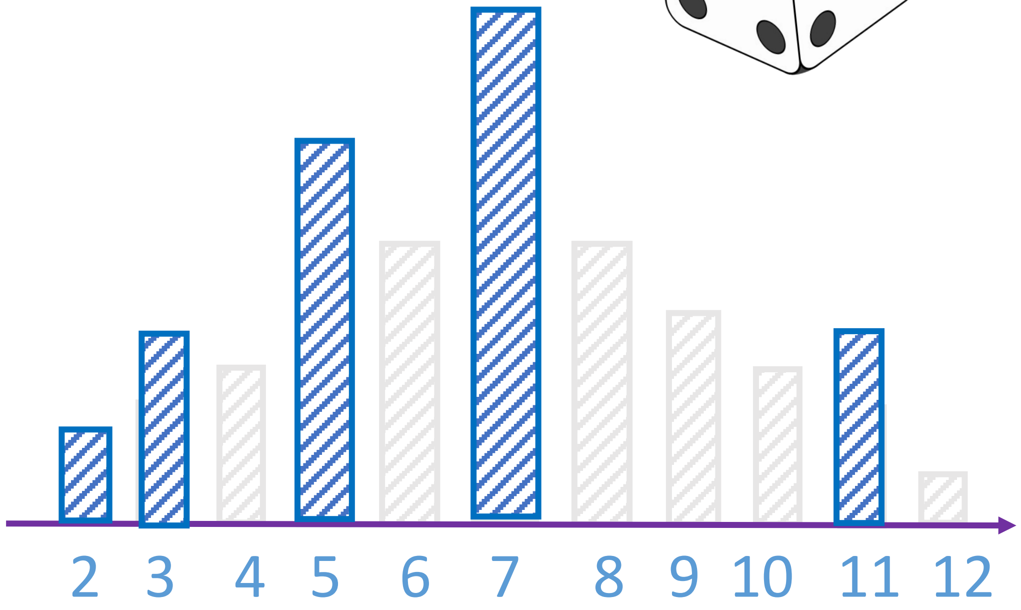
Conditional Random Variables (Examples)

Example 1 (Rolling two dice).

Let X be the sum of two dice. E be X is prime.







Conditional Random Variables (Examples)

Example 2 (Toss 100 coins).

Let X be the number of heads. E be X is odd.

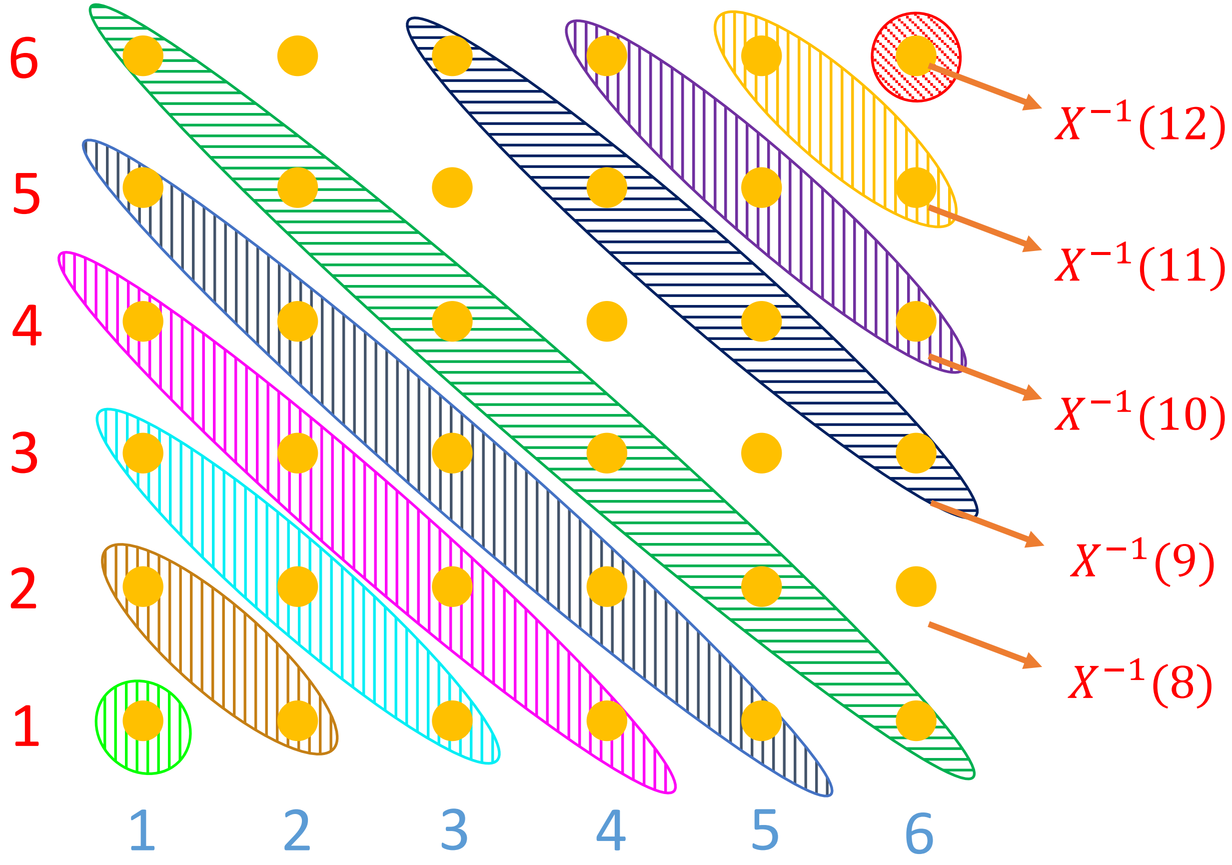
$\mathbb{P}[E] = \frac{1}{2}$ (consider tossing first 99 coins, then the last one.)

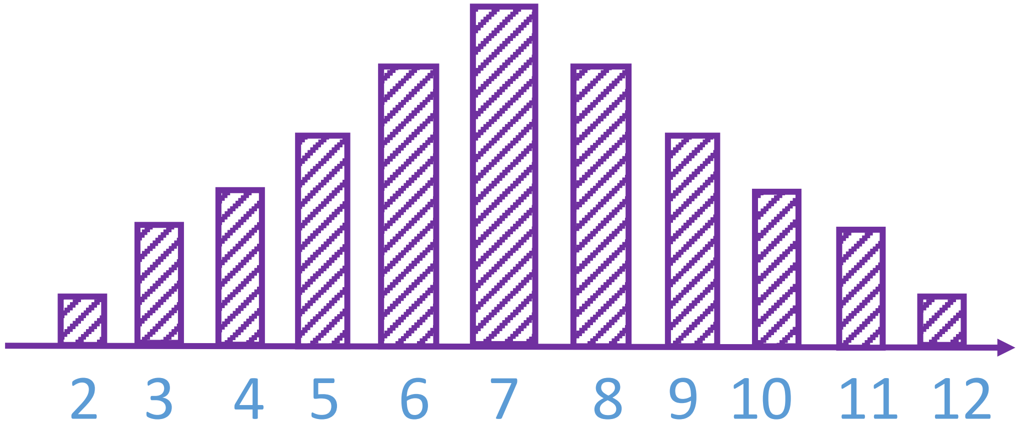
$\mathbb{P}[X = a \mid E] = 2 \cdot \frac{\binom{100}{a}}{2^{100}}$ for odd a. For even a, $\mathbb{P}[X = a \mid E]$ =0.

Prior distribution

Definition.

The prior distribution of X is its distribution before conditioning.

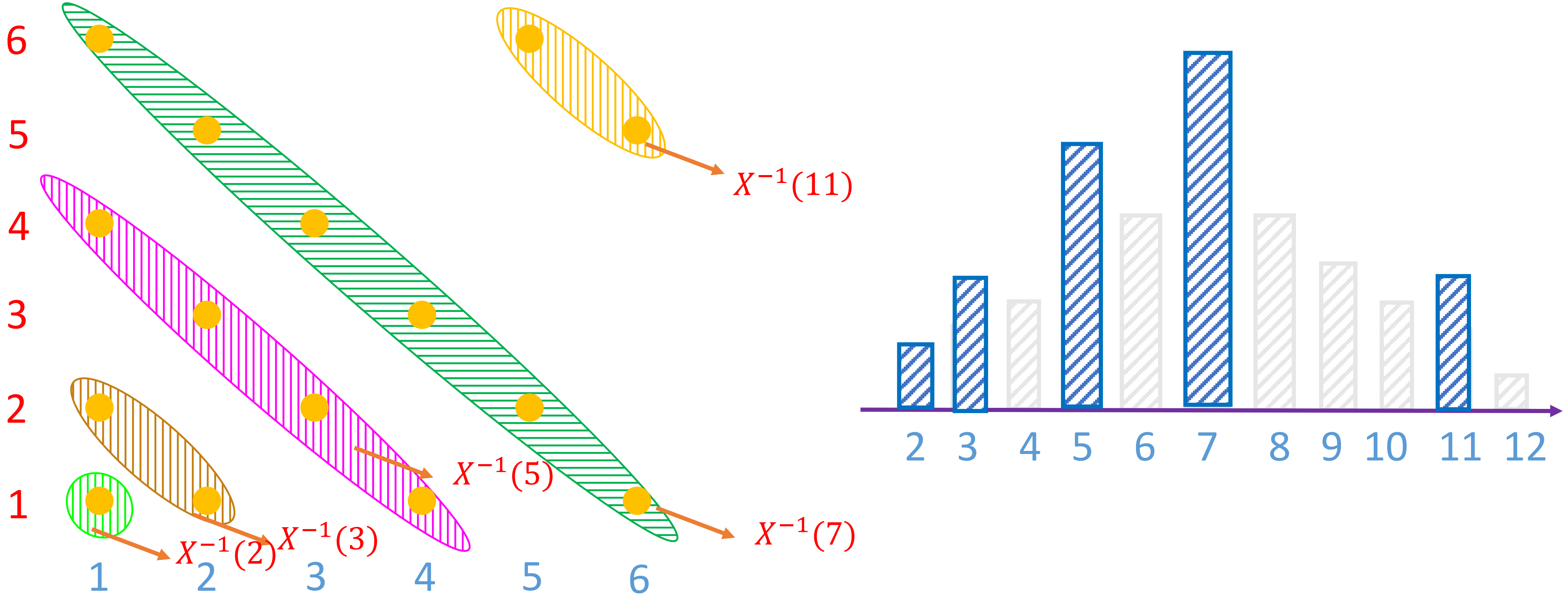




Posterior distribution

Definition.

The posterior distribution of X is its distribution after conditioning.



Example: Estimate parameter of a coin

Example.

Suppose X has probability P of being 1.

and probability $1-P$ of being 0.

We know with probability 1/3, $P = 1/2$.

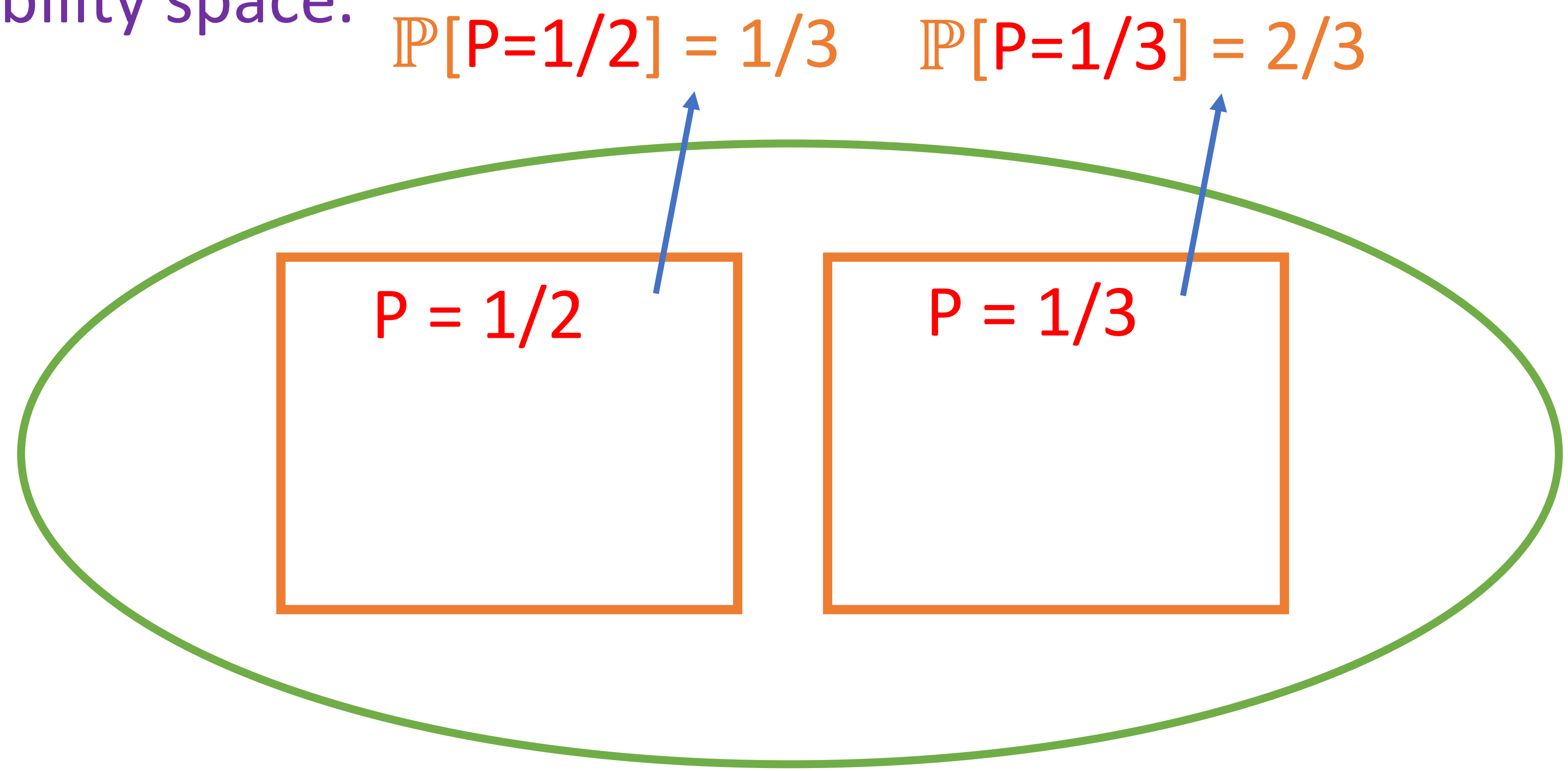
with probability 2/3, $P = 2/3$.

} Prior distribution of P

Now we observe that $X=0$. What is our belief for P?

Example: Estimate parameter of a coin

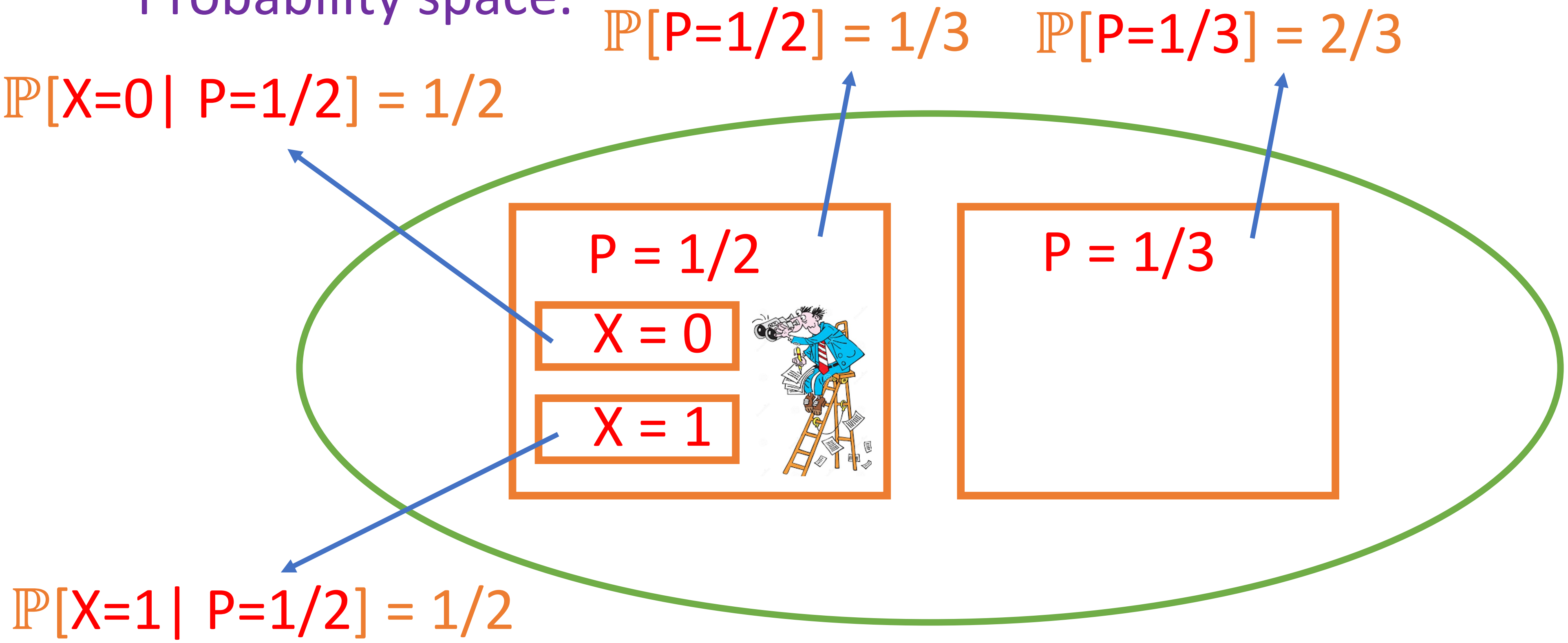
Probability space.



If you only care about P, there are two possible worlds.

Example: Estimate parameter of a coin

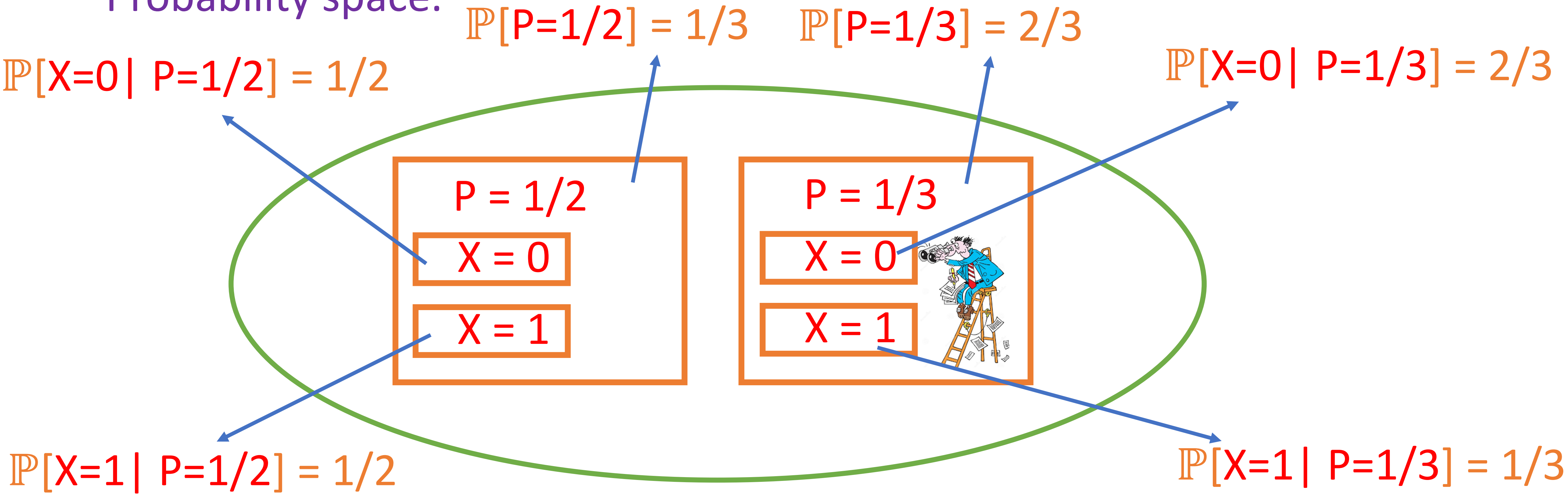
Probability space.



Within each of them, there are two possible worlds of X.

Example: Estimate parameter of a coin

Probability space.



Within each of them, there are two possible worlds of X.

Example: Estimate parameter of a coin

Probability space.

$\mathbb{P}$[X=0, P=1/2] = 1/6

$\mathbb{P}$[X=0, P=1/3] = 4/9

X = 0, P=1/2

X = 0, P=1/3

X = 1, P=1/2

X = 1, P=1/3

$\mathbb{P}$[X=1, P=1/2] = 1/6

$\mathbb{P}$[X=1, P=1/3] = 2/9

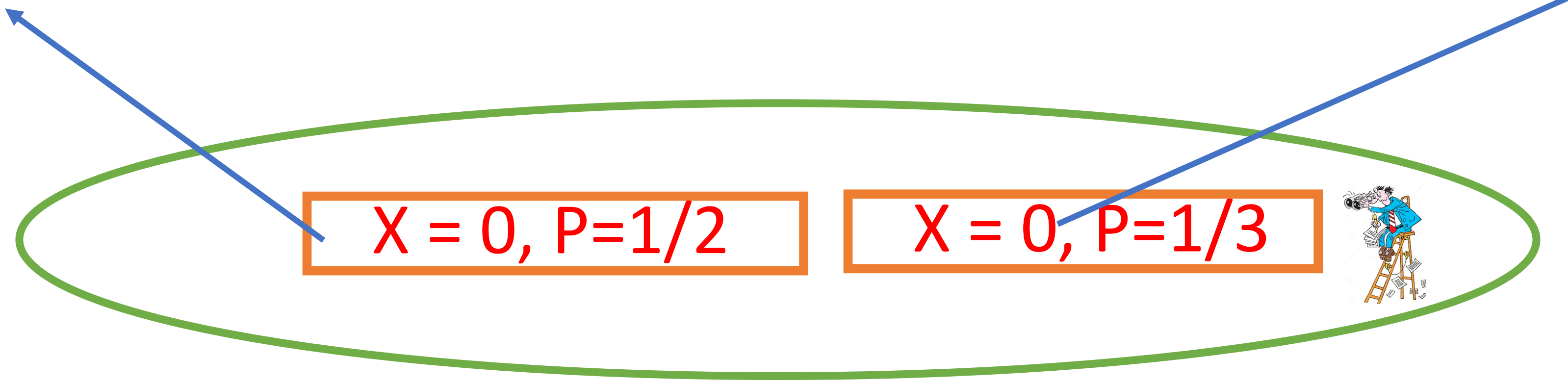
Equivalently, there are four possible worlds of X and P.

Example: Estimate parameter of a coin

Probability space.

$\mathbb{P}$[X=0, P=1/2] = 1/6

$\mathbb{P}$[X=0, P=1/3] = 4/9



Posterior distribution of P:

$$\mathbb{P}[P=1/2 \mid X = 0] = \frac{\frac{1}{6}}{\frac{1}{6}+\frac{4}{9}} = \frac{3}{11}$$

Now we know X = 0. the space shrinks after conditioning.

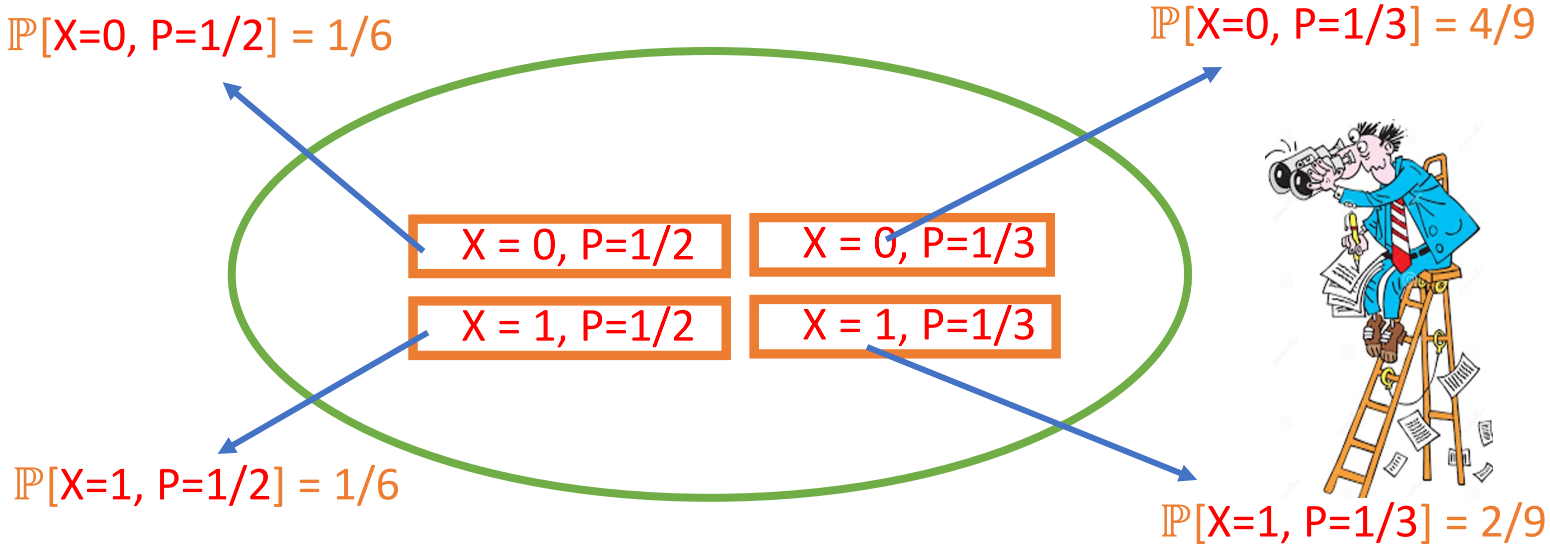
Example: Estimate parameter of a coin

Probability space.

When we analyze this process and list the probabilities, X is a random variable. It also has a prior distribution.

$\mathbb{P}[X=0] = 1/6+4/9 = 11/18$

$\mathbb{P}[X=1] = 1/6+2/9 = 7 / 18$



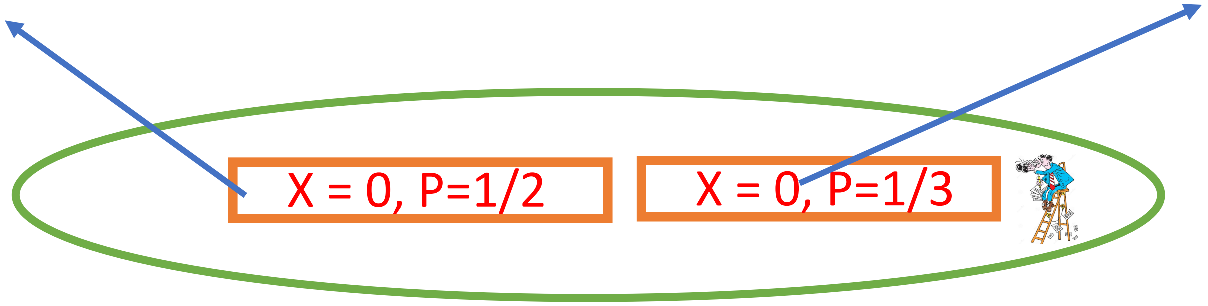
Example: Estimate parameter of a coin

Probability space.

When we observed the value of X, it is a fixed value (realization). P has what we call posterior distribution.

$\mathbb{P}[X=0, P=1/2] = 1/6$

$\mathbb{P}[X=0, P=1/3] = 4/9$



$$\mathbb{P}[P=1/2 \mid X=0] = 1/6 \,/\, (1/6+4/9)$$

$$\mathbb{P}[P=1/3 \mid X=0] = 4/9 \,/\, (1/6+4/9)$$

Debate: People vs. Collins

Trial [edit]

After a mathematics instructor testified about the multiplication rule for probability, though ignoring conditional probability, the prosecutor invited the jury to consider the probability that the accused (who fit a witness's description of a black male with a beard and mustache and a Caucasian female with a blond ponytail, fleeing in a yellow car) were not the robbers, suggesting that they estimated the probabilities as:

Black man with beard	1 in 10
Man with mustache	1 in 4
White woman with pony tail	1 in 10
White woman with blond hair	1 in 3
Yellow motor car	1 in 10
Interracial couple in car	1 in 1,000

The jury returned a guilty verdict.[1]

What witness said

What prosecutor did

"Prosecutor's fallacy"--- we might talk about it In 7/24 lecture