Probabilistic Robotics Course

Introduction to Probabilities

Giorgio Grisetti
grisetti@diag.uniroma1.it

Department of Computer Control and Management Engineering
Sapienza University of Rome

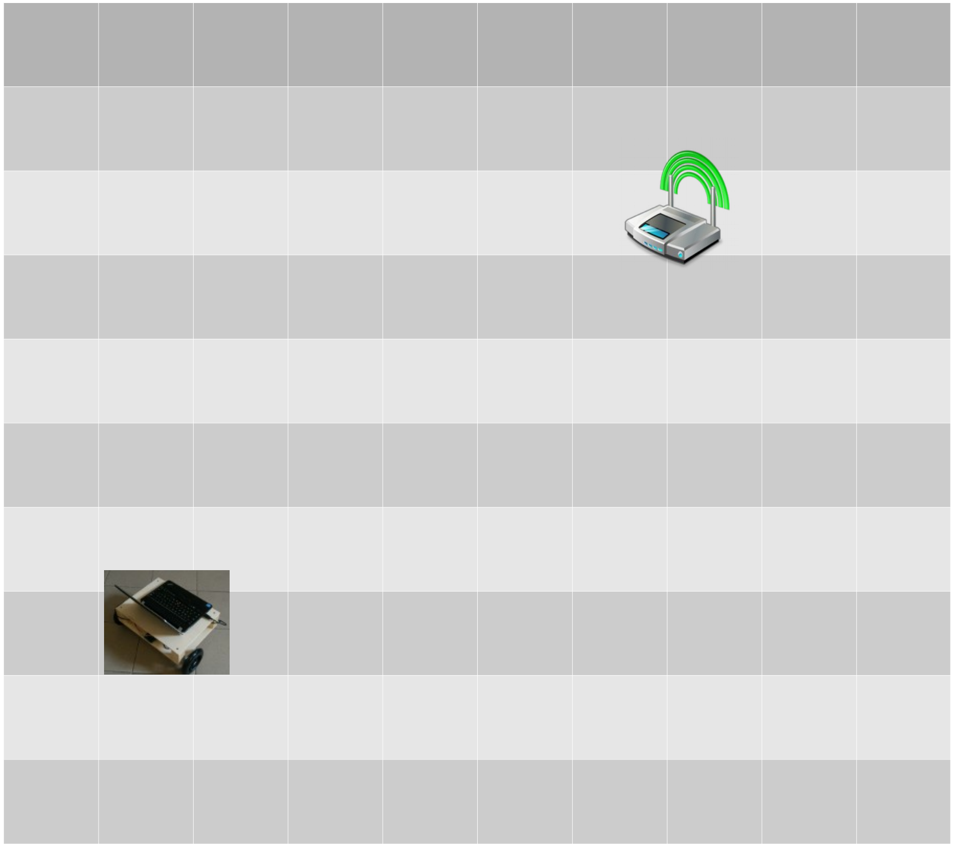
1. $P(E) \geq 0$
2. $P(\bigcup_i E_i) = \sum_i P(E_i)$
3. $P(\Omega) = 1$

Outline

- Random Phenomena
- Events
- Axioms
- Conditional Probability
- Chain Rule
- Marginalization
- Functions of Random Variables

Scenario

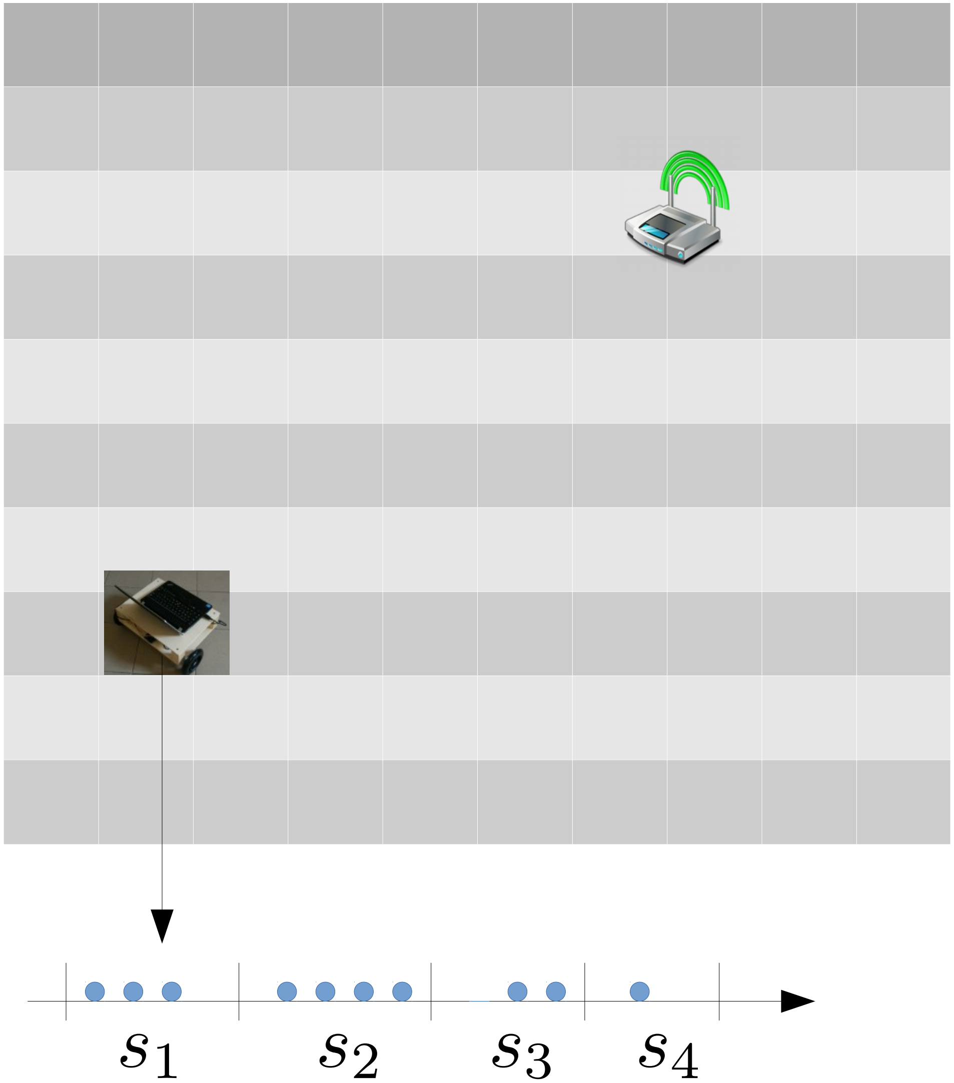
- Orazio is located in a building
- In the building there is an access point
- We want to know the WiFi signal strength that orazio will sense at different locations



Scenario

- We put Orazio in a location, and start recording samples of measured signal strengths
- This will give us a statistic about the signal strength at a location
- We define a set of intervals on the strengths
- To each interval we assign a number

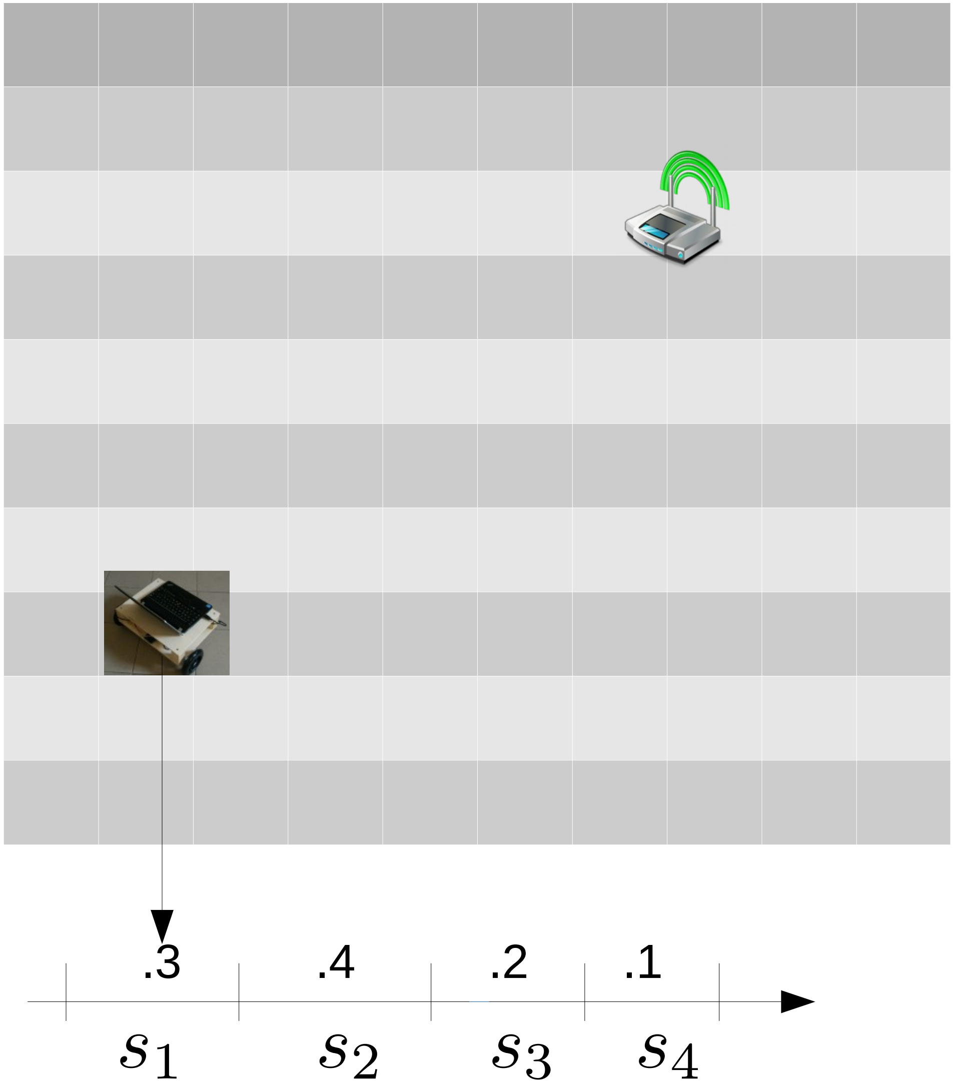
$$P_{l_j}(S = s_i) = \frac{1}{N} N_i$$



Scenario

- We put Orazio in a location, and start recording samples of measured signal strengths
- This will give us a statistic about the signal strength at a location
- We define a set of intervals on the strengths
- To each interval we assign a number

$$P_{l_j}(S = s_i) = \frac{1}{N} N_i$$

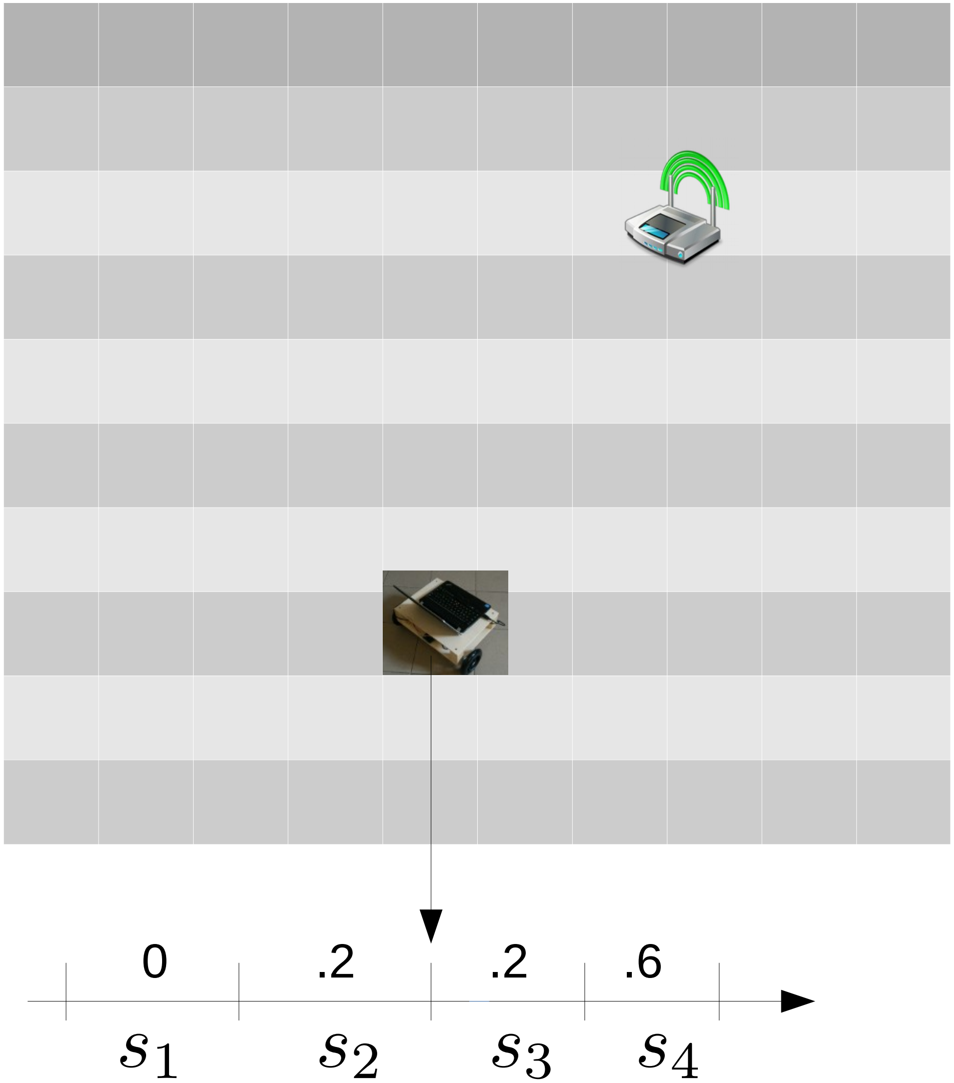


Scenario

- We put Orazio in a location, and start recording samples of measured signal strengths
- This will give us a statistic about the signal strength at a location
- We define a set of intervals on the strengths
- To each interval we assign a number

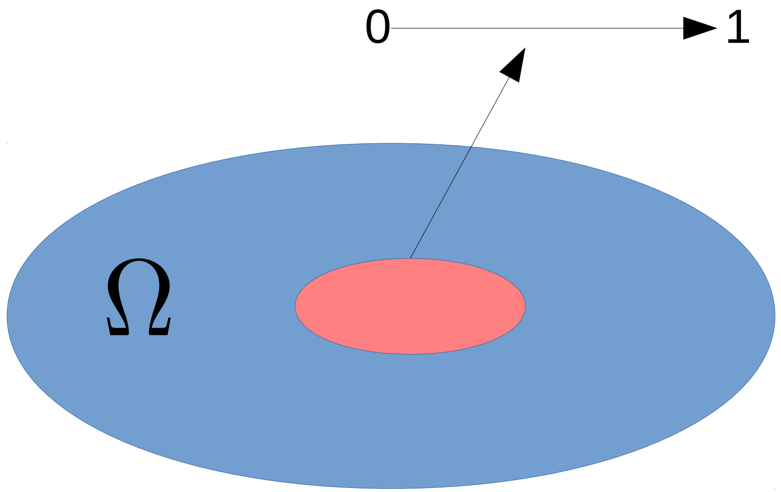
$$P_{l_j}(S = s_i) = \frac{1}{N} N_i$$

- We repeat the process for each location



Events and Probability

- The statement "the signal strength S falls in the interval s_i " is an **event**.
- We will compactly write S_i to denote the event $S = s_i$
- $P_{l_j}(S_i)$ is the a number denoting the probability of the event of observing a signal strenght, given a location l_i
- Formally a probability is a function going from each measurable subset of the event space Ω to the interval [0,1]



Axioms

Not all functions are valid probability assignments, only those that fulfill the following axioms:

$$1.\ \ P(E) \geq 0$$

1. The probability of an event is greater or equal than 0

$$2.\ \ P(\bigcup_i E_i) = \sum_i P(E_i)$$

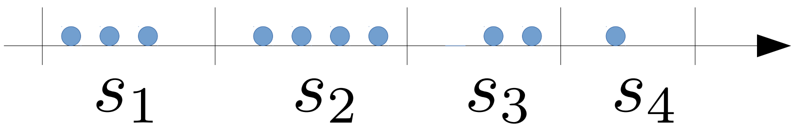
2. The probability of the union of a set of disjoint events, is the sum of probabilities of the events

$$3.\ \ P(\Omega) = 1$$

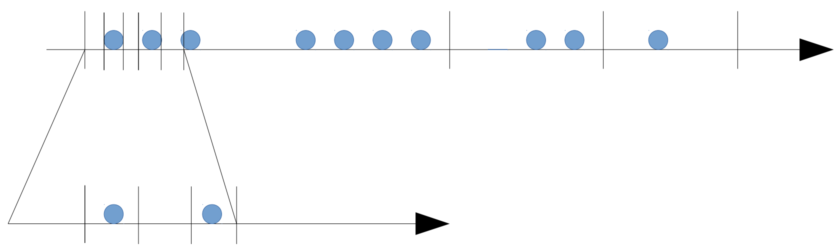
3. The probability of the outcome being in the set of possible outcomes is 1

Continuous Domains

In our previous example, the signal strengths are continuous



Recovering the probability through a statistic becomes hard, since we would have to use infinitesimally small intervals



Only the bins that have at least one sample will get a non 0 probability

Continuous Domains

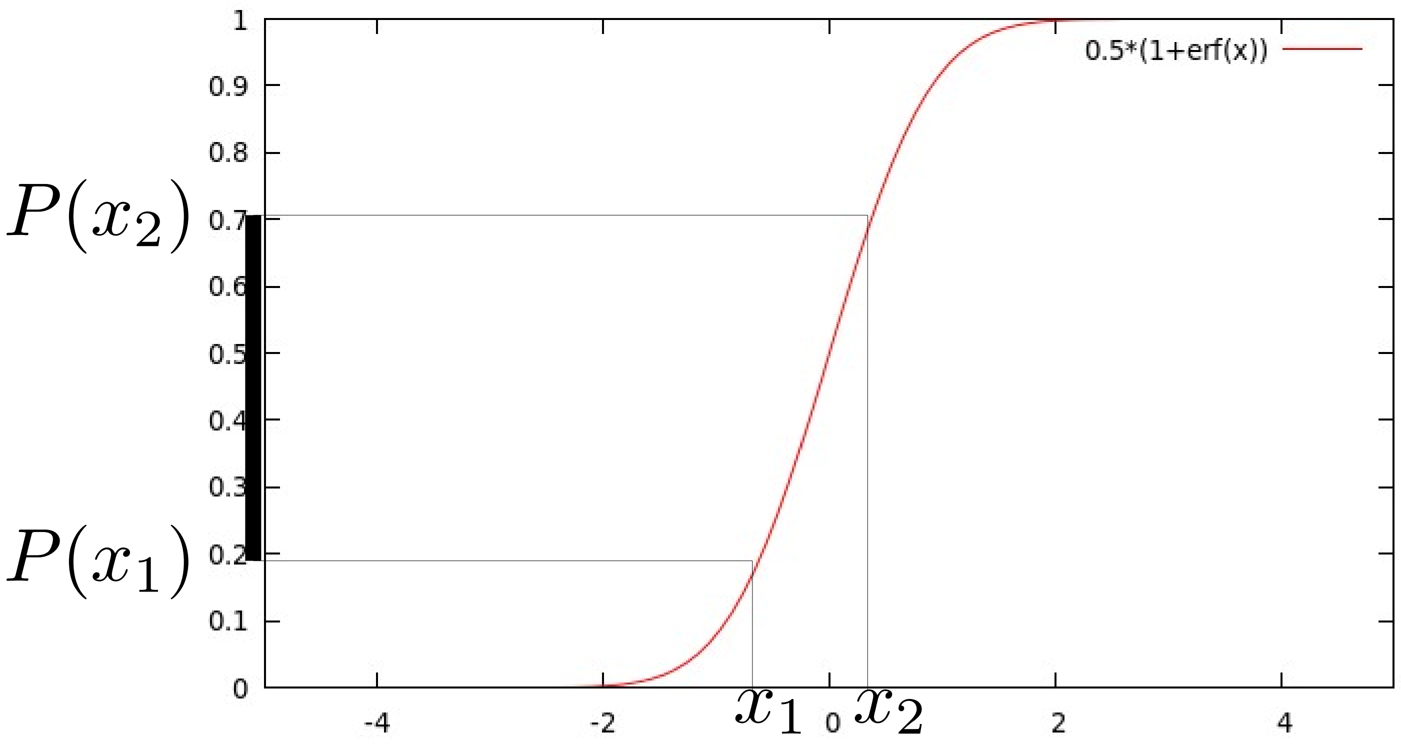
In a continuous domain, the probability of the outcome assuming exactly a specific value x is, in general 0

Yet we can define the event $P(X < x)$ that is true when the outcome of the event is lower than a certain value

We will denote this function for continuous domains as $P(x)$ and call it cumulative density function

The cumulative density is monotonically increasing

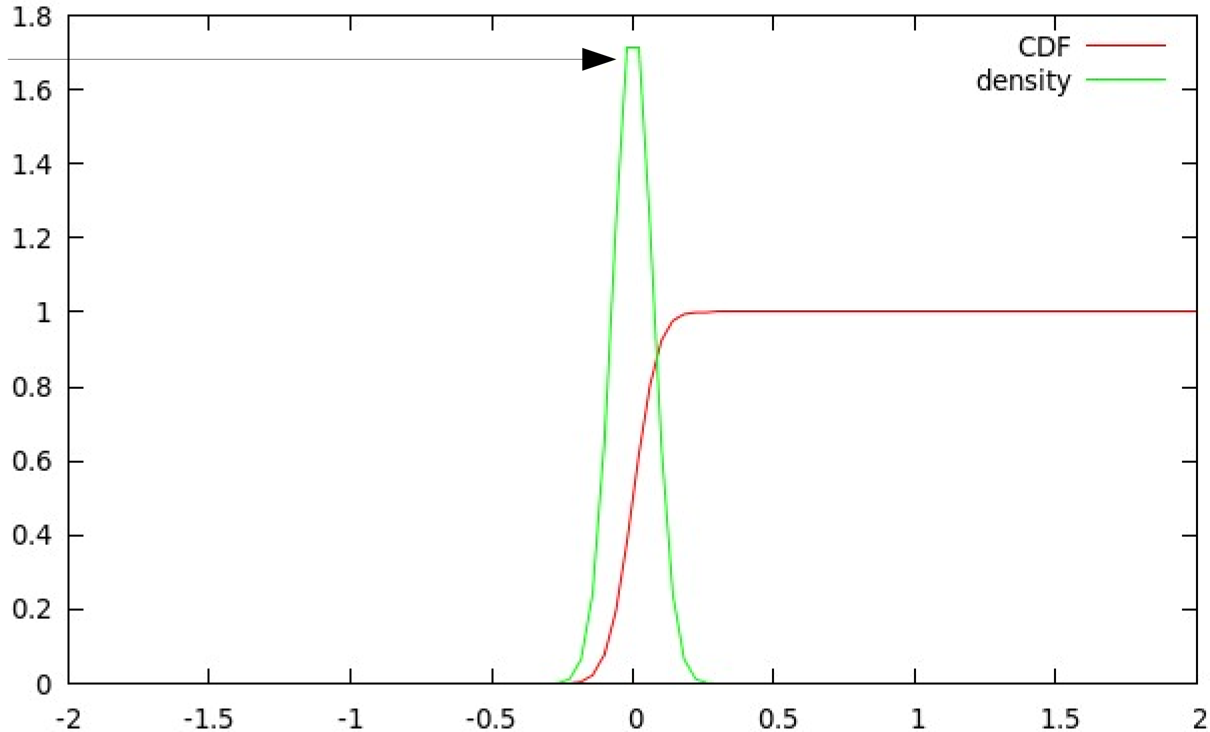
Continuous Domains



- The probability that an element falls in an interval $[x_1, x_2]$ is $P(x_2) - P(x_1)$
- As $x_2 \to x_1$ this value tends to 0

Probability Density

density can be > 1



For continuous distributions it makes sense to consider the probability density $p(x)$ defined as

$$p(x) = \frac{\partial P(x)}{\partial x}$$

Probability Density

Computing the probability for a generic event $E \subseteq \Omega$

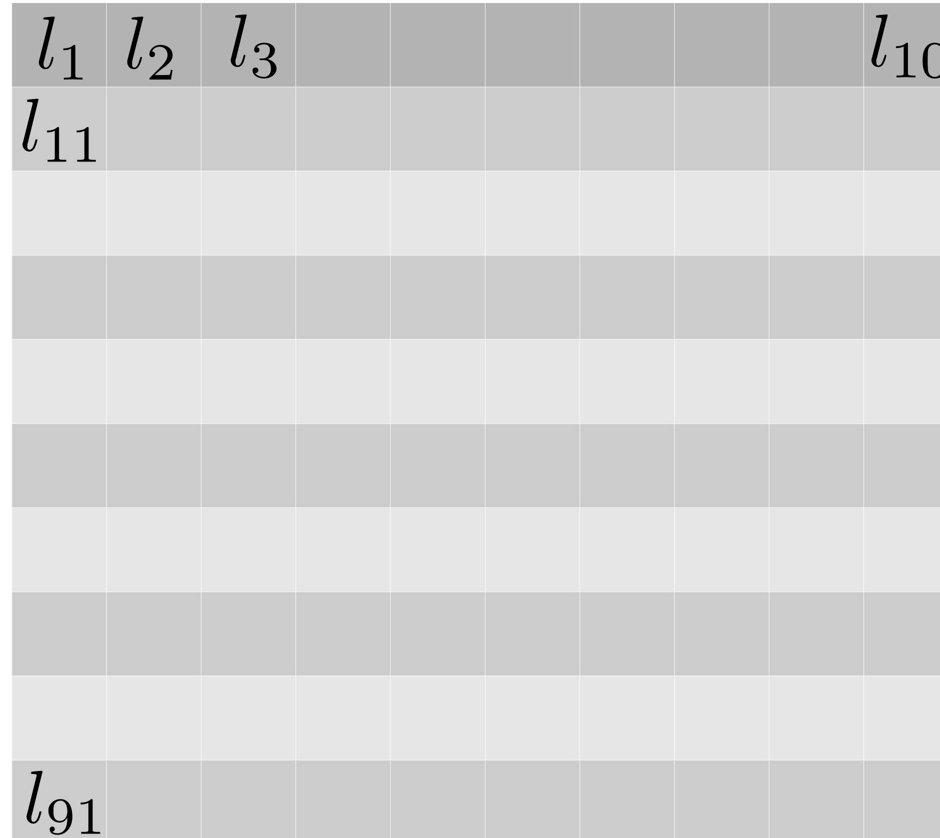
is done by integration:

$$P(E) = \int_{x \in E} p(x) dx$$

This corresponds to summing up all infinitesimal disjoint events that cover E.

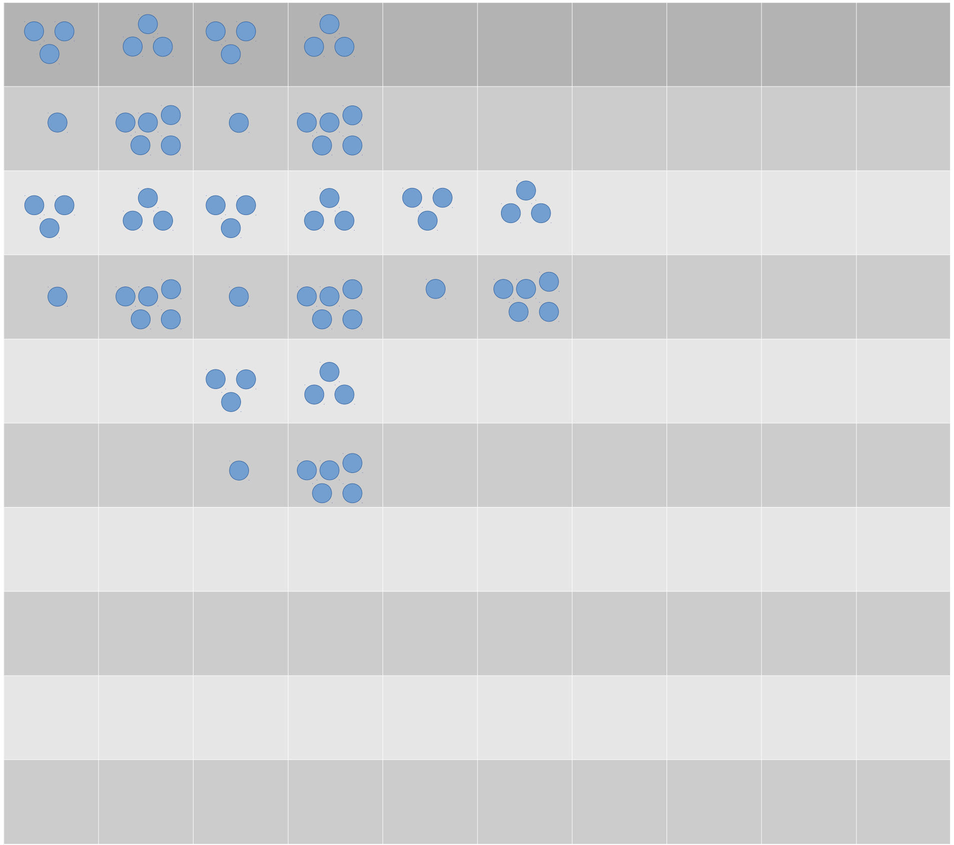
Back to Orazio

- Similar to what we have done for the signal strength, we can acquire statistics about the locations occupied by orazio during its missions
- We count how often orazio visits each location



Back to Orazio

- Similar to what we have done for the signal strength, we can acquire statistics about the locations occupied by orazio during its missions
- We count how often orazio visits each location



Location Statistics

.03	.03	.03							
.01	.05	.01							
			...						

- The statistics give us $P(L_i)$
- When we were acquiring stats about location, we ignored the signal strength

Conditional Probability

- During our experiment on the signal strength, we acquired independent statistics at each location.
- We have a separate distribution over signal strength, given the location.
- We write this as

$$P_{l_j}(S = s_i) := P(S_i|L_j)$$

- This symbol denotes a distribution over the signal strengths, **GIVEN** the knowledge of the location, and serves us to pick up the correct instance of an experiment.

Chain Rule

How likely is it that orazio sits at location L_2 and from there it measures the signal strength S_3?

$$P(S_3, L_2) = ?$$

Certainly it is less likely than being at location L_2 and measuring **any** signal strength:

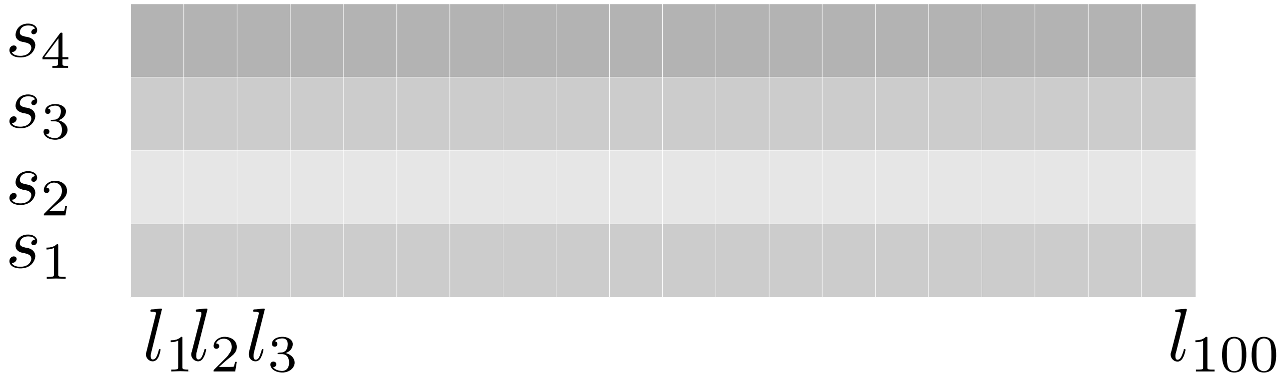
$$P(S_3, L_2) \leq P(L_2)$$

It is as much likely as being in L_2 **and**, from L_2 measuring a signal strength:

$$P(S_3, L_2) = P(S_3 \mid L_2) \cdot P(L_2)$$

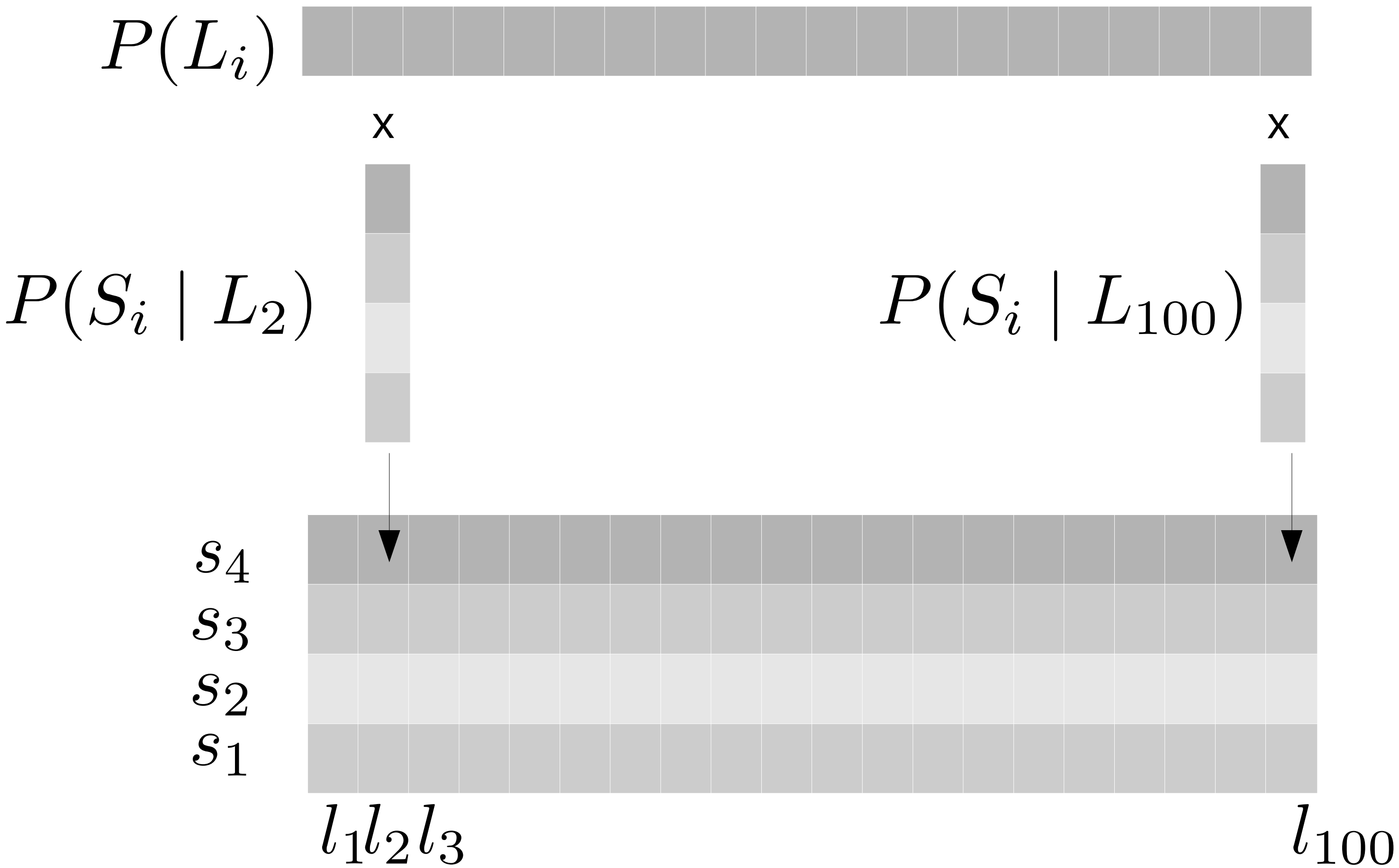
Joint Distributions

- $P(S_3, L_2)$ is a joint distribution, whose domain is all possible pairs of signal strengths and locations
- We can imagine it as a grid, whose columns correspond to locations, and whose rows correspond to signal strengths



Joint from Conditional

Multiply each element of the conditional by the probability of the conditioner:

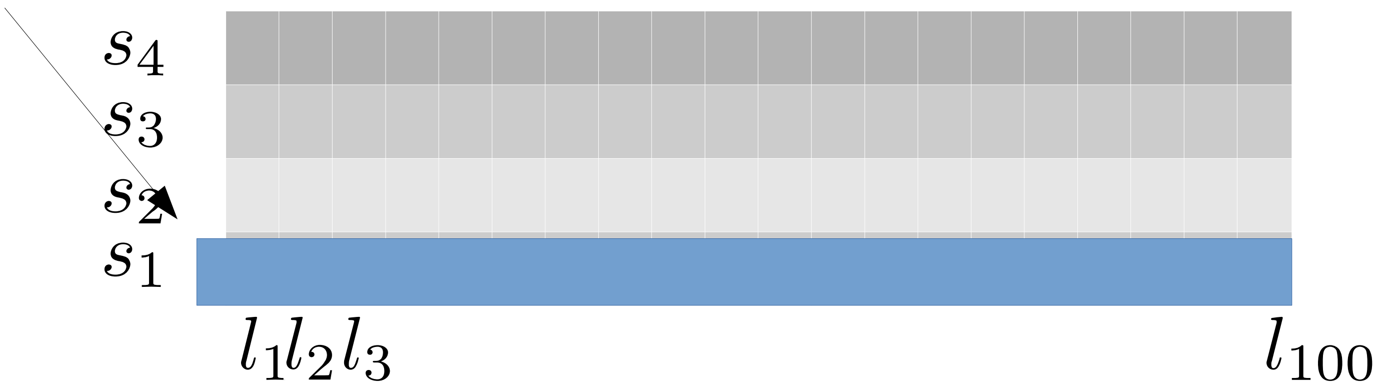


Marginals

We want to compute the probability of sensing a very weak signal strength s_1:

$$P(S_1)$$

This is the probability of measuring s_1 from any location and corresponds to the following row in the joint table:

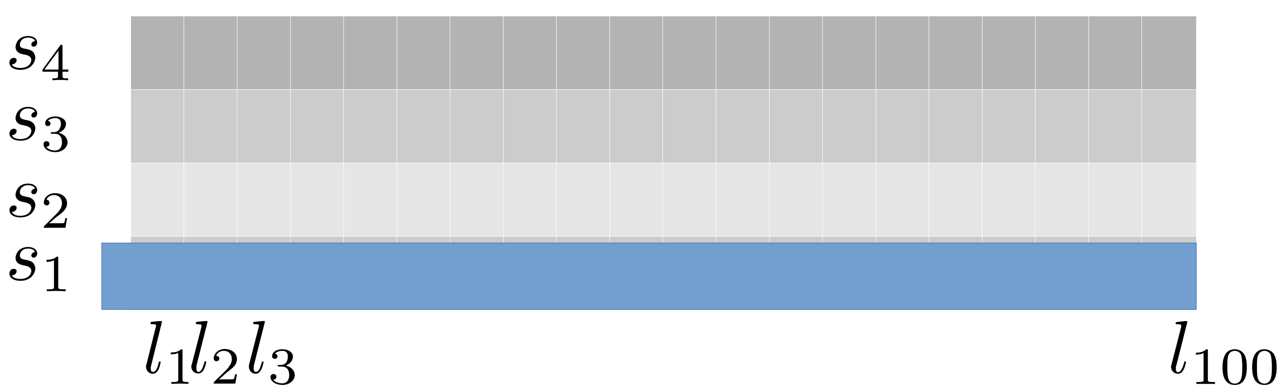


Marginals

- The events $< S_i, L_j >$ are disjoint :)

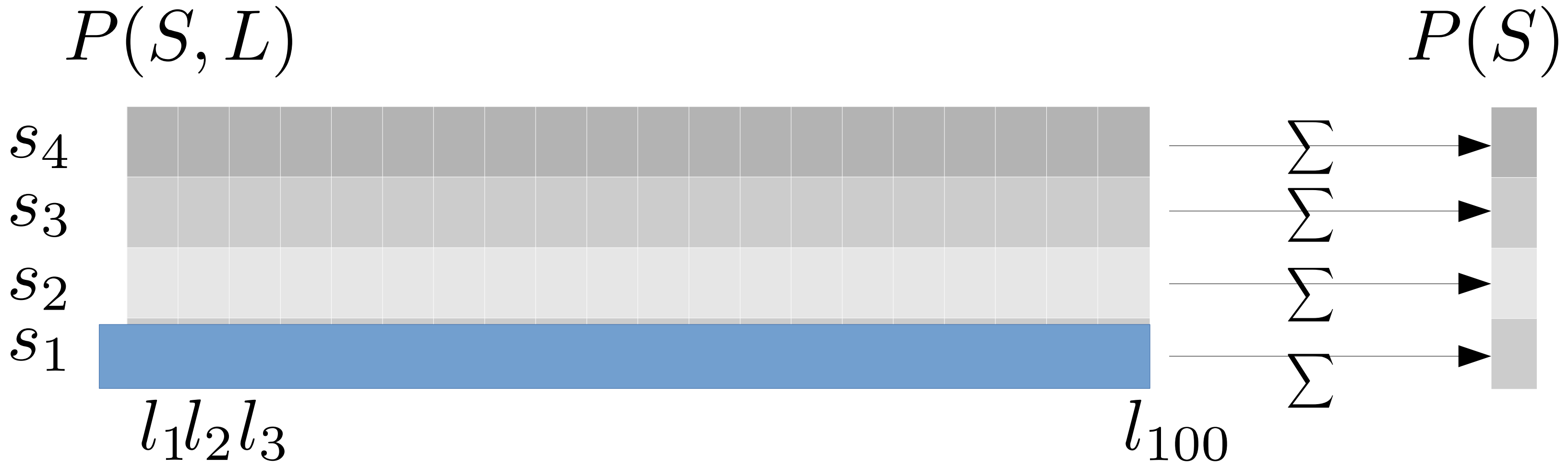
- Use the second axiom

$$P(S_1) = P\left(\bigcup_j < S_1, L_j >\right) = \sum_j P(S_1, L_j)$$



Marginals

- This is called **marginalization**
- This process consists of suppressing a variable from a table.
- Visually, it works as follows:



Independence

In reality not all phenomena are correlated we can express this as:

$$P(A \mid B) = P(A)$$

This means: knowing B tells nothing about A.

In the Orazio world, an independent phenomenon might be the floor at which the elevator is located and signal strength.

Joint of Independent

The chain rule applies also to independent variables

$$P(A, B) = P(A|B)P(B) = P(A)P(B)$$

The joint distribution of independent variables is the product of their distributions.

Independence test

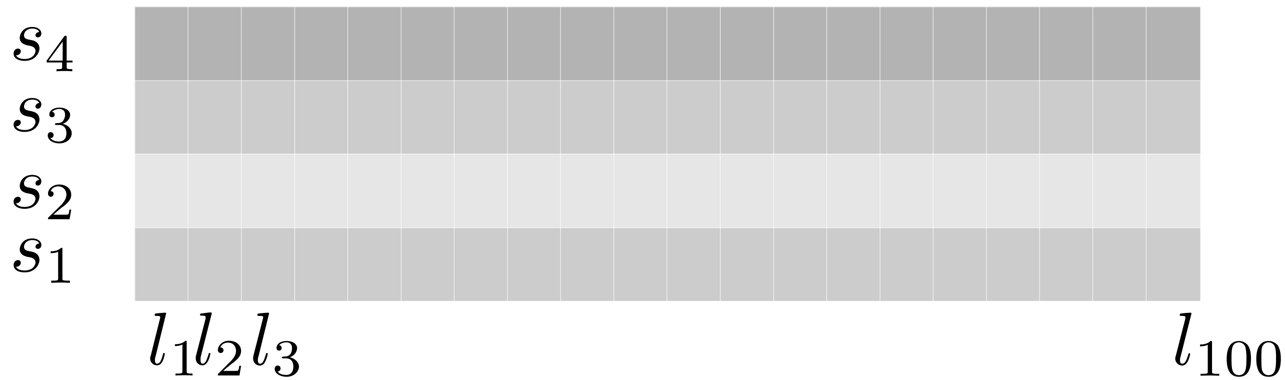
To figure out if two variables are independent, given their table we

- compute the marginal over the first variable
- compute the marginal over the second variable
- compute the joint distribution from the marginals as if the variables were independent

If this resulting table is the same as the original one, the variables are independent, otherwise they are not

Localizing Orazio

We are given the joint table over locations and strengths



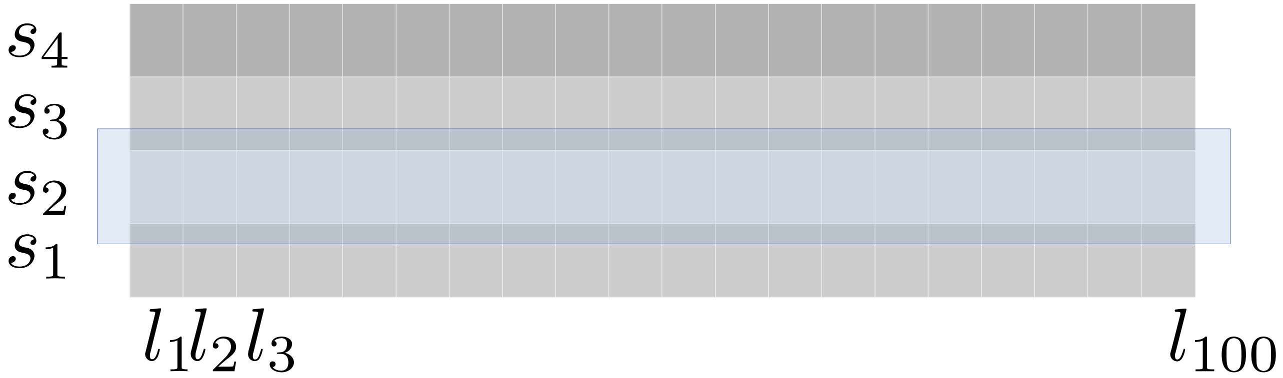
We measure signal strength s_2

We want to figure out where Orazio could be, given the measured strength

$$P(L|S_2)$$

Localizing Orazio

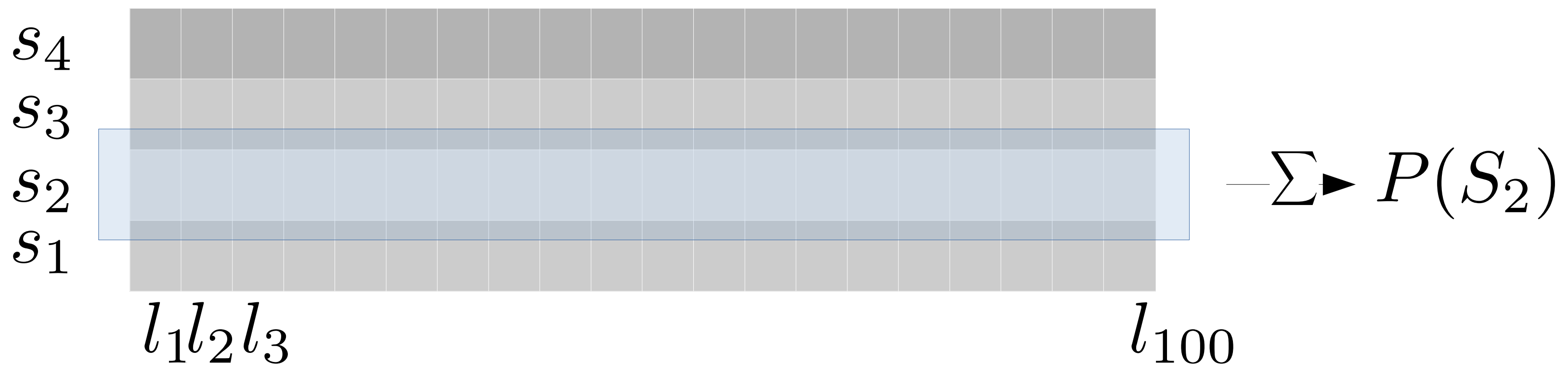
- We measure the signal s_2 , thus all cases where our signal is different are not relevant for our inference



- This row of the table, however is not a valid distribution (it does not sum up to one)
- Yet it comprises all the knowledge we can cosider, given the evidence

Localizing Orazio

The probability mass in the row corresponds to measuring the evidence

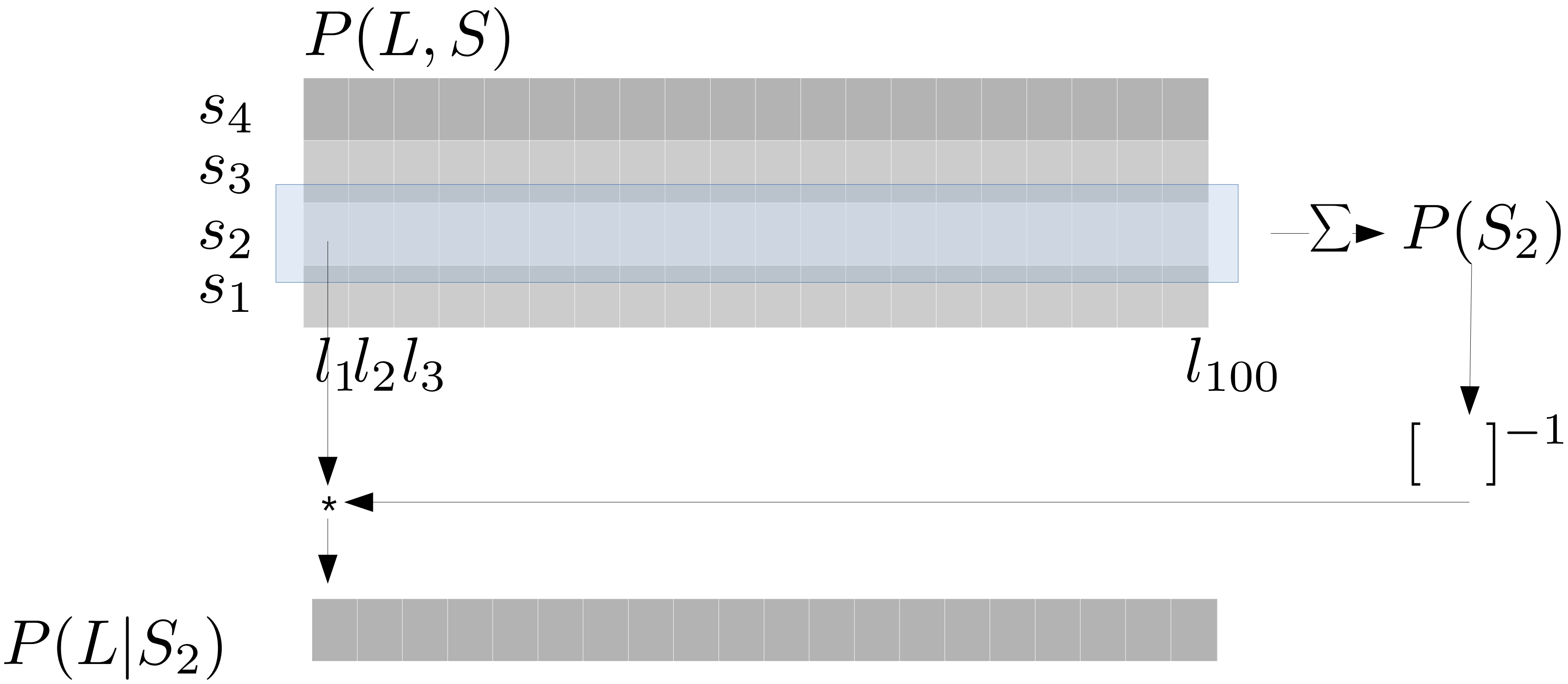


Dividing each element of the row by the (constant) value of $P(S_2)$ renders the row a valid distribution over the possible locations

$$P(L_j|S_2) = \frac{P(S_2, L_j)}{P(S_2)}$$

Localizing Orazio

Visually



$$P(L_j|S_2) = \frac{P(S_2, L_j)}{P(S_2)}$$

Functions

X is a random variable distributed according to $P(X)$

Let $y = f(x)$ be a generic function defined on the possible values of X

Y is a random variable, whose distribution is the following

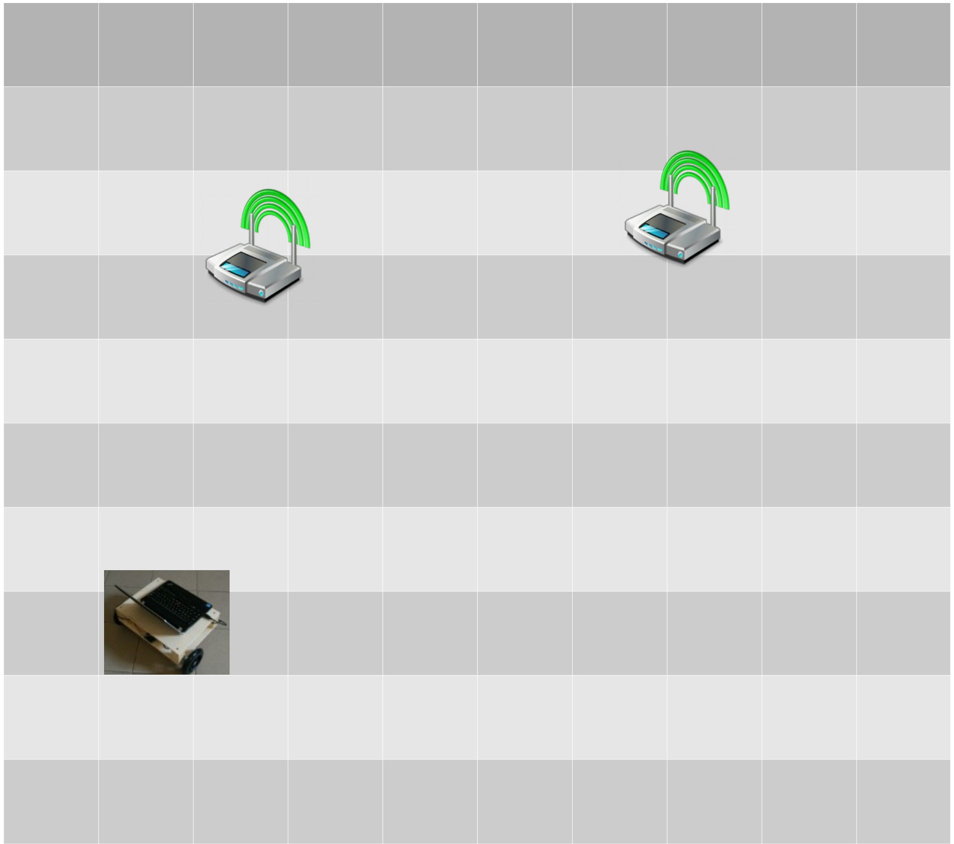
$$P(Y) = \sum_{X_i = f^{-1}(Y)} P(X_i)$$

Conditional Independence

We buy another access point

We now have 2 networks in the environment and we can compute the statistics for the strength of first and second signal

Let us assume that the only element influencing the signal strength is the location (the signals do not interfere)



$$P(S^A \mid L)$$

$$P(S^B \mid L)$$

Conditional Independence

What about $P(S^A, S^B)$

Are the two variables independent?

I should answer the question

“if I know S^A can I tell anything about S^B ?“

We know that the strength depends on the location. This means that the location and the strength are correlated.

If the strength of the 1st signal influences the location, and the 2nd signal is influenced by the location, then the signals are correlated through the location

Conditional Independence

But what if I know the location?

$$P(S^A, S^B|L) = P(S^A|L)P(S^B|L)$$

The two signals are independent if I know the location.

Exploiting the structure of the problem, we figure out if two variables are correlated or not

Probability Identities

A probability equation obtained

- by applying the axioms and
- exploiting conditional independence from domain

holds regardless the "shape" of the densities

Consequence: you can add the same conditioning term to *all* terms of the equation and it still holds.

$$P(A,B) = P(A|B)P(B) \Rightarrow P(A,B|C) = P(A|B,C)P(B|C)$$

$$P(A|B) = \frac{P(A,B)}{P(B)} \Rightarrow P(A|B,C) = \frac{P(A,B|C)}{P(B|C)}$$

$$P(A) = \sum_i P(A,B_i) \Rightarrow P(A|C) = \sum_i P(A,B|C)$$

Summary

- Axioms

1. $P(E) \geq 0$

2. $P(\bigcup_i E_i) = \sum_i P(E_i)$

3. $P(\Omega) = 1$

- Chain Rule

$$P(A, B) = P(A|B)P(B)$$

- Marginalization

$$P(A) = \sum_i P(A, B_i)$$

- Conditioning

$$P(A|B) = \frac{P(A, B)}{P(B)}$$