Lecture 1: probability concepts I.

Bayesian probabilities in your non-academic life:

Example: The Monty Hall or Let's Make a Deal Problem

- Three doors
- Car (prize) behind one door
- You pick a door, but don't open it yet
- Monty then opens one of the other doors, <u>always revealing no car</u> (he knows where it is)
- You now get to switch doors if you want
- Should you?
- Most people reason: Two remaining doors were equiprobable before, and nothing has changed. So doesn't matter whether you switch or not.

Bayes' theorem

Let's work a couple of examples using Bayes Law:

Example: Trolls Under the Bridge

Trolls are bad. Gnomes are benign.
Every bridge has 5 creatures under it:

20% have TTGGG (H_1)
20% have TGGGG (H_2)
60% have GGGGG (benign) (H_3)

Before crossing a bridge, a knight captures one of the 5 creatures at random. It is a troll. "I now have an 80% chance of crossing safely," he reasons, "since only the case

20% had TTGGG (H1) → now have TGGG

is still a threat."

machine learning example

A CASE STUDY: CHANGE-POINT DETECTION

task is to detect partitions in a sequence of observations, in order for the data in each block to be statistically "similar," in other words, to be distributed according to a common probability distribution.

Figure 1 shows the number of deadly accidents per year in the coal mines in England spanning the years 1851-1962. Looking at the graph, it is readily observed that the "front" part of the graph looks different from its "back" end, with a change around 1890-1900. As a matter of fact, in 1890, new health and safety regulations were introduced, following pressure from the coal miners' unions. We will use the poisson distribution.

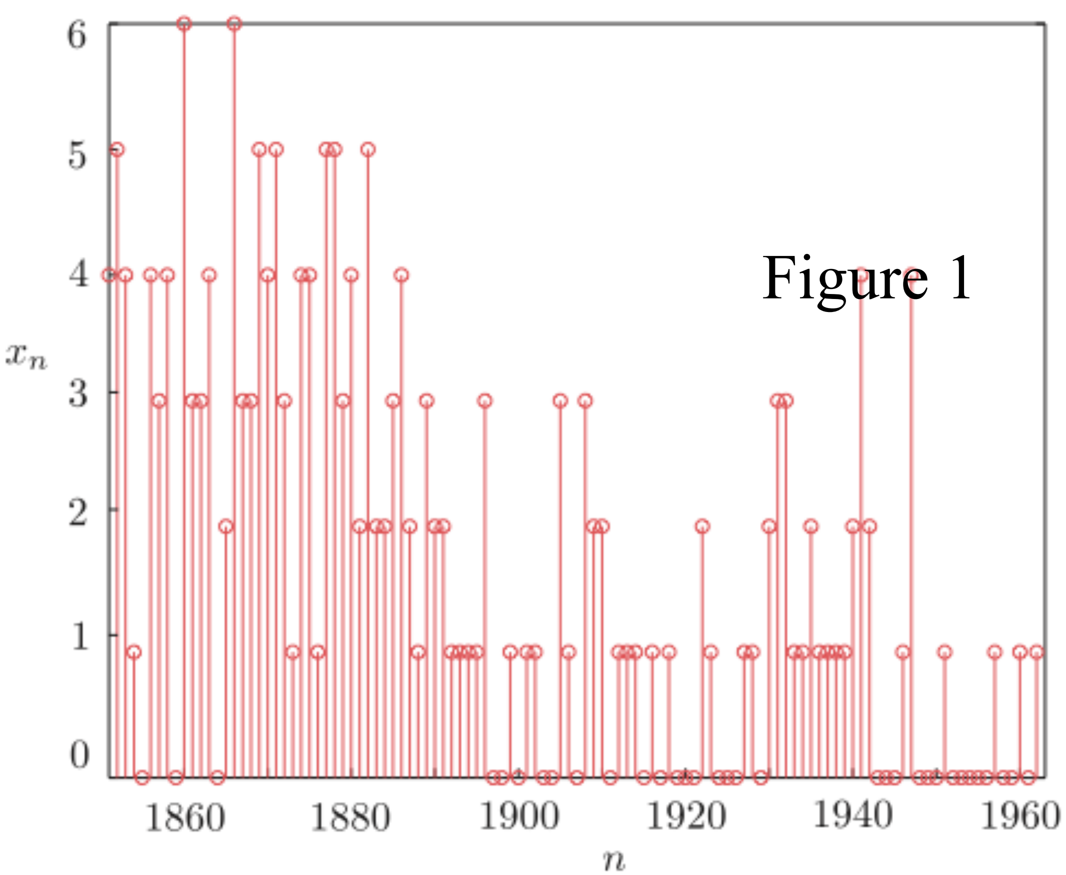
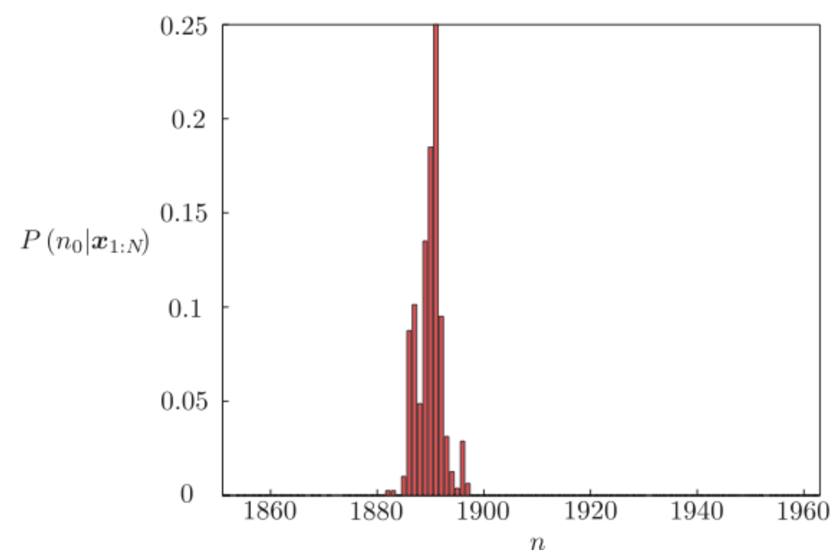
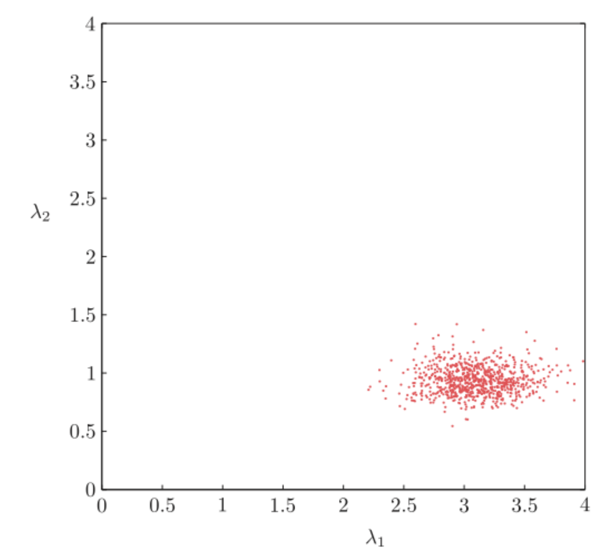


Figure 1

Coal mining data

x=[4 5 4 1 0 4 3 4 0 6 3 3 4 0 2 6 3 3 5 4 5 3 1 4 4 1 5 5 3 4 2 5 2 2 ...
3 4 2 1 3 2 2 1 1 1 1 3 0 0 1 0 1 1 0 0 3 1 0 3 2 2 0 1 1 1 0 1 0 1 0 ...
0 0 2 1 0 0 0 1 1 0 2 3 3 1 1 2 1 1 1 1 2 4 2 0 0 0 1 4 0 0 0 1 0 0 0 ...
0 0 1 0 0 1 0 1]





Laws of Probability

"There is this thing called *probability*. It obeys the laws of an axiomatic system. When identified with the real world, it gives (partial) information about the future."

- What axiomatic system?
- How to identify to real world?
  - Bayesian or frequentist viewpoints are somewhat different "mappings" from axiomatic probability theory to the real world
  - yet both are useful

"*And, it gives a consistent and complete calculus of inference.*"

First, warmup exercise about frequentist notion of probabilities

Probability Theory

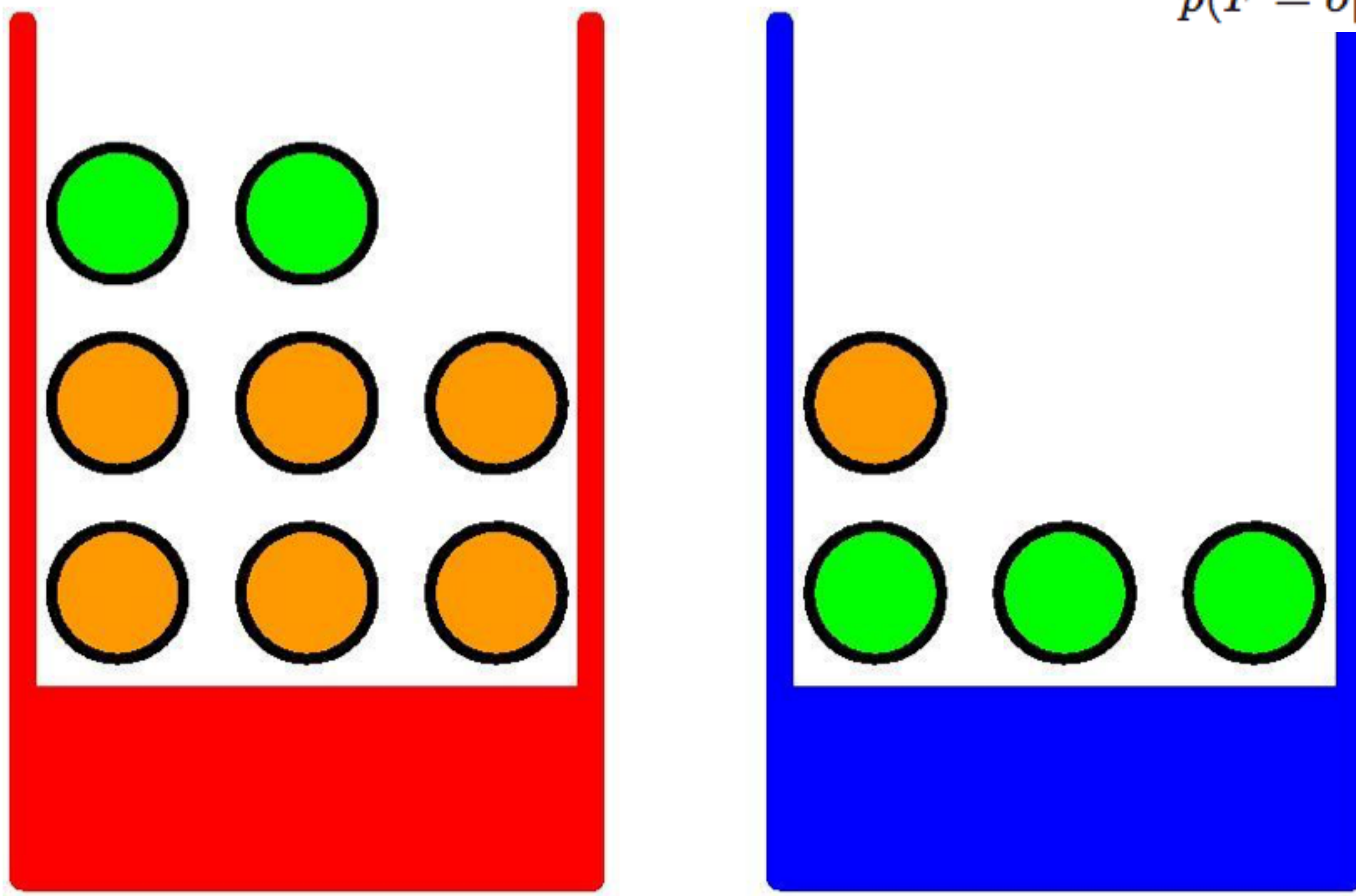
joint probabilities
X and Y random variables

Apples and Oranges

$$p(B=r) = 4/10$$ picking from red

$$p(B=b) = 6/10$$ picking from blue

$$p(F=a|B=r) = 1/4$$
$$p(F=o|B=r) = 3/4$$
$$p(F=a|B=b) = 3/4$$
$$p(F=o|B=b) = 1/4$$



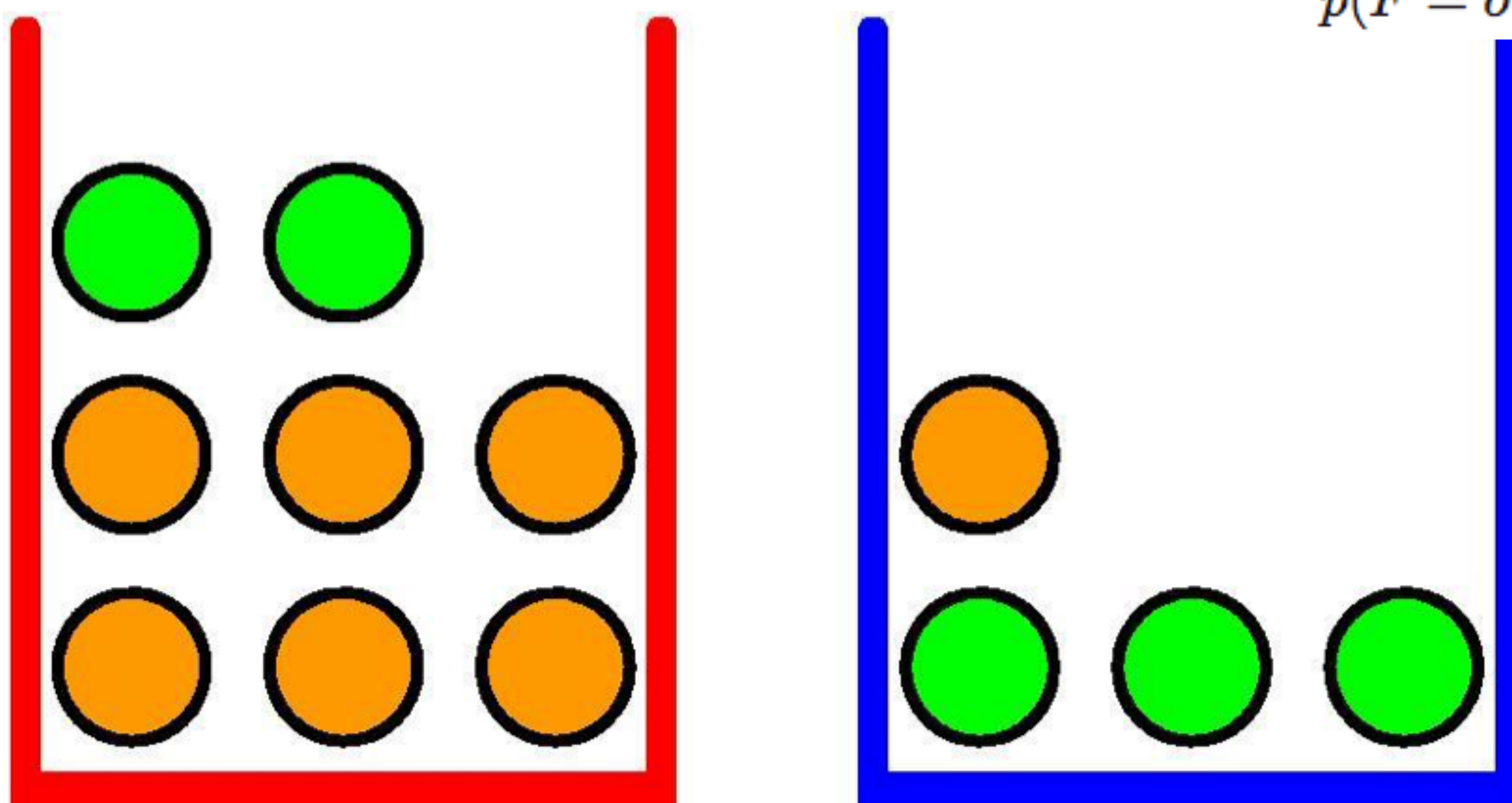
Probability Theory

joint probabilities
X and Y random variables

Apples and Oranges

$$p(B=r) = 4/10$$ picking from red
$$p(B=b) = 6/10$$ picking from blue

$$p(F=a|B=r) = 1/4$$
$$p(F=o|B=r) = 3/4$$
$$p(F=a|B=b) = 3/4$$
$$p(F=o|B=b) = 1/4$$



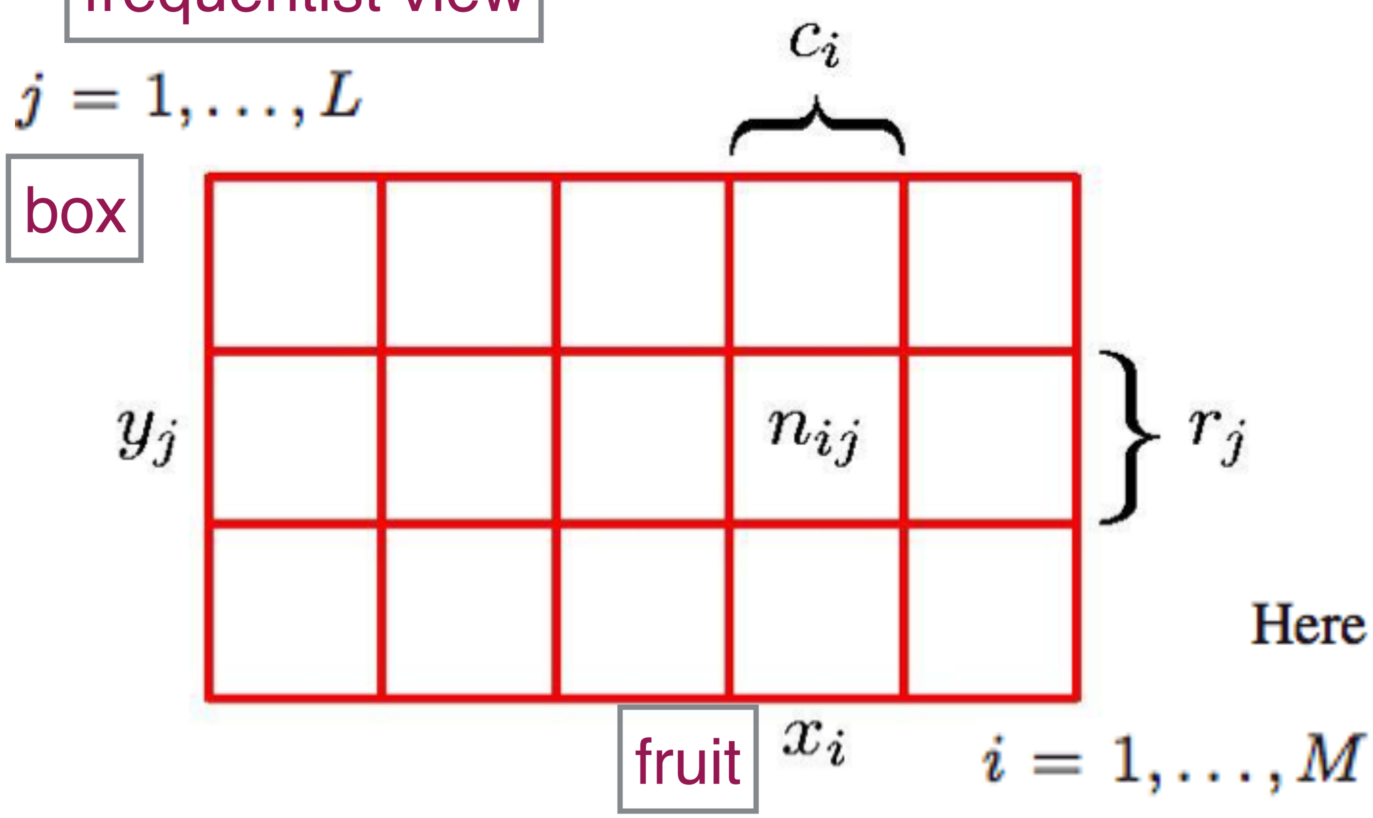
what is the probability to pick apple?
if orange, what is the probability that it came from blue box?

two elementary rules in probability theory help: sum rule and product rule

Probability Theory

joint probabilities
X and Y random variables

frequentist view



Marginal Probability

$$p(X = x_i) = \frac{c_i}{N}.$$

Here we are implicitly considering the limit $N \to \infty$

Joint Probability

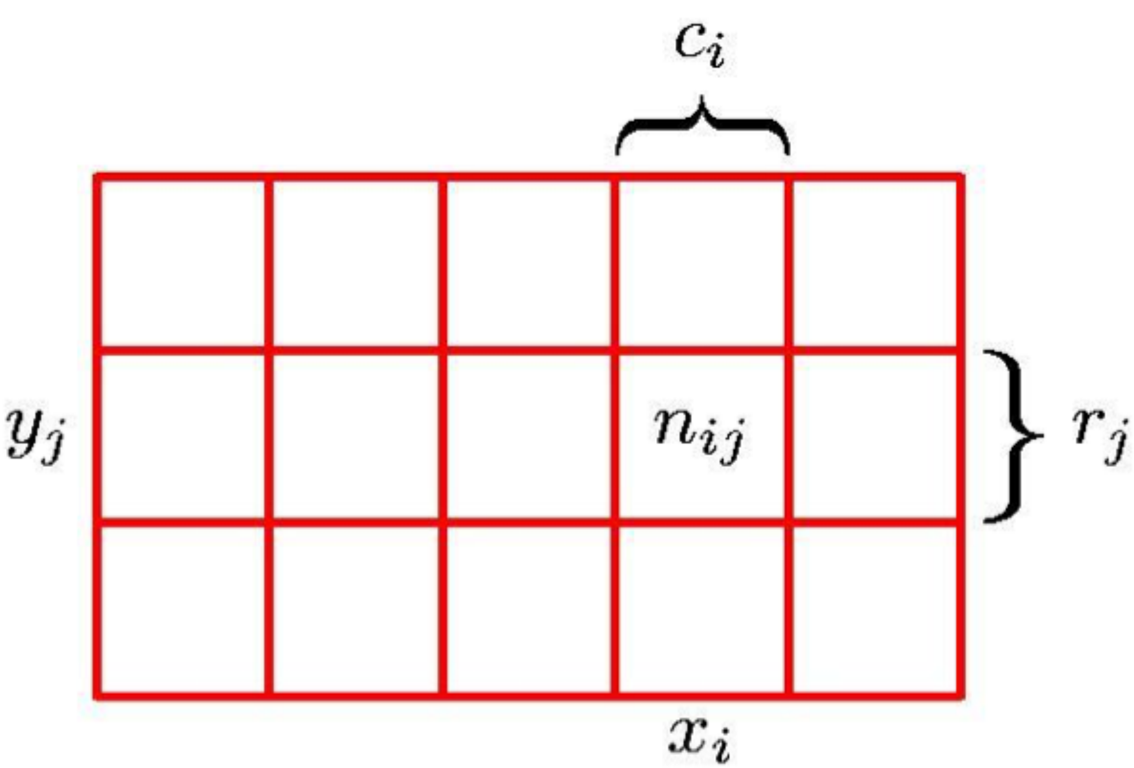
$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

Conditional Probability

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

Probability Theory

joint probabilities
X and Y random variables



Sum Rule

$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N}\sum_{j=1}^{L} n_{ij}$$

$$= \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

Product Rule

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N}$$

$$= p(Y = y_j | X = x_i) p(X = x_i)$$

The Rules of Probability

joint probabilities

X and Y random variables

Sum Rule $$p(X) = \sum_Y p(X, Y)$$

Product Rule $$p(X, Y) = p(Y|X)p(X)$$

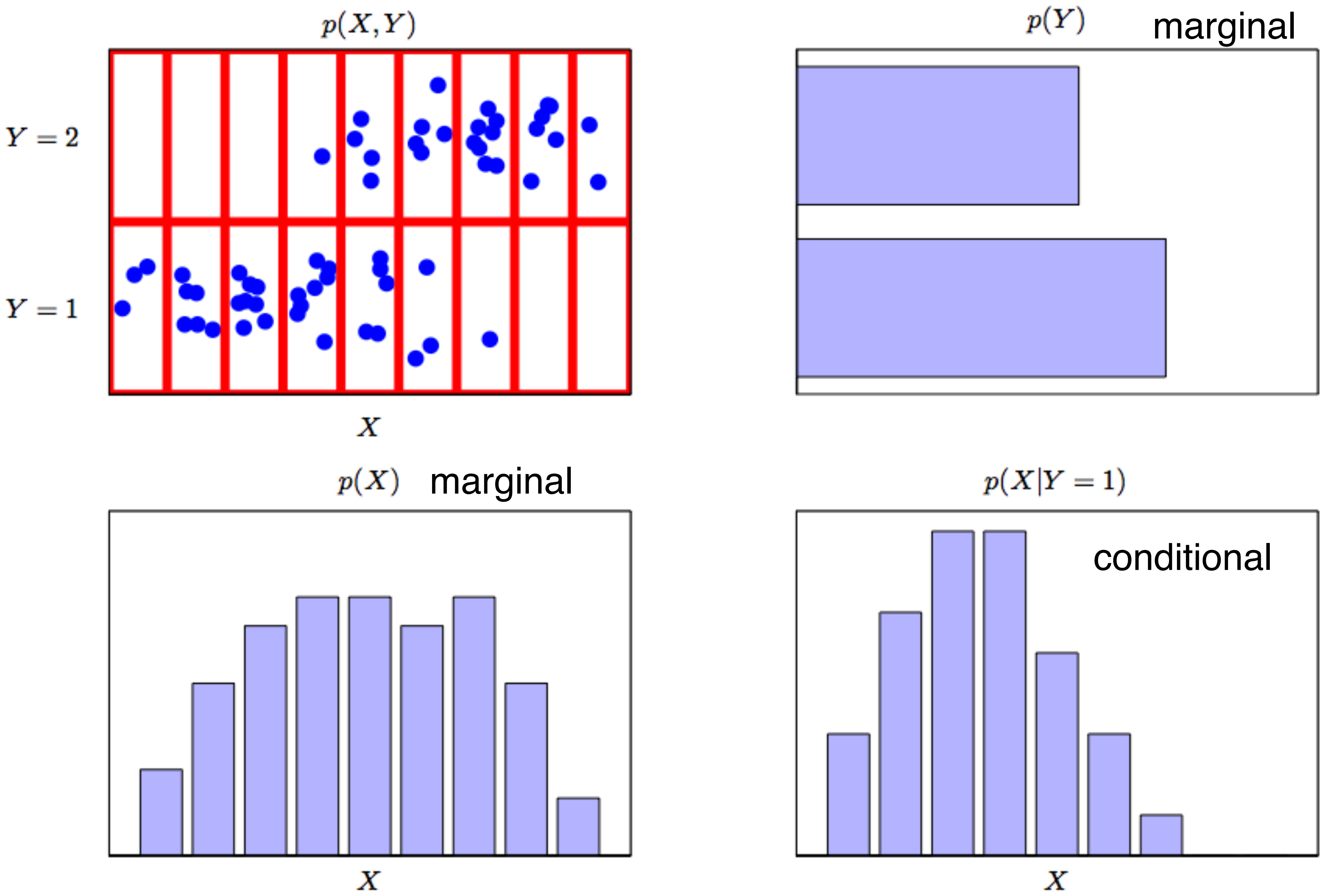
Bayes' Theorem

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

$$p(X) = \sum_Y p(X|Y)p(Y)$$ normalization

posterior $\propto$ likelihood × prior

tool: histogram of 60 events — joint probability distribution



return to the problem of two boxes with fruits

$$\begin{aligned} p(B=r) &= 4/10 \\ p(B=b) &= 6/10 \end{aligned}$$

marginal

$$p(B=r)+p(B=b)=1$$

normalization

$$\begin{aligned} p(F=a|B=r) &= 1/4 \\ p(F=o|B=r) &= 3/4 \\ p(F=a|B=b) &= 3/4 \\ p(F=o|B=b) &= 1/4 \end{aligned}$$

conditional

$$p(F=a|B=r)+p(F=o|B=r)=1$$

normalization

$$p(F=a|B=b)+p(F=o|B=b)=1$$

$$\begin{aligned} p(F=a) &= p(F=a|B=r)p(B=r)+p(F=a|B=b)p(B=b) \\ &= \frac{1}{4}\times\frac{4}{10}+\frac{3}{4}\times\frac{6}{10}=\frac{11}{20} \end{aligned}$$

picking apple

$$p(F=o)=1-11/20=9/20$$

picking orange

return to the problem of two boxes with fruits

if orange was picked, what was the probability of the box color ?

using Bayes' theorem, we can reverse the conditional probabilities:

$$p(B=r|F=o) = \frac{p(F=o|B=r)p(B=r)}{p(F=o)} = \frac{3}{4} \times \frac{4}{10} \times \frac{20}{9} = \frac{2}{3}$$

and from the sum rule:

$$p(B=b|F=o) = 1 - 2/3 = 1/3.$$

return to the problem of two boxes with fruits

if orange was picked, what was the probability of the box ?

using Bayes' theorem, we can reverse the conditional probabilities:

$$p(B=r|F=o) = \frac{p(F=o|B=r)p(B=r)}{p(F=o)} = \frac{3}{4} \times \frac{4}{10} \times \frac{20}{9} = \frac{2}{3}$$

and from the sum rule:

$$p(B=b|F=o) = 1 - 2/3 = 1/3.$$

interpretation of Bayes' theorem:
p(B) *prior probability*, if we are told that blue box was chosen
available before we observe the fruit

Once we are told it was orange, we can use Bayes' theorem to
calculate p(BIF) which is the *posterior probability*