02

Probabilities and Inference

Notice

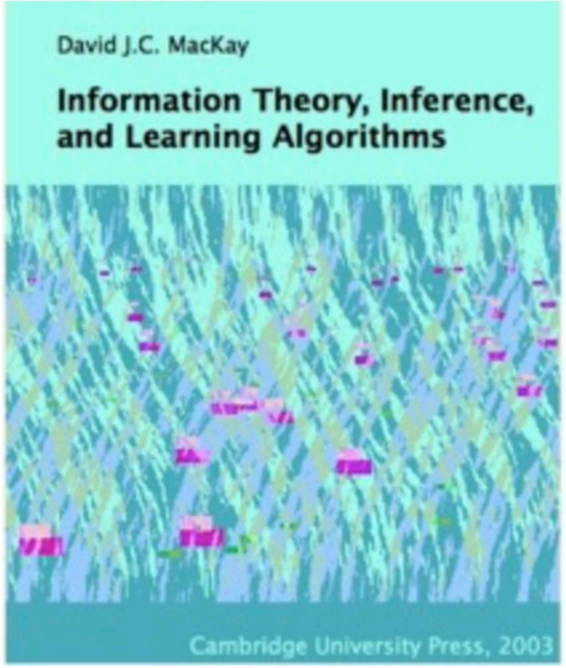
- **Author**
  - João Moura Pires (jmp@fct.unl.pt)

- This material can be freely used for personal or academic purposes without any previous authorization from the author, provided that this notice is maintained/kept.

- For commercial purposes the use of any part of this material requires the previous authorization from the author.

Bibliography

- **Many examples are extracted and adapted from:**



David J.C. MacKay

Information Theory, Inference, and Learning Algorithms

Cambridge University Press, 2003

Information Theory, Inference, and Learning Algorithms
David J.C. MacKay
2005, Version 7.2

- **And some slides were based on Iain Murray course**
  - http://www.inf.ed.ac.uk/teaching/courses/it/2014/

Table of Contents

Basics on Probabilities and Notation

Ensemble

- An **ensemble** X is a triple (x, A_X, P_X), where the *outcome* x is the value of a random variable, which takes on one of a set of possible values, $A_X = \{a_1, a_2, \ldots, a_i, \ldots, a_I\}$, having probabilities $P_X = \{p_1, p_2, \ldots, p_I\}$, with $P(x = a_i) = p_i$, $p_i \geq 0$ and $\sum_{a_i \in X} P(x = a_i) = 1$
  - A_X is the *alphabet* of X
  - $P(x = a_i)$ may be written as $P(a_i)$ or as $P(x)$.

Ensemble

- An **ensemble** ***X*** is a triple (x, A_X, P_X), where the *outcome* x is the value of a random variable, which takes on one of a set of possible values, $A_X = \{a_1, a_2, \ldots, a_i, \ldots, a_I\}$, having probabilities $P_X = \{p_1, p_2, \ldots, p_I\}$, with $P(x = a_i) = p_i$, $p_i \geq 0$ and $\sum_{a_i \in X} P(x = a_i) = 1$
  - A_X is the *alphabet* of X
  - $P(x = a_i)$ may be written as $P(a_i)$ or as $P(x)$.

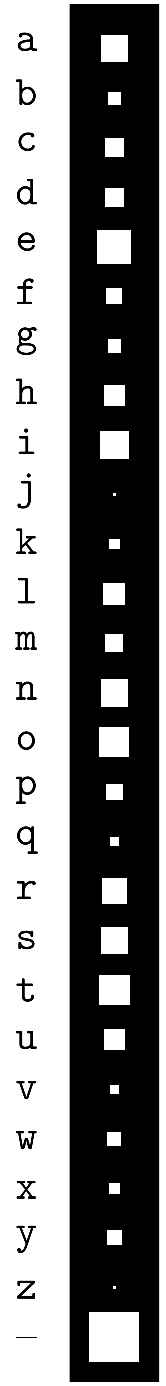
- **Probability of a subset.** If T is a subset of A_X then:

$$P(T) = P(x \in T) = \sum_{a_i \in T} P(x = a_i)$$

Ensemble: an example

- A letter that is randomly chosen from an English document.
- There are twenty-seven possible letters:
  - a–z,
  - space
  - character ‘-’.

i	a_i	p_i
1	a	0.0575
2	b	0.0128
3	c	0.0263
4	d	0.0285
5	e	0.0913
6	f	0.0173
7	g	0.0133
8	h	0.0313
9	i	0.0599
10	j	0.0006
11	k	0.0084
12	l	0.0335
13	m	0.0235
14	n	0.0596
15	o	0.0689
16	p	0.0192
17	q	0.0008
18	r	0.0508
19	s	0.0567
20	t	0.0706
21	u	0.0334
22	v	0.0069
23	w	0.0119
24	x	0.0073
25	y	0.0164
26	z	0.0007
27	–	0.1928

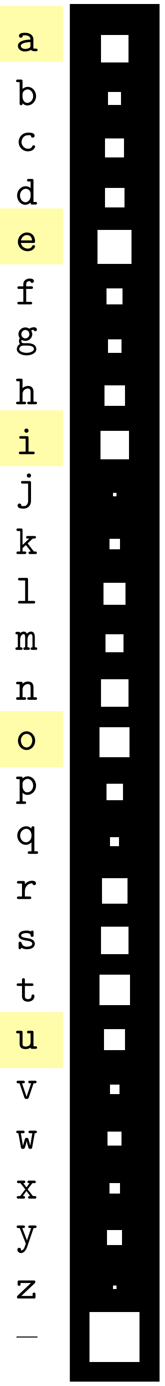


* estimated from “The Frequently Asked Questions Manual for Linux”

Ensemble: an example

- A letter that is randomly chosen from an English document.
- There are twenty-seven possible letters:
  - a–z,
  - space
  - character ‘-’.
- Let V be the set of vowels $V = \{a, e, i, o, u\}$.

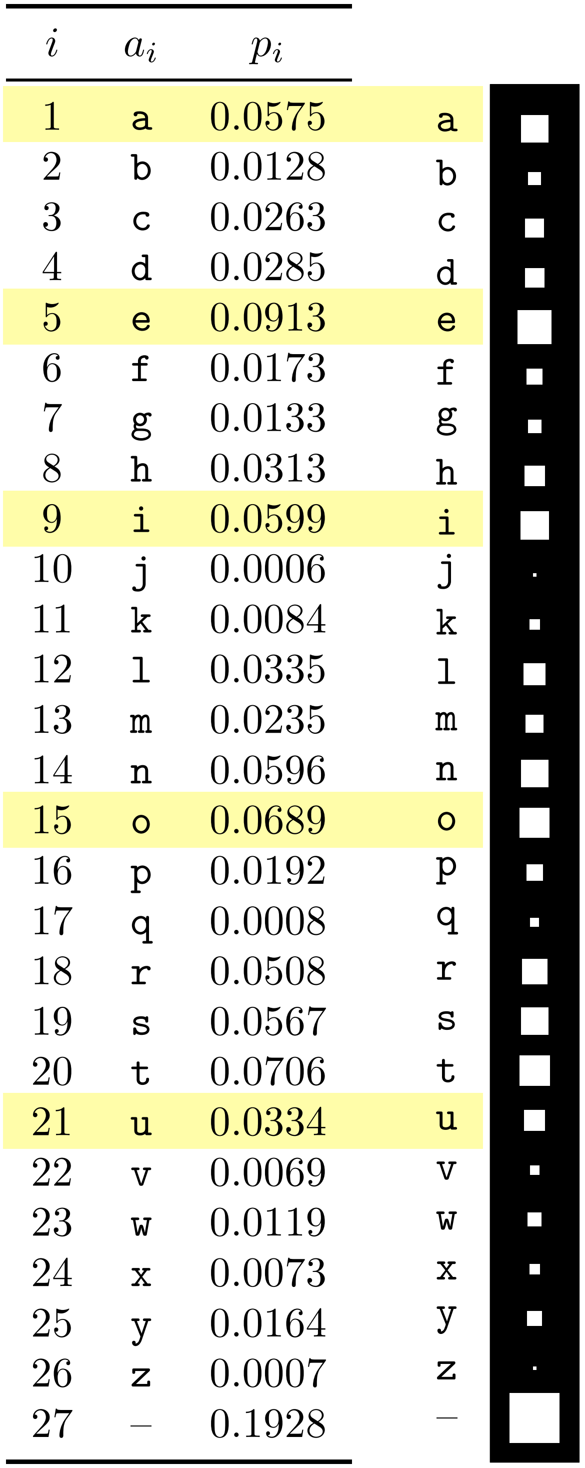
i	a_i	p_i
1	a	0.0575
2	b	0.0128
3	c	0.0263
4	d	0.0285
5	e	0.0913
6	f	0.0173
7	g	0.0133
8	h	0.0313
9	i	0.0599
10	j	0.0006
11	k	0.0084
12	l	0.0335
13	m	0.0235
14	n	0.0596
15	o	0.0689
16	p	0.0192
17	q	0.0008
18	r	0.0508
19	s	0.0567
20	t	0.0706
21	u	0.0334
22	v	0.0069
23	w	0.0119
24	x	0.0073
25	y	0.0164
26	z	0.0007
27	–	0.1928



* estimated from “The Frequently Asked Questions Manual for Linux”

Ensemble: an example

- A letter that is randomly chosen from an English document.
- There are twenty-seven possible letters:
  - a–z,
  - space
  - character ‘-’.
- Let V be the set of vowels $V = \{a, e, i, o, u\}$.
  - $P(V) = 0.06+0.09+0.06+0.07+0.03=0.31$



i	a_i	p_i
1	a	0.0575
2	b	0.0128
3	c	0.0263
4	d	0.0285
5	e	0.0913
6	f	0.0173
7	g	0.0133
8	h	0.0313
9	i	0.0599
10	j	0.0006
11	k	0.0084
12	l	0.0335
13	m	0.0235
14	n	0.0596
15	o	0.0689
16	p	0.0192
17	q	0.0008
18	r	0.0508
19	s	0.0567
20	t	0.0706
21	u	0.0334
22	v	0.0069
23	w	0.0119
24	x	0.0073
25	y	0.0164
26	z	0.0007
27	–	0.1928

* estimated from “The Frequently Asked Questions Manual for Linux”

Joint Ensemble

- An **joint ensemble** ***XY*** is an ensemble in which each outcome is an **ordered** pair x, y with

  $x \in A_x = \{a_1, \ldots, a_I\}$ and $y \in A_y = \{b_1, \ldots, a_J\}$.

  We call $P(x, y)$ the joint probability of x and y.

- Notes:
  - xy «-» x, y
  - the two variables are **not necessarily independent**.

Joint Ensemble

- An **joint ensemble** ***XY*** is an ensemble in which each outcome is an **ordered** pair x, y with

  $x \in A_x = \{a_1,\ldots,a_I\}$ and $y \in A_y = \{b_1,\ldots,a_J\}$.

  We call $P(x, y)$ the joint probability of x and y.

- Notes:
  - xy «-» x, y
  - the two variables are **not necessarily independent**.

- Example: bigrams xy in English from

  “The Frequently Asked Questions Manual for Linux”

  - In general $P(x, y) \neq P(y, x)$

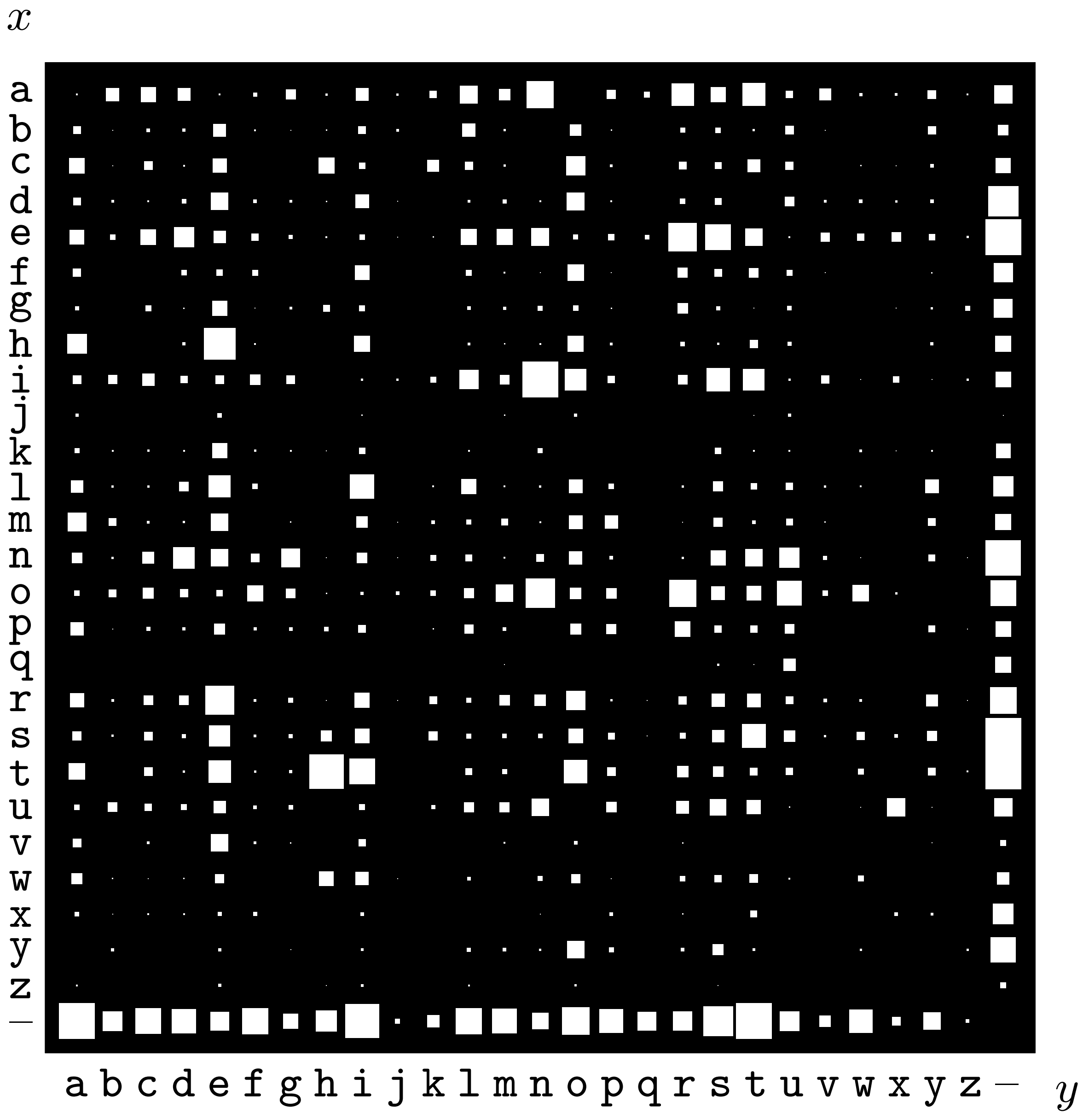
Joint Ensemble

- An **joint ensemble *XY*** is an ensemble in which each outcome is an **ordered** pair *x*, *y* with

  $x \in A_x = \{a_1,\ldots,a_I\}$ and $y \in A_y = \{b_1,\ldots,a_J\}$.

  We call $P(x, y)$ the joint probability of x and y.

- Notes:
  - xy «-» x, y
  - the two variables are **not necessarily independent**.



- Example: bigrams *xy* in English from

  “The Frequently Asked Questions Manual for Linux”

  - In general $P(x, y) \neq P(y, x)$

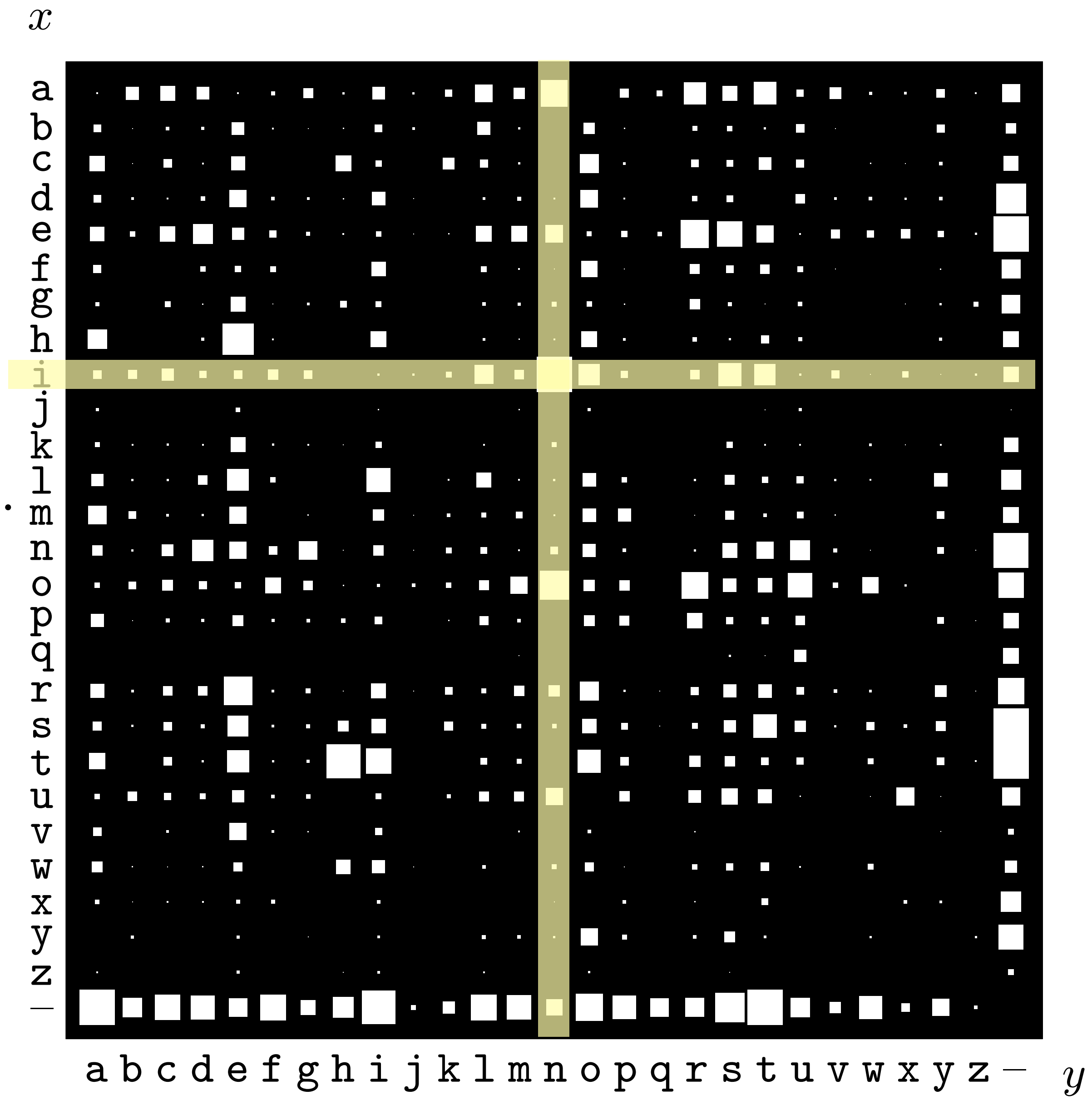
Joint Ensemble

- An **joint ensemble *XY*** is an ensemble in which each outcome is an **ordered** pair *x*, *y* with

  $x \in A_x = \{a_1, \ldots, a_I\}$ and $y \in A_y = \{b_1, \ldots, a_J\}$.

  We call $P(x, y)$ the joint probability of x and y.

- Notes:
  - *xy* «-» *x*, *y*
  - the two variables are **not necessarily independent**.



- Example: bigrams *xy* in English from

  “The Frequently Asked Questions Manual for Linux”

  - In general $P(x, y) \neq P(y, x)$

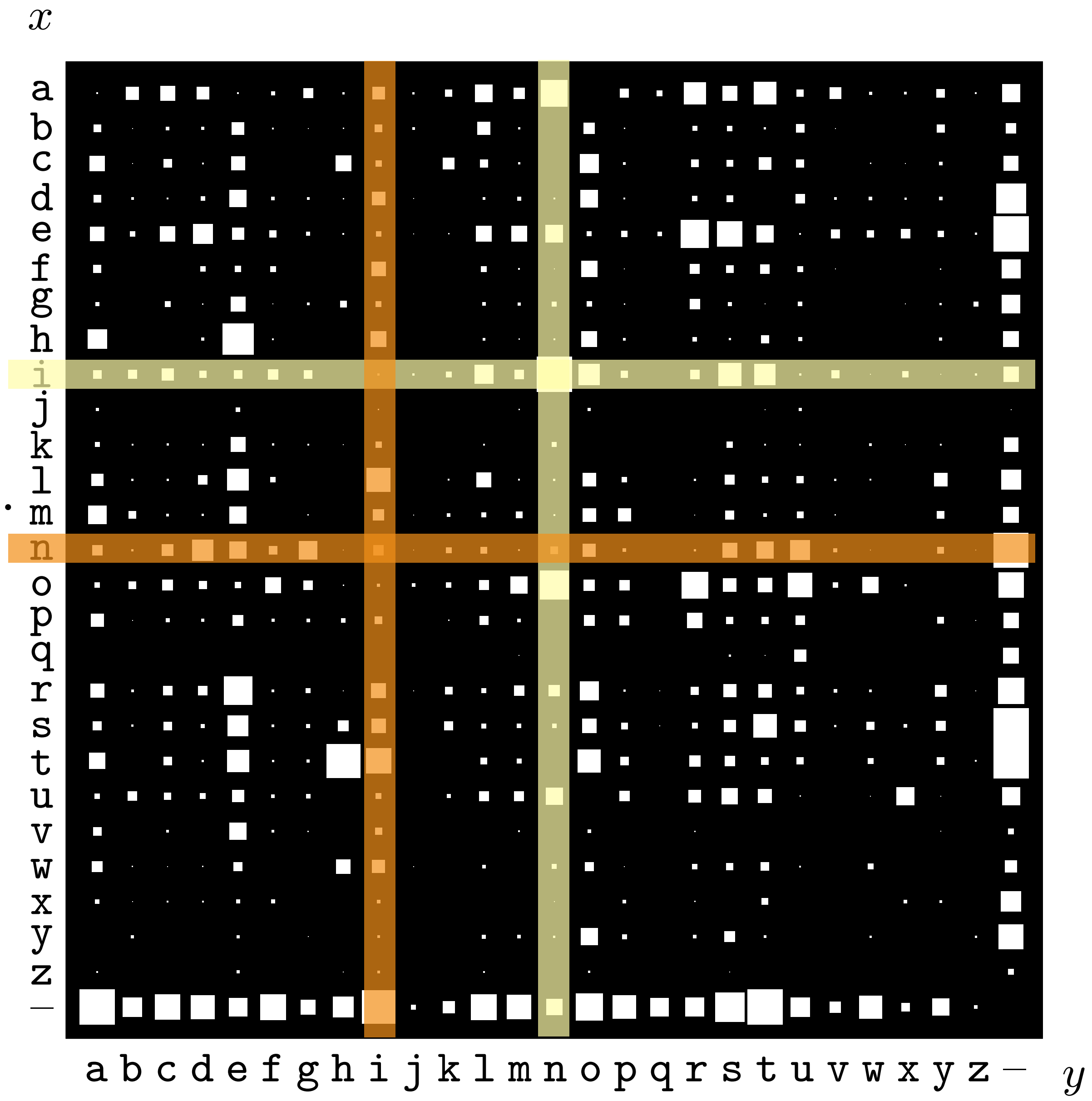
Joint Ensemble

- An **joint ensemble** ***XY*** is an ensemble in which each outcome is an **ordered** pair *x*, *y* with

  $x \in A_x = \{a_1, \ldots, a_I\}$ and $y \in A_y = \{b_1, \ldots, a_J\}$.

  We call $P(x, y)$ the joint probability of *x* and *y*.

- Notes:
  - *xy* «-» *x*, *y*
  - the two variables are **not necessarily independent**.



- Example: bigrams *xy* in English from

  “The Frequently Asked Questions Manual for Linux”

  - In general $P(x, y) \neq P(y, x)$

Marginal Probability

- Marginal probability $P(x)$ from the joint probability $P(x, y)$

$$P(x=a_i)=\sum_{y\in A_Y} P(x=a_i,y) \qquad P(x)=\sum_{y\in A_Y} P(x,y)$$

- Marginal probability $P(y)$ from the joint probability $P(x, y)$

$$P(y)=\sum_{x\in A_X} P(x,y)$$

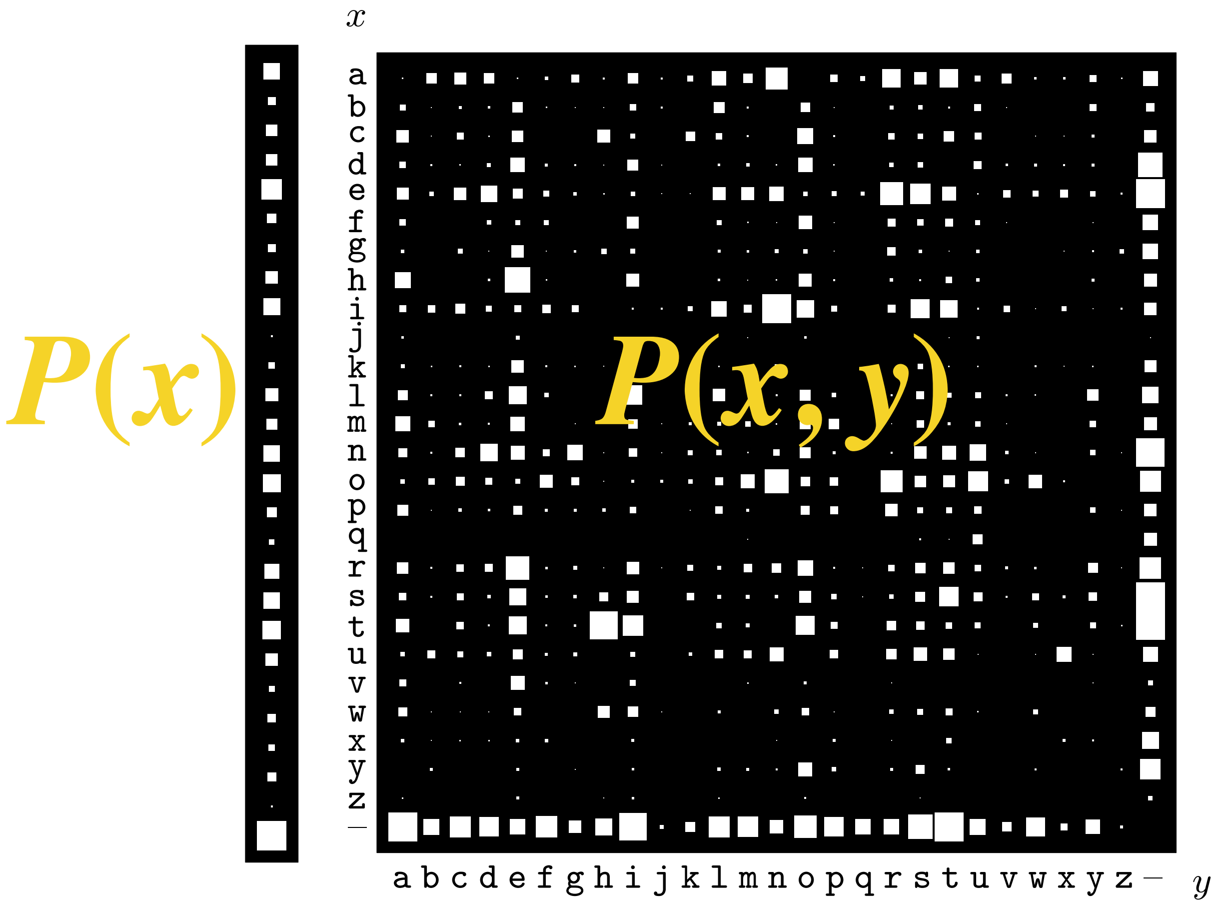
Marginal Probability

- Marginal probability $P(x)$ from the joint probability $P(x, y)$

$$P(x = a_i) = \sum_{y \in A_Y} P(x = a_i, y) \qquad P(x) = \sum_{y \in A_Y} P(x, y)$$

- Marginal probability $P(y)$ from the joint probability $P(x, y)$

$$P(y) = \sum_{x \in A_X} P(x, y)$$



Marginal Probability

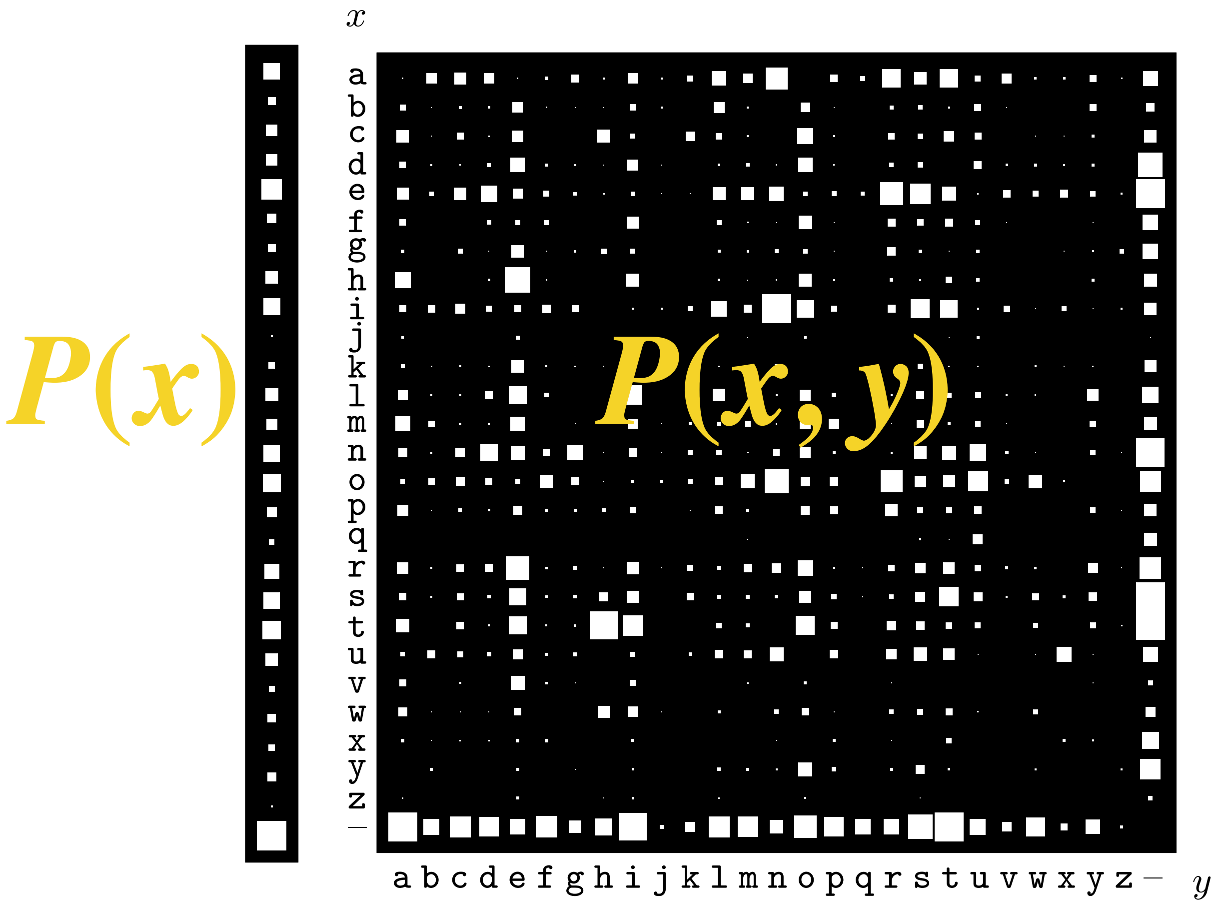
- Marginal probability $P(x)$ from the joint probability $P(x, y)$

$$P(x=a_i)=\sum_{y\in A_Y} P(x=a_i,y) \qquad P(x)=\sum_{y\in A_Y} P(x,y)$$

- Marginal probability $P(y)$ from the joint probability $P(x, y)$

$$P(y)=\sum_{x\in A_X} P(x,y)$$

- This joint ensemble has the special property that its two marginal distributions, $P(x)$ and P(y), are identical.



Condicional Probability

- $P(x = a_i \mid y = b_j)$ is the probability that x equals a_i, given y equals b_j:

$$P(x=a_i \mid y=b_j) = \frac{P(x=a_i, y=b_j)}{P(y=b_j)} \quad if \quad P(y=b_j) \neq 0$$

  - **If $P(y = b_j) = 0$ then $P(x = a_i \mid y = b_j)$ is undefined**

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Condicional Probability

- $P(x = a_i \mid y = b_j)$ is the probability that x equals a_i, given y equals b_j:

$$P(x=a_i \mid y=b_j) = \frac{P(x=a_i, y=b_j)}{P(y=b_j)} \quad \mathit{if} \quad P(y=b_j) \neq 0$$

  - **If $P(y = b_j) = 0$ then $P(x = a_i \mid y = b_j)$ is undefined**

- From $P(x, y)$ we can obtain $P(x \mid y)$ and $P(y \mid x)$

  - To obtain $P(y \mid x)$ we normalize the rows by dividing each $P(x, y)$ in a row by $P(x)$.

$$\sum_y P(y \mid x) = 1$$

Condicional Probability

- $P(x = a_i \mid y = b_j)$ is the probability that x equals a_i, given y equals b_j:

$$P(x = a_i \mid y = b_j) = \frac{P(x = a_i, y = b_j)}{P(y = b_j)} \quad \mathit{if} \quad P(y = b_j) \neq 0$$

  - **If $P(y = b_j) = 0$ then $P(x = a_i \mid y = b_j)$ is undefined**

- From $P(x, y)$ we can obtain $P(x \mid y)$ and $P(y \mid x)$
  - To obtain $P(y \mid x)$ we normalize the rows by dividing each $P(x, y)$ in a row by $P(x)$.

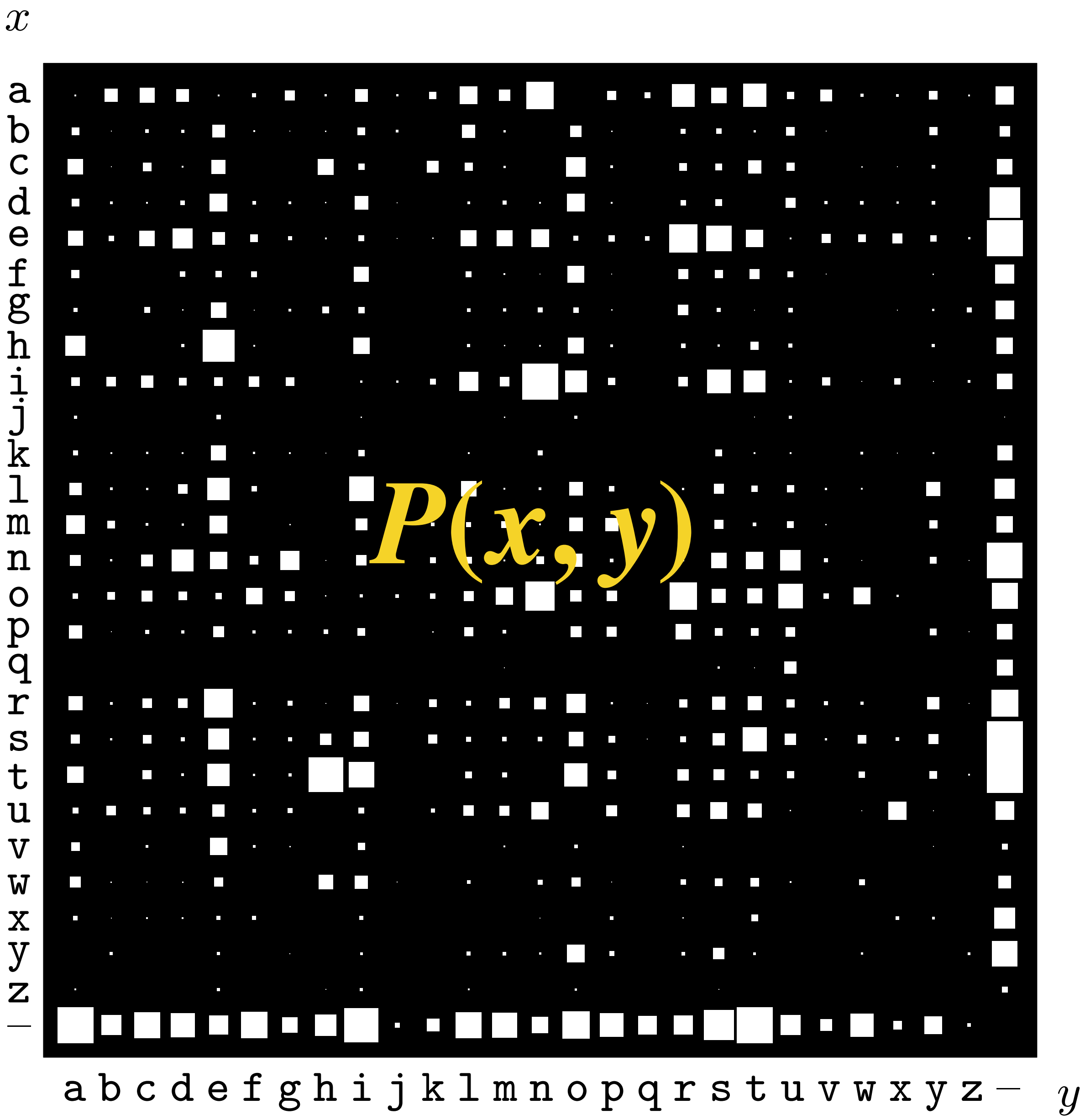
$$\sum_y P(y \mid x) = 1$$

  - To obtain $P(x \mid y)$ we normalize the columns by dividing each $P(x, y)$ in a column by $P(y)$.

$$\sum_x P(x \mid y) = 1$$

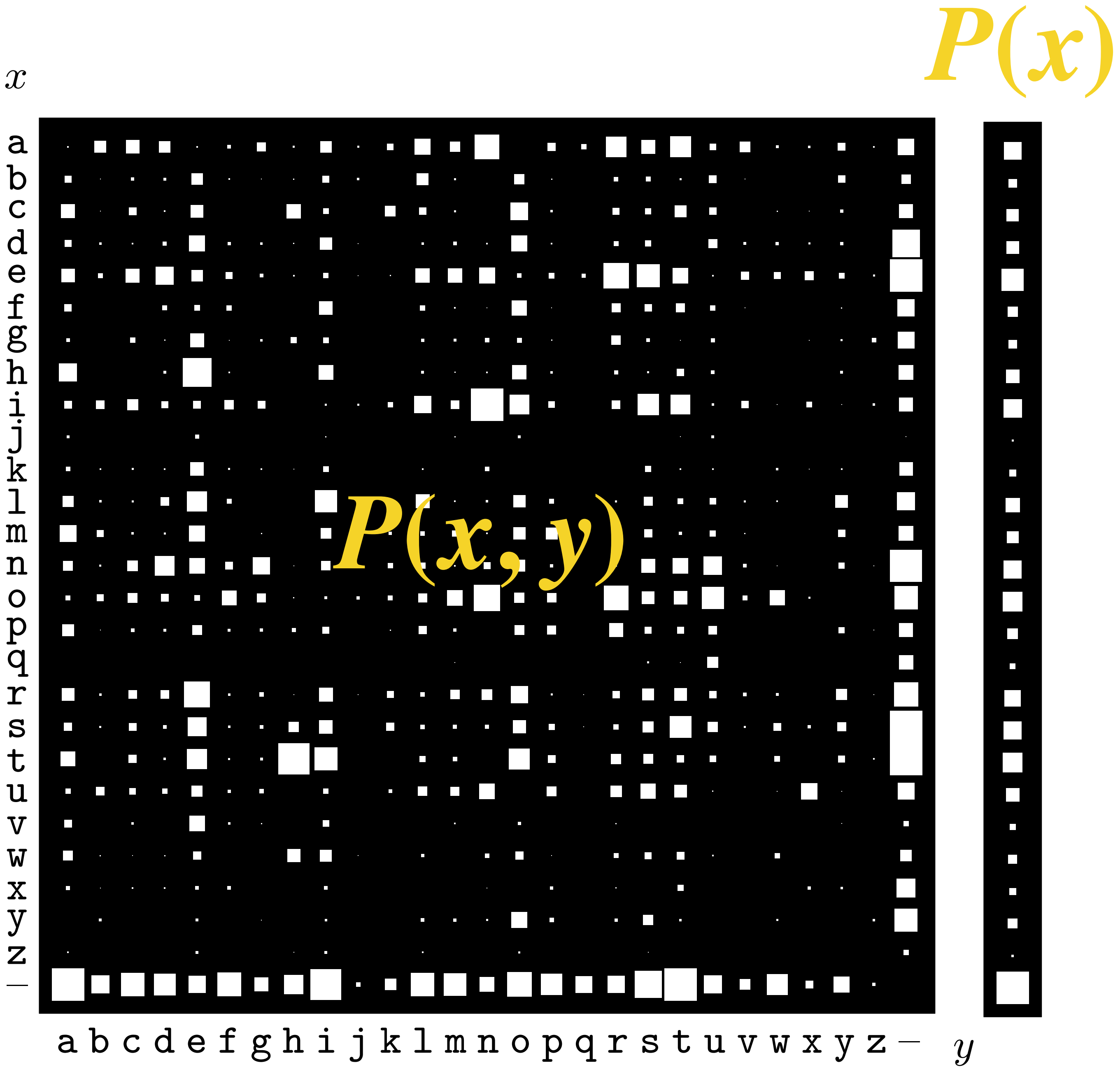
Condicional Probability: Example

- To obtain $P(y \mid x)$ we normalize the rows by **dividing each $P(x, y)$ in a row by $P(x)$**.



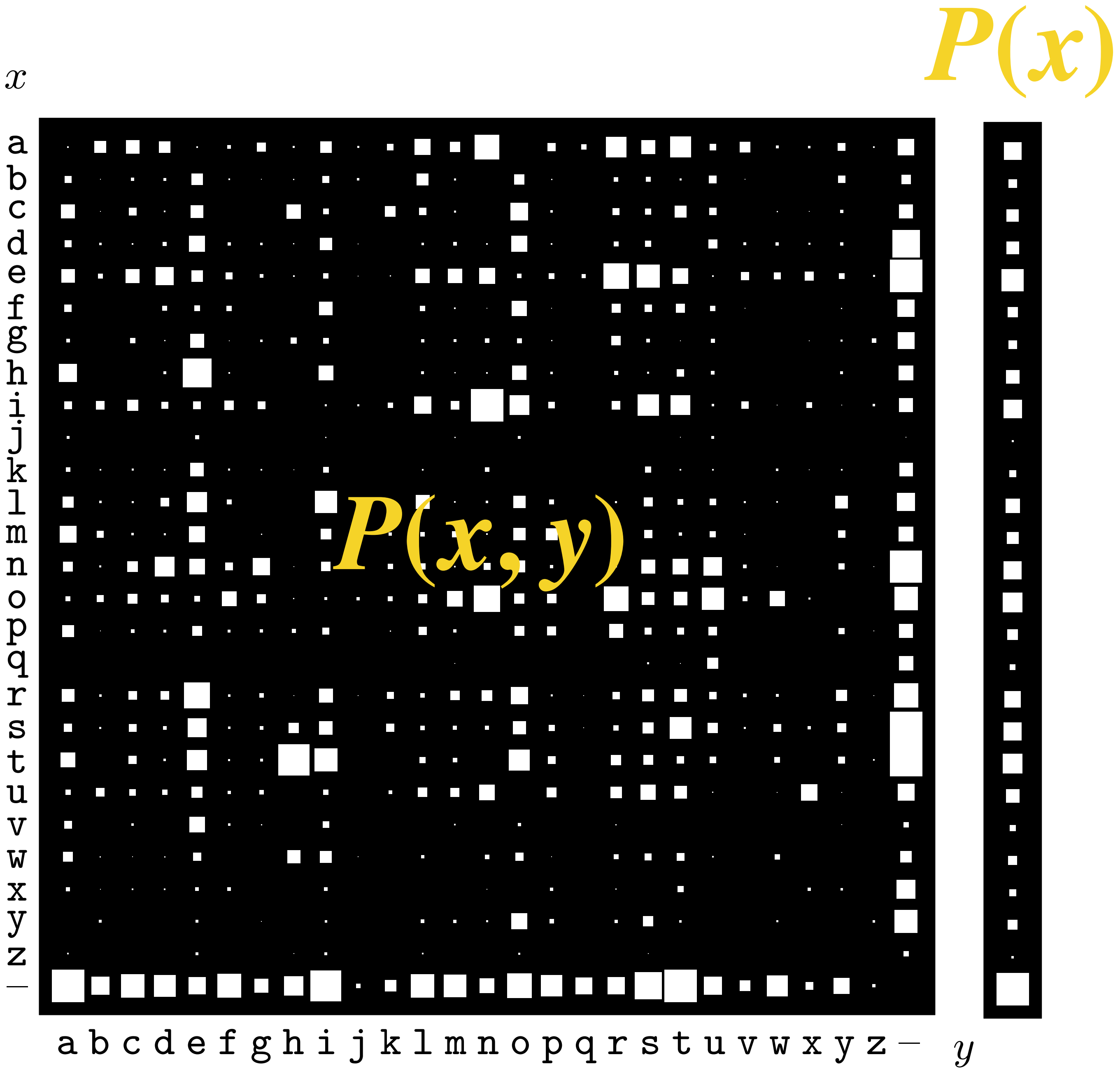
Condicional Probability: Example

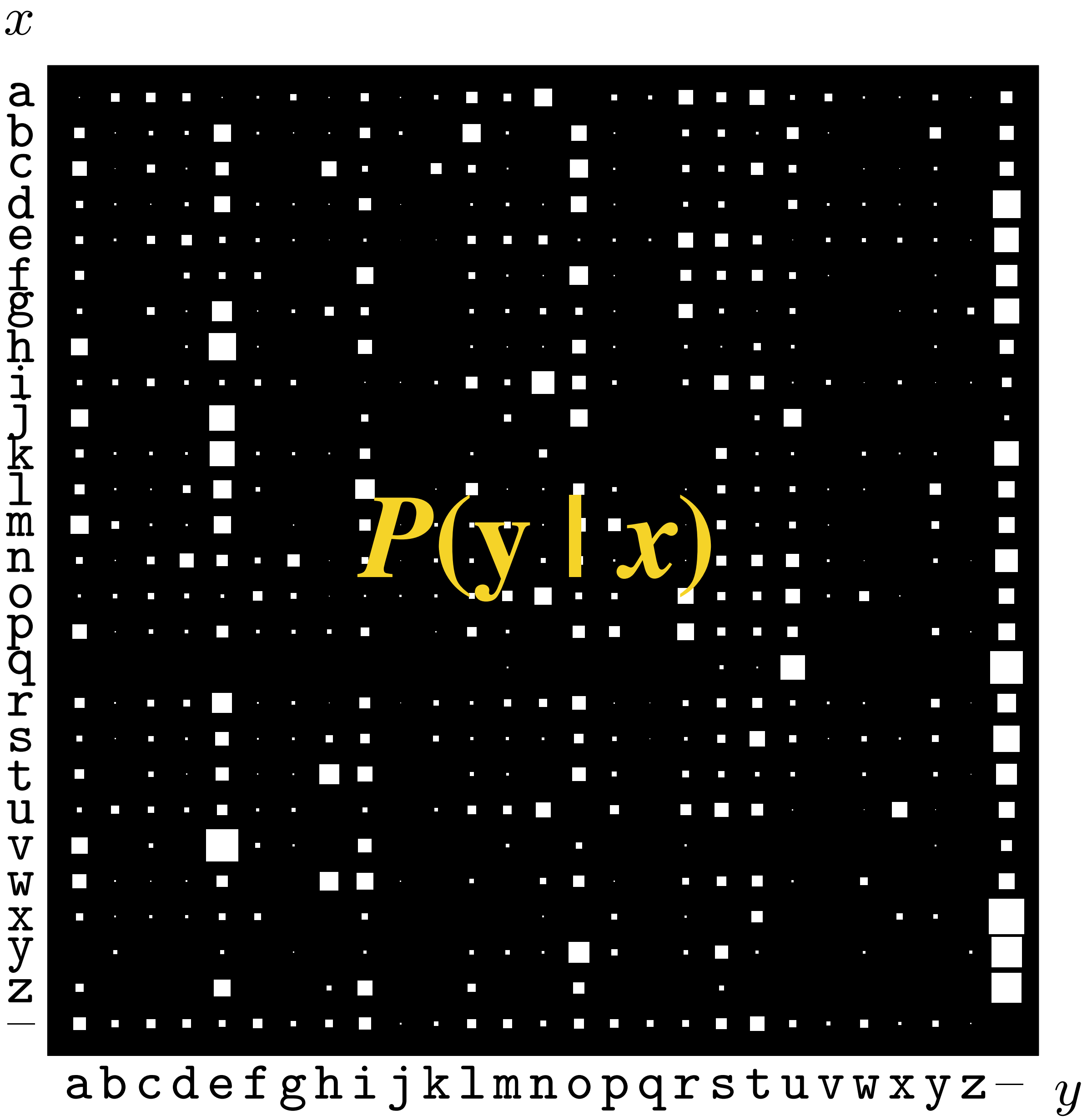
- To obtain $P(y \mid x)$ we normalize the rows by **dividing each $P(x, y)$ in a row by $P(x)$**.



Condicional Probability: Example

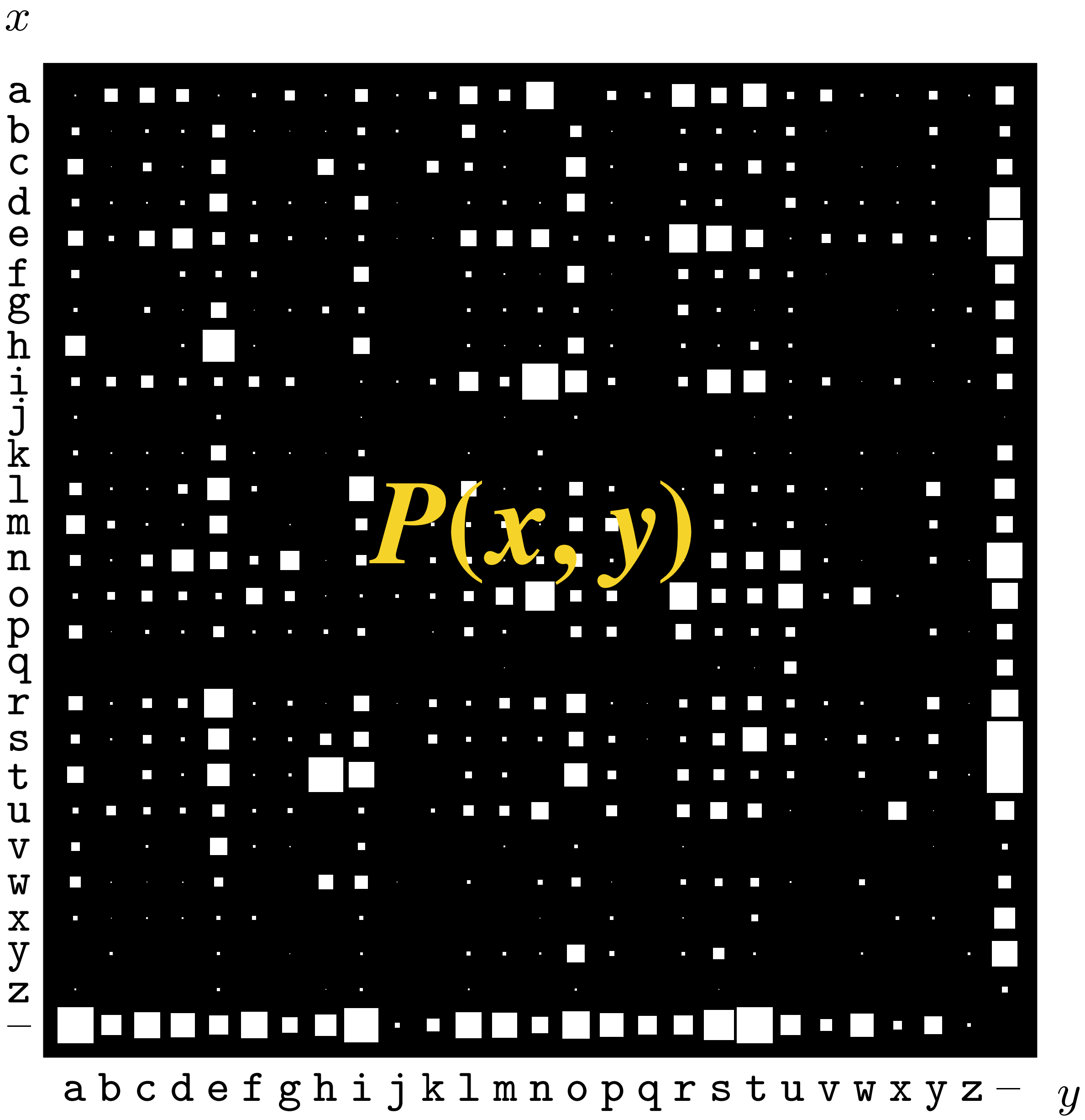
- To obtain $P(y \mid x)$ we normalize the rows by **dividing each $P(x, y)$ in a row by $P(x)$**.





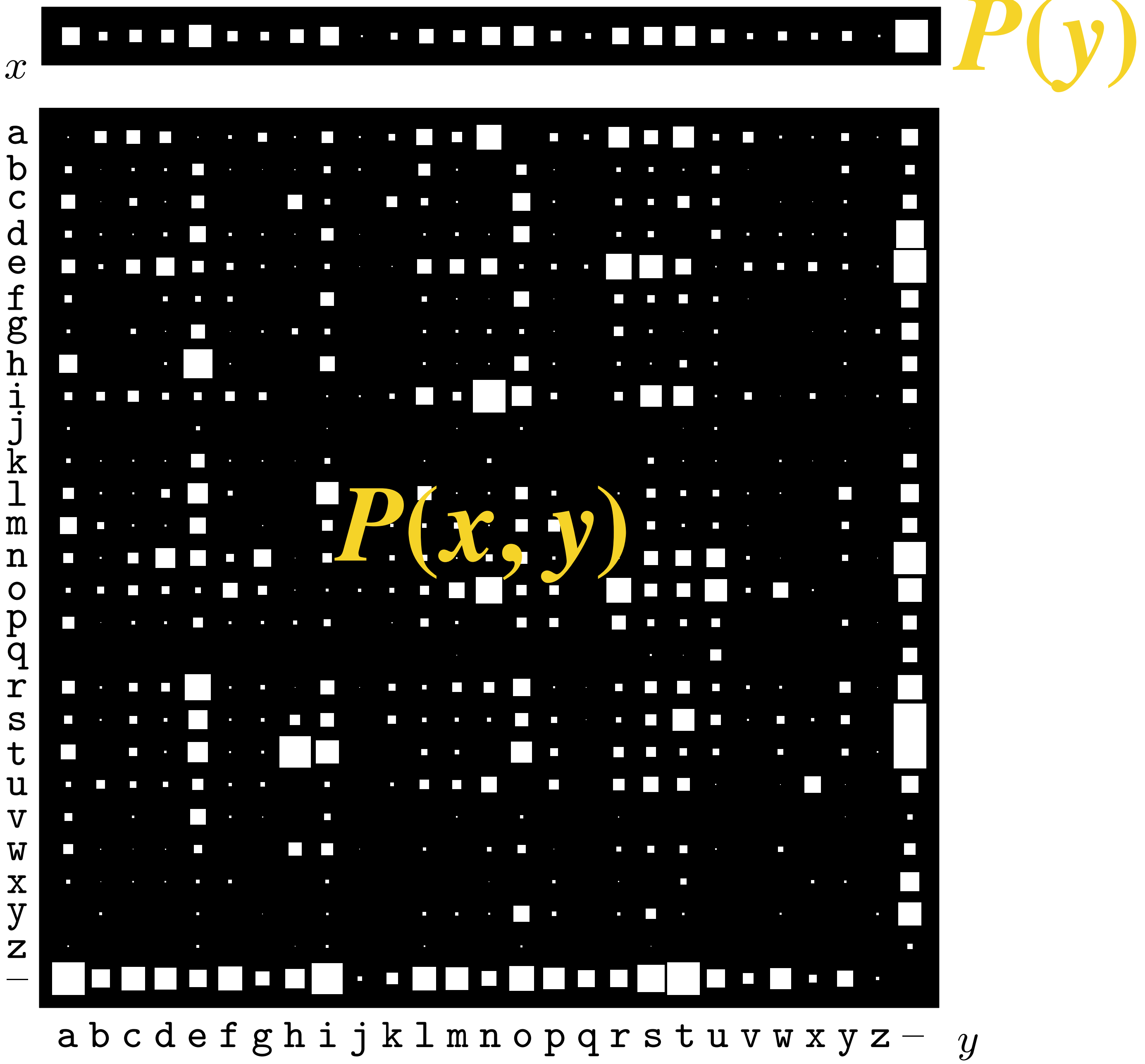
Condicional Probability: Example

- To obtain $P(x \mid y)$ we normalize the rows by **dividing each $P(x, y)$ in a column by $P(y)$**.



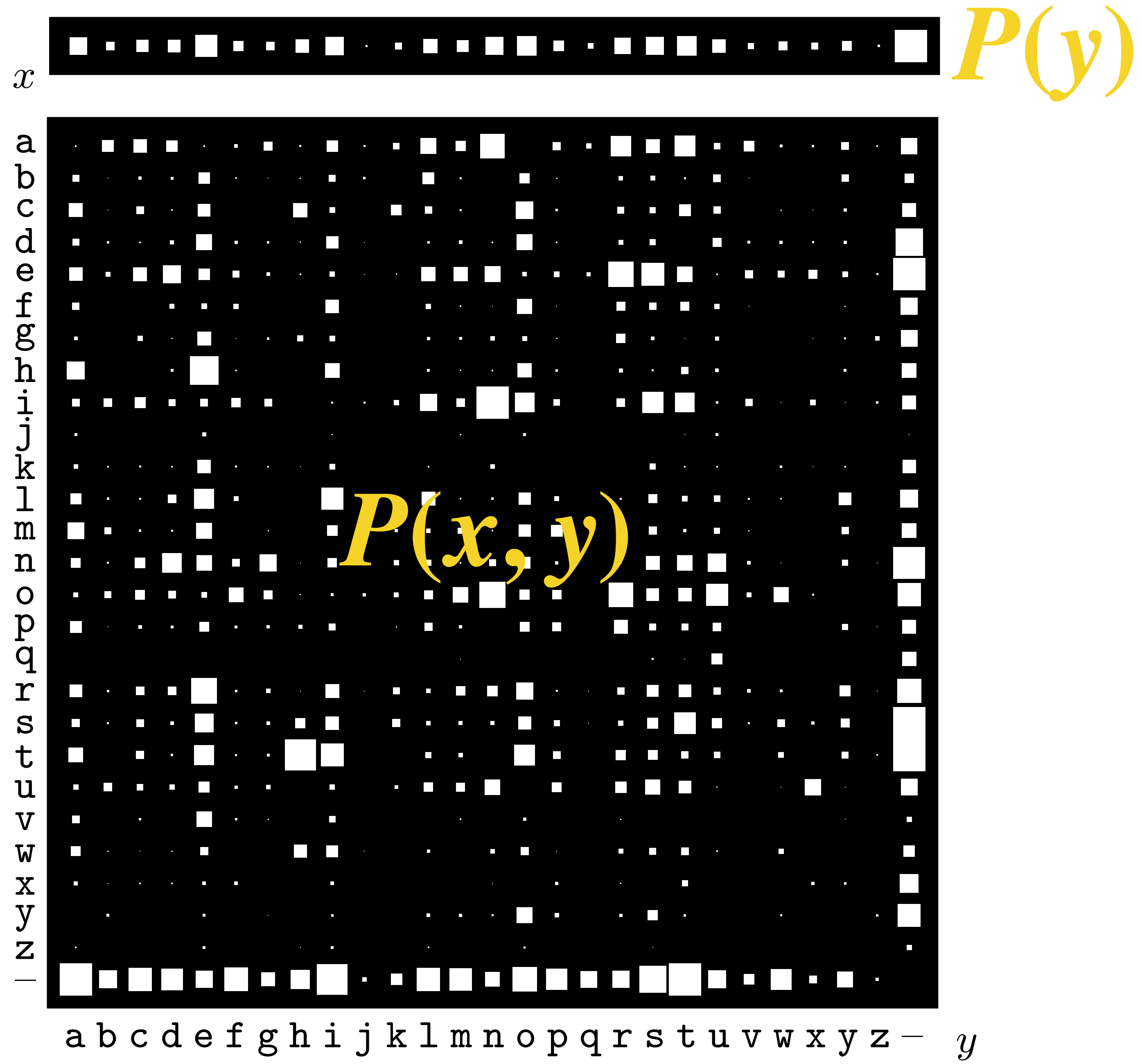
Condicional Probability: Example

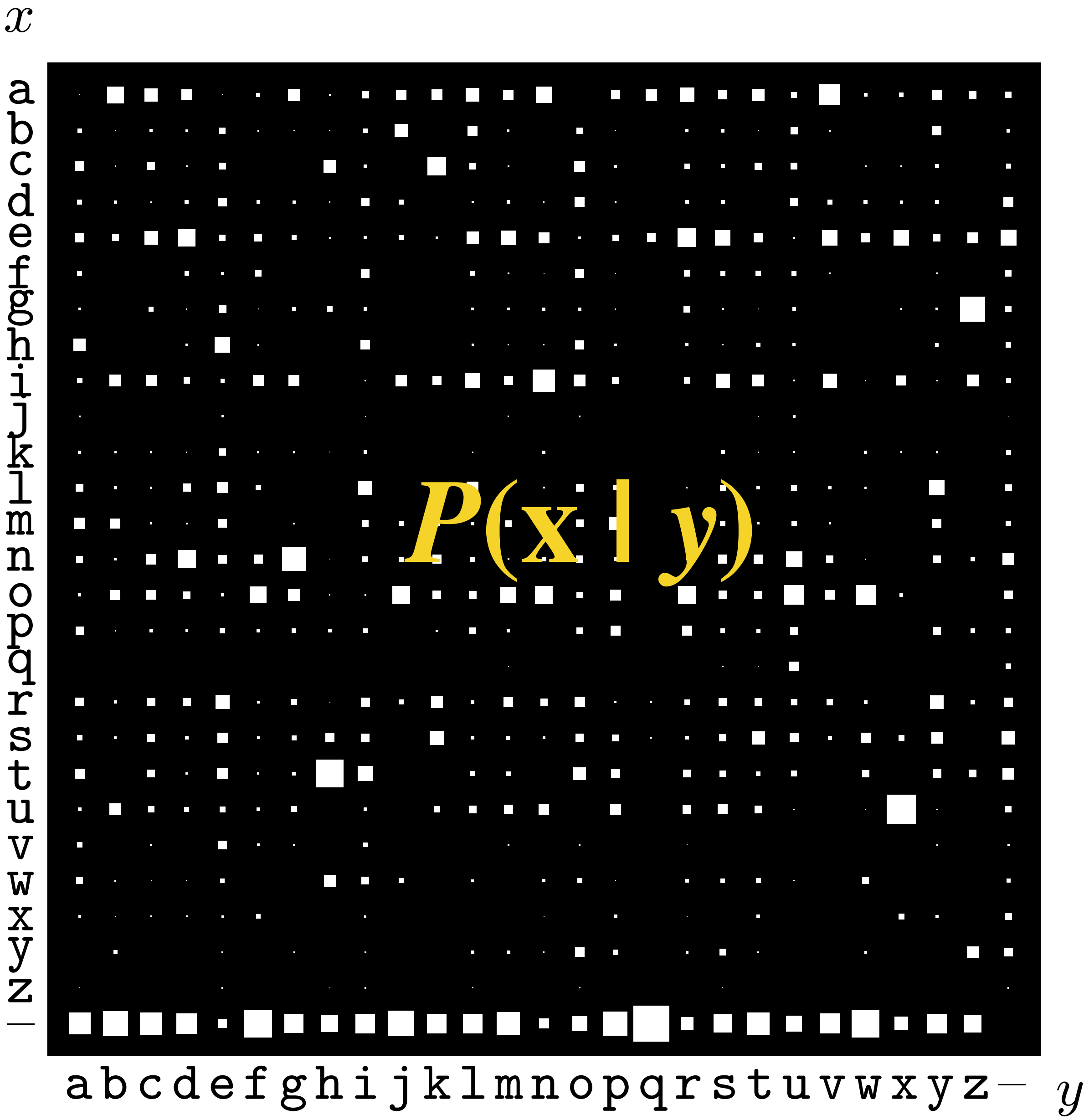
- To obtain $P(x \mid y)$ we normalize the rows by **dividing each $P(x, y)$ in a column by $P(y)$**.



Condicional Probability: Example

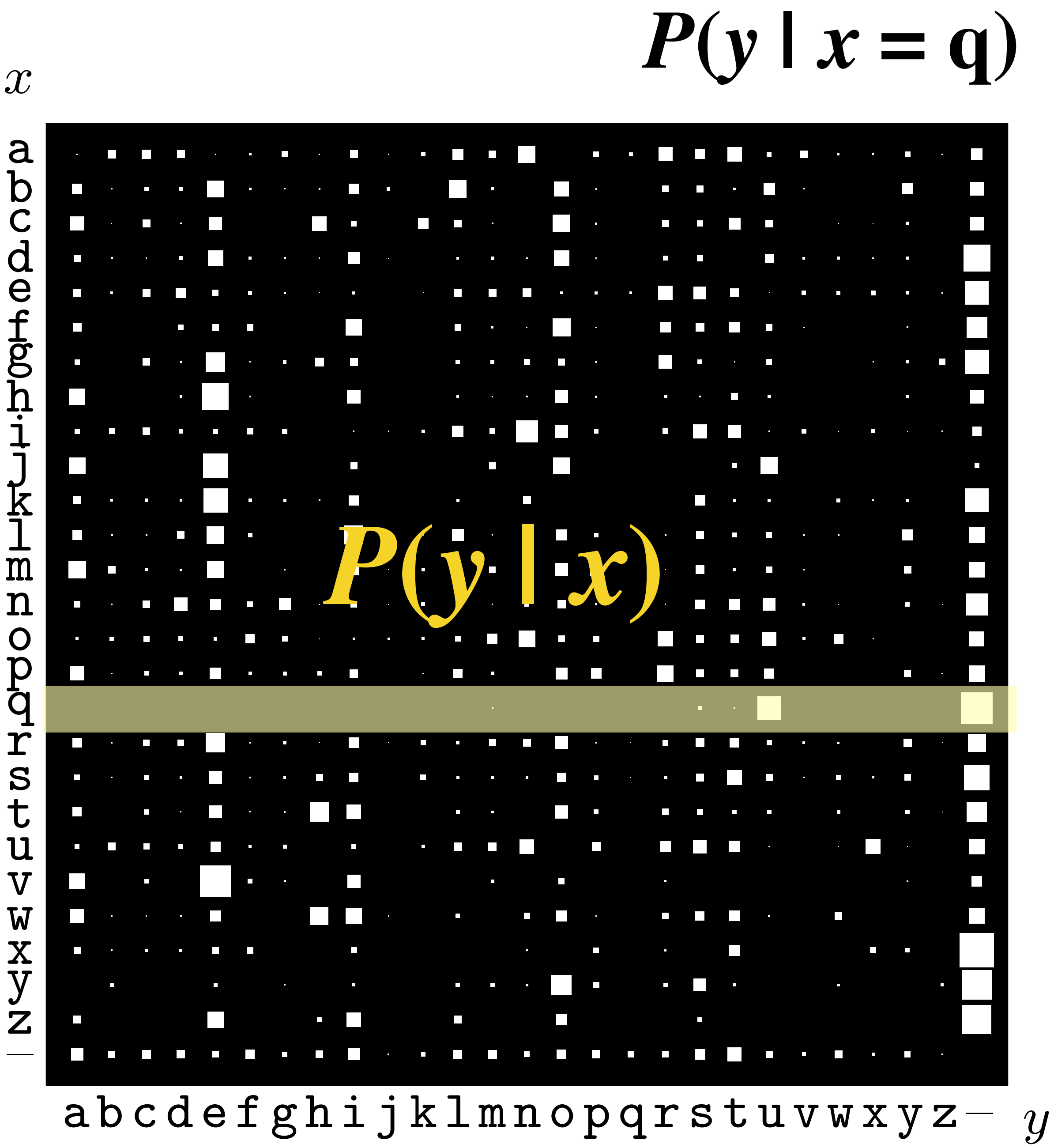
- To obtain $P(x \mid y)$ we normalize the rows by **dividing each $P(x, y)$ in a column by $P(y)$**.





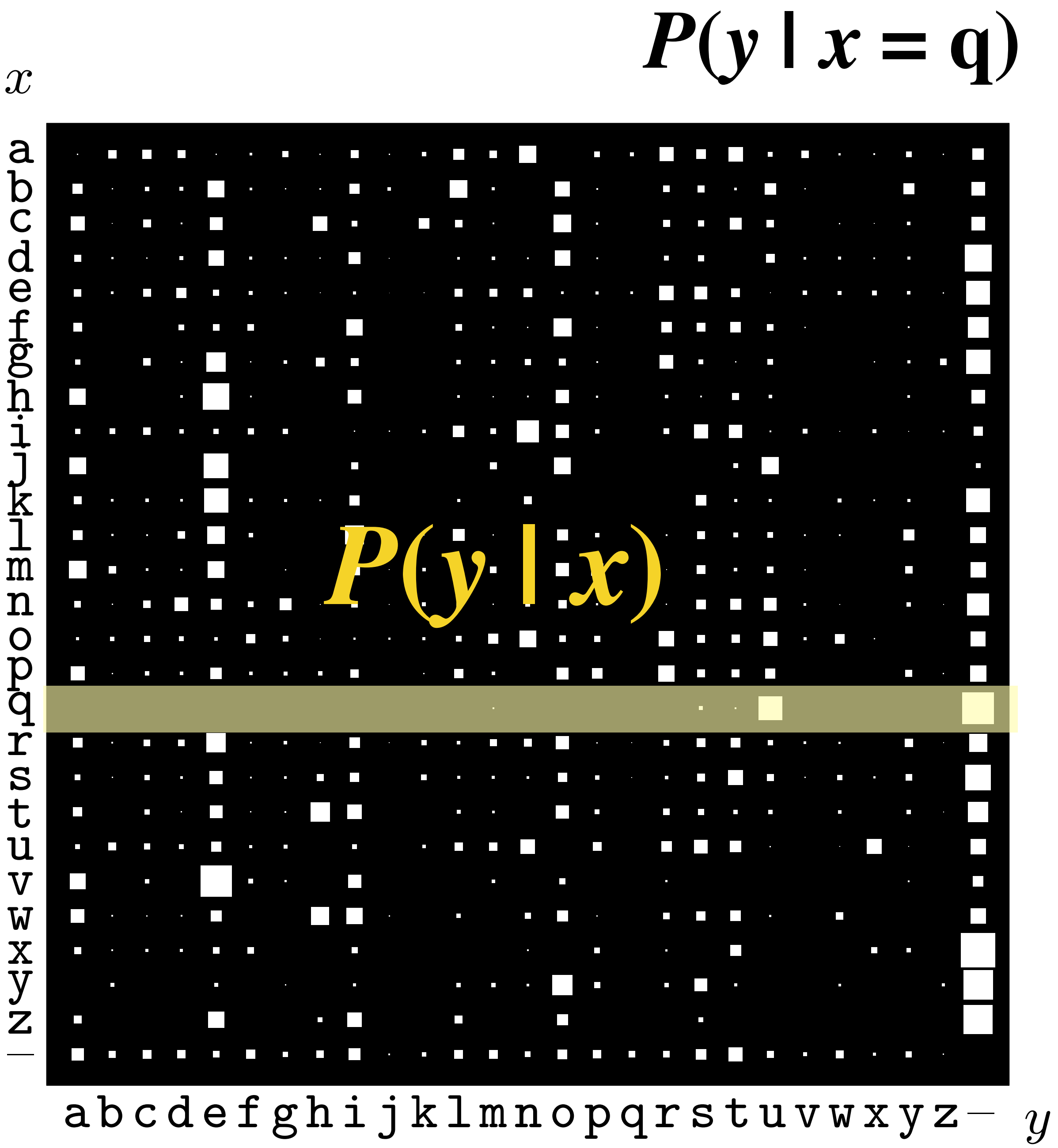
Condicional Probability: Example

- For each x, $P(y \mid x)$ is a probability distribution.



Condicional Probability: Example

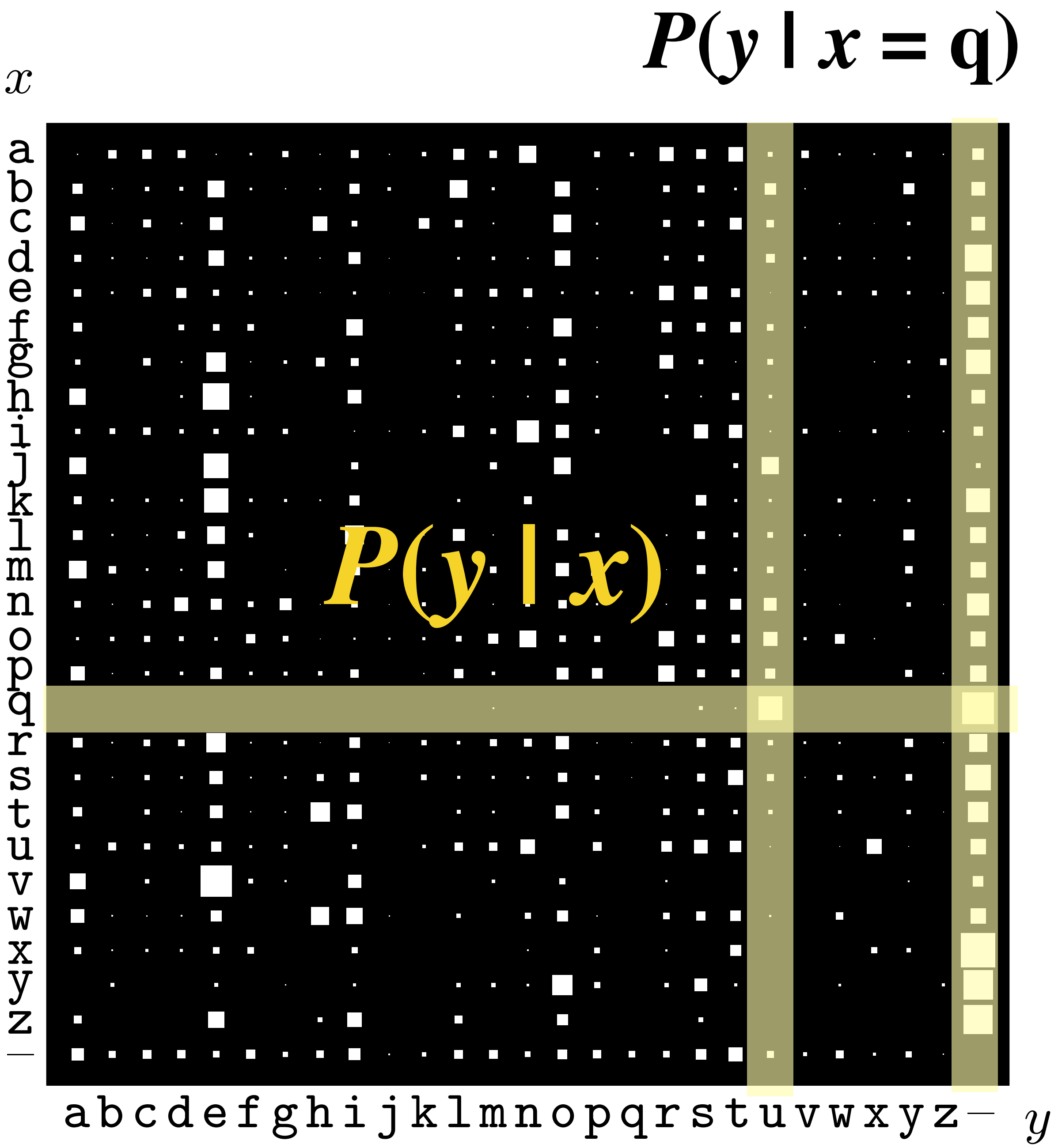
- For each *x*, *P*(*y* | *x*) is a probability distribution.



- The two most probable values for the second letter *y* given that the first letter *x* is **q** are **u** and **-**.

Condicional Probability: Example

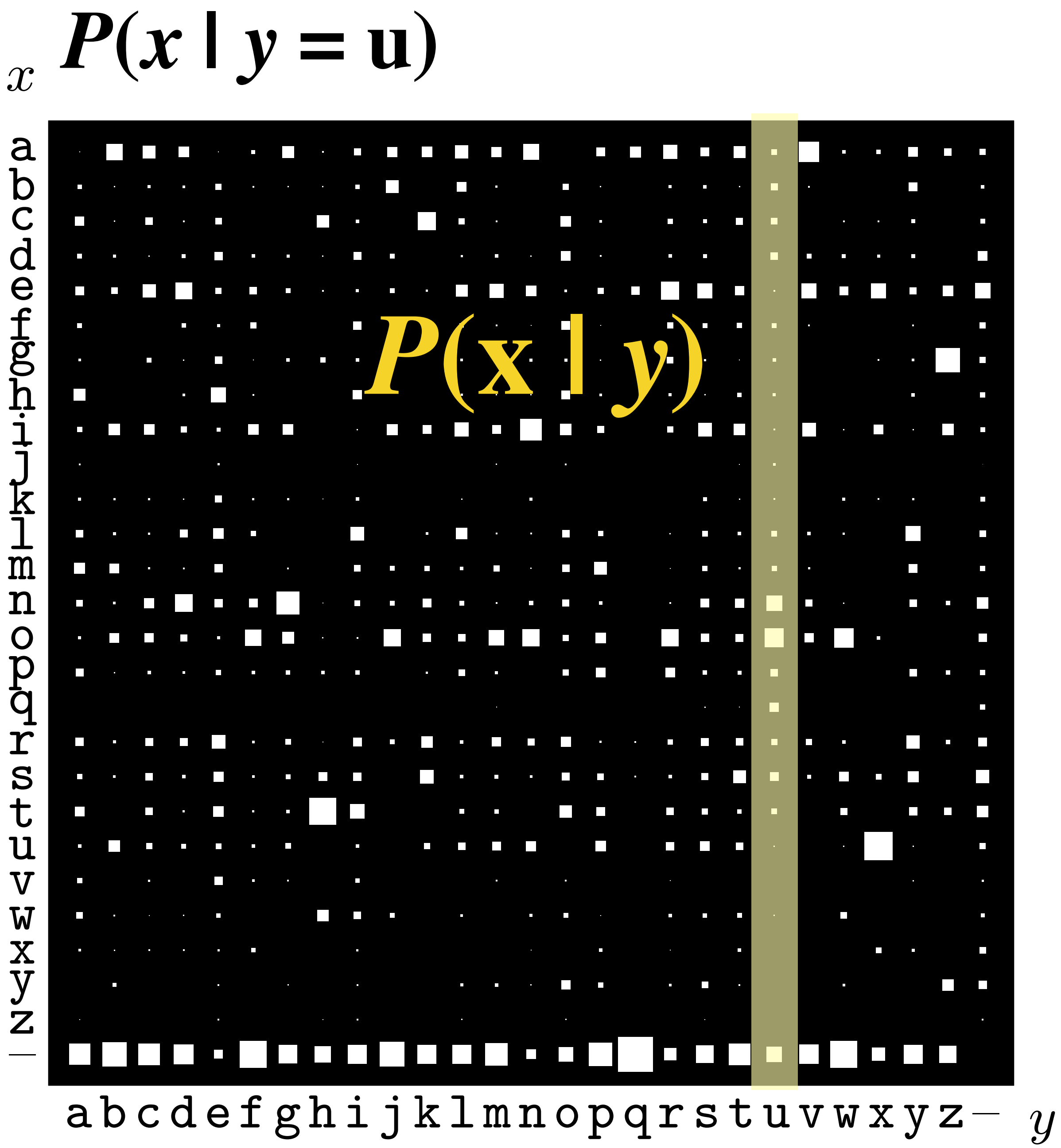
- For each x, $P(y \mid x)$ is a probability distribution.



- The two most probable values for the second letter y given that the first letter x is **q** are **u** and **-**.

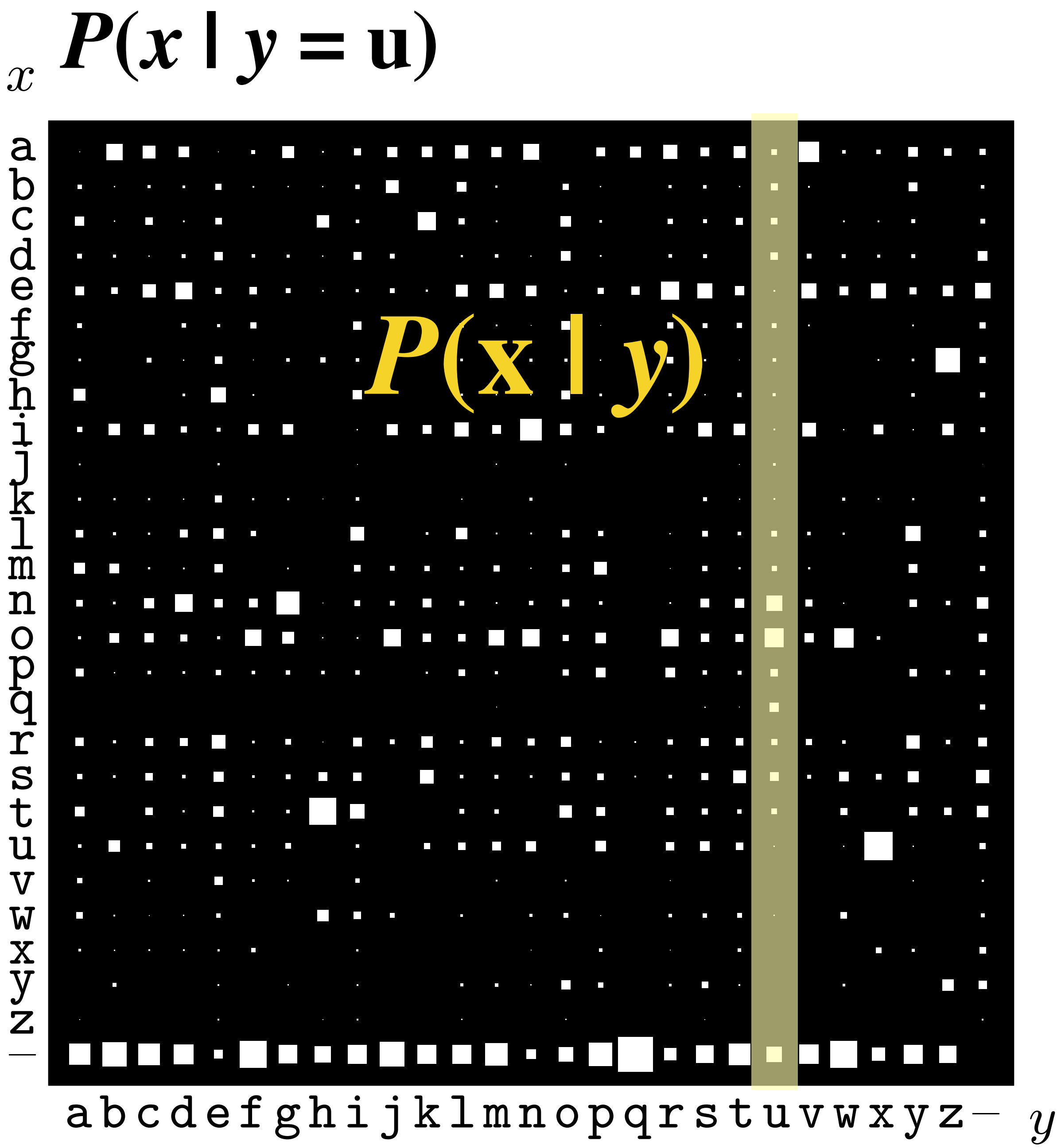
Condicional Probability: Example

- For each y, $P(x \mid y)$ is a probability distribution.



Condicional Probability: Example

- For each y, $P(x \mid y)$ is a probability distribution.

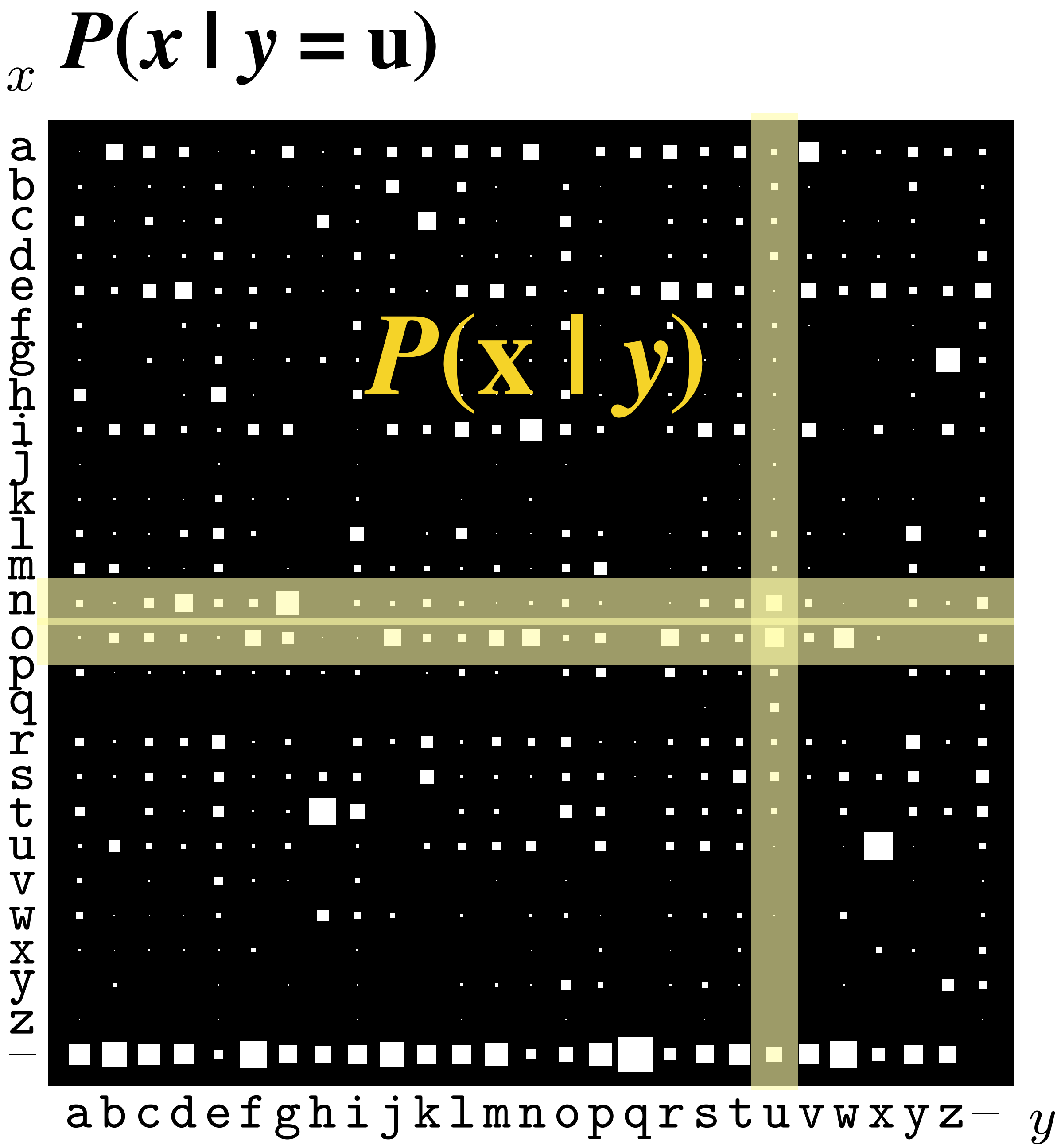


- The two most probable values for the first letter x given that the second letter y is **u** are **n** and **o**.

Condicional Probability: Example

- For each y, $P(x \mid y)$ is a probability distribution.

- The two most probable values for the first letter x given that the second letter y is **u** are **n** and **o**.



Product and Sum rules

- $P(.\,|\,\boldsymbol{H})$ - $\boldsymbol{H}$ denotes assumptions on which the probabilities are based.

Product and Sum rules

- $P(. \mid \boldsymbol{H})$ - $\boldsymbol{H}$ denotes assumptions on which the probabilities are based.
- **Product rule** (or chain rule)
  - $P(x,y \mid \boldsymbol{H}) = P(x \mid y, \boldsymbol{H})\ P(y \mid \boldsymbol{H}) = \mathrm{P}(y \mid x, \boldsymbol{H})\ P(x \mid \boldsymbol{H})$
  - Obtained from the definition of conditional probability

Product and Sum rules

- $P(. \mid \boldsymbol{H})$ - $\boldsymbol{H}$ denotes assumptions on which the probabilities are based.
- **Product rule** (or chain rule)
  - $P(x,y \mid \boldsymbol{H}) = P(y \mid \boldsymbol{H})\, P(x \mid y, \boldsymbol{H}) = P(x \mid \boldsymbol{H})\, \mathrm{P}(y \mid x, \boldsymbol{H})$
  - Obtained from the definition of conditional probability

Product and Sum rules

- $P(. \mid \boldsymbol{H})$ - $\boldsymbol{H}$ denotes assumptions on which the probabilities are based.
- **Product rule** (or chain rule)
  - $P(x,y \mid \boldsymbol{H}) = P(y \mid \boldsymbol{H})\, P(x \mid y, \boldsymbol{H}) = P(x \mid \boldsymbol{H})\, \mathrm{P}(y \mid x, \boldsymbol{H})$
  - Obtained from the definition of conditional probability
- **Sum rule**
  - A rewriting of the marginal probability definition
  - $P(x \mid \boldsymbol{H}) = \sum_y P(x,y \mid \boldsymbol{H})$
  -

Product and Sum rules

- $P(.\,|\,\boldsymbol{H})$ - $\boldsymbol{H}$ denotes assumptions on which the probabilities are based.
- **Product rule** (or chain rule)
  - $P(x,y\,|\,\boldsymbol{H}) = P(y\,|\,\boldsymbol{H})\,P(x\,|\,y,\boldsymbol{H}) = P(x\,|\,\boldsymbol{H})\,\mathrm{P}(y\,|\,x,\boldsymbol{H})$
  - Obtained from the definition of conditional probability
- **Sum rule**
  - A rewriting of the marginal probability definition
  - $P(x\,|\,\boldsymbol{H}) = \sum_{y} P(x,y\,|\,\boldsymbol{H})$
  - $P(x\,|\,\boldsymbol{H}) = \sum_{y} P(x\,|\,y,\boldsymbol{H})P(y\,|\,\boldsymbol{H})$

We can define an ensemble in terms of a **collection of conditional probabilities**

Bayes' Theorem

- Essential to integrate new pieces of evidence.

$$P(y \mid x, \boldsymbol{H}) = \frac{P(x \mid y, \boldsymbol{H}) P(y \mid \boldsymbol{H})}{P(x \mid \boldsymbol{H})}$$

Bayes' Theorem

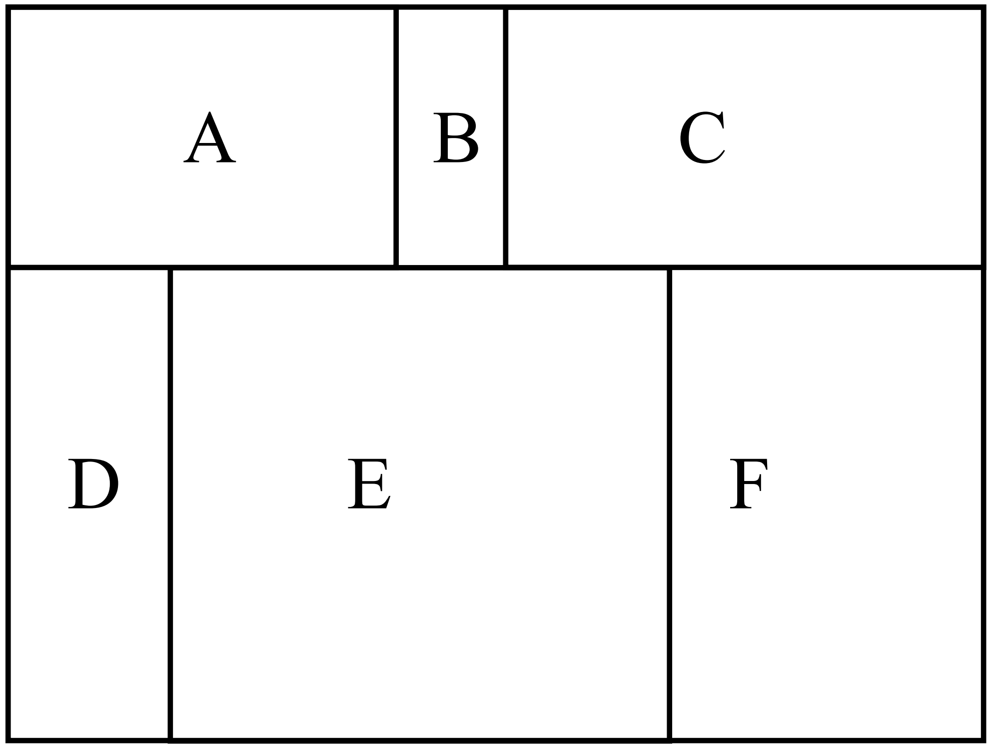
- Essential to integrate new pieces of evidence.

$$P(y \mid x, \boldsymbol{H}) = \frac{P(x \mid y, \boldsymbol{H}) P(y \mid \boldsymbol{H})}{P(x \mid \boldsymbol{H})}$$

$$P(y \mid x, \boldsymbol{H}) = \frac{P(x \mid y, \boldsymbol{H}) P(y \mid \boldsymbol{H})}{\sum_{y'} P(x \mid y', \boldsymbol{H}) P(y' \mid \boldsymbol{H})}$$

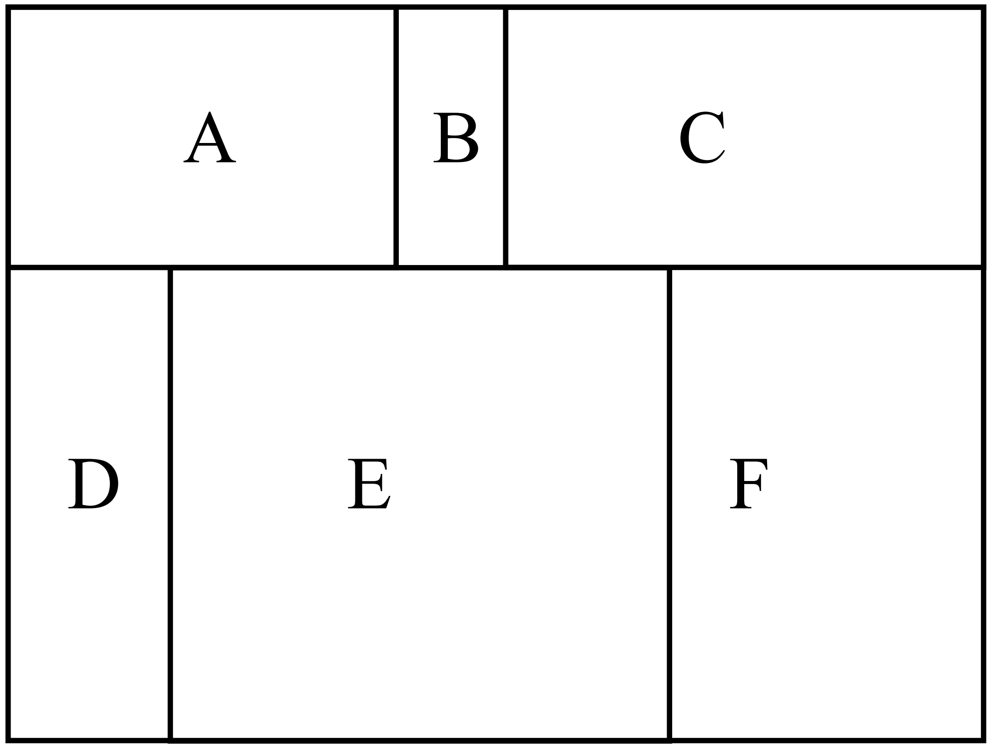
Bayes' Theorem: a geometric interpretation

- Consider this modified version of a Venn Diagram where the area of each event A, …, F is proportional to its probability



Bayes' Theorem: a geometric interpretation

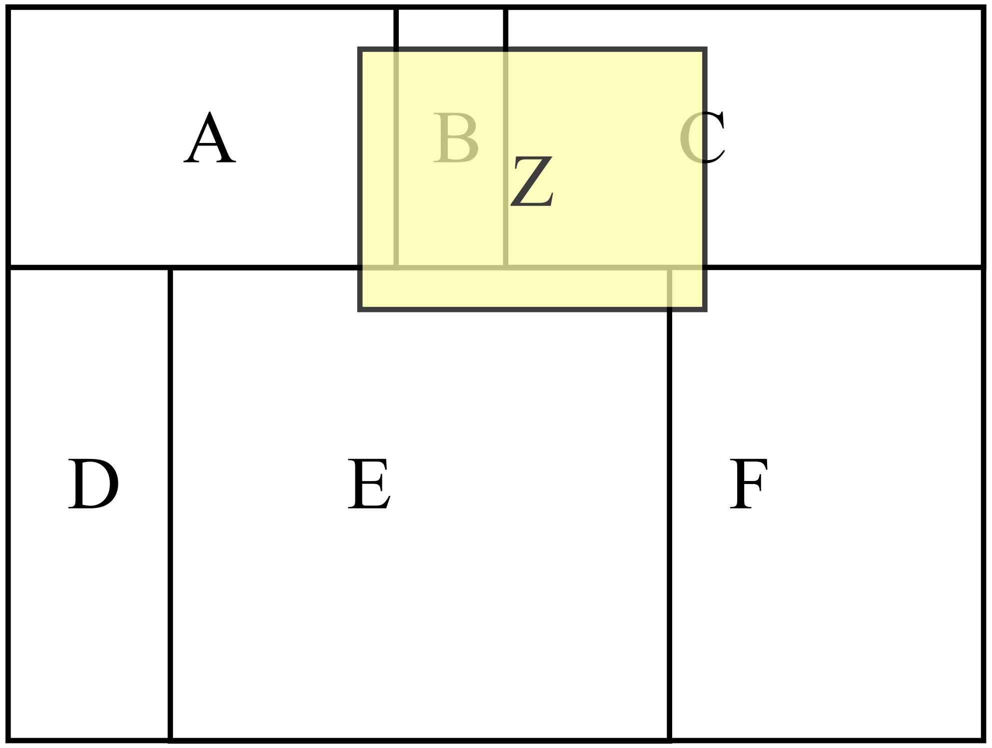
- Consider this modified version of a Venn Diagram where the area of each event A, …, F is proportional to its probability



- $P(\text{A}) + \ldots . P(\text{F}) = 1$
- P(E) is the most probable event
- P(B) is the less probable event

Bayes' Theorem: a geometric interpretation

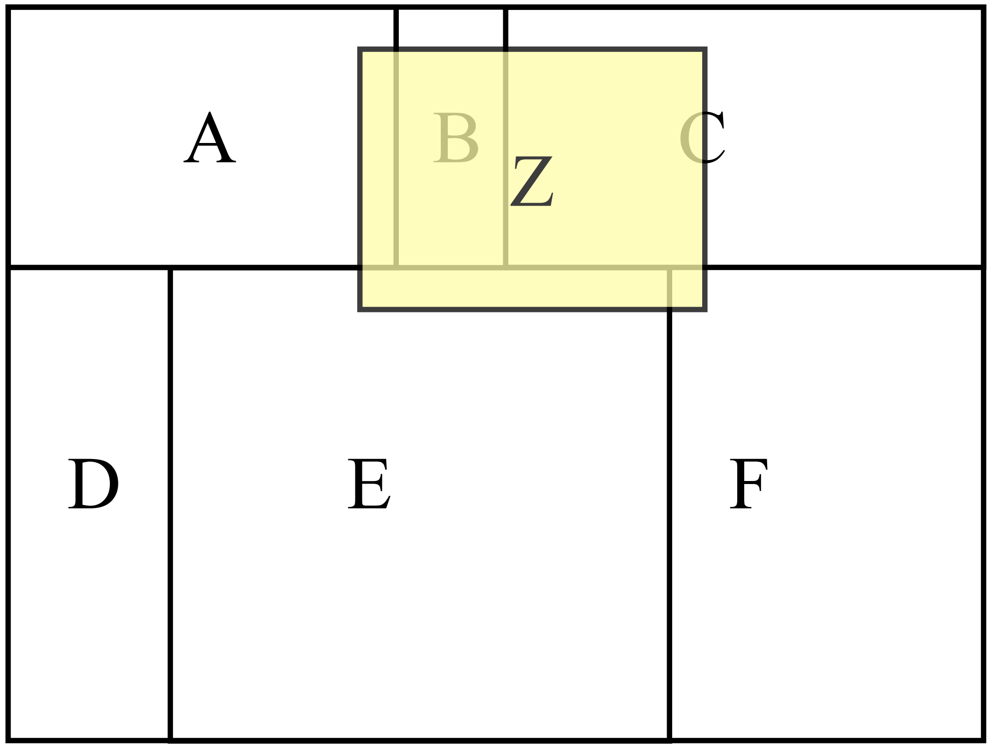
- Consider this modified version of a Venn Diagram where the area of each event A, …, F is proportional to its probability



- P(A) + …. P(F) = 1
- P(E) is the most probable event
- P(B) is the less probable event
- Now lets get a new evidence.
  - **Z happen !**

Bayes' Theorem: a geometric interpretation

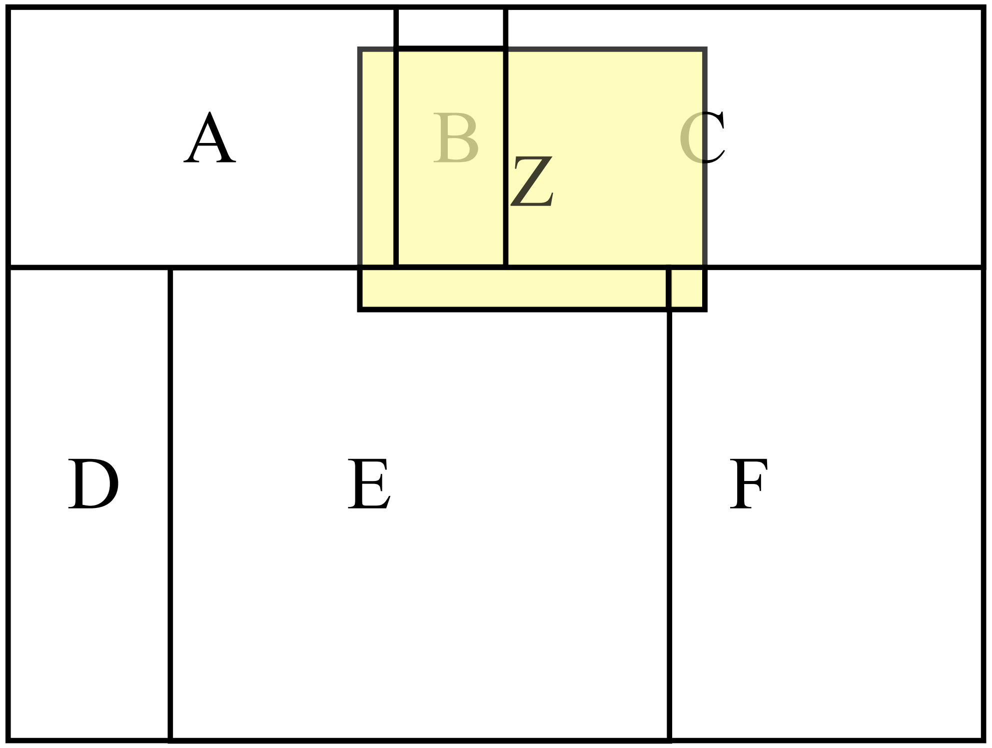
- Consider this modified version of a Venn Diagram where the area of each event A, …, F is proportional to its probability



- $P(A) + \ldots. P(F) = 1$
- P(E) is the most probable event
- P(B) is the less probable event
- Now lets get a new evidence.
  - **Z happen !**
- What are now the new probabilities?

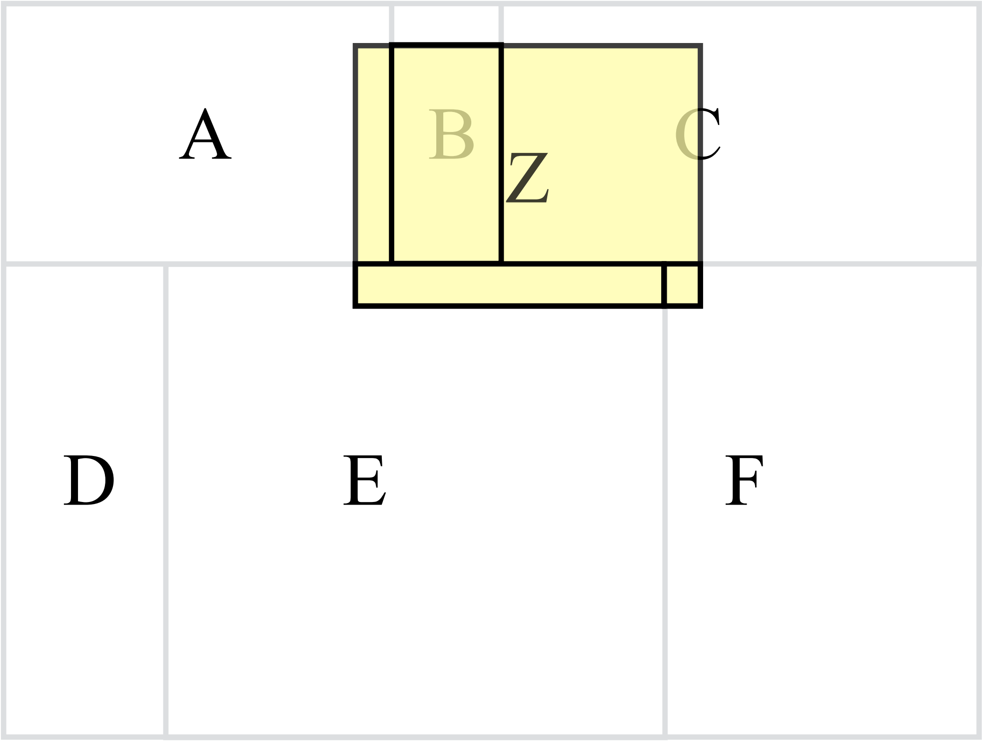
Bayes’ Theorem: a geometric interpretation

- If we now that Z happen, then $P(Z) = 1$
  - So nothing else outside Z matters !



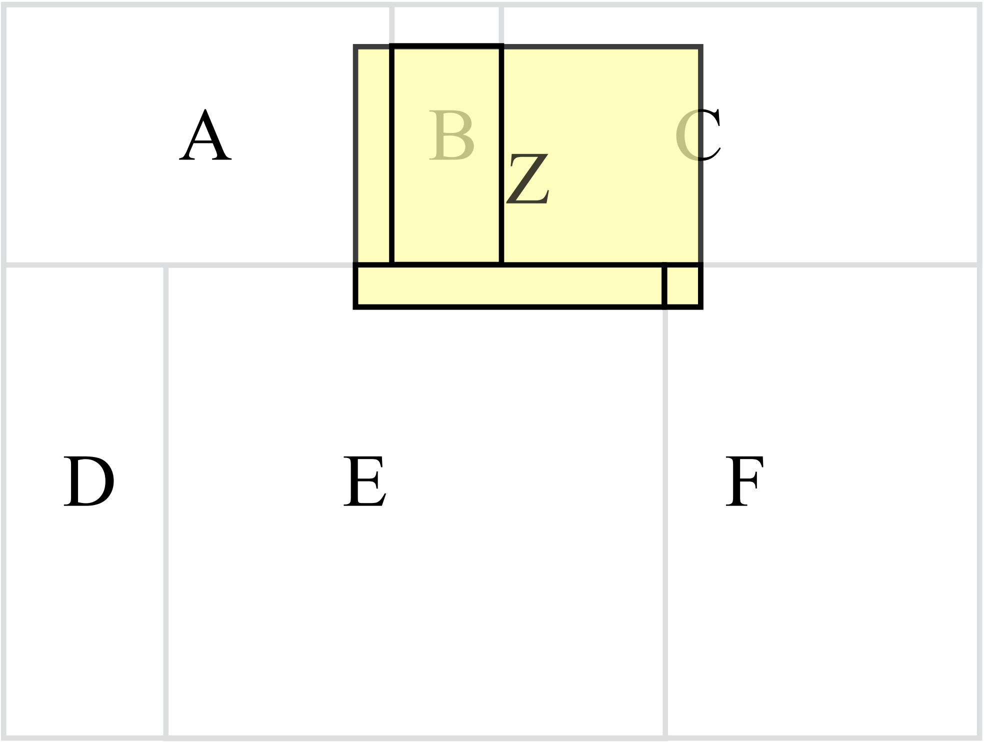
Bayes' Theorem: a geometric interpretation

- If we now that Z happen, then $P(Z) = 1$
  - So nothing else outside Z matters !



Bayes' Theorem: a geometric interpretation

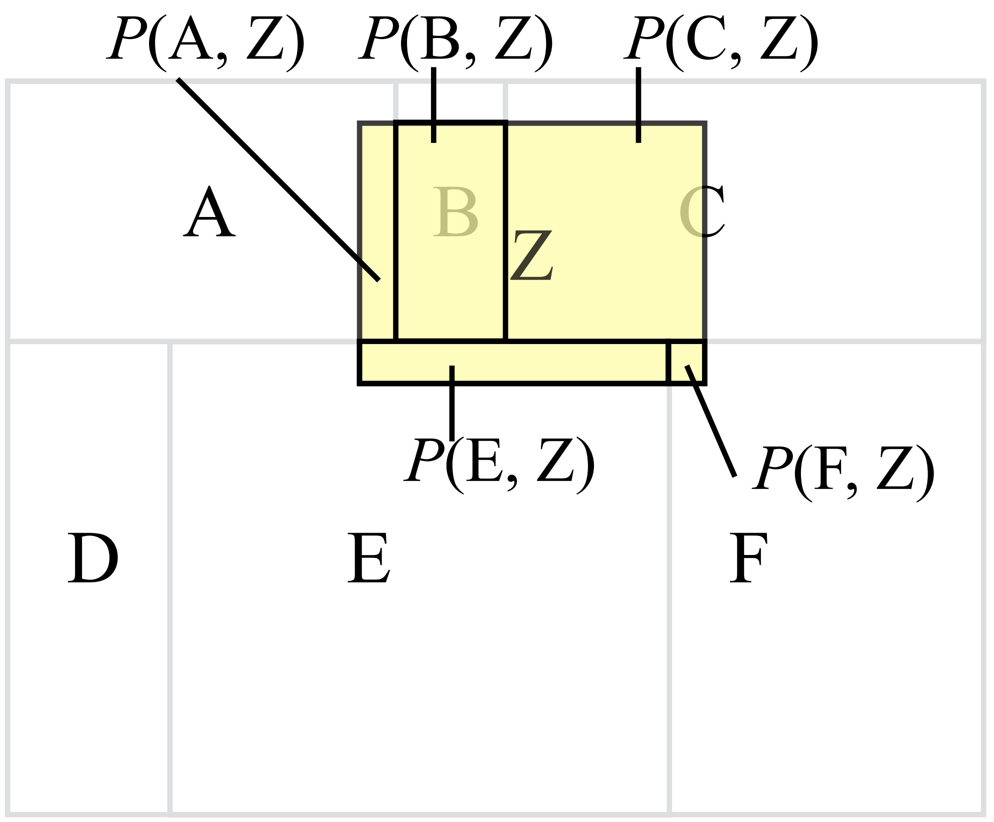
- If we now that Z happen, then *P*(Z) = 1
  - So nothing else outside Z matters !



- What are the probabilities A, …., F, that are inside Z, before Z happen?

Bayes' Theorem: a geometric interpretation

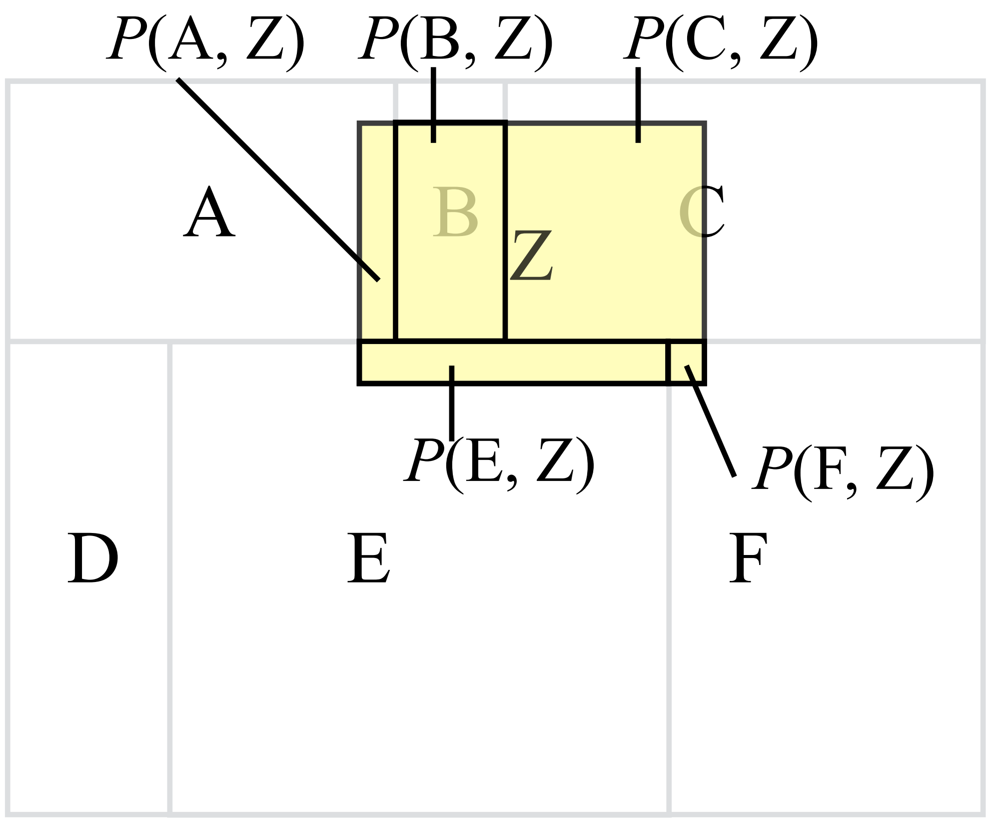
- If we now that Z happen, then $P(Z) = 1$
  - So nothing else outside Z matters !



- What are the probabilities A, …., F, that are inside Z, before Z happen?

Bayes' Theorem: a geometric interpretation

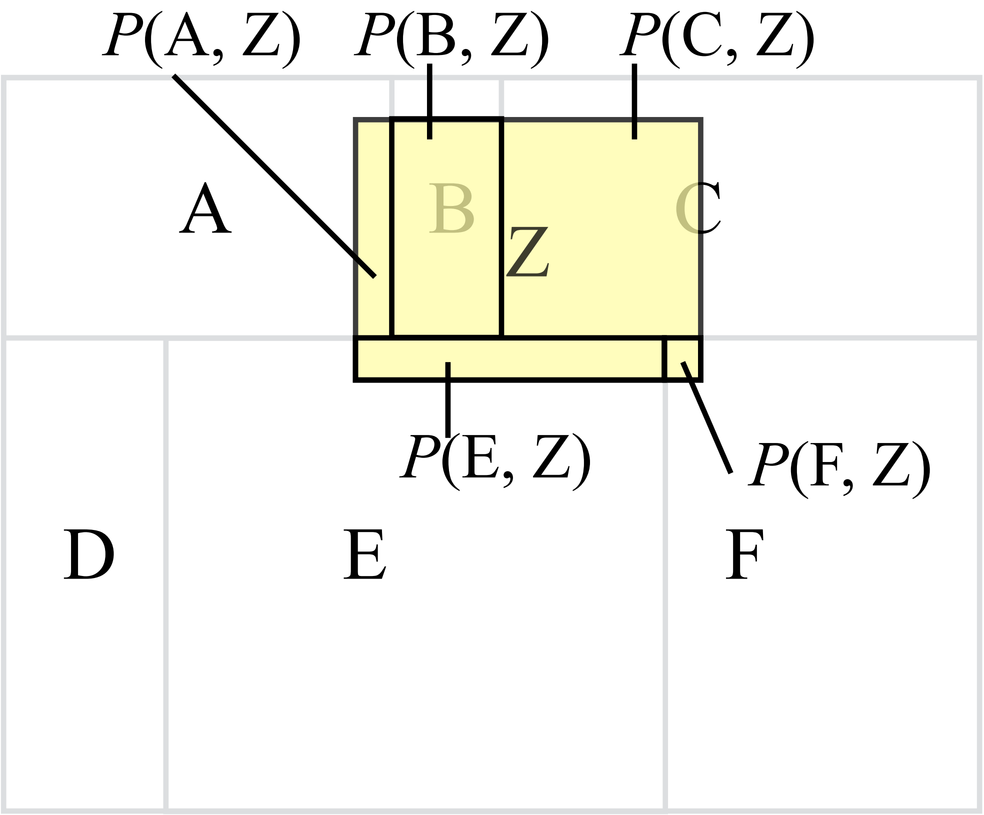
- If we now that Z happen, then *P*(Z) = 1
  - So nothing else outside Z matters !



- What are the probabilities A, …., F, that are inside Z, before Z happen?
- Notice that ∑P(_, Z) = P(Z)

Bayes' Theorem: a geometric interpretation

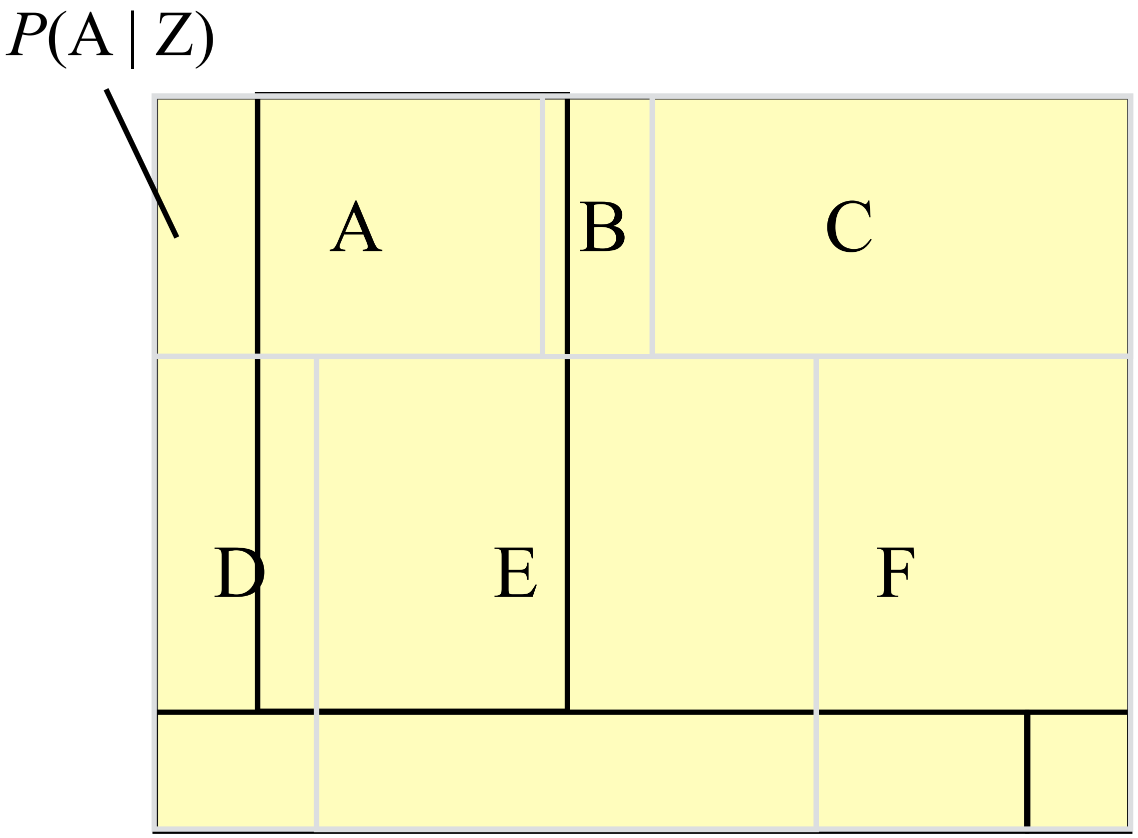
- If we now that Z happen, then *P*(Z) = 1
  - So nothing else outside Z matters !



- What are the probabilities A, …., F, that are inside Z, before Z happen?
- Notice that ∑P(_, Z) = P(Z)
- Now, We just have to scale those probabilities such that P(Z) = 1, i.e, such that P(A, Z) + P(B, Z) + P(C, Z) + P(E, F) + P(F, Z) = 1

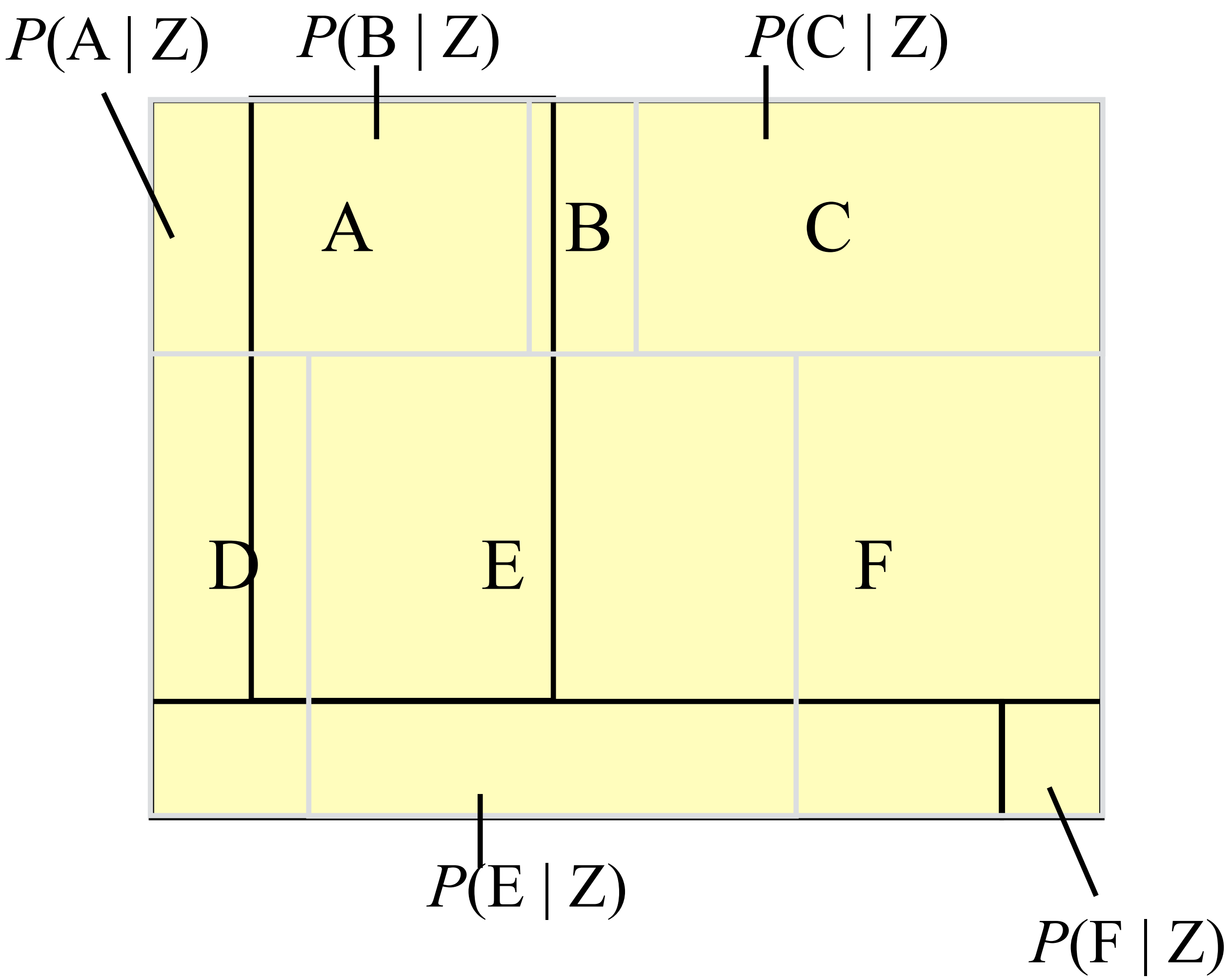
Bayes' Theorem: a geometric interpretation

- If we now that Z happen, then $P(Z) = 1$
  - So nothing else outside Z matters !



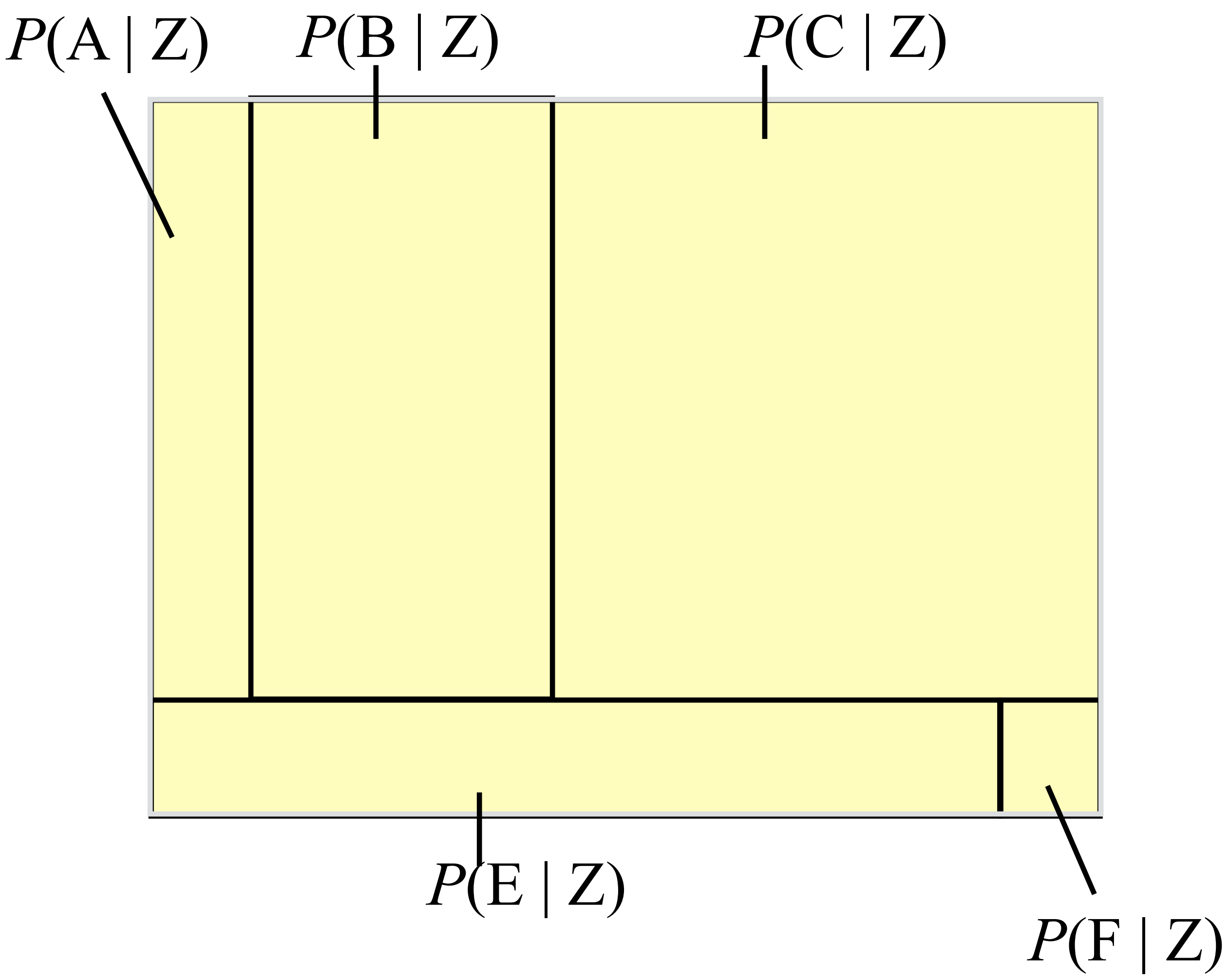
Bayes' Theorem: a geometric interpretation

- If we now that Z happen, then $P(Z) = 1$
  - So nothing else outside Z matters !



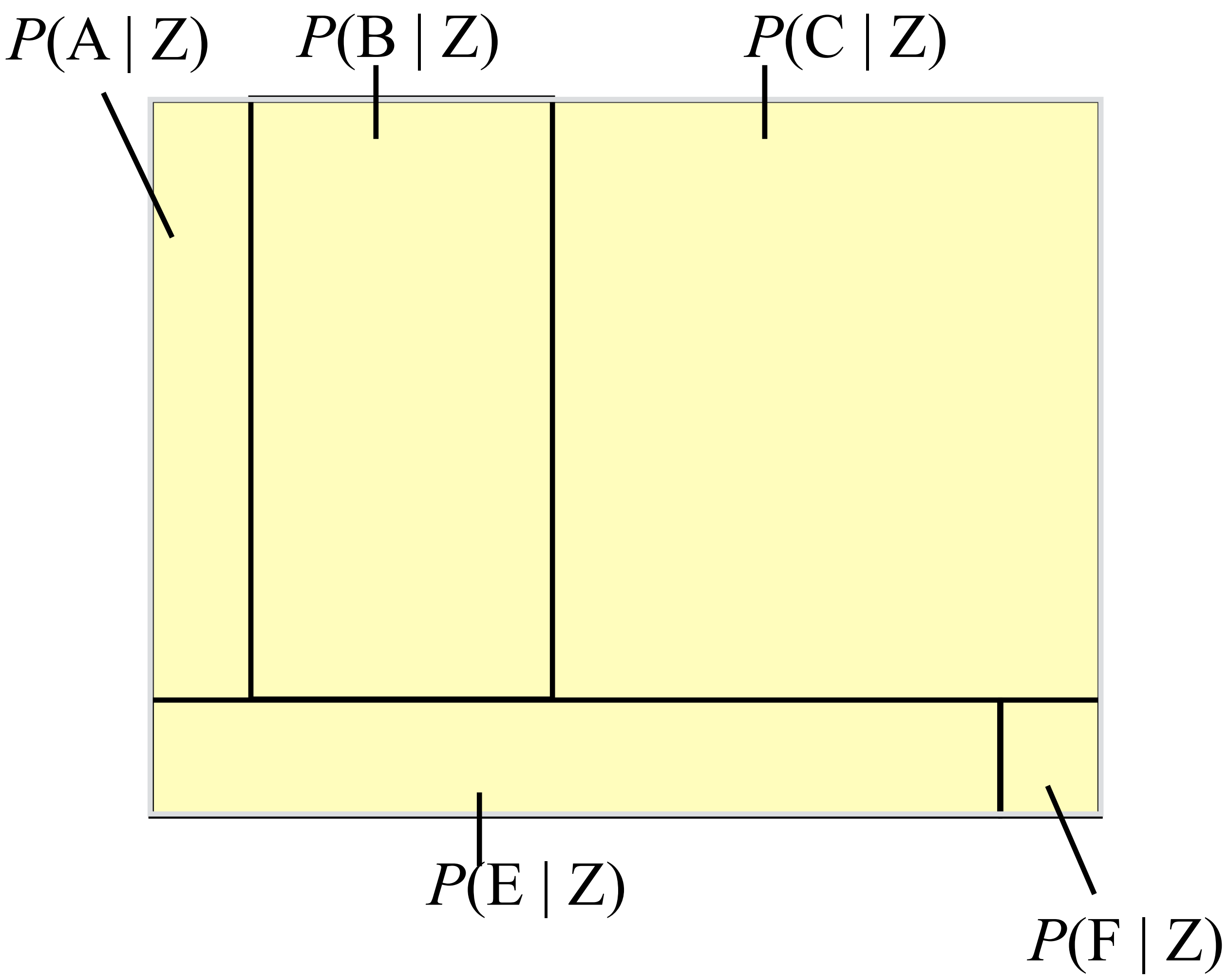
Bayes' Theorem: a geometric interpretation

- If we now that Z happen, then $P(Z) = 1$
  - So nothing else outside Z matters !



Bayes' Theorem: a geometric interpretation

- If we now that Z happen, then $P(Z) = 1$
  - So nothing else outside Z matters !



- P(A, Z) -> $P(A \mid Z) = P(A, Z) / P(Z)$
- P(B, Z) -> $P(B \mid Z) = P(B, Z) / P(Z)$
- P(C, Z) -> $P(C \mid Z) = P(C, Z) / P(Z)$
- P(E, Z) -> $P(E \mid Z) = P(E, Z) / P(Z)$
- P(F, Z) -> $P(F \mid Z) = P(F, Z) / P(Z)$
- $P(D \mid Z) = 0$

FACULTADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Bayes' Theorem

- Essencial para integrar novas evidências.

$$P(y \mid x, \boldsymbol{H}) = \frac{P(x \mid y, \boldsymbol{H}) P(y \mid \boldsymbol{H})}{P(x \mid \boldsymbol{H})}$$

Bayes' Theorem

- Essencial para integrar novas evidências.

$$P(y \mid x, \boldsymbol{H}) = \frac{P(x \mid y, \boldsymbol{H}) P(y \mid \boldsymbol{H})}{P(x \mid \boldsymbol{H})}$$

$$P(y \mid x, \boldsymbol{H}) = \frac{P(x \mid y, \boldsymbol{H}) P(y \mid \boldsymbol{H})}{\sum_{y'} P(x \mid y', \boldsymbol{H}) P(y' \mid \boldsymbol{H})}$$

An example: the presentation

- Jo has a test for a **nasty disease**.

An example: the presentation

- Jo has a test for a **nasty disease**.
  - Jo’s **state of health** by the variable a,

An example: the presentation

- Jo has a test for a **nasty disease**.
  - Jo's **state of health** by the variable *a*,

a = 1	Jo has the disease
a = 0	Jo does not have the disease

An example: the presentation

- Jo has a test for a **nasty disease**.
  - Jo's **state of health** by the variable *a*,
  - The **test result** by *b*.

a = 1	Jo has the disease
a = 0	Jo does not have the disease

An example: the presentation

- Jo has a test for a **nasty disease**.
  - Jo's **state of health** by the variable *a*,
  - The **test result** by *b*.

a = 1	Jo has the disease
a = 0	Jo does not have the disease

- The **result of the test** is either '**positive**' (b = 1) or '**negative**' (b = 0)

An example: the presentation

- Jo has a test for a **nasty disease**.
  - Jo’s **state of health** by the variable *a*,
  - The **test result** by *b*.

a = 1	Jo has the disease
a = 0	Jo does not have the disease

- The **result of the test** is either ‘**positive**’ (b = 1) or ‘**negative**’ (b = 0)
  - The test is 95% reliable:

An example: the presentation

- Jo has a test for a **nasty disease**.
  - Jo's **state of health** by the variable *a*,
  - The **test result** by *b*.

a = 1	Jo has the disease
a = 0	Jo does not have the disease

- The **result of the test** is either '**positive**' (b = 1) or '**negative**' (b = 0)
  - The test is 95% reliable:
    - in 95% of cases of people who really have the disease, a positive result is returned;
    - in 95% of cases of people who do not have the disease, a negative result is obtained.

FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE NOVA DE LISBOA

An example: the presentation

- Jo has a test for a **nasty disease**.
  - Jo's **state of health** by the variable *a*,
  - The **test result** by *b*.

a = 1	Jo has the disease
a = 0	Jo does not have the disease

- The **result of the test** is either '**positive**' (b = 1) or '**negative**' (b = 0)
  - The test is 95% reliable:
    - in 95% of cases of people who really have the disease, a positive result is returned;
    - in 95% of cases of people who do not have the disease, a negative result is obtained.
- Background information:
  - 1% of people of Jo's age and background have the disease.

An example: the presentation

- Jo has a test for a **nasty disease**.
  - Jo's **state of health** by the variable *a*,
  - The **test result** by *b*.

a = 1	Jo has the disease
a = 0	Jo does not have the disease

- The **result of the test** is either '**positive**' (b = 1) or '**negative**' (b = 0)
  - The test is 95% reliable:
    - in 95% of cases of people who really have the disease, a positive result is returned;
    - in 95% of cases of people who do not have the disease, a negative result is obtained.
- Background information:
  - 1% of people of Jo's age and background have the disease.
- **Jo has the test, and the result is positive. What is the probability that Jo has the disease?**

An example: capture the probabilities

- Variables

An example: capture the probabilities

- Variables

Jo’s state of the health	
$a = 1$	Jo has the disease
$a = 0$	Jo does not have the disease

An example: capture the probabilities

- Variables

Jo’s state of the health	
$a = 1$	Jo has the disease
$a = 0$	Jo does not have the disease

Result of the test	
$b = 1$	result of the test is positive
$b = 0$	result of the test is negative

An example: capture the probabilities

- Variables

Jo's state of the health	
$\boldsymbol{a} = \mathbf{1}$	Jo has the disease
$\boldsymbol{a} = \mathbf{0}$	Jo does not have the disease

Result of the test	
$\boldsymbol{b} = \mathbf{1}$	result of the test is positive
$\boldsymbol{b} = \mathbf{0}$	result of the test is negative

- The test reliability specifies the conditional probability of b given a:

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

An example: capture the probabilities

- Variables

Jo's state of the health	
$a = 1$	Jo has the disease
$a = 0$	Jo does not have the disease

Result of the test	
$b = 1$	result of the test is positive
$b = 0$	result of the test is negative

- The test reliability specifies the conditional probability of b given a:
  - The test is 95% reliable:
    - in 95% of cases of **people who really have the disease**, a **positive result is returned**;
    - in 95% of cases of **people who do not have the disease**, a **negative result is obtained**.

An example: capture the probabilities

- Variables

Jo's state of the health	
a = 1	Jo has the disease
a = 0	Jo does not have the disease

Result of the test	
b = 1	result of the test is positive
b = 0	result of the test is negative

- The test reliability specifies the conditional probability of b given a:
  - The test is 95% reliable:
    - in 95% of cases of **people who really have the disease**, a **positive result is returned**;
    - in 95% of cases of **people who do not have the disease**, a **negative result is obtained**.

$$\begin{aligned} P(b=1\,|\,a=1) &= 0.95 \qquad & P(b=1\,|\,a=0) &= 0.05 \\ P(b=0\,|\,a=1) &= 0.05 \qquad & P(b=0\,|\,a=0) &= 0.95; \end{aligned}$$

An example: capture the probabilities

- Variables

Jo's state of the health	
$a = 1$	Jo has the disease
$a = 0$	Jo does not have the disease

Result of the test	
$b = 1$	result of the test is positive
$b = 0$	result of the test is negative

- The test reliability specifies the conditional probability of b given a:

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

$$
\begin{array}{ll}
P(b\!=\!1 \mid a\!=\!1) = 0.95 & P(b\!=\!1 \mid a\!=\!0) = 0.05 \\
P(b\!=\!0 \mid a\!=\!1) = 0.05 & P(b\!=\!0 \mid a\!=\!0) = 0.95;
\end{array}
$$

An example: capture the probabilities

- Variables

Jo's state of the health	
***a* = 1**	Jo has the disease
***a* = 0**	Jo does not have the disease

Result of the test	
***b* = 1**	result of the test is positive
***b* = 0**	result of the test is negative

- The test reliability specifies the conditional probability of b given a:

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

$$\begin{array}{llll} P(b=1 \mid a=1) = 0.95 & & P(b=1 \mid a=0) = 0.05 \\ P(b=0 \mid a=1) = 0.05 & & P(b=0 \mid a=0) = 0.95; \end{array}$$

1% of people of Jo's age and background have the disease.

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

An example: capture the probabilities

- Variables

Jo's state of the health	
***a* = 1**	Jo has the disease
***a* = 0**	Jo does not have the disease

Result of the test	
***b* = 1**	result of the test is positive
***b* = 0**	result of the test is negative

- The test reliability specifies the conditional probability of b given a:

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

$$P(b=1 \mid a=1) = 0.95 \qquad P(b=1 \mid a=0) = 0.05$$
$$P(b=0 \mid a=1) = 0.05 \qquad P(b=0 \mid a=0) = 0.95;$$

- The disease prevalence tells us about the marginal probability of a:
  - $P(a = 1) = 0.01$
  - $P(a = 0) = 0.99$

1% of people of Jo's age and background have the disease.

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

An example: compute the joint probabilities

- Variables (*a* - Jo's state of the health; *b* - Result of the test)

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

An example: compute the joint probabilities

- Variables (*a* - Jo's state of the health; *b* - Result of the test)

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

$$P(a, b) = P(b \mid a)\ P(a)$$

An example: compute the joint probabilities

- Variables (a - Jo's state of the health; b - Result of the test)

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

$$P(a, b) = P(b \mid a)\, P(a)$$

$P(a, \mathrm{b})$	$a = 1$	$a = 0$
$b = 1$	**0.0095**	**0.0495**
$b = 0$	**0.0005**	**0.9405**

An example: compute the joint probabilities

- Variables (*a* - Jo's state of the health; *b* - Result of the test)

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

$$P(a, b) = P(b \mid a)\, P(a)$$

$$P(b) = \sum_{a} P(a,b)$$

$P(a, \mathrm{b})$	$a = 1$	$a = 0$
$b = 1$	**0.0095**	**0.0495**
$b = 0$	**0.0005**	**0.9405**

An example: compute the joint probabilities

- Variables (a - Jo’s state of the health; b - Result of the test)

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

$$P(a, b) = P(b \mid a)\, P(a)$$

$P(a, \text{b})$	$a = 1$	$a = 0$
$b = 1$	**0.0095**	**0.0495**
$b = 0$	**0.0005**	**0.9405**

$$P(b) = \sum_a P(a,b)$$

$P(b = 1)$	**0.059**
$P(b = 0)$	**0.941**

An example: compute the joint probabilities … etc.

- Variables (*a* - Jo’s state of the health; *b* - Result of the test)

P(*b* | *a*)	*a* = 1	*a* = 0
b = 1	**0.95**	**0.05**
b = 0	**0.05**	**0.95**

P(*a* = 1)	*P*(*a* = 0)
0.01	**0.99**

$$P(a, b) = P(b \mid a)\, P(a)$$

P(*a*, b)	*a* = 1	*a* = 0
b = 1	**0.0095**	**0.0495**
b = 0	**0.0005**	**0.9405**

$$P(b) = \sum_{a} P(a,b)$$

P(*b* = 1)	**0.059**
P(*b* = 0)	**0.941**

$$P(a \mid b) = P(a, b) \,/\, P(b)$$

An example: compute the joint probabilities … etc.

- Variables (*a* - Jo's state of the health; *b* - Result of the test)

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

$$P(a, b) = P(b \mid a)\, P(a)$$

$$P(b) = \sum_a P(a,b)$$

$P(a, b)$	$a = 1$	$a = 0$
$b = 1$	**0.0095**	**0.0495**
$b = 0$	**0.0005**	**0.9405**

$P(b = 1)$	**0.059**
$P(b = 0)$	**0.941**

$P(a \mid b)$	$a = 1$	$a = 0$
$b = 1$	**0.161**	**0.839**
$b = 0$	**0.001**	**0.999**

$$P(a \mid b) = P(a, b) \,/\, P(b)$$

An example: answer the question

- **Jo has the test, and the result is positive**. **What is the probability that Jo has the disease?**
  - We are looking for $P(a = 1 \mid b = 1)$

$$P(a=1 \mid b=1) = \frac{P(a=1, b=1)}{P(b=1)}$$

An example: answer the question

- **Jo has the test, and the result is positive**. **What is the probability that Jo has the disease?**
  - We are looking for $P(a = 1 \mid b = 1)$

$$P(a=1 \mid b=1)=\frac{P(a=1,b=1)}{P(b=1)}$$

$$P(a=1 \mid b=1)=\frac{P(b=1 \mid a=1)P(a=1)}{P(b=1,a=1)+P(b=1,a=0)}$$

An example: answer the question

- **Jo has the test, and the result is positive**. **What is the probability that Jo has the disease?**
  - We are looking for $P(a = 1 \mid b = 1)$

$$P(a=1 \mid b=1) = \frac{P(a=1,b=1)}{P(b=1)}$$

$$P(a=1 \mid b=1) = \frac{P(b=1 \mid a=1)P(a=1)}{P(b=1,a=1)+P(b=1,a=0)}$$

$$P(a=1 \mid b=1) = \frac{P(b=1 \mid a=1)P(a=1)}{P(b=1 \mid a=1)P(a=1)+P(b=1 \mid a=0)P(a=0)}$$

An example: answer the question

- **Jo has the test, and the result is positive**. **What is the probability that Jo has the disease?**
  - We are looking for $P(a = 1 \mid b = 1)$

$$P(a=1 \mid b=1) = \frac{P(a=1, b=1)}{P(b=1)}$$

$$P(a=1 \mid b=1) = \frac{P(b=1 \mid a=1)P(a=1)}{P(b=1, a=1) + P(b=1, a=0)}$$

$$P(a=1 \mid b=1) = \frac{P(b=1 \mid a=1)P(a=1)}{P(b=1 \mid a=1)P(a=1) + P(b=1 \mid a=0)P(a=0)}$$

$$P(a=1 \mid b=1) = \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.05 \times 0.99} = 0.16$$

An example: answer the question

- **Jo has the test, and the result is positive**. **What is the probability that Jo has the disease?**
  - We are looking for $P(a = 1 \mid b = 1)$

$$P(a=1 \mid b=1) = \frac{P(a=1, b=1)}{P(b=1)}$$

$$P(a=1 \mid b=1) = \frac{P(b=1 \mid a=1)P(a=1)}{P(b=1, a=1) + P(b=1, a=0)}$$

$$P(a=1 \mid b=1) = \frac{P(b=1 \mid a=1)P(a=1)}{P(b=1 \mid a=1)P(a=1) + P(b=1 \mid a=0)P(a=0)}$$

$$P(a=1 \mid b=1) = \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.05 \times 0.99} = 0.16$$

$$P(a=1) = 0.01$$

Statistical Independence

- Two random variables *X* and *Y* are independent if and only if

$$P(x,y) = p(x)p(y)$$

Statistical Independence

- Two random variables *X* and *Y* are independent if and only if

$$P(x,y) = p(x)p(y)$$

- In general:

$$P(x,y) = p(x \mid y)p(y) \qquad P(x,y) = p(y \mid x)p(x)$$

Statistical Independence

- Two random variables *X* and *Y* are independent if and only if

$$P(x,y) = p(x)p(y)$$

- In general:

$$P(x,y) = p(x \mid y)p(y) \qquad P(x,y) = p(y \mid x)p(x)$$

- When *X* and *Y* are independent

$$P(x \mid y) = p(x) \qquad P(y \mid x) = p(y)$$

Statistical Independence

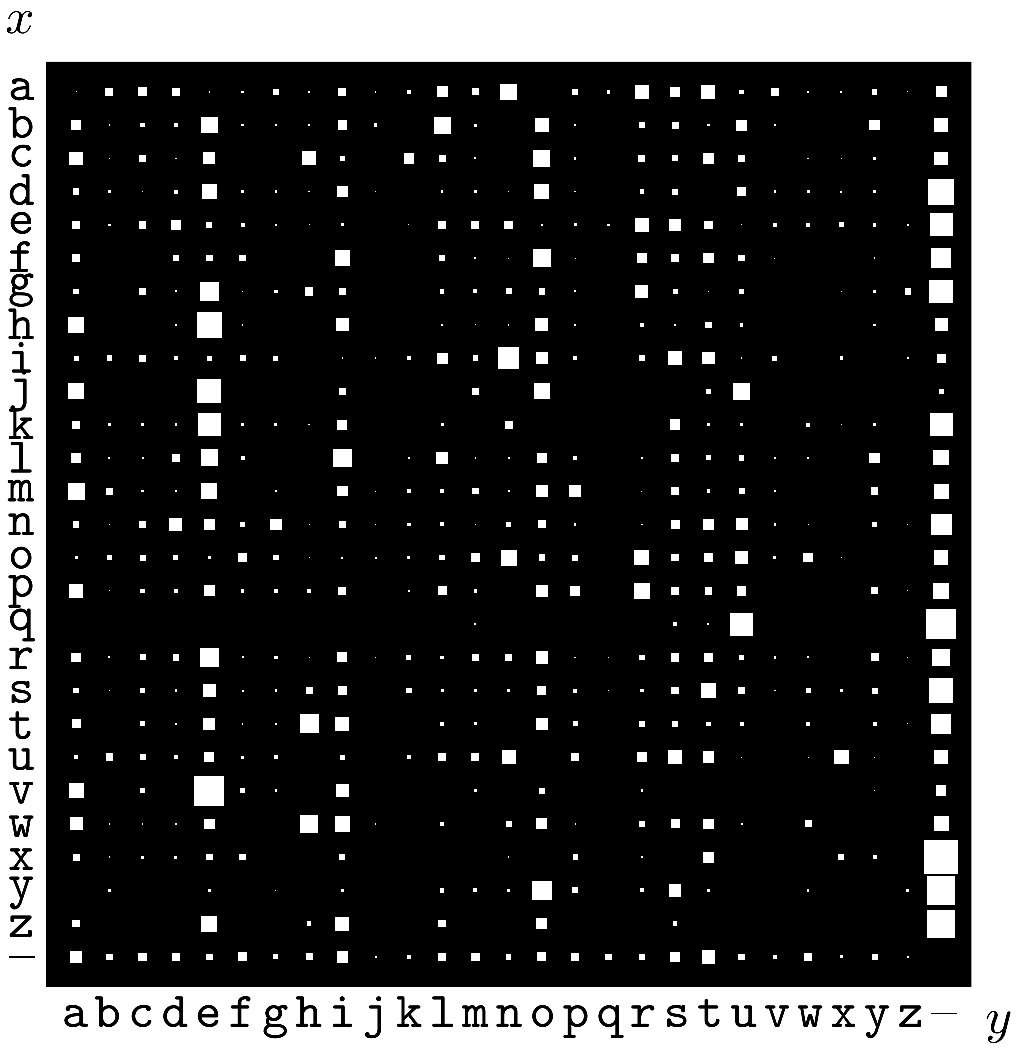
- When *X* and *Y* are independent

$$P(x \mid y) = p(x)$$

$$P(y \mid x) = p(y)$$

Are they independent?

$$P(\mathbf{y} \mid x)$$



The Meaning of Probabilities

The Meaning of Probabilities

- Probabilities can describe **frequencies of outcomes in random experiments.**

The Meaning of Probabilities

- Probabilities can describe **frequencies of outcomes in random experiments.**
- Probabilities can also be used to describe **degrees of belief in propositions** that do not involve random variables.

The Meaning of Probabilities

- Probabilities can describe **frequencies of outcomes in random experiments.**
- Probabilities can also be used to describe **degrees of belief in propositions** that do not involve random variables.

- Degrees of belief can be mapped onto probabilities if they satisfy simple consistency rules:
  - **Axiom 1**: Degrees of belief can be ordered; if $B(x)$ is ‘greater’ than $B(y)$, and $B(y)$ is ‘greater’ than $B(z)$, then $B(x)$ is ‘greater’ than $B(z)$.

The Meaning of Probabilities

- Probabilities can describe **frequencies of outcomes in random experiments.**
- Probabilities can also be used to describe **degrees of belief in propositions** that do not involve random variables.

- Degrees of belief can be mapped onto probabilities if they satisfy simple consistency rules:
  - **Axiom 1**: Degrees of belief can be ordered; if $B(x)$ is 'greater' than $B(y)$, and $B(y)$ is 'greater' than $B(z)$, then $B(x)$ is 'greater' than $B(z)$.
  - **Axiom 2**: There is a function f such that $B(x) = f[\ B(\sim x)\ ]$

The Meaning of Probabilities

- Probabilities can describe **frequencies of outcomes in random experiments.**
- Probabilities can also be used to describe **degrees of belief in propositions** that do not involve random variables.

- Degrees of belief can be mapped onto probabilities if they satisfy simple consistency rules:
  - **Axiom 1**: Degrees of belief can be ordered; if $B(x)$ is 'greater' than $B(y)$, and $B(y)$ is 'greater' than $B(z)$, then $B(x)$ is 'greater' than $B(z)$.
  - **Axiom 2**: There is a function f such that $B(x) = f[\ B(\sim x)\ ]$
  - **Axiom 3**: There is a function g such that $B(x, y) = g[\ B(x|y), \mathrm{B}(y)\ ]$

Forward Probabilities and Inverse Probabilities

Forward Probabilities and Inverse Probabilities

- Probability calculations often fall into one of two categories:
  - Forward probability
  - Inverse probability.

Forward Probabilities and Inverse Probabilities

- Probability calculations often fall into one of two categories:
  - Forward probability
  - Inverse probability.
- **Forward probability problems** involve a generative model that describes a process that is assumed to give rise to some data; The task is to **compute the probability distribution** or **expectation of some quantity** that **depends on the data**.

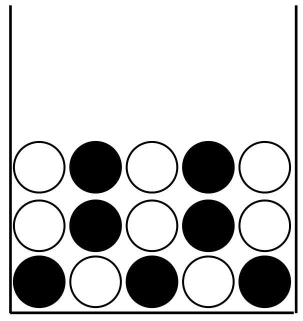
Forward Probabilities and Inverse Probabilities

- Probability calculations often fall into one of two categories:
  - Forward probability
  - Inverse probability.
- **Forward probability problems** involve a generative model that describes a process that is assumed to give rise to some data; The task is to **compute the probability distribution** or **expectation of some quantity** that **depends on the data**.
- **Inverse probability problems** involve a generative model of a process, but instead of computing the probability distribution of some quantity produced by the process, we **compute the conditional probability of one or more of the unobserved variables** in the process, given the observed variables. **This invariably requires the use of Bayes' theorem**

Forward Probability Example: *N* draws **with replacement**

Forward Probability Example: *N* draws **with replacement**

- An urn contains K balls, of which B are black and $W = K - B$ are white.



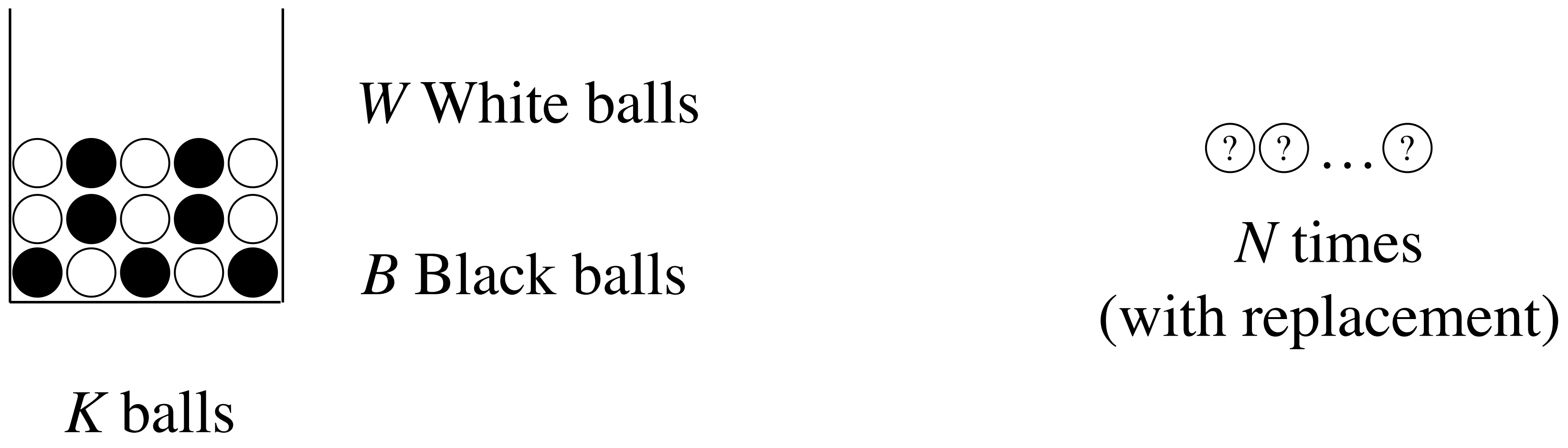
W White balls

B Black balls

K balls

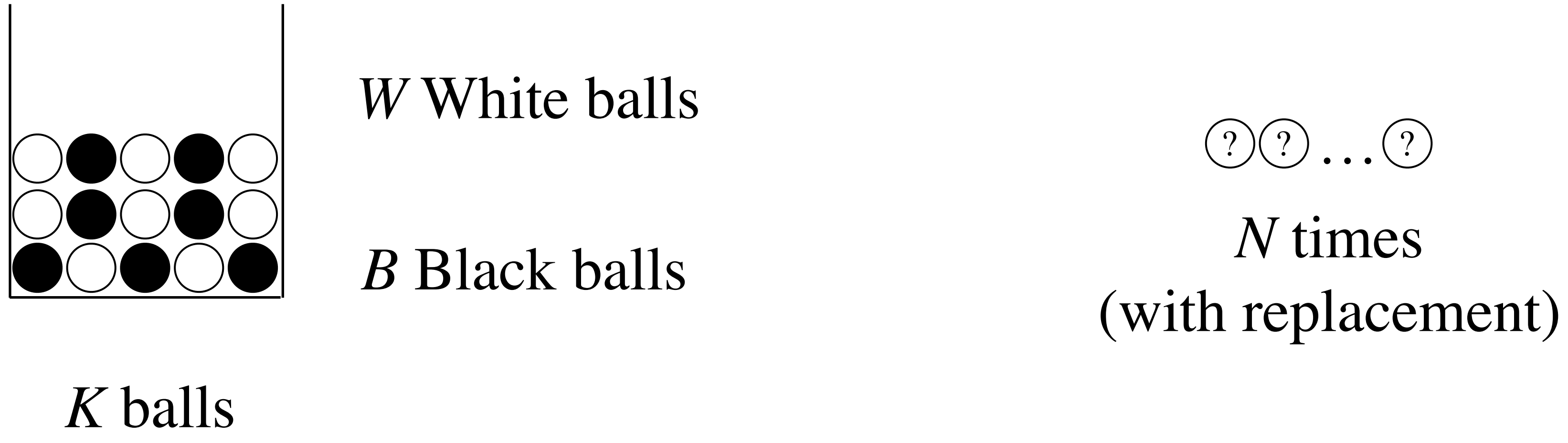
Forward Probability Example: *N* draws **with replacement**

- An urn contains K balls, of which B are black and $W = K - B$ are white.
- Fred draws a ball at random from the urn and replaces it, N times.



Forward Probability Example: *N* draws **with replacement**

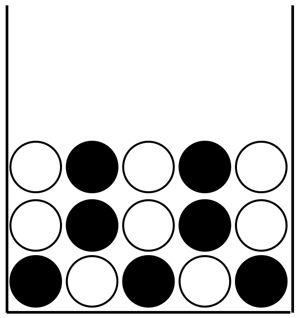
- An urn contains K balls, of which B are black and $W = K - B$ are white.
- Fred draws a ball at random from the urn and replaces it, N times.



- **What is the probability distribution** of the **number of times a black ball is drawn**, n_B?

Forward Probability Example: *N* draws **with replacement**

- An urn contains K balls, of which B are black and $W = K - B$ are white.
- Fred draws a ball at random from the urn and replaces it, N times.

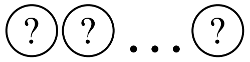


W White balls

B Black balls

K balls

$\boxed{f_B = B / K \text{ - the fraction of black balls}}$

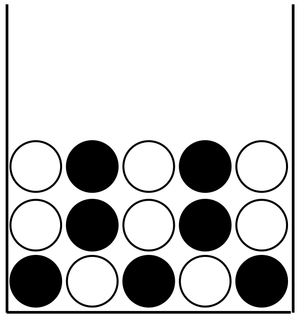


N times
(with replacement)

- **What is the probability distribution** of the **number of times a black ball is drawn**, n_B?

Forward Probability Example: *N* draws **with replacement**

- An urn contains K balls, of which B are black and $W = K - B$ are white.
- Fred draws a ball at random from the urn and replaces it, N times.

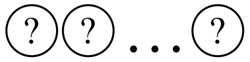


W White balls

B Black balls

K balls

$\boxed{f_B = B / K \text{ - the fraction of black balls}}$

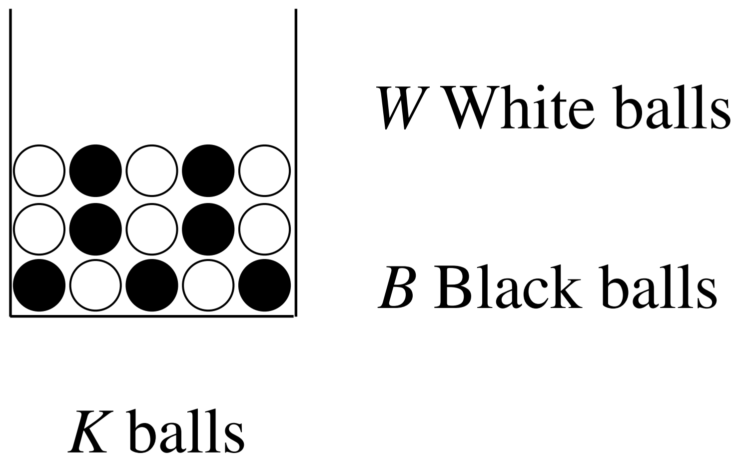


N times
(with replacement)

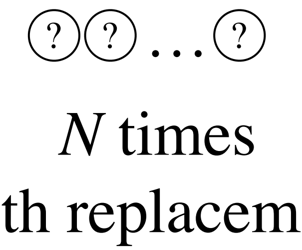
- **What is the probability distribution** of the **number of times a black ball is drawn**, n_B?
- n_B has a binomial distribution:

Forward Probability Example: *N* draws **with replacement**

- An urn contains *K* balls, of which *B* are black and *W* = *K* - *B* are white.
- Fred draws a ball at random from the urn and replaces it, *N* times.



$\boldsymbol{f_B} = B\,/\,K$ *- the fraction of black balls*



- **What is the probability distribution** of the **number of times a black ball is drawn**, n_B?
- n_B has a binomial distribution:

$$P(n_B \mid f_B, N) = \binom{N}{n_B} f_B^{\,n_B} (1 - f_B)^{N - n_B}$$

Binomial Distribution

- Lets *f* be the probability of one outcome of a random experiment. Let *r* be a random variable that represents the number of times the outcome occurs in *N* independent experiments.

$$P(r \mid f, N) = \binom{N}{r} f^r (1-f)^{N-r}$$

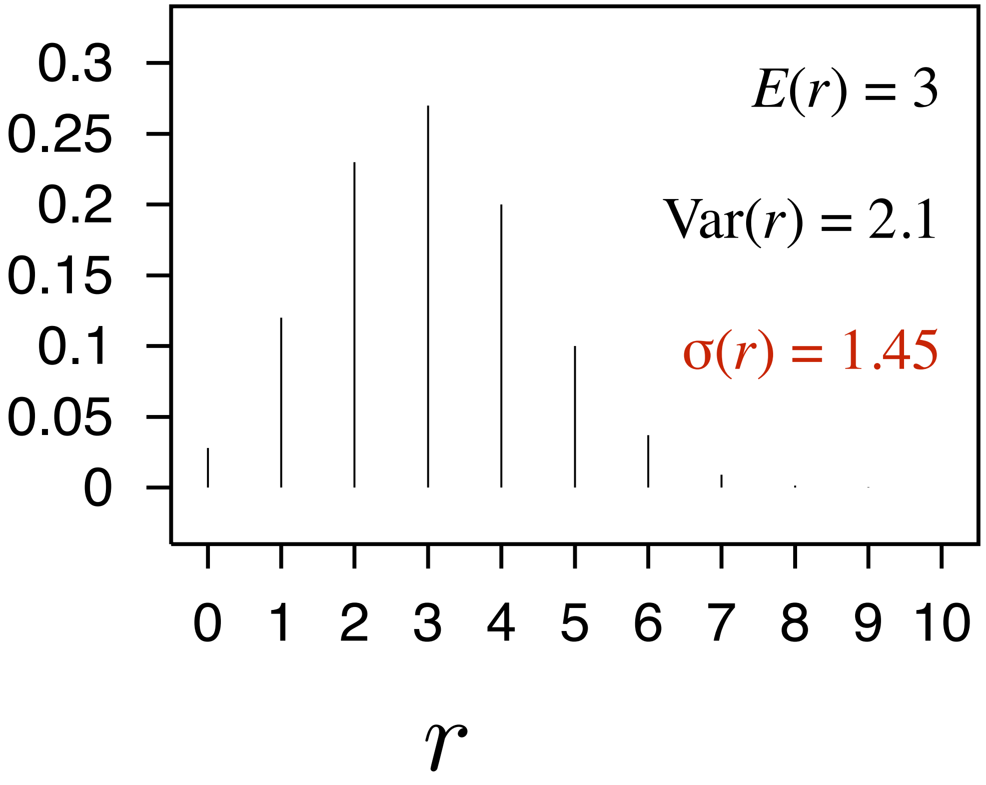
- The Mean

$$E(r) = \sum_{r=0}^{N} r.P(r \mid f, N) = Nf$$

- The Variance

$$Var(r) = E((r - E(r))^2) = Nf(1-f)$$

$f = 0.3, N = 10$



Forward Probability Example: *N* draws with replacement

- **What is the probability distribution** of the **number of times a black ball is drawn**, n_B?
- **What is the expectation** of n_B? What is the variance of n_B? What is the standard deviation of n_B?

$$E(n_B) = Nf_B \qquad Var(n_B) = Nf_B(1 - f_B) \qquad \sigma(n_B) = \sqrt{Nf_B(1 - f_B)}$$

Forward Probability Example: *N* draws with replacement

- **What is the probability distribution** of the **number of times a black ball is drawn**, n_B?
- **What is the expectation** of n_B? What is the variance of n_B? What is the standard deviation of n_B?

$$E(n_B) = Nf_B \qquad Var(n_B) = Nf_B(1 - f_B) \qquad \sigma(n_B) = \sqrt{Nf_B(1 - f_B)}$$

  - Give numerical answers for the cases $N = 5$ and $N = 400$, when $B = 2$ and $K = 10$.

Forward Probability Example: *N* draws with replacement

- **What is the probability distribution** of the **number of times a black ball is drawn**, n_B?
- **What is the expectation** of n_B? What is the variance of n_B? What is the standard deviation of n_B?

$$E(n_B) = Nf_B \qquad Var(n_B) = Nf_B(1 - f_B) \qquad \sigma(n_B) = \sqrt{Nf_B(1 - f_B)}$$

  - Give numerical answers for the cases $N = 5$ and $N = 400$, when $B = 2$ and $K = 10$.

$$f_B = B / K = 1/5$$

Forward Probability Example: *N* draws with replacement

- **What is the probability distribution** of the **number of times a black ball is drawn**, n_B?
- **What is the expectation** of n_B? What is the variance of n_B? What is the standard deviation of n_B?

$$E(n_B) = Nf_B \qquad Var(n_B) = Nf_B(1 - f_B) \qquad \sigma(n_B) = \sqrt{Nf_B(1 - f_B)}$$

  - Give numerical answers for the cases $N = 5$ and $N = 400$, when $B = 2$ and $K = 10$.

$$f_B = B / K = \tfrac{1}{5}$$

  - $N = 5$

$$E(n_B) = 1 \qquad Var(n_B) = \tfrac{4}{5} \qquad \sigma(n_B) = 0.89$$

Forward Probability Example: *N* draws with replacement

- **What is the probability distribution** of the **number of times a black ball is drawn**, n_B?
- **What is the expectation** of n_B? What is the variance of n_B? What is the standard deviation of n_B?

$$E(n_B) = Nf_B \qquad Var(n_B) = Nf_B(1 - f_B) \qquad \sigma(n_B) = \sqrt{Nf_B(1 - f_B)}$$

  - Give numerical answers for the cases $N = 5$ and $N = 400$, when $B = 2$ and $K = 10$.

$$f_B = B / K = 1/5$$

  - $N = 5$

$$E(n_B) = 1 \qquad Var(n_B) = 4/5 \qquad \sigma(n_B) = 0.89$$

  - $N = 400$

$$E(n_B) = 80 \qquad Var(n_B) = 64 \qquad \sigma(n_B) = 8$$

Inverse Probability Example

Inverse Probability Example

- There are eleven urns labelled by $u \in \{0, 1, 2, \ldots, 10\}$, each containing 10 balls.

Inverse Probability Example

- There are eleven urns labelled by $u \in \{0, 1, 2, \ldots, 10\}$, each containing 10 balls.
- Urn $\boldsymbol{u}$ contains $\boldsymbol{u}$ **black balls** and $10 - u$ white balls.

Inverse Probability Example

- There are eleven urns labelled by $u \in \{0, 1, 2, \ldots, 10\}$, each containing 10 balls.
- Urn $\boldsymbol{u}$ contains $\boldsymbol{u}$ **black balls** and $10 - u$ white balls.

- Fred selects an urn u at random and draws N times with replacement from that urn, obtaining n_B blacks and $N - n_B$ whites.

Inverse Probability Example

- There are eleven urns labelled by $u \in \{0, 1, 2, \ldots, 10\}$, each containing 10 balls.
- Urn $\boldsymbol{u}$ contains $\boldsymbol{u}$ **black balls** and $10 - u$ white balls.

- Fred selects an urn u at random and draws N times with replacement from that urn, obtaining n_B blacks and $N - n_B$ whites.
- Fred's friend, Bill, looks on. **If after $N = 10$ draws $n_B = 3$ blacks have been drawn, what is the probability that the urn Fred is using is urn u, from Bill's point of view?**

Inverse Probability Example

- The variables:
  - The chosen urn: $u \in \{0, 1, 2, \ldots, 10\}$
  - The number of black balls from N draws: $n_B \in \{0, 1, \ldots, N\}$

Inverse Probability Example

- The variables:
  - The chosen urn: $u \in \{0, 1, 2, . . . , 10\}$
  - The number of black balls from N draws: $n_B \in \{0, 1, . . . , N\}$
- Fred's friend, Bill, looks on. **If after $N = 10$ draws $n_B = 3$ blacks have been drawn, what is the probability that the urn Fred is using is urn u, from Bill's point of view?**

Inverse Probability Example

- The variables:
  - The chosen urn: $u \in \{0, 1, 2, \ldots, 10\}$
  - The number of black balls from N draws: $n_B \in \{0, 1, \ldots, N\}$
- Fred's friend, Bill, looks on. **If after $N = 10$ draws $n_B = 3$ blacks have been drawn, what is the probability that the urn Fred is using is urn u, from Bill's point of view?**

$$\boxed{P(u \mid n_B, N) = ?}$$

Inverse Probability Example

- The variables:
  - The chosen urn: $u \in \{0, 1, 2, \ldots, 10\}$
  - The number of black balls from N draws: $n_B \in \{0, 1, \ldots, N\}$
- Fred's friend, Bill, looks on. **If after $N = 10$ draws $n_B = 3$ blacks have been drawn, what is the probability that the urn Fred is using is urn *u*, from Bill's point of view?**

$$\boxed{P(u \mid n_B, N) = ?}$$

- The joint distribution of u and n_B

Inverse Probability Example

- The variables:
  - The chosen urn: $u \in \{0, 1, 2, \ldots, 10\}$
  - The number of black balls from N draws: $n_B \in \{0, 1, \ldots, N\}$
- Fred's friend, Bill, looks on. **If after $N = 10$ draws $n_B = 3$ blacks have been drawn, what is the probability that the urn Fred is using is urn u, from Bill's point of view?**

$$\boxed{P(u \mid n_B, N) = ?}$$

- The joint distribution of u and n_B

$$P(u, n_B \mid N) = ?$$

Inverse Probability Example

- The variables:
  - The chosen urn: $u \in \{0, 1, 2, \ldots, 10\}$
  - The number of black balls from N draws: $n_B \in \{0, 1, \ldots, N\}$
- Fred's friend, Bill, looks on. **If after $N = 10$ draws $n_B = 3$ blacks have been drawn, what is the probability that the urn Fred is using is urn *u*, from Bill's point of view?**

$$\boxed{P(u \mid n_B, N) = ?}$$

- The joint distribution of u and n_B

$$P(u, n_B \mid N) = ?$$

$$P(u, n_B \mid N) = P(n_B \mid u, N) P(u)$$

Inverse Probability Example

- The variables:
  - The chosen urn: $u \in \{0, 1, 2, \ldots, 10\}$
  - The number of black balls from N draws: $n_B \in \{0, 1, \ldots, N\}$
- Fred's friend, Bill, looks on. **If after $N = 10$ draws $n_B = 3$ blacks have been drawn, what is the probability that the urn Fred is using is urn u, from Bill's point of view?**

$$\boxed{P(u \mid n_B, N) = ?}$$

- The joint distribution of u and n_B

$$P(u, n_B \mid N) = ?$$

$$P(u, n_B \mid N) = P(n_B \mid u, N)\overset{\checkmark}{P(u)}$$

Inverse Probability Example

- The variables:
  - The chosen urn: $u \in \{0, 1, 2, \dots, 10\}$
  - The number of black balls from N draws: $n_B \in \{0, 1, \dots, N\}$
- Fred's friend, Bill, looks on. **If after $N = 10$ draws $n_B = 3$ blacks have been drawn, what is the probability that the urn Fred is using is urn u, from Bill's point of view?**

$$\boxed{P(u \mid n_B, N) = ?}$$

- The joint distribution of u and n_B

$$P(u, n_B \mid N) = ?$$

$$P(u, n_B \mid N) = \overset{\surd}{P(n_B \mid u, N)}\,\overset{\surd}{P(u)}$$

Inverse Probability Example: joint probability

- The variables:
  - The chosen urn: $u \in \{0, 1, 2, \ldots, 10\}$
  - The number of black balls from N draws: $n_B \in \{0, 1, \ldots, N\}$
- $P(u, n_B \mid N) = P(n_B \mid u, N) P(u)$

Inverse Probability Example: joint probability

- The variables:
  - The chosen urn: $u \in \{0, 1, 2, \ldots, 10\}$
  - The number of black balls from N draws: $n_B \in \{0, 1, \ldots, N\}$
- $P(u, n_B \mid N) = P(n_B \mid u, N) P(u)$

$$P(u) = 1/11$$

Inverse Probability Example: joint probability

- The variables:
  - The chosen urn: $u \in \{0, 1, 2, \ldots, 10\}$
  - The number of black balls from N draws: $n_B \in \{0, 1, \ldots, N\}$
- $P(u, n_B \mid N) = P(n_B \mid u, N) P(u)$

$$P(u) = \frac{1}{11}$$

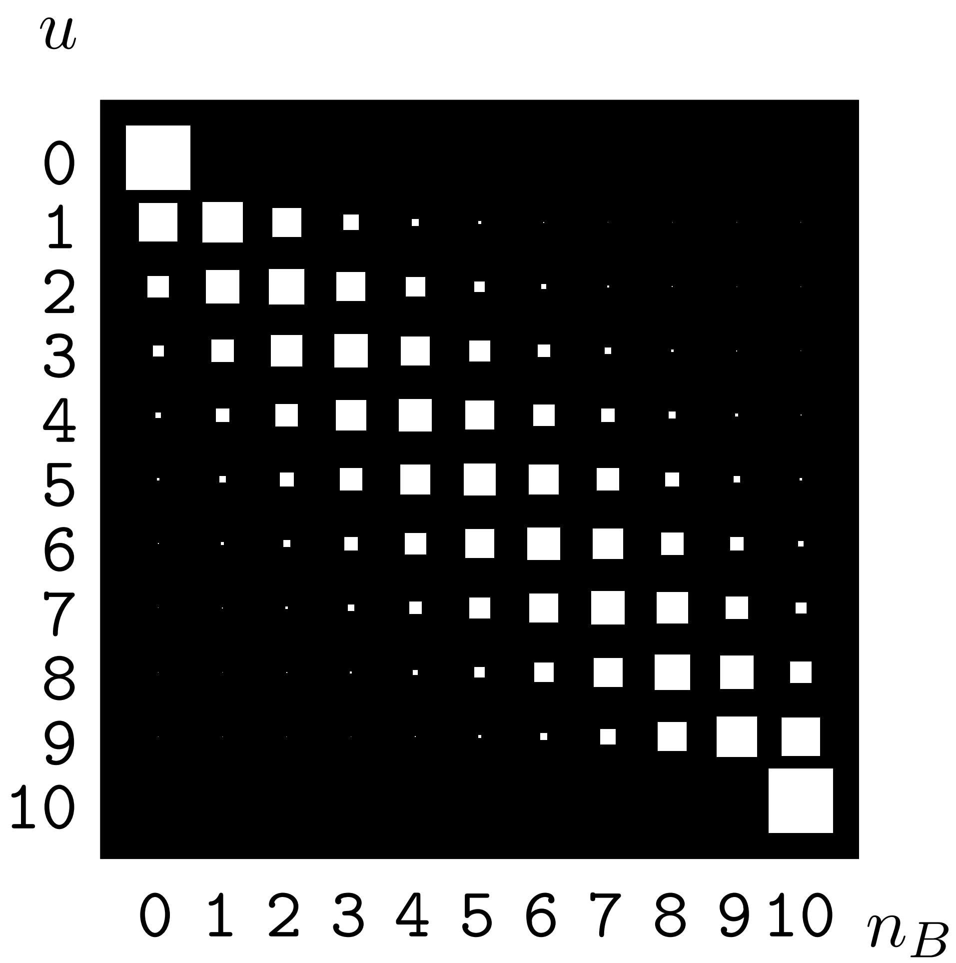
$$N = 10$$

Inverse Probability Example: joint probability

- The variables:
  - The chosen urn: $u \in \{0, 1, 2, \ldots, 10\}$
  - The number of black balls from N draws: $n_B \in \{0, 1, \ldots, N\}$
- $P(u, n_B \mid N) = P(n_B \mid u, N) P(u)$

  $P(u) = \frac{1}{11}$

  $N = 10$



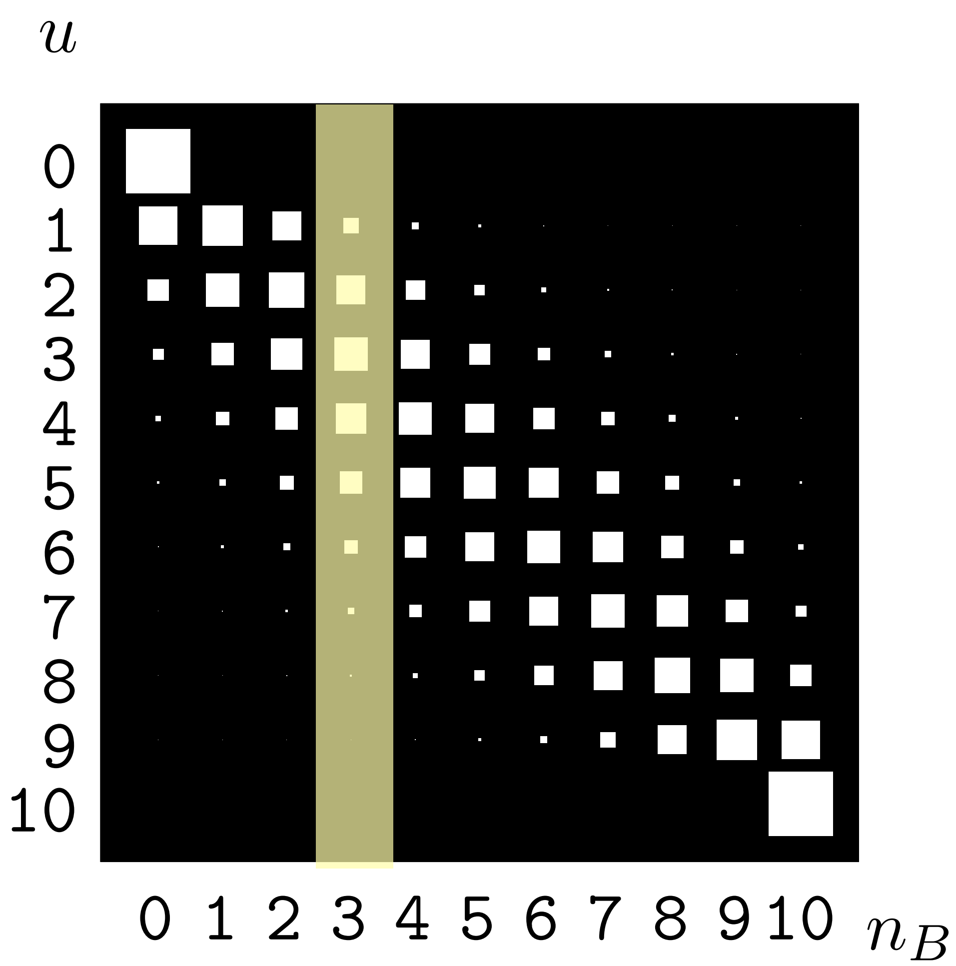
Inverse Probability Example: joint probability

- The variables:
  - The chosen urn: $u \in \{0, 1, 2, \ldots, 10\}$
  - The number of black balls from N draws: $n_B \in \{0, 1, \ldots, N\}$
- $P(u, n_B \mid N) = P(n_B \mid u, N) P(u)$

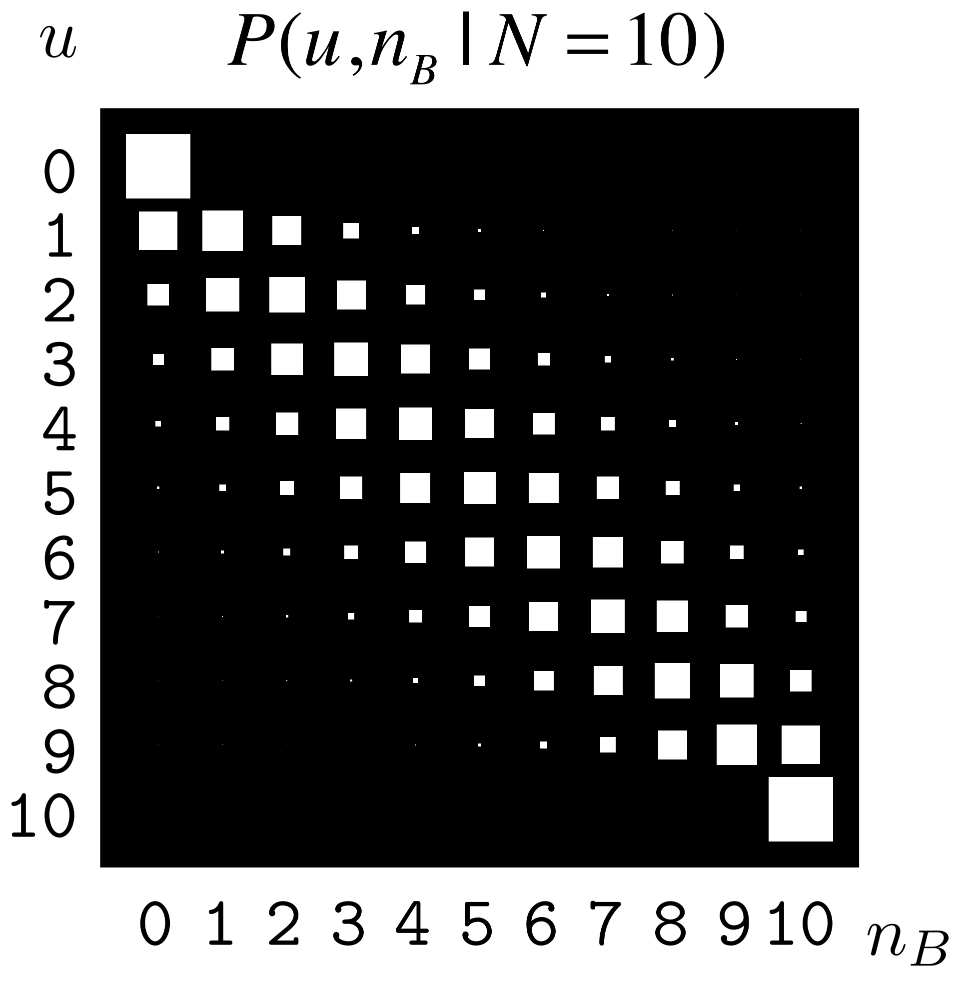
  $P(u) = 1/11$

  $N = 10$

  $n_B = 3$

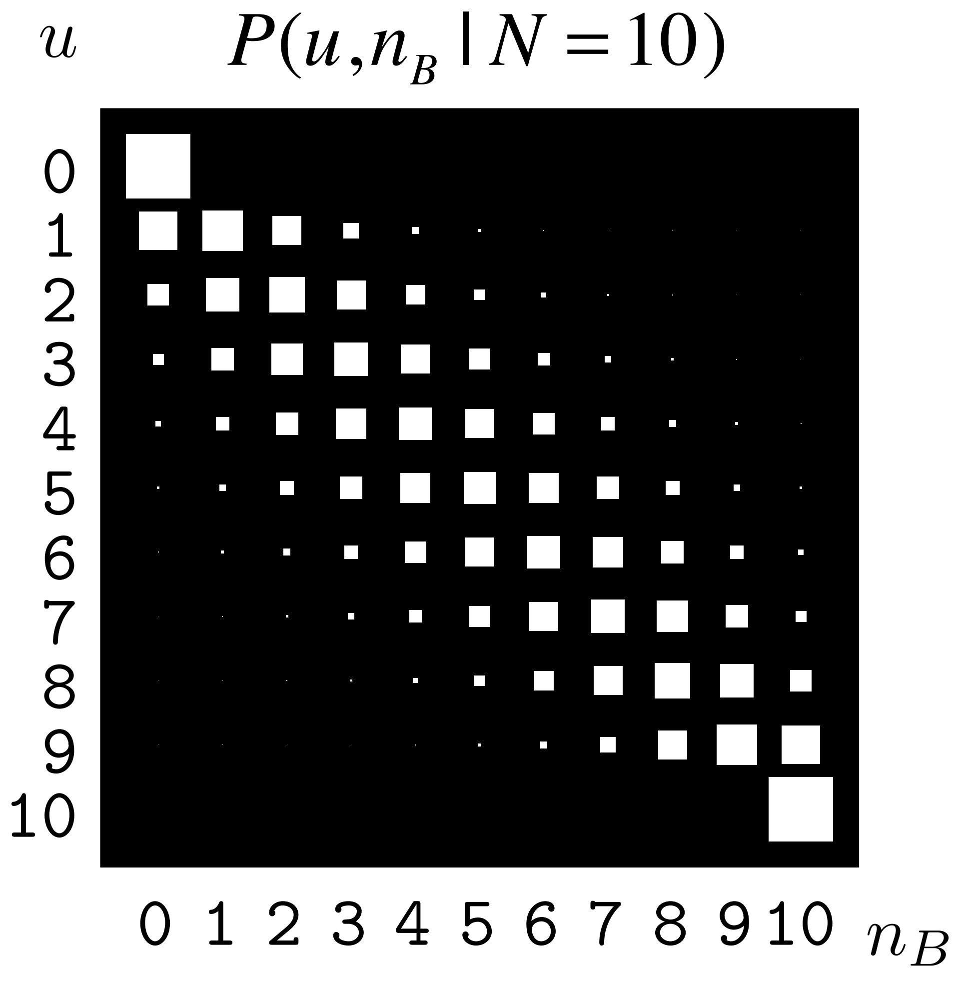


Inverse Probability Example: joint probability



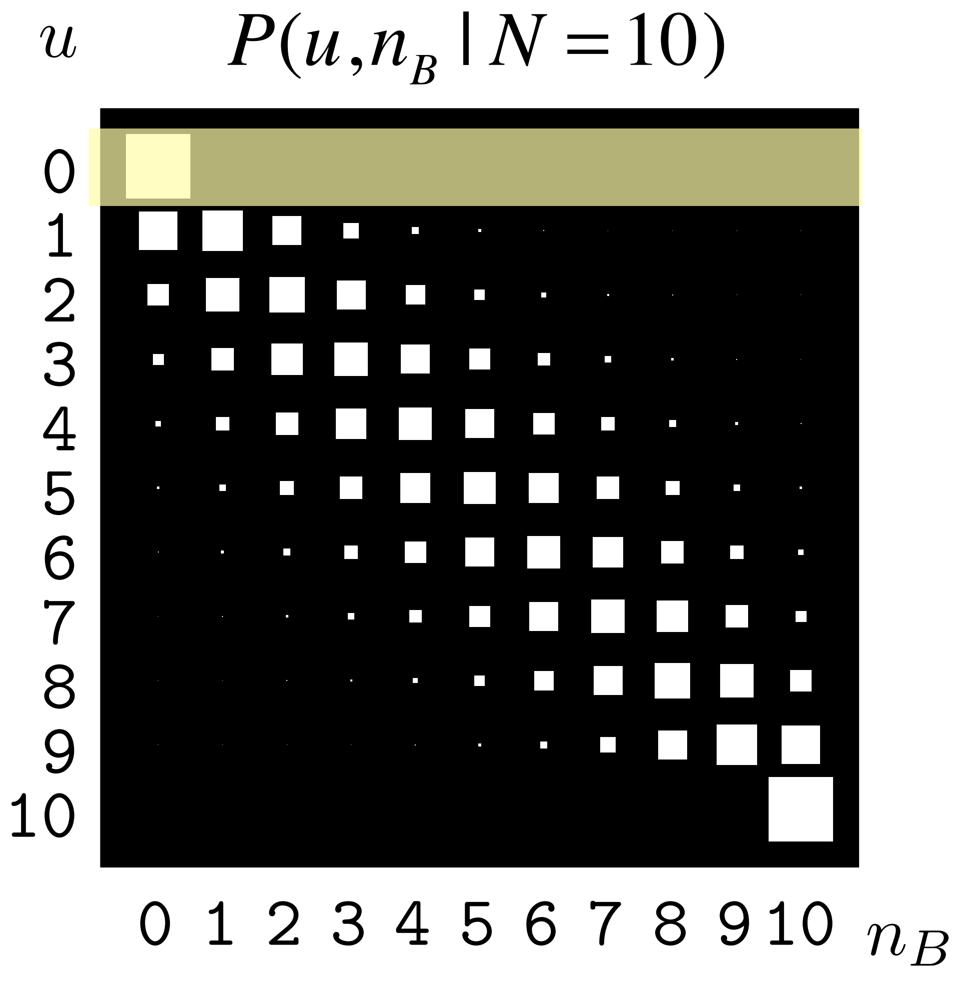
Inverse Probability Example: joint probability

- If the chosen urn is $u = 0$, all balls are white and n_B has to 0



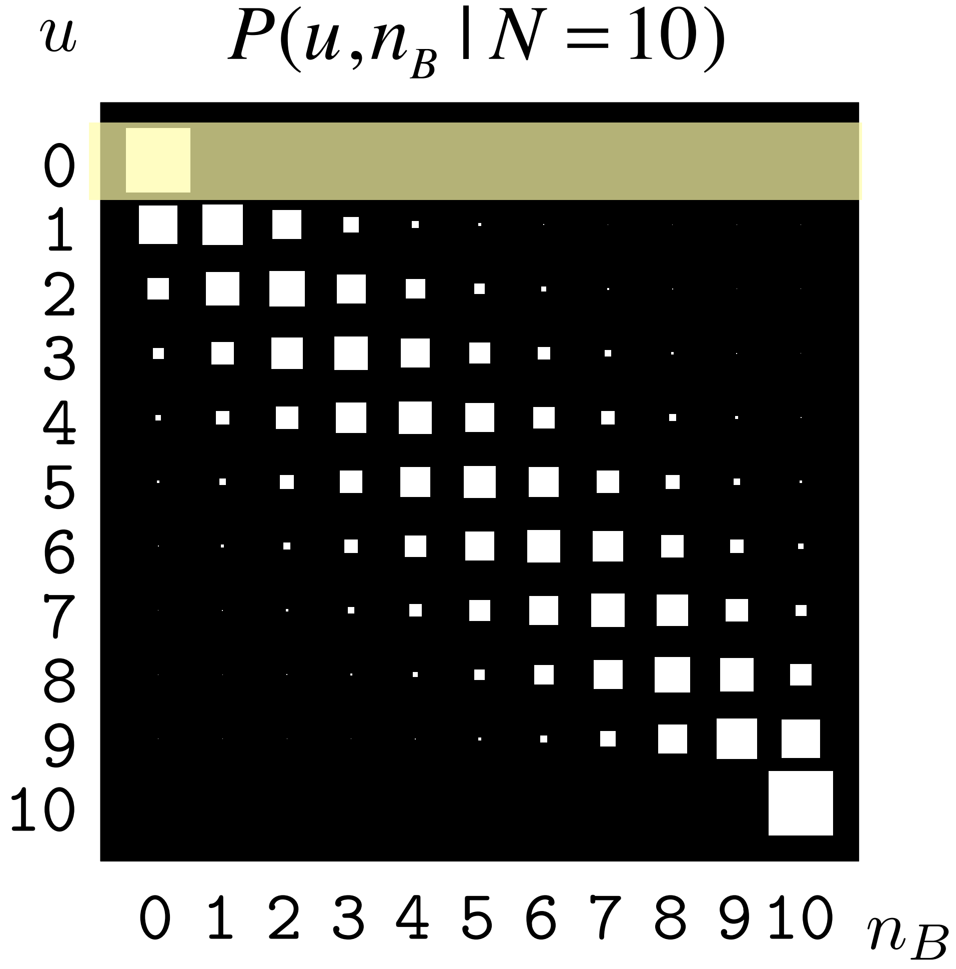
Inverse Probability Example: joint probability

- If the chosen urn is $u = 0$, all balls are white and n_B has to 0
    - $P(u = 0, n_B \neq 0) = 0$ and we can say that $P(n_B = 0 \mid u = 0) = 1$;



Inverse Probability Example: joint probability

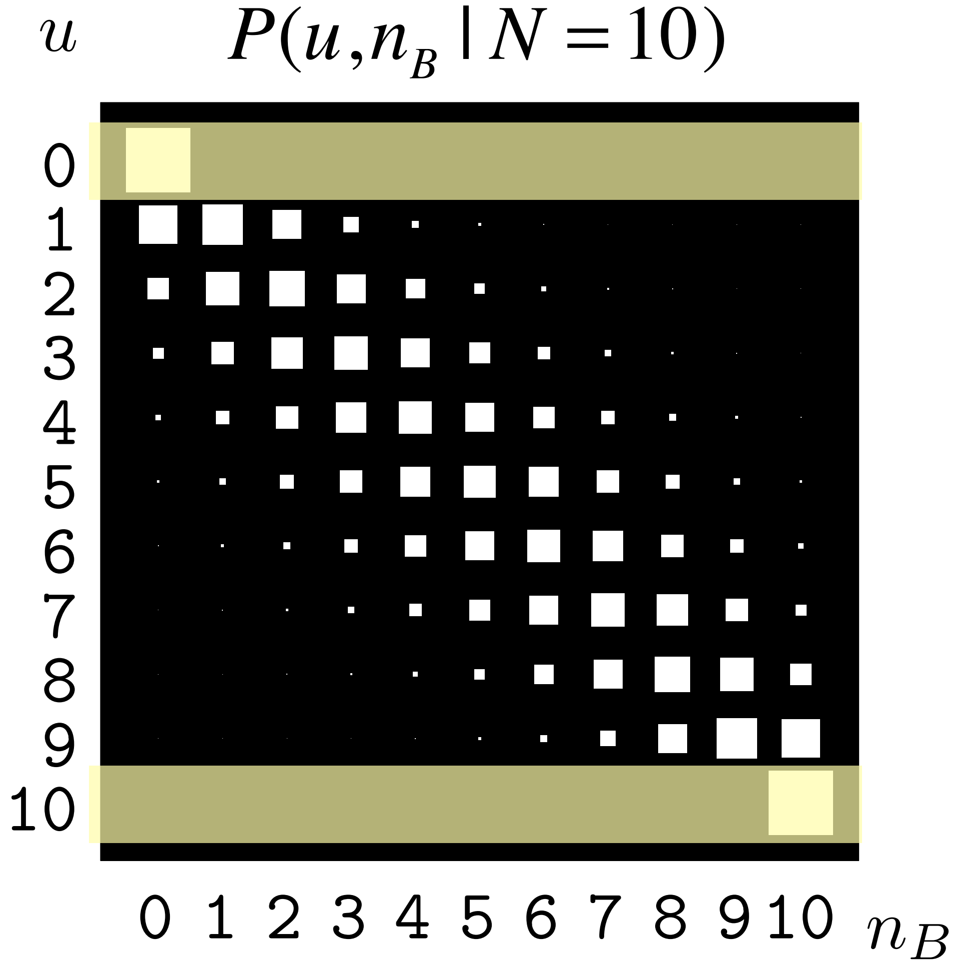
- If the chosen urn is $u = 0$, all balls are white and n_B has to 0
  - $P(u = 0, n_B \neq 0) = 0$ and we can say that $P(n_B = 0 \mid u = 0) = 1$;
- If the chosen urn is $u = 10$, all balls are black and n_B has to the number of draws ($N = 10$)



Inverse Probability Example: joint probability

- If the chosen urn is $u = 0$, all balls are white and n_B has to 0
    - $P(u = 0, n_B \neq 0) = 0$ and we can say that $P(n_B = 0 \mid u = 0) = 1$;
- If the chosen urn is $u = 10$, all balls are black and n_B has to the number of draws ($N = 10$)
    - $P(u = 10, n_B \neq 10) = 0$ and we can say

      that $P(n_B = 10 \mid u = 10) = 1$;



Inverse Probability Example

- Fred's friend, Bill, looks on. **If after $N = 10$ draws $n_B = 3$ blacks have been drawn, what is the probability that the urn Fred is using is urn u, from Bill's point of view?**

$$\boxed{P(u \mid n_B, N) = ?}$$

- The joint distribution of u and n_B:

$$P(u, n_B \mid N) = P(n_B \mid u, N) P(u)$$

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Inverse Probability Example

- Fred’s friend, Bill, looks on. **If after N = 10 draws n_B = 3 blacks have been drawn, what is the probability that the urn Fred is using is urn u, from Bill’s point of view?**

$$\boxed{P(u \mid n_B, N) = ?}$$

- The joint distribution of u and n_B:

$$P(u, n_B \mid N) = P(n_B \mid u, N) P(u)$$

- We can obtain the conditional distribution of u given n_B:

Inverse Probability Example

- Fred’s friend, Bill, looks on. **If after N = 10 draws n_B = 3 blacks have been drawn, what is the probability that the urn Fred is using is urn u, from Bill’s point of view?**

$$\boxed{P(u \mid n_B, N) = ?}$$

- The joint distribution of u and n_B:

$$P(u, n_B \mid N) = P(n_B \mid u, N) P(u)$$

- We can obtain the conditional distribution of u given n_B:

$$P(u \mid n_B, N) = \frac{P(u, n_B \mid N)}{P(n_B \mid N)}$$

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Inverse Probability Example

- Fred's friend, Bill, looks on. **If after N = 10 draws n_B = 3 blacks have been drawn, what is the probability that the urn Fred is using is urn u, from Bill's point of view?**

$$\boxed{P(u \mid n_B, N) = ?}$$

- The joint distribution of u and n_B:

$$P(u, n_B \mid N) = P(n_B \mid u, N) P(u)$$

- We can obtain the conditional distribution of u given n_B:

$$P(u \mid n_B, N) = \frac{P(u, n_B \mid N)}{P(n_B \mid N)}$$

$$P(u \mid n_B, N) = \frac{P(n_B \mid u, N) P(u)}{P(n_B \mid N)}$$

Inverse Probability Example

$$\boxed{P(u \mid n_B, N) = ?}$$

$$P(u \mid n_B, N) = \frac{P(n_B \mid u, N) P(u)}{P(n_B \mid N)}$$

Inverse Probability Example

$$\boxed{P(u \mid n_B, N) = ?}$$

$$P(u \mid n_B, N) = \frac{P(n_B \mid u, N)P(u)}{P(n_B \mid N)}$$

- $P(u) = 1/11$ for all u since we assume that all urns are equiprobable

Inverse Probability Example

$$\boxed{P(u \mid n_B, N) = ?}$$

$$P(u \mid n_B, N) = \frac{P(n_B \mid u, N)P(u)}{P(n_B \mid N)}$$

- $P(u) = 1/11$ for all u since we assume that all urns are equiprobable
- $P(n_B \mid u, N)$ can be calculated as the previous example.

Inverse Probability Example

$$\boxed{P(u \mid n_B, N) = ?}$$

$$P(u \mid n_B, N) = \frac{P(n_B \mid u, N) P(u)}{P(n_B \mid N)}$$

- $P(u) = 1/11$ for all u since we assume that all urns are equiprobable
- $P(n_B \mid u, N)$ can be calculated as the previous example.
    - For the urn u, u will be also the number of black balls in that urn,

Inverse Probability Example

$$\boxed{P(u \mid n_B, N) = ?}$$

$$P(u \mid n_B, N) = \frac{P(n_B \mid u, N) P(u)}{P(n_B \mid N)}$$

- $P(u) = 1/11$ for all u since we assume that all urns are equiprobable
- $P(n_B \mid u, N)$ can be calculated as the previous example.
  - For the urn u, u will be also the number of black balls in that urn,
  - $f_u = u/10$ (10 balls in each urn),

$$P(n_B \mid u, N) = \binom{N}{n_B} f_u^{\,n_B} (1 - f_u)^{N-n_B}$$

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Inverse Probability Example

$$\boxed{P(u \mid n_B, N) = ?}$$

$$P(u \mid n_B, N) = \frac{P(n_B \mid u, N) P(u)}{P(n_B \mid N)}$$

- $P(u) = 1/11$ for all u since we assume that all urns are equiprobable
- $P(n_B \mid u, N)$ can be calculated as the previous example.
  - For the urn u, u will be also the number of black balls in that urn,
  - $f_u = u/10$ (10 balls in each urn),

$$P(n_B \mid u, N) = \binom{N}{n_B} f_u^{\,n_B} (1 - f_u)^{N-n_B}$$

  - And how to compute $P(n_B \mid N)$?

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Inverse Probability Example

$$\boxed{P(u \mid n_B, N) = ?}$$

$$P(u \mid n_B, N) = \frac{P(n_B \mid u, N) P(u)}{P(n_B \mid N)}$$

- $P(u) = 1/11$ for all u since we assume that all urns are equiprobable
- $P(n_B \mid u, N)$ can be calculated as the previous example.
  - For the urn u, u will be also the number of black balls in that urn,
  - $f_u = u/10$ (10 balls in each urn),

$$P(n_B \mid u, N) = \binom{N}{n_B} f_u^{\,n_B} (1 - f_u)^{N - n_B}$$

  - And how to compute $P(n_B \mid N)$?
    - This a marginal from the joint probability $P(u, n_B \mid N)$

Inverse Probability Example

- And how to compute $P(n_B \mid N)$?

$$\boxed{P(u \mid n_B, N) = \frac{P(n_B \mid u, N) P(u)}{P(n_B \mid N)}}$$

$$P(n_B \mid N) = \sum_u P(u, n_B \mid N) = \sum_u P(u) P(n_B \mid u, N)$$

Inverse Probability Example

- And how to compute $P(n_B \mid N)$?

$$\boxed{P(u \mid n_B, N) = \frac{P(n_B \mid u, N)P(u)}{P(n_B \mid N)}}$$

$$P(n_B \mid N) = \sum_u P(u, n_B \mid N) = \sum_u P(u)P(n_B \mid u, N)$$

- Since …

Inverse Probability Example

- And how to compute $P(n_B | N)$?

$$\boxed{P(u | n_B, N) = \frac{P(n_B | u, N) P(u)}{P(n_B | N)}}$$

$$P(n_B | N) = \sum_u P(u, n_B | N) = \sum_u P(u) P(n_B | u, N)$$

- Since …

$$P(n_B | u, N) = \binom{N}{n_B} f_u^{\,n_B} (1 - f_u)^{N - n_B}$$

Inverse Probability Example

- And how to compute $P(n_B \mid N)$?

$$\boxed{P(u \mid n_B, N) = \frac{P(n_B \mid u, N) P(u)}{P(n_B \mid N)}}$$

$$P(n_B \mid N) = \sum_u P(u, n_B \mid N) = \sum_u P(u) P(n_B \mid u, N)$$

- Since …

$$P(n_B \mid u, N) = \binom{N}{n_B} f_u^{\,n_B} (1 - f_u)^{N - n_B}$$

$$P(u) = \tfrac{1}{11}$$

Inverse Probability Example

$$\boxed{P(u \mid n_B, N) = \frac{P(n_B \mid u, N) P(u)}{P(n_B \mid N)}}$$

- And how to compute $P(n_B \mid N)$?

$$P(n_B \mid N) = \sum_u P(u, n_B \mid N) = \sum_u P(u) P(n_B \mid u, N)$$

- Since …

$$P(n_B \mid u, N) = \binom{N}{n_B} f_u^{\,n_B} (1 - f_u)^{N - n_B}$$

$$P(u) = 1/_{11}$$

$$\boxed{P(u \mid n_B, N) = \frac{1}{P(n_B \mid N)} \tfrac{1}{11} \binom{N}{n_B} f_u^{\,n_B} (1 - f_u)^{N - n_B}}$$

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

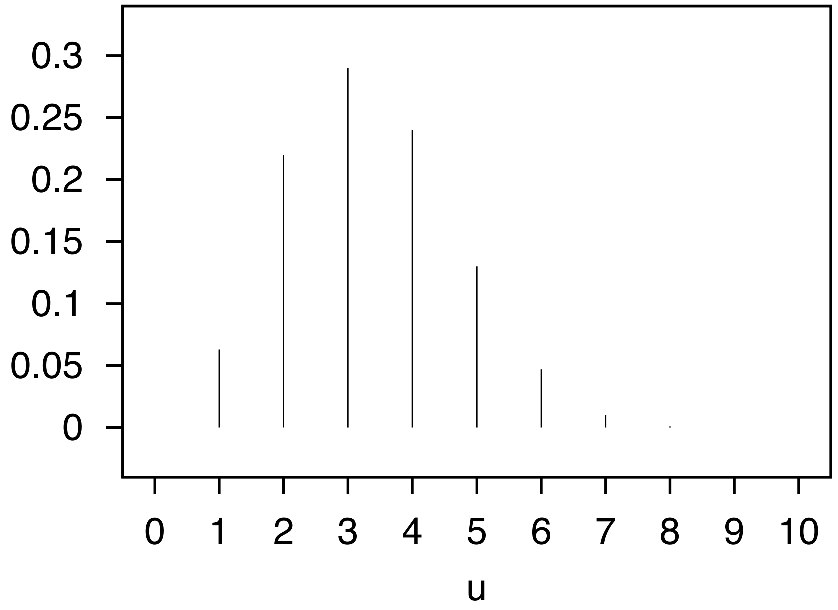
Inverse Probability Example

$$P(u \mid n_B, N) = \frac{1}{P(n_B \mid N)} \frac{1}{11} \binom{N}{n_B} f_u^{\,n_B} (1 - f_u)^{N - n_B}$$

- For $n_B = 3$ (and $N = 10$) the normalizing constant, the the marginal probability of $n_{B,}$

  $P(n_B \mid N) = 0.083$

- The posteriori probability $P(u \mid n_B = 3, N = 10)$



u	$P(u \mid n_B = 3, N)$
0	0
1	0.063
2	0.22
3	0.29
4	0.24
5	0.13
6	0.047
7	0.0099
8	0.00086
9	0.0000096
10	0

Terminology of inverse Probability

$$P(x \mid y, \boldsymbol{H}) = \frac{P(y \mid x, \boldsymbol{H}) P(x \mid \boldsymbol{H})}{P(y \mid \boldsymbol{H})}$$

Terminology of inverse Probability

$\boldsymbol{H}$ the overall hypothesis

$$P(x \mid y, \boldsymbol{H}) = \frac{P(y \mid x, \boldsymbol{H}) P(x \mid \boldsymbol{H})}{P(y \mid \boldsymbol{H})}$$

Terminology of inverse Probability

- $P(x \mid \boldsymbol{H})$ or just $P(x)$ - ***prior* probability of *x***.

H the overall hypothesis

prior probability of x

$$P(x \mid y,\boldsymbol{H}) = \frac{P(y \mid x,\boldsymbol{H})P(x \mid \boldsymbol{H})}{P(y \mid \boldsymbol{H})}$$

Terminology of inverse Probability

$\boldsymbol{H}$ the overall hypothesis

- $P(x \mid \boldsymbol{H})$ or just $P(x)$ - ***prior* probability of $\boldsymbol{x}$**.
- $P(y \mid x, \boldsymbol{H})$ or just $P(y \mid x)$ - ***likelihood* of $\boldsymbol{x}$**.
  - For a fixed x, $P(y \mid x, \boldsymbol{H})$ defines a probability over y.
  - For a fixed y, $P(y \mid x, \boldsymbol{H})$ defines the ***likelihood*** of x.

$$P(x \mid y, \boldsymbol{H}) = \frac{\overbrace{P(y \mid x, \boldsymbol{H})}^{\textit{likelihood of } x}\,\overbrace{P(x \mid \boldsymbol{H})}^{\textit{prior} \text{ probability of } x}}{P(y \mid \boldsymbol{H})}$$

Terminology of inverse Probability

$\boldsymbol{H}$ the overall hypothesis

- $P(x \mid \boldsymbol{H})$ or just $P(x)$ - ***prior* probability of $\boldsymbol{x}$**.
- $P(y \mid x, \boldsymbol{H})$ or just $P(y \mid x)$ - ***likelihood* of $\boldsymbol{x}$**.
  - For a fixed x, $P(y \mid x, \boldsymbol{H})$ defines a probability over y.
  - For a fixed y, $P(y \mid x, \boldsymbol{H})$ defines the ***likelihood*** of x.
- $P(x \mid y, \boldsymbol{H})$ or just $P(x \mid y)$ - ***posterior* probability of $\boldsymbol{x}$ given $\boldsymbol{y}$**.

likelihood of x — *prior* probability of x

$$P(x \mid y, \boldsymbol{H}) = \frac{P(y \mid x, \boldsymbol{H})\,P(x \mid \boldsymbol{H})}{P(y \mid \boldsymbol{H})}$$

posterior probability of x

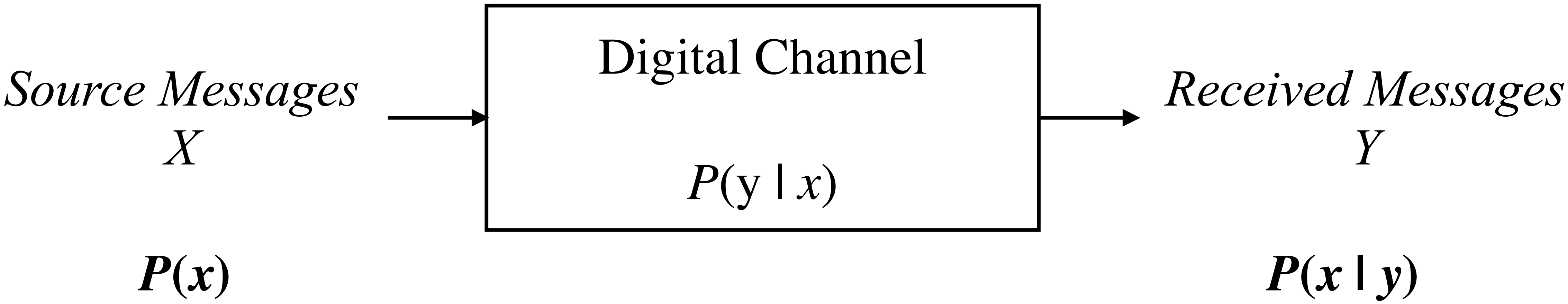
Terminology of inverse Probability

H the overall hypothesis

- $P(x \mid \boldsymbol{H})$ or just $P(x)$ - ***prior* probability of *x***.
- $P(y \mid x, \boldsymbol{H})$ or just $P(y \mid x)$ - ***likelihood* of *x***.
  - For a fixed x, $P(y \mid x, \boldsymbol{H})$ defines a probability over y.
  - For a fixed y, $P(y \mid x, \boldsymbol{H})$ defines the ***likelihood*** of x.
- $P(x \mid y, \boldsymbol{H})$ or just $P(x \mid y)$ - ***posterior* probability of *x* given *y***.
- $P(y \mid \boldsymbol{H})$ - is the normalizing constant, know as ***evidence*** or the ***marginal likelihood***.

$$\underbrace{P(x \mid y, \boldsymbol{H})}_{\textit{posterior} \text{ probability of } x} = \frac{\overbrace{P(y \mid x, \boldsymbol{H})}^{\textit{likelihood of } x}\,\overbrace{P(x \mid \boldsymbol{H})}^{\textit{prior} \text{ probability of } x}}{P(y \mid \boldsymbol{H})}$$

Terminology of inverse Probability



Source Messages

X

Digital Channel

$P(y \mid x)$

Received Messages

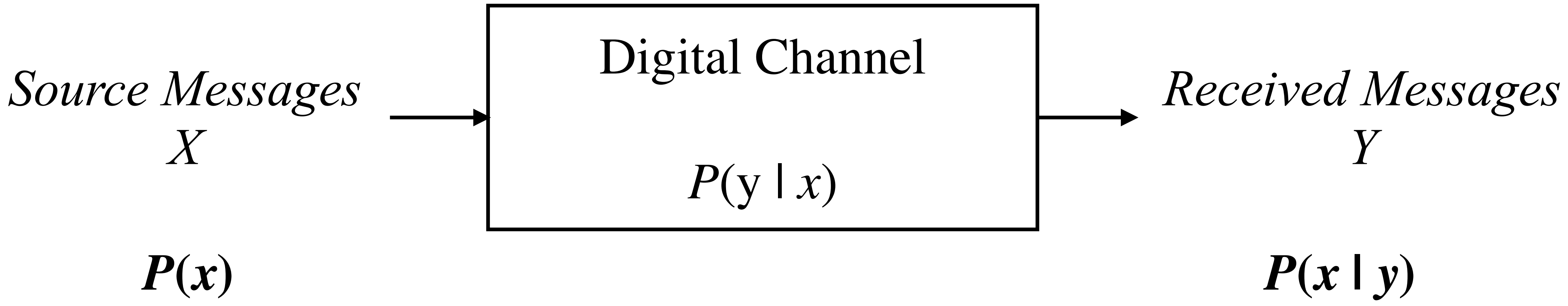
Y

$\boldsymbol{P(x)}$

$\boldsymbol{P(x \mid y)}$

Terminology of inverse Probability

H the overall hypothesis



Source Messages
X

Digital Channel

$P(y \mid x)$

Received Messages
Y

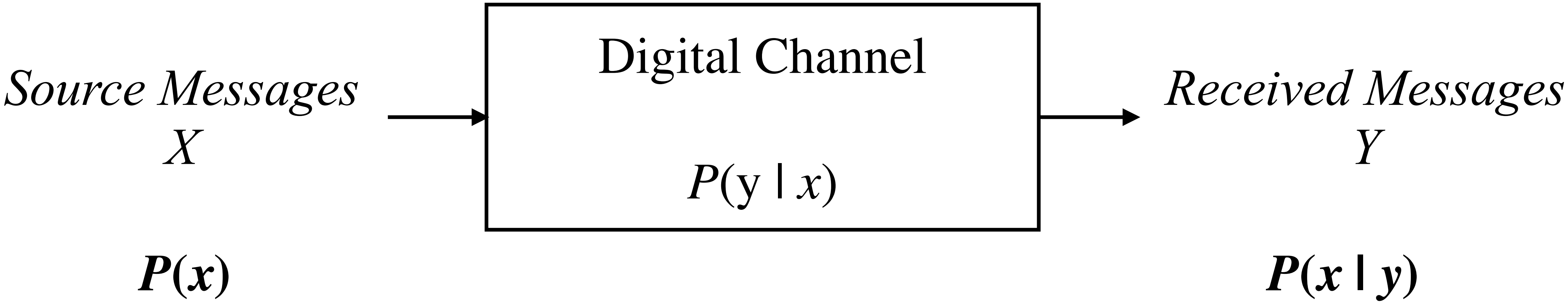
$\boldsymbol{P(x)}$

$\boldsymbol{P(x \mid y)}$

Terminology of inverse Probability

- $P(x \mid \boldsymbol{H})$ or just $P(x)$ - ***prior* probability of *x***.

H the overall hypothesis



Source Messages
X

Digital Channel

$P(y \mid x)$

Received Messages
Y

$\boldsymbol{P(x)}$

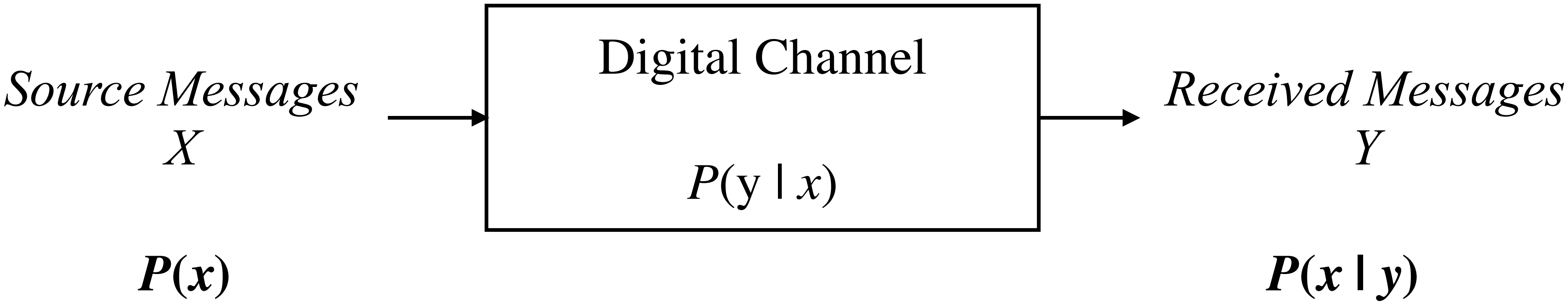
$\boldsymbol{P(x \mid y)}$

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Terminology of inverse Probability

- $P(x \mid \boldsymbol{H})$ or just $P(x)$ - ***prior* probability of *x***.
- $P(y \mid x, \boldsymbol{H})$ or just $P(y \mid x)$ - ***likelihood* of *x***.
  - For a fixed x, $P(y \mid x, \boldsymbol{H})$ defines a probability over *y*.
  - For a fixed y, $P(y \mid x, \boldsymbol{H})$ defines the likelihood of *x*.

H the overall hypothesis



Source Messages
X

$\boldsymbol{P(x)}$

Digital Channel

$P(\mathrm{y} \mid x)$

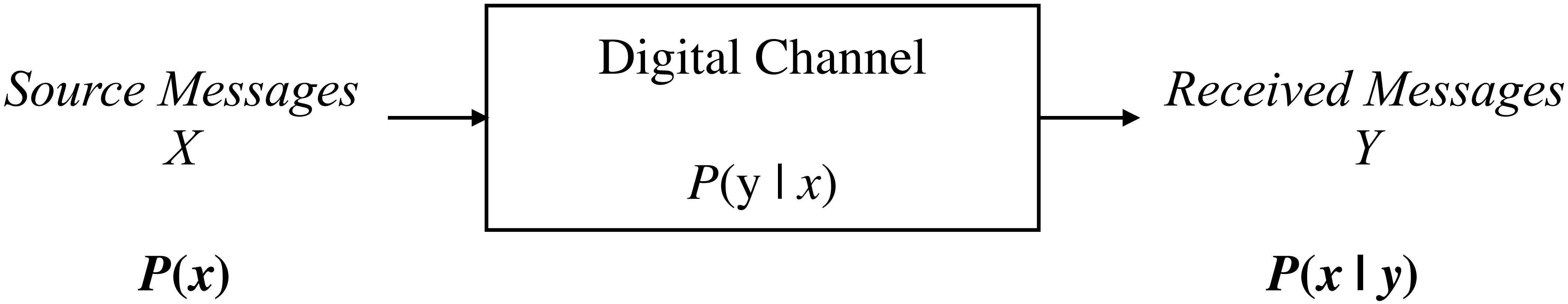
Received Messages
Y

$\boldsymbol{P(x \mid y)}$

Terminology of inverse Probability

- $P(x \mid \boldsymbol{H})$ or just $P(x)$ - ***prior* probability of *x***.
- $P(y \mid x, \boldsymbol{H})$ or just $P(y \mid x)$ - ***likelihood* of *x***.
  - For a fixed x, $P(y \mid x, \boldsymbol{H})$ defines a probability over y.
  - For a fixed y, $P(y \mid x, \boldsymbol{H})$ defines the likelihood of x.
- $P(x \mid y, \boldsymbol{H})$ or just $P(x \mid y)$ - ***posterior* probability of *x* given *y***.

$\boldsymbol{H}$ the overall hypothesis



Source Messages
X

Digital Channel

$P(\mathrm{y} \mid x)$

Received Messages
Y

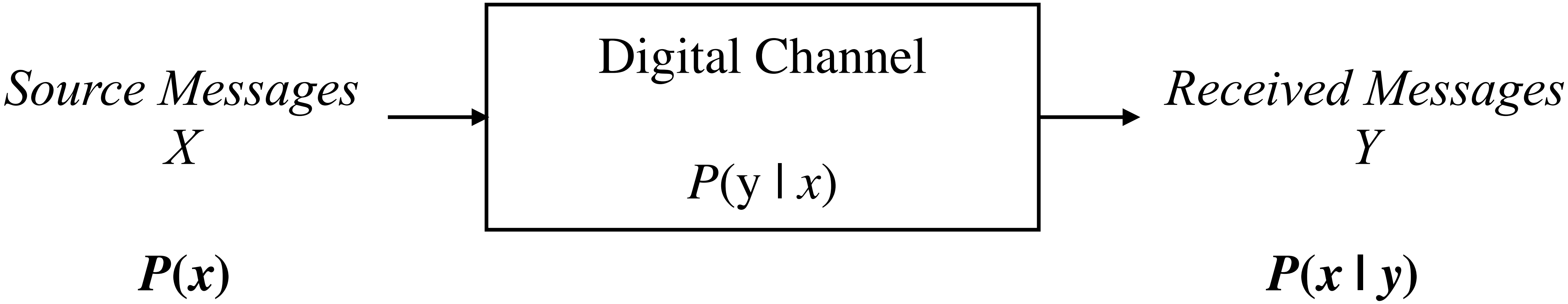
$\boldsymbol{P(x)}$

$\boldsymbol{P(x \mid y)}$

Terminology of inverse Probability

$\boldsymbol{H}$ the overall hypothesis

- $P(x \mid \boldsymbol{H})$ or just $P(x)$ - ***prior* probability of *x***.
- $P(y \mid x, \boldsymbol{H})$ or just $P(y \mid x)$ - ***likelihood* of *x***.
  - For a fixed x, $P(y \mid x, \boldsymbol{H})$ defines a probability over y.
  - For a fixed y, $P(y \mid x, \boldsymbol{H})$ defines the likelihood of x.
- $P(x \mid y, \boldsymbol{H})$ or just $P(x \mid y)$ - ***posterior* probability of *x* given *y***.
- $P(y \mid \boldsymbol{H})$ - is the normalizing constant, know as ***evidence*** or the ***marginal likelihood***.



Source Messages
X

Digital Channel

$P(y \mid x)$

Received Messages
Y

$\boldsymbol{P(x)}$

$\boldsymbol{P(x \mid y)}$

Terminology of inverse Probability

- **In general** *Θ* represents the **unknown parameters**
- *D* represents the **data**
- ***H*** the overall hypothesis

$$P(\theta \mid D, \boldsymbol{H}) = \frac{P(D \mid \theta, \boldsymbol{H}) P(\theta \mid \boldsymbol{H})}{P(D \mid \boldsymbol{H})}$$

$$posterior = \frac{likelihood \times prior}{evidence}$$

Terminology of inverse Probability: an example

- Variables (*a* - Jo's state of the health; *b* - Result of the test)

Initial information

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

Terminology of inverse Probability: an example

- Variables (*a* - Jo's state of the health; *b* - Result of the test)

Initial information

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

→ ***prior*** **probability of *a***

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

Terminology of inverse Probability: an example

- Variables (*a* - Jo's state of the health; *b* - Result of the test)

Initial information

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

→ ***prior* probability of *a***

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

→ ***likelihood* of *a given b* = 1**

Terminology of inverse Probability: an example

- Variables (a - Jo's state of the health; b - Result of the test)

Initial information

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

→ ***prior* probability of a**

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

→ ***likelihood* of a *given* $b = 1$**

→ ***likelihood* of a *given* $b = 0$**

Terminology of inverse Probability: an example

- Variables (a - Jo's state of the health; b - Result of the test)

Initial information

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

→ ***prior* probability of a**

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

→ ***likelihood* of a *given* $b = 1$**

→ ***likelihood* of a *given* $b = 0$**

↓

***probability* of b *given* $a = 0$**

Terminology of inverse Probability: an example

- Variables (a - Jo's state of the health; b - Result of the test)

Initial information

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

prior **probability of** ***a***

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

likelihood **of *a given b* = 1**

likelihood **of *a given b* = 0**

probability **of *b given a* = 0**

probability **of *b given a* = 1**

Terminology of inverse Probability: an example

- Variables (a - Jo’s state of the health; b - Result of the test)

Initial information

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

→ *prior* **probability of** a

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

→ *likelihood* **of** a *given* $b = 1$

→ *likelihood* **of** a *given* $b = 0$

$$P(a, b) = P(b \mid a)\ P(a)$$

$P(a, \mathrm{b})$	$a = 1$	$a = 0$
$b = 1$	**0.0095**	**0.0495**
$b = 0$	**0.0005**	**0.9405**

$$P(b) = \sum_{a} P(a,b)$$

$P(b = 1)$	**0.059**
$P(b = 0)$	**0.941**

Terminology of inverse Probability: an example

- Variables (*a* - Jo’s state of the health; *b* - Result of the test)

Initial information

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

→ ***prior* probability of *a***

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

→ ***likelihood* of *a given b* = 1**

→ ***likelihood* of *a given b* = 0**

$$P(a, b) = P(b \mid a)\, P(a)$$

$P(a, \mathrm{b})$	$a = 1$	$a = 0$
$b = 1$	**0.0095**	**0.0495**
$b = 0$	**0.0005**	**0.9405**

joint distribution

$$P(b) = \sum_a P(a,b)$$

$P(b = 1)$	**0.059**
$P(b = 0)$	**0.941**

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Terminology of inverse Probability: an example

- Variables (*a* - Jo’s state of the health; *b* - Result of the test)

Initial information

P(*a* = 1)	*P*(*a* = 0)
0.01	**0.99**

→ ***prior* probability of *a***

P(*b* | *a*)	*a* = 1	*a* = 0
b = 1	**0.95**	**0.05**
b = 0	**0.05**	**0.95**

→ ***likelihood* of *a given b* = 1**

→ ***likelihood* of *a given b* = 0**

$$P(a, b) = P(b \mid a)\ P(a)$$

P(*a*, b)	*a* = 1	*a* = 0
b = 1	**0.0095**	**0.0495**
b = 0	**0.0005**	**0.9405**

joint distribution

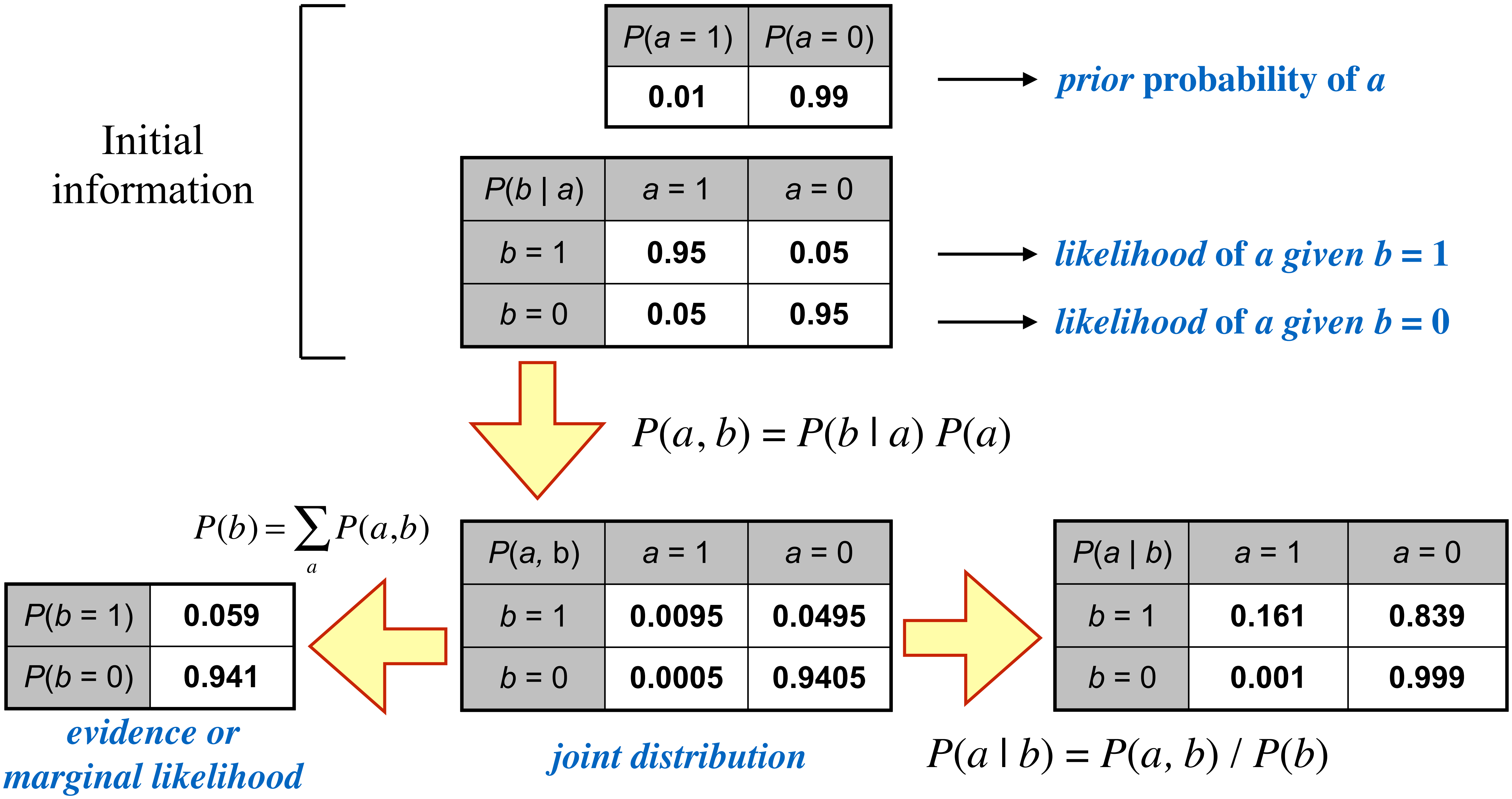
$$P(b) = \sum_a P(a,b)$$

P(*b* = 1)	**0.059**
P(*b* = 0)	**0.941**

evidence or marginal likelihood

Terminology of inverse Probability: an example

- Variables (*a* - Jo's state of the health; *b* - Result of the test)



Initial information

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

→ ***prior* probability of *a***

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

→ ***likelihood* of *a given b* = 1**

→ ***likelihood* of *a given b* = 0**

$$P(a, b) = P(b \mid a)\, P(a)$$

$P(a, \mathrm{b})$	$a = 1$	$a = 0$
$b = 1$	**0.0095**	**0.0495**
$b = 0$	**0.0005**	**0.9405**

joint distribution

$$P(b) = \sum_a P(a,b)$$

$P(b = 1)$	**0.059**
$P(b = 0)$	**0.941**

evidence or marginal likelihood

$P(a \mid b)$	$a = 1$	$a = 0$
$b = 1$	**0.161**	**0.839**
$b = 0$	**0.001**	**0.999**

$$P(a \mid b) = P(a, b) \,/\, P(b)$$

Terminology of inverse Probability: an example

- Variables (*a* - Jo's state of the health; *b* - Result of the test)

Initial information

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

→ ***prior* probability of *a***

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

→ ***likelihood* of *a given b* = 1**

→ ***likelihood* of *a given b* = 0**

$$P(a, b) = P(b \mid a)\, P(a)$$

$$P(b) = \sum_a P(a,b)$$

$P(b = 1)$	**0.059**
$P(b = 0)$	**0.941**

evidence or marginal likelihood

$P(a, \mathrm{b})$	$a = 1$	$a = 0$
$b = 1$	**0.0095**	**0.0495**
$b = 0$	**0.0005**	**0.9405**

joint distribution

posterior* probability of *a

$P(a \mid b)$	$a = 1$	$a = 0$
$b = 1$	**0.161**	**0.839**
$b = 0$	**0.001**	**0.999**

$$P(a \mid b) = P(a, b) \,/\, P(b)$$

Terminology of inverse Probability: an example

- Variables (a - Jo’s state of the health; b - Result of the test)

Initial information

$P(a = 1)$	$P(a = 0)$
0.01	**0.99**

→ ***prior* probability of *a***

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

→ ***likelihood* of *a given b* = 1**

→ ***likelihood* of *a given b* = 0**

$$P(a, b) = P(b \mid a)\, P(a)$$

$P(a, \mathrm{b})$	$a = 1$	$a = 0$
$b = 1$	**0.0095**	**0.0495**
$b = 0$	**0.0005**	**0.9405**

joint distribution

$$P(b) = \sum_a P(a,b)$$

$P(b = 1)$	**0.95**
$P(b = 0)$	**0.05**

evidence or marginal likelihood

***posterior* probability of *a* given *b* = 1**

$P(a \mid b)$	$a = 1$	$a = 0$
$b = 1$	**0.161**	**0.839**
$b = 0$	**0.001**	**0.999**

Terminology of inverse Probability: an example

- Variables (*a* - Jo's state of the health; *b* - Result of the test)

Initial information

$P(a=1)$	$P(a=0)$
0.01	**0.99**

→ ***prior*** **probability of *a***

$P(b \mid a)$	$a = 1$	$a = 0$
$b = 1$	**0.95**	**0.05**
$b = 0$	**0.05**	**0.95**

→ ***likelihood*** **of *a given b* = 1**

→ ***likelihood*** **of *a given b* = 0**

$$P(a,b) = P(b \mid a)\,P(a)$$

$P(a, \mathrm{b})$	$a = 1$	$a = 0$
$b = 1$	**0.0095**	**0.0495**
$b = 0$	**0.0005**	**0.9405**

joint distribution

$$P(b) = \sum_a P(a,b)$$

$P(b=1)$	**0.95**
$P(b=0)$	**0.05**

evidence or marginal likelihood

posterior **probability of *a* given *b* = 1**

$P(a \mid b)$	$a = 1$	$a = 0$
$b = 1$	**0.161**	**0.839**
$b = 0$	**0.001**	**0.999**

posterior **probability of *a* given *b* = 0**

Inverse probability and **prediction**

- Assuming again that Bill has observed $n_B = 3$ blacks in $N = 10$ draws, let Fred draw another ball from the **same urn**. **What is the probability that the next drawn ball is a black?**

Inverse probability and **prediction**

- Assuming again that Bill has observed $n_B = 3$ blacks in $N = 10$ draws, let Fred draw another ball from the **same urn**. **What is the probability that the next drawn ball is a black?**
- What we have to calculate $P(ball_{N+1} = black \mid n_B, N) = ?$

Inverse probability and **prediction**

- Assuming again that Bill has observed $n_B = 3$ blacks in $N = 10$ draws, let Fred draw another ball from the **same urn**. **What is the probability that the next drawn ball is a black?**
- What we have to calculate $P(ball_{N+1} = black \mid n_B, N) = ?$

$$P(ball_{N+1} = black \mid n_B, N) = \sum_u P(ball_{N+1} = black \mid u, n_B, N) P(u \mid n_B, N)$$

Inverse probability and prediction

- Assuming again that Bill has observed $n_B = 3$ blacks in $N = 10$ draws, let Fred draw another ball from the **same urn**. **What is the probability that the next drawn ball is a black?**
- What we have to calculate $P(ball_{N+1} = black \mid n_B, N) = ?$

$$P(ball_{N+1} = black \mid n_B, N) = \sum_u P(ball_{N+1} = black \mid u, n_B, N) P(u \mid n_B, N)$$

- Since the balls are drawn with replacement from the chosen urn u, the probability that the next drawn ball is a black, given the urn u, is just dependent on u, whatever n_B and N are.

$$P(ball_{N+1} = black \mid u, n_B, N) = f_u = {}^{u}\!/_{10}$$

Inverse probability and **prediction**

- Assuming again that Bill has observed $n_B = 3$ blacks in $N = 10$ draws, let Fred draw another ball from the **same urn**. **What is the probability that the next drawn ball is a black?**
- What we have to calculate $P(ball_{N+1} = black \mid n_B, N) = ?$

$$P(ball_{N+1} = black \mid n_B, N) = \sum_u P(ball_{N+1} = black \mid u, n_B, N) P(u \mid n_B, N)$$

- Since the balls are drawn with replacement from the chosen urn u, the probability that the next drawn ball is a black, given the urn u, is just dependent on u, whatever n_B and N are.

$$P(ball_{N+1} = black \mid u, n_B, N) = f_u = {}^{u}\!/_{10}$$

$$P(ball_{N+1} = black \mid n_B, N) = \sum_u f_u P(u \mid n_B, N)$$

Inverse probability and prediction

- Assuming again that Bill has observed $n_B = 3$ blacks in $N = 10$ draws, let Fred draw another ball from the **same urn**. **What is the probability that the next drawn ball is a black?**
- What we have to calculate $P(ball_{N+1} = black \mid n_B, N) = ?$

$$P(ball_{N+1} = black \mid n_B, N) = \sum_u P(ball_{N+1} = black \mid u, n_B, N) P(u \mid n_B, N)$$

- Since the balls are drawn with replacement from the chosen urn u, the probability that the next drawn ball is a black, given the urn u, is just dependent on u, whatever n_B and N are.

$$P(ball_{N+1} = black \mid u, n_B, N) = f_u = {}^{u}\!/_{10}$$

$$P(ball_{N+1} = black \mid n_B, N) = \sum_u f_u P(u \mid n_B, N)$$

- Since the values of $P(u \mid n_B, N)$ were previously calculated,

$$P(ball_{N+1} = black \mid n_B, N) = 0.333$$

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Inverse probability and prediction

u
0
1
2
3
4
5
6
7
8
9
10

Inverse probability and prediction

u	$P(u \mid n_B, N)$
0	0
1	0,063
2	0,22
3	0,29
4	0,24
5	0,13
6	0,047
7	0,0099
8	0,00086
9	0,0000096
10	0

Inverse probability and prediction

$$P(ball_{N+1} = black \mid u, n_B, N) = f_u = {}^{1}\!/_{10}$$

u	$P(u \mid n_B, N)$	$f_u = u/10$
0	0	0
1	0,063	0,1
2	0,22	0,2
3	0,29	0,3
4	0,24	0,4
5	0,13	0,5
6	0,047	0,6
7	0,0099	0,7
8	0,00086	0,8
9	0,0000096	0,9
10	0	1

Inverse probability and prediction

u	$P(u \mid n_B, N)$	$f_u = u/10$	$P(.)$
0	0	0	0
1	0,063	0,1	0,0063
2	0,22	0,2	0,044
3	0,29	0,3	0,087
4	0,24	0,4	0,096
5	0,13	0,5	0,065
6	0,047	0,6	0,0282
7	0,0099	0,7	0,00693
8	0,00086	0,8	0,000688
9	0,0000096	0,9	0,00000864
10	0	1	0

$$P(ball_{N+1} = black \mid u, n_B, N) = f_u = {}^{1}\!/_{10}$$

$$P(ball_{N+1} = black, u, n_B, N) =$$
$$P(ball_{N+1} = black \mid u, n_B, N) P(u \mid n_B, N)$$

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Inverse probability and prediction

$$P(ball_{N+1} = black \,|\, u, n_B, N) = f_u = {}^1\!/_{10}$$

$$P(ball_{N+1} = black, u, n_B, N) =$$
$$P(ball_{N+1} = black \,|\, u, n_B, N) P(u \,|\, n_B, N)$$

u	$P(u \mid n_B, N)$	$f_u = u/10$	$P(.)$
0	0	0	0
1	0,063	0,1	0,0063
2	0,22	0,2	0,044
3	0,29	0,3	0,087
4	0,24	0,4	0,096
5	0,13	0,5	0,065
6	0,047	0,6	0,0282
7	0,0099	0,7	0,00693
8	0,00086	0,8	0,000688
9	0,0000096	0,9	0,00000864
10	0	1	0

$$P(ball_{N+1} = black \,|\, n_B, N) = 0.333$$

$$P(ball_{N+1} = black \,|\, n_B, N) = \sum_u f_u P(u \,|\, n_B, N)$$

Inverse probability and prediction vs the classical way

$$P(ball_{N+1} = black \mid u, n_B, N) = f_u = {}^1\!/_{10}$$

$$P(ball_{N+1} = black, u, n_B, N) =$$
$$P(ball_{N+1} = black \mid u, n_B, N) P(u \mid n_B, N)$$

u	$P(u \mid n_B, N)$	$f_u = u/10$	$P(.)$
0	0	0	0
1	0,063	0,1	0,0063
2	0,22	0,2	0,044
3	0,29	0,3	0,087
4	0,24	0,4	0,096
5	0,13	0,5	0,065
6	0,047	0,6	0,0282
7	0,0099	0,7	0,00693
8	0,00086	0,8	0,000688
9	0,0000096	0,9	0,00000864
10	0	1	0

$$P(ball_{N+1} = black \mid n_B, N) = 0.333$$

$$P(ball_{N+1} = black \mid n_B, N) = \sum_u f_u P(u \mid n_B, N)$$

Inverse probability and prediction vs the classical way

$$P(ball_{N+1} = black \mid u, n_B, N) = f_u = {}^{1}\!/_{10}$$

$$P(ball_{N+1} = black, u, n_B, N) =$$
$$P(ball_{N+1} = black \mid u, n_B, N) P(u \mid n_B, N)$$

u	$P(u \mid n_B, N)$	$f_u = u/10$	$P(.)$
0	0	0	0
1	0,063	0,1	0,0063
2	0,22	0,2	0,044
3	0,29	0,3	0,087
4	0,24	0,4	0,096
5	0,13	0,5	0,065
6	0,047	0,6	0,0282
7	0,0099	0,7	0,00693
8	0,00086	0,8	0,000688
9	0,0000096	0,9	0,00000864
10	0	1	0

Selecting the most plausible hypothesis

$$P(ball_{N+1} = black \mid n_B, N) = 0.333$$

$$P(ball_{N+1} = black \mid n_B, N) = \sum_u f_u P(u \mid n_B, N)$$

Inverse probability and prediction vs the classical way

$$P(ball_{N+1} = black \mid u, n_B, N) = f_u = \frac{1}{10}$$

$$P(ball_{N+1} = black, u, n_B, N) =$$
$$P(ball_{N+1} = black \mid u, n_B, N) P(u \mid n_B, N)$$

u	$P(u \mid n_B, N)$	$f_u = u/10$	$P(.)$
0	0	0	0
1	0,063	0,1	0,0063
2	0,22	0,2	0,044
3	$\boxed{0{,}29}$	0,3	0,087
4	0,24	0,4	0,096
5	0,13	0,5	0,065
6	0,047	0,6	0,0282
7	0,0099	0,7	0,00693
8	0,00086	0,8	0,000688
9	0,0000096	0,9	0,00000864
10	0	1	0

Selecting the most plausible hypothesis

$$P(ball_{N+1} = black \mid n_B, N) = 0.333$$

$$P(ball_{N+1} = black \mid n_B, N) = \sum_u f_u P(u \mid n_B, N)$$

Inverse probability and prediction vs the classical way

$$P(ball_{N+1} = black \mid u, n_B, N) = f_u = {}^{1}\!/_{10}$$

$$P(ball_{N+1} = black, u, n_B, N) =$$
$$P(ball_{N+1} = black \mid u, n_B, N) P(u \mid n_B, N)$$

u	$P(u \mid n_B, N)$	$f_u = u/10$	$P(.)$
0	0	0	0
1	0,063	0,1	0,0063
2	0,22	0,2	0,044
3	0,29	0,3	0,087
4	0,24	0,4	0,096
5	0,13	0,5	0,065
6	0,047	0,6	0,0282
7	0,0099	0,7	0,00693
8	0,00086	0,8	0,000688
9	0,0000096	0,9	0,00000864
10	0	1	0

Selecting the most plausible hypothesis

$\boldsymbol{u = 3}$

$$P(ball_{N+1} = black \mid n_B, N) = 0.333$$

$$P(ball_{N+1} = black \mid n_B, N) = \sum_u f_u P(u \mid n_B, N)$$

Inverse probability and prediction vs the classical way

$$P(ball_{N+1} = black \mid u, n_B, N) = f_u = {}^1\!/_{10}$$

$$P(ball_{N+1} = black, u, n_B, N) =$$
$$P(ball_{N+1} = black \mid u, n_B, N) P(u \mid n_B, N)$$

u	$P(u \mid n_B, N)$	$f_u = u/10$	$P(.)$
0	0	0	0
1	0,063	0,1	0,0063
2	0,22	0,2	0,044
3	0,29	0,3	0,087
4	0,24	0,4	0,096
5	0,13	0,5	0,065
6	0,047	0,6	0,0282
7	0,0099	0,7	0,00693
8	0,00086	0,8	0,000688
9	0,0000096	0,9	0,00000864
10	0	1	0

Selecting the most plausible hypothesis

$$u = 3$$

then making the predictions assuming that hypothesis to be true

$$P = 0.3$$

$$P(ball_{N+1} = black \mid n_B, N) = 0.333$$

$$P(ball_{N+1} = black \mid n_B, N) = \sum_u f_u P(u \mid n_B, N)$$

The likelihood principle

The likelihood principle: given a generative model for data d given parameters $\boldsymbol{\theta}$, $P(d \,|\, \boldsymbol{\theta})$, and having observed a particular outcome d_1, all inferences and predictions should depend only on the function $P(d_1 \,|\, \boldsymbol{\theta})$.

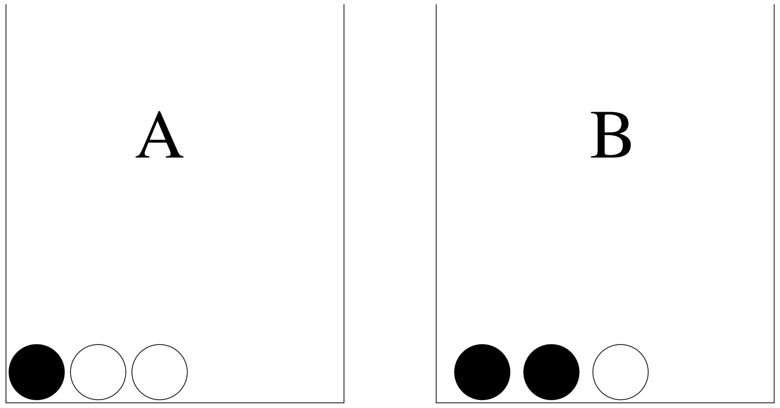
The likelihood principle

The likelihood principle: given a generative model for data d given parameters $\boldsymbol{\theta}$, $P(d \,|\, \boldsymbol{\theta})$, and having observed a particular outcome d_1, all inferences and predictions should depend only on the function $P(d_1 \,|\, \boldsymbol{\theta})$.

- The details of the other possible outcomes and their probabilities are irrelevant.
- All that matters is the probability of the outcome that actually happened given the different hypotheses.
- We need only to know the **likelihood**, i.e., **how the probability of the data that happened varies with the hypothesis.**

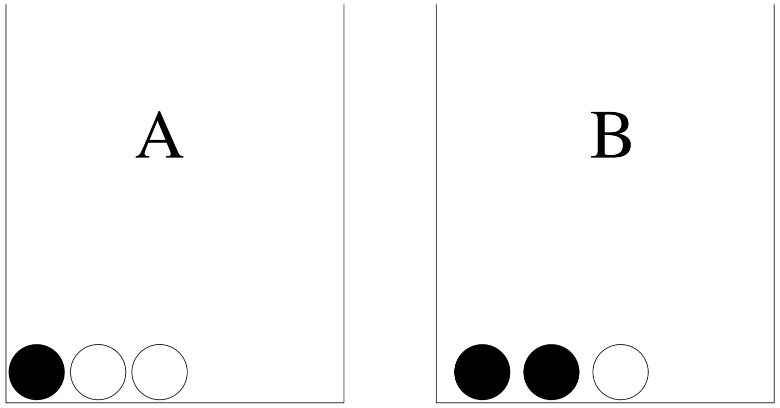
The likelihood principle

- Urn A contains three balls: one black, and two white; urn B contains three balls: two black, and one white.



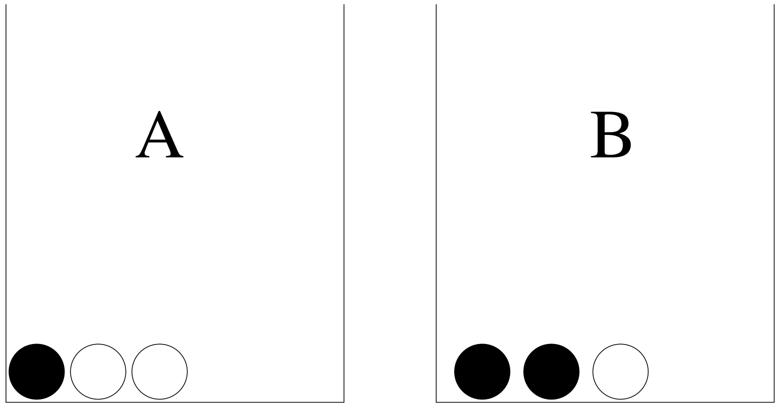
The likelihood principle

- Urn A contains three balls: one black, and two white; urn B contains three balls: two black, and one white.
- One of the urns is selected at random and one ball is drawn. The ball is black.



The likelihood principle

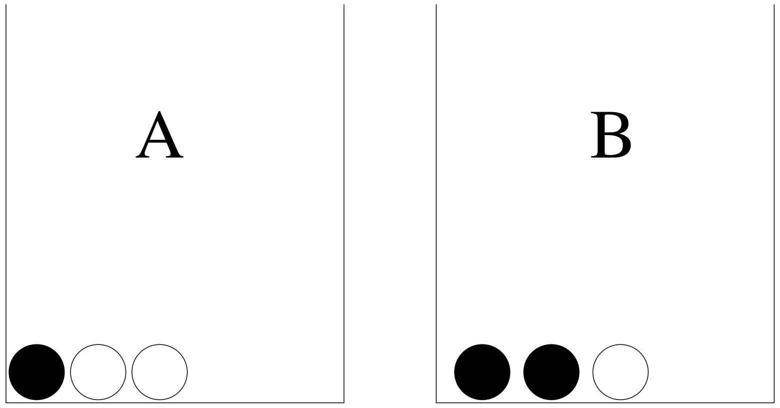
- Urn A contains three balls: one black, and two white; urn B contains three balls: two black, and one white.
- One of the urns is selected at random and one ball is drawn. The ball is black.
- **What is the probability that the selected urn is urn A**?



$$\boxed{P(u = A \mid b = bl) = ?}$$

The likelihood principle

- Urn A contains three balls: one black, and two white; urn B contains three balls: two black, and one white.

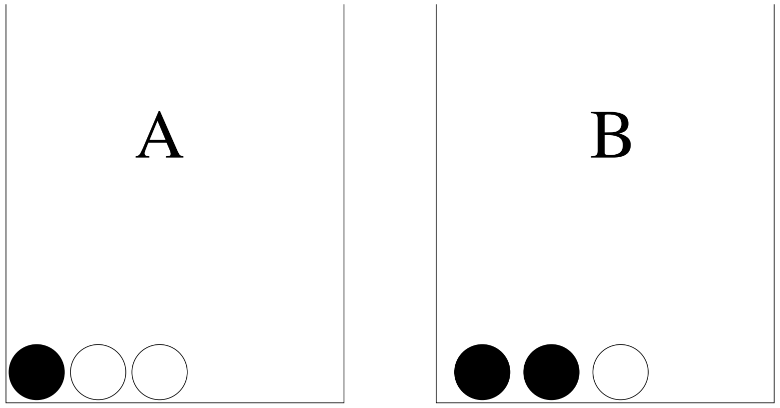


- One of the urns is selected at random and one ball is drawn. The ball is black.
- **What is the probability that the selected urn is urn A**?
- $u = \{A, B\}$; $P(u = A) = P(u = B) = 1/2$

$$\boxed{P(u = A \,|\, b = bl) = ?}$$

The likelihood principle

- Urn A contains three balls: one black, and two white; urn B contains three balls: two black, and one white.

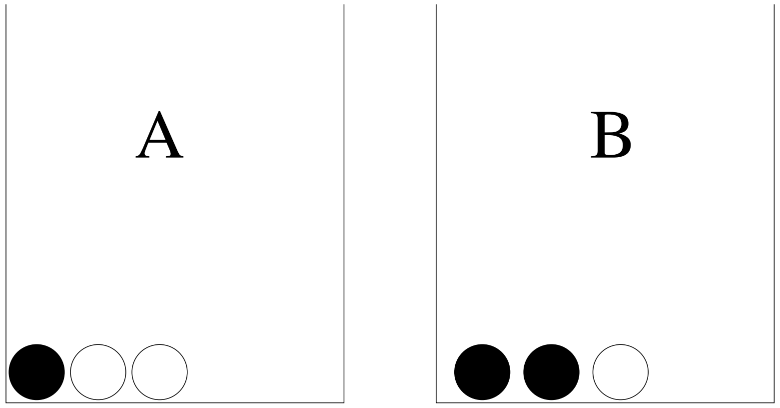


- One of the urns is selected at random and one ball is drawn. The ball is black.
- **What is the probability that the selected urn is urn A**?
- $u = \{\mathrm{A}, \mathrm{B}\}$; $P(u = \mathrm{A}) = P(u = \mathrm{B}) = 1/2$
- $b = \{\mathrm{bl}, \mathrm{wh}\}$ (black; white).

$$\boxed{P(u = A \mid b = bl) = ?}$$

The likelihood principle

- Urn A contains three balls: one black, and two white; urn B contains three balls: two black, and one white.
- One of the urns is selected at random and one ball is drawn. The ball is black.

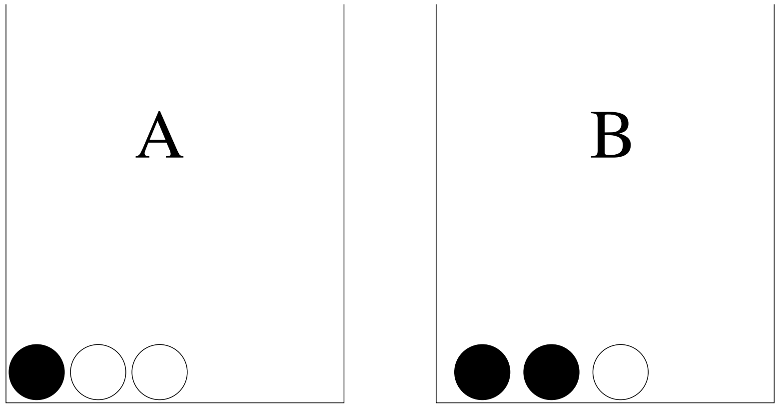


- **What is the probability that the selected urn is urn A**?
- u = {A, B}; $P(u$ = A) = $P(u$ = B) = 1/2
- b = {bl, wh} (black; white).
  - $P(b$ = bl | u = A) = 1/3; P(b = wh | u = A) = 2/3

$$\boxed{P(u = A \mid b = bl) = ?}$$

The likelihood principle

- Urn A contains three balls: one black, and two white; urn B contains three balls: two black, and one white.
- One of the urns is selected at random and one ball is drawn. The ball is black.



- **What is the probability that the selected urn is urn A**?
- $u = \{A, B\}$; $P(u = A) = P(u = B) = 1/2$
- $b = \{bl, wh\}$ (black; white).
  - $P(b = bl \mid u = A) = 1/3$; $P(b = wh \mid u = A) = 2/3$
  - $P(b = bl \mid u = B) = 2/3$; $P(b = wh \mid u = A) = 1/3$

$$\boxed{P(u = A \mid b = bl) = ?}$$

The likelihood principle

- One of the urns is selected at random and one ball is drawn.

  The ball is black

- $u = \{\mathrm{A}, \mathrm{B}\}$; $P(u = \mathrm{A}) = P(u = \mathrm{B}) = 1/2$
- $b = \{\mathrm{bl}, \mathrm{wh}\}$ (black; white).
  - $P(b = \mathrm{bl} \mid u = \mathrm{A}) = 1/3$; $\mathrm{P}(b = \mathrm{wh} \mid u = \mathrm{A}) = 2/3$
  - $P(b = \mathrm{bl} \mid u = \mathrm{B}) = 2/3$; $\mathrm{P}(b = \mathrm{wh} \mid u = \mathrm{A}) = 1/3$

The likelihood principle

- One of the urns is selected at random and one ball is drawn.

  The ball is black

- $u = \{A, B\}$; $P(u = A) = P(u = B) = 1/2$
- $b = \{bl, wh\}$ (black; white).
  - $P(b = bl \mid u = A) = 1/3$; $P(b = wh \mid u = A) = 2/3$
  - $P(b = bl \mid u = B) = 2/3$; $P(b = wh \mid u = A) = 1/3$

$$P(u = A \mid b = bl) = \frac{P(u = A, b = bl)}{P(b = bl)} = \frac{P(b = bl \mid u = A)P(u = A)}{P(b = bl)}$$

The likelihood principle

- One of the urns is selected at random and one ball is drawn.

  The ball is black

- u = {A, B}; $P(u$ = A) = $P(u$ = B) = 1/2
- b = {bl, wh} (black; white).
  - $P(b$ = bl | u = A) = 1/3; P(b = wh | u = A) = 2/3
  - $P(b$ = bl | u = B) = 2/3; P(b = wh | u = A) = 1/3

$$P(u = A \mid b = bl) = \frac{P(u = A, b = bl)}{P(b = bl)} = \frac{P(b = bl \mid u = A)P(u = A)}{P(b = bl)}$$

$$P(u = A \mid b = bl) = \frac{P(b = bl \mid u = A)P(u = A)}{P(b = bl \mid u = A)P(u = A) + P(b = bl \mid u = B)P(u = B)}$$

The likelihood principle

The likelihood principle: given a generative model for data d given parameters $\boldsymbol{\theta}$, $P(d \mid \boldsymbol{\theta})$, and having observed a particular outcome d_1, all inferences and predictions should depend only on the function $P(d_1 \mid \boldsymbol{\theta})$.

- The details of the other possible outcomes and their probabilities are irrelevant.
- All that matters is the probability of the outcome that actually happened given the different hypotheses.
- We need only to know the **likelihood**, i.e., **how the probability of the data that happened varies with the hypothesis.**

The likelihood principle

- One of the urns is selected at random and one ball is drawn.

  The ball is black

- u = {A, B}; $P(u$ = A) = $P(u$ = B) = 1/2
- b = {bl, wh} (black; white).
  - $P(b$ = bl | u = A) = 1/3; P(b = wh | u = A) = 2/3
  - $P(b$ = bl | u = B) = 2/3; P(b = wh | u = A) = 1/3

$$P(u = A \mid b = bl) = \frac{P(u = A, b = bl)}{P(b = bl)} = \frac{P(b = bl \mid u = A)P(u = A)}{P(b = bl)}$$

$$P(u = A \mid b = bl) = \frac{P(b = bl \mid u = A)P(u = A)}{P(b = bl \mid u = A)P(u = A) + P(b = bl \mid u = B)P(u = B)}$$

The likelihood principle

- One of the urns is selected at random and one ball is drawn.

The ball is black

- u = {A, B}; $P(u = \mathrm{A}) = P(u = \mathrm{B}) = 1/2$
- b = {bl, wh} (black; white).
  - $P(b = \mathrm{bl} \mid u = \mathrm{A}) = 1/3$; $\mathrm{P}(b = \mathrm{wh} \mid u = \mathrm{A}) = 2/3$
  - $P(b = \mathrm{bl} \mid u = \mathrm{B}) = 2/3$; $\mathrm{P}(b = \mathrm{wh} \mid u = \mathrm{A}) = 1/3$

how the probability of the **data that happened** varies with the hypothesis

$$P(u = A \mid b = bl) = \frac{P(u = A, b = bl)}{P(b = bl)} = \frac{P(b = bl \mid u = A)P(u = A)}{P(b = bl)}$$

$$P(u = A \mid b = bl) = \frac{P(b = bl \mid u = A)P(u = A)}{P(b = bl \mid u = A)P(u = A) + P(b = bl \mid u = B)P(u = B)}$$

The likelihood principle

- One of the urns is selected at random and one ball is drawn.

The ball is black

- $u = \{\mathrm{A}, \mathrm{B}\}$; $P(u = \mathrm{A}) = P(u = \mathrm{B}) = 1/2$
- $b = \{\mathrm{bl}, \mathrm{wh}\}$ (black; white).
  - $P(b = \mathrm{bl} \mid u = \mathrm{A}) = 1/3$; $\mathrm{P}(b = \mathrm{wh} \mid u = \mathrm{A}) = 2/3$
  - $P(b = \mathrm{bl} \mid u = \mathrm{B}) = 2/3$; $\mathrm{P}(b = \mathrm{wh} \mid u = \mathrm{A}) = 1/3$

how the probability of the **data that happened** varies with the hypothesis

$$P(u = A \mid b = bl) = \frac{P(u = A, b = bl)}{P(b = bl)} = \frac{P(b = bl \mid u = A) P(u = A)}{P(b = bl)}$$

$$P(u = A \mid b = bl) = \frac{P(b = bl \mid u = A) P(u = A)}{P(b = bl \mid u = A) P(u = A) + P(b = bl \mid u = B) P(u = B)}$$

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Further Reading and Summary

Q&A

Further Reading

- **Recommend Readings**
  - Information Theory, Inference, and Learning Algorithms from David MacKay, 2015, pages 22 - 32.
- **Supplemental readings:**

What you should know

- Joint Probability; Marginal Probability; Condicional Probability.
- Product and Sum rules.
- Bayes' Theorem
- Statistical Independence
- The two common but different interpretations for the meaning of Probabilities
- Forward Probabilities and Inverse Probabilities
- Terminology of inverse Probability: prior probability; likelihood; likelihood; evidence
- The likelihood principle
- To address Inverse probability problems

Further Reading and Summary

Q&A