Bayes' Theorem

Jan Kracík

Department of Applied Mathematics
FEECS, VŠB - TU Ostrava

Introduction

Bayes' theorem

- fundamental theorem in probability theory
- named after reverend Thomas Bayes (1701–1761)
- discovered in Bayes' work and published by Richard Price (1723-1791)
- further developed by Pierre-Simon Laplace (1749-1827)
- forms a cornerstone of statistical learning

Measure - motivation

Area of a plane region
Volume of a solid
Weight of a solid
Count of elements
Price of goods

Measure - motivation

Area of a plane region
Volume of a solid
Weight of a solid
Count of elements
Price of goods

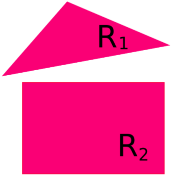
} **measures** in a common sense

Measure - area

Area of a plane region

- $A(R) \ldots$ area of a region R (set)
- $A(R) \geq 0$ for any region R
- $A(\emptyset) = 0$
- If $R_1 \cap R_2 = \emptyset$ then

$$A(R_1 \cup R_2) = A(R_1) + A(R_2)$$

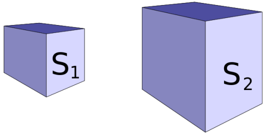


Measure - volume

Volume of a solid

- $V(S)\ldots$ volume of a solid S (set)
- $V(S) \geq 0$ for any solid S
- $V(\emptyset) = 0$
- If $S_1 \cap S_2 = \emptyset$ then

$$V(S_1 \cup S_2) = V(S_1) + V(S_2)$$

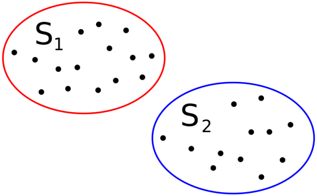


Measure - count

Count of elements of a set

- $C(S)\ldots$ count of elements of a set S
- $C(S) \geq 0$ for any set S
- $C(\emptyset) = 0$
- If $S_1 \cap S_2 = \emptyset$ then

  $C(S_1 \cup S_2) = C(S_1) + C(S_2)$



Measure

What are the common properties of the “measures”?

Measure

What are the common properties of the “measures”?

Let Ω be an arbitrary set and $\mathcal{F}$ be a set of (all) subsets of Ω.

Measure

What are the common properties of the "measures"?

Let Ω be an arbitrary set and $\mathcal{F}$ be a set of (all) subsets of Ω.

Measure is a function $M : \mathcal{F} \to [0, +\infty) \cup \{+\infty\}$ such that

- for all $S \in \mathcal{F}$: $M(S) \geq 0$,
- $M(\emptyset) = 0$,
- if $S_1, S_2 \in \mathcal{F}$ and $S_1 \cap S_2 = \emptyset$ then $M(S_1 \cup S_2) = M(S_1) + M(S_2)$

Probability

From mathematical point of view,

a probability is a measure.

Probability

Random experiment - outcome of the experiment cannot be completely determined beforehand.

Probability

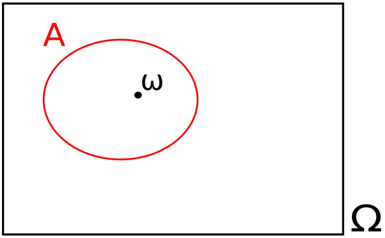
Random experiment - outcome of the experiment cannot be completely determined beforehand.

- Sample space Ω . . . set of all possible outcomes
- Random event $A \subset \Omega$
  A occurs if and only if the outcome $\omega \in A$

Probability

Random experiment - outcome of the experiment cannot be completely determined beforehand.

- Sample space Ω ... set of all possible outcomes
- Random event $A \subset \Omega$
  A occurs if and only if the outcome $\omega \in A$

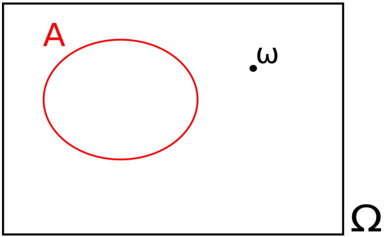


A occurs

Probability

Random experiment - outcome of the experiment cannot be completely determined beforehand.

- Sample space Ω . . . set of all possible outcomes
- Random event $A \subset \Omega$
  A occurs if and only if the outcome $\omega \in A$

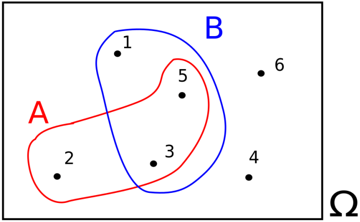


A does not occur

Random experiment - example

Random experiment: rolling a dice

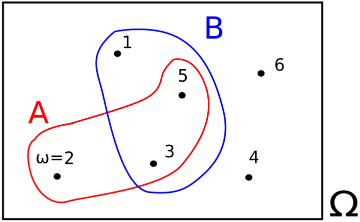
- Sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Random event A ... prime number is rolled
- Random event B ... odd number is rolled



Random experiment - example

Random experiment: rolling a dice

- Sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Random event A ... prime number is rolled
- Random event B ... odd number is rolled



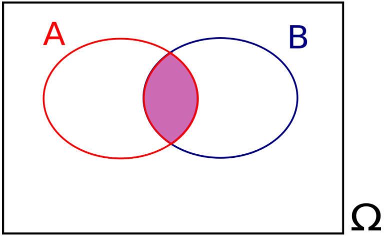
Result: $\omega = 2$. A occurred; B did not occur.

Let A, B be random events.

- $A \cap B$ occurs if and only if A occurs **and** B occurs (intersection)
- $A \cup B$ occurs if and only if A occurs **or** B occurs (union)
- $\overline{A}$ occurs if and only if A **does not** occur (complement)

Let A, B be random events.

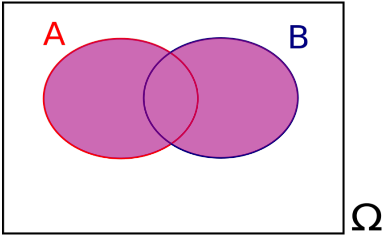
- $A \cap B$ occurs if and only if A occurs **and** B occurs (intersection)
- $A \cup B$ occurs if and only if A occurs **or** B occurs (union)
- $\overline{A}$ occurs if and only if A **does not** occur (complement)



$A \cap B$

Let A, B be random events.

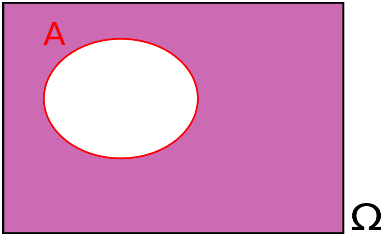
- $A \cap B$ occurs if and only if A occurs **and** B occurs (intersection)
- $A \cup B$ occurs if and only if A occurs **or** B occurs (union)
- $\overline{A}$ occurs if and only if A **does not** occur (complement)



$A \cup B$

Let A, B be random events.

- $A \cap B$ occurs if and only if A occurs **and** B occurs (intersection)
- $A \cup B$ occurs if and only if A occurs **or** B occurs (union)
- $\overline{A}$ occurs if and only if A **does not** occur (complement)



$\overline{A}$

Let $\mathcal{F}$ be the set of all random events in a sample space Ω.

Let $\mathcal{F}$ be the set of all random events in a sample space Ω.

Probability

Probability is a measure on $\mathcal{F}$ such that $P(\Omega) = 1$.

Let $\mathcal{F}$ be the set of all random events in a sample space Ω.

Probability

Probability is a measure on $\mathcal{F}$ such that $P(\Omega) = 1$.

$P(A)$ is interpreted as a measure of belief that A occur.

Let $\mathcal{F}$ be the set of all random events in a sample space Ω.

Probability

Probability is a measure on $\mathcal{F}$ such that $P(\Omega) = 1$.

$P(A)$ is interpreted as a measure of belief that A occur.

From the properties of measure it follows:

- $P(\emptyset) = 0$
- $P(\overline{A}) = 1 - P(A)$
- ...

Let $\mathcal{F}$ be the set of all random events in a sample space Ω.

Probability

Probability is a measure on $\mathcal{F}$ such that $P(\Omega) = 1$.

$P(A)$ is interpreted as a measure of belief that A occur.

From the properties of measure it follows:

- $P(\emptyset) = 0$
- $P(\overline{A}) = 1 - P(A)$
- ......

Let $\mathcal{F}$ be the set of all random events in a sample space Ω.

Probability

Probability is a measure on $\mathcal{F}$ such that $P(\Omega) = 1$.

$P(A)$ is interpreted as a measure of belief that A occur.

From the properties of measure it follows:

- $P(\emptyset) = 0$
- $P(\overline{A}) = 1 - P(A)$
- ....

Let $\mathcal{F}$ be the set of all random events in a sample space Ω.

Probability

Probability is a measure on $\mathcal{F}$ such that $P(\Omega) = 1$.

$P(A)$ is interpreted as a measure of belief that A occur.

From the properties of measure it follows:

- $P(\emptyset) = 0$
- $P(\overline{A}) = 1 - P(A)$
- ....

Let $\mathcal{F}$ be the set of all random events in a sample space Ω.

Probability

Probability is a measure on $\mathcal{F}$ such that $P(\Omega) = 1$.

$P(A)$ is interpreted as a measure of belief that A occur.

From the properties of measure it follows:

- $P(\emptyset) = 0$
- $P(\overline{A}) = 1 - P(A)$
- ....

Let $\mathcal{F}$ be the set of all random events in a sample space Ω.

Probability

Probability is a measure on $\mathcal{F}$ such that $P(\Omega) = 1$.

$P(A)$ is interpreted as a measure of belief that A occur.

From the properties of measure it follows:

- $P(\emptyset) = 0$
- $P(\overline{A}) = 1 - P(A)$
- ....

Let $\mathcal{F}$ be the set of all random events in a sample space Ω.

Probability

Probability is a measure on $\mathcal{F}$ such that $P(\Omega) = 1$.

$P(A)$ is interpreted as a measure of belief that A occur.

From the properties of measure it follows:

- $P(\emptyset) = 0$
- $P(\overline{A}) = 1 - P(A)$
- ....

Let $\mathcal{F}$ be the set of all random events in a sample space Ω.

Probability

Probability is a measure on $\mathcal{F}$ such that $P(\Omega) = 1$.

$P(A)$ is interpreted as a measure of belief that A occur.

From the properties of measure it follows:

- $P(\emptyset) = 0$
- $P(\overline{A}) = 1 - P(A)$
- ....

Let $\mathcal{F}$ be the set of all random events in a sample space Ω.

Probability

Probability is a measure on $\mathcal{F}$ such that $P(\Omega) = 1$.

$P(A)$ is interpreted as a measure of belief that A occur.

From the properties of measure it follows:

- $P(\emptyset) = 0$
- $P(\overline{A}) = 1 - P(A)$
- ....

Conditional probability

Let A, B be random events and $P(B) > 0$.

Conditional probability of A given B

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

$P(A|B)$ is a measure of belief that A occurs if B **occurred.**

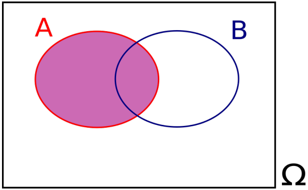
Conditional probability

Let A, B be random events and $P(B) > 0$.

Conditional probability of A given B

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

$P(A|B)$ is a measure of belief that A occurs if B **occurred.**



$P(A)\ldots$ "size" of A (relative to the "size" of Ω)

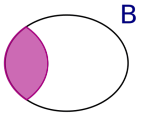
Conditional probability

Let A, B be random events and $P(B) > 0$.

Conditional probability of A given B

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

$P(A|B)$ is a measure of belief that A occurs if B **occurred.**



$P(A|B)\ldots$ "size" of $A \cap B$ relative to the "size" of B

Bayes' Theorem

Let A, B be random events and $P(A) > 0$, $P(B) > 0$.

Bayes' Theorem

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

Bayes' Theorem

Let A, B be random events and $P(A) > 0$, $P(B) > 0$.

Bayes' Theorem

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

Bayes' Theorem

Let A, B be random events and $P(A) > 0$, $P(B) > 0$.

Bayes' Theorem

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

To evaluate $P(A|B)$ we need $P(A)$, $P(B|A)$, and $P(B|\overline{A})$.

Bayes' Theorem

Let A, B be random events and $P(A) > 0, P(B) > 0$.

Bayes' Theorem

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

To evaluate $P(A|B)$ we need $P(A)$, $P(B|A)$, and $P(B|\overline{A})$.

Prior and posterior probability

$P(A)$ ... **prior** belief that A occur
$P(A|B)$ ... **posterior** belief that A occur if B *occurred*

Occurrence of B (partial information) brings an information about A.

Bayes' Theorem

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

Bayes' Theorem represents a fundamental principle of statistical learning.

Example 2: Image Processing - Finding Charlie

Example 2: Image Processing - Finding Charlie

Charlie was kidnapped last night. Please help!

Example: Image Processing

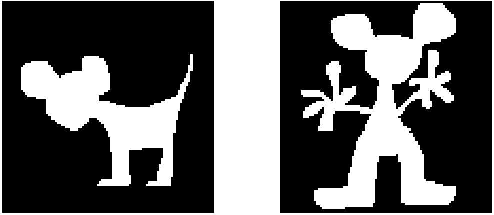
Consider a binary digital image (black and white pixels) of a scene.

- White areas . . . foreground objects
- Black areas . . . background of the scene.
- Real-world scene

Example: Image Processing

Consider a binary digital image (black and white pixels) of a scene.

- White areas . . . foreground objects
- Black areas . . . background of the scene.
- Real-world scene: small number of large objects is more likely than large number of small objects



For example ...

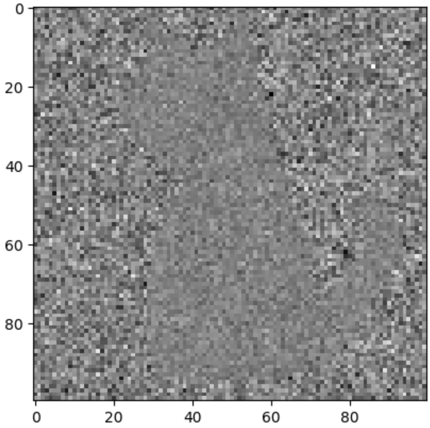
- Original binary image (unobserved): 100×100 pixels

- Observed gray-scale image: 100×100 pixels

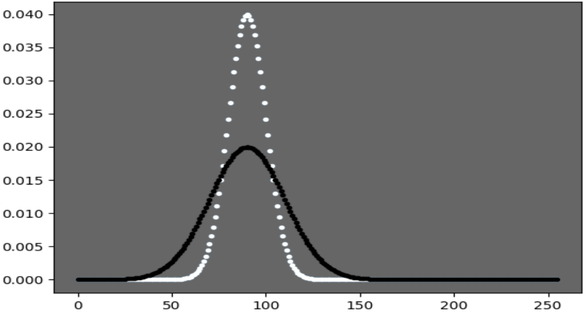
- Original binary image (unobserved): 100×100 pixels
  Random events: $A_1, A_2, \ldots, A_{2^{10000}}$
  A_k occurs if and only if the original image is the k-th one.
- Observed gray-scale image: 100×100 pixels

- Original binary image (unobserved): 100×100 pixels
  Random events: $A_1, A_2, \ldots, A_{2^{10000}}$
  A_k occurs if and only if the original image is the k-th one.
- Observed gray-scale image: 100×100 pixels
  Random events: $B_1, B_2, \ldots, B_{256^{10000}}$
  B_n occurs if and only if the observed image is the n-th one.

Observed image . . . B_O



- Select **observation model:** $P(B_O|A_k)$ for $k = 1, 2, \ldots, 2^{10000}$
  relates the observation to the original image
  (different Gaussian noise for black and white pixels)



Naive approach: Uniform prior

$$P(A_1) = P(A_2) = \cdots = P(A_{2^{10000}}) = \frac{1}{2^{10000}}$$

The image with highest posterior probability

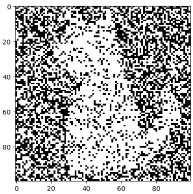
$$P(A_E|B_O) = \max_k P(A_k|B_O)$$

Naive approach: Uniform prior

$$P(A_1) = P(A_2) = \cdots = P(A_{2^{10000}}) = \frac{1}{2^{10000}}$$

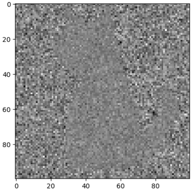
The image with highest posterior probability

$$P(A_E|B_O) = \max_k P(A_k|B_O)$$



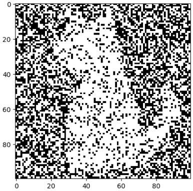
Naive prior:

$$P(A_1) = P(A_2) = \cdots = P(A_{2^{10000}}) = \frac{1}{2^{10000}}$$



original

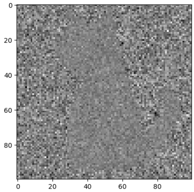
$\times$



result

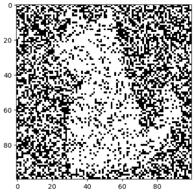
Naive prior:

$$P(A_1) = P(A_2) = \cdots = P(A_{2^{10000}}) = \frac{1}{2^{10000}}$$



original

×



result

:-(

Sophisticated approach: select prior probabilities so that they reflect the prior knowledge:

Small number of large objects is more likely than large number of small objects.

Sophisticated approach: select prior probabilities so that they reflect the prior knowledge:

Small number of large objects is more likely than large number of small objects.

$$P(A_k) = c \exp(-H(A_k))$$

- c ... constant
- $H(A_k)$ is the number of adjacent pixels with different color in image A_k
- $H(A_k)$ equals to the length of the border between black and white areas

Visualized posterior probability (of individual pixels being white)

3

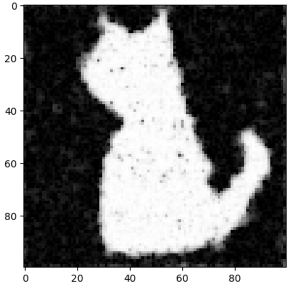
Visualized posterior probability (of individual pixels being white)

2

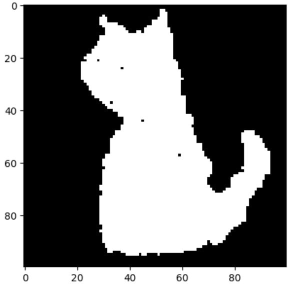
Visualized posterior probability (of individual pixels being white)

1

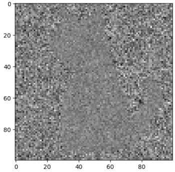
Visualized posterior probability (of individual pixels being white)

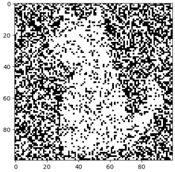


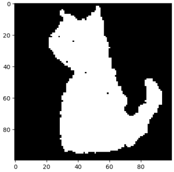
Visualized estimated picture



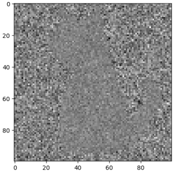
Compare the results

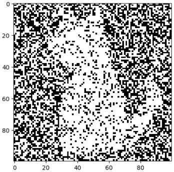


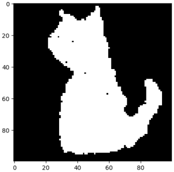




Compare the results

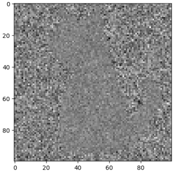


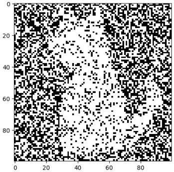


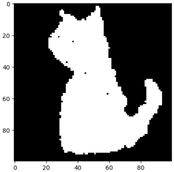


:-)

Compare the results







Thank you for your attention!

The original image