Information Theory

By: Prof. Mohamed Hamada
Software Engineering Lab.
The University of Aizu

Evaluation

- Attendance and class activities 20%
- Midterm Exam 35%
- Final Exam 45%

Contact

- Send e-mail to

  hamada@u-aizu.ac.jp
- Office: Room 346-C
- Course materials at

www.u-aizu.ac.jp/~hamada/education.html

Check every week for update

Goals

- Understand the concepts of information entropy and channel capacity
- Understand the digital communication model and its components
- Understand how the components operate
- Understand data compression
- Understand error detection and correction

Course Outline

- Introduction to set theory & probability
- Introduction to information theory
- Coding techniques & data compression
- Information Entropy
- Communication Channel
- Error Detection and Correction

Today's Outline

Introduction to Set Theory and Probability

- 1. Sets, Operations on sets
- 2. Trial, Probability space, Events
- 3. Random variables, Probability distribution
- 4. Expected values, Variance
- 5. Conditional Probability
- 6. Bayes Theory

Sets

Sets

A set is a collection of objects without repetition. The order of elements is irrelevant.

A set can be expressed by writing all elements

Example : even = $\{ 0, 2, 4, 6, 8, 10 \}$

OR can be expressed by suing a common property of its elements

Example : even = $\{x : 0 \leq x \leq 10 \text{ and } x \text{ is even} \}$

Sets

Elements of sets :

We use the symbols $\in$ and $\notin$ to show that an element belongs to a set or not

Example : even = { 0, 2, 4, 6, 8, 10 }

$2 \in$ even

$3 \notin$ even

Subsets : A subset of a set A is a collection of elements that all belongs to A

Example : S = {2,4} $\Rightarrow$ S $\subset$ even

T = {1, 4 } $\Rightarrow$ T $\not\subset$ even

Empty Set : is the set of no elements $\phi = \{ \ \}$

Operations on sets

Set Union

$$A \cup B = \{ x \mid x \in A \text{ or } x \in B \}$$

Commutative $A \cup B = B \cup A$

Example : even = {0, 2, 4, 6 }
odd = {1, 3, 5 }

even $\cup$ odd = {0, 1, 2, 3, 4, 5, 6}

Operations on sets

Set Intersection

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}$$

Commutative $A \cap B = B \cap A$

Example: even = $\{0, 2, 4, 6\}$
odd = $\{1, 3, 5\}$

even $\cap$ odd $= \phi$

even $\cap \{1, 2, 3\} = \{2\}$

Operations on sets

Set Subtraction

$$A - B = \{x \mid x \in A \text{ and } x \notin B\}$$

Non commutative $\quad A - B \neq B - A$

Example : even = {0, 2, 4, 6 }
odd = {1, 3, 5 }

even – odd = even

even – {1, 2, 3} = {0, 4, 6 }

Operations on sets

Power Set

Power Set: is the set of all subsets

$$P(A) = 2^A = \{B \mid B \subseteq A\}$$

Example: odd = {1, 3, 5 }

$$P(odd) = 2^{odd} = \{ \phi, \{1\}, \{3\}, \{5\}, \{1,3\}, \{1,5\}, \{3,5\}, odd \}$$

Cardinality

Cardinality is the number (#) of elements in a set

Example: $\# (odd) = 3$, $\# (2^{odd}) = 8 = 2^3 = 2^{\#(odd)}$

Trials

Many problems in probabilities and statistics involve situations in which an experiment with the possible outcomes is repeated several times.

Each repetition of the experiment is called a ***trail***.

Example : throw a dice

Probability

Probability:

A probability is a number associated with or assigned to a set in order to measure it in some sense

Probability Space:

Consider the symbols Ω and P where

Ω : universal set ;

P : is a probability function mapping power set to reals in the interval [0,1]; i.e. $P: 2^{\Omega} \rightarrow [0,1]$

With the following 3 properties:

(1) $P(\emptyset) = 0$

(2) $P(\Omega) = 1$

(3) $\forall A = \{a_1, a_2, \ldots, a_m\} \subset \Omega$,

$$P(A) = \sum_{n=1}^{m} P(a_n),$$

where $P(a_n) = P(\{a_n\})$

Event

The subset of Ω to which a probability has been assigned is called an event

An event is a classification of trial outcome.

Example :

The set of events of throwing a dice is {1,2,3,4,5,6}

Random Variables

Random Variable is a numerical quantity whose value depends on chance.

Example : A person to be selected at random

For a universal set Ω and a $\in \Omega$

The function $X : \Omega \to$ set of numerical values is called a random variable

Example : Throwing 2 dice we get

$\Omega = \{ (m,n) \mid m,n \in \{1,2,\ldots, 6\} \}$

$$\Omega = \begin{matrix} (1,1), (1,2), (1,3), (1,4), (1,5), (1,6) \\ (2,1), (2,2), (2,3), (2,4), (2,5), (2,6) \\ (3,1), (3,2), (3,3), (3,4), (3,5), (3,6) \\ (4,1), (4,2), (4,3), (4,4), (4,5), (4,6) \\ (5,1), (5,2), (5,3), (5,4), (5,5), (5,6) \\ (6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \end{matrix}$$

Random Variables

Proposition

If X and Y are random variables then $X+Y$, $X-Y$, XY and X/Y $(Y \neq 0)$ are random variables

Probability Distribution

The probability that the random variable X takes the value a_n is

$$p_n = P(X = a_n)$$
$$p_n \geq 0, \quad \sum p_n = 1$$

If certainty is 100% then probability = 1 (unity)

In general $\forall n \qquad 0 \leq p_n \leq 1$

Example : fair coin front ↑

back ↓

$$\begin{pmatrix} \uparrow & \downarrow \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Probability Distribution

Example : fair dice

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

Example of random variable : sum of two dice throws

Consider the random variable X_1 as outcome of the first dice:
$X_1((m, n)) = m$

Consider the random variable X_2 as outcome of the second dice:
$X_2((m, n)) = n$

Consider the random variable X_3 as sum of X_1 and X_2 ($X_3=X_1+X_2$):
$X_3((m, n)) = m + n$

+	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Probability distribution of X_1

Probability distribution of X_2

Probability distribution of X_3

Expected value (mean or average value)

The Expected value is designed as fallows :

Take the value of (the random variable) X at each point $a \in \Omega$, multiply it by the probability of that point (i.e P(a)) and sum over all $a \in \Omega$

$$E(X) = \sum_{a \in \Omega} X(a) \; P(a)$$

Note that:

$$E(X + Y) = E(X) + E(Y)$$

$$E(cX) = c\,E(X) \qquad \text{(for constant c)}$$

Expected value (mean or average value)

$$E(X) = \sum_{a \in \Omega} X(a) \; P(a)$$

Example

Consider a land with 100 acre with 50% of price \$150 per acre, 30% of price \$100 per acre and 20% of price \$50 per acre.

30% \$100	20% \$50
50%	\$150

What is the average (expected or mean) price per acre for the whole land?

Answer:

Ω = {50%, 30%, 20% } = {50/100, 30/100, 20/100}

E (X) = 150 * 50/100 + 100 * 30/100 + 50 * 20/100 = 115 \$ per acre

Variance

The Variance of a variable X (denoted by V(X) or σ^2 (X) is defined by

$$V(X) = E[(X - E(X))^2]$$

Note that: for a constant c

$$\begin{aligned} V(c) &= E[(c - E(c))^2] \\ &= E[(c - c)^2] \\ &= 0 \end{aligned}$$

$$\begin{aligned} V(cX) &= E[(cX - E(cX))^2] \\ &= E[(cX - cE(X))^2] \\ &= E[c^2(X - E(X))^2] \\ &= c^2 E[(X - E(X))^2] \\ &= c^2 V(X) \end{aligned}$$

Variance

Example:

For a fair coin toss (with front denoted by ↑ and back denoted by ↓) we have

	Ω				
	(↑,↑)	(↑,↓)	(↓,↑)	(↓,↓)	∑
p	¼	¼	¼	¼	1
x = # (↑)	2	1	1	0	4
x P(x)	2/4	¼	¼	0/4	E[X] = 1
(x- E[x])^2 P(x)	$(2-1)^2/4$	$(1-1)^2/4$	$(1-1)^2/4$	$(0-1)^2/4$	V[X] = 1/2

Exercise

Prove this equality

$$V(X) = E(X^2) - E(X)^2$$

Exercise: Answer

Prove this equality $V(X) = E(X^2) - E(X)^2$

Answer:

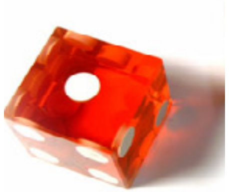
$$
\begin{aligned}
V(X) &= E[(X - E(X))^2] \\
&= E[(X - E(X))(X - E(X))] \\
&= E[X^2 - 2XE(X) + E(X)^2] \\
&= E(X^2) - 2E(X)^2 + E(X)^2 \\
&= E(X^2) - E(X)^2
\end{aligned}
$$

Conditional Probability

Conditional probability is the probability of A when B is known

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

Example:



If a dice is rolled:

Let an event A = the dice comes up with 5 = {5}

Let an event B = the dice comes up odd = {1,3,5}

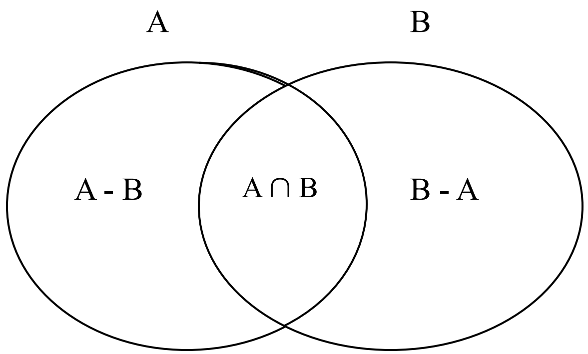
Then: P(A)=1/6, P(B)=P({1,3,5})=P({1})+ P({3})+ P({5})=1/6+ 1/6+ 1/6=1/2

$A \cap B = \{5\}$, $P(A \cap B)=P(\{5\})=1/6$

Hence $P(A \mid B) = P(A \cap B) / P(B) = (1/6) / (1/2) = 1/3$

Conditional Probability

Note that :



$$P(B) = P(A \cap B) + P(B - A) \geq P(A \cap B) \quad \Rightarrow \quad P(A \mid B) \leq 1$$

$$\text{If } B \subseteq A \Rightarrow A \cap B = B \Rightarrow P(A \mid B) = 1$$

Bayes Theorem

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

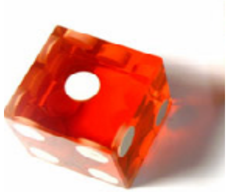
Knowing the outcome of a particular situation, one (using Bayes theorem) can find the probability that the outcome occurred as a result of a particular previous event.

Note that:

1. in general $P(A \mid B) \neq P(B \mid A)$
2. $P(A \mid B) = P(B \mid A)$ only if $P(A) = P(B)$

Bayes Theorem

Example: **Throw a single dice: let A = { x | x odd }, B = { 1,2 }**



$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{2/6} = 1/2$$

$$P(B \mid A) = \frac{P(B)\,P(A \mid B)}{P(A)} = \frac{(2/6)\,(1/2)}{3/6} = 1/3$$

Bayes Theorem

Properties:

1. $$P(\phi|B) = \frac{P(\phi \cap B)}{P(B)} = \frac{P(\phi)}{P(B)} = \frac{0}{P(B)} = 0$$

2. $$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$

3. $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\sum_{a\in A} P(\{a\} \cap B)}{P(B)} = \sum_{a\in A} \frac{P(\{a\} \cap B)}{P(B)} = \sum_{a\in A} P(\{a\}|B)$$

Independence of Random Variables

Two discrete random variables, X_1 and X_2 are independent if

$$\forall \text{ value } j, k \quad P(X_1 = j \cap X_2 = k) = P(X_1 = j)\ P(X_2 = K)$$

Note that:

If X_1 and X_2 are independent random variables then it is easy to show that:

$$\mathbf{P(X_1 = j \mid X_2) = P(X_1 = j)}$$

END