Formal Modeling in Cognitive Science

Lecture 19: Application of Bayes' Theorem; Discrete Random Variables; Distributions

Steve Renals (notes by Frank Keller)

School of Informatics
University of Edinburgh
s.renals@ed.ac.uk

22 February 2007

1. Application of Bayes' Theorem
   - Background
   - Application to Diagnosis
   - Base Rate Neglect
2. Discrete Random Variables
3. Distributions
   - Probability Distributions
   - Cumulative Distributions

Background

Let's look at an application of Bayes' theorem to the analysis of cognitive processes. First we need to introduce some data.

Research on human decision making investigates, e.g., how physicians make a *medical diagnosis* (Casscells et al. 1978):

Example

If a test to detect a disease whose prevalence is 1/1000 has a false-positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person's symptoms or signs?

Background

Most frequent answer: 95%

Reasoning: if false-positive rate is 5%, then test will be correct 95% of the time.

Correct answer: 2%

Reasoning: assume you test 1000 people; the test will be positive in 50 cases (5%), but only one person actually has the disease. Hence the chance that a person with a positive result has the disease is $1/50 = 2\%$.

Only 12% of subjects give the correct answer.

Mathematics underlying the correct answer: Bayes' Theorem.

Bayes' Theorem

We need to think about Bayes' theorem slightly differently to apply it to this problem (and the terms have special names now):

Bayes' Theorem (for hypothesis testing)

Given a hypothesis h and data D which bears on the hypothesis:

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

$P(h)$: independent probability of h: *prior probability*
$P(D)$: independent probability of D
$P(D|h)$: conditional probability of D given h: *likelihood*
$P(h|D)$: conditional probability of h given D: *posterior probability*

Application to Diagnosis

In Casscells et al.'s (1978) examples, we have the following:

- h: person tested has the disease;
- $\bar{h}$: person tested doesn't have the disease;
- D: person tests positive for the disease.

The following probabilities are known:

$$P(h) = 1/1000 = 0.001 \qquad P(\bar{h}) = 1 - P(h) = 0.999$$
$$P(D|\bar{h}) = 5\% = 0.05 \qquad P(D|h) = 1 \text{ (assume perfect test)}$$

Compute the probability of the data (rule of total probability):

$$P(D) = P(D|h)P(h) + P(D|\bar{h})P(\bar{h}) = 1 \cdot 0.001 + 0.05 \cdot 0.999 = 0.05095$$

Compute the probability of correctly detecting the illness:

$$P(h|D) = \frac{P(h)P(D|h)}{P(D)} = \frac{0.001 \cdot 1}{0.05095} = 0.01963$$

Base Rate Neglect

Base rate: the probability of the hypothesis being true in the absence of any data (i.e., prior probability).

Base rate neglect: people have a tendency to ignore base rate information (see Casscells et al.'s (1978) experimental results).

- base rate neglect has been demonstrated in a number of experimental situations;
- often presented as a fundamental bias in decision making;
- however, experiments show that subjects use base rates in certain situations;
- it has been argued that base rate neglect is only occurs in artificial or abstract mathematical situations.

Base Rates and Experience

Potential problems with in Casscells et al.'s (1978) study:

- subjects were simply told the statistical facts;
- they had no first-hand experience with the facts (through exposure to many applications of the test);
- providing subjects with experience has been shown to reduce or eliminate base rate neglect.

Medin and Edelson (1988) tested the role of experience on decision making in medical diagnosis.

Base Rates and Experience

Medin and Edelson (1988) trained subjects on a diagnosis task in which diseases varied in frequency:

- subjects were presented with pairs of symptoms and had to select one of six diseases;
- feedback was provided so that they learned symptom/disease associations;
- base rates of the diseases were manipulated;
- once subjects had achieved perfect diagnosis accuracy, they entered the transfer phase;
- subjects now made diagnoses for combinations of symptoms they had not seen before; *made use of base rate information.*

Discrete Random Variables

Definition: Random Variable

If S is a sample space with a probability measure and X is a real-valued function defined over the elements of S, then X is called a random variable.

We will denote random variable by capital letters (e.g., X), and their values by lower-case letters (e.g., x).

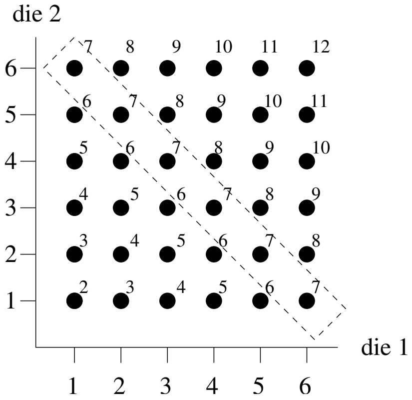
Example

Given an experiment in which we roll a pair of dice, let the random variable X be the total number of points rolled with the two dice.

For example $X = 7$ picks out the set $\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$.

Discrete Random Variables

This can be illustrated graphically:



For each outcome, this graph lists the value of X, and the dotted area corresponds to $X = 7$.

Discrete Random Variables

Example

Assume a balanced coin is flipped three times. Let X be the random variable denoting the total number of heads obtained.

Outcome	Probability	x
HHH	$\frac{1}{8}$	3
HHT	$\frac{1}{8}$	2
HTH	$\frac{1}{8}$	2
THH	$\frac{1}{8}$	2

Outcome	Probability	x
TTH	$\frac{1}{8}$	1
THT	$\frac{1}{8}$	1
HTT	$\frac{1}{8}$	1
TTT	$\frac{1}{8}$	0

Hence, $P(X = 0) = \frac{1}{8}$, $P(X = 1) = P(X = 2) = \frac{3}{8}$, $P(X = 3) = \frac{1}{8}$.

Probability Distributions

Definition: Probability Distribution

If X is a discrete random variable, the function given by $f(x) = P(X = x)$ for each x within the range of X is called the probability distribution of X.

Theorem: Probability Distribution

A function can serve as the probability distribution of a discrete random variable X if and only if its values, $f(x)$, satisfy the conditions:

1. $f(x) \geq 0$ for each value within its domain;
2. $\sum_x f(x) = 1$, where x over all the values within its domain.

Probability Distributions

Example

For the probability function defined in the previous example:

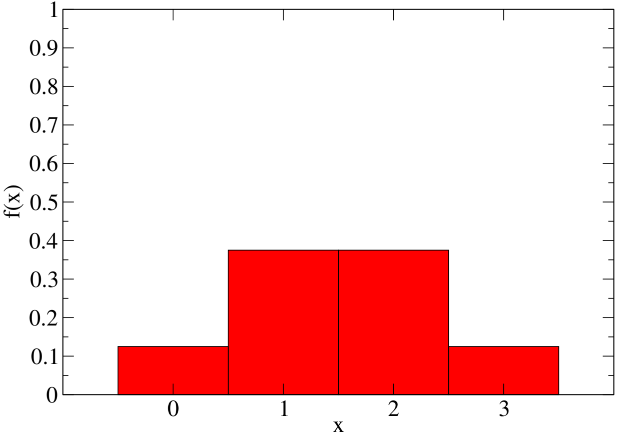
x	$f(x) = P(X = x)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

This function can be written more concisely as:

$$f(x) = \frac{4 - |3 - 2x|}{8}$$

Probability Distributions

A probability distribution is often represented as a *probability histogram*. For the previous example:



Cumulative Distribution

In many cases, we're interested in the probability for values $X \leq x$, rather than for $X = x$.

Definition: Cumulative Distribution

If X is a discrete random variable, the function given by:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t) \text{ for } -\infty < x < \infty$$

where $f(t)$ is the value of the probability distribution of X at t, is the cumulative distribution of X.

Cumulative Distributions

Example

Consider the probability distribution $f(x) = \frac{4-|3-2x|}{8}$ from the previous example. The values of the cumulative distribution are:

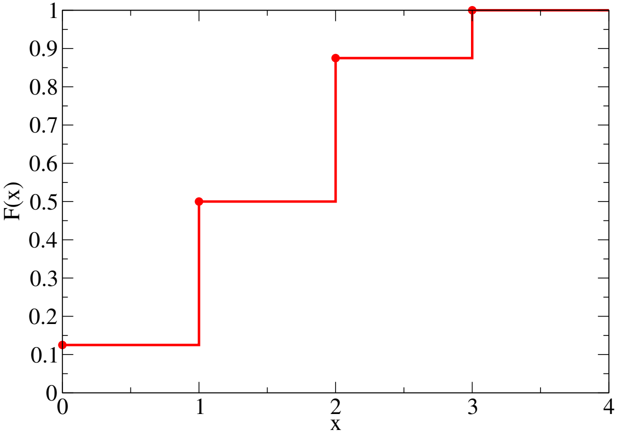
x	$f(x)$	F(x)
0	$\frac{1}{8}$	$\frac{1}{8}$
1	$\frac{3}{8}$	$\frac{4}{8}$
2	$\frac{3}{8}$	$\frac{7}{8}$
3	$\frac{1}{8}$	$\frac{8}{8}$

Note that $F(x)$ is defined for all real values of x:

$$F(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{8} & \text{for } 0 \leq x < 1 \\ \frac{4}{8} & \text{for } 1 \leq x < 2 \\ \frac{7}{8} & \text{for } 2 \leq x < 3 \\ \frac{8}{8} & \text{for } x \geq 3 \end{cases}$$

Cumulative Distributions

The cumulative distribution can be graphed; for the previous example:



Cumulative Distributions

Theorem: Cumulative Distributions

The values $F(x)$ of the cumulative distribution of a discrete random variable X satisfies the conditions:

1. $F(-\infty) = 0$ and $F(\infty) = 1$;
2. if $a < b$, then $F(a) \leq F(b)$ for any real numbers a and b.

Example

Consider the example of $F(x)$ on the previous slide:

1. $F(-\infty) = 0$ as $F(0) < 0$ by definition; $F(\infty) = 1$ as $F(\infty) \geq 3$ by definition;
2. $F(a) < F(b)$ holds for $(0, 1)$, $(1, 2)$, $(2, 3)$ by definition; $F(a) = F(b)$ holds for all other values of a and b.

Summary

- There are many applications of Bayes' theorem in cognitive science (here: medical diagnosis);
- base rate neglect: experimental subjects ignore information about prior probability;
- a random variable picks out a subset of the sample space;
- a probability distribution returns a probability for each value of a random variable.
- a cumulative distribution sums all the values of a probability up to a threshold.

References

Casscells, W., A. Schoenberger, and T. Grayboys. 1978. Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine* 299(18):999–1001.

Medin, D. L. and S. M. Edelson. 1988. Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General* 117(1):68–85.