PROBABILITY THEORY

Probability

Probability Theory

We wish to attach probabilities to different kinds of events (or hypotheses or propositions):

- Event A: the next card is an Ace.
- Event R: it will rain tomorrow.
- Event C: the suspect left the crime stain.

Probabilities

Assign probabilities to events: Pr(A) or p_A or even p means "the probability that event A is true." All probabilities are conditional on some information I, so should write Pr(A|I) for "the probability that A is true given that I is known."

No matter how probabilities are defined, they need to follow some mathematical laws in order to lead to consistent theories.

First Law of Probability

 $0 \leq \Pr(A|I) \leq 1$

$\Pr(A|A,I) = 1$

If A is the event that a die shows an even face (2, 4, or 6), what is I? What is Pr(A|I)?

Second Law of Probability

If A, B are mutually exclusive given I

```
\Pr(A \text{ or } B|I) = \Pr(A|I) + \Pr(B|I)
```

```
so \Pr(\bar{A}|I) = 1 - \Pr(A|I)
```

 $(\overline{A} \text{ means not-}A).$

If A is the event that a die shows an even face, and B is the event that the die shows a 1, verify the Second Law.

Third Law of Probability

$\Pr(A \text{ and } B|I) = \Pr(A|B,I) \times \Pr(B|I)$

If A is event that die shows an even face, and B is the event that the die shows a 1, verify the Third Law.

Will generally omit the I from now on.

Independent Events

Events A and B are independent if knowledge of one does not affect probability of the other:

$$Pr(A|B) = Pr(A)$$
$$Pr(B|A) = Pr(B)$$

Therefore, for independent events

$$\Pr(A \text{ and } B) = \Pr(A) \Pr(B)$$

This may be written as

 $\Pr(AB) = \Pr(A)\Pr(B)$

Law of Total Probability

Because B and \overline{B} are mutually exclusive and exhaustive:

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

If A is the event that die shows a 3, B is the event that the die shows an even face, and \overline{B} the event that the die shows an odd face, verify the Law of Total Probability.

Odds

The odds O(A) of an event A are the probability of the event being true divided by the probability of the event not being true:

$$O(A) = \frac{\Pr(A)}{\Pr(\overline{A})}$$

This can be rearranged to give

$$\Pr(A) = \frac{O(A)}{1 + O(A)}$$

Odds of 10 to 1 are equivalent to a probability of 10/11.

Bayes' Theorem

The third law of probability can be used twice to reverse the order of conditioning:

$$\Pr(B|A) = \frac{\Pr(B \text{ and } A)}{\Pr(A)}$$
$$= \frac{\Pr(A|B)\Pr(B)}{\Pr(A)}$$

Odds Form of Bayes' Theorem

From the third law of probability

$$\Pr(B|A) = \Pr(A|B) \Pr(B) / \Pr(A)$$

$$\Pr(\bar{B}|A) = \Pr(A|\bar{B}) \Pr(\bar{B}) / \Pr(A)$$

Taking the ratio of these two equations:

$$\frac{\Pr(B|A)}{\Pr(\bar{B}|A)} = \frac{\Pr(A|B)}{\Pr(A|\bar{B})} \times \frac{\Pr(B)}{\Pr(\bar{B})}$$

Posterior odds = likelihood ratio \times prior odds.

Birthday Problem

Forensic scientists in Arizona looked at the 65,493 profiles in the Arizona database and reported that two profiles matched at 9 loci out of 13. They reported a "match probability" for those 9 loci of 1 in 754 million. Are the numbers 65,493 and 754 million inconsistent?

Troyer et al., 2001. Proc Promega 12th Int Symp Human Identification.

To begin to answer this question suppose that every possible profile has the same profile probability P and that there are N profiles in a database (or in a population). The probability of at least one pair of matching profiles in the database is one minus the probability of no matches.

Birthday Problem

Choose profile 1. The probability that profile 2 does not match profile 1 is (1-P). The probability that profile 3 does not match profiles 1 or 2 is (1-2P), etc. So, the probability P_M of at least one matching pair is

$$P_M = 1 - \{1(1-P)(1-2P) \cdots [1-(N-1)P]\}$$

$$\approx 1 - \prod_{i=0}^{N-1} e^{-iP} \approx 1 - e^{-N^2 P/2}$$

If P = 1/365 and N = 23, then $P_M = 0.51$. So, approximately, in a room of 23 people there is greater than a 50% probability that two people have the same birthday.

Birthday Problem

If P = 1/(754 million) and N = 65,493, then $P_M = 0.98$ so it is highly probable there would be a match. There are other issues, having to do with the four non-matching loci, and the possible presence of relatives in the database.

If $P = 10^{-16}$ and N = 300 million, then $P_M =$ is essentially 1. It is almost certain that two people in the US have the same rare DNA profile.

Statistics

- Probability: For a given model, what do we expect to see?
- Statistics: For some given data, what can we say about the model?
- Example: A marker has an allele A with frequency p_A .
 - Probability question: If $p_A = 0.5$, and if alleles are independent, what is the probability of AA?
 - Statistics question: If a sample of 100 individuals has 23 AA's, 48 Aa's and 29 aa's, what is an estimate of p_A ?