We hope that the reader has seen a little basic probability theory previously.

We will give a very quick review; some references for further reading appear

at the end of the chapter. A variable A represents an event (a subset of the

space of possible outcomes). Equivalently, we can represent the subset via a

RANDOM VARIABLE random variable, which is a function from outcomes to real numbers; the sub-

set is the domain over which the random variable A has a particular value.

Often we will not know with certainty whether an event is true in the world.

We can ask the probability of the event 0 ≤ P(A) ≤ 1. For two events A and

B, the joint event of both events occurring is described by the joint probabil-

ity P(A, B). The conditional probability P(A|B) expresses the probability of

event A given that event B occurred. The fundamental relationship between

CHAIN RULE joint and conditional probabilities is given by the chain rule:

(11.1) P(A, B) = P(A ∩ B) = P(A|B)P(B) = P(B|A)P(A)

Without making any assumptions, the probability of a joint event equals the

probability of one of the events multiplied by the probability of the other

event conditioned on knowing the first event happened.

Writing P(A) for the complement of an event, we similarly have:

(11.2) P(A, B) = P(B|A)P(A)

PARTITION RULE Probability theory also has a partition rule, which says that if an event B can

be divided into an exhaustive set of disjoint subcases, then the probability of

B is the sum of the probabilities of the subcases. A special case of this rule

gives that:

(11.3) P(B) = P(A, B) + P(A, B)

BAYES’ RULE From these we can derive Bayes’ Rule for inverting conditional probabili-

ties:

P(A|B) = P(B|A)P(A)

P(B)

=

"

P(B|A)

∑X∈{A,A}

P(B|X)P(X)

#

(11.4) P(A)

This equation can also be thought of as a way of updating probabilities. We

start off with an initial estimate of how likely the event A is when we do

PRIOR PROBABILITY not have any other information; this is the prior probability P(A). Bayes’ rule

POSTERIOR lets us derive a posterior probability P(A|B) after having seen the evidence B,

PROBABILITY

Online edition (c)

2009 Cambridge UP

11.2 The Probability Ranking Principle 221

based on the likelihood of B occurring in the two cases that A does or does not

hold.1

ODDS Finally, it is often useful to talk about the odds of an event, which provide

a kind of multiplier for how probabilities change:

Odds: O(A) = P(A)

P(A)

=

P(A)

1 − P(A)