PRACTICAL WORK 2.

21.11 Exercises

Exercise *2.1* [★]

This exercise indicates the kind of facility with set theory needed for this book, and summarizes a few useful results in probability theory. Use set theory and the axioms defining a probability function to show that:

a. $P(A\ u\ B) = P(A) + P(B) - P(A \cap B)$ [the addition rule]

b. $P(0) = 0$

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

d. $A \subseteq B \Rightarrow P(A) \leq P(B)$

e. $P(B - A) = P(B) - P(A \cap B)$

Exercise *2.2* [★]

Assume the following sample space:

(2.23) Ω = {is-noun, has-plural-s, is-adjective, is-verb]

and the function $f : 2^{\Omega} \rightarrow [0, 1]$ with the following values:

x	$f(x)$
{ is-noun }	0.45
{ has-plural-s }	0.2
{ is-adjective }	0.25
{ is-verb }	0.3

Can f be extended to all of 2^{Ω} such that it is a well-formed probability distribution? If not, how would you model these data probabilistically?

Exercise 2.3 [★]

Compute the probability of the event 'A period occurs after a three-letter word and this period indicates an abbreviation (not an end-of-sentence marker),' assuming the following probabilities.

(2.24) P(is-abbreviation] three-letter-word) = 0.8

(2.25) P(three-letter-word) = 0.0003

Exercise 2.4 [★]

Are X and Y as defined in the following table independently distributed?

x	0	0	1	1
Y	0	1	0	1
$p(X = x, Y = y)$	0.32	0.08	0.48	0.12

Exercise 2.5 [★]

In example 5, we worked out the expectation of the sum of two dice in terms of the expectation of rolling one die. Show that one gets the same result if one calculates the expectation for two dice directly.

Exercise 2.6 [★★]

Consider the set of grades you have received for courses taken in the last two years. Convert them to an appropriate numerical scale. What is the appropriate distribution for modeling them?

Exercise 2.7 [★★]

Find a linguistic phenomenon that the binomial distribution is a good model for. What is your best estimate for the parameter p?

Exercise 2.8 [★★]

For $i = 8$ and $j = 2$, confirm that the maximum of equation (2.15) is at 0.8, and that the maximum of equation (2.17) is 0.75. Suppose our prior belief had instead been captured by the equation:

$$P(\mu_m) = 30m^2(1-m)^2$$

What then would the MAP probability be after seeing a particular sequence of 8 heads and 2 tails? (Assume the theory μ_m and a prior belief that the coin is fair.)