Directly applying section 3.3.1,

$$\begin{aligned} P(B \mid A) &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{P(\{(4,4)\})}{P(\{(4,4)\}) + P(\{(2,6)\}) + P(\{(6,2)\}) + P(\{(3,5)\}) + P(\{(5,3)\})} \\ &= \frac{\frac{1}{6} \cdot \frac{1}{6}}{5 \cdot \frac{1}{6} \cdot \frac{1}{6}} \\ &= \frac{1}{5}. \end{aligned}$$

We don't count the event $\{(4,4)\}$ twice, but we do count both $\{(3,5)\}$ and $\{(5,3)\}$, since they are distinct events. We say "order matters" for these types of events.

3.4 Bayes' Theorem

T9

Given two events A and B, **Bayes' theorem** (aka Bayes' rule) states that

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

Sometimes this is written

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid \neg A)P(\neg A)} \tag{3.1}$$

$$= \frac{1}{1 + \dfrac{P(B \mid \neg A)}{P(B \mid A)} \cdot \dfrac{P(\neg A)}{P(A)}}. \tag{3.2}$$

This is a useful theorem for determining a test's effectiveness. If a test is performed to determine whether an event has occurred, we might as questions like "if the test indicates that the event has occurred, what is the probability it has actually occurred?" Bayes' theorem can help compute an answer.

3.4.1 Testing Outcomes

The test can be either positive or negative, meaning it can either indicate or not indicate that A has occurred. Furthermore, this result can be either *true* ☺ or *false* ☹.

There are four options, then. Consider an event A and an event that is a test result B indicating that event A has occurred. table 3.1 shows these four possible test outcomes. The event A occurring can lead to a true positive or a false negative, whereas $\neg A$ can lead to a true negative or a false positive.

Table 3.1: Test outcome B for event A.

		A		$\neg A$	
positive (B)	■	true	☺	false	☹
negative ($\neg B$)	■	false	☹	true	☺

Terminology is important, here.

- $P(\{\text{true positive}\}) = P(B \mid A)$, aka **sensitivity** or **detection rate**,
- $P(\{\text{true negative}\}) = P(\neg B \mid \neg A)$, aka **specificity**,
- $P(\{\text{false positive}\}) = P(B \mid \neg A)$,
- $P(\{\text{false negative}\}) = P(\neg B \mid A)$.

Clearly, the desirable result for any test is that it is *true*. However, no test is true 100 percent of the time. So sometimes it is desirable to err on the side of the false positive, as in the case of a medical diagnostic. Other times, it is more desirable to err on the side of a false negative, as in the case of testing for defects in manufactured balloons (when a false negative isn't a big deal).

3.4.2 Posterior Probabilities

Returning to Bayes' theorem, we can evaluate the **posterior probability** $P(A \mid B)$ of the event A having occurred given that the test B is positive, given information that includes the **prior probability** $P(A)$ of A. The form in equation (3.1) or equation (3.2) is typically useful because it uses commonly known test probabilities: of the true positive $P(B \mid A)$ and of the false positive $P(B \mid \neg A)$. We calculate $P(A \mid B)$ when we want to interpret test results.

Some interesting results can be found from this. For instance, if we let $P(B \mid A) = P(\neg B \mid \neg A)$ (sensitivity equal specificity) and realize that $P(B \mid \neg A) + P(\neg B \mid \neg A) = 1$ (when $\neg A$, either B or $\neg B$), we can derive the expression

$$P(B \mid \neg A) = 1 - P(B \mid A).$$

Using this and $P(\neg A) = 1 - P(A)$ in equation (3.2) gives (recall we've assumed sensitivity equals specificity!)

$$\begin{aligned} P(A \mid B) &= \frac{1}{1 + \dfrac{1 - P(B \mid A)}{P(B \mid A)} \cdot \dfrac{1 - P(A)}{P(A)}} \\ &= \frac{1}{1 + \left(\dfrac{1}{P(B \mid A)} - 1\right)\left(\dfrac{1}{P(A)} - 1\right)} \end{aligned}$$

This expression is plotted in figure 3.1. See that a positive result for a rare event (small $P(A)$) is hard to trust unless the sensitivity $P(B \mid A)$ *and* specificity $P(\neg B \mid \neg A)$ are very high, indeed!

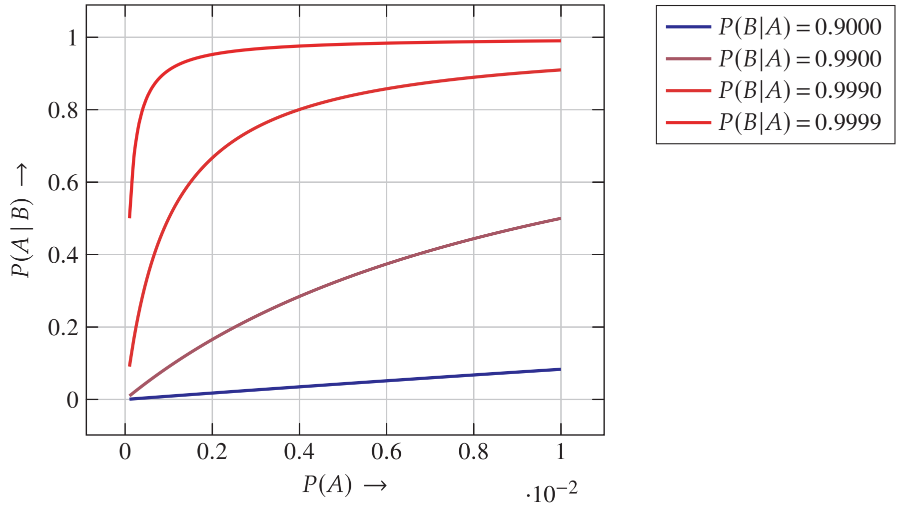

Figure 3.1. For different high-sensitivities, the probability that an event A occurred given that a test for it B is positive versus the probability that the event A occurs, under the assumption the specificity equals the sensitivity.

Example 3.4

Suppose 0.1 percent of springs manufactured at a given plant are defective. Suppose you need to design a test that, when it indicates a deffective part, the part is actually defective 99 percent of the time. What sensitivity should your test have assuming it can be made equal to its specificity?

We proceed in Python.

```
from sympy import * # for symbolics
import numpy as np # for numerics
import matplotlib.pyplot as plt # for plots
```

Define symbolic variables.

```
var('p_A,p_nA,p_B,p_nB,p_B_A,p_B_nA,p_A_B',real=True)
```

```
(p_A, p_nA, p_B, p_nB, p_B_A, p_B_nA, p_A_B)
```

Beginning with Bayes' theorem and assuming the sensitivity and specificity are equal by section 3.4.2, we can derive the following expression for the posterior probability $P(A \mid B)$.

```
p_A_B_e1 = Eq(p_A_B,p_B_A*p_A/p_B).subs(
  {
    p_B: p_B_A*p_A+p_B_nA*p_nA, # conditional prob
    p_B_nA: 1-p_B_A, # Eq (3.5)
    p_nA: 1-p_A
  }
)
print(p_A_B_e1)
```

$$p_{AB} = \frac{p_A p_{BA}}{p_A p_{BA} + (1 - p_A)(1 - p_{BA})}$$

Solve this for $P(B \mid A)$, the quantity we seek.

```
p_B_A_sol = solve(p_A_B_e1,p_B_A,dict=True)
p_B_A_eq1 = Eq(p_B_A,p_B_A_sol[0][p_B_A])
print(p_B_A_eq1)
```

$$p_{BA} = \frac{p_{AB}(1 - p_A)}{-2p_A p_{AB} + p_A + p_{AB}}$$

Now let's substitute the given probabilities.

```
p_B_A_spec = p_B_A_eq1.subs(
  {
    p_A: 0.001,
    p_A_B: 0.99,
  }
)
print(p_B_A_spec)
```

$$p_{BA} = 0.999989888981011$$

That's a tall order!

3.5 Random Variables

13

Probabilities are useful even when they do not deal strictly with events. It often occurs that we measure something that has randomness associated with it. We use random variables to represent these measurements.

A **random variable** $X : \Omega \rightarrow \mathbb{R}$ is a function that maps an outcome ω from the sample space Ω to a real number $x \in \mathbb{R}$, as shown in figure 3.2. A random variable will be denoted with a capital letter (e.g. X and K) and a specific value that it maps to (the value) will be denoted with a lowercase letter (e.g. x and k).

A **discrete random variable** K is one that takes on discrete values. A **continuous random variable** X is one that takes on continuous values.

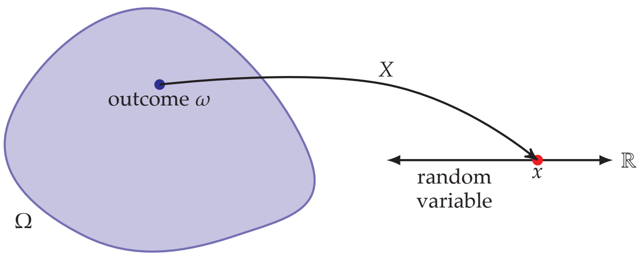

Figure 3.2. A random variable X maps an outcome $\omega \in \Omega$ to an $x \in \mathbb{R}$.

Example 3.5

Roll two unbiased dice. Let K be a random variable representing the sum of the two. Let $P(k)$ be the probability of the result $k \in K$. Plot and interpret $P(k)$.

Figure 3.3 shows the probability of each sum occurring.

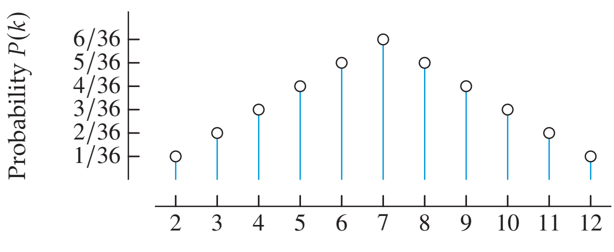

Figure 3.3. PMF for the summ of two dice rolled.