2.1.4 Medical screening revisited

Let us now return to our cancer screening example and apply the sum and product rules of probability to answer our two questions. For clarity, when working through this example, we will once again be explicit about distinguishing between the random variables and their instantiations. We will denote the presence or absence of cancer by the variable C, which can take two values: $C = 0$ corresponds to 'no cancer' and $C = 1$ corresponds to 'cancer'. We have assumed that one person in a hundred in the population has cancer, and so we have

$$
\begin{align}
p(C = 1) &= 1/100 \tag{2.12}\\
p(C = 0) &= 99/100, \tag{2.13}
\end{align}
$$

respectively. Note that these satisfy $p(C = 0) + p(C = 1) = 1$.

Now let us introduce a second random variable T representing the outcome of a screening test, where $T = 1$ denotes a positive result, indicative of cancer, and $T = 0$ a negative result, indicative of the absence of cancer. As illustrated in Figure 2.3, we know that for those who have cancer the probability of a positive test result is 90%, while for those who do not have cancer the probability of a positive test result is 3%. We can therefore write out all four conditional probabilities:

$$
\begin{align}
p(T = 1|C = 1) &= 90/100 \tag{2.14}\\
p(T = 0|C = 1) &= 10/100 \tag{2.15}\\
p(T = 1|C = 0) &= 3/100 \tag{2.16}\\
p(T = 0|C = 0) &= 97/100. \tag{2.17}
\end{align}
$$

Again, note that these probabilities are normalized so that

$$
p(T = 1|C = 1) + p(T = 0|C = 1) = 1 \tag{2.18}
$$

and similarly

$$
p(T = 1|C = 0) + p(T = 0|C = 0) = 1. \tag{2.19}
$$

We can now use the sum and product rules of probability to answer our first question and evaluate the overall probability that someone who is tested at random will have a positive test result:

$$
\begin{align}
p(T = 1) &= p(T = 1|C = 0)p(C = 0) + p(T = 1|C = 1)p(C = 1)\\
&= \frac{3}{100} \times \frac{99}{100} + \frac{90}{100} \times \frac{1}{100} = \frac{387}{10,000} = 0.0387. \tag{2.20}
\end{align}
$$

We see that if a person is tested at random there is a roughly 4% chance that the test will be positive even though there is a 1% chance that they actually have cancer. From this it follows, using the sum rule, that $p(T = 0) = 1 - 387/10,000 = 9613/10,000 = 0.9613$ and, hence, there is a roughly 96% chance that the do not have cancer.

Now consider our second question, which is the one that is of particular interest to a person being screened: if a test is positive, what is the probability that the person