2.2.2.1 Probability of a union of two events

Given two events, A and B, we define the probability of A or B as follows:

$$p(A \vee B) = p(A) + p(B) - p(A \wedge B)$$
(2.1)

$$= p(A) + p(B)$$
 if A and B are mutually exclusive (2.2)

2.2.2.2 Joint probabilities

We define the probability of the joint event A and B as follows:

$$p(A,B) = p(A \land B) = p(A|B)p(B) \tag{2.3}$$

This is sometimes called the **product rule**. Given a **joint distribution** on two events p(A, B), we define the **marginal distribution** as follows:

$$p(A) = \sum_{b} p(A,B) = \sum_{b} p(A|B=b)p(B=b)$$
(2.4)

where we are summing over all possible states of B. We can define p(B) similarly. This is sometimes called the **sum rule** or the **rule of total probability**.

The product rule can be applied multiple times to yield the **chain rule** of probability:

$$p(X_{1:D}) = p(X_1)p(X_2|X_1)p(X_3|X_2, X_1)p(X_4|X_1, X_2, X_3) \dots p(X_D|X_{1:D-1})$$
(2.5)

where we introduce the Matlab-like notation 1:D to denote the set $\{1,2,\ldots,D\}$.

2.2.2.3 Conditional probability

We define the **conditional probability** of event A, given that event B is true, as follows:

$$p(A|B) = \frac{p(A,B)}{p(B)} \text{ if } p(B) > 0$$
 (2.6)

2.2.3 Bayes rule

Combining the definition of conditional probability with the product and sum rules yields **Bayes** rule, also called **Bayes Theorem**²:

$$p(X = x|Y = y) = \frac{p(X = x, Y = y)}{p(Y = y)} = \frac{p(X = x)p(Y = y|X = x)}{\sum_{x'} p(X = x')p(Y = y|X = x')}$$
(2.7)

2.2.3.1 Example: medical diagnosis

As an example of how to use this rule, consider the following medical diagonsis problem. Suppose you are a woman in your 40s, and you decide to have a medical test for breast cancer called a **mammogram**. If the test is positive, what is the probability you have cancer? That obviously depends on how reliable the test is. Suppose you are told the test has a **sensitivity**

^{2.} Thomas Bayes (1702-1761) was an English mathematician and Presbyterian minister.