

APPENDIX J

THE BASE RATE FALLACY

William Stallings

J.1	CONDITIONAL PROBABILITY AND INDEPENDENCE.....	2
J.2	BAYES' THEOREM.....	3
J.3	THE BASE RATE FALLACY DEMONSTRATED	5
J.4	REFERENCES.....	6

Copyright 2014
Supplement to
Computer Security, Third Edition
Pearson 2014
<http://williamstallings.com/ComputerSecurity>

We begin with a review of important results from probability theory, then demonstrate the base-rate fallacy.

J.1 CONDITIONAL PROBABILITY AND INDEPENDENCE

We often want to know a probability that is conditional on some event. The effect of the condition is to remove some of the outcomes from the sample space. For example, what is the probability of getting a sum of 8 on the roll of two dice if we know that the face of at least one die is an even number? We can reason as follows. Because one die is even and the sum is even, the second die must show an even number. Thus, there are three equally likely successful outcomes: (2, 6), (4, 4) and (6, 2), out of a total set of possibilities of $[36 - (\text{number of events with both faces odd})] = 36 - 3 \times 3 = 27$. The resulting probability is $3/27 = 1/9$.

Formally, the **conditional probability** of an event A assuming the event B has occurred, denoted by $\Pr[A|B]$, is defined as the ratio

$$\Pr[A|B] = \frac{\Pr[AB]}{\Pr[B]}$$

where we assume $\Pr[B]$ is not zero.

In our example, $A = \{\text{sum of 8}\}$ and $B = \{\text{at least one die even}\}$. The quantity $\Pr[AB]$ encompasses all of those outcomes in which the sum is 8 and at least one die is even. As we have seen, there are three such outcomes. Thus, $\Pr[AB] = 3/36 = 1/12$. A moment's thought should convince you that $\Pr[B] = 3/4$. We can now calculate

$$\Pr[A|B] = \frac{1/12}{3/4} = \frac{1}{9}$$

This agrees with our previous reasoning.

Two events A and B are called **independent** if $\Pr[AB] = \Pr[A]\Pr[B]$. It can easily be seen that if A and B are independent, $\Pr[A|B] = \Pr[A]$ and $\Pr[B|A] = \Pr[B]$.

J.2 BAYES' THEOREM

One of the most important results from probability theory is known as Bayes' theorem. First we need to state the total probability formula. Given a set of mutually exclusive events E_1, E_2, \dots, E_n , such that the union of these events covers all possible outcomes, and given an arbitrary event A , then it can be shown that

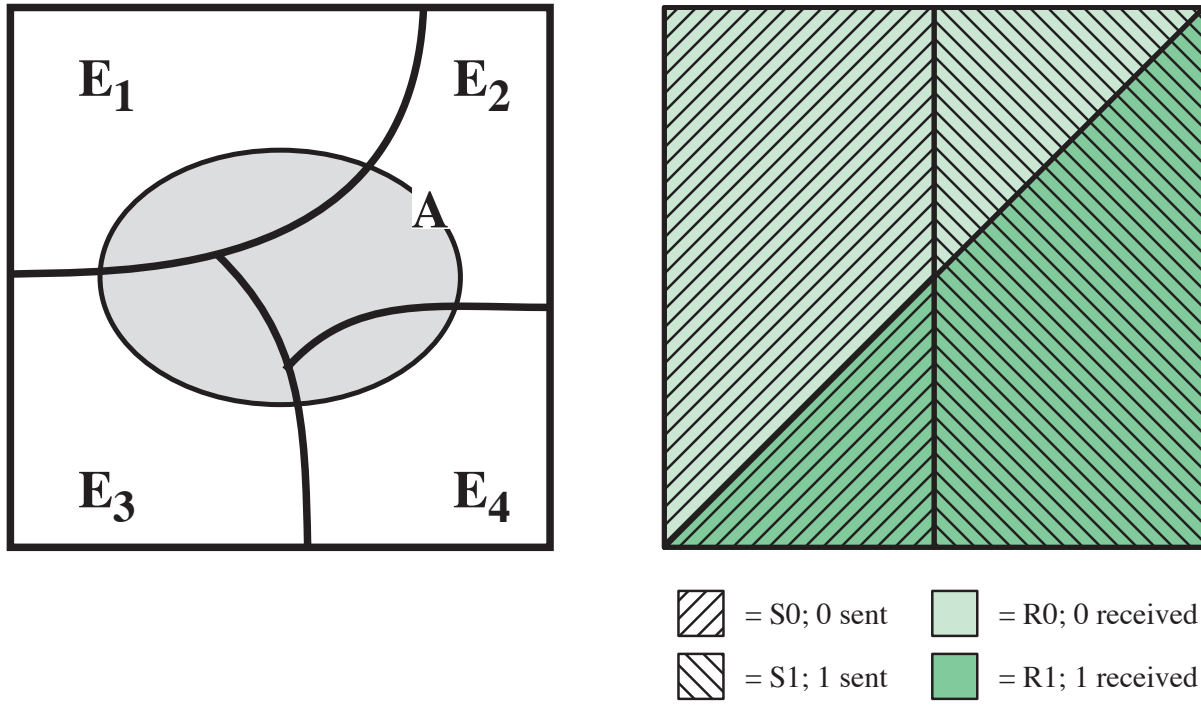
$$\Pr[A] = \sum_{i=1}^n \Pr[A|E_i]\Pr[E_i] \quad (\text{J.1})$$

Bayes' theorem may be stated as follows:

$$\Pr[E_i|A] = \frac{\Pr[A|E_i]\Pr[E_i]}{\Pr[A]} = \frac{\Pr[A|E_i]\Pr[E_i]}{\sum_{j=1}^n \Pr[A|E_j]\Pr[E_j]} \quad (\text{J.2})$$

Figure J.1a illustrates the concepts of total probability and Bayes' theorem.

Bayes' theorem is used to calculate "posterior odds," that is, the probability that something really is the case, given evidence in favor of it. For example, suppose we are transmitting a sequence of zeroes and ones over a noisy transmission line. Let S_0 and S_1 be the events a zero is sent at a given time and a one is sent, respectively, and R_0 and R_1 be the events that a zero is received and a one is received. Suppose we know the



(a) Diagram to illustrate concepts

(b) Example

Figure J.1 Illustration of Total Probability and Bayes' Theorem

probabilities of the source, namely $\Pr[S1] = p$ and $\Pr[S0] = 1 - p$. Now the line is observed to determine how frequently an error occurs when a one is sent and when a zero is sent, and the following probabilities are calculated: $\Pr[R0|S1] = p_a$ and $\Pr[R1|S0] = p_b$. If a zero is received, we can then calculate the conditional probability of an error, namely the conditional probability that a one was sent given that a zero was received, using Bayes' theorem:

$$\Pr[S1|R0] = \frac{\Pr[R0|S1]\Pr[S1]}{\Pr[R0|S1]\Pr[S1] + \Pr[R0|S0]\Pr[S0]} = \frac{p_a p}{p_a p + (1 - p_b)(1 - p)}$$

Figure J.1b illustrates the preceding equation. In the figure, the sample space is represented by a unit square. Half of the square corresponds to S_0

and half to S1, so $\Pr[S0] = \Pr[S1] = 0.5$. Similarly, half of the square corresponds to R0 and half to R1, so $\Pr[R0] = \Pr[R1] = 0.5$. Within the area representing S0, one quarter of that area corresponds to R1, so $\Pr[R1/S0] = 0.25$. Other conditional probabilities are similarly evident.

J.3 THE BASE RATE FALLACY DEMONSTRATED

Consider the following situation. A patient has a test for some disease that comes back positive (indicating he has the disease). You are told that

- The accuracy of the test is 87% (i.e., if a patient has the disease, 87% of the time, the test yields the correct result, and if the patient does not have the disease, 87% of the time, the test yields the correct result).
- The incidence of the disease in the population is 1%.

Given that the test is positive, how probable is it that the patient does not have the disease? That is, what is the probability that this is a false alarm? We need Bayes' theorem to get the correct answer:

$$\begin{aligned} \Pr[\text{well/positive}] &= \frac{\Pr[\text{positive/well}]\Pr[\text{well}]}{\Pr[\text{positive/disease}]\Pr[\text{disease}] + \Pr[\text{positive/well}]\Pr[\text{well}]} \\ &= \frac{(0.13)(0.99)}{(0.87)(0.01) + (0.13)(0.99)} = 0.937 \end{aligned}$$

Thus, in the vast majority of cases, when a disease condition is detected, it is a false alarm.

This problem, used in a study [PIAT91, PIAT94], was presented to a number of people. Most subjects gave the answer 13%. The vast majority, including many physicians, gave a number below 50%. Many physicians who

guessed wrong lamented, "If you are right, there is no point in making clinical tests!" The reason most people get it wrong is that they do not take into account the basic rate of incidence (the base rate) when intuitively solving the problem. This error is known as the *base-rate fallacy* [BARH80].

How could this problem be fixed? Suppose we could drive both of the correct result rates to 99.9%. That is, suppose we have $\text{Pr}[\text{positive}/\text{disease}] = 0.999$ and $\text{Pr}[\text{negative}/\text{well}] = 0.999$. Plugging these numbers into the Equation (J.2), we get $\text{Pr}[\text{well}/\text{positive}] = 0.09$. Thus, if we can accurately detect disease and accurately detect lack of disease at a level of 99.9%, then the rate of false alarms will be 9%. This is much better, but still not ideal. Moreover, again assume 99.9% accuracy, but now suppose that the incidence of the disease in the population is only $1/10000 = 0.0001$. We then end up with a rate of false alarms of 91%. In actual situations, [AXEL00] found that the probabilities associated with IDSs were such that the false alarm rate was unsatisfactory.

J.4 REFERENCES

AXEL00 Axelsson, S. "The Base-Rate Fallacy and the Difficulty of Intrusion Detection." *ACM Transactions and Information and System Security*, August 2000.

BARH80 Bar-Hillel, M. "The Base-Rate Fallacy in Probability Judgements." *Acta Psychologica*, May 1980.

PIAT91 Piattelli-Palmarini, M. "Probability: Neither Rational nor Capricious." *Bostonia*, March 1991.

PIAT94 Piattelli-Palmarini, M. *Inevitable Illusions: How Mistakes of Reason Rule Our Minds*. New York: Wiley, 1994.