Appendix G

Conditional Probability and Likelihood

A random variable is defined by the result of a rule (e.g., a function) that associates a real number with each outcome in a sample space, S. Consider two discrete random variables, A and B, that are associated with events in the sample space. The probabilities of A and B, denoted by $P(A)$ and $P(B)$, respectively, take values between zero and one and, by definition, $P(S) = 1$. Conditional probability is denoted by $P(A \mid B)$ and read as "the probability of A given B." Given that B has occurred, the sample space shrinks from S to B.

We expect that $P(A \mid B)$ is proportional to $P(A \cap B)$, which is the intersection probability that both A and B occur. Since $P(B \mid B)$ must be unity and $P(B \cap B) = P(B)$, the constant of proportionality must be $1/P(B)$:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \tag{G.1}$$

Similarly,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} \tag{G.2}$$

Using (G.2) to replace $P(A \cap B)$ in (G.1) gives **Bayes's Theorem**,

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)} \tag{G.3}$$

Likelihood

With conditional probability, we are given specific information about an event that affects the probability of another event. For example, knowing that $B = b$ allows us to "update" the probability of A by $P(A \mid B = b)$ through (G.3). Conversely, if we view the conditional probability as a function of the *second* argument, B, we get a likelihood function

$$L(b \mid A) = \alpha P(A \mid B = b) \tag{G.4}$$

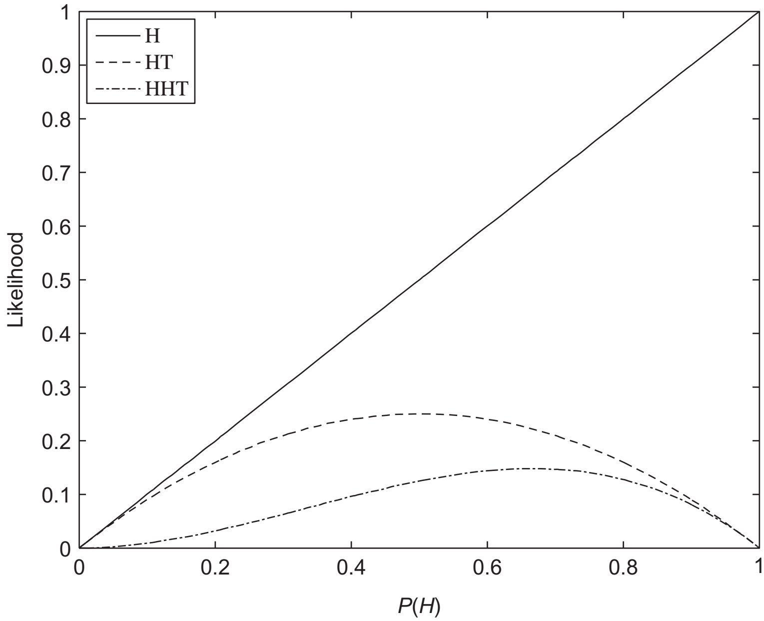

FIGURE G.1 Likelihood as a function of the probability of "H" on a two-sided coin given observations of heads "H" and tails "T."

where α is a constant parameter. For example, consider a sequence of observations of a two-sided coin with outcomes "H" and "T." If we flip the coin once and observe "H," we can ask, "What is the likelihood that the true probability $P(H)$ is 0.5?" Figure G.1 shows that $L = 0.5$, and in fact the most likely value is $P(H) = 1$; note that $L(0) = 0$ since "H" has been observed and therefore must have a nonzero probability. Suppose flipping the coin again also gives "H." Maximum likelihood is still located at $P(H) = 1$, and the likelihood of smaller $P(H)$ is diminished. Remember that we can turn this around and ask, "If we *know* that $P(H) = 0.5$ (i.e., the coin is fair), what is the probability of observing 'HH'?" This is conditional probability, which in this case is 0.25 (the same as $L(0.5 \mid HH)$. If, on the other hand, one of the observations is also "T," then both $L(0)$ and $L(1)$ must be zero (Figure G.1). Note that, unlike conditional probability likelihood is not a probability since it may take values larger than unity and it does not integrate to one.