Entropy As a Measure of Randomness: A Fundamental Concept in the Mathematical Sciences

by Professor Richard S. Ellis

At the Fifteenth Annual Mathematics Competition Awards Ceremony, held on April 27, 2000, I spoke on entropy as a measure of randomness. Because of the importance and beauty of the ideas in the talk, I am happy to be able to share them with a wider audience.

The many faces of entropy form a cluster of fundamental concepts in numerous areas of the mathematical sciences, including probability, statistics, information theory, and statistical mechanics. In this article I will describe how entropy as a measure of randomness can be used to solve the following gambling problem. If one suspects that a die is loaded, but is given only certain minimal information—namely, the average value of a large number of tosses of the die—then how can one systematically determine the probabilities of the six faces?

In answering this question, we will see how an ambiguous, ill-posed problem is given a precisely formulation to which the tools of mathematical analysis can be applied. The entropy concept used in this context can also be applied to solve problems in completely different areas; e.g., in statistical mechanics, which analyzes macroscopic physical systems by applying statistical principles to their microscopic constituents. This discovery of commonalities between completely different areas illustrates the power of mathematical reasoning to isolate and analyze unifying concepts that underlie apparently unrelated phenomena.

Before stating the gambling problem, we first consider a fair die, for which the probability of each of the six faces is $1/6$. If one tosses a fair die a large number of times, one expects the average value of the tosses to be close to, but not necessarily equal to, the theoretical mean, which equals $3.5 = \frac{1}{6}(1+2+3+4+5+6) = \frac{1}{6}\sum_{i=1}^{6} i.$ We write $\bar{y}$ for the average value of the tosses. It follows that if $\bar{y}$ differs markedly from 3.5, then we have reason to suspect that the die is loaded. A basic question, which this article does not address, is how many times should the die be tossed so that the conclusion $\bar{y} \neq 3.5$ can be trusted as a valid criterion for a loaded die.

The gambling problem can now be stated. Given $\bar{y} \neq 3.5$, we seek to determine the probabilities of each of the six faces. For $i = 1,\dots,6,$ let p_i be the probability of the i'th face. Then the p_i satisfy the following three constraints:

$$p_i \geq 0, \;\; \sum_{i=1}^{6} p_i = 1, \text{ and } \sum_{i=1}^{6} i \cdot p_i = \bar{y}.$$

The last equality expresses the consistency of the p_i with the information that the average value of the tosses is the given number $\bar{y}$. Unfortunately, for any value of $\bar{y}$ satisfying $0 < \bar{y} < 6,$ there are infinitely many choices of p_i that satisfy these three constraints. In this formulation, the gambling problem is said to be ill-posed.

In order to give the problem a precise formulation, we consider the entropy of the probabilities $p = (p_1,\dots, p_6)$, following Ludwig Boltzmann, who introduced this concept in statistical mechanics. The entropy of p is defined by $h(p) = -\sum_{i=1}^{6} p_i \log p_i$, where logarithms are calculated to the base e. This function has a number of easily verified properties. The first property is that $h(p) \geq 0$. The second property is that $h(p)$ equals its minimum value of 0 if and only if one of the p_i equals 1 and the other p_i equal 0; in some sense, this choice of probabilities represents the "least random" $p = (p_1, \ldots, p_6)$. An example is $p_6 = 1$ and $p_1 = \cdots = p_5 = 0$, which corresponds to the nonrandom situation of obtaining a 6 every time the die is tossed. The third property is that $h(p)$ equals its maximum value $(\log 6)$ if and only if all the $p_i = 1/6$; this choice of probabilities corresponds to a fair die and in some sense represents the "most random" choice of p.

The second and third properties of $h(p)$ justify, in a weak sense, calling $h(p)$ a measure of the randomness in $p = (p_1, \ldots, p_6)$. Indeed, according to the second property, $h(p)$ equals its minimum value of 0 for the least random choices $p = (1,0,0,0,0,0), (0,1,0,0,0,0), \ldots, (0,0,0,0,0,1)$, while according to the second property, $h(p)$ equals its maximum value of $\log 6$ for the most random choice (1/6, 1/6, 1/6, 1/6, 1/6, 1/6). As I mention at the end of this article, a much more convincing justification can be given for designating $h(p)$ a measure of the randomness in p.

Our original, ill-posed problem—namely, given $\bar{y} \neq 3.5$, determine the probabilities $p = (p_1, \ldots, p_6)$—can now be given a precise formulation by adding one crucial ingredient; namely, the entropy $h(p) = -\sum_{i=1}^{6} p_i \log p_i$. Using $h(p)$ as a measure of the randomness of $p = (p_1, \ldots, p_6)$, we seek to determine the most random choice of p that is consistent with the information $\bar{y} \neq 3.5$. This is easily translated into mathematical terms as the following constrained maximization problem: maximize $h(p)$, where p satisfies the three constraints $p_i \geq 0, \sum_{i=1}^{6} p_i = 1$, and $\sum_{i=1}^{6} i \cdot p_i = \bar{y}$.

Although there are infinitely many p that satisfy the three constraints, one can prove that this constrained maximization problem has a unique solution, which we write as $\hat{p}$. Although this unique solution cannot be written down explicitly, it can be computed numerically and exhibited graphically. One can also prove that in the class of all p that satisfy the three constraints, $\hat{p}$ becomes, in a certain well-defined sense, the optimal solution as the number of tosses grows larger and larger.

At the Mathematics Competition Awards Ceremony, I completed my presentation by justifying the designation of $h(p)$ as a measure of the randomness in p. Since this cannot be done convincingly without introducing considerably more notation, the details will be omitted here. It is important to point out, however, that the justification rests on the following rather profound observation: the entropy $h(p)$ is a bridge between the highly simplified description of the dice tossing game in terms of the single number $\bar{y}$ and the extremely complicated description of the game in terms of sequences of individual tosses. This facet of entropy—a bridge between two levels of description, one highly simplified and the other extremely complicated—underlies numerous applications of the concept in the mathematical sciences.