The Normal Distribution ... or Another Distribution

Adam Brandenburger*
Shellwyn Weston†

Version 12/29/15

1 Introduction

The normal — also known as Gaussian — distribution is surely the probability distribution most studied in introductory courses in probability and statistics.

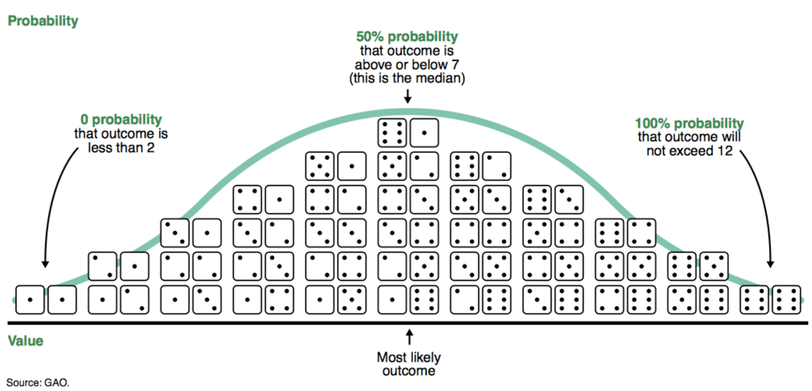

Figure 1

One reason so much attention is given to the normal distribution is that the Central Limit Theorem tells us that if a quantity is made up of the sum or average of a number of independently varying quantities, where the means and variances of these latter quantities are sufficiently 'well-behaved,' then the first quantity will be approximately normally distributed. To state a simple version of the **Central Limit Theorem** more formally, let $X_1, X_2, \ldots$ be a sequence of random variables which are independently and identically distributed, with common mean μ and common variance σ^2. Then, if we form the statistic:

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma},$$

*Stern School of Business, Tandon School of Engineering, NYU Shanghai, New York University, New York, NY 10012, U.S.A., adam.brandenburger@stern.nyu.edu, adambrandenburger.com.
†Institute for the Interdisciplinary Study of Decision Making, New York University, New York, NY 10012, U.S.A., sweston@stern.nyu.edu.

the distribution of this statistic will approach that of the normal distribution with mean 0 and variance 1, as n tends to ∞.[1] Equivalently, we can say that, for large n, the quantity $\sum_{i=1}^{n} X_i$ is approximately equal to $n\mu$ plus an error term which is normally distributed with mean 0 and variance $n\sigma^2$. Figure 1 depicts the probability distribution of the total score obtained by tossing a fair die twice (independently). The shape of the graph is a little like the **Bell curve** exhibited by the normal distribution. See Figure 2.[2] The Central Limit Theorem tells us that the shape of the graph will become closer and closer to the Bell curve as the number of tosses increases.

Phenomena which are often assumed (or found) to follow a normal distribution include the heights of members of a biological population and the observational errors in a physical experiment. A bonus is that the mathematical form of the normal distribution often makes for tractable analysis and closed-form solutions to calculations.

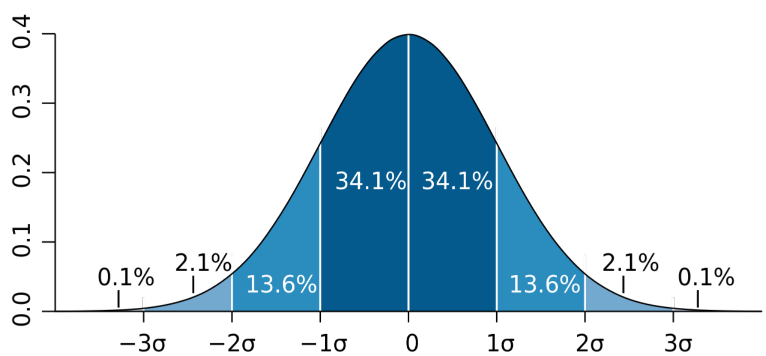

Figure 2

There may be a neural basis for being inclined to employ a normal distribution. In certain environments, animals (including humans) learn via what is known as **reinforcement learning** — more precisely, via what is called temporal difference reinforcement learning.[3] In this mode, an animal adjusts its behavior based on the difference between its expected reward and the reward it actually receives. A well-behaved measure of this difference is variance. This leads to the normal distribution because it is, in a precise sense, the unique probability distribution defined solely by mean and variance.[4] Pursuing this line of argument, one can even hypothesize that our evolutionary default is to perceive uncertainty in terms of the normal distribution (or similarly shaped distributions).[5]

A full-fledged **expected utility model of learning** also depends on distributional assumptions. Indeed, what is the optimal decision at any moment may differ significantly, depending on these assumptions, especially with respect to **rare** (low probability-high consequence) events. If a decision maker employs a normal distribution, either explicitly or by default, then s/he is making some strong assumptions. Refer again to Figure 2. Observe that a good number of outcomes are close to the average — approximately 68 percent of outcomes lie within one standard deviation of the mean. In addition, almost all outcomes lie within four standard deviations of the mean — only 1 in 15,787 observations will be a four-sigma event. The odds of a ten-sigma event are approximately 130,000,000,000,000,000,000,000 to 1. A decision maker who uses a normal distribution is assuming that rare events are not just rare, so to speak, but exceedingly unlikely.

2 How Normal is Normal?

Many real-world contexts of great importance to decision makers appear not to exhibit normally-distributed uncertainty. Here are some examples: sales of branded products, such as books, pharmaceuticals, and software; volumes of financial transactions, web hits, phone calls, and internet traffic; sizes of cities and networks; prices of ventures, financial assets, and markets; magnitudes of environmental or technological disasters such as hurricanes, earthquakes, oil spills, and nuclear plant failures.[6]

An unwarranted assumption of the normal distribution can lead to highly biased judgments of the consequences of an action or strategy. For example, if single-day financial-market drops of 10 percent were normally distributed, they would be expected to occur once every 500 years.[7] In actuality, they occur two orders of magnitude more frequently, about once every five years.[8]

3 Heavy-Tailed Distributions

The quantities in the example of the previous section have been found to be well- or approximately modeled by using, not the normal distribution, but various members of the family of heavy-tailed distributions. There is some disagreement as to what definition of heavy-tailedness should be. Here is an authoritative choice.[9] Say that a random variable X has sub-exponential tails there are numbers $C, c, \alpha > 0$ such that

$$\Pr\left(|X| \geq \lambda\right) \leq C e^{-c\lambda^{\alpha}}$$

for all $\lambda > 0$. (The notation $|X|$ means the absolute value of X.) Roughly speaking, this definition says that the probability mass in the tails of the probability distribution decreases rapidly (faster than exponentially) as the distance from the center of the distribution increases. A **heavy-tailed distribution** is one where the tails are not sub-exponential; in other words, the tails are heavy.

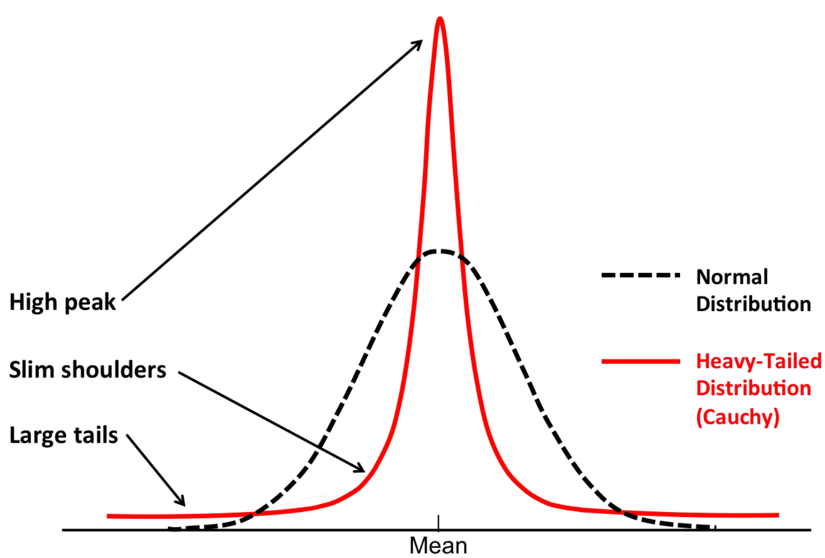

Figure 3

4 How Useful is Experience?

Figure 3 compares the normal distribution with the Cauchy distribution, an important symmetric heavy-tailed distribution.[10] We can see some very significant qualitative differences between the two curves. The greater mass in the tails of a heavy-tailed distribution must come from somewhere. (Recall from basic probability theory that the total area under each of the curves must be one.) Indeed, this mass has to come from the shoulders of the distribution, and we can see this effect in the **slim shoulders** and **high peak** exhibited by the heavy-tailed distribution relative to the normal distribution.

These characteristic features of heavy-tailed distributions make such contexts very treacherous. Sampling from these distributions will, with high probability, yield outcomes which are tightly clustered around the center of the distribution. If this is how we gain experience — from observing outcomes of the distribution for some while — then our experience may be highly misleading. We will be led to underestimate the frequency of low-probability events. Even worse, if these low-probability events are also of high consequence, we cannot afford to learn the hard way. Decision makers who do not develop the capacity to identify heavy-tailed contexts will undervalue or dismiss potential blockbuster opportunities, sell assets too cheaply, and fail to insure against catastrophic loss.

Einstein said of Nature's workings: "God is subtle but he is not malicious." Just perhaps, Einstein was wrong when it comes to heavy-tailed distributions!

Notes

[1]By "approach" is meant "converge in distribution to." This is the Lindeberg-Lévy version of the Central Limit Theorem. There are various stronger versions of the theorem; see any standard textbook on probability theory for details.

[2]Source: https://commons.wikimedia.org/wiki/File:Standard_deviation_diagram.svg.

[3]See Schultz, W., P. Dayan, and P.R. Montague, "A Neural Substrate of Prediction and Reward," *Science*, 275, 1593-1599, for evidence on the neural basis for this type of learning; and Sutton, R, and A. Barto, *Reinforcement Learning: An Introduction*, MIT Press, 1998, for a general treatment of reinforcement learning.

[4]The precise statement is that the normal distribution can be characterized as the entropy-maximizing probability distribution subject to the constraints of a given mean and a given variance. This is a standard result in information theory.

[5]Weston, S., "Envisioning the Improbable: Judgment and Strategy in Heavy-Tailed Contexts," *Academy of Management Proceedings*, 2014, 12849.

[6]See Andriani, P., and B. McKelvey, "Power Law Phenomena in Organizations: Redirecting Organization Studies," Academy of Management Meeting, Honolulu, Hawaii, August 2005; Newman, M., "Power Laws, Pareto Distributions and Zipf's Law," *Contemporary Physics* 46, 2005, 323-351; and Sornette, D., *Critical Phenomena in Natural Sciences*, Springer, 2001.

[7]Buchanan, M., "Power Laws and the New Science of Complexity Management," *Strategy and Business*, 34, Spring 2004, http://www.strategy-business.com/article/04107?pg=all.

[8]Mandelbrot, B., and R. Hudson, *The Mis(Behavior) of Markets*, Basic Books, 2004.

[9]Tao, T., "A Review of Probability Theory," at https://terrytao.wordpress.com/2010/01/01/254a-notes-0-a-review-of-probability-theory/.

[10]See Weston, op.cit..