# Half of a Coin: Negative Probabilities

**Gábor J. Székely**
Bowling Green State University, Bowling Green, OH
Hungarian Academy of Sciences, Budapest, Hungary

**Abstract:** Half coins are strange objects with infinitely many sides. They are numbered with $0, 1, 2, \ldots$ and the positive even numbers are taken with negative probabilities. Two half coins make a complete coin in the sense that if we flip two half coins then the sum of the outcomes is 0 or 1 with 1/2 probability as if we simply flipped a fair coin. In this paper we clarify the meaning and interpretation of negative probabilities and illustrate their importance in finance.

## 1 Kolmogorov's Bible

Around the time when Kolmogorov published his most influential book in 1933 (A. N. Kolmogorov 1933), the Nobel Laureate physicist, E. Wigner in 1932 published the following claim (joint with L. Szilárd, E. Wigner 1932): in quantum theory the joint density function $P(x, p)$ of the location and the momentum of a particle cannot be nonnegative everywhere: it is always real but its integral over the whole space is 0. Another Nobel Laureate physicist, P. Dirac (1942), also emphasized the necessity of negative probabilities. We can continue this list with R. P. Feynman (1987), M. S. Bartlett (1945), etc. The views of many outstanding physicists have not really touched the heart of most mathematicians. For them the notion of probability is codified in Kolmogorov's 'Bible' therefore probabilities are real numbers in the interval [0,1], nothing else makes sense 'by definition'. Most experts claim that even if we could define negative probabilities in a consistent way, nobody needs them. Negative probabilities simply have no applications.

In this paper we shall see that from a mathematical point of view negative probabilities are the same type of natural extensions of classical probabilities as the negative numbers are natural extensions of the nonnegative ones. So, in order to use negative probabilities we do not need to leave the 'Paradise' of Kolmogorov's theory, all we need is to use it in a more flexible way. Concerning the applicability of negative probabilities we discuss an example in finance. This example shows the close connection between negative probabilities and debits.

## 2 What is the Half of a Coin?

Let us start our journey in the world of negative probabilities with a fair coin. It has two sides: head and tail, we can also denote them by 0 and 1. Fair coins are random variables that take 0 and 1 with the same 1/2 probability. The most convenient analytic tool for studying the addition of coins or any other integer valued random variables is the *probability generating function (p.g.f.)*. It is defined by the formula $f(z) = \sum_n p_n z^n$, where $p_n$ is the probability of taking the integer number $n$. The p.g.f. of a fair coin is $f(z) = 1/2 + z/2$. If we drop the condition of nonnegativity of the numbers $p_n$ and suppose only that the sequence $p_n$ sequence is normed, that is $\sum_n p_n = 1$ and $\sum_n |p_n| < \infty$, then we simply say that $f(z)$ is a *generating function (g.f.)* and $(p_n)$ is a *generalized distribution*. In the same sense we can speak of *generalized random variables*.

The absolute convergence $\sum_n |p_n| < \infty$ guarantees that a generalized random variable $X$ which takes the value $n$ with probability $p_n$ can be interpreted and also simulated as a classical random variable that takes the value $n$ with frequency $\approx |p_n| / \sum_{n=0}^{\infty} |p_n|$. But we must not forget that if $p_n < 0$ then we need to interpret this unusual phenomenon. We return to this problem at the end of this paper.

The addition of independent random variables corresponds to the multiplication of their p.g.f.'s (see e.g. W. Feller 1967 Ch. XI.) thus the p.g.f. of the sum of two fair coins is $((1+z)/2)^2 = 1/4 + z/2 + z^2/4$. It seems natural to define the half coin via the generalized g.f.

$$\sqrt{\frac{1+z}{2}} = \sum_{n=0}^{\infty} p_n z^n. \tag{1}$$

According to the binomial theorem

$$\sqrt{\frac{1+z}{2}} = \frac{1}{\sqrt{2}} \sum_{n=0}^{\infty} \binom{1/2}{n} z^n.$$

We can get rid of the somewhat strange choice number $\binom{1/2}{n}$ with the help of Catalan numbers defined by the formula

$$C_n = \frac{\binom{2n}{n}}{n+1}, \quad n = 0, 1, \ldots$$

This number sequence 1, 1, 2, 5, 14, 42, 132, 429, ... occurs in very many seemingly unrelated situations (see H. W. Gould 1985). Since

$$\binom{1/2}{n} = \frac{(1/2)(-1/2)(-3/2)\ldots(-(2n-1)/2)}{n!} = \frac{(-1)^{n-1} 2C_{n-1}}{4^n}$$

for $n = 1, 2, \ldots,$ half coins are random variables that take the value $n$ with probability

$$p_n = (-1)^{n-1} \sqrt{2} \frac{C_{n-1}}{4^n} \; n = 0, 1, \ldots.$$

($C_{-1} = -1/2$ is a convenient definition.) This formula is free from the strange choice function $\binom{1/2}{n}$ but an even more mystical phenomenon appears. In this sequence, every other number (meaning every other probability) is negative.

Formula (1) with $z = 1$ shows that $\sum_n p_n = 1$. We also need to check $\sum_n |p_n| < \infty$. Apply the same formula (1) with $z = -1$ to see that $0 = \sum_{n=0}^{\infty} (-1)^n p_n$, and therefore $\sum_{n=1}^{\infty} |p_n| = p_0 = 1/\sqrt{2} = \sqrt{2}/2$, hence $\sum_{n=0}^{\infty} |p_n| = \sqrt{2}$.

## 3 Fractions of Biased Coins and Dice

The p.g.f. of a biased coin is $p + qz$, where $0 < p = 1 - q < 1$.

Half of this coin can be defined via the generalized g.f. $\sqrt{p+qz}$. Its Taylor expansion

$$\sqrt{p+qz} = \sqrt{p} \sum_{n=0}^{\infty} \left(\frac{q}{p}\right)^n \binom{1/2}{n} z^n$$

shows that for $p \geq q$ the sequence of signed probabilities $p_n = \sqrt{p}\left(\frac{q}{p}\right)^n \binom{1/2}{n}$ is absolute convergent.

The same holds for the cubic root, ... or any other $n$-th root of $p + qz$. Thus it makes sense to speak not only about half coins but also about third coins, etc. Similarly we can deal with half, third, ... dice with as many sides as we wish. A generalized die with $n + 1$ sides ($n = 1, 2, \ldots$ is a random variable $X$ with possible values $0, 1, \ldots, n$ such that the the sequence of probabilities $P(X = k) = p_k > 0 \;\; k = 0, 1, \ldots n$ is non-increasing. It is not hard to show that for every $m = 1, 2, 3, \ldots$ the $m$-th part of a generalized die is meaningful in the sense that if $f(z) = \sum_{k=0}^{n} p_k z^k$ is the p.g.f. of a generalized die and

$$f(z)^{1/m} = \sum_{n=0}^{\infty} a_n^{(m)} z^n, \quad m = 1, 2, 3, \ldots,$$

then for $m = 2, 3, \ldots \sum_{n=0}^{\infty} |a_n^{(m)}| < \infty$.

## 4 The Fundamental Theorem of Negative Probabilities

In the sequence of probabilities $p_n, \;\; n = 1, 2, \ldots$ of a half coin every other number is negative. Does this make any sense, do these generalized random variables have any meaning? Alfred North Whitehead said that, "the point about zero is that we do not need to use it in the operations of daily life. No one goes out to buy zero fish." Similarly, we cannot remove a five-acre swath from a three-acre field, but nothing prevents us from subtracting five from two. In this section we prove that our mystical random variables have the same 'operational meaning' than the negative numbers.

The summation of independent random variables corresponds to the product of their generating functions. Not so long ago we proved (see I. Z. Ruzsa and G. J. Székely 1983 or I. Z. Ruzsa and G. J. Székely 1988) the following.

**Fundamental theorem:** *For every generalized g.f. $f$ (of a signed probability distribution) there exist two p.d.f.'s $g$ and $h$ (of ordinary nonnegative probability distributions) such that the product $fg = h$.*

Thus if $f$ is the generalized g.f. of a half coin $C$, a third of a die, (or any other related mystical object), then we can always find two ordinary coins, ordinary dice (ordinary random object) $C_1, C_2$ such that if we flip $C$ and $C_1$, their sum is $C_2$. In this sense every generalized (signed) distribution is a kind of difference ('so-called convolution difference') of two non-signed (ordinary) probability distributions. This result justifies the application of signed probabilities in the same sense as we use negative numbers.

## 5 An Application of Negative Probabilities in Finance

In the theory of interest (see e.g. S. G. Kellison 1991) the following notation is used traditionally. Payments at times $t = 1, 2, \ldots, n$ are denoted by $R_1, R_2, \ldots, R_n$, and the *discount factor* is $v = (1+i)^{-1}$ where $i$ the *effective* *rate of interest*. The *duration* is by definition

$$d = \frac{\sum_{t=1}^{n} t v^t R_t}{\sum_{t=1}^{n} v^t R_t}.$$

This is clearly an expected value type quantity: if the random variable $T$ takes the values $t = 1, 2, \ldots, n$ with probabilities

$$p_t = P(T = t) = \frac{v^t R_t}{\sum_{t=1}^{n} v^t R_t},$$

then $d$ is the expected value of $T$. What is the advantage of calling the weights $p_t$ probability and the weighted average $d$ expected value? The probability interpretation suggests that the variance of $T$ and other probabilistically or statistically important notions might play some role in this context. And they do. For example the derivative of $d$ with respect to the interest rate $i$ is $-v\sigma^2$, where $\sigma^2$ is the variance of the random variable $T$. Since both $v$ and the variance $\sigma^2$ is nonnegative, $d$ as a function of $i$ cannot increase. This is important in practice. On the other hand it does make sense to suppose that $R_t$ can take negative values; just think of a negative payment which is a withdraw of money. In this case the variance $\sigma^2$ can easily be negative. Negative probabilities and negative variances directly correspond to negative payments, and there is nothing more natural than negative payments. We do it every day when we withdraw money from our bank account.

## 6 The Operational Meaning of Negative Probabilities

Garret Birkhoff, mathematics professor at Harvard, said that 'Everybody speaks about Probability, but no one is able to clearly explain to others what a meaning has Probability according to his own conception.' Bertrand Russel was equally critical in a 1929 lecture when he said 'Probability is the most important concept in modern science, especially as nobody has the slightest notion what is means.' In case of negative probabilities the meaning is even less obvious.

What seems to be helpful if we separate the *operational value of probability* and the *interpretation of this operational value*. Probabilities, even if they are between 0 and 1 need interpretations. For example the so-called frequentists interpret probabilities as limit of relative frequencies. Bayesians have their own interpretation based on the ratio of bets you would be willing to offer. What we claim is that the operational value of probability (the value we work with, say add, multiply, ...) is not necessarily a number between 0 and 1. The operational value can be negative, complex (like the value of the state function in quantum physics) or even more abstract (see G. J. Székely 1976). Then we need to interpret these abstract 'creatures' (similarly to Max Born's interpretation of complex valued state functions in quantum physics). In case of negative probabilities or signed distributions our fundamental theorem mentioned above suggests a natural interpretation. If a random variable $X$ has a signed distribution, we can always find two other random variables $Y, Z$ with ordinary (not signed) distributions such that $X$ and $Y$ are independent and $X + Y = Z$ in distribution. Thus $X$ can be interpreted as the 'difference' of two 'ordinary' $Z$ and $Y$.

Negative probabilities can also have more than one interpretations. Here is a direct and natural one. If $(p_n)$ is a sequence of signed probabilities such that $\sum |p_n| < \infty$ , then $a_n = |p_n| / \sum |p_n|$ is a traditional nonnegative probability distribution but if $p_n < 0$, then we must not forget that in the operational sense $p_n$ is on the other side of the scale, $p_n < 0$ is a debit type probability.

## REFERENCES


■ Barnett, M. S. (1945). Negative probability. *Proc. Cambridge Phil. Soc.*, **41**, 71–73.

■ Dirac, P. A. M. (1942). The physical interpretation of quantum mechanics. *Proc. Roy. Soc. London*, **A 180**, 1–39.

■ Feller, W. (1967). *An Introduction to Probability Theory and Its Application*. 3d ed. Wiley, New York.

■ Feynman, R. P. (1987). *Negative Probability in: Quantum Implications, Essays in Honor of David Bohm*. Eds. B. J. Hiley and F. D. Peat. Roiutledge and Kegan Paul, London, 235–246.

■ Gould, H. W. (1985). *Bell & Catalan Numbers: Research Bibliography of Two Special Number Sequences*. Sixth Edition, Morgantown, WV.

■ Kellison, S. G. (1991). *The Theory of Interest*. Irwin/McGraw-Hill, Boston.

■ Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin. Reprinted: *Foundations of Probability Theory*. Chelsea, New York (1956).

■ Ruzsa, I. Z., Szēkely, G. J. (1983). Convolution quotients of nonnegative definite functions. *Monatshefte für Mathematics* **95**, 235–239.

■ Ruzsa, I. Z., Szēkely, G. J. (1988). *Algebraic Probability Theory*. Wiley, New York.

■ Szēkely, G. J. (1976). Probabilities in operator structures. *Annales Univ. Sci. Budapest, Sect. Math*. **19**, 141–142.

■ (2005, to appear) Székely, G. J. and Rizzo, M. L. *The Mathematical Psychology of Finance*. Springer, New York.

■ Wigner, E. (1932) *Phys. Rev.* **40**, 749. see also in Quantum-mechanical distribution functions revisited. *in: Perspectives in quantum theory*. Eds Yourgrau, W. and , A.van der Merwe, MIT Press, Cambridge MA (1971).