ing up lemma has been a standard tool in IT, but an information theoretic proof of it was found only in 1986, by Marton. Recently, she was able to prove measure concentration theorems, via her IT method, under substantially weaker assumptions than independence; other approaches have not made this possible, so far.

Shannon and Investment

Thomas M. Cover
Stanford University

The theory of growth rate optimal investment is beginning to look more and more like information theory. Although Shannon never published in this area, he gave a well-attended talk on the subject in the mid 1960s at MIT, and the influence of information theory has been substantial.

The first work with an information theoretic flavor in investment theory was Kelly's [1956] BSTJ paper in which he proved that the increase in the growth rate of wealth in a horse race due to side information was equal to the mutual information between the winner of the horse race and the side information. Breiman [1961] generalized this result and proved that Kelly gambling (gambling in proportion to the probability of winning) had a higher growth rate of wealth than any other investment scheme and that it minimized the time necessary for the wealth to achieve a distant goal. In the mid 1960s, Shannon gave a lecture on maximizing the growth rate of wealth and gave a geometric Wiener example.

At about this time, Shannon and Samuelson (a Nobel Prize winner-to-be in economics) held a number of evening discussion meetings on information theory and economics. It is not clear what was said in these meetings, but Samuelson seems to have become set in his views. He published several papers arguing strongly against maximizing the expected logarithm as an acceptable investment criterion. (It happens that maximizing the expected logarithm is the prescription for the growth-rate optimal portfolio.)

For example, Samuelson [1969] wrote, "Our analysis enables us to dispel a fallacy that has been borrowed into portfolio theory from information theory of the Shannon type." Samuelson goes on to argue that growth rate optimal policies do not achieve maximum utility unless one has a logarithmic uility for money. Of course this is the case, but it does not deny the fact that log optimal wealth has an objective property: it has a better growth rate than that achieved by any other strategy. Since growth rate optimal policies achieve a demonstrably desirable goal, growth rate optimal portfolios should only have a utility interpretation as an afterthought. In fact, Samuelson [1979] wrote a paper entitled, "Why we should not make mean log of wealth big though years to act are long." This is a two page paper in words of one syllable that makes the point that maximizing the expected log of wealth is not appropriate. The growth optimal portfolio literature has been slow to develop. It is possible that Samuelson's eloquent admonitions had their effect.

Thorp [1969] proved, among other things, that the growth rate optimal portfolio is not necessarily on the efficient frontier, thus showing the incompatibility of log optimality and the mean-variance theory of Markowitz. Thinking that a portfolio with several good properties must have more, Bell and Cover [1980, 1988] showed that the log optimal portfolio is also competitively optimal. Thus, the long run optimal portfolio is also optimal in the short run in the sense that it outperforms the wealth induced by any other portfolio, at least half the time. This leads to a similar result for Shannon coding, proving that the ideal code lengths log $(1/p(x))$ are competitively optimal as well as expected length optimal [Cover, 1991b].

The value of side information, first investigated by Kelly, was generalized by Barron and Cover [1988] to show that the increase in the growth rate of wealth is less than or equal to the mutual information.

Algoet and Cover [1988], generalizing Breiman, showed that the conditionally log optimal portfolio is asymptotically optimal in growth rate in ergodic markets and the growth rate converges to a constant. As a special case of this convergence argument, a new sandwich proof was given of the AEP.

There have also been some results in establishing universal portfolios—the counterpart to universal data compression. A universal portfolio attempts to do as well on the fly as if one had known ahead of time the precise sequence of stock market returns, and used the best constant rebalanced portfolio. (This restriction of the best constant rebalanced portfolio is naturally motivated since they are optimal for markets which have independent identically distributed investment opportunities.) Cover and Gluss [1986], Cover [1991a], and Ordentlich and Cover [1996, 1998] have proved at various levels of generality, that there exists a universal portfolio achieving a wealth $\hat{S}_n$ at time n such that $\hat{S}_n / S_n^* \geq 2/\sqrt{n+1}$ for every stock market sequence and for every n, where S_n^* is the wealth generated by the best constant rebalanced portfolio with hindsight. In fact, the lower bound corresponds to the associated minimax regret lower bound for universal data compression. And since S_n^*, the best wealth achievable in hindsight, is expected to grow exponentially, the $\sqrt{n}$ term is asymptotically negligible in the exponent. Thus, one has the same asymptotic growth rate of wealth as if one had known the exact stock market sequence ahead of time.

The growth rate optimal portfolio theory development has gone hand in hand with the counterpart-theorems in data compression and universal data compression. Shannon's influence on portfolio theory, although entirely indirect, has been substantial. If one identifies a result as being information theoretic if it involves entropy, mutual information, channel capacity, and asymptotic equipartition properties, one would have to say that this growing segment of portfolio theory is a proper subject of information theory.

References

[1] J. Kelly, "A new interpretation of information rate," *Bell System Tech. Journal*, pp.917-926, 1956.

[2] H. Latané, "Criteria for choice among risk ventures," *Journal of Political Economy*, 67:144-155, 1959.

[3] L. Breiman, "Optimal gambling systems for favorable games," Fourth Berkeley Symposium, 1:65-78, 1961.

[4] P. Samuelson, "General proof that diversification pays," *Journal of Financial and Quantitative Analysis*, 2:1-13, 1967.

[5] P. Samuelson, "Lifetime portfolio selection by dynamic stochastic programming," *Rev. Econom. Statist.*, pp.239-246, 1969.

[6] E. Thorp, "Optimal gambling systems for favorable games," *Rev. Internat. Statist.* 37:273-293, 1969.

[7] P. Samuelson, "Why we should not make mean log of wealth big though years to act are long," *Journal of Banking and Finance III*, pp. 305-307, 1979.

[8] R. Bell and T. Cover. Competitive Optimality of Logarithmic Investment. *Mathematics of Operations Research*, 5(2):161–166, May 1980.

[9] T. Cover and D. Gluss. Empirical Bayes Stock Market Portfolios. *Advances in Applied Mathematics*, (7):170-181, 1986. Summary of this paper appears in: Proceedings of Conference Honoring Herbert Robbins, Springer-Verlag, 1986. Abstract and Summary appears in "Adaptive Statistical Procedures and Related Topics," IMS Lecture Notes Monograph Series, Vol. 8, ed. by J. Van Ryzin.

[10] A. Barron and T. Cover, A Bound on the Financial Value of Information. *IEEE Transactions of Information Theory*, 34(5): 1097-1100, September 1988.

[11] P. Algoet and T. Cover. Asymptotic Optimality and Asymptotic Equipartition Properties of Log-Optimum Investment. *The Annals of Probability*, 16(2):876-898, 1988.

[12] R. Bell and T. Cover. Game-Theoretic Optimal Portfolios. *Management Science*, 34(6): 724-733, June 1988.

[13] T. Cover. Universal Portfolios. *Mathematical Finance*, 1(1): 1-29, January 1991.

[14] T. Cover, On the Competitive Optimality of Huffman Codes. *IEEE Transactions on Information Theory*, 37(1): 172-174, January 1991.

[15] T. Cover and E. Ordentlich. Universal Portfolios with Side Information. *IEEE Transactions on Information Theory*, 42(2):348-363, March 1996.

[16] E. Ordentlich and T. Cover. The Cost of Achieving the Best Portfolio in Hindsight. To appear in *Mathematics of Operations Research*.

Shannon and Cryptography

James L. Massey
Prof. em., ETH-Zürich

Martin Hellman, co-founder with Whitfield Diffie of "public-key cryptography", attributes his interest in cryptography to three sources, one of which was the fact that in 1970, Prof. Peter Elias of MIT introduced him to "Shannon's virtually forgotten 1949 paper on cryptography" (citation from the September 1981 IT-Newsletter article reporting that Diffie and Hellman had received the Donald G. Fink Prize Paper Award). Hellman has stated elsewhere that it was the following words of Shannon from that paper [1] which put him on the path to public-key cryptography: "The problem of good cipher design is essentially one of finding difficult problems, subject to certain other conditions. We may construct our cipher in such a way that breaking it is equivalent to (or requires at some point in the process) the solution of some problem known to be laborious."

But Shannon's contribution to cryptography go deeper that just being the godfather to public-key cryptography. His paper [1] is now generally credited with transforming cryptography from an art to a science. He was the first to give a system diagram of a secrecy system showing the message M, the cryptogram E (which was Shannon's notation for the Enciphered message) and the key K in the manner that (1) M and K determine E, (2) E and K determine M (as is necessary for decryption), and (3) the "enemy cryptanalyst" has access only to E but wants to find M. (To Shannon, M was always the totality of plaintext that is enciphered before the key is changed.)

One great service of Shannon was to say precisely what it means for a cipher to be unbreakable. More precisely, he defined *perfect secrecy* to mean that M and E are statistically independent or, equivalently, that $H(M \mid E) = H(M)$. As my yellowed handwritten notes from one of Shannon's lectures in 1961 at MIT state, "It is obvious that $H(M \mid E) \leq H(K \mid M)$." Hence for perfect secrecy we must have $H(M) \leq H(K \mid M) \leq H(K)$, i.e., the length of the key in binary digits must be at least as great as the number of bits of information in the message that is being hidden. Patent offices around the world will be forever grateful for this proof that there is no unbreakable cipher with a short key. Shannon also demonstrated that the Vernam cipher or "one-time pad" in which a coin-tossing key is added bit-by-bit modulo-two to the message is unbreakable. Vernam knew this, too. In his 1926 paper [2] describing this cipher, he wrote that its unbreakability had been confirmed by field trials with the U.S. Army Signal Corps.

Shannon defined the *unicity distance* of a cipher as the amount of ciphertext that determines the key K essentially uniquely and showed, for what he called a "random cipher", that this was closely approximated by $\frac{H(K)}{D}$, where D is the