Real Estate Review

Volume 47, Number 1 Spring 2018

THOMSON REUTERS®

For Customer Assistance Call 1-800-328-4880

Mat #42028465

The Dilemma[*]

Roger J. Brown, Ph.D.[**]

I. Introduction

If you own your house you have an estate in land known as a "fee interest." Here, the word "fee" has nothing to do with the more common use of the term meaning a charge on the invoice from your lawyer. It comes from the old English word "infeftment" from feudal times, which means bestowing inheritable land on another person.

An intractable mathematical problem can sometimes be solved by a transformation. In matrix algebra this could be a change of basis; in signal processing you might employ an inverse Fourier transform. Doing so permits you to achieve theoretical solutions in the transform space that might not otherwise be possible.

Each of the two paragraphs above represent a challenge to the average person. Both reflect a particular argot for two fields of thought—the first law, the second mathematics—which impact the lives of nearly everyone in civilized society. There is a growing rift between these two. It is unclear how, or if, the gap can be bridged. It is less clear what will happen if it is not.

In this article we describe a problem that society faces and suggest a fruitful direction. In order to frame the issues and contain the problem to some manageable degree, the focus will be on the intersection of law and economics. We don't want to quantify everything,[1] but we appear to be digitizing everything. The purpose of this monograph is to explore the limits of two fields that resist reconciliation.

In 1948, Claude Shannon[2] connected mathematics and communication.[3] Not only did this garner Shannon considerable fame, 70 years later its wisdom is everywhere around us. He taught us that a signal sent through a noisy channel could be transmitted in a way that reduced the error to an arbitrarily small factor. He showed that by using discrete, binary (one and zero) pulses to communicate we can be assured of transmitting information to a close approximation of perfection at the receiving end of the channel. If you received this article by e-mail you received the entire article, all of its characters, in the proper order. The article was not garbled. You have Claude Shannon to thank for that. The theory is elegant and simple. On the other hand, the consequence, that information is a function of probability, is anything but simple.

If we are to negotiate an uneasy peace between lawyers and mathematicians it will be necessary for each side to accept the importance the other side plays in conducting a civil

---

[*] This is the second part of a three part series on decision trees and information theory. The first part is "The Case for Decision Trees in Partition Actions", *The California Real Property Journal*, 2017, Vol. 35, Issue 2, pp. 27-35.

[**] Colvin Institute of Real Estate Development, University of Maryland at College Park, College Park, Maryland, USA. The author would like to thank Tom Compton for very helpful comments. All errors remain the author's solely.

society. Before we embark on that heroic task let's take a short side trip in search of a good motivation for doing so.

II. The Internet of Everything

Try the following experiment in any major city. At your computer's keyboard type the words "bank losses due. . ." into your browser search field. Almost instantly you will be offered an auto-complete that finishes the query with ". . .to cybercrime."[4] Clicking on that link brings up 422,000 results in just over half a second. Having done that, find the telephone number for your local district attorney's office. Dial that number during regular business hours. Imagine that you wish to report a crime or inquire about a trial. You don't have to worry about what you will say when someone answers your call as it is unlikely anyone will. You will encounter an array of prompts and recorded messages. The point of this exercise is not to make public servants look bad. The point is to highlight the difference in efficiency that occurs when you send a signal to Google and when you send a signal to a legal institution. Equally important is the fact that the signal you send Google may be received at or answered from your next door neighbor's garage or India. It may happen at any hour. The call to your district attorney was a local call during the daytime. The digital age is agnostic about territory, national boundaries, and time.

In this context we contemplate informational efficiency. Google uses a highly developed algorithm commonly known as "page rank" to instantly calculate the probability that the answer placed at the top of the search results is the information you seek. How good is that calculation? Think of your own experience with Google searches. How often is what you seek at or near the top of the list of alternatives? In the search described above the 422,000 "hits" constitute many thousands of computer screens or electronic "pages" for you to scroll through. How often do you ever go beyond the first page?

Perhaps we can agree that sound mathematics underlying information theory benefit millions of people every day. We are finished debating the value of Shannon's theory. We have been for a long time.

Contrast what we have just described with the nature of a well-functioning, market based, reasonably secure first-world society resting on laws that align incentives with payoffs through a system of rules. The hallmark of this society, indeed the reason U.S. society is viewed as "safe," is that things change very slowly. Variance (risk) is minimized when speed is reduced. Drive slower and you drive safer. This is hardly consistent with the goal of speed and efficiency delivered by computers.

The central core of the dilemma could be viewed as an extension of the oft quoted remark "Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom."[5] Our question is: Taking information as the most general, when and under what conditions does it become evidence? When is it truth?

Shannon viewed everything as a channel. Taking the legal system as a channel, practically everyone outside of it would agree that it is among the noisiest of channels. It may even be possible that those whose livelihoods depend on their system are adding noise.

III. How data become a decision

Perhaps one of the most used words in the

English language today is "algorithm." Everywhere we turn some "system" is imposing an algorithm on us. Home loans are underwritten with algorithms. Credit card fraud is uncovered by algorithms. Insurance of all kinds is granted and claims settled based on algorithms. An algorithm is just a sequence of steps based on rules that leads to a result. Two simple examples from algebra might be

$$2 + 5 * 5 = 27 \tag{1}$$

$$(2 + 5) * 5 = 35 \tag{2}$$

The reason the answers differ despite each equation having the same digits on the left is the rule that says that multiplication is performed before addition but operations in parentheses are performed before multiplication.

Computers and the internet have made data ubiquitous. Every mouse click is a data point. Your surfing habits and preferences are constantly monitored, recorded and analyzed. Today these data form the basis of widely used algorithms employed in the branch of artificial intelligence known as machine learning. One tool of machine learning is a decision tree. The reason each person on an airplane has paid a different price for his or her ticket is because a decision tree optimized air carrier revenue. If you are unfortunate enough to develop a brain tumor, the shape you see on the film attached to the light box hanging in your doctor's office was made using a decision tree.[6]

This section introduces the methodology by which data turn into decisions. Humans do this every day by employing historical knowledge and experiences to form rules that govern how they act in the future. "Experience is what you get when you don't get what you want" is a popular cliché. Computers just analyze experiential data faster and tirelessly. In this section we develop a simple and stylized example showing how data fed to a machine become a decision. The reader should set appropriate expectations. It is all too easy to become disappointed with the results below. Reactions like "I could have done better than that" are unproductive. The master tailor can usually produce a custom product that out shines the random garment bought off the rack. How many people retain a master tailor? The goal is to understand how we manage a small planet occupied by more than seven billion people in the digital age, not to second guess the "toy" version of an algorithm after the fact. Doing so not only misses the learning experience, in the long run loses the battle. Algorithms persist because they work.

The Shannon insight, that information is a function of probability, leads to the important concept of *information entropy*, a number that results from applying an equation to data.[7] For instance, the entropy of the characters necessary to convey this paragraph to the reader is 4.198. One casual definition of entropy is a measure of uncertainty, the lower it is the less uncertainty in the message to be resolved.[8] What will seem foreign in what follows is crucial: any stream of ones and zeros provides information that can be quantified on the basis of how it improves one's position from the situation which existed prior to the receipt of the information. Let's use some simple examples to start.

Suppose I retain you to draw a contract for me. You do so and the following three passages appear in the document. For each, we compute its entropy in Figure 1.

PHRASE	ENTROPY
party of the first part	3.36031
Brown	2.32193
he	1.

Figure 1

All three may describe the same person. The fact that entropy falls with the length of the phrase is incidental.[9] The important point is what happens when we combine two of the phrases. Knowing nothing of the facts, "party of the first part" leaves much to the imagination. But when you subsequently add one of the other two words you *may* know the sex of the party or the party's name. The combination of two pieces of information results, naturally, in information gain.[10]

The previous might be called "anecdotal evidence" frowned upon as a means to reach general conclusions or to form a universal rule. Data are just anecdotal evidence in bulk. The next step is to take a critical look at some amorphous data and how they might be organized and used to create a rule. We chose real property title law as the setting. The data size is small to make our example manageable as a teaching tool. As you consider what follows it is useful to keep in mind that the property ownership recording system has millions of data points. Imagining how the predictive value of the model improves with more data is important to appreciating the example below.[11]

IV. The Simplest Decision Tree

Figure 2 shows our data having ten observations (rows) of persons who requested title insurance. Each person has four attributes and an outcome.[12] We will use a decision tree and entropy to create rules from these data based on how information improves as we analyze it.

Data

Attributes				Classes
Type	**Exceptions**	**Endorsements**	**Lender**	**Outcome**
Residential	0	Few	None	Issued
Residential	1	Few	Institutional	Issued
Commercial	0	Few	None	Issued
Residential	1	Few	Institutional	Issued
Commercial	1	Few	Institutional	Loss
Commercial	1	Many	Private	Cancelled
Residential	2	Many	Institutional	Cancelled
Commercial	2	Many	Private	Cancelled
Residential	0	Usual	Institutional	Loss
Commercial	1	Usual	Institutional	Loss

Figure 2

Just as we know millions of people use the public records system, we can also agree that they each will exhibit far more than four characteristics. One could create a column to record such trivial things as owner's age, or more meaningful descriptors such as lot size. In any event we accept that many more columns could be added but for now we will keep only these four.[13]

Next we need to measure how our situation improves with more information. We have data describing ten outcomes in the far right column, IIIILCCCLL, a string for which the entropy is 1.571.[14] Surely we are less uncertain when we know something about the property in question? But how much less uncertain are we? The answer is the difference in entropy before and after. This is the information gained by the having mutual information. Mechanically, we split the data along the lines of each attribute and choose for our decision rule the one providing the most information gain.

We begin by inquiring which attribute is most useful. Figure 3 shows that Endorsements offers the most information gain. It is that attribute that becomes the "root" of our decision tree.

Gain	Type	Exceptions	Endorsements	Lender
Entropy	0.125	0.534	1.210	0.695

Figure 3

We now explore how the answer for Endorsements, 1.210, was derived. First we isolate the Endorsement attribute paired with the outcomes, then we split and aggregate subsets in line with the three possible values Endorsements can take (Few, Many and Usual) and calculate entropy for each separate further subset in Figure 4.

Attribute	Classes
Endorsements	**Outcome**
Few	Issued
Few	Issued
Few	Issued
Few	Issued
Few	Loss
Many	Cancelled
Many	Cancelled
Many	Cancelled
Usual	Loss
Usual	Loss

4I, 3L, 3C or (IIIILLLCCC)

Entropy 1.571

Endorsements	Classes
Few	Issued
Few	Issued
Few	Issued
Few	Issued
Few	Loss

4I, 1L or (IIIIL)

Entropy 0.722

Endorsements	Classes
Many	Cancelled
Many	Cancelled
Many	Cancelled

3C or (CCC)

Entropy 0.000

Endorsements	Classes
Usual	Loss
Usual	Loss

2L or (LL)

Entropy 0.000

Figure 4

When a sub-set of attributes has all the same outcomes we call that "pure," which means that it has no variation, making entropy zero and offering no information gain, which is the case for Many and Usual.[15] Because that might not always be the case we show all three sub-attributes in Equation 3, which is beginning entropy reduced by a weighted average of the entropies for the individual sub-attributes.[16]

$$\underset{\text{Original Uncertainty}}{1.571} - \underset{\text{Revised Uncertainty}}{\left(.722 * {}^{5}\!/_{10} + 0 * {}^{3}\!/_{10} + 0 * {}^{2}\!/_{10}\right)} = \underset{\text{Information Gain}}{1.210} \tag{3}$$

The information gain is the right side of the equal sign Equation 3. In the parentheses is the uncertainty *after* we have considered the new information.

This process is repeated for the other attributes until we reach the point where the remainder is pure and the tree terminates. The result is the decision tree shown in Figure 5. We refer to boxes from which there are further splits into separate branches as "nodes" and the final boxes from which there are no further steps as "leaves." Note that all leaves represent choices from the data column "outcome." From the decision tree we can consider sets of attributes not shown in our data and make predictions. Construct a new observation: the buyer of a residential property with many endorsements and two exceptions who uses an institutional lender. We can follow the steps in the decision tree to make predictions about the outcome this new person may expect.

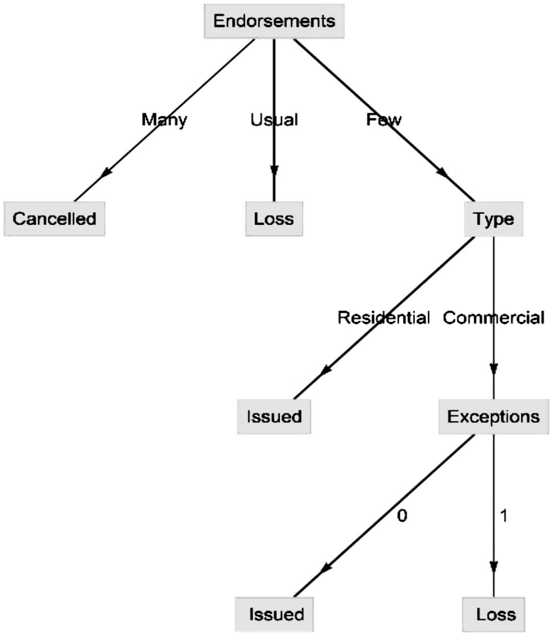


Figure 5

The decision tree produces the following rules:

1. If there are many endorsements the request is cancelled;

2. If there are the usual number of endorsements the policy results in a loss;
3. If there are few endorsements *and* type is residential the policy is issued;
4. If there are few endorsements *and* the type is commercial *and* there are no exceptions the policy will be issued; and
5. If there are few endorsements *and* the type is commercial *and* there is one exception the policy will result in a loss.[17]

The alert reader will likely experience some dissatisfaction at this point. For instance, there is no mention in our decision tree of the lender. The reason for this occurs in the steps we skipped over in sequentially splitting the data. In this case the last node could have been either exceptions or lender as they each offered the exact same information gain. Either could have been chosen. Exceptions was chosen arbitrarily for exposition purposes. Also, our new player from above is counterintuitive. In choosing a lender, what difference does it make to the residential borrower if the title policy has many or few endorsements? In any event our decision tree results in cancellation if there are many endorsements regardless of exceptions or lender. There are a number of reasons for this discomfiture that deserve mention.

1. Ten observations are barely a dataset, more like a collection of ten anecdotal snapshots. The problems with small datasets are many and the best answer to most of them is to collect more data.
2. There is always opportunity to question the model. Decision trees are simple but not optimal. It was chosen for this article because it is relatively easy to explain. Numerous algorithms are available which improve on the decision tree or offer different approaches.[18]
3. The example we have been using is known as a "greedy" algorithm in that its focus is on local optimization, not global optimization which is the goal of most machine learning.

What is important here is not the shortcomings of decision trees. Rather it is to recognize that algorithms exist, simple ones are easy to understand, there is a way to quantify information and that such a quantification leads to improved predictive accuracy. Apply this knowledge to a large health insurer, credit card issuer or retailer with millions of customers, each providing thousands of mouse clicks a minute. Under those conditions many of the problems we found here go away, especially with advanced mathematics and programming talent.

V. What does it all mean?

Suppose I ask for your acquiescence to everything I have written here. My guess is that many would not oblige, rationally and appropriately disagreeing with much if not all of what I have said. On the other hand, it is possible that *all* readers have at one time or another encountered and clicked on the icon shown in Figure 6.

Figure 6

Underlying each of the variants[19] of Figure 6 are legal documents that purport to bind the user to agreements. It is nearly certain the user has not read these. Agreements involve rights, rarely in these documents are they given; more likely they are taken away. Waivers may be included that are non-waivable as a matter of public policy in some jurisdictions. None of this matters to the authors of terms and conditions. Nor does it matter to the user. Should one actually navigate to the printed area of the terms and conditions it is likely what is found is so one-sided it makes a Landlord's Lease look like a Valentine. But the point is that no one cares.[20]

If Shannon did not intend that his Theory of Communication involve meaning he is close to getting his wish. Contracts that are not read and cannot be enforced lean in the direction of meaningless.

All contracts involve costs and benefits. Comparing costs to benefits to determine if the benefits outweigh the costs for each party is the essence of a rational market transaction. The reality is that millions, perhaps billions, of people each day proclaim that they do not care. Engaging in a sort of virtual lottery, they uncritically click on a button behind which may be either bounteous reward or ruinous consequences in order to obtain immediate delivery of a stream of ones and zeros. If humans are rational this can only mean that no one fears the consequence of agreeing. Their decision tree looks like Figure 7. We must presume that the greatest information gain is by agreeing, that large numbers of people believe that more uncertainty will be resolved by proceeding than by refusing to accept the terms and conditions. Those large number constitute a market for information seeking and that market has spoken.

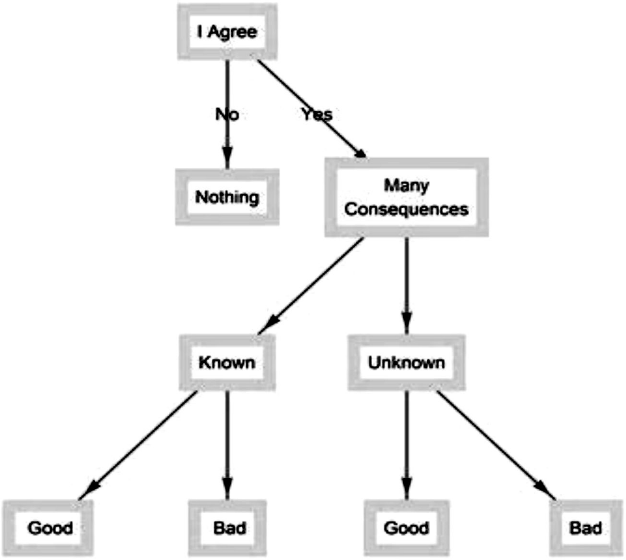


Figure 7

An alternate interpretation is that using the Internet amounts to the cyber equivalent of Pascal's Wager.[21]

VI. Conclusion

Lawyers shun mathematics.[22] The separation of church and state is barely more ingrained in our society. Yet time and again mathematics are found in the courtroom, sometimes with disastrous results.[23] We once divided the world into alpha and numeric with the former the domain of the law. Now even the alpha side of life is digitized. With territorial boundaries erased by satellite transmission jurisdictions blur at best, disappear at worst.[24]

The dilemma that lies at the heart of this paper is the acceptance that information is a function of probability. The thread between that Law of Nature and the laws of man is frayed by a digital age that elevates probability over physical limits. We may be a Nation of Laws but we are a Universe of Algorithms. Heretofore the laws of men have been somewhat enforceable when national boundaries are clear and difficult to penetrate.

That has changed. Consider the phrase:

> "In the context of computer crime, the inexorable connection between the Internet and interstate commerce may sometimes be sufficient to satisfy the jurisdictional element of the statute at issue."[25]

The string above has entropy of 3.98. What would it take to reduce that number and provide more certainty? Removing "may sometimes be sufficient to" is one way.[26] Contemplate how much time the hacker in the Ukraine spent pondering his future after he read those words. What is evident from a superficial search of any source is that cybercrimes are very hard to prosecute, not least because the alleged perpetrator and the victim are in different countries.

How long will it be until the word "country" has little or no meaning? To have meaning, the phrase "A Nation of Laws" requires a Nation.

The Dilemma

Lawyers are no strangers to probability. They are risk minimizers. Implicitly they know the glass being half empty is a function of probability. The menu of prompts at the DA's office is an interactive decision tree. Malpractice insurance is priced based on conditional probability. Every lawyer assesses the probability of prevailing at trial and updates that assessment with each newly *discovered* fact. Becoming acquainted with the mathematical formalities of probability offers an added and useful dimension to the practice.

Cyber-math is a force to reckon with. It threatens the law. On the other hand, slowly evolving laws of men make for a safe physical environment in which computers and geeks can thrive. The truce one imagines has the law speeding up and the digital age slowing down such that a peaceful and orderly convergence is possible. This is decidedly not what either party has in mind. But the alternative is a train wreck and it is not a pretty sight.

NOTES:

[1] The law covers such things as contractual disputes which lend themselves to quantification. It also covers child custody which resists the same treatment on the individual level but becomes quantitative as a budgetary issue at the policy level when foster care and local agency child protective services come into play.

[2] Soni, J. and Goodman, R., A Mind at Play: How Claude Shannon Invented the Information Age, 2017, Simon and Schuster, NY.

[3] Shannon, C. E. (1948). "A Mathematical Theory of Communication." *Bell System Technical Journal* Vol. 27 No. 3, pp. 379-422.

[4] This, for reasons that should become clear, is constantly changing. The result described was at the time the first draft of this paper was being written. When the final draft was being prepared the result was ". . .to data theft." As more information become available the result will and should change again.

[5] Clifford Stoll, Ph.D., astronomer, author and pundit.

[6] The actual routine is random forests, for which decision trees are the progenitor. On September 7, 2016, a Google search on the terms "digital imaging random forests" yielded 293,000 hits.

[7] For those technically inclined, information entropy, computing in units of "bits", $H = -\sum_{i=1}^{n} P(x_i) Log_2 P(x_i)$ where $P(x_i)$ is the probability of an individual character, x. Cf. Brown, R. "Entropy – What Kind of Bet is Real Estate – REALLY?" *Journal of Property Investment and Finance*, 2017, Vol 35, No. 3, pp 341-351.

[8] What is to be resolved is only whether the same message that was sent arrives in the same form. It is important to point out that "less uncertainty in the message" does NOT equate to "less uncertainty in *the meaning of* the message." Much has been written concerning whether Shannon ever intended to include "meaning" in his work on communication.

[9] A phrase of any length with the same repeating character ("ccccccccccccccccccc" for instance) has entropy of zero because it has no variation.

[10] Also known as mutual information.

[11] The example uses the ID3 algorithm and is similar to one found involving family law described in Hunter, D. and Zeleznikow, J. Building Intelligent Legal Information Systems (Computer Law, No. 13) 1994, Springer Verlag, Heidelberg.

[12] Requesting title insurance can result in the request being cancelled or the policy being issued. The third outcome, Loss, must assume the policy was issued. In the list of outcomes we could have had two titles, one "Issued – no claim", another "Issued – Loss" which only presents a labeling problem as both begin with the same letter. This does highlight a matter of greater consequence. There are a number of branching options. One of these could be "Claim – no loss" and "Claim – Loss." As always, an important part of the process is proper set up of the problem.

[13] Adding columns leads to what data scientists call "over-fitting," which is related but not the same as what statisticians call "over-determined," subjects of concern but beyond the scope of this article.

[14] .4 $Log_2(.4)+.3\ Log_2(.3)+.3\ Log_2(.3)=1.571$ where the decimals are the fraction of time each character appears in the string.

[15] When there is no variation we have certainty which is probability = 1 (see fn 9). Recall that the Log of 1 is zero, making the result of H zero (see fn 11).

[16] Note the fractions are composed of the proportions the sub-attributes represent and add to one.

[17] It is left as an exercise for the reader to determine why those with two exceptions do not appear in the tree.

[18] Some keywords to search on for more information in this area include "OLS Regression", "Logistic", "k-Nearest Neighbors", "Clusters", "Hidden Markov", and "Feature Extraction" among many others.

[19] Including but not limited to "I accept the terms and conditions", "I consent to the terms of service", "I accept the privacy policy" and the like.

[20] This is admittedly overstatement. The firm who hired and paid their attorney to write the terms and conditions certainly cared. The vanishingly small percentage of cases where the terms and conditions are actually litigated attest that someone, somewhere cares, but rarely.

[21] Pascal, B., Pensees, Section 233, (1670) available at Project Gutenberg http://www.gutenberg.org/ebooks/18269?msg=welcome_stranger#p_233. See also Hajek, Alan, Pascal's Wager," Stanford Encyclopedia of Philosophy ht

tps://plato.stanford.edu/entries/pascal-wager/index.html. Some have characterized Pascal's Wager as a matrix describing decision under uncertainty which is approximately a tabular form of a decision tree.

[22] Perhaps the best indictment of courtroom mathematics is Tribe, L. H. *Trial by Mathematics: Precision and Ritual in the Legal Process*, Harvard Law Review V. 84, No. 6, April, 1971. Tribe is a lawyer whose undergraduate work was in mathematics. Others followed: Schneps, L. and Colmez, C., Math on Trial 2013 Basic Books, NY; O'Neil, C., Weapons of Math Destruction, How Big Data Increases Inequality and Threatens Democracy, 2016 Crown publishers, NY; Ellenberg, J. How Not to be Wrong: The Power of Mathematical Thinking, 2015 Penguin Books, NY and Kaye, David H., Double Helix and the Law of Evidence, 2010 Harvard University Press, Cambridge MA.

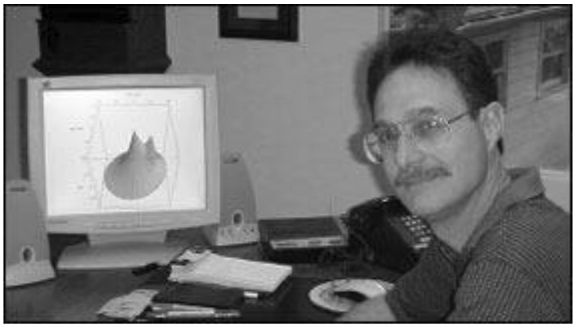
[23] The most famous and tragic case being that of Sally Clark. See Batt, John, Stolen Innocence, 2005 Ebury Press, London. UK and Brown, R. *Sally Clark – What Went Wrong*, 2014 MCLE Course manual summarized at Mathestate.com http://www.mathestate.com/Sally%20Clark%20-%20What%20went%20wrong.pdf.

[24] Cf. Julian Assange, Bradley (Chelsea) Manning, Edward Snowden and Hillary Clinton.

[25] Scott Eltringham, Ed., Prosecuting Computer Crimes, Office of Legal Education Executive Office for United States Attorneys, p. 113.

[26] Which leaves entropy of the remaining, somewhat ungrammatical, sentence at 3.97. This is offered tongue-in-cheek as we again stray, as it is so hard to avoid doing, into the quasi-forbidden territory of "meaning".

About the Authors



Roger J. Brown, Ph.D.

Roger Brown has 48 years' experience negotiating investment real estate sales, exchanges and leases; originated and serviced mortgage loans; fiduciary responsibilities of general partner; expert witness in civil litigation; writing and teaching in the field of real estate finance, appraisal, taxation, analysis and management. General Manager of firm established in 1967 that has had as many as 100 employees. Supervised all aspects of finance, legal and accounting administration for investment portfolio.