And the Winner is . . .*
After 250 years the Rev. Bayes finally finds Redemption

*Roger J. Brown Ph.D.***

This article elaborates the computations required to optimize using the Principle of Maximum Entropy.

I. Introduction

In the first of this series we discussed the structure of decision trees and how outcomes occurred at the end of a string of sequential, mutually exclusive and collectively exhaustive binary events, each choice associated with a probability. In the second part we described how decision trees, entropy and information gain combined to turn a dataset into a decision. This final part extends those concepts further, elaborating the computations required to optimize using the Principle of Maximum Entropy ("ME").

As background the reader is referred to the long-running debate between two camps of probabilists: The Frequentists and the Bayesians.[1] This monograph hopes neither to resolve that controversy, nor even sway either side. However, this effort *is* motivated by a comparison of those methodologies in the context of real estate decision making.

We assume that the reader is familiar with classical statistical reasoning, the Law of Large numbers, the Central Limit Theorem, parameter estimation, hypothesis testing and the like.[2] That suite of tools has dominated for so long that it may be referred to as The Party Line. We will *not* assume an understanding of the alternative: The work of the Loyal Opposition, commonly known as The Bayesians. At the risk of not giving Bayes his due, a short introduction of his method follows.

II. The Bayesian Approach

Thomas Bayes (1701-1761) was a clergyman and self-taught mathematician. His now famous theorem, published after his death,[3] is Equation 1:

*This is the third 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. The second part is "The Dilemma," *Real Estate Review*, Spring, 2018 Vol. 47, Issue 1, pp. 75-88.

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

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \quad (1)$$

Where:

- A and B are events, $P(B) \neq 0$;
- $P(A|B)$ is the conditional probability of event A occurring given that B is true;
- $P(B|A)$ is the conditional probability of event B occurring given that A is true; and
- $P(A)$ and $P(B)$ are probabilities of observing A and B independently of each other.

Equations, even ones involving simple algebra, can be forbidding. A little intuition can overcome this. Suppose you get into your car to drive from your home to work, a trip you have made many times. From experience you know the fastest route and the time it takes. On average you arrive at your office 14 minutes after you leave your home. Because of street lights, weather and other uncertainty you know that the 14 minute trip can vary by two minutes either way. Your experience, just described, is data. This may easily be transformed into a probability distribution. The 14-minute average is your expectation (also known as the arithmetic mean) and the four-minute (two minutes on each side of the mean) variance is, well, the variance from your expectation.[4] This is the classic frequentist approach most of us learned in school.

Rev. Bayes saw things differently. Starting with history and its expectation, Bayes viewed your data as a probability distribution formed around a series of successes or failures to arrive in 14 minutes. You either did or did not arrive in that period of time.[5]

Herein lies the clash between the Frequentists and Bayesians. Clergyman Bayes viewed probability as *a degree of belief* that could be adjusted with the advent of new information. Thus, upon leaving you believed *to some degree* that you would arrive in 14 minutes. But if, as you left your driveway, you heard on the radio that there was a blockage (accident or construction) on the road ahead, reducing the four lanes to two lanes, that new information (new data) causes you to adjust your belief that you will arrive in 14 minutes. Bayes described the belief you held as you entered your car as your prior conditional probability (often referred to as "your prior") and he named the altered belief you computed after hearing news about traffic conditions your posterior conditional probability (many times just "your posterior"). This example is useful to anyone who has driven a car and has had the experience described.[6] The genius of Bayes is the generalization of this concept found in the transformation process that causes your degree of belief to change from prior to posterior. That transformation represents a kind of learning. As one accumulates data one learns more about a hypothesis and changes one's belief about the probability of any outcome implied by that hypothesis.

Examination of the components of Equation 1 in Figure 1 discloses the remarkable simplicity of Bayes' Theorem.

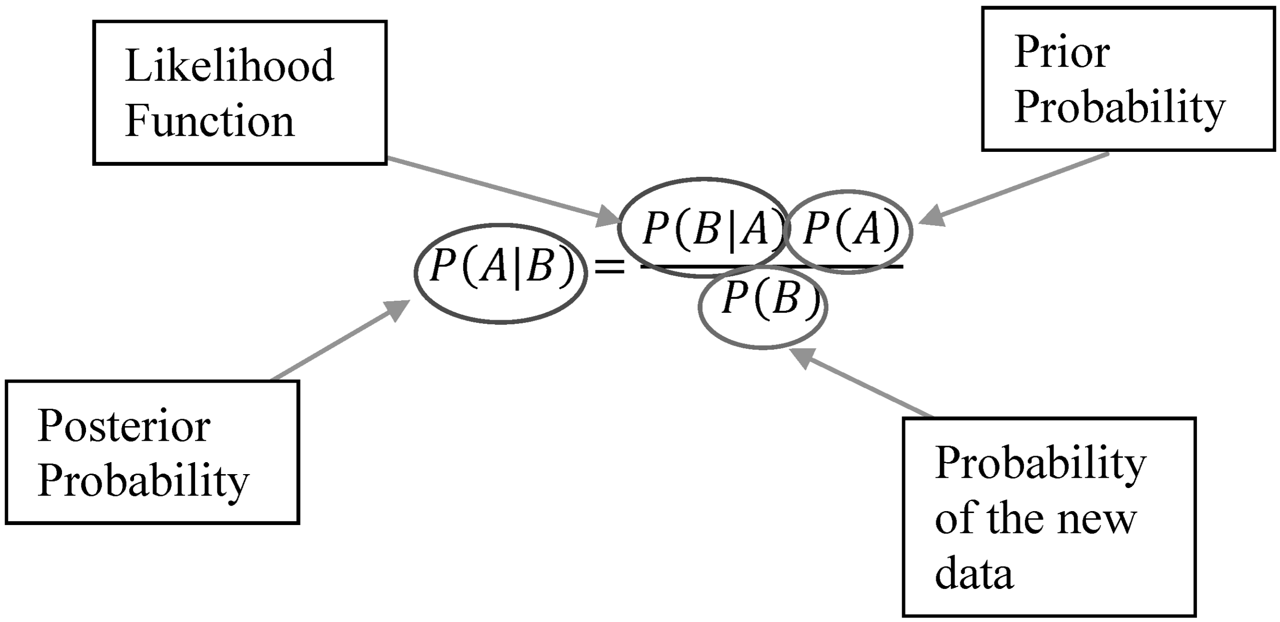

Figure 1

In words, Figure 1 tells us that our process of learning, manifested by the posterior probability (updated prior) on the left-hand side, is the result of multiplying our prior probability by the ratio of the likelihood function[7] to the probability of the data (the latter acting as a normalizing constant). Getting to work on time, like most things in life, is a real estate problem. It requires a simple, understandable algorithm everyone can use.

In Figure 1 our prior is the probability of event A, P(A), with A being the event that we will arrive in 14 minutes. Event B is the blockage, so our posterior is the probability of arriving in 14 minutes, given that there is a blockage, P(A|B). Suppose you kept track and knew that of the last 100 times you drove to work you arrived in 14 minutes 95 times. Therefore P(A) = .95, the simple ratio of successes, 95, to attempts, 100. Being delayed can result from a number of causes, weather, a flat tire, returning home to get the file you forgot, or a blockage due to accident or construction. In your 100 trips you were delayed once due to a blockage on the road. Again the ratio of delays due to blockage, 1, to total attempts, 100, makes P(B|A) = .01. The denominator or normalizing constant, P(B), involves the probability of a blockage anywhere and is a little more involved. For this we need to know the general or global delay factor. Suppose we have data on all freeways over many different 100 days from which we learn that, on average, the chance of blockage on *any* route on *any* day is 1 in 50 or .02. We then carefully collect all the *possibilities.* We can either arrive on time or not and for different reasons. All must be accounted for in the denominator. These are the sum of the probability of being delayed by a blockage along our route (.95* .01 = 0.0095) plus the probability of being delayed for some other reason, .02 * (1-.95) = .001. Adding these we finally have the denominator of Bayes equation, .0095 + .001 = 0.0105 = P(B). Putting it all together into Equation 1 we have

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} = \frac{.01 * .95}{.0105} = 0.905$$

The act of moving the probability you will arrive at work in 14 minutes from 95 percent to 90.5 percent as a result of new information (data) is known colloquially among Bayesians as "revising your prior." It is one of the most common acts of mankind, unavoidable in nearly every aspect of life.[8]

III. A Real Estate Example of Bayesian Computations

Placing this in a real estate investment risk context, imagine that you own a residential rental property, say a 50-unit apartment building, in a neighborhood where tenant turnover is common. You must choose tenants carefully to avoid lease defaults. In your initial months of ownership you screen tenants based on the cash in the bank they show on their application. After a short period of time you have 30 units turnover but of those only 10 of the tenants perform their leases, the remainder default. You think of this as 10 successes in 30 draws. Naturally you are not happy with a success rate less than half.

Next door to your building is a much larger apartment building containing 450 units. One day you meet the owner and compare notes on tenants, operations, and other management practices. He informs you that, in his earlier days and over a similar period of time, he had also used bank savings as his primary screening tool. He informs you that his experience for 220 units that turned over was that only 40 tenants chosen that way performed their leases and 180 did not. He says that he switched to employment longevity as his primary screening tool and that his tenant defaults had fallen dramatically. Respecting his longer and more robust experience you decide to change your tenant selection method to an employment based metric and find that you enjoy similar results.

Reasoning with conditional probability is one of the most powerful and ubiquitous intellectual constructs available. Because of computers and the internet, this wizard-behind-the-curtain now dominates our lives. Every mouse click is a data point. Every "If-Then" instruction in a spreadsheet is conditional probability. The input-response variables in econometrics are another example. It is literally all around us, incrementing continuously and instantaneously. It also is nicely illustrated with decision trees. Figure 2 describes the landlord dilemma in symbols where "T" is Tenant, "Sm" is the small building, "Lg" is the large building, "Perf" indicates the tenant performed and "Def" means that the tenant defaulted. The first pair of branches represents our two buildings, the next branching from each building node applies the condition of the previous branch (which building the tenant came from) to the outcome. Finally, on the far right are the probabilities (which sum to 1) of the only four possible outcomes.

Bayesian reasoning permits us to update our naïve prior with new information to form an improved posterior. Using the information in Figure 2 we can answer the question, given that a tenant defaulted, what is the probability that he lived in the larger (or smaller) building?[9]

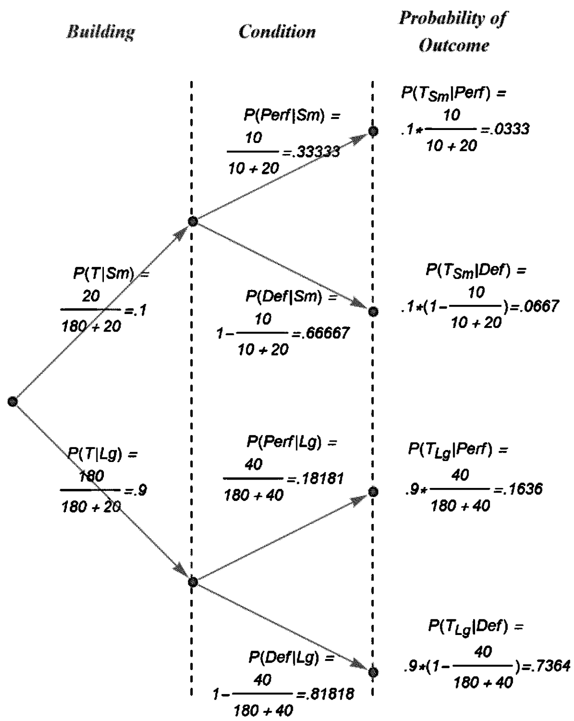

Figure 2

Figure 2 provides all the information we need to compute posterior probabilities. Importantly, the computation is simple, composed of a series of fractions. Finally, it is intuitively reasonable in that it simulates our usual learning experience. First we learn there are two buildings, then we learn the number of occupants in each, then we learn the default histories in each, finally we compute all four possibilities. Tenants are either in the large or small building and they either default or they perform. At each step new data arrives, we learn something new, update our information and recalculate our probabilities.

If you listen closely you can hear the howls of protest from the Frequentists. "Of course," they say, "what do you expect when you have more data? That is the consequence of the law of large numbers that underlies classical statistics. As n grows without bound the estimated and actual probabilities converge."

Indeed, there is a continuous analog to the two landlord story. Suppose the owner of the smaller property models his turnover risk as a Beta[10,20] distribution.[10] The probability density function ("pdf") would look like Figure 3 where we illustrate an arbitrary .21 outcome. Consulting the Cumulative Distribution Function ("CDF") for the same parameterized distribution tells us we have a 6.5659 percent chance of a .21 outcome.

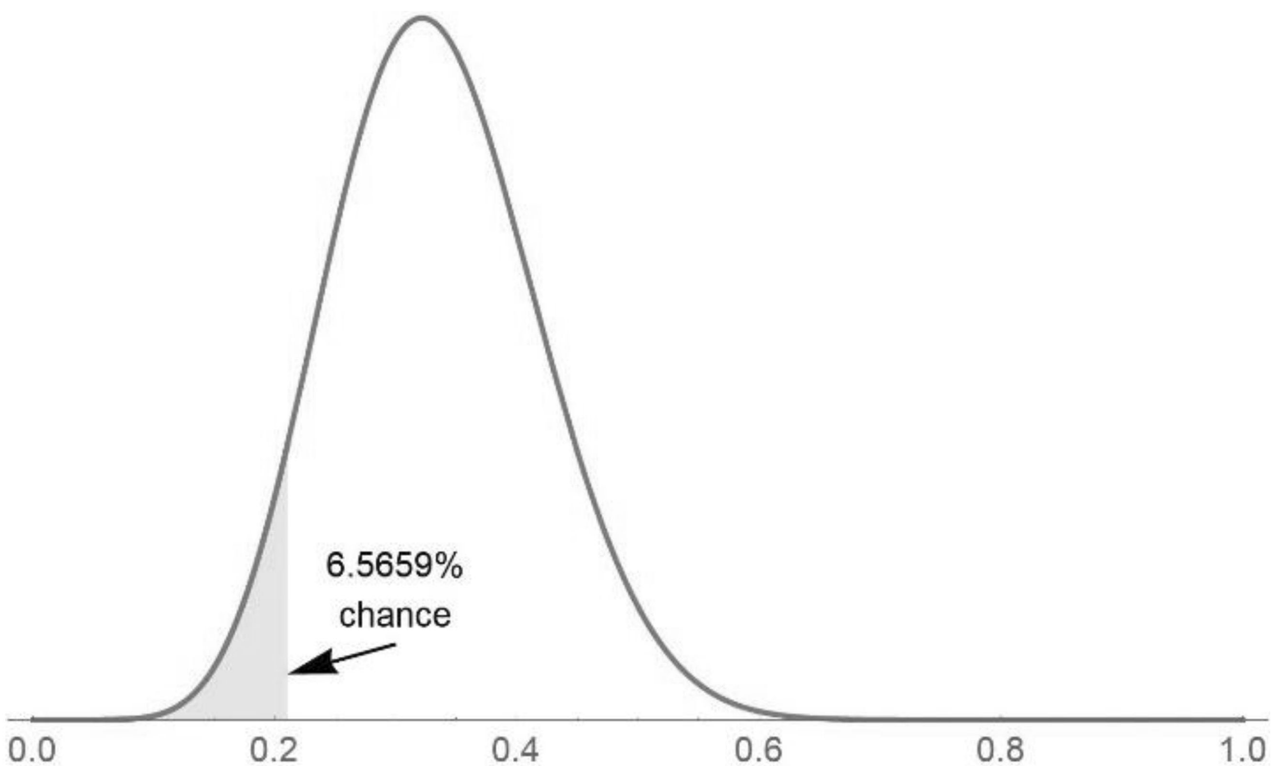

Figure 3

When we update our prior with the experience and data from the larger building the distribution changes dramatically, as in Figure 4 which shows the posterior distribution and the prior on the same plot. Consulting the CDF, this time for the posterior distribution, again using an arbitrary ..21 outcome we find the probability to be 16.9905 percent rather than 6.5659 percent. We have thus improved our prediction by updating our prior with new information.

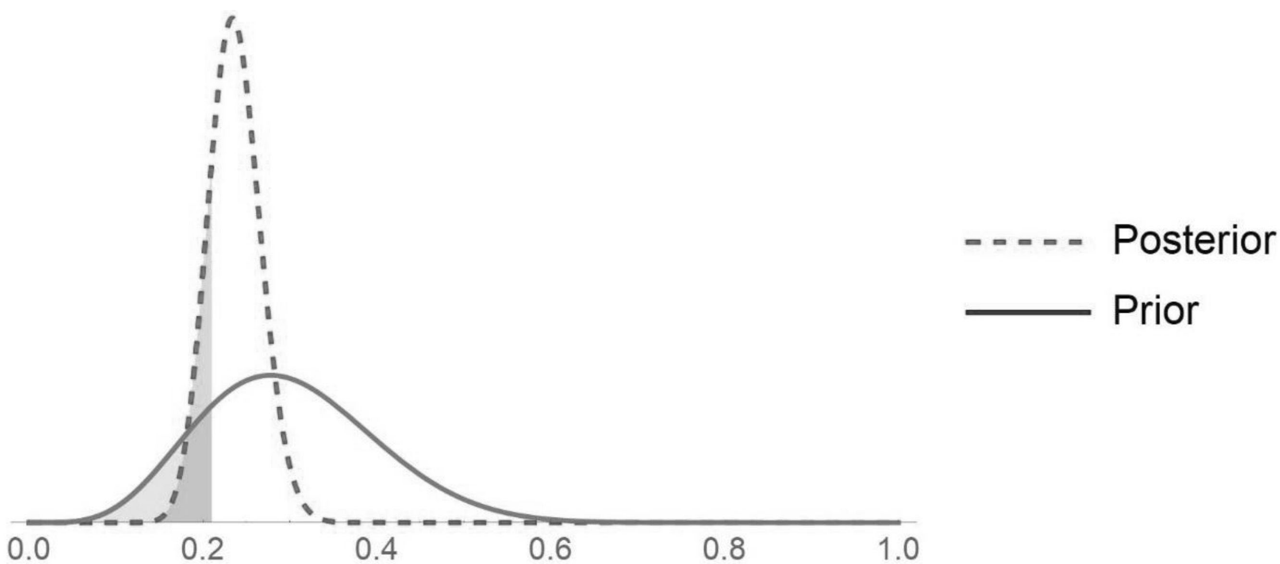

Figure 4

While both academically correct and mathematically elegant, this equivalent presentation suffers on two fronts. First, it requires calculus to compute and possibly to understand, something fewer consumers of information are equipped to do. Real estate brokers are just trying to get to work on time. Second, and more important, are the philosophical underpinnings of the Frequentist approach which holds that the answer is correct *in the limit*, meaning that if *n* is allowed to increase to infinity that is the only answer you can count on. Waiting on infinity is a lonely vigil. Godot will arrive first. Most mortals tiptoe through life gathering information as they go, updating their priors, improving their predictive ability and making decisions without ever arriving at infinity. Infinity is the place where you have *all* the information. You never do. The art of business is making decisions before all the information is in. That is what real estate professionals do.

IV. The Principle of Maximum Entropy

Frequentist or Bayesian, rational people make decisions about a variety of choices, many mutually exclusive, requiring one to forego otherwise attractive alternatives. The common tool for making these decisions efficiently is optimization in the face of constraints where one maximizes some potential uncertain gain or minimizes some potential uncertain loss while limited by a set of exogenous factors.

Often in investment applications one maximizes expected value, another elegant model not easy to put into practice. Expected value is a tool of the Frequentists. It has the (false for Real Estate) underlying assumption that the true final answer emerges at the end of a large number of (independent) repeated trials. Those enamored of coin tossing to support their theories should flip a building in the air a few times and count the number of times it lands on its foundation.[11] Frequentism is marginally possible with financial assets provided one is willing to accept some fairly strong assumptions. For Bayesians and real estate it is impossible because of the small datasets, the uniqueness of the asset, its fixed location and the irreversibility arising from constructing a single alternative on the land.

The preferred optimization tool for the Information Age is Maximum Entropy ("ME").[12] The general idea is to segregate our knowledge from our ignorance very carefully so that the update of our prior is as pure as possible. That is, update of the prior is unsullied by any bias that may spill over from the information upon which we formed that prior. Equation 2 is the entropy function to be maximized.[13]

$$S = -\sum p_i Log_2[p_i] \qquad (2)$$

V. A Second Real Estate Example

Assume we have 25 acres to develop. Possible zoning classes are Residential, Industrial or Commercial. Each zone has different costs (studies, reports, bribes, waivers etc.) associated with bringing them to the local authority for approval. Approval is based on the project size and the usual NIMBY provincialism of land use politics. Industrial absorbs all the available land, the alternate uses leave room for other adjacent buildings or open space, all factors that affect the outcome. The voting records and land use preferences for individual City Council Members are known. There are nine council members, eight are equally divided, four may be expected to be reliably against; the other four normally in favor of development proposals. The ninth swing vote has an erratic but documented voting history for various land use types. We have the distribution of past votes. The summary is shown in Table 1.

Zoning	Cost	Acres	Approval Probability	Disapproval Probability
Residential	$1000	15	0.5	0.5
Industrial	$2000	25	0.8	0.2
Commercial	$3000	20	0.9	0.1

Table 1

Our interest is in the probability that a particular land use will be approved, using only this data. We have two constraints at present. The first is that all the probabilities must add to 1.[14] We also have data showing that, on the average, developers spend $1,750.00 to get a project approved.[15] Thus, our two constraints are Equations 3 and 4.[16]

$$1 = P(Res) + P(Ind) + P(Com) \quad (3)$$

$$1750 = 1000\,P(Res) + 2000\,P(Ind) + 3000\,P(Com) \quad (4)$$

With three unknowns and two equations we cannot solve directly for the probabilities. But some algebraic re-arrangement solves for one of the variables in terms of the other variables to produce Equation 5 ready for optimization.

$$S = -(0.25 + P(Com))Log_2[0.25 + P(Com)] - (0.75 - 2\,P(Com))Log_2[0.75 - 2\,P(Com)] - P(Com)Log_2[P(Com)] \quad (5)$$

Using Newton's method in a root search we discover that S is maximized at 1.51652 when P(Com) = 0.21624, shown as the point at the apex of the plot in Figure 5.

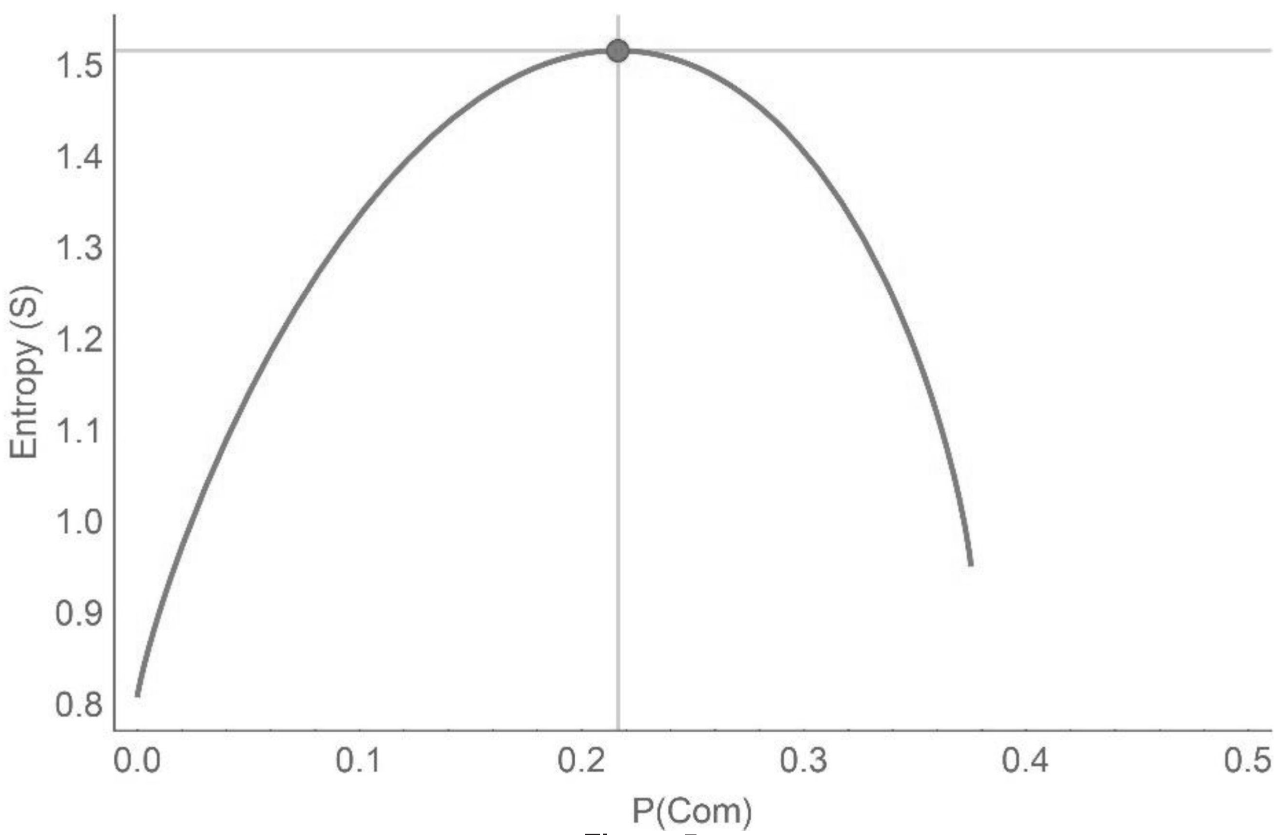

Figure 5

Taking the first term in Equation 5 as Residential and the second term as Industrial we have all three probabilities in Table 2 which we are pleased to note add to 1.

P(Com)	0.21624
P(Res)	0.46624
P(Ind)	0.317521

Table 2

Continuing along our developer's learning curve we now consider what happens when a competing political faction introduces a fourth land use, open space, essentially leaving the land undeveloped. This would take the entire land mass out of production. This political development is very expensive to combat. The developer's new condition is shown in Table 3.

Zoning	Cost	Acres	Approval Probability	Disapproval Probability
Res	$1000	15	0.5	0.5
Ind	$2000	25	0.8	0.2
Com	$3000	20	0.9	0.1
Open	$8000	25	0.5	0.5

Table 3

Surveying our new conditions and consulting our data we find that, in other communities under similar circumstances, the average cost for dealing with these four uses is $2,500. Note that the algebraic rearrangement that allowed Equation 5 fails when the number of constraints increase. The method of LaGrange multiplier is used for this optimization problem, producing the probabilities shown in Table 4.

P(Com)	0.354626
P(Res)	0.296438
P(Ind)	0.247798
P(Open)	0.101138

Table 4

Much has been written about the ability of human beings to make predictions, especially when their own lives are affected by the outcome.[17] Real estate investors are generally an optimistic group. Developers may be among the most optimistic. They should benefit from analysis that helps them predict probabilities to guide them to the right choice of land use. Having a mechanism that saves them from assuming they know more than they do can be a valuable tool. Bayesian reasoning optimized with Entropy is that mechanism.

VI. Conclusion

This struggle was not about pitting those comfortable with common fractions against those who prefer calculus. Major disagreement between the two camps turns on a more fundamental issue: subjectivity. Frequentists claim that the anonymity of large datasets offers objectivity and that the Bayesian posterior is heavily influenced by an inherently subjective prior. About this we have little argument. But what of it? Every set of facts will appear different to two different observers. Averaging out errors attendant to that reality is what the law of large numbers delivers, to be sure. One could equally claim that Frequentists are just a big crowd of aggregated Bayesians. George Box, reporting that all models are wrong but some are useful, commended us all to Occam's Razor in the interest of economy.[18] Gregory (2005) points out that Bayesian analysis has a built-in razor automatically penalizing complicated models, setting data complexity against model accuracy as a tradeoff.[19] Both the promise and the threat of cyberspace is that algorithms let the data choose the model.

The Rev. Bayes was interested in a theological question. It seemed to him that the probability of a deity was, as Frequentists would design it, a question with a black or white answer. Yet believers and atheists separately seemed to have the same amount of evidence.[20] Perhaps he noticed that there were fewer atheists in nursing homes than nurseries and concluded that people updated their priors as they grew closer to the time they needed the answer. Ours is not to sort out whether or how probability affects mens' souls. However, there are some investors who consider real estate a religion.[21] Those souls can make good use of Bayes' equation.

Real estate entrepreneurs resist standardization. Every property is unique, every building different, every owner exceptional. Averaging is anathema to the industry. Entropy is useful because measures the information that is lost by averaging.[22] The successful real estate professional knows he is in the information business and what he sells is not land and building, but clear thinking about the people and problems which exist upon the land. That clear thinking requires an algorithm such as the one we have discussed here.

What is called for is a manageable method for estimating probabilities and making predictions on the ground and in the near term. The Party Line and its progeny, portfolio theory, has failed real estate. What is needed is tools for better on-site management of an asset that requires hands-on effort. People need to get to work on time. The Bayesian techniques described here are suitable. Simple and intuitive methods are useful in the hands of real estate practitioners. With those tools, a little more each nanosecond, they join the new breed of data scientists awash in Big Data.

It's a convergence we can live with.

NOTES:

[1]On January 7, 2018 a Google search of the term "Frequentists vs. Bayesians" produced 28,900 results.

[2]This may be a daunting assumption for those whose training was not in the quantitative realm. The sentence containing it has keywords that make for useful online searching into this fascinating area of life.

[3]Bayes, Thomas; Price, Richard. (1763). "An Essay towards solving a Problem in the Doctrine of Chances." *Philosophical Transactions of the Royal Society of Lon-*

don. 53(0): 370-418.

[4]Stylized examples, simplified for expository purposes are vulnerable to criticism. At this point it is best not to question too closely how “event” is defined.

[5]For those precision artists who want to quibble that there are seconds involved in minutes and there are always increments of time around any specific instant, one may define a success as occurring within a window of say 12-16 minutes. However valid the approach, this is a different problem from the example we contemplate here.

[6]This is similar to how GPS works for those who drive cars with computer assistance.

[7]The term “likelihood” is technically different from “probability” despite the two words having very similar common meanings.

[8]For a more dramatic, and tragic, example see Nate Silver’s (2012) equivalent computation involving the probability that the planes that hit the World Trade Center on September 11, 2001 were a terrorist attack. On that day 2,977 innocent people experienced a blockage that ended their lives. The probability that the first plane was a terrorist attack was 38 percent but when the second plane hit that prior was revised to 99 percent.

[9]Care must be taken not to confuse this question with its converse: Given that the tenant lived in the larger (or smaller) building, what is the probability he defaulted. To contrast with a more extreme example, think about the difference between the probability that a person speaks English, given that he is reading this article (very high) vs. the probability that a person is reading this article given that he speaks English (very low).

[10]Beta is a two-parameter (a, b) model which is a common start for Bayesian analysis. It has some interesting and useful properties. An excellent discussion of this using baseball batting averages may be found at http://varianceexplained.org/statistics/beta_distribution_and_baseball/. Another common starting point is the Uniform distribution where every outcome is equi-probable. The Beta[1,1] distribution is identical to the Uniform distribution.

[11]This is even harder with land.

[12]Sivia (2006), referring to Skilling (1998), at page 112 provides an excellent discussion as to why ME is the appropriate objective function to optimize due to its lack of implied correlation.

[13]The units are “bits” due to using Log base 2.

[14]This is also known as The Sum Rule.

[15]Add the number of zeros appropriate for your local government then add nine more for California.

[16]In this illustration we are not using the data regarding past voting records. It is left as an exercise for the reader to design a method employing these data.

[17]Cf. the work of Kahneman and Tversky on Confirmation Bias and related Behavioral Economics subjects that lead to the 2002 Nobel Prize in Economics for Daniel Kahneman. The 2017 Nobel Prize went to Richard Thaler who wrote how limited rationality and lack of self-control affect economic decisions of human beings.

[18]Box, G. E. P., “Science and Statistics,” *Journal of the American Statistical Association,* 1976, 71: 791–799.

[19]Perhaps the most powerful commercial application of this is the symbolic regression offered in Data Modeler which may be found at www.evolved-analytics.com.

[20]All of it or none of it, depending on your viewpoint. It has been observed that atheism is a faith-based religion.

[21]Whether it is remains an open question. Consider all wars. They are nominally about religion but actually about real estate. The combatants may be talking about the Lord but they keep score in the land.

[22]Eigen and Winkler (1981) at p. 144.

REFERENCES:

Eigen, Manfred and Winkler, Ruthild, *Laws of the Game, How the Principles of Nature Govern Chance*, Princeton University Press, Princeton, NJ, 1981.
Gregory, Phil, *Bayesian Logical Data Analysis for the Physical Sciences*, CUP, Cambridge, 2005.
Gigerenzer, Gerd, *Calculated Risks*, Simon and Schuster, NY, NY 2002
Jaynes, Edwin T., *Probability Theory, the Logic of Science*, Cambridge UP, Cambridge, 2003.
Jeffreys, W. H. and Berger, "J. O. Ockham's razor and Bayesian analysis," *American Scientist*, Vol 80, pp. 64-72.
Harrington, Peter, *Machine Learning in Action*, Manning Publications Co. Shelter Island, NY, 2012.
Kahneman, Daniel, *Thinking Fast and Slow*, Farrar, Straus and Giroux, NY, NY, 2013.
Kosco, Bart, *Fuzzy Thinking, the New Science of Fuzzy Logic*, Hyperion, NY, 1993.
Kullback, Solomon, *Information Theory and Statistics*, Dover Publications, Inc. Mineola, NY, 1997.
Silver, Nate. *The Signal and the Noise: Why So Many Predictions Fail-but Some Don't*, Penguin Publishing Group, New York, NY, 2012.
Sivia, D. S. and Skilling, J., *Data Analysis, a Bayesian Tutorial*, Oxford UP, Oxford, 2006.
Skilling, J., *Massive Inference and Maximum Entropy.* In *Maximum Entropy and Bayesian Methods*, (ed. R. Fischer, R. Preuss and U. von Toussaint), Kluwer, Dordrecht. 1998.
Young, M. S., & Graff, R. A. (1995). "Real estate is not normal: A fresh look at real estate return distributions." *Journal of Real Estate Finance and Economics*, 10(3), 225-259 (1995).