Decoding Entropy

Mhaskar, Tejas

20 May 2013

Decoding Entropy

(A Credit Risk Modelling Perspective)

Tejas Mhaskar

"Entropy Is Simple - If We Avoid The Briar Patches!" - Frank L. Lambert

Abstract: Since its evolution, the concept of Entropy has been applied in various fields like Computer Science, Quantitative Finance, Physics etc. The definition of Entropy has slightly different meanings depending on the field of science to which it is being applied. This paper aims to examine the concept of Entropy and its application in Credit Risk Model Development and Validation.

Non-Technical Introduction

Entropy (Shannon's Entropy) is a probabilistic measure of uncertainty, whereas, Information is measure of reduction in that Entropy. Let us take an example of a "fair coin" to distinguish between these two terms. Before we flip the coin, we are uncertain about what will happen once it is flipped. After the coin is flipped, the uncertainty about the flip immediately drops to zero since we now know the response i.e. heads or tails. This in turn means that we have gained information.

Now, how does one measure the uncertainty about the response (heads or tails) before the coin is flipped? How does one quantify this uncertainty about the flip? This is where 'Entropy' comes into picture.

Technical Introduction

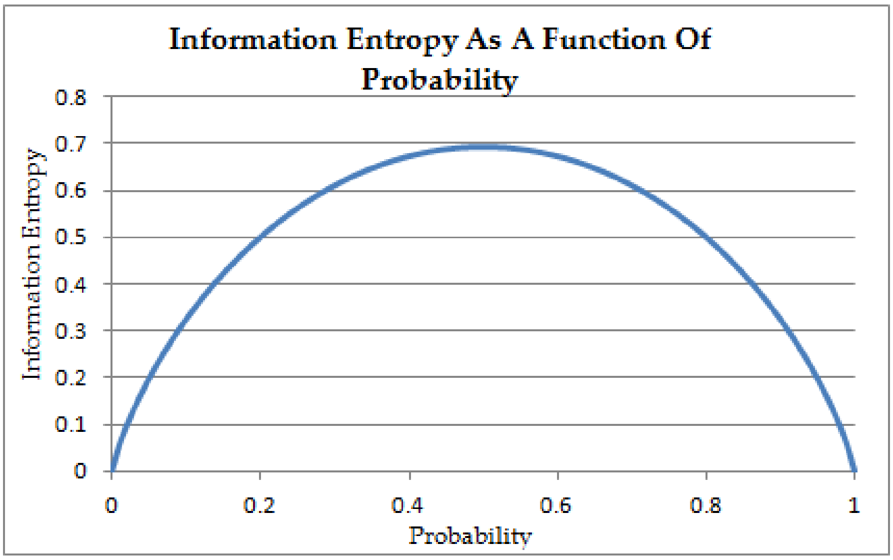
The concept of Thermodynamic Entropy (analogous to Information Entropy) was originated by Rudolph Clausius. Rudolph earlier coined it as "Equivalence-Value". To understand the true meaning of Entropy, we must first recall the "Second Law of Thermodynamics". The second law of thermodynamics helps us understand why the world works the way it does. It says that energy of all kinds spreads out if there is absence of any hindrance. It explains wearing out of car tyres, cooling down of hot pans etc. Thermodynamic Entropy quantitatively measures any such kind of a spontaneous process.

This concept was first applied in the field of Information Theory by Claude Shannon. From the perspective of Information Theory, Entropy is a measure of the uncertainty in a random variable. Information Entropy (also known as Shannon's Entropy) quantifies the expected value of information contained in a message. Therefore, Entropy is a measure of unpredictability. In order to understand this more clearly, let us take an example of a fair coin. When a fair coin is flipped, the probability of heads is the same as the probability of tails. Therefore, the Entropy of the coin flip is as high as it could be.

When taken from a finite sample, the entropy can explicitly be written as:

$$H(X) = \sum_i P(x_i)\,I(x_i) = -\sum_i P(x_i)\log_b P(x_i) = -\sum_i \frac{n_i}{N}\log_b \frac{n_i}{N} = \log_b N - \frac{1}{N}\sum_i n_i \log_b n_i,$$

The following chart plots the Entropy as a function of Probability:



Application of Entropy in Risk Model Validation

Since Information Entropy is a measure of unpredictability, it has become a common technique in Validation of Credit Risk Models which estimate the creditworthiness of borrowers by assigning a Probability of Default (PD) to them.

Let us try to understand this application using the previous "coin flipping" illustration. Let us take an example of a Rating Model which assigns a PD of 50% to all the borrowers (a Random model). As mentioned earlier in the "coin flipping" example, the probability of default is the same as the probability of survival in this case. Therefore, such a rating model will have the highest Entropy. It can be noticed from the chart that as the assigned PD moves away from 50% (in either direction), one becomes more certain about the credit event and hence the entropy reduces.

The Unconditional Entropy of any Rating Model is simply Information Entropy of the total PD of the population, whereas, the Conditional Entropy of the Rating Model is quantified using PD's assigned to each Rating Grade in the model. At any given point, the Conditional Entropy cannot exceed the Information Entropy. The difference between Unconditional Entropy and Conditional Entropy measures the decrease in uncertainty about the credit event caused by the introduction of Rating Grades in the model.

Following are the common Entropy measures used in Validation of Credit Rating Models:

- Kullback-Leibler Divergence (KL)
- Conditional Information Entropy Ratio (CIER)

KL measures the distance between Unconditional and Conditional Entropy.
CIER measures the ratio of distance between Unconditional and Conditional Entropy to Unconditional Entropy.

Application of Entropy in Retail Pooling under IRB Approach of Basel II Guidelines

The Internal Rating Based (IRB) approach under the Basel II Guidelines for Capital Computation requires banks to create homogenous pools for their Retail portfolios. The term 'homogeneity' is used in the context of risk characteristics or default behaviour. These pools then form the basis for estimating a PD for exposures in the pool.

The idea behind creating homogenous pools is that such a creation would ideally reduce the uncertainty prevailing in the overall retail portfolio. This boils down to quantifying the reduction in uncertainty caused by the introduction of homogenous pools that subdivide the overall portfolio. This is where Entropy based measures come into play. Conditional Information Entropy Ratio (CIER) can be used for evaluating the quality of the retail pools.

Retail Pools can also be created using Decision Tree analysis. Decision Trees make use of statistical methods to subdivide the data into different groups. Some of the common algorithms used in Decision Tree building are Classification And Regression Tree (CART), LearnDT, Chi-squared Automated Interaction Detector (CHAID) etc. LearnDT algorithm uses Information Gain (synonymous to Kullback-Leibler Divergence) as a splitting criterion for building a tree.

A decision tree can be constructed top-down using the Information Gain as follows:

- Begin at the root node
- Determine the variable with the highest information gain which is not already used as an ancestor node
- Add a child node for each possible value of that variable
- Attach all examples to the child node where the variable values of the examples are identical to the variable value attached to the node
- If all examples attached to the child node can be classified uniquely add that classification to that node and mark it as leaf node
- Go back to step two if there are unused variables left, otherwise add the classification of most of the examples attached to the child node

Alternative for Entropy based measures

An alternative measure for Information Gain in Decision Tree analysis is "Gini Impurity". Gini Impurity is used as a splitting criterion in Classification And Regression Trees (CART) algorithm mentioned above. Gini Index for a given node t is computed using the following formula:

$$\text{Gini}(t) = 1 - \sum_j [\, p(j|t) \,]^2$$

where $p(j|t)$ is the relative frequency of class j at node t

Gini Index is

- Maximum when records are equally distributed among all classes, implying least interesting information
- Minimum when all records belong to one class, implying most interesting information

Conclusion

This paper discussed the concept of Entropy and its use in Credit Risk Modelling. Entropy based measures are robust and easy to compute. They can be used as a measure of performance for Credit Rating Models. They can also be used for Retail Pooling under IRB approach for Basel II. Gini Impurity can be used as a substitute for Information Gain (which is an Entropy based measure) in the process of Decision Tree building.

Appendix 1: Calculating Entropy and Entropy Based Measures for Credit Rating Models

Unconditional and Conditional Entropy

Unconditional Entropy is calculated using the following formula

$$H_0 = - (PD * \log_2 (PD) + (1-PD) * \log_2 (1-PD))$$

For each rating class c, conditional entropy is H_c:

$$h_c = - (PDc * \log_2 (PD_c) + (1- PD_c) * \log_2 (1- PD_c))$$

Where PD_c is % of defaulters within rating class c.

Across all rating classes in a model, the conditional entropy H1 (averaged using the observed frequencies of the individual rating classes' p_c) is defined as:

$$H_1 = -\sum p_c * h_c$$

Kullback-Leibler Divergence (KL) and Conditional Information Entropy Ratio (CIER)

First, we calculate H0 (Unconditional Entropy) from the overall default probability. Then, we calculate hc for each rating grade to arrive at H1 (Conditional Entropy).

Kullback-Leibler Distance = H0 - H1

CIER = (H0 - H1)/ H0

References

[1] R. M. Gray, Entropy and Information Theory

[2] Ben-Naim, Arieh (2007). Entropy Demystified. World Scientific

[3] Callen, Herbert, B (2001). Thermodynamics and an Introduction to Thermostatistics

[4] Entropy for beginners – A wikibook

[5] Shannon, Claude E. (July/October 1948). "A Mathematical Theory of Communication".

[6] Borda, Monica (2011). Fundamentals in Information Theory and Coding

[7] G. Christodoulakis, S. Satchell, The Analytics of Risk Model Validation

[8] Rating models and validation, Oesterreichische Nationalbank (OeNB)

[9] Martin, Nathaniel F.G. & England, James W. (2011). Mathematical Theory of Entropy

[10] Working Paper 14 - Studies on the Validation of Internal Rating Systems, Basel Committee on Banking Supervision (BCBS)

[11] Mitchell, Tom M., Machine Learning

[12] R. J. Lewis, An Introduction to Classification And Regression Tree (CART) Analysis