Transfer Entropy

Applied in Financial market

Under Professor Trilochan Tripathi

19 December 2023

By Abhinav Sharma

Documentation

Table of content

Probability and Expected Value

The probability of different outcomes occurring in a system influences the entropy. If all outcomes are equally probable, the system has maximum uncertainty, indicating maximum entropy.

The expected value (mean) of a random variable is essential in entropy calculations. It represents the average information content or uncertainty associated with the variable.

The entropy formula involves the expected value of the logarithm of the probabilities. The expected value is a way of summarising the overall information content of a probability distribution.

A.] Unbiased Event (Coin Toss):

Probability:
In an unbiased coin toss, there are two possible outcomes – heads or tails. Since the coin is unbiased, each outcome has an equal probability of occurring. Therefore, the probability of getting heads or tails is both 0.5 or 50%.

Expected Value:
The expected value (or mean) of an unbiased coin toss is calculated by taking the sum of each possible outcome multiplied by its probability:

$E(X)=(0.5\times \text{Heads})+(0.5\times \text{Tails})$

Since the probabilities are equal, the expected value is the average of the possible outcomes, and in this case, it is 0.5.

B.] Biased Event (Stock Price Up-Down):

Probability:
In the case of a biased event like stock price movement, probabilities are not necessarily equal. Let's consider a simplified example where a stock can either go up (U) or down (D). The probabilities of each outcome may be represented as
$P(U)+P(D)=1$.

Expected Value:
The expected value for a biased event is calculated similarly to the unbiased event:

$$E(X)=(P(U)\times Up)+(P(D)\times Down)$$

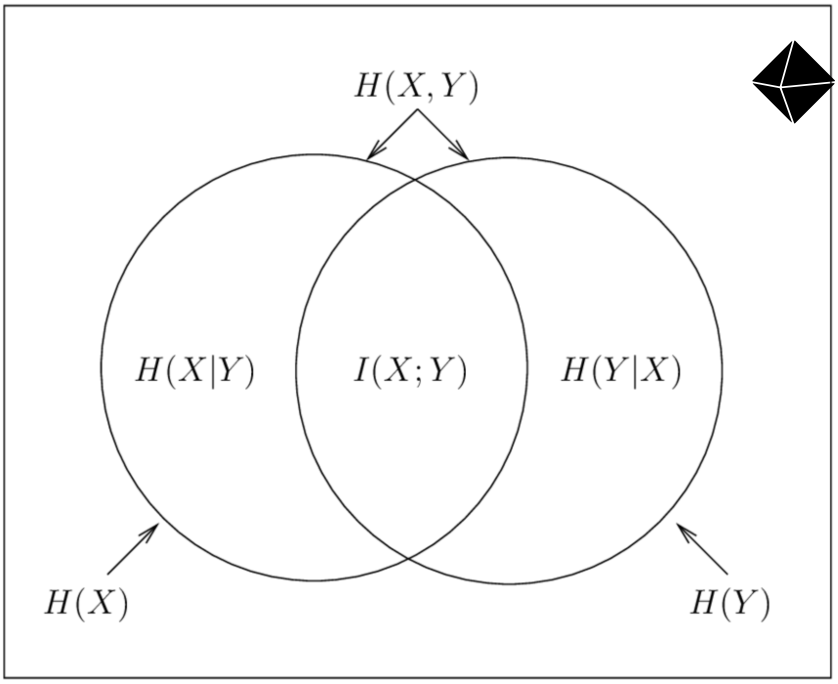
This formula takes into account the probabilities of each outcome and their associated values. The expected value provides an average value considering the likelihood of each outcome.

Code snippet_1.1

Below is link to Python program that prompts the user to input the possible outcomes, there associated values, and the probabilities for an event. The program then calculates and prints the expected value.

Documentation

Mutual Information and Entropy



$$H(X) = -\sum_x P(x) \cdot \log_2(P(x))$$

Where:

- $H(X)$ is the self entropy of random variable X,
- $P(x)$ is the probability of outcome x.

These are different type of Entropy Self, Joint and Conditional Entropy.

The formula for joint entropy $H(X,Y)$ of two random variables X and Y is given by:

$$H(X,Y) = -\sum_x \sum_y P(x,y) \cdot \log_2(P(x,y))$$

Where:

- $H(X,Y)$ is the joint entropy of random variables X and Y,
- $P(x,y)$ is the joint probability of $X = x$ and $Y = y$.

This formula measures the uncertainty or information content associated with the joint distribution of the two random variables. It quantifies the overall unpredictability when considering the combined outcomes of X and Y.

$$H(X|Y) = H(X,Y) - H(Y)$$

This formula indicates that the conditional entropy of X given Y is equal to the joint entropy of X and Y minus the entropy of Y. It represents the reduction in uncertainty about X when Y is known.

Documentation

Concept of Surpriseness and Randomness

Surprisal and the Concept of Randomness:

Surprisal:

Surprisal is a measure of how unexpected or surprising an event is. It is the negative logarithm of the probability of the event. Mathematically, it is represented as:

$$S(x) = -\log(P(x))$$

Where:

- $S(x)$ is the surprisal of event x,
- $P(x)$ is the probability of event x.

The higher the surprisal, the more surprising or unexpected the event is. If an event has a high probability (close to 1), the surprisal is low. Conversely, if the probability is low (close to 0), the surprisal is high.

Connection to Randomness:

The concept of surprisal is closely tied to the notion of randomness. In a truly random process, each outcome is equally likely, leading to a uniform distribution of probabilities. In such cases, the surprisal is maximized for rare events and minimized for common events.

When dealing with randomness, we often encounter events that are less predictable, leading to higher surprisal values. The logarithmic function is used to emphasize the exponential nature of the surprise associated with rare events. Taking the logarithm helps in scaling down the values and making the surprisal more interpretable.

The expected value of surprisal, also known as the expected information or entropy, is a measure of the average amount of surprise or uncertainty associated with a random variable. The mathematical formula for the expected value of surprisal (H), often denoted as $E[S]$, is given by:

$$H(X) = -\sum_i P(x_i) \cdot \log_2(P(x_i))$$

Here:

- $H(X)$ is the expected value of surprisal or entropy of the random variable X.
- $P(x_i)$ is the probability of the event x_i.
- The sum is taken over all possible values x_i of the random variable X.

This formula represents the average amount of information (in bits, when using base 2 logarithm) needed to specify the outcome of the random variable X. It quantifies the uncertainty associated with the random variable.

It's worth noting that this formula is consistent with the definition of entropy in information theory. The concept of entropy is closely related to the expected value of surprisal, and in many contexts, the terms are used interchangeably.

Entropy of Discrete and Continuous process

Discrete Random Variable:

For a discrete random variable X with a probability mass function $P(x)$, the surprisal for an event x_i is given by:

$$S(x_i) = -\log_2(P(x_i))$$

The expected value of surprisal (entropy) is then calculated as:

$$H(X) = -\sum_i P(x_i) \cdot \log_2(P(x_i))$$

Continuous Random Variable:

For a continuous random variable X with a probability density function $f(x)$, the surprisal for an event x is often replaced with differential entropy:

$$h(x) = -\log_2(f(x))$$

The expected value of the differential entropy is given by:

$$H(X) = -\int_{-\infty}^{\infty} f(x) \cdot \log_2(f(x))\, dx$$

It's important to note that the interpretation of differential entropy can be more subtle than that of entropy for discrete variables. Differential entropy doesn't have the same direct interpretation as the average amount of "surprisal" or "information" as in the discrete case. Additionally, differential entropy can be unbounded and may not have the same intuitive properties as entropy for discrete variables.

Entropy of complex continuous process

Estimating Differential Entropy through Histogram Analysis

In situations where the probability density function (PDF) of a continuous random variable is not analytically tractable, and only empirical data samples are available, the histogram approach serves as a practical method to estimate the differential entropy. This methodology proves particularly valuable in scenarios where direct access to the underlying PDF is challenging or unavailable.

Methodology:

1. **Data Collection:**
   - Obtain a representative sample of empirical data from the distribution under consideration.
2. **Histogram Construction:**
   - Divide the observed data range into discrete bins, forming a histogram. The height of each bin corresponds to the frequency of data points falling within that interval.
3. **PDF Estimation:**
   - Normalize the histogram such that the area beneath the distribution equals unity. This normalization transforms the histogram into an empirical estimate of the probability density function (PDF).
4. **Differential Entropy Estimation:**
   - Utilize the estimated PDF in the differential entropy formula, which is adapted to the discrete nature of the histogram:

$H(X) = -\int_{-\infty}^{\infty} f(x) \cdot \log_2(f(x))\, dx$

In the case of a histogram, this integral is replaced by a sum over the bins:

$H(X) \approx -\sum_i p_i \cdot \log_2(p_i)$

where p_i is the normalized height of the i-th bin.

5. **Negation:**
   - Introduce a negative sign to the computed result, aligning with the convention of differential entropy.

Considerations:

- The accuracy of the estimation depends on the size of the sample and the selection of bin sizes in the histogram. Finer bin resolutions generally lead to more precise estimates but may introduce increased variance.
- The histogram approach provides a pragmatic means of estimating differential entropy when analytical expressions for the underlying PDF are elusive or impractical, making it particularly applicable in real-world scenarios.

This approach offers a robust alternative for researchers dealing with complex distributions and emphasizes the adaptability of the histogram in approximating the PDF for subsequent differential entropy calculations. The methodology is inherently suited to scenarios where direct access to the mathematical formulation of the distribution is challenging or unattainable.

Code snippet_1.2

The Python code simulates a discrete financial process with random price movements, calculates the probabilities, and then computes the surprisal for each event and the self-entropy of the entire process. The results, including the simulated process, probabilities, surprisal values, and self-entropy, are then printed.

Documentation

Transfer Entropy

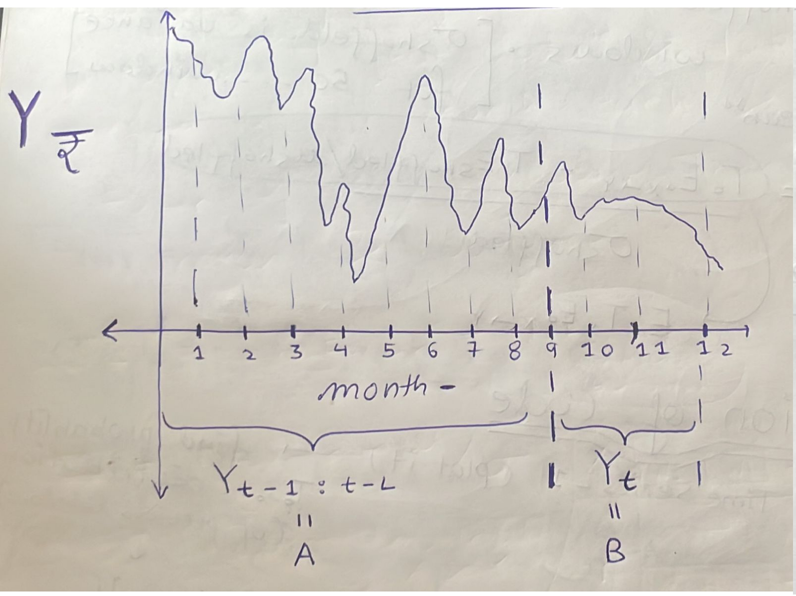
The formula for Transfer Entropy (TE) from variable X to variable Y is expressed as the conditional entropy of the future state of Y (Y_t) given its own past ($Y_{t-1:t-L}$) minus the conditional entropy of Y_t given both its past and the past of X ($X_{t-1:t-L}$):

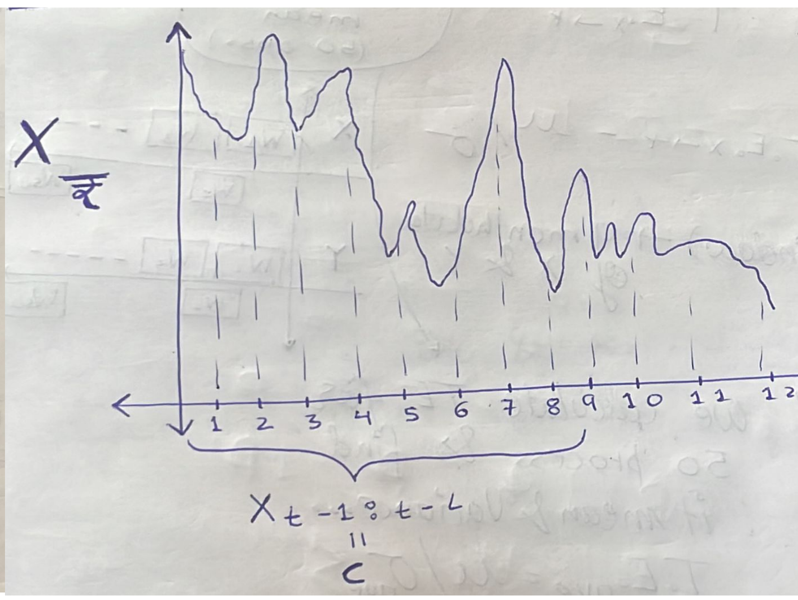
$$TE_{X \to Y} = H(Y_t|Y_{t-1:t-L}) - H(Y_t|Y_{t-1:t-L}, X_{t-1:t-L})$$

Here:

- $TE_{X \to Y}$ is the Transfer Entropy from X to Y,
- $H(Y_t|Y_{t-1:t-L})$ is the conditional entropy of the future state Y_t given its own past $Y_{t-1:t-L}$,
- $H(Y_t|Y_{t-1:t-L}, X_{t-1:t-L})$ is the conditional entropy of Y_t given both its past $Y_{t-1:t-L}$ and the past of X $X_{t-1:t-L}$,
- L represents the time lag or delay.

Transfer Entropy is a measure used in information theory to quantify the amount of information that the past of one variable (X) provides about the future of another variable (Y), beyond what is already known from the past of Y itself. It is commonly employed in the analysis of information flow or causality between time series data.





Documentation

Let's substitute Y_t with B, $Y_{t-1:t-L}$ with A, and $X_{t-1:t-L}$ with C in the Transfer Entropy formula and calculate the relevant entropies:

1. **Substitute Variables:**
   $TE_{X \to Y} = H(B|A) - H(B|A, C)$
2. **Calculate Entropies:**
   - $H(A)$: Conditional entropy of A
   - $H(B)$: Conditional entropy of B
   - $H(B|A)$: Conditional entropy of B given A
   - $H(B|A, C)$: Conditional entropy of B given both A and C
3. **Transfer Entropy:**
   - Calculate $TE_{X \to Y}$ using the formula.

These entropies can be calculated using probability distributions or data samples. If you have the joint probabilities or samples for A, B, and C, you can apply the entropy formulas accordingly.

You can express the conditional entropy $H(B|A)$ using the joint entropy $H(A, B)$ and the marginal entropy $H(A)$ with the formula:

$$H(B|A) = H(A, B) - H(A)$$

Here's how you can rewrite the conditional entropy:

$$H(B|A) = -\sum_{a,b} P(A = a, B = b) \cdot \log_2 \left(\frac{P(A=a,B=b)}{P(A=a)}\right) + \sum_a P(A = a) \cdot \log_2(P(A = a))$$

This expression simplifies to:

$$H(B|A) = H(A, B) - H(A)$$

This relation allows you to calculate the conditional entropy of B given A using the joint entropy of A, B and the marginal entropy of A.

The joint entropy $H(A, B)$ of two random variables A and B is given by the formula:

$$H(A, B) = -\sum_{a,b} P(A = a, B = b) \cdot \log_2(P(A = a, B = b))$$

Where:

- $H(A, B)$ is the joint entropy of random variables A and B,
- $P(A = a, B = b)$ is the joint probability of $A = a$ and $B = b$.

This formula measures the amount of uncertainty or information content associated with the joint distribution of A and B. It quantifies the overall unpredictability when considering the combined outcomes of both random variables.

1. **Entropy of A:**
   $$H(A) = -\sum_{a} P(A = a) \cdot \log_2(P(A = a))$$
2. **Entropy of B:**
   $$H(B) = -\sum_{b} P(B = b) \cdot \log_2(P(B = b))$$

The conditional entropy $H(B|(A, C))$ of variable B given the joint information of variables A and C can be expressed using the joint entropy $H(A, B, C)$ and the joint entropy $H(A, C)$ with the formula:

$$H(B|(A, C)) = H(A, B, C) - H(A, C)$$

This formula indicates that the conditional entropy of B given both A and C is equal to the joint entropy of A, B, C minus the joint entropy of A, C.

Documentation

Conclusion

If Transfer Entropy (TE) is greater than 0 ($TE > 0$) between two variables, say from variable X to variable Y (denoted as $TE_{X \to Y}$), it implies that there is an information flow from the past of X to the future of Y beyond what is already known from the past of Y itself.

In other words, a positive Transfer Entropy suggests that the knowledge of the past values of X provides additional information or helps in predicting the future values of Y more accurately than considering only the past values of Y.

This can be interpreted in terms of causality or information transfer between the two variables. A positive TE indicates that changes or patterns in X are related to or influence the future behavior of Y.

The formula for the estimated Transfer Entropy (ETE) can be expressed as the difference between the calculated Transfer Entropy (TE) and the mean of the Transfer Entropy obtained from multiple iterations:

$$ETE_{X \to Y} = TE_{X \to Y} - \mathrm{Mean}_{TE}$$

Where:

- $ETE_{X \to Y}$ is the estimated Transfer Entropy from X to Y,
- $TE_{X \to Y}$ is the Transfer Entropy calculated from the actual data,
- Mean_{TE} is the mean of the Transfer Entropy values obtained from multiple iterations.

This formula helps in assessing whether the observed Transfer Entropy is significantly different from what would be expected by chance, considering the variability obtained from repeated calculations. If $ETE_{X \to Y}$ is significantly greater than zero, it suggests that there is a non-random information transfer from X to Y.

Documentation

To estimate Transfer Entropy (TE) from data, especially in a time series context, you often perform the following steps:

1. **Collect Data:**
   - Gather time series data for variables X and Y.
2. **Discretize Data (if needed):**
   - If the data is continuous, you might need to discretize it into bins.
3. **Calculate Transfer Entropy Multiple Times:**
   - For each iteration:
     - Randomly shuffle the time points.
     - Divide the data into past (A), present (B), and future (C) segments.
     - Calculate $TE_{X \to Y}$ using the formula: $TE_{X \to Y} = H(B|A) - H(B|A, C)$.
4. **Calculate Mean and Variance:**
   - Compute the mean and variance of the TE values obtained from multiple iterations.
5. **Statistical Analysis:**
   - Analyze the mean and variance to understand the central tendency and variability of the TE estimates.
6. **Interpret Results:**
   - If the mean TE is significantly different from zero and the variance is within an acceptable range, it suggests the presence of information transfer.

The idea is to perform this process multiple times to account for variations in the data and obtain a more robust estimate of Transfer Entropy. Subtracting the mean from the $TE_{X \to Y}$ calculated from your actual data helps in assessing whether the observed TE is statistically significant.

Joint Entropy

Let's assume the bins for X are $\{a, b, c, d\}$ with corresponding probabilities $\{P(a), P(b), P(c), P(d)\}$, and for Y, the bins are $\{e, f, g, h\}$ with probabilities $\{P(e), P(f), P(g), P(h)\}$.

The joint probability $P(X = x_i, Y = y_j)$, where x_i is the i-th bin of X and y_j is the j-th bin of Y, is given by the product of the probabilities of the individual bins:

$$P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$$

In your example, if X has bins $\{a, b, c, d\}$ and Y has bins $\{e, f, g, h\}$, then the joint probability matrix $P(X, Y)$ would be:

$$\begin{array}{ccccc} & e & f & g & h \\ a & a \cdot e & a \cdot f & a \cdot g & a \cdot h \\ b & b \cdot e & b \cdot f & b \cdot g & b \cdot h \\ c & c \cdot e & c \cdot f & c \cdot g & c \cdot h \\ d & d \cdot e & d \cdot f & d \cdot g & d \cdot h \end{array}$$

Here, each cell represents the joint probability of the corresponding combination of bins from X and Y.

Mathematically, this normalization condition is expressed as:

$$\sum_i P(X = x_i) = 1$$
$$\sum_j P(Y = y_j) = 1$$

If both of these conditions are satisfied, then the sum of all the products in the joint probability matrix:

$$\sum_i \sum_j P(X = x_i, Y = y_j) = 1$$

This ensures that the joint probability distribution is properly normalized, and the total probability across all possible outcomes of X and Y is 1.