**Surprise Presentation**

From the paper I’ve attached to this lecture, which can serve as a motivating example for why we’d like to measure surprise, this presentation serves as a way to inform everyone about one way to quantify surprise that I found interesting.

**Important Facts:**

There are a few important properties about surprise that are worth mentioning – namely:

* Currently there is no widely accepted mathematical theory to quantify surprise.
* The model I’m about to present is the closest thing to an existing model.

**Definitions:**

With that being said, here’s my definition I’d like to use as a way to think of surprise:

**Surprise:** Quantifies how data effects observers by measuring the difference between the beliefs before and after an event takes place.

Before we get to forming a mathematical model to quantify this, two things need to be noted:

* Surprise exists only in the presence of uncertainty
* Surprise can only be quantified relative to the expectations of a behavior

**A Mathematical Model**:

Since we need to factor in beliefs both before and after an event occurs, we’re going to work with the Bayesian model for probability.

We can then use Bayes’ theorem to relate these two ideas.

For starters, we can define the background knowledge of an observer to be:

β = {P(M)} for M ∈ 𝝡, for 𝝡 being the set of possible outcomes

We can then measure the effect of our actual data based on our model space M and our observed data D.

We can get these values by applying Bayes’ Theorem:

ⱯM ∈ 𝝡, P(M | D) =

Where:

* P(M) represents our prior belief on what will happen
* P(D) represents the probability that our outcome actually occurs without surprise
* P(D | M) represents the probability that our outcome occurs given out expectations P(M)

From here, we can call our Surprise function, which will take in a data set 𝝡 and our outcome D.

S(D, 𝝡) = KL(P(M | D), P(M)) =

\*Note that the log is log2\*

**Kullback-Liebler Divergence**

In order to transform our surprise equation to an integral, we have to use something called Kullback-Liebler Divergence. I won’t go into the details in the notes, but below is a quick run down on how this concept can be applied.

We first have to define an important term to apply to our work:

**Entropy:** The way of measuring randomness in a variable. The equation for entropy is:

\*Entropy is normally measured in bits\*

We can then find KL(p | q) by finding H(p, q) – H(p). Since we are dealing with a sum of elements in a set, this becomes an integral and eventually reduces to our surprise equation that we derived above!

If you’d like to learn more about Surprise in mathematics, feel free to visit:

[Formal Bayesian Theory of Surprise Home Page (usc.edu)](http://ilab.usc.edu/surprise/#:~:text=Therefore%20we%20formally%20measure%20surprise%20elicited%20by%20quantifying,defined%20by%20the%20average%20of%20the%20log-odd%20ratio%3A)

If you’d like to learn more about Kullback-Liebler Divergence, feel free to visit:

[Kullback-Leibler Divergence - GeeksforGeeks](https://www.geeksforgeeks.org/kullback-leibler-divergence/)

**Lecture notes by Andrew Meyer**