SCIENTIFIC MACHINE LEARNING

Mark Temple-Raston

Decision Machine, LLC, New York, USA

ABSTRACT

Scientific Machine Learning implements the science-of-counting to analytically process any time-series and produce a complete set of thermodynamic measurements that define the state of the system. The scientific measurements are illustrated with a time-series of closing-prices for a stock (GE). Exact scientific measurements from Scientific Machine Learning (SML) are then used to create time-series decision services without model or bias. Services are built for a large class of Open allocation problems and applied to real optimal sales data for a consumer product good.

KEYWORDS

Machine Learning, Risk Analysis, Non-Equilibrium Thermodynamics, Information Theory, Decision Theory

1. INTRODUCTION

Emerging in the first half of the 19th century, the science-of-counting generated new and fruitful scientific disciplines, for example, thermodynamics, the theory of electricity, and physical chemistry, that generated many successful practical applications during the industrial revolution. Unlike mechanics (constant total energy and no external interactions), the science-of-counting applies to all time-series, mechanical or otherwise, that count states, events, or a unit of measure, counting natural number multiples (1, 2, 3, ...).

The science-of-counting (SoC) is founded on the principle of maximum information entropy, a method of logical inference used to define both science and machine learning, through enforced constraints on information [1]. The simplest, unsupervised, non-trivial and solvable machine learning is derived from the science-of-counting [2]. Scientific Machine Learning (SML) ingests time-series and returns scientific measurements. To make the instrumentation available to all, a web-based service is built to provide unbiased, model-free scientific measurements for any input time-series subject to modest data quality requirements.

Examples of scientific measurements of a time-series are presented below for six months of closing price data for General Electric (GE) stock, consisting of timestamps for the trading day and the closing prices (adjusted). The time-series input is plotted in Figure 1. The arithmetic mean (average) of the closing prices is superimposed (dash red curve).

David C. Wyld et al. (Eds): NLPCL, BIGML, AISC – 2025
pp. 65-71, 2025. CS & IT - CSCP 2025 DOI: 10.5121/csit.2025.151807

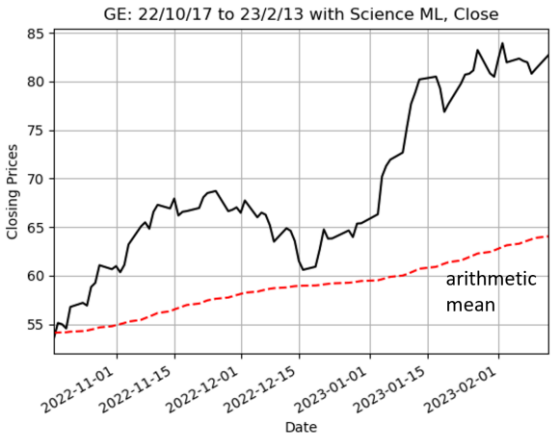


Figure 1. Closing Prices for General Electric (GE) with arithmetic mean.

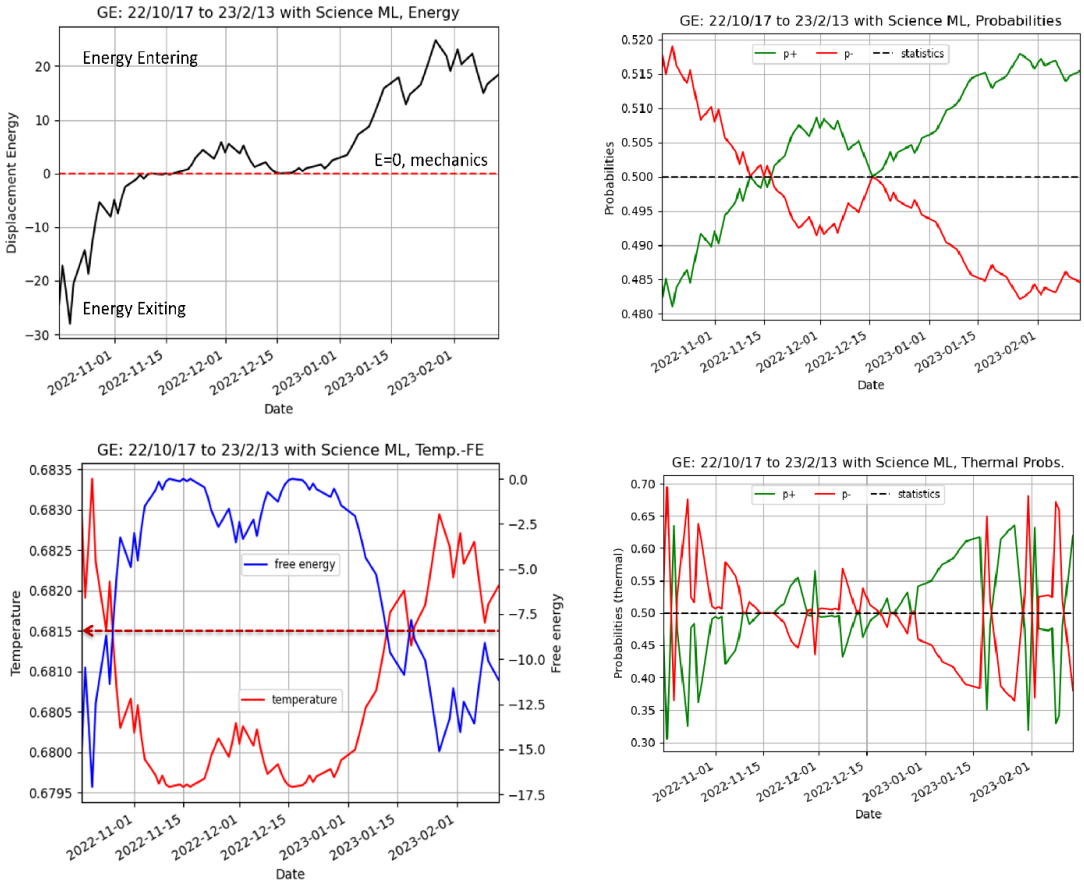


Figure 2. SML Measurements for GE (Binary).Presents time-series measurements for energies, probabilities and temperature. If the time-series behaved mechanically or statistically, the plot of the energy would coincide or lay close to the red dotted line. Take away: the time-series of closing prices does not behave statistically.

Several scientific measurements are plotted (Figure 2) from GE's historical closing prices (Figure 1). The measurement service for binary states counts the number of up or down states based on closing prices. Figure 2a plots the displacement energy, the total energy minus the mechanical energy of the two-state system, equivalently, the energy entering (positive) or exiting (negative) the system. In practice, it is observed that real-world time-series deviate significantly from mechanics and statistical equilibrium, and that systems routinely store and release energy (dissipation of entropic heat). When the displacement energy vanishes, E=0, the system reduces to mechanics and the binary states behave statistically, $p_+=p_-=1/2$. Figure 2b plots the probabilities, p_+ (green) and p_- (red), that GE stock will go up on the next trading day (green), or, close down on the next trading day (red), respectively. Figure 2c presents a double-sided plot with the time-series measurements for the (entropic) temperature and free-energy (Helmholtz), the energy available to do price movement work. As the free energy is observed doing price movement work, that is, decreases, the temperature of the system increases. Observe the dissipative structure in the time-series in Figure 2c. When the internal temperature equals the body temperature, the system is in thermal equilibrium, and the measurement of body temperature is possible. Finally in Figure 2d, we plot the elevated thermal probabilities for the two-state system for closing prices, calculated from non-equilibrium thermodynamics.

Counting is an independent, self-sufficient, and exact human act that is pure experience; Counting precedes the introduction of subject and object, or, of judgment, because counting is independent of subject. For the same reason, Scientific ML measures time-series before models or bias can be introduced. Note that counting groups of people, places or things, can include concepts, ideas or qualities that exist as thought or feeling rather than as tangible object.

The next section examines a large class of practical decision problems: Open System Allocations (OSA). OSA generalizes multi-armed bandits to open environments with time-series. Using SML, the timeseries determines an exact probability distribution function (pdf) for each “arm". SML is also applied to sales time-series for a Consumer Product Good (CPG) to calculate optimal energy allocations at each timestamp to get the best sales performance for the least energy/effort.

2. OPEN SYSTEM ALLOCATIONS

Assume resources for allocation to multiple decision options are made available, but with uncertain results. With experience allocating resources and monitoring the results, we presume that our decisions and future results will both see improvement. Recognizing that the environment is Open, so that environmental conditions change as energy enters or exits the system, SML produces the scientific measurement that supports Open analysis. We define a very large class of allocation problems: Open System Allocations (OSA). In the special and somewhat unrealistic case where probabilities and probability distributions for each decision option are constant in time, then Open System Allocations reduce to Multi-Armed Bandits (MAB). Many MAB applications are in use: online advertising (which ad to display), clinical trials (allocating patients to different treatment arms), recommender systems (product or service suggestions), portfolio management (adjusting investments based on reward probabilities), adaptive routing (best traffic route), product development (features selection), and so on. Note that all these applications assume that the environment is effectively static on the timescales considered critical for the problem. But, evidently, none of these applications are in truth static, and the MAB models must be replaced when environmental conditions change appreciably.

2.1. Decision Machine

We have seen that SML provides exact scientific measurements that quantify our current position or state in the environment. In the context of a Decision Machine, SML measurements provide the means to adaptive allocations as the environment changes (see Figure 4). One of SML's scientific measurements is the exact probability distribution function (pdf) for a time-series, the sampling distribution. Two examples from SML appear in Figure 3: the pdf for GE's closing prices (Figure 1); and the pdf for CPG Sales (Figure 5a, details in the next section).

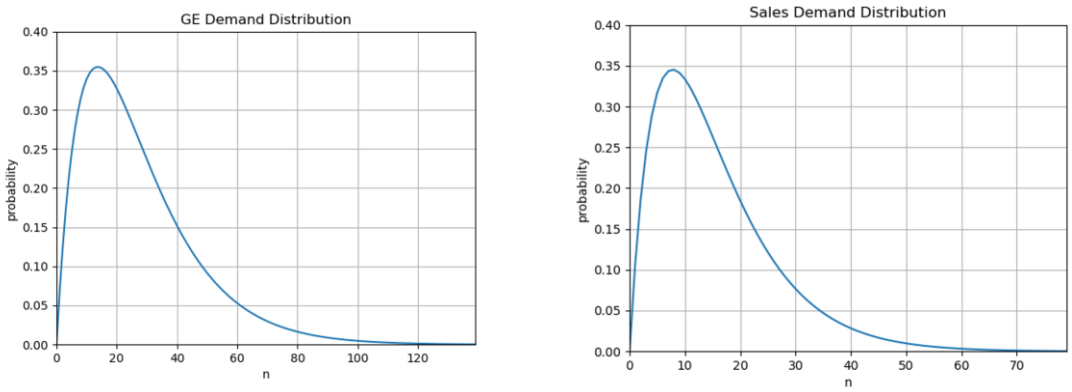


Figure 3. Exact PDFs for GE and Sales Time-Series (Units) derived from time-series. The left pane is the price displacement pdf from the average for GE closing prices; the right pane is the sales displacement pdf from the average for the Consumer Product Good in section 2.2.

We define an N-armed Open bandit to be a set of N one-armed Open bandits. An Open one-armed bandit is a slot machine that plays differently based on the environment. Each arm, indexed by n<N, has a finite number of (mutually exclusive) winning states, S_n, with returns R^n=(R_1,...,R_{S_n}). There are also a finite number of losing states, L_n, that return nothing. SML open bandits calculate the exact probabilities and probability distributions, given the time-series. The PDF is observed to change with timestamp.

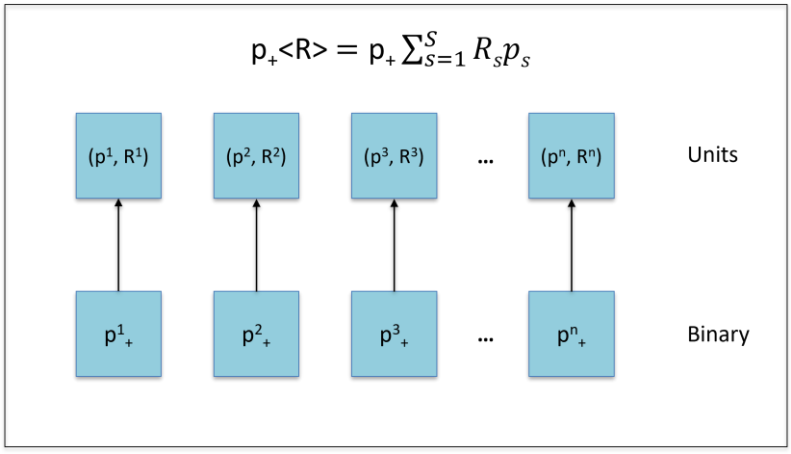


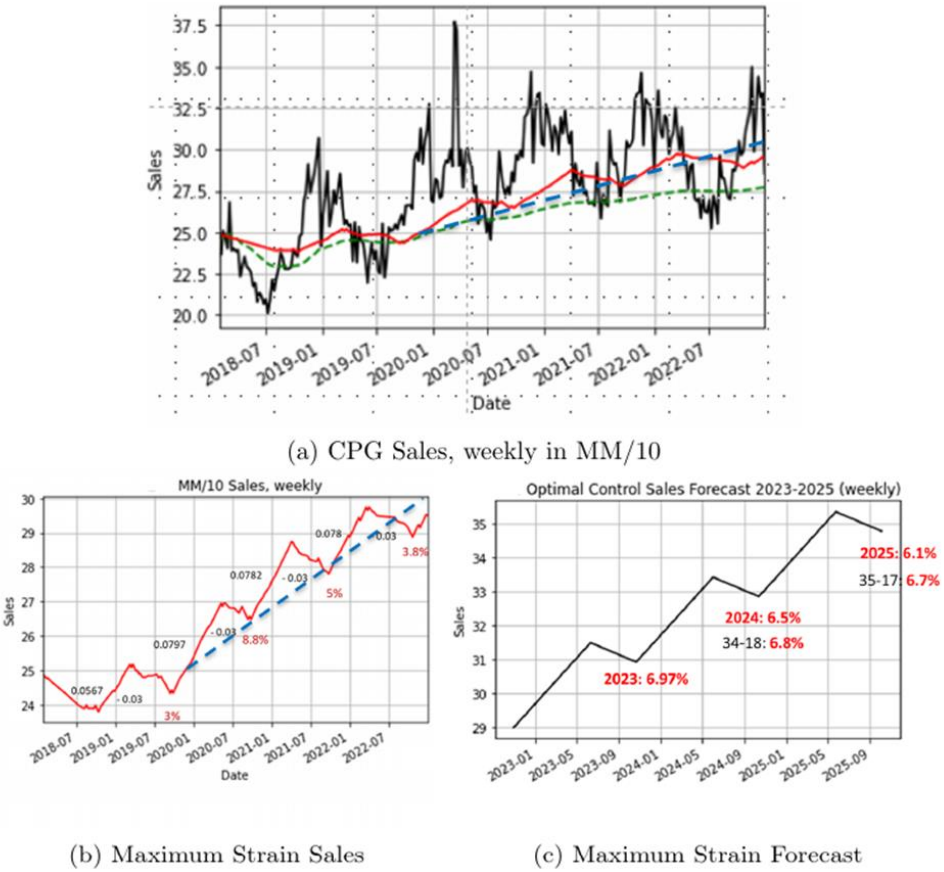
Figure 4. Decision Machine (Binary and Units)

A Decision Machine for Open System Allocations is structured in Figure 4, where the n^{th}-arm probability of being rewarded anything, p+^n, is deduced from the binary state analysis of the

time-series. The calculation of the expected return uses multiples of the unit of measure as states in the top row. In risk applications, the bottom row is the PD (probability of default) and the top row is the EAD (exposure at default). With the PDFs for each one-armed Open bandit, the best analytical decision is the one-armed Open bandit with the largest or least top row value in Figure 4, calculated exactly using SML.

2.2. Cultivation

For a Consumer Product Good (CPG), Figure 5a plots the weekly consumer sales figures (black curve) and the running arithmetic mean (green, dashed) from 2016 to 2023 (in units of MM/10).



(a) CPG Sales, weekly in MM/10

(b) Maximum Strain Sales

(c) Maximum Strain Forecast

Figure 5 : Max Improvement

When the sales time-series is processed by SML, two scientific quantities are evaluated exactly: the expected value <n>, and the expected strain, -λ<ε>. However, when energy is either entering or exiting the ecosystem (E≠0), the natural coordinates that emerge from the science-of-counting are not the expected value, <n>, or the expected strain, -λ<ε>, but the expected demand, $<η>$, and the expected supply, $<ξ>$, defined by the Gibbs matrix equation:

$$\begin{bmatrix} < \eta > \\ < \xi > \end{bmatrix} = \begin{bmatrix} 1 & E \\ v/p & 1 \end{bmatrix} \begin{bmatrix} < n > \\ -\lambda < \epsilon > \end{bmatrix}. \quad (1)$$

The optimal path for increasing expected demand is calculated explicitly from the eigenvalues and eigenvectors, e1 and e2, of the Gibbs matrix equation (1). The Gibbs matrix acting on blue eigenvector e1 can only scale without altering the direction---this corresponds to the blue path in Figure 5a and 5b. In Figure 6, we summarize the coordinate systems that are defined by SML and from this calculate the angle θ that defines thermal equilibrium and the optimal path decision.

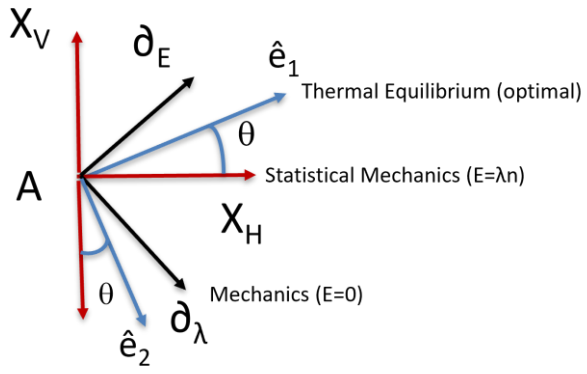


Figure 6. Optimal Path Decision Machine. Three different coordinate systems are presented: geometric, Gibbs, and eigenaxes (optimal).

Efficiency at this level is likely beyond our operational reach. But there is a lot of beneficial middle ground between what we currently do and what is most efficient. The red paths in Figure 5 begin the cultivation of the middle ground.

The red sales path in Figure 5a is the path of maximum strain/supply due to an increase in displacement energy. The linearity can be understood geometrically and the critical angle calculated. More difficult to control are sales paths steeper than the critical angle, because the energy can be unpredictably dissipated. The piece-wise linearity in the red path reflects seasonality in sales: there is a high season and low season, with durations defined by sales operations (in weeks that add up to 52 weeks). The slopes of the triangular demand curve are determined empirically, from the historical time-series (see figure 5b). Using the slopes calculate for 2020 for future returns, year-over-year percentage increases in sales for the optimized paths are calculated (see figure 5c): 2023: 6.97\%, 2024: 6.8\%, 2025: 6.7\%.

3. CONCLUSIONS

Scientific ML (SML) is built on the science-of-counting to produce exact scientific measurements of a thermodynamic nature from any time-series, subject to modest data quality requirements. SML realizes both scientific deduction (what is necessarily true) and scientific induction (hypothesis testing, that which is most likely true). SML is implemented as cloud services to provide unbiased, model-free instruments of scientific measure for any input time-series to a broad audience of users.

Thermodynamic metaphors in finance and sales can be helpful to understand and manage business operations. However, assumptions about the operating environment that are false distort and undermined efforts to manage. For example, metaphors that leverage the Ideal Gas Law (PV = nRT) applied to markets implicitly assume that there are no interactions between market participants. With SML measurements available, thermodynamic metaphors that are insightful

may be redesigned and structured so that a metaphor might be promoted to scientific decision framework.

SML is the foundational technology for the manufacture of Decision Machines, that provide the best analytical support possible for human decisions based on time-series. In the previous section we constructed three Decision Machines for Open systems: the best allocation decision among a set of decisions starting with no initial information; and two optimal decisions, best result for minimum effort, and best result (maximum strain) without overheating.

Many sources of time-series remain to be investigated.

REFERENCES

[1] Jaynes, E.T. (2003) "Probability Theory: The Logic of Science", Cambridge University Press.
[2] Temple-Raston, M (2025) "Learning in the Science-of-Counting", https://www.academia.edu/129814915/Learning_in_the_Science_of_Counting

AUTHOR

Mark Temple-Raston is founder, CIO and chief data scientist of Decision Machine, a state-of-the-art alternative data and self-service machine learning company that produces unbiased scientific measurements from time-series. He also serves as CIO and chief data scientist for BRANDthro, a neuromarketing consultancy that uses emotion AI and neuroscience for enhanced market understanding. Mark has a doctorate from Cambridge University, England, in Theoretical Physics.