Towards a Frequentist Interpretation of Sets of Measures

Pablo I. Fierens and Terrence L. Fine *
School of Electrical Engineering
Rhodes Hall 388
Cornell University
Ithaca, NY 14853

Abstract

We explore an objective, frequentist-related interpretation for a set of measures $\mathcal{M}$ such as would determine upper and lower envelopes; $\mathcal{M}$ also specifies the classical frequentist concept of a compound hypothesis. However, in contrast to the compound hypothesis case, in which there is a true measure $\mu_{\theta_0} \in \mathcal{M}$ that is assumed either unknown or random selected, we do not believe that any single measure is the true description for the random phenomena in question. Rather, it is the whole set $\mathcal{M}$, itself, that is the appropriate imprecise probabilistic description. Envelope models have hitherto been used almost exclusively in subjective settings to model the uncertainty or strength of belief of individuals or groups. Our interest in these imprecise probability representations is as mathematical models for those objective frequentist phenomena of engineering and scientific significance where what is known may be substantial, but relative frequencies, nonetheless, lack (statistical) stability.

A full probabilistic methodology needs not only an appropriate mathematical probability concept, enriched by such notions as expectation and conditioning, but also an interpretive component to identify data that is typical of the model and an estimation component to enable inference to the model from data and background knowledge. Our starting point is this first task of determining typicality. Kolmogorov complexity is used as the key non-probabilistic idea to enable us to create simulation data from an envelope model in an attempt to identify "typical" sequences. First steps in finite sequence frequentist modeling will also be taken towards inference of the set $\mathcal{M}$ from finite frequentist data and then applied to data on vowel production from an Internet message source.

Keywords. lower envelopes, frequentist interpretation, simulation.

* corresponding author, research of both authors partially supported by AFOSR Grant F30602-00-2-0558

1 Motivating Viewpoint

A sound enterprise of imprecise probability absolutely requires an interpretation. Jacob Bernoulli's years-long struggle to interpret probability and Bruno de Finetti's struggles to develop and gain acceptance for a subjective interpretation both attest to the importance and difficulty of establishing an interpretation. Focus on mathematical formal development, while of value in providing deep enlightenment, also can seduce and distract us—difficult as mathematics can be, it poses more familiar challenges than those of probabilistic reasoning and interpretation.

A paradigm of new departures in, say, applied mathematics is to find an anomalous physical phenomenon, study it, and then model it by a new mathematical approach and/or concept. However, in fact, much of our ability to recognize anomalous phenomena is conditioned by our mathematical constructs. In some sense we are only likely to find what we expect to see. A telling instance of this is provided by the 20th Century history of the count of human chromosome pairs— an insistence on their being 24 pairs long after there were photographs and experiments clearly showing only 23 pairs (see Ridley [9]). New concepts of probability are needed to open our intuition to new perceptions of phenomena.

Where there is merit in subjective interpretations, in concepts of degrees of belief, there is a physical reality independent of our individual thoughts. We deny the everyday practices of science and engineering at our peril. There is logical room for intrinsically imprecise physical phenomena, imprecise in that the underlying empirical relational systems are not homomorphic to mathematical relations on the real numbers (see Krantz et al. [6] for this viewpoint). The notion of physical phenomena sharing the precision of the real number system is so well-entrenched that identifications of imprecise phenomena are difficult to make and hard to defend.

Our strategy has been to work backwards. First, we introduce a new way of "seeing" by adopting description by the mathematical concept of envelopes familiar to many. Second, we ask what kind of mathematical data (not necessarily generated by any known physical phenomenon) could lead us to model it by envelopes? How could envelopes display themselves in the data, under ideal circumstances. Third, we attempt to close the circle of probabilistic modeling by proposing inference from such data sets to envelope models. We then are faced with self-consistency questions that are analogous to the laws of large numbers and goodness-of-fit tests. Our understanding of these issues is in its infancy. Fourth, and finally, we study the empirical phenomenon of vowel generation to see if it provides a real-world example of an imprecise probabilistic phenomenon.

2 Sets of Measures and Envelopes

Our long-term interest in objective frequentist interpretations for upper and lower probabilities has led us, through a series of papers, to conclude (e.g., see Papamarcou and Fine [8], Sadrolhefazi and Fine [10]) that such models could only be stationary if we used undominated lower probabilities. Furthermore, this prior research was concerned primarily with asymptotic considerations (but see [10]) in an attempt to isolate behavior that could not be encompassed by standard numerical probability. The somewhat pathological nature of undominated lower probability has led us to consider the more tractable concept of upper and lower envelopes. Prior frequentist attempts to deal with upper and lower envelopes include Cozman and Chrisman [2] and Walley and Fine [12], although these kindred researches were focused on asymptotic considerations. Our approach relates more closely to that of Walley and Fine in that we focus on the sequence of relative frequencies and not, as do Cozman and Chrisman, on the data sequence that gave rise to the relative frequencies. There is some justification in doing the latter, but it is not the direction we follow.

Walley [13], Theorem 3.3.3, informs us that a lower prevision $\underline{P}$ is coherent if and only if it is the lower envelope of a set $\mathcal{M}$ of finitely additive measures. Theorem 3.6.1 further asserts that $\mathcal{M}$ can be taken to be convex, although Walley [13], p. 505, asserts that "The properties of compactness and convexity of $\mathcal{M}$ do not appear to have any behavioural significance." Nonetheless, we are encouraged by Theorem 3.6.1 to focus on convex $\mathcal{M}$. This condition on coherent $\underline{P}$ can also be recast in terms of three simple analytical conditions (Walley [13], Theorem 2.5.5) previsions (expectations):

$$\underline{P}X \geq \inf_{\omega} X(\omega);$$

$$\lambda > 0 \Rightarrow \underline{P}\lambda X = \lambda \underline{P}X;$$

$$\underline{P}(X+Y) \geq \underline{P}X + \underline{P}Y.$$

Coherent upper previsions $\bar{P}$ are determined by

$$\bar{P}X = -\underline{P}(-X).$$

Characterizations of imprecise knowledge in terms of a set of measures $\mathcal{M}$ is also a mainstay of classical, frequentist statistics in which an "unknown" parameter $\theta \in \Theta$ indexes a set of measures $\mathcal{M} = \{\mu_\theta : \theta \in \Theta\}$. Curiously, frequentist statisticians provide no frequentist interpretation for their concept of the unknown parameter. They do, of course, assume that there is a true parameter value θ_0 and the measure μ_{θ_0} governs the observation X. Bayesians grandly dispense with ignorance by requiring a precise prior π on Θ.

Our goal is to provide an objective frequentist understanding of $\mathcal{M}$ developed in the fundamental case of a random variable X taking on only finitely many values. We can either choose to elaborate a view based upon a single time series realization or upon an ensemble of such time series. Our non-exclusive focus is on the case of relative frequencies calculated along a single time series $\{X_i, i = 1 : n\}$ of finite length n. We share with Kolmogorov [5] the belief that

> (1) The frequency concept based on the notion of *limiting frequency* as the number of trials increases to infinity, does not contribute anything to substantiate the applicability of the results of probability theory to real practical problems where we have always to deal with a finite number of trials.

More needs to be said, however, about what we mean by "finite". First of all, our sequences cannot be "short". If they are not long enough, then, as we shall see from the considerations of Section 4, there will not be an arithmetic opportunity for the generation of the necessary variety of relative frequencies. Furthermore, the concept of Kolmogorov complexity, reviewed in Section 6, is only meaningful for sequences that are sufficiently long compared to certain constants (e.g., $g_{\mathcal{S},\Psi}$) that are implicit in the defining choice of universal Turing machine. There is also an issue with the sequences being "very long" or with the perspective of the "long-run". Our investigation is based upon a premise of instability or variety in the world. The exceeding complexity of the world interacting with the data sources we propose to model (e.g., natural language sources of text or speech, file sizes transferred over the Internet) is one reason for there to be unpredictable temporal variations even over fairly long time periods. However, over very long time periods, one can expect aging effects and perhaps other phenomena that change or destroy the data source under consideration. This effect, while well-known, is conveniently overlooked in the fiction of the "long-run" essential to the frequentist views of conventional probability. Having said this, we can be somewhat quantitative about what we mean by "short" but not about the real-world dependent meaning of "very long".

3 Desiderata for Typical Sequences

We seek an operational interpretation of the family $\mathcal{M}$ in which there is no true measure governing outcomes. Put otherwise, what objective, frequentist data is typical or representative of such a specification of imprecise probability? We restrict ourselves to sequences drawn from a finite alphabet corresponding to a finite sample space. The essence of the problem appears most clearly in the finite alphabet case and its exploration is freed from complicating technicalities and vitiating idealizations. If the data is to describe $\mathcal{M}$ then relative frequencies calculated along the sequence (we are not taking an ensemble view) should come close to each of the measures in $\mathcal{M}$. Each measure in $\mathcal{M}$ should be approximated in a persistent fashion rather than transiently. If we allowed an infinitely long sequence, then this notion of persistence would be well-captured by recurrence infinitely often, and we would not need to commit to a specification of the rate of occurrence (e.g., see Walley and Fine [12]). However, we agree with Kolmogorov that our earthly concerns are with finite sequences. It does us no good in practice to have an essentially asymptotic interpretation that cannot be made sense of in the finite case.

Finally, a burden of foundations of probability is that numerical probability and an objective interpretation that is frequentist is so well-embedded that no alternative can compete on an equal footing. An alternative probability model has a chance of acceptance only if any competing standard probability model is so complex that it is cut off by Ockam's Razor. This is particularly challenging in the setting of finite sequences of lengths that are "long" but not "very long".

Hence, given a family of measures $\mathcal{M}$, we seek a construction for generating:

- (necessarily) long **finite** sequences,
- that exhibit observably **persistent oscillations** of relative frequencies,
- **closely visiting given measures and no others**,
- with such oscillations occurring in a highly **complex pattern** of transition times,
- that contraindicates an underlying deterministic (relatively simple) mechanism governing such transition times.
- The resulting pattern of relative frequencies should **discourage alternative explanations**.

4 Requirements for Transitions between Relative Frequencies

We first expose the arithmetic considerations that force us to consider long finite sequences. This development will also underlie subsequent approaches to approximation and complexity constraints. Assume that at time n_0 we have achieved a relative frequency of measure P_0 and wish to (nearly) achieve $P_1 \neq P_0$ at some minimal later time n_1. Let $\beta = n_1/n_0$ and note that $\beta > 1, n_1 > n_0 > 0$. Let P_2 be the measure representing the composition of outcomes on trials from n_0+1 to n_1. Hence, equality of compositions requires that

$$\beta P_1 = (\beta - 1)P_2 + P_0,$$

where P_0, P_1 are given and the minimal β and necessary measure P_2 are to be found. Unit normalization for P_2 follows immediately from that for P_0, P_1. Letting α denote the alphabet or sample space of possible elementary outcomes $\omega \in \alpha$, $\Lambda(\omega) = P_0(\omega)/P_1(\omega)$, we see that

$$\frac{\beta - \Lambda}{\beta - 1} = \frac{P_2}{P_1} \geq 0,$$

with the last inequality following from the required nonnegativity of P_2. It follows that nonnegativity implies the constraint

$$\beta \geq \max_{\omega \in \alpha} \Lambda(\omega),$$

and the minimal value of β is this maximum of Λ. Arithmetic round-off issues (βn_0 is usually not an integer n_1) can prevent us from exactly achieving a prescribed relative frequency for any finite sequence length.

We see that arithmetic requires that the sequence of waiting times, $\{n_i - n_{i-1}\}$, for successive transitions measures $P_{\mu_{i-1}}$ and P_{μ_i} grows at least exponentially fast. This suggests that these models will be appropriate when dealing with such long sequences as might arise from heavy-tailed waiting times. Heavy-tailed phenomena are conventionally modeled by distributions having infinite higher order moments, and often even all moments are infinite. While it is possible to construct (asymptotically stationary) simulation models having persistently fluctuating relative frequencies by using a block-length distribution having an **infinite mean**, we propose a method that does not require infinite mean random variables.

5 Approximating $\mathcal{M}$ by a Finite Covering $\hat{\mathcal{M}}$

In order to define what we mean by a complex pattern of relative frequencies closely visiting the members of $\mathcal{M}$, we first convert the typically uncountable $\mathcal{M}$ into a finite subset $\hat{\mathcal{M}}$. A further important reason for doing so is that Kolmogorov complexity K of sequences, discussed in the next section, does not differentiate with respect to the size of the differences between successive elements. The paradox is that, because there is a one-to-one correspondence between sequences $\underline{x}_n$, over an alphabet given by the set α, and their relative frequencies $\underline{r}_n$, highly complex α-valued sequences $\underline{x}$ can correspond to very simple sequences of relative frequencies if one ignores the tiny variations in relative frequencies that are the only ones possible for large n. Very small differences in the sequence of relative frequencies can yield high complexity,

$$(\exists c)(\forall \underline{x}_n) K(\underline{r}_n) > K(\underline{x}_n) - c.$$

We deal with this by quantizing the set $\mathcal{M}$ of measures determining a lower envelope through a finite approximating set $\hat{\mathcal{M}}$ of m measures that cover $\mathcal{M}$ to within a (small) value of an appropriately selected distance measure d. By requiring transitions between measures in $\hat{\mathcal{M}}$, we force the relative frequencies at transition times to have at least a minimum change. A natural choice for the distance between measures $d(P_0, P_1)$ is based upon the minimum ratio $\beta = n_1/n_0$ needed to transition from P_0 at n_0 to P_1 at n_1,

$$\log(\beta) = \max_{\omega\in\alpha}\{\log(P_0(\omega)) - \log(P_1(\omega))\} \leq$$

$$d(P_0, P_1) = \max_{\omega\in\alpha} |\log(P_0(\omega)) - \log(P_1(\omega))|.$$

It is readily verified that d possesses the properties of a metric on the probability simplex. Henceforth, we approximate the minimal β by the upper bound e^d and require the constraint

$$n_1 \geq e^d n_0. \qquad (*)$$

6 Maximum Kolmogorov Complexity

Our approach is based upon the notion of Kolmogorov complexity (see Li and Vitanyi [7]) and generates a simulation with the desired properties. While the resulting simulation must appear to be nonstationary, **nothing is assumed a priori about nonstationarity** beyond the constraint of Eq. (*). In Section 7 we will couple Kolmogorov complexity to the generation of the sequences we desire. A straightforward application of maximum Kolmogorov complexity to the data sequence itself would lead merely to a Bernoulli sequence for probability one-half.

We first introduce the complexity concept itself. Let $\mathcal{S}(\hat{\mathcal{M}})$ denote a recursive (so that it is effectively computable) countable set of finite length, strings over an alphabet α of size, say, κ. Let $\mathcal{S}_n(\hat{\mathcal{M}})$ denote the finite subset of $\mathcal{S}$ of strings of length n; $\mathcal{S}_n$ will be further identified in Section 7 as specifying the essential characteristics or state of the α-valued strings whose relative frequencies interest us. For purposes of simulation we wish to identify the most complex, least structured, most pattern-free strings in $\mathcal{S}_n$, subject to the constraints we will impose in Section 7 to select α-valued strings whose associated sequences of relative frequencies persistently approximate to measures in $\hat{\mathcal{M}}$.

The Kolmogorov complexity $K_\Psi(s|\mathcal{S}_n)$ of an element $s \in \mathcal{S}_n$ is the length $|p|$ of the shortest binary-valued string p that is a program for a universal Turing machine (UTM) Ψ whose output is the desired string $s \in \mathcal{S}_n$ (see Li and Vitanyi [7] for details). There exists a finite $g_{\mathcal{S},\Psi}$ for the UTM Ψ, such that if $||\mathcal{S}_n||$ denotes the finite ($\leq \kappa^n$) cardinality of the set $\mathcal{S}_n$,

$$\lceil \log_2 ||\mathcal{S}_n|| \rceil \leq \max_{s\in\mathcal{S}_n} K_\Psi(s|\mathcal{S}_n) \leq g_{\mathcal{S},\Psi} + \lceil \log_2 ||\mathcal{S}_n|| \rceil.$$

Hence, for large n, maximum Kolmogorov complexity strings in $\mathcal{S}_n$ have complexity about $\log_2 ||\mathcal{S}_n||$.

Kolmogorov complexity is not effectively computable (is not itself a recursive function), and this is closely related to the well-known undecidability of the halting problem.

If we select m and a string s^* at random from the set $\mathcal{S}_n$, then

$$P(K_\Psi(s^*|\mathcal{S}_n) \geq \log_2 ||\mathcal{S}_n|| - m) \geq 1 - 2^{-m}.$$

Hence, strings randomly selected from $\mathcal{S}_n$ will with high probability be nearly maximally complex. This provides a "practical" method for identifying nearly maximally complex strings and constructing simulations that are highly probable to generate appropriate sequences.

7 A Simulation Algorithm for Lower Envelopes

Let k denote the number of transitions between measures in $\hat{\mathcal{M}}$. Let m_ϵ denote the starting point for examination of relative frequencies, selected to ignore the initial necessarily large fluctuations in relative frequencies for small sample sizes.

Define

$$\Sigma_k = \{(\mu_i, n_i) : \text{for } i = 1, \ldots, k,$$
$$m_\epsilon \le n_i < n_{i+1} \le n, n_{k+1} = n,$$
$$e^{d(P_{\mu_i}, P_{\mu_{i+1}})} n_i \le n_{i+1}\},$$

as a sequence of k pairs specifying that measure $P_{\mu_i} \in \hat{\mathcal{M}}$ is achieved at sample position n_i.

7.1 Algorithm

1. Choose k and Σ_k having the jointly maximum Kolmogorov complexity.
2. This can be assured with high probability by random selection. (These should then be the most irregular and least predictable choices of pairs of measures and transition times satisfying the given constraints.)
3. Interpolate so as to achieve the successive measures required at the transition times, as determined by Step 1.
4. This interpolation, between, say, initial P_{μ_i} at n_i and final $P_{\mu_{i+1}}$ at n_{i+1}, will have a high probability of being of maximum complexity if it is carried out by an independent and identically distributed sequence of discrete random variables governed by a measure Q_{i+1} satisfying

$$\gamma_{i+1} = \frac{n_{i+1}}{n_i}, \quad Q_{i+1} = \frac{\gamma_{i+1} P_{\mu_{i+1}} - P_{\mu_i}}{\gamma_{i+1} - 1}.$$

   Success is assured by satisfaction of the constraint that

$$\gamma_{i+1} \ge e^{d(P_{\mu_i}, P_{\mu_{i+1}})} \ge \beta_{i+1}.$$

8 Statistical Analysis

Random selection requires knowing, for each k, the number $R(k)$ of sequences of type Σ_k. An approximation to the number $R(k)$ of sequences of length n with first transition point occuring no earlier than m_ϵ, there being exactly k transition points on a set of m measures, and we consider transitions only to the ν nearest neighbor measures lying at a distance $\underline{d}$ from the current measure, is given by

$$R(k) = m\nu^{k-1} \frac{(n - e^{\underline{d}(k-1)} m_\epsilon)^k}{k! e^{\frac{1}{2}\underline{d}k(k-1)}}. \qquad (**)$$

The entropy $H(k)$ of the process generating sequences of type Σ_k, conditional upon k, is just $\log(R(k))$.

The probability π_k of choosing a particular number of transitions k is given by

$$R = \sum_j R(j), \quad \pi_k = P(K = k) = \frac{R(k)}{R}.$$

Further analysis reveals that the number k of transition points has a probability mass function that is approximately normally distributed.

The overall entropy of the process generating sequences of type Σ_k is

$$H = \log(R),$$

which is a good approximation to the Kolmogorov complexity (see Section 6).

9 Examples of Ternary-valued Simulation Sequences

Throughout this simulation we have $\alpha = \{1, 2, 3\}$, and $\mathcal{M}$ is defined as the convex hull of the following three extreme point measures described as the rows of the matrix

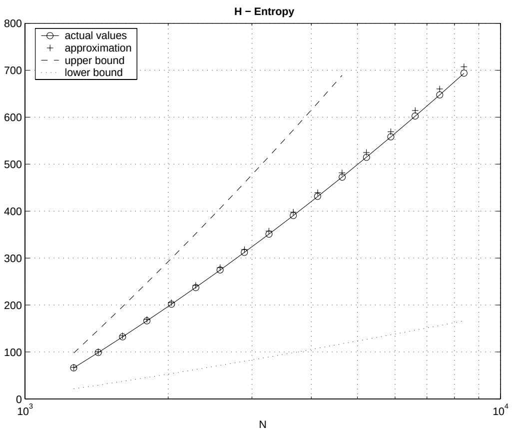
$$\mathbf{P} = \begin{pmatrix} 1/3 - 1/36 & 1/3 + 1/72 & 1/3 + 1/72 \\ 1/3 + 1/72 & 1/3 - 1/36 & 1/3 + 1/72 \\ 1/3 + 1/72 & 1/3 + 1/72 & 1/3 - 1/36 \end{pmatrix}.$$

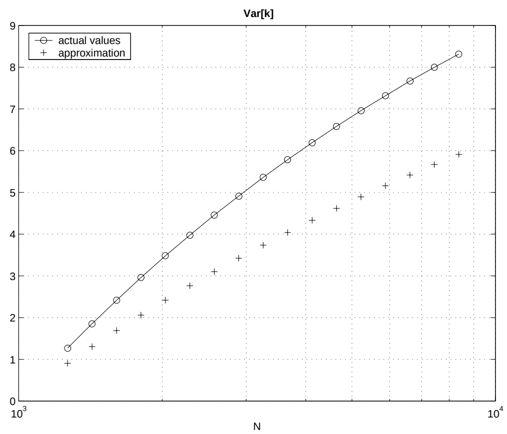
Furthermore, we start making transitions only after $m_\epsilon = 1000$, to ignore unavoidably large fluctuations.

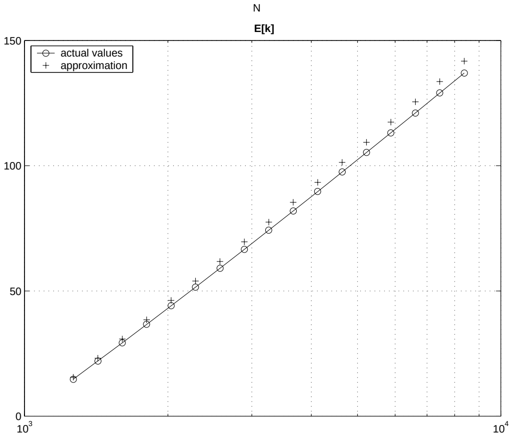
The two simulations displayed differ only in the density of the covering of $\mathcal{M}$ by $\hat{\mathcal{M}}$. The maximum sequence length is governed by Matlab's largest integer $R = 1e300$ for which $\log(R) = 691$. In each of the two examples below, we provide the following five plots: comparisons of the actual entropy H of the transition pairs $\{(\mu_i, n_i), i = 1 : k\}$ to the approximation based on Eq. (**), and the variance and mean of the number k of transition points as a function of sequence length together with our analytical approximations to them; the relative frequencies for the three outcomes from 1000:end; the relative frequency trajectory shown with the approximating measures.

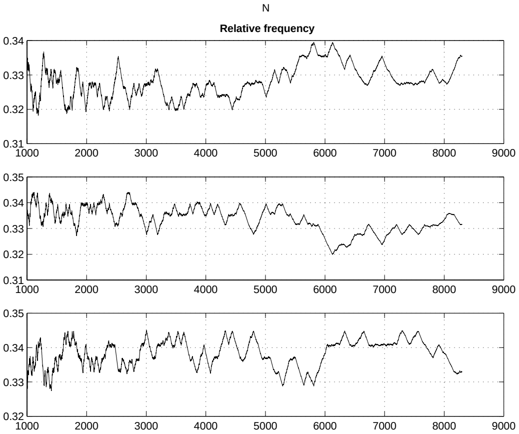
In the first example, using a coarse level of approximation,

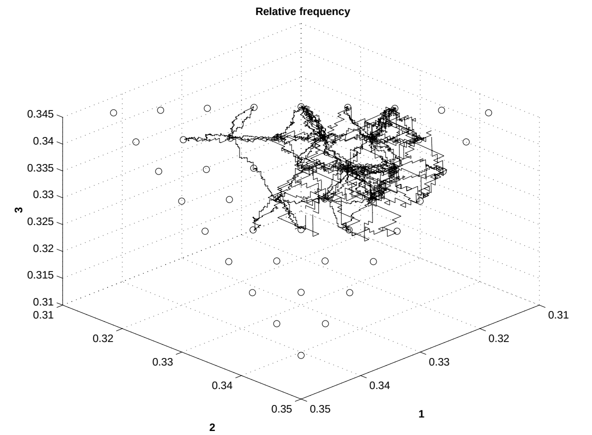
$$d = 1.2e - 2 \Rightarrow m = 45.$$





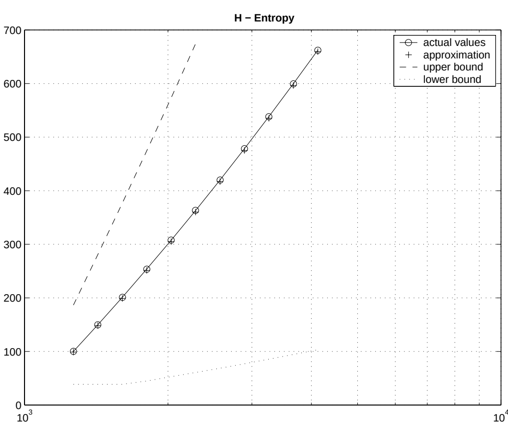


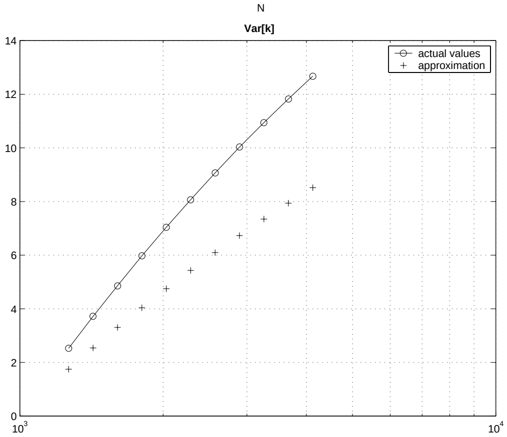


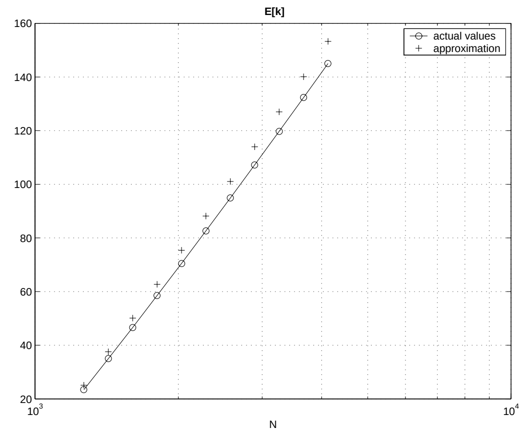


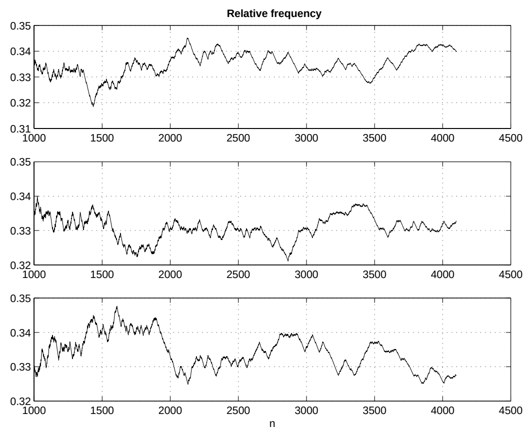
In the second example, using a finer level of approximation,

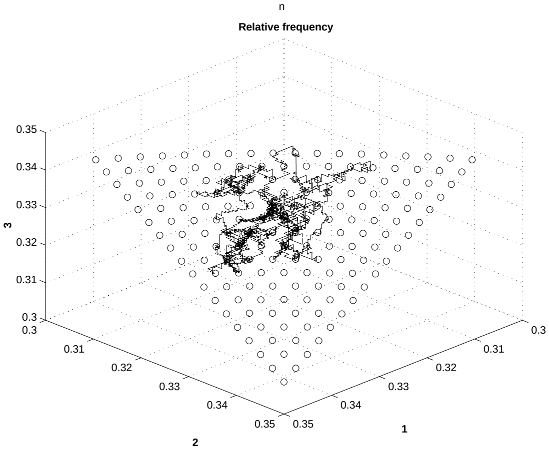
$$d = 7.1e-3 \Rightarrow m = 171.$$











10 Modeling

We turn to the inverse problem to simulation, inference from data to a model. How might we model given long-run unstable frequentist data? An approach is illustrated by the following

10.1 Algorithm

1. From the finite sequences of relative frequencies $\{r_n(\omega)\}$ on an alphabet (sample space) α of size κ, compute the relative frequency trajectory $\mathcal{T}$ lying in the probability simplex in $\mathbb{R}^\kappa$. Do so only for $n \geq m_\epsilon >> 1$.

2. Estimate $\mathcal{M}$ by the convex hull of $\mathcal{T}$, or by a smoother version defined by the hyperplanes derived from a restricted set of random variables (gambles) $\{X_i\}$ of particular interest. These random variables induce hyperplanes with normal $\pm X_i$ and thresholds $\min_{k \geq m_\epsilon} E_{r_k} X_i$ and $\max_{k \geq m_\epsilon} E_{r_k} X_i$ via
$$\{\mu : \min_{k \geq m_\epsilon} E_{r_k} X_i \leq E_\mu X_i \leq \max_{k \geq m_\epsilon} E_{r_k} X_i\}.$$

3. A particular choice for the random variables $\{X_i\}$ is the set of indicator functions for events. In this case, the indicated minimum and maximum expectations are the lower and upper probabilities.

4. Enlarge the convex set constructed above by some form of "discounting" (e.g., Fierens [4], Shafer [11], Walley [13], Sec. 5.3.5) to cover the possibility that the data sequence is not long enough to exhibit all of the measures in $\mathcal{M}$.

At this time we have little to provide by way of an analytical study of the properties of this algorithm. We note that there are asymptotic analyses of learning $\mathcal{M}$ from frequentist data in Cozman and Chrisman [2] and Walley and Fine [12]. What, for example, is the analog of traditional "goodness-of-fit" in this setting?

11 Application to Vowels Data

Data to which to apply our modeling approach should exhibit long-run instability of relative frequencies, without admitting a fairly simple explanation for this instability. Possible examples, drawn from such Internet variables as lengths of web files requested and sizes of packets and delays in their transmission, are discussed in Crovella, Taqqu, and Bestavros [3] and Willinger and Paxson [14], albeit these authors argue for conventional but heavy-tailed statistical models. We apply the above approach to model data on the occurrences of the vowels `a,e,i,o,u` in messages drawn from Internet job postings in the first half of December 2000. We restrict our attention to just the 5.5 million occurrences of these vowels, ignoring all other ascii characters in the approximately 12000 individual messages. In order to render intelligible the displays given below, we clustered the five vowels into three classes by grouping `[a,u]` and `[i,o]`. For typographical reasons, all the plots for this section follow the References.

We divided the vowels data into twenty consecutive long blocks of lengths 250000 (this being deemed "long" but not "very long"). We calculated the relative frequencies of the three clusters and display their trajectory starting from trial 1000 (to eliminate the usual large fluctuations of relative frequencies in short initial sequences). The trajectory for the first such block is shown first, followed by the trajectories for the first nine blocks. The bounding box shown in the second figure is the convex set determined by just the lower and upper probabilities of elementary outcomes. The third figure repeats the second figure, but this time for blocks 11 through 19. The fourth figure then compares the two convex bounding boxes for both data sets, circles indicating the box corresponding to the first nine trajectories. Note that the boxes overlap substantially but fall short of the coincidence one might desire. This suggests that the vowel data has substantial temporal variability, and if we wished the convex set estimated, say, from Blocks 1:9 to contain the trajectories generated by Blocks 11:19, then discounting is needed.

Finally, using the convex estimate based upon the 30 nontrivial events for five vowels, we then simulated a trajectory, shown in Figure 5, for samples sizes 1e3 to 5e3.

In the paper to be presented in June we expect to have additional vowel data and comparison of the results obtained from our envelope methodology with those from a traditional stochastic model.

References

[1] de Cooman, G., F. Cozman, S. Moral, P. Walley, eds. [1999], *Proceedings of the First International Symp. on Imprecise Probabilities and Their Applications,* Imprecise Probabilities Project, Universiteit Gent, Zwijnaarde, Belgium.

[2] Cozman, F., L. Chrisman [1997], *Learning Convex Sets of Probability from Data,* CMU-RI-TR 97-25, Carnegie-Mellon University, Pittsburgh, PA.

[3] Crovella, M., M. Taqqu, A. Bestavros [1998], Heavy-tailed probability distributions in the world wide web, in R. Adler, R. Feldman and M. Taqqu, eds., *A Practical Guide to Heavy Tails: Statistical Techniques and Applications,* Birkhauser, 3–25.

[4] Fierens, P.I. [2000], *Interval-valued Probability Models of Unix File Sizes,* Sec. 4.5.2, M.S. dissertation, Cornell University, Ithaca, NY.

[5] Kolmogorov, A.N. [1963], On tables of random numbers, *Sankhya: The Indian Journal of Statistics, Series A,* p.369.

[6] Krantz, D., R. Luce, P. Suppes, A. Tversky [1971], *Foundations of Measurement,* vol. I, Academic Press, NY.

[7] Li, M. and P. Vitanyi [1997], *An Introduction to Kolmogorov Complexity and Its Applications, Second Ed.,* Springer-Verlag, NY.

[8] Papamarcou, Adrianos, T.L. Fine [1991], Unstable collectives and envelopes of probability measures, *Annals of Probability,* **19**, 893–906.

[9] Ridley, M. [1999], *Genome: The Autobiography of a Species in 23 Chapters,* pp. 23, 24, HarperCollins Pubs., New York.

[10] Sadrolhefazi, A. and T. Fine [1994], Finite-dimensional distributions and tail behavior in stationary interval-valued probability models, *The Annals of Statistics,* **22**, 1840–1870.

[11] Shafer, G. [1976], *A Mathematical Theory of Evidence,* Princeton University Press, Princeton, NJ, 251–255.

[12] Walley, P. and T. Fine [1982], Towards a frequentist theory of upper and lower probability, *The Annals of Statistics,* **10**, 742–761.

[13] Walley, P. [1991], *Statistical Reasoning with Imprecise Probabilities,* Chapman and Hall Pubs., London.

[14] Willinger, W., V. Paxson [1998], Where the mathematics meets the Internet, *Notices Amer. Math. Soc.,* **45**, 961–970.

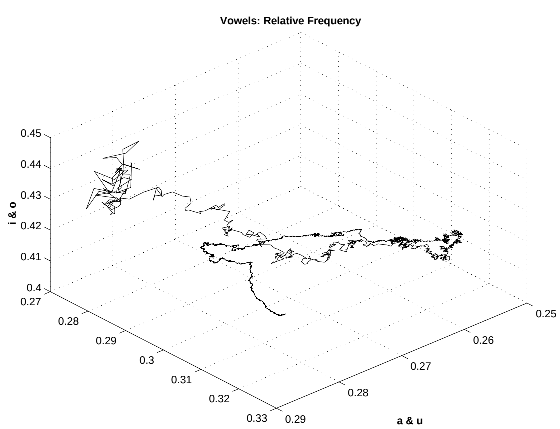


Figure 1: Trajectory of Block 1 of Vowels

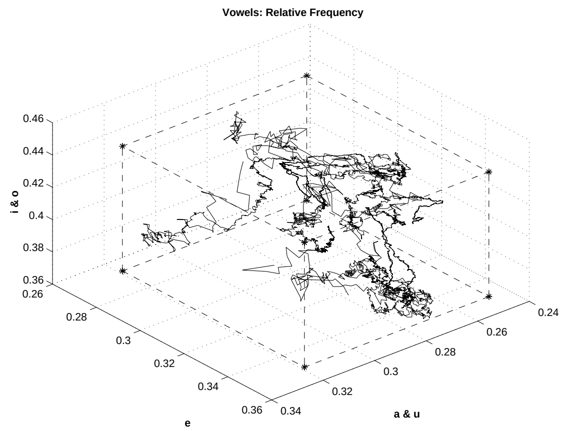


Figure 2: Trajectories of Blocks 1:9 of Vowels

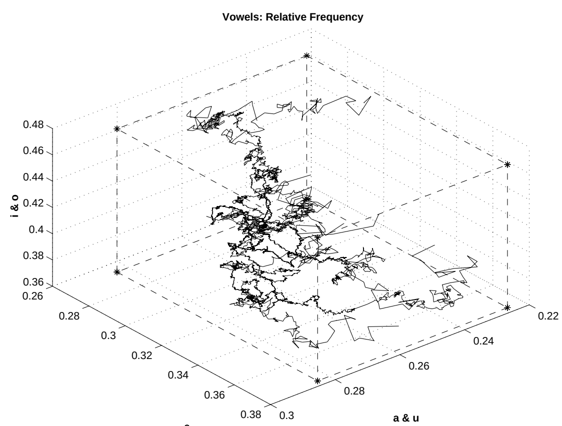


Figure 3: Trajectories of Blocks 11:19 of Vowels

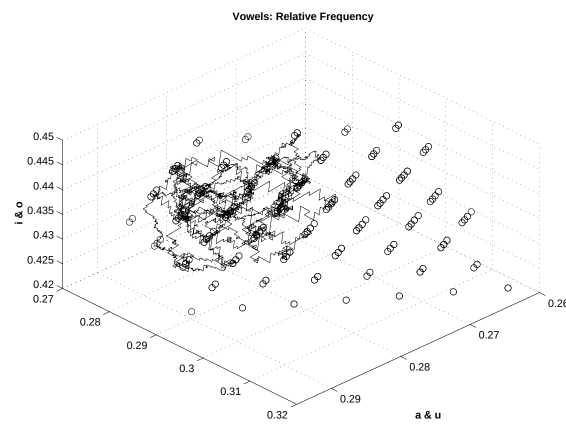


Figure 5: Simulated Vowel Data

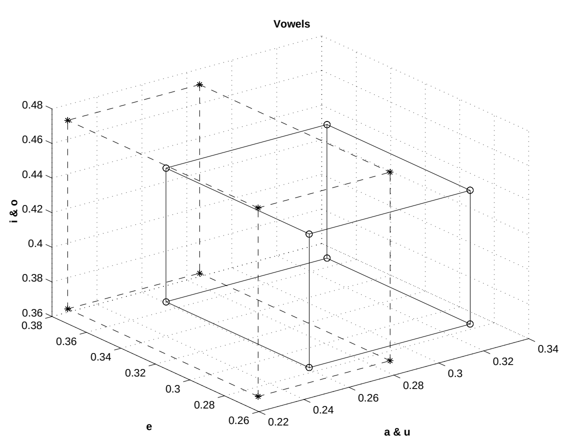


Figure 4: Two IVP Bounding Sets