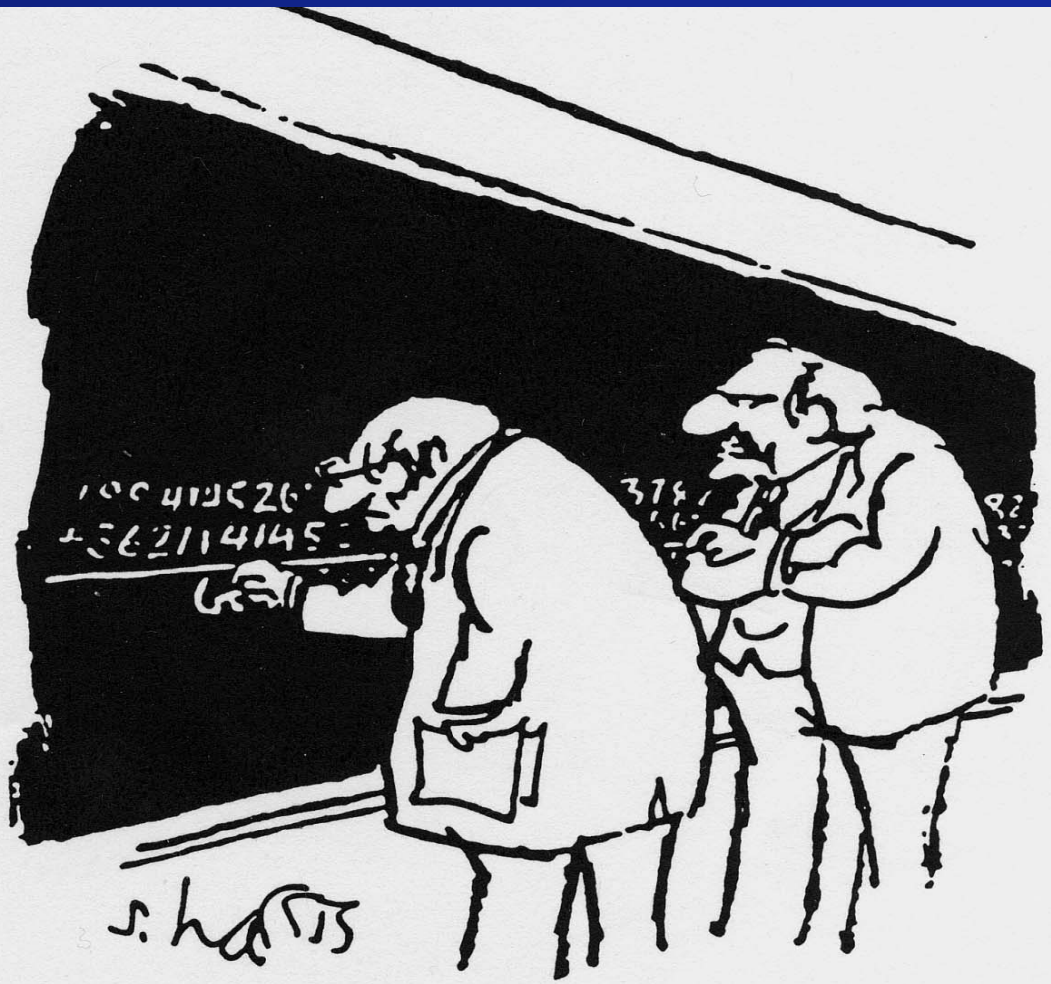
Randomness and Serial Irregularity

Steve Pincus

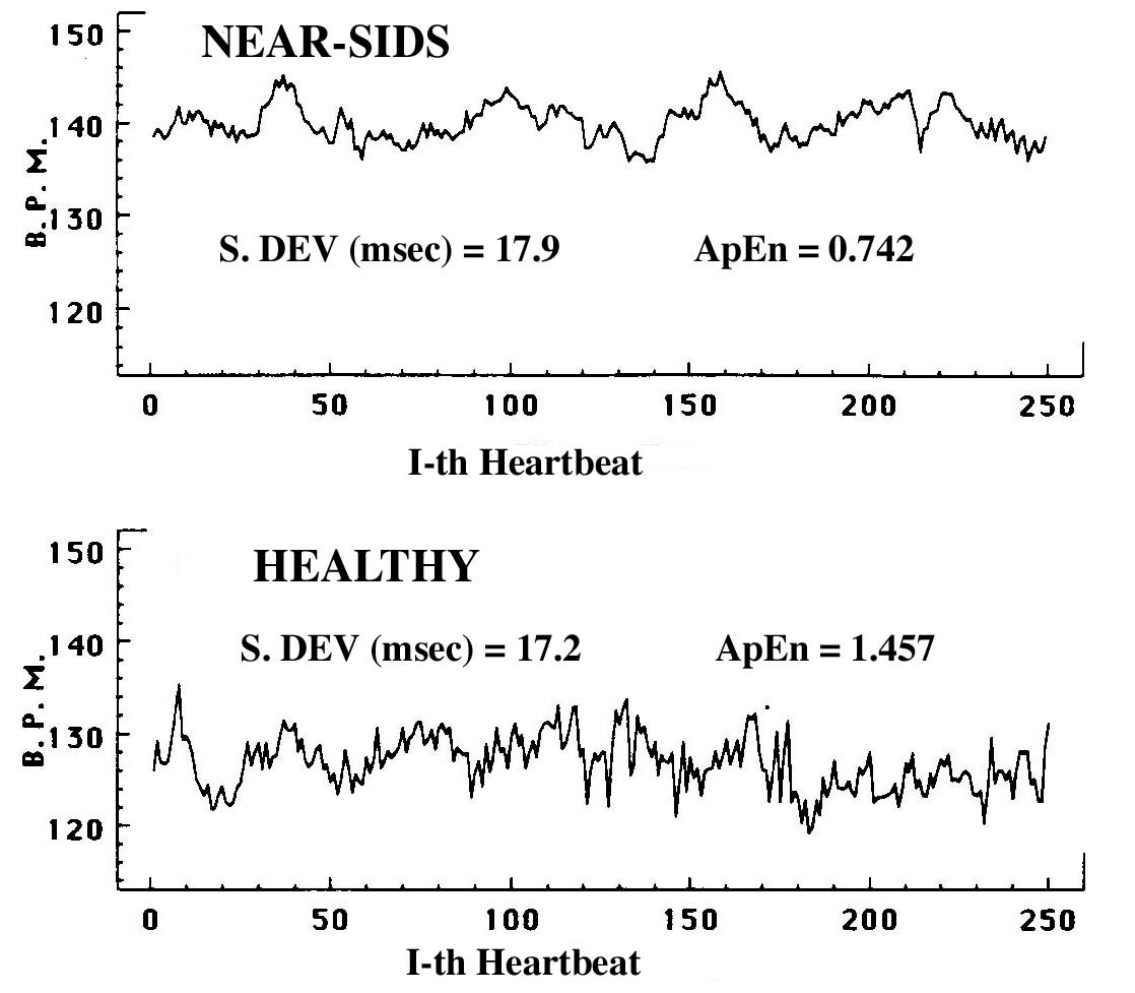


"Adding two numbers which probably have never been added before is <u>not</u> considered a mathematical breakthrough"

Prehistory: Biomedical Applications

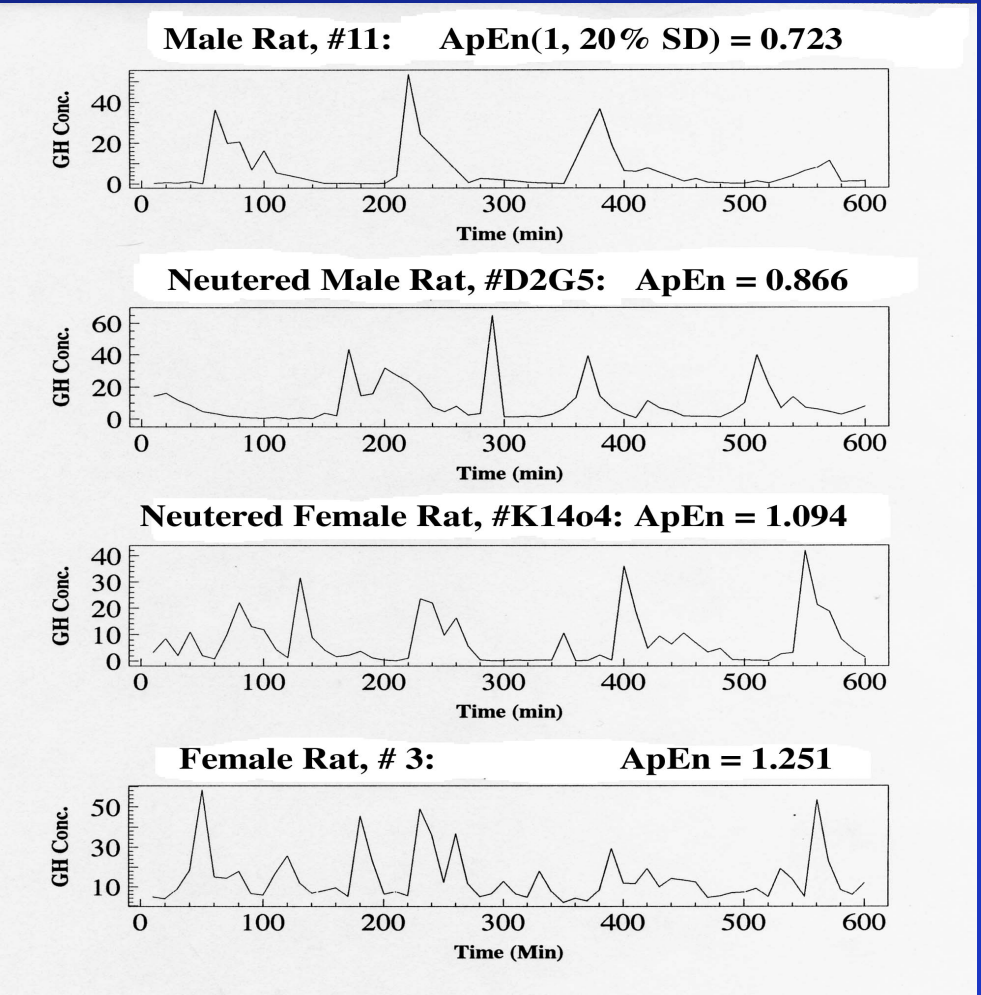
- Reports of changing complexity in biomedical series (dimension, largest Lyapunov exponent, entropy)
- Visual evaluation of heart rate, EEG data confirmatory, suggest changing `extent of irregularity` w. pathology
- But direct complexity assessments did not typically agree with visuals
- Critical opportunities (among many):
  - Fetal heart rate monitoring utility in doubt
    - Primary determinant of C-section vs. vaginal delivery
  - EEG not routinely used in clinical settings
    - LA evaluations of concussions, response to antidepressants, depth of anesthesia …
- Needed: more from extant data, not (necessarily) more technology

2 Infant Heart Rate Tracings



Pincus SM, Cummins TR, Haddad GG, *Am J Physiol* 264, R638-646 (1993)

Gender Differences, GH: Entire & Neutered Rats



Gevers E, Pincus SM, Robinson ICAF, Veldhuis JD, *Am J Physiol* 274, R437-444 (1998)

MOTIVATION FOR ApEn DEVELOPMENT

- **Applications oriented -- Finite, noisy serial data**
- **Sensibly grade 'more random', not just Y/N iid.**
- **Calibrate subtle, difficult to `eyeball` differences**
- **Chaos, complexity algorithms *per se* generally inappropriate to biologic data**
  - **Models likely complicated mixed/random networks**
  - **Nonreplicable findings (huge "error bars"): curse of dimensionality**

MOTIVATION FOR ApEn DEVELOPMENT(2)

- Thematically faithful modification required
  - Visual intuition should match numerics
  - 'Model-independent' (broad class) applications
  - Replicable findings
  - Revisit Info. Theory, 'rate of entropy' -- with a twist

- Oftentimes modeling nearly impossible; however, full modeling not required to discriminate, classify

- Spectrum, correlation more limited than perceived
  - estimation, interpretation issues
  - for broad-banded [non-linear/non-Gaussian processes & composite] systems, often don't efficiently capture system change

(Descriptive) Definition of ApEn

- **Given time-series u(i), fix m and r**

ENGLISH: ApEn measures the (logarithmic) likelihood that runs of patterns that are close (within tolerance r) for (window length) m observations remain close on next incremental comparisons.

TECHNICAL: ApEn(m,r) = - (average over i of ln [conditional probability that $|u(j+m)-u(i+m)| \leq r$, given that $|u(j+k)-u(i+k)| \leq r$ for k=0,1,..., m-1].

- **Greater likelihood of remaining close & more recurrent features, regularity, produce smaller ApEn values.**
- **Conversely, greater apparent randomness yields larger ApEn values**

Pincus SM, *Proc Natl Acad Sci USA* 88; 2297-2301: 1991

Definition of ApEn

- (A) Fix integer $m \geq 0$, real $r > 0$, and sequence length N
- (B) For a sequence of real numbers $\underline{u} := (u(1), u(2), \ldots u(N))$
  - (i) define block $\underline{x}(i) = (u(i), u(i+1), \ldots u(i+m-1))$
  - (ii) define $d(\underline{x}(i), \underline{x}(j)) = \max_{k=1,2,\ldots,m} (|u(i+k-1) - u(j+k-1)|)$
- (C) Let $C_i^m(r) =$ (no. of $j \leq N-m+1$ such that $d(\underline{x}(i), \underline{x}(j)) \leq r$) / $(N-m+1)$

Remark: The numerator of $C_i^m(r)$ counts, to within a resolution r, the number of blocks of consecutive values of length m that are approximately the same as a given block of consecutive values.

- (D) Define $\Phi^m(r) = \dfrac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log C_i^m(r)$

- (E) Define $ApEn(m, r, N)(\underline{u}) = \Phi^m(r) - \Phi^{m+1}(r)$, $m \geq 1$;
  $ApEn(0, r, N)(\underline{u}) = -\Phi^1(r)$.

- ApEn(m, r, N)(u) measures the log. frequency with which blocks of length m that are close together remain close for blocks augmented by one position

- Greater likelihood of remaining close & more recurrent features, regularity, produce smaller ApEn values, and conversely

Pincus SM, *Proc Natl Acad Sci USA* 88; 2297-2301: 1991

ApEn is a RELATIVE MEASURE

- Must fix
  - m (window length)
  - r (tolerance level)
  - N (length of time-series)
  - Typical setting: m = 1 or 2; r = 0.2 SD; 50 $\leq$ N $\leq$ 5000

- ApEn
  - is a relative measure of regularity
  - encapsulates *marginal probability* information
  - is NOT intended as an estimate of K-S entropy (which is often infinity)

Primary Statistical Properties

- **Small `error bars` (good replicability):**
  ApEn SD < 0.06 for diverse classes of processes
- **Asymptotic normality, & LIL (Law of Iterated Log)**
  - **Symmetrized version of ApEn ($ApEn_w$, defined by periodic extension) is a U-Statistic, implying CLT & LIL**
  - **Easy estimate for upper bound, ApEn - $ApEn_w$ (vanishes in limit)**
- **χ^2 limiting distribution for finite alphabet, discrete state** (Alhakim A, Molchanov S, 2001; Rukhin AL, 2000)
- **ApEn has higher power against runs test for short (binary) series** (Chatterjee, Yilmaz, Habibullah, & Laudato, 2000)

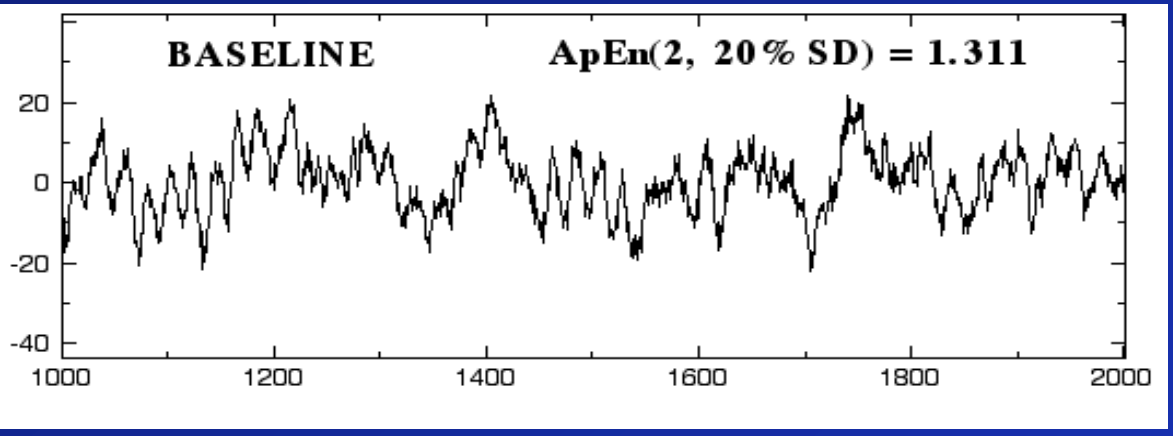
Tradeoffs & Familiar Friends

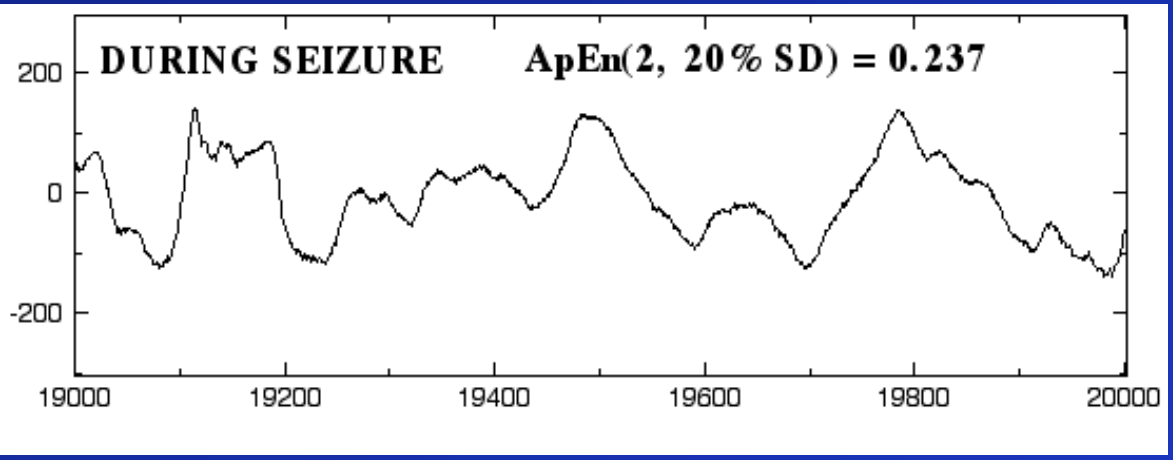
- ApEn agrees with `familiar` measures in specified contexts
  - [Discrete state] For Markov Chains, ApEn converges to rate of entropy as $N \to \infty$ [Deterministic dynamical systems] ApEn (m(N), r(N)) converges to Kolmogorov-Sinai entropy as $N \to \infty$, with m(N) and r(N) converging (suitably) to ∞ and 0, respectively

- Thus ApEn extends these measures to settings of
  - Finite lengths (non-asymptotic)
  - Unknown, composite or other (most of reality) models

- m, r specification balances tradeoffs between data length N & resolution
  - Bigger m, smaller r $\Longrightarrow$ sharper parameter info.
  - But for fixed N, the accompanying (conditional probability) estimates are less robust
  - Balance resolution and replicability

2 EEG RECORDINGS (T4): Seizure Anticipation

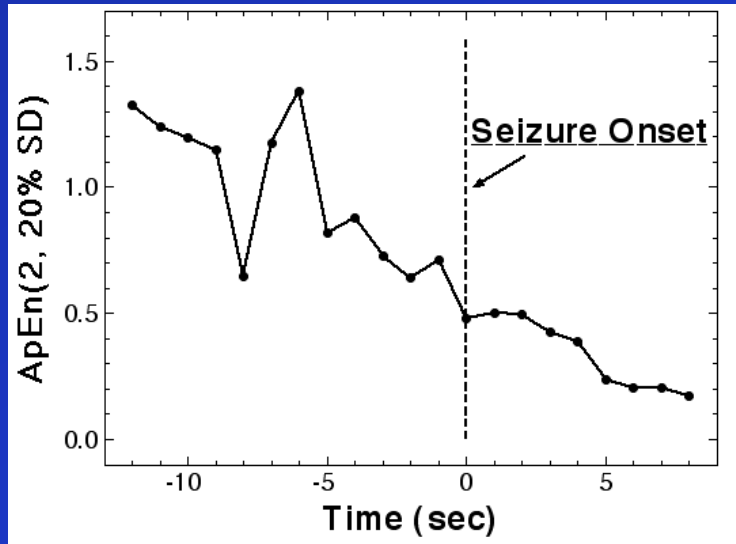




- MTLE (Mesial Temporal)
- Scalp, Quiet Wake

  (~, sharper for intracranial)
- Anticipation long before event
- Steady Acute Event ApEn Decrease

Zaveri H, Novotny E, Duckrow R, Spencer S/ Yale Neurology



Application of ApEn to Biomedical Time-Series

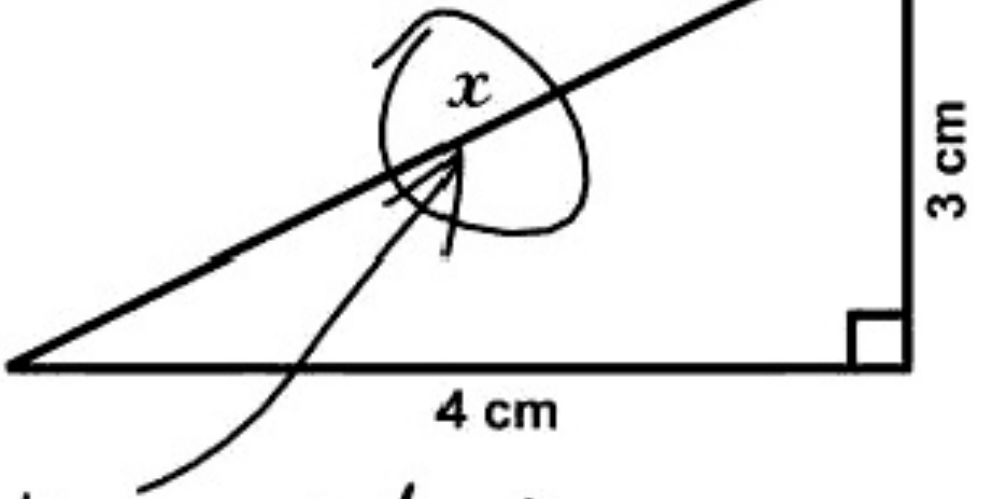
- Numerous studies
- Subclinical detection, & changes earlier than moments
- Evaluate efficacy of therapy
- Multiple signals & Applications:
  - Heart rate / ECG: Occult CAD; A. & V. Fibrillation; Hypertension; Drug abuse; Fetal distress/sheep;A-SIDS
  - Blood pressure: Syncope
  - Hormone secretory pulsatility: LH, FSH, testosterone, GH, Insulin, … $\Longrightarrow$ Endocrine tumors; Gender & Aging diffs; Diabetes Onset; ERT Patch v. Pill
  - Balance platform, EMG: Parkinson's; Alzheimer's; tremor; motor control
  - EEG: Seizure; depth of anesthesia; mental disorders; stroke
  - Mood ratings (psychiatry): PMS, depression
  - Neuronal firing: burstiness of ion channels
  - MRI: noise reduction in imaging
  - DNA: [potential] classifying `junk` DNA, noncoding regions

Limitations Imposed by Continuous State

- Information-theoretic entropy measures most natural in discrete state (e.g., binary or integers) -- Shannon onwards
- Continuous (cts) state constructions impose `partitions`, then compare `entropies` on each partition
- Entropies on non-nested partitions (even in limit) need not relate nicely
  - Manifested by *lim sup* (not lim) specification in Ornstein continuous state entropy formulation
  - *Flip-flop pair* ** discloses similar conundrum in ApEn context (issue the continuum, not ApEn formulation): Process A may appear more random than B on many choices of partitions, but not necessarily on all partitions of suitably small diameter
- Conclusion: in cts state, relative (to partition, or (m,r) choice) regularity is all that is possible, in general
- Usefully, typically a relative consistency in ApEn, both mathematically & experimentally:
  $ApEn(m,r)(A) \leq ApEn(m,r)(B) \Longrightarrow ApEn(n,s)(A) \leq ApEn(n,s)(B)$ & conversely for wide (m,r), (n,s) range
- **Suggests discrete state studies ahead**

** Pincus SM & Huang WM, *Commun Statist*. 21; 3061-3077: 1992

3. Find x.



Here it is

Reduction to Binary

- Many possible directions for further ApEn theory & applications
- Binary state space {0, 1} chosen
- Cleanest, yet discloses essential perspectives
  - Single sequence evaluations of extent of irregularity, algorithmic & combinatorial
    - Both finite & infinite
  - Yet consistent with classical probability, statistics & information theory
  - (!) Develop `randomness` bottom-up from small finite blocks, rather than infinite limits
  - Produces explicit specifications of maximally irregular sequences
  - Grades sequences by proximity to maximal irregularity
  - Refines (delineates) classical notions of independence, normality
  - Key extensions:
    - Bivariate (conditional irregularity; applicable as asynchrony measure)
    - Spatial arrays (vector-ApEn identifies, corrects flaw in RA Fisher experimental design)
  - Everything computable

Simpler Expression for ApEn in Binary State

Set state space as binary {0, 1}, with r < 1 our measure of resolution

- **note that we then monitor exact matches in blocks**.

- Thus **ApEn** = $\mathbf{H_m - H_{m+1}}$:

  ` difference between m-block and m+1- block (`classical Shannon`) entropies, interpreted for single seq.`

Ex: m=1: $ApEn(1)(\underline{u}) = \{f_{\{0\}} \log f_{\{0\}} + f_{\{1\}} \log f_{\{1\}}\} - \{f_{\{0,0\}} \log f_{\{0,0\}} + f_{\{0,1\}} \log f_{\{0,1\}} + f_{\{1,0\}} \log f_{\{1,0\}} + f_{\{1,1\}} \log f_{\{1,1\}}\}$.

Chaitin Example

- Motivation for (noncomputable) Algorithmic Complexity

- Two binary sequences (length 20, Sci Am 232, p. 47, 1975)
  (A) 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1
  (B) 0 1 1 0 1 1 0 0 1 1 0 1 1 1 1 0 0 0 1 0

- "Each represents event with prob. 2^{-20} (classical prob.)"

- Present perspective -- think about 2-blocks

- Observe computable difference
  ApEn(1) = 0 for A
  ApEn(1) = 0.6774 for B

- As aside, Chaitin did pretty well for B
  max ApEn(1) ≈ ln 2 = 0.693

CENTRAL RESULT: `Recipe for Randomness` (1)

- Constructive algorithm for normal numbers (asymptotically equidistributed, all block lengths)

- Addresses near void, despite "almost all numbers" normal
  -- Champernowne # .123456789101112…, & binary analog, variants primary examples

- Building blocks: finite max. irregular sequences – liaison between finite, infinite length worlds

- Choose L(n), nondecreasing & integer-valued, $L(n) \to \infty$

- For all n, choose max. irregular v_n, length $v_n = L(n)$.

- Form $w_m = v_1 \vee v_2 \vee \ldots v_m$ ($\vee$ = concatenate/paste together)

- Theorem(recipe): If L(n) non-exponential*, then $\lim_{m\to\infty} w_m$ normal (base 2)

  * Thm. 10, Pincus S & Singer BH, *Proc Natl Acad Sci USA* 95, 10367-72 (1998):

    $nL(n) = O(S_n)$, where $S_n := \sum_{i=1}^{n} L(i)$

  * If growth exponential, fine at breakpoints, but possibly intermediate `toxic` excursions

Design Your Own Normal Number

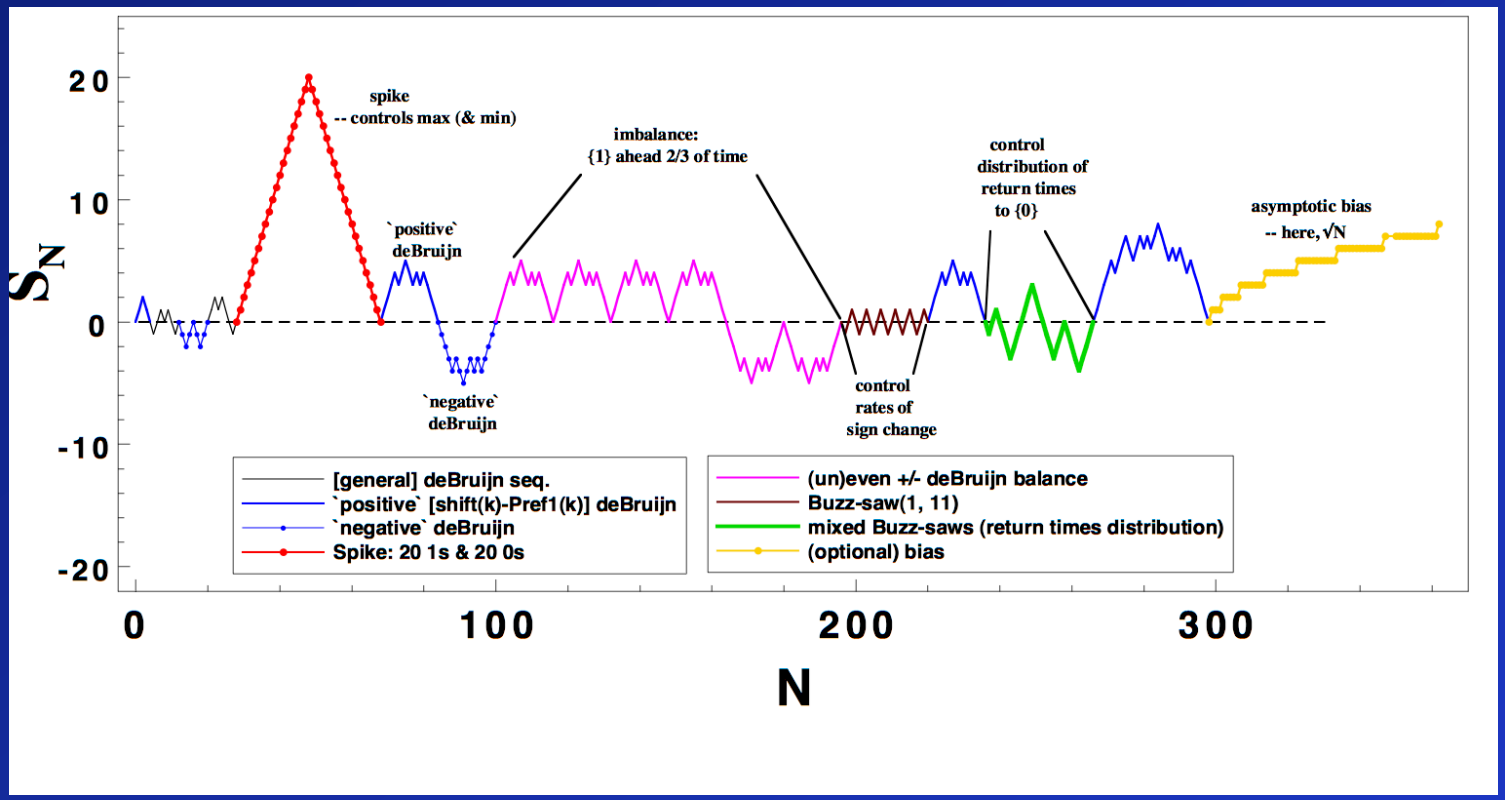


Figure 1. `Design your own normal number`. This schematic sample path indicates various techniques to precisely control distinct attributes of normal number production. The rubric to guarantee normality is to concatenate finite maximally irregular sequences of non-decreasing lengths L(i) (subject to a very modest growth rate restriction), together with possible insertions or perturbations that in aggregate are o(N). (A) L(i) controls the rates of convergence of block frequencies to equidistribution. (B) deBruijn sequences provide maximal irregularity for concatenate lengths a power of 2, with shifted Prefer 1 (`positive`) deBruijn sequences and their negations ensuring local positivity or negativity of S_N; more generally, the selection of maximal elements greatly influences S_N. (C) spike insertions (local runs of 1s followed by 0s, or conversely) control extremal values [Theorem 2]. (D) Imbalance in positive vs. negative deBruijn concatenates can tilt the `fairness` of the game (% of time that 1 leads 0) [Theorem 9]. (E) Fixed height Buzz-saw insertions control the rate of sign changes of S_N [Theorem 7]. (F) Mixed height Buzz-saw insertions control the distribution of return times to {0}[Theorem 8]. (G) embedded insertions can provide any o(N) rate of bias, if desired [Theorem 2, ref. 15].

ApEn Summary

- Sequential theory of 'randomness'
  - Finite and infinite lengths
  - Model independent
  - Both short seqs. ($N \geq 5$) & asymptotics
  - Notion of maximal irregularity
- Refines " i.i.d." & normal #'s into classes
- Analytic formulas
- Rigorous statistical development
- Spatial & bivariate extensions
- Computability (vs. Alg. complexity)
- Mechanistic interpretation for models
- Numerous, diverse applications

- SM Pincus, *Proc Natl Acad Sci* 88; 2297-2301: 1991
- SM Pincus & WM Huang, Comm Stat - Thy Meth 21; 3061-77: 1992
- SM Pincus, *Math Biosci* 122; 161-181: 1994
- SM Pincus & BH Singer, *Proc Natl Acad Sci* 93; 2083-88: 1996
- SM Pincus *et al*, *Proc Natl Acad Sci* 93; 14100-105: 1996
- SM Pincus & RE Kalman, *Proc Natl Acad Sci* 94; 3513-18: 1997
- BH Singer & SM Pincus, *Proc Natl Acad Sci* 95; 1363-68: 1998
- SM Pincus & BH Singer, *Proc Natl Acad Sci* 95; 10367-72: 1998
- SM Pincus & BH Singer, *Proc Natl Acad Sci* 109; 19145-50: 2012
- SM Pincus & BH Singer, *Proc Natl Acad Sci* 111; 5485-90: 2014

Fit with Main Q's & Working Group Objectives

- I'm worried – myriad queueing & traffic model forms (among other types of networks) decidedly non-hierarchical
- Given recognition that different levels of spatial resolution can invert ranking by complexity (`flip-flop pair` as an example), regard analyses at specified levels as distinct
- Systems without natural hierarchies may be considered more `complex`, but not from all formulations of complexity
- Consider, e.g., the constant e (also re Working Gp Q's)
  - Algorithmically e very simple (sum of 1/ n!)
  - Yet e highly irregular / `unpredictable`, quite possibly normal
- Also, same number in different bases can be qualitatively different in `complexity` / irregularity: W. Schmidt (1960) showed order of the continuum of numbers normal in one base, not even simply normal in another, non-commensurate base
- Conclusion: live with each given formulation of complexity as distinct, & analyze *de novo*

"He was working on a theory of entropy, and developed a severe case of it himself."