An Axiomatic Approach to The Law of Small Numbers

Jawwad Noor [1] Fernando Payró [2]

[1]Boston University

[2]UAB, BSE & CODE

EEA-ESEM 2023

Motivation

Studies show that people do not understand randomness. A prominent finding is that people do not expect "streaks" to persist

- Gambler's fallacy
  - Probability of heads decreases after a streak of heads:

    $$P(HHT) > P(HHH)$$

  - Evidence: Terrell and Farmer (1996), Croson and Sundali(2005), Suetens et al (2016), Mueller et al (2021),...
- Excessive Alternation
  - Sequences that have too many alternations and too few streaks are more likely than they should:

    $$P(HTHTTH) > P(HHHTTT)$$

  - Evidence: Rapoport and Budescu (1997), Bar Hillel and Wagenaar (1991),..

Literature

Psychology:

- The leading theory is Kahneman and Tversky (1971,1974) "Law of Small Numbers"

  "Even small samples are highly representative of the populations from which they are drawn"

  "People expect that the essential characteristics of the process will be represented, not only globally in the entire sequence but also locally in each of its parts"

Economics:

- Rabin (2002): $P(x_1, x_2, ...) = P(x_1, x_2)P(x_3, x_4)\cdots$
  - $P(x_i, x_{i+1})$ sample without replacement from an urn
  - Urn with one head ball and one tail ball: $P(HT) = \frac{1}{2}(1), P(HH) = \frac{1}{2}(0)$
- Literature provides tractable models and direction

This Paper

We are interested in addressing foundational questions:

- Can the Law of Small Numbers be formalized?
  - Each small sample is highly representative of the fairness of the coin
  - Does this mean that each segment of the [sequence] is highly representative of the “fairness” of the coin”? What is a “segment”?
  - Can it be identified from behavior?
- Are there general implications for LSN?
  - E.g. Rabin’s model implies an LLN type of result. Is this because of LSN or the specific parametric assumption?

This Paper

- Is LSN a characteristic that should have been eliminated by evolution?
  - Is it the case that "rational" agents would outperform LSN agents in an evolutionary race?
- What is a streak?
  - How many contiguous heads is a streak?
  - Shouldn't it matter what happened before the "streak"?
    - is H to be expected after HHHTHHHHHHTTT?

This Paper

Environment: Canonical coin tossing with true bias θ^*

- First Model: Formalize LSN as Mean Reversion: sample mean stays close to θ^* **on path**
  - Gambler's Fallacy: $P(HHT) > P(HHH)$
  - Excessive Alternation: $P(HTHTTH) > P(HHHTTT)$
- Second Model: Formalize LSN as Local Mean Reversion: Mean Reversion on the last k-throws which we call a "segment"
  - Nests Rabin (2002)

This Paper

- Derive General Learning implications for LSN
  - Mean Reversion agents NEVER rule out the true bias
  - Local Mean Reversion agents may learn the wrong parameter (Rabin (2002))
- Evolutionary Survival:
  - Mean Reversion agents will survive an evolutionary race against IID agents
- Use sample/segment mean to define a streak
  - Formalize Streak Aversion through an Axiom
  - Local Mean Reversion implies Streak Aversion

1. Motivation ✓
2. Related Literature
3. Axiomatization
    - Mean Reversion
    - Local Mean Reversion
4. Bayesian Learning
5. Evolutionary Race
6. Streak Aversion and Other Evidence

Related literature

Economics

- Rabin (2002) “Inference by Believers in the Law of Small Numbers”
- Rabin and Vayanos (2009) “The Gambler’s and Hot-Hand Fallacies: Theory and Applications”
- Benjamin, Rabin and Raymond (2016) “A Model of Nonbelief in the Law of Large Numbers”

Psychology

- Rapoport and Budescu (1997) “Randomization in individual choice behavior”
- Bar Hillel and Wagenaar (1991) “The perception of randomness”

Canonical coin tossing environment

Primitives:

- Possible realizations of toss $\Omega = \{0, 1\}$
  - Also write $\Omega = \{T, H\}$
  - Can be extended to finite set of outcomes
- The space of realizations of $n \leq \infty$ tosses is Ω^n

Beliefs

- $(P^n)_{n=1}^{\infty}$ where $P^n \in \Delta(\Omega^n)$ for each $n < \infty$
- Assume they have full support (exposition)

Preliminaries

Notation:

- Sample mean (number of heads) of $x \in \Omega^n$

$$\overline{x}^n := \frac{\sum_{i=1}^n x_i}{n}$$

- Distance between the sample mean of x^n and the bias θ^*,

$$d(x^n) = |\overline{x}^n - \theta^*|,$$

Interpret coin tosses as objectively i.i.d.

$$Q(x^n) = \prod_{i=1}^n \theta^{*x_i}(1-\theta^*)^{1-x_i}$$

- Relevance: Beliefs are misspecified

Axioms

We impose three axioms on $(P^n)_{n=1}^{\infty}$

- Mean Reversion axiom is inspired by KT quotes applied to the Law of Large Numbers

Intuitively

- LLN: In large samples, the sample mean concentrates near the population mean
- LSN: sample mean concentrates near population in **every** small sample

Therefore, the sample mean should evolve close to the mean

Axioms

Intuition

	$HTHTTH$	$HHHTTT$
$n=1$	1	1
$n=2$	$\frac{1}{2}$	1
$n=3$	$\frac{2}{3}$	1
$n=4$	$\frac{1}{2}$	$\frac{3}{4}$
$n=5$	$\frac{2}{5}$	$\frac{3}{5}$
$n=6$	$\frac{1}{2}$	$\frac{1}{2}$

Weak Mean Reversion For any N and $x, y \in \Omega^\infty$ such that $\overline{x}^{N-1} = \overline{y}^{N-1}$,

$$d(x^n) \leq d(y^n) \text{ for all } n \leq N \implies P^N(x^N) \geq P^N(y^N)$$

Consistency Axioms

Applications typically require a bit more structure:

Marginal Consistency For any n and any event $A^n \subset \Omega^n$,

$$P^n(A^n) = P^{n+1}(A^n \Omega_{n+1})$$

- Ensures there exists $P^\infty \in \Delta(\Omega^\infty)$ such that $P^n(A) = P(A \times \Omega \times \Omega \times ...)$
- Allows us to interpret the family of beliefs as a single belief
- Need it to answer: Are anomalous beliefs about randomness consistent with a single belief?

Consistency Axioms

MR Independence For any n and $x^n, y^n \in \Omega^n$ and any $x_{n+1} \in \Omega$,

$$\overline{x}^n = \overline{y}^n \implies \frac{P^{n+1}(x^n x_{n+1})}{P^n(x^n)} = \frac{P^{n+1}(y^n x_{n+1})}{P^n(y^n)}$$

- Can be interpreted as follows: Conditional probability of x_{n+1} only depends on $\overline{x}^n$
- Relates P^{n+1} with P^n

Main Result

Theorem

A family of full support beliefs $\{P^n\}_{n=1}^{\infty}$ satisfies Weak Mean Reversion, MR Independence and Marginal Consistency iff for every $i \geq 1$, there exists $g^i : [0,1]^2 \to (0,1]$ that is weakly decreasing in its first argument, s.t. for any n and $x^n \in \Omega^n$,

$$P^n(x^n) = \prod_{i=1}^{n} (\theta_{i,\overline{x}^{i-1}})^{x_i} (1 - \theta_{i,\overline{x}^{i-1}})^{1-x_i}$$

where

$$\theta_{i,\overline{x}^{i-1}} = g^i(d(x^{i-1}1), \overline{x}^{i-1})$$
$$1 - \theta_{i,\overline{x}^{i-1}} = g^i(d(x^{i-1}0), \overline{x}^{i-1}).$$

Interpretation

Consider the case in which $(P^n)_{n=1}^{\infty}$ is the product of independent Bernoulli measures:

$$P^n(x^n) = \prod_{i=1}^{n} (\theta_i)^{x_i} (1-\theta_i)^{1-x_i},$$

- Over here, the bias changes with the number of throws, but is independent of the number of heads in the previous throws

$$P^n(x^n) = \prod_{i=1}^{n} (\theta_{i,\overline{x}^{i-1}})^{x_i} (1-\theta_{i,\overline{x}^{i-1}})^{1-x_i},$$

- Bias depends on the number of heads in the previous throws
- It is a **self-correcting** bias representation

An Equivalent Representation

Corollary

$(P^n)_{n=1}^{\infty}$ satisfy Weak MR and MR Independence if and only if for each i there exists $g^i : [0,1]^2 \to [0,1]$ that is weakly decreasing in its first argument such that for any n and $x^n \in \Omega^n$,

$$P^n(x^n) = \prod_{i=1}^{n} g^i(d(x^i), \overline{x}^{i-1}))$$

Furthermore, $(P^n)_{n=1}^{\infty}$ satisfies Marginal Consistency iff $g^i(d(x^{i-1}1), \overline{x}^{i-1})) + g^i(d(x^{i-1}0), \overline{x}^{i-1})) = 1$ for all x, i

Examples

Freedman Urn

- Urn starts with one head ball and one tail ball
- Each time we draw a head ball (resp. tail ball) we add m tail balls (resp. head ball)
- Satisfies MR, MC, MR Independence, and even Knowledge of Bias!

$$g^i(|x^i - \frac{1}{2}|, \overline{x}^{i-1}) = \begin{cases} \frac{(i-1)\overline{x}^{i-1}+1}{i} & |\overline{x}^i - \frac{1}{2}| = |\frac{(i-1)\overline{x}^i}{i} - \frac{1}{2}| \\ \frac{(i-1)(1-\overline{x}^{i-1})+1}{i} & |\frac{(i-1)\overline{x}^i+1}{i} - \frac{1}{2}| \end{cases}$$

Examples

One parameter specification

$$g^i(|\overline{x}^i-\theta^*|)=\frac{1}{Z^i}\left(\frac{1}{1+|\overline{x}^i-\theta^*|}\right)^{\lambda_i}$$

- Parameter is identified

$$\frac{P^i(x^{i-1}1)}{P^i(x^{i-1}0)}=(\frac{1+d(x^{i-1}0)}{1+d(x^{i-1}1)})^{\lambda i}$$

- λ_i parametrices the sensitivity of P^i with respect to d
- Empirical: Does sensitivity increase/decrease with I?

1. Motivation ✓
2. Related Literature ✓
3. Axiomatization ✓
   - Mean Reversion ✓
   - Local Mean Reversion
4. Bayesian Learning
5. Evolutionary Race
6. Streak Aversion and Other Evidence

Local Mean Reversion

LSN: "each segment of the [sequence] is highly representative of the "fairness" of the coin" (TK (1971))

- Can be interpreted as only a mean in a segment matters
- MR is defined on the segment $\{1, 2, ..., n\}$
- Literature violates Mean Reversion
- They satisfy it "Locally"

Generalize Mean Reversion to allow only apply to a recent "segment"

Segment

Define a Segment $W_n = \{k_n, ..., n\}$ where $k_n \leq n$

Segment Regularity For all n,

$$W_{n+1} \backslash \{n+1\} \subseteq W_n$$

- Segment regularity implies $k_n \leq k_{n+1}$

Let $\overline{x}^n(W_n) = \frac{\sum_{i \geq k_n} x_i}{|W_n|}$ and $d_{W_n}(x^n) = |\overline{x}^n(W_n) - \theta^*|$

- We can adapt our axioms to segments

Segment Axioms

Weak Local MR For any n and $x, y \in \Omega^n$ s.t. $\overline{x}^{n-1}(W_{n-1}) = \overline{y}^{n-1}(W_{n-1})$,

$$d_{W_i}(x^i) \leq d_{W_i}(y^i) \text{ for all } i \leq n \implies P^n(x) \geq P^n(y)$$

Local MR Independence For all n

$$\overline{x}^n(W_n) = \overline{y}^n(W_n) \implies \frac{P^{n+1}(x^n x_{n+1})}{P^n(x^n)} = \frac{P^{n+1}(y^n x_{n+1})}{P^n(y^n)}$$

Locally Self-Correcting Representation

Theorem

A family of full support beliefs $\{P^n\}_{n=1}^{\infty}$ satisfies Marginal Consistency, Weak Local MR, Local MR Independence, and Segment Regularity iff there exists a regular family of segments $\{W_n\}_{n=1}^{\infty}$ and for each $i \geq 1$ there exists a continuous function $g^i : [0,1]^2 \to (0,1]$ that is weakly decreasing in its first argument such that for any n and $x^n \in \Omega^n$,

$$P^n(x^n) = \prod_{i=1}^{n} (\theta_{i,\overline{x}^{i-1}(W_{i-1})})^{x_i} (1 - \theta_{i,\overline{x}^{i-1}(W_{i-1})})^{1-x_i}$$

where

$$\theta_{i,\overline{x}^{i-1}(W_{i-1})} = g^i(d_{W_i}(x^{i-1}1), \overline{x}^{i-1}(W_{i-1}))$$
$$1 - \theta_{i,\overline{x}^{i-1}(W_{i-1})} = g^i(d_{W_i}(x^{i-1}0), \overline{x}^{i-1}(W_{i-1}))$$

Discussion

How do you identify segments from behavior?

Definition

For any n, a *behavioral segment at n* is a set of contiguous indices $W_n = \{k_n, .., n\} \subseteq \{1, .., n\}$ containing n and satisfying: for all $x, y \in \Omega^\infty$,

$$\frac{P^n(x_1 \ldots x_{k_n-1} x_{k_n} \ldots x_{n-1} x_n)}{P^{n-1}(x_1 \ldots x_{k_n-1} x_{k_n} \ldots x_{n-1})} = \frac{P^n(y_1 \ldots y_{k_n-1} x_{k_n} \ldots x_{n-1} x_n)}{P^{n-1}(y_1 \ldots y_{k_n-1} x_{k_n} \ldots x_{n-1})}$$

- The smallest segment seems natural (experimental question)
- Axioms suggest an experimental design to test MR vs Local MR

1. Motivation ✓
2. Related Literature ✓
3. Axiomatization ✓
    - Mean Reversion ✓
    - Local Mean Reversion ✓
4. Bayesian Learning
5. Evolutionary Race
6. Streak Aversion and Other Evidence

Bayesian Inference

Primitives:

- $\Theta = \{\theta_1, ..., \theta_n\}$ parameter space
- $\mu \in \Delta(\Theta)$ prior
- $P_\theta^n(x^n)$ ex-ante probability of a sequence x^n conditional on the true bias being θ

$$P_\theta^n(x) = \prod_{i=1}^{n} g^i(d_\theta(x^i), \overline{x}^{i-1})$$

- **Only satisfies Weak MR and MR Independence**

Then, her ex-ante beliefs over sequences of length n is given by

$$P^n(x^n) = \sum_\theta P_\theta^n(x^n)\mu(\theta)$$

- $P^n(\theta|x^n)$ denote her Bayesian posterior after observing x^n:

$$P^n(\theta|x^n) = \frac{P_\theta^n(x^n)\mu(\theta)}{\sum_{\theta' \in \Theta} P_{\theta'}^n(x^n)\mu(\theta')}$$

Results

Theorem

Assume $\mu \in \Delta(\Theta)$ and each $P^n \in \Delta(\Omega^n)$ have full support. Then,

$$\liminf_n P^n(\theta^*|x^n) > 0 \ a.s\text{-}p^{\theta^*}$$

- Theorem establishes that regardless of whether there is convergence to a degenerate belief, the agent always places a non-vanishing probability on the true parameter
- Very different than the existing literature
  - Local MR nests Rabin (2002) $\implies$ Local MR can learn the wrong parameter
- What about actually learning the parameter?

Results

Theorem

Suppose $\mu \in \Delta(\Theta)$ *and each* $P^n \in \Delta(\Omega^n)$ *have full support, and that* g^i *is strictly decreasing in its first argument and continuous in its second for each* i

1. *If* $g^i \to c$ *uniformly faster than* $\frac{1}{n^2} \to 0$,[a] *where* $c > 0$ *is a constant function, then* $p^{\theta^*}(\lim_{n\to\infty} P^n(\theta^*|x^n) \in (0,1)) = 1$, *that is,*

$$0 < \lim_{n\to\infty} P^n(\theta^*|x^n) \neq 1 \text{ a.s.- } p^{\theta^*}$$

2. *If* $g^i = g$ *for all* $i > 1$, *then* $p^{\theta^*}(\lim_{n\to\infty} P^n(\theta^*|x^n) \to 1) = 1$, *that is,*

$$\lim_{n\to\infty} P^n(\theta^*|x^n) \to 1 \text{ a.s.- } p^{\theta^*}$$

[a] There exists N such that for all $n > N$, $|g^n(a,\theta) - c| < \frac{1}{n^2}$ for all a, θ in the support of g^n.

1. Motivation ✓
2. Related Literature ✓
3. Axiomatization ✓
   - Mean Reversion ✓
   - Local Mean Reversion ✓
4. Bayesian Learning ✓
5. Evolutionary Race
6. Streak Aversion and Other Evidence

Evolutionary Race

Consider two populations of agents. An "IID agent" has an accurate perception of i.i.d. sequences. An "LSN" agent follows the ϵ^*-specification for g:

$$g^i_{\epsilon^*}(|\overline{x}^i - \theta^*|) = \begin{cases} \frac{\alpha}{Z^i} & |\overline{x}^i - \theta^*| \leq \epsilon^* \\ \frac{1-\alpha}{Z^i} & otherwise \end{cases}$$

- Continuum of "safe" hunting grounds: small reward $r = 1$ (a "rabbit")
- Continuum of "risky" hunting grounds: a fraction $\theta \in \Theta = \{\hat{\theta}, 1 - \hat{\theta}\}$, where $\hat{\theta} > \frac{1}{2}$, contains a large reward, $r = 2$ (a "deer"), and the remaining fraction contain no reward, $r = 0$
- The fraction θ is unknown and both have a common prior about it

Information and Payoffs

In every period, one risky ground is randomly chosen and publicly sampled by both LSN and IID agent types

- The agents update their beliefs about θ based on whether a deer is sighted in the sampling hunting ground
- Let $x_i = 0$ (resp. $x_i = 1$) denote that the deer was not present (resp. present), in which case we can write the reward in period i as $r = 2x_i$
- Each type $A = IID, LSN$ determines the fraction of its population, $k^A_{x^i} \in [0, 1]$, that hunts in the risky grounds in period i conditional on having observed x^i signals, which the remainder fraction $1 - k^A_{x^i}$ hunting in the safe ground

Population

Letting Λ_{i-1}^A denote the population of type A at the start of period i. The total reward per capita received by type A is

$$c_{x^i}^A := \frac{R_{x^i}^A}{\Lambda_{i-1}^A} = k_{x^i}^A(2\theta) + (1 - k_{x^i}^A)$$

- Both maximize expected utility using a common strictly increasing strictly concave utility index u to determine the optimal $(k_{x^i}^A, 1 - k_{x^i}^A)$ based on history of deer sightings x^i from the sampled hunting ground
- The population of type A agents grows by a factor of $\lambda^{c_{x^i}^A}$ in period i, where $\lambda > 1$
- Object if interest

$$\prod_{i=1}^{n} \frac{\lambda^{c_{x^i}^{LSN}}}{\lambda^{c_{x^i}^{IID}}}$$

Proposition

The following hold for LSN agents with $0 < \epsilon^* < 2\hat{\theta} - 1$
(i) The LSN agents are eventually more confident about the true parameter, a.s.:

$$P^n_{LSN}(\theta|x^n) \geq P^n_{IID}(\theta|x^n) \text{ for all sufficiently large } n \in \mathbb{N} \text{ a.s.-}p^\theta$$

(ii) The population of LSN agents never vanishes, a.s.:

$$\lim \inf_{n\to\infty} \prod_{i=1}^{n} \frac{\lambda^{c^{LSN}_{x^i}}}{\lambda^{c^{IID}_{x^i}}} > 0 \quad a.s\text{-}p^\theta$$

(iii) Assume $\hat{\theta} \geq \frac{3}{4}$ *and* $\epsilon = \hat{\theta} - \frac{1}{2}$*. Then the population of LSN dominates, with probability greater than* $\frac{1}{2}$*:*

$$P(\lim \inf_{n\to\infty} \prod_{i=1}^{n} \frac{\lambda^{c^{LSN}_{x^i}}}{\lambda^{c^{IID}_{x^i}}} = \infty) > \frac{1}{2}$$

1. Motivation ✓
2. Related Literature ✓
3. Axiomatization ✓
   - Mean Reversion ✓
   - Local Mean Reversion ✓
4. Bayesian Learning ✓
5. Evolutionary Race ✓
6. Streak Aversion and Other Evidence

Streak Aversion

Intuitively, Local MR can explain

- Gambler's fallacy: $P(HHT) > P(HHH)$
- Excessive Alternation: $P(HTHTTH) > P(HHHTTT)$

These terms are not formal

- View GF and EA as manifestations of Streak Aversion
- Need to define first what is a streak

Discussion

We define a streak as too many heads in a segment

- For any sequence $x \in \Omega^\infty$ we say it has a streak at x^{n-1} if $\frac{\sum_{i\in W_n\setminus\{n+1\}} x_i}{|W_n|} \geq \theta^*$
- Example: If a segment has length 3
  - THH exhibits a streak at toss 3
  - HHT does not exhibit a streak at toss 3

Streak Aversion For any n, x^n and outcome $\omega \in \{0, 1\}$

$$x^n \text{ contains a streak of } \omega \text{ at outcome } n-1$$
$$\implies P^n(x^{n-1}(1-\omega)) \geq P^n(x^{n-1}\omega)$$

Trivially

- Local MR implies Streak Aversion
- Local MR "implies" GF and EA

Other Evidence on Beliefs about Randomness

The model has issues with other evidence on beliefs about randomness:

- Aversion to patterns: $P(HTHTTH) > P(HTHTHT)$
  - Tune (1964), Wagenaar (1970), Kahneman and Tversky (1972)
- Sample-Size Neglect/ Non Belief in LLN: $P(\overline{x}^n \geq \kappa)$ insensitive to n
  - Evidence: Kahneman and Tversky (1972), Benjamin, Moore and Rabin (2018)

Discussion

Aversion to Patterns

- Inconsistent with Mean Reversion: HTHTHTHT is one of the most likely sequence according to MR
- Suggest another core property of beliefs: Order Aversion (future research)

Sample-Size Neglect/NBLLN

- Can be at odds with Mean Reversion
    - Freedman Urns generate LLN
    - Not every LSN model generates LLN

Discussion

Theorem

If $(P^n)_{n=1}^{\infty}$ satisfies Weak Mean Reversion then the Law of Large Numbers is not implied, that is, it may not be the case that for all

$$\lim_{n\to\infty} P^n(|x^n - \theta^*| > \epsilon) = 0$$

if the limit exists

- The proof constructs an example of a Weak Mean Reversion model that satisfies Exchangeability and Marginal Consistency, but generically fails MR-Independence
- Limiting frequencies are a non-degenerate random variable

Hot Hand Effect

Subjects sometimes expect a streak to continue

- Uncertainty/Inference about the parameter can accommodate such behavior
- Agent expects the mean to remain close to bias, hence finds a streak of heads "too unlikely" for any parameter that is not close to 1
  - Intuition familiar from psychology (Gilovich, Vallone and Tversky (1985), Rabin (2002), Rabin and Vayanos (2010))
- Can be demonstrated with Local MR in inference setting

St. Petersburg Paradox

Recall the famous St. Petersburg Paradox

- Lottery pays 2^n if first head occurs at toss n
- Expected value is $\sum_{n=1}^{\infty} \frac{1}{2^n} 2^n = \infty$ but certainty equivalent is finite
- MR can accommodate that $P^n(T...TH) \to 0$ faster than $\frac{1}{2^n} \to 0$

1. Motivation ✓
2. Related Literature ✓
3. Axiomatization ✓
   - Mean Reversion ✓
   - Local Mean Reversion ✓
4. Bayesian Learning ✓
5. Evolutionary Race ✓
6. Streak Aversion and Other Evidence ✓

Concluding Remarks

Today

- Evidence: Gambler's Fallacy, Excessive Alternation
- Hypothesis: beliefs are sensitive to the evolution of sample mean, not contiguous streak per see
- Formulate two nested theories of Streak Aversion
- Local Mean Reversion captures the spirit of the literature
- Mean Reversion has desirable properties in learning and evolutionary settings

Concluding Remarks

Future

- Deeper exploration of Streak Aversion:
  - Does the Gambler's Fallacy depend on pre-streak outcomes? Would subjects pay to see what happened before a streak?
- Building blocks of beliefs about randomness?
  - Local Mean Reversion explains evidence on Streak Aversion
  - "Order Aversion" is needed to explain disbelief in patterns
  - CS/Econ? Model as pseudorandomness?

Thank You!