Monotonicity vs Periodicity in the Law of Large Numbers

Dingchen Sha[2], Gabby Digan[2], Jack Shao[1], Nathan Sell[1], Tony Ruan[3]

[1]Department of Mathematics [2]Department of Statistics [3]Department of Computer Science

Hypothetical Situation

You are playing bean bag toss and you have an **80%** probability of successfully making a toss. To win the game, you need to make at least **75%** of your tosses. ***What number of tosses [1, 12] will maximize your probability to achieve at least 75% of your tosses?***

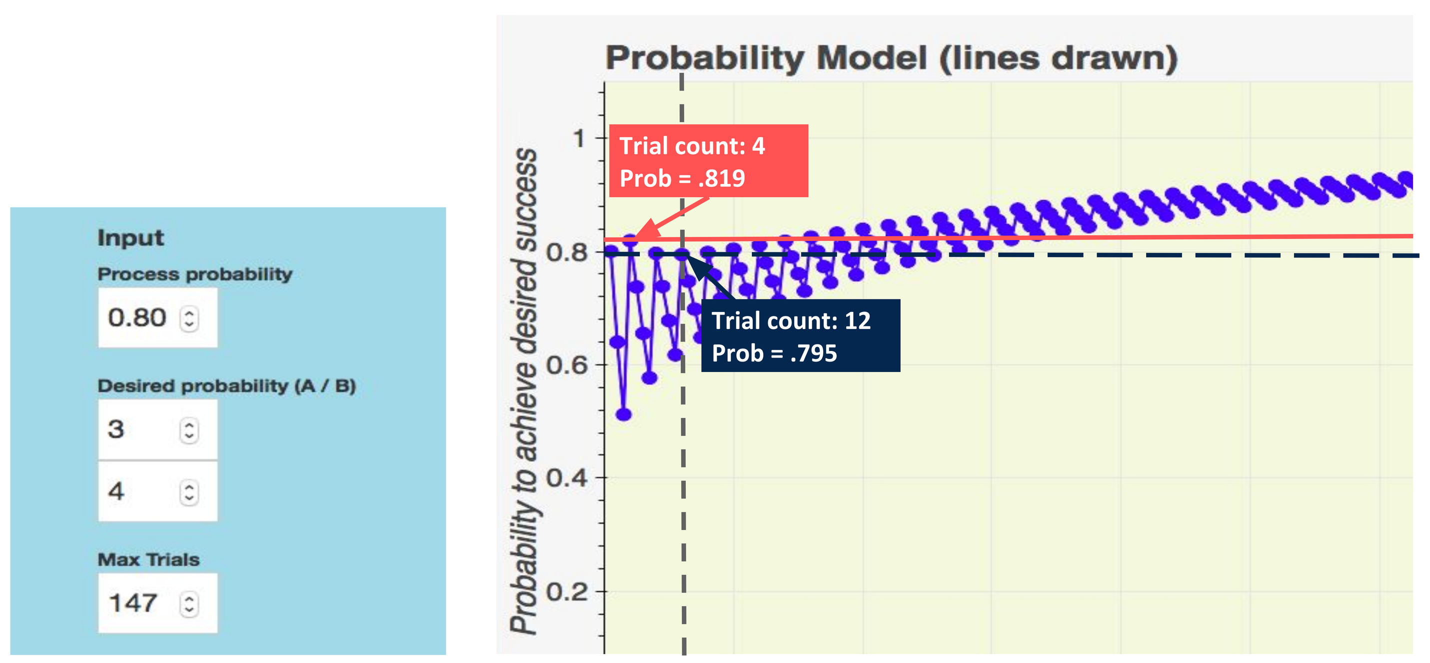

Figure 1: Probability of attaining at least a 75% proportion of successes for any n trials when the probability of success for any given trial is 0.80

Motivating Thoughts

Suppose we have a process that succeeds X% of the time and all experiments are independent. Let's call X the process probability. We want to achieve at least Y% success when running the process over a set number of trials n. Let's call Y the desired probability, and assume Y is rational.

What value of n will give us the greatest probability to achieve *at least* our desired probability?

Background

Definition 1.1. *Let $p, q \in (0,1)$ $n \geq 1$, and we define*

$$A(p,q,n) := \mathbb{P}(\mathrm{Bin}(n,p) \geq qn) = \sum_{k \geq qn} \mathbb{P}(\mathrm{Bin}(n,p) = k) = \sum_{k \geq qn} \binom{n}{k} p^k (1-p)^{n-k}.$$

Probability of getting some portion of your independent Bernoulli trials as successes

- **n** = Number of trials
- **p** = Probability of a success for a given trial, (Process Probability)
- **q** = Desired proportion of success, (Desired Proportion) -> (A/B)

Theorem (The Law of Large Numbers). *Let $X_1, X_2, ..., X_N$ be $N \in \mathbb{N}$ IID random variables, $\bar{X} = \frac{S_N}{N} = \frac{X_1+X_2+...+X_N}{N}$ be the average of these random variables, and $\mu = E(X_i)$ for any $i \in [1, N]$. Then for all $\mathbb{R} \ni \epsilon > 0$ $P(|\bar{X} - \mu| < \epsilon) = 1$.*

As we ***increase*** the sample size of an experiment, the average of of the results will be ***closer*** to the expected value

Monotonic Conjecture

Conjecture 3.1. *Let $q = \frac{\alpha}{\beta}$ where $\alpha, \beta \in \mathbb{N}$, and let $k \in \mathbb{N}$. Then there exists an $\epsilon \in \mathbb{R}$, with $\epsilon > 0$, such that whenever $|p - q| \geq \epsilon$, $A(p, q = \frac{\alpha}{\beta}, k\beta)$ is monotone.*

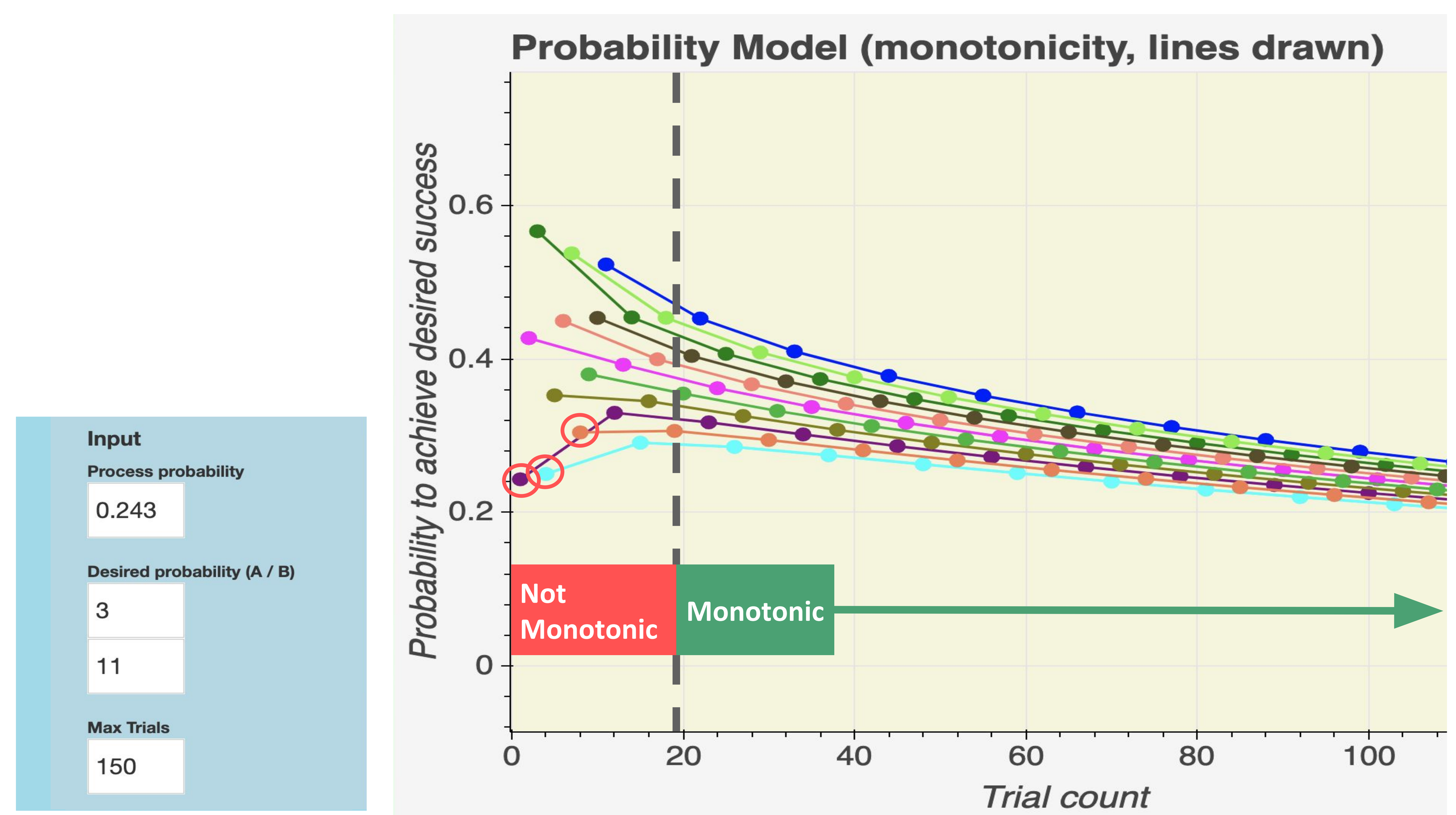

Figure 2: Graph where the points not adhering to monotonic behavior are circled in red and the dashed line indicates where proper monotonic behavior begins

For all trial counts n in the same congruence class mod B, the function A(p,q,n) exhibits monotonic behavior **for a trial count n that is greater than or equal to k, where k is a sufficiently large positive integer.**

Overall Trends & Maximization

On the local level, in trial counts between [wB, xB], where w and x are nonzero and 1 < w < x...

p < q

- Process probability less than our desired proportion
- Resulting probabilities converge to 0
- Overall trend: decreasing
- **wB** results as the maximum

p = q

- Process probability equal to our desired proportion
- Resulting probabilities converge to 0.5
- Overall trend: increasing below 0.5 and decreasing above 0.5
- **wB** results as the maximum

p > q

- Process probability greater than our desired proportion
- Resulting probabilities converge to 1
- Overall trend is increasing
- **xB** results as the maximum

Periodic Behavior

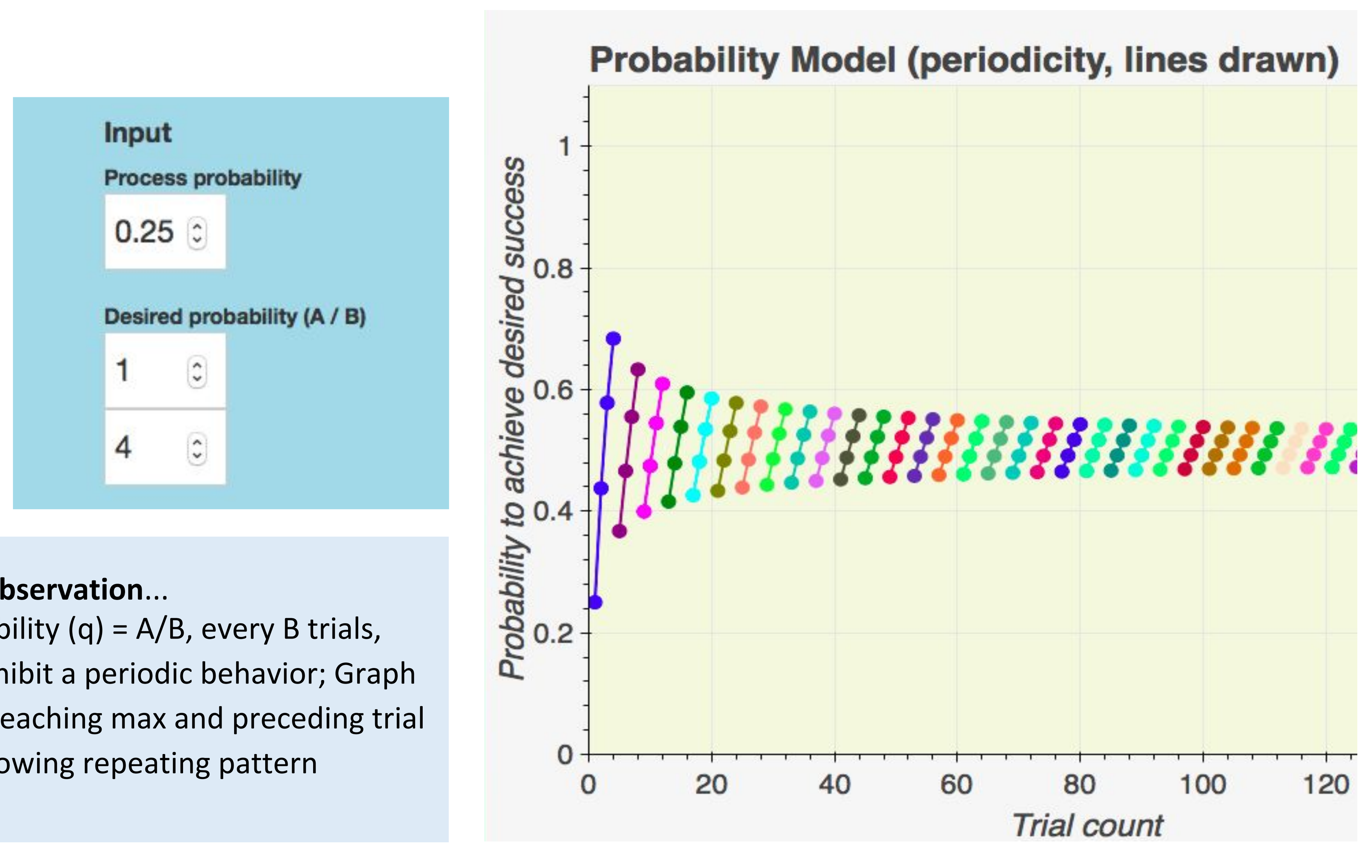

Figure 3: General Observation...
With desired probability (q) = A/B, every B trials, the probabilities exhibit a periodic behavior; Graph with trial count Bk reaching max and preceding trial count in B trials following repeating pattern

Definition 1.1. *Let $p, q \in (0,1)$, $n, \alpha, \beta \in \mathbf{N}$, $q = \frac{\alpha}{\beta}$ and let $A(p,q,n)$ be defined as usual. Then we define $n_k = \left\lfloor \frac{k}{q} \right\rfloor = \left\lfloor \frac{k}{\frac{\alpha}{\beta}} \right\rfloor = \left\lfloor \frac{\beta k}{\alpha} \right\rfloor$, with $n_0 = 0$, for $k \in \mathbf{N} \cup 0$.*

Taking consecutive differences of n_k, provides us with monotone subsets of A(p,q,n) Explaining the number of points in a subsection of a given group of B trials

$$\text{Let } q = \frac{3}{11} \text{ so that } n_k = \left\lfloor \frac{\beta k}{\alpha} \right\rfloor = \left\lfloor \frac{11k}{3} \right\rfloor.$$

$$n_1 = \lfloor 11/3 \rfloor = 3, n_2 = \lfloor 22/3 \rfloor = 7, n_3 = \lfloor 33/3 \rfloor = 11$$

With a longer sequence being

$$3, 7, 11, 14, 18, 22, 25, 29, 33, 36, 40, 44, 47, 51, 55, 58, 62, 66, 69, 73, ...$$

Accounting for $n_0 = 0$ at the beginning of the sequence and taking the difference of consecutive terms

$$d_1 = 3, d_2 = 4, d_3 = 4, d_4 = 3, d_5 = 4, d_6 = 4, ...$$

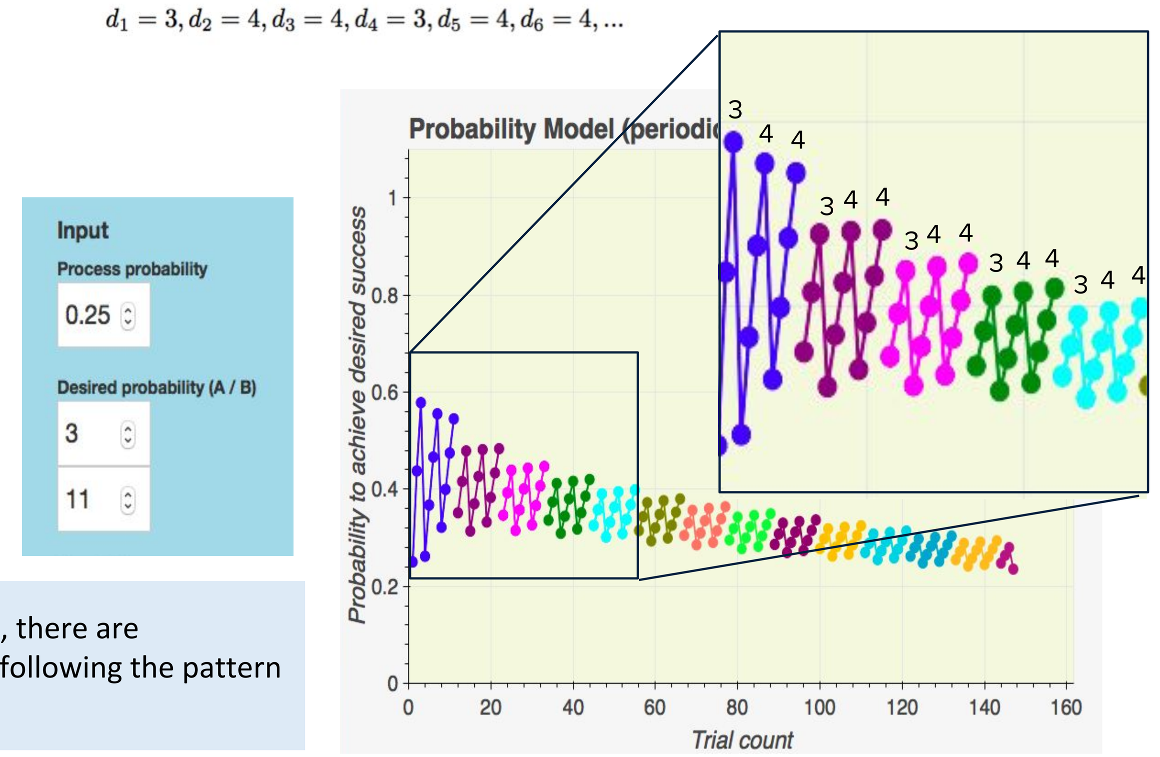

Figure 4: Just like we expected, there are subsections in each 11 groups following the pattern 3,4,4

Key Theorem

Theorem. *Let $p > q = \frac{\alpha}{\beta}$ for $\alpha, \beta \in \mathbb{N}$, and define $n_k = \left\lfloor \frac{k}{q} \right\rfloor$. Then*

$$A(p,q,n+1) < A(p,q,n) \implies n = n_k \ \exists\, k \in \mathbb{N}$$

Defining n_k in this manner provides us with what we can call increment points -- points where we are guaranteed that, for trial counts that aren't increment points, i.e. $n \neq n_k$ for all k, we have $\lceil q(n+1) \rceil = \lceil qn \rceil$