Macroscopic patterns emerge from random individual behaviours

Linglong Yuan
University of Liverpool

AIMS-UoL Joint Postgraduate Conference

16.06.2023

Plan

Introduction

Example: Central Limit Theorem and large deviations

Example: Kingman's model of selection and mutation

Macroscopic phenomena in complex systems

How do the following phenomena happen?

- Water becomes ice at degree zero
- A magnet loses magnetism above certain temperature
- Free market is more efficient in productivity
- Richer gets richer
- Species extinction
- Covid-19 spreads exponentially at outbreak

Why stochastic models?

In stochastic models, we assume

- there are many individuals in a population
- the population is in a certain environment with constant or evolving characteristics
- individuals interact randomly with each other under the constraints from the environment
- although we do not dictate how each individual should behave (it is completely random), macroscopic/collective phenomena will appear

What is the fate of the population given the random behaviours of individuals?

Plan

Central Limit Theorem

Theorem

Let $X_1, X_2, \ldots$ be independent and identically distributed (i.i.d) random variables with mean μ and variance σ^2. Let $S_n = \sum_{i=1}^{n} X_i$. Then

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \text{ is approximately standard normal as } n \to \infty.$$

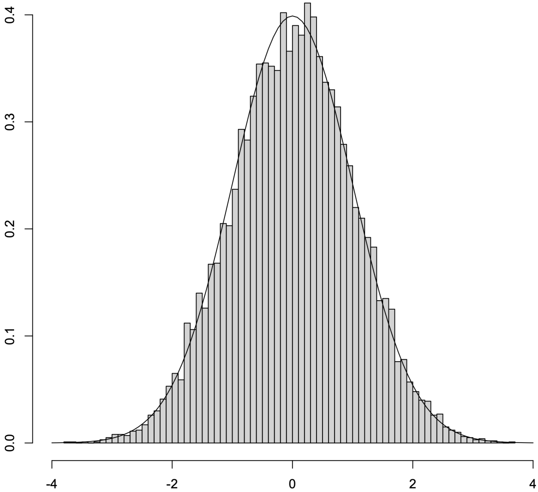


Figure: Histogram of a sample of data $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ vs. the pdf of a standard normal distribution

Remark Although each random variable behaves independently of any other, they collectively fall in the attraction of standard normal

One step further: large deviations

Theorem (Nagaev, 1979)

Let $X_1, X_2, \ldots$ be independent and identically distributed (i.i.d) random variables with mean μ and variance σ^2. Let $S_n = \sum_{i=1}^n X_i$. Assume that

- *the tail probability function $\overline{F}(t) := \mathbb{P}(X_1 \geq t)$ is regularly varying with index $-\beta < -2$*
- *there exists $\delta > 0$ such that $\mathbb{E}[|X_1|^{2+\delta}] < \infty$*

Then for any $x_n \geq \sqrt{n}$,

$$\mathbb{P}(S_n - \mu n \geq x_n) \sim \overline{\Phi}(\tfrac{x_n}{\sigma\sqrt{n}}) + n\overline{F}(x_n), \quad n \to \infty$$

where $\overline{\Phi}$ is the tail probability function of the standard normal distribution.

Two scenarios

Let $M_n = \max\{X_1, X_2, \ldots, X_n\}$. We can write

$$\mathbb{P}(S_n - \mu n \geq x_n) = \mathbb{P}(S_n - \mu n \geq x_n, M_n < x_n) + \mathbb{P}(S_n - \mu n \geq x_n, M_n \geq x_n)$$

Then,

- **normal scenario:** $\mathbb{P}(S_n - \mu n \geq x_n, M_n < x_n) \sim \overline{\Phi}(\frac{x_n}{\sigma\sqrt{n}})$
- **one-big-jump scenario:**

$$\begin{aligned}\mathbb{P}(S_n - \mu n \geq x_n, M_n \geq x_n) &\sim \mathbb{P}(M_n \geq x_n) \\ &\sim n\overline{F}(x_n)\end{aligned}$$

What happens in the one-big-jump scenario?

Proposition 1 (Berger, Birkner, Y, 23)

Let $(x_n)_{n\geq 1}$ *be a sequence satisfying*

- $\lim_{n\to\infty} n\overline{F}(x_n) = 0$,
- $\overline{F}(x_n) > 0$ *for all* n.

Then we have

What happens in the one-big-jump scenario?

Proposition 1 (Berger, Birkner, Y, 23)

Let $(x_n)_{n\geq 1}$ *be a sequence satisfying*

- $\lim_{n\to\infty} n\overline{F}(x_n) = 0$,
- $\overline{F}(x_n) > 0$ *for all* n.

Then we have

$$\lim_{n\to\infty} d_{\mathrm{TV}}\Big(\mathscr{L}\left(R(\xi_1,\ldots,\xi_n)\,\middle|\, M_n \geq x_n\right),\ \left(\mathscr{L}(\xi)\right)^{\otimes(n-1)}\Big) = 0$$

where d_{TV} *=total variation distance;* $R(\cdots)$ *is to remove the largest element.*

What happens in the one-big-jump scenario?

Proposition 1 (Berger, Birkner, Y, 23)

Let $(x_n)_{n\geq 1}$ *be a sequence satisfying*

- $\lim_{n\to\infty} n\overline{F}(x_n) = 0$,
- $\overline{F}(x_n) > 0$ *for all* n.

Then we have

$$\lim_{n\to\infty} d_{\mathrm{TV}}\left(\mathscr{L}\left(R(\xi_1,\ldots,\xi_n)\,\middle|\, M_n \geq x_n\right),\ \left(\mathscr{L}(\xi)\right)^{\otimes(n-1)}\right) = 0$$

where d_{TV} *=total variation distance;* $R(\cdots)$ *is to remove the largest element.*

Remark 2

Note that this result requires no structural conditions on the distribution of the ξ's.

The phase transition

Depending on how large x_n is,

- if $\overline{\Phi}(\frac{x_n}{\sigma\sqrt{n}}) \sim n\overline{F}(x_n)$:

  with probability $\frac{\overline{\Phi}(\frac{x_n}{\sigma\sqrt{n}})}{\mathbb{P}(S_n - \mu n \geq x_n)}$, normal scenario occurs

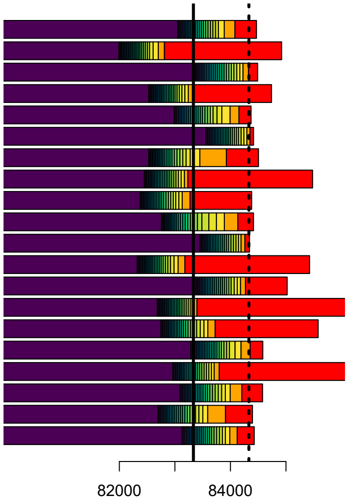
  with probability $\frac{n\overline{F}(x_n)}{\mathbb{P}(S_n - \mu n \geq x_n)}$, one-big-jump scenario occurs

- if x_n is much smaller, only normal scenario occurs

- if x_n is much larger, only one-big-jump scenario occurs

Simulation: $\overline{F}(x) = x^{-2.5}, x \geq 1; \quad n = 50000$

Not centralised; total length is the sum; length of the red segment is the largest summand; x_n is the distance between vertical lines



Plan

Kingman's model (1978)

Kingman considers an infinite population with discrete generations, and fitness values of an indivdual within $[0, 1]$.

Selection: At each generation, the number of offspring of an individual in the next generation depends on its fitness. If it is fitter, then more offspring will be produced.

Mutation:

- For each child, with probability b, it is mutated, and its fitness will be sampled randomly from a common distribution Q
- with probability $1 - b$, it inherits the fitness of its parent

Maths formulation

Kingman's model uses probability measures to describe the evolution of the population.

It has three parameters (P_0, Q, b) and the dynamics is defined as:

$$P_{n+1}(dx) = (1-b)\underbrace{\frac{xP_n(dx)}{\int_0^1 yP_n(dy)}}_{\text{selection}} + b\underbrace{Q(dx)}_{\text{mutation}}, \quad n \geq 0. \qquad (1)$$

- Q, P_n are probability measures on $[0, 1]$,
- $b \in (0, 1)$ is deterministic.

What questions to ask?

- Will (P_n) converge?
- What does the limit of P_n look like?
- How does the limit of P_n depend on the three parameters (P_0, Q, b)?

Kingman (1978): convergence and condensation

Define $\zeta := 1 - b \int \frac{Q(dy)}{1-y}$.

Theorem

(1)-Mutation dominates Selection:
If $\zeta \leq 0$, then $(P_n)_{n\geq 0}$ converges strongly to

$$\frac{b\theta Q(dx)}{\theta - (1-b)x},$$

with θ being the unique solution of $\int \frac{b\theta Q(dx)}{\theta-(1-b)x} = 1$.

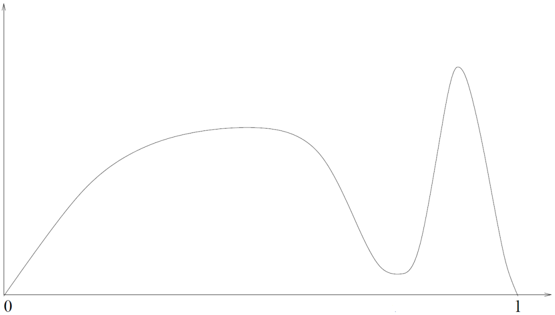
(2)-Selection dominates Mutation:
If $\zeta > 0$, then $(P_n)_{n\geq 0}$ converges weakly to

$$\frac{bQ(dx)}{1-x} + \zeta\delta_1(dx),$$

here $\delta_1(dx)$ is the Dirac measure at 1. Condensation occurs.

Regimes

Meritocracy or Aristocracy: if condensation will occur
Democracy: if condensation will not occur



A random model

In the original model, the mutation probability b is fixed for all generations.

If we say the mutation probability for generation n is b_n such that (b_n) is an i.i.d. sequence with

$$\mathbb{E}[b_n] = b, \quad \forall n \geq 1.$$

How will such noise affect the condensate size?

In other words, if you want to reduce or increase the condensate size, would you add the noise or not?

Comparison: main result

Theorem (Y, 2020,2022)

The sequence (P_n) in the random model will converge to a limit.

The limit will less likely have a condensate, and if it does, the condensate size will be smaller than that from the Kingman's model.

References

- Quentin Berger, Matthias Birkner and Y. Collective vs. individual behaviour for sums of i.i.d. random variables: appearance of the one-big-jump phenomenon. ArXiv (2023)
- Sergey V. Nagaev. Large deviations of sums of independent random variables. The Annals of Probability (1979).
- Linglong Yuan. Kingman's model with random mutation probabilities: convergence and condensation II. Journal of Statistical Physics (2020).
- Linglong Yuan. Kingman's model with random mutation probabilities: convergence and condensation I. Advances in Applied Probability (2022).

THANK YOU