Some probabilistic toy models for biological evolution

Iddo Ben-Ari, University of Connecticut

Σ Σeμiνaρ, October 2017

~~Some~~A probabilistic toy model~~s~~ for biological evolution

Iddo Ben-Ari, University of Connecticut

Σ Σeμiνaρ, October 2017

Law of Large Numbers

If you repeatedly sample from the UCONN student population, each time independently of the past, then the proportion of graduate students you have sampled will tend to proportion of graduate students in the population.

More generally,

If you repeatedly sample from the UCONN student population, each time independently of the past and calculate the average GPA of the students you have sampled, then these averages will converge to the average population GPA.

The corresponding mathematical statement is called the Law of Large Numbers:

Theorem 1

- *Suppose that $X_1, X_2, \ldots$ are IID and $E[|X_1|] < \infty$.*

Then

$$\underbrace{\frac{X_1 + \cdots + X_n}{n}}_{\text{empirical average}} \to \underbrace{E[X_1]}_{\text{population average}}, \text{ a.s. and in } L^1.$$

Comments:

- This allows to estimate population averages through "small samples".

Glivenko-Cantelli

If you repeatedly sample from the UCONN student population, each time independently of the past and keep track of the GPA's of students you have sampled, then the proportion of students whose GPA is at most f will tend to the actual corresponding proportions among entire student population.

$$\underbrace{\frac{\#\text{ students sampled with GPA } \le f}{\#\text{ students sampled}}}_{\text{empirical distribution function}} \to \underbrace{\text{proportion students with GPA } \le f}_{\text{actual distribution function}}.$$

The corresponding mathematical statement is known as the Glivenko-Cantelli lemma:

Theorem 2

- *Suppose* $X_1, X_2, \ldots$ *are IID with distribution function* $F(f) = P(X_1 \le f)$.
- *Define the empirical distribution function* $\hat{F}_n(f) = \dfrac{\#\{X_j : j \le n, X_j \le f\}}{n}$.

Then

$$\hat{F}_n \to F, \text{ uniformly, a.s.}$$

Comments:

- This is a corollary to the Law of Large Numbers.
- (in some cases gives) Approximation of F with relatively small number of samples.

A toy model to biological evolution

Population size dynamics Population

- increases by 1 with probability p, independently of past.
- decreases by 1 (if possible) with probability $1 - p$, independently of past.

Fitness dynamics

- When population increases, the new member is assigned a fitness value $\mathrm{U}[0, 1]$, independently of everything else.
- When population decreases, the least fit individual is eliminated from the population.

Question. How would population landscape look in the long run ?

What does model do ?

Here's a simulation.

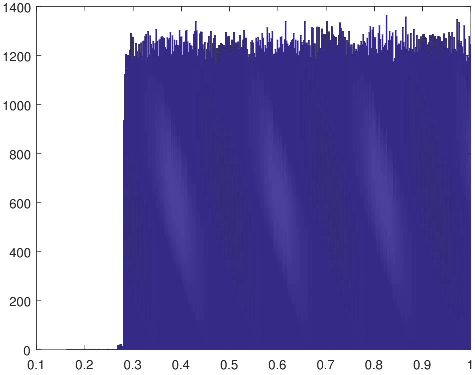


Figure: # of members for fitness values, after $T = 10^6$ iterations. Here $p = 0.8$ ($f_c = 0.25$)

What does model do ?

Explanation.

Let's look at the increments in population size of those with fitness below f:

- When population increases goes up by 1 with probability f.
  $\Rightarrow +1$ with probability pf.
- When population decreases goes down by 1 with probability 1, or stays put if population is zero.
  $\Rightarrow -1$ with probability $(1-p)$.

Now

- Write $X_t(f)$ for the increment between time t and time $t-1$.
- Forget that "stays put" part.
- Then adding the increments, LLN tells us

$$\text{Population below } f \text{ at time } T = \sum_{1 \leq j \leq T} X_t(f) \sim TE[X_1] = T(pf - (1-p)).$$

What does model do ?

We've observed:

- Population below f at time $T = \sum_{1 \leq j \leq T} X_t(f) \sim TE[X_1] = T(pf - (1-p))$.
- Let $f_c = \frac{1-p}{p}$.

From this, we observe that if

- $f > f_c$, # those with fitness $\leq f$ goes to $+\infty$, and eventually will never gets back to 0.
- if $f < f_c$, # those with fitness $\leq f$ goes to $-\infty$.

So going back to original model:

- After some (random, but finite) time all those with fitness $> f_c$ will live forever.
- The part of population below f_c will be repeatedly wiped out.

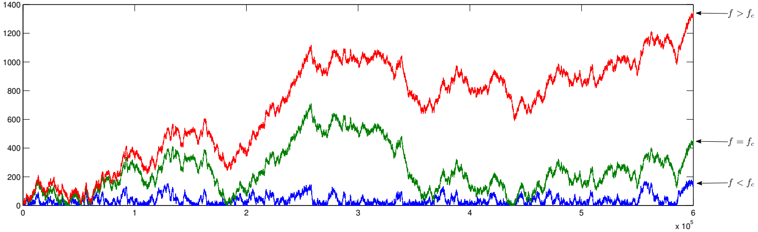


Figure: # members with fitnesses below certain values

G-K for the model

The above considerations with a little bit of math, give:

Theorem 3

- *Suppose* $p > \frac{1}{2}$, *and let* $f_c = \frac{1-p}{p} \in [0,1)$.
- *Define the empirical fitness distribution function*

$$\hat{F}_T(f) = \frac{\#\ \textit{members at time } T \textit{ with fitness} \leq f}{\#\ \textit{members at time } T}.$$

- *Let* F *be the distribution function of* $U[f_c, 1]$.

Then

$$\hat{F}_T \to F \textit{ uniformly, a.s.}$$

Comments:

- The effect of removing lowest fitnesses leads to a censored fitness distribution.
- This is therefore a "skewed" version of Glivenko-Cantelli, and coincides with G-K when $p = 1$.

Fluctuations

LLN tells us that:

$$X_1 + \cdots + X_n \sim nE[X_1].$$

Or, upon "centering", we can look at the "fluctuations"

$$(X_1 - E[X_1]) + \cdots + (X_n - E[X_1]) = o(n).$$

Question. what is the "real" order of "fluctuations" ?

- Assuming $E[X_1^2] \in (0, \infty)$, the second moment is $\Theta(n)$.
- Reasonable to assume fluctuations of order $\sqrt{n}$.

Let's turn to Mike and his home-made Galton Board

Fluctuations

LLN tells us that:

$$X_1 + \cdots + X_n \sim nE[X_1].$$

Or, upon "centering", we can look at the "fluctuations"

$$(X_1 - E[X_1]) + \cdots + (X_n - E[X_1]) = o(n).$$

Question. what is the "real" order of "fluctuations" ?

- Assuming $E[X_1^2] \in (0, \infty)$, the second moment is $\Theta(n)$.
- Reasonable to assume fluctuations of order $\sqrt{n}$.

Let's turn to Mike and his home-made Galton Board

What did we just see ?

- The fluctuations exhibit some structure.
- We didn't get the order, but feel free to analyze more videos...

This deserves a theorem.

Central Limit theorem

Theorem 4

- *Suppose $X_1, X_2, \ldots$ are IID.*
- *(finite) Expectation $E[X_1] = \mu$.*
- *Variance $E[(X_1 - \mu)^2] = \sigma^2 \in (0, \infty)$.*

Then

$$P\left(\frac{\sum_{j=1}^{n}(X_j - \mu)}{\sigma\sqrt{n}} \le x\right) \to \Phi(x) \text{ where } \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}dy$$

is the distribution function of Gaussian (normal) with mean zero and variance 1 (AKA $N(0,1)$).

Comments:

- The fluctuations are of order of $\sqrt{n}$; and
- **" Universality"**: after scaling, fluctuations tend to a normal distribution, regardless of the original distribution of the X_j's.

Let's prove it.

Warning: this is a deep result. Its proof is non-trivial.

- It would be convenient to center and normalize $X_1, \ldots,$ by replacing $X_j \to \frac{X_j - \mu}{\sigma}$ (look again at the statement of the theorem)
- The resulting RVs have mean zero and variance 1.

Proof of CLT, I

"Universality"

- If there exists a limit, then it is the same whatever the D is, as long as $E[X_1] = 0$ and $E[X_1^2] = 1$.

Consider another such IID sequence $Y_1, Y_2, \ldots$, independent of the X_j's.
Let

$$\hat{S}^k = \frac{Y_1 + \cdots + Y_{k-1} + Y_k + X_{k+1} + \cdots + X_n}{\sqrt{n}} \text{ and}$$

$$\hat{S}^{k,0} = \frac{Y_1 + \cdots + Y_{k-1} + \cancel{Y_k} + X_{k+1} + \cdots + X_n}{\sqrt{n}}.$$

Then

$$\begin{aligned}
f(\hat{S}^n) - f(\hat{S}^0) &= \sum_{j=1}^{n} f(\hat{S}^j) - f(\hat{S}^{j-1}) \\
&= \sum_{j=1}^{n} \left(f(\hat{S}^{j,0}) + f'(\hat{S}^{j,0}) \frac{Y_j}{\sqrt{n}} + f''(\hat{S}^{j,0}) \frac{Y_j^2}{2n} + \ldots \right) \\
&- \sum_{j=1}^{n} \left(f(\hat{S}^{j,0}) + f'(\hat{S}^{j,0}) \frac{X_j}{\sqrt{n}} + f''(\hat{S}^{j,0}) \frac{X_j^2}{2n} + \ldots) \right) \\
&= \sum_{j=1}^{n} \left(f'(\hat{S}^{j,0}) \frac{Y_j - X_j}{\sqrt{n}} + f''(\hat{S}^{j,0}) \frac{X_j^2 - Y_j^2}{2n} + \ldots) \right)
\end{aligned}$$

Proof of CLT, I

"Universality"

- If there exists a limit, then it is the same whatever the D is, as long as $E[X_1] = 0$ and $E[X_1^2] = 1$.

Consider another such IID sequence $Y_1, Y_2, \ldots,$ independent of the X_j's.
Let

$$\hat{S}^k = \frac{Y_1 + \cdots + Y_{k-1} + Y_k + X_{k+1} + \cdots + X_n}{\sqrt{n}} \text{ and}$$

$$\hat{S}^{k,0} = \frac{Y_1 + \cdots + Y_{k-1} + \cancel{Y_k} + X_{k+1} + \cdots + X_n}{\sqrt{n}}.$$

Now take expectation.

$$\begin{aligned}
E[f(\hat{S}^n) - f(\hat{S}^0)] &= E[\sum_{j=1}^{n} f(\hat{S}^j) - f(\hat{S}^{j-1})] \\
&= E[\sum_{j=1}^{n} \left(f(\hat{S}^{j,0}) + f'(\hat{S}^{j,0}) \frac{Y_j}{\sqrt{n}} + f''(\hat{S}^{j,0}) \frac{Y_j^2}{2n} + \ldots \right) \\
&\quad - \sum_{j=1}^{n} \left(f(\hat{S}^{j,0}) + f'(\hat{S}^{j,0}) \frac{X_j}{\sqrt{n}} + f''(\hat{S}^{j,0}) \frac{X_j^2}{2n} + \ldots \right)] \\
&= \sum_{j=1}^{n} \left(E[f'(\hat{S}^{j,0}) \frac{Y_j - X_j}{\sqrt{n}}] + [f''(\hat{S}^{j,0}) \frac{X_j^2 - Y_j^2}{2n}] + O(n^{-3/2}) \right)
\end{aligned}$$

Proof of CLT, I

"Universality"

- If there exists a limit, then it is the same whatever the D is, as long as $E[X_1] = 0$ and $E[X_1^2] = 1$.

Consider another such IID sequence $Y_1, Y_2, \ldots,$ independent of the X_j's.
Let

$$\hat{S}^k = \frac{Y_1 + \cdots + Y_{k-1} + Y_k + X_{k+1} + \cdots + X_n}{\sqrt{n}} \text{ and}$$
$$\hat{S}^{k,0} = \frac{Y_1 + \cdots + Y_{k-1} + \cancel{Y_k} + X_{k+1} + \cdots + X_n}{\sqrt{n}}.$$

$$\begin{aligned}
E[f(\hat{S}^n) - f(\hat{S}^0)] &= \sum_{j=1}^{n} \left(E[f'(\hat{S}^{j,0}) \frac{Y_j - X_j}{\sqrt{n}}] + E[f''(\hat{S}^{j,0}) \frac{X_j^2 - Y_j^2}{2n}] + O(n^{-3/2}) \right) \\
&= \sum_{j=1}^{n} E[f'(\hat{S}^{j,0})] E[\frac{Y_j - X_j}{\sqrt{n}}] + \sum_{j=1}^{n} E[f''(\hat{S}^{j,0})] E[\frac{X_j^2 - Y_j^2}{2n}] + O(n^{-1/2}) \\
&= 0 + 0 + O(n^{-1/2}) \\
&= O(n^{-1/2}) \to 0 \text{ as } n \to \infty.
\end{aligned}$$

This shows universality.

Proof of CLT, II

Existence of limit

- Enough to show that there exists a choice for $Y_1, Y_2, \ldots$ for which the distribution of $\hat{S}^n$ has a limit.

How ?

- We know from undergrad probability that if $Y_1 \sim N(0,1)$, the distribution of $\hat{S}^n$ is $N(0,1)$ for all n, so we've found one...
- We can show limit of the form in the Theorem by a rather elementary combinatorial calculation for random variables taking only two values.

  This is done by an application of Stirling's formula, and is the oldest version of the CLT, known as deMoivre-Laplace theorem (1738).

CLT for Glivenko-Cantelli ?

Let's do some recalling:

- When $X_1, X_2, \dots$ IID, Glivenko-Cantelli states that empirical distribution function converges to distribution function:

$$\hat{F}_n \to F \text{ uniformly, a.s.}$$

- That is

$$\frac{\#\{j \leq n, X_j \leq f\}}{n} \to P(X_1 \leq f).$$

- Therefore the CLT for each fixed f is

$$\sqrt{n}(\hat{F}_n(f) - F(f)) \to \sigma_f N(0,1).$$

- Let's assume $X_1, X_2, \dots$ are $U[0,1]$. Then

$$\sigma_f^2 = f(1-f).$$

Question. for each n, LHS of $(*)$ is a process in f. Do these processes "converge" to a process ?

CLT for Glivenko-Cantelli

Theorem 5

- *Suppose* $X_1, X_2, \ldots$ *are IID* $U[0,1]$ *and let* F *be the distribution function.*
- *Let* $\hat{F}_n(f)$ *be the empirical distribution function.*
- *Let* $\Delta_n(f) = \sqrt{n}(\hat{F}_n(f) - F(f))$.

Then $\Delta_n(\cdot)$ *converges to a Brownian bridge, that is, a centered continuous Gaussian process* B *with covariance function* $E[B_s B_t] = s(1-t), 0 \leq s \leq t \leq 1$.

Results of this type are called functional CLTs.
A lot to explain here.

- A process is an indexed set of RVs. In this case the indices are the set $[0,1]$.
- A Gaussian process is a processes whose finite joint distributions are multivariate Gaussian (normal).
- A process is centered if all RVs are centered.
- The distribution of centered Gaussian processes are determined by their covariance function.
- The limit process is known as Brownian Bridge: Standard Brownian motion conditioned to be at 0 at time 1.

CLT for Glivenko-Cantelli

Theorem 5

- *Suppose* $X_1, X_2, \ldots$ *are IID* $U[0,1]$ *and let* F *be the distribution function.*
- *Let* $\hat{F}_n(f)$ *be the empirical distribution function.*
- *Let* $\Delta_n(f) = \sqrt{n}(\hat{F}_n(f) - F(f))$.

Then $\Delta_n(\cdot)$ *converges to a Brownian bridge, that is, a centered continuous Gaussian process* B *with covariance function* $E[B_s B_t] = s(1-t), 0 \leq s \leq t \leq 1$.

Results of this type are called functional CLTs.
As for the convergence:

- The mode of convergence is a topic for another discussion (weak convergence on right-continuous processes with left limits).
- This mode guarantees convergence of joint distributions and more (e.g. maximum, integrals)

CLT for Glivenko-Cantelli

Theorem 5

- *Suppose* $X_1, X_2, \ldots$ *are IID* $U[0,1]$ *and let* F *be the distribution function.*
- *Let* $\hat{F}_n(f)$ *be the empirical distribution function.*
- *Let* $\Delta_n(f) = \sqrt{n}(\hat{F}_n(f) - F(f))$.

Then $\Delta_n(\cdot)$ *converges to a Brownian bridge, that is, a centered continuous Gaussian process* B *with covariance function* $E[B_s B_t] = s(1-t), 0 \leq s \leq t \leq 1$.

Results of this type are called functional CLTs.
Here's simulation of Brownian Bridge.

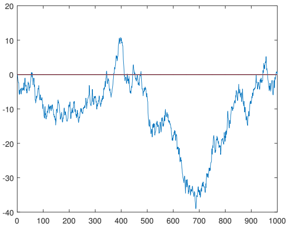


Figure: Brownian Bridge

Back to toy our model

Here we also have a functional CLT. The expression is more involved though. Let

$$\hat{\Delta}_n = \hat{F}_n - F$$

Processes appearing in limit

- W_1 standard BM, and the corresponding bridge Br_1:

$$\mathrm{Br}_1(f) := W_1(f) - fW_1(1).$$

- If $f_c = 0$, choose $\widetilde{W_1} \equiv 0$.
- If $f_c > 0$: $\widetilde{W_1}$ standard BM derived from W_1 as follows :
  - $U \sim U[f_c, 1]$, independent of W_1.
  - An "interval" $\widetilde{A}_t$ of length $(1 - f_c)t$, shifted by U.
  - $\widetilde{W_1}(t) := \frac{1}{\sqrt{f_c(1-f_c)}}\left((1-f_c)W_1(f_c t) + f_c \int \mathbf{1}_{\widetilde{A}_t}(s)dW_1(s)\right).$

- W_2, standard BM, independent of W_1, U.

Back to toy our model

Here we also have a functional CLT. The expression is more involved though. Let

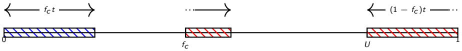
$$\hat{\Delta}_n = \hat{F}_n - F$$

Processes appearing in limit

- W_1 standard BM, and the corresponding bridge Br_1:

$$\mathrm{Br}_1(f) := W_1(f) - fW_1(1).$$

- If $f_c = 0$, choose $\widetilde{W_1} \equiv 0$.
- If $f_c > 0$: $\widetilde{W_1}$ standard BM derived from W_1 as follows :
    - $U \sim U[f_c, 1]$, independent of W_1.
    - An "interval" $\widetilde{A}_t$ of length $(1 - f_c)t$, shifted by U.
    - $\widetilde{W_1}(t) := \frac{1}{\sqrt{f_c(1 - f_c)}} \left((1 - f_c)W_1(f_c t) + f_c \int \mathbf{1}_{\widetilde{A}_t}(s) dW_1(s) \right).$



- W_2, standard BM, independent of W_1, U.

Toy model functional CLT

For a path $\omega \in D[0,1]$, let $\Psi(\omega) := \omega(1) - \inf_{0 \le t \le 1} \omega(t)$

Theorem 6 (B '13)

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{p}\begin{pmatrix} \overbrace{\sigma_1 Br_1 + f_c W_2(1)(1-F)}^{\textit{Gaussian process}} \\ \underbrace{\Psi(\widetilde{\sigma}_1 \widetilde{W}_1 + f_c W_2)}_{\textit{Positive RV}} \end{pmatrix},$$

with $\sigma_1 = \sqrt{p}$, $\widetilde{\sigma}_1 = \sqrt{f_c(2p-1)}$, *and the convergence is* $D(f_c,1] \times \mathbb{R}$.

Marginals

$$\sqrt{n}\widehat{\Delta}_n(f) \Rightarrow \begin{cases} \sigma(f \wedge 1)N(0,1) & f > f_c; \\ \sigma(f_c)|N(0,1)| & f = f_c; \\ 0 & f < f_c, \end{cases}$$

where $\sigma(f) := \frac{1}{p}\sqrt{f(1-f)p + \left(\frac{1-f}{1-f_c}\right)^2 f_c}$

Toy model functional CLT

For a path $\omega \in D[0,1]$, let $\Psi(\omega) := \omega(1) - \inf_{0 \le t \le 1} \omega(t)$

Theorem 6 (B '13)

$$\sqrt{n} \begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{p} \begin{pmatrix} \overbrace{\sigma_1 Br_1 + f_c W_2(1)(1-F)}^{\textit{Gaussian process}} \\ \underbrace{\Psi(\widetilde{\sigma}_1 \widetilde{W}_1 + f_c W_2)}_{\textit{Positive RV}} \end{pmatrix},$$

with $\sigma_1 = \sqrt{p}$, $\widetilde{\sigma}_1 = \sqrt{f_c(2p-1)}$, *and the convergence is* $D(f_c, 1] \times \mathbb{R}$.

Marginals

$$\sqrt{n}\widehat{\Delta}_n(f) \Rightarrow \begin{cases} \sigma(f \wedge 1) N(0,1) & f > f_c; \\ \sigma(f_c)|N(0,1)| & f = f_c; \\ 0 & f < f_c, \end{cases}$$

where $\sigma(f) := \frac{1}{p}\sqrt{f(1-f)p + \left(\frac{1-f}{1-f_c}\right)^2 f_c}$

Toy model CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{p}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + f_c W_2(1) g \\ \Psi(\widetilde{\sigma_1 W_1} + f_c W_2) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while it limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

Toy model CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{p}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + f_c W_2(1) g \\ \Psi(\widetilde{\sigma_1 W_1 + f_c W_2}) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while it limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

Toy model CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{p}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + f_c W_2(1) g \\ \Psi(\widetilde{\sigma_1 W_1} + f_c W_2) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while it limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

Toy model CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{p}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + f_c W_2(1) g \\ \Psi(\widetilde{\sigma_1 W_1 + f_c W_2}) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while it limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

Toy model CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{p}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + f_c W_2(1) g \\ \Psi(\widetilde{\sigma}_1 \widetilde{W}_1 + f_c W_2) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while it limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

Toy model CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{p}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + f_c W_2(1) g \\ \Psi(\widetilde{\sigma}_1 \widetilde{W_1} + f_c W_2) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while it limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

Toy model CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{p}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + f_c W_2(1) g \\ \Psi(\widetilde{\sigma}_1 \widetilde{W_1} + f_c W_2) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while it limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

Toy model CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{p}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + f_c W_2(1) g \\ \Psi(\widetilde{\sigma}_1 \widetilde{W_1} + f_c W_2) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while it limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

Toy model CLT Discussion

Recall

$$\sqrt{n}\begin{pmatrix} \widehat{\Delta}_n(\cdot)|_{(f_c,1]} \\ \widehat{\Delta}_n(f_c) \end{pmatrix} \Rightarrow \frac{1}{p}\begin{pmatrix} \sigma_1 \mathrm{Br}_1 + f_c W_2(1) g \\ \Psi(\widetilde{\sigma}_1 \widetilde{W_1} + f_c W_2) \end{pmatrix}$$

Origin of terms

1. Bridge arising from empirical process associated with births.
   Only surviving term when $f_c = 0$, recovering classical CLT for empirical processes.
2. Fluctuations from bridge due to randomness of births, and existence of deaths
3. Population with fitness $\leq f_c$ is null recurrent random walk above its running minimum, hence Ψ.
   Note that it's of order $\sqrt{n}$, hence only appearing in CLT.
   a. Scaling limit for the births.
   b. Fluctuations from randomness of births, and negative increments.

Discontinuity

- The limit process is not in $D[f_c, 1]$, because its distribution at f_c is $\sigma(f_c)|N(0,1)| > 0$ a.s., while it limit from the right is $\sigma(f_c)N(0,1)$.
- The standard normal random variables above are NOT the same.

Thank you.