Extinction times in the stochastic logistic epidemic

Malwina Luczak[1]

School of Mathematical Sciences
Queen Mary University of London
e-mail: m.luczak@qmul.ac.uk

2 April 2014
Durham

[1]Supported by EPSRC Leadership Fellowship, grant reference EP/J004022/2. Joint work with Graham Brightwell, LSE.

A simple SIS epidemic model

- Each individual in a population of size N is either *infective* or *susceptible*.
- $X^N(t)$ represents the number of infectives at time t.
- Each infective encounters a random other member of the population at rate λ (infection rate); if the other individual is currently susceptible, they become infective.
- Each infective recovers at rate μ (recovery rate); once recovered they become susceptible again.
- This model for the spread of an *SIS epidemic* in a population is due to Feller (1939), Bartlett (1957) and Weiss and Dishon (1971).

SIS epidemic model: the underlying Markov chain

- $(X^N(t))_{t\geq 0}$ evolves as a continuous-time Markov chain with state space $\{0,\dots,N\}$.
- The transitions are as follows:

$$\begin{array}{lllll} x & \to & x+1 & \text{at rate} & \lambda x(1-x/N), \\ x & \to & x-1 & \text{at rate} & \mu x. \end{array}$$

Other application areas

This is a very simple stochastic process, that also appears in the contexts of:

- metapopulation models,
- spread of rumours,
- chemical reactions.

Alternative interpretation: metapopulation model

- There are N patches, and $X^N(t)$ is the number of patches occupied at time t, for $t \geq 0$. Then $X^N(t)$ has the following dynamics.
- The number $X^N(t)$ increases by 1 at rate $\frac{\lambda X^N(t)(N - X^N(t))}{N}$: each occupied patch attempts to "colonise" another patch at rate λ, and the probability that the colonised patch is currently unoccupied is $(N - X^N(t))/N$.
- The number $X^N(t)$ decreases by 1 at rate $\mu X^N(t)$: each colony is wiped out at rate μ.

How does the epidemic behave?

- We are interested in limiting behaviour as $N \to \infty$.
- One might expect the behaviour of this stochastic process to be related to the solution of the differential equation

$$\frac{dz}{dt} = \lambda z(1-z) - \mu z \qquad z \in [0,1].$$

  Here $z(t)$ represents the proportion of infectives at time t.
- For $\lambda \leq \mu$, the equation has a unique attractive fixed point at $z = 0$.
- For $\lambda > \mu$, the fixed point at 0 is repulsive, and there is an attractive fixed point at $z = 1 - \mu/\lambda$.

Basic reproduction number

- The behaviour of the Markov chain $X^N(t)$ also depends on whether λ is greater than, equal to, or less than μ. In other words, a key parameter is the ratio λ/μ, and whether it is greater or less than 1.
- In the context of an epidemic, the ratio λ/μ is called the *basic reproduction number*, and denoted R_0. It means the average number of cases one case generates over the course of its infectious period.
- If $R_0 \leq 1$, then the probability of an epidemic becoming established tends to 0 as the population size $N \to \infty$. If $R_0 > 1$, then this probability is asymptotically positive.

Long-term behaviour of the stochastic model

- The stochastic model we introduced is a continuous-time Markov chain, with a finite state space $\{0, \ldots, N\}$.
- There is an *absorbing state*, namely 0. Once the Markov chain enters this state, it stays there.
- With probability 1, the Markov chain will eventually enter the absorbing state: the epidemic will die out, even when $R_0 > 1$ (i.e. even when $\lambda > \mu$, unlike the deterministic version).

Extinction time: $\lambda > \mu$

- For $x_0 = X^N(0)$, let $T_e^{X^N}(x_0)$ be the time to extinction for (X_t^N). i.e., the hitting time of the absorbing state 0.
- Whenever $x_0 = X^N(0) \to \infty$, we have

$$\mathbb{E}\, T_e^{X^N}(x_0) = \sqrt{2\pi}\frac{\lambda}{(\lambda-\mu)^2}\frac{e^{\gamma N}}{\sqrt{N}}(1-o(1)),$$

  as $N \to \infty$, where $\gamma = \log\lambda - \log\mu - \frac{\lambda-\mu}{\lambda} > 0$.
- Moreover, the time to extinction is asymptotically an exponential random variable.
- See: Barbour (1976), Andersson and Djehiche (1998), Nåsell (2011).

$\lambda > \mu$

- Conditioning on the event that the chain has not entered state 0 by time t, one obtains a limiting distribution, called the *quasi-stationary distribution*, centred around the attractive fixed point of the differential equation.
- Starting from a fixed state, the chain converges (presumably rapidly) to the quasi-stationary distribution.
- Moving from near the fixed point to 0 is a *rare event*. The expected time until the rare event occurs can be estimated very precisely, as above. The time to extinction is exactly exponential if the chain is started in the quasi-stationary distribution.

Starting from a small state

If the chain starts in a fixed state x_0, there is a positive probability (asymptotically $(\mu/\lambda)^{x_0}$) that the epidemic dies out in constant time.

Conditioned on this not occurring, the extinction time is distributed asymptotically as above.

Extinction time: $\lambda < \mu$

In this case, the time to extinction is approximately $\log N/(\mu - \lambda)$.

This is the focus of our work, and more details will follow.

Extinction time: $\lambda = \mu$

If $\lambda = \mu$, the time to extinction is somewhere in between (time of order $N^{1/2}$, it turns out).

Doering, Sargsyan and Sander (2005) give a formula for the expectation of the extinction time, starting from a state x_0 of order N:

$$\mathbb{E} T_e^{X^N}(x_0) = \left(\frac{\pi}{2}\right)^{3/2} \sqrt{N} + \log x_0 + O(1).$$

A critical window

- Suppose $\lambda = \lambda(N)$ and $\mu = \mu(N)$.
- There is a "critical window" where $|\mu - \lambda| = O(N^{-1/2})$.
- If $(\mu - \lambda)N^{1/2} \to C$ $(-\infty < C < \infty)$ and $x_0 N^{-1/2} \to b$ $(b > 0)$, then the expected time to extinction is asymptotically $f(C, b)N^{1/2}$, for some function f, and the time to extinction is not well-concentrated. See Dolgoarshinnyk and Lalley (2006).

Inside the critical window

- Nåsell (2011, and earlier papers) shows that, within the window, the expected time to extinction starting from a state of order $\sqrt{N}$ is of order $\sqrt{N}$, whereas the expected extinction time starting from state 1 is of order $\log N$.
- It follows via our methods that the time to extinction is of order $\sqrt{N}$, even if the starting state is of order larger than $\sqrt{N}$.

Scaling window

Thinking of a scaling window gives a more sophisticated picture. Suppose again $\lambda = \lambda(N)$ and $\mu = \mu(N)$.

- If $\lambda - \mu \to 0$, and $(\lambda - \mu)\sqrt{N} \to \infty$ (sufficiently fast), the epidemic takes a long time to die out (time of order roughly $\exp(N(\lambda - \mu)^2/2\lambda^2))$. See work of Nåsell.
- If $(\mu - \lambda)\sqrt{N} \to \infty$, the epidemic dies out quickly (time of order $\frac{1}{\mu-\lambda}\log[N(\mu - \lambda)^2]$ if we start from a state of order N). More details will follow later in the talk.

$\lambda < \mu$

- The following formula appears in two papers from the literature. Here the chain starts in state $x_0 = z_0 N$, it is assumed that λ and μ are constants with $\lambda < \mu$, and $T_e = T_e^{X^N}(x_0)$ is the (random) time to extinction.

  $$(\mu - \lambda(1 - z_0))T_e - (\log N + \log z_0 + \log(\mu - \lambda + \lambda z_0) - \log \mu) \to W,$$

  in distribution, where W has the *standard Gumbel distribution*: $\mathbb{P}(W \leq w) = e^{-e^{-w}}$.
- Unfortunately, this formula is incorrect.

What is wrong?

Formula in the literature:

$$(\mu - \lambda(1 - z_0))T_e - (\log N + \log z_0 + \log(\mu - \lambda + \lambda z_0) - \log \mu) \to W.$$

- The first-order asymptotics are $T_e^{X^N}(z_0 N) \simeq \dfrac{\log N}{\mu - \lambda(1 - z_0)}$.
- But the constant in front of the $\log N$ surely should not depend on z_0.
- The formula appears to say that $T_e^{X^N}(z_0 N) = O(\log N)$ in the case $\mu = \lambda$, which cannot be right.
- Moreover, the term $\big(\log N + \cdots\big)$ does not behave as it should as $\mu - \lambda$ decreases to 0.

Previous results and open problems

- Doering, Sargsyan and Sander (2005) give an asymptotic formula for the mean extinction time, in the case where $\lambda < \mu$ are fixed constants, in the form $\frac{1}{\mu-\lambda}(\log N + \log z_0) + O(1)$, and note that "our formulas do not agree with... but they do agree with the numerical results."
- Nåsell (2011) leaves as an open problem the asymptotic estimation of the expected time to extinction in the cases where (a) $\mu - \lambda$ is bounded away from zero, or tends to zero slowly ("subcritical regime"), (b) $|\mu - \lambda| = O(N^{-1/2})$ ("critical regime").

Theorem [B. & Luczak, 2014+]

Suppose $\lambda = \lambda(N)$ and $\mu = \mu(N)$, and

- $(\mu - \lambda)N^{1/2} \to \infty$,
- $X^N(0)/N \to z_0$, with $z_0 \in (0, 1]$.

Then $T_e = T_e^{X^N}(z_0 N)$ satisfies

$$(\mu-\lambda) T_e - \big(\log N + \log z_0 + 2\log(\mu-\lambda) - \log(\mu-\lambda+\lambda z_0) - \log\mu \big) \to W,$$

in distribution, where W has the standard Gumbel distribution.

This formula seems not to have previously appeared in the literature, even in the case where λ and μ are constants.

Our result gives the asymptotic distribution of the extinction time throughout the entire subcritical regime.

Consequences

Formula for distribution of extinction time:

$$(\mu-\lambda)T_e-\big(\log N+\log z_0+2\log(\mu-\lambda)-\log(\mu-\lambda+\lambda z_0)-\log\mu\big)\to W.$$

- $T_e = T_e^{X^N}(z_0 N)$ is sharply concentrated around $\dfrac{\log N + 2\log(\mu-\lambda)}{\mu-\lambda}$, throughout this regime.
- We can easily derive an asymptotic formula for $\mathbb{E}\,T_e^{X^N}(z_0 N)$.
- The formula "runs out" when $\mu-\lambda$ is of order about $N^{-1/2}$, at the transition into the critical regime.

Other starting states

- We do not need to assume that the starting state x_0 is of order N.
- Whenever $(\mu-\lambda)\sqrt{N} \to \infty$, and $x_0(\mu-\lambda) \to \infty$,

$$(\mu-\lambda)T_e - \Big(\log x_0 + \log(\mu-\lambda) - \log\Big(1+\frac{\lambda x_0}{(\mu-\lambda)N}\Big) - \log\mu\Big) \to W,$$

in distribution, where W has the standard Gumbel distribution.

Simpler formulae

- The formula on the previous slide can be simplified in certain regimes.
- If $x_0(\mu - \lambda) \to \infty$ and $x_0 = o((\mu - \lambda)N)$, then

$$(\mu - \lambda) T_e^{X^N}(x_0) - \big(\log x_0 + \log(\mu - \lambda) - \log \mu \big) \to W,$$

  in distribution, where W has the standard Gumbel distribution.
- If $x_0/(\mu - \lambda)N \to \infty$, then

$$(\mu - \lambda) T_e^{X^N}(x_0) - \big(\log N + 2\log(\mu - \lambda) - 2\log \mu \big) \to W,$$

  in distribution, where W has the standard Gumbel distribution.

A special case

In particular, if $\mu - \lambda = o(1)$ and the starting state $x_0 = X^N(0)$ is of order N, then

$$(\mu - \lambda) T_e^{X^N}(x_0) - (\log N + 2\log(\mu - \lambda) - 2\log\mu) \to W$$

in distribution.

Note that this formula does not involve the initial value $z_0 = X^N(0)/N$.

Informal description

- From any starting state above about $(\mu - \lambda)N$, $X^N(t)$ moves rapidly to a state of order $(\mu - \lambda)N$.
- The bulk of the time to extinction is spent moving from a state of order $(\mu - \lambda)N$ to a state of order about $1/(\mu - \lambda)$.
- However, most of the variability in the extinction time comes from the final phase, from a state around $1/(\mu - \lambda)$ to extinction.

Formula for expected extinction time

- The following formula for the *expected* extinction time, starting from state x, is originally due to Leigh (1981), and has been rediscovered several times since.

$$\mathbb{E}T_e^{X^N}(x) = \frac{1}{\mu}\sum_{j=0}^{N-1}\left(\frac{\lambda}{\mu N}\right)^j \sum_{s=1}^{\min(x,N-j)} \frac{(N-s)!}{(N-s-j)!}\frac{1}{s+j}.$$

  This formula is valid for all values of the parameters.
- Deriving the precise asymptotics of this sum in various different regimes is quite challenging.
- Moreover, we are interested in the distribution of the extinction time, not just the expectation.

A different approach: proof overview

- We assume that our starting state x_0 satisfies $x_0(\mu - \lambda) \to \infty$.
- Our intermediate results are stated in terms of a function $\omega(N)$ tending slowly to infinity. (We can take $\omega(N) = (\mu - \lambda)^{1/4} N^{1/8}$.)
- Our proof proceeds by analysing the Markov chain in three phases, corresponding to the previous rough description of the course of the epidemic. For some starting states, we don't need all the phases.

The three phases

- In the first phase, we show that, in some time t that is small compared to the overall duration of the epidemic, $X^N(t)$ drops to a state below $N(\mu - \lambda)\omega(N)$ with high probability.
- In the second phase, we show that $X^N(t)/N$ closely follows a solution to the differential equation, starting from a state at most $N(\mu - \lambda)\omega(N)$, until it reaches a state below about $N^{1/2}\omega(N)$.
- In the third phase, once $X^N(t) \leq N^{1/2}\omega(N)$, we show that $X^N(t)/N$ behaves like a linear birth and death chain, whose behaviour is well-understood.

Differential equation approximation

- Recall the differential equation

$$\frac{dz}{dt} = \lambda z(1-z) - \mu z \qquad z \in [0,1],$$

derived from the average drifts.

- The general theory of Kurtz (1971) tells us that, if $X^N(t)/N$ starts close to a solution $z(t)$ of the differential equation, then it remains close to this solution over a time interval of constant length.

- Our method shows that $X^N(t)/N$ in fact follows the differential equation closely over longer time intervals.

Solution of the differential equation

- The equation $\frac{dz}{dt} = \lambda z(1-z) - \mu z$ can be solved explicitly:

$$z(t) = \frac{z_0(\mu-\lambda)e^{-(\mu-\lambda)t}}{\mu - \lambda + z_0\lambda(1-e^{-(\mu-\lambda)t})},$$

where $z_0 = z(0)$.

- We shall make use of the inverse of the function $z(t)$:

$$t(z) = \frac{\log(z_0/z) + \log\left(1+\frac{\lambda}{\mu-\lambda}z\right) - \log\left(1+\frac{\lambda}{\mu-\lambda}z_0\right)}{\mu-\lambda},$$

where $t(z_0) = 0$.

First phase: upper bound

It turns out that, for any starting state x_0 and any values of the parameters,

$$\mathbb{E}_{x_0} X_t^N \leq Nz(t),$$

where $z(t)$ is the solution of the differential equation with $z(0) = x_0/N$.

Applying this with $t = \dfrac{1}{\omega(N)^{1/2}\lambda(\mu - \lambda)}$, for any $x_0 \leq N$, we find

$$\mathbb{E}_{x_0} X_t^N \leq N(\mu - \lambda)\omega(N)^{1/2},$$

and therefore $\mathbb{P}(X_t^N \geq N(\mu - \lambda)\omega(N)) = o(1)$.

Second phase: following the differential equation

If $(X^N(t)/N)$ were to follow the differential equation *exactly*, starting from $z_0 = x_0/N$, then the hitting time of state $x^* = N^{1/2}\omega(N)$ would be $t(x^*/N)$, which is equal to

$$\frac{\log x_0 - \log x^* + \log(\mu - \lambda) - \log\left(\mu - \lambda + \lambda z_0\right) + o(1)}{\mu - \lambda}.$$

(Note that $\log\left(1 + \frac{\lambda}{\mu-\lambda}\frac{N^{1/2}\omega(N)}{N}\right) = o(1)$ by our assumption on $\mu - \lambda$ and choice of $\omega(N)$.)

How to prove this law of large numbers?

We use a concentration of measure result from L. (2013). This is designed for use in combination with a *coupling* of two copies of a Markov chain, and is especially useful if the coupling is *contractive*, i.e., the expected distance between the two copies decreases on one step of the coupled chain.

Let P be the transition matrix of a discrete-time Markov chain, and let g be a function on its state space. Then let $(Pg)(x)$ denote the expectation of $g(Y)$, where Y is chosen by taking one step of the chain from x.

Concentration of measure inequality – L. (2013)

Let P be the transition matrix of a discrete-time Markov chain (X_i) with discrete state space S. Let $f : S \to \mathbb{R}$ be a function. Suppose $S_0 \subseteq S$ and functions $a_{x,i}$ on S satisfy $|\mathbb{E}_x[f(X_i)] - \mathbb{E}_y[f(X_i)]| \leq a_{x,i}(y)$ whenever $x, y \in S_0$, and $P(x,y) > 0$. Let $S_0^0 = \{x \in S_0 : y \in S_0$ whenever $P(x,y) > 0\}$. Assume that, for some sequence $(\alpha_i)_{i \in \mathbb{Z}^+}$ of positive constants, $\sup_{x \in S_0^0}(Pa_{x,i}^2)(x) \leq \alpha_i^2$. Let $k > 0$, and let $\beta = 2\sum_{i=0}^{k-1} \alpha_i^2$. Suppose also that $2\sup_{0 \leq i \leq k-1} \sup_{x \in S_0^0, P(x,y)>0} a_{x,i}(y) \leq \alpha$. Let $A_k = \{\omega : X_i(\omega) \in S_0^0 : 0 \leq i \leq k-1\}$. Then, for all $a > 0$,

$$\mathbb{P}_{x_0}\left(\left\{|f(X_k) - \mathbb{E}_{x_0}[f(X_k)]| \geq a\right\} \cap A_k\right) \leq 2e^{-a^2/(2\beta + 2\alpha a/3)}.$$

Applying the theorem

- We apply this result to a discretised version $\hat{X}_k$ of $(X^N(t))$, with transition probabilities given by:

$$p_{x,x+1} = \frac{\lambda x(1 - x/N)}{K(\mu + \lambda)N}; \quad p_{x,x-1} = \frac{\mu x}{K(\mu + \lambda)N};$$

  and $p_{x,x} = 1 - p_{x,x+1} - p_{x,x-1}$, where K is a large constant. So $(X^N(\frac{k}{K(\mu+\lambda)N}))$ is approximated by $\hat{X}_k$.
- We define a coupling of two copies $\hat{X}$ and $\hat{Y}$ as follows.
  - If $\hat{X}_k = \hat{Y}_k$, then the two copies move together at the next step.
  - Otherwise, at most one of the two copies moves (so they never "cross").

A contractive coupling

We prove that this coupling is contractive, as long as $\mu > \lambda$:

$$\mathbb{E}(|\hat{X}_{k+1} - \hat{Y}_{k+1}| \mid \mathcal{F}_k) \leq |\hat{X}_k - \hat{Y}_k| \left(1 - \frac{\mu - \lambda}{K(\mu + \lambda)N}\right),$$

for all k, where $(\mathcal{F}_k)$ is the natural filtration of the coupling.

Applying the theorem

- We apply the theorem with $S_0 = \{0, 1, \ldots, 2x_0\}$.
- We obtain that

$$\mathbb{P}_{x_0}\left(|\hat{X}_k - \mathbb{E}_{x_0}\hat{X}_k| \geq a\right) \leq 2\exp\left(-\frac{a^2}{4x_0(\lambda+\mu)/(\lambda-\mu)+\frac{4}{3}a}\right) + e^{-(\mu-\lambda)x_0/\mu},$$

  for all a and k.
- The last term above is an upper bound for the probability that the chain leaves S_0, i.e., it reaches $2x_0$ before 0.

How do we use concentration of measure?

- Consider $E_k = \frac{\mathbb{E}\hat{X}_k}{N} - z\left(\frac{k}{K(\mu+\lambda)N}\right)$.
- A short calculation gives

$$
\begin{aligned}
&\mathbb{E}(\hat{X}_{k+1} - \hat{X}_k) \\
&= \frac{1}{K(\lambda+\mu)N}\Big(- (\mu-\lambda)\mathbb{E}\hat{X}_k - \lambda \frac{\mathbb{E}\hat{X}_k^2}{N}\Big) \\
&= \frac{1}{K(\lambda+\mu)N}\Big(- (\mu-\lambda)\mathbb{E}\hat{X}_k - \lambda \frac{(\mathbb{E}\hat{X}_k)^2}{N} - \lambda \frac{\mathbb{E}(\hat{X}_k - \mathbb{E}\hat{X}_k)^2}{N}\Big).
\end{aligned}
$$

- The fact that $\hat{X}_k$ is well-concentrated implies that this last term can be bounded, uniformly in k. In fact, $\mathbb{E}(\hat{X}_k - \mathbb{E}\hat{X}_k)^2 \leq \frac{30x_0(\lambda+\mu)}{\mu-\lambda}$

Controlling the error E_k

- Also, for some $u \in \left(\frac{k}{K(\mu+\lambda)N}, \frac{k+1}{K(\mu+\lambda)N}\right)$,

$$
\begin{aligned}
z\Big(\frac{k+1}{K(\mu+\lambda)N}\Big) - z\Big(\frac{k}{K(\mu+\lambda)N}\Big) \\
= \quad \frac{1}{K(\lambda+\mu)N}\Big[-(\mu-\lambda)z\Big(\frac{k}{K(\mu+\lambda)N}\Big) - \lambda z\Big(\frac{k}{K(\mu+\lambda)N}\Big)^2\Big] \\
+ \frac{1}{2K^2(\mu+\lambda)^2N^2} z''(u).
\end{aligned}
$$

- We then deduce that

$$
|E_{k+1}| \leq |E_k|\Big(1 - \frac{\mu-\lambda}{K(\mu+\lambda)N}\Big) + \frac{30\lambda x_0}{(\mu-\lambda)KN^3} + \frac{\sup_u |z''(u)|}{2K^2(\mu+\lambda)^2N^2},
$$

where the supremum is over $u \in [0,1]$. The last term is negligible for large K.

Deducing the law of large numbers

- The calculation above gives a uniform bound:

$$E_k = \frac{\mathbb{E}\hat{X}_k}{N} - z\Big(\frac{k}{K(\mu+\lambda)N}\Big) \leq \frac{70x_0\lambda\mu}{N^2(\mu-\lambda)^2}.$$

- Then the concentration of measure result is applied again to show that, with high probability, $\hat{X}_k/N$ is close to its expectation, and therefore close to the solution of the differential equation for all time.
- This is exactly what we want, provided "close" means small with respect to our target value $N^{1/2}\omega(N)$.

Third phase: approximation

Once we reach a point where $X^N(t)/N$ is less than $N^{1/2}\omega(N)/N \ll \mu - \lambda$, downward steps occur at rate $\mu X^N(t)$, and upward steps at rate

$$\lambda X^N(t)(1 - X^N(t)/N) = X^N(t)\big(\mu - (\mu - \lambda) - \lambda X^N(t)/N\big).$$

So, roughly speaking, the "logistic correction" $-\lambda X^N(t)^2/N$ is a negligible contribution to the downward drift.

Thus the behaviour of the logistic process is essentially the same as that of the *linear birth and death chain* $(Y(t))$ taking steps up at rate $\lambda Y(t)$ and down at rate $\mu Y(t)$.

The linear birth and death chain

The extinction time of the chain $(Y(t))$ can be analysed exactly, using probability generating functions.
If $Y(0) = y_0$, then, for $t \geq 0$:

$$\mathbb{P}(Y(t) = 0) = \left(\frac{\mu - \mu e^{-(\mu-\lambda)t}}{\mu - \lambda e^{-(\mu-\lambda)t}} \right)^{y_0}.$$

Let $T_e^Y(y_0)$ be the time to extinction for $(Y(t))$. Then:

$$(\mu - \lambda) T_e^Y(y_0) - \big(\log y_0 + \log(\mu - \lambda) - \log \mu \big) \to W,$$

in distribution, where W has the standard Gumbel distribution, provided $y_0(\mu - \lambda) \to \infty$.

Extinction time in the third phase

We show that, if $(X^N(t))$ is started at any state near $x^* = N^{1/2}\omega(N)$, then the time to extinction has the same asymptotic distribution as $T_e^Y(x^*)$, i.e.,

$$(\mu - \lambda) T_e^{X^N}(x^*) - \big(\log x^* + \log(\mu - \lambda) - \log \mu \big) \to W,$$

in distribution, where W has the standard Gumbel distribution.

Combining the phases

- The time for the first phase is $o((\mu - \lambda)^{-1})$.
- The time for the second phase is

$$\frac{\log x_0 - \log x^* + \log(\mu - \lambda) - \log\left(\mu - \lambda + \lambda x_0/N\right) + o(1)}{\mu - \lambda}.$$

- The time for the third phase is distributed as

$$\frac{\log x^* + \log(\mu - \lambda) - \log \mu + W}{\mu - \lambda}.$$

- So the total time to extinction from state x_0 is distributed as

$$\frac{\log x_0 + 2\log(\mu - \lambda) - \log\left(\mu - \lambda + \lambda x_0/N\right) - \log \mu + W}{\mu - \lambda}.$$