Supplemental Material for "Asymptotic absorption-time distributions in extinction-prone Markov processes"

David Hathcock[1] and Steven H. Strogatz[2]

[1]*Department of Physics, Cornell University, Ithaca, New York 14853, USA*
[2]*Department of Mathematics, Cornell University, Ithaca, New York 14853, USA*

(Dated: April 19, 2022)

This Supplemental Material provides rigorous mathematical derivations supporting the results discussed in the main text and gives details of the various example models used to demonstrate the Gaussian and Gumbel classes. Section S1 provides discussion of additional cases (with different initial or boundary conditions) where our results still apply. Section S2 gives a detailed derivation of the absorption-time cumulants (Eq. (2) in the main text), including the weighting factors w^n and their properties. Sections S3 and S4 provide the details of the asymptotic analysis used to characterize Markov processes with normal and Gumbel absorption-time distributions respectively and discuss the leading corrections and transition matrix spectrum for each case. Section S5 shows the classification of processes with transition rates that decay as a power-law with exponent p near the boundary. We show this power-law decay leads to Gaussian distributions ($p \leq 1/2$), or a skewed family of distributions ($p > 1/2$). Finally, in Section S6 we describe the dynamics for the evolutionary game, the susceptible-infectious-susceptible model, the logistic model, and the autocatalytic chemical reaction model. For each case we derive the Markov transition probabilities and explain how the model falls into one of our universality classes.

S1. DISTRIBUTIONS ARE ROBUST TO CHANGES IN INITIAL AND BOUNDARY CONDITIONS

In the main text, we specialize to Markov chains with a finite state space of size N, a reflecting upper boundary, and initial condition at the maximal state $p_m(0) = \delta_{m,N}$. Our asymptotic absorption-time distributions, however, should be robust to changes in initial and boundary conditions. Because the dynamics are extinction-prone, the system quickly progresses toward the absorbing state, spending negligible time near the reflecting boundary. Therefore, if the initial condition m_0 is sufficiently large ($m_0 \sim N$ for large N), corrections due to variation in the initial condition will be sub-dominant as $N \to \infty$. By the same argument, we expect the same asymptotic distributions to occur for infinite systems with free boundary conditions and no maximal state N, but large initial condition. On the other hand, if the upper boundary is absorbing, our result describes the absorption-time distribution, given that the absorbing state at 0 is reached (i.e. if we ignore all trajectories that are absorbed at the upper boundary) [1]. Finally, our results can also be used to determine the first-passage-time distribution to an arbitrary state m, since the first-passage problem can be solved by making the target state absorbing [2].

S2. ABSORPTION-TIME CUMULANTS AND WEIGHTING FACTOR PROPERTIES

A. Derivation of the absorption-time cumulants

In this section we derive a general formula for the absorption-time cumulants, shown as Eq. (2) in the main text. This derivation follows Ref. [1], but we generalize to Markov chains where the ratio $r_m = b_m/d_m$ is non-constant. We start from the master equation, Eq. (1) in the main text, and restrict our attention to the transient (non-absorbing) states, $m > 0$, since these determine the time it takes to reach absorption. The master equation for these states can be expressed as $\dot{\mathbf{p}}(t) = \Omega \cdot \mathbf{p}(t)$, where Ω is the transient transition matrix with elements

$$\Omega_{mn} = b_n\delta_{m,n+1} + d_n\delta_{m,n-1} - (b_n + d_n)\delta_{m,n} \tag{S1}$$

for $m, n = 1, \ldots, N$ and $\mathbf{p}(t)$ is the vector of transient state occupancy probabilities.

The entire first-passage process can be characterized in terms of the transition matrix Ω. In fact, the first-passage distribution $p(t)$ can be written in terms of an element of the matrix exponential, $p(t) = d_1[\exp(\Omega t)]_{1,N}$ and the moments of T are

$$\mu_n \coloneqq E[T^n] = (-1)^n n!\mathbf{1}\Omega^{-n}\mathbf{p}(0), \tag{S2}$$

where $\mathbf{1}$ is a row vector containing all 1's and E denotes expected value.

As discussed in the main text, to proceed it is useful to introduce the following decomposition of the transition matrix: $\Omega = \Omega_{RW} D$, where D is a diagonal matrix $D_{mm} = b_m + d_m$ and

$$[\Omega_{RW}]_{mn} = \frac{r_n}{1+r_n}\delta_{m,n+1} + \frac{1}{1+r_n}\delta_{m,n-1} - \delta_{m,n}, \tag{S3}$$

with $r_n = b_n/d_n$. The rates D_{mm} determine how long the system waits in state m before taking a step and r_m is the relative probability of stepping forward versus backward along the chain. Defining $V = -\Omega_{RW}^{-1}$, the elements V_{ij} are the average number of visits to state i before absorption starting from an initial state j.

With the above decomposition we can easily invert the transition matrix,

$$[-\Omega^{-1}]_{ij} = \frac{V_{ij}}{b_i + d_i}, \tag{S4}$$

where visit numbers V_{ij} are given by

$$V_{ij} = (1+r_i) \sum_{n=1}^{\min(i,j)} \prod_{m=n}^{i-1} r_m. \tag{S5}$$

Then, using Eq. (S2) the moments can be expressed as

$$\mu_n = n! \sum_{i_1,i_2,\ldots i_n=1}^{N} \frac{V_{i_1 i_2} V_{i_2 i_3} \cdots V_{i_{n-1} i_n} V_{i_n N}}{(b_{i_1}+d_{i_1})(b_{i_2}+d_{i_2})\cdots(b_{i_n}+d_{i_n})}. \tag{S6}$$

To compute the cumulants, we use the standard conversion formulas: $\kappa_1 = \mu_1$, $\kappa_2 = \mu_2 - \mu_1^2$, $\kappa_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3$, and so on. Since the relation between cumulants and moments is polynomial, if we collect terms with common denominators it follows that the cumulants have the form quoted in the main text,

$$\kappa_n(\{r_j\}, N) = \sum_{1\le i_1\le i_2\le\cdots\le i_n\le N} \frac{w^n_{i_1 i_2\cdots i_n}(\{r_j\})}{(b_{i_1}+d_{i_1})\cdots(b_{i_n}+d_{i_n})}. \tag{S7}$$

where the weights w^n depend on the visit numbers V_{ij} (and hence are functions of only the ratios $\{r_j\}$). Note that we sum over $i_1 \le i_2 \le \cdots \le i_n$, so that each product in the denominator of Eq. (S7) appears exactly once. The weights are determined using Eq. (S6) and the moment-cumulant conversion formulas. For example, the second and third cumulants are

$$\kappa_2 = \sum_{i,j=1}^{N} \frac{2V_{ij}V_{jN} - V_{iN}V_{jN}}{(b_i+d_i)(b_j+d_j)} \tag{S8}$$

$$\kappa_3 = \sum_{i,j,k=1}^{N} \frac{6V_{ij}V_{jk}V_{kN} - 6V_{ij}V_{jN}V_{kN} + 2V_{iN}V_{jN}V_{kN}}{(b_i+d_i)(b_j+d_j)(b_k+d_k)}. \tag{S9}$$

From here we can read off the weights w^n: they are simply the numerators in the above expressions, summed over distinct permutations of the indices (since these terms have the same denominators). Carrying out the sum we obtain,

$$w^2_{ij} = \sum_{\sigma\in\Pi_2} 2V_{\sigma_1\sigma_2}V_{\sigma_2 N} - V_{\sigma_1 N}V_{\sigma_2 N} \tag{S10}$$

$$w^3_{ijk} = \sum_{\sigma\in\Pi_3} 6V_{\sigma_1\sigma_2}V_{\sigma_2\sigma_3}V_{\sigma_3 N} - 6V_{\sigma_1\sigma_2}V_{\sigma_2 N}V_{\sigma_3 N} + 2V_{\sigma_1 N}V_{\sigma_2 N}V_{\sigma_3 N}, \tag{S11}$$

where Π_2 is the set of distinct permutations of indices $\{i,j\}$ and Π_3 is the set of distinct permutations of $\{i,j,k\}$.

B. Properties of the weighting factors w^n

In the main text, we noted that the weights w^n have a few convenient properties. In particular, they are positive, $w^n_{i_1 i_2\cdots i_n}(\{r_j\}) \ge 0$ and increasing functions of each of the r_j. To show these properties, we use the fact that $V_{ii} = V_{ij}$

for any $i < j$. This is easy to see from Eq. (S5), but also has an intuitive physical interpretation. Since the system is eventually absorbed at the boundary state 0, if it starts from a state $j > i$ it must visit i before absorption. After the first visit, the statistics of the random walk are identical to a walk initialized in state i. Using this property the sum over permutations above dramatically simplifies. For $i < j$ we have

$$\begin{aligned} w^2_{ij} &= (2V_{ij}V_{jN} - V_{iN}V_{jN} + 2V_{ji}V_{iN} - V_{jN}V_{iN}) \\ &= 2V_{ji}V_{iN}. \end{aligned} \tag{S12}$$

Similarly, after simplification we find

$$w^3_{ijk} = 3!\,(V_{kj}V_{ji}V_{iN} + V_{ki}V_{iN}V_{jN}) \tag{S13}$$

$$\begin{aligned} w^4_{ijkl} = 4!\,(&V_{lk}V_{kj}V_{ji}V_{iN} + V_{lk}V_{ki}V_{iN}V_{jN} + V_{lj}V_{ji}V_{iN}V_{kN} \\ &+ V_{lj}V_{jN}V_{ki}V_{iN} + V_{li}V_{iN}V_{kj}V_{jN} + V_{li}V_{iN}V_{kN}V_{iN}) \end{aligned} \tag{S14}$$

when $i < j < k < l$. When some indices are identical, these results still hold, but they must be divided by the number of permutations of the identical indices, e.g. $w^2_{ii} = V_{ii}V_{iN}$ (notice this differs from Eq. (S12) by a factor of 2). The important feature of these expressions is that they are *positive* sums of products of the visit numbers V_{ij}. We conjecture that the weights at every order can also be written as a positive sums of products of the visit numbers (though we omit the expressions here, we have checked this is true up to order $n = 6$). If this is the case, it immediately follows that the weights w^n are positive and increasing functions of each r_j because the visit numbers, Eq. (S5), themselves also have these properties.

S3. ASYMPTOTIC ANALYSIS FOR THE GAUSSIAN UNIVERSALITY CLASS

A. Cumulant bounds

To estimate the asymptotics of the cumulants we start from Eq. (S7) derived above. Since the weights w^n are increasing functions of the r_j, we argued in the main text that

$$\kappa_n(r_{\min}, N) \le \kappa_n(\{r_j\}, N) \le \kappa_n(r_{\max}, N), \tag{S15}$$

where $r_{\max} = \lim_{N\to\infty} \max_{1<j<N} r_j$ and $r_{\min} = \lim_{N\to\infty} \min_{1<j<N} r_j$. The cumulants $\kappa_n(r_{\max}, N)$ and $\kappa_n(r_{\min}, N)$ correspond to auxiliary Markov chains where $b_j + d_j$ is unchanged, but $r_j = r_{\max}$ or $r_j = r_{\min}$ respectively.

Following Ref. [1], we provide asymptotic bounds on $\kappa(r, N)$ that lead to an analytic criterion for the Gaussian universality class. Since the diagonal elements of the weights w^n are greater than 1, we can bound the cumulant κ_n from below by a sum of the unweighted diagonal elements $(b_i + d_i)^{-n}$. To bound from above we can take the maximum value of $(b_i + d_i)^{-n}$ times the sum of the weighting factors. The sum over weighting factors $w^n_{i_1 i_2 \cdots i_n}(r)$ is precisely the n^{th} cumulant for a biased random walk (with $b_i + d_i = 1$ and uniform r). This sum can be computed exactly using eigenvalues of the transition matrix [1]. In particular, the sum is $\mathcal{O}(N)$ for any n as long as $r < 1$ and asymptotically can be represented in the integral form given below. Note that $r_{\max} < 1$ as long as $r_j < 1 - \epsilon$ for all j and some $\epsilon > 0$: this condition was the first requirement for the Gaussian universality class quoted in the main text. Altogether we have,

$$\begin{aligned} \sum_{n=1}^{N} \frac{1}{(b_i + d_i)^n} \le \kappa_n(r, N) &\le \left(\max_{1\le i\le N} \frac{1}{b_i + d_i}\right)^n \times \frac{N}{\pi} \int_0^\pi \frac{(n-1)!}{(1 - 2\sqrt{r}/(1+r)\cos x)^n} \mathrm{d}x \\ &= \left(\max_{1\le i\le N} \frac{1}{b_i + d_i}\right)^n \times \mathcal{O}(N). \end{aligned} \tag{S16}$$

We can now read off the second condition for the Gaussian universality class. To nail down the asymptotics of $\kappa_n(r, N)$ we want the upper and lower bounds in Eq. (S16) to have the same scaling for large N. Specifically, we require

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{(b_i + d_i)^n} \sim c_n \left(\max_{1\le i\le N} \frac{1}{b_i + d_i}\right)^n, \tag{S17}$$

for some N-independent constant c_n. Setting $n = 1$ in this equation leads to the condition Eq. (3) quoted in the main text. We can make this simplification because when Eq. (S17) is satisfied for $n = 1$, it is also satisfied for

$n > 1$. To see this fact, first note that $\langle (b_i + d_i)^{-n} \rangle < \max_i (b_i + d_i)^{-n}$ trivially. Furthermore, we can write the left hand side of Eq. (S17) as $N^{-1} ||(b+d)^{-1}||_p^p$, where $||\cdot||_p$ is the p-norm and $(b+d)^{-1}$ is the vector containing elements $(b_i + d_i)^{-1}$. Using p-norm inequalities, we have $N^{-1}||(b+d)^{-1}||_1 < N^{-1/n}||(b+d)^{-1}||_n$ for $n > 1$. Then if $c \cdot \max_i (b_i + d_i)^{-1} < \langle (b_i + d_i)^{-1} \rangle$ as $N \to \infty$ for some constant c, it follows that $c^n \max_i (b_i + d_i)^{-n} < \langle (b_i + d_i)^{-n} \rangle$ in this limit as well. Thus, it is sufficient to check Eq. (S17) holds for $n = 1$, since this implies the condition holds for all $n > 1$.

As discussed in the main text, the condition Eq. (S17) can be interpreted as the waiting times being 'flat' in the following sense: all (or at least a significant fraction) of the $(b_i + d_i)^{-1}$ are the same order as their maximum value. If this condition holds, then for large N we have that $\kappa_n(r, N) \sim c_n(r) f(N)^n N$ where $f(N) \sim \max_{1 \le i \le N} (b_i + d_i)^{-1}$ as $N \to \infty$. Since these asymptotics hold for both $r = r_{\min}$ and $r = r_{\max}$, it follows from Eq. (S15) that $\kappa_n(\{r_j\}, N) \sim c_n(\{r_j\}) f(N)^n N$ as well, possibly with a different constant $c_n(r_{\min}) < c_n(\{r_j\}) < c_n(r_{\max})$. These asymptotics imply that the higher-order cumulants are dominated by the variance and hence the distribution looks normal for large N, i.e. the standardized cumulants $\tilde{\kappa}_n = \kappa_n / \kappa_2^{n/2} \to 0$ as $N \to \infty$.

B. Leading correction to the Gaussian

The leading correction to the Gaussian distribution for finite N comes from the skew, $\tilde{\kappa}_3 = \tilde{c}_3 / \sqrt{N}$. Here we will give a bound on the magnitude of the skew, that can be used to predict when finite systems will have a nearly Gaussian absorption-time distribution. First, define

$$K_2 = \lim_{N \to \infty} \frac{1}{N f(N)^2} \sum_{i=1}^{N} \frac{1}{(b_i + d_i)^2}, \tag{S18}$$

where $f(N) \sim \max_{1 \le i \le N} (b_i + d_i)^{-1}$ as $N \to \infty$ as above. Then from Eq. (S16) that $\kappa_2 \ge K_2 N f(N)^2$. Evaluating the integral in Eq. (S16) we have $\kappa_3 \le 2 f(N)^3 N (r_{\max} + 1)^3 (r_{\max}^2 + 4 r_{\max} + 1)/(1 - r_{\max})^5$. Putting these together,

$$\tilde{\kappa}_3 \le \frac{2(r_{\max} + 1)^3 (r_{\max}^2 + 4 r_{\max} + 1)}{(1 - r_{\max})^5 K_2^{3/2}} \frac{1}{\sqrt{N}}. \tag{S19}$$

The convergence is slowest (i.e. the coefficient of $1/\sqrt{N}$ is large), when the conditions for the universality class are pushed to their limits: if the system is barely extinction-prone, $r_{\max} \approx 1$, or the waiting times are not very uniform, $K_2 \ll 1$ (the sum in Eq. (S17) is nearly dominated by the maximal term). Finally, we note that this is a rough upper bound; in many cases the convergence is much faster, e.g., if only a few $r_j \approx 1$ but the rest are very small. Replacing $r_{\max}$ with the average r_j in Eq. (S19) may often give a better estimate of the actual skew for a given system, even if it does not give a strict upper bound.

C. Transition matrix spectrum

As noted in the main text, we can also connect the spectrum of the transition matrix to the Gaussian universality class. Specifically, if Eq. (S17) is satisfied with $b_i + d_i$ replaced by the eigenvalues λ_i of the negative transition matrix $-\Omega$, the absorption-time distribution will be Gaussian. To show this, we use the spectral representation of the absorption-time cumulants [1, 3],

$$\kappa_n = (n-1)! \sum_{i=1}^{N} \lambda_i^{-n}. \tag{S20}$$

If Eq. (S17) is satisfied for the eigenvalues, we have

$$\sum_{i=1}^{N} \lambda_i^{-n} \sim c_n \left(\max_{1 \le i \le N} \lambda_i^{-1} \right)^n, \tag{S21}$$

Since the left-hand side of this expression is exactly the cumulant κ_n (up to the constant $(n-1)!$), it immediately follows that $\kappa_n \sim c_n g(N)^n N$ where $g(N) \sim \max_{1 \le i \le N} \lambda_i^{-1}$ as $N \to \infty$. Just as in the main text, this scaling implies that the standardized cumulants vanish for large N: $\tilde{\kappa}_n \sim \tilde{c}_n N^{1 - n/2}$ and the distribution is asymptotically Gaussian.

More generally, the distribution approaches a Gaussian as long as $\tilde{\kappa}_n \to 0$ as $N \to \infty$. This condition with Eq. (S20) describes a broader class of eigenvalue spectra that give rise to Gaussian absorption-time distributions. Specifically, we need

$$\left(\sum_{i=1}^{N} \lambda_i^{-n}\right) \Bigg/ \left(\sum_{i=1}^{N} \lambda_i^{-2}\right)^{n/2} \xrightarrow{N\to\infty} 0. \tag{S22}$$

While this condition is difficult to interpret, we consider two examples that illustrate the type of spectra that can give rise to Gaussian absorption-time distributions. First, if $\lambda_m = m^p$, the above condition is satisfied for $p \leq 1/2$. This result is related to the emergence of Gaussian distributions for the systems considered in Section S5.B, which have transition rates that decay as a power-law with $p \leq 1/2$. Also, if $\lambda_m = P(m)/Q(m)$ for some polynomials P and Q, the condition is satisfied when the degree of Q is greater than that of P.

S4. ASYMPTOTIC ANALYSIS FOR THE GUMBEL UNIVERSALITY CLASS

A. Cumulant bounds

For the Gumbel universality class we require $b_m + d_m = f(N)m[1 + \mathcal{O}(m/N)]$, $b_{\alpha N} + d_{\alpha N}$ be of order at least $\mathcal{O}(Nf(N))$ for any $0 < \alpha < 1$, and $r_m = r + \mathcal{O}(m/N)$ for large N. These properties are sufficient to guarantee that the absorption-time cumulants are asymptotic to those for an exactly solvable canonical model (for which the above equalities hold exactly, not just to leading order). Following Ref. [1], we restrict two of the indices in Eq. (S7) to be $\mathcal{O}(N)$ away from the absorbing state, $\alpha N \leq i_{n-1} \leq i_n \leq N$. With this restriction we can bound the sums,

$$\sum_{\substack{1\leq i_1\leq i_2\leq\cdots\leq i_{n-1}\\ \alpha N\leq i_{n-1}\leq i_n\leq N}} \frac{w^n_{i_1 i_2\cdots i_n}(\{r_j\})}{(b_{i_1}+d_{i_1})\cdots(b_{i_n}+d_{i_n})} \leq \frac{1}{f(N)^n N^2} \sum_{1\leq i_1\leq i_2\leq\cdots\leq i_n\leq N} w^n_{i_1 i_2\cdots i_n}(r). \tag{S23}$$

In the previous section, we established that the sum over the weighting factors is $\mathcal{O}(N)$, so this portion of the sum is $\mathcal{O}(f(N)^{-n}N^{-1})$. We now consider indices $1 < i_1 < \alpha N$ and $\alpha N < i_n < N$,

$$\sum_{\substack{i_1\leq i_2\leq\cdots\leq i_n\\ 1\leq i_1\leq \alpha N\leq i_n\leq N}} \frac{w^n_{i_1 i_2\cdots i_n}(\{r_j\})}{(b_{i_1}+d_{i_1})\cdots(b_{i_n}+d_{i_n})} \leq \frac{1}{f(N)^n N} \sum_{\substack{i_1\leq i_2\leq\cdots\leq i_n\\ 1\leq i_1\leq \alpha N\leq i_n\leq N}} w^n_{i_1 i_2\cdots i_n}(r). \tag{S24}$$

Since the weighting factors decay exponentially away from the diagonal elements, the sum on the right hand side of Eq. (S24) is $\mathcal{O}(1)$ and this portion of the sum is also $\mathcal{O}(f(N)^{-n}N^{-1})$.

Since the same bounds also apply for any other pair of the indices, the only remaining portion of the cumulant sum Eq. (S7) is that where all indices are near 0. Here the approximations that $b_m + d_m$ is linear and r_m is constant become asymptotically exact so that,

$$\sum_{1\leq i_1\leq i_2\leq\cdots\leq i_n\leq \alpha N} \frac{w^n_{i_1 i_2\cdots i_n}(\{r_j\})}{(b_{i_1}+d_{i_1})\cdots(b_{i_n}+d_{i_n})} \sim \frac{1}{f(N)^n} \sum_{1\leq i_1\leq i_2\leq\cdots\leq i_n\leq \alpha N} \frac{w^n_{i_1 i_2\cdots i_n}(r)}{i_1 i_2\cdots i_n}. \tag{S25}$$

The right hand side of Eq. (S25) is at least $\mathcal{O}(f(N)^{-n})$ and therefore this region of the cumulant sum dominates asymptotically compared to the $\mathcal{O}(f(N)^{-n}N^{-1})$ terms estimated above. In other words, the absorption process is entirely dominated by the coupon collection behavior near the absorbing state. Furthermore, we can freely extend the upper limit of the sum to N (instead of αN) since this will only add subdominant terms. Finally, we obtain the result quoted in the main text,

$$\kappa_n(\{r_j\}, N) \sim \frac{1}{f(N)^n} \sum_{1\leq i_1\leq i_2\leq\cdots\leq i_n\leq N} \frac{w^n_{i_1 i_2\cdots i_n}(r)}{i_1 i_2\cdots i_n}. \tag{S26}$$

Thus, for any Markov chain satisfying the conditions at the beginning of this section, the cumulants are asymptotic to those for the "canonical model" with $b_m + d_m = f(N)m$ and $r_m = r$ exactly. In Section S4.C we show this model has an asymptotically Gumbel absorption-time distribution.

B. Leading correction to the Gumbel

The leading correction $\delta\kappa_2$ to the standard deviation comes from the quadratic term in the transition rates, $b_m + d_m \approx f(N)m(1+m/N)$. Plugging this into Eq. (S26) for one set of rates in the denominator and using the partial fraction decomposition $1/i(j+j^2/N) = 1/ij + 1/i(j+N)$ leads to

$$\begin{aligned}\delta\kappa_2 &\sim \frac{1}{f(N)^2}\sum_{1\le i\le j\le \alpha N}\frac{w_{i,j}^2(r)}{i(j+N)} \\ &= \frac{1}{f(N)^2}\sum_{i=1}^{\alpha N}\frac{(1+r)^2(1-r^i)^2}{(1-r)^2 i(i+N)} + 2\sum_{i=1}^{j-1}\sum_{j=1}^{\alpha N}\frac{r^{j-i}(1+r)^2(1-r^i)^2}{(1-r)^2 i(j+N)}\end{aligned} \tag{S27}$$

where in the second line we make use of the fact that $r_m \approx r$ is approximately constant to write the explicit expression for $w_{i,j}^2$ (obtained from Eqs. (S5) and (S12)). The sums in the second line can be evaluated explicitly in terms of special functions, including harmonic numbers and the Lerch transcendent. The first sum is asymptotically dominant, leading to $\delta\kappa_2 \sim f(N)^{-2}N^{-1}\ln(N)$. More generally, the asymptotics above hold if $b_m + d_m \approx f(N)m(1+m/g(N))$ as long as the function $g(N)\to\infty$ as $N\to\infty$. An analogous calculation shows that for this case $\delta\kappa_2 \sim [\ln g(N)]/g(N)f(N)^2$. If $g(N)$ grows superlinearly, however, $[\ln g(N)]/g(N)f(N)^2$ is dominated by the corrections due to Eq. (S23) and (S24) computed above, leading to $\delta\kappa_2 \sim f(N)^{-2}N^{-1}$.

The higher-order cumulants can be analyzed in similar fashion. Since the weights decay exponentially away from the diagonal, the terms with $i_1 = i_2 = \cdots = i_n \equiv i$ is asymptotically dominant. For these elements $w^n_{i_1 i_2\cdots i_n}(\{r_j\}) = (n-1)!V_{ii}^n$ and it is straightforward to show that $\delta\kappa_n \sim N^{-1}f(N)^{-n}$. for the standardized cumulants $\tilde\kappa_n = \kappa_n/\kappa_2^{3/2}$, the factors of $f(N)$ in the asymptotics cancel and we are left with $\mathcal{O}(N^{-1}\ln N)$ corrections from the standard deviation. In other words, the deviations from the Gumbel cumulants scale like $\delta\tilde\kappa_n = \tilde\kappa_n - \tilde\kappa_n^{\text{Gumbel}} \sim C_n N^{-1}\ln N$ for large N. For the more general case, where the quadratic term in the rates is suppressed by $g(N)$, the scaling is $\delta\tilde\kappa_n \sim C_n[\ln g(N)]/g(N)$ for sublinear $g(N)$ and $\kappa_n \sim C_n/N$ otherwise.

C. Large-N limit for the canonical model

The canonical Markov model with coupon-collection behavior has $b_m + d_m = f(N)m$ and $r_m = r$. As noted in the main text, $f(N)$ simply sets the time scale for the process and does not affect the shape of the absorption-time distribution. Therefore, for convenience, we will rescale time $t \to t(r+1)/f(N)$ so that $b_m = rm$ and $d_m = m$. For this system, the absorption-time distribution $p(t)$ has been computed exactly using generating functions [4, Appendix B],

$$p(t) = \frac{Ne^{\nu t}\nu^2}{(e^{\nu t}-1)^2(1+\frac{\nu}{e^{\nu t}-1})^{N+1}} \tag{S28}$$

where $\nu = 1-r$. To derive the asymptotic form of the distribution we standardize to zero mean and unit variance. The standardized distribution is simply $\sigma p(\sigma t+\mu)$, where $\mu \sim (\ln N + \ln\nu + \gamma)/\nu$ and $\sigma \sim \pi/\nu\sqrt{6}$ are the mean and standard deviation of the absorption time. Here $\gamma \approx 0.5772$ is the Euler-Mascheroni constant. Plugging in this transformation and taking $N\to\infty$, we find

$$\sigma p(\sigma t+\mu) \xrightarrow{N\to\infty} \frac{\pi}{\sqrt{6}}\exp\left(-\gamma - \pi t/\sqrt{6} - e^{-\gamma-\pi t/\sqrt{6}}\right), \tag{S29}$$

which is precisely the standardized Gumbel distribution.

D. Transition matrix spectrum

The Gumbel distribution also arises if the transition matrix eigenvalues decay linearly. For instance, suppose $\lambda_m = bm$. Then, using Eq. (S20) and taking $N\to\infty$, we have that

$$\tilde\kappa_n = \left((n-1)!\sum_{m=1}^{\infty}(bm)^{-n}\right)\Big/\left(\sum_{m=1}^{\infty}(bm)^{-2}\right)^{n/2} = (n-1)!\frac{\zeta(n)}{\zeta(2)^{n/2}}, \tag{S30}$$

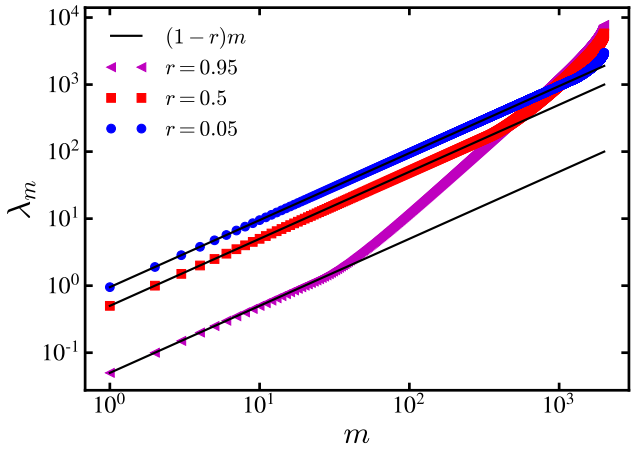

FIG. S1. The eigenvalues of the transition matrix for the canonical model $b_m = rm$, $d_m = m$ with $N = 2000$ and $r = 0.05$, 0.5, and 0.95 plotted on a log-log scale. The black lines show $(1-r)m$ for each value of r. The eigenvalues closely follow this linear relation up to a cut-off m_c that is dependent on r. Since the leading eigenvalues are linear the absorption-time distribution is Gumbel.

which are precisely the cumulants for a standardized Gumbel distribution. The result is unchanged if the dominant eigenvalues are approximately linear, i.e. $\lambda_m \approx bm$ for $m < \alpha N$ where α is a constant $0 < \alpha < 1$. In this case, the standardized cumulants are still $\tilde{\kappa}_n = (n-1)!\zeta(n)/\zeta(2)^{n/2}$ with the larger eigenvalues contributing $\mathcal{O}(1/N)$ corrections that vanish asymptotically.

This second case appears to be what happens in practice: for $N \gg 1$ the eigenvalues become linear up to a cutoff. We have carried out numerical calculations of the spectrum for the canonical model from the previous section, $b_m = rm$, $d_m = m$ for a few values of $r < 1$. As shown in Figure S1, the leading eigenvalues in the spectrum become equally spaced: $\lambda_m \approx (1-r)m$ for indices below a cutoff m_c. Numerical tests indicate m_c is approximately a constant proportion of N, i.e. $m_c \approx \alpha(r)N$, where $0 < \alpha(r) < 1$. Above this cutoff the eigenvalues grow super-linearly. The above calculation illustrates how the Gumbel absorption-time distribution arises in this model from the perspective of the eigenvalue spectrum.

S5. ASYMPTOTIC ANALYSIS FOR THE POWER-LAW PROCESSES

In this section we generalize the Gumbel criteria discussed above. Consider Markov processes with transition rates that satisfy $b_m + d_m = f(N)m^p[1 + \mathcal{O}(m/N)]$. Moreover, suppose that $b_{\alpha N} + d_{\alpha N}$ is of order at least $\mathcal{O}(N^p f(N))$ for any $0 < \alpha < 1$, and $r_m = r + \mathcal{O}(m/N)$ for large N. In other words, this process has transition rates that vanish as a power-law m^p near the boundary. In the main text we claimed that $p \leq 1/2$ gives rise to Gaussian absorption times, while $p > 1/2$ leads to a skewed family of distributions (whose shape depends on the parameters p and r). We rigorously justify these claims in the following subsections.

A. Skewed family for $p > 1/2$

When $p > 1/2$ the transition rates decay quickly enough that the process is dominated by slowdown near the boundary (similar to coupon collection), giving rise to skewed distributions. To analyze this case, we can apply similar asymptotic analysis to that given in Section S4.A for the Gumbel class. Repeating the bounds in Eqs. (S23) and (S24) for the power-law process, we find that

$$\sum_{\substack{1\leq i_1\leq i_2\leq\cdots\leq i_{n-1}\\ \alpha N\leq i_{n-1}\leq i_n\leq N}} \frac{w^n_{i_1 i_2\cdots i_n}(\{r_j\})}{(b_{i_1}+d_{i_1})\cdots(b_{i_n}+d_{i_n})} = \mathcal{O}(f(N)^{-n}N^{1-2p}) \tag{S31}$$

$$\sum_{\substack{i_1\leq i_2\leq\cdots\leq i_n\\ 1\leq i_1\leq\alpha N\leq i_n\leq N}} \frac{w^n_{i_1 i_2\cdots i_n}(\{r_j\})}{(b_{i_1}+d_{i_1})\cdots(b_{i_n}+d_{i_n})} = \mathcal{O}(f(N)^{-n}N^{-p}). \tag{S32}$$

As long as $p > 1/2$, these terms are each asymptotically dominated by the indices near 0,

$$\sum_{1\le i_1\le i_2\le\cdots\le i_n\le \alpha N} \frac{w^n_{i_1 i_2\cdots i_n}(\{r_j\})}{(b_{i_1}+d_{i_1})\cdots(b_{i_n}+d_{i_n})} \sim \frac{1}{f(N)^n} \sum_{1\le i_1\le i_2\le\cdots\le i_n\le \alpha N} \frac{w^n_{i_1 i_2\cdots i_n}(r)}{i_1^p i_2^p\cdots i_n^p}, \tag{S33}$$

which are at least of order $\mathcal{O}(f(N)^{-n})$. Similar to the Gumbel class, the absorption process is dominated by the slow behavior near the absorbing state, where the transition rates decay. Extending the upper limit on the sum from αN to N (which only adds subdominant terms), we find that the cumulants κ_n satisfy

$$\kappa_n(\{r_j\},N) \sim \frac{1}{f(N)^n} \sum_{1\le i_1\le i_2\le\cdots\le i_n\le N} \frac{w^n_{i_1 i_2\cdots i_n}(r)}{i_1^p i_2^p\cdots i_n^p}. \tag{S34}$$

Notice that this asymptotic formula for the cumulants is identical to Eq. (4) in the main text, but with the denominator $(i_1 i_2\cdots i_m)$ raised to the power p. Thus, we have shown the absorption-time cumulants for a general Markov process, with $b_m+d_m = f(N)m^p[1+\mathcal{O}(m/N)]$, are asymptotic to those for the minimal model $b_m = rd_m = rm^p$ (after rescaling time so that $f(N) = r+1$). The absorption-time distributions for the minimal model were explored numerically in Figs. 1(d) and 2 in the main text. For $p > 1/2$, we find a family of distributions that become more skewed as p increases. The shape of the distributions depends subtly on r except when $p = 1$, where the distribution is Gumbel, as revealed by our analysis above.

B. Gaussian distributions for $p \le 1/2$

To show the normality of the absorption-time distribution for $p \le 1/2$, we show that the variance κ_2 diverges at least as fast as the higher-order cumulants. Using the asymptotic estimate from the previous section, we have that

$$\kappa_n(\{r_j\},N) \sim \frac{1}{f(N)^n} \sum_{1\le i_1\le i_2\le\cdots\le i_n\le N} \frac{w^n_{i_1 i_2\cdots i_n}(r)}{i_1^p i_2^p\cdots i_n^p} + \mathcal{O}(f(N)^{-n}N^{1-2p}). \tag{S35}$$

As noted above, when $p < 1/2$, we can not guarantee that the first term is dominant. If the second term is dominant than the cumulants scale like $\kappa_n \sim c_n f(N)^{-n}N^{1-2p}$. Then the standardized cumulants asymptotically vanish, $\tilde{\kappa}_n = \kappa_n/\kappa_2^{n/2} \propto N^{(2-n)(1-2p)/2} \to 0$ as $N\to\infty$. Hence, the absorption-time distribution is asymptotically normal. On the other hand, if the first term in Eq. (S35) is dominant, we can show the distribution is still Gaussian. Since the weight factors fall off exponentially away from the diagonal, the diagonal terms are asymptotically dominant. Using the fact that $w^n_{i_1 i_2\cdots i_n}(\{r_j\}) = (n-1)!V^n_{ii} \ge 1$, when $i_1 = i_2 = \cdots = i_n \equiv i$ together with Eq. (S5), we have

$$\kappa_n(\{r_j\},N) \sim \frac{(1+r)^n}{(1-r)^n f(N)^n} \sum_{i=1}^{N} \frac{(1-r^i)^n}{i^{np}}. \tag{S36}$$

Notice that when $p = 1/2$, the sum in this expression diverges as $N\to\infty$ for the variance ($n = 2$), but converges for the higher-order cumulants ($n > 2$). More generally, for any $p \le 1/2$, it is straightforward to show that the sum in Eq. (S36) always diverges faster with N for the variance than for the higher-order cumulants. As above, this scaling leads to $\kappa_n/\kappa_2^{n/2} \to 0$ for large N, so that the distribution asymptotically approaches a Gaussian. Figures 1(d) and 2 in the main text show this result is confirmed in numerical simulations: the distribution for $p \le 1/2$ looks approximately normal and the skew approaches 0.

S6. EXAMPLE MODELS

In this section we provide details of the evolutionary game, SIS, logistic, and autocatalytic chemical reaction models, each of which exhibit Gaussian, Gumbel, or exponential absorption-time distributions in different parameter regimes. Parameters used for the simulations presented in the main text are provided in Table S1.

A. Evolutionary games

In the main text, we present absorption-time distributions measured via simulations of a two-strategy evolutionary game. In this game, two types of individuals, mutants (M) and wild-types (W), compete and have *frequency dependent*

TABLE S1. Parameter choices for the simulations used to measure absorption-time distributions shown in Figures 1(a)-(c) and 3(b) in the main text. See Section S6 for model and parameter definitions. Evolutionary games use well-mixed population structure except in Figure 1(a). In Figure 1(c) the relative weighting of the convolution of Gumbel distributions is $s = (1 + e^{\beta(c-a)})/(1 + e^{\beta(b-d)}) \approx 0.73$ for both sets of parameters.

Figure	Model	Parameters
1(a)	1D Evolutionary Game	$\beta = 1$, $a = 2$, $b = 4$, $c = 1$, $d = 0.1$
1(b)	SIS Model	$\Lambda = 0.5$
	Logistic Model	$B = 0.5$, $K = 1$
	Chemical Reaction Model	$k_1 = 1$, $k_2 = 0.75$, $k_3 = 1.25$
1(c)	Evolutionary Game (black circles)	$\beta = 1$, $a = 1$, $b = 0.5$, $c = 0.8$, $d = 0.1$
	Evolutionary Game (red circles)	$\beta = 2$, $a = 0.3$, $b = 1.3$, $c = 0.06$, $d = 1.2$
3(b)	SIS Model	$\Lambda = 1.4$
	Logistic Model	$B = 1.4$, $K = 1$
	Chemical Reaction Model	$k_1 = 1$, $k_2 = 1.35$, $k_3 = 0.14$
	Evolutionary Game	$\beta = 1$, $a = 1$, $b = 1.5$, $c = 1.2$, $d = 1$

fitness, which means that an individual's fitness depends on the identity of its neighbors. This dependence is encoded by the payoff matrix,

	M	W
M	a	b
W	c	d

(S37)

For example, a mutant (M) with 2 mutant neighbors and 3 wild-type neighbors will have payoff $\pi = 2a + 3b$. The fitness is then $\exp(\beta\pi)$, where the parameter β, the selection intensity, controls how strongly payoff influences fitness. This choice is known as the exponential fitness mapping [3]; we note that other fitness mappings do not change the qualitative behavior discussed below. The dynamics of the model are as follows: an individual is chosen randomly, proportional to their fitness. The selected individual gives birth to an offspring of the same time (M or W) which replaces a random neighbor (selected uniformly). We will let m denote the number of wild-types in the population. Thus, when $m = 0$, the mutants have taken over the population (in the jargon, the mutation becomes fixed). We focus on cases in which the mutation becomes fixed, ignoring those when the mutation dies out (which have infinite absorption time). We consider evolution in two types of network populations: a one-dimensional (1D) ring of individuals and a well-mixed (complete graph) population. Each exhibits different absorption-time behavior.

1. 1D ring population structure

First we consider individuals connected in a 1D periodic ring [5]. Assuming a single initial mutant, the mutant population grows as a connected chain. Any changes in the population must occur at the boundary between mutants and wild-types. The boundary mutants and wild-types have payoff $a+b$ and $c+d$ respectively (they have one of each type as a neighbor). Thus the probability b_m of removing a mutant, and the probability d_m of adding a mutant, are given by

$$b_m = e^{\beta(c+d)}/F_m\,, \qquad d_m = e^{\beta(a+b)}/F_m\,, \qquad \text{for } 1 < n < N-1, \tag{S38}$$

where F_m is the average fitness:

$$F_m = 2e^{\beta(a+b)} + (N-m-2)e^{\beta 2a} + 2e^{\beta(c+d)} + (m-2)e^{\beta 2d}. \tag{S39}$$

The rates are slightly different for $m = 1$ and $m = N-1$ [5]. For example, when $m = 1$ there is a single wild-type with two mutant neighbors. These transition rates are:

$$b_1 = \frac{e^{\beta 2c}}{2e^{\beta(a+b)} + (N-3)e^{\beta 2a} + e^{\beta 2c}} \qquad d_1 = \frac{e^{\beta(a+b)}}{2e^{\beta(a+b)} + (N-3)e^{\beta 2a} + e^{\beta 2c}} \tag{S40}$$

$$b_{N-1} = \frac{e^{\beta(c+d)}}{e^{\beta 2b} + 2e^{\beta(c+d)} + (N-3)e^{\beta 2d}} \qquad d_{N-1} = \frac{e^{\beta 2b}}{e^{\beta 2b} + 2e^{\beta(c+d)} + (N-3)e^{\beta 2d}}. \tag{S41}$$

For large N, however, these changes to the transition rates do not affect the absorption-time distribution. One can check that these transition rates satisfy the requirements of the Gaussian universality class if $(a+b) > (c+d)$.

2. *Well-mixed population*

If the population is well-mixed, every individual has contact with every other, and hence their fitness depends simply on the fraction of mutants in the population. The payoffs (per contact) for mutants and wild-types respectively are $\pi_M = a(N-m-1)/(N-1) + bm/(N-1)$ and $\pi_W = c(N-m)/(N-1) + d(m-1)/(N-1)$, where again a, b, c, and d are elements of the payoff matrix Eq. (S37) and m is the number of wild-types in the population. The rates at which the wild-type population increases or decreases are [3]

$$b_m = \frac{m\, e^{\beta\pi_W}}{m\, e^{\beta\pi_W} + (N-m)\, e^{\beta\pi_M}} \frac{(N-m)}{N-1}, \qquad d_m = \frac{(N-m)\, e^{\beta\pi_M}}{m\, e^{\beta\pi_W} + (N-m)\, e^{\beta\pi_M}} \frac{m}{N-1}. \tag{S42}$$

For the birth (death) rate the first fraction represents the probability of choosing a wild-type (mutant) to give birth, while the second fraction is the probability of the offspring replacing a mutant (wild-type) in the populations.

Probability flows toward the absorbing state ($r_m < 1$) if $b > d$ and $a > c$. From the transition probabilities it is clear $b_m + d_m$ decays linearly near both $m = 0$ and $m = N$. Expanding around these points, $b_m + d_m = f(N)(N-m) + \mathcal{O}(m)$ and $b_m + d_m = \tilde{f}(N)(N-m) + \mathcal{O}((N-m)^2)$. Our theory predicts the distribution will be a convolution of Gumbel distributions with relative weighting $s = \lim_{N\to\infty} f(N)/\tilde{f}(N)$. Taking this limit for the transition probabilities in Eq. (S42) we find

$$s = \frac{1 + e^{\beta(c-a)}}{1 + e^{\beta(b-d)}}. \tag{S43}$$

B. SIS model

The stochastic susceptible-infected-susceptible (SIS) model of epidemiology [6] describes the spread of an infectious disease through a population. The population is broken into two groups, those susceptible to the disease and those currently infected. This model describes diseases that do not confer immunity following recovery (or the immunity lasts only for a short time compared to the time scale on which the disease spreads). The rate per contact at which the disease is transmitted between individuals is Λ/N, and we set the time-scale so that the recovery rate is 1. Letting m represent the number of infected individuals, there are $m(N-m)$ contacts between infected and susceptible people in a well-mixed population. The m infected individuals each recover at rate 1. Thus the rates at which the infected population increases and decreases are respectively

$$b_m = \Lambda m(1 - m/N), \qquad d_m = m. \tag{S44}$$

This system has the vanishing transition probabilities near $m = 0$, indicating coupon collection behavior (it is also straightforward to explicitly check it satisfies our requirements for the universality class as long as $\Lambda < 1$). Our simulations show that it has the expected Gumbel distribution of times for the infection to die out (Fig. 1(b) in the main text).

C. Logistic model

The stochastic logistic model describes the dynamics and fluctuations of an ecological population [7]. The model assumes a constant birth rate B per individual as well as a constant death rate (which we set to 1 by choosing the appropriate time scale) when the population is sparse. For higher populations, competition between individuals increases the death rate quadratically. The transition rates are

$$b_m = Bm, \qquad d_m = m + Km^2/N, \tag{S45}$$

where the parameter K controls how strongly competition influences death rates (this parameter is related to the carrying capacity of the ecosystem). Again the transition rates vanish linearly near $m = 0$ and this model belongs to the Gumbel universality class as long as $B < 1$.

D. Autocatalytic chemical reaction model

Our final example model describes a stochastic autocatalytic chemical reaction [8],

$$2X + A \underset{k_2}{\overset{k_1}{\rightleftharpoons}} 3X, \qquad X \xrightarrow{k_3} B, \tag{S46}$$

where k_i are the reaction rates. The concentrations of species A and B are fixed at saturation levels and we want to describe the dynamics and fluctuations of m, the number of particles of species X. This is a variation of the Schlögl model where the reaction $X \to B$ is irreversible and the reactions cease when no particles of X remain. Applying our results to this model we will classify the distribution of reaction times: how long does the reaction proceed before the supply of X is exhausted.

The birth-death transition rates for the reaction given above are

$$b_m = \frac{k_1}{N} m(m-1), \qquad d_m = \frac{k_2}{N^2} m(m-1)(m-2) + k_3 m. \tag{S47}$$

Again, the transition rates decay linearly near the absorbing boundary at $m = 0$, indicating the Gumbel universality class; it is straightforward to check that the required expansions hold. The conditions that guarantee $r_m = b_m/d_m < 1$ are more intricate. In particular, if $k_3 > k_2$, then $r_m < 1$ as long as $k_1 < k_2 + k_3$. On the other hand, if $k_3 < k_2$ we need $k_2 < 2\sqrt{k_2 k_3}$. With either of these conditions satisfied the autocatalytic reaction times will be Gumbel distributed. Note that this model has an infinite state space: the number of X particles can be any positive integer. As discussed in the main text, we expect our results to apply to this class of models as long as the initial condition is large. The simulation shown in Figure 1(b) indicates this expectation is indeed borne out.

[1] D. Hathcock and S. H. Strogatz, Phys. Rev. E **100**, 012408 (2019).
[2] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, Vol. 1 (Elsevier, 1992).
[3] P. Ashcroft, *The Statistical Physics of Fixation and Equilibration in Individual-Based Models* (Springer, 2016).
[4] S. Azaele, S. Suweis, J. Grilli, I. Volkov, J. R. Banavar, and A. Maritan, Rev. Mod. Phys. **88**, 035003 (2016).
[5] P. M. Altrock, A. Traulsen, and M. A. Nowak, Phys. Rev. E **95**, 022407 (2017).
[6] J. A. Jacquez and C. P. Simon, Mathematical Biosciences **117**, 77 (1993).
[7] J. Grasman and R. HilleRisLambers, Ecological Modelling **103**, 71 (1997).
[8] C. R. Doering, K. V. Sargsyan, L. M. Sander, and E. Vanden-Eijnden, Journal of Physics: Condensed Matter **19**, 065145 (2007).