ETH zürich EPFL

Master thesis - MSc Physics

The hidden perils of (over-)control in complex systems

Author:
Charles Mitais

Supervisors:
Pr. Didier Sornette (ETHZ)
Pr. Sandro Lera (SUSTech)
Pr. Paolo De Los Rios (EPFL)

Start date:
27.09.2021

End date:
25.03.2022

ABSTRACT

Eurich and Pawelzik [1] and Patzelt et al. [2] proposed a model of closed-loop human motor control. The algorithm uses "optimal" parameter estimation based on past observations to predict next deviation. Long tails observed in experiments are reproduced by this model, questioning its optimality. We show that when control has access to few past observations only, deviations from the origin are power law distributed: $P(z) \sim z^{-(1+\mu)}$. Indeed, optimality is based on the properties of Maximum Likelihood Estimator, which requires a sufficiently large amount of data to estimate parameters precisely. If it is not the case, we show that shrinking deviations leads the estimator to diverge, such that the resulting system self-organizes into a critical regime. The exponent $\mu = 1$ is derived analytically in the extreme case where control has access to one past position only. By means of asymptotic theory we relate qualitatively the exponent μ to the length of memory. With increasing memory, control eventually becomes optimal in the sense that final distribution converges to noise distribution. This convergence is faster than the central limit theorem (CLT), as extreme deviations vanish for $n \gtrsim 10$ past observations. To prevent criticality with very short memory controllers, we propose a natural modification of the algorithm that simply replaces the estimator by a constant when the ratio of two consecutive steps exceeds a given threshold. For well-chosen parameters this regularization tames wild fluctuations and produces normally distributed time series. The choice of parameters is however not straightforward. Reducing the number of free parameters down to one gives similar results and allows optimizing numerically its value more easily.

CONTENTS

1

INTRODUCTION

Complex systems, sometimes described intuitively as "wholes that are more than the sum of their parts", are systems where large collections of locally dependent elements may spontaneously self-organize to produce non-trivial collective phenomenon. Those systems exhibit extreme events much more frequently than expected with normally distributed sequences. Their tail distribution is described by power laws $P(z) \sim z^{-(1+\mu)}$, which present very specific properties. They reflect self-similar structures thanks to scale invariance and, more importantly here, posses a well-defined average (resp. variance) only for $\mu > 1$ (resp. $\mu > 2$).

Such distributions appear in numerous natural and artificial phenomenon. In physics for example, they describe the behavior of order parameters at the critical points for second order phase transitions. Alternatively, Beggs and Plenz [3] demonstrated that propagation of spontaneous activity in cortical networks is described by the same equations than avalanches, with exponent $\mu = -1/2$. In economics, Pareto [4] observed that wealth distribution is heavy tailed, introducing the now famous $80 - 20$ principle. Few years later, Estoup [5] remarked that the rank frequency distribution of words also follows a power law, later formalized by Zipf [6]. In particular, experiments of visuomotor control of a virtual target in Bormann et al. [7] demonstrated that fluctuations are power law distributed. By studying the balancing of a stick at the fingertip Cabrera and Milton [8] found the same distribution for fluctuations around some target value, characteristic of on-off intermittency.

To model those experiments, a simplified one dimensional control algorithm employing "optimal" estimation of an unknown excitation parameter based on past observations was proposed by Eurich and Pawelzik [1]. Resulting system self-organizes into a critical regime, producing heavy tailed fluctuations. This scaling behavior was attributed to multiplicative noise as a result of optimal parameter estimation.

Taking this model as starting point we investigate the validity of control methodology, in particular it's so-called optimality, to explain the deep origin of the power law distributed fluctuations. Besides, we study the complex dependence of final distribution on the length of the observation period.

Optimal parameter estimation takes its names from the use of well-known Maximum Likelihood Estimator (MLE), which indeed yields very good results when the number of samples is large, thanks to the Central Limit Theorem (CLT, Cox and Hinkley [9]). In our case, when memory is short (typically less than 10 past observations), those properties clearly do not hold. The estimator involves sample variance of the time series on which the control is learned at its denominator. When fluctuations get close to the origin, the denominator thus goes to 0 and the estimator diverges. For the extreme case with shortest memory, we re-derive the exponent $\mu = 1$ already computed analytically in Patzelt et al. [2] and recover the postulated linear relation between the length of observation n with the scaling exponent μ for small n.

However, with more memory numerical simulations demonstrate that this control strategy generates normally distributed time series with zero mean and same variance as noise; which defines of optimal control. In this context the random map takes the form of a Kesten process, a mixture of multiplicative and additive processes, which are known to produce power law distributed sequences (Kesten [10]). This asymptotic theory provides a qualitative way to relate the tail exponent of the distribution with memory length. As power laws, Kesten processes arise in various fields, including economics [11] or physics of disordered systems [12]. More interestingly here, Statman et al. [13] applied it to model qualitatively synaptic size dynamics. They were able to recover the shape and long-term stability of experimentally observed synaptic size distributions.

Finally, we propose a natural modification of the control strategy for the very short memory case, by replacing the original estimator with a constant when the time series starts diverging. Despite its apparent naivety, adding this regularization prevents extreme deviations for well-chosen parameters.

2

INITIAL MODEL

Consider a one dimensional time-discrete random map

$$y_{t+1} = \alpha_0 y_t + \beta_t, \tag{1}$$

where the dynamical variable y_t denotes the deviation of a system from some target value at time t ($t = 0, 1, \dots$). α_0 is a system parameter unknown to the controller and assumed to be a constant larger than 1, therefore making the system unstable. $\beta_t \sim \mathcal{N}(0, \sigma_0^2)$ is a Gaussian random variable describing non-predictable fluctuations; its variance $\sigma_0^2 = \text{const}$ is a second hidden parameter.

The controller is assumed to know the form of the dynamical equation (1). A control strategy consists in computing an estimate $\hat{\alpha}_0^{(n)}$ of α_0 from y_t and n previous observations $y_{t-1}, y_{t-2}, \dots y_{t-n}$. At each time step the controller removes the predicted deviation $c_t y_t$ to the dynamics (1), with $c_t = \hat{\alpha}_0^{(n)}$. When control is switched on, the map is replaced by

$$z_{t+1} = (\alpha_0 - c_t) z_t + \beta_t, \tag{2}$$

where z_t denotes the controlled dynamics. Control is said optimal if controlled dynamics (2) follows noise distribution.

Estimates $\hat{\alpha}_0^{(n)}$ are be obtained by Maximum Likelihood Estimation (MLE), as in [1]. The likelihood is the joint probability of observing z_t and the past realizations $\{z_{t-i-1}\}_{i=0}^{n-1}$ given $\{c_{t-i-1}\}_{i=0}^{n-1} \equiv c_i^n$, as a function of α_0:

$$\mathcal{L}(\{z_{t-i-1}\}_{i=0}^{n-1}; \alpha_0, c_i^n) \equiv P(z_t, z_{t-1}, \dots, z_{t-n} | \alpha_0; c_i^n). \tag{3}$$

Iterating Bayes' law gives

$$\mathcal{L}(\{z_{t-i-1}\}_{i=0}^{n}; \alpha_0, c_i^n) = \prod_{i=0}^{n-1} P(z_{t-i} | z_{t-i-1}, \dots z_{t-n}; \alpha_0, c_i^n) P(z_{t-n} | \alpha_0; c_i^n), \tag{4}$$

and the log likelihood is

$$l(\{z_{t-i-1}\}_{i=0}^{n}; \alpha_0, c_i^n) = \sum_{i=0}^{n-1} \log\big(P(z_{t-i} | z_{t-i-1}, \dots z_{t-n}; \alpha_0, c_i^n)\big) + \log\big(P(z_{t-n} | \alpha_0; c_i^n)\big). \tag{5}$$

Next we drop the last term by assuming that without knowing previous deviations, distribution of z_{t-n} is independent of α_0: $\partial_{\alpha_0} P(z_{t-n} | \alpha_0; c_i^n) = 0$. Indeed, without knowledge of the starting point z_{t-1}, z_t could be anywhere, irrespective of α_0 and c_{t-1} (this assumption is numerically validated in [1]). The estimator of α_0 is then

$$\hat{\alpha}_0^{(n)} = \underset{\alpha_0}{\operatorname{argmax}} \sum_{i=0}^{n-1} \log(P(z_{t-i} | z_{t-i-1}, \dots z_{t-n}; \alpha_0, c_i^n)). \tag{6}$$

Given (2), the dynamic is conditionally Gaussian:

$$P(z_{t-i} | z_{t-i-1}, \dots z_{t-n}, \alpha_0; c_i^n) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}\left(z_{t-i} - (\alpha_0 - c_{t-i-1}) z_{t-i-1}\right)^2}. \tag{7}$$

Plugging into (6):

$$\hat{\alpha}_0^{(n)} = \underset{\alpha_0}{\operatorname{argmin}} \left[\frac{1}{2\sigma_0^2} \sum_{i=0}^{n-1} \left(z_{t-i} - (\alpha_0 - c_{t-i-1}) z_{t-i-1} \right)^2 + \frac{1}{2} \log\left(2\pi\sigma_0^2\right) \right]. \tag{8}$$

The first order condition reads

$$0 = \frac{\partial}{\partial \alpha_0}\bigg|_{\hat{\alpha}_0^{(n)}} \log(P(z_{t-i}|z_{t-i-1}, \dots z_{t-n}; \alpha_0, c_i^n)) = \frac{1}{\sigma_0^2} \sum_{i=0}^{n-1} z_{t-i-1} z_{t-i} - \alpha_0 z_{t-i-1}^2 + c_{t-i-1} z_{t-i-1}^2, \tag{9}$$

and finally the estimator is

$$\hat{\alpha}_0^{(n)} = \frac{\sum_{i=0}^{n-1} c_{t-i-1} z_{t-i-1}^2 + z_{t-i-1} z_{t-i}}{\sum_{i=0}^{n-1} z_{t-i-1}^2}. \tag{10}$$

MLE estimator is asymptotically normal [9]: $(\hat{\alpha}_0^{(n)} - \alpha_0) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, \frac{1}{n}\mathcal{I}^{-1}\right)$, where

$$\mathcal{I} = \mathbb{E}[-\partial_{\alpha_0}^2|_{\hat{\alpha}_0^{(n)}} \log\left(P(z_{t-i}|z_{t-i-1}, \dots z_{t-n}; \alpha_0, c_i^n)\right)] = \frac{1}{\sigma_0^2} \left\langle \sum_{i=0}^{n-1} z_{t-i-1}^2 \right\rangle \tag{11}$$

is the Fischer information. It measures the amount of information that $\{z_{t-i-1}\}_{i=0}^n$ carries about α_0. The variance of $\hat{\alpha}_0^{(n)}$ (see section A.1 for a more detailed derivation) is then

$$s_n^2 = \frac{\sigma_0^2}{n^2 \left\langle \sum_{i=0}^{n-1} z_{t-i-1}^2 \right\rangle}. \tag{12}$$

When n is sufficiently large, $\langle z_t^2 \rangle$ is well defined such that (12) becomes $s_n^2 = \sigma_0^2/(n^2 \hat{\sigma}_z^2)$ and vanishes with n. Fluctuations of long memory systems are then normally distributed, with variance σ_0^2:

$$z_{t+1} \equiv (\alpha_0 - c_t) z_t + \beta \underset{n\gg 1}{\sim} \mathcal{N}\left(0, \sigma_0^2\right). \tag{13}$$

2.1 NUMERICAL RESULTS

Numerical solution off $P(z_t)$ for $n = 50$ approaches noise (Figure 1), with $\langle z_t \rangle = 8.74 \cdot 10^{-4}$ and $\langle z_t^2 \rangle = 0.81$. To test Goodness-of-Fit we compute first the Kolmogorov-Smirnov (KS) statistics [14], which tests the null hypothesis that $z_t \sim \mathcal{N}(0, \sigma_0^2)$. With the numerical solution, KS statistics is $1.19 \cdot 10^{-3}$ with $p-$value 0.53. The null hypothesis is therefore acceptable. We also present results from Anderson-Darling test gives more weight to the tail [15], which is relevant in our case. The simulation yields a critical value $D = 0.441$, inferior to the critical value with significance level 15% for a normal distribution, 0.576 [16]. Again, normality assumption is accepted.

For very small n, those properties are not guaranteed. The variance of the estimator (12) $\sigma_0^2/(n^2 \langle z_t^2 \rangle)$ diverges as $z_t \to 0$ because Fischer information (knowledge about α_0 in the past observations) decreases as control gets better. Trajectory z_t simulated with $n = 1$ (Figure 2) exhibits extreme deviations, with $\max_t\{z_t\} = 3.29 \cdot 10^5$ while last percentile is $\mathcal{Q}(99\%) = 54.52$. The difference between them enlightens the presence of a heavy tail. Even though 83.3% of the

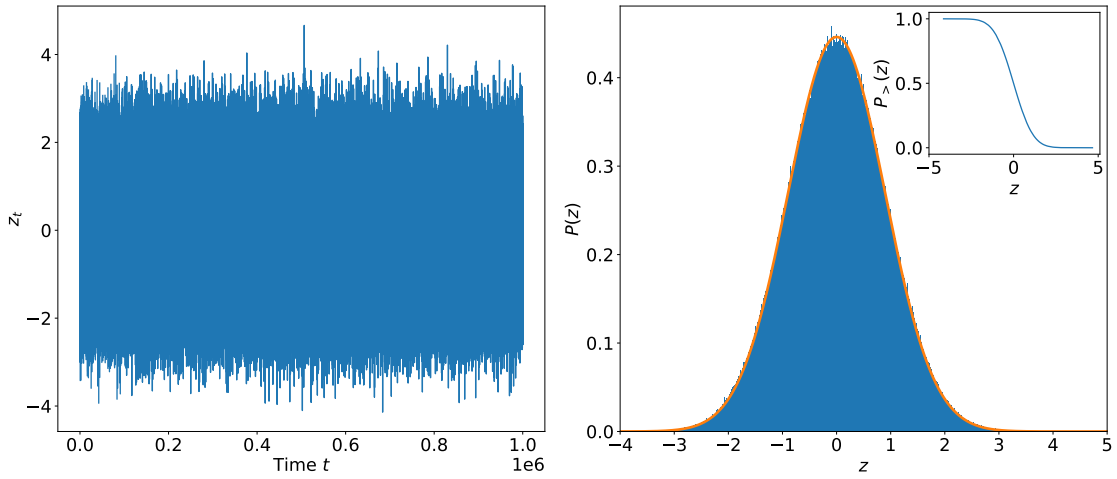

Figure 1: Left: solution of (2) with controller (10) and $n = 50$ look-backs. Right: associated histogram, compared with $P_\beta \sim \mathcal{N}(0, \sigma_0^2)$ (orange) and complementary cumulative distribution function (CCDF, inset). $N = 10^6$ time-steps, $\alpha_0 = 2$, $\sigma_0^2 = 0.8$, $z_0 = 1$. $\langle z \rangle = 8.74 \cdot 10{-}4$ and $\langle z^2 \rangle = 0.81$.

fluctuations are within the expected range for a normal distribution, remaining tail dominates the statistics. Besides, the complementary cumulative distribution function (CCDF, Figure 2) demonstrate a power-tail with exponent $\mu = 1$ as suggested in [1, 2]:

$$P(z) = \frac{C}{z^{1+\mu}}, \quad P_>(z) = P(Z > z) = \frac{C}{z^\mu}. \tag{14}$$

where $P_>$ is the tail distribution or CCDF.

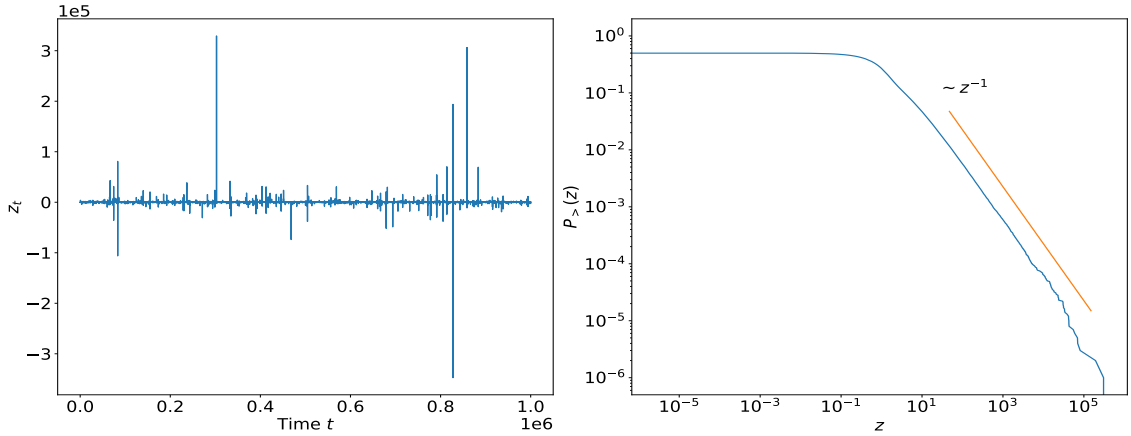

Figure 2: Left: solution of(2) with controller (10) and $n = 1$ look-back. Right: log-log plot of CCDF $P_>(z)$ with reference line z^{-1}. $N = 10^6$ time-steps, $\alpha_0 = 2$, $\sigma_0^2 = 0.8$, $z_0 = 1$.

2.2 CONVERGENCE OF THE TAIL

This behavior comes from the estimation of α_0 (10), which involves the sample variance in the denominator. In the case $n = 1$ the dynamics reads

$$z_{t+1} = \left(\alpha_0 - c_{t-1} - \frac{z_t}{z_{t-1}}\right) z_t + \beta_t. \tag{15}$$

When the dynamics is well controlled the estimator c_{t-1} is close to the ground truth α_0 and z_{t-1} is close to 0. If besides $(\alpha_0 - c_{t-1})z_{t-1} \sim \mathcal{O}(1)$, z_t is also of order 1 and thus $\frac{z_t}{z_{t-1}}$ diverges. Using conservation of probability:

$$P(z_{t+1}) = P(z_{t-1})\left|\frac{\mathrm{d}z_{t-1}}{\mathrm{d}z_{t+1}}\right|. \tag{16}$$

Replacing $z_{t-1} = \frac{-z_t^2}{(z_{t+1}-\beta_t)-(\alpha_0-c_{t-1})z_t} \underset{z_{t+1}\gg 1}{\longrightarrow} 0$ and $\left|\frac{\mathrm{d}z_{t-1}}{\mathrm{d}z_{t+1}}\right| = \frac{z_t^2}{[(z_{t+1}-\beta_t)-(\alpha_0-c_{t-1})z_t]^2}$, the tail of the distribution is

$$P(z_{t+1}) \propto \frac{1}{z_{t+1}^2} \tag{17}$$

as $P(z_{t-1})$ becomes a constant when z_{t-1} goes to[1] 0. Control still works as z_t stays bounded but is not optimal because of the heavy tailed time series, which generates large deviations and a divergent variance. With larger n the estimator (10) gives better results, and the tail distribution $P_>(z_t)$ gets thinner.

This estimator is a sum of random variables divided by the sample variance of z, which resembles the origin of a Student's $t-$distribution with n degree of freedom [17]:

$$f_n(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \tag{18}$$

which has asymptotic tail $|x|^{-(n+1)}$; fitting with the exponent $\mu^{(n)} = n$ obtained numerically Figure (3) and suggested in [1].

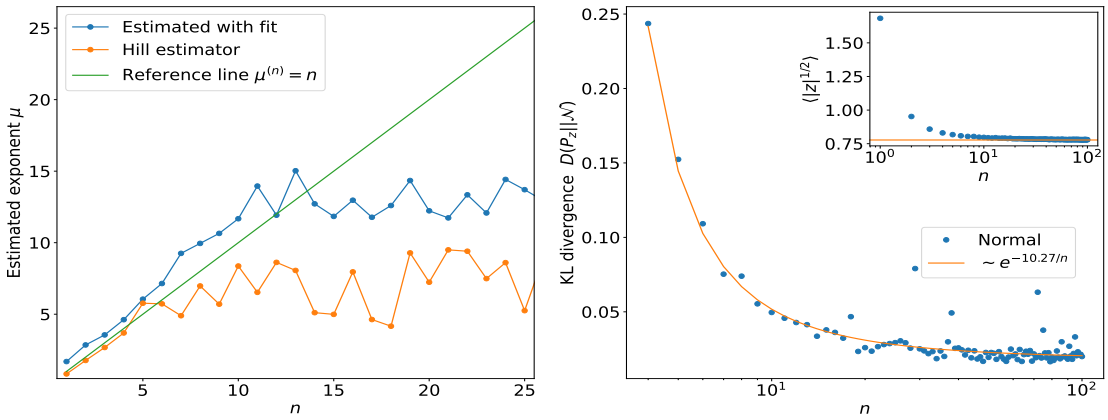

Figure 3: Left: Tail exponent for different n, obtained by fitting the CCDF (blue) or with Hill estimator (orange, [18]). Green line is the postulated exponent $\mu^{(n)} = n$ [1, 2]. Right: Convergence toward a Gaussian. Kullback-Leibler divergence between simulated distribution $P(z_t)$ and a centered normal distribution $\mathcal{N}(0, \sigma_0^2)$ (blue). In orange reference line $e^{-C/n}$ with $C \approx -10.27$, indicating a threshold at $n^* \approx 10$. After n^* empirical distribution is close to the normal one, and adding more look-backs does not change the behavior. Inset: absolute half moment $\left\langle |z|^{1/2} \right\rangle$ compared with theoretical value $\left\langle |X|^{1/2} \right\rangle \approx 0.778$ (orange line) for $\mathcal{N}(0, \sigma_0^2)$. $N = 10^5$ time-steps, $\alpha_0 = 2$, $\sigma_0^2 = 0.8$, $z_0 = 1$.

[1] Power laws are constant below some threshold $z_m > 0$.

This asymptotic power law behavior is stable upon convolution with additive scaling factor [17, chap. 4]. The tail of a sum of n random variables following (14) with $\mu = n$ is thus distributed as

$$P_n(z) \sim \frac{n}{z^{1+\mu}}, \quad z \to \infty \tag{19}$$

However, if $\mu > 2$ variance $\langle z^2 \rangle$ is mathematically defined so the sequence converges to a Gaussian distribution with variance $n \langle z^2 \rangle$. This convergence occurs for the centre of the pdf (small values) but is not guaranteed beyond a certain threshold, where it stays a power law [17, chap. 3]. The threshold between those two regimes is obtained by matching distributions:

$$P_n(z) \sim \frac{1}{\sqrt{2\pi n}} e^{-\frac{z^2}{2n}}, \quad z < z_0 \tag{20a}$$

$$P_n(z) \sim \frac{n}{z^{1+\mu}}, \quad z > z_0. \tag{20b}$$

Equating them, one gets $z_0(n) \sim \sqrt{n \log n}$. If $z_{max} = \max_t\{z_t\} < z_0(n)$ then the time series exhibits no deviations outside of Gaussian regime. With power law distributed time series, the typical maximum after N steps is $m_\mu \sim N^{1/\mu}$, while it is $m_\mathcal{N} \sim \sigma\sqrt{2\log(N)}$ for a normally distributed sequence (see section A.2). Numerical simulations indeed reveal two regimes for z_{max} (Figure 4): for small n it follows exactly the typical maximum of a power law distribution $m_\mu(n)$ with exponent $\mu^{(n)} = n$. But once the maximum deviation gets smaller than $z_0(n)$, z_{max} becomes independent of n, equal to $m_\mathcal{N} \approx 4.27$ (for $N = 10^5$). Figure 14 gives CCDF for some values on n; highlighting the convergence towards a normal distribution. Figure 3 shows the Kullback-Leibler (KL) divergence between the empirical distribution and noise distribution $\mathcal{N}(0, \sigma_0^2)$ versus n. It decreases fast for small n and stays constant close to 0 for $n \gtrsim 10$, implying that z_t reaches its final distribution. Figure 3 also shows the absolute half moment[2] $\left\langle |z|^{1/2} \right\rangle$ versus n, which tends to the theoretical value[3] at $n \gtrsim 10$. For $N = 10^5$, the maximum of the normal distribution $m_\mathcal{N}$ equals $z_0(n)$ for $n \approx 8.57$, justifying the change of regime observed Figure 4.

2.3 ANALYSIS AS A KESTEN PROCESS

When n is sufficiently large, $\hat{\alpha}_0^{(n)}$ is normally distributed around the ground truth α_0 with variance s_n^2. It is then possible to rewrite

$$c_t = \alpha_0 - s_n \xi_t \sim \mathcal{N}(\alpha_0, s_n^2), \tag{21}$$

where $\xi_t \sim \mathcal{N}(0,1)$. Replacing in the dynamical equation (2) one gets

$$z_{t+1} = a_t z_t + \beta_t, \quad \text{with } a_t = s_n \xi_t \sim \mathcal{N}(0, s_n^2). \tag{22}$$

This stochastic equation, known as a Kesten process [10], is a mixture of multiplicative and additive process, with a_t and β_t i.i.d. normally distributed random

[2] Moments with $k \geq 1$ are not defined for power laws with $\mu = 1$ hence we use $k = 1/2$ moment.

[3] $\mathbb{E}[|X|^{1/2}] = \int \mathrm{d}x \frac{|x|^{1/2}}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{x^2}{2\sigma_0^2}} = \sqrt{\frac{\sigma_0\sqrt{2}}{\pi}} \Gamma\left(\frac{3}{4}\right) \approx 0.778$ for $\sigma_0^2 = 0.8$.

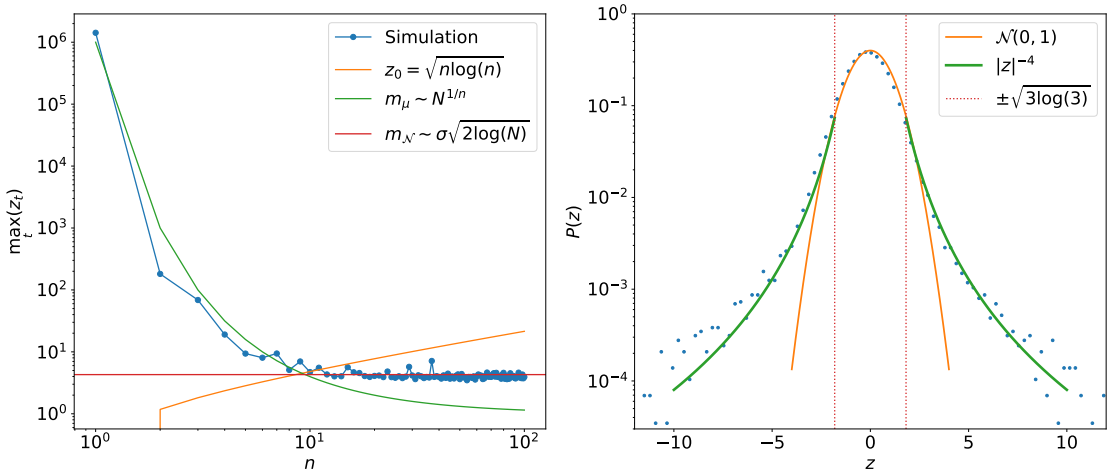

Figure 4: Left: Convergence of the tail. Maximum value of the time series z_{max} (blue), with trendline for the typical maximum of a sequence of power laws $m_n \sim N^{1/n}$ (green) and normally distributed variables $m_\mathcal{N} \sim \sigma_0\sqrt{2\log(N)}$ (red, section A.2). Orange line gives the typical threshold beyond which convergence toward a Gaussian is invalid: once $z_{max} < z_0(n)$ the whole distribution becomes normally distributed, without a remaining power tail. Equating $m_\mu(n) = m_\mathcal{N}(n)$ with $N = 10^5$ gives the threshold at $n \approx 9$, which fits with the observations on both graphs. Right: Illustration of convergence in the centre. Histogram of z_t for $n = 3$, with logarithmic y axis. In the region in between $\pm\sqrt{n\log(n)}$ (red borders), z_t follows closely a normal distribution (orange), while the tail still follows a power law $P(z) \sim |z|^{-4}$ after (green). $N = 10^5$ time-steps, $\alpha_0 = 2$, $\sigma_0^2 = 0.8$, $z_0 = 1$.

variables. This process has been used in different disciplines, such as in the modeling of 1D disordered systems [12], the study of synaptic size dynamics [19, 13] and economics [11].

The average of (22) squared,

$$\left\langle z_{t+1}^2 \right\rangle = \left\langle a_t^2 \right\rangle \left\langle z_t^2 \right\rangle + \left\langle \beta_t^2 \right\rangle, \tag{23}$$

admits a stationary solution for $\left\langle a_t^2 \right\rangle < 1$

$$\left\langle z_t^2 \right\rangle = \frac{\sigma_0^2}{1 - \left\langle a_t^2 \right\rangle}. \tag{24}$$

In absence of noise ($\beta_t \equiv 0$) the process converges to 0 if $\langle \log|a_t| \rangle < 0$ (similar to the condition $\left\langle a_t^2 \right\rangle < 1$) [20]. Given $a_t = s_n\xi_t$ and $s_n \propto 1/n$, this condition is easily met[4]. Besides, a_t can also take values larger than 1 with probability $P(|a_t| > 1) = P(|\xi_t| > 1/s_n) \approx 0.54$ for $s_n = 1$, giving birth to intermittent amplifications. Turning noise on pushes back the solution from the origin, acting as a reinjection of the dynamics. Those two elements generate intermittency in the solution, with a power law distribution for z_t [21]. Using renewal theory, Kesten and Goldie proved [10, 22] the following results:

1. If a_t and ξ_t are i.i.d. are real-valued random variables such that $\langle \log|a_t| \rangle < 0$, then z_t converges in distribution to a unique limiting distribution.

[4] From Jensen's inequality applied to $e^X = e^{\log|a_t|}$: $e^{\langle \log|a_t| \rangle} \leq \left\langle e^{\log|a_t|} \right\rangle = \left\langle e^{\log(s_n) + \frac{1}{2}\log\left(\frac{2}{\pi}\right)} \right\rangle$. Taking the log on both sides gives $\langle |a_t| \rangle \leq \log(s_n) + \frac{1}{2}\log\frac{2}{\pi} < 0$ for $s_n \lessapprox 1.25$.

2. If, besides, $\beta_t/(1-a_t)$ is non degenerate (that is β_t is not proportional to a_t), and if there exists a positive constant μ such that
   (a) $0 < \langle |\beta_t|^\mu \rangle < +\infty$,
   (b) μ is defined by
   $$\langle |a_t|^\mu \rangle = 1, \tag{25}$$
   (c) and $\langle |a_t|^\mu \log^+ |a_t| \rangle < \infty$,

   then the tail of this distribution is a power law:
   $$P_>(z) \sim C z^{-\mu}. \tag{26}$$

Condition 2a simply states that β_t doesn't have a fatter tail than $x^{-(1+\mu)}$, which is the case for Gaussian noise. The third condition is also met as a_t is normally distributed. Equation (25), which holds for $\mu < 2$ [23], provides an effective way to compute the exponent μ. Replacing $a_t = s_n \xi_t$ gives

$$1 \overset{!}{=} \langle |a_t|^\mu \rangle = \langle s_n^\mu |\xi_t|^\mu \rangle . \tag{27}$$

This development is valid when $\hat{\alpha}_0^{(n)}$ is normally distributed, in which case $s_n = \sigma_0/(n\sigma_z)$. The condition becomes

$$1 \overset{!}{=} \langle |a_t|^\mu \rangle = s_n^\mu \langle |\xi_t|^\mu \rangle = \frac{1}{n^\mu} \left(\frac{\sigma_0}{\sigma_z} \right)^\mu \langle |\xi_t|^\mu \rangle . \tag{28}$$

The last factor is the absolute moment of a centered Gaussian random variable, which is given by

$$\langle |\xi_t|^\mu \rangle = \frac{2^{\mu/2} \Gamma\left(\frac{\mu+1}{2}\right)}{\sqrt{\pi}}, \tag{29}$$

and is valid for any real $\mu > -1$ [24]. Inverting equation (28), we obtain n such that the limiting distribution of z has exponent μ:

$$n^\mu = \left(\frac{\sigma_0}{\sigma_z} \right)^\mu \frac{2^{\mu/2} \Gamma\left(\frac{\mu+1}{2}\right)}{\sqrt{\pi}}, \tag{30}$$

i.e. $n \propto \left[\frac{1}{\sqrt{\pi}} \Gamma\left(\frac{\mu+1}{2}\right) \right]^{1/\mu}$. Using Stirling's approximation:

$$\Gamma\left(\frac{\mu+1}{2}\right) = \sqrt{\frac{\pi}{\mu+1}} \left(\frac{\mu+1}{2e} \right)^{(\mu+1)/2} + \mathcal{O}\left(\frac{1}{\mu}\right), \tag{31}$$

which gives

$$n \sim (\mu+1)^{-1/2\mu + (\mu+1)/2\mu} \sim (\mu+1)^{1/2}. \tag{32}$$

This result seems in contradiction with Figure 4, but one should remind that it is valid only for large n, as it relies on Gaussian approximation of the estimator. However, numerical simulations have shown a shrinking tail for large n. Indeed, this predicted power law distribution holds for the tail only. As n increases, the cutoff separating the tail gets larger, until it gets larger than the maximum deviation observed. At this point $P(z_t)$ is normally distributed. The final result (32) of this derivation is consequently not usable in practice, but still provides a theoretical basis demonstrating the intrinsic scaling behavior of (2). Besides, even though c_t is not necessarily normally distributed for smaller n, equation (25) holds as long the distribution of $\alpha_0 - c_t$ respects the conditions stated above. This approach thus justifies the power law distribution qualitatively, and allows to compute the exponent if the distribution of c_t is known.

3 MODIFIED CONTROL

For controls that takes into account few past observations only, simulations are heavy tailed. Taking the extreme case $n = 1$ we propose to add a regularization, to hinder extreme deviations. When $|z_t/z_{t-1}|$ exceeds a threshold $\varepsilon > 0$ it is replaced by a positive constant η. This constant step must be taken in the right direction (i.e. such that c_t gets closer to α_0), which is given by the sign of z_t/z_{t-1} (see Figure 5). The new map reads:

$$z_{t+1} = (\alpha_0 - c_t)z_t + \beta_t \tag{33a}$$

$$c_t = \begin{cases} c_{t-1} + \frac{z_t}{z_{t-1}}, & \left|\frac{z_t}{z_{t-1}}\right| < \varepsilon \\ c_{t-1} + \eta \operatorname{sign}\left(\frac{z_t}{z_{t-1}}\right), & \left|\frac{z_t}{z_{t-1}}\right| > \varepsilon. \end{cases} \tag{33b}$$

(33b) is similar to a line search strategy, where $\operatorname{sign}(z_t/z_{t-1})$ is the descent direction and η a constant learning rate.

Final distribution depends heavily on the choice of those two parameters. From simple heuristic arguments one can reduce the range of parameters to consider: First, when ε is too large, regularization is rarely engaged hence results are unchanged. We therefore restrict the analysis to $\varepsilon < 1$, meaning that original controller is replaced only when the time series is growing ($|z_t/z_{t-1}| > 1$). Secondly, η gives the maximal precision the estimator can achieve on α_0, and should thus be small enough. As for ε, we focus on the values $\eta < 1$.

3.1 NUMERICAL RESULTS

Figure 6 shows the maximum value and 99%−quantile of time series simulated with (33). As before, computing sample mean or variance does not make sense as they are not defined for a power law with exponent $\mu = 1$. Results are similar with both metrics: they depend on η only for $\varepsilon \lesssim 0.8$, while they follow a pattern along the line $\varepsilon \sim \eta$ when the threshold gets larger. This first observation allows to fix ε to a value below 0.8, reducing the number of free parameters.

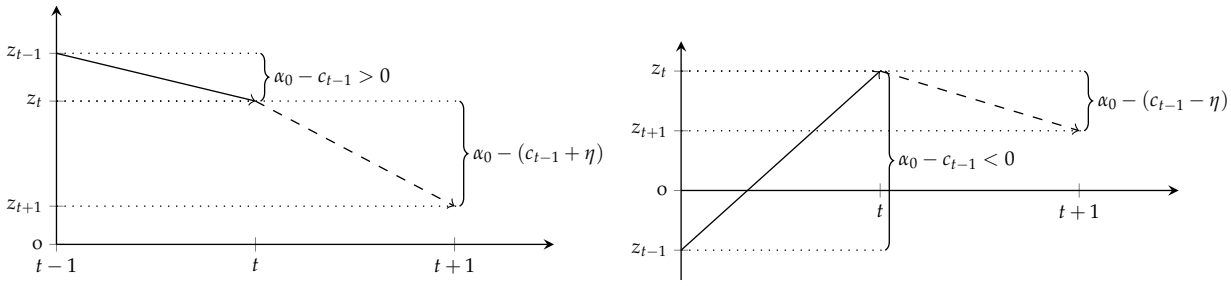

Figure 5: Schematic representation of the modification proposed. Left: $\operatorname{sign}\left(\frac{z_t}{z_{t-1}}\right) = 1$, Right: $\operatorname{sign}\left(\frac{z_t}{z_{t-1}}\right) = -1$. At each time step a constant is either added or subtracted to the previous estimate, in order to minimize the error $|\alpha_0 - c_t|$.

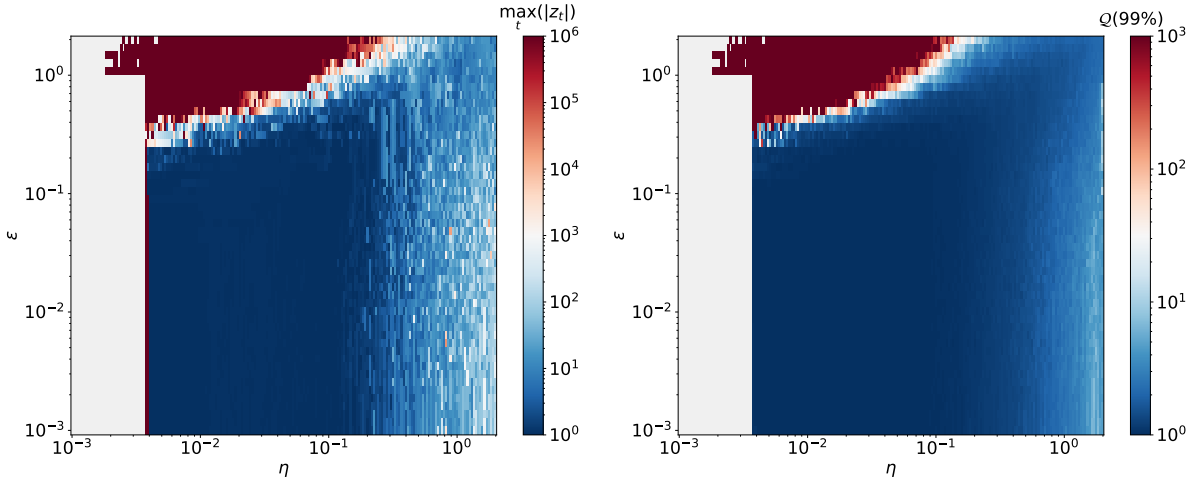

Figure 6: Summary statistics of time series simulated with (33): maximum value (Left) and 99%−quantile (Right) for logarithmically spaced η and ε. Difference between them enlightens the magnitude of extreme (top 1%) deviations. Both quantities are normalized by their theoretical value for a normal distribution $\mathcal{N}(0, \sigma_0^2)$, $m_{\mathcal{N}} = 3.84$ and $\mathcal{Q}_{\mathcal{N}}(99\%) = 2.08$ (see section A.2). $N = 10^4$, $\alpha_0 = 2$, $\sigma_0^2 = 0.8$, $z_0 = 1$.

Both the maximum and last percentile exhibit an abrupt transition at $\eta^* = 4 \cdot 10^{-3}$. They diverge below, and reach the theoretical value (see section A.2) for a normal distribution[1] $\mathcal{N}(0, \sigma_0^2)$ above. With smaller learning rate the estimator (33b) is updated by smaller steps and, therefore, takes more time to converge towards α_0. With the initial condition $z_0 = 1$ and $\alpha_0 = 2$ the time series starts increasing, engaging (33b) from the beginning. As it stays above the origin, the dynamics reduces to $z_{t+1} = (\alpha_0 - t \cdot \eta) z_t + \beta_t$. It becomes on average convergent ($\langle z_t / z_{t-1} \rangle < 1$) after $\sim (\alpha_0 - 1)/\eta$ time-steps (see Figure 7).

If η is to small the excitation $(\alpha_0 - c_t)$ therefore stays larger than one during many time-steps, putting the time series on an explosive track. Above η^*, the controller gets closer to α_0 sufficiently early and prevents the time series from diverging, even though it may take a long time reach a stationary regime. For Figure 6, the first 1000 time-steps (10% of total time) are omitted to focus on the stationary regime only. This threshold obviously depends on α_0, in a piecewise linear way (Figure 8). Noise dependence exhibits a transition at $\sigma_0^2 = 1$. Above, noise is strong enough to compensate a large excitation factor $(\alpha_0 - c_t)$, allowing z_t to leave the initially divergent trajectory.

Figure 9 shows again the maximum and 99%−quantile, for $\varepsilon = 0.1$. A second transition becomes visible at $\eta \approx 1.5 \cdot 10^{-1}$. Maximum value and last quantile both start increasing, but much slower for the latter. The increasing spread between them indicates that extreme deviations increase more rapidly than smaller ones.

From those observations we have seen that several regime arise given ε and η. In the "central" region $\varepsilon < 1.5 \cdot 10^{-1}$ and $4 \cdot 10^{-3} < \eta \lesssim 1.5 \cdot 10^{-1}$ both the maximum and last percentile of the time series are constant, close to their theoretical value for $\mathcal{N}(0, \sigma_0^2)$. For larger learning rate the tail of the distribution gets heavier; and the solution becomes totally divergent below η^*. Finally for $\varepsilon \gtrsim 0.8$ distribution becomes closer to the power law obtained with original control (2). To confirm this observation Figure 10 shows the estimated tail

[1] $m_{\mathcal{N}} = 3.84$ for $N = 10^4$ and $\mathcal{Q}_{\mathcal{N}}(99\%) = 2.08$

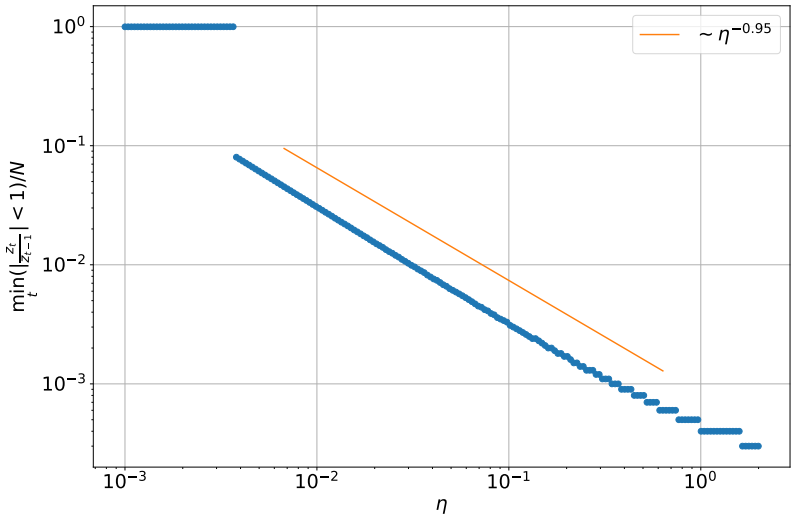

Figure 7: Relative duration of the transient regime before the solution of (33) starts decreasing ($\min_t\{|z_t/z_{t-1}| < 1\}/N$), given η. In orange, reference line $\eta^{-0.95}$ fitting the predicted duration $(\alpha_0 - 1)/\eta$. The initial flat regime corresponds to diverging part: the dynamics is stuck on an explosive trajectory. $N = 10^4$, $\alpha_0 = 2$, $\sigma_0^2 = 0.8$, $z_0 = 1$ and $\varepsilon = 0.1$.

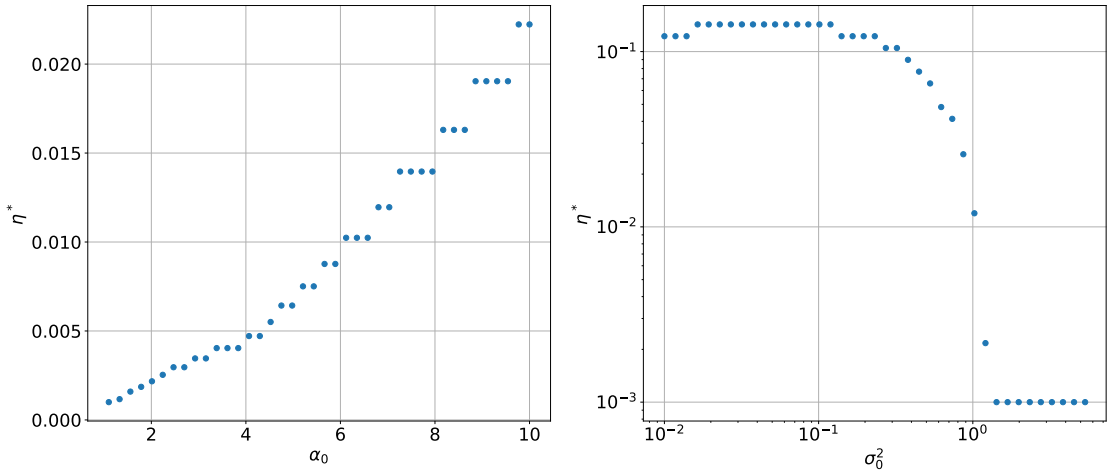

Figure 8: Threshold η^* below which the dynamics always diverge, versus excitation parameter α_0 (Left) and noise intensity σ_0^2 (Right). The linear relationship $\eta^* \sim \alpha_0$ comes from the estimator (33b), but the transition around $\sigma_0^2 = 1$ is less trivial. After this threshold, noise is large enough to compensate the error due to a smaller learning rate. $N = 10^4$, $\alpha_0 = 2$, $\sigma_0^2 = 0.8$, $z_0 = 1$ and $\varepsilon = 0.1$.

exponent. The same behavior appear : in the "central" region the exponent is almost independent and close to 7, indicating a rather thin tail[2]. In the "outer" region ($\varepsilon \gtrsim 0.8$ or $\eta \gtrsim 0.15$) it drops to 1, as with the initial model.

All those observations suggest to focus on the "central" region, where statistics is close to Gaussian distribution. For a deeper analysis of the final distribution we select $\varepsilon = 0.1$ and $\eta = 0.05$. Figure 11 compares the solution of the original

[2] Exponent is estimated with Hill estimator, but the distribution in this region is closer to a Gaussian i.e. without a power tail.

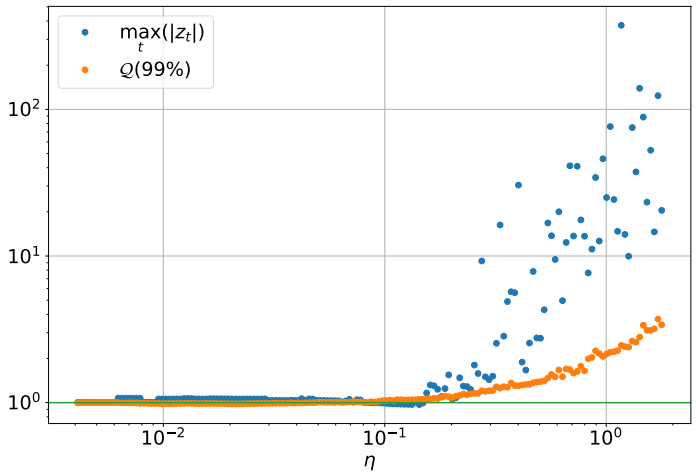

Figure 9: Sectional view of Figure 6 for $\varepsilon = 0.1$ and $\eta > \eta^* = 4 \cdot 10^{-3}$, showing the transition at $\eta = 1.5 \cdot 10^{-1}$. The increasing spread between last percentile and maximum value indicates the appearance of a heavy tail. Both quantities are normalized by their theoretical value with a normal distribution for comparison ($m_{\mathcal{N}} = 3.84$ and $\mathcal{Q}_{\mathcal{N}}(99\%) = 2.08$, see section A.2). $N = 10^4$, $\alpha_0 = 2$, $\sigma_0^2 = 0.8$, $z_0 = 1$.

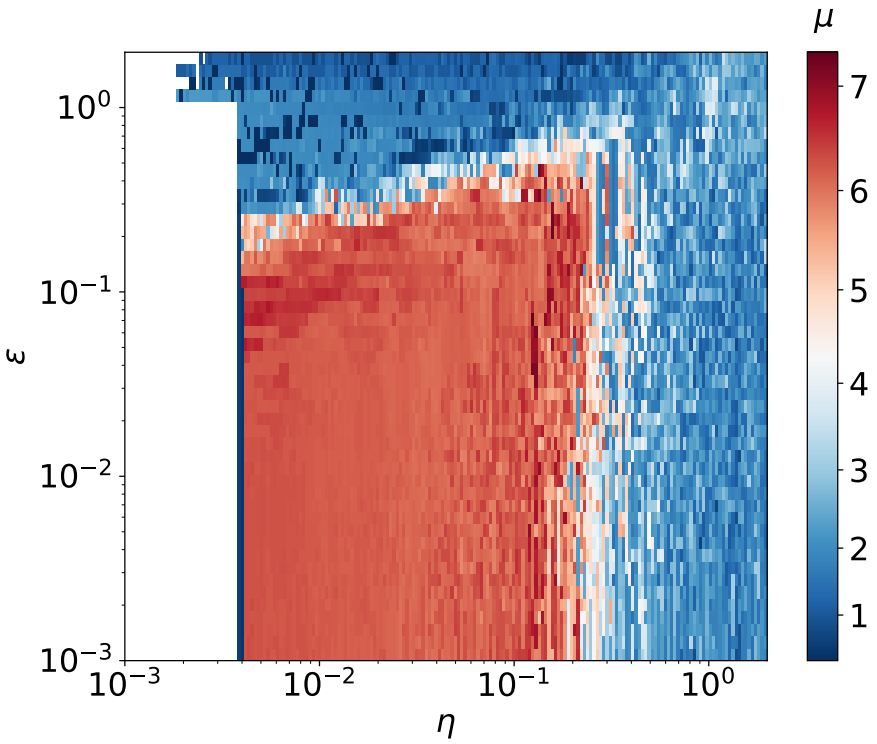

Figure 10: Estimated exponent of the final distribution from (33), using Hill estimator [18]. In the upper and right regions $P(z_t)$ is close to the power law obtained with original control (17), with exponent $\mu \approx 1$. In the central region the distribution gets closer to a Gaussian therefore the exponent is much larger. $N = 10^4$, $\alpha_0 = 2$, $\sigma_0^2 = 0.8$, $z_0 = 1$.

MLE control with the proposed control. Figure 12 shows $P_>(z_t)$ in both case. The distribution with modified control is indeed Gaussian. Kolmogorov-Smirnov test

gives a distance $1.17 \cdot 10^{-3}$ and a $p-$value of 0.49, meaning that the solution is very likely normally distributed.

The modification proposed achieves optimal control, yielding a Gaussian distribution without large deviations ($\max z_t = 4.8$).

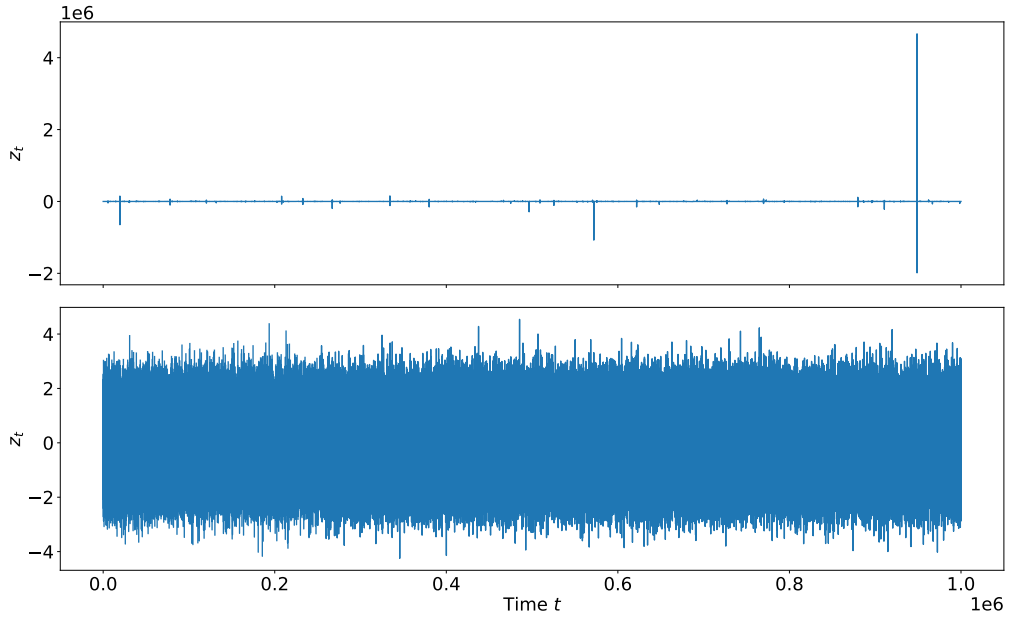

Figure 11: Numerical simulations of (2) with the original MLE controller (10) (Upper) and proposed modification (33) (Lower), for $\varepsilon = 0.1$ and $\eta = 0.05$. Added regularization achieves to prevent large deviations. All parameters are identical: $\alpha_0 = 2$, $\sigma_0^2 = 0.8$, $z_0 = 1$, $N = 10^6$.

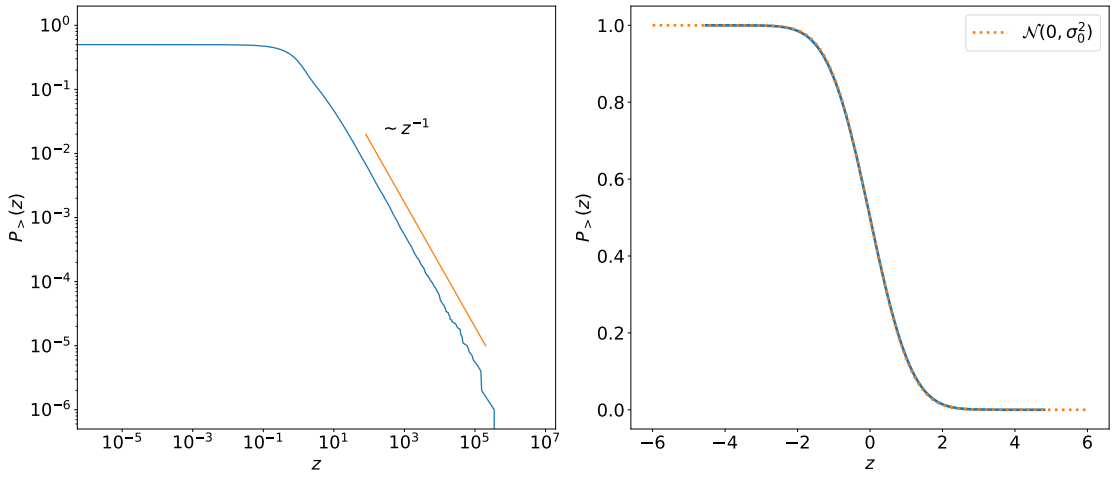

Figure 12: CCDF of simulations from Figure 11. Original MLE controller (Left) and proposed modification (Right), with $\varepsilon = 0.1$ and $\eta = 0.05$. The final distribution with regularization is very close to a $\mathcal{N}(0, \sigma_0^2)$ (orange), indicating optimal control. All parameters are identical: $\alpha_0 = 2$, $\sigma_0^2 = 0.8$, $z_0 = 1$, $N = 10^6$. Left plot is in log-log to make the power law visible.

3.2 SIMPLIFICATION OF THE MODEL

All those observations confirm that the modification proposed is indeed able to make the controller optimal, in the sense that $z_t \sim P_\beta = \mathcal{N}(0, \sigma_0^2)$. However this result depends heavily on the choice of ε and η. Even though we obtained good result with $(\varepsilon, \eta) = (0.1, 0.5)$, this choice is arbitrary. By reducing to one the number of free parameters, it is easier to numerically optimize the value.

A first simplification could be to set $\eta = \varepsilon$, as results are typically better along this line for large ε (data not shown). The map then reduces to

$$z_{t+1} = \left[\alpha_0 - c_{t-1} - \operatorname{sign}\left(\frac{z_t}{z_{t-1}}\right) \min\left(\eta, \left|\frac{z_t}{z_{t-1}}\right|\right)\right] z_t + \beta_t. \tag{34}$$

Similarly, simulations have shown that below some value the distribution is independent of ε. Another simplification could thus be to forget totally the original controller, by setting $\varepsilon = 0$

$$z_{t+1} = \left[\alpha_0 - c_{t-1} - \eta \operatorname{sign}\left(\frac{z_t}{z_{t-1}}\right)\right] z_t + \beta_t. \tag{35}$$

This control is still adaptive as it keeps track of the descent direction sign (z_t/z_{t-1}), but always moves by constant steps η. In both case, having a unique parameter makes the model simpler. It is numerically possible to evaluate η^*, that minimize Kolmogorov-Smirnov distance D_{KS} with $\mathcal{N}(0, \sigma_0^2)$ (Figure 13). The first model ($\eta = \varepsilon$) gives an acceptable $p-$value for a wider range of η, but both achieve optimal control.

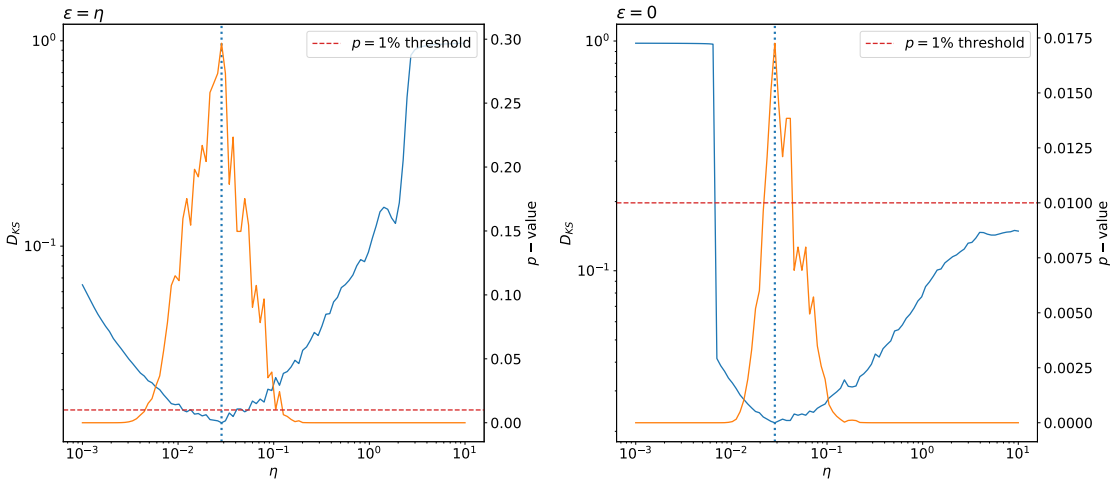

Figure 13: KS distance (left axis, blue) and $p-$value (right axis, orange) for the two simplifications proposed (34, 35). Red dashed line indicates the threshold $p = 1\%$, above which the hypothesis that $P(z_t)$ is normally distributed is acceptable. Distance between the two distributions are similar in both case, but the first proposition $\varepsilon = \eta$ is normally distributed for a broader range of η, with better $p-$value. The optimal learning rates are respectively $\eta_{opt} = 1.35 \cdot 10^{-2}$ and $\eta_{opt} = 1.96 \cdot 10^{-2}$. $N = 10^4$, $\alpha_0 = 2$, $\sigma_0^2 = 0.8$, $z_0 = 1$.

Distribution $P(z_t)$ is not a power law anymore, thus we can also study its variance. In the case $\varepsilon = 0$ it reads

$$\langle z_{t+1}^2 \rangle = \left\langle \left[\left(\alpha_0 - c_{t-1} - \eta \operatorname{sign}\left(\frac{z_t}{z_{t-1}}\right)\right) z_t + \beta_t\right]^2 \right\rangle. \tag{36}$$

Thanks to the properties of noise we expand it as

$$\left\langle z_{t+1}^2 \right\rangle = \left\langle \left(\alpha_0 - c_{t-1} - \eta \operatorname{sign}\left(\frac{z_t}{z_{t-1}} \right) \right)^2 z_t^2 \right\rangle + \sigma_0^2 \tag{37}$$

$$= \left\langle \left(\alpha_0 - c_{t-1} \right)^2 z_t^2 \right\rangle + \eta^2 \left\langle z_t^2 \right\rangle - 2\eta \left\langle z_t^2 \operatorname{sign}\left(\frac{z_t}{z_{t-1}} \right) \left(\alpha_0 - c_{t-1} \right) \right\rangle + \sigma_0^2. \tag{38}$$

The minimal value is reached for

$$0 = \left. \frac{\partial \left\langle z_{t+1}^2 \right\rangle}{\partial \eta} \right|_{\eta_{opt}} = 2\eta_{opt} \left\langle z_t^2 \right\rangle - 2 \left\langle z_t^2 \operatorname{sign}\left(\frac{z_t}{z_{t-1}} \right) \left(\alpha_0 - c_{t-1} \right) \right\rangle, \tag{39}$$

i.e.

$$\eta_{opt} = \sigma_0^2 \frac{\left\langle z_t^2 \operatorname{sign}\left(\frac{z_t}{z_{t-1}} \right) \left(\alpha_0 - c_{t-1} \right) \right\rangle}{\left\langle z_t^2 \right\rangle}. \tag{40}$$

This value cannot be computed ex ante as it depends on hidden parameters. If available, one could use a "train" set to optimize the parameters first.

CONCLUSION

With short memory, control algorithm proposed in [1] achieves to keep the time series bounded, but yields power law distributed fluctuations, questioning the term "optimal control". Indeed, optimality supposedly comes from the Maximum Likelihood Estimation of α_0. If the time series on which parameter is estimated is not long enough, MLE properties breaks down. When fluctuations get close to the origin information about the dynamics gets lost into noise (recall the variance of the estimator $\sigma_0^2/n^2\hat{\sigma}_z^2$). In the extreme case $n = 1$ the estimator is per se divergent, because of the inverse factor z_{t-1}^{-1}. Recasting the process in Kesten form demonstrates that this model intrinsically produces power laws. With more memory however tail events have more probability to be captured by the estimator and averaged out, thanks to the convergence of tail distribution towards a normal distribution, which requires more samples than the centre. With this first part we therefore describe in details the properties of the initial model, and explain the convergence towards optimal control as memory increases. This results contradicts the claim that criticality originates from optimal control itself. Extreme fluctuations simply come from a naive use of MLE. As the model fits well with experimental data [2], understanding deeply its characteristics gives interesting insights on how human performs visuomotor closed-loop control. Besides, the same underlying stochastic process was also found in describing synaptic size dynamics [25].

Considering the very short memory case ($n = 1$), we propose a modification that prevents divergence of the estimator: When the ratio of two consecutive deviations gets larger than some threshold, MLE estimator is replaced by a constant increment in the descent direction. This replacement is still adaptive as this direction is obtained from previous observations. Numerical simulations show that this naive modification leads to optimal control for well-chosen parameters. In particular, the threshold must be small enough to ensure that the original estimator is replaced sufficiently often. Otherwise, former power law behavior dominates the distribution and results are the same as before. Without knowledge of hidden parameters α_0 and σ_0^2, the choice of parameters is not straightforward. Hopefully, simulations show that using the same parameters for both the threshold and the increment yields good results, reducing the complexity of the model. Another choice with satisfying results is to set the threshold to zero, that is getting rid of the original model. One major inconvenient of this modification is that for certain initial conditions and parameters the dynamics can be stuck in an unstable regime, before reaching stationarity. Knowing α_0 and σ_0^2 it is possible to numerically optimize the parameters, by minimizing KS statistics or sample variance. When possible, the best would thus be to use a training set to optimize η before using the controller.

Despite its optimality, our proposed modification is indeed not of practical use in the context of the initial model. More generally, strategies of adaptive control in engineering also often rely on parameter estimation [26], but are more complex. MLE is not commonly use, as it requires large datasets. Kalman filters

are for example widely used to denoise past observations, and could have been a great alternative to our naive modification.

APPENDIX

A.1 VARIANCE OF THE ESTIMATOR $\alpha_0^{(n)}$

MLE estimator is asymptotically normal [9], meaning

$$\sqrt{n}(\hat{\alpha}_0^{(n)} - \alpha_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}^{-1}) \tag{A.1.1}$$

where $\mathcal{I}$ is the Fisher information:

$$\mathcal{I} \equiv \left\langle -\frac{\partial^2 \log \mathcal{L}(\{z_{t-i-1}\}_{i=0}^n; \alpha_0, c_i^n)}{\partial \alpha_0^2} \right\rangle. \tag{A.1.2}$$

The log-likelihood is given by

$$\log \mathcal{L}(\{z_{t-i-1}\}_{i=0}^n; \alpha_0, c_i^n) = \sum_{i=0}^{n-1} \log\big(P(z_{t-i}|z_{t-i-1}, \dots z_{t-n}; \alpha_0, c_i^n)\big) + \log\big(P(z_{t-n}|\alpha_0; c_i^n)\big), \tag{A.1.3}$$

where $P(z_{t-i}|z_{t-i-1}, \dots z_{t-n}; \alpha_0, c_i^n)$ is Gaussian:

$$P(z_{t-i}|z_{t-i-1}, \dots z_{t-n}; \alpha_0, c_i^n) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(z_{t-i}-(\alpha_0 - c_{t-i-1})z_{t-i-1})^2}. \tag{A.1.4}$$

Differentiating (A.1.3) twice gives

$$\frac{\partial^2 \log \mathcal{L}(\{z_{t-i-1}\}_{i=0}^n; \alpha_0, c_i^n)}{\partial \alpha_0^2} = -\frac{1}{\sigma_0^2} \sum_{i=0}^{n-1} z_{t-i-1}^2. \tag{A.1.5}$$

Fischer information thus reads

$$\mathcal{I} = \frac{1}{\sigma_0^2} \left\langle \sum_{i=0}^{n-1} z_{t-i-1}^2 \right\rangle, \tag{A.1.6}$$

and the variance of $\hat{\alpha}_0^{(n)}$ is

$$s_n^2 = \frac{1}{n\mathcal{I}} = \frac{\sigma_0^2}{n \left\langle \sum_{i=0}^{n-1} z_{t-i-1}^2 \right\rangle}. \tag{A.1.7}$$

Defining the sample variance (provided that it exists) $\sigma_z^2 = \left\langle z_{t-i-1}^2 \right\rangle$, Fisher information reads

$$\mathcal{I} = \frac{n\sigma_z^2}{\sigma_0^2}, \tag{A.1.8}$$

and the variance of the estimator is

$$s_n^2 = \frac{\sigma_0^2}{n^2 \sigma_z^2}. \tag{A.1.9}$$

For $n \leq 2$, variance is not defined so (A.1.7) should be used instead.

A.2 MAXIMUM AND QUANTILES OF DISTRIBUTION

Let's consider a sequence $\{X_i\}_i$ of N iid random variables and $X_{max} = \max_i\{X_i\}$ its maximum value. We are interested by the typical value of X_{max} and last percentile $\mathcal{Q}(99\%)$ of power laws and normal distributions.

A.2.1 *Power law*

If X_i follows a power law with exponent μ, its distribution and tail function are:

$$P(x) = \frac{C}{x^{1+\mu}}, \quad P_>(x) = P(X > x) = \frac{C}{x^\mu}. \tag{A.2.1}$$

The probability that a given value m is larger than $X_{max} = \max_i\{X_i\}$ is

$$\Pi_<(m) = P(\max_i\{X_i\} < m) = P(X_1 < m, X_2 < m, \dots X_N < m). \tag{A.2.2}$$

Thanks to independence, $\Pi_<(m) = [P_<(m)]^N = [1 - P_>(m)]^N$, which we can rewrite

$$\Pi_<(m) = e^{N\ln(1-P_>(m))} \simeq e^{-N\ln(P_>(m))}, \tag{A.2.3}$$

where we expanded the logarithm for small values of $P_>(m)$, as they are the only one relevant in the exponential for large N. Using (A.2.1) we get

$$\Pi_<(m) = e^{Nm^{-\mu}} \tag{A.2.4}$$

and the maximum value m attained with probability p is $m \sim (N\ln(p))^{1/\mu}$, linear in N for $\mu = 1$. With $N = 10^6$ as in Figure 2, a typical maximum is 10^6 with probability $p = 1/e \approx 37\%$. For power laws like (A.2.1), quantiles function is $\mathcal{Q}(p) = x_m(1-p)^{-1/\mu}$, where x_m marks the beginning of the power tail. Unlike moments, quantiles are defined for any $\mu > 0$ and is thus a more convenient statistical tool.

A.2.2 *Normal distribution*

Now if the sequence follows $\mathcal{N}(0, \sigma_0^2)$, we take a different approach, by computing an upper bound for the expected maximum value $\langle X_{max}\rangle$. By Jensen's inequality:

$$e^{t\langle X_{max}\rangle} \leq \left\langle e^{tX_{max}} \right\rangle = \left\langle \max_i e^{tX_i} \right\rangle. \tag{A.2.5}$$

Then we use $\max_i e^{tX_i} \leq \sum_i e^{tX_i}$ to get

$$e^{t\langle X_{max}\rangle} \leq \left\langle \sum_i e^{tX_i} \right\rangle. \tag{A.2.6}$$

The RHS is a sum of moment generating functions:

$$e^{t\langle X_{max}\rangle} \leq \left\langle \sum_i e^{tX_i} \right\rangle = \sum_i e^{\frac{t^2\sigma_0^2}{2}} = Ne^{\frac{t^2\sigma_0^2}{2}}. \tag{A.2.7}$$

Taking the logarithm on both sides we obtain

$$\langle X_{max}\rangle \leq \frac{\log(N)}{t} + \frac{t\sigma_0^2}{2}. \tag{A.2.8}$$

Choosing $t^* = \frac{\sqrt{2\log N}}{\sigma}$ to maximize the expression[3], the expected maximum is bounded by $m_{\mathcal{N}} = \sigma_0\sqrt{2\ln N}$. Quantile function of the normal distribution is $\mathcal{Q}(p) = \mu + \sigma_0\sqrt{2}\text{erf}^{-1}(2p-1)$. For $\mathcal{N}(0,\sigma_0^2)$, the last percentile is $\mathcal{Q}(99\%) \approx 2.08$.

[3] One can check that $\partial_t^2|_{t^*} \frac{\log(N)}{t} + \frac{t\sigma_0^2}{2} > 0$.

A.3 TAIL DISTRIBUTION VERSUS n

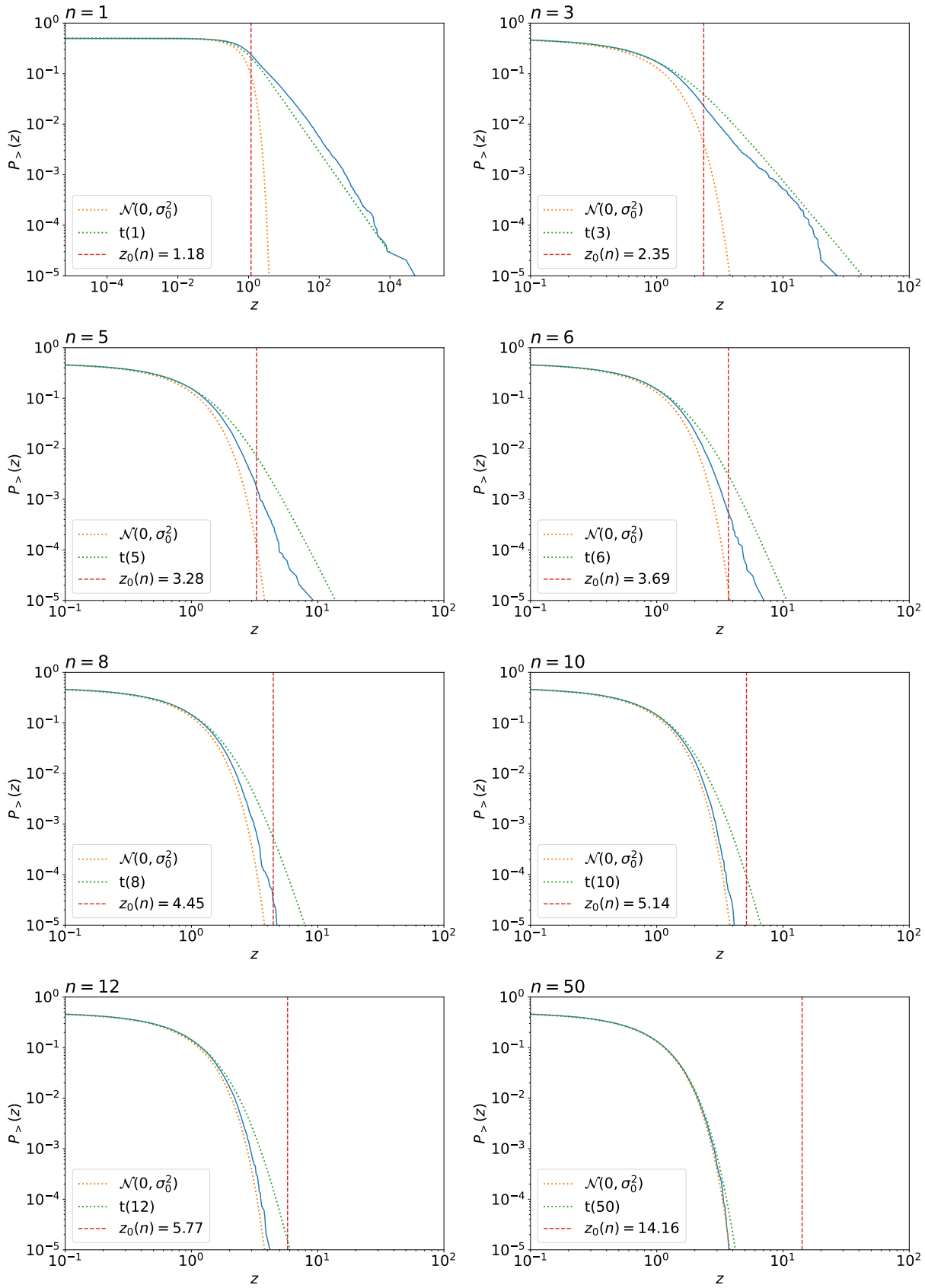

Figure 14: Tail distribution of time series simulated with (2) and (10). In orange and green, CCDF of $\mathcal{N}(0, \sigma_0^2)$ and $t(n)$ for reference. Red dashed line indicates the region of convergence toward a Gaussian $z_0(n)$. After this threshold the power tail remains (20).

BIBLIOGRAPHY

[1] Christian W. Eurich and Klaus Pawelzik. Optimal control yields power law behavior. pages 365–370, 2005.

[2] Felix Patzelt, Markus Riegel, Udo Ernst, and Klaus Pawelzik. Self-organized critical noise amplification in human closed loop control. *Frontiers in Computational Neuroscience*, 1:4, 2007. ISSN 1662-5188. doi: 10.3389/neuro.10.004.2007.

[3] John M. Beggs and Dietmar Plenz. Neuronal avalanches in neocortical circuits. *Journal of Neuroscience*, 23(35):11167–11177, 2003. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.23-35-11167.2003.

[4] Vilfredo Pareto. *Écrits sur la courbe de la répartition de la richesse. Œuvres complètes : T. III*. Librairie Droz, Genève, 1967. ISBN 9782600040211.

[5] J. B. Estoup. Les gammes stenographiques. *Institut Stenographique de France*, 1916.

[6] George Kingsley Zipf. *Human behavior and the principle of least effort.* Addison-Wesley Press, Oxford, England, 1949.

[7] R. Bormann, J.-L. Cabrera, J.G. Milton, and C.W. Eurich. Visuomotor tracking on a computer screen—an experimental paradigm to study the dynamics of motor control. *Neurocomputing*, 58-60:517–523, 2004. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2004.01.089. Computational Neuroscience: Trends in Research 2004.

[8] Juan L. Cabrera and John G. Milton. On-off intermittency in a human balancing task. *Phys. Rev. Lett.*, 89:158702, Sep 2002. doi: 10.1103/PhysRevLett.89.158702.

[9] David R. Cox and David V. Hinkley. *Theoretical Statistics*. Chapman & Hall, London, England, 1974.

[10] Harry Kesten. Random difference equations and Renewal theory for products of random matrices. *Acta Mathematica*, 131(none):207 – 248, 1973. doi: 10.1007/BF02392040.

[11] Xavier Gabaix. Power laws in economics and finance. Working Paper 14299, National Bureau of Economic Research, September 2008.

[12] Laurens de Haan, Sidney I. Resnick, Holger Rootzén, and Casper G. de Vries. Extremal behaviour of solutions to a stochastic difference equation with applications to arch processes. *Stochastic Processes and their Applications*, 32(2):213–224, 1989. ISSN 0304-4149. doi: https://doi.org/10.1016/0304-4149(89)90076-8.

[13] Adiel Statman, Maya Kaufman, Amir Minerbi, Noam E. Ziv, and Naama Brenner. Synaptic size dynamics as an effectively stochastic process. *PLOS Computational Biology*, 10(10):1–17, 10 2014. doi: 10.1371/journal.pcbi.1003846.

[14] David Sheskin. *Handbook of parametric and nonparametric statistical procedures*. Chapman and Hall/CRC, Boca Raton [etc, 2nd ed. edition, 2000. ISBN 158488133X.

[15] B. W. Yap and C. H. Sim. Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12):2141–2155, 2011.

[16] M. A. Stephens. Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347):730–737, 1974. ISSN 01621459.

[17] Didier Sornette. *Critical phenomena in natural sciences : chaos, fractals, selforganization, and disorder : concepts and tools / Didier Sornette.* Springer series in synergetics. Springer, Berlin ;, 2nd ed. edition, 2004. ISBN 3540407545.

[18] Bruce M. Hill. A Simple General Approach to Inference About the Tail of a Distribution. *The Annals of Statistics*, 3(5):1163 – 1174, 1975. doi: 10.1214/aos/1176343247.

[19] Liran Hazan and Noam E. Ziv. Activity dependent and independent determinants of synaptic size diversity. *Journal of Neuroscience*, 40(14):2828–2848, 2020. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.2181-19.2020.

[20] Didier Sornette and Rama Cont. Convergent multiplicative processes repelled from zero: Power laws and truncated power laws. *Journal de Physique I*, 7(3):431–444, 1997. doi: 10.1051/jp1:1997169.

[21] Didier Sornette. Linear stochastic dynamics with nonlinear fractal properties. *Physica A-statistical Mechanics and Its Applications*, 250:295–314, 1998.

[22] Charles M. Goldie. Implicit Renewal Theory and Tails of Solutions of Random Equations. *The Annals of Applied Probability*, 1(1):126 – 166, 1991. doi: 10.1214/aoap/1177005985.

[23] Hideki Takayasu, Aki-Hiro Sato, and Misako Takayasu. Stable infinite variance fluctuations in randomly amplified langevin systems. *Phys. Rev. Lett.*, 79:966–969, Aug 1997. doi: 10.1103/PhysRevLett.79.966.

[24] Andreas Winkelbauer. Moments and absolute moments of the normal distribution, 2014.

[25] Simone Holler, German Köstinger, Kevan A C Martin, Gregor F P Schuhknecht, and Ken J Stratford. Structure and function of a neocortical synapse. *Nature*, 591(7848):111–116, 2021. ISSN 0028-0836. doi: 10.1038/s41586-020-03134-2.

[26] Anuradha M. Annaswamy and Alexander L. Fradkov. A historical perspective of adaptive control and learning. *Annual Reviews in Control*, 52:18–41, 2021. ISSN 1367-5788. doi: https://doi.org/10.1016/j.arcontrol.2021.10.014.