Stability and Rare Events in Stochastic Models

Sergey Foss

Heriot-Watt University, Edinburgh and

Institute of Mathematics, Novosibirsk

ANSAPW

University of Queensland

8-11 July, 2013

Outline

- (I) Fluid Approximation Approach for stability of queueing models
- (II) Stability of multi-component processes
- (III) Stochastic sequences with a regenerative structure that may depend both on the future and on the past
- (IV) Supremum of a random walk: the most likely way to exceed a high level.

(I) The Fluid Approximation Approach

- (I.1) Stability and instability criteria in terms of fluid limits
- (I.2) Extension: Random fluid limits
- (I.3) Not always applicable: Example
- (I.4) Measure-valued fluid limits: Open problem

(I.1) Fluid Limits

Assume that a dynamical behaviour of a stochastic system (queueing system or, more generally, stochastic network) may be described by discrete-time Markov chain $X_n, n = 0, 1, 2, \ldots$ taking values in state space $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$.

Assume further that $|\cdot|$ is some "(semi-)norm" and that $\{x \ : \ |x| \le c\}$ is a "compact set", for any $c > 0$.

Say, $\mathcal{X} = \mathcal{R}_+^d$ and $|x| = \sum x_i$ is the $\mathcal{L}_1$-norm.

Denote by $X_n^{(x)}$ a Markov chain that starts from initial value $X_0^{(x)} = x$ with $|x| > 0$.

Consider the following linear scaling, both in time and in space:

$$\overline{X}^{(x)}(t) = \frac{X^{(x)}_{[|x|t]}}{|x|}, \quad t \ge 0.$$

Clearly, $|\overline{X}^{(x)}(0)| = 1$.

Here $[t]$ is the biggest integer that does not exceed t.

Definition

Let sequence $x = x_n$ be such that $|x_n| \to \infty$ as $n \to \infty$.

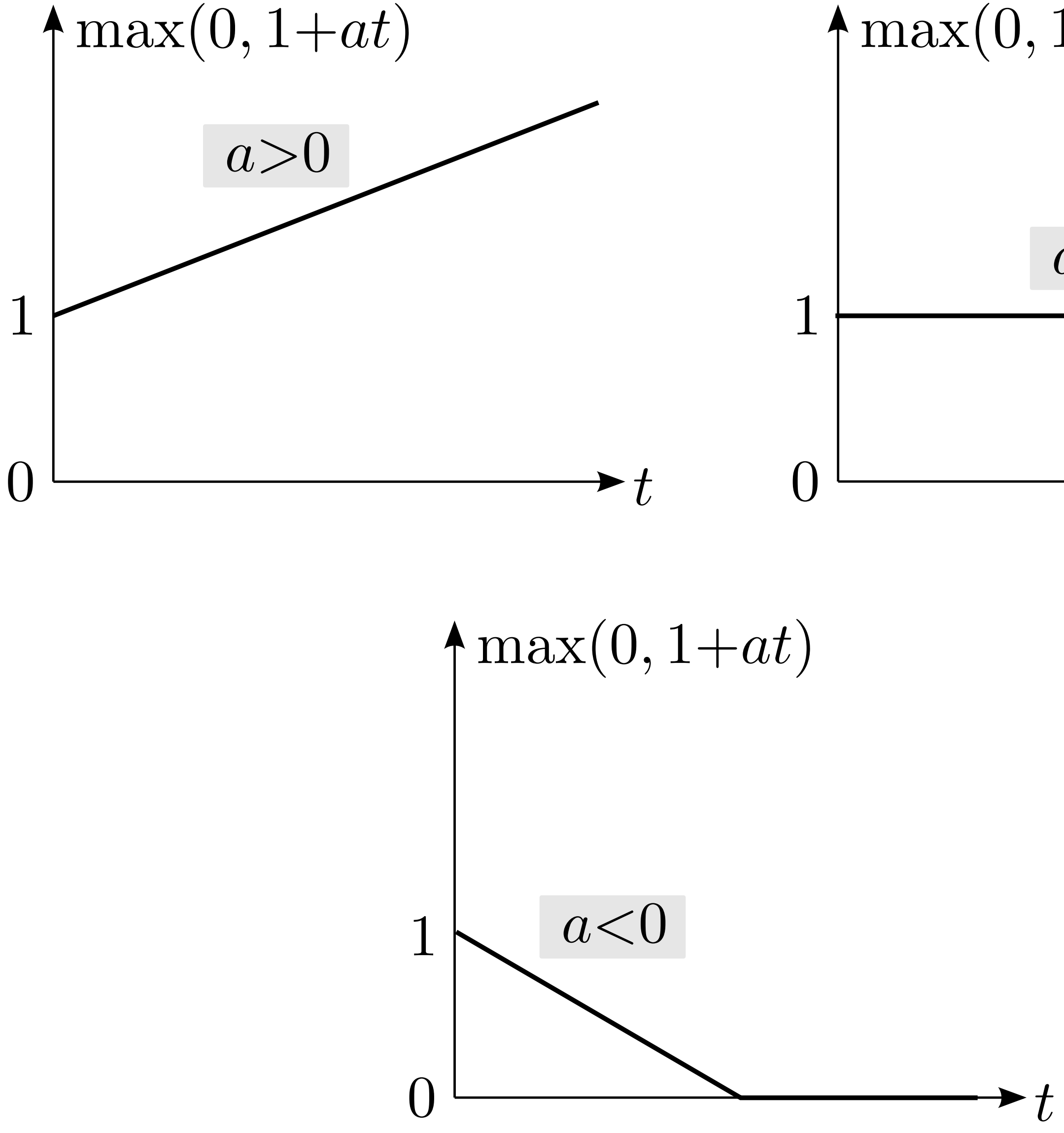
A *fluid limit* $Z(t), t \geq 0$ is a stochastic process such that, for any $t_0 > 0$, its restriction to time interval $[0, t_0]$ is a weak limit of a sequence of random processes $\{\overline{X}^{(x_n)}(t), 0 \leq t \leq t_0\}$.

A collection of *all* fluid limits is a *fluid model.*

Example. Single-server queue: exponential times $\{t_n\}$ with mean $1/\lambda$, exponential service times $\{\sigma_n\}$ with mean $1/\mu$, drift $a = 1/\mu - 1/\lambda$.

("Exponential" assumptions are for simplicity).

X_n is a waiting time (or workload) at the nth arrival instant.


max$(0, 1+at)$
$a>0$
1
0
t
max$(0, 1+at)$
$a=0$
1
0
t
max$(0, 1+at)$
$a<0$
1
0
t

Example. Two single-server queues in tandem, $M/M/1 \to M/1$.

Exponential interarrival times $\{t_n\}$ with mean $1/\lambda$, exponential service times $\{\sigma_{1,n}\}$ with mean $1/\mu_1$ in queue 1 and exponential service times $\{\sigma_{2,n}\}$ with mean $1/\mu_2$ in queue 2.

Assume $\lambda < \min(\mu_1, \mu_2)$.

Remark. In these examples, a fluid limit is

- unique, up to the initial value,
- deterministic and
- piece-wise linear.

This is frequent in queueing and other stochastic networks.

For models with deterministic fluid limits, the following stability criterion is useful.

Stability Criterion via Fluid Limits

(Rybko-Stolyar, Dai, Stolyar, ...)

Criterion.

Under some technical conditions, the following holds:

If there exists time T and number $\varepsilon \in (0,1)$ such that, for any fluid limit $Z(t)$, we have $|Z(T)| \leq 1 - \varepsilon$ a.s.,

then the stochastic system is *stable*.

This means that the underlying Markov chain $X_n, n = 0, 1, \ldots$ is *positive recurrent*.

Under a further minorization condition, this implies convergence to a/the limiting distribution in the total variation norm.

However, for models with random fluid limits, this criterion may never work because, for any such limit, random variable $\sup_t |Z(t)|$ may have unbounded support.

(I.2) Example of a random fluid limit

Assume that the system starts with N customers. Assume that, in addition, there is a Poisson input stream with rate λ such that

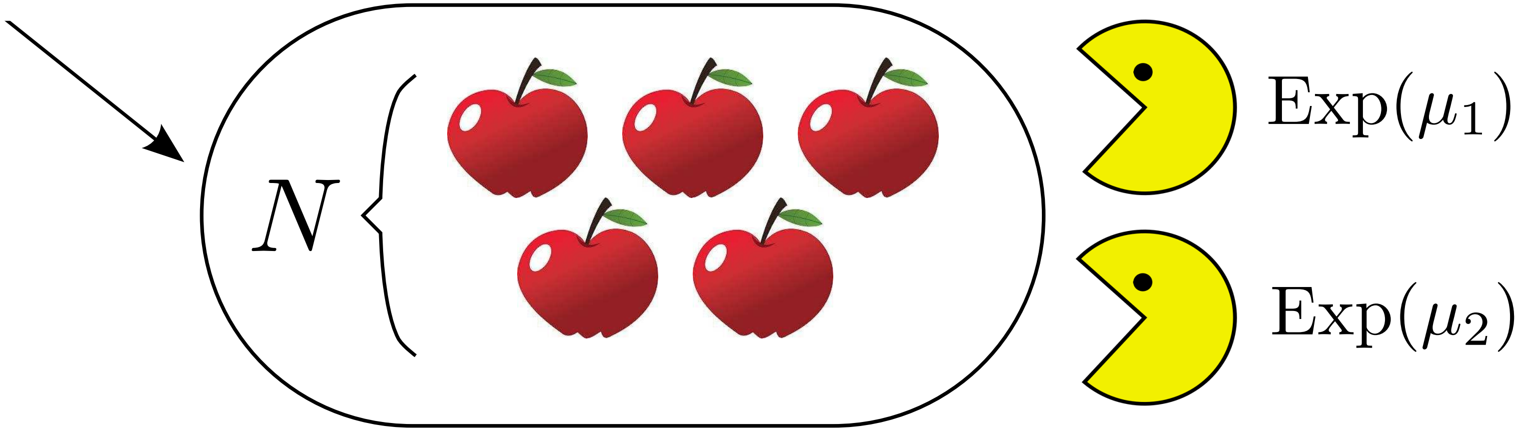
$$\mu_1 + \mu_2 > \lambda > \max(\mu_1, \mu_2).$$

Assume that a server “wins” and leaves the system if he does not find a customer to serve.

Example

$\lambda < \mu_1 + \mu_2$

$\lambda > \mu_1,\ \mu_2$



N

$\mathrm{Exp}(\mu_1)$

$\mathrm{Exp}(\mu_2)$

1st wins w.p. $\frac{\mu_1}{\mu_1+\mu_2}$

2nd wins w.p. $\frac{\mu_2}{\mu_1+\mu_2}$

Assume now that if a second server does not find a customer to serve, he also leaves the system.

If, say, the first server “wins” and leaves the system, then the second server has two chances, either also to leave the system or to stay (since $\mu_2 < \lambda$). This is a simple birth-and-death process, and probability to leave is μ_2/λ.

Then we may conclude that the following may occur: either

- both servers leave the system, this occurs with probability

$$\frac{\mu_1}{\mu_1+\mu_2} \cdot \frac{\mu_2}{\lambda} + \frac{\mu_2}{\mu_1+\mu_2} \cdot \frac{\mu_1}{\lambda} = \frac{2\mu_1\mu_2}{\lambda(\mu_1+\mu_2)}$$

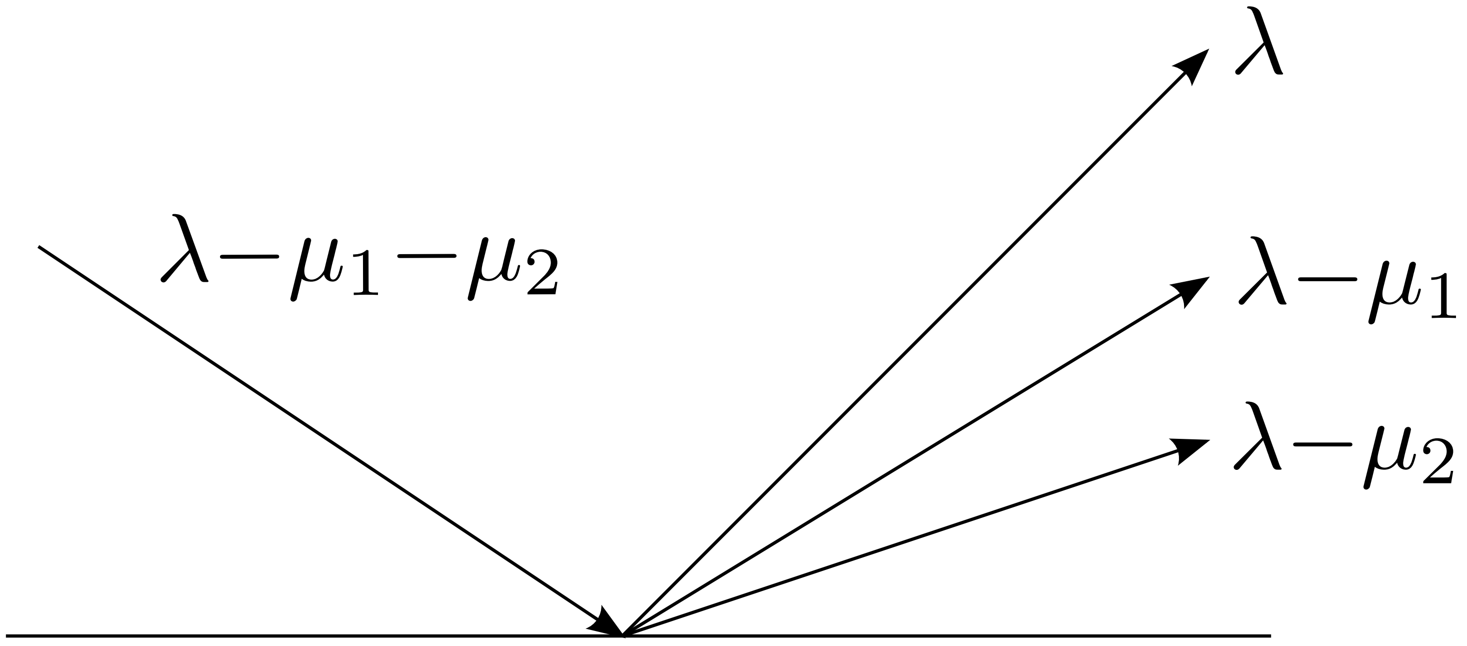
- or only the first server leaves, this occurs with probability

$$\frac{\mu_1}{\mu_1+\mu_2} \cdot \frac{1-\mu_2}{\lambda}$$

- or only the the second server leaves, this occurs with probability

$$\frac{\mu_2}{\mu_1+\mu_2} \cdot \frac{1-\mu_1}{\lambda}$$

When we turn to the fluid limit, we get:



Stability criterion it terms of random fluid limits

Criterion for RFL.

Under some technical conditions, the following holds:

If, for any fluid limit Z, there exists stopping time T_Z and number $\varepsilon \in (0,1)$ such that

(a) the family of stopping times $\{T_z\}$ is uniformly integrable and,

for any fluid limit Z, $\mathbf{E}|Z(T_Z)| \le 1-\varepsilon$,

then the stochastic system is *stable*.

This means that the underlying Markov chain $X_n, n = 0, 1, \ldots$ is *positive recurrent*.

Remark By Jensen's inequality,

$$\log \mathbf{E} Z_1(T) > \mathbf{E} \log Z_1(T)$$

if this is a non-degenerate random variable. Then the following is possible:

$$\log \mathbf{E} Z_1(T) > 0 > \mathbf{E} \log Z_1(T).$$

Significant difference between conditions for positive recurrence and for recurrence

(I.3) Example of a model whole stability condition depends on a whole distribution of a service time

Consider a polling system with 2 stations and a single server that consequently visits the stations and serve customers there. Customers arrive at station $k = 1, 2$ in a Poisson stream of intensity λ_k. Service times at station k form an i.i.d. sequence with a general distribution B_k with a positive finite mean b_k. It takes an exponential time with mean γ for the server to travel from station 1 to station 2 and also from station 2 to station 1. The server follows the *exhaustive* policy at station 2 and an *adaptive limited* policy at station 1. In more detail, the server works in “cycles”, and each cycle starts when the server arrives at station 2.

If the server finds customers there, he starts to serve them one-by-one (including new arrivals) until he empties the queue. Then he travels to station 1 (during an exponential time), serves $\min(1, q)$ customers at station 1 if there are q customers there, and then travels (for another exponential time) to station 2. This is a *standard* cycle.

If, upon his arrival to station 2, the server finds it empty, the new cycle is *modified*: the server is allowed to serve m extra customers at station 1. More precisely, now he starts the cycle with his travel to station 1, serves $\min(1 + m, q)$ customers there and then travels back to station 2. Here m is a fixed non-negative integer.

Theorem The underlying Markov chain is positive recurrent if and only if inequality

$$\rho + 2\gamma\lambda_1 < 1 + mV(1 - \rho),$$

where $\rho = \lambda_1 b_1 + \lambda_2 b_2$ and V a positive constant which is a (known) function of $\mathbf{E}e^{-\lambda_2 \sigma_1}$.

Here σ_1 is a typical service time at queue 1.

(I.4) Measure-valued fluid limits

(2) Multi-component processes

We assume that $\{X_n\}$ is semi-Markov:

it can be “markovized”, (X_n, Y_n) is Markov.

We look for conditions for $\{X_n\}$ to be stable.

Here Y_n may tend to “infinity” or the pair may be stable (in a certain sense). Also, Y_n may be “null-recurrent”.

(II.1) Example: tandem queue

Assume $\mu_1 < \lambda < \mu_2$.

Let X_n be the waiting time of customer n in front of queue 2, and Y_n its waiting time in front of queue 1.

Here $Y_n \to \infty$ and X_n converges to the stationary waiting time in a single-server queue with “inter-arrival” times $\{\sigma_{1,n}\}$ and service times $\{\sigma_{2,n}\}$.

Other examples: I. Adan and G. Weiss

(II.2) Y_n is a *driving* sequence

We assume that Y_n is a Markov chain itself, with a positive recurrent atom.

Two examples: with drift condition and with monotone condition.

(II.3) Open problem: stability of a multiple-access model with harvesting

Jeongho Jeon, Anthony Ephremides

$http : //arxiv.org/abs/1112.5995$

Remarks

(III) Stochastic sequences with a regenerative structure that may depend both on the future and on the past

Convergence of *functionals* of stochastic recursions.

Examples:

- Random walk with positive drift
- Contact process in discrete time
- Infinite bin model
- Continious-space version of infinite bin model
- Harris ergodic Markov chain

See S Foss and S Zachary, Stochastic sequences with a regenerative structure that may depend both on the future and on the past

http://arxiv.org/abs/1212.1475

(to appear in Adv Appl Probab)

for general statements and more “advanced” examples.

(IV) Supremum of a random walk: the most likely way to exceed a high level

Let $S_0 = 0$, $S_n = \sum_1^n \xi_i$ where $\{\xi_i\}$ are i.i.d.

Let $M = \sup_{n \geq 0} S_n$ and assume M is finite a.s.

Asymptotics for $\mathbf{P}(M > x)$, as $x \to \infty$

(ONLY ASYMPTOTICS!)

There are 5 cases, two “heavy-tailed” and three “light-tailed”.

- (HT1) (“Classical HT”): $\mathbf{E}|\xi| < \infty$, $\mathbf{E}\xi < 0$, and $\mathbf{E}e^{c\xi} = \infty$, for any $c > 0$. The principle of a *single big jump*.

- (HT2) $\mathbf{E}|\xi| = \infty$. The principle of a *single big jump*, but for non-linear trajectory.

- (LT1) (“Classical LT”): $\mathbf{E}|\xi| < \infty$, $\mathbf{E}\xi < 0$, then $\gamma := \sup\{c \geq 0 \ : \ \mathbf{E}e^{c\xi} \leq 1\} > 0$ and, further, $\mathbf{E}e^{\gamma\xi} = 1$ and $b := \mathbf{E}\xi e^{\gamma\xi} < \infty$. *Solidarity* property.

- (LT2) (“S-gamma LT”): $\mathbf{E}|\xi| < \infty$, $\mathbf{E}\xi < 0$, then $\gamma := \sup\{c \geq 0 \ : \ \mathbf{E}e^{c\xi} \leq 1\} > 0$ and $\mathbf{E}e^{\gamma\xi} < 1$. The principle of a *single big jump*, but at *fixed* times.

- (LT3) (Intermediate): $\mathbf{E}|\xi| < \infty$, $\mathbf{E}\xi < 0$, then $\gamma := \sup\{c \geq 0 \ : \ \mathbf{E}e^{c\xi} \leq 1\} > 0$ and, further, $\mathbf{E}e^{\gamma\xi} = 1$ and $b := \mathbf{E}\xi e^{\gamma\xi} = \infty$. Conditional “compound Poisson”.

Open problem: Continuous spetrum...

Similar results may be (and some of them have been) obtained in queueing networks (tandem queues, multi-server queues, etc.).

Open problems: Asymptotics for the stationary sojourn time in a generalized Jackson network in cases (HT1, HT2, LT2, LT3).

ADVERT:

SECOND EDITION of

S Foss, D Korshunov and S Zachary

An Introduction to Heavy-Tailed and Subexponential Distributions,

Springer, June 2013

with new Sections and many exercises