a tutorial on: LÉVY-DRIVEN QUEUES

Michel Mandjes

Korteweg-de Vries Institute for Mathematics, University of Amsterdam, the Netherlands &
CWI, Amsterdam, the Netherlands &
EURANDOM, Eindhoven, the Netherlands

Carleton University, Ottawa
Conference on
Approximations, Asymptotics and Resource Management for Stochastic Networks
(sponsored by the Fields Institute and MITACS)

August, 2010

Thanks!

Thanks!

Thanks, Yiqiang, for inviting me!

Thanks!

Thanks, Yiqiang, for inviting me!

Thanks, Minyi and Yiqiang, for organizing this event!

Organization

This mini-course will take 6 hours (4 times 1.5 hours), spread over Wed-Sat.

Organization

This mini-course will take 6 hours (4 times 1.5 hours), spread over Wed-Sat.

Structure:

- ⋆ What are Lévy processes? What can you use them for?
- ⋆ What is a queue with Lévy input? Why are they relevant?
- ⋆ Results on stationary and transient behavior of Lévy-driven queues;
- ⋆ Asymptotics;
- ⋆ Variants of the standard queue;
- ⋆ Lévy-driven networks (multiple queues).

PART I:

WHAT ARE LÉVY PROCESSES?

What are Lévy processes?

Definition:

Lévy processes are stochastic processes with stationary independent increments.

What are Lévy processes?

Definition:

Lévy processes are stochastic processes with stationary independent increments.

Lévy process $(X_t)_t$, in continuous time (i.e., $t \in \mathbb{R}$):

- ⋆ Stationary increments: distribution of $X_{t+s} - X_t$ only depends on s (*length* of the interval), and not on t (*position* of the interval).
- ⋆ Independent increments: $X_{t+s} - X_t$ does not depend on X_t, for all $s \geq 0$.

Lévy processes: examples

Lévy processes: examples

Brownian motion!

Lévy processes: examples

Brownian motion!

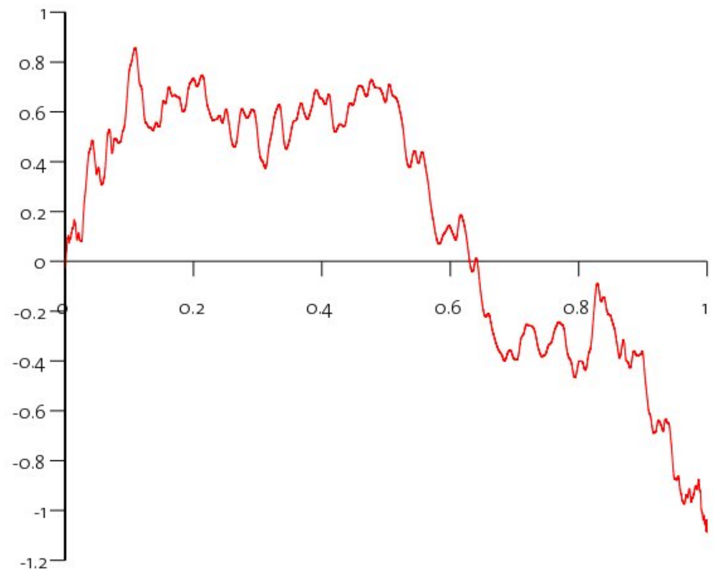
$(X_t)_t$ is Normally distributed with mean μt and variance $\sigma^2 t$.

Lévy processes: examples

Brownian motion!

$(X_t)_t$ is Normally distributed with mean μt and variance $\sigma^2 t$.

Sample path:



Lévy processes: examples

Brownian motion: $(X_t)_t$ is Normally distributed with mean μt and variance $\sigma^2 t$.

$$\mathbb{E}e^{-\alpha X_t} = \int_{-\infty}^{\infty} e^{-\alpha x} \frac{1}{\sqrt{2\pi}\sigma t} \exp\left(-\frac{(x-\mu t)^2}{2\sigma^2}\right) \mathrm{d}x.$$

Lévy processes: examples

Brownian motion: $(X_t)_t$ is Normally distributed with mean μt and variance $\sigma^2 t$.

$$\mathbb{E}e^{-\alpha X_t} = \int_{-\infty}^{\infty} e^{-\alpha x} \frac{1}{\sqrt{2\pi}\sigma t} \exp\left(-\frac{(x-\mu t)^2}{2\sigma^2}\right) \mathrm{d}x.$$

Easier though: $X_t \stackrel{\mathrm{d}}{=} \mu t + \sigma\sqrt{t}\,U$, with U standard Normal!

Hence: $\mathbb{E}e^{-\alpha X_t} = e^{-\alpha\mu t}\,\mathbb{E}e^{-\alpha\sigma\sqrt{t}U}$, and

$$\begin{aligned}
\mathbb{E}e^{-\alpha U} &= \int_{-\infty}^{\infty} e^{-\alpha x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \\
&= e^{\alpha^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x+\alpha)^2}{2}\right) \mathrm{d}x = e^{\alpha^2/2}.
\end{aligned}$$

Lévy processes: examples

Brownian motion: $(X_t)_t$ is Normally distributed with mean μt and variance $\sigma^2 t$.

$$\mathbb{E}e^{-\alpha X_t} = \int_{-\infty}^{\infty} e^{-\alpha x} \frac{1}{\sqrt{2\pi}\sigma t} \exp\left(-\frac{(x-\mu t)^2}{2\sigma^2}\right) \mathrm{d}x.$$

Easier though: $X_t \stackrel{\mathrm{d}}{=} \mu t + \sigma\sqrt{t}\,U$, with U standard Normal!

Hence: $\mathbb{E}e^{-\alpha X_t} = e^{-\alpha\mu t}\,\mathbb{E}e^{-\alpha\sigma\sqrt{t}U}$, and

$$\begin{aligned} \mathbb{E}e^{-\alpha U} &= \int_{-\infty}^{\infty} e^{-\alpha x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \\ &= e^{\alpha^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x+\alpha)^2}{2}\right) \mathrm{d}x = e^{\alpha^2/2}. \end{aligned}$$

Conclude:

$$\mathbb{E}e^{-\alpha X_t} = e^{-\alpha\mu t}\, e^{\alpha^2\sigma^2 t}.$$

Lévy processes: examples

Brownian motion:

$$\mathbb{E}e^{-\alpha X_t} = e^{-\alpha\mu t}\, e^{\alpha^2\sigma^2 t}.$$

This can be rewritten as:

$$\mathbb{E}e^{-\alpha X_t} = e^{\varphi(\alpha)t},$$

with

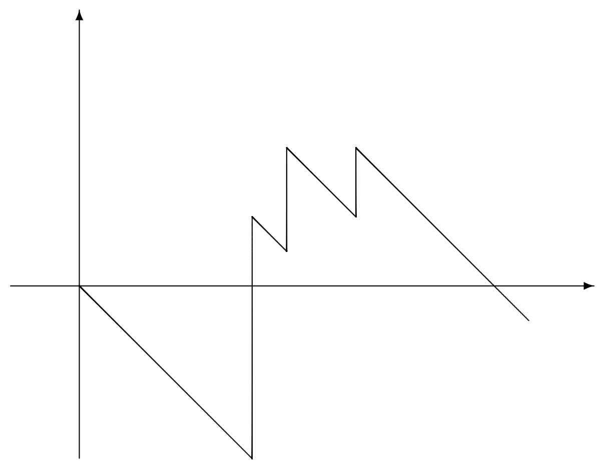
$$\varphi(\alpha) := -\alpha\mu + \frac{1}{2}\alpha^2\sigma^2.$$

We write: $X \in \mathbb{B}\mathrm{m}(\mu, \sigma^2)$.

Lévy processes: examples

Compound Poisson with drift!

Sample path:



Lévy processes: examples

Compound Poisson with drift:

- ⋆ Jobs arrive according to a Poisson process with rate λ;
- ⋆ The jobs are i.i.d. samples from a (nonnegative) distribution B, with Laplace transform

$$b(\alpha) := \mathbb{E}e^{-\alpha B}.$$

- ⋆ the storage system is depleted at rate r.

Lévy processes: examples

Compound Poisson with drift:

$\mathbb{E}e^{-\alpha X_t}$ can be computed by conditioning on N_t, i.e., the number of jumps in $[0,t]$:

$$\mathbb{P}(N_t = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}.$$

Hence,

$$\mathbb{E}e^{-\alpha X_t} = e^{r\alpha t}\sum_{k=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^k}{k!}(b(\alpha))^k = e^{r\alpha t}\exp(-\lambda t(1-b(\alpha))).$$

Lévy processes: examples

Compound Poisson with drift:

$$\mathbb{E}e^{-\alpha X_t} = e^{r\alpha t} \cdot \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} (b(\alpha))^k = e^{r\alpha t} \exp(-\lambda t(1 - b(\alpha)).$$

Note that we again have that

$$\mathbb{E}e^{-\alpha X_t} = e^{\varphi(\alpha) t},$$

but now with

$$\varphi(\alpha) := r\alpha - \lambda + \lambda b(\alpha).$$

We write: $X \in \mathbb{C}\mathrm{P}(r, \lambda, b(\cdot))$.

Sample paths

Observe:

- ⋆ There are continuous Lévy processes (Brownian motion),
- ⋆ but also processes with jumps.

Sample paths

Observe:

- ⋆ There are continuous Lévy processes (Brownian motion),
- ⋆ but also processes with jumps.

The class of Lévy processes is broad and versatile.

Infinite divisibility

X_t is, for any t, *infinitely divisible*:

we have the distributional equality, with $X_t^{(i)}$ i.i.d. copies of X_t:

$$X_t \stackrel{\mathrm{d}}{=} \sum_{i=1}^{n} X_{t/n}^{(i)},$$

for any $n \in \mathbb{N}$.

Infinite divisibility

X_t is, for any t, *infinitely divisible*:

we have the distributional equality, with $X_t^{(i)}$ i.i.d. copies of X_t:

$$X_t \stackrel{\mathrm{d}}{=} \sum_{i=1}^{n} X_{t/n}^{(i)},$$

for any $n \in \mathbb{N}$.

Each Lévy process can be associated with an infinitely divisible distribution, and vice versa.

Characteristic triplet

Alternatively, for any value of t,

$$\log \mathbb{E} e^{sX_t} = t \log \mathbb{E} e^{sX_1},$$

where $s \in \mathbb{C}$.

Characteristic triplet

Alternatively, for any value of t,

$$\log \mathbb{E}e^{sX_t} = t \log \mathbb{E}e^{sX_1},$$

where $s \in \mathbb{C}$.

More specific characterization of Lévy processes:

the so-called *Lévy exponent* $\log \mathbb{E}e^{sX_1}$ is necessarily of the form

$$\log \mathbb{E}e^{sX_1} = sd + \frac{1}{2}s^2\sigma^2 + \int_{-\infty}^{\infty} (e^{sx} - 1 - sx1_{[0,1)}(|x|))\Pi(\mathrm{d}x),$$

where (d, σ^2, Π) is commonly referred to as the *characteristic triplet*.

Characteristic triplet

$$\log \mathbb{E} e^{sX_1} = sd + \frac{1}{2}s^2\sigma^2 + \int_{-\infty}^{\infty}(e^{sx} - 1 - sx1_{[0,1)}(|x|))\Pi(\mathrm{d}x).$$

Suppose $X \in \mathbb{B}\mathrm{m}(\mu, \sigma^2)$.

Then $d = \mu$, $\sigma^2 = \sigma^2$, and $\Pi \equiv 0$.

Characteristic triplet

$$\log \mathbb{E} e^{sX_1} = sd + \frac{1}{2} s^2 \sigma^2 + \int_{-\infty}^{\infty} (e^{sx} - 1 - sx 1_{[0,1)}(|x|)) \Pi(\mathrm{d}x).$$

Suppose $X \in \mathbb{B}\mathrm{m}(\mu, \sigma^2)$.

Then $d = \mu$, $\sigma^2 = \sigma^2$, and $\Pi \equiv 0$.

Suppose $X \in \mathbb{C}\mathrm{P}(r, \lambda, b(\cdot))$.

Then

$$d = -r + \lambda \int_0^1 x \Pi(\mathrm{d}x),$$

$\sigma^2 = 0$, and $\Pi(\mathrm{d}x) = \lambda \, \mathrm{d}\mathbb{P}(B \leq x)$ on $[0, \infty)$.

Spectrally one-sided Lévy processes

Let $(X_t)_{t\geq 0}$ be a Lévy process, with drift $\mathbb{E}X_1 < 0$.

Two special cases:

(A) $(X_t)_{t\geq 0}$ has no negative jumps, or is *spectrally positive*;

(B) $(X_t)_{t\geq 0}$ has no positive jumps, or is *spectrally negative*.

Spectrally positive Lévy processes

$(X_t)_{t\geq 0}$ has no negative jumps, or is *spectrally positive*; we write $X \in \mathscr{S}_+$.

Spectrally positive Lévy processes

$(X_t)_{t\geq 0}$ has no negative jumps, or is *spectrally positive*; we write $X \in \mathscr{S}_+$.

Then the *Laplace exponent* is given by the function $\varphi(\cdot) : [0,\infty) \mapsto [0,\infty)$, i.e.,

$$\varphi(\alpha) := \log \mathbb{E} e^{-\alpha X_1}.$$

It is known that $\varphi(\cdot)$ is increasing and convex on $[0,\infty)$, with slope

$$\varphi'(0) = \lim_{\alpha \downarrow 0} \frac{\mathbb{E}(-X_1 e^{-\alpha X_1})}{\mathbb{E} e^{-\alpha X_1}} = -\mathbb{E} X_1 > 0$$

in the origin.

Spectrally positive Lévy processes

$(X_t)_{t\geq 0}$ has no negative jumps, or is *spectrally positive*; we write $X \in \mathscr{S}_+$.

Then the *Laplace exponent* is given by the function $\varphi(\cdot) : [0,\infty) \mapsto [0,\infty)$, i.e.,

$$\varphi(\alpha) := \log \mathbb{E} e^{-\alpha X_1}.$$

It is known that $\varphi(\cdot)$ is increasing and convex on $[0,\infty)$, with slope

$$\varphi'(0) = \lim_{\alpha \downarrow 0} \frac{\mathbb{E}(-X_1 e^{-\alpha X_1})}{\mathbb{E} e^{-\alpha X_1}} = -\mathbb{E} X_1 > 0$$

in the origin.

Therefore the inverse $\psi(\cdot)$ of $\varphi(\cdot)$ is well-defined on $[0,\infty)$.

Spectrally positive Lévy processes

$(X_t)_{t\geq 0}$ has no negative jumps, or is *spectrally positive*; we write $X \in \mathscr{S}_+$.

Then the *Laplace exponent* is given by the function $\varphi(\cdot) : [0,\infty) \mapsto [0,\infty)$, i.e.,

$$\varphi(\alpha) := \log \mathbb{E} e^{-\alpha X_1}.$$

It is known that $\varphi(\cdot)$ is increasing and convex on $[0,\infty)$, with slope

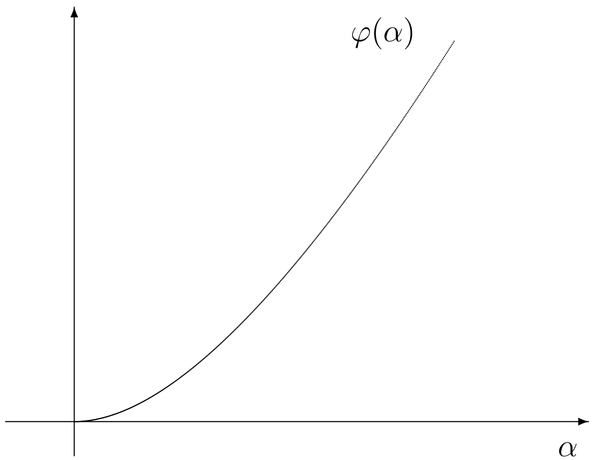
$$\varphi'(0) = \lim_{\alpha \downarrow 0} \frac{\mathbb{E}(-X_1 e^{-\alpha X_1})}{\mathbb{E} e^{-\alpha X_1}} = -\mathbb{E} X_1 > 0$$

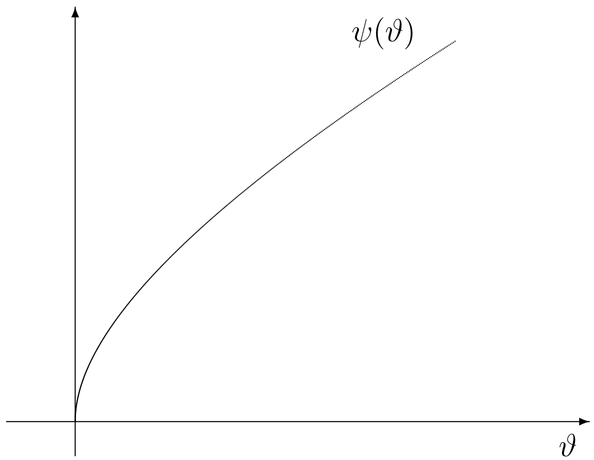
in the origin.

Therefore the inverse $\psi(\cdot)$ of $\varphi(\cdot)$ is well-defined on $[0,\infty)$.

(Technical note: In the sequel we also require that X_t is not a *subordinator*, i.e., a monotone process; thus X_1 has probability mass on the positive half-line, which implies that $\lim_{\alpha \to -\infty} \varphi(\alpha) = \infty$.)

Spectrally positive Lévy processes





Spectrally negative Lévy processes

$(X_t)_{t \geq 0}$ has no positive jumps, or is *spectrally negative*; we write $X \in \mathscr{S}_-$.

Spectrally negative Lévy processes

$(X_t)_{t\geq 0}$ has no positive jumps, or is *spectrally negative*; we write $X \in \mathscr{S}_-$.

Now we define $\Phi(\beta) := \log \mathbb{E} e^{\beta X_1}$, which is well-defined for any $\beta \geq 0$.

Spectrally negative Lévy processes

$(X_t)_{t\geq 0}$ has no positive jumps, or is *spectrally negative*; we write $X \in \mathscr{S}_-$.

Now we define $\Phi(\beta) := \log \mathbb{E}e^{\beta X_1}$, which is well-defined for any $\beta \geq 0$.

Realize that

$$\Phi'(0) = \lim_{\beta \downarrow 0} \frac{\mathbb{E}(X_1 e^{\beta X_1})}{\mathbb{E}e^{\beta X_1}} = \mathbb{E}X_1 < 0;$$

hence $\Phi(\beta)$ is *no* bijection on $[0, \infty)$.

Spectrally negative Lévy processes

$(X_t)_{t\geq 0}$ has no positive jumps, or is *spectrally negative*; we write $X \in \mathscr{S}_-$.

Now we define $\Phi(\beta) := \log \mathbb{E}e^{\beta X_1}$, which is well-defined for any $\beta \geq 0$.

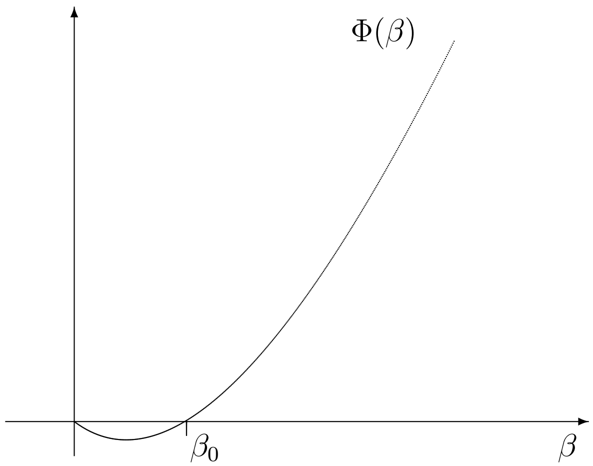
Realize that

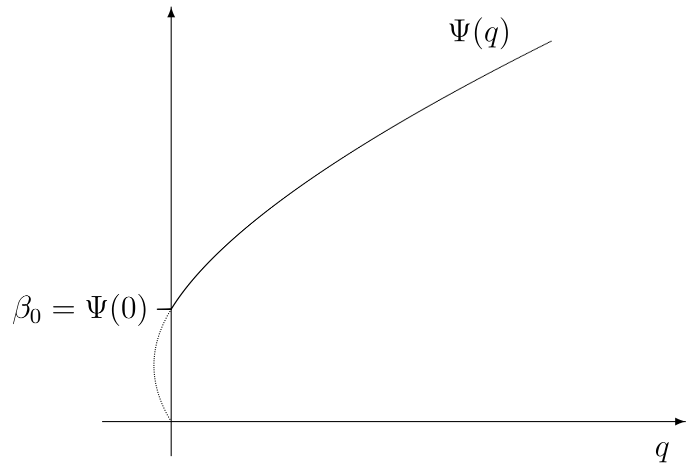
$$\Phi'(0) = \lim_{\beta \downarrow 0} \frac{\mathbb{E}(X_1 e^{\beta X_1})}{\mathbb{E}e^{\beta X_1}} = \mathbb{E}X_1 < 0;$$

hence $\Phi(\beta)$ is *no* bijection on $[0, \infty)$.

Define the *right* inverse through $\Psi(q) := \sup\{\beta \geq 0 : \Phi(\beta) = q)$. Realize that $\beta_0 := \Psi(0) > 0$.

Spectrally negative Lévy processes





Spectrally one-sided Lévy processes

Brownian motion:

$$\mathbb{B}\mathrm{m}(\mu, \sigma^2) \subset \mathscr{S}_+, \quad \text{but also} \quad \mathbb{B}\mathrm{m}(\mu, \sigma^2) \subset \mathscr{S}_-.$$

Compound Poisson:

$$\mathbb{C}\mathrm{P}(r, \lambda, b(\cdot)) \subset \mathscr{S}_+.$$

α-stable Lévy motion

A random variable Y has a stable distribution if for any $a, b > 0$ there are a $c > 0$ and $d \in \mathbb{R}$ such that

$$aY' + bY'' \stackrel{\mathrm{d}}{=} cY + d,$$

where Y' and Y'' are independent copies of Y.

α-stable Lévy motion

A random variable Y has a stable distribution if for any $a, b > 0$ there are a $c > 0$ and $d \in \mathbb{R}$ such that

$$aY' + bY'' \stackrel{\mathrm{d}}{=} cY + d,$$

where Y' and Y'' are independent copies of Y.

It turns out that the characteristic function of Y now necessarily looks like

$$\log \mathbb{E}e^{\mathrm{i}\theta Y} = \begin{cases} -\sigma^\alpha|\theta|^\alpha(1 - \mathrm{i}\beta \mathrm{sign}(\theta)\tan(\pi\alpha/2) + \mathrm{i}m\theta & \alpha \neq 1; \\ -\sigma|\theta|(1 + \mathrm{i}\beta\pi/2\mathrm{sign}(\theta)\log|\theta| + \mathrm{i}m\theta & \alpha = 1, \end{cases}$$

where $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\sigma \in [0, \infty)$, $m \in \mathbb{R}$, and $\mathrm{sign}(x) = 1_{(0,\infty)}(x) - 1_{(-\infty,0)}(x)$.

α-stable Lévy motion

A random variable Y has a stable distribution if for any $a, b > 0$ there are a $c > 0$ and $d \in \mathbb{R}$ such that

$$aY' + bY'' \stackrel{\mathrm{d}}{=} cY + d,$$

where Y' and Y'' are independent copies of Y.

It turns out that the characteristic function of Y now necessarily looks like

$$\log \mathbb{E} e^{\mathrm{i}\theta Y} = \begin{cases} -\sigma^\alpha |\theta|^\alpha (1 - \mathrm{i}\beta \mathrm{sign}(\theta) \tan(\pi\alpha/2) + \mathrm{i}m\theta & \alpha \neq 1; \\ -\sigma|\theta|(1 + \mathrm{i}\beta\pi/2 \mathrm{sign}(\theta) \log|\theta| + \mathrm{i}m\theta & \alpha = 1, \end{cases}$$

where $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\sigma \in [0, \infty)$, $m \in \mathbb{R}$, and $\mathrm{sign}(x) = 1_{(0,\infty)}(x) - 1_{(-\infty,0)}(x)$.

Write: Y is distributed $S_\alpha(\sigma, \beta, m)$.

α-stable Lévy motion

Meaning of the parameters:

- α: *index of stability*; directly related to the 'heaviness' of the tail.

  In particular, if $\alpha \in (0, 1]$, then $\mathbb{E}|Y| = \infty$ (for $\alpha = 1$ we have the Cauchy distribution).

  If $\alpha = 2$ we obtain the Normal distribution.

- β: *skewness*. The extreme cases are $\beta = 1$, corresponding to a random variable with nonnegative support, and $\beta = -1$, in which case the support is nonpositive.

  Choosing $\beta = 0$ and $m = 0$ leads to a symmetric distribution.

- σ: *scale parameter*.

- For $\alpha \in (1, 2]$, we have that $\mathbb{E}Y = m$. This explains why m is called the *shift parameter*.

α-stable Lévy motion

Asymptotics:

Let $Y \stackrel{\mathrm{d}}{=} S_\alpha(\sigma, \beta, m)$. Then, as $u \to \infty$,

$$\mathbb{P}(Y > u)u^\alpha \to C_{\alpha,\sigma}\left(\frac{1+\beta}{2}\right),$$

where

$$C_{\alpha,\sigma} := \begin{cases} \sigma^\alpha(1-\alpha)/\left(\Gamma(2-\alpha)\cos(\pi\alpha/2)\right) & \alpha \neq 1; \\ 2\sigma/\pi & \alpha = 1. \end{cases}$$

α-stable Lévy motion

Having defined stable distribution, we now introduce α-stable Lévy motions.

X_t is an *α-stable Lévy motion* if $(X_t)_t$ has stationary independent increments such that

$$X_t \overset{\mathrm{d}}{=} S_\alpha(t^{1/\alpha}, \beta, m);$$

we write $X \in \mathbb{S}(\alpha, \beta, m)$.

If $\beta = \pm 1$, then $X \in \mathscr{S}_\pm$.

α-stable Lévy motion

One could say that α-stable Lévy motions are *self-similar*:

Picking $m = 0$, and writing $(X_t^{(\alpha)})_t$ to stress the dependence on α, one has

$$X_{Mt}^{(\alpha)} \stackrel{\mathrm{d}}{=} M^{1/\alpha} X_t^{(\alpha)}$$

(unless $\alpha = 1$, $\beta \neq 0$).

In other words: when zooming in, one essentially sees the same pattern, given that one adjusts the axes in a suitable fashion.

Application areas

- ⋆ Financial models;
- ⋆ Communication networks.

Application areas

Financial models:

Lévy processes have turned out to accurately match all kinds of financial processes.

Applications in option pricing, credit risk, etc. (useful: Lévy processes allow *jumps*).

Compound Poisson with drift is a classical model in ruin and insurance theory.

Application areas

Communication networks:

Under very general conditions, the input process of a broad class of (short-range dependent) queueing systems converges to Brownian motion (cf. functional Central Limit Theorem) —

see e.g. book by Whitt.

If input traffic has heavy-tailed characteristics (e.g. on-off sources with heavy-tailed on-times), then there is convergence to α-stable Lévy motion —

see e.g. Taqqu *et al.*, Mikosch *et al.*

A useful lemma

Consider $X \in \mathscr{S}_+$, and let

$$\tau(x) := \inf\{t \geq 0 : X_t \leq -x\}.$$

Observe that $e^{-\varphi(\alpha)t}\,\mathbb{E}e^{-\alpha X_t}$ is a mean-1 martingale.

A useful lemma

Consider $X \in \mathscr{S}_+$, and let

$$\tau(x) := \inf\{t \geq 0 : X_t \leq -x\}.$$

Observe that $e^{-\varphi(\alpha)t}\,\mathbb{E}e^{-\alpha X_t}$ is a mean-1 martingale.

- ⋆ Only positive jumps: $X_{\tau(x)} = -x$;
- ⋆ assuming that $\mathbb{E}X_1 < 0$, it holds that $\tau(x) < \infty$ almost surely.

A useful lemma

Consider $X \in \mathscr{S}_+$, and let

$$\tau(x) := \inf\{t \geq 0 : X_t \leq -x\}.$$

Observe that $e^{-\varphi(\alpha)t}\,\mathbb{E}e^{-\alpha X_t}$ is a mean-1 martingale.

- ⋆ Only positive jumps: $X_{\tau(x)} = -x$;
- ⋆ assuming that $\mathbb{E}X_1 < 0$, it holds that $\tau(x) < \infty$ almost surely.

Then 'optional sampling' implies

$$1 = \mathbb{E}e^{-\varphi(\alpha)\tau(x)}\,\mathbb{E}e^{-\alpha X_{\tau(x)}} = e^{-\alpha x}\,\mathbb{E}e^{-\varphi(\alpha)\tau(x)}.$$

A useful lemma

Consider $X \in \mathscr{S}_+$, and let

$$\tau(x) := \inf\{t \geq 0 : X_t \leq -x\}.$$

Observe that $e^{-\varphi(\alpha)t}\,\mathbb{E}e^{-\alpha X_t}$ is a mean-1 martingale.

- ⋆ Only positive jumps: $X_{\tau(x)} = -x$;
- ⋆ assuming that $\mathbb{E}X_1 < 0$, it holds that $\tau(x) < \infty$ almost surely.

Then 'optional sampling' implies

$$1 = \mathbb{E}e^{-\varphi(\alpha)\tau(x)}\,\mathbb{E}e^{-\alpha X_{\tau(x)}} = e^{-\alpha x}\,\mathbb{E}e^{-\varphi(\alpha)\tau(x)}.$$

Hence:

Lemma: Let $X \in \mathscr{S}_+$, and $\mathbb{E}X_1 < 0$. Then

$$\mathbb{E}e^{-\vartheta\tau(x)} = e^{-\psi(\vartheta)x}.$$

PART II:

WHAT ARE LÉVY-DRIVEN QUEUES?
STATIONARY BEHAVIOR

Queues in continuous time

Lévy-driven queue: continuous-time counterpart of the classical discrete-time queue.

In discrete time, a queue can be described through the well-known Lindley recursion: with $Q_0 = x$,

$$Q_{n+1} = \max\{Q_n + Y_n, 0\}.$$

Iterating: $Q_{n+1} = \max\{Q_{n-1} + Y_{n-1} + Y_n, Y_n, 0\}$.

With $X_n := \sum_{i=0}^{n} Y_i$, this eventually leads to

$$Q_n = X_n + \max\left\{x, \max_{0\leq i\leq n} -X_i\right\}.$$

Queues in continuous time

Discrete time:

$$Q_n = X_n + \max\left\{x, \max_{0\le i\le n} -X_i\right\}.$$

Take continuous-time counterpart:

$$Q_t = X_t + \max\{x, L_t\},\ t \ge 0,$$

with

$$L_t := \sup_{0\le s\le t} -X_u = -\inf_{0\le s\le t} X_u;$$

this increasing process L_t is often referred to as *local time*.

Queues in continuous time

Assuming queue started at $-\infty$, one can alternatively write

$$Q_t = \sup_{s \le t}(X_t - X_s);$$

assume $\mathbb{E}X_1 < 0$ to ensure stability.

Queues in continuous time

Assuming queue started at $-\infty$, one can alternatively write

$$Q_t = \sup_{s \le t}(X_t - X_s);$$

assume $\mathbb{E}X_1 < 0$ to ensure stability.

As the input process X_t is reversible:

$$Q \stackrel{\mathrm{d}}{=} \sup_{t \ge 0} X_t$$

(Reich).

Queues in continuous time

Remarkably, steady-state distribution of reflected process is distributed as supremum of free process!

Hence: close relation between queueing probabilities and ruin probabilities!

Queues in continuous time

Alternative: solution of a so-called *Skorokhod problem*; then $(Q_t)_t$ is reflection of $(X_t)_t$ at 0.

Queues in continuous time

Alternative: solution of a so-called *Skorokhod problem*; then $(Q_t)_t$ is reflection of $(X_t)_t$ at 0.

Let $(L_t^\star)_t$ be a nondecreasing right-continuous process such that

(A) $(Q_t)_t$, given by $Q_0 = x$ and $Q_t = X_t + L_t^\star$, is non-negative for all $t \geq 0$;

(B) $L_t^\star$ can only increase when $Q_t = 0$, that is

$$\int_0^T Q_t \mathrm{d}L_t^\star = 0, \quad \text{for all } T > 0.$$

Natural conditions for a queueing process!

Queues in continuous time

Alternative: solution of a so-called *Skorokhod problem*; then $(Q_t)_t$ is reflection of $(X_t)_t$ at 0.

Let $(L_t^\star)_t$ be a nondecreasing right-continuous process such that

(A) $(Q_t)_t$, given by $Q_0 = x$ and $Q_t = X_t + L_t^\star$, is non-negative for all $t \geq 0$;

(B) $L_t^\star$ can only increase when $Q_t = 0$, that is

$$\int_0^T Q_t \mathrm{d} L_t^\star = 0, \quad \text{for all } T > 0.$$

Natural conditions for a queueing process!

Then it can be proven that the only process satisfying these conditions is $L_t^\star = \max\{x, L_t\}$, so that

$$Q_t = X_t + \max\{x, L_t\}$$

for $t \geq 0$, with L_t as before.

Stationary workload

Can stationary workload be determined?

Stationary workload

Can stationary workload be determined?

Cumbersome in general, but . . .

. . . nice expressions in spectrally one-sided case!

Stationary workload: spectrally-positive case

Stationary workload: spectrally-positive case

We consider compound Poisson input and constant depletion rate r; assume $\lambda\,\mathbb{E}B < r$.

$f_Q(\cdot)$ density of the steady-state workload.

For any $x > 0$, due to rate conservation

$$r f_Q(x) = \lambda \left(\int_{(0,x)} f_Q(y)\mathbb{P}(B > x - y)\mathrm{d}y + p_0\mathbb{P}(B > x) \right),$$

with $p_0 := \mathbb{P}(Q = 0)$

Stationary workload: spectrally-positive case

Now

$$r f_Q(x) = \lambda \left(\int_{(0,x)} f_Q(y)\mathbb{P}(B > x - y)\mathrm{d}y + p_0 \mathbb{P}(B > x) \right)$$

implies

$$\begin{aligned} \bar{\kappa}(\alpha) &:= \int_{(0,\infty)} e^{-\alpha x} f_Q(x)\mathrm{d}x \\ &= \frac{1}{r} \int_{(0,\infty)} e^{-\alpha x} \lambda \left(\int_{(0,x)} f_Q(y)\mathbb{P}(B > x - y)\mathrm{d}y + p_0 \mathbb{P}(B > x) \right) \mathrm{d}x, \end{aligned}$$

which after elementary calculus reduces to

$$r\bar{\kappa}(\alpha) = \lambda \left(\bar{\kappa}(\alpha) + p_0 \right) \frac{1 - b(\alpha)}{\alpha}.$$

Realize that $\kappa(\alpha) := \mathbb{E}e^{-\alpha Q} = p_0 + \bar{\kappa}(\alpha)$ and $\kappa(\alpha) \to 1$ as $\alpha \downarrow 0$.

Stationary workload: spectrally-positive case

Theorem: [Pollaczek-Khintchine] Let $X \in \mathbb{CP}(r, \lambda, b(\cdot))$. For $\alpha \geq 0$,

$$\mathbb{E}e^{-\alpha Q} = \frac{r\alpha p_0}{r\alpha - \lambda(1 - b(\alpha))} = \frac{\alpha(r - \lambda\,\mathbb{E}B)}{r\alpha - \lambda(1 - b(\alpha))}.$$

Stationary workload: spectrally-positive case

Let $B_1^{\rm res}, B_2^{\rm res}, \ldots$ be i.i.d. samples from the residual lifetime distribution of B, that is

$$\mathbb{P}(B^{\rm res} \leq x) = \frac{1}{\mathbb{E}B} \int_0^x \mathbb{P}(B > y)\mathrm{d}y.$$

Realizing that $b^{\rm res}(\alpha) := \mathbb{E}e^{-\alpha B^{\rm res}} = (1 - b(\alpha))/(\alpha\,\mathbb{E}B)$, Pollaczeck-Khinchine can alternatively be written as

$$\mathbb{E}e^{-\alpha Q} = \left(1 - \frac{\lambda\,\mathbb{E}B}{r}\right) \sum_{n=0}^{\infty} \left(\frac{\lambda\,\mathbb{E}B}{r}\right)^n (b^{\rm res}(\alpha))^n.$$

As a consequence, with $\varrho := \lambda\,\mathbb{E}B/r$,

$$\mathbb{P}(Q \leq x) = \mathbb{P}\left(\sum_{n=1}^{N} B_n^{\rm res} \leq x\right),$$

where $\mathbb{P}(N = n) = (1 - \varrho)\varrho^n$.

Steady-state workload Q can be interpreted as a geometric number of residuals of the job size B.

Stationary workload: spectrally-positive case

Now we have solved the compound Poisson case;

idea: approximate spectrally positive by compound Poisson!

Stationary workload: spectrally-positive case

Now we have solved the compound Poisson case;

idea: approximate spectrally positive by compound Poisson!

For $X \in \mathscr{S}_+$, there are d, $\sigma^2 \geq 0$, and measure $\Pi_\varphi(\cdot)$ such that $\int_{(0,\infty)} \min\{1, x^2\}\Pi_\varphi(\mathrm{d}x) < \infty$, that the Laplace exponent reads,

$$\varphi(\alpha) = \alpha d + \frac{1}{2}\alpha^2\sigma^2 + \int_{(0,\infty)} (e^{-\alpha x} - 1 + \alpha x\, 1_{(0,1)}(x))\Pi_\varphi(\mathrm{d}x).$$

Now define, with $\varepsilon_n \to 0$,

$$\varphi_n(\alpha) := \left(d + \int_{\varepsilon_n}^1 x\Pi_\varphi(\mathrm{d}x) + \frac{\sigma^2}{\varepsilon_n}\right)\alpha + \frac{\sigma^2}{\varepsilon_n^2}\left(e^{-\alpha\varepsilon_n} - 1\right) + \int_{\varepsilon_n}^\infty (e^{-\alpha x} - 1)\Pi_\varphi(\mathrm{d}x).$$

Stationary workload: spectrally-positive case

$$\varphi(\alpha) = \alpha d + \frac{1}{2}\alpha^2\sigma^2 + \int_{(0,\infty)} (e^{-\alpha x} - 1 + \alpha x\, 1_{(0,1)}(x))\Pi_\varphi(\mathrm{d}x)$$

and

$$\varphi_n(\alpha) := \left(d + \int_{\varepsilon_n}^1 x\Pi_\varphi(\mathrm{d}x) + \frac{\sigma^2}{\varepsilon_n}\right)\alpha + \frac{\sigma^2}{\varepsilon_n^2}\left(e^{-\alpha\varepsilon_n} - 1\right) + \int_{\varepsilon_n}^\infty (e^{-\alpha x} - 1)\Pi_\varphi(\mathrm{d}x).$$

We have: $\varphi_n(s) \to \varphi(s)$ as $n \to \infty$, whereas, for all $n \in \mathbb{N}$, $\varphi_n'(0) = \varphi'(0)$.

Stationary workload: spectrally-positive case

Important: $\varphi_n(\alpha)$ is *the Laplace exponent of a compound Poisson process*!

Stationary workload: spectrally-positive case

Important: $\varphi_n(\alpha)$ is *the Laplace exponent of a compound Poisson process*!

- ⋆ The drift term of this compound Poisson process is

$$d_n := d + \int_{\varepsilon_n}^{1} x\Pi_\varphi(\mathrm{d}x) + \frac{\sigma^2}{\varepsilon_n} > 0.$$

Stationary workload: spectrally-positive case

Important: $\varphi_n(\alpha)$ is *the Laplace exponent of a compound Poisson process*!

- ⋆ The drift term of this compound Poisson process is

$$d_n := d + \int_{\varepsilon_n}^{1} x\Pi_\varphi(\mathrm{d}x) + \frac{\sigma^2}{\varepsilon_n} > 0.$$

- ⋆ Then, the term $\sigma^2/\varepsilon_n^2 \cdot (e^{-\alpha\varepsilon_n} - 1)$ can be interpreted as the contribution of a Poisson stream (arrival rate $\lambda_{1,n} := \sigma^2/\varepsilon_n^2$) of jobs of deterministic size $\beta_{1,n} := \varepsilon_n$.

Stationary workload: spectrally-positive case

Important: $\varphi_n(\alpha)$ is *the Laplace exponent of a compound Poisson process*!

- ⋆ The drift term of this compound Poisson process is

$$d_n := d + \int_{\varepsilon_n}^{1} x\Pi_\varphi(\mathrm{d}x) + \frac{\sigma^2}{\varepsilon_n} > 0.$$

- ⋆ Then, the term $\sigma^2/\varepsilon_n^2 \cdot (e^{-\alpha\varepsilon_n} - 1)$ can be interpreted as the contribution of a Poisson stream (arrival rate $\lambda_{1,n} := \sigma^2/\varepsilon_n^2$) of jobs of deterministic size $\beta_{1,n} := \varepsilon_n$.

- ⋆ Finally,

$$\int_{\varepsilon_n}^{\infty}(e^{-\alpha x} - 1)\Pi_\varphi(\mathrm{d}x) = \Pi_\varphi([\varepsilon_n, \infty)) \int_{\varepsilon_n}^{\infty}(e^{-\alpha x} - 1)\frac{\Pi_\varphi(\mathrm{d}x)}{\Pi_\varphi([\varepsilon_n, \infty))},$$

  which is the contribution of a Poisson stream (arrival rate $\lambda_{2,n} := \Pi_\varphi([\varepsilon_n, \infty))$) of jobs, whose sizes are i.i.d. samples from a 'truncated distribution' with density $\Pi_\varphi(\mathrm{d}x)/\Pi_\varphi([\varepsilon_n, \infty))$, for $x \geq \varepsilon_n$, and mean

$$\beta_{2,n} := \int_{\varepsilon_n}^{\infty} x\frac{\Pi_\varphi(\mathrm{d}x)}{\Pi_\varphi([\varepsilon_n, \infty))}.$$

Stationary workload: spectrally-positive case

Q_n: steady state workload if input were compound Poisson process with Laplace exponent $\varphi_n(\alpha)$.

Due to $\varphi_n(\alpha) \to \varphi(\alpha)$ it is conceivable that $\mathbb{E}e^{-\alpha Q_n} \to \mathbb{E}e^{-\alpha Q}$. From Pollaczek-Khinchine:

$$\begin{aligned}\mathbb{E}e^{-\alpha Q_n} &= \alpha(d_n - \lambda_{1,n}\beta_{1,n} - \lambda_{2,n}\beta_{2,n}) \Big/ \left(d_n\alpha - \frac{\sigma^2}{\varepsilon_n^2}\left(1 - e^{-\alpha\varepsilon_n}\right) - \int_{\varepsilon_n}^{\infty} (1 - e^{-\alpha x})\Pi_\varphi(\mathrm{d}x) \right) \\ &\to \frac{\alpha\varphi'(0)}{\varphi(\alpha)} \text{ as } n \to \infty;\end{aligned}$$

the convergence follows from straightforward algebra.

Stationary workload: spectrally-positive case

Q_n: steady state workload if input were compound Poisson process with Laplace exponent $\varphi_n(\alpha)$.

Due to $\varphi_n(\alpha) \to \varphi(\alpha)$ it is conceivable that $\mathbb{E}e^{-\alpha Q_n} \to \mathbb{E}e^{-\alpha Q}$. From Pollaczek-Khinchine:

$$\begin{aligned}\mathbb{E}e^{-\alpha Q_n} &= \alpha(d_n - \lambda_{1,n}\beta_{1,n} - \lambda_{2,n}\beta_{2,n}) \Big/ \left(d_n\alpha - \frac{\sigma^2}{\varepsilon_n^2}\left(1 - e^{-\alpha\varepsilon_n}\right) - \int_{\varepsilon_n}^{\infty} (1 - e^{-\alpha x})\Pi_\varphi(\mathrm{d}x) \right) \\ &\to \frac{\alpha\varphi'(0)}{\varphi(\alpha)} \text{ as } n \to \infty;\end{aligned}$$

the convergence follows from straightforward algebra.

Hence, if we can prove that $\mathbb{E}e^{-\alpha Q_n} \to \mathbb{E}e^{-\alpha Q}$, we have established the following result (Zolotarev).

Theorem: [generalized Pollaczek-Khintchine] Let $X \in \mathscr{S}_+$. For $\alpha \geq 0$,

$$\mathbb{E}e^{-\alpha Q} = \frac{\alpha\varphi'(0)}{\varphi(\alpha)}.$$

The convergence $\mathbb{E}e^{-\alpha Q_n} \to \mathbb{E}e^{-\alpha Q}$ is a technical issue that lies beyond the scope of this survey.

Stationary workload: spectrally-positive case

Alternative proofs rely on martingale techniques, most notably the *Kella-Whitt martingale*.

With

$$L_t(x) := \max\{x, L_t - x\} = \max\left\{x, -\inf_{0\le s\le t} X_s\right\},$$

it can be shown using stochastic integration theory that, for $X \in \mathscr{S}_+$,

$$K_t := \varphi(\alpha)\int_0^t e^{-\alpha Q_s}\mathrm{d}s + e^{-\alpha x} - e^{-\alpha Q_t} - \alpha L_t(x)$$

is a martingale.

Stationary workload: spectrally-positive case

$$K_t := \varphi(\alpha) \int_0^t e^{-\alpha Q_s} \mathrm{d}s + e^{-\alpha x} - e^{-\alpha Q_t} - \alpha L_t(x)$$

is a martingale.

Assume that the queue is in stationarity at time 0; then

$$0 = \mathbb{E}K_1 = \varphi(\alpha)\mathbb{E}e^{-\alpha Q} + \mathbb{E}e^{-\alpha Q} - \mathbb{E}e^{-\alpha Q} - \alpha\mathbb{E}L_1,$$

so that

$$\mathbb{E}e^{-\alpha Q} = \frac{\alpha\,\mathbb{E}L_1}{\varphi(\alpha)}.$$

Now realizing that $\mathbb{E}e^{-\alpha Q} \to 1$ as $\alpha \downarrow 0$, we retrieve the generalized Pollaczek-Khintchine result.

Stationary workload: spectrally-positive case

Example: Suppose $X \in \mathbb{B}\mathrm{m}(\mu, \sigma^2)$ for some $\mu < 0$. Then, with $\nu := -\mu/\sigma^2 > 0$,

$$\mathbb{E}e^{-\alpha Q} = \frac{\alpha \varphi'(0)}{\varphi(\alpha)} = \frac{\nu}{\nu + \alpha}.$$

Stationary workload: spectrally-positive case

Example: Suppose $X \in \mathbb{B}\mathrm{m}(\mu, \sigma^2)$ for some $\mu < 0$. Then, with $\nu := -\mu/\sigma^2 > 0$,

$$\mathbb{E}e^{-\alpha Q} = \frac{\alpha\varphi'(0)}{\varphi(\alpha)} = \frac{\nu}{\nu + \alpha}.$$

Hence: steady-state workload in a Brownian queue has an exponential distribution with mean $1/\nu$.

Stationary workload: spectrally-positive case

Possible to obtain all moments of the steady-state queue Q!

$$\mu := \mathbb{E}Q = -\frac{\mathrm{d}}{\mathrm{d}\alpha}\,\frac{\alpha\varphi'(0)}{\varphi(\alpha)}\bigg|_{\alpha\downarrow 0} = \frac{\varphi''(0)}{2\varphi'(0)},$$

and similarly

$$v := \mathbb{V}\mathrm{ar}Q = \frac{1}{4}\left(\frac{\varphi''(0)}{\varphi'(0)}\right)^2 - \frac{1}{3}\frac{\varphi'''(0)}{\varphi'(0)}.$$

Stationary workload: spectrally-negative case

Stationary workload: spectrally-negative case

Way easier!

Stationary workload: spectrally-negative case

Observe that $\mathbb{E}e^{\beta_0 X_t}$ is a martingale, with $\beta_0 := \Psi(0) > 0$.

Stationary workload: spectrally-negative case

Observe that $\mathbb{E}e^{\beta_0 X_t}$ is a martingale, with $\beta_0 := \Psi(0) > 0$.

'Optional sampling' gives, for any positive x,

$$\mathbb{P}(\exists t \geq 0 : X_t > x)e^{\beta_0 x} = 1$$

(use that, due to $X \in \mathscr{S}_-$, given a certain level $x > 0$ is reached, it is attained with equality).

Stationary workload: spectrally-negative case

Observe that $\mathbb{E}e^{\beta_0 X_t}$ is a martingale, with $\beta_0 := \Psi(0) > 0$.

'Optional sampling' gives, for any positive x,

$$\mathbb{P}(\exists t \geq 0 : X_t > x)e^{\beta_0 x} = 1$$

(use that, due to $X \in \mathscr{S}_-$, given a certain level $x > 0$ is reached, it is attained with equality).

As Q is distributed as the supremum over $t \geq 0$ of X_t ('Reich's identity'), we obtain:

Theorem: Let $X \in \mathscr{S}_-$. Then Q is exponentially distributed with mean $1/\beta_0$.

PART II:

TRANSIENT BEHAVIOR

Transient workload

We consider four metrics:

- ⋆ transient distribution;
- ⋆ busy period;
- ⋆ correlation function
- ⋆ infimum over given time interval.

Transient workload distribution: spectrally-positive case

We start with $X \in \mathscr{S}_+$.

Kella-Whitt martingale:

$$K_t := \varphi(\alpha)\int_0^t e^{-\alpha Q_s}\mathrm{d}s + e^{-\alpha x} - e^{-\alpha Q_t} - \alpha L_t(x).$$

T: exponentially distributed with mean $1/\vartheta$.

Then:

$$0 = \mathbb{E}K_T = \varphi(\alpha)\int_0^\infty \int_0^t \vartheta e^{-\vartheta t} e^{-\alpha Q_s}\mathrm{d}s\mathrm{d}t - e^{-\alpha x} - \mathbb{E}_x e^{-\alpha Q_T} - \alpha \mathbb{E}L_T(x).$$

Transient workload distribution: spectrally-positive case

$$0 = \mathbb{E}K_T = \varphi(\alpha)\int_0^\infty\int_0^t \vartheta e^{-\vartheta t} e^{-\alpha Q_s}\mathrm{d}s\mathrm{d}t - e^{-\alpha x} - \mathbb{E}_x e^{-\alpha Q_T} - \alpha \mathbb{E}L_T(x).$$

The first term reads:

$$\varphi(\alpha)\int_0^\infty\int_s^\infty \vartheta e^{-\vartheta t} e^{-\alpha Q_s}\mathrm{d}t\mathrm{d}s = \frac{\varphi(\alpha)}{\vartheta}\,\mathbb{E}_x e^{-\alpha Q_T}.$$

Now $\mathbb{E}_x e^{-\alpha Q_T}$ can be solved, and we obtain an expression in which unknown term $\mathbb{E}L_T(x)$ appears in numerator, and in which denominator equals $\vartheta - \varphi(\alpha)$.

Then: root of denominator (i.e., $\alpha = \psi(\vartheta)$) should be a root of the numerator as well (otherwise the transform equals ∞). This yields $\mathbb{E}L_T(x)$.

Transient workload distribution: spectrally-positive case

Eventually, we obtain:

Theorem: Let $X \in \mathscr{S}_+$, and let T be exponentially distributed with mean $1/\vartheta$, independently of X. Then

$$\mathbb{E}_x e^{-\alpha Q_T} = \vartheta \int_0^\infty e^{-\vartheta t} \mathbb{E}_x e^{-\alpha Q_t} = \frac{\vartheta}{\vartheta - \varphi(\alpha)} \left(e^{-\alpha x} - \frac{\alpha}{\psi(\vartheta)} e^{-\psi(\vartheta) x} \right).$$

Transient workload distribution: spectrally-positive case

Eventually, we obtain:

Theorem: Let $X \in \mathscr{S}_+$, and let T be exponentially distributed with mean $1/\vartheta$, independently of X. Then

$$\mathbb{E}_x e^{-\alpha Q_T} = \vartheta \int_0^\infty e^{-\vartheta t} \mathbb{E}_x e^{-\alpha Q_t} = \frac{\vartheta}{\vartheta - \varphi(\alpha)} \left(e^{-\alpha x} - \frac{\alpha}{\psi(\vartheta)} e^{-\psi(\vartheta) x} \right).$$

Hence: we have uniquely characterized the distribution of the workload after an exponential time, for an arbitrary starting level.

Alternative technique: level-crossing.

Transient workload distribution: spectrally-positive case

This result implies 'generalized Pollaczek-Khintchine' in at least two ways:

(a) let $\vartheta \downarrow 0$, so that T corresponds with some epoch infinitely far away, and use elementary calculus (L'Hôpital);

(b) find $\mathbb{E}e^{-\alpha Q_T}$ by deconditioning, use that in stationarity $\mathbb{E}e^{-\alpha Q_T}$ should coincide with $\mathbb{E}e^{-\alpha Q_0}$, and then solve $\mathbb{E}e^{-\alpha Q_0}$.

Transient workload distribution: spectrally-positive case

The special case of $X \in \mathbb{B}\mathrm{m}(\mu, \sigma^2)$ can be solved explicitly.

It turns out that

$$\mathbb{P}(Q_t \le y \,|\, Q_0 = x) = 1 - \Phi_{\mathrm{N}}\left(\frac{-y + x + \mu t}{\sigma\sqrt{t}}\right) - e^{2\mu y/\sigma^2}\Phi_{\mathrm{N}}\left(\frac{-y - x - \mu t}{\sigma\sqrt{t}}\right),$$

with $\Phi_{\mathrm{N}}(\cdot)$ denoting the distribution function of a standard Normal random variable.

(Sending $t \to \infty$ gives the exponential distribution.)

Transient workload distribution: spectrally-negative case

Now $X \in \mathscr{S}_{-}$.

Transient workload distribution: spectrally-negative case

Now $X \in \mathscr{S}_-$.

q-scale functions:

Let $W^{(q)}(x)$ be a strictly increasing and continuous function whose Laplace transform satisfies

$$\int_0^\infty e^{-\beta x} W^{(q)}(x) \mathrm{d}x = \frac{1}{\Phi(\beta) - q}, \qquad \beta > \Psi(q). \tag{1}$$

In addition,

$$Z^{(q)}(x) := 1 + q \int_0^x W^{(q)}(y) \mathrm{d}y. \tag{2}$$

Transient workload distribution: spectrally-negative case

Again: goal is to characterize distribution after an exponential time.

Transient workload distribution: spectrally-negative case

Again: goal is to characterize distribution after an exponential time.

Pistorius (2004): with mild abuse of notation, the transform (with respect to t) of the density of Q_t, given that $Q_0 = x$:

$$\int_0^\infty e^{-qt}\mathbb{P}_x(Q_t = y)\mathrm{d}t = e^{-\Psi(q)y}\frac{\Psi(q)}{q}Z^{(q)}(x) - W^{(q)}(x-y).$$

Transient workload distribution: spectrally-negative case

Again: goal is to characterize distribution after an exponential time.

Pistorius (2004): with mild abuse of notation, the transform (with respect to t) of the density of Q_t, given that $Q_0 = x$:

$$\int_0^\infty e^{-qt}\mathbb{P}_x(Q_t = y)\mathrm{d}t = e^{-\Psi(q)y}\frac{\Psi(q)}{q}Z^{(q)}(x) - W^{(q)}(x-y).$$

Straightforward calculus: this leads to, with T denoting an exponential rv with mean q^{-1},

$$\int_0^\infty e^{-\beta x}\mathbb{E}_x e^{-\alpha Q_T}\mathrm{d}x = I_1 - I_2;$$

where

$$\begin{aligned} I_1 &:= \int_0^\infty \int_0^\infty q e^{-\beta x} e^{-\alpha y} e^{-\Psi(q)y}\frac{\Psi(q)}{q}Z^{(q)}(x)\mathrm{d}x\mathrm{d}y, \\ I_2 &:= \int_0^\infty \int_0^\infty q e^{-\beta x} e^{-\alpha y} W^{(q)}(x-y)\mathrm{d}x\mathrm{d}y. \end{aligned}$$

Transient workload distribution: spectrally-negative case

We now compute $I_1 \equiv I_1(\alpha, \beta, q)$ and $I_2 \equiv I_2(\alpha, \beta, q)$ explicitly.

Transient workload distribution: spectrally-negative case

We now compute $I_1 \equiv I_1(\alpha, \beta, q)$ and $I_2 \equiv I_2(\alpha, \beta, q)$ explicitly.

Using the definitions of the q-scale functions:

$$\begin{aligned} I_1(\alpha, \beta, q) &= \frac{\Psi(q)}{\Psi(q)+\alpha} \int_0^\infty e^{-\beta x} Z^{(q)}(x) \mathrm{d}x \\ &= \frac{\Psi(q)}{\Psi(q)+\alpha} \left(\frac{1}{\beta} + \int_0^\infty \int_y^\infty q W^{(q)}(y) e^{-\beta x} \mathrm{d}x \mathrm{d}y \right) = \frac{\Psi(q)}{\Psi(q)+\alpha} \frac{1}{\beta} \left(1 + \frac{q}{\Phi(\beta) - q} \right). \end{aligned}$$

Likewise,

$$I_2(\alpha, \beta, q) = \int_0^\infty q e^{-(\alpha+\beta)y} \frac{1}{\Phi(\beta) - q} \mathrm{d}y = \frac{q}{\alpha+\beta} \frac{1}{\Phi(\beta) - q}.$$

Transient workload distribution: spectrally-negative case

Theorem: Let $X \in \mathscr{S}_-$, and let T be exponentially distributed with mean $1/q$, independently of X. Then

$$\int_0^\infty e^{-\beta x}\mathbb{E}_x e^{-\alpha Q_T}\mathrm{d}x = \frac{1}{\beta}\left(\frac{\Psi(q)}{\Psi(q)+\alpha} + \frac{q}{\Phi(\beta)-q}\frac{\Psi(q)-\beta}{\Psi(q)+\alpha}\frac{\alpha}{\alpha+\beta}\right).$$

Transient workload distribution: spectrally-negative case

Theorem: Let $X \in \mathscr{S}_-$, and let T be exponentially distributed with mean $1/q$, independently of X. Then

$$\int_0^\infty e^{-\beta x}\mathbb{E}_x e^{-\alpha Q_T}\mathrm{d}x = \frac{1}{\beta}\left(\frac{\Psi(q)}{\Psi(q)+\alpha} + \frac{q}{\Phi(\beta)-q}\frac{\Psi(q)-\beta}{\Psi(q)+\alpha}\frac{\alpha}{\alpha+\beta}\right).$$

Very implicit result: double transform.

Busy period

Second transient characteristic: busy period.

How long does it take before the queue idles?

Busy period

Second transient characteristic: busy period.

How long does it take before the queue idles?

More precisely: define

$$\tau := \inf\{t \geq 0 : Q_t = 0\},$$

where Q_0 is distributed according to the stationary distribution.

Busy period

Second transient characteristic: busy period.

How long does it take before the queue idles?

More precisely: define

$$\tau := \inf\{t \geq 0 : Q_t = 0\},$$

where Q_0 is distributed according to the stationary distribution.

We write: $p(t) := \mathbb{P}(\tau > t)$; we derive the Laplace transform of $p(\cdot)$.

Busy period: spectrally-positive case

Recall:

Lemma: Let $X \in \mathscr{S}_+$, and $\mathbb{E}X_1 < 0$. Then

$$\mathbb{E}e^{-\vartheta\tau(x)} = e^{-\psi(\vartheta)x}.$$

Busy period: spectrally-positive case

Recall:

Lemma: Let $X \in \mathscr{S}_+$, and $\mathbb{E}X_1 < 0$. Then

$$\mathbb{E}e^{-\vartheta\tau(x)} = e^{-\psi(\vartheta)x}.$$

Hence, using integration by parts:

$$\int_0^\infty e^{-\vartheta t}\mathbb{P}(\tau(x) > t)\mathrm{d}t \; = \; \int_0^\infty \mathbb{P}(\tau(x) > t)\mathrm{d}\left(-\frac{1}{\vartheta}e^{-\vartheta t}\right)$$

Busy period: spectrally-positive case

Recall:

Lemma: Let $X \in \mathscr{S}_+$, and $\mathbb{E}X_1 < 0$. Then

$$\mathbb{E}e^{-\vartheta\tau(x)} = e^{-\psi(\vartheta)x}.$$

Hence, using integration by parts:

$$\begin{aligned}
\int_0^\infty e^{-\vartheta t}\mathbb{P}(\tau(x) > t)\mathrm{d}t &= \int_0^\infty \mathbb{P}(\tau(x) > t)\mathrm{d}\left(-\frac{1}{\vartheta}e^{-\vartheta t}\right) \\
&= \left[-\mathbb{P}(\tau(x) > t)\frac{1}{\vartheta}e^{-\vartheta t}\right]_0^\infty + \frac{1}{\vartheta}\int_0^\infty e^{-\vartheta t}\mathrm{d}\mathbb{P}(\tau(x) > t) \\
&= \frac{1}{\vartheta} - \frac{1}{\vartheta}\int_0^\infty e^{-\vartheta t}\mathrm{d}\mathbb{P}(\tau(x) \le t) = \frac{1}{\vartheta}\left(1 - e^{-\psi(\vartheta)x}\right).
\end{aligned}$$

Busy period: spectrally-positive case

We thus find:

$$\int_0^\infty e^{-\vartheta t} p(t)\mathrm{d}t = \int_0^\infty \left(\int_0^\infty e^{-\vartheta t}\mathbb{P}(\tau(x) > t)\mathrm{d}t \right) \mathrm{d}\mathbb{P}(Q_0 < x)$$
$$= \frac{1}{\vartheta} \int_0^\infty \left(1 - e^{-\psi(\vartheta)x} \right) \mathrm{d}\mathbb{P}(Q_0 < x).$$

Busy period: spectrally-positive case

We have

$$\int_0^\infty e^{-\vartheta t} p(t) \mathrm{d}t = \frac{1}{\vartheta} \int_0^\infty \left(1 - e^{-\psi(\vartheta)x}\right) \mathrm{d}\mathbb{P}(Q_0 < x),$$

but the latter expression equals:

$$\frac{1}{\vartheta}\left(1 - \mathbb{E}e^{-\psi(\vartheta)Q_0}\right),$$

which we can evaluate with 'generalized Pollaczek-Khinchine'.

Busy period: spectrally-positive case

We have

$$\int_0^\infty e^{-\vartheta t} p(t) \mathrm{d}t = \frac{1}{\vartheta} \int_0^\infty \left(1 - e^{-\psi(\vartheta)x}\right) \mathrm{d}\mathbb{P}(Q_0 < x),$$

but the latter expression equals:

$$\frac{1}{\vartheta}\left(1 - \mathbb{E}e^{-\psi(\vartheta)Q_0}\right),$$

which we can evaluate with 'generalized Pollaczek-Khinchine'.

'Generalized Pollaczek-Khinchine':

$$\mathbb{E}e^{-\alpha Q_0} = \frac{\alpha\varphi'(0)}{\varphi(\alpha)}.$$

Busy period: spectrally-positive case

Conclude:

Proposition: Let $X \in \mathscr{S}_+$. Then

$$\int_0^\infty e^{-\vartheta t} p(t) \mathrm{d}t = \frac{1}{\vartheta} - \varphi'(0) \frac{\psi(\vartheta)}{\vartheta^2}.$$

Busy period: spectrally-positive case

Special case: $X \in \mathbb{C}\mathrm{P}(r, \lambda, b(\cdot))$.

Then the notion of a busy period *starting in 0 is well-defined*; denote this by τ^0.

Let $\pi(\vartheta) := \mathbb{E}e^{-\vartheta\tau^0}$.

Busy period: spectrally-positive case

Let $\pi(\vartheta) := \mathbb{E}e^{-\vartheta\tau^0}$.

Known (Takács): $\pi(\vartheta) = \beta(\vartheta + \lambda - \lambda\pi(\vartheta))$, after renormalizing time such that $r = 1$.

Busy period: spectrally-positive case

Let $\pi(\vartheta) := \mathbb{E}e^{-\vartheta\tau^0}$.

Known (Takács): $\pi(\vartheta) = \beta(\vartheta + \lambda - \lambda\pi(\vartheta))$, after renormalizing time such that $r = 1$.

Recall: $\varphi(\alpha) = \alpha - \lambda + \lambda\beta(\alpha)$.

Therefore

$$0 = \beta(\vartheta + \lambda - \lambda\pi(\vartheta)) - \pi(\vartheta) = \frac{1}{\lambda}\varphi(\vartheta + \lambda - \lambda\pi(\vartheta)) - \frac{\vartheta}{\lambda},$$

and hence $\varphi(\vartheta + \lambda - \lambda\pi(\vartheta)) = \vartheta$. Apply $\psi(\cdot)$ to both sides, and we obtain the following result.

Busy period: spectrally-positive case

Let $\pi(\vartheta) := \mathbb{E}e^{-\vartheta\tau^0}$.

Known (Takács): $\pi(\vartheta) = \beta(\vartheta + \lambda - \lambda\pi(\vartheta))$, after renormalizing time such that $r = 1$.

Recall: $\varphi(\alpha) = \alpha - \lambda + \lambda\beta(\alpha)$.

Therefore

$$0 = \beta(\vartheta + \lambda - \lambda\pi(\vartheta)) - \pi(\vartheta) = \frac{1}{\lambda}\varphi(\vartheta + \lambda - \lambda\pi(\vartheta)) - \frac{\vartheta}{\lambda},$$

and hence $\varphi(\vartheta + \lambda - \lambda\pi(\vartheta)) = \vartheta$. Apply $\psi(\cdot)$ to both sides, and we obtain the following result.

Proposition: Let $X \in \mathbb{C}\mathrm{P}(1, \lambda, b(\cdot))$. Then

$$\pi(\vartheta) = \frac{\lambda + \vartheta}{\lambda} - \frac{1}{\lambda}\psi(\vartheta).$$

Busy period: spectrally-negative case

Without proof we state:

Proposition: Let $X \in \mathscr{S}_-$. Then

$$\int_0^\infty e^{-qt} p(t) \mathrm{d}t = \frac{1}{q}\left(1 - \frac{\Psi(0)}{\Psi(q)}\right).$$

Correlation function

We now examine the Laplace transform $\hat{r}(\cdot)$ corresponding to the correlation of the workload process.

Assume system is in steady-state at time 0.

Then

$$r(t) := \mathbb{C}\mathrm{orr}(Q_0, Q_t) = \frac{\mathbb{C}\mathrm{ov}(Q_0, Q_t)}{\sqrt{\mathbb{V}\mathrm{ar}Q_0 \cdot \mathbb{V}\mathrm{ar}Q_t}} = \frac{\mathbb{E}(Q_0 Q_t) - (\mathbb{E}Q_0)^2}{\mathbb{V}\mathrm{ar}Q_0}.$$

Correlation function: spectrally-positive case

Let T be exponentially distributed with mean $1/\vartheta$.

Realize that

$$\mathbb{E}(e^{-\alpha Q_T} \mid Q_0 = q) = \int_0^\infty \vartheta e^{-\vartheta t} \mathbb{E}(e^{-\alpha Q_t} \mid Q_0 = q) \mathrm{d}t.$$

Correlation function: spectrally-positive case

Let T be exponentially distributed with mean $1/\vartheta$.

Realize that

$$\mathbb{E}(e^{-\alpha Q_T} \mid Q_0 = q) = \int_0^\infty \vartheta e^{-\vartheta t}\mathbb{E}(e^{-\alpha Q_t} \mid Q_0 = q)\mathrm{d}t.$$

We know that

$$\mathbb{E}_x e^{-\alpha Q_T} = \vartheta \int_0^\infty e^{-\vartheta t}\mathbb{E}_x e^{-\alpha Q_t} = \frac{\vartheta}{\vartheta - \varphi(\alpha)}\left(e^{-\alpha x} - \frac{\alpha}{\psi(\vartheta)}e^{-\psi(\vartheta)x}\right).$$

Correlation function: spectrally-positive case

Let T be exponentially distributed with mean $1/\vartheta$.

Realize that

$$\mathbb{E}(e^{-\alpha Q_T} \mid Q_0 = q) = \int_0^\infty \vartheta e^{-\vartheta t} \mathbb{E}(e^{-\alpha Q_t} \mid Q_0 = q) \mathrm{d}t.$$

We know that

$$\mathbb{E}_x e^{-\alpha Q_T} = \vartheta \int_0^\infty e^{-\vartheta t} \mathbb{E}_x e^{-\alpha Q_t} = \frac{\vartheta}{\vartheta - \varphi(\alpha)} \left(e^{-\alpha x} - \frac{\alpha}{\psi(\vartheta)} e^{-\psi(\vartheta) x} \right).$$

By differentiation with respect to α and subsequently letting $\alpha \downarrow 0$, we obtain

$$\int_0^\infty \vartheta e^{-\vartheta t} \mathbb{E}(Q_t \mid Q_0 = q) \mathrm{d}t = -\frac{\varphi'(0)}{\vartheta} + q + \frac{e^{-\psi(\vartheta) q}}{\psi(\vartheta)}. \tag{3}$$

Correlation function: spectrally-positive case

Concentrate on the Laplace transform $\gamma(\vartheta)$ of $\mathbb{C}\mathrm{ov}(Q_0, Q_t)$.

Straightforward calculus reveals that

$$\gamma(\vartheta) := \int_0^\infty \mathbb{C}\mathrm{ov}(Q_0, Q_t) e^{-\vartheta t} \mathrm{d}t = \int_0^\infty (\mathbb{E}(Q_0 Q_t) - \mu^2) e^{-\vartheta t} \mathrm{d}t$$
$$= \int_0^\infty \int_0^\infty q \cdot \mathbb{E}(Q_t \mid Q_0 = q) \cdot e^{-\vartheta t} \mathrm{d}\mathbb{P}(Q_0 \leq q) \mathrm{d}t - \frac{\mu^2}{\vartheta};$$

(use queue is in stationarity at time 0, and hence also at t).

Recall μ and v are mean and variance of Q.

Correlation function: spectrally-positive case

By invoking (3) we find that this equals

$$\int_0^\infty \frac{q}{\vartheta}\left(-\frac{\varphi'(0)}{\vartheta}+q+\frac{e^{-\psi(\vartheta)q}}{\psi(\vartheta)}\right)\mathrm{d}\mathbb{P}(Q_0\leq q)-\frac{\mu^2}{\vartheta}$$
$$= -\frac{\mu\varphi'(0)}{\vartheta^2}+\frac{v}{\vartheta}+\frac{1}{\vartheta\psi(\vartheta)}\mathbb{E}(Q_0e^{-\psi(\vartheta)Q_0}).$$

Correlation function: spectrally-positive case

By invoking (3) we find that this equals

$$\int_0^\infty \frac{q}{\vartheta}\left(-\frac{\varphi'(0)}{\vartheta}+q+\frac{e^{-\psi(\vartheta)q}}{\psi(\vartheta)}\right)\mathrm{d}\mathbb{P}(Q_0\leq q)-\frac{\mu^2}{\vartheta}$$
$$= -\frac{\mu\varphi'(0)}{\vartheta^2}+\frac{v}{\vartheta}+\frac{1}{\vartheta\psi(\vartheta)}\mathbb{E}(Q_0e^{-\psi(\vartheta)Q_0}).$$

From 'generalized Pollaczek-Khinchine' we obtain by differentiating

$$\mathbb{E}(Q_0e^{-\alpha Q_0})=\varphi'(0)\left(-\frac{1}{\varphi(\alpha)}+\alpha\frac{\varphi'(\alpha)}{(\varphi(\alpha))^2}\right).$$

Correlation function: spectrally-positive case

By invoking (3) we find that this equals

$$\int_0^\infty \frac{q}{\vartheta}\left(-\frac{\varphi'(0)}{\vartheta}+q+\frac{e^{-\psi(\vartheta)q}}{\psi(\vartheta)}\right)\mathrm{d}\mathbb{P}(Q_0\leq q)-\frac{\mu^2}{\vartheta}$$
$$= \; -\frac{\mu\varphi'(0)}{\vartheta^2}+\frac{v}{\vartheta}+\frac{1}{\vartheta\psi(\vartheta)}\mathbb{E}(Q_0e^{-\psi(\vartheta)Q_0}).$$

From 'generalized Pollaczek-Khinchine' we obtain by differentiating

$$\mathbb{E}(Q_0e^{-\alpha Q_0})=\varphi'(0)\left(-\frac{1}{\varphi(\alpha)}+\alpha\frac{\varphi'(\alpha)}{(\varphi(\alpha))^2}\right).$$

Inserting this relation, in addition to the explicit expression for μ:

$$\gamma(\vartheta):=\int_0^\infty \mathbb{C}\mathrm{ov}(Q_0,Q_t)e^{-\vartheta t}\mathrm{d}t=-\frac{\varphi''(0)}{2\vartheta^2}+\frac{v}{\vartheta}+\frac{\varphi'(0)}{\vartheta^2}\left(\frac{1}{\vartheta\psi'(\vartheta)}-\frac{1}{\psi(\vartheta)}\right).$$

Correlation function: spectrally-positive case

We finally obtain:

Theorem: Let $X \in \mathscr{S}_+$. Then, for any $\vartheta \geq 0$,

$$\hat{r}(\vartheta) := \int_0^\infty r(t)\, e^{-\vartheta t} \mathrm{d}t = \frac{\gamma(\vartheta)}{v} = \frac{1}{\vartheta} - \frac{\varphi''(0)}{2v\vartheta^2} + \frac{\varphi'(0)}{v\vartheta^2} \left[\frac{1}{\vartheta \psi'(\vartheta)} - \frac{1}{\psi(\vartheta)} \right].$$

Correlation function: spectrally-negative case

Recall (i) that we have the double transform of Q_t:

$$\int_0^\infty e^{-\beta x}\mathbb{E}_x e^{-\alpha Q_T}\mathrm{d}x = \frac{1}{\beta}\left(\frac{\Psi(q)}{\Psi(q)+\alpha} + \frac{q}{\Phi(\beta)-q}\frac{\Psi(q)-\beta}{\Psi(q)+\alpha}\frac{\alpha}{\alpha+\beta}\right),$$

and (ii) that Q_0 is exponentially distributed with mean $1/\beta_0$.

Correlation function: spectrally-negative case

Recall (i) that we have the double transform of Q_t:

$$\int_0^\infty e^{-\beta x}\mathbb{E}_x e^{-\alpha Q_T}\mathrm{d}x = \frac{1}{\beta}\left(\frac{\Psi(q)}{\Psi(q)+\alpha} + \frac{q}{\Phi(\beta)-q}\frac{\Psi(q)-\beta}{\Psi(q)+\alpha}\frac{\alpha}{\alpha+\beta}\right),$$

and (ii) that Q_0 is exponentially distributed with mean $1/\beta_0$.

This leads to:

Theorem: Let $X \in \mathscr{S}_+$. Then, for any $q \geq 0$,

$$\hat{r}(q) := \int_0^\infty r(t)\, e^{-qt}\mathrm{d}t = \frac{1}{q} + \frac{\beta_0^2}{q^2}\Phi'(\beta_0)\left(\frac{1}{\Psi(q)} - \frac{1}{\beta_0}\right).$$

Correlation function: structural properties

Proposition: Let $X \in \mathscr{S}_+$ or $X \in \mathscr{S}_-$. Then $r(\cdot)$ is positive, decreasing, and convex.

Correlation function: structural properties

Proposition: Let $X \in \mathscr{S}_+$ or $X \in \mathscr{S}_-$. Then $r(\cdot)$ is positive, decreasing, and convex.

Proof through *completely monotone functions*, to be thought of as Laplace transforms of nonnegative random variables.

We demonstrate this concept for the case $X \in \mathscr{S}_-$.

Correlation function: structural properties

$\mathscr{C}$: class of completely monotone functions.

Correlation function: structural properties

$\mathscr{C}$: class of completely monotone functions.

The concept of complete monotonicity is easy to work with, as one can use a set of practical properties.

Correlation function: structural properties

Lemma: The following properties apply:

(1) $\mathscr{C}$ is closed under addition: if $f(\alpha) \in \mathscr{C}$ and $g(\alpha) \in \mathscr{C}$, then $f(\alpha) + g(\alpha) \in \mathscr{C}$. This extends to: if $f_x(\alpha) \in \mathscr{C}$ for $x \in \Xi$, then $\int_{x\in\Xi} f_x(\alpha)\mu(\mathrm{d}x) \in \mathscr{C}$ for any measure $\mu(\cdot)$.

(2) $\mathscr{C}$ is closed under multiplication: if $f(\alpha) \in \mathscr{C}$ and $g(\alpha) \in \mathscr{C}$, then $f(\alpha)g(\alpha) \in \mathscr{C}$.

(3) Properties of composite $\mathscr{C}$ functions: if $f(\alpha) \in \mathscr{C}$ and $g(\alpha) \geq 0$ with $g'(\alpha) \in \mathscr{C}$, then $f(g(\alpha)) \in \mathscr{C}$.

(4) Let $U(\alpha)$ non-decreasing on $[0, \infty)$, and $U(0) = 0$, $u := \lim_{\alpha\to\infty} U(\alpha) < \infty$, and

$$f(\alpha) := \int_{[0,\infty)} e^{-\alpha x}\mathrm{d}U(x);$$

clearly $f(\alpha) \in \mathscr{C}$ and $u = f(0)$. Then also

$$g(\alpha) := \frac{f(0) - f(\alpha)}{\alpha} \in \mathscr{C}.$$

(5) $\mathscr{C}$ closed under differentiation: if $f(\alpha) \in \mathscr{C}$, then $-f'(\alpha) \in \mathscr{C}$.

Correlation function: structural properties

Let $X \in \mathscr{S}_-$.

Next step:

$$\Psi(0)/\Psi(q) \in \mathscr{C}.$$

Correlation function: structural properties

Let $X \in \mathscr{S}_-$.

Next step:

$$\Psi(0)/\Psi(q) \in \mathscr{C}.$$

Reason:

$$\int_0^\infty e^{-qt} p(t) \mathrm{d}t = \frac{1}{q}\left(1 - \frac{\Psi(0)}{\Psi(q)}\right).$$

implies that

$$\mathbb{E} e^{-q\tau} = \Psi(0)/\Psi(q).$$

Correlation function: structural properties

Integration by parts:

$$\rho^{(1)}(q) := \int_0^\infty e^{-qt} r'(t)\mathrm{d}t = \frac{\beta_0^2}{q}\Phi'(\beta_0)\left(\frac{1}{\Psi(q)} - \frac{1}{\beta_0}\right);$$

$$\rho^{(2)}(q) := \int_0^\infty e^{-qt} r''(t)\mathrm{d}t = -r'(0) + \beta_0^2\Phi'(\beta_0)\left(\frac{1}{\Psi(q)} - \frac{1}{\beta_0}\right).$$

Recall that $\Psi(0)/\Psi(q) \in \mathscr{C}$.

Conclude: $\rho^{(2)}(q)$ is in $\mathscr{C}$, and hence $r''(\cdot)$ is positive, i.e., $r(\cdot)$ is convex.

Known: $f(q) \in \mathscr{C}$ implies that, with $g(q) := (f(0) - f(q))/q$, also $g(q) \in \mathscr{C}$.

Taking $f(q) = \rho^{(2)}(q)$, we have $-\rho^{(1)}(q)$ is in $\mathscr{C}$, and hence $r'(\cdot)$ is negative, i.e., $r(\cdot)$ is decreasing.

Similarly, $\rho(q)$ is in $\mathscr{C}$, and hence $r(\cdot)$ is positive. $\square$

Correlation function: structural properties

Proof for $X \in \mathscr{S}_+$ is more involved.

Correlation function: structural properties

Proof for $X \in \mathscr{S}_+$ is more involved.

Crucial: $-\psi(\cdot)$ is the Laplace exponent of an *increasing* Lévy process.

Hence this Lévy process does not have a Brownian component, and it entails that $\psi'(\cdot) \in \mathscr{C}$.

Delicate manipulation with Laplace transforms: proof of the Laplace transform of $r''(\cdot)$ being is completely monotone, and therefore $r(\cdot)$ is convex.

Procedure to do statements about $r(\cdot)$ is decreasing and positive: similar to the case $X \in \mathscr{S}_-$.

Infimum over given time interval

Last transient performance metric:

$$M_t := \inf_{s \in [0,t]} Q_s.$$

Infimum over given time interval

Last transient performance metric:

$$M_t := \inf_{s \in [0,t]} Q_s.$$

Observe that $M_t > u$ corresponds to $Q_0 + \inf_{s \in [0,t]} X_s > u$.

Infimum over given time interval

Last transient performance metric:

$$M_t := \inf_{s\in[0,t]} Q_s.$$

Observe that $M_t > u$ corresponds to $Q_0 + \inf_{s\in[0,t]} X_s > u$. Hence:

$$\begin{aligned}
&\int_0^\infty e^{-\vartheta t} \int_0^\infty e^{-\alpha u} \mathbb{P}(M_t > u) \mathrm{d}u \mathrm{d}t \\
&= \int_0^\infty e^{-\vartheta t} \int_0^\infty e^{-\alpha u} \int_u^\infty \mathbb{P}\left(\inf_{s\in[0,t]} X_s > u - q \right) \mathrm{d}\mathbb{P}(Q_0 \leq q) \mathrm{d}u \mathrm{d}t \\
&= \int_0^\infty \int_0^q e^{-\alpha u} \int_0^\infty e^{-\vartheta t} \mathbb{P}(\tau(q-u) > t) \mathrm{d}t \mathrm{d}u \mathrm{d}\mathbb{P}(Q_0 \leq q).
\end{aligned}$$

Infimum over given time interval

By using integration by parts we have that

$$\int_0^\infty \int_0^q e^{-\alpha u} \int_0^\infty e^{-\vartheta t}\mathbb{P}(\tau(q-u) > t)\mathrm{d}t\mathrm{d}u\mathrm{d}\mathbb{P}(Q_0 \leq q)$$

equals

$$\int_0^\infty \int_0^q e^{-\alpha u}\frac{1}{\vartheta}\left(1 - \mathbb{E}e^{-\vartheta\tau(q-u)}\right)\mathrm{d}u\mathrm{d}\mathbb{P}(Q_0 \leq q).$$

Infimum over given time interval

Now we have to distinguish between $X \in \mathscr{S}_+$ and $\mathscr{S}_-$.

In the former case: we know $\mathbb{E}e^{-\vartheta\tau(x)}$, and we have to apply 'generalized Pollaczek-Khinchine'. We obtain the following result.

Proposition: Let $X \in \mathscr{S}_+$. Then

$$\int_0^\infty e^{-\vartheta t} \int_0^\infty e^{-\alpha u} \mathbb{P}(M_t > u) \mathrm{d}u \mathrm{d}t = \frac{1}{\vartheta}\left(\frac{1}{\alpha} - \frac{\varphi'(0)}{\varphi(\alpha)}\right) - \frac{\varphi'(0)}{(\alpha - \psi(\vartheta))\vartheta}\left(\frac{\psi(\vartheta)}{\vartheta} - \frac{\alpha}{\varphi(\alpha)}\right).$$

Infimum over given time interval

In the latter case, Q_0 has an exponential distribution with parameter β_0.

Interchanging the order of integration, and applying a certain factorization identity, we obtain the following result.

Proposition: Let $X \in \mathscr{S}_-$. Then

$$\int_0^\infty e^{-qt} \int_0^\infty e^{-\beta u} \mathbb{P}(M_t > u) \mathrm{d}u \mathrm{d}t = \frac{1}{\beta + \beta_0} \frac{\Psi(q)}{\Psi(q) + \beta_0}.$$

PART III:

ASYMPTOTICS

Asymptotics

⋆ Tail of workload distribution;

⋆ Tail of busy period distribution;

⋆ Joint transient distribution;

⋆ Rare-event simulation, importance sampling.

Workload asymptotics

Goal: characterize $\mathbb{P}(Q > u)$ for u large.

Workload asymptotics

Goal: characterize $\mathbb{P}(Q > u)$ for u large.

Depends on the *heaviness* of the 'upper tail'!

Workload asymptotics

Goal: characterize $\mathbb{P}(Q > u)$ for u large.

Depends on the *heaviness* of the 'upper tail'!

Three cases:

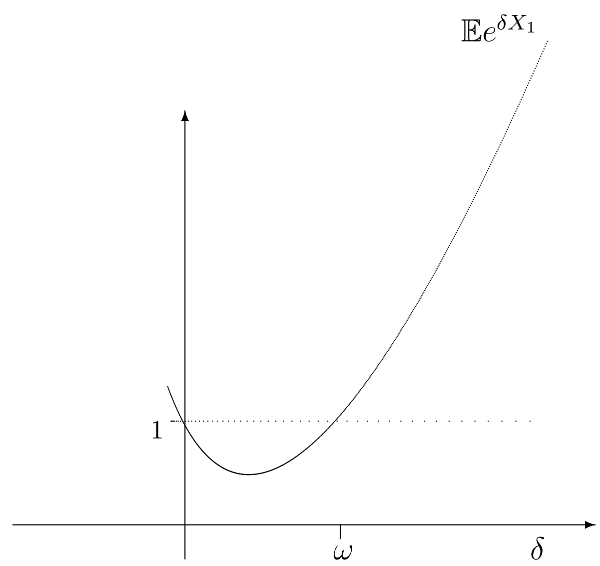
- $\star$ light-tailed regime;
- $\star$ intermediate regime;
- $\star$ heavy-tailed regime.

Workload asymptotics: light-tailed regime

$\mathscr{L}$: class of Lévy processes such that there is an $\omega > 0$ such that

$$\mathbb{E}e^{\omega X_1} = 1 \quad \text{and} \quad \mathbb{E}X_1 e^{\omega X_1} < \infty.$$

Workload asymptotics: light-tailed regime



Workload asymptotics: light-tailed regime

Examples: Brownian case, compund Poisson with light-tailed jobs, . . .

Workload asymptotics: light-tailed regime

For ease: start with $X \in \mathbb{CP}(1, \lambda, b(\cdot))$; we consider more the general case of $X \in \mathscr{L}$ later.

Workload asymptotics: light-tailed regime

For ease: start with $X \in \mathbb{CP}(1, \lambda, b(\cdot))$; we consider more the general case of $X \in \mathscr{L}$ later.

Let $\rho := \lambda \, \mathbb{E}B < 1$. Then ω solves $\mathbb{E}e^{\omega X_1} = 1$, or, equivalently, $\varphi(-\omega) = 0$.

More concretely:

$$\lambda + \omega = \lambda b(-\omega).$$

Workload asymptotics: light-tailed regime

We now introduce alternative probability measure.

Original probability measure: $\mathbb{P}$;

alternative measure $\mathbb{Q}$ characterized as $\mathbb{CP}(1, \lambda + \omega, \bar{b}(\cdot))$, where

$$\bar{b}(\alpha) := b(\alpha - \omega)/b(-\omega).$$

Workload asymptotics: light-tailed regime

We now introduce alternative probability measure.

Original probability measure: $\mathbb{P}$;

alternative measure $\mathbb{Q}$ characterized as $\mathbb{CP}(1, \lambda + \omega, \bar{b}(\cdot))$, where

$$\bar{b}(\alpha) := b(\alpha - \omega)/b(-\omega).$$

Workload asymptotics: light-tailed regime

Convexity: $\varphi'(-\omega) = 1 + \lambda b'(-\omega) < 0$.

Therefore

$$(\lambda + \omega)\,\mathbb{E}_{\mathbb{Q}} B = (\lambda + \omega) \left(-\frac{b'(-\omega)}{b(-\omega)} \right) = -\lambda b'(-\omega) =: \rho_{\mathbb{Q}} > 1,$$

so that under $\mathbb{Q}$ the queue is *unstable*.

Workload asymptotics: light-tailed regime

Under $\mathbb{Q}$ the queue is *unstable*.

Realize: $\mathbb{P}(Q > u)$ equals $\mathbb{P}(\exists t \geq 0 : X_t > u) = \mathbb{P}(\sigma(u) < \infty)$, where $\sigma(u)$ is defined as the hitting time of level u, i.e.,

$$\sigma(u) := \inf\{t : X_t \geq u\}.$$

Hence: under $\mathbb{Q}$ we have that $\sigma(u) < \infty$ almost surely, for any $u > 0$.

Workload asymptotics: light-tailed regime

Change of measure:

$$\mathbb{P}(Q > u) = \mathbb{E}_{\mathbb{Q}}\left(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} 1_{\{\sigma(u)<\infty\}}\right).$$

Using that $\sigma(u) < \infty$ almost surely under $\mathbb{Q}$, in conjunction with $\mathbb{E}e^{\omega X_t} = 1$ for all $t \geq 0$, it is a standard that

$$\mathbb{P}(Q > u) = \mathbb{E}_{\mathbb{Q}} e^{-\omega X_{\sigma(u)}}.$$

Workload asymptotics: light-tailed regime

$$\mathbb{P}(Q > u) = \mathbb{E}_{\mathbb{Q}} e^{-\omega X_{\sigma(u)}}.$$

Realize: $X_{\sigma(u)} = u + R_u$, where R_u is the *overshoot* over level u.

Workload asymptotics: light-tailed regime

$$\mathbb{P}(Q > u) = \mathbb{E}_{\mathbb{Q}} e^{-\omega X_{\sigma(u)}}.$$

Realize: $X_{\sigma(u)} = u + R_u$, where R_u is the *overshoot* over level u.

Let L_n be the n-th *ladder height*, i.e., the difference between the n-th and $(n-1)$-st record; these random variables are positive and i.i.d., and nondefective (why?).

Workload asymptotics: light-tailed regime

$$\mathbb{P}(Q > u) = \mathbb{E}_{\mathbb{Q}} e^{-\omega X_{\sigma(u)}}.$$

Realize: $X_{\sigma(u)} = u + R_u$, where R_u is the *overshoot* over level u.

Let L_n be the n-th *ladder height*, i.e., the difference between the n-th and $(n-1)$-st record; these random variables are positive and i.i.d., and nondefective (why?).

Renewal theory: R_u converges to a limiting random variable R, where

$$\mathbb{Q}(R \leq v) = \frac{1}{\mathbb{E}_{\mathbb{Q}} L} \int_0^v (1 - \mathbb{Q}(L \leq y))\mathrm{d}y,$$

with L denoting a ladder height.

Workload asymptotics: light-tailed regime

Due to the definition of $\mathbb{Q}$:

$$\mathrm{d}\mathbb{Q}(L \leq y) = e^{\omega y}\mathrm{d}\mathbb{P}(L \leq y) = e^{\omega y}\lambda\mathbb{P}(B > y)\mathrm{d}y;$$

it follows from the definition of ω that this density integrates to 1.

Workload asymptotics: light-tailed regime

Due to the definition of $\mathbb{Q}$:

$$\mathrm{d}\mathbb{Q}(L \le y) = e^{\omega y}\mathrm{d}\mathbb{P}(L \le y) = e^{\omega y}\lambda\mathbb{P}(B > y)\mathrm{d}y;$$

it follows from the definition of ω that this density integrates to 1.

Combining the above, we obtain that, as $u \to \infty$,

$$\mathbb{P}(Q > u)e^{\omega u} \to \frac{1}{\mathbb{E}_{\mathbb{Q}}L}\int_0^\infty e^{-\omega y}(1 - \mathbb{Q}(L \le y))\mathrm{d}y.$$

Straightforward calculus now yields the classical Cramér-Lundburg asymptotics.

Theorem: Let $X \in \mathbb{C}\mathrm{P}(1, \lambda, b(\cdot)) \cap \mathscr{L}$. Then, as $u \to \infty$,

$$\mathbb{P}(Q > u)e^{\omega u} \to \frac{1-\rho}{\rho_{\mathbb{Q}} - 1}.$$

Workload asymptotics: light-tailed regime

In passing, we also proved that, for all $u \geq 0$,

$$\mathbb{P}(Q > u) \leq e^{-\omega u}$$

(realize that $R_u \geq 0$).

This uniform bound applies for all $X \in \mathscr{L}$, i.e., not just for compound Poisson; the proof relies on a change-of-measure argument.

Corollary: Let $X \in \mathscr{L}$. Then $\mathbb{P}(Q > u) \leq e^{-\omega u}$.

Workload asymptotics: light-tailed regime

Now consider more general $X \in \mathscr{L}$: is it for instance possible to extend the asymptotics to $\mathscr{S}_+$?

Recall: we have Laplace transform of Q, viz. $\alpha\varphi'(0)/\varphi(\alpha)$.

Can we use this to obtain asymptotics?

Workload asymptotics: light-tailed regime

Transform $\rightarrow$ asymptotics: so-called *Heaviside principle*.

Note that

$$\int_0^\infty e^{-\alpha Q}\mathbb{P}(Q > x)\mathrm{d}x = \frac{1}{\alpha} - \frac{\varphi'(0)}{\varphi(\alpha)}.$$

Now observe that when $X \in \mathscr{L}$, $\varphi(\cdot)$ has a pole in $-\omega$, and

$$\lim_{\alpha \downarrow -\omega} \int_0^\infty e^{-\alpha Q}\mathbb{P}(Q > x)\mathrm{d}x = \frac{\varphi'(0)}{-\varphi'(-\omega)} > 0;$$

note that we assumed that the denominator of the last expression is finite (definition of $\mathscr{L}$).

Now the Heaviside principle yields that, as $u \rightarrow \infty$,

$$\mathbb{P}(Q > u)e^{\omega u} \rightarrow \frac{\varphi'(0)}{-\varphi'(-\omega)}.$$

Workload asymptotics: light-tailed regime

$$\mathbb{P}(Q > u)e^{\omega u} \to \frac{\varphi'(0)}{-\varphi'(-\omega)}.$$

Poisson case: it indeed gives Cramér-Lundberg.

Caveat: Heaviside principle, although well established in the literature and frequently used, lacks full mathematical rigor.

Workload asymptotics: light-tailed regime

The most general result is due to Bertoin and Doney:

tail asymptotics for $\mathbb{P}(Q > u)$ are derived for the full class $\mathscr{L}$.

Workload asymptotics: light-tailed regime

The most general result is due to Bertoin and Doney:

tail asymptotics for $\mathbb{P}(Q > u)$ are derived for the full class $\mathscr{L}$.

Of the form $Ce^{-\omega u}$, where ω solves $\mathbb{E}e^{\omega X_1} = 1$, but with some rather involved expression for C.

Workload asymptotics: intermediate regime

Define

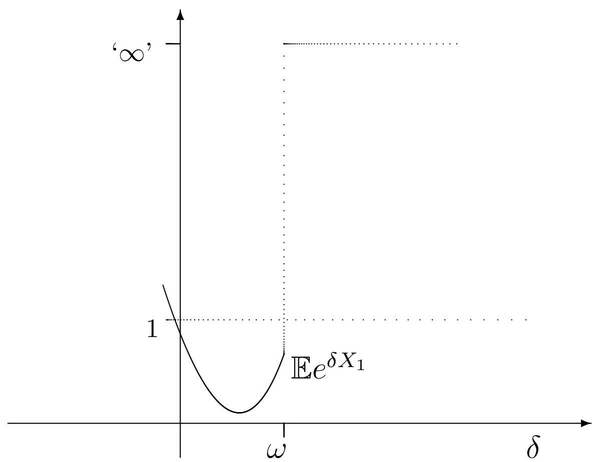
$$\omega := \sup\{\delta \geq 0 : \mathbb{E}e^{\delta X_1} < \infty\}.$$

We say that $X \in \mathscr{I}$ if

$$\omega \in (0, \infty) \quad \text{and} \quad \mathbb{E}e^{\omega X_1} < 1.$$

At $\delta = \omega$, moment generating function $\mathbb{E}e^{\delta X_1} < 1$ jumps from a value strictly smaller than 1 to ∞.

Workload asymptotics: intermediate regime



Workload asymptotics: intermediate regime

Again change-of-measure technique can be used to find a uniform upper bound.

Define: $M(\delta) := \mathbb{E}e^{\delta X_1}$.

Identify with $\mathbb{Q}(\vartheta)$ the Lévy process that obeys

$$\mathbb{E}_{\mathbb{Q}(\vartheta)}e^{\delta X_1} = \frac{M(\delta+\vartheta)}{M(\vartheta)}.$$

As before, for all $\vartheta < \omega$,

$$\mathbb{P}(Q > u) = \mathbb{E}_{\mathbb{Q}(u)}\left(e^{-\vartheta X_{\sigma(u)}} \cdot (M(\vartheta))^{\sigma(u)}\right) \leq e^{-\vartheta u}.$$

We obtain the following bound.

Corollary: Let $X \in \mathscr{I}$. Then $\mathbb{P}(Q > u) \leq e^{-\omega u}$.

Workload asymptotics: intermediate regime

Without proof:

Proposition: Let $X \in \mathscr{I}$. Then, as $u \to \infty$,

$$\frac{\mathbb{P}(Q > u)}{\mathbb{P}(X_1 > u)} \to \frac{\mathbb{E}e^{\omega Q}}{M(\omega)\log M(\omega)}.$$

Workload asymptotics: intermediate regime

Without proof:

Proposition: Let $X \in \mathscr{I}$. Then, as $u \to \infty$,

$$\frac{\mathbb{P}(Q > u)}{\mathbb{P}(X_1 > u)} \to \frac{\mathbb{E}e^{\omega Q}}{M(\omega) \log M(\omega)}.$$

Interestingly, they show that for $X \in \mathscr{I}$ the tail distribution of Q is proportional to that of X_1!

Workload asymptotics: heavy-tailed regime

Now: Levy processes for which $\mathbb{E}e^{\delta X_1} = \infty$ for all $\delta > 0$.

Important subclass: regularly varying Lévy processes $\mathscr{R}$.

Workload asymptotics: heavy-tailed regime

Now: Levy processes for which $\mathbb{E}e^{\delta X_1} = \infty$ for all $\delta > 0$.

Important subclass: regularly varying Lévy processes $\mathscr{R}$.

Considering the class of compound Poisson inputs, regular variation refers to the tail of the distribution of the jobs: for an index α and all $y > 0$

$$\frac{\mathbb{P}(B > yx)}{\mathbb{P}(B > x)} \to y^{\alpha}.$$

Workload asymptotics: heavy-tailed regime

We now give 'recipe' to find the tail asymptotics $\mathbb{P}(Q > u)$ for u large

Key idea: in these heavy-tailed scenarios a large workload is (with overwhelming probability) due to a *single* big job.

Workload asymptotics: heavy-tailed regime

We now give 'recipe' to find the tail asymptotics $\mathbb{P}(Q > u)$ for u large

Key idea: in these heavy-tailed scenarios a large workload is (with overwhelming probability) due to a *single* big job.

The approach consists of

- ⋆ lower bound, in which the probability of this most likely scenario is evaluated, and
- ⋆ upper bound in which it is shown that the contributions of other scenarios (e.g. no big job, multiple big jobs) can be neglected.

Workload asymptotics: heavy-tailed regime

We here demonstrate how the lower bound is derived.

Consider $X \in \mathbb{C}\mathrm{P}(r, \lambda, b(\cdot))$; denote $\varrho := \lambda\,\mathbb{E}B$.

Due the the law of large numbers, we can find (for any $\delta, \varepsilon > 0$) a $t_{\delta,\varepsilon}$ such that for all $t \geq t_{\delta,\varepsilon}$,

$$\mathbb{P}(X_t > (\varrho - \varepsilon)t) > 1 - \delta.$$

To have that Q_0 exceeding u it suffices that

- ⋆ a job of size at least $u + (r - \varrho)t + \varepsilon t$ arrived at time $-t$, and
- ⋆ that between $-t$ and 0 at least $(\varrho - \varepsilon)t$ arrived;

former event is rare, as opposed to the latter.

Workload asymptotics: heavy-tailed regime

Hence:

$$\begin{aligned}\mathbb{P}(Q>u) &\geq \int_{t_{\delta,\varepsilon}}^{\infty} \lambda \mathbb{P}(B>u+(r-\varrho)t+\varepsilon t)\mathbb{P}(-X_{-t}>(\varrho-\varepsilon)t)\mathrm{d}t \\ &\geq (1-\delta)\int_{t_{\delta,\varepsilon}}^{\infty} \lambda \mathbb{P}(B>u+(1-\varrho)t+\varepsilon t)\mathrm{d}t \\ &= (1-\delta)\frac{\varrho}{r-\varrho+\varepsilon}\mathbb{P}(B^{\mathrm{res}}>u+t_{\delta,\varepsilon}) \sim \frac{(1-\delta)\varrho}{r-\varrho+\varepsilon}\mathbb{P}(B^{\mathrm{res}}>u);\end{aligned}$$

last step due to the definition of regular variation.

Workload asymptotics: heavy-tailed regime

Hence:

$$\begin{aligned}
\mathbb{P}(Q>u) &\geq \int_{t_{\delta,\varepsilon}}^{\infty} \lambda\mathbb{P}(B>u+(r-\varrho)t+\varepsilon t)\mathbb{P}(-X_{-t}>(\varrho-\varepsilon)t)\mathrm{d}t \\
&\geq (1-\delta)\int_{t_{\delta,\varepsilon}}^{\infty} \lambda\mathbb{P}(B>u+(1-\varrho)t+\varepsilon t)\mathrm{d}t \\
&= (1-\delta)\frac{\varrho}{r-\varrho+\varepsilon}\mathbb{P}(B^{\mathrm{res}}>u+t_{\delta,\varepsilon}) \sim \frac{(1-\delta)\varrho}{r-\varrho+\varepsilon}\mathbb{P}(B^{\mathrm{res}}>u);
\end{aligned}$$

last step due to the definition of regular variation.

Now let $\delta,\varepsilon\downarrow 0$. After proving the corresponding upper bound, the following theorem is obtained.

Theorem: Let $X\in\mathbb{CP}(r,\lambda,b(\cdot))\cap\mathscr{R}$. Then, as $u\to\infty$,

$$\mathbb{P}(Q>u)\sim\frac{\varrho}{r-\varrho}\mathbb{P}(B^{\mathrm{res}}>u).$$

Workload asymptotics: heavy-tailed regime

Alternative approach if Laplace transform is available: *Tauberian inversion*.

Workload asymptotics: heavy-tailed regime

Alternative approach if Laplace transform is available: *Tauberian inversion*.

Define the following notion.

Definition: We say that $f(x) \in \mathscr{R}_\delta(n, \eta)$, with $\delta \in (n, n+1)$, for $x \downarrow 0$, if

$$f(x) = \sum_{i=0}^{n} \frac{f^{(i)}(0)}{i!} x^i + \eta x^\delta L(1/x), \quad x \downarrow 0,$$

for a slowly varying function $L(\cdot)$, i.e., $L(x)/L(tx) \to 1$ for $x \to \infty$, for any t.

Workload asymptotics: heavy-tailed regime

Suppose now that $\varphi(\alpha) \in \mathscr{R}_\nu(n, \eta)$, it is readily checked that

$$\mathbb{E}e^{-\alpha Q} = \frac{\alpha\varphi'(0)}{\varphi(\alpha)} \in \mathscr{R}_{\nu-1}\left(n-1, \frac{\zeta}{\varphi'(0)}\right).$$

Tauberian theorem (Bingham, Goldie, and Teugels) now yields:

Theorem: Let $X \in \mathscr{S}_+ \cap \mathscr{R}$, with $\varphi(\alpha) \in \mathscr{R}_\nu(n, \eta)$. Then, as $u \to \infty$,

$$\mathbb{P}(Q > u) \sim \frac{(-1)^n}{\Gamma(2-\nu)} \cdot \left(\frac{\eta}{\varphi'(0)}\right) u^{1-\nu} L(u).$$

Workload asymptotics: heavy-tailed regime

Example: Consider $X \in \mathbb{C}\mathrm{P}(1, \lambda, b(\cdot))$. Suppose $\mathbb{P}(B > x) \sim x^{-\delta} L(x)$.

From $\varphi(\alpha) = \alpha + \lambda b(\alpha) - \lambda$, it follows that $\varphi(\alpha) \in \mathscr{R}_\delta(n, \lambda\Gamma(1-\delta)(-1)^n)$ by applying 'Tauber'.

Then the above theorem (for $X \in \mathscr{S}_+ \cap \mathscr{R}$) confirms the result for compound Poisson. $\diamondsuit$

Workload asymptotics: heavy-tailed regime

Define the class of heavy-tailed (or: *subexponential*) Lévy processes, as follows.

- ⋆ First introduce the notion of *subexponential distribution functions*:

  with $D(\cdot)$ being a distribution function on $[0,\infty)$ and $D^{(2)}$ the convolution of D with itself, we say that D is subexponential if $1-D^{(2)}(x)\sim 2(1-D(x))$ as $x\to\infty$.

- ⋆ For a measure $\mu(\cdot)$ we say that it is subexponential if (i) $\mu([1,\infty)<\infty$, and (ii) $\mu([1,\cdot])/\mu([1,\infty))$ is subexponential.

- ⋆ Then define

$$\Pi_I((x,\infty)) := \int_x^\infty \Pi((y,\infty))\mathrm{d}y.$$

- ⋆ We say that $X\in\mathscr{H}$ if $\Pi_I(\cdot)$ is a subexponential.

Workload asymptotics: heavy-tailed regime

Without proof:

Theorem: Let $X \in \mathscr{H}$. Then, as $u \to \infty$,

$$\mathbb{P}(Q > u) \sim \frac{1}{-\mathbb{E}X_1} \int_u^\infty \mathbb{P}(X_1 > x) \mathrm{d}x.$$

Workload asymptotics: heavy-tailed regime

Without proof:

Theorem: Let $X \in \mathscr{H}$. Then, as $u \to \infty$,

$$\mathbb{P}(Q > u) \sim \frac{1}{-\mathbb{E}X_1} \int_u^\infty \mathbb{P}(X_1 > x) \mathrm{d}x.$$

Cf.: single big job.

Workload asymptotics: heavy-tailed regime

Class of α-stable Lévy motions belongs to $\mathscr{H}$.

Recall: Let $Y \stackrel{\mathrm{d}}{=} S_\alpha(\sigma, \beta, m)$. Then, as $u \to \infty$,

$$\mathbb{P}(Y > u)u^\alpha \to C_{\alpha,\sigma}\left(\frac{1+\beta}{2}\right),$$

where

$$C_{\alpha,\sigma} := \begin{cases} \sigma^\alpha(1-\alpha)/\left(\Gamma(2-\alpha)\cos(\pi\alpha/2)\right) & \alpha \neq 1; \\ 2\sigma/\pi & \alpha = 1. \end{cases}$$

Workload asymptotics: heavy-tailed regime

Following result is immediate consequence of result for $X \in \mathscr{H}$, asymptotics of $\mathbb{P}(X_1 > u)$, and Karamata's theorem; recall that $m < 0$.

Proposition: Let $X \in \mathbb{S}(\alpha, \beta, m)$, with $\alpha \in (1, 2)$. Then

$$\mathbb{P}(Q > u) \sim \frac{1}{(-m)} \int_u^\infty x^{-\alpha} C_{\alpha,1} \left(\frac{1+\beta}{2} \right) \mathrm{d}x \sim \frac{1}{(-m)} \frac{1}{\alpha - 1} u^{-\alpha+1} C_{\alpha,1} \left(\frac{1+\beta}{2} \right).$$

Busy-period asymptotics

Busy-period asymptotics

First consider the light-tailed case.

Busy-period asymptotics

First consider the light-tailed case.

- ⋆ Then $\mathbb{E}e^{-sX_1} = 1$ has a negative root, say $\omega < 0$.
- ⋆ Hence: $\mathbb{E}e^{-sX_1}$ has a minimizer somewhere between ω and 0.

Relying on Heaviside heuristics, we now study the tail of $\mathbb{P}(\tau > t)$ ($r(t)$ works similarly).

Busy-period asymptotics

Considering $X \in \mathscr{S}_+ \cap \mathscr{L}$, assume $\varphi(\alpha) = 0$ has a negative root.

Recall:

$$\int_0^\infty e^{-\vartheta t} p(t) \mathrm{d}t = \frac{1}{\vartheta} - \varphi'(0) \frac{\psi(\vartheta)}{\vartheta^2}.$$

Hence transform holds for any positive ϑ, but we can consider the analytic continuation up to the branching point $\vartheta^\star < 0$ of $\psi(\cdot)$.

Busy-period asymptotics

$\zeta < 0$ denotes the minimizer of $\varphi(\cdot)$, so that $\varphi(\zeta) = \vartheta^\star < 0$ (notice that $v_\varphi := \varphi''(\zeta) > 0$).

Then write, for $\vartheta \downarrow \vartheta^\star$,

$$\psi(\vartheta) - \zeta \sim \sqrt{2/v_\varphi} \cdot \sqrt{\vartheta - \vartheta^\star}.$$

Busy-period asymptotics

$\zeta < 0$ denotes the minimizer of $\varphi(\cdot)$, so that $\varphi(\zeta) = \vartheta^\star < 0$ (notice that $v_\varphi := \varphi''(\zeta) > 0$).

Then write, for $\vartheta \downarrow \vartheta^\star$,

$$\psi(\vartheta) - \zeta \sim \sqrt{2/v_\varphi} \cdot \sqrt{\vartheta - \vartheta^\star}.$$

Hence, around $\vartheta^\star$, we have, for some (irrelevant) constant κ,

$$\int_0^\infty e^{-\vartheta t}\mathbb{P}(\tau > t)\mathrm{d}t = \frac{1}{\vartheta} - \varphi'(0)\frac{\psi(\vartheta)}{\vartheta^2} \sim \kappa + A_\varphi\sqrt{\vartheta - \vartheta^\star}; \quad A_\varphi := -\frac{\varphi'(0)}{(\vartheta^\star)^2}\sqrt{\frac{2}{v_\varphi}} < 0.$$

Busy-period asymptotics

$\zeta < 0$ denotes the minimizer of $\varphi(\cdot)$, so that $\varphi(\zeta) = \vartheta^\star < 0$ (notice that $v_\varphi := \varphi''(\zeta) > 0$).

Then write, for $\vartheta \downarrow \vartheta^\star$,

$$\psi(\vartheta) - \zeta \sim \sqrt{2/v_\varphi} \cdot \sqrt{\vartheta - \vartheta^\star}.$$

Hence, around $\vartheta^\star$, we have, for some (irrelevant) constant κ,

$$\int_0^\infty e^{-\vartheta t}\mathbb{P}(\tau > t)\mathrm{d}t = \frac{1}{\vartheta} - \varphi'(0)\frac{\psi(\vartheta)}{\vartheta^2} \sim \kappa + A_\varphi\sqrt{\vartheta - \vartheta^\star}; \quad A_\varphi := -\frac{\varphi'(0)}{(\vartheta^\star)^2}\sqrt{\frac{2}{v_\varphi}} < 0.$$

'Heaviside': tail distribution of the busy period is

$$\mathbb{P}(\tau > t) \sim \frac{A_\varphi}{\Gamma(-\frac{1}{2})} \cdot \frac{e^{\vartheta^\star t}}{t\sqrt{t}}.$$

Busy-period asymptotics

Considering $X \in \mathscr{S}_+ \cap \mathscr{L}$: works similarly.

Heavy-tailed case with compound Poisson input has also been analyzed.

Asymptotics of joint transient distribution

Focus is on probabilities of the type

$$\mathbb{P}(Q_0 > pu, Q_{T(u)} > qu),$$

for $p, q > 0$ and functions $T(\cdot)$.

We summarize the main results.

Asymptotics of joint transient distribution

Focus is on probabilities of the type

$$\mathbb{P}(Q_0 > pu, Q_{T(u)} > qu),$$

for $p, q > 0$ and functions $T(\cdot)$.

We summarize the main results.

A. Under certain conditions probability of interest is dominated by 'most demanding event':

$$\mathbb{P}(Q_0 > pu, Q_{T(u)} > qu) \sim \mathbb{P}(Q > \max\{p, q\}u)$$

for u large, where Q denotes the steady-state workload.

These conditions turn out to reduce to $T(u)$ being sublinear (i.e., $T(u)/u \to 0$ as $u \to \infty$).

Asymptotics of joint transient distribution

B. Under another condition the probability 'decouples':

$$\mathbb{P}(Q_0 > pu, Q_{T(u)} > qu) \sim \mathbb{P}(Q > pu)\mathbb{P}(Q > qu).$$

Here crucial role is played by Q^D, for $D > \mathbb{E}X_1$, which is distributed as $\sup_{t\geq 0}(X_t - Dt)$; as a result Q^D resembles the original queue Q, but the drain rate is adapted by D.

Decoupling condition: for all $\eta > 0$, $D > \mathbb{E}X_1$,

$$\lim_{u\to\infty} \frac{\mathbb{P}(Q^D > \eta T(u))}{\mathbb{P}(Q > pu)\mathbb{P}(Q > qu)} = 0.$$

Asymptotics of joint transient distribution

B. Decoupling condition: for all $\eta > 0$, $D > \mathbb{E}X_1$,

$$\lim_{u\to\infty} \frac{\mathbb{P}(Q^D > \eta T(u))}{\mathbb{P}(Q > pu)\mathbb{P}(Q > qu)} = 0.$$

For various Lévy inputs 'decoupling condition' reduces to requiring that T_u is superlinear (i.e., $T_u/u \to \infty$ as $u \to \infty$); for instance if tails of Q and Q^D decay exponentially.

Asymptotics of joint transient distribution

B. Decoupling condition: for all $\eta > 0$, $D > \mathbb{E}X_1$,

$$\lim_{u\to\infty} \frac{\mathbb{P}(Q^D > \eta T(u))}{\mathbb{P}(Q > pu)\mathbb{P}(Q > qu)} = 0.$$

For various Lévy inputs 'decoupling condition' reduces to requiring that T_u is superlinear (i.e., $T_u/u \to \infty$ as $u \to \infty$); for instance if tails of Q and Q^D decay exponentially.

Condition does *not* hold, however, for $X \in \mathscr{R}$: then 'decoupling' reduces to $T(u)/u^2 \to \infty$.

Rationale: for $T(u)$ increasing subquadratically with overwhelming probability it suffices to have a *single* big jump to cause overflow over pu at time 0, and over qu at time $T(u)$; whereas 'decoupling' would correspond to *two* big jumps.

Asymptotics of joint transient distribution

B. Decoupling condition: for all $\eta > 0$, $D > \mathbb{E}X_1$,

$$\lim_{u\to\infty} \frac{\mathbb{P}(Q^D > \eta T(u))}{\mathbb{P}(Q > pu)\mathbb{P}(Q > qu)} = 0.$$

For various Lévy inputs 'decoupling condition' reduces to requiring that T_u is superlinear (i.e., $T_u/u \to \infty$ as $u \to \infty$); for instance if tails of Q and Q^D decay exponentially.

Condition does *not* hold, however, for $X \in \mathscr{R}$: then 'decoupling' reduces to $T(u)/u^2 \to \infty$.

Rationale: for $T(u)$ increasing subquadratically with overwhelming probability it suffices to have a *single* big jump to cause overflow over pu at time 0, and over qu at time $T(u)$; whereas 'decoupling' would correspond to *two* big jumps.

Hence for $X \in \mathscr{R}$ there is a *third* regime, viz. $T(u)$ increasing superlinearly but subquadratically.

Asymptotics of joint transient distribution

Special interesting case: $T(u) = Ru$ for some $R > 0$;

for $X \in \mathscr{L}$ intuitively appealing asymptotics are known, based on sample-path large deviations results.

The regimes obtained can be interpreted in terms of most likely paths to overflow.

Asymptotics of joint transient distribution

⋆ If R small (that is, fulfilling explicit criterion in terms of p, q, and characteristics of the Lévy process $(X_t)_t$), then asymptotics are of type $\mathbb{P}(Q > \max\{p, q\}u)$.

⋆ If this condition does not apply, two cases are possible:

- for large R most likely scenario is that buffer first builds up pu, then drains, remains empty for a while, and starts building up relatively short before R.

  In this case asymptotics look like $\mathbb{P}(Q > pu)\mathbb{P}(Q > qu)$.

- for moderate R buffer remains (most likely) nonempty between 0 and R.

Asymptotics of joint transient distribution

Hence: there are (uniquely characterized) $\bar{R}$ and $\check{R}$ such that for all R smaller than $\bar{R}$,

$$\lim_{u\to\infty} \frac{1}{u} \log \mathbb{P}(Q_0 > pu, Q_{Ru} > qu) = -\max\{p,q\}\omega,$$

for R between $\bar{R}$ and $\check{R}$,

$$\lim_{u\to\infty} \frac{1}{u} \log \mathbb{P}(Q_0 > pu, Q_{Ru} > qu) = -p\omega - R\cdot \sup_{\delta} \left(\delta\left(\frac{q-p}{R}\right) - \log \mathbb{E} e^{\delta X_1}\right),$$

and for R larger than $\check{R}$,

$$\lim_{u\to\infty} \frac{1}{u} \log \mathbb{P}(Q_0 > pu, Q_{Ru} > qu) = -(p+q)\omega,$$

where ω solves $\mathbb{E} e^{\omega X_1} = 1$.

Rare-event simulation, importance sampling

What to do if you're not sure the asymptotic regime has kicked in?

Rare-event simulation, importance sampling

What to do if you're not sure the asymptotic regime has kicked in?

Simuation!

Rare-event simulation, importance sampling

What to do if you're not sure the asymptotic regime has kicked in?

Simuation!

Here: estimation of

- ⋆ Tail of workload distribution;
- ⋆ Tail of busy-period distribution;
- ⋆ Tail of workload correlation function;

Rare-event simulation: tail of workload distribution

General statement:

number of simulation runs needed to obtain an estimate with predefined precision (expressed in terms of the ratio of the width of the confidence interval and the estimate), is *inversely proportional to the probability to be estimated*.

Rare-event simulation: tail of workload distribution

General statement:

number of simulation runs needed to obtain an estimate with predefined precision (expressed in terms of the ratio of the width of the confidence interval and the estimate), is *inversely proportional to the probability to be estimated*.

Suppose $X \in \mathscr{L}$.

Number of runs needed to estimate $\mathbb{P}(Q > u)$ grows exponentially in u

Objective: speed up the simulation.

Rare-event simulation: tail of workload distribution

Let ω solve $\mathbb{E}e^{\omega X_1} = 1$.

Rare-event simulation: tail of workload distribution

Let ω solve $\mathbb{E}e^{\omega X_1} = 1$.

Idea: do not perform simulation under the original measure $\mathbb{P}$, corresponding to the characteristic triplet (d, σ^2, Π),

but under an alternative measure $\mathbb{Q}$ under which the event of interest occurs more frequently.

After weighing simulation output with appropriate likelihood ratios: *importance sampling.*

Rare-event simulation: tail of workload distribution

This $\mathbb{Q}$ is *exponentially twisted* version of $\mathbb{P}$, that is, $\mathbb{Q}$ is such that, in self-evident notation,

$$\mathbb{E}_{\mathbb{Q}} e^{\delta X_1} = \mathbb{E} e^{(\delta+\omega) X_1}$$

(is a Laplace transform!).

Rare-event simulation: tail of workload distribution

This $\mathbb{Q}$ is *exponentially twisted* version of $\mathbb{P}$, that is, $\mathbb{Q}$ is such that, in self-evident notation,

$$\mathbb{E}_{\mathbb{Q}} e^{\delta X_1} = \mathbb{E} e^{(\delta+\omega) X_1}$$

(is a Laplace transform!).

Rare-event simulation: tail of workload distribution

$$\mathbb{E}_{\mathbb{Q}} e^{\delta X_1} = \mathbb{E} e^{(\delta+\omega)X_1}$$

Elementary to check that $\mathbb{Q}$ also corresponds to a Lévy process, with triplet

$$\left(d + \sigma^2 \omega + \int_{-1}^{1} x(e^{\omega x} - 1)\Pi(\mathrm{d}x), \sigma^2, e^{\omega x}\Pi(\mathrm{d}x) \right).$$

Rare-event simulation: tail of workload distribution

$$\mathbb{E}_{\mathbb{Q}} e^{\delta X_1} = \mathbb{E} e^{(\delta+\omega) X_1}$$

Elementary to check that $\mathbb{Q}$ also corresponds to a Lévy process, with triplet

$$\left(d + \sigma^2 \omega + \int_{-1}^{1} x(e^{\omega x} - 1)\Pi(\mathrm{d}x), \sigma^2, e^{\omega x}\Pi(\mathrm{d}x) \right).$$

Convexity of $\mathbb{E} e^{\delta X_1}$ implies that

$$\mathbb{E}_{\mathbb{Q}} X_1 = \mathbb{E} X_1 e^{\omega X_1} > 0,$$

so that the random variable

$$T := \inf\{t : X_t \geq u\}$$

becomes nondefective under $\mathbb{Q}$.

Rare-event simulation: tail of workload distribution

Hence, as before:

$$\mathbb{P}(Q > u) = \mathbb{E}_{\mathbb{Q}} e^{-\omega X_T};$$

cf. the change-of-measure arguments used for 'Cramér-Lundberg'.

Rare-event simulation: tail of workload distribution

Hence, as before:

$$\mathbb{P}(Q > u) = \mathbb{E}_{\mathbb{Q}} e^{-\omega X_T};$$

cf. the change-of-measure arguments used for 'Cramér-Lundberg'.

Idea: simulate under $\mathbb{Q}$ until T, record the value Y_i of $e^{-\omega X_T}$ in each run i, perform n runs, and estimate $\mathbb{P}(Q > u)$ by

$$t_n := \frac{1}{n} \sum_{i=1}^{n} y_i,$$

with y_i realizations of Y_i.

Unbiased estimator!

Rare-event simulation: tail of workload distribution

Estimator:

$$t_n := \frac{1}{n}\sum_{i=1}^{n} y_i,$$

with y_i realizations of Y_i.

Observe: Y_i are bounded by $e^{-\omega u}$.

Rare-event simulation: tail of workload distribution

Estimator:

$$t_n := \frac{1}{n}\sum_{i=1}^{n} y_i,$$

with y_i realizations of Y_i.

Observe: Y_i are bounded by $e^{-\omega u}$.

Also, as variances are positive,

$$\mathbb{E}\left(\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2\right) \geq \left(\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)\right)^2 = \left(\mathbb{P}(Q > u)\right)^2,$$

so that

$$\liminf_{u\to\infty}\frac{1}{u}\log\mathbb{E}\left(\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2\right) \geq \liminf_{u\to\infty}\frac{2}{u}\log\mathbb{P}(Q > u) = -2\omega.$$

Rare-event simulation: tail of workload distribution

Our estimator actually achieves this lower bound:

$$\mathbb{E}\left(\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2\right) \le e^{-2\omega u},$$

so that

$$\limsup_{u\to\infty} \frac{1}{u} \log \mathbb{E}\left(\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2\right) \le -2\omega.$$

We call the estimator asymptotically efficient.

Rare-event simulation: tail of busy-period distribution

Recall: $\tau := \inf\{t \geq 0 : Q_t = 0\}$, where Q_0 has stationary distribution.

$p(t) := \mathbb{P}(\tau > t)$.

Rare-event simulation: tail of busy-period distribution

Earlier we found striking feature:

transforms have the same *branching point* as the transforms of the workload correlation function!!

Spectrally positive, light tails ($\exists \alpha < 0 : \varphi(\alpha) = 0$): we roughly have

$$r(t) \sim p(t) \sim e^{\vartheta^\star t},$$

where ζ is the minimizer of $\varphi(\cdot)$ and $\vartheta^\star = \varphi(\zeta)$ the branching point of $\psi(\cdot)$.

Idea: develop importance sampling technique for $p(t)$ that can be reused for $r(t) = \mathbb{C}\mathrm{orr}(Q_0, Q_t)$.

Rare-event simulation: tail of busy-period distribution

Naïve simulation: estimate $p(t)$ by

$$S_n^{(\mathrm{NS})}(t) := \frac{1}{n}\sum_{i=1}^{n} 1\{\tau_i > t\}.$$

Rare-event simulation: tail of busy-period distribution

Naïve simulation: estimate $p(t)$ by

$$S_n^{(\mathrm{NS})}(t) := \frac{1}{n}\sum_{i=1}^{n} 1\{\tau_i > t\}.$$

Number of runs needed to obtain estimate of given precision: roughly of order $1/p(t)$, i.e., exponentially increasing...

Rare-event simulation: tail of busy-period distribution

Again: fast algorithm based on importance sampling.

- $\star$ Let, in the interval $(0, t]$, the Lévy process be twisted with $-\zeta = -\psi(\vartheta^\star) > 0$.
  Meaning: $\varphi(\vartheta)$ replaced by $\bar{\varphi}(\vartheta) := \varphi(\vartheta + \zeta) - \varphi(\zeta)$.
- $\star$ But what about distribution of Q_0?
  Simulate Q_0 from a κ-twisted version, i.e., a distribution with LT $\mathbb{E}e^{-(\alpha-\kappa)Q_0}/\mathbb{E}e^{\kappa Q_0}$.

Call new measure $\mathbb{Q}_\kappa$.

Rare-event simulation: tail of busy-period distribution

We simulate the process under $\mathbb{Q}_\kappa$ till time t. Likelihood $L := L_A \cdot L_B$, where

- contribution due to the twisted Lévy process between 0 and t:

$$L_A := e^{\psi(\vartheta^\star)X_t} \cdot \mathbb{E}e^{-\psi(\vartheta^\star)X_t} = e^{\psi(\vartheta^\star)X_t} \cdot e^{\vartheta^\star t}.$$

- contribution due to the twisted queue at time 0 (use 'Pollaczek-Khinchine'):

$$L_B := e^{-\kappa Q_0} \cdot \mathbb{E}e^{\kappa Q_0} = e^{-\kappa Q_0} \cdot \frac{-\kappa\varphi'(0)}{\varphi(-\kappa)}.$$

Estimate $p(t)$ by, sampling under $\mathbb{Q}_\kappa$,

$$S_n^{(\mathrm{IS})}(t) := \frac{1}{n}\sum_{i=1}^{n} L_i 1\{\tau_i > t\}.$$

Rare-event simulation: tail of busy-period distribution

$$L = e^{\psi(\vartheta^\star)X_t} \cdot e^{\vartheta^\star t} \cdot e^{-\kappa Q_0} \cdot \frac{-\kappa\varphi'(0)}{\varphi(-\kappa)}.$$

First option: not twisting Q_0 at all (i.e., choosing $\kappa = 0$).

This does *not* work well: recalling that a necessary condition for $\{\tau > t\}$ is $\{Q_0 + X_t > 0\}$, we find

$$\mathbb{E}_{\mathbb{Q}_\kappa} L^2 1\{\tau > t\} \leq \left(-\frac{\kappa\varphi'(0)}{\varphi(-\kappa)}\right)^2 e^{2\vartheta^\star t}\mathbb{E}_{\mathbb{Q}_\kappa} e^{-2\kappa Q_0} e^{-2\psi(\vartheta^\star)Q_0}.$$

Rare-event simulation: tail of busy-period distribution

$$L = e^{\psi(\vartheta^\star)X_t} \cdot e^{\vartheta^\star t} \cdot e^{-\kappa Q_0} \cdot \frac{-\kappa\varphi'(0)}{\varphi(-\kappa)}.$$

First option: not twisting Q_0 at all (i.e., choosing $\kappa = 0$).

This does *not* work well: recalling that a necessary condition for $\{\tau > t\}$ is $\{Q_0 + X_t > 0\}$, we find

$$\mathbb{E}_{\mathbb{Q}_\kappa} L^2 1\{\tau > t\} \leq \left(-\frac{\kappa\varphi'(0)}{\varphi(-\kappa)}\right)^2 e^{2\vartheta^\star t}\mathbb{E}_{\mathbb{Q}_\kappa} e^{-2\kappa Q_0} e^{-2\psi(\vartheta^\star)Q_0}.$$

Asymptotic efficiency, meaning that the number of replications needed to obtain an estimate with a certain fixed precision grows subexponentially in the 'rarity parameter' t:

$$\limsup_{t\to\infty} t^{-1} \log \mathbb{E}_{\mathbb{Q}_\kappa} L^2 1\{\tau > t\} \leq 2\vartheta^\star.$$

In other words: when picking $\kappa = 0$ we need to have $\mathbb{E}_{\mathbb{Q}_0} e^{-2\psi(\vartheta^\star)Q_0} < \infty$ for logarithmic efficiency... *Not* a priori clear....

Rare-event simulation: tail of busy-period distribution

$$L = e^{\psi(\vartheta^\star)X_t} \cdot e^{\vartheta^\star t} \cdot e^{-\kappa Q_0} \cdot \frac{-\kappa\varphi'(0)}{\varphi(-\kappa)}.$$

First option: not twisting Q_0 at all (i.e., choosing $\kappa = 0$).

This does *not* work well: recalling that a necessary condition for $\{\tau > t\}$ is $\{Q_0 + X_t > 0\}$, we find

$$\mathbb{E}_{\mathbb{Q}_\kappa} L^2 1\{\tau > t\} \leq \left(-\frac{\kappa\varphi'(0)}{\varphi(-\kappa)}\right)^2 e^{2\vartheta^\star t}\mathbb{E}_{\mathbb{Q}_\kappa} e^{-2\kappa Q_0} e^{-2\psi(\vartheta^\star)Q_0}.$$

Rare-event simulation: tail of busy-period distribution

$$L = e^{\psi(\vartheta^\star)X_t} \cdot e^{\vartheta^\star t} \cdot e^{-\kappa Q_0} \cdot \frac{-\kappa\varphi'(0)}{\varphi(-\kappa)}.$$

Second option: twisting with $\kappa = -\zeta > 0$.

Easy to see that we do get logarithmic efficiency here!

Rare-event simulation: workload correlation function

But can we come up with an efficient simulation algorithm for $r(t)$?

Rare-event simulation: workload correlation function

But can we come up with an efficient simulation algorithm for $r(t)$?

Remember:

$$r(t) = \frac{\mathbb{E}Q_0Q_t - \mu^2}{v},$$

with $\mu := \mathbb{E}Q$ and $v := \mathbb{V}\mathrm{ar}Q$ known...

We can estimate $\mathbb{E}Q_0Q_t - \mu^2$ by

$$T_n^{(\mathrm{NS})}(x) := \frac{1}{n}\sum_{i=1}^{n} Q_0^{(i)}Q_t^{(i)} - \mu^2.$$

How many runs needed?

Rare-event simulation: workload correlation function

Variance of this estimator:

$$\frac{1}{n} \cdot \mathbb{V}\mathrm{ar}\,(Q_0 Q_t) = \frac{\mathbb{E}(Q_0^2 Q_t^2) - (\mathbb{E}(Q_0 Q_t))^2}{n} \to \frac{(\mathbb{E}Q^2)^2 - (\mathbb{E}Q)^4}{n};$$

Conclude: number of runs needed roughly proportional to $1/r(t)^2$!!!

Rare-event simulation: workload correlation function

Solution: coupling

Rare-event simulation: workload correlation function

We construct a coupling as follows.

Write:

$$r(t) = \frac{1}{v} \cdot \mathbb{E}(Q_0 \cdot (Q_t - Q_t^\star)),$$

where both Q and $Q^\star$ are stationary versions of the workload, and $Q_t^\star$ is *independent* of Q_0.

Construct this as follows: generate Q_0 and $Q_0^\star$ independently, sampled from the stationary distribution of the workload. Now use exactly the same driving Lévy process X_t over $(0, t]$ to drive both Q_t and $Q_t^\star$ from their two independently generated initial conditions.

This makes Q_t and Q_0 correlated but $Q_t^\star$ and Q_0 independent.

Rare-event simulation: workload correlation function

We can estimate $\mathbb{E}Q_0Q_t - \mu^2$ by

$$T_n^{(\mathrm{CS})}(x) := \frac{1}{n}\sum_{i=1}^{n} Q_0^{(i)}(Q_t^{(i)} - Q_t^{\star(i)}).$$

What is performance of this estimator?

Rare-event simulation: workload correlation function

Split $\mathbb{E}(Q_0 \cdot (Q_t - Q_t^\star))$ into four terms, as follows.

Recall $M_t = \inf_{s\in(0,t]} X_s$. Then

$$r(t) = r_{++}(t) + r_{+-}(t) + r_{-+}(t) + r_{--}(t),$$

where

$$\begin{aligned}
r_{++}(t) &:= \mathbb{E}(Q_0 \cdot (Q_t - Q_t^\star) \cdot 1\{Q_0 + M_t > 0, Q_0^\star + M_t > 0\}),\\
r_{+-}(t) &:= \mathbb{E}(Q_0 \cdot (Q_t - Q_t^\star) \cdot 1\{Q_0 + M_t > 0, Q_0^\star + M_t < 0\}),\\
r_{-+}(t) &:= \mathbb{E}(Q_0 \cdot (Q_t - Q_t^\star) \cdot 1\{Q_0 + M_t < 0, Q_0^\star + M_t > 0\}),\\
r_{--}(t) &:= \mathbb{E}(Q_0 \cdot (Q_t - Q_t^\star) \cdot 1\{Q_0 + M_t < 0, Q_0^\star + M_t < 0\}).
\end{aligned}$$

Rare-event simulation: workload correlation function

Split $\mathbb{E}(Q_0 \cdot (Q_t - Q_t^\star))$ into four terms, as follows.

Recall $M_t = \inf_{s\in(0,t]} X_s$. Then

$$r(t) = r_{++}(t) + r_{+-}(t) + r_{-+}(t) + r_{--}(t),$$

where

$$\begin{aligned}
r_{++}(t) &:= \mathbb{E}(Q_0 \cdot (Q_t - Q_t^\star) \cdot 1\{Q_0 + M_t > 0, Q_0^\star + M_t > 0\}),\\
r_{+-}(t) &:= \mathbb{E}(Q_0 \cdot (Q_t - Q_t^\star) \cdot 1\{Q_0 + M_t > 0, Q_0^\star + M_t < 0\}),\\
r_{-+}(t) &:= \mathbb{E}(Q_0 \cdot (Q_t - Q_t^\star) \cdot 1\{Q_0 + M_t < 0, Q_0^\star + M_t > 0\}),\\
r_{--}(t) &:= \mathbb{E}(Q_0 \cdot (Q_t - Q_t^\star) \cdot 1\{Q_0 + M_t < 0, Q_0^\star + M_t < 0\}).
\end{aligned}$$

It is evident that $r_{--}(t) = 0$ as both queues have been empty (and this happens most of the time!)

Rare-event simulation: workload correlation function

Key observation: $|Q_t - Q_t^\star| \le |Q_0 - Q_0^\star|$.

We therefore have:

$$\mathbb{V}\mathrm{ar}\,(Q_0(Q_t - Q_t^\star)) \le \mathbb{E}Q_0^2(Q_t - Q_t^\star)^2 \le \mathbb{E}Q_0^2(Q_0 - Q_0^\star)^2.$$

In addition:

$$\begin{aligned}\mathbb{E}Q_0^2(Q_0 - Q_0^\star)^2 \;\le\; & \mathbb{E}(Q_0^2(Q_0 - Q_0^\star)^2 \cdot 1\{Q_0 + M_t > 0, Q_0^\star + M_t > 0\}) + \\ & + \mathbb{E}(Q_0^2(Q_0 - Q_0^\star)^2 \cdot 1\{Q_0 + M_t > 0, Q_0^\star + M_t \le 0\}) \\ & + \mathbb{E}(Q_0^2(Q_0 - Q_0^\star)^2 \cdot 1\{Q_0 + M_t \le 0, Q_0^\star + M_t > 0\})\end{aligned}$$

Rare-event simulation: workload correlation function

Lemma: in the spectrally-positive case

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}(Q_0^k 1\{\tau > t\}) \le \vartheta^\star$$

(and $\ldots \le q^\star$ in the spectrally-negative case).

Hence,

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{V}\mathrm{ar}\,(Q_0(Q_t - Q_t^\star)) \le \vartheta^\star.$$

Rare-event simulation: workload correlation function

Lemma: in the spectrally-positive case

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}(Q_0^k 1\{\tau > t\}) \leq \vartheta^\star$$

(and $\ldots \leq q^\star$ in the spectrally-negative case).

Hence,

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{V}\mathrm{ar}\,(Q_0(Q_t - Q_t^\star)) \leq \vartheta^\star.$$

Consequently,

$$\frac{\sqrt{\mathbb{V}\mathrm{ar}\, T_n^{(\mathrm{CS})}(x)}}{r(t)} \approx \frac{\sqrt{e^{\vartheta^\star t}/n}}{e^{\vartheta^\star t}},$$

so that number of runs needed grows roughly as $1/r(t)$.

Substantial improvement!

Rare-event simulation: workload correlation function

Augment coupling algorithm with importance sampling (as for busy period),

and we even get an asymptotically efficient algorithm (i.e., number of runs grows *subexponentially*).

Example: estimation of $r(t)$ for reflected Brownian motion

Take $\mu = -1$, $\sigma^2 = 1$; remember

$$Q_t = X_t + \max\left\{-\inf_{0\leq s\leq t} X_s, Q_0\right\}.$$

Q_0 has an exponential distrbution with mean $\frac{1}{2}$.

Example: estimation of $r(t)$ for reflected Brownian motion

Take $\mu = -1$, $\sigma^2 = 1$; remember

$$Q_t = X_t + \max\left\{-\inf_{0\le s\le t} X_s, Q_0\right\}.$$

Q_0 has an exponential distrbution with mean $\frac{1}{2}$.

Then we sample X_t from a normal distribution with mean $-t$ and variance t; say it has value z. Using Brownian Bridge:

$$\mathbb{P}\left(-\inf_{0\le s\le t} X_s \le x \,\middle|\, X_t = z\right) = \exp\left(-2\frac{x}{t}(x+z)\right).$$

Then it can be verified that

$$Y_z := \left(-\inf_{0\le s\le t} X_s \,\middle|\, X_t = z\right) \stackrel{\mathrm{d}}{=} -\frac{z}{2} + \frac{1}{2}\sqrt{z^2 - 2t\log U},$$

where U has a uniform distribution over $(0, 1]$.

Hence: easy simulation of Q_t, requiring just three random numbers!

⋆ Perform 10^8 runs per experiment;

⋆ the table gives the relative errors.

		Naive	Coupling	IS
$t = 10$	$7.91 \cdot 10^{-4}$	35%	0.85%	0.038%
$t = 12$	$2.21 \cdot 10^{-4}$	75%	1.50%	0.042%
$t = 14$	$6.75 \cdot 10^{-5}$	133%	2.82%	0.045%
$t = 16$	$2.17 \cdot 10^{-5}$	151%	4.99%	0.049%
$t = 18$	$6.83 \cdot 10^{-6}$	160%	8.4%	0.054%
$t = 20$	$2.27 \cdot 10^{-6}$	188%	11.9%	0.057%

⋆ Perform 10^8 runs per experiment;

⋆ the table gives the relative errors.

		Naive	Coupling	IS
$t = 10$	$7.91 \cdot 10^{-4}$	35%	0.85%	0.038%
$t = 12$	$2.21 \cdot 10^{-4}$	75%	1.50%	0.042%
$t = 14$	$6.75 \cdot 10^{-5}$	133%	2.82%	0.045%
$t = 16$	$2.17 \cdot 10^{-5}$	151%	4.99%	0.049%
$t = 18$	$6.83 \cdot 10^{-6}$	160%	8.4%	0.054%
$t = 20$	$2.27 \cdot 10^{-6}$	188%	11.9%	0.057%

⋆ Under importance sampling the relative error is more or less constant!

PART IV:

VARIANTS OF THE STANDARD QUEUE

Variants of the standard queue

- $\star$ Finite-buffer queues;
- $\star$ Models with feedback;
- $\star$ Vacation and polling models;
- $\star$ Models with Markov-additive input.

Finite-buffer queues

Consider a Lévy-driven queue in which workload cannot exceed level $K > 0$.

Corresponding Skorokhod problem can be formulated, in which Q_t is expressed in terms of

- $\star$ local time at 0 (as before),
- $\star$ but now also the local time at K.

Finite-buffer queues

Assuming for ease $Q_0 = 0$. Then

$$Q_t = X_t + L_t - \bar{L}_t,$$

with L_t ($\bar{L}_t$) the local time at 0 (at K, respectively);

popularly speaking, L_t only increases when $Q_t = 0$, whereas $\bar{L}_t$ only increases when $Q_t = K$.

Finite-buffer queues

Assuming for ease $Q_0 = 0$. Then

$$Q_t = X_t + L_t - \bar{L}_t,$$

with L_t ($\bar{L}_t$) the local time at 0 (at K, respectively); popularly speaking, L_t only increases when $Q_t = 0$, whereas $\bar{L}_t$ only increases when $Q_t = K$.

Then Q_t can be explicitly solved:

$$Q_t = X_t - \sup_{s\in[0,t]} \left(\max\left\{ \min\left\{ X_s - K, \inf_{u\in[0,t]} X_u \right\}, \inf_{u\in[s,t]} X_u \right\} \right),$$

whereas an alternative solution is

$$Q_t = \sup_{s\in[0,t]} \max\left\{ X_t - X_s, \inf_{u\in[s,t]} (K + X_t - X_u) \right\}.$$

Finite-buffer queues

First part of following result characterizes steady-state workload Q in terms of a first-passage time (not required anymore that $\mathbb{E}X_1 < 0$).

Second part assumes $X \in \mathscr{S}_-$, but realize $X \in \mathscr{S}_+$ can be dealt with analogously.

Recall (implicit) definition of $W^{(0)}(\cdot)$: a strictly increasing and continuous function whose Laplace transform satisfies

$$\int_0^\infty e^{-\beta x} W^{(0)}(x)\mathrm{d}x = \frac{1}{\Phi(\beta)}, \qquad \beta > \Psi(0).$$

Finite-buffer queues

Write $\pi_K(u) := \mathbb{P}(Q < u)$.

Proposition: (i) For $u \in [0, K]$,

$$1 - \pi_K(u) = \mathbb{P}(X_{\tau[y-K,y)} \le y),$$

where $\tau[u, v) := \inf\{t \ge 0 : X_t \notin [u, v)\}$, for $u \le 0 \le v$.

(ii) Let $X \in \mathscr{S}_-$. Then, for $u \in [0, K]$,

$$1 - \pi_K(u) = \frac{W^{(0)}(K - x)}{W^{(0)}(K)}.$$

Finite-buffer queues

Write $\pi_K(u) := \mathbb{P}(Q < u)$.

Proposition: (i) For $u \in [0, K]$,

$$1 - \pi_K(u) = \mathbb{P}(X_{\tau[y-K,y)} \le y),$$

where $\tau[u, v) := \inf\{t \ge 0 : X_t \notin [u, v)\}$, for $u \le 0 \le v$.

(ii) Let $X \in \mathscr{S}_-$. Then, for $u \in [0, K]$,

$$1 - \pi_K(u) = \frac{W^{(0)}(K - x)}{W^{(0)}(K)}.$$

As we know the transform of $W^{(0)}(\cdot)$, this result characterizes $\mathbb{P}(Q \ge u)$.

For the case of Brownian input, it turns out that Q has a truncated exponential distribution, as is easily checked.

Finite-buffer queues

In finite-buffer models: notion of a *loss rate*, defined by, in self-evident notation,

$$\ell^K := \mathbb{E}_{\pi_K} \bar{L}_1.$$

Proposition: If $\int_1^\infty y\Pi(\mathrm{d}y) = \infty$, then $\ell^K = \infty$, and otherwise

$$\ell^K = \frac{\mathbb{E}X_1}{K} \int_0^K x\pi_k(\mathrm{d}x) + \frac{\sigma^2}{2K} + \frac{1}{2K} \int_0^K \int_{-\infty}^{\infty} k(x,y)\Pi(\mathrm{d}y)\pi_K(\mathrm{d}x),$$

where $k(x,y) := -(x^2 + 2xy)$ for $y \leq -x$, $k(x,y) := y^2$ for $-x < y < K - x$, and $k(x,y) := 2y(K-x) - (K-x)^2$ for $y \geq K - x$.

Finite-buffer queues

For $X \in \mathscr{L}$ possible to find asymptotics of ℓ^K for K large.

Are of the form $Ce^{-\omega K}$, for some rather complicated C, and ω solving $\mathbb{E}e^{\omega X_1} = 1$.

Models with feedback

So far: input stream was *not* affected by the current level of the workload.

Now we *do* allow such dependencies.

Models with feedback

First model: input is

$$\mathbb{C}\mathrm{P}(r(x), \lambda(x), b(\cdot))$$

when the current workload level is $x \geq 0$; note that the distribution of the jobs B does *not* depend on x.

Rate conservation argument: density $f_Q(\cdot)$ of stationary workload obeys

$$r(x) f_Q(x) = \int_{(0,x)} \lambda(x) f_Q(y) \mathbb{P}(B > x - y) \mathrm{d}y + \lambda(0) p_0 \mathbb{P}(B > x),$$

with $p_0 := \mathbb{P}(Q = 0)$.

Models with feedback

Special case that jobs have an exponential distribution with mean $1/\mu$:

after multiplication with $e^{\mu x}$ we get the differential equation

$$g'(x) = g(x)\lambda(x)/r(x),$$

with $g(x) := e^{\mu x} r(x) f_Q(x)$.

For the case $p_0 > 0$ we obtain by an elementary separation of variables argument that

$$f_Q(x) = \frac{\lambda(0)p_0}{r(x)} \exp\left(\int_0^x \left(\frac{\lambda(y)}{r(y)} - \mu\right) \mathrm{d}y\right),$$

under appropriate integrability conditions; the case $p_0 = 0$ should be dealt with separately.

Models with feedback

Other model: queue fed by a spectrally-positive Lévy process, where feedback information about the workload level may lead to *adaptation* of the Lévy exponent.

One possiblity: models in which workload can only be observed at Poisson instants; at these Poisson instants, the Lévy exponent may be adapted based on the amount of work present at that time.

Vacation and polling models

Lévy-driven queue with server vacations is studied: stochastic storage process alternatingly experiencing active and passive (vacation) periods.

⋆ During active periods, work is generated according to Lévy process $X_{\rm D}(\cdot) \in \mathscr{S}_+$ with negative drift, until workload reaches zero (i.e., the storage reservoir is empty).

Vacation and polling models

Lévy-driven queue with server vacations is studied: stochastic storage process alternatingly experiencing active and passive (vacation) periods.

- ⋆ During active periods, work is generated according to Lévy process $X_{\rm D}(\cdot) \in \mathscr{S}_+$ with negative drift, until workload reaches zero (i.e., the storage reservoir is empty).
- ⋆ From then on, storage level behaves according to second Lévy process $X_{\rm U}(\cdot)$, assumed to be non-decreasing.

  As during this period work accumulates in the queue, it may be interpreted as a vacation; it lasts $aI + bV$, where I is a function of the length of the preceding active period, and V is an independent vacation time, and a and b are given nonnegative scalars.

  The case in which the workload is still zero after $aI + bV$, has to be treated separately: then the vacation period is extended until work is generated by $X_{\rm U}(\cdot)$.

Vacation and polling models

Lévy-driven queue with server vacations is studied: stochastic storage process alternatingly experiencing active and passive (vacation) periods.

- ⋆ During active periods, work is generated according to Lévy process $X_{\rm D}(\cdot) \in \mathscr{S}_+$ with negative drift, until workload reaches zero (i.e., the storage reservoir is empty).
- ⋆ From then on, storage level behaves according to second Lévy process $X_{\rm U}(\cdot)$, assumed to be non-decreasing.

  As during this period work accumulates in the queue, it may be interpreted as a vacation; it lasts $aI + bV$, where I is a function of the length of the preceding active period, and V is an independent vacation time, and a and b are given nonnegative scalars.

  The case in which the workload is still zero after $aI + bV$, has to be treated separately: then the vacation period is extended until work is generated by $X_{\rm U}(\cdot)$.
- ⋆ Subsequently a new active period starts; etc.

Vacation and polling models

Consider sequence of epochs right before an active period starts.

Transform of storage level at such embedded epoch can be expressed in terms of transform at previous embedded epoch.

Vacation and polling models

Consider sequence of epochs right before an active period starts.

Transform of storage level at such embedded epoch can be expressed in terms of transform at previous embedded epoch.

These transforms should be identical in equilibrium $\longrightarrow$ transform of the stationary storage level at those embedded epochs.

Vacation and polling models

Consider sequence of epochs right before an active period starts.

Transform of storage level at such embedded epoch can be expressed in terms of transform at previous embedded epoch.

These transforms should be identical in equilibrium $\longrightarrow$ transform of the stationary storage level at those embedded epochs.

Using Kella-Whitt martingale, they can be translated into transform of workload at arbitrary epoch.

Vacation and polling models

Consider sequence of epochs right before an active period starts.

Transform of storage level at such embedded epoch can be expressed in terms of transform at previous embedded epoch.

These transforms should be identical in equilibrium $\longrightarrow$ transform of the stationary storage level at those embedded epochs.

Using Kella-Whitt martingale, they can be translated into transform of workload at arbitrary epoch.

Interestingly, these vacation models can be related to so-called *polling models*, in which a single server visits multiple queues according to some predefined discipline.

Vacation and polling models

Lévy-driven polling systems can be considered in very general context:

- ⋆ N-queue polling model with switchover times;
- ⋆ Each of the queues is fed by nondecreasing Lévy process, which can be different during each of the consecutive periods within the server's cycle.
- ⋆ The N-dimensional Lévy processes obtained in this fashion are described by their (joint) Laplace exponent, thus allowing for *non-independent* input streams.

Vacation and polling models

For this general Lévy-driven polling system analysis is same as before:

⋆ First step: steady-state distribution of the workload is determined at embedded epochs (which are now polling and switching instants);
  importantly *joint* transform of all N workloads is found.

⋆ As before, application of Kella-Whitt martingale yields the steady-state distribution at arbitrary epoch.

Results are so general that they cover most important polling disciplines, like exhaustive and gated.

Models with Markov-additive input

Markov-additive processes (MAP s): Markov-modulated version of Lévy processes.

Models with Markov-additive input

A MAP (for ease only spectrally positive case $\mathscr{S}_+^{\mathrm{MAP}}$) is a bivariate Markovian process (X_t, J_t) that is defined as follows.

Models with Markov-additive input

A MAP (for ease only spectrally positive case $\mathscr{S}_+^{\mathrm{MAP}}$) is a bivariate Markovian process (X_t, J_t) that is defined as follows.

- ⋆ Let $(J_t)_t$ be irreducible continuous-time Markov chain with finite state space $E = \{1, \ldots, N\}$, transition rate matrix $Q = (q_{ij})$ and (unique) stationary distribution $\boldsymbol{\pi}$.

Models with Markov-additive input

A MAP (for ease only spectrally positive case $\mathscr{S}_+^{\mathrm{MAP}}$) is a bivariate Markovian process (X_t, J_t) that is defined as follows.

- ⋆ Let $(J_t)_t$ be irreducible continuous-time Markov chain with finite state space $E = \{1, \ldots, N\}$, transition rate matrix $Q = (q_{ij})$ and (unique) stationary distribution $\boldsymbol{\pi}$.
- ⋆ For each state i that J_t can attain, let $(X_t^{(i)})_t$ be a Lévy process with Laplace exponent

$$\varphi_i(\alpha) = \log \mathbb{E} \exp(-\alpha X_1^{(i)}).$$

Models with Markov-additive input

A MAP (for ease only spectrally positive case $\mathscr{S}_+^{\mathrm{MAP}}$) is a bivariate Markovian process (X_t, J_t) that is defined as follows.

- ⋆ Let $(J_t)_t$ be irreducible continuous-time Markov chain with finite state space $E = \{1, \ldots, N\}$, transition rate matrix $Q = (q_{ij})$ and (unique) stationary distribution $\boldsymbol{\pi}$.
- ⋆ For each state i that J_t can attain, let $(X_t^{(i)})_t$ be a Lévy process with Laplace exponent

  $$\varphi_i(\alpha) = \log \mathbb{E} \exp(-\alpha X_1^{(i)}).$$

- ⋆ Letting T_n and T_{n+1} be two successive transition epochs of J_t, and given that J_t jumps from state i to state j at $t = T_n$, we define the additive process X_t in the time interval $[T_n, T_{n+1})$ through

  $$X_t = X_{T_n-} + U_{ij}^n + [X_t^{(j)} - X_{T_n}^{(j)}],$$

  where $(U_{ij}^n)_n$ is a sequence of i.i.d. random variables with Laplace transform

  $$b_{ij}(\alpha) = \mathbb{E} e^{-\alpha U_{ij}^1},$$

  where $U_{ii}^1 \equiv 0$, describing the jumps at transition epochs.

Models with Markov-additive input

To make the MAP spectrally positive, it is required that $U^1_{ij} \geq 0$ (for all $i, j \in \{1, \ldots, N\}$) and that $X_t^{(i)}$ is allowed to have only positive jumps (for all $i \in \{1, \ldots, N\}$).

Models with Markov-additive input

Observe: modulating Markov chain does not jump in $[t, t+h)$ with probability $1 + q_{jj}h + o(h)$, given $J_t = j$ (recall that $q_{jj} < 0$), and jumps to k with probability $q_{jk}h + o(h)$.

Therefore, in self evident notation, with

$$\Xi_{ij}(\alpha, t) := \mathbb{E}_i(e^{-\alpha X_t}, J_t = j),$$

we obtain

$$\begin{aligned} \Xi_{ij}(\alpha, t+h) &= (1 + q_{jj}h)\Xi_{ij}(\alpha, t)\mathbb{E}e^{-\alpha X_h^{(j)}} + \sum_{k \neq j} q_{kj}h \cdot \Xi_{ik}(\alpha, t) b_{kj}(\alpha) + o(h) \\ &= (1 + \varphi_i(\alpha))\Xi_{ij}(\alpha, t) + h \sum_{k=1}^{N} \Xi_{ik}(\alpha, t) q_{kj} b_{kj}(\alpha) + o(h). \end{aligned}$$

Models with Markov-additive input

$$\Xi_{ij}(\alpha, t+h) = (1+\varphi_i(\alpha))\Xi_{ij}(\alpha, t) + h\sum_{k=1}^{N}\Xi_{ik}(\alpha, t)q_{kj}b_{kj}(\alpha) + o(h).$$

Subtract $\Xi_{ij}(\alpha, t)$ from both sides; divide by h: we obtain system of linear differential equations.

Models with Markov-additive input

$$\Xi_{ij}(\alpha, t+h) = (1+\varphi_i(\alpha))\Xi_{ij}(\alpha, t) + h\sum_{k=1}^{N} \Xi_{ik}(\alpha, t) q_{kj} b_{kj}(\alpha) + o(h).$$

Subtract $\Xi_{ij}(\alpha, t)$ from both sides; divide by h: we obtain system of linear differential equations.

Its solution is given in following proposition, which shows some sort of infinite-divisibility, but now at matrix level.

MAP can be regarded as a genuine matrix-counterpart of the Lévy process!

Proposition: The matrix $(\Xi_{ij}(\alpha, t))_{ij}$ equals $e^{M(\alpha)t}$, where

$$M_{ij}(\alpha) := 1_{\{i=j\}}\varphi_i(\alpha) + q_{ij} b_{ij}(\alpha).$$

Models with Markov-additive input

Just as in Lévy case: MAP-driven queues.

Stable under assumption that

$$\mathbb{E}X_1 = \sum_{i=1}^{N} \pi_i \mathbb{E}X_1^{(i)} + \sum_{i \neq j} \pi_i q_{ij} \mathbb{E}U_{ij} < 0.$$

All issues we have addressed so far for the Lévy-driven queue (stationary distribution, transience, busy periods, tail probabilities, etc.) can be studied for the MAP-driven queue as well!

Models with Markov-additive input

Now: only short account of main findings on the stationary distribution.

Under stability condition:

$$\mathbb{E}(e^{-\alpha Q}, J = j) = \left(\alpha \boldsymbol{\ell} (M(\alpha))^{-1}\right)_j ,$$

where $\boldsymbol{\ell}$ is a row vector.

Models with Markov-additive input

Now: only short account of main findings on the stationary distribution.

Under stability condition:

$$\mathbb{E}(e^{-\alpha Q}, J = j) = \left(\alpha \boldsymbol{\ell}(M(\alpha))^{-1}\right)_j ,$$

where $\boldsymbol{\ell}$ is a row vector.

Models with Markov-additive input

Now: only short account of main findings on the stationary distribution.

Under stability condition:

$$\mathbb{E}(e^{-\alpha Q}, J = j) = \left(\alpha \boldsymbol{\ell}(M(\alpha))^{-1}\right)_j,$$

where $\boldsymbol{\ell}$ is a row vector.

Interesting: compare structure of this result with 'generalized Pollaczek-Khinchine': it is essentially its MAP-counterpart!

Left: methods to determine $\boldsymbol{\ell}$. Several techniques have been developed.

Case $X \in \mathscr{S}_{-}^{\mathrm{MAP}}$ is also dealt with: then Q has phase-type distribution.

PART V:

NETWORKS

Tandem queue

Consider two *concatenated* Lévy-driven queues: a Lévy-driven tandem queue.

The output of the 1st (upstream) queue is immediately transferred to the 2nd (downstream) queue.

Let r_1 ($r_2 > 0$) be the output rates at upstream (downstream, respectively) node respectively.

In order to avoid degeneracy: assume $r_2 < r_1$.

Suppose that Lévy process J_t feeds into the first queue, with $\mathbb{E}J_1 < r_2$ (stationarity condition).

Tandem queue

Q_1, Q_2: be the stationary workload at first/second node, respectively.

Q: *total stationary* workload contained in stations $1, 2$. Note that $Q_2 = Q - Q_1$.

Tandem queue

Consider $(X_{1,t}, X_{2,t})'_{t\geq 0}$, with $X_{1,t} := J_t - r_1 t$ and $X_{2,t} := J_t - r_2 t$.

Tandem queue

Consider $(X_{1,t}, X_{2,t})'_{t\geq 0}$, with $X_{1,t} := J_t - r_1 t$ and $X_{2,t} := J_t - r_2 t$.

Then, due to 'Reich':

$$Q_1 \stackrel{\mathrm{d}}{=} \sup_{t\geq 0} X_{1,t}$$

and

$$Q \stackrel{\mathrm{d}}{=} \sup_{t\geq 0} X_{2,t}.$$

Tandem queue

Consider $(X_{1,t}, X_{2,t})'_{t\geq 0}$, with $X_{1,t} := J_t - r_1 t$ and $X_{2,t} := J_t - r_2 t$.

Then, due to 'Reich':

$$Q_1 \stackrel{\mathrm{d}}{=} \sup_{t\geq 0} X_{1,t}$$

and

$$Q \stackrel{\mathrm{d}}{=} \sup_{t\geq 0} X_{2,t}.$$

Hence following representation for the joint stationary workload holds:

$$(Q_1, Q_2) \stackrel{\mathrm{d}}{=} \left(\sup_{t\geq 0} X_{1,t}, \sup_{t\geq 0} X_{2,t} - \sup_{t\geq 0} X_{1,t} \right).$$

We are interested in the distribution of Q_2 as well as in the joint distribution.

Tandem queue

To shorten the notation, let

$$\bar{X}_{i,S} := \sup_{t \in S} X_{i,t},$$

and

$$\bar{X}_i = \bar{X}_{i,[0,\infty)}.$$

Also let

$$G_i := G_{X_i} = \arg\sup_{t \geq 0} X_{i,t}$$

be the (first) epoch that $(X_{i,t})_{t\geq 0}$ attains its maximum, for $i = 1, 2$ and $S \subset \mathbb{R}$.

Tandem: representation of downstream workload

Focus on downstream queue.

Note that it holds that Q_2 is distributed as $\sup_{t\geq 0} X_{2,t} - \sup_{t\geq 0} X_{1,t}$, but ...

Tandem: representation of downstream workload

Focus on downstream queue.

Note that it holds that Q_2 is distributed as $\sup_{t\geq 0} X_{2,t} - \sup_{t\geq 0} X_{1,t}$, but ...

$\ldots (X_{1,t})_{t\geq 0}$ and $(X_{2,t})_{t\geq 0}$ are strongly dependent (note that $X_{1,t} - X_{2,t} = (r_2 - r_1)t$).

Tandem: representation of downstream workload

Still we can find a nice representation, as follows.

Define $t_u := u/(r_1 - r_2)$, i.e., minimal time needed for second queue to exceed level u, starting empty.

Lemma: $G_1 \leq t_u \leq G_2$ a.s.

Tandem: representation of downstream workload

Lemma: $G_1 \leq t_u \leq G_2$ a.s.

Proof: two parts.

(i) $G_2 \geq t_u$. As follows:

Suppose $Q_2 > u$ and $G_2 < t_u$. Then, using $r_1 > r_2$,

$$
\begin{aligned}
Q_2 &= \sup_{t \in [0, t_u)} X_{2,t} - \sup_{s \geq 0} X_{1,s} \\
&\leq \sup_{t \in [0, t_u)} (J_t - r_2 t) - (J_t - r_2 t) = (r_1 - r_2) t_u = u.
\end{aligned}
$$

Contradiction!

Tandem: representation of downstream workload

(ii) $G_1 \leq t_u$. As follows:

Partition $\mathbb{P}(Q_2 > u)$ into

$$\mathbb{P}(\bar{X}_{2,[t_u,\infty)} - \bar{X}_{1,[0,\infty)} > u; G_1 > t_u) + \mathbb{P}(\bar{X}_{2,[t_u,\infty)} - \bar{X}_{1,[0,\infty)} > u; G_1 \leq t_u).$$

Tandem: representation of downstream workload

(ii) $G_1 \leq t_u$. As follows:

Partition $\mathbb{P}(Q_2 > u)$ into

$$\mathbb{P}(\bar{X}_{2,[t_u,\infty)} - \bar{X}_{1,[0,\infty)} > u; G_1 > t_u) + \mathbb{P}(\bar{X}_{2,[t_u,\infty)} - \bar{X}_{1,[0,\infty)} > u; G_1 \leq t_u).$$

The latter probability trivially equals

$$\mathbb{P}(\bar{X}_{2,[t_u,\infty)} - \bar{X}_{1,[0,t_u]} > u; G_1 \leq t_u).$$

Tandem: representation of downstream workload

(ii) $G_1 \leq t_u$. As follows:

Partition $\mathbb{P}(Q_2 > u)$ into

$$\mathbb{P}(\bar{X}_{2,[t_u,\infty)} - \bar{X}_{1,[0,\infty)} > u; G_1 > t_u) + \mathbb{P}(\bar{X}_{2,[t_u,\infty)} - \bar{X}_{1,[0,\infty)} > u; G_1 \leq t_u).$$

The latter probability trivially equals

$$\mathbb{P}(\bar{X}_{2,[t_u,\infty)} - \bar{X}_{1,[0,t_u]} > u; G_1 \leq t_u).$$

Considering the former probability, observe that under $G_1 > t_u$,

$$\bar{X}_{2,[t_u,\infty)} - \bar{X}_{1,[0,\infty)} \geq (J_{G_1} - r_2 G_1) - (J_{G_1} - r_1 G_1) = (r_1 - r_2)G_1 > (r_1 - r_2)t_u = u.$$

But this probability is not reduced when replacing $\bar{X}_{1,[0,\infty)}$ by $\bar{X}_{1,[0,t_u]}$:

$$\left\{\bar{X}_{2,[t_u,\infty)} - \bar{X}_{1,[0,\infty)} > u\right\} \subseteq \left\{\bar{X}_{2,[t_u,\infty)} - \bar{X}_{1,[0,t_u)} > u\right\},$$

so that the former probability equals

$$\mathbb{P}(\bar{X}_{2,[t_u,\infty)} - \bar{X}_{1,[0,t_u]} > u; G_1 > t_u).$$

Tandem: representation of downstream workload

Adding the two probabilities up yields

$$\mathbb{P}(Q_2 > u) = \mathbb{P}(\bar{X}_{2,[t_u,\infty)} - \bar{X}_{1,[0,t_u]} > u).$$

In other words: we could have taken $G_1 \leq t_u$.

Thus

$$\mathbb{P}(Q_2 > u) = \mathbb{P}\left(\bar{X}_{2,[t_u,\infty)} - \bar{X}_{1,[0,t_u]} > u\right).$$

Hence, using that $X_{1,t_u} - X_{2,t_u} = u$,

$$\bar{X_2}_{,[t_u,\infty)} - \bar{X_1}_{,[0,t_u]} = (\bar{X_2}_{,[t_u,\infty)} - X_{2,t_u}) - (\bar{X_1}_{,[0,t_u]} - X_{1,t_u}) + u.$$

Tandem: representation of downstream workload

In view of stationarity and independence of increments of $(X_{i,t})_{t\geq 0}$, we obtain:

Theorem: Let $(X_t^{(1)})_{t\geq 0}, (X_t^{(2)})_{t\geq 0}$ be independent copies of $(X_{1,t})_{t\geq 0}, (X_{2,t})_{t\geq 0}$ respectively. Then, for each $u > 0$,

$$\mathbb{P}(Q_2 > u) = \mathbb{P}\left(\sup_{t\in[0,\infty)} X_t^{(2)} > \sup_{t\in[0,t_u]} -X_t^{(1)} \right).$$

Tandem: distribution of downstream workload

Goal: find Laplace transform $\mathbb{E}e^{-\alpha Q_2}$ for $J \in \mathscr{S}_+$.

Tandem: distribution of downstream workload

Goal: find Laplace transform $\mathbb{E}e^{-\alpha Q_2}$ for $J \in \mathscr{S}_+$.

Let $\varphi_1(\alpha) := \mathbb{E}e^{-\alpha X_{1,1}}$. Also,

$$\bar{\tau}(x) := \inf\{t \geq 0 : -X_t^{(1)} > x\}.$$

Then, for each $y \geq 0$,

$$\mathbb{P}\left(\sup_{t\in[0,t_u]}(-X_t^{(1)}) < y\right) = \mathbb{P}(\bar{\tau}(y) > t_u)$$

and, as seen before,

$$\mathbb{E}e^{-\vartheta\tau(x)} = e^{-x\varphi_1^{-1}(\vartheta)}.$$

Obviously,

$$\sup_{t\in[0,\infty)} X_t^{(2)} \overset{\mathrm{d}}{=} Q.$$

Tandem: distribution of downstream workload

Application of representation of downstream workload, with $\psi_1(\cdot) := \varphi_1^{-1}(\cdot)$,

$$\begin{aligned}
\int_0^\infty e^{-\alpha u}\mathbb{P}(Q_2 > u)\mathrm{d}u &= \int_0^\infty e^{-\alpha u} \int_0^\infty \mathbb{P}(\bar{\tau}(y) > t_u)\mathrm{d}\mathbb{P}(Q \le y)\mathrm{d}u \\
&= (r_1 - r_2)\int_0^\infty \int_0^\infty e^{-\alpha(r_1-r_2)v}\mathbb{P}(\bar{\tau}(y) > v)\mathrm{d}v\mathrm{d}\mathbb{P}(Q \le y) \\
&= \frac{1}{\alpha}\left(1 - \int_0^\infty \int_0^\infty e^{-\alpha(r_1-r_2)v}\mathrm{d}\mathbb{P}(R_y \le v)\mathrm{d}\mathbb{P}(Q \le y)\right) \\
&= \frac{1}{\alpha}\left(1 - \mathbb{E}e^{-\psi_1(\alpha(r_1-r_2))Q}\right).
\end{aligned}$$

Tandem: distribution of downstream workload

As a consequence:

$$\mathbb{E}e^{-\alpha Q_2} = \mathbb{E}e^{-\psi_1(\alpha(r_1 - r_2))Q},$$

which, combined with 'generalized Pollaczek-Khinchine', gives the following result.

Tandem: distribution of downstream workload

As a consequence:

$$\mathbb{E}e^{-\alpha Q_2} = \mathbb{E}e^{-\psi_1(\alpha(r_1-r_2))Q},$$

which, combined with 'generalized Pollaczek-Khinchine', gives the following result.

Theorem: Let $J \in \mathscr{S}_+$ with $\mathbb{E}J_1 < r_2 < r_1$. Then, for each $\alpha > 0$,

$$\mathbb{E}e^{-\alpha Q_2} = \frac{-\mathbb{E}X_{2,1}}{r_1 - r_2} \frac{\psi_1(\alpha(r_1 - r_2))}{\alpha - \psi_1(\alpha(r_1 - r_2))}.$$

Tandem: distribution of downstream workload

Suppose $J \in \mathbb{B}\mathrm{m}(0,1)$.

Then density of $\sup_{t\in[0,t_u]} -X_t^{(1)}$ equals

$$\begin{aligned}\varrho(x) &:= \frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}\left(\sup_{t\in[0,t_u]} -X_t^{(1)}\} \leq x\right) \\ &= \sqrt{\frac{2}{\pi t_u}}\exp\left(-\frac{(x-r_1t_u)^2}{2t_u}\right) - 2r_1e^{2r_1x}\left(1-\Phi_{\mathrm{N}}\left(\frac{x+r_1t_u}{\sqrt{t_u}}\right)\right).\end{aligned}$$

After some standard calculus, for each $u \geq 0$,

$$\begin{aligned}&\mathbb{P}(Q_2 > u) = \\ &= \frac{r_1-2r_2}{r_1-r_2}e^{-2r_2u}\Phi_{\mathrm{N}}\left(\frac{r_1-2r_2}{\sqrt{r_1-r_2}}\sqrt{u}\right) + \frac{r_1}{r_1-r_2}\left(1-\Phi_{\mathrm{N}}\left(\frac{r_1}{\sqrt{r_1-r_2}}\sqrt{u}\right)\right).\end{aligned}$$

Tandem: distribution of downstream workload

Suppose $J \in \mathbb{B}\mathrm{m}(0,1)$.

After lengthy but standard calculus, we obtain following asymptotics, as $u \to \infty$:

(i) if $r_1 > 2r_2$, then

$$\mathbb{P}(Q_2 > u)e^{2r_2 u} \to \frac{r_1 - 2r_2}{r_1 - r_2};$$

(ii) if $r_1 = 2r_2$, then

$$\mathbb{P}(Q_2 > u)\sqrt{u}e^{2r_2 u} \to \frac{1}{\sqrt{2\pi r_2}};$$

(iii) if $r_1 < 2r_2$, then

$$\mathbb{P}(Q_2 > u)\left(\frac{u}{r_1 - r_2}\right)^{3/2} \exp\left(\frac{r_1^2}{2(r_1 - r_2)}u\right) \to \frac{1}{\sqrt{2\pi}} \frac{4r_2}{r_1^2(r_1 - 2r_2)^2}.$$

These extend to $J \in \mathscr{L} \cap \mathscr{S}_+$ ('Heaviside').

Tandem: distribution of downstream workload

Applying 'Tauber' to the transform of Q_2:

Theorem: Assume that $X_1 \in \mathscr{S}_+ \cap \mathscr{R}$, with $\varphi_1(\alpha) \in \mathscr{R}_\nu(n, \eta)$. Then, as $u \to \infty$,

$$\mathbb{P}(Q_2 > u) = \left(\frac{-\mathbb{E}X_{1,1}}{r_1 - r_2} \right)^{1-\nu} \mathbb{P}(Q > u)(1 + o(1)) =$$

$$= \frac{(-1)^{n+1}}{\Gamma(2-\nu)} \frac{\eta}{-\mathbb{E}X_{2,1}} \left(\frac{-\mathbb{E}X_{1,1}}{r_1 - r_2} \right)^{1-\nu} u^{1-\nu} L(u)(1 + o(1)).$$

Tandem: distribution of downstream workload

Extensions:

Tandem: distribution of downstream workload

Extensions:

I. Joint distribution:

$$\mathbb{E}e^{-\alpha Q_1-\bar{\alpha}Q_2}=\frac{-\mathbb{E}X_{2,1}\bar{\alpha}}{\bar{\alpha}-\psi_1((r_1-r_2)\bar{\alpha})}\frac{\psi_1((r_1-r_2)\bar{\alpha})-\alpha}{(r_1-r_2)\bar{\alpha}-\varphi_1(\alpha)}.$$

Tandem: distribution of downstream workload

Extensions:

I. Joint distribution:

$$\mathbb{E}e^{-\alpha Q_1 - \bar{\alpha} Q_2} = \frac{-\mathbb{E}X_{2,1}\bar{\alpha}}{\bar{\alpha} - \psi_1((r_1 - r_2)\bar{\alpha})} \frac{\psi_1((r_1 - r_2)\bar{\alpha}) - \alpha}{(r_1 - r_2)\bar{\alpha} - \varphi_1(\alpha)}.$$

II. Bivariate asymptotics:

$$\mathbb{P}(Q_1 > Au, Q_2 > (1 - A)u)$$

as $u \to \infty$ and $A \in (0, 1)$.

Tandem: distribution of downstream workload

Extensions:

I. Joint distribution:

$$\mathbb{E}e^{-\alpha Q_1-\bar{\alpha}Q_2} = \frac{-\mathbb{E}X_{2,1}\bar{\alpha}}{\bar{\alpha}-\psi_1((r_1-r_2)\bar{\alpha})}\frac{\psi_1((r_1-r_2)\bar{\alpha})-\alpha}{(r_1-r_2)\bar{\alpha}-\varphi_1(\alpha)}.$$

II. Bivariate asymptotics:

$$\mathbb{P}(Q_1 > Au, Q_2 > (1-A)u)$$

as $u \to \infty$ and $A \in (0,1)$.

III. More sophisticated systems: multihop tandems and intree networks.

EPILOGUE

A few conclusions

- ⋆ Lévy-driven queues are a practically relevant concept;
- ⋆ fairly explicit analysis is possible;
- ⋆ a broad variety of techniques can be used (transforms, rate conservation, asymptotic techniques, importance sampling, martingales, . . .