Structure of Markov jump processes

Infinitesimal Generator

$\forall x, y,\ y \neq x \in E$, limits $q_x := \lim_{t\to 0} \frac{1-p_{xx}(t)}{t}$, $q_{xy} = \lim_{t\to 0} \frac{p_{xy}(t)}{t}$ exist in $\mathbb{R}_+$ and satisfy $\sum_{y\neq x} q_{xy} = q_x$

q_{xy}: **Jump rate** from x to y

$Q := \{q_{xy}\}_{x,y\in E}$ where $q_{xx} = -q_x$: **Infinitesimal Generator** of process $\{X_t\}_{t\in\mathbb{R}_+}$

Formally: $Q = \lim_{h\to 0} \frac{1}{h}[P(h) - I]$ where I: identity matrix

Structure of Markov jump processes

Infinitesimal Generator

$\forall x, y,\ y \neq x \in E$, limits $q_x := \lim_{t\to 0} \frac{1-p_{xx}(t)}{t}$, $q_{xy} = \lim_{t\to 0} \frac{p_{xy}(t)}{t}$ exist in $\mathbb{R}_+$ and satisfy $\sum_{y\neq x} q_{xy} = q_x$

q_{xy}: **Jump rate** from x to y

$Q := \{q_{xy}\}_{x,y\in E}$ where $q_{xx} = -q_x$: **Infinitesimal Generator** of process $\{X_t\}_{t\in\mathbb{R}_+}$

Formally: $Q = \lim_{h\to 0} \frac{1}{h}[P(h) - I]$ where I: identity matrix

Structure of Markov jump processes

Sequence $\{Y_n\}_{n\in\mathbb{N}}$ of visited states: Markov chain with transition matrix $p_{xy} = \mathbb{I}_{x\neq y} \frac{q_{xy}}{q_x}$

Conditionally on $\{Y_n\}_{n\in\mathbb{N}}$, sojourn times $\{\tau_n\}_{n\in\mathbb{N}}$ in successive states Y_n: independent, with distributions $\mathrm{Exp}(q_{Y_n})$

Examples

- for Poisson process (λ): only non-zero jump rate $q_{x,x+1} = \lambda = q_x, \; x \in \mathbb{N}$

Examples

- for Poisson process (λ): only non-zero jump rate $q_{x,x+1} = \lambda = q_x,\ x \in \mathbb{N}$
- For continuous-time random walk on graph $G = (V, E)$, non-zero rates: $q_{i,j} = \mathbb{I}_{i \sim j}$, hence $q_i = d_i$. Generator Q is opposite of so-called Laplacian matrix $L(G)$ of graph G

Examples

- for Poisson process (λ): only non-zero jump rate $q_{x,x+1} = \lambda = q_x, \ x \in \mathbb{N}$

- For continuous-time random walk on graph $G = (V, E)$, non-zero rates: $q_{i,j} = \mathbb{I}_{i \sim j}$, hence $q_i = d_i$. Generator Q is opposite of so-called Laplacian matrix $L(G)$ of graph G

- For FIFO $M/M/1/\infty$ queue, non-zero rates: $q_{x,x+1} = \lambda$, $q_{x,x-1} = \mu \mathbb{I}_{x>0}, \ x \in \mathbb{N}$ hence $q_x = \lambda + \mu \mathbb{I}_{x>0}$

Examples

- for Poisson process (λ): only non-zero jump rate $q_{x,x+1} = \lambda = q_x,\ x \in \mathbb{N}$
- For continuous-time random walk on graph $G = (V, E)$, non-zero rates: $q_{i,j} = \mathbb{I}_{i \sim j}$, hence $q_i = d_i$. Generator Q is opposite of so-called Laplacian matrix $L(G)$ of graph G
- For FIFO $M/M/1/\infty$ queue, non-zero rates: $q_{x,x+1} = \lambda$, $q_{x,x-1} = \mu \mathbb{I}_{x>0},\ x \in \mathbb{N}$ hence $q_x = \lambda + \mu \mathbb{I}_{x>0}$
- For $M/M/\infty/\infty$ queue, non-zero rates: $q_{x,x+1} = \lambda$, $q_{x,x-1} = \mu x,\ x \in \mathbb{N}$ hence $q_x = \lambda + \mu x$

Structure of Markov jump processes (continued)

Let $T_n := \sum_{k=0}^{n-1} \tau_k$: time of n-th jump.

If $T_\infty = +\infty$ almost surely: trajectory determined on $\mathbb{R}_+$, hence generator Q determines law of process $\{X_t\}_{t\in\mathbb{R}_+}$

Structure of Markov jump processes (continued)

Let $T_n := \sum_{k=0}^{n-1} \tau_k$: time of n-th jump.

If $T_\infty = +\infty$ almost surely: trajectory determined on $\mathbb{R}_+$, hence generator Q determines law of process $\{X_t\}_{t \in \mathbb{R}_+}$

Process is called **explosive** if instead $T_\infty < +\infty$ with positive probability. Then process not completely characterized by generator

Structure of Markov jump processes (continued)

Let $T_n := \sum_{k=0}^{n-1} \tau_k$: time of n-th jump.

If $T_\infty = +\infty$ almost surely: trajectory determined on $\mathbb{R}_+$, hence generator Q determines law of process $\{X_t\}_{t\in\mathbb{R}_+}$

Process is called **explosive** if instead $T_\infty < +\infty$ with positive probability. Then process not completely characterized by generator

Sufficient conditions for non-explosiveness:

- $\sup_{x\in E} q_x < +\infty$
- Recurrence of induced chain $\{Y_n\}_{n\in\mathbb{N}}$
- For **Birth and Death** processes (i.e. $E = \mathbb{N}$, only non-zero rates: $\beta_n = q_{n,n+1}$, birth rate; $\delta_n = q_{n,n-1}$, death rate), non-explosiveness holds if
$$\sum_{n>0} \frac{1}{\beta_n + \delta_n} = +\infty$$

Kolmogorov's forward and backward equations

Formal differentiation of $P(t+h) = P(t)P(h) = P(h)P(t)$ yields

$\frac{d}{dt}P(t) = P(t)Q$ Kolmogorov's forward equation

$\frac{d}{dt}p_{xy}(t) = \sum_{z\in E} p_{xz}(t)q_{zy}$

$\frac{d}{dt}P(t) = QP(t)$ Kolmogorov's backward equation

$\frac{d}{dt}p_{xy}(t) = \sum_{z\in E} q_{xz}p_{zy}(t)$

Kolmogorov's forward and backward equations

Formal differentiation of $P(t+h) = P(t)P(h) = P(h)P(t)$ yields

$$\frac{d}{dt}P(t) = P(t)Q \qquad \text{Kolmogorov's forward equation}$$
$$\frac{d}{dt}p_{xy}(t) = \sum_{z\in E} p_{xz}(t)q_{zy}$$

$$\frac{d}{dt}P(t) = QP(t) \qquad \text{Kolmogorov's backward equation}$$
$$\frac{d}{dt}p_{xy}(t) = \sum_{z\in E} q_{xz}p_{zy}(t)$$

Follow directly from $Q = \lim_{h\to 0}\frac{1}{h}[P(h) - I]$ for finite E, in which case $P(t) = \exp(tQ),\ t \geq 0$

Hold more generally–in particular for non-explosive processes–with a more involved proof (justifying exchange of summation and differentiation)

Stationary distributions and measures

Definition

Measure $\{\pi_x\}_{x\in E}$ is **stationary** if it satisfies $\pi^T Q = 0$, or equivalently the **global balance equations**

$$\forall x \in E,\ \underbrace{\pi_x \sum_{y\neq x} q_{xy}}_{\text{flow out of } x} = \underbrace{\sum_{y\neq x} \pi_y q_{yx}}_{\text{flow into } x}$$

Stationary distributions and measures

Definition

Measure $\{\pi_x\}_{x\in E}$ is **stationary** if it satisfies $\pi^T Q = 0$, or equivalently the **global balance equations**

$$\forall x \in E,\ \underbrace{\pi_x \sum_{y\neq x} q_{xy}}_{\text{flow out of } x} = \underbrace{\sum_{y\neq x} \pi_y q_{yx}}_{\text{flow into } x}$$

Kolmogorov's equations suggest that, if $X_0 \sim \pi$ for stationary π then $X_t \sim \pi$ for all $t \geq 0$

Stationary distributions and measures

Definition

Measure $\{\pi_x\}_{x\in E}$ is **stationary** if it satisfies $\pi^T Q = 0$, or equivalently the **global balance equations**

$$\forall x \in E,\ \underbrace{\pi_x \sum_{y\neq x} q_{xy}}_{\text{flow out of } x} = \underbrace{\sum_{y\neq x} \pi_y q_{yx}}_{\text{flow into } x}$$

Kolmogorov's equations suggest that, if $X_0 \sim \pi$ for stationary π then $X_t \sim \pi$ for all $t \geq 0$

EXAMPLE: stationarity for birth and death processes

$$\begin{aligned}\pi_0\beta_0 &= \pi_1\delta_1,\\ \pi_x(\beta_x+\delta_x) &= \pi_{x-1}\beta_{x-1}+\pi_{x+1}\delta_{x+1},\ x\geq 1\end{aligned}$$

Irreducibility, recurrence, invariance

Definition

- Process $\{X_t\}_{t\in\mathbb{R}_+}$ is **irreducible** (respectively, **irreducible recurrent**) if induced chain $\{Y_n\}_{n\in\mathbb{N}}$ is.

- State x is **positive recurrent** if $\mathbb{E}_x(R_x) < +\infty$, where

$$R_x = \inf\{t > \tau_0 : X_t = x\}.$$

- Measure π is **invariant** for process $\{X_t\}_{t\in\mathbb{R}_+}$ if for all $t > 0$, $\pi^T P(t) = \pi^T$, i.e.

$$\forall x \in E, \sum_{y\in E} \pi_y p_{yx}(t) = \pi_x.$$

Limit theorems 1

Theorem

For irreducible recurrent $\{X_t\}_{t\in\mathbb{R}_+}$, $\exists$ invariant measure π, unique up to some scalar factor. It can be defined as, for any $x \in E$:

$$\forall y \in E,\ \pi_y = \mathbb{E}_x \int_0^{R_x} \mathbb{I}_{X_t=y} dt,$$

or alternatively with $T_x := \inf\{n > 0 : Y_n = x\}$,

$$\forall y \in E,\ \pi_y = \frac{1}{q_y}\mathbb{E}_x \sum_{n=1}^{T_x} \mathbb{I}_{Y_n=y}.$$

Limit theorems 1

Theorem

For irreducible recurrent $\{X_t\}_{t\in\mathbb{R}_+}$, $\exists$ invariant measure π, unique up to some scalar factor. It can be defined as, for any $x \in E$:

$$\forall y \in E,\ \pi_y = \mathbb{E}_x \int_0^{R_x} \mathbb{I}_{X_t=y}dt,$$

or alternatively with $T_x := \inf\{n > 0 : Y_n = x\}$,

$$\forall y \in E,\ \pi_y = \frac{1}{q_y}\mathbb{E}_x \sum_{n=1}^{T_x} \mathbb{I}_{Y_n=y}.$$

Corollaries

- $\{\hat{\pi}_y\}$ invariant for $\{Y_n\}_{n\in\mathbb{N}} \Leftrightarrow \{\hat{\pi}_y/q_y\}$ invariant for $\{X_t\}_{t\in\mathbb{R}_+}$. Thus invariance $\Leftrightarrow$ stationarity.
- For irreducible recurrent $\{X_t\}_{t\in\mathbb{R}_+}$, either all or no state $x \in E$ is positive recurrent.

Limit theorems 2

Theorem

$\{X_t\}_{t\in\mathbb{R}_+}$ is **ergodic** *(i.e. irreducible, positive recurrent) iff it is irreducible, non-explosive and such that $\exists\pi$, probability distribution satisfying global balance equations.*
Then π is also the unique invariant probability distribution.

Limit theorems 2

Theorem

*$\{X_t\}_{t\in\mathbb{R}_+}$ is **ergodic** (i.e. irreducible, positive recurrent) iff it is irreducible, non-explosive and such that $\exists\pi$, probability distribution satisfying global balance equations.*
Then π is also the unique invariant probability distribution.

Theorem

For ergodic $\{X_t\}_{t\in\mathbb{R}_+}$ with stationary distribution π, any initial distribution for X_0 and π-integrable f,

$$\text{almost surely } \lim_{t\to\infty}\frac{1}{t}\int_0^t f(X_s)ds = \sum_{x\in E}\pi_x f(x) \quad \text{(ergodic theorem)}$$

and in distribution $X_t \xrightarrow{\mathcal{D}} \pi$ as $t\to\infty$.

Limit theorems 3

Theorem

For irreducible, non-ergodic $\{X_t\}_{t\in\mathbb{R}_+}$, any initial distribution for X_0, then for all $x \in E$,

$$\lim_{t\to\infty} \mathbb{P}(X_t = x) = 0.$$

Time reversal and reversibility

For stationary ergodic $\{X_t\}_{t\in\mathbb{R}}$ with stationary distribution π,
time-reversed process $\tilde{X}_t = X_{-t}$:
Markov with transition rates $\tilde{q}_{xy} = \frac{\pi_y q_{yx}}{\pi_x}$

Time reversal and reversibility

For stationary ergodic $\{X_t\}_{t\in\mathbb{R}}$ with stationary distribution π,
time-reversed process $\tilde{X}_t = X_{-t}$:
Markov with transition rates $\tilde{q}_{xy} = \frac{\pi_y q_{yx}}{\pi_x}$

Definition

Stationary ergodic $\{X_t\}_{t\in\mathbb{R}}$ with stationary distribution π **reversible** iff distributed as time-reversal $\{\tilde{X}_t\}_{t\in\mathbb{R}}$, i.e.

$$\forall x \neq y \in E, \qquad \underbrace{\pi_x q_{xy}}_{\text{flow from } x \text{ to } y} = \underbrace{\pi_y q_{yx}}_{\text{flow from } y \text{ to } x},$$

detailed balance equations.

Time reversal and reversibility

For stationary ergodic $\{X_t\}_{t\in\mathbb{R}}$ with stationary distribution π,
time-reversed process $\tilde{X}_t = X_{-t}$:
Markov with transition rates $\tilde{q}_{xy} = \frac{\pi_y q_{yx}}{\pi_x}$

Definition

Stationary ergodic $\{X_t\}_{t\in\mathbb{R}}$ with stationary distribution π **reversible** iff distributed as time-reversal $\{\tilde{X}_t\}_{t\in\mathbb{R}}$, i.e.

$$\forall x \neq y \in E, \qquad \underbrace{\pi_x q_{xy}}_{\text{flow from } x \text{ to } y} = \underbrace{\pi_y q_{yx}}_{\text{flow from } y \text{ to } x},$$

detailed balance equations.

Detailed balance, i.e. reversibility for π implies global balance for π.
EXAMPLE: for birth and death processes, detailed balance always holds for stationary measure.

Reversibility and truncation

Proposition

Let generator Q on E admit reversible measure π. Then for subset $F \subset E$, **truncated generator** $\hat{Q}$:

$$\begin{aligned} \hat{Q}_{xy} &= Q_{xy},\ x \neq y \in F, \\ \hat{Q}_{xx} &= -\sum_{y \neq x} \hat{Q}_{xy},\ x \in F \end{aligned}$$

admits $\{\pi_x\}_{x \in F}$ as reversible measure.

Erlang's model of telephone network

- Call types $s \in \mathcal{S}$: type-s calls arrive at instants of Poisson (λ_s) process, last (if accepted) for duration Exponential (μ_s)
- type-s calls require one circuit (unit of capacity) per link $\ell \in s$
- Link ℓ has capacity C_ℓ circuits

Erlang's model of telephone network

- Call types $s \in \mathcal{S}$: type-s calls arrive at instants of Poisson (λ_s) process, last (if accepted) for duration Exponential (μ_s)
- type-s calls require one circuit (unit of capacity) per link $\ell \in s$
- Link ℓ has capacity C_ℓ circuits

Stationary probability distribution:

$$\pi_x = \frac{1}{Z} \prod_{s \in \mathcal{S}} \frac{\rho_s^{x_s}}{x_s!} \prod_\ell \mathbb{I}_{\sum_{s \ni \ell} x_s \leq C_\ell},$$

where: $\rho_s = \lambda_s / \mu_s$, Z: normalizing constant.

Erlang's model of telephone network

- Call types $s \in \mathcal{S}$: type-s calls arrive at instants of Poisson (λ_s) process, last (if accepted) for duration Exponential (μ_s)
- type-s calls require one circuit (unit of capacity) per link $\ell \in s$
- Link ℓ has capacity C_ℓ circuits

Stationary probability distribution:

$$\pi_x = \frac{1}{Z} \prod_{s \in \mathcal{S}} \frac{\rho_s^{x_s}}{x_s!} \prod_\ell \mathbb{I}_{\sum_{s \ni \ell} x_s \leq C_\ell},$$

where: $\rho_s = \lambda_s / \mu_s$, Z: normalizing constant.

Basis for dimensioning studies of telephone networks (prediction of call rejection probabilities)
More recent application: performance analysis of peer-to-peer systems for video streaming.

Jackson networks

- Stations $i \in I$ receive external arrivals at Poisson rate $\overline{\lambda}_i$
- Station i when processing x_i customers completes service at rate $\mu_i \phi_i(x_i)$ (e.g.: $\phi_i(x) = \min(x_i, n_i)$: queue with n_i servers and service times Exponential (μ_i))
- After completing service at station i, customer joins station j with probability $p_{ij}, j \in I$, and leaves system with probability $1 - \sum_{j \in I} p_{ij}$
- Matrix $P = (p_{ij})$: sub-stochastic, such that $\exists (I - P)^{-1}$

Jackson networks

- Stations $i \in I$ receive external arrivals at Poisson rate $\overline{\lambda}_i$
- Station i when processing x_i customers completes service at rate $\mu_i \phi_i(x_i)$ (e.g.: $\phi_i(x) = \min(x_i, n_i)$: queue with n_i servers and service times Exponential (μ_i))
- After completing service at station i, customer joins station j with probability $p_{ij}, j \in I$, and leaves system with probability $1 - \sum_{j \in I} p_{ij}$
- Matrix $P = (p_{ij})$: sub-stochastic, such that $\exists (I - P)^{-1}$

Traffic equations

$$\forall i \in I, \ \lambda_i = \overline{\lambda}_i + \sum_{j \in I} \lambda_j p_{ji}$$

or $\lambda = (I - P^T)^{-1} \overline{\lambda}$

Jackson networks (continued)

Stationary measure:

$$\pi_x = \prod_{i \in I} \frac{\rho_i^{x_i}}{\prod_{m=1}^{x_i} \phi_i(m)},$$

where $\rho_i = \lambda_i / \mu_i$, and λ_i: solutions of traffic equations

Jackson networks (continued)

Stationary measure:

$$\pi_x = \prod_{i \in I} \frac{\rho_i^{x_i}}{\prod_{m=1}^{x_i} \phi_i(m)},$$

where $\rho_i = \lambda_i / \mu_i$, and λ_i: solutions of traffic equations Application: process ergodic when π has finite mass. e.g. for $\phi_i(x) = \min(x, n_i)$, ergodicity iff $\forall i \in I,\ \rho_i < n_i$.

Jackson networks (continued)

Stationary measure:

$$\pi_x = \prod_{i \in I} \frac{\rho_i^{x_i}}{\prod_{m=1}^{x_i} \phi_i(m)},$$

where $\rho_i = \lambda_i/\mu_i$, and λ_i: solutions of traffic equations Application: process ergodic when π has finite mass. e.g. for $\phi_i(x) = \min(x, n_i)$, ergodicity iff $\forall i \in I,\ \rho_i < n_i$.

Proof: verify **partial balance** equations for all $x \in \mathbb{N}^I$:

$$\begin{array}{ll} \forall i \in I, & \\ \pi_x[\sum_{j \neq i} q_{x,x-e_i+e_j} + q_{x,x-e_i}] & = \sum_{j \neq i} \pi_{x-e_i+e_j} q_{x-e_i+e_j,x} + \pi_{x-e_i} q_{x-e_i,x}, \\ \pi_x \sum_{i \in I} q_{x,x+e_i} & = \sum_{i \in I} \pi_{x+e_i} q_{x+e_i,x}, \end{array}$$

which imply global balance equations

$$\pi_x[\sum_{i \in I}(q_{x,x-e_i} + q_{x,x+e_i} + \sum_{j \neq i} q_{x,x-e_i+e_j})] = \\ \sum_{i \in I}(\pi_{x-e_i} q_{x-e_i,x} + \pi_{x+e_i} q_{x+e_i,x} + \sum_{j \neq i} \pi_{x-e_i+e_j} q_{x-e_i+e_j,x})$$

Foster-Lyapunov criterion – continuous time

Theorem

Assume (i) Process $\{X_t\}_{t\in\mathbb{R}_+}$ *irreducible non explosive;*

(ii) There is a function $V : E \to \mathbb{R}_+$, *a finite set* $K \subset E$ *and constants* $b, \epsilon > 0$ *such that*

$$\forall x \in E, \sum_{y \neq x} q_{xy}[V(y) - V(x)] \leq -\epsilon + b\mathbb{I}_{x\in K}.$$

Then $\{X_t\}_{t\in\mathbb{R}_+}$ *is ergodic.*

Proof steps

- Induced chain $\{Y_n\}_{n\in\mathbb{N}}$ such that

$$\mathbb{E}(V(Y_{n+1}) - V(Y_n)|Y_n = x) \leq -\frac{\epsilon}{q_x} + \frac{b}{q_x}\mathbb{I}_{x\in K}$$

Proof steps

- Induced chain $\{Y_n\}_{n\in\mathbb{N}}$ such that

$$\mathbb{E}(V(Y_{n+1}) - V(Y_n)|Y_n = x) \leq -\frac{\epsilon}{q_x} + \frac{b}{q_x}\mathbb{I}_{x\in K}$$

- Hence for $N := \inf\{n > 0 : Y_n \in K\}$,

$$\mathbb{E}[V(Y_N) - V(Y_0)|Y_0 = x] \leq -\epsilon\mathbb{E}\left[\sum_{n=0}^{N-1}\frac{1}{q_{Y_n}}|Y_0 = x\right] + \frac{b}{q_x}$$

Proof steps

- Induced chain $\{Y_n\}_{n\in\mathbb{N}}$ such that

$$\mathbb{E}(V(Y_{n+1}) - V(Y_n)|Y_n = x) \leq -\frac{\epsilon}{q_x} + \frac{b}{q_x}\mathbb{I}_{x\in K}$$

- Hence for $N := \inf\{n > 0 : Y_n \in K\}$,

$$\mathbb{E}[V(Y_N) - V(Y_0)|Y_0 = x] \leq -\epsilon\mathbb{E}\left[\sum_{n=0}^{N-1}\frac{1}{q_{Y_n}}|Y_0 = x\right] + \frac{b}{q_x}$$

- Yields, letting $R(K) := \inf\{t > \tau_0 : X_t \in K\} = T_N$, return time to set K,

$$\mathbb{E}[R(K)|X_0 = x] \leq \frac{1}{\epsilon}\left[V(x) + \frac{b}{q_x}\right]$$

Proof steps

- Induced chain $\{Y_n\}_{n\in\mathbb{N}}$ such that

$$\mathbb{E}(V(Y_{n+1})-V(Y_n)|Y_n=x)\leq -\frac{\epsilon}{q_x}+\frac{b}{q_x}\mathbb{I}_{x\in K}$$

- Hence for $N:=\inf\{n>0:Y_n\in K\}$,

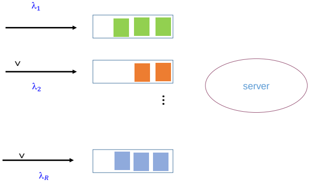
$$\mathbb{E}[V(Y_N)-V(Y_0)|Y_0=x]\leq -\epsilon\mathbb{E}\left[\sum_{n=0}^{N-1}\frac{1}{q_{Y_n}}|Y_0=x\right]+\frac{b}{q_x}$$

- Yields, letting $R(K):=\inf\{t>\tau_0:X_t\in K\}=T_N$, return time to set K,

$$\mathbb{E}[R(K)|X_0=x]\leq\frac{1}{\epsilon}\left[V(x)+\frac{b}{q_x}\right]$$

- Implies, reasoning on chain $\{Z_n\}_{n\in\mathbb{N}}$ of visits of $\{Y_n\}_{n\in\mathbb{N}}$ to set K, that $\mathbb{E}_x(R_x)<+\infty$ for all $x\in K$, hence ergodicity

Priority queue



- R customer types
- Infinite queue, single server with unit capacity
- Policy: always serve customer with highest priority (lowest class index)
  Interrupt lower priority service upon higher priority arrival
  Resume interrupted service where it was stopped (FIFO per class)
- Poisson λ_r arrivals in class r; Exponential μ_r service times
  Loads: $\rho_r := \lambda_r/\mu_r$

- $X_r(t)$: number of class-r customers present at time t
- A Markov jump process with only non-zero rates

$$q_{x,x+e_r} = \lambda_r, \quad q_{x,x-e_r} = \mu_r \mathbb{I}_{x_r>0}\mathbb{I}_{x_1=\cdots=x_{r-1}=0}$$

Proposition

Process ergodic if $\rho := \sum_r \rho_r < 1$, transient if $\rho > 1$

Assume $\mu_r \equiv \mu$ and ergodicity.
Then mean number of customers at equilibrium:

$$\mathbb{E}(X_r) = \frac{\rho_r}{(1-\sum_{s<r}\rho_s)(1-\sum_{s\leq r}\rho_s)}$$

Proof elements – priority queue

- Process clearly irreducible non-explosive
  Foster's criterion with $V(x) := \sum_r x_r/\mu_r \Rightarrow$ ergodic if $\rho < 1$

Proof elements – priority queue

- Process clearly irreducible non-explosive
  Foster's criterion with $V(x) := \sum_r x_r/\mu_r \Rightarrow$ ergodic if $\rho < 1$
- $\rho > 1$: with $X(0) = 0$, $X_r(t) = N_r(t) - D_r(t)$ (arrivals minus departures)
  Work spent on class r by time t: $W_r(t) \geq \sum_{m=1}^{D_r(t)} \sigma_{r,m}$
  for i.i.d. service times $\sigma_{r,m}$

Proof elements – priority queue

- Process clearly irreducible non-explosive
  Foster's criterion with $V(x) := \sum_r x_r/\mu_r \Rightarrow$ ergodic if $\rho < 1$
- $\rho > 1$: with $X(0) = 0$, $X_r(t) = N_r(t) - D_r(t)$ (arrivals minus departures)
  Work spent on class r by time t: $W_r(t) \geq \sum_{m=1}^{D_r(t)} \sigma_{r,m}$
  for i.i.d. service times $\sigma_{r,m}$
- Law of large numbers for Poisson processes: almost surely, $\lim_{t\to\infty} N_r(t)/t = \lambda_r$

Proof elements – priority queue

- Process clearly irreducible non-explosive
  Foster's criterion with $V(x) := \sum_r x_r/\mu_r \Rightarrow$ ergodic if $\rho < 1$
- $\rho > 1$: with $X(0) = 0$, $X_r(t) = N_r(t) - D_r(t)$ (arrivals minus departures)
  Work spent on class r by time t: $W_r(t) \geq \sum_{m=1}^{D_r(t)} \sigma_{r,m}$
  for i.i.d. service times $\sigma_{r,m}$
- Law of large numbers for Poisson processes: almost surely, $\lim_{t\to\infty} N_r(t)/t = \lambda_r$
- If for some r, $D_r(t) \leq \lambda_r t/2$ then $X_r(t) \geq \lambda_r t/2 + o(t)$
  Else, by Law of large numbers for $\sigma_{r,m}$, $\forall r$, $W_r(t) \geq D_r(t)/\mu_r + o(t)$

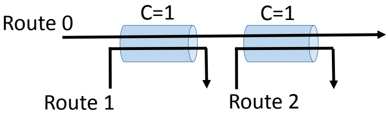
Proof elements – priority queue

- Process clearly irreducible non-explosive
  Foster's criterion with $V(x) := \sum_r x_r/\mu_r \Rightarrow$ ergodic if $\rho < 1$
- $\rho > 1$: with $X(0) = 0$, $X_r(t) = N_r(t) - D_r(t)$ (arrivals minus departures)
  Work spent on class r by time t: $W_r(t) \geq \sum_{m=1}^{D_r(t)} \sigma_{r,m}$
  for i.i.d. service times $\sigma_{r,m}$
- Law of large numbers for Poisson processes: almost surely, $\lim_{t\to\infty} N_r(t)/t = \lambda_r$
- If for some r, $D_r(t) \leq \lambda_r t/2$ then $X_r(t) \geq \lambda_r t/2 + o(t)$
  Else, by Law of large numbers for $\sigma_{r,m}$, $\forall r, W_r(t) \geq D_r(t)/\mu_r + o(t)$
- Since $\sum_r W_r(t) \leq t$, implies

$$\sum_r X_r(t)/\mu_r \geq \rho t - t + o(t)$$

In both cases $\max_r X_r(t) \to \infty$ almost surely

Internet flow control



- Network links $\ell \in \mathcal{L}$ with capacity C_ℓ
- $X_t(r)$ transmissions of type $r \in \mathcal{R}$, use links $\ell \in r$
- Each gets allocation $\lambda_r \geq 0$, solving

$$\begin{array}{ll} \text{Max} & \sum_{r \in \mathcal{R}} X_t(r) U_r(\lambda_r) \\ \text{such that} & \forall \ell \in \mathcal{L}, \ \sum_{r \ni \ell} X_t(r) \lambda_r \leq C_\ell \end{array}$$

- Utility function $U_r(\lambda) = w_r \frac{\lambda^{1-\alpha}}{1-\alpha}$ if $\alpha \neq 1$, $w_r \log(\lambda)$ for $\alpha = 1$ $\rightarrow (w, \alpha)$-fairness (TCP: approximately $w_r = 1/T_r^2, \alpha = 2$)

Flow dynamics

- Requests for type r-transmissions arrive at (Poisson) rate ν_r
- Volume to be served: Exp(μ_r). Denote $\rho_r := \nu_r/\mu_r$
- Schedulable region $\mathcal{C}$: $x \in \mathbb{R}_+^{\mathcal{R}}$ such that $\forall \ell \in \mathcal{L}, \sum_{r \ni \ell} x_r \leq C_\ell$

Flow dynamics

- Requests for type r-transmissions arrive at (Poisson) rate ν_r
- Volume to be served: Exp(μ_r). Denote $\rho_r := \nu_r / \mu_r$
- Schedulable region $\mathcal{C}$: $x \in \mathbb{R}_+^{\mathcal{R}}$ such that $\forall \ell \in \mathcal{L}, \sum_{r \ni \ell} x_r \leq C_\ell$

Theorem

For positive (w, α) and (w, α)-fair sharing, if for some $\epsilon > 0$, $(1 + \epsilon)\rho \in \mathcal{C}$, then process $\{X_t\}_{t \in \mathbb{R}_+}$ is ergodic.
Conversely, if $\rho \notin \mathcal{C}$, for any feasible bandwidth allocation policy ((w, α)-fair or otherwise), process $\{X_t\}_{t \in \mathbb{R}_+}$ is transient.

Proof elements: Foster's criterion in continuous time

- Take Lyapunov function $V(x) := \sum_{r\in\mathcal{R}} \frac{1}{\mu_r} \int_0^{x_r} U_r'\left(\frac{\rho_r}{x}\right) dx$
  "Drift" of Lyapunov function:

$$
\begin{aligned}
\Delta \quad & := \sum_{r\in\mathcal{R}} \nu_r [V(x+e_r) - V(x)] + \mu_r x_r \lambda_r [V(x-e_r) - V(x)] \\
& \approx \sum_{r\in\mathcal{R}} (\nu_r - \mu_r x_r \lambda_r) \frac{\partial}{\partial x_r} V(x) \\
& = \sum_{r\in\mathcal{R}} (\rho_r - x_r \lambda_r) U_r' \left(\frac{\rho_r}{x_r}\right) \\
& = \sum_{r\in\mathcal{R}} (\rho_r - x_r \lambda_r) w_r \left(\frac{\rho_r}{x_r}\right)^{-\alpha}
\end{aligned}
$$

Proof elements: Foster's criterion in continuous time

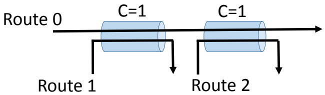
- Take Lyapunov function $V(x) := \sum_{r\in\mathcal{R}} \frac{1}{\mu_r} \int_0^{x_r} U_r'\left(\frac{\rho_r}{x}\right) dx$
  "Drift" of Lyapunov function:

$$
\begin{aligned}
\Delta \quad & := \sum_{r\in\mathcal{R}} \nu_r[V(x+e_r) - V(x)] + \mu_r x_r \lambda_r [V(x-e_r) - V(x)] \\
& \approx \sum_{r\in\mathcal{R}} (\nu_r - \mu_r x_r \lambda_r) \frac{\partial}{\partial x_r} V(x) \\
& = \sum_{r\in\mathcal{R}} (\rho_r - x_r\lambda_r) U_r'\left(\frac{\rho_r}{x_r}\right) \\
& = \sum_{r\in\mathcal{R}} (\rho_r - x_r\lambda_r) w_r \left(\frac{\rho_r}{x_r}\right)^{-\alpha}
\end{aligned}
$$

- $(1+\epsilon)\rho \in \mathcal{C} \Rightarrow$ allocations $\tilde{\lambda}_r = (1+\epsilon)\rho_r/x_r$ feasible
- Rates λ_r maximize $F(\lambda) := \sum_r w_r x_r \frac{\lambda_r^{1-\alpha}}{1-\alpha}$ over feasible allocations
- Hence concave function $t \in [0,1] \to F(t\lambda + (1-t)\tilde{\lambda})$ maximal at $t = 1$

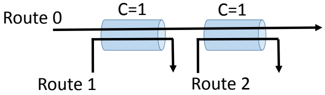
$$
\begin{aligned}
& \Rightarrow \sum_{r\in\mathcal{R}} [\lambda_r - \tilde{\lambda}_r] w_r x_r \tilde{\lambda}_r^{-\alpha} \geq 0 \\
& \Leftrightarrow \sum_{r\in\mathcal{R}} w_r [x_r\lambda_r - (1+\epsilon)\rho_r] \left(\frac{\rho_r}{x_r}\right)^{-\alpha} \geq 0 \\
& \Rightarrow \Delta \leq -\epsilon \sum_{r\in\mathcal{R}} w_r \rho_r^{1-\alpha} x_r^{\alpha}
\end{aligned}
$$

A suboptimal (unfair) allocation



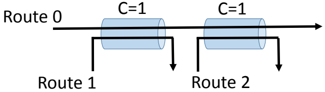
- Two-link network: ergodic under fair allocations if $\rho_0 + \rho_1 < 1, \rho_0 + \rho_2 < 1$

A suboptimal (unfair) allocation



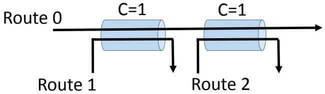
- Two-link network: ergodic under fair allocations if $\rho_0 + \rho_1 < 1, \rho_0 + \rho_2 < 1$
- Alternative allocation: give capacity 1 to types 1 and 2 if $x_1 + x_2 \geq 1$; give capacity 1 to type 0 only if $x_1 + x_2 = 0$

A suboptimal (unfair) allocation



- Two-link network: ergodic under fair allocations if $\rho_0 + \rho_1 < 1, \rho_0 + \rho_2 < 1$
- Alternative allocation: give capacity 1 to types 1 and 2 if $x_1 + x_2 \geq 1$; give capacity 1 to type 0 only if $x_1 + x_2 = 0$
  $\Rightarrow$ New condition for ergodicity: $\rho_0 < (1 - \rho_1)(1 - \rho_2)$
  e.g. Network unstable for $\rho_i \equiv 2/5, i = 0, 1, 2$

A suboptimal (unfair) allocation



- Two-link network: ergodic under fair allocations if $\rho_0 + \rho_1 < 1, \rho_0 + \rho_2 < 1$
- Alternative allocation: give capacity 1 to types 1 and 2 if $x_1 + x_2 \geq 1$; give capacity 1 to type 0 only if $x_1 + x_2 = 0$
  $\Rightarrow$ New condition for ergodicity: $\rho_0 < (1 - \rho_1)(1 - \rho_2)$
  e.g. Network unstable for $\rho_i \equiv 2/5, i = 0, 1, 2$
- Not an unrealistic allocation: results from differentiated service with priority to packets on short routes in network's routers and rate reduction by reactive control (TCP) at sender for longer route

Takeaway messages

Markov jump processes:
i) generator Q characterizes distribution if not explosive
ii) Balance equation characterizes invariant measure if irreducible non-explosive
iii) Limit theorems: stationary distribution reflects long-term performance

Exactly solvable models include reversible processes, plus several other important classes (e.g. Jackson networks)

Takeaway messages

- Ergodicity can be established with Foster's criterion and adequate Lyapunov function even when stationary distribution not known explicitly
- Several models for which **schedulable region** characterizes set of traffic parameters (loads per class) which make system ergodic, and for which known simple policy achieves ergodicity whenever possible with no explicit inference of traffic parameters
- Even though ergodicity a "first order" property (saying delays stay finite, not their magnitude), can yield useful insights, e.g. potential problems due to prioritizing packet service in Internet routers