Stochastic relations of random variables and processes

Lasse Leskelä

Helsinki University of Technology

7th World Congress in Probability and Statistics
Singapore, 18 July 2008

Fundamental problem of applied probability

$$\mathrm{E}\, f(X(t)) = ?$$

What if X is complex?

- Asymptotics
- Simulation
- Bounds

Stochastic bounds

Let X_1 and X_2 be (irreducible, positive recurrent) Markov processes with stationary distributions μ_1 and μ_2.

Problem

Can we show that $\mu_1 \leq_{\mathrm{st}} \mu_2$ without explicitly knowing μ_1 or μ_2?

Recall that μ_1 is stochastically less than μ_2, denoted $\mu_1 \leq_{\mathrm{st}} \mu_2$, if $\int f \, d\mu_1 \leq \int f \, d\mu_2$ for all positive increasing f.

Sufficient condition

Theorem (Whitt 1986; Massey 1987)

A sufficient condition for $\mu_1 \leq_{\mathrm{st}} \mu_2$ is that the transition rate kernels of X_1 and X_2 satisfy for all $x \leq y$:

- $Q_1(x, B) \leq Q_2(y, B)$ *for all upper sets B such that $x, y \notin B$*
- $Q_2(x, B) \geq Q_2(y, B)$ *for all lower sets B such that $x, y \notin B$*

The above condition is not sharp in general. Can we do any better?

Outline

Stochastic relations

Preservation of stochastic relations

Maximal subrelations

Outline

Stochastic relations

Preservation of stochastic relations

Maximal subrelations

Coupling

A coupling of random elements X and Y is a bivariate random element $(\hat{X}, \hat{Y})$ such that:

- $\hat{X}$ has the same distribution as X
- $\hat{Y}$ has the same distribution as Y

A coupling of probability measures μ on S_1 and ν on S_2 is a probability measure λ on $S_1 \times S_2$ having marginals μ and ν.

Remark

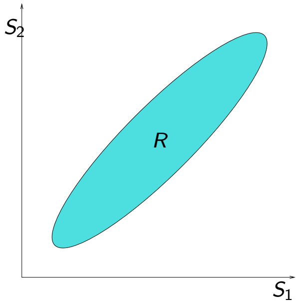
$(\hat{X}, \hat{Y})$ is a coupling of X and Y if and only if $\mathsf{P}((\hat{X}, \hat{Y}) \in \cdot)$ is a coupling of $\mathsf{P}(X \in \cdot)$ and $\mathsf{P}(Y \in \cdot)$.

Stochastic relations

Any meaningful distributional relation should have a coupling counterpart (Hermann Thorisson).

Stochastic relations

Any meaningful distributional relation should have a coupling counterpart (Hermann Thorisson).



Denote

- $x \sim y$, if $(x, y) \in R$
- $X \sim_{\mathrm{st}} Y$, if there exists a coupling $(\hat{X}, \hat{Y})$ of X and Y such that $\hat{X} \sim \hat{Y}$ almost surely.
- $\mu \sim_{\mathrm{st}} \nu$, if there exists a coupling λ of μ and ν such that $\lambda(R) = 1$.

$R_{\mathrm{st}} = \{(\mu, \nu) : \mu \sim_{\mathrm{st}} \nu\}$ is the stochastic relation generated by R.

- For Dirac measures, $\delta_x \sim_{\mathrm{st}} \delta_y$ if and only if $x \sim y$.

Functional characterization

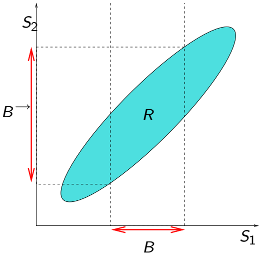
Theorem (Strassen 1965; L. 2008+)

The following are equivalent:

(i) $\mu \sim_{\mathrm{st}} \nu$

(ii) $\mu(B) \le \nu(B^{\rightarrow})$ *for all compact* $B \subset S_1$

(iii) $\int_{S_1} f\, d\mu \le \int_{S_2} f^{\rightarrow}\, d\nu$ *for all upper semicontinuous compactly supported* $f : S_1 \to \mathbb{R}_+$



$$B^{\rightarrow} = \cup_{x_1 \in B} \{x_2 \in S_2 : x_1 \sim x_2\}$$

$$f^{\rightarrow}(x_2) = \sup_{x_1 : x_1 \sim x_2} f(x_1).$$

Functional characterization

Theorem (Strassen 1965; L. 2008+)

The following are equivalent:

(i) $\mu \sim_{\mathrm{st}} \nu$

(ii) $\mu(B) \leq \nu(B^{\rightarrow})$ *for all compact* $B \subset S_1$

(iii) $\int_{S_1} f \, d\mu \leq \int_{S_2} f^{\rightarrow} \, d\nu$ *for all upper semicontinuous compactly supported* $f : S_1 \to \mathbb{R}_+$

Remark

If R is an order (reflexive and transitive) relation on S, then conditions (ii) and (iii) are equivalent to

(ii') $\mu(B) \leq \nu(B)$ for all measurable upper sets B,

(iii') $\int_S f \, d\mu \leq \int_S f \, d\nu$ for all increasing measurable $f : S_1 \to \mathbb{R}_+$.

(Strassen 1965; Kamae, Krengel, O'Brien 1977)

Examples

- Stochastic equality. Let $=_{\mathrm{st}}$ be the stochastic relation generated by the equality $=$. Then $X =_{\mathrm{st}} Y$ if and only if X and Y have the same distribution.
- Stochastic ϵ-distance. Define $x \approx y$ by $|x - y| \leq \epsilon$. Two real random variables satisfy $X \approx_{\mathrm{st}} Y$ if and only if for all x the corresponding c.d.f.'s satisfy $F_Y(x - \epsilon) \leq F_X(x) \leq F_Y(x + \epsilon)$.
- Stochastic induced order. Define $x \leq^{f,g} y$ by $f(x) \leq g(y)$. Then $\mu \leq^{f,g}_{\mathrm{st}} \nu$ if and only if $\mu(f^{-1}((\alpha, \infty))) \leq \nu(g^{-1}((\alpha, \infty)))$ for all real numbers α (Doisy 2000).

Outline

Stochastic relations

Preservation of stochastic relations

Maximal subrelations

Monotonicity vs. relation-preservation

Order relations $\rightsquigarrow$ monotone functions f:

$$x \leq y \implies f(x) \leq f(y)$$

General relations $\rightsquigarrow$ relation-preserving pairs of functions (f, g):

$$x \sim y \implies f(x) \sim g(y)$$

Stochastic relations $\rightsquigarrow$ stochastically relation-preserving pairs of probability kernels (random functions) (F, G):

$$x \sim y \implies F(x, \cdot) \sim_{\mathrm{st}} G(y, \cdot)$$

Preservation of stochastic relations

A pair of probability kernels (P_1, P_2) stochastically preserves a relation R, if

$$x_1 \sim x_2 \implies P_1(x_1, \cdot) \sim_{\mathrm{st}} P_2(x_2, \cdot)$$

or equivalently,

$$\mu_1 \sim_{\mathrm{st}} \mu_2 \quad \implies \mu_1 P_1 \sim_{\mathrm{st}} \mu_2 P_2.$$

Preservation of stochastic relations

A pair of probability kernels (P_1, P_2) stochastically preserves a relation R, if

$$x_1 \sim x_2 \implies P_1(x_1, \cdot) \sim_{\mathrm{st}} P_2(x_2, \cdot)$$

or equivalently,

$$\mu_1 \sim_{\mathrm{st}} \mu_2 \quad \implies \mu_1 P_1 \sim_{\mathrm{st}} \mu_2 P_2.$$

Theorem (Zhang 1998; L. 2008+)

A pair (P_1, P_2) stochastically preserves R if and only if there exists a probability kernel P on $S_1 \times S_2$ such that:

(i) *$P(x, \cdot)$ couples $P_1(x_1, \cdot)$ and $P_2(x_2, \cdot)$ for all $x = (x_1, x_2)$.*

(ii) $x \in R \implies P(x, R) = 1.$

Stochastic relations of Markov processes

A pair of Markov processes stochastically preserve a relation R, if

$$x \sim y \quad \Longrightarrow \quad X(x,t) \sim_{\mathrm{st}} Y(y,t) \text{ for all } t,$$

or equivalently,

$$\mu \sim_{\mathrm{st}} \nu \quad \Longrightarrow \quad X(\mu,t) \sim_{\mathrm{st}} Y(\nu,t) \text{ for all } t.$$

Stochastic relations of Markov processes

A pair of Markov processes stochastically preserve a relation R, if

$$x \sim y \quad \Longrightarrow \quad X(x,t) \sim_{\mathrm{st}} Y(y,t) \text{ for all } t,$$

or equivalently,

$$\mu \sim_{\mathrm{st}} \nu \quad \Longrightarrow \quad X(\mu,t) \sim_{\mathrm{st}} Y(\nu,t) \text{ for all } t.$$

Remark

A Markov process is stochastically monotone, if

$$x \leq y \quad \Longrightarrow \quad X(x,t) \leq_{\mathrm{st}} X(y,t) \text{ for all } t.$$

Relation-preserving Markov processes

Let X_1 and X_2 be discrete-time Markov processes with transition probability kernels P_1 and P_2.

Theorem (L. 2008+)

The following are equivalent:

(i) X_1 *and* X_2 *stochastically preserve the relation* R*.*

(ii) $P_1(x_1, B) \leq P_2(x_2, B^{\rightarrow})$ *for all* $x_1 \sim x_2$ *and compact* $B \subset S_1$*.*

(iii) *There exists a Markovian coupling of* X_1 *and* X_2 *for which* R *is invariant.*

Relation-preserving Markov processes

Let X_1 and X_2 be discrete-time Markov processes with transition probability kernels P_1 and P_2.

Theorem (L. 2008+)

The following are equivalent:

(i) *X_1 and X_2 stochastically preserve the relation R.*

(ii) *$P_1(x_1, B) \le P_2(x_2, B^{\rightarrow})$ for all $x_1 \sim x_2$ and compact $B \subset S_1$.*

(iii) *There exists a Markovian coupling of X_1 and X_2 for which R is invariant.*

Remarks

- If R is an order, (ii) can be replaced by
  (ii') $P_1(x_1, B) \le P_2(x_2, B)$ for all $x_1 \le x_2$ and upper sets B
  (Kamae, Krengel, O'Brien 1977).
- An analogous result holds for nonexplosive Markov jump processes, generalizing the result of Whitt and Massey.

Outline

Stochastic subrelations

Recall our starting point:

Problem

Can we show that the stationary distributions μ_1 *and* μ_2 *of Markov processes* X_1 *and* X_2 *satisfy* $\mu_1 \leq_{\mathrm{st}} \mu_2$ *without explicitly knowing* μ_1 *or* μ_2*?*

Stochastic subrelations

Recall our starting point:

Problem

Can we show that the stationary distributions μ_1 and μ_2 of Markov processes X_1 and X_2 satisfy $\mu_1 \leq_{\mathrm{st}} \mu_2$ without explicitly knowing μ_1 or μ_2?

- The sufficient condition of Whitt and Massey essentially says that X_1 and X_2 stochastically preserve the order relation $R_{\leq} = \{(x, y) : x \leq y\}$.

Stochastic subrelations

Recall our starting point:

Problem

Can we show that the stationary distributions μ_1 and μ_2 of Markov processes X_1 and X_2 satisfy $\mu_1 \leq_{\mathrm{st}} \mu_2$ without explicitly knowing μ_1 or μ_2?

- The sufficient condition of Whitt and Massey essentially says that X_1 and X_2 stochastically preserve the order relation $R_{\leq} = \{(x, y) : x \leq y\}$.
- A less stringent sufficient condition: Show that X_1 and X_2 stochastically preserve a nontrivial subrelation of $R_{\leq}$.

Subrelation algorithm

Given a closed relation R and continuous probability kernels P_1 and P_2, define a sequence of relations by $R^{(0)} = R$,

$$R^{(n+1)} = \left\{ (x, y) \in R^{(n)} : \ (P_1(x, \cdot), P_2(y, \cdot)) \in R^{(n)}_{\mathrm{st}} \right\},$$

and let $R^* = \bigcap_{n=0}^{\infty} R^{(n)}$.

Subrelation algorithm

Given a closed relation R and continuous probability kernels P_1 and P_2, define a sequence of relations by $R^{(0)} = R$,

$$R^{(n+1)} = \left\{ (x,y) \in R^{(n)} : \ (P_1(x,\cdot), P_2(y,\cdot)) \in R^{(n)}_{\mathrm{st}} \right\},$$

and let $R^* = \bigcap_{n=0}^{\infty} R^{(n)}$.

Theorem (L. 2008+)

The relation R^* is the maximal closed subrelation of R that is stochastically preserved by (P_1, P_2). Especially, the pair (P_1, P_2) preserves a nontrivial subrelation of R if and only if $R^* \neq \emptyset$.

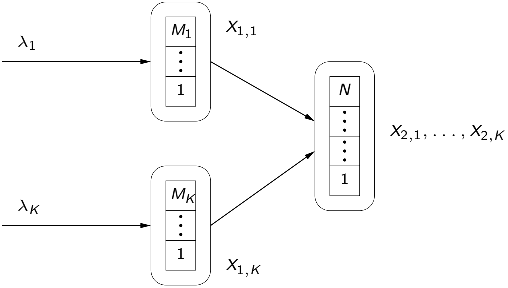
Remark

A modified algorithm works for Markov jump processes.

Application: Multilayer loss network

Multiclass loss network with

- M_k servers dedicated to class-k jobs (layer 1)
- N multiclass servers processing the overflow traffic (layer 2)



Application: Multilayer loss network

Modified system $Y = (Y_{1,1}, \ldots, Y_{1,K}; Y_{2,1}, \ldots, Y_{2,K})$

- One class-1 server replaced by a shared server
- Can we show that $\mathsf{E} \sum_{i,k} X_{i,k} \leq \mathsf{E} \sum_{i,k} Y_{i,k}$ in steady state?

Define the relation $x \sim y$ by $\sum_{i,k} x_{i,k} \leq \sum_{i,k} y_{i,k}$.

- $\sim$ is not an order (different state spaces)
- X and Y do not preserve $\sim_{\rm st}$
- But maybe (X, Y) preserves some subrelation of $\sim_{\rm st}$?

Application: Multilayer loss network

Example

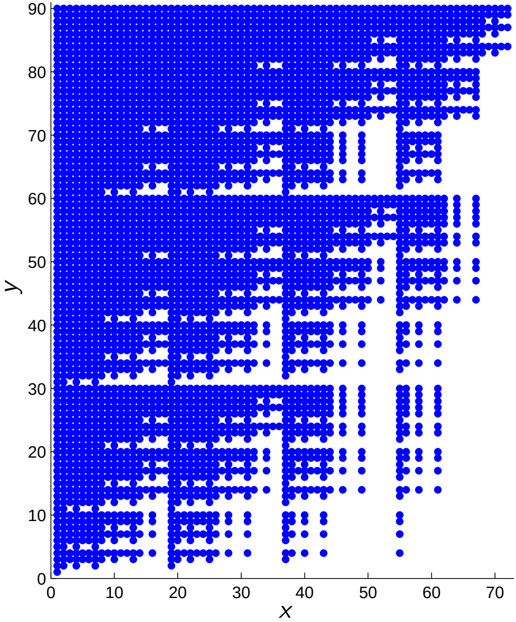
Two customer classes

- Server configuration: $M_1 = 3$, $M_2 = 2$, $N = 2$
- Arrival rates $\lambda_1 = 1$, $\lambda_2 = 2$
- Service rate $\mu = 1$

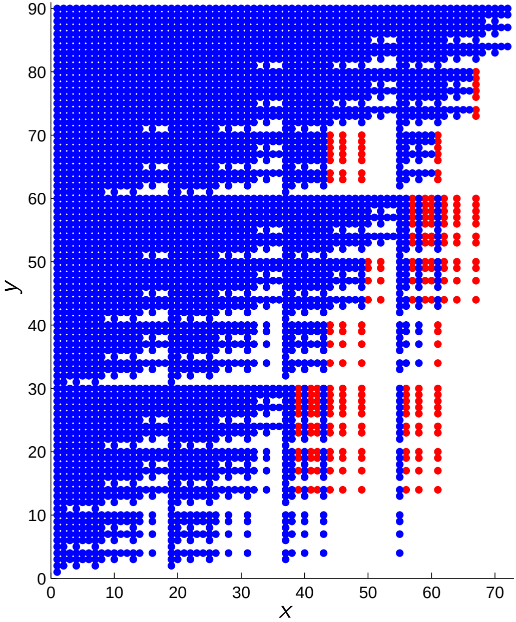
How many iterations do we need to compute R_∞?

- X has 72 possible states
- Y has 90 possible states

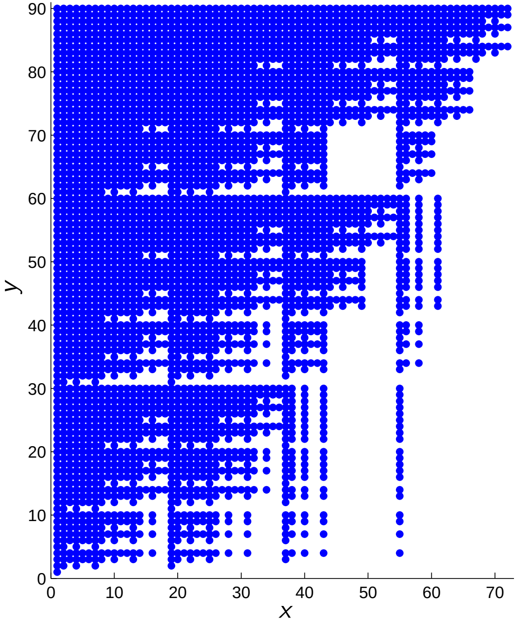
Application: Multilayer loss network



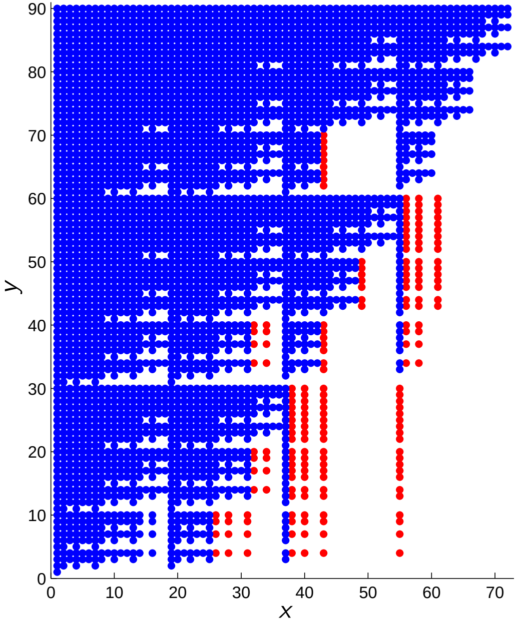
Application: Multilayer loss network



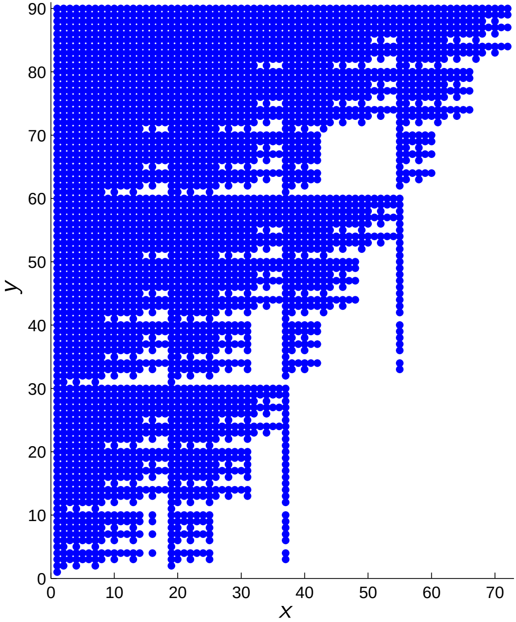
Application: Multilayer loss network



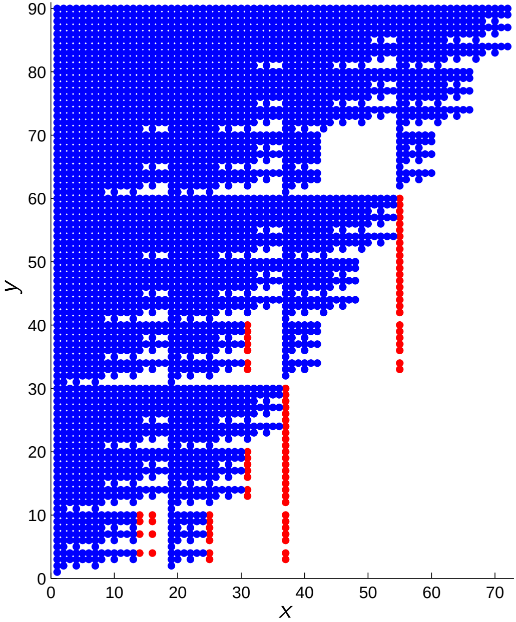
Application: Multilayer loss network



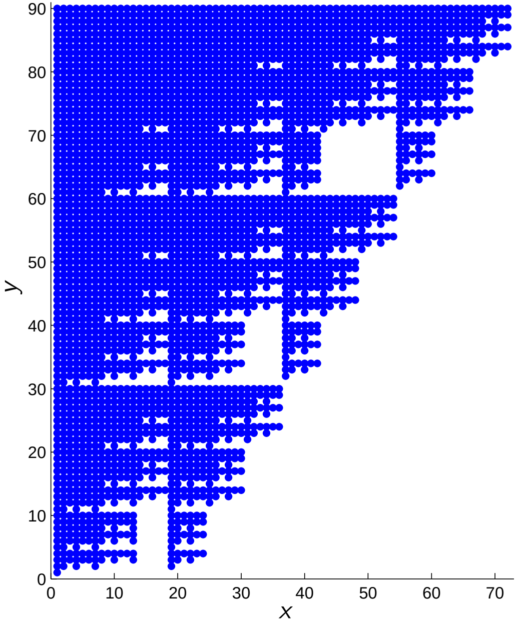
Application: Multilayer loss network



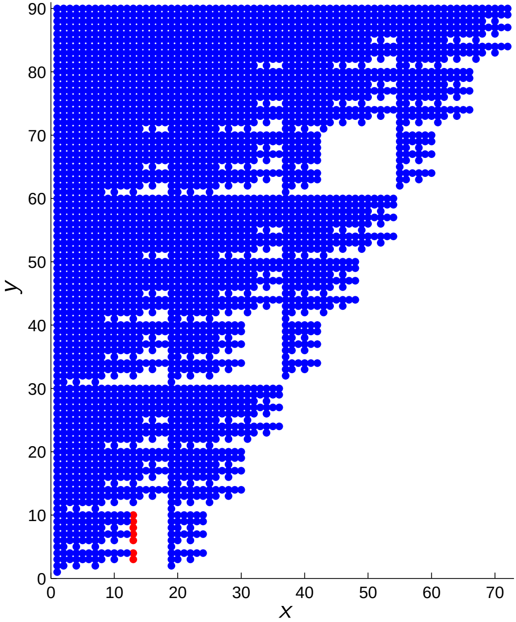
Application: Multilayer loss network



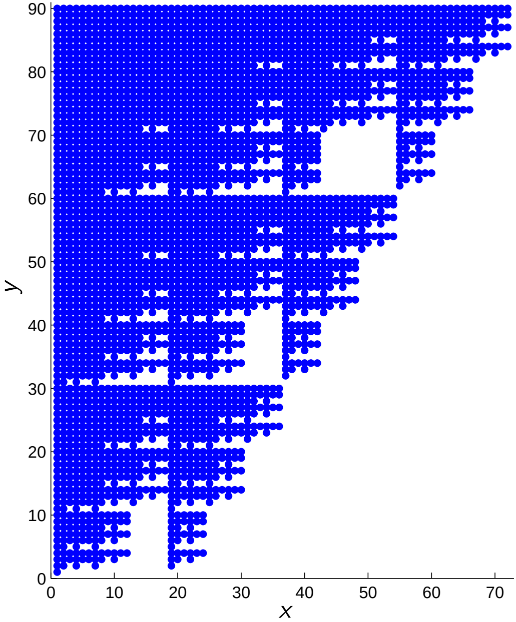
Application: Multilayer loss network



Application: Multilayer loss network



Application: Multilayer loss network



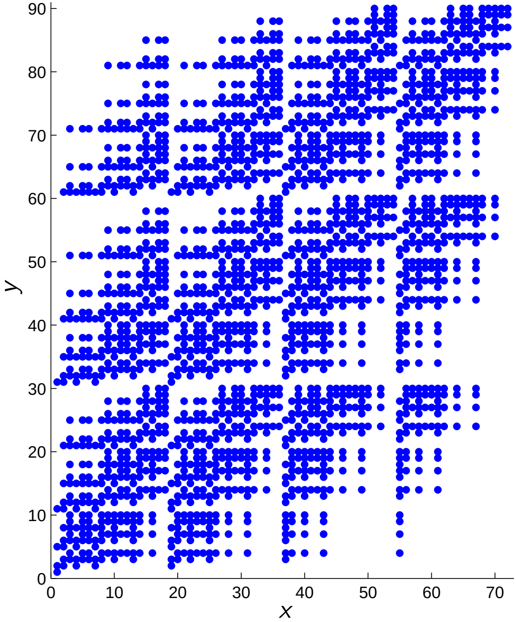
Application: Multilayer loss network

What if we started with a stricter relation?

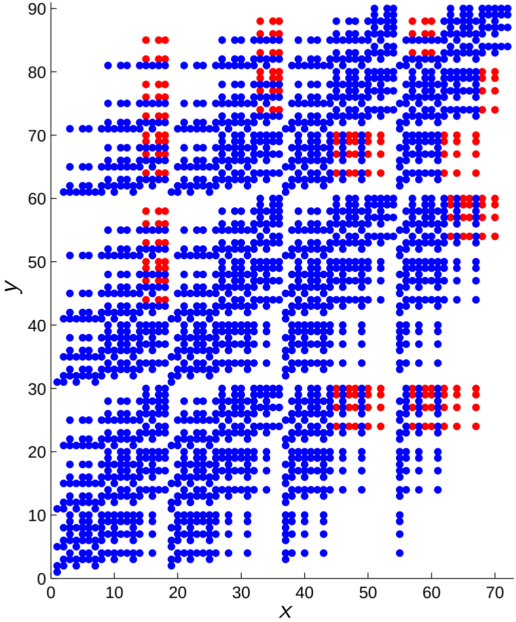
Redefine $x \sim y$ by

$$0 \leq \sum_{i,k} y_{i,k} - \sum_{i,k} x_{i,k} \leq 1$$

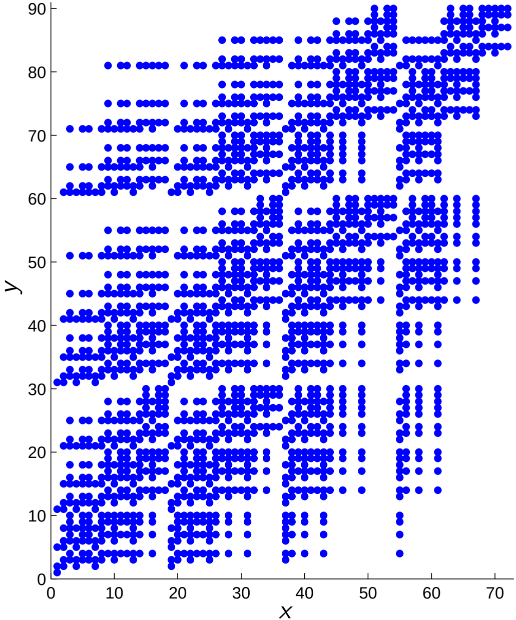
Application: Multilayer loss network



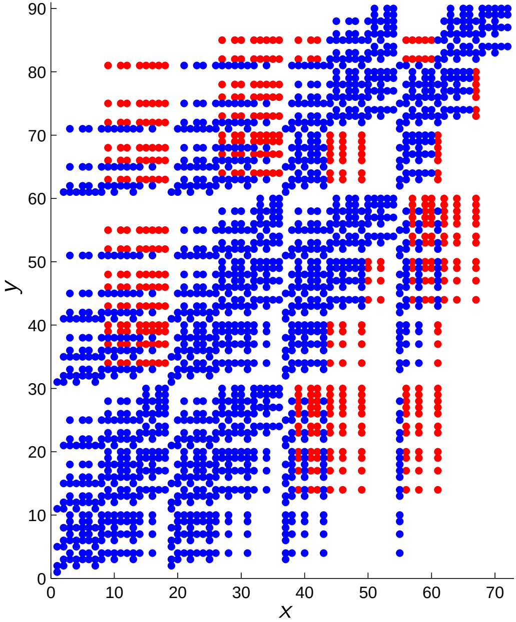
Application: Multilayer loss network



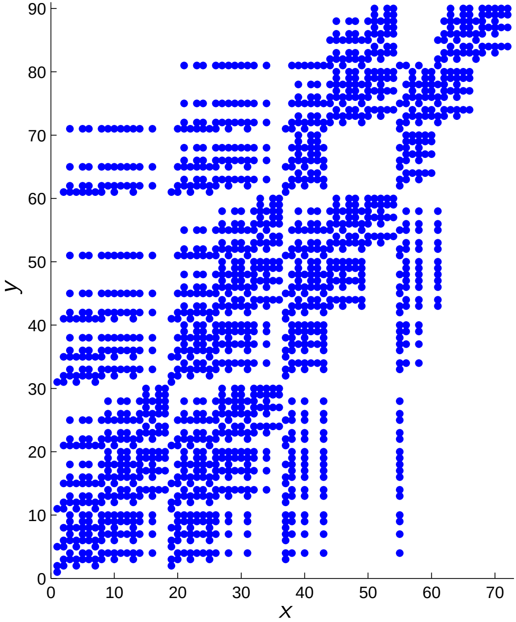
Application: Multilayer loss network



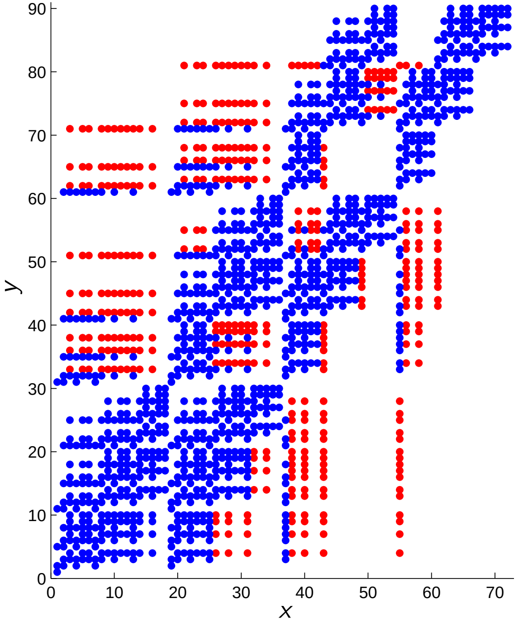
Application: Multilayer loss network



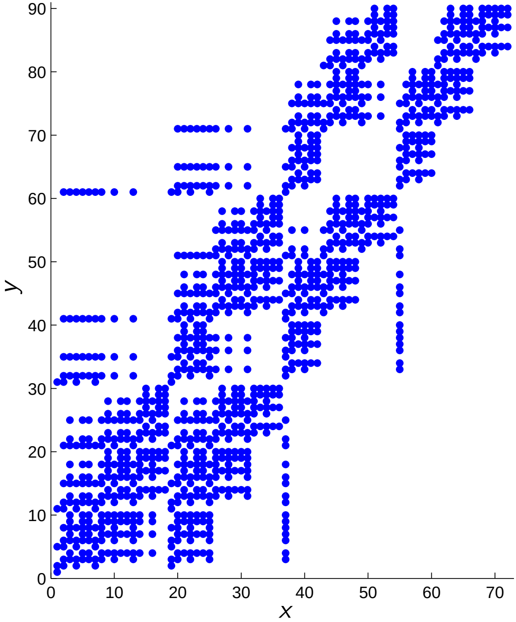
Application: Multilayer loss network



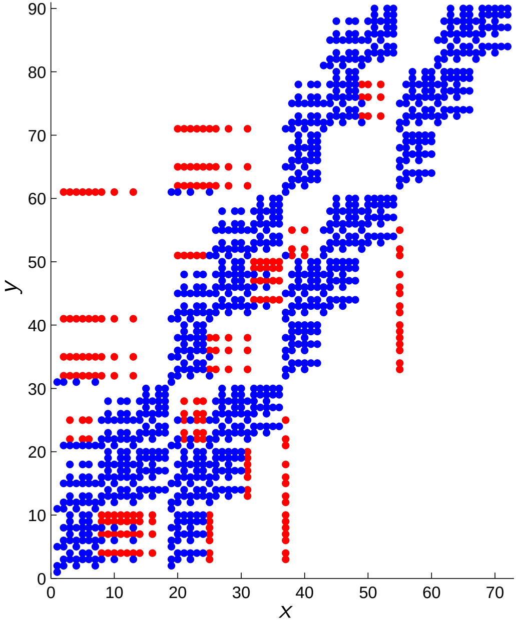
Application: Multilayer loss network



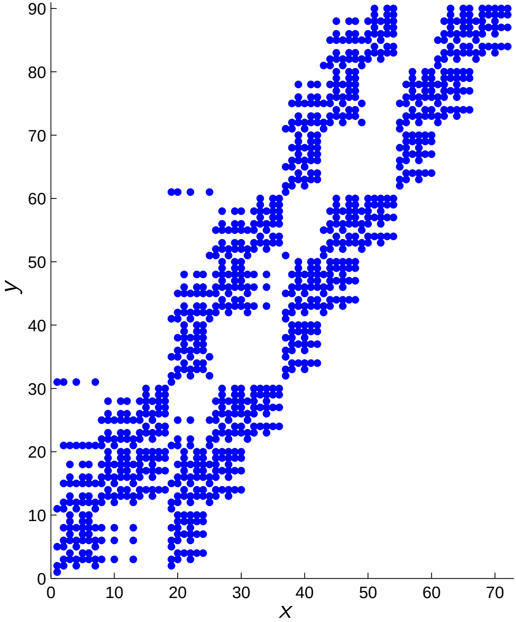
Application: Multilayer loss network



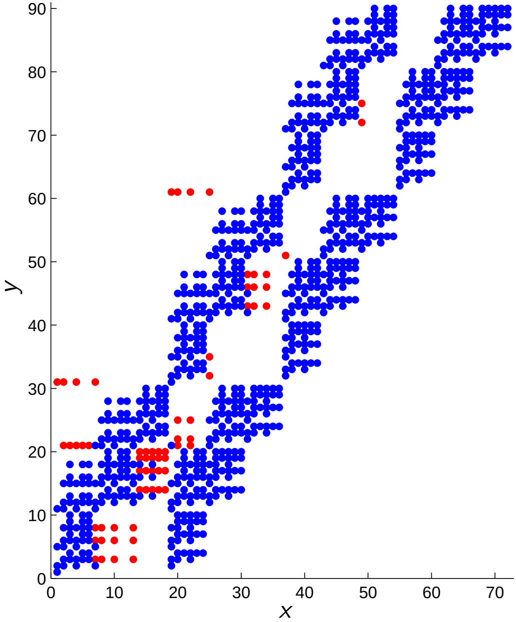
Application: Multilayer loss network



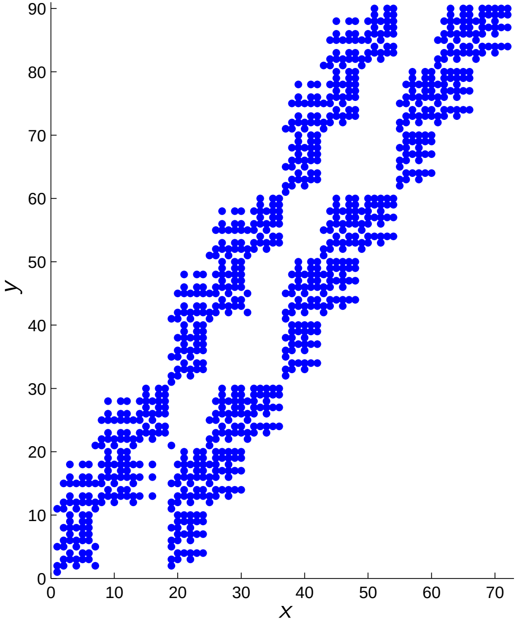
Application: Multilayer loss network



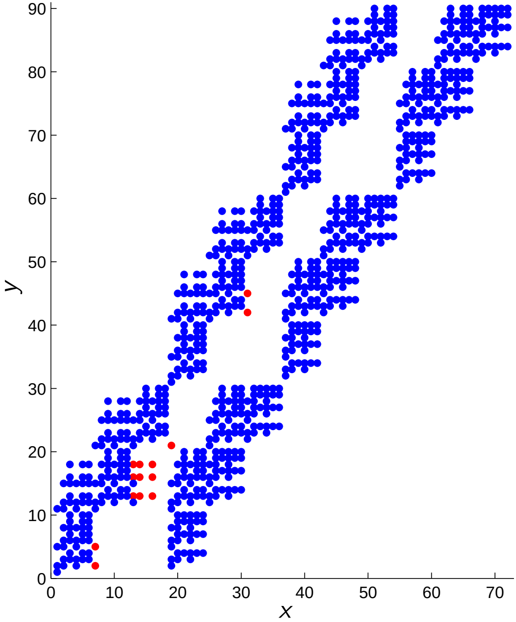
Application: Multilayer loss network



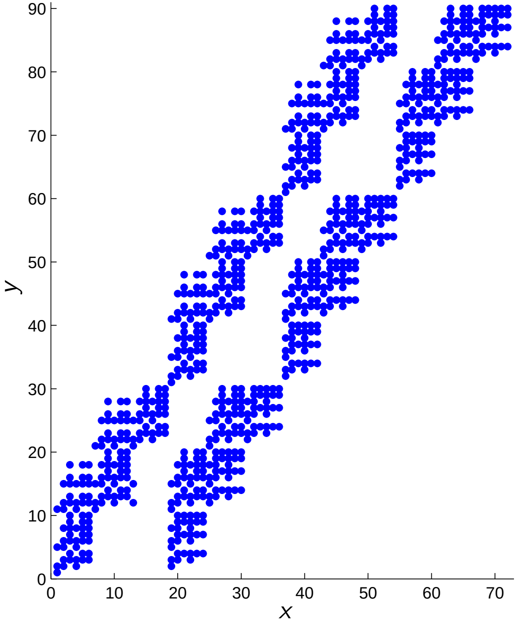
Application: Multilayer loss network



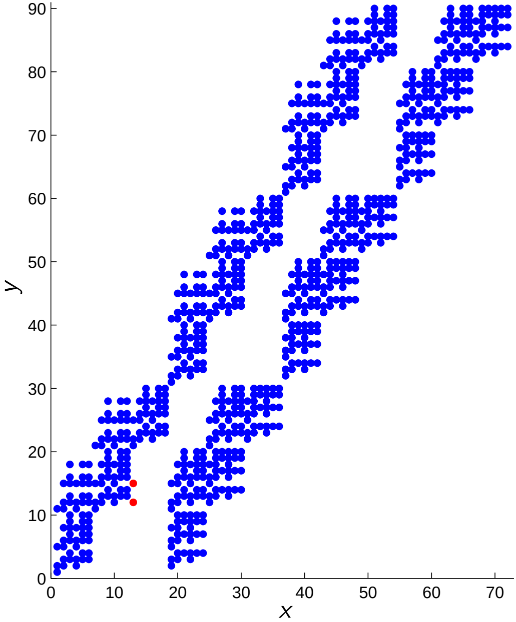
Application: Multilayer loss network



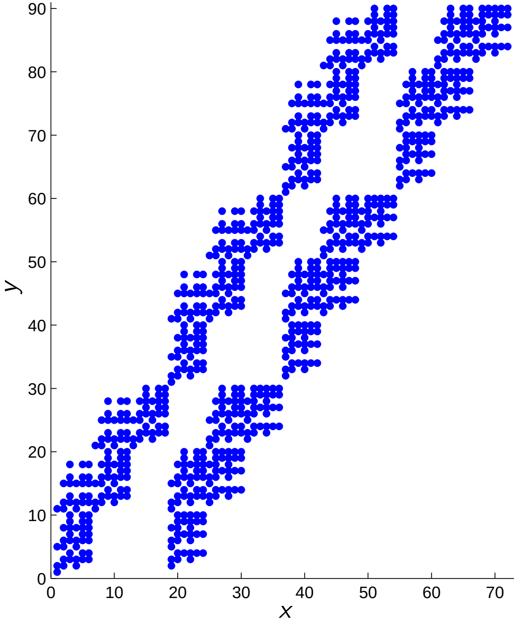
Application: Multilayer loss network



Application: Multilayer loss network



Application: Multilayer loss network



Application: Multilayer loss network

Theorem (Jonckheere & L. 2008)

The processes X and Y stochastically preserve the relation $R = \{(x, y) : |x - y| \in \Delta\}$, where

$$\Delta = \{0,\ e_2,\ e_2 - e_{1,1},\ 2e_2 - e_{1,1}\}.$$

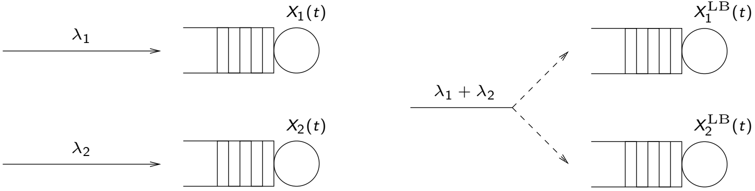
Especially, the stationary distributions of the processes satisfy

$$|Y| - 1 \le_{\mathrm{st}} |X| \le_{\mathrm{st}} |Y|,$$

and

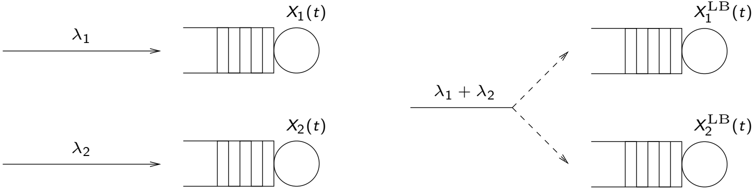
$$\begin{aligned} X_{1,1} &\ge_{\mathrm{st}} Y_{1,1}, \\ X_{1,k} &=_{\mathrm{st}} Y_{1,k} \quad \text{for all } k \neq 1, \\ \sum_k X_{2,k} &\le_{\mathrm{st}} \sum_k Y_{2,k}. \end{aligned}$$

Application: Load balancing



Common sense: $\mathsf{E}(X_1^{\mathrm{LB}}(t) + X_2^{\mathrm{LB}}(t)) \leq \mathsf{E}(X_1(t) + X_2(t))$

Application: Load balancing



Common sense: $\mathsf{E}(X_1^{\mathrm{LB}}(t) + X_2^{\mathrm{LB}}(t)) \le \mathsf{E}(X_1(t) + X_2(t))$

The rate kernel pair (Q^{LB}, Q) does not stochastically preserve:

- $R^{\mathrm{nat}} = \{(x,y) : x_1 \le y_1,\ x_2 \le y_2\}$
- $R^{\mathrm{sum}} = \{(x,y) : |x| \le |y|\}$, where $|x| = x_1 + x_2$

How about a subrelation of R^{sum}?

Application: Load balancing

Theorem (L. 2008+)

The subrelation algorithm started from R^{sum} *yields*

$$R^{(n)} = \left\{(x,y) : |x| \le |y| \text{ and } x_1 \vee x_2 \le y_1 \vee y_2 + (y_1 \wedge y_2 - n)^+\right\}$$
$$\downarrow$$
$$R^* = \{(x,y) : |x| \le |y| \text{ and } x_1 \vee x_2 \le y_1 \vee y_2\}.$$

Especially, (Q^{LB}, Q) *stochastically preserves the relation* R^**.*

Remark

- R^* is the weak majorization order on $\mathbb{Z}_+^2$
- $X \sim_{\mathrm{st}}^* Y$ if and only if $\mathrm{E}\, f(X) \le \mathrm{E}\, f(Y)$ for all coordinatewise increasing Schur-convex functions f (Marshall & Olkin 1979).

Conclusions

Algorithmic probability

- Computational methods for analytical results
- Comparison without ordering
- State space reduction

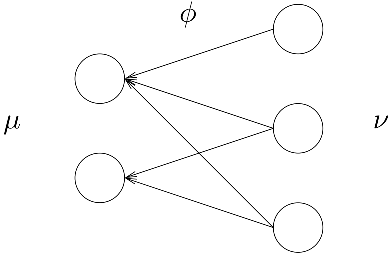
Open problems:

- Numerical methods for finite Markov chains
- Subrelations versus dependence orderings
- Diffusions, Feller processes, martingales, ...

Discussion: Coupling vs. mass transportation

$$W_\phi(\mu,\nu) = \inf_{\lambda \in K(\mu,\nu)} \int_{S_1 \times S_2} \phi(x_1, x_2)\, \lambda(dx)$$

- $K(\mu,\nu)$ is the set of couplings of μ and ν



- W_ϕ is a Wasserstein metric, if ϕ is a metric.
- $\mu \sim_{\mathrm{st}} \nu$ if and only if $W_\phi(\mu,\nu) = 0$ for $\phi(x_1, x_2) = 1(x_1 \not\sim x_2)$.

(Monge 1781, Kantorovich 1942, Wasserstein 1969, Chen 2005)

M.-F. Chen.
Eigenvalues, Inequalities, and Ergodic Theory.
Springer, 2005.

M. Doisy.
A coupling technique for stochastic comparison of functions of Markov processes.
J. Appl. Math. Decis. Sci., 4(1):39–64, 2000.

M. Jonckheere and L. Leskelä.
Stochastic bounds for two-layer loss systems.
To appear in Stoch. Models, arXiv:0708.1927, 2008+.

T. Kamae, U. Krengel, and G. L. O'Brien.
Stochastic inequalities on partially ordered spaces.
Ann. Probab., 5(6):899–912, 1977.

L. Leskelä.
Stochastic relations of random variables and processes.
Submitted. Preprint: `http://www.iki.fi/lsl/`, 2008+.

W. A. Massey.

Stochastic orderings for Markov processes on partially ordered spaces.
Math. Oper. Res., 12(2):350–367, 1987.

V. Strassen.
The existence of probability measures with given marginals.
Ann. Math. Statist., 36(2):423–439, 1965.

H. Thorisson.
Coupling, Stationarity, and Regeneration.
Springer, 2000.

W. Whitt.
Stochastic comparisons for non-Markov processes.
Math. Oper. Res., 11(4):608–618, 1986.

S.-Y. Zhang.
Existence of ρ-optimal coupling operator for jump processes.
Acta Math. Sin. (Chinese series), 41(2):393–398, 1998.