technische universität
dortmund

A Partial Order Approach to Optimization under Uncertainty

Günter Rudolph

Lehrstuhl für Algorithm Engineering

Fakultät für Informatik

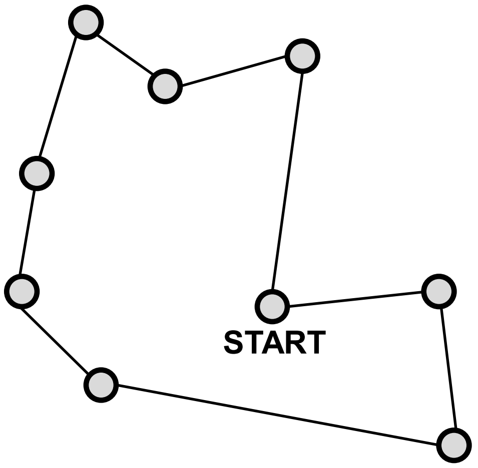
TU Dortmund

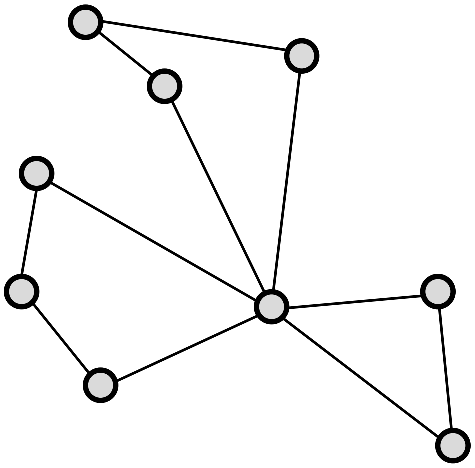
Bourglinster
26-MAY-11

Outline

- Introduction: From Deterministic to Stochastic Models
- Traditional Approach to Uncertainty
- Partial Order Approach to Uncertainty
- Future Work

Deterministic Models of the World





TSP (Travelling Salesperson Problem)

minimize tour length among n cities;

but deliver all goods to all cities!

TSP is NP-hard already!

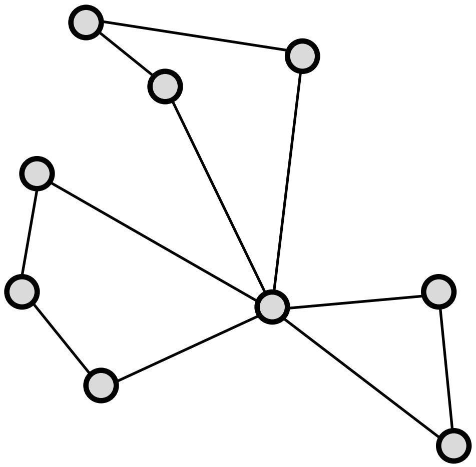
VRP (Vehicle Routing Problem)

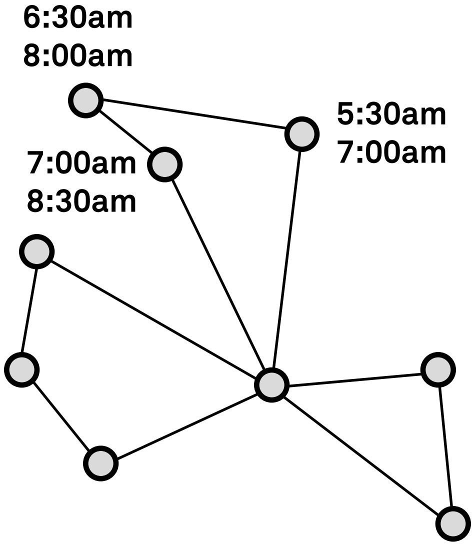
arises from capacity constraint

on goods to be loaded on one truck;

VRP more diffucult than TSP.

Deterministic Models of the World





VRP (Vehicle Routing Problem) arises from capacity constraint on goods to be loaded on one truck;

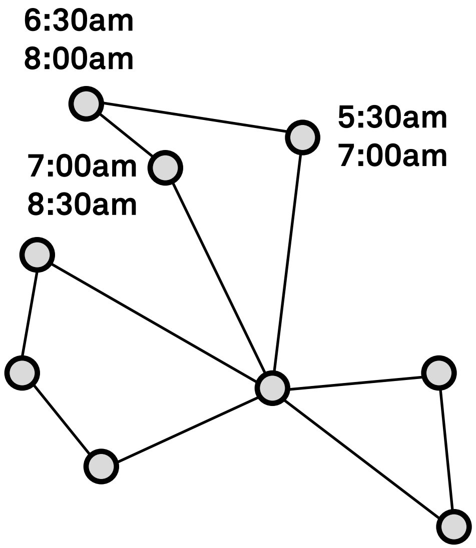
VRPTW (VRP with Time Windows) Adding time windows specifying when goods may be delivered!

VRP more difficult than TSP.

VRPTW more difficult than VRP.

From Deterministic to Stochastic Models

Deterministic Models of the World



VRPTW (VRP with Time Windows)
Adding time windows specifying when
goods may be delivered!

VRPTW more difficult than VRP.

Caveat!

Optimal solutions assume:
distance between two cities
can be driven with some
mean velocity

⇓

required time to travel
a certain distance
is **proportional**
to the **distance**!

Is that true?

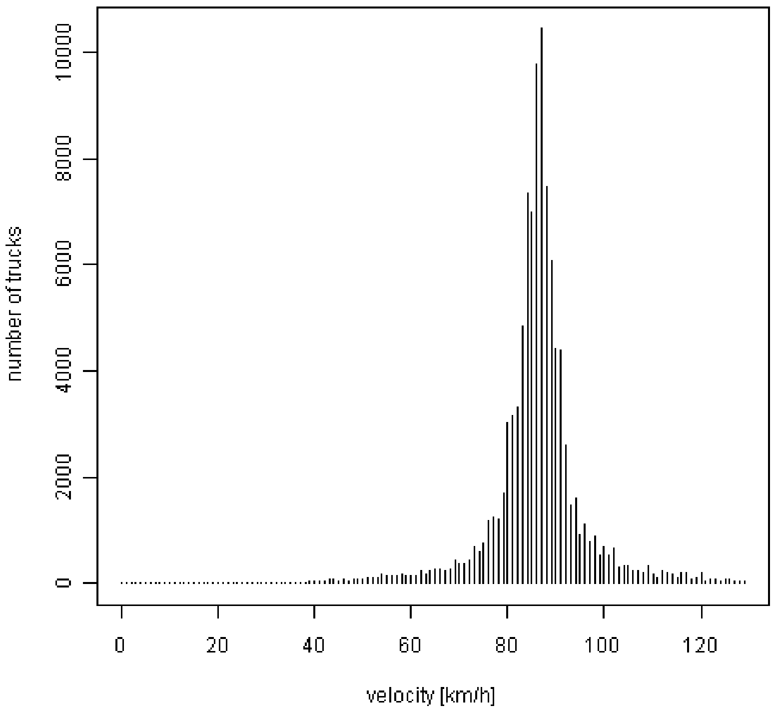
From Deterministic to Stochastic Models

Real world data: Measure speed of trucks on German motorways (60k/day)

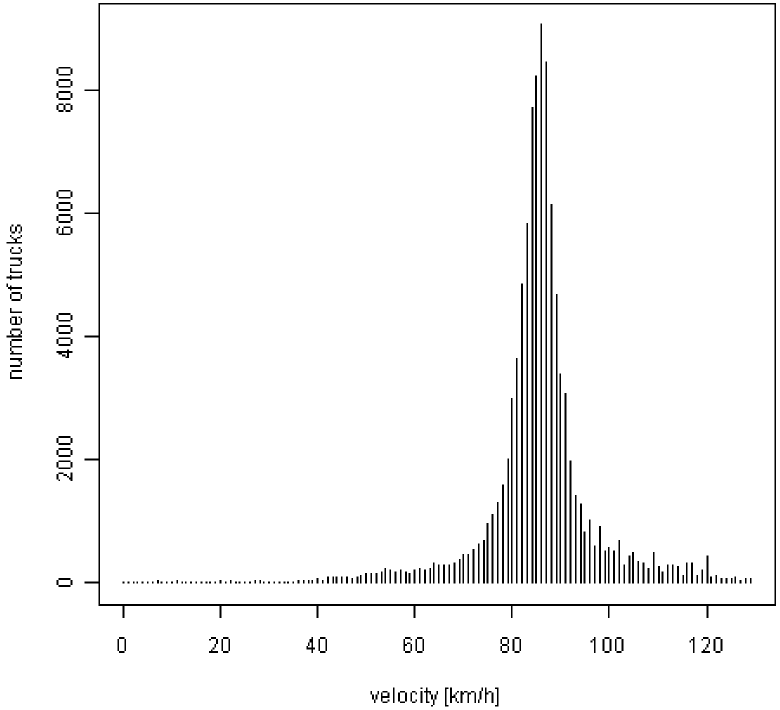
Monday 30-AUG-2004

speed limit 80 km/h

0 am – 4 am



4 am – 8 am



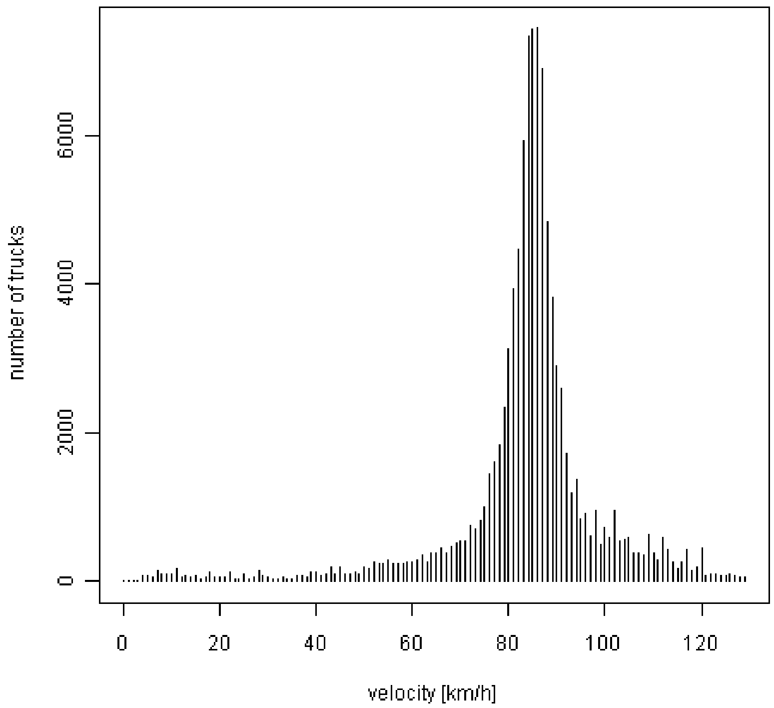
From Deterministic to Stochastic Models

Real world data: Measure speed of trucks on German motorways

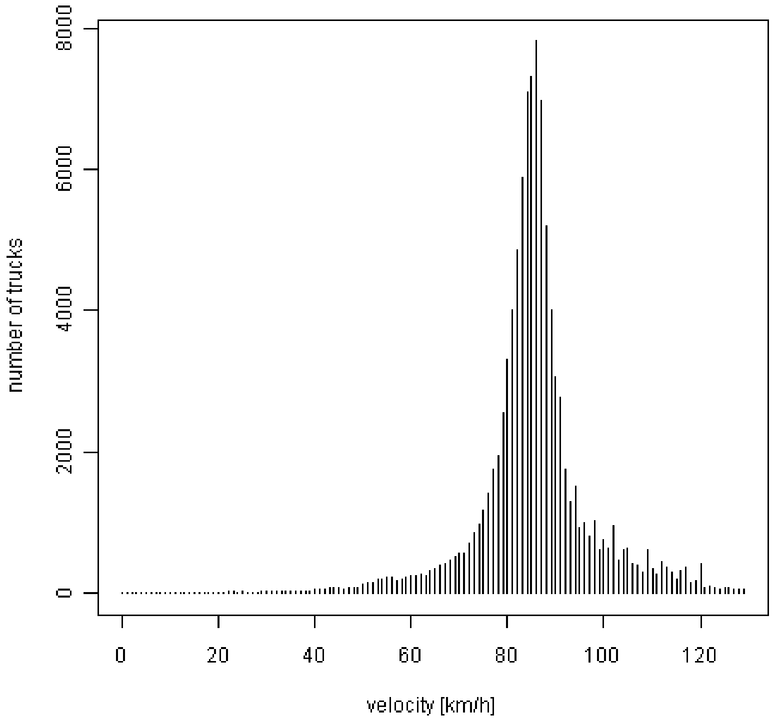
Monday 30-AUG-2004

speed limit 80 km/h

8 am – 12 am



12 am – 4 pm



From Deterministic to Stochastic Models

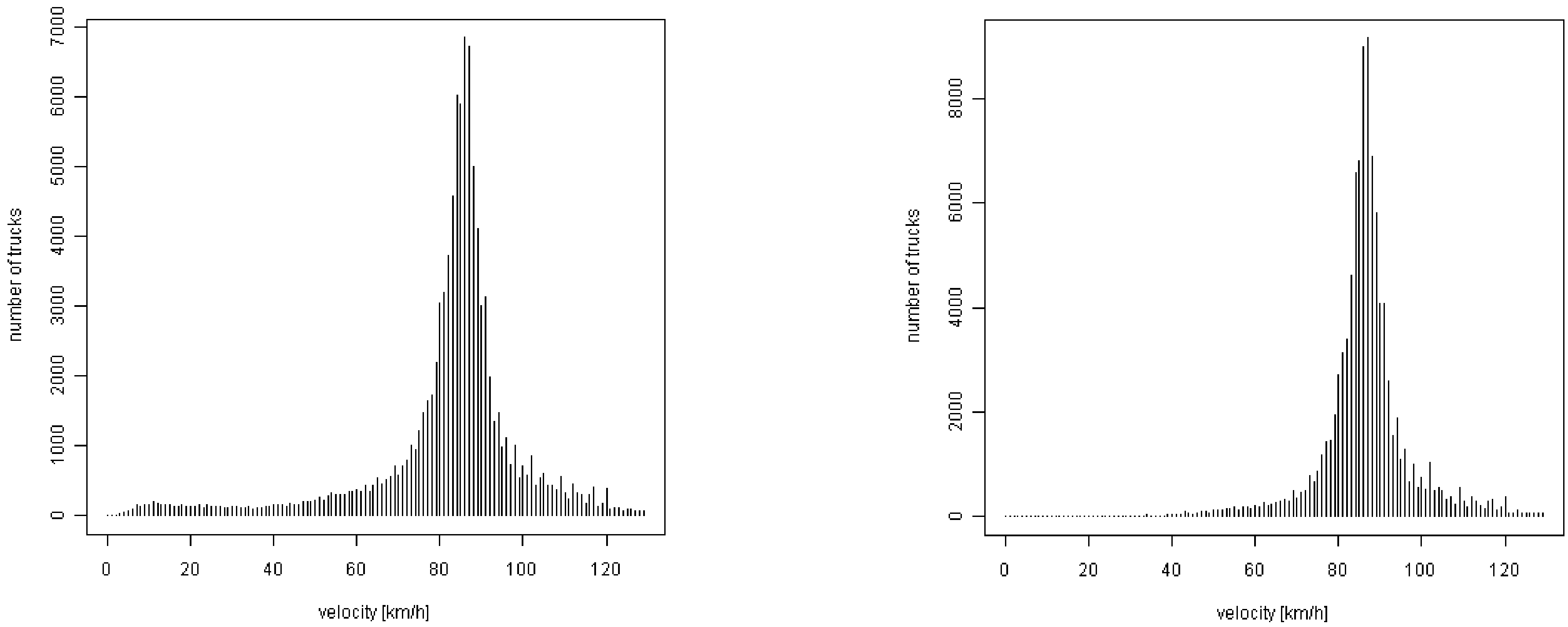
Real world data: Measure speed of trucks on German motorways

Monday 30-AUG-2004

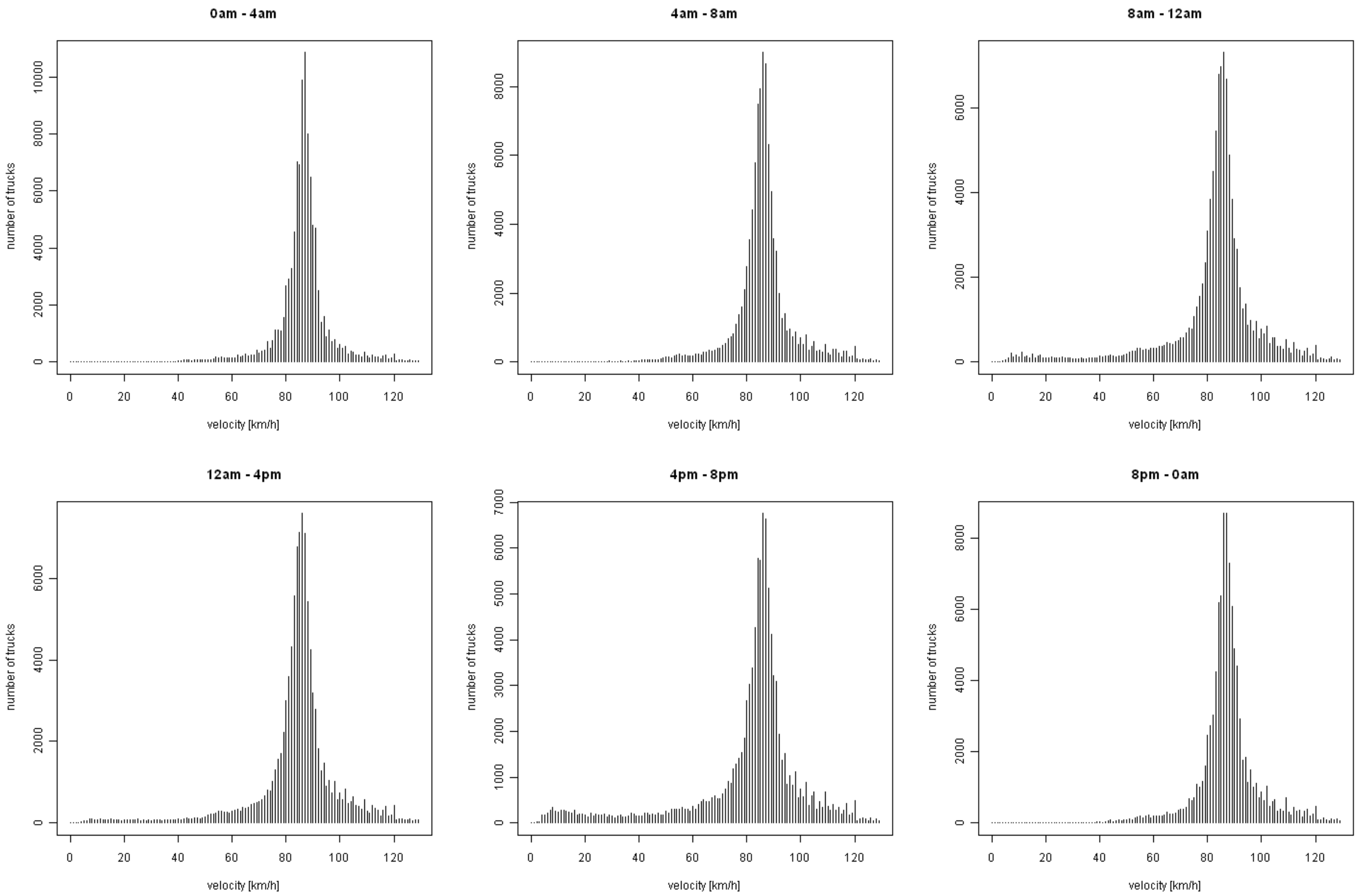
speed limit 80 km/h

4pm – 8 pm

8 am – 12 pm

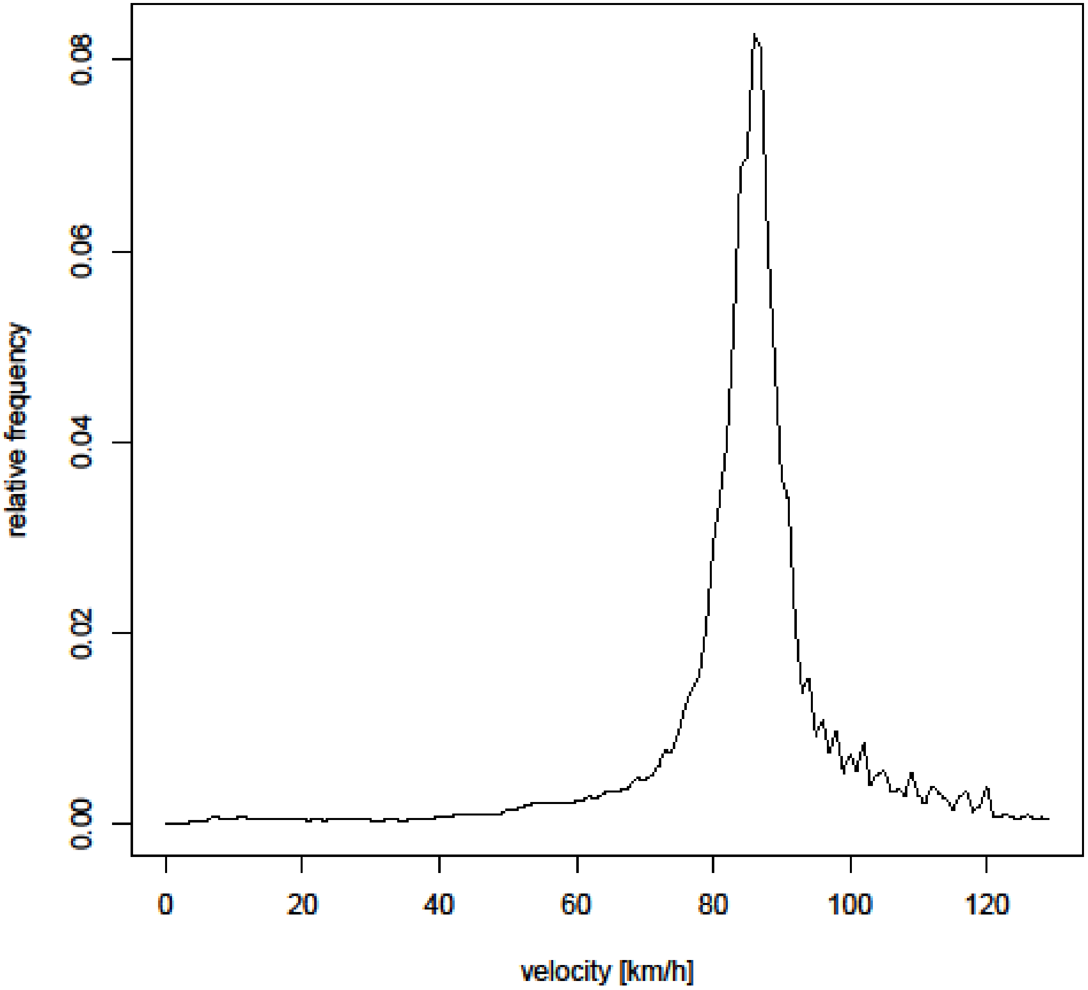


From Deterministic to Stochastic Models



Thursday 02-SEP-2004

From Deterministic to Stochastic Models



Estimated from data:

$$\begin{aligned} \hat{\mu} &= 84.65 \\ \hat{\sigma}^2 &= 207.19 \end{aligned}$$

distribution?

- Gauss ☒
- Laplace ☒☑
- Cauchy ☑

required time to drive a certain distance is a **random variable**

From Deterministic to Stochastic Models

Deterministic Optimization

typically:
- deterministic parameters
- static, i.e. time-invariant parameters

decision variables, here: $x \in \mathbb{R}^n \times \mathbb{Z}^m$, $n \geq 1$, $m \geq 0$

Stochastic Optimization

Parameters = random variables (distributions known)

⇒ objective function is random variable

⇒ minimization of random variable is mathematically senseless

Surrogate / Auxiliary Models

Idea: make problem deterministic

original problem: $f(x, \omega) \rightarrow$ "min! " , $\omega \in \Omega$ random parameter

surrogate problems: $E[\, f(x, \omega)\,] \rightarrow \text{min!}$

$P\{\, f(x, \omega) \leq \delta \,\} \rightarrow \text{max!}$

$E[\, f(x, \omega)\,] + \xi \cdot V[\, f(x, \omega)\,] \rightarrow \text{min!}$

Difficulties

- higher moments increase nonlinearity!

$f(x, \omega) = \omega_1 x^2 + \omega_2 x \quad \Rightarrow \quad V[\, f(x, \omega)\,] = x^4 V[\, \omega_1] + x^2 V[\omega_2\,]$

- analytically unsolvable integrals may show up

$E[\, \max\{\, x^2 \omega_i - x\, \omega_i^2 ,\, i = 1, \ldots , 10 \,\}\,] \;,\; \omega_i \sim N(0,1)$

$\Rightarrow$ metaheuristics come into play ...

Traditional Approach

$$F(x) = E[\, f(x, \omega)\,] = \int \underset{\substack{\uparrow \\ \text{known}}}{f(x, \omega)}\; \underset{\substack{\uparrow \\ \text{known}}}{p_{\omega}(\omega)}\, d\omega$$

Idea: numerical integration

Problem: finite discretization of support → approximation

$$F(x) = \underbrace{g(x; K)}_{\text{deterministic}} + \underbrace{Z(K)}_{\text{“noise“}}, \quad K \text{ number of steps}$$

Metaheuristics and Control of Noise

Metaheuristics (like ES) can cope with noise!

ES far away from optimum:

→ noise does not affect ES taking correct selection decisions

ES close to optimum:

→ Control of “noise“ strength by better num. integration

⇒ saves many function evaluations!

Metaheuristics and Metamodels

Traditional Approach

Metamodel =

- model of true objective function (model of a model)
- can be evaluated mucher quicker than true objective function

→ 2nd layer of surrogate model (here!)

$F(x, \omega)$	original
$F(x) = E[\, f(x, \omega) \,]$	1st layer
$M(x) = \text{Metamodel}_o f\, (\, F(x) \,)$	2nd layer

→ approximation error = *noise*

Special Scenario ⇒ Specialized Metaheuristic

Nonlinear two-stage stochastic problem with recourse

$\min \{ f(x) + E[\, g(x, \omega) \,] \; : x \in X \}$

with $g(x, \omega) = \min\{ h(x, y, \omega) : y \in Y \}$ and random $\omega \in \Omega$

⇒ **nested metaheuristics (hybrid / memetic)**

Outer method optimizes over X

Inner method optimizes over Y (using ideas presented previously)

Back to the introductory example …

Two objectives:

1. Minimize tour length per truck (under given constraints)
2. Minimize risk that a truck misses time window

$\Rightarrow$ Biobjective deterministic problem!

Resolves the uncertainty only partially …

Incomparable solutions:
small length / high risk vs. longer tour / less risk

Hypothesis

(strict) partial order captures uncertainty!

additional problem classes / sources of uncertainty:

- multiple objectives (conflicting goals)
- unter risk (e.g noisy objective function)
- under incomplete knowledge
  (e.g. interval-values objective function without measure)
- ...

If true then we need only 1 optimization algorithm

that can cope with strict partial orders (leads to less extensive theory)

Working Plan

Choose general problem class

that contains previously mentioned problem classes as special cases

Develop theory for general problem class

We obtain:

- theory for special cases with small effort
- revealing insight into structural properties of problem

Quick Reminder on Partial Orders (I)

$(\mathcal{F}, \preceq)$. partially ordered set,

if binary relation $\preceq$ is partial order relation on set $\mathcal{F}$, i.e.

- reflexive $a \preceq a$
- antisymmetric $a \preceq b \wedge b \preceq a \Rightarrow a = b$
- transitive $a \preceq b \wedge b \preceq c \Rightarrow a \preceq c$

strict partial order: $a \prec b \Leftrightarrow a \preceq b \wedge a \neq b$

- antireflexive $a \prec b \Rightarrow a \neq b$
- asymmetric $a \prec b \Rightarrow \neg(b \prec a)$
- transitive $a \prec b \wedge b \prec c \Rightarrow a \prec c$

Quick Reminder on Partial Orders (II)

two elements $a, b \in \mathcal{F}$ are

(a) comparable, if $a \preceq b$ or $b \preceq a$

(b) incomparable if neither $a \preceq b$ nor $b \preceq a$, denoted $a \parallel b$

if all pairs of distinct elements from $\mathcal{A} \subseteq \mathcal{F}$ are

(a) comparable, then $\mathcal{A}$ is a chain;

(b) incomparable, then $\mathcal{A}$ is an antichain.

Quick Reminder on Partial Orders (III)

$a^* \in \mathcal{F}$ minimal element $\Leftrightarrow \nexists\, a \in \mathcal{F} : a \prec a^*$

set of minimal elements $\mathcal{M}(\mathcal{F}, \preceq)$

complete $\forall\, a \in \mathcal{F} : \exists\, a^* \in \mathcal{M}(\mathcal{F}, \preceq) : a^* \preceq a$

Quick Reminder on Partial Orders (IV)

map $f : \mathcal{X} \to \mathcal{F}$

decision set $\mathcal{X}$, partially ordered objective set $(\mathcal{F}, \preceq)$

induces order on objective set:

$$x \prec_f y \quad :\Leftrightarrow \quad f(x) \prec f(y)$$

$$x \sim_f y \quad :\Leftrightarrow \quad f(x) = f(y)$$

$$x \preceq_f y \quad :\Leftrightarrow \quad x \prec_f y \lor x \sim_f y$$

minimal elements in objective set:

$$\mathcal{M}_f(\mathcal{A}, \preceq_f) := \{\, a \in \mathcal{A} : f(a) \in \mathcal{M}(\mathcal{F}, \preceq) \,\}, \quad \mathcal{A} \subseteq \mathcal{X}$$

Base Algorithms

Optimization task: Find subset of $\mathcal{M}_f(\mathcal{X}, \preceq)$

How to measure progress in optimization?

let $|\mathcal{X}| < \infty; \quad A, B \subseteq \mathcal{X} \Rightarrow$

$$d(A, B) = |A \cup B| - |A \cap B| \quad \text{(metric)}$$

$$\delta_B(A) = |A| - |A \cap B|$$

Convergence to set of minimal elements:

$d(f(P_t), \mathcal{F}^*) \to 0$ or $\delta_{\mathcal{F}^*}(f(P_t)) \to 0$

Plan for developing a joint theory:

1. Design probabilistic base algorithms with desired properties under weakest preconditions
2. Qualitative behavior of base algorithm only depends on structural properties of transition matrix/function
3. Product decomposition of transition matrix into transition matrices of single (genetic) operators
4. Qualitative behavior depends on structural properties of single transition matrices (and their combination)
5. Instantiation of base algorithm for special problem class
6. Prove that that transition matrices of single operators possess these structural properties

Base Algorithms

Algorithm A1

```
$B(0)$ drawn at random from $\mathcal{X}^n$
$A(0) = \mathcal{M}_f(B(0), \preceq)$
$t = 0$
repeat
  $B(t+1) = \text{generate}(B(t))$
  $A(t+1) = \mathcal{M}_f(A(t) \cup B(t+1), \preceq)$
  $t \leftarrow t+1$
until stopping criterion fulfilled
```

Theorem:

If $(B_t)_{t>0}$ homogeneous finite Markov chain with irreducible transition matrix then

$$d(f(A_t), \mathcal{F}^*) \to 0$$

with probability 1.

Base Algorithms

Algorithm A2

```
$B(0)$ drawn at random from $\mathcal{X}^n$
$A(0) = \mathcal{M}_f(B(0), \preceq)$
$t = 1$
repeat
  $B(t) = \text{generate}(B(t-1))$
  $B^*(t) = \mathcal{M}_f(B(t), \preceq)$
  $C(t) = \emptyset$
  foreach $b \in B^*(t)$ do
    $D_b = \{a \in A(t) : f(b) \prec f(a)\}$
    if $D_b \neq \emptyset$ then $A(t) \leftarrow (A(t) \setminus D_b) \cup \{b\}$
    if $\forall a \in A(t) : f(a) \parallel f(b)$ then $C(t) \leftarrow C(t) \cup \{b\}$
  endfor
  $k = \min\{m - |A(t)|, |C(t)|\}$
  $A(t+1) = A(t) \cup \text{draw}(k, C(t))$
  $t \leftarrow t + 1$
until stopping criterion fulfilled
```

Partial Order Approach

Theorem:

If $(B_t)_{t>0}$ homogeneous finite Markov chain with irreducible transition matrix then

$$\delta_{\mathcal{F}^*}(f(A_t)) \to 0$$

and

$$|A_t| \to \min\{\, m, |\mathcal{F}^*|\,\}$$

with probability 1.

Remark:

A1 and A2 identical if target set is a chain.

Base Algorithms

Algorithm B1

```
$B(0)$ drawn at random from $\mathcal{X}^n$
$A(0) = \mathcal{M}_f(B(0), \preceq)$
$t = 0$
repeat
  $B(t+1) = \text{generate}(A(t))$
  $A(t+1) = \mathcal{M}_f(A(t) \cup B(t+1), \preceq)$
  $t \leftarrow t+1$
until stopping criterion fulfilled
```

Theorem:

If transition matrix from A_t to B_{t+1} positive, then

$$d(f(A_t), \mathcal{F}^*) \to 0$$

with probability 1.

Base Algorithms

Algorithm B2

```
$B(0)$ drawn at random from $\mathcal{X}^n$
$A(0) = \mathcal{M}_f(B(0), \preceq)$
$t = 1$
repeat
  $B(t) =$ generate$(A(t-1))$
  $B^*(t) = \mathcal{M}_f(B(t), \preceq)$
  $C(t) = \emptyset$
  foreach $b \in B^*(t)$ do
    $D_b = \{a \in A(t) : f(b) \prec f(a)\}$
    if $D_b \neq \emptyset$ then $A(t) \leftarrow (A(t) \setminus D_b) \cup \{b\}$
    if $\forall a \in A(t) : f(a) \parallel f(b)$ then $C(t) \leftarrow C(t) \cup \{b\}$
  endfor
  $k = \min\{m - |A(t)|, |C(t)|\}$
  $A(t+1) = A(t) \cup$ draw$(k, C(t))$
  $t \leftarrow t + 1$
until stopping criterion fulfilled
```

Theorem:

If transition matrix from A_t to B_{t+1} positive, then

$$\delta_{\mathcal{F}^*}(f(A_t)) \to 0$$

and

$$|A_t| \to \min\{\, m, |\mathcal{F}^*|\,\}$$

with probability 1.

Remark:

B1 and B2 identical id target set is a chain:
Both EAs turn to EA with (1+n)-selection.

Partial Order Approach

Lemma:

Let B, P, I, D, C be stochastic matrices, where

- B arbitrary $b_{ij} \geq 0 \wedge \forall i : \Sigma_j \, b_{ij} = 1$
- P positive $\forall i, j : p_{ij} > 0$
- I irreducible $\exists k : I^k$ itiv
- D diagonal-positive $\forall i : d_{ii} > 0$
- C column-allowable $\forall j : \exists i : c_{ij} > 0$

Then these statements are true:

(a) ID and DI are irreducible,

(b) BP and PC are positive.

Partial Order Approach

Theorem:

(a) $\left(\prod_{i=1}^{u} B_i\right) \; P \; \left(\prod_{j=1}^{v} C_j\right)$ is positive.

(b) $\left(\prod_{i=1}^{u} D_i\right) \; I \; \left(\prod_{j=1}^{v} \widetilde{D}_j\right)$ is irreducible.

→ Instantiation: $X = \{ 0, 1 \}^n$

Base Algorithms

Crossover	Mutation	Preselection
B arbitrary	P bit inversion with 0 < prob. < 1	C (D) no preselection
D mating selection „with repetition“	I 1-bit-inversion with probability 1	D antichain hierarchy with ranks according to #dom. individuals, Tournament selection on ranks

Base Approach

Plan:

1. Definie partial order relation on objective set
   (output: better, worse, equally good, don‘t know)
2. Prove: reflexive, antisymmetric, transitive
3. Characterize minimal elements
4. Verify: Are minimal elements solution that we really want?
5. Transcription of general theory for special case
6. If applies, simplify base algorithm

multiobjective problems:

- partial order: $a \preceq b \;\Leftrightarrow\; \forall i : a_i \leq b_i \;\wedge\; a \neq b$
- minimal elements: Pareto-optimal objective vectors

$\Rightarrow$ convergence to pareto-optimal set ✓

Partial Order Approach

Interval-valued objective function:

- partial order: $[u_1, u_2] \prec [v_1, v_2] \;\Leftrightarrow\; [u_1, u_2] \cap [v_1, v_2] = \emptyset$
- minimal elements:

(a) Intervals, that contain f* ; ✓

(b) Intervals, that intersect above intervals ✗

→ Deploy methods for interval size reduction

Partial Order Approach

Noisy objective function: $\tilde{f}(x) = f(x) + Z,\ Z \in [-\delta, \delta]$

- partial order: $\tilde{f}(x) \prec \tilde{f}(y) \ \Leftrightarrow\ \tilde{f}(x) + \delta < \tilde{f}(y) - \delta$
- minimal elements:

ε- optimal solution with ε=3δ ✓

$\rightarrow$ Deploy methods for ε- reduction

Partial Order Approach

Noisy objective function: $\tilde{f}(x) = f(x) + Z,\ Z$ arbitrary

- partial order: $\tilde{f}(x) \prec \tilde{f}(y) \ \Leftrightarrow$

$$\mathsf{P}\{\,\tilde{f}(x) < \tilde{f}(y) \mid f(x) < f(y)\} \geq 1 - \alpha$$

realized by statistical test
(with k samples)

- minimal elements:

ε- optimal solutions with error probability α ✓

Conclusions / Future Work

Some evidence that hypthesis is true:
Uncertainty can be captured by partial order

Theory for base algorithms with partially ordered sets yields

- limit theory for many problem subclasses (for free)
- user-friendly conditions based on single operators

Future work: extension w.r.t. base algorithms & problems classes