Accuracy vs. Complexity: the stochastic bound approach[1]

F. Aït Salaht [1] J. Cohen[1] H. Castel Taleb [2]
J.M. Fourneau [1] N. Pekergin [3]

[1]PRiSM, Univ. Versailles St Quentin, UMR CNRS 8144, Versailles France

[2]SAMOVAR, UMR 5157, Télécom Sud Paris, Evry, France

[3]LACL, Univ. Paris Est, Créteil, France

Wodes 2012, October 2012

[1] Digiteo Project: MARINA

Motivation

• Solving some problems of performance evaluation which deal with discrete distributions.

Motivation

• Solving some problems of performance evaluation which deal with discrete distributions.

- **Problem:**
  - Numerical solutions are computationally hard;
  - Distribution size increases multiplicatively.

Motivation

• Solving some problems of performance evaluation which deal with discrete distributions.

- **Problem:**
  - Numerical solutions are computationally hard;
  - Distribution size increases multiplicatively.
- **Proposition:**
  - Use the stochastic bound theory to reduce the size of the distribution at each step of the computation;

    Stochastic bound $\Longrightarrow$ Result is a bound of the exact distribution;
  - Control the distribution size $\Longrightarrow$ Control of the complexity ;
  - Develop an algorithmic approach to obtain stochastic bounds.

Basic assumptions

We consider:

- $\boldsymbol{d}$: Discrete probability distribution on totally ordered state space $\mathcal{H}$, $|\mathcal{H}| = N$, $\boldsymbol{d}(i) > 0$ *for* $i \in \mathcal{H}$;
- $\boldsymbol{r}$: Positive increasing reward; $R_{\boldsymbol{d}} = \sum \boldsymbol{r}(i)\boldsymbol{d}(i)$;

Goal:

- Compute distribution $\boldsymbol{db}$ on support $\mathcal{F}$ with K states such that $K << N$;
- $\boldsymbol{db}$ is $\begin{cases} \textit{the best approximation of } \boldsymbol{d} \textit{ for } \boldsymbol{r}; \\ \textit{stochastic lower (resp. upper) bound.} \end{cases}$

- Let $\mathcal{G} = \mathcal{H} \cup \mathcal{F}$.
- Totally ordered and finite state space $\longrightarrow$ minimal and maximal state, denoted as *MinState* and *MaxState*.

Stochastic bounds

- $\mathcal{G} = \{1, 2, \ldots, n\}$ a finite state space.
- X, Y: discrete distributions over $\mathcal{G}$;
- $p_X(i) = prob(X = i)$ and $p_Y(i) = prob(Y = i)$ for $i \in \mathcal{G}$.

Definition ($\leq_{st}$ order)

$$X \leq_{st} Y \text{ iff } \quad \sum_{k=i}^{n} p_X(k) \leq \sum_{k=i}^{n} p_Y(k), \qquad \forall i.$$

Stochastic bounds

- $\mathcal{G} = \{1, 2, \ldots, n\}$ a finite state space.
- X, Y: discrete distributions over $\mathcal{G}$;
- $p_X(i) = prob(X = i)$ and $p_Y(i) = prob(Y = i)$ for $i \in \mathcal{G}$.

Definition ($\leq_{st}$ order)

$$X \leq_{st} Y \;\; iff \quad \sum_{k=i}^{n} p_X(k) \leq \sum_{k=i}^{n} p_Y(k), \qquad \forall i.$$

Comparison of non decreasing functionals

$$X \leq_{st} Y \iff \mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$$

for all non decreasing functions $f : \mathcal{G} \to \mathbb{R}^+$ whenever expectations exist.

Most accurate stochastic bounds

For $\boldsymbol{d}$ defined over $\mathcal{H}$, compute $\boldsymbol{d}1$ and $\boldsymbol{d}2$ such that:

1. $\boldsymbol{d}2 \leq_{st} \boldsymbol{d} \leq_{st} \boldsymbol{d}1$,
2. $\boldsymbol{d}1$ and $\boldsymbol{d}2$ have only K states (not necessarily the same set, but of the same size),
3. $$\sum_{i \in \mathcal{G}} \boldsymbol{r}(i)\boldsymbol{d}(i) - \sum_{i \in \mathcal{G}} \boldsymbol{r}(i)\boldsymbol{d}2(i)$$ is minimal among the lower bounding distributions of $\boldsymbol{d}$ with K states,
4. $$\sum_{i \in \mathcal{G}} \boldsymbol{r}(i)\boldsymbol{d}1(i) - \sum_{i \in \mathcal{G}} \boldsymbol{r}(i)\boldsymbol{d}(i)$$ is minimal among the upper bounding distributions of $\boldsymbol{d}$ with K states.

Most accurate stochastic bounds

Proposition

$\boldsymbol{r}$ is increasing and $\boldsymbol{d}2 \leq_{st} \boldsymbol{d}$:

$$\sum_{i\in\mathcal{G}} \boldsymbol{r}(i)\boldsymbol{d}(i) - \sum_{i\in\mathcal{G}} \boldsymbol{r}(i)\boldsymbol{d}2(i) \text{ is positive.}$$

Proposition

If ***d2*** is the more accurate lower bound, then
$\boldsymbol{d2}(MinState) > 0$ and $\boldsymbol{d}(MinState) > 0, \quad MinState \in \mathcal{F} \cap \mathcal{H}$.

Most accurate stochastic bounds

Proposition

$\boldsymbol{r}$ is increasing and $\boldsymbol{d}2 \leq_{st} \boldsymbol{d}$:

$$\sum_{i \in \mathcal{G}} \boldsymbol{r}(i)\boldsymbol{d}(i) - \sum_{i \in \mathcal{G}} \boldsymbol{r}(i)\boldsymbol{d}2(i) \text{ is positive.}$$

Proposition

If $\boldsymbol{d2}$ is the more accurate lower bound, then
$\boldsymbol{d2}(MinState) > 0$ and $\boldsymbol{d}(MinState) > 0, \quad MinState \in \mathcal{F} \cap \mathcal{H}$.

Lemma

$\boldsymbol{d}2$ is optimal distribution solution.

- For $\boldsymbol{d}$ defined over $\mathcal{H}$ and $\boldsymbol{d}2$ over $\mathcal{F}$, then

$$\mathcal{F} \subset \mathcal{H} \quad (i.e. \;\; \mathcal{G} = \mathcal{H}).$$

A Greedy Algorithm

Compute a lower distribution over K points.

Algorithm 1 Greedy (sometimes optimal) Lower Bounding

1: Begin with $\boldsymbol{d}2 = \boldsymbol{d}$ and $\mathcal{F} = \mathcal{H}$;

2: Compute $\boldsymbol{d}(i)(\boldsymbol{r}(i) - \boldsymbol{r}(\Gamma^-_{\mathcal{H}}(i)^2)), \quad \forall i \in \mathcal{H} \backslash \{MinState\}$;

3: Sort the results in increasing order;

4: Select the $(N - K)$ first states out of N to define **SelectSet**;

5: $\forall j \in$ **SelectSet**, $\quad \boldsymbol{d2}(\Gamma^-_{\mathcal{F}}(j)) = \boldsymbol{d2}(\Gamma^-_{\mathcal{F}}(j)) + \boldsymbol{d2}(\Gamma^-_{\mathcal{H}}(j))$;
Remove state j from $\mathcal{F}$ ($\boldsymbol{d}2(j)$ is not defined anymore).

[2] Predecessor of x ($\Gamma^-_{\mathcal{G}}(x)$): Biggest state in $\mathcal{G}$ smaller than x.

A Greedy Algorithm

Theorem

Algorithm provides ***d2*** which is a strong stochastic lower bound of ***d*** with support $\mathcal{F}$.

Complexity: $O(NlogN) \implies$ Sort operation.

A Greedy Algorithm

Theorem

Algorithm provides ***d2*** which is a strong stochastic lower bound of ***d*** with support $\mathcal{F}$.

Complexity: $O(NlogN) \Longrightarrow$ Sort operation.

- *But what about the optimality of the algorithm?*

Lemma

Removing two adjacent nodes $\Longrightarrow$ cumulated rewards costs more than two independent deletions.

$\Longrightarrow$ Optimality criterion is not always satisfied.

Optimal Algorithm based on dynamic programming

- Graph theory problem.
- Consider the weighted graph $G = (V, E)$ with:
  $$w(e) = \sum_{j \in \mathcal{H}: u < j < v} \boldsymbol{d}(j)(\boldsymbol{r}(j) - \boldsymbol{r}(u)) \quad \text{if} \quad v \in \mathcal{H}.$$
- Compute a shortest path P in G from state *MinState* to state *EndState* with K arcs.

Lemma

$\boldsymbol{d}_P$ defined over $\mathcal{F}$ such that $\boldsymbol{d}_P \leq_{st} \boldsymbol{d}$. The path P from state *MinState* to state *EndState* through all elements of $\mathcal{F}$ has weight:

$$\sum_{i \in \mathcal{H}} \boldsymbol{r}(i)\boldsymbol{d}(i) - \sum_{i \in \mathcal{F}} \boldsymbol{r}(i)\boldsymbol{d}(i).$$

Optimal Algorithm based on dynamic programming

- Graph theory problem.
- Consider the weighted graph $G = (V, E)$ with:
  $$w(e) = \sum_{j \in \mathcal{H}: u<j<v} \boldsymbol{d}(j)(\boldsymbol{r}(j) - \boldsymbol{r}(u)) \quad \text{if} \quad v \in \mathcal{H}.$$
- Compute a shortest path P in G from state *MinState* to state *EndState* with K arcs.

Lemma

$\boldsymbol{d}_P$ defined over $\mathcal{F}$ such that $\boldsymbol{d}_P \leq_{st} \boldsymbol{d}$. The path P from state *MinState* to state *EndState* through all elements of $\mathcal{F}$ has weight:

$$\sum_{i \in \mathcal{H}} \boldsymbol{r}(i)\boldsymbol{d}(i) - \sum_{i \in \mathcal{F}} \boldsymbol{r}(i)\boldsymbol{d}(i).$$

Algorithm Optimal Lower Bound

Guérin and Orda (2002): algorithm based on dynamic programming;

Complexity: $O(N^2K)$ and cubic when K has the same order as N.

An Example

- Well-known problem in performance evaluation: *Distribution of the completion time in a stochastic task graph*.

- Application of the proposed methodology but not an extensive comparison for stochastic task graphs.

An Example

- Well-known problem in performance evaluation: *Distribution of the completion time in a stochastic task graph*.

> • Application of the proposed methodology but not an extensive comparison for stochastic task graphs.

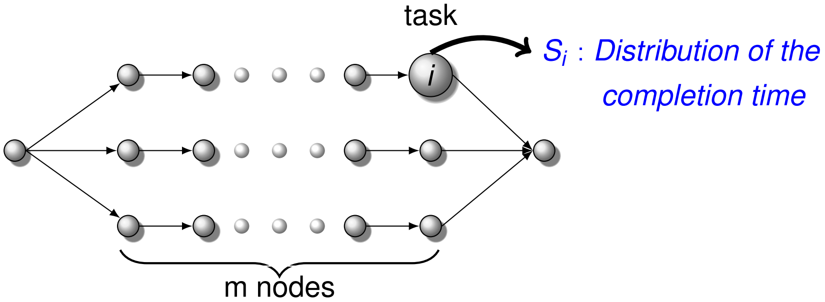

Figure: Task graph

The distribution of the completion time in a stochastic task graph

Task completion times: $T_i = \max_{j \in \Gamma_i^-} \{T_j\} + S_i$.

The distribution of the completion time in a stochastic task graph

Task completion times: $T_i = \max_{j \in \Gamma_i^-} \{T_j\} + S_i$.

Computation of the distribution require two operations:

$$\begin{cases} \bullet \ \textit{Addition} \longrightarrow \text{ Convolution;} \\ \bullet \ \textit{Maximum of random variables} \longrightarrow \text{ product of underlying pmf.} \end{cases}$$

Monotonicity of $(max, +)$ operations

Let x, y and z discrete random variables:

addition: $x \leq_{st} y \Longrightarrow x \otimes z \leq y \otimes z$.

Max: $x \leq_{st} y \Longrightarrow max(x, z) \leq max(x, z)$.

The distribution of the completion time in a stochastic task graph

Convolution example

We consider

- X, Y two independent random variables over $\mathcal{G}_X$ and $\mathcal{G}_Y$ resp.;
  $\mathcal{G}_X = \{1, 3, 5\}$ and $\mathcal{G}_Y = \{2, 5\}$;
  Probability distributions: $p_X = [0.2, 0.5, 0.3]$ and $p_Y = [0.6, 0.4]$.

- Resulting distribution
  $p_Z = p_X \otimes p_Y = [0.12, 0.3, 0.08, 0.18, 0.2, 0.12]$ defined over $\mathcal{G}_Z = \{3, 5, 6, 7, 8, 10\}$.

Convolution requires $O(|\mathcal{G}_X| \times |\mathcal{G}_Y|)$ operations (+) and at most $|\mathcal{G}_X| \times |\mathcal{G}_Y|$ states for the resulting distribution.

$\Longrightarrow$ Explosion on the size of the distribution of the results.

Analytic result

m	L	T	R_d
4	12160	0.7383	37.1455
5	46256	7.8542	43.3317
6	188416	415.1603	46.3308
7	785504	$8.3653\ 10^3$	46.5201
8	2974896	$2.4244\ 10^5$	56.1796

Table : Exact results

m	K	Greedy		(Locally)-Optimal	
		T	R_{d2}	T	R_{d2}
4	25	0.1125	35.6090	0.5781	36.3648
	50	0.1705	36.5403	3.7996	36.8294
5	25	0.1412	41.4151	0.8191	42.2156
	50	0.2484	42.5091	6.0513	42.8496
6	25	0.1793	43.6972	1.0872	45.0021
	50	0.3083	45.0599	8.1150	45.7225
7	25	0.2134	42.9925	1.3683	44.7492
	50	0.3697	45.0117	10.1021	45.7387
8	25	0.2552	52.2219	1.4004	53.9880
	50	10.5801	54.1708	13.4566	55.0026

Table : Bounds

Analytic result

m	L	T	R_d
4	12160	0.7383	37.1455
5	46256	7.8542	43.3317
6	188416	415.1603	46.3308
7	785504	$8.3653\ 10^3$	46.5201
8	2974896	$2.4244\ 10^5$	56.1796

Table : Exact results

m	K	Greedy		(Locally)-Optimal	
		T	R_{d2}	T	R_{d2}
4	25	0.1125	35.6090	0.5781	36.3648
	50	0.1705	36.5403	3.7996	36.8294
5	25	0.1412	41.4151	0.8191	42.2156
	50	0.2484	42.5091	6.0513	42.8496
6	25	0.1793	43.6972	1.0872	45.0021
	50	0.3083	45.0599	8.1150	45.7225
7	25	0.2134	42.9925	1.3683	44.7492
	50	0.3697	45.0117	10.1021	45.7387
8	25	0.2552	52.2219	1.4004	53.9880
	50	10.5801	54.1708	13.4566	55.0026

Table : Bounds

Analytic result

m	L	T	R_d
4	12160	0.7383	37.1455
5	46256	7.8542	43.3317
6	188416	415.1603	46.3308
7	785504	$8.3653\ 10^3$	46.5201
8	2974896	$2.4244\ 10^5$	**56.1796**

Table : Exact results

m	K	Greedy		(Locally)-Optimal	
		T	R_{d2}	T	R_{d2}
4	25	0.1125	35.6090	0.5781	36.3648
	50	0.1705	36.5403	3.7996	36.8294
5	25	0.1412	41.4151	0.8191	42.2156
	50	0.2484	42.5091	6.0513	42.8496
6	25	0.1793	43.6972	1.0872	45.0021
	50	0.3083	45.0599	8.1150	45.7225
7	25	0.2134	42.9925	1.3683	44.7492
	50	0.3697	45.0117	10.1021	45.7387
8	25	0.2552	**52.2219**	1.4004	**53.9880**
	50	10.5801	**54.1708**	13.4566	**55.0026**

Table : Lower Bounds

Conclusion

The proposed method consists to:

- Controls distribution sizes;
- Make a trade-off between accuracy and speed by changing distribution sizes.

Perspective:

Develop new applications in networks performance evaluation based on discretized histogram model.