Achieving COSMOS

A Metric for Determining When to Give up and When to Reach for the Stars

Emma Tosch[1] Lee Spector[1,2]

[1]University of Massachusetts
Amherst, MA 01003

[2]Hampshire College
Amherst, MA 01002

7 July 2012

Achieving COSMOS : A Metric for Determining When to Give up and When to Reach for the Stars

Is 100 runs enough?

Is 100 runs enough?
Or:

Is 100 runs enough?

Or:

Have I done enough runs to make reliable inferences *using this system*?

Koza's Measures

M : population size/indexable array of ordered individual programs
s : success predicate (problem-specific)

Name	Description	Formula/Estimator
$\mathcal{I}(M, i, \delta)$	Computational Effort	$\mathcal{R}(\delta) \times i \times M$
$\mathcal{R}(\delta)$	Number of independent runs needed to satisfy s with probability δ	$\mathcal{R}(\delta) = \left\lceil \frac{\log \epsilon}{\log(1-P(M,i))} \right\rceil$
$\mathcal{P}(M, i)$	Cumulative Probability of Success	$\sum_i \mathcal{Y}(M, i)$
$\mathcal{Y}(M, i)$	Point Probability of Success	$\sum_k s(M_i(0))$

Koza's Measures

M : population size/indexable array of ordered individual programs
s : success predicate (problem-specific)

Name	Description	Formula/Estimator
$\mathcal{R}(\delta)$	Number of independent runs needed to satisfy s with probability δ	$\mathcal{R}(\delta) = \left\lceil \frac{\log \epsilon}{\log(1-P(M,i))} \right\rceil$

COSMOS

COSMOS

Convergence

of

COSMOS

Convergence

of

Sample

Means for

COSMOS

Convergence

of

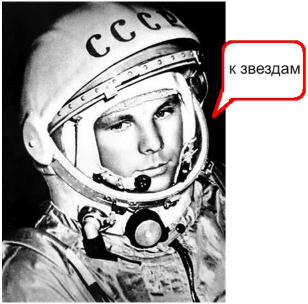
Sample

Means for

Order

Statistics

COSMOS



Convergence

of

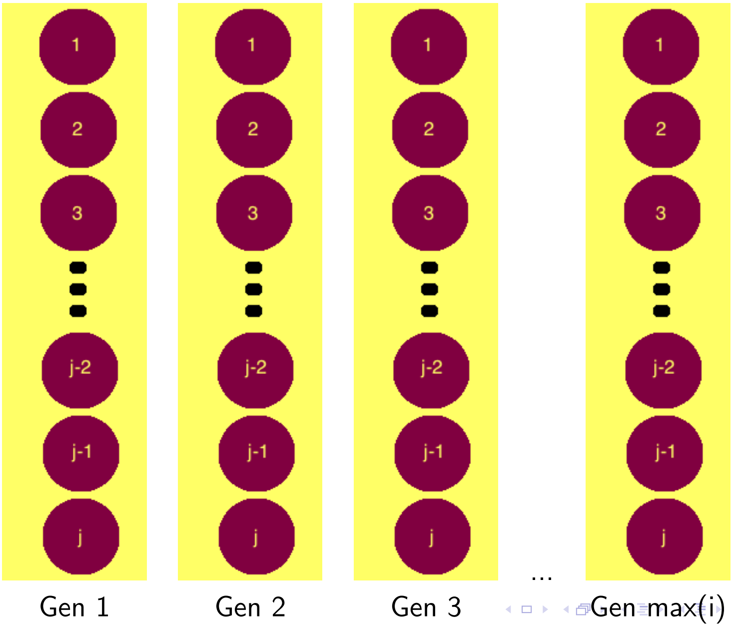
Sample

Means for

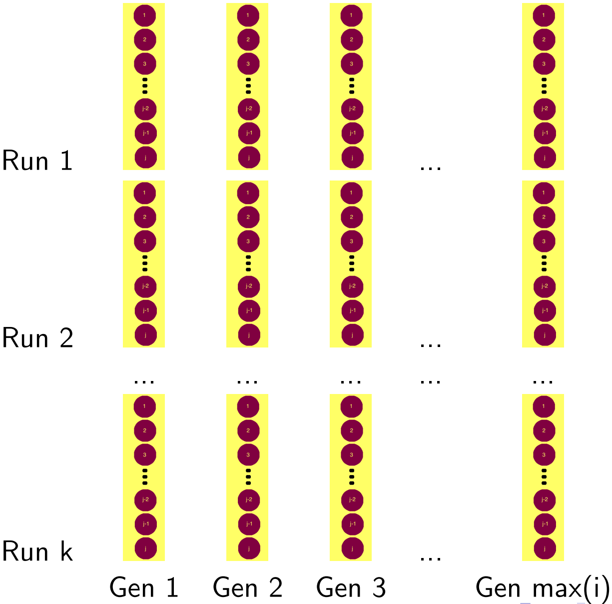
Order

Statistics

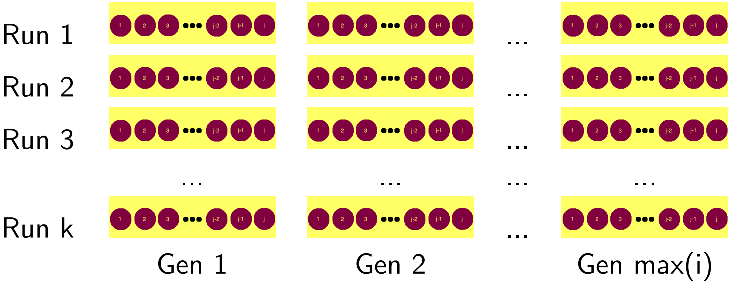
Generational Model



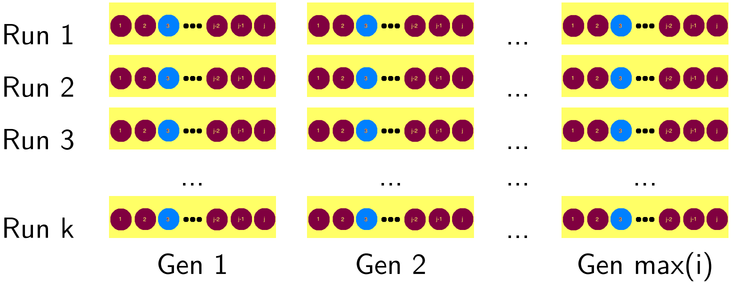
Generational Model across *k* runs



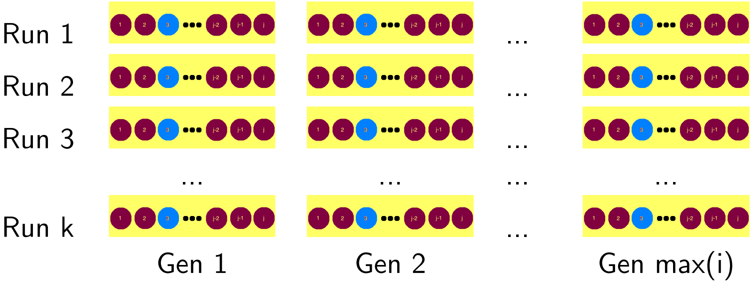
Generational Model across k runs



Select the third ordered individual...



...and calculate the sample mean



$\frac{\sum_k error(3,1,k)}{k}$ $\frac{\sum_k error(3,2,k)}{k}$ $\frac{\sum_k error(3,\max(i),k)}{k}$

What might convergence for some sample mean might look like?

What might convergence for some sample mean might look like?

...say, the third ordinal of the second generation?

How to define Convergence?

Convergence for ordinal q in generation i

Let:

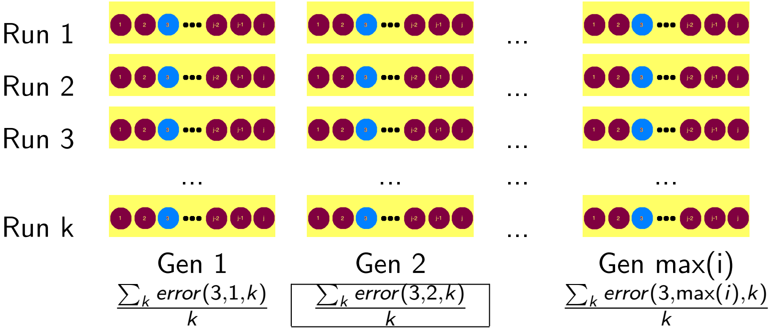
$\overline{X}_k$	$\overline{X}_{k+1}$
Mean of k random samples	Mean of $k+1$ random samples

We want:

$$\overline{X}_k = (1 \pm \epsilon)\overline{X}_{k+1}$$

for some error ϵ with high probability.

Recall...


Run 1
Run 2
Run 3
...
Run k
Gen 1
$\frac{\sum_k error(3,1,k)}{k}$
Gen 2
$\frac{\sum_k error(3,2,k)}{k}$
Gen max(i)
$\frac{\sum_k error(3,\max(i),k)}{k}$

Convergence for ordinal 3 in generation 2

Let:

Convergence for ordinal 3 in generation 2

Let:

$$\overline{X}_k = \frac{\sum_k error(3, 2, k)}{k}$$

$$\overline{X}_{k+1} = \frac{\sum_{k+1} error(3, 2, k+1)}{k+1}$$

Convergence for ordinal 3 in generation 2

Let:

$$\overline{X}_k = \frac{\sum_k error(3,2,k)}{k}$$

$$\overline{X}_{k+1} = \frac{\sum_{k+1} error(3,2,k+1)}{k+1}$$

We want:

$$\overline{X}_k = (1 \pm \epsilon)\overline{X}_{k+1}$$

for some error ϵ with high probability.

Idea: Want to detect *when* doing more runs doesn't give us any new information.

So why Order Statistics?

So why Order Statistics?

- We know nothing about the underlying distribution

So why Order Statistics?

- We know nothing about the underlying distribution
- Means not that useful for most distributions

So why Order Statistics?

- We know nothing about the underlying distribution
- Means not that useful for most distributions
- Commonly used in nonparametric statistics for their robustness

Selected Order Statistics

- min
- “lower quartile”
- “median”
- “upper quartile”
- max

COSMOS Estimator

For each generation:

1. Create an array to hold the sample means of the ordinals.
2. For each ordinal:
   a. Start with some minimum number of runs and compute its mean.
   b. Add one run to the total and recompute the mean.
   c. Take the ratio of the two sample means and enter it into the array.
3. If the ratio is within 1 $\pm\epsilon$ for all ordinals, return the number of runs.
4. Else, repeat.

Return the maximum number of recommended runs.

Context

Repeating experiments from (Luke 2001) and (Daida 2001)

Context

Repeating experiments from (Luke 2001) and (Daida 2001)

- All experiments used a GP System (ECJ)

Context

Repeating experiments from (Luke 2001) and (Daida 2001)

- All experiments used a GP System (ECJ)
- Fitness values were default scaled errors

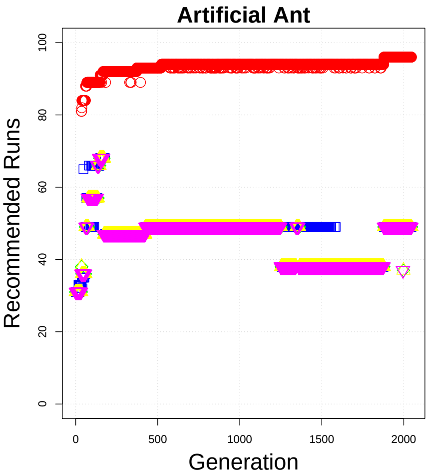
Context

Repeating experiments from (Luke 2001) and (Daida 2001)

- All experiments used a GP System (ECJ)
- Fitness values were default scaled errors
- GP system was generational (steady-state version is simpler!)

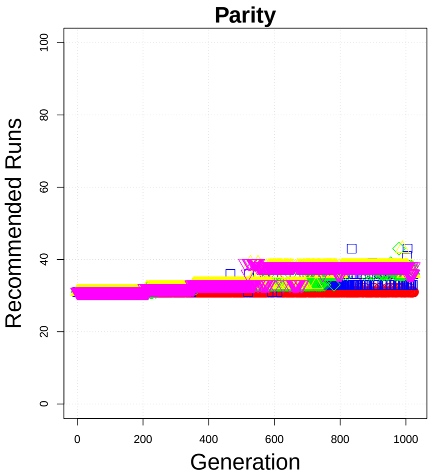
Artificial Ant

Ordinals	
○	Minimum
□	Lower Quartile
◇	Median
△	Upper Quartile
▽	Maximum



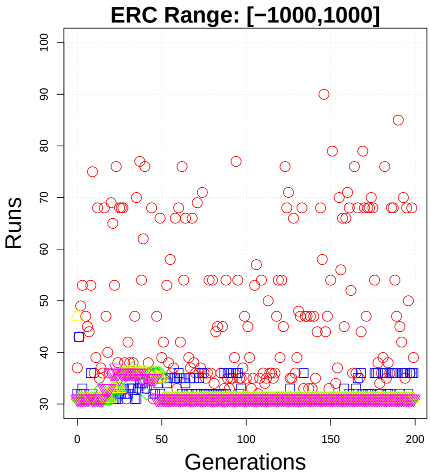
Parity

Ordinals	
○	Minimum
□	Lower Quartile
◇	Median
△	Upper Quartile
▽	Maximum



Binomial-3 ERC Range: [-1000,1000]

Ordinals	
○	Minimum
□	Lower Quartile
◇	Median
△	Upper Quartile
▽	Maximum



Future Work

Future Work

- Obvious shortcomings of the estimator

Future Work

- Obvious shortcomings of the estimator
- Proving bounds on the estimator

Future Work

- Obvious shortcomings of the estimator
- Proving bounds on the estimator
- Lower bounds on runs

Future Work

- Obvious shortcomings of the estimator
- Proving bounds on the estimator
- Lower bounds on runs
- Utility of non-convergence?

Acknowledgements

- Workshop coordinators
- My advisor, Lee Spector
- Alexandre Passos, Luke Vilnis, Kendra Paradis

This work is supported in part by the National Science Foundation under NSF grants #CNS-0619337 and #1017817.