Simulation optimization via bootstrapped Kriging

Jack P.C. Kleijnen

Department of Information Management /
Center for Economic Research (CentER)
Tilburg School of Economics & Management (TiSEM)
Tilburg University, Tilburg, Netherlands

Workshop on deterministic & stochastic simulation
Isaac Newton Institute, Cambridge, U.K., 5-9 Sept. 2011

Overview

Scope: *Deterministic* & *stochastic* simulations

Focus: Optimization via *Kriging* (meta)model
  Analysis of Kriging model: *Bootstrapping*

Two *types* of bootstrapping:

- Stochastic simulation: Replicates → *distribution-free* bootstrap
- Deterministic simulation: No replicates → *parametric* bootstrap (multivariate Gaussian); parameters estimated from simulation I/O

Kriging: Stochastic simulation

1. *Geo*statistics: *Nugget* / *Measurement* error
2. *Deterministic* simulation: *Numerical* noise

Sub 1 & 2: Kriging model Y(x) = μ + Z(x) + e

with GP Z(x) & White Noise e ~ NIID[0, σ(e)]

3. *Random* simulation: σ(e) → σ[x(i)] & CRN

Sub 3 (1 & 2): Predictor for new point x(n+1):

Y^[x(n+1)] = μ + Σ'(n+1)[Σ + Σ(ebar)]$^{-1}$(ybar - 1μ)

with ybar average of m(i) replicates of point i

Σ(ebar) diagonal, unless CRN

Not an exact *interpolator* of n averages

Kriging parameters: MLE

MLE for (hyper)parameters: μ, σ^2, $\Sigma(\theta)$, $\Sigma(ebar)$

Plug-in MLE: *Non-linea*r Kriging predictor

$Y\hat{}\hat{}[x(n+1)] = \mu\hat{} + \Sigma\hat{}'(n+1)[\Sigma\hat{} + \Sigma\hat{}(ebar)]^{-1}(ybar - 1\mu\hat{})$

Biased estimator of predictor variance:

$E[s^2\{Y\hat{}(x)\}] \neq \sigma^2\{Y\hat{}\hat{}[x(n+1)]\}$

Random simulation: Estimate $\Sigma(e)$ from replicates

CRN: # replicates m > # combinations n

$s^2\{Y\hat{}(x)\}$ and $s^2\{Y\hat{}\hat{}(x)\} = s^{2*}$ not *max* at same x;

see bootstrap s^{2*} (slide 5) and EGO (slide 6)

Parametric bootstrap for $s^2\{Y\hat{}\hat{}(x)\}$

1. *Original* I/O (**X**, **y**) gives original MLE (μ^, σ^, **θ**^)
2. *Sample* (y*(1), …, y*(n), y*(n + 1)) from N(***μ***^, **Σ**^) with
   μ^: all (n + 1) elements equal to μ^
   Σ^: (n + 1) x (n + 1) matrix; see paper, eq. (9)
3. *Bootstrapped* (y*(1), …, y*(n)) (Step 2) with **X** (Step 1***)***
   gives bootstrap MLE μ*****, σ*, **θ*** (see Step 1)
4. Use Step 3 to compute bootstrap *predictor* y*^^(n + 1)
5. Use Steps 4 & 2 to compute Squared Error
   SE = [y*^^(n + 1) - y*(n + 1)]^2
6. Repeat Steps 2-5, B times: $s^{2*} = \Sigma_b$ SE(b) / B

Example: Circuit simulator in Sacks et al. (1989), EGO

EGO/EI with bootstrap variance

EGO: Global 'exploration' / 'local' exploitation

Assume: *Deterministic* simulation; *single* output

1. Find y^o, minimum among n old outputs
2. Find $\mathbf{x}^o$, maximizer $\mathbf{x}$ of
   EI($\mathbf{x}$) = $E[y^o - y(\mathbf{x}) \mid y(\mathbf{x}) < y^o]$ with
   $y(\mathbf{x}) \sim N(\hat{\hat{y}}, s^2\{\hat{y}(x)\})$
   Find $\mathbf{x}^o$ via candidate set or Genetic Algorithm
3. Simulate $\mathbf{x}^o$; refit Kriging; go to 1 until EI ≈ 0

Alternative: *Bootstrapped* estimator s^{2*} (slide 5)
   Result: Better in 3 of 4 test functions; one tie

Constrained optimization

Example: Call center (min. cost, such that service > .95)

Goal output y(0): Min E[y(0**, x**)] (e.g., cost)

Other r – 1 constrained outputs: E[y(h, **x**)] ≥ c(h)

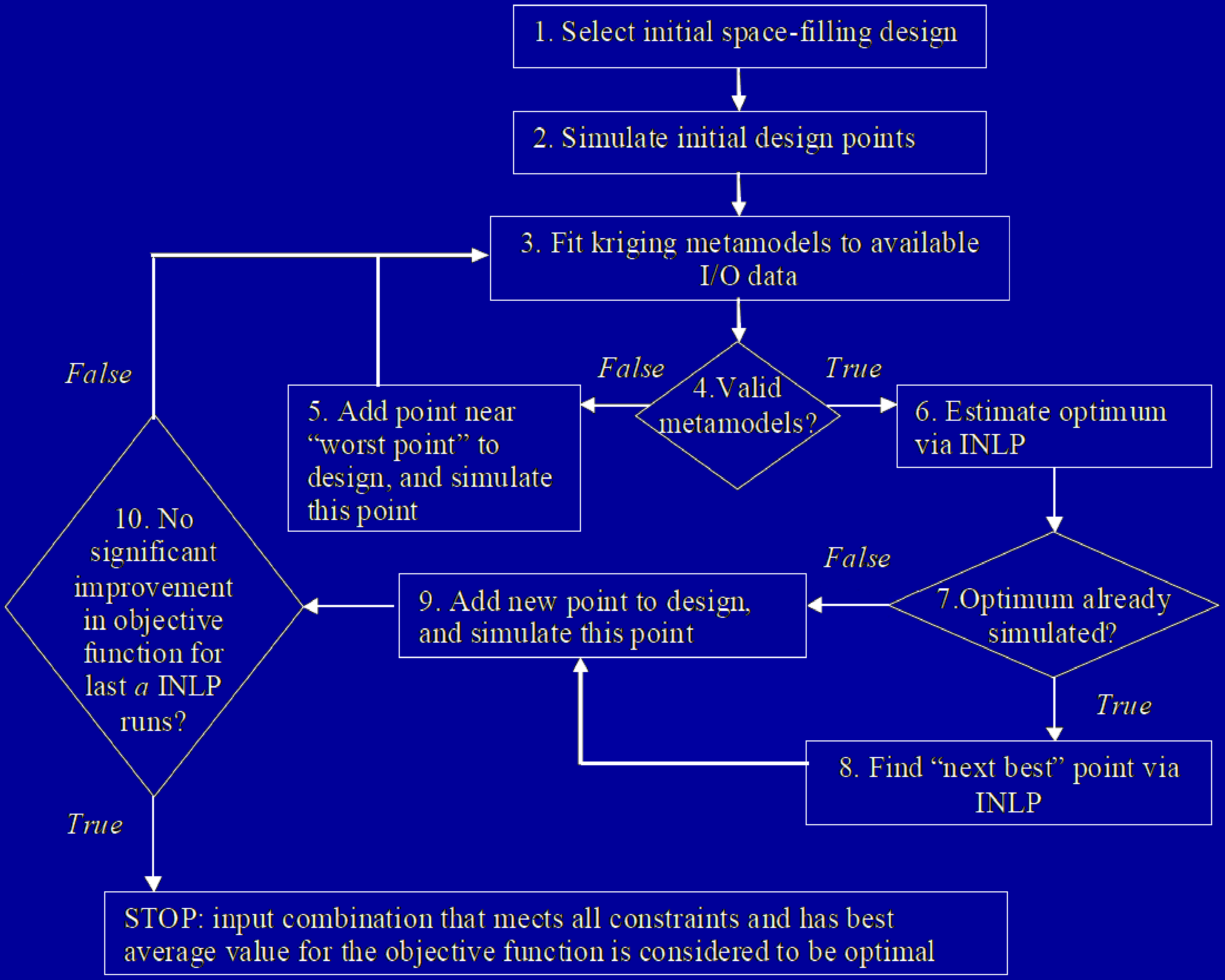
s constraints for d inputs: f(g)[x(1),…,x(d)] ≥ c(g)

Non-negative integer inputs: x(j) ∊ N

Solution combines (see next slide)

- Sequential DOE (like EGO)
- Kriging & bootstrap (like EGO)
- Integer Non-Linear Programming (INLP)

Note: Adapt for *continuous* inputs, incl. *gradients* & *deterministic* outputs


1. Select initial space-filling design
2. Simulate initial design points
3. Fit kriging metamodels to available I/O data
4.Valid metamodels?
False
True
5. Add point near “worst point” to design, and simulate this point
6. Estimate optimum via INLP
7.Optimum already simulated?
False
True
8. Find “next best” point via INLP
9. Add new point to design, and simulate this point
10. No significant improvement in objective function for last *a* INLP runs?
False
True
STOP: input combination that meets all constraints and has best average value for the objective function is considered to be optimal

Monotonic bootstrapped Kriging

Practice: I/O function known to be *monotonic*
Example: Queuing simulation's mean & quantile

Assume: *Random* simulation with *replications*

Distribution-free bootstrapped Kriging (next slide)

Result: *Confidence intervals* with higher coverage and similar width

Future research:

- Replace classic Kriging by *stochastic* Kriging
- Preserve *convexity* or *nonnegativity*
- *Deterministic* simulation: Parametric bootstrap

Procedure for monotonic Kriging

1. *Resample* -- with replacement -- the m IID original w(i, r): *w*(i, r)* [i = 1, …, n; r = 1, …, m]
2. From w*(i, r) compute the *average* wbar*(i)
3. From (**X**, **wbar***) compute *Kriging* y(θ*)
4. Accept only *monotonically* increasing Kriging: Positive *gradients* (∇y(i)* > 0 with i = 1,…,n)
5. Repeat B times; *sort* B' predictions y*[**x**(n + 1)]

Point estimator: *Median* of B' predictions

90% confidence interval (CI):

Lower/upper limit: *5 / 95% quantile* of B' pred.

Asymmetric CI; *positive* lower limit

Conclusions

- *EGO*: Parametric bootstrap estimator of variance of Kriging predictor
- *Constrained* opt. in *random* simulation: Distribution-free bootstrap for validation of Kriging model (giving opt. via INLP)
- *Monotonic* bootstrapped Kriging

The End

Kriging: Basics

Kriging: *Global* model

$Y(x) = \mu + Z(x)$

with stationary GP $Z(x)$ with zero mean

$\text{corr}[Y\{x(i)\}, Y\{x(j)\}] = \prod \exp[-\theta(k)\{x(ik) - x(jk)\}^2)$

Linear predictor for point $x(n+1)$:

$Y\hat{}[x(n+1)] = \mu + r'R^{-1}(y-1\mu)$

Exact interpolator: $y\hat{}[x(i)] = y[x(i)]$ with $i = 1, \ldots, n$

Predictor variance:

$$\sigma^2[1- r'R^{-1}r + \{(1 -1'R^{-1}r)^2)/(1'R^{-1}1)\}]$$

Parametric bootstrap: Basics

Data driven statistical method

Examples: Give n IID observations y(i) (i = 1, …,n)

a. Mean E(y) of y(i) ~ Exp(λ)
b. Skewness: $\sum (y(i) - ybar)^3/[(n - 1)s^3]$

Sub a:

1. Estimate λ^ = 1/ybar
2. Sample y* from Exp(λ^): *Parametric* bootstrap
3. Estimate mean: y*bar = ∑ y*(i)}/n
4. Repeat Steps 2-3: y*bar(b) (b = 1, …, B)
5. Sort y*bar(b): y*bar(1) < …< y*bar(B)
6. 90% CI for mean: y*bar(0.05B), y*bar(0.95B)

Distribution-free bootstrap: Basics

1. Resampling with replacement y(i) gives y*(i)
   Example: y(1) is sampled 0, 1, …, n times
2. Estimate mean: y*bar = ∑ y*(i)/n
3. Repeat Steps 2-3: y*bar(b) (b = 1, …, B)
4. Sort y*bar(b): y*bar(1) < …< y*bar(B)
5. 90% CI for mean: y*bar(0.05B), y*bar(0.95B)

Bootstrap: Applications

Bootstrap: simple idea; yet, “art” of modeling

Aplications:

1. CI for estimated skewness (Example b)
2. Validation of simulation models
3. Ranking of journals on quality (citations)
4. See other slides

Constrained optimization: details

- “Enough” replicates per point; see Law (2007)
- CRN
- Fit Kriging to averages

Global/local: Steps 5 /9 in flowchart

Cross-validate n(cv) points incl. bootstrap: Step 4
Complications in bootstrap:
Multiple outputs, non-constant m(i), CRN

Studentized prediction error: Divide by √ of bootstrapped variance + replicate var. estimate

Apply *Bonferroni*’s inequality: $\alpha/[r \times n(cv)]$

Constrained optimization: details

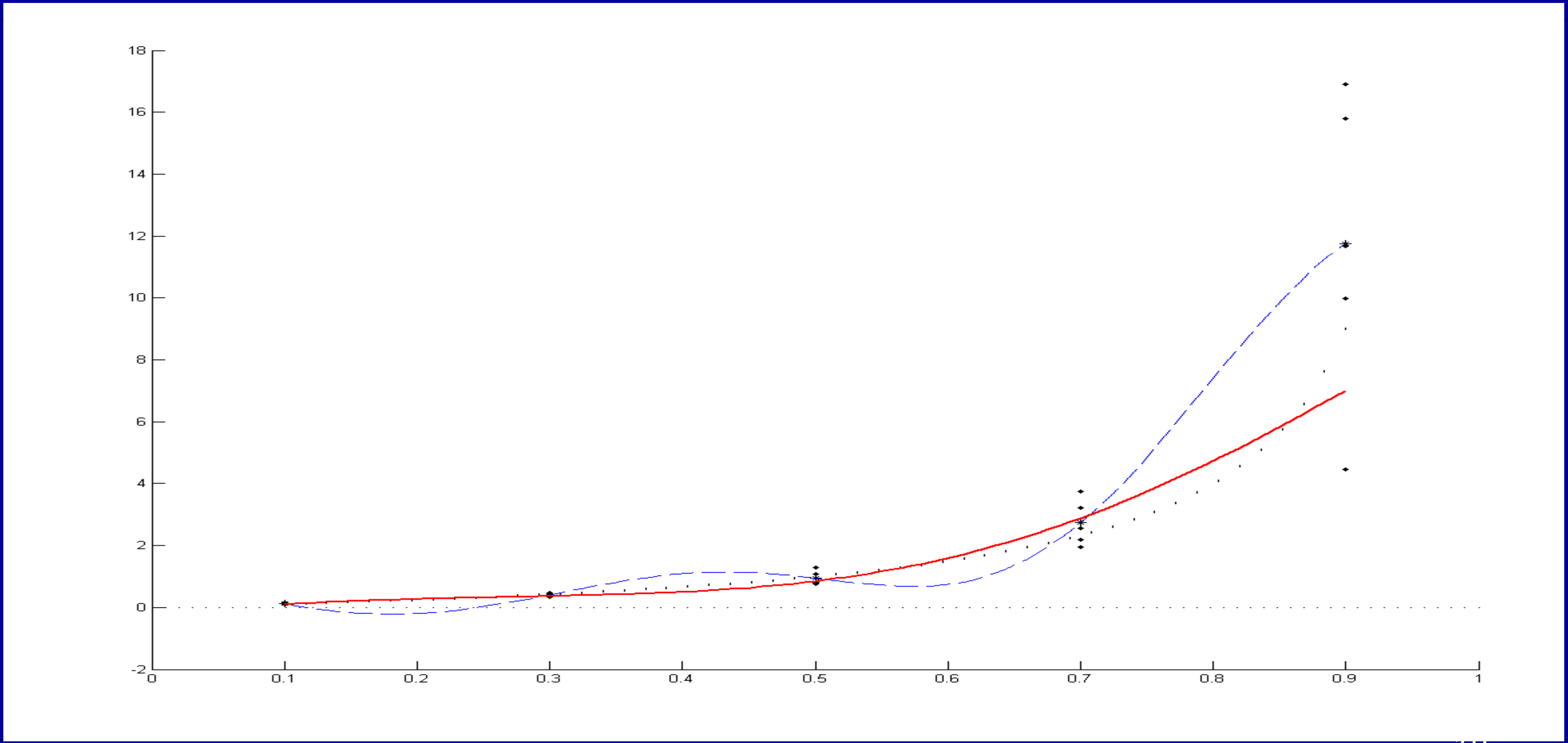
Step 5: If Kriging is rejected, then add point halfway worst point and nearest neighbor

Applications:

- Academic *inventory* system (s, S)
- Realistic *call center*
- ‘Better’ than *OptQuest* (in Arena)

M/M/1 example:
Wiggling versus monotonic Kriging

m = 5; no CRN; wbar(i) < wbar(i + 1);
Gaussian correlation function



Bootstrap monotonic vs Ordinary Kriging