Some Topics in Optimization for Simulation

Michael Fu
University of Maryland

The ACNW OPTIMIZATION TUTORIALS
June 13, 2005

Tres Tapas

- A Soft Overview:
  Optimization for Simulation (JOC & Annals OR papers)

- A Little Tutorial:
  Stochastic Gradient Estimation (book, handbook chapter)

- Something New:
  Global Optimization Algorithm (paper under review at OR)

Part I: Overview of Simulation Optimization

- General Problem Setting
- Motivating Examples:
  - Service Sector: Call Center Design
  - Financial Engineering: Pricing of Derivatives
  - Academic: Single-Server Queue, (s,S) Inventory System
- Difference from deterministic optimization
- Research: Main Methods
- Practice: Implemented Software

Problem Setting

- Find the best parameter settings
  to minimize (or maximize)
  an OBJECTIVE FUNCTION
  [possibly subject to constraints]

  $$\min_{\theta \varepsilon \Theta} J(\theta)$$

- Key: OBJECTIVE FUNCTION contains quantities that must be estimated from stochastic simulation output: $J(\theta) = E[L(\theta,\omega)]$

Example: Call Center Design/Operation

- Multiple sources of jobs (multichannel contact)
  - voice, e-mail, fax, interactive Web
- Multiple classes of jobs
  - e.g., address change vs. account balance or payment
  - customer segmentation according to priorities or preferences
- Stochastic elements:
  - arrivals (timing and type)
  - service (length of time, efficiency of operator)

Example: Call Center (cont.)

- OBJECTIVE FUNCTION: Performance Measures …
  - Customers: waiting time, abandonment rate, % blocked calls
  - Service facility: operator wages & efficiency, trunk utilization

- Controllable parameters
  - agents: number and type (training)
  - routing of calls (FCFS, priority, complex algorithms)
  - configuration of call center (possibly a network)

- Basic tradeoff: customer service vs. cost of providing service

Example: Options Pricing

- American-style (aka Bermudan) Call Options
  - the right (but not the obligation) to buy an asset at a certain (strike) price on certain (exercise) dates
- Optimization Problem: when to exercise the right
  - Objective: maximize expected payoff
  - Decision: exercise or hold for each exercisable date
- Optimal Stopping Problem in Stochastic Dynamic Programming
  - one solution approach: parameterize exercise boundary
  - optimize w.r.t. parameters

Academic ("Toy") Examples

- Single-Server Queue
  - Minimize $E[W(\theta)] + c/\theta$,
  - W is waiting time, θ is mean service time
- (s,S) Inventory System
  - When inventory falls below s, order up to S
  - Minimize total cost: order, holding, backlogging

What makes simulation optimization hard?

- Ordinary optimization can concentrates on the search.
- Due to the stochastic nature of the problem, there is both search and evaluation.

- Trade-off between
  finding more candidate solutions
  vs. obtaining a better estimate of current solutions

  i.e.,

  finding $\arg\min_{\theta \varepsilon \Theta} J(\theta)$ vs. estimating $J(\theta)$

How else does it differ from “Ordinary” Optimization?

- Key Characteristic:
  Output performance measures estimated via stochastic simulation that is <u>EXPENSIVE</u>, (nonlinear, possibly nondifferentiable)

  i.e., a *single* simulation replication
  may take as long to run
  as a typical LP model

Simulation Optimization

- Approaches (Banks et al. 2000)
  - Guarantee asymptotic convergence to optimum.
  - Guarantee optimality under deterministic counterpart.
  - Guarantee a prespecified probability of correct selection.
  - Robust heuristics.
- Theory has concentrated on the first three.
- Practice has implemented many variations of the last one, with some implementations of the next to the last one in the case of a finite set of pre-selected alternatives.

Main Approaches

- Stochastic Approximation (Gradient-Based)
- Response Surface Methodology
- Sample Path Optimization (Robinson et al., Shapiro et al.)
  (aka Stochastic Counterpart, Sample Average Approximation)
- Random Search Algorithms
- Ordinal Optimization
- Others: Nested Partitions (Shi et al.), COMPASS (Hong/Nelson),
  Statistical Ranking & Selection, Multiple Comparisons
- Approaches in Deterministic Optimization
  - Genetic Algorithms (evolutionary approaches)
  - Tabu Search
  - Neural Networks
  - Simulated Annealing

Some Basics: Convergence Rates

$$1/\sqrt{n}$$

n = # simulation replications (samples)

- 100 times the effort for additional decimal place of accuracy
- also best possible asymptotic convergence rate for SA

let Z_n denote the sample mean (bar{Y}_n)

- CLT: $Z_n \rightarrow N(J,\sigma^2/n)$ in distribution
- large deviations: $P(|Z_n-J|>\varepsilon) \rightarrow \exp[-nR(\varepsilon)]$

Stochastic Approximation

$$\theta_{n+1} = \prod_{\Theta}(\theta_n + a_n \hat{g}(\theta_n))$$

Key point: How to find the gradient estimate $\hat{g}(\theta_n)$?

PREVIEW OF COMING ATTRACTIONS (Part II)

Direct Estimators

- Unbiased: PA, LR, WD

Indirect Estimators

- brute-force finite differences (FD, SD)
- simultaneous perturbations: (SPSA)

all parameters *simultaneously randomly* perturbed

Stochastic Gradient Estimation Approaches

approach	# simulations	key features	disadvantages
IPA	1	highly efficient, easy to implement	limited applicability
other PA	often > 1	model-specific	more difficult to apply
LR/SF	1	requires only model input distributions	possibly high variance
WD	2*(# appearances of parameter)	requires only model input distributions	possibly large # simulations
SD FD (one-sided)	2*p p+1 (dimension)	widely applicable, model-free	noiser, biased, large # simulations
SP	2	widely applicable, model-free	noiser, biased

Response Surface Methodology (RSM)

- sequential procedure (vs. metamodeling)
- design of experiments
- regression model

- basic form:
  - linear model and move in gradient direction until gradient approximately zero
  - quadratic model to find optimum

- Stay tuned: more details from our next speaker!

Sample Path Optimization

- aka stochastic counterpart, sample average approximation

- basic form:
  - run many replications (samples) of system
  - store sample paths in memory
  - optimize using arsenal of tools from mathematical programming, nonlinear optimization

- Stay tuned: discussion in afternoon panel?

Random Search Algorithms

- Key element: definition of neighborhood

- basic form:
  - Initilize: select initial best θ^*.
  - Select another θ_i according to prob dist on neighborhood.
  - Perform simulations to estimate $J(\theta^*)$ and $J(\theta_i)$.
  - Increase counter n_θ for the better one (and update current best θ^* if necessary).
  - Return arg max n_θ (i.e., θ with highest counter)

Ordinal Optimizaton

- Main Idea:
  comparison is easier than estimation
  (no difference in relative and absolute performance for deterministic opt)

- recall estimation convergence rate of $1/\sqrt{n}$
  vs. large deviations exponential convergence rate:
  – which is better? ⇔ difference < 0 or difference > 0

    prob of making wrong decision → 0 exp fast

Practice

- At present, nearly every commercial discrete-event simulation software package contains a module that performs some sort of ``optimization'' rather than just pure statistical estimation. Contrast this with the status in 1990, when none of the packages included such an option.
- The most recent editions of two widely used discrete-event simulation textbooks have added new sections on the topic.
- “Simulation Optimization” is a new entry in the 2nd edition, Encyclopedia of Operations Research & Management Science.

Commercial Software

- OptQuest (Arena, Crystal Ball, et al.)
  - standalone module, most widely implemented
  - scatter search, tabu search, neural networks
- AutoStat (AutoMod from Autosimulations, Inc.)
  - part of a complete statistical output analysis package
  - dominates semiconductor industry
  - evolutionary (variation of genetic algorithms)
- OPTIMIZ (SIMUL8): neural networks
- SimRunner (ProModel): evolutionary
- Optimizer (WITNESS): simulated annealing, tabu search

Theory vs. Practice

- Practice
  - Robust heuristics
  - Concentration on search
  - Family of solutions
  - Use of memory
- Theory (predominantly)
  - Provable convergence
  - Sophisticated mathematics
  - Single point

Closing on Part I: Op-Ed on Needs

- Better Integration of Search/Opt & Evaluation (instead of separated, exploit interaction)
- Statistical Statements for Metaheuristics
  - Recent Work of Barry Nelson with OptQuest?
- Inclusion of Variance Reduction
- Efficient Allocation of Simulation Budget
- Other Uses of Gradient Estimation
  (lead in to Part II)

- More discussion in afternoon panel session!

Part II: Stochastic Gradient Estimation

- Simulation & the Law of the Unconscious Statistician
- Derivatives of Random Variables & Measures
- Techniques
  — Perturbation Analysis (PA): IPA and SPA
  — Likelihood Ratio (LR) Method
  — Weak Derivatives (WD)

  — Simultaneous Perturbations Stochastic Approximation (SPSA)
- Examples
  — Simple Random Variables
  — Single-Server Queue

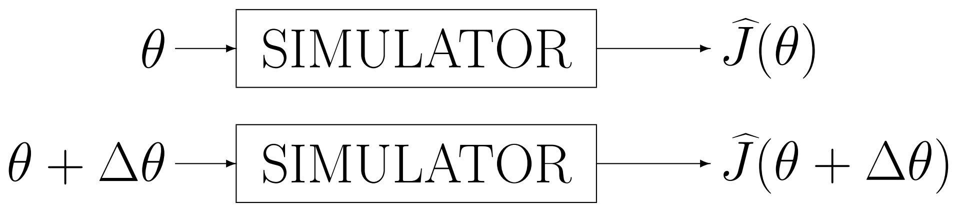
GRADIENT ESTIMATION PROBLEM

$$
\begin{aligned}
J(\theta) &= E[Y(\theta,\omega)] \quad = \quad \text{performance measure}, \\
Y(\theta,\omega) &= \text{sample performance}, \\
\omega &= \text{stochastic effects (sample path)}, \\
\theta &= \text{controllable vector of } p \text{ parameters}.
\end{aligned}
$$

GOAL: Find (unbiased/consistent) estimator for $\nabla_\theta J(\theta)$.

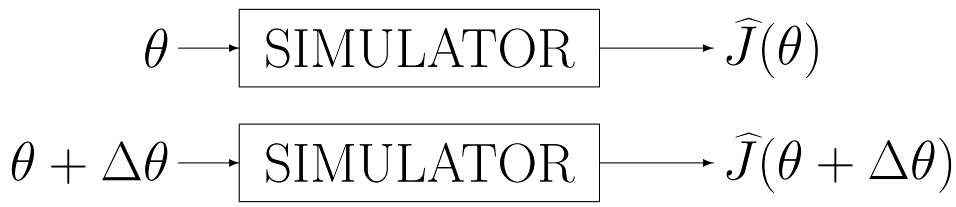
Example: Queueing Systems
Y is average time in queue over some number of customers served,
θ is parameters in the interarrival and service time distributions,
or routing probabilities.

TRADITIONAL Approach
(“BRUTE FORCE”)

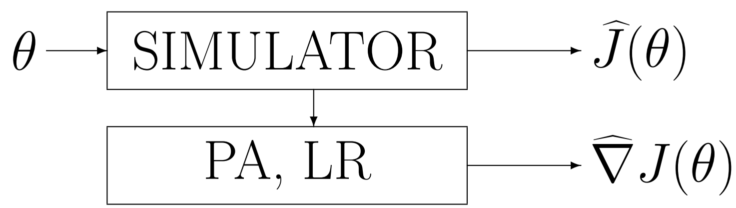


SPSA provides an opportunity for substantial improvement.

INDIRECT GRADIENT ESTIMATION



DIRECT GRADIENT ESTIMATION



WD does NOT fall into this category!
SPA may require additional simulation, too.

Simultaneous Perturbation Stochastic Approximation (SPSA)

- uses two simulations per iteration, regardless of dimension!
- model independent (treats simulation as black box)
- easy to implement, **simultaneously** perturb all parameters, e.g., symmetric Bernoulli r.v. ($\pm\Delta$)

$$\text{old SD}: \quad \frac{Y(\theta_n + c_n e_i) - Y(\theta_n - c_n e_i)}{2c_n},$$

$$\text{new SP}: \quad \frac{Y(\theta_n + c_n \Delta_n) - Y(\theta_n - c_n \Delta_n)}{2c_n (\Delta_n)_i}.$$

Simulation & the Law of the Unconscious Statistician

$$J(\theta) = E[Y(\theta)] = E[Y(X_1, \ldots, X_T)],$$

$\{X_i\}$ input r.v.s, T fixed constant.

Simulation Goal: Estimate

$$E[Y] = \int y \cdot dF_Y(y), \qquad F_Y \text{ UNKNOWN}$$

Law of the Unconscious Statistician:

$$E[Y(\mathbf{X})] = \int Y(\mathbf{x}) dF_{\mathbf{X}}(\mathbf{x}), \qquad F_{\mathbf{X}} \text{ KNOWN.}$$

Simulator **GENERATES** $Y(\mathbf{X})$

Where is Spot (aka θ)?

$$E[Y(X)] = \begin{cases} \int_{-\infty}^{\infty} Y(x) f(x;\theta) dx \\ \\ \int_0^1 Y(X(\theta;u)) du \end{cases}$$

Example: Queueing Systems
f is p.d.f. of service time,
X is inverse transform for service time ($-\theta \ln u$ for exponential).

Temporarily, for expositional purposes, assume

- θ **scalar**;
- **single** input r.v. X
- X continuous valued with **density** (p.d.f.) f.

Taking Derivatives

$$E[Y(X)] = \begin{cases} \int_{-\infty}^{\infty} Y(x) f(x;\theta) dx \\ \\ \int_0^1 Y(X(\theta;u)) du \end{cases}$$

Assuming interchange of expectation and differentiation:

$$\frac{dE[Y(X)]}{d\theta} = \begin{cases} \int_{-\infty}^{\infty} Y(x) \frac{df(x;\theta)}{d\theta} dx \\ \int_0^1 \frac{dY(X(\theta;u))}{d\theta} du, \end{cases}$$

What needs to be smooth w.r.t. parameter θ?
(dominated convergence theorem)
(i) input distributions,
(ii) sample performance function.

Infinitesimal Perturbation Analysis (IPA): Sample (Path) Derivatives & the Chain Rule

$$\frac{dE[Y(X)]}{d\theta} = \int_0^1 \frac{dY(X(\theta;u))}{d\theta} du = \int_0^1 \frac{\partial Y}{\partial X} \frac{dX(\theta;u)}{d\theta} du.$$

Estimator:

$$\frac{\partial Y(X)}{\partial X} \frac{dX}{d\theta}$$

Many input random variables:

$$\sum_i \frac{\partial Y(X)}{\partial X_i} \frac{dX_i}{d\theta}$$

KEY OBSERVATION: smoothness conditions on Y

IPA

KEY IDEA: small changes in parameter cause small changes in sample performance, i.e., sample performance function a.s. continuous; allow dominated convergence theorem to exchange limit (derivative) and integral (expectation).

Sources of Discontinuity:

- system $\rightarrow$ **the commuting condition**
- performance measure (e.g., indicator functions)
- parameter (e.g., in certain discrete distributions)

Smoothed Perturbation Analysis (SPA)

KEY IDEAS:

- conditional expectation smoothes:

$$E\left[\nabla E[L(\theta)|Z]\right] = \nabla E[E[L(\theta)|Z]] = \nabla J(\theta)$$

- ESTIMATOR:
  IPA term + conditional contribution:
  $\boxed{\text{jump rate}} \times \boxed{\text{jump conditional expectation}}$

- how to pick Z s.t. $E[L(\theta)|Z]$ can be computed
  and dominated convergence theorem is applicable

Example: Single-Server Queue

$$Y = \frac{1}{N} \sum_{n=1}^{N} T_n$$

T_n = system time, A_n = interarrival time, X_n = service time.

Lindley recursion:

$$T_{n+1} = X_{n+1} + (T_n - A_{n+1})^+$$

a.s. continuous under mild conditions (one little kink).

Differentiating:

$$\frac{dT_{n+1}}{d\theta} = \frac{dX_{n+1}}{d\theta} + \frac{dT_n}{d\theta} \mathbf{1}\{T_n \geq A_{n+1}\}$$

Brief Lexicon

- **IPA** — Infinitesimal Perturbation Analysis
- **FPA** — Finite Perturbation Analysis
- **EPA** — Extended IPA
- **SPA** — Smoothed Perturbation Analysis
- **RPA** — Rare Perturbation Analysis
- **SIPA** — Structural IPA
- **DPA** — Discontinuous Perturbation Analysis
- **APA** — Augmented IPA

Derivatives of Measures: Likelihood Ratio (LR) Method

$$\frac{dE[Y(X)]}{d\theta} = \int Y(x)\frac{df(x;\theta)}{d\theta}dx = \int_{-\infty}^{\infty} Y(x)\frac{\partial \ln f(x;\theta)}{\partial\theta}f(x)dx$$

Estimator:

$$Y(X)\frac{\partial \ln f(X;\theta)}{\partial\theta}$$

Many input random variables:

$$Y(X)\sum_i \frac{\partial \ln f_i(X_i;\theta)}{\partial\theta}$$

KEY OBSERVATION: linear growth of variance

Derivatives of Measures: Weak Derivatives (WD)

$$\frac{dE[Y(X)]}{d\theta} = \int Y(x) \frac{df(x;\theta)}{d\theta} dx$$
$$= c(\theta) \int_0^\infty Y(x) \left(f^{(2)}(x;\theta) - f^{(1)}(x;\theta) \right) dx,$$

$(c(\theta), f^{(2)}, f^{(1)})$ is called the weak derivative for f.

Estimator:

$$c(\theta) \left(Y(X^{(2)}) - Y(X_1^{(1)}) \right), \qquad X^{(1)} \sim f^{(1)}, X^{(2)} \sim f^{(2)}$$

Many input random variables:

$$\sum_i c_i(\theta) \left(Y(X_1, \ldots, X_i^{(2)}, \ldots, X_T) - Y(X_1, \ldots, X_i^{(1)}, \ldots, X_T) \right),$$

KEY OBSERVATION: two simulations per partial derivative

What are the Primary Ingredients for Implementation?

IPA: $\dfrac{dX}{d\theta}$, sample path derivatives $\dfrac{dY}{dX}$

SPA: IPA ingredients plus conditioning quantities and ability to compute resulting conditional expectation

LR: $\dfrac{d \ln f(X)}{d\theta}$

WD: $(c, F^{(2)}, F^{(1)})$ NOT unique!

Derivatives for some common input distributions

input dist	WD	IPA	LR/SF
$X \sim F$	$(c, F^{(2)}, F^{(1)})$	$\frac{dX}{d\theta}$	$\frac{d \ln f(X)}{d\theta}$
$Ber(\theta; a, b)$	$(1, a, b)$	NA	$\frac{1}{\theta}\mathbf{1}\{X = a\}$ $-\frac{1}{1-\theta}\mathbf{1}\{X = b\}$
$Ber(p; \theta, b)$	NA	$\mathbf{1}\{X = \theta\}$	NA
$geo(\theta)$	$(\frac{1}{\theta}, geo(\theta), 2 + negbin(2, \theta))$	NA	$\frac{1}{\theta} + \frac{1-X}{1-\theta}$
$bin(n, \theta)$	$(n, 1 + bin(n-1, \theta), bin(n-1, \theta))$	NA	$\frac{X}{\theta} - \frac{n-X}{1-\theta}$
$Poi(\theta)$	$(1, Poi(\theta) + 1, Poi(\theta))$	NA	$\frac{X}{\theta} - 1$
$N(\theta, \sigma^2)$	$\left(\frac{1}{\sqrt{2\pi}\sigma}, \theta + Wei(2, \frac{1}{2\sigma^2}), \theta - Wei(2, \frac{1}{2\sigma^2})\right)$	1	$\frac{X-\theta}{\sigma^2}$
$N(\mu, \theta^2)$	$(\frac{1}{\theta}, Mxw(\mu, \theta^2), N(\mu, \theta^2))$	$\frac{X-\mu}{\theta}$	$\frac{1}{\theta}\left[\left(\frac{x-\mu}{\theta}\right)^2 - 1\right]$
$U(0, \theta)$	$(\frac{1}{\theta}, \theta, U(0, \theta))$	$\frac{X}{\theta}$	NA
$U(\theta-\gamma, \theta+\gamma)$	$(\frac{1}{2\gamma}, \theta+\gamma, \theta-\gamma)$	1	NA
$U(\mu-\theta, \mu+\theta)$	$(\frac{1}{\gamma}, Ber(0.5; \mu-\theta, \mu+\theta), U(\mu-\theta, \mu+\theta))$	$\frac{X-\mu}{\sigma}$	NA
$exp(\theta)$	$(\frac{1}{\theta}, Erl(2, \theta), exp(\theta))$	$\frac{X}{\theta}$	$\frac{1}{\theta}\left(\frac{X}{\theta} - 1\right)$
$Wei(\alpha, \theta)$	$(\frac{\alpha}{\theta}, F^*(\alpha, \theta), Wei(\alpha, \theta))$	$\frac{X}{\theta}$	$\frac{1}{\theta}\left[\left(\frac{X}{\theta}\right)^\alpha - \alpha\right]$
$gam(\alpha, \theta)$	$(\frac{\alpha}{\theta}, gam(\alpha+1, \theta)), gam(\alpha, \theta)$	$\frac{X}{\theta}$	$\frac{1}{\theta}\left(\frac{X}{\theta} - \alpha\right)$

Types of Parameters: Distributional vs. Structural

Examples:

- queueing: interarrival and service times
- simulation (GSMP): interevent times
- queueing: routing parameters
- inventory: re-order points
- options: exercise thresholds
- statistical process control: control limits

Structural parameters may be more difficult to handle.
PA: differentiate directly (without chain rule; θ does not enter through X).
LR/WD: move parameter into a distribution (change of variables).

PART III: Model Reference Adaptive Search

- Solution space $\chi \subseteq \Re^n$

  continuous or discrete (combinatorial)

- Objective: find $x^* \in \chi$

  $$\text{s.t.} \quad x^* \in \arg\max_{x \in \chi} H(x)$$

- Assumptions: existence, uniqueness (but possibly many local optima)

Overview of Global Optimization Approaches

- **Instance-based approaches:** search for new solutions depends *directly* on previously generated solutions
  - **simulated annealing**
  - **genetic algorithms (GAs)**
  - **tabu search**
  - **nested partitions**
- **Model-based search methods:** new solutions generated via *probability distribution* (*model*) updated from previously generated solutions (*indirect* dependence)
  - **cross-entropy method (CE)**
  - **estimation of distribution algorithms (EDAs)**

Brief Review of Genetic Algorithms (GAs)

- works with **population** of solutions
- update population (generate new generation):
  - **operators**, e.g., crossover, mutation,
    - often *probabilistic*
    - produces new candidates
  - **selection** (from old and new)

Estimation of Distribution Algorithms (EDAs)

- works with sequence of **probability distributions** over solution space (continuous pdf, discrete pmf)

Main Steps of typical procedure:

- initialization: starting distribution g_0
- until stopping rule satisfied, iterate the following:
  - generate population from current distribution
  - **evaluate** newly generated solutions and **select** some subset to **update** distribution

EDAs (continued)

similarities to GAs

- uses a population
- selection process
- randomized algorithm,
  but uses "model" (distribution) instead of operators

aka

- probabilistic model building genetic algorithms (PMBGAs)
- distribution estimation algorithms (DEAs)
- iterated density estimation algorithms (IDEAs)

EDAs (continued)

BIG QUESTION:

How to update distribution?

- MRAS approach:

  sequence of implicit model reference distributions

- traditional EDAs use an explicit construction,

  can be difficult & computationally expensive

- cross-entropy method uses

  single fixed reference distribution

Model Reference Adaptive Search (MRAS)

- **Main idea:** Next distribution obtained by tilting previous

$$g_{k+1}(x) = \frac{H(x)g_k(x)}{E_{g_k}[H(X)]}, \ \forall x \in \chi.$$

Properties:

$$E_{g_{k+1}}[H(X)] \geq E_{g_k}[H(X)], \text{ and}$$

$$\lim_{k\to\infty} E_{g_k}[H(X)] = H(x^*).$$

- **Obvious Difficulties**
  - requires enumerating all points in solution space
  - $g_k(x)$ may not be computationally tractable
- **Proposed Approach**
  - Monte Carlo (sampling) version
  - use parameterized distributions $\{f(\cdot,\theta)\}$
    projection of $g_k(\cdot)$, which are *implicitly* generated

MRAS (deterministic version) components

- positive continuous strictly increasing function S needed to insure positive values, preserving order
- parameterized family of distributions $\{f(\cdot,\theta)\}$
- selection parameters ρ and non-decreasing $\{\gamma_k\}$, affecting distribution updates,
  specifically, in iteration k, only solutions better than γ_k are used in determining θ_{k+1}

MRAS parameter updates (deterministic version)

- $\gamma_{k+1} = \sup_l \{ l : P_{\theta_k}(H(X) \geq l) \geq \rho \}$ *(1-ρ)* quantiles w.r.t. f
- $\theta_{k+1} = \arg\max_{\theta \in \Theta} \int_{x \in \chi} [S(H(x))]^k I\{H(x) > \gamma_{k+1}\} \ln f(x, \theta) dx$

Lemma. θ_{k+1} as computed above minimizes KL-divergence between g_{k+1} and f,

i.e., $$\theta_{k+1} = \arg\min_{\theta \in \Theta} D(g_{k+1} \mid f(\cdot, \theta)) := \arg\min_{\theta \in \Theta} E_{g_{k+1}}\left[\ln \frac{g_{k+1}(X)}{f(X, \theta)} \right]$$

where $$g_{k+1}(x) = \frac{S(H(x)) I_{\{H(x) \geq \gamma_{k+1}\}} g_k(x)}{E_{g_k}[S(H(X)) I_{\{H(X) \geq \gamma_{k+1}\}}]},$$

$$g_1(x) := \frac{I_{\{H(x) \geq \gamma_1\}}}{E_{\theta_0}[I_{\{H(X) \geq \gamma_1\}} / f(X, \theta_0)]}$$

MRAS Basic Algorithm (deterministic version)

Initialization: Specify $\rho \in (0,1],\ S(\cdot): \Re \to \Re^+,\ f(x,\theta_0) > 0\ \forall x \in \chi$

- **Repeat** until a specified stopping rule is satisfied:
  - Calculate (1-ρ)-quantile

$$\gamma_{k+1} = \sup_{l}\{l : P_{\theta_k}(H(X) \geq l) \geq \rho\}$$

  - Update parameter

$$\theta_{k+1} = \arg\max_{\theta \in \Theta} \int_{x \in \chi} [S(H(x))]^k I\{H(x) > \gamma_{k+1}\} \ln f(x,\theta) dx$$

ROBERT H. SMITH SCHOOL OF BUSINESS
The Institute for Systems Research
A. JAMES CLARK SCHOOL OF ENGINEERING

Theory

- Restriction to natural exponential family (NEF)
  - covers pretty broad class of distributions
  Examples: Gaussian, Poisson, binomial, geometric

- **Global convergence** can be established under some mild regularity conditions
  - both deterministic and Monte Carlo versions, and stochastic optimization version

Current Research Interests

- Simulation Methodology
  - stochastic gradient estimation
  - computing budget allocation
  - importance sampling
- Markov Decision Processes (with Steve Marcus)
  - modeling and solution methodologies
  - simulation-based (sampling)
  - population-based (evolutionary, analytical)
  - global optimization
- Financial Engineering (with Dilip Madan)
  - pricing of American-style derivatives
  - credit risk (default)
- Other: fluid models for traffic network optimization, call centers

Probing Further

- M.C. Fu, “Optimization via Simulation: A Review,” *Annals of Operations Research*, Vol.53, 199-248, 1994.
- M.C. Fu, “Optimization for Simulation: Theory vs. Practice" (Feature Article), *INFORMS Journal on Computing*, 2002.
- M.C. Fu and J.Q. Hu, *Conditional Monte Carlo: Gradient Estimation and Optimization Applications*, Kluwer Academic Publishers, 1997.
- M.C. Fu, “Stochastic Gradient Estimation,” Chapter 19 in *Handbook of OR/MS: Simulation*, edited by Shane Henderson and Barry Nelson, 2005.
- M.C. Fu, “Simulation Optimization” and “Perturbation Analysis”, in *Encyclopedia of Operations Research and Management Science*, 2nd ed., 2001.
- M.C. Fu et al., “Integrating Optimization and Simulation: Research and Practice,” Proceedings of the 2000 Winter Simulation Conference, 610-616.
- J.Hu, M.C. Fu, S.I. Marcus, “A Model Reference Adaptive Search Algorithm for Global Optimization,” submitted to *Operations Research*, TR 2005-81 available at http://techreports.isr.umd.edu/ARCHIVE/dsp_reportList.php?year=2005&center=ISR.