Useful Recent Trends in Simulation Methodology

Shane Henderson
Cornell University
Thanks: NSF DMI 0400287

e.g., Inbound call centre staffing

- Min staffing costs
  s/t customers happy “enough”

- \( \min c^T x \)
  s/t \( g_i(x) \geq 0, \quad i = 1, \ldots, \text{num periods} \)

- \( \max \min g_i(x) \)
  s/t \( c^T x \leq B \)

From Simulation Model

Simulation Optimization
Inbound Call Centre Staffing

- What is the arrival rate of customers?

**Input Uncertainty**

- Conversations with HR folks suggest dependence of call volumes between periods

**Dependence Modeling**

Outline

- Part I: Simulation Optimization
  - Discuss, argue, argue
- Part II: Input Uncertainty
  - Discuss, argue, argue
- Part III: Dependence Modeling
  - Run out of time / discuss, argue, argue

Caveat Emptor: I will be giving my views on these subjects, which are diametrically opposed to common sense and good judgement.
Part I
Simulation Optimization

The Generic Problem

- **Min** $f(x)$
  - **s/t** $g(x) \geq 0$

- **Here**
  - $f(x) = E f(x, Y)$
  - $g(x) = E g(x, Y)$

- **Example 1: Newsvendor**
  - $Y$ = demand
  - $x$ = amount to stock
  - $f(x, Y) = cx - s \min(Y, x) - v \max(x - Y, 0)$
Example 2: Call Center

- Min staff costs, s/t satisfactory service
- $x_i$ = number of staff working shift $i$
- $Y$ = interarrival, service, abandonment times etc for one day's operation
- $f(x) = c^T x$ = staffing cost
- $g_i(x, Y) = \text{Num happy customers}$
  - 0.8 Num arrivals

Example 3, 4

- Example 3: Ambulance base location
  - $x$ = vector of ambulance base locations
  - $Y$ = call locations, times, scene times etc over time horizon
  - $f(x, Y) = \text{fraction of calls reached in } \leq 10 \text{ mins}$
- Example 4: Ambulance redeployment
  - $x$ = parameters of redeployment policy
  - $Y, f$ = same as in Example 3
Sample Average Approximation

- AKA external sampling, sample path opt, stochastic counterpart. Strong links with retrospective optimization
- Replace $\mathbb{E}(\cdot)$ with sample average of $\cdot$
- Fix the random samples beforehand
- Apply a deterministic optimization algorithm
- Doesn’t always work ... some weird counterexamples

Example: Newsvendor

- $f(x, Y) = cx - s \min(Y, x) - v \max(x - Y, 0)$
- Generate demands $Y_1, \ldots, Y_n$
  \[
  \min f_n(x) = \frac{1}{n} \sum_{i=1}^{n} f(x, Y_i)
  \]
- $f_n$ is piecewise linear and convex
- Can apply a convex optimization code
**SAA General Results**

- Lots known about $x_n$, the optimal solution
- Under reasonable conditions, including unique minimum:
  \[ x_n \to x^* \text{ a.s.} \]
  \[ \sqrt{n}(f_n(x_n) - f(x^*)) \Rightarrow N(0, \sigma^2) \]
  
- If multiple optima, related results hold

---

**Gradient Estimation**

- Derivatives of $f_n(x)$ boost computation
- Same thing as IPA derivatives
- Sometimes need extra smoothing when $f_n$ is not differentiable
- Easily found for newsvendor. What about harder problems?
Example 3: Base Location (Mason)

- \( f_n \) not continuous, but is piecewise differentiable
- \( \mathbb{E} f'_n \neq f' \)
- But does that matter?

Stochastic Approximation

- Time-honoured approach. Dates to 50's
- Let \( G(x) \) be estimate of \( \nabla f(x) \)
- \( x_{n+1} = x_n - a_n G(x_n) \)
- \( \{a_n\} \) is gain (step length) sequence
- Essentially steepest descent
- Works well only if choose \( a_n \)'s carefully
- Can prove lots of theorems, but ...
- Use only if SAA isn't efficient
Selecting the Best System

- Finite number of $x$'s
- Allows us to make statements like
  \[ P(x_n \text{ is truly best of } x\text{'s visited}) \geq 0.95 \]
- Use after search procedure is complete
- "Clean up afterwards"
- Additional runs will probably be needed

Bird’s-Eye View

- Apply SAA if good optimization algorithms available for SAA problem
  - Usually need structured problem
  - IPA (true) gradients help a lot, not essential
  - Requires careful control of random number streams
  - Samples don't need to be iid... can apply variance reduction methods
- If deterministic problems are tough, then anything goes. (Random search, metaheuristics)
- Selection of the best can "clean up" afterwards
Where to Find Out More

- SAA
  - Alex Shapiro (GA Tech), esp. MC sampling methods chapter in “Handbook of Stochastic Programming”, and Linderoth, Shapiro and Wright paper
- Gradient estimation: Glasserman, Fu and Hu books, L’Ecuyer papers, reviews by Michael Fu
- Selecting the best: Chapter by Kim and Nelson in forthcoming “Handbook of Simulation”

Part II
Input Uncertainty
The Bruce Lee bakery

- Open from 6am till 3pm every day
- Service times i.i.d., \( \exp(\mu) \)
- How many staff are needed so that 90% of customers wait \( \leq 1 \) minute in queue?
- Poisson arrival process, rate \( \Lambda \), on day \( i \)
- \((\Lambda; \ i \geq 0)\) i.i.d. \( \mathcal{G}(\alpha, \beta) \)

\[
\lim_{\ell \to \infty} \frac{\sum_{i=1}^{\ell} S_i}{\sum_{i=1}^{\ell} N_i} = \frac{ES_1}{EN_1}
\]

\[
EN_1 = EE[N_1|\Lambda_1] = E[9\Lambda_1] = 9E\Lambda_1
\]

- Similar for \( E S_i \) so standard approach works
The Steve Russell wine store
Open from 11am till 8pm every day
- Service times i.i.d., exp(\( \mu \))
- How many staff are needed so that 90% of customers wait \( \leq 1 \) minute in queue?
- Poisson arrival process, rate \( \lambda \)
- Don’t know \( \lambda \)
- Uncertainty in \( \lambda \) well modeled as \( \mathcal{G}(\alpha, \beta) \)

Performance measure still \( E S_1 / E N_1 \)
- \( E N_1 = 9 \lambda = ? \)
- Which value of \( \lambda \) to use?
- Same problem with \( E S_1 \)
- What should we do?
**Terminology**

- Simulation model needs input distributions and parameters
- Model uncertainty: which family?
- Parameter uncertainty: what parameters?
- Input uncertainty: Both problems

**Why do we care?**

- $M/M/1$ queue, arrival rate $\lambda_0 = 9$ (unknown), service rate $\mu_0 = 10$
- Expected time in system is $f(\lambda_0)$ where

$$f(\lambda) = \begin{cases} (\mu_0 - \lambda)^{-1} & \text{if } \lambda < \mu_0, \\ \infty & \text{if } \lambda \geq \mu_0 \end{cases}$$

- True performance = $f(\lambda_0) = 1$
Estimating $\lambda_0$

$$\hat{\lambda}_n = 1/\overline{U}_n 
\approx N(\lambda_0, \lambda_0^2/n)$$

- Assume this holds exactly
- Possibly negative ... ignore
- Possibly bigger than $\mu_0$
- Assume perfect simulation that reports $f(\hat{\lambda}_n)$

The induced distribution of $f(\hat{\lambda}_n)$
Back to the wine store

- By ignoring input uncertainty, we can make one of two errors:
  - Overestimate $\lambda_0$: Hire too many staff
  - Super service, expensive

  - Underestimate $\lambda_0$: Hire too few staff
  - Poor service, cheap

- Not so bad at a wine store, but what about at a 911 call center?

What should we do about it?

- Any method should be
  - Transparent to users
  - Statistically valid
  - Easily implemented
  - Computationally efficient

- Approach 0: Do “nothing”

- OK so long as
  - users understand that the simulation analyzes the chosen system
  - a sense of model bias is obtained

- Sensitivity analysis, uncertainty analysis
Suppose \( \theta = (\theta_1, \theta_2) \) (e.g., arrival, service rates)

\[ \begin{align*}
\theta_2 \\
\theta_1 \\
\end{align*} \]

And here \((L_b, U_b)\)

Simulate here \((L_s, U_b)\)

\( f \) increases

Bayesian methods

- Chick and Ng (2002)
- Ng and Chick (2001)
- Lots of work on “Bayesian Model Average”
Induced distribution methods

- Study distribution of $f(\tilde{m}, \tilde{\theta})$
- Mean: Andradottir and Glynn (2003)
- Density: Steckley and Henderson (2003), Steckley (2005)
- Helton: $G(x; m, \theta) = P(\tilde{X}_1(m, \theta) \leq x)$
  Look at distribution of $G(x; \tilde{m}, \tilde{\theta})$

Bird’s Eye View and More Info

- Cheng and Holland for small numbers of parameters
- Induced distribution if you can get your head around the idea
- Bayesian model average once the tools are automated and in software
- In the meantime: sensitivity and uncertainty analysis
- More info: Henderson tutorial, WSC ’03
Part III
Modeling Dependence

Dependence

- Some of the random variables in the simulation are related.
- Can dramatically alter performance, decisions based on model.
  - # arrivals in morning vs afternoon in call centre.
  - Wind fields in yacht match race simulation.
  - Travel speeds of ambulances.
Modeling Philosophy

- What needs to be modeled?
- Use simple models to help decide.

1. Capture the physics
2. Use reasonable joint distribution with well understood pluses and minuses
3. Use partially specified distributions

Call Centre Arrivals

- Palm-Khintchine thm suggests arrivals follow a nonhomogeneous Poisson process
  - # arrivals in a period is Poisson
  - different time intervals are independent
- Call centre HR folks and data:
  - # arrivals has a bigger variance than Poisson
  - busier than usual at 9am ⇒ busier at 10am
- Uh oh
Randomly Varying Arrival Rates

- Vijay Mehrotra, Sam Steckley
- One explanation:
  - Arrival rate function is random
  - Once arrival rate function realized, arrivals follow a NHPP with that rate function
- Why is this at least plausible?
  - Mechanism generating arrivals can vary due to weather, promotions, etc
- Trying to capture the unknown physics

Yacht Match Race Simulation

- Andy Philpott, Hamish Sheild, U of Auckland
- 2 yachts
- Design & tactical Qs
- Model wind speed and direction on a grid
- Spatial/temporal dependence
- We use a vector AR
- Well-understood joint distribution
Vector AR Process

Ambulance Travel Speeds

- Call locations are random
- So are travel speeds!
  - Traffic lights
  - Weather (visibility, road conditions)
  - Congestion
- Should we explicitly model random travel speeds?
Randomness in travel speeds can improve performance!
What About Queueing Effects?

- Service time of a call = travel + nontravel
- Travel time = distance / speed and speed depends on macro and micro factors
- Macro factors affect all calls, so average of service time changes
- Micro factors influence individual calls, so variability of service time changes
- Looked at M/G/1 using transform inversion and M/G/∞ models

Size Matters

Fraction on time

80%

5 ambulances

30 ambulances

Utilization
Random Travel Speeds

- Help a bit if calls are clustered near the boundary of reachable calls
- Dependence structure matters: One form increases variability. The other increases mean service time
- Hurt heavily utilized fleets more than lightly-utilized fleets

More Info

- Henderson tutorial at WSC ’05 discusses these examples in more depth
- Biller and Ghosh chapter in forthcoming “Handbook of Simulation”
- Prediction: Spatial/temporal models will become more common – need some convenient model classes
Last Words

- Are you a tourist wanting to hook up with simulation folks?
  - INFORMS Simulation Society business meeting tonight 6:15pm in the Taylor room
- Are you a colleague who thinks I misrepresented things? Speak now!
- Are you a researcher looking for a new area? Sim Opt in an application setting is a great area!

The End