ORIE 4580/5580: Simulation Modeling and Analysis

ORIE 5581: Monte Carlo Simulation

Unit 13: Output Analysis in DES

Sid Banerjee

School of ORIE, Cornell University

output analysis

estimate performance measures of interest from a simulation model

- from output analysis viewpoint, two types of simulation models:
  - terminating
  - steady-state (nonterminating)

terminating simulation

- a simulation that has a natural time at which it should stop
- this duration can be random and/or controlled
- examples:
  - modeling a grocery store from 8:00 am to 5:00 pm
  - modeling an epidemic till it dies out
  - option pricing (end at exercise date of the option)

steady-state (nonterminating) simulation

- sim that runs continuously without any stopping condition
- usually interested in the "long-run" behavior
- what matters is *relative rate*:
  for fast dynamics (order of μs), sim over few minutes $\approx$ infinite horizon
- examples:
  – finding throughput of a production facility assuming conditions do not fluctuate over time
  – rideshare system during non-peak hours
  – telecommunications network where arrival rate is constant over short periods of time, and packets arrive frequently.

output analysis for terminating simulation

- standard replication + CI methodology works!
- example: let $\{Y(t) : t \in [0, T]\}$ be a trajectory of the simulation model
  suppose we are interested in performance measures

$$\mu = \mathbb{E}\left[\frac{1}{T}\int_0^T Y(t)\ dt\right]$$

$$\theta = \mathbb{E}\left[\max_{t\in[0,T]} Y(t)\right]$$

- run the simulation model for r different replications
  - replications are independent
  - start from same initial conditions
- let $\{Y_r(t) : t \in [0, t]\}$ denote the trajectory of replication r

output analysis for terminating simulation

terminating simulation: sample statistics

- from r simulation runs, we obtain r estimates of μ

$$\bar{X}_r = \frac{1}{T}\int_0^T Y_r(t)\ dt \quad , \quad \bar{Z}_r = \max_{t \le T} Y_r(t)$$

- estimate μ, θ by $\hat{\mu}_r = \frac{1}{r}\sum_{r=1}^{r} \bar{X}_r, \hat{\theta}_r = \frac{1}{r}\sum_{r=1}^{r} \bar{Z}_r$
- $Var(\bar{X}_r)$ is unknown. estimate it by

$$s^2_{\bar{X}_r} = \frac{1}{r-1}\sum_{r=1}^{r}\left(\bar{X}_r - \hat{\mu}_r\right)^2 .$$

- $Var(\hat{\mu}_r) = Var\left(\frac{1}{r}\sum_{r=1}^{r} \bar{X}_r\right) = \frac{1}{r^2}\, r\, Var(\bar{X}_1) \approx \frac{s^2_{\bar{X}_r}}{r}$
- use this to build CIs

terminating simulation: entity-based statistics

same method works when θ is "observation-based" rather than "time-weighted" average.

example: $\theta =$ empirical average waiting time per customer in a queueing system over interval $[0, T]$

- let $\{Y_1, Y_2, \ldots, Y_{N_T}\}$ be the waiting times of the customers over the time interval $[0, t]$, then

$$\theta = \mathbb{E}\left[\frac{1}{N_T}\sum_{i=1}^{N_T} Y_i\right]$$

- run r simulation runs and let $\{Y_1^r, Y_2^r, \ldots, Y_{N_T^r}^r\}$ be the customer waiting times in run r.
- set $\bar{\theta}_r = \frac{1}{N_T^r}\sum_{i=1}^{N_T^r} Y_i^r$ ...

output analysis for steady-state simulation

- $\{Y(t) : t \geq 0\}$ is the trajectory of a steady-state simulation.
- $Y(t)$ is said to have a well-defined steady-state mean if

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t Y(s)\, ds = \theta,$$

where θ is a deterministic constant.

output analysis for steady-state simulation

- example: $Y(t)$ is the length of a stable $M/M/1$ queue at t

  $\lim_{t\to\infty} \frac{1}{t}\int_0^t Y(s)\,ds$ is the long-run average queue length

  $\frac{1}{t}\int_0^t Y(s)\,ds$ converges to a deterministic constant as t grows

- example: W_i is waiting time of the i-th customer

  $\frac{1}{n}\sum_{i=1}^n W_i$ converges to a deterministic constant as n grows

steady-state means: notes

1. when assessing the steady-state performance of a system, the choice of the initial conditions is irrelevant
   however, over a finite horizon, the initial conditions matter
2. steady-state means can provide approximations for performance over finite time intervals
   if $Y(s)$ is the cost incurred at time s, then
   $$\int_0^t Y(s)\, ds \approx \dots$$
3. steady-state performance measures are usually easier to compute analytically than finite-horizon performance measures

challenges of steady-state simulation

model is initialized with arbitrary initial conditions.

two problems:

- the initial conditions can affect the **relaxation time.**
- the output obtained from a steady-state simulation model over non-overlapping time intervals can be **correlated.**

initial transient problem

- when interested in the long-run performance of the system, we usually initialize the simulation model by using arbitrary initial conditions
- this introduces a “bias” into the simulation
- the initial segment of the simulation is not representative of the steady-state behavior
- if we run the model for an infinite period of time, then the impact of the initial conditions will disappear
  but . . .

relaxation time

given that we have to run the simulation model for a finite period of time, several questions arise:

1. how long does the initial transient period persist?
2. how can we identify the end of the initial transient period?
3. how can we mitigate the effect of the initial transient period?

relaxation time: single-server queue

$M/M/1$ queue with $Exp(\lambda)$ interarrival time and $Exp(\mu)$ service time

- if $\lambda < \mu$, the system has a well-defined steady state
- let $Y(t)$ be the number of customers in the system at time t

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t Y(s)\, ds = \frac{\rho}{1-\rho},$$

where $\rho = \lambda/\mu$.

- $\lambda = 1/100s^{-1}$ and $\mu = 1/90s^{-1}$, $\lim_{t\to\infty} \frac{1}{t}\int_0^t Y(s)\, ds = 9$.
- find number in system at 1, 5, 10, 50, 100 and 500 hours
- run 100 replications to build CI for average number in system.

relaxation time: single-server queue

t (hours)	95% confidence interval for $\lim_{t\to\infty} \frac{1}{t}\int_0^t Y(s)\, ds$
0	$[0.00,\ 0.00] = 0.00 \mp 0.00$
1	$[2.94, 3.76] = 3.35 \mp 0.41$
5	$[5.74, 7.22] = 6.48 \mp 0.74$
10	$[6.68, 8.14] = 7.41 \mp 0.73$
50	$[8.24, 9.58] = 8.91 \mp 0.67$
100	$[8.38, 9.26] = 8.82 \mp 0.44$
500	$[8.68, 9.12] = 8.90 \mp 0.22$
∞	$[9.00,\ 9.00] = 9.00 \mp 0.00$

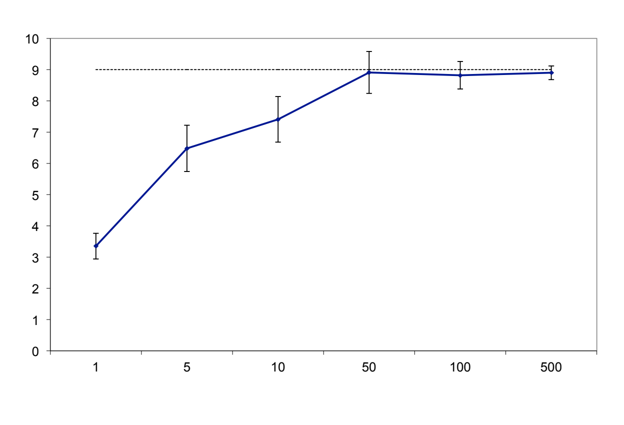

Figure 1: average 'dragged down' by low starting queues (initial transient)

clicker question: population variance

to find the steady-state average waiting time in an $m/m/1$ queue, we run one replication over $[0, T]$ and compute $\bar{\theta} = \frac{1}{N_T}\sum_{i=1}^{N_T} Y_i$
can approximate the variance of θ as

$$Var(\theta) = \frac{1}{N_T - 1}\sum_{i=1}^{N_T}(Y_i - \bar{\theta})^2$$

(a) you saved a lot of replications - well done!

(b) this is ok if instead of N_T, you average over a random number of agents n chosen independently of the simulation

(c) this is ok if you fix the number of agents you average over to a constant n

(d) this sounds like a bad idea...

autocorrelation problem

- want long-run average time customers spend in the system
- let $Y_1, Y_2, \ldots$ be time spent in system by successive customers

  $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} Y_i$ converges to average time a customer spends in system

autocorrelation problem

time spent in system by successive customers are not independent!

if current customer spends a long time, then it is likely that the next customer spends a long time in the system...

– cannot apply standard CI methodology.

replication-deletion method for steady-state simulation

warm-up period

to remove the bias introduced by the initial transient

- run model for a "warm up" period where no statistics are collected
- once warm up period is completed, start to collect statistics

replication-deletion

to remove autocorrelation effects

- run simulation model for r independent replications
  $\{Y_r(t) : t \in [0, t]\}$ is the trajectory of replication r
- warm-up: ignore up to time point d in each replication
- use estimate $\bar{Y}_r = \dfrac{1}{t-d} \displaystyle\int_d^t Y_r(s) d_s$

replication-deletion method for steady-state simulation

choosing the warm-up period

- one way to choose warm-up period d is so that the average in Figure 1 looks to have settled down
  this is conservative! why?

- better way: plot $\mathbb{E}[Y(t)]$ vs t and look for where it levels off
  (can be different for different metrics – take the maximum)

choosing the warmup period

t (hours)	95% CI for $\mathbb{E}[Y(t)]$
0	$[0.00,\ 0.00] = 0.00 \mp 0.00$
1	5.4 ∓ 1.0
5	9.4 ∓ 1.6
10	7.9 ∓ 1.6
50	8.3 ∓ 1.7
100	8.7 ∓ 1.8
500	9.7 ∓ 1.7
∞	9.00 ∓ 1.7

replication-deletion method: wrapup

- easy to implement
- uses standard CI methodology
- every replication repeats initial transient period
  $\Longrightarrow$ waste of computational effort
- is there a way to not repeat the initial transient period many times?

batch means in steady state simulation

- the idea is to obtain independent observations, so that we can apply the standard ci methodology
- method of batch means is applied to one loooong simulation run
- we go through the initial transient phase only once

method of batch means for steady state simulation

- $\{Y(t) : t \in [0, t + d]\}$ is the trajectory of the simulation model
- d is the warm up period
- divide interval $[d, t + d]$ into k batches, each with length $\ell = t/k$

method of batch means:

1. compute the batch means as

$$\bar{b}_i = \frac{1}{\ell} \int_{d+\ell(i-1)}^{d+\ell i} Y(s)\, ds \qquad \text{for } i = 1, \ldots, k.$$

method of batch means for steady state simulation

2. compute point estimate of $\lim_{t\to\infty} \frac{1}{t}\int_0^t Y(s)\,ds$ as mean of batch means:

$$\bar{Y} = \frac{1}{k}\sum_{i=1}^{k} \bar{B}_i.$$

3. estimate variance of $\bar{B}_i$ as $s_k^2 = \dfrac{1}{k-1}\displaystyle\sum_{i=1}^{k}(\bar{B}_i - \bar{Y})^2$.

4. approximate CI for $\lim_{t\to\infty} \frac{1}{t}\int_0^t Y(s)\,ds$ is

$$\bar{Y} \mp t_{\alpha/2,k-1}\,\frac{s_k}{\sqrt{k}}.$$