Simulations midterm

good luck

Before we Sim

Prob Review

- we need to review some basic probability terms...
- Random Variable → a random phenomenon where function X provides a real # to every outcome $P(X=x)$
- we can define random variables w/ Cumulative Distribution Function (CDF)...

$$F(x)=P(X\le x)=\begin{cases}\sum_{x\le X}P(X=x), & \text{if discrete}\\ \int_{-\infty}^{X} f(x)\,dx, & \text{if continuous}\end{cases}$$

- we also like to know the Expected Value of $g(X)$...

$$E[g(X)]=\sum_{x\in X} g(x)P(X=x) \text{ if discrete} \quad E[g(X)]=\int_{-\infty}^{\infty} g(x)f(x)\,dx \text{ if continuous}$$

how we recognize common RV

- The covariance of X and Y is...

$$Cov(X,Y)=E[(X-\mu_x)(Y-\mu_y)]=E[XY]-E[X]E[Y]$$

and... $Cov(X,X)=Var(X)$

and if X and Y are independent ⇒ $Cov(X,Y)=\emptyset$

The Correlation between X and Y...

$$\rho(X,Y)=Cov(X,Y)/\sqrt{Var(X)Var(Y)}$$

- Central Limit Theorem → w/ a sequence of iid rv's $\{X_n\}$ ... $Z_n=\frac{\bar{X}_n-\mu}{\sqrt{\sigma^2/n}}$, w/ $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$

$Z_n \approx N(0,1)$ as $n\to\infty$, and $S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2 \to \sigma^2$ as $n\to\infty$

- We can use this to create Confidence Intervals which give us a range that our true μ is w/ some % confidece. Given by...

$$\bar{X}_n \pm 1.96\sqrt{S_n^2/n}$$

Random # Gen.

- Our simulations packages need random # generation to operate. But computers can't actually generate trully random outputs like say a coin toss would. The create Pseudo Random Numbers
- Random Number → a variable from a Unif(0,1)
- Random Variable → random outcomes of all other distributions. Can be obtained by transforming random #.
- main method computers use... Linear Congruential Generators ... a (multiplier) c (incriment) m (modulus) z (seed)

generates sequence of integers between ∅ and m-1...

$$Z_n=(aZ_{n-1}+c) \bmod m$$

use that to get U_i in [0,1]...

$$U_n=\frac{Z_n}{m}$$

- cycle lengths of Z_n's will be ≤m. Obviously not really random. We pick $m\ge 10^9$ to make look more random.
- We like LCG's cause they're fast, low storage, and reproducible, and have good statistical properties based on param.
- We measure an LCG's period, or how long (n) it takes for $Z_n=Z_0$. Largest period ≡ Full Period ≡ m-1
- Two types of LCG's... multiplicative or mixed

multiplicative $c=0$

$$Z_n=(aZ_{n-1}) \bmod m$$

has full period (desirable) when m is a prime # and a is a primitive root of m.

↳ $a>1$ and $(a^{m-1}-1) \bmod m=0$ and for all $i<m-1$, $(a^i-1) \bmod m\ne 0$

↳ easier thm...

* a multiplicative LCG w/ $m=2^b$ has a max period of 2^{b-2} achieved if... ☐ Z_0 is odd ☐ a has form 8k+3 or 8k+5 for some integer k

mixed $c>0$

$$Z_n=(aZ_{n-1}+c) \bmod m$$

has full period if c and m are relatively prime (only pos. integer divides both is 1) and q is prime and divides m.

↳ easier thm...

* a mixed LCG w/ $m=2^k$, has full period if... ☐ c is odd ☐ a has form 4j+1 for some integer j

Test Rand. #'s

- for our random #'s we can look at empirical tests to test their actual randomness.
- Chi-squared Test → look at histogram of observations vs fitted density function. Look at how many obsv. out of n are in each bin. null hypothesis H_0: X_i's are iid w/ dist. func. F. k bins. N_j = # of obsv in bin. p_j = prob of outcome according to F. test stat. $\chi^2=\sum_{j=1}^{k}(N_j-np_j)^2/np_j$, if H_0 true. dist should match χ^2. for Uniform [0,1] $\chi^2=\frac{k}{n}\sum_{j=1}^{k}(N_j-\frac{n}{k})^2$
- Serial Test → a d dimensional χ^2 test to test uniformity and indp. between succesive obsv. H_0: U_i's are iid U(0,1)
- for both χ^2 and Serial test the prob. of rejecting a false H_0 is low.
- Autocorrelation Test → test autocorrelation of a sequence at various lag j's. We can't get the true autocor. we need to estimate... $\hat{\rho}_j=\frac{12}{h+1}\sum_{k=0}^{h}(U_{1+kj}U_{1+(k+1)j})-3$ $\quad\hat{v}_j=\frac{13h+7}{(h+1)^2}$ $\quad h=\lfloor\frac{n-1}{j}\rfloor-1$

w/ H_0: $\rho_j=\emptyset$

for large j test stat... $Z_j=\hat{\rho}_j/\sqrt{\hat{v}_j} \sim N(0,1)$

Sims

- We categorize simulations in 3 ways. Static vs Dynamic, Deterministic vs Stochastic, and Continuous vs Discrete

Static Sim

- A static simulation is one where time plays no role. It relys on random # or random variat generation for things to happen.
- When solving problems we need to remember we'r not getting optimal solutions. We're getting output.
- examples → coin flip, newsvender problem, prg. managment

Discrete Event Sim

- DES → when a system evolves over time where state var. change instantaneously at sep. points in time.
- Single-Server Queuing System →

customers $\xrightarrow{\text{arrival}}$ queue $\xrightarrow[\text{dispatching}]{}$ server → departure

$\underbrace{\text{queue} \to \text{server}}_{\text{residence time}}$

common components of single server queuing system...

- event → instaneous occurence that may change state of system
- simulation clock → variable that keeps track of time in model
- entity → obj. explicitly represented in a model.
- attribute → property of an entity
- activity → duration of time of specified length that we know upfront
- delay → duration of unspecified length
- set → ordered lst of associated entities

things we need to know in begining...

- customer interarrival times
- service times
- queue type
- stopping rule

quantities we could want to find...

- Expected Avg. Delay in Queue → wait time for cust.

$$d(n)=E\left[\frac{\sum_{i=1}^{n}D_i}{n}\right]$$

wait time; D_i's delays for each cust.

$$\hat{d}(n)=\frac{\sum_{i=1}^{n}d_i}{n} \to \text{estimator}$$

- Expected Avg. # customers in Queue → $Q(t)$ = # in queue at time t. $T(n)$ = time n^{th} customer exits queue

$$\hat{q}(n)=\frac{\int_0^{T(n)}Q(t)\,dt}{T(n)-\emptyset}$$

estimator

- Expected Server Utilization →

$$B(t)=\begin{cases}1 & \text{if busy at } t\\ 0 & \text{if not}\end{cases}$$

server status

$$u(n)=\frac{\int_0^{T(n)}B(t)\,dt}{T(n)-\emptyset} \to \text{estimator}$$

- If we are running simulations by hand, there are 2 strategies to do so Process-Interaction Approach & Event Scheduling Approach.

Output of Sims

- What kind of output analysis we do depends on the simulation...
- we want to use CI to account for sampling error, but simulations will never produce iid Normal output, so we can't apply classical statistical techniques.
- So what can we do with output...

Terminating Sim

- means we know when the system will end and we want to observe the system at an end.
- we could be working w/ discrete Y_n or continuous $Y(t)$ data and our goal is to find Unbiased estimators $\bar{Y}_n$ for some property (ie avg wait time)
- we'd also like an estimator for $Var(\bar{Y}_n)$, but because our Y_i's are correlated w get extremely biased output. So we use..

Method of Independent Replications

↳ conduct R indp. replications, each w/ n observations, to ensure indp.

↳ $Z_i=\frac{1}{n}\sum_{j=1}^{n}Y_{i,j}$, $i=1...R$

↳ each initialized under same conditions

rep1	$Y_{11}\ Y_{12} \cdots Y_{1n}$	Z_1
rep2	$Y_{21}\ Y_{22} \cdots Y_{2n}$	Z_2
⋮	⋮	⋮
rep R	$Y_{R1}\ Y_{R2} \cdots Y_{Rn}$	Z_R

Y_{ij} ith rep jth obsv; sample means

- The Z_i's from the independent replications are iid and can use them to find a grand sample mean, sample variance $(\hat{V}=\frac{1}{R-1}\sum_{i=1}^{R}(Z_i-\bar{Z}_R)^2)$ $(Var(\bar{Z}_R)=\frac{\hat{V}}{R})$, and a confidence interval $\bar{Z}_R \pm 1.96\sqrt{\hat{V}/R}$

Steady State Sim

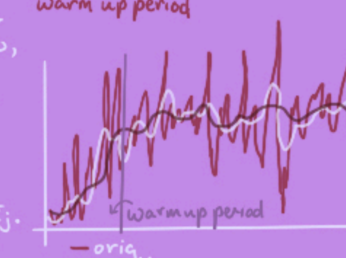
- we want to know about a system during it's run, while it's in it's normal state we want it's steady state mean μ.
- a problem ⇒ Initialization bias → we don't care about the time it took the system to warm up to steady state.
- Easiest way to deal is to truncate the warm up period. so...

$$\bar{X}_{n,\ell}=\frac{1}{n-\ell}\sum_{i=\ell+1}^{n}X_i$$

- but how do we choose truncate point ℓ...
- Welch's Method → 1) we start w/ the avg. of k rep's $\bar{X}_j$'s. 2) Then for chosen window w, moving averages

↳ $$\bar{X}_j(w)=\begin{cases}\frac{1}{2w+1}\sum_{m=-w}^{w}\bar{X}_{j+m}, & w+1\le j\le n-w\\ \frac{1}{2j-1}\sum_{m=-j+1}^{j-1}\bar{X}_{j+m}, & 1\le j\le w\end{cases}$$

this finds the average of a window sized w, rather than an avg for each $\bar{X}_j$. Find averages for many w's, 3) plot, and 4) visually find warm up period.



- We prefer to err on the side of the truncation point being too large, even it makes slightly bigger CI's...
- we make CI's of steady state μ in lof 2 ways...
- Multiple Replications Method → determine truncation ℓ, run k indp. reps of length n, cutting ℓ off each. Can find CI for μ... $\bar{Y}+1.96\sqrt{\hat{V}/k}$

grand mean of replications; var estimator

↳ note... ℓ↑ for fixed n ⇒ systematic error ↓ sampling error ↑

n↑ for fixed ℓ ⇒ systematic + sampling error ↓

- Batch Means Method → take a very long sim and cut off ℓ once at the begining, then split into b nonoverlapping batches of size m. (n=bm) We find a CI for μ ... $\bar{Y}\pm 1.96\sqrt{\hat{V}_B/b}$

grand mean of estimates; var estimator; # batches

but how to choose m? rec. $10\le b\le 30$