Discrete event simulation:

Main program:

① Invokes initialization: time clock=0, initialize system state and counters, initialize event list.

② Invokes timing routine: determines next event type, advances simulation clock

③ Invokes event routine: updates system state, updates statistical counters, generates future events and adds them to the event list.

(3.1) Calls library routines for generating random variables

Modeling can be:

- activity based
- event based
- process based

Random number generators:

Desired properties:

- Generated numbers $\sim U(0,1)$
- Numbers should exhibit no correlation
- fast
- low memory
- portable
- separate streams
- reproducible streams

Linear Congruential Generators (LCG)

Recursive formula

$$z_i = (a z_{i-1} + c) \pmod{m}$$

multiplier (a), increment (c), modulus (m)

$m > 0$

$a < m$ $\qquad$ z_0 seed

$c < m$

obtain Random numbers as $U_i = \frac{z_i}{m}$

looping occurs and cycle length is period and depends on seed

Objections:

1. Not random $\quad z_i = \left[a^i z_0 + \frac{c(a^i - 1)}{a-1} \right] \pmod{m}$

2. U_i only can take rational values.

However with right parameters it can look IID $U(0,1)$

THM: For LCG to have full period:

1. m and c are relatively prime
2. If q is a prime number that divides m, then q divides $a-1$
3. If 4 divides m, then 4 divides $a-1$

m large, full period helps with uniformity generic case mixed generators $c > 0$

Full period possible
like to have $m = 2^b$

common case multiplicative generators

$$z_i = (a z_{i-1}) \pmod m$$

Full period not possible, period $= m-1$ is. Often m is largest prime number smaller than 2^b

Extensions

$$z_i = g(z_{i-1}, z_{i-2}, \dots) \pmod m$$

Feedback shift register generators

$$b_i = (c_1 b_{i-1} + c_2 b_{i-2} + \dots + c_q b_{i-q}) \pmod 2$$

$$b_i = b_{i-r} \oplus b_{i-q}$$

$$W_i = W_{i-r} \oplus W_{i-q}$$

$$U = \frac{W_i}{2^l}$$

Generalized feedback shift register GFSR

Fitting distribution functions

Very important to use the right distribution

Options:

1) Use data directly
2) Define empirical distribution
3) Fit theoretical distribution

1) Limited to obs, tails

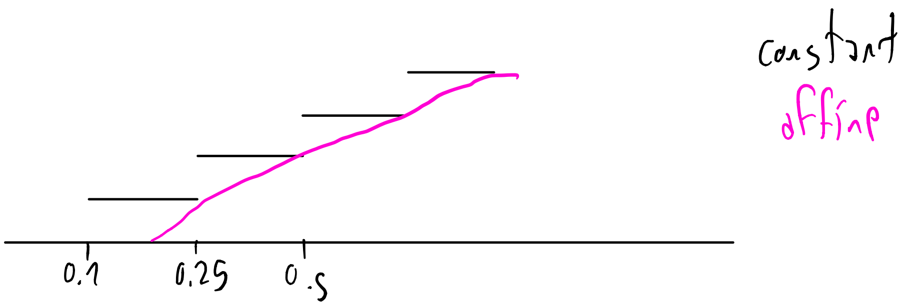
2) Empirical distribution functions:

continuous case

Piecewise constant

Given $x_1, \dots, x_n$

we construct empirical distr. function $F_n(x) = \frac{\# x_i \leq x}{n}$



Constant

1) Let $\{X_{(i)}\}$ be sorted sequence of obs
2) Then $F(x_{(i)}) = i/n$

Affine:

1) Let $\{X_{(i)}\}$ be sorted sequence of obs
2) then
$$F_n(X) = \begin{cases} 0 & X < X_{(1)} \\ \dfrac{i-1+\dfrac{x-x_{(i)}}{(x_{i+1})-x_i}}{n-1} & x_{(i)} \leq x < x_{(i+1)} \\ 1 & x \geq x_{(n)} \end{cases}$$

discrete case

assume grouped data in k adjacent intervals
j^{th} interval contains n_j observations, n in total

Empirical piecewise constant distr:

$$g(x) = \frac{\sum_{i=1}^{j} n_j}{n} \quad a_{j-1} \le x < a_j$$

3) Fit theoretical distribution to observed data

Data analysis process
Steps:
1) Hypothesize family of distribution by looking at data through summary statistics (mean, median, variance, coeff. var., skewness, Or creating histograms/ quantile summaries
2) Estimate parameters: using an estimator like Maximum Likelihood Estimators (MLE)
3) Determining how representative fitted distr. are
Heuristic:
- Density-histogram plots frequencies → χ^2 test
- Distribution function differences plot → Kolmogorov-Smirnov test

$\hat{F}(x) - F_n(x)$
hypothesis, empirical data

Goodness of fit tests:

- Find test statistic S which is function of $x_1, \dots, x_n$ s.t. distr. S known
- Choose interval (a,b) such that $P((a<S<b) \mid H_0 \text{ is true}) = 1-\alpha$
- Actual test: take sample and calculate test statistic, if $s \notin (a,b)$ then reject H_0.

χ^2 test

a) Make n observations and take N_i as number of times x_i is found. For continuous number of times measured value falls in bin i.

b) Expected number of observations is np_i:
- For discrete p_i is given by prob. mass function
- For continuous $\quad p_i = \int_{a_{i-1}}^{a_i} \hat{f}(x)\, dx$

c) test statistic

$$\chi^2_{k-1-m} = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} = \frac{1}{n} \sum_{i=1}^{k} \left(\frac{N_i^2}{p_i} \right) - n$$

d) with α and dof you can find critical value

K-S test:

$$D_n = \sup_n |F_n(x) - \hat{F}(x)| \qquad \text{So:}$$

$$D^+ = \max_{1 \le i \le n} \left\{ \frac{i}{n} - \hat{F}(X_{(i)}) \right\} \qquad D = \max\{D^+, D^-\}$$

$$D^- = \max_{1 \le i \le n} \left\{ \hat{F}(X_{(i)}) - \frac{i-1}{n} \right\}$$

Poisson processes: continuous-time counting process.

Required properties:

- Independent increments: number of arrivals in disjoint intervals are independent
- Stationarity: For every $s \geq 0$ the number of arrivals in the interval $[s, s+t)$ is equal to $N(t)$ it only depends on interval length t, but not on s
- No simultaneous events

The arrival rate (or intensity) λ of the Poisson process denotes the average number of arrivals per time unit.

THM: For a Poisson process, the variable $N(t)$ has a discrete Poisson distribution with parameter λt: its probability mass function is given by:

$$\Pr[N(t) = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, 2, \ldots$$

For constant rate λ and interval length t and at time s, the number of arrivals in $[s, s+t)$ is increment process

$$N(s+t) - N(s) \text{ for which:}$$

$$\Pr[N(s+t) - N(s) = k] = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

$$E[N(t)] = \lambda t$$

Interarrival times: an event takes place if first arrival takes place at time T, happens to be greater than t. Must have no arrivals in $[0, t)$

$$\Pr[T_1 > t] = \Pr[N(t) = 0] = e^{-\lambda t}$$

$$\Pr[T_1 \leq t] = 1 - e^{-\lambda t} \qquad \text{exponential distr.}$$

pdf $\Pr[X=t]$

$f(t) = \lambda e^{-\lambda t}$

cdf $\Pr[X \leq t]$

$F(t) = \int_0^t \lambda e^{-\lambda \tau} d\tau = -e^{-\tau} \Big|_0^t = 1 - e^{-\lambda t}$

$E[X] = \frac{1}{\lambda}$ $\quad$ $Var[X] = \frac{1}{\lambda^2}$

Memoryless property: $\Pr[X \leq T+t \mid X \geq T] = \Pr[X \leq t]$

Arriving at time t:

Poisson: no arrival in $[0, t)$, arrival in $[t, t+\delta t)$

$$\Pr[N(t)=0] \Pr[\underbrace{N(\delta t) \geq 1}_{1-\Pr[N(\delta t)=0]}] = e^{-\lambda t}(1-e^{-\lambda \delta t})$$

Exponential:

Prob of event after t · probably of event before $t+\delta t$ given it happens after t

$$= (1-F(t)) \overset{\text{memoryless}}{F(\delta t)} = e^{-\lambda t}(1-e^{-\lambda \delta t})$$

Erlang:

Waiting time till n^{th} event/arrival

$S_n = \sum_{i=1}^{n} T_i \quad n \geq 1$ sum of exponential random variables

pdf $f(t) = \lambda e^{-\lambda t} (\lambda t)^{n-1} / (n-1)!$

Properties of poisson process:

- **superposition property**: two poisson processes with intensities λ_1 and λ_2 can be combined into one with rate $\lambda_1 + \lambda_2$
- **conditioning on number of arrivals**: if in interval $[0, t)$ the number of arrivals $N(t)$ is equal to n then these are each uniformly distr. on $[0, t)$
- **random selection**: if random selection is made with probability of accepting p from process at rate λ then this constitutes new poisson process at rate λp $\quad Pr = \frac{\lambda(t)}{\lambda^*}$
- **random split** $\quad Pr[X_1 < X_2] = \frac{\lambda_1}{\lambda_1 + \lambda_2}$
- **PASTA** (Poisson Arrival See Time Averages): consider system that spends time in states E_j. Arrivals at rate λ induce state changes. In equilibrium with state E_j we have two prob:
  - M_j system is in E_j at random instant
  - M_j^* " just before arrival
  - in general $M_j \neq M_j^*$

Hitchhiker paradox: you arrive in the middle of interarrival but your expected waiting time is $\lambda = 1/2\,(\beta + \frac{\sigma^2}{\beta}) = 1/2\,(\beta + \beta) = \beta$

Variance of exponential distribution is $(\bar{X})^2$ thus $\overline{X^2} = (\bar{X})^2 + Var(x) = 2(\bar{x})^2$ hence $\bar{w} = \frac{1}{2}\frac{\overline{X^2}}{\bar{X}} = \frac{1}{2}\frac{2(\bar{x})^2}{\bar{x}} = \bar{x}$

Generating random variates

We want to generate random variates from a selected distribution

Considerations:

- Exactness
- Efficiency
- Complexity
- Robustness

Inverse Transformation method (ITM):

CDF dice

$F(x) = Pr[X \leq x] = \sum P(x_i)$

1) Generate $U \sim U(0,1)$
2) determine smallest possible integer I s.t. $U \leq F(x_I)$ and return $X = X_I$

1) Generate $U \sim (0,1)$
2) return $X = F^{-1}(U)$

$$f(x) \begin{cases} f(x) & a \leq x < b \\ g(x) & b \leq x < c \\ 1 & x > c \end{cases} \Rightarrow F(x) \begin{cases} 0 & x < 0 \\ \int f(x) + C & a \leq x < b \\ \int g(x) + K & b \leq x < c \\ 1 & x > c \end{cases}$$

$$\Downarrow$$

$F^{-1}(x) \{$ inverses

Bootstrapping

Make simulations, get outcomes and consider them RV

Consider mean of observations $\bar{x}(n) = \frac{1}{n} \sum_{i=1}^{n} x_i$

If normally distributed and known variance σ^2

then α confidence interval $= \bar{x}(n) \pm z_{1-\frac{\alpha}{2}} \sqrt{\sigma^2}$

If variance unknown extract from data $s^2(n) = \sum_{i=1}^{n} \frac{(X_i - \bar{X}(n))^2}{n-1}$

this adds to uncertainty so student t-distribution for c_i with $n-1$ dof

If x_n correlated then biased variance. Observations should be independent and normally distributed.

Central Limit Theorem $x_1, \dots, x_n$ IID RV with μ finite and $\sigma^2 > 0$ then

$$Z_n = \frac{\bar{x}(n) - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$ then $\lim_{n \to \infty} Z_n$ st. norm. distr.

Non normal data: bootstrapping

t-test

$H_0 : \mu = 0$

$$t = \frac{\bar{x}(n)}{\sqrt{Var(\bar{x}(n))}} = \frac{\bar{x}(n)}{\sqrt{\frac{s^2(n)}{n}}}$$

std. normally distr. under assumption that H_0 is true

Comparing alternative system configurations

Paired approach

If observations of each system are independent and normally distributed

$$z_i = X_{1i} - X_{2i}$$

H_0 hypothesis $\bar{z}(n) = 0$

Option 1: construct c.i.

$$\bar{z}(n) \pm t_{1-\alpha/2, n-1} \sqrt{\frac{s_z^2(n)}{n}}$$

if 0 in c.i. no reason to reject H_0, else reject

Option 2: t-test

$$t = \frac{\bar{z}(n)}{\sqrt{\frac{s_z^2(n)}{n}}}$$

test statistic that follows a n-1 dof t-distr

Unpaired approach

Welch's approach / unpooled variance

Can have different sample sizes

mean of system 1

$$\bar{X}_1(n_1) = \sum_{i=1}^{n} \frac{x_{1i}}{n_1}$$

mean of system 2

$$\bar{X}_2(n_2) = \sum_{i=1}^{n} \frac{x_{2i}}{n_2}$$

$$s_1^2(n_1) = \frac{\sum (x_{1i} - \bar{X}_1(n_1))^2}{n_1 - 1}$$

$$Var[\bar{X}_1(n_1)] = \frac{s_1^2(n_1)}{n_1}$$

Variance of $\bar{X}_1(n_1) - \bar{X}_2(n_2)$?

$$\hat{Var}\left[\bar{x}_1(n_1) - \bar{x}_2(n_2)\right] = \hat{Var}\left[\bar{x}_1(n_1)\right] + \hat{Var}\left[\bar{x}_2(n_2)\right] - 2Cov\left(\bar{x}_1(n_1) - \bar{x}_2(n_2)\right)$$

assume independence

C.I.:

$$\bar{x}_1(n_1) - \bar{x}_2(n_2) \pm t_{1-\alpha/2, \hat{f}} \sqrt{\frac{S_1^2(n_1)}{n_1} + \frac{S_2^2(n_2)}{n_2}}$$

$\hat{f}$ = longass formula on formula sheet

For t-test = c_i to 0

test normality with Lilliefors test
Wilcoxon test if not normal

Ranking and selection (Dudewicz & Dalal)

assumptions:

- independence across obs / runs
- independence across systems
- normally distributed obs

Procedure:

- For each of k systems obtain $n_0 = 20$ or 40 obs yielding

$$\bar{X}_i^{(i)}(n_0) = \sum_{j=1}^{n_0} \frac{X_{ij}}{n_0}$$

$$S_i^2(n_0) = \sum_{j=1}^{n_0} \frac{(X_{ij} - \bar{X}_j^{(i)}(n_0))^2}{n_0 - 1}$$

total number of repl needed $N_i = \max\left\{ n_0 + 1, \left\lceil \frac{h_1^2 S_i^2(n_0)}{d^*} \right\rceil \right.$

d^* minimum important difference
probability of selecting actual best system is p^*

check table in formula sheet

- make for each system i, $N_i - n_0$ additional replications and compute mean from them
- compute weights

$$W_{i1} = \frac{n_0}{N_i} \left\{ 1 + \sqrt{1 - \frac{N_i}{n_0} \left[1 - \frac{(N_i - n_0)(d^*)^2}{h_1^2 \, S_1^2(n_0)} \right]} \right\}$$

$$W_{i2} = 1 - W_i$$

- create grand total sampling means

$$\bar{X}_i(N_i) = W_{i1} \bar{X}_i^{(1)}(n_0) + W_{i2} \bar{X}_i^{(2)}(N_i - n_0)$$

and select best one

Variance reduction

Common Random numbers (CRN)

Taking into account Covariance to narrow c.i.

Pair system's observations $E[u^k] = \int_0^1 u^k du = \frac{1}{k+1}$

Antithetic Variables (AV)

offset small obs by large ones

average $x_j^{(1)}$ normal run using random numbers u
$x_j^{(2)}$ 〃 $1-u$