# Lecture 4 : Simulation 

Prof. Hee Yong Youn<br>College of Software<br>Sungkyunkwan University<br>Suwon, Korea<br>youn7147@skku.edu

## Modeling and evaluation

$\square$ Modeling and evaluation
$\square$ Two approaches :

- analytical modeling or simulation
$\square$ Analytical modeling :
- math oriented, abstract, compact, simple, and economical
$\square$ Simulation :
- Rule oriented mimicing the phenomenon modeled.

Relies on $\qquad$ number of iterations to extract useful information.

Detailed, complex, flexible, and expensive

## Simulation and emulation

$\square$ Simulation vs. emulation
$\square$ Emulation : simulation at the level of $\qquad$ code
$\square$ Simulation : much ___ detailed than emulation
$\square$ Generating random numbers
$\square$ Required to simulate $\qquad$ phenomenon with random variates
$\square$ Using physical process or computer
$\square$ Random number generation by computer

- Pseudo-random due to its $\qquad$
- Appears to be $\qquad$ statistically
$\square$ Midsquare/ linear congruential/ additive congruential


## Midsquare method

$\square$ Midsquare method
$\square(\mathrm{ex})$ seed $=5318$

$$
\begin{aligned}
& 5318^{2}=28 \underline{281124} \\
& 2811^{2}=07 \underline{001721} \\
&{9017^{2}}^{=} 81306289 \\
& \vdots
\end{aligned}
$$

- Flaw: possibly stuck
(ex) 99, 80, 40, 60, 60, ...


## Linear congruential method

$\square$ Linear congruential method
$\square z_{n+1}=\left(a z_{n}+c\right) \bmod m$
$\square$ The number of random numbers in a sequence cycle $\left(N_{r}\right) \leq m$
$\square N_{r}=\boldsymbol{m}$ (means the longest cycle) iff
i) $\operatorname{gcd}(m, c)=1$
ii) $(a-1)$ mod (all prime factors of $m)=0$
iii) if $(\boldsymbol{m} \bmod 4)=0$ then $((a-1) \bmod 4)=0$

## Linear congruential method

$\square$ Constant selection ( $w=$ word length)

- $m=\mathbf{2}^{\text {w }}$
- $c=2^{w-1} \pm 1$
- $a=p^{k}\left(p:\right.$ prime and $\left.\left(p^{k}-1\right) \bmod 4=0\right)$ (ex) $p=13$
$\left(13^{4}-1\right) \bmod 4=28560 \bmod 4=0$
$\left(13^{5}-1\right) \bmod 4=371292 \bmod 4=0$
$\left(13^{6}-1\right) \bmod 4=4826809 \bmod 4=0$


## Mixed congruential method

$\square$ Mixed congruential method
$\square$ For linear congruential, when $c>$ $\qquad$
$\square(E x 4.1)$ Generate a sequence with $m=16, a=13, c=11$, seed $=1$
$\square$ Check the conditions of $m, a, c$ for max. period
$\square(13 \times 1+11) \bmod 16=8$
$(13 \times 8+11) \bmod 16=115 \bmod 16=3$
$(13 \times 3+11) \bmod 16=50 \bmod 16=2$
$\vdots$

## Multiplicative congruential method

For linear congruential, with $c=$ $\qquad$
$\square$ Max. period $=\frac{m}{4} \quad$ if $z_{0}$ is odd and $a=8 k+1$ for some $k$
(ex) $m=2^{4}=16, z_{0}=3, a=9$
$\left.\begin{array}{l}(9 \times 3) \bmod 16=11 \\ (9 \times 11) \bmod 16=3\end{array}\right\}$
$3,11,3,11, \ldots($ period $=$ $\qquad$
if $z_{0}=10,(9 \times 10) \bmod 16=10,10, \ldots$ (So, seed selection is important!)
$\square(E x 4.2)$ Generate a sequence with $m=16, a=9$, seed $=7$
$\square(9 \times 7) \bmod 16=63 \bmod 16=15$
$(9 \times 15) \bmod 16=135 \bmod 16=7($ period $=$ $\qquad$ )

- If $a=11$,
$(11 \times 7) \bmod 16=77 \bmod 16=13$
$(11 \times 13) \bmod 16=143 \bmod 16=15$
$(11 \times 15) \bmod 16=165 \bmod 16=5$
$(11 \times 5) \bmod 16=55 \bmod 16=7($ period $=$ $\qquad$ ) (So, $a$ is important!)


## Multiplicative congruential method

$\square$ Good selection of ' $a$ ' and seed
$\square a \approx \sqrt{\mathrm{~m}}$ and the last three bits are ' 011 ' or ' 101 '
$\square$ Also, the number of 1's should be as $\qquad$ as possible
$\square$ Seed must be $\qquad$

$$
\text { (ex) } a=0101=5 \text {, seed }=3
$$

$\left.\begin{array}{rl}(5 \times 3) \bmod 16 & =15 \\ (5 \times 15) \bmod 16 & =11 \\ (5 \times 11) \bmod 16 & =7 \\ (5 \times 7) \bmod 16 & =3 \\ (5 \times 3) \bmod 16 & =15\end{array}\right\} \frac{16}{4}$
$\square$ Good choice for 16-bit computer

$$
\begin{aligned}
& m=2^{16}, a=2^{8}+5=261=(100000101)_{2} \\
& z_{0}=129=(10000001)_{2}
\end{aligned}
$$

## Additive congruential

$\square$ For random number sequence whose period is greater than $2^{w}$
$\square$ Use an existing sequence
$\square z_{n}=\left(z_{n-1}+z_{n-k}\right) \bmod m$

- If $\boldsymbol{k}=\mathbf{2}$, Fibonacci generator
$\square$ Generalization of additive congruential
$\square z_{n}=\left(\sum_{j=1}^{k} a_{j} z_{n-j}\right) \bmod m$
$\square$ The $\qquad$ is as large as $m^{k}-1$
$\square$ Generate a sequence with $0,7,10,9$, and $k=4, a_{i}=1(1 \leq i \leq 4)$

$$
\begin{aligned}
& z_{4}=\left(z_{3}+z_{0}\right) \bmod 16=(0+9) \bmod 16=9 \\
& z_{5}=(7+9) \bmod 16=0 \\
& z_{6}=(10+0) \bmod 16=10 \\
& z_{7}=(9+10) \bmod 16=3 \\
& \vdots
\end{aligned}
$$

$\square$ Are the numbers in the sequence unique?

## Tausworthe's bitwise manipulation

$\square b_{n}=\left(b_{n-r}+b_{n-q}\right) \bmod 2=b_{n-r} \oplus b_{n-q}$
$\square$ The $\qquad$ is independent of $\boldsymbol{w}$
$\square(E x) \boldsymbol{b}_{\boldsymbol{n}}=\boldsymbol{b}_{\boldsymbol{n - 1}} \oplus \boldsymbol{b}_{\boldsymbol{n}-3}$, seed $=5(\mathbf{0 1 0 1})$

$$
b_{0} b_{1} b_{2} b_{3} b_{4} b_{5} b_{6} b_{7} b_{8} b_{9}
$$

0101
5
1010
10
0
4
$1 \quad 9$

13
17
014

## Validation techniques

$\square$ Validation techniques
$\square$ A random number sequence is checked for

- ___ distribution


## Check for uniform distribution

$\square \chi^{2}$ ( chi-squared ) method
$\square$ Comparison of two $\qquad$ $s$

- A $\qquad$
$\qquad$ generated from the two r.v.'s compared
$\square$ If $\chi^{2} \rightarrow \mathbf{0}$, it is said that the two r.v.'s have the same limiting distribution with probability $1-F\left(\chi^{2}\right)$
- $\chi^{2} \equiv \sum_{n=1}^{N} \frac{\left(O_{n}-E_{n}\right)^{2}}{E_{n}}$
$O_{n}$ : number of observed occurrences in a subrange $n$
$E_{n}$ : number of expected occurrences by the assumed density for the subrange $n$
- Why divide by $\boldsymbol{E}_{\boldsymbol{n}}$ ?


## Table lookup method

$\square N$ : Number of subranges
$\square R$ : Number of parameters in the PDF
$R=1$ for uniform, poisson, exponential
$R=\mathbf{2}$ for normal and Weibull
$\square$ D.F. $($ degree of freedom) $=N-R$

$\square$ "If $\chi^{2} \leq x$, then the distribution cannot be rejected with probability F. In other words, accepted with probability 1-F."

## Table lookup method ( $\chi^{2}$ table)

Chi-Squared Distribution

| $n F$ | .010 | .050 | .100 | .250 | .500 | .750 | .900 | .950 | .990 |
| ---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 3 | .115 | .352 | .584 | 1.21 | 2.37 | 4.11 | 6.25 | 7.81 | 11.3 |
| 4 | .297 | .711 | 1.06 | 1.92 | 3.36 | 5.39 | 7.78 | 9.49 | 13.3 |
| 5 | .554 | 1.15 | 1.61 | 2.67 | 4.35 | 6.63 | 9.24 | 11.1 | 15.1 |
| 6 | .872 | 1.64 | 2.20 | 3.45 | 5.35 | 7.84 | 10.6 | 12.6 | 16.8 |
| 7 | 1.24 | 2.17 | 2.83 | 4.25 | 6.35 | 9.04 | 12.0 | 14.1 | 18.5 |
| 8 | 1.65 | 2.73 | 3.49 | 5.07 | 7.34 | 10.2 | 13.4 | 15.5 | 20.1 |
| 9 | 2.09 | 3.33 | 4.17 | 5.90 | 8.34 | 11.4 | 14.7 | 16.9 | 21.7 |
| 10 | 2.56 | 3.94 | 4.87 | 6.74 | 9.34 | 12.5 | 16.0 | 18.3 | 23.2 |
| 11 | 3.05 | 4.57 | 5.58 | 7.58 | 10.3 | 13.7 | 17.3 | 19.7 | 24.7 |
| 12 | 3.57 | 5.23 | 6.30 | 8.44 | 11.3 | 14.8 | 18.5 | 21.0 | 26.2 |
| 13 | 4.11 | 5.89 | 7.04 | 9.30 | 12.3 | 16.0 | 19.8 | 22.4 | 27.7 |
| 14 | 4.66 | 6.57 | 7.79 | 10.2 | 13.3 | 17.1 | 21.1 | 23.7 | 29.1 |
| 15 | 5.23 | 7.26 | 8.55 | 11.0 | 14.3 | 18.2 | 22.3 | 25.0 | 30.6 |
| 16 | 5.81 | 7.96 | 9.31 | 11.9 | 15.3 | 19.4 | 23.5 | 26.3 | 32.0 |
| 17 | 6.41 | 8.67 | 10.1 | 12.8 | 16.3 | 20.5 | 24.8 | 27.6 | 33.4 |
| 18 | 7.01 | 9.39 | 10.9 | 13.7 | 17.3 | 21.6 | 26.0 | 28.9 | 34.8 |
| 19 | 7.63 | 10.1 | 11.7 | 14.6 | 18.3 | 22.7 | 27.2 | 30.1 | 36.2 |
| 20 | 8.26 | 10.9 | 12.4 | 15.5 | 19.3 | 23.8 | 28.4 | 31.4 | 37.6 |
| 21 | 8.90 | 11.6 | 13.2 | 16.3 | 20.3 | 24.9 | 29.6 | 32.7 | 38.9 |
| 22 | 9.54 | 12.3 | 14.0 | $17: 2$ | 21.3 | 26.0 | 30.8 | 33.9 | 40.3 |
| 23 | 10.2 | 13.1 | 14.8 | 18.1 | 22.3 | 27.1 | 32.0 | 35.2 | 41.6 |
| 24 | 10.9 | 13.8 | 15.7 | 19.0 | 23.3 | 28.2 | 33.2 | 36.4 | 43.0 |
| 25 | 11.5 | 14.6 | 16.5 | 19.9 | 24.3 | 29.3 | 34.4 | 37.7 | 44.3 |
| 26 | 12.2 | 15.4 | 17.3 | 20.8 | 25.3 | 30.4 | 35.6 | 38.9 | 45.6 |
| 27 | 12.9 | 16.2 | 18.1 | 21.7 | 26.3 | 31.5 | 36.7 | 40.1 | 47.0 |
| 28 | 13.6 | 16.9 | 18.9 | 22.7 | 27.3 | 32.6 | 37.9 | 41.3 | 48.3 |
| 29 | 14.3 | 17.7 | 19.8 | 23.6 | 28.3 | 33.7 | 39.1 | 42.6 | 49.6 |
| 30 | 15.0 | 18.5 | 20.6 | 24.5 | 29.3 | 34.8 | 40.3 | 43.8 | 50.9 |
|  |  |  |  |  |  |  |  |  |  |

This table is reprinted with permission from Standard Mathematical Tables © 1976 CRC Press, Boca Raton, FL.

## Table lookup method (test procedure)

$\square$ Input is the table containing the sample data $\left(O_{n}\right)$

1. Predict the distribution based on the sample data (also decide the parameter values)
2. Obtain the expected number of occurrences for each subrange using the predicted distribution $\left(E_{n}\right)$
3. Obtain $\chi^{2}$
4. Make a decision

## Table lookup method

$\square$ (Em 4.8)

| No. of <br> broken parts | No. of <br> Shipments | Predicted no. <br> of shipments |
| :---: | :---: | :--- |
| 0 | 10 |  |
| 1 | 41 |  |
| 2 | 57 |  |
| 3 | 55 |  |
| 4 | 43 |  |
| 5 | 27 |  |
| 6 | 11 |  |

1. Assume $\qquad$ distribution
Total shipments $=244$
Expected no of broken parts/shipment $=$ $(0 \times 10+1 \times 41+2 \times 57+\ldots+6 \times 11) / 244=2.84$
2. $P[k=0]=\left(2.84^{0} / 0!\right) \mathrm{e}^{-2.84}=0.0584$,
$\mathrm{E}_{0}=244 \times 0.0584=14.3$
$P[k=1]=\left(2.84^{1} / 1!\right) \mathrm{e}^{-2.84}=0.17$,
$E_{1}=244 \times 0.17=40.5$
3. $\chi^{2}=(10-14.3)^{2} / 14.3+(41-40.5)^{2} / 40.5+\ldots$

$$
=3.034
$$

4. D.F. $=7-1=6$, from $\chi^{2}$ table, accept the model with probability of $\mathbf{0 . 7 5}$.

## Table lookup method

$\square$ (Ex 4.4) Would you accept with $\mathbf{9 0 \%}$ confidence, a RNG of the following table?

| Range | Occurrence |
| :---: | :---: |
| $0.0 \sim 0.1$ | 21 |
| $0.1 \sim 0.2$ | 20 |
| $0.2 \sim 0.3$ | 19 |
| $0.3 \sim 0.4$ | 17 |
| $0.4 \sim 0.5$ | 22 |
| $0.5 \sim 0.6$ | 21 |
| $0.6 \sim 0.7$ | 20 |
| $0.7 \sim 0.8$ | 18 |
| $0.8 \sim 0.9$ | 21 |
| $0.9 \sim 1.0$ | 21 |

1. Assume $\qquad$ distribution
Total occurrence $=200$
Expected no occurrences/subrange = $200 / 10=20$
2. $\mathrm{E}_{0}=20$
$\mathrm{E}_{1}=20$

+ 

3. $\chi^{2}=(21-20)^{2} / 20+(20-20)^{2} / 20+\ldots$

$$
=22 / 20=1.1
$$

4. D.F. $=10-1=9$, from $\chi^{2}$ table, accept the model with probability of $\mathbf{0 . 9 9}$.

## Check for independence

$\square$ Serial test
$\square$ Run-up test
$\square$ Relative distance test
$\square$ Serial-correlation test
$\square$ Serial test
$\square$ Check increasing $n$-tuples of the sequence for $\qquad$ with $\chi^{2}$ test method

## Run-up test

$\square$ Count the no. of $\qquad$ increasing sequences of length $1,2, \ldots$, or $\geq 6$
$\square$ Construct an r.v. from the data, and test it for independency with $\chi^{\mathbf{2}}$
$\square R=\frac{1}{N} \sum_{j=1}^{6} \sum_{k=1}^{6} A_{j k}\left(r_{j}-N B_{j}\right)\left(r_{k}-N B_{k}\right)$
( $r$ : No. of run-ups of length $i, N$ : length of the sequence, $\mathrm{DF}=6$ )
$\square(E m 4.10) \frac{0710}{3} \frac{9}{1} \frac{41114}{3} \frac{14}{1} \frac{515}{2} \frac{812}{2} \frac{7}{1} \frac{5}{1}$

$$
r_{1}=\ldots, r_{2}=\ldots, r_{3}=\ldots, r_{4}=0, r_{5}=0, r_{6}=0
$$

$\square R=2.59$, and conclude that the sequence is independent with confidence 0.75

## Relative distance(RD) test

$\square$

$\square \mathbf{R D}=\left\{\begin{array}{l}\left(1-z_{n}\right)+z_{n+1} \\ z_{n+1}-z_{n}\end{array}\right.$
if $z_{n+1}<z_{n}$ otherwiseTest RD sequence for $\qquad$ with $\chi^{2}$

## Serial-correlation test

$\square$ Covariance of r.v.'s

$$
\begin{aligned}
\operatorname{cov}(X, Y) & =E[(X-E[X])(Y-E[Y])] \\
& =E[X Y]-2 E[X] E[Y]+E[X] E[Y] \\
& =E[X Y]-E[X] E[Y]
\end{aligned}
$$

$\square$ If $X$ and $Y$ are independent, then $E[X Y]=E[X] E[Y] \rightarrow \operatorname{cov}(X, Y)=0$
$\square$ If $\operatorname{cov}(X, Y) \geq 0$, then $X, Y$ are $\qquad$
$\square \boldsymbol{R}_{\boldsymbol{k}}$ : autocovariance at $\log k$ between $x_{n}$ and $x_{n+k}$
$\square R_{k}=\frac{1}{n-k} \sum_{i=1}^{n-k}\left(z_{i}-\mathbf{1} / 2\right)\left(z_{i+k}-1 / 2\right)$
$n$ : No. of elements in the sequence

$$
E[Z]=1 / 2
$$

$\square$ When $n \rightarrow \infty, R_{k} \rightarrow N\left(0,\left(\frac{1}{12 \sqrt{n-k}}\right)^{2}\right)$

## Serial-correlation test

(ex) $x_{n}=7^{5} x_{n-1} \bmod \left(2^{31}-1\right), n=10,000$

|  |  | Standard <br> Lag <br> k | Autocovariance <br> $\mathrm{R}_{\mathrm{k}}$ | Deviation <br> of $\mathrm{R}_{\mathrm{k}}$ |
| :---: | :---: | :---: | :---: | :---: |
|  |  |  | Upper Limit |  |
| 1 | -0.000038 | 0.000833 | -0.001409 | 0.001333 |
| 2 | -0.001017 | 0.000833 | -0.002388 | 0.000354 |
| 3 | -0.000489 | 0.000833 | -0.001860 | 0.000882 |
| 4 | -0.000033 | 0.000834 | -0.001404 | 0.001339 |
| 5 | -0.000531 | 0.000834 | -0.001902 | 0.000840 |
| 6 | -0.001277 | 0.000834 | -0.002648 | 0.000095 |
| 7 | -0.000385 | 0.000834 | -0.001757 | 0.000986 |
| 8 | -0.000207 | 0.000834 | -0.001579 | 0.001164 |
| 9 | 0.001031 | 0.000834 | -0.000340 | 0.002403 |
| 10 | -0.000224 | 0.000834 | -0.001595 | 0.001148 |

## Serial-correlation test (procedure)

1. Get $R_{k}$
2. Obtain standard deviation by $\frac{1}{12 \sqrt{n-k}}$
3. Obtain confidence interval
a. $90 \%$ confidence $\rightarrow \alpha=0.1, p=1-\alpha / 2=0.95, Z_{0.95}=1.645$

4. If the interval does not include 0 , then the correlation is $\qquad$ .
$R_{k}$ is the $\qquad$ point of the limits.

## Recommended random number generation

$\square$ For 32-bit

- $z_{n}=A z_{n-1} \bmod \left(2^{31}-1\right)$ $A=16807$ (minimal standard with period of $2^{31}-2$ ), 48271, or 69621
$\square$ For 16-bit
- $z(i)_{n}=A_{i} z(i)_{n-1} \bmod B_{i}$

$$
\begin{array}{lll}
A_{1}=157 & A_{2}=146 & A_{3}=142 \\
B_{1}=32363 & B_{2}=31727 & B_{3}=\mathbf{3 1 6 5 7}
\end{array}
$$

- $z_{n}=\left(z(1)_{n}-z(2)_{n}-z(3)_{n}\right) \bmod 32362$


## Overflow problem

$\square$ Overflow problem of $A z_{n-1}$ with LCG method

$$
\operatorname{ax} \bmod m=g(x)+m h(x)
$$

$$
\left\{\begin{array}{l}
g(x)=a(x \bmod q)-r(x \operatorname{div} q) \\
h(x)=(x \operatorname{div} q)-(\operatorname{ax} \operatorname{div} m)
\end{array}\right.
$$

$q=m \operatorname{div} a, r=m \bmod a$
If $\boldsymbol{q}>\boldsymbol{r}$, then

$$
h(x)= \begin{cases}1 & \text { if } g(x)<0 \\ 0 & \text { otherwise }\end{cases}
$$

$\square$ We must decide $a$ such that $q>r$ to avoid the computation of $h(x)$
$\square(\mathrm{Ex}) a=3, x=6, m=15$

## Seed selection and use of sequence

$\square$ Avoid $\qquad$ values
$\square$ Do not $\qquad$ a stream
$\square$ Do not $\qquad$ streams
$\square$ Do not use $\qquad$ seeds
$\square$ Since no $\qquad$

- No guarantee of avoidance of overlap


## Generating random numbers of nonuniform distribution

$\square$ Given : $\qquad$ of desired r.v.
$\square$ Obtained: random numbers of the desired distribution
$\square$ Inversion/ composition/ rejection method

## Inversion method

$\square$ Inversion method
$Z \equiv F(Y)(Y$ is desired r.v.)

$$
\begin{aligned}
G(z) & =P[Z \leq z] \\
& =P[F(Y) \leq z] \\
& =P\left[Y \leq F^{-1}(z)\right] \\
& =F\left(F^{-1}(z)\right) \\
& =z, 0 \leq z<1
\end{aligned}
$$

## Inversion method

$\square($ ex) exponential distribution


$$
\begin{aligned}
& \begin{array}{l}
F(y)=1-\mathrm{e}^{-\lambda y}=z \\
-\mathrm{e}^{-\lambda y}=z-1 \\
-\lambda y=\ln (1-z) \\
y=-\frac{1}{\lambda} \ln (1-z)
\end{array} \\
& \text { Since } 0 \leq z<1, y=-\frac{1}{\lambda} \ln z
\end{aligned}
$$

## Composition method

$\square$ Generate a random variable using the composite of another random variable of a different distribution
$\square$ (ex) Poisson distribution ( $\lambda$ ) ( $T$ is given input constant)

- Interarrival time is $\qquad$ distribution
$\square$ The value of target random variable is the number of random variables of exponential distribution whose sum $\geq T$
$\square$ Procedure
i) Generate random numbers of $\qquad$ distribution, $x_{1}, x_{2}, \ldots$
ii) Stop when $x_{1}+\cdots+x_{m} \geq T$
iii) $k=m-1$


## Composition method (Poisson distribution)

$\square$ Since $x_{i}=-\frac{1}{\lambda} \ln z_{i}$,

$$
\begin{aligned}
& -\frac{1}{\lambda}\left(\ln z_{1}+\ln z_{2}+\ldots+\ln z_{m}\right)=-\frac{1}{\lambda} \ln \left(z_{1} z_{2} \ldots z_{m}\right) \geq T \\
& \ln \left(z_{1} z_{2} \cdots z_{m}\right) \leq-\lambda T \\
& z_{1} z_{2} \cdots z_{m} \leq e^{-\lambda T}
\end{aligned}
$$

So i) Generate $\qquad$ r.v. $z_{1}, z_{2}, \ldots$
ii) Stop when $z_{1} z_{2} \ldots z_{m} \leq e^{-\lambda T}$
iii) $k=m-1$

## Composition method (Poisson distribution)

ㅁ (Proof)
$\operatorname{Prob}\left[\left(x_{1}+\ldots+x_{m}\right) \geq T\right]=\frac{1}{(m-1)!} \int_{\lambda}^{\infty} t^{m-1} e^{-t} d t($ if $T=1)$
$\operatorname{Prob}[k=n]=\frac{1}{n!} \int_{\lambda}^{\infty} t^{n} e^{-t} d t-\frac{1}{(n-1)!} \int_{\lambda}^{\infty} t^{n-1} e^{-t} d t$
$=\boldsymbol{e}^{-\lambda} \frac{\lambda^{n}}{n!}$

## Composition method

$\square$ (ex) Geometric distribution(p)
$\square$ Procedure 1: count the number of consecutive trials up to the $\qquad$ success
$\square$ Procedure 2: $k=\left\lceil\frac{\ln z}{\ln (1-p)}\right\rceil$
$\left\lceil\frac{\ln z}{\ln (1-p)}\right\rceil=n$ iff $\quad(n-1)<\frac{\ln z}{\ln (1-p)} \leq n$

$$
n \ln (1-p)<\ln z<(n-1) \ln (1-p)
$$

$$
(1-p)^{n}<z \leq(1-p)^{n-1} \equiv \text { event } A
$$

$$
\operatorname{Prob}[A]=(1-p)^{n-1} p
$$

## Composition method

$\square$ (ex) Binomial distribution(p,n)

- Procedure
i) Generate $z_{1}, \ldots, z_{\mathrm{n}}$
ii) $k=$ no. of $z_{i}^{\prime}$ 's which are $<p$
- Procedure for small $p$
i) $\operatorname{sum}=0 ; i=1$
ii) $G_{i}(p)=\left\lceil\frac{\ln z_{i}}{\ln (1-p)}\right\rceil ; \operatorname{sum}=\operatorname{sum}+G_{i}(p)$
ii) If $\operatorname{sum}<n$, then $i=i+1$; go to ii)

$$
\text { else } k=i-1
$$

## Composition method (Binomial distribution)

$\square$ Inversion method
i) Compute $F(x), x=0,1,2, \ldots, n$ and store.
ii) Generate $z_{i}$, and find $x$ such that

$$
F(x) \leq z_{i}<F(x+1) ; k=x
$$

## Rejection method

$\square$ Generate r.v. $X$ of density $f(t)$ using another r.v. $Y$ of density $g(t)$, where $f(t) \leq \operatorname{cg}(t)$ for all $t(c$ is a positive constant. $c g(t)$ covers $f(t))$
$\square$ Good for small $c$
$\square$ Procedure
i) Generate $X$ using $g(t)$
ii) Generate $z$
iii) if $z \geq \frac{\mathrm{f}(\mathrm{x})}{\operatorname{cg}(\mathrm{x})}$, then reject $X$, go to i)
else $X$ is accepted

## Rejection method

$\square$ (ex) normal distribution $(\mu, \sigma)$
i) Generate $z_{1}, x=-\ln z_{1} \quad(\operatorname{exponential}$ distribution with $\lambda=1)$
ii) Generate $z_{2}$
iii) If $z_{2}>\mathrm{e}^{-(\mathrm{x}-1)^{2} / 2}$, then go to i)
iv) Generate $z_{3}$ (to decide left or right after accepted)
v) If $z_{3}>0.5$, then return $\mu+\sigma x$
else return $\mu-\sigma x$

## Simulations

$\square$ Simulation time: time expired in $\qquad$ system
$\square$ Run time: time expired for the $\qquad$ of simulation
$\square$ Events: operations causing a $\qquad$ of state

## Types of simulation

$\square$ Time-based/event based
$\square$ Time-based simulation
$\square$ $\qquad$ control
$\square$ Good for the problems having $\qquad$ events at any moment
$\square$ Highly likely an $\qquad$ at each clock
$\square$ Event-based simulation
$\square$ Execution of main control loop represents a single $\qquad$
$\square$ Use event queue

## Types of simulation



FIGURE 4.4 Event-Based Simulation Control Flow

## Comparison of time vs. event-based simulation

$\square$ (ex) Simulation of Poisson arrival and the processing of the jobs

- Time-based simulation
while $t<t_{\text {max }}$ do
- next job arrives $t_{j}$ time later according to $\qquad$ distribution
- $t_{n}=t+t_{j}$
- until $t_{n}<t$ do $t=t+\Delta t$
- process new job
- collect statistical data
report statistics


## Comparison of time vs. event-based simulation

- Event-based simulation
while $t<t_{\text {max }}$ do
- $t=t+\Delta t$
- $N$ jobs arrive during $\Delta t$ according to $\qquad$ distribution
- Process $N$ new jobs
- collect statistical data
report statistics


## Comparison of time vs. event-based simulation

## EX 4.15.Time-based simulation for instruction execution speed

## Get input parameter


(Input)
-Frequency of inst length
( $1 / 2 / 3$ ) in PROBIN
-Memory speed (MEMSTEP)
-Prob of memory access with
each length of inst (PROBMEM)
(Output)
-No of instruction execution per sec $=$
EXECI/(MAXSTEP x Clock period)

## Comparison of time vs. event-based simulation

EX 4.16. Event-based simulation for disk access time

(Input)
-P: prob of accessing the same track
-No of tracks: 75
-LATENCY: one rotation time
-SEEK: time taken for the arm to move one track
(Output)
-AVG.DELAY: time taken for one file access

## Accumulating statistics

$\square$ Accumulating statistics
$\square$ Trace or logs : too voluminous
$\square$ Statistics : useful behavior characteristics

- Mean

$$
\bar{k}=\mathrm{E}[k]=\sum_{k=-\infty}^{\infty} k f_{k}
$$

- Standard deviation

$$
\begin{aligned}
& \sigma_{k}=\left(\mathbf{E}\left[k^{2}\right]-(\mathrm{E}[k])^{2}\right)^{1 / 2} \\
& \mathbf{E}\left[k^{2}\right]=\sum_{k=-\infty}^{\infty} k^{2} f_{k}
\end{aligned}
$$

## Accumulating statistics

$\square$ Any particular simulation run is a particular $\qquad$ of the stochastic process
$\square$ Calculation $\qquad$ does not require to store all sample points but the running sum

## Accumulating statistics

- (ex) simulation data: $1,3,0,2,1,4,2$

Frequency table:

| $K$ | $\mathbf{0}$ | $\mathbf{1}$ | 2 | 3 | 4 |
| :--- | :--- | :--- | :--- | :--- | :--- |
| No. of observations | 1 | 2 | 2 | 1 | 1 |
| PDF | $1 / 7$ | $2 / 7$ | $2 / 7$ | $1 / 7$ | $1 / 7$ |

$\left\{\begin{array}{l}\bar{k}=\mathrm{E}[k]=1 \times \frac{2}{7}+2 \times \frac{2}{7}+3 \times \frac{1}{7}+4 \times \frac{1}{7}=\frac{13}{7}, ~\end{array}\right.$
Calculation on the fly :

$$
\frac{1+3+0+2+1+4+2}{7}=\frac{13}{7}
$$

$$
\mathbf{E}\left[k^{2}\right]=1 \times \frac{2}{7}+4 \times \frac{2}{7}+9 \times \frac{1}{7}+16 \times \frac{1}{7}=\frac{35}{7}
$$

Calculation on the fly :

$$
\frac{1^{2}+3^{2}+0^{2}+2^{2}+1^{2}+4^{2}+2^{2}}{7}=\frac{35}{7}
$$

## Accumulating statistics

$\square$ Problems of running sum
$\square$ No additional statistics can be extracted later

- Overflow or precision problem
$\square$ Solution
If the range is small and known, use $\qquad$ .
$\square$ Otherwise, use running sum with $\qquad$ frequency array.


## Accumulating statistics

$\square$ (ex) waiting time $\left(t_{w}\right)$

## Customer


$-t_{w}=\frac{\sum_{\mathrm{j}=1}^{5}\left(\mathrm{C}_{\mathrm{j} 0}-\mathrm{C}_{\mathrm{ji}}\right)}{5}$

- Running sum approach

$$
t_{w}=\frac{t_{1} \times 1+t_{2} \times 2+t_{3} \times 1+t_{4} \times 1+t_{5} \times 2+t_{6} \times 3+t_{7} \times 2+t_{8} \times 1}{5}
$$

## Analyzing simulation results

$\square$ Data collection range determines simulation accuracy and length of run time
$\square$ Startup transients


## Analyzing simulation results

$\square$ Approaches for maximizing simulation accuracy

- Run the simulation $\qquad$ (expensive and nondeterministic)
- Truncate data up to certain point ( $\qquad$ state)
$\square$ Start at $\qquad$ point
$\square$ Use running average crossings



## Confidence intervals

$\square$ Repeating simulation with different seeds yields different results
$\square$ If runs are long enough to justify the assumption of $\qquad$ distribution, the $\qquad$ distribution can be used to provide a confidence level and range of values of the results
$\square$ (ex) 5 runs provide $x_{\boldsymbol{k}}$ 's as $11,12,10,14,13$

$$
\begin{aligned}
& \bar{x}=\frac{1}{n} \sum_{k=1}^{n} x_{k}=(11+12+10+14+13) / 5=12 \\
& s^{2}=\frac{1}{n-1} \sum_{k=1}^{n}\left(x_{k}-\bar{x}\right)^{2}=(1+0+4+4+1) / 4=2.5
\end{aligned}
$$

For confidence of $\mathbf{9 8 \%}, v=n-1, t_{4,0.98}=3.747$

$$
\varepsilon=3.747\left(\frac{\mathrm{~s}^{2}}{\mathrm{n}}\right)^{0.5}=2.65
$$

Thus, $x=12 \pm 2.65$ with $98 \%$ confidence

## Student- $t$ distribution

Student-t Distribution

|  | Confidence Probability |  |  |  |
| ---: | :---: | :---: | :---: | ---: |
| $v$ | .80 | .90 | .96 | .98 |
| 1 | 3.078 | 6.314 | 15.895 | 31.821 |
| 2 | 1.886 | 2.920 | 4.849 | 6.965 |
| 3 | 1.638 | 2.353 | 3.482 | 4.541 |
| 4 | 1.533 | 2.132 | 2.999 | 3.747 |
| 5 | 1.476 | 2.015 | 2.757 | 3.365 |
| 6 | 1.440 | 1.943 | 2.612 | 3.143 |
| 7 | 1.415 | 1.895 | 2.517 | 2.998 |
| 8 | 1.397 | 1.860 | 2.449 | 2.896 |
| 9 | 1.383 | 1.833 | 2.398 | 2.821 |
| 10 | 1.372 | 1.812 | 2.359 | 2.764 |
| 11 | 1.363 | 1.796 | 2.328 | 2.718 |
| 12 | 1.356 | 1.782 | 2.303 | 2.681 |
| 13 | 1.350 | 1.771 | 2.282 | 2.650 |
| 14 | 1.345 | 1.761 | 2.264 | 2.624 |
| 15 | 1.341 | 1.753 | 2.249 | 2.602 |
| 16 | 1.337 | 1.746 | 2.235 | 2.583 |
| 17 | 1.333 | 1.740 | 2.224 | 2.567 |
| 18 | 1.330 | 1.734 | 2.214 | 2.552 |
| 19 | 1.328 | 1.729 | 2.205 | 2.539 |
| 20 | 1.325 | 1.725 | 2.197 | 2.528 |
| 25 | 1.316 | 1.708 | 2.167 | 2.485 |
| 30 | 1.310 | 1.697 | 2.147 | 2.457 |
| 40 | 1.303 | 1.684 | 2.123 | 2.423 |
| 50 | 1.299 | 1.676 | 2.109 | 2.403 |
| 75 | 1.293 | 1.665 | 2.090 | 2.377 |
| 100 | 1.290 | 1.660 | 2.081 | 2.364 |
| $\infty$ | 1.282 | 1.645 | 2.054 | 2.326 |

This table is reprinted with permission from Standard Mathematical Tables © 1976 CRC Press, Boca Raton, FL.

## Regenerative simulation

$\square$ If we know the stochastic process for its $\qquad$ points, we can start and stop the simulation at those points.
$\square$ For $n$ of these runs with average length $N$,

$$
\varepsilon=\frac{t_{\alpha, p}}{N}\left(\frac{s^{2}}{n}\right)^{\frac{1}{2}}
$$

$\square$ Due to independency of the r.v.'s ( $n$ of them), if the number of intervals is large, the output variables tend toward a $\qquad$ distribution
$\square$ In practice, regenerative simulation is difficult since finding renewal points is very difficult

