

EE/CpE 345

Modeling and Simulation

Fall 2003

Class 8

Topics

- Finish Chapter 7 material on testing runs, c^2 test
- Chapter 8 – generating random variables with a desired distribution

Example 7.7

- Chi-square test

$$c^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

- O_i is the observed number in the i^{th} class
- E_i is the expected number in the i^{th} class
- n is number of classes
- pick the classes so each has at least 5 in class

- Use 10 uniform width classes $[0,.1)$, $[.1,.2)$, ... $[.9,1.0)$, so $E_i = 10$

- From data in example,
 $O_i = (8, 8, 10, 9, 12, 8, 10, 14, 10, 11)$
- $c^2 = 3.4$, which is less than critical value $c^2_{0.05,9} = 16.9$

Why use Frequency Tests?

- Consider the sequence of random numbers:

.03, .01, .04, .01,

.15, .19, .12, .16,

.25, .24, .23, .21,

.34, .31, .35, .39,

.42, .46, .45, .44,

...

.93, .91, .94, .91

- They would pass the K-S and chi-square test, but are not suitable as random numbers
- Frequency tests will find their deficiencies

Run Tests

- A *run* is a succession of similar events

- Coin flipping example:

HTTHHTTTHT

– six runs are marked

- For sequences of random numbers, define *up* runs and *down* runs, depending on whether successive numbers are increasing or decreasing. E.g.:

.87 .15 .23 .45 .69 .32 .30 .19 .24 .18 .65 .82 .93 .22 .81
↓ ↑ ↓ ↑ ↓ ↓ ↓ ↑ ↓ ↑ ↑ ↑ ↓ ↑

.08 .18 .23 .36 .42 .55 .63 .72 .89 .91
↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑

.08 .93 .15 .96 .26 .84 .28 .79 .36 .57
↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓ ↑



These don't look like
any realistic random sequence

Testing Runs

- For a truly random sequence with N samples, mean and variance of number of runs, a , are:

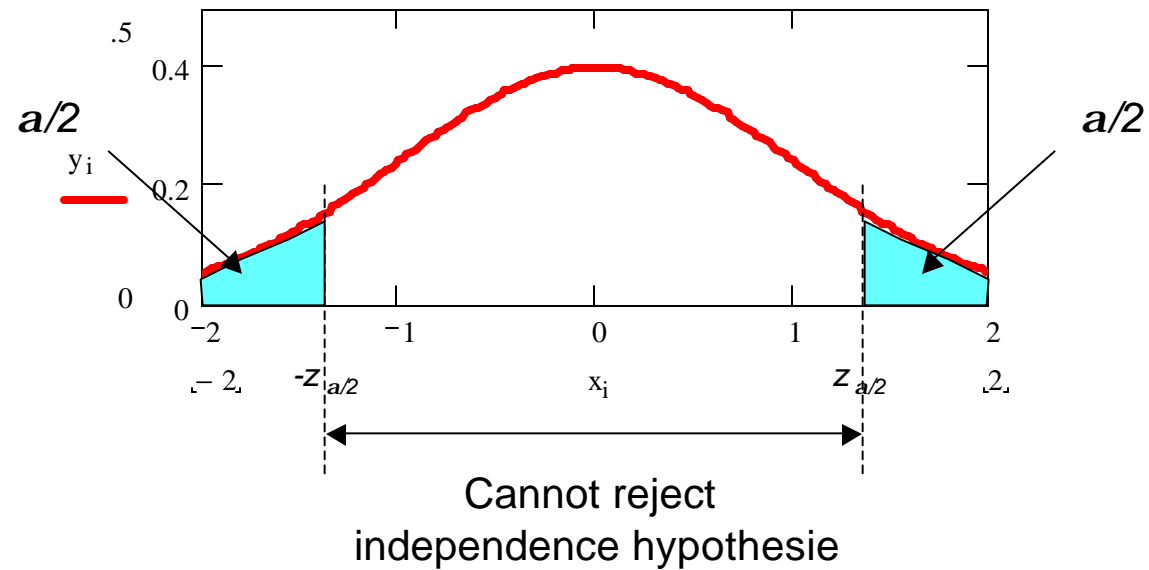
$$\mathbf{m}_a = \frac{2N-1}{3}$$
$$\mathbf{s}_a^2 = \frac{16N-29}{90}$$

- For $N > 20$, this can be approximated by a normal distribution, $N(\mathbf{m}_a, \mathbf{s}_a^2)$
- Observe the number of runs in the data, a
- Calculate test statistic:

$$Z_0 = \frac{a - \mathbf{m}_a}{\mathbf{s}_a} = \frac{a - [(2N-1)/3]}{\sqrt{(16N-29)/90}}$$

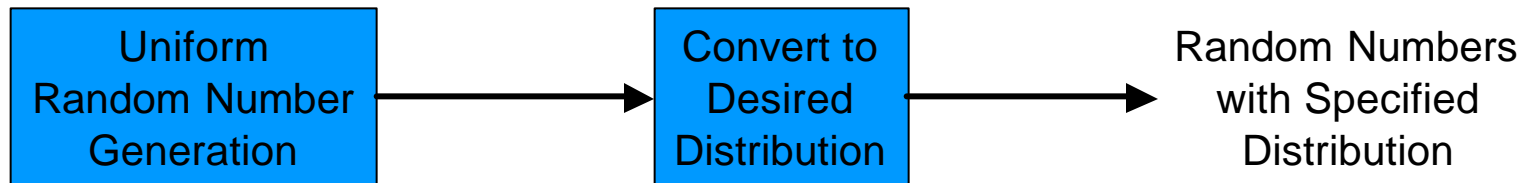
- Compare to a normal distribution

Testing Runs



- Find $z_{a/2}$ for significance $a/2$ from Normal distribution tables
- Check to see if statistic exceeds $z_{a/2}$

Random-Variate Generation



- Given a process to generate uniformly distributed random numbers, how to generate any arbitrary distribution
 - continuous and discrete valued R.V.s
- Techniques:
 - Inverse Transform
 - Convolution Method
 - Acceptance-Rejection
- Doesn't the simulation environment have the distributions needed?
 - Not always, especially for special distributions
 - It's useful to know how they work to understand constraints

Inverse Transform Technique

- Use exponential distribution to illustrate technique
- Given R_i drawn from $U(0, 1)$, generate X_i drawn from exponential distribution
- Most useful when the c.d.f., $F(x)$ can be readily inverted

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$\begin{aligned} 1 - e^{-\lambda X} &= R \\ e^{-\lambda X} &= 1 - R \\ -\lambda X &= \ln(1 - R) \\ X &= -\frac{1}{\lambda} \ln(1 - R) \end{aligned}$$

- Distribution of R and $1-R$ are identical, allowing simplification

$$X = -\frac{1}{\lambda} \ln(R)$$

Inverse Transform Technique

Exponential Distribution

$N := 200$

$i := 0.. N - 1$ $\lambda := 1$

$R := \text{runif}(N, 0, 1)$

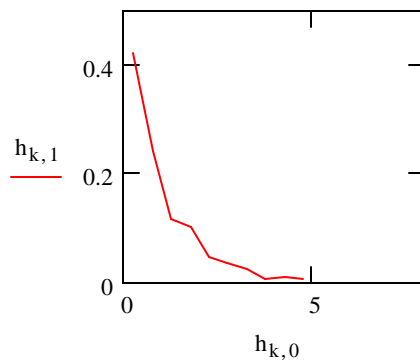
$$X := -\left(\frac{1}{\lambda}\right) \ln(R)$$

$M := 10$

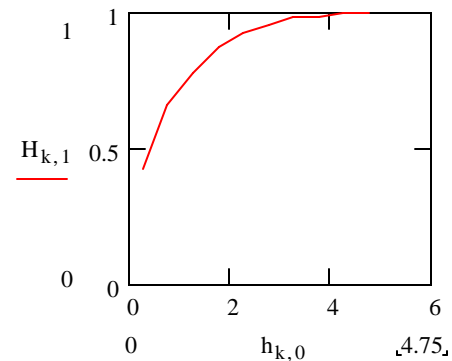
$h := \text{histogram}(M, X)$

$k := 0.. M - 1$

$$h_{k,1} := \frac{h_{k,1}}{N}$$



- Example 8.1: Generate 200 exponentially distributed random numbers, plot histogram
- Generating c.d.f. from histogram:



Verifying that Inverse Transform Technique Generates R.V.s with Correct Distribution

$$P(X_1 \leq x_0) = P(R_1 \leq F(x_0)) = F(x_0)$$

- X_i is generated from R_i , transformed by $F(\cdot)$
- R_i is uniformly distributed on $(0,1)$

Applying Inverse Transform Technique to Other Distributions

- Weibull

$$f(x) = \begin{cases} \frac{b}{a^b} x^{b-1} e^{-(x/a)^b} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-(x/a)^b} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$X = a [-\ln(R)]^{1/b}$$

- Triangular

$$f(x) = \begin{cases} x, & 0 \leq x \leq 1 \\ 2-x, & 1 < x \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{x^2}{2}, & 0 < x \leq 1 \\ 1 - \frac{(2-x)^2}{2}, & 1 < x \leq 2 \\ 1, & x > 2 \end{cases}$$

$$X = \begin{cases} \sqrt{2R}, & 0 \leq R \leq \frac{1}{2} \\ 2 - \sqrt{2(1-R)}, & \frac{1}{2} < R \leq 1 \end{cases}$$

Examples of Inverse Transform Technique

Weibull Distribution

N := 2000

i := 0.. N - 1

$\alpha := 1$

R := runif(N, 0, 1)

$\beta := 1.5$

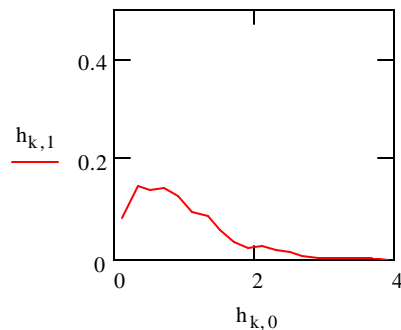
$$X := \alpha \cdot (-\ln(R))^{\frac{1}{\beta}}$$

M := 20

h := histogram(M, X)

k := 0.. M - 1

$$h_{k,1} := \frac{h_{k,1}}{N}$$



Triangular Distribution

N := 20000

i := 0.. N - 1

R := runif(N, 0, 1)

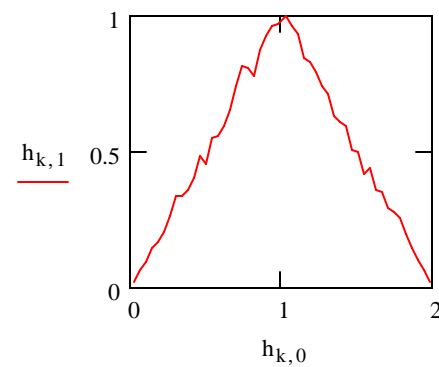
$$X_i := \text{if} \left[R_i \leq .5, \sqrt{2 \cdot R_i}, 2 - \sqrt{2 \cdot (1 - R_i)} \right]$$

M := 50

h := histogram(M, X)

k := 0.. M - 1

$$h_{k,1} := \frac{h_{k,1} \cdot \frac{M}{2}}{N}$$



Empirical Continuous Distributions

- If observed distribution is believed to be discrete, we can use lookup table method previously discussed. What if the distribution is known to be continuous?
 - Interpolate intermediate values: Example 8.2: 5 samples of data are available

i	Interval $x_{(i-1)} < x \leq x_{(i)}$	Probability $1/n$	Cumulative Probability, i/n	Slope, a_i
1	$0.0 < x \leq 0.80$	0.2	0.2	4.00
2	$0.80 < x \leq 1.24$	0.2	0.4	2.20
3	$1.24 < x \leq 1.45$	0.2	0.6	1.05
4	$1.45 < x \leq 1.83$	0.2	0.8	1.90
5	$1.83 < x \leq 2.76$	0.2	1.0	4.65

- Line segment slopes:

$$a_i = \frac{x_{(i)} - x_{(i-1)}}{i/n - (i-1)/n} = \frac{x_{(i)} - x_{(i-1)}}{1/n}$$

- Inverse c.d.f.:

$$X = x_{(i-1)} + a_i \left(R - \frac{(i-1)}{n} \right) \quad \text{when } (i-1)/n < R \leq i/n$$

Empirical Continuous Distributions

- Sometimes a large number of data samples are available to generate empirical distribution
 - not efficient, or always necessary to generate large number of interpolation segments
 - summarize available data into frequency distribution with smaller number of bins
 - fit continuous empirical c.d.f. to frequency distribution

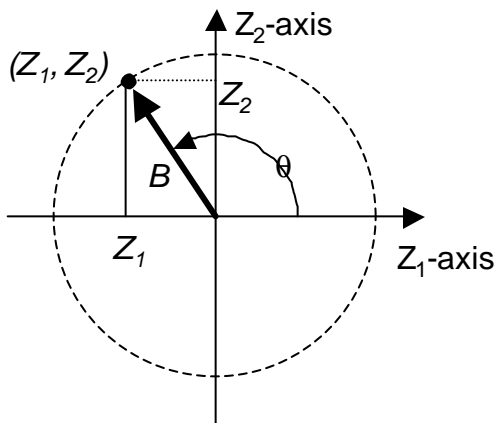
Direct Transform for Normal Distributions

- Normal c.d.f.:

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2p}} e^{-\frac{t^2}{2}} dt$$

- Inverse transform technique cannot be applied - inverse c.d.f. cannot be expressed in closed form.

Direct Transform for Normal Distributions



- Z_1 and Z_2 are normal R.V.s. In polar coordinates:

$$Z_1 = B \cos \mathbf{q}$$

$$Z_2 = B \sin \mathbf{q}$$

$$B^2 = Z_1^2 + Z_2^2$$

- B^2 has a chi-square distribution with 2 degrees of freedom - equivalent to an exponential distribution with mean 2. Using Inverse Transform technique:

$$B = \sqrt{-2 \ln R}$$

- \mathbf{q} is uniformly distributed on $(0, 2\mathbf{p})$ and is independent of B .
- Given independent, uniformly distributed R_1 and R_2 , generate normal Z_1, Z_2 :

$$Z_1 = \sqrt{-2 \ln R_1} \cos(2\mathbf{p} R_2)$$

$$Z_2 = \sqrt{-2 \ln R_1} \sin(2\mathbf{p} R_2)$$

Direct Transform for Normal Distributions

- Computation of $\sqrt{\ln(R)}$ is CPU intensive, but is reused for Z_1 and Z_2
- To obtain normal variates X_i with mean \mathbf{m} and variance \mathbf{s}^2 :

$$X_i = \mathbf{m} + \mathbf{s} Z_i$$

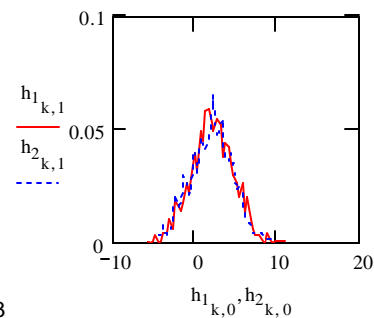
Normal Distribution using Direct Transform

```

N := 1000
i := 0..N - 1
R1 := runif(N, 0, 1)
R2 := runif(N, 0, 1)
Z1_i := sqrt(-2 * ln(R1)) * cos(2 * pi * R2_i)
Z2_i := sqrt(-2 * ln(R1)) * sin(2 * pi * R2_i)
X1 := mu + sigma * Z1
X2 := mu + sigma * Z2
M := 50
h1 := histogram(M, X1)
h2 := histogram(M, X2)
k := 0..M - 1

```

$$h_{1,k,1} := \frac{h_{1,k,1}}{N} \quad h_{2,k,1} := \frac{h_{2,k,1}}{N}$$



$\text{mean}(X_1) = 2.092$

$\text{mean}(X_2) = 2.015$

$\text{var}(X_1) = 6.584$

$\text{var}(X_2) = 6.491$

Convolution Method

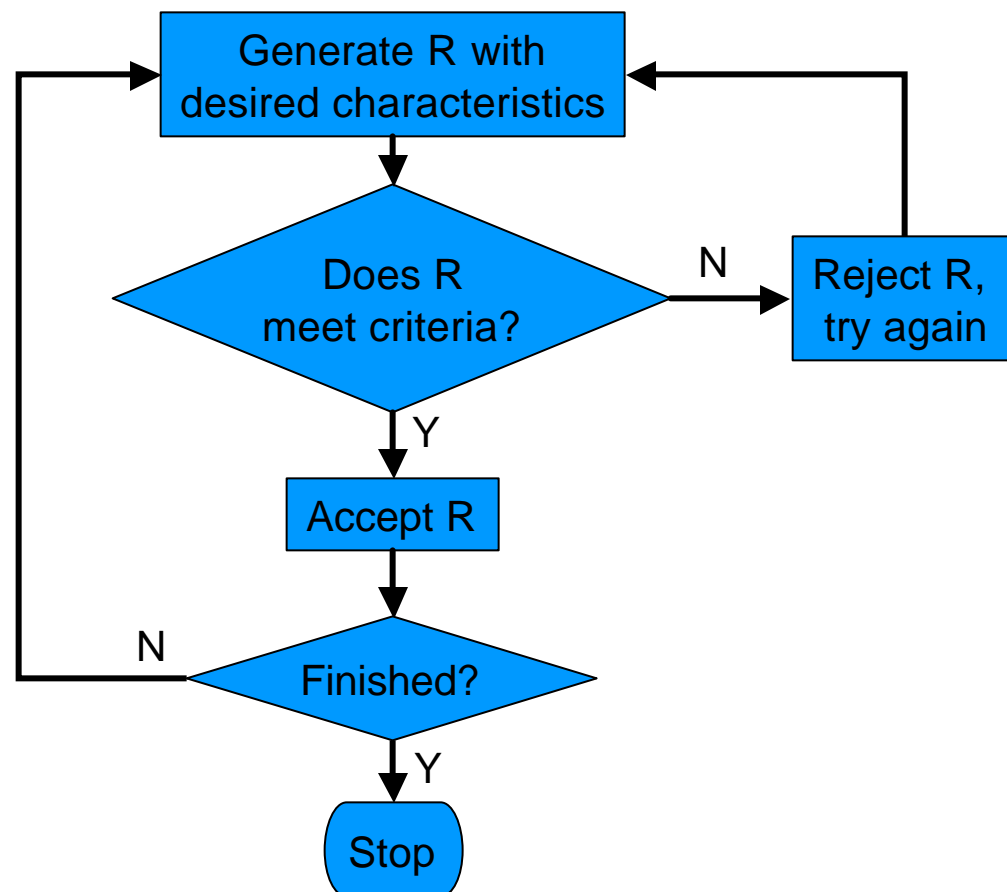
- Probability distribution of sum of independent R.V.'s is convolution of distributions of variables
- Erlang R.V. X with parameters (K, q) is sum of K independent exponential R.V.s X_i each with mean $1/Kq$

$$X = \sum_{i=1}^K X_i$$

- From Inverse Transform Technique, each X_i is generated with $1/I = 1/Kq$

$$\begin{aligned} X &= \sum_{i=1}^K -\frac{1}{Kq} \ln R_i \\ &= -\frac{1}{Kq} \ln \left(\prod_{i=1}^K R_i \right) \end{aligned}$$

Acceptance-Rejection Technique



- Use the Acceptance-Rejection Technique when other methods have no straightforward solution (e.g., no closed form solution)
- Efficiency depends on fraction of generated random numbers that are rejected

Using Acceptance-Rejection Technique for Poisson Distribution

- Poisson R.V., N , with mean a has a probability mass function

$$p(n) = P(N = n) = \frac{e^{-a} a^n}{n!}, \quad n = 0, 1, 2, \dots$$

- N is the number of arrivals from Poisson arrival process in one unit of time
- Interarrival times A_1, A_2, \dots of successive customers are exponentially distributed with rate a
- We know how to generate an exponential distribution from Inverse Transform Technique
- Relationship between the discrete Poisson process and continuous exponential distribution:

$$N = n$$

if and only if

$$A_1 + A_2 + \dots + A_n \leq 1 < A_1 + \dots + A_n + A_{n+1}$$

- Generate $n+1$ exponential interarrival times until some arrival occurs after $t=1$, then set $N=n$

Using Acceptance-Rejection Technique for Poisson Distribution

- Generate arrival times

$$A_1 + A_2 + \dots + A_n \leq 1 < A_1 + \dots + A_n + A_{n+1}$$

or
$$\sum_{i=1}^n -\frac{1}{\mathbf{a}} \ln R_i \leq 1 < \sum_{i=1}^{n+1} -\frac{1}{\mathbf{a}} \ln R_i$$

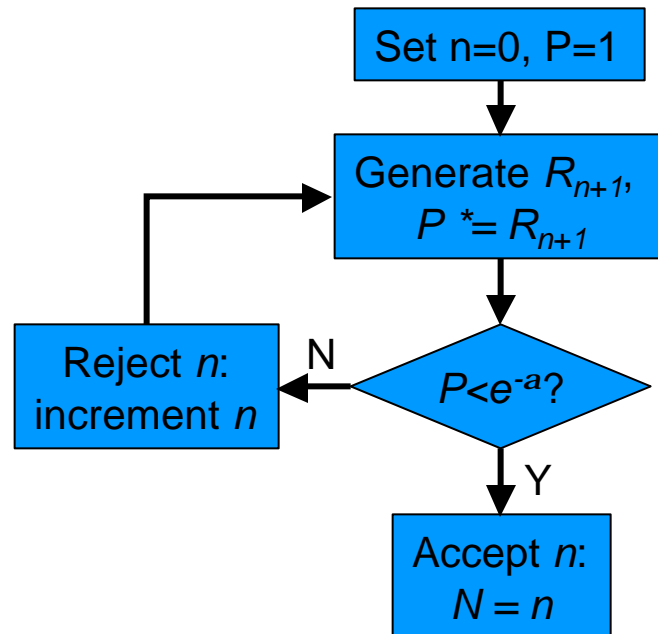
or
$$\sum_{i=1}^n \ln R_i \geq -\mathbf{a} > \sum_{i=1}^{n+1} \ln R_i$$

or
$$\ln \prod_{i=1}^n R_i \geq -\mathbf{a} > \ln \prod_{i=1}^{n+1} R_i$$

or
$$\prod_{i=1}^n R_i \geq e^{-\mathbf{a}} > \prod_{i=1}^{n+1} R_i$$

Using Acceptance-Rejection Technique for Poisson Distribution

- Given a :



Using Acceptance-Rejection Technique for Poisson Distribution

- Example 8.11: Bus with Poisson arrival process, arrival rate $a=4$ per hour. Generate number of buses arriving during a 1 hour period
- With high acceptance rate, this technique works well. As acceptance rate drops (e.g., for larger a), this technique becomes inefficient.
- An approximation for the Poisson process for $a > 15$ is:

$$Z = \sqrt{-2 \ln R_1} \cos(2\pi R_2)$$

$$N = \begin{cases} 0 & \text{if } a + \sqrt{a}Z - .5 < 0 \\ \text{ceil}(a + \sqrt{a}Z - .5) & \text{otherwise} \end{cases}$$

$$\alpha_{\text{bus}} := 4$$

$$e^{-\alpha_{\text{bus}}} = 0.018$$

```

N(α) :=
  n ← 0
  P ← 1
  m ← e-α
  while P ≥ m
    P ← P.rnd(1)
    n ← n + 1
  return n
    
```

How many arrivals per hour during a given 8 hour period?

$$i := 0..7$$

$$\text{Arrivals}_i := N(\alpha_{\text{bus}})$$

$$\text{Arrivals} = \begin{pmatrix} 3 \\ 4 \\ 7 \\ 3 \\ 4 \\ 2 \\ 4 \\ 3 \end{pmatrix}$$

$$\text{mean}(\text{Arrivals}) = 3.75$$

Homework

- Generate 1000 random numbers between 0 and 1, based on the linear congruential techniques covered last week. Use a modulus size of at least 1000. Find the number of up-runs and down-runs in the sequence. Using the run-test method discussed, see if the sequence generated appears independent with a 5% level of significance.
- Ch. 8 exercises 3, 17