CS 475/575 Slide Set 6

M. Overstreet
Spring 2005

1

Random Variates

- Problem: choose to model some component/behavior of simulated system as random -- but with a distribution that is "right."
- Usually never know for certain what "right" is.
- If have data from referent system, usually want to match that data.
- May want *empirical* or *theoretical* distribution.
- Often want to choose a theoretical distribution that best "fits" raw data (since lots of sim. languages make this easy).

2

Choosing "right" distributions

- May know something about system that suggests a particular distribution (so all that's needed is parameters), e.g., normal or exponential.

1. Known characteristic of data help:
   - Discrete or continuous
   - Range: bounded, nonnegative, infinite

3

Choosing right dist - (cont.)

2. Known characteristics of system:
   Arrivals often exponential
   Measure error often normal
   Interfailure times cannot be negative.
3. More formal: use standard "goodness-of-fit" test to match empirical to theoretical.
   (We will do with Arena)

4

choosing right dist. (cont.)

4. Less formal: pick something, then (if you have validation data) do sensitivity testing to see how much your choice effects results.
5. Simple heuristics (see table 4.2 in text, pg. 164):
   - If nothing is known, but range is bounded, use uniform.
   - If little is know, and range is bounded, try triangular if relative frequencies suggest this.
   - If high variance, bounded on left, try exponential

5

choosing right dist. (cont.)

Characteristic	Distribution
High variance Bounded on left Unbounded on right	Exponential
Symmetric or non-symmetric Bounded on both sides	Triangular
All Values equally likely Bounded on both sides	Uniform
Symmetric Unbounded on both sides Can be negative	Normal

6

theoretical distributions

- ⌘Thus far we've (mostly) focused on generating U(0,1) random variables.
- ⌘What about other distributions?
- ⌘Several techniques, but first need to review (or cover) some basic statistics.
- ⌘Basic idea is that we are sampling from a (potentially) infinite population.
  - (how many numbers are between 0 and 1?)
  - Assume we are always working with a sample
  - Some samples are "typical," some are not.

7

theoretical dists (cont.)

- ⌘ Might toss a coin 100 times and get, say more than 65% heads, but not often (this is "unusual" but certainly possible)
- ⌘ But we we do this often and keep score, we will sometimes get "too many heads"
  - (if we do things often enough, it would be unusual if the unusual never happened)
- ⌘ As n (the number of tosses -- or samples or observations) approaches infinity, then if we plotted the frequencies as a histogram (as relative percentages of total), we would generate an approximation of the

  Probability density function (PDF)

  for the distribution.

8

PDFs & CDFs

- ⌘Each distribution (normal, exponential, uniform, etc.) has its own characteristic PDF.
- ⌘Each has a total area of 1.
- ⌘If we integrate the PDF from -infinity to some point, we get the

  Cummulative Density Function (CDF)
- ⌘This is useful we we want to generate random data with the same distribution.

9

Example

- Suppose we are simulating activity at a local restaurant. People arrive in groups of various sizes. Want a simulation that produces similar behavior.
- Suppose we have gathered the data given in the next table.

10

restaurant arrival data

Arrivals Per Party	Frequency	Relative Frequency	Cummulative Frequency
1	30	0.10	0.10
2	110	0.37	0.47
3	45	0.15	0.62
4	71	0.24	0.86
5	12	0.04	0.90
6	13	0.04	0.94
7	7	0.02	0.96
8	12	0.04	1.00

11

sample code

```
x = uniform(0,1)
if 0 <= x < 0.10 return 1
if 0.1 <= x < 0.47 return 2
if 0.47 <= x < 0.62 return 3
if 0.62 <= x < 0.86 return 4
  ...
```

- Claim: “over the long run” this code will produce the empirical data exactly.
  - lemma: model behavior will be no better than the data
- Where is the PDF and CDF in the above?

12

poisson processes

- ⌘Widely used in simulation
- ⌘Assume:
  1. Arrivals occur one at a time.
  2. The number of arrivals between t and t+s depends only on s, not on t

     (This is called a *stationary process;* arrival frequencies do not change over time)
  3. The numbers of arrivals in nonoverlapping time intervals are independent random variables

     (This is the *memoryless* property; lots of arrivals in one interval are not made up for in the next.)

13

poisson (cont.)

- ⌘If so, then the process is said to be Poisson.
- ⌘Let N(t) be the expected number of arrivals in time duration t. Then we could show:

$$P[N(t) = n] = \frac{e^{-\lambda t}\lambda t}{n!}$$

14

reading:

- ⌘Arena text, chapter 4.
- ⌘Reference: Leemis, chapter 6

15