Last Lecture - Assorted Topics

Assorted Items

1) Grading Status – ISCA deadline; Disk Crash

2) Presentation Signup

3) Presentation Judging Form

4) My News

5) Sampling Technique Comparison , Yi HPCA 2005

6) Markov Sequence

5) Confidence Interval calculation

EE 382M

Performance Evaluation and Benchmarking

Presentation Judging

Name of Presenter:

Judge Name: (If you come late you cannot judge; you lose participation points)

Technical content - preparation and contents – background, objectives - results are not the focus, but motivation, methodology, prelim results, inferences (2)

1 2 3 4 5 6 7 8____________________

Slide organization - visual quality of slides - good flow - picking the right words - do not overcrowd – use big font - good illustrations- (1.5)

1 2 3 4 5 6 ____________________

presentation skills - voice/eye contact etc - (1)

0 1 2 3 4 ____________________

time mngmt – too short or long (0.5)

0 1 2 ____________________

Total ____________________(out of 20)

TOP500 will replace LINPACK with Conjugate gradient

1) TOP500 list – announced every 6 months since 1993.
2) New list expected this week at Supercomputing 2014
3) Oak Ridge National Lab and Univ Tennessee
4) The top machine in the latest listing (June 2014) was the Tianhe-2 (MilkyWay-2) at the National Super Computer Center in Guangzhou, China.
5) 3,120,000 cores to achieve 33,862,700 gigaFLOPS (33,862.7 teraFLOPS, or almost 34 petaFLOPS).
6) Number one in June 1993, was a 1,024-core machine at the Los Alamos National Laboratory that achieved 59.7 gigaFLOPS, i.e. six orders of magnitude in 21 years.

CONJUGATE GRADIENT

an iterative method of solving certain linear equations

emphasize data access instead of calculation

conjugate gradients involve moving data in large matrices, rather than performing dense calculations.

combines both inexact and exact calculations – approximate computing - the new conjugate gradients benchmark gives opportunity to show benefits of using full precision selectively

energy required to reach a solution with a combination of exact and inexact computation is reduced by IBM by almost 300.

Reference: http://www.cio.com/article/2685219/hardware/beyond-flops-the-co-evolving-world-of-computer-benchmarking.html

Comparison of 6 Prevailing Simulation Techniques

1) SimPoint [18]

2) Reduced input sets (MinneSPEC and SPEC test/train)

3) Simulating the first Z million instructions only

4) Fast-forwarding X million instructions and then simulating the next Z million

5) Fast-forwarding X million, warming-up for the next Y million, then simulating the next Z million instructions

6) SMARTS, a rigorous, statistical sampling technique

Periodic Sampling as in SMARTS

Simulates selected portions of the dynamic instruction execution at fixed

Intervals

The sampling frequency and the length of each sample are used to control the overall simulation time

SMARTS (Sampling Microarchitectural Simulation) [Wunderlich, ISCA 2003]

Random Sampling

SMARTS uses statistical sampling theory to estimate the CPI error of the

sampled simulation versus the reference simulation.

If the estimated error is higher than the user-specified confidence

interval, then SMARTS recommends a higher sampling frequency.

SMARTS also uses “functional warming” to maintain branch predictor

and cache state.

SIMPOINT (Representative Sampling) vs SMARTS

Simpoint uses fewer chunks than SMARTS

SMARTS uses large numbers of small chunks

Simpoint uses few chunks of large chunks

SMARTS uses statistical sampling and provides error bounds

SIMPOINT uses clustering

SIMPOINT does not provide error bounds

SIMPOINT analyzes the data to determine the chosen samples

SMARTS uses random samples

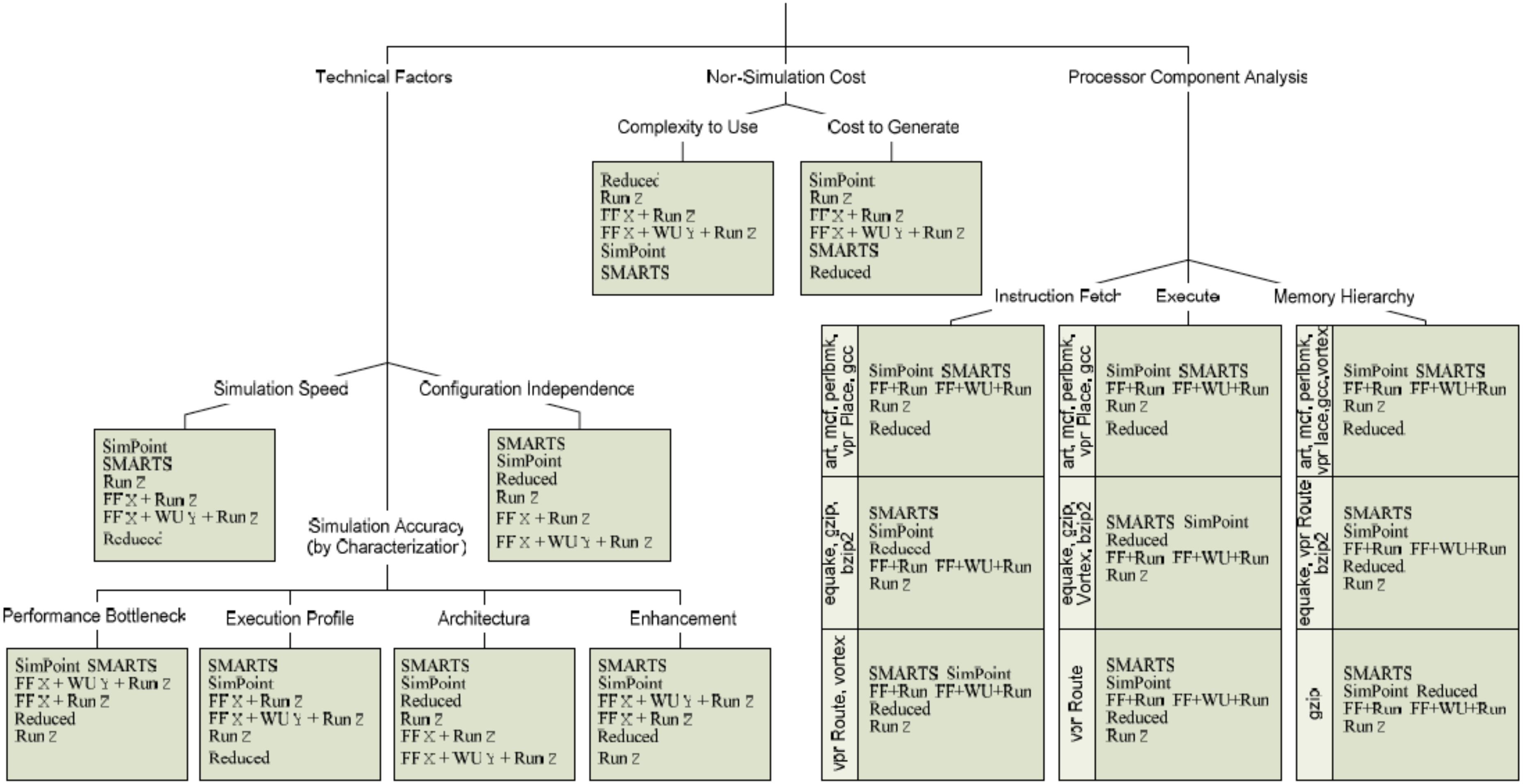
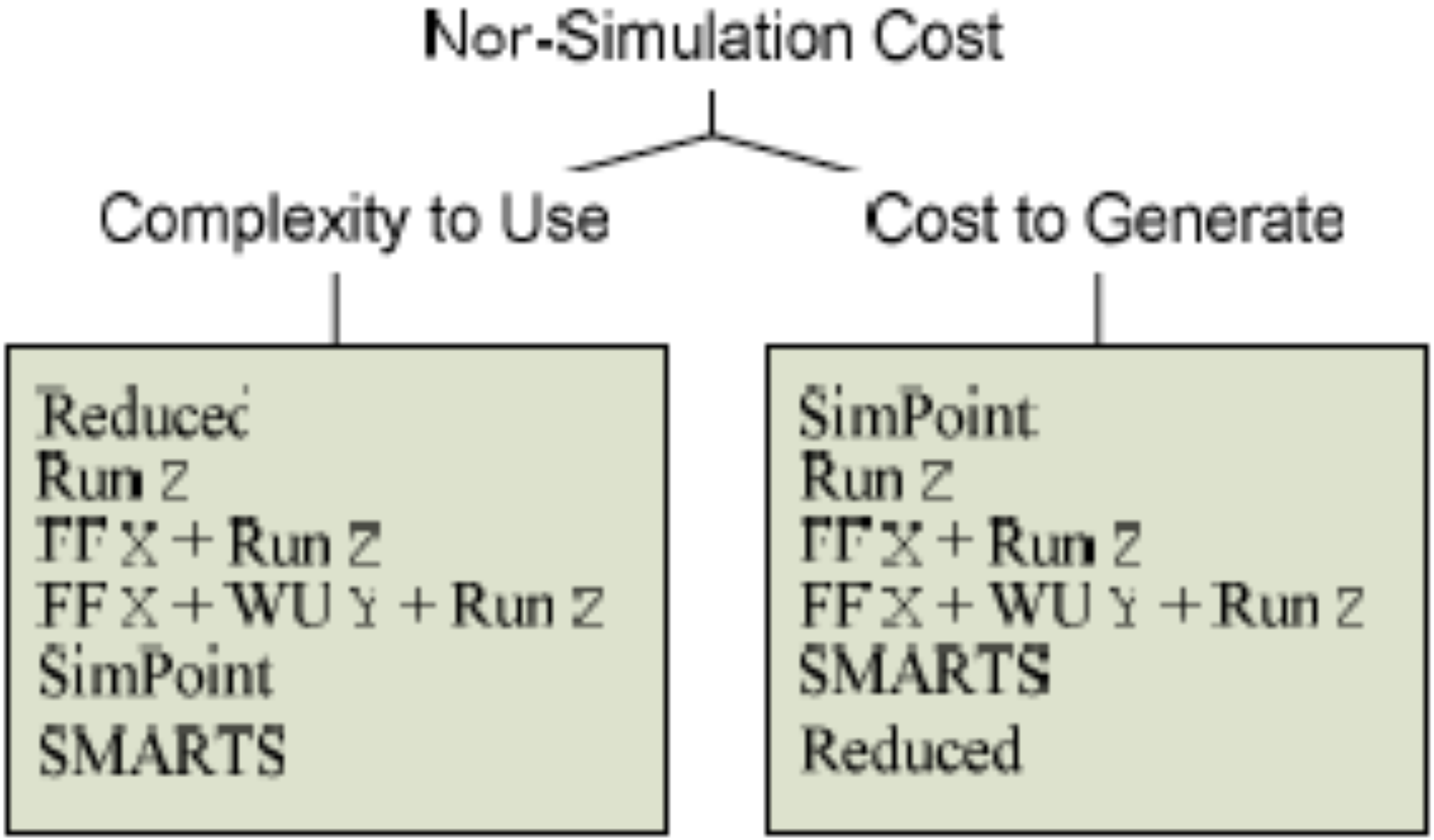
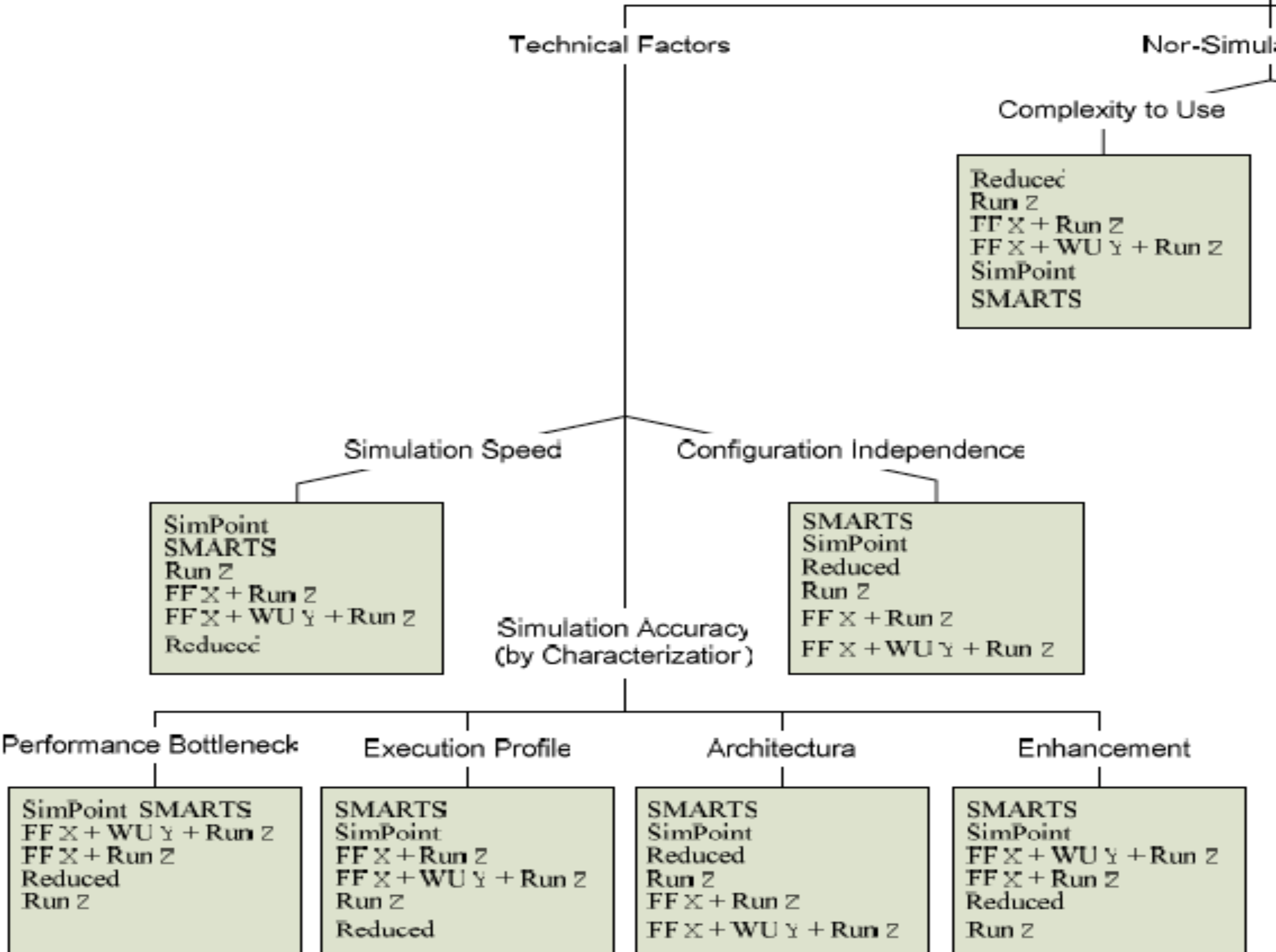
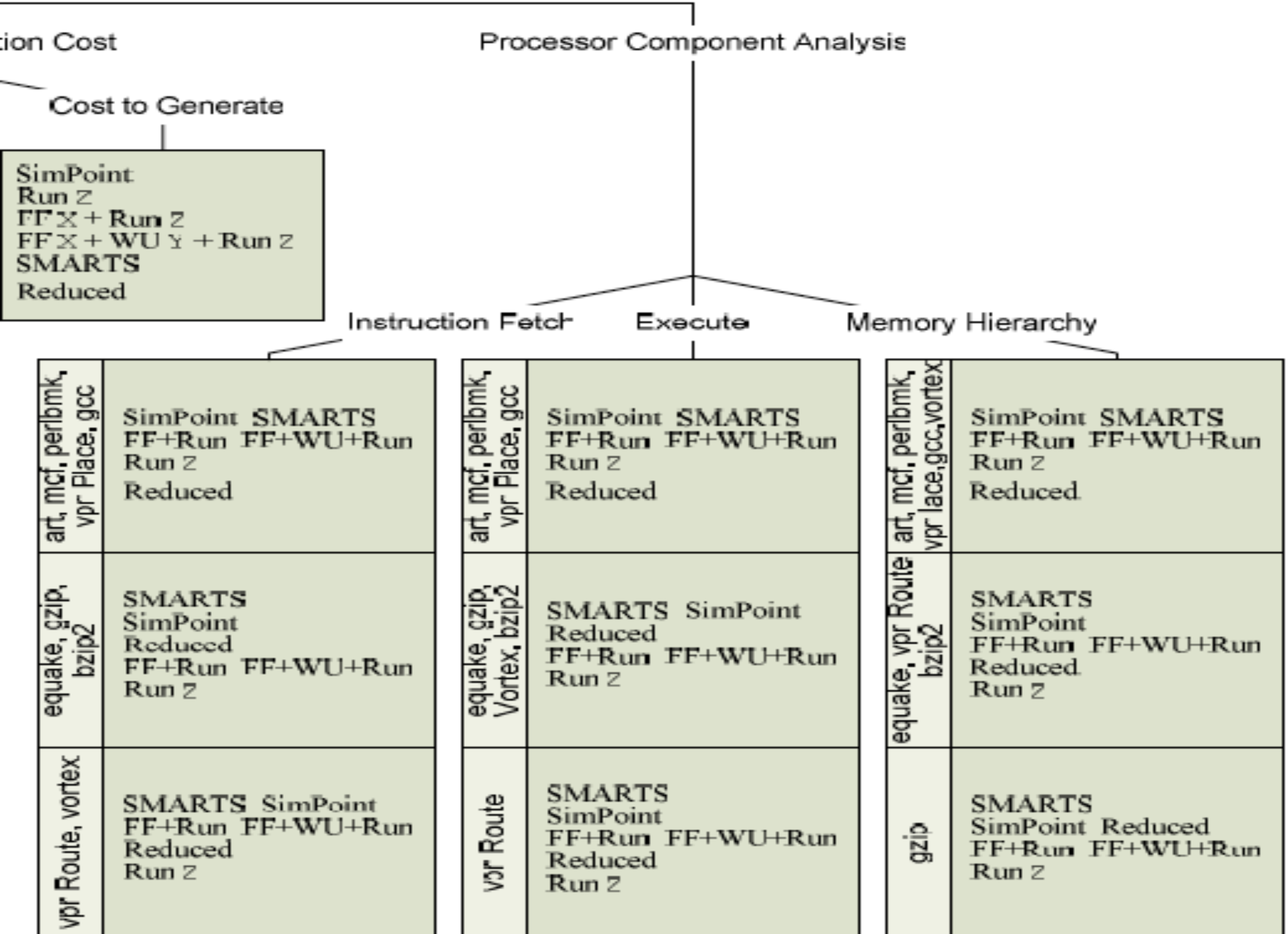


Figure 6. Decision Tree for the Selection of a Simulation Technique


Nor-Simulation Cost
Complexity to Use
Cost to Generate
Reduced
Run Z
FF X + Run Z
FF X + WU Y + Run Z
SimPoint
SMARTS
SimPoint
Run Z
FF X + Run Z
FF X + WU Y + Run Z
SMARTS
Reduced


Technical Factors
Nor-Simul
Complexity to Use
Reducec
Run Z
FF X + Run Z
FF X + WU Y + Run Z
SimPoint
SMARTS
Simulation Speed
Configuration Independence
SimPoint
SMARTS
Run Z
FF X + Run Z
FF X + WU Y + Run Z
Reducec
SMARTS
SimPoint
Reduced
Run Z
FF X + Run Z
FF X + WU Y + Run Z
Simulation Accuracy
(by Characterizatior)
Performance Bottleneck
Execution Profile
Architectura
Enhancement
SimPoint SMARTS
FF X + WU Y + Run Z
FF X + Run Z
Reduced
Run Z
SMARTS
SimPoint
FF X + Run Z
FF X + WU Y + Run Z
Run Z
Reduced
SMARTS
SimPoint
Reduced
Run Z
FF X + Run Z
FF X + WU Y + Run Z
SMARTS
SimPoint
FF X + WU Y + Run Z
FF X + Run Z
Reduced
Run Z


tion Cost
Processor Component Analysis
Cost to Generate
SimPoint
Run Z
FF X + Run Z
FF X + WU Y + Run Z
SMARTS
Reduced
Instruction Fetch
Execute
Memory Hierarchy
art, mcf, perlbmk, vpr Place, gcc
SimPoint SMARTS
FF+Run FF+WU+Run
Run Z
Reduced
equake, gzip, bzip2
SMARTS
SimPoint
Reduced
FF+Run FF+WU+Run
Run Z
vpr Route, vortex
SMARTS SimPoint
FF+Run FF+WU+Run
Reduced
Run Z
art, mcf, perlbmk, vpr Place, gcc
SimPoint SMARTS
FF+Run FF+WU+Run
Run Z
Reduced
equake, gzip, Vortex, bzip2
SMARTS SimPoint
Reduced
FF+Run FF+WU+Run
Run Z
vpr Route
SMARTS
SimPoint
FF+Run FF+WU+Run
Reduced
Run Z
art, mcf, perlbmk, vpr lace,gcc,vortex
SimPoint SMARTS
FF+Run FF+WU+Run
Run Z
Reduced
equake, vpr Route bzip2
SMARTS
SimPoint
FF+Run FF+WU+Run
Reduced
Run Z
gzip
SMARTS
SimPoint Reduced
FF+Run FF+WU+Run
Run Z

What is a Markov Process?

A Markov process can be thought of as a 'memoryless' process.

A process satisfies the Markov property if one can make predictions for the future of the process based solely on its present state just as well as one could knowing the process's full history. i.e., conditional on the present state of the system, its future and past are independent

i.e. probability of being in a state depends only on the immediately previous state and not the entire history

EXAMPLE MARKOV CHAIN

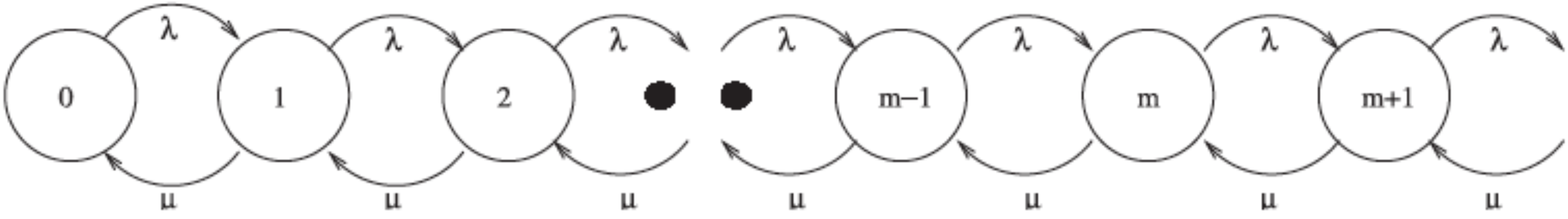


Figure 10.2 State transition diagram of the *M/M/1* system.

What are *statistics*?

- "**A branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.**"

  Merriam-Webster

→ **We are most interested in analysis and interpretation here.**

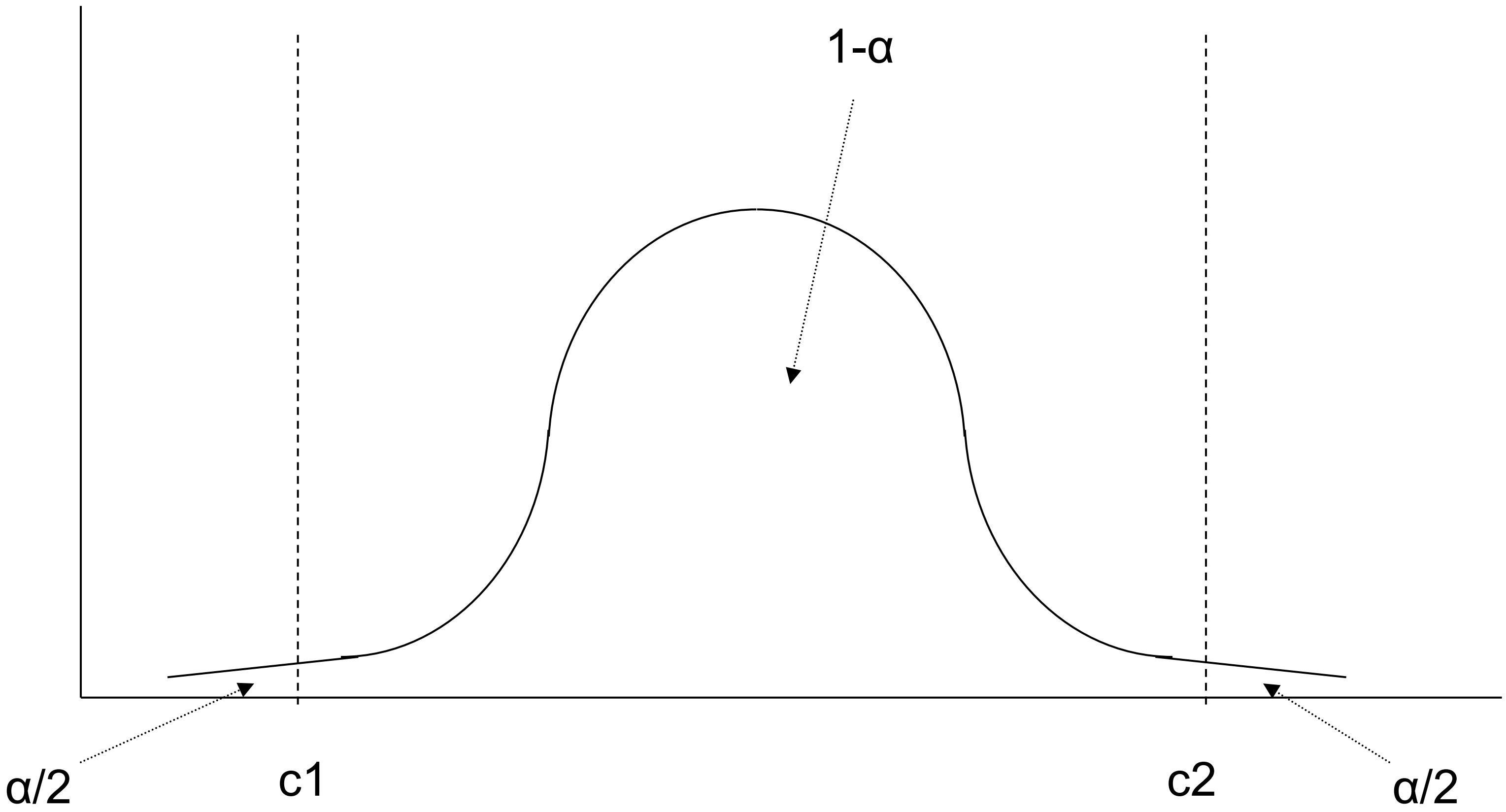
- "**Lies, damn lies, and statistics!**"

Goals

- **Provide intuitive conceptual background for some standard statistical tools.**
  - Draw meaningful conclusions in presence of noisy measurements.
  - Allow you to correctly and intelligently apply techniques in new situations.

→ Don't simply plug and crank from a formula.

Confidence Interval for the Mean



Normalize *x*

$$z = \frac{\bar{x} - x}{s / \sqrt{n}}$$

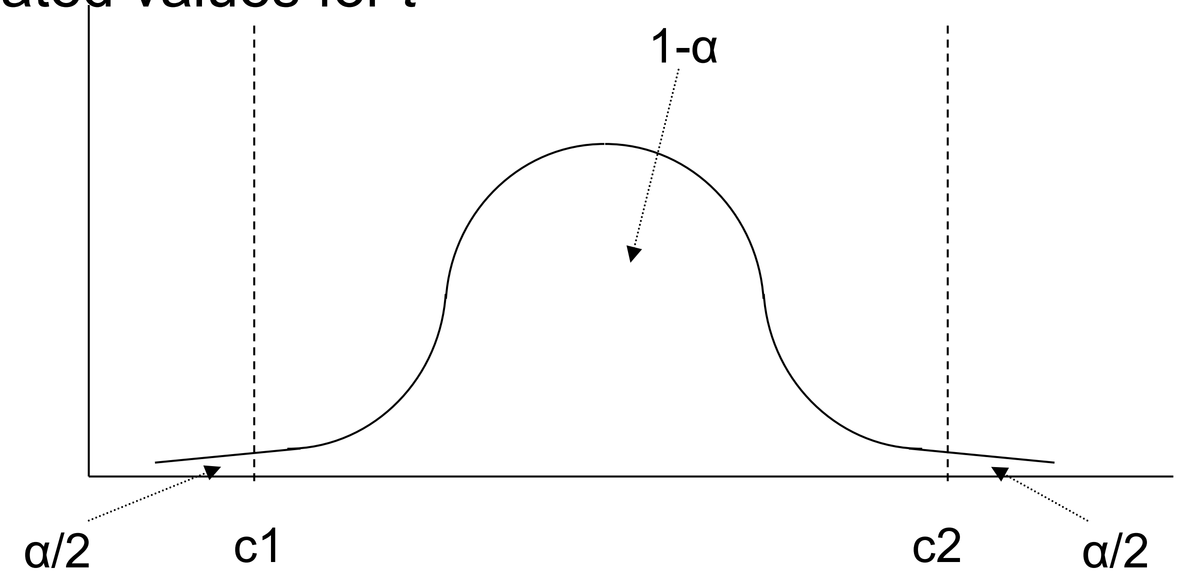
$$n = \text{number of measurements}$$

$$\bar{x} = \text{mean} = \sum_{i=1}^{n} x_i$$

$$s = \text{standard deviation} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} (x_i - \bar{x})^2}{n - 1}$$

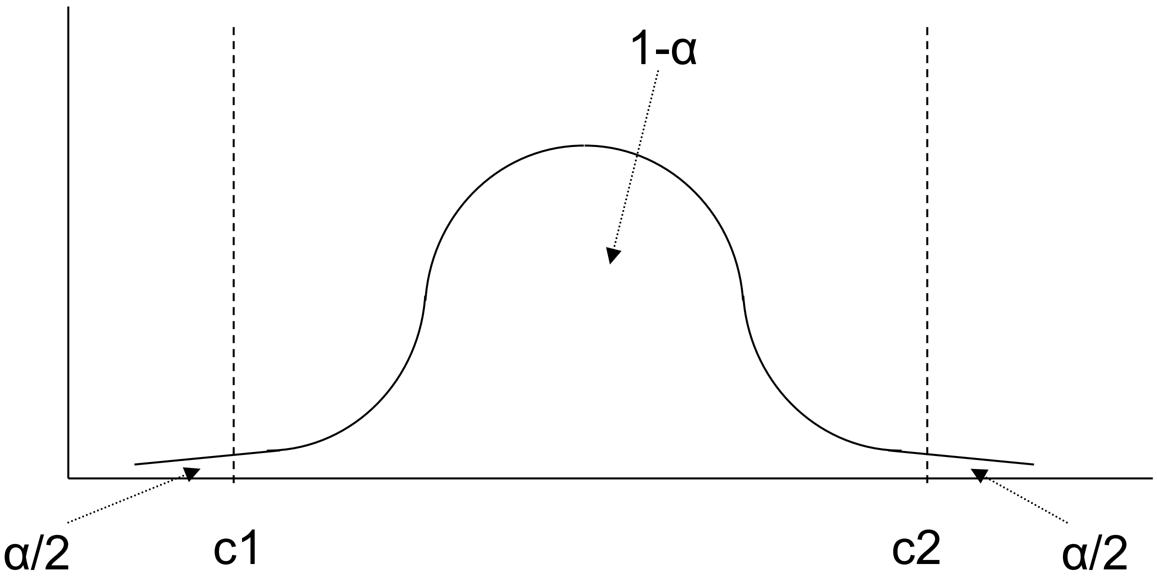
Confidence Interval for the Mean

- **Normalized *z* follows a Student's *t* distribution**
  - (*n*-1) degrees of freedom
  - Area left of $c_2 = 1 - \alpha/2$
  - Tabulated values for *t*



Confidence Interval for the Mean

- **As n → ∞, normalized distribution becomes Gaussian (normal)**



Confidence Interval for the Mean

$$c_1 = \bar{x} - t_{1-\alpha/2;n-1} \frac{s}{\sqrt{n}}$$

$$c_2 = \bar{x} + t_{1-\alpha/2;n-1} \frac{s}{\sqrt{n}}$$

Then,

$$\Pr(c_1 \le x \le c_2) = 1 - \alpha$$

An Example

Experiment	Measured value
1	8.0 s
2	7.0 s
3	5.0 s
4	9.0 s
5	9.5 s
6	11.3 s
7	5.2 s
8	8.5 s

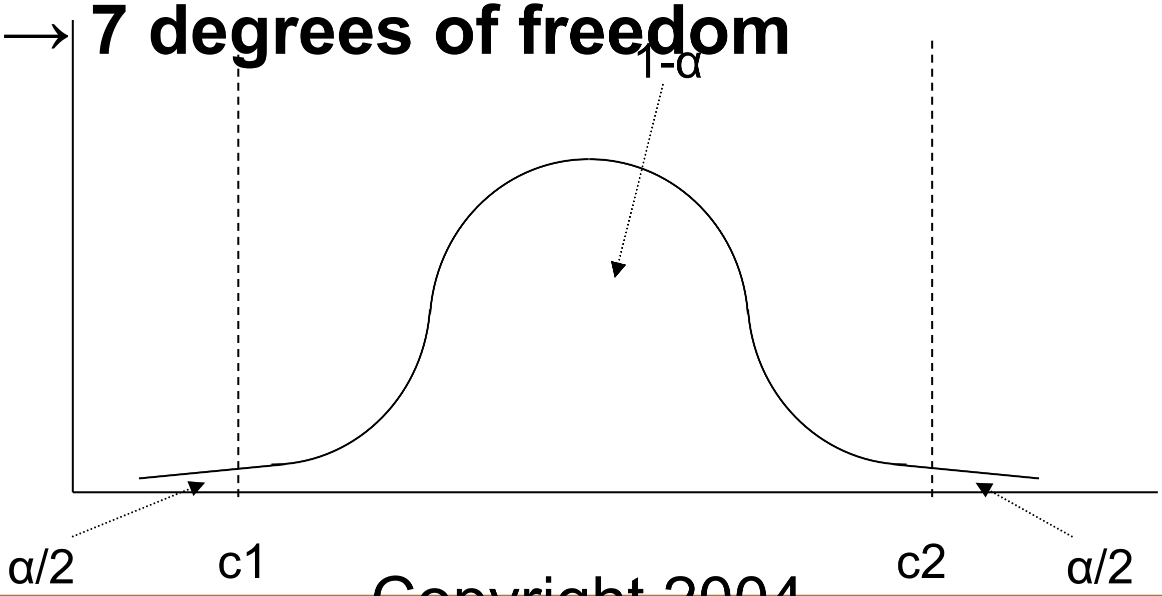
An Example (cont.)

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 7.94$$

$$s = \text{sample standard deviation} = 2.14$$

An Example (cont.)

- **90% CI → 90% chance actual value in interval**
- **90% CI → α = 0.10**
  - 1 - α /2 = 0.95
- ***n* = 8 → 7 degrees of freedom**



90% Confidence Interval

$$a = 1 - \alpha / 2 = 1 - 0.10 / 2 = 0.95$$

$$t_{a;n-1} = t_{0.95;7} = 1.895$$

$$c_1 = 7.94 - \frac{1.895(2.14)}{\sqrt{8}} = 6.5$$

$$c_2 = 7.94 + \frac{1.895(2.14)}{\sqrt{8}} = 9.4$$

	a		
n	0.90	0.95	0.975
…	…	…	…
5	1.476	2.015	2.571
6	1.440	1.943	2.447
7	1.415	1.895	2.365
…	…	…	…
∞	1.282	1.645	1.960

95% Confidence Interval

$$a = 1 - \alpha / 2 = 1 - 0.10 / 2 = 0.975$$

$$t_{a;n-1} = t_{0.975;7} = 2.365$$

$$c_1 = 7.94 - \frac{2.365(2.14)}{\sqrt{8}} = 6.1$$

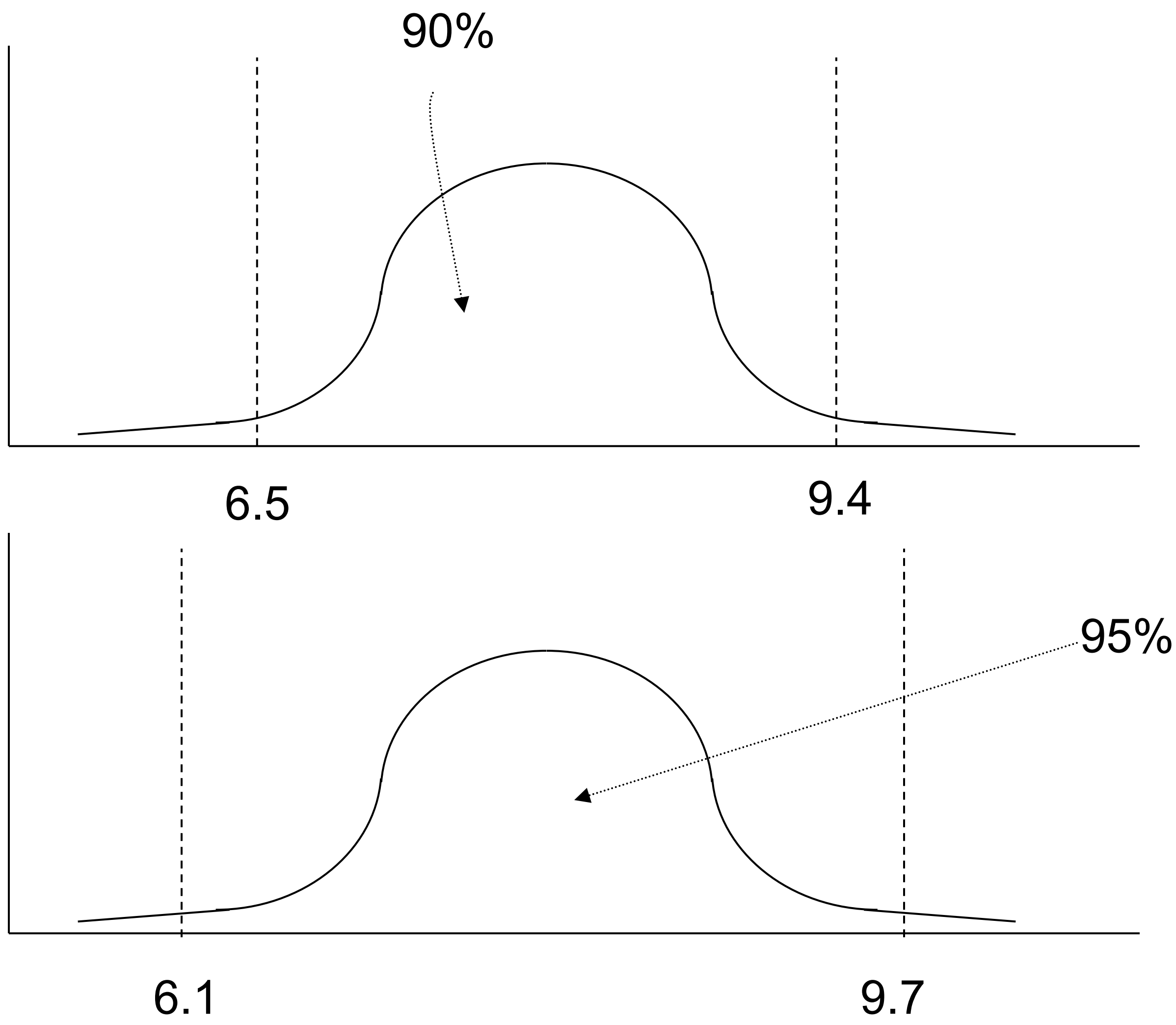
$$c_2 = 7.94 + \frac{2.365(2.14)}{\sqrt{8}} = 9.7$$

	a		
n	0.90	0.95	0.975
…	…	…	…
5	1.476	2.015	2.571
6	1.440	1.943	2.447
7	1.415	1.895	2.365
…	…	…	…
∞	1.282	1.645	1.960

What does it mean?

- **90% CI = [6.5, 9.4]**
  - 90% chance real value is between 6.5, 9.4
- **95% CI = [6.1, 9.7]**
  - 95% chance real value is between 6.1, 9.7
- **Why is interval wider when we are more confident?**

Higher Confidence → Wider Interval?



Coefficient of Variation (COV)

$$COV = \frac{s}{\bar{x}}$$

- **Dimensionless**
- **Compares relative size of variation to mean value**

Quantifying Variability

- **Variance**
- **Sample Variance**
- **When actual mean is known, divide by n**
- **When actual mean is unknown, divide by n-1 (instead of n) (i.e. estimating the population's variance)**

DEFINITION 2.3: Define $f^T(x)$, the *interreference temporal density function*, $f^T(x)$,, to be the probability of there being x unique references between successive references to the same item,

$$f^T(x) = \sum_t P\Big[u(w(t)) = x\Big].$$

2000,2004,2008,2012,(2004,2008,2012^9),

ft(inf)=1/31; ft(1)=0 ft(2)=26/31;

f(infinity)=4/31

2000,2000,2000,2000,2000

ft(0)=1