UCF Stands For Opportunity

*CDA6530: Performance Models of Computers and Networks*

# *Chapter 9:Statistical Analysis of Simulated Data and Confidence Interval*

SCHOOL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCE

# Sample Mean

- r.v. X: E[X]=$\theta$, Var[X]=$\sigma^2$
- Q: how to use simulation to derive?
  - Simulate X repeatedly
- $X_1, \cdots, X_n$ are i.i.d., $=_{\text{statistic}}$ X
- Sample mean: $\bar{X} \equiv \sum_{i=1}^{n} \frac{X_i}{n}$

$$E[\bar{X}] = \theta \qquad Var(\bar{X}) = \frac{\sigma^2}{n}$$

# *Sample Variance*

- $\sigma^2$ unknown in simulation
  - Hard to use $Var(\bar{X}) = \frac{\sigma^2}{n}$ to measure simulation variance
  - Thus we need to estimate $\sigma^2$
- Sample variance $S^2$:

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

  - n-1 instead of n is to provide unbiased estimator $E[S^2] = \sigma^2$

# Estimate Error

- Sample mean $\bar{X}$ is a good estimator of $\theta$, but has an error
  - How confidence we are sure that the sample mean is within an acceptable error?
- From central limit theorm:

$$\sqrt{n}\frac{(\bar{X}-\theta)}{\sigma} \sim N(0,1)$$

- It means that:

$$Z \equiv \sqrt{n}\frac{(\bar{X}-\theta)}{S} \sim N(0,1)$$

# *Confidence Interval*

- R.v. Z$\sim$ N(0,1), for 0<$\alpha$<1, define:
  - P(Z>$z_\alpha$) = $\alpha$
  - From normal lookup table:
    - $z_{0.025}$ = 1.96 for $\alpha$ = 0.025
  - P(- 1.96 <Z< 1.96) =1-2$\alpha$ = 0.95

$$P(\bar{X} - z_\alpha \frac{S}{\sqrt{n}} < \theta < \bar{X} + z_\alpha \frac{S}{\sqrt{n}}) \approx 1 - 2\alpha$$

$$P(\bar{X} - 1.96 \frac{S}{\sqrt{n}} < \theta < \bar{X} + 1.96 \frac{S}{\sqrt{n}}) \approx 0.95$$

**95% confidence interval ($\alpha$ = 0.025) of an estimate is:**

$$(\bar{X} \pm 1.96 S/\sqrt{n})$$

# *When to stop a simulation?*

- Repeatedly generate data (sample) until 100(1-2$\alpha$) percent confidence interval estimate of $\theta$ is less than I
  - Generate at least 100 data values.
  - Continue generate, until you generated k values such that $2z_\alpha S/\sqrt{k} < I$
  - The 100(1-2$\alpha$) percent confidence interval of estimate is

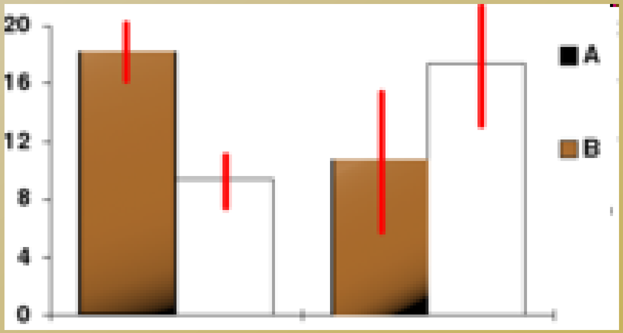
$$(\bar{X} - z_\alpha S/\sqrt{k}, \bar{X} + z_\alpha S/\sqrt{k})$$

# *Fix no. of simulation runs*

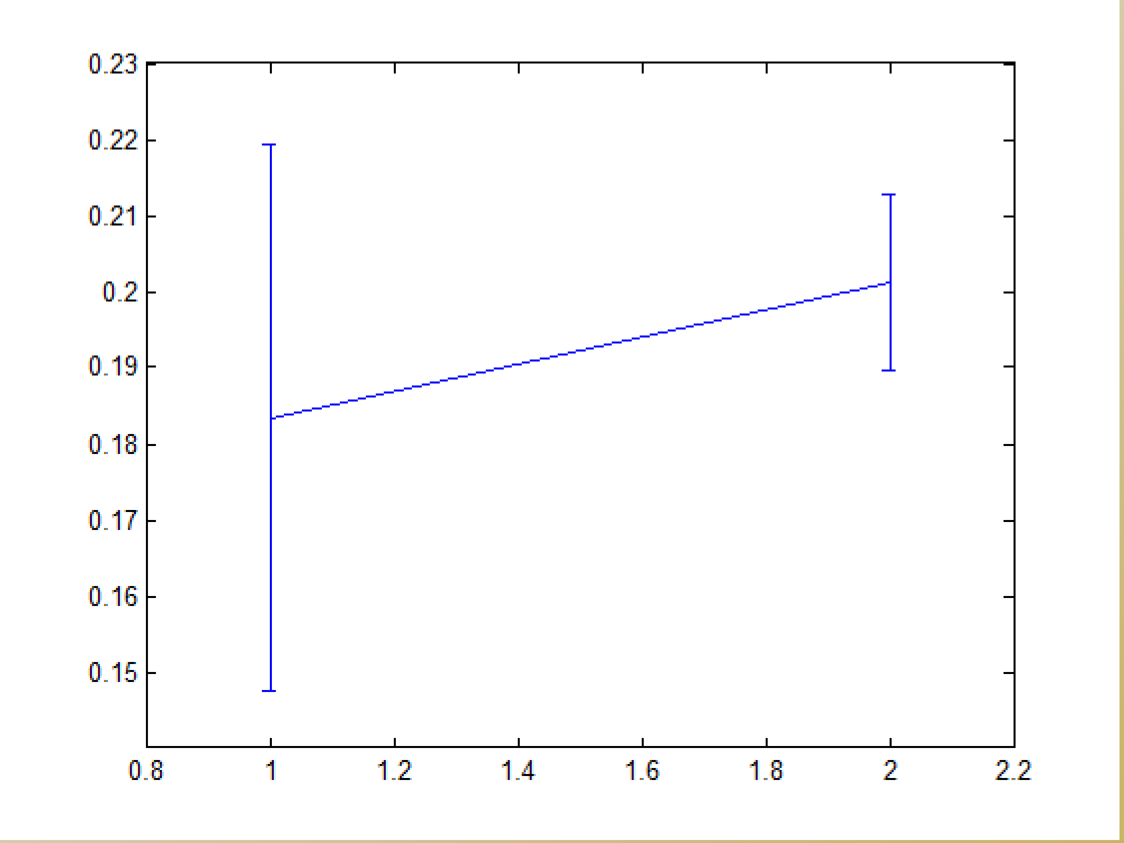
- Suppose we only simulate 100 times
  - k=100
- What is the 95% ($\alpha$=0.025)confidence interval?

$$(\bar{X} - z_\alpha S/\sqrt{k}, \bar{X} + z_\alpha S/\sqrt{k})$$

$$(\bar{X} - 1.96S/\sqrt{k}, \bar{X} + 1.96S/\sqrt{k})$$



# Example: Generating Expo. Distribution




**Compare 100 samples and 1000 samples confidence intervals**