COMP9334
Capacity Planning for Computer Systems and Networks

Week 6A: Discrete event simulation (2). Independent replications. Confidence interval.

Weeks 4B_2 and 5A

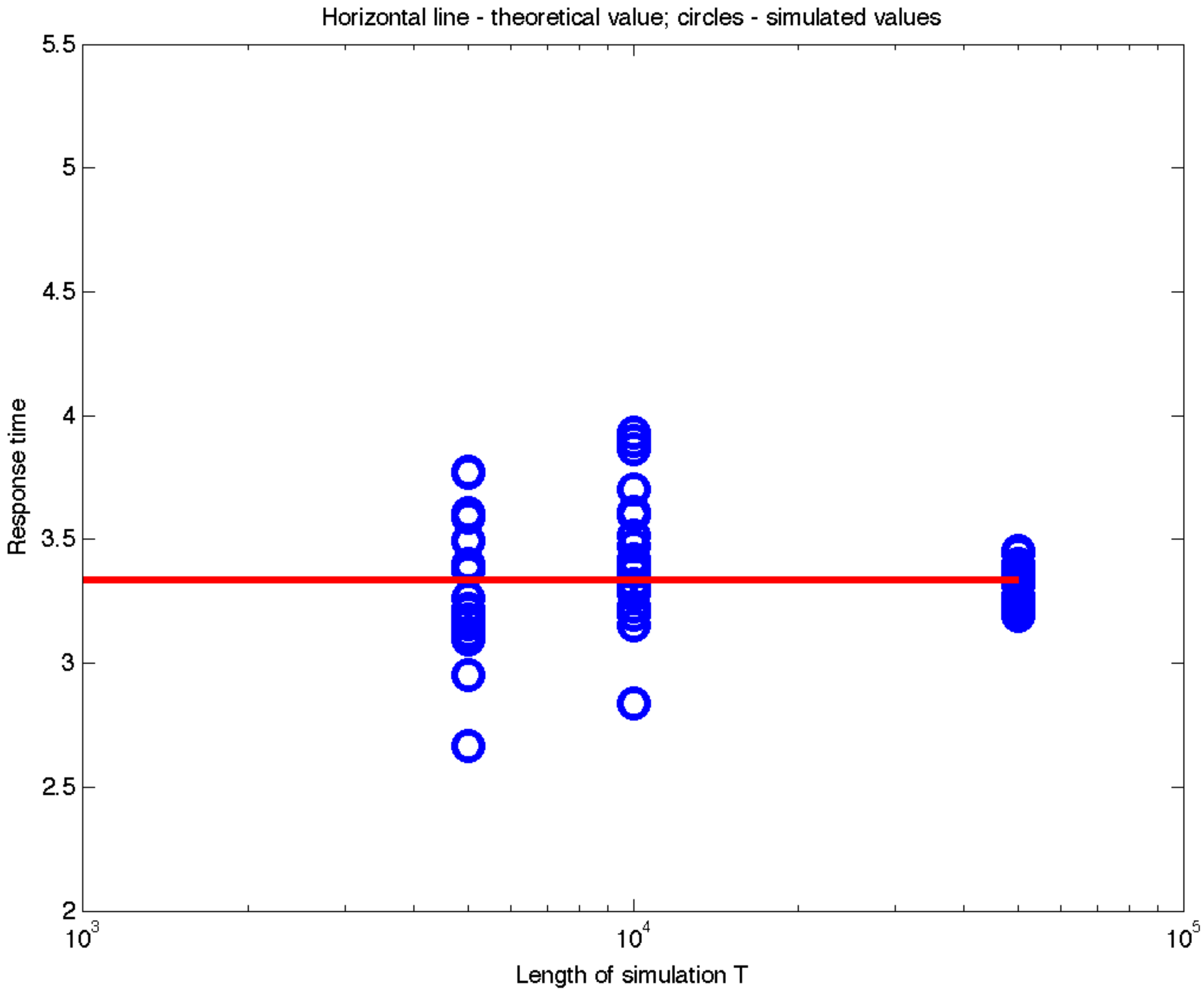
- Two topics
  - How to general random numbers of a given probability distribution by using inverse transform method
  - How to perform discrete event simulation of queues

- You should be able to simulate
  - Many types of queues
    - Single-server or multi-server
    - Different queueing disciplines
  - Many inter-arrival time and service time distributions

- However, there are a number of problems …

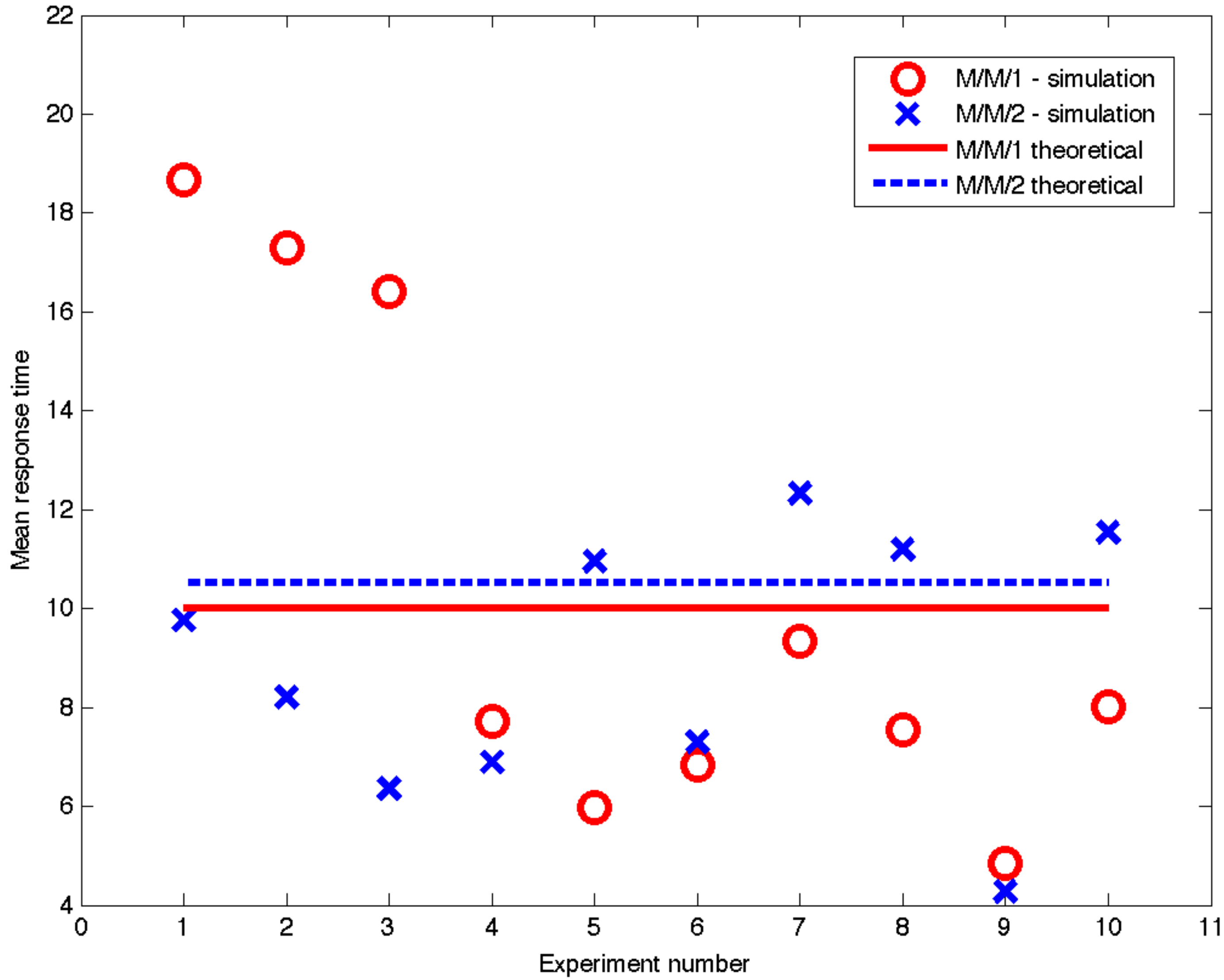
Problem: data interpretation, simulation length

- Figure from
- Week 5A's revision problem #1



Problem: How do we compare 2 alternative choices?

- Week 5A's Revision Problem, #2
- From Queueing theory, we expect the M/M/1 system to be better but simulation doesn't seem to suggest that?



Analysis of simulation results

- A very important topic but it is very often ignored
- Simulation is *not* simply about
  - Writing a computer program
  - Verifying the correctness of the program
  - Running the simulation once and present the results
- Verifying the correctness of the simulation program is important
- It is equally important to do **sound statistical analysis** on the simulation results obtained

This lecture

- Analysis of simulation results
- How to choose simulation parameters?
  - How long should I simulate for?
  - How many times should I repeat the simulation?
- Confidence interval

Analysis of simulation data

- There are many statistical methods to analyse data depending on the situation

- We will focus on analysing *steady state mean value* only

- For example, we are interested to find the *steady state mean response time* of a queue

- Recall that we talked about
  - Transient and steady state behaviour of queue in Week 2B
  - Steady state of Markov chain in Week 3B

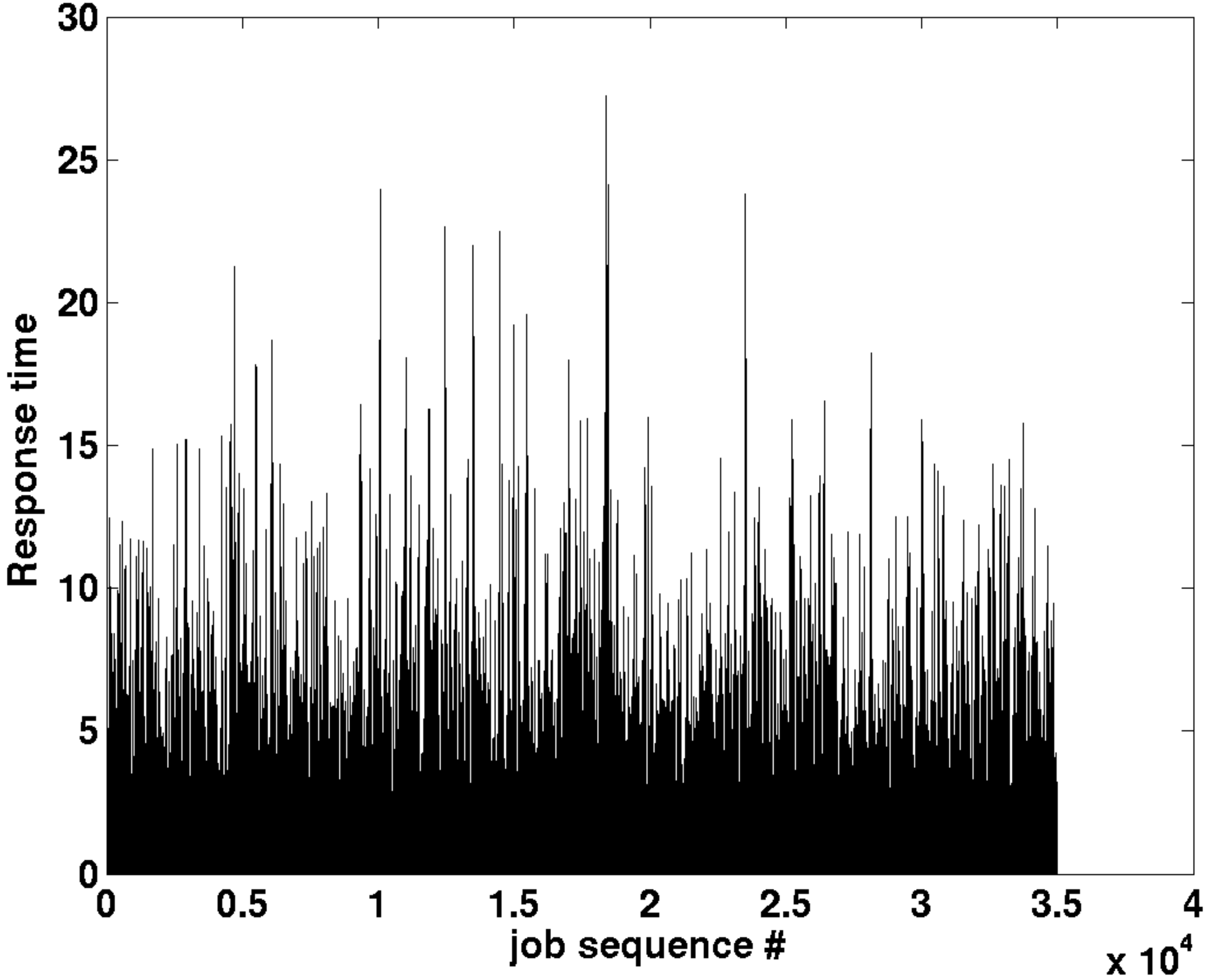
What is steady state? (1)

- Let us simulate an M/M/1 queue with
  - Arrival rate $\lambda = 0.7$
  - Service rate $\mu = 1$
  - Simulation ends when master clock is 50000s
- In this simulation, we record the response time for each job
  - Let X(k) = Response time of k^{th} job
  - The next page shows X(k) changes continuously
- Let N denote the number of jobs in the simulation
  - N = 35000 for our simulation
- In Week 5A, we computed the mean response time using

$$\frac{X(1)+X(2)+\ldots+X(N)}{N}$$

Response time continuously changes over time

- This graph shows response time of X(k) of the *k*-th job where k = 1 to 35000
- Note response time continuously varies
- Response time *does not* settle to a constant value
- But *mean* response time *does* settle



What is steady state? (2)

- Let us instead compute the running mean M(k) where

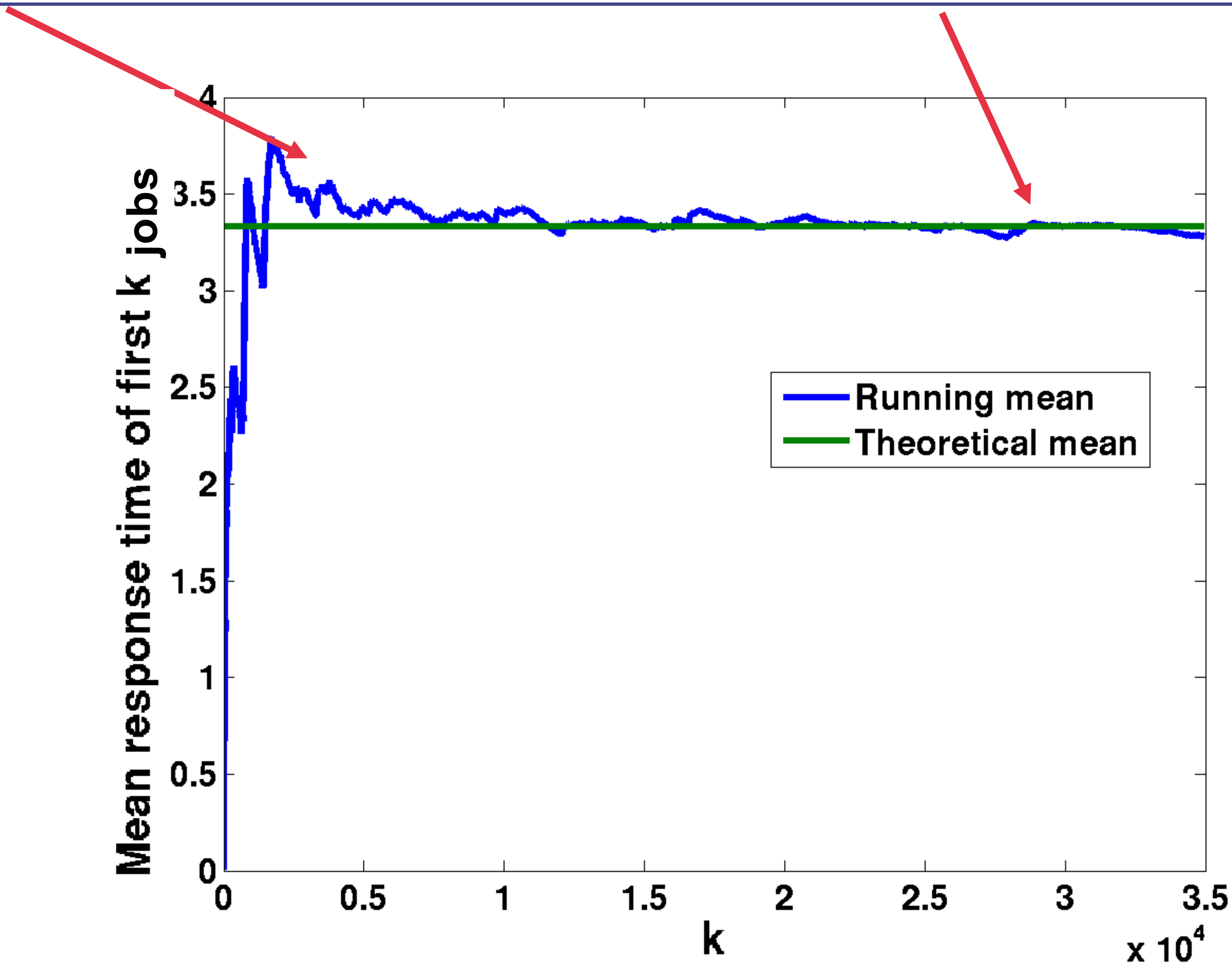
$$M(k) = \frac{X(1)+X(2)+\ldots+X(k)}{k}$$

- For example, if k = 5, then

$$M(5) = \frac{X(1)+X(2)+\ldots+X(5)}{5}$$

- Thus M(5) is the mean response time of the first 5 jobs
- In general, M(k) is the mean response time of the first *k* jobs
- Let us plot M(k) - see the next slide

Transient behaviour versus steady state behaviour



Transient removal: Introduction (1)

- The early part of the simulation displays transient (= non-steady state behaviour)
- The later part of the simulation converges or fluctuates around the steady state value
- Since we are interested in the steady state value, we should not use the transient part of the data to compute the steady state value
- We should remove the transient part and only use the steady state part to compute the mean
- One method to identify the transient part is to use visual inspection
  - Note: In the previous slide, we have the theoretical value to guide us but in practice you don't, you will learn a transient removal method based on batch means in Revision Problem 5B

Transient removal: Introduction (2)

- Let us assume that the first *m* jobs constitute the transient part and there are *N* jobs altogether, we should revise the formula to compute the mean to

$$\frac{X(m+1) + X(m+2) + \ldots + X(N)}{N-m}$$

- Note: We used too simple a method to compute the mean in Week 5A but I didn't want to complicate things at that time!

- **Important: You must run the simulation long enough so that you have a good number of data points (or jobs) in the steady state part.**

Independent replications

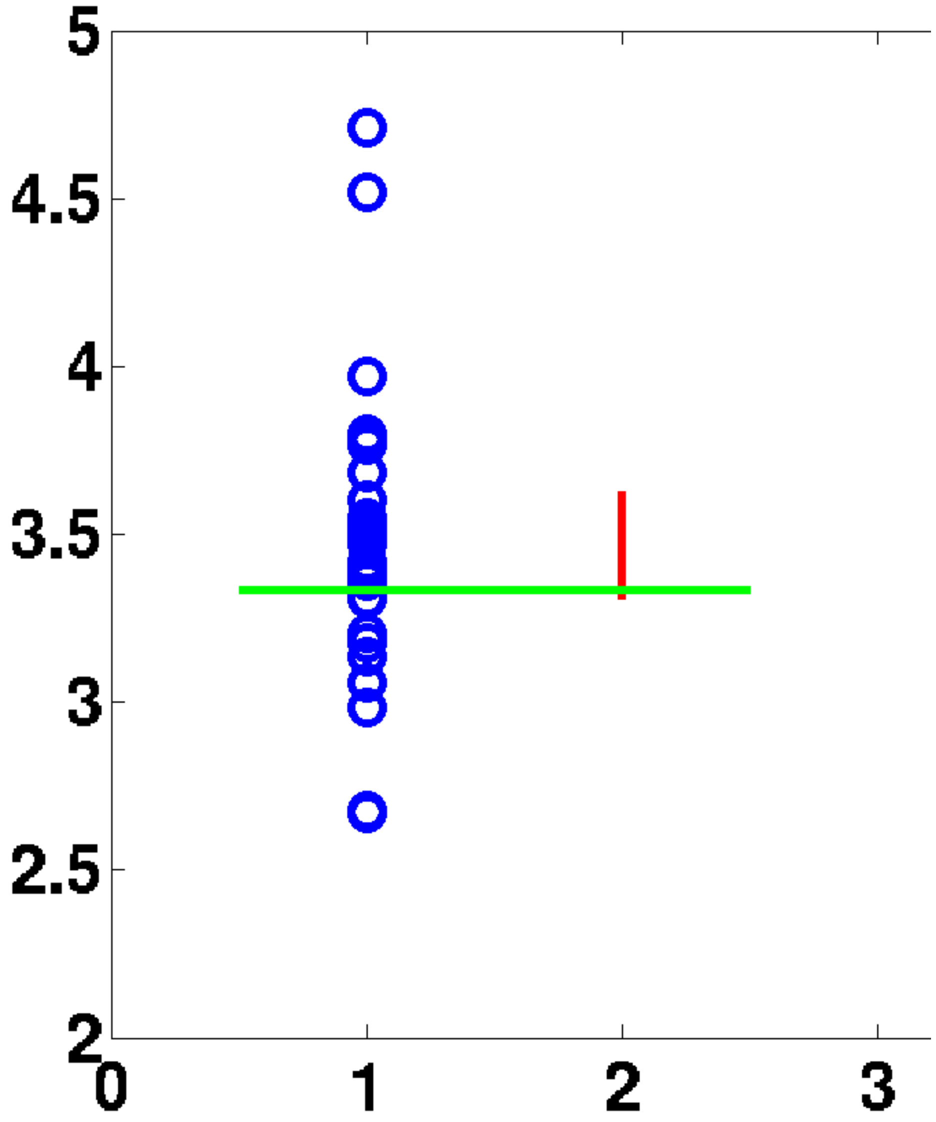
- Assume that we carry out simulations to find out what the steady state mean response time of a queueing system is
  - Important note: *We cannot get exact answer from simulation*
  - We express our simulation results as e.g. there is a probability of 95% that the mean response time is in the interval [3.1,3.3].
  - We call the interval [3.1,3.3] the 95% *confidence interval*.
- Independent replications: Repeat the simulation a number of times using ***different*** sets of random numbers
- Why independent replications?
  - Independent replications allow us to use statistical method to estimate a confidence interval of steady state mean response time

Example: Independent replications

- We want to use simulation to estimate the mean response time of an M/M/1 queue with
  - Arrival rate λ = 0.7
  - Service rate μ = 1
  - Simulation ends when master clock is 16000s
- We repeat the experiment 30 times using different sets of random numbers
- For each independent experiments
  - We record the response time of all the jobs
  - Remove the transient part
  - Compute the mean response time using the steady state section
- We obtain 30 different estimates of the mean response time, one from each independent experiment
- These independent estimates allow us to find a confidence interval

Example (Cont'd)

- The blue circles show the estimated mean response time from the 30 independent experiments
- The red line is the 95% confidence interval
  - There is a 95% probability that the true mean response time that we want to estimate is in the interval [3.30,3.62]
- The green line is the theoretical mean response time (which you should not normally know).



Computing the confidence interval (1)

- Assume that you do *n* independent replications
- In each replication, you remove the transient part and compute an estimate of the mean steady state response time
  - Let us call your estimate from the k^th replication, T(k)
- Compute the sample mean

$$\hat{T} = \frac{\sum_{i=1}^{n} T(i)}{n}$$

- And the sample standard deviation

$$\hat{S} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{T} - T(i))^2}{n-1}}$$

Note: for sample standard deviation, *(n-1)* is in the denominator, *not n*.

Computing confidence interval (2)

- There is a probability (1-α) that the mean response time that you want to estimate lies in the interval

$$[\hat{T} - t_{n-1,1-\frac{\alpha}{2}} \frac{\hat{S}}{\sqrt{n}}, \hat{T} + t_{n-1,1-\frac{\alpha}{2}} \frac{\hat{S}}{\sqrt{n}}]$$

where $t_{n-1,1-\frac{\alpha}{2}}$ is the upper $(1 - \dfrac{\alpha}{2})$

critical point for the Student t distribution

with $(n-1)$degrees of freedom

Computing confidence interval (3)

- The value $t_{n-1,1-\frac{\alpha}{2}}$ can be obtained from looking up

  the Student t distribution table
  - Note: A Student t table has been provided on the web site

- There are also programs that compute it
  - In Matlab, you can use tinv(1-alpha/2,n-1)

Example: Independent replications (cont'd) $t_{n-1,1-\frac{\alpha}{2}}$

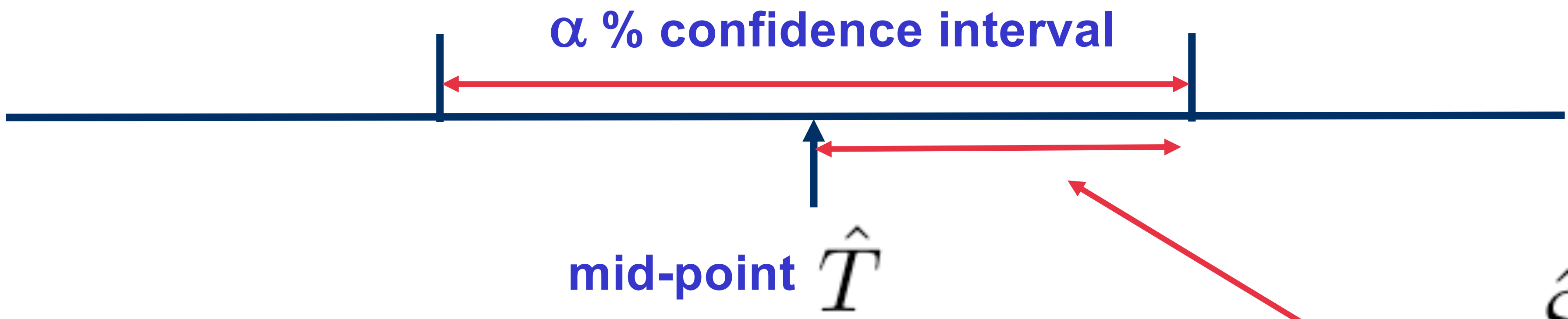
- From the example on p.15
- The sample mean of (n =) 30 replications = 3.47
- The sample standard deviation of 30 replications is 0.43
- If we want to compute the 95% confidence interval, α = 0.05
- Since we did 30 independent experiments and want 95% confidence interval, we use $t_{29,0.975}$
- From the t-distribution table, the value of $t_{29,0.975}$ is 2.0452, the 95% confidence interval is

$$[3.47 - 2.0452\frac{0.43}{\sqrt{30}}, 3.47 + 2.0452\frac{0.43}{\sqrt{30}}] = [3.30, 3.62]$$

More on confidence interval

- Confidence interval

$$[\hat{T} - t_{n-1,1-\frac{\alpha}{2}}\frac{\hat{S}}{\sqrt{n}}, \hat{T} + t_{n-1,1-\frac{\alpha}{2}}\frac{\hat{S}}{\sqrt{n}}]$$



$$t_{n-1,1-\frac{\alpha}{2}}\frac{\hat{S}}{\sqrt{n}}$$

$$n \uparrow \; \Rightarrow t_{n-1,1-\frac{\alpha}{2}} \downarrow$$

$$\Rightarrow t_{n-1,1-\frac{\alpha}{2}}\frac{\hat{S}}{\sqrt{n}} \downarrow$$

What can we get from simulation?

- If your queueing problem has a mathematical solution, you will get *one* value for the steady state mean response time
- If you simulate a queue to try to estimate the mean response time, you will *not* know the exact value of the steady state mean response time
- Simulation can only give you a confidence interval of what you want to estimate
- You can reduce the confidence interval by doing many independent replications!

Choice of simulation parameters (1)

- Simulation parameters
  - Length of simulation
  - Number of replications
  - Accuracy

- Unfortunately, there are *no* hard rules to choose them. You will need to do some trial and error
  - If the length of simulation is not long enough, you will need to increase it
  - If the number of replications is not enough to give you the desired accuracy, you will need to increase it

Choice of simulation parameters (2)

- Length of simulation
  - Must be longer than the transient
  - Should have a good number of data point in the steady state part
    - Hard to say what "good" is. Get a few hundred if you can. The more the better but of course your simulation will run longer

- Number of replications
  - You may want to have 5 replications to start with
  - After removing the transient, compute the confidence interval for your estimate.
  - Compare the width of your confidence interval with your desired accuracy. If the confidence interval that you have obtained is too wide, you will need to increase the number of replications.
  - Progressively (basically by trial-and-error), increase the number of replications until you get the desired level of accuracy

Summary

- Simulation is not just a computer programming exercise
- You need to make sure that your program is correct
- It is also important to analyse your results statistically
- Methods discussed include
  - Transient removal technique
  - Confidence interval
  - Determining number of replications

References

- The primary reference is Law and Kelton, “Simulation Modelling and Analysis”
  - Transient removal, Sections 9.1, 9.2 and 9.5
  - Replication method, Section 9.5.2
- Raj Jain, “The Art of Computer Systems Performance Analysis” has materials on
  - Transient removal methods, Section 25.3
  - Calculating confidence interval, Section 13.2
- If you are interested to know the mathematical background on confidence interval, student t-distribution etc., a possible reference is Wackerly et al, “Mathematical Statistics with Applications”.