Lecture 9: Output Data Analysis

Holger Füßler

Course overview

1. Introduction	7. NS-2: Fixed networks
2. Building block: RNG	8. NS-2: Wireless networks
3. Building block: Generating random variates I and modeling examples	9. Output analysis
4. Building block: Generating random variates II and modeling examples	10. Output analysis: comparing different configurations
5. Algorithmics: Management of events	11. Omnet++ / OPNET
6. NS-2: Introduction	12. Simulation lifecycle, summary

Part A: Output Data Analysis for a Single System

Holger Füßler

Holger Füßler
Universität Mannheim

Structure – Part A

- Part A.I: Problem statement ‘output analysis’
- Part A.II: Some probability theory and statistics
- Part A.III: Types of discrete event simulations
- Part A.IV: Credibility of simulation studies

A.I Problem statement

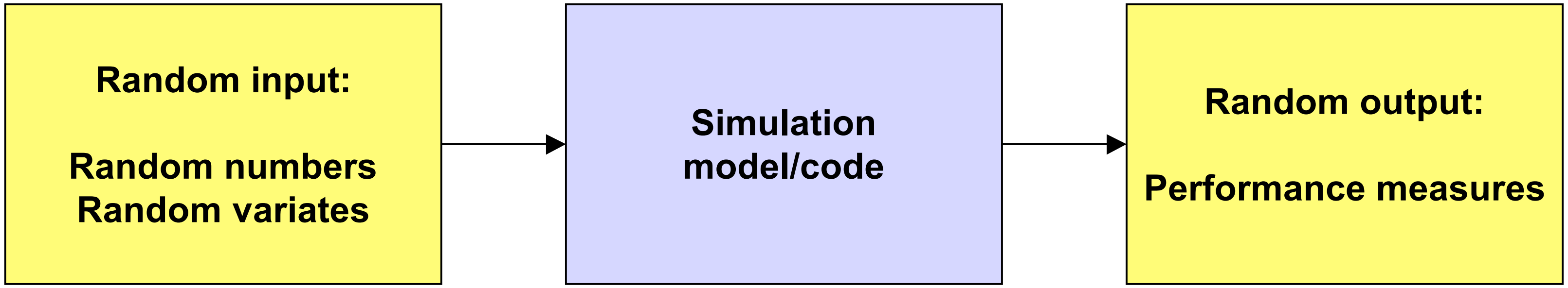
- By performing a simulation study one ‘observes’ a system.
- As output one gets a collection of data (statistics, traces).
- What ‘features’ can be inferred about the system?

“In many simulation studies a great deal of time and money is spent on model development and ‘programming,’ but little effort is made to analyze the simulation output data appropriately.”

“… a simulation is a computer-based statistical sampling experiment.”

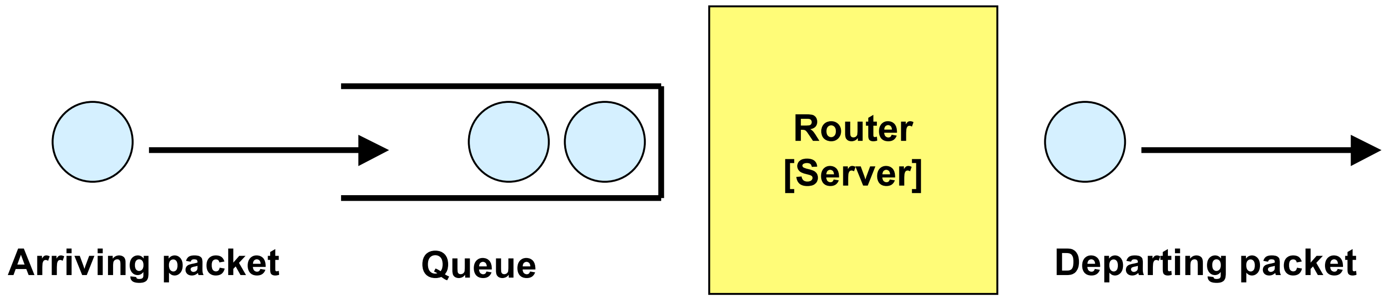
[LK2000, Chap. 9]

A.I Problem statement cont‘d



- Do not get ‚exact‘ answers
- Two different runs of the same model: different numerical results

A.I Example 1: results for M/M/1 queue



Varying seeds (NetSim lab 3, change argument of lcgrand):

Replication	Average delay	Average number in queue	Server utilization
1	0.606	0.060	0.192
2	0.554	0.057	0.206
3	0.586	0.057	0.200
4	0.452	0.046	0.197
5	0.490	0.050	0.199

A.I Example 1: results for M/M/1 queue cont‘d

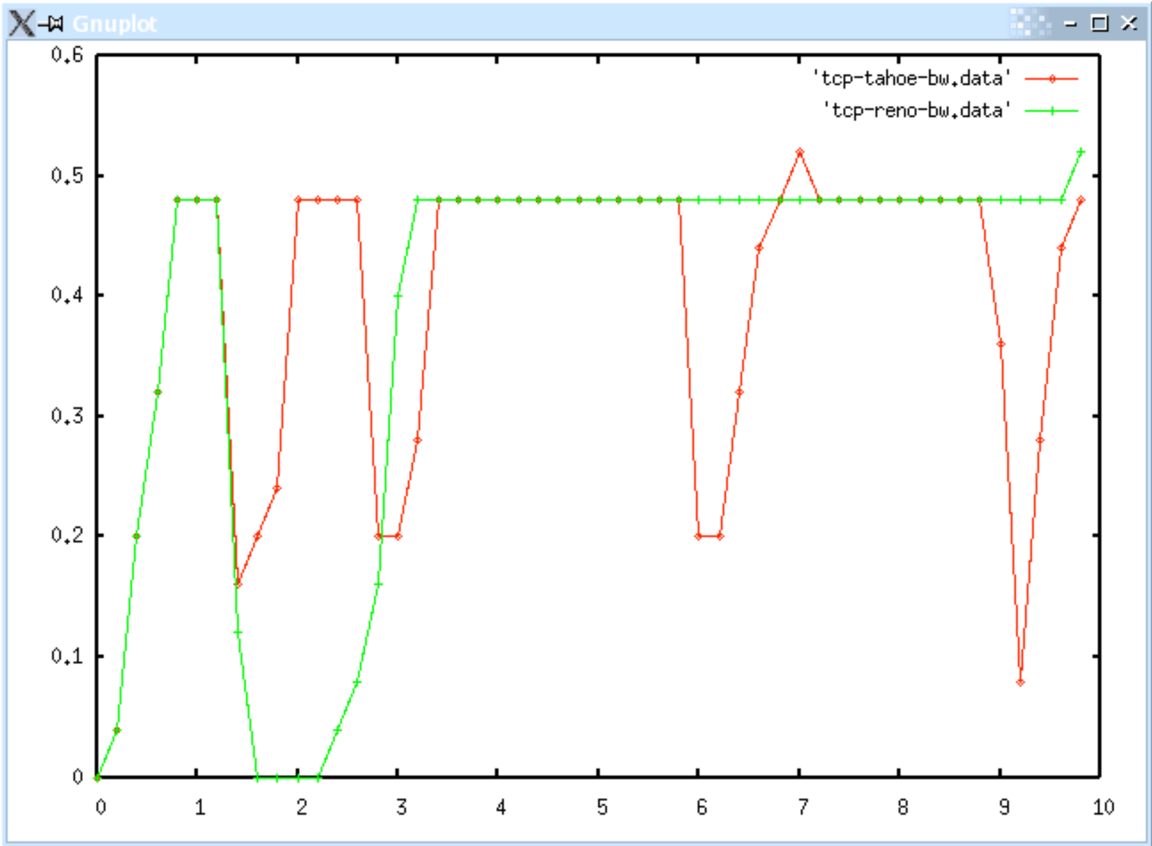
» Varying number of packets (NetSim lab 3, change parameter in mm1.in)

Number of packets	Average delay	Average number in queue	Server utilization
1000	0.606	0.060	0.192
2000	0.554	0.057	0.206
3000	0.586	0.057	0.200
4000	0.452	0.046	0.197
5000	0.490	0.050	0.199

» What is the ‚true‘ value?

» How much does the obtained result differ from the ‚true‘ value?

A.I Example 2: TCP average achievable bandwidth



- **How long is the warm-up period?**
- **Assume that we have random bit/link errors: what is the achievable bandwidth?**

A.I Example 2: TCP achievable bandwidth

```
+ 7.805824 3 2 ack 40 ------- 0 3.0 0.0 115 268
- 7.805824 3 2 ack 40 ------- 0 3.0 0.0 115 268
r 7.809248 3 2 ack 40 ------- 0 3.0 0.0 114 265
+ 7.809248 2 1 ack 40 ------- 0 3.0 0.0 114 265
- 7.809248 2 1 ack 40 ------- 0 3.0 0.0 114 265
r 7.81728 0 1 tcp 1040 ------- 0 0.0 3.0 124 266
+ 7.81728 1 2 tcp 1040 ------- 0 0.0 3.0 124 266
- 7.81728 1 2 tcp 1040 ------- 0 0.0 3.0 124 266
r 7.81744 1 2 tcp 1040 ------- 0 0.0 3.0 117 258
+ 7.81744 2 3 tcp 1040 ------- 0 0.0 3.0 117 258
- 7.81744 2 3 tcp 1040 ------- 0 0.0 3.0 117 258
r 7.818944 0 1 tcp 1040 ------- 0 0.0 3.0 125 267
+ 7.818944 1 2 tcp 1040 ------- 0 0.0 3.0 125 267
r 7.822464 2 3 tcp 1040 ------- 0 0.0 3.0 116 257
+ 7.822464 3 2 ack 40 ------- 0 3.0 0.0 116 269
- 7.822464 3 2 ack 40 ------- 0 3.0 0.0 116 269
r 7.825888 3 2 ack 40 ------- 0 3.0 0.0 115 268
…
```

» **Conversion to meaningful results?**

A.I General set-up

m Observations

n Experiments

$$\left\{\begin{matrix} Y_{11} & \dots & Y_{1i} & \dots & Y_{1m} \\ Y_{21} & \dots & Y_{2i} & \dots & Y_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ Y_{n1} & \dots & Y_{ni} & \dots & Y_{nm} \end{matrix}\right.$$

Warm-up period

Challenge:
Choose n, m and warm-up period approriately!

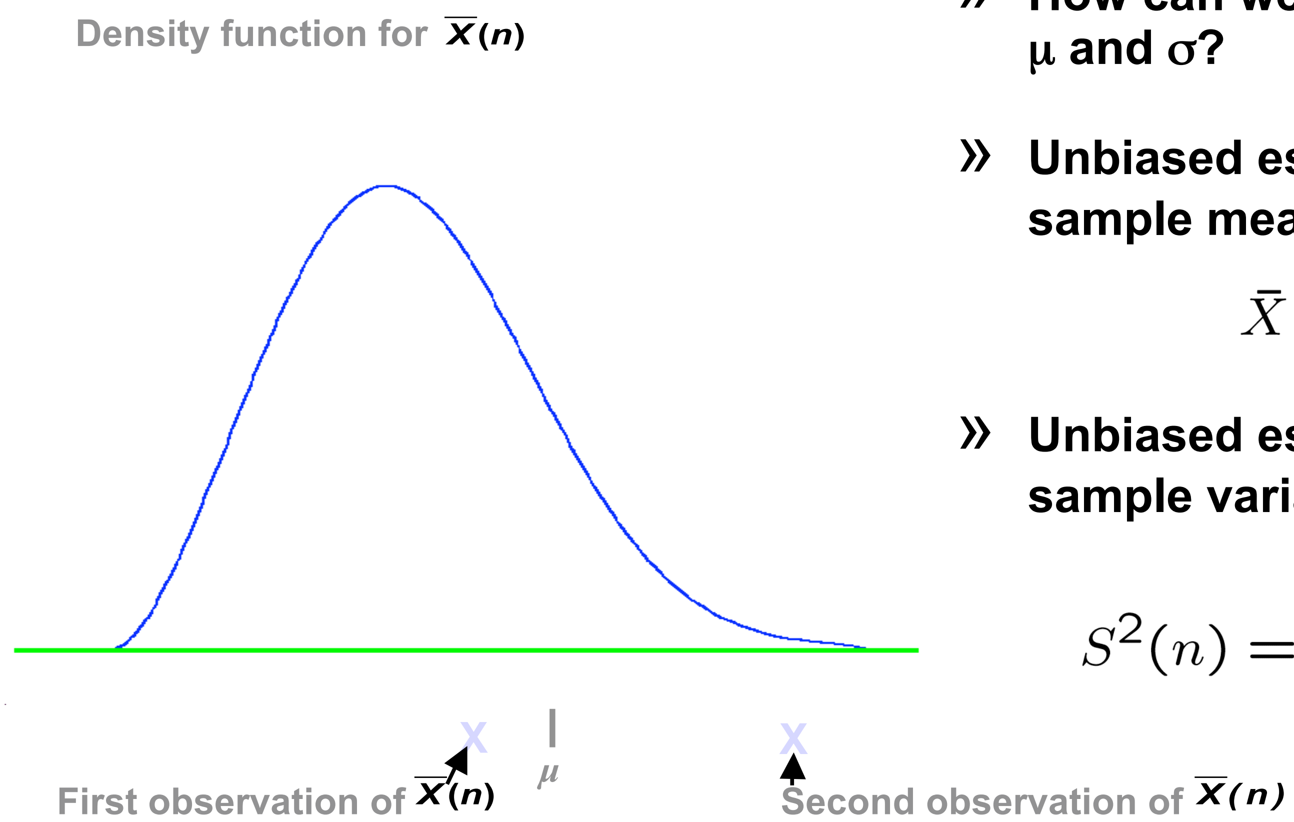
A.II Estimates of means and variances

- Suppose that $X_1, X_2, \ldots, X_n$ are IID random variables with finite mean μ and finite variance σ^2.
- How can we get an estimate for μ and σ?
- Unbiased estimator for μ: sample mean

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

- Unbiased estimator for σ^2: sample variance

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



A.II Estimates of means and variances (Calculation)

let X_j^{Σ} be the sum of all values until the j-th element and $X_j^{\Sigma^2}$ the sum of the respective squares, i.e.

$$X_j^{\Sigma} = \sum_{i=1}^{j} X_i \qquad X_j^{\Sigma^2} = \sum_{i=1}^{j} X_i^2$$

then, the estimators can be calculated as

$$\bar{X}(n) = \frac{X_n^{\Sigma}}{n}$$

$$S^2(n) = \frac{1}{n-1}\left(X_n^{\Sigma^2} - 2 \cdot (\bar{X}(n)) \cdot X_n^{\Sigma} + n \cdot (\bar{X}(n))^2\right)$$

→ **we only need the accumulated sum / squared sum of each variable**

A.II Confidence intervals for the mean

- Again, assume $X_1, X_2, \ldots, X_n$ are IID random variables with finite mean μ and finite variance σ^2 greater 0.
- Central limit theorem states: $\bar{X}$(n) is approximately distributed as a normal random variable with mean μ and variance σ^2/n.
- For sufficiently large n, an approximate 100(1-α) percent confidence interval for μ is given by

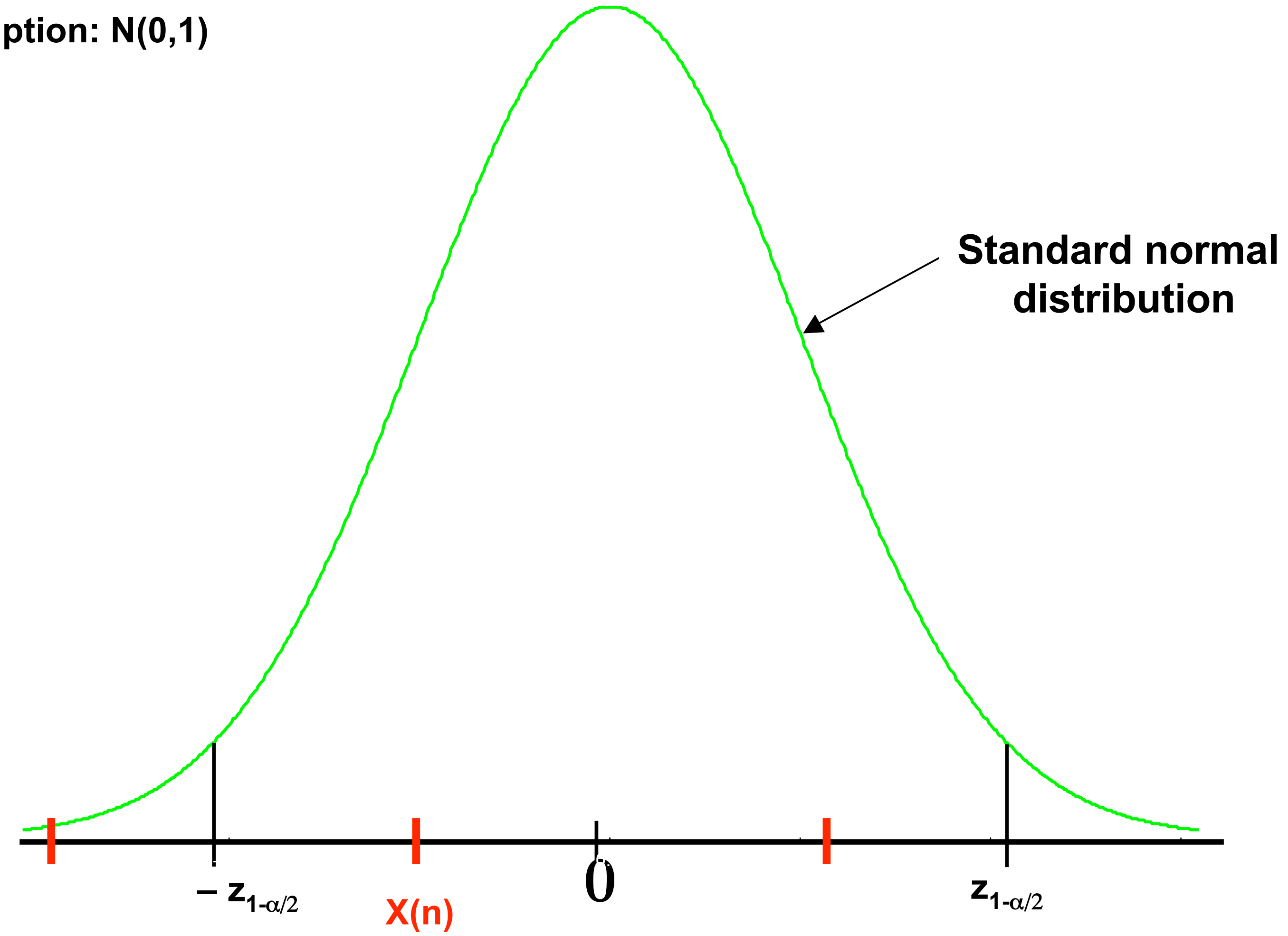
$$\bar{X}(n) \pm z_{1-\alpha/2}\sqrt{\frac{S^2(n)}{n}}$$

- Interpretation of 'confidence interval': "… in 100(1-α) percent of all cases the true parameter μ is within the interval.
- Why 'approximate'?
  - It is only asymptotically correct

Warning: Only „good“ for sample size of approx. > 50

A.II Confidence intervals for the mean: illustration

Assumption: N(0,1)



A.II Example: M/M/1 queue

Replication	Average delay
1	0.606
2	0.554
3	0.586
4	0.452
5	0.490
6	0.548
7	0.519
8	0.498
9	0.366
10	0.364
Average (.3digits)	0.498

- **Experiments for 1000 packets each**
- **$S^2(10) = 0.007$**
- **95% confidence interval:**

  $z \approx 2.0$
- **0.498 ± 0.014**

A.II Selecting the sample size

» **Result so far (under a lot of assumptions): for sufficiently large n, an approximate 100(1-α) percent confidence interval for μ is given by**

$$\bar{X}(n) \pm \underbrace{z_{1-\alpha/2}\sqrt{\frac{S^2(n)}{n}}}_{\text{Let us bound this by an absolute value v}}$$

How do we have to select the number of samples n?

» **Thus, given $S^2(n)$ (or σ^2 if it is known) and value v and $z_{1-\alpha/2}$, one can solve for number of samples n.**

A.II Example: calculating required sample sizes

- Let Y_i, i=1, 2, …, be IID Bernoulli random variables with parameter p.
- What is the sample size necessary to estimate p within 0.05 with probability .95?
- Assume no information is given w.r.t. the variance.
- Max. variance: 0.025
- $0.025 \cdot (2/0.05)^2 = n$
- n=400

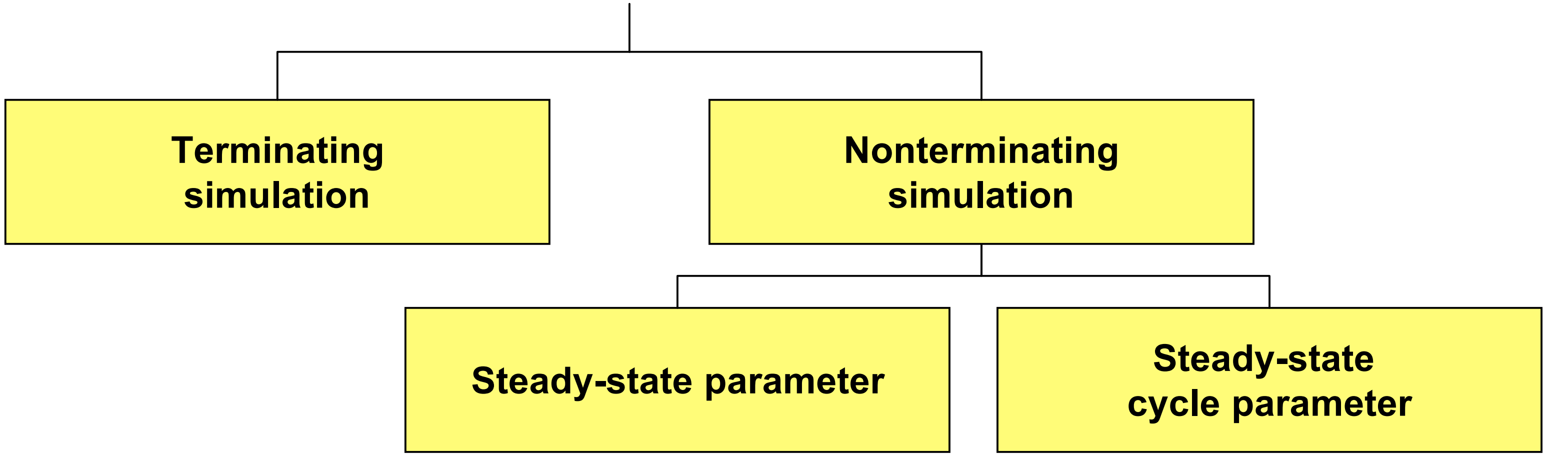
A.II Experiment: Estimated coverages

- Coverage: proportion of confidence intervals that contain the ‚true' parameter μ.
- Should be 1 - α for ‚n sufficiently large'
- Can be checked for known distributions.

Distribution	Skewness	M=5	M=10	M=20	M=40
Normal	0.00	0.910	0.902	0.898	0.900
Exponential	2.00	0.854	0.878	0.870	0.890
Chi square	2.83	0.810	0.830	0.848	0.890
Lognormal	6.18	0.758	0.768	0.842	0.852
Hyperexp.	6.43	0.584	0.586	0.682	0.774

Estimated coverages for 90 percent confidence intervals based on 500 independent experiments for each of the sample sizes [Source: Law/Kelton]

A.III Types of simulations w.r.t. output analysis



Terminating: Parameters to be estimated are defined relative to specific initial and stopping conditions that are part of the model

Nonterminating: There is no natural and realistic event that terminates the model

Interested in “long-run” behavior characteristic of “normal” operation

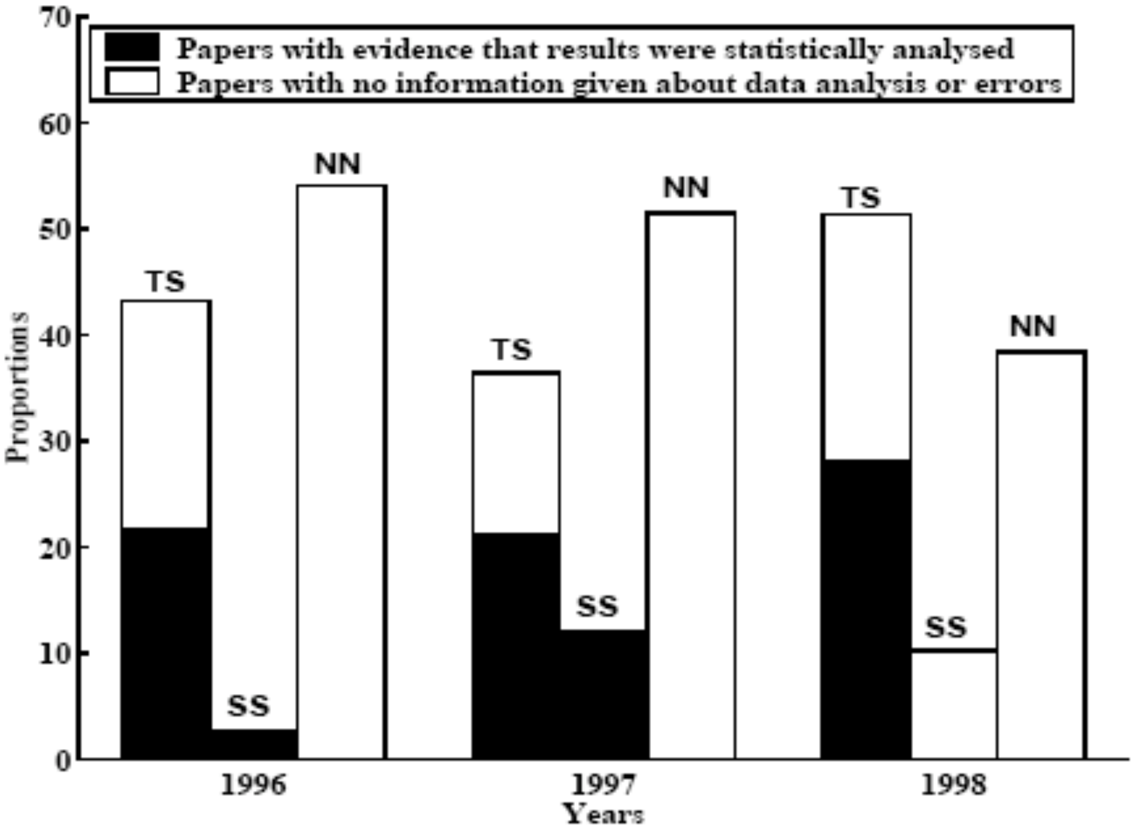
If the performance measure of interest is a characteristic of a steady-state distribution of the process, it is a *steady-state* parameter of the model

Not all nonterminating systems are steady-state: there could be a periodic “cycle” in the long run, giving rise to *steady-state cycle* parameters

A.III Types of simulations w.r.t. output analysis

- Terminating simulations: „9 to 5 scenarios“
- Example: M/M/1 queue
  - Initial condition: empty queue
  - Terminating condition: time elapsed
- Statistics for terminating simulations: see Part II of this lecture
- Challenge: steady-state simulations
  - How to get rid of impact of initial condition?
  - When to stop simulation?

A.IV „Crisis of credibility“

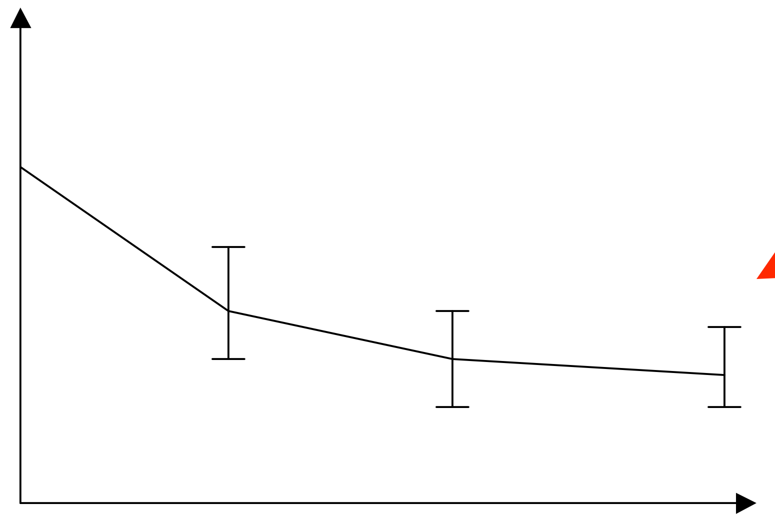


(c) IEEE/ACM Transactions on Networking

K. Pawlikowski, H.-D. Joshua Jeong, and J.-S. Ruth Lee, *On credibility of simulation studies of telecommunication networks*, IEEE Communications Magazine 40(1).

A.IV Recommendations (Pawlikowski et al.)

» Reported simulation experiments should be repeatable
  - Give information about
    - The PRNG(s) used during the simulation
    - The type of simulation
    - The method of analysis of simulation output data
    - The final statistical errors associated with the results



Indicate confidence interval (CI)
for specified confidence level (CL)

A.IV Comments for ‚best practices'

- Independence or covariance-stationarity rarely encountered in practice ☺
- But: if the number of replications, samples etc. is too low even under the assumptions of independence or covariance-stationarity, something is probably flawed …
  - We need mathematical results to check
- In reality, also time and space constraints can severely impact achievable confidence intervals
  - But this should be specified
- Trace and plot as many variables as possible to cross-check correctness ☺

A.IV Wrap-up

- Any stochastic computer simulation (using RNGs/(PRNGs) has to be regarded as a (simulated) statistical experiment.
- Statistics background:
  - estimating means and variances
  - confidence levels and intervals
  - hypothesis testing
- Transient and steady-state behavior
- Terminating, steady-state and cyclic steady state simulations
- The issue of credibility

References – Part A

- Averill M. Law, W. David Kelton: "Simulation Modeling and Analysis", McGraw-Hill, 3rd edition, 2000
  - Chapters 4 and 9

Part B: Comparing Different Configurations

Holger Füßler
Universität Mannheim

Where we are …

Why network simulations?

» **Educational use**
  – See protocol in action
  – Does it work as intended?

» **Get some quantitative results for a single configuration**
  – E.g., how long does it take to find a route?

» **Compare different configuration and decide which one is ‘better’**
  – Trade-offs

Design phase

Tuning phase

Decision phase

Getting harder

Lecture overview – Part B

» **Part B.I: Scope of this Lecture - Motivation**

» **Part B.II: Comparison of different configurations**

» **Part B.III: Variance reduction techniques**

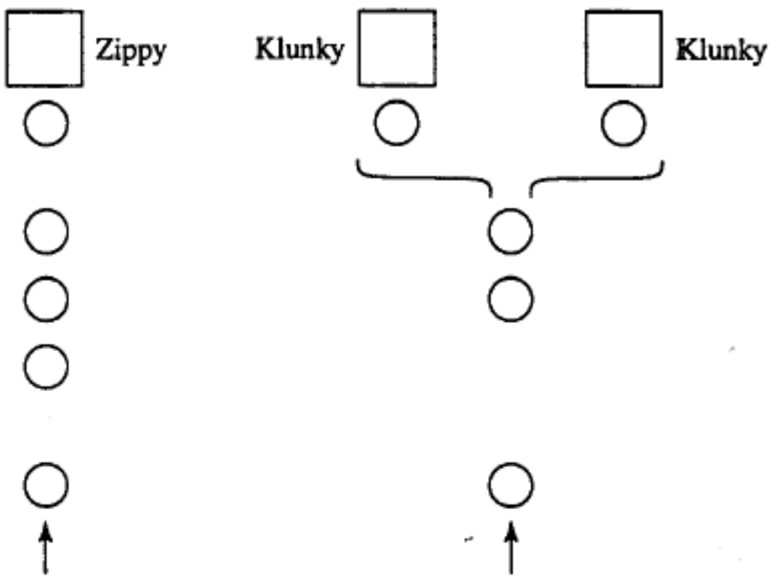
B.I Comparing different configurations: e.g. Different Routing Protocols

» **Which one is ‘better’?**
  - **Answer depends on metric(s)**
  - **Answer depends on scenario**
    - **Mobility pattern**
    - **Communication pattern**
    - **Caching strategies**
    - **Flooding strategies**
    - **…**

» **Metrics:**
  - **Packet delivery ratio**
  - **Route acquisition time**
  - **End-to-end delay**
  - **Overhead costs**

B.II Comparison of two different configurations

- **Let's go back to some simple scenario**
- **Compare M/M/1 queue (service time: exponential with mean 0.9) with M/M/2 queue (service time: exponential with mean 1.8 each)**

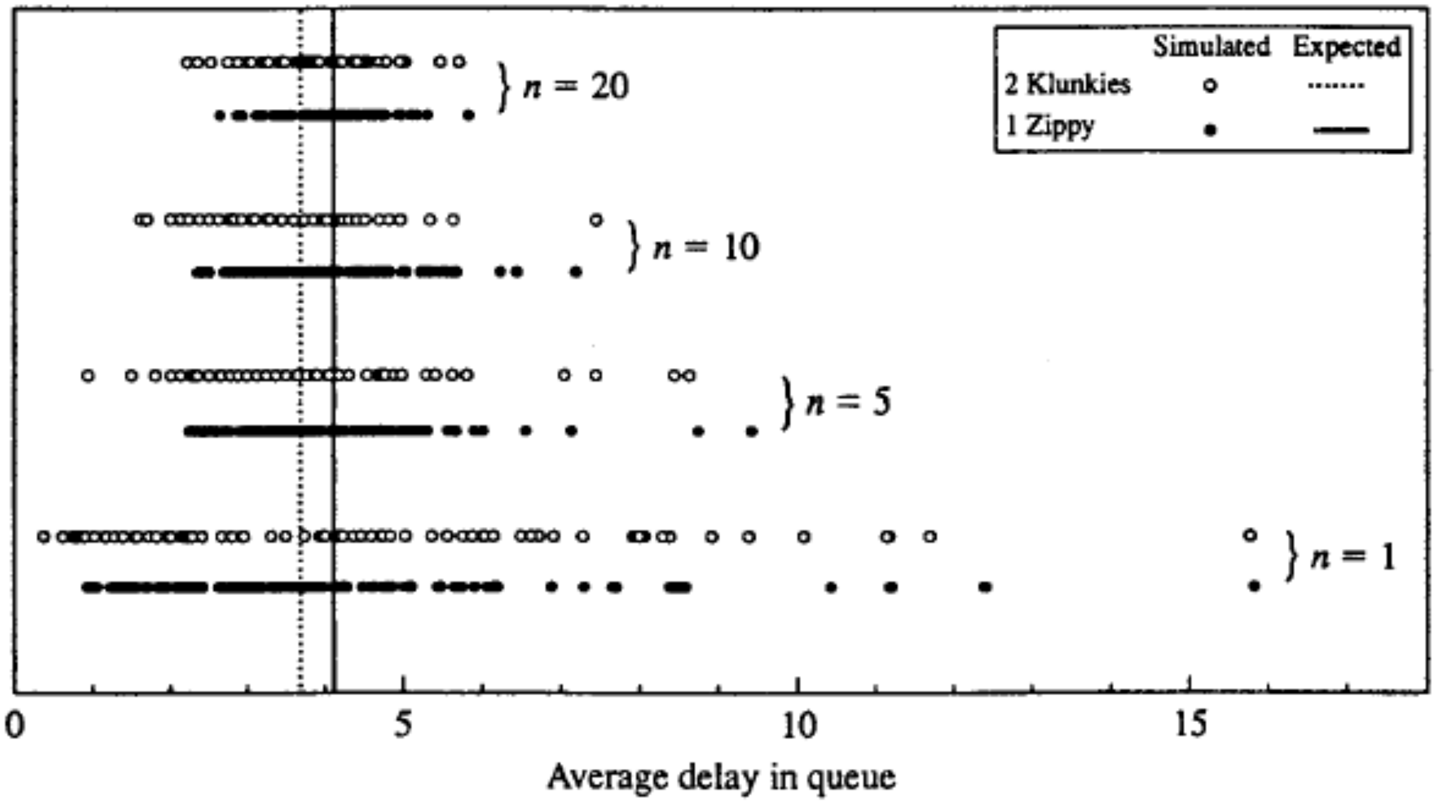


B.II Motivation by example 2

Experiment	$\hat{d}_Z(100)$	$\hat{d}_K(100)$	Recommendation	
1	3.80	4.60	Zippy	(wrong)
2	3.17	8.37	Zippy	(wrong)
3	3.96	4.18	Zippy	(wrong)
4	1.91	5.77	Zippy	(wrong)
5	1.71	2.23	Zippy	(wrong)
6	6.16	4.72	Klunky	(right)
7	5.67	1.39	Klunky	(right)
⋮	⋮	⋮	⋮	
98	8.40	9.39	Zippy	(wrong)
99	7.70	1.54	Klunky	(right)
100	4.64	1.17	Klunky	(right)

n	P(wrong answer)
1	0.52
5	0.43
10	0.38
20	0.34

B.II Motivation by example 3



B.II Confidence interval for the difference between two systems

- **Two alternative simulated systems (*i* = 1, 2), μ_i = expected performance measure from system *i***
- **Take "sample" of n_i observations (replications) from system *i***
- **X_{ij} = observation *j* from system *i***
- **Want: confidence interval on $z = \mu_1 - \mu_2$**
- **If interval misses 0, conclude there is a *statistical* difference between the systems**
- **Is the difference *practically* significant? Must use judgment in context.**

B.II Paired confidence interval

- Assume $n_1 = n_2$ (=n, say)
- For a fixed j, X_{1j} and X_{2j} need not be independent
  - Important for variance reduction techniques (next part)
- Let $Z_j = X_{1j} – X_{2j}$
- Problem reduced to 'single system problem'
- Find confidence interval for $E[Z_j]$
- Previous example (10 runs):

$$\bar{Z}(10) = 0.376$$
$$S^2(10)/90 \approx 1.25$$
$$\text{Confidence interval for CL 95\%: } 0.376 \pm 2.24$$

B.II Issues not covered in this part

- Other comparison methods
- Comparing more than two systems
- Ranking and selection

B.III Variance reduction techniques

- **Main drawback of using simulation to study stochastic models:**

  Results are uncertain — have *variance* associated with them

- **Would like to have as little variance as possible — more precise results**

- **One sure way to decrease the variance:**

  Run it some more (longer runs, additional replications)

- **Sometimes can manipulate simulation to reduce the variance of the output at little or no additional cost — *not* just by running it some more**

- **Another way of looking at it — try to achieve a desired level of precision (e.g., confidence-interval smallness) with less simulating — *Variance-reduction technique* (VRT)**

B.III Common random numbers (CRN)

- When comparing two or more alternative system configurations
- Basic idea: compare alternative configurations ‘under similar experimental conditions’ – use random numbers ‘for same purpose’
  - Often used ‘unconsciously’
  - E.g. use same movement and communication pattern when comparing two ad-hoc routing protocols
- Example of ‘what can go wrong’:

k	U_k	Usage in $M/M/1$	Usage in $M/M/2$	Agree?
1	0.401	A	A	Yes
2	0.614	A	A	Yes
3	0.434	S	S	Yes
4	0.383	A	A	Yes
5	0.506	S	S	Yes
6	0.709	A	A	Yes
7	0.185	S	S	Yes
8	0.834	A	A	Yes
9	0.646	A	S	No
10	0.376	A	A	Yes

B.III Mathematical basis for CRN

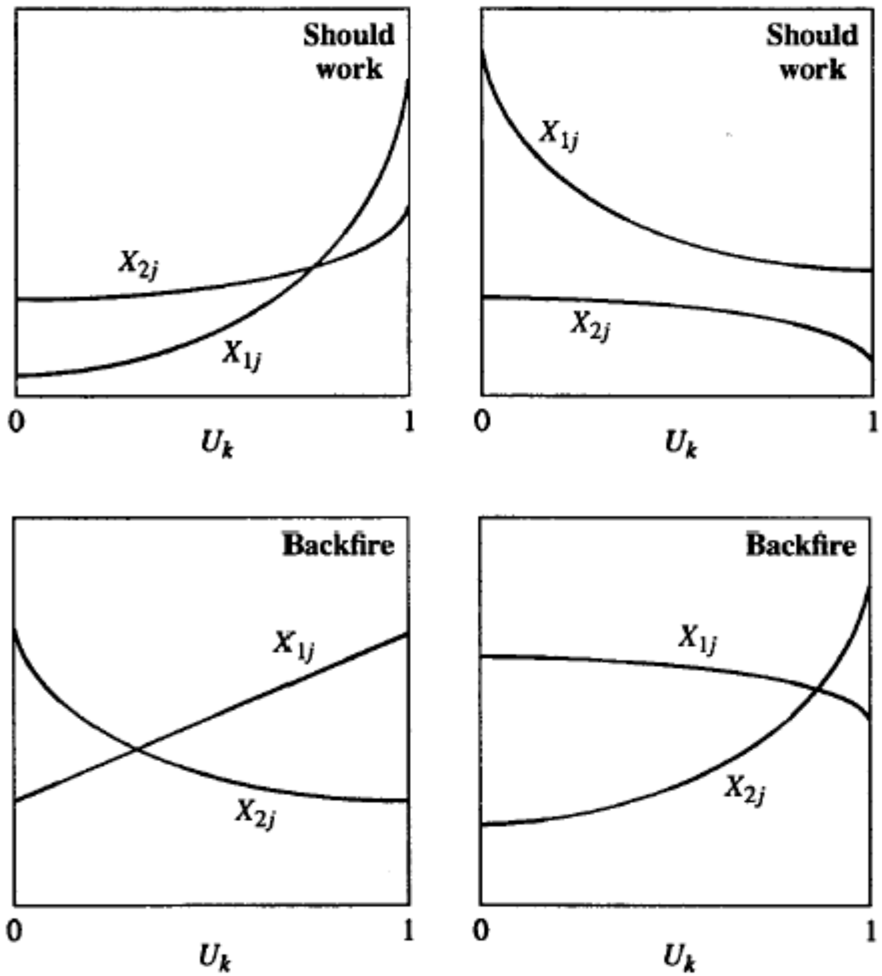
We have two alternatives, where X_{1j} and X_{2j} are the observations from the first and second configuration on the jth independent replication.

Again, let $Z_j = X_{1j} - X_{2j}$.

$$Var[\bar{Z}(n)] = \frac{Var(Z_j)}{n} = \frac{Var(X_{1j}) + Var(X_{2j}) - 2Cov(X_{1j}, X_{2j})}{n}$$

When we can induce some positive correlation, we can make … smaller.

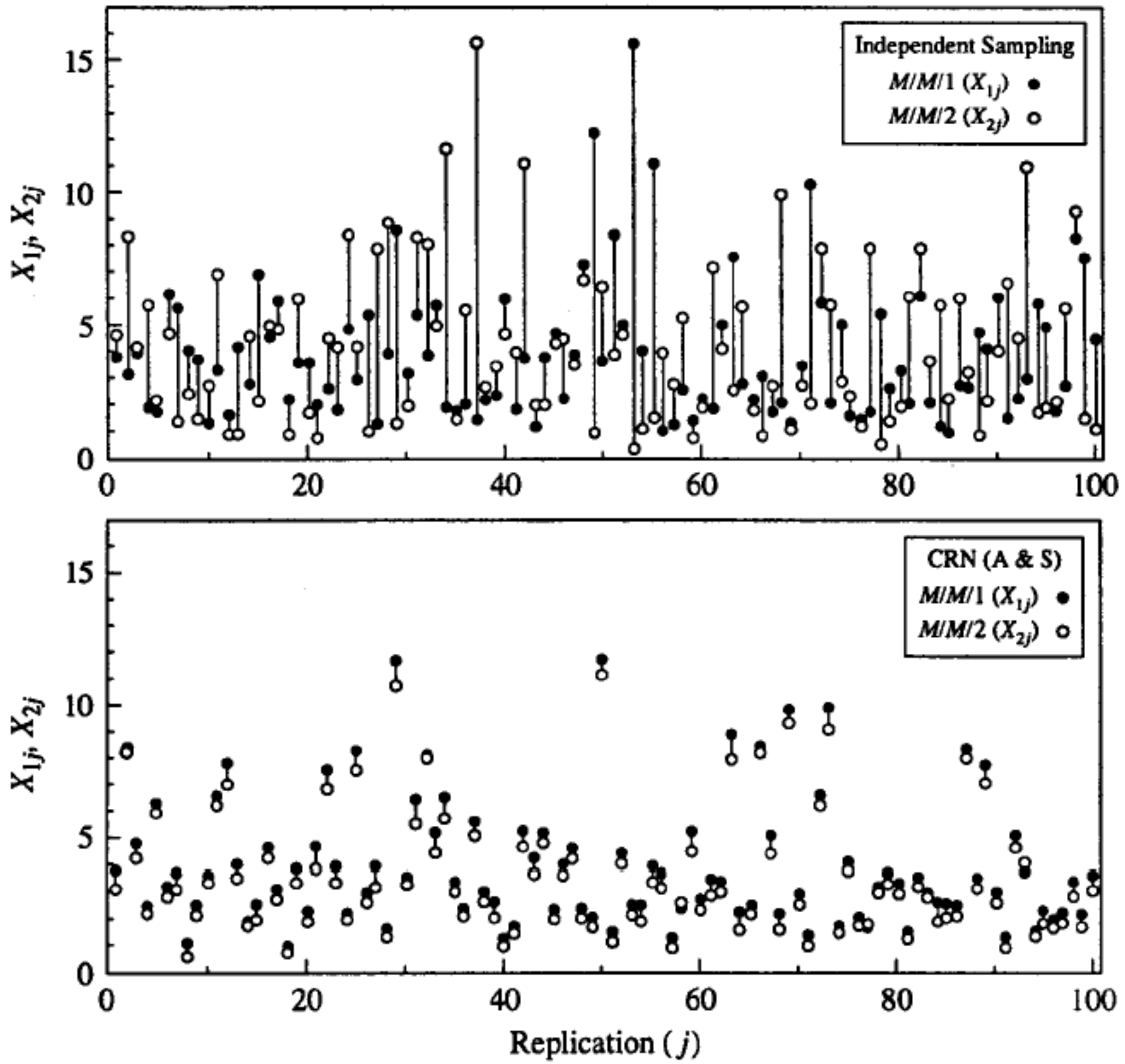
B.III Applicability of CRN



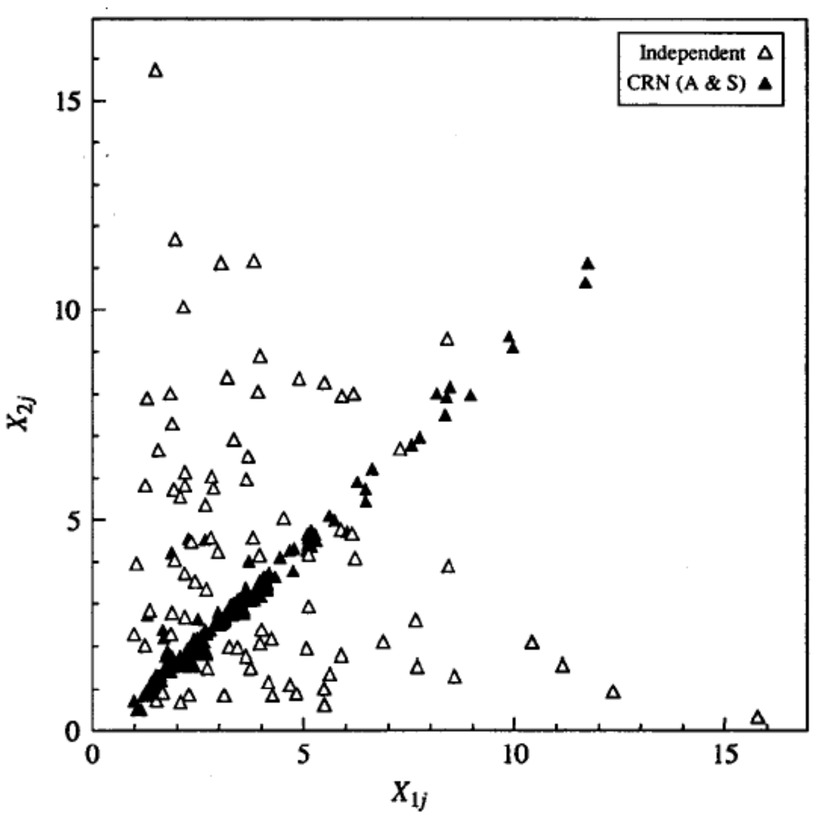
B.III CRN – Synchronization techniques

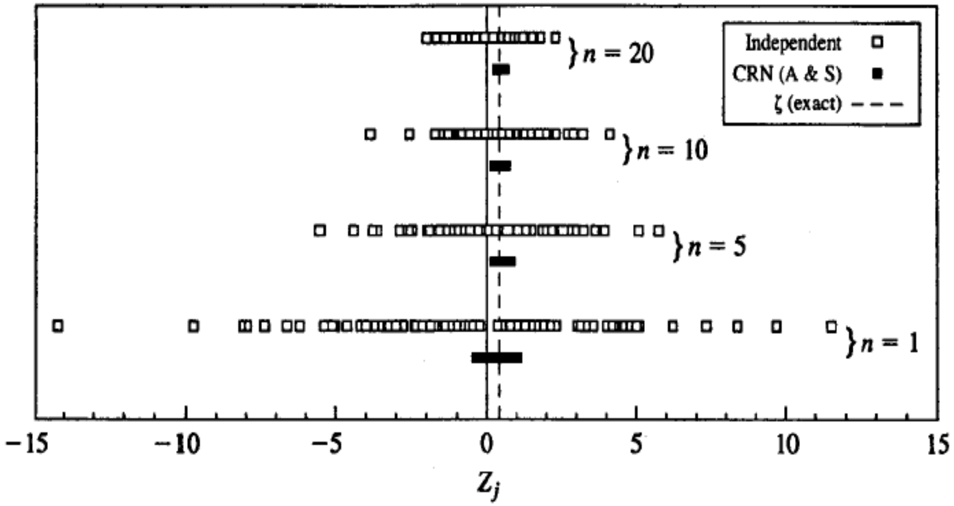
- » Use of dedicated random number streams
- » Use of inverse transform
  - – But: inverse transform not always the most efficient choice …
- » Compute random numbers in advance
  - – Costs some memory
- » … or waste some random numbers
  - – … to keep things synchronized

B.III CRN at work 1



B.III CRN at work 2





B.IV Application to network simulations

- When comparing two alternatives
  - Use ‘same’ topology
    - E.g. preprocessed movement pattern, same radio transmission range
    - E.g. same (preprocessed) link error patterns
  - Use ‘same’ communication pattern
- What else is ‘random’ and can affect results?

Wrap-up Part B

- Today's focus: comparison of two alternative configurations
- Problem reduced to finding a confidence interval of a 'single' system
- Confidence intervals: computation for specified precision
- Precision corresponds to variance (of sample mean): variance reduction techniques needed
  - Common random numbers

References – Part B

- Law, Kelton: Chapters 10 and 11