Università degli Studi di Messina
DIPARTIMENTO DI INGEGNERIA

Dependable Computing
Modeling and Simulation

Discrete event simulation: output analysis

Master degree in
Engineering in Computer Science

Types of simulations

- Terminating simulation
  - Simulation runs for a duration of time T_E
  - E is a specified event
  - Also called *transient* simulation
- Steady-state simulation
  - The system runs continuously or over a long period
  - A state not influenced by initial conditions

Measures of interest

- Often simulation output is a sequence of *discrete time* data

$$\{Y_1, Y_2, \ldots, Y_n\}$$

- They are collected to estimate a measure θ
- It could also be a measure derived from *continuous time* data

Type of data

- Discrete time data
  - Sequence of points $\{Y_1, Y_2, \ldots, Y_n\}$
  - E.g.: mean time spent in a queue
- Continuous time data
  - Data are of the form $\{Y(t), 0 \leq t \leq T_E\}$
  - Queue length at time t

Point estimation

- In the case of discrete-time data ($\{Y_1, \ldots, Y_n\}$), the point estimator of θ is

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

- The estimator is error free (unbiased) if

$$E(\hat{\theta}) = \theta$$

- Usually this is not true: $\left(E(\hat{\theta}) \neq \theta\right)$
- We would like to have

$$E(\hat{\theta}) = \theta$$

Point estimation

- In the case of continuous time data ($\{Y(t), 0 \le t \le T_E\}$), the point estimator is

$$\hat{\Phi} = \frac{1}{T_E} \int_0^{T_E} Y(t)\, dt$$

(time average)

- The estimator is error free (unbiased) if

$$E(\hat{\Phi}) = \Phi$$

Estimation of a measure

- All the measures fall into one of the two categories
- E.g.:
  - Fraction of time during which a queue length is greater then *k*
  - We define the function $Y(t)=\begin{cases} 1 & if \;\; L_Q(t)>k \\ 0 & otherwise \end{cases}$
- Quantile estimation is treated differently

Statistical background

- When the estimator is unbiased, the r.v.

$$t = \frac{\hat{\theta} - \theta}{\sigma(\hat{\theta})}$$

  is distributed as a *Student-t* distribution with *n* (number of collected data) degrees of freedom

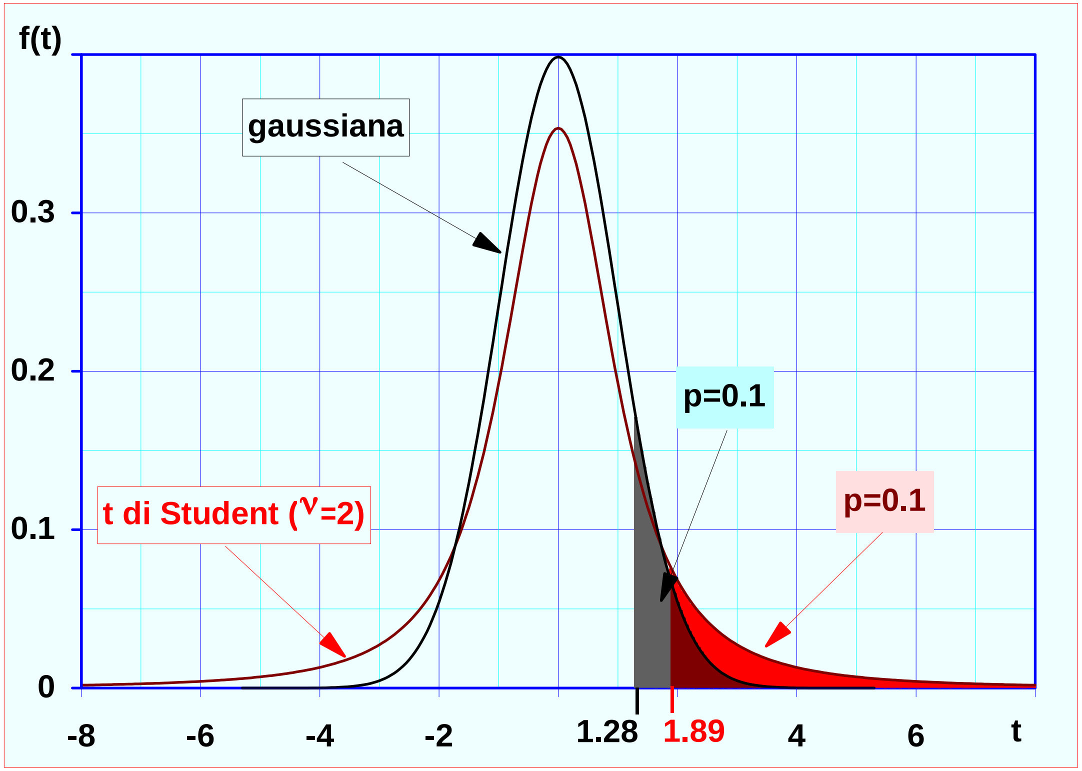
- The sample variance with *n-1* degrees of freedom could be used instead of standard deviation

$$\hat{\sigma} = \frac{s}{\sqrt{n}} \qquad s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(\theta_i - \hat{\theta}\right)^2$$

Student t distribution

A *Students t* distribution tends to a Gaussian distribution when *n* increases

It has greater quantile than the corresponding Gaussion distribution when the degrees of freedom are few



Confidence level

- We need to know the interval such that

$$P[\theta \in I] = \alpha$$

- α is said *confidence level*
- I is said *confidence interval*

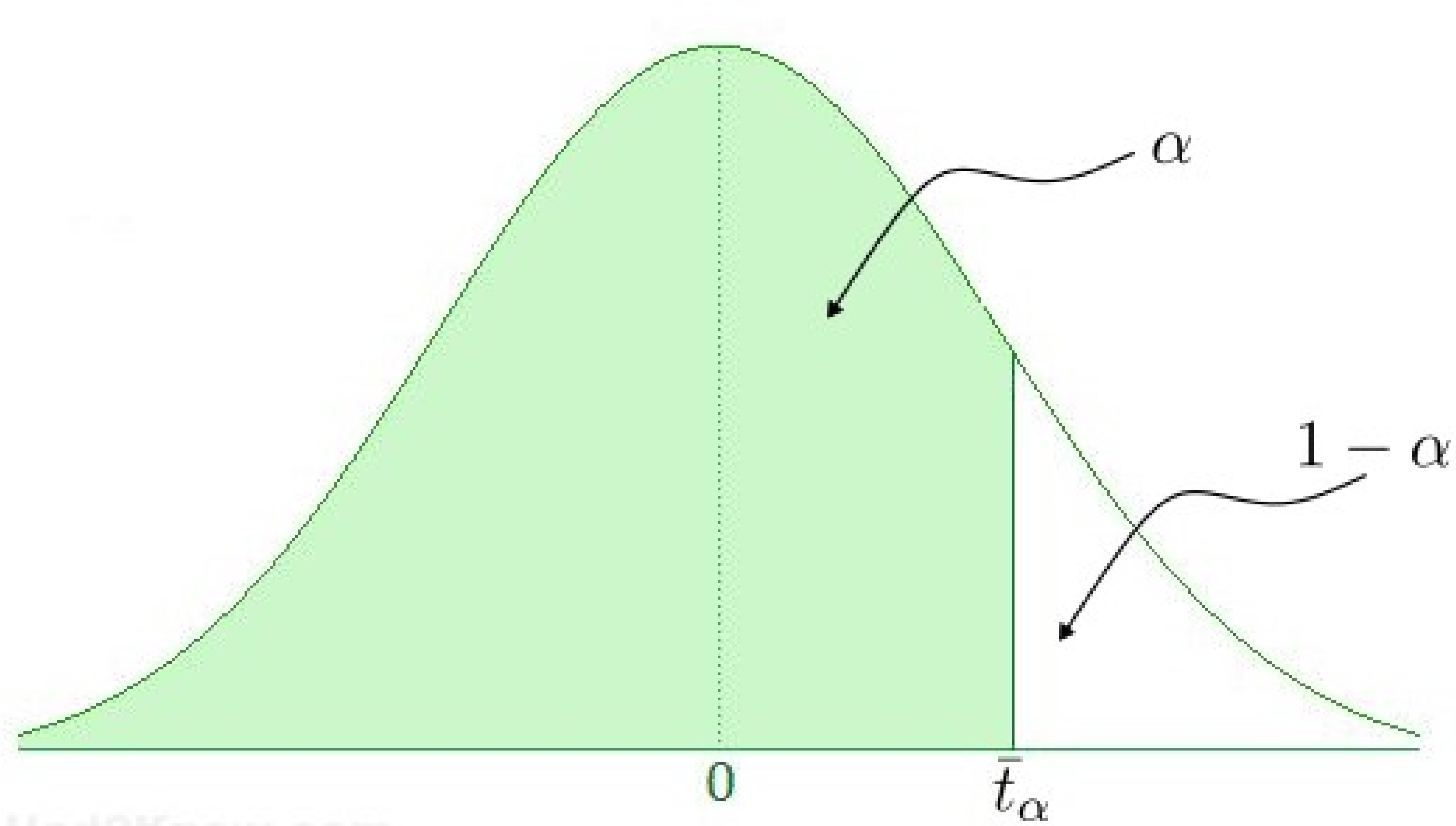
Evaluating I

-$\bar{t}_\alpha \leq t \leq \bar{t}_\alpha$: $P[-\bar{t}_\alpha \leq t \leq \bar{t}_\alpha] = \alpha$

$$
\begin{aligned}
P[-\bar{t}_\alpha \leq t \leq \bar{t}_\alpha] &= P[-\bar{t}_\alpha \leq \frac{\widehat{\theta} - \theta}{\sigma} \leq \bar{t}_\alpha] \\
&= P[-\bar{t}_\alpha \sigma \leq (\widehat{\theta} - \theta) \leq \bar{t}_\alpha \sigma] \\
&= P[\widehat{\theta} - \bar{t}_\alpha \sigma \leq \theta \leq \widehat{\theta} + \bar{t}_\alpha \sigma] = \alpha
\end{aligned}
$$

$$I = \{x \in \mathbb{R} \mid \widehat{\theta} - \bar{t}_\alpha \sigma \leq x \leq \widehat{\theta} + \bar{t}_\alpha \sigma\}$$

Quantile



Is not the confidence level we previously denoted

Percentage points of Student-t

					Probability					
ν	.7500	.8000	.8500	.9000	.9500	.9750	.9900	.9950	.9990	.9995
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.22	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850

Percentage points (cont.)

	Probability									
ν	.7500	.8000	.8500	.9000	.9500	.9750	.9900	.9950	.9990	.9995
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.160	3.373
250	0.675	0.843	1.039	1.285	1.651	1.969	2.341	2.596	3.123	3.330
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
INF	0.675	0.842	1.036	1.282	1.645	1.960	2.327	2.576	3.091	3.291

Confidence interval

- The confidence interval with percentage probability 100(1-α)% is

$$\hat{\theta} - t_{\frac{\alpha}{2},n}\,\hat{\sigma} \leqslant \theta \leqslant \hat{\theta} + t_{\frac{\alpha}{2},n}\,\hat{\sigma}$$

where $t_{\frac{\alpha}{2},n}$ is the 1-$\alpha/2$ percentile of Student-t distribution with n freedom degrees

Confidence interval

- When *R* independent values are known, estimation can be computed as their mean

$$\bar{Y} = \sum_{i=1}^{R} \bar{Y}_i$$

- $\bar{Y}$ has an error with respect to the real value
- If Y_i are independent and equally distributed, estimator is unbiased
- The measure of the error is called *confidence interval*

Error estimation

- Sample variance over *R* measure

$$S^2 = \frac{1}{R-1} \sum_{i=1}^{R} (Y_i - \bar{Y})^2$$

- The Y_i values are assumed gaussian distributed, thus

Half sized confidence interval width

$$\bar{Y} \pm t_{\frac{\alpha}{2}, R-1} \frac{S}{\sqrt{R}} \Rightarrow \bar{Y} \pm t_{\frac{\alpha}{2}, R-1} \sqrt{\frac{1}{(R-1)R} \sum_{i=1}^{R} (Y_i - \bar{Y})^2}$$

Error estimation (2)

- $t_{\alpha/2,R-1}$ is the quantile of the *t-student* distribution with *R-1* freedom degrees such that an area equal to $\alpha/2$ is cut-off by each tails
- Confidence interval limit the error with *100(1-α)%* probability
- *(1-α)100%* is said ***confidence***

Example

- Let us suppose we want to compute the expected cycle time of a production cycle
- Computed estimation after $R=120$ simulations is $5.80\ h$, and its standard deviation is $S=1.60\ h$
- Which is the *95%* confidence interval?

Example (cont.)

- $\alpha = \frac{(100-95)}{100} = \frac{5}{100} = 0.05$
- $\frac{\alpha}{2} = 0.025 \Rightarrow 1 - \frac{\alpha}{2} = 0.975$
- $t_{\frac{\alpha}{2}} = 1.98$ is derived by the table
- The mean cycle time interval

$$5.80 \pm 1.98 \frac{1.60}{\sqrt{120}} = 5.80 \pm 0.29\, h$$

Terminating simulation

- Simulated time from 0 to T_E
- The i-th simulation cycle gives n_i samples of the measure
- The samples could not be
  - independent
  - identically distributed

Output data

$$
\text{R cycles}\left\{\begin{matrix} Y_{11} & Y_{12} & \cdots & Y_{1n_1} \\ Y_{21} & Y_{22} & \cdots & Y_{2n_2} \\ \vdots & \vdots & \cdots & \vdots \\ Y_{R1} & Y_{R2} & \cdots & Y_{Rn_R} \end{matrix}\right. \qquad \begin{matrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_R \end{matrix}
$$

$$\bar{Y}$$

Measure estimation

Confidence interval

Given the confidence level (1-α), all the quantities are computed

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_j$$

$$\bar{Y} = \frac{1}{R} \sum_{i=1}^{R} \bar{Y}_i$$

$$S^2 = \frac{1}{R-1} \sum_{i=1}^{R} \left(\bar{Y}_i - \bar{Y}\right)^2$$

$$H = t_{\frac{\alpha}{2}, R-1} \frac{S}{\sqrt{R}}$$

100(1-α)% confidence interval is

$$\bar{Y} - H \leqslant \theta \leqslant \bar{Y} + H$$

Continuous data output

- When the estimator is an integral

$$
\begin{array}{lll}
Y_1(t) & 0 \leqslant t \leqslant T_{E_1} & \bar{Y}_1 \\
Y_2(t) & 0 \leqslant t \leqslant T_{E_2} & \bar{Y}_2 \\
\vdots & \vdots & \vdots \\
Y_{R1}(t) & 0 \leqslant t \leqslant T_{E_R} & \bar{Y}_R \\
 & & \bar{Y}
\end{array}
$$

$$
\bar{Y}_i = \frac{1}{T_{E_i}} \int_0^{T_{E_i}} Y_i(t)\, dt
$$

Specified precision

- When precision is specified (ε) with high probability

$$P\left(\left|\bar{Y}-\theta\right|<\varepsilon\right)\geqslant 1-\alpha$$

- R_0 independent replications ($R_0 \geq 2$, *generally* $R_0>10$)
- Estimation of $R \geq R_0$

$$H=t_{\frac{\alpha}{2},R-1}\frac{S_0}{\sqrt{R}}\leqslant\varepsilon \quad\Rightarrow\quad R\geq\left(t_{\frac{\alpha}{2},R-1}\frac{S_0}{\varepsilon}\right)^2$$

Specified precision (2)

$$t_{\frac{\alpha}{2},R-1} \geqslant z_{\frac{\alpha}{2}} \Rightarrow R \geqslant \left(z_{\frac{\alpha}{2}} \cdot \frac{S_0}{\epsilon} \right)^2$$

Where $z_{\alpha/2}$ is the *100(1-α/2)%* quantile of the standard normal distribution

1. Compute *R*
2. Collect *(R-R₀)* samples
3. Compute the confidence interval
4. If it is greater than ε go to step 1.

Steady-state simulation

- Given a sequence with n samples $\{Y_1, Y_2, \ldots, Y_n\}$ steady-state measure is estimated as

$$\theta = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} Y_i$$

- Example: Y_i = permanence time of i-th process into a system => θ is the mean permanence time of a generic process

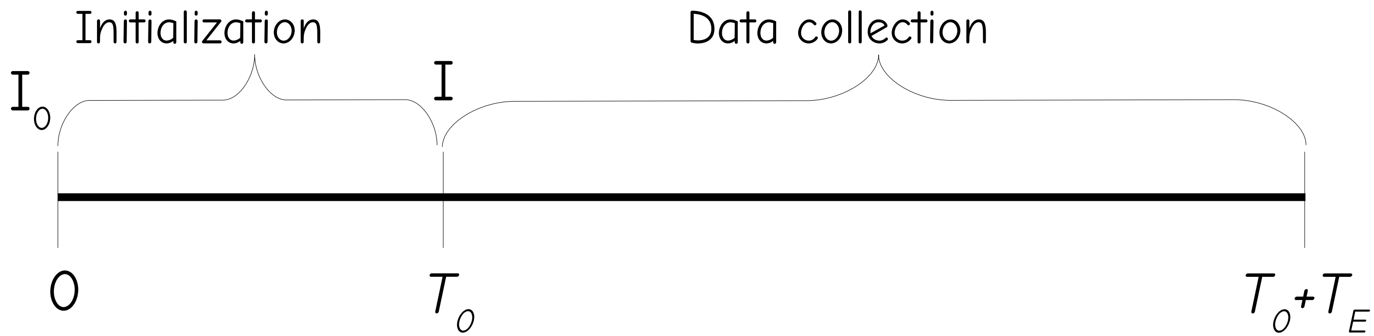
Features

- Determine n or T_E
- Identical distributed data are needed
- Initialize simulation in a state near the long-run conditions
  - Simplified models

Initialization

- Simulation is split into two phases
  - *initialization*: no data are collected up to time T_0
  - Data are collected in time interval $[T_0 , T_0+T_E]$
- T_0 has to be chosen

Establishing T_0



- State I is a r.v.
- Statistical methods => difficult to be applied
- Empirical methods => less general

Independent replications

- Deviation due to the transitory is negligible
- Over *n* observations, the first *d* are not considerate ($[0, T_0]$)
- The estimation *i* is

$$\bar{Y}_{i,d_i} = \frac{1}{n_i - d_i} \sum_{j=d+1}^{n_i} Y_j$$

T_E length

- Well established empirical rules
  - Time length after T_0 is such that the collected data are *10* times the erased data

$$n-d>10\,d\,,\quad T_E\geqslant 10\,T_0$$

  - $R \leq 25$

Measure estimation

- Given a confidence of *100(1-α)%* error is to large
  - Replications are increased
  - *T_E is increased*
- New replications are estimated *(R-R_0)*
- T_E is proportionally increased

Independent replications (2)

- Given the $\bar{Y}_{i,d_i}$ measures, same method for computing confidence intervals in the case of terminating simulations is used
- Replication length beyond T_0 has to be ten times T_0 at least

$$(n_i - d_i) \geqslant 10\, d_i \Rightarrow T_E \geq 10\, T_0$$

- Greatest possible values of R, but $R \leq 25$

Specified precision

- How to obtain a specified precision ε?
- Increasing the number of replicas R
- Increasing the time T_E

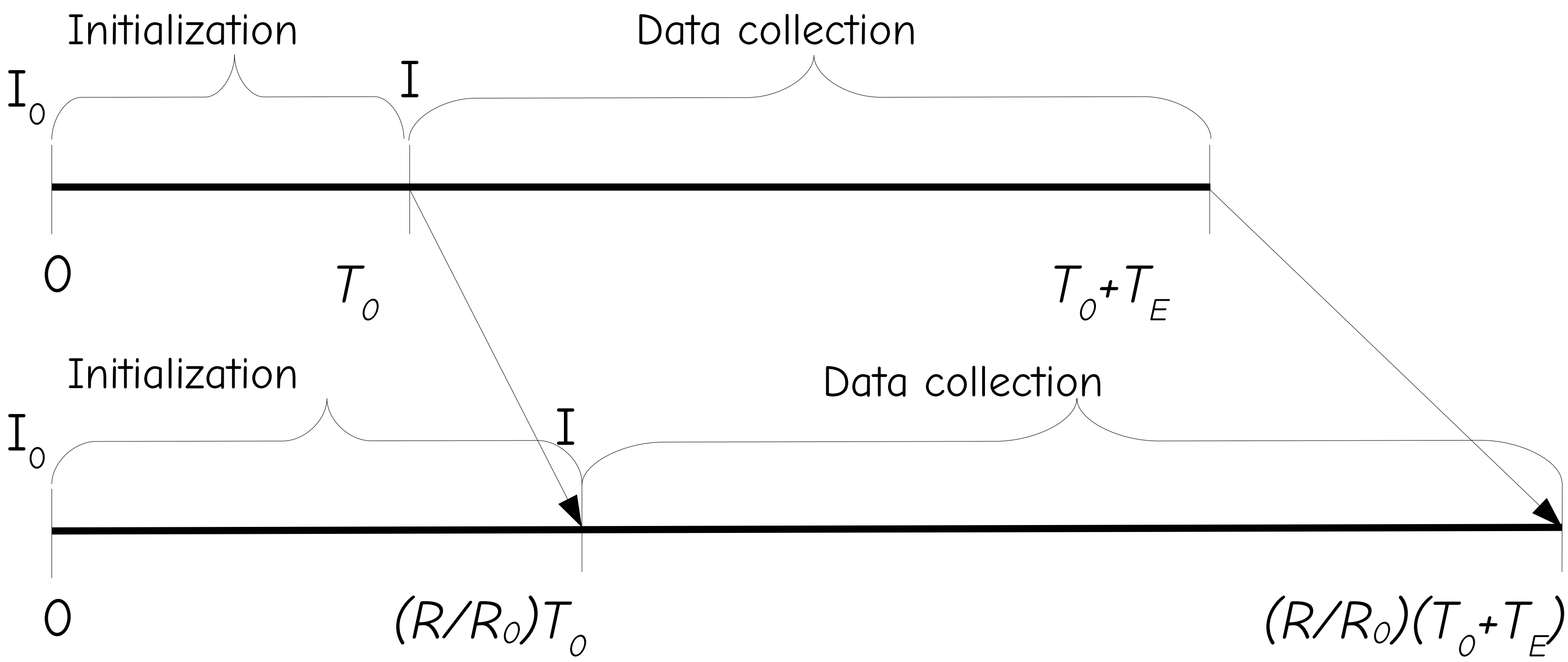
Collecting data

- Replications are increased of R-R_0
- Simulated time T_E-T_0 is increased proportionally to R/R_0
- New length: $(R/R_0)(T_0+T_E)$
- Erase data from 0 to $(R/R_0)T_0$

Observations

- Advantages:
  - reduce due to the transitory
- Disadvantage:
  - the state at T_0+T_E must be stored
  - simulation has to restart for the increased time

Observation period



Deviation is more sensitive to the initial state than the number of replications

The book

- Chapter 1
  - Section 1.11
- Chapter 3
  - Sections 3.1, 3.2
- Chapter 7
  - Section 7.1, 7.2, 7.3
- Chapter 11
  - Sections 11.1, 11.2, 11.3, 11.4 (without 11.4.3 and 11.4.4), 11.5