Eberhard O. Voit

A First Course in Systems Biology

Chapter 5
Parameter Estimation

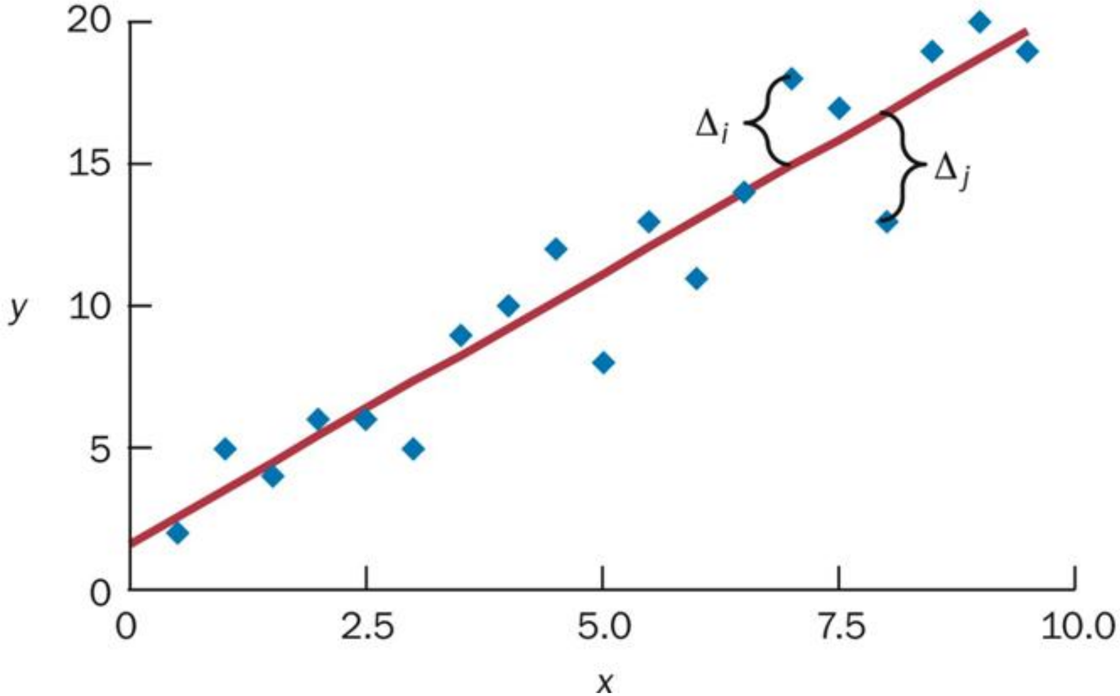
Parameter Estimation

Parameter estimation refers to the process of using sample data (in reliability engineering, usually times-to-failure or success data) to estimate the parameters of the selected distribution.

Linear regression

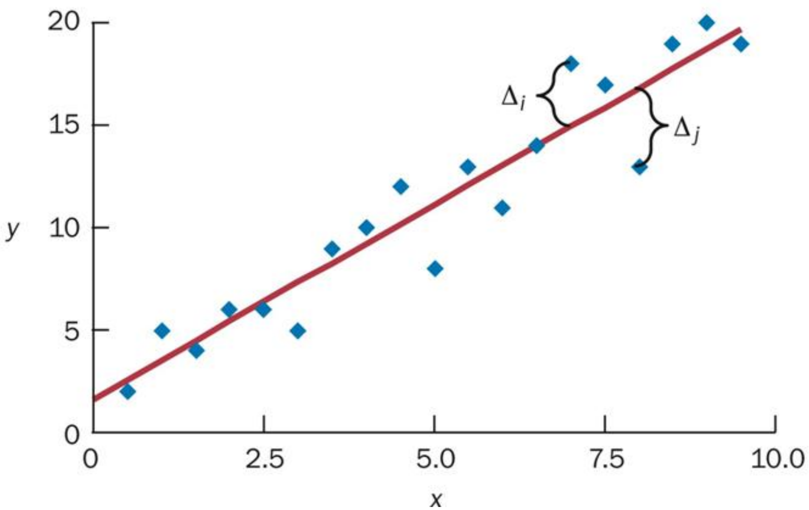
Regression analysis is a set of statistical processes for estimating the relationships among variables.

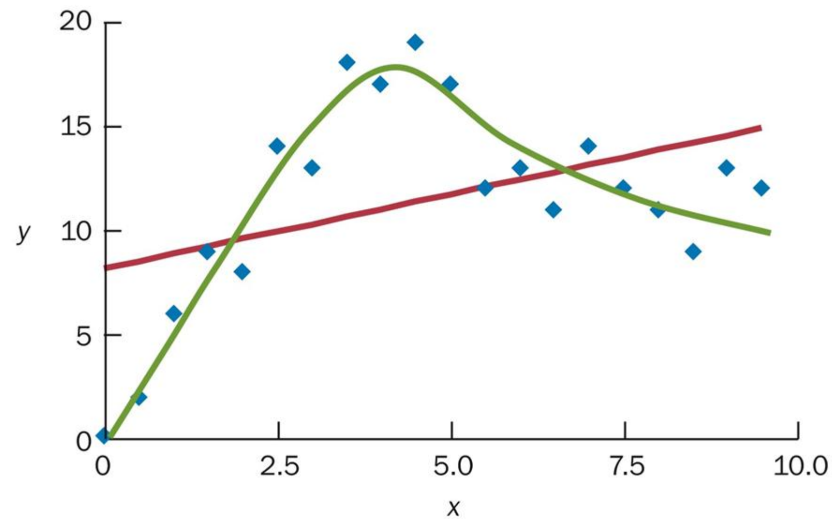
Linear regression is a linear approach for modeling the relationship between a scalar dependent variable y and one or more independent variables denoted **x**.



y = 1.89x + 1.63

Minimize sum of squared errors (SSE) through least-square methods.





Correct use vs Wrong use

Linear regression with two variables

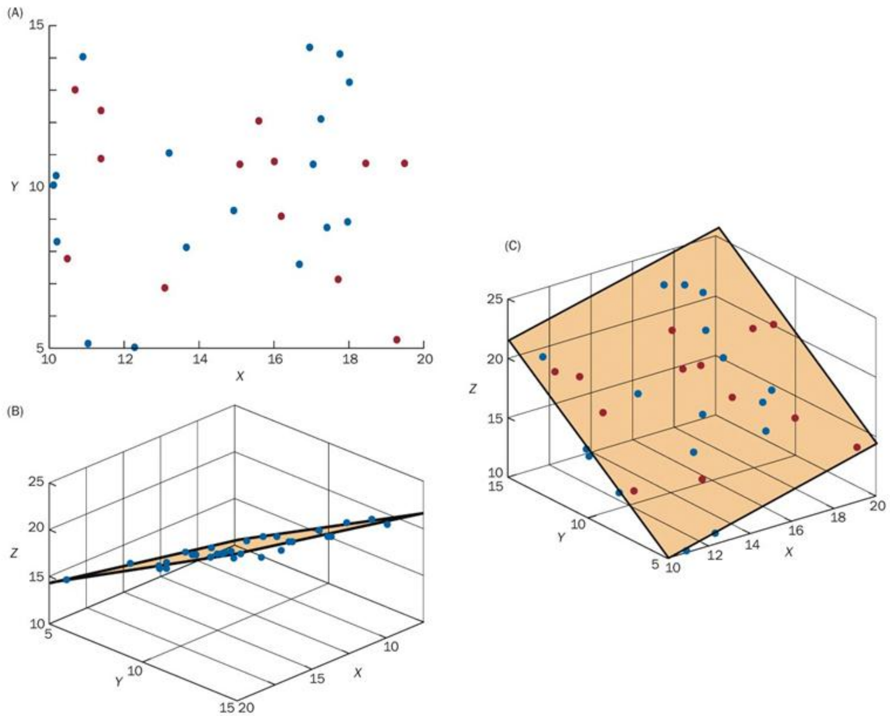
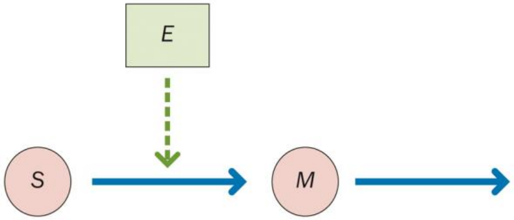


TABLE 5.1: DATA FOR MULTIPLE LINEAR REGRESSION AND CORRESPONDING VALUES OF THE REGRESSION PLANE			
X	***Y***	***Z* (data)**	***Z* (regression)**
17.43	8.72	16.73	17.38
10.49	7.75	13.58	13.28
16.69	7.58	15.47	15.77
17.06	10.69	19.32	19.46
13.10	6.85	13.82	13.39
18.46	10.71	21.06	20.07
17.27	12.10	20.55	21.13
16.04	10.77	19.15	19.10
14.95	9.25	16.68	16.90
17.74	7.11	16.30	15.70
17.98	8.89	17.38	17.82
13.67	8.12	14.89	15.07
15.09	10.68	19.42	18.58
10.22	8.28	13.27	13.75
11.04	5.12	10.08	10.54
11.39	12.34	19.60	18.85
11.40	10.87	17.57	17.19
10.71	12.98	19.93	19.28
10.20	10.33	15.45	16.06
12.30	5.02	10.95	10.98
16.97	14.32	22.96	23.51
17.79	14.10	22.68	23.62
13.22	11.05	18.06	18.19
10.91	14.01	20.35	20.53
19.30	5.26	14.31	14.29
15.60	12.04	21.42	20.34
18.02	13.23	22.52	22.73
10.12	10.05	15.15	15.71
19.50	10.73	20.77	20.55
16.22	9.07	17.56	17.25

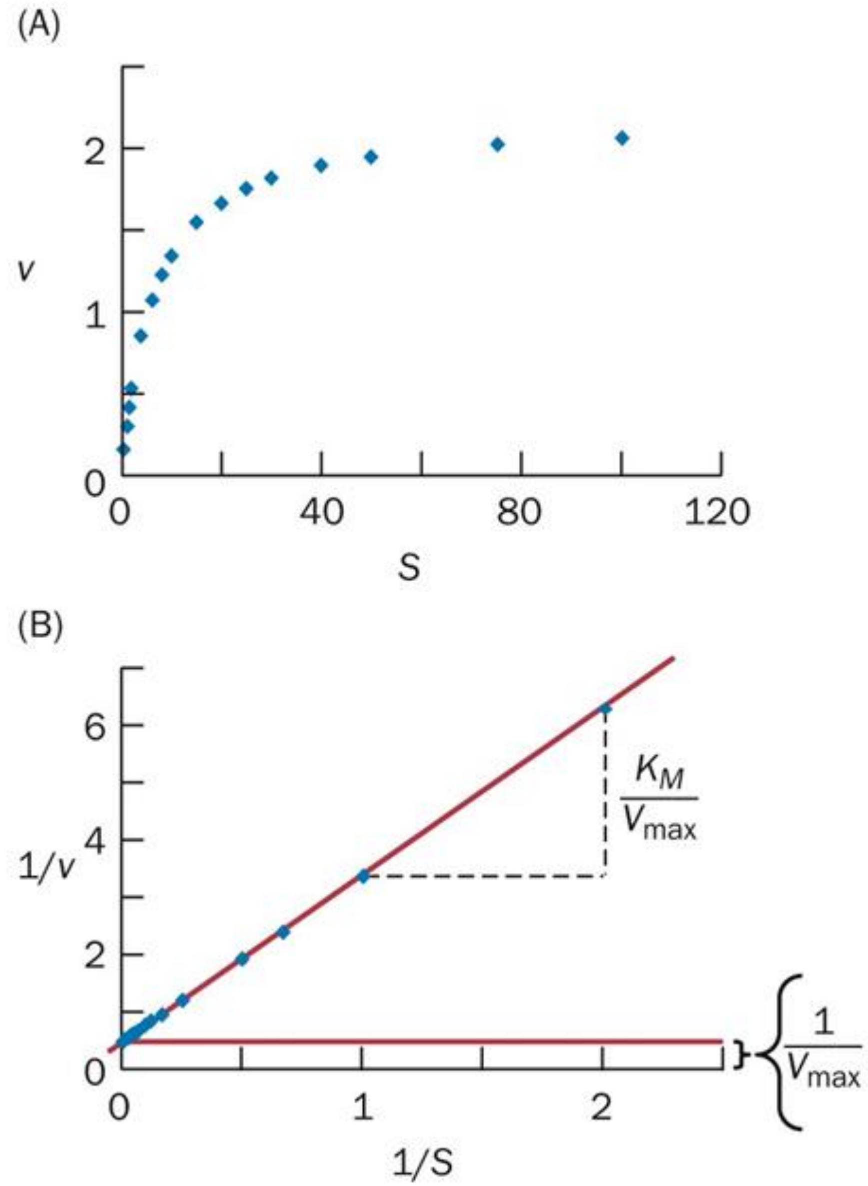
Table 5-1 A First Course in Systems Biology (© Garland Science 2013)

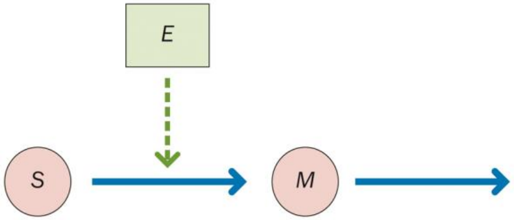


Michaelis-Menten equation

$$\frac{dM}{dt} = V_{max}\frac{S}{K_M + S} - cM$$

K_M and V_{max} can be estimated using linear regression

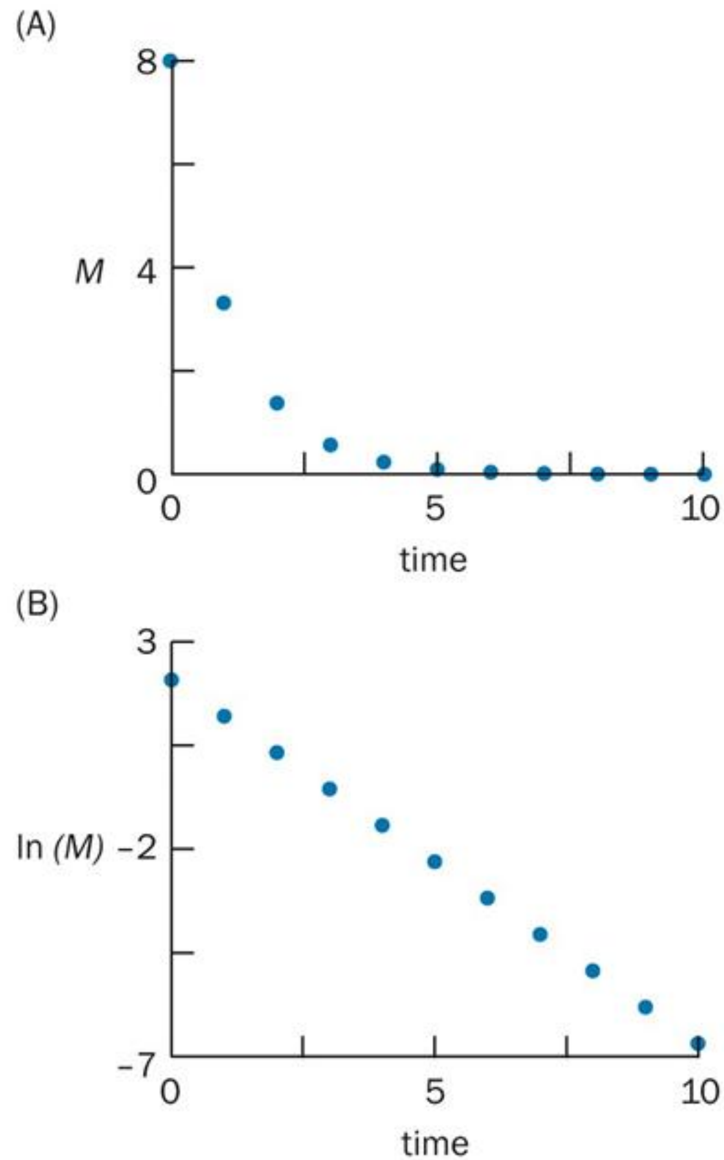




Michaelis-Menten equation

$$\frac{dM}{dt} = V_{max}\frac{S}{K_M + S} - cM$$

c can be estimated using separate experiment



Parameter estimation in non-linear systems

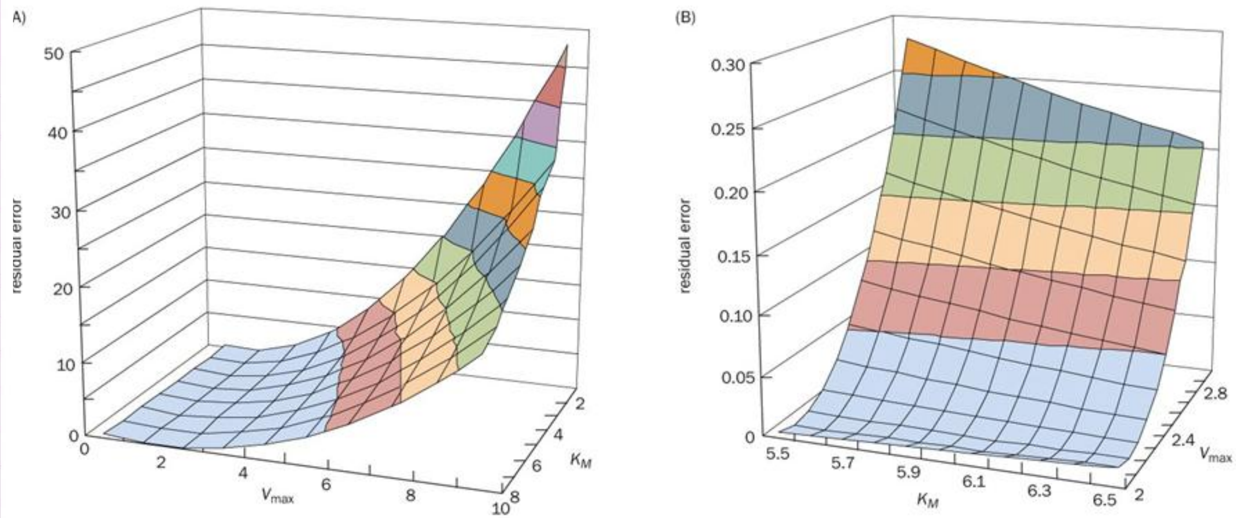
Three classes of search algorithms:

1. Exhaustive search
2. Gradient search (steepest descent/hill-climb)
3. Evolutionary algorithms

Comprehensive grid search

TABLE 5.2: MEASUREMENTS OF REACTION SPEED v VERSUS SUBSTRATE CONCENTRATION S

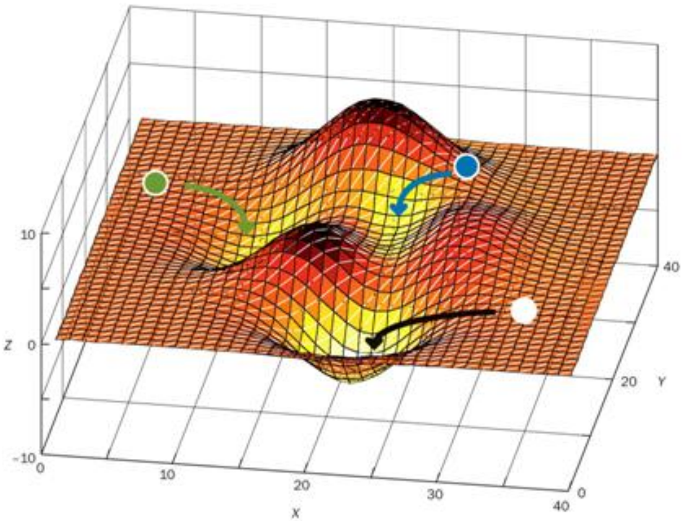
S	v
0.5	0.16
1	0.30
1.5	0.42
2	0.52
4	0.85
6	1.06
8	1.22
10	1.34
15	1.54
20	1.67
25	1.75
30	1.81
40	1.90
50	1.95
75	2.03
100	2.07

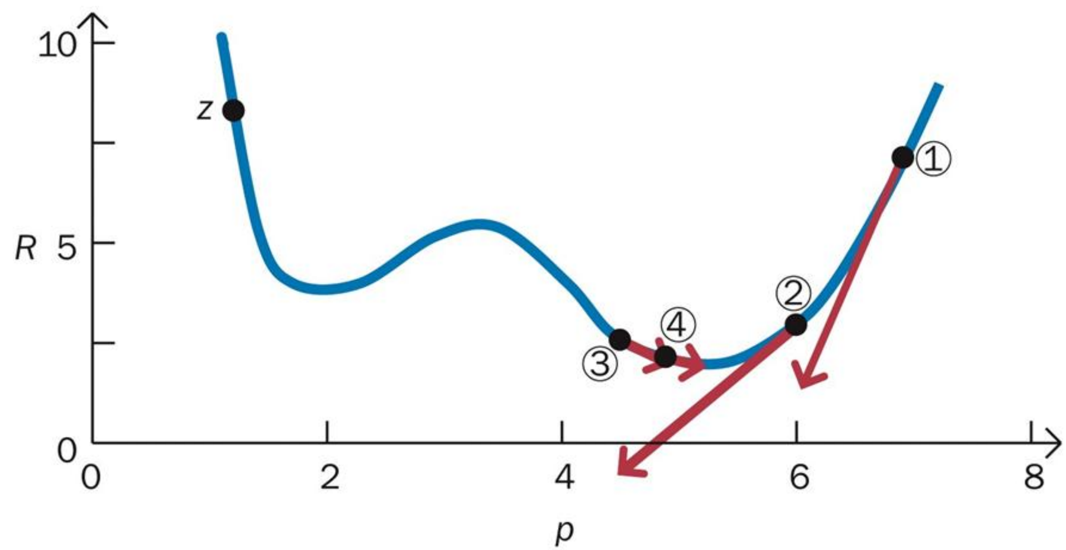


Residual errors (SSEs) in a grid search

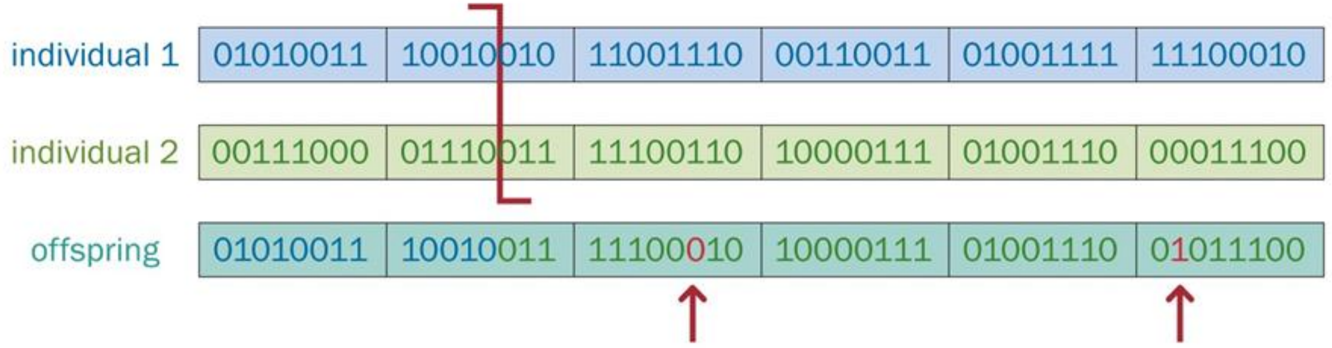
Heavy computation is a serious problem for finer search. Sophesticated approaches such as **Latin hypercube sampling** and **branch-and-bound** methods could be useful.

Non-linear regression



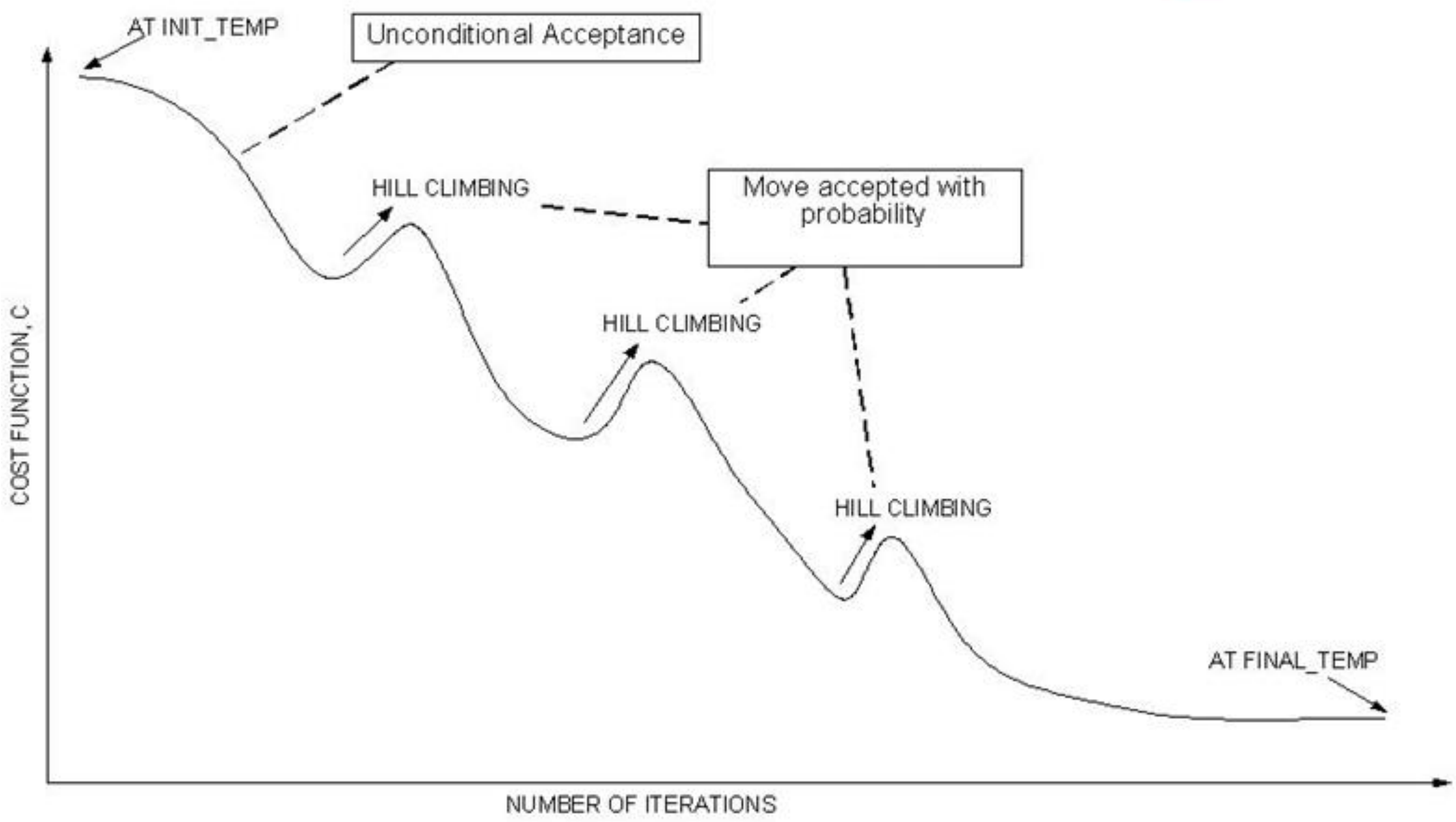


Genetic Algorithm



Other Stochastic Algorithms

Simulated Annealing

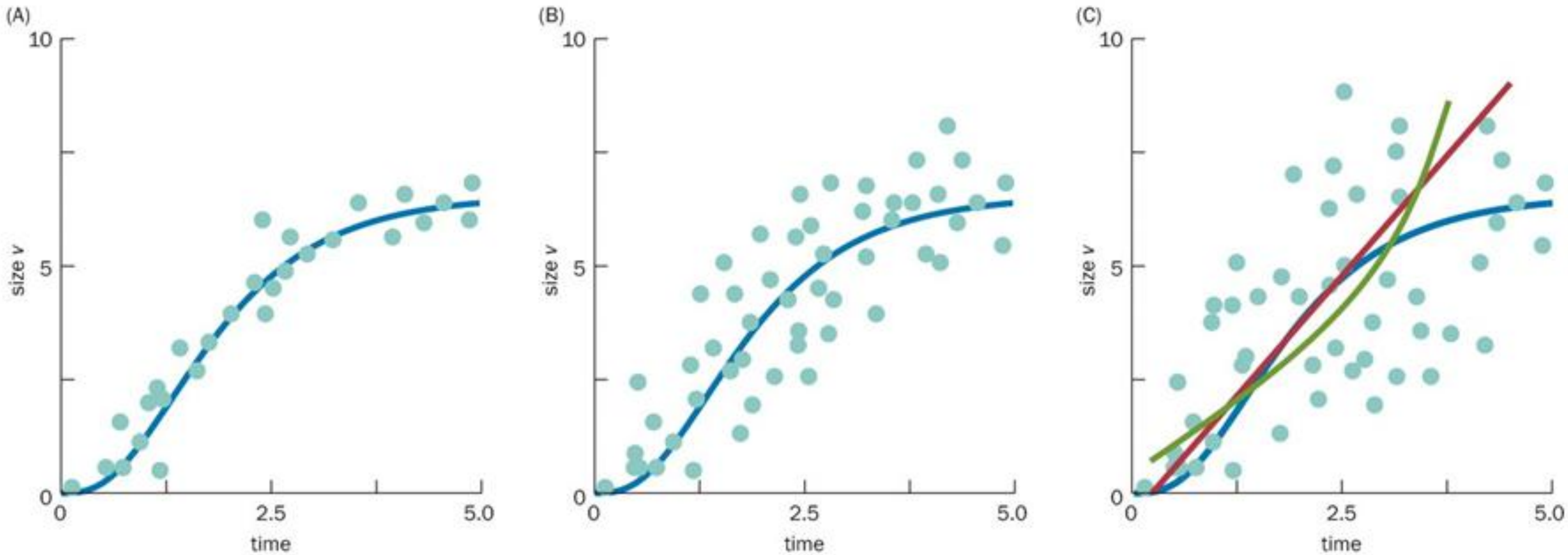


- At a fixed temperature *T*:
- Perturb (randomly) the current state to a new state
- ΔE is the difference in energy between current and new state
- If $\Delta E < 0$ (new state is lower), accept new state as current state
- If $\Delta E \geq 0$, accept new state with probability

  $$Pr\ (accepted) = exp\ (- \Delta E / k_B . T)$$
- Eventually the systems evolves into thermal equilibrium at temperature *T*; then the formula mentioned before holds
- When equilibrium is reached, temperature *T* can be lowered and the process can be repeated

Challenges:

1. Noise



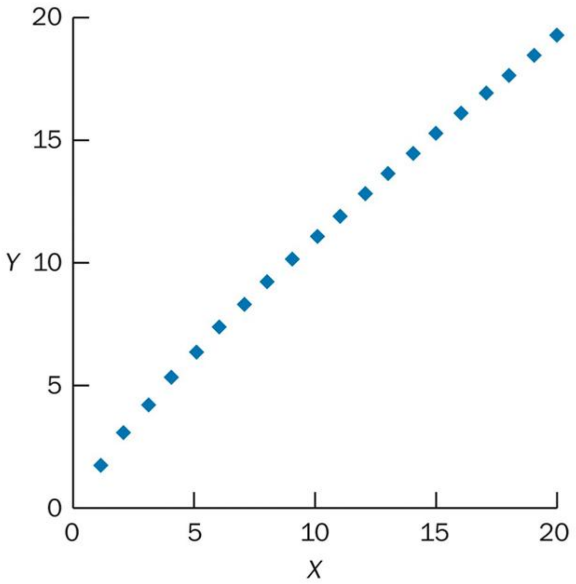
Challenges:

2. Multiple solutions for same SSE

Example 1:

$p_1 p_2 = 12$ with $p_1 = 3$, $p_2 = 4$

Example 2:



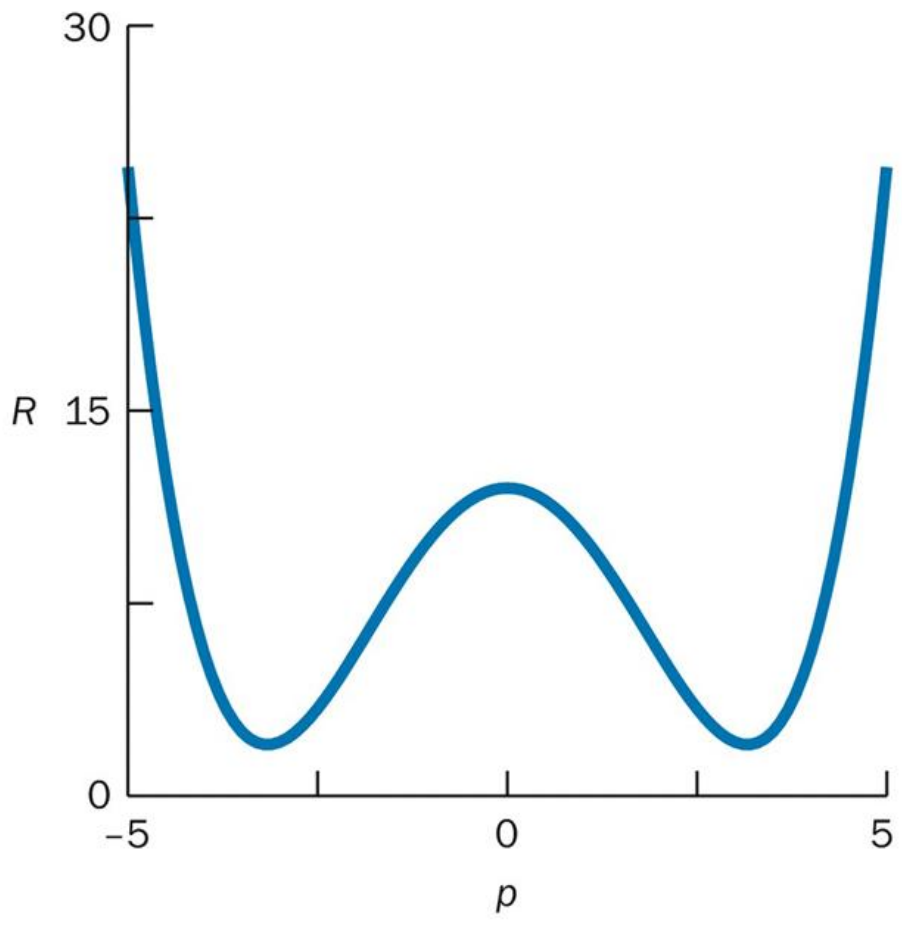
Consider $f = pX^a Y^b$

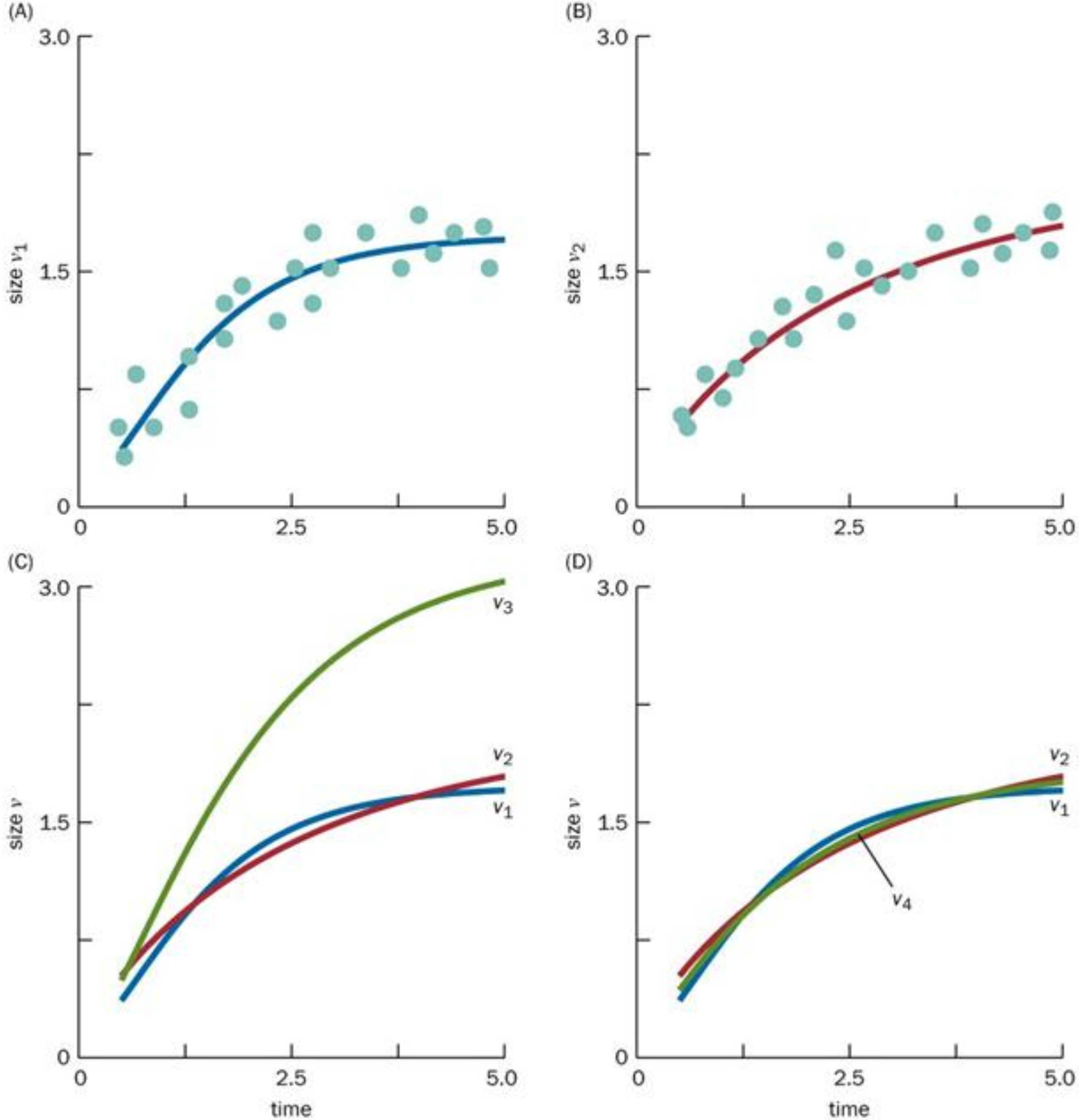
If $Y = \alpha X^{\gamma}$, then the following two solutions give identical results with $\alpha = 1.75$ and $\gamma = 0.8$:

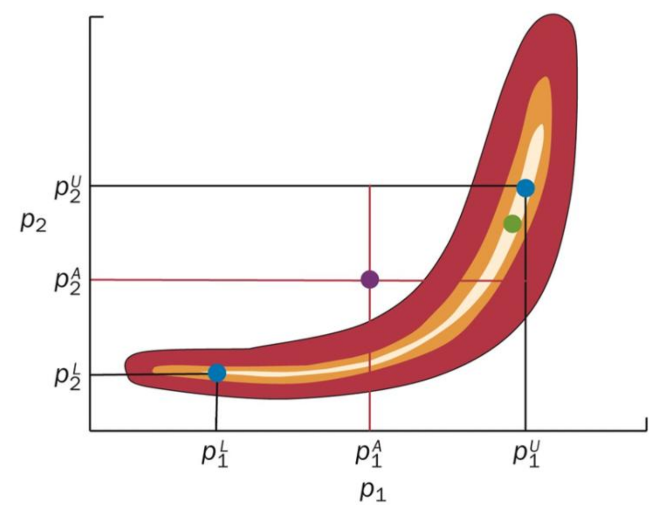
(2.45, 1.2, -0.3) and (4.2875, 2, -1.3)

Challenges:

3. Sloppiness




(A)
(B)
(C)
(D)
size v_1
size v_2
size v
time
3.0
1.5
0
2.5
5.0
v_1
v_2
v_3
v_4


p_2
p_2^U
p_2^A
p_2^L
p_1^L
p_1^A
p_1^U
p_1

Parameter Estimation for Systems of Differential Equations

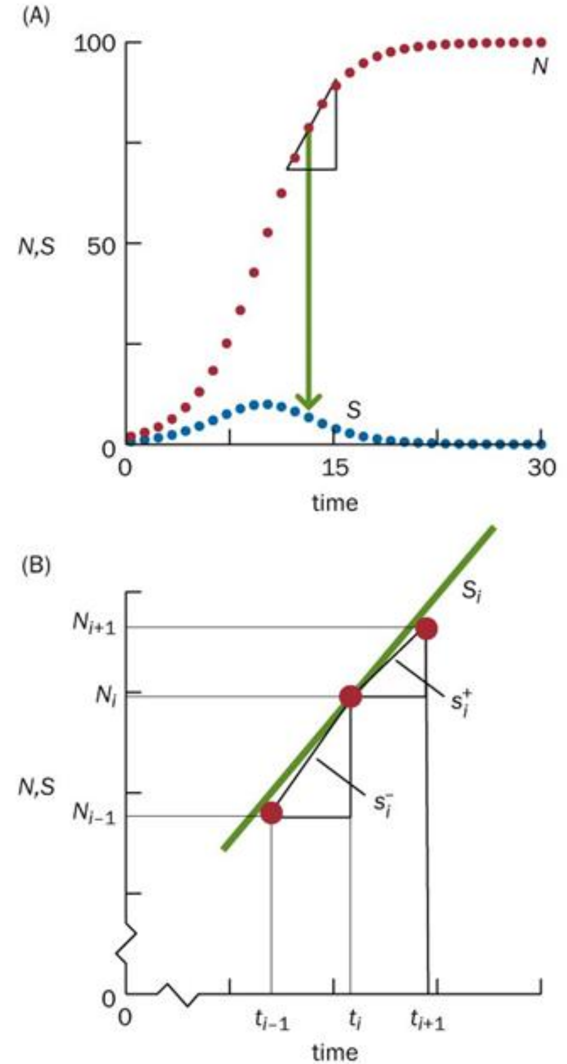
Main reasons for the estimation is harder:

1. Usually contain more parameters than individual functions
2. Necessity for comparing time-dependent experimental data with solutions of differential equations, requiring lot of computation.

Parameter Estimation for Systems of Differential Equations

One solution: Avoid computation (related to integration)!

$$\dot{N} = aN - bN^2$$



$$\dot{N} = aN - bN^2$$

$1.80 = 4.34a - 4.34b^2$
$3.51 = 9.18a - 9.18b^2$

$a = 0.3899$ ($\sim a = 0.4$)
$b = 0.0039$ ($\sim b = 0.004$)

TABLE 5.6: DATASET USED FOR ESTIMATING THE PARAMETERS IN (5.12)*

t	N	S (true)	S (3 – point)
0	2	0.78†	
2	4.34	1.66	1.80
4	9.18	3.33	3.51
6	18.36	6.00	6.05
8	33.36	8.89	8.58
10	52.70	9.97	9.47
12	71.26	8.19	7.99
14	84.66	5.19	5.30
16	92.47	2.78	2.95
18	96.47	1.36	1.48
20	98.38	0.64	0.70
22	99.27	0.29	0.32
24	99.67	0.13	0.15
26	99.85	0.06	0.07
28	99.93	0.03	0.03
30	99.97	0.01†	

* The data consist of the numbers of bacteria (in units of millions), the true slopes, which are usually not known, and the slopes estimated with the three-point method.

† Slopes at times $t = 0$ and $t = 30$ cannot be obtained with this method.