

# Estimation of Parameters from Data

Ross Swaney\* and James B. Rawlings<sup>†</sup>  
Department of Chemical & Biological Engineering  
University of Wisconsin-Madison  
Madison, WI 53706

September 1, 2014

Copyright © 2014 by Ross E. Swaney and James B. Rawlings

## 1 Introduction

The purpose of process modeling is to predict the behavior of chemical and physical phenomena. More often than not, it is not practical to predict the behaviors involved with *ab initio* models. Instead, most process models employ phenomenological models to represent the complex underlying physical and molecular behavior. These are compact models giving the required characteristics in a manageable form.

Typical examples of models for intrinsic behavior include those for thermodynamic properties, transport properties, and chemical kinetic rates. Models are also often used to embody empirical solutions to complex flow situations, for example the models of volume-integrated behavior used to predict pressure drops and heat transfer coefficients.

These models by nature rely on empirical data. They involve two components: (1) a model form, often an algebraic relation, containing one or more adjustable constants referred to as “parameters”, and (2) the values to be used for these parameters. The parameter values are adjusted so that the model predictions match the available empirical data. Such data-based models are sometimes also referred to as “correlations”.

The sections below discuss methods for determining the unknown model parameters to obtain a good fit between the model predictions and the set of available data. The first section addresses basic model fitting without statistical arguments. In many cases the quantity of data available is limited, and the model fitting process relies upon judgment and physical intuition. When larger quantities of data are available, statistical methods become useful. These techniques introduce probability arguments to analyze the magnitudes of the errors, assess models, and predict probable ranges of the parameters.

---

\*swaney@engr.wisc.edu

<sup>†</sup>rawlings@engr.wisc.edu

## 2 Random Variables and Probability

Let  $x$  be a random variable taking real values and the function  $F(x)$  denote the **probability distribution function** of the random variable so that

$$F(a) = \Pr(x \leq a)$$

i.e.  $F(x)$  at  $x = a$  is the probability that the random variable  $x$  takes on a value less than or equal to  $a$ .

We next define the **probability density function**, denoted  $p(x)$ , such that

$$F(x) = \int_{-\infty}^x p(s) ds, \quad -\infty < x < \infty \quad (1)$$

Also, we can define the density function for discrete (integer valued) as well as continuous random variables. Alternatively, we can replace the integral in Equation 1 with a sum over an integer-valued function. The random variable may be a coin toss or a dice game, which takes on values from a discrete set contrasted to a temperature or concentration measurement, which takes on a values from a continuous set. The density function has the following properties

$$p(x) \geq 0 \quad \text{all } x$$

$$\int_{-\infty}^{\infty} p(x) dx = 1$$

and the interpretation in terms of probability

$$\Pr(x_1 \leq x \leq x_2) = \int_{x_1}^{x_2} p(x) dx$$

### 2.1 Mean and Variance

The **mean** or **expectation** of a random variable  $x$  is defined as

$$\bar{x} = \int_{-\infty}^{\infty} xp(x) dx \quad (2)$$

The **variance** is defined as

$$\text{var}(x) = \int_{-\infty}^{\infty} (x - \bar{x})^2 p(x) dx$$

The standard deviation is the square root of the variance

$$\sigma(x) = \sqrt{\text{var}(x)}$$

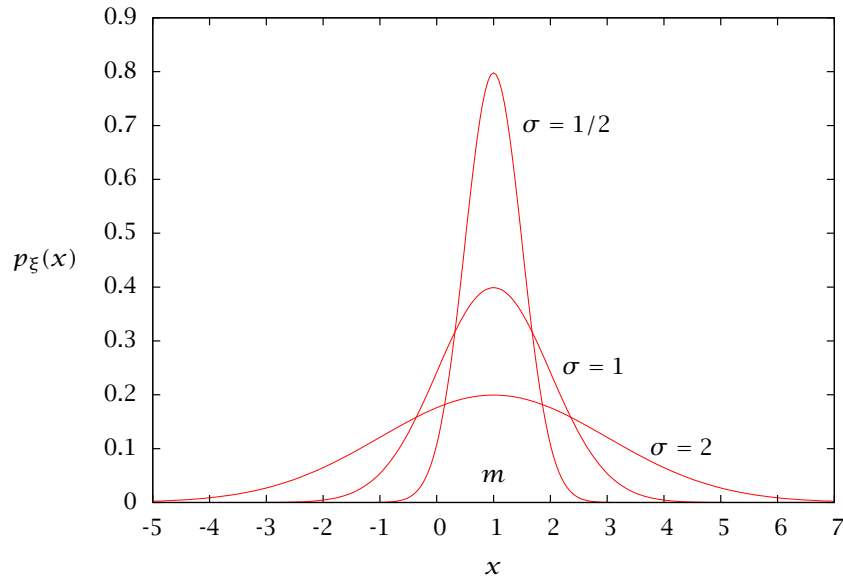


Figure 1: Normal distribution,  $p_{\xi}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(x-m)^2}{\sigma^2}\right)$ . Mean is one and standard deviations are 1/2, 1 and 2.

## 2.2 Review of the Normal Distribution

Probability and statistics provide one useful set of tools to model the uncertainty in experimental data. It is appropriate to start with a brief review of the normal distribution, which plays a central role in analyzing data. The normal or Gaussian distribution is ubiquitous in applications. It is characterized by its mean,  $m$ , and variance,  $\sigma^2$ , and is given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(x-m)^2}{\sigma^2}\right) \quad (3)$$

Figure 1 shows the univariate normal with mean zero and unit variance. We adopt the following notation to write Equation 3 more compactly

$$x \sim N(m, \sigma^2)$$

which is read “the random variable  $x$  is distributed as a normal with mean  $m$  and variance  $\sigma^2$ .” Equivalently, the probability density  $p(x)$  for random variable  $x$  is given by Equation 3.

For distributions in more than one variable, we let  $\mathbf{x}$  be an  $n_p$ -vector and the generalization of the normal is

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n_p/2} |\mathbf{P}|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x}-\mathbf{m})' \mathbf{P}^{-1} (\mathbf{x}-\mathbf{m})\right]$$

$$p(\mathbf{x}) = \exp\left(-1/2\left(3.5x_1^2 + 2(2.5)x_1x_2 + 4.0x_2^2\right)\right)$$

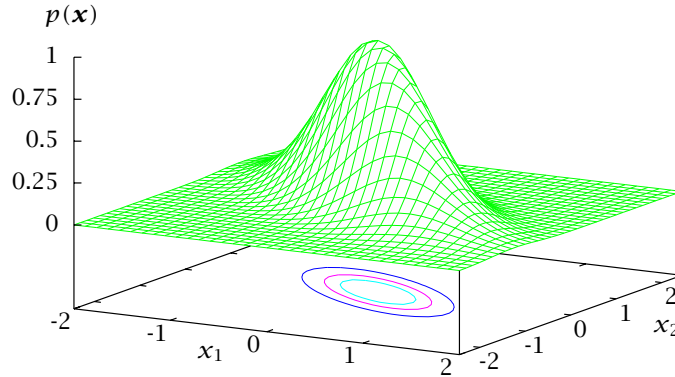


Figure 2: Multivariate normal for  $n = 2$ .

in which the  $n_p$ -vector  $\mathbf{m}$  is the mean and the  $n_p \times n_p$ -matrix  $\mathbf{P}$  is called the covariance matrix. The notation  $|\mathbf{P}|$  denotes determinant of  $\mathbf{P}$ . We also can write for the random variable  $\mathbf{x}$  vector

$$\mathbf{x} \sim N(\mathbf{m}, \mathbf{P})$$

The matrix  $\mathbf{P}$  is a real, *symmetric* matrix. Figure 2 displays a multivariate normal for

$$\mathbf{P}^{-1} = \begin{bmatrix} 3.5 & 2.5 \\ 2.5 & 4.0 \end{bmatrix}$$

As displayed in Figure 2, lines of constant probability in the multivariate normal are lines of constant

$$(\mathbf{x} - \mathbf{m})' \mathbf{P}^{-1} (\mathbf{x} - \mathbf{m})$$

To understand the geometry of lines of constant probability (ellipses in two dimensions, ellipsoids or hyperellipsoids in three or more dimensions) we examine the eigenvalues and eigenvectors of the  $\mathbf{P}$  matrix.

### 2.3 Eigenvalues and Eigenvectors

An eigenvector of a matrix  $\mathbf{A}$  is a nonzero vector  $\mathbf{v}$  such that when multiplied by  $\mathbf{A}$ , the resulting vector points in the same direction as  $\mathbf{v}$ , and only its magnitude is rescaled. The

rescaling factor is known as the corresponding eigenvalue  $\lambda$ . Therefore the eigenvalues and eigenvectors satisfy the relation

$$A\mathbf{v} = \lambda\mathbf{v}, \quad \mathbf{v} \neq \mathbf{0}$$

We normalize the eigenvectors so

$$\mathbf{v}'\mathbf{v} = \sum_i v_i^2 = 1$$

The eigenvectors show us the orientation of the ellipse given by the normal distribution. Consider the ellipse in the two-dimensional  $\mathbf{x}$  coordinates given by the quadratic

$$\mathbf{x}'A\mathbf{x} = b$$

If we march along a vector  $\mathbf{x}$  pointing in the eigenvector  $\mathbf{v}$  direction, we calculate how far we can go in this direction until we hit the ellipse  $\mathbf{x}'A\mathbf{x} = b$ . Substituting  $\alpha\mathbf{v}$  for  $\mathbf{x}$  in this expression yields

$$(\alpha\mathbf{v}')A(\alpha\mathbf{v}) = b$$

Using the fact that  $A\mathbf{v} = \lambda\mathbf{v}$  for the eigenvector gives

$$\alpha^2\lambda\mathbf{v}'\mathbf{v} = b$$

because the eigenvectors are of unit length, we solve for  $\alpha$  and obtain

$$\alpha = \sqrt{\frac{b}{\lambda}}$$

which is shown in Figure 3. Each eigenvector of  $A$  points along one of the axes of the ellipse. The eigenvalues show us how stretched the ellipse is in each eigenvector direction.

If we want to put simple bounds on the ellipse, then we draw a box around it as shown in Figure 3. Notice the box contains much more area than the corresponding ellipse and we have lost the correlation between the elements of  $\mathbf{x}$ . This loss of information means we can put different tangent ellipses of quite different shapes inside the same box. The size of the bounding box is given by

$$\text{length of } i\text{th side} = \sqrt{b\tilde{A}_{ii}}$$

in which

$$\tilde{A}_{ii} = (i, i) \text{ element of } A^{-1}$$

Figure 3 displays these results: the eigenvectors are aligned with the ellipse axes and the eigenvalues scale the lengths. The lengths of the sides of the box that is tangent to the ellipse are proportional to the square root of the diagonal elements of  $A^{-1}$ .

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = b$$

$$\mathbf{A} \mathbf{v}_i = \lambda_i \mathbf{v}_i$$

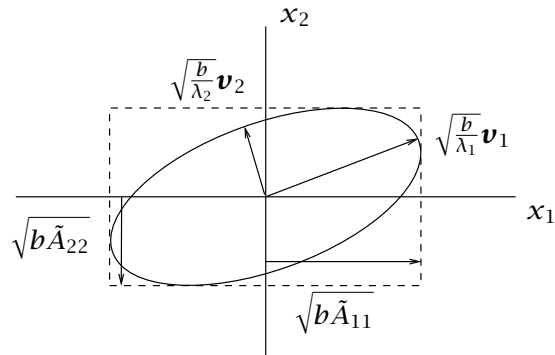


Figure 3: The geometry of quadratic form  $\mathbf{x}' \mathbf{A} \mathbf{x} = b$ .

## 2.4 Sampling random variables

Often we do not know the probability density of a random variable, but we have collected samples of the random variable. Given enough samples, we can sometimes approximate the probability density or any of the properties of the probability density such as the mean and variance. For example, the MATLAB function `rand` generates samples of a uniformly distributed random variable on the interval  $[0, 1]$ , and `randn` generates samples of a normally distributed random variable with unit variance and zero mean. The MATLAB function `hist` plots a histogram of the samples.

### Example 1: Histogram of samples from a normal density

Use MATLAB to generate 10,000 samples of a normally distributed random variable with zero mean and unit variance and plot the histogram of the samples. How does the histogram compare to Figure 1? Compare this result to the histogram with 10,000 samples.

### Solution

The two commands:

```
x = randn(10000,1);
hist(x,50)
```

generates the 10,000 samples and plots the histogram using 50 bins. The default is 10 bins if we leave off the second argument to `hist`. Type `help hist` and `help randn` to learn more about these functions.  $\square$

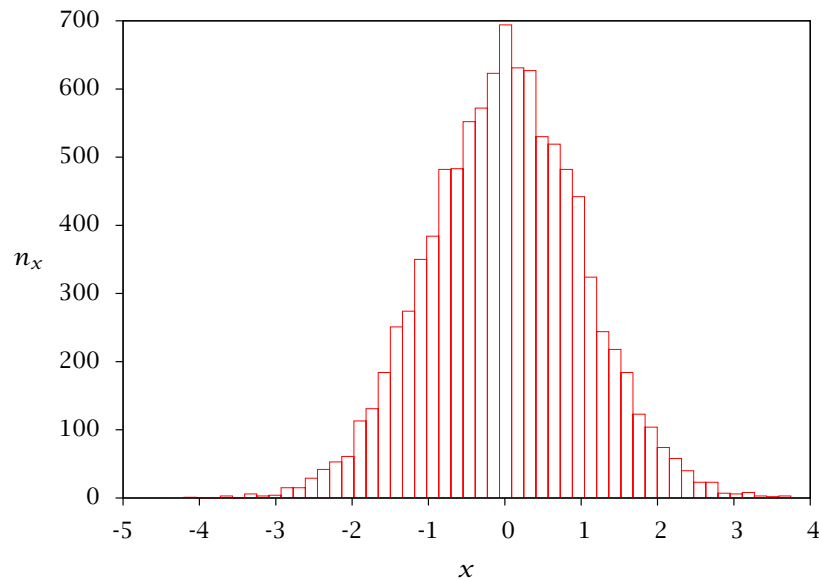


Figure 4: Histogram of 10,000 samples of MATLAB's randn function.

### Sample mean and variance

We can use the samples to generate an approximation of the mean and variance of the random variable. These are known as the *sample mean* and *sample variance*, respectively, to distinguish the approximation computed from samples from the true mean and variance. Let the samples of  $x$  be denoted  $x_i, i = 1, 2, \dots, n$  in which we have  $n$  samples. The sample mean and variance are given by

$$m = \frac{1}{n} \sum_{i=1}^n x_i$$
$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2$$

Notice the sample mean is the formula you always use to compute an average of a collection of numbers. The sample variance sums the squares of the distances of the samples from the mean, and then divides by  $n - 1$  (rather than  $n$ ). The division by  $n - 1$  reflects the fact that we cannot compute a variance or spread when we have only a single sample.

We can extend the sample mean and variance when  $\mathbf{x}$  is a vector of random variables.

Let  $\mathbf{x}_i$  denote the  $i$ th sample of the random  $\mathbf{x}$  vector. The formulas are

$$\mathbf{m} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$$

$$\mathbf{P} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})'$$

Notice in the sample variance, we take the (outer) product of two vectors  $\mathbf{x}\mathbf{x}'$ , which is an  $n \times n$  matrix, and not the usual (inner) product  $\mathbf{x}'\mathbf{x}$ , which is a scalar. The MATLAB functions `mean` and `cov` compute the sample mean and covariance for samples of a vector-valued random variable

### Example 2: Sample mean and covariance for three different measurement types

Say we measure three constant variables in an experimental system: the temperature, pressure, and concentration of some key component. Say the mean temperature is 20 °C, the mean pressure is 1 bar, and the mean concentration is 2 mol/L. Say we know the measurement error standard deviations are 0.5 °C, 0.2 bar, and 0.1 mol/L, respectively. We also know these measurement errors are independent random variables. Use `randn` to generate 15 samples of these three measurements and use `mean` and `cov` to compute the sample mean and covariance. How close are the sample mean and covariance to the true mean and covariance used to generate the samples?

### Solution

We generate the samples and compute the sample mean and covariance using

```
nsam = 15;

Tsam = 20 + 0.5*randn(nsam,1);
Psam = 1 + 0.2*randn(nsam,1);
Csam = 2 + 0.1*randn(nsam,1);

Measmat = [Tsam, Psam, Csam]

smean = mean(Measmat)
sP = cov(Measmat)
plot(Measmat)
```

which generates the following output

```
Measmat =

    20.15075    0.92840    2.12553
    19.37189    1.31131    2.00060
    19.40432    1.31351    1.82824
```



20.16828	0.76685	1.88849
20.34552	0.91807	1.94029
20.41108	0.89530	2.03677
20.16328	0.96543	1.86885
18.86462	1.00509	1.94699
20.35392	1.03841	2.14621
20.17987	0.96683	1.84220
19.98065	0.64298	1.87331
20.07565	0.94577	2.06438
20.67500	0.93168	2.11631
19.57186	0.94377	2.03733
20.21466	0.98150	1.84896

smean =

19.99542	0.97033	1.97096
----------	---------	---------

sP =

2.3235e-01	-3.8051e-02	1.3868e-02
-3.8051e-02	2.8680e-02	7.6420e-05
1.3868e-02	7.6420e-05	1.2442e-02

Notice the mean is fairly accurate. The true covariance matrix is

$$\mathbf{P} = \begin{bmatrix} (0.5)^2 & 0 & 0 \\ 0 & (0.2)^2 & 0 \\ 0 & 0 & (0.1)^2 \end{bmatrix}$$

The covariance matrix is diagonal because the measurement errors are not correlated with each other. Notice the sample covariance and the true covariance are fairly different from each other. The diagonal elements are reasonably close but the off-diagonal elements in the sample covariance are not very close to zero. It is generally true that sample means are accurate with a small number of samples, but sample variances require a much larger number of samples before they are accurate. The data are plotted in Figure 5.

□

## 2.5 Linear transformation of a random variable

If we multiply a scalar random variable  $x$  by a constant  $a$  and define that to be a new random variable  $y$ , then we alter both the mean and variance of  $y$  compared to  $x$ . The formulas are given as follows

$$y = ax$$

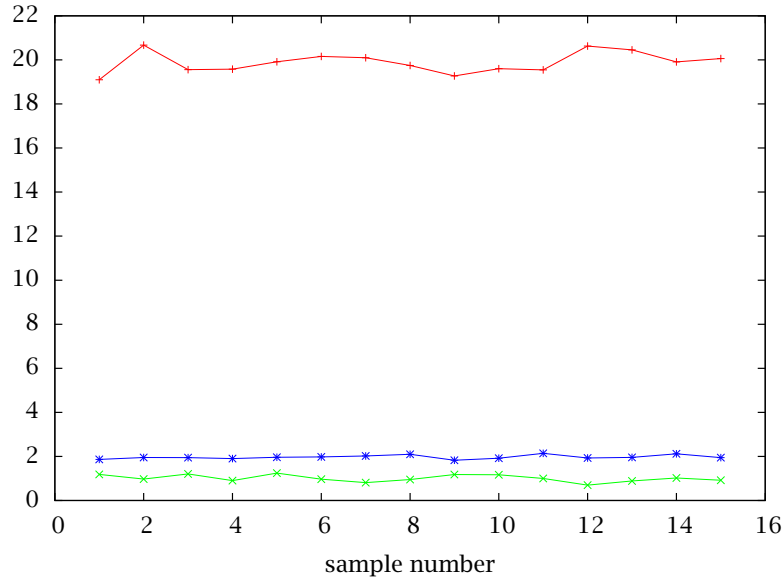


Figure 5: The 15 samples of measurement, pressure, and concentration.

$$m_y = a m_x \quad \text{var}(y) = a^2 \text{var}(x)$$

You should test this result by generating some samples of a uniform or normal distribution and multiplying by a scalar.

For normal distributions, we know more. If  $x$  is normal, then  $y$  is also normal, and for  $x \sim N(m_x, \sigma_x^2)$ , then  $y \sim N(m_y, \sigma_y^2)$ , and  $m_y = a m_x$  and  $\sigma_y^2 = a^2 \sigma_x^2$ .

**Vector of random variables.** The corresponding formulas for vector-valued random variables are as follows.

$$\begin{aligned} \mathbf{y} &= \mathbf{A}\mathbf{x} \\ \mathbf{m}_y &= \mathbf{A}\mathbf{m}_x \quad \text{var}(\mathbf{y}) = \mathbf{A}\text{var}(\mathbf{x})\mathbf{A}' \end{aligned} \quad (4)$$

In particular if  $\mathbf{x} \sim N(\mathbf{m}_x, \mathbf{P}_x)$  then  $\mathbf{y} \sim N(\mathbf{m}_y, \mathbf{P}_y)$  in which

$$\mathbf{m}_y = \mathbf{A}\mathbf{m}_x \quad \mathbf{P}_y = \mathbf{A}\mathbf{P}_x\mathbf{A}'$$

This result provides a handy way to generate normal distributions with desired covariance  $\mathbf{P}$ . The function `randn` provides random variables with unit variance,  $\mathbf{P}_x = \mathbf{I}$ . If we generate samples of  $\mathbf{y}$  by multiplication by the *square root* of the matrix  $\mathbf{P}$ , then we have  $\mathbf{y} = \sqrt{\mathbf{P}}\mathbf{x}$ , and the variance for  $\mathbf{y}$  is given by Equation 4

$$\mathbf{P}_y = \sqrt{\mathbf{P}} \mathbf{I} (\sqrt{\mathbf{P}})' = \mathbf{P}$$

The MATLAB command for the square root of a matrix is `sqrtn`.

## 2.6 Central limit theorem.

Put a statement of the central limit theorem here.

### Example 3: Adding ten uniformly distributed random variables.

Consider ten uniformly and independently distributed random variables,  $x_1, x_2, \dots, x_{10}$ . Consider a new random variable  $y$ , which is the sum of the ten  $x$  random variables

$$y = x_1 + x_2 + \dots + x_{10}$$

Even though the ten  $x$  random variables are uniformly distributed, and their probability distribution looks nothing like a normal distribution, let's explore the probability distribution of the resulting  $y$  random variable. According to the central limit theorem, it may appear to be normally distributed.

The  $x$  random variables are distributed as

$$x \sim U(0, 1)$$

Computing the mean and variance gives

$$\begin{aligned}\bar{x} &= \int_0^1 x dx = \frac{x^2}{2} \Big|_0^1 = \frac{1}{2} \\ \text{var}(x) &= \int_0^1 (x - \bar{x})^2 dx = \frac{1}{3} (x - 1/2)^3 \Big|_0^1 = \frac{1}{12}\end{aligned}$$

If we stack the  $x$  variables in a vector

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{10} \end{bmatrix}$$

we can write the  $y$  random variable as the linear transformation of the  $x$ 's

$$y = Ax \quad A = [1 \quad 1 \quad \dots \quad 1]$$

Using the previous results on linear transformations we know that  $y$ 's mean and variance are given by

$$\begin{aligned}\bar{y} &= A\bar{x} & \bar{y} &= 5 \\ \text{var}(y) &= A\text{var}(x)A' & \text{var}(y) &= \frac{5}{6}\end{aligned}$$

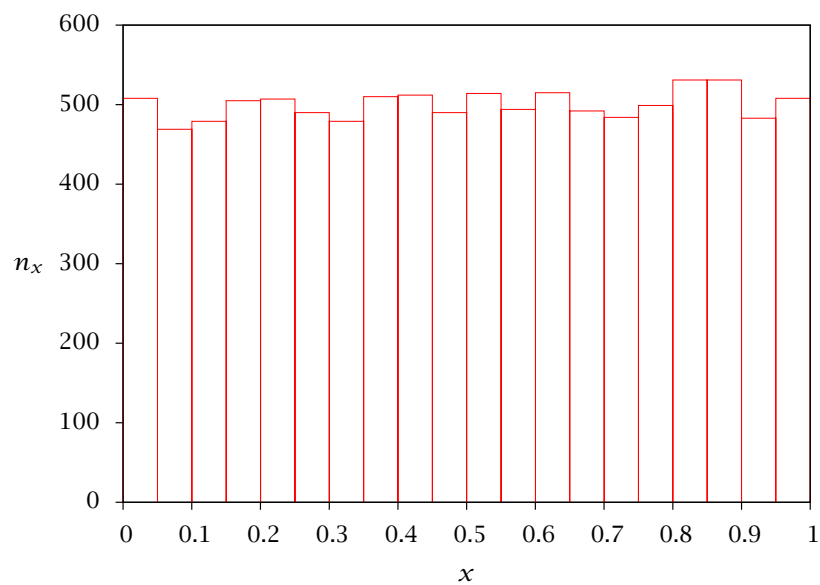


Figure 6: Histogram of 10,000 samples of uniformly distributed  $x$ .

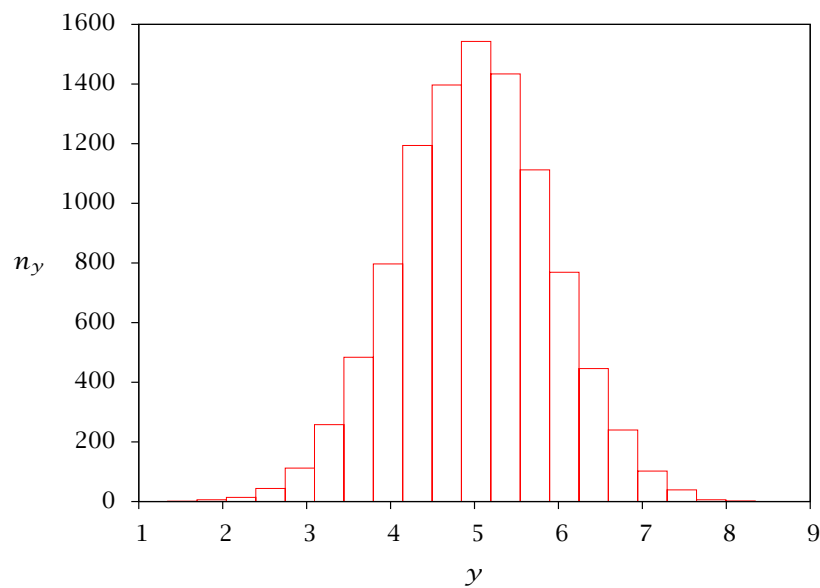


Figure 7: Histogram of 10,000 samples of  $y = \sum_{i=1}^{10} x_i$ .

So, if the central limit theorem is in force with only ten random variables in the sum, we might expect  $y$  to be distributed as

$$y \sim N(5, 5/6)$$

We use the following Octave code (MATLAB code is similar) to generate 10,000 samples of the 10 uniformly distributed  $x$  random variables and add them together to make  $y$ .

```

nsam = 10000;
nsum = 10;
x = rand(nsam, nsum);
y = sum(x,2);
figure(1); hist(x,50)
figure(2); hist(y,50)
mx = mean(x(1,:))
varx = var(x(1,:))
my = mean(y)
vary = var(y)

```

Executing this code and then examining the means and variances of the first  $x$ ,  $x_1$ , and  $y$  gives

```

mx = 0.48859
varx = 0.12250
my = 4.9949
vary = 0.80278

```

A histogram of the 10,000 samples of  $x_1$  and  $y$  are shown in Figures 6 and 7. It is clear that even ten uniformly distributed  $x$  random variables produce nearly a normal distribution for their sum  $y$ .

□

## 2.7 Confidence intervals for the normal distribution

Assume  $\mathbf{x}$  is normally distributed with mean  $\mathbf{m}$  and covariance  $\mathbf{P}$ . Figure 8 shows 1000 samples for

$$\mathbf{m} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \mathbf{P} = \begin{bmatrix} 2 & 0.75 \\ 0.75 & 0.5 \end{bmatrix}$$

The samples were generated with the following code

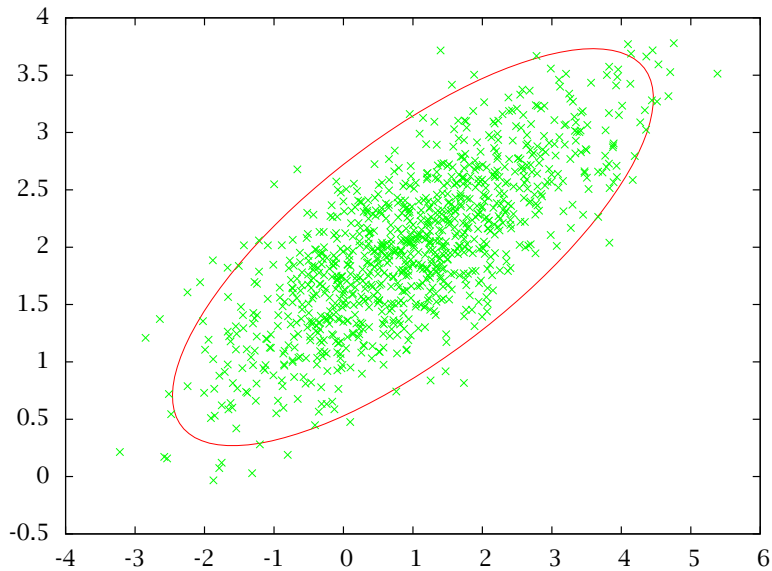


Figure 8: One thousand samples of the random variable  $\mathbf{x} \sim N(\mathbf{m}, \mathbf{P})$  and 95% probability contour.

```

m = [1;2];
n = length(m);
P = [2, 0.75; 0.75, 0.5];
nsam = 1000;

%% generate the samples x is N(m, P)
x = repmat(m, 1, nsam) + sqrtm(P)*randn(n,nsam);

%% compute the 95% confidence interval ellipse
alpha = 0.95;
A = inv(P);
b = chi2inv(alpha, n);
[xe, ye] = ellipse(A, b, 100, m);

%% plot samples and 95% confidence ellipse
plot(x(1,:), x(2,:), 'o', xe, ye)

```

The  $\alpha$ -level probability ellipse for the normal is given by

$$(\mathbf{x} - \mathbf{m})' \mathbf{P}^{-1} (\mathbf{x} - \mathbf{m}) \leq \chi^2(n, \alpha)$$

Notice from the code sample that the MATLAB function to return the value of  $\chi^2(n, \alpha)$  is<sup>1</sup>

```
chi2inv(alpha, n)
```

For  $\alpha = 0.95$  and  $n = 2$ , we have  $\chi^2(n, \alpha) = 5.99$ , and each sample of  $\mathbf{x}$  has 95% probability of lying inside this ellipse. We use the function `ellipse(A, b, 100, m)` to generate the ellipse corresponding to  $(\mathbf{x} - \mathbf{m})'A(\mathbf{x} - \mathbf{m}) \leq b$ . Notice we can make this plot only when  $n = 2$ . In higher dimensions we report the box containing the ellipse as shown in Figure 3.

Figure 8 shows 1000 samples and we see that the samples indeed generate an elliptical region. For 1000 samples and  $\alpha = 0.95$ , we would expect that  $1000(1 - 0.95) = 50$  samples would lie outside the 95% probability region. In Figure 8, the actual number is 44. You can compute the number outside the ellipse with the code

```
e = x - repmat(m, 1, nsam);
sum( diag(e'*inv(P)*e) > chi2inv(alpha, n) )
```

The first command moves the mean of the samples to zero and the second command measures how far each sample is from zero. The ones outside the ellipse have values of  $\mathbf{e}'\mathbf{P}^{-1}\mathbf{e}$  greater than  $\chi^2(n, \alpha)$  and the “greater than” test generates 1 for these samples and 0 for the others. The `sum` command then counts how many lie outside the ellipse.

In parameter estimation problems, the  $\alpha$ -level probability ellipse is also known as an  $\alpha$ -level confidence interval. It is customary to report 95% confidence intervals, although reporting several values of  $\alpha$  may be appropriate depending on the application.

### 3 Least-Squares Estimation

Consider again the problem of fitting a straight line to data

$$y_i = mx_i + b$$

in which  $y_i$  is the measurement at  $x_i$ ,  $i = 1, \dots, n_d$  and  $n_d$  is the number of data points. Using matrix vector notation, we can write the equation for all the data as

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta}$$

in which the parameters to be estimated are placed in the  $\boldsymbol{\theta}$  vector

$$\boldsymbol{\theta} = \begin{bmatrix} m \\ b \end{bmatrix}$$

and the  $\mathbf{y}$  vector and  $\mathbf{A}$  matrix are given by

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_d} \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_{n_d} & 1 \end{bmatrix}$$

<sup>1</sup>In particular, notice that the order of the arguments  $(n, \alpha)$  is reversed.

We do not expect the best fit line to pass through all the data points, so we modify the model to account for measurement error<sup>2</sup>

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta}_0 + \mathbf{e} \quad (5)$$

in which  $\boldsymbol{\theta}_0$  is the (unknown!) true parameter value vector and  $\mathbf{e}$  is a random variable. We model the measurement error as a normal distribution with mean  $\mathbf{0}$  and variance  $\sigma^2\mathbf{I}$ .

$$\mathbf{e} \sim N(\mathbf{0}, \sigma^2\mathbf{I}) \quad (6)$$

The best estimate of  $\boldsymbol{\theta}$  in a least-squares sense is given by

$$\hat{\boldsymbol{\theta}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y} \quad (7)$$

a formula that we used in Module 2. However, we also can examine the distribution of parameter estimates given the observed measurements corrupted by the measurement errors. Imagine we create replicate datasets by drawing measurement errors  $\mathbf{e}$  from the distribution given in Equation 6. For each dataset we apply Equation 7 and produce a parameter estimate. The distribution of measurement errors creates a distribution of parameter estimates. In fact, for models linear in the parameters, we can show the parameter estimates also are normally distributed

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}_0, \mathbf{P})$$

in which the mean is the true value of the parameters and the covariance is

$$\mathbf{P} = \sigma^2(\mathbf{A}'\mathbf{A})^{-1}$$

We also can calculate the parameter “confidence intervals.” We merely compute the size of the ellipse containing a given probability of the multivariate normal. As discussed previously, that can be shown to be the chi-square probability function (Box and Tiao, 1973, p. 116). Given the number of estimated parameters,  $n_p$ , and the confidence level,  $\alpha$ , then

$$\frac{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \mathbf{A}' \mathbf{A} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})}{\sigma^2} \leq \chi^2(n_p, \alpha) \quad (8)$$

The  $\chi^2$  distribution is tabulated in many statistics handbooks Box, Hunter, and Hunter (1978) and is available in many computing environments<sup>3</sup>

<sup>2</sup>Notice the model *structure* is usually in error also, e.g., the true relationship between  $y$  and  $x$  may be nonlinear, variables other than  $x$  may be required to predict  $y$ , and so on. The procedure outlined here lumps structural error into  $\mathbf{e}$  as well, but structural error is not accounted for correctly in this way. If the structure is in serious doubt, one may pose instead model discrimination tests to choose between competing models with different structures Stewart, Henson, and Box (1996); Stewart, Shon, and Box (1998).

<sup>3</sup>The MATLAB command is `chi2inv(alpha, n)`.



## 4 Basic Model Fitting

Consider a model represented as follows:

$$y = f(x, \theta) \quad (9)$$

The model provides a prediction for the value of  $y$  corresponding to the given input  $x$ . The model output could be a vector of values  $y = \{y_1, \dots, y_{N_y}\}$ , although most often it is a single variable. Similarly, the model input  $x = \{x_1, \dots, x_{N_x}\}$  might be one or several variables. The adjustable parameter(s) are contained in the vector  $\theta = \{\theta_1, \dots, \theta_{N_\theta}\}$ .

### Example 4: Antoine Model for Vapor Pressure

One of the simplest models used to correlate vapor pressure data is the Antoine equation:

$$\ln P_{\text{sat}} = A + \frac{B}{T + C} \quad (10)$$

This model outputs a predicted value for the vapor pressure  $P_{\text{sat}}$  of a substance for a given temperature  $T$ .  $A$ ,  $B$  and  $C$  are parameters in the model, determined by fitting to experimental vapor pressure data.  $\square$

The data used in parameter estimation consists of the results of several experiments, each providing a set of corresponding, simultaneously-observed values for the “input” and “output” variables of the model. Each experimental result is termed an “observation” (or more commonly, a “data point”). Each observation  $k$  consists of the values

$$x^k = \{x_1^k, \dots, x_{N_x}^k\}, \quad y^k = \{y_1^k, \dots, y_{N_y}^k\}$$

The distinction between the input and output variables in the model is generally not important from the standpoint of parameter estimation, since all variable values must be observed. Note the model form need not match experimental circumstances. The independent variables  $x$  in the model form need not be variables that are independently adjustable in the experiment. For instance, in Example 4, one might perhaps set the pressure  $P_{\text{sat}}$  for each data point and then measure the corresponding temperature  $T$ .

### 4.1 “Nonlinear Least Squares”

With more data points than unknown parameters, it is not possible for the model to fit each data point exactly. The question is how to obtain the best compromise so that the model predictions on the whole are close to the data values. Closeness for any single observation may be measured by the square of the distance between the model prediction and the observation. Closeness averaged over the entire data set is often measured by the sum of the squares of the distances for each individual observation. Minimizing this combined measure of distance makes the model predictions as close as possible to the data in an average sense, and methods based on this idea are referred to as “least squares” methods.

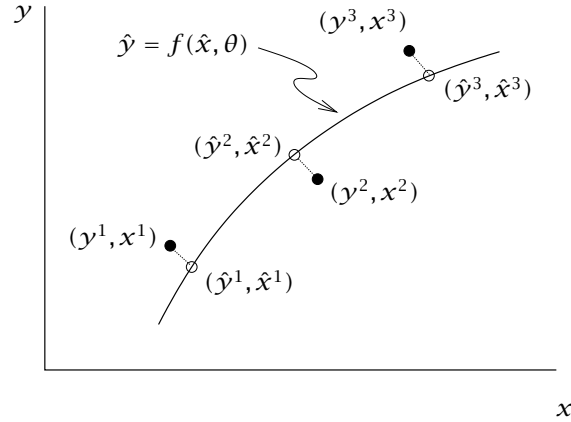


Figure 9: Distances between data points and model-fitted points  $(\hat{y}, \hat{x})$ .

The model will not fit the observed data points  $(y^k, x^k)$ ,  $k = 1, \dots, K$  exactly, so we consider the corresponding set of points  $(\hat{y}^k, \hat{x}^k)$ ,  $k = 1, \dots, K$  that do fit the model exactly and are closest to each data point (see Figure 9). For each data point, the values  $\hat{y}$  are those predicted by the model for inputs  $\hat{x}$ , and  $\hat{x}$  is determined so as to make  $(\hat{y}, \hat{x})$  as close as possible to the data point values  $(y, x)$ .

The sum-of-squares measure is as follows:

$$SS = \sum_{k=1}^K \left[ \sum_{i=1}^{N_y} \left( \frac{\hat{y}_i^k - y_i^k}{\sigma_{y_i}} \right)^2 + \sum_{j=1}^{N_x} \left( \frac{\hat{x}_j^k - x_j^k}{\sigma_{x_j}} \right)^2 \right] \quad (11)$$

$$\text{where } \hat{y}^k = f(\hat{x}^k, \theta), \quad k = 1, \dots, K$$

In the above, the quantities  $\hat{y}_i^k - y_i^k = \epsilon_{y_i}^k$  and  $\hat{x}_j^k - x_j^k = \epsilon_{x_j}^k$  represent the “errors” between the value fitted to the model and each observation value. The denominator quantities  $\sigma_{y_i}$  and  $\sigma_{x_j}$  are constants selected to scale each type of variable  $y_i$  or  $x_j$ . This is done in order to put the errors on a comparable basis so it will be meaningful to add them all together into the sum in (11). Typically these error scaling constants are chosen as estimates of the accuracy in the experimental measurements of each variable type. For instance, if  $y$  were a pressure and  $x$  were a temperature, it would make little sense to add the square of a pressure error measured in bars to the square of a temperature error in Kelvin. However, if the accuracy of the pressure measurements is estimated to be  $\pm 0.02$  bar and the accuracy of the temperature measurements is estimated to be  $\pm 0.5$  K, squares of the scaled errors could be meaningfully added using  $\sigma_y = 0.02$  bar and  $\sigma_x = 0.5$  K. Note that the values of the error scales  $\sigma_{y_i}$  and  $\sigma_{x_j}$  have units, so that the scaled errors, and SS, are dimensionless. More can be said about these error scales in the context of statistical methods, but even when doing basic model fitting with a limited number of data points, the data fitter can and should provide estimates for them.

Once the sum-of-squares in (11) is constructed, the least squares parameter estimate can be computed by solving the following minimization problem:

$$\min_{\theta, \hat{x}^1, \dots, \hat{x}^K} SS(\theta, \hat{x}^1, \dots, \hat{x}^K) \quad (12)$$

While a simple calculation can find the minimum for the special case where  $f(x, \theta)$  is a linear function, models encountered in chemical engineering are often nonlinear. As with most nonlinear problems, solving (12) then requires an iterative solution and a starting estimate for the unknown variables. (12) is a particular kind of minimization problem where the objective function to be minimized is a sum of squares, and MATLAB has a specific function `lsqnonlin` that can be applied to solve problems of this form. `lsqnonlin` performs a minimization of the general form “  $\min_{X_1, \dots, X_M} \sum_{\ell=1}^{N_{\text{tot}}} [F_{\ell}(X)]^2$  ”. It obtains values of the vector of errors  $F(X)$  (not squared) by calling a function supplied by the user. It may be applied to (12) by constructing the vector  $F$  to contain one element for each error term in the sum (11), and by taking  $X = \{\theta, \hat{x}^1, \dots, \hat{x}^K\}$ .

When the minimum in (12) is found, the resulting value  $SS_{\text{min}}$  offers some indication of the average error between the fitted model predictions and the data values. There are  $K \cdot (N_y + N_x)$  error terms in the sum (11), while there are  $N_{\theta} + K \cdot N_x$  adjustable variables in (12), leaving the net number of error degrees of freedom in  $SS$  equal to  $K \cdot N_y - N_{\theta}$ . Therefore an average scaled error can be computed as  $\sqrt{\frac{SS_{\text{min}}}{K \cdot N_y - N_{\theta}}}$ . It is usually not possible to ascribe this residual error in a specific way to either uncertainties in the data or inaccuracy of the model form itself, as both reasons may contribute. However, the mean error value does indicate how well the model and data agree, and suggests the level of accuracy that might be expected when the fitted model is put to use.

#### Example 5: Fitting the Antoine Model

Fit the Antoine model in Example 4 to the following set of values reported<sup>4</sup> for the vapor pressure of acetone.

$T$ (°C)	$P_{\text{sat}}$ (mmHg)
-59.4	1
-40.5	5
-31.1	10
-20.8	20
-9.4	40
-2.0	60
7.7	100
22.7	200
39.5	400
56.5	760

<sup>4</sup>See (Perry, Green, and Maloney, 1997, Table 2-8, p.2-61)

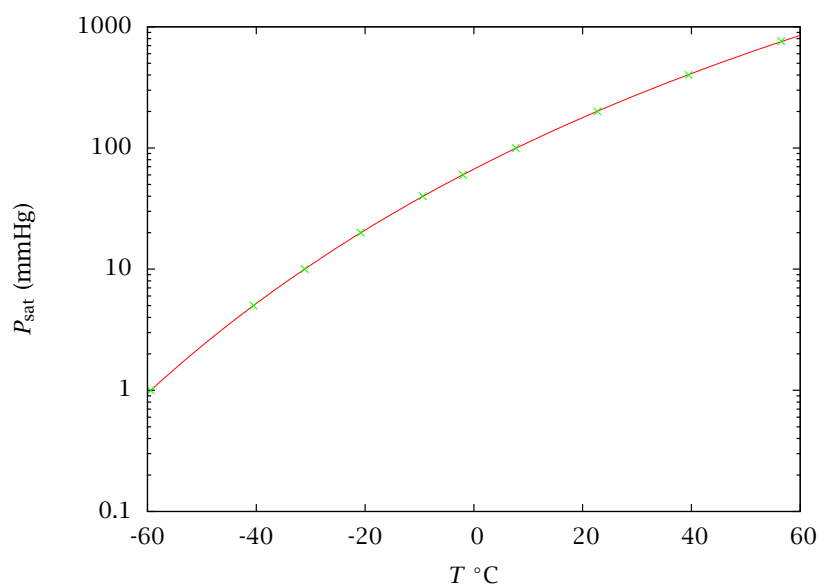


Figure 10: Antoine model fitted to the acetone vapor pressure data of Example 5.

Assume the reported temperature values have accuracies of  $\pm 0.1^\circ\text{C}$  and that the pressure values have accuracies of  $\pm 2\%$ . Plot the fitted model along with the data points. What is the average scaled error between the model predictions and the data? Interpret this value as an accuracy of the model predictions expressed in terms of (a) predicted temperature for a given pressure, and (b) predicted pressure for a given temperature.

### Solution

MATLAB code is contained in script `fitacetone` together with the functions `nlls` and `Antoine`. When using `lsqnonlin` (which is called by `nlls`), you need to provide a sufficiently good initial guess for the parameters, which you provide when calling `nlls`. Once an acceptable fit is obtained with an acceptably close initial guess, it sometimes helps to run the numerical minimization again using the result of the previous calculation as the new initial guess. In this case the resulting fit is quite good (Figure 10). The fact that the average scaled error of the fit is less than unity suggests that the values being fitted were smoothed before being listed in the table in the handbook.

□

## 5 Parameter Estimation with Differential Equation Models

Now we turn to the single most important parameter estimation problem in chemical reactor modeling: determining reaction-rate constants given dynamic concentration measurements. To get started, we consider a simple reactor model consisting of a single differential equation with a single experimentally measured quantity

$$\frac{dx}{dt} = f(x; \theta) \quad (13)$$

$$x(0) = g(x_0; \theta) \quad (14)$$

$$y = h(x) \quad (15)$$

in which  $x$  is the single material balance of interest,  $\theta$  are the unknown model parameters,  $x_0$  is the initial condition, and  $y$  is the experimentally measurable quantity. For simplicity let us assume here that  $x$  itself is measured, in which case  $h(x) = x$ . As we see in several of the examples, it may be necessary to include some of the initial conditions also as unknown parameters. Often  $t$  is time, but in steady-state tubular PFRs, reactor volume or length can take the place of time without changing the structure of the parameter-estimation problem.

As before, we define a least-squares objective to measure our fit to the data

$$\Phi(\theta) = \sum_i (\tilde{x}_i - x_i)^2 \quad (16)$$

in which  $\tilde{x}_i$  is the experimental measurement at time  $t_i$ , and  $x_i$  is the solution to the model at time  $t_i$ ,  $x_i = x(t_i; \theta)$ . Note  $x_i$  is the only part of the objective function that depends on the model parameters. Again, we minimize this objective function to obtain our parameter estimates

$$\min_{\theta} \Phi(\theta) \quad (17)$$

subject to Equations 13–15.

The major change is that the model constraint consists of nonlinear differential equations rather than linear or nonlinear algebraic equations as in the previous sections. The differential equations make it much more expensive to evaluate the constraints while solving the optimization problem.

### Parameter estimation algorithm:

- ① Guess initial parameter values.
- ② Using an appropriate ODE solver, solve the model given the current parameter values. Compute  $x_i$ .
- ③ Evaluate  $\Phi$  using Equation 16.
- ④ Update parameter values to minimize  $\Phi$ . This step and the next are usually controlled by an optimization package.

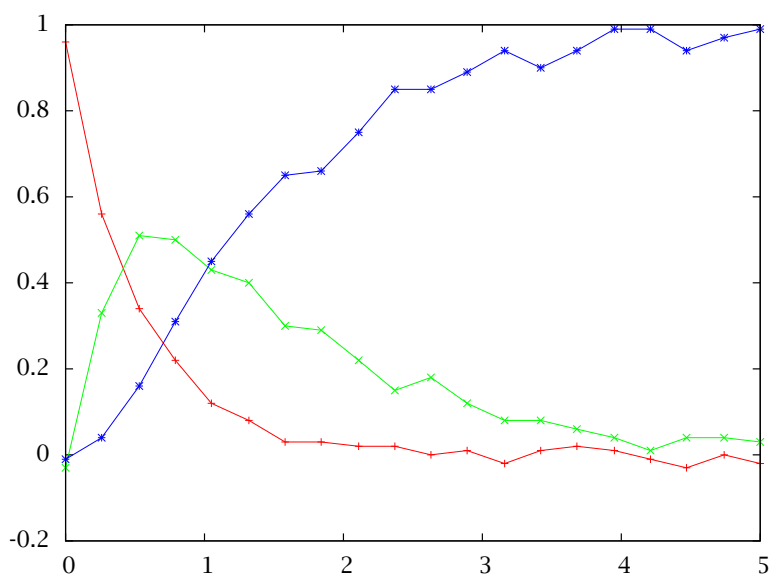


Figure 11: Measurements of species concentrations in Reactions 18 versus time.

- ⑤ Check convergence criteria. If not converged, go to ②.
- ⑥ On convergence, set  $\hat{\theta}$  to current parameter values. Calculate approximate confidence intervals.

### Example 6: Estimating two rate constants in reaction $A \rightarrow B \rightarrow C$

Consider the irreversible series reactions



We wish to estimate the two rate constants  $k_1$  and  $k_2$  from the measurements shown in Figure 11. We would also like to know how much confidence to place in these estimated parameters.

#### Solution

We use the function `parest.m` to estimate parameters in differential equation models. Type at the command line `help parest` to get an idea of the many options available. For this problem we require only the following simple options.

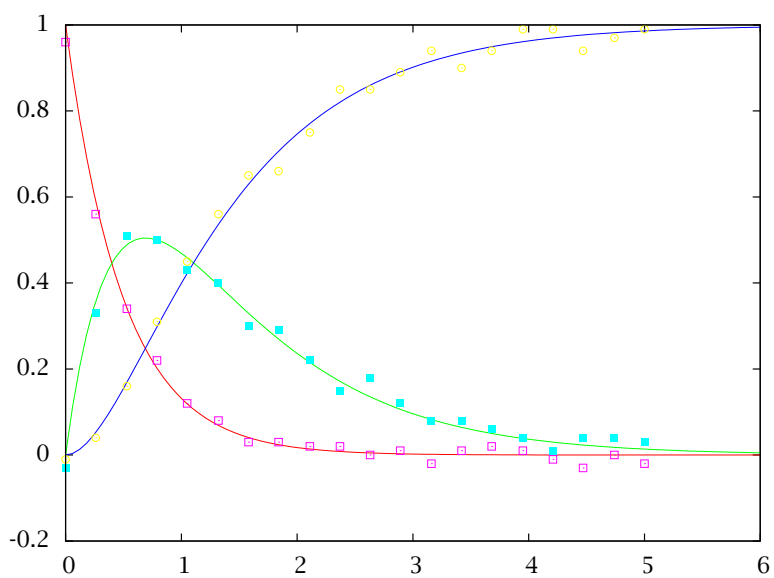


Figure 12: Fit of model to measurements using estimated parameters.

```
%% ABC.m
%% Estimate parameters for the A->B->C model
%% jbr, 11/2008

% make sure ensure sundials/cvode is in the path
startup_STB

k1 = 2; k2 = 1;
ca0 = 1; cb0 = 0; cc0 = 0;
thetaac = [k1; k2];
tfinal = 6; nplot = 100;

model.odefcn = @massbal_ode;
model.tplot = linspace(0, tfinal, nplot)';
model.param = thetaac;
model.ic = [ca0; cb0; cc0];

objective.estflag = [1, 2];
objective.panic = [0.5; 3];
objective.parlb = [1e-4; 1e-4;];
objective.parub = [10; 10];

measure.states = [1,2,3];
%% load measurements from a file
table = load ('ABC_data.dat');
measure.time = table(:,1);
measure.data = table(:,2:end);

%% estimate the parameters
estimates = parest(model, measure, objective);

disp('Estimated Parameters and Bounding Box')
[estimates.parest estimates.bbox]

%%plot the model fit to the noisy measurements
figure(1)
plot(model.tplot, estimates.x, measure.time, measure.data, 'o');
```

We also require the function `massbal_ode` to define the right-hand side of the differential equations



```
function xdot = massbal_ode(x, t, theta)
    ca = x(1);
    cb = x(2);
    cc = x(3);
    k1 = theta(1);
    k2 = theta(2);
    r1 = k1*ca;
    r2 = k2*cb;
    xdot = [-r1; r1-r2; r2];
end
```

Running the `ABC.m` file produces the following parameter estimates and confidence intervals.

$$\hat{\theta} = \begin{bmatrix} 2.03 \\ 0.989 \end{bmatrix} \pm \begin{bmatrix} 0.229 \\ 0.0823 \end{bmatrix} \quad \theta_0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

The estimates are close to the true values  $\theta_0$  used to generate the data. The confidence intervals are reasonably tight given the three species measurements with the noise level indicated in Figure 11. The fit of the model using the estimated parameters is shown in Figure 12.

□

## 6 Exercises

### Exercise 1: Estimating rate constants in series reactions with different measurements

Consider again Example 6. Download the program `parest.m` from the class website and do the following parameter estimation problems.

- Consider measurements of all three species to be available and reproduce the estimates and approximate confidence intervals given in the Example 6.
- Next consider measurements of only species B and C to be available. Estimate the two rate constants using only these measurements. What happens to the confidence intervals? Do you have sufficient information to estimate both rate constants with B and C measurements?
- Next consider measurement of only species B to be available and estimate both rate constants using this measurement. What happens to the confidence intervals? Do you have sufficient information to estimate both rate constants with only the B measurement?

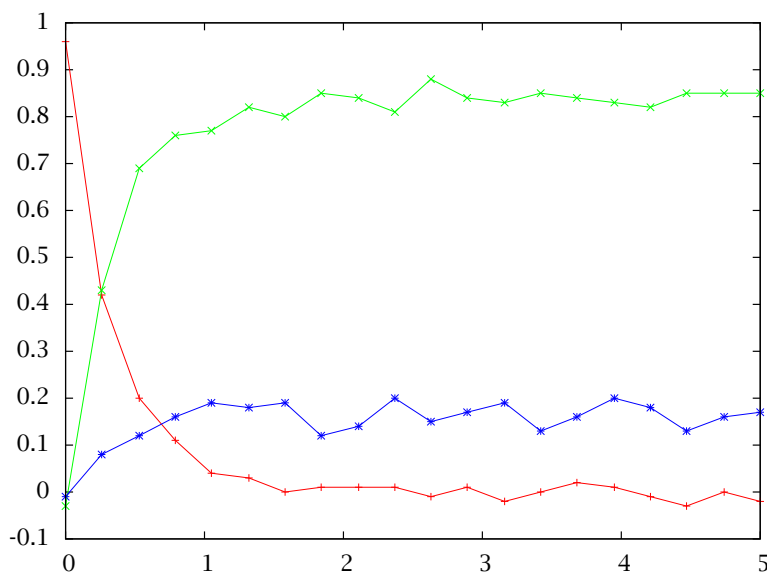


Figure 13: Measurements of species concentrations versus time for parallel reactions.

- (d) Next consider the measurement of only species A to be available and estimate both rate constants using this measurement. What happens to the confidence intervals? Do you have sufficient information to estimate both rate constants with only the A measurement?
- (e) Finally consider the measurement of only species A to be available, but estimate only the first rate constant. Assume the second rate constant is known to be  $k_2 = 1$ . What happens to the confidence intervals? Do you have sufficient information to estimate only the first rate constant with only the A measurement?

### Exercise 2: Estimating rate constants in parallel reactions with different measurements

Consider the parallel reaction scheme



and the measurements shown in Figure 13. The reactor is initially filled with only reactant A.

- (a) Fit both rate constants using all three species measurements: (A,B,C). Comment on the model fit and the uncertainty in the parameters. Are these measurements adequate to estimate the two rate constants.

- (b) Repeat for all combinations of two species measurements; (A,B), (A,C), and (B,C). Are two measurements sufficient for estimating the two rate constants?
- (c) Repeat for all single species measurements; (A), (B), and (C). Is one measurement sufficient for estimating the two rate constants?

**Exercise 3: Linear transformation of a normal is normal (Rawlings and Ekerdt, 2012, Exercise 9.14)**

Let  $\mathbf{x}$  be a normally distributed random variable with mean  $\mathbf{m}_x$  and covariance  $\mathbf{P}_x$

$$\mathbf{x} \sim N(\mathbf{m}_x, \mathbf{P}_x)$$

Let a new random variable  $\mathbf{z}$  be defined as a linear (affine) transformation of  $\mathbf{x}$

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

in which  $\mathbf{A}$  is a constant matrix and  $\mathbf{b}$  is a constant vector. Then one can show that  $\mathbf{z}$  is also normally distributed with mean  $\mathbf{m}_z$  and covariance  $\mathbf{P}_z$  given by

$$\mathbf{z} \sim N(\mathbf{m}_z, \mathbf{P}_z) \quad \mathbf{m}_z = \mathbf{A}\mathbf{m}_x + \mathbf{b} \quad \mathbf{P}_z = \mathbf{A}\mathbf{P}_x\mathbf{A}' \quad (20)$$

- (a) Generate 1000 samples of  $\mathbf{x}$  for

$$\mathbf{x} \sim N(\mathbf{m}_x, \mathbf{P}_x) \quad \mathbf{m}_x = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \mathbf{P}_x = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$$

Plot the 95% probability contour (ellipse) for the random variable  $\mathbf{x}$  and plot the 1000 samples on the same plot. How many of your samples are inside the ellipse?

- (b) Transform your 1000 samples of  $\mathbf{x}$  using

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{b} \quad \mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} -10 \\ 12 \end{bmatrix}$$

What is the mean and covariance for  $\mathbf{z}$  according to Equation 20. Plot the 95% probability contour (ellipse) for the random variable  $\mathbf{z}$  and plot the 1000 transformed samples on the same plot. How many of your  $\mathbf{z}$  samples are inside the ellipse?

- (c) Repeat for

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} -10 \end{bmatrix}$$

Since  $z$  is a scalar for this case, make a histogram plot of the  $z$  samples.

- (d) Consider again the linear model and least-squares estimate

$$\mathbf{y} = \mathbf{A}\boldsymbol{\theta}_0 + \mathbf{e} \quad \hat{\boldsymbol{\theta}} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{y}$$

in which

$$\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

Use the above result on linear transformation of normals to show that the parameter estimates are distributed as given in Section 3

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}_0, \mathbf{P}) \quad \mathbf{P} = \sigma^2 (\mathbf{A}' \mathbf{A})^{-1}$$

#### Exercise 4: Estimating rate constant and activation energy from rate constant measurements at different temperatures

Assume a reaction rate has been measured at several different temperatures in the range  $300 \text{ K} \leq T \leq 500 \text{ K}$ . Model the rate (rate constant) as

$$k = k_0 \exp(-E/T) \quad (21)$$

in which  $k_0$  ( $\text{min}^{-1}$ ) is the pre-exponential factor and  $E$  (K) is the activation energy scaled by the gas constant. To make the estimation problem linear, transform the data by taking the logarithm of Equation 21

$$\ln k = \ln k_0 - E/T$$

- (a) Estimate the parameters  $\ln k_0$  and  $E$  using least squares given the following single experiment consisting of nine rate constant measurements at the following temperatures

T	k
300	0.01658
325	0.06151
350	0.09822
375	0.2751
400	0.9982
425	2.113
450	4.401
475	4.505
500	13.42

What are the least squares estimates of  $\ln k_0$  and  $E$  for this experiment?

- (b) Plot the data and least squares fit on two plots: one plot of  $k$  versus  $T$  and a separate plot of  $\ln k$  versus  $1/T$ . How well does Equation 21 represent these data?
- (c) Next we wish to analyze the following 50 replicate experiments in which each experiment consists of nine measurements of  $k$  at different  $T$ . The first four of these experiments are shown here

T1	k1	T2	k2	T3	k3	T4	k4
303	0.01145	302	0.0175	302	0.01208	301	0.00955
323	0.0483	321	0.04742	326	0.06483	329	0.05799
351	0.167	345	0.1456	354	0.2289	354	0.1729
376	0.3197	379	0.5643	375	0.436	376	0.2612
402	1.221	396	0.9718	401	0.877	404	0.8787
429	1.278	425	2.166	427	2.152	422	1.052
453	3.844	447	2.853	452	2.354	448	3.882
479	4.519	476	5.098	477	4.685	475	5.931
501	13.8	504	12.05	496	8.29	496	6.235

Download these data from the file

kt\_lotsmeas.dat

Calculate a least squares estimates for each experiment. Plot the 50 estimates using the estimate of  $\ln k_0$  as the abscissa and the estimate of  $E$  as the ordinate.

- (d) Next combine all 50 experiments into one dataset of  $k, T$  values and estimate  $\ln k_0$  and  $E$  using all 450 measurements. What are the values of the estimates?
- (e) Plot all 450 measurement and the least squares fit on two plots: one plot of  $k$  versus  $T$  and a separate plot of  $\ln k$  versus  $1/T$ . How well does Equation 21 represent all of the measurements.

### Exercise 5: Fitting heat capacity data

The following data are available for the heat capacity of ethanol in the ideal gas state:

$T(K)$	$C_p$ (kJ/kmol·K)
250	61.0
300	69.5
350	69.9
400	85.8
450	90.4
500	91.9
550	100.0
600	108.5
650	118.1
700	123.2

- (a) Heat capacities are often fitted to simple polynomial models. Fit the above data to the model

$$C_p = A + BT + CT^2 + DT^3 + ET^4$$

Plot the fitted equation over the temperature range from 250–700 K, and in the same plot also show the data points. Estimate the accuracy of predictions made with the fitted model.

- (b) In Perry's Handbook<sup>5</sup>, ideal gas heat capacities are given by equations having the following form:

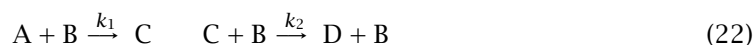
$$C_p = C_1 + C_2 \left[ \frac{C_3/T}{\sinh(C_3/T)} \right]^2 + C_4 \left[ \frac{C_5/T}{\cosh(C_5/T)} \right]^2$$

Fit this model to the above data. Plot the fitted model along with the polynomial fit and the original data over the range from 250–700 K. Estimate the accuracy of predictions made with the fitted model.

- (c) From molecular physics it is known that gas heat capacities plateau at a minimum value at lower temperatures, rise over an intermediate temperature range, and plateau again at a maximum value at high temperatures. Plot the predictions of the two fitted models obtained above over the range from 100–1500 K. Comment on the suitability of the functional forms assumed for each.
- (d) From the molecular structure of ethanol we can determine the lower and upper plateau values for the ideal gas heat capacity  $C_p$  to be  $4R$  and  $25R$  respectively. Propose a nonlinear model form with a function shape that varies between these two plateaus. Fit your model to the above data and plot the fitted model and the original data over the temperature range from 100–1500 K. Estimate the accuracy of predictions made with your model.
- (e) Why fit a model, why not just use the data you have and interpolate? Prepare an interpolant for the above data using the MATLAB function `polyfit` with  $n = K - 1$ . Then use function `polyval` to make a plot of the interpolation over the range from 250–700 K, marking the original data points in the plot.

### Exercise 6: Does B catalyze the second reaction?

We have collected the laboratory data shown in Figure 14 and are considering the following two proposed kinetic mechanisms to explain these data



<sup>5</sup>See (Perry et al., 1997, Table 2-198)

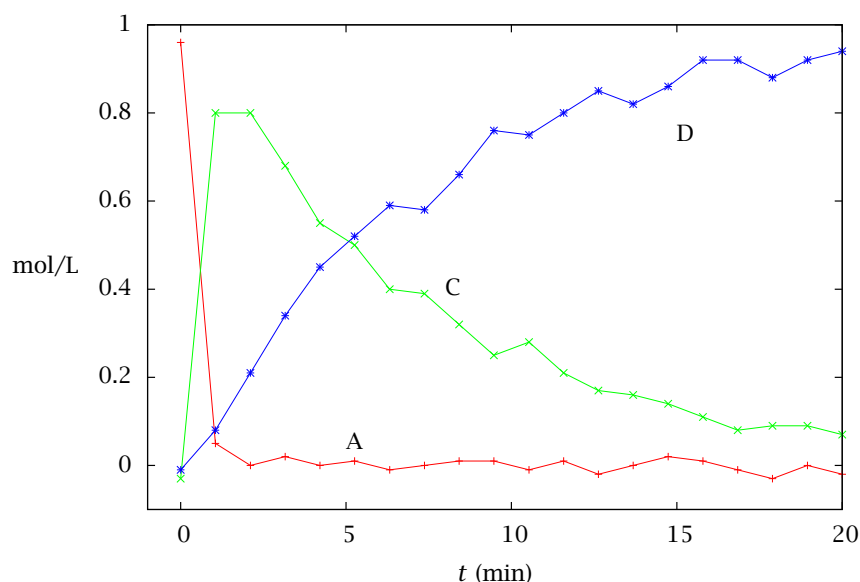


Figure 14: Measurements of species concentrations versus time for Exercise 6.

We have measurements of species A, C, and D, but not species B. Although we cannot measure species B, we do know that we charged the reactor with

$$c_A(0) = 1 \text{ mol/L} \quad c_B(0) = 10 \text{ mol/L} \quad c_C(0) = 0 \text{ mol/L} \quad c_D(0) = 0 \text{ mol/L}$$

We wish to find out whether or not species B catalyzes the second reaction.

- Estimate the kinetic parameters for Reactions 22 and show the best parameter estimates and their confidence intervals. Be sure to report the units of the rate constants. Plot the fit to the experimental data using the estimated parameters.
- Repeat for Reactions 23.
- Which mechanism better explains the data? Explain your answer.

### Exercise 7: Can you hear me now?

Consider again the two kinetic mechanisms of Exercise 6.

- Repeat Exercise 6, but with the experimental data shown in Figure 15 that were collected after charging the reactor with

$$c_A(0) = 4 \text{ mol/L} \quad c_B(0) = 3 \text{ mol/L} \quad c_C(0) = 1 \text{ mol/L} \quad c_D(0) = 0 \text{ mol/L}$$

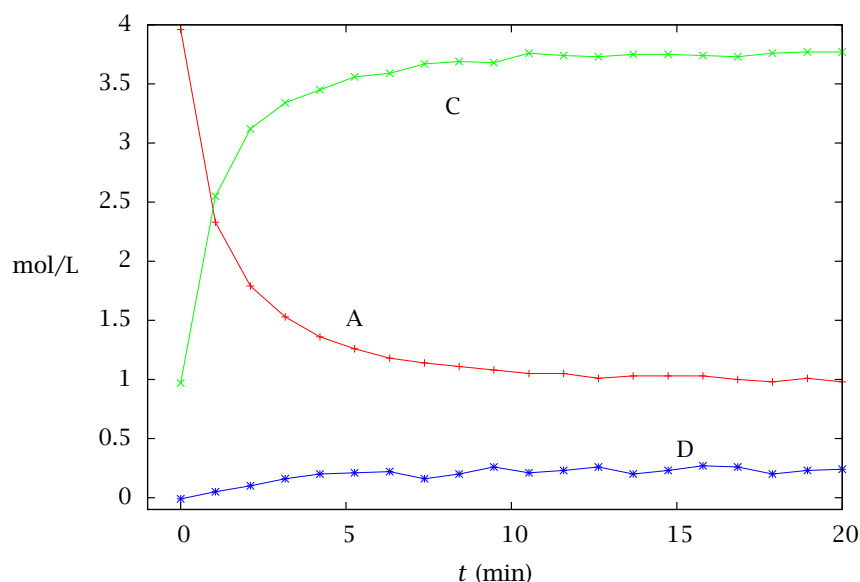


Figure 15: Measurements of species concentrations versus time for Exercise 7.

- (b) Make a plot of the residuals of the model fits for the catalyzed and uncatalyzed mechanisms. Use `help parest` to see how the residuals are returned by `parest` so you can plot them.

What do you conclude from these residual plots? Does your conclusion change about which mechanism better explains these data?

- (c) Explain any differences in your conclusions from the two different data sets. Although we are measuring the same species in both experiments, do the two experiments contain different information? Explain why or why not.

### Exercise 8: Finding rate constant and order

We want to estimate the rate constant  $k$  and order  $n$  for a rate expression by fitting it to some rate versus concentration data for the reaction



We have measurements of  $r$  for different  $c_A$  given in Figure 16. We assume the values of  $c_A$  are known exactly and all of the experimental measurement error is in  $r$ .

- (a) Consider a log transformation of the rate expression  $\ln(r) = \ln(k) + n \ln(c_A)$  with



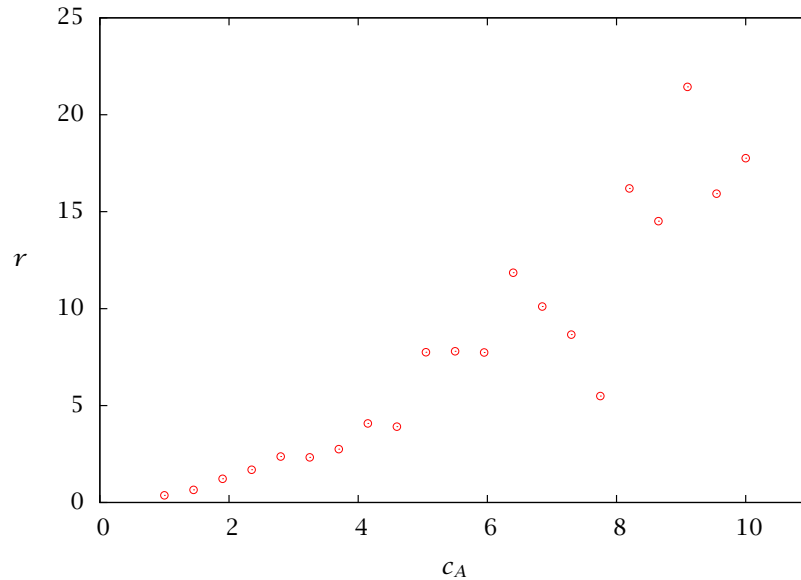


Figure 16: Rate versus concentration of A.

model parameters

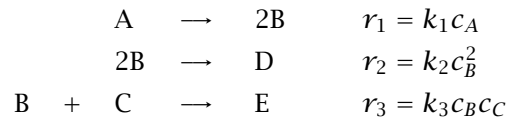
$$\boldsymbol{\theta} = \begin{bmatrix} \ln(k) \\ n \end{bmatrix}$$

Develop a model that is linear in the parameters  $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\theta}$ . What are  $\boldsymbol{y}$  and  $\boldsymbol{A}$  for this problem?

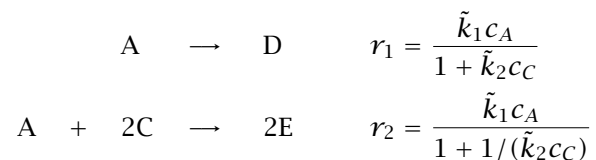
- (b) What are the least squares estimates of  $\ln(k)$  and  $n$ ? Plot the data and your model fit for  $r$  versus  $c_A$ . Plot the data and your model fit for  $\ln(r)$  versus  $\ln(c_A)$ . Comment on the quality of the fit of the model to the data.
- (c) Notice in this problem we are *not* told the value of the measurement error variance  $\sigma^2$ . Look at your model fit to the  $\ln(r)$  versus  $\ln(c_A)$  data. How might you estimate the variance of the measurement error in  $\ln(r)$  from the residuals of the model fit? What is a reasonable value of  $\sigma^2$  for your measurement error based on the residuals?
- (d) Given this value of  $\sigma^2$ , plot the 95% confidence ellipse for the least squares estimates of  $\ln(k)$  and  $n$ . Your least squares estimates given above should be the center of the 95% confidence ellipse. Are the parameters  $\ln(k)$  and  $n$  strongly correlated in this model? Explain why or why not based on your plot.

**Exercise 9: Choose the best reaction model**

We are considering two different kinetic models to represent some experimental pilot plant reactor data.

**Model 1**

Notice that the kinetic events in this model are elementary reactions and the rate expressions are consistent with the reaction stoichiometry. We call such a model a *reaction mechanism*.

**Model 2**

Notice that the kinetic events in this model are *not* elementary reactions. The rate expressions are not given by molecular collision arguments. But we can still estimate the model parameters from experimental data.

The data from a well-stirred batch reactor are shown in Figure 17. No B is initially charged to the reactor.

- Fit parameters  $k_1, k_2, k_3$  for Model 1 to the data. What are the best parameters and the 95% confidence intervals for this model. Plot the model fit and the residuals for Model 1.
- Fit parameters  $\tilde{k}_1, \tilde{k}_2$  for Model 2 to the data. What are the best parameters and the 95% confidence intervals for this model? Plot the model fit and the residuals for Model 2.
- Which model would you recommend for this system? Justify your choice.

**Exercise 10: Noise in the Antoine model measurements**

The data in Example 5 were seen to fit the Antoine model very well. These values may have been smoothed prior to being reported.

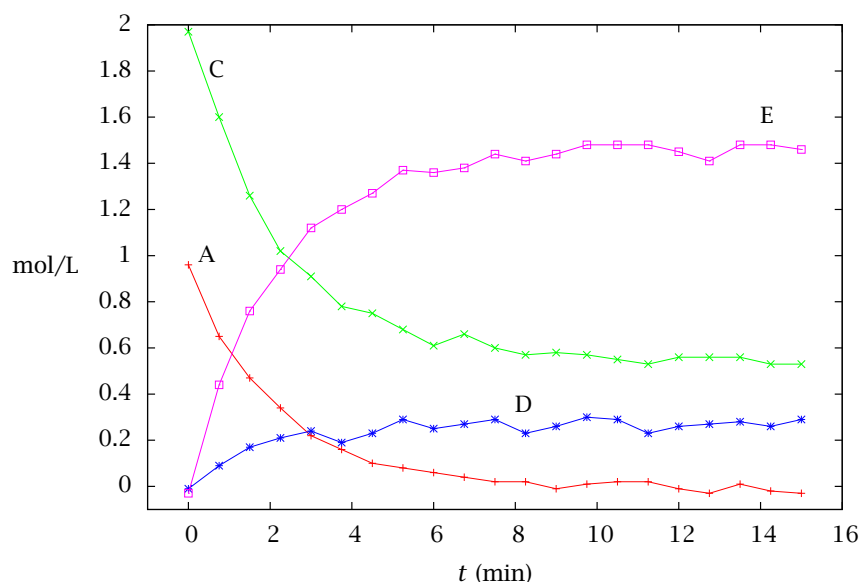
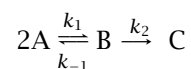


Figure 17: Concentrations of species A, C, D, and E versus time.

- (a) Add simulated experimental errors to the measurements: Look up the MATLAB function `rand` and generate a vector of the same length as your vector of vapor pressure observations containing random numbers distributed uniformly on the interval  $[0, 1]$ . Convert these to a distribution on the interval  $[-1, 1]$  by multiplying each by 2 and subtracting 1. Then add simulated errors of  $\approx \pm 15\%$  to the vector of vapor pressure values by adding 0.15 times the  $[-1, 1]$  random numbers to the natural log of the reported values.
- (b) Repeat Example 5 to fit the model to the more-noisy data. Plot the model predictions along with the noisy data.

### Exercise 11: Estimating kinetic parameters for reversible series reactions

The following reactions take place in a well-mixed batch reactor. The kinetics are described by the system of ODEs shown.



$$\begin{aligned}\frac{dc_A}{dt} &= -2k_1c_A^2 + 2k_{-1}c_B \\ \frac{dc_B}{dt} &= k_1c_A^2 - k_{-1}c_B - k_2c_B \\ \frac{dc_C}{dt} &= k_2c_B\end{aligned}$$

$$k_1 = 6 \times 10^{-3}, \quad k_{-1} = 1 \times 10^{-2}, \quad k_2 = 5 \times 10^{-2}$$

- (a) First, manufacture some simulated experimental data:
- Using the parameter values shown, run `ode15s` to obtain model predictions for the three state variables  $c_A$ ,  $c_B$ ,  $c_C$  at 15 time values evenly spaced from  $t = 0$  to  $t = 100$ . Use initial values  $c_{A0} = 10$ ,  $c_{B0} = 0$ ,  $c_{C0} = 0$ .
  - Using the procedure described in Exercise 10, add simulated errors of  $\approx \pm 15\%$  to the values of  $\mathcal{Y} = \{c_A, c_B, c_C\}$  and errors of  $\approx \pm 1$  time unit to the values of  $t$ .
- (b) Use the nonlinear least-squares estimation procedure of Section 4 to estimate values of the parameters  $\theta = \{k_1, k_{-1}, k_2\}$  from the simulated data. Use the `nlls` function. (You may or may not wish to make modifications to it, depending on your approach.)
- You will need to construct a function `f(xhat, theta)` to pass to `nlls`. `f` will call `ode15s` to calculate the model predictions  $\hat{\mathcal{Y}}$  for  $c_A$ ,  $c_B$ ,  $c_C$ , given arguments  $\hat{\mathcal{X}}$  ( $= t$ ) and  $\theta$ .
- (Note this calculation may take a few minutes, since it involves many repeated ODE integrations.)
- (c) Plot a curve of the predictions of the model using these estimated parameters over the range  $0 \leq t \leq 100$  along with the simulated data points. Also plot the curve given by the original model parameters.

## References

- G. E. P. Box and G. C. Tiao. *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, Massachusetts, first edition, 1973.
- G. E. P. Box, W. G. Hunter, and J. S. Hunter. *Statistics for Experimenters*. John Wiley & Sons, New York, 1978.
- R. H. Perry, D. W. Green, and J. O. Maloney, editors. *Perry's Chemical Engineers' Handbook*. McGraw-Hill, New York, seventh edition, 1997.
- J. B. Rawlings and J. G. Ekerdt. *Chemical Reactor Analysis and Design Fundamentals*. Nob Hill Publishing, Madison, WI, 2nd edition, 2012. 664 pages, ISBN 978-0-9759377-2-3.

W. E. Stewart, T. L. Henson, and G. E. P. Box. Model discrimination and criticism with single-response data. *AIChE J.*, 42(11):3055-3062, 1996.

W. E. Stewart, Y. Shon, and G. E. P. Box. Discrimination and goodness of fit of multiresponse mechanistic models. *AIChE J.*, 44(6):1404-1412, 1998.