10.34, Numerical Methods Applied to Chemical Engineering
Professor William H. Green
Lecture #25: Conclude Models vs. Data

Parameter Estimation

1) Model definition/Formulation: choosing $\underline{\theta}$
2) Compile/Assess what you already knew before adjusting $\underline{\theta}$
    a. estimate parameters, error bars
    b. initial guess $\underline{\theta}$
3) Adjust $\underline{\theta}$
    a. Determine if Model is Consistent with Data: if inconsistent, you have learned something important
    b. $\underline{\theta}_{bestfit}$ at $\underline{\theta}_{localminima}$ of $\chi^2(\underline{\theta})$
4) Refine $p(\underline{\theta})$: narrow range for the parameters
    a. summarize what we have learned

4-STEP PROCESS IS ALSO CALLED: "LEAST-SQUARES FITTING"

1) Repeat measurement "i" $N_{replicates}$ times $Y_i^{(j)}$ $j = 1, N_{replicates}$

$$Y_i^{data} = \frac{\sum_{j=1}^{N_{rep}} Y_i^{(j)}}{N_{rep}} \qquad \sigma_i = \sqrt{\frac{\sum_{j=1}^{N_{rep}} \left(Y_i^{(j)} - Y_i^{data}\right)^2}{N_{rep} - 1}}$$

$$\chi^2 = \sum_{i=1}^{N_{data}} \left(\frac{Y_i^{data} - Y_i^{\text{model}}(\underline{\theta})}{\sigma_i} \right)^2$$

Quantitative Definition of "Consistent"

calc $\chi^2(\underline{\theta}; \underline{Y}^{data})$

Prob(measure $\underline{Y}^{data}$ with $\chi^2 \geq \chi^2_{expt}$) {if small, unlikely that our model is right}

$$\rightarrow \int_{x^2_{\text{expt}}}^{\infty} \frac{t^{\frac{v}{2}-1} e^{-\frac{t}{2}}}{2^{N/2}\Gamma\left(\frac{v}{2}\right)} dt = \frac{\Gamma\left(\frac{v}{2}, \frac{x^2_{\text{expt}}}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}$$

$v = N_{data} - N_{params\_adjusted}$

GAMMA FUNCTION

If $\dfrac{\Gamma\left(\frac{\nu}{2},\frac{\chi^2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} < 0.01$ → very unlikely model is consistent with data

$\chi^2(\theta) < \chi^2_{max}$ for consistency

If you have more data, you have more confidence. Need lots more data than number of $\underline{\theta}$'s.

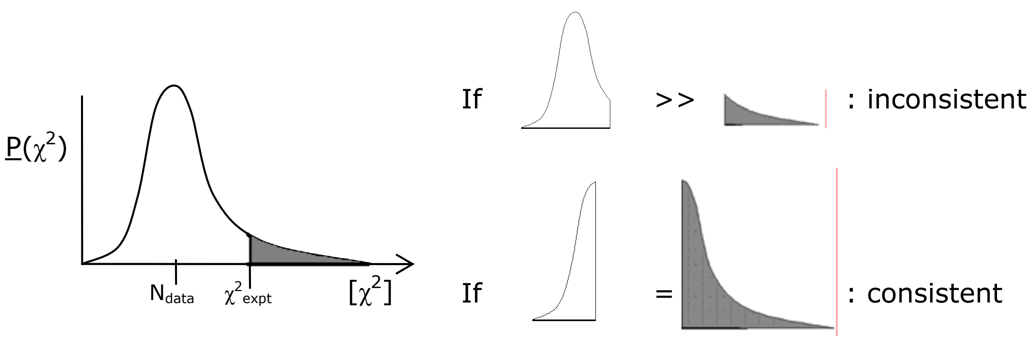
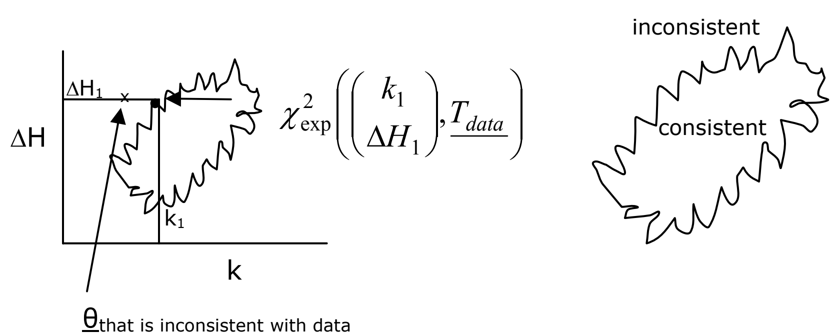


Figure 1. Chi-squared distribution tests.

MatLab: chi2cdf.m



$\chi^2_{exp}\left(\begin{pmatrix}k_1\\ \Delta H_1\end{pmatrix}, \underline{T_{data}}\right)$

Figure 2. An example of two parameter fitting.

neglect

$\chi^2(\underline{\theta}) = \chi^2(\underline{\theta}_{bestfit}) + \frac{1}{2}(\underline{\theta} - \underline{\theta}_{best})^T * \underline{\underline{H}}(\underline{\theta}_{best}) * (\underline{\theta} - \underline{\theta}_{best}) + O(\Delta\theta^3)$

$\underline{\underline{H}} \approx \underline{\underline{J}}^T \underline{\underline{J}}$ $\qquad J_{in} = \left.\dfrac{\partial Y_i^{\text{model}}}{\partial \theta_n}\right|_{\underline{\theta}_{best}}$

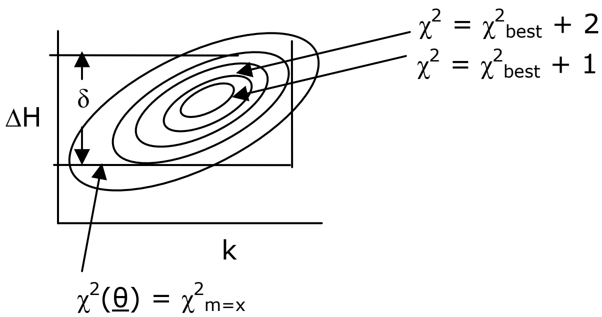


Figure 3. Contours around best fit.

People want these contours to be circles

Range of parameters that are acceptable: $\Delta H \pm \delta(\Delta H)$

$$k = k_{best} + \delta k$$

Covariance Matrix

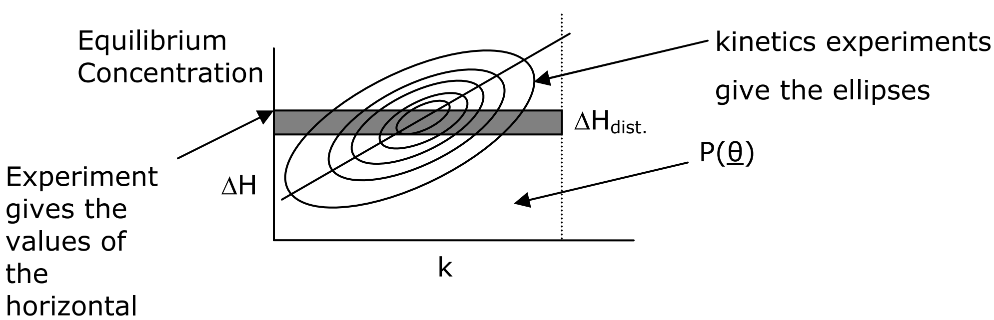


Figure 4. Each experiment tells you about different "cuts" or ellipses and where they all intersect is the answer.

BAYESIAN: store $p(\underline{\theta})$

STORE ALL THE DATA: Thermochemistry Active Tables, PrIMe