10.34, Numerical Methods Applied to Chemical Engineering
Professor William H. Green
Lecture #22: Introduction: Models vs. Data.

Models vs. Data

Engineers think of practical problems and efficient solutions from the top-down.

Scientists use a micro-view and can neglect the big picture in the bottom-up analysis.

Models are always wrong.

But experiments also never match.

more important

$\underline{Y}_{model}(\underline{x}(\pm\delta\underline{x}),\ \theta\pm\delta\theta)$

- $\underline{x}$: "knobs" parameters in model we can physically adjust
- θ: all other parameters that affect results (cannot control) — more important
- $\pm\delta\theta$: we generally know bounds on $\underline{\theta}$ ← prior information about θ

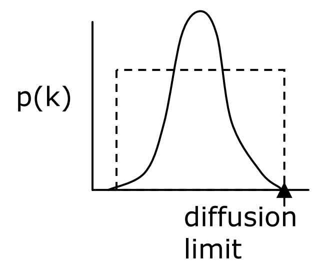
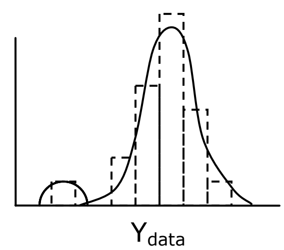


Figure 1. Normal distribution.

$\underline{Y}_{data}^{(1)}(\underline{x})$

$\underline{Y}_{data}^{(2)}(\underline{x})$

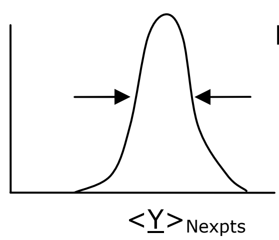


seldom have sampling capable of making true distribution curve

Average Value

$<\underline{Y}_{data}>_{Nexpts}$

$\sigma_{exp.data|Nexpts}$



$P(<Y_{data}>,\sigma_{data}) \approx$

$$\frac{1}{\sigma\sqrt{2\pi}}\exp\left[\frac{-\left(<Y>-<Y>_{true}\right)^2}{2\sigma_{mean}^2}\right]$$

$$\sigma_{mean} \approx \sigma_{data} \Big/ \sqrt{N_{\exp ts}}$$

Figure 3. Normal distribution curve showing 1 standard deviation.

Why do we compare the model to data?

- Is The Model Consistent With The Data?

$|<Y_{data}> - Y_{model}| >> \sigma_{mean}$ means Inconsistent (akin to confidence interval: $CI \cong t_{\alpha,\nu} \cdot \sigma_{mean}$)

- Model Discrimination

  Often more than 1 model: If they are consistent, would like to be able to pick one closer to the data or say that either model works fine.

- Parameter Refinement

  How narrow can you make the range on θ?

- Experimental Design

  Identify which $\{\theta_i\}$ are not determined by data. A few θ_i often control the fit. Some θ_i cannot be determined well by experiment (poorly conditioned matrices).

Introduction to Chi-Squared Analysis

Assume all error is Gaussian.

$$\chi^2 \equiv \sum_{i}^{N_{action}} \frac{\left| < \underline{Y}_n > (\underline{x}_n) - \underline{Y}_{model}(x_n, \underline{\theta}) \right|^2}{\sigma_n^2} \sim N_{data}$$ for the "true" model

$$\sigma_{mean} \sim \frac{S.D._{exp}}{\sqrt{N_{expts}}}$$

parameter refinement: $\chi^2(\underline{\theta})$ ← minimize χ^2 by changing $\underline{\theta}$

experimental design: derivatives of χ^2 with respect to $\underline{\theta}$.

Bayesian View

Prior knowledge p($\underline{\theta}$, σ) → posterior p($\underline{\theta}$, σ; $\underline{Y}_{data}$)

More knowledge after experiment. Use to narrow error bars.

Conditional Probability

$$P(E_1 \cap E_2) = P(E_1)\underbrace{P(E_2 \mid E_1)}_{\text{if independent: } P(E_2)}$$: probability of E_2 knowing E_1 happened (correlation)

$$P_{posterior}(\underline{\theta},\sigma|\underline{Y}_{data}) = \frac{\overbrace{P(\underline{Y}_{data}|\underline{\theta},\sigma)}^{\text{model}}\overbrace{P(\underline{\theta},\sigma)}^{\text{prior}}}{\underbrace{\int\int d\theta\, d\sigma\, P(\underline{Y}|\underline{\theta},\sigma)P(\underline{\theta},\sigma)}_{\text{normalize}} \approx P(\underline{Y}_{data})}$$

$P(\underline{Y}_{data}|\underline{\theta},\sigma)$: probability of observing $\underline{Y}_{data}$ we really observed if $\underline{\theta}$, σ are true values.

We do not know $\underline{\theta}$ and σ exactly.