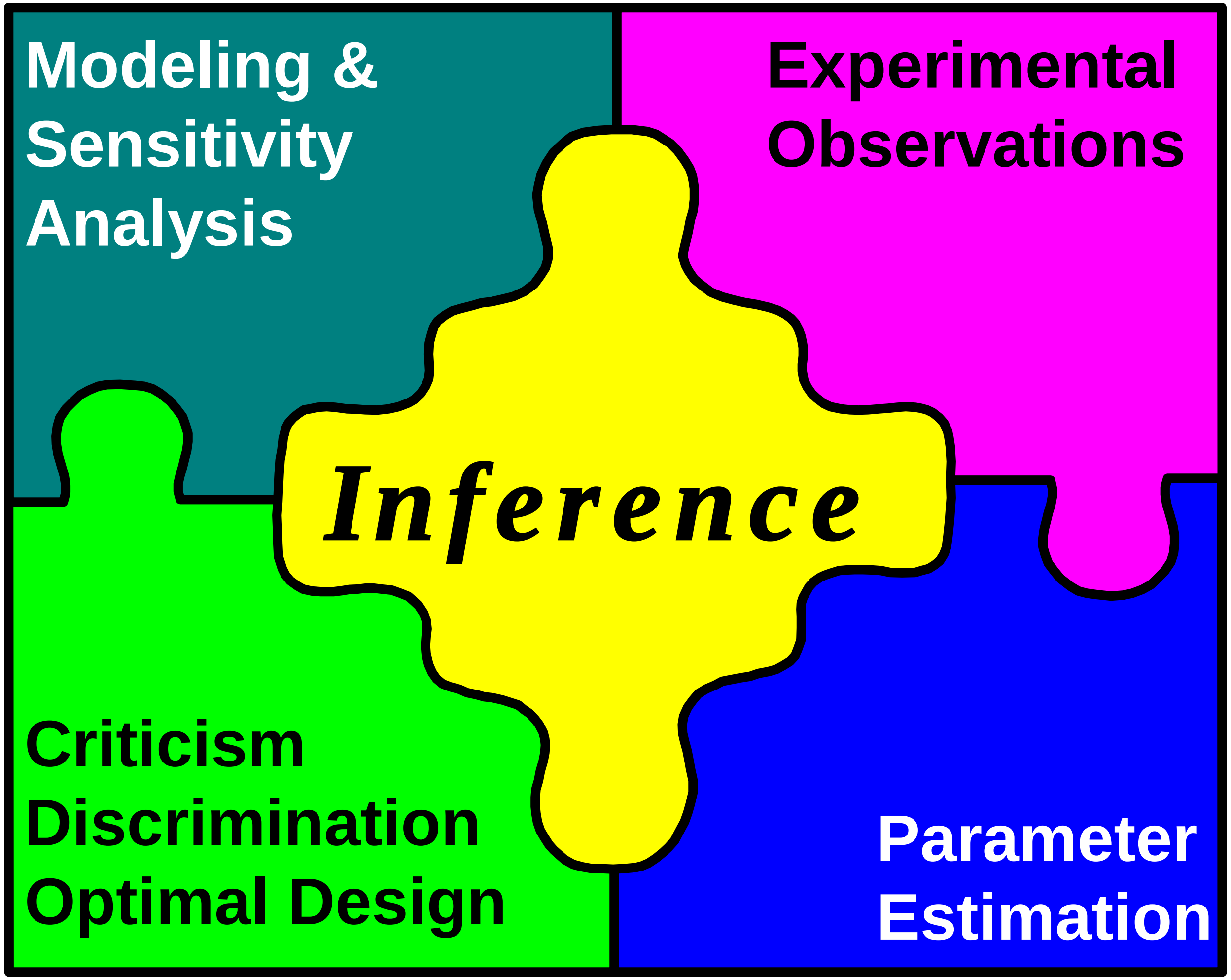
Modeling and Nonlinear Parameter Estimation
Discrimination-Identification and Criticism
Prediction under Uncertainty

@Michael Caracotsios, 2002

The engineering science puzzle



About models & inference

to produce a satisfactory model it is not necessary that it be exactly right…
(no model or procedure is perfect)
but rather that it not be grossly wrong in the context in which it is to be used…

By statistical inference we mean inference about the state of nature made in terms of probability; usually the state of nature is described by the value of one or more parameters. Such a parameter θ could, for instance, be the rate constant of a chemical reaction or the thermal conductivity of a certain alloy. Thus a solution to the inference problem is supplied by a posterior distribution $p(\theta|y)$ from the data y given a relevant prior state of knowledge represented by $p(\theta)$

About models and inference

Consider the reaction and data below. Estimate the uncertainty in the calculation of the catalyst mass required for production of B.

Design experiments to improve the estimation of k_2 and k_3.

What can you do about k_4 ?

$$A \underset{k_3}{\overset{k_1}{\rightleftharpoons}} B \underset{k_4}{\overset{k_2}{\rightleftharpoons}} C$$

$$\ln k_1 = -1.545161 \pm 0.11$$

$$\ln k_2 = -0.711056 \pm 0.1394$$

$$\ln k_3 = -4.164588 \pm 3.501$$

$$\ln k_4 = -20.0$$

About the basics

- Adequacy of the Expectation Function (Model)
  - Lack-of-fit and Discrimination
- Constancy of error variance from one observation to another
  - Transformation of Data
- Normality of the distributions of the observations
  - Exponential Power Distributions
- Independence of these distributions
  - Violation leads to dramatic consequences
  - Relaxation leads to time series and dynamic models
- Lack of “aberrant” observations
  - Anomalies in the experimental setup
  - Observations generated from alternative models with large bias or very large variance

Explicit Expectation

Error and Expectation

$y_u = E(y_u) + \varepsilon_u$

$E(y_u) = \eta\left(\xi_u; \theta\right)$ $\quad \dfrac{\partial y_u}{\partial \theta_k} = \dfrac{\partial \eta}{\partial \theta_k}$

$\varepsilon = N_n(0, \sigma^2 \mathbf{I}_n)$

Prediction and Uncertainty

$p = \eta\left(\xi; \hat{\theta}\right) \pm \Delta p$

Example: Differential reactor

$$2A \rightleftharpoons B + C$$

$$y_1 = r = \frac{k_1\left(p_A - \sqrt{\frac{p_B p_C}{K}}\right)}{1 + K_A p_A p_C + K_B p_B + K_C p_C}$$

Implicit Expectation

Error and Expectation

$$\mathbf{E}(t,\mathbf{u};\theta)\frac{\partial \mathbf{u}}{\partial t}=\mathbf{F}\left(t,\mathbf{x},\nabla\mathbf{u},\nabla\left(\Im\nabla\mathbf{u}\right);\theta\right)$$

$$\mathbf{y}_u = E(\mathbf{y}_u)+\varepsilon_u$$

$$E(\mathbf{y}_u)=\eta\left(\xi_u,\mathbf{u}_u;\theta\right) \quad \frac{\partial y_u}{\partial \theta_k}=\frac{\partial \eta}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \theta_k}+\frac{\partial \eta}{\partial \theta_k}$$

$$\varepsilon = \mathbf{N}_n(\mathbf{0},\sigma^2\mathbf{I}_n)$$

Prediction and Uncertainty

$$\mathrm{p}=\eta\left(\xi;\hat{\theta}\right)\pm\Delta\mathrm{p}$$

Example: Integral reactor

$$2A \rightleftharpoons B + C \qquad r = \frac{k_1\left(p_A - \sqrt{\frac{p_B p_C}{K}}\right)}{1 + K_A p_A p_C + K_B p_B + K_C p_C}$$

$$\frac{\partial F_A}{\partial V} = \Re_A = -2r \quad \frac{\partial F_B}{\partial V} = \Re_B = r \quad \frac{\partial F_C}{\partial V} = \Re_C = r \quad \frac{\partial F_I}{\partial V} = 0$$

$$\frac{\partial F\tilde{H}}{\partial V} = U\alpha\left(T_a - T\right) \qquad \tilde{H}\left(\mathbf{F}, P, T\right) = 0$$

$$y_1 = \frac{F_{Ao} - F_A}{F_{Ao}} \quad y_2 = \frac{M_B F_B}{M_A F_A + M_B F_B + M_C F_C} \quad y_3 = T$$

Solution of Science and Engineering Models

- Algebraic Models
- Initial Value Models
- Boundary Value Models
- Initial-Boundary Value Models
- Models with the 3rd Dimension

Algebraic Models

Systems of Algebraic Equations

$$\mathbf{F}(\mathbf{u};\theta) = 0$$

$$\mathbf{J}(\mathbf{u}_k;\theta)\Delta\mathbf{u}_k = -\mathbf{F}(\mathbf{u}_k;\theta)$$

$$\mathbf{u}_{k+1} = \mathbf{u}_k + \lambda\Delta\mathbf{u}_k \qquad 0 < \lambda \leq 1$$

$$\left\|\mathbf{J}^{-1}(\mathbf{u}_k)\mathbf{F}(\mathbf{u}_{k+1})\right\| \leq \left\|\mathbf{J}^{-1}(\mathbf{u}_k)\mathbf{F}(\mathbf{u}_k)\right\|$$

and

$$\left\|\mathbf{F}(\mathbf{u}_{k+1})\right\| \leq \left\|\mathbf{F}(\mathbf{u}_k)\right\|$$

$$\mathbf{u}(\theta_{\mathbf{new}}) = \mathbf{u}(\theta_{\mathbf{old}}) + \frac{\partial\mathbf{u}}{\partial\theta}(\theta_{\mathbf{new}} - \theta_{\mathbf{old}})$$

$$\mathbf{J}(\mathbf{u}^*;\theta)\frac{\partial\mathbf{u}^*}{\partial\theta} = -\frac{\partial\mathbf{F}(\mathbf{u}^*;\theta)}{\partial\theta}$$

$$\mathbf{J}(\mathbf{u}^*;\theta)\frac{\partial\mathbf{u}^*}{\partial\ln\theta} = -\theta\frac{\partial\mathbf{F}(\mathbf{u}^*;\theta)}{\partial\theta}$$

- **Newton's Method**
  - **Adaptive strategy**
  - **Non-Negativity Conditions**
  - **Convergence form poor guess**
  - **Homotopy techniques**
- **Sensitivity Analysis**
  - **un-normalized coefficients**
  - **semi-normalized**
  - **Critical for the estimation**

Initial Value Models

Systems of Differential Equations

$$\mathbf{E}\mathbf{u}' = \mathbf{E}\frac{d\mathbf{u}}{dt} = \mathbf{F}(\mathbf{u};\phi,\theta)$$

$$\mathbf{u}'(t_{n+1}) = \frac{\alpha}{h_{n+1}}\mathbf{u}(t_{n+1}) + \mathbf{c}_{n+1} \qquad \alpha = 1 + \frac{1}{2} + \cdots + \frac{1}{k}$$

$$\mathbf{E}\frac{\alpha}{h_{n+1}}\mathbf{u}(t_{n+1}) + \mathbf{E}\mathbf{c}_n - \mathbf{F}(\mathbf{u}(t_{n+1});\ \phi,\theta) = \mathbf{0}$$

$$\mathbf{E}\frac{d}{dt}\frac{\partial\mathbf{u}}{\partial\theta} = \frac{\partial\mathbf{F}(\mathbf{u};\phi,\theta)}{\partial\mathbf{u}}\frac{\partial\mathbf{u}}{\partial\theta} + \frac{\partial\mathbf{F}(\mathbf{u};\phi,\theta)}{\partial\theta}$$

$$\mathbf{E}\frac{d}{dt}\frac{\partial\mathbf{u}}{\partial\ln\theta} = \frac{\partial\mathbf{F}(\mathbf{u};\phi,\theta)}{\partial\mathbf{u}}\frac{\partial\mathbf{u}}{\partial\ln\theta} + \theta\frac{\partial\mathbf{F}(\mathbf{u};\phi,\theta)}{\partial\theta}$$

- **Fixed Leading Coefficient**
  - **Adaptive Newton's Method**
  - **Non-Negativity Conditions**
- **Parametric Continuation**
- **Sensitivity Analysis**
  - **Un-normalized coefficients**
  - **Semi-normalized coefficients**
  - **Critical for estimation**
- **Explicit discontinuities**
- **Implicit discontinuities**

Boundary Value Models

Model at the Interior of Domain

$$\mathbf{F}(x,\mathbf{u},\mathbf{u}_x,\mathbf{u}_{xx};\theta)=0 \quad @\ \alpha<x<b$$

Model at the Left Boundary

$$\mathbf{g}(\alpha,\mathbf{u},\mathbf{u}_x;\theta)=0 \quad @\ x=\alpha$$

Model at the Right Boundary

$$\mathbf{g}(b,\mathbf{u},\mathbf{u}_x;\theta)=0 \quad @\ x=b$$

- **Discretization of the spatial derivatives**
  - **Finite Differences (uniform and non-uniform grids)**
  - **Global Orthogonal Collocation**
  - **Orthogonal Collocation on Finite Elements**

Initial-Boundary Value Models

Model at the Interior of Domain

$$\mathbf{E}\frac{\partial \mathbf{u}}{\partial t} = \mathbf{F}(t, x, \mathbf{u}, \mathbf{u}_x, \mathbf{u}_{xx}; \theta) = 0 \quad @ \ \alpha < x < b \qquad \mathbf{u}_x = \frac{\partial \mathbf{u}}{\partial x} \qquad \mathbf{u}_{xx} = \frac{1}{x^m}\frac{\partial}{\partial x}\left(x^m \frac{\partial \mathbf{u}}{\partial x}\right)$$

Model at the Left Boundary

$$\mathbf{E}\frac{\partial \mathbf{u}}{\partial t} = \mathbf{g}(t, \alpha, \mathbf{u}, \mathbf{u}_x; \theta) = 0 \quad @ \ x = \alpha$$

Model at the Right Boundary

$$\mathbf{E}\frac{\partial \mathbf{u}}{\partial t} = \mathbf{g}(t, b, \mathbf{u}, \mathbf{u}_x; \theta) = 0 \quad @ \ x = b$$

- **Discretization of the spatial derivatives**
  - **Finite Differences (uniform and non-uniform grids)**
  - **Global Orthogonal Collocation**
  - **Orthogonal Collocation on Finite Elements**
- **Fixed Leading Coefficient within the Method of Lines**

Models with the 3rd Dimension

$$-\frac{\partial F_A}{\partial V} = k_m a_{\upsilon}\left(C_A^b - C_A^s\Big|_{r=R}\right)$$

$$D_{\mathbf{eff}}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial C_A^s}{\partial r}\right) = \rho_s r_A$$

$$C_A = C_A\left(F_A, F_B, \ldots F, T, P\right)$$

$$z = 0 \qquad C_A^b = C_{A0}^b$$

$$r = 0 \qquad \frac{\partial C_A^s}{\partial r} = 0$$

$$r = R \qquad -D_{\mathbf{eff}}\frac{\partial C_A^s}{\partial r} = k_m\left(C_A^s\Big|_{r=R} - C_A^b\right)$$

Discretization of the 3rd dimension

- **Global Orthogonal Collocation**
- **Orthogonal Collocation on Finite Elements**

Simultaneous solution of the 3rd dimension residuals with the parent equations

Statistical Inferences

- Estimation
- Discrimination
- Criticism & Design
- Prediction
- Uncertainty

Statistical Inference

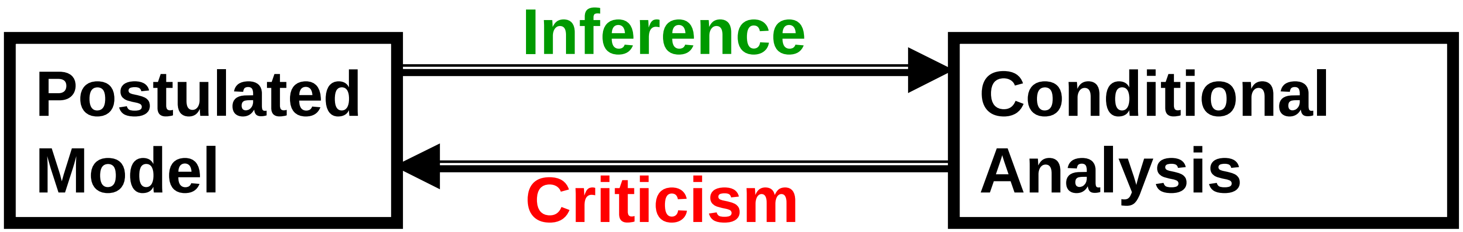
- **Sampling Theory**

  Inferences are made in terms of the sampling distributions of statistics, which are functions of the observations.The probabilities refer to the frequency with which different values of statistics (from sets of data other than those that have actually happened) could occur for some *fixed but unknown* values of the parameters. (Significance tests, confidence intervals, Neyman-Pearson theory of Hypothesis testing comprise the Sampling Theory approach.)

- **Bayes' Theorem**

  Inferences are based on probabilities associated with *different* values of parameters which could have given rise to the *fixed* set of data which has actually occurred. In calculating these probabilities we must make assumptions about *prior distributions* that express a state of knowledge or ignorance about the parameters.

Model Building



- **Inference [Sponsor]**
- Inference deals with the estimation of the parameters of the postulated model.
- **Criticism [Critic]**
- Criticism seeks to see if a fitted model is faulty

- Criticism seeks to see if a fitted model is faulty

SINGLE RESPONSE MECHANISTIC MODELS

Bayes' Theorem

The probability distribution for θ and σ posterior to the data **y** is proportional to the product of the distribution for θ and σ prior to the data and the likelihood for θ and σ given **y.**

$$p(\theta,\sigma\,|\,\mathbf{y}) = \ell(\theta,\sigma\,|\,\mathbf{y})\; p(\theta,\sigma)$$

$p(\theta,\sigma\,|\,\mathbf{y})$: **Posterior Distribution**

$\ell(\theta,\sigma\,|\,\mathbf{y})$: **Likelihood Function**

$p(\theta,\sigma)$: **Prior Distribution**

Non-Informative Prior Distribution

$$p(\theta,\sigma) \propto \sigma^{-1}$$

The critical assumptions

$$y_u = E(y_u) + \varepsilon_u$$

$$E(y_u) = \eta(\xi_u, \mathbf{u}_u; \theta)$$

- **The observations are independently distributed**
  - **Very sensitive assumption**
- **Each observation has the same variance σ^2**
  - **Equal weight for each observation**
  - **Data transformation (ex. Log, Inverse, Square Root)**
- **The observations are *Normally* distributed**
  - **Central limit theorem**
  - **Data are Spherically Normal**

Derivation of the Posterior Probability Distribution p(θ,σ|y)

$$y_u = \eta(\xi_u, \mathbf{u}_u; \theta) + \varepsilon_u$$

Define: $\mathbf{S}(\theta) = \sum_{u=1}^{n} \left(y_u - \eta(\xi_u, \mathbf{u}_u; \theta) \right)^2$

$$\ell(\theta, \sigma \mid \mathbf{y}) \propto \sigma^{-n} \exp\left(-\frac{1}{2\sigma^2} S(\theta) \right)$$

$$p(\theta, \sigma) \propto \sigma^{-1}$$

$$p(\theta, \sigma \mid \mathbf{y}) \propto \sigma^{-(n+1)} \exp\left(-\frac{1}{2\sigma^2} S(\theta) \right)$$

- **The posterior *p*(θ,σ|y) distribution function contains everything we know about the parameters θ**
- **Our objective is to maximize the posterior probability function**
- **This is done by minimizing S(θ) [also know as the sum of squares of residuals]**
- **Once the parameter θ is obtained we obtain the *marginal posterior* distribution for θ by integrating out the nuisance parameter σ**

Minimization of S(θ): Explicit Models

$$y_u = \eta(\xi_u;\theta) + \varepsilon_u$$

$$S(\theta) = \sum_{u=1}^{n} \left(y_u - \eta(\xi_u;\theta)\right)^T \left(y_u - \eta(\xi_u;\theta)\right)$$

$$\theta_L \le \theta \le \theta_U$$

Normal Equations

$$\mathbf{X}^T \mathbf{X}\left(\theta^{k+1} - \theta^k\right) = \mathbf{X}^T \mathbf{E}\left(\theta^k\right)$$

$$\mathbf{E}_u(\theta) = y_u - \eta(\xi_u;\theta)$$

$$\mathbf{X}_{ur} = \frac{\partial \mathbf{E}_u(\theta)}{\partial \theta_r} = -\frac{\partial \eta(\xi_u;\theta)}{\partial \theta_r}$$

Hessian Matrix and SQP/GRG Methods

$$\mathbf{X}^T \mathbf{X}\left(\theta^{k+1} - \theta^k\right) \approx \mathbf{B}_k\left(\theta^{k+1} - \theta^k\right) = -\nabla S\left(\theta^k\right) \quad \mathbf{B}_o = s\mathbf{I}$$

MINIMIZATION ALGORITHMS

- GregPlus in Athena uses the Newton or the Gauss-Newton Algorithm
- The Marquardt and Levenberg Algorithm
- Sequential Quadratic Programming [SQP] Algorithms
  - Popular in Simulation Software
- General Reduced Gradient [GRG] Algorithms
  - Excel Implementation

Minimization of S(θ): Implicit Models

$$\mathbf{E}\frac{d\mathbf{u}}{dt} = \mathbf{F}\left(t,\mathbf{u};\boldsymbol{\theta}\right)$$

$$y_u = \boldsymbol{\eta}(\boldsymbol{\xi}_u,\mathbf{u}_u;\boldsymbol{\theta}) + \boldsymbol{\varepsilon}_u$$

$$S\left(\boldsymbol{\theta}\right) = \sum_{u=1}^{n}\left(y_u - \boldsymbol{\eta}(\boldsymbol{\xi}_u,\mathbf{u}_u;\boldsymbol{\theta})\right)^T\left(y_u - \boldsymbol{\eta}(\boldsymbol{\xi}_u,\mathbf{u}_u;\boldsymbol{\theta})\right)$$

$$\boldsymbol{\theta}_L \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}_U$$

Normal Equations

$$\mathbf{X}^T\mathbf{X}\left(\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^k\right) = \mathbf{X}^T\mathbf{E}\left(\boldsymbol{\theta}^k\right)$$

$$\mathbf{E}_u\left(\boldsymbol{\theta}\right) = y_u - \boldsymbol{\eta}(\boldsymbol{\xi}_u,\mathbf{u}_u;\boldsymbol{\theta})$$

$$\mathbf{X}_{ur} = \frac{\partial\mathbf{E}_u\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}_r} = -\frac{\partial\boldsymbol{\eta}(\boldsymbol{\xi}_u,\mathbf{u}_u;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_r} - \frac{\partial\boldsymbol{\eta}(\boldsymbol{\xi}_u,\mathbf{u}_u;\boldsymbol{\theta})}{\partial\mathbf{u}_u}\frac{\partial\mathbf{u}_u}{\partial\boldsymbol{\theta}_r}$$

@ **END OF MINIMIZATION**

- **A set of estimated parameters θ,**
- **a set of constrained parameters,**
- **and a set of undetermined parameters**

- **For the set of the estimated parameters θ we may obtain the *marginal* *posterior* distribution for θ by integrating out the nuisance parameter σ**

Marginal Posterior Distribution

$$S(\theta) = S(\hat{\theta}) + (\theta - \hat{\theta})^T \mathbf{X}^T \mathbf{X} (\theta - \hat{\theta})$$

$$p(\theta \mid \mathbf{y}) = \frac{\Gamma\left[\frac{1}{2}(\nu + k)\right] \left|\mathbf{X}^T \mathbf{X}\right|^{\frac{1}{2}} s^{-k}}{\left[\Gamma\left(\frac{1}{2}\right)\right]^k \Gamma\left(\frac{1}{2}\nu\right) \left(\sqrt{\nu}\right)^k} \left[1 + \frac{(\theta - \hat{\theta})^T \mathbf{X}^T \mathbf{X} (\theta - \hat{\theta})}{\nu s^2}\right]^{-\frac{1}{2}(\nu + k)}$$

$$s^2 = \frac{1}{v} S(\hat{\theta}) \qquad \nu = n - k$$

$p(\theta \mid \mathbf{y})$ **is the multivariate** t **distribution (1954)**

$$p(\theta \mid \mathbf{y}) = t_k\left(\hat{\theta}, s^2 \left(\mathbf{X}^T \mathbf{X}\right)^{-1}, \nu\right)$$

Marginal Posterior Distribution for θ_i

$$p(\theta_i \mid \mathbf{y}) = \frac{\Gamma\left[\frac{1}{2}(\nu+1)\right]\left|\mathbf{C}_{ii}^{-1}\right|^{\frac{1}{2}} s^{-1}}{\left[\Gamma\left(\frac{1}{2}\right)\right]^{1}\Gamma\left(\frac{1}{2}\nu\right)\left(\sqrt{\nu}\right)^{1}}\left[1+\frac{\left(\theta_i-\hat{\theta}_i\right)^{T}\mathbf{C}_{ii}^{-1}\left(\theta_i-\hat{\theta}_i\right)}{\nu s^{2}}\right]^{-\frac{1}{2}(\nu+1)}$$

$$\mathbf{C} = \left(\mathbf{X}^{T}\mathbf{X}\right)^{-1}$$

$p(\theta_i \mid \mathbf{y})$ **is the univariate** t **distribution (1908)**

$$p(\theta_i \mid \mathbf{y}) = t\left(\hat{\theta}_i, s^{2}\mathbf{C}_{ii}, \nu\right)$$

Covariance and variance of parameters. HPD intervals with probability content 1-α

$$\mathbf{Cov}(\theta) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \frac{S(\hat{\theta})}{\nu}$$

$$\mathrm{C} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}$$

$$\mathbf{Var}(\theta_i) = \sigma^2_{\theta_i} = \mathbf{C}_{ii} \frac{S(\hat{\theta})}{\nu}$$

Degrees of freedom

$$\nu = n - k$$

$$-t\left(\frac{\alpha}{2}, \nu\right) \le \frac{\theta_i - \hat{\theta}_i}{\sigma_{\theta_i}} \le t\left(\frac{\alpha}{2}, \nu\right)$$

A useful choice is α=0.05, which gives 95% probability HPD interval for each of the estimated set of parameters

Marginal posterior distribution for parameter transformations

$$k = k_0 \exp\left(-\frac{E}{RT}\right)$$

$$\ln k = \ln k_B + \frac{E}{RT_B}\left(1 - \frac{T_B}{T}\right)$$

Option A:

$$\theta_1 = \ln k_B$$

$$\theta_2 = \frac{E}{RT_B}$$

$$k = \exp\left(\theta_1 + \theta_2\left(1 - \frac{T_B}{T}\right)\right)$$

Option B:

$$\theta_1 = k_B$$

$$\theta_2 = \frac{E}{RT_B}$$

$$k = \theta_1 \exp\left(\theta_2\left(1 - \frac{T_B}{T}\right)\right)$$

$p(\theta \mid \mathbf{y})$ **is the multivariate t-distribution**

$$p(\theta \mid \mathbf{y}) = t_k\left(\hat{\theta}, s^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}, \nu\right)$$

$$\phi = \phi(\theta) \qquad \mathbf{D} = \frac{\partial \phi}{\partial \theta}$$ **: Jacobian Matrix**

$p(\phi \mid \mathbf{y})$ **is the multivariate t-distribution**

$$p(\phi \mid \mathbf{y}) = t_k\left(\mathbf{D}\hat{\theta}, s^2\mathbf{D}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{D}^T, \nu\right)$$

Goodness-of-Fit Analysis Experimental and Lack-of-fit Error

$$S(\hat{\theta}) = S_E(\hat{\theta}) + S_L(\hat{\theta})$$

$S_E(\hat{\theta})$: **Experimental Error**

$S_L(\hat{\theta})$: **Lack-of-Fit Error**

$$S_E(\hat{\theta}) = \sum_{g=1}^{G}\sum_{u=1}^{N_g}\left(y_u - \bar{y}_G\right)^2 \qquad \bar{y}_G = \frac{\sum_{u=1}^{N_G} y_u}{N_G} \qquad \nu_E = \sum_{g=1}^{G}\left(N_g - 1\right)$$

$$S_L(\hat{\theta}) = S(\hat{\theta}) - S_E(\hat{\theta}) \qquad \nu_L = n - k - \nu_E$$

- **Extreme Cases**
- **Lack-of-Fit=0**
  - **Perfect Model**
- **Lack-of-Fit=S(θ)**
  - **Worst Model**

Lack-of-Fit Analysis
How Good is my Model?

$S(\hat{\theta}) = S_E(\hat{\theta}) + S_L(\hat{\theta})$

$s_E^2 = \frac{S_E(\hat{\theta})}{\nu_E}$: **Sample Variance for Experimental Error**

$s_L^2 = \frac{S_L(\hat{\theta})}{\nu_L}$: **Sample Variance for Lack-of-Fit Error**

$F = \frac{s_L^2}{s_E^2}$: **Ratio of Sample Variances**

$$\textbf{Pr}\left[\left(\frac{s_L^2}{s_E^2}\right) > F\right] = \alpha(F,\nu_L,\nu_E) = \frac{1}{B\left(\frac{\nu_L}{2},\frac{\nu_E}{2}\right)} \int_F^{\infty} \left(\frac{\nu_L}{\nu_E}\right)^{\frac{\nu_E}{2}} x^{\frac{\nu_L-2}{2}} \left(1+\frac{\nu_L}{\nu_E}x\right)^{-\frac{\nu_L+\nu_E}{2}} dx$$

Model Acceptance Test

$\alpha \le 0.01$ **casts doubt on the model**

$\alpha \approx 1.00$ **supports the model**

Discrimination amongst rival models

Unknown variance $\sigma_{\mathbf{J}}^2$

$S_{\mathbf{J}}(\hat{\theta}) = S_E(\hat{\theta}) + S_{L\mathbf{J}}(\hat{\theta})$ **of Model $M_{\mathbf{J}}$**

Posterior Probability of Model $M_{\mathbf{J}}$

$$p(M_{\mathbf{J}} \mid \mathbf{y}) \propto p(M_{\mathbf{J}})\, 2^{-\frac{k_j}{2}} \left(S_{\mathbf{J}}(\hat{\theta})\right)^{-\frac{\nu_E}{2}}$$

Posterior Probability Share held by Model $M_{\mathbf{J}}$

$$\pi(M_{\mathbf{J}} \mid \mathbf{y}) = \frac{p(M_{\mathbf{J}} \mid \mathbf{y})}{\sum_{\mathbf{J}=1}^{\mathbf{M}} p(M_{\mathbf{J}} \mid \mathbf{y})}$$

Optimal Experimental Design for Parameter Estimation

Consider the function $Q(\theta)$:

$$Q(\theta)=\left(\theta-\hat{\theta}\right)^T \mathbf{X}^T\mathbf{X}\left(\theta-\hat{\theta}\right)$$

Then:

$\frac{Q(\theta)}{ks^2}=F(k,\nu,\alpha)$ is distributed a *posteriori* as F

$$\nu=n-k$$

$$s^2=\frac{1}{\nu}S(\hat{\theta})$$

Volume of the Ellipsoid $Q(\theta)$

$$\text{Vol}\left(Q(\theta)\right)\propto\sqrt{\det\left(\mathbf{X}^T\mathbf{X}\right)^{-1}}$$

Trace of the inverse

$$\text{Tr}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}=\sum_{i=1}^{k}\left(\mathbf{X}^T\mathbf{X}\right)_{ii}^{-1}$$

A useful choice is α=0.05, which gives 95% probability HPD region for the estimated set of parameters

Optimal Design Basis

Minimization of the ellipsoid volume

Minimization of the trace of the inverse of the Hessian Matrix

Optimal Experimental Design for Model Discrimination

The Concept[Kullback's Total Measure of Info]

Maximize [Entropy at Input-Entropy at Output]

Discrimination Criterion $D(\theta)$

$$D(\theta) = -\sum_{i=1}^{m} \sum_{j=i+1}^{m} \Pi_{i,n-1} \Pi_{j,n-1} \left(\int p_i \ln \frac{p_i}{p_j} dy_n + \int p_j \ln \frac{p_j}{p_i} dy_n \right)$$

where

$\Pi_{i,n-1}$: **probability of model i after n-1 observations**

p_i: **probability density function of the n-th observation under model i**

Optimal Experimental Design for Model Discrimination

$$D(\theta) = -\sum_{i=1}^{m}\sum_{j=i+1}^{m}\Pi_{i,n-1}\Pi_{j,n-1}\left(\int p_i \ln\frac{p_i}{p_j}dy_n + \int p_j \ln\frac{p_j}{p_i}dy_n\right)$$

$$p_i\left(y_n\right) = \frac{\left(s^i\sqrt{1+b^i}\right)^{-1}}{B\left(\frac{\nu^i}{2},\frac{1}{2}\right)\sqrt{\nu^i}}\left[1+\frac{1}{\nu^i}\left(\frac{y_n-\tilde{y}_n^i}{s^i\sqrt{1+b^i}}\right)^2\right]^{-\frac{\nu^i+1}{2}}$$

$$b^i = \mathbf{x}_n^i\left(\mathbf{X}_i^T\mathbf{X}_i\right)\mathbf{x}_n^{i\,T}$$

MULTI-RESPONSE MECHANISTIC MODELS

Bayes' Theorem

The probability distribution for θ and Σ posterior to the data $\mathbf{y}$ is proportional to the product of the distribution for θ and Σ prior to the data and the likelihood for θ and Σ given $\mathbf{y}$.

$$p(\theta, \Sigma \mid \mathbf{y}) = \ell(\theta, \Sigma \mid \mathbf{y})\, p(\theta, \Sigma)$$

$p(\theta, \Sigma \mid \mathbf{y})$: **Posterior Distribution**

$\ell(\theta, \Sigma \mid \mathbf{y})$: **Likelihood Function**

$p(\theta, \Sigma)$: **Prior Distribution**

Non-Informative Prior Distribution

$$p(\theta, \Sigma) \propto \Sigma^{-\frac{m+1}{2}}$$

Derivation of the Posterior Probability Distribution p(θ,Σ|y)

$$\mathbf{y}_u = \eta(\xi_u, \mathbf{u}_u; \theta) + \boldsymbol{\varepsilon}_u$$

Define: $$\upsilon_{ij}(\theta) = \sum_{u=1}^{n} \left(\mathbf{y}_{ui} - \eta(\xi_u, \mathbf{u}_u; \theta)\right)\left(\mathbf{y}_{uj} - \eta(\xi_u, \mathbf{u}_u; \theta)\right)$$

$$\ell(\boldsymbol{\theta}, \Sigma \mid \mathbf{y}) \propto \Sigma^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\mathbf{tr}\left[\Sigma^{-1}\upsilon(\theta)\right]\right)$$

$$p(\boldsymbol{\theta}, \boldsymbol{\sigma}) \propto \Sigma^{-\frac{m+1}{2}}$$

$$p(\boldsymbol{\theta}, \Sigma \mid \mathbf{y}) \propto \Sigma^{-\frac{n+m+1}{2}} \exp\left(-\frac{1}{2}\mathbf{tr}\left[\Sigma^{-1}\upsilon(\theta)\right]\right)$$

TASKS AT HAND

- **The posterior *p*(θ,Σ|y) distribution function contains everything we know about the parameters θ**
- **Our objective is to maximize the posterior probability function**
- **This is done by using the conditional maximum-density estimate for Σ at each value of θ.**

Derivation of the Posterior Probability Distribution p(θ,Σ|y)

$$\mathbf{y}_u = \eta(\xi_u, \mathbf{u}_u; \theta) + \boldsymbol{\varepsilon}_u$$

Moment Matrix of Residuals

$$\upsilon_{ij}(\theta) = \sum_{u=1}^{n} \left(\mathbf{y}_{ui} - \eta_j(\xi_u, \mathbf{u}_u; \theta)\right)\left(\mathbf{y}_{uj} - \eta_j(\xi_u, \mathbf{u}_u; \theta)\right)$$

$$p(\theta \mid \mathbf{y}) \propto \left|\upsilon(\theta)\right|^{-\frac{n+m+1}{2}}$$

$$\hat{\Sigma} = \frac{\upsilon(\hat{\theta})}{n+m+1}$$

Definition of the Objective

$$S(\theta) = -2 \ln p(\theta \mid \mathbf{y}) + c \Rightarrow$$

$$S(\theta) = (n+m+1) \ln \left|\upsilon(\theta)\right|$$

TASKS AT HAND

- **Once the parameter Σ is obtained we obtain the *marginal posterior* distribution for θ.**
- **We then proceed to maximize the posterior density for the model parameters**
- **This is accomplished by minimizing the logarithm of the determinant of the moment matrix of the residuals, i.e., the function *S*(θ).**

Minimization of the Objective Function

Model: $\mathbf{y}_u = f_u(\mathbf{x}_u, \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_u$ **Moment Matrix:** $\upsilon_{ij}(\boldsymbol{\theta}) = \sum_{u=1}^{n} \left(\mathbf{y}_{ui} - f_i(\mathbf{x}_u, \boldsymbol{\theta})\right)\left(\mathbf{y}_{uj} - f_j(\mathbf{x}_u, \boldsymbol{\theta})\right)$

Objective Function: $S(\boldsymbol{\theta}) = (n+m+1)\ln\left|\boldsymbol{\upsilon}(\boldsymbol{\theta})\right|$

Normal Equations: $\mathbf{A}\left(\boldsymbol{\theta}^{k+1} - \boldsymbol{\theta}^{k}\right) = \nabla S\left(\boldsymbol{\theta}^{k}\right)$

Gradient: $\nabla S(\boldsymbol{\theta})_r = \dfrac{\partial S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_r} = (n+m+1)\sum_{i}\sum_{j\le i}\left(2-\boldsymbol{\delta}_{ij}\right)\boldsymbol{\upsilon}^{ij}\dfrac{\partial \boldsymbol{\upsilon}_{ij}}{\partial \boldsymbol{\theta}_r}$

Hessian:

$$\mathrm{A}_{rs} = (n+m+1)\sum_{i}\sum_{j\le i}\left(2-\boldsymbol{\delta}_{ij}\right)\boldsymbol{\upsilon}^{ij}\frac{\partial^2 \boldsymbol{\upsilon}_{ij}}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_s} - \frac{1}{2}(n+m+1)\sum_{i}\sum_{j\le i}\sum_{k}\sum_{\ell\le k}\left(2-\boldsymbol{\delta}_{ij}\right)\left(2-\boldsymbol{\delta}_{k\ell}\right)\left[\boldsymbol{\upsilon}^{ik}\boldsymbol{\upsilon}^{\ell j} + \boldsymbol{\upsilon}^{i\ell}\boldsymbol{\upsilon}^{kj}\right]\frac{\partial \boldsymbol{\upsilon}_{k\ell}}{\partial \boldsymbol{\theta}_s}\frac{\partial \boldsymbol{\upsilon}_{ij}}{\partial \boldsymbol{\theta}_r}$$

$$\frac{\partial \boldsymbol{\upsilon}_{ij}}{\partial \boldsymbol{\theta}_r} = -\sum_{u}\left[\frac{\partial f_{ui}}{\partial \boldsymbol{\theta}_r}\left(y_{uj} - f_{uj}\right) + \left(y_{ui} - f_{ui}\right)\frac{\partial f_{uj}}{\partial \boldsymbol{\theta}_r}\right]$$

$$\frac{\partial^2 \boldsymbol{\upsilon}_{ij}}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_s} = -\sum_{u}\left[\frac{\partial f_{ui}}{\partial \boldsymbol{\theta}_r}\frac{\partial f_{uj}}{\partial \boldsymbol{\theta}_s} + \frac{\partial f_{ui}}{\partial \boldsymbol{\theta}_s}\frac{\partial f_{uj}}{\partial \boldsymbol{\theta}_r}\right]$$

Marginal Posterior Distribution

We expand the objective function

$$S(\theta) \approx S(\hat{\theta}) + (\theta - \hat{\theta})^T \mathbf{A} (\theta - \hat{\theta}) \qquad \mathbf{A} = \frac{1}{2}(m+n+1) \left. \frac{\partial^2 \ln |\upsilon(\theta)|}{\partial \theta \partial \theta^T} \right|_{\theta=\hat{\theta}}$$

We also have the definition

$$S(\theta) = -2 \ln p(\theta \mid \mathbf{y}) + c$$

Combining the two equations

$$p(\theta \mid \mathbf{y}) \propto \exp\left[-\frac{1}{2} (\theta - \hat{\theta})^T \mathbf{A} (\theta - \hat{\theta}) \right]$$

$p(\theta \mid \mathbf{y})$ is the multivariate Normal distribution

$$p(\theta \mid \mathbf{y}) = \mathbb{N}\left\{\hat{\theta}, \mathbf{A}^{-1}\right\}$$

$p(\theta_i \mid \mathbf{y})$ is the univariate Normal distribution

$$p(\theta_i \mid \mathbf{y}) = \mathbb{N}\left\{\hat{\theta}_i, \mathbf{A}_{ii}^{-1}\right\}$$

Covariance and variance of parameters. HPD intervals with probability content 1-α

$$\mathbf{A} = \frac{1}{2}(m+n+1)\frac{\partial^2 \ln\left|\upsilon(\theta)\right|}{\partial\theta\partial\theta^T}\Bigg|_{\theta=\hat{\theta}}$$

$$\mathbf{Cov}(\theta) = \mathbf{A}^{-1}$$

$$\mathbf{C} = \mathbf{A}^{-1}$$

$$\mathbf{Var}(\theta_i) = \sigma^2_{\theta_i} = \mathbf{C}_{ii}$$

$$-U\left(\frac{\alpha}{2},\nu\right) \le \frac{\theta_i - \hat{\theta}_i}{\sigma_{\theta_i}} \le U\left(\frac{\alpha}{2},\nu\right)$$

A useful choice is α=0.05, which gives 95% probability HPD interval for each of the estimated set of parameters

Marginal posterior distribution for parameter transformations

$$k = k_0 \exp\left(-\frac{E}{RT}\right)$$

$$\ln k = \ln k_B + \frac{E}{RT_B}\left(1 - \frac{T_B}{T}\right)$$

Option A:

$$\theta_1 = \ln k_B$$

$$\theta_2 = \frac{E}{RT_B}$$

$$k = \exp\left(\theta_1 + \theta_2\left(1 - \frac{T_B}{T}\right)\right)$$

Option B:

$$\theta_1 = k_B$$

$$\theta_2 = \frac{E}{RT_B}$$

$$k = \theta_1 \exp\left(\theta_2\left(1 - \frac{T_B}{T}\right)\right)$$

$p(\boldsymbol{\theta} \mid \mathbf{y})$ **is the multivariate Normal distribution**

$$p(\boldsymbol{\theta} \mid \mathbf{y}) = \mathbb{N}\left\{\hat{\boldsymbol{\theta}}, \mathbf{A}^{-1}\right\}$$

$$\boldsymbol{\phi} = \boldsymbol{\phi}(\boldsymbol{\theta}) \qquad \mathbf{D} = \frac{\partial \boldsymbol{\phi}}{\partial \boldsymbol{\theta}}$$ **: Jacobian Matrix**

$p(\boldsymbol{\phi} \mid \mathbf{y})$ **is the multivariate Normal distribution**

$$p(\boldsymbol{\phi} \mid \mathbf{y}) = \mathbb{N}\left\{\mathbf{D}\hat{\boldsymbol{\theta}}, \mathbf{D}\mathbf{A}^{-1}\mathbf{D}^T\right\}$$

where: $\mathbf{A} = \frac{1}{2}(m+n+1)\left.\frac{\partial^2 \ln\left|\upsilon(\boldsymbol{\theta})\right|}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$

Lack-of-Fit Analysis
How Good is my Model?

Unknown variance Σ_J of Model M_J

$$\upsilon_{ij}(\theta) = \sum_{u=1}^{n} \left(\mathbf{y}_{ui} - f_i(\mathbf{x}_u,\theta)\right)\left(\mathbf{y}_{uj} - f_j(\mathbf{x}_u,\theta)\right)$$

Goodness-of-Fit Measure for Model M_J

$$\Im_J = \nu_J \ln\left|\frac{\upsilon_J(\hat{\theta})}{\nu_J}\right| - (\nu_J - \nu_e)\ln\left|\frac{\upsilon_J(\hat{\theta}) - \upsilon_e(\hat{\theta})}{\nu_J - \nu_e}\right| - \nu_e \ln\left|\frac{\upsilon_e(\hat{\theta})}{\nu_e}\right|$$

Model Acceptance Test

$\Pr\left(\Im > \Im_J\right) \le 0.01$ **casts doubt on the model**

$\Pr\left(\Im > \Im_J\right) \approx 1.00$ **supports the model**

Strict levels of acceptance or rejection are not advocated

Discrimination amongst rival models
Multiresponse mechanistic models

Unknown variance Σ_J of Model M_J

$$\upsilon_{ij}(\theta) = \sum_{u=1}^{n} \left(\mathbf{y}_{ui} - f_i(\mathbf{x}_u, \theta)\right)\left(\mathbf{y}_{uj} - f_j(\mathbf{x}_u, \theta)\right)$$

Posterior Probability of Model M_J

$$p(M_{\mathbf{J}} \mid \mathbf{y}) \propto p(M_{\mathbf{J}})\, 2^{-\frac{k_J}{2}} \left|\upsilon_{\mathbf{J}}(\hat{\theta})\right|^{-\frac{\nu_E}{2}}$$

$\nu_E =$ **number of replicate events-number of settings of replications**

Posterior Probability Share held by Model M_j

$$\pi(M_{\mathbf{J}} \mid \mathbf{y}) = \frac{p(M_{\mathbf{J}} \mid \mathbf{y})}{\sum_{\mathbf{J}=1}^{\mathbf{M}} p(M_{\mathbf{J}} \mid \mathbf{y})}$$

Optimal Experimental Design for Parameter Estimation

Consider the function $Q(\boldsymbol{\theta})$:

$$Q(\boldsymbol{\theta}) = \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)^T \mathbf{A} \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right) \qquad \mathbf{A} = \frac{1}{2}(m+n+1)\left.\frac{\partial^2 \ln\left|\boldsymbol{\upsilon}(\boldsymbol{\theta})\right|}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

Then:

$Q(\boldsymbol{\theta}) = \chi^2(k,\alpha)$ **is distributed a** *posteriori* **as** χ^2

Volume of the Ellipsoid $Q(\boldsymbol{\theta})$

$$\mathbf{Vol}\left(Q(\boldsymbol{\theta})\right) \propto \sqrt{\det \mathbf{A}^{-1}}$$

Trace of the inverse

$$\mathbf{Tr}\mathbf{A}^{-1} = \sum_{i=1}^{k} \mathbf{A}_{ii}^{-1}$$

A useful choice is α=0.05, which gives 95% probability HPD region for the estimated set of parameters

Optimal Design Basis

Minimization of the ellipsoid volume

Minimization of the trace of the inverse of the Hessian Matrix

The End: Prediction under Uncertainty

Reaction Mechanism

$$A \underset{k_3}{\overset{k_1}{\rightleftharpoons}} B \underset{k_4}{\overset{k_2}{\rightleftharpoons}} C$$

Postulated Model

$$\frac{dC_A}{dt} = -k_1 C_A + k_3 C_B$$

$$\frac{dC_B}{dt} = k_1 C_A - k_2 C_B - k_3 C_B + k_4 C_C$$

$$\frac{dC_C}{dt} = k_2 C_B - k_4 C_C$$

$$t = 0 \quad C_A = C_{A0} \quad C_B = C_{B0} \quad C_C = C_{C0}$$

Estimation

- **Parameter Estimation**
- **The Covariance Matrix**
- **The HPD Intervals**

↓

Prediction

- **Uncertainty in Rate Mode**
- **Uncertainty in Design Mode**

The End: Prediction under Uncertainty

$p(\theta \mid \mathbf{y})$ **is the multivariate Normal distribution**

$$p(\theta \mid \mathbf{y}) = \mathbb{N}\left\{\hat{\theta}, \mathbf{A}^{-1}\right\}$$

$$\phi = \phi(\theta) \qquad \mathbf{D} = \frac{\partial \phi}{\partial \theta}$$ **: Jacobian Matrix**

$p(\phi \mid \mathbf{y})$ **is the multivariate Normal distribution**

$$p(\phi \mid \mathbf{y}) = \mathbb{N}\left\{\mathbf{D}\hat{\theta}, \mathbf{D}\mathbf{A}^{-1}\mathbf{D}^T\right\}$$

where: $\mathbf{A} = \frac{1}{2}(m+n+1)\left.\frac{\partial^2 \ln\left|\upsilon(\theta)\right|}{\partial\theta\partial\theta^T}\right|_{\theta=\hat{\theta}}$

Reaction Mechanism

$$A \underset{k_3}{\overset{k_1}{\rightleftharpoons}} B \underset{k_4}{\overset{k_2}{\rightleftharpoons}} C$$

Postulated Model

$$\frac{dC_A}{dt} = -k_1 C_A + k_3 C_B$$

$$\frac{dC_B}{dt} = k_1 C_A - k_2 C_B - k_3 C_B + k_4 C_C$$

$$\frac{dC_C}{dt} = k_2 C_B - k_4 C_C$$

$$t = 0 \quad C_A = C_{A0} \quad C_B = C_{B0} \quad C_C = C_{C0}$$