<table>
<thead>
<tr>
<th>Week</th>
<th>Description</th>
<th>Reading Material</th>
</tr>
</thead>
</table>
| 13-14 Mar 30-Apr 11 | Parameter Estimation  
• Least Squares (non-linear)  
• Maximum Likelihood  
• Search Methods (Simplex, Gradient, Simulated Annealing)  
• Choice of parameters  
• Multiple types of observations; GLS and weighting errors  
• Bayesian methods for parameter estimation  
  - GLUE  
  - Metropolis-Hastings (MCMC methods)  
  - Parameter uncertainty and correlation | Wallach et al. (2006) Ch 4  
Makowski et al. (2002)  
Wang et al. (2005)  
Beven and Binley (1992) (see below for refs) |
Weeks 13-14 Objectives

1) Become familiar with concepts of parameter estimation for dynamic models
2) Learn about some of the basic methods
Random Model Components: Dynamic Equations & Response Function

**Dynamic Equations:**

\[ U_i(t+\Delta t) = U_i(t) + g_i[U(t),X(t); \theta] \Delta t + \eta_i \]

Where \( \eta_i \) is a random variable for rate equation \( i \)

Also, parameters \( \theta \) may be uncertain as could inputs \( X(t) \)

**Response functions:**

\[ Y = f(X,\theta) + \varepsilon \]

Where \( \varepsilon \) is a random variable
Parameters in Models

• Dynamic biophysical models may have many parameters (θ)
• We may know values for some of the parameters but not others, and even those that we think we know are only estimates that have uncertainty
• One typically has to estimate parameters before applying a model to a particular situation
Parameters

- A parameter (θ) is a numerical value that is a characteristic of the biological and physical components in a model.
- Parameters are not computed by the model nor they observed inputs.
- However, the same variable may be a parameter in one situation and a measured input in another (such as initial soil nitrogen).
- We seldom know the true parameter values and thus parameters are random variables.
Model Error

- Model error ($\varepsilon$) is the difference between observed and predicted output
  \[ \varepsilon = y - f(x; \theta) \]
- This error is a random variable, for example, $\varepsilon$ may be normally distributed with mean of 0.0 and variance of $\sigma^2$
- If this is true, then the model errors are centered on 0, and $E(y|x) = f(x; \theta)$ for a given set of inputs ($x$) and true parameters ($\theta$).
Parameter Estimation

- The problem is that we do not know the true parameter values for many reasons (what are some?)
- However, to apply the model, we want for it to predict, on average, the measured values in a particular situation or in many situations
- Thus, we need to estimate the set of parameter values that provide predictions that accurately reproduce measured values
Estimating Parameters

• Experiments that isolate a particular process so that direct observations on the system can be used to estimate a parameter or a few parameters
• What are examples of this approach?
• What are examples where this approach would not work?
• There are actually more reasons why this may not work in practice with many of the biophysical (ecological, hydrology, etc.) models. What are some?
Other Approaches

• There are many methods
• These methods recognize that the model parameters, however estimated initially, are really random variables and for particular applications would need to be estimated to achieve desired results for specific objectives
• Similar problem as regression analysis where regression coefficients have to be estimated for a particular equation form using a specific set of data
Is Parameter Estimation = Regression Analysis?

- Both rely on statistical notions
- Numerical values of parameters and coefficients are uncertain (random variables)
- Procedures for estimating parameters of dynamic biophysical models are similar to non-linear regression (most biophysical models are non-linear, and may have to be simulated to predict functions (as noted earlier)
Is Parameter Estimation = Regression Analysis?

But, there are important differences. What are some?
General Methods

• Based on measurements alone and a "frequentist" statistical approach
• Based on measurements and on prior knowledge of parameter values using a Bayesian statistical approach.
• We will introduce examples of each.
Ordinary Least Squares

• If there are N measurements of a model output, parameter values are selected that minimizes the sum of the squared errors between model predictions and observations ($Z_{OLS}$):

$$Z_{OLS} = \sum_{i=1}^{N} [y_i - f(x; \theta)]^2$$

• There are a number of ways to find the $\theta$ that results in the least value of ($Z_{OLS}$)
Generalized Least Squares

This is like OLS, but in this case, the variances for $N$ different observations are not equal. This results in a “weighting” of the sum of square errors as follows:

$$Z_{GLS} = \sum_{i=1}^{N} \frac{[y_i - f(x; \theta)]^2}{\sigma_i^2}$$

An example will be shown of this later.
Maximum Likelihood

Likelihood is the probability of the observations for given values of Y for given value of θ, assuming model errors are independent and have variance of $\sigma^2$

Where

$Y = y_1, \ldots y_N$, the observations

$N = \text{number of measurements (} y_i \text{) taken}$

$\theta = \text{set of parameter values}$

$P(Y|\theta) = \text{Likelihood function}$

\[
P(Y | \theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{[y_i - f(\theta)]^2}{2\sigma^2}\right]
\]

The problem then is to find $\theta$ that maximizes this Likelihood function
Log Likelihood

If one takes the Log of the likelihood function shown earlier, a more simple equation is produced

Where

\( Y = y_1, \ldots y_N \), the observations

\( N = \) number of measurements \((y_i)\) taken

\( \theta = \) set of parameter values

\( \ln(P(Y|\theta)) = \) Log Likelihood function

\[
\ln[P(Y|\theta)] = -\frac{N}{2} \ln(2\pi \sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{N} [y_i - f(\theta)]^2
\]

The problem then is to find \( \theta \) that maximizes the Log Likelihood
More General Maximum Likelihood

Probability of the observations for given values of $\theta$ and $V$

Where

$\theta = \text{set of parameter values}$

$V = \text{NxN variance-covariance matrix of model errors}$

$L = \text{Likelihood function}$

$N = \text{number of measurements ($y_i$) taken}$

$$L(y_1,\ldots,y_N,V) = P(y_1,\ldots,y_N | \theta,V)$$

$$= (2\pi)^{-N/2} |V|^{-1/2} \exp\{-1/2[Y - F(\theta)]^T V^{-1}[Y - F(\theta)]\}$$
Obtaining Parameter Estimates

- **Criterion**
  - OLS
  - Generalized Least Squares
  - Max Likelihood
- **Solution, finding best set of parameters for complex models**
  - Search methods
    - Brute Force
    - Gradient
    - Simplex
    - Random search
  - Bayesian methods
    - GLUE
    - Metropolis-Hastings, other Markov Chain Monte Carlo (MCMC) methods
Visual

- Vary parameters and select the combination that gives the best visual fit to the observations
- Not a good way to do this, but this is frequently done
- Need to use Z or L or some criteria and allow a procedure to select the parameter set
Systematic Search

- Set a range over which to search for each parameter to be estimated
- Select \( m \) equally-spaced values within the range for each parameter
- Simulate model output for all combinations of parameter values
- Compute \( Z \) (or \( L \)) for each
- Select parameter set that minimizes \( Z \) or maximizes \( L \); \( Z \) is a measure of residual error
- This is the same as a complete factorial design in sensitivity analysis
- Too costly to be practical in most cases (i.e., when number of parameters is larger than about 3 to 5.)
Simplex Method

With $n$ parameters, generate $n+1$ combinations and simulate $Y$ for each.

Compute $Z$ (or $L$), the criterion used to determine “best parameter set”.

Find the maximum $Z$ and replace the parameter set that gave this $Z$ with another.

Still have $n+1$ sets, repeat the above, reducing the changes as iterations increase.

Procedure illustrated here.

Advantage: no derivatives are needed compared with gradient search methods.

See *Numerical Recipes* book..
Originally developed by Nelder and Mead (1961)
Gradient Search Methods

$Z(\theta)$ is, for example, sum of square errors for the OLS method and $\theta$ is a vector of parameters

Have starting vector, $\theta^0$ and compute $Z^0$

Compute $\delta Z(\theta)/\delta \theta$, the gradient of $Z$ or $\nabla Z$, for each $\theta_i$ at the $\theta^0$ point

Choose a new point, $\theta^1 = \theta^0 - \alpha (\nabla Z)$

compute $Z^1$ at the point $\theta^1$

If $Z^1$ is less than $Z^0$, continue along gradient computed earlier

But, if $Z^1$ is greater than $Z^0$, compute gradient again at $\theta^j$

Continue until $Z^{j+1}$ is within a tolerance of $Z^j$

The parameter set $\theta^j$ is the least square estimate, and $Z^j$ is the residual error between predicted and observed at the optimal set of parameter values
Gradient Search

This is shown schematically here (2 parameters)
The contour lines are lines of equal $Z$, the minimal value is in the center
Vertical axis shows values of one parameter and the horizontal shows the other parameters
The arrows are in the direction of the gradient, which is computed 3 times in this example
Gradient Search Problems

Local minimums can result in a set of parameters that are not optimal.
Note what happens when initial set of parameters was A vs. B.
This is particularly true when the number of parameters to estimate are more than 3 or 4, but it is true in all cases.
Simulated Annealing

- Bouncing ball idea, random jumps to different parts of the search space to avoid getting trapped (Monte Carlo method)
- Size of jumps decreases with number of iterations so that the search converges to a minimum (or maximum) point
- Solution keeps record of m points (say m = 100) that are the lowest (highest) values obtained so far as the solution proceeds
- At the end, one has m solutions for $\theta$, including the “optimal” one; one can think of these as a cloud of points, each with the same Z to a desired level of accuracy
- The initial jump bounds (temperature) and the rate of decrease of this temperature (annealing) is the origin of the name, like in annealing metals
- If $\theta$ is composed of 7 variables (i.e., $\theta_1 \ldots, \theta_7$), then one can estimate the expected value from the m possibilities that are “optimal” or within a small tolerance of the optimal point.
- Also, correlations among parameters can be estimated from the m best solutions
Simulated Annealing Algorithm

The structure of the simulated annealing algorithm
Simulated Annealing Example

Two parameter Rosenbrock function with 2 parameters $(x_1$ and $x_2$):

$$f(x) = [(1 - x_1)^2 + 100(x_2 - x_1^2)]^2$$

Find $x_1$ and $x_2$ that minimize $f$, with $x_1$ and $x_2$ varying between $-1.5$ and $1.5$ each.
Simulated Annealing Example (Continued)

Progression of reduction in objective criteria $Z$ vs. number of trials

Objective function reduction
What Parameters in a Complex Model Should be Estimated?

- Selection based on literature or expert knowledge
- Selection to avoid identifiability problem
- Selection based on sensitivity analysis
- Selection to achieve least RMSEP (prediction)
Selection to Avoid Identifiability Problem

- Occurs when there is no unique solution
- Analysis of model equations may identify parameters that cannot be estimated
- This may occur, for example, when two parameters are multiplied together in a model equation and only appear in that one place, as shown in the AZODYN model example (p.117):

\[
DM_j = DM_{j-1} + EBMAX \cdot f_{t_{j-1}} \cdot EIMAX \cdot [1 - \exp(-K \cdot D \cdot NUC_{j-1})] \cdot C \cdot g_{r_{j-1}}
\]

- The parameters \{EBMAX, EIMAX, C\} cannot be simultaneously estimated, nor can parameters \{K, D\}
Selection Using Sensitivity Analysis

- Morris Method, for example (see sensitivity analysis lecture & chapter 3)
- A parameter with high mean sensitivity indicator $d_i$ indicates an important influence
- A parameter with high variance of sensitivity indicator $d_i$ means that the $i^{th}$ factor interacts strongly with other factors or that the response to this factor in non-linear – or both.
Selection to Minimize Prediction Error

- Start by estimating one parameter at a time, using cross validation
- Select parameter that gives the best MSEP
- Combine this parameter with each of the remaining parameters (i.e., sets of 2)
- Select the next parameter that gives the best MSEP, using cross validation
- Continue this process until all parameters are estimated and MSEP values are available for 1, 2, 3, … N parameters
- Typically, one will find that MSEP is lower for some subset of parameters
- Caution – this will take a LONG time, so one may want to use this in a limited way to narrow down a set that was selected in some other prior way
Example

• ASODYN model with 3 state variables and 18 parameters
• Estimate 4 parameters (EBMAX, K, D, and VMAX)
• All other parameters were set to their nominal or most likely values
• Six years of measurements were available
• Measurements of dry matter, nitrogen uptake, and LAI, with 3 replicates taken on from 8 to 18 different dates in the six years
• Variances of measurements vary with time
• Parameters were estimated using weighted least squares (based on Generalized least squares and equation (9), p. 110)
4. Parameter estimation for crop models

Table 2. Model parameters, initial values, and ranges of variation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
<th>Initial value</th>
<th>Range of variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBMX</td>
<td>Radiation use efficiency</td>
<td>3.3 g MJ$^{-1}$</td>
<td>1.8–4</td>
</tr>
<tr>
<td>K</td>
<td>Radiation extinction coefficient</td>
<td>0.72</td>
<td>0.6–0.8</td>
</tr>
<tr>
<td>D</td>
<td>LAI/critical nitrogen uptake</td>
<td>0.028</td>
<td>0.02–0.045</td>
</tr>
<tr>
<td>VMAX</td>
<td>Maximal rate of nitrogen uptake</td>
<td>0.5 kg ha$^{-1}$·°Cd$^{-1}$</td>
<td>0.2–0.7</td>
</tr>
<tr>
<td>C</td>
<td>Photosynthetically active radiation/global radiation</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>Tmin</td>
<td>Minimal temperature for photosynthesis</td>
<td>0°C</td>
<td></td>
</tr>
<tr>
<td>Topt</td>
<td>Optimal temperature for photosynthesis</td>
<td>15°C</td>
<td></td>
</tr>
<tr>
<td>Tmax</td>
<td>Maximal temperature for photosynthesis</td>
<td>40°C</td>
<td></td>
</tr>
<tr>
<td>EIMAX</td>
<td>Ratio of intercepted to incident radiation</td>
<td>0.96</td>
<td></td>
</tr>
<tr>
<td>Tep-flo</td>
<td>Sum of temperature between earing and flowering</td>
<td>150°Cd</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Parameter of the critical nitrogen concentration function</td>
<td>1.55 t ha$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>Parameter of the critical nitrogen concentration function</td>
<td>4.4%</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>Parameter of the critical nitrogen concentration function</td>
<td>5.35%</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>Parameter of the critical nitrogen concentration function</td>
<td>−0.442</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>Parameter of the maximal nitrogen concentration function</td>
<td>2 t ha$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>Parameter of the maximal nitrogen concentration function</td>
<td>6%</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>Parameter of the maximal nitrogen concentration function</td>
<td>8.3%</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>Parameter of the maximal nitrogen concentration function</td>
<td>−0.44</td>
<td></td>
</tr>
</tbody>
</table>
Example:
Summary graphs of data

e 6. Variances of winter wheat dry matter measurements (kg ha\(^{-2}\)) obtained in Grignon in at different dates between end-of-winter and flowering.

e 7. Measurements of dry matter, nitrogen uptake, and LAI obtained in Grignon in 1999. Each
4. Parameter estimation for crop models

\[ Z_{WLS}(\theta) = \sum_{i=1}^{6} \sum_{j=1}^{N_i} \frac{[y_{ij}^{DM} - f^{DM}(x_{ij}; \theta)]^2}{\text{vâr}(y_{ij}^{DM})} + \sum_{i=1}^{6} \sum_{j=1}^{N_i} \frac{[y_{ij}^{NU} - f^{NU}(x_{ij}; \theta)]^2}{\text{vâr}(y_{ij}^{NU})} + \sum_{i=1}^{6} \sum_{j=1}^{N_i} \frac{[y_{ij}^{f}(x_{ij}; \theta)]^2}{\text{vâr}(y_{ij}^{f})} \]

where \( f^{DM}(x_{ij}; \theta) \), \( f^{NU}(x_{ij}; \theta) \), \( f^{f}(x_{ij}; \theta) \) are the simulated values of dry matter, nitrogen uptake, and LAI for year \( i \), \( i = 1, \ldots, 6 \) and time \( t_j \), \( j = 1, \ldots, N_i \). \( \text{vâr}(y_{ij}^{DM}) \), \( \text{vâr}(y_{ij}^{NU}) \), \( \text{vâr}(y_{ij}^{f}) \) are the empirical variances (calculated from the replicates) for each of the three types of measurements at time \( t_j \), \( j = 1, \ldots, N_i \). The variance \( \text{vâr}(y_{ij}^{s}) \) is calculated as

\[ \text{vâr}(y_{ij}^{s}) = \frac{1}{R(R - 1)} \sum_{k=1}^{R} [y_{ijk}^{s} - y_{ij}^{s}]^2 \]
Estimated Parameters

Table 3. Initial parameter values and values estimated by weighted least squares. The standard errors associated with the parameter estimates are presented between brackets.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial value</th>
<th>Estimated value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$EBMAX$ (g MJ$^{-1}$)</td>
<td>3.3</td>
<td>3.29 (0.11)</td>
</tr>
<tr>
<td>$K$</td>
<td>0.72</td>
<td>0.74 (0.06)</td>
</tr>
<tr>
<td>$D$</td>
<td>0.028</td>
<td>0.028 (0.001)</td>
</tr>
<tr>
<td>$VMAX$ (kg ha$^{-1}$Cd$^{-1}$)</td>
<td>0.5</td>
<td>0.38 (0.02)</td>
</tr>
</tbody>
</table>
Table 4. RMSE values obtained with initial parameter values and with values estimated by weighted least squares.

<table>
<thead>
<tr>
<th>Parameter values</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dry matter (kg ha(^{-1}))</td>
</tr>
<tr>
<td>Initial values</td>
<td>614.9</td>
</tr>
<tr>
<td>Estimated values</td>
<td>607.2</td>
</tr>
</tbody>
</table>
GLUE: A Bayesian Method

References


General Ideas

• Bayes’ Equation

\[ P(\theta|y) = \frac{P(y|\theta)P(\theta)}{\int_{\theta} P(y|\theta)P(\theta)} \]

• A prior distribution of parameters \([P(\theta)]\) must be known
• A likelihood function \([=P(y|\theta)]\) is needed
• \(\theta\) is considered to be a random variable; its estimate includes a distribution, not a single value
• Then, we compute the probability of \(\theta\), given a set of observations
Two distributions are Needed

• **Prior parameter distribution** = probability distribution describing our initial knowledge about parameter values.

\[ P(\theta) \]

Using Monte Carlo to create \( N \) possible parameter vectors, each has the same probability, \( p(\theta) \)

• **Likelihood function** = function relating data to parameters.

\[ P(y|\theta) \]
Measurements

Prior Information about parameter values

Bayesian method

Combined to estimate parameter distribution
Prior probability distribution

Posterior probability distribution

Likelihood function
Generalized Likelihood Uncertainty Estimator (GLUE)

• Bayesian Approach for estimating posterior distribution \( P(\theta|Y) \) based on prior distribution of \( \theta \), \( P(\theta) \). \( P(\theta|Y) \) will be a discrete probability distribution

• Posterior distribution estimate is \( \{ (\theta_j, p_j), j=1, \ldots, J \} \) where \( J \) is the number of discrete possibilities \( \theta_j \).

\[
P(\theta|y) = \frac{P(y|\theta)P(\theta)}{\int_{\theta} P(y|\theta)P(\theta)}
\]
GLUE Steps

1. Randomly generate $N$ vectors of $\theta$ [$\theta_i$, $i=1,N$] from the prior distribution of $\theta$, $P(\theta)$

2. Calculate the Likelihood values [$P(Y|\theta_i)$].

3. Calculate $p_i$ for each $\theta_i$, remembering that each $\theta_i$ has $J$ parameter values

$$p_i = \frac{P(Y | \theta_i)}{\sum_{i=1}^{N} P(Y | \theta_i)}$$

4. The pairs ($\theta_i,p_i$), $i = 1, \ldots, N$ describe the posterior distribution of $\theta$ and can be used to estimate expected value for each of the $J$ parameters, also variance and covariance among parameters

--- Where is $p(\theta)$, the prior distribution of $\theta$? What is $p_i$?
Comment on step 1

How large a sample $N$?

- The accuracy of the results depends on the sample size $N$.

- $N$ values that are too small may lead to inaccurate results. Very large $N$ values may require too much computing time.

- A simple test is to look at averages as a function of sample size.

- A more dependable test is to run the algorithm several times with different starting points.
Comments on step 4
How to approximate the posterior distribution

• The set of pairs \((\theta_1, p_1^*), \ldots, (\theta_N, p_N^*)\) represents a probability distribution.

• This distribution is an approximation of the posterior distribution.

• The posterior mean can be estimated as

\[ \hat{\mu}_{post} = \sum_{i=1}^{N} p_i^* \theta_i \]

• The posterior variance can be estimated as

\[ \hat{\tau}^2_{post} = \sum_{i=1}^{N} p_i^* (\theta_i - \hat{\mu}_{post})^2 \]

• An estimate of covariance between parameters can also be computed
Part 2: Likelihood function and prior distributions

Example (continued)

Discussion of the posterior distribution

1. Result is a probability distribution (posterior distr.)
2. Posterior mean is intermediate between prior mean and observation.
3. Weight of each depends on prior variance and measurement error.
4. Posterior variance is lower than both prior variance and measurement error variance.
5. Use just one data point and still get estimator.
Example 2 - Model

• Non linear model predicting relative growth as a function of some factor $x$ (e.g. $x = \text{nitrogen fertilizer}$).

$$f(x; \theta) = [1 - \exp(-\theta \times x)]$$

• One parameter $\theta$: the growth rate.
Example 2 – Prior distribution

Let’s suppose that the prior distribution is defined from expert knowledge as:

\[ P(\theta) = N(0.03, 0.015^2) \]
• Let’s suppose that two measurements of relative growth $y_1$ and $y_2$ are available:

\[ y_1 = 0.83 \text{ for } x_1 = 100, \]

\[ y_2 = 0.95 \text{ for } x_2 = 200. \]

• The two measurements are assumed to be independent.

• Their standard error is assumed known and equal to 0.02.

• The statistical model is then $y = f(x; \theta) + \varepsilon$, $\varepsilon \sim N(0,(0.02)^2)$ with independence between the two values of $\varepsilon$. 
Example 2 - Likelihood

\[ P(y_1, y_2 | \theta) = P(y_1 | \theta) \times P(y_2 | \theta) \]

\[ = \frac{1}{\sqrt{2\pi \sigma}} \exp\left\{ - \frac{[y_1 - (1 - \exp(-\theta \times x_1))]^2}{2\sigma^2} \right\} \times \frac{1}{\sqrt{2\pi \sigma}} \exp\left\{ - \frac{[y_2 - (1 - \exp(-\theta \times x_2))]^2}{2\sigma^2} \right\} \]

We have assumed that \( \sigma \) is known and equal to 0.02.
What if it is not known????
Example

**Step 0:** Use the prior distribution as the proposal distribution.

\[ g(\theta) = \text{N}(0.03, 0.015^2). \]

**Step 1:** Generate \( N \) parameter vectors, in this problem there is only 1 parameter, so generate \( N \) values \( \theta_1, \theta_2, \ldots, \theta_N \). (i.e., \( N \) may be 1,000)

**Step 2:** Calculate a « weight » for each parameter value \( w_1, w_2, \ldots, w_N \).

The weight is equal to the likelihood value,

\[
w_i = P(y_1 | \theta_i) \times P(y_2 | \theta_i)
\]

\[
w_i = \frac{1}{\sqrt{2\pi} \sigma} \exp \left\{ - \frac{(y_1 - (1 - \exp(-\theta_i \times x_1)))^2}{2\sigma^2} \right\} \times \frac{1}{\sqrt{2\pi} \sigma} \exp \left\{ - \frac{(y_2 - (1 - \exp(-\theta_i \times x_2)))^2}{2\sigma^2} \right\}
\]

**Step 3:** Calculate normalized weights \( w_1^*, \ldots, w_N^* \).

\[
W_i^* = \frac{w_i}{\sum_{i=1}^{N} w_i}
\]

**Step 4:** Use the sample of parameter values and the normalized weights to approximate the posterior distribution.
Example Results with $N=10000$

Plot of normalized weights

Parameter values drawn from posterior (after resampling)

Estimated posterior mean: 0.0176
Estimated posterior standard deviation: $1.17 \times 10^{-3}$
Example 2 – How large a sample $N$?

- The algorithm is run five times (with different seeds) for two different $N$ values:
  - $N=100$
  - $N=10000$

- The posterior mean and posterior variance are computed after each run.

- The stability of the result is analyzed.
## Example 2 – How many simulations?

<table>
<thead>
<tr>
<th>N</th>
<th>Run</th>
<th>Posterior mean</th>
<th>Posterior standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>1</td>
<td>0.0178</td>
<td>9.99E-04</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>0.0173</td>
<td>9.48E-04</td>
</tr>
<tr>
<td>100</td>
<td>3</td>
<td>0.0173</td>
<td>1.10E-03</td>
</tr>
<tr>
<td>100</td>
<td>4</td>
<td>0.0176</td>
<td>1.04E-03</td>
</tr>
<tr>
<td>100</td>
<td>5</td>
<td>0.0176</td>
<td>9.91E-04</td>
</tr>
<tr>
<td>10000</td>
<td>1</td>
<td>0.0176</td>
<td>1.17E-03</td>
</tr>
<tr>
<td>10000</td>
<td>2</td>
<td>0.0176</td>
<td>1.14E-03</td>
</tr>
<tr>
<td>10000</td>
<td>3</td>
<td>0.0176</td>
<td>1.17E-03</td>
</tr>
<tr>
<td>10000</td>
<td>4</td>
<td>0.0176</td>
<td>1.14E-03</td>
</tr>
<tr>
<td>10000</td>
<td>5</td>
<td>0.0176</td>
<td>1.14E-03</td>
</tr>
</tbody>
</table>

The estimation of the posterior mean is very accurate with $N=10000$. 
Example of Generated Parameter Sets & Posterior Distribution

- \((\text{EBMAX}, K, D, \text{and VMAX})\) 
  \(p_i\)
- \((3.30 \quad 0.720 \quad 0.0280 \quad 0.500)\) \(i=1\) \(0.1810\)
- \((3.11 \quad 0.770 \quad 0.0265 \quad 0.503)\) \(i=2\) \(0.0020\)
- \((3.39 \quad 0.817 \quad 0.0291 \quad 0.487)\) \(i=3\) \(0.0482\)
- \((3.26 \quad 0.661 \quad 0.0277 \quad 0.523)\) \(i=4\) \(0.0006\)
- \((3.46 \quad 0.840 \quad 0.0270 \quad 0.507)\) \(i=N\) \(0.0131\)
Markov Chain Monte Carlo (MCMC)

• Metropolis-Hastings is an example
• Similar to GLUE
• Use prior distribution to generate a new one
• Thus, generate candidate parameter vectors “on the fly” based on current estimate
• Makowski showed that the Metropolis-Hastings MCMC method was somewhat better than GLUE for one particular problem
Discussion

• Sampling methods have gained favor over other search methods, particularly for complex models
• Likelihood function – various choices that can affect results
• If prior distribution is not good, results will be affected, regardless of method (GLUE or MCMC)
• Benefit from:
  – estimation of posterior distribution of parameters,
  – small sample size needed to estimate them
  – ability to perform uncertainty analysis as part of the process
Discussion

- Philosophical issues
- Choice of parameters to estimate
- Limitations and precautions
- Dealing with correlated factors
- Approaches
  - Bayesian vs. frequentist approaches
  - Search methods
  - Prior distribution need
  - Design of experiment – dependent on method
- Computer resources needed
- Interpretation of results
Homework

• Retrieve files from Class Web site
  – WORD document with problems to work
  – EXCEL Spreadsheet with data
• Due on April 17 or any time before then