Sigmoid curves and a case for close-to-linear nonlinear models

Charles Y. Tan  
MBSW  
May 19th, 2009

All models are wrong, some are more useful than others.
Outline

Introduction
Nonlinear Models
Sigmoid Curves
Assess the Approximation
Numerical Case Study
Conclusions
Sigmoid curves are common in biological sciences

- Quantitative bioanalytical methods
  - Immunoassays
  - Bioassays
  - Hill equation (1910)

- Pharmacology
  - Concentration-effect or dose-response curves
  - Emax model (1964)

- Growth curves
  - (Population or organ) size as function of time
  - Mechanistic and empirical
  - Autocatalytic model (1838, 1908)
Statistics: old favorite and new question

- Classic models: (four-parameter) logistic models
  - Hill equation, Emax model, and autocatalytic model are the *same* models: logistic models.
  - They’re symmetric.
Statistics: old favorite and new question

- Classic models: (four-parameter) logistic models
  - Hill equation, Emax model, and autocatalytic model are the same models: logistic models.
  - They’re symmetric.

- New question: what model to use when data are asymmetric
  - Answer from some quarters: “five-parameter logistic (5PL)” (Richards model)
Statistics: old favorite and new question

- Classic models: (four-parameter) logistic models
  - Hill equation, Emax model, and autocatalytic model are the same models: logistic models.
  - They’re symmetric.
- New question: what model to use when data are asymmetric
  - Answer from some quarters: “five-parameter logistic (5PL)” (Richards model)
Nonlinear regression

\[ y_i = f(x_i; \theta) + \varepsilon_i, \quad i = 1, 2, \ldots, n, \]

- Nonlinearity of \( f \) with respect to \( \theta \): defining characteristics
- Nonlinearity of \( f \) with respect to \( x \): incidental
Nonlinear regression

\[ y_i = f(x_i; \theta) + \varepsilon_i, \quad i = 1, 2, \ldots, n, \]

- Nonlinearity of \( f \) with respect to \( \theta \): defining characteristics
- Nonlinearity of \( f \) with respect to \( x \): incidental
- Homogeneous variance: \( \varepsilon_i \)'s are i.i.d. \( N(0, \sigma^2) \)

Maximum Likelihood = Least Squares

Objective function:

\[ S(\theta) = (y - f(\theta))^T (y - f(\theta)) \]
1st order approximation of the model (w.r.t. parameter)

\[ f(\theta) \approx f(\theta^*) + F_\bullet(\theta - \theta^*), \]

where

\[ F_\bullet = F_\bullet(\theta^*) = \left( \frac{\partial f(x_i; \theta)}{\partial \theta_j} \bigg|_{\theta=\theta^*} \right)_{n \times k} \]

Plug it in the definition of \( S(\theta) \), we have a partial 2nd order expansion of \( S(\theta) \) near \( \theta^* \):

\[ S(\theta) \approx \varepsilon' \varepsilon - 2\varepsilon' F_\bullet (\theta - \theta^*) + (\theta - \theta^*)' F_\bullet' F_\bullet (\theta - \theta^*) \]
Common framework for inference

\[ S(\theta^*) - S(\hat{\theta}) \approx (\hat{\theta} - \theta^*)'F_\theta F_\theta (\hat{\theta} - \theta^*) \approx \varepsilon' F_\theta (F_\theta' F_\theta)^{-1} F_\theta' \varepsilon \]

\[ S(\hat{\theta}) \approx \varepsilon' (I - F_\theta(F_\theta' F_\theta)^{-1} F_\theta') \varepsilon \]
Common framework for inference

\[ S(\theta^*) - S(\hat{\theta}) \approx (\hat{\theta} - \theta^*)'F_\bullet F_\bullet (\hat{\theta} - \theta^*) \approx \varepsilon'F_\bullet (F_\bullet F_\bullet)^{-1}F_\bullet \varepsilon \]

\[ S(\hat{\theta}) \approx \varepsilon'(I - F_\bullet (F_\bullet F_\bullet)^{-1}F_\bullet)\varepsilon \]

Since \( F_\bullet (F_\bullet F_\bullet)^{-1}F_\bullet \) is idempotent
Common framework for inference

\[ S(\theta^*) - S(\hat{\theta}) \approx (\hat{\theta} - \theta^*)'F_\cdot F_\cdot(\hat{\theta} - \theta^*) \approx \varepsilon' F_\cdot (F_\cdot F_\cdot)^{-1} F_\cdot \varepsilon \]

\[ S(\hat{\theta}) \approx \varepsilon' (I - F_\cdot (F_\cdot F_\cdot)^{-1} F_\cdot) \varepsilon \]

Since \( F_\cdot (F_\cdot F_\cdot)^{-1} F_\cdot \) is idempotent

Local inference:

\[ \frac{(\hat{\theta} - \theta^*)'F_\cdot F_\cdot(\hat{\theta} - \theta^*)}{S(\hat{\theta})} \sim \frac{k}{n - k} F_{k,n-k} \]

Global inference:

\[ \frac{S(\theta^*) - S(\hat{\theta})}{S(\hat{\theta})} \sim \frac{k}{n - k} F_{k,n-k} \]

Charles Y. Tan MBSW May 19th, 2009

Sigmoid curves and a case for close-to-linear nonlinear models
Implications of Local and Global Inferences

- Local: always produces ellipsoids in the parameter space
  - Implicitly assumes $S(\theta)$ is an elliptic paraboloid near $\theta^*$
  - Both the shape and cutoff depend on the approximation
  - Is $S(\theta)$ always elliptic paraboloid like near $\theta^*$?

- Global ($S(\theta) \leq c$): shape varies case by case
  - Follows the Likelihood Principle (“Exact”)
  - The cutoff $c$ depends on the approximation
  - What if $S(\theta)$ has multiple minimums and the difference in their $S$ values is small?
Intrinsic and parameter-effect curvatures

Expectation surface or solution locus: \( f(\theta) \in \mathbb{R}^n \)

Its approximation:

\[
f(\theta) \approx f(\theta^*) + F_\theta(\theta - \theta^*)
\]
Intrinsic and parameter-effect curvatures

Expectation surface or solution locus: \( f(\theta) \in \mathbb{R}^n \)

Its approximation:

\[
f(\theta) \approx f(\theta^*) + F_\theta(\theta - \theta^*)
\]

- Planar assumption
  - The expectation surface is close to its tangent plane.
  - Intrinsic curvature: deviation at \( f(\hat{\theta}) \).
Intrinsic and parameter-effect curvatures

Expectation surface or solution locus: \( f(\theta) \in \mathbb{R}^n \)
Its approximation:

\[
f(\theta) \approx f(\theta^*) + F_\theta(\theta - \theta^*)
\]

- **Planar assumption**
  - The expectation surface is close to its tangent plane.
  - Intrinsic curvature: deviation at \( f(\hat{\theta}) \).

- **Uniform-coordinate assumption**
  - Straight parallel equispaced lines in the parameter space \( \mathbb{R}^k \) map into straight parallel equispaced lines in the expectation surface (as they do in the tangent plane).
  - Parameter-effect curvature: deviation at \( f(\hat{\theta}) \).
Curvatures (nonlinearity) are local properties

- The model $f$
- The parameters $\theta$
  - Parameterization
  - Values
- The design $x$
  - Sample size
  - Values
- The particular realization of $\epsilon$
Finite sample property: close-to-linear

- Asymptotically, i.e., $n \to \infty$ or $\sigma \to 0$, all nonlinear models behave like linear models.

- A nonlinear model is **close-to-linear** if it behaves like a linear model under relative small $n$ and moderate $\sigma$ (Ratkowsky).
Let $x$ denote the independent variable. Let $\theta$ be either $(a, b, c, d)$ for four-parameter models or $(a, b, c, d, g)$ for five-parameter models. Let $u = f(x; \theta)$. We impose following conditions on the independent variable and parameters:

I. The curve is sigmoid when $u$ is plotted against $x$;

II. When $x = c$, $u = (a + d)/2$;

III. When $b > 0$, $d$ is the left asymptote and $a$ is the right asymptote;

IV. When $b < 0$, $a$ is the left asymptote and $d$ is the right asymptote;

V. $u$ is a function of $x$ through $b(x - c)$.  

Charles Y. Tan  MBSW  May 19th, 2009

Sigmoid curves and a case for close-to-linear nonlinear models
A sigmoid curve is *symmetric* if and only if $\partial f / \partial x$ is an even function centered at the mid point $c$.

*Inflection point* is where $\partial f / \partial x$ reaches a (local) minimum or maximum.

A necessary, but not sufficient, condition for symmetry: the inflection point is unique and coincides with the mid point $c$. 
Four-parameter logistic (4PL) curve

The model:

\[ f(x; a, b, c, d) = d + \frac{a - d}{1 + e^{-b(x-c)}} \]
Four-parameter logistic (4PL) curve

- The model:
  \[ f(x; a, b, c, d) = d + \frac{a - d}{1 + e^{-b(x-c)}} \]

- Linearizing function:
  \[ \text{logit} \left( \frac{u - d}{a - d} \right) = b(x - c) \]
Four-parameter logistic (4PL) curve

- The model:

\[ f(x; a, b, c, d) = d + \frac{a - d}{1 + e^{-b(x-c)}} \]

- Linearizing function:

\[ \text{logit} \left( \frac{u - d}{a - d} \right) = b(x - c) \]

- Since \( f(x; a, b, c, d) \) is the same curve as \( f(x; d, -b, c, a) \), the condition of \( a > d \) or \( a < d \) is needed to resolve the identifiability problem.
Richards model ("5PL")

The model:

\[ f(x; a, b, c, d, g) = d + \frac{a - d}{1 + \left(2^{1/g} - 1\right)e^{-b(x-c)}}^g \]

For \( g = 1 \), Richards model is reduced to 4PL.

For \( g \neq 1 \), Richards model is asymmetric.
Richards model ("5PL")

- The model:

\[ f(x; a, b, c, d, g) = d + \frac{a - d}{(1 + (2^{1/g} - 1)e^{-b(x-c)})^g} \]

- Linearizing function:

\[
\log \left( \frac{2^{1/g} - 1}{\left( \frac{u - d}{a - d} \right)^{-1/g} - 1} \right) = b(x - c)
\]

For \( g = 1 \), Richards model is reduced to 4PL.

For \( g \neq 1 \), Richards model is asymmetric.
Richards model ("5PL")

The model:

\[ f(x; a, b, c, d, g) = d + \frac{a - d}{1 + (2^{1/g} - 1)e^{-b(x - c)})^g} \]

Linearizing function:

\[ \log \left( \frac{2^{1/g} - 1}{\left(\frac{u - d}{a - d}\right)^{-1/g} - 1} \right) = b(x - c) \]

- For \( g = 1 \), Richards model is reduced to 4PL.
- For \( g \neq 1 \), Richards model is asymmetric.
Richards model: flexibility and “identification problem”

- Four distinctive segments of the parameter space

  R1. $b > 0$ and $a > d$: increasing function of $x$; as $g : 0 \to +\infty$, the inflection point: $+\infty \to \log(\log 2)/b + c < c$;

  R2. $b > 0$ and $a < d$: decreasing function of $x$; as $g : 0 \to +\infty$, the inflection point: $+\infty \to \log(\log 2)/b + c < c$;

  R3. $b < 0$ and $a > d$: decreasing function of $x$; as $g : 0 \to +\infty$, the inflection point: $-\infty \to \log(\log 2)/b + c > c$;

  R4. $b < 0$ and $a < d$: increasing function of $x$; as $g : 0 \to +\infty$, the inflection point: $-\infty \to \log(\log 2)/b + c > c$. 
Richards model: flexibility and “identification problem”

- Four distinctive segments of the parameter space
  - **R1.** $b > 0$ and $a > d$: increasing function of $x$; as $g : 0 \to +\infty$, the inflection point: $+\infty \to \log(\log 2)/b + c < c$;
  - **R2.** $b > 0$ and $a < d$: decreasing function of $x$; as $g : 0 \to +\infty$, the inflection point: $+\infty \to \log(\log 2)/b + c < c$;
  - **R3.** $b < 0$ and $a > d$: decreasing function of $x$; as $g : 0 \to +\infty$, the inflection point: $-\infty \to \log(\log 2)/b + c > c$;
  - **R4.** $b < 0$ and $a < d$: increasing function of $x$; as $g : 0 \to +\infty$, the inflection point: $-\infty \to \log(\log 2)/b + c > c$.

- Flexibility: each pair, R1/R4 and R2/R3, is capable to model an inflection point anywhere in $\mathbb{R}$.
Richards model: flexibility and “identification problem”

- Four distinctive segments of the parameter space
  
  R1. $b > 0$ and $a > d$: increasing function of $x$; as $g : 0 \rightarrow +\infty$, the inflection point: $+\infty \rightarrow \log(\log 2)/b + c < c$;
  
  R2. $b > 0$ and $a < d$: decreasing function of $x$; as $g : 0 \rightarrow +\infty$, the inflection point: $+\infty \rightarrow \log(\log 2)/b + c < c$;
  
  R3. $b < 0$ and $a > d$: decreasing function of $x$; as $g : 0 \rightarrow +\infty$, the inflection point: $-\infty \rightarrow \log(\log 2)/b + c > c$;
  
  R4. $b < 0$ and $a < d$: increasing function of $x$; as $g : 0 \rightarrow +\infty$, the inflection point: $-\infty \rightarrow \log(\log 2)/b + c > c$.

- Flexibility: each pair, R1/R4 and R2/R3, is capable to model an inflection point anywhere in $\mathbb{R}$

- “Identification problem”: pairs of curves that are not identical, but very similar (same asymptotes, same mid point, same inflection point), yet far apart in the parameter space.
Four-parameter Gompertz (4PG) curve

The model:

\[ f(x; a, b, c, d) = d + \frac{a - d}{2^{\exp(-b(x-c))}} \]
Four-parameter Gompertz (4PG) curve

- The model:
  \[ f(x; a, b, c, d) = d + \frac{a - d}{2^{\exp(-b(x-c))}} \]

- Linearizing function:
  \[ -\log\left(-\log_2\left(\frac{u - d}{a - d}\right)\right) = b(x - c) \]
Four-parameter Gompertz (4PG) curve

The model:

\[ f(x; a, b, c, d) = d + \frac{a - d}{2^{\exp(-b(x-c))}} \]

Linearizing function:

\[ - \log \left( - \log_2 \left( \frac{u - d}{a - d} \right) \right) = b(x - c) \]

Asymmetric sigmoid curve

GenLinMod
4PG: distinctive but not quite flexible

- Four distinctive segments of the parameter space
  - **G1.** $b > 0$ and $a > d$: increasing function of $x$; the inflection point is at $\log(\log 2)/b + c < c$;
  - **G2.** $b > 0$ and $a < d$: decreasing function of $x$; the inflection point is at $\log(\log 2)/b + c < c$;
  - **G3.** $b < 0$ and $a > d$: decreasing function of $x$; the inflection point is at $\log(\log 2)/b + c > c$;
  - **G4.** $b < 0$ and $a < d$: increasing function of $x$; the inflection point is at $\log(\log 2)/b + c > c$. 

G1–G4 can be thought as the limiting version of R1–R4 as $g \to +\infty$. 

$f(x; a, b, c, d)$ and $f(x; d, -b, c, a)$ have the same asymptotes, the same mid point, and their inflection points are equal distance from mid point.
4PG: distinctive but not quite flexible

- Four distinctive segments of the parameter space
  - **G1.** $b > 0$ and $a > d$: increasing function of $x$; the inflection point is at $\log(\log 2)/b + c < c$;
  - **G2.** $b > 0$ and $a < d$: decreasing function of $x$; the inflection point is at $\log(\log 2)/b + c > c$;
  - **G3.** $b < 0$ and $a > d$: decreasing function of $x$; the inflection point is at $\log(\log 2)/b + c > c$;
  - **G4.** $b < 0$ and $a < d$: increasing function of $x$; the inflection point is at $\log(\log 2)/b + c > c$.

- **G1–G4** can be thought as the limiting version of R1–R4 as $g \to +\infty$. 

Charles Y. Tan  MBSW  May 19th, 2009
4PG: distinctive but not quite flexible

- Four distinctive segments of the parameter space
  - G1. $b > 0$ and $a > d$: increasing function of $x$; the inflection point is at $\log(\log 2)/b + c < c$;
  - G2. $b > 0$ and $a < d$: decreasing function of $x$; the inflection point is at $\log(\log 2)/b + c < c$;
  - G3. $b < 0$ and $a > d$: decreasing function of $x$; the inflection point is at $\log(\log 2)/b + c > c$;
  - G4. $b < 0$ and $a < d$: increasing function of $x$; the inflection point is at $\log(\log 2)/b + c > c$.

- G1–G4 can be thought as the limiting version of R1–R4 as $g \to +\infty$.

- $f(x; a, b, c, d)$ and $f(x; d, -b, c, a)$ have the same asymptotes, the same mid point, and their inflection points are equal distance from mid point.
The new model: mixing two 4PG curves up linearly

\[
f(x) = g \left( d + \frac{a - d}{2^{\exp\left(-b(x-c)\right)}} \right) + (1 - g) \left( a + \frac{d - a}{2^{\exp\left(b(x-c)\right)}} \right)
\]
The new model: mixing two 4PG curves up linearly

The model:

\[ f(x) = g \left( d + \frac{a - d}{2^{\exp(-b(x-c))}} \right) + (1-g) \left( a + \frac{d - a}{2^{\exp(b(x-c))}} \right) \]

Linearizing function:

\[ \Psi^{-1} \left( \frac{u - gd - (1-g)a}{a - d}; g \right) = b(x - c) \]

where \( \Psi(t; g) = \frac{g}{2^{\exp(-t)}} - \frac{1-g}{2^{\exp(t)}} \).
The new model: mixing two 4PG curves up linearly

The model:

$$f(x) = g \left( d + \frac{a - d}{2 \exp \left( -b(x - c) \right)} \right) + (1 - g) \left( a + \frac{d - a}{2 \exp \left( b(x - c) \right)} \right)$$

Linearizing function:

$$\psi^{-1} \left( \frac{u - gd - (1 - g)a}{a - d}; g \right) = b(x - c)$$

where $$\psi(t; g) = \frac{g}{2 \exp(-t)} - \frac{1 - g}{2 \exp(t)}$$. 

$$f(x; a, b, c, d, g) = f(x; d, -b, c, a, 1 - g)$$: either $$a > d$$ or $$a < d$$ would resolve the identifiability issue without any loss.
The new model: flexible and distinctive

Theorem

- When \( g = 1/2 \), it is a symmetric;
- When \( 1/2 < g \leq 1 \), the inflection point is unique and between \( \log(\log 2)/b + c \) and \( c \);
- When \( 0 \leq g < 1/2 \), the inflection point is unique and between \( c \) and \( -\log(\log 2)/b + c \);
- When \( g > 1 \), there are multiple inflection points, one of which is less than \( \log(\log 2)/b + c \) for \( b > 0 \) or greater than \( \log(\log 2)/b + c \) for \( b < 0 \);
- When \( g < 0 \), there are multiple inflection points, one of which is greater than \( -\log(\log 2)/b + c \) for \( b > 0 \) or less than \( -\log(\log 2)/b + c \) for \( b < 0 \).
Use the “complete” to assess the “partial”

Original objective function: \( S(\theta) = (y - f(\theta))' (y - f(\theta)) \)
Use the “complete” to assess the “partial”

- Original objective function: \( S(\theta) = (y - f(\theta))' (y - f(\theta)) \)

- Partial 2nd order expansion of the objective:

\[
S(\theta) \approx \varepsilon'\varepsilon - 2\varepsilon' F(\theta - \theta^*) + (\theta - \theta^*)' F F' (\theta - \theta^*)
\]
Use the “complete” to assess the “partial”

- Original objective function: \( S(\theta) = (y - f(\theta))' (y - f(\theta)) \)

- Complete 2nd order expansion of the objective:
  \[
  S(\theta) \approx \varepsilon'\varepsilon - 2\varepsilon'F_\bullet(\theta - \theta^*) + (\theta - \theta^*)'H(\theta - \theta^*)
  \]
  where
  \[
  H = \frac{1}{2} \nabla^2 S(\theta^*) = F'_\bullet F \bullet - [\varepsilon'] [F_{\bullet\bullet}]
  \]
  \[
  [\varepsilon'] [F_{\bullet\bullet}] = \left( \sum_{i=1}^{n} \varepsilon_i \frac{\partial^2 f(x_i; \theta)}{\partial \theta_r \partial \theta_s} \bigg|_{\theta=\theta^*} \right)_{k \times k}
  \]

- Partial 2nd order expansion of the objective:
  \[
  S(\theta) \approx \varepsilon'\varepsilon - 2\varepsilon'F_\bullet(\theta - \theta^*) + (\theta - \theta^*)'F'_\bullet F_\bullet(\theta - \theta^*)
  \]
Quantify close-to-linear-ness by comparing $H$ to $F' F$. 

$$H = F' F - [\epsilon'] [F_{\cdot \cdot}]$$

- For linear models: $F_{\cdot \cdot} = 0$, hence $H = F' F$. 

Charles Y. Tan  MBSW  May 19th, 2009
Quantify close-to-linear-ness by comparing $H$ to $F'F$. 

$$H = F'F - [\epsilon'] [F_{\epsilon\epsilon}]$$

- For linear models: $F_{\epsilon\epsilon} = 0$, hence $H = F'F$. 
- As $\sigma \to 0$: $H \to F'F$ almost surely 
- As $n \to \infty$: $H \to F'F$ almost surely
Quantify close-to-linear-ness by comparing $H$ to $F' \cdot F$.

$H = F' \cdot F - [\epsilon'] [F_{oo}]$

- For linear models: $F_{oo} = 0$, hence $H = F' \cdot F$.
- As $\sigma \to 0$: $H \to F' \cdot F$ almost surely.
- As $n \to \infty$: $H \to F' \cdot F$ almost surely.
- For any $\sigma$ and $n$: $\mathcal{E}(H) = F' \cdot F$. 
Geometry of $S(\theta)$ and eigenvalues of $H$

- All eigenvalues are positive: $S(\theta)$ near $\theta^*$ is **elliptic** paraboloid like and has a minimum.

- Some of the eigenvalues are negative: $S(\theta)$ near $\theta^*$ is **hyperbolic** paraboloid like (non-informative).

- The whole $S(\theta)$ is unbounded from below, no LS or ML solution: at least warned.

- $S(\theta)$ has (multiple) minimum(s) away from the true value $\theta^*$, nominal LS or ML solution can be found: misleading.
Geometry of $S(\theta)$ and eigenvalues of $H$

- All eigenvalues are positive: $S(\theta)$ near $\theta^*$ is **elliptic** paraboloid like and has a minimum.

- Some of the eigenvalues are negative: $S(\theta)$ near $\theta^*$ is **hyperbolic** paraboloid like (*non-informative*).
Geometry of $S(\theta)$ and eigenvalues of $H$

- All eigenvalues are positive: $S(\theta)$ near $\theta^*$ is elliptic paraboloid like and has a minimum.

- Some of the eigenvalues are negative: $S(\theta)$ near $\theta^*$ is hyperbolic paraboloid like (*non-informative*).

- The whole $S(\theta)$ is unbounded from below, no LS or ML solution: at least warned.
Geometry of $S(\theta)$ and eigenvalues of $H$

- All eigenvalues are positive: $S(\theta)$ near $\theta^*$ is elliptic paraboloid like and has a minimum.

- Some of the eigenvalues are negative: $S(\theta)$ near $\theta^*$ is hyperbolic paraboloid like (non-informative).

- The whole $S(\theta)$ is unbounded from below, no LS or ML solution: at least warned.

- $S(\theta)$ has (multiple) minimum(s) away from the true value $\theta^*$, nominal LS or ML solution can be found: misleading.
How close is $H$ to $F' \cdot F'$ overall

Define relative information content $\tau$ as

$$\tau = \begin{cases} 
\frac{\det(H)}{\det(F' \cdot F')}, & \text{if } H \text{ is positive definite;} \\
-m, & \text{if } m \text{ eigen values of } H \leq 0
\end{cases}$$
How close is $\mathbf{H}$ to $\mathbf{F}^\prime \mathbf{F}$ overall

- Define \textit{relative information content} $\tau$ as
  \[
  \tau = \begin{cases} 
  \frac{\det(\mathbf{H})}{\det(\mathbf{F}^\prime \mathbf{F})}, & \text{if } \mathbf{H} \text{ is positive definite;} \\
  -m, & \text{if } m \text{ eigen values of } \mathbf{H} \leq 0
  \end{cases}
  \]

- Define \textit{probability of model failure} as $\xi = \Pr\{\tau < 0\}$
How close is $\mathbf{H}$ to $\mathbf{F}'\mathbf{F}_*$ overall

- Define *relative information content* $\tau$ as

$$\tau = \begin{cases} \frac{\det(\mathbf{H})}{\det(\mathbf{F}'\mathbf{F}_*)}, & \text{if } \mathbf{H} \text{ is positive definite;} \\ -m, & \text{if } m \text{ eigen values of } \mathbf{H} \leq 0 \end{cases}$$

- Define *probability of model failure* as $\xi = \Pr\{\tau < 0\}$

- Define *deviation from unity* $\eta$ as $\eta^2 = E\left[ (\tau - 1)^2 | \tau > 0 \right]$
How close is $\mathbf{F} \cdot \mathbf{H}^{-1} \mathbf{F}'$ to idempotency

From $S(\theta) \approx \epsilon'\epsilon - 2\epsilon'\mathbf{F} \cdot (\theta - \theta^*) + (\theta - \theta^*)' \mathbf{H}(\theta - \theta^*)$, we obtain more rigorous approximations:

- $\mathbf{S}(\theta^*) - \mathbf{S}(\hat{\theta}) \approx \epsilon'\mathbf{F} \cdot \mathbf{H}^{-1} \mathbf{F}'\epsilon$
  - compared with $\epsilon'\mathbf{F} \cdot (\mathbf{F}'\mathbf{F} \cdot)^{-1} \mathbf{F}'\epsilon$

- $\mathbf{S}(\hat{\theta}) \approx \epsilon'(\mathbf{I} - \mathbf{F} \cdot \mathbf{H}^{-1} \mathbf{F}')\epsilon$
  - compared with $\epsilon'((\mathbf{I} - \mathbf{F} \cdot (\mathbf{F}'\mathbf{F} \cdot)^{-1} \mathbf{F}'))\epsilon$

- Dependence of $\mathbf{S}(\theta^*) - \mathbf{S}(\hat{\theta})$ and $\mathbf{S}(\hat{\theta})$ is measured by $\|\mathbf{F} \cdot \mathbf{H}^{-1} \mathbf{F}'(\mathbf{I} - \mathbf{F} \cdot \mathbf{H}^{-1} \mathbf{F}')\|$ (after normalization)
  - compared with independence
Let \( t_1 = \text{tr}(F\cdot H^{-1}F') \), \( t_2 = \text{tr}((F\cdot H^{-1}F')^2) \),
\( t_3 = \text{tr}((F\cdot H^{-1}F')^3) \) and \( t_4 = \text{tr}((F\cdot H^{-1}F')^4) \).

Define **effective degree of freedom of the model** as

\[
\alpha = \frac{t_1^2}{t_2}
\]
Three effective degrees

Let $t_1 = \text{tr}(F \cdot H^{-1} F')$, $t_2 = \text{tr}((F \cdot H^{-1} F')^2)$, $t_3 = \text{tr}((F \cdot H^{-1} F')^3)$ and $t_4 = \text{tr}((F \cdot H^{-1} F')^4)$.

- Define effective degree of freedom of the model as
  $$\alpha = \frac{t_1^2}{t_2}$$

- Define effective degree of freedom of the residuals as
  $$\beta = \frac{(n - t_1)^2}{n - 2t_1 + t_2}$$
Three effective degrees

Let \( t_1 = \text{tr}(F \cdot H^{-1} F') \), \( t_2 = \text{tr}((F \cdot H^{-1} F')^2) \), \( t_3 = \text{tr}((F \cdot H^{-1} F')^3) \) and \( t_4 = \text{tr}((F \cdot H^{-1} F')^4) \). 

- Define **effective degree of freedom of the model** as
  \[
  \alpha = \frac{t_1}{t_2}
  \]

- Define **effective degree of freedom of the residuals** as
  \[
  \beta = \frac{(n - t_1)^2}{n - 2t_1 + t_2}
  \]

- Define **effective degree of dependence** as
  \[
  \gamma = \sqrt{\frac{t_2 - 2t_3 + t_4}{t_2(n - 2t_1 + t_2)}}
  \]
Four particular curves: from a cell based bioassay

<table>
<thead>
<tr>
<th>Model</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>4P Logistic</td>
<td>2500</td>
<td>−1.7</td>
<td>log(30)</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>Richards</td>
<td>2500</td>
<td>−1.3</td>
<td>log(30)</td>
<td>400</td>
<td>3</td>
</tr>
<tr>
<td>4P Gompertz</td>
<td>2500</td>
<td>−1.1</td>
<td>log(30)</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>New Model</td>
<td>2500</td>
<td>−1.1</td>
<td>log(30)</td>
<td>400</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Competitive alternatives for the same data

Charles Y. Tan MBSW May 19th, 2009

Sigmoid curves and a case for close-to-linear nonlinear models
Spectral decomposition of $\mathbf{F}'\mathbf{F}_0$: four-parameter models

<table>
<thead>
<tr>
<th>Model</th>
<th>Eigenvalues</th>
<th>Eigenvectors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$a$ $b$ $c$ $d$</td>
</tr>
<tr>
<td>4PL</td>
<td>$2.7 \times 10^6$</td>
<td>0 0 1.0 0</td>
</tr>
<tr>
<td></td>
<td>$4.0 \times 10^5$</td>
<td>0 1.0 0 0</td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>0.88 0 0 0.48</td>
</tr>
<tr>
<td></td>
<td>0.74</td>
<td>0.48 0 0 -0.88</td>
</tr>
<tr>
<td>4PG</td>
<td>$2.6 \times 10^6$</td>
<td>0 0.14 0.99 0</td>
</tr>
<tr>
<td></td>
<td>$1.1 \times 10^6$</td>
<td>0 -0.99 0.14 0</td>
</tr>
<tr>
<td></td>
<td>1.8</td>
<td>0.36 0 0 0.93</td>
</tr>
<tr>
<td></td>
<td>0.74</td>
<td>0.93 0 0 -0.36</td>
</tr>
</tbody>
</table>
Spectral decomposition of $\mathbf{F'}\mathbf{F}$: five-parameter models

<table>
<thead>
<tr>
<th>Model</th>
<th>Eigenvalues</th>
<th>Eigenvectors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$a$</td>
</tr>
<tr>
<td>Richards</td>
<td>$2.6 \times 10^6$</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>$7.4 \times 10^5$</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>$5.4 \times 10^2$</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>−0.76</td>
</tr>
<tr>
<td></td>
<td>0.34</td>
<td>0.65</td>
</tr>
<tr>
<td>New</td>
<td>$2.6 \times 10^6$</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>$1.0 \times 10^6$</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>$1.4 \times 10^5$</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.87</td>
<td>−0.76</td>
</tr>
<tr>
<td></td>
<td>0.36</td>
<td>0.65</td>
</tr>
</tbody>
</table>
Sigmoid curves and a case for close-to-linear nonlinear models
Deviations from unity $\eta$

![Graph showing four different curves: 4PL, Richards, 4PG, and New, with values of $\eta$ ranging from $5 \times 10^{-4}$ to $2 \times 10^{-1}$ on the x-axis and from 0 to 4 on the y-axis.](image)
At a given $\sigma$: model $\alpha$ (x-axis) and residuals $\beta$ (y-axis)
Closeup of effective degrees: $\alpha$ and $\beta$ when $\gamma < 0.1$
New paradigm: close-to-linear nonlinear models

- Nonlinear regressions in general
  - Nonlinearity is complex and exceedingly local: \( H = F' F - [\epsilon'] [F_{..}] \)
  - Close-to-linear model is an unstated prerequisite for most statistical methods and numerical algorithms. Exception: bootstrapping.
  - Extending model for flexibility should only be done with sufficient justifications since the cost could be high.
New paradigm: close-to-linear nonlinear models

- Nonlinear regressions in general
  - Nonlinearity is complex and exceedingly local:
    \[ H = F' \cdot F - [\epsilon'] [F_{..}] \]
  - Close-to-linear model is an unstated prerequisite for most statistical methods and numerical algorithms. Exception: bootstrapping.
  - Extending model for flexibility should only be done with sufficient justifications since the cost could be high.

- Sigmoid curves in particular
  - Richards model (“5PL”) is NOT close-to-linear and its routine use is unjustifiable.
  - The proposed new model is (more) flexible and close-to-linear.
  - 4PL and 4PG are close-to-linear.
The model:

\[ f(x; a, b, c, d) = d + (a - d)\Phi(b(x - c)). \]

Linearizing function:

\[ \Phi^{-1}\left(\frac{u - d}{a - d}\right) = b(x - c) \]

Since \( f(x; a, b, c, d) \) is the same curve as \( f(x; d, -b, c, a) \), the condition of \( a > d \) or \( a < d \) is needed to resolve the identifiability problem.
Backup

Generalized linear models vs sigmoid curves

- Link function: link mean to linear predictor
  - Logit link
  - Probit link
  - Log-log link
- IRLS works.
- Profile likelihood is preferred over Wald’s.

- Linearization function: linearize standardized response to linear regressor
  - Logit curve
  - Probit curve
  - Gompertz curve
- Close-to-linear
- Some PE curvature when design and parameterization mismatch.
Let $A$ be a square matrix and $\epsilon \sim N(0, \sigma^2 I)$, then $\mathbb{E}(\epsilon' A \epsilon / \sigma^2) = \text{tr}(A)$ and $\mathbb{V}(\epsilon' A \epsilon / \sigma^2) = 2 \text{tr}(A^2)$.

- $A$ is idempotent: $\epsilon' A \epsilon / \sigma^2 \sim \chi^2(r)$ and $r = \text{tr}(A) = \text{rank}(A)$
- $A$ is not idempotent: $(s_1/s_2)(\epsilon' A \epsilon / \sigma^2)$ matches the first two moments of $\chi^2(s_1^2/s_2)$, where $s_1 = \text{tr}(A)$ and $s_2 = \text{tr}(A^2)$.

If $AB = 0$, then $\epsilon' A \epsilon$ is independent of $\epsilon' B \epsilon$. 
For any matrix norm: \( \mathbf{C} = 0 \iff \| \mathbf{C} \| = 0 \)

Frobenius norm:  
\[
\| \mathbf{C} \| = \sum_i \sum_j c_{ij}^2 = \text{tr}(\mathbf{C}^2)
\]

\( \gamma \) is normalized so that 0 \( \leq \gamma \leq 1 \)

\[
\gamma = \frac{\| \mathbf{F} \mathbf{H}^{-1} \mathbf{F}' (\mathbf{I} - \mathbf{F} \mathbf{H}^{-1} \mathbf{F}') \|}{\| \mathbf{F} \mathbf{H}^{-1} \mathbf{F}' \| \| \mathbf{I} - \mathbf{F} \mathbf{H}^{-1} \mathbf{F}' \|} = \sqrt{\frac{t_2 - 2t_3 + t_4}{t_2(n - 2t_1 + t_2)}}
\]
Flexibility of the new model: the effect of $g$

- - - $g = 1.5$
- - - $g = 1$
- - - $g = 0.75$

- - - $g = 0.5$
- - - $g = 0.25$
- - - $g = 0$
- - - $g = -0.5$

Sigmoid curves and a case for close-to-linear nonlinear models