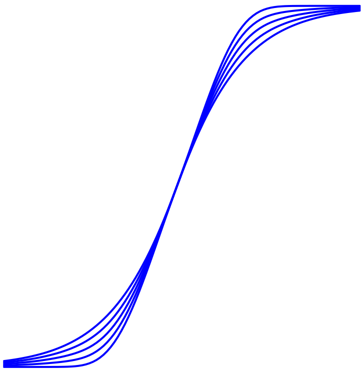
Sigmoid curves and a case for close-to-linear nonlinear models

Charles Y. Tan
charles_tan@merck.com

Merck Research Laboratories / MSD
West Point, Pennsylvania

Nonclinical Statistics Conference 2008

Outline



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)
  - Ratkowsky (1983, 1990): "significant intrinsic curvature", "a particularly unfortunate model", "abuse of Occams Razor"
  - Seber and Wild (1989): "Bad ill-conditioning and convergence problems"

Nonlinear regression

$$y_i = f(x_i; \boldsymbol{\theta}) + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

- Nonlinearity of f with respect to $\boldsymbol{\theta}$: defining characteristics
- Nonlinearity of f with respect to x: incidental

Nonlinear regression

$$y_i = f(x_i;\boldsymbol{\theta}) + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

- Nonlinearity of f with respect to $\boldsymbol{\theta}$: defining characteristics
- Nonlinearity of f with respect to x: incidental
- Homogeneous variance: ε_i's are i.i.d. $N(0, \sigma^2)$
  Maximum Likelihood = Least Squares
  Objective function:

$$S(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}))' (\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}))$$

1st order approximation of the model

$$\mathbf{f}(\boldsymbol{\theta}) \approx \mathbf{f}(\boldsymbol{\theta}^*) + \mathbf{F}_\bullet(\boldsymbol{\theta} - \boldsymbol{\theta}^*),$$

where

$$\mathbf{F}_\bullet = \mathbf{F}_\bullet(\boldsymbol{\theta}^*) = \left(\left. \frac{\partial f(x_i; \boldsymbol{\theta})}{\partial \theta_j} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right)_{n \times k}$$

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

$$S(\boldsymbol{\theta}) \approx \varepsilon' \varepsilon - 2\varepsilon' \mathbf{F}_\bullet(\boldsymbol{\theta} - \boldsymbol{\theta}^*) + (\boldsymbol{\theta} - \boldsymbol{\theta}^*)' \mathbf{F}_\bullet' \mathbf{F}_\bullet (\boldsymbol{\theta} - \boldsymbol{\theta}^*)$$

Common framework for inference

$$S(\theta^*) - S(\hat{\theta}) \approx (\hat{\theta} - \theta^*)'\mathbf{F}_\bullet'\mathbf{F}_\bullet(\hat{\theta} - \theta^*) \approx \varepsilon'\mathbf{F}_\bullet(\mathbf{F}_\bullet'\mathbf{F}_\bullet)^{-1}\mathbf{F}_\bullet'\varepsilon$$

$$S(\hat{\theta}) \approx \varepsilon'(\mathbf{I} - \mathbf{F}_\bullet(\mathbf{F}_\bullet'\mathbf{F}_\bullet)^{-1}\mathbf{F}_\bullet')\varepsilon$$

Common framework for inference

$$S(\theta^*) - S(\hat{\theta}) \approx (\hat{\theta} - \theta^*)'\mathbf{F}_\bullet'\mathbf{F}_\bullet(\hat{\theta} - \theta^*) \approx \varepsilon'\mathbf{F}_\bullet(\mathbf{F}_\bullet'\mathbf{F}_\bullet)^{-1}\mathbf{F}_\bullet'\varepsilon$$

$$S(\hat{\theta}) \approx \varepsilon'(\mathbf{I} - \mathbf{F}_\bullet(\mathbf{F}_\bullet'\mathbf{F}_\bullet)^{-1}\mathbf{F}_\bullet')\varepsilon$$

Since $\mathbf{F}_\bullet(\mathbf{F}_\bullet'\mathbf{F}_\bullet)^{-1}\mathbf{F}_\bullet'$ is idempotent

Common framework for inference

$$S(\boldsymbol{\theta}^*) - S(\hat{\boldsymbol{\theta}}) \approx (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)' \mathbf{F}_\bullet' \mathbf{F}_\bullet (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \approx \boldsymbol{\varepsilon}' \mathbf{F}_\bullet (\mathbf{F}_\bullet' \mathbf{F}_\bullet)^{-1} \mathbf{F}_\bullet' \boldsymbol{\varepsilon}$$

$$S(\hat{\boldsymbol{\theta}}) \approx \boldsymbol{\varepsilon}' (\mathbf{I} - \mathbf{F}_\bullet (\mathbf{F}_\bullet' \mathbf{F}_\bullet)^{-1} \mathbf{F}_\bullet') \boldsymbol{\varepsilon}$$

Since $\mathbf{F}_\bullet (\mathbf{F}_\bullet' \mathbf{F}_\bullet)^{-1} \mathbf{F}_\bullet'$ is idempotent

Local inference:

$$\frac{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)' \mathbf{F}_\bullet' \mathbf{F}_\bullet (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)}{S(\hat{\boldsymbol{\theta}})} \sim \frac{k}{n-k} F_{k,n-k}$$

Global inference:

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

Intrinsic and parameter-effect curvatures

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

$$\mathbf{f}(\boldsymbol{\theta}) \approx \mathbf{f}(\boldsymbol{\theta}^*) + \mathbf{F}_{\bullet}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)$$

Intrinsic and parameter-effect curvatures

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

$$\mathbf{f}(\boldsymbol{\theta}) \approx \mathbf{f}(\boldsymbol{\theta}^*) + \mathbf{F}_{\bullet}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)$$

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

Intrinsic and parameter-effect curvatures

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

$$\mathbf{f}(\boldsymbol{\theta}) \approx \mathbf{f}(\boldsymbol{\theta}^*) + \mathbf{F}_\bullet(\boldsymbol{\theta} - \boldsymbol{\theta}^*)$$

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

Curvatures (nonlinearity) are local properties

- The model f
- The parameters $\boldsymbol{\theta}$
  - Parameterization
  - Values
- The design $\mathbf{x}$
  - Sample size
  - Values
- The particular realization of ε

Ratkowsky's concept: close-to-linear

- Asymptotically, i.e., $n \longrightarrow \infty$ or $\sigma \longrightarrow 0$, 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 σ.

Shared parameterization to make a fair comparison

Let x denote the independent variable. Let $\boldsymbol{\theta}$ be either (a, b, c, d) for four-parameter models or (a, b, c, d, g) for five-parameter models. Let $u = f(x; \boldsymbol{\theta})$. 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 $b(x - c)$.

Symmetry and inflection point

- 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:

$$\operatorname{logit}\left(\frac{u - d}{a - d}\right) = b(x - c)$$

Four-parameter logistic (4PL) curve

- The model:

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

- Linearizing function:

$$\mathrm{logit}\left(\frac{u - d}{a - d}\right) = b(x - c)$$

- 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:

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

Richards model ("5PL")

- The model:

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

- Linearizing function:

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

Richards model ("5PL")

- The model:

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

- 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 $g : 0 \to +\infty$, 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 $\mathbb{R}$
- “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:

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

Four-parameter Gompertz (4PG) curve

- The model:

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

- Linearizing function:

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

Four-parameter Gompertz (4PG) curve

- The model:

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

- Linearizing function:

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

- Asymmetric sigmoid curve

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 $\log(\log 2)/b + c > c$.

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 $g \to +\infty$.

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

- The model:

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

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

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: $S(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}))' (\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}))$

Use the "complete" to assess the "partial"

- Original objective function: $S(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}))' (\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}))$

- Partial 2nd order expansion of the objective:

$$S(\boldsymbol{\theta}) \approx \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - 2\boldsymbol{\varepsilon}'\mathbf{F}_{\bullet}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) + (\boldsymbol{\theta} - \boldsymbol{\theta}^*)'\mathbf{F}_{\bullet}'\mathbf{F}_{\bullet}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)$$

Use the "complete" to assess the "partial"

- Original objective function: $S(\boldsymbol{\theta}) = (\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}))' (\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}))$
- *Complete* 2nd order expansion of the objective:

$$S(\boldsymbol{\theta}) \approx \varepsilon'\varepsilon - 2\varepsilon'\mathbf{F}_{\bullet}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) + (\boldsymbol{\theta} - \boldsymbol{\theta}^*)'\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)$$

where

$$\mathbf{H} = \frac{1}{2}\nabla^2 S(\boldsymbol{\theta}^*) = \mathbf{F}_{\bullet}'\mathbf{F}_{\bullet} - [\varepsilon'] [\mathbf{F}_{\bullet\bullet}]$$

$$[\varepsilon'] [\mathbf{F}_{\bullet\bullet}] = \left(\sum_{i=1}^{n} \varepsilon_i \frac{\partial^2 f(x_i; \boldsymbol{\theta})}{\partial\theta_r \partial\theta_s} \bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \right)_{k \times k}$$

- Partial 2nd order expansion of the objective:

$$S(\boldsymbol{\theta}) \approx \varepsilon'\varepsilon - 2\varepsilon'\mathbf{F}_{\bullet}(\boldsymbol{\theta} - \boldsymbol{\theta}^*) + (\boldsymbol{\theta} - \boldsymbol{\theta}^*)'\mathbf{F}_{\bullet}'\mathbf{F}_{\bullet}(\boldsymbol{\theta} - \boldsymbol{\theta}^*)$$

Quantify close-to-linear-ness by comparing $\mathbf{H}$ to $\mathbf{F}_\bullet'\mathbf{F}_\bullet$

$$\mathbf{H} = \mathbf{F}_\bullet'\mathbf{F}_\bullet - [\varepsilon'] [\mathbf{F}_{\bullet\bullet}]$$

- For linear models: $\mathbf{F}_{\bullet\bullet} = \mathbf{0}$, hence $\mathbf{H} = \mathbf{F}_\bullet'\mathbf{F}_\bullet$

Quantify close-to-linear-ness by comparing $\mathbf{H}$ to $\mathbf{F}'_{\bullet}\mathbf{F}_{\bullet}$

$$\mathbf{H} = \mathbf{F}'_{\bullet}\mathbf{F}_{\bullet} - [\varepsilon'] [\mathbf{F}_{\bullet\bullet}]$$

- For linear models: $\mathbf{F}_{\bullet\bullet} = \mathbf{0}$, hence $\mathbf{H} = \mathbf{F}'_{\bullet}\mathbf{F}_{\bullet}$
- As $\sigma \to 0$: $\mathbf{H} \to \mathbf{F}'_{\bullet}\mathbf{F}_{\bullet}$ almost surely
- As $n \to \infty$: $\mathbf{H} \to \mathbf{F}'_{\bullet}\mathbf{F}_{\bullet}$ almost surely

Quantify close-to-linear-ness by comparing $\mathbf{H}$ to $\mathbf{F}_\bullet'\mathbf{F}_\bullet$

$$\mathbf{H} = \mathbf{F}_\bullet'\mathbf{F}_\bullet - \left[\varepsilon'\right]\left[\mathbf{F}_{\bullet\bullet}\right]$$

- For linear models: $\mathbf{F}_{\bullet\bullet} = \mathbf{0}$, hence $\mathbf{H} = \mathbf{F}_\bullet'\mathbf{F}_\bullet$
- As $\sigma \to 0$: $\mathbf{H} \to \mathbf{F}_\bullet'\mathbf{F}_\bullet$ almost surely
- As $n \to \infty$: $\mathbf{H} \to \mathbf{F}_\bullet'\mathbf{F}_\bullet$ almost surely
- For any σ and n: $\mathcal{E}(\mathbf{H}) = \mathbf{F}_\bullet'\mathbf{F}_\bullet$

Geometry of $S(\theta)$ and eigenvalues of $\mathbf{H}$

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

Geometry of $S(\theta)$ and eigenvalues of $\mathbf{H}$

- All eigenvalues are positive: $S(\boldsymbol{\theta})$ near $\boldsymbol{\theta}^*$ is elliptic paraboloid like and has a minimum.
- Some of the eigenvalues are negative: $S(\boldsymbol{\theta})$ near $\boldsymbol{\theta}^*$ is hyperbolic paraboloid like (*non-informative*).

Geometry of $S(\boldsymbol{\theta})$ and eigenvalues of $\mathbf{H}$

- All eigenvalues are positive: $S(\boldsymbol{\theta})$ near $\boldsymbol{\theta}^*$ is elliptic paraboloid like and has a minimum.
- Some of the eigenvalues are negative: $S(\boldsymbol{\theta})$ near $\boldsymbol{\theta}^*$ is hyperbolic paraboloid like (*non-informative*).
  - The whole $S(\boldsymbol{\theta})$ is unbounded from below, no LS or ML solution: at least warned.

Geometry of $S(\theta)$ and eigenvalues of **H**

- All eigenvalues are positive: $S(\boldsymbol{\theta})$ near $\boldsymbol{\theta}^*$ is elliptic paraboloid like and has a minimum.
- Some of the eigenvalues are negative: $S(\boldsymbol{\theta})$ near $\boldsymbol{\theta}^*$ is hyperbolic paraboloid like (*non-informative*).
  - The whole $S(\boldsymbol{\theta})$ is unbounded from below, no LS or ML solution: at least warned.
  - $S(\boldsymbol{\theta})$ has (multiple) elliptic paraboloid like "pockets" away from the true value $\boldsymbol{\theta}^*$, nominal LS or ML solution can be found: misleading.

How close is $\mathbf{H}$ to $\mathbf{F}_\bullet'\mathbf{F}_\bullet$ overall

- Define *relative information content* τ as

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

How close is $\mathbf{H}$ to $\mathbf{F}_\bullet'\mathbf{F}_\bullet$ overall

- Define *relative information content* τ as

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

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

How close is $\mathbf{H}$ to $\mathbf{F}'_\bullet \mathbf{F}_\bullet$ overall

- Define *relative information content* τ as

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

- Define *probability of model failure* as $\xi = \Pr\{\tau < 0\}$
- Define *deviation from unity* η as $\eta^2 = E\left[(\tau - 1)^2 | \tau > 0\right]$

How close is $\mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}'_\bullet$ to idempotency

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

- $S(\boldsymbol{\theta}^*) - S(\hat{\boldsymbol{\theta}}) \approx \boldsymbol{\varepsilon}'\mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}'_\bullet \boldsymbol{\varepsilon}$
  - compared with $\boldsymbol{\varepsilon}'\mathbf{F}_\bullet(\mathbf{F}'_\bullet \mathbf{F}_\bullet)^{-1}\mathbf{F}'_\bullet \boldsymbol{\varepsilon}$
- $S(\hat{\boldsymbol{\theta}}) \approx \boldsymbol{\varepsilon}'(\mathbf{I} - \mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}'_\bullet)\boldsymbol{\varepsilon}$
  - compared with $\boldsymbol{\varepsilon}'(\mathbf{I} - \mathbf{F}_\bullet(\mathbf{F}'_\bullet \mathbf{F}_\bullet)^{-1}\mathbf{F}'_\bullet)\boldsymbol{\varepsilon}$
- Dependence of $S(\boldsymbol{\theta}^*) - S(\hat{\boldsymbol{\theta}})$ and $S(\hat{\boldsymbol{\theta}})$ is measured by $\|\mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}'_\bullet(\mathbf{I} - \mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}'_\bullet)\|$ (after normalization)
  - compared with independence

Three effective degrees

Let $t_1 = \mathrm{tr}(\mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}'_\bullet)$, $t_2 = \mathrm{tr}\big((\mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}'_\bullet)^2\big)$,
$t_3 = \mathrm{tr}\big((\mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}'_\bullet)^3\big)$ and $t_4 = \mathrm{tr}\big((\mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}'_\bullet)^4\big)$

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

$$\alpha = \frac{t_1^2}{t_2}$$

Three effective degrees

Let $t_1 = \mathrm{tr}(\mathbf{F_\bullet H^{-1}F'_\bullet})$, $t_2 = \mathrm{tr}\left((\mathbf{F_\bullet H^{-1}F'_\bullet})^2\right)$, $t_3 = \mathrm{tr}\left((\mathbf{F_\bullet H^{-1}F'_\bullet})^3\right)$ and $t_4 = \mathrm{tr}\left((\mathbf{F_\bullet H^{-1}F'_\bullet})^4\right)$

- 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 $t_1 = \mathrm{tr}(\mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}'_\bullet)$, $t_2 = \mathrm{tr}\big((\mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}'_\bullet)^2\big)$, $t_3 = \mathrm{tr}\big((\mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}'_\bullet)^3\big)$ and $t_4 = \mathrm{tr}\big((\mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}'_\bullet)^4\big)$

- 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

$$\beta = \frac{(n - t_1)^2}{n - 2t_1 + t_2}$$

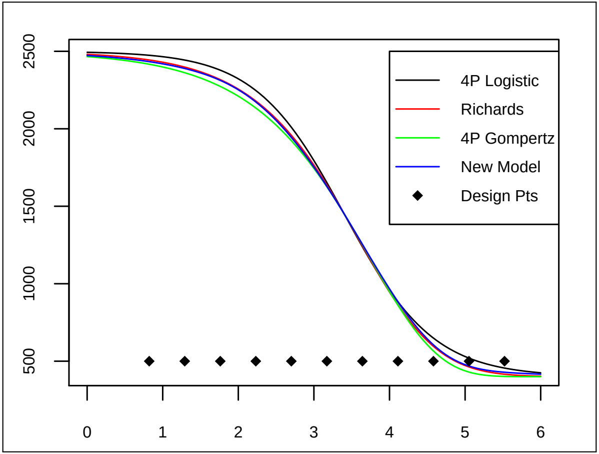
- Define *effective degree of dependence* as

$$\gamma = \sqrt{\frac{t_2 - 2t_3 + t_4}{t_2(n - 2t_1 + t_2)}}$$

Four particular curves: from a cell based bioassay

Model	a	b	c	d	g
4P Logistic	2500	-1.7	$\log(30)$	400	
Richards	2500	-1.3	$\log(30)$	400	3
4P Gompertz	2500	-1.1	$\log(30)$	400	
New Model	2500	-1.1	$\log(30)$	400	0.8

Competitive alternatives for the same data



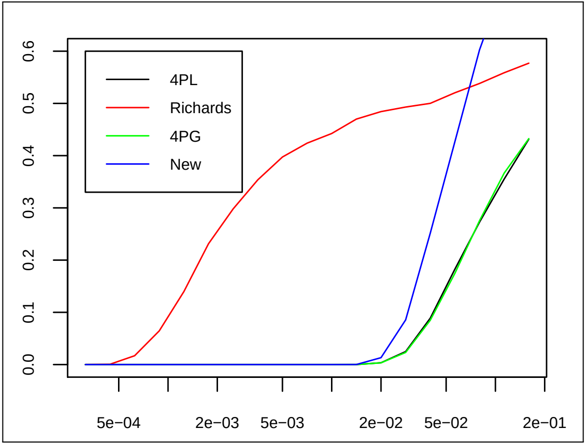
Spectral decomposition of $\mathbf{F}_\bullet'\mathbf{F}_\bullet$: four-parameter models

Model	Eigen-values	Eigenvectors *a*	*b*	*c*	*d*
4PL	2.7×10^6	0	0	1.0	0
	4.0×10^5	0	1.0	0	0
	2.0	0.88	0	0	0.48
	0.74	0.48	0	0	−0.88
4PG	2.6×10^6	0	0.14	0.99	0
	1.1×10^6	0	−0.99	0.14	0
	1.8	0.36	0	0	0.93
	0.74	0.93	0	0	−0.36

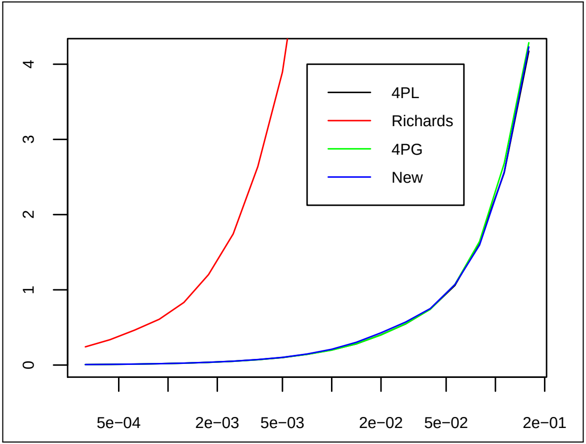
Spectral decomposition of $\mathbf{F}_\bullet'\mathbf{F}_\bullet$: five-parameter models

Model	Eigenvalues	Eigenvectors *a*	*b*	*c*	*d*	*g*
Richards	2.6×10^6	0	0	1.0	0	0
	7.4×10^5	0	−1.0	0	0	0
	5.4×10^2	0	0	0	0	1.0
	0.80	−0.76	0	0	0.65	0
	0.34	0.65	0	0	0.76	0
New	2.6×10^6	0	0	0.98	0	−0.17
	1.0×10^6	0	1.0	0	0	0
	1.4×10^5	0	0	−0.18	0	−0.98
	0.87	−0.76	0	0	0.65	0
	0.36	0.65	0	0	0.76	0

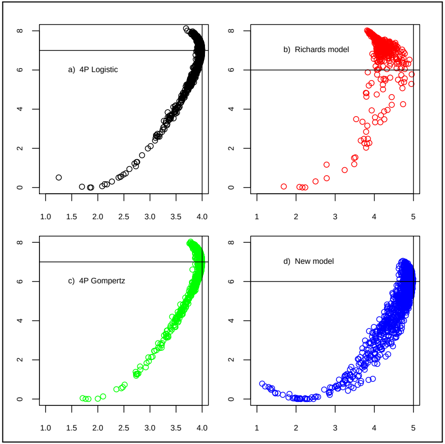
Probability of model failure ξ



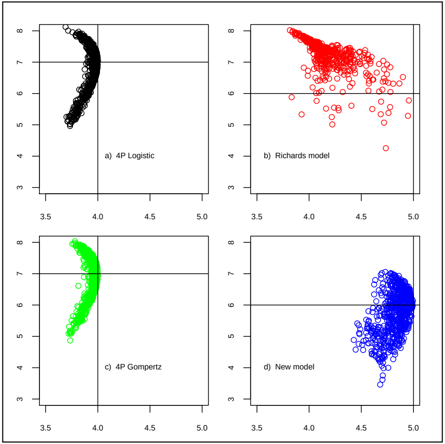
Deviation from unity η



At a given σ: model α (x-axis) and residuals β (y-axis)



Closeup of effective degrees: α and β when $\gamma < 0.1$



New paradigm: close-to-linear nonlinear models

- Nonlinear regressions in general
  - Nonlinearity is complex and exceedingly local: $\mathbf{H} = \mathbf{F}'_{\bullet}\mathbf{F}_{\bullet} - [\varepsilon'][\mathbf{F}_{\bullet\bullet}]$
  - 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:
    $\mathbf{H} = \mathbf{F}'_{\bullet}\mathbf{F}_{\bullet} - [\varepsilon'] [\mathbf{F}_{\bullet\bullet}]$
  - 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.

Four-parameter probit (4PP) curve

- The model:

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

- Linearizing function:

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

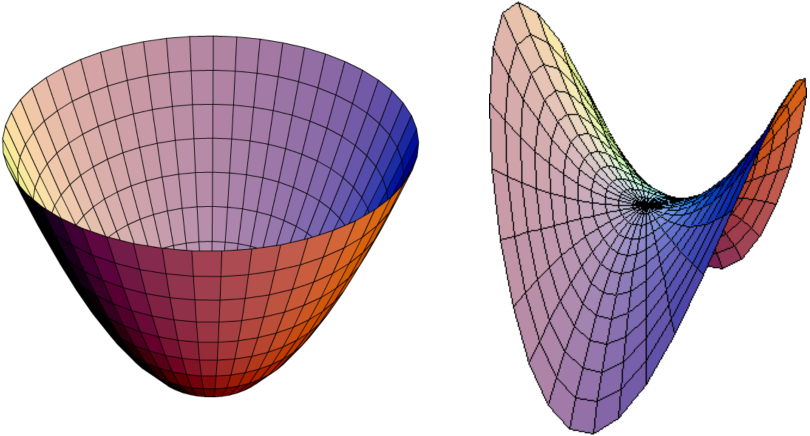
- Since $f(x; a, b, c, d)$ is the same curve as $f(x; d, -b, c, a)$, the condition of $a > d$ or $a < d$ is needed to resolve the identifiability problem.

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.

Paraboloid: elliptic (left) and hyperbolic (right)



Distribution of quadratic forms

Let $\mathbf{A}$ be a square matrix and $\varepsilon \sim N(0, \sigma^2 \mathbf{I})$, then $\mathcal{E}(\varepsilon' \mathbf{A} \varepsilon / \sigma^2) = \mathrm{tr}(\mathbf{A})$ and $\mathcal{V}(\varepsilon' \mathbf{A} \varepsilon / \sigma^2) = 2\mathrm{tr}(\mathbf{A}^2)$.

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

Usual matrix norm: Frobenius norm

- For any matrix norm: $\mathbf{A} = \mathbf{0} \iff \|\mathbf{A}\| = 0$
- Frobenius norm: $\|\mathbf{A}\| = \sum_i \sum_j a_{ij}^2 = \mathrm{tr}(\mathbf{A}^2)$
- γ is normalized so that $0 \leq \gamma \leq 1$

$$\gamma = \frac{\|\mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}_\bullet' (\mathbf{I} - \mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}_\bullet')\|}{\|\mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}_\bullet'\| \|\mathbf{I} - \mathbf{F}_\bullet \mathbf{H}^{-1} \mathbf{F}_\bullet'\|} = \sqrt{\frac{t_2 - 2t_3 + t_4}{t_2(n - 2t_1 + t_2)}}$$

Flexibility of the new model: the effect of g