Some time series models (not in RWC)

There are ≥ 3 different-looking methods for analyzing time series:

- frequency-domain methods;
- autoregressive-moving-average (ARMA) models, aka Box-Jenkins;
- state-space models, aka dynamic linear models (DLMs).

We'll focus on DLMs, a close relative of Kalman filters.

A DLM can be used as a

- filter: estimate a system's state in real time.
- smoother: estimate a system's state *post hoc* at a series of times.

In this course, we talk about DLMs used as smoothers.

ARMA models can be written as DLMs.

Example of a DLM: Linear Growth Model

We have a series of observations $y_i, i = 0, \ldots, T$.

A DLM has two parts:

Observation equation: Model y_t as a function of the state μ_t:

$$y_t = \mu_t + n_t, \quad t = 0, \cdots, T,$$

$n_t \sim N(0, \sigma_n^2)$ is observation error.

State equation: Model the state μ_t's evolution over time:

$$\begin{aligned} \mu_t &= \mu_{t-1} + \theta_{t-1} + w_{1,t}, \quad t = 1, \cdots, T, \\ \theta_t &= \theta_{t-1} + w_{2,t}, \quad t = 1, \cdots, T-1, \end{aligned}$$

μ_t is the current level, θ_t is the current slope or time trend, $w_{1,t} \sim N(0, \sigma_1^2)$, $w_{2,t} \sim N(0, \sigma_2^2)$ are evolution "error"s.

The DLM literature customarily adds

$(\mu_0, \theta_0) \sim$ Normal with specified mean and variance.

This is sometimes called a prior distribution even by those who do a maximum-likelihood analysis.

Such a prior is necessary for filtering, specifically to allow Bayesian updating of the posterior for the state (μ_t, θ_t) at each time t.

In smoothing, μ_0 and θ_0 are often given mean zero and large variances.

This DLM in constraint-case form

It is easier to write this in constraint-case form than as a MLM:

$$\begin{array}{rclllll} y_t & = & \mu_t & & +n_t, & t = 0, \cdots, T \\ 0 & = & \mu_{t-1} - \mu_t & +\theta_{t-1} & +w_{1,t} & t = 1, \cdots, T \\ 0 & = & & +\theta_{t-1} - \theta_t & +w_{2,t} & t = 1, \cdots, T-1. \end{array}$$

Note: the index t has different ranges in the three equations.

Hodges (2014) writes this as one large equation.

This DLM written as a mixed linear model

Observation equation: $y_t = \mu_t + n_t, \quad t = 0, \cdots, T,$

State equation:

$$\begin{aligned}
\mu_t &= \mu_{t-1} + \theta_{t-1} + w_{1,t}, \quad t = 1, \cdots, T, \\
\theta_t &= \theta_{t-1} + w_{2,t}, \quad t = 1, \cdots, T-1,
\end{aligned}$$

Re-parameterize θ_t:

$$\begin{aligned}
\theta_1 &= \theta_0 + w_{2,1} \\
\theta_2 &= \theta_1 + w_{2,2} = \theta_0 + \sum_{i=1}^{2} w_{2,i} \\
\theta_3 &= \theta_2 + w_{2,3} = \theta_0 + \sum_{i=1}^{3} w_{2,i} \\
&\vdots \\
\theta_t &= \theta_0 + \sum_{i=1}^{t} w_{2,i}.
\end{aligned}$$

State equation:

$$\begin{aligned}
\mu_t &= \mu_{t-1} + \theta_{t-1} + w_{1,t}, \quad t = 1, \cdots, T, \\
\theta_t &= \theta_0 + \sum_{i=1}^{t} w_{2,i}, \qquad t = 1, \cdots, T-1
\end{aligned}$$

Now re-parameterize μ_t:

$$\begin{aligned}
\mu_1 &= \mu_0 + w_{1,1} + \theta_0 \\
\mu_2 &= \mu_1 + w_{1,2} + \theta_1 = \mu_0 + \sum_{i=1}^{2} w_{1,i} + 2\theta_0 + w_{2,1} \\
\mu_3 &= \mu_2 + w_{1,3} + \theta_2 = \mu_0 + \sum_{i=1}^{3} w_{1,i} + 3\theta_0 + 2w_{2,1} + w_{2,2} \\
&\vdots \\
\mu_t &= \mu_0 + \sum_{i=1}^{t} w_{1,i} + t\theta_0 + \sum_{i=1}^{t-1} (t-i) w_{2,i}.
\end{aligned}$$

$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ \vdots & \vdots \\ 1 & T \end{pmatrix} \begin{pmatrix} \mu_0 \\ \theta_0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & \cdots & & & 0 \\ 1 & 0 & \cdots & & & 0 \\ 1 & 1 & 0 & \cdots & & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & & \ddots & \vdots \\ 1 & 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} w_{11} \\ w_{12} \\ \vdots \\ w_{1T} \end{pmatrix}$$

$$+ \begin{pmatrix} 0 & 0 & \cdots & & & 0 \\ 0 & 0 & \cdots & & & 0 \\ 1 & 0 & 0 & \cdots & & 0 \\ 2 & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & & \ddots & \vdots \\ T-1 & T-2 & T-3 & T-4 & \cdots & 1 \end{pmatrix} \begin{pmatrix} w_{21} \\ w_{22} \\ \vdots \\ w_{2,T-1} \end{pmatrix} + \begin{pmatrix} n_0 \\ n_1 \\ \vdots \\ n_T \end{pmatrix}$$

This is almost in MLM form: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{Z}_2\mathbf{u}_2 + \boldsymbol{\epsilon}$.

This model is doubly saturated; users often set $w_{1t} = 0$, then this is ...

DLM for r-dimensional outcome $\mathbf{y}_t$

Observation equation: $\mathbf{y}_t = \mathbf{F}_t \boldsymbol{\theta}_t + \mathbf{n}_t, \quad \mathbf{n}_t \sim N_r(0, \Sigma_t^n)$,

$\mathbf{F}_t$ $r \times p$ is known; $\boldsymbol{\theta}$ $p \times 1$, $\mathbf{n}_t$ $r \times 1$ are unknown; Σ_t^n $r \times r$ is either.

State equation: $\boldsymbol{\theta}_t = \mathbf{H}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N_p(0, \Sigma_t^w)$,

$\mathbf{H}_t$ and Σ_t^w are $p \times p$; $\mathbf{H}_t$ is known, Σ_t^w is known or unknown.

$\boldsymbol{\theta}_0$ usually has a fully specified p-variate normal prior.

This defines a huge class of models with covariates, intervention effects, flexible cyclic and quasi-cyclic effects. Example coming next slide!

The linear growth model has $r = 1$, $p = 2$, $\boldsymbol{\theta}_t = (\mu_t, \theta_t)'$, $\mathbf{F}_t = [1\ 0]$,

$$\mathbf{H}_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \Sigma_t^n = \sigma_n^2, \text{ and } \Sigma_t^w = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}.$$

West & Harrison (1997) is an encyclopedia of DLMs.

Example: Localizing epileptic activity (Lavine et al)

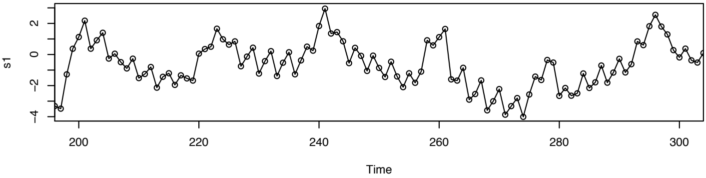
y_t = % change in average pixel value for light of wavelength 535 nm,
$t = 0, \ldots, 649$, with time steps of 0.28 sec.

Stimulus was applied during time steps $t = 75$ to 94

Object: Estimate the response to the stimulus.

Complication: artifacts from heartbeat and breathing (respiration), with periods of 2–4 and 15–25 time steps.

Here is ∼100 time steps:



Use a DLM to filter out the artifacts, smooth the response

y_t = % change in average pixel value for light of wavelength 535 nm, $t = 0, \ldots, 649$, with time steps of 0.28 sec.

Stimulus was applied during time steps $t = 75$ to 94

Model: a DLM with observation equation

$$y_t = s_t + h_t + r_t + v_t$$

- s_t is the smoothed response, the object of this analysis;
- h_t, r_t are heartbeat and respiration respectively;
- v_t is iid $N(0, W_v)$ error.

State equations for s_t, h_t, r_t

State equation for s_t is the linear growth model:

$$\begin{pmatrix} s_t \\ \text{slope}_t \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} s_{t-1} \\ \text{slope}_{t-1} \end{pmatrix} + \mathbf{w}_{s,t},$$

$$\mathbf{w}'_{s,t} = (0, w_{slope,t}) \text{ and } w_{slope,t} \sim \text{ iid } N(0, W_s).$$

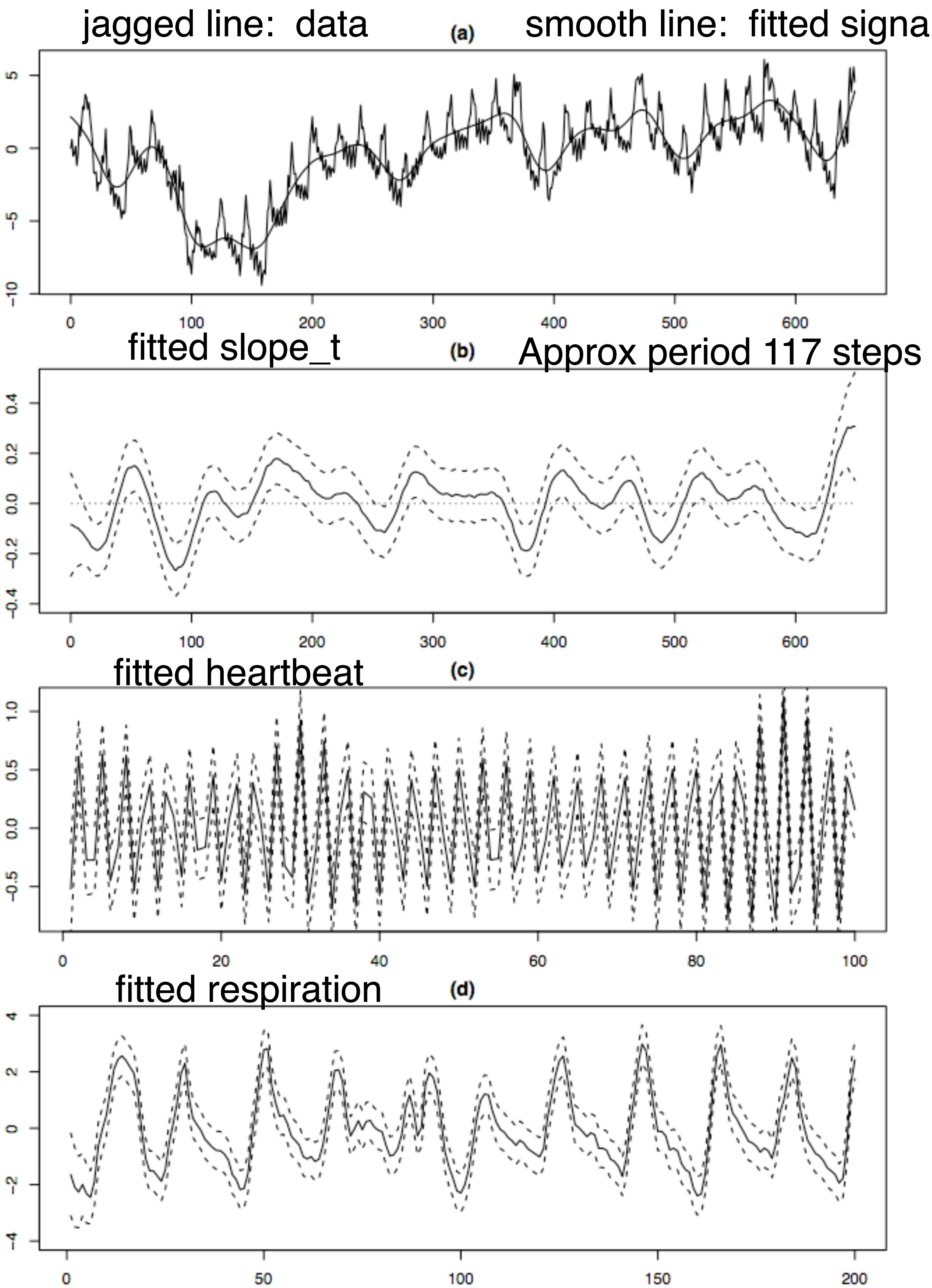
State equation for quasi-cyclic components (this is for heartbeat):

$$\begin{pmatrix} b_t \cos\alpha_t \\ b_t \sin\alpha_t \end{pmatrix} = \begin{bmatrix} \cos\delta_h & \sin\delta_h \\ -\sin\delta_h & \cos\delta_h \end{bmatrix} \begin{pmatrix} b_{t-1}\cos\alpha_{t-1} \\ b_{t-1}\sin\alpha_{t-1} \end{pmatrix} + \mathbf{w}_{h,t},$$

$$\mathbf{w}'_{h,t} = (w_{h1,t}, w_{h2,t}) \sim \text{ iid } N_2(0, \mathbf{W}_h) \text{ for } \mathbf{W}_h = W_h\mathbf{I}_2.$$

Periods: Heartbeat 2.78 time steps ($\delta_h = 1/2.78$); respiration 18.75.

Here's the fit of this model:



Alternative syntaxes for richly-parameterized models

Main syntax: Mixed linear models.

Key idea: Write models as mixed linear models by clever choice of $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{G}$, and $\mathbf{R}$.

Key tools:

- Mixed linear model theory, methods, and computing, and ideas adapted from simple linear models.
- The conventional analysis uses the restricted likelihood, large-sample approximations, and bootstrapping.
- Bayesian methods rely on MCMC, INLA, or variational Bayes (aka approximate Bayes computing).

Alternative #1: Gaussian Markov random fields (Rue & Held 2005)

Key idea: Represent components of models and priors as Gaussian Markov random fields (GMRFs), using conditional dependence.

Key tools:

- Model the mean structure in a modular fashion, with components being GMRFs or simple effects (e.g., fixed effects).
- Rue & Held emphasize Bayesian analyses.
- "Exact" MCMC exploits sparse precision matrices.
- Approximate analyses use INLA.
- Many models can be represented in this syntax, at worst closely analogous to models we've expressed as MLMs.

Alternative #2: Likelihood inference for models with unobservables (Lee et al 2006)

Key ideas: Extend generalized linear models in several directions using likelihood-like functions.

Key tools:

- Modular modelling of the observation error (exponential family), the linear predictor, error dispersion, random effects dispersion.
- Can handle unobservable random variables other than random effects, e.g., missing data or predictions.
- Analysis: Estimates are maxima of likelihood-like functions; uncertainty is described using the curvature at the maximum.

Alternative #1: GMRFs, the key idea

If $\mathbf{x} = (x_1, \ldots, x_n)' \sim$ MV Normal, the precision matrix expresses conditional dependence and independence.

If $\mathbf{x} \sim$ Normal with mean $\boldsymbol{\mu}$ and precision matrix $\mathbf{P}$, then

$$\text{cov}(x_i, x_j | \mathbf{x}_{(-ij)}) = 0 \Leftrightarrow P_{ij} = 0$$

where • $\mathbf{x}_{(-ij)}$ is $\mathbf{x}$ without its i^{th} and j^{th} elements,

• P_{ij} is the $(i,j)^{th}$ element of $\mathbf{P}$.

If many $P_{ij} = 0$, $\mathbf{P}$ is "sparse" and computing can exploit this.

Many familiar models have sparse $\mathbf{P}$.

Examples: One-way random effects model, ICAR model

One-way RE model: Model $y_{ij} = \theta_i + \epsilon_{ij}, \quad \theta_i = \mu + \delta_i$,
with ϵ_{ij}, $\delta_i \sim$ Normal and mutually independent.

$$\begin{aligned} \text{cov}(y_{ij}, \theta_{i'}|\theta_i) &= 0 \text{ for } i \neq i' \\ \text{cov}(y_{ij}, \mu|\theta_i) &= 0 \\ \text{cov}(\theta_{i'}, \theta_i|\mu) &= 0. \end{aligned}$$

If $\mathbf{x}$ includes $\mathbf{y}$, $\boldsymbol{\theta}$, and $\mu \Rightarrow \mathbf{x}$'s precision matrix $\mathbf{P}$ is sparse.

ICAR: Model $y_i = \delta_i + \epsilon_i$ with $\epsilon_i \sim$ iid N(0, σ_e^2),
indep't of $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)' \sim$ ICAR with precision $\mathbf{Q}/\sigma_s^2$.

$$\begin{aligned} \text{cov}(y_i, \delta_j|\delta_i) &= 0 \text{ for } i \neq j \\ \text{cov}(\delta_j, \delta_i|\boldsymbol{\delta}_{(-ij)}) &= 0 \text{ if } i \text{ and } j \text{ are not neighbors.} \end{aligned}$$

Neighbor pairs are relatively few $\Rightarrow \mathbf{Q}$ is sparse $\Rightarrow \mathbf{P}$ is sparse.

Example: Autoregressive model of order 1 (AR1)

Suppose $x_t = \phi x_{t-1} + \epsilon_t$ with $|\phi| < 1$ and $\epsilon_t \sim$ iid N(0,1).

$\Rightarrow x_t | x_1, \ldots, x_{t-1} \sim N(\phi x_{t-1}, 1)$
$x_t | x_{t-1}$ is conditionally independent of $x_1, \ldots, x_{t-2}$
$x_t | x_{t-1}, x_{t+1}$ is independent of $x_{t'}$ for $t' \notin \{t-1, t, t+1\}$.

If $x_1 \sim N(0, (1-\phi)^{-1}) \Rightarrow \mathbf{x}$ is a GMRF with precision matrix

$$\begin{bmatrix} 1 & -\phi & 0 & 0 & & 0 & 0 & 0 \\ -\phi & 1+\phi^2 & -\phi & 0 & \ldots & 0 & 0 & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi & & 0 & 0 & 0 \\ & \vdots & & & \ddots & & \vdots & \\ 0 & 0 & 0 & 0 & & -\phi & 1+\phi^2 & -\phi \\ 0 & 0 & 0 & 0 & \ldots & 0 & -\phi & 1 \end{bmatrix}.$$

Example: Dynamic linear model (DLM)

Observation equation $y_t = \mathbf{F}_t \boldsymbol{\theta}_t + \mathbf{n}_t$ with $\mathbf{n}_t \sim N_r(0, \Sigma_t^n)$ independently.

State equation $\boldsymbol{\theta}_t = \mathbf{H}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t$ with $\mathbf{w}_t \sim N_p(0, \Sigma_t^w)$ independently.

This gives a sparse $\mathbf{P}$ for the data y_t and $\boldsymbol{\theta}_t$:

$$\begin{aligned}
\text{cov}(y_t, \boldsymbol{\theta}_{t'} | \boldsymbol{\theta}_t) &= 0 \text{ if } t \neq t', \\
\text{cov}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t'} | \boldsymbol{\theta}_{t-1}) &= 0 \text{ if } t' < t - 1, \text{ and} \\
\text{cov}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t'} | \boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_{t+1}) &= 0 \text{ if } t' \notin \{t-1, t, t+1\}.
\end{aligned}$$

Any MLM has at least an analogous model here

Modeling = adding components for different features of the data.

In the combined vector of outcomes **y** and unknowns, the components are unconditionally independent of each other.

Simple random effect = a GMRF with a diagonal precision matrix.

Time series: DLMs are GMRFs; ARMA models can be written as DLMs.

Longitudinal analyses: See “Simple RE” and “Time Series”.

Graphical models: An edge between 2 nodes = conditional dependence.

Penalized splines: Rue & Held (2005) propose GMRFs using differences and the Weiner process.

Geostatistical models: GPs can be represented as GMRFs for computing, but this is not identical to the original GP.

Alternative #2: Likelihood Inference for Models with Unobservables

Generalized linear models (GLMs) have these key elements:

- Error distribution (exponential family);
- Linear predictor and link function; and
- Analysis using maximum likelihood, standard large-sample approximations, and IRLS for computing.

This system extends GLMs by:

- Adding random effects to the linear predictor.
- Modeling the error distribution's dispersion parameter with its own GLM and random effects.
- Modelling "unobservables," e.g., missing data and predictions.
- Analysis using the so-called h-likelihood; a model with all these pieces is analyzed as a series of linked GLMs.

This approach has generated some controversy

Commentators (e.g., on Lee & Nelder 2009):

- Lee et al propose some new models and unify existing models.
- They mainly disagree with Lee et al's claims about the value of their unified analytic approach.
- The model syntax and computing method are not controversial; the theory of analysis is.

Lee et al say about their theory of analysis:

- It's a principled extension of the Likelihood Principle.
- It "avoids prior probability" $\Rightarrow$ it's superior to Bayes.
- It solves all problems in analysis apart from minor technical issues they can solve with 2^{nd} order approximations.

My 2 cents worth on this

I and discussants of Lee & Nelder (2009) find these claims overstated.

Some simple points:

- The analysis approach is an *ad hoc* patchwork.
  - They do different things for different unknowns to avoid Bayes and avoid known problems.

- *Ad hoc*kery is OK if it performs; these methods cannot perform as claimed because of:
  - Multiple maxima
  - Maxima at boundary values
  - Measures of uncertainty defined using curvature at the maximum, rationalized by large-sample theory.

Lee et al (2006), Lee & Nelder (2009) do not mention these problems.

Summing up the first part of the course

A theory of a class of models like MLMs has two parts:

- A syntax for expressing many models.
- Tools for understanding analyses of models expressed in that syntax.

The right syntax, expressing many models, allows:

- Powerful, unified computing for a large class of models.
- Theory for many models simultaneously; precedents include linear models and generalized linear models.

What do I want in a theory of MLMs?

The obvious place to start is the tools we get from the powerful, beautiful theory of linear models:

- Find discrepant features of the data (residuals/outliers).
- Seek deviations form model assumptions (residuals: non-linearity, non-constant variance, transformations of $\mathbf{y}$).
- Seek data features with large influence on the results.
- Assess evidence for adding predictors (added variable plots).
- Understand indeterminate results and competition among predictors (collinearity).

We'll begin by looking at simple extensions of these ideas from linear model theory to MLMs.

BUT before we do that . . .

MLMs provide a whole new set of ways to generate mysteries and complications, and they're much more complicated than linear models.

$\Rightarrow$ We need to consider a different style for learning about our methods, a scientific style, complementing the traditional mathematical style.

The next lecture will demonstrate this scientific style on a problem that arises in fitting the "random regressions" model.

The rest of the course will use both styles.