Analysis of DNA Damage and Repair in Colonic Crypts

Jeffrey S. Morris
University of Texas, MD Anderson Cancer Center

Marina Vannucci and Raymond J. Carroll,
Texas A&M University
Outline

- Introduction
  - Colon Architecture and Carcinogenesis Studies
  - Hierarchical Functional Model
- Wavelet-based Estimation of Hierarchical Functions
- Kernel-based Estimation of Correlations
- Bayesian Modeling of Crypt Cell Architecture
- Conclusion
Architecture of Colon

- **Stem Cells**: Mother cells near crypt base
- **Depth in crypt**: ~ age of cells
- **Relative Cell Position**
  (Depth w/in crypt): \( t \in (0,1) \)
Colon Carcinogenesis Studies

- Rats fed different diets, exposed to carcinogen, then euthanized.
- DNA adducts, DNA repair, apoptosis quantified as staining intensities
- Data has hierarchical structure
  - Treatment groups - rats - crypts - cells/pixels
  - Hierarchical Longitudinal Data
Hierarchical Functional Model

2-level HF model:

\[ Y_{abc} = g_{abc}(t) + e_{abc}, \]
\[ g_{abc}(t) = g_{ab}(t) + \eta_{abc}(t) \]
\[ g_{ab}(t) = g_{a}(t) + \xi_{ab}(t) \]

where \( e_{abc} \sim MVN(0, \sigma_e^2 I) \),

\( \eta_{abc}(\bullet) \) and \( \xi_{ab}(\bullet) \): mean 0 with covariance matrices \( \Sigma_1(t_1, t_2) \) and \( \Sigma_2(t_1, t_2) \).
DNA Repair Data

DNA Repair Enzyme for Selected Crypts

Fish, T 3, R 1, C 6

Fish, T 9, R 3, C 8

Corn, T 9, R 3, C 19

Fish, T 3, R 1, C 23

Fish, T 9, R 3, C 19

Corn, T 9, R 3, C 20
Wavelet Space Model

Wavelets: families of orthonormal basis functions

\[ d_{abc} = \{d_{abc}^{j,k}\} = W \ y_{abc} \]

Discrete Wavelet Transform

\[
\begin{align*}
    d_{abc}^{j,k} &\sim N(\theta_{abc}^{j,k}, \sigma_e^2) \\
    \theta_{abc}^{j,k} &\sim N(\theta_{ab}^{j,k}, \sigma_{1,j}^2) \\
    \theta_{ab}^{j,k} &\sim N(\theta_{a}^{j,k}, \sigma_{2,j}^2)
\end{align*}
\]
“Shrinkage” Prior

Prior on $\theta_a^{j,k} \Rightarrow$ Normal-0 mixture

$$\theta_a^{j,k} \sim N(0, \gamma_a^{j,k} \tau_j^2)$$

$$\gamma_a^{j,k} \sim \text{Bernoulli}(p_j)$$

Nonlinear shrinkage -- denoises data

$p_j$ and $\tau_j^2$ - regularization parameters (scale dependent shrinkage)

Smaller $p_j$ : More smoothing at scale $2^{j-1}$
Model Fitting

- Hierarchical model fit using MCMC
  - Marginalization done to get stationary distribution of treatment-level coefficients.
  - IDWT done on set of coefficient estimates to obtain posterior distributions of $g_a(t)$.
- Relative variability at hierarchical levels, measured by trace/determinant, given directly by wavelet space results.
- Possible to also obtain individual and subsample-level curve predictors.
## Results: DNA Repair

Estimates & 90% posterior bounds by diet/time

<table>
<thead>
<tr>
<th>Diet</th>
<th>0 h</th>
<th>3 h</th>
<th>6 h</th>
<th>9 h</th>
<th>12 h</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fish Oil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corn Oil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary

- Method to fit hierarchical longitudinal data
- Nonparametrically estimate mean profiles for:
  - 1. Treatments
  - 2. Individuals
  - 3. Subsampling units
- Do Bayesian inference
- Handles spatially hetero. data, unbalanced designs
- Estimate of relative variability at hier. levels
- Paper under review -- tech. report available.
Estimating Correlations

- Assess interrelationships between 2 responses as function of cell position
  - e.g. DNA adducts and DNA repair

- Modeled as correlation: \( \rho(t_1, t_2) \)
Model

Response 1, e.g. DNA Adducts

\[ Y_{abc} = g_{abc}(t) + e_{abc}, \]
\[ g_{abc}(t) = g_{ab}(t) + \eta_{abc}(t), \]
\[ g_{ab}(t) = g_a(t) + \xi_{ab}(t). \]

Response 2, e.g. DNA Repair

\[ Z_{abc} = f_{abc}(t) + u_{abc}, \]
\[ f_{abc}(t) = f_{ab}(t) + \theta_{abc}(t), \]
\[ f_{ab}(t) = f_a(t) + \psi_{ab}(t). \]

\[ \rho(t_1, t_2) = \text{Corr}\{\xi_{ab}(t_1), \psi_{ab}(t_2)\} \]
Nonparametric Estimation

- Nonparametric estimation method:
  - Estimate crypt-level functions $g_{abc}(t), f_{abc}(t)$ on grid
  - Fit mixed model at each grid point
  - Correlation estimated from variance components

- Undersmooth at crypt level

- Virtually no loss of efficiency from nonparametric estimation for our data
  - Asymptotics: Same dominating term as estimator with $g_{abc}(t), f_{abc}(t)$ known
Parametric vs. Nonparametric Methods

- Simulation study done to compare parametric and nonparametric methods
- Study confirmed efficiency of nonparametric method relative to parametric method under certain conditions
Summary

- Method to assess linear relationships between responses that depend on longitudinal variable
- Example where little cost in doing nonparametric estimation
- For details -- see paper in 11/2001 *JASA*
Modeling Crypt Cell Architecture

When measurements taken ‘by cell’,

\[ t = \frac{\text{distance from each cell’s midpoint to crypt base}}{\text{length of crypt}} \]

- Relative cell positions not observed, since tedious to measure
- **SURROGATE OBSERVED:**

\[ u = \frac{\# \text{ of cells below each cell}}{\# \text{ of cells in crypt}} \]

Results in missing longitudinal variable
Modeling Crypt Cell Architecture

- Model for relative cell positions $t$
  - $t \sim$ o.s. of Beta, or mixture of Betas
- Given validation data set $(t,u)$:
  - Can efficiently generate missing $t$ using EM or within MCMC step
  - Characterize crypt architecture in colonic regions

*Biostatistics* paper -- to appear
Conclusion

- Cell-based colon carcinogenesis studies
  - Hierarchical Longitudinal data
  - Rich in information -- challenging to extract
- New methods developed
  - Wavelet regression methods for long. data
  - Kernel methods for longitudinal correlations
  - Method for missing data with order constraints

Email: jeffmo@mdanderson.org