Bayesian Group Lasso for Analyzing Contingency Tables

Sudhir Raman, Thomas J. Fuchs, Peter J. Wild, Edgar Dahl, Volker Roth
Talk Outline

- Motivation
- Introduction
- Feature Selection
- Contingency Tables
- Application - Breast Cancer
- Results
- Summary
Motivation

- Applications involving
  - Categorical variables - leading to sparsity in groups.
  - Count data – Frequently encountered in medical applications.
- Meaningful Variance estimates
Consider a standard linear regression problem

\[ y_i = x_i^t \beta + \varepsilon_i \quad s.t. \quad g(\beta) \leq \kappa. \]

A Bayesian version:

\[ p(\beta, \theta | y, X) \propto p(y | X, \beta, \theta) p(\beta | \theta) p(\theta) \]

Posterior  Likelihood  Prior
Feature Selection
Feature Selection

Lasso

\[ \|y - X \beta\|_2^2 \quad \text{subject to} \quad \|\beta\|_1 \leq \kappa. \]

Prior (Park & Casella, 2008):

\[ p(\beta|0, k^{-1}) = \prod_{i=1}^{D} \frac{k}{2} \exp(-k|\beta_i|) \]

Group-Lasso

minimize \( l(\beta) \) \( \text{s.t.} \) \( \sum_{g=1}^{G} \|\beta_g\|_2 \leq \kappa. \)

Prior:

\[ \prod_{g=1}^{G} \text{M-Laplace}(\beta_g|0, c^{-1}) \propto c^{pg/2} \exp(-c\|\beta_g\|_2), \]
Hierarchical model

\[
p(\beta_g | \rho) = \int_0^\infty N(\beta_g | 0, \sigma^2 \lambda_g^2 I) p(\lambda_g^2 | \rho) d\lambda_g^2
\]

Normal-Gamma model:

\[
p(\beta_g | \rho) = \int_0^\infty N(\beta_g | 0, \sigma^2 \lambda_g^2) \text{Gamma}(\lambda_g^2 | \frac{p_g + 1}{2}, \frac{2}{a_g}) d\lambda_g^2
\]

\[\propto \text{M-Laplace}(\beta_g | 0, (a_g/\sigma^2)^{-\frac{1}{2}})\].

where \(a_g = p_g \rho\) and \(b_g = ||\beta_g||^2/\sigma^2\) (for each group \(g\)) , using the generalized inverse gaussian distribution.
Generalized linear models

- **Stochastic component:**
  - $Z$ distributed based on mean $\theta$.
  
  \[ Z \sim f(\theta) \]

- **Random Effect:**
  \[ \eta_i = x_i^t \beta + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2) \]

- **Link Function:**
  \[ g(\theta) = \eta \]
Contingency tables for count data
Contingency Tables

- Contingency tables
  - Count data
  - Categorical variables

- Poisson for Modeling:
  - Random counts, fixed time period.
    
    \[ z_i | \theta_i \sim \text{Poisson}(\theta_i) = \frac{\theta_i^{z_i} e^{-\theta_i}}{z_i!}, \]

    with the Poisson mean \( \theta_i = e^{\eta_i} \), and \( \eta_i \sim N(x_i^T \beta, \sigma^2) \).

- Example – a clinical study of fixed time period.
What is X?

Dummy Coding - Example

\[ X = \begin{bmatrix}
1 & 1 & 0 & 0 & 0 \\
2 & 0 & 1 & 0 & 0 \\
3 & 0 & 0 & 1 & 0 \\
4 & 0 & 0 & 0 & 1 \\
\end{bmatrix} \]

Too many columns
Construction of $X$

- **Higher order interaction terms.**

\[ X = [X^{C_1}, \ldots, X^{C_d}, X^{C_1:C_2}, \ldots, X^{C_{d-1}:C_d}, \ldots, X^{C_1:\cdots:C_Q+1}, \ldots, X^{C_{d-Q}:\cdots:C_d}] \]

- **Polynomial Contrast Codes**
  - Used for ordered variables.
  - Avoids over-parameterization.
  - Orthogonal in nature – results in an orthogonal design matrix
    - \((X^T X = I)\).
Inference:

- **Gibbs Sampling:**
  - Highly efficient (no matrix inversion).

- **Standard posterior conditionals.**

- **Sampling the $\eta$ variable:**
  - Adaptive Rejection sampling.
  - Laplace approximation:

\[
p(\eta_i | \beta, \sigma^2, X, z) \propto \exp \left[ \sum_i \eta_i z_i \exp(\eta_i) - \frac{1}{2\sigma^2} (x_i^t \beta - \eta_i)^2 \right].
\]
Experiments
Computational Pathology
Application – Breast Cancer

- Breast Cancer – Leading cause for tumor-related death of women, particularly in Western countries.
  - Finding biomarkers for prediction and prognosis is important.

- Immunohistochemistry:
  - Labeling proteins in tissue sections using antibody-antigen interaction.
  - Cost effective, used on a routine basis.

- Tissue Microarray Technology:
  - TMA: allows simultaneous in situ analysis of 1000 primary tumors.
  - Promises to significantly accelerate studies seeking for biomarkers.
Primary samples are taken from cancerous breast tissue. Tissue cylinders of size 0.6mm are arrayed in a paraffin block. Slices of 0.6µm are cut off and are stained. TMA spot from a single patient with breast cancer stained with the YB-1 antigen.
Results – Significant Interactions

- Two patient groups – Find difference:
  - Low/High risk groups.
- Intensity levels for 7 proteins, with higher-order interactions up to order 2.

Low Risk Group

High Risk Group
Summary

- Bayesian Group-Lasso to deal with
  - Count data – Poisson Model
  - Feature selection with categorical variables

- Detection of novel *compound* bio-markers in the breast cancer dataset.

- Advantages:
  - Average over solutions, meaningful variance estimates.
  - Higher-order interactions.

- Future work:
  - Extending to other types of data (Weibull, Beta, Dirichlet).
  - Going beyond Group-Lasso – Applying more sparse constraints, based on [Caron & Doucet, 2008]).
  - Clustering.
Thank you for your attention.