Bayesian Hierarchical Methods for Medical Fraud Assessment

Tahir Ekin

McCoy College of Business
Texas State University
San Marcos Texas

McCoy College of Business
Woods Award Seminar
San Marcos, TX

April 2018
Research Portfolio Overview

Applied (Bayesian) statistics and operations research

Research portfolio consists of two core areas:

1. Simulation based stochastic optimization
2. Medical fraud analytics

Interdisciplinary collaboration:


Simulation based stochastic optimization


Medical fraud analytics


Medical fraud analytics: Work in progress


Why healthcare? Why now?

- U.S. health care spending: $3.2 trillion, or $9,990 per person in 2015
- Medicare: Mainly serves older than 65 and disabled, 55 million users, more than 40% of total inpatient hospital costs, federally funded
- Medicaid: Mainly serves low income people; 65 million users, funded by the state and federal governments, controlled by states
- Medical fraud: Intentional deception or misrepresentation made by a person or an entity, with the knowledge that the deception could result in some kinds of unauthorized benefits, NHCAA (2012)
- Three to ten percent lost to fraud, waste and abuse
  - Complexity, heterogeneity of the system, unconditional trust on providers, lack of resources for investigations
  - Low probability of detection, relatively low probability of being convicted once detected and the low severity of punishment even if convicted
Figure: viPS 2010 Industry Survey Results
Types of Healthcare fraud

- Identity fraud
- Providing incorrect care or manipulating billing rules
- Managed care fraud
- **Improper coding**: upcoding, unbundling, multiple (double) billing, phantom (ghost) billing
- Providing unnecessary care
- Kickback schemes and self-referrals
Improper coding and providing unnecessary care

- Home health agencies: Dr. Jacques Roy from Rockwall, Texas billed more home health services through Medicare than any other medical practice in the U.S between 2006 and 2011 for more than $350 million.

- Florida, Dr. Salomon Melgen: Aggressive Medicine or Malpractice? Incorrect diagnoses and falsified tests, he made $8.4 million (vs $6,061) during the six years by treating people with lasers and $57.3 million (vs $3 million) by treating patients with Lucentis injection.

- Texas, 2005- : 160 ongoing Medicaid dental fraud investigations, Texas spending more on braces than other 49 states combined, Nevaeh Hall, a 4-year-old, got overtreated and had complications (brain damage) during a routine appointment at a local dental clinic.
Kickback schemes and self-referrals

- New Orleans, 2015: $50 million 6 year scheme by two doctors and a registered nurse where they referred patients to four sham companies for medically unnecessary home health services and treatment.

- Florida, Arizona, Texas: power wheelchairs
  - Cooper Medical Supply: fraudulent prescriptions and medical documents to submit false claims to Medicare for expensive, high-end power wheelchairs. More than 80 percent of the beneficiaries lived over 100 miles away from Cooper Medical Supply, and most were not even given the wheelchairs.
  - Positive Home Oxygen and Dr. Robert Lyle Cleveland: In exchange of Dr. Cleveland signing Certificates of Medical Necessity for power wheelchairs for patients who did not meet the coverage requirements, patients would be referred to him.
Fraud, waste and abuse (over-payments) result with:
- Direct cost implications to the government and to the tax-payers
- Diminished ability of the medical systems to provide quality care to the deserving patients

Audits and investigations by CMS and Office of Inspector General

Statistical issues in fraud assessment:
- Sampling and over-payment estimation
- Use of data mining for fraud detection
Figure: Overview of the Fraud Prevention System
Unsupervised Data Mining Motivation

- Medical claims data
- Dynamic nature of fraud and expensive labeling: Li et al. (2008)
- Hierarchical/layered nature of medical claims data

- Unsupervised methods mainly used for initial screening
  - Anomaly detection: Ng et al. (2010), Capelleveen (2013), Iyengar et al. (2013)
  - Clustering based on geographic information: Musal (2010)
  - Lorenz curve: Ekin et al. (2017)
  - Bayesian coclustering: Ekin et al. (2013)

- Identify the hidden patterns among providers (doctors), procedures and patients
- Reveal billing behaviors and find providers that are outliers
Latent Dirichlet Allocation (LDA): Motivation

- Blei et al. (2003); Blei and Lafferty (2009)
- Search/links vs zoom in/out
- Analysis of the words of the original texts to discover the themes, their relationship with each other and evolution of topics over time
- Probabilistic model: treatment of data as arising from a generative process that includes hidden variables
- Compute the conditional (posterior) distribution of hidden variables given the observed variables
- Organization and summary of archives
- Advanced descriptive statistics
LDA for Medical Fraud Detection

- **Topics**: Collection of Medical Procedures
- **Documents**: Doctors (or a set of physicians, hospital)
- **Words**: Procedures

\[
p(\beta, \theta, z, w) = \left( \prod_{i=1}^{K} \rho(\beta_i | \eta) \right) \left( \prod_{d=1}^{D} \rho(\theta_d | \alpha) \prod_{n=1}^{N} \rho(z_{d,n} | \theta_d) \rho(w_{d,n} | \beta_{1:K}, z_{d,n}) \right)
\]
LDA: Generative process

- Observed words, $W_{d,n}$ and a given number of topics, $K$
- Aim: reveal the hidden topic structure

Choose distributions for each topic: ($\beta_k \sim \text{Dir}(\eta); k=1,...,K$)

For each document $d$:
  - Draw topic proportions ($\theta_d \sim \text{Dir}(\alpha)$)

For each ($n^{th}$) word in the ($d^{th}$) document:
  - Draw a topic, $k$ with respect to the proportions; $Z_{d,n} \sim \text{Mult}(\theta_d)$
  - For that particular ($k^{th}$) topic, you already have the topic distribution, $\beta_k \sim \text{Dir}(\eta)$
  - Draw a word with respect to the chosen topic's distribution; $W_{d,n} \sim \text{Mult}(\beta_{Z_{d,n}})$
Collapsed Gibbs LDA

\[ p(Z|W) \propto \int_{\theta} \int_{\beta} \left( \prod_{d=1}^{D} \prod_{n=1}^{N} p(z_{d,n}|\theta_d) p(W_{d,n}|\beta_{1:K,z_{d,n}}) \right) d\beta d\theta \] (1)

\[ p(z_{d,n} = k|Z_{(-d,n)}, W) = \frac{\left( q_k(-d,n) + \alpha_0 \right) \sum_{n=1}^{N} (q_k(-d,n) + \beta_0)}{\sum_{t=1}^{K} \left( q_t(-d,n) + \alpha_0 \right) \sum_{n=1}^{N} (q_t(-d,n) + \beta_0)} \] (2)

\[
E[\theta_d,k] = \frac{q_k(d,.) + \alpha_0}{\sum_{t=1}^{K} [q_t(.,.) + \alpha_0]}
\]

\[
E[\beta_{k,n}] = \frac{q_k(.,n) + \eta_0}{\sum_{t=1}^{N} [q_t(.,n) + \eta_0]}
\]
Steps of Collapsed Gibbs LDA

- The $Z$ variables are initialized to determine the initial state of the Markov chain.
- The chain is then run for a number of iterations, each time finding a new state by sampling each $z_{d,n}$ from the specified distribution in equation (2).
- After a fixed number of iterations, convergence of the Markov chain is checked.
- When convergence is attained, the respective counts of $Z$ are recorded and the posterior mean values of $\theta$ and $\beta$ are computed.
Which procedures are frequently billed together? : $\beta_k$
Can we detect doctors that have unusual behavior? : $\theta_d$
Which doctors are similar? : Similarity assessment via Hellinger distance (Hellinger, 1909)

Data: CMS Medicare 2012 Part B claims; $\geq 9$ M unique providers and 27 attributes for each
Focus is on Vermont: 1,493,224 seperately billed claims, 72 provider types
1,055 procedure codes that are billed by 2,268 unique doctors
10%: test data

Model choice: number of topics

\[
\text{perplexity}(D_{\text{test}}) = \exp\left\{ - \frac{\sum_{d=1}^{D} \log(p(W_{d,n}))}{ND} \right\}
\]

For \( K \) values of 5, 10, 15, 20, the perplexity values are 5.57, 5.05, 4.82, 4.62

MCMC implemented via a C++ program

Python and R (tm and wordcloud packages) for data pre-processing and analysis
“g9008” : “Coordinated Care Fee, Physician Coordinated Care Oversight Services”
“a0425” : “Ground Mileage, Per Statute Mile”.
## LDA Results: Analysis of Topics

<table>
<thead>
<tr>
<th>HCPCs Code</th>
<th>Description</th>
<th>$\beta_4$,</th>
</tr>
</thead>
<tbody>
<tr>
<td>92014</td>
<td>Eye exam and treatment</td>
<td>0.351</td>
</tr>
<tr>
<td>92012</td>
<td>Eye exam established patient</td>
<td>0.136</td>
</tr>
<tr>
<td>92083</td>
<td>Visual field examination(s)</td>
<td>0.075</td>
</tr>
<tr>
<td>66984</td>
<td>Cataract surgery w/iol 1 stage</td>
<td>0.061</td>
</tr>
<tr>
<td>92133</td>
<td>Computerized ophthalmic imaging optic nerve</td>
<td>0.060</td>
</tr>
</tbody>
</table>

**Table:** List of procedures in Topic 4, a portion of $\beta_4$ with terms sorted descendingly

...
LDA Results: Analysis of Topics

Table: List of procedures in Topic 4, a portion of $\beta_4$ with terms sorted descendingly

<table>
<thead>
<tr>
<th>HCPCs Code</th>
<th>Description</th>
<th>$\beta_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>92014</td>
<td>Eye exam and treatment</td>
<td>0.351</td>
</tr>
<tr>
<td>92012</td>
<td>Eye exam established patient</td>
<td>0.136</td>
</tr>
<tr>
<td>92083</td>
<td>Visual field examination(s)</td>
<td>0.075</td>
</tr>
<tr>
<td>66984</td>
<td>Cataract surgery w/iol 1 stage</td>
<td>0.061</td>
</tr>
<tr>
<td>92133</td>
<td>Computerized ophthalmic imaging optic nerve</td>
<td>0.060</td>
</tr>
</tbody>
</table>

- Procedures related to “eye exam and treatment”
- Which specialty of doctors bill for this topic?
Table: List of procedures in Topic 4, a portion of $\beta_4$ with terms sorted descendingly

- Procedures related to “eye exam and treatment”
- Which specialty of doctors bill for this topic?
- Ophthalmologists: eye exams, diagnose and treat disease, prescribe medications and perform eye surgery
- Optometrists: certain eye problems and diseases, and may participate in your pre- and post-operative care
LDA Results: Analysis of Topics

- 105 providers of which 100 are “eye” related providers
- 32 out of the 36 ophthalmologists and 68 out of the 78 optometrists.
- Remaining five: a physician assistant, an interventional radiologist, an ambulatory surgical center and diagnostic radiologists: Potential red flags

- What about the rest of the optometrists and ophthalmologists who have not billed for Topic 4 the most?
- 4 ophthalmologists billed for Topic 1 the most.
- 2, 1, 7 optometrists have billed the most for topics 1, 8 and 10
Table: List of procedures in Topic 1, a portion of $\beta_1$ with terms sorted descendingly

<table>
<thead>
<tr>
<th>HCPCs Code</th>
<th>Description</th>
<th>$\beta_1, \beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>J3300</td>
<td>Triamcinolone A inj PRS-free</td>
<td>0.443</td>
</tr>
<tr>
<td>92226</td>
<td>Special eye exam subsequent</td>
<td>0.127</td>
</tr>
<tr>
<td>92134</td>
<td>Cptr ophth dx img post segmt</td>
<td>0.108</td>
</tr>
<tr>
<td>90805</td>
<td>Psytx off 20-30 min with E &amp; M</td>
<td>0.103</td>
</tr>
<tr>
<td>67028</td>
<td>Injection eye drug</td>
<td>0.067</td>
</tr>
</tbody>
</table>

- Procedures 92226 and 67028 within Topic 1 are related to eye exams and injections.
- Topic 10 includes code “142” that corresponds to “lens surgery”: 7 optometrists
LDA Results: Analysis of Topics

<table>
<thead>
<tr>
<th>HCPCs Code</th>
<th>Description</th>
<th>$\beta_8,$</th>
</tr>
</thead>
<tbody>
<tr>
<td>99204</td>
<td>Office/outpatient visit new</td>
<td>0.304</td>
</tr>
<tr>
<td>92557</td>
<td>Comprehensive hearing test</td>
<td>0.070</td>
</tr>
<tr>
<td>99203</td>
<td>Office/outpatient visit new</td>
<td>0.067</td>
</tr>
<tr>
<td>45380</td>
<td>Colonoscopy and biopsy</td>
<td>0.060</td>
</tr>
<tr>
<td>43239</td>
<td>Upper gi endoscopy biopsy</td>
<td>0.052</td>
</tr>
</tbody>
</table>

Table: List of procedures in Topic 8, a portion of $\beta_8$ with terms sorted descendingly

- How to explain the 1 optometrist that does not behave like peers?
LDA Results: Analysis of Topics

- Can reveal characteristics that are different than the overall pattern

**Figure**: Select topic proportions, $\theta_d$ for the peer group (left) and maximum topic proportions $\theta_d$ (right)
LDA Results: Similarity Assessment

- **Hellinger Distance between the topic proportions of two doctors**

\[ d_{ij} = \frac{1}{\sqrt{2}} \sqrt{\sum_{k=1}^{K} \left( \sqrt{\theta_{i,k}} - \sqrt{\theta_{j,k}} \right)^2} \]

**Figure:** Hellinger Distance between the outlier and peers

- Average: 0.899, maximum: 0.973 and minimum: 0.510.
- Reveals further insights: outliers, teamwork
LDA: Limitations and Extensions

- How to sample from the posterior distribution:
  - Sampling based algorithms (MCMC): MC, dependent sequence of RV, with a limiting distribution as posterior: Steyvers and Griffiths, 2006
  - Variational algorithms which find the distribution that closely mimics the posterior distribution via optimization: Blei et al. (2003), Hoffman et al. (2010)

- Model evaluation: Perplexity, lack of labeled data
- One type of cluster, does not address dyadic data: Co-clustering
- Correlation among topics
- Co-variates: topic proportions by provider type
Prescription Fraud and Opioid Abuse

“The level of urgency is greater than ever to develop creative solutions based on exploiting modern data mining and communication proficiencies.”

President’s Commission on Combating Drug Addiction and the Opioid Crisis (2017)

**Figure:** Growth in Part D spending of commonly abused opioids

**NEARLY 1 IN 3** beneficiaries received a commonly abused opioid

These opioids accounted for **OVER $4 BILLION** in Part D spending

The commonly abused opioids with the **HIGHEST PART D SPENDING** were:

- OxyContin
- Hydrocodone-Acetaminophin
- Oxycodone-Acetaminophin
- Fentanyl
In 2015, controlled substances accounted for $8.4 billion, 6 percent of all Part D spending.

Opioids include narcotics intended to manage pain from surgery, injury, or illness.

More than half a million Part D beneficiaries were found to receive high amounts of opioids in 2016 (OIG, 2017)

Opioids can create euphoric effects which make them vulnerable to abuse and misuse despite the common adverse side effects such as respiratory depression, confusion, tolerance, and physical dependence.

91 Americans die every day from an opioid overdose corresponding to more than 60 percent of total overdose deaths in the U.S. (CDC, 2017)
Figure: Cumulative changes in spending for opioids vs other drugs
Bayesian Hierarchical Methods with covariates

Structural topic models: Roberts et al. (2016)

Figure: Structural Topic Model Graphical Plate
Bayesian Hierarchical Methods with covariates

Documents: $D$ providers
Words: $N_d$ drugs, $V$ unique drugs (vocabulary)
Topics: $K$ collections of drug prescriptions

Logistic Normal model for topic prevalence, $\theta$ and log transformed deviations from the baseline word rates $m$ based on differences in topic, covariates ($Y$) and topic-covariate interactions for topic content, $\beta$
Finding drug associations
Finding drug prescription patterns across medical specialties
Finding the providers with different drug prescription distributions versus their specialty peers

Data: CMS Medicare 2015 Part D prescriptions; 24.5 million observations for more than 850,000 medical providers
Focus is on New Hampshire: among highest opioid use/overdose pp
5701 medical providers with 88 different specialties billing for more than 800 K prescriptions
**Application**

---

**Figure:** Topics’ Expected Proportion and the Top Two Drugs (left), Topic Quality (right)
Figure: Topic Correlations
Figure: Expected Proportion of Main Drug Categories Across Topics
Application: Distribution of Topic 15 Across Specialties

Figure: Distribution of Topic 15 Across Specialties
Figure: Topic Proportions for "Interventional Pain Management" Specialists (red dots are averages)
Figure 8: Lorenz Curves (top) and Topic Distribution of the Outlier Provider (bottom)
Incorporation into decision making framework: semi-automated system

Poisson co-clustering: Counts of billings for dyadic pair of providers and procedures

Use of co-variates such as provider type, counts, monetary amount

Model stability: Issues with convergence and inference

Network analysis
Main Takeaways

- Approximately a minimum of $297 per person/year medical overpayment
- Complexity and size help fraudsters to find a way to game the system
- There is a need for adaptive statistical methods
- Great potential for unsupervised Bayesian data mining approaches
- Real time fraud assessment needs to be more frequent in the near future

- Need for an integrated team of health professionals and data analysts
- Unsupervised model evaluation with labeled data
- Integrated (semi-automated) decision making
Thank you

Contact: t.e18@txstate.edu; www.tahirekin.com
Medical fraud vs other types of fraud

Figure: Medical fraud vs other types of fraud, FICO 2009
LDA: Generative process

- Assume \((\mu, \Sigma)\) be a K-dimensional mean and covariance matrix
- Choose distributions for each topic: \((\beta_k \sim Dir(\eta); k=1,\ldots,K)\)
- For each document \(d\):
  - Draw \((\eta_d|\mu, \Sigma) \sim N(\mu, \Sigma)\)
- For each \((n^{th})\) word in the \((d^{th})\) document:
  - Draw a topic, \(k\) with respect to the proportions; \(Z_{d,n} \sim Mult(f(\eta))\) where
    \[
    f(\eta_{d,k}) = \frac{\exp(\eta_{d,k})}{\sum_j \exp(\eta_{d,j})}
    \]
  - For that \((k^{th})\) topic, you already have the topic distribution, \(\beta_k \sim Dir(\eta)\)
  - Draw a word with respect to the chosen topic's distribution; \(W_{d,n} \sim Mult(\beta_{Z_{d,n}})\)
LDA Collapsed Gibbs notation

- $Z_{(-d,n)}$: vector of all other topic assignments other than the one for $n^{th}$ medical procedure for the $d^{th}$ doctor
- $q_k(d,n)$: count of assignments to $k^{th}$ topic for the $n^{th}$ medical procedure in the $d^{th}$ document
- $q_k(-d,n)$: count of all assignments to $k^{th}$ topic excluding $q_k(d,n)$
- $q_k(d,\cdot)$: total count of assignments to the $k^{th}$ topic for all $N$ procedures and the $d^{th}$ doctor
- $q_k(\cdot,n)$: total count of assignments to the $k^{th}$ topic for all $D$ doctors and the $n^{th}$ procedure
LDA Results: Analysis of Topics

Figure: List of procedures in Topic 2, a portion of $\beta_2$ with terms sorted descendingly

<table>
<thead>
<tr>
<th>HCPCS Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>93010</td>
<td>Electrocardiogram report</td>
</tr>
<tr>
<td>93306</td>
<td>w/Doppler complete</td>
</tr>
<tr>
<td>93000</td>
<td>Electrocardiogram complete</td>
</tr>
<tr>
<td>99213</td>
<td>Office/outpatient visit</td>
</tr>
<tr>
<td>93018</td>
<td>Cardiovascular stress test</td>
</tr>
<tr>
<td>99214</td>
<td>Office/outpatient visit</td>
</tr>
<tr>
<td>93288</td>
<td>Pm device eval. in person</td>
</tr>
<tr>
<td>93016</td>
<td>Cardiovascular stress test</td>
</tr>
<tr>
<td>99223</td>
<td>Initial hospital care</td>
</tr>
<tr>
<td>93227</td>
<td>Ecg monit/reprt up to 48 hrs</td>
</tr>
</tbody>
</table>

First 5 terms have a total weight of 71%

Expected focus on cardio related codes but of a basic nature
LDA Results: Topic 2 Analysis

- Group of procedures that have similarity
- Set of unique doctors that bill for this particular topic

The doctors who billed for this topic the most:

- 12 Diagnostic Radiologists (DR): 12 out of possible 41
- 57 Family Practitioners (FP): 57 out of possible 257
- 42 Internal Medicine Physicians (IM): 22 out of possible 139
- 10 Anesthesiologists (Anesth.): 10 out of possible 44
- 4 Cardiologists (Card.): 54 out of possible 120
LDA Results: Topic 3 Analysis

<table>
<thead>
<tr>
<th>HCPCS Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>J0588</td>
<td>IncobotulinumtoxinA</td>
</tr>
<tr>
<td>J0586</td>
<td>AbobotulinumtoxinA</td>
</tr>
<tr>
<td>77421</td>
<td>Stereoscopic x-ray guidance</td>
</tr>
<tr>
<td>77300</td>
<td>Radiation therapy dose plan</td>
</tr>
<tr>
<td>77427</td>
<td>Radiation tx management x5</td>
</tr>
<tr>
<td>45380</td>
<td>Colonoscopy and biopsy</td>
</tr>
<tr>
<td>45385</td>
<td>Lesion removal colonoscopy</td>
</tr>
<tr>
<td>43239</td>
<td>Upper gi endoscopy biopsy</td>
</tr>
<tr>
<td>77334</td>
<td>Radiation treatment aid(s)</td>
</tr>
<tr>
<td>J2930</td>
<td>Methylprednisolone injection</td>
</tr>
</tbody>
</table>

**Figure:** List of procedures in Topic 3, a portion of $\beta_3$ with terms sorted descendingly

- The first two codes account for 80% of the weight, used as pain relievers.
- This topic appears to be focused on DR and IM related codes.
- The doctors who billed for this topic the most: 19 DR, 10 FP, 9 IM, 3 Anesth., 1 Card.
- Cardiologist raises flags.
Hellinger Distance between the topic proportions of two doctors $i$ and $j$:

$$d_{ij} = \frac{1}{\sqrt{2}} \sqrt{\sum_{k=1}^{K} [(\sqrt{\theta_{i,k}} - \sqrt{\theta_{j,k}})^2]}$$

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>$\theta_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr.Trost</td>
<td>IM</td>
<td>0.5300 0.0015 0.0015 0.3290 0.0015 0.0015 0.0456 0.0456 0.0456 0.0015</td>
</tr>
<tr>
<td>Dr.Grant</td>
<td>DR</td>
<td>0.9712 0.0019 0.0015 0.0019 0.0019 0.0019 0.0019 0.0019 0.0019 0.0140 0.0019</td>
</tr>
<tr>
<td>Dr.Christensen</td>
<td>Card.</td>
<td>0.0029 0.0029 0.9121 0.0029 0.0029 0.0029 0.0029 0.0029 0.0029 0.0450 0.0225 0.0029</td>
</tr>
<tr>
<td>Dr.Spencer</td>
<td>DR</td>
<td>0.8982 0.0719 0.0025 0.0049 0.0049 0.0025 0.0025 0.0049 0.0225 0.0025 0.0049</td>
</tr>
<tr>
<td>Dr.Schwartz</td>
<td>DR</td>
<td>0.0110 0.0026 0.7621 0.0026 0.0026 0.0026 0.0026 0.0026 0.1919 0.0110 0.0110</td>
</tr>
</tbody>
</table>

Table: Topic proportion vectors for five given doctors

- $d_{2,4} = 0.178; d_{2,3} = 0.931; d_{3,5} = 0.183; d_{2,5} = 0.903; d_{1,2} = 0.452$
- Reveals further insights: outliers, teamwork
Correlated topic models

- Draw topic proportions from a logistic normal, where topic occurrences can exhibit correlation.
- Better prediction vs increased computational cost
Figure: Diagnostic Values by Number of Topics and initialization
Bayesian co-clustering

- U.S. Medicare Part B (outpatient) claims data
  - Anesthesiologists in Texas that provide services in a facility
  - Providers that have billed for at least 10 unique procedures and the procedures that are billed by at least 20 unique providers
  - Binary billing matrix that lists whether each of $J=94$ procedures are billed by $I=376$ providers or not

- Objective is to reveal common billing patterns
- Bayesian co-clustering allows mixed membership for clusters of providers and of procedures; so called soft-clustering
Bayesian co-clustering

- Wang et al. (2010), Ekin et al. (2013)
- Procedures-Providers and Providers-Patients (Beneficiaries)
Bayesian co-clustering: Model

\[ X_{ij} = 1 \text{ if provider } i \text{ bills for procedure } j \]

\[ X_{ij} | Z_{1i} = k, Z_{2j} = l, \beta_{kl} \sim Ber(\beta_{kl}) \]

Priors:

\[ \theta_1 \sim Dir(\alpha_{1k}; k = 1, \ldots, K), \theta_2 \sim Dir(\alpha_{2l}; l = 1, \ldots, L) \]

\[ \beta_{kl} \sim Beta(a_{kl}, b_{kl}), k = 1, \ldots, K, l = 1, \ldots, L. \]

Posterioris:

\[ \beta_{kl} | Z_1, Z_2, X \sim Beta\left( a_{kl} + \sum_{i,j} X_{ij} I(Z_{1i} = k, Z_{2j} = l), b_{kl} + \sum_{i,j} (1 - X_{ij}) I(Z_{1i} = k, Z_{2j} = l) \right) \]
Bayesian co-clustering: Model

\[
\theta_1 | Z_1 \sim \text{Dir} \left( \alpha_{1k} + \sum_{i,j} I(Z_{1i} = k); k = 1, \ldots, K \right),
\]

\[
\theta_2 | Z_2 \sim \text{Dir} \left( \alpha_{2l} + \sum_{i,j} I(Z_{2j} = l); l = 1, \ldots, L \right).
\]

The full conditionals of \((Z_{1i}, Z_{2j})\) can be obtained as

\[
p(Z_{1i} = k, Z_{2j} = l | \theta_1, \theta_2, \beta, X_{ij}) = \frac{\beta_{kl} X_{ij} (1 - \beta_{kl})^{1-X_{ij}} \theta_{1k} \theta_{2l}}{\sum_{r=1}^{K} \sum_{c=1}^{L} \beta_{rc} X_{ij} (1 - \beta_{rc})^{1-X_{ij}} \theta_{1r} \theta_{2c}}.
\]
Bayesian co-clustering Results

\[ \beta = \begin{bmatrix}
0.145 & 0.143 \\
0.140 & 0.142 \\
0.149 & 0.146
\end{bmatrix} \]

- The most frequent occurrences between provider-procedure pairs are in co-cluster (3, 1).
- The provider-procedure pairs that are in co-cluster (3, 1) are 6.4% more likely to bill compared to the pairs in co-cluster (2, 1).
Bayesian co-clustering Results

- Billing of the Procedure 64941 that corresponds to facet joint injection by Provider with ID 100.
- The posterior modes are $Z_1 = 3$ and $Z_2 = 1$ in line with $\beta_{3,1}$

Figure: Posterior distributions of the memberships for Provider 100 and Procedure 64941
Bayesian co-clustering: Potential insights

- Can we detect providers and procedure pairs that have unusual behavior?
  - $\beta_{kl}$ provides insights for associations
  - The higher the $\beta_{kl}$, the more likely is the probability that a member of provider cluster $k$ bills for the members of procedure cluster $l$.
  - Discrepancies between the expected behavior and the actual behavior of a given provider can provide investigative leads.
  - For a given billing; if the provider does not behave similar to his co-cluster; this may reveal a potential fraudulent behavior.

- Identification of associations among providers and patients
  - Potential flags for unusual memberships in provider and procedure clusters
  - $\theta_1$ for providers, $\theta_2$ for patients
  - Can be useful for conspiracy fraud and kickback schemes
Providers-Patients co-clustering

\[ X_{ij} = 1 \text{ if provider } i \text{ serves patient } j \]
Figure: Posterior distributions of memberships of provider 18 and beneficiary 5
Providers-Patients co-clustering

Figure: Posterior distribution of membership probabilities