The Blessings of Multiple Causes

David M. Blei
Departments of Computer Science and Statistics
Data Science Institute
Columbia University
We have *complicated data*; we want to *make sense* of it.
What is *complicated data*?

- many data points; many dimensions
- elaborate structures and relationships (e.g., text)
- different interconnected modalities (e.g., images, links, text, clicks)
What is *making sense of data*?

- make predictions about the future
- identify interpretable patterns
- do science: confirm, elaborate, form causal theories
PROBABILISTIC MACHINE LEARNING

- ML methods that connect domain knowledge to data.
- A methodology for articulating assumptions and computing with them.
- Goal: Make probabilistic ML expressive, scalable, easy to develop.
BAYESIAN STATISTICS

- Statistical methods that connect domain knowledge to data.
- A methodology for articulating assumptions and computing with them
- Goal: Make Bayesian statistics expressive, scalable, easy to develop
Communities discovered in a 3.7M node network of U.S. Patents

[Gopalan and Blei PNAS 2013]
Neuroscience analysis of 220 million fMRI measurements

[Manning+ PLOS ONE 2014]
<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Game</td>
<td>Life</td>
<td>Film</td>
<td>Book</td>
<td>Wine</td>
</tr>
<tr>
<td>Season</td>
<td>Know</td>
<td>Movie</td>
<td>Life</td>
<td>Street</td>
</tr>
<tr>
<td>Team</td>
<td>School</td>
<td>Show</td>
<td>Books</td>
<td>Hotel</td>
</tr>
<tr>
<td>Coach</td>
<td>Street</td>
<td>Life</td>
<td>Novel</td>
<td>House</td>
</tr>
<tr>
<td>Play</td>
<td>Man</td>
<td>Television</td>
<td>Man</td>
<td>Room</td>
</tr>
<tr>
<td>Points</td>
<td>Family</td>
<td>Films</td>
<td>Story</td>
<td>Night</td>
</tr>
<tr>
<td>Games</td>
<td>Says</td>
<td>Director</td>
<td>War</td>
<td>Place</td>
</tr>
<tr>
<td>Giants</td>
<td>House</td>
<td>Man</td>
<td>Author</td>
<td>Restaurant</td>
</tr>
<tr>
<td></td>
<td>Children</td>
<td>Story</td>
<td>House</td>
<td>Park</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>War</td>
<td>Garden</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bush</td>
<td>Building</td>
<td>Won</td>
<td>Yankees</td>
<td>Government</td>
</tr>
<tr>
<td>Campaign</td>
<td>Street</td>
<td>Team</td>
<td>Game</td>
<td>War</td>
</tr>
<tr>
<td>Clinton</td>
<td>Square</td>
<td>Second</td>
<td>Mets</td>
<td>Military</td>
</tr>
<tr>
<td>Republican</td>
<td>House</td>
<td>Race</td>
<td>Season</td>
<td>Officials</td>
</tr>
<tr>
<td>House</td>
<td>House</td>
<td>Round</td>
<td>Run</td>
<td>Iraq</td>
</tr>
<tr>
<td>Party</td>
<td>Buildings</td>
<td>Cup</td>
<td>League</td>
<td>Forces</td>
</tr>
<tr>
<td>Democratic</td>
<td>Space</td>
<td>Open</td>
<td>Baseball</td>
<td>Iraqi</td>
</tr>
<tr>
<td>Political</td>
<td>Percent</td>
<td>Game</td>
<td>Team</td>
<td>Army</td>
</tr>
<tr>
<td>Democrats</td>
<td>Real</td>
<td>Play</td>
<td>Games</td>
<td>Troops</td>
</tr>
<tr>
<td>Senator</td>
<td></td>
<td></td>
<td></td>
<td>Soldiers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children</td>
<td>Stock</td>
<td>Church</td>
<td>Art</td>
<td>Police</td>
</tr>
<tr>
<td>School</td>
<td>Percent</td>
<td>War</td>
<td>Museum</td>
<td>Yesterday</td>
</tr>
<tr>
<td>Women</td>
<td>Companies</td>
<td>Life</td>
<td>Show</td>
<td>Man</td>
</tr>
<tr>
<td>Family</td>
<td>Fund</td>
<td>Black</td>
<td>Gallery</td>
<td>Officer</td>
</tr>
<tr>
<td>Parents</td>
<td>Market</td>
<td>Political</td>
<td>Works</td>
<td>Officers</td>
</tr>
<tr>
<td>Banks</td>
<td></td>
<td>Catholic</td>
<td>Artists</td>
<td>Case</td>
</tr>
<tr>
<td>Child</td>
<td>Investors</td>
<td>Street</td>
<td>Artists</td>
<td>Found</td>
</tr>
<tr>
<td>Life</td>
<td>Bank</td>
<td>Political</td>
<td>Street</td>
<td>Charged</td>
</tr>
<tr>
<td>Says</td>
<td>Funds</td>
<td>Government</td>
<td>Exhibition</td>
<td>Shot</td>
</tr>
<tr>
<td>Help</td>
<td>Financial</td>
<td>Jewish</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother</td>
<td>Business</td>
<td>Pope</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Topics found in 1.8M articles from the New York Times**

[Hoffman+ JMLR 2013]
Population analysis of 2 billion genetic measurements

[Gopalan+ Nature Genetics 2016]
(Fancy) discrete choice analysis of 5.7M purchases

[Ruiz+ 2017]
The probabilistic pipeline

- Customized data analysis is important to many fields.
- Pipeline separates **assumptions, computation, application**
- Eases collaborative solutions to statistics/ML problems
Causal inference from observational data

- How can we understand the world through observation?
- Important to genetics, economics, physics, medicine, finance, ...
- Today: Use probabilistic machine learning for causal inference
The probabilistic pipeline

- Customized data analysis is important to many fields.
- Pipeline separates **assumptions, computation, application**
- Eases collaborative solutions to statistics/ML problems
Causal inference from observational data

How can we understand the world through observation?

- Important to genetics, economics, physics, medicine, finance, ...
- Today: Use probabilistic machine learning for causal inference
This is joint work with Yixin Wang (Statistics)
Credit → Yixin
(Blame → Dave)
An introduction to the deconfounder
A frivolous causal inference problem

- Data about movies: casts and revenue
- Goal: Understand the causal effect of putting an actor in a movie
- Causal: “What will the revenue be if we make a movie with a particular cast?”
The naive solution

<table>
<thead>
<tr>
<th>Title</th>
<th>Cast</th>
<th>Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avatar</td>
<td>[Sam Worthington, Zoe Saldana, Sigourney Weaver, Stephen Lang, ...]</td>
<td>$2788M</td>
</tr>
<tr>
<td>Titanic</td>
<td>[Kate Winslet, Leonardo DiCaprio, Frances Fisher, Billy Zane, ...]</td>
<td>$1845M</td>
</tr>
<tr>
<td>The Avengers</td>
<td>[Robert Downey Jr., Chris Evans, Mark Ruffalo, Chris Hemsworth, ...]</td>
<td>$1520M</td>
</tr>
<tr>
<td>Jurassic World</td>
<td>[Chris Pratt, Bryce Dallas Howard, Irrfan Khan, Vincent D’Onofrio, ...]</td>
<td>$1514M</td>
</tr>
<tr>
<td>Furious 7</td>
<td>[Vin Diesel, Paul Walker, Dwayne Johnson, Michelle Rodriguez, ...]</td>
<td>$1306M</td>
</tr>
<tr>
<td>Avengers: Age of Ultron</td>
<td>[Robert Downey Jr., Chris Hemsworth, Mark Ruffalo, Chris Evans, ...]</td>
<td>$1405M</td>
</tr>
<tr>
<td>Frozen</td>
<td>[Kristen Bell, Idina Menzel, Jonathan Groff, Josh Gad, ...]</td>
<td>$1274M</td>
</tr>
<tr>
<td>Iron Man 3</td>
<td>[Robert Downey Jr., Gwyneth Paltrow, Don Cheadle, Guy Pearce, ...]</td>
<td>$1215M</td>
</tr>
<tr>
<td>Minions</td>
<td>[Sandra Bullock, Jon Hamm, Michael Keaton, Allison Janney, ...]</td>
<td>$1157M</td>
</tr>
<tr>
<td>Captain America: Civil</td>
<td>[Chris Evans, Robert Downey Jr., Scarlett Johansson, Sebastian Stan, ...]</td>
<td>$1153M</td>
</tr>
</tbody>
</table>

- Naive solution: Fit a regression (or use deep learning)
- Actors are features; revenue is the response
- Estimates revenue as a function of which actors are cast
The naive solution

<table>
<thead>
<tr>
<th>Title</th>
<th>Cast</th>
<th>Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avatar</td>
<td>[Sam Worthington, Zoe Saldana, Sigourney Weaver, Stephen Lang, …]</td>
<td>$2788M</td>
</tr>
<tr>
<td>Titanic</td>
<td>[Kate Winslet, Leonardo DiCaprio, Frances Fisher, Billy Zane, …]</td>
<td>$1845M</td>
</tr>
<tr>
<td>The Avengers</td>
<td>[Robert Downey Jr., Chris Evans, Mark Ruffalo, Chris Hemsworth, …]</td>
<td>$1520M</td>
</tr>
<tr>
<td>Jurassic World</td>
<td>[Chris Pratt, Bryce Dallas Howard, Irrfan Khan, Vincent D’Onofrio, …]</td>
<td>$1514M</td>
</tr>
<tr>
<td>Furious 7</td>
<td>[Vin Diesel, Paul Walker, Dwayne Johnson, Michelle Rodriguez, …]</td>
<td>$1506M</td>
</tr>
<tr>
<td>Avengers: Age of Ultron</td>
<td>[Robert Downey Jr., Chris Hemsworth, Mark Ruffalo, Chris Evans, …]</td>
<td>$1405M</td>
</tr>
<tr>
<td>Frozen</td>
<td>[Kristen Bell, Idina Menzel, Jonathan Groff, Josh Gad, …]</td>
<td>$1274M</td>
</tr>
<tr>
<td>Iron Man 3</td>
<td>[Robert Downey Jr., Gwyneth Paltrow, Don Cheadle, Guy Pearce, …]</td>
<td>$1215M</td>
</tr>
<tr>
<td>Minions</td>
<td>[Sandra Bullock, Jon Hamm, Michael Keaton, Allison Janney, …]</td>
<td>$1157M</td>
</tr>
<tr>
<td>Captain America: Civil War</td>
<td>[Chris Evans, Robert Downey Jr., Scarlett Johansson, Sebastian Stan, …]</td>
<td>$1153M</td>
</tr>
</tbody>
</table>

- But standard ML does not (necessarily) provide causal inferences
- Whether an actor was cast is different from casting an actor
- Causal inference is about prediction under intervention
- James Bond movies are about James Bond, a British spy
- Cast James Bond, M, Q, Ms. Moneypenny
- M, Q, Ms Moneypenny only appear in Bond movies
- Bond movies always do well at the box office
The naive solution

- James Bond-ness is an **unobserved confounder**.
- Confounders affect both the cast ("causes") and the revenue ("effect")
- Confounders bias "passive ML," when used to predict interventions.
  - Some actors overestimated; others are underestimated
Unobserved confounders are everywhere.
What is causal inference?

- Causal inference is about **prediction under intervention**.
  
  [Hernan and Robins 2019; Imbens and Rubin 2015; Pearl 2009]

- “What will the revenue be if we make a movie with a particular cast?”

- Challenge: Unobserved confounders (like James Bond-ness)
The classical solution

THINK ABOUT CONFOUNDERS

MEASURE CONFOUNDERS
\{\omega_1, \ldots, \omega_n\}

ESTIMATE CAUSAL EFFECTS

\[ E[y \mid do(a)] = E[E[y \mid w, a=a]] \]

DATA

\{actors_1, revenue_1, actors_2, revenue_2, \ldots, actors_n, revenue_n\}
This approach requires that we find and measure **sufficient confounders**.

But whether we included sufficient confounders is **untestable**.

The classical solution rests on **hope**. (And it makes us **worry**.)
<table>
<thead>
<tr>
<th>Title</th>
<th>Cast</th>
<th>Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avatar</td>
<td>(Sam Worthington, Zoe Saldana, Sigourney Weaver, Stephen Lang, ...)</td>
<td>$2788M</td>
</tr>
<tr>
<td>Titanic</td>
<td>(Kate Winslet, Leonardo DiCaprio, Frances Fisher, Billy Zane, ...)</td>
<td>$1845M</td>
</tr>
<tr>
<td>The Avengers</td>
<td>(Robert Downey Jr., Chris Evans, Mark Ruffalo, Chris Hemsworth, ...)</td>
<td>$1520M</td>
</tr>
<tr>
<td>Jurassic World</td>
<td>(Chris Pratt, Bryce Dallas Howard, Irrfan Khan, Vincent D’Onofrio, ...)</td>
<td>$1514M</td>
</tr>
<tr>
<td>Furious 7</td>
<td>(Vin Diesel, Paul Walker, Dwayne Johnson, Michelle Rodriguez, ...)</td>
<td>$1506M</td>
</tr>
<tr>
<td>Avengers: Age of Ultron</td>
<td>(Robert Downey Jr., Chris Hemsworth, Mark Ruffalo, Chris Evans, ...)</td>
<td>$1405M</td>
</tr>
<tr>
<td>Frozen</td>
<td>(Kristen Bell, Idina Menzel, Jonathan Groff, Josh Gad, ...)</td>
<td>$1274M</td>
</tr>
<tr>
<td>Iron Man 3</td>
<td>(Robert Downey Jr., Gwyneth Paltrow, Don Cheadle, Guy Pearce, ...)</td>
<td>$1215M</td>
</tr>
<tr>
<td>Minions</td>
<td>(Sandra Bullock, Jon Hamm, Michael Keaton, Allison Janney, ...)</td>
<td>$1157M</td>
</tr>
<tr>
<td>Captain America: Civil War</td>
<td>(Chris Evans, Robert Downey Jr., Scarlett Johansson, Sebastian Stan, ...)</td>
<td>$1153M</td>
</tr>
</tbody>
</table>

- But our problem is not classical.
- There are many causes (one per actor)—**multiple causal inference**
- *Multiple causes helps construct a variable that contains confounders.*
The deconfounder

MODEL
ASSIGNED
CAUSES

\[ \{Z_1, \ldots, Z_n\} \]
\[ \hat{Z}_i = \mathbb{E}[Z_i | A_i = a_i] \]

ESTIMATE
SUBSTITUTE
CONFOUNDERS

\[ \mathbb{E}[Y | do(a)] = \mathbb{E}[\mathbb{E}[Y | Z, A = a]] \]

ESTIMATE
CAUSAL
EFFECTS

DATA

\{\text{actors}_1, \text{revenue}_1, \ldots, \text{actors}_m, \text{revenue}_m\}
The deconfounder

Find, fit, and check a factor model of the assigned causes.
Use the model to form substitute confounders for each individual.
Use the substitute confounders in a causal model of the outcome.
The deconfounder

- Find, fit, and check a **probabilistic matrix factorization** of movie casts.
- Use the model to infer the **per-movie variables** in the matrix factorization.
- Use these variables in a **regression from casts to earnings**.
Case study: Actors

- “Overestimated”:

- “Underestimated”:

- Most “corrected”:
Intuition and assumptions

- Intuition: “Multi-cause confounders” induce dependence among the causes.
- That dependence is encoded in the data; we can capture it with a factor model.
- **Assumption:** No unobserved single-cause confounders
  - But this is weaker than “no unobserved confounders”
Intuition and assumptions

Intuition: “Multi-cause confounders” induce dependence among the causes.
That dependence is encoded in the data; we can capture it with a factor model.

Assumption: No unobserved single-cause confounders
  - But this is weaker than “no unobserved confounders”
How do genes affect a trait?

- The causes are genetic variation
- The effect is a trait
- Confounder: Each person’s ancestry induces correlation in *multiple genes.*
How do sports players affect how well the team is doing?

- The causes are who is in the game.
- The effect is the points scored in the game.
- Confounder: The coach uses *multiple players* together.
How do prices of items affect how much money is spent?

- The causes are the prices of each item for sale.
- The effect is how much money is spent by consumers.
- Confounder: Holidays affect the prices and demand of multiple items.
The deconfounder in more detail
Multiple causal inference

<table>
<thead>
<tr>
<th>Title</th>
<th>Cast</th>
<th>Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avatar</td>
<td>[Sam Worthington, Zoe Saldana, Sigourney Weaver, Stephen Lang, …]</td>
<td>$2788M</td>
</tr>
<tr>
<td>Titanic</td>
<td>[Kate Winslet, Leonardo DiCaprio, Frances Fisher, Billy Zane, …]</td>
<td>$1845M</td>
</tr>
<tr>
<td>The Avengers</td>
<td>[Robert Downey Jr., Chris Evans, Mark Ruffalo, Chris Hemsworth, …]</td>
<td>$1520M</td>
</tr>
<tr>
<td>Jurassic World</td>
<td>[Chris Pratt, Bryce Dallas Howard, Irrfan Khan, Vincent D’Onofrio, …]</td>
<td>$1514M</td>
</tr>
<tr>
<td>Furious 7</td>
<td>[Vin Diesel, Paul Walker, Dwayne Johnson, Michelle Rodriguez, …]</td>
<td>$1506M</td>
</tr>
<tr>
<td>Avengers: Age of Ultron</td>
<td>[Robert Downey Jr., Chris Hemsworth, Mark Ruffalo, Chris Evans, …]</td>
<td>$1405M</td>
</tr>
<tr>
<td>Frozen</td>
<td>[Kristen Bell, Idina Menzel, Jonathan Groff, Josh Gad, …]</td>
<td>$1274M</td>
</tr>
<tr>
<td>Iron Man 3</td>
<td>[Robert Downey Jr., Gwyneth Paltrow, Don Cheadle, Guy Pearce, …]</td>
<td>$1215M</td>
</tr>
<tr>
<td>Minions</td>
<td>[Sandra Bullock, Jon Hamm, Michael Keaton, Allison Janney, …]</td>
<td>$1157M</td>
</tr>
<tr>
<td>Captain America: Civil War</td>
<td>[Chris Evans, Robert Downey Jr., Scarlett Johansson, Sebastian Stan, …]</td>
<td>$1153M</td>
</tr>
</tbody>
</table>

- Observed dataset \( D = \{(a_1, y_1), \ldots, (a_n, y_n)\} \)
  - assigned causes \( a_i = \{a_{i1}, \ldots, a_{im}\} \)
  - outcome \( y_i \)

- Goal: Do causal inference, \( \mathbb{E}[Y; \text{do}(a)] \)
  - “The expectation of \( Y \) in the model where we intervened on \( a \)."
Multiple causal inference

<table>
<thead>
<tr>
<th>Title</th>
<th>Cast</th>
<th>Revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avatar</td>
<td>[Sam Worthington, Zoe Saldana, Sigourney Weaver, Stephen Lang, …]</td>
<td>$2788M</td>
</tr>
<tr>
<td>Titanic</td>
<td>[Kate Winslet, Leonardo DiCaprio, Frances Fisher, Billy Zane, …]</td>
<td>$1845M</td>
</tr>
<tr>
<td>The Avengers</td>
<td>[Robert Downey Jr., Chris Evans, Mark Ruffalo, Chris Hemsworth, …]</td>
<td>$1520M</td>
</tr>
<tr>
<td>Jurassic World</td>
<td>[Chris Pratt, Bryce Dallas Howard, Irrfan Khan, Vincent D’Onofrio, …]</td>
<td>$1514M</td>
</tr>
<tr>
<td>Furious 7</td>
<td>[Vin Diesel, Paul Walker, Dwayne Johnson, Michelle Rodriguez, …]</td>
<td>$1506M</td>
</tr>
<tr>
<td>Avengers: Age of Ultron</td>
<td>[Robert Downey Jr., Chris Hemsworth, Mark Ruffalo, Chris Evans, …]</td>
<td>$1405M</td>
</tr>
<tr>
<td>Frozen</td>
<td>[Kristen Bell, Idina Menzel, Jonathan Groff, Josh Gad, …]</td>
<td>$1274M</td>
</tr>
<tr>
<td>Iron Man 3</td>
<td>[Robert Downey Jr., Gwyneth Paltrow, Don Cheadle, Guy Pearce, …]</td>
<td>$1215M</td>
</tr>
<tr>
<td>MInions</td>
<td>[Sandra Bullock, Jon Hamm, Michael Keaton, Allison Janney, …]</td>
<td>$1157M</td>
</tr>
<tr>
<td>Captain America: Civil War</td>
<td>[Chris Evans, Robert Downey Jr., Scarlett Johansson, Sebastian Stan, …]</td>
<td>$1153M</td>
</tr>
</tbody>
</table>

- If there are unobserved confounders then
\[ \mathbb{E}[Y; \text{do}(a)] \neq \mathbb{E}[Y | A = a]. \]

- We can calculate the right term from data, but it's not equal to the left term.
The deconfounder

Find, fit, and check a **factor model** of the movie casts.

Use the factor model to form **substitute confounders** for each movie.

Use the substitute confounders in a **causal model** of movie revenue.
Fit a probabilistic factor model

A probabilistic factor model has the following form,

\[ \beta_j \sim p(\beta_j) \quad j = 1, \ldots, m \]
\[ z_i \sim p(z_i) \quad i = 1, \ldots, n \]
\[ a_{ij} \sim p(a_{ij} | z_i, \beta_j) \]

E.g., mixtures, matrix factorization, deep generative models, topic models, ...
Poisson factorization [Gopalan+ 2015]

\[ \beta_{jk} \sim \text{Gam}(a, b) \quad i \in \{1, \ldots, n\} \]
\[ z_{ik} \sim \text{Gam}(a, b) \quad j \in \{1, \ldots, m\} \]
\[ a_{ij} \sim \text{Poi}(z_i^T \beta_j) \quad k \in \{1, \ldots, d\} \]

- Provides a generative model of the assigned causes \( a_{ij} \).
- Can be approximated on large datasets with variational methods.
- A Bayesian form of non-negative matrix factorization [Lee and Seung 1999]
Poisson factorization [Gopalan+ 2015]

\[\beta_{jk} \sim \text{Gam}(a, b) \quad i \in \{1, \ldots, n\}\]

\[z_{ik} \sim \text{Gam}(a, b) \quad j \in \{1, \ldots, m\}\]

\[a_{ij} \sim \text{Poi}(z_i^\top \beta_j) \quad k \in \{1, \ldots, d\}\]

- Consider the dataset of casts \(a_{1:n}\).
- Approximate the posterior distribution \(p(z_{1:n}, \beta_{1:m} \mid a_{1:n})\).
- We only model the actors \(a_i\); the outcome is not involved.
Check the factor model

- Star Wars
- Nemo
- Unforgiven
- 8 Mile
- Blade Runner
- American Beauty
- Raiders of the Lost Ark
- Forrest Gump
- Before Sunset

- Estimate the local latent variable $\hat{z}_i = \mathbb{E}_{\text{model}}[Z | a_i, \beta]$.
- Check how well $\hat{z}_i$ captures the distribution of the actors.
- E.g., use a **predictive check** on actors. (No need for exact inference.)
Poisson factorization [Gopalan+ 2015]

\[
\begin{align*}
\beta_{jk} & \sim \text{Gam}(a, b) & i & \in \{1, \ldots, n\} \\
z_{ik} & \sim \text{Gam}(a, b) & j & \in \{1, \ldots, m\} \\
a_{ij} & \sim \text{Poi}(z_i^\top \beta_j) & k & \in \{1, \ldots, d\}
\end{align*}
\]

- Provides a generative model of the assigned causes \( a_{ij} \).
- Can be approximated on large datasets with variational methods.
- A Bayesian form of non-negative matrix factorization [Lee and Seung 1999]
Check the factor model

- Star Wars
- Nemo
- Unforgiven

- 8 Mile
- Blade Runner
- American Beauty

- Raiders of the Lost Ark
- Forrest Gump
- Before Sunset

- Estimate the local latent variable $\hat{z}_i = E_{\text{model}}[Z | a_i, \beta]$.
- Check how well $\hat{z}_i$ captures the distribution of the actors.
- E.g., use a **predictive check** on actors. (No need for exact inference.)
Check the factor model

<table>
<thead>
<tr>
<th>Model</th>
<th>Predictive score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilistic PCA</td>
<td>0.14</td>
</tr>
<tr>
<td>Poisson factorization</td>
<td>0.16</td>
</tr>
<tr>
<td>Mixtures</td>
<td>0.01</td>
</tr>
<tr>
<td>Deep exponential families</td>
<td>0.19</td>
</tr>
</tbody>
</table>
Check the factor model

Estimate the local latent variable $\hat{z}_i = \mathbb{E}_{\text{model}}[Z \mid a_i, \beta]$. 

Check how well $\hat{z}_i$ captures the distribution of the actors.

E.g., use a predictive check on actors. (No need for exact inference.)
Check the factor model

<table>
<thead>
<tr>
<th>Model</th>
<th>Predictive score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probabilistic PCA</td>
<td>0.14</td>
</tr>
<tr>
<td>Poisson factorization</td>
<td>0.16</td>
</tr>
<tr>
<td>Mixtures</td>
<td>0.01</td>
</tr>
<tr>
<td>Deep exponential families</td>
<td>0.19</td>
</tr>
</tbody>
</table>
Do causal inference

- \{Sam Worthington, Zoe Saldana, Sigourney Weaver, Stephen Lang, ...\} $2788M
- \{Kate Winslet, Leonardo DiCaprio, Frances Fisher, Billy Zane, ...\} $1845M
- \{Robert Downey Jr., Chris Evans, Mark Ruffalo, Chris Hemsworth, ...\} $1520M
- \{Chris Pratt, Bryce Dallas Howard, Irrfan Khan, Vincent D'Onofrio, ...\} $1514M
- \{Vin Diesel, Paul Walker, Dwayne Johnson, Michelle Rodriguez, ...\} $1506M
- \{Robert Downey Jr., Chris Hemsworth, Mark Ruffalo, Chris Evans, ...\} $1405M
- \{Kristen Bell, Idina Menzel, Jonathan Groff, Josh Gad, ...\} $1274M
- \{Robert Downey Jr., Gwyneth Paltrow, Don Cheadle, Guy Pearce, ...\} $1215M
- \{Sandra Bullock, Jon Hamm, Michael Keaton, Allison Janney, ...\} $1157M
- \{Chris Evans, Robert Downey Jr., Scarlett Johansson, Sebastian Stan, ...\} $1153M

- The estimated local variables $\hat{z}_i$ are substitute confounders.
- They are latent attributes of movie casts that the factorization has discovered.
- Form an augmented dataset of triplets $(a_i, y_i, \hat{z}_i)$. 
Do causal inference

- Use the substitute confounders in a causal inference.
- E.g., fit regression from casts and confounders to revenue,
  \[
  \mathbb{E}[Y \mid a, \hat{Z}] = \beta^T a + \eta^T \hat{Z}.
  \]
- Use adjustment/the g-formula to perform causal inference,
  \[
  \mathbb{E}[Y \mid \text{do}(a)] \approx \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[Y \mid a, \hat{Z}].
  \]
The deconfounder

Find, fit, and check a factor model of the movie casts.
Use the factor model to form substitute confounders for each movie.
Use the substitute confounders in a causal model of movie revenue.
Case study: Actors

- Overestimated:

- Underestimated:

- Most corrected:
A little theory
The deconfounder

- Find, fit, and check a **factor model** of the movie casts.
- Use the factor model to form **substitute confounders** for each movie.
- Use the substitute confounders in a **causal model** of movie revenue.
A little theory
The deconfounder

Suppose we fit a **good factor model** of the assigned causes (the actors).

Then its local latent variable will contain **multi-cause confounders**.

Main assumption: No single cause confounders.
Intuition (through graphical models)

If we find a good factor model then

\[ p(a_{i1}, \ldots, a_{im} | z_i, \beta_{1:m}) = \prod_{j=1}^m p(a_{ij} | z_i, \beta_j) \]
Intuition (through graphical models)

There cannot be an unobserved multi-cause confounder.

Contradiction: If one existed then the independence statement would not hold.
Intuition (through graphical models)

Note: there still might be a single-cause confounder

This is a weaker assumption than “strong ignorability.”
Theory: It works

THEOREM: THE DECONFOUNDER

Suppose $p_{\text{true}}(a)$ can be written $\int p(z) \prod_j p(a_j | z, \beta) dz$.

Then $Z$ blocks the backdoor path between the causes and the effect.

This implies that,

$$\mathbb{E} [Y ; \text{do}(a)] = \mathbb{E}_Z [\mathbb{E}_Y [Y | Z, a]].$$

Thus we can estimate the interventional expectation.

(It's a little more nuanced than this; ask me later...)
Simulation study
We did many simulations and studies
GWAS is a problem of multiple causal inference

How is genetic variation causally connected to a trait?

For each individual: a trait and many measurements of the genome (SNPs).
Example: Genome-wide association studies (GWAS)

- Multiple-cause confounding is a problem.
- Non-causal SNPs may be highly correlated to causal SNPs
- Misestimates causal effects
Simulation study

<table>
<thead>
<tr>
<th>ID (i)</th>
<th>SNP_1 ((a_{i,1}))</th>
<th>SNP_2 ((a_{i,2}))</th>
<th>SNP_3 ((a_{i,3}))</th>
<th>SNP_4 ((a_{i,4}))</th>
<th>SNP_5 ((a_{i,5}))</th>
<th>SNP_6 ((a_{i,6}))</th>
<th>SNP_7 ((a_{i,7}))</th>
<th>SNP_8 ((a_{i,8}))</th>
<th>SNP_9 ((a_{i,9}))</th>
<th>...</th>
<th>SNP_100K ((a_{i,100K}))</th>
<th>Height (feet) ((y_i))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>...</td>
<td>5.73</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td>2</td>
<td>5.26</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>6.24</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td></td>
<td>0</td>
<td>5.78</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td>1</td>
<td>5.09</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- Generate SNPs \(a_{ij}\), where each individual belongs to a latent group \(c_i\).
- The true outcome is a trait \(y_i\), drawn from
  \[
  y_i = \sum_j \beta_j a_{ij} + \lambda_{c_i} + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma_{c_i}),
  \]
  where many \(\beta_j\) are zero, i.e., non-causal SNPs.
- **Confounded**: the intercept \(\lambda_{c_i}\) and error \(\varepsilon_i\) are connected to the latent group.
## Simulation study

<table>
<thead>
<tr>
<th>Method</th>
<th>Pred. score</th>
<th>Real-valued outcome (RMSE × 10^2)</th>
<th>Binary outcome (RMSE × 10^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No control</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Control for confounders*</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(G)LMM</td>
<td>—</td>
<td>0.14</td>
<td>—</td>
</tr>
<tr>
<td>PPCA</td>
<td>0.14</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>PF</td>
<td>0.15</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>LFA</td>
<td>0.14</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Mixture</td>
<td>0.00</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>DEF</td>
<td>0.20</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*We fit many factor models; none was the true model.*

*Each provides different levels of predictive performance.*

*All computation done in Edward [Tran+ 2018].*
Simulation study

<table>
<thead>
<tr>
<th></th>
<th>pred. score</th>
<th>Real-valued outcome RMSE×10²</th>
<th>Binary outcome RMSE×10²</th>
</tr>
</thead>
<tbody>
<tr>
<td>No control</td>
<td>—</td>
<td>58.82</td>
<td>29.50</td>
</tr>
<tr>
<td>Control for confounders*</td>
<td>—</td>
<td>25.32</td>
<td>25.77</td>
</tr>
<tr>
<td>(G)LMM</td>
<td>—</td>
<td>35.18</td>
<td>28.87</td>
</tr>
<tr>
<td>PPCA</td>
<td>0.14</td>
<td>33.32</td>
<td>26.70</td>
</tr>
<tr>
<td>PF</td>
<td>0.15</td>
<td>33.38</td>
<td>26.84</td>
</tr>
<tr>
<td>LFA</td>
<td>0.14</td>
<td>33.93</td>
<td>26.83</td>
</tr>
<tr>
<td>Mixture</td>
<td>0.00</td>
<td>57.59</td>
<td>29.96</td>
</tr>
<tr>
<td>DEF</td>
<td>0.20</td>
<td><strong>26.47</strong></td>
<td><strong>25.91</strong></td>
</tr>
</tbody>
</table>

- Also fit outcome models with no control and with observed confounders
- The deconfounder provides good causal estimates.
- Predictive checks indicate downstream causal performance.
Explains and justifies existing methods for GWAS

- Linear mixed models [Yu+ 2006; Kang+ 2008; etc.]
- Principal component analysis [Price+ 2006]
- Logistic factor analysis [Song+ 2015; Hao+ 2015]
- Mixed-membership models [Pritchard+ 2000a,b; Falush+ 2003; Falush+ 2007]
- Deep generative models [Tran and Blei 2018]
### Simulation study

<table>
<thead>
<tr>
<th></th>
<th>pred. score</th>
<th>Real-valued outcome $\text{RMSE} \times 10^2$</th>
<th>Binary outcome $\text{RMSE} \times 10^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No control</td>
<td>—</td>
<td>58.82</td>
<td>29.50</td>
</tr>
<tr>
<td>Control for confounders*</td>
<td>—</td>
<td>25.32</td>
<td>25.77</td>
</tr>
<tr>
<td>(G)LMM</td>
<td>—</td>
<td>35.18</td>
<td>28.87</td>
</tr>
<tr>
<td>PPCA</td>
<td>0.14</td>
<td>33.32</td>
<td>26.70</td>
</tr>
<tr>
<td>PF</td>
<td>0.15</td>
<td>33.38</td>
<td>26.84</td>
</tr>
<tr>
<td>LFA</td>
<td>0.14</td>
<td>33.93</td>
<td>26.83</td>
</tr>
<tr>
<td>Mixture</td>
<td>0.00</td>
<td>57.59</td>
<td>29.96</td>
</tr>
<tr>
<td>DEF</td>
<td>0.20</td>
<td><strong>26.47</strong></td>
<td><strong>25.91</strong></td>
</tr>
</tbody>
</table>

- Also fit outcome models with no control and with observed confounders
- The deconfounder provides good causal estimates.
- Predictive checks indicate downstream causal performance.
Explains and justifies existing methods for GWAS

- Linear mixed models [Yu+ 2006; Kang+ 2008; etc.]
- Principal component analysis [Price+ 2006]
- Logistic factor analysis [Song+ 2015; Hao+ 2015]
- Mixed-membership models [Pritchard+ 2000a,b; Falush+ 2003; Falush+ 2007]
- Deep generative models [Tran and Blei 2018]
Discussion
Causal inference from observational data

- How can we understand the world through observation?
- Important to genetics, economics, physics, medicine, finance, ...
- Today: Use probabilistic machine learning for causal inference
The deconfounder

- Find, fit, and check a **factor model** of the assigned causes.
- Use the factor model to form **substitute confounders** for each individual.
- Use the substitute confounders in a **causal model** of the outcome.
The deconfounder

Suppose we fit a **good factor model** of the assigned causes (the actors).

Then its local latent variable will contain **multi-cause confounders**.

(There are assumptions.)
The deconfounder

- Uses **probabilistic machine learning** for **causal inference**.
- Can employ **approximate inference** and **Bayesian model checking**.
- Requires **weaker assumptions** than classical causal inference.
Further reading and current research

https://arxiv.org/abs/1805.06826

▶ Other readings
- Tran and Blei (2018), ICLR
- Ranganath and Perotte (2018), arXiv 1805.08273

▶ Current research about the deconfounder
- SEMs and the causal graphical view
- testing with the deconfounder
- understanding the bias-variance trade-off of the deconfounder
- latent mediators & mechanisms
- many applications (medicine, recommendation, sports, fairness, ...)

Extra slides
Identification
On identification

- A causal quantity is identifiable if it can be written as a function of the observed variables.
- If the causal quantity changes, so does the distribution of the observed data.
- D'Amour (2019) gives two examples where $\mathbb{E}[Y; \text{do}(a)]$ is not identifiable.
- These results help flesh out the theory of multiple causal inference.
- But identification is still possible (with assumptions).
On identification

- Assume we pinpoint a substitute confounder \( \hat{z} = f(a) \), e.g., many causes.

- (Theorem) Differences of complete interventions are
  \[
  E[Y; do(a)] - E[Y; do(a')] .
  \]
  They are nonparametrically identifiable when the outcome separates contributions from the unobserved confounders and causes.

- (Theorem) Consider a subset of causes \( B \). The subset intervention is
  \[
  E[Y; do(a_B)].
  \]
  It is identifiable with overlap on the subset, \( p(a_B | z) > 0 \).