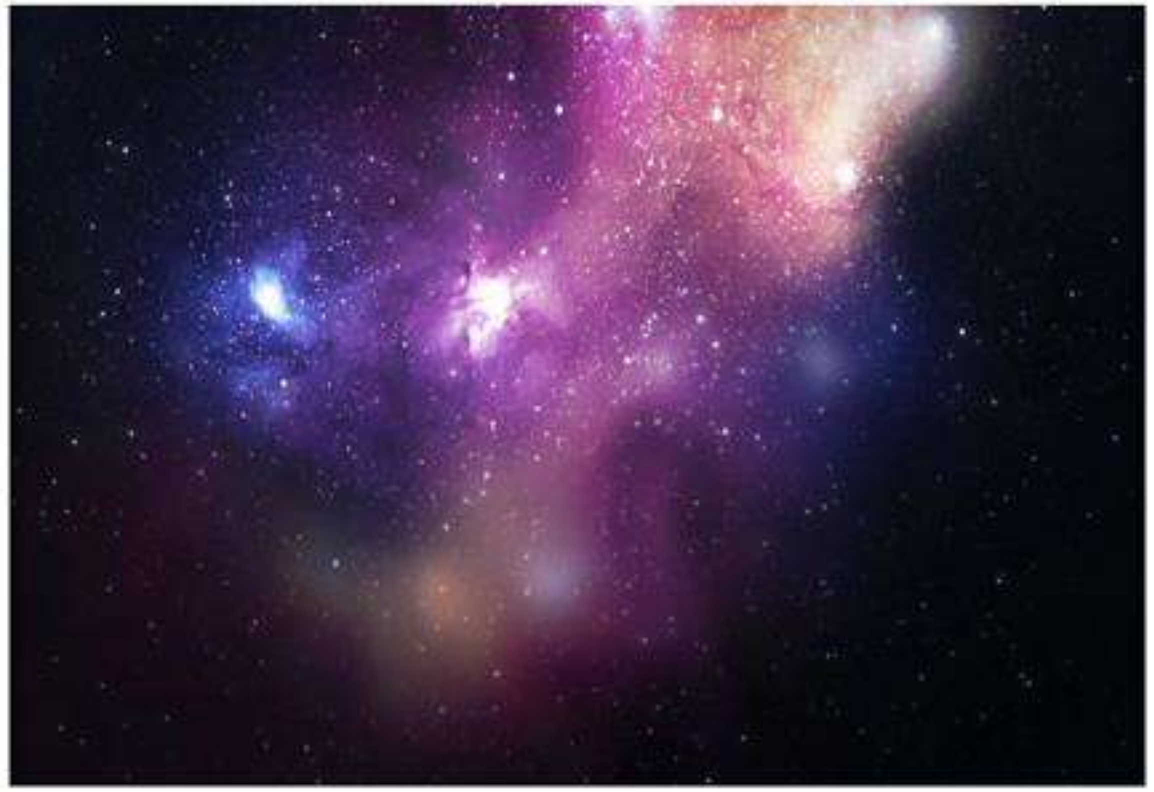
The Blessings of Multiple Causes

David M. Blei
Departments of Computer Science and Statistics
Data Science Institute
Columbia University

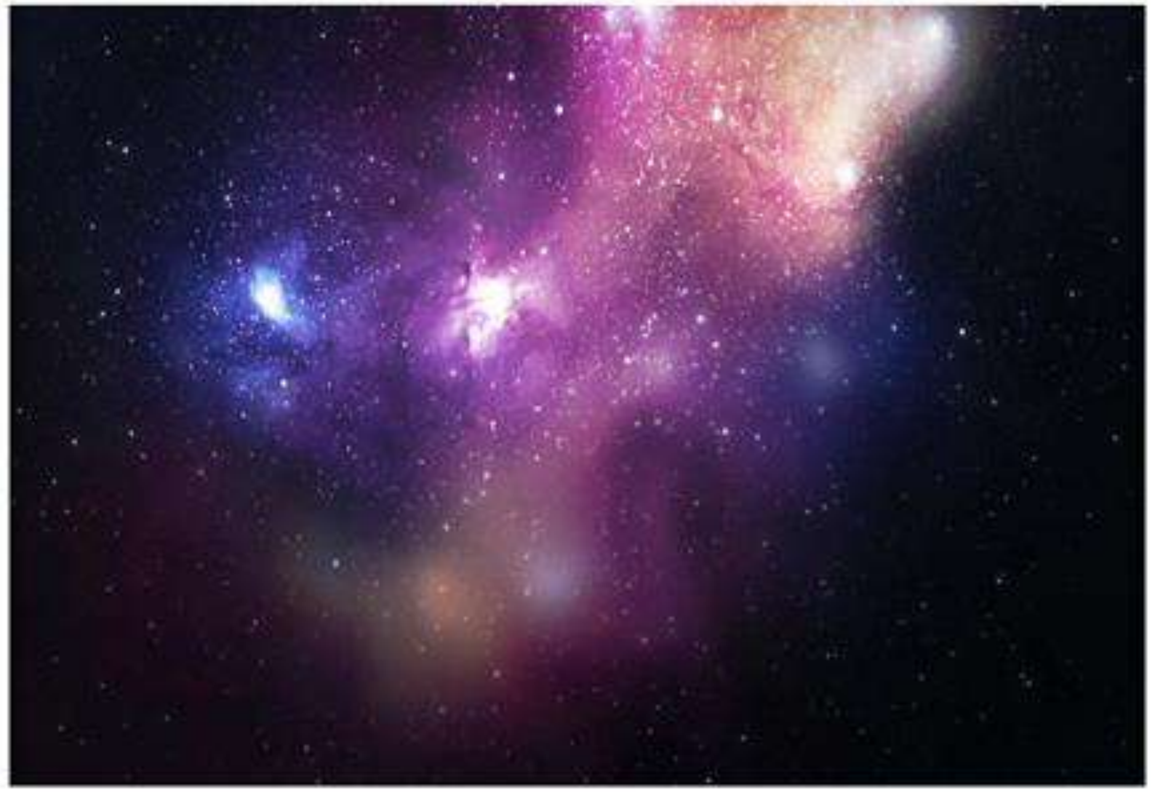



We have ***complicated data***; we want to ***make sense*** of it.



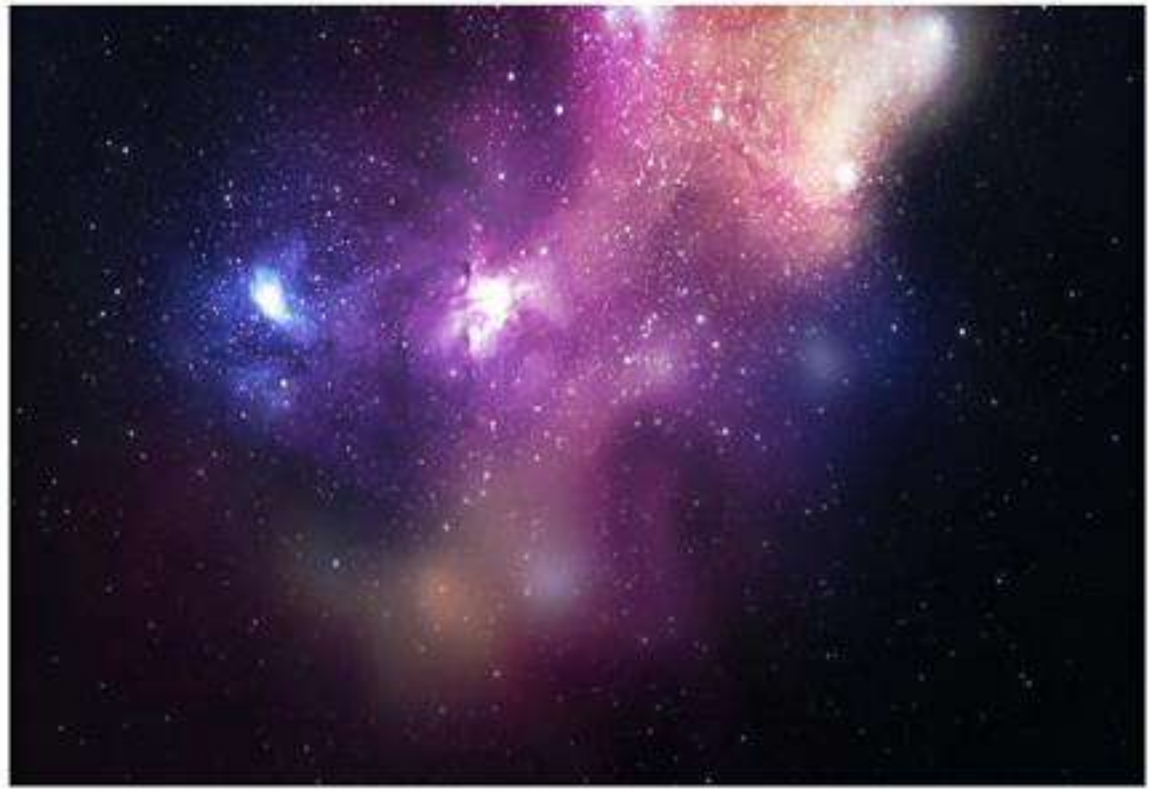
What is ***complicated data***?

- many data points; many dimensions
- elaborate structures and relationships (e.g., text)
- different interconnected modalities (e.g., images, links, text, clicks)



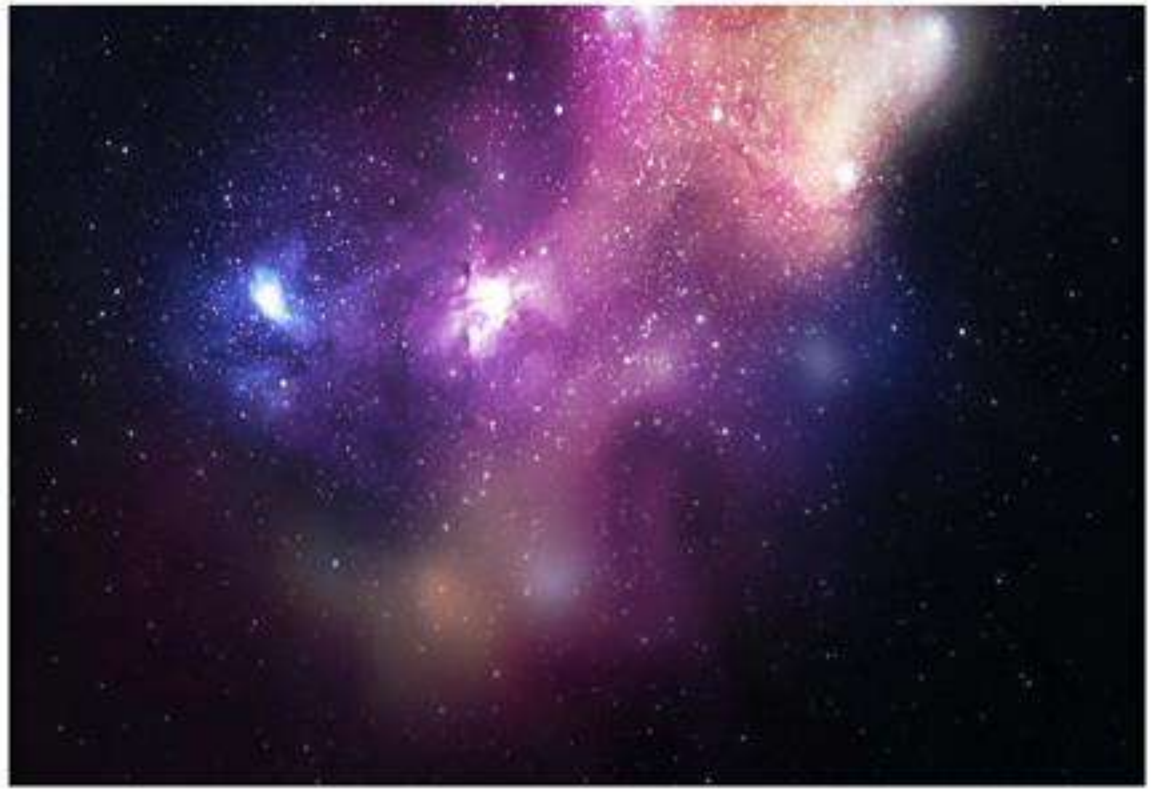
What is ***making sense of data***?

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



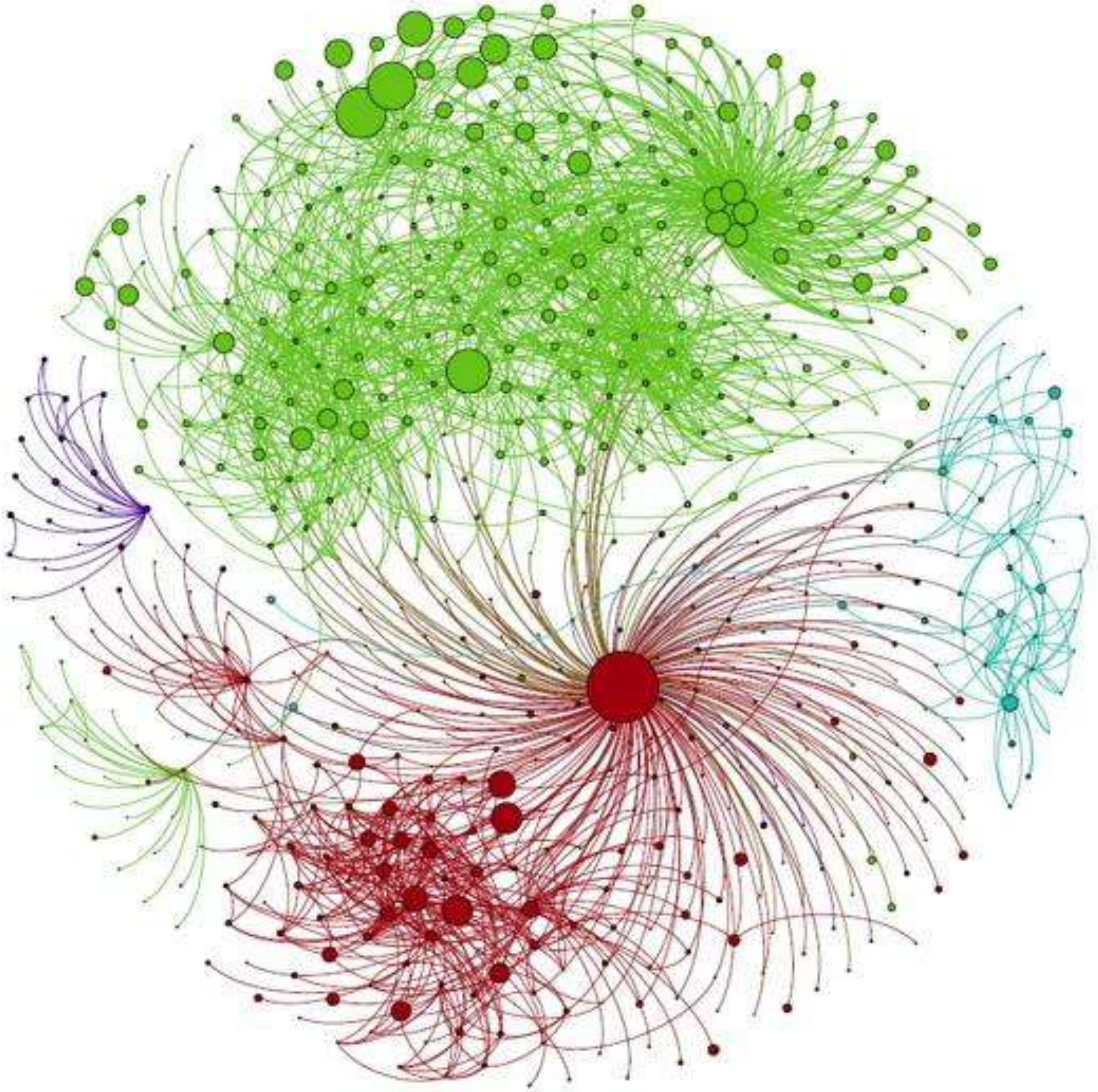
PROBABILISTIC MACHINE LEARNING

- 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*



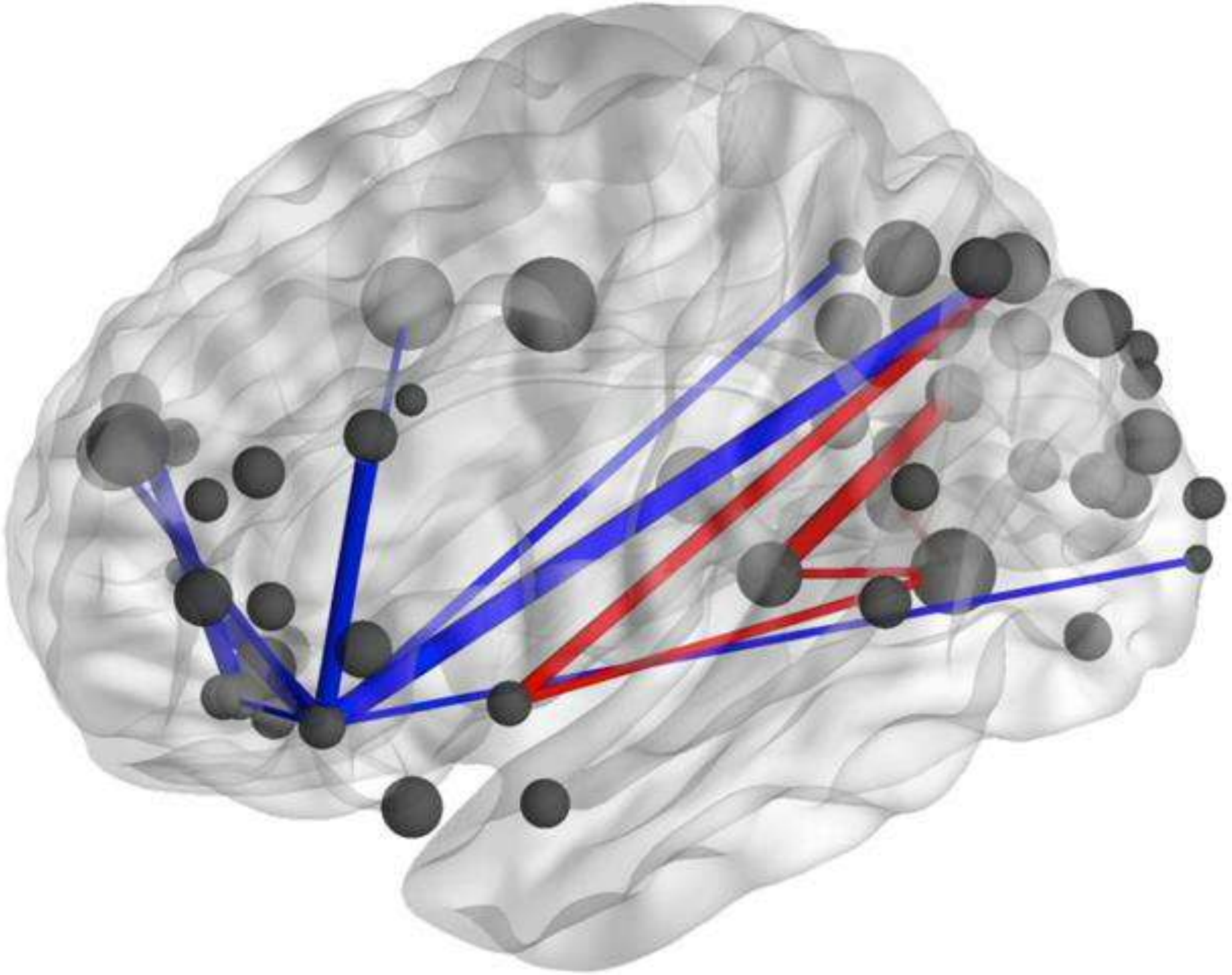
BAYESIAN STATISTICS

- 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]

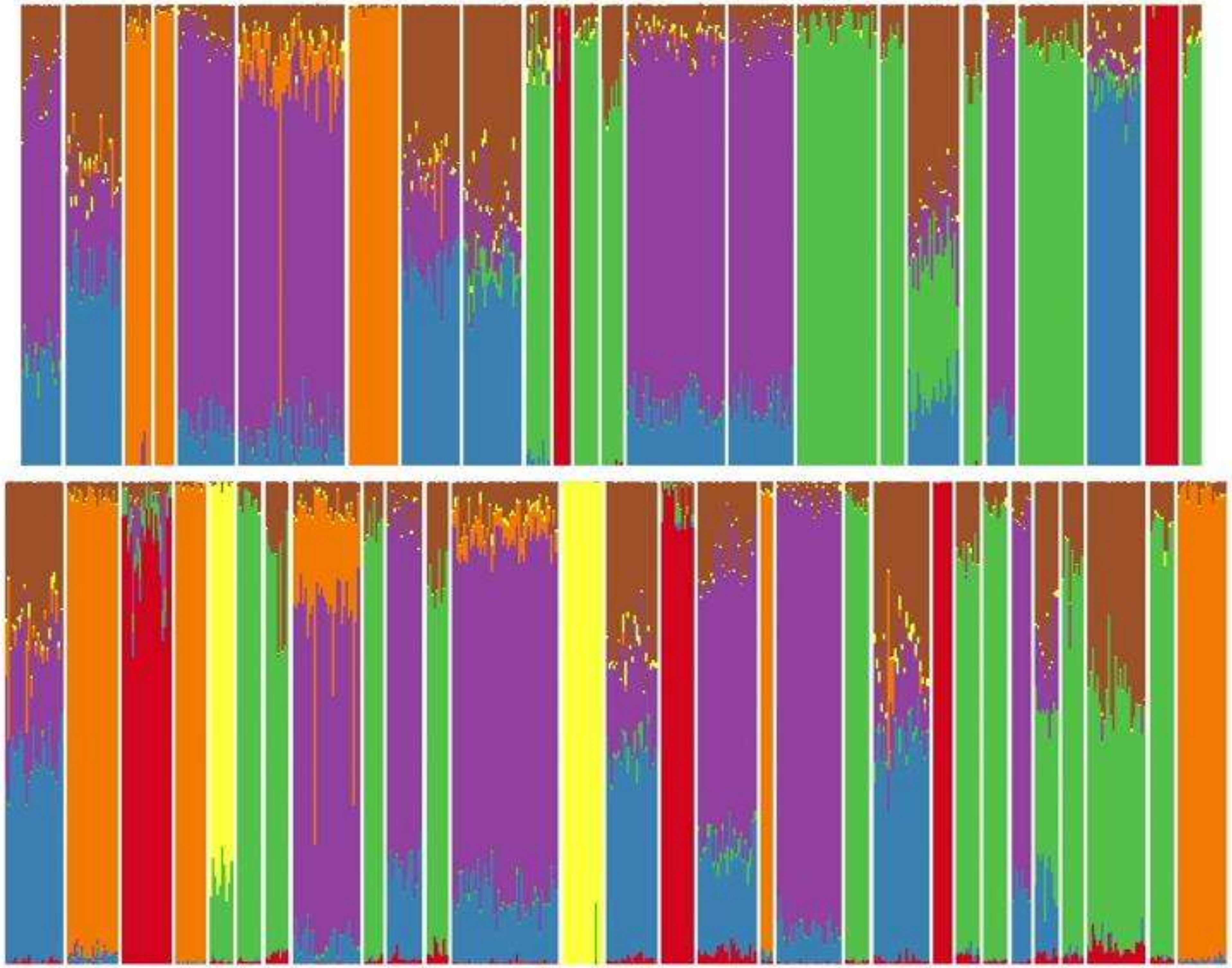
1	2	3	4	5
Game	Life	Film	Book	Wine
Season	Know	Movie	Life	Street
Team	School	Show	Books	Hotel
Coach	Street	Life	Novel	House
Play	Man	Television	Story	Room
Points	Family	Films	Man	Night
Games	Says	Director	Author	Place
Giants	House	Man	House	Restaurant
Second	Children	Story	War	Park
Players	Night	Says	Children	Garden

6	7	8	9	10
Bush	Building	Won	Yankees	Government
Campaign	Street	Team	Game	War
Clinton	Square	Second	Mets	Military
Republican	Housing	Race	Season	Officials
House	House	Round	Run	Iraq
Party	Buildings	Cup	League	Forces
Democratic	Development	Open	Baseball	Iraqi
Political	Space	Game	Team	Army
Democrats	Percent	Play	Games	Troops
Senator	Real	Win	Hit	Soldiers

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

Topics found in 1.8M articles from the New York Times

[Hoffman+ JMLR 2013]



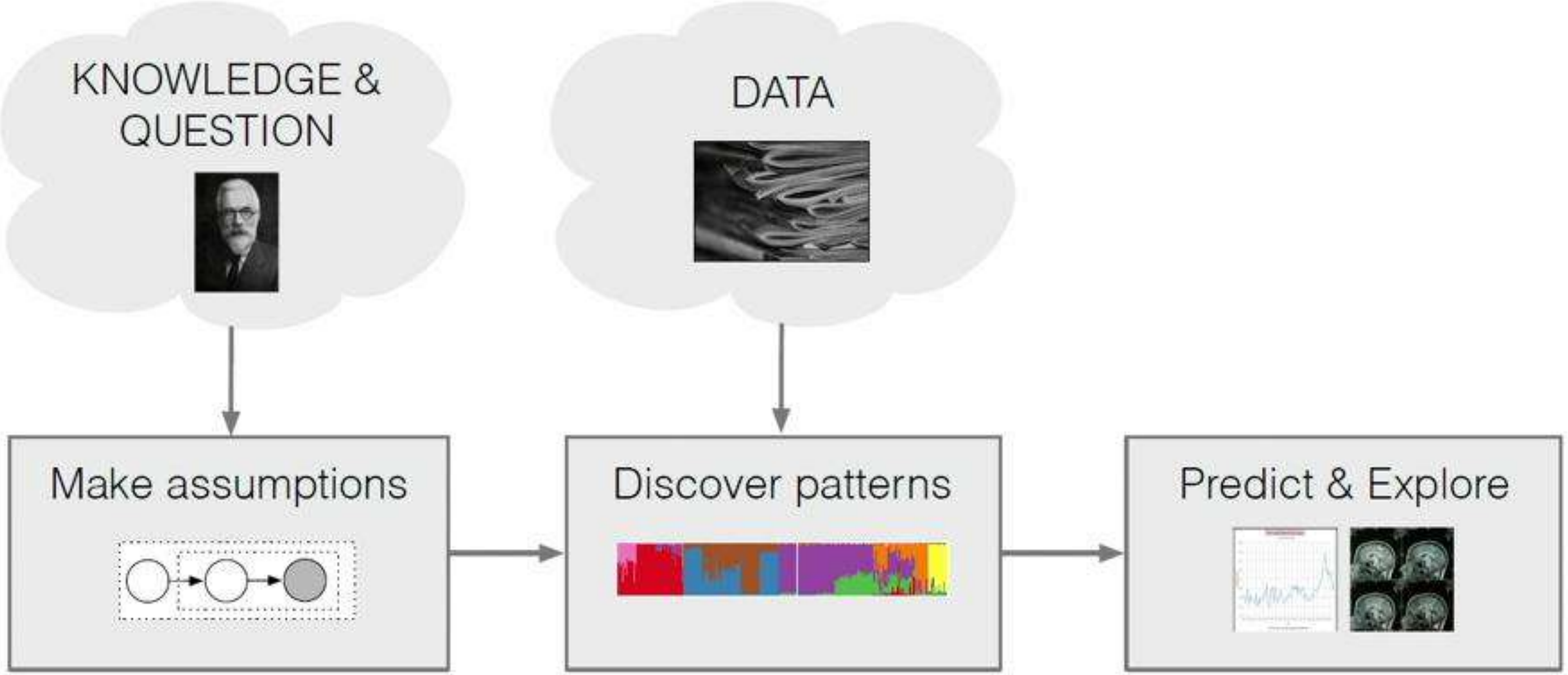
Population analysis of 2 billion genetic measurements

[Gopalan+ Nature Genetics 2016]

(Fancy) discrete choice analysis of 5.7M purchases

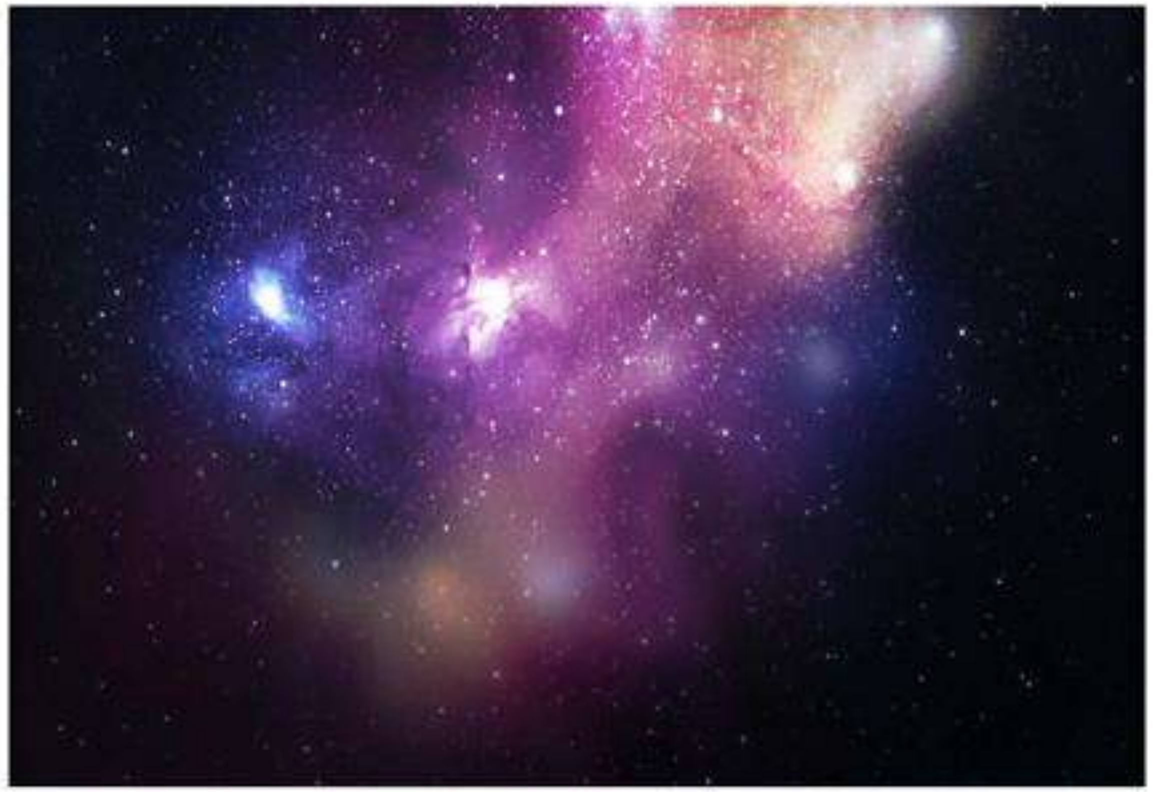
[Ruiz+ 2017]

The probabilistic pipeline



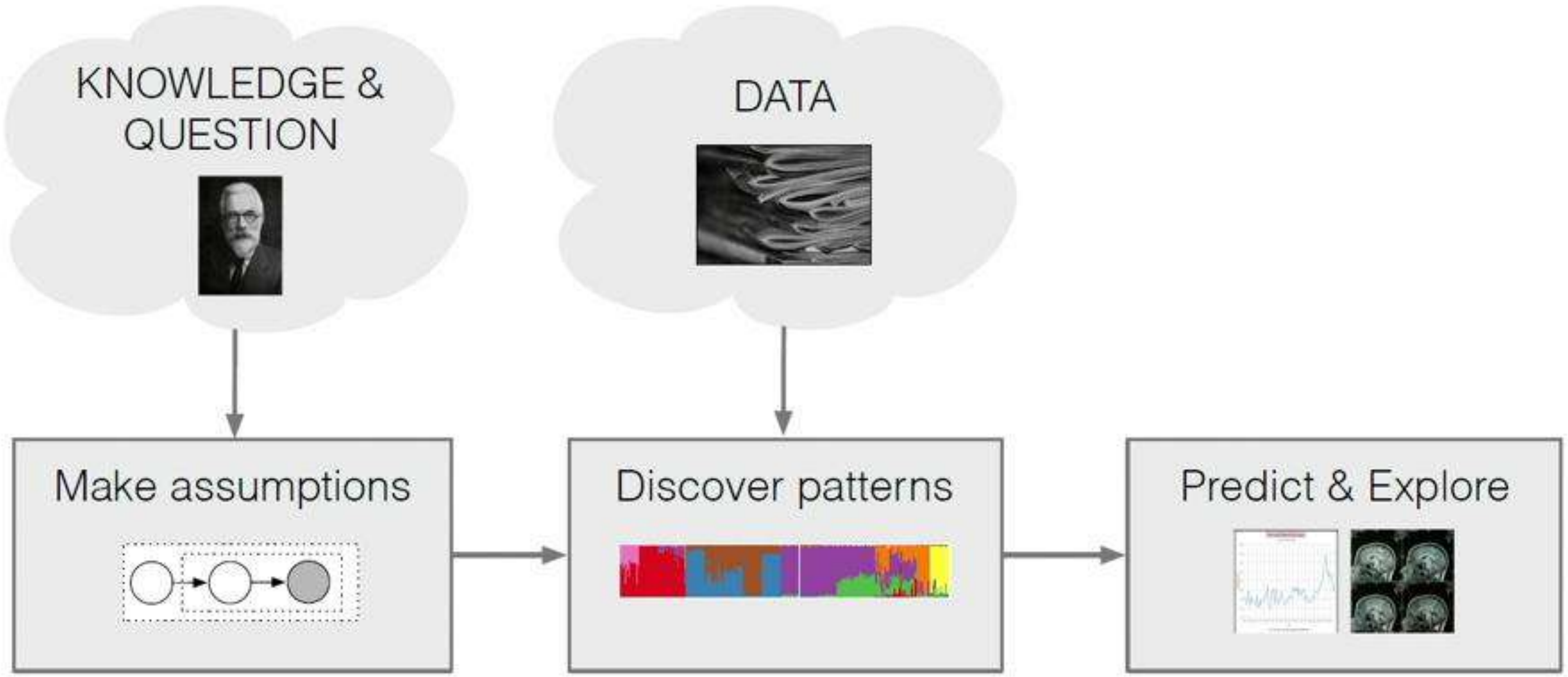
- Customized data analysis is important to many fields.
- Pipeline separates **assumptions**, **computation**, **application**
- Eases collaborative solutions to statistics/ML problems

Causal inference from observational data



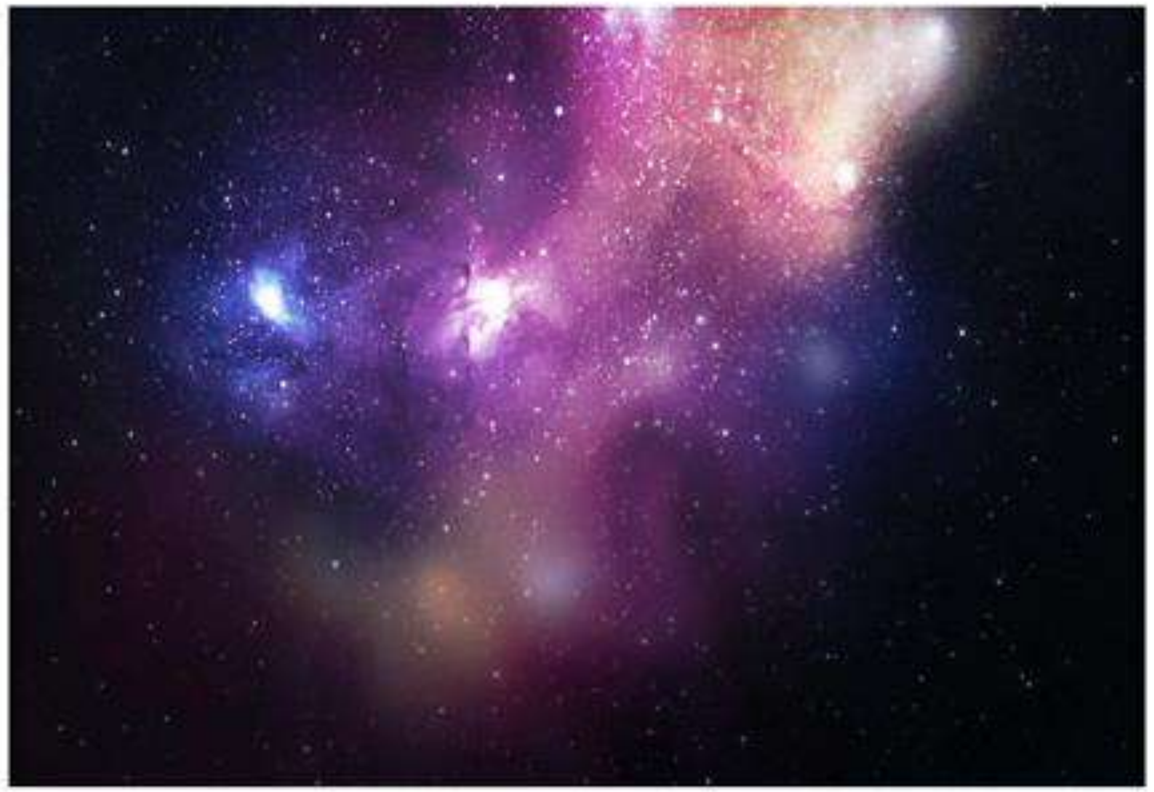
- **How can we understand the world through observation?**
- Important to genetics, economics, physics, medicine, finance, ...
- Today: Use probabilistic machine learning for causal inference

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

Credit

- This is joint work with Yixin Wang (Statistics)
- Credit $\rightarrow$ Yixin
- (Blame $\rightarrow$ 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

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

- 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

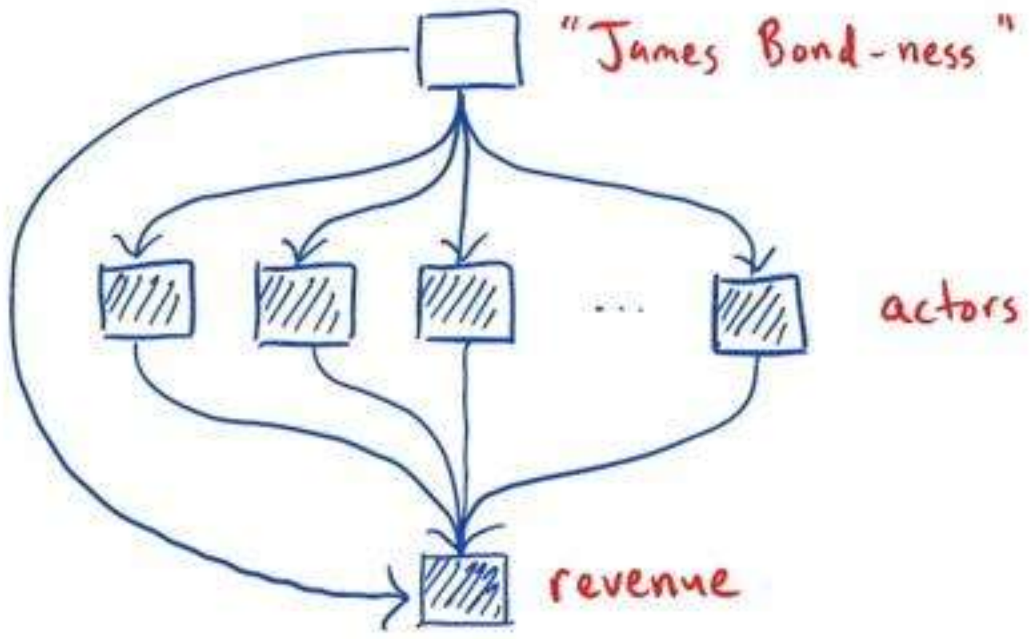
Title	Cast	Revenue
Avatar	{Sam Worthington, Zoe Saldana, Sigourney Weaver, Stephen Lang, ... }	\$2788M
Titanic	{Kate Winslet, Leonardo DiCaprio, Frances Fisher, Billy Zane, ... }	\$1845M
The Avengers	{Robert Downey Jr., Chris Evans, Mark Ruffalo, Chris Hemsworth, ... }	\$1520M
Jurassic World	{Chris Pratt,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

- 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**

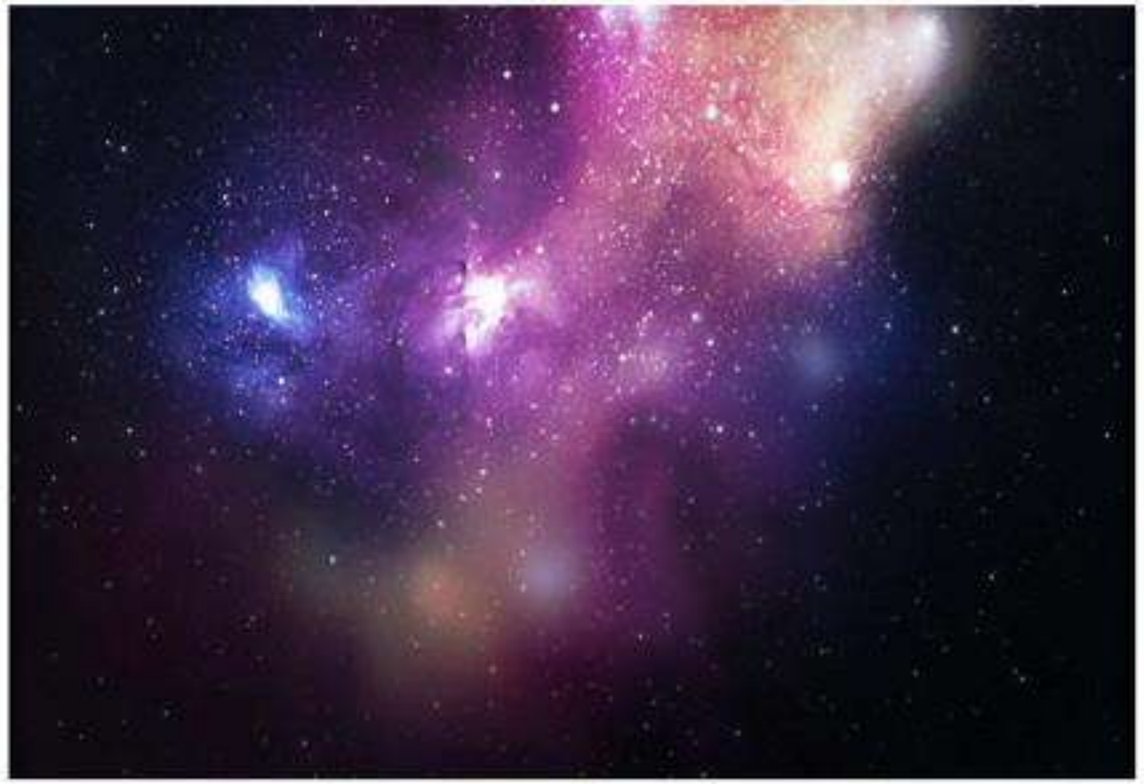

Metro Goldwyn Mayer
TRADE
MARK


Metro Goldwyn Mayer
TRADE
MARK
- 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

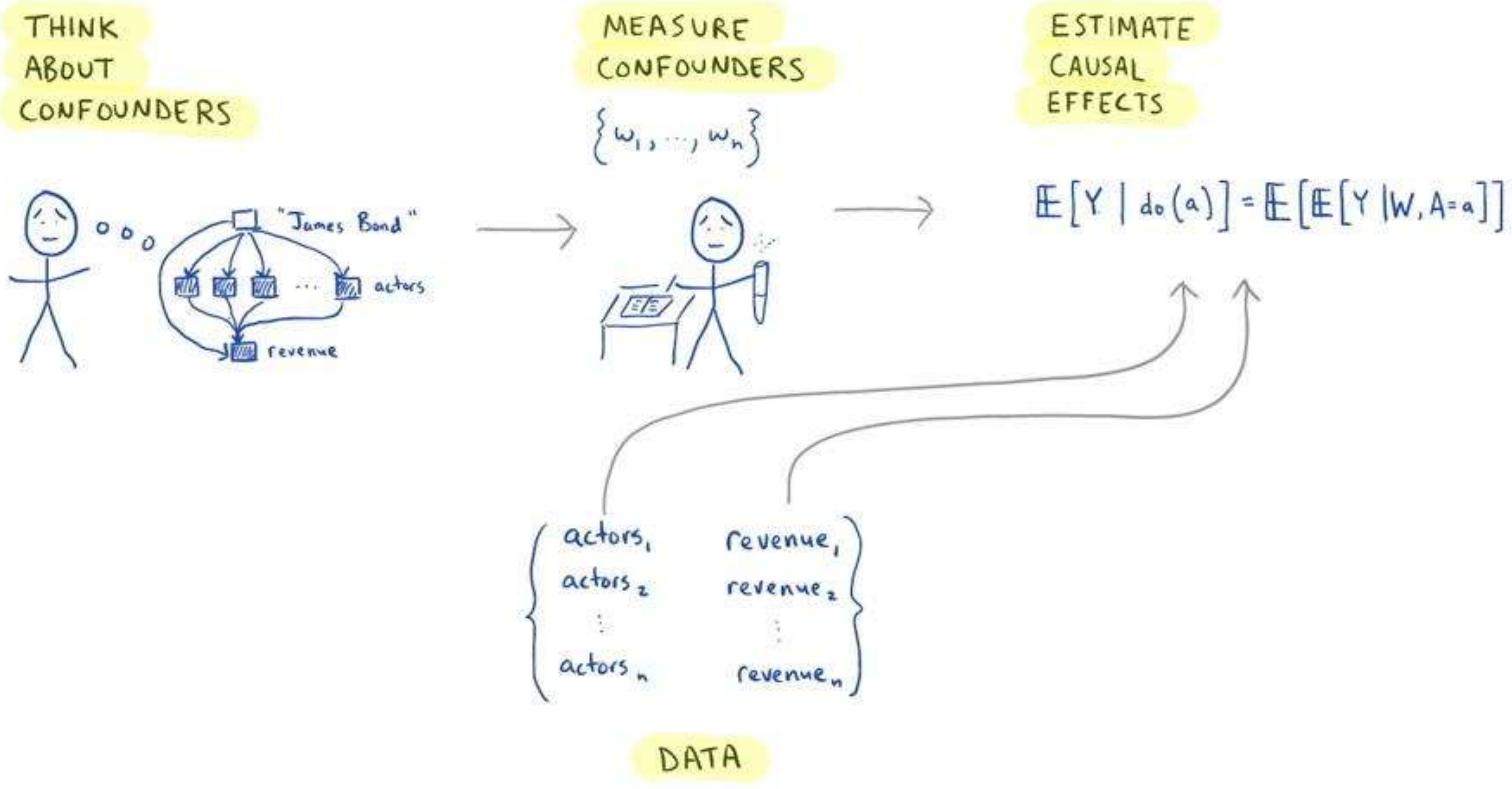



Unobserved confounders are everywhere.

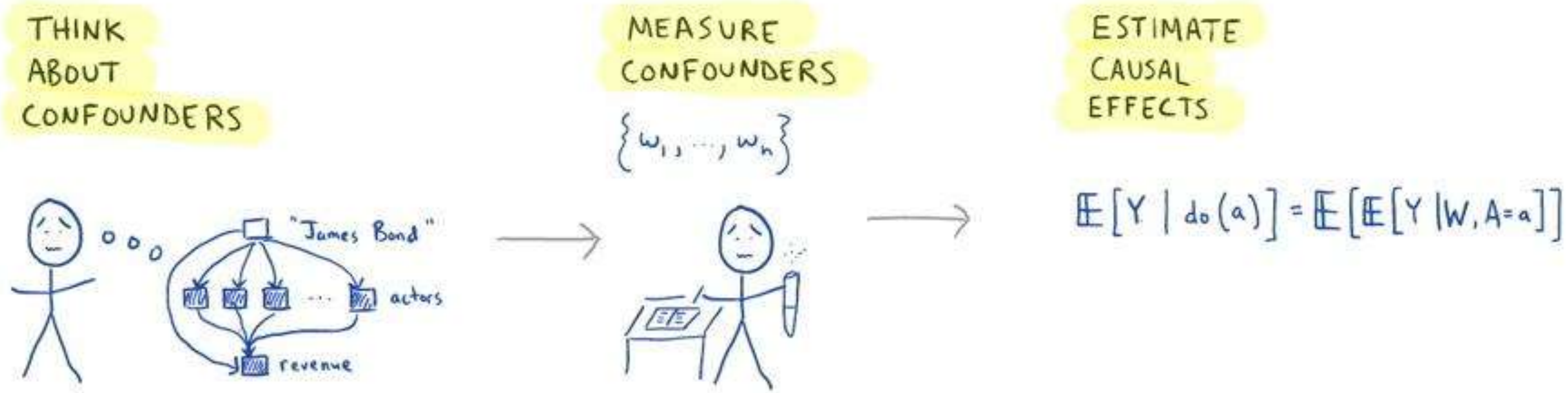
What is causal inference?

- 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



The classical solution



- 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**.)

Multiple causal inference

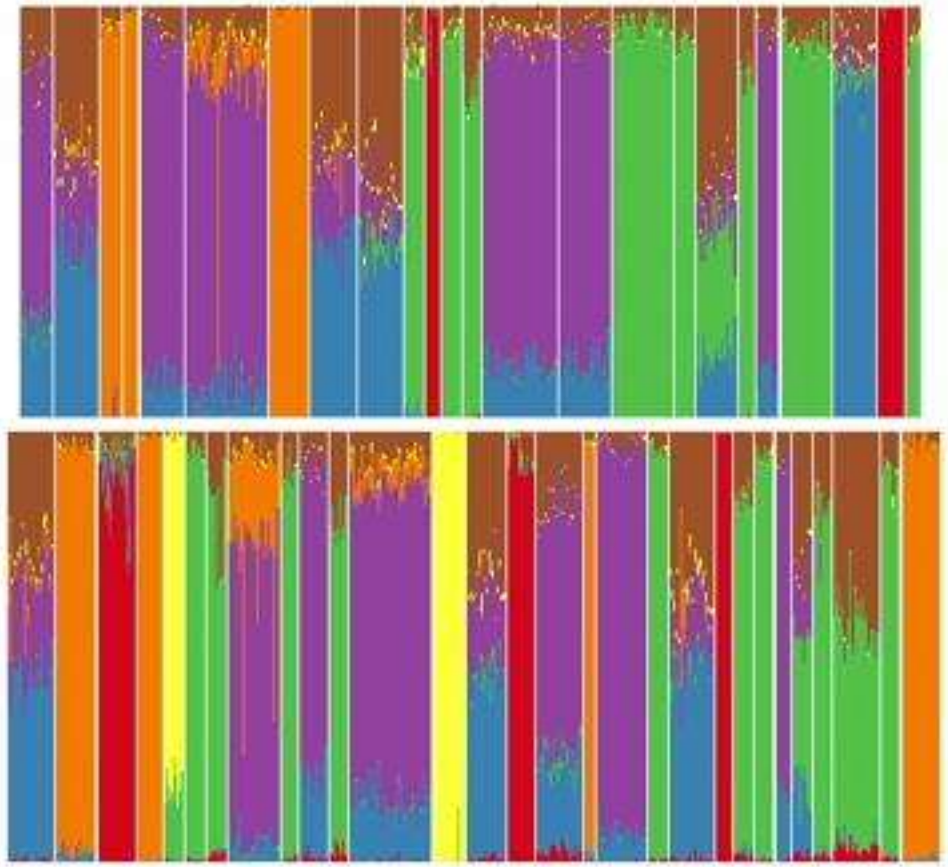
Title	Cast	Revenue
Avatar	{Sam Worthington, Zoe Saldana, Sigourney Weaver, Stephen Lang, ... }	\$2788M
Titanic	{Kate Winslet, Leonardo DiCaprio, Frances Fisher, Billy Zane, ... }	\$1845M
The Avengers	{Robert Downey Jr., Chris Evans, Mark Ruffalo, Chris Hemsworth, ... }	\$1520M
Jurassic World	{Chris Pratt,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

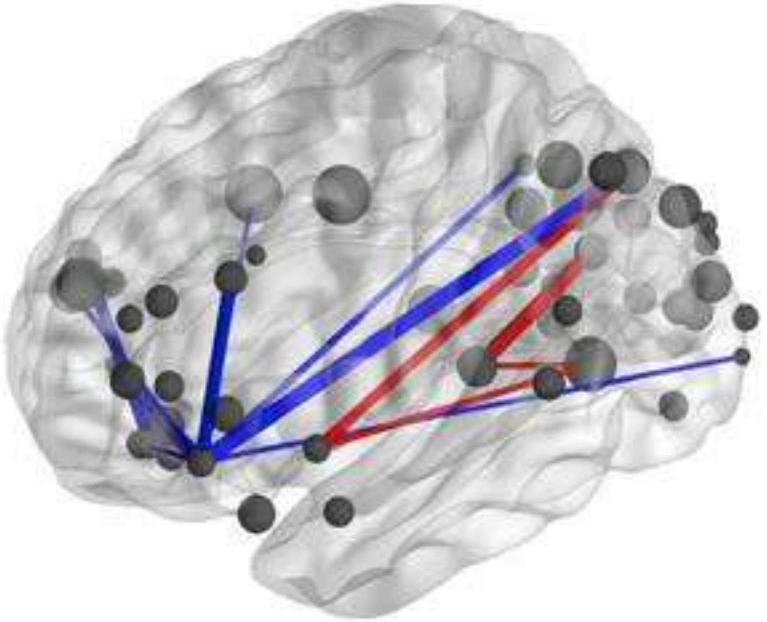
- But our problem is not classical.
- There are many causes (one per actor)—**multiple causal inference**
- ***Multiple causes helps construct a variable that contains confounders.***

1	2	3	4	5
Game	Life	Film	Book	Wine
Season	Know	Movie	Life	Street
Team	School	Show	Books	Hotel
Coach	Street	Life	Novel	House
Play	Man	Television	Story	Room
Points	Family	Films	Man	Night
Games	Says	Director	Author	Place
Giants	House	Man	House	Restaurant
Second	Children	Story	War	Park
Players	Night	Says	Children	Garden

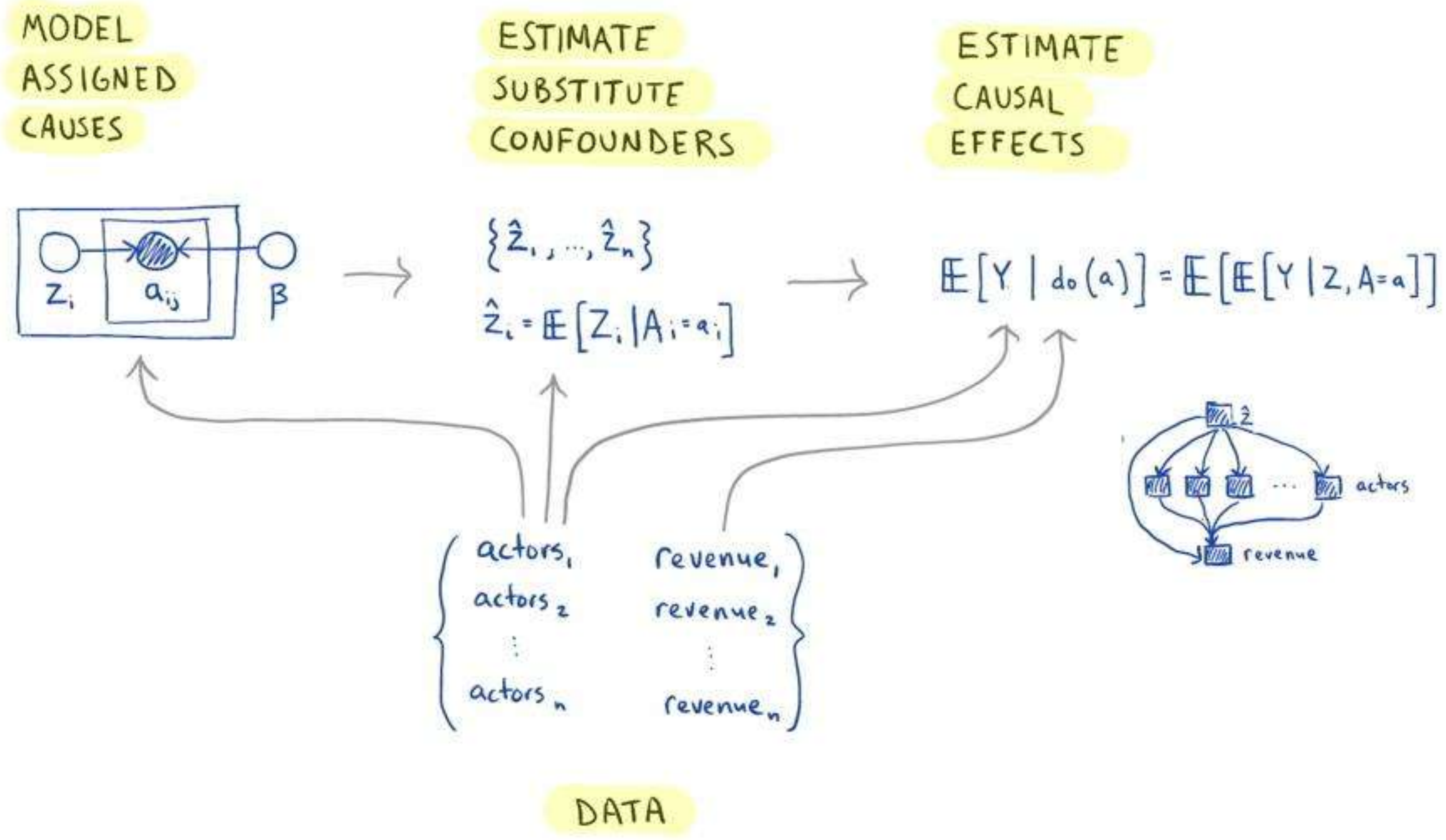
6	7	8	9	10
Bush	Building	Won	Yankees	Government
Campaign	Street	Team	Game	War
Clinton	Square	Second	Mets	Military
Republican	Housing	Race	Season	Officials
House	House	Round	Run	Iraq
Party	Buildings	Cup	League	Forces
Democratic	Development	Open	Baseball	Iraqi
Political	Space	Game	Team	Army
Democrats	Percent	Play	Games	Troops
Senator	Real	Win	Hit	Soldiers

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

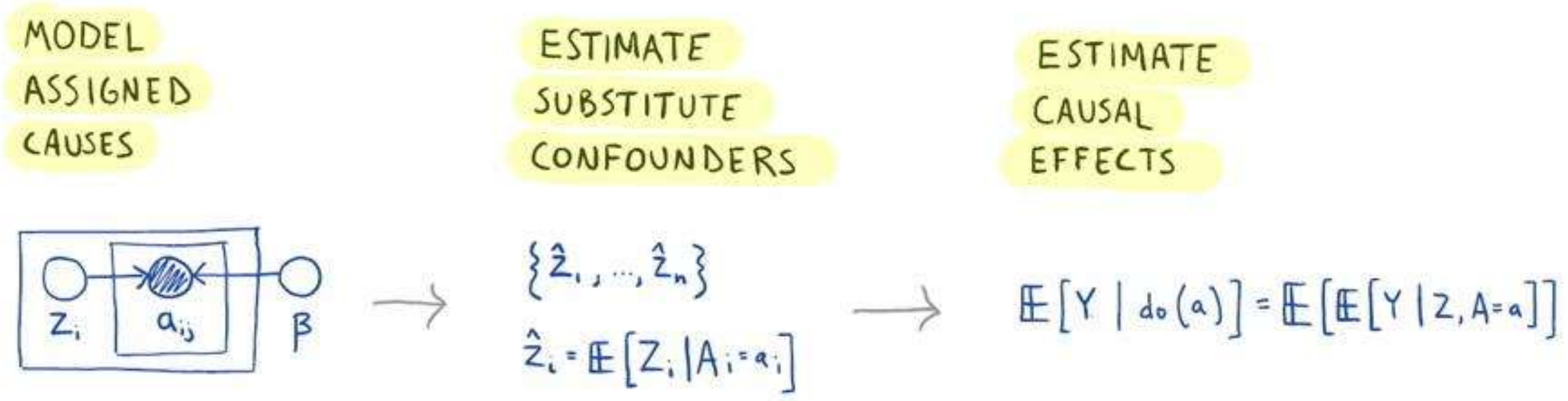




The deconfounder

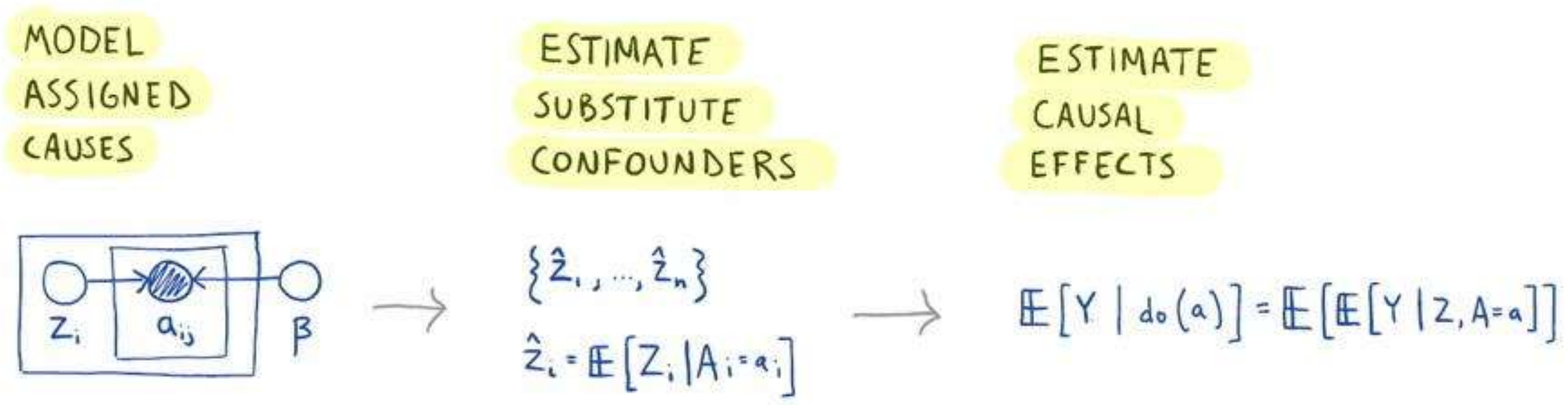


The deconfounder



- 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**.


Metro Goldwyn Mayer
TRADE
MARK

Case study: Actors

- "Overestimated":



- "Underestimated":

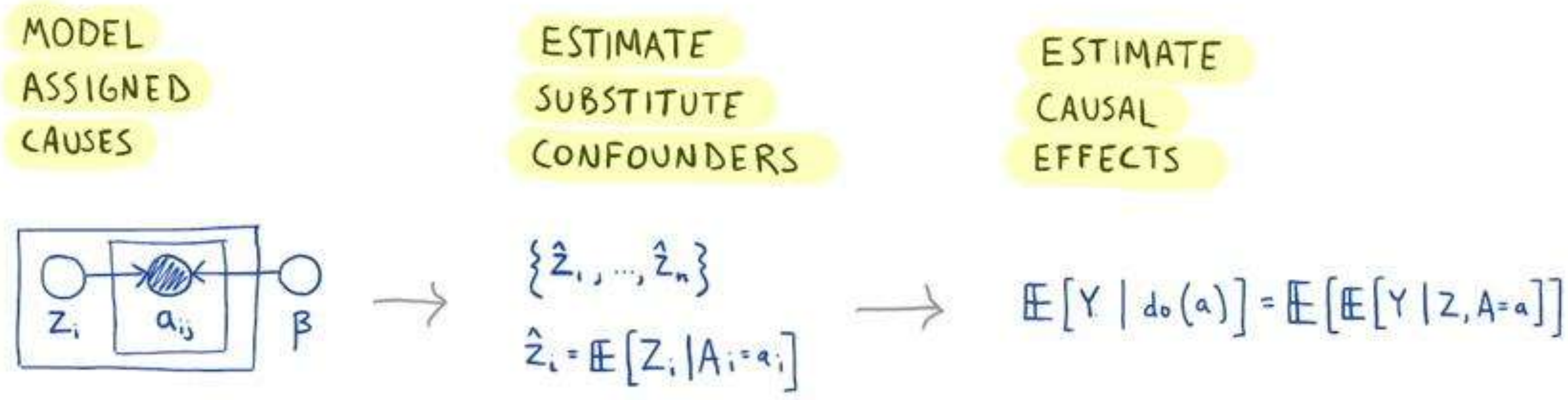



- Most "corrected":





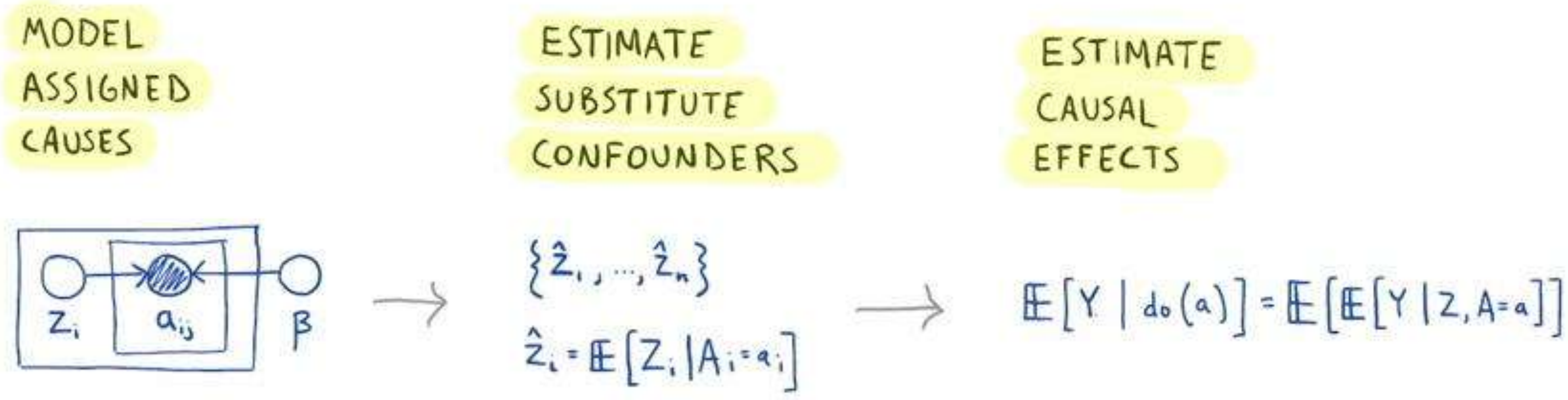
Intuition and assumptions



- 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”


Metro Goldwyn Mayer
TRADE
MARK

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”

Beyond James Bond

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.*

Beyond James Bond

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.

Beyond James Bond

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

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

- Observed dataset $\mathcal{D} = \{(\mathbf{a}_1, y_1), \ldots, (\mathbf{a}_n, y_n)\}$
  - assigned causes $\mathbf{a}_i = \{a_{i1}, \ldots, a_{im}\}$
  - outcome y_i
- Goal: Do causal inference, $\mathbb{E}[Y; \mathrm{do}(\mathbf{a})]$
  - "The expectation of Y in the model where we intervened on $\mathbf{a}$."

Multiple causal inference

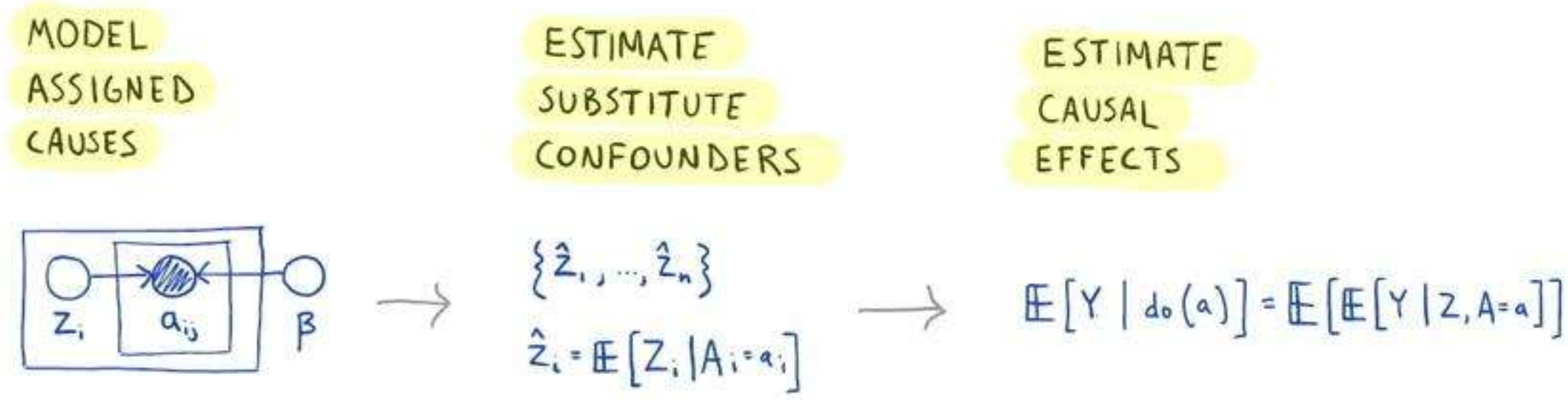
Title	Cast	Revenue
Avatar	{Sam Worthington, Zoe Saldana, Sigourney Weaver, Stephen Lang, ... }	\$2788M
Titanic	{Kate Winslet, Leonardo DiCaprio, Frances Fisher, Billy Zane, ... }	\$1845M
The Avengers	{Robert Downey Jr., Chris Evans, Mark Ruffalo, Chris Hemsworth, ... }	\$1520M
Jurassic World	{Chris Pratt,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

- If there are unobserved confounders then

$$\mathbb{E}[Y\,;\mathrm{do}(\mathbf{a})] \neq \mathbb{E}[Y \mid A = \mathbf{a}].$$

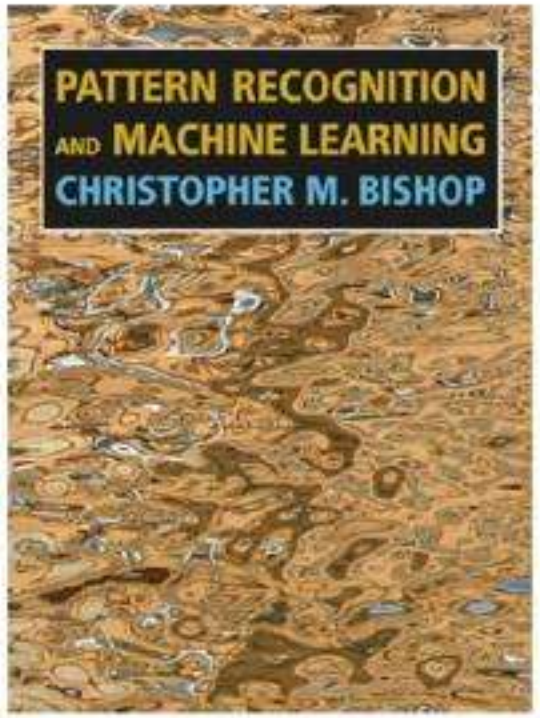
- We can calculate the right term from data, but it's not equal to the left term.

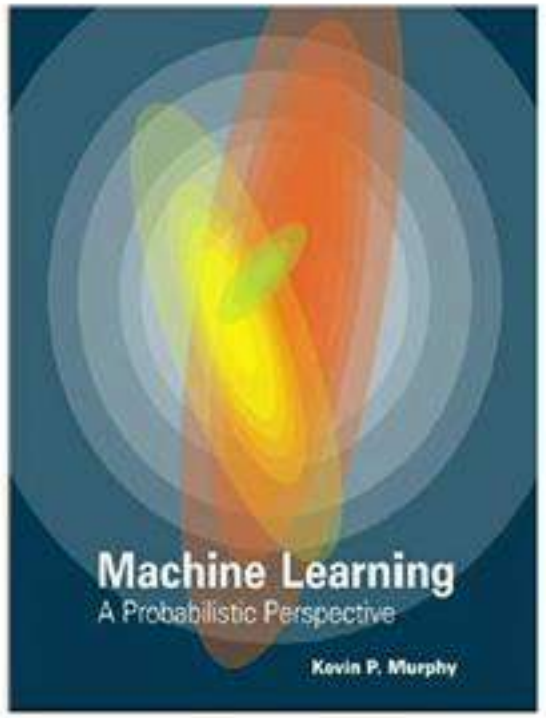
The deconfounder

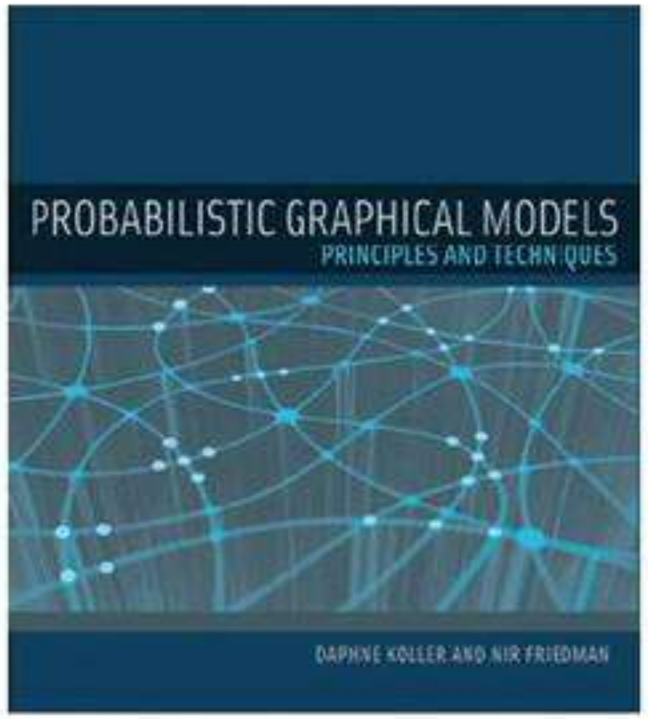


- Find, fit, and check a **factor model** of the movie casts.
- Use the factor model to form **substitute confounders** for each movie.
- Use the substitute confounders in a **causal model** of movie revenue.

Fit a probabilistic factor model





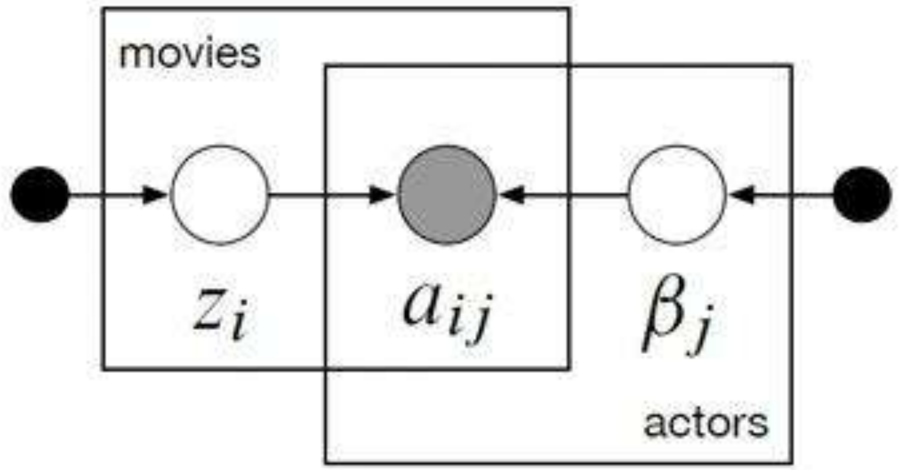


- A probabilistic factor model has the following form,

$$\begin{aligned} \beta_j &\sim p(\beta_j) & j &= 1, \ldots, m \\ z_i &\sim p(z_i) & i &= 1, \ldots, n \\ a_{ij} &\sim p(a_{ij} \mid z_i, \beta_j). \end{aligned}$$

- E.g., mixtures, matrix factorization, deep generative models, topic models, ...

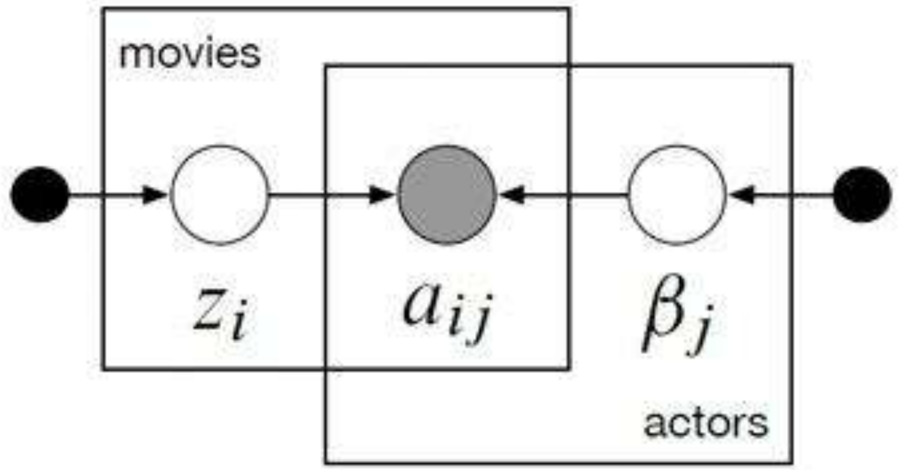
Poisson factorization [Gopalan+ 2015]



$$\beta_{jk} \sim \mathrm{Gam}(a, b) \qquad i \in \{1, \ldots n\}$$
$$z_{ik} \sim \mathrm{Gam}(a, b) \qquad j \in \{1, \ldots, m\}$$
$$a_{ij} \sim \mathrm{Poi}(z_i^\top \beta_j) \qquad k \in \{1, \ldots, d\}$$

- 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]

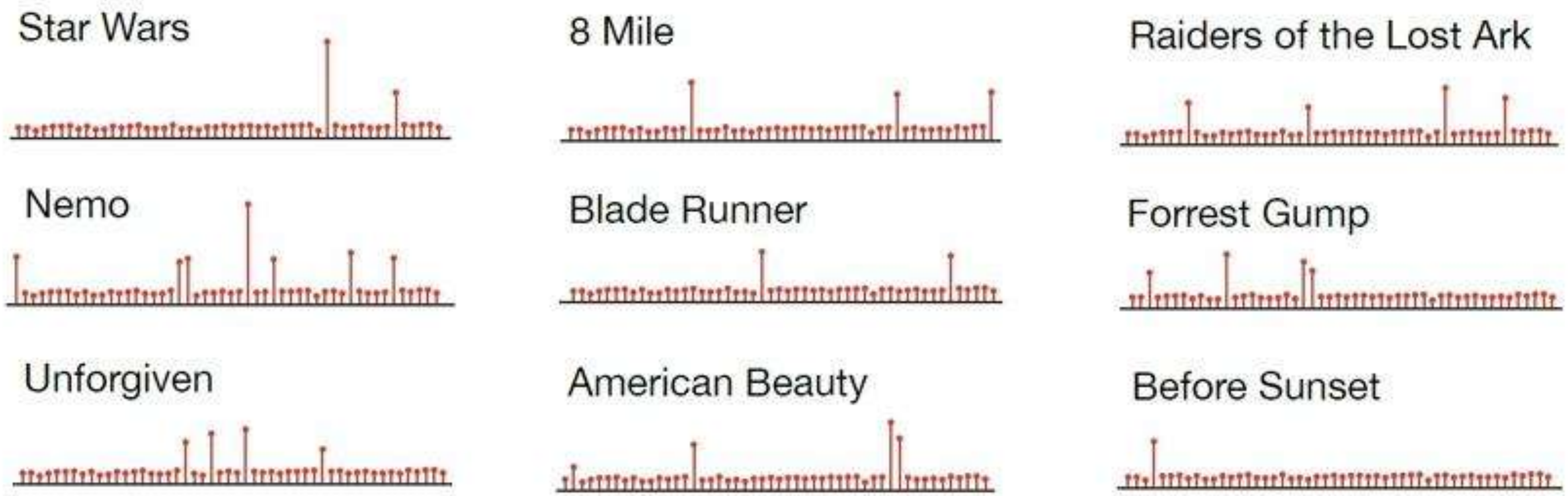
Poisson factorization [Gopalan+ 2015]



$$
\begin{aligned}
\beta_{jk} &\sim \text{Gam}(a, b) & \quad i &\in \{1, \dots n\} \\
z_{ik} &\sim \text{Gam}(a, b) & \quad j &\in \{1, \dots, m\} \\
a_{ij} &\sim \text{Poi}(z_i^\top \beta_j) & \quad k &\in \{1, \dots, d\}
\end{aligned}
$$

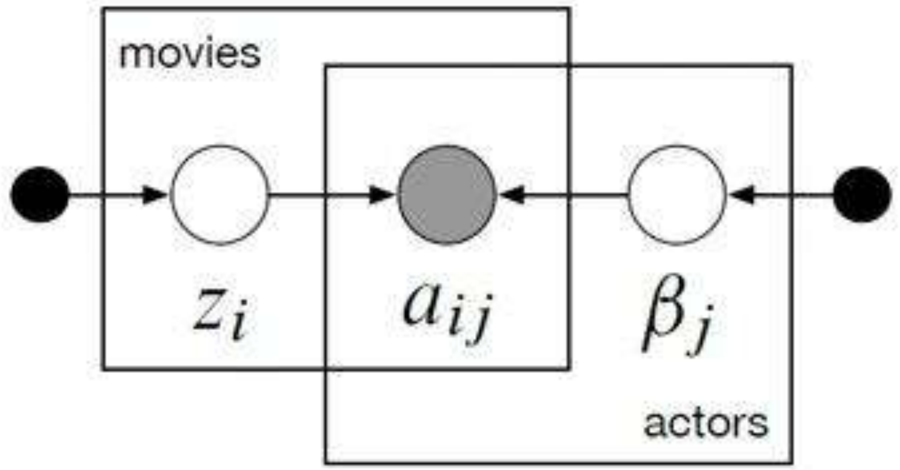
- Consider the dataset of casts $\mathbf{a}_{1:n}$.
- Approximate the posterior distribution $p(z_{1:n}, \beta_{1:m} \mid \mathbf{a}_{1:n})$.
- **We only model the actors a_i; the outcome is not involved.**

Check the factor model



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

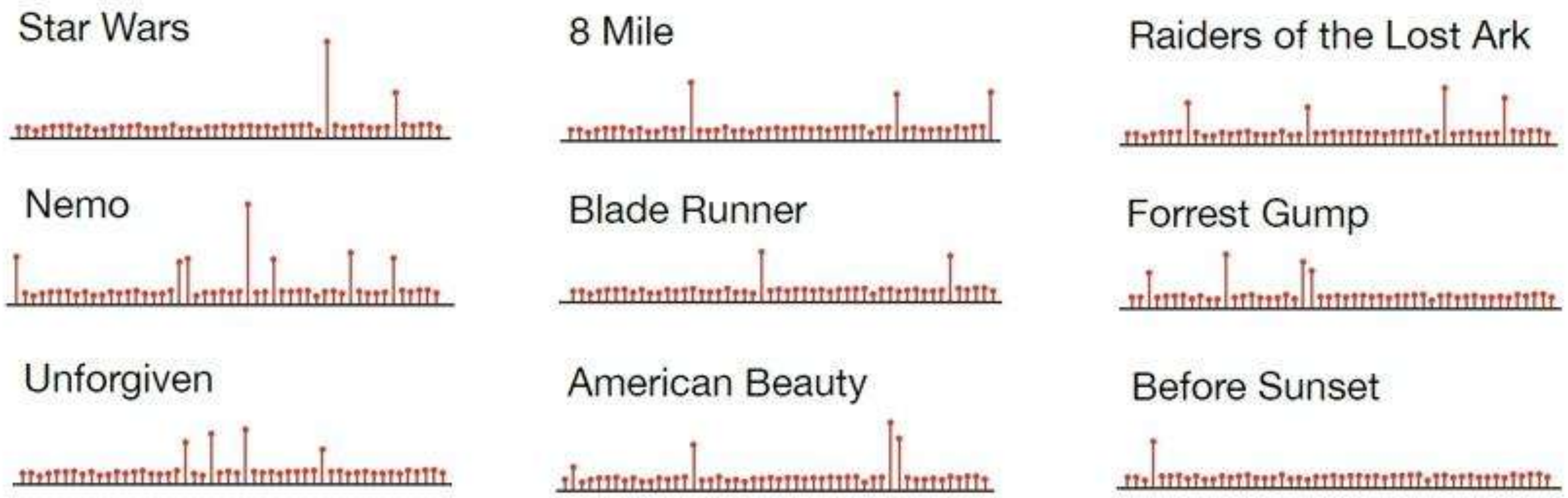
Poisson factorization [Gopalan+ 2015]



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

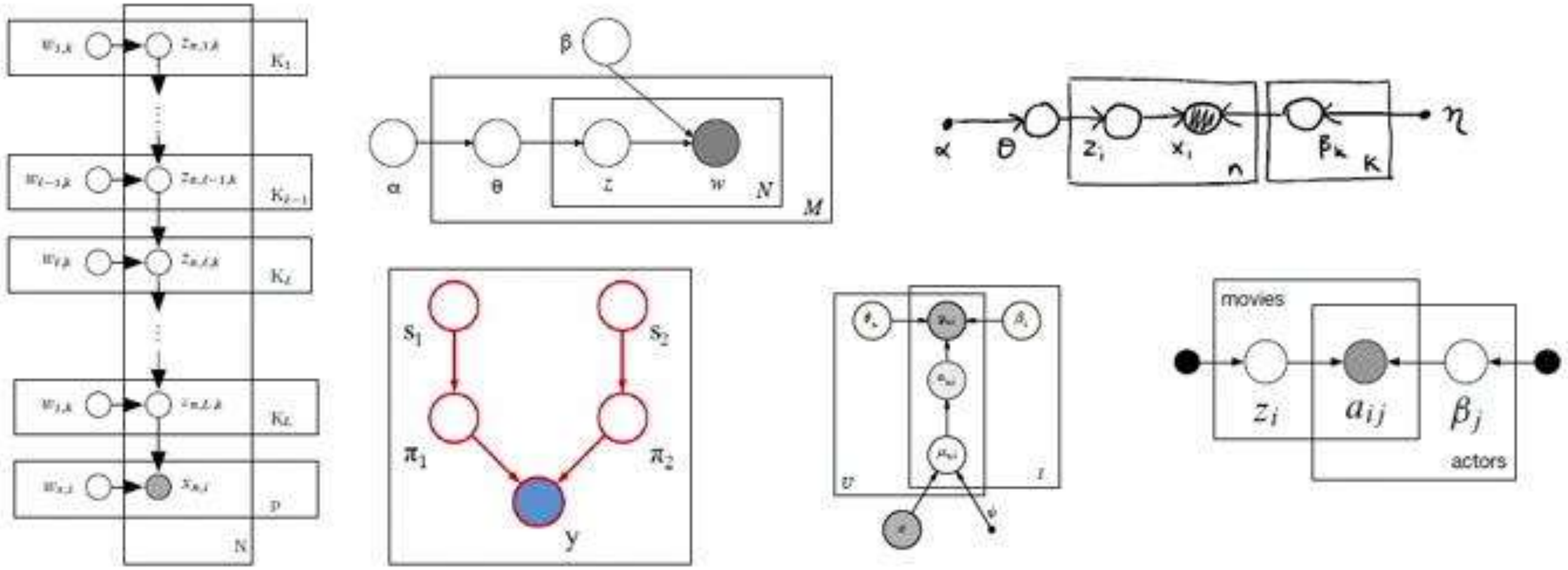
- Provides a generative model of the assigned causes a_{ij}.
- Can be approximated on large datasets with variational methods
- A Bayesian form of non-negative matrix factorization [Lee and Seung 1999]

Check the factor model



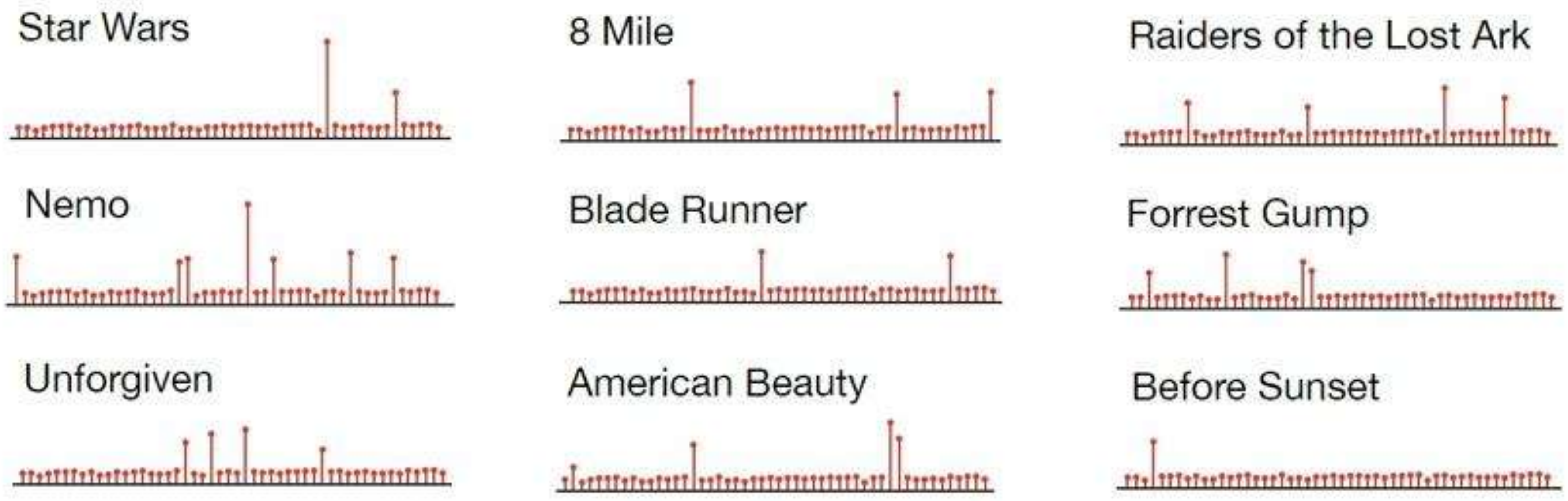
- Estimate the local latent variable $\hat{z}_i = \mathbb{E}_{\text{model}}[Z \mid \mathbf{a}_i, \boldsymbol{\beta}]$.
- Check how well $\hat{z}_i$ captures the distribution of the actors.
- E.g., use a **predictive check** on actors. (No need for exact inference.)

Check the factor model



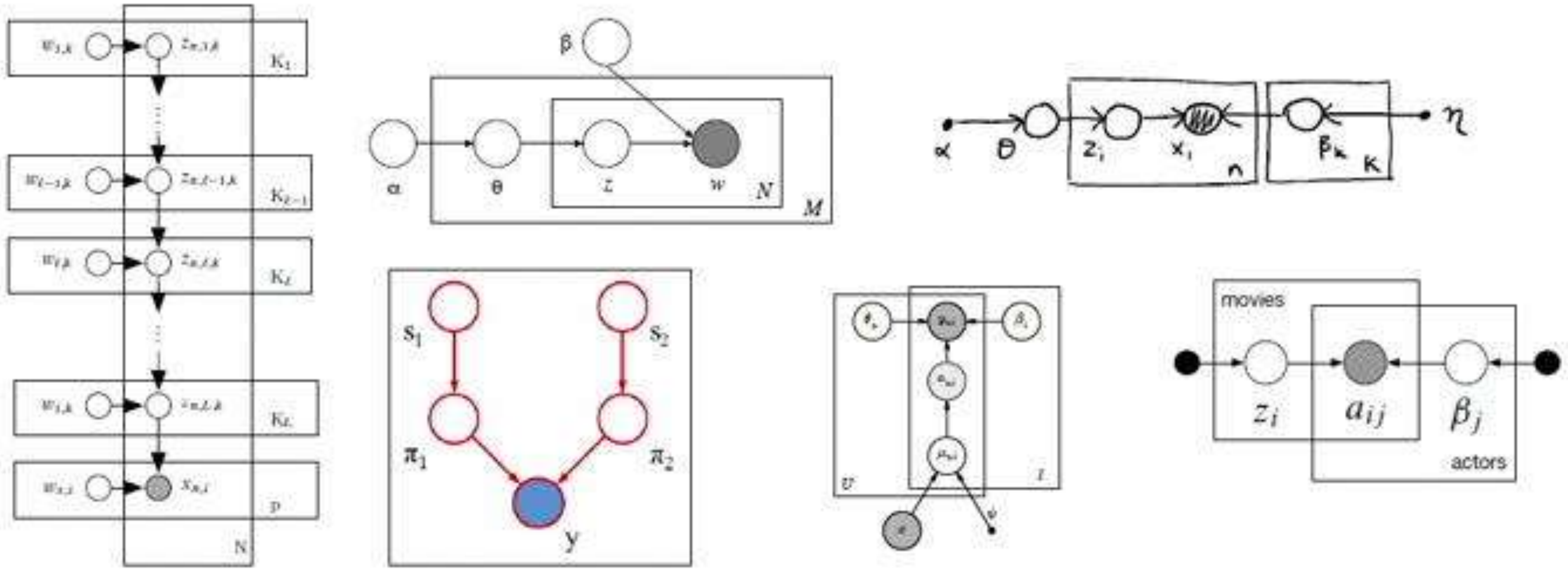
Model	Predictive score
Probabilistic PCA	0.14
Poisson factorization	0.16
Mixtures	0.01
Deep exponential families	0.19

Check the factor model



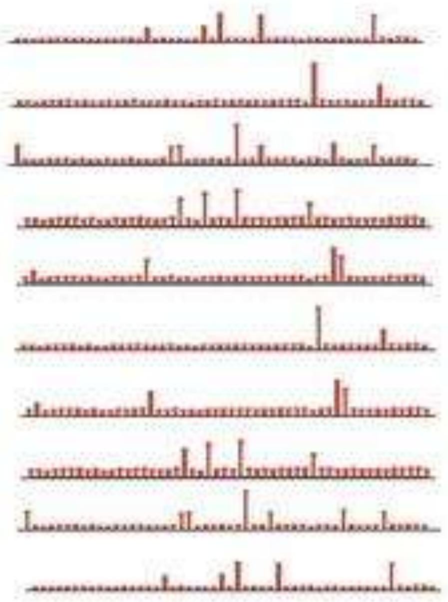
- Estimate the local latent variable $\hat{z}_i = \mathbb{E}_{\text{model}}[Z \mid \mathbf{a}_i, \boldsymbol{\beta}]$.
- Check how well $\hat{z}_i$ captures the distribution of the actors.
- E.g., use a **predictive check** on actors. (No need for exact inference.)

Check the factor model



Model	Predictive score
Probabilistic PCA	0.14
Poisson factorization	0.16
Mixtures	0.01
Deep exponential families	0.19

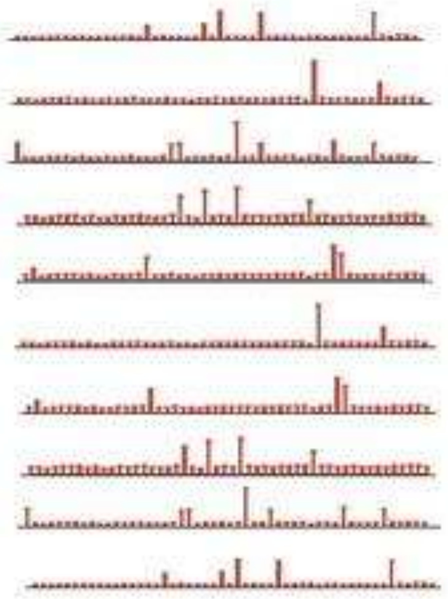
Do causal inference



{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 $(\mathbf{a}_i, y_i, \hat{z}_i)$.

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

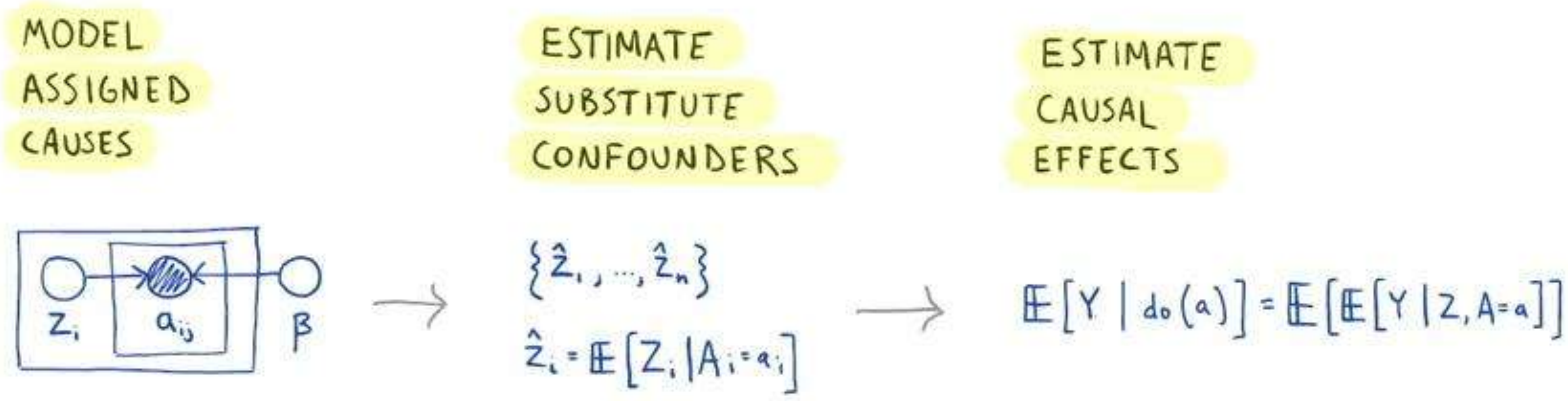
- Use the substitute confounders in a **causal inference**.
- E.g., fit regression from casts and confounders to revenue,

$$\mathbb{E}[Y \mid \mathbf{a}, \hat{z}] = \beta^\top \mathbf{a} + \eta^\top \hat{z}.$$

- Use adjustment/the g-formula to perform causal inference,

$$\mathbb{E}[Y \,;\, \mathrm{do}(\mathbf{a})] \approx \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[Y \mid \mathbf{a}, \hat{Z}].$$

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":



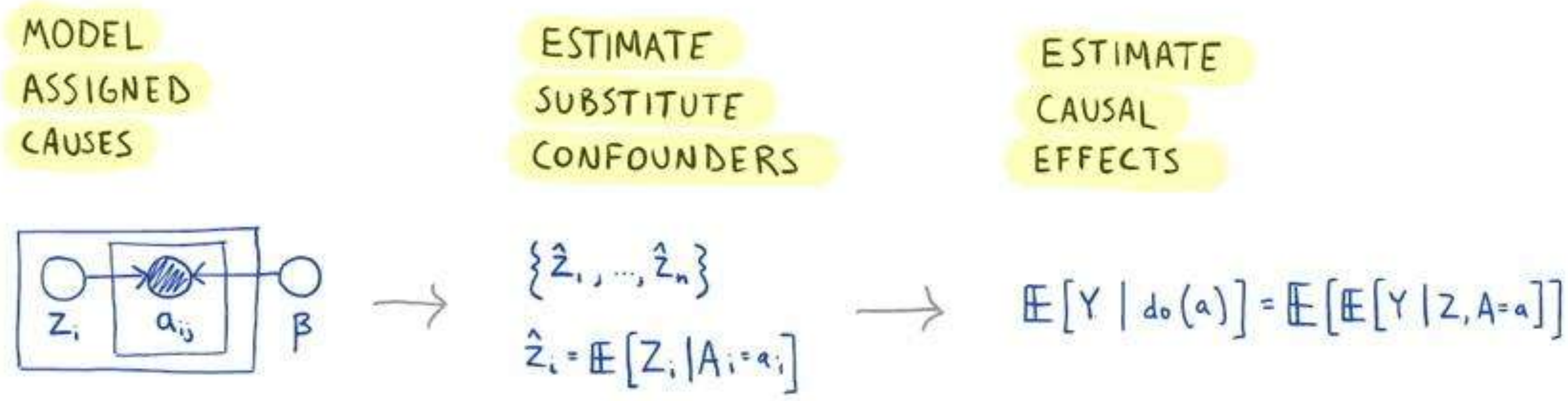

- Most "corrected":





A little theory

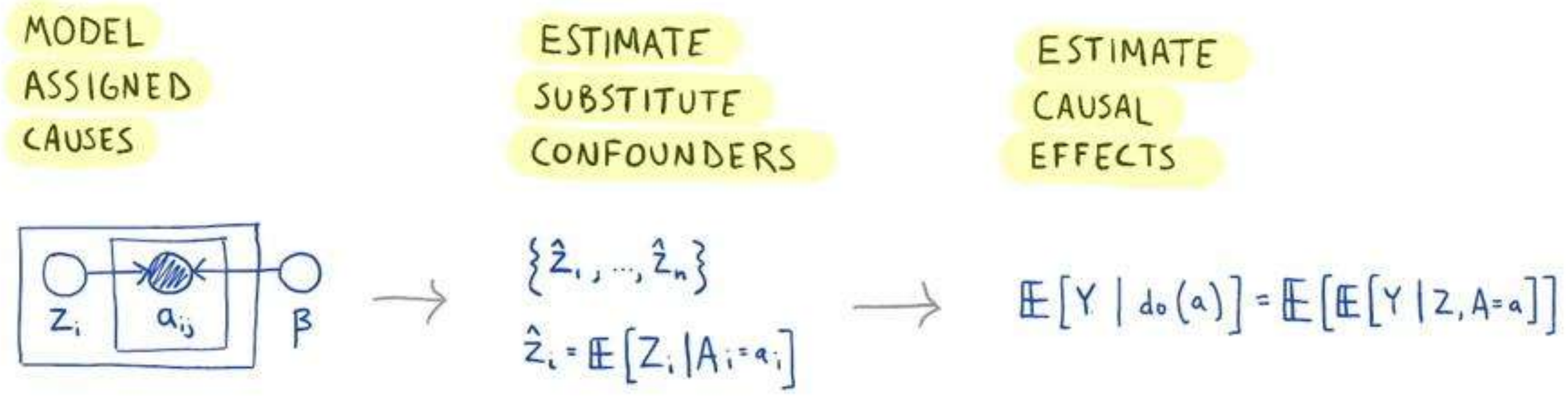
The deconfounder



- 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.

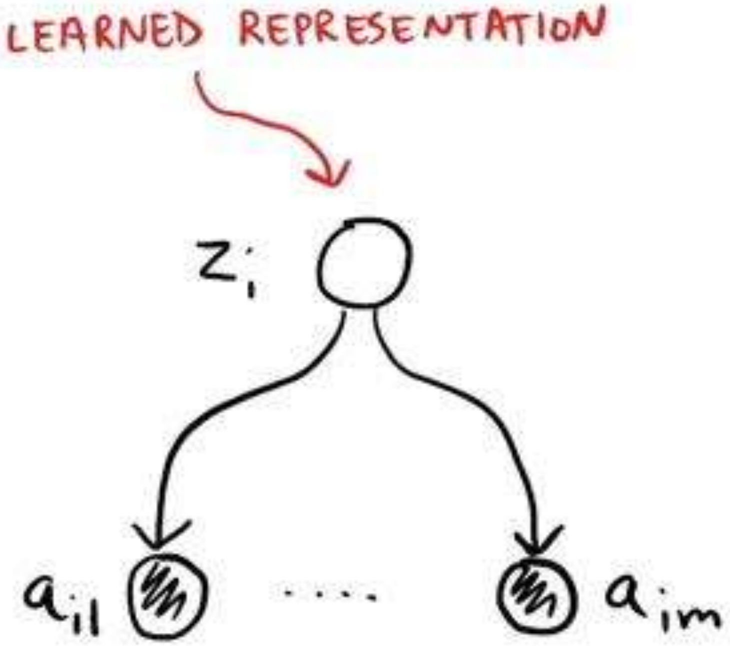
A little theory

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.

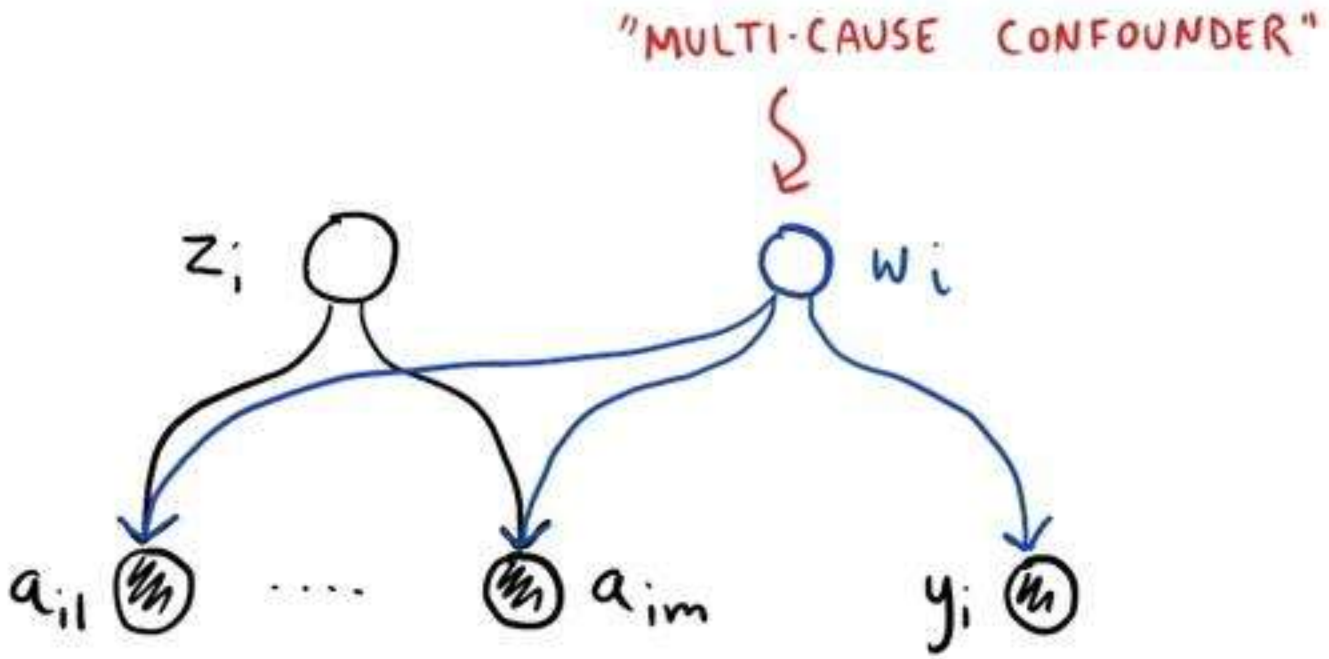
Intuition (through graphical models)



If we find a good factor model then

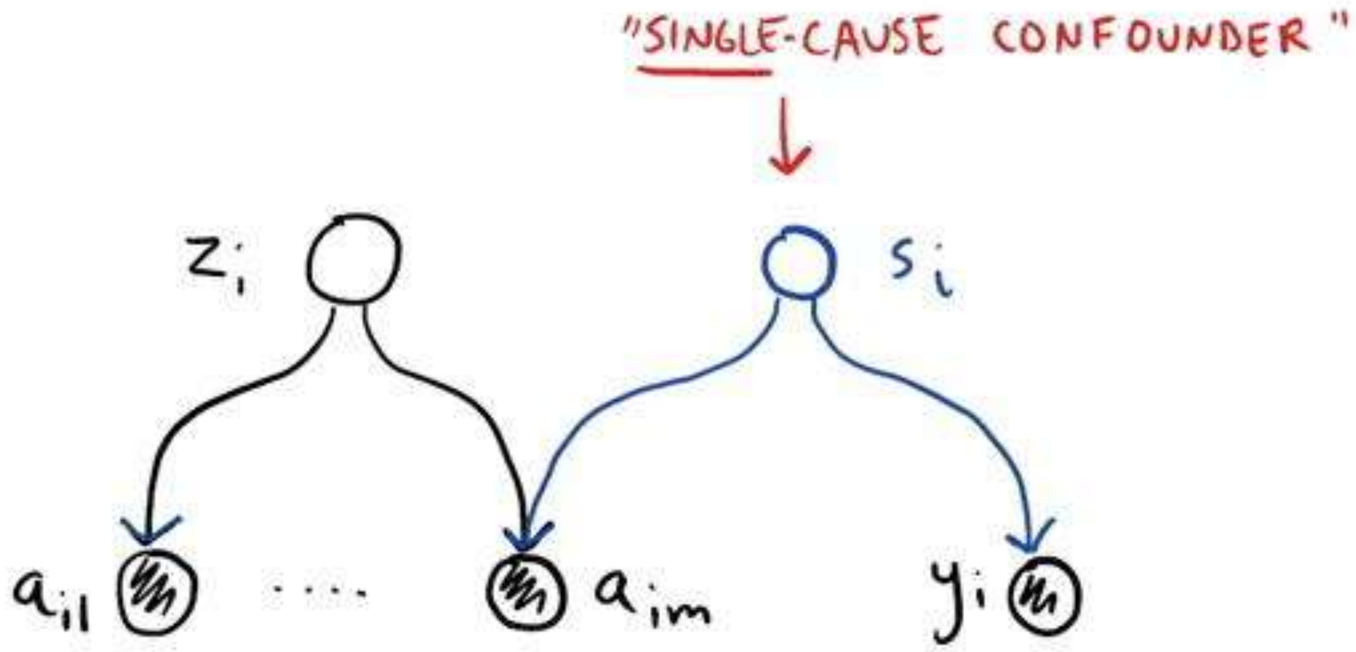
$$p(a_{i1}, \ldots, a_{im} \mid z_i, \beta_{1:m}) = \prod_{j=1}^{m} p(a_{ij} \mid z_i, \beta_j)$$

Intuition (through graphical models)



- There cannot be an unobserved **multi-cause confounder**.
- Contradiction: If one existed then the independence statement would not hold.

Intution (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 $p_{\text{true}}(\mathbf{a})$ can be written $\int p(z) \prod_j p(a_j \mid z, \boldsymbol{\beta}) dz$.

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

This implies that,

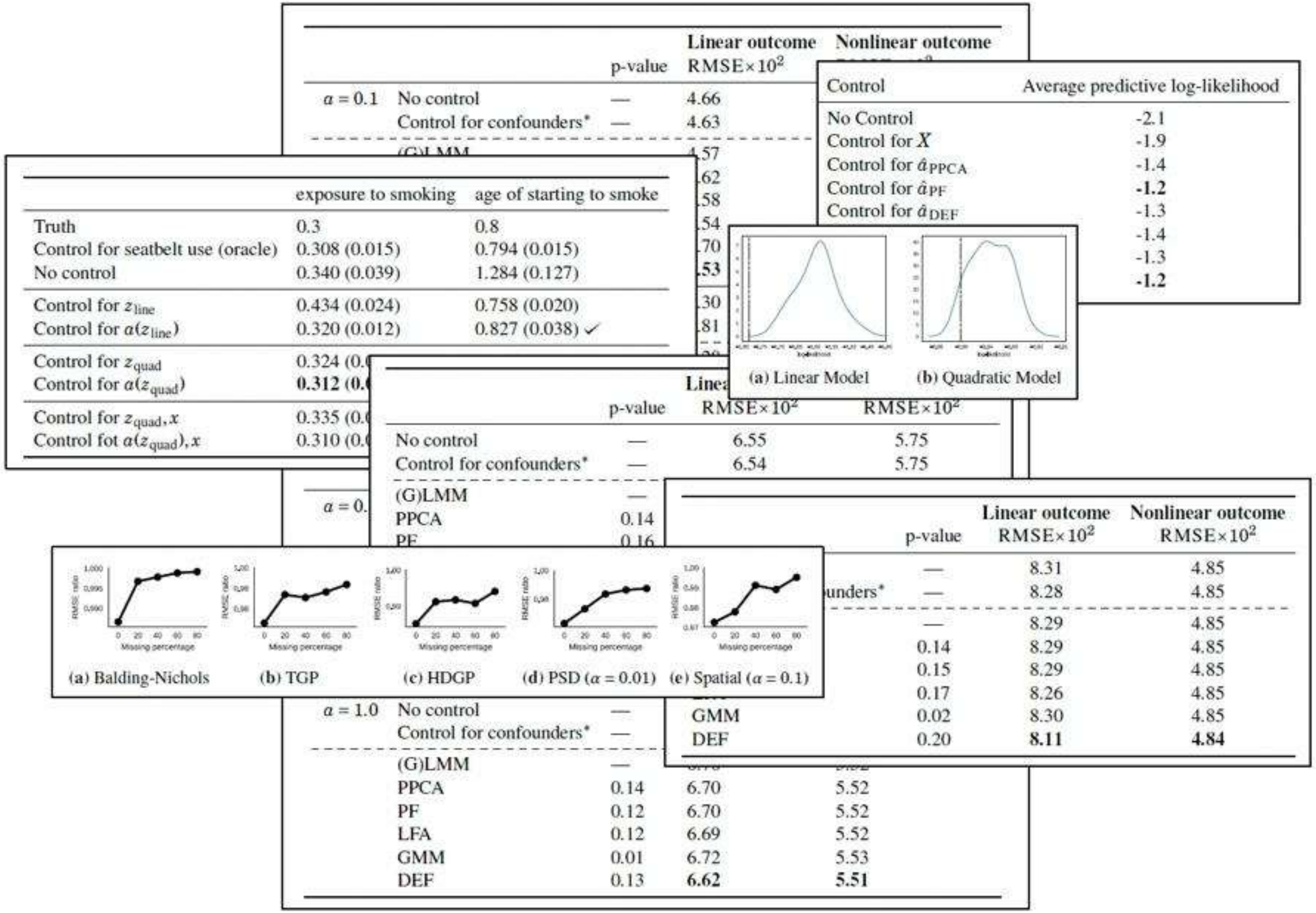
$$\mathbb{E}\left[Y\, ;\, \mathrm{do}(\mathbf{a})\right] = \mathbb{E}_Z\left[\mathbb{E}_Y\left[Y \mid Z, \mathbf{a}\right]\right].$$

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



Example: Genome-wide association studies (GWAS)

- 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

ID (i)	SNP_1 ($a_{i,1}$)	SNP_2 ($a_{i,2}$)	SNP_3 ($a_{i,3}$)	SNP_4 ($a_{i,4}$)	SNP_5 ($a_{i,5}$)	SNP_6 ($a_{i,6}$)	SNP_7 ($a_{i,7}$)	SNP_8 ($a_{i,8}$)	SNP_9 ($a_{i,9}$)	$\cdots$	SNP_100K ($a_{i,100K}$)	Height (feet) (y_i)
1	1	0	0	1	0	0	1	2	0	$\cdots$	0	5.73
2	1	2	2	1	2	1	1	0	1	$\cdots$	2	5.26
3	2	0	1	1	0	1	0	1	1	$\cdots$	2	6.24
4	0	0	0	1	1	0	1	2	0	$\cdots$	0	5.78
5	1	2	1	1	1	0	1	0	0	$\cdots$	1	5.09
$\vdots$						$\vdots$						$\vdots$

- Generate SNPs a_{ij}, where each individual belongs to a latent group c_i.
- The true outcome is a trait y_i, drawn from

$$y_i = \sum_j \beta_j a_{ij} + \lambda_{c_i} + \varepsilon_i \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_{c_i}),$$

where many β_j are zero, i.e., non-causal SNPs.

- *Confounded*: the intercept λ_{c_i} and error ε_i are connected to the latent group.

Simulation study

	pred. score	**Real-valued outcome** RMSE$\times 10^2$	**Binary outcome** RMSE$\times 10^2$
No control	—		
Control for confounders*	—		
(G)LMM	—		
PPCA	0.14		
PF	0.15		
LFA	0.14		
Mixture	0.00		
DEF	0.20		

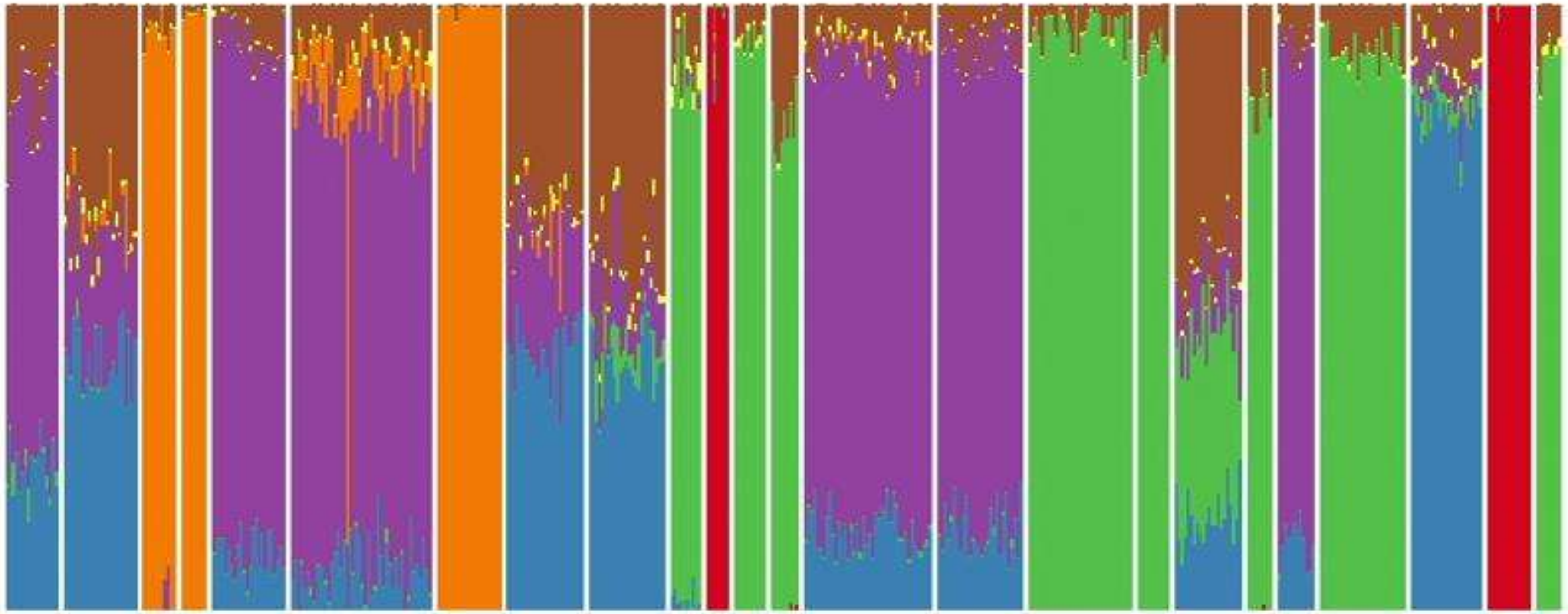
- We fit many factor models; none was the true model.
- Each provides different levels of predictive performance.
- All computation done in Edward [Tran+ 2018] .

Simulation study

	pred. score	**Real-valued outcome** RMSE$\times 10^2$	**Binary outcome** RMSE$\times 10^2$
No control	—	58.82	29.50
Control for confounders*	—	25.32	25.77
(G)LMM	—	35.18	28.87
PPCA	0.14	33.32	26.70
PF	0.15	33.38	26.84
LFA	0.14	33.93	26.83
Mixture	0.00	57.59	29.96
DEF	0.20	**26.47**	**25.91**

- 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



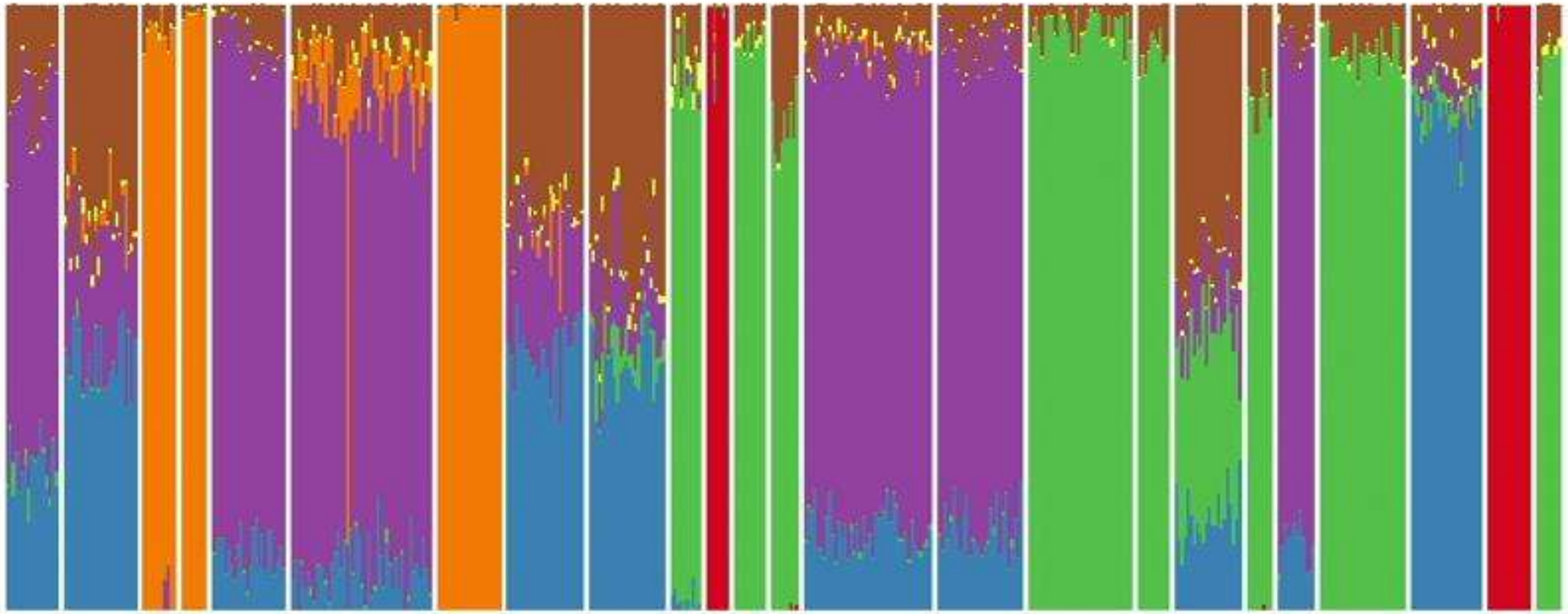
- Linear mixed models [Yu+ 2006; Kang+ 2008; etc.]
- Principal component analysis [Price+ 2006]
- Logistic factor analysis [Song+ 2015; Hao+ 2015]
- Mixed-membership models [Pritchard+ 2000a,b; Falush+ 2003; Falush+ 2007]
- Deep generative models [Tran and Blei 2018]

Simulation study

	pred. score	**Real-valued outcome** RMSE$\times 10^2$	**Binary outcome** RMSE$\times 10^2$
No control	—	58.82	29.50
Control for confounders*	—	25.32	25.77
(G)LMM	—	35.18	28.87
PPCA	0.14	33.32	26.70
PF	0.15	33.38	26.84
LFA	0.14	33.93	26.83
Mixture	0.00	57.59	29.96
DEF	0.20	**26.47**	**25.91**

- Also fit outcome models with no control and with observed confounders
- The deconfounder provides good causal estimates.
- Predictive checks indicate downstream causal performance.

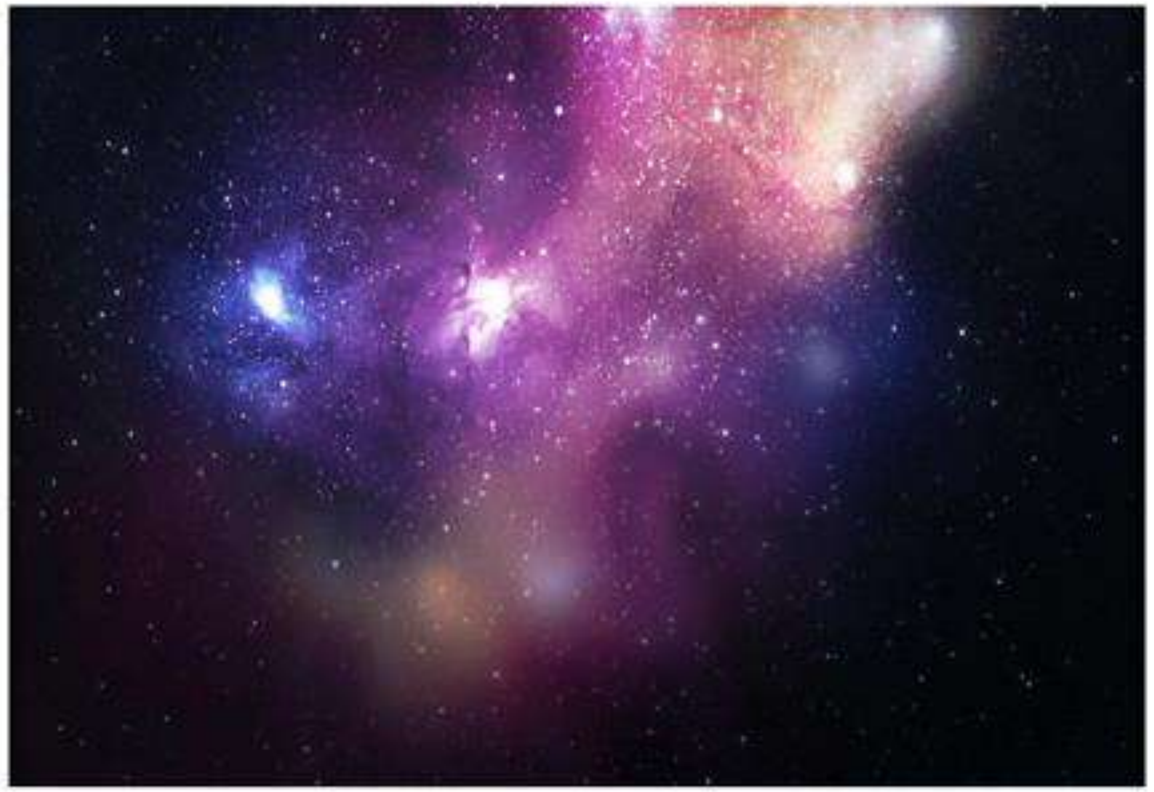
Explains and justifies existing methods for GWAS



- 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]

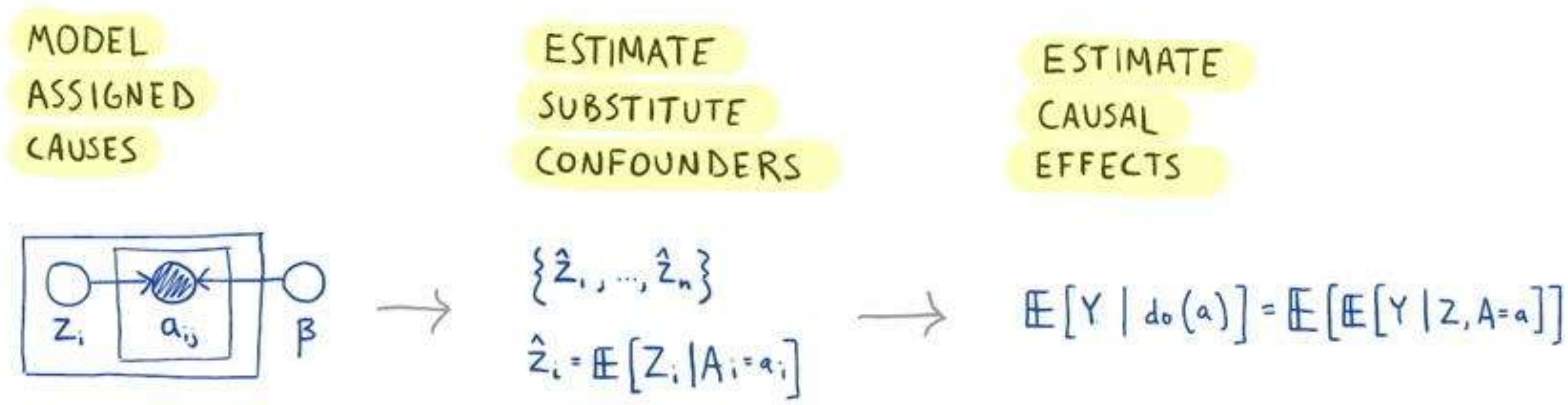
Discussion

Causal inference from observational data



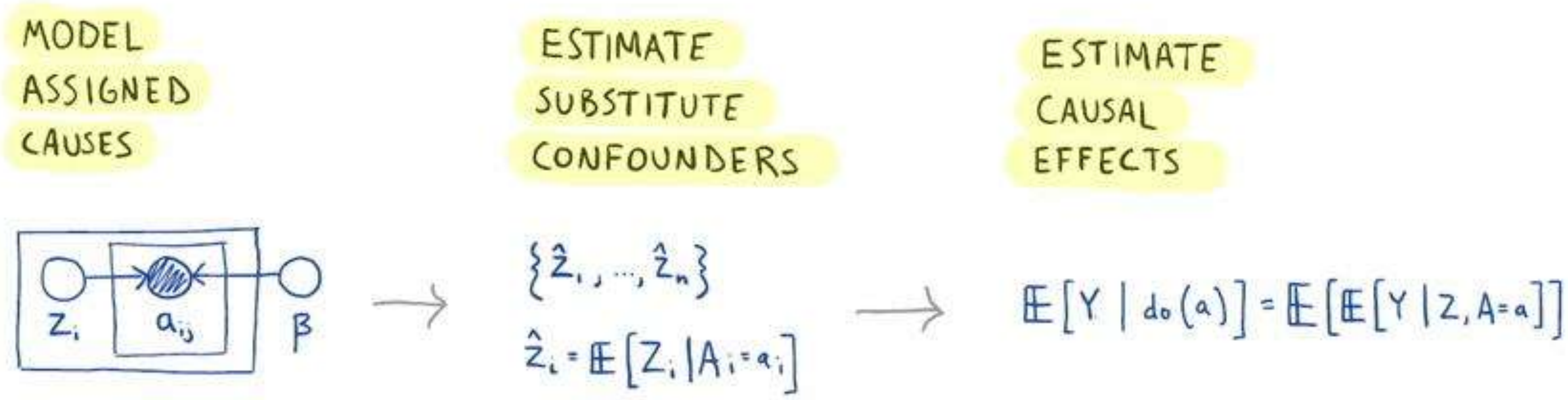
- **How can we understand the world through observation?**
- Important to genetics, economics, physics, medicine, finance, ...
- Today: Use probabilistic machine learning for causal inference

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

Y. Wang and D. Blei. *The Blessings of Multiple Causes*, 2018.
`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 $\mathbb{E}[Y\,;\mathrm{do}(\mathbf{a})]$ 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 $\hat{z} = f(\mathbf{a})$, e.g., many causes.
- (Theorem) Differences of complete interventions are

$$\mathbb{E}\left[Y \,;\, \mathrm{do}(\mathbf{a})\right] - \mathbb{E}\left[Y \,;\, \mathrm{do}(\mathbf{a}')\right].$$

  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

$$\mathbb{E}\left[Y \,;\, \mathrm{do}(\mathbf{a}_B)\right].$$

  It is identifiable with overlap on the subset, $p(\mathbf{a}_B \mid z) > 0$.