Let's chop them up!
(A brief survey on SIR techniques)

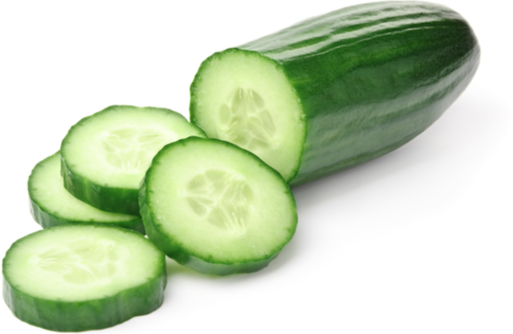
C. Tao [1]

[1] School of Mathematical Sciences, Fudan University

[2] Department of Computer Science, University of Warwick

January 21, 2016, University of Warwick

This ain’t a culinary lecture!



Outline

1. Sliced inverse regression
2. The Bayesian partition model
3. Other recent developments [optional]
4. Concluding remarks

Background

The challenge

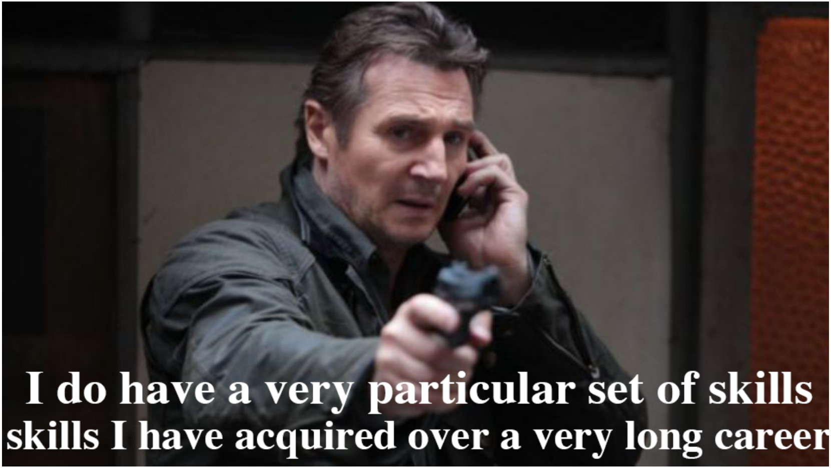
Say X are some high-dimensional predictors and Y are some responses of interest, one would like to have a low-dimensional summary $\widetilde{X}$ of X that is informative about Y.

Examples

X: genetic makeup,	Y: disease risk
X: historic quotes on stocks,	Y: future prices
X: brain activations,	Y: psychological status

Potential gain

- Better model generalizability and interpretability
- More efficient computations


I do have a very particular set of skills
skills I have acquired over a very long career

Common solutions

Two summarizing strategies

- Dimension reduction: $\widetilde{X}$ is a transformation of X
  - CCA, PLS, RRR
- Variable selection: $\widetilde{X}$ is a subset of X
  - LASSO, penGAM, ISIS (not those terrorists)

Measuring informativeness

- Parametric measures
  - Predictive power of $\widetilde{X}$ on Y
  - Model consistency (likelihood)
  - Association
- Nonparametric measure
  - Shared information

Aren’t they good enough?

Limitations

- Validity of the model assumptions
- Data consuming
- Computationally challenging
  - Applies to both para. and non-para. solutions

Any more appealing alternatives?

Regression revisited

Forward regression

$$\mathbb{E}[Y|X] = \phi(X)$$

- Estimate $\hat{\phi}_n$ with empirical sample $(\boldsymbol{X}_n, \boldsymbol{Y}_n)$

Cons

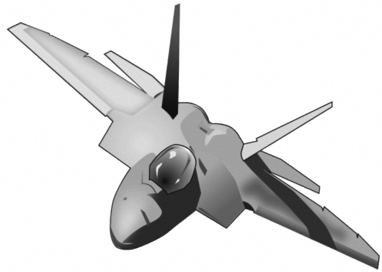
- The family of ϕ may not be known *apriori*
- Estimation often relies on the distribution of $Y = \psi(X, E)$
    - ψ the data generating mechanism
    - E the randomness involved

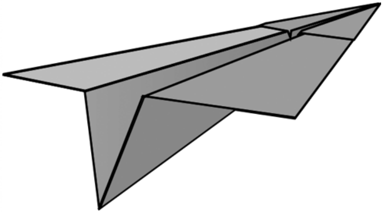
The catch

- We don't really need ϕ to characterize the dependency
- And we do not need to know the distribution of Y either

An simple analogy

- You learn the basic laws of aerodynamics from a paper plane
- But it takes a lot more to build an F22 raptor
- Basics is suffice for us, let's stick with it!!!





Sliced inverse regression (SIR)

Inverse regression

$$\mathbb{E}[X|Y] = \eta(Y)$$

Assuming the following general data generation mechanism

$$Y = \psi(X^\top \boldsymbol{\beta}_1, \cdots, X^\top \boldsymbol{\beta}_K, E). \tag{1}$$

Theorem (Li, 1991)

Under model (1), and assume X follows elliptical distributions, the centered inverse regression curve $\bar{\eta}(Y) = \mathbb{E}[X|Y] - \mathbb{E}[X]$ is contained in the linear subspace spanned by $\Sigma_{XX}\boldsymbol{\beta}_k$ ($k = 1, \cdots, K$), where Σ_{XX} denotes the covariance matrix of X.

Sketch of proof.

$$\begin{aligned}
\mathbb{E}[X|Y] &= \mathbb{E}[\mathbb{E}[X|\boldsymbol{\eta}^T X, Y]|Y] \\
&= \mathbb{E}[\mathbb{E}[X|\boldsymbol{\eta}^T X]|Y] \\
&= \mathbb{E}[\mathbb{E}[P_\eta X + Q_\eta X|\boldsymbol{\eta}^T X]|Y] \\
&= \mathbb{E}[P_\eta X|Y] + \mathbb{E}[\mathbb{E}[Q_\eta X|\boldsymbol{\eta}^T X]|Y]
\end{aligned}$$

Since for the elliptical distribution $\mathbb{E}[Q_\eta X|\boldsymbol{\eta}^T X] = 0$, thus the theorem holds. □

$\mathbb{E}[\text{cov}[Z|Y]] = \text{cov}[Z] - \text{cov}[\mathbb{E}[Z|Y]]$ also could be used to extract information of βs.

SIR estimation

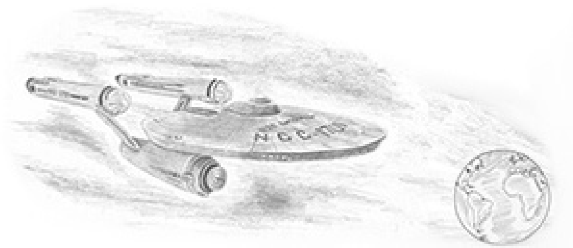
In the case of one-dimensional Y

Algorithm

1. Standardizing X
2. Partitioning the whole data into several slices according to the value of Y
3. Calculate the slice mean of X accordingly
4. Run principal component analysis on slice means of X
5. Locating the most important K-dimensional subspace for tracking the inverse regression curve $\mathbb{E}[X|Y]$

Take home messages

1. Don’t rely on the models, let the data talk
2. The conditional distribution of X given Y encodes vital information about dependencies



Bayesian partitioning for eQTL analysis

What is eQTL?

- eQTL: expression quantitative trait loci
- To correlate variations in the gene expression with DNA
- cQTL: clinical QTL (traditional GWAS)
- Finding co-localize eQTL and cQTL identifies a list of candidate genes for follow-up studies of the disease

For imaging-genetic studies

- eQTL ⇒ activations, structural images, connectivities, *etc.*
- To identify a list of genes and imaging traits that correlate with the clinical symptoms.

Terminologies explained

- cis-acting and trans-acting
  - on the gene or not
- epistatic and pleiotropic effects
  - many to one and one to many

Some historical comments

- eQTL analysis dates back to a time genome-wide dense sequencing is technically impossible, so it utilizes the LD structure of the genetic markers to identify causal locus.

Bayesian partitioning (BP) models for eQTL

Highlights

- Integrates eQTL, cQTL and SIR
- Distribution based, indep. of specific interactions
- Accounting for association structures (LD, co-expression)
- Dynamic clustering
- Improved sensitivity for weak couplings

The full model is overwhelmingly sophisticated, so I'll try to capitalize only the key ideas in this talk.

A peek of causal modeling

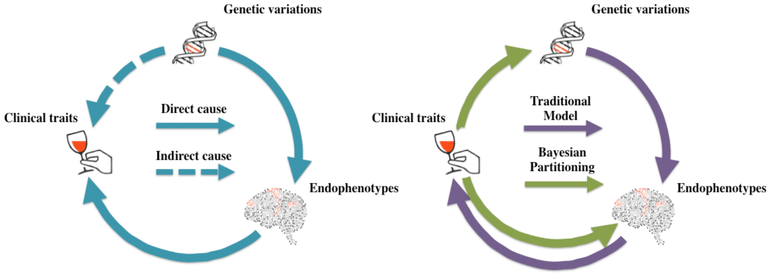


Figure : (**Left**) Ground truth causal network (**Right**) Bayesian causal network used by traditional model (purple) and Bayesian partitioning model (green). Endophenotypes can include gene expression, brain activation, etc.

YOU KNOW NOTHING,
JON SNOW
quickmeme.com

Key question for traditional bayesian model

Which models are most consistent with the data **under our assumptions**?

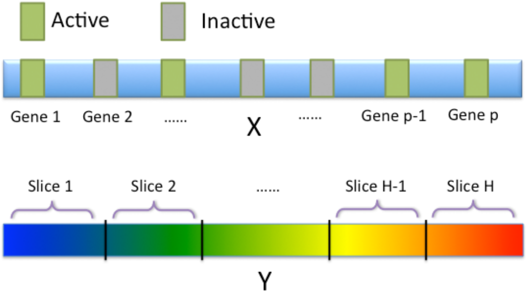
Key question for Bayesian partition

Which **partition schemes** and **conditional distributions** that are most consistent with the data we observe?

BP for single quantitive trait

Basic notations

- X: categorical variables (SNPs), $X_j \in [1:K]$
- Y: quantitive trait (gene expression)
- $S(Y)$: slice membership, $h \in [1:H]$
- $\mathcal{A}$: QTL locus set



Dirichlet-multinomial model condition on partition

$$X_{\mathcal{A}}|S(Y) = h \sim \text{Multinomial}(1, \boldsymbol{\theta}_{\mathcal{A}}^{(h)})$$
$$\boldsymbol{\theta}_{\mathcal{A}}^{(h)} \sim \text{Dirichlet}(\tfrac{\alpha_0}{K^{|\mathcal{A}|}}, \cdots, \tfrac{\alpha_0}{K^{|\mathcal{A}|}})$$

Dynamic partitioning

The slicing prior $Pr(S(Y)) = \pi_0^{|S|-1}(1-\pi_0)^{n-|S|}$

Compute $Pr(X_A|S(Y))$ by integrating out $\boldsymbol{\theta}_{\mathcal{A}}^{(h)}$

$$Pr(X_{\mathcal{A}}|Y) = \textstyle\sum_{S(Y)\in\Omega} Pr(X_{\mathcal{A}}|S(Y))Pr(S(Y))$$

Can be computed in $O(n^2)$, draw slicing schemes from $Pr(S(Y)|X_{\mathcal{A}}, Y)$ via forward - summation - backward - sampling if needed

Grouping the genes

I: indicator function of active gene set $\mathcal{A}$

Saturated NULL model and posterior distribution

$$Pr(X_{\mathcal{A}^c}|X_{\mathcal{A}}, Y) = Pr(X_{\mathcal{A}^c}|X_{\mathcal{A}}) = \frac{Pr_{null}(X)}{Pr_{null}(X_{\mathcal{A}})}$$

$$Pr(I) \sim Bernoulli(\eta_I, p, |\mathcal{A}|)$$

$$P(I|Y,X) \propto P(X_{\mathcal{A}}|Y)P(X_{\mathcal{A}^c}|X_{\mathcal{A}})P(I) \propto \frac{Pr(X_{\mathcal{A}}|Y)}{Pr_{null}(X_{\mathcal{A}})}\left(\frac{\eta_I}{1-\eta_I}\right)^{|\mathcal{A}|}$$

Bayesian factor and Gibbs sampling

$$BF(\mathcal{A}|Y) = \frac{Pr(X_{\mathcal{A}}|Y)}{Pr_{null}(X_{\mathcal{A}})} = \sum_{S(Y)\in\Omega} BF(X_{\mathcal{A}}|S(Y))Pr(S(Y))$$

$$BF(X_{\mathcal{A}}|S(Y)) = \frac{Pr(X_{\mathcal{A}}|S(Y))}{Pr_{null}(X_{\mathcal{A}})}$$

$$Pr(I_k = 1|I_{[-k]}, X, Y) = \frac{\eta_I BF(\mathcal{A}_{[-k]} \cup \{k\}|Y)}{(1-\eta_I)BF(\mathcal{A}_{[-k]}|Y)+\eta_I BF(\mathcal{A}_{[-k]} \cup \{k\}|Y)}$$

Multiple conditionally indep. QTL groups

$\mathcal{A}_1, \cdots, \mathcal{A}_M$: conditionally indep. associated gene groups

$P_m(X_{\mathcal{A}}|S(Y)) = \prod_{m=1}^{M} P(X_{\mathcal{A}_m}|S(Y))$

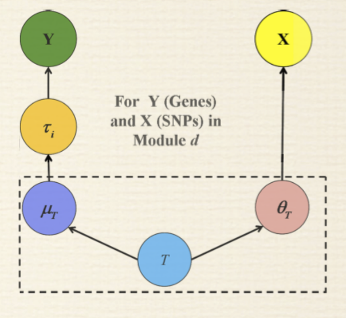
Partition follows Chinese restaurant process

Modeling block structure of LD

- L: genetic location
- $B, \mathcal{B}_h$: LD block partition and indicator
- $X_{\mathcal{B}_h} \sim \text{Multinomial}(1, \boldsymbol{\theta}_{\mathcal{B}}^{(h)}), \boldsymbol{\theta}_{\mathcal{B}}^{(h)} \sim \text{Dirichlet}(\frac{\alpha_b}{K^{|\mathcal{B}_h|}}, \cdots, \frac{\alpha_b}{K^{|\mathcal{B}_h|}})$
- $P_{blk}(X_{\mathcal{B}_h}), P_{blk}(X|B) = \prod_{h=1}^{|B|} P_{blk}(X_{\mathcal{B}_h})$
- $P_{blk}(X) = \sum P_{blk}(X|B)P(B), P_{blk}(X_{\mathcal{A}^c}|X_{\mathcal{A}}) = \frac{P_{blk}(X)}{P_{blk}(X_{\mathcal{A}})}$

Augmented partitioning, gene clustering and multiple modules

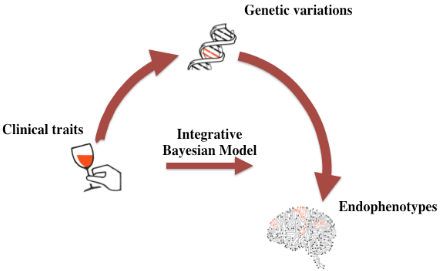
- R, T: auxiliary ranking and associated slicing
- $Y_{i,j}$: gene expressions for subject i, gene j
- $C_j, \mathcal{G}_c$: gene cluster membership
- $Y_{i,j}|C_j = c \sim N(\tau_{i,c}, \sigma_c^2)$, $\tau_{i,c}|T_i = t \sim N(\mu_{t,c}, \sigma_c^2/\kappa_1)$
- $\mu_{t,c} \sim N(0, \sigma_c^2/\kappa_2), \sigma_c^2 \sim Inv\chi^2(\nu_0, \sigma_0^2)$



Comparison with integrative Bayesian model

Overview of the model in [FC Stingo, 2013, JASA]

- $X \in \mathbb{R}^p$ imaging features, $Z \in \mathbb{R}^q$ genetic covariates
- $G \in \{1, \cdot, K\}$ group indicator
- Latent labels for discriminatory features/covariates
  - $\gamma \in \{0, 1\}^p$ feature label
  - $\delta \in \{0, 1\}^q$ covariate label



Modeling

- Feature modeling
  - Nondiscriminatory: $f_0(X_j; \theta_{0j}) \sim N(0, \sigma^2_{0j})$
  - Discriminatory (group k): $f_k(X_j; \theta_{kj}) \sim N(\mu_{kj}, \sigma^2_{kj})$
- Covariate effect modeling
  - $\mu_{kj} = \mu_{0k} + \beta^{\mathsf{T}}_{kj} Z$, μ_{0k} the random effects
  - Sparsity priors on $\boldsymbol{\beta}_{k(\gamma)}$
- MRF priors for spatial structure

Comparisons

- Commonalities
  - Sample the latent indicator for feature and covariate
  - Split sample into groups
- Disparities
  - Deterministic VS agnostic grouping
  - Generative VS nongenerative modeling

Other recent developments

What we learnt from BP

- SIR is nonparametric, the rest are parametric
- A blend of para. and non-para. ideas might prove useful

Sliced inverse regression with interaction detection (SIRI)

- Variable selection for active set $\mathcal{A}$

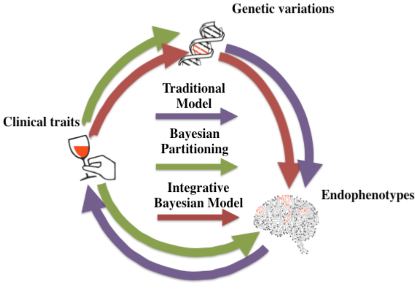
$$X_{\mathcal{A}} | Y \in S_h \sim \mathrm{MVN}(\boldsymbol{\mu}_h, \boldsymbol{\Sigma})$$
$$X_{\mathcal{A}^c} | (X_{\mathcal{A}}, Y \in S_h) \sim \mathrm{MVN}(\alpha + \beta^\top X_{\mathcal{A}}, \boldsymbol{\Sigma}_0)$$

- $\boldsymbol{\mu}_h \in \mathbb{V}^q \Longleftrightarrow$ SIR
- Likelihood ratio test to compare models
- Forward - addition - backward - deletion

Concluding remarks

Limitations

- Where is the p-value
- Difficult to implement and estimate
- Not accounting for the covariate effect
- One dimensional auxiliary ranking


Genetic variations
Clinical traits
Endophenotypes
Traditional Model
Bayesian Partitioning
Integrative Bayesian Model