Conditional Dependence Estimation via Shannon Capacity: Axioms, Estimators and Applications

Pramod Viswanath

University of Illinois at Urbana-Champaign

09/23/2016

Motivation

- A measure $\tau(P_{X,Y})$ captures how X influences Y.
- Choices of $\tau(P_{X,Y})$:
  - Linear: Pearson Correlation Coefficient: $\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$.
  - Nonlinear: Mutual Information $I(X;Y) = \mathbb{E}\left[\log \frac{p_{xy}}{p_x p_y}\right]$.
- Sometimes they don't work well.

Motivation

- Let X, Y be binary random variables; X denoting smoking and Y denoting lung-cancer. $X \to Y$.
- Hypothetical city: P_X is small but $P_{Y|X}$ is large.
- Mutual Information is very small.

Motivation

- Let X, Y be binary random variables; X denoting smoking and Y denoting lung-cancer. $X \to Y$.
- Hypothetical city: P_X is small but $P_{Y|X}$ is large.
- Mutual Information is very small.
- Depends on joint distribution $P_{X,Y}$: Factual influence measure.
- Depends only on conditional distribution $P_{Y|X}$: Potential influence measure.

Potential Influence Measures

- Focus of this talk: Potential influence measures.
- Let the channel from X to Y be given as $P_{Y|X}$.
- A potential influence measure on the space of conditional distributions.

$$\tau : P_{Y|X} \mapsto \mathbb{R}^+.$$

UMI: Uniform Mutual Information

- $UMI \triangleq I(U_X P_{Y|X})$.

$$X \sim U_X \longrightarrow P_{Y|X} \longrightarrow Y.$$

- Pros: Potential influence measure.
- Cons: Requires support of X to be discrete or compact.

CMI: Capacitated Mutual Information

- Shannon capacity $CMI \triangleq \max_{Q_X} I(Q_X P_{Y|X})$.

$$X \sim Q_X \longrightarrow P_{Y|X} \longrightarrow Y.$$

Estimators

- Question: How to estimate UMI and CMI from **samples** of (X, Y)?
- Real valued, high dimensional, (X, Y).
- **Key Issue**: Samples are drawn from $P_X P_{Y|X}$ but need to estimate MI for $U_X P_{Y|X}$.
- For CMI: need to do optimization and estimation jointly.

Possible Approaches

- Joint kernel density estimation for $P_{X,Y}$.
  - Need numerical integration.
  - May overkill.
- Discretization based algorithms.
  - May be sensitive.

Mutual Information Estimators

- $\widehat{I}(X;Y) = \widehat{H}(X) + \widehat{H}(Y) - \widehat{H}(X,Y)$.
- Pros: We have good entropy estimators.
  - Nearest neighbor approach.
- Cons: Not adaptable to UMI and CMI.
- Want other MI estimators to be adaptable to UMI and CMI.

Adapting to UMI and CMI

- **Inspired work** by [Kraskov, Stögbauer and Grassberger, 2004]:

$$\widehat{I}_{\mathrm{KSG}}(X;Y) = \log(k - \frac{1}{2}) + \log(N) - \frac{1}{N}\sum_{i=1}^{N}\big(\log(n_{x,i}) + \log(n_{y,i})\big).$$

- $n_{x,i}$: number of samples within ℓ_∞ distance of k-nearest distance $\rho_{k,i}$ in X-dimension.

Adapting to UMI and CMI

- **Inspired work** by [Kraskov, Stögbauer and Grassberger, 2004]:

$$\widehat{I}_{\mathrm{KSG}}(X;Y) = \log(k - \frac{1}{2}) + \log(N) - \frac{1}{N}\sum_{i=1}^{N}\big(\log(n_{x,i}) + \log(n_{y,i})\big).$$

- $n_{x,i}$: number of samples within ℓ_∞ distance of k-nearest distance $\rho_{k,i}$ in X-dimension.

- **Importance sampling:**
  - Compute functional of $Q_X P_{Y|X}$ whereas samples from $P_X P_{Y|X}$.
  - **Weights**: $w_i = Q_X(X_i)/P_X(X_i)$.

Adapting to UMI and CMI

- **Inspired work** by [Kraskov, Stögbauer and Grassberger, 2004]:

$$\widehat{I}^{(w)}_{\mathrm{KSG}}(X;Y) = \log(k - \frac{1}{2}) + \log(N) - \frac{1}{N}\sum_{i=1}^{N} w_i \big(\log(n_{x,i}) + \log(n_{y,i}) \big).$$

- $n_{x,i}$: number of samples within ℓ_∞ distance of k-nearest distance $\rho_{k,i}$ in X-dimension.

- **Importance sampling:**
  - Compute functional of $Q_X P_{Y|X}$ whereas samples from $P_X P_{Y|X}$.
  - **Weights**: $w_i = Q_X(X_i)/P_X(X_i)$.

Adapting to UMI and CMI

- **Inspired work** by [Kraskov, Stögbauer and Grassberger, 2004]:

$$\widehat{I}^{(w)}_{\mathrm{KSG}}(X;Y) = \log(k-\frac{1}{2}) + \log(N) - \frac{1}{N}\sum_{i=1}^{N} w_i \big(\log(\widetilde{n}_{x,i}) + \log(\widetilde{n}_{y,i})\big).$$

- $\widetilde{n}_{x,i}$: **Weighted** number of samples within ℓ_∞ distance of k-nearest distance $\rho_{k,i}$ in X-dimension.

- **Importance sampling:**
  - Compute functional of $Q_X P_{Y|X}$ whereas samples from $P_X P_{Y|X}$.
  - **Weights**: $w_i = Q_X(X_i)/P_X(X_i)$.

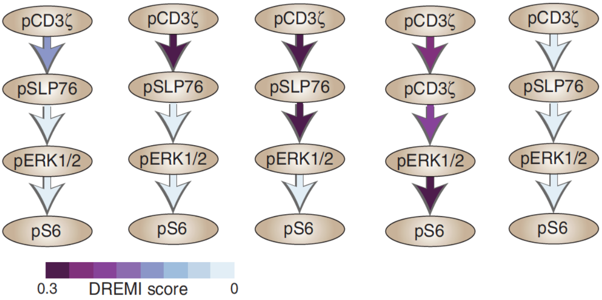
Adapting to UMI and CMI

- UMI : Let $Q_X = U_X$, weights are **inversely proportional** to density estimate.
- **Optimize** over weights for CMI:

$$\widehat{I}_{\mathrm{CMI}}(X;Y) = \max_{w_i : \sum_{i=1}^{n} w_i = 1} \widehat{I}^{(w)}_{\mathrm{KSG}}(X;Y).$$

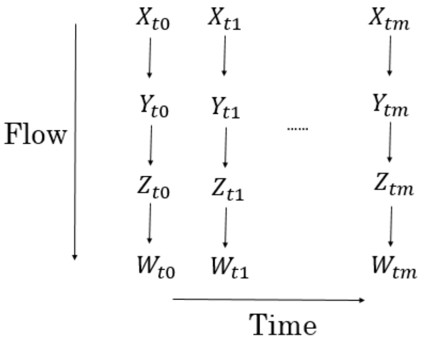
Application

- Single Cell Flow Cytometry [Krishnaswamy et al, Science 2014]:



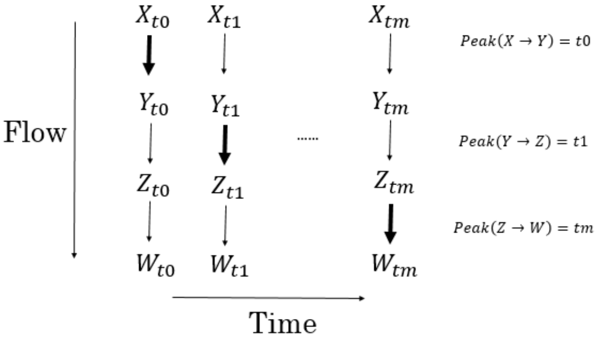
Single Cell Cytometry Application

- Single Cell Flow Cytometry:



Single Cell Cytometry Application

- Single Cell Flow Cytometry:



- Succeed if $\mathrm{Peak}(X \to Y) \le \mathrm{Peak}(Y \to Z) \le \mathrm{Peak}(Z \to W)$.

Single Cell Cytometry Application

- Significant reduction in **sample complexity**.

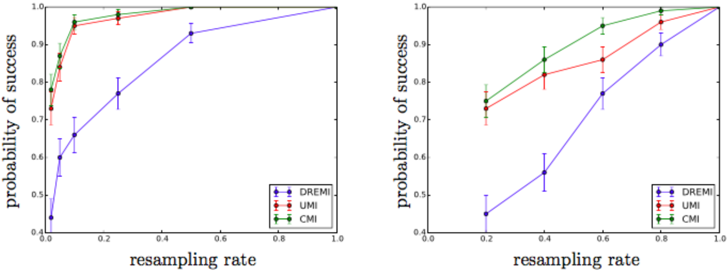


Figure: Figure 6(a,b) of [Krishnaswamy, et. al.]

Axiomatic View of potential influence measure

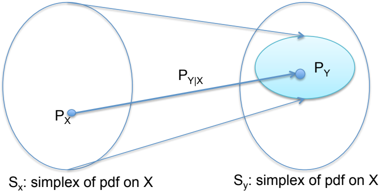
- Why choose UMI and CMI as potential influence measures?

(0) **Independence**.

(1) **Data Processing**.

(2) **Additivity**.

(3) **Monotonicity**: τ monotonic function of the range. Range = Convex Hull ($P_{Y|X=x}$).



Axiomatic View of influence measure τ

- UMI satisfies independence, data-processing, additivity axioms.
- UMI is **NOT** monotonic.
- CMI satisfies all the axioms.
- Other measures?

Summary

- **Novel potential** measures of influence of X on Y.
  - UMI, CMI.
- **Estimators** of UMI and CMI from sample.
  - Nearest neighbor methods.
  - KSG mutual information estimator.
  - Importance sampling.

Collaborators

Weihao Gao — Sreeram Kannan — Sewoong Oh