Link to paper:
goo.gl/GARoJH

Nonparanormal Information Estimation
A Framework for Realistic Dependence Estimation

Shashank Singh and Barnabás Póczos
Machine Learning Department & Department of Statistics
Carnegie Mellon University

Introduction

- Estimating dependence between variables is a fundamental subproblem in machine learning.
- Mutual information (MI) is a popular measure of dependence.
- Previous MI estimators need strong assumptions or low dimensionality.
- We propose/study *nonparanormal estimators* to bridge this gap.

Information Estimation

- **Multivariate Mutual Information:** Given a D-dimensional random variable $X = (X_1, ..., X_D)$ with joint density $p = p_1 \times \cdots p_D$,
$$I(X) := \mathop{\mathbb{E}}_{X\sim p}\left[\log\left(\frac{p(X)}{\prod_{j=1}^D p_j(X_j)}\right)\right] = D_{KL}\left(p \middle\| \prod_{j=1}^D p_j\right),$$
where D_{KL} denotes KL-divergence.
  - $I(X)$ measures dependency/redundancy between $X_1, ..., X_D$.
  - Pairwise mutual information, conditional mutual information, transfer entropy, etc. can be written in terms of I.
- **Information Estimation** refers to the problem of estimating $I(X)$, given n IID samples of a random variable X.
- **Gaussian case:** If X is known to be Gaussian, the minimax mean squared error for information estimation is essentially $2D/n$. [1]
  - consistent if $D \in o(n)$, but Gaussianity is very restrictive
  - fails if X is heavy-tailed, multi-modal, skewed, nonlinear
- **Nonparametric Case:** If the density of X is known to be s-times differentiable, the minimax rate is $\asymp n^{-\frac{8s}{4s+D}}$. [2]
  - mild assumptions, but consistency requires $D \in o(\log n)$
  - in practice, fails if D is bigger than **4-6**

Research Question

- **Question: Can we estimate dependence in high dimensions without Gaussian assumptions?**
- **Our Answer: Yes, using a nonparanormal model!**

References

T Tony Cai, Tengyuan Liang, and Harrison H Zhou.
Law of log determinant of sample covariance matrix and optimal estimation of differential entropy for high-dimensional Gaussian distributions.
J. of Multivariate Analysis, 137:161–172, 2015.

Kirthevasan Kandasamy, Akshay Krishnamurthy, Barnabas Poczos, Larry Wasserman, and James M. Robins.
Nonparametric von mises estimators for entropies, divergences and mutual informations.
In *Advances in Neural Information Processing Systems*, pages 397–405, 2015.

LF Kozachenko and Nikolai N Leonenko.
Sample estimate of the entropy of a random vector.
Problemy Peredachi Informatsii, 23(2):9–16, 1987.

The Nonparanormal Distribution

- A D-dimensional random variable X taking values in $\mathcal{X}^D$ has a **nonparanormal** (or Gaussian copula) distribution, denoted $X \sim \mathcal{NPN}(\Sigma; f)$ if there exist differentiable monotone functions $f_1, ..., f_D : \mathcal{X} \to \mathbb{R}$ such that
$$f(X) = (f_1(X_1), ..., f_D(X_D)) \sim \mathcal{N}(0, \Sigma).$$

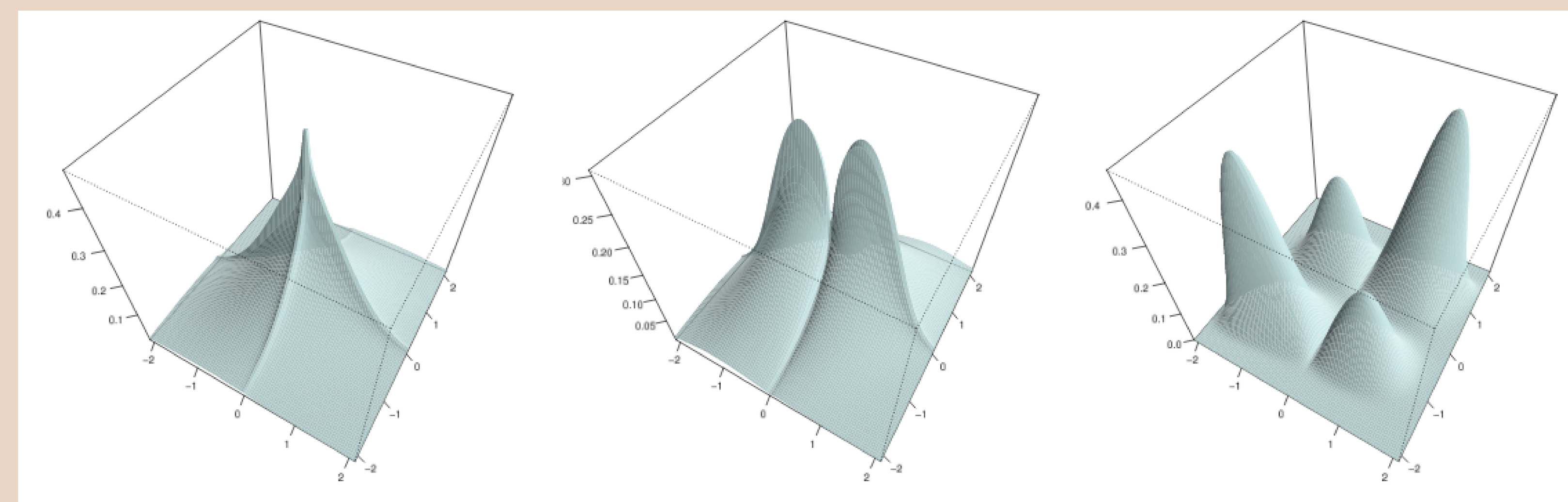


Figure 1: Examples of nonparanormal densities.

- Two perspectives:
  1. Marginal transformation of Gaussian
  2. 2^{nd}-order additive model for densities: $p(x) \propto e^{-f^T(x)\Sigma f(x)}$.
     - 1^{st}-order model is independent: $p(x) \propto e^{w \cdot f(x)}$

Estimating Nonparanormal Mutual Information

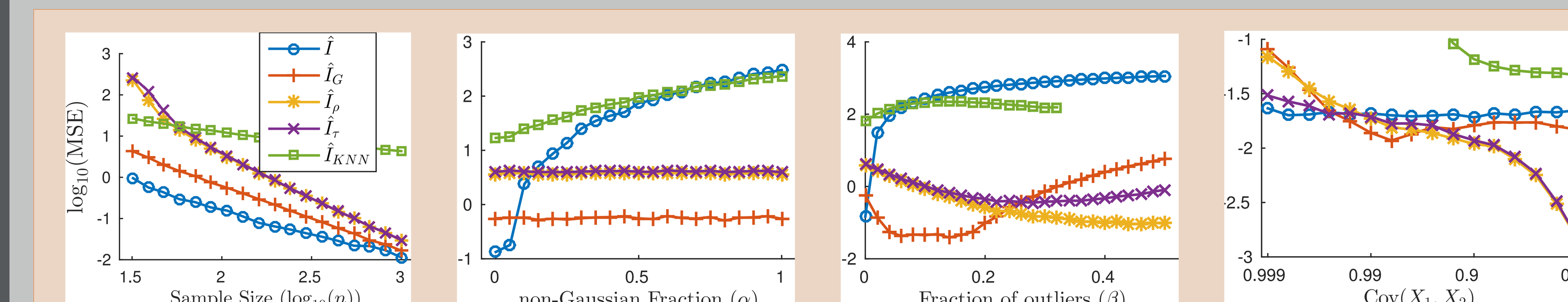
- **Basic Lemma:** If $X \sim \mathcal{NPN}(\Sigma; f)$, then
$$I(X) = -\frac{1}{2}\log|\Sigma|, \qquad (1)$$
where $|\Sigma|$ denotes the determinant of Σ. Hence,
  1. $I(X)$ doesn't depend on f.
  2. we can plug an estimate of Σ into Eq. (1).
- We propose **3** distinct estimators for Σ:
  - *Gaussianization Estimator* $\widehat{I}_G$ transforms data to have asymptotically Gaussian marginals and then estimates the covariance directly.
  - *Spearman* $\widehat{I}_\rho$ and *Kendall* $\widehat{I}_\tau$ estimators transform estimated rank-correlation, based on the identities
$$\Sigma = 2\sin\left(\frac{\pi}{6}\rho\right) \quad \text{and} \quad \Sigma = \sin\left(\frac{\pi}{2}\tau\right),$$
where ρ and τ are Spearman's and Kendall's rank correlation matrices.
- $\widehat{\Sigma}_G$, $\widehat{\Sigma}_\rho$, $\widehat{\Sigma}_\tau$ may not be positive definite (so $\mathbb{P}\left[\log|\widehat{\Sigma}| = \infty\right] > 0$).
  - Regularize estimate of Σ to have minimum eigenvalue $\lambda_D(\widehat{\Sigma}) \geq z > 0$, where z is a tuning parameter. i.e., use
$$\widehat{\Sigma}_{T,z} := \mathop{\mathrm{argmin}}_{\Sigma : \lambda_D(\Sigma) \geq z} \left\|\Sigma - \widehat{\Sigma}_T\right\|_F \quad \text{for} \quad T \in \{G, \rho, \tau\}.$$

Theoretical Results

- **Theorem 1:** If $z \leq \lambda_D(\Sigma)$, there exists a universal constant C s.t.
$$\mathbb{E}\left[\left(\widehat{I}_{\rho,z} - I\right)^2\right] \leq \frac{CD^2}{z^2 n}.$$
- For Gaussian X, the distribution of $\widehat{I} - I$ is independent of Σ [1]
  - Quite surprising, since $I \to \infty$ as $\lambda_D(\Sigma) \to 0$!
- **Theorem 2:** There exists a constant $C_{n,D}$ such that
$$\inf_{\widehat{I}} \sup_{\Sigma : \lambda_D(\Sigma) \geq \lambda} \mathbb{E}\left[\left(\widehat{I} - I\right)^2\right] \geq -C_{n,D}\log^2(\lambda).$$

Experimental Results

- We compare **5** estimators:
  - Debiased (optimal) Gaussian estimator $\widehat{I}$ [1]
  - Our proposed estimators $\widehat{I}_{G,z}$, $\widehat{I}_{\rho,z}$, $\widehat{I}_{\tau,z}$, with $z = 10^{-3}$
  - Nonparametric k-nearest neighbors estimator $\widehat{I}_{KNN}$ [3], with $k = 2$



Experiment 1: If X is Gaussian, NPN estimators approach $\widehat{I}$.

Experiment 2: If we transform marginals, $\widehat{I}$ diverges.

Experiment 3: NPN estimators are robust to outliers.

Experiment 4: NPN estimators (but not $\widehat{I}$) error depends on Σ.

Experiment details: All results are averaged over **100** IID trials. In each trial, $\Sigma \in \mathbb{R}^{25\times 25}$ is randomly sampled from a Wishart distribution. Experiment 2 transforms a fraction α of dimensions according to $z \mapsto e^z$. Experiment 3 replaces a fraction β of data uniformly at random from $\{-5, +5\}$.

Conclusions

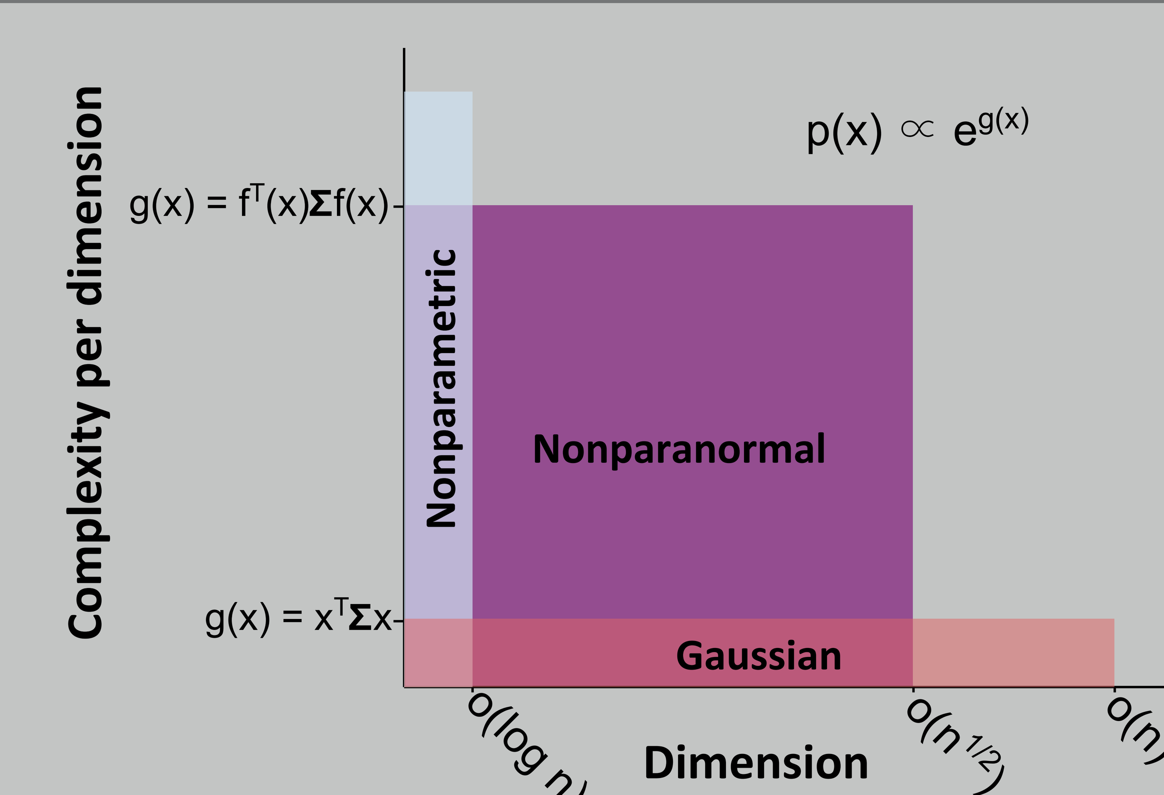


Figure 2: Phase diagram showing when each type of estimator is consistent.

http://www.andrew.cmu.edu/user/sss1/
August 8, 2017
sss1@andrew.cmu.edu