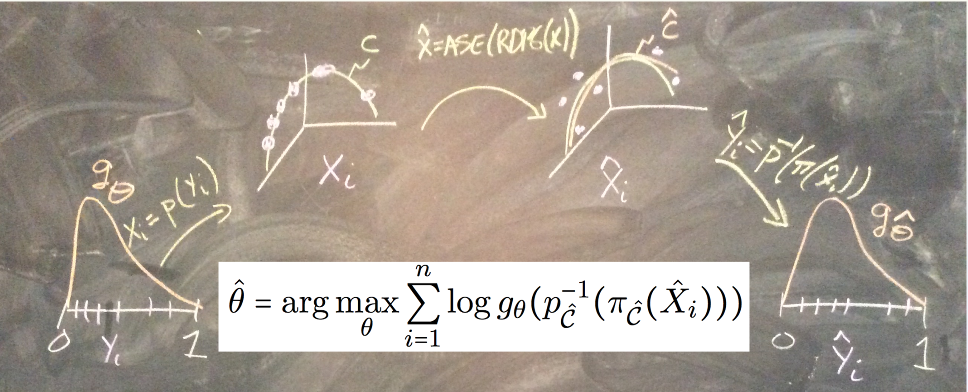
# Manifold Learning for Subsequent Inference


Carey E. Priebe
Johns Hopkins University


June 20, 2018
DARPA Fundamental Limits of Learning (FunLoL)
Los Angeles, California




$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \log g_\theta(p_{\hat{\mathcal{C}}}^{-1}(\pi_{\hat{\mathcal{C}}}(\hat{X}_i)))$$

http://arxiv.org/abs/1806.01401

Submitted to Statistical Science

# On estimation and inference in latent structure random graphs

**Avanti Athreya, Minh Tang, Younger Park, and Carey E. Priebe**

Department of Applied Mathematics & Statistics, Johns Hopkins University

Key Contribution 1: Latent Structure Model
Key Contribution 2: $2 \to \infty$ & Donsker CLT $\implies$ Efficiency
Key Contribution 3: Manifold Learning for Subsequent Inference

Acknowledgment: DARPA FunLoL via SIMPLEX

# Latent Structure Model

## Definition

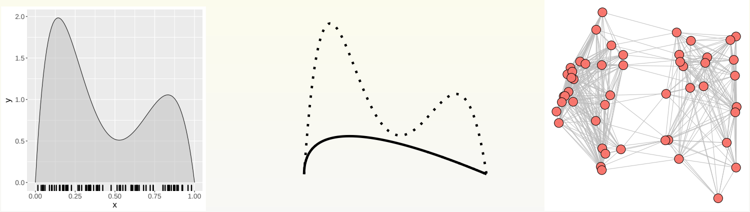
Let $\mathcal{C}$ be an LSM-regular curve of minimal subspace dimension $d$. Let $p(t) : [0,1] \to \mathcal{C}$ denote the arclength reparameterization of $\mathcal{C}$. Let $G$ be a family of distributions on $[0,1]$, and let $F$ denote the induced distributions on $\mathcal{C}$: that is, $\mu_F(B) = \mu_G(p^{-1}(B))$ for any set $B \subset \mathcal{C}$, where $\mu_F$ and $\mu_G$ are the distribution measures of associated to $F$ and $G$. We say that an RDPG with i.i.d. latent position matrix $\mathbf{X}$ is a

*latent structure random graph with parametric underlying distribution $G$ and known univariate support $\mathcal{C}$*

if the latent position vectors $X_i$ are distributed according to $F = G(p^{-1})$ where $G$ belongs to some regular parametric family $G_\Theta = \{G_\theta; \theta \in \Theta \subset \mathbb{R}^l\}$ on $[0,1]$ and $p$ and $\mathcal{C}$ are known.
We write

$$X_i \sim^{i.i.d.} F = G_\theta(p^{-1}), \theta \in \Theta;\ \operatorname{supp} F = \mathcal{C}.$$

# Latent Structure Model




Left: *Unobserved* density on $[0,1]$.
Center: *Unobserved* latent position density on $\mathcal{C}$.
Right: *Observed* graph generated from latent positions on $\mathcal{C}$.

# Latent Structure Model

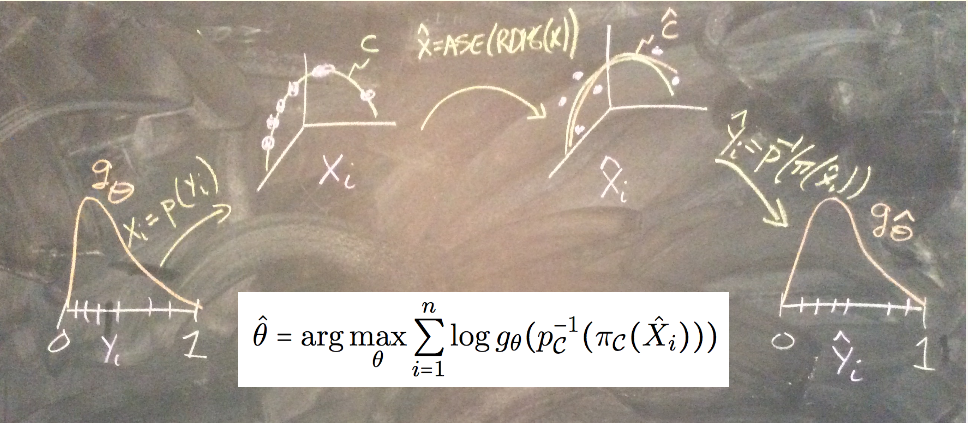
## Definition

Let $\mathcal{C}$ be an LSM-regular curve of minimal subspace dimension $d$. Let $p(t) : [0,1] \to \mathcal{C}$ denote the arclength reparameterization of $\mathcal{C}$. Let $G$ be a family of distributions on $[0,1]$, and let $F$ denote the induced distributions on $\mathcal{C}$: that is, $\mu_F(B) = \mu_G(p^{-1}(B))$ for any set $B \subset \mathcal{C}$, where $\mu_F$ and $\mu_G$ are the distribution measures of associated to $F$ and $G$. We say that an RDPG with i.i.d. latent position matrix $\mathbf{X}$ is a *latent structure random graph with parametric underlying distribution $G$ and known univariate support* $\mathcal{C}$ if the latent position vectors $X_i$ are distributed according to $F = G(p^{-1})$ where $G$ belongs to some regular parametric family $G_\Theta = \{G_\theta ; \theta \in \Theta \subset \mathbb{R}^l\}$ on $[0,1]$ and $p$ and $\mathcal{C}$ are known. We write

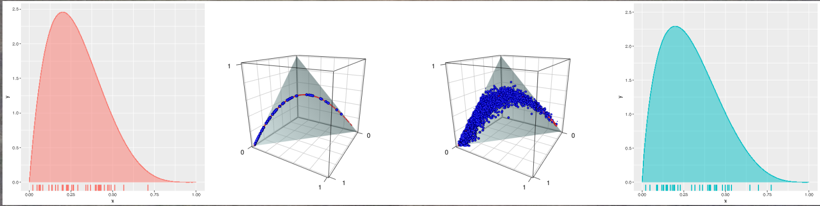
$$X_i \sim^{i.i.d.} F = G_\theta(p^{-1}), \theta \in \Theta;\ \operatorname{supp} F = \mathcal{C}.$$

*latent structure random graph*
*with parametric/nonparametric underlying distribution $G$*
*and known/parametric/nonparametric support $\mathcal{C}$*

# Asymptotic Efficiency in the Latent Structure Model




$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \log g_\theta(p_{\mathcal{C}}^{-1}(\pi_{\mathcal{C}}(\hat{X}_i)))$$



# Adjacency Spectral Embedding

http://arxiv.org/abs/1709.05454

(Journal of Machine Learning Research)

## Statistical inference on random dot product graphs: a survey

Avanti Athreya, Donniell E. Fishkind, Keith Levin, Vince Lyzinski, Youngser Park, Yichen Qin, Daniel L. Sussman, Minh Tang, Joshua T. Vogelstein, Carey E. Priebe

*(Submitted on 16 Sep 2017)*

The random dot product graph (RDPG) is an independent-edge random graph that is analytically tractable and, simultaneously, either encompasses or can successfully approximate a wide range of random graphs, from relatively simple stochastic block models to complex latent position graphs. In this survey paper, we describe a comprehensive paradigm for statistical inference on random dot product graphs, a paradigm centered on spectral embeddings of adjacency and Laplacian matrices. We examine the analogues, in graph inference, of several canonical tenets of classical Euclidean inference: in particular, we summarize a body of existing results on the consistency and asymptotic normality of the adjacency and Laplacian spectral embeddings, and the role these spectral embeddings can play in the construction of single- and multi-sample hypothesis tests for graph data. We investigate several real-world applications, including community detection and classification in large social networks and the determination of functional and biologically relevant network properties from an exploratory data analysis of the Drosophila connectome. We outline requisite background and current open problems in spectral graph inference.

| | |
|---|---|
| Comments: | An expository survey paper on a comprehensive paradigm for inference for random dot product graphs, centered on graph adjacency and Laplacian spectral embeddings. Paper outlines requisite background; summarizes theory, methodology, and applications from previous and ongoing work; and closes with a discussion of several open problems |
| Subjects: | **Methodology (stat.ME)**; Statistics Theory (math.ST); Machine Learning (stat.ML) |
| MSC classes: | 62FXX, 62GXX, 62HXX, 05CXX |
| Journal reference: | Journal of Machine Learning Research, 2018 |
| Cite as: | arXiv:1709.05454 **[stat.ME]** |

# Theorem 1

We now describe a consistency result in the $2 \to \infty$ norm that provides uniform control of deviations between the estimated and true latent positions. This uniform control can matter significantly for subsequent inference. An analogous result is available for much more general random matrix perturbations.

**Theorem (Lyzinski et al., *IEEE TNSE*, 2017; Cape et al., *Annals of Statistics*, forthcoming)**

*Let $\mathbf{A}_n \sim \mathrm{RDPG}(\mathbf{X}_n)$ for $n \geqslant 1$ be a sequence of random dot product graphs where the $\mathbf{X}_n$ is assumed to be of rank $d$ for all $n$ sufficiently large. Let $\mathbf{P}_n = \mathbf{X}_n \mathbf{X}_n^\top$ and let $\delta_n = \max_i \sum_j \mathbf{P}_{n,ij}$ be the maximum expected degree. Denote by $\hat{\mathbf{X}}_n$ the adjacency spectral embedding of $\mathbf{A}_n$ and let $(\hat{\mathbf{X}}_n)_i$ and $(\mathbf{X}_n)_i$ be the $i$-th row of $\hat{\mathbf{X}}_n$ and $\mathbf{X}_n$, respectively. Let $E_n$ be the event that there exists an orthogonal transformation $\mathbf{W}_n \in \mathbb{R}^{d \times d}$ such that for some $C > 0$*

$$\max_i \|(\hat{\mathbf{X}}_n)_i - \mathbf{W}_n (\mathbf{X}_n)_i\| \leqslant \frac{C d^{1/2} \log^2 n}{\delta_n^{1/2}}.$$

*Then $E_n$ occurs asymptotically almost surely: $\mathbb{P}(E_n) \to 1$.*

# Theorem 1

**Theorem (Lyzinski et al., *IEEE TNSE*, 2017; Cape et al., *Annals of Statistics*, forthcoming)**

*Let* $\mathbf{A}_n \sim \mathrm{RDPG}(\mathbf{X}_n)$ *for* $n \geqslant 1$ *be a sequence of random dot product graphs where the* $\mathbf{X}_n$ *is assumed to be of rank* $d$ *for all* $n$ *sufficiently large. Let* $\mathbf{P}_n = \mathbf{X}_n \mathbf{X}_n^\top$ *and let* $\delta_n = \max_i \sum_j \mathbf{P}_{n,ij}$ *be the maximum expected degree. Denote by* $\hat{\mathbf{X}}_n$ *the adjacency spectral embedding of* $\mathbf{A}_n$ *and let* $(\hat{\mathbf{X}}_n)_i$ *and* $(\mathbf{X}_n)_i$ *be the* $i$*-th row of* $\hat{\mathbf{X}}_n$ *and* $\mathbf{X}_n$*, respectively. Let* $E_n$ *be the event that there exists an orthogonal transformation* $\mathbf{W}_n \in \mathbb{R}^{d \times d}$ *such that for some* $C > 0$

$$\max_i \|(\hat{\mathbf{X}}_n)_i - \mathbf{W}_n (\mathbf{X}_n)_i\| \leqslant \frac{C d^{1/2} \log^2 n}{\delta_n^{1/2}}.$$

*Then* $E_n$ *occurs asymptotically almost surely:* $\mathbb{P}(E_n) \to 1$.

# Theorem 2

Having established that the estimated latent positions are consistent in $2 \to \infty$, we next point out that the latent position estimates are asymptotically normal. Specifically, for a $d$-dimensional random dot product graph with i.i.d. latent positions, there exists a sequence of orthogonal matrices $\mathbf{W}_n$ such that for any row index $i$, $\sqrt{n}(\mathbf{W}_n(\hat{\mathbf{X}}_n)_i - (\mathbf{X}_n)_i)$ converges to a mixture of multivariate normals.

## Theorem (Athreya et al., *Sankhya*, 2016)

*Let* $(\mathbf{A}_n, \mathbf{X}_n) \sim \mathrm{RDPG}(F)$ *be a sequence of adjacency matrices and associated latent positions of a* $d$*-dimensional random dot product graph according to an inner product distribution* $F$*. Let* $\Phi(\mathbf{x}, \Sigma)$ *denote the cdf of a (multivariate) Gaussian with mean zero and covariance matrix* $\Sigma$*, evaluated at* $\mathbf{x} \in \mathbb{R}^d$*. Then there exists a sequence of orthogonal* $d$*-by-*$d$ *matrices* $(\mathbf{W}_n)_{n=1}^{\infty}$ *such that for all* $z \in \mathbb{R}^d$ *and for any fixed index* $i$*,*

$$\lim_{n\to\infty} P\left[n^{1/2}(\mathbf{W}_n(\hat{\mathbf{X}}_n)_i - (\mathbf{X}_n)_i) \leqslant z \mid \mathbf{X}_i = x\right] = \Phi\left(z, \Sigma(\mathbf{x})\right),$$

*where*

$$\Sigma(\mathbf{x}) = \Delta^{-1}\mathbb{E}\left[(\mathbf{x}^\top X_1 - (\mathbf{x}^\top X_1)^2)X_1X_1^\top\right]\Delta^{-1}, \Delta = \mathbb{E}[X_1X_1^\top], X_1 \sim F.$$

# Theorem 2

## Theorem (Athreya et al., *Sankhya*, 2016)

*Let* $(\mathbf{A}_n, \mathbf{X}_n) \sim \mathrm{RDPG}(F)$ *be a sequence of adjacency matrices and associated latent positions of a* $d$*-dimensional random dot product graph according to an inner product distribution* $F$*. Let* $\Phi(\mathbf{x}, \Sigma)$ *denote the cdf of a (multivariate) Gaussian with mean zero and covariance matrix* $\Sigma$*, evaluated at* $\mathbf{x} \in \mathbb{R}^d$*. Then there exists a sequence of orthogonal* $d$*-by-*$d$ *matrices* $(\mathbf{W}_n)_{n=1}^{\infty}$ *such that for all* $\boldsymbol{z} \in \mathbb{R}^d$ *and for any fixed index* $i$*,*

$$\lim_{n\to\infty} P\left[n^{1/2}(\mathbf{W}_n(\hat{\mathbf{X}}_n)_i - (\mathbf{X}_n)_i) \leqslant \boldsymbol{z} \mid \mathbf{X}_i = \boldsymbol{x}\right] = \Phi\left(\boldsymbol{z}, \Sigma(\mathbf{x})\right),$$

*where*

$$\Sigma(\mathbf{x}) = \Delta^{-1}\mathbb{E}\left[(\mathbf{x}^\top X_1 - (\mathbf{x}^\top X_1)^2)X_1X_1^\top\right]\Delta^{-1}, \Delta = \mathbb{E}[X_1X_1^\top], X_1 \sim F.$$

# Theorem 3

We now state our key empirical process result that provides us uniform convergence of scaled sums of differences of functions of estimated and true latent positions, provided the functions belong to a sufficiently regular class.

We first recall certain definitions, which we reproduce from van der Vaart and Wellner. Let $X_i, 1 \leqslant i \leqslant n$ be identically distributed random variables on a measure space $(\mathcal{X}, \mathcal{B})$, and let $P_n$ be their associated *empirical measure*; that is, $P_n$ is the discrete random measure defined, for any $E \in \mathcal{B}$, by

$$P_n(E) = \frac{1}{n} \sum_{i=1}^{n} 1_E(X_i).$$

Let $P$ denote the common distribution of the random variables $X_i$, and suppose that $\mathcal{F}$ is a class of measurable, real-valued functions on $\mathcal{X}$. The *$\mathcal{F}$-indexed empirical process* $\mathbb{G}_n$ is the stochastic process

$$f \mapsto \mathbb{G}_n(f) = \sqrt{n}(P_n - P)f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big( f(X_i) - \mathbb{E}[f(X_i)] \Big).$$

Under certain conditions, the empirical process $\{\mathbb{G}_n(f) : f \in \mathcal{F}\}$ can be viewed as a map into $\ell^\infty(\mathcal{F})$, the collection of all uniformly bounded real-valued functionals on $\mathcal{F}$. In particular, let $\mathcal{F}$ be a class of functions for which the empirical process $\mathbb{G}_n = \sqrt{n}(P_n - P)$ converges to a limiting process $\mathbb{G}$ where $\mathbb{G}$ is a tight Borel-measurable element of $\ell^\infty(\mathcal{F})$ (more specifically a Brownian bridge). Then $\mathcal{F}$ is said to be a *$P$-Donsker class*.

# Theorem 3

## Theorem (Tang et al., *Bernoulli*, 2017)

*Let* $(\mathbf{X}_n, \mathbf{A}_n)$ *for* $n = 1, 2, \ldots,$ *be a sequence of* $d$*-dimensional* $\mathrm{RDPG}(F)$*. Let* $\mathcal{F}$ *be a collection of (at least) twice continuously differentiable functions on* $\operatorname{supp} F$ *with*

$$\sup_{f \in \mathcal{F}, X \in \operatorname{supp} F} \|(\partial f)(X)\| < \infty; \qquad \sup_{f \in \mathcal{F}, X \in \operatorname{supp} F} \|(\partial^2 f)(X)\| < \infty.$$

*Furthermore, suppose* $\mathcal{F}$ *is such that* $\mathbb{G}_n = \sqrt{n}(P_n - P)$ *converges to* $\mathbb{G}$*, a* $P$*-Brownian bridge on* $\ell^\infty(\mathcal{F})$*. Then there exists a sequence of orthogonal matrices* $\mathbf{W}_n$ *such that as* $n \to \infty$*,*

$$\sup_{f \in \mathcal{F}} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big( f(\mathbf{W}_n \hat{X}_i) - f(X_i) \Big) \Big| \to 0,$$

*where* $\{\hat{X}_i\}_{i=1}^n$ *are the rows of* $\hat{\mathbf{X}}_n$*. Therefore, the* $\mathcal{F}$*-indexed empirical process*

$$f \in \mathcal{F} \mapsto \hat{\mathbb{G}}_n f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big( f(\mathbf{W}_n \hat{X}_i) - \mathbb{E}[f(X_i)] \Big)$$

*also converges to* $\mathbb{G}$ *on* $\ell^\infty(\mathcal{F})$*.*

# Theorem 3

## Theorem (Tang et al., *Bernoulli*, 2017)

*Let* $(\mathbf{X}_n, \mathbf{A}_n)$ *for* $n = 1, 2, \ldots,$ *be a sequence of* $d$*-dimensional* $\mathrm{RDPG}(F)$*. Let* $\mathcal{F}$ *be a collection of (at least) twice continuously differentiable functions on* $\operatorname{supp} F$ *with*

$$\sup_{f \in \mathcal{F}, X \in \operatorname{supp} F} \|(\partial f)(X)\| < \infty; \qquad \sup_{f \in \mathcal{F}, X \in \operatorname{supp} F} \|(\partial^2 f)(X)\| < \infty.$$

*Furthermore, suppose* $\mathcal{F}$ *is such that* $\mathbb{G}_n = \sqrt{n}(P_n - P)$ *converges to* $\mathbb{G}$*, a* $P$*-Brownian bridge on* $\ell^\infty(\mathcal{F})$*. Then there exists a sequence of orthogonal matrices* $\mathbf{W}_n$ *such that as* $n \to \infty$*,*

$$\sup_{f \in \mathcal{F}} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big( f(\mathbf{W}_n \hat{X}_i) - f(X_i) \Big) \Big| \to 0,$$

*where* $\{\hat{X}_i\}_{i=1}^n$ *are the rows of* $\hat{\mathbf{X}}_n$*. Therefore, the* $\mathcal{F}$*-indexed empirical process*

$$f \in \mathcal{F} \mapsto \hat{\mathbb{G}}_n f = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big( f(\mathbf{W}_n \hat{X}_i) - \mathbb{E}[f(X_i)] \Big)$$

*also converges to* $\mathbb{G}$ *on* $\ell^\infty(\mathcal{F})$*.*

# Theorem 3

**Theorem (Tang et al., *Bernoulli*, 2017)**

*[...]*

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big( f(\mathbf{W}_n \hat{X}_i) - f(X_i) \Big) \right| \to 0$$

*[...]*

This theorem is in essence a functional central limit theorem for the estimated latent positions $\{\hat{X}_i\}$ in the RDPG setting, and we emphasize that for any $n$, the $\{\hat{X}_i\}_{i=1}^{n}$ are not jointly independent random variables, i.e., this is a functional central limit theorem for *dependent* data. Due to the non-identifiability of random dot product graphs, there is an explicit dependency on a sequence of orthogonal matrices $\mathbf{W}_n$. The main technical result from this theorem is used to show the asymptotic normality of $M$-estimation for the parameters of LSMs.

# Main Result: Theorem 4

**Theorem (Athreya et al., *Statistical Science*, forthcoming)**

*Suppose $X_i$ i.i.d. $F_{\theta_0}$ are latent positions of a latent structure model satisfying the regularity assumptions delineated above. Let $\mathbf{A}$ be the adjacency matrix of the random dot product with latent positions $\mathbf{X}$, and let $\hat{\mathbf{X}}$ be the suitably-rotated adjacency spectral embedding of $\mathbf{A}$. Let $\hat{\theta}_{MLE}$ and $\hat{\theta}$ satisfy*

$$\hat{\theta}_{MLE} = \arg\max\Big(\sum_{i=1}^{n} \log g(p^{-1}(\pi(X_i)), \theta)\Big),$$

$$\hat{\theta} = \arg\max\Big(\sum_{i=1}^{n} \log g(p^{-1}(\pi(\hat{X}_i)), \theta)\Big).$$

*Then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \to \mathcal{N}(0, I^{-1}(\theta_0)).$$

# Main Result: Theorem 4

## Theorem (Athreya et al., *Statistical Science*, forthcoming)

*Suppose $X_i$ i.i.d. $F_{\theta_0}$ are latent positions of a latent structure model satisfying the regularity assumptions delineated above. Let $\mathbf{A}$ be the adjacency matrix of the random dot product with latent positions $\mathbf{X}$, and let $\hat{\mathbf{X}}$ be the suitably-rotated adjacency spectral embedding of $\mathbf{A}$. Let $\hat{\theta}_{MLE}$ and $\hat{\theta}$ satisfy*

$$\hat{\theta}_{MLE} = \arg\max(\sum_{i=1}^{n} \log g(p^{-1}(\pi(X_i)), \theta)),$$

$$\hat{\theta} = \arg\max(\sum_{i=1}^{n} \log g(p^{-1}(\pi(\hat{X}_i)), \theta)).$$

*Then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \to \mathcal{N}(0, I^{-1}(\theta_0)).$$

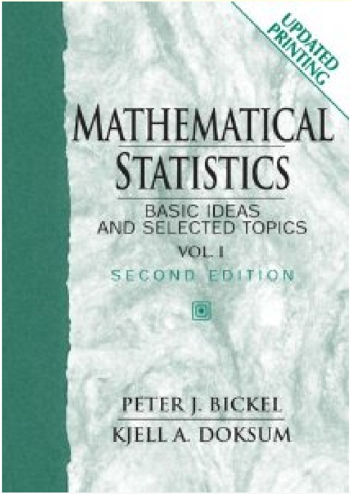
Proof:

Observe that $\sqrt{n}(\hat{\theta}_{MLE} - \theta_0)$ converges to a normal distribution with mean zero and variance $I^{-1}(\theta_0)$.

Thus, it remains to show that

$$\sqrt{n}(\hat{\theta} - \hat{\theta}_{MLE}) \to 0$$

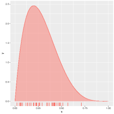
in probability ...

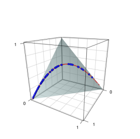


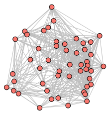

### 5.4.2 Asymptotic Normality of Minimum Contrast and $M$-Estimates

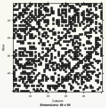
**Remark 5.4.2.** Solutions to (5.4.20) are called $M$-estimates as well as estimating equation estimates—see Section 2.2.1. Our arguments apply to $M$-estimates. Nothing in the arguments require that $\bar{\theta}_n$ be a minimum contrast as well as an $M$-estimate (i.e., that $\psi = \frac{\partial \rho}{\partial \theta}$ for some $\rho$).

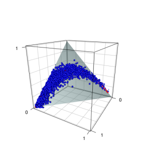
$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \log g_{\theta}(p_{\mathcal{C}}^{-1}(\pi_{\mathcal{C}}(\hat{X}_i)))$$

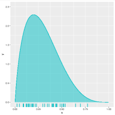










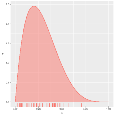
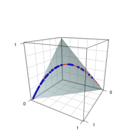
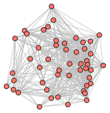
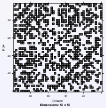
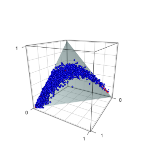
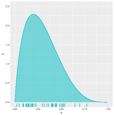


Top: *Unobserved* density on $[0, 1]$ and latent positions on $\mathcal{C}$.
Center: *Observed* graph generated and adjacency matrix.
Bottom: Estimated latent positions around $\mathcal{C}$ and $M$-estimate.

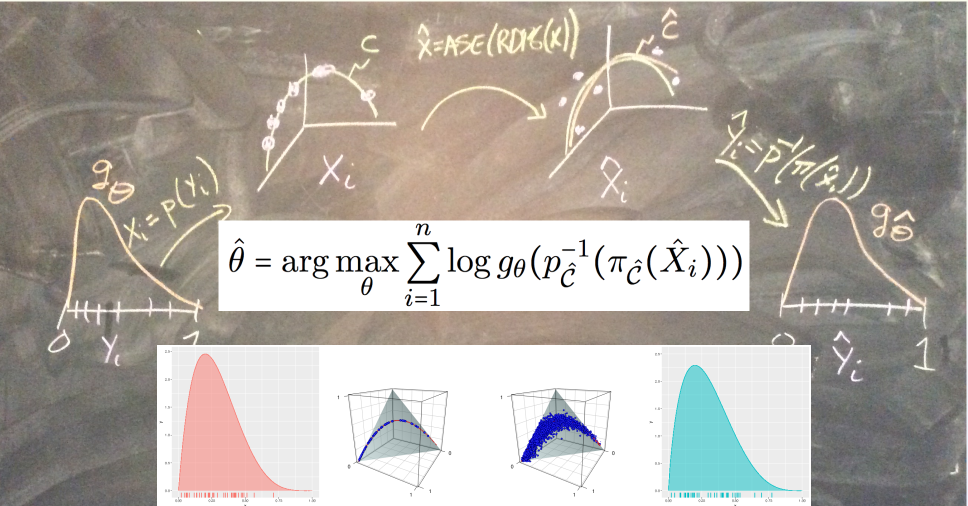
# Asymptotic Efficiency in the Latent Structure Model
# Beta($a = 1, b = 2$) H-W LSM

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \log g_{\theta}(p_{\mathcal{C}}^{-1}(\pi_{\mathcal{C}}(\hat{X}_i)))$$

| | (MSE($\widehat{a}$), MSE($\widehat{b}$)) | |
|---|---|---|
| | $n = 1000$ | $n = 8000$ |
| using $\mathbf{X}$ | (0.00068, 0.00280) | (0.00014,0.00097) |
| using $\hat{\mathbf{X}}$ | (0.004,0.019) | (0.00015,0.00120) |

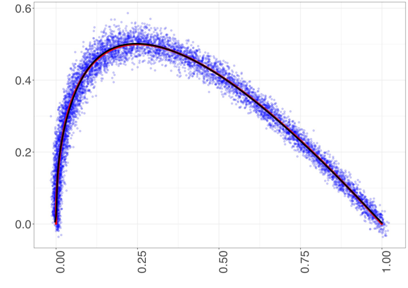







# Manifold Learning for Subsequent Inference: Parametric Rate of Convergence in the Latent Structure Model




$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \log g_\theta(p_{\hat{\mathcal{C}}}^{-1}(\pi_{\hat{\mathcal{C}}}(\hat{X}_i)))$$

# Manifold Learning for Subsequent Inference: Beta($a = 1, b = 1$) H-W LSM ($n = 8000$)



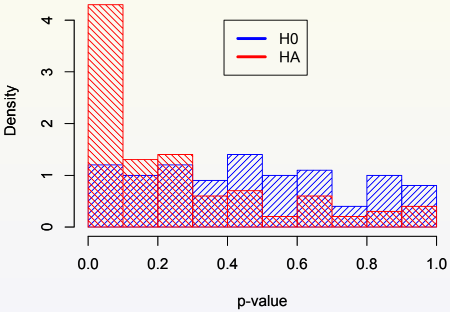

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \log g_{\theta}(p_{\hat{\mathcal{C}}}^{-1}(\pi_{\hat{\mathcal{C}}}(\hat{X}_i)))$$

| | MSE($\widehat{a}$) | MSE($\widehat{b}$) |
|---|---|---|
| using $\mathbf{X}$ and $\mathcal{C}$ | 0.00015 | 0.00008 |
| using $\hat{\mathbf{X}}$ and $\mathcal{C}$ | 0.00023 | 0.00010 |
| using $\hat{\mathbf{X}}$ and $\hat{\mathcal{C}}$ | 0.0011 | 0.0011 |

# two-sample testing

- $G$ parametric & $\mathcal{C}$ {known ; parametric}
  above: theory & simulation

- $G$ nonparametric & $\mathcal{C}$ {known ; parametric ; nonparametric}
  simulation results:

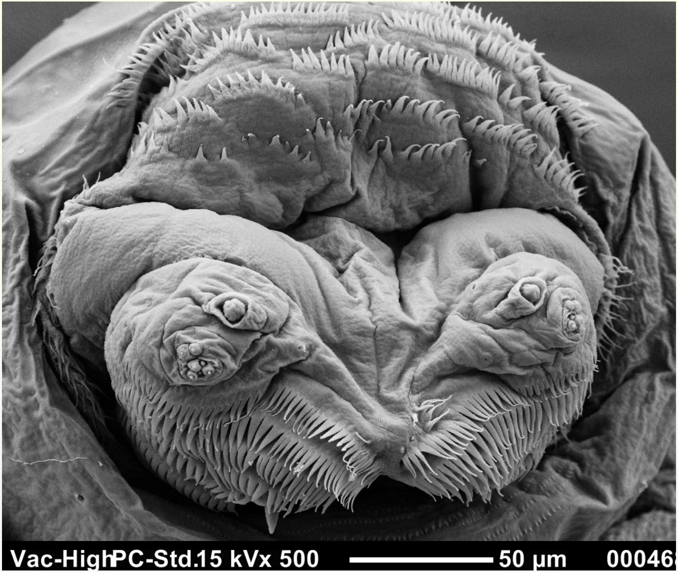



- $G$ nonparametric & $\mathcal{C}$ nonparametric
  below: real data = connectome re bilateral homology

## Yogi Berra (purportedly):

*"In theory there is no difference between theory and practice. In practice, there is."*

(cf. *"That's all well and good in practice, but how does it work in theory?"*)




scanning electron microscope image of a Drosophila larva face
(Tigran Norekian @ Friday Harbor Lab,
in collaboration with Ben Cocanougher & Leonid Moroz)

# Connectome

nature
International journal of science

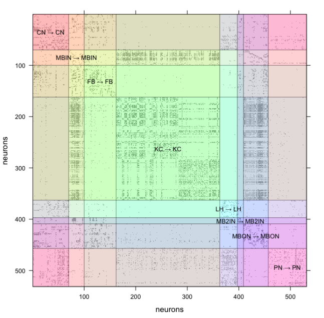
PDF

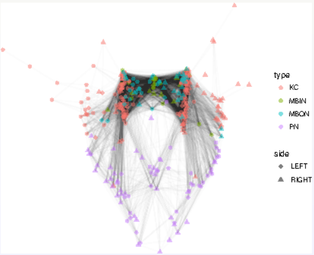
Article | Published: 09 August 2017

## The complete connectome of a learning and memory centre in an insect brain

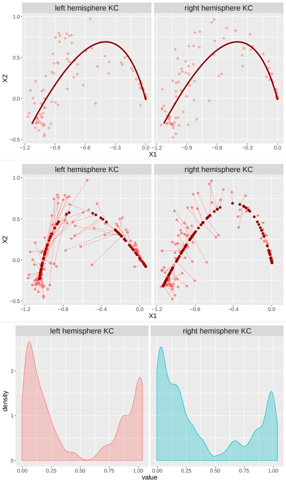
Katharina Eichler, Feng Li, Ashok Litwin-Kumar, Youngser Park, Ingrid Andrade, Casey M. Schneider-Mizell, Timo Saumweber, Annina Huser, Claire Eschbach, Bertram Gerber, Richard D. Fetter, James W. Truman, Carey E. Priebe, L. F. Abbott, Andreas S. Thum, Marta Zlatic & Albert Cardona

*Nature* **548**, 175–182 (10 August 2017) | Download Citation



# Bilateral Homology

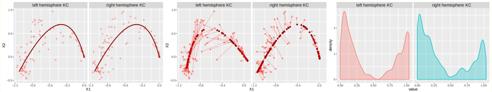



Top: Estimated structural support.

Center: Projection onto this estimated support.

Bottom: Density estimates for the underlying distribution.

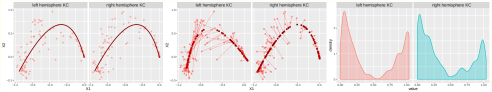
# Bilateral Homology




Purpose of work from the perspective of a biologist:
The mushroom body is the seat of memory for the Drosophila larva. The structure of this portion of the brain is well studied. It is assumed that the left and right side of the mushroom body are bilaterally homologous, that is, for each neuron that appears in the mushroom body on the left, there is an equivalent and mirror symmetrical neuron on the right in function and morphology. In practice, such perfect symmetry is not observed, but this is believed to be due to developmental history (small differences in environment experienced by each neuron in the animal and at a larger scale by the experience of each animal, which influence neural structure and connectivity).

The conjecture that the left and right side are equivalent is a convenience that doesn't prevent restful sleep for neuroscientists but does keep statisticians up at night. Testing whether the left and right side are truly structurally similar is, mathematically, a difficult problem. Furthermore, in order to detect changes in the left and right side of a mushroom body (in the case, for example, where a memory is written into the left but not right side of a brain and needs to be detected), a hypothesis test needs to be developed and implemented.

# Bilateral Homology

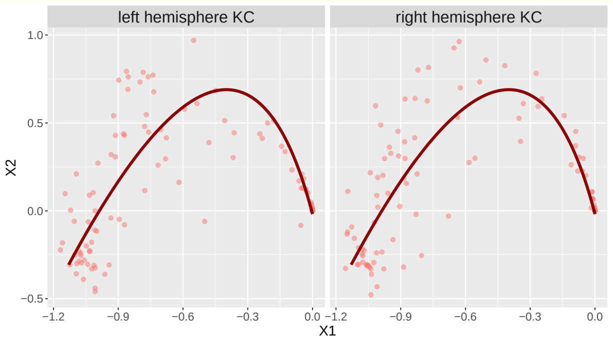


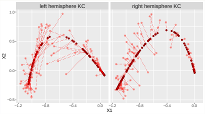

Purpose of work from the perspective of a biologist:

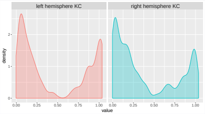
The mushroom body is the seat of memory for the Drosophila larva. The structure of this portion of the brain is well studied. It is assumed that the left and right side of the mushroom body are bilaterally homologous, that is, for each neuron that appears in the mushroom body on the left, there is an equivalent and mirror symmetrical neuron on the right in function and morphology. In practice, such perfect symmetry is not observed, but this is believed to be due to developmental history (small differences in environment experienced by each neuron in the animal and at a larger scale by the experience of each animal, which influence neural structure and connectivity).

The conjecture that the left and right side are equivalent is a convenience that doesn't prevent restful sleep for neuroscientists but does keep statisticians up at night. Testing whether the left and right side are truly structurally similar is, mathematically, a difficult problem. Furthermore, in order to detect changes in the left and right side of a mushroom body (in the case, for example, where a memory is written into the left but not right side of a brain and needs to be detected), a hypothesis test needs to be developed and implemented.

# Bilateral Homology

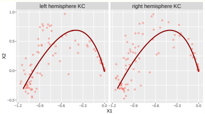


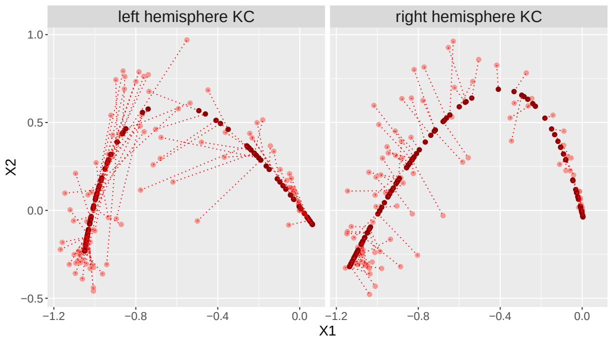



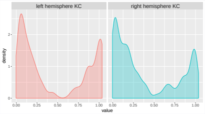




Top: Estimated structural support.

# Bilateral Homology

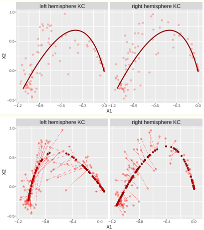


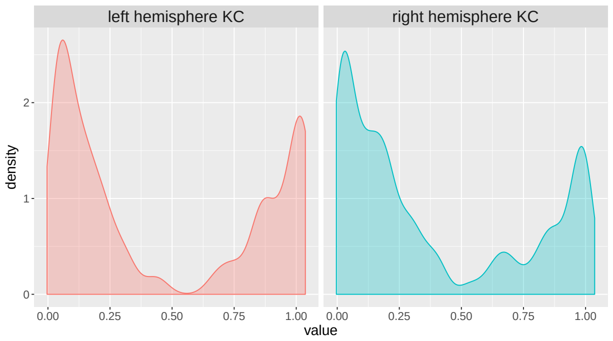




Center: Projection onto this estimated support.

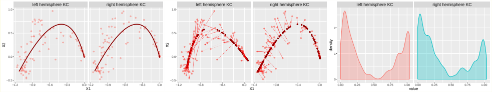
# Bilateral Homology




Bottom: Density estimates for the underlying distribution.

# Bilateral Homology




## Purpose of work from the perspective of a biologist:

The mushroom body is the seat of memory for the Drosophila larva. The structure of this portion of the brain is well studied. It is assumed that the left and right side of the mushroom body are bilaterally homologous, that is, for each neuron that appears in the mushroom body on the left, there is an equivalent and mirror symmetrical neuron on the right in function and morphology. In practice, such perfect symmetry is not observed, but this is believed to be due to developmental history (small differences in environment experienced by each neuron in the animal and at a larger scale by the experience of each animal, which influence neural structure and connectivity). The conjecture that the left and right side are equivalent is a convenience that doesn't prevent restful sleep for neuroscientists but does keep statisticians up at night. Testing whether the left and right side are truly structurally similar is, mathematically, a difficult problem. Furthermore, in order to detect changes in the left and right side of a mushroom body (in the case, for example, where a memory is written into the left but not right side of a brain and needs to be detected), a hypothesis test needs to be developed and implemented.

In summary, this LSM work promises two neuroscience advances. First, it gives mathematical grounding to the belief that the left and right hemisphere of the mushroom body consist of bilaterally homologous neurons.
Second, it provides a suitable hypothesis test for future experiments in which the structure of the mushroom body is altered by experience (such as the storage of long term memory which differently affects structure on the left and right side).

http://arxiv.org/abs/1806.01401

Submitted to Statistical Science

# On estimation and inference in latent structure random graphs

**Avanti Athreya, Minh Tang, Younger Park, and Carey E. Priebe**

Department of Applied Mathematics & Statistics, Johns Hopkins University



Key Contribution 1: Latent Structure Model
Key Contribution 2: $2 \to \infty$ & Donsker CLT $\implies$ Efficiency
Key Contribution 3: Manifold Learning for Subsequent Inference

Acknowledgment: DARPA FunLoL via SIMPLEX

## Leopold Kronecker to Hermann von Helmholtz (1888):

*"The wealth of your practical experience*
*with sane and interesting problems*
*will give to mathematics*
*a new direction and a new impetus."*


Kronecker

Helmholtz