Scoring Bayesian Networks of Mixed Variables

Bryan Andrews, MS
Joesph Ramsey, PhD
and
Greg Cooper, MD, PhD

August 14, 2017

Learning Bayesian Networks (BNs)

- BNs constitute a widely used graphical framework for representing probabilistic relationships
- Many application in Bayesian Inference and Causal Discovery
- Learning structure is crucial
  - Limited work has been done in the presence of both discrete and continuous variables

Learning Bayesian Networks (BNs)

- BNs constitute a widely used graphical framework for representing probabilistic relationships
- Many application in Bayesian Inference and Causal Discovery
- Learning structure is crucial
  - Limited work has been done in the presence of both discrete and continuous variables

Goal: Provide scalable solutions for learning BNs in the presence of both discrete and continuous variables

Outline

- Bayesian Information Criterion (BIC)
- Mixed Variable Polynomial (MVP) score
- Conditional Gaussian (CG) score
- Adaptations
- Simulations and empirical results

Outline

- Bayesian Information Criterion (BIC)
- Mixed Variable Polynomial (MVP) score
- Conditional Gaussian (CG) score
- Adaptations
- Simulations and empirical results

The Bayesian Information Criterion

Let M be a model and D be a dataset

BIC is an approximation for log $p(M|D)$

The Bayesian Information Criterion

Let M be a model and D be a dataset

BIC is an approximation for log $p(M|D)$

$$\log p(M|D) \approx -2\, lik + dof \log n$$

Where *lik* is the log likelihood, *dof* are the degrees of freedom, and *n* is the number of samples

The Bayesian Information Criterion

Let M be a model and D be a dataset

BIC is an approximation for log $p(M|D)$

$$\log p\left(M|D\right) \approx -2\, lik + dof\, \log n$$

Where *lik* is the log likelihood, *dof* are the degrees of freedom, and *n* is the number of samples

Scores a BN as the sum over all BIC calculations for each node given its parents

Outline

- Bayesian Information Criterion (BIC)
- Mixed Variable Polynomial (MVP) score
- Conditional Gaussian (CG) score
- Adaptations
- Simulations and empirical results

The Mixed Variable Polynomial (MVP) score

- Use higher order polynomials to estimate relationships between variables
  - Allows for nonlinear relationships between continuous variables
  - Allows for complicated PMFs for discrete variables
    - Approximates Logistic Regression

The Mixed Variable Polynomial (MVP) score

- Use higher order polynomials to estimate relationships between variables
  - Allows for nonlinear relationships between continuous variables
  - Allows for complicated PMFs for discrete variables
    - Approximates Logistic Regression
- Calculate a log-likelihood and degrees of freedom for BIC

Modeling a Continuous Child

- Partition according to the discrete parents
  - Splits the data into subsets

Modeling a Continuous Child

- Partition according to the discrete parents
  - Splits the data into subsets
- Perform regression with the continuous parents for each partition
  - Calculate a log likelihood and degrees of freedom for each subset

Modeling a Continuous Child

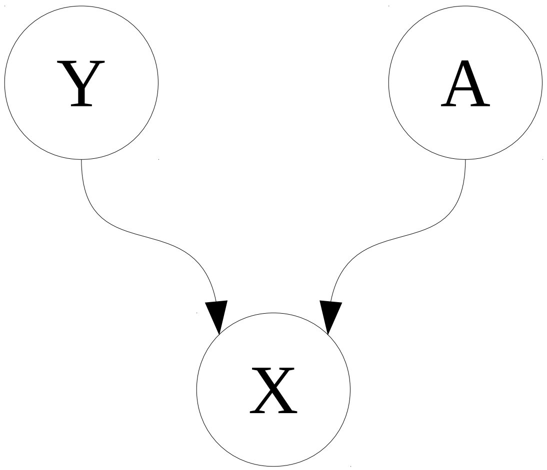
- Partition according to the discrete parents
  - Splits the data into subsets
- Perform regression with the continuous parents for each partition
  - Calculate a log likelihood and degrees of freedom for each subset
- Aggregate the log likelihood and degrees of freedom terms from each subset together

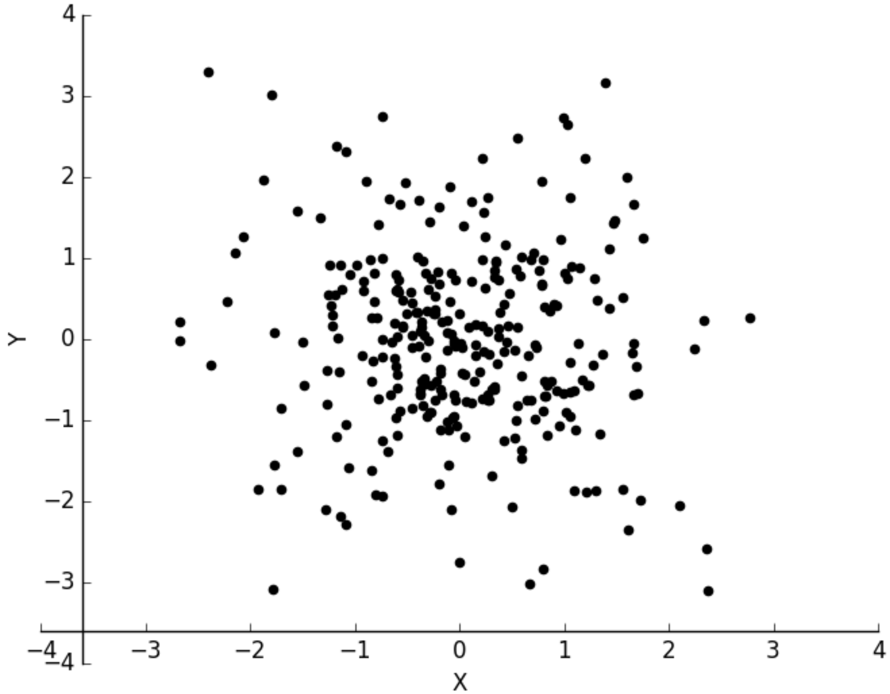
Modeling a Continuous Child

- Partition according to the discrete parents
  - Splits the data into subsets
- Perform regression with the continuous parents for each partition
  - Calculate a log likelihood and degrees of freedom for each subset
- Aggregate the log likelihood and degrees of freedom terms from each subset together
- Score continuous child using BIC

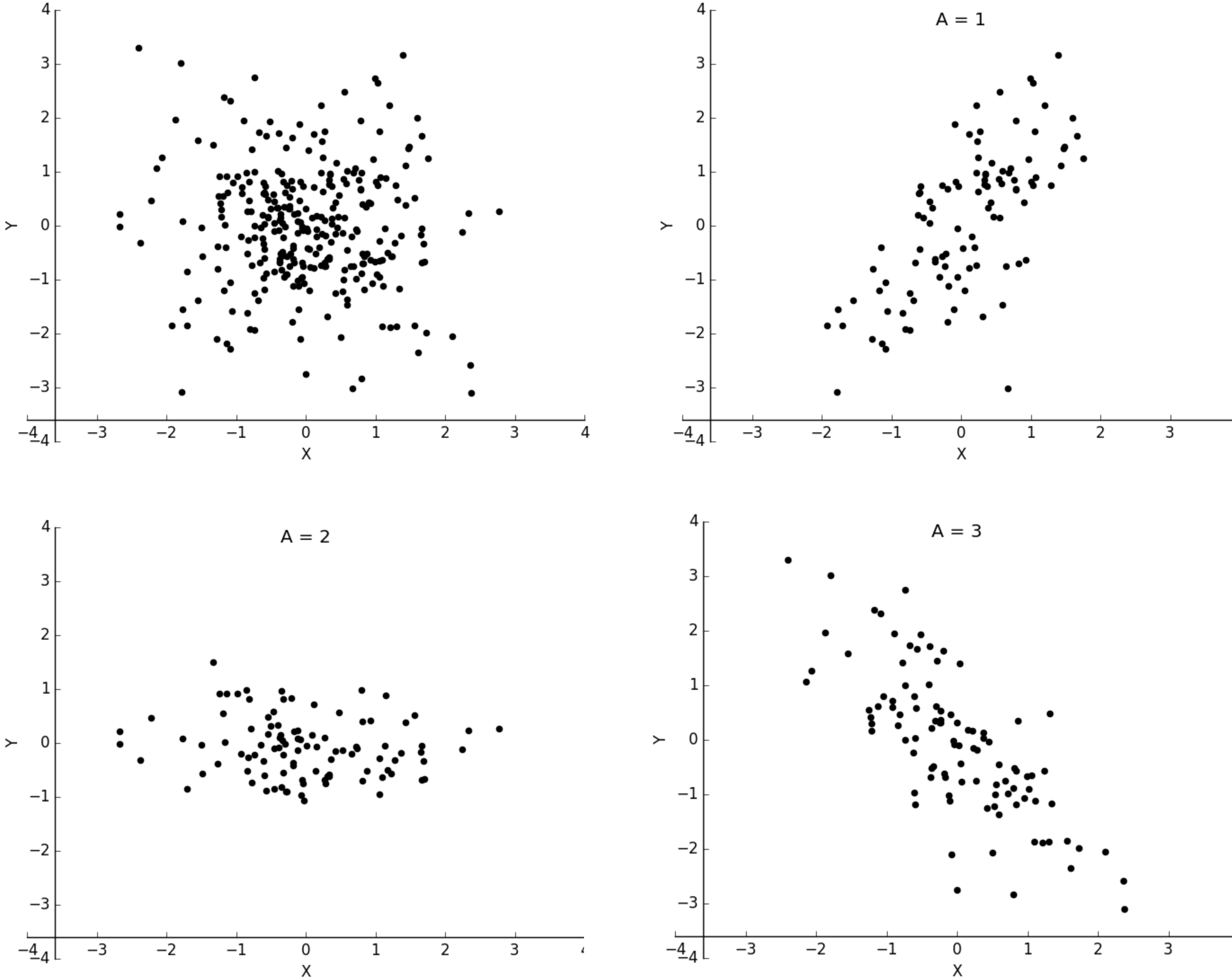
Modeling a Continuous Child

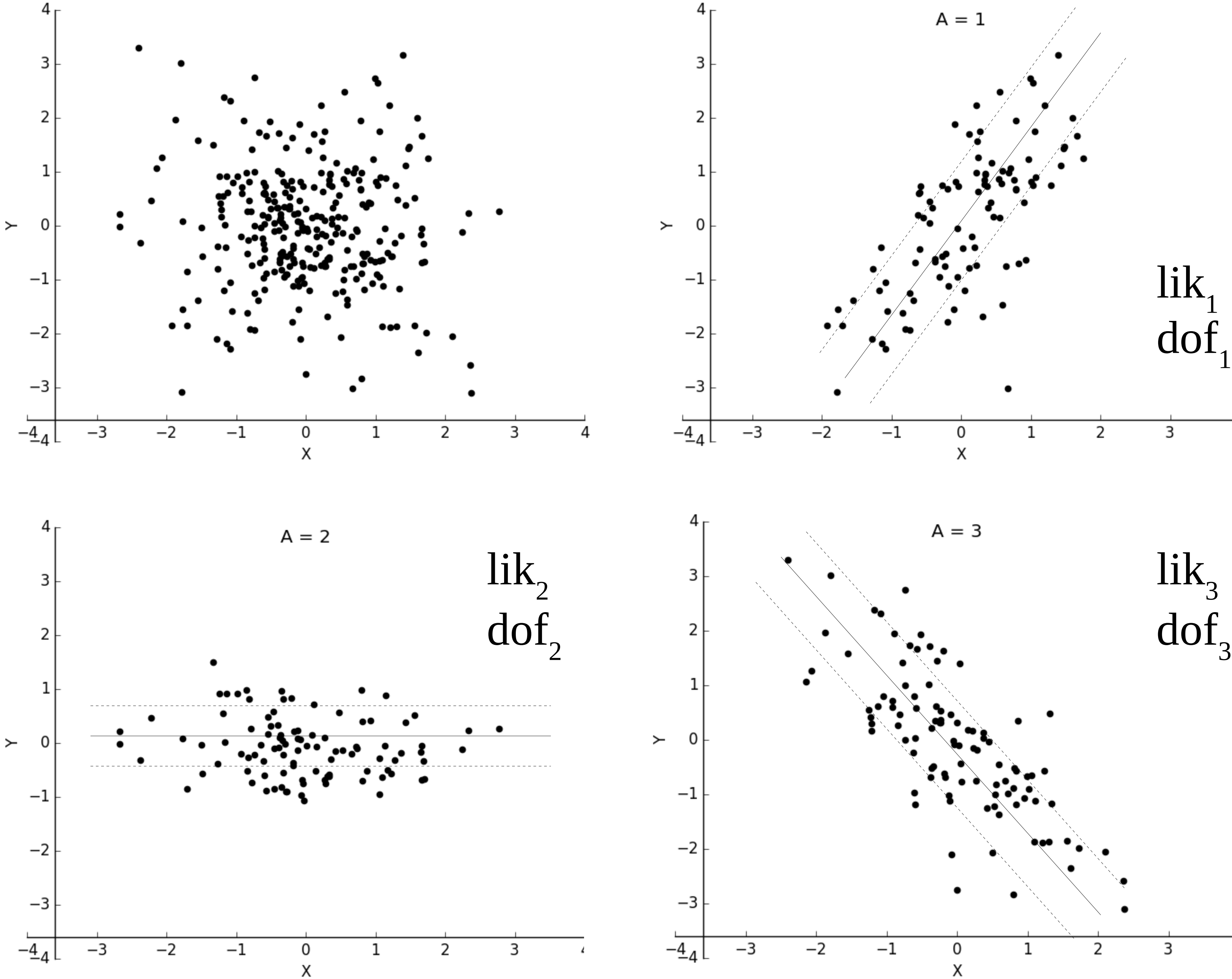
- Let X, Y be continuous
- Let A be discrete (|A| = 3)
- Want: $\text{lik}_{X \mid Y, A}$, $\text{dof}_{X \mid Y, A}$





4
3
2
1
0
−1
−2
−3
−4
Y
−4
−3
−2
−1
0
1
2
3
4
X

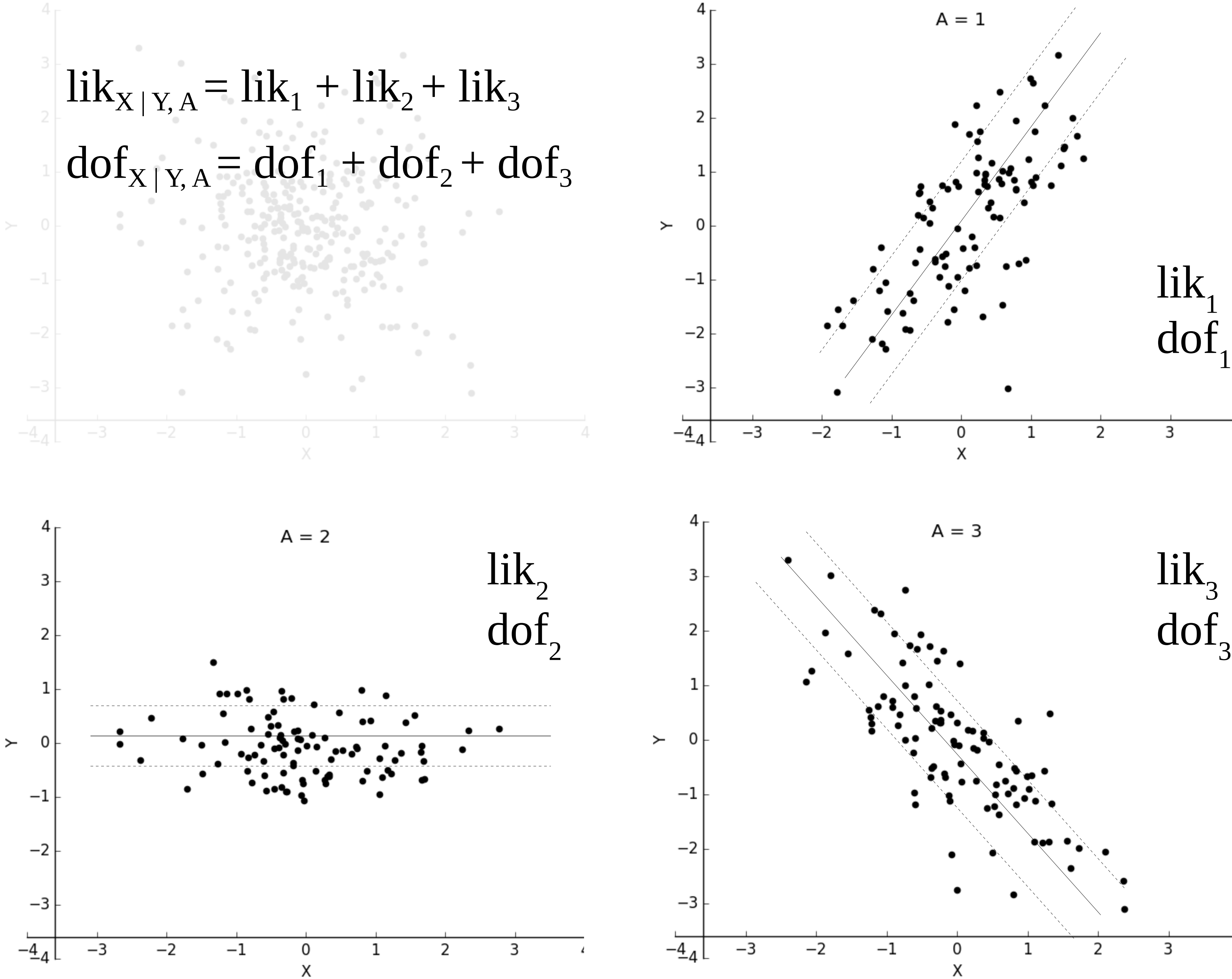

A = 1
A = 2
A = 3
X
Y



A = 1
lik1
dof1
A = 2
lik2
dof2
A = 3
lik3
dof3
X
Y

$$\text{lik}_{X\,|\,Y,A} = \text{lik}_1 + \text{lik}_2 + \text{lik}_3$$

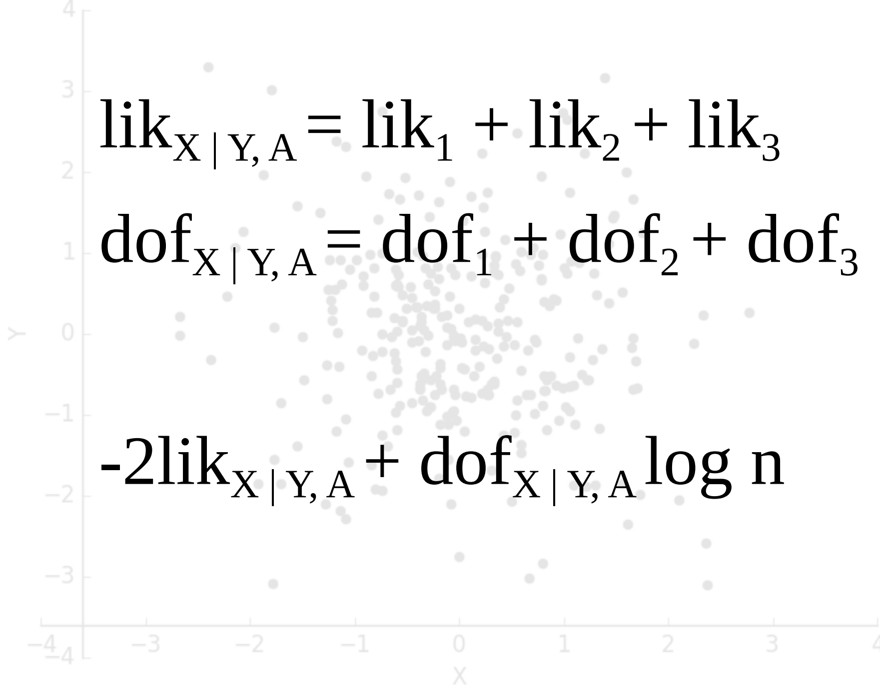
$$\text{dof}_{X\,|\,Y,A} = \text{dof}_1 + \text{dof}_2 + \text{dof}_3$$

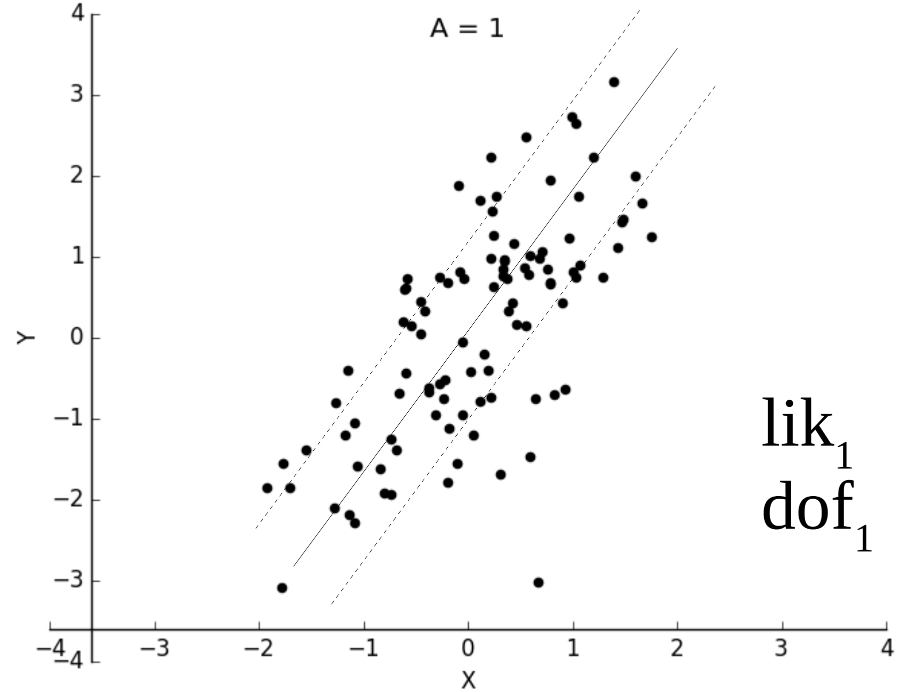


$$\text{lik}_{X\mid Y,A} = \text{lik}_1 + \text{lik}_2 + \text{lik}_3$$

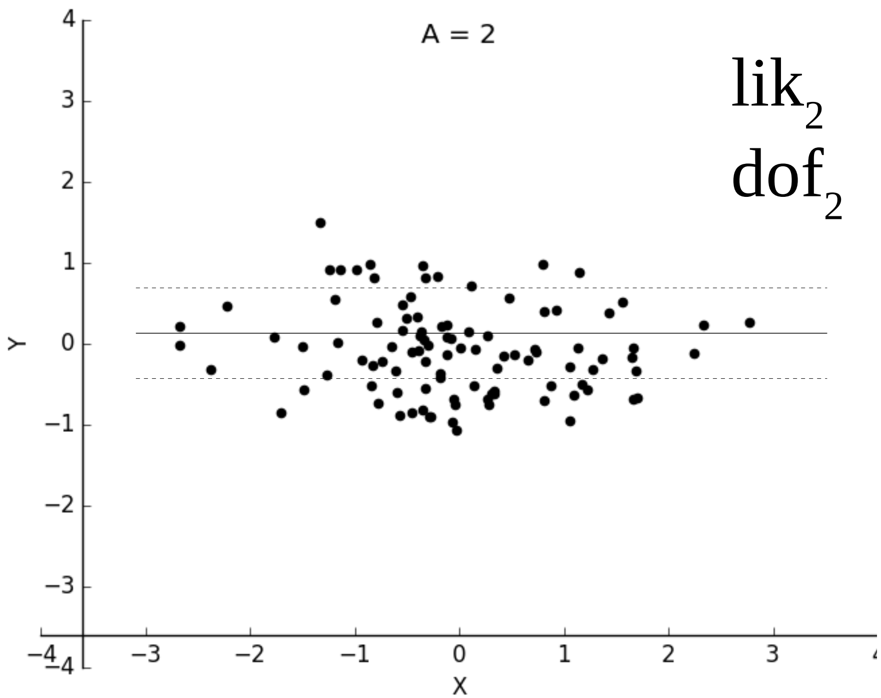
$$\text{dof}_{X\mid Y,A} = \text{dof}_1 + \text{dof}_2 + \text{dof}_3$$

$$-2\text{lik}_{X\mid Y,A} + \text{dof}_{X\mid Y,A} \log n$$

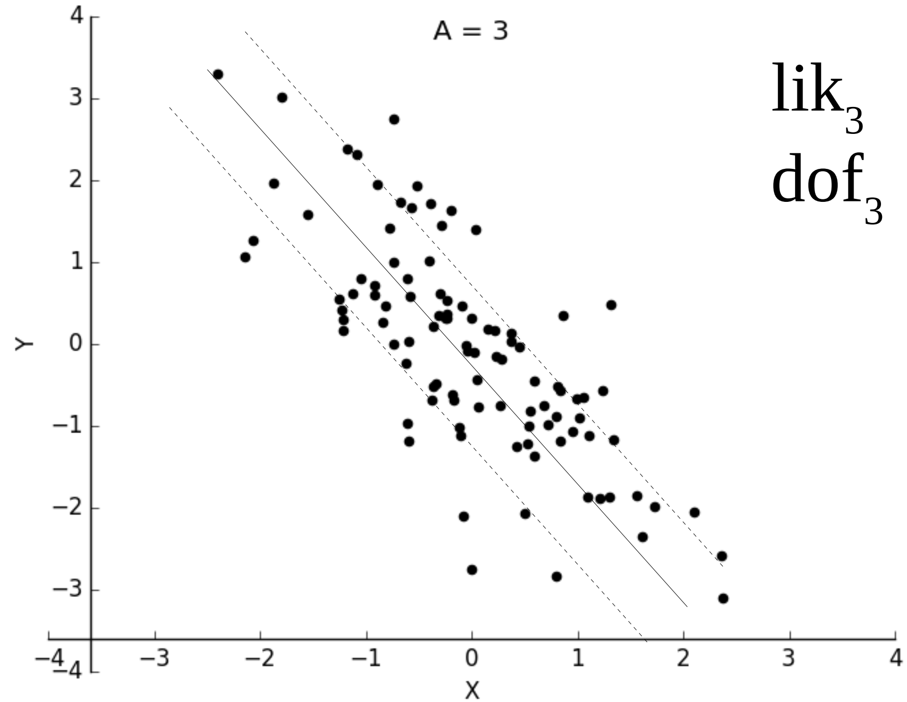




lik_1
dof_1



lik_2
dof_2



lik_3
dof_3

Modeling a Discrete Child

- Binarize the child A into d (0, 1) variables where $d = |A|$

Modeling a Discrete Child

- Binarize the child A into d (0, 1) variables where $d = |A|$
- Partition according to the discrete parents
  - Splits the data into subsets

Modeling a Discrete Child

- Binarize the child *A* into *d* (0, 1) variables where $d = |A|$
- Partition according to the discrete parents
  - Splits the data into subsets
- Perform regression with the continuous parents for each partition
  - Treat the regression lines a components to PMFs for *A*

Modeling a Discrete Child

- Binarize the child A into d (0, 1) variables where $d = |A|$
- Partition according to the discrete parents
  - Splits the data into subsets
- Perform regression with the continuous parents for each partition
  - Treat the regression lines a components to PMFs for A
- Calculate a log likelihood and degrees of freedom for each subset

Modeling a Discrete Child

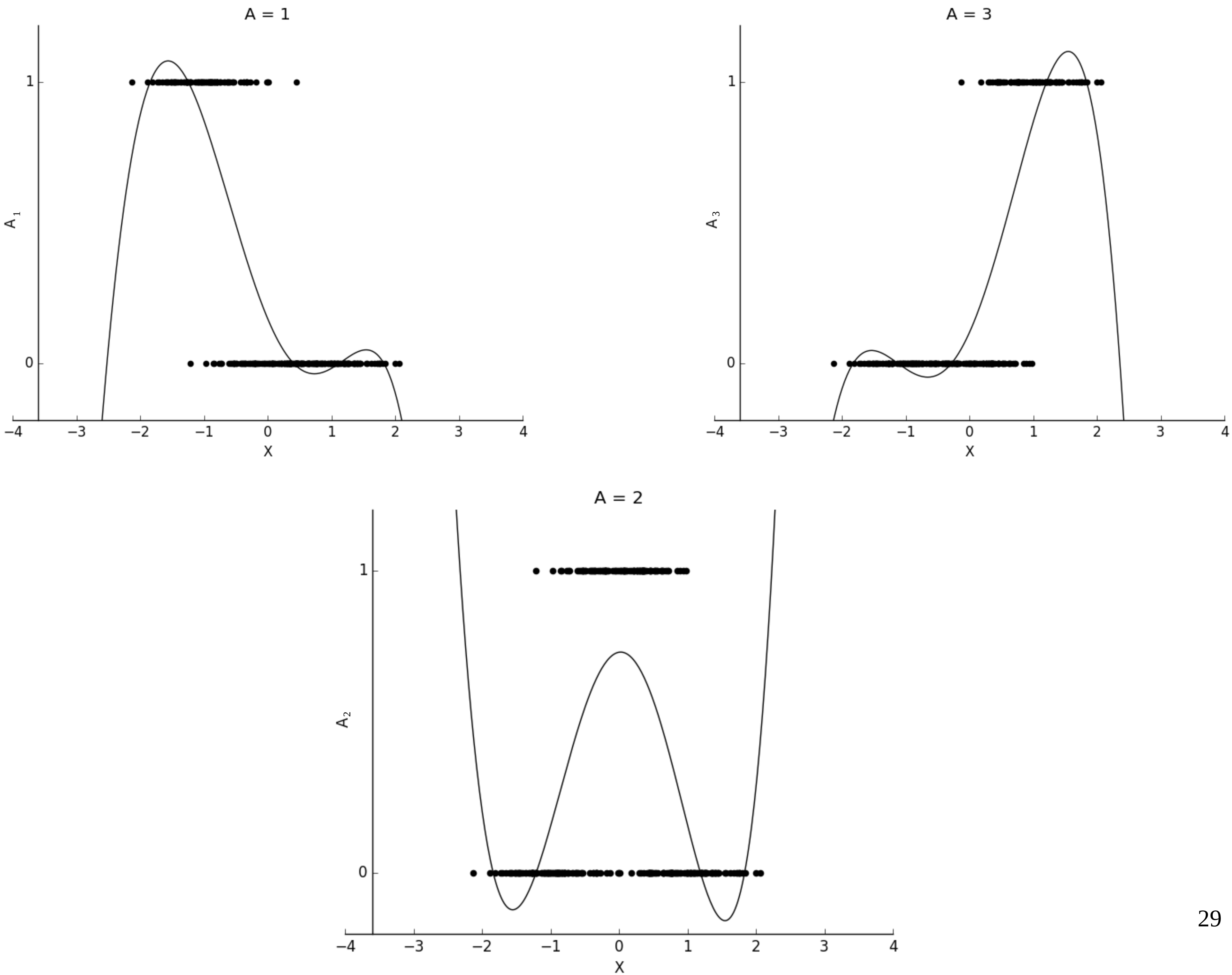
- Binarize the child A into d (0, 1) variables where $d = |A|$
- Partition according to the discrete parents
  - Splits the data into subsets
- Perform regression with the continuous parents for each partition
  - Treat the regression lines a components to PMFs for A
- Calculate a log likelihood and degrees of freedom for each subset
- Aggregate the log likelihood and degrees of freedom terms from each subset together

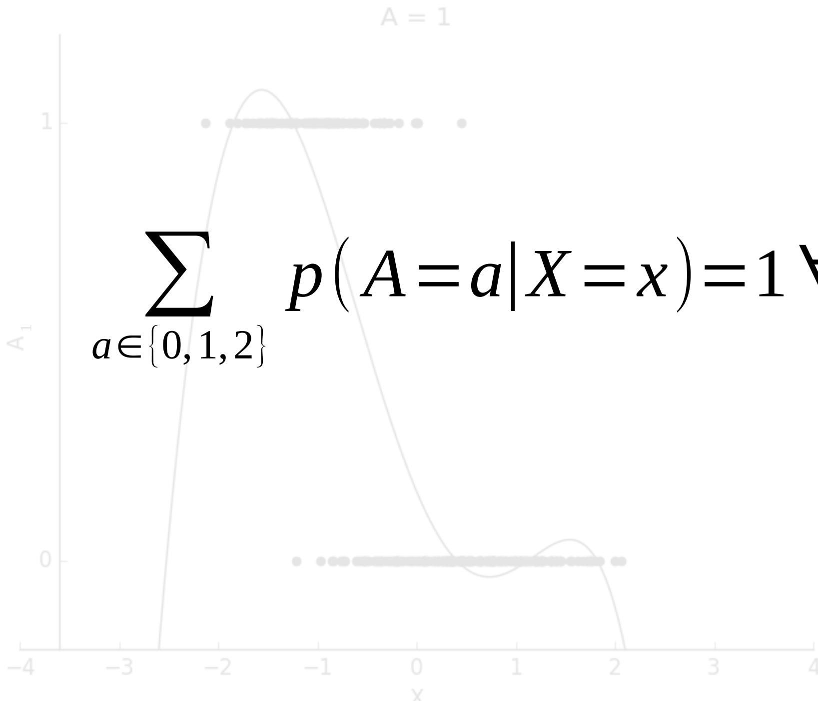
Modeling a Discrete Child

- Binarize the child A into d (0, 1) variables where $d = |A|$
- Partition according to the discrete parents
  - Splits the data into subsets
- Perform regression with the continuous parents for each partition
  - Treat the regression lines a components to PMFs for A
- Calculate a log likelihood and degrees of freedom for each subset
- Aggregate the log likelihood and degrees of freedom terms from each subset together
- Score discrete child using BIC

Modeling a Discrete Child

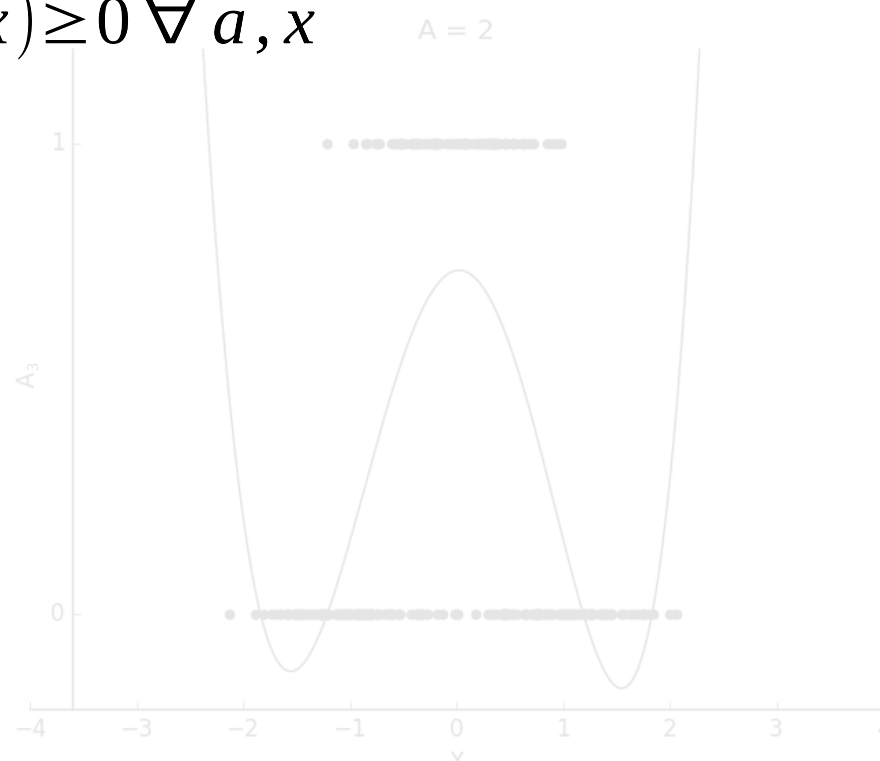
- Let X be continuous
- Let A be discrete ($|A| = 3$)
- Want: $lik_{A|X}$, $dof_{A|X}$

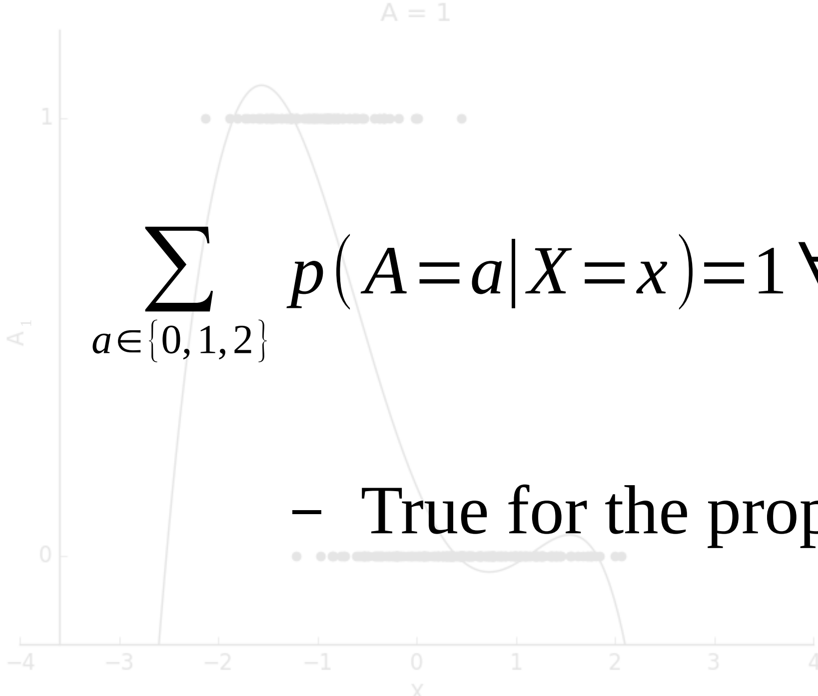

A = 1
A₁
X
−4 −3 −2 −1 0 1 2 3 4
0
1
A = 3
A₃
X
−4 −3 −2 −1 0 1 2 3 4
0
1
A = 2
A₂
X
−4 −3 −2 −1 0 1 2 3 4
0
1

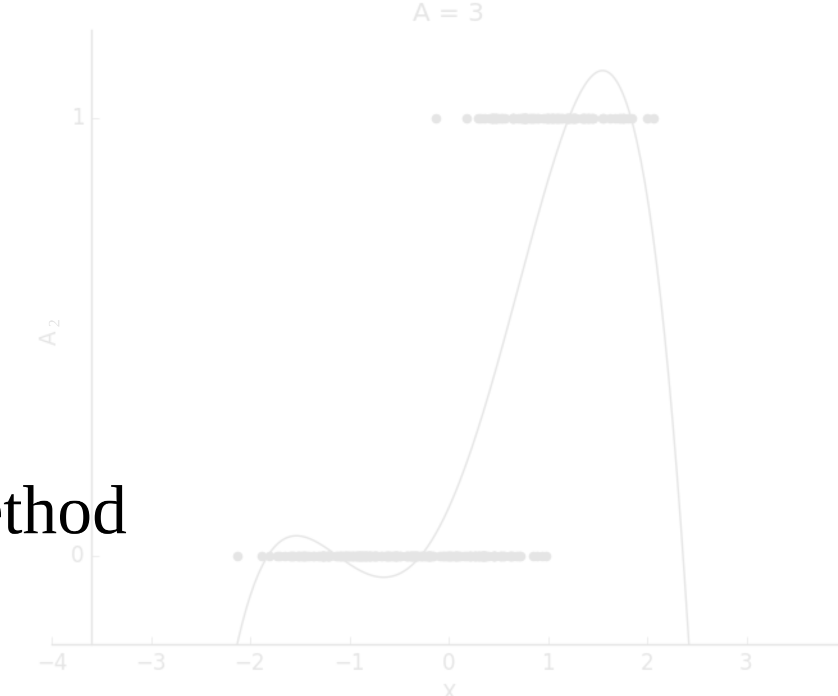


$$\sum_{a \in \{0,1,2\}} p(A=a|X=x)=1 \; \forall \, x$$

$$p(A=a|X=x) \geq 0 \; \forall \, a, x$$



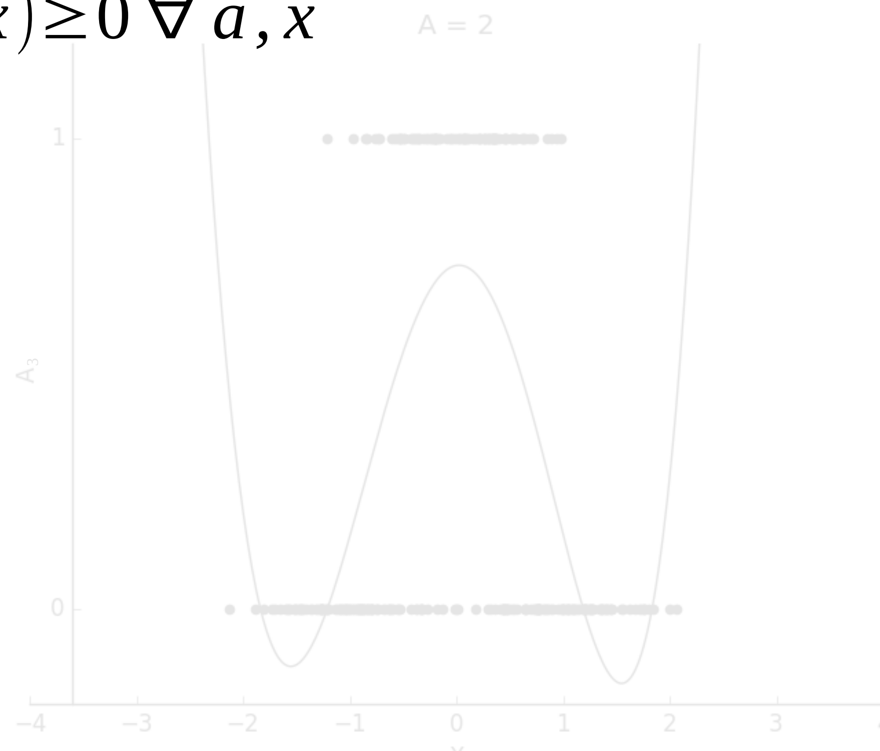


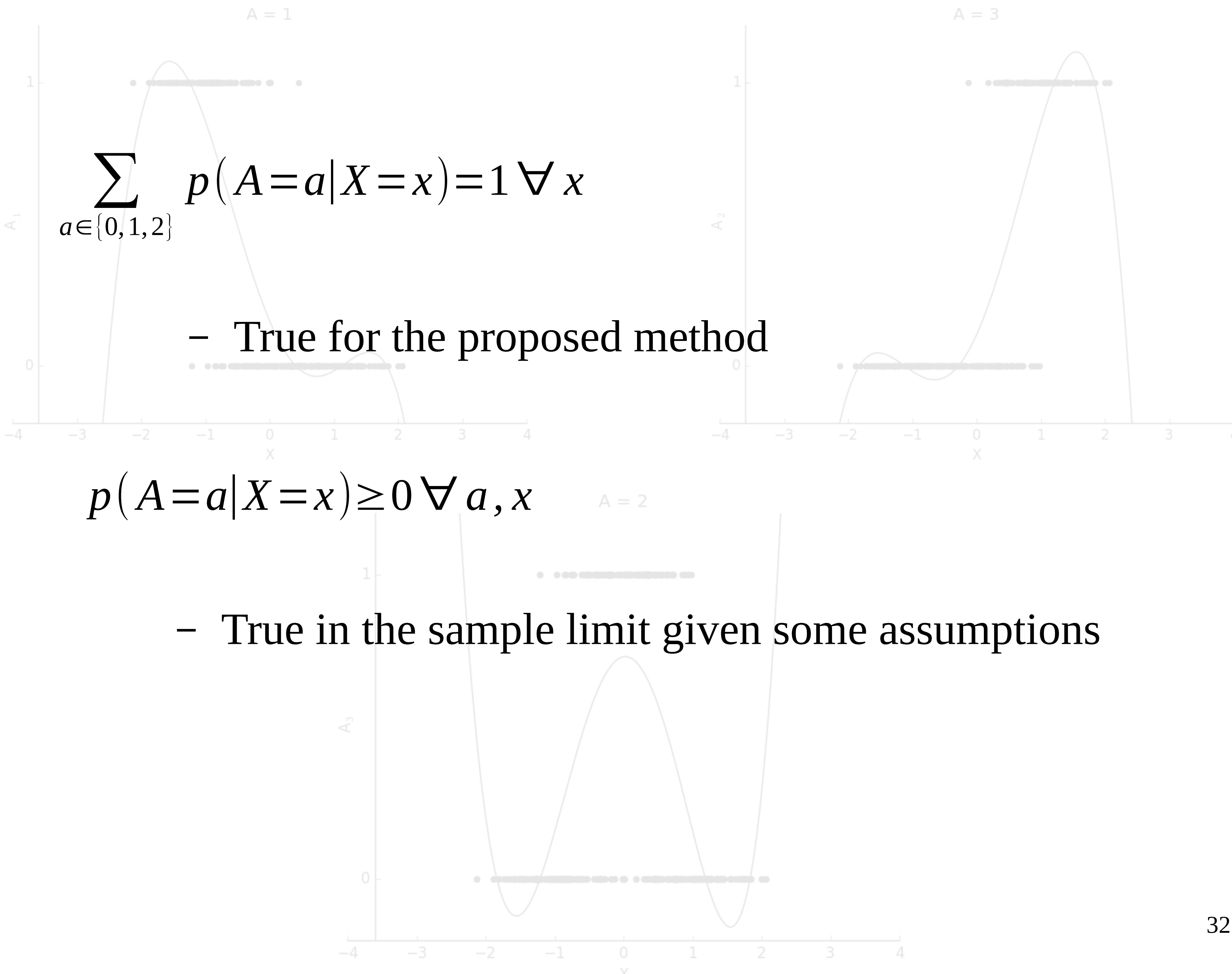


$$\sum_{a \in \{0,1,2\}} p(A=a|X=x) = 1 \, \forall \, x$$

– True for the proposed method

$$p(A=a|X=x) \geq 0 \, \forall \, a, x$$





$$\sum_{a \in \{0,1,2\}} p\left(A=a \mid X=x\right)=1 \;\forall\, x$$

– True for the proposed method

$$p\left(A=a \mid X=x\right) \geq 0 \;\forall\, a, x$$

– True in the sample limit given some assumptions

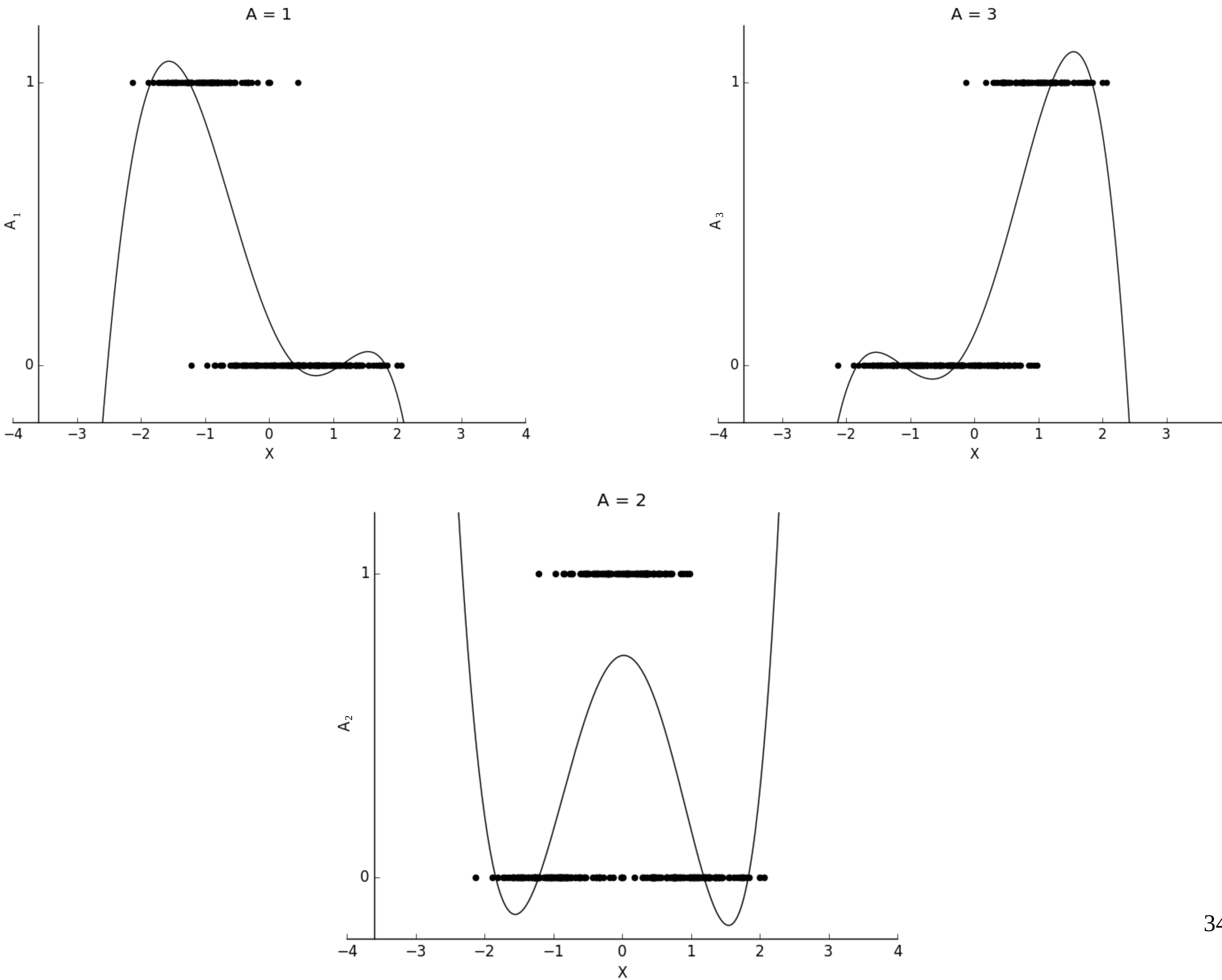
$$\sum_{a \in \{0,1,2\}} p(A=a|X=x)=1 \; \forall x$$

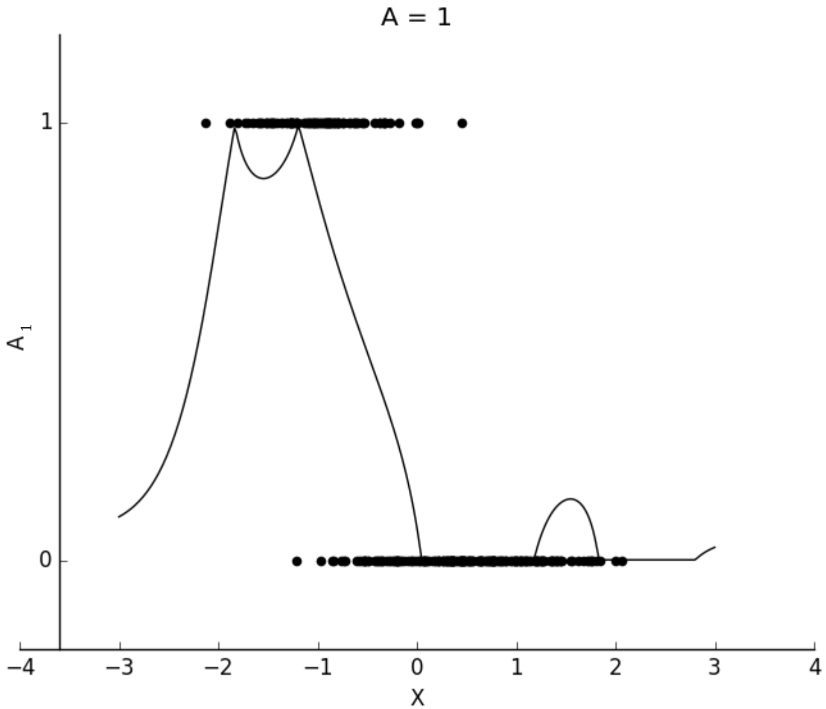
– True for the proposed method

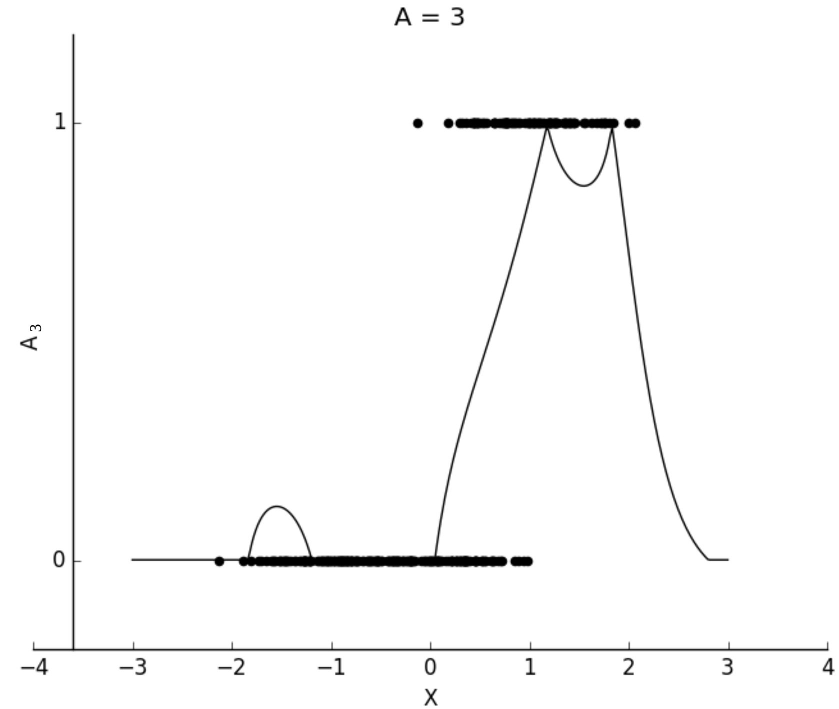
$$p(A=a|X=x) \geq 0 \; \forall a, x$$

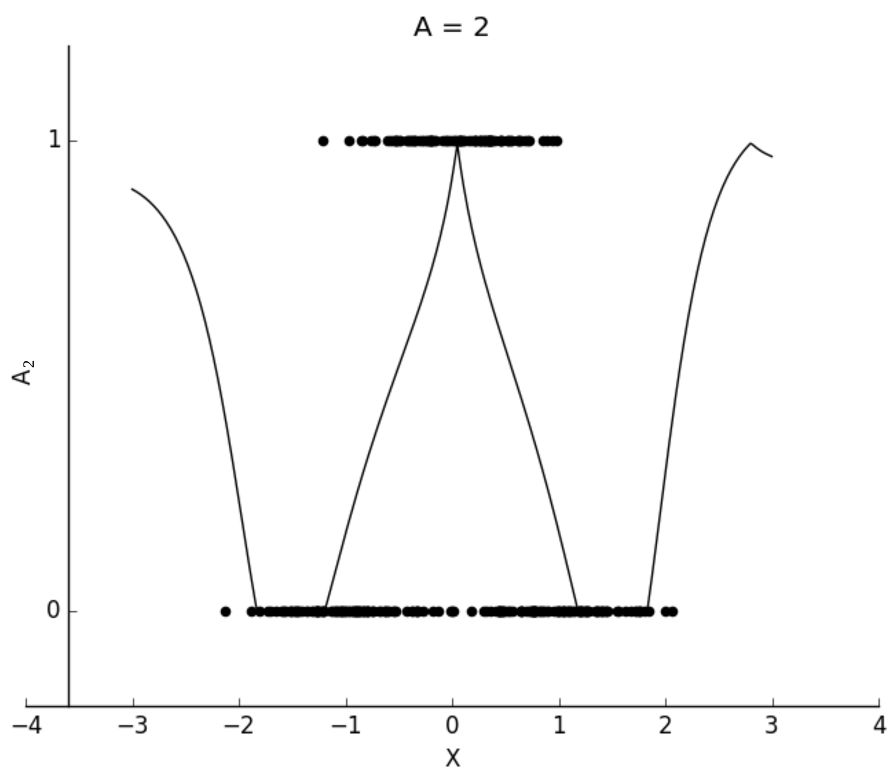
– True in the sample limit given some assumptions

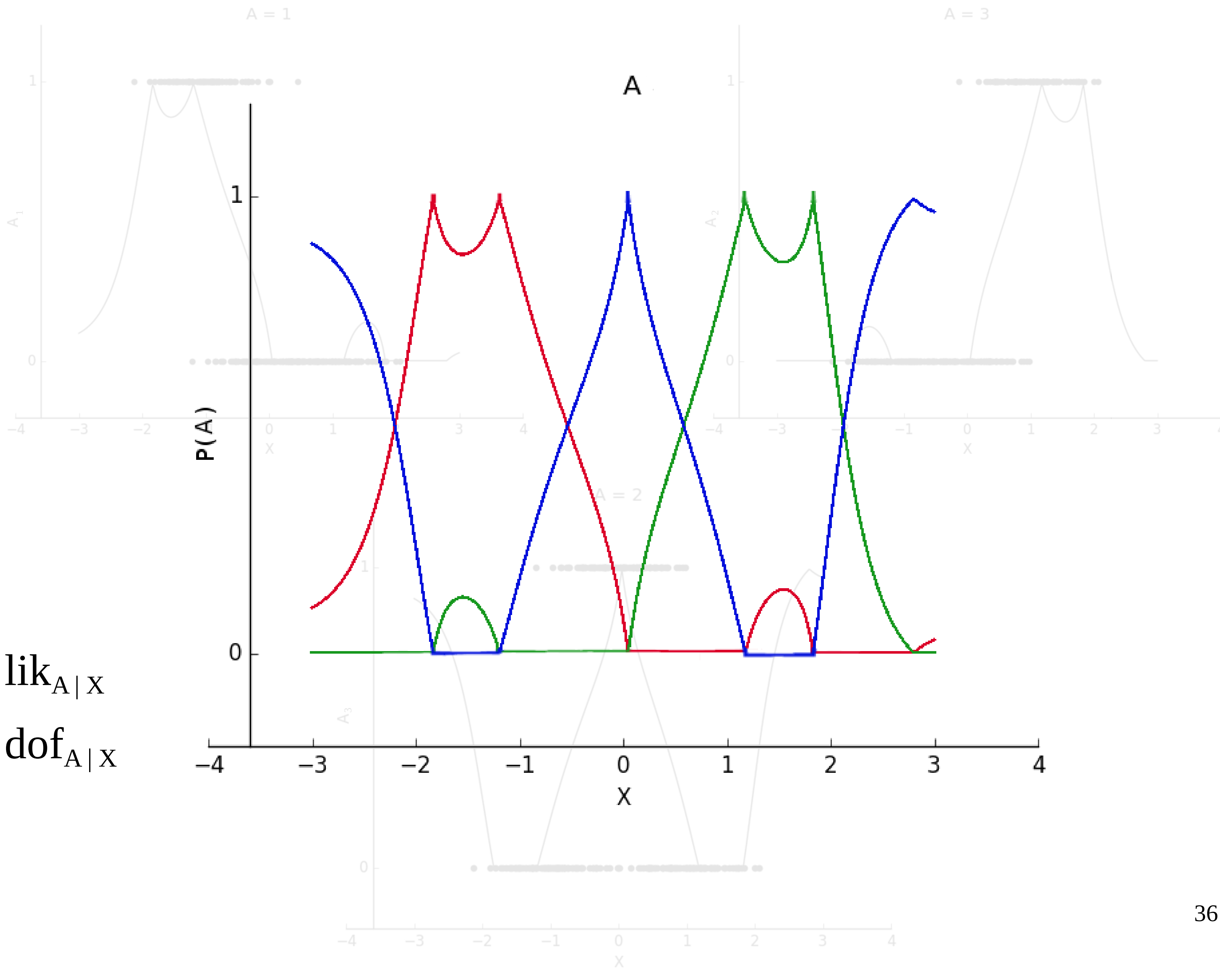
Define a procedure to shrink illegal distributions back into the domain of probabilities


A = 1
A₁
X
A = 3
A₃
X
A = 2
A₂
X


A = 1
1
0
A_1
−4
−3
−2
−1
0
1
2
3
4
X

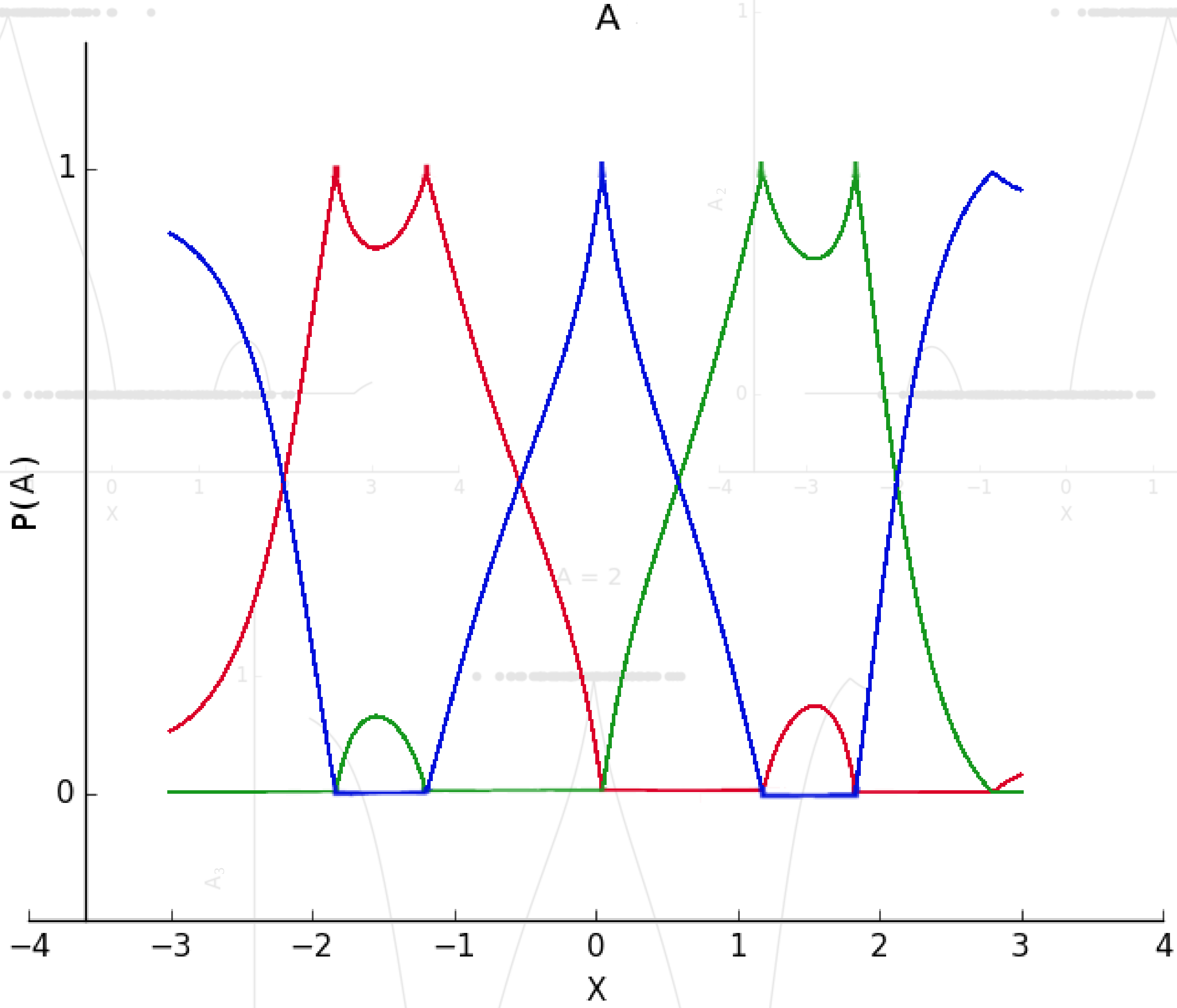

A = 3
1
0
A_3
−4
−3
−2
−1
0
1
2
3
4
X


A = 2
1
0
A_2
−4
−3
−2
−1
0
1
2
3
4
X



$\text{lik}_{A\,|\,X}$

$\text{dof}_{A\,|\,X}$



$lik_{A|X}$

$dof_{A|X}$

$-2lik_{A|X} + dof_{A|X} \log n$

Outline

- Bayesian Information Criterion (BIC)
- Mixed Variable Polynomial (MVP) score
- Conditional Gaussian (CG) score
- Adaptations
- Simulations and empirical results

The Conditional Gaussian (CG) score

- Move all the continuous variables to the left and all the discrete variables to the right of the conditioning bar
  - Calculate the desired probability using partitioned Gaussian and Multinomial distributions

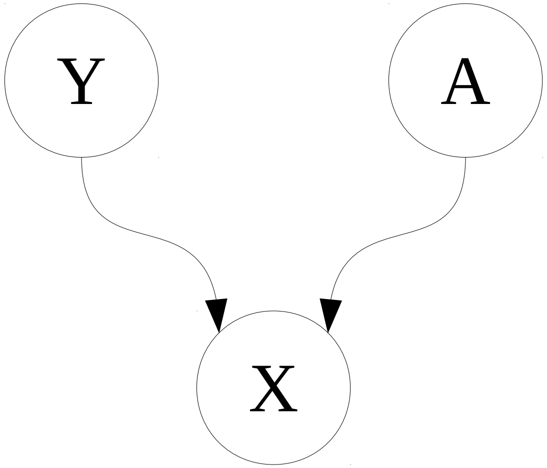
The Conditional Gaussian (CG) score

- Move all the continuous variables to the left and all the discrete variables to the right of the conditioning bar
  - Calculate the desired probability using partitioned Gaussian and Multinomial distributions
- Calculate a log-likelihood and degrees of freedom for BIC

Modeling a Continuous Child

Assume Y, A are parents of X

Let X, Y be continuous
Let A be discrete

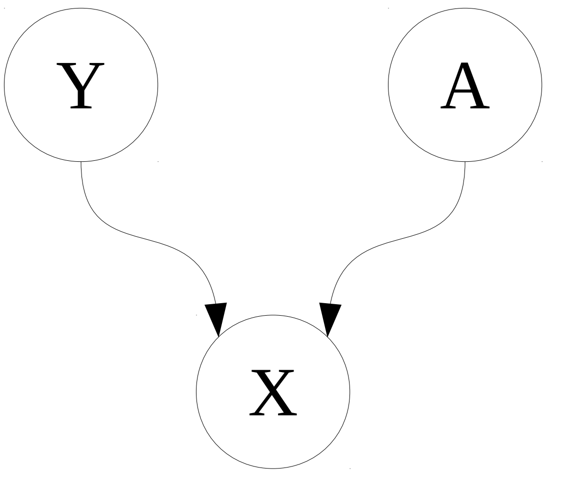


Modeling a Continuous Child

Assume Y, A are parents of X

Let X, Y be continuous
Let A be discrete

$$p(X|Y,A)=\frac{p(X,Y,A)}{p(Y,A)}$$



Modeling a Continuous Child

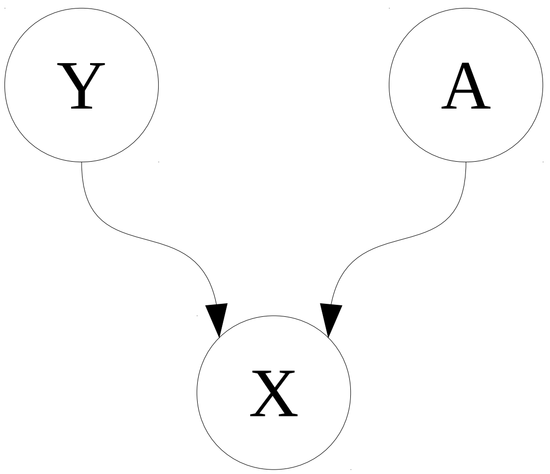
Assume Y, A are parents of X

Let X, Y be continuous
Let A be discrete

$$p(X|Y,A) = \frac{p(X,Y,A)}{p(Y,A)}$$

$$= \frac{p(X,Y|A)\,p(A)}{p(Y|A)\,p(A)}$$

$$= \frac{p(X,Y|A)}{p(Y|A)}$$



Modeling a Continuous Child

Assume Y, A are parents of X

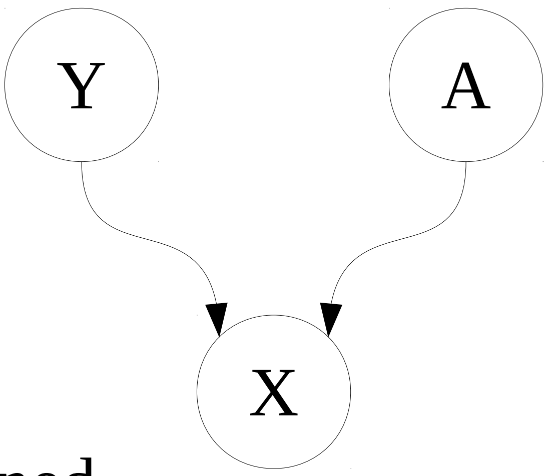
Let X, Y be continuous
Let A be discrete

$$p(X|Y,A) = \frac{p(X,Y,A)}{p(Y,A)}$$

$$= \frac{p(X,Y|A)\,p(A)}{p(Y|A)\,p(A)}$$

$$= \frac{p(X,Y|A)}{p(Y|A)}$$

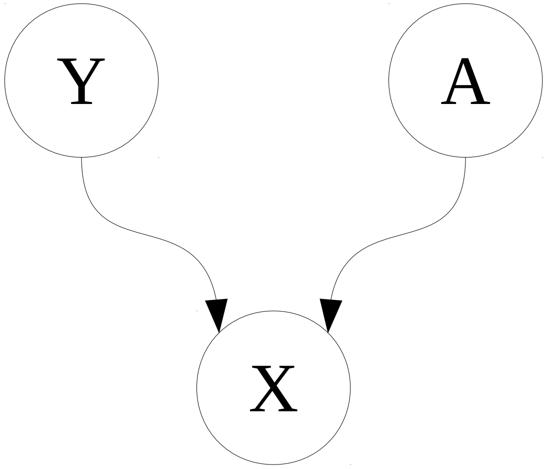
Partitioned Gaussians



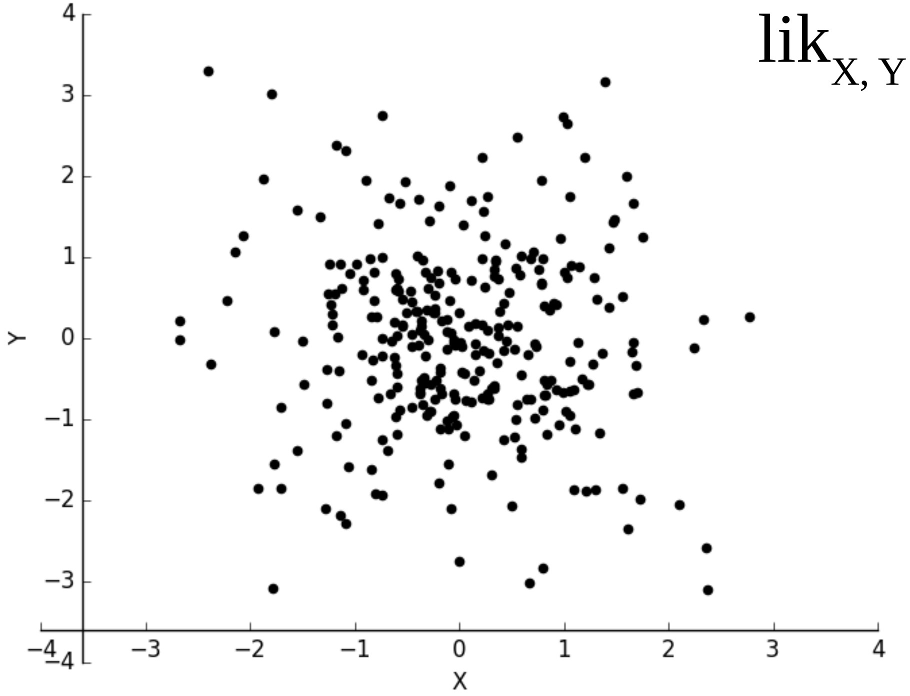
Modeling a Continuous Child

- Want: $\text{lik}_{X, Y|A}$, $\text{dof}_{X, Y|A}$
  $\text{lik}_{Y|A}$, $\text{dof}_{Y|A}$

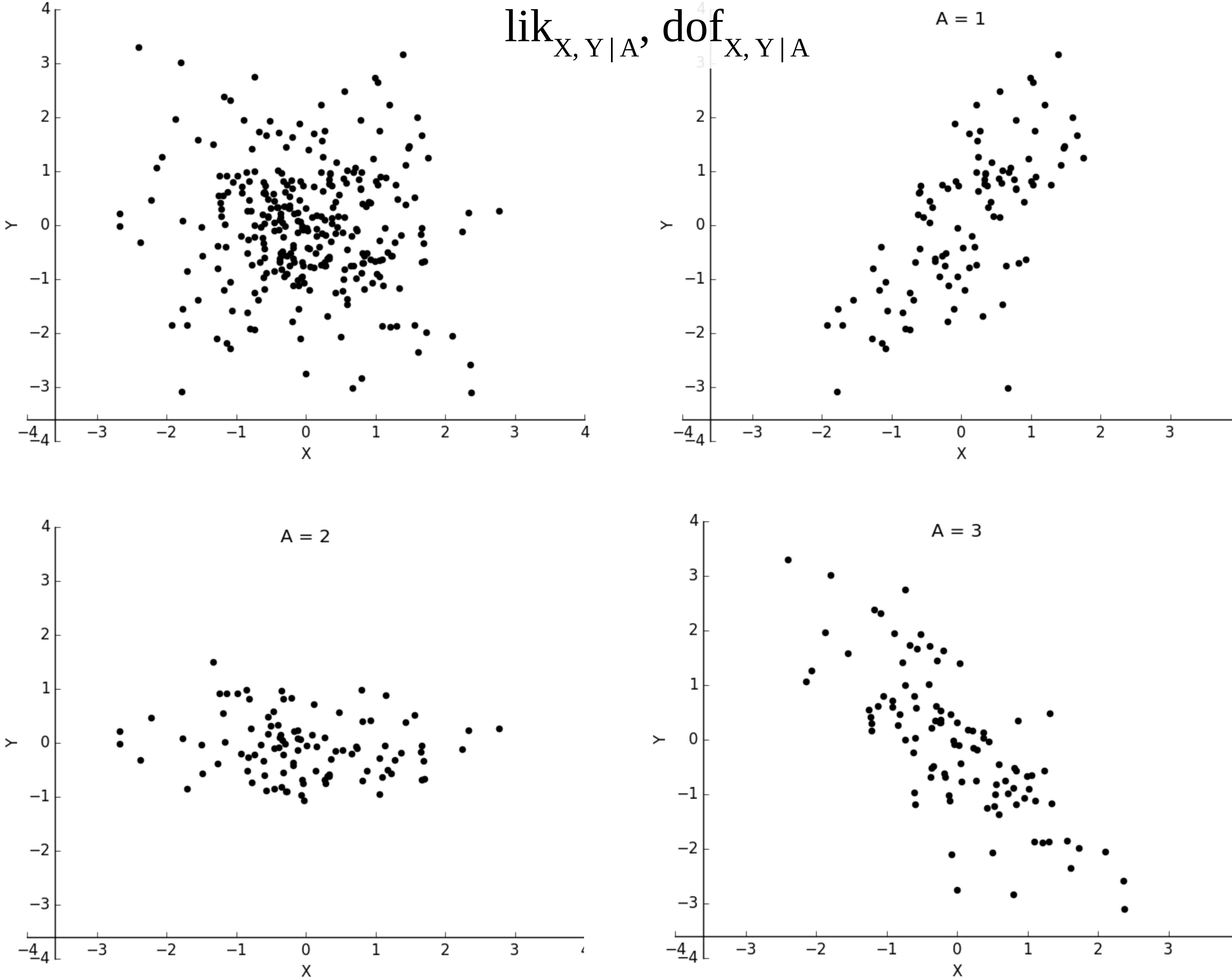
$$\frac{p(X, Y|A)}{p(Y|A)}$$

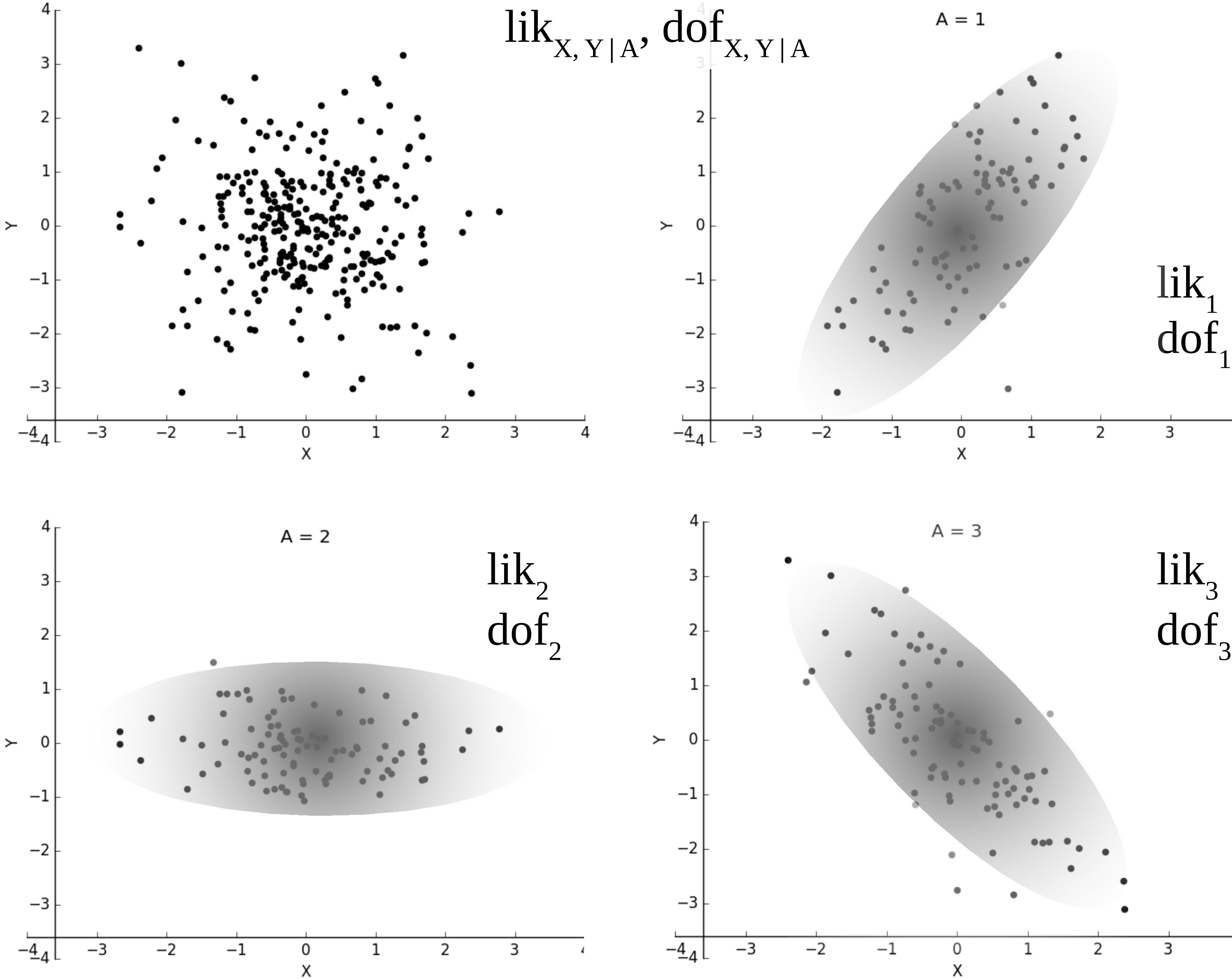


$lik_{X, Y|A}$, $dof_{X, Y|A}$



$\text{lik}_{X,Y|A}$, $\text{dof}_{X,Y|A}$

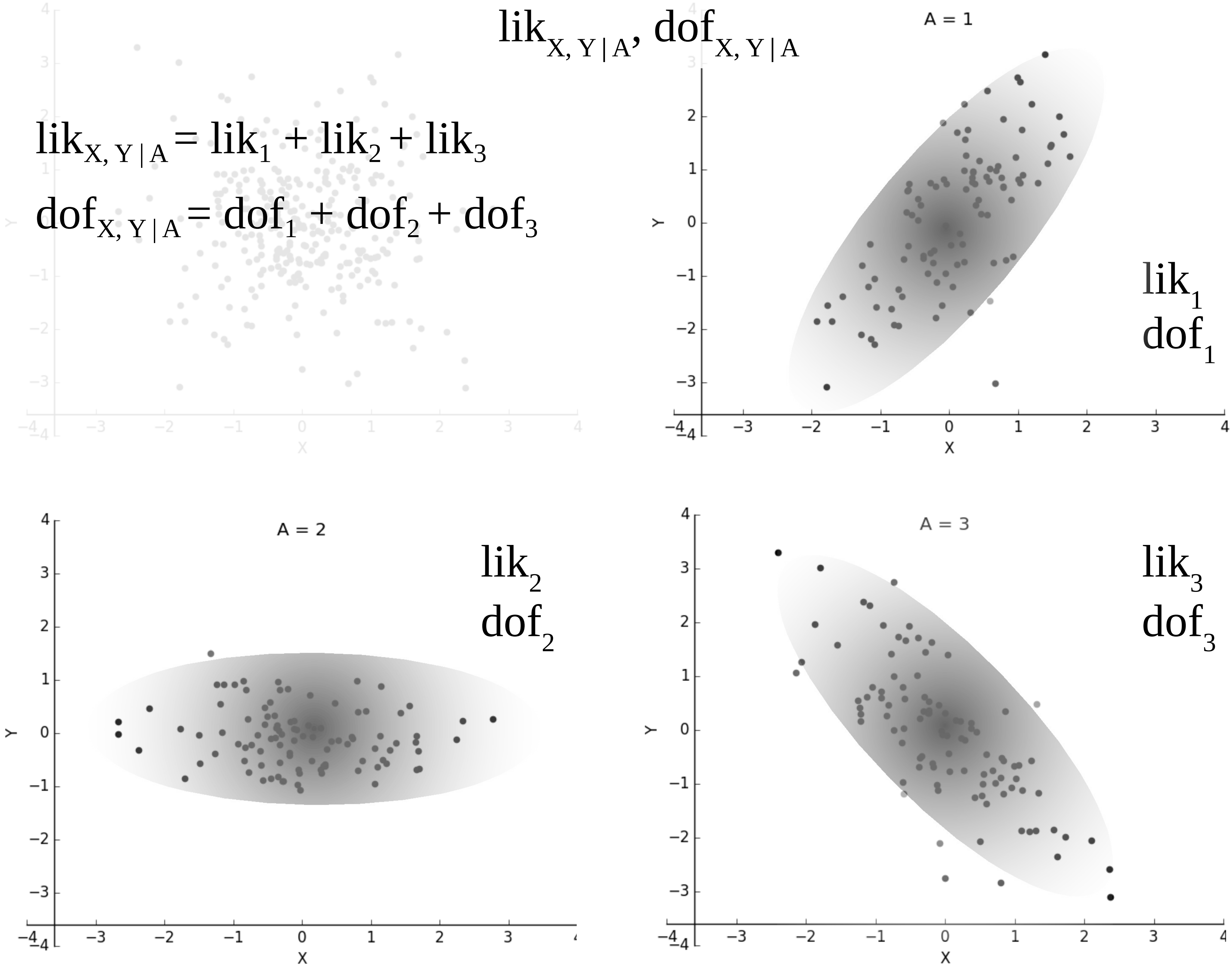



$\text{lik}_{X, Y | A}$, $\text{dof}_{X, Y | A}$
A = 1
lik_1
dof_1
A = 2
lik_2
dof_2
A = 3
lik_3
dof_3
X
Y

$\mathrm{lik}_{X,\,Y\,|\,A}$, $\mathrm{dof}_{X,\,Y\,|\,A}$

$$\mathrm{lik}_{X,\,Y\,|\,A} = \mathrm{lik}_1 + \mathrm{lik}_2 + \mathrm{lik}_3$$

$$\mathrm{dof}_{X,\,Y\,|\,A} = \mathrm{dof}_1 + \mathrm{dof}_2 + \mathrm{dof}_3$$

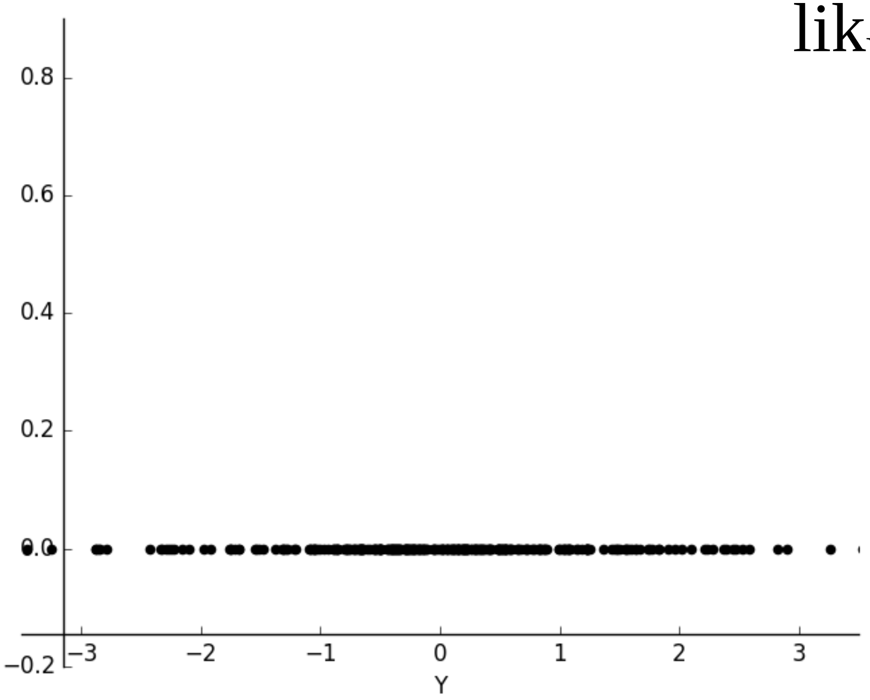


lik_1
dof_1

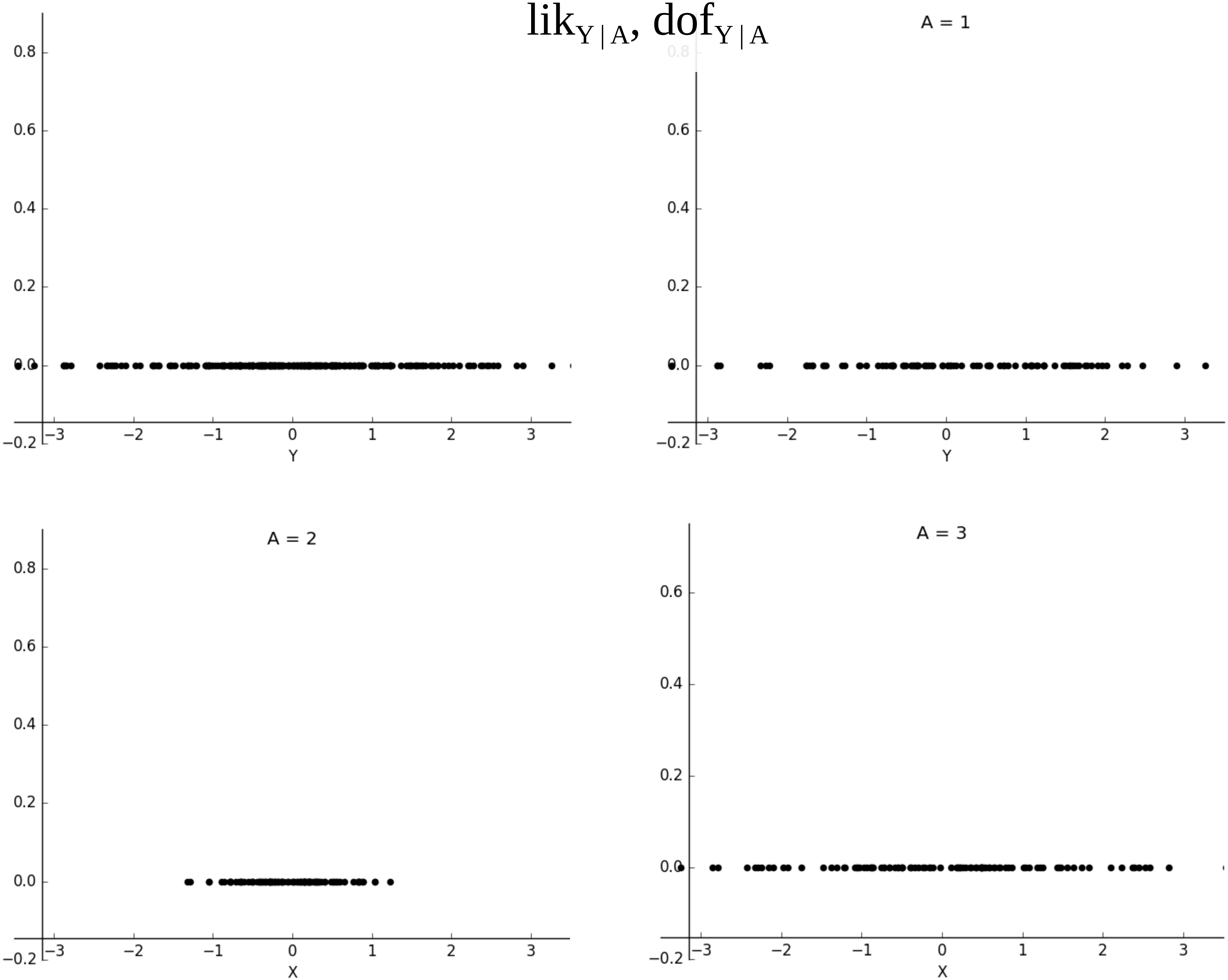
lik_2
dof_2

lik_3
dof_3

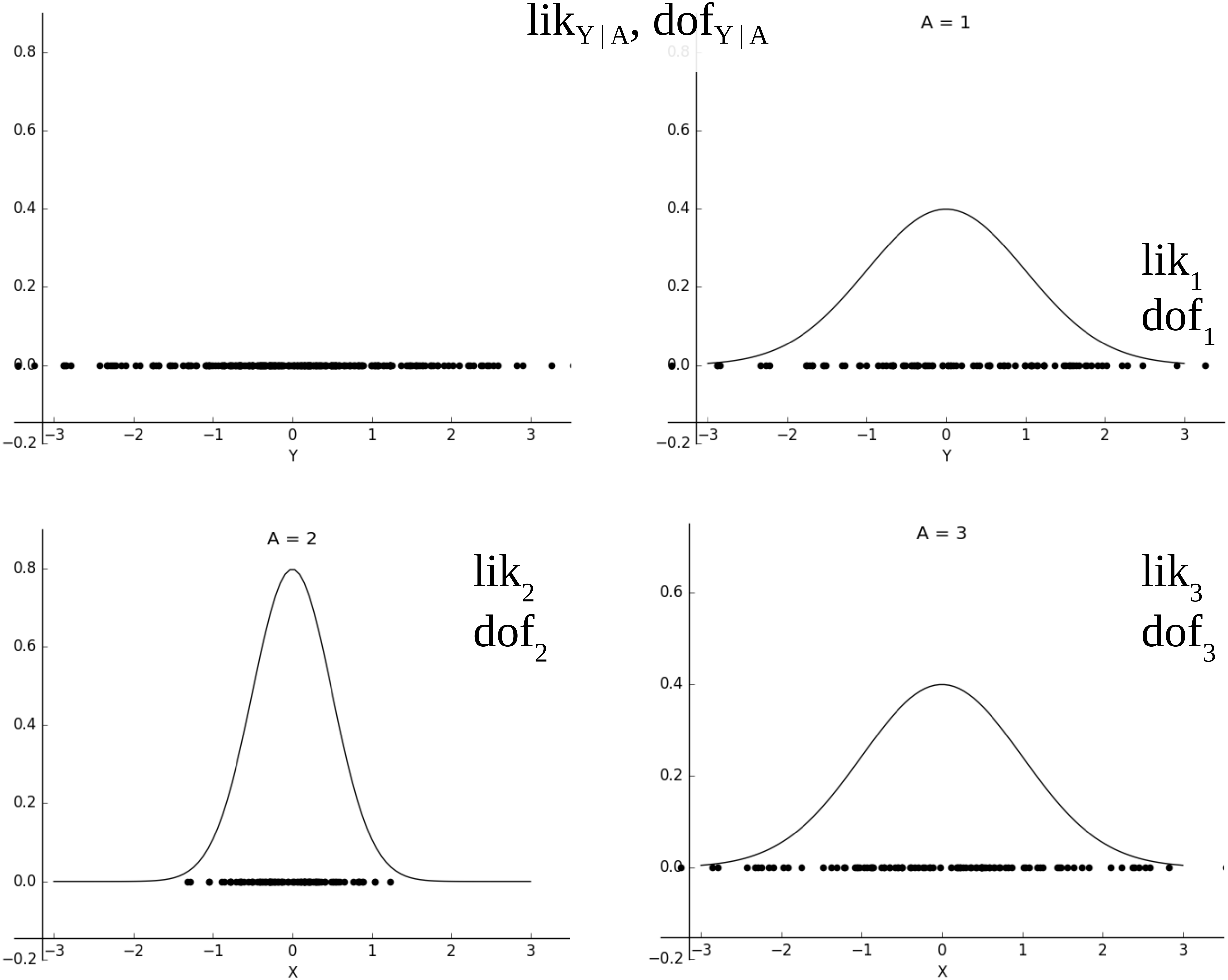
$\text{lik}_{Y|A}, \text{dof}_{Y|A}$



$\text{lik}_{Y\,|\,A}$, $\text{dof}_{Y\,|\,A}$



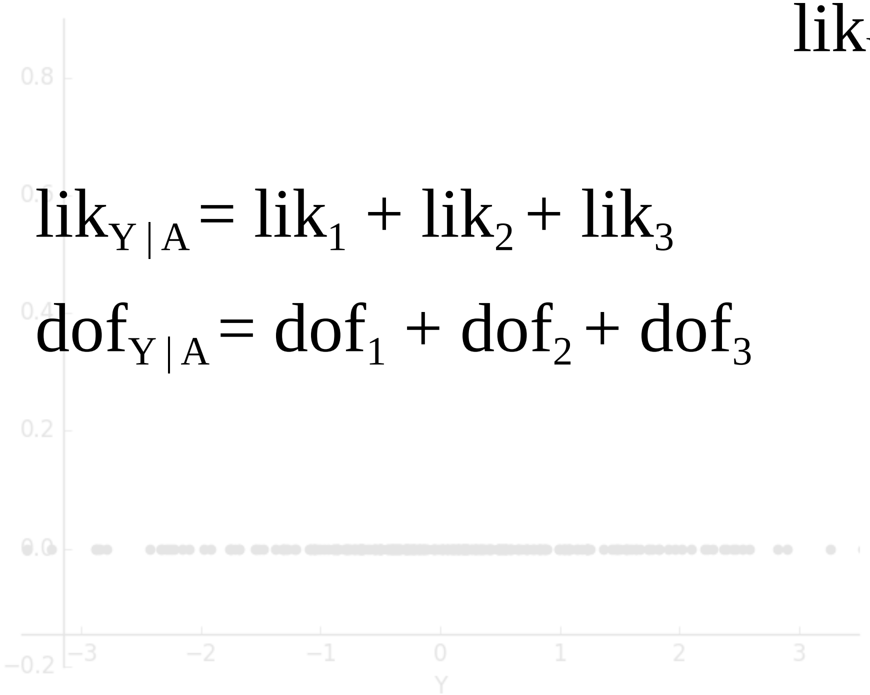
$\mathrm{lik}_{Y|A}$, $\mathrm{dof}_{Y|A}$

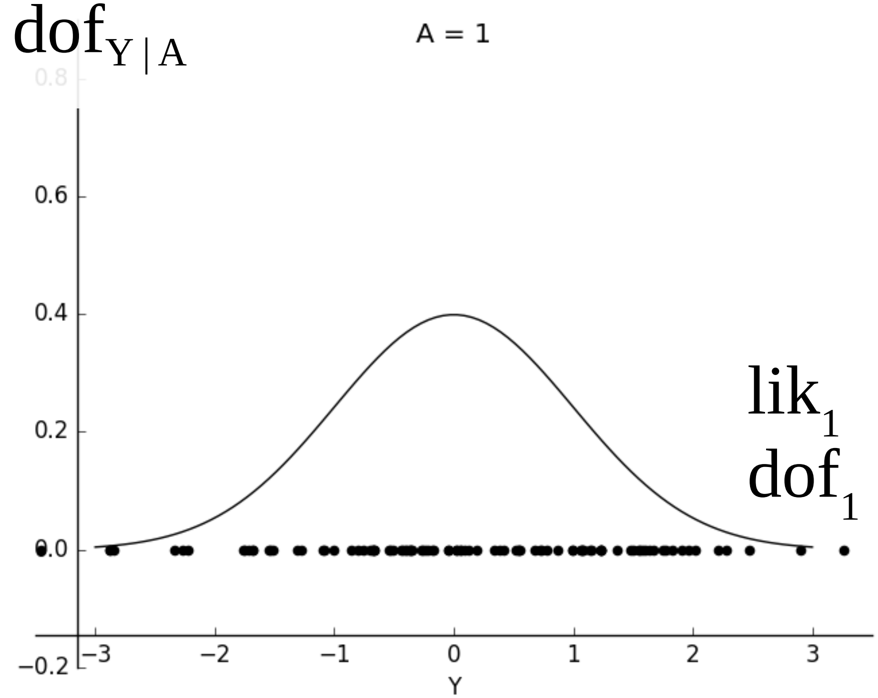


$lik_{Y|A}$, $dof_{Y|A}$

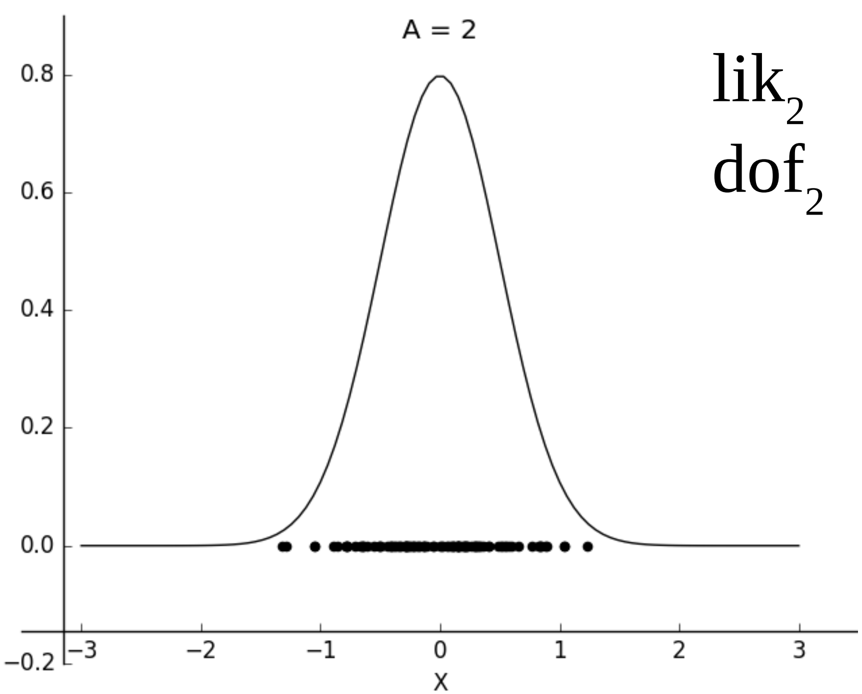
$$lik_{Y|A} = lik_1 + lik_2 + lik_3$$

$$dof_{Y|A} = dof_1 + dof_2 + dof_3$$

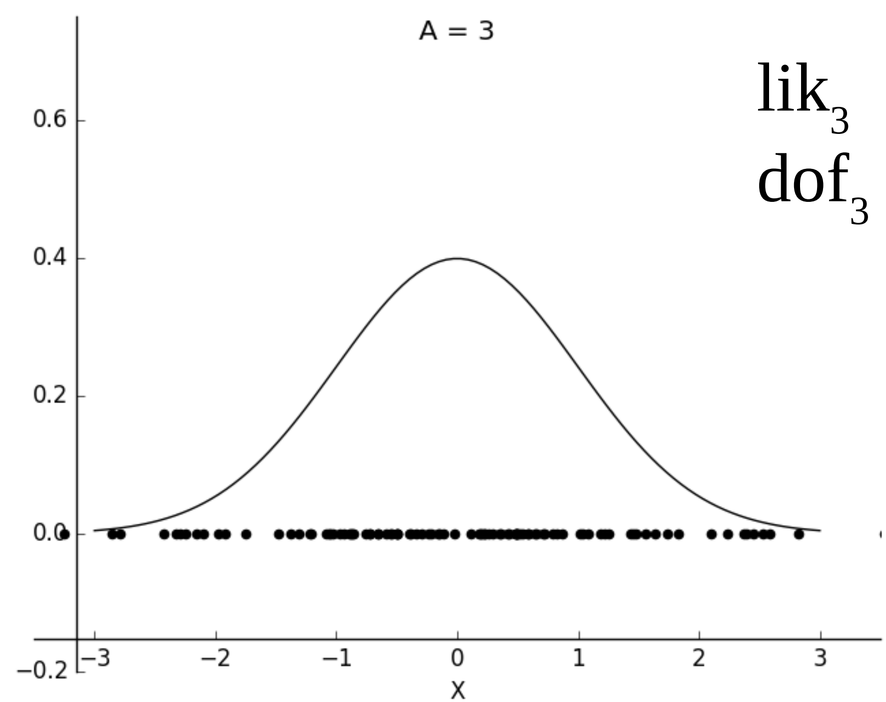




lik_1
dof_1



lik_2
dof_2



lik_3
dof_3

Modeling a Continuous Child

Have: $\text{lik}_{X,\,Y\,|\,A}$, $\text{dof}_{X,\,Y\,|\,A}$

$\text{lik}_{Y\,|\,A}$, $\text{dof}_{Y\,|\,A}$

$$\frac{p(X, Y|A)}{p(Y|A)}$$

Modeling a Continuous Child

Have: $\text{lik}_{X,\,Y\,|\,A}$, $\text{dof}_{X,\,Y\,|\,A}$

$\text{lik}_{Y\,|\,A}$, $\text{dof}_{Y\,|\,A}$

$$\frac{p(X, Y|A)}{p(Y|A)}$$

$$\text{lik}_{X\,|\,Y,\,A} = \text{lik}_{X,\,Y\,|\,A} - \text{lik}_{Y\,|\,A}$$
$$\text{dof}_{X\,|\,Y,\,A} = \text{dof}_{X,\,Y\,|\,A} - \text{dof}_{Y\,|\,A}$$

Modeling a Continuous Child

Have: $\text{lik}_{X,\,Y\,|\,A}$, $\text{dof}_{X,\,Y\,|\,A}$

$\text{lik}_{Y\,|\,A}$, $\text{dof}_{Y\,|\,A}$

$$\frac{p(X,Y|A)}{p(Y|A)}$$

$$\text{lik}_{X\,|\,Y,\,A} = \text{lik}_{X,\,Y\,|\,A} - \text{lik}_{Y\,|\,A}$$

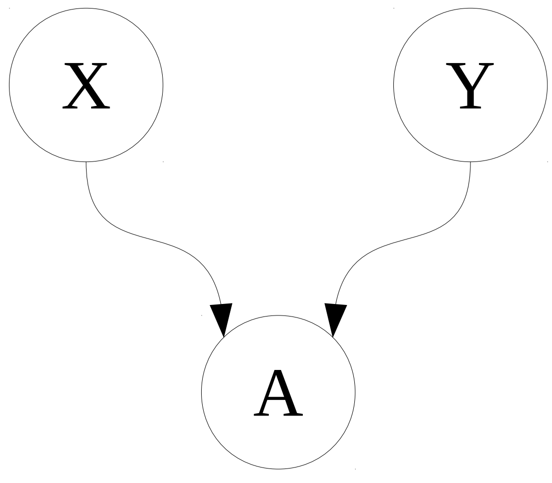
$$\text{dof}_{X\,|\,Y,\,A} = \text{dof}_{X,\,Y\,|\,A} - \text{dof}_{Y\,|\,A}$$

$$-2\text{lik}_{X\,|\,Y,A} + \text{dof}_{X\,|\,Y,\,A} \log n$$

Modeling a Discrete Child

Assume X, Y are
parents of A

Let X, Y be continuous
Let A be discrete

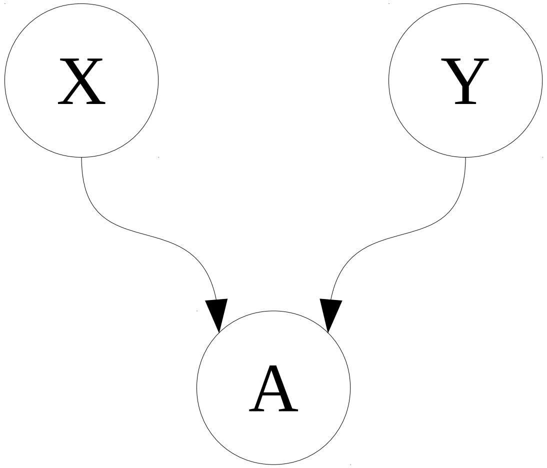


Modeling a Discrete Child

Assume X, Y are parents of A

Let X, Y be continuous
Let A be discrete

$$p(A|X,Y)=\frac{p(X,Y,A)}{p(X,Y)}$$



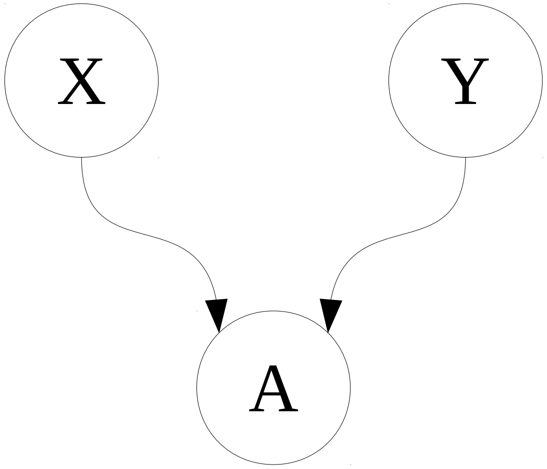
Modeling a Discrete Child

Assume X, Y are parents of A

Let X, Y be continuous
Let A be discrete

$$p(A|X,Y)=\frac{p(X,Y,A)}{p(X,Y)}$$

$$=\frac{p(X,Y|A)\,p(A)}{p(X,Y)}$$



Modeling a Discrete Child

Assume X, Y are parents of A

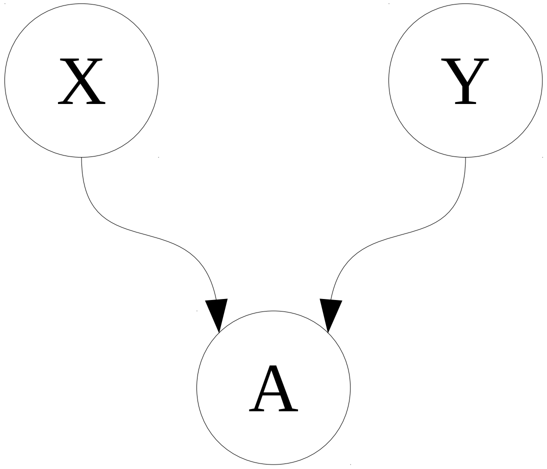
Let X, Y be continuous
Let A be discrete

$$p(A|X,Y) = \frac{p(X,Y,A)}{p(X,Y)}$$

$$= \frac{p(X,Y|A)\,p(A)}{p(X,Y)}$$

Partitioned Gaussians

Multinomial



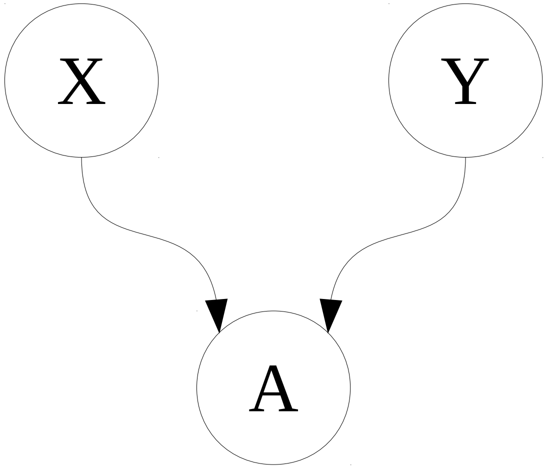
Modeling a Discrete Child

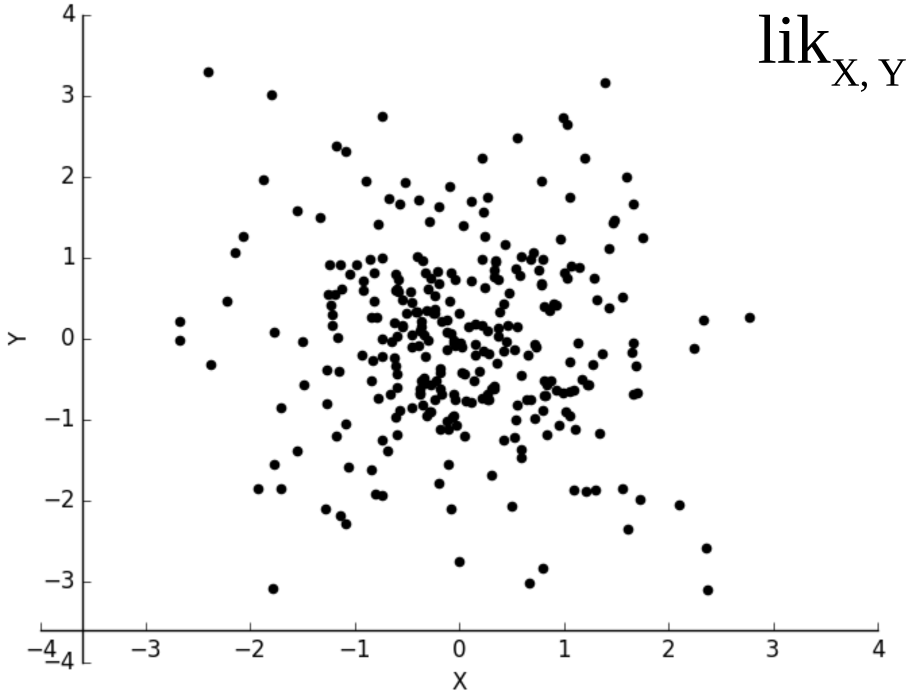
- Want: $\text{lik}_{X,\,Y\,|\,A}$, $\text{dof}_{X,\,Y\,|\,A}$

  lik_{A}, dof_{A}

  $\text{lik}_{X,\,Y}$, $\text{dof}_{X,\,Y}$

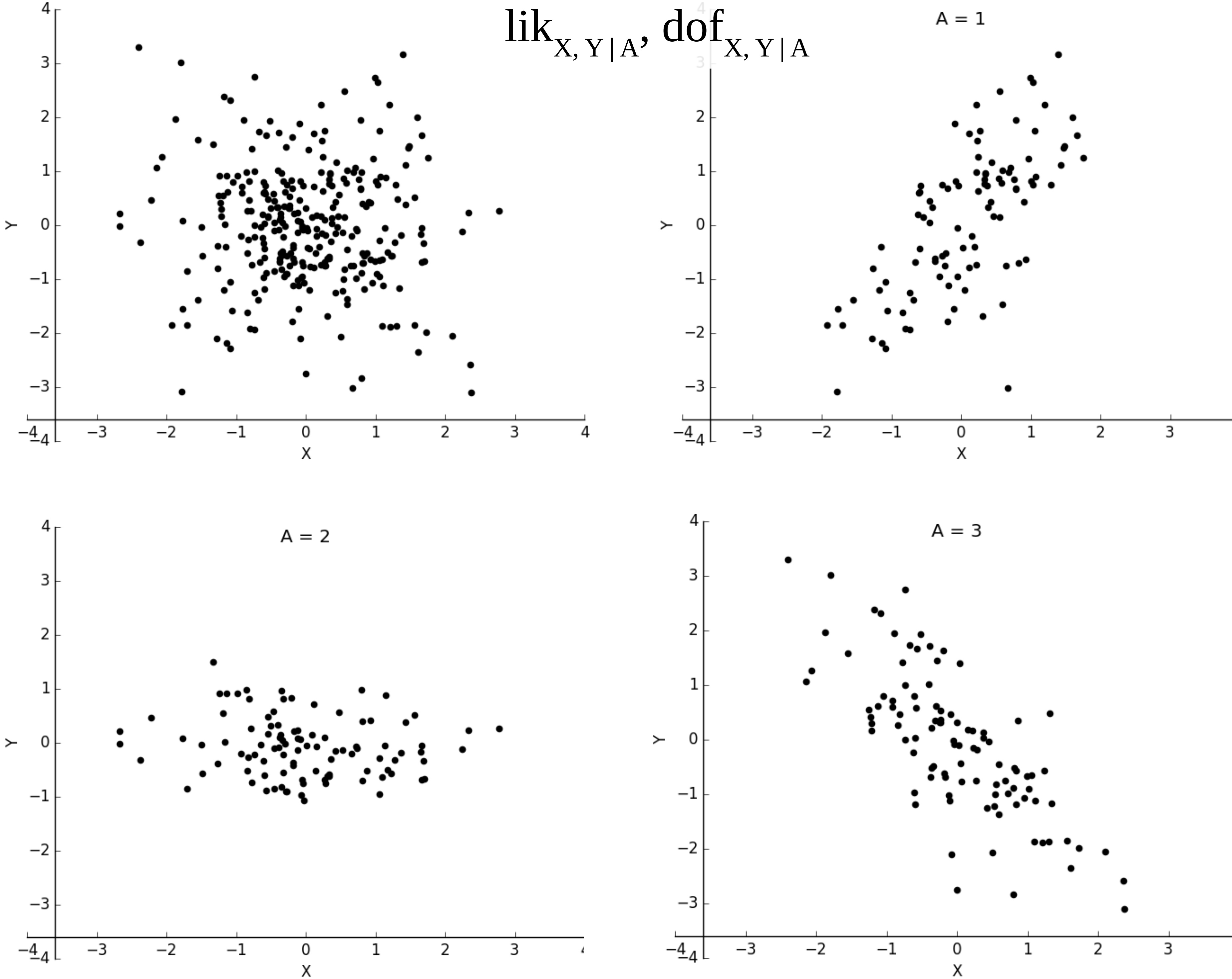
$$\frac{p(X,Y|A)\,p(A)}{p(X,Y)}$$

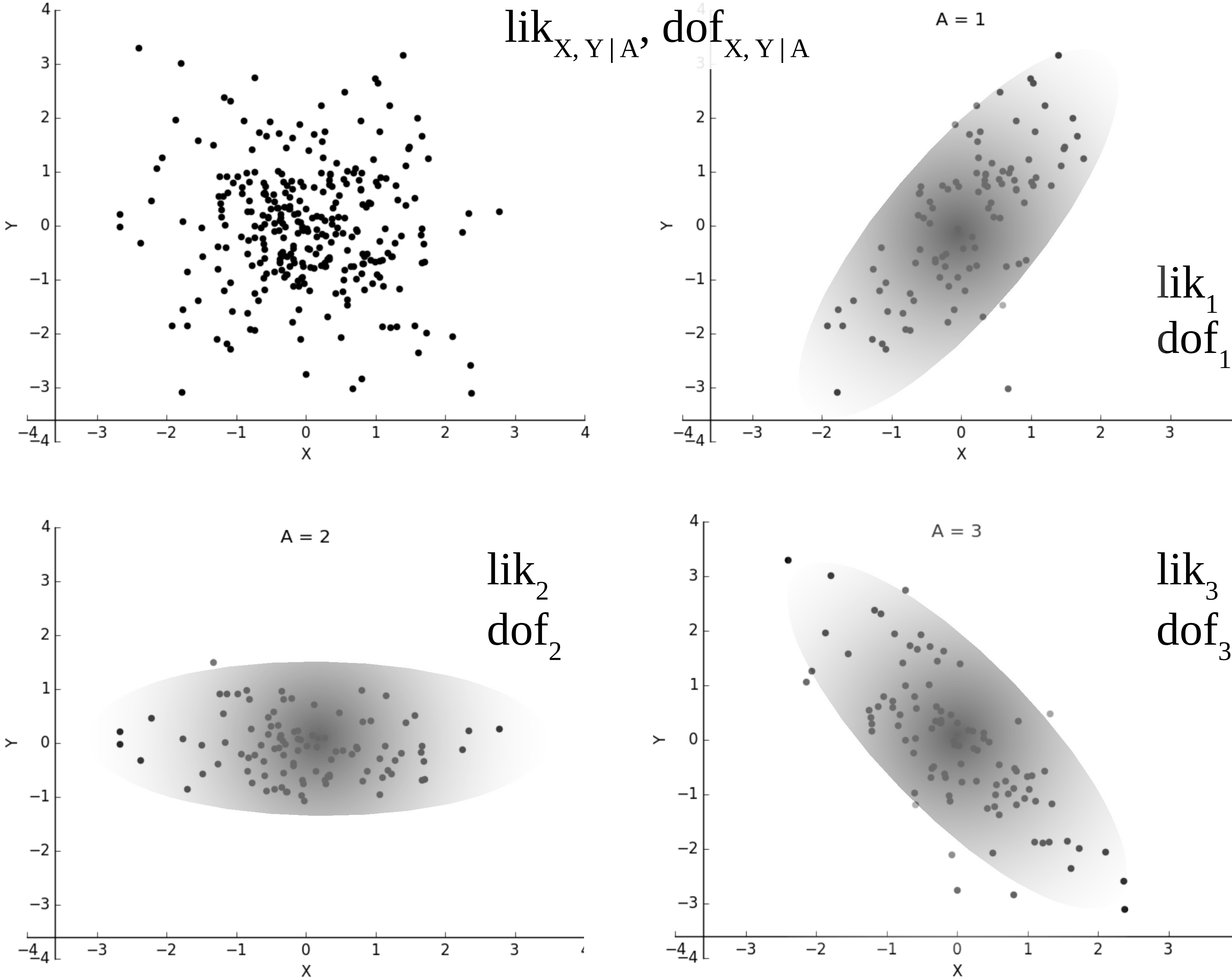




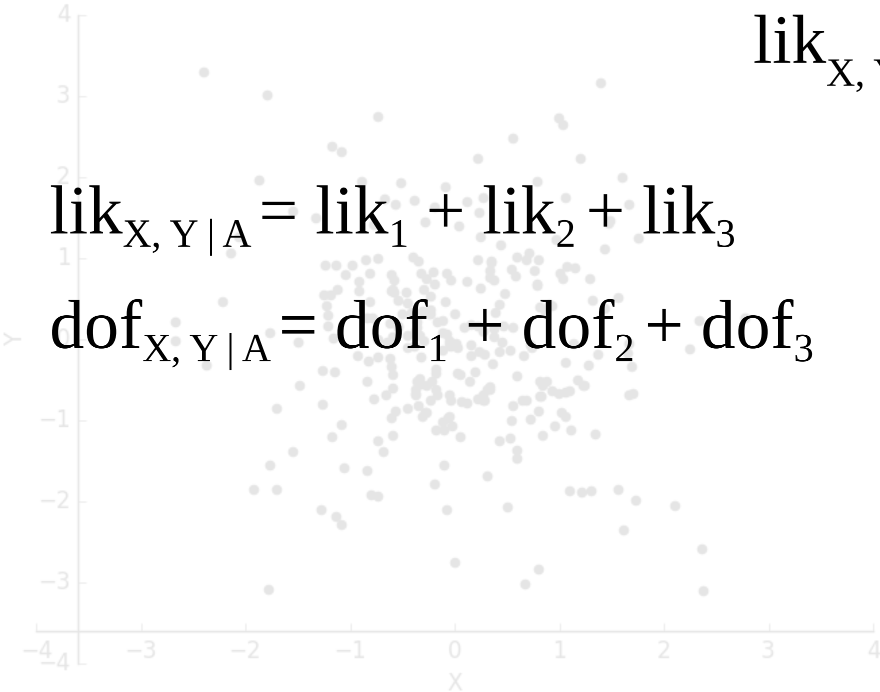
$lik_{X, Y | A}$, $dof_{X, Y | A}$

$lik_{X,\,Y\,|\,A}$, $dof_{X,\,Y\,|\,A}$



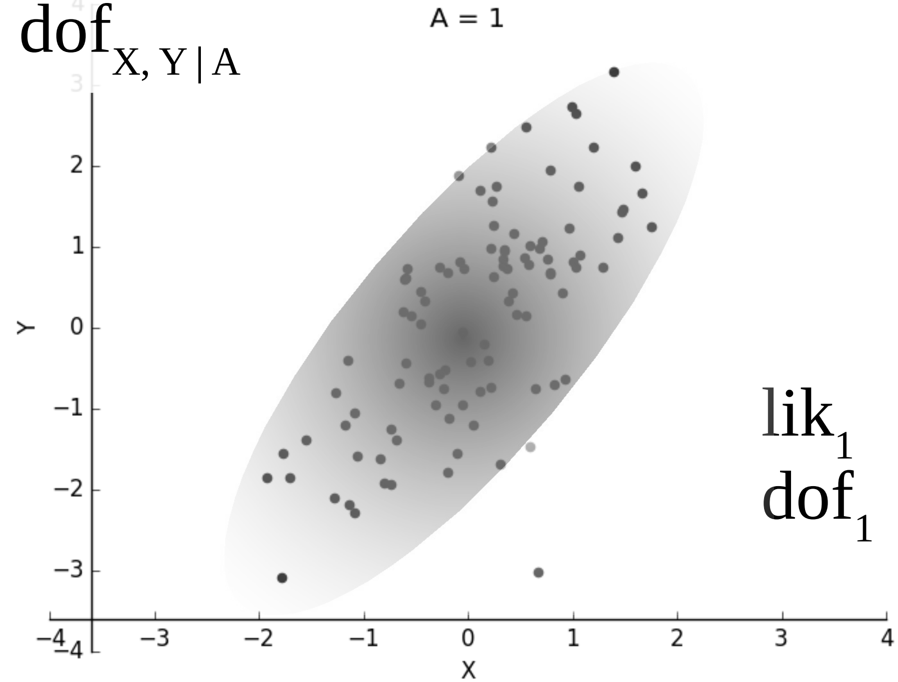

lik_{X, Y | A}, dof_{X, Y | A}
A = 1
lik_1
dof_1
A = 2
lik_2
dof_2
A = 3
lik_3
dof_3
X
Y

$\text{lik}_{X,Y|A}$, $\text{dof}_{X,Y|A}$

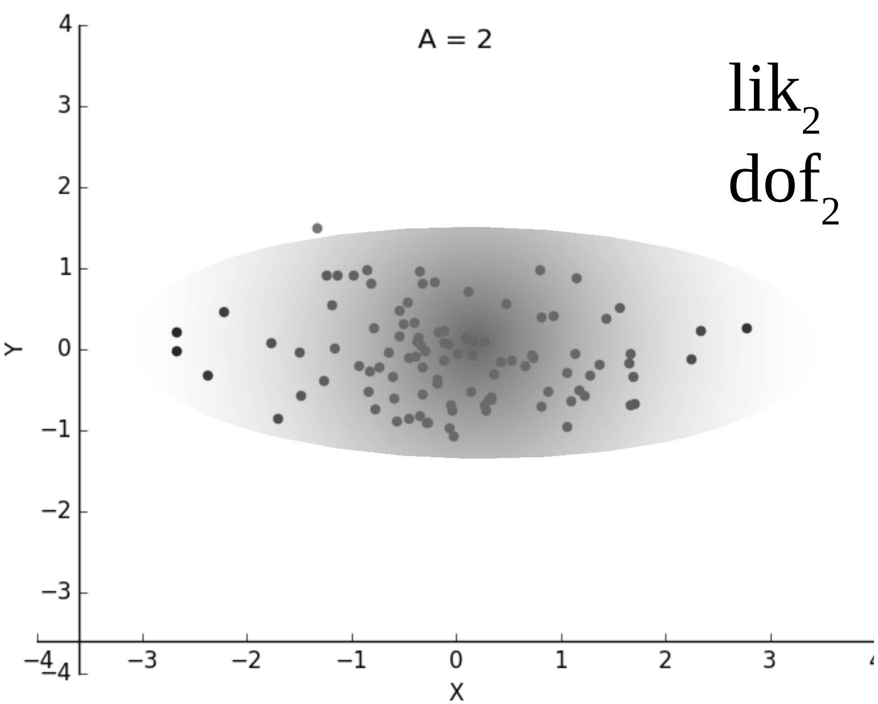


$$\text{lik}_{X,Y|A} = \text{lik}_1 + \text{lik}_2 + \text{lik}_3$$

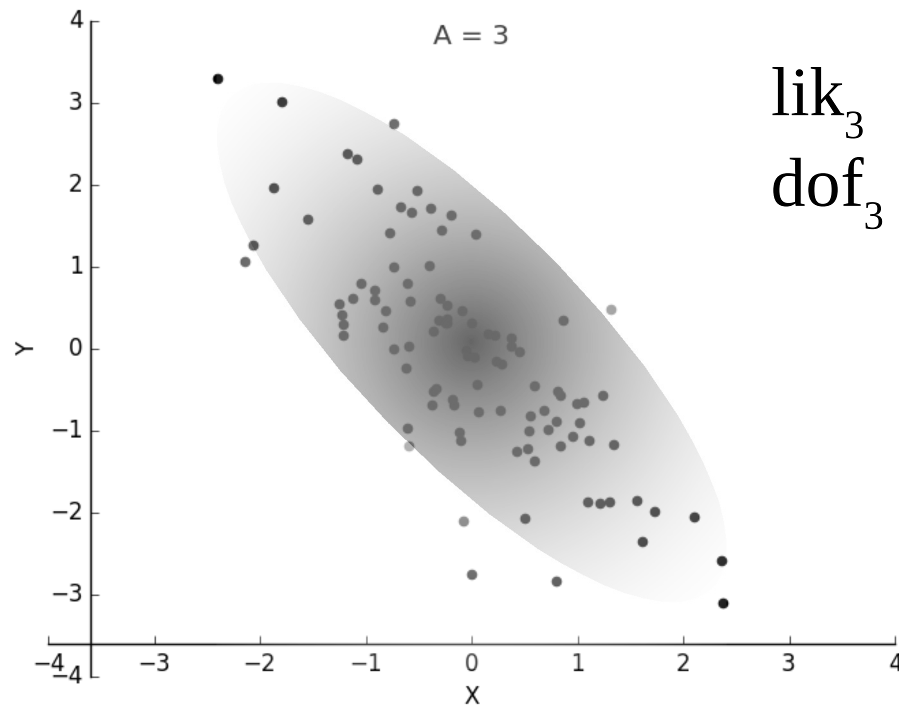
$$\text{dof}_{X,Y|A} = \text{dof}_1 + \text{dof}_2 + \text{dof}_3$$



lik_1
dof_1

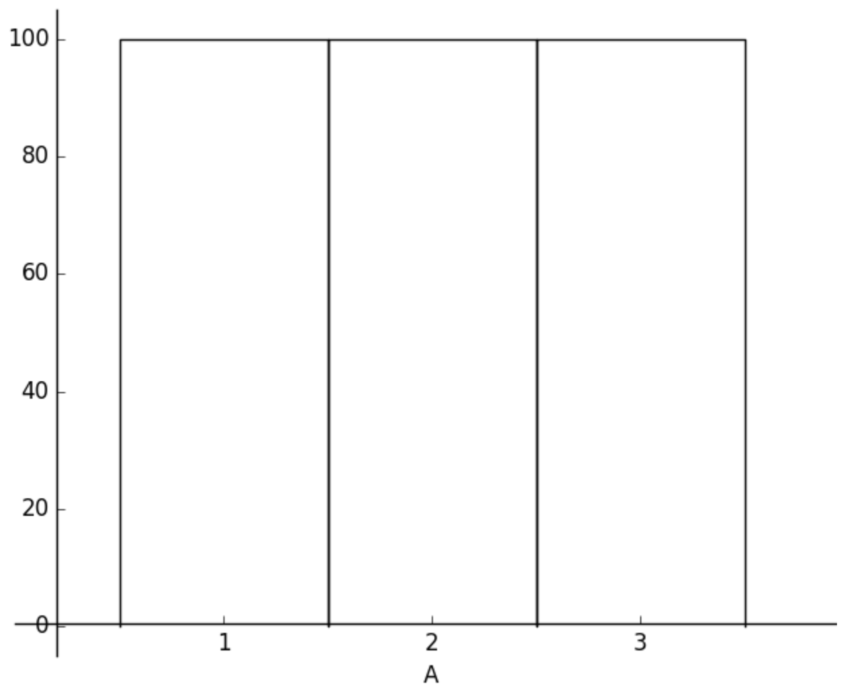


lik_2
dof_2

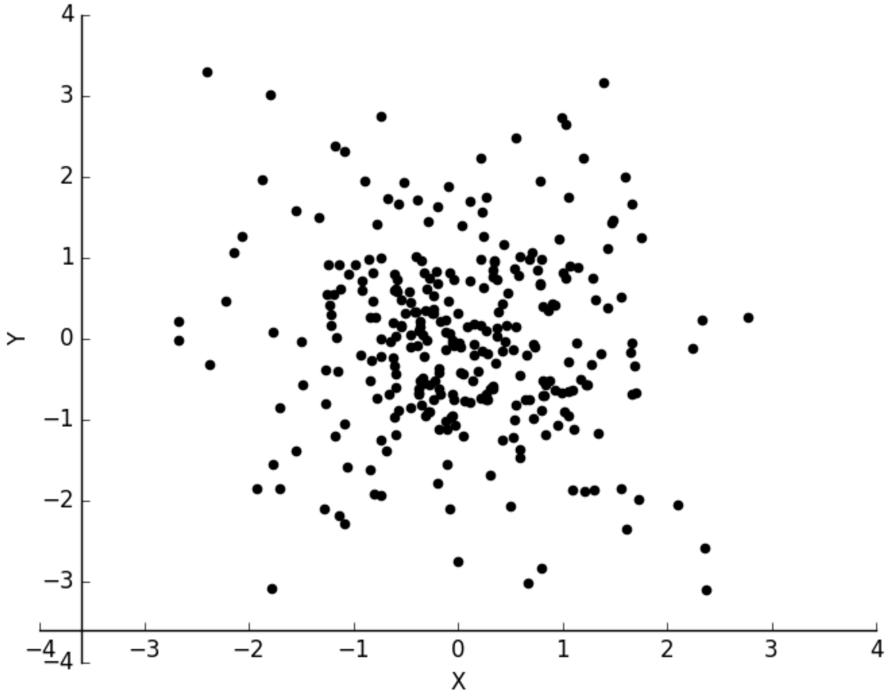


lik_3
dof_3

lik_A, dof_A

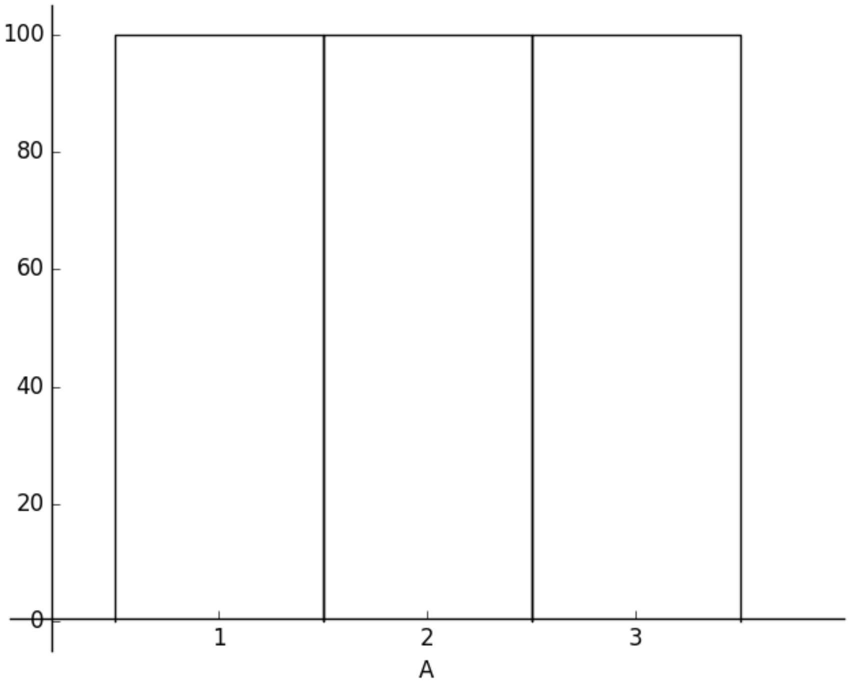


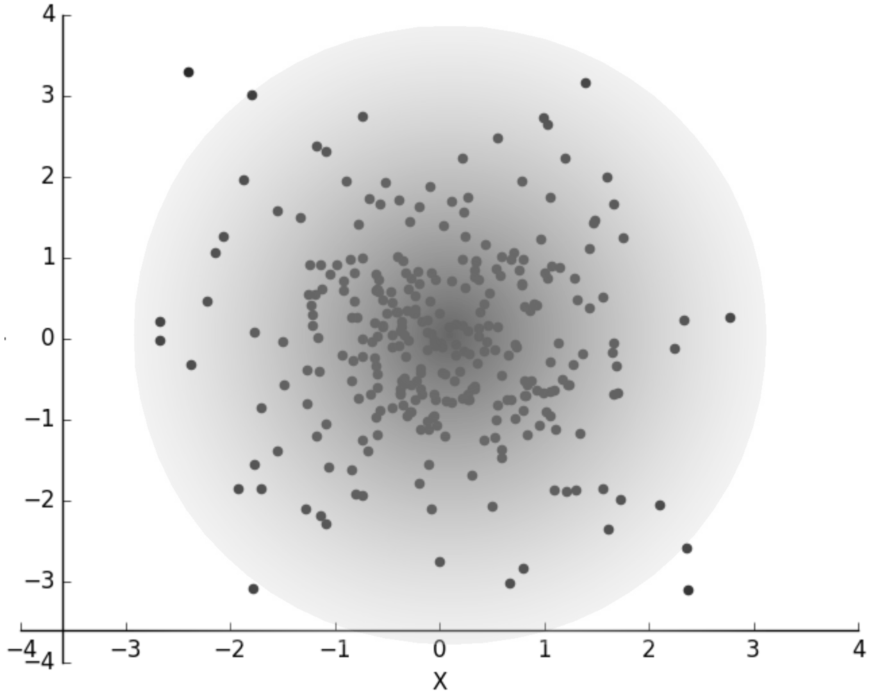
$lik_{X, Y}, dof_{X, Y}$



lik_A, dof_A

$lik_{X,\,Y}$, $dof_{X,\,Y}$





lik_A

dof_A

$lik_{X,\,Y}$

$dof_{X,\,Y}$

Modeling a Discrete Child

Have: $\text{lik}_{X, Y \mid A}$, $\text{dof}_{X, Y \mid A}$

lik_{A}, dof_{A}

$\text{lik}_{X, Y}$, $\text{dof}_{X, Y}$

$$\frac{p(X, Y \mid A)\, p(A)}{p(X, Y)}$$

Modeling a Discrete Child

Have: $\text{lik}_{X,\,Y\,|\,A}$, $\text{dof}_{X,\,Y\,|\,A}$

lik_{A}, dof_{A}

$\text{lik}_{X,\,Y}$, $\text{dof}_{X,\,Y}$

$$\frac{p(X,Y|A)\,p(A)}{p(X,Y)}$$

$$\text{lik}_{A\,|\,X,\,Y} = \text{lik}_{X,\,Y\,|\,A} + \text{lik}_{A} - \text{lik}_{X,\,Y}$$

$$\text{dof}_{A\,|\,X,\,Y} = \text{dof}_{X,\,Y\,|\,A} + \text{dof}_{A} - \text{dof}_{X,\,Y}$$

Modeling a Discrete Child

Have: $\text{lik}_{X,\,Y\,|\,A}$, $\text{dof}_{X,\,Y\,|\,A}$

lik_{A}, dof_{A}

$\text{lik}_{X,\,Y}$, $\text{dof}_{X,\,Y}$

$$\frac{p(X,Y|A)\,p(A)}{p(X,Y)}$$

$$\text{lik}_{A\,|\,X,\,Y} = \text{lik}_{X,\,Y\,|\,A} + \text{lik}_{A} - \text{lik}_{X,\,Y}$$

$$\text{dof}_{A\,|\,X,\,Y} = \text{dof}_{X,\,Y\,|\,A} + \text{dof}_{A} - \text{dof}_{X,\,Y}$$

$$-2\text{lik}_{A\,|\,X,\,Y} + \text{dof}_{A\,|\,X,\,Y} \log n$$

Outline

- Bayesian Information Criterion (BIC)
- Mixed Variable Polynomial (MVP) score
- Conditional Gaussian (CG) score
- Adaptations
- Simulations and empirical results

Adaptations

- Binomial Structure Prior
  - Treat the addition of each parent as an independent random trial
  - Model the prior probability of each parent-child model using a Binomial distribution
- Discretization Heuristic
  - Discretize continuous parents of discrete children in order to use multinomial scoring

Outline

- Bayesian Information Criterion (BIC)
- Mixed Variable Polynomial (MVP) score
- Conditional Gaussian (CG) score
- Adaptations
- Simulations and empirical results

Conditional Gaussian Simulation

- Randomly generate a set of variables and edges

Conditional Gaussian Simulation

- Randomly generate a set of variables and edges
- Specify a causal ordering over the variables

Conditional Gaussian Simulation

- Randomly generate a set of variables and edges
- Specify a causal ordering over the variables
- In causal order, simulate one variable at a time

Conditional Gaussian Simulation

- Randomly generate a set of variables and edges
- Specify a causal ordering over the variables
- In causal order, simulate one variable at a time
  - Use multinomial relationships with discretized continuous parents for discrete children

Conditional Gaussian Simulation

- Randomly generate a set of variables and edges
- Specify a causal ordering over the variables
- In causal order, simulate one variable at a time
  - Use multinomial relationships with discretized continuous parents for discrete children
  - Use partitioned linear Gaussian relationships for continuous children

Conditional Gaussian Simulation

- Randomly generate a set of variables and edges
- Specify a causal ordering over the variables
- In causal order, simulate one variable at a time
  - Use multinomial relationships with discretized continuous parents for discrete children
  - Use partitioned linear Gaussian relationships for continuous children

Note: For all simulation we simulate discrete-continuous at a 50-50 split where discrete variables have a random number of categories between 2 and 5

Non-linear Simulation

- Randomly generate a set of variables and edges
- Specify a causal ordering over the variables
- In causal order, simulate one variable at a time
  - Use multinomial relationships with discretized continuous parents for discrete children
  - Use partitioned polynomial regression with Gaussian noise for continuous children

Note: For all simulation we simulate discrete-continuous at a 50-50 split where discrete variables have a random number of categories between 2 and 5

Algorithms

CG – Conditional Gaussian

CG d – Conditional Gaussian w/ Discretization Heuristic

MVP 1 – Mixed Variable Polynomial w/ linear basis

MVP log n – Mixed Variable Polynomial w/ polynomial basis

LR 1 – Logistic Regression w/ linear basis

LR log n – Logistic Regression w/ polynomial basis

Statistics

AP – Adjacency Precision

correctly predicted adjacent / predicted adjacent

AR – Adjacency Recall

correctly predicted adjacent / true adjacent

AHP – Arrowhead Precision

correctly predicted arrowhead / predicted arrowhead

AHR – Arrowhead Recall

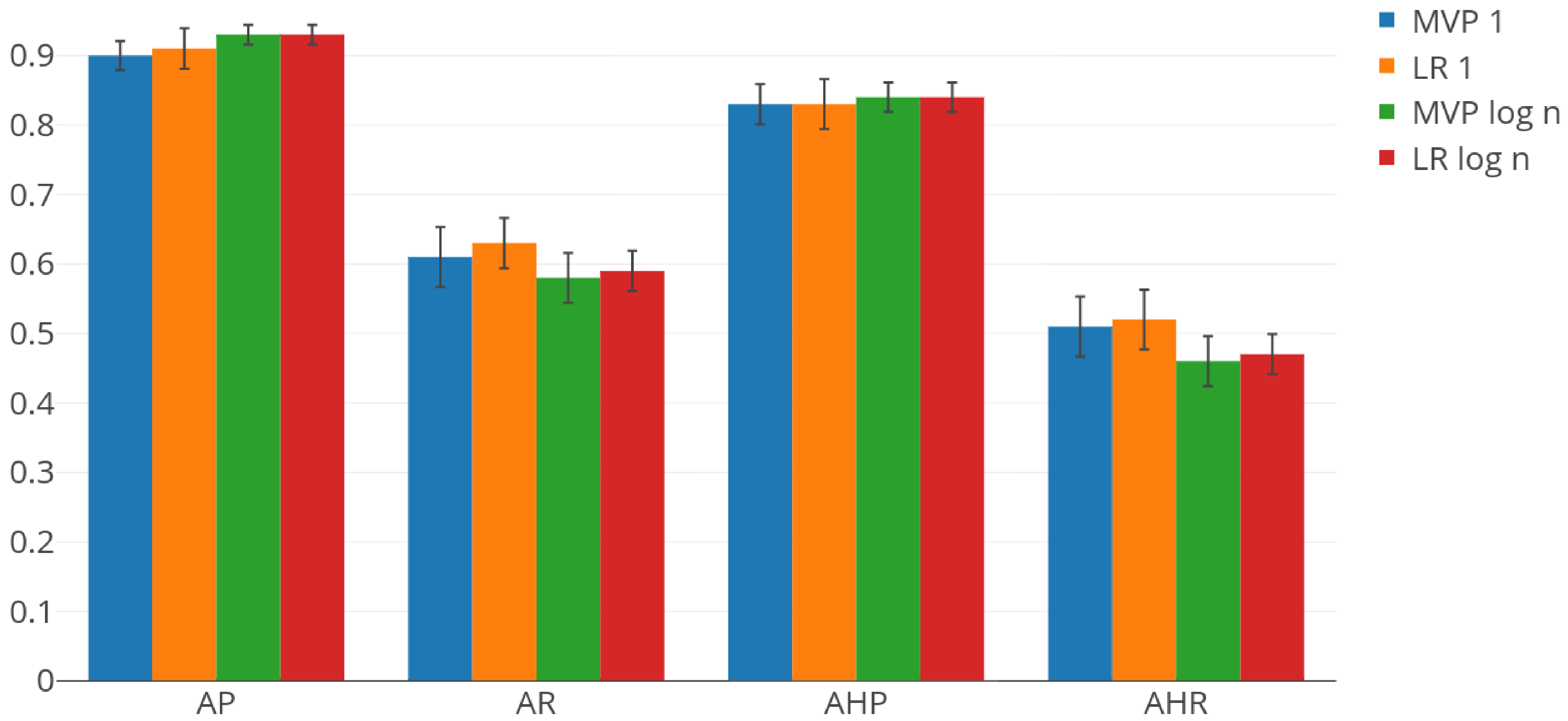
correctly predicted arrowhead / true arrowhead

T (s) – Computation time (in seconds)

All statistics are averaged over 10 runs on networks of 1000 instances

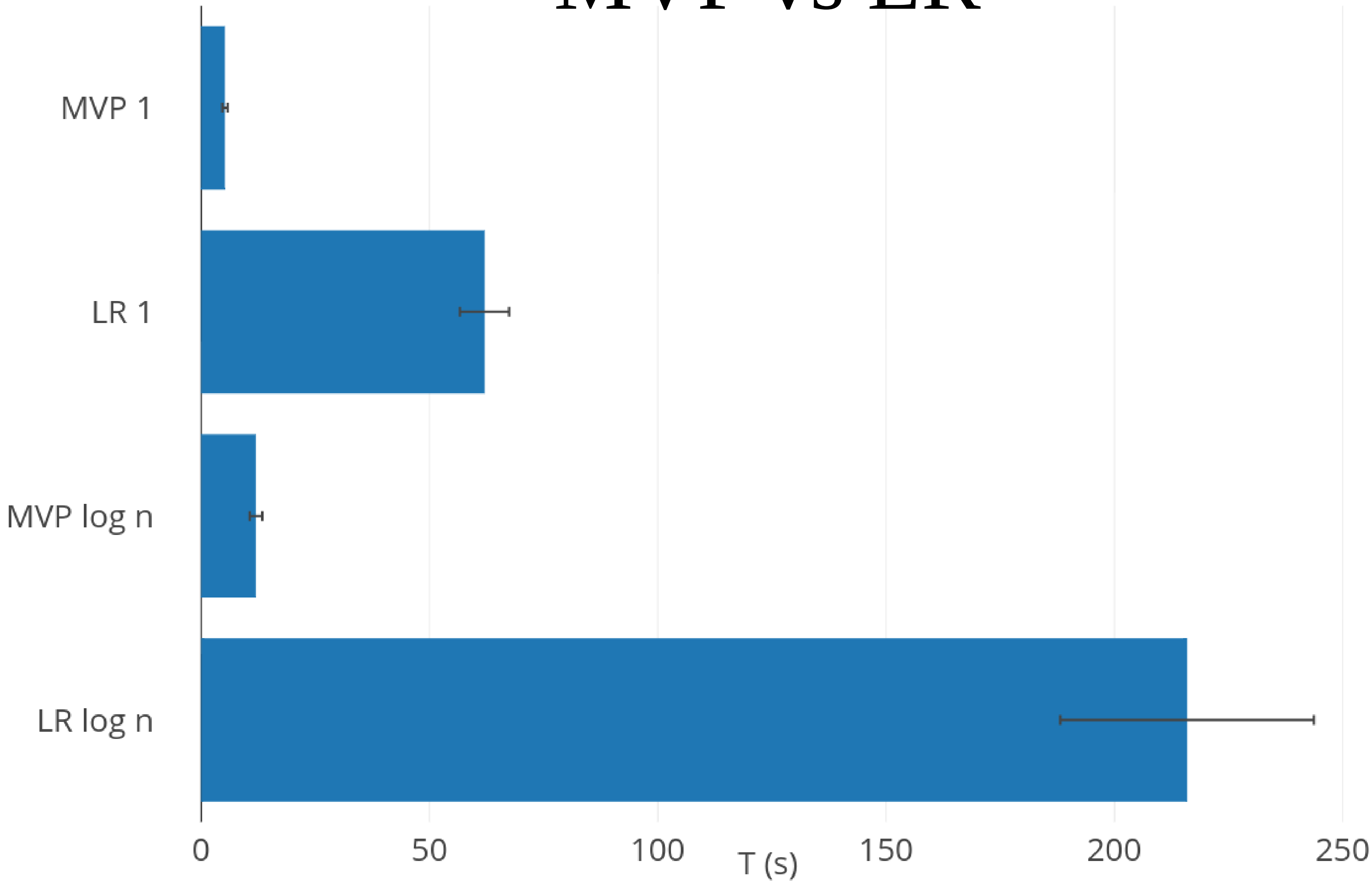
As a search fGES was used (Ramsey 2017) (Chickering 2002)

MVP vs LR



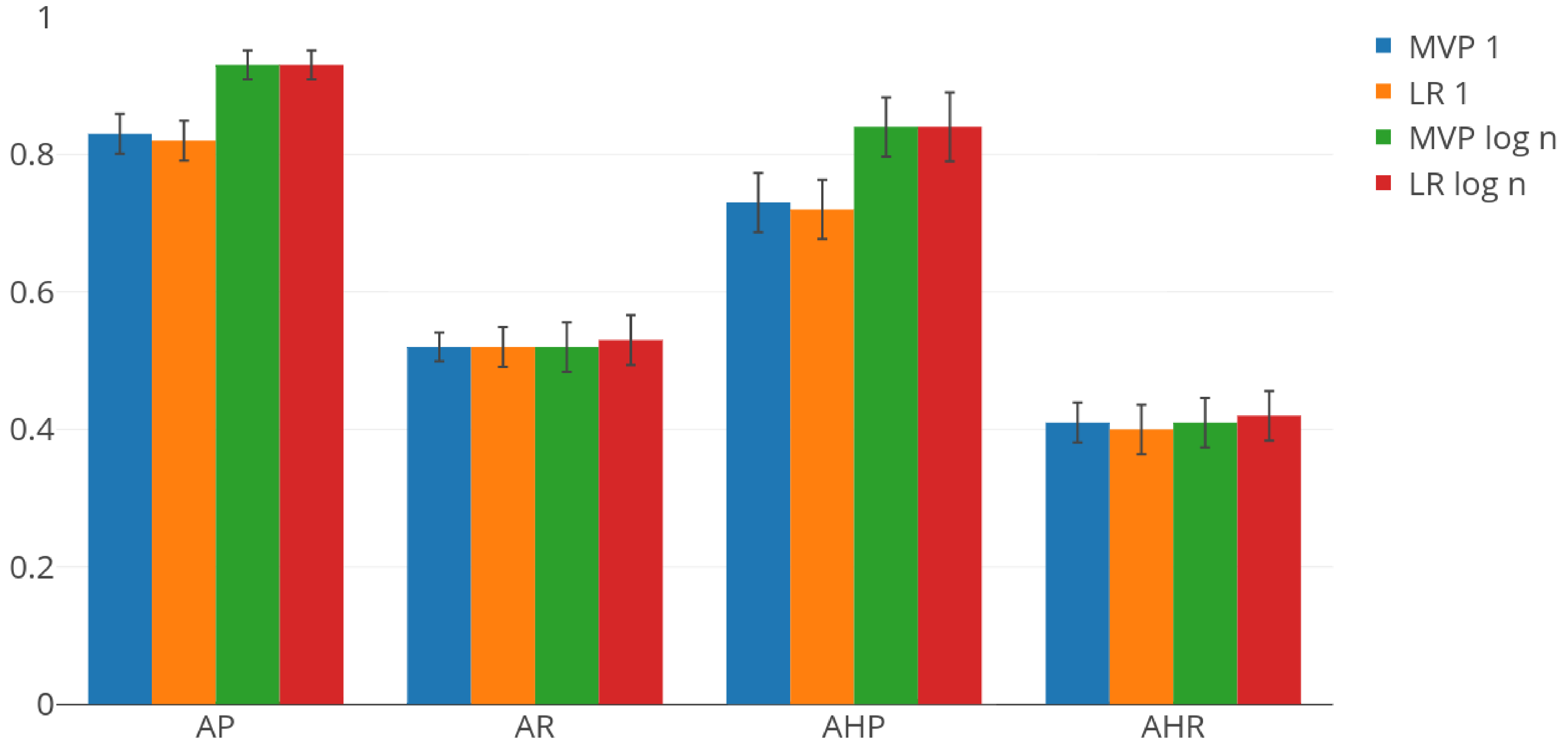
Avg Deg 4 | 100 Measured | Linear Simulation

MVP vs LR



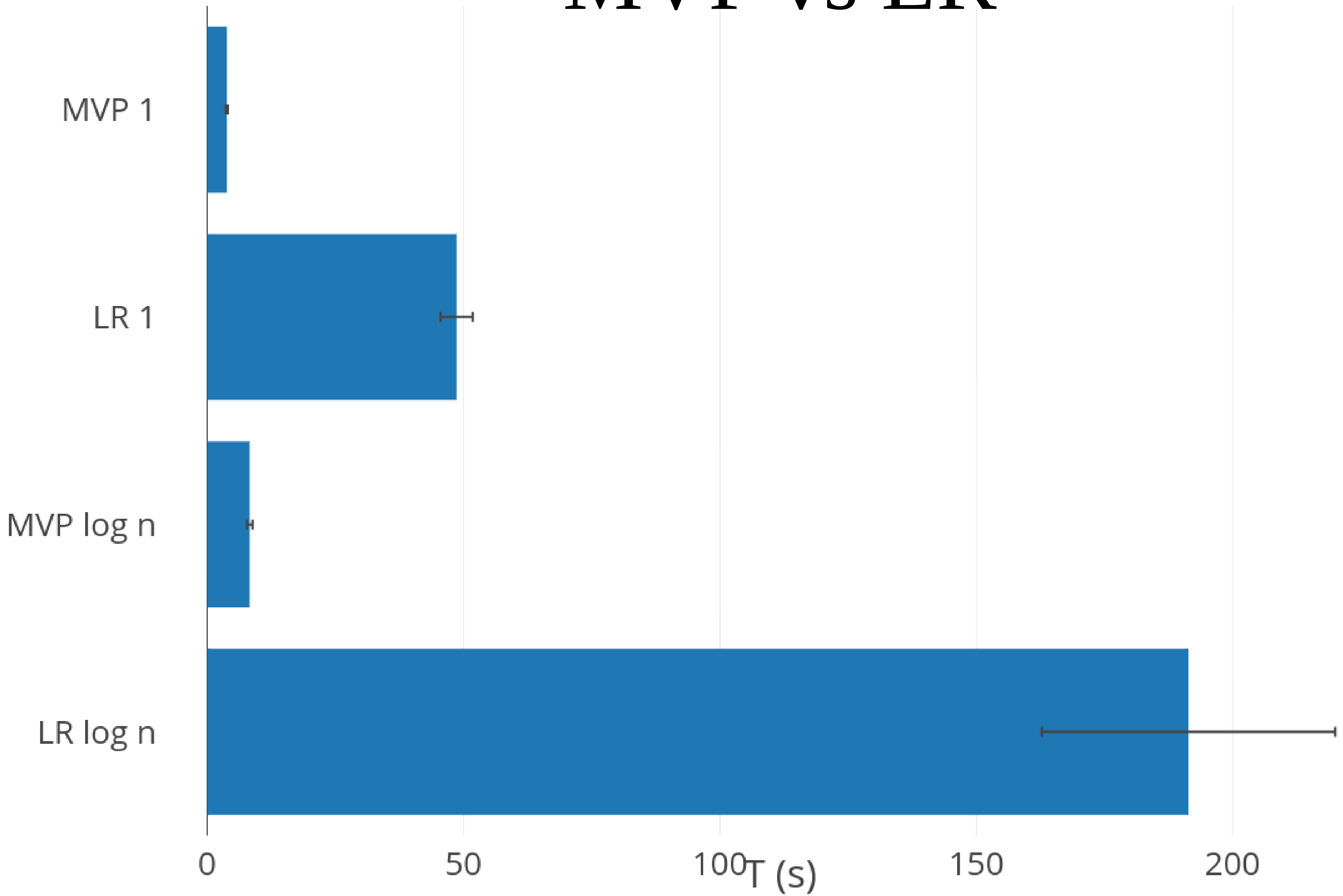
Avg Deg 4 | 100 Measured | Linear Simulation

MVP vs LR



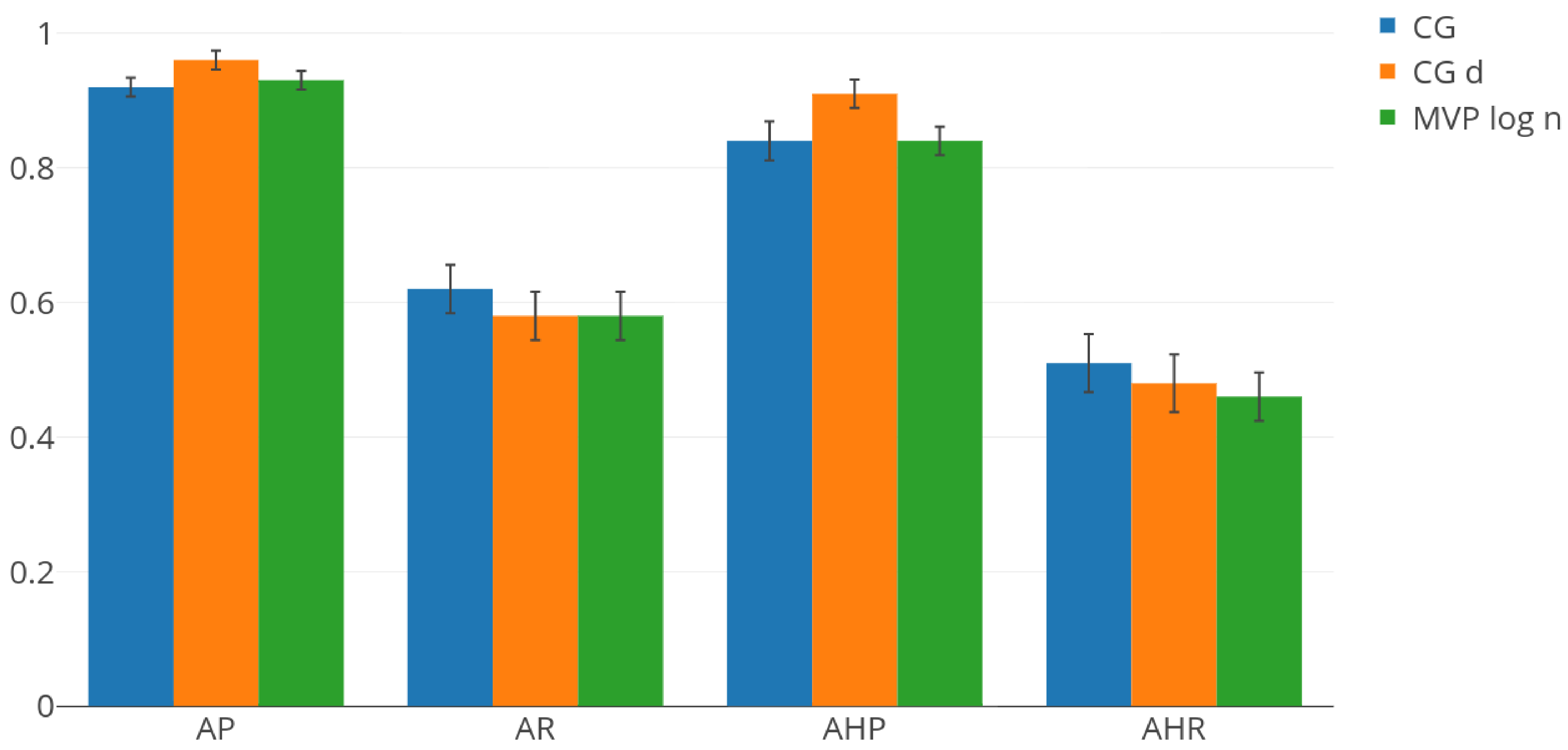
Avg Deg 4 | 100 Measured | Non-Linear Simulation

MVP vs LR



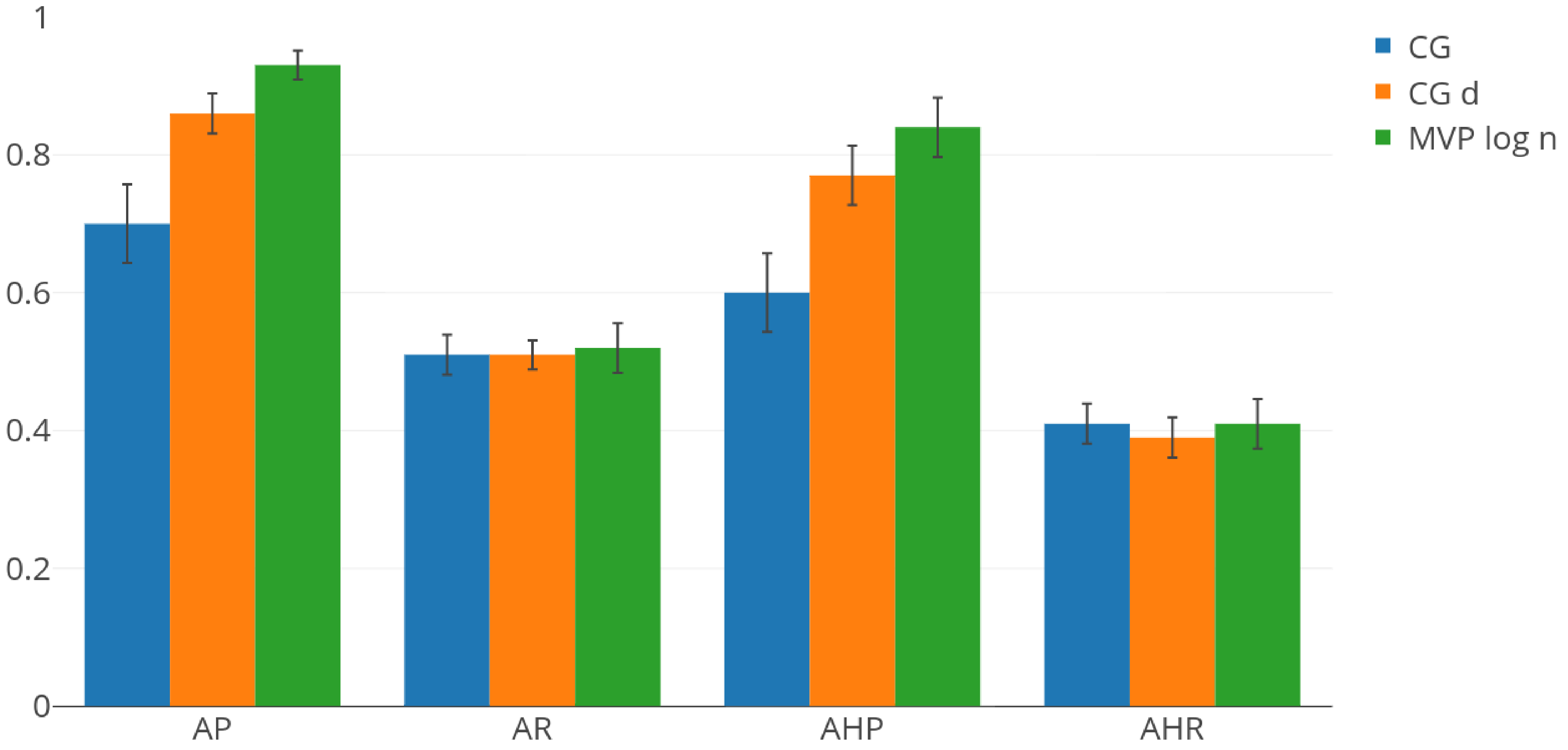
Avg Deg 4 | 100 Measured | Non-Linear Simulation

MVP vs CG



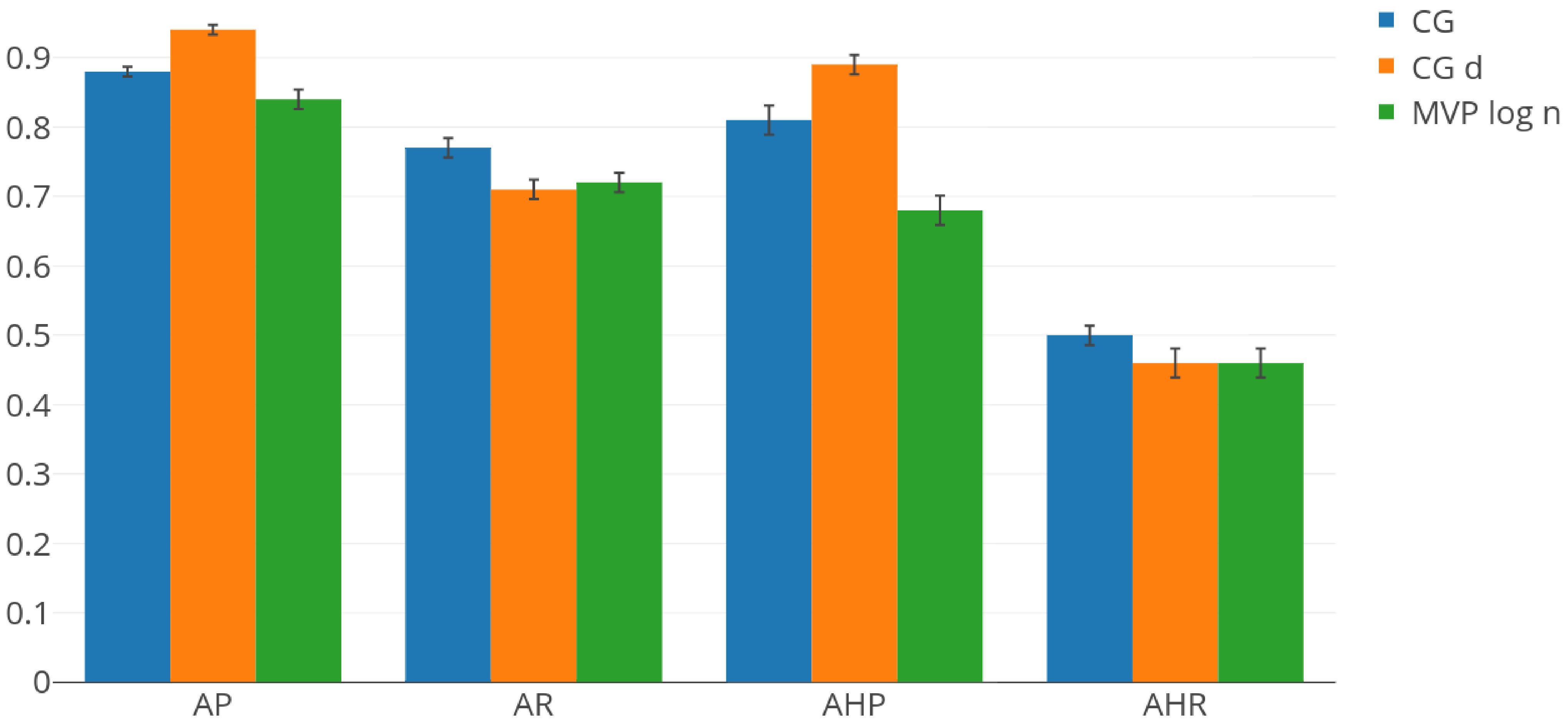
Avg Deg 4 | 100 Measured | Linear Simulation

MVP vs CG



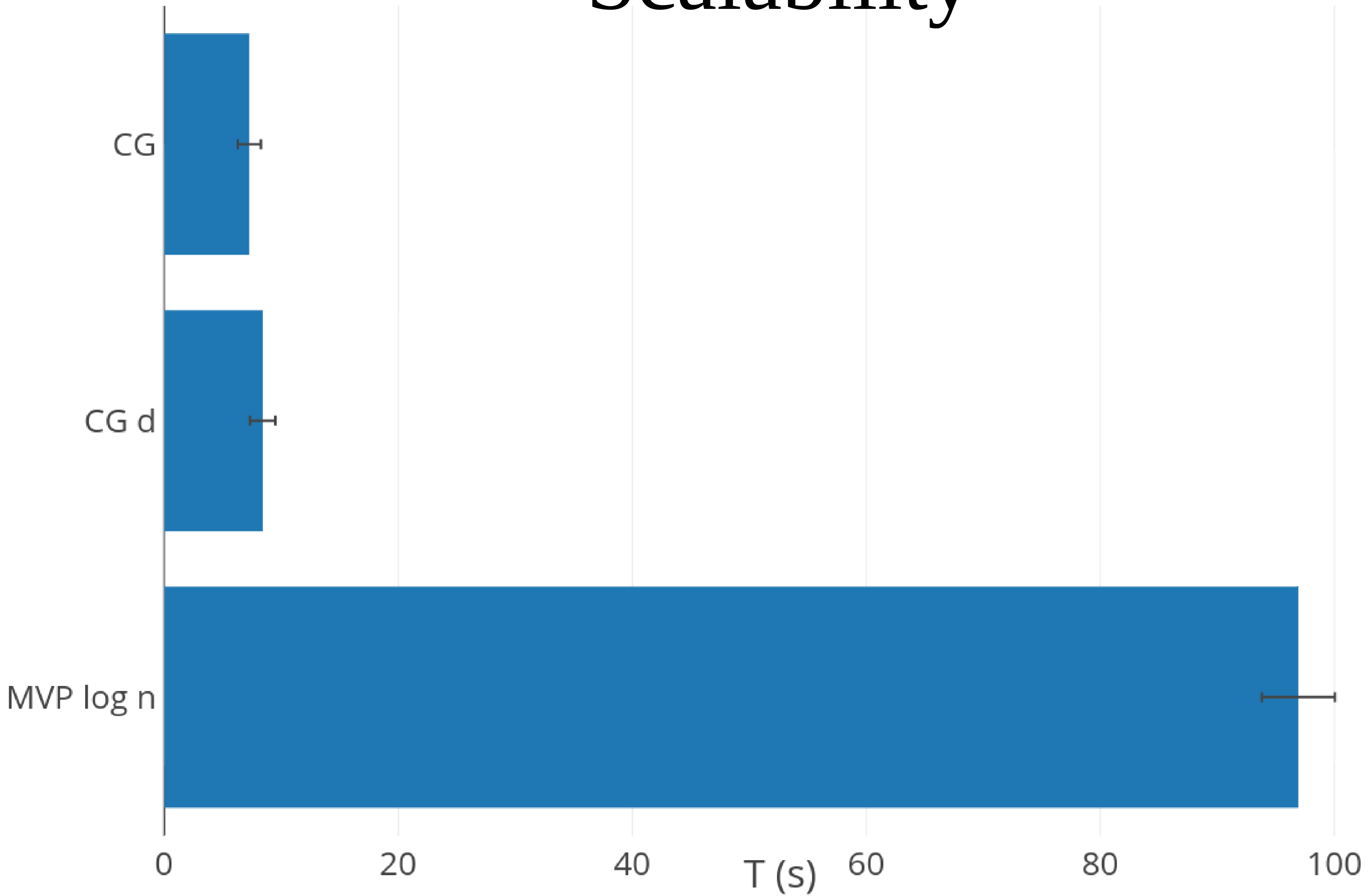
Avg Deg 4 | 100 Measured | Non-Linear Simulation

Scalability



Avg Deg 2 | 500 Measured | Linear Simulation

Scalability



Avg Deg 2 | 500 Measured | Linear Simulation

Conclusions

- We present two novel scoring methods for learning BNs in the presence of both continuous and discrete variables
  - Mixed Variable Polynomial (MVP)
    - Similar performance to LR but 10-20 times faster
    - Allows for a more general class of relationship
  - Conditional Gaussian (CG)
    - Quick and effective

Conclusions

- We present two novel scoring methods for learning BNs in the presence of both continuous and discrete variables
  - Mixed Variable Polynomial (MVP)
    - Similar performance to LR but 10-20 times faster
    - Allows for a more general class of relationship
  - Conditional Gaussian (CG)
    - Quick and effective
- Both scores perform well on simulated data (linear and non-linear) and scale to networks of at least 500 variables

Thank You

All presented methods are available
on Tetrad

`https://github.com/cmu-phil/tetrad`