Data thinning to overcome double dipping

Anna Neufeld

April, 2023

What is double dipping?

Classical statistical methods assume that we only ever test pre-specified hypotheses about pre-specified models.

What is double dipping?

Classical statistical methods assume that we only ever test pre-specified hypotheses about pre-specified models.

In reality, we explore our data, fit several models, select our favorite model, then test hypotheses about this model.

What is double dipping?

Classical statistical methods assume that we only ever test pre-specified hypotheses about pre-specified models.

In reality, we explore our data, fit several models, select our favorite model, then test hypotheses about this model.

Double Dipping: Using the same data for two tasks, such as:

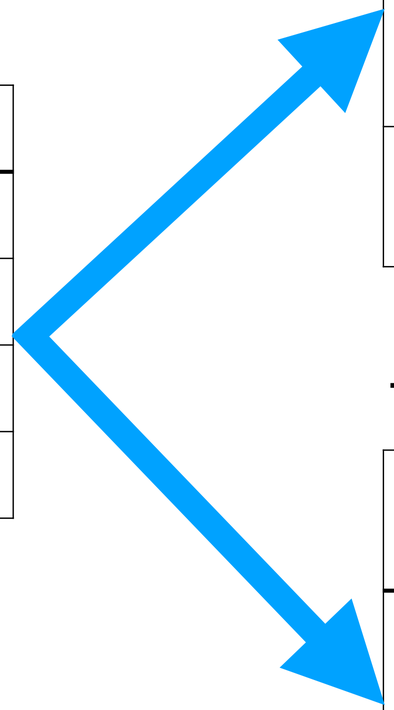
1. Generating and testing a null hypothesis.
2. Fitting and evaluating a model.

One possible solution: sample splitting

	Feature 1	Feature 2
Obs. 1	12	6
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34

One possible solution: sample splitting

	Feature 1	Feature 2
Obs. 1	12	6
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34



Train

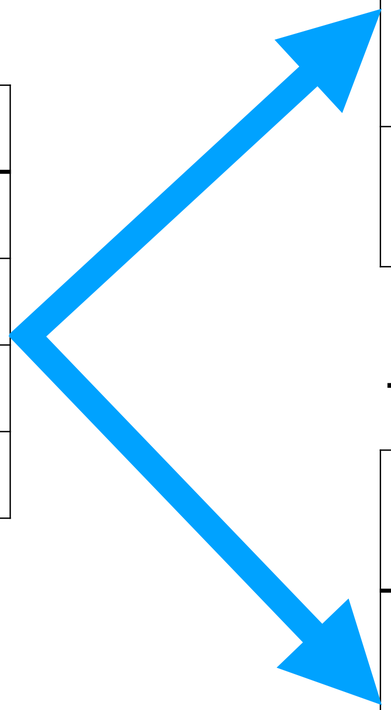
	Feature 1	Feature 2
Obs. 1	12	6
Obs. 2	31	8

Test

	Feature 1	Feature 2
Obs. 3	11	31
Obs. 4	22	34

One possible solution: sample splitting

	Feature 1	Feature 2
Obs. 1	12	6
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34



Train

	Feature 1	Feature 2
Obs. 1	12	6
Obs. 2	31	8

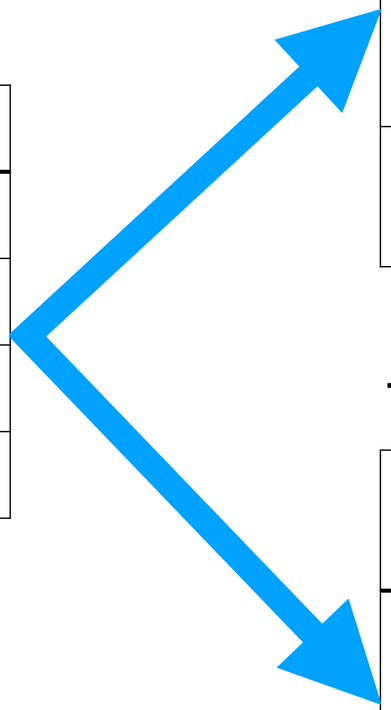
Test

	Feature 1	Feature 2
Obs. 3	11	31
Obs. 4	22	34

Select hypothesis.

One possible solution: sample splitting

	Feature 1	Feature 2
Obs. 1	12	6
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34



Train

	Feature 1	Feature 2
Obs. 1	12	6
Obs. 2	31	8

Select hypothesis.

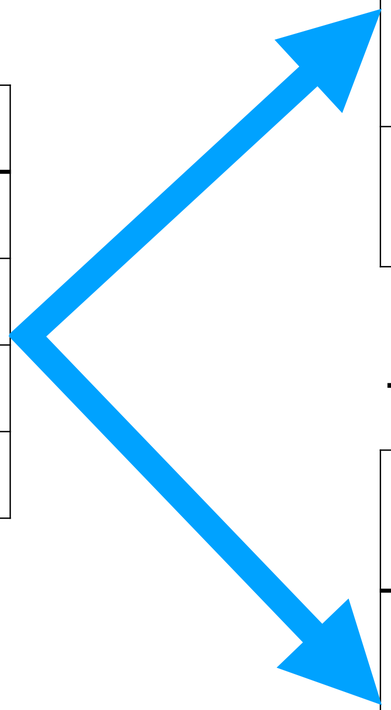
Test

	Feature 1	Feature 2
Obs. 3	11	31
Obs. 4	22	34

Test hypothesis.

One possible solution: sample splitting

	Feature 1	Feature 2
Obs. 1	12	6
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34



Train

	Feature 1	Feature 2
Obs. 1	12	6
Obs. 2	31	8

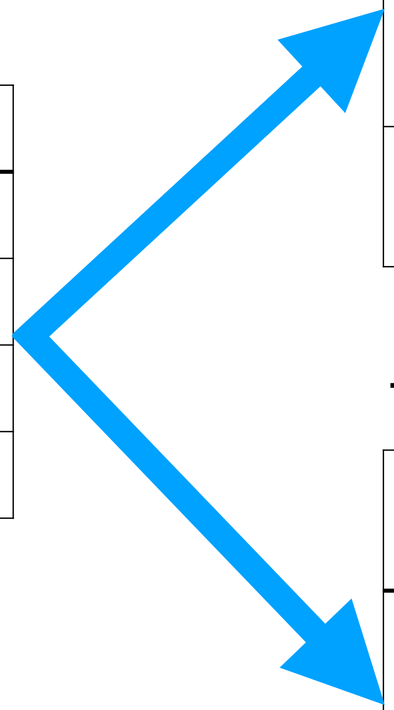
Fit model.

Test

	Feature 1	Feature 2
Obs. 3	11	31
Obs. 4	22	34

One possible solution: sample splitting

	Feature 1	Feature 2
Obs. 1	12	6
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34



Train

	Feature 1	Feature 2
Obs. 1	12	6
Obs. 2	31	8

Fit model.

Test

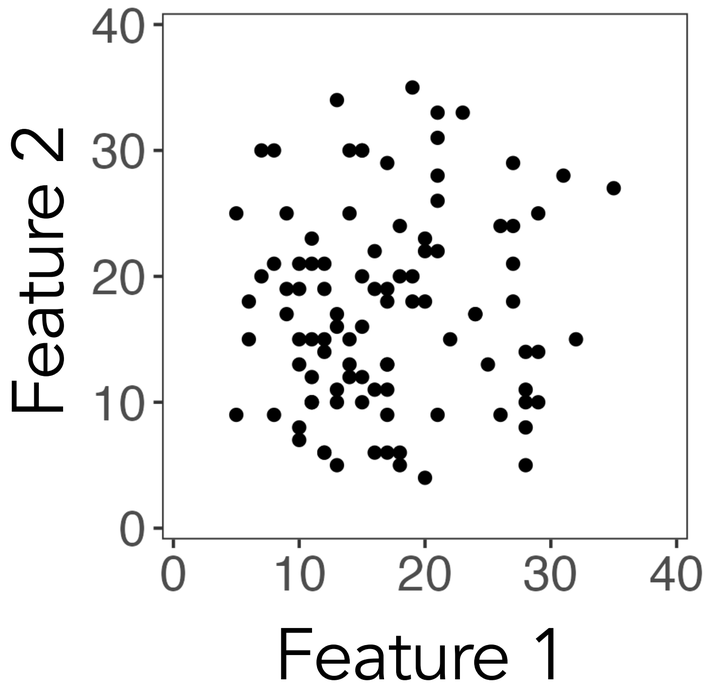
	Feature 1	Feature 2
Obs. 3	11	31
Obs. 4	22	34

Evaluate model.

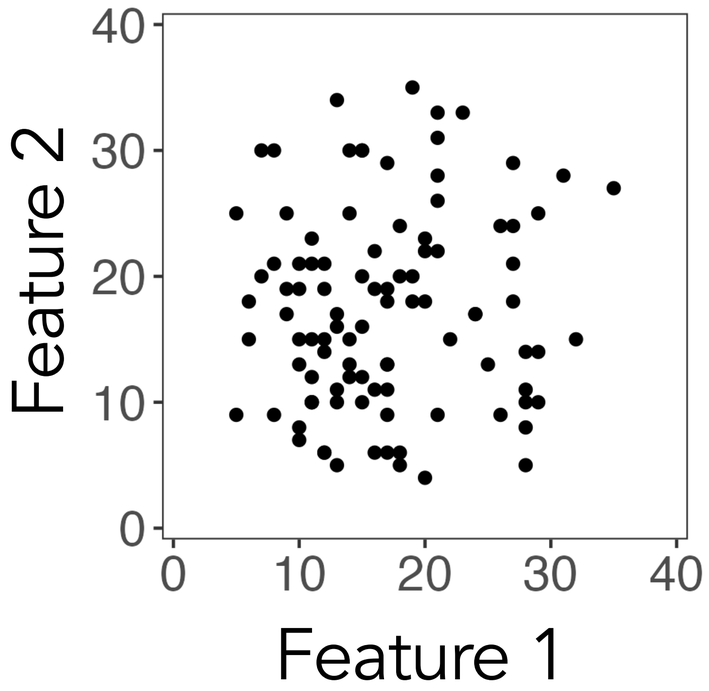
Outline

1. **Motivation: sample splitting doesn’t always work**
2. Poisson thinning
3. Data thinning
4. Generalized data thinning
5. Application to changepoint validation
6. Ongoing work

Example 1: using the same data to generate and test a hypothesis

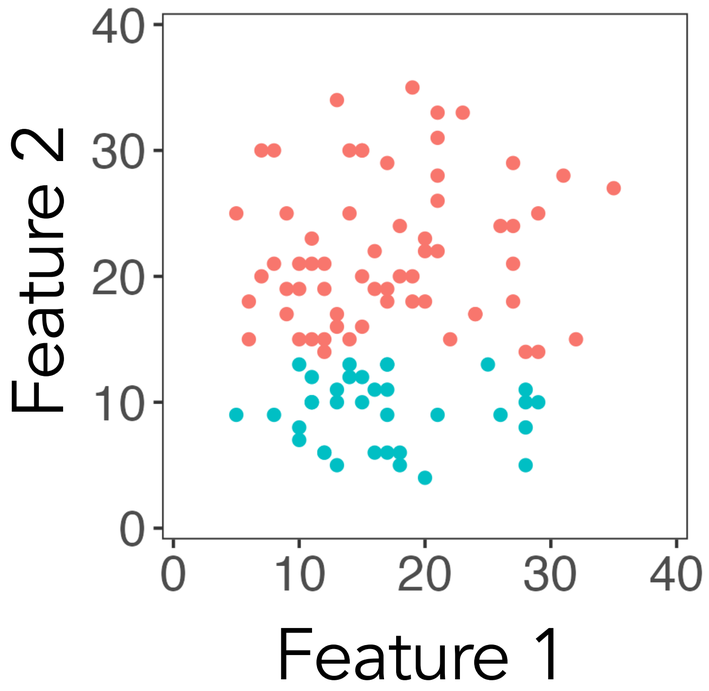


Example 1: using the same data to generate and test a hypothesis



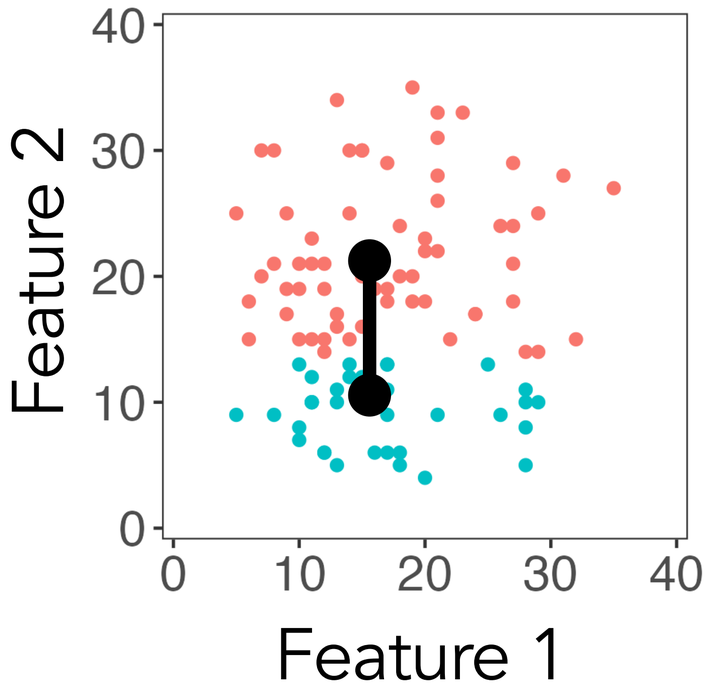
Step 1: cluster the observations.

Example 1: using the same data to generate and test a hypothesis



Step 1: cluster the observations.

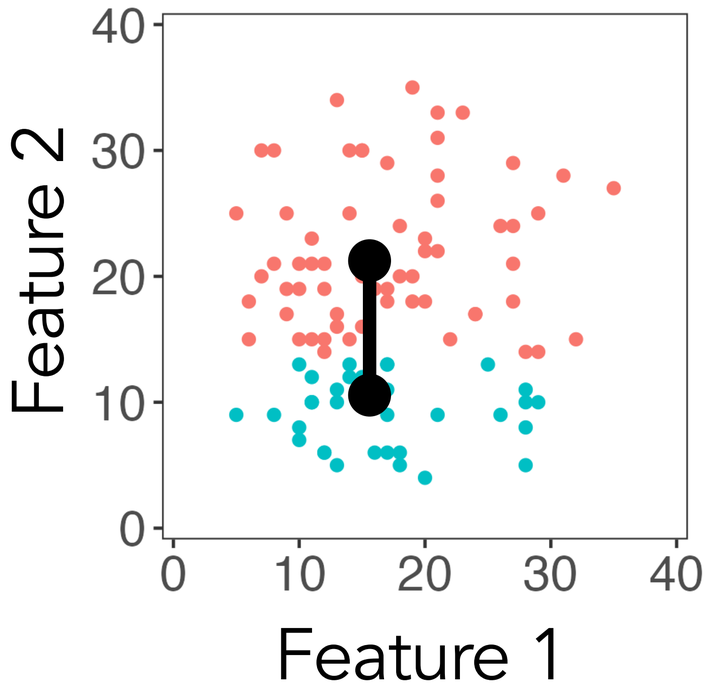
Example 1: using the same data to generate and test a hypothesis



Step 1: cluster the observations.

Generate H_0 :"the expected value of Feature 2 is the same between red observations and the blue observations."

Example 1: using the same data to generate and test a hypothesis

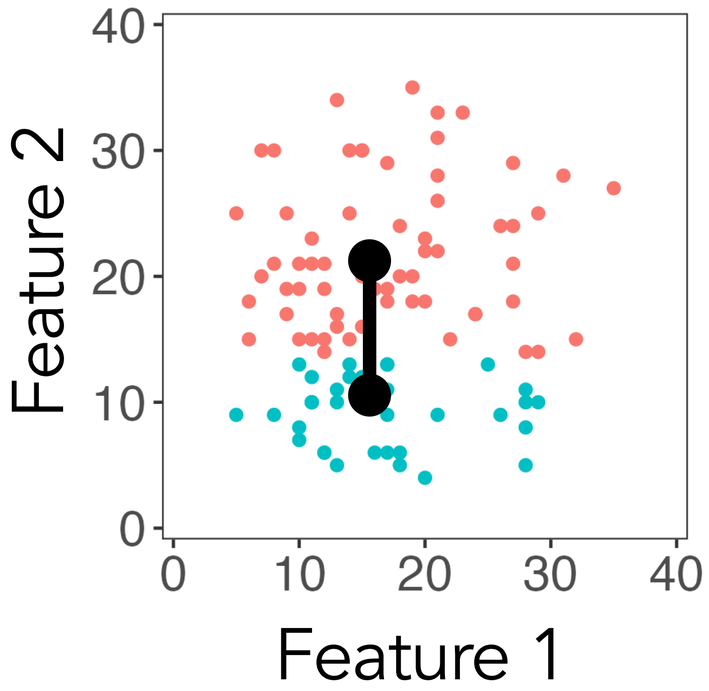


Step 1: cluster the observations.

Generate H_0 :"the expected value of Feature 2 is the same between red observations and the blue observations."

Step 2: test H_0 with a t-test.

Example 1: using the same data to generate and test a hypothesis



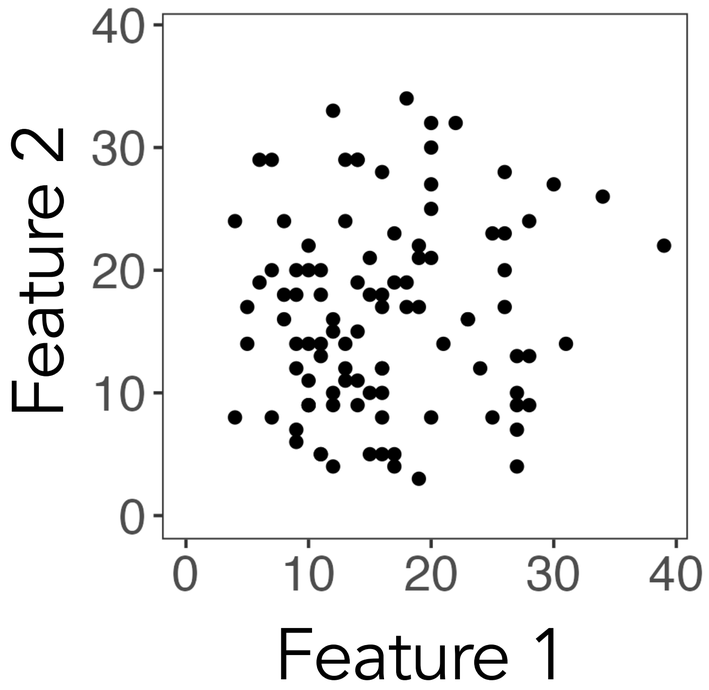
Step 1: cluster the observations.

Generate H_0 :"the expected value of Feature 2 is the same between red observations and the blue observations."

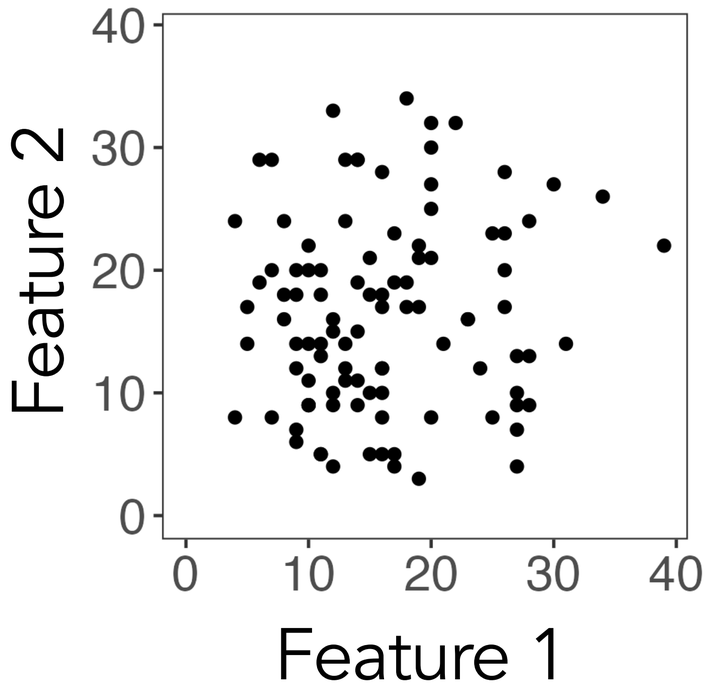
$p < 10^{-10}$ 😱

Step 2: test H_0 with a t-test.

Example 2: using the same data to fit and evaluate a model

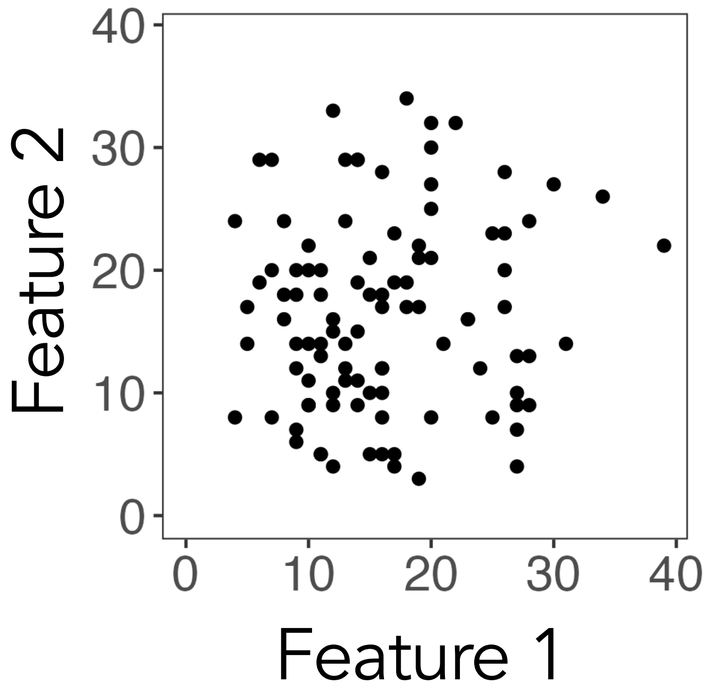


Example 2: using the same data to fit and evaluate a model



Goal: how many clusters are in this data?

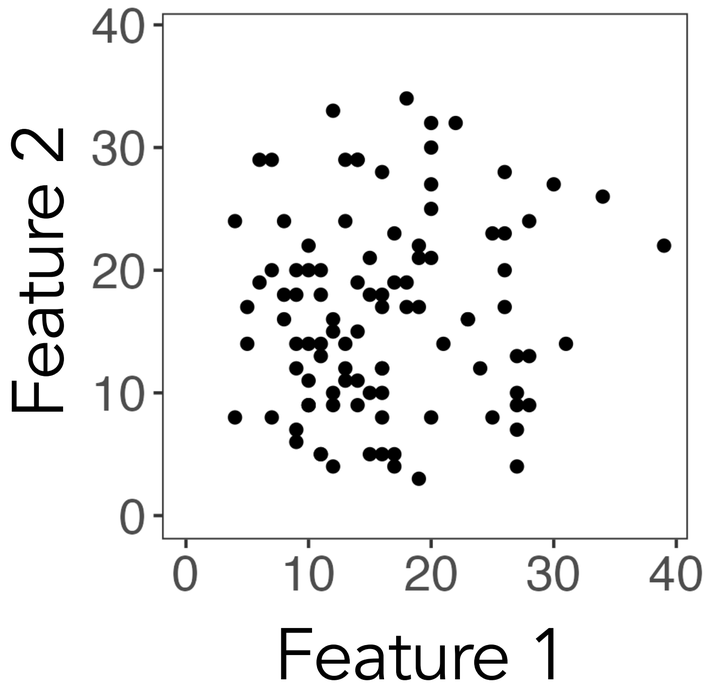
Example 2: using the same data to fit and evaluate a model



Goal: how many clusters are in this data?

For several values of k:

Example 2: using the same data to fit and evaluate a model

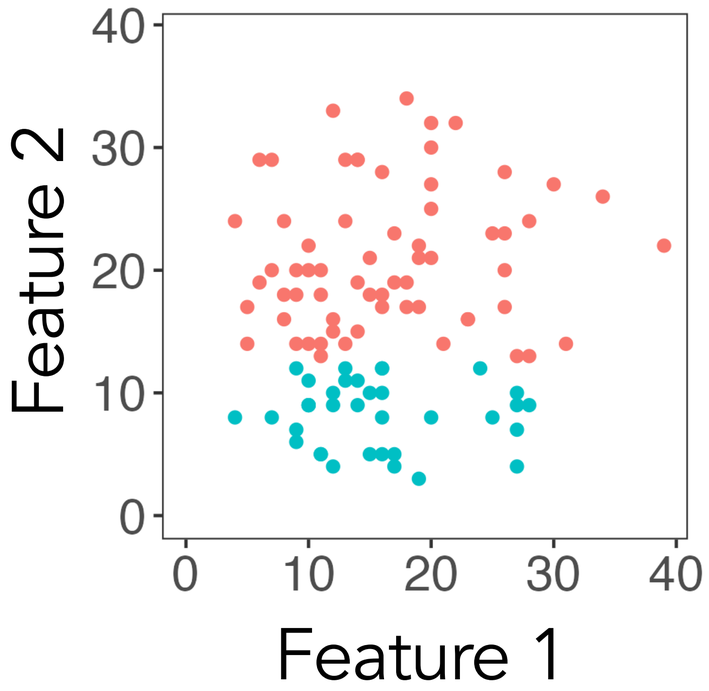


Goal: how many clusters are in this data?

For several values of k:

Step 1: fit a model with k clusters.

Example 2: using the same data to fit and evaluate a model

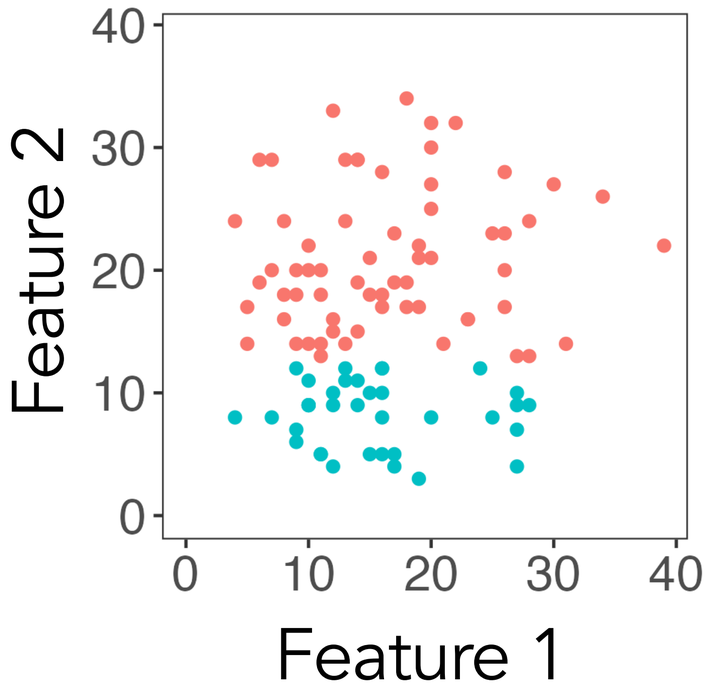


Goal: how many clusters are in this data?

For several values of k:

Step 1: fit a model with k clusters.

Example 2: using the same data to fit and evaluate a model



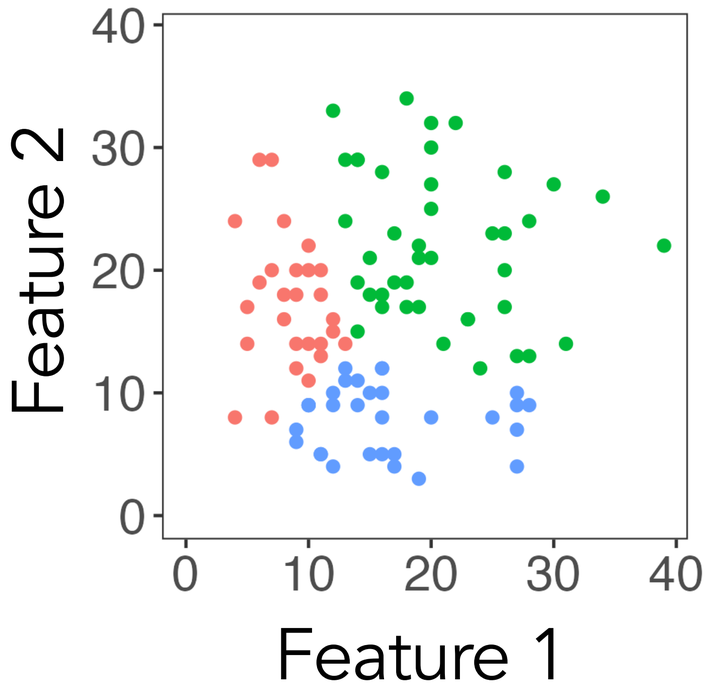
Goal: how many clusters are in this data?

For several values of k:

Step 1: fit a model with k clusters.

Step 2: evaluate model using a loss function.

Example 2: using the same data to fit and evaluate a model



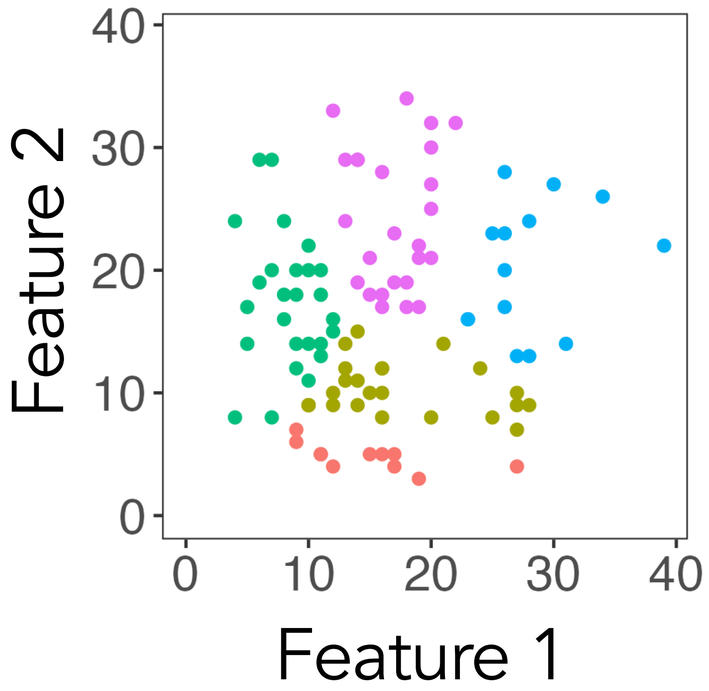
Goal: how many clusters are in this data?

For several values of k:

Step 1: fit a model with k clusters.

Step 2: evaluate model using a loss function.

Example 2: using the same data to fit and evaluate a model



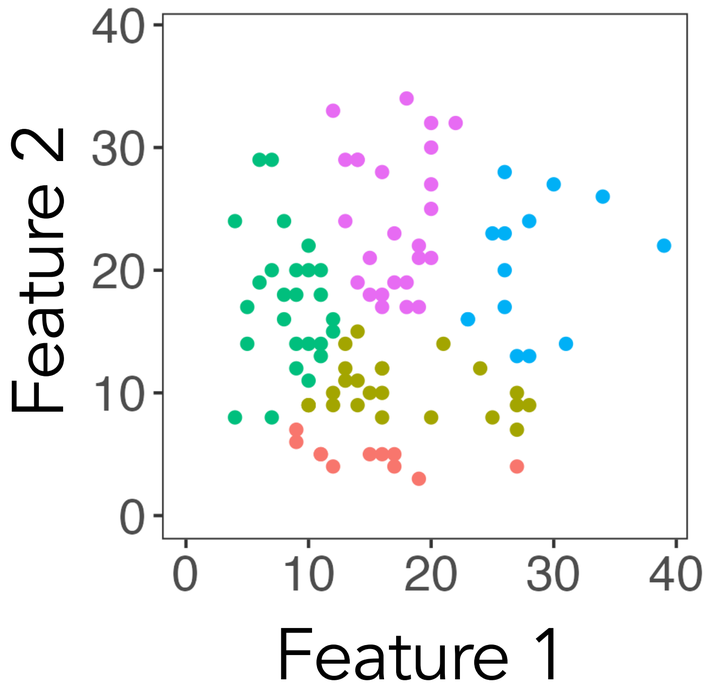
Goal: how many clusters are in this data?

For several values of k:

Step 1: fit a model with k clusters.

Step 2: evaluate model using a loss function.

Example 2: using the same data to fit and evaluate a model

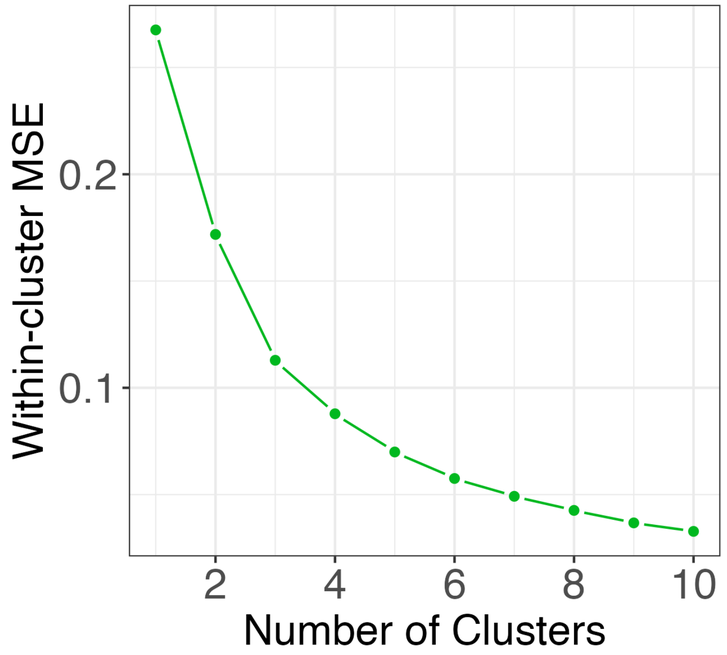


Goal: how many clusters are in this data?

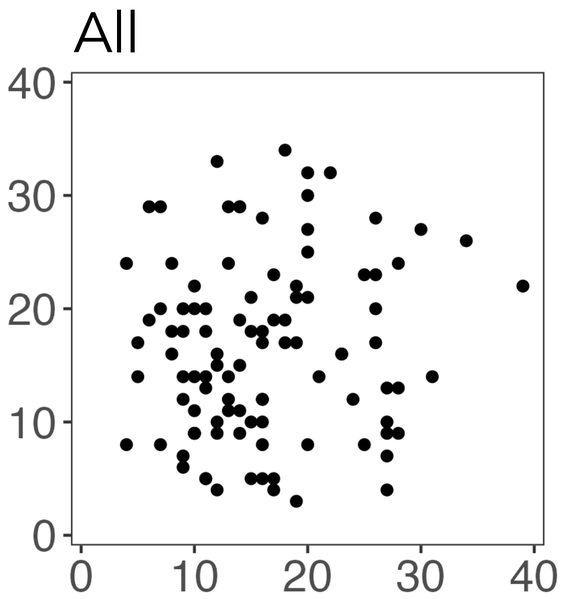
For several values of k:

Step 1: fit a model with k clusters.

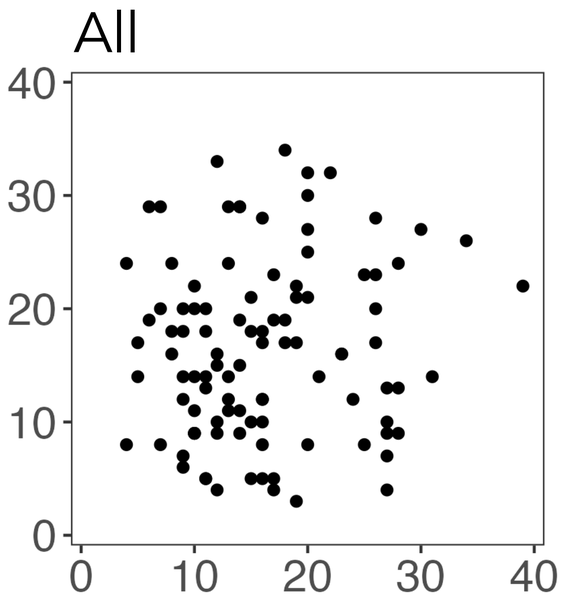
Step 2: evaluate model using a loss function.

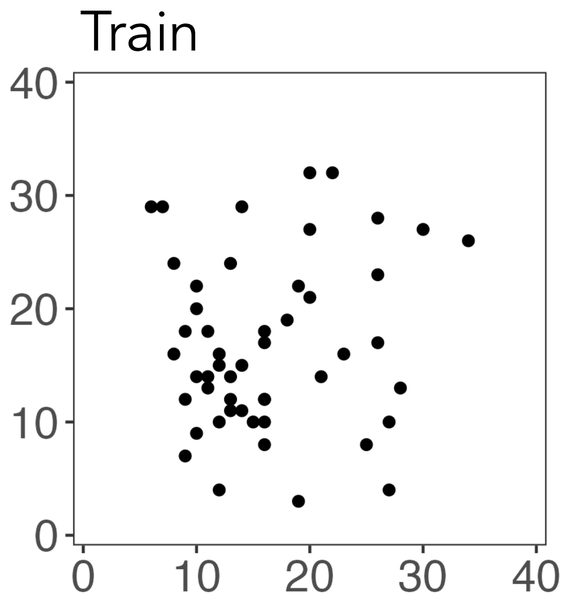


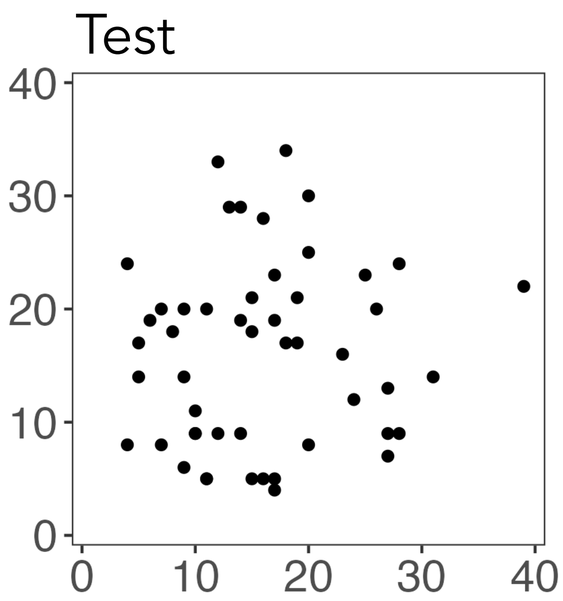
Sample splitting cannot be used for our motivating examples



Sample splitting cannot be used for our motivating examples

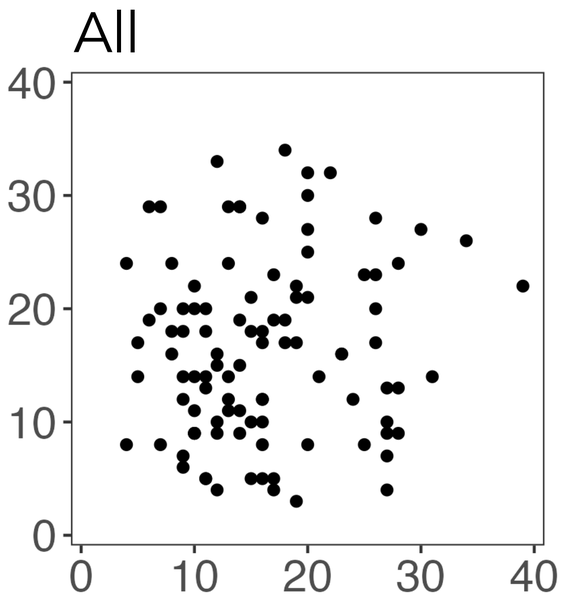


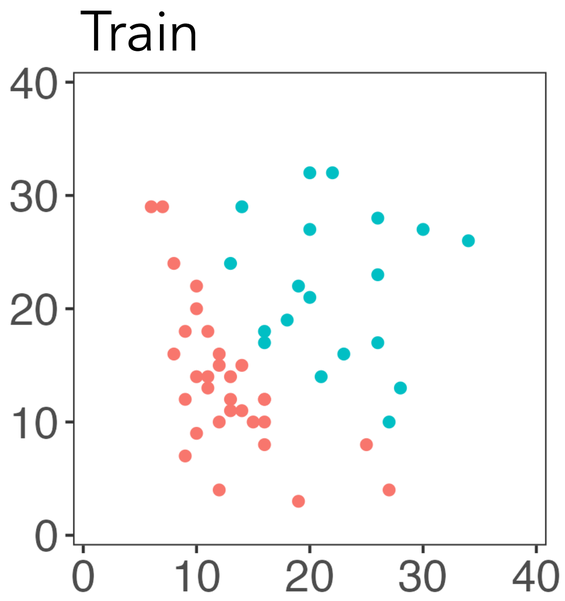


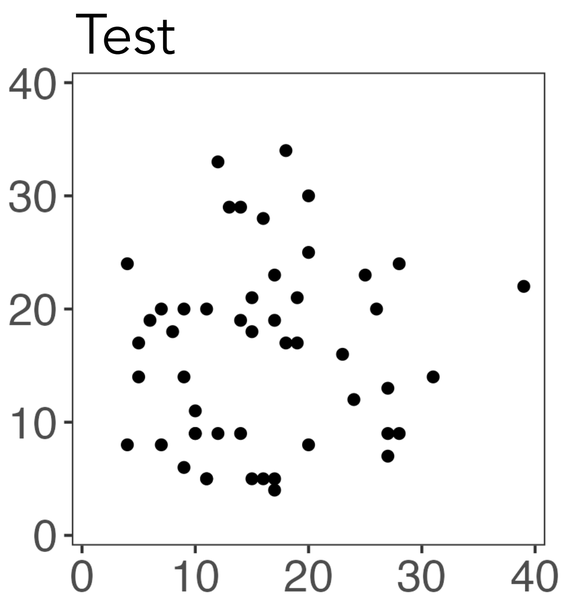


Step 1: split observations into train/test.

Sample splitting cannot be used for our motivating examples



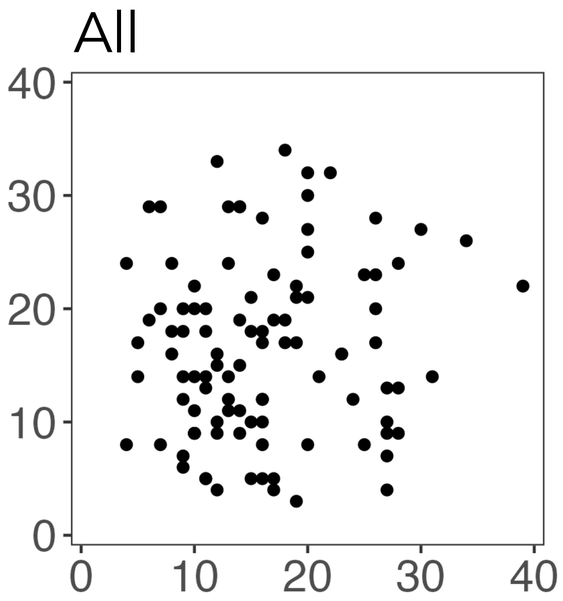


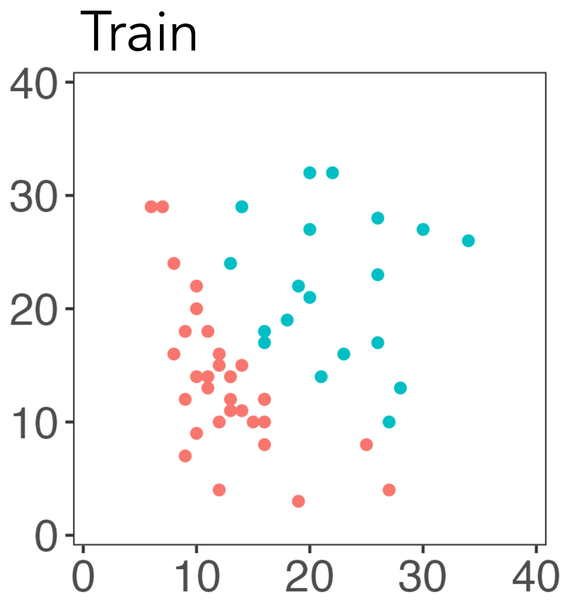


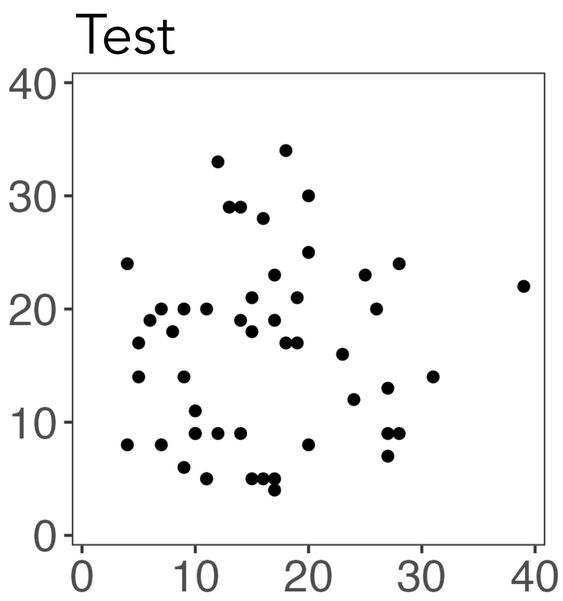
Step 1: split observations into train/test.

Step 2: cluster the training set.

Sample splitting cannot be used for our motivating examples





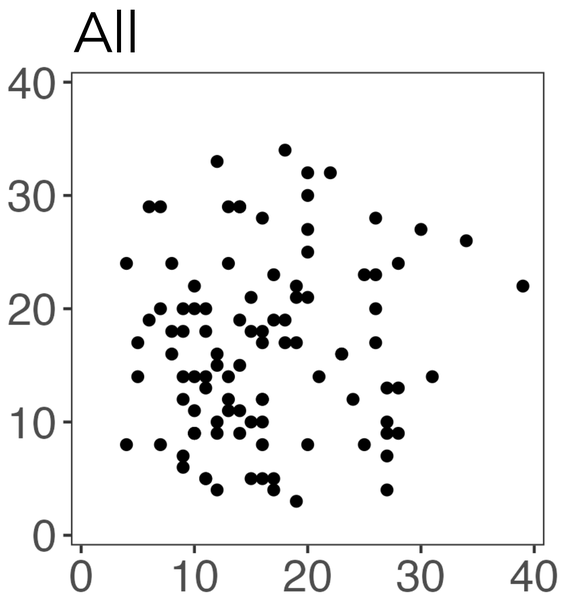


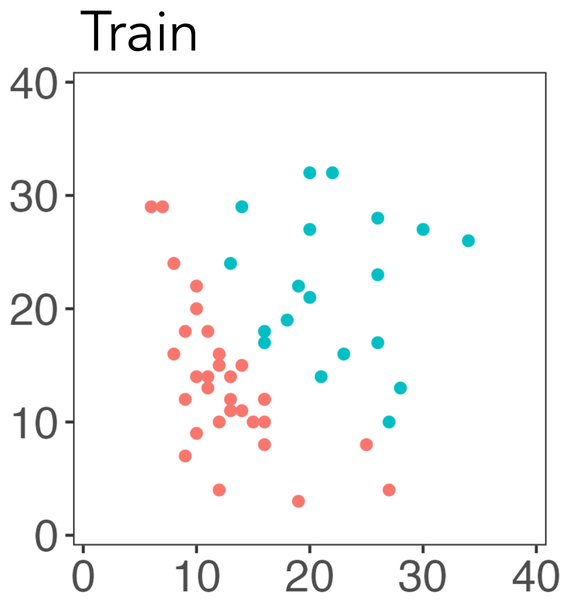
Step 1: split observations into train/test.

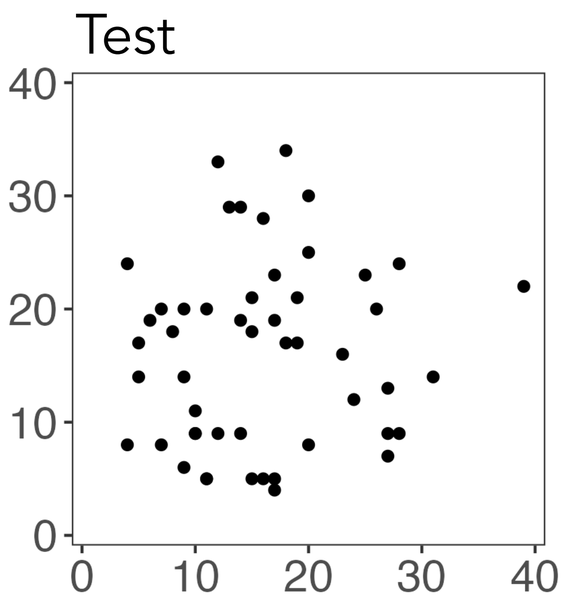
Step 2: cluster the training set.

Step 3: evaluate clusters or test for difference in means using test set.

Sample splitting cannot be used for our motivating examples







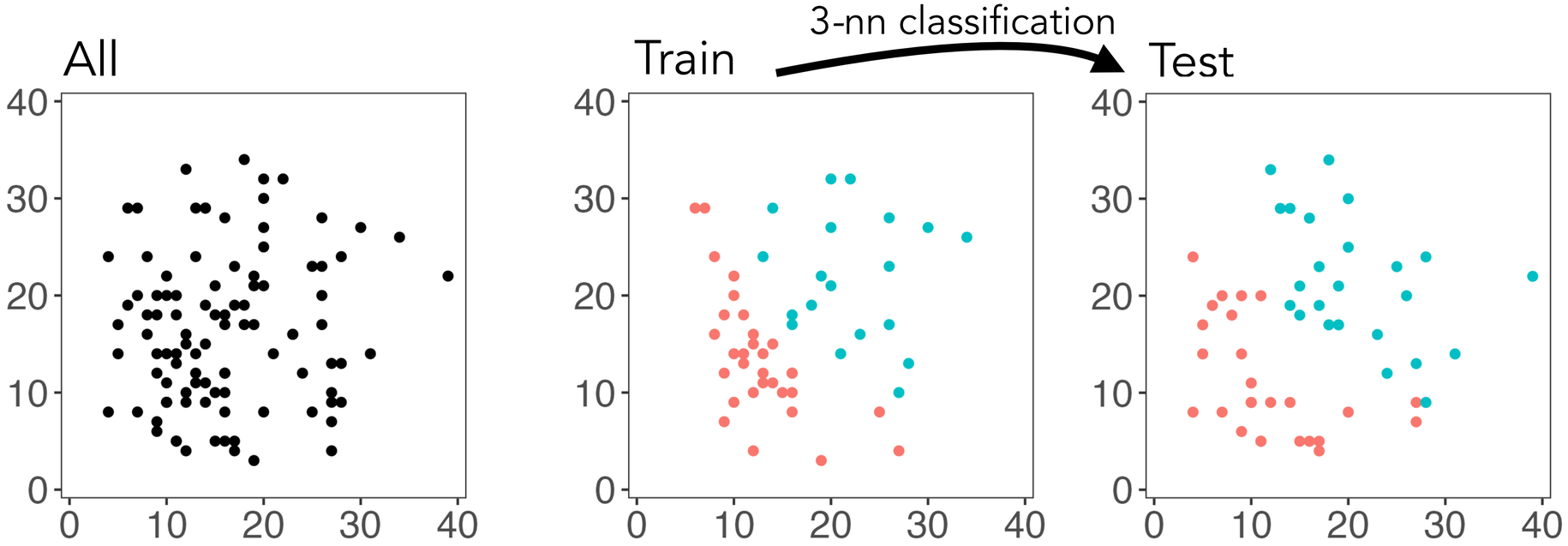
Step 1: split observations into train/test.

Step 2: cluster the training set.

Step 2.5: assign labels to observations in test set.

Step 3: evaluate clusters or test for difference in means using test set.

Sample splitting cannot be used for our motivating examples



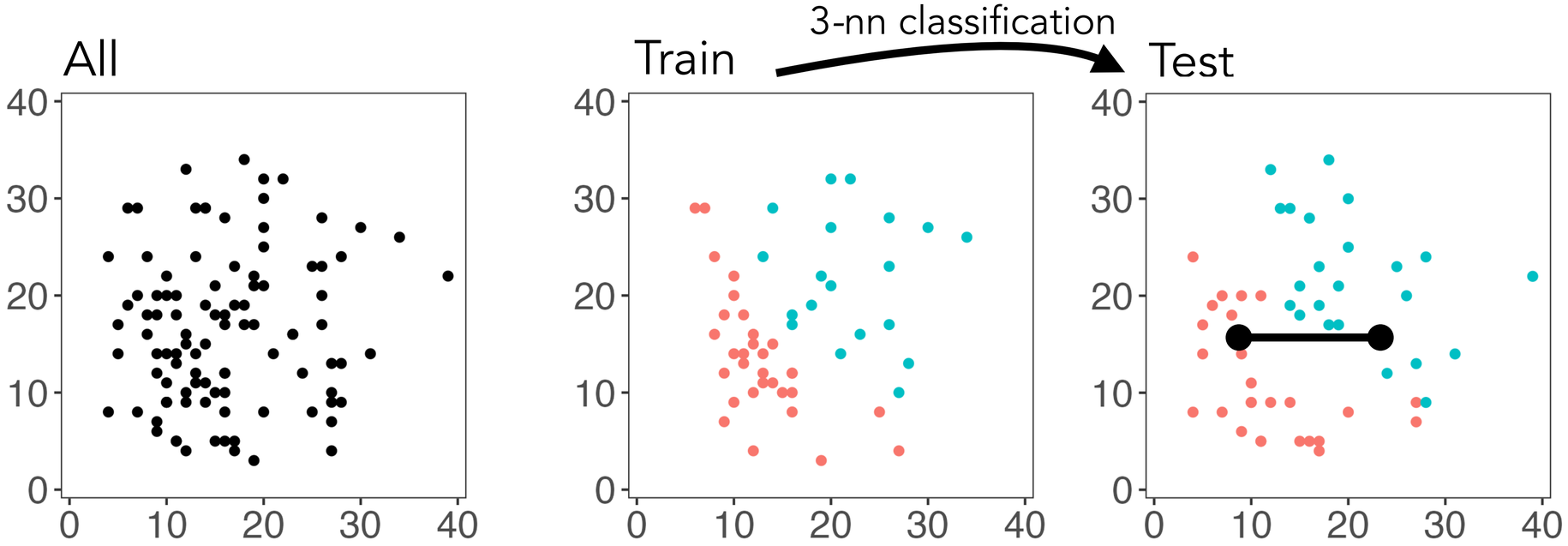
Step 1: split observations into train/test.

Step 2: cluster the training set.

Step 2.5: assign labels to observations in test set.

Step 3: evaluate clusters or test for difference in means using test set.

Sample splitting cannot be used for our motivating examples



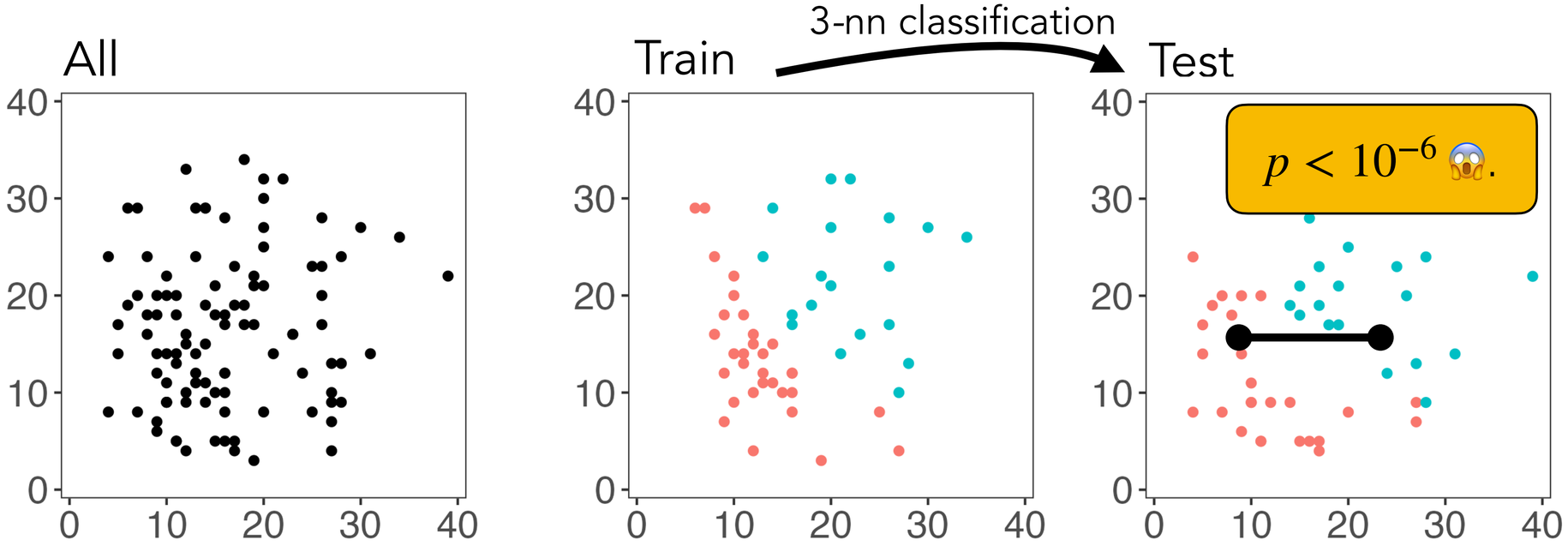
Step 1: split observations into train/test.

Step 2: cluster the training set.

Step 2.5: assign labels to observations in test set.

Step 3: evaluate clusters or test for difference in means using test set.

Sample splitting cannot be used for our motivating examples



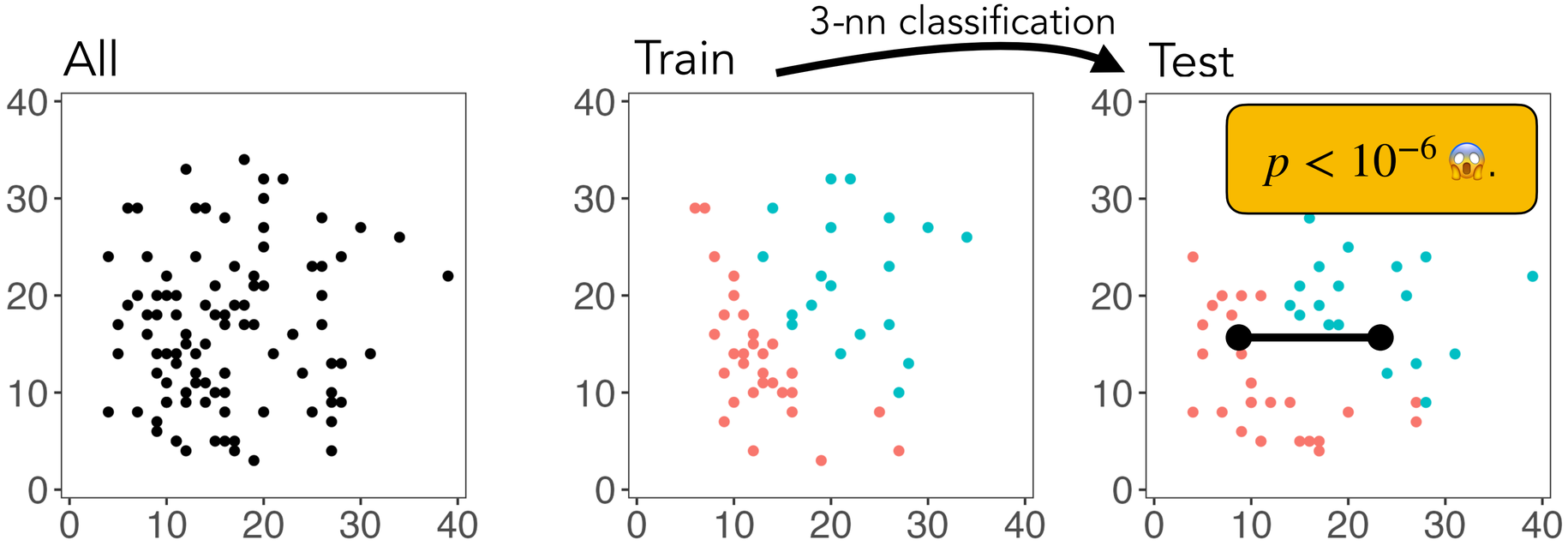
Step 1: split observations into train/test.

Step 2: cluster the training set.

Step 2.5: assign labels to observations in test set.

Step 3: evaluate clusters or test for difference in means using test set.

Sample splitting cannot be used for our motivating examples



Step 1: split observations into train/test.

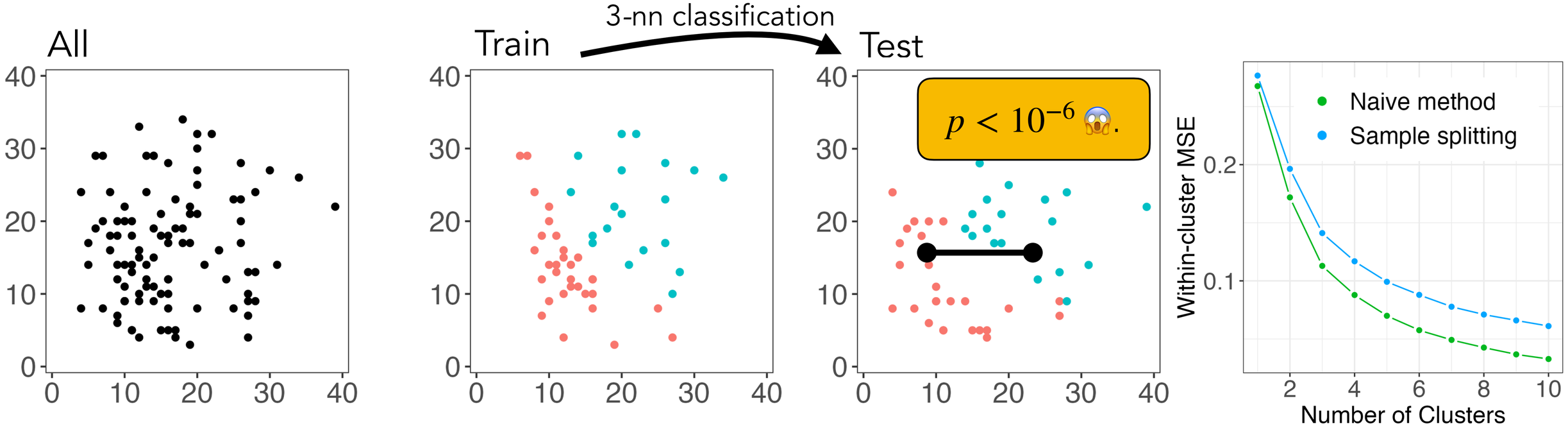
Step 2: cluster the training set.

Step 2.5: assign labels to observations in test set.

Step 3: evaluate clusters or test for difference in means using test set.

Gao, Bien, and Witten, 2022 (JASA).

Sample splitting cannot be used for our motivating examples



Step 1: split observations into train/test.

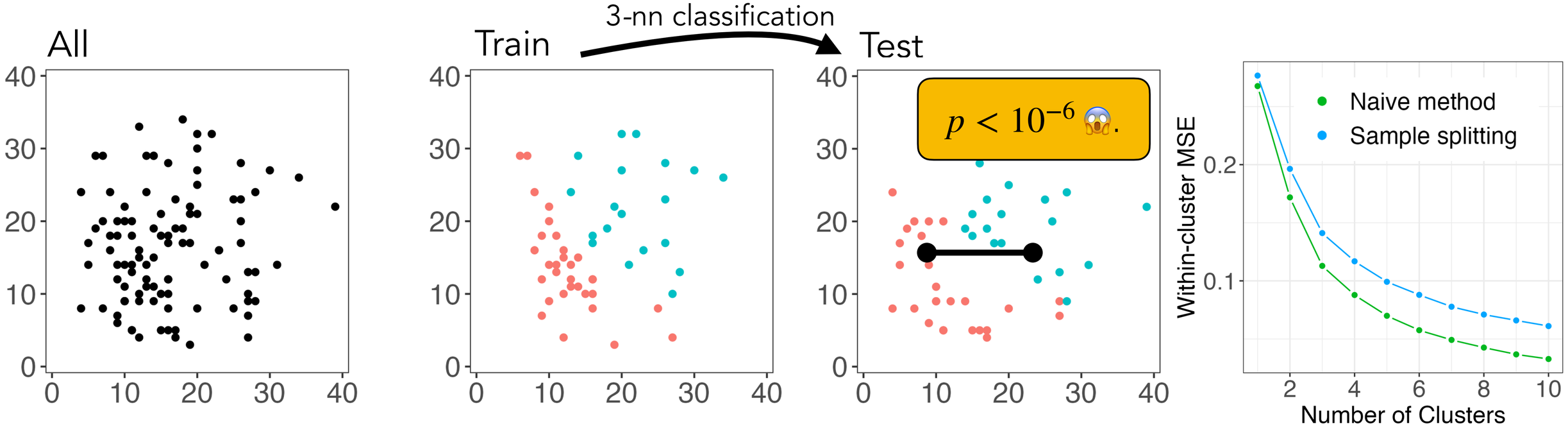
Step 2: cluster the training set.

Step 2.5: assign labels to observations in test set.

Step 3: evaluate clusters or test for difference in means using test set.

Gao, Bien, and Witten, 2022 (JASA).

Sample splitting cannot be used for our motivating examples



Step 1: split observations into train/test.

Step 2: cluster the training set.

Step 2.5: assign labels to observations in test set.

Step 3: evaluate clusters or test for difference in means using test set.

Gao, Bien, and Witten, 2022 (JASA).
Fu and Perry, 2020 (JCGS).

Other situations in which sample splitting is not a good option

1. Fixed-X regression settings.

2. Non-IID data.

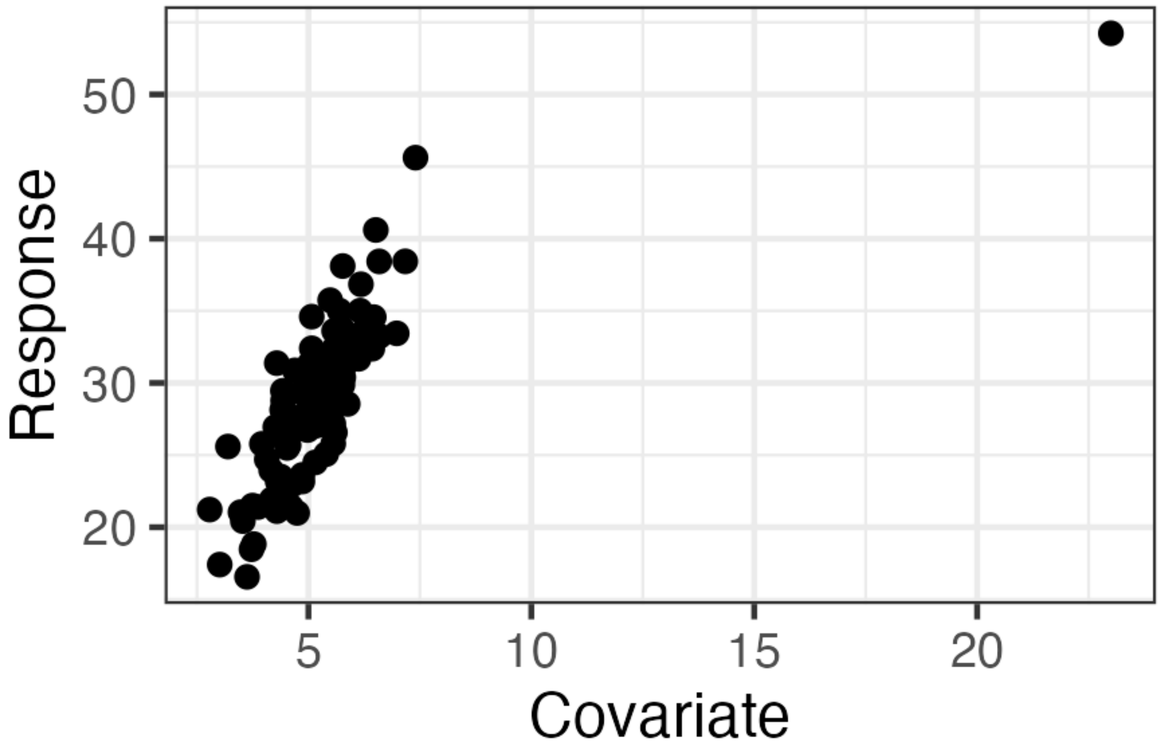
3. Data with outliers or
   influential points.

Other situations in which sample splitting is not a good option

1. Fixed-X regression settings.

2. Non-IID data.

3. Data with outliers or influential points.



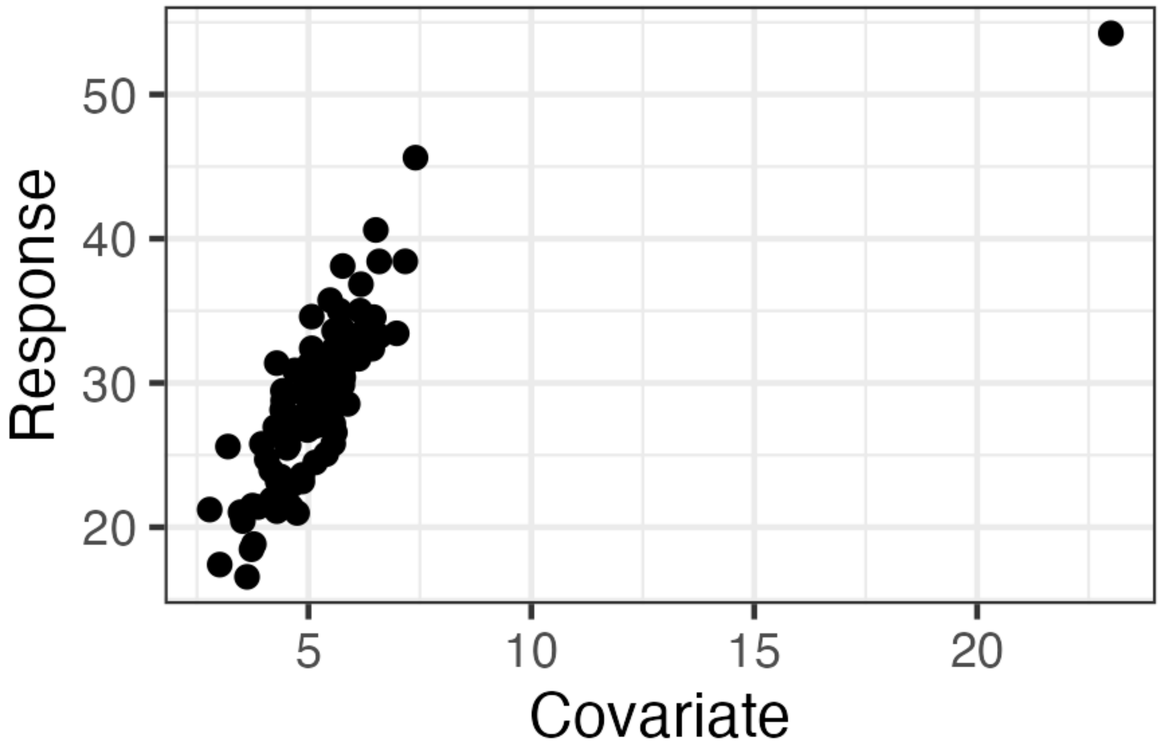
Other situations in which sample splitting is not a good option

1. Fixed-X regression settings.

2. Non-IID data.

3. Data with outliers or influential points.

Application later: changepoint detection.



Outline

1. Motivation: sample splitting doesn’t always work
2. **Poisson thinning**
3. Data thinning
4. Generalized data thinning
5. Application to changepoint validation
6. Ongoing work

Poisson thinning

X

	Feature 1	Feature 2
Obs. 1	18	6
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34

Poisson thinning

X

	Feature 1	Feature 2
Obs. 1	18	6
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34

$X^{(1)}$

	Feature 1	Feature 2
Obs. 1	14	1
Obs. 2	10	6
Obs. 3	5	17
Obs. 4	6	25

$X^{(2)}$

	Feature 1	Feature 2
Obs. 1	4	5
Obs. 2	21	2
Obs. 3	6	14
Obs. 3	16	9

Poisson thinning

X

	Feature 1	Feature 2
Obs. 1	18	6
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34

$X^{(1)}$

	Feature 1	Feature 2
Obs. 1	14	1
Obs. 2	[illegible]	6
Obs. 3	5	17
Obs. 4	6	25

$X^{(2)}$

	Feature 1	Feature 2
Obs. 1	[illegible]	5
Obs. 2	21	2
Obs. 3	6	14
Obs. 3	16	9

Poisson thinning

X

	Feature 1	Feature 2
Obs. 1	18	6
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34

X_{ij}

$X^{(1)}$

	Feature 1	Feature 2
Obs. 1	14	1
Obs. 2	10	6
Obs. 3	5	17
Obs. 4	6	25

$X^{(2)}$

	Feature 1	Feature 2
Obs. 1	4	5
Obs. 2	21	2
Obs. 3	6	14
Obs. 3	16	9

Poisson thinning

X

	Feature 1	Fea
Obs. 1	18	
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34

X_{ij}

$$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, \epsilon)$$

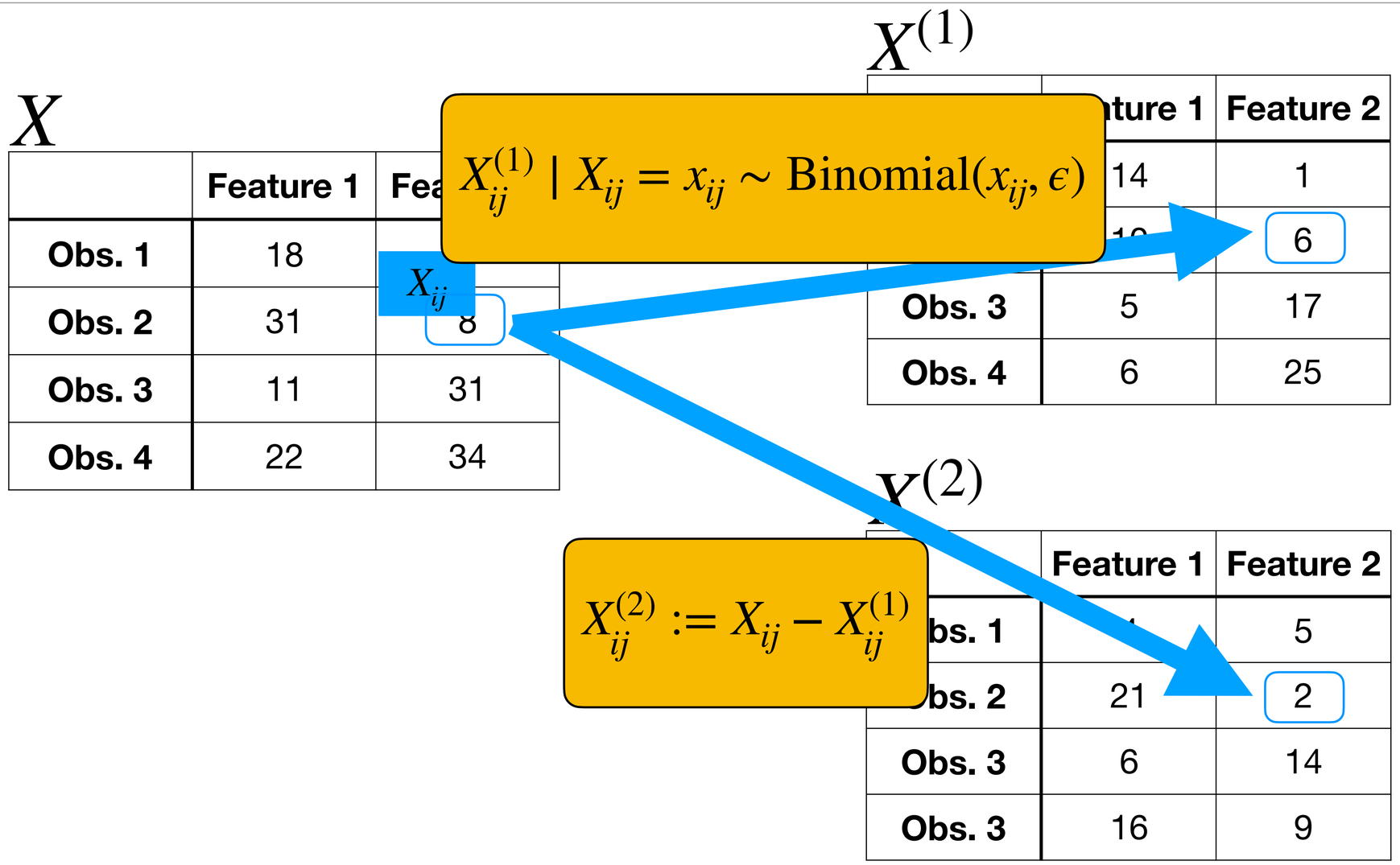
$X^{(1)}$

	ture 1	Feature 2
	14	1
		6
Obs. 3	5	17
Obs. 4	6	25

$X^{(2)}$

	Feature 1	Feature 2
Obs. 1		5
Obs. 2	21	2
Obs. 3	6	14
Obs. 3	16	9

Poisson thinning



Poisson thinning

X

	Feature 1	Fea
Obs. 1	18	
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34

X_{ij}

$$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, \epsilon)$$

$X^{(1)}$

	ture 1	Feature 2
	14	1
		6
Obs. 3	5	17
Obs. 4	6	25

$$X_{ij}^{(2)} := X_{ij} - X_{ij}^{(1)}$$

$X^{(2)}$

	Feature 1	Feature 2
bs. 1		5
bs. 2	21	2
Obs. 3	6	14
Obs. 3	16	9

If $X_{ij} \sim \text{Poisson}(\Lambda_{ij})$, then:

1. $X_{ij}^{(1)} \sim \text{Poisson}(\epsilon\Lambda_{ij})$
2. $X_{ij}^{(2)} \sim \text{Poisson}((1-\epsilon)\Lambda_{ij})$
3. $X_{ij}^{(1)} \perp\!\!\!\perp X_{ij}^{(2)}$

Poisson thinning

X

	Feature 1	Fea
Obs. 1	18	
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34

X_{ij}

$$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, \epsilon)$$

$X^{(1)}$

	ture 1	Feature 2
	14	1
		6
Obs. 3	5	17
Obs. 4	6	25

$$X_{ij}^{(2)} := X_{ij} - X_{ij}^{(1)}$$

$X^{(2)}$

	Feature 1	Feature 2
bs. 1		5
bs. 2	21	2
Obs. 3	6	14
Obs. 3	16	9

If $X_{ij} \sim \text{Poisson}(\Lambda_{ij})$, then:

1. $X_{ij}^{(1)} \sim \text{Poisson}(\epsilon\Lambda_{ij})$
2. $X_{ij}^{(2)} \sim \text{Poisson}((1-\epsilon)\Lambda_{ij})$
3. $X_{ij}^{(1)} \perp\!\!\!\perp X_{ij}^{(2)}$

A very well-known result: binomial thinning of the Poisson distribution.

Poisson thinning

X

	Feature 1	Fea
Obs. 1	18	
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34

X_{ij}

$$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, \epsilon)$$

$X^{(1)}$

	ture 1	Feature 2
	14	1
		6
Obs. 3	5	17
Obs. 4	6	25

Select hypothesis.

$$X_{ij}^{(2)} := X_{ij} - X_{ij}^{(1)}$$

$X^{(2)}$

	Feature 1	Feature 2
bs. 1		5
bs. 2	21	2
Obs. 3	6	14
Obs. 3	16	9

If $X_{ij} \sim \text{Poisson}(\Lambda_{ij})$, then:

1. $X_{ij}^{(1)} \sim \text{Poisson}(\epsilon\Lambda_{ij})$
2. $X_{ij}^{(2)} \sim \text{Poisson}((1-\epsilon)\Lambda_{ij})$
3. $X_{ij}^{(1)} \perp\!\!\!\perp X_{ij}^{(2)}$

A very well-known result: binomial thinning of the Poisson distribution.

Poisson thinning

X

	Feature 1	Fea
Obs. 1	18	
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34

X_{ij}

$$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, \epsilon)$$

$X^{(1)}$

	ture 1	Feature 2
	14	1
		6
Obs. 3	5	17
Obs. 4	6	25

Select hypothesis.

$$X_{ij}^{(2)} := X_{ij} - X_{ij}^{(1)}$$

$X^{(2)}$

	Feature 1	Feature 2
bs. 1		5
bs. 2	21	2
Obs. 3	6	14
Obs. 3	16	9

Test hypothesis.

If $X_{ij} \sim \text{Poisson}(\Lambda_{ij})$, then:

1. $X_{ij}^{(1)} \sim \text{Poisson}(\epsilon\Lambda_{ij})$
2. $X_{ij}^{(2)} \sim \text{Poisson}((1-\epsilon)\Lambda_{ij})$
3. $X_{ij}^{(1)} \perp\!\!\!\perp X_{ij}^{(2)}$

A very well-known result: binomial thinning of the Poisson distribution.

Poisson thinning

X

	Feature 1	Fea
Obs. 1	18	
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34

X_{ij}

$$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, \epsilon)$$

$X^{(1)}$

	ture 1	Feature 2
	14	1
		6
Obs. 3	5	17
Obs. 4	6	25

Fit model.

$$X_{ij}^{(2)} := X_{ij} - X_{ij}^{(1)}$$

$X^{(2)}$

	Feature 1	Feature 2
bs. 1		5
bs. 2	21	2
Obs. 3	6	14
Obs. 3	16	9

If $X_{ij} \sim \text{Poisson}(\Lambda_{ij})$, then:

1. $X_{ij}^{(1)} \sim \text{Poisson}(\epsilon\Lambda_{ij})$
2. $X_{ij}^{(2)} \sim \text{Poisson}((1-\epsilon)\Lambda_{ij})$
3. $X_{ij}^{(2)} \perp\!\!\!\perp X_{ij}^{(2)}$

A very well-known result: binomial thinning of the Poisson distribution.

Poisson thinning

X

	Feature 1	Fea
Obs. 1	18	
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34

X_{ij}

$$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, \epsilon)$$

$X^{(1)}$

	ture 1	Feature 2
	14	1
		6
Obs. 3	5	17
Obs. 4	6	25

Fit model.

$$X_{ij}^{(2)} := X_{ij} - X_{ij}^{(1)}$$

$X^{(2)}$

	Feature 1	Feature 2
bs. 1		5
bs. 2	21	2
Obs. 3	6	14
Obs. 3	16	9

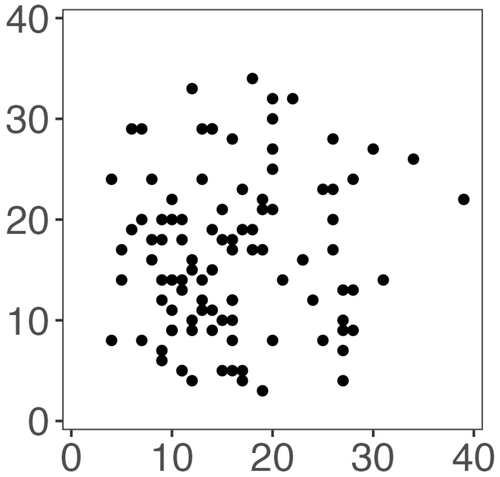
Evaluate model.

If $X_{ij} \sim \text{Poisson}(\Lambda_{ij})$, then:

1. $X_{ij}^{(1)} \sim \text{Poisson}(\epsilon\Lambda_{ij})$
2. $X_{ij}^{(2)} \sim \text{Poisson}((1-\epsilon)\Lambda_{ij})$
3. $X_{ij}^{(2)} \perp\!\!\!\perp X_{ij}^{(2)}$

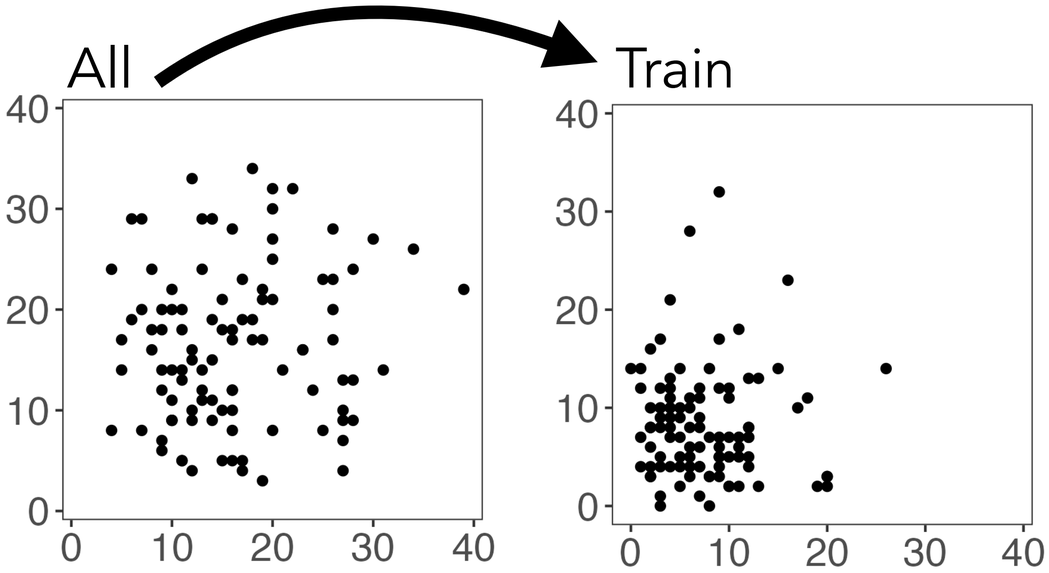
A very well-known result: binomial thinning of the Poisson distribution.

Thinning avoids the pitfall of sample splitting on our motivating examples



Thinning avoids the pitfall of sample splitting on our motivating examples

$$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, 0.5)$$

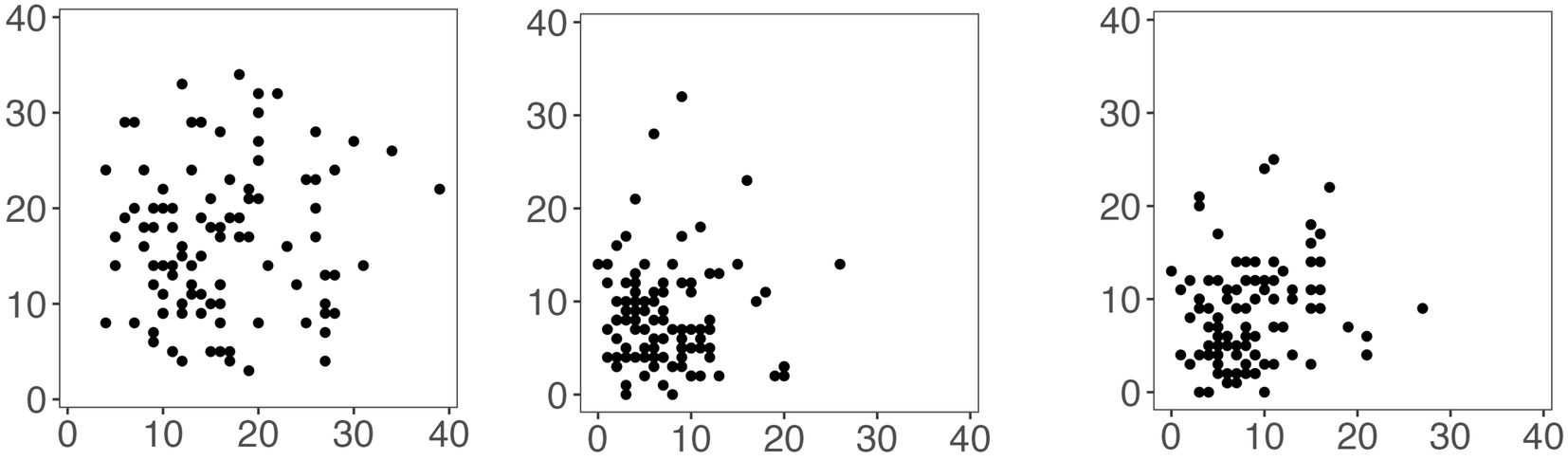


Step 1: thin observations into train/test.

Thinning avoids the pitfall of sample splitting on our motivating examples

$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, 0.5)$

$X_{ij}^{(2)} = X_{ij} - X_{ij}^{(1)}$

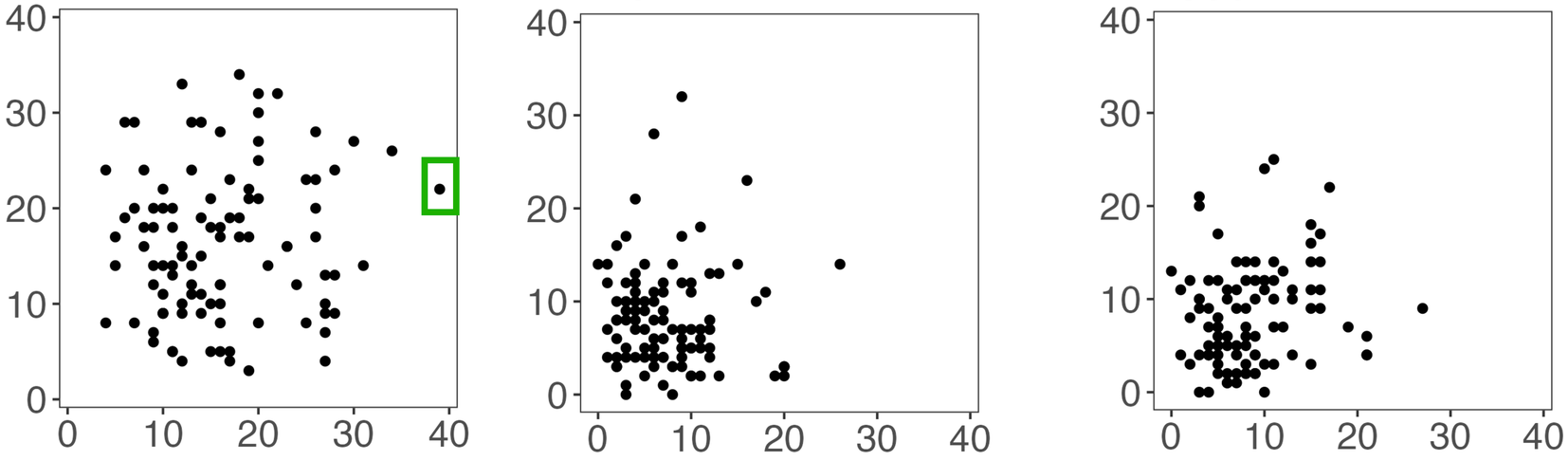


Step 1: thin observations into train/test.

Thinning avoids the pitfall of sample splitting on our motivating examples

$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, 0.5)$

$X_{ij}^{(2)} = X_{ij} - X_{ij}^{(1)}$

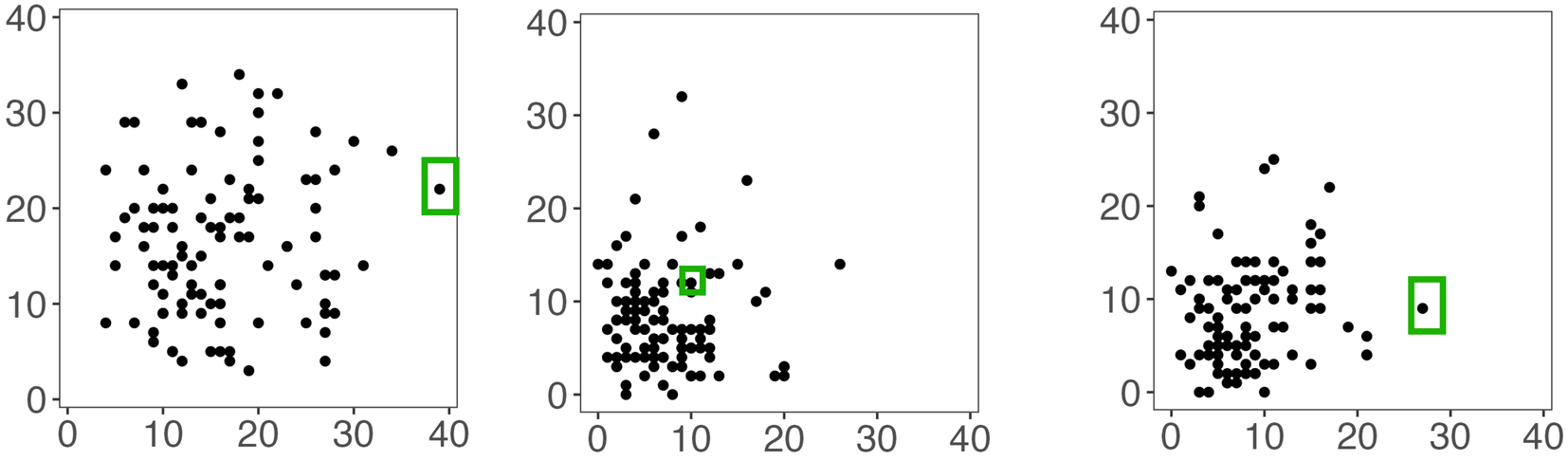


Step 1: thin observations into train/test.

Thinning avoids the pitfall of sample splitting on our motivating examples

$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, 0.5)$

$X_{ij}^{(2)} = X_{ij} - X_{ij}^{(1)}$

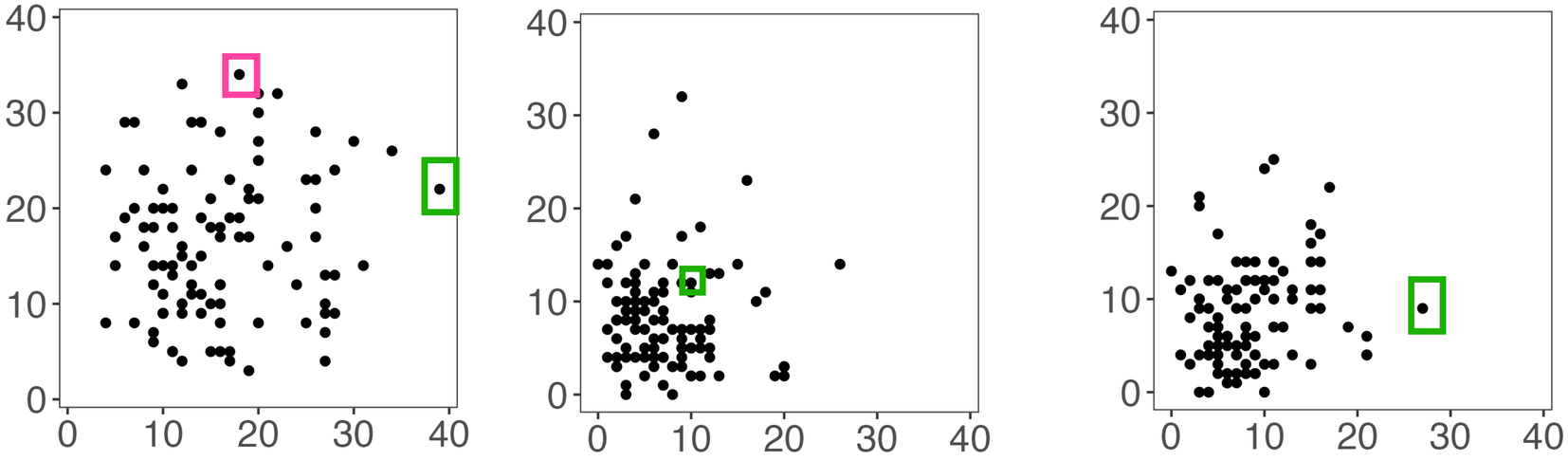


Step 1: thin observations into train/test.

Thinning avoids the pitfall of sample splitting on our motivating examples

$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, 0.5)$

$X_{ij}^{(2)} = X_{ij} - X_{ij}^{(1)}$

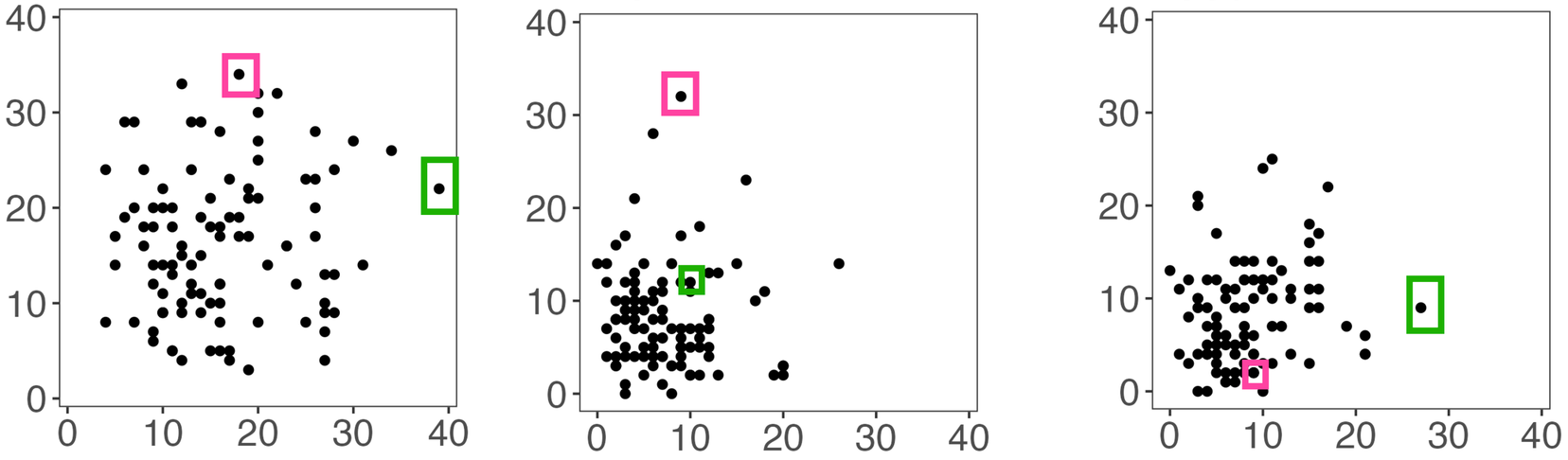


Step 1: thin observations into train/test.

Thinning avoids the pitfall of sample splitting on our motivating examples

$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, 0.5)$

$X_{ij}^{(2)} = X_{ij} - X_{ij}^{(1)}$

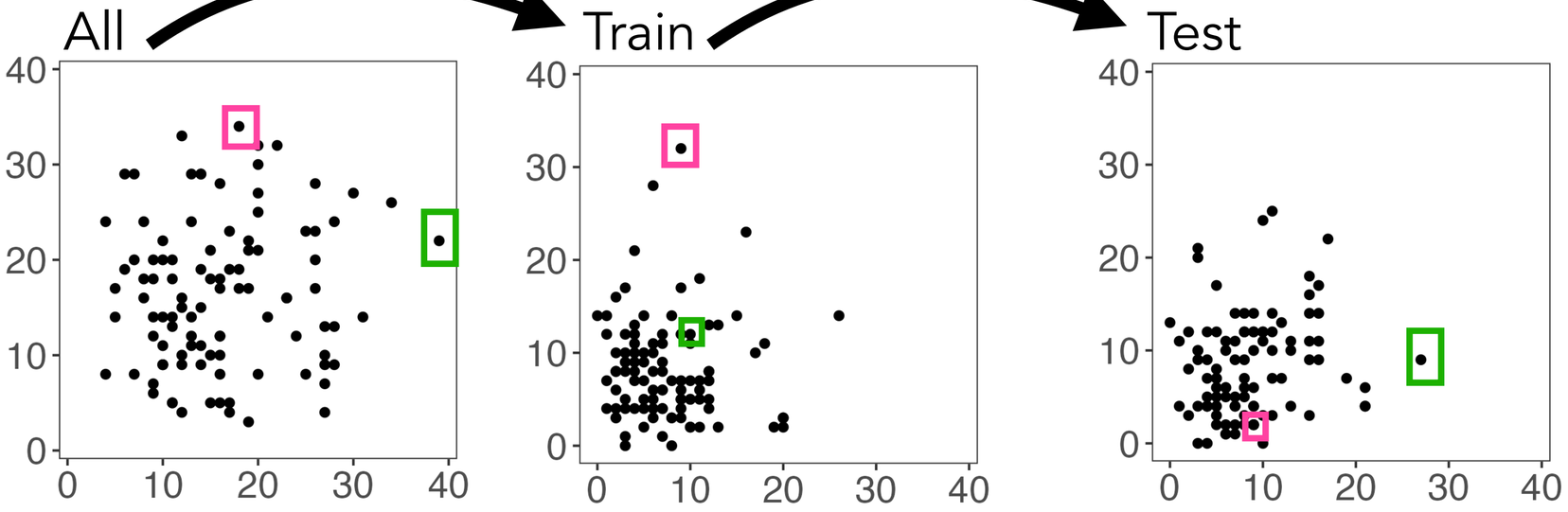


Step 1: thin observations into train/test.

Thinning avoids the pitfall of sample splitting on our motivating examples

$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, 0.5)$

$X_{ij}^{(2)} = X_{ij} - X_{ij}^{(1)}$



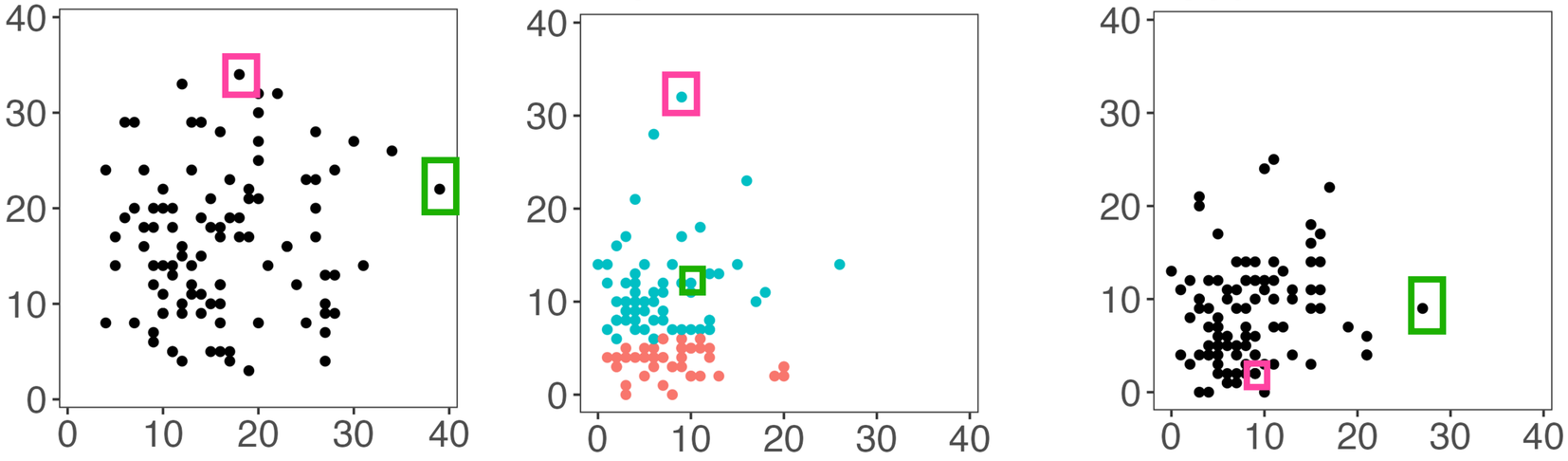
Step 1: thin observations into train/test.

Step 2: cluster the training set.

Thinning avoids the pitfall of sample splitting on our motivating examples

$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, 0.5)$

$X_{ij}^{(2)} = X_{ij} - X_{ij}^{(1)}$



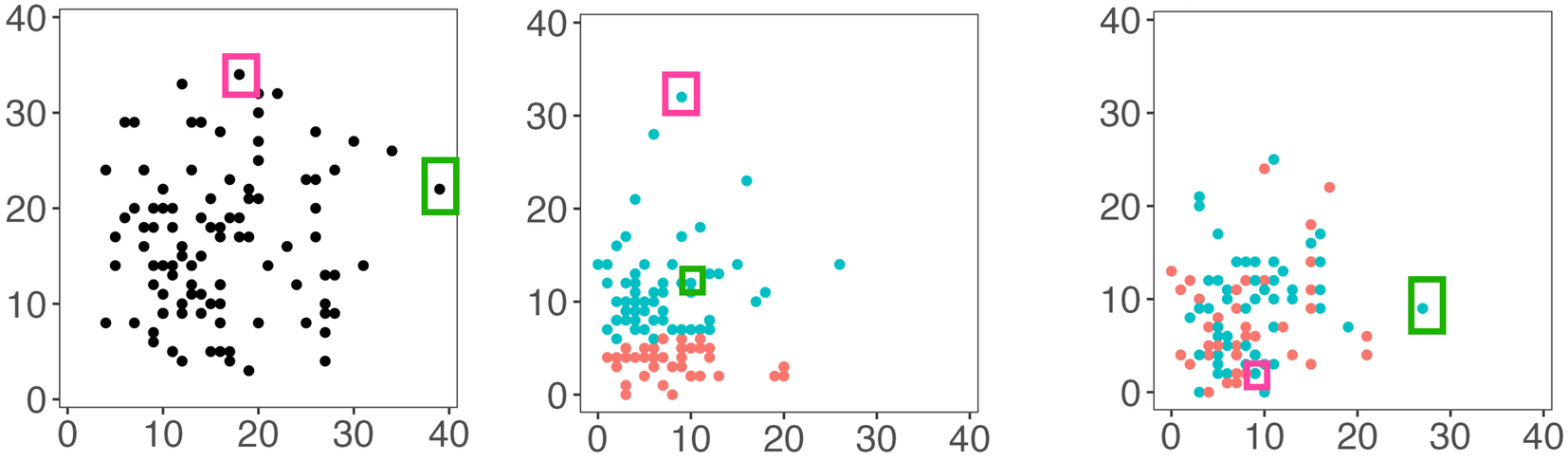
Step 1: thin observations into train/test.

Step 2: cluster the training set.

Thinning avoids the pitfall of sample splitting on our motivating examples

$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, 0.5)$

$X_{ij}^{(2)} = X_{ij} - X_{ij}^{(1)}$



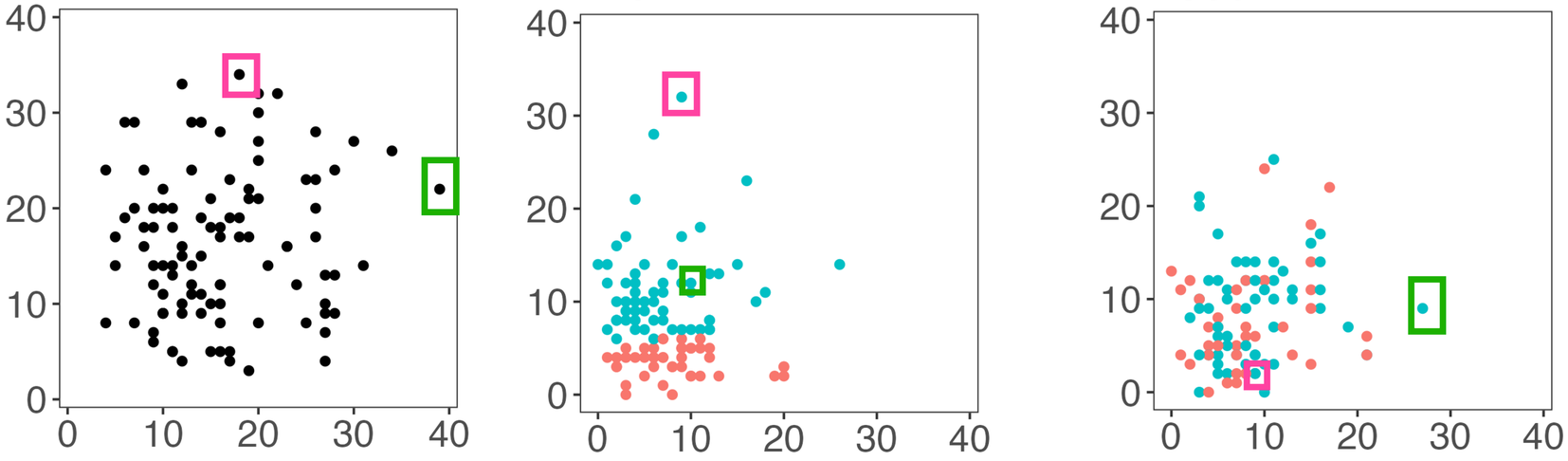
Step 1: thin observations into train/test.

Step 2: cluster the training set.

Thinning avoids the pitfall of sample splitting on our motivating examples

$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, 0.5)$

$X_{ij}^{(2)} = X_{ij} - X_{ij}^{(1)}$



Step 1: thin observations into train/test.

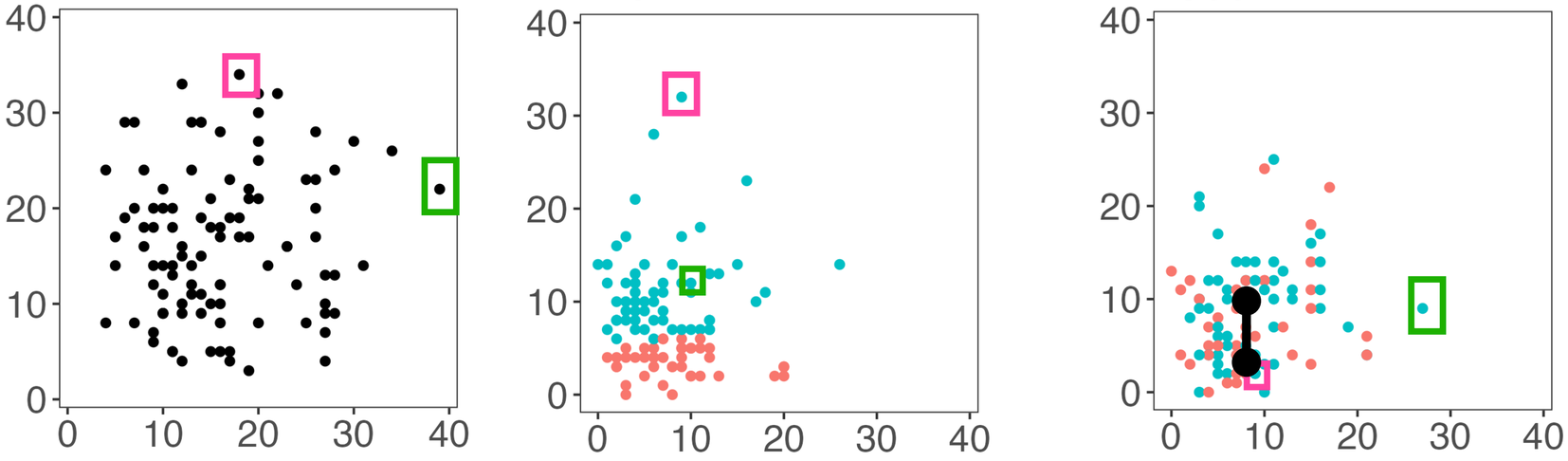
Step 2: cluster the training set.

Step 3: evaluate clusters or test for difference in means on test set.

Thinning avoids the pitfall of sample splitting on our motivating examples

$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, 0.5)$

$X_{ij}^{(2)} = X_{ij} - X_{ij}^{(1)}$



Step 1: thin observations into train/test.

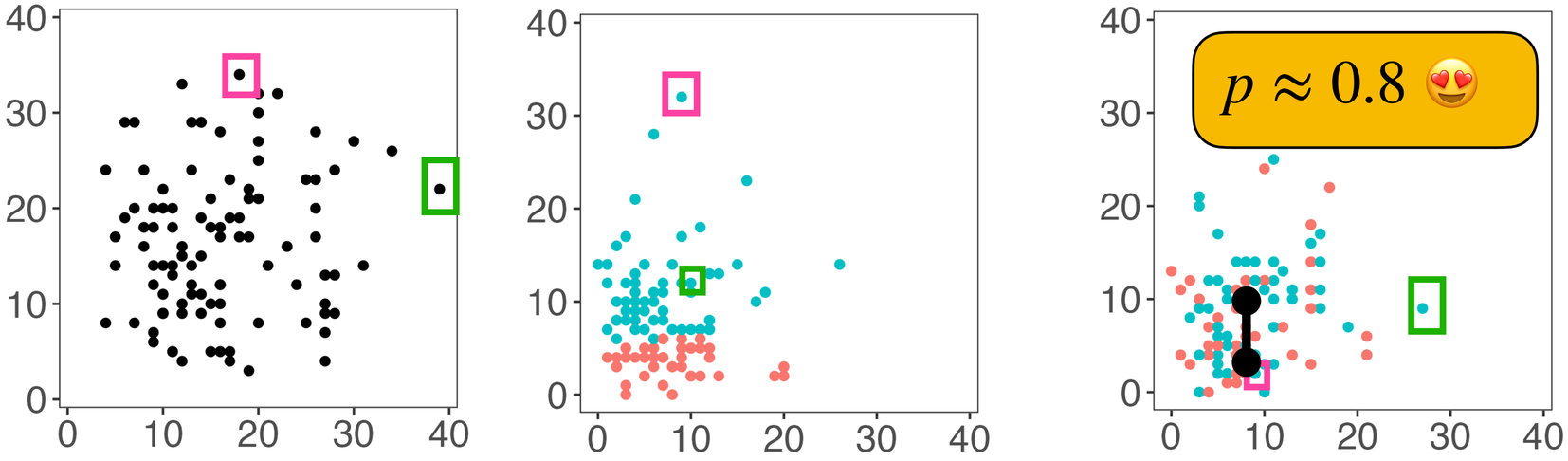
Step 2: cluster the training set.

Step 3: evaluate clusters or test for difference in means on test set.

Thinning avoids the pitfall of sample splitting on our motivating examples

$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, 0.5)$

$X_{ij}^{(2)} = X_{ij} - X_{ij}^{(1)}$



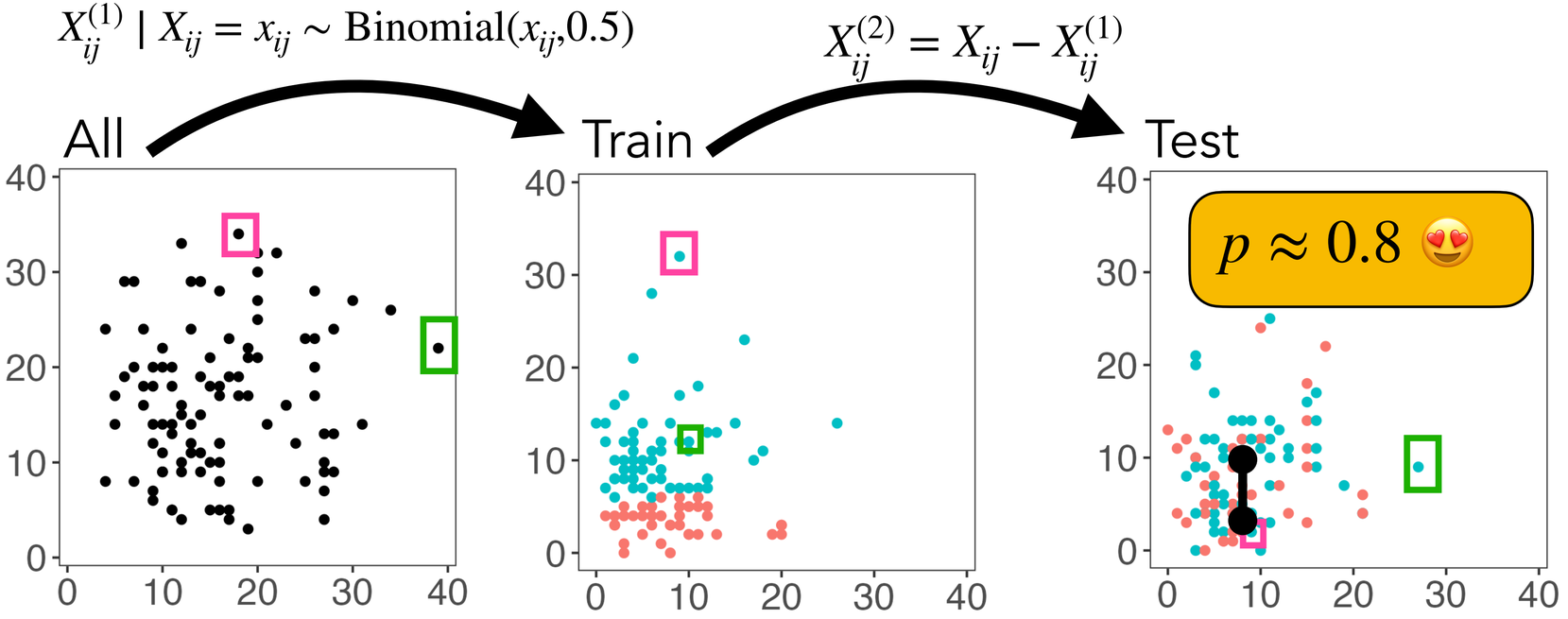
Step 1: thin observations into train/test.

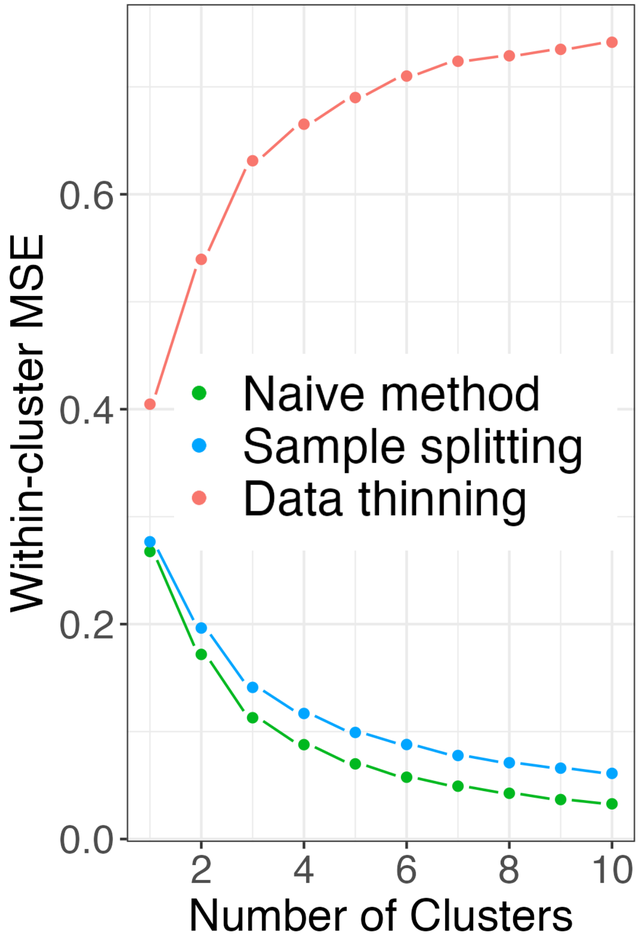
Step 2: cluster the training set.

Step 3: evaluate clusters or test for difference in means on test set.

Thinning avoids the pitfall of sample splitting on our motivating examples

$$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, 0.5) \qquad X_{ij}^{(2)} = X_{ij} - X_{ij}^{(1)}$$



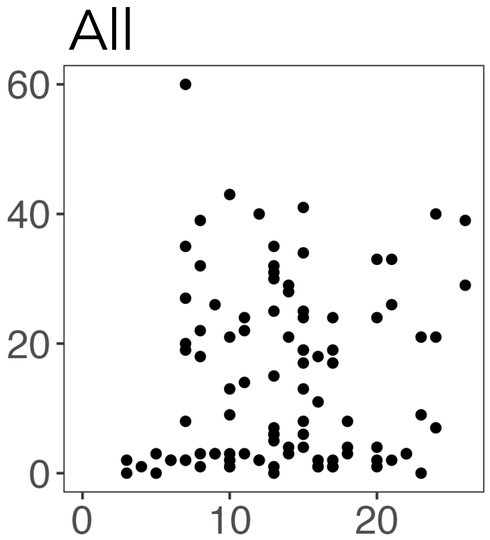


Step 1: thin observations into train/test.

Step 2: cluster the training set.

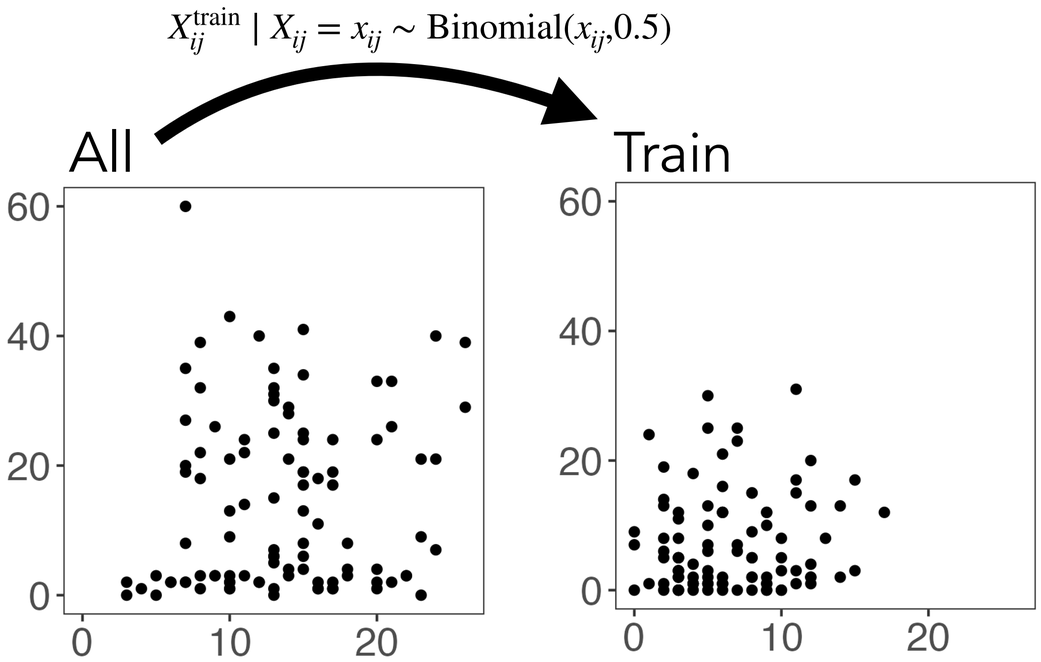
Step 3: evaluate clusters or test for difference in means on test set.

Thinning avoids the pitfall of sample splitting on our motivating examples



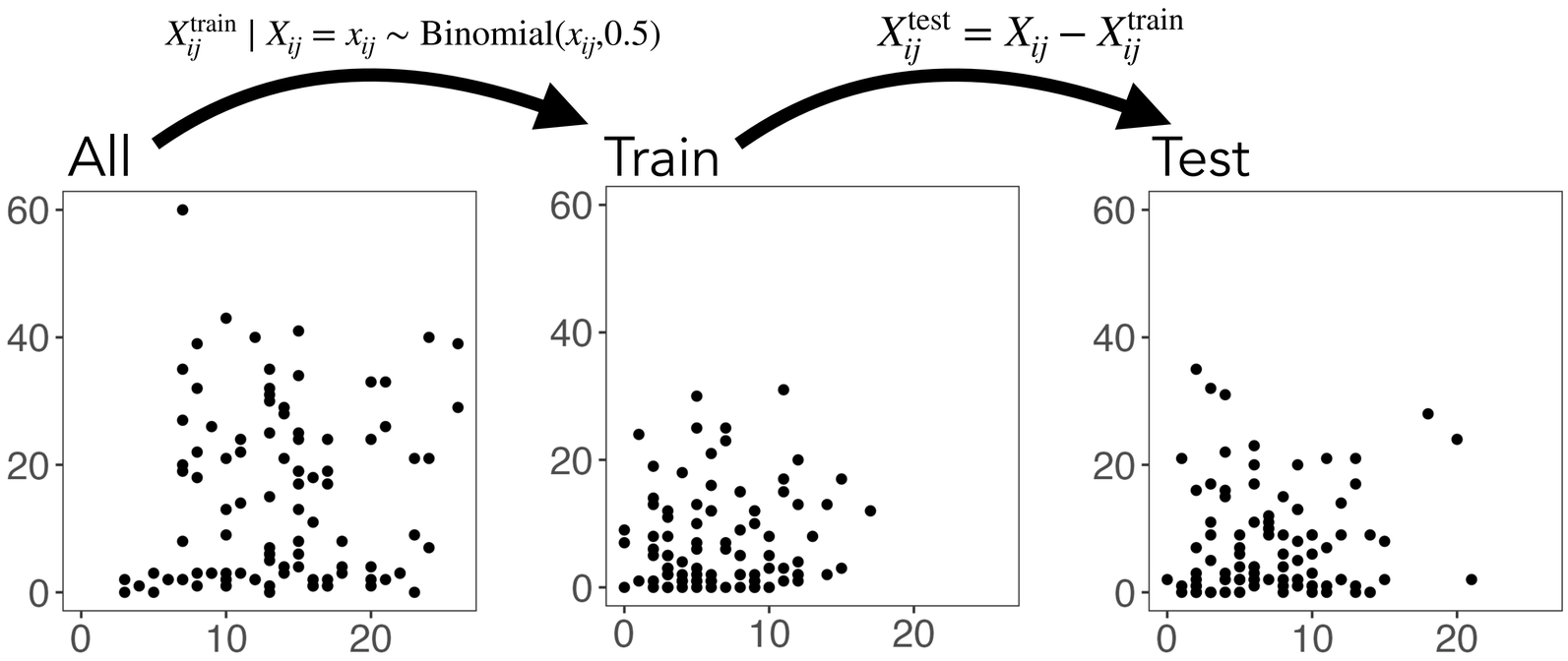
$$X_{i2} \sim \begin{cases} \text{Poisson}(3) & \text{if } i \leq 50 \\ \text{Poisson}(25) & \text{if } i > 50 \end{cases}$$

Thinning avoids the pitfall of sample splitting on our motivating examples



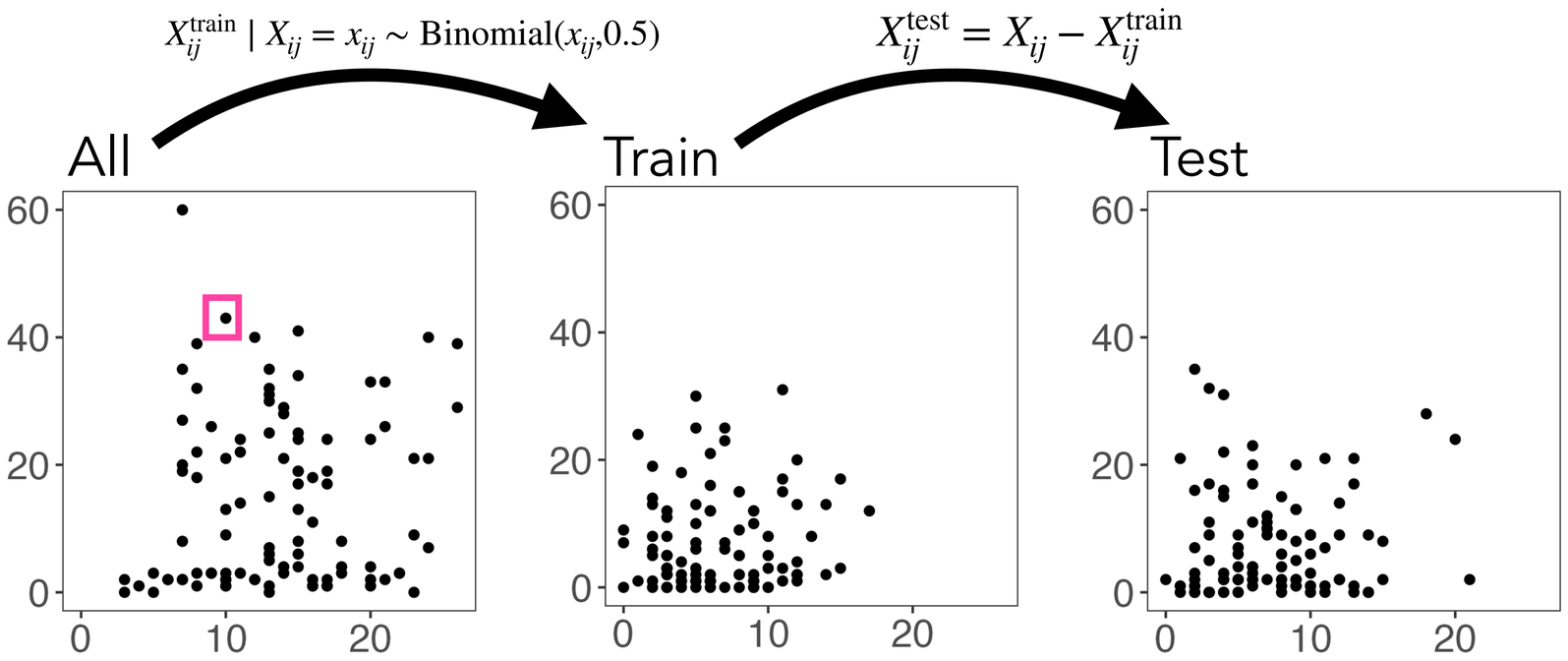
$$X_{i2} \sim \begin{cases} \text{Poisson}(3) & \text{if } i \leq 50 \\ \text{Poisson}(25) & \text{if } i > 50 \end{cases}$$

Thinning avoids the pitfall of sample splitting on our motivating examples



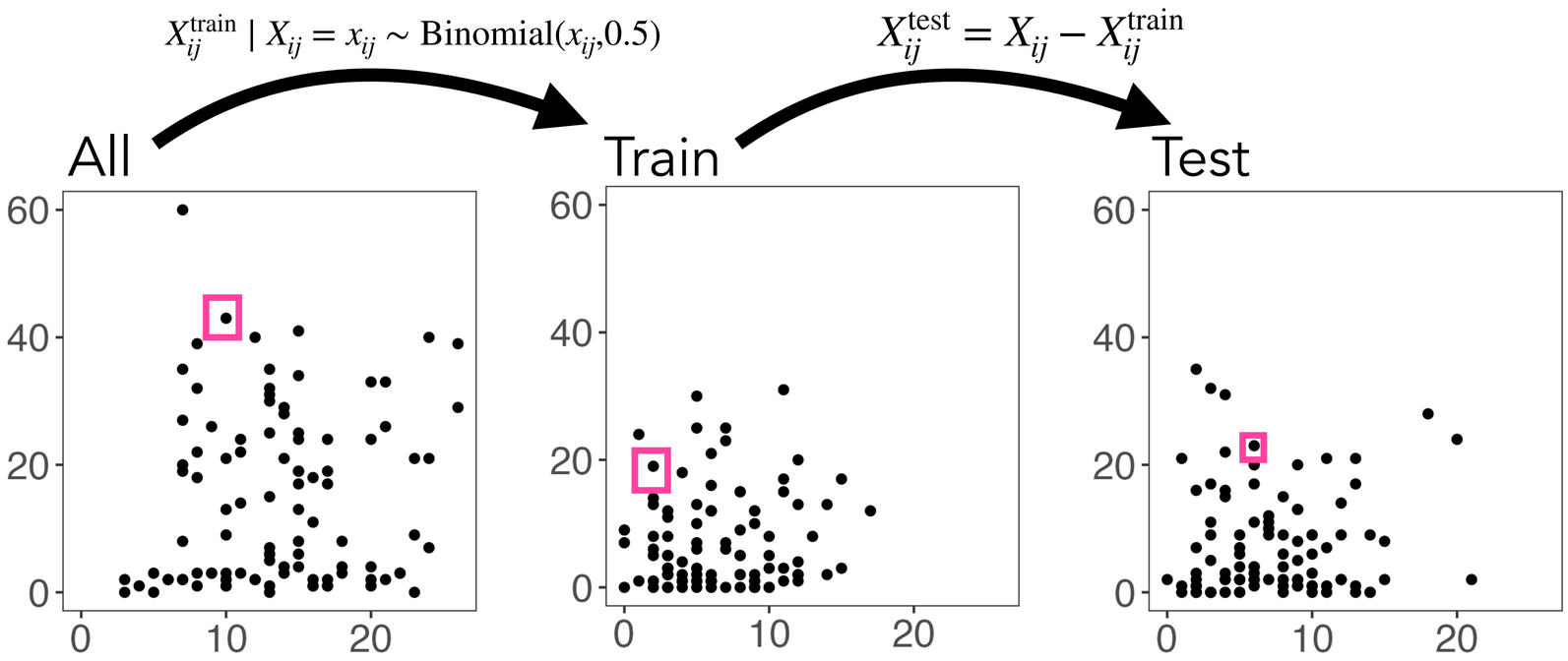
$$X_{i2} \sim \begin{cases} \text{Poisson}(3) & \text{if } i \leq 50 \\ \text{Poisson}(25) & \text{if } i > 50 \end{cases}$$

Thinning avoids the pitfall of sample splitting on our motivating examples



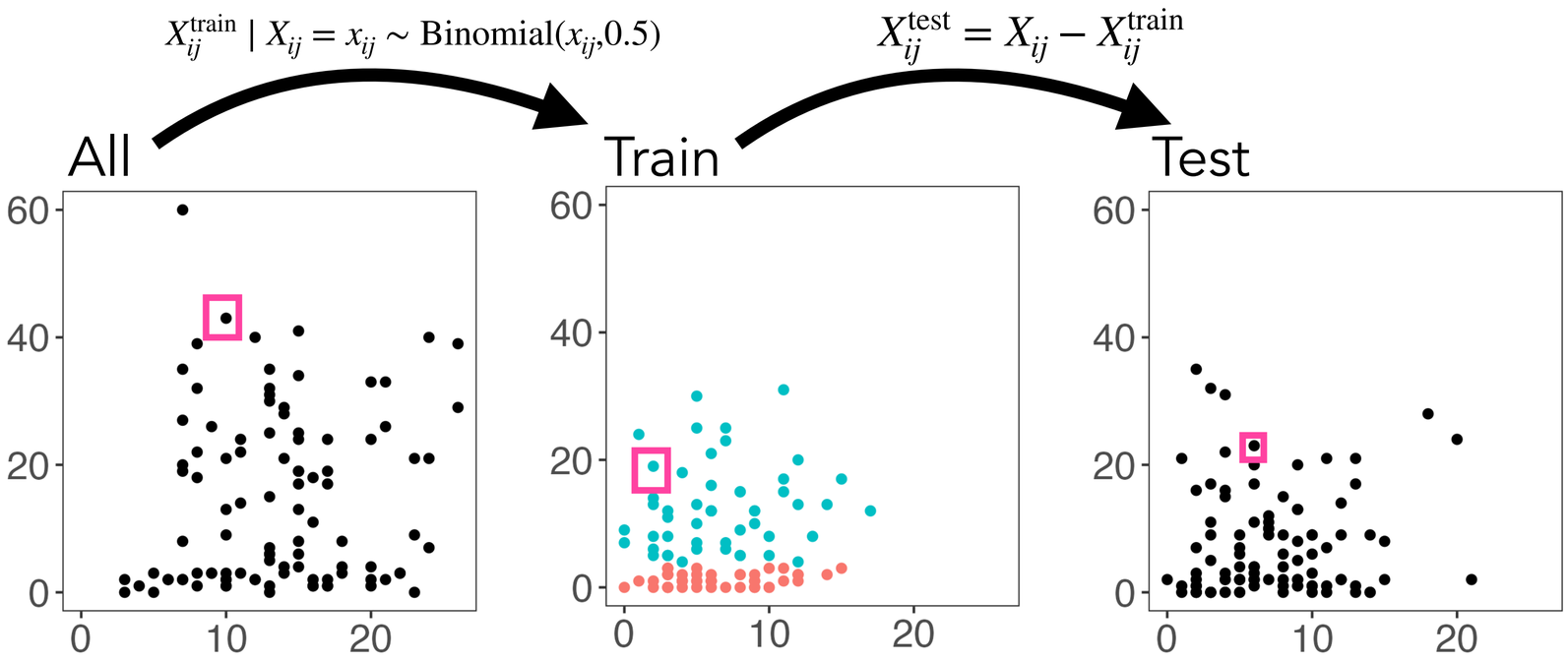
$$X_{i2} \sim \begin{cases} \text{Poisson}(3) & \text{if } i \leq 50 \\ \text{Poisson}(25) & \text{if } i > 50 \end{cases}$$

Thinning avoids the pitfall of sample splitting on our motivating examples



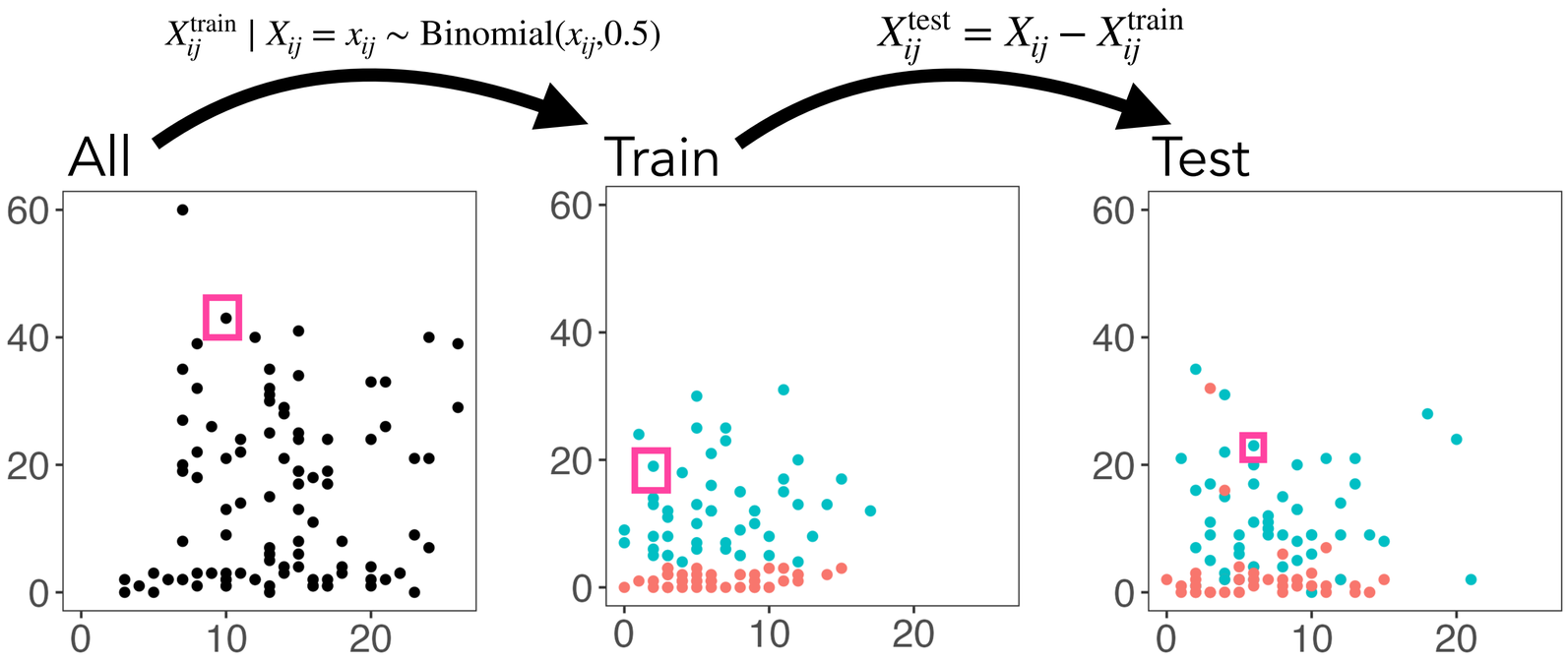
$$X_{i2} \sim \begin{cases} \text{Poisson}(3) & \text{if } i \leq 50 \\ \text{Poisson}(25) & \text{if } i > 50 \end{cases}$$

Thinning avoids the pitfall of sample splitting on our motivating examples



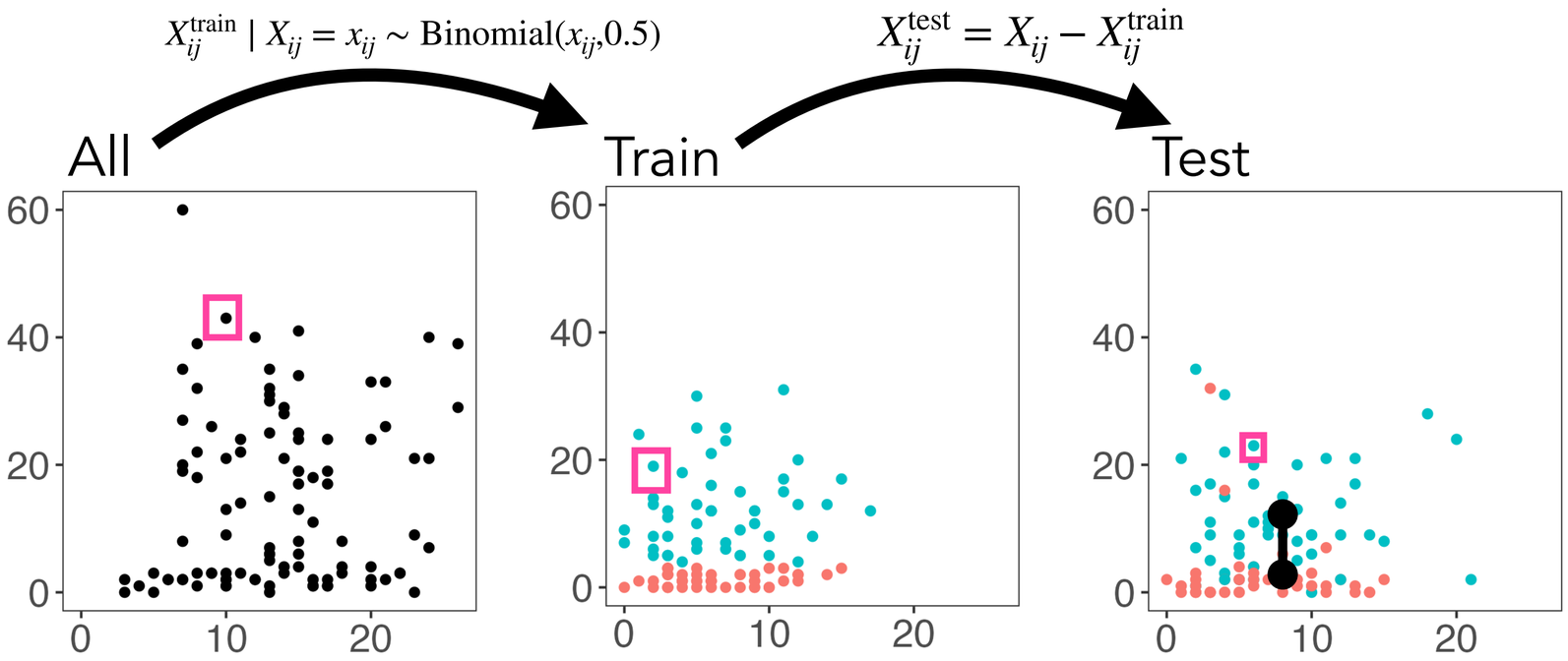
$$X_{i2} \sim \begin{cases} \text{Poisson}(3) & \text{if } i \leq 50 \\ \text{Poisson}(25) & \text{if } i > 50 \end{cases}$$

Thinning avoids the pitfall of sample splitting on our motivating examples



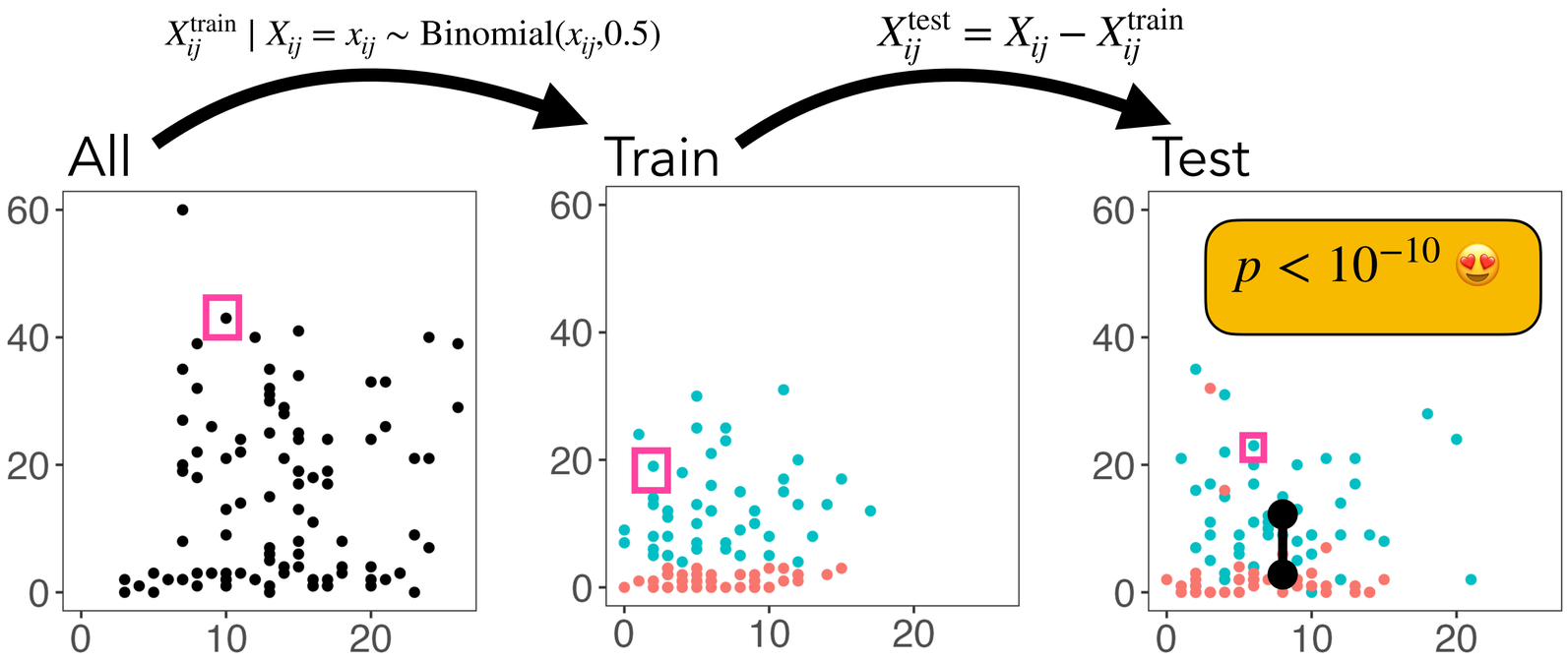
$$X_{i2} \sim \begin{cases} \text{Poisson}(3) & \text{if } i \leq 50 \\ \text{Poisson}(25) & \text{if } i > 50 \end{cases}$$

Thinning avoids the pitfall of sample splitting on our motivating examples



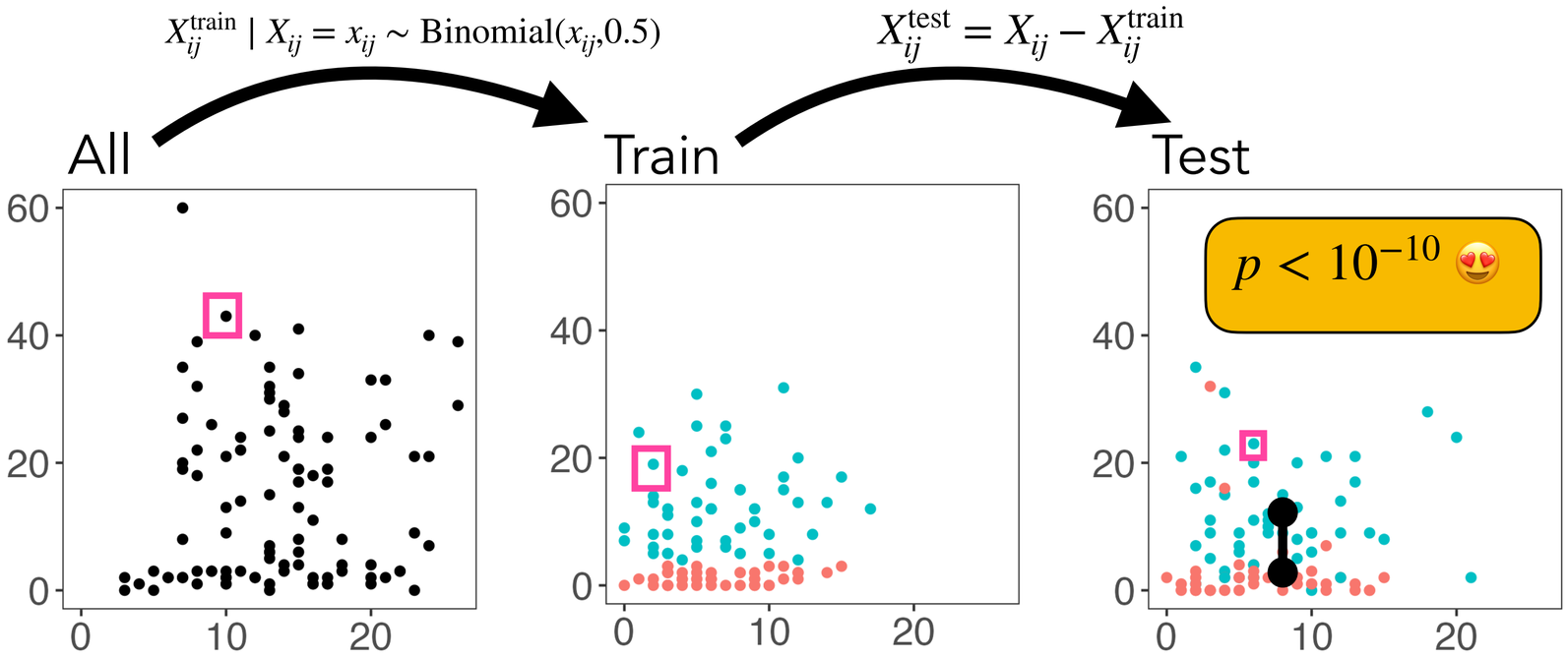
$$X_{i2} \sim \begin{cases} \text{Poisson}(3) & \text{if } i \leq 50 \\ \text{Poisson}(25) & \text{if } i > 50 \end{cases}$$

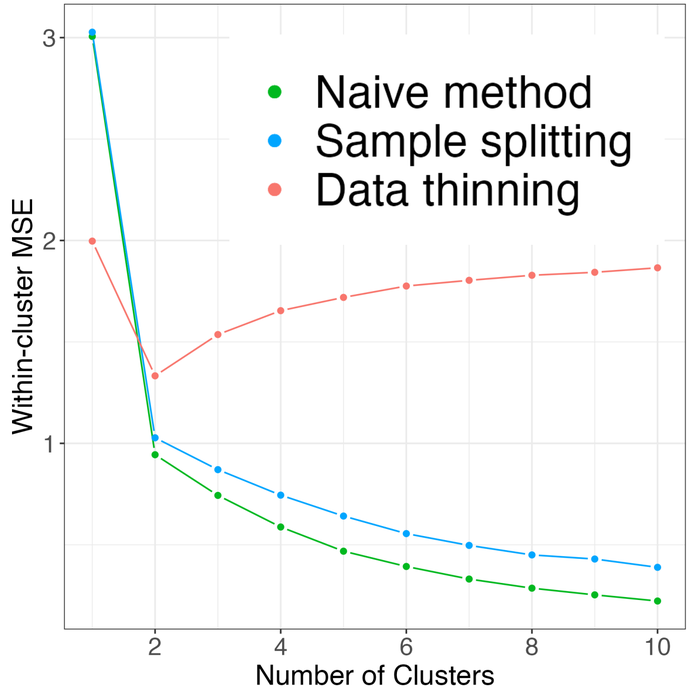
Thinning avoids the pitfall of sample splitting on our motivating examples



$$X_{i2} \sim \begin{cases} \text{Poisson}(3) & \text{if } i \leq 50 \\ \text{Poisson}(25) & \text{if } i > 50 \end{cases}$$

Thinning avoids the pitfall of sample splitting on our motivating examples





$$X_{i2} \sim \begin{cases} \text{Poisson}(3) & \text{if } i \leq 50 \\ \text{Poisson}(25) & \text{if } i > 50 \end{cases}$$

Outline

1. Motivation: sample splitting doesn’t always work
2. Poisson thinning
3. **Data thinning**
4. Generalized data thinning
5. Application to changepoint validation
6. Ongoing work

What did we like about Poisson thinning?

We split a single observation X into $X^{(1)}$ and $X^{(2)}$ such that:

(1) $X^{(1)}$ and $X^{(2)}$ have the same distribution as X, up to a parameter scaling.

(2) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

What did we like about Poisson thinning?

We split a single observation X into $X^{(1)}$ and $X^{(2)}$ such that:

(1) $X^{(1)}$ and $X^{(2)}$ have the same distribution as X, up to a parameter scaling.

(2) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

Can we achieve these same properties when X is not Poisson?

Data thinning

Goal: split a single observation X into $X^{(1)}$ and $X^{(2)}$ such that:

(1) $X^{(1)}$ and $X^{(2)}$ have the same distribution as X, up to a parameter scaling.

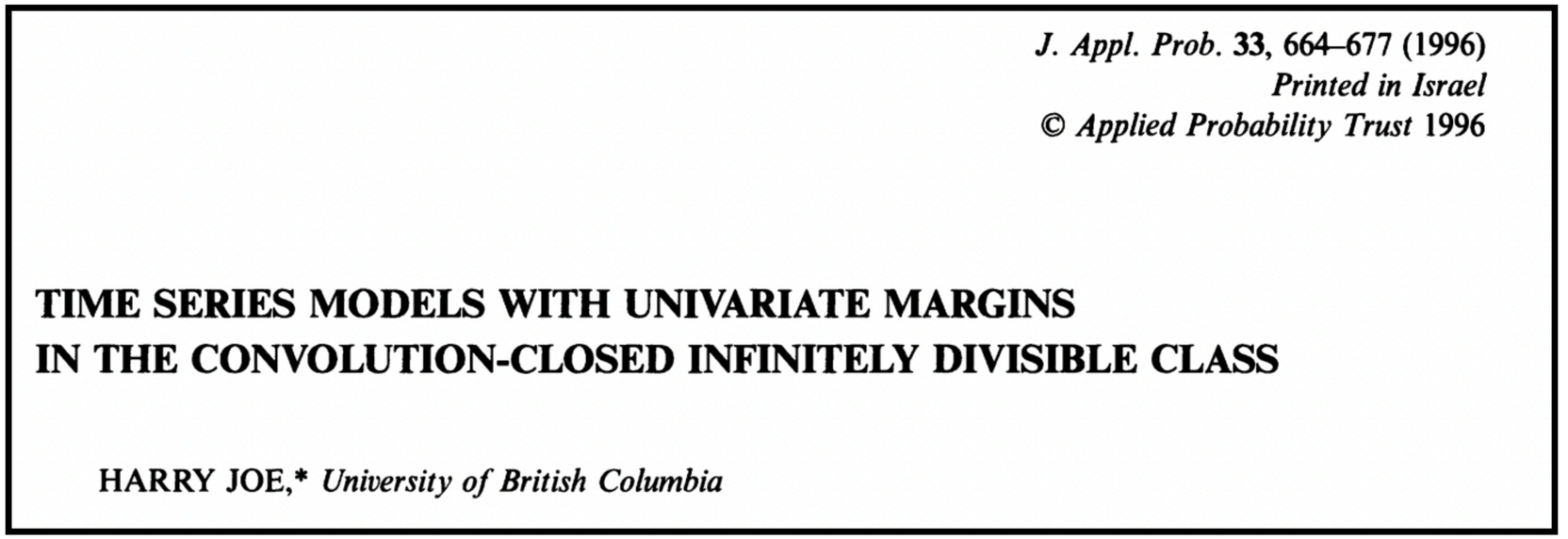
(2) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

Data thinning

Goal: split a single observation X into $X^{(1)}$ and $X^{(2)}$ such that:

(1) $X^{(1)}$ and $X^{(2)}$ have the same distribution as X, up to a parameter scaling.

(2) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.



J. Appl. Prob. **33**, 664–677 (1996)
Printed in Israel

TIME SERIES MODELS WITH UNIVARIATE MARGINS IN THE CONVOLUTION-CLOSED INFINITELY DIVISIBLE CLASS

HARRY JOE,* *University of British Columbia*

Convolution-closed distributions

A family of distributions F_λ is "convolution-closed" in parameter λ if

- $X' \sim F_{\lambda_1}$,
- $X'' \sim F_{\lambda_2}$,
- $X' \perp\!\!\!\perp X''$

together imply that $X' + X'' \sim F_{\lambda_1+\lambda_2}$.

Convolution-closed distributions

A family of distributions F_λ is "convolution-closed" in parameter λ if

- $X' \sim F_{\lambda_1}$,
- $X'' \sim F_{\lambda_2}$,
- $X' \perp\!\!\!\perp X''$

together imply that $X' + X'' \sim F_{\lambda_1+\lambda_2}$.

Distribution	Convolution-closed in:
$X \sim \text{Poisson}(\lambda)$	λ
$X \sim \text{N}(\mu, \sigma^2)$	(μ, σ^2)
$X \sim \text{NegativeBinomial}(\mu, b)$	(μ, b)
$X \sim \text{Gamma}(\alpha, \beta)$	α, if β is fixed
$X \sim \text{Binomial}(r, p)$	r, if p is fixed
$X \sim \text{N}_k(\mu, \Sigma)$.	(μ, Σ).
$X \sim \text{Multinomial}_k(r, p)$	r, if p is fixed
$X \sim \text{Wishart}_p(n, \Sigma)$	n, if p and Σ are fixed.

Convolution-closed distributions

A family of distributions F_λ is "convolution-closed" in parameter λ if

- $X' \sim F_{\lambda_1'}$
- $X'' \sim F_{\lambda_2'}$
- $X' \perp\!\!\!\perp X''$

together imply that $X' + X'' \sim F_{\lambda_1+\lambda_2}$.

Distribution	Convolution-closed in:
$X \sim \text{Poisson}(\lambda)$	λ
$X \sim \text{N}(\mu, \sigma^2)$	(μ, σ^2)
$X \sim \text{NegativeBinomial}(\mu, b)$	(μ, b)
$X \sim \text{Gamma}(\alpha, \beta)$	α, if β is fixed
$X \sim \text{Binomial}(r, p)$	r, if p is fixed
$X \sim \text{N}_k(\mu, \Sigma)$.	(μ, Σ).
$X \sim \text{Multinomial}_k(r, p)$	r, if p is fixed
$X \sim \text{Wishart}_p(n, \Sigma)$	n, if p and Σ are fixed.

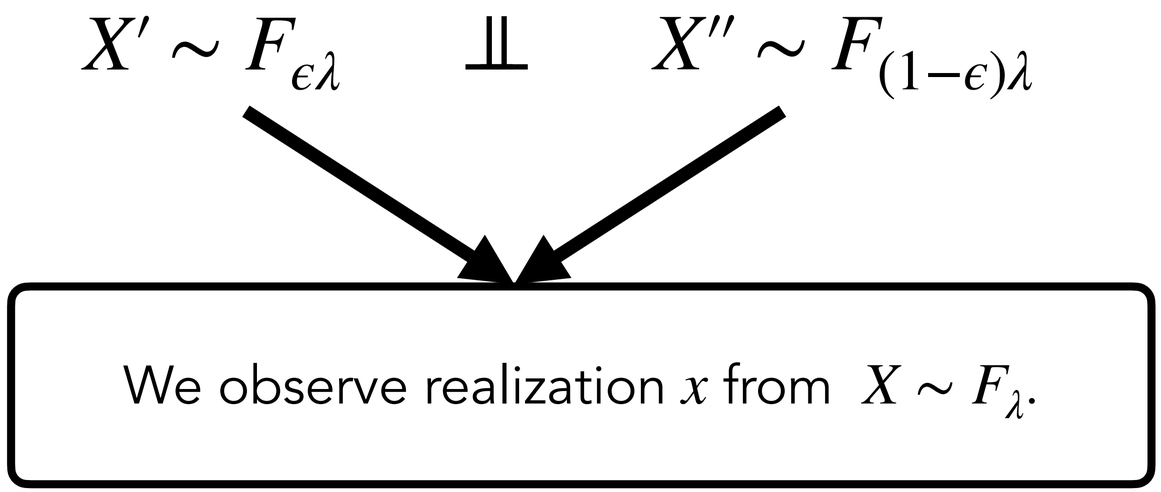
Jorgensen and Song (1998, Journal of Applied Probability) unify notation for any **exponential dispersion family.**

Data thinning for convolution-closed distributions

We observe realization x from $X \sim F_\lambda$.

Data thinning for convolution-closed distributions

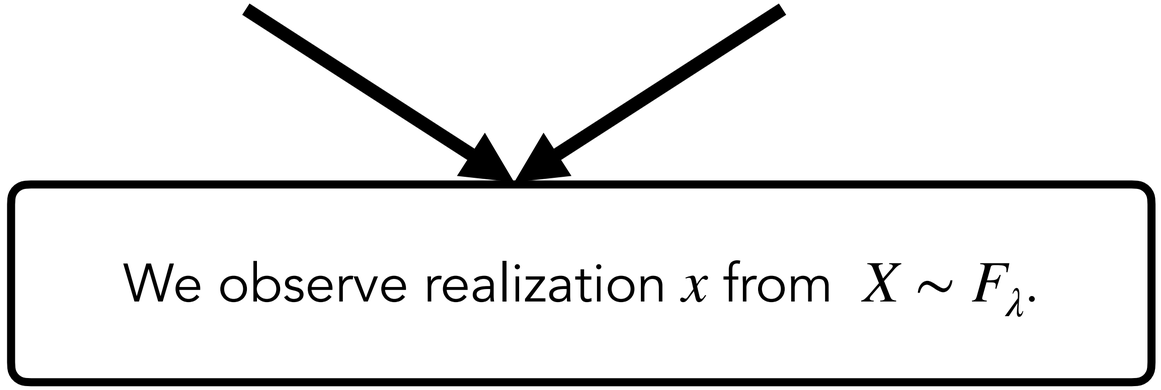
We know x could have arisen as $x' + x''$, where



Data thinning for convolution-closed distributions

We know x could have arisen as $x' + x''$, where

$$X' \sim F_{\epsilon\lambda} \quad \perp\!\!\!\perp \quad X'' \sim F_{(1-\epsilon)\lambda}$$



We observe realization x from $X \sim F_\lambda$.

If we had observed x' and x'', we would have satisfied our goal of data thinning!

Data thinning for convolution-closed distributions

We know x could have arisen as $x' + x''$, where

$$X' \sim F_{\epsilon\lambda} \quad \perp\!\!\!\perp \quad X'' \sim F_{(1-\epsilon)\lambda}$$

We observe realization x from $X \sim F_\lambda$.

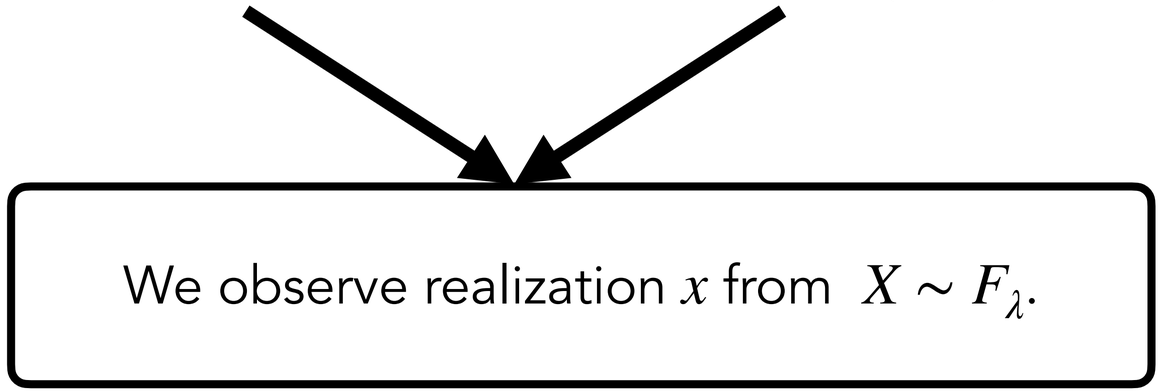
If we had observed x' and x'', we would have satisfied our goal of data thinning!

Can we work backwards to recover x' and x''?

Data thinning for convolution-closed distributions

We know x could have arisen as $x' + x''$, where

$$X' \sim F_{\epsilon\lambda} \quad \perp\!\!\!\perp \quad X'' \sim F_{(1-\epsilon)\lambda}$$



We observe realization x from $X \sim F_\lambda$.

If we had observed x' and x'', we would have satisfied our goal of data thinning!

Can we work backwards to recover x' and x''?

Let $G_{\epsilon,x}$ be the conditional distribution of $X' \mid X = x$.

Data thinning for convolution-closed distributions

We know x could have arisen as $x' + x''$, where

$$X' \sim F_{\epsilon\lambda} \quad \perp\!\!\!\perp \quad X'' \sim F_{(1-\epsilon)\lambda}$$

We observe realization x from $X \sim F_\lambda$.

Algorithm:

Draw $X^{(1)} \mid X = x$ from $G_{\epsilon,x}$. Let $X^{(2)} = X - X^{(1)}$.

If we had observed x' and x'', we would have satisfied our goal of data thinning!

Can we work backwards to recover x' and x''?

Let $G_{\epsilon,x}$ be the conditional distribution of $X' \mid X = x$.

Data thinning for convolution-closed distributions

We know x could have arisen as $x' + x''$, where

$$X' \sim F_{\epsilon\lambda} \quad \perp\!\!\!\perp \quad X'' \sim F_{(1-\epsilon)\lambda}$$

We observe realization x from $X \sim F_\lambda$.

Algorithm:

Draw $X^{(1)} \mid X = x$ from $G_{\epsilon,x}$. Let $X^{(2)} = X - X^{(1)}$.

Theorem:

$X^{(1)} \sim F_{\epsilon\lambda}$, $X^{(2)} \sim F_{(1-\epsilon)\lambda}$, $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

If we had observed x' and x'', we would have satisfied our goal of data thinning!

Can we work backwards to recover x' and x''?

Let $G_{\epsilon,x}$ be the conditional distribution of $X' \mid X = x$.

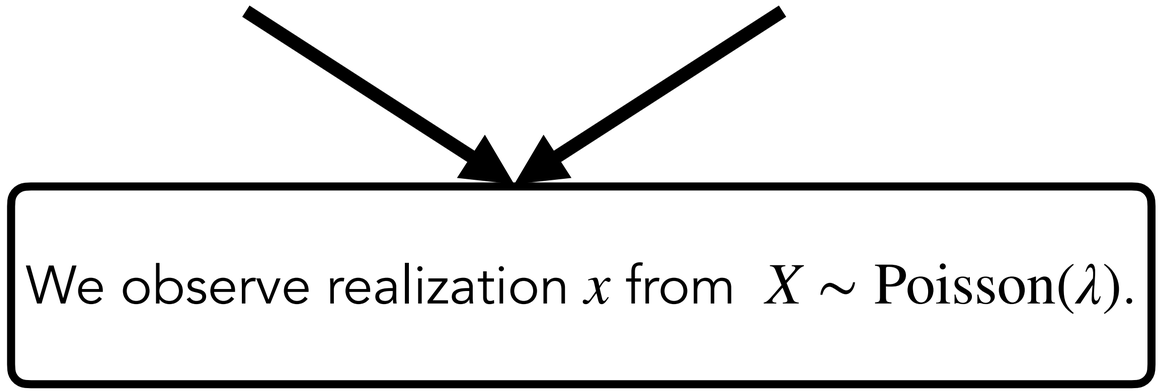
Data thinning recipe for the Poisson distribution

We observe realization x from $X \sim \text{Poisson}(\lambda)$.

Data thinning recipe for the Poisson distribution

We know x could have arisen as $x' + x''$, where

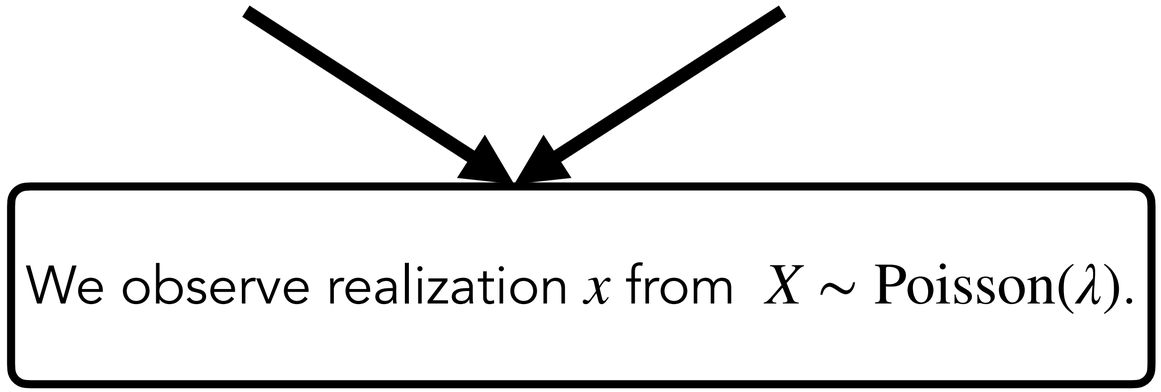
$$X' \sim \text{Poisson}(\epsilon\lambda) \quad \perp\!\!\!\perp \quad X'' \sim \text{Poisson}((1-\epsilon)\lambda)$$



Data thinning recipe for the Poisson distribution

We know x could have arisen as $x' + x''$, where

$$X' \sim \text{Poisson}(\epsilon\lambda) \quad \perp\!\!\!\perp \quad X'' \sim \text{Poisson}((1-\epsilon)\lambda)$$



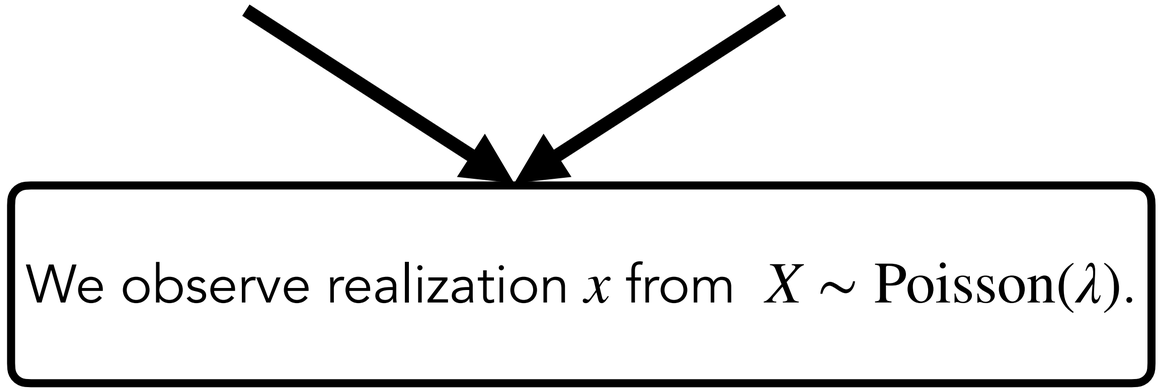
We observe realization x from $X \sim \text{Poisson}(\lambda)$.

Can we work backwards to recover x' and x''?

Data thinning recipe for the Poisson distribution

We know x could have arisen as $x' + x''$, where

$$X' \sim \text{Poisson}(\epsilon\lambda) \quad \perp\!\!\!\perp \quad X'' \sim \text{Poisson}((1-\epsilon)\lambda)$$



We observe realization x from $X \sim \text{Poisson}(\lambda)$.

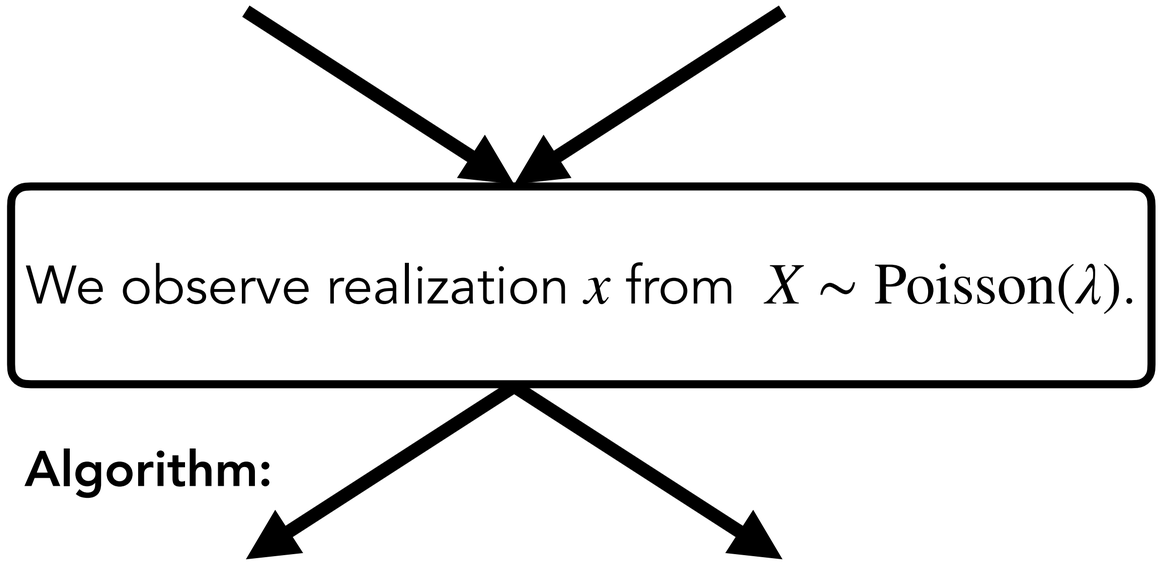
Can we work backwards to recover x' and x''?

Can show that the conditional distribution of $X' \mid X = x$ is $\text{Binomial}(x, \epsilon)$.

Data thinning recipe for the Poisson distribution

We know x could have arisen as $x' + x''$, where

$$X' \sim \text{Poisson}(\epsilon\lambda) \quad \perp\!\!\!\perp \quad X'' \sim \text{Poisson}((1-\epsilon)\lambda)$$



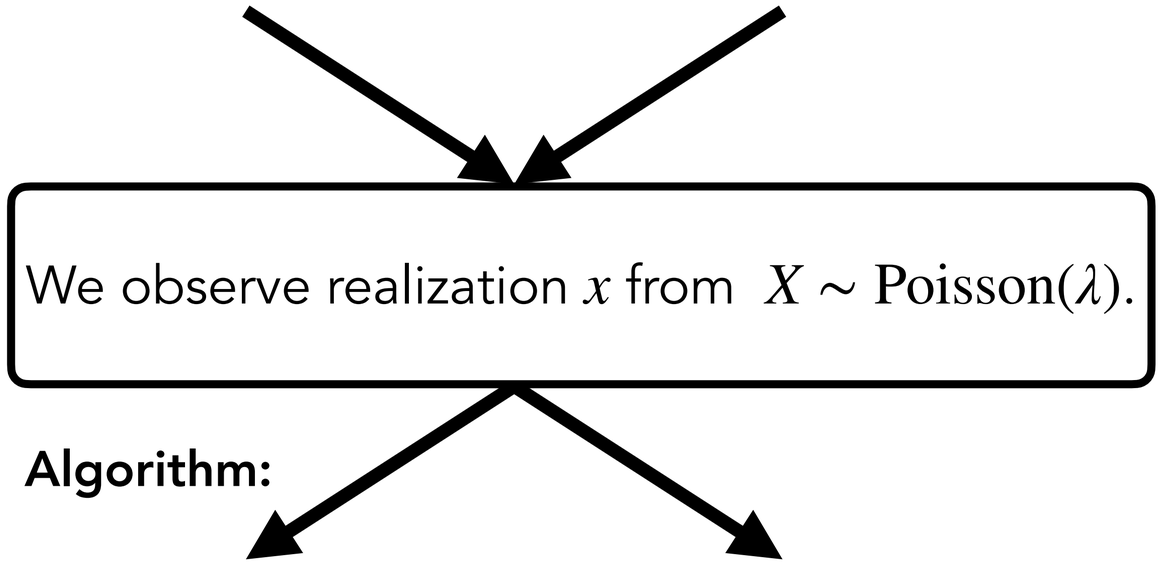
Can we work backwards to recover x' and x''?

Can show that the conditional distribution of $X' \mid X = x$ is $\text{Binomial}(x, \epsilon)$.

Data thinning recipe for the Poisson distribution

We know x could have arisen as $x' + x''$, where

$$X' \sim \text{Poisson}(\epsilon\lambda) \quad \perp\!\!\!\perp \quad X'' \sim \text{Poisson}((1-\epsilon)\lambda)$$



Can we work backwards to recover x' and x''?

Can show that the conditional distribution of $X' \mid X = x$ is $\text{Binomial}(x, \epsilon)$.

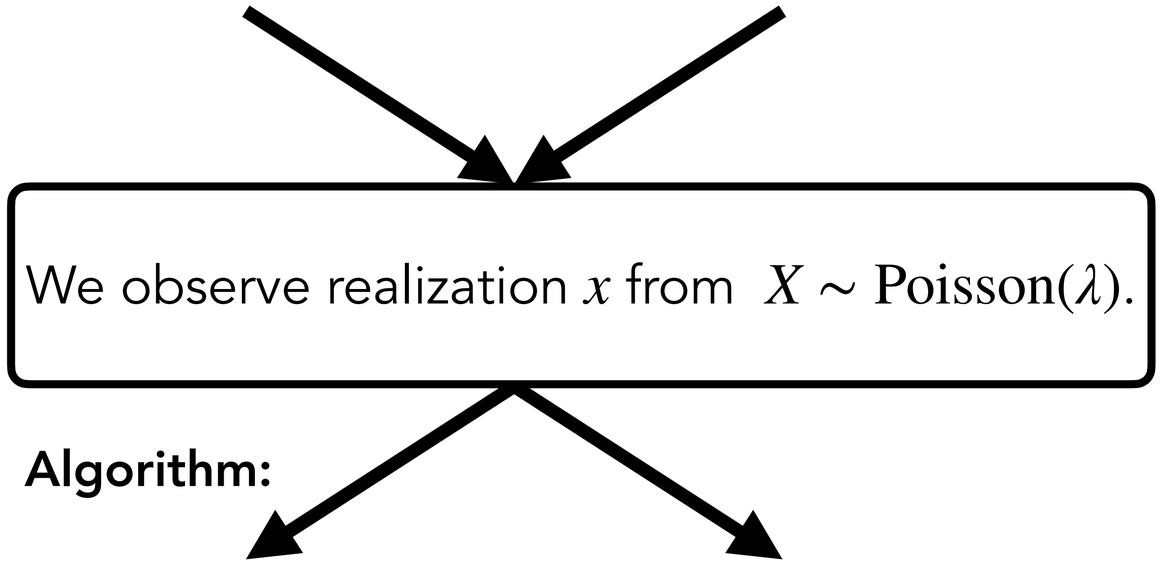
Theorem:

$X^{(1)} \sim \text{Poisson}(\epsilon\lambda)$, $X^{(2)} \sim \text{Poisson}((1-\epsilon)\lambda)$, $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

Data thinning recipe for the Poisson distribution

We know x could have arisen as $x' + x''$, where

$$X' \sim \text{Poisson}(\epsilon\lambda) \quad \perp\!\!\!\perp \quad X'' \sim \text{Poisson}((1-\epsilon)\lambda)$$



Theorem:
$X^{(1)} \sim \text{Poisson}(\epsilon\lambda)$, $X^{(2)} \sim \text{Poisson}((1-\epsilon)\lambda)$, $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

Can we work backwards to recover x' and x''?

Can show that the conditional distribution of $X' \mid X = x$ is $\text{Binomial}(x, \epsilon)$.

We recover the well-known Poisson thinning recipe.

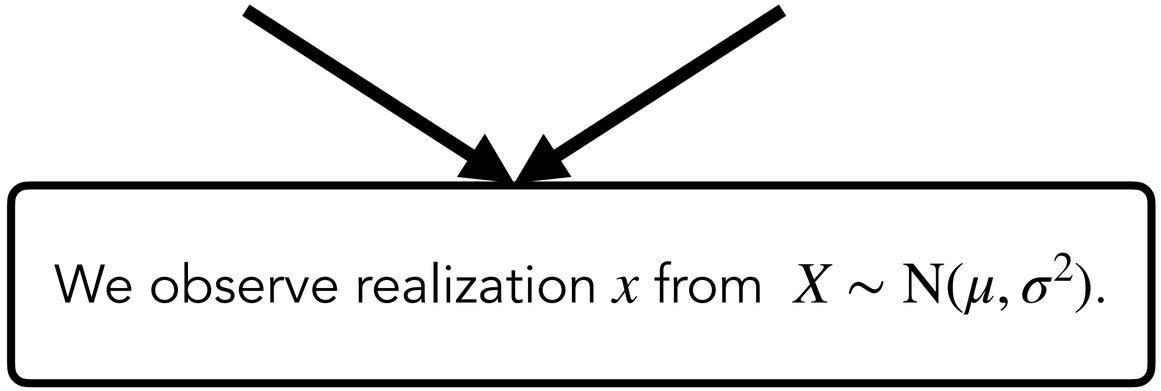
Data thinning recipe for the Gaussian distribution

We observe realization x from $X \sim \mathrm{N}(\mu, \sigma^2)$.

Data thinning recipe for the Gaussian distribution

We know x could have arisen as $x' + x''$, where

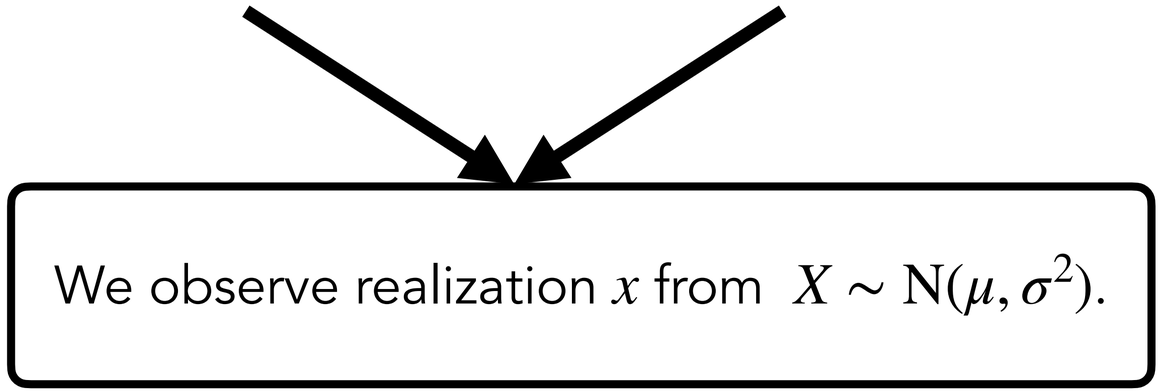
$$X' \sim \mathrm{N}(\epsilon\mu, \epsilon\sigma^2) \quad \perp\!\!\!\perp \quad X'' \sim \mathrm{N}((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$$



Data thinning recipe for the Gaussian distribution

We know x could have arisen as $x' + x''$, where

$$X' \sim \mathrm{N}(\epsilon\mu, \epsilon\sigma^2) \quad \perp\!\!\!\perp \quad X'' \sim \mathrm{N}((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$$

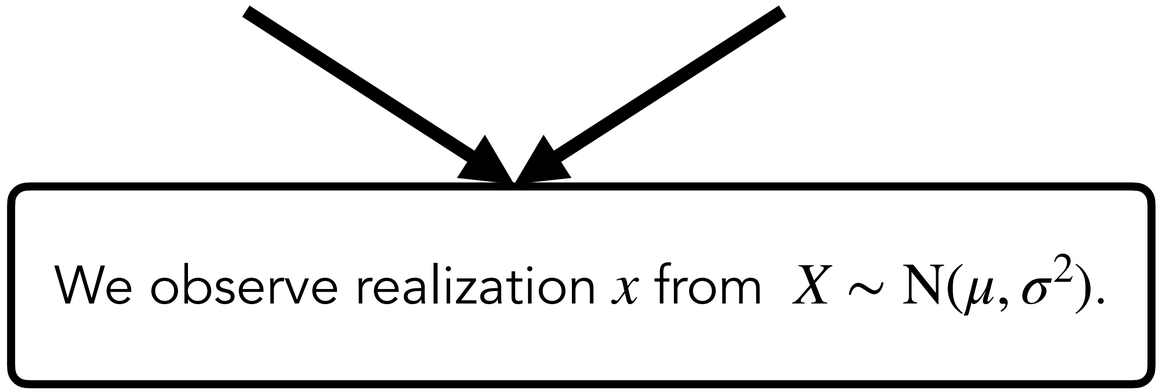


Can we work backwards to recover x' and x''?

Data thinning recipe for the Gaussian distribution

We know x could have arisen as $x' + x''$, where

$$X' \sim \mathrm{N}(\epsilon\mu, \epsilon\sigma^2) \quad \perp\!\!\!\perp \quad X'' \sim \mathrm{N}((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$$



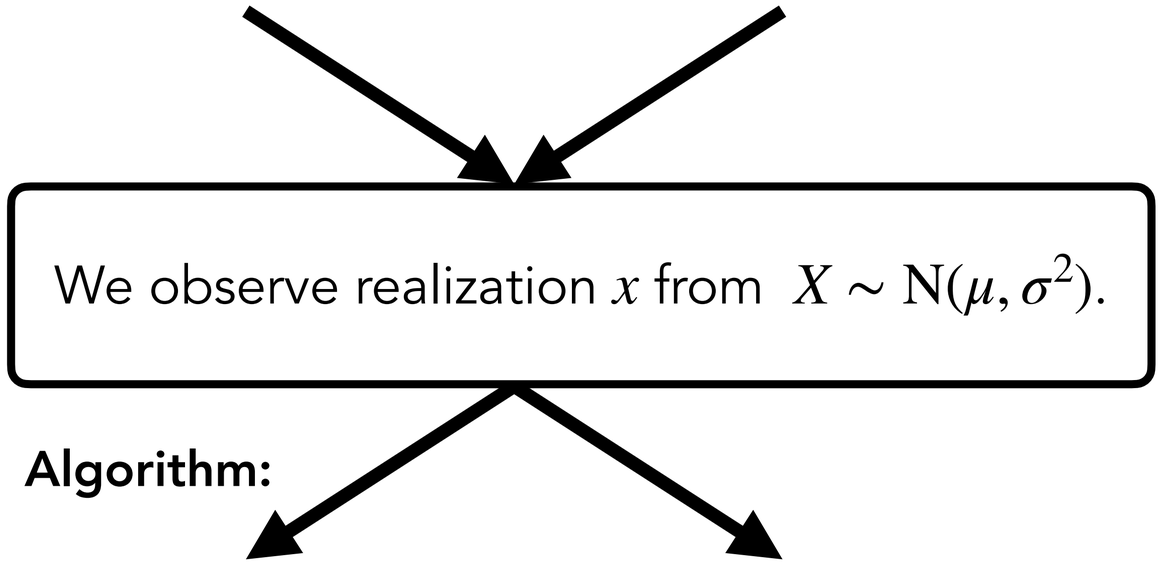
Can we work backwards to recover x' and x''?

Can show that the conditional distribution of $X' \mid X = x$ is $\mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\sigma^2)$.

Data thinning recipe for the Gaussian distribution

We know x could have arisen as $x' + x''$, where

$$X' \sim \mathrm{N}(\epsilon\mu, \epsilon\sigma^2) \quad \perp\!\!\!\perp \quad X'' \sim \mathrm{N}((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$$



Can we work backwards to recover x' and x''?

Can show that the conditional distribution of $X' \mid X = x$ is $\mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\sigma^2)$.

Data thinning recipe for the Gaussian distribution

We know x could have arisen as $x' + x''$, where

$$X' \sim \mathrm{N}(\epsilon\mu, \epsilon\sigma^2) \quad \perp\!\!\!\perp \quad X'' \sim \mathrm{N}((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$$

We observe realization x from $X \sim \mathrm{N}(\mu, \sigma^2)$.

Algorithm:

Draw $X^{(1)} \mid X = x \sim \mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\sigma^2)$. Let $X^{(2)} = X - X^{(1)}$.

Theorem:

$X^{(1)} \sim \mathrm{N}(\epsilon\mu, \epsilon\sigma^2)$, $X^{(2)} \sim \mathrm{N}((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$, $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

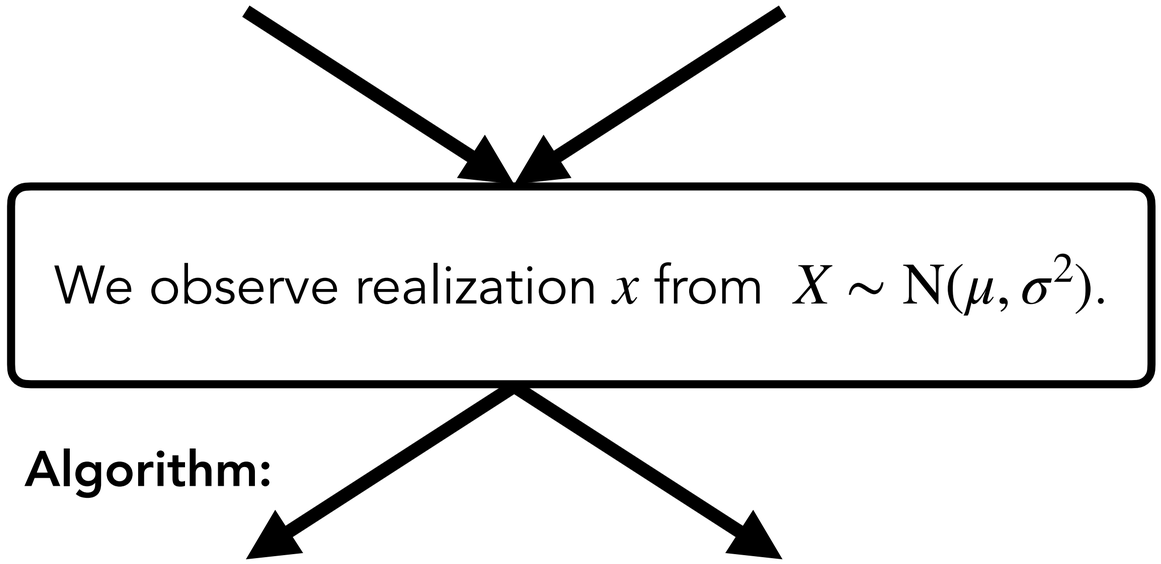
Can we work backwards to recover x' and x''?

Can show that the conditional distribution of $X' \mid X = x$ is $\mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\sigma^2)$.

Data thinning recipe for the Gaussian distribution

We know x could have arisen as $x' + x''$, where

$$X' \sim \mathrm{N}(\epsilon\mu, \epsilon\sigma^2) \quad \perp\!\!\!\perp \quad X'' \sim \mathrm{N}((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$$



Theorem:

$X^{(1)} \sim \mathrm{N}(\epsilon\mu, \epsilon\sigma^2)$, $X^{(2)} \sim \mathrm{N}((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$, $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

Can we work backwards to recover x' and x''?

Can show that the conditional distribution of $X' \mid X = x$ is $\mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\sigma^2)$.

Recovers a scaled version of another well-known result.

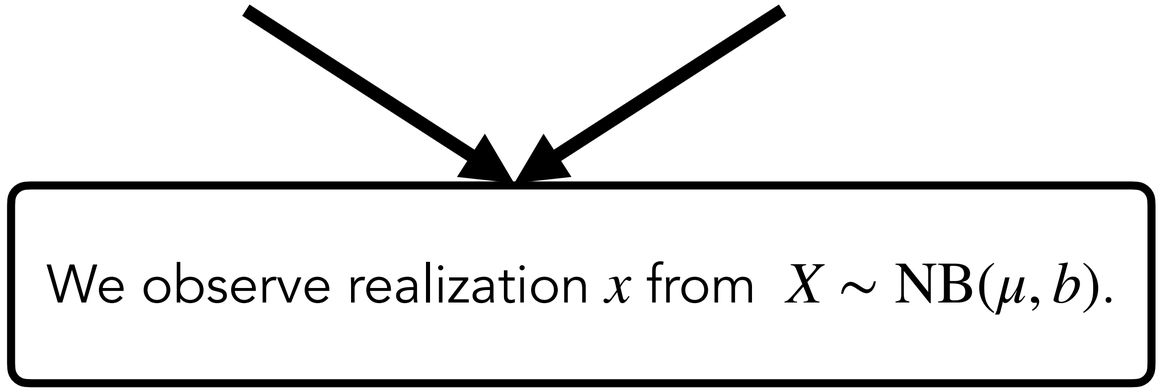
Data thinning recipe for the negative binomial distribution

We observe realization x from $X \sim \text{NB}(\mu, b)$.

Data thinning recipe for the negative binomial distribution

We know x could have arisen as $x' + x''$, where

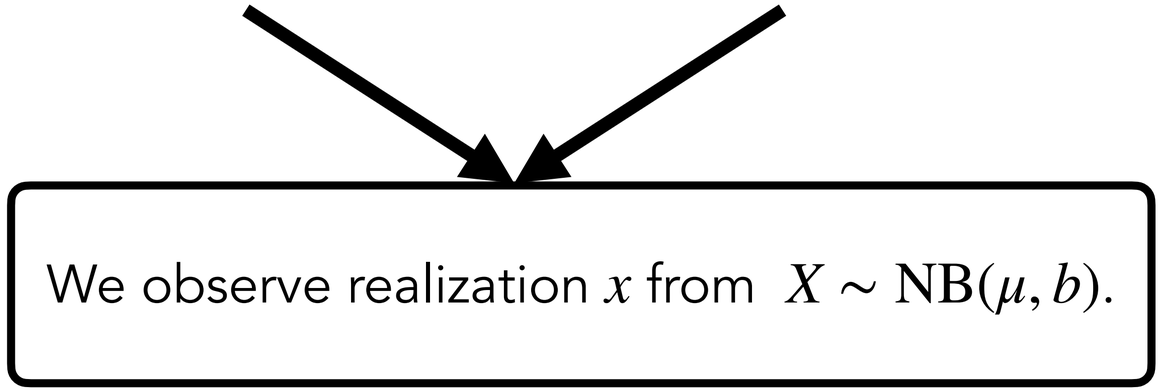
$$X' \sim \text{NB}(\epsilon\mu, \epsilon b) \quad \perp\!\!\!\perp \quad X'' \sim \text{NB}((1-\epsilon)\mu, (1-\epsilon)b)$$



Data thinning recipe for the negative binomial distribution

We know x could have arisen as $x' + x''$, where

$$X' \sim \mathrm{NB}(\epsilon\mu, \epsilon b) \quad \perp\!\!\!\perp \quad X'' \sim \mathrm{NB}((1-\epsilon)\mu, (1-\epsilon)b)$$



Can we work backwards to recover x' and x''?

Data thinning recipe for the negative binomial distribution

We know x could have arisen as $x' + x''$, where

$$X' \sim \text{NB}(\epsilon\mu, \epsilon b) \quad \perp\!\!\!\perp \quad X'' \sim \text{NB}((1-\epsilon)\mu, (1-\epsilon)b)$$

We observe realization x from $X \sim \text{NB}(\mu, b)$.

Can we work backwards to recover x' and x''?

Can show that the conditional distribution of $X' \mid X = x$ is $\text{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.

Data thinning recipe for the negative binomial distribution

We know x could have arisen as $x' + x''$, where

$X' \sim \text{NB}(\epsilon\mu, \epsilon b) \quad \perp\!\!\!\perp \quad X'' \sim \text{NB}((1-\epsilon)\mu, (1-\epsilon)b)$

We observe realization x from $X \sim \text{NB}(\mu, b)$.

Algorithm:

Draw
$X^{(1)} \mid X = x \sim \text{BetaBin}(x, \epsilon b, (1-\epsilon)b)$.

Let $X^{(2)} = X - X^{(1)}$.

Can we work backwards to recover x' and x''?

Can show that the conditional distribution of $X' \mid X = x$ is $\text{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.

Data thinning recipe for the negative binomial distribution

We know x could have arisen as $x' + x''$, where

$X' \sim \text{NB}(\epsilon\mu, \epsilon b) \quad \perp\!\!\!\perp \quad X'' \sim \text{NB}((1-\epsilon)\mu, (1-\epsilon)b)$

We observe realization x from $X \sim \text{NB}(\mu, b)$.

Algorithm:

Draw
$X^{(1)} \mid X = x \sim \text{BetaBin}(x, \epsilon b, (1-\epsilon)b)$.

Let $X^{(2)} = X - X^{(1)}$.

Theorem:

$X^{(1)} \sim \text{NB}(\epsilon\mu, \epsilon b)$, $X^{(2)} \sim \text{NB}((1-\epsilon)\mu, (1-\epsilon)b)$, $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

Can we work backwards to recover x' and x''?

Can show that the conditional distribution of $X' \mid X = x$ is $\text{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.

Data thinning recipe for the negative binomial distribution

We know x could have arisen as $x' + x''$, where

$X' \sim \text{NB}(\epsilon\mu, \epsilon b) \quad \perp\!\!\!\perp \quad X'' \sim \text{NB}((1-\epsilon)\mu, (1-\epsilon)b)$

We observe realization x from $X \sim \text{NB}(\mu, b)$.

Algorithm:

Draw
$X^{(1)} \mid X = x \sim \text{BetaBin}(x, \epsilon b, (1-\epsilon)b)$.

Let $X^{(2)} = X - X^{(1)}$.

Theorem:

$X^{(1)} \sim \text{NB}(\epsilon\mu, \epsilon b)$, $X^{(2)} \sim \text{NB}((1-\epsilon)\mu, (1-\epsilon)b)$, $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

Can we work backwards to recover x' and x''?

Can show that the conditional distribution of $X' \mid X = x$ is $\text{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.

This is a new result!

For many common distributions, the distribution $G_{\epsilon,x}$ has a simple form

Distribution of X:	Draw $X^{(1)} \mid X = x$ from $G_{\epsilon,x}$, where $G_{\epsilon,x}$ is:	Distribution of $X^{(1)}$:	Distribution of $X^{(2)}$, where $X^{(2)} = X - X^{(1)}$:
$\text{Poisson}(\lambda)$	$\text{Binomial}(x, \epsilon)$	$\text{Poisson}(\epsilon\lambda)$	$\text{Poisson}((1-\epsilon)\lambda)$

For many common distributions, the distribution $G_{\epsilon,x}$ has a simple form

Distribution of X:	Draw $X^{(1)} \mid X = x$ from $G_{\epsilon,x}$, where $G_{\epsilon,x}$ is:	Distribution of $X^{(1)}$:	Distribution of $X^{(2)}$, where $X^{(2)} = X - X^{(1)}$:
$\text{Poisson}(\lambda)$	$\text{Binomial}(x, \epsilon)$	$\text{Poisson}(\epsilon\lambda)$	$\text{Poisson}((1-\epsilon)\lambda)$

Related work on Poisson thinning:

- Sarkar and Stephens, 2021, Nature Genetics.
- Chen et al., 2021, arXiv:2108.03336
- Leiner et al., 2021, arXiv:2112.11079.
- Neufeld et al., 2022, Biostatistics.
- Oliveira, Lei, and Tibshirani, 2022, arXiv:2212.01943.

For many common distributions, the distribution $G_{\epsilon,x}$ has a simple form

Distribution of X:	Draw $X^{(1)} \mid X = x$ from $G_{\epsilon,x}$, where $G_{\epsilon,x}$ is:	Distribution of $X^{(1)}$:	Distribution of $X^{(2)}$, where $X^{(2)} = X - X^{(1)}$:
$\text{Poisson}(\lambda)$	$\text{Binomial}(x, \epsilon)$	$\text{Poisson}(\epsilon\lambda)$	$\text{Poisson}((1-\epsilon)\lambda)$
$\text{N}(\mu, \sigma^2)$	$\text{N}(\epsilon x, \epsilon(1-\epsilon)\sigma^2)$	$\text{N}(\epsilon\mu, \epsilon\sigma^2)$	$\text{N}((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$

For many common distributions, the distribution $G_{\epsilon,x}$ has a simple form

Distribution of X:	Draw $X^{(1)} \mid X = x$ from $G_{\epsilon,x}$, where $G_{\epsilon,x}$ is:	Distribution of $X^{(1)}$:	Distribution of $X^{(2)}$, where $X^{(2)} = X - X^{(1)}$:
$\text{Poisson}(\lambda)$	$\text{Binomial}(x, \epsilon)$	$\text{Poisson}(\epsilon\lambda)$	$\text{Poisson}((1-\epsilon)\lambda)$
$\text{N}(\mu, \sigma^2)$	$\text{N}(\epsilon x, \epsilon(1-\epsilon)\sigma^2)$	$\text{N}(\epsilon\mu, \epsilon\sigma^2)$	$\text{N}((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$

Related work on Gaussian thinning:

- Tian and Taylor, 2018, Annals of Statistics.
- Rasines and Young, 2022, Biometrika.
- Leiner et al., 2021, arXiv:2112.11079.
- Tian, 2020, Annals of Statistics.
- Oliveira, Lei, and Tibshirani, 2021, arXiv:2111.09447.

For many common distributions, the distribution $G_{\epsilon,x}$ has a simple form

Distribution of X:	Draw $X^{(1)} \mid X = x$ from $G_{\epsilon,x}$, where $G_{\epsilon,x}$ is:	Distribution of $X^{(1)}$:	Distribution of $X^{(2)}$, where $X^{(2)} = X - X^{(1)}$:
$\text{Poisson}(\lambda)$	$\text{Binomial}(x, \epsilon)$	$\text{Poisson}(\epsilon\lambda)$	$\text{Poisson}((1-\epsilon)\lambda)$
$\text{N}(\mu, \sigma^2)$	$\text{N}(\epsilon x, \epsilon(1-\epsilon)\sigma^2)$	$\text{N}(\epsilon\mu, \epsilon\sigma^2)$	$\text{N}((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$
$\text{NegativeBinomial}(\mu, b)$	$\text{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.	$\text{NegativeBinomial}(\epsilon\mu, \epsilon b)$	$\text{NegativeBinomial}((1-\epsilon)\mu, (1-\epsilon)b)$

For many common distributions, the distribution $G_{\epsilon,x}$ has a simple form

Distribution of X:	Draw $X^{(1)} \mid X = x$ from $G_{\epsilon,x}$, where $G_{\epsilon,x}$ is:	Distribution of $X^{(1)}$:	Distribution of $X^{(2)}$, where $X^{(2)} = X - X^{(1)}$:
$\text{Poisson}(\lambda)$	$\text{Binomial}(x, \epsilon)$	$\text{Poisson}(\epsilon\lambda)$	$\text{Poisson}((1-\epsilon)\lambda)$
$\text{N}(\mu, \sigma^2)$	$\text{N}(\epsilon x, \epsilon(1-\epsilon)\sigma^2)$	$\text{N}(\epsilon\mu, \epsilon\sigma^2)$	$\text{N}((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$
$\text{NegativeBinomial}(\mu, b)$	$\text{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.	$\text{NegativeBinomial}(\epsilon\mu, \epsilon b)$	$\text{NegativeBinomial}((1-\epsilon)\mu, (1-\epsilon)b)$
$\text{Binomial}(r, p)$	$\text{Hypergeometric}(\epsilon r, (1-\epsilon)r, x)$.	$\text{Binomial}(\epsilon r, p)$	$\text{Binomial}((1-\epsilon)r, p)$

For many common distributions, the distribution $G_{\epsilon,x}$ has a simple form

Distribution of X:	Draw $X^{(1)} \mid X = x$ from $G_{\epsilon,x}$, where $G_{\epsilon,x}$ is:	Distribution of $X^{(1)}$:	Distribution of $X^{(2)}$, where $X^{(2)} = X - X^{(1)}$:
$\text{Poisson}(\lambda)$	$\text{Binomial}(x, \epsilon)$	$\text{Poisson}(\epsilon\lambda)$	$\text{Poisson}((1-\epsilon)\lambda)$
$\text{N}(\mu, \sigma^2)$	$\text{N}(\epsilon x, \epsilon(1-\epsilon)\sigma^2)$	$\text{N}(\epsilon\mu, \epsilon\sigma^2)$	$\text{N}((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$
$\text{NegativeBinomial}(\mu, b)$	$\text{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.	$\text{NegativeBinomial}(\epsilon\mu, \epsilon b)$	$\text{NegativeBinomial}((1-\epsilon)\mu, (1-\epsilon)b)$
$\text{Binomial}(r, p)$	$\text{Hypergeometric}(\epsilon r, (1-\epsilon)r, x)$.	$\text{Binomial}(\epsilon r, p)$	$\text{Binomial}((1-\epsilon)r, p)$
$\text{Gamma}(\alpha, \beta)$	$x \cdot \text{Beta}(\epsilon\alpha, (1-\epsilon)\alpha)$.	$\text{Gamma}(\epsilon\alpha, \beta)$	$\text{Gamma}((1-\epsilon)\alpha, \beta)$
$\text{Exponential}(\lambda)$	$x \cdot \text{Beta}(\epsilon, (1-\epsilon))$.	$\text{Gamma}(\epsilon, \lambda)$	$\text{Gamma}(1-\epsilon, \lambda)$
$\text{N}_k(\mu, \Sigma)$	$\text{N}(\epsilon x, \epsilon(1-\epsilon)\Sigma)$.	$\text{N}_k(\epsilon\mu, \epsilon\Sigma)$	$\text{N}_k((1-\epsilon)\mu, (1-\epsilon)\Sigma)$
$\text{Multinomial}_k(r, p)$	$\text{MultivarHypergeom}(x_1, \ldots, x_K, \epsilon r)$	$\text{Multinom}_k(\epsilon r, p)$	$\text{Multinomial}_k((1-\epsilon)r, p)$
$\text{Wishart}_p(n, \Sigma)$.	$x^{1/2} Z x^{1/2}$, where . $Z \sim \text{MatrixBeta}_p(\epsilon n/2, (1-\epsilon)n/2)$	$\text{Wishart}_p(\epsilon n, \Sigma)$	$\text{Wishart}_p((1-\epsilon)n, \Sigma)$

What if we get a nuisance parameter wrong?

Gaussian thinning algorithm

Suppose $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$.

Draw
$X^{(1)} \sim \mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\,\sigma^2)$ and
$X^{(2)} = X - X^{(1)}$.
Then:

1) $X^{(1)} \sim \mathrm{N}\left(\epsilon\mu, \epsilon\sigma^2\right)$
2) $X^{(2)} \sim \mathrm{N}\left((1-\epsilon)\mu, (1-\epsilon)\sigma^2\right)$
3) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

What if we get a nuisance parameter wrong?

Gaussian thinning algorithm

Suppose $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$.

Draw

$X^{(1)} \sim \mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\,\tilde{\sigma}^2)$ and

$X^{(2)} = X - X^{(1)}$.

Then:

1) $X^{(1)} \sim \mathrm{N}\left(\epsilon\mu, \epsilon\sigma^2\right)$
2) $X^{(2)} \sim \mathrm{N}\left((1-\epsilon)\mu, (1-\epsilon)\sigma^2\right)$
3) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

What if we get a nuisance parameter wrong?

Gaussian thinning algorithm

Suppose $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$.

Draw

$X^{(1)} \sim \mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\,\tilde{\sigma}^2)$ and

$X^{(2)} = X - X^{(1)}$.

Then:

~~1) $X^{(1)} \sim \mathrm{N}\left(\epsilon\mu, \epsilon\sigma^2\right)$~~

~~2) $X^{(2)} \sim \mathrm{N}\left((1-\epsilon)\mu, (1-\epsilon)\sigma^2\right)$~~

~~3) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.~~

What if we get a nuisance parameter wrong?

Gaussian thinning algorithm

Suppose $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$.

Draw

$X^{(1)} \sim \mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\,\tilde{\sigma}^2)$ and

$X^{(2)} = X - X^{(1)}$.

Then:

1) $X^{(1)} \sim \mathrm{N}\left(\epsilon\mu, \epsilon^2\sigma^2 + \epsilon(1-\epsilon)\tilde{\sigma}^2\right)$
2) $X^{(2)} \sim \mathrm{N}\left((1-\epsilon)\mu, (1-\epsilon)^2\sigma^2 + \epsilon(1-\epsilon)\tilde{\sigma}^2\right)$
3) $\mathrm{Cov}\left(X^{(1)}, X^{(2)}\right) = \epsilon(1-\epsilon)\left(\sigma^2 - \tilde{\sigma}^2\right)$.

What if we get a nuisance parameter wrong?

Gaussian thinning algorithm

Suppose $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$.

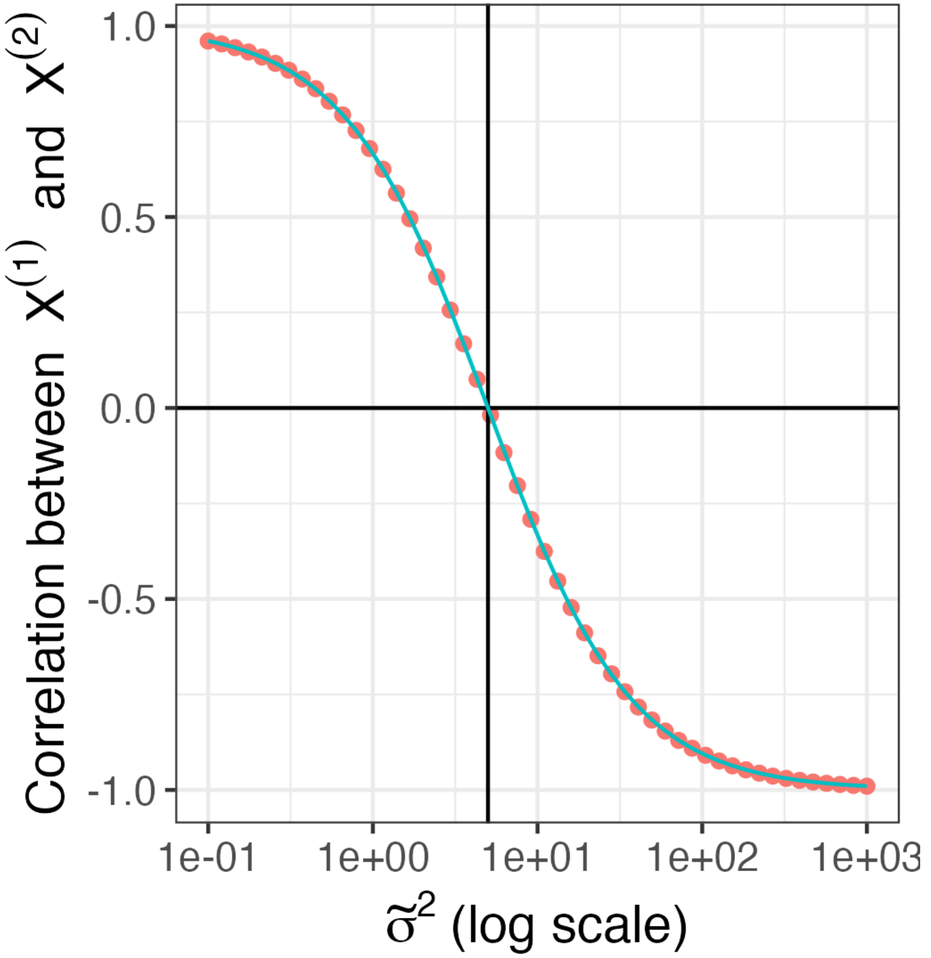
Draw

$X^{(1)} \sim \mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\ \tilde{\sigma}^2)$ and

$X^{(2)} = X - X^{(1)}$.

Then:

1) $X^{(1)} \sim \mathrm{N}\left(\epsilon\mu, \epsilon^2\sigma^2 + \epsilon(1-\epsilon)\tilde{\sigma}^2\right)$
2) $X^{(2)} \sim \mathrm{N}\left((1-\epsilon)\mu, (1-\epsilon)^2\sigma^2 + \epsilon(1-\epsilon)\tilde{\sigma}^2\right)$
3) $\mathrm{Cov}\left(X^{(1)}, X^{(2)}\right) = \epsilon(1-\epsilon)\left(\sigma^2 - \tilde{\sigma}^2\right)$.



What if we get a nuisance parameter wrong?

Negative binomial thinning algorithm

Suppose $X \sim \text{NegBin}\left(\mu, b\right)$.

Draw

$X^{(1)} \sim \text{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$,

$X^{(2)} = X - X^{(1)}$, then:

1) $X^{(1)} \sim \text{NegBin}\left(\epsilon\mu, \epsilon b\right)$.
2) $X^{(2)} \sim \text{NegBin}\left((1-\epsilon)\mu, (1-\epsilon)b\right)$
3) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

What if we get a nuisance parameter wrong?

Negative binomial thinning algorithm

Suppose $X \sim \text{NegBin}\left(\mu, b\right)$.

Draw

$X^{(1)} \sim \text{BetaBinomial}(x, \epsilon\tilde{b}, (1-\epsilon)\tilde{b}\ )$,

$X^{(2)} = X - X^{(1)}$, then:

1) $X^{(1)} \sim \text{NegBin}\left(\epsilon\mu, \epsilon b\right)$.
2) $X^{(2)} \sim \text{NegBin}\left((1-\epsilon)\mu, (1-\epsilon)b\right)$
3) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

What if we get a nuisance parameter wrong?

Negative binomial thinning algorithm

Suppose $X \sim \text{NegBin}\left(\mu, b\right)$.

Draw

$X^{(1)} \sim \text{BetaBinomial}(x, \epsilon\tilde{b}, (1-\epsilon)\tilde{b}\ )$,

$X^{(2)} = X - X^{(1)}$, then:

~~1) $X^{(1)} \sim \text{NegBin}\left(\epsilon\mu, \epsilon b\right)$.~~

~~2) $X^{(2)} \sim \text{NegBin}\left((1-\epsilon)\mu, (1-\epsilon)b\right)$~~

~~3) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.~~

What if we get a nuisance parameter wrong?

Negative binomial thinning algorithm

Suppose $X \sim \text{NegBin}\left(\mu, b\right)$.

Draw

$X^{(1)} \sim \text{BetaBinomial}(x, \epsilon\tilde{b}, (1-\epsilon)\tilde{b}\,)$,

$X^{(2)} = X - X^{(1)}$, then:

1) $\text{E}[X^{(1)}] = \epsilon\mu.$
2) $\text{E}[X^{(2)}] = (1-\epsilon)\mu$
3) $\text{Cov}\left(X^{(1)}, X^{(2)}\right) = \epsilon(1-\epsilon)\dfrac{\mu^2}{b}\left(1 - \dfrac{b+1}{\tilde{b}+1}\right).$

What if we get a nuisance parameter wrong?

Negative binomial thinning algorithm

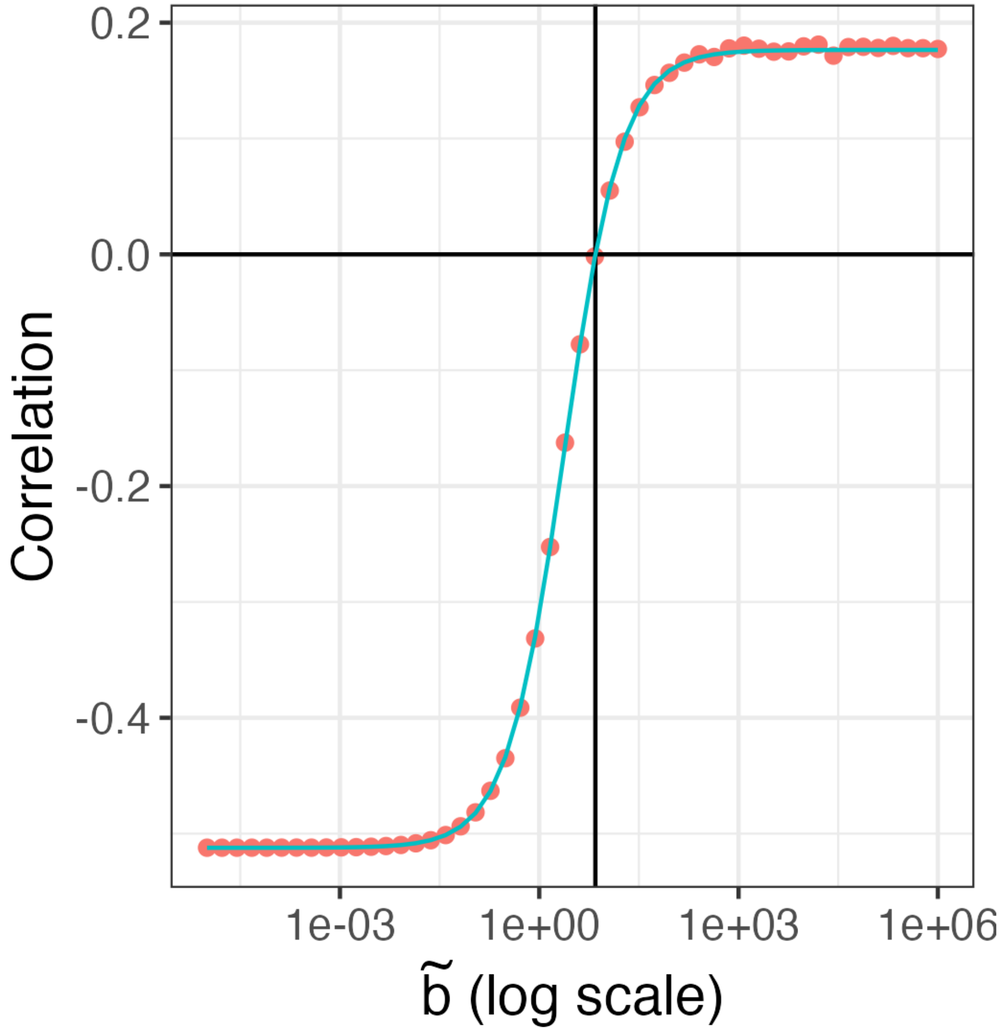
Suppose $X \sim \text{NegBin}\left(\mu, b\right)$.

Draw

$X^{(1)} \sim \text{BetaBinomial}(x, \epsilon\tilde{b}, (1-\epsilon)\tilde{b}\,)$,

$X^{(2)} = X - X^{(1)}$, then:

1) $\text{E}[X^{(1)}] = \epsilon\mu.$
2) $\text{E}[X^{(2)}] = (1-\epsilon)\mu$
3) $\text{Cov}\left(X^{(1)}, X^{(2)}\right) = \epsilon(1-\epsilon)\dfrac{\mu^2}{b}\left(1 - \dfrac{b+1}{\tilde{b}+1}\right).$



The tuning parameter ϵ governs an information tradeoff

Gaussian thinning algorithm

Suppose $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$.

Draw
$X^{(1)} \sim \mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\,\sigma^2)$ and
$X^{(2)} = X - X^{(1)}$.
Then:

1) $X^{(1)} \sim \mathrm{N}\left(\epsilon\mu, \epsilon\sigma^2\right)$
2) $X^{(2)} \sim \mathrm{N}\left((1-\epsilon)\mu, (1-\epsilon)\sigma^2\right)$
3) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

The tuning parameter ϵ governs an information tradeoff

Gaussian thinning algorithm

Suppose $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$.

Draw
$X^{(1)} \sim \mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\,\sigma^2)$ and
$X^{(2)} = X - X^{(1)}$.
Then:

1) $X^{(1)} \sim \mathrm{N}\left(\epsilon\mu, \epsilon\sigma^2\right)$
2) $X^{(2)} \sim \mathrm{N}\left((1-\epsilon)\mu, (1-\epsilon)\sigma^2\right)$
3) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

Theorem: Fisher Information in X about μ is divided between $X^{(1)}$ and $X^{(2)}$ with proportions ϵ and $1-\epsilon$.

The tuning parameter ϵ governs an information tradeoff

Gaussian thinning algorithm

Suppose $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$.

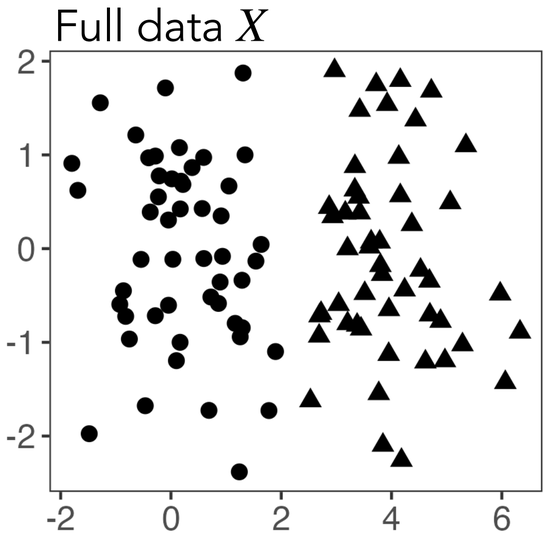
Draw
$X^{(1)} \sim \mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\,\sigma^2)$ and
$X^{(2)} = X - X^{(1)}$.
Then:

1) $X^{(1)} \sim \mathrm{N}\left(\epsilon\mu, \epsilon\sigma^2\right)$
2) $X^{(2)} \sim \mathrm{N}\left((1-\epsilon)\mu, (1-\epsilon)\sigma^2\right)$
3) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

Theorem: Fisher Information in X about μ is divided between $X^{(1)}$ and $X^{(2)}$ with proportions ϵ and $1-\epsilon$.

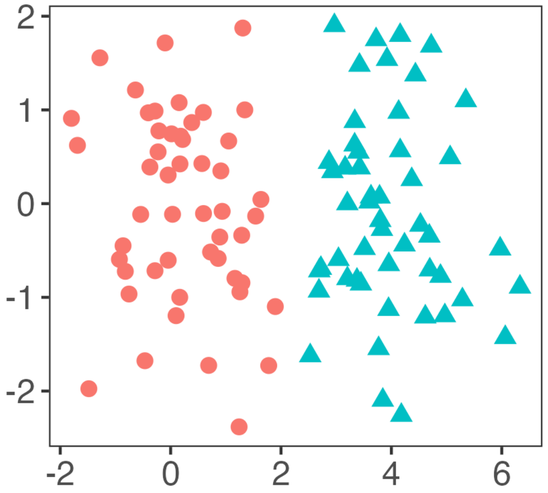
Similar results can be derived for other decompositions.

Visualizing the role of ϵ



Visualizing the role of ϵ

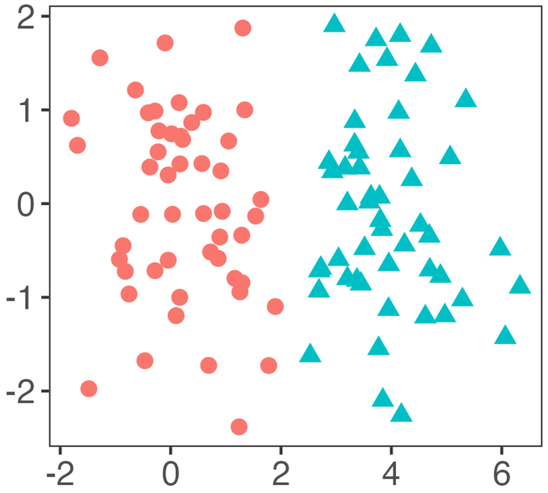
Full data X



Adjusted Rand index = 1

Visualizing the role of ϵ

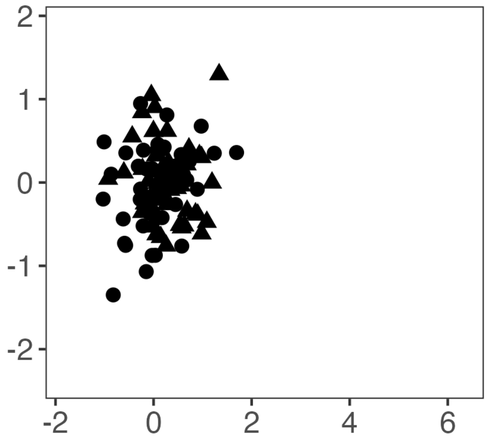
Full data X



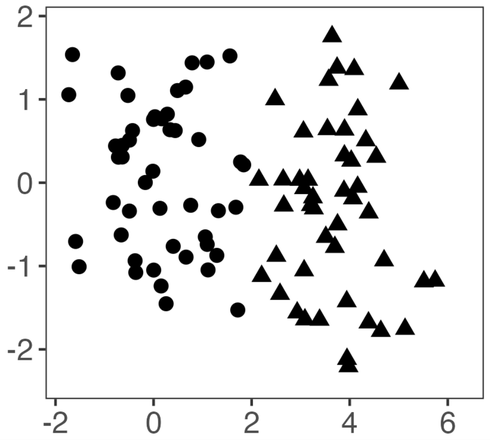
Adjusted Rand index = 1

$\epsilon = 0.1$

$X^{(1)}$

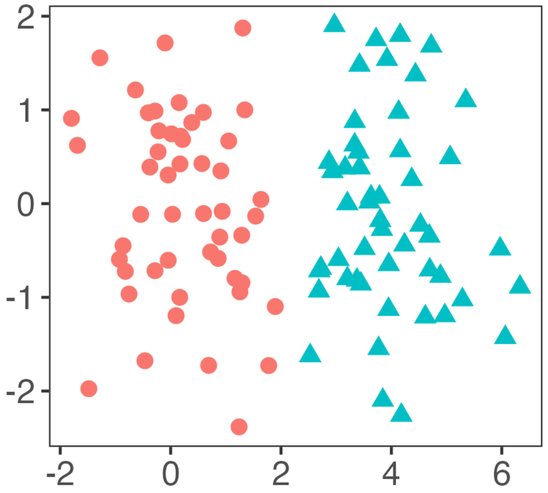


$X^{(2)}$



Visualizing the role of ϵ

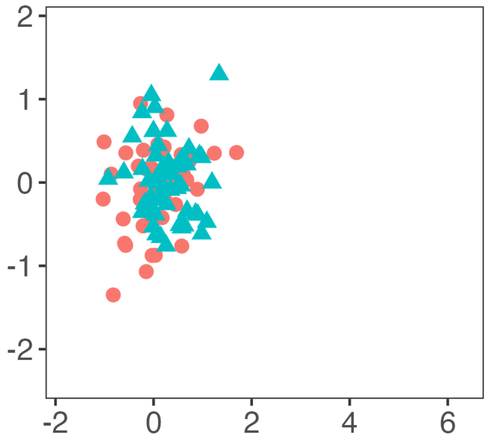
Full data X



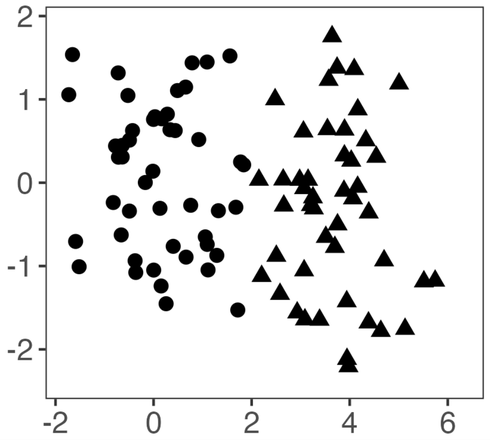
Adjusted Rand index = 1

$\epsilon = 0.1$

$X^{(1)}$

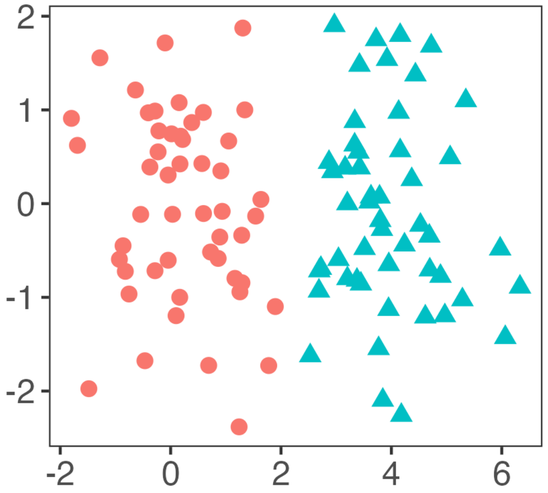


$X^{(2)}$



Visualizing the role of ϵ

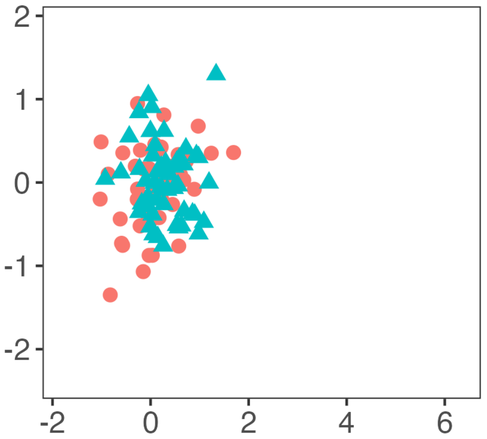
Full data X



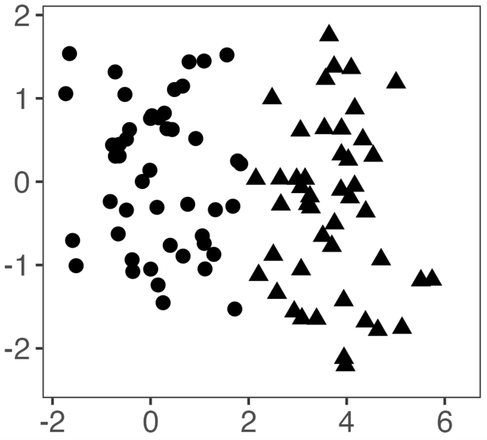
Adjusted Rand index = 1

$\epsilon = 0.1$

$X^{(1)}$, ARI = 0.03

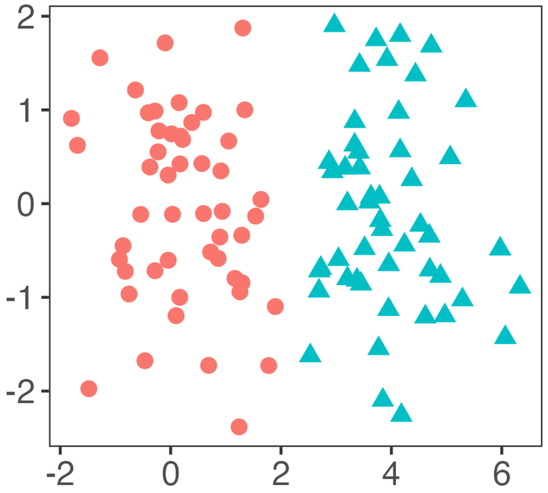


$X^{(2)}$



Visualizing the role of ϵ

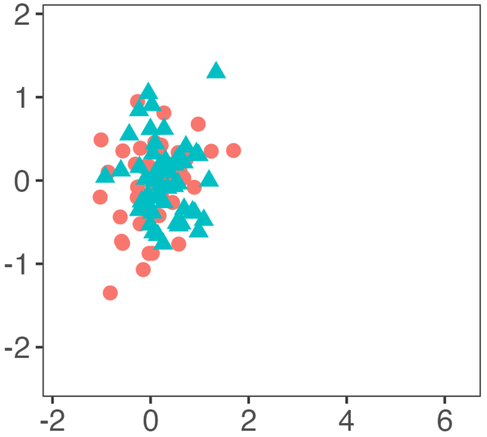
Full data X



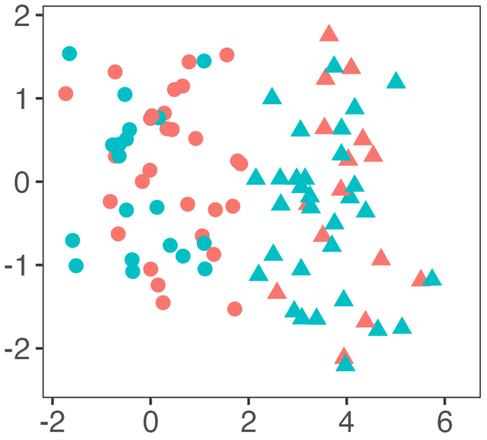
Adjusted Rand index = 1

$\epsilon = 0.1$

$X^{(1)}$, ARI = 0.03

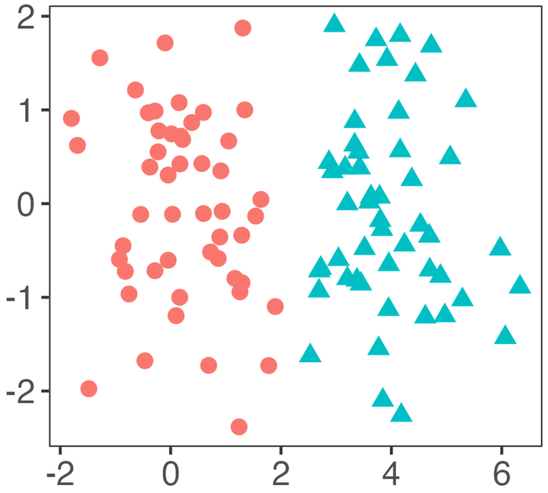


$X^{(2)}$



Visualizing the role of ϵ

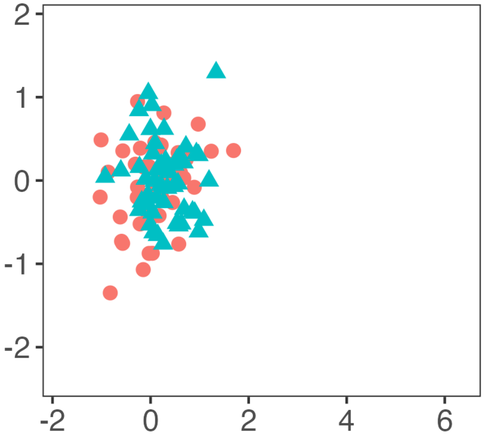
Full data X



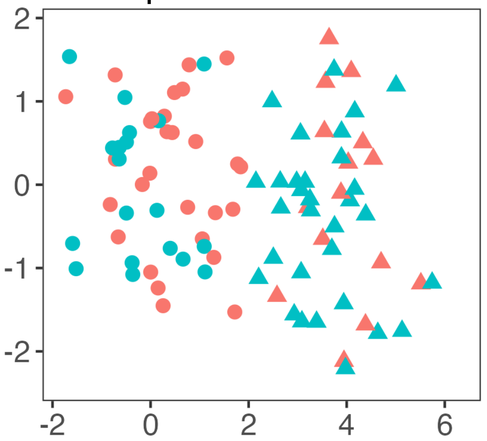
Adjusted Rand index = 1

$\epsilon = 0.1$

$X^{(1)}$, ARI = 0.03

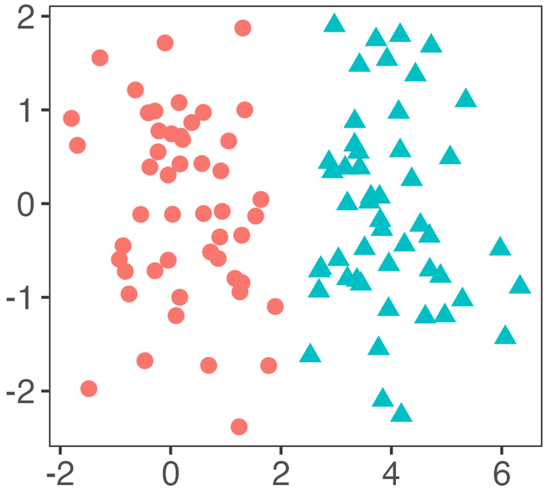


$X^{(2)}$, p-value = 0.36



Visualizing the role of ϵ

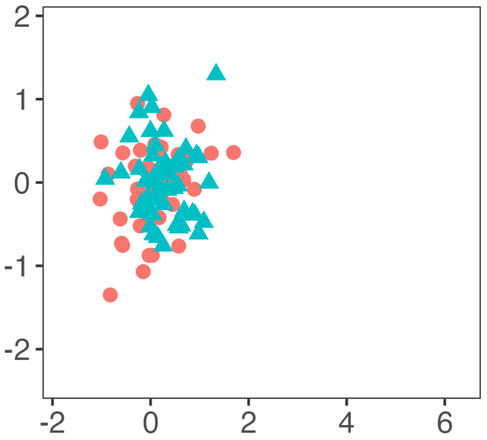
Full data X



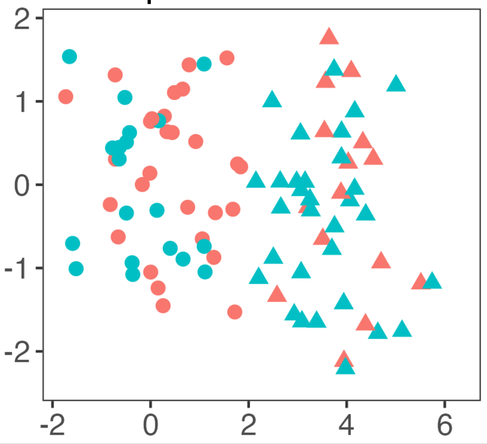
Adjusted Rand index = 1

$\epsilon = 0.1$

$X^{(1)}$, ARI = 0.03

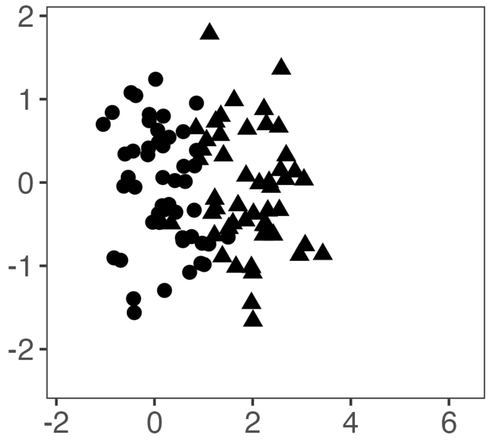


$X^{(2)}$, p-value = 0.36

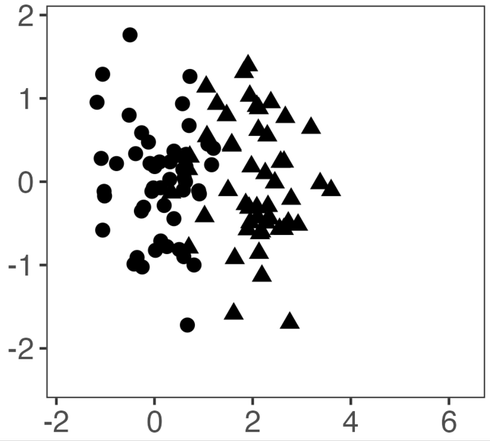


$\epsilon = 0.5$

$X^{(1)}$

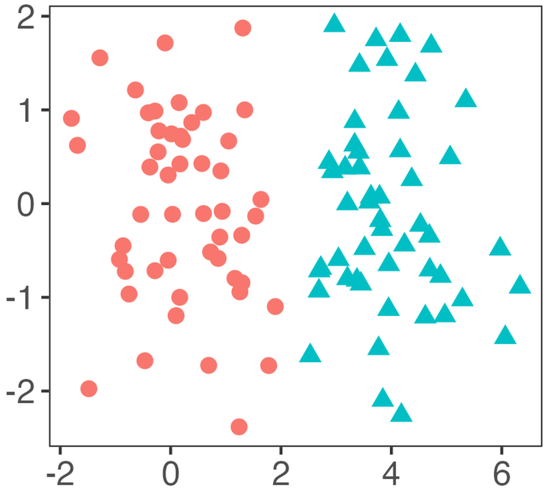


$X^{(2)}$



Visualizing the role of ϵ

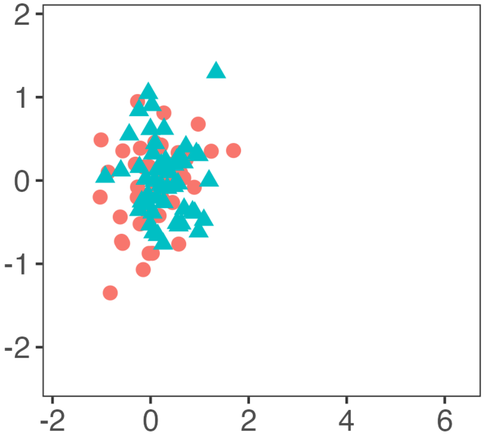
Full data X



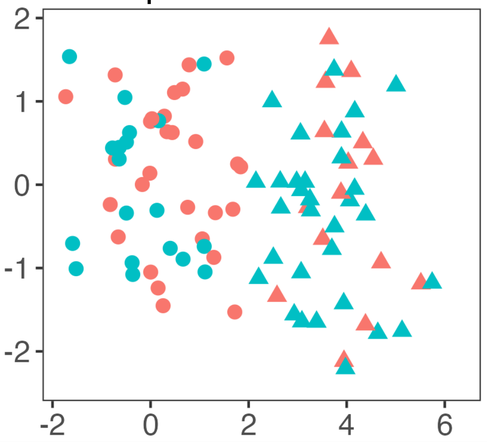
Adjusted Rand index = 1

$\epsilon = 0.1$

$X^{(1)}$, ARI = 0.03

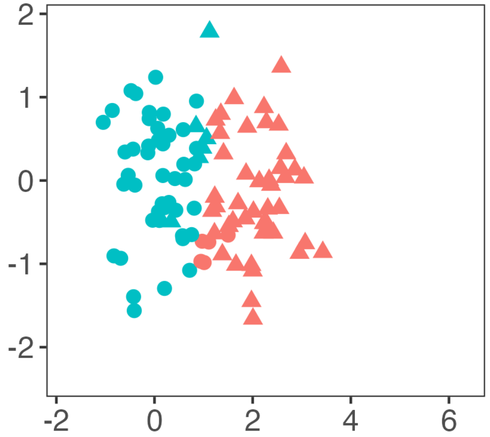


$X^{(2)}$, p-value = 0.36

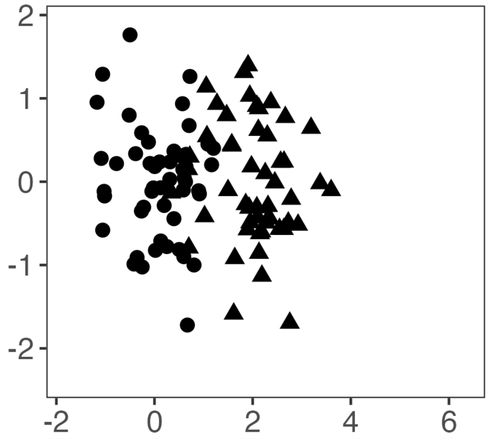


$\epsilon = 0.5$

$X^{(1)}$, ARI = 0.65

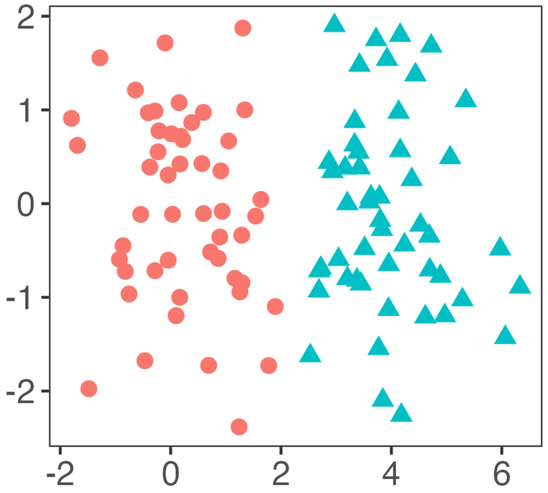


$X^{(2)}$



Visualizing the role of ϵ

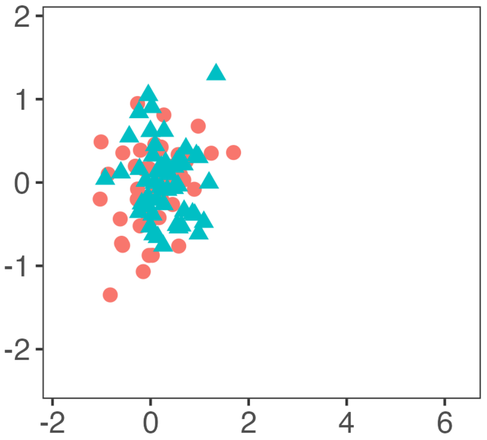
Full data X



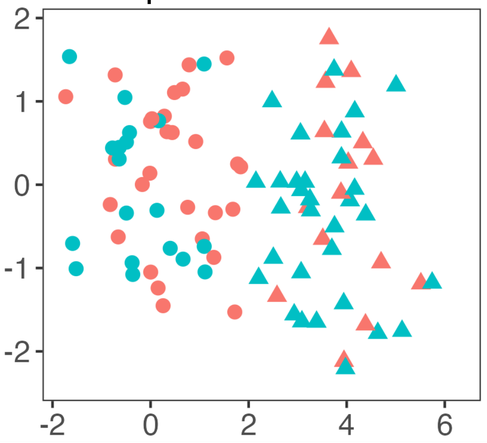
Adjusted Rand index = 1

$\epsilon = 0.1$

$X^{(1)}$, ARI = 0.03

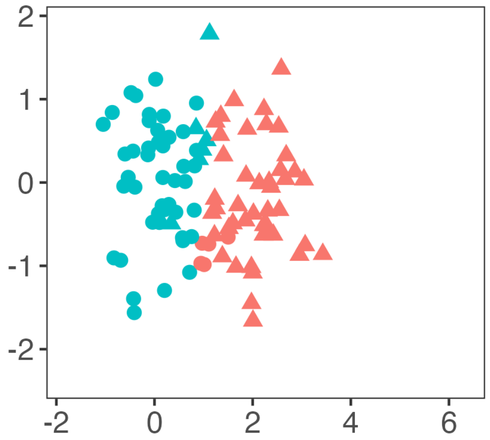


$X^{(2)}$, p-value = 0.36

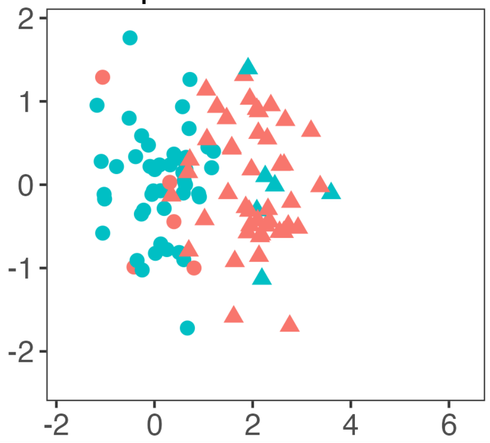


$\epsilon = 0.5$

$X^{(1)}$, ARI = 0.65

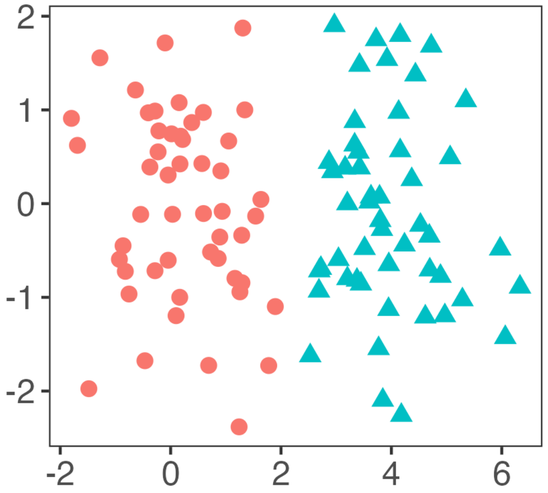


$X^{(2)}$, p-value = 0.02



Visualizing the role of ϵ

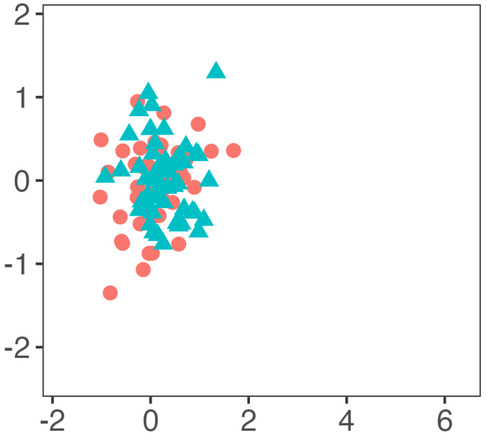
Full data X



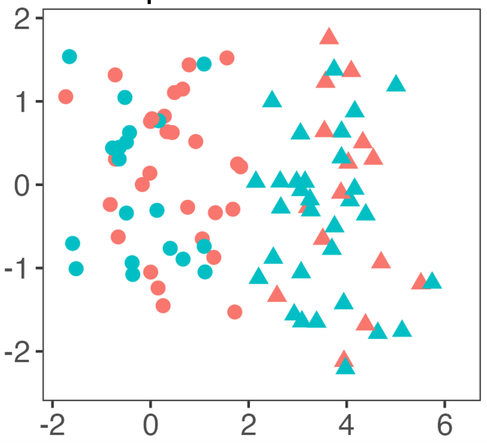
Adjusted Rand index = 1

$\epsilon = 0.1$

$X^{(1)}$, ARI = 0.03

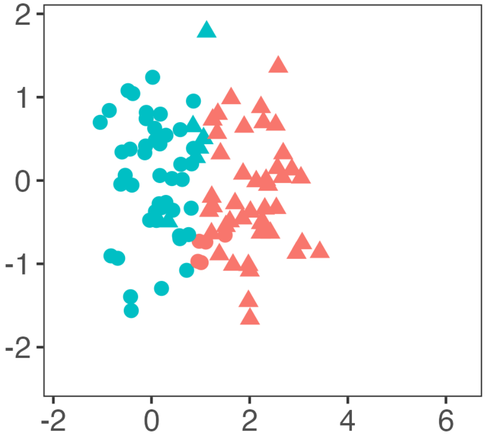


$X^{(2)}$, p-value = 0.36

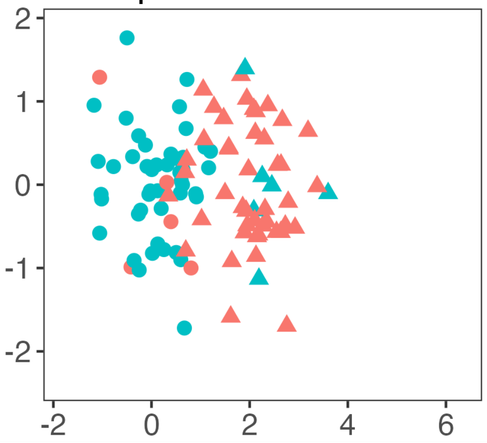


$\epsilon = 0.5$

$X^{(1)}$, ARI = 0.65

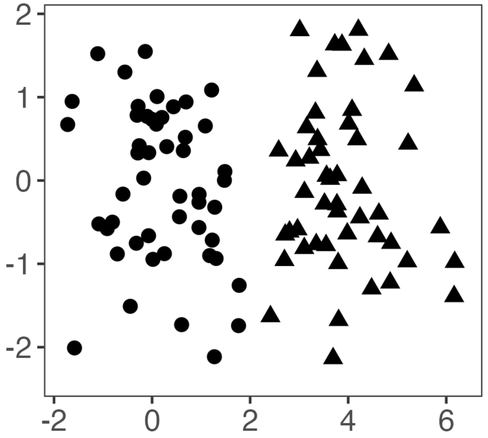


$X^{(2)}$, p-value = 0.02

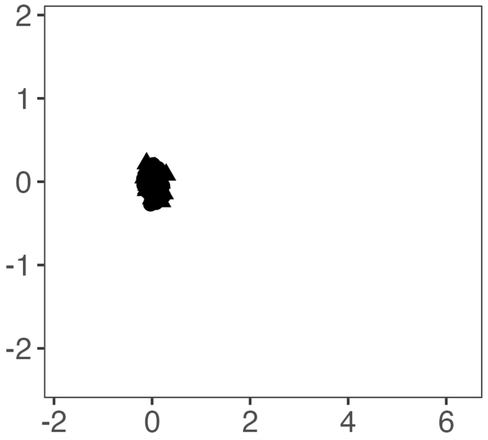


$\epsilon = 0.99$

$X^{(1)}$

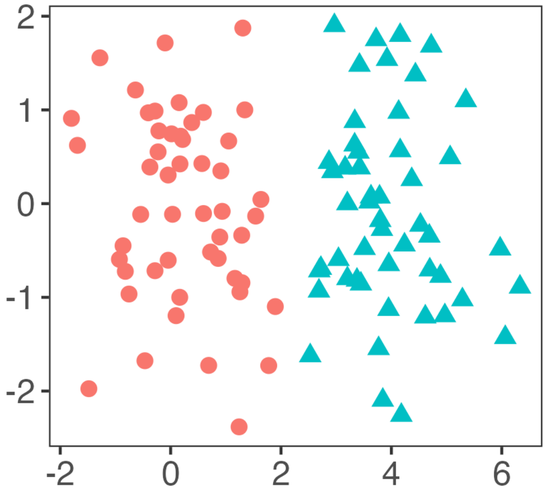


$X^{(2)}$



Visualizing the role of ϵ

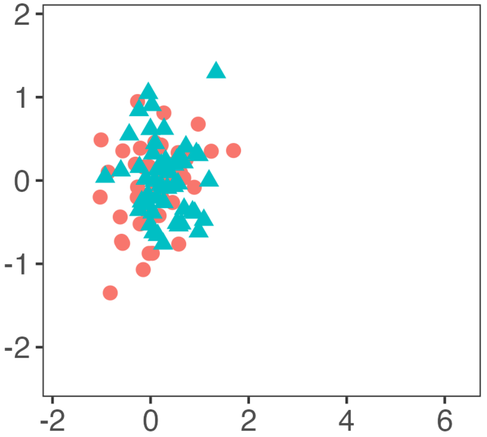
Full data X



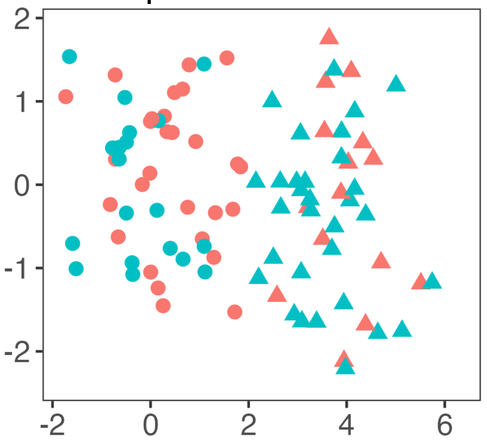
Adjusted Rand index = 1

$\epsilon = 0.1$

$X^{(1)}$, ARI = 0.03

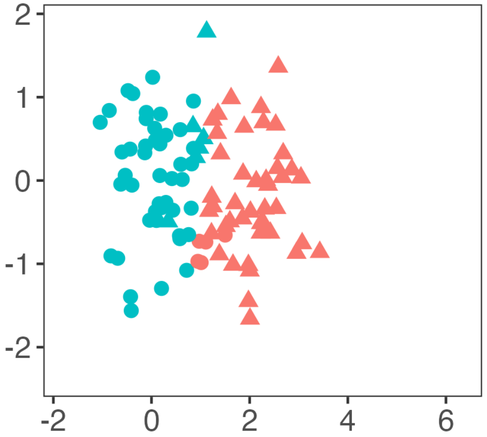


$X^{(2)}$, p-value = 0.36

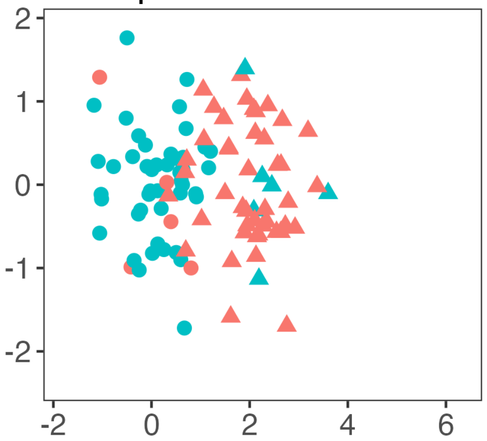


$\epsilon = 0.5$

$X^{(1)}$, ARI = 0.65

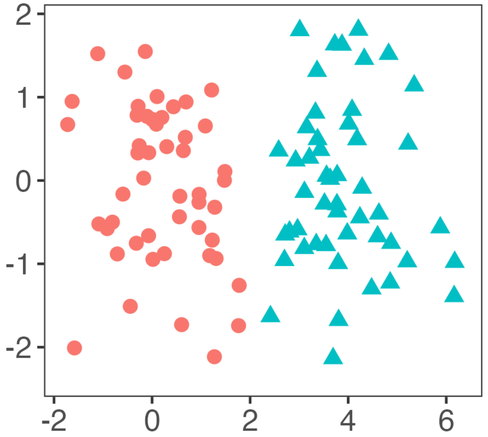


$X^{(2)}$, p-value = 0.02

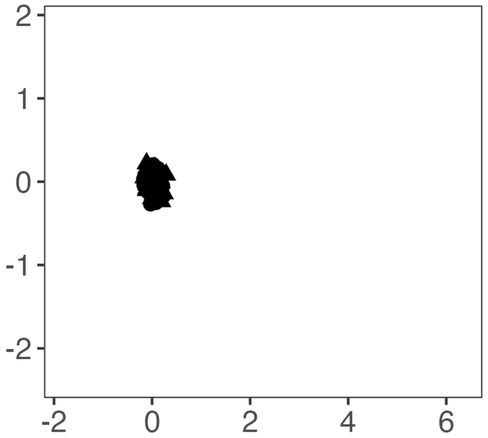


$\epsilon = 0.99$

$X^{(1)}$, ARI = 1

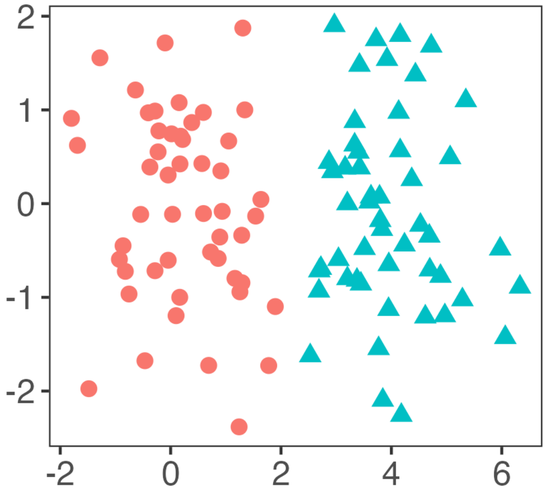


$X^{(2)}$



Visualizing the role of ϵ

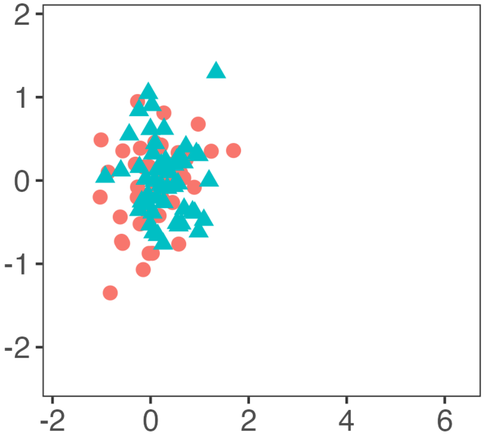
Full data X



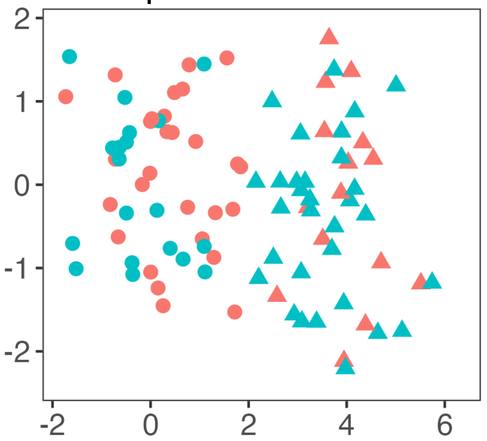
Adjusted Rand index = 1

$\epsilon = 0.1$

$X^{(1)}$, ARI = 0.03

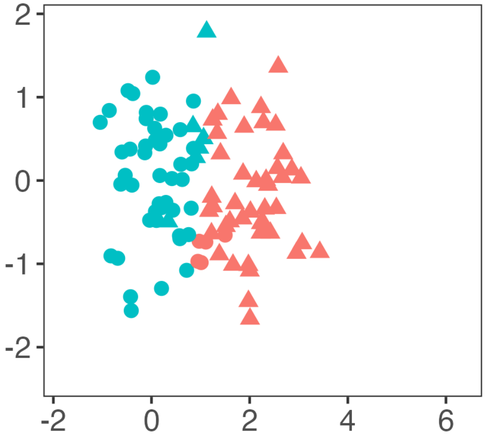


$X^{(2)}$, p-value = 0.36

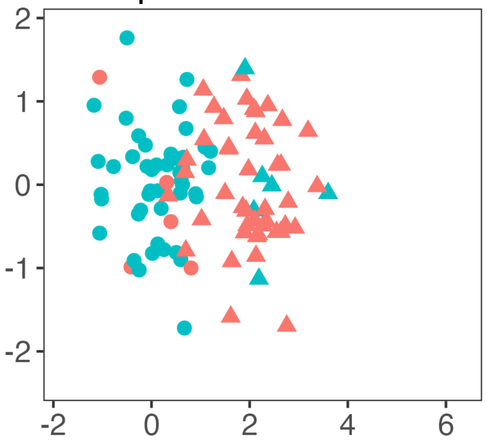


$\epsilon = 0.5$

$X^{(1)}$, ARI = 0.65

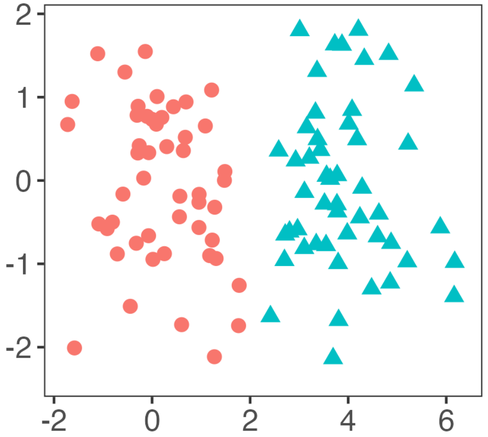


$X^{(2)}$, p-value = 0.02

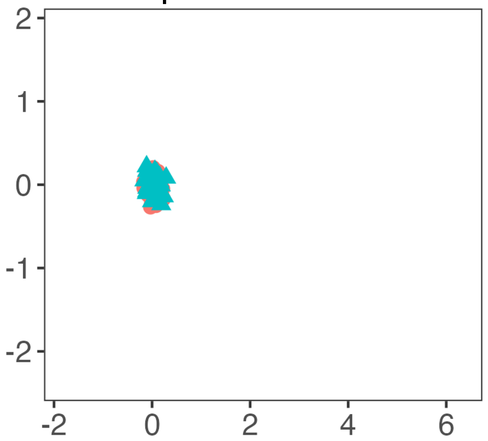


$\epsilon = 0.99$

$X^{(1)}$, ARI = 1



$X^{(2)}$, p-value = 0.11



Our recipe extends naturally to splitting into M>2 folds

Goal: split a single observation X into $\left(X^{(1)}, \ldots, X^{(M)}\right)$ such that:

(1) Each $X^{(m)}$ has the same distribution as X, up to a parameter scaling.

(2) The $X^{(m)}$ are mutually independent.

For many distributions, the multifold recipe has a simple form

Distribution of X	Draw $\left(X^{(1)}, \ldots, X^{(M)}\right) \mid X = x$ from:	Distribution of $X^{(m)}$
$\text{Poisson}(\lambda)$	$\text{Multinomial}(x, \epsilon_1, \ldots, \epsilon_M)$	$\text{Poisson}(\epsilon_m \lambda)$

For many distributions, the multifold recipe has a simple form

Distribution of X	Draw $\left(X^{(1)}, \ldots, X^{(M)}\right) \mid X = x$ from:	Distribution of $X^{(m)}$
$\text{Poisson}(\lambda)$	$\text{Multinomial}(x, \epsilon_1, \ldots, \epsilon_M)$	$\text{Poisson}(\epsilon_m \lambda)$
$\text{N}(\mu, \sigma^2)$	$\text{N}_M(\epsilon\mu, \sigma^2 \text{diag}(\epsilon) - \sigma^2 \epsilon\epsilon^T)$.	$\text{N}(\epsilon_m \mu, \epsilon_m \sigma^2)$

For many distributions, the multifold recipe has a simple form

Distribution of X	Draw $\left(X^{(1)}, \ldots, X^{(M)}\right) \mid X = x$ from:	Distribution of $X^{(m)}$
$\text{Poisson}(\lambda)$	$\text{Multinomial}(x, \epsilon_1, \ldots, \epsilon_M)$	$\text{Poisson}(\epsilon_m \lambda)$
$\text{N}(\mu, \sigma^2)$	$\text{N}_M(\epsilon\mu, \sigma^2 \text{diag}(\epsilon) - \sigma^2 \epsilon\epsilon^T)$.	$\text{N}(\epsilon_m \mu, \epsilon_m \sigma^2)$
$\text{NegativeBinomial}(\mu, b)$	$\text{DirichletMultinomial}(x, \epsilon_1 b, \ldots, \epsilon_M b)$.	$\text{NegativeBinomial}(\epsilon_m \mu, \epsilon_m b)$

For many distributions, the multifold recipe has a simple form

Distribution of X	Draw $\left(X^{(1)}, \ldots, X^{(M)}\right) \mid X = x$ from:	Distribution of $X^{(m)}$
$\text{Poisson}(\lambda)$	$\text{Multinomial}(x, \epsilon_1, \ldots, \epsilon_M)$	$\text{Poisson}(\epsilon_m \lambda)$
$\text{N}(\mu, \sigma^2)$	$\text{N}_M(\epsilon\mu, \sigma^2 \text{diag}(\epsilon) - \sigma^2 \epsilon\epsilon^T)$.	$\text{N}(\epsilon_m \mu, \epsilon_m \sigma^2)$
$\text{NegativeBinomial}(\mu, b)$	$\text{DirichletMultinomial}(x, \epsilon_1 b, \ldots, \epsilon_M b)$.	$\text{NegativeBinomial}(\epsilon_m \mu, \epsilon_m b)$
$\text{Gamma}(\alpha, \beta)$	$x \cdot \text{Dirichlet}\left(\epsilon_1 \alpha, \ldots, \epsilon_M \alpha\right)$	$\text{Gamma}(\epsilon_m \alpha, \beta)$
$\text{Exponential}(\lambda)$	$x \cdot \text{Dirichlet}\left(\epsilon_1, \ldots, \epsilon_M\right)$	$\text{Gamma}(\epsilon_m, \lambda)$
$\text{Binomial}(r, p)$	$\text{MultivariateHypergeometric}(\epsilon_1 r, \ldots, \epsilon_M r, x)$.	$\text{Binomial}(\epsilon_m r, p)$

Statistics > Methodology

[Submitted on 18 Jan 2023]

Data thinning for convolution-closed distributions

Anna Neufeld, Ameer Dharamshi, Lucy L. Gao, Daniela Witten

We propose data thinning, a new approach for splitting an observation into two or more independent parts that sum to the original observation, and that follow the same distribution as the original observation, up to a (known) scaling of a parameter. This proposal is very general, and can be applied to any observation drawn from a "convolution closed" distribution, a class that includes the Gaussian, Poisson, negative binomial, Gamma, and binomial distributions, among others. It is similar in spirit to -- but distinct from, and more easily applicable than -- a recent proposal known as data fission. Data thinning has a number of applications to model selection, evaluation, and inference. For instance, cross-validation via data thinning provides an attractive alternative to the "usual" approach of cross-validation via sample splitting, especially in unsupervised settings in which the latter is not applicable. In simulations and in an application to single-cell RNA-sequencing data, we show that data thinning can be used to validate the results of unsupervised learning approaches, such as k-means clustering and principal components analysis.

R package and tutorials: https://anna-neufeld.github.io/datathin/

Outline

1. Motivation: sample splitting doesn’t always work
2. Poisson thinning
3. Data thinning
4. **Generalized data thinning**
5. Application to changepoint validation
6. Ongoing work

Revisiting the goals of data thinning

Goal: split a single observation X into $X^{(1)}$ and $X^{(2)}$ such that:

(1) $X^{(1)}$ and $X^{(2)}$ have the same distribution as X, up to a parameter scaling.

(2) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

Revisiting the goals of data thinning

Goal: split a single observation X into $X^{(1)}$ and $X^{(2)}$ such that:

(1) $X^{(1)}$ and $X^{(2)}$ have the same distribution as X, up to a parameter scaling.

(2) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

In our previous recipe:

(3) $X = X^{(1)} + X^{(2)}$.

Revisiting the goals of data thinning

Goal: split a single observation X into $X^{(1)}$ and $X^{(2)}$ such that:

(1) $X^{(1)}$ and $X^{(2)}$ have the same distribution as X, up to a parameter scaling.

(2) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

In our previous recipe:

~~**(3)** $X = X^{(1)} + X^{(2)}$.~~

Revisiting the goals of data thinning

Goal: split a single observation X into $X^{(1)}$ and $X^{(2)}$ such that:

(1) $X^{(1)}$ and $X^{(2)}$ have the same distribution as X, up to a parameter scaling.

(2) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

In our previous recipe:

~~**(3)** $X = X^{(1)} + X^{(2)}$.~~ **(3)** $X = T(X^{(1)}, X^{(2)})$.

Revisiting the goals of data thinning

Goal: split a single observation X into $X^{(1)}$ and $X^{(2)}$ such that:

~~**(1)** $X^{(1)}$ and $X^{(2)}$ have the same distribution as X, up to a parameter scaling.~~

(2) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

In our previous recipe:

~~**(3)** $X = X^{(1)} + X^{(2)}$.~~ **(3)** $X = T(X^{(1)}, X^{(2)})$.

Revisiting the goals of data thinning

Goal: split a single observation X into $X^{(1)}$ and $X^{(2)}$ such that:

~~**(1)** $X^{(1)}$ and $X^{(2)}$ have the same distribution as X, up to a parameter scaling.~~

(2) $X^{(1)} \perp\!\!\!\perp X^{(2)}$.

In our previous recipe:

~~**(3)** $X = X^{(1)} + X^{(2)}$.~~ **(3)** $X = T(X^{(1)}, X^{(2)})$.

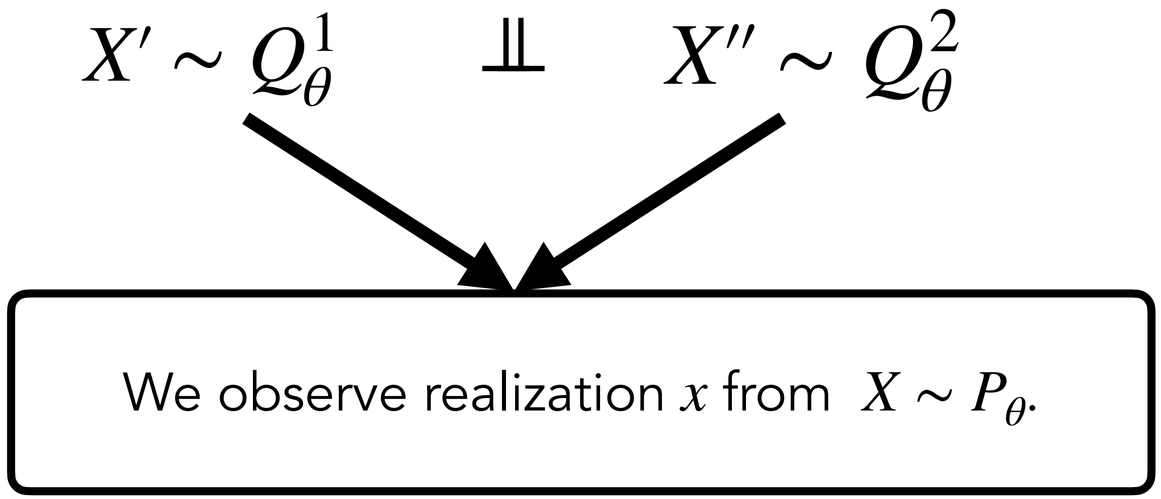
Suppose we know that, if $(X', X'') \sim Q_\theta^1 \times Q_\theta^2$, then $T(X', X'') \sim P_\theta$.
Then we can "unwind" this transformation.

Generalized thinning with non-additive decompositions

We observe realization x from $X \sim P_\theta$.

Generalized thinning with non-additive decompositions

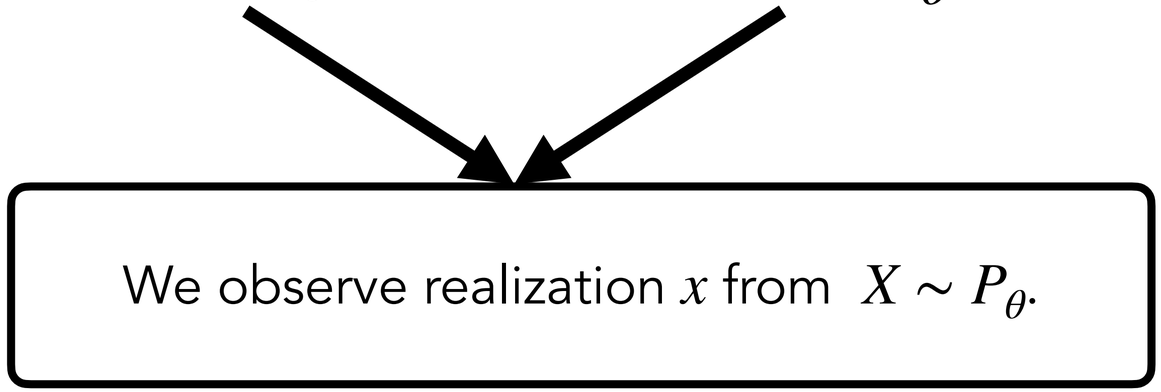
We know x could have arisen as $T(x', x'')$, where



Generalized thinning with non-additive decompositions

We know x could have arisen as $T(x', x'')$, where

$$X' \sim Q_\theta^1 \quad \perp\!\!\!\perp \quad X'' \sim Q_\theta^2$$

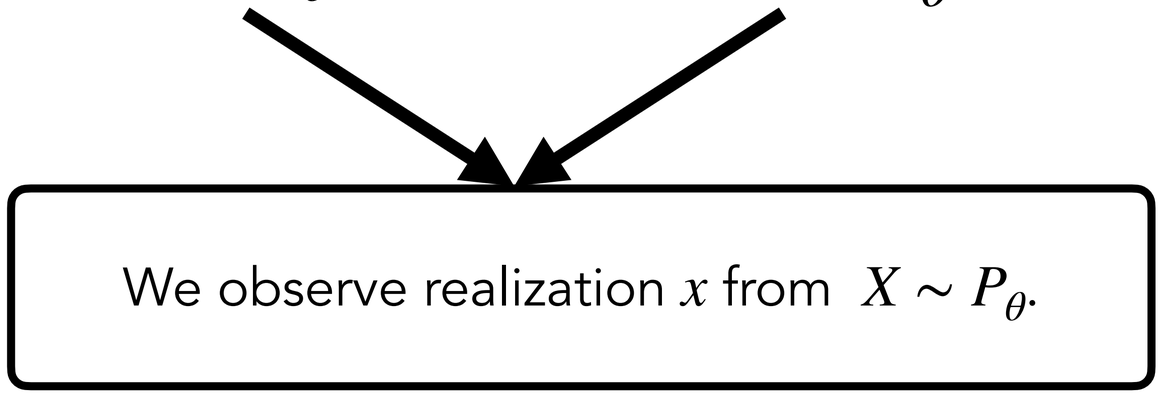


Can we work backwards to recover x' and x''?

Generalized thinning with non-additive decompositions

We know x could have arisen as $T(x', x'')$, where

$$X' \sim Q_\theta^1 \quad \perp\!\!\!\perp \quad X'' \sim Q_\theta^2$$



Can we work backwards to recover x' and x''?

Let $G_{\theta,x}$ be the conditional distribution of $(X', X'') \mid X = x$.

Generalized thinning with non-additive decompositions

We know x could have arisen as $T(x', x'')$, where

$$X' \sim Q_\theta^1 \quad \perp\!\!\!\perp \quad X'' \sim Q_\theta^2$$

We observe realization x from $X \sim P_\theta$.

Algorithm:

Draw $\left(X^{(1)}, X^{(2)}\right) \mid X = x$ from $G_{\theta,x}$.

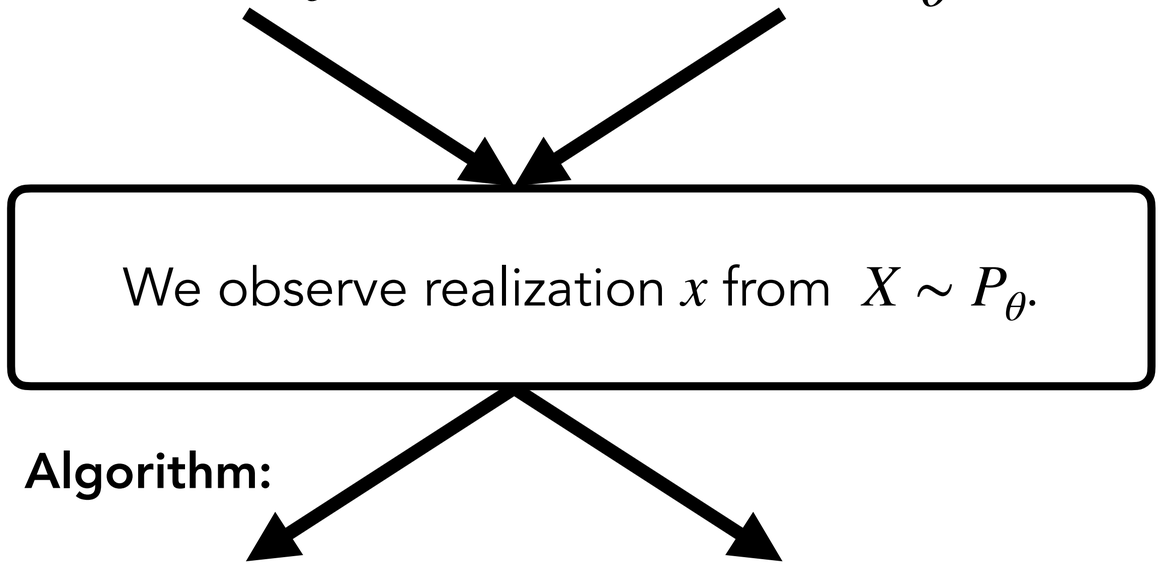
Can we work backwards to recover x' and x''?

Let $G_{\theta,x}$ be the conditional distribution of $(X', X'') \mid X = x$.

Generalized thinning with non-additive decompositions

We know x could have arisen as $T(x', x'')$, where

$$X' \sim Q_\theta^1 \quad \perp\!\!\!\perp \quad X'' \sim Q_\theta^2$$



Theorem:

$X^{(1)} \sim Q_\theta^1, \quad X^{(2)} \sim Q_\theta^2, \quad X^{(1)} \perp\!\!\!\perp X^{(2)}$.

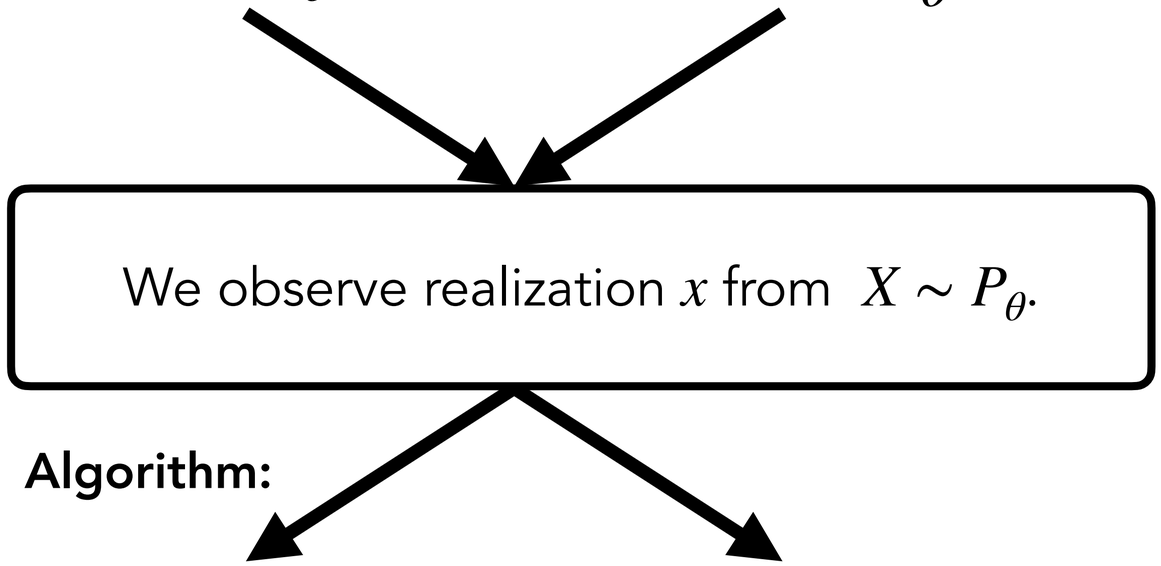
Can we work backwards to recover x' and x''?

Let $G_{\theta,x}$ be the conditional distribution of $(X', X'') \mid X = x$.

Generalized thinning with non-additive decompositions

We know x could have arisen as $T(x', x'')$, where

$$X' \sim Q_\theta^1 \quad \perp\!\!\!\perp \quad X'' \sim Q_\theta^2$$



Algorithm:

Draw $\left(X^{(1)}, X^{(2)}\right) \mid X = x$ from $G_{\theta,x}$.

Theorem:

$X^{(1)} \sim Q_\theta^1, \quad X^{(2)} \sim Q_\theta^2, \quad X^{(1)} \perp\!\!\!\perp X^{(2)}$.

Can we work backwards to recover x' and x''?

Let $G_{\theta,x}$ be the conditional distribution of $(X', X'') \mid X = x$.

Key idea: If $X = T(X', X'')$ is sufficient for θ in the joint of (X', X''), then $G_{\theta,x}$ does not depend on θ.

The list of distributions we can thin is extensive

Family	Distribution P_θ, where $X \sim P_\theta$.	Distribution $Q_\theta^{(k)}$ where $X^{(k)} \overset{ind.}{\sim} Q_\theta^{(k)}$.	Sufficient statistic T (sufficient for θ)
Natural exponential family (in parameter θ)	$N(\theta, \sigma^2)$	$\mathrm{N}(\epsilon_k\theta, \epsilon_k\sigma^2)$	$\sum_{k=1}^K X^{(k)}$
	$\mathrm{Poisson}(\theta)$	$\mathrm{Poisson}(\epsilon_k\theta)$	
	$\mathrm{NegBin}(r, \theta)$	$\mathrm{NegBin}(\epsilon_k r, \theta)$	
	$\mathrm{Binomial}(r, \theta)$	$\mathrm{Binomial}(\epsilon_k r, \theta)$	
	$\mathrm{Gamma}(\alpha, \theta)$	$\mathrm{Gamma}(\epsilon_k\alpha, \theta)$	
	$\mathrm{N}_p(\boldsymbol{\theta}, \Sigma)$	$N_p(\epsilon_k\boldsymbol{\theta}, \epsilon_k\Sigma)$	
	$\mathrm{Multinomial}_p(r, \boldsymbol{\theta})$	$\mathrm{Multinomial}_p(\epsilon_k r, \boldsymbol{\theta})$	
	$\mathrm{Gamma}(K/2, \theta)$	$N(0, \frac{1}{2\theta})$	$\sum_{k=1}^K \left(X^{(k)}\right)^2$
	$\mathrm{Gamma}(K, \theta)$	$\mathrm{Weibull}(\theta^{-\frac{1}{\nu}}, \nu)$	$\sum_{k=1}^K \left(X^{(k)}\right)^\nu$
General exponential family (in parameter θ)	$\mathrm{Beta}(\theta, \beta)$	$\mathrm{Beta}\left(\frac{1}{K}\theta + \frac{k-1}{K}, \frac{1}{K}\beta\right)$	$\left(\Pi_{k=1}^K X^{(k)}\right)^{1/K}$
	$\mathrm{Beta}(\alpha, \theta)$	$\mathrm{Beta}\left(\frac{1}{K}\alpha, \frac{1}{K}\theta + \frac{k-1}{K}\right)$	$\left(\Pi_{k=1}^K \left(1 - X^{(k)}\right)\right)^{1/K}$
	$\mathrm{Gamma}(\theta, \beta)$	$\mathrm{Gamma}(\frac{1}{K}\theta + \frac{k-1}{K}, \frac{1}{K}\beta)$	$\left(\Pi_{k=1}^K X^{(k)}\right)^{1/K}$
	$\mathrm{Weibull}(\theta, \nu)$	$\mathrm{Gamma}(\frac{1}{K}, \theta^{-\nu})$	$\left(\sum_{k=1}^K X^{(k)}\right)^{1/\nu}$
	$\mathrm{Pareto}(\nu, \theta)$	$\mathrm{Gamma}(\frac{1}{K}, \theta)$	$\nu \times \mathrm{Exp}\left(\sum_{k=1}^K X^{(k)}\right)$
	$\mathrm{N}(0,\theta)$	$\mathrm{Gamma}(\frac{1}{2K}, \frac{1}{2\theta})$	$X^2 = \sum_{k=1}^K X^{(k)}$
	$\mathrm{N}_K(\theta_1 1_K, \theta_2 I_K)$	$N(\theta_1, \theta_2)$	sample mean and variance
Truncated support family	$\mathrm{Unif}(0, \theta)$	$\theta \cdot \mathrm{Beta}(\frac{1}{K}, 1)$	$\max\left(X^{(1)}, \ldots, X^{(K)}\right)$
	$\theta \cdot \mathrm{Beta}(\alpha, 1)$	$\theta \cdot \mathrm{Beta}(\frac{\alpha}{K}, 1)$	
	$\theta + \mathrm{Exp}(\lambda)$	$\theta + \mathrm{Exp}(\lambda/\mathrm{K})$	$\min\left(X^{(1)}, \ldots, X^{(K)}\right)$
Non-parametric	F^n	F^{n_k}	$\mathrm{sort}(X^{(1)}, \ldots, X^{(K)})$

The list of distributions we can thin is extensive

Family	Distribution P_θ, where $X \sim P_\theta$.	Distribution $Q_\theta^{(k)}$ where $X^{(k)} \overset{ind.}{\sim} Q_\theta^{(k)}$.	Sufficient statistic T (sufficient for θ)
Natural exponential family (in parameter θ)	$N(\theta, \sigma^2)$	$\mathrm{N}(\epsilon_k\theta, \epsilon_k\sigma^2)$	$\sum_{k=1}^K X^{(k)}$
	$\mathrm{Poisson}(\theta)$	$\mathrm{Poisson}(\epsilon_k\theta)$	
	$\mathrm{NegBin}(r, \theta)$	$\mathrm{NegBin}(\epsilon_k r, \theta)$	
	$\mathrm{Binomial}(r, \theta)$	$\mathrm{Binomial}(\epsilon_k r, \theta)$	
	$\mathrm{Gamma}(\alpha, \theta)$	$\mathrm{Gamma}(\epsilon_k\alpha, \theta)$	
	$\mathrm{N}_p(\boldsymbol{\theta}, \Sigma)$	$N_p(\epsilon_k\boldsymbol{\theta}, \epsilon_k\Sigma)$	
	$\mathrm{Multinomial}_p(r, \boldsymbol{\theta})$	$\mathrm{Multinomial}_p(\epsilon_k r, \boldsymbol{\theta})$	
	$\mathrm{Gamma}(K/2, \theta)$	$N(0, \frac{1}{2\theta})$	$\sum_{k=1}^K \left(X^{(k)}\right)^2$
	$\mathrm{Gamma}(K, \theta)$	$\mathrm{Weibull}(\theta^{-\frac{1}{\nu}}, \nu)$	$\sum_{k=1}^K \left(X^{(k)}\right)^\nu$
General exponential family (in parameter θ)	$\mathrm{Beta}(\theta, \beta)$	$\mathrm{Beta}\left(\frac{1}{K}\theta + \frac{k-1}{K}, \frac{1}{K}\beta\right)$	$\left(\Pi_{k=1}^K X^{(k)}\right)^{1/K}$
	$\mathrm{Beta}(\alpha, \theta)$	$\mathrm{Beta}\left(\frac{1}{K}\alpha, \frac{1}{K}\theta + \frac{k-1}{K}\right)$	$\left(\Pi_{k=1}^K \left(1 - X^{(k)}\right)\right)^{1/K}$
	$\mathrm{Gamma}(\theta, \beta)$	$\mathrm{Gamma}(\frac{1}{K}\theta + \frac{k-1}{K}, \frac{1}{K}\beta)$	$\left(\Pi_{k=1}^K X^{(k)}\right)^{1/K}$
	$\mathrm{Weibull}(\theta, \nu)$	$\mathrm{Gamma}(\frac{1}{K}, \theta^{-\nu})$	$\left(\sum_{k=1}^K X^{(k)}\right)^{1/\nu}$
	$\mathrm{Pareto}(\nu, \theta)$	$\mathrm{Gamma}(\frac{1}{K}, \theta)$	$\nu \times \mathrm{Exp}\left(\sum_{k=1}^K X^{(k)}\right)$
	$\mathrm{N}(0,\theta)$	$\mathrm{Gamma}(\frac{1}{2K}, \frac{1}{2\theta})$	$X^2 = \sum_{k=1}^K X^{(k)}$
	$\mathrm{N}_K(\theta_1 1_K, \theta_2 I_K)$	$N(\theta_1, \theta_2)$	sample mean and variance
Truncated support family	$\mathrm{Unif}(0, \theta)$	$\theta \cdot \mathrm{Beta}(\frac{1}{K}, 1)$	$\max\left(X^{(1)}, \ldots, X^{(K)}\right)$
	$\theta \cdot \mathrm{Beta}(\alpha, 1)$	$\theta \cdot \mathrm{Beta}(\frac{\alpha}{K}, 1)$	
	$\theta + \mathrm{Exp}(\lambda)$	$\theta + \mathrm{Exp}(\lambda/\mathrm{K})$	$\min\left(X^{(1)}, \ldots, X^{(K)}\right)$
Non-parametric	F^n	F^{n_k}	$\mathrm{sort}(X^{(1)}, \ldots, X^{(K)})$

The list of distributions we can thin is extensive

Family	Distribution P_θ, where $X \sim P_\theta$.	Distribution $Q_\theta^{(k)}$ where $X^{(k)} \overset{ind.}{\sim} Q_\theta^{(k)}$.	Sufficient statistic T (sufficient for θ)
Natural exponential family (in parameter θ)	$N(\theta, \sigma^2)$	$\mathrm{N}(\epsilon_k\theta, \epsilon_k\sigma^2)$	$\sum_{k=1}^K X^{(k)}$
	$\mathrm{Poisson}(\theta)$	$\mathrm{Poisson}(\epsilon_k\theta)$	
	$\mathrm{NegBin}(r, \theta)$	$\mathrm{NegBin}(\epsilon_k r, \theta)$	
	$\mathrm{Binomial}(r, \theta)$	$\mathrm{Binomial}(\epsilon_k r, \theta)$	
	$\mathrm{Gamma}(\alpha, \theta)$	$\mathrm{Gamma}(\epsilon_k\alpha, \theta)$	
	$\mathrm{N}_p(\boldsymbol{\theta}, \Sigma)$	$N_p(\epsilon_k\boldsymbol{\theta}, \epsilon_k\Sigma)$	
	$\mathrm{Multinomial}_p(r, \boldsymbol{\theta})$	$\mathrm{Multinomial}_p(\epsilon_k r, \boldsymbol{\theta})$	
	$\mathrm{Gamma}(K/2, \theta)$	$N(0, \frac{1}{2\theta})$	$\sum_{k=1}^K \left(X^{(k)}\right)^2$
	$\mathrm{Gamma}(K, \theta)$	$\mathrm{Weibull}(\theta^{-\frac{1}{\nu}}, \nu)$	$\sum_{k=1}^K \left(X^{(k)}\right)^\nu$
General exponential family (in parameter θ)	$\mathrm{Beta}(\theta, \beta)$	$\mathrm{Beta}\left(\frac{1}{K}\theta + \frac{k-1}{K}, \frac{1}{K}\beta\right)$	$\left(\Pi_{k=1}^K X^{(k)}\right)^{1/K}$
	$\mathrm{Beta}(\alpha, \theta)$	$\mathrm{Beta}\left(\frac{1}{K}\alpha, \frac{1}{K}\theta + \frac{k-1}{K}\right)$	$\left(\Pi_{k=1}^K \left(1 - X^{(k)}\right)\right)^{1/K}$
	$\mathrm{Gamma}(\theta, \beta)$	$\mathrm{Gamma}(\frac{1}{K}\theta + \frac{k-1}{K}, \frac{1}{K}\beta)$	$\left(\Pi_{k=1}^K X^{(k)}\right)^{1/K}$
	$\mathrm{Weibull}(\theta, \nu)$	$\mathrm{Gamma}(\frac{1}{K}, \theta^{-\nu})$	$\left(\sum_{k=1}^K X^{(k)}\right)^{1/\nu}$
	$\mathrm{Pareto}(\nu, \theta)$	$\mathrm{Gamma}(\frac{1}{K}, \theta)$	$\nu \times \mathrm{Exp}\left(\sum_{k=1}^K X^{(k)}\right)$
	$\mathrm{N}(0,\theta)$	$\mathrm{Gamma}(\frac{1}{2K}, \frac{1}{2\theta})$	$X^2 = \sum_{k=1}^K X^{(k)}$
	$\mathrm{N}_K(\theta_1 1_K, \theta_2 I_K)$	$N(\theta_1, \theta_2)$	sample mean and variance
Truncated support family	$\mathrm{Unif}(0, \theta)$	$\theta \cdot \mathrm{Beta}(\frac{1}{K}, 1)$	$\max\left(X^{(1)}, \ldots, X^{(K)}\right)$
	$\theta \cdot \mathrm{Beta}(\alpha, 1)$	$\theta \cdot \mathrm{Beta}(\frac{\alpha}{K}, 1)$	
	$\theta + \mathrm{Exp}(\lambda)$	$\theta + \mathrm{Exp}(\lambda/\mathrm{K})$	$\min\left(X^{(1)}, \ldots, X^{(K)}\right)$
Non-parametric	F^n	F^{n_k}	$\mathrm{sort}(X^{(1)}, \ldots, X^{(K)})$

Statistics > Methodology

[Submitted on 22 Mar 2023]

Generalized Data Thinning Using Sufficient Statistics

Ameer Dharamshi, Anna Neufeld, Keshav Motwani, Lucy L. Gao, Daniela Witten, Jacob Bien

Outline

1. Motivation: sample splitting doesn’t always work
2. Poisson thinning
3. Data thinning
4. Generalized data thinning
5. **Application to changepoint validation**
6. Ongoing work

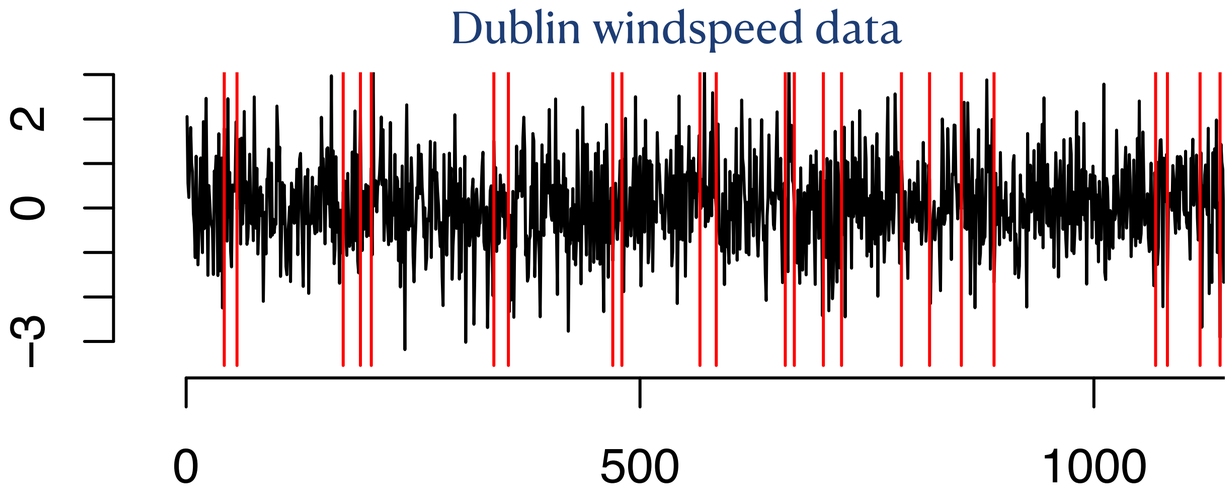
Application: changepoint detection

Application: changepoint detection

Goal: Identify points in a sequence where the distribution changes.

Application: changepoint detection

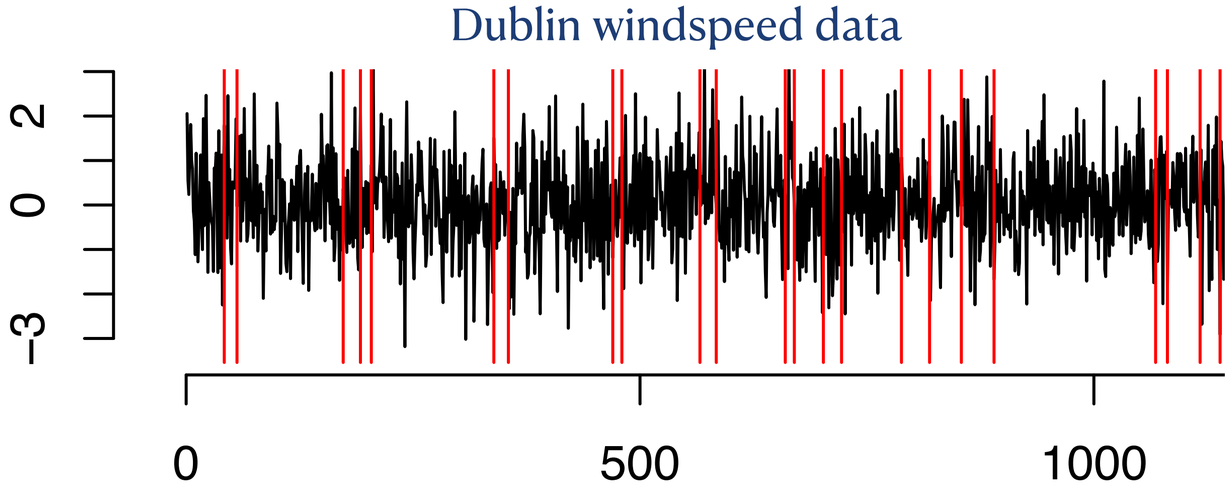
Goal: Identify points in a sequence where the distribution changes.



Application: changepoint detection

Goal: Identify points in a sequence where the distribution changes.

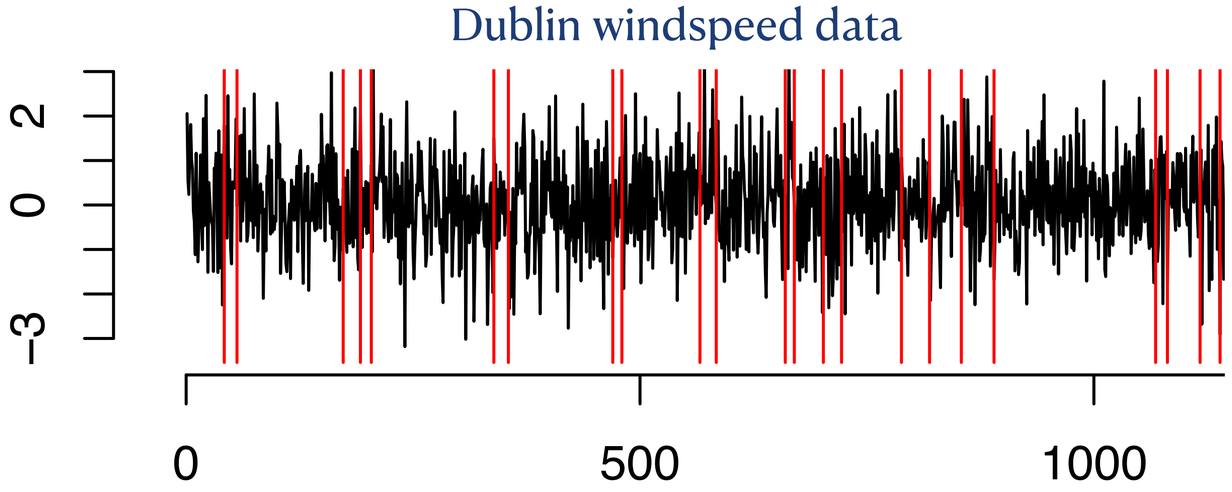
Setting: Differences in wind speeds have previously been modeled as $X_i \sim N(0, \theta_i)$. How can we validate an estimated changepoint?



Application: changepoint detection

Goal: Identify points in a sequence where the distribution changes.

Setting: Differences in wind speeds have previously been modeled as $X_i \sim N(0, \theta_i)$. How can we validate an estimated changepoint?



Can we thin a Gaussian with unknown variance?

We can't apply this because the variance is unknown

Family	Distribution P_θ, where $X \sim P_\theta$.	Distribution $Q_\theta^{(k)}$ where $X^{(k)} \overset{ind.}{\sim} Q_\theta^{(k)}$.	Sufficient statistic T (sufficient for θ)
Natural exponential family (in parameter θ)	$N(\theta, \sigma^2)$	$\mathrm{N}(\epsilon_k\theta, \epsilon_k\sigma^2)$	$\sum_{k=1}^K X^{(k)}$
	$\mathrm{Poisson}(\theta)$	$\mathrm{Poisson}(\epsilon_k\theta)$	
	$\mathrm{NegBin}(r, \theta)$	$\mathrm{NegBin}(\epsilon_k r, \theta)$	
	$\mathrm{Binomial}(r, \theta)$	$\mathrm{Binomial}(\epsilon_k r, \theta)$	
	$\mathrm{Gamma}(\alpha, \theta)$	$\mathrm{Gamma}(\epsilon_k\alpha, \theta)$	
	$\mathrm{N}_p(\boldsymbol{\theta}, \Sigma)$	$N_p(\epsilon_k\boldsymbol{\theta}, \epsilon_k\Sigma)$	
	$\mathrm{Multinomial}_p(r, \boldsymbol{\theta})$	$\mathrm{Multinomial}_p(\epsilon_k r, \boldsymbol{\theta})$	
	$\mathrm{Gamma}(K/2, \theta)$	$N(0, \frac{1}{2\theta})$	$\sum_{k=1}^K \left(X^{(k)}\right)^2$
	$\mathrm{Gamma}(K, \theta)$	$\mathrm{Weibull}(\theta^{-\frac{1}{\nu}}, \nu)$	$\sum_{k=1}^K \left(X^{(k)}\right)^\nu$
General exponential family (in parameter θ)	$\mathrm{Beta}(\theta, \beta)$	$\mathrm{Beta}\left(\frac{1}{K}\theta + \frac{k-1}{K}, \frac{1}{K}\beta\right)$	$\left(\Pi_{k=1}^K X^{(k)}\right)^{1/K}$
	$\mathrm{Beta}(\alpha, \theta)$	$\mathrm{Beta}\left(\frac{1}{K}\alpha, \frac{1}{K}\theta + \frac{k-1}{K}\right)$	$\left(\Pi_{k=1}^K \left(1 - X^{(k)}\right)\right)^{1/K}$
	$\mathrm{Gamma}(\theta, \beta)$	$\mathrm{Gamma}(\frac{1}{K}\theta + \frac{k-1}{K}, \frac{1}{K}\beta)$	$\left(\Pi_{k=1}^K X^{(k)}\right)^{1/K}$
	$\mathrm{Weibull}(\theta, \nu)$	$\mathrm{Gamma}(\frac{1}{K}, \theta^{-\nu})$	$\left(\sum_{k=1}^K X^{(k)}\right)^{1/\nu}$
	$\mathrm{Pareto}(\nu, \theta)$	$\mathrm{Gamma}(\frac{1}{K}, \theta)$	$\nu \times \mathrm{Exp}\left(\sum_{k=1}^K X^{(k)}\right)$
	$\mathrm{N}(0,\theta)$	$\mathrm{Gamma}(\frac{1}{2K}, \frac{1}{2\theta})$	$X^2 = \sum_{k=1}^K X^{(k)}$
	$\mathrm{N}_K(\theta_1 1_K, \theta_2 I_K)$	$N(\theta_1, \theta_2)$	sample mean and variance
Truncated support family	$\mathrm{Unif}(0, \theta)$	$\theta \cdot \mathrm{Beta}(\frac{1}{K}, 1)$	$\max\left(X^{(1)}, \ldots, X^{(K)}\right)$
	$\theta \cdot \mathrm{Beta}(\alpha, 1)$	$\theta \cdot \mathrm{Beta}(\frac{\alpha}{K}, 1)$	
	$\theta + \mathrm{Exp}(\lambda)$	$\theta + \mathrm{Exp}(\lambda/\mathrm{K})$	$\min\left(X^{(1)}, \ldots, X^{(K)}\right)$
Non-parametric	F^n	F^{n_k}	$\mathrm{sort}(X^{(1)}, \ldots, X^{(K)})$

But, we do know how to thin a Gamma!

Family	Distribution P_θ, where $X \sim P_\theta$.	Distribution $Q_\theta^{(k)}$ where $X^{(k)} \overset{ind.}{\sim} Q_\theta^{(k)}$.	Sufficient statistic T (sufficient for θ)
Natural exponential family (in parameter θ)	$N(\theta, \sigma^2)$	$\mathrm{N}(\epsilon_k\theta, \epsilon_k\sigma^2)$	$\sum_{k=1}^K X^{(k)}$
	$\mathrm{Poisson}(\theta)$	$\mathrm{Poisson}(\epsilon_k\theta)$	
	$\mathrm{NegBin}(r, \theta)$	$\mathrm{NegBin}(\epsilon_k r, \theta)$	
	$\mathrm{Binomial}(r, \theta)$	$\mathrm{Binomial}(\epsilon_k r, \theta)$	
	$\mathrm{Gamma}(\alpha, \theta)$	$\mathrm{Gamma}(\epsilon_k\alpha, \theta)$	
	$\mathrm{N}_p(\boldsymbol{\theta}, \Sigma)$	$N_p(\epsilon_k\boldsymbol{\theta}, \epsilon_k\Sigma)$	
	$\mathrm{Multinomial}_p(r, \boldsymbol{\theta})$	$\mathrm{Multinomial}_p(\epsilon_k r, \boldsymbol{\theta})$	
	$\mathrm{Gamma}(K/2, \theta)$	$N(0, \frac{1}{2\theta})$	$\sum_{k=1}^K \left(X^{(k)}\right)^2$
	$\mathrm{Gamma}(K, \theta)$	$\mathrm{Weibull}(\theta^{-\frac{1}{\nu}}, \nu)$	$\sum_{k=1}^K \left(X^{(k)}\right)^\nu$
General exponential family (in parameter θ)	$\mathrm{Beta}(\theta, \beta)$	$\mathrm{Beta}\left(\frac{1}{K}\theta + \frac{k-1}{K}, \frac{1}{K}\beta\right)$	$\left(\Pi_{k=1}^K X^{(k)}\right)^{1/K}$
	$\mathrm{Beta}(\alpha, \theta)$	$\mathrm{Beta}\left(\frac{1}{K}\alpha, \frac{1}{K}\theta + \frac{k-1}{K}\right)$	$\left(\Pi_{k=1}^K \left(1 - X^{(k)}\right)\right)^{1/K}$
	$\mathrm{Gamma}(\theta, \beta)$	$\mathrm{Gamma}(\frac{1}{K}\theta + \frac{k-1}{K}, \frac{1}{K}\beta)$	$\left(\Pi_{k=1}^K X^{(k)}\right)^{1/K}$
	$\mathrm{Weibull}(\theta, \nu)$	$\mathrm{Gamma}(\frac{1}{K}, \theta^{-\nu})$	$\left(\sum_{k=1}^K X^{(k)}\right)^{1/\nu}$
	$\mathrm{Pareto}(\nu, \theta)$	$\mathrm{Gamma}(\frac{1}{K}, \theta)$	$\nu \times \mathrm{Exp}\left(\sum_{k=1}^K X^{(k)}\right)$
	$\mathrm{N}(0,\theta)$	$\mathrm{Gamma}(\frac{1}{2K}, \frac{1}{2\theta})$	$X^2 = \sum_{k=1}^K X^{(k)}$
	$\mathrm{N}_K(\theta_1 1_K, \theta_2 I_K)$	$N(\theta_1, \theta_2)$	sample mean and variance
Truncated support family	$\mathrm{Unif}(0, \theta)$	$\theta \cdot \mathrm{Beta}(\frac{1}{K}, 1)$	$\max\left(X^{(1)}, \ldots, X^{(K)}\right)$
	$\theta \cdot \mathrm{Beta}(\alpha, 1)$	$\theta \cdot \mathrm{Beta}(\frac{\alpha}{K}, 1)$	
	$\theta + \mathrm{Exp}(\lambda)$	$\theta + \mathrm{Exp}(\lambda/\mathrm{K})$	$\min\left(X^{(1)}, \ldots, X^{(K)}\right)$
Non-parametric	F^n	F^{n_k}	$\mathrm{sort}(X^{(1)}, \ldots, X^{(K)})$

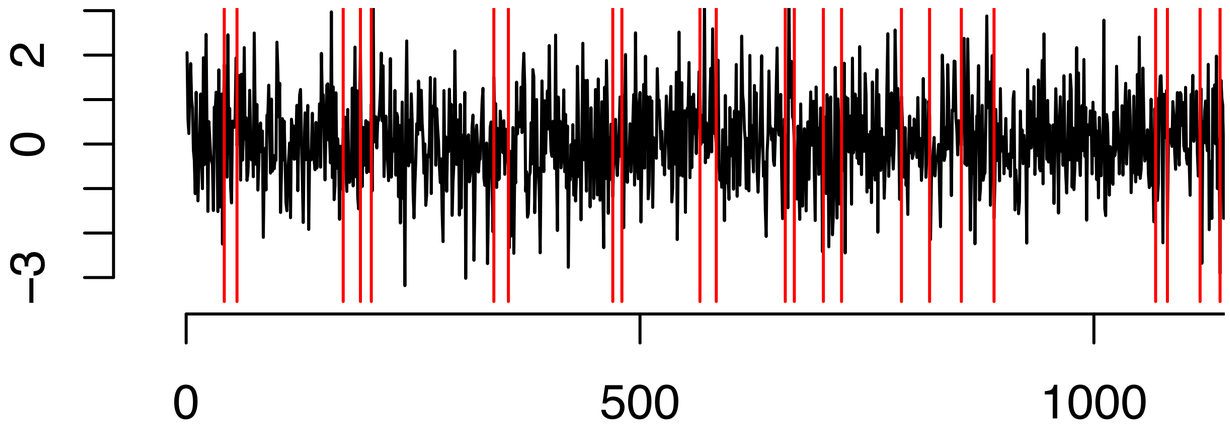
Fact: if $X_i \sim N(0,\theta_i)$, then

$$X_i^2 \sim \mathrm{Gamma}(\frac{1}{2}, \frac{1}{2\theta_i^2}).$$

Application: changepoint detection

Goal: Identify points in a sequence where the distribution changes.

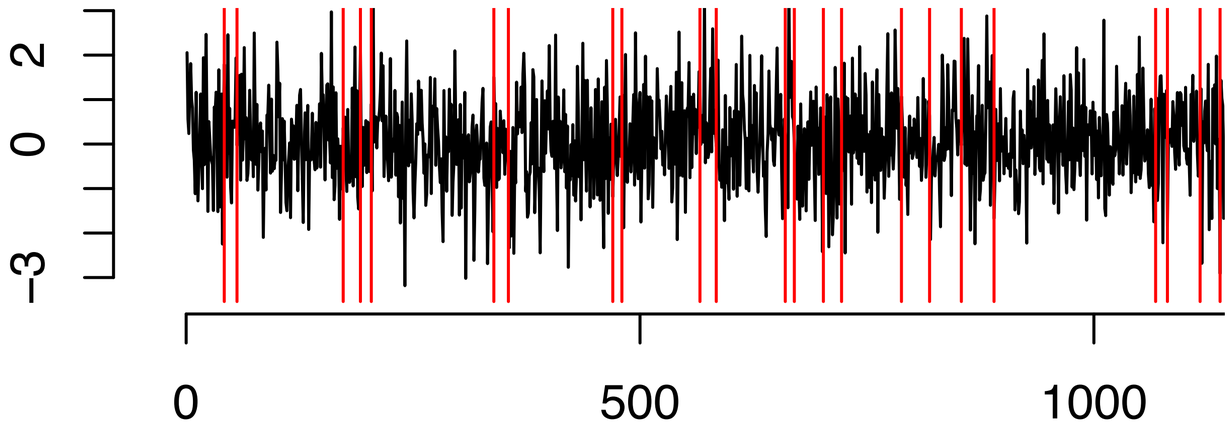
Setting: Differences in wind speeds have previously been modeled as $X_i \sim N(0, \theta_i)$. How can we validate an estimated changepoint?



Application: changepoint detection

Goal: Identify points in a sequence where the distribution changes.

Setting: Differences in wind speeds have previously been modeled as $X_i \sim N(0, \theta_i)$. How can we validate an estimated changepoint?

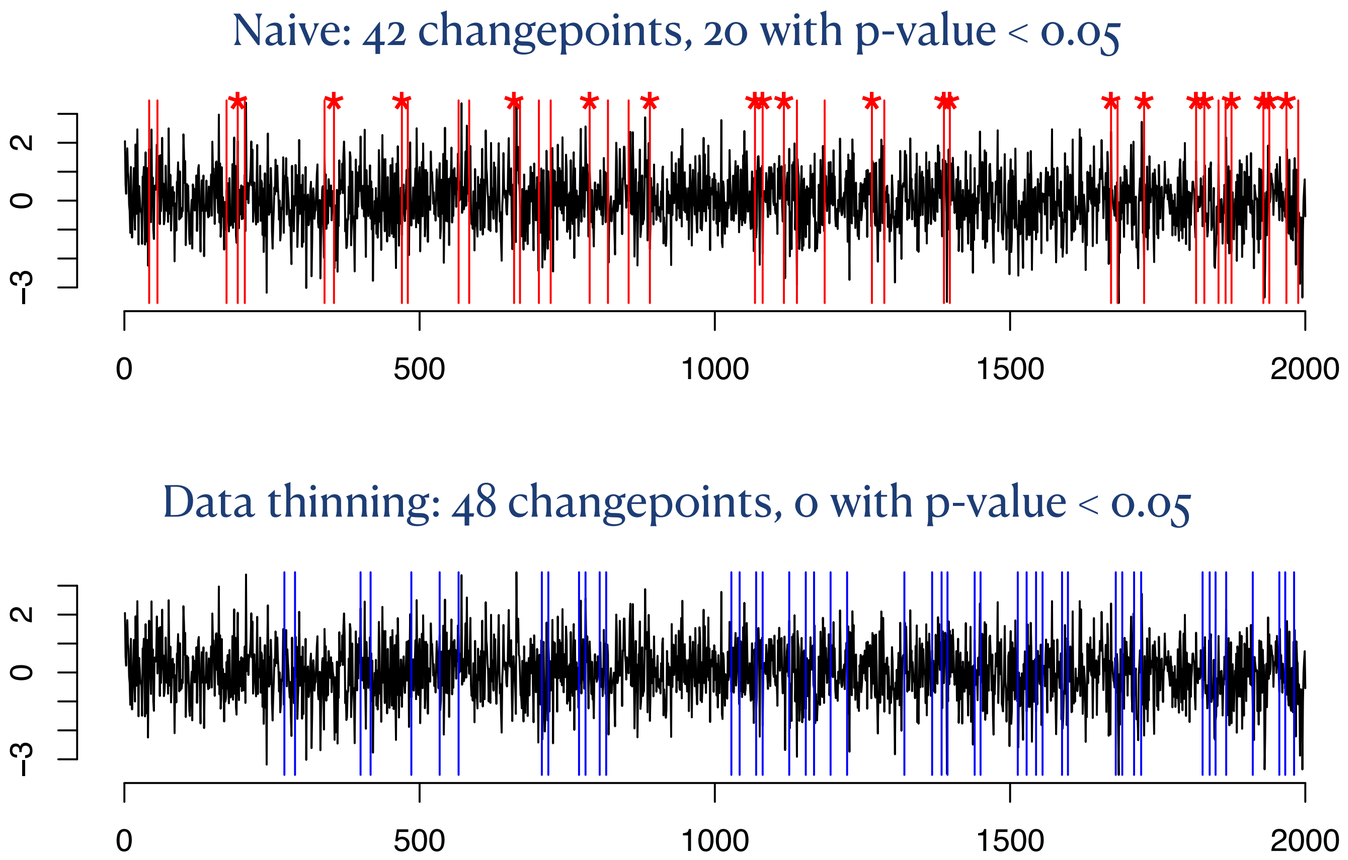


Define $Z_i := X_i^2$. Note that $Z_i \sim \text{Gamma}\left(0.5, 0.5/\theta_i^2\right)$.

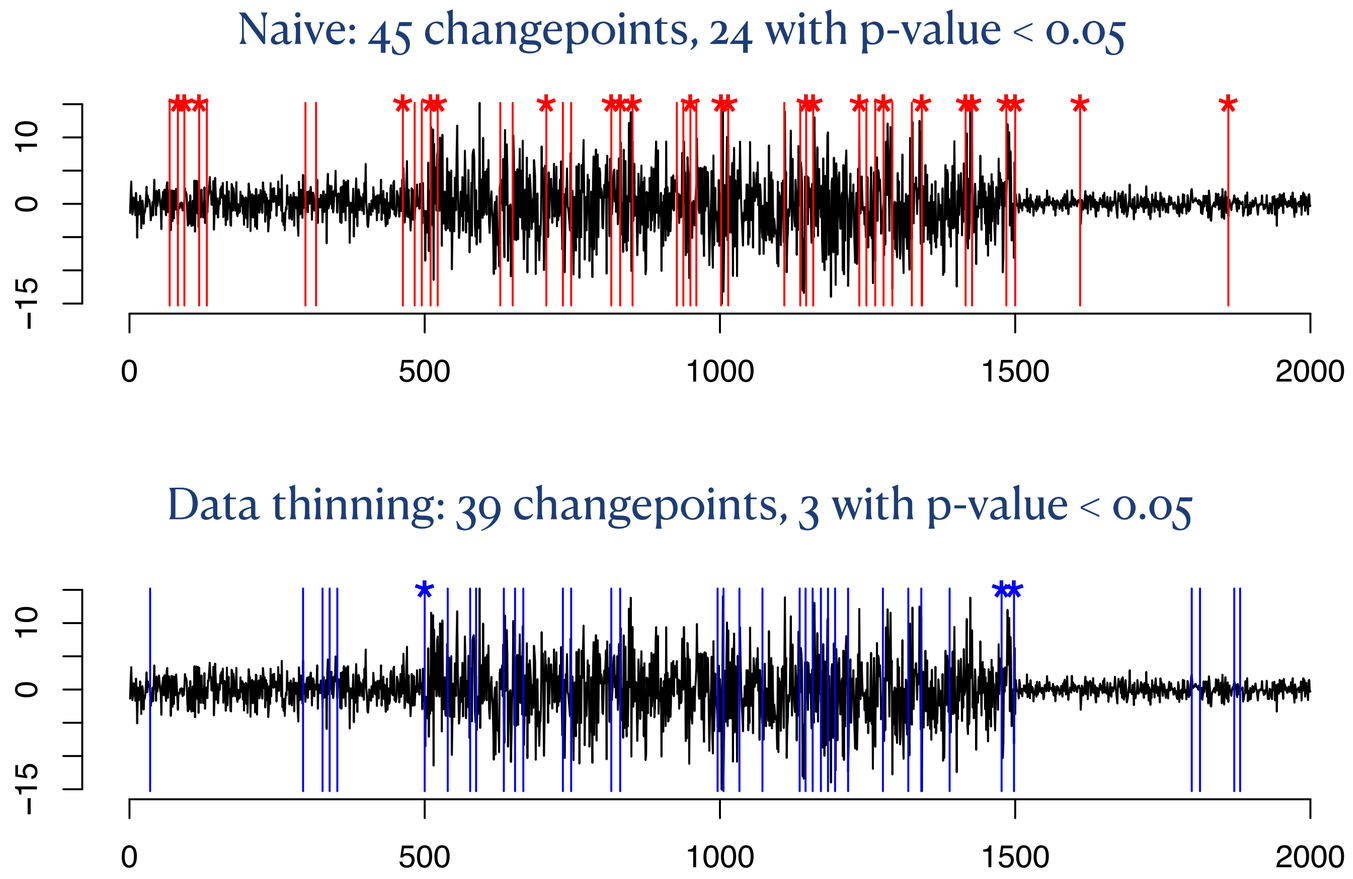
Option 1: Estimate and test changepoints on Z_i.

Option 2: Thin Z_i into $Z_i^{(1)} \sim \text{Gamma}\left(0.25, 0.5/\theta_i^2\right)$ and $Z_i^{(2)} \sim \text{Gamma}\left(0.25, 0.5/\theta_i^2\right)$.
Identify changepoints using $Z_i^{(1)}$, and test using $Z_i^{(2)}$.

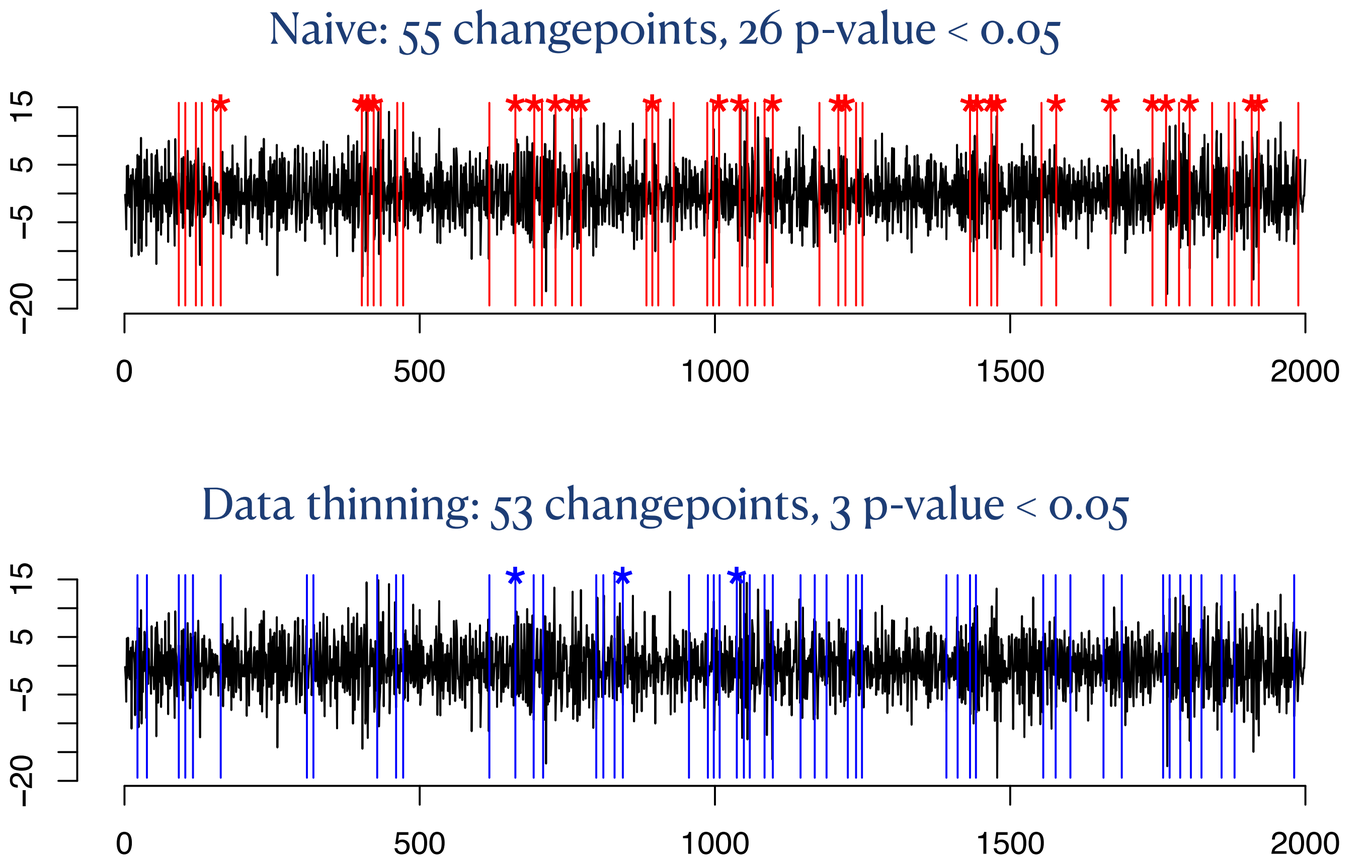
Quick detour to simulated data with no changepoints



Quick detour to simulated data with two true changepoints



Dublin wind speed data

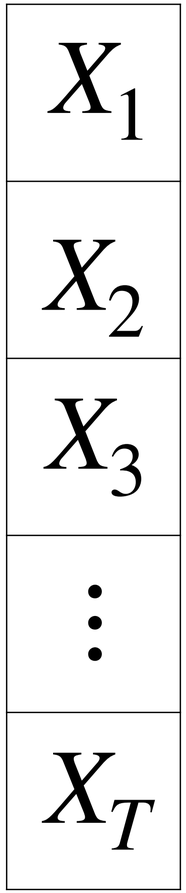


Outline

1. Motivation: sample splitting doesn’t always work
2. Poisson thinning
3. Data thinning
4. Generalized data thinning
5. Application to changepoint validation
6. **Ongoing work**

What if we want to thin time series data?

$$X_t \sim N(\mu_t, \sigma^2)$$

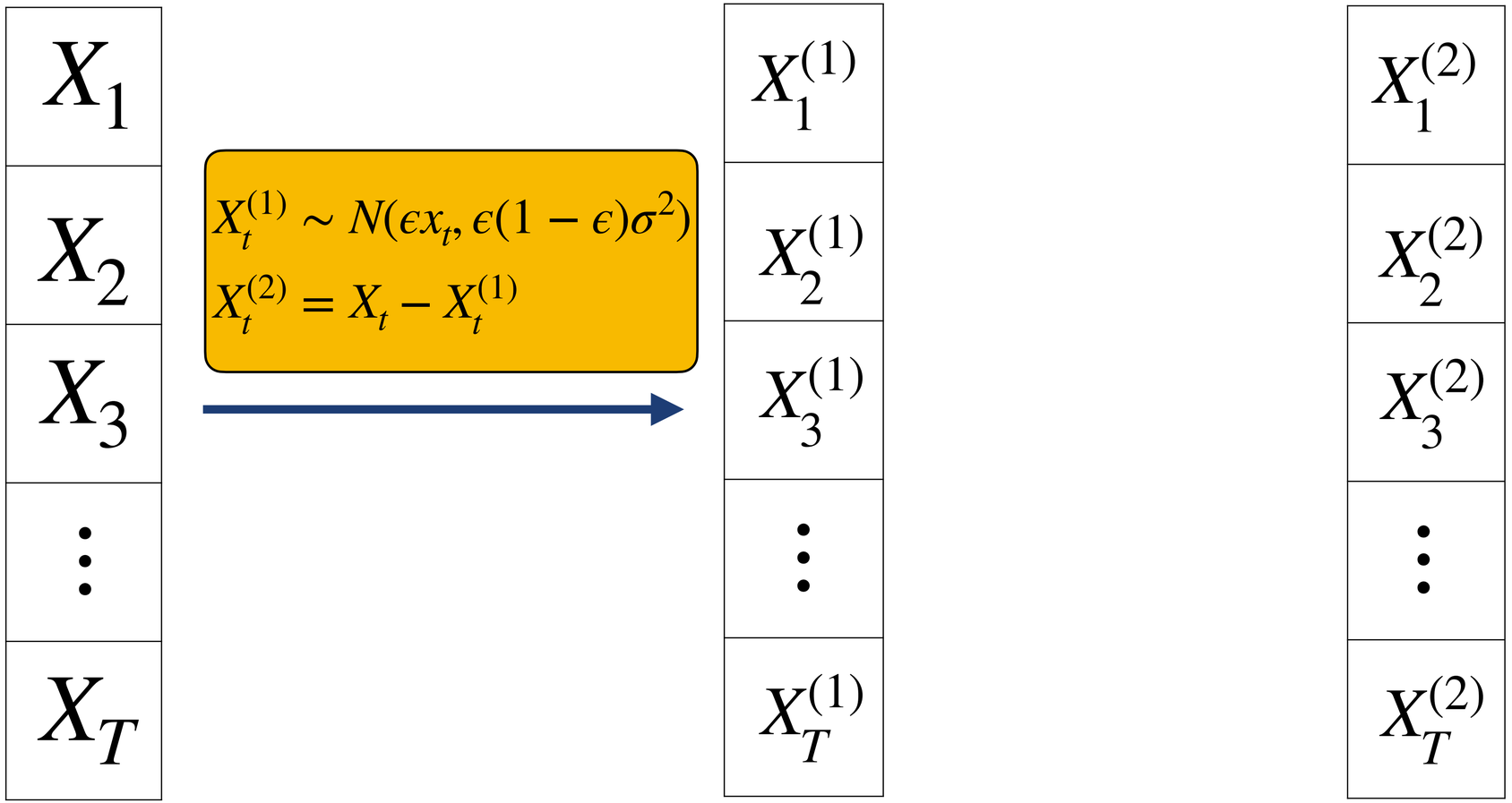


What if we want to thin time series data?

$X_t \sim N(\mu_t, \sigma^2)$

$X_t^{(1)} \sim N(\epsilon\mu_t, \epsilon\sigma^2)$

$X_t^{(2)} \sim N((1-\epsilon)\mu_t, (1-\epsilon)\sigma^2)$

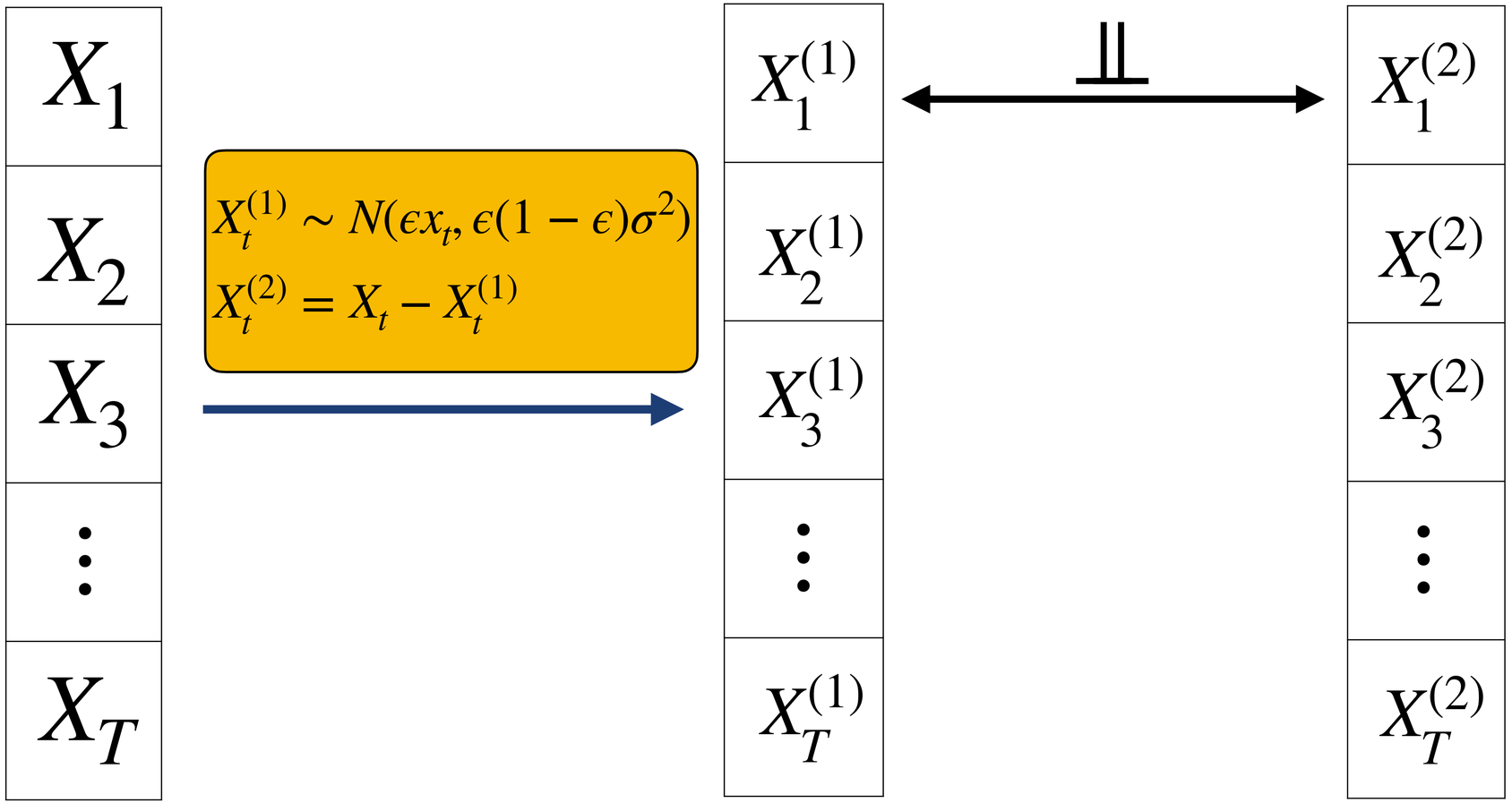


What if we want to thin time series data?

$X_t \sim N(\mu_t, \sigma^2)$

$X_t^{(1)} \sim N(\epsilon\mu_t, \epsilon\sigma^2)$

$X_t^{(2)} \sim N((1-\epsilon)\mu_t, (1-\epsilon)\sigma^2)$

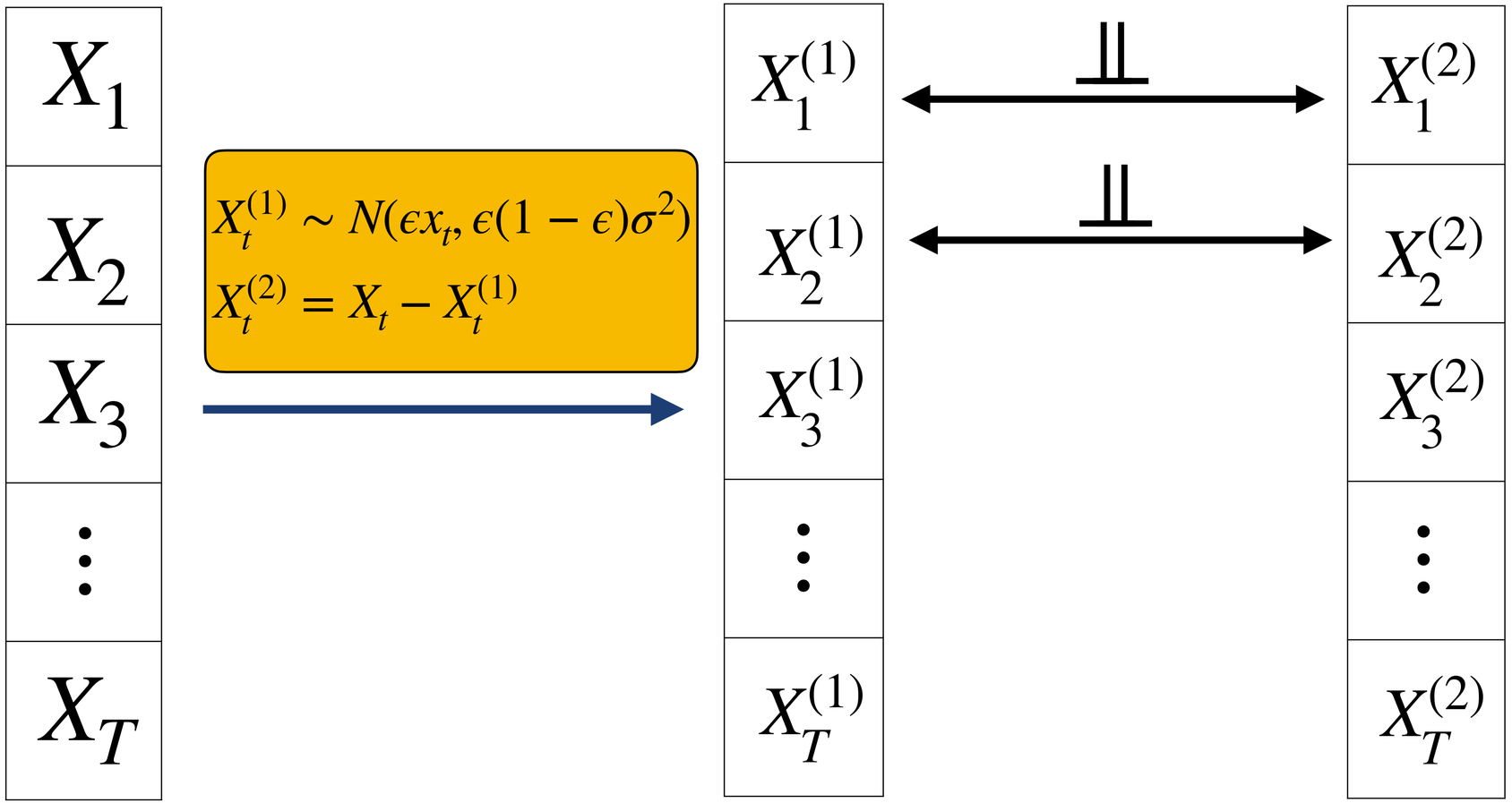


What if we want to thin time series data?

$X_t \sim N(\mu_t, \sigma^2)$

$X_t^{(1)} \sim N(\epsilon\mu_t, \epsilon\sigma^2)$

$X_t^{(2)} \sim N((1-\epsilon)\mu_t, (1-\epsilon)\sigma^2)$

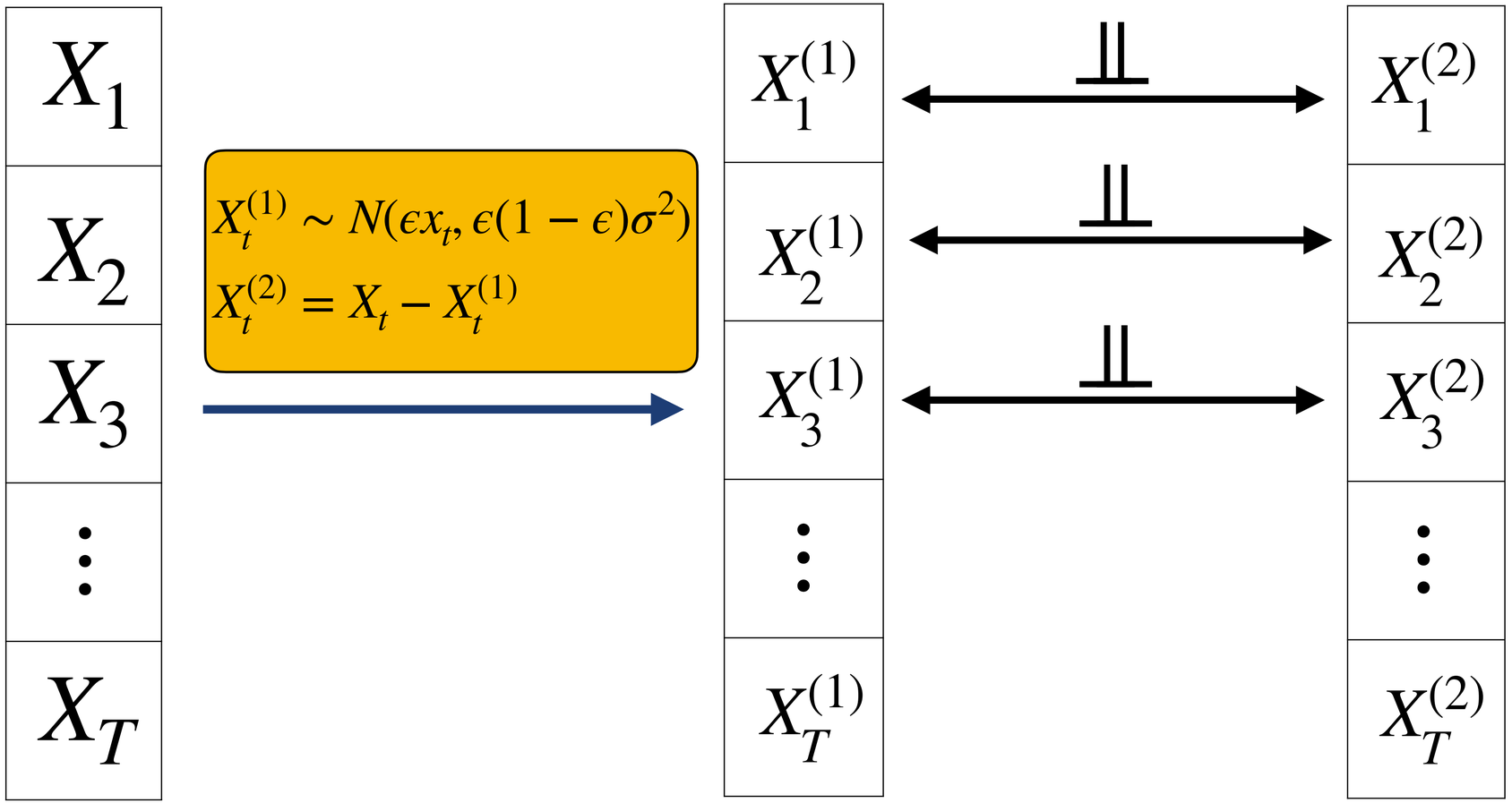


What if we want to thin time series data?

$X_t \sim N(\mu_t, \sigma^2)$

$X_t^{(1)} \sim N(\epsilon\mu_t, \epsilon\sigma^2)$

$X_t^{(2)} \sim N((1-\epsilon)\mu_t, (1-\epsilon)\sigma^2)$

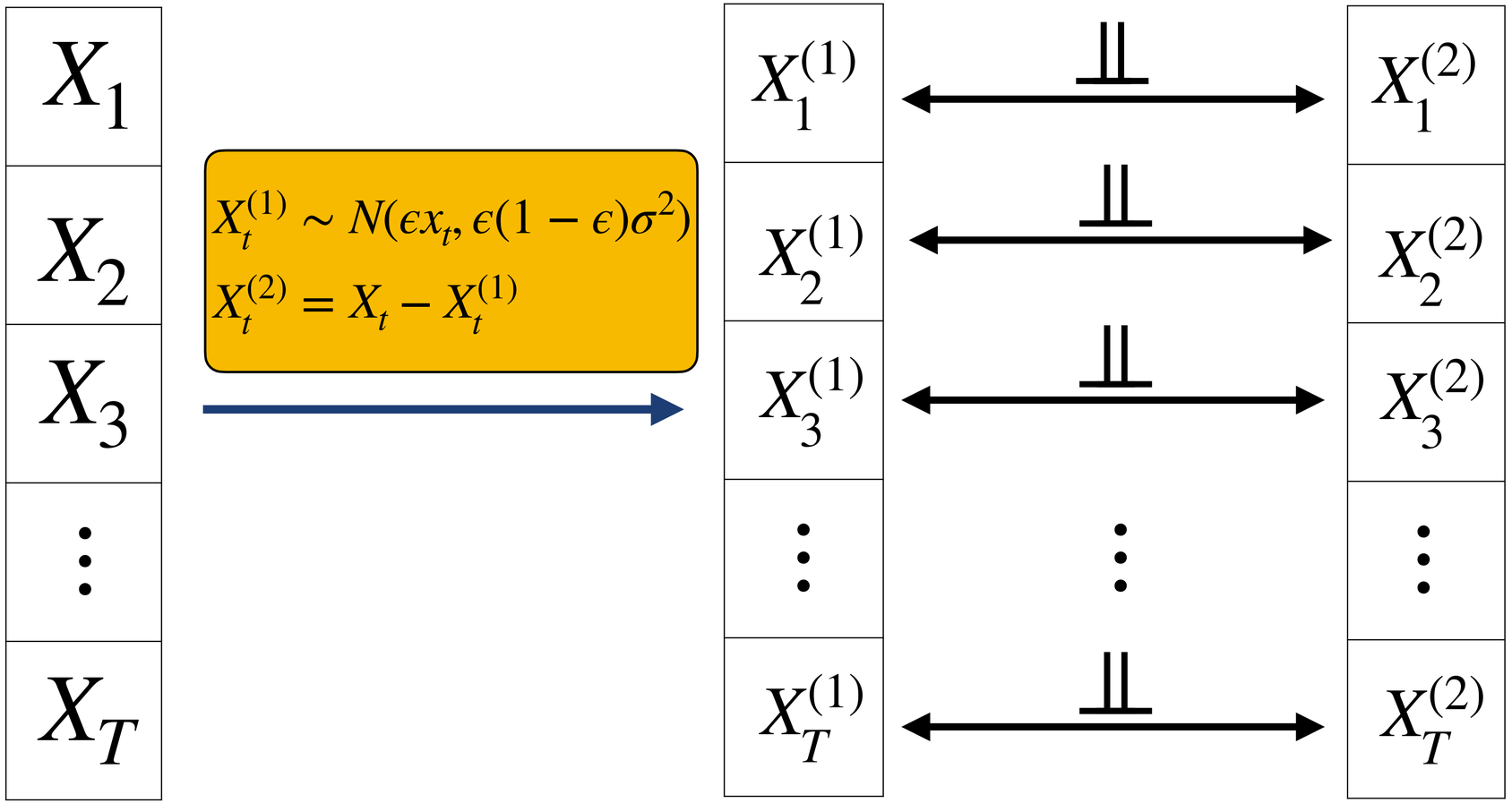


What if we want to thin time series data?

$X_t \sim N(\mu_t, \sigma^2)$

$X_t^{(1)} \sim N(\epsilon\mu_t, \epsilon\sigma^2)$

$X_t^{(2)} \sim N((1-\epsilon)\mu_t, (1-\epsilon)\sigma^2)$

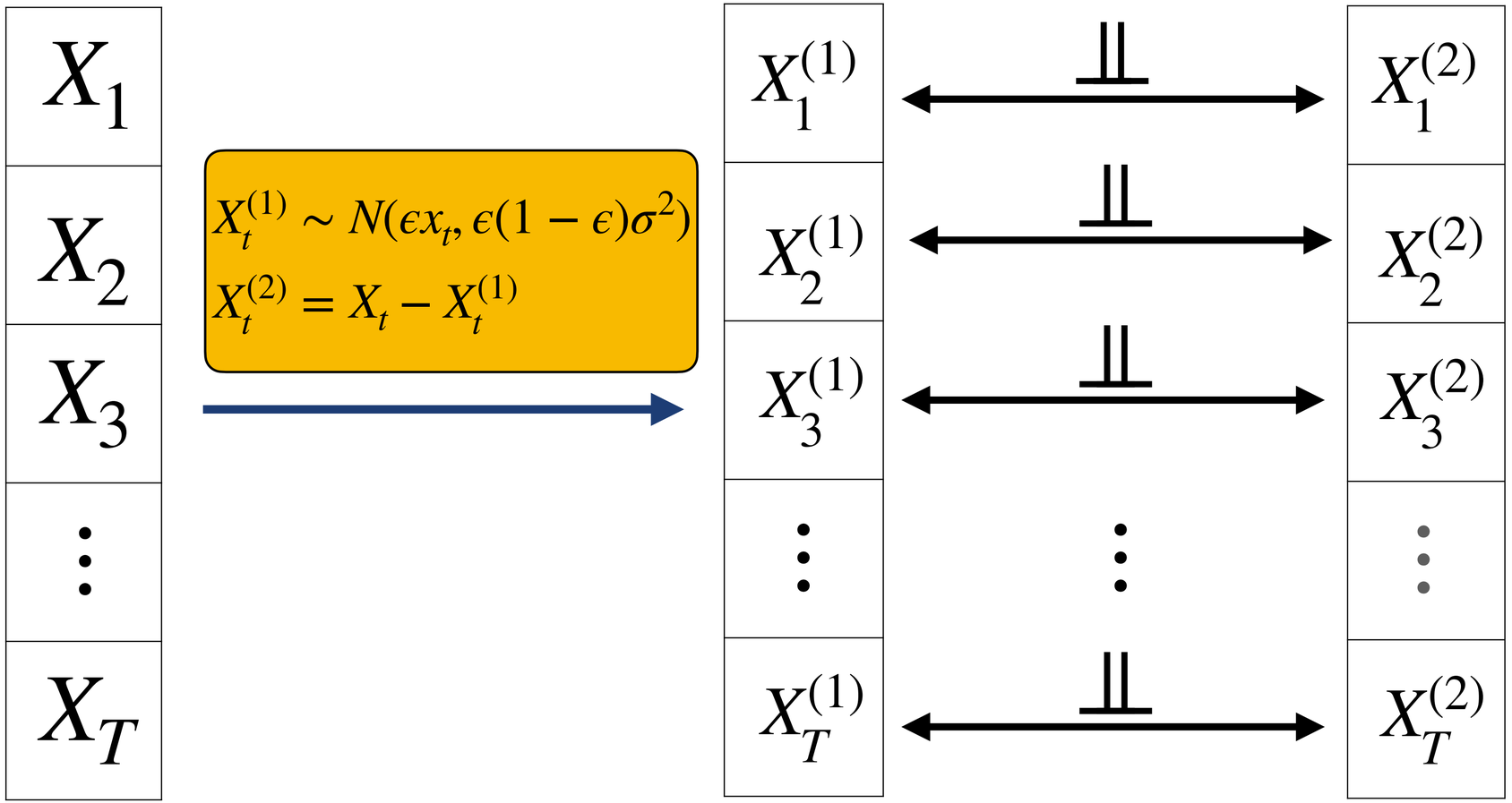


What if we want to thin time series data?

$X_t \sim N(\mu_t, \sigma^2)$

$X_t^{(1)} \sim N(\epsilon\mu_t, \epsilon\sigma^2)$

$X_t^{(2)} \sim N((1-\epsilon)\mu_t, (1-\epsilon)\sigma^2)$

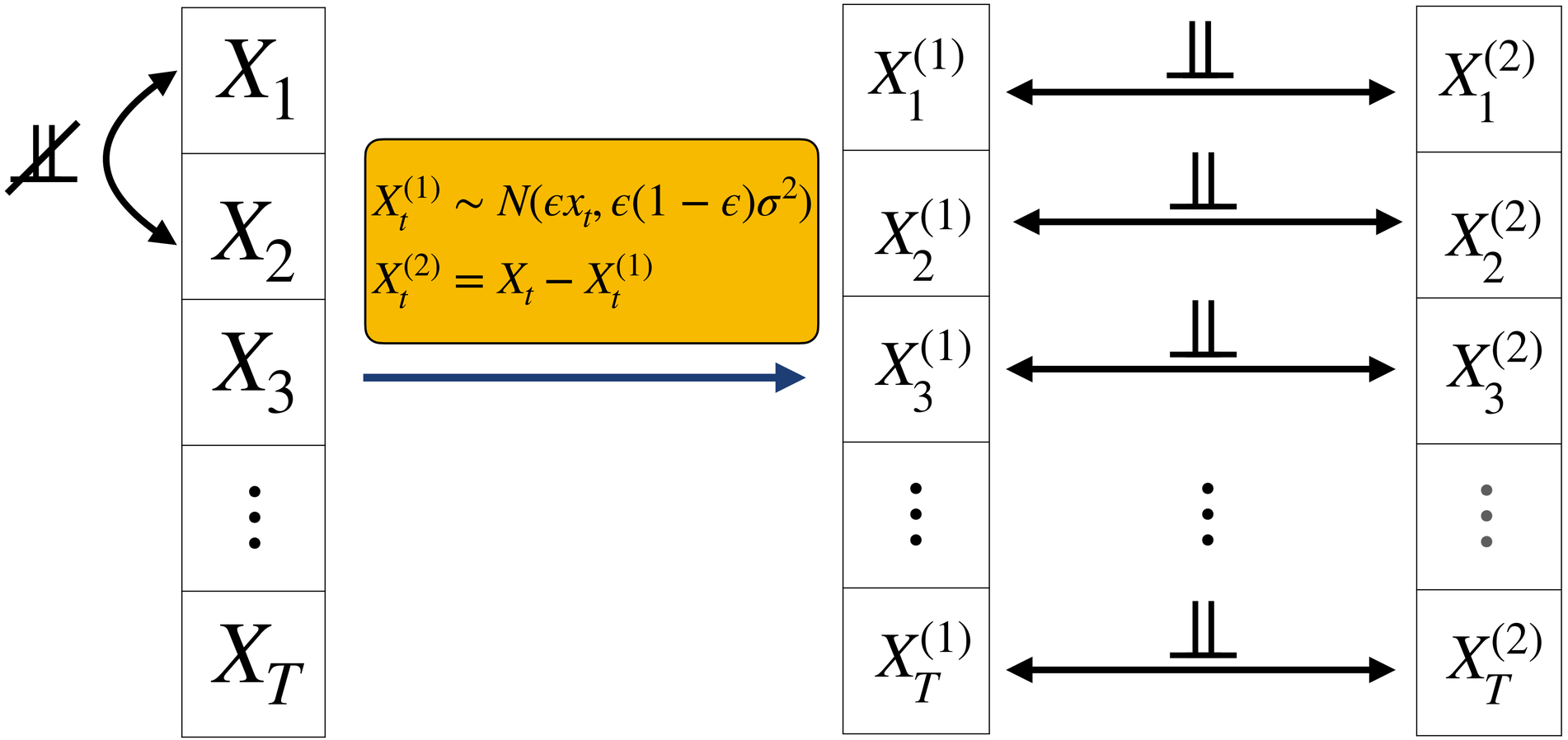


What if we want to thin time series data?

$X_t \sim N(\mu_t, \sigma^2)$

$X_t^{(1)} \sim N(\epsilon\mu_t, \epsilon\sigma^2)$

$X_t^{(2)} \sim N((1-\epsilon)\mu_t, (1-\epsilon)\sigma^2)$

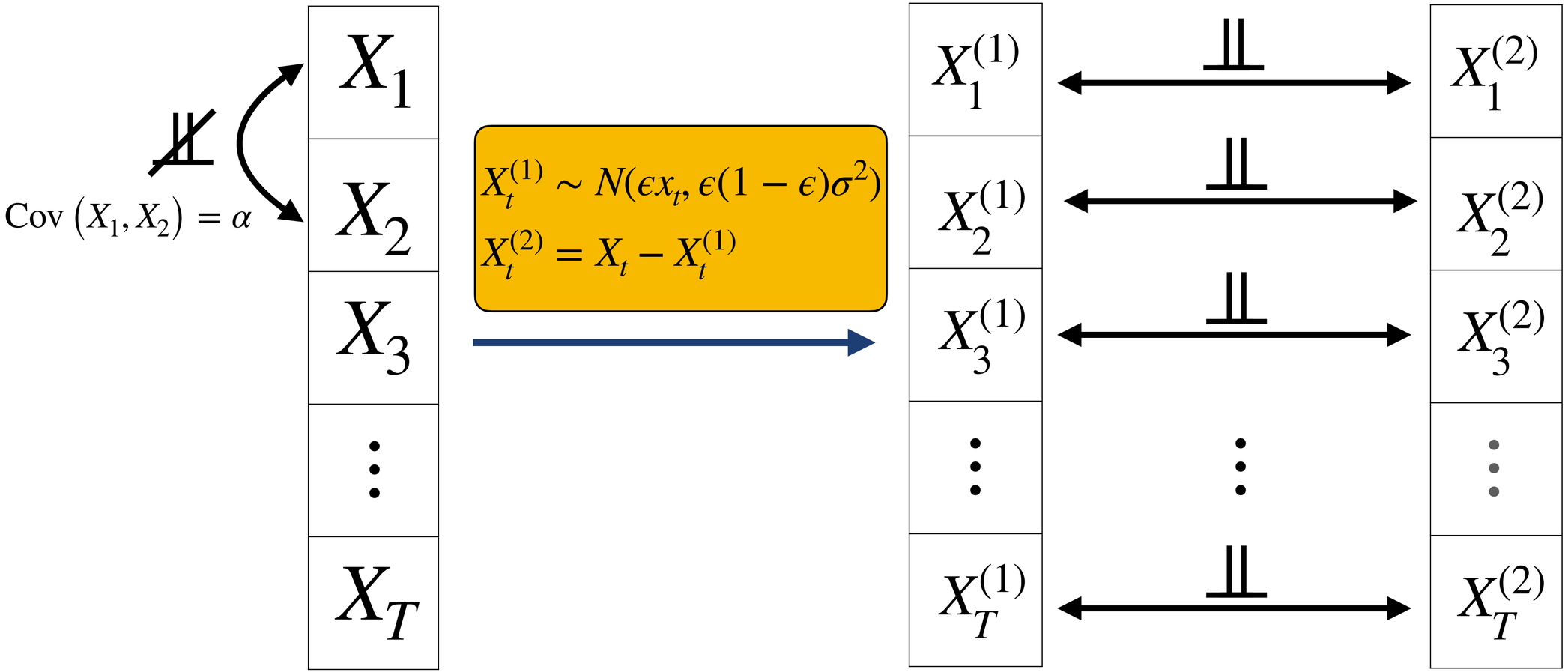


What if we want to thin time series data?

$X_t \sim N(\mu_t, \sigma^2)$

$X_t^{(1)} \sim N(\epsilon\mu_t, \epsilon\sigma^2)$

$X_t^{(2)} \sim N((1-\epsilon)\mu_t, (1-\epsilon)\sigma^2)$

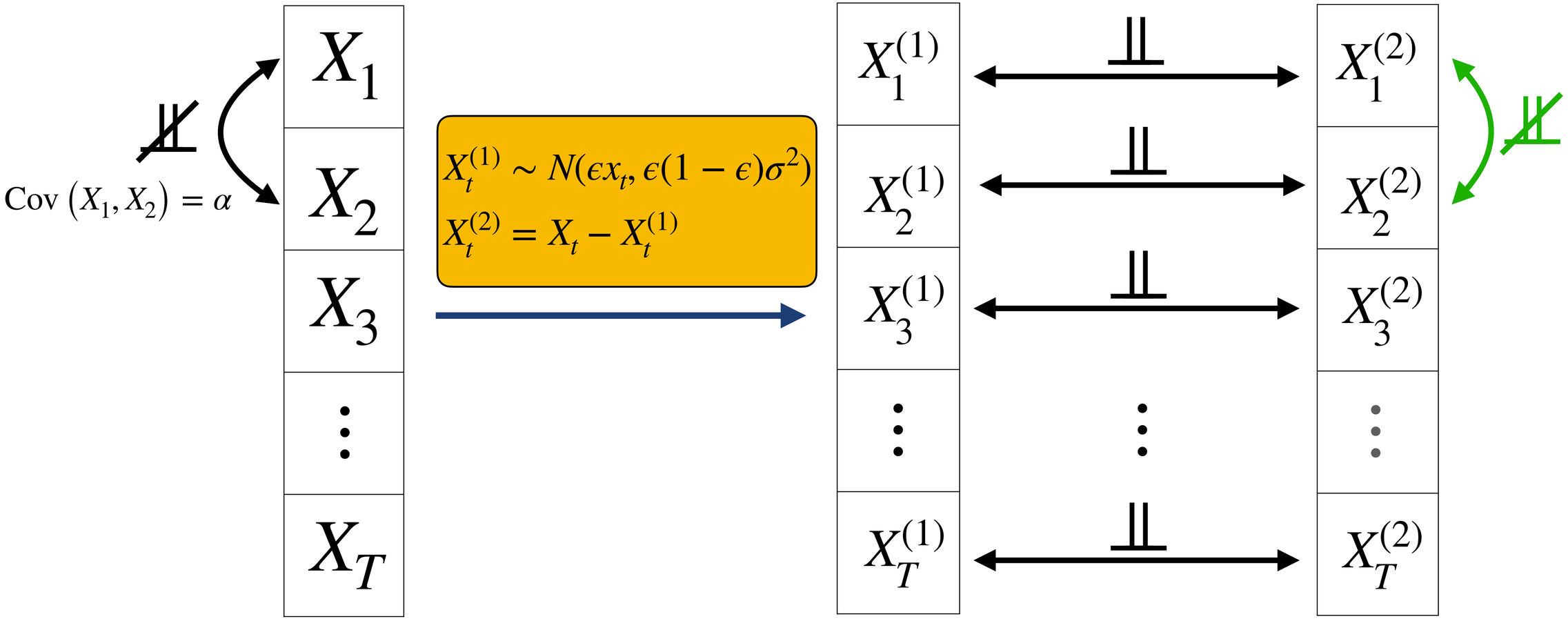


What if we want to thin time series data?

$X_t \sim N(\mu_t, \sigma^2)$

$X_t^{(1)} \sim N(\epsilon\mu_t, \epsilon\sigma^2)$

$X_t^{(2)} \sim N((1-\epsilon)\mu_t, (1-\epsilon)\sigma^2)$

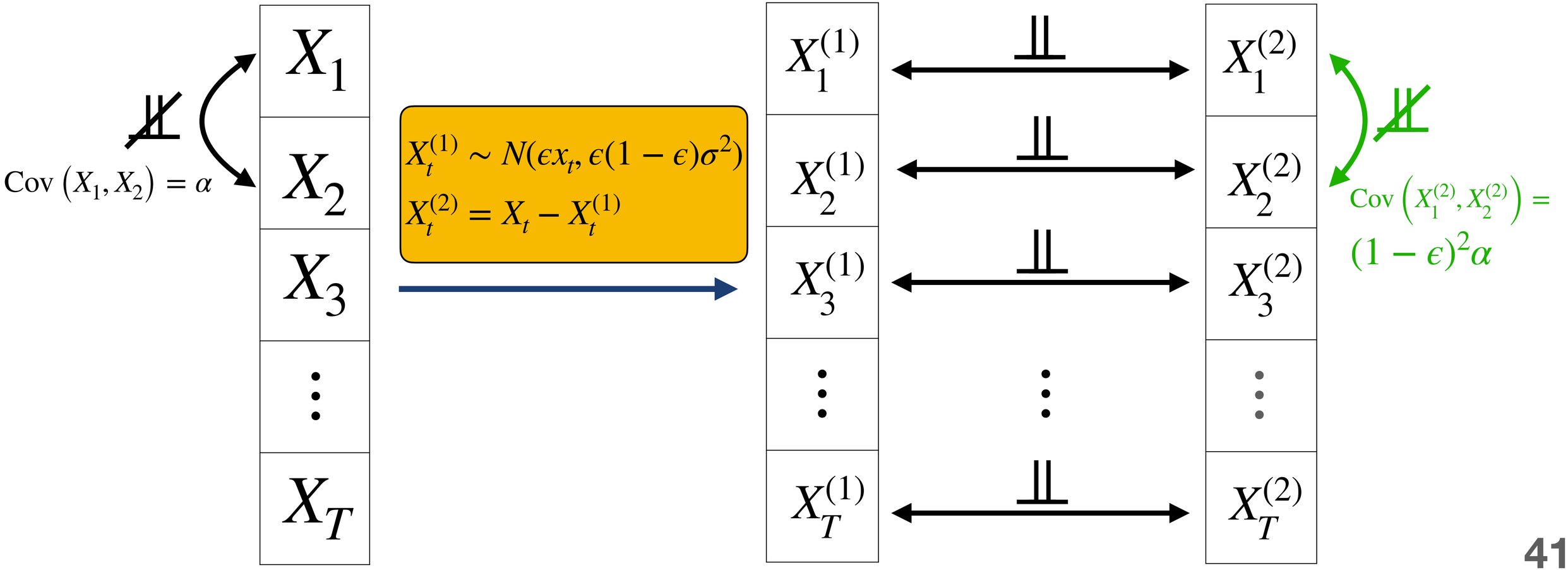


What if we want to thin time series data?

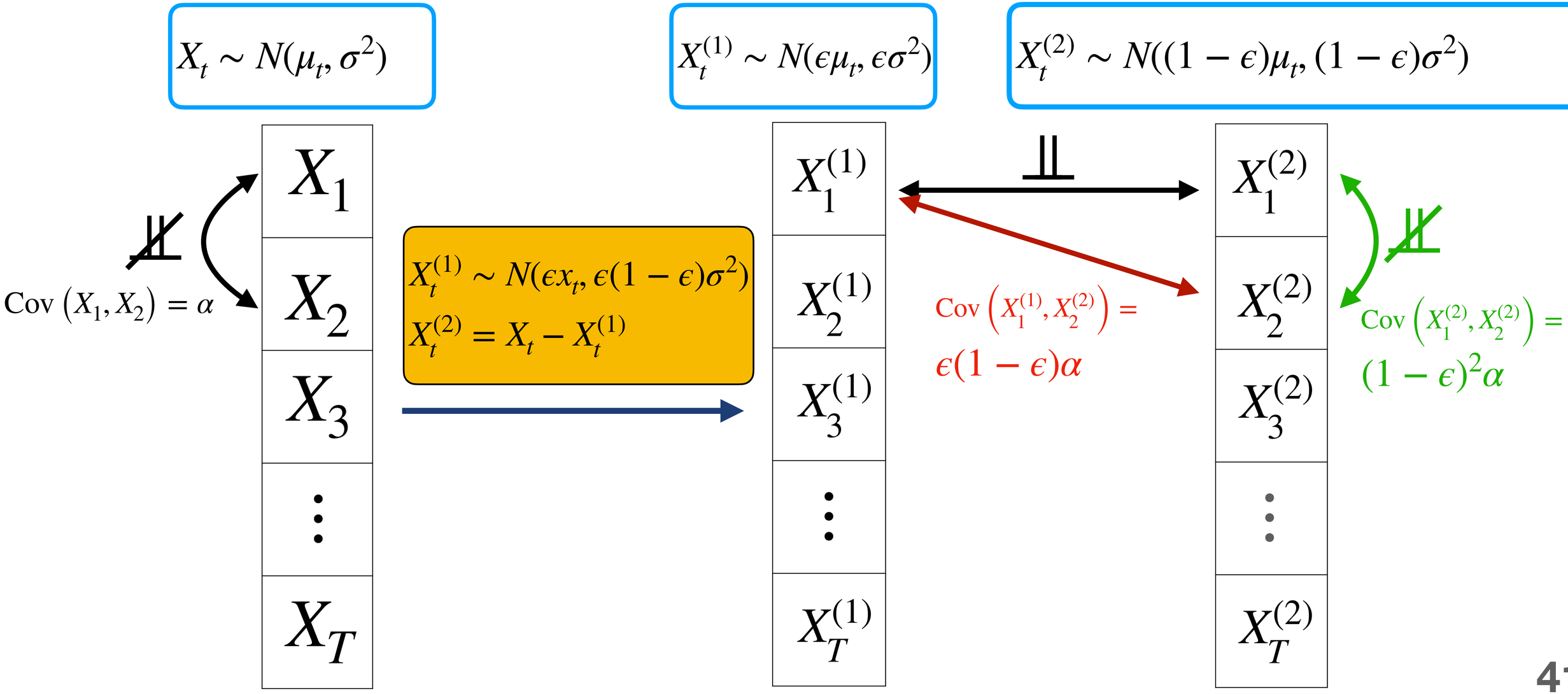
$X_t \sim N(\mu_t, \sigma^2)$

$X_t^{(1)} \sim N(\epsilon\mu_t, \epsilon\sigma^2)$

$X_t^{(2)} \sim N((1-\epsilon)\mu_t, (1-\epsilon)\sigma^2)$



What if we want to thin time series data?



A new approach for thinning time series?

A new approach for thinning time series?

Let $X = (X_1, \ldots, X_T)$, where each $X_t \sim N(0, \sigma^2)$ and $\mathrm{Cov}(X_t, X_{t+1}) \neq 0$.

A new approach for thinning time series?

Let $X = (X_1, \ldots, X_T)$, where each $X_t \sim N(0, \sigma^2)$ and $\mathrm{Cov}(X_t, X_{t+1}) \neq 0$.

A new approach for thinning time series?

Let $X = \left(X_1, \ldots, X_T\right)$, where each $X_t \sim N(0,\sigma^2)$ and $\mathrm{Cov}(X_t, X_{t+1}) \neq 0$.

We can treat X as a *single realization* from $\mathrm{N}_T(0,\Sigma)$ and use multivariate thinning:

A new approach for thinning time series?

Let $X = (X_1, \ldots, X_T)$, where each $X_t \sim N(0, \sigma^2)$ and $\text{Cov}(X_t, X_{t+1}) \neq 0$.

We can treat X as a *single realization* from $\mathbf{N}_T(0, \Sigma)$ and use multivariate thinning:

1. If Σ is known, then we've already shown how to do this.

A new approach for thinning time series?

Let $X = (X_1, \ldots, X_T)$, where each $X_t \sim N(0, \sigma^2)$ and $\text{Cov}(X_t, X_{t+1}) \neq 0$.

We can treat X as a *single realization* from $\text{N}_T(0, \Sigma)$ and use multivariate thinning:

1. If Σ is known, then we've already shown how to do this.
2. If Σ is unknown, we can try to thin $W = XX^\top \sim \text{Wishart}_T(1, \Sigma)$.

A new approach for thinning time series?

Let $X = (X_1, \ldots, X_T)$, where each $X_t \sim N(0, \sigma^2)$ and $\text{Cov}(X_t, X_{t+1}) \neq 0$.

We can treat X as a *single realization* from $\text{N}_T(0, \Sigma)$ and use multivariate thinning:

1. If Σ is known, then we've already shown how to do this.
2. If Σ is unknown, we can try to thin $W = XX^\top \sim \text{Wishart}_T(1, \Sigma)$.
   - **Challenge:** this is a singular Wishart, and it's hard to work with.

Acknowledgements

Daniela Witten
University of Washington

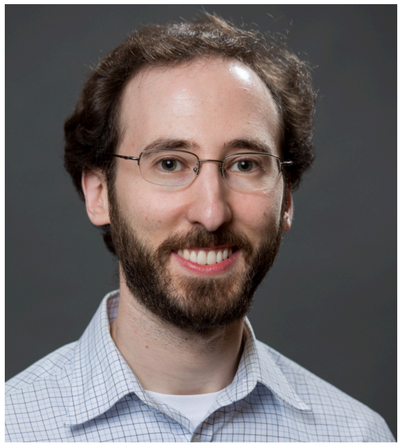
Lucy Gao
University of British Columbia

Ameer Dharamshi
University of Washington

Keshav Motwani
University of Washington

Alexis Battle
Johns Hopkins

Joshua Popp
Johns Hopkins



Jacob Bien
USC

Questions?

Are the results of our data analysis stable?