Data thinning for convolution-closed distributions

Anna Neufeld

June, 2023

We often wish that we had access to two independent datasets, drawn from the same data-generating mechanism

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

	Feature 1	Feature 2
Obs. 1	9	3
Obs. 2	33	9
Obs. 3	12	29
Obs. 4	25	41

We often wish that we had access to two independent datasets, drawn from the same data-generating mechanism

Train

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

Test

	Feature 1	Feature 2
Obs. 1	9	3
Obs. 2	33	9
Obs. 3	12	29
Obs. 4	25	41

We often wish that we had access to two independent datasets, drawn from the same data-generating mechanism

Train

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

Fit model.

Test

	Feature 1	Feature 2
Obs. 1	9	3
Obs. 2	33	9
Obs. 3	12	29
Obs. 4	25	41

We often wish that we had access to two independent datasets, drawn from the same data-generating mechanism

Train

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

Fit model.

Test

	Feature 1	Feature 2
Obs. 1	9	3
Obs. 2	33	9
Obs. 3	12	29
Obs. 4	25	41

Validate model.

As we typically have access to only one dataset, a common solution is sample splitting

Original data

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

As we typically have access to only one dataset, a common solution is sample splitting

Original data

	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

As we typically have access to only one dataset, a common solution is sample splitting

Original data

	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

As we typically have access to only one dataset, a common solution is sample splitting

Original data

	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

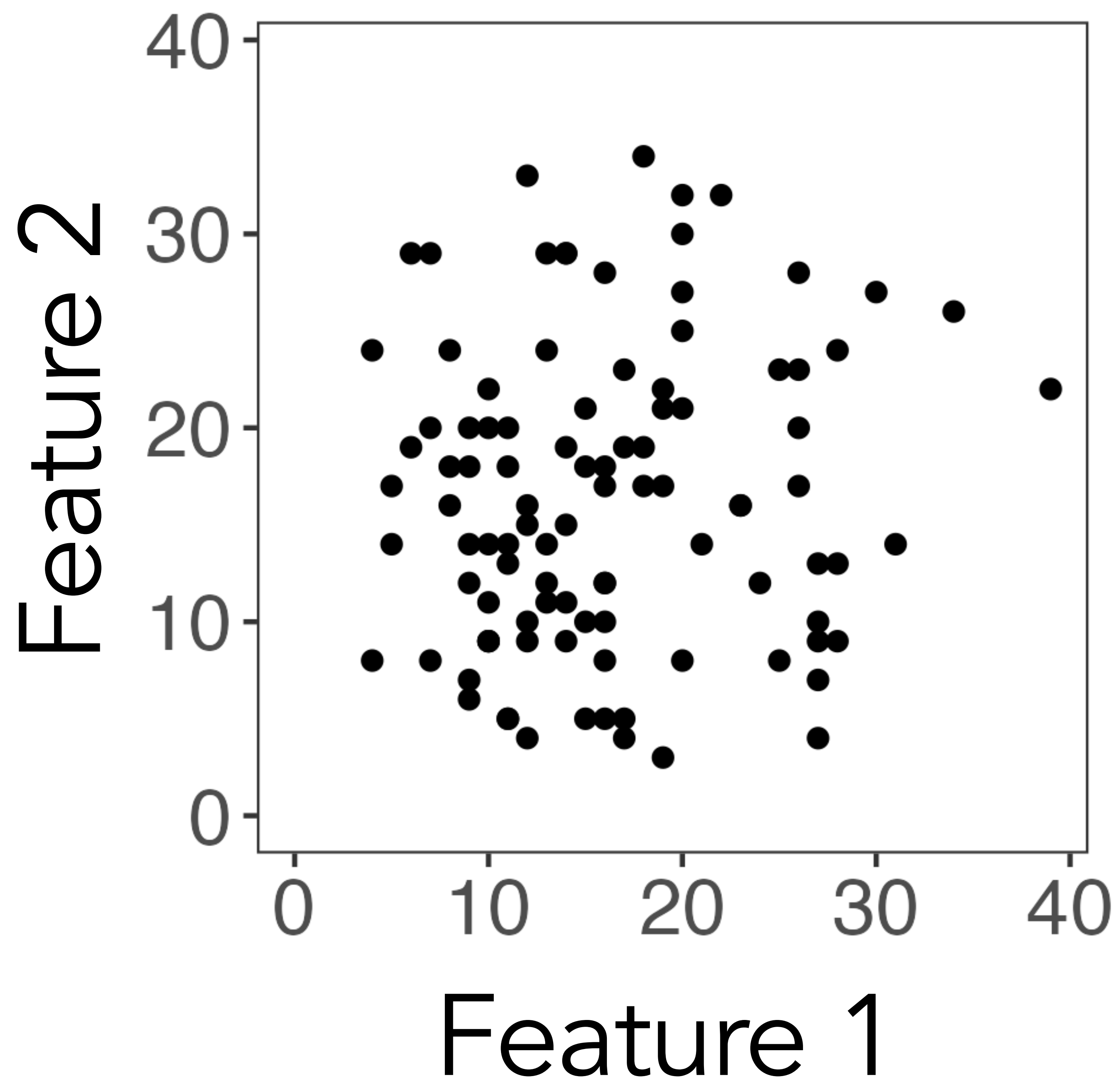
	Feature 1	Feature 2
Obs. 3	11	31
Obs. 4	22	34

Validate model.

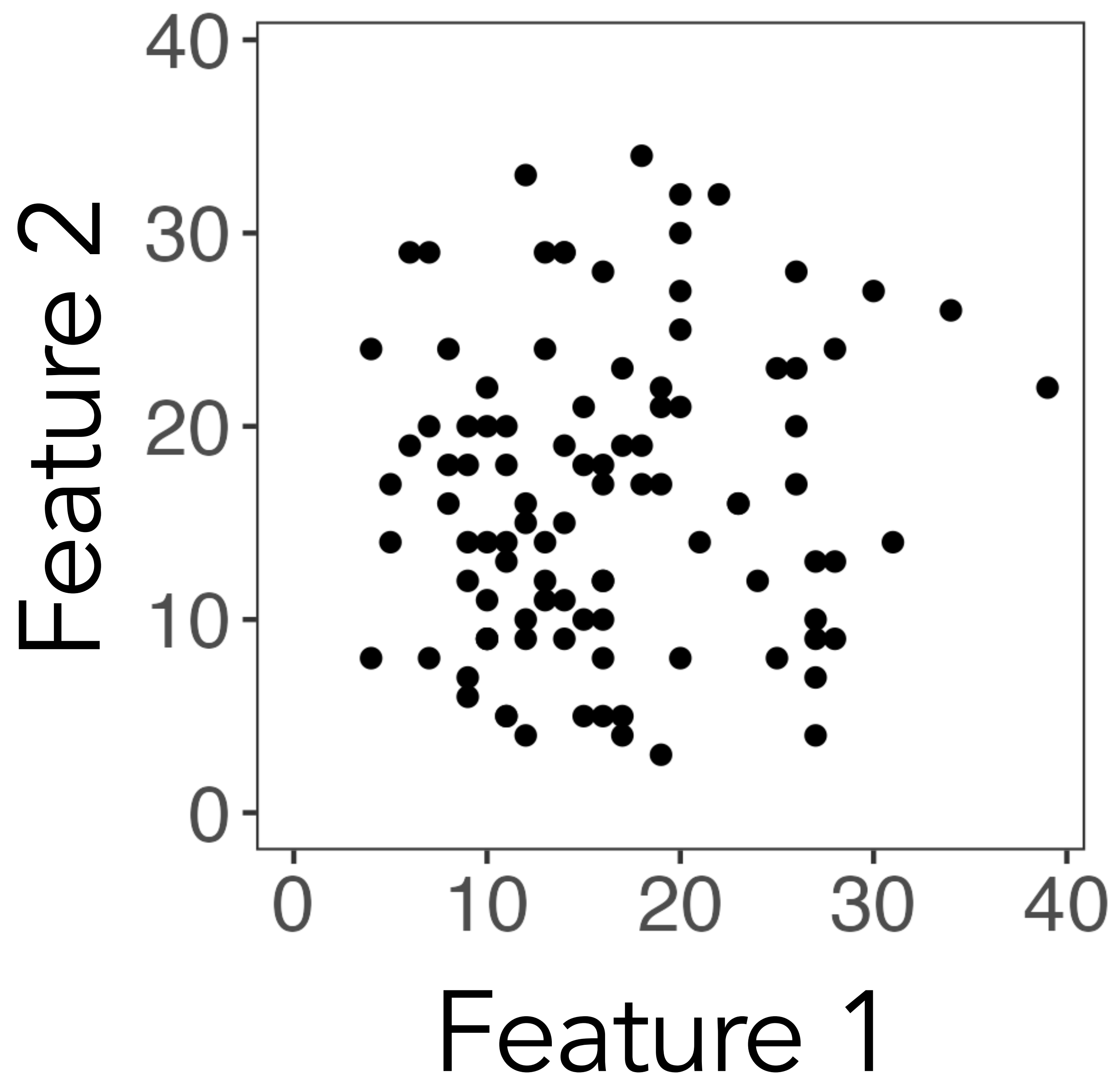
Outline

1. **Motivation: sample splitting doesn't always work**
2. Poisson thinning
3. Data thinning
4. Application to changepoint detection
5. Generalized data thinning

Example: how many clusters are in this data?



Example: how many clusters are in this data?

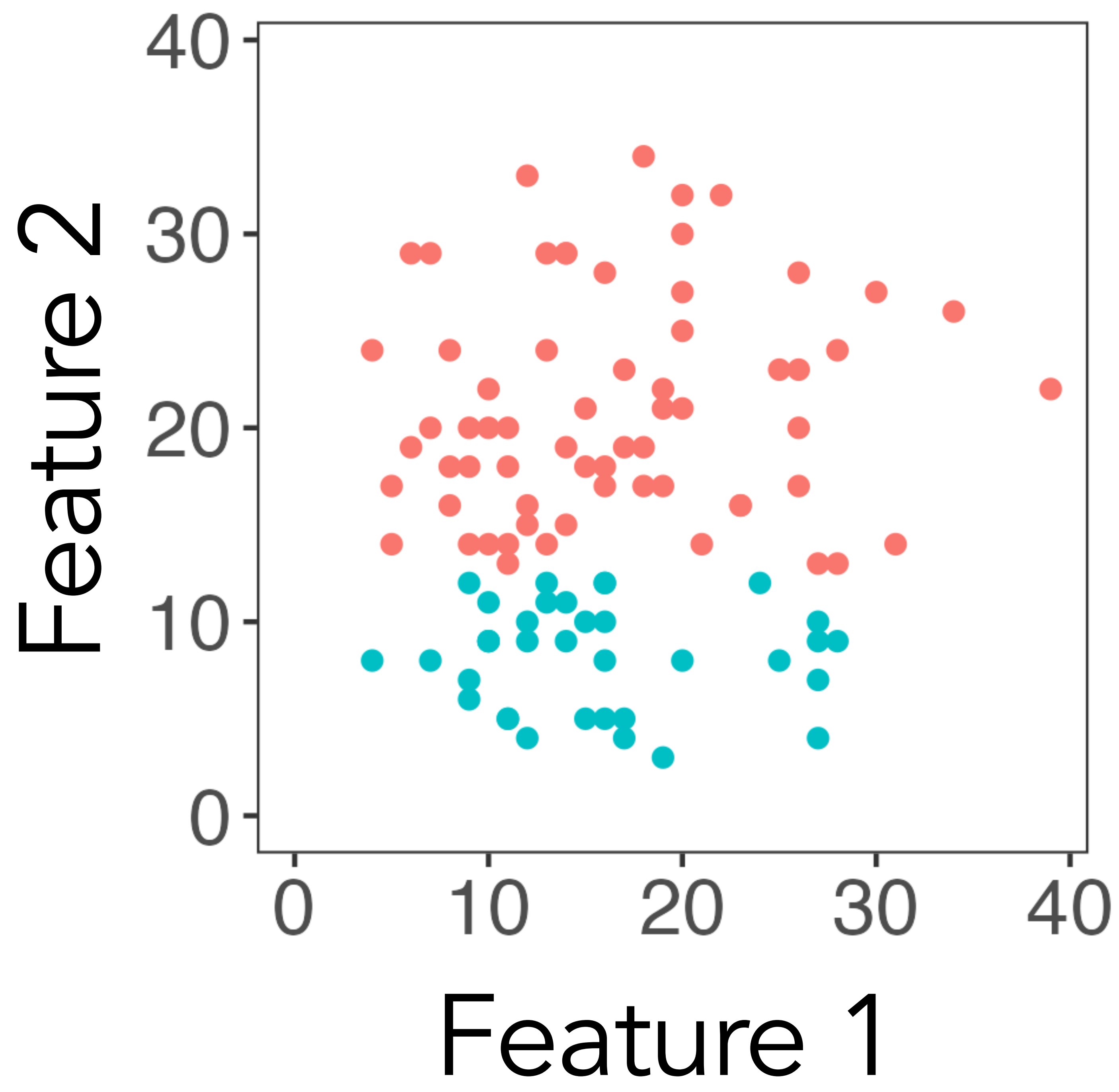


For several values of k:

Step 1: fit a model with k clusters.

Step 2: evaluate model using a loss function.

Example: how many clusters are in this data?

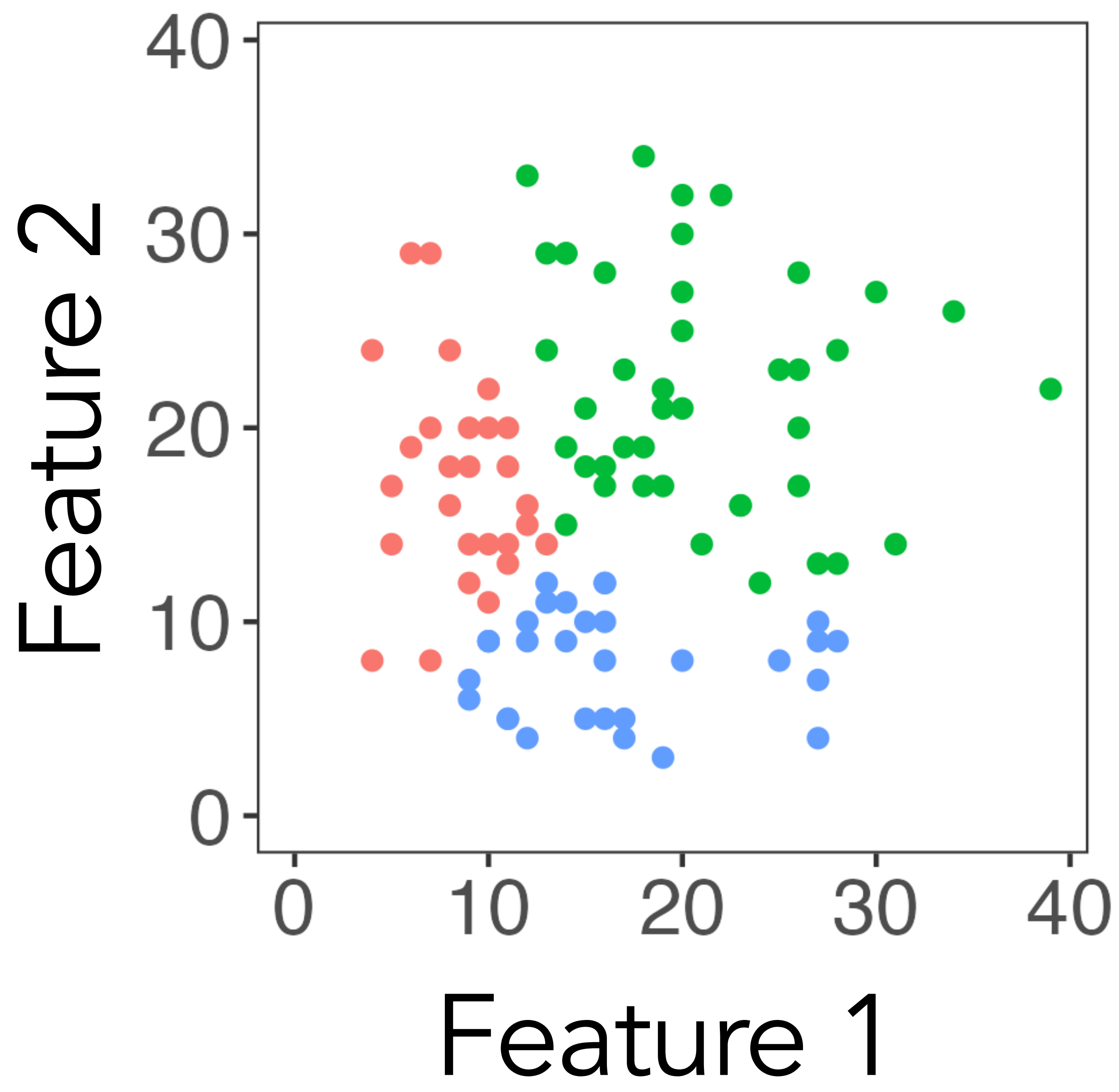


For several values of k:

Step 1: fit a model with k clusters.

Step 2: evaluate model using a loss function.

Example: how many clusters are in this data?

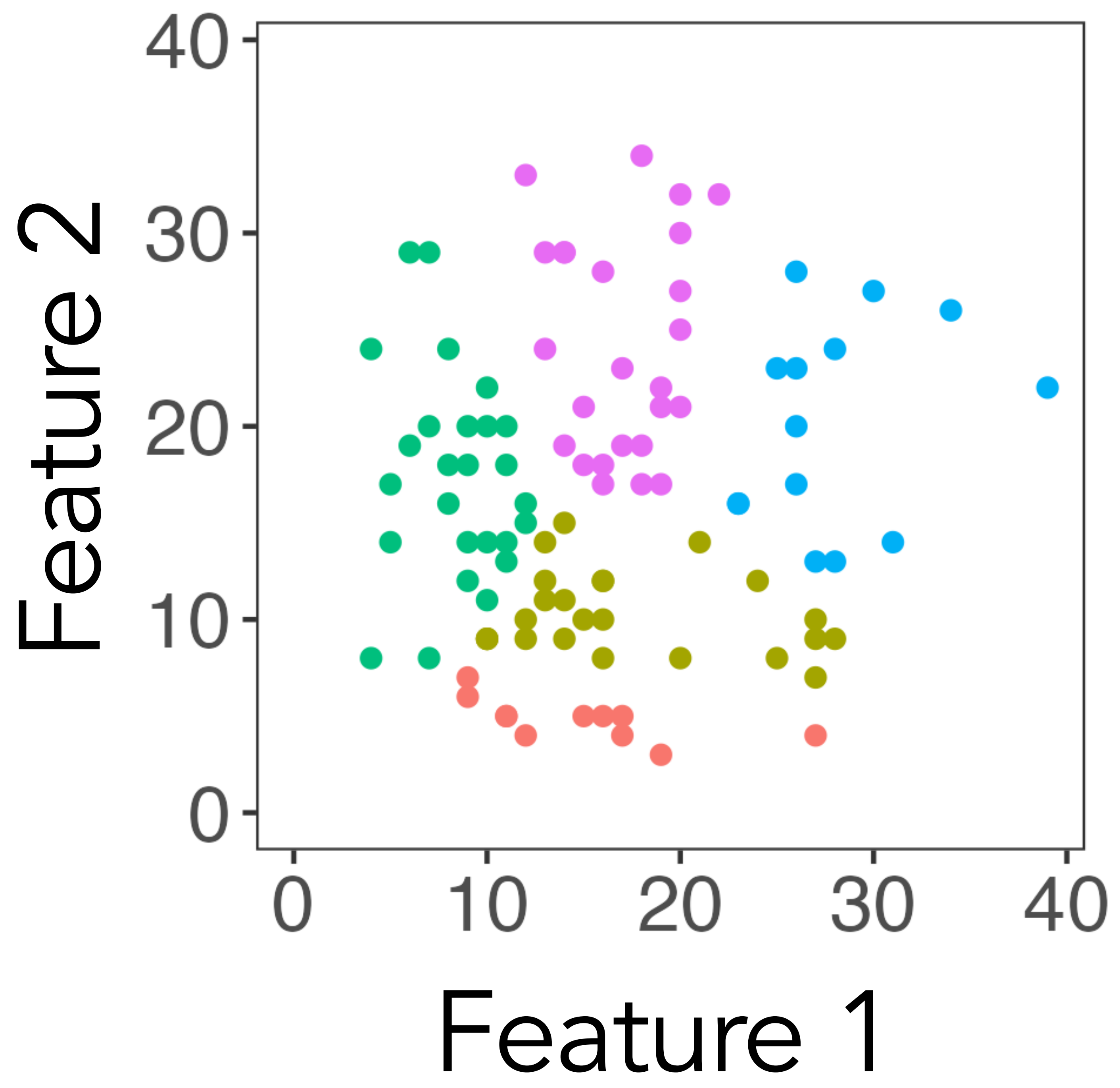


For several values of k:

Step 1: fit a model with k clusters.

Step 2: evaluate model using a loss function.

Example: how many clusters are in this data?

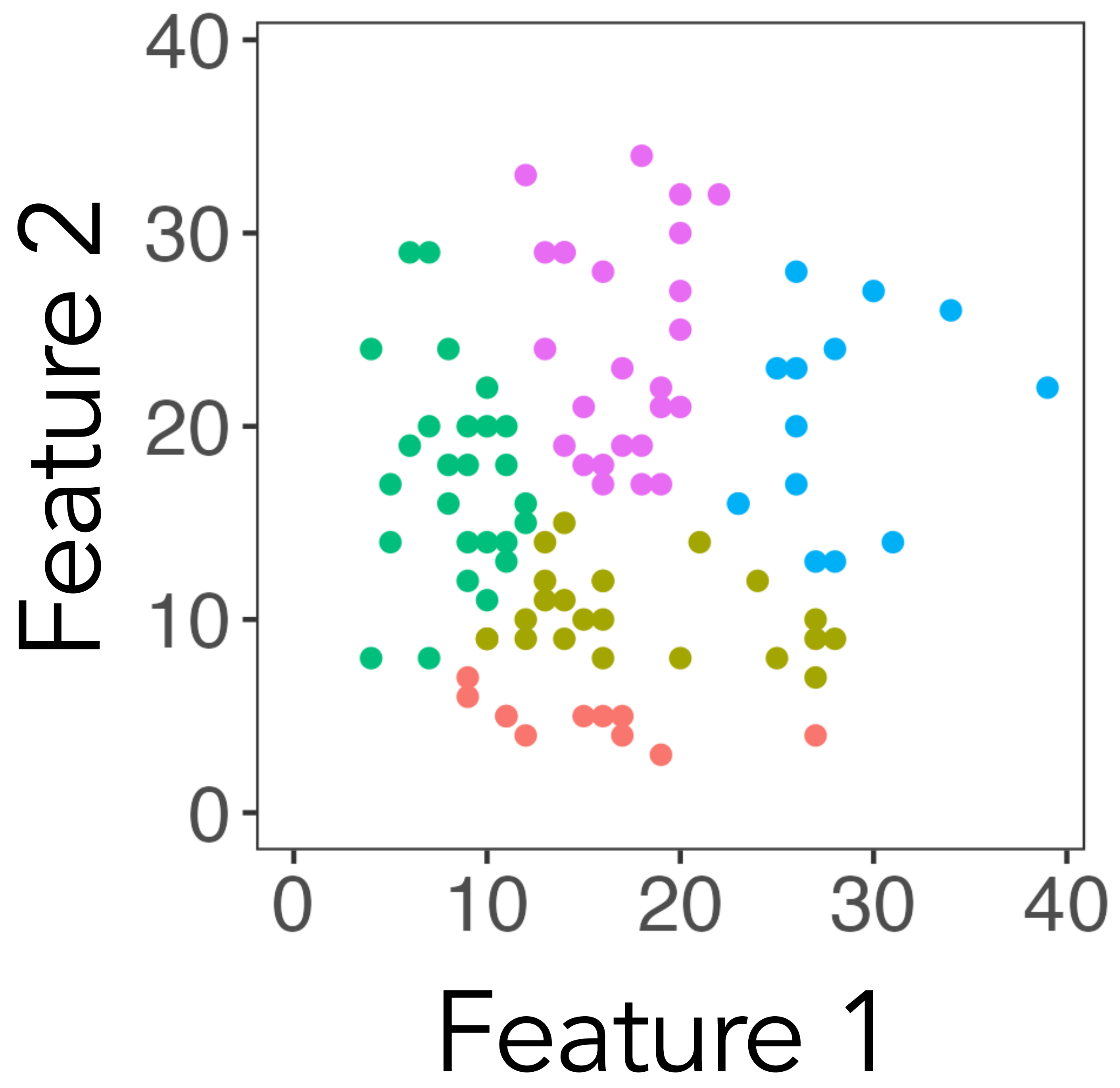


For several values of k:

Step 1: fit a model with k clusters.

Step 2: evaluate model using a loss function.

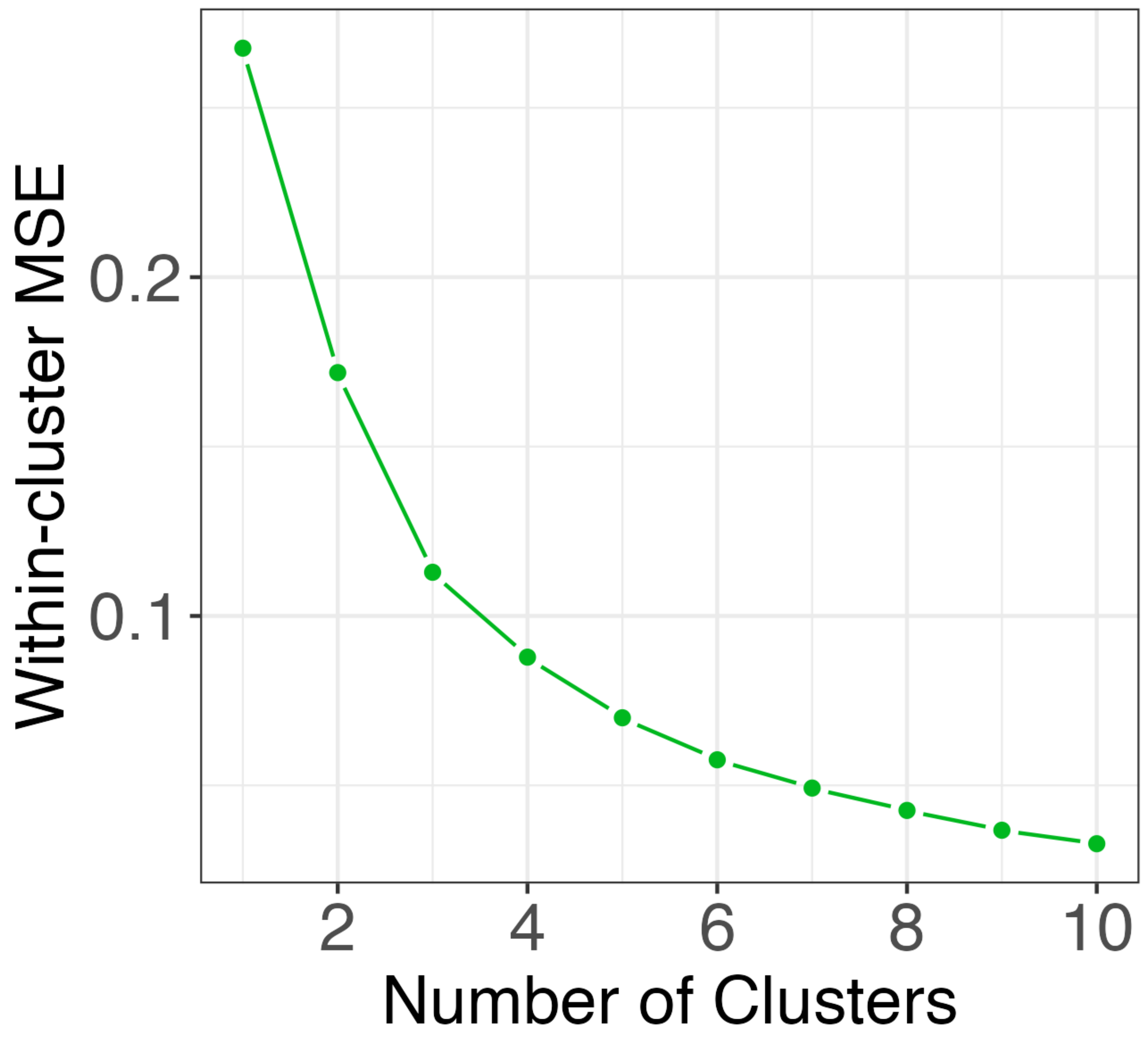
Example: how many clusters are in this data?



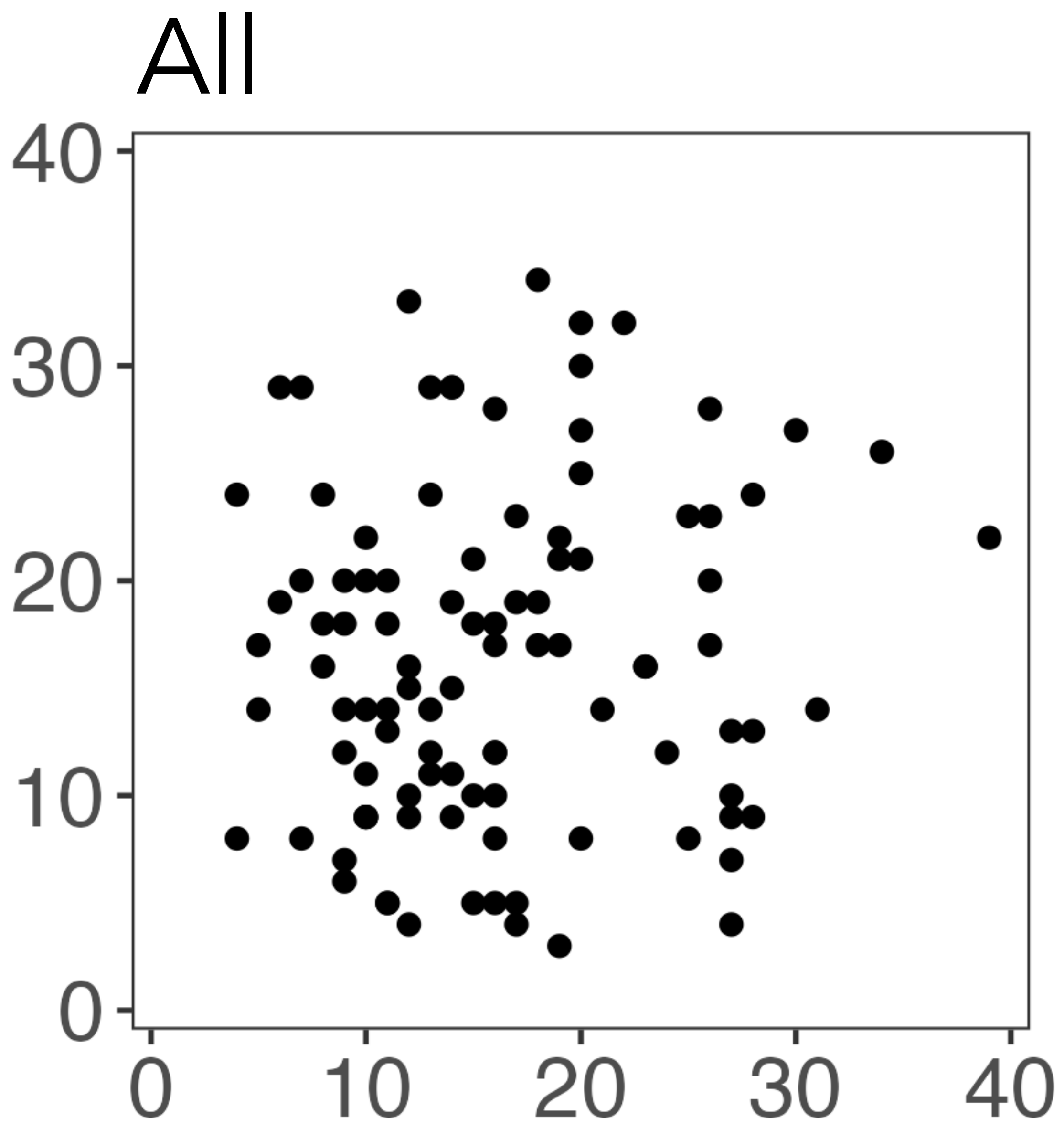
For several values of k:

Step 1: fit a model with k clusters.

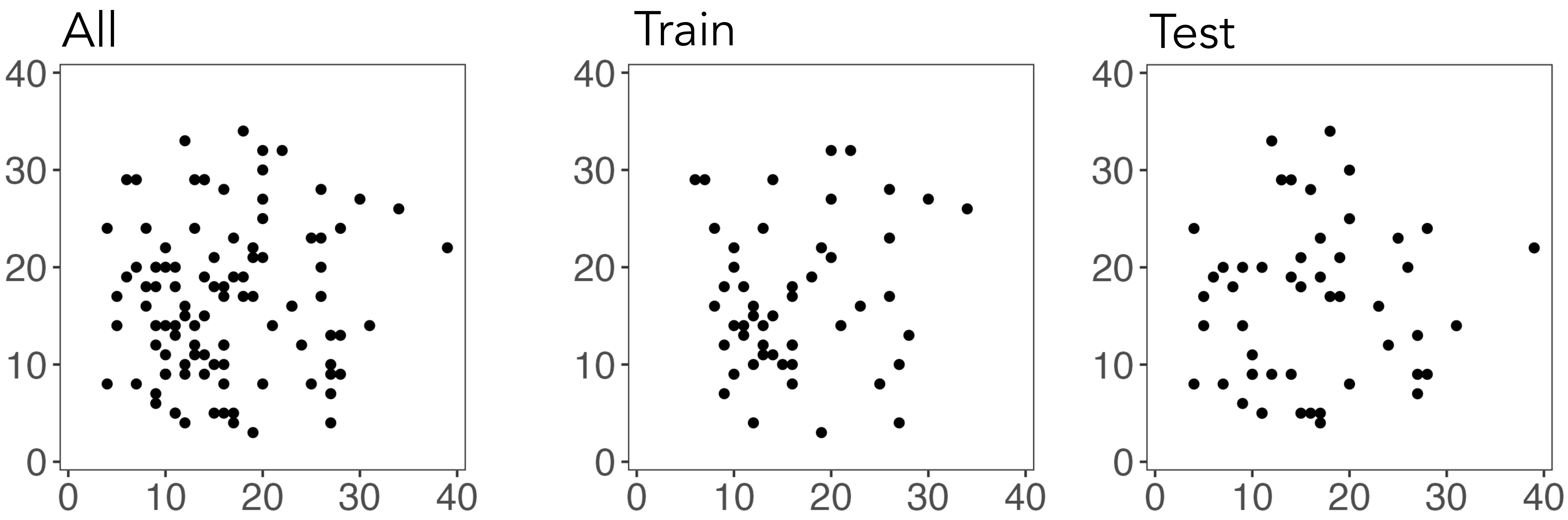
Step 2: evaluate model using a loss function.



Sample splitting cannot be used to validate the results of clustering

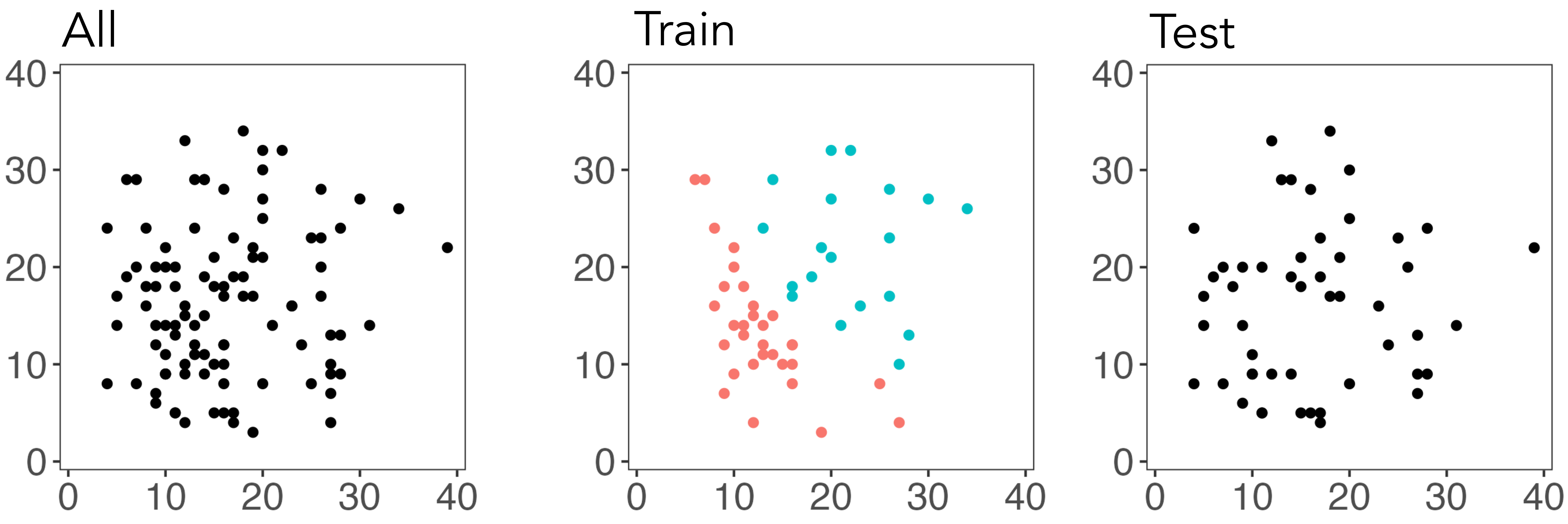


Sample splitting cannot be used to validate the results of clustering



Step 1: split observations into train/test.

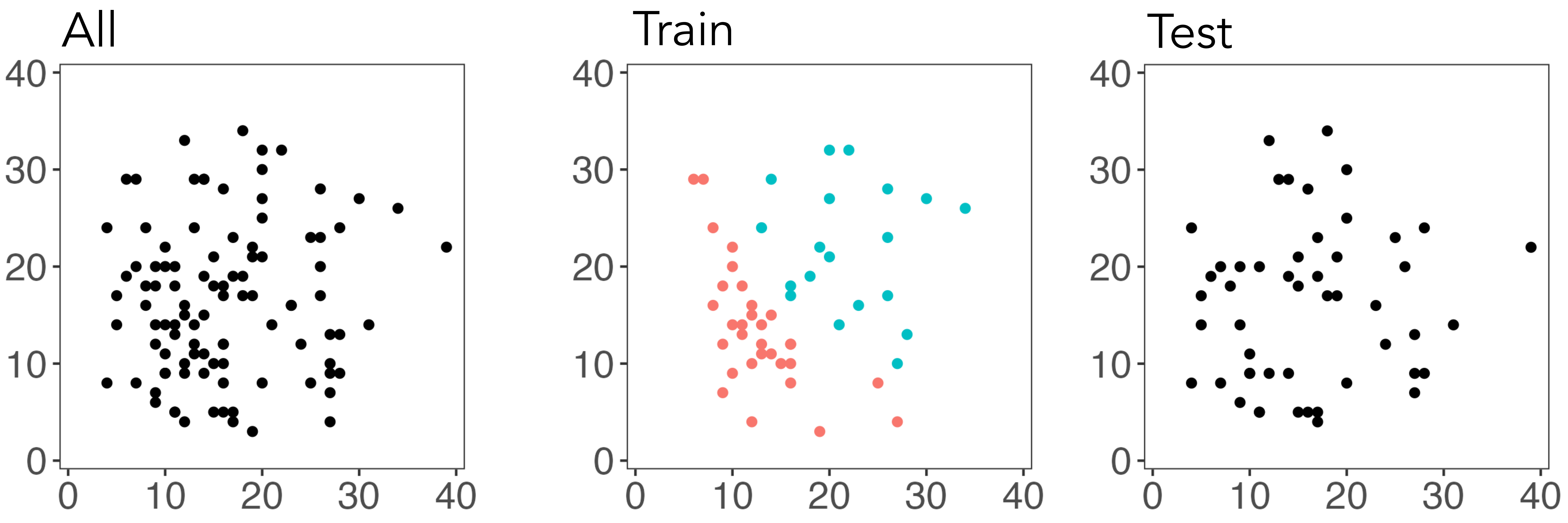
Sample splitting cannot be used to validate the results of clustering



Step 1: split observations into train/test.

Step 2: cluster the training set.

Sample splitting cannot be used to validate the results of clustering

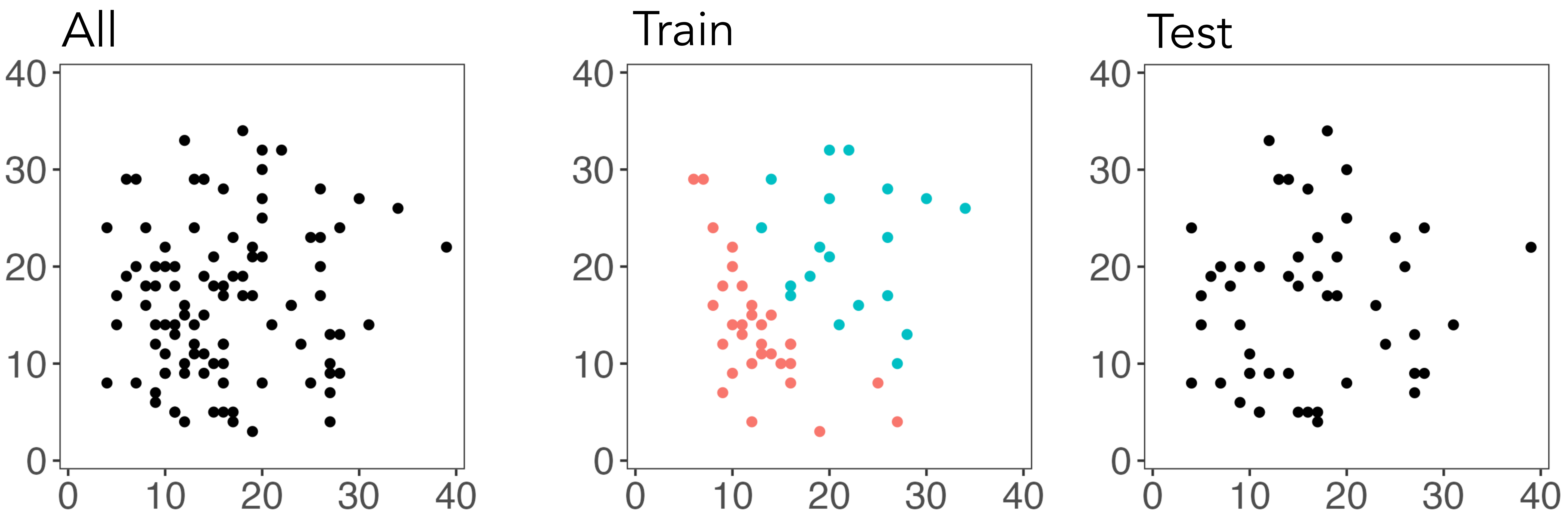


Step 1: split observations into train/test.

Step 2: cluster the training set.

Step 3: evaluate clusters using test set.

Sample splitting cannot be used to validate the results of clustering



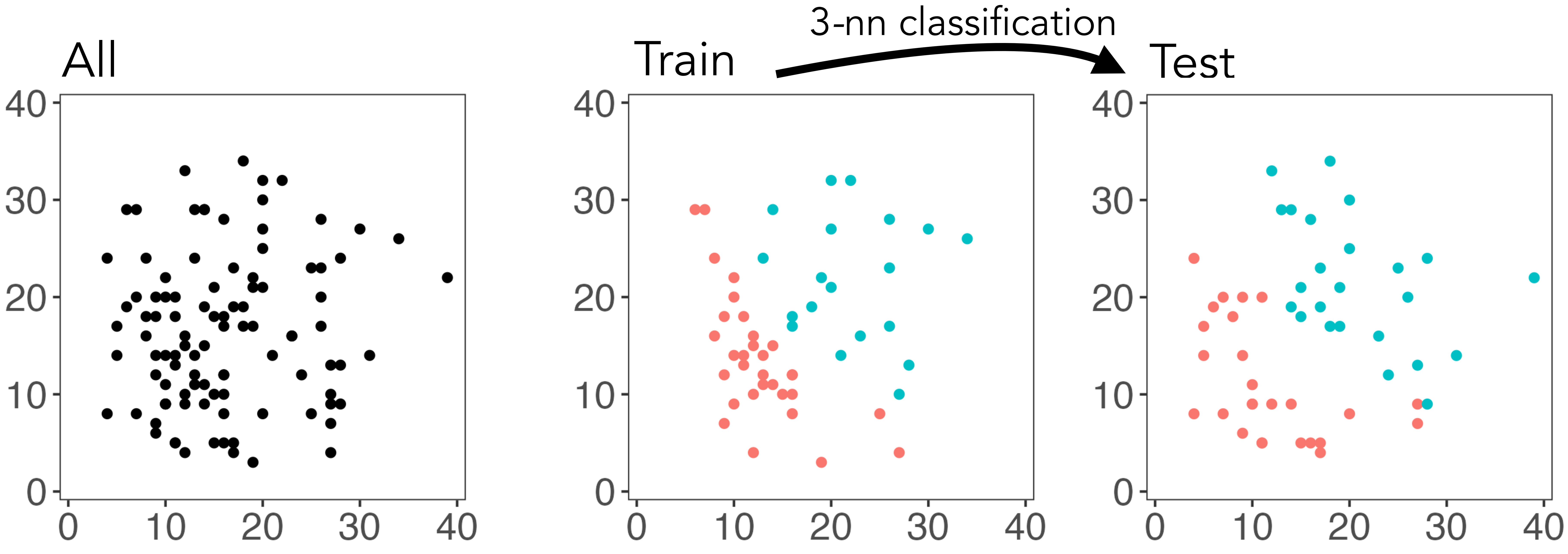
Step 1: split observations into train/test.

Step 2: cluster the training set.

Step 2.5: assign labels to observations in test set.

Step 3: evaluate clusters using test set.

Sample splitting cannot be used to validate the results of clustering



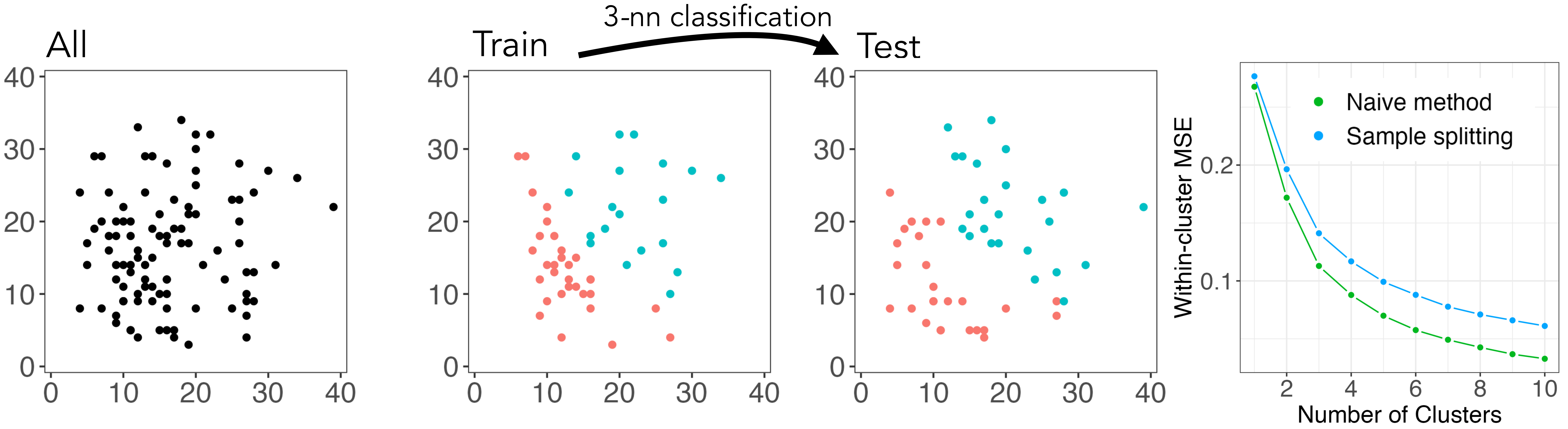
Step 1: split observations into train/test.

Step 2: cluster the training set.

Step 2.5: assign labels to observations in test set.

Step 3: evaluate clusters using test set.

Sample splitting cannot be used to validate the results of clustering



Step 1: split observations into train/test.

Step 2: cluster the training set.

Step 2.5: assign labels to observations in test set.

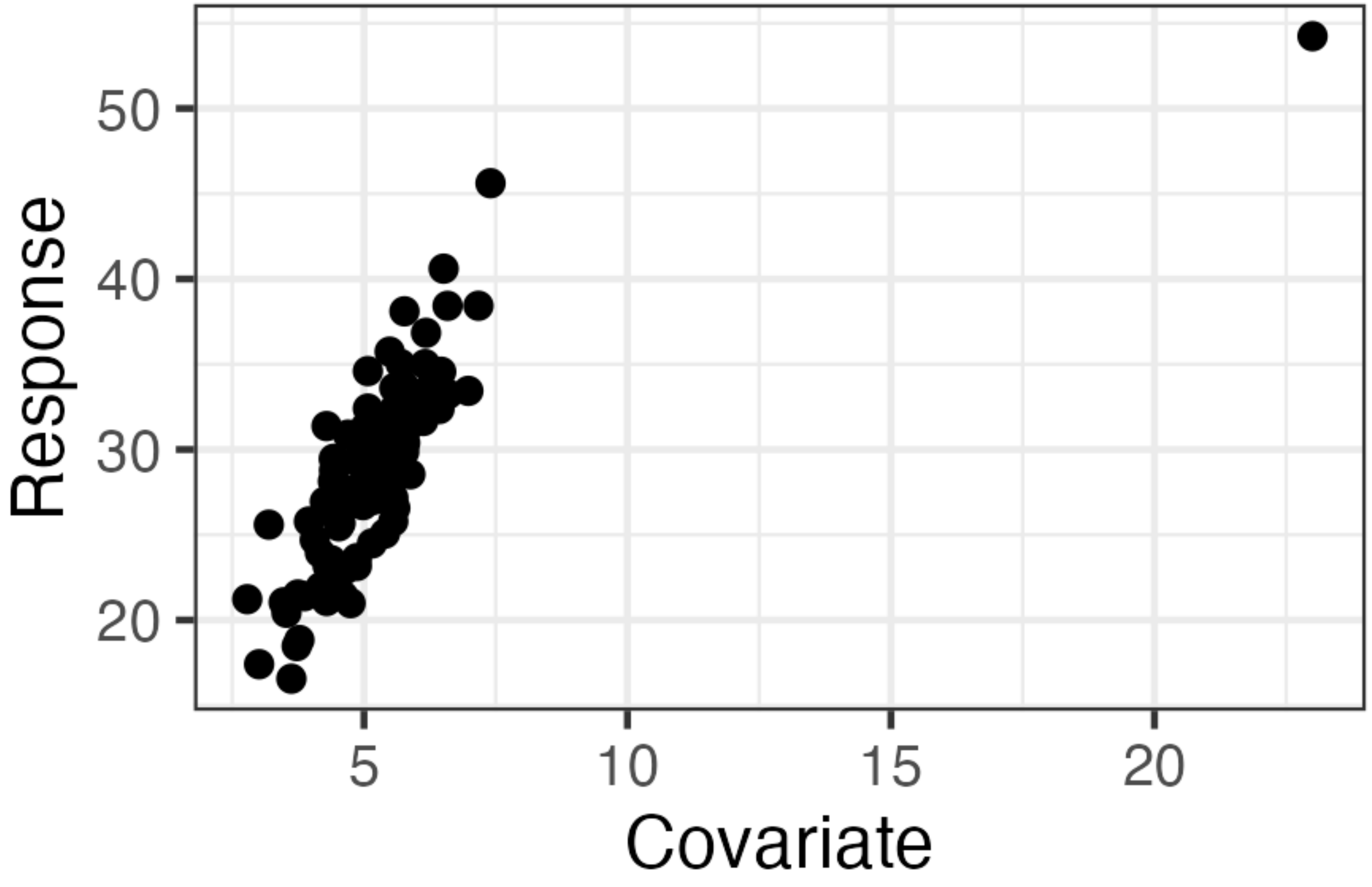
Step 3: evaluate clusters using test set.

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.



Outline

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

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. 4	16	9

Poisson thinning

X

	Feature 1	Feature 2
Obs. 1	18	6
Obs. 2	31	$\boxed{8}$
Obs. 3	11	31
Obs. 4	22	34

X_{ij}

$X^{(1)}$

	Feature 1	Feature 2
Obs. 1	14	1
Obs. 2	10	$\boxed{6}$
Obs. 3	5	17
Obs. 4	6	25

$X^{(2)}$

	Feature 1	Feature 2
Obs. 1	[illegible]	5
Obs. 2	21	$\boxed{2}$
Obs. 3	6	14
Obs. 4	16	9

Poisson thinning

X

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

X_{ij}

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

$X^{(1)}$

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

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

$X^{(2)}$

	Feature 1	Feature 2
[illegible]bs. 1	[illegible]	5
[illegible]bs. 2	21	2
Obs. 3	6	14
Obs. 4	16	9

Poisson thinning

X

	Feature 1	Feature 2
Obs. 1	18	[illegible]
Obs. 2	31	8
Obs. 3	11	31
Obs. 4	22	34

X_{ij}

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

$X^{(1)}$

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

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

$X^{(2)}$

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

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

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

A very well-known result.

Poisson thinning

X

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

X_{ij}

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

$X^{(1)}$

	ture 1	Feature 2
[illegible]	14	1
[illegible]	[illegible]	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	[illegible]	5
bs. 2	21	2
Obs. 3	6	14
Obs. 4	16	9

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

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

A very well-known result.

Poisson thinning

X

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

X_{ij}

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

$X^{(1)}$

	[illegible]ture 1	Feature 2
[illegible]	14	1
[illegible]	[illegible]	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
[illegible]bs. 1	[illegible]	5
[illegible]bs. 2	21	2
Obs. 3	6	14
Obs. 4	16	9

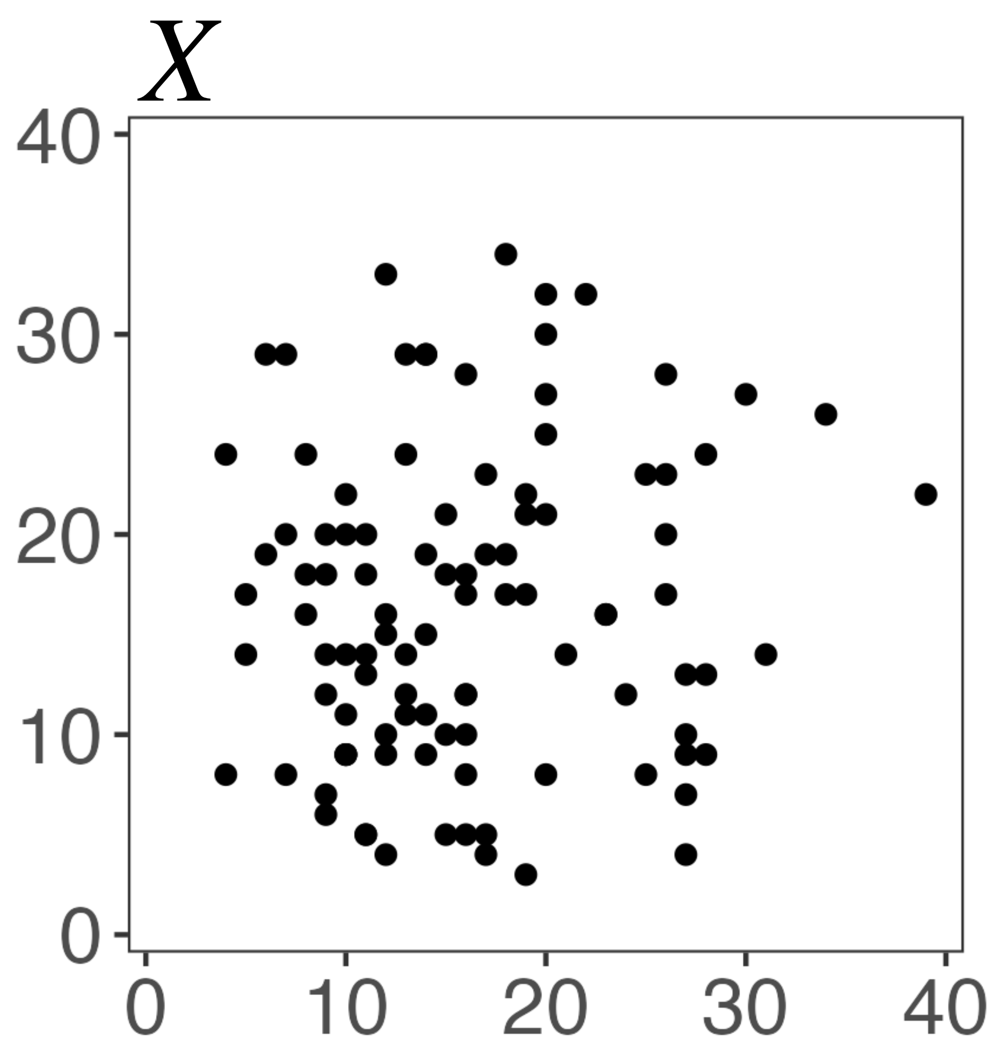
Validate model.

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

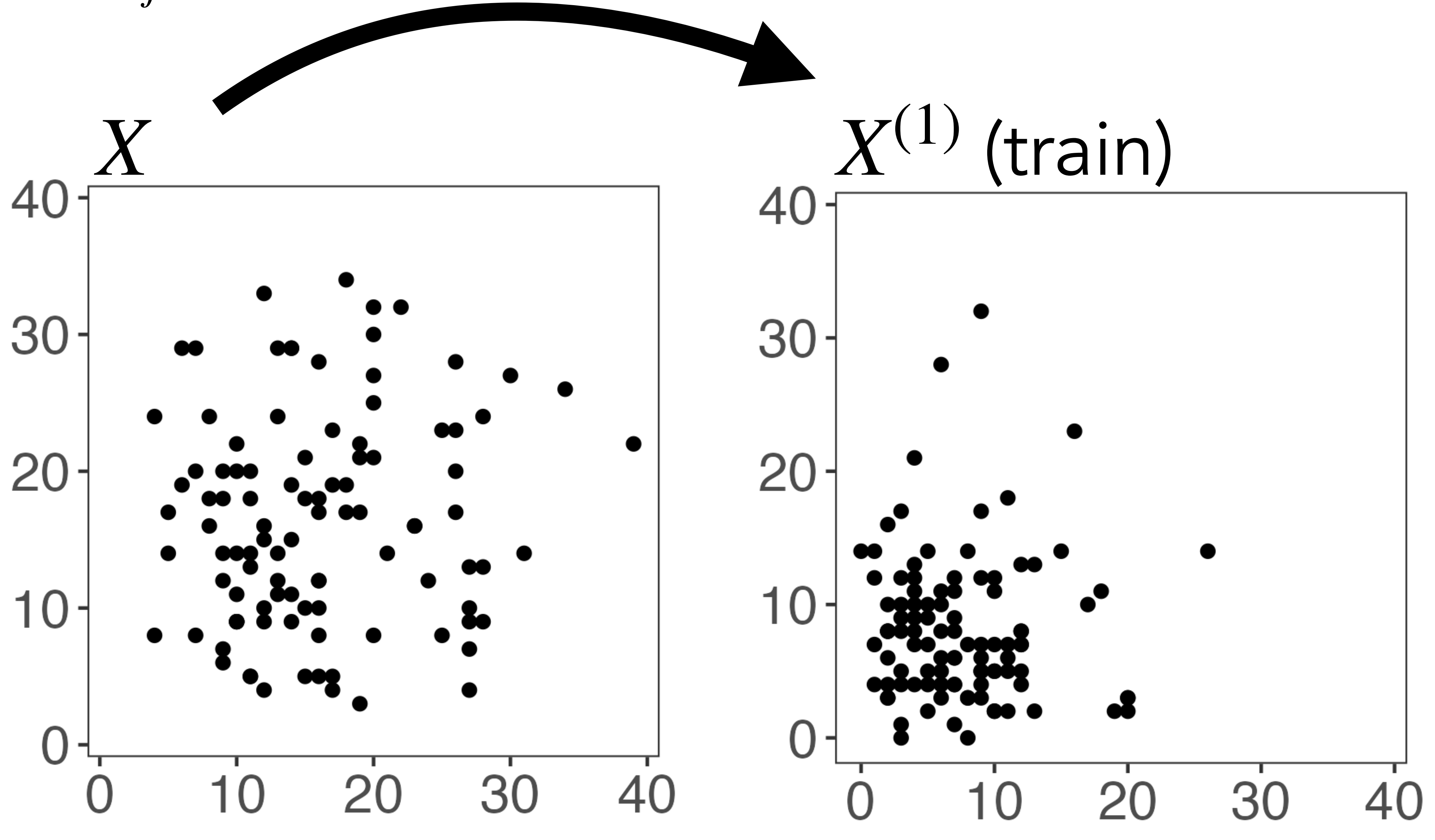
A very well-known result.

Thinning avoids the pitfall of sample splitting on our motivating example



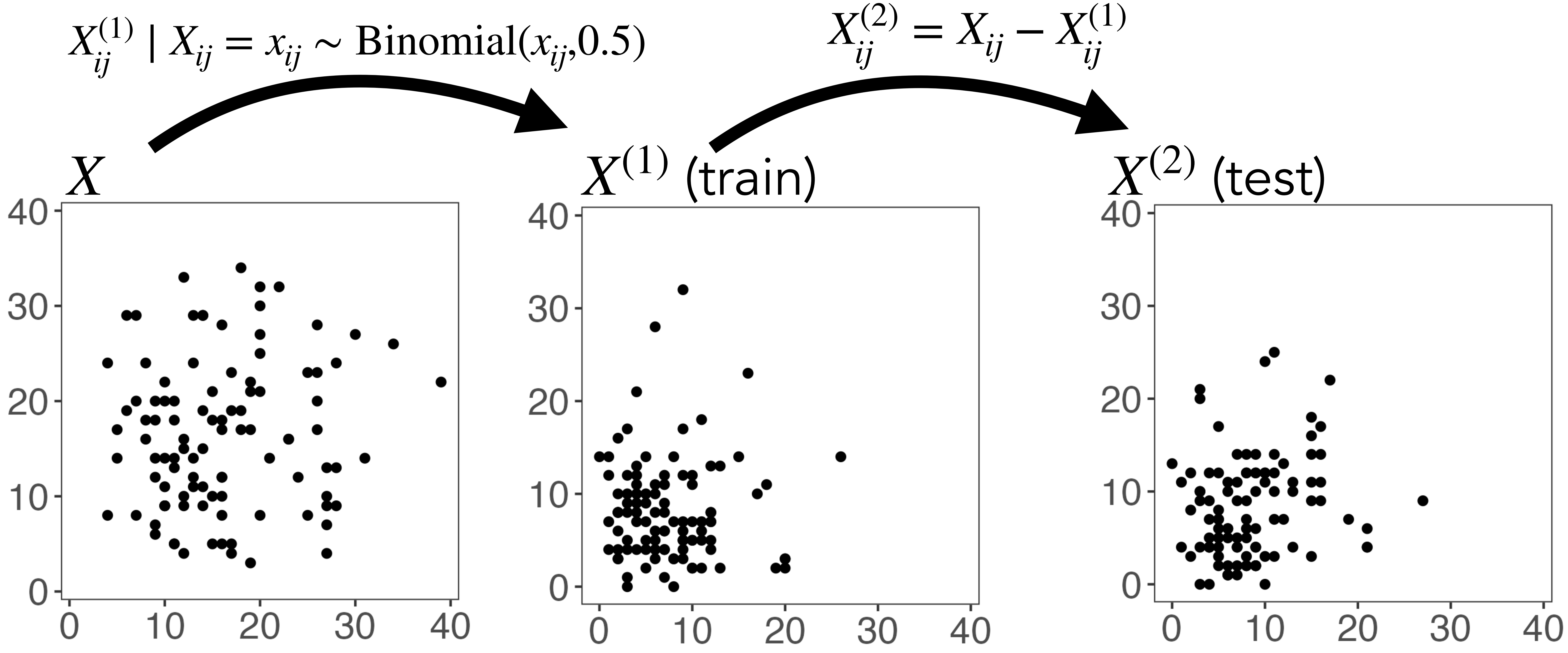
Thinning avoids the pitfall of sample splitting on our motivating example

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



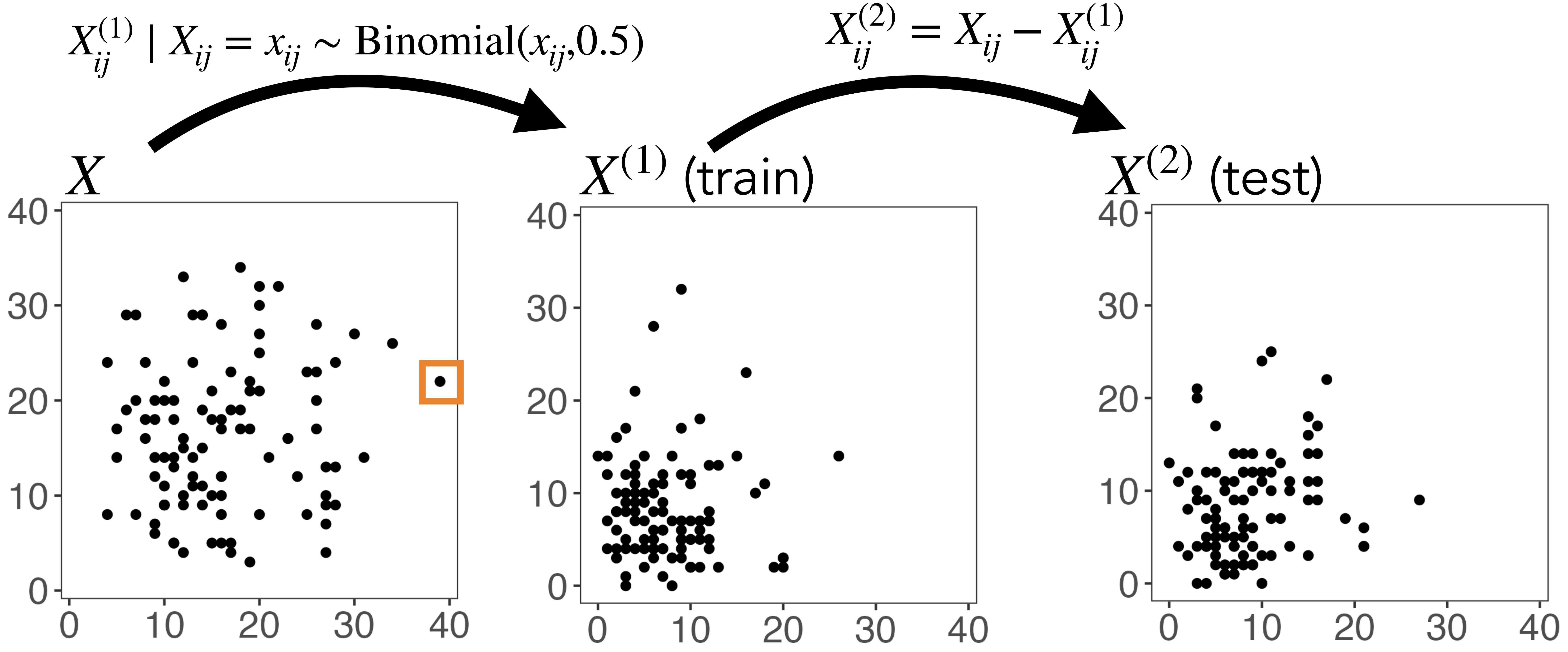
Step 1: thin observations into train/test.

Thinning avoids the pitfall of sample splitting on our motivating example



Step 1: thin observations into train/test.

Thinning avoids the pitfall of sample splitting on our motivating example

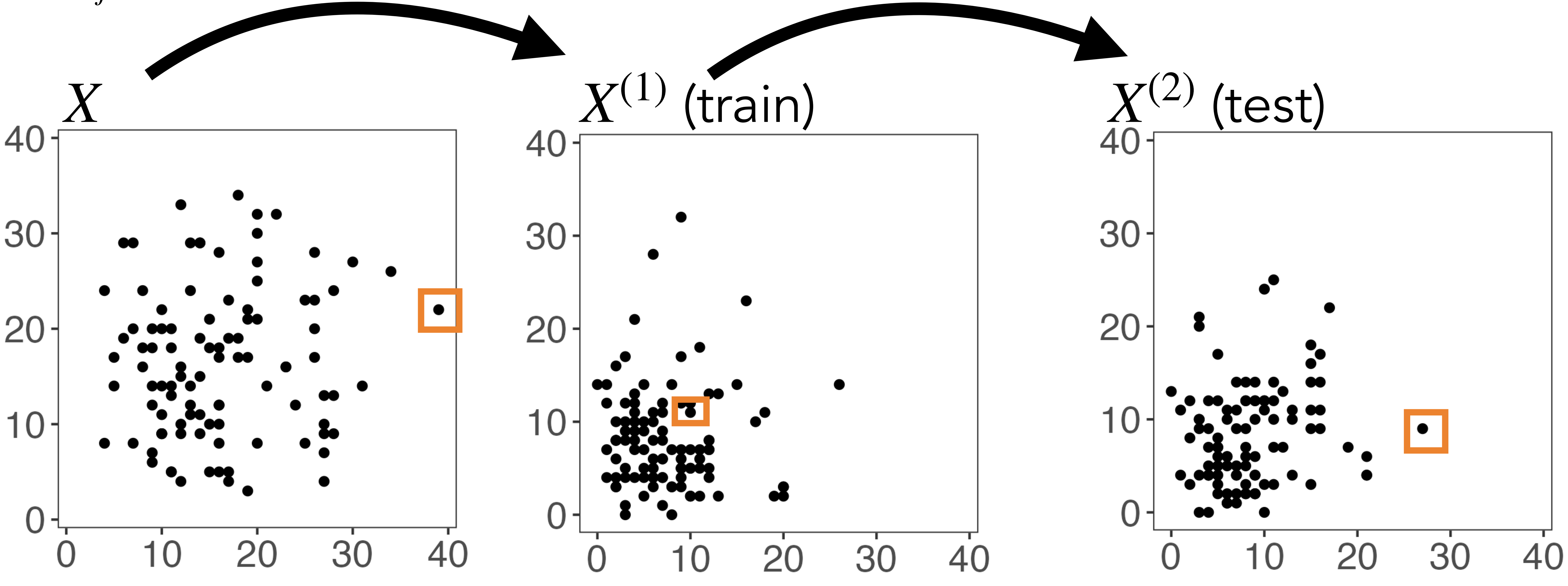


Step 1: thin observations into train/test.

Thinning avoids the pitfall of sample splitting on our motivating example

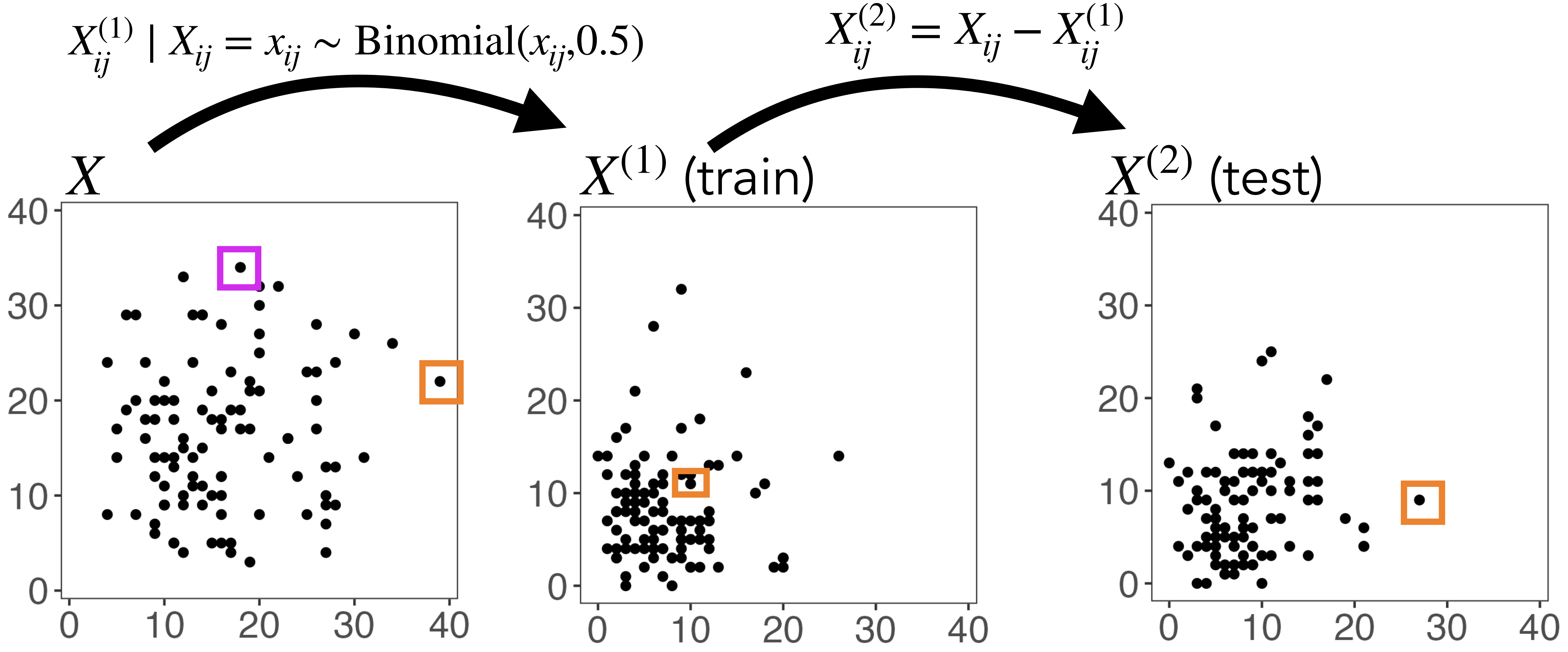
$X_{ij}^{(1)} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, 0.5)$

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



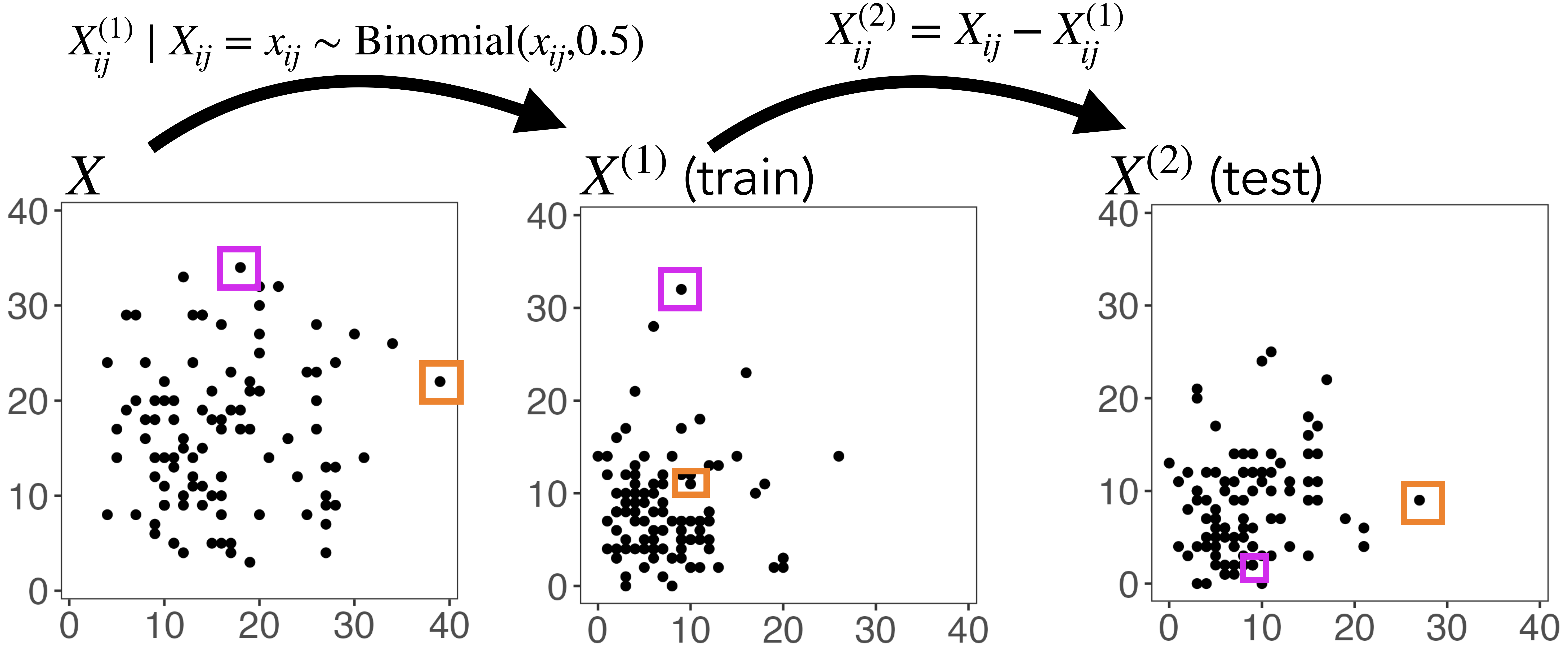
Step 1: thin observations into train/test.

Thinning avoids the pitfall of sample splitting on our motivating example



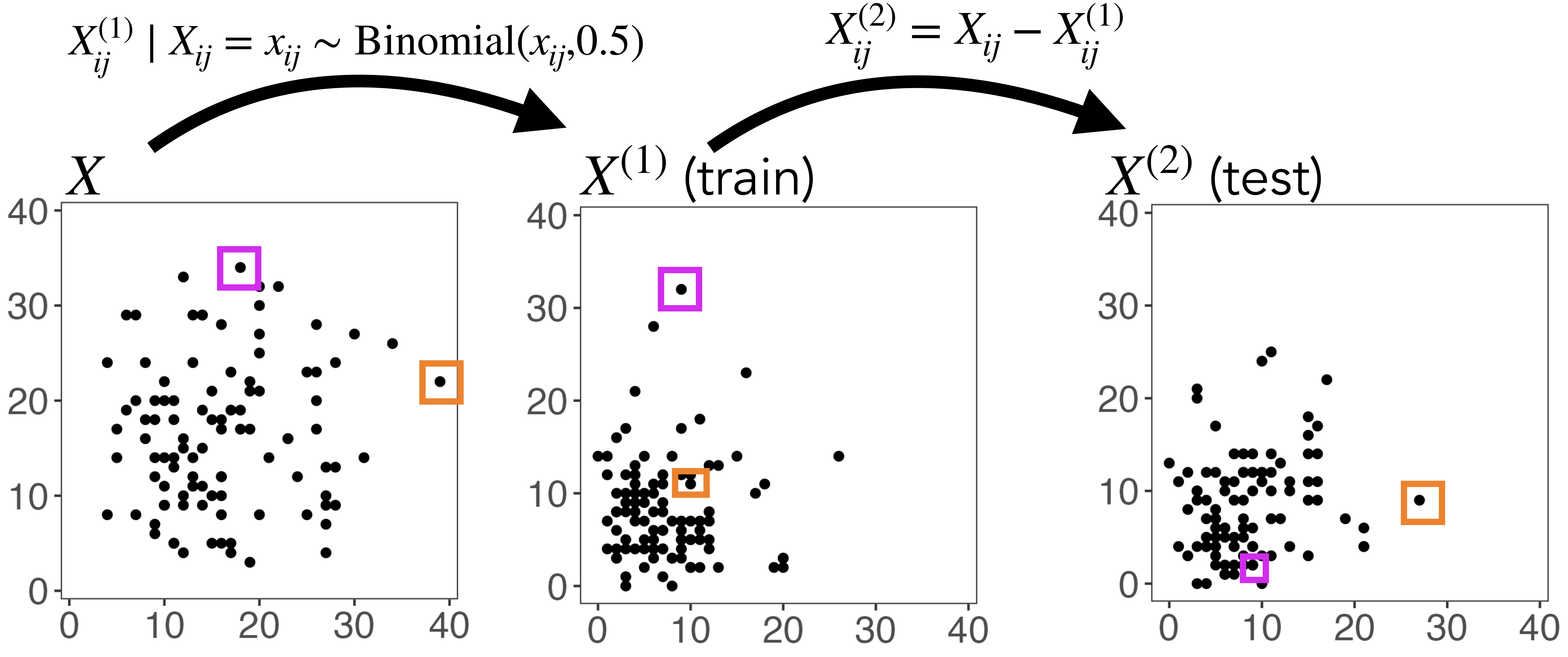
Step 1: thin observations into train/test.

Thinning avoids the pitfall of sample splitting on our motivating example



Step 1: thin observations into train/test.

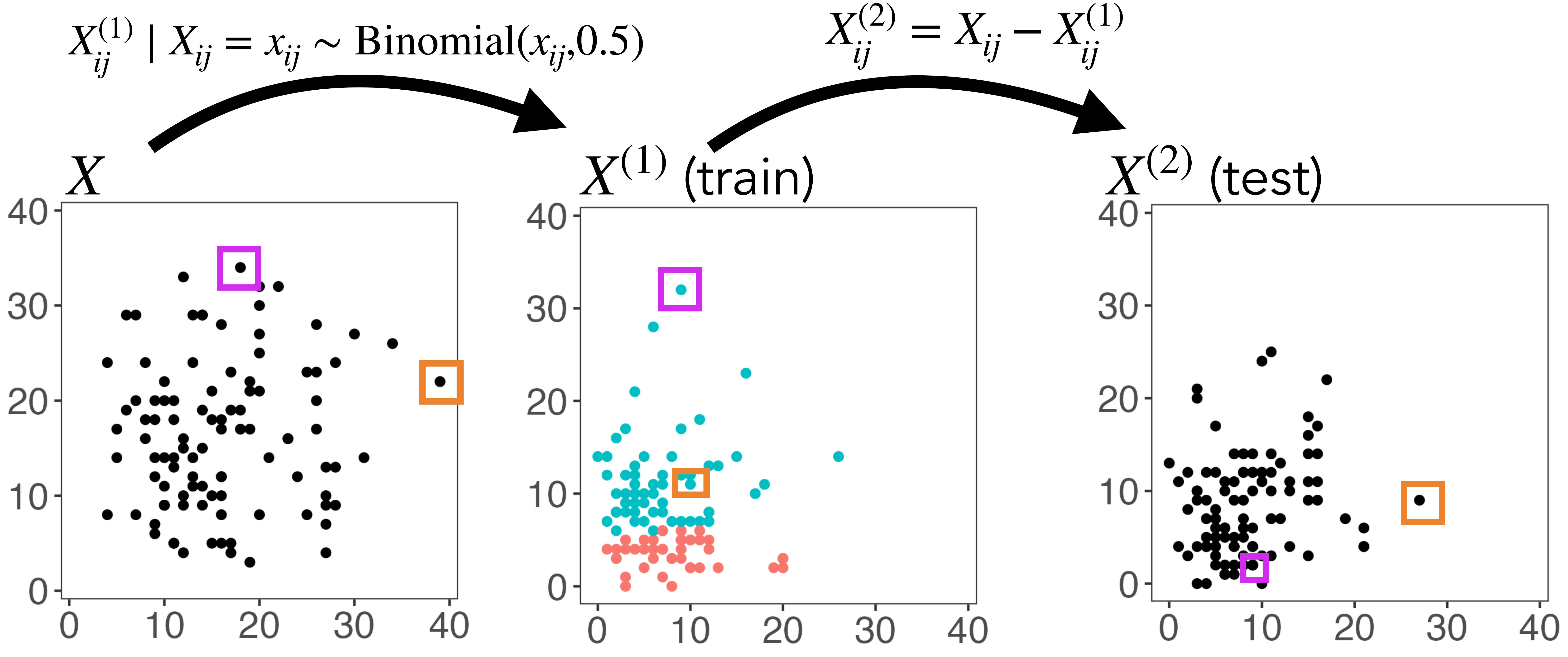
Thinning avoids the pitfall of sample splitting on our motivating example



Step 1: thin observations into train/test.

Step 2: cluster the training set.

Thinning avoids the pitfall of sample splitting on our motivating example



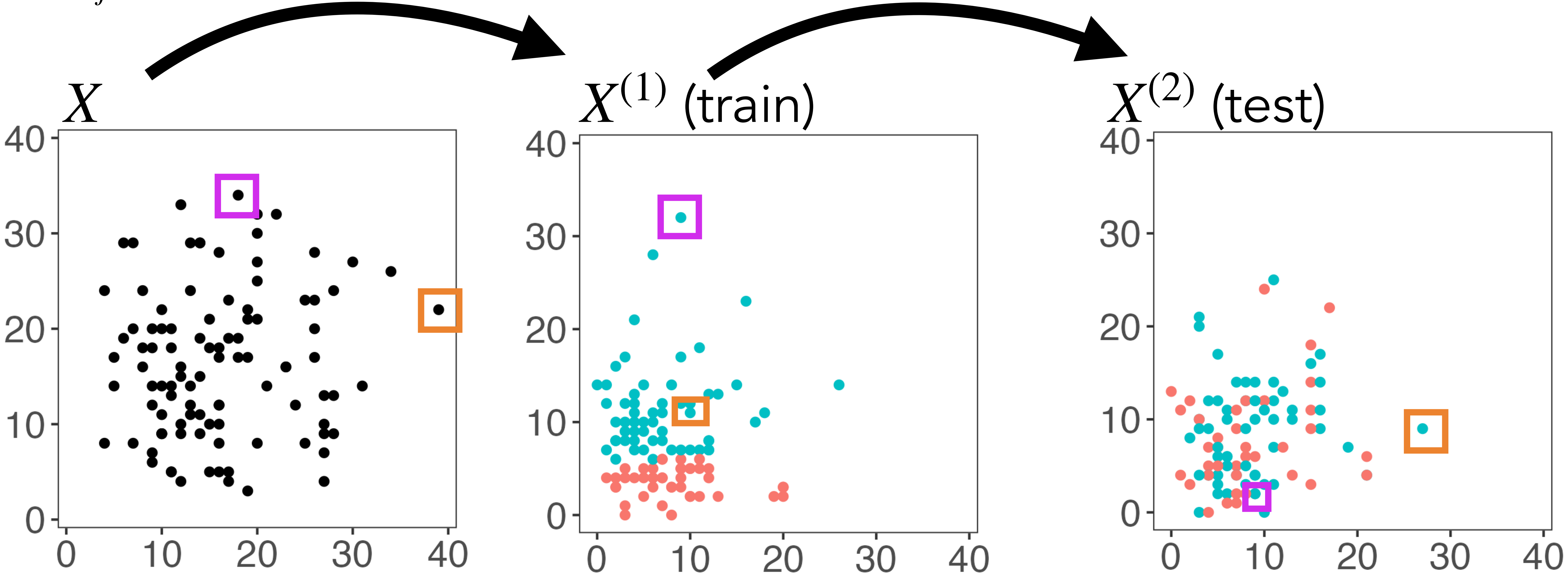
Step 1: thin observations into train/test.

Step 2: cluster the training set.

Thinning avoids the pitfall of sample splitting on our motivating example

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

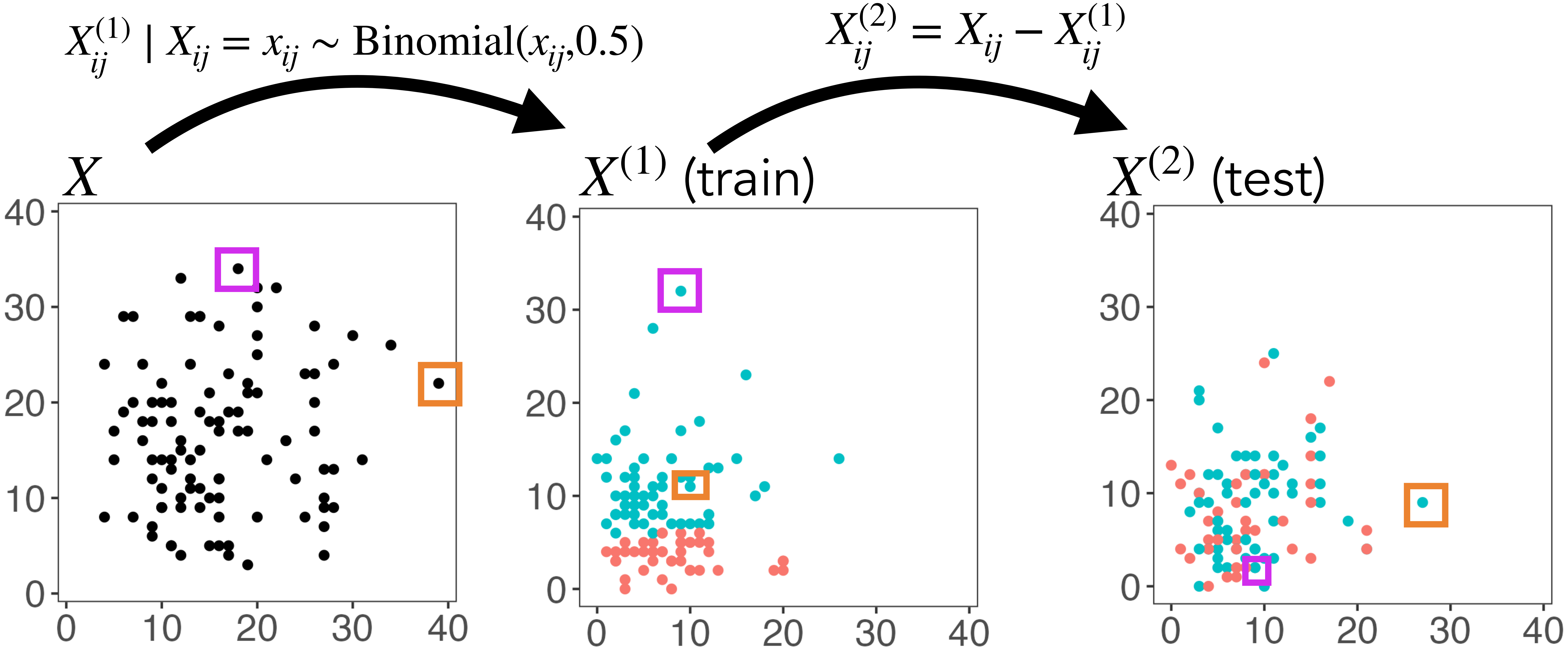
$$X_{ij}^{(2)} = X_{ij} - X_{ij}^{(1)}$$



Step 1: thin observations into train/test.

Step 2: cluster the training set.

Thinning avoids the pitfall of sample splitting on our motivating example



Step 1: thin observations into train/test.

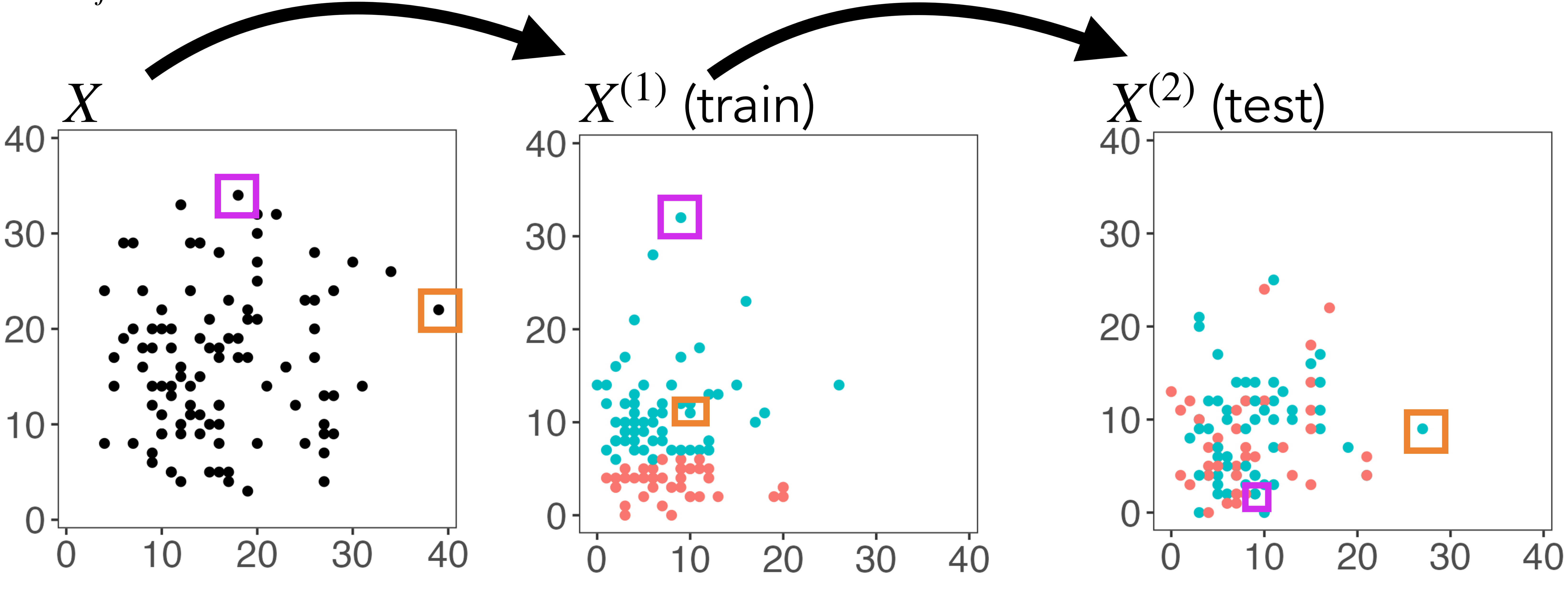
Step 2: cluster the training set.

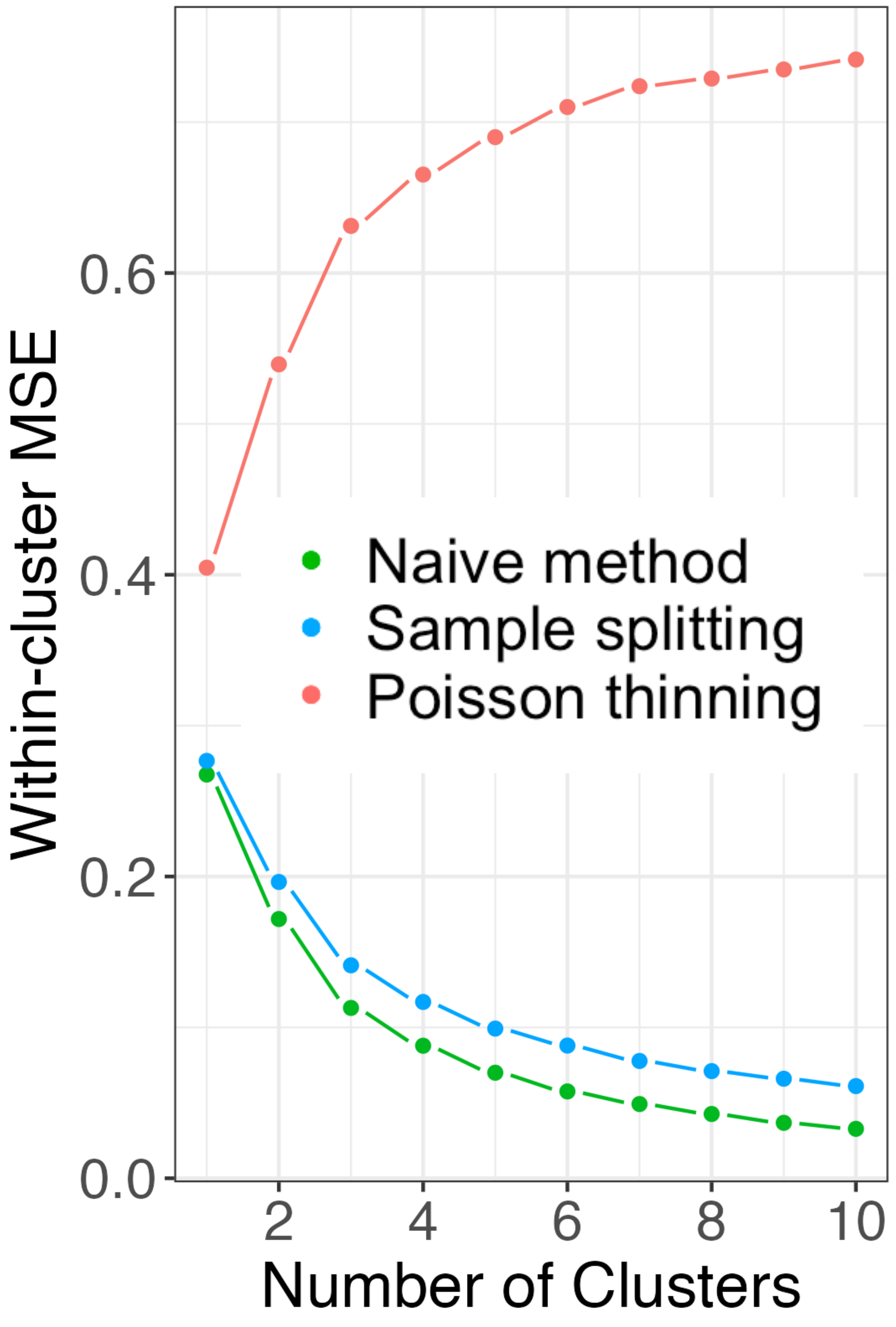
Step 3: evaluate clusters on test set.

Thinning avoids the pitfall of sample splitting on our motivating example

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

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



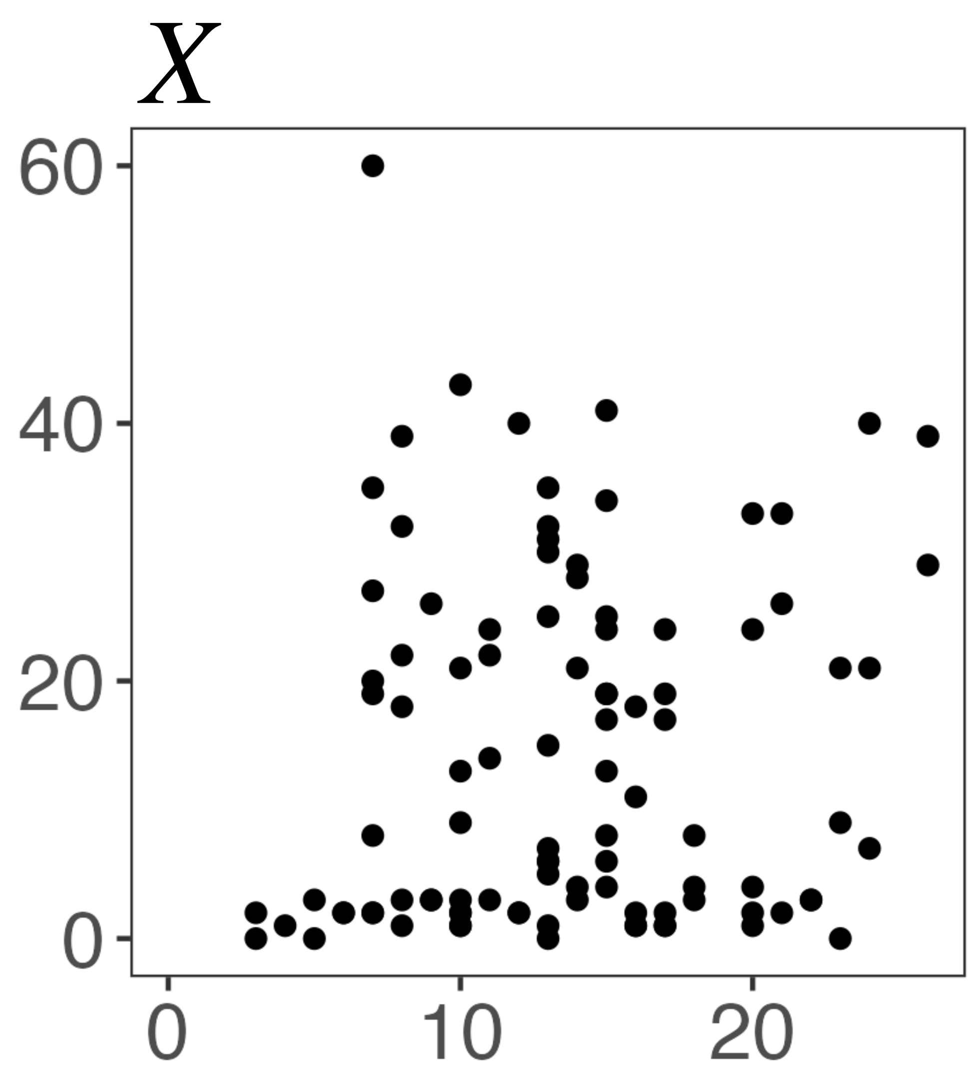


Step 1: thin observations into train/test.

Step 2: cluster the training set.

Step 3: evaluate clusters on test set.

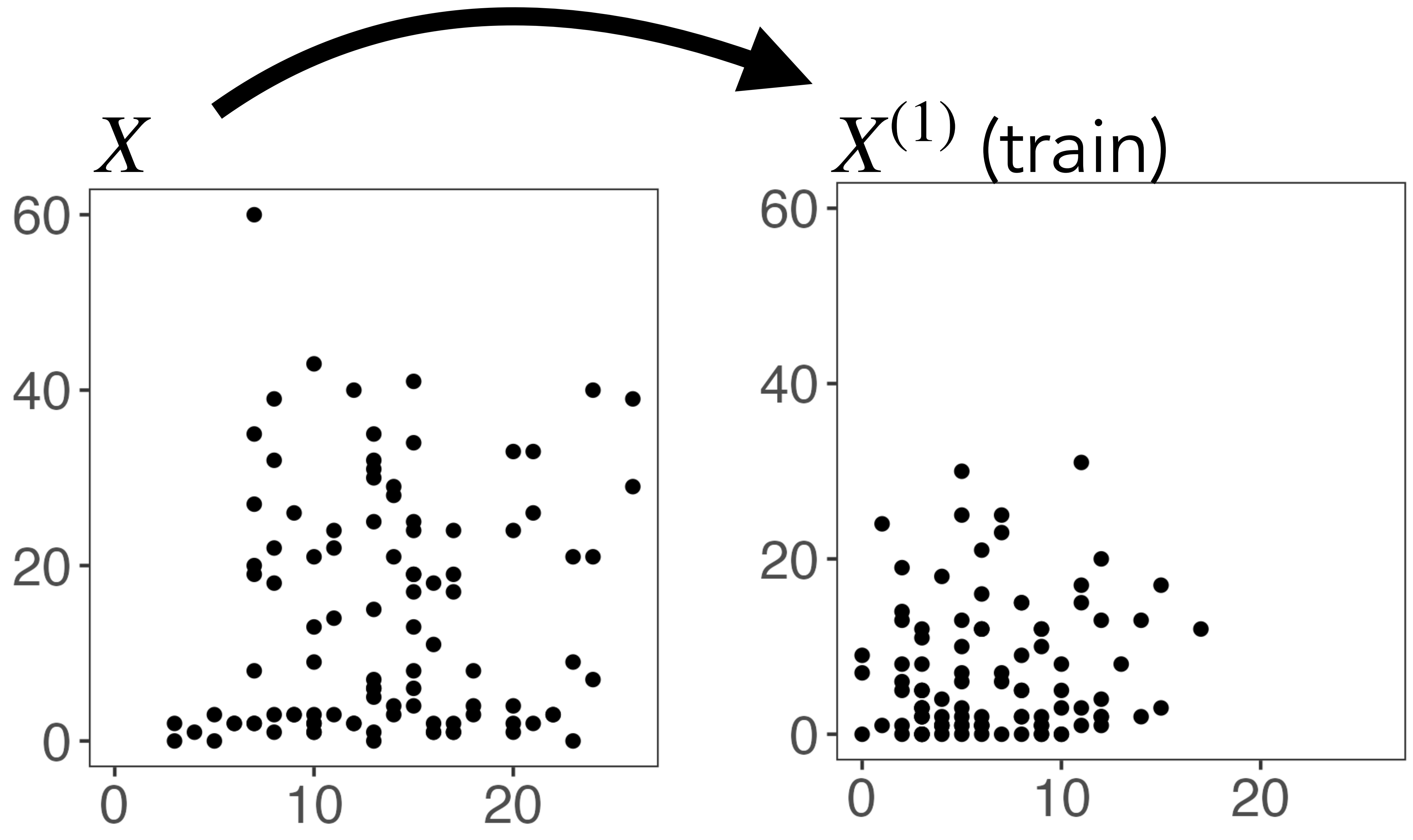
Thinning avoids the pitfall of sample splitting on our motivating examples



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

Thinning avoids the pitfall of sample splitting on our motivating examples

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

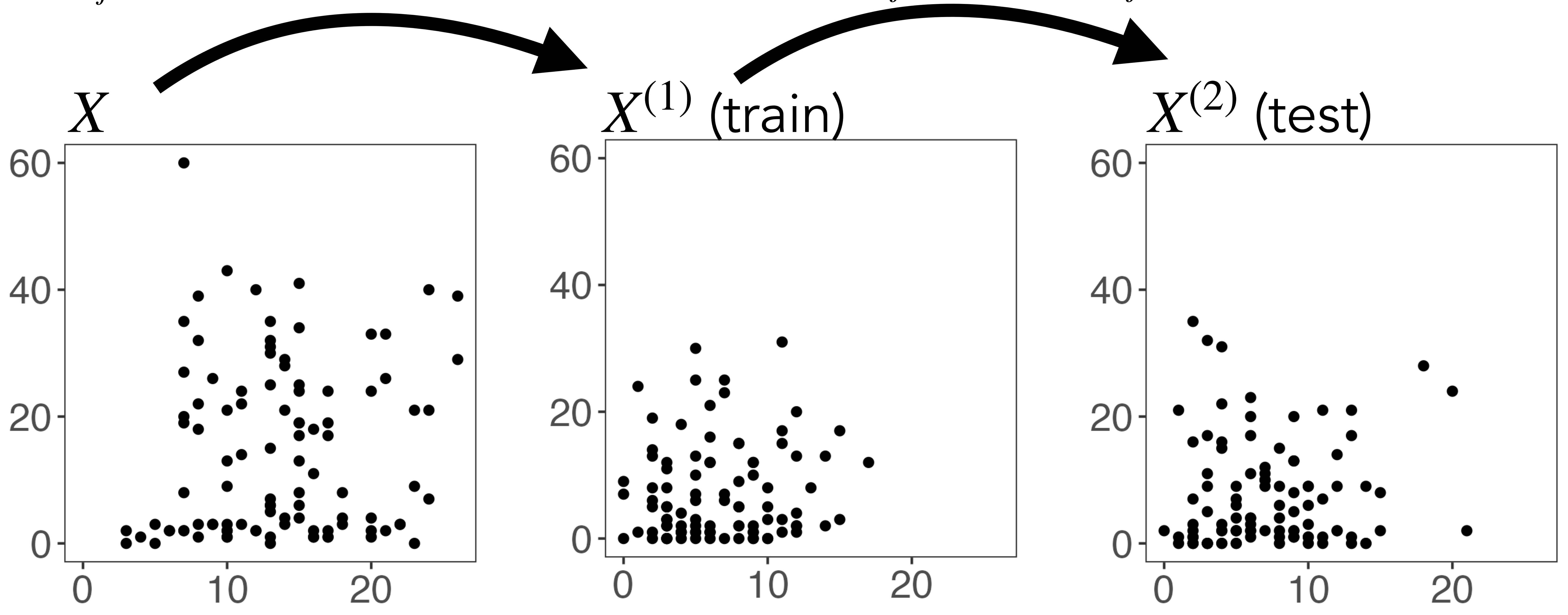


$$X_{i2} \sim \begin{cases} \text{Poisson}(3) & \text{if } i \leq 50 \\ \text{Poisson}(25) & \text{if } i > 50 \end{cases}$$

Thinning avoids the pitfall of sample splitting on our motivating examples

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

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

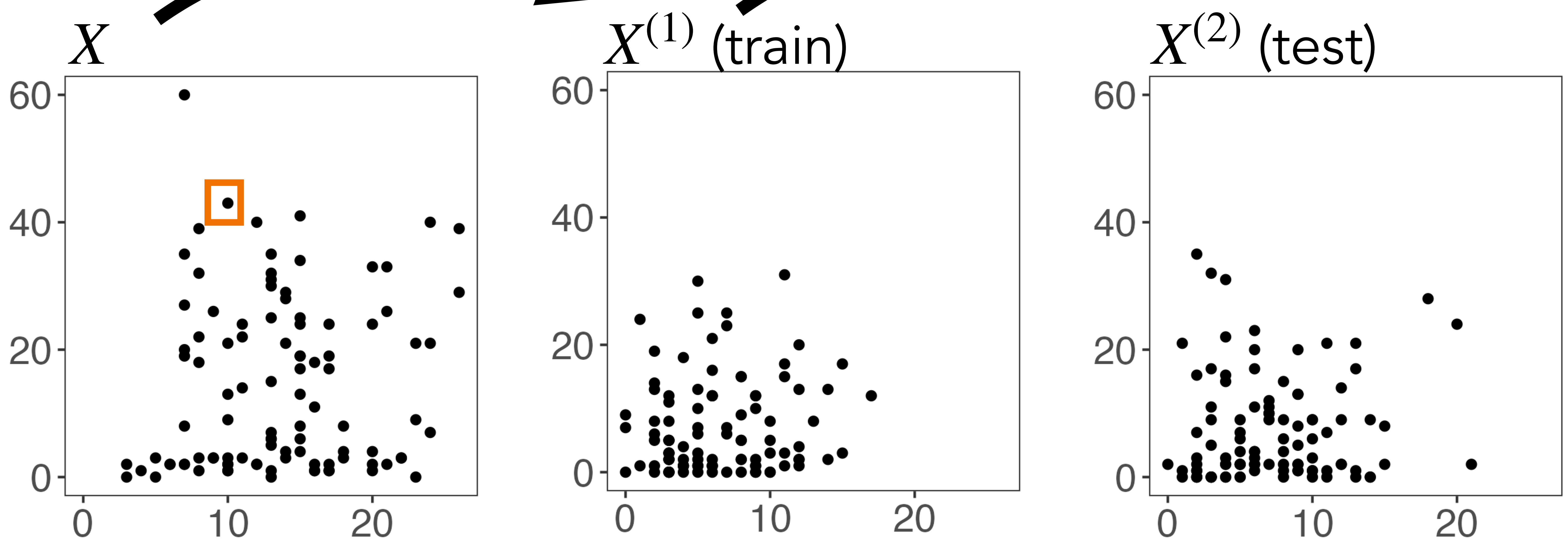


$$X_{i2} \sim \begin{cases} \text{Poisson}(3) & \text{if } i \leq 50 \\ \text{Poisson}(25) & \text{if } i > 50 \end{cases}$$

Thinning avoids the pitfall of sample splitting on our motivating examples

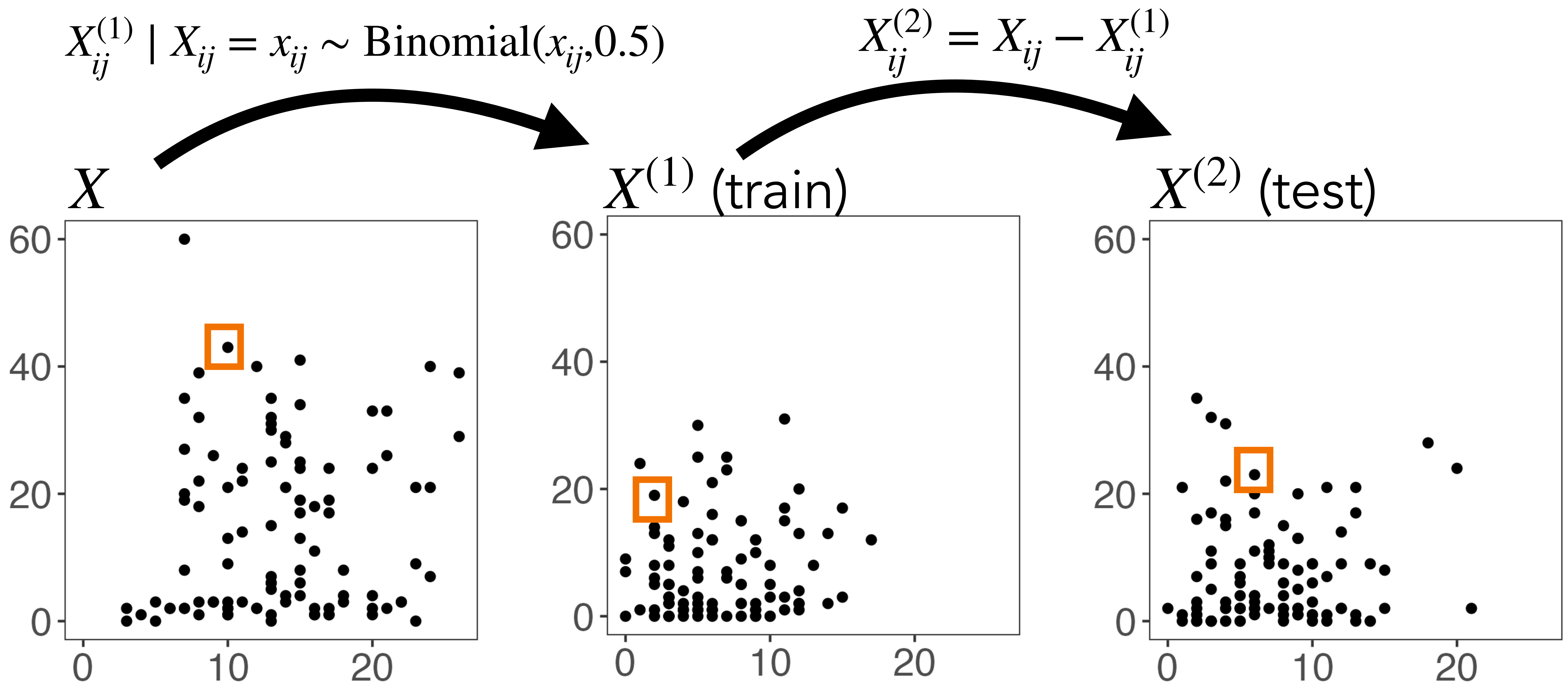
$X^{(1)}_{ij} \mid X_{ij} = x_{ij} \sim \text{Binomial}(x_{ij}, 0.5)$

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



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

Thinning avoids the pitfall of sample splitting on our motivating examples

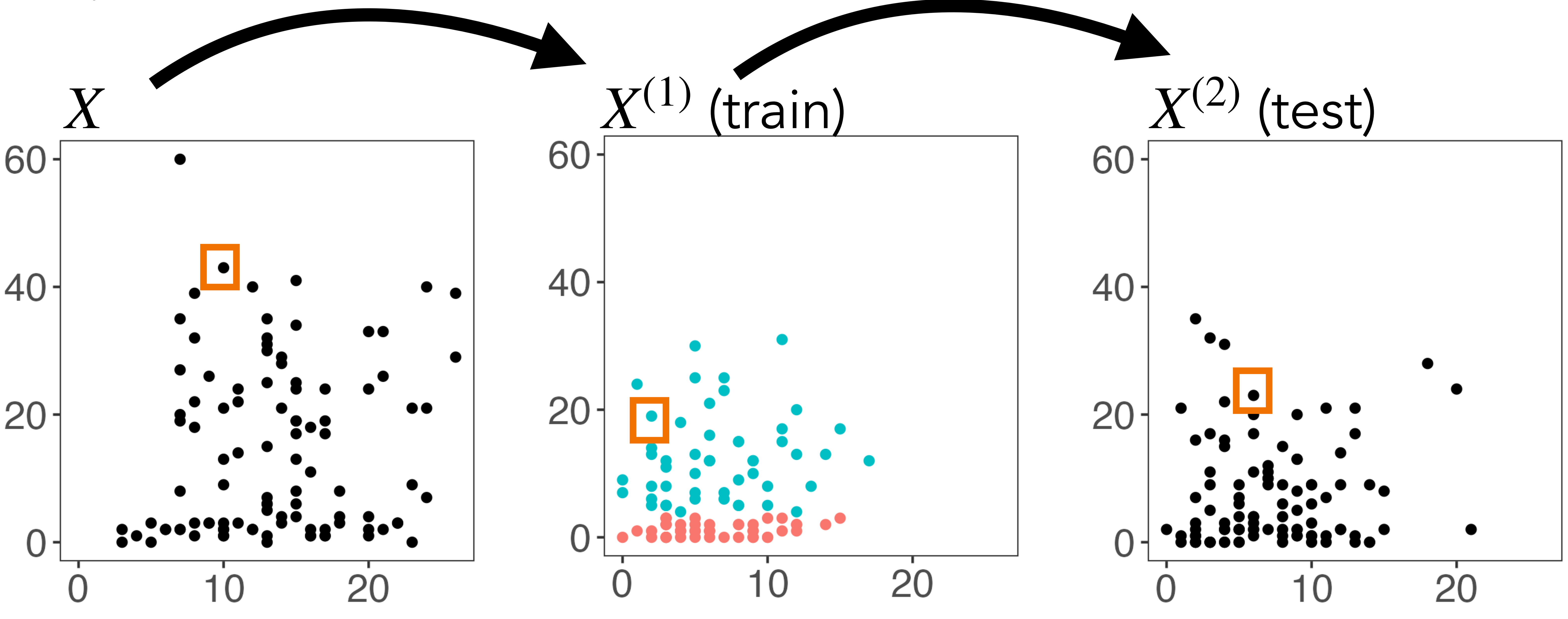


$$X_{i2} \sim \begin{cases} \text{Poisson}(3) & \text{if } i \leq 50 \\ \text{Poisson}(25) & \text{if } i > 50 \end{cases}$$

Thinning avoids the pitfall of sample splitting on our motivating examples

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

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

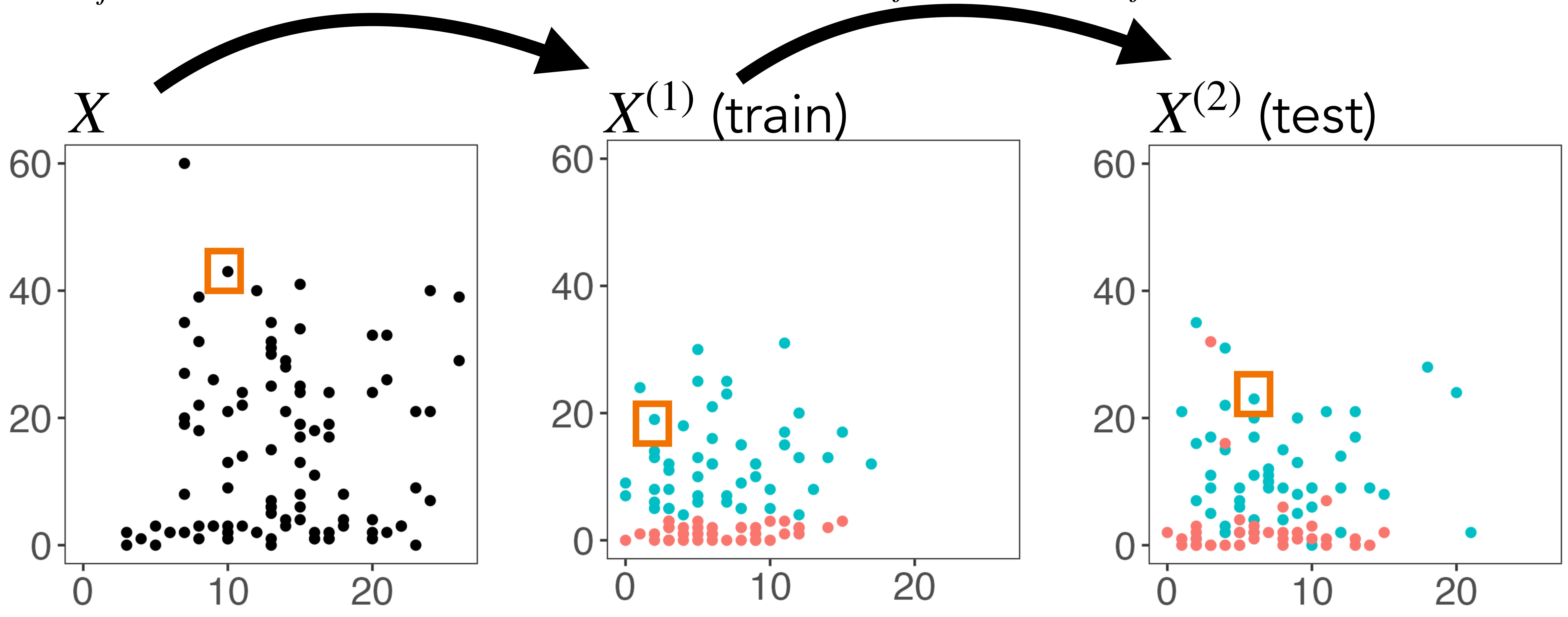


$$X_{i2} \sim \begin{cases} \text{Poisson}(3) & \text{if } i \leq 50 \\ \text{Poisson}(25) & \text{if } i > 50 \end{cases}$$

Thinning avoids the pitfall of sample splitting on our motivating examples

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

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

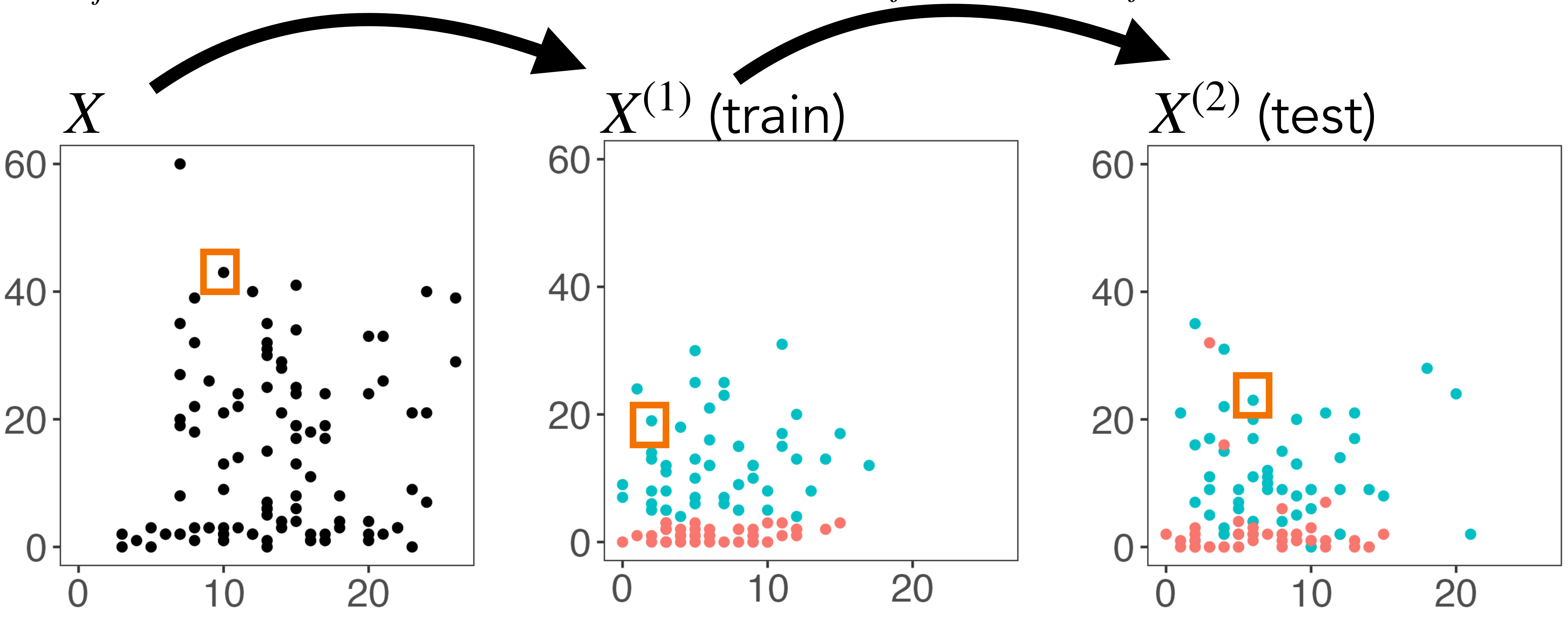


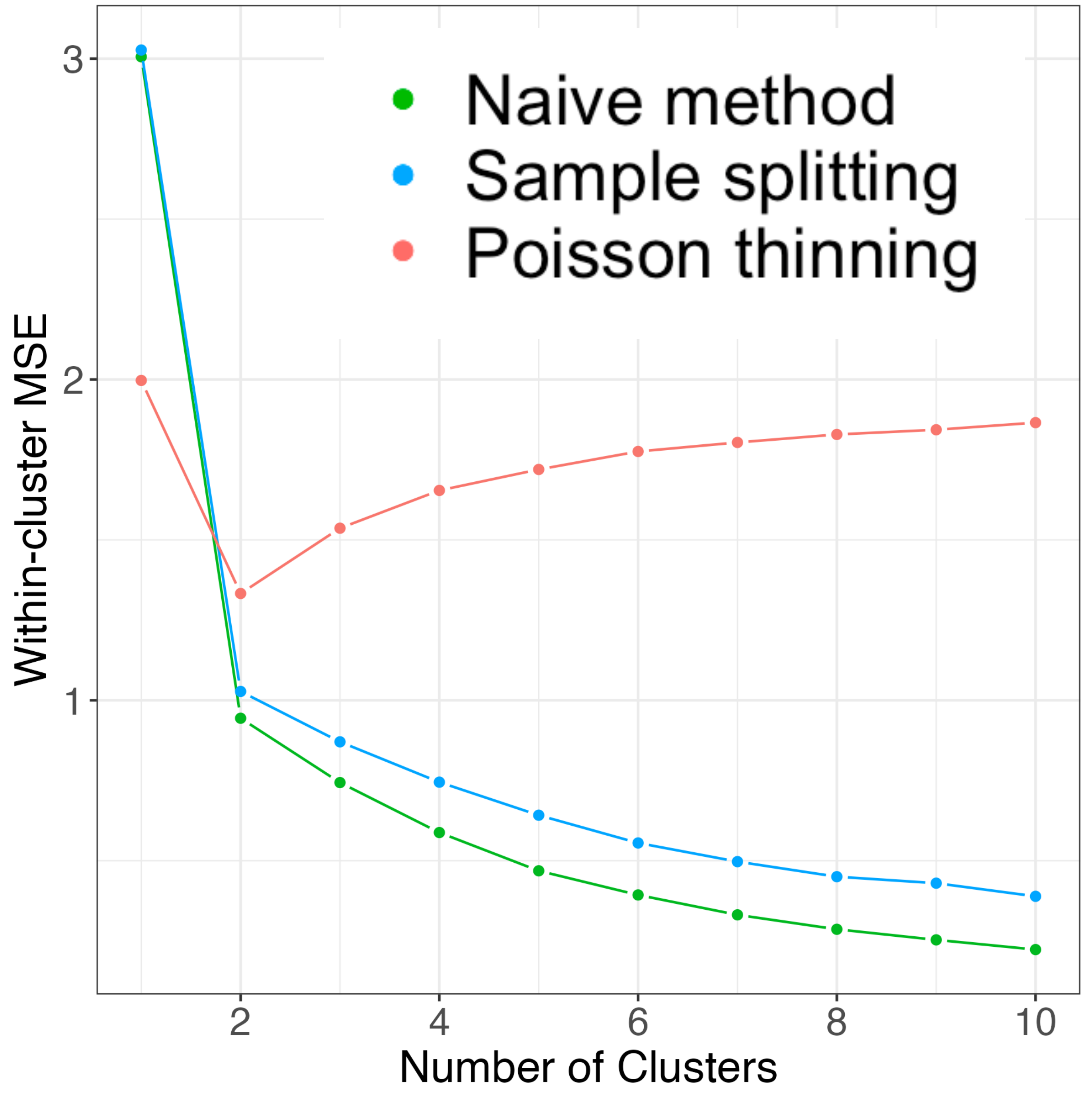
$$X_{i2} \sim \begin{cases} \text{Poisson}(3) & \text{if } i \leq 50 \\ \text{Poisson}(25) & \text{if } i > 50 \end{cases}$$

Thinning avoids the pitfall of sample splitting on our motivating examples

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

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





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

Outline

1. Motivation: sample splitting doesn't always work
2. Poisson thinning
3. **Data thinning**
4. Application to changepoint detection
5. Generalized data thinning

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.

Data thinning for convolution-closed distributions

Data thinning for convolution-closed distributions

$$\boxed{\text{We observe realization } x \text{ from } X \sim F_\lambda.}$$

Data thinning for convolution-closed distributions

We know x could have arisen as $x' + x''$, where
$X' \sim F_{\epsilon\lambda}, \;\; X'' \sim F_{(1-\epsilon)\lambda}, \;\; X' \perp\!\!\!\perp X''$.

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

$\Downarrow$

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

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

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

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

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

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

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

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

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

Draw $X^{(1)}$ 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 for the Poisson distribution

Data thinning for the Poisson distribution

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

Data thinning for the Poisson distribution

We know x could have arisen as $x' + x''$, where
$X' \sim \text{Pois}(\epsilon\lambda),\ \ X'' \sim \text{Pois}((1-\epsilon)\lambda),\ \ X' \perp\!\!\!\perp X''.$

$\downarrow$

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

Data thinning for the Poisson distribution

We know x could have arisen as $x' + x''$, where
$X' \sim \text{Pois}(\epsilon\lambda),\ \ X'' \sim \text{Pois}((1-\epsilon)\lambda),\ \ X' \perp\!\!\!\perp X''$.

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

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

Data thinning for the Poisson distribution

We know x could have arisen as $x' + x''$, where $X' \sim \text{Pois}(\epsilon\lambda),\ \ X'' \sim \text{Pois}((1-\epsilon)\lambda),\ \ X' \perp\!\!\!\perp X''$.

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

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 the Poisson distribution

We know x could have arisen as $x' + x''$, where $X' \sim \text{Pois}(\epsilon\lambda),\ X'' \sim \text{Pois}((1-\epsilon)\lambda),\ X' \perp\!\!\!\perp X''$.

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

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''?

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

Data thinning for the Poisson distribution

We know x could have arisen as $x' + x''$, where $X' \sim \text{Pois}(\epsilon\lambda),\ \ X'' \sim \text{Pois}((1-\epsilon)\lambda),\ \ X' \perp\!\!\!\perp X''$.

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

Draw $X^{(1)}$ from $\text{Binomial}(x, \epsilon)$. 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''?

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

Data thinning for the Poisson distribution

We know x could have arisen as $x' + x''$, where $X' \sim \text{Pois}(\epsilon\lambda),\ X'' \sim \text{Pois}((1-\epsilon)\lambda),\ X' \perp\!\!\!\perp X''$.

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

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

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

Draw $X^{(1)}$ from $\text{Binomial}(x, \epsilon)$. Let $X^{(2)} := X - X^{(1)}$.

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

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

Data thinning for the Poisson distribution

We know x could have arisen as $x' + x''$, where $X' \sim \text{Pois}(\epsilon\lambda),\ \ X'' \sim \text{Pois}((1-\epsilon)\lambda),\ \ X' \perp\!\!\!\perp X''$.

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

Draw $X^{(1)}$ from $\text{Binomial}(x, \epsilon)$. Let $X^{(2)} := X - X^{(1)}$.

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

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''?

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

We have recovered Poisson thinning!

Data thinning recipe for the negative binomial distribution

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

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

Data thinning recipe for the negative binomial distribution

We know x could have arisen as $x' + x''$, where $X' \sim \mathrm{NB}(\epsilon\mu, \epsilon b), X'' \sim \mathrm{NB}((1-\epsilon)\mu, (1-\epsilon)b), X' \perp\!\!\!\perp X''$.

$\downarrow$

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

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

Data thinning recipe for the negative binomial distribution

We know x could have arisen as $x' + x''$, where $X' \sim \text{NB}(\epsilon\mu, \epsilon b), X'' \sim \text{NB}((1-\epsilon)\mu, (1-\epsilon)b), X' \perp\!\!\!\perp X''$.

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

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 recipe for the negative binomial distribution

We know x could have arisen as $x' + x''$, where $X' \sim \mathrm{NB}(\epsilon\mu, \epsilon b), X'' \sim \mathrm{NB}((1-\epsilon)\mu, (1-\epsilon)b), X' \perp\!\!\!\perp X''$.

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

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''?

The conditional distribution of $X' \mid X = x$ is $\mathrm{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.

Data thinning recipe for the negative binomial distribution

We know x could have arisen as $x' + x''$, where
$X' \sim \mathrm{NB}(\epsilon\mu, \epsilon b), X'' \sim \mathrm{NB}((1-\epsilon)\mu, (1-\epsilon)b), X' \perp\!\!\!\perp X''$.

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

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

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

Draw $X^{(1)}$ from $\mathrm{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.
Let $X^{(2)} := X - X^{(1)}$.

The conditional distribution of $X' \mid X = x$ is $\mathrm{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.

Data thinning recipe for the negative binomial distribution

We know x could have arisen as $x' + x''$, where $X' \sim \mathrm{NB}(\epsilon\mu, \epsilon b), X'' \sim \mathrm{NB}((1-\epsilon)\mu, (1-\epsilon)b), X' \perp\!\!\!\perp X''$.

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

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

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

Draw $X^{(1)}$ from $\mathrm{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$. Let $X^{(2)} := X - X^{(1)}$.

The conditional distribution of $X' \mid X = x$ is $\mathrm{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.

Theorem:

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

Data thinning recipe for the negative binomial distribution

We know x could have arisen as $x' + x''$, where $X' \sim \mathrm{NB}(\epsilon\mu, \epsilon b), X'' \sim \mathrm{NB}((1-\epsilon)\mu, (1-\epsilon)b), X' \perp\!\!\!\perp X''$.

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

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

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

Draw $X^{(1)}$ from $\mathrm{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$. Let $X^{(2)} := X - X^{(1)}$.

The conditional distribution of $X' \mid X = x$ is $\mathrm{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.

Theorem:

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

This is a new result!

Data thinning recipe for the negative binomial distribution

We know x could have arisen as $x' + x''$, where $X' \sim \mathrm{NB}(\epsilon\mu, \epsilon b), X'' \sim \mathrm{NB}((1-\epsilon)\mu, (1-\epsilon)b), X' \perp\!\!\!\perp X''$.

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

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

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

Draw $X^{(1)}$ from $\mathrm{BetaBinomial}(x, \epsilon\tilde{b}, (1-\epsilon)\tilde{b})$. Let $X^{(2)} := X - X^{(1)}$.

The conditional distribution of $X' \mid X = x$ is $\mathrm{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.

Theorem:

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

This is a new result!

Data thinning recipe for the negative binomial distribution

We know x could have arisen as $x' + x''$, where
$X' \sim \mathrm{NB}(\epsilon\mu, \epsilon b), X'' \sim \mathrm{NB}((1-\epsilon)\mu, (1-\epsilon)b), X' \perp\!\!\!\perp X''$.

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

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

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

Draw $X^{(1)}$ from $\mathrm{BetaBinomial}(x, \epsilon\tilde{b}, (1-\epsilon)\tilde{b})$.
Let $X^{(2)} := X - X^{(1)}$.

The conditional distribution of $X' \mid X = x$ is $\mathrm{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.

This is a new result!

Data thinning recipe for the negative binomial distribution

We know x could have arisen as $x' + x''$, where $X' \sim \text{NB}(\epsilon\mu, \epsilon b), X'' \sim \text{NB}((1-\epsilon)\mu, (1-\epsilon)b), X' \perp\!\!\!\perp X''$.

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

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

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

Draw $X^{(1)}$ from $\text{BetaBinomial}(x, \epsilon\tilde{b}, (1-\epsilon)\tilde{b})$. Let $X^{(2)} := X - X^{(1)}$.

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

$$\text{Cov}\left(X^{(1)}, X^{(2)}\right) = \epsilon(1-\epsilon)\frac{\mu^2}{b}\left(1 - \frac{b+1}{\tilde{b}+1}\right).$$

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 $X^{(2)} = X - X^{(1)}$:
Poisson(λ)	Binomial(x, ϵ)	Poisson$(\epsilon\lambda)$	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 $X^{(2)} = X - X^{(1)}$:
Poisson(λ)	Binomial(x, ϵ)	Poisson$(\epsilon\lambda)$	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 $X^{(2)} = X - X^{(1)}$:
$\mathrm{Poisson}(\lambda)$	$\mathrm{Binomial}(x, \epsilon)$	$\mathrm{Poisson}(\epsilon\lambda)$	$\mathrm{Poisson}((1-\epsilon)\lambda)$
$\mathrm{N}(\mu, \sigma^2)$	$\mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\sigma^2)$	$\mathrm{N}(\epsilon\mu, \epsilon\sigma^2)$	$\mathrm{N}((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$

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.
- Tian, 2020, Annals of Statistics.
- Rasines and Young, 2022, Biometrika.
- Leiner et al., 2022, arXiv:2112.11079.
- Oliveira, Lei, and Tibshirani, 2022, 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 $X^{(2)} = X - X^{(1)}$:
$\mathrm{Poisson}(\lambda)$	$\mathrm{Binomial}(x, \epsilon)$	$\mathrm{Poisson}(\epsilon\lambda)$	$\mathrm{Poisson}((1-\epsilon)\lambda)$
$\mathrm{N}(\mu, \sigma^2)$	$\mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\sigma^2)$	$\mathrm{N}(\epsilon\mu, \epsilon\sigma^2)$	$\mathrm{N}((1-\epsilon)\mu, (1-\epsilon)\sigma^2)$
$\mathrm{NegativeBinomial}(\mu, b)$	$\mathrm{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.	$\mathrm{NegativeBinomial}(\epsilon\mu, \epsilon b)$	$\mathrm{NegativeBinomial}((1-\epsilon)\mu, (1-\epsilon)b)$
$\mathrm{Binomial}(r, p)$	$\mathrm{Hypergeometric}(\epsilon r, (1-\epsilon)r, x)$.	$\mathrm{Binomial}(\epsilon r, p)$	$\mathrm{Binomial}((1-\epsilon)r, p)$
$\mathrm{Gamma}(\alpha, \beta)$	$x \cdot \mathrm{Beta}\,(\epsilon\alpha, (1-\epsilon)\alpha)$.	$\mathrm{Gamma}(\epsilon\alpha, \beta)$	$\mathrm{Gamma}((1-\epsilon)\alpha, \beta)$
$\mathrm{Exponential}(\lambda)$	$x \cdot \mathrm{Beta}(\epsilon, (1-\epsilon))$.	$\mathrm{Gamma}(\epsilon, \lambda)$	$\mathrm{Gamma}(1-\epsilon, \lambda)$
$\mathrm{N}_k(\mu, \Sigma)$	$\mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\Sigma)$.	$\mathrm{N}_k(\epsilon\mu, \epsilon\Sigma)$	$\mathrm{N}_k((1-\epsilon)\mu, (1-\epsilon)\Sigma)$
$\mathrm{Multinomial}_k(r, p)$	$\mathrm{MultivarHypergeom}(x_1, \ldots, x_K, \epsilon r)$	$\mathrm{Multinom}_k(\epsilon r, p)$	$\mathrm{Multinomial}_k((1-\epsilon)r, p)$
$\mathrm{Wishart}_p(n, \Sigma)$.	$x^{1/2} Z x^{1/2}$, where . $Z \sim \mathrm{MatrixBeta}_p(\epsilon n/2, (1-\epsilon)n/2)$	$\mathrm{Wishart}_p(\epsilon n, \Sigma)$	$\mathrm{Wishart}_p((1-\epsilon)n, \Sigma)$

Our recipe extends naturally to splitting into M>2 folds

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.

Our recipe extends naturally to splitting into M>2 folds

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

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.

Our recipe extends naturally to splitting into M>2 folds

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

Our recipe extends naturally to splitting into M>2 folds

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

Our recipe extends naturally to splitting into M>2 folds

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

Outline

1. Motivation: sample splitting doesn't always work
2. Poisson thinning
3. Data thinning
4. **Application to changepoint detection**
5. Generalized data thinning

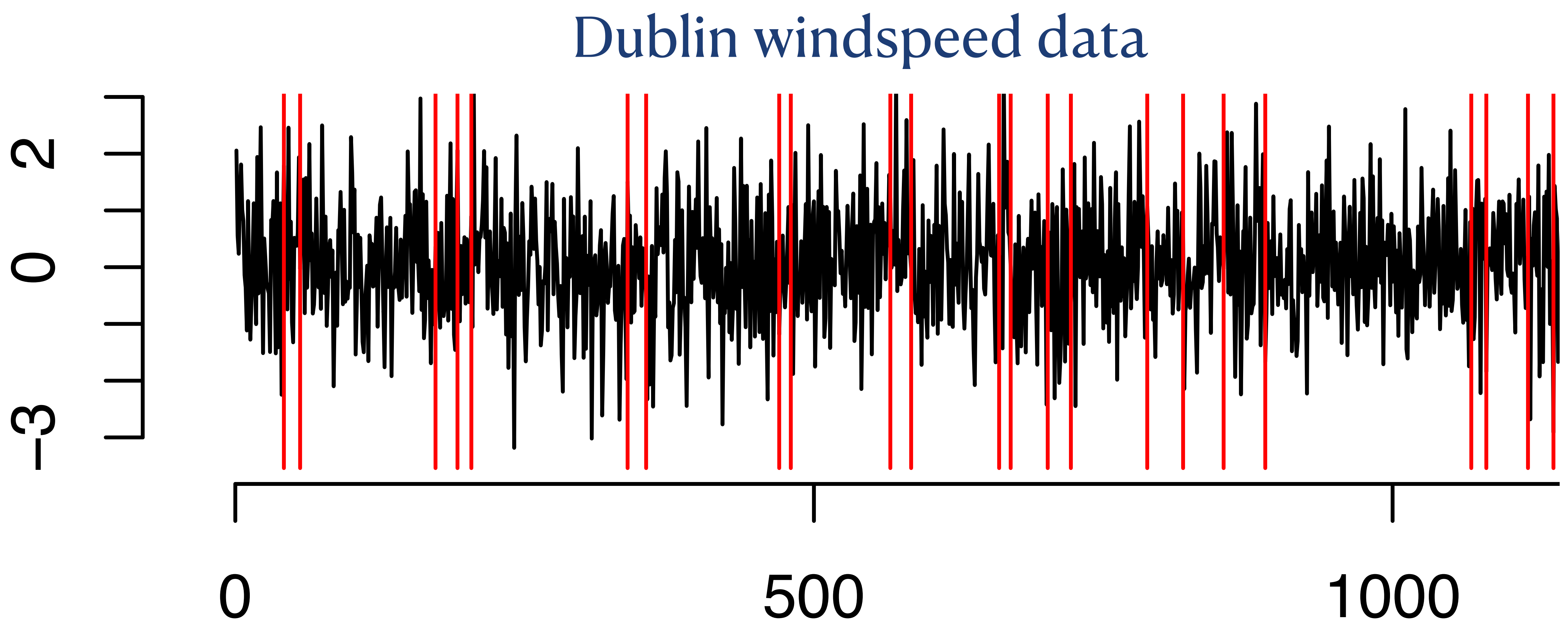
Application: changepoint detection

Application: changepoint detection

Settings: Previous authors modeled differences in wind speeds as $X_i \sim N(0, \theta_i)$ for times $i = 1, \ldots, T$, and estimated points in the sequence where the variance changes.

Application: changepoint detection

Settings: Previous authors modeled differences in wind speeds as $X_i \sim N(0, \theta_i)$ for times $i = 1, \ldots, T$, and estimated points in the sequence where the variance changes.

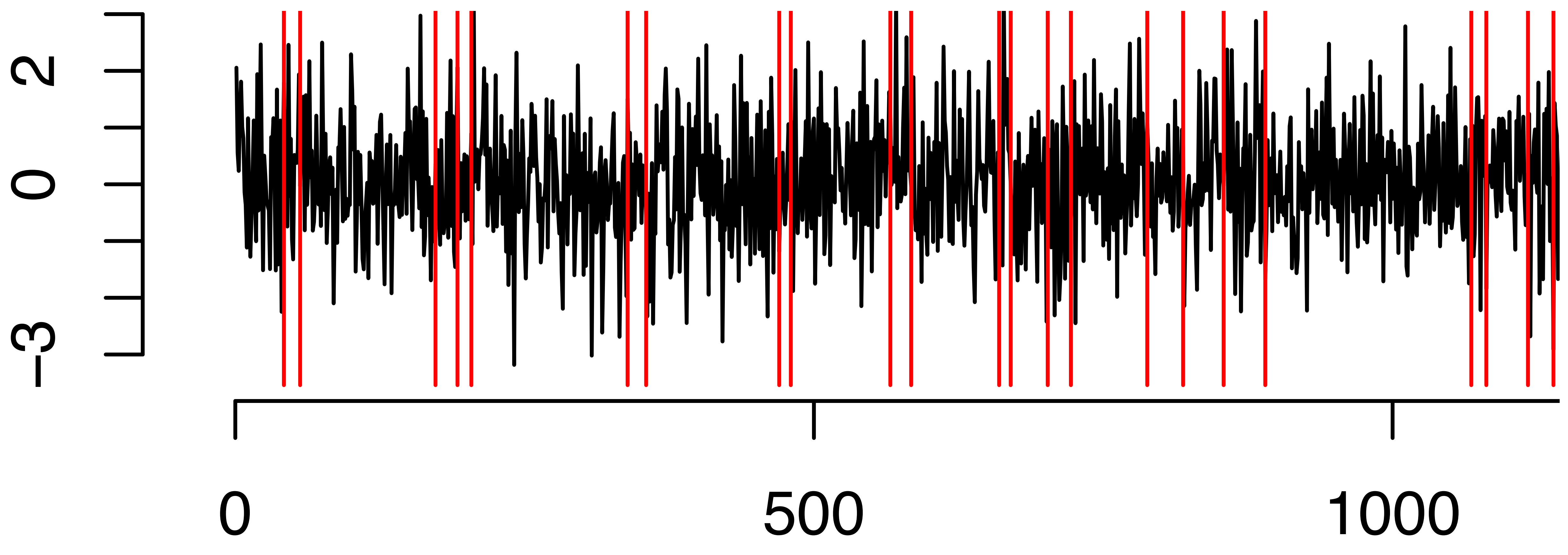


Application: changepoint detection

Settings: Previous authors modeled differences in wind speeds as $X_i \sim N(0, \theta_i)$ for times $i = 1, \ldots, T$, and estimated points in the sequence where the variance changes.

Goal: Can we validate these estimated changepoints?

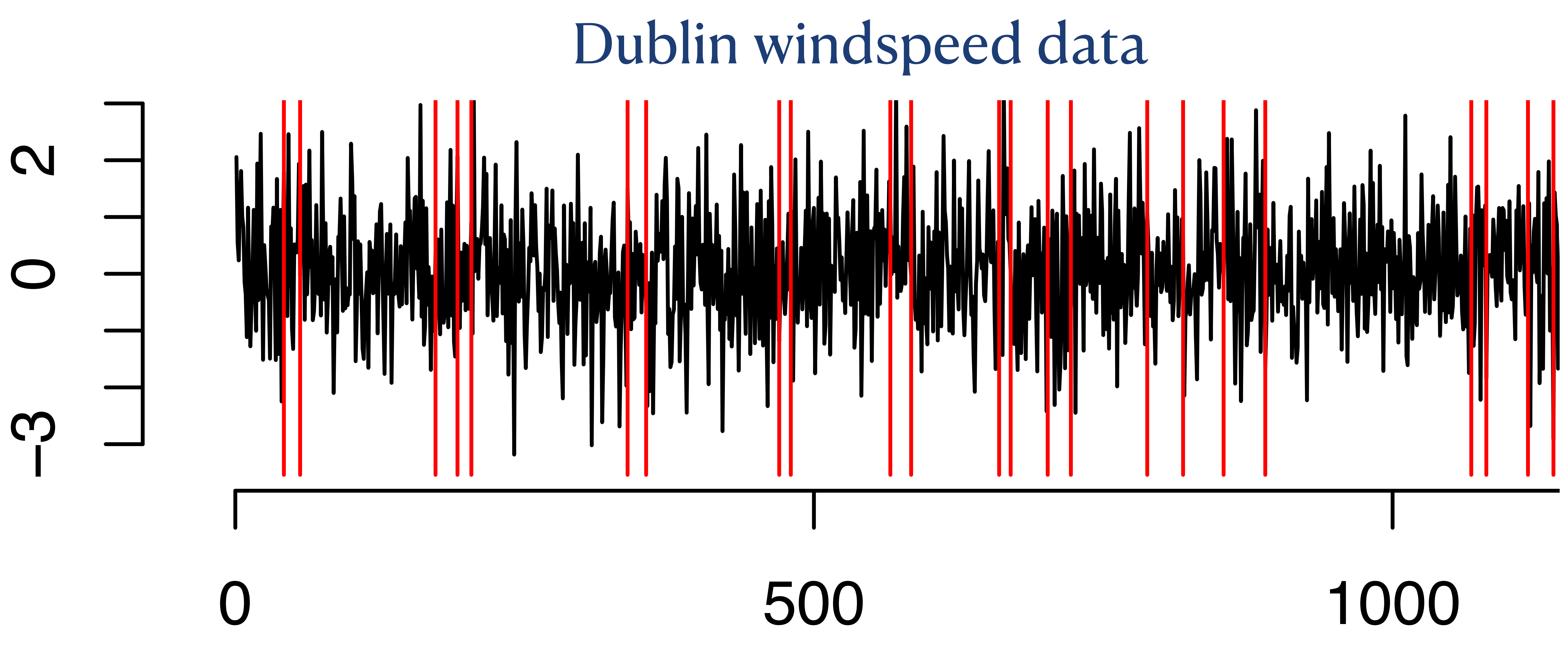
Dublin windspeed data



Application: changepoint detection

Settings: Previous authors modeled differences in wind speeds as $X_i \sim N(0, \theta_i)$ for times $i = 1, \ldots, T$, and estimated points in the sequence where the variance changes.

Goal: Can we validate these estimated changepoints?



Idea: thin the data, such that we can estimate changepoints on training set and evaluate significance on test set.

We do not know how to thin a Gaussian with unknown variance

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 . $Z \sim \text{MatrixBeta}_p(\epsilon n/2, (1-\epsilon)n/2)$	$\text{Wishart}_p(\epsilon n, \Sigma)$	$\text{Wishart}_p((1-\epsilon)n, \Sigma)$

We do not know how to thin a Gaussian with unknown variance

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 $X^{(2)} = X - X^{(1)}$:
$\mathrm{Poisson}(\lambda)$	$\mathrm{Binomial}(x, \epsilon)$	$\mathrm{Poisson}(\epsilon\lambda)$	$\mathrm{Poisson}((1-\epsilon)\lambda)$
$\boxed{\mathrm{N}(\mu, \sigma^2)}$	$\boxed{\mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\sigma^2)}$	$\mathrm{N}(\epsilon\mu, \epsilon\sigma^2)$	[illegible]
$\mathrm{NegativeBinomial}(\mu, b)$	$\mathrm{BetaBinomial}(x, \epsilon b, (1-\epsilon)b)$.	NegativeBinomial([illegible]	[illegible] $1-\epsilon)b)$
$\mathrm{Binomial}(r, p)$	$\mathrm{Hypergeometric}(\epsilon r, (1-\epsilon)r, x)$.	$\mathrm{Binomial}(\epsilon r, p)$	[illegible]
$\mathrm{Gamma}(\alpha, \beta)$	$x \cdot \mathrm{Beta}\,(\epsilon\alpha, (1-\epsilon)\alpha)$.	$\mathrm{Gamma}(\epsilon\alpha, \beta)$	[illegible]
$\mathrm{Exponential}(\lambda)$	$x \cdot \mathrm{Beta}(\epsilon, (1-\epsilon))$.	$\mathrm{Gamma}(\epsilon, \lambda)$	[illegible]
$\mathrm{N}_k(\mu, \Sigma)$	$\mathrm{N}(\epsilon x, \epsilon(1-\epsilon)\Sigma)$.	$\mathrm{N}_k(\epsilon\mu, \epsilon\Sigma)$	[illegible]
$\mathrm{Multinomial}_k(r, p)$	$\mathrm{MultivarHypergeom}(x_1, \ldots, x_K, \epsilon r)$	$\mathrm{Multinom}_k(\epsilon r, p)$	$\mathrm{Multinomial}_k((1-\epsilon)r, p)$
$\mathrm{Wishart}_p(n, \Sigma)$.	$x^{1/2} Z x^{1/2}$, where . $Z \sim \mathrm{MatrixBeta}_p(\epsilon n/2, (1-\epsilon)n/2)$	$\mathrm{Wishart}_p(\epsilon n, \Sigma)$	$\mathrm{Wishart}_p((1-\epsilon)n, \Sigma)$

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

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

We do not know how to thin a Gaussian with unknown variance

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

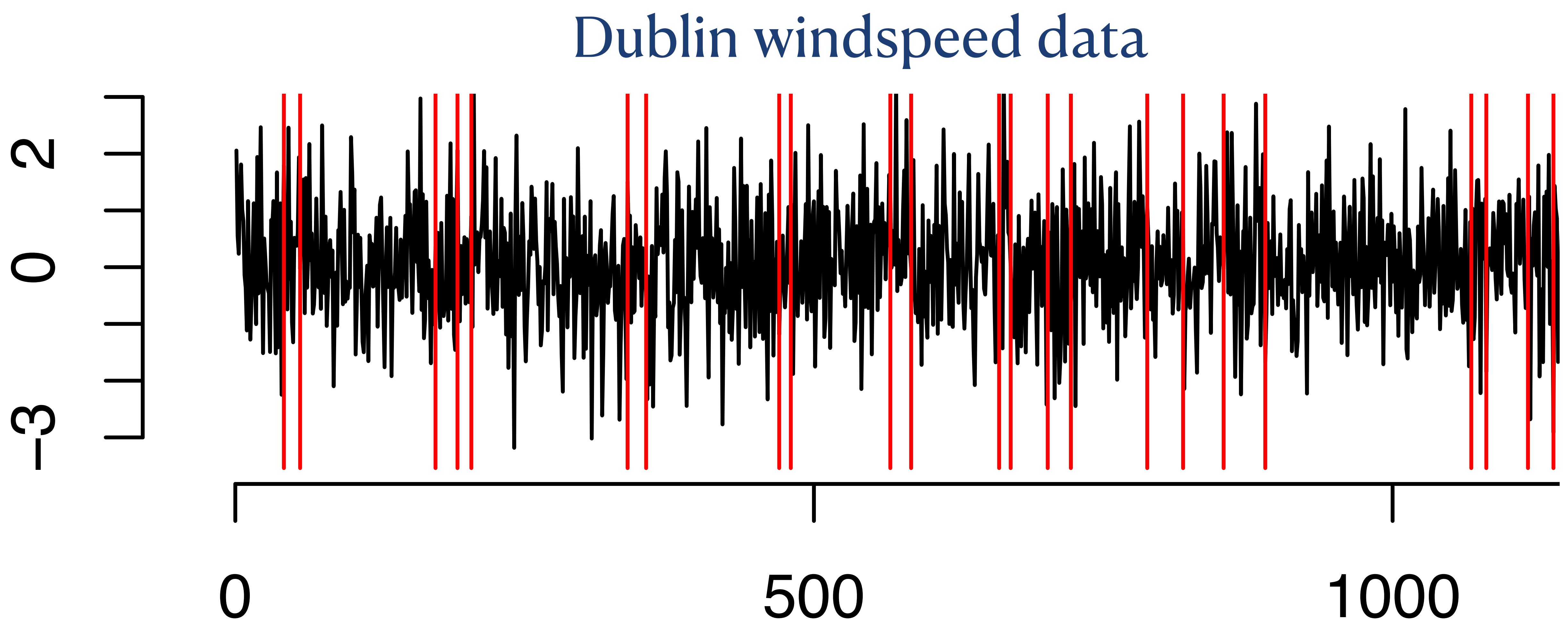
Fact: if $X_i \sim N(0, \theta_i)$, then

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

Application: changepoint detection

Settings: Previous authors modeled differences in wind speeds as $X_i \sim N(0, \theta_i)$ for times $i = 1, \ldots, T$, and estimated points in the sequence where the variance changes.

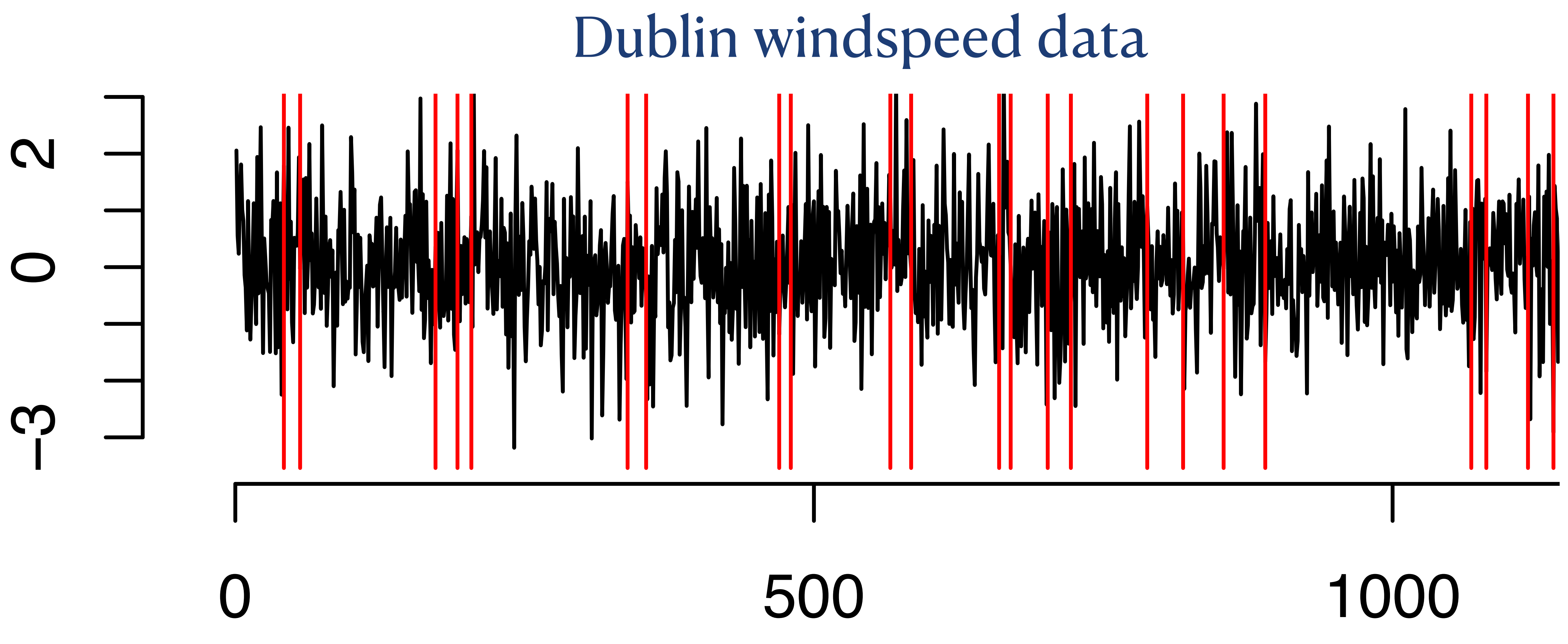
Goal: Can we validate these estimated changepoints?



Application: changepoint detection

Settings: Previous authors modeled differences in wind speeds as $X_i \sim N(0, \theta_i)$ for times $i = 1, \ldots, T$, and estimated points in the sequence where the variance changes.

Goal: Can we validate these estimated changepoints?



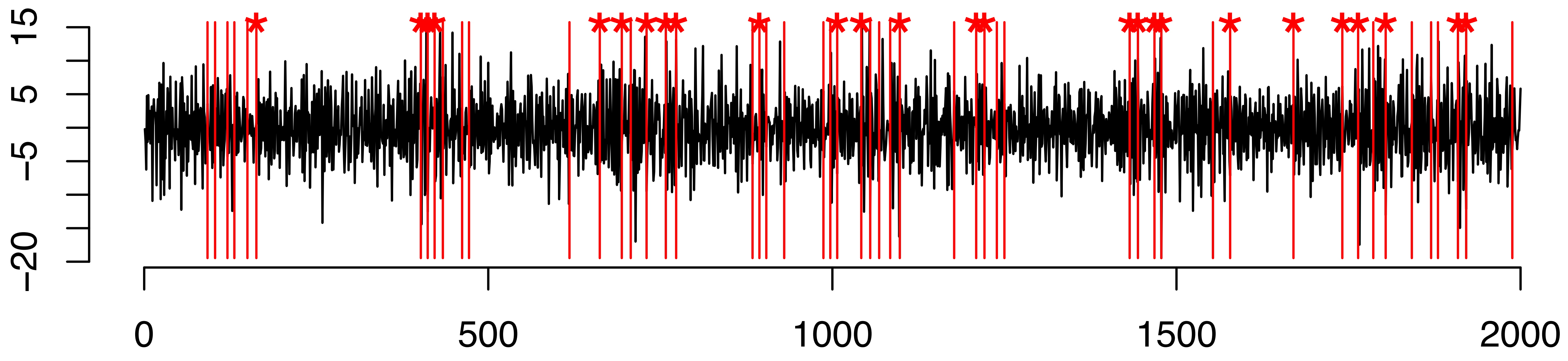
Define $Z_i := X_i^2$. Note that $Z_i \sim$ Gamma $\left(0.5, 0.5/\theta_i^2\right)$.

Naive method: Estimate and test changepoints on Z_i.

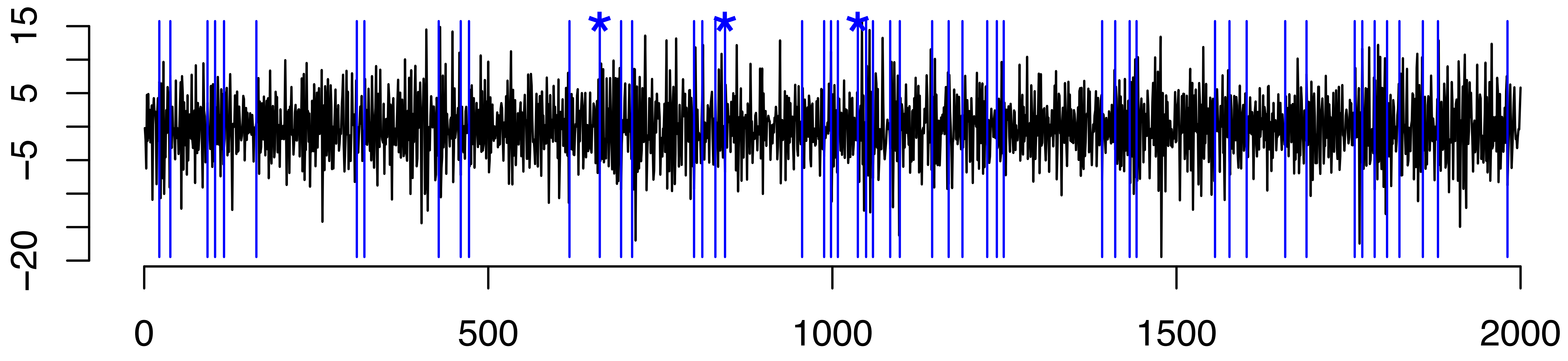
Thinning: Thin Z_i into $Z_i^{(1)} \sim$ Gamma $\left(0.25, 0.5/\theta_i^2\right)$ and $Z_i^{(2)} \sim$ Gamma $\left(0.25, 0.5/\theta_i^2\right)$.
Identify changepoints using $Z^{(1)}$, and test using $Z^{(2)}$.

Dublin wind speed data

Naive: 55 changepoints, 26 p-value < 0.05



Data thinning: 53 changepoints, 3 p-value < 0.05



Outline

1. Motivation: sample splitting doesn't always work
2. Poisson thinning
3. Data thinning
4. Application to changepoint detection
5. **Generalized data thinning**

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

Generalized thinning with non-additive decompositions

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

$\Downarrow$

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
$X' \sim Q^1_\theta, \;\; X'' \sim Q^2_\theta, \;\; X' \perp\!\!\!\perp X''.$

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

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 $X' \sim Q^1_\theta, \;\; X'' \sim Q^2_\theta, \;\; X' \perp\!\!\!\perp X''$.

↓

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

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

Let $G_{x,\theta}$ 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 $X' \sim Q_\theta^1,\ X'' \sim Q_\theta^2,\ X' \perp\!\!\!\perp X''$.

↓

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

↓

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

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

Let $G_{x,\theta}$ 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
$X' \sim Q_\theta^1, \;\; X'' \sim Q_\theta^2, \;\; X' \perp\!\!\!\perp X''$.

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

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

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 $G_{x,\theta}$ 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 $X' \sim Q^1_\theta, \;\; X'' \sim Q^2_\theta, \;\; X' \perp\!\!\!\perp X''$.

↓

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

↓

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

Theorem:
$X^{(1)} \sim Q^1_\theta, \quad X^{(2)} \sim Q^2_\theta, \quad X^{(1)} \perp\!\!\!\perp X^{(2)}$.

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

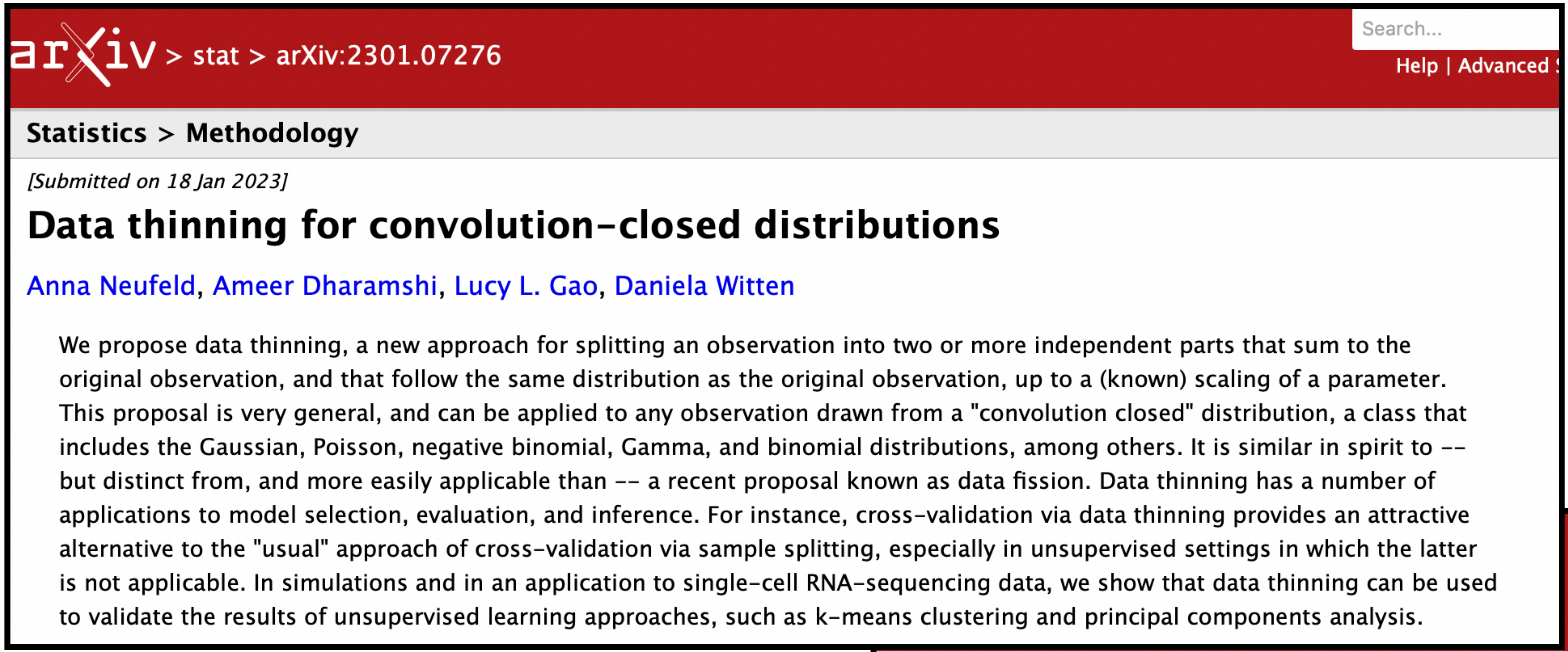
Let $G_{x,\theta}$ be the conditional distribution of $(X', X'') \mid X = x$.

Key idea: If $X = T(X', X'')$ is sufficient for θ in the joint of (X', X''), then $G_{x,\theta}$ 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)	$N(\theta, \sigma^2)$	$\mathrm{N}(\epsilon_k\theta, \epsilon_k\sigma^2)$	$\sum_{k=1}^K X^{(k)}$
	$\text{Poisson}(\theta)$	$\text{Poisson}(\epsilon_k\theta)$	
	$\text{NegBin}(r, \theta)$	$\text{NegBin}(\epsilon_k r, \theta)$	
	$\text{Binomial}(r, \theta)$	$\text{Binomial}(\epsilon_k r, \theta)$	
	$\text{Gamma}(\alpha, \theta)$	$\text{Gamma}(\epsilon_k\alpha, \theta)$	
	$\mathrm{N}_p(\boldsymbol{\theta}, \Sigma)$	$N_p(\epsilon_k\boldsymbol{\theta}, \epsilon_k\Sigma)$	
	$\text{Multinomial}_p(r, \boldsymbol{\theta})$	$\text{Multinomial}_p(\epsilon_k r, \boldsymbol{\theta})$	
	$\text{Gamma}(K/2, \theta)$	$N(0, \frac{1}{2\theta})$	$\sum_{k=1}^K \left(X^{(k)}\right)^2$
	$\text{Gamma}(K, \theta)$	$\text{Weibull}(\theta^{-\frac{1}{\nu}}, \nu)$	$\sum_{k=1}^K \left(X^{(k)}\right)^\nu$
General exponential family (in parameter θ)	$\text{Beta}(\theta, \beta)$	$\text{Beta}\left(\frac{1}{K}\theta + \frac{k-1}{K}, \frac{1}{K}\beta\right)$	$\left(\Pi_{k=1}^K X^{(k)}\right)^{1/K}$
	$\text{Beta}(\alpha, \theta)$	$\text{Beta}\left(\frac{1}{K}\alpha, \frac{1}{K}\theta + \frac{k-1}{K}\right)$	$\left(\Pi_{k=1}^K \left(1 - X^{(k)}\right)\right)^{1/K}$
	$\text{Gamma}(\theta, \beta)$	$\text{Gamma}(\frac{1}{K}\theta + \frac{k-1}{K}, \frac{1}{K}\beta)$	$\left(\Pi_{k=1}^K X^{(k)}\right)^{1/K}$
	$\text{Weibull}(\theta, \nu)$	$\text{Gamma}(\frac{1}{K}, \theta^{-\nu})$	$\left(\sum_{k=1}^K X^{(k)}\right)^{1/\nu}$
	$\text{Pareto}(\nu, \theta)$	$\text{Gamma}(\frac{1}{K}, \theta)$	$\nu \times \text{Exp}\left(\sum_{k=1}^K X^{(k)}\right)$
	$\mathrm{N}(0,\theta)$	$\text{Gamma}(\frac{1}{2K}, \frac{1}{2\theta})$	$X^2 = \sum_{k=1}^K X^{(k)}$
	$\mathrm{N}_K(\theta_1 1_K, \theta_2 I_K)$	$N(\theta_1, \theta_2)$	sample mean and variance
Truncated support family	$\text{Unif}(0, \theta)$	$\theta \cdot \text{Beta}(\frac{1}{K}, 1)$	$\max\left(X^{(1)}, \ldots, X^{(K)}\right)$
	$\theta \cdot \text{Beta}(\alpha, 1)$	$\theta \cdot \text{Beta}(\frac{\alpha}{K}, 1)$	
	$\theta + \text{Exp}(\lambda)$	$\theta + \text{Exp}(\lambda/\mathrm{K})$	$\min\left(X^{(1)}, \ldots, X^{(K)}\right)$
Non-parametric	F^n	F^{n_k}	$\text{sort}(X^{(1)}, \ldots, X^{(K)})$

References


arXiv > stat > arXiv:2301.07276

Search...

Help | Advanced

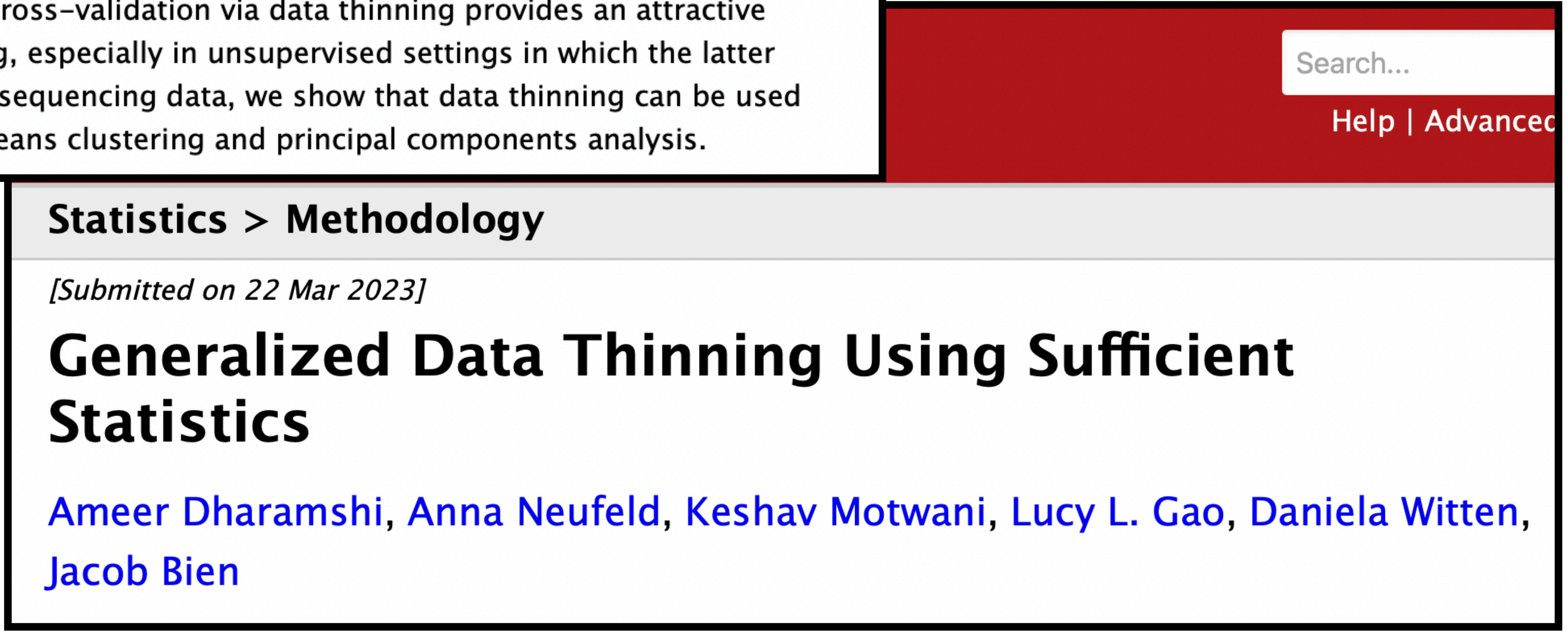
Statistics > Methodology

[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.


Search...

Help | Advanced

Statistics > Methodology

[Submitted on 22 Mar 2023]

Generalized Data Thinning Using Sufficient Statistics

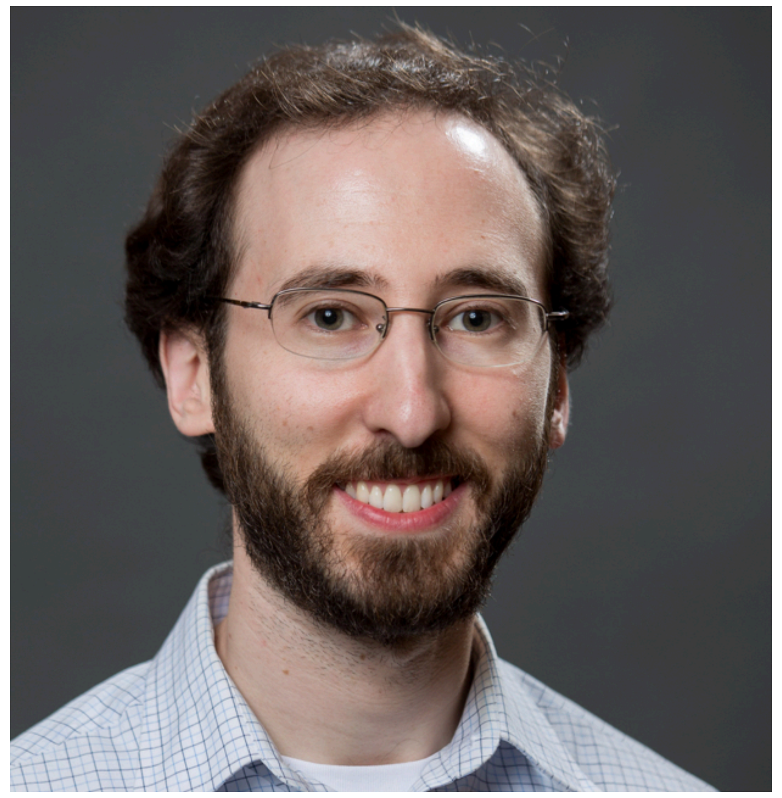
Ameer Dharamshi, Anna Neufeld, Keshav Motwani, Lucy L. Gao, Daniela Witten, Jacob Bien

Acknowledgements

Daniela Witten
University of Washington

Lucy Gao
University of British Columbia

Ameer Dharamshi
University of Washington

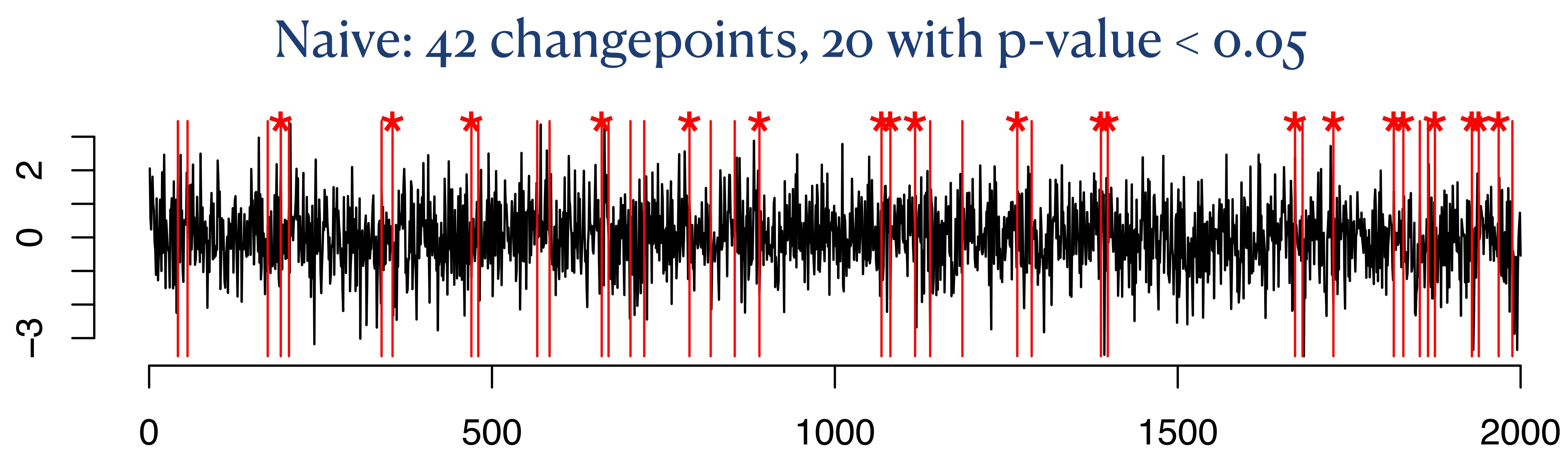


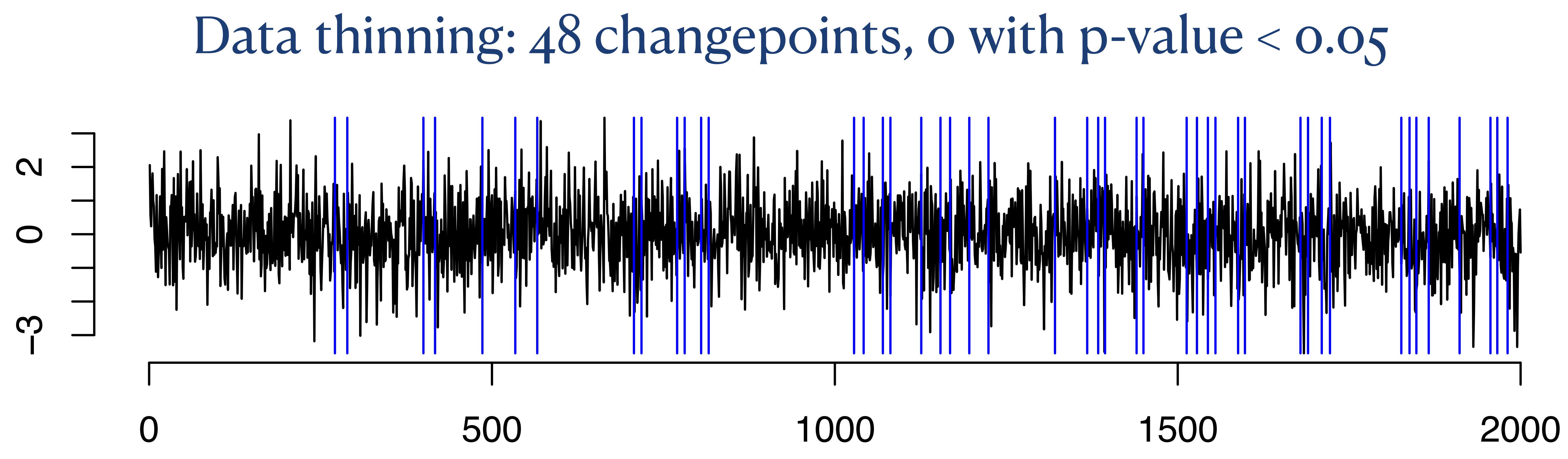
Jacob Bien
University of Southern California

Keshav Motwani
University of Washington

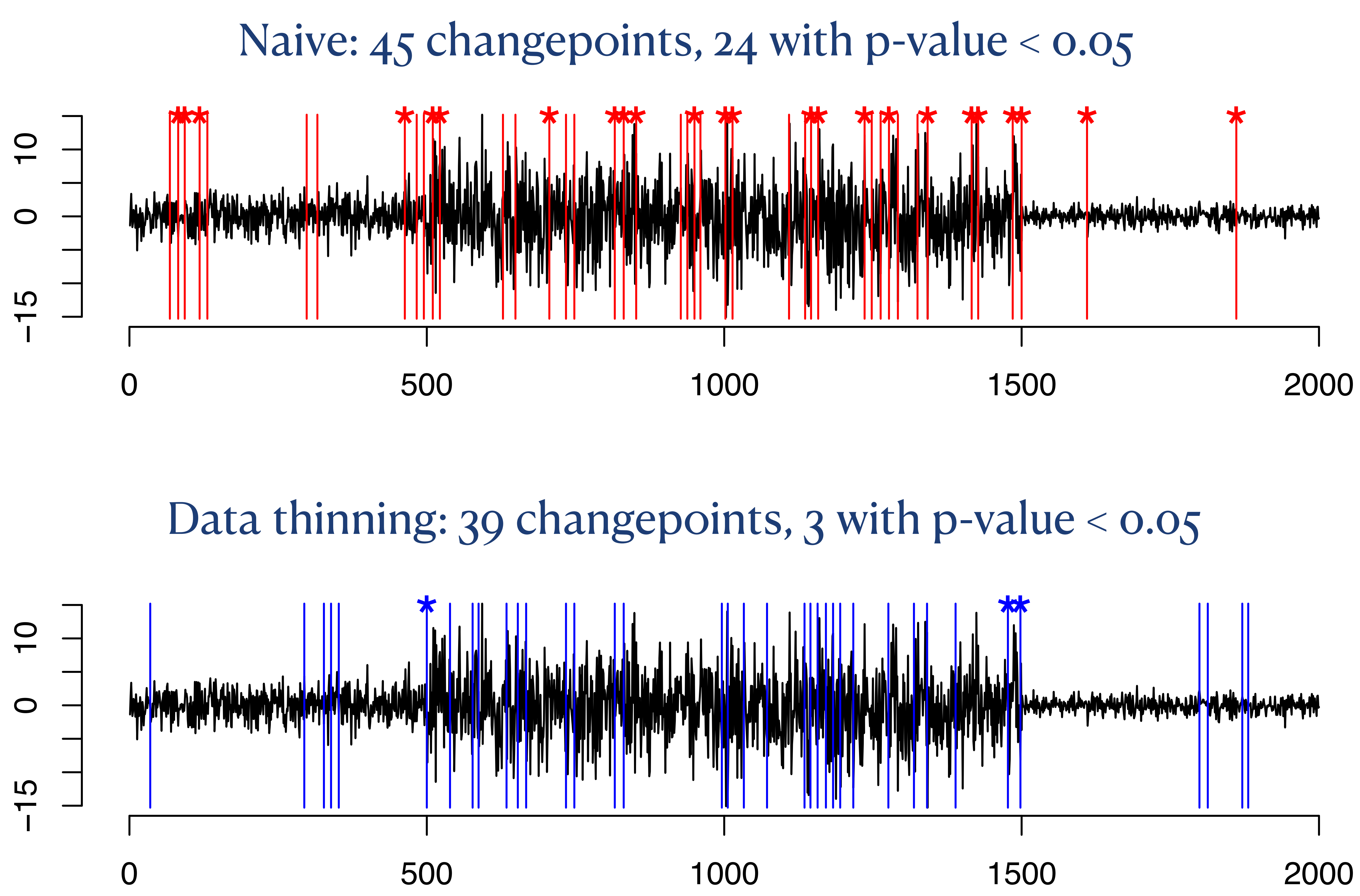
Questions?

Simulated data with no changepoints





Simulated data with two true changepoints



Are the results of our data analysis stable?