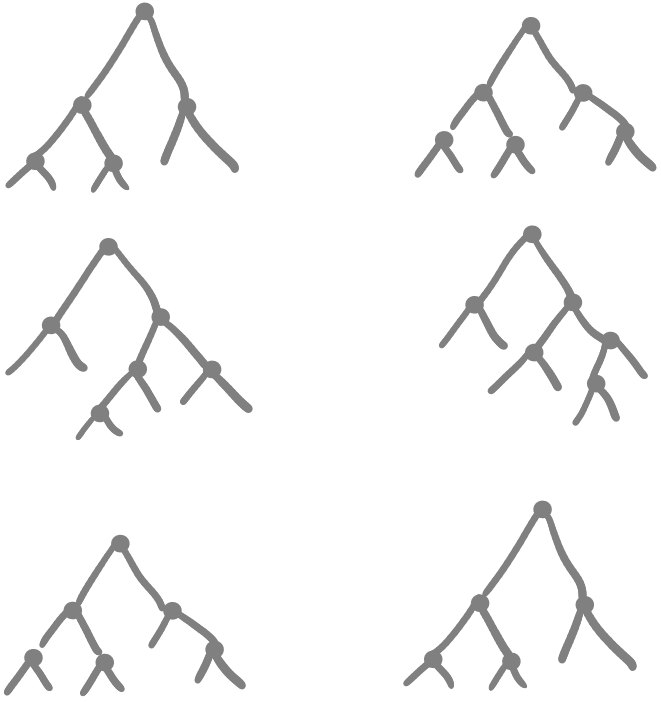
How to sample nodes from already existing forest?

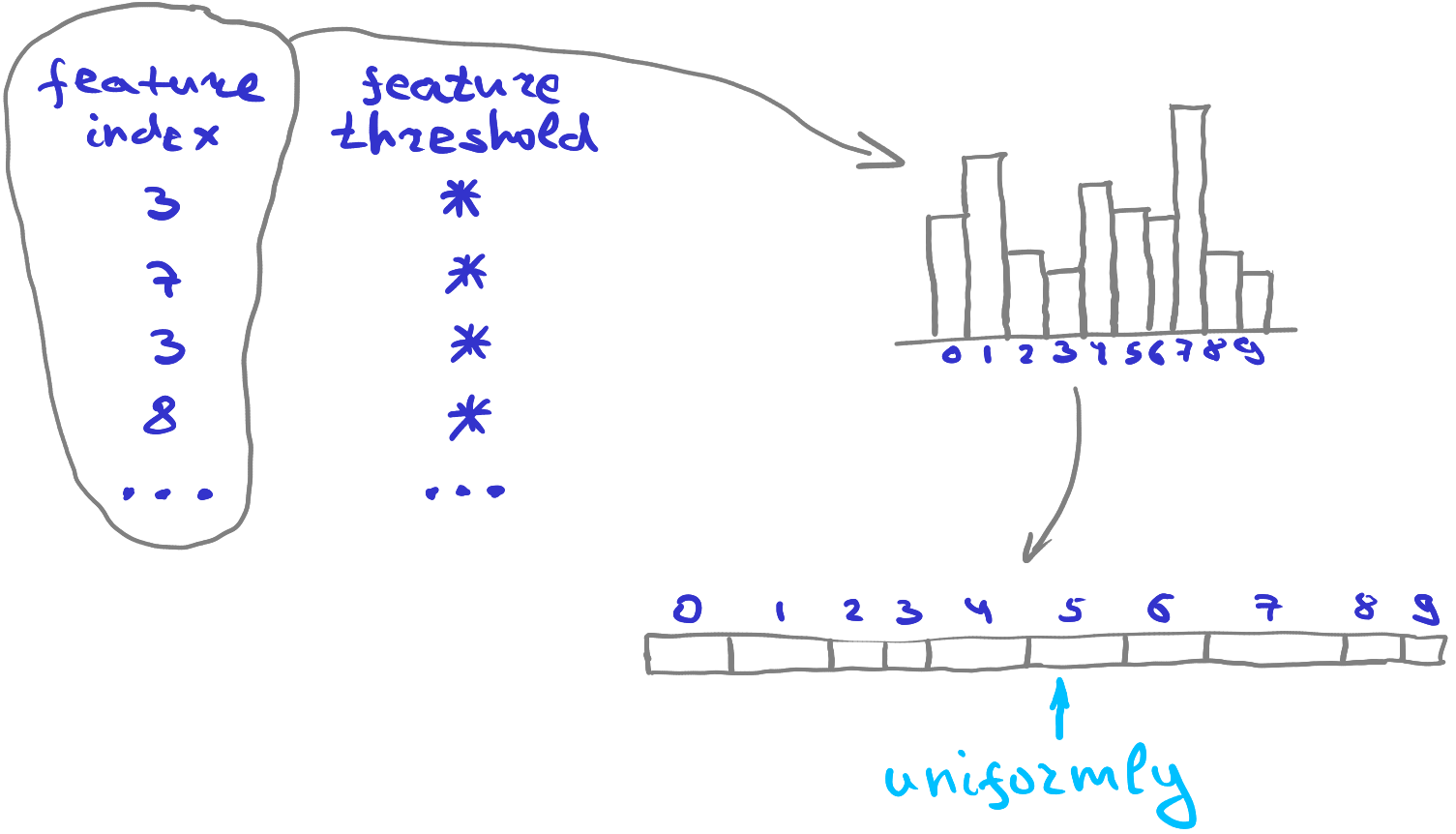


Nodes:

feature index	feature threshold
3	0.1
7	0.2
3	−1.4
8	23.7

...

Feature index statistics: sampling is easy



Feature threshold statistics: sampling is (a bit) harder

pickup a feature:
let it be 3

feature index	feature threshold
3	0.1
*	*
3	-1.4
*	*
...	...

?

Feature threshold statistics: sampling is (a bit) harder

Option 1: build a hystogram

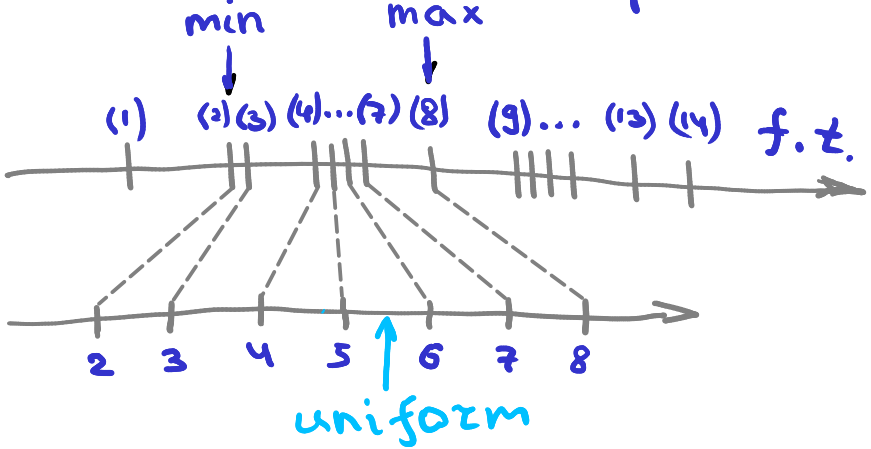
But:

1. How many bins?
2. What bin width is?
3. Should bin width be uniform?

Feature threshold statistics: sampling is (a bit) harder

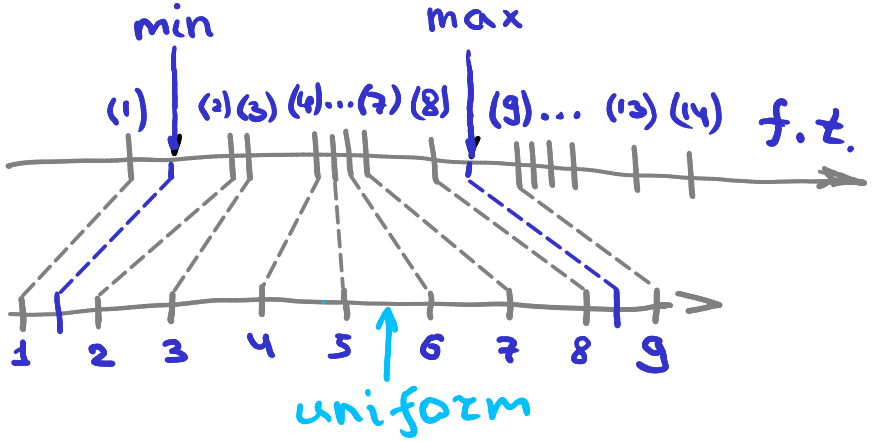
Option 2 : sample from data itself.

Simple case: min and max present in data.



Feature threshold statistics: sampling is (a bit) harder

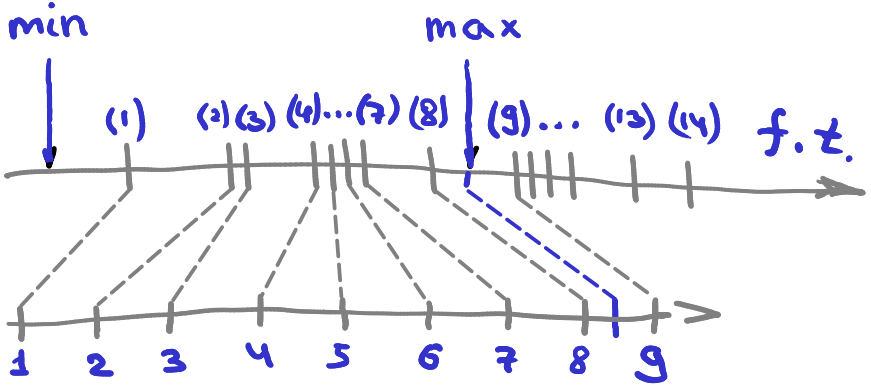
Less simple case: min and max doesn't match samples.



ends of sampling interval are not integers anymore, but that's not a problem

Feature threshold statistics: sampling is (a bit) harder

Even more less simple case:
min or max are outside of samples.



should we extrapolate?