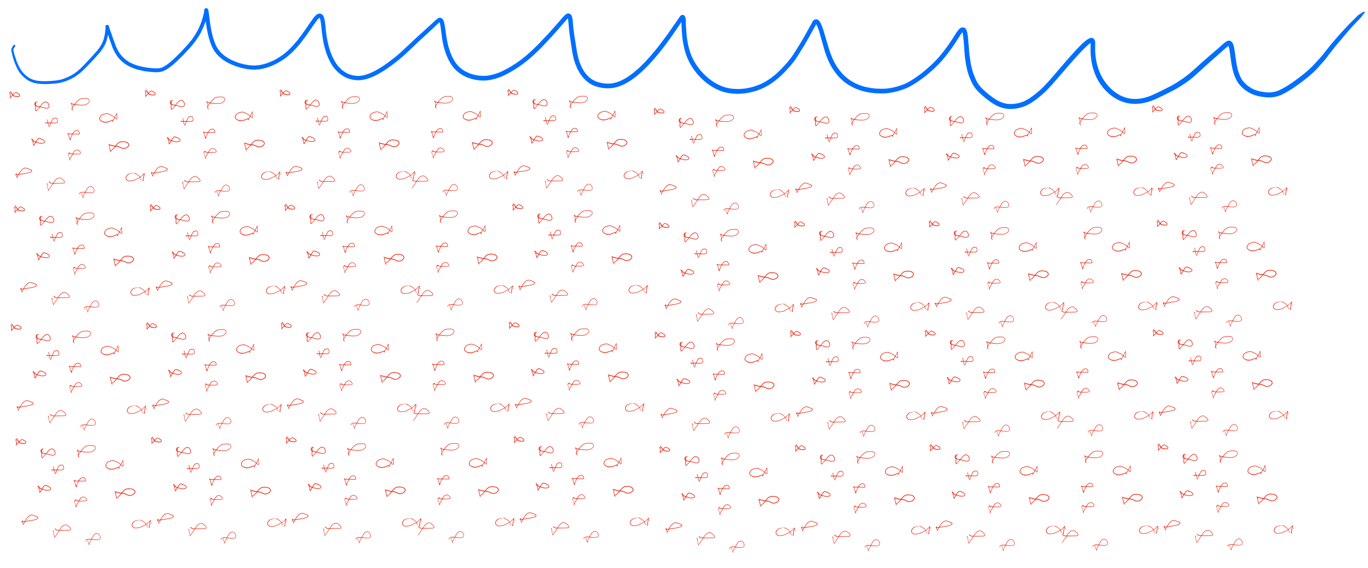
The Central Limit Theorem

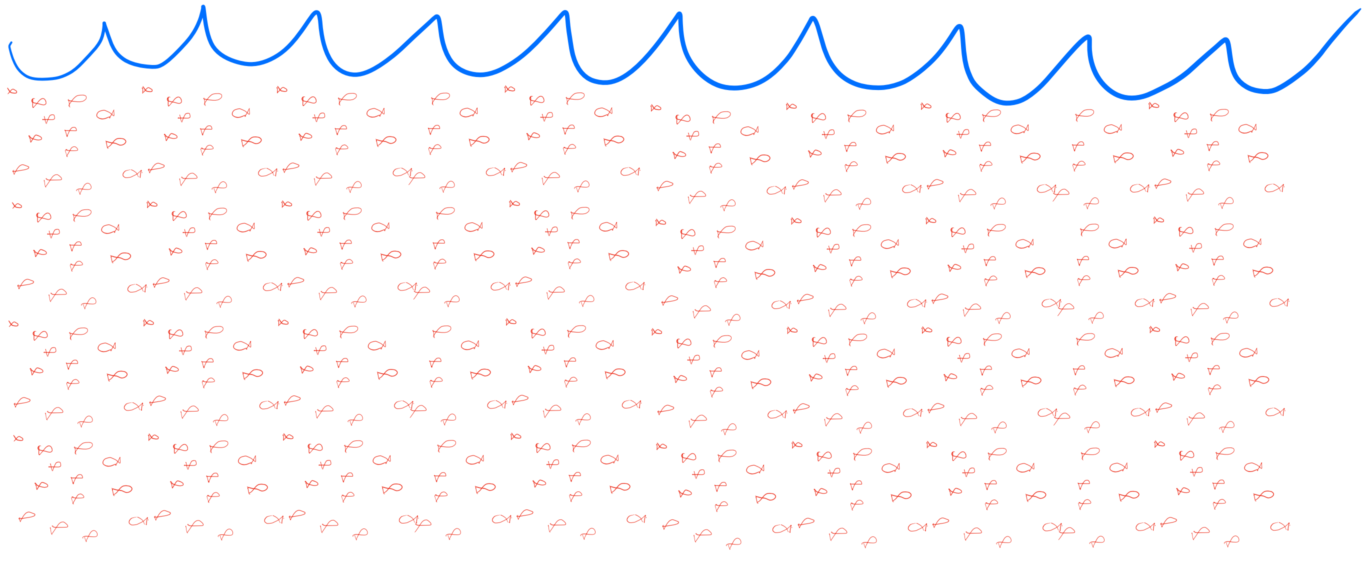
The Central Limit Theorem

(with pictures!)

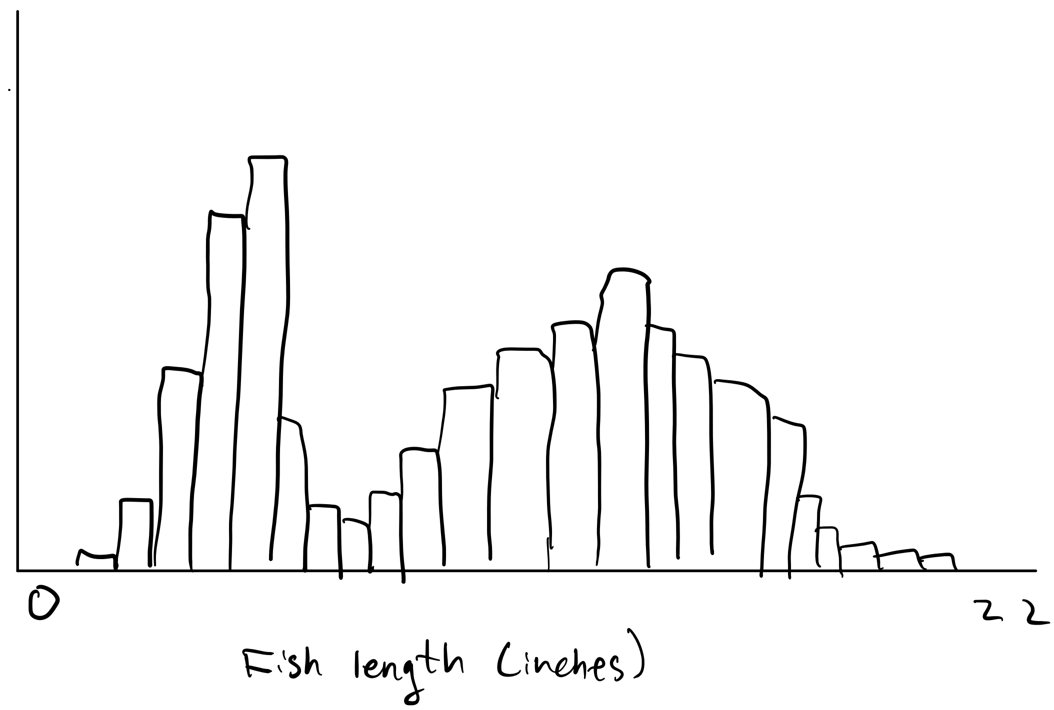
~Imagine~ we want to estimate the average length of fish in a pond...

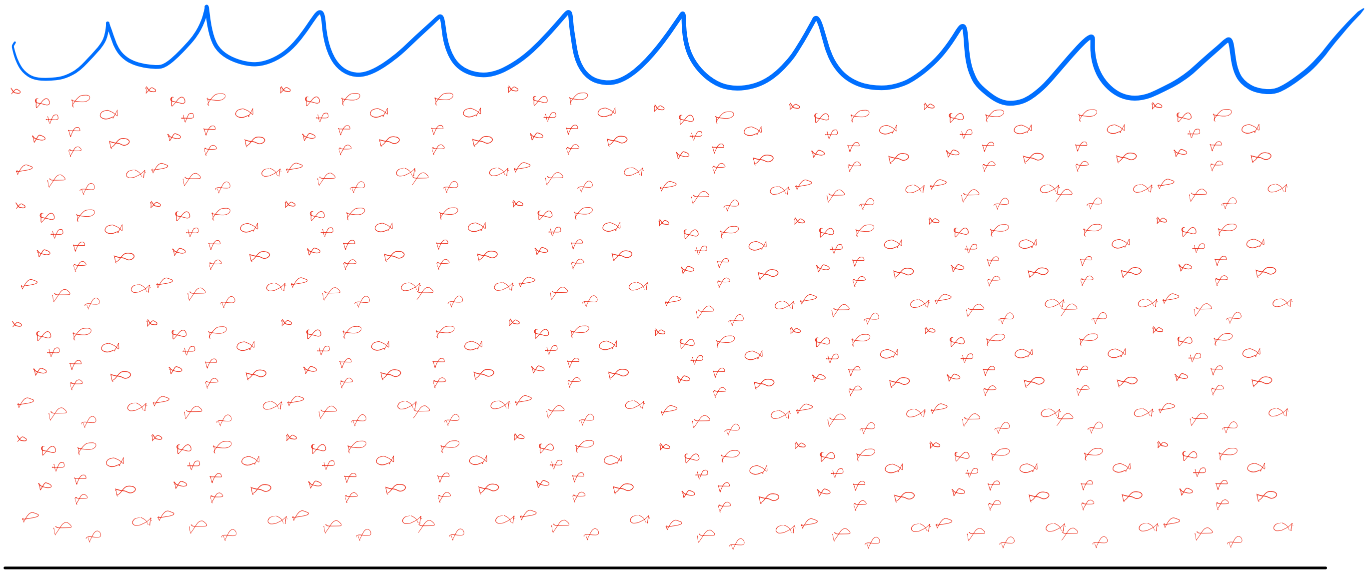
~Imagine~ we want to estimate the average length of fish in a pond...





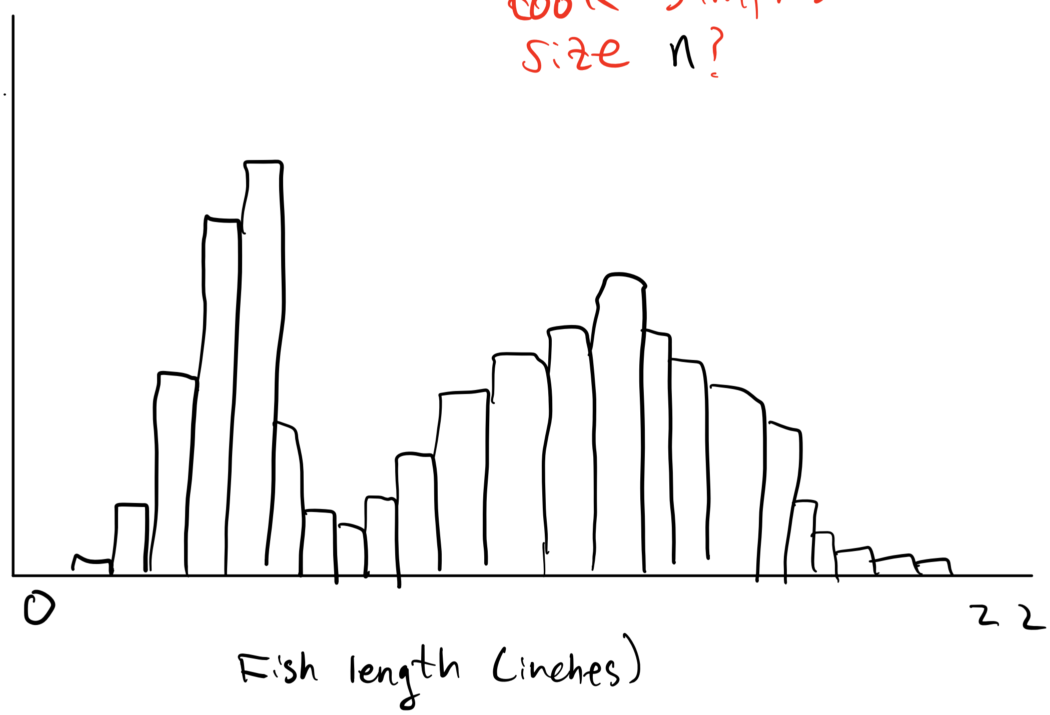
The distribution of **ALL** these lengths looks a little funky . . .

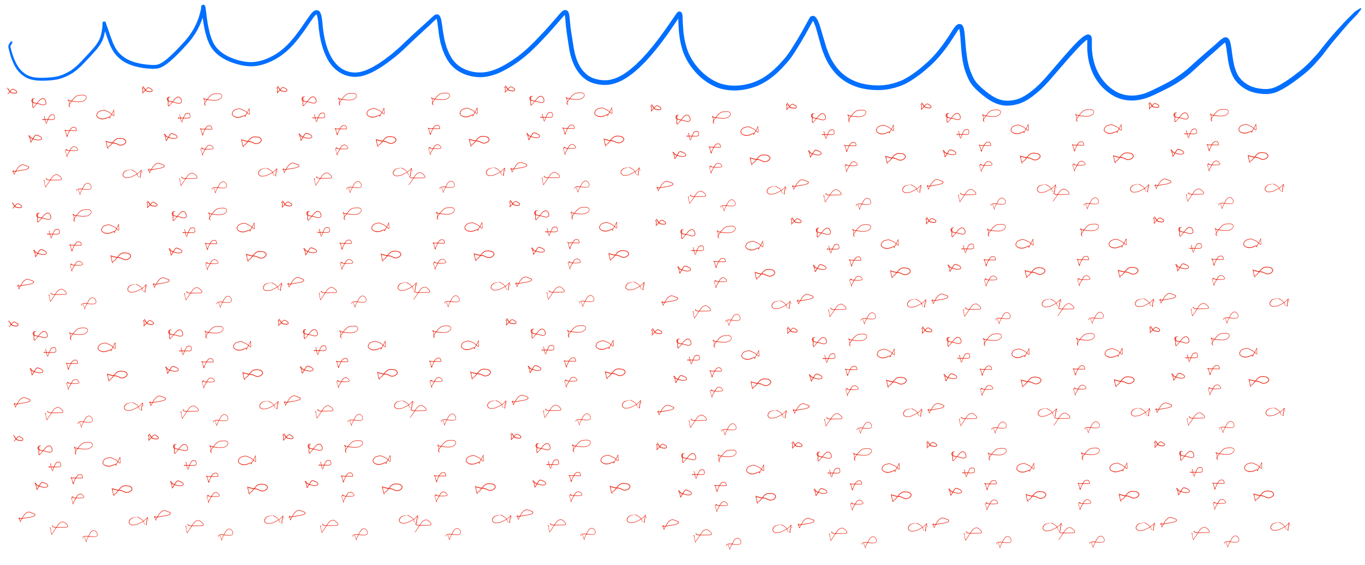




The distribution of **ALL** these lengths looks a little funky...

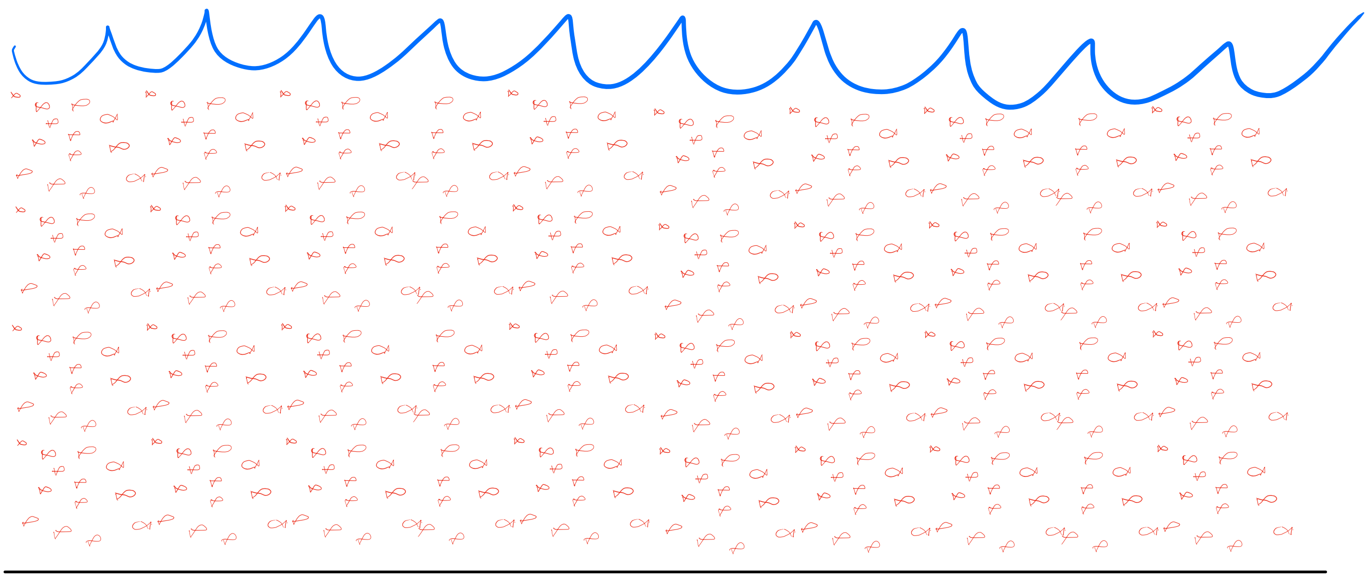
What if we repeatedly took samples of size n?





The distribution of ALL these lengths looks a little funky...

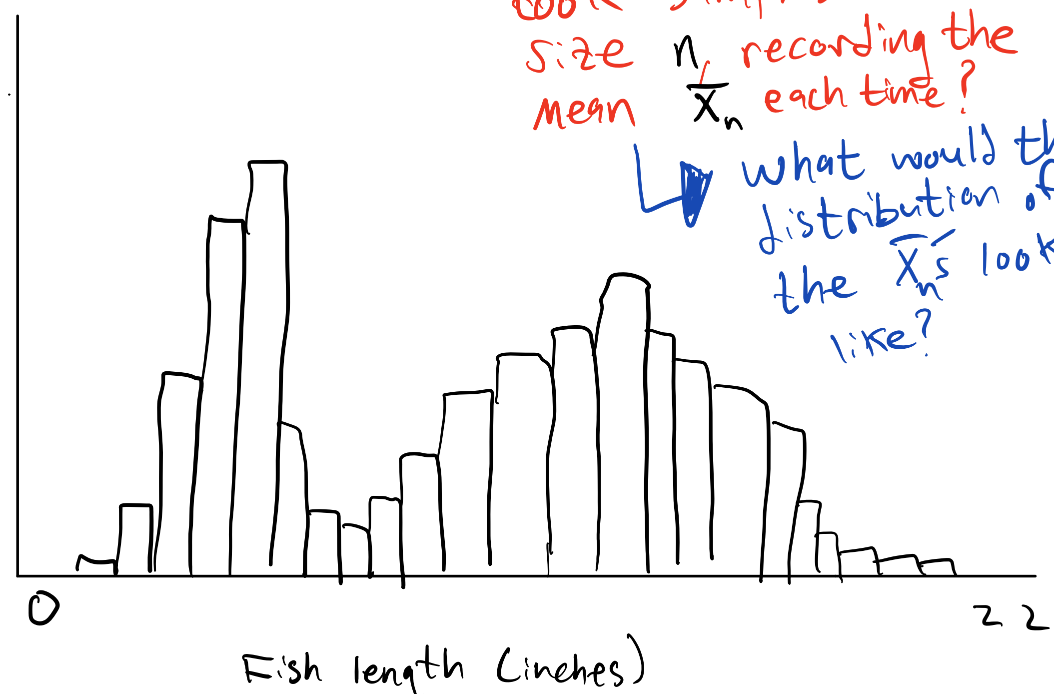
What if we repeatedly took samples of size n, recording the mean $\bar{X}_n$ each time?

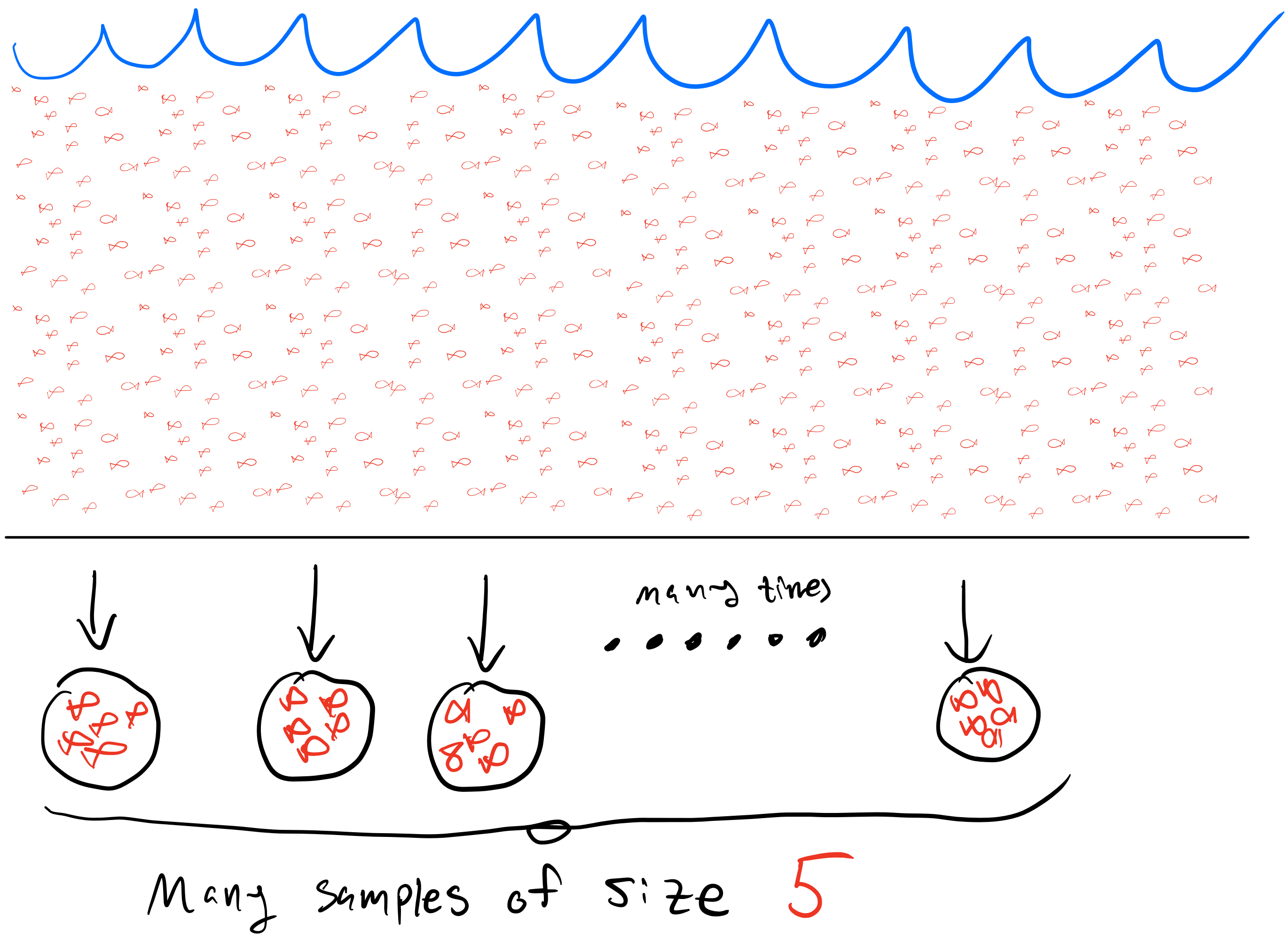


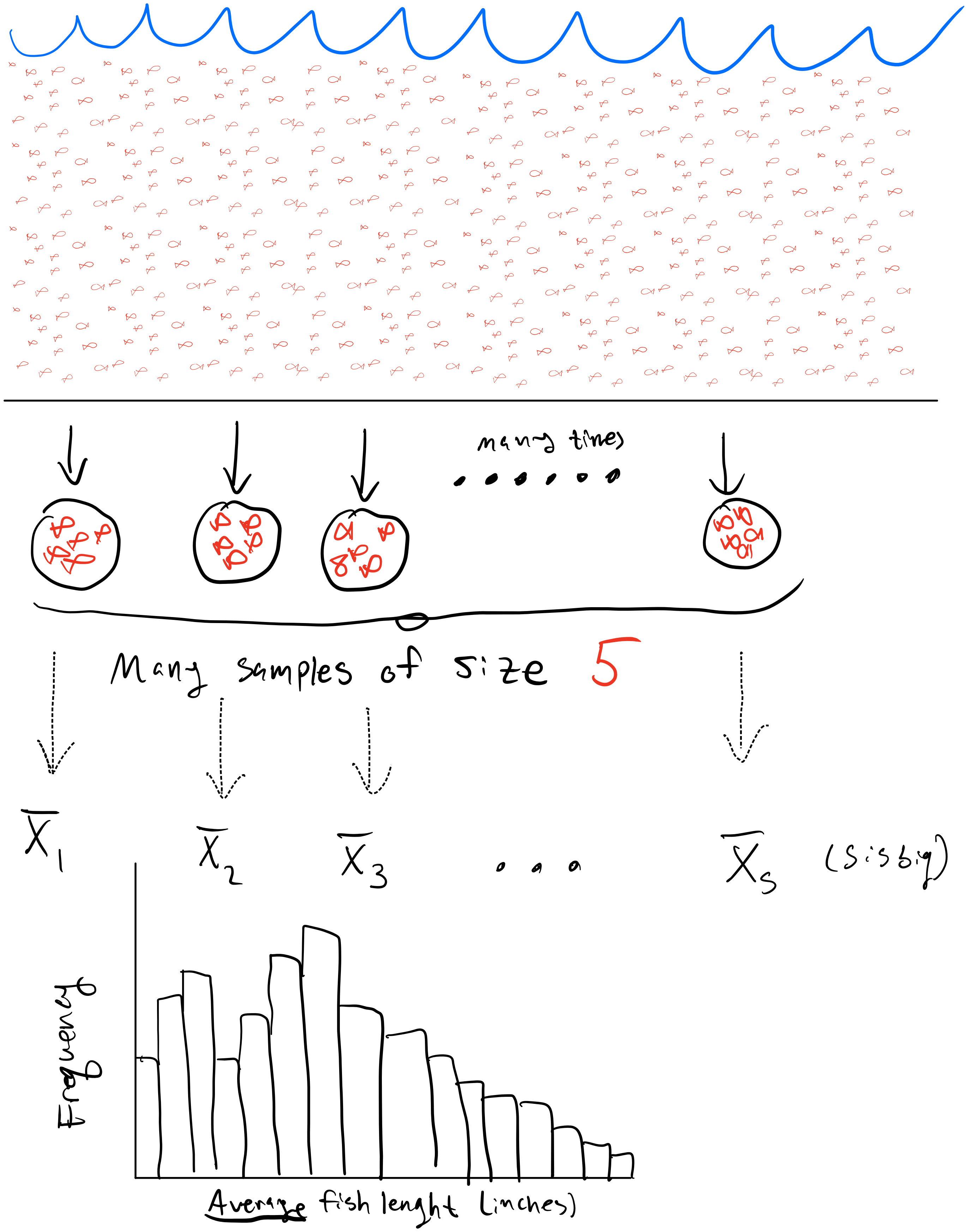
The distribution of ALL these lengths looks a little funky...

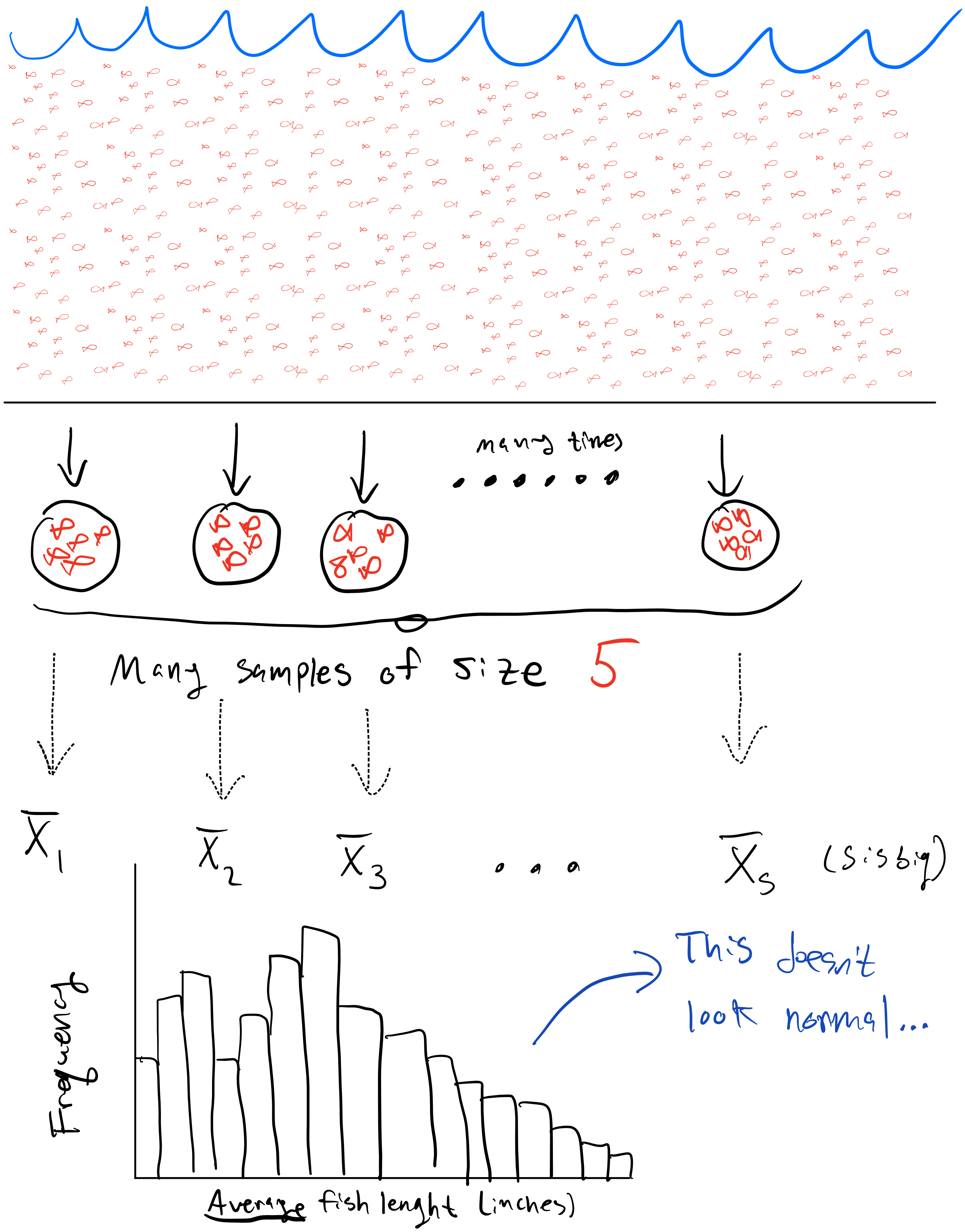
What if we repeatedly took samples of size n, recording the mean $\bar{X}_n$ each time?

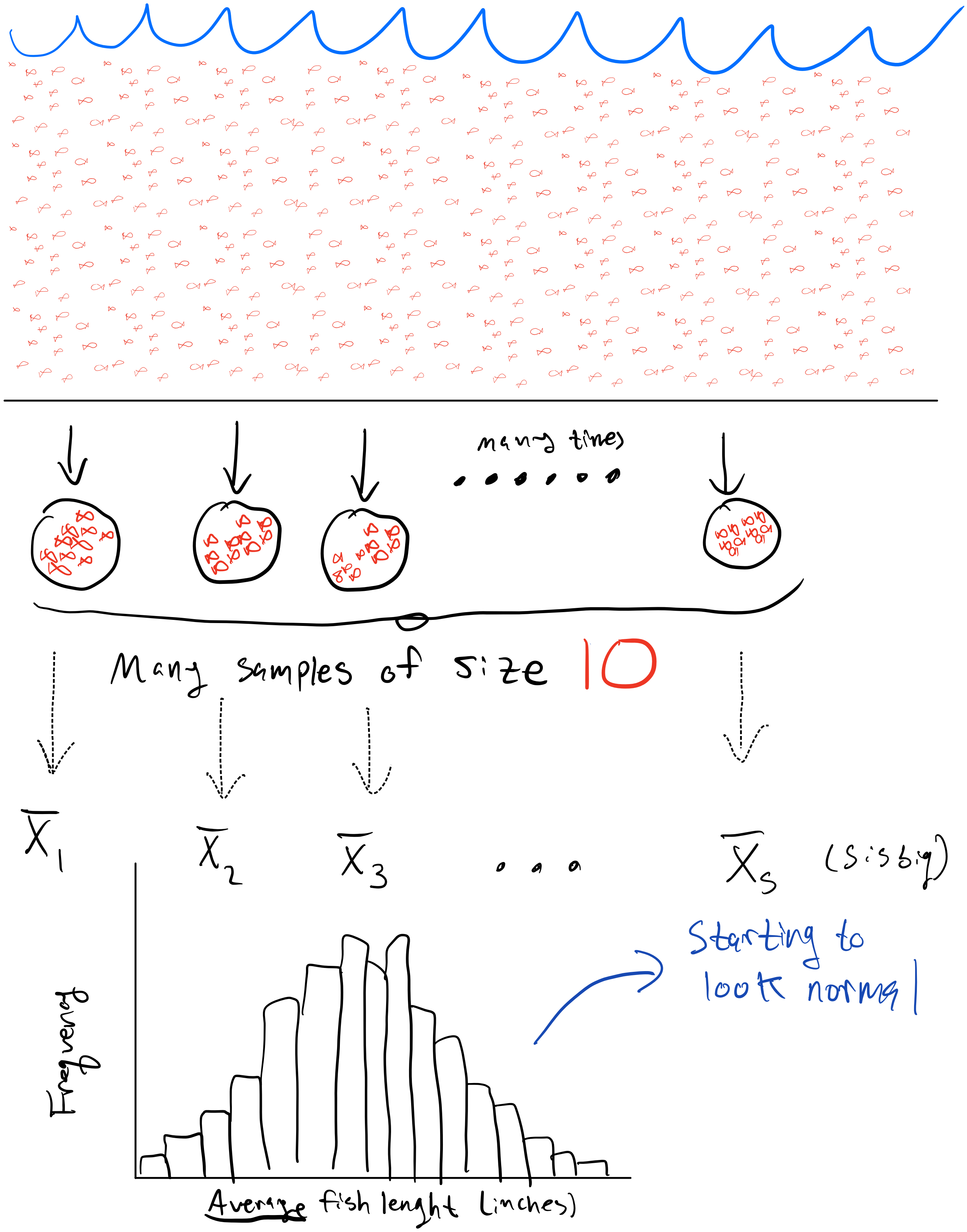
↳ what would the distribution of the $\bar{X}_n$'s look like?

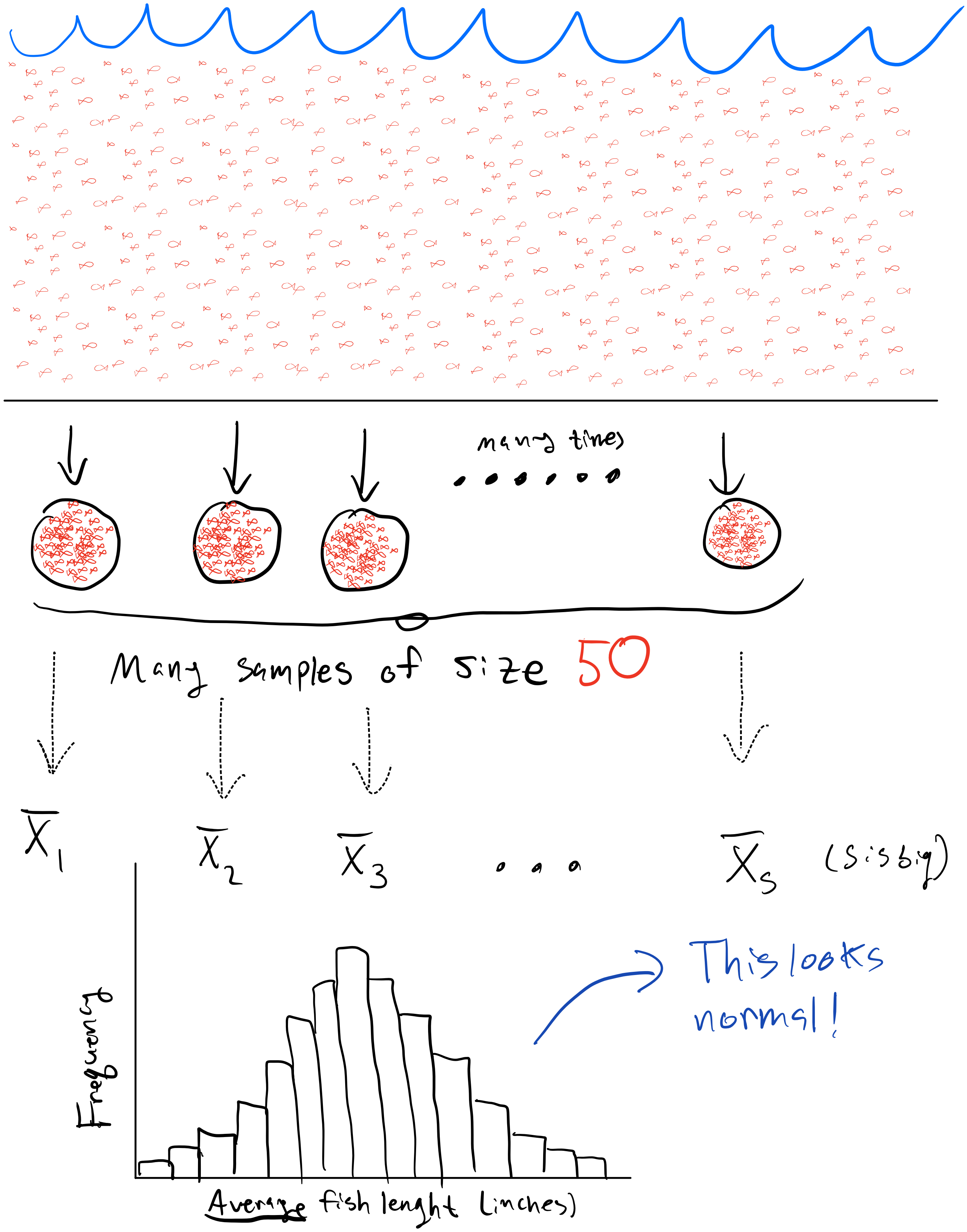



many times
Many samples of size 5


many times
Many samples of size 5
$\bar{X}_1$
$\bar{X}_2$
$\bar{X}_3$
$\bar{X}_s$
(s is big)
Frequency
Average fish lenght (inches)


many times
Many samples of size 5
X̄1
X̄2
X̄3
. . .
X̄s (s is big)
Frequency
Average fish lenght (inches)
This doesn't look normal...


many times
Many samples of size 10
$\bar{X}_1$
$\bar{X}_2$
$\bar{X}_3$
$\bar{X}_s$
(s is big)
Starting to look normal
Frequency
Average fish lenght (inches)


many times
Many samples of size 50
$\bar{X}_1$
$\bar{X}_2$
$\bar{X}_3$
$\bar{X}_s$
(s is big)
Frequency
Average fish lenght (inches)
This looks normal!

What we saw

- We had a funky distribution of a random variable (fish length)

- When we took large samples, over and over, and recorded the mean each time, we saw that the distribution of those means began to look normal as n got large

- This is the essence of the CLT!

No matter how wacky the initial distribution of X, the distribution of $\frac{1}{n}\sum_{i=1}^{n} X_i$ approaches the normal distribution as n gets big

- Typically, $n=30$ is "large enough"
  - But we should always be careful before assuming the CLT has "kicked in"