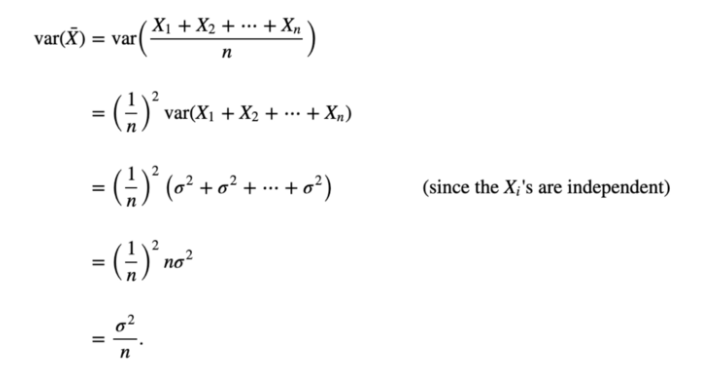
* a lot of machine learning mathematical problems involve extrapolating a subset of data to infer a global population
* For example
  + We may only get 100 replies on a survey for a new website
  + But the target audience is for 10m customers
  + It is too hard to ask for 10m customers therefore we have to infer from 100
* Probability distribution explain the likelihood of different things happening
  + They can tell if event A is probably not going to happen but Event B is much more likely to happen
* To calculate the expectation of any statistic you
  + Always take the expectation

* For example if you have 100 customers from a population of 10m independent customers
  + The mean of this msall sample is a very good estimate of the population mean
* The second moment of the distribution is the sample mean



* The variance function can feed through the sample mean statistic to extract 1/n
* The end results tells us that the variance of each sample mean is equivalent to the variance of the underlying data divided by the number of data points in each sample
* This is another huge result as it tells us by how much the variance of our mean estimate decreases as N increases
* For example
  + Lets says we have normally distributed data (variance = 1)
  + When N = 100 samples we can now say within +/- 0.1
* The central limit theorem proves that when random variables are added the distribution converges toward a normal distribution even if the original variables themselves are not normally distributed
  + This normally happens with N>25
* Now that we have proven that the sample mean statistic has a mean of and a variance of sigma^2/n
* Bootstrap methods use Monte Carlo simulations to approximate a distribution i
* For example
  + If you have 10,000 samples from a random number generator
  + If you calculate from the sample of 100 numbers