confidence sets

- Let C be a set that is constructed from $X_1, ...X_n$. C is a $1-\alpha$ confidence set for θ if $P(\theta \epsilon C) \geq 1-\alpha$. Note that θ is fixed, unknown. It is not a random variable. C is random. This distinction is subtle and will become clearer as we go.
- when the set is an interval, its called a confidence interval.
- In the case of an interval $[L, U]$, it is the same as saying $P(L \leq \theta \leq U)$. if $L = \overline{X} - c$ and $U = \overline{X} + c$, then this is the same as $P(\overline{X} - c \leq \theta \leq \overline{X} + c)$ which is $P(\overline{-}c \leq \theta - \overline{X} \leq c)$ which is the same as $P(\frac{-\sqrt{n}c}{\sigma} \leq \frac{\sqrt{n}(\overline{X}-\theta)}{\sigma} \leq \frac{\sqrt{n}c}{\sigma})$. This is just a standard normal and we know how to do a math. Given a target α, we can compute c in order to achieve that target.
- methods to construct confidence intervals:
  - probability inequality
  - inverting a test
  - pivots
  - large sample approximations

  these are more themes rather than exclusive categories. we will see overlap between them.
- an example of the probability inequality test is as follows: consider hoeffding's inequality $P(|\hat{p} - p|) \leq 2e^{-2n\epsilon^2}$. This directly gives us a confidence interval $P(\hat{p} - \epsilon \leq p \leq \hat{p} + \epsilon)$ and we can achieve desired α by setting ϵ according to heoffding's. this method turns out not work in all situations, so we'll see other methods.
- inverting a test:
  - do a size α hypothesis test for all candidate θ's and only keep the θ's that you don't reject from the test. this set of θ's is a $1-\alpha$ confidence set. why? because lets say the true value is one of the rejected ones. the probability of that being rejected, by construction $\leq \alpha$. QED.
- Larry doesn't do pivots method of confidence intervals
- confidence sets by large sample approximation:
  - we've seen that for the MLE, $\frac{\hat{\theta}_n - \theta}{\hat{se}}$ ~ $N(0,1)$. this means the confidence interval c_n is $\hat{\theta} \pm z_{\frac{\alpha}{2}} \hat{se}$.
- The most useful confidence sets tend to be those that give the smallest intervals → its called an optimal confidence set. we won't cover that here, but remember that small confidence sets are better cuz they give tighter bounds.
- In practice, a useful tactic is to use different data to get different confidence sets and then take the intersection of those to get a smaller confidence set.
- think about this → every test (hypothesis test) is basically a confidence set. the reverse is also true! if we already have an interval/set, we can setup a test that rejects the null with size α. this is called the duality between confidence sets and hypothesis testing.
- A confidence interval is more informative than the result of a hypothesis test. Instead of just saying reject/don't reject, the confidence interval specifies the size of the interval and a smaller size with low α is informative, instead of just a test result.
- THE REMAINDER OF THE CLASS HAS VERY FEW/SPARSE VIDEOS, SO IT ISN'T REALLY POSSIBLE TO WATCH THE LECTURES AND GLEAN THE MATERIAL. HOWEVER, THE MATERIAL IN THIS LAST THIRD OF THE COURSE THAT'S RELEVANT TO MACHINE LEARNING IS COVERED BETWEEN MY NOTES IN THE MISCELLANEOUS/OTHER FOLDER AND MY NOTES ON THE STANFORD CS 229 MACHINE LEARNING COURSE.