Bootstrap

So far, our confidence intervals have relied on asymptotic normality

- We rely on being able to calculate asymptotic variance based on:
  - CLT if $\hat{\theta} = \overline{X_n}$
  - Delta method if $\hat{\theta} = g(\overline{X_n})$
  - Fisher information if $\hat{\theta} = \hat{\theta}^{\text{MLE}}$
- However, what if it is not in one of these three forms?
- What if we tried to get an estimator for the mean of a Poisson variable?
  - We could estimate λ with $\overline{X_n}$
  - θ is the integer such that:
    - $\sum_{k=0}^{\theta-1} p(k) < 0.5$
    - $\sum_{k=0}^{\theta} p(k) \geq 0.5$
  - Therefore we can write our estimate of median as $\hat{\theta} = g(\overline{X_n})$
  - But this g is not differentiable!
- The key idea is that to estimate the variance of our estimator $\hat{\theta}$, we need multiple estimates
  - But we used up all n of our samples computing a single estimator
  - Imagine we had $B \cdot n$ samples: then we could compute B estimators and then take their variance

Bootstrap Sample

Given n samples X_i, a **bootstrap sample** is a collection of m samples X_i^* such that:

$$X_i^* \sim \hat{\mathbb{P}}_n = \text{Unif}(X_1, \ldots, X_n)$$

This distribution is known as the **empirical distribution** of the data

- The corresponding CDF is:

$$\hat{F}_n(x) = \frac{1}{n} \sum 1(X_i \leq x)$$

- With the LLN, one can show this CDF converges in probability to $F(x)$ as n goes to ∞

Then, we can compute B bootstrap samples each of size n and can then use the sample variance v_{boot}

- Note there are two approximations:

$$v_{\text{boot}} \approx \mathbb{V}_{\hat{\mathbb{P}}_n}[\hat{\theta}] \approx \mathbb{V}_{\mathbb{P}}[\hat{\theta}]$$

- The first one can be made accurate by just increasing B
- The second one relies on the number of samples n which we can't change

Properties of Bootstrap Sample

-