So far:

3 approaches to parameter estimation:

1) Method of Moments
- set first few sample moments equal to corresponding moments of distribution being estimated; solve for parameters

2) Maximum Likelihood
- Likelihood function gives probability of observed data as a function of unknown ~~p~~ model parameters
- Maximize to find parameter values for which probability of observed data is largest
- For simple examples, can maximize analytically
- More often, need numerical optimization methods like Newton's method
  ↳ approximate log-likelihood with 2nd order Taylor series, maximize approximation, repeat.

3) Bayesian
- Formulate prior distribution expressing state of knowledge about parameter(s) before observing data.
- Bayes' rule tells us how state of knowledge is updated after observing data
- For simple examples with conjugate priors, can do Bayes' rule calculations exactly to get to posterior
- More often, need numerical integration methods like MCMC for Monte Carlo integration.
  ↳ draw a sample from the posterior, use sample means to approximate quantities from posterior
- If we need a single best guess of parameter value, often summarize posterior distribution with the posterior mean.
- For interval estimates, use posterior credible intervals:

$P(\theta \in [a,b] \mid X_1, \ldots, X_n) = 0.95$ ← or whatever level you set.

Estimators are random variables:

- Each sample you take has different observed data
  - ⇒ From each sample you get a different estimate.
- Distribution of an estimator is its sampling distribution.
- Compare estimators by their bias, variance, and MSE
  - want all 3 small
  - relative importance of bias & MSE a matter of opinion
- Often Bayesian estimators can be framed as

shrinkage estimators:
  - shrink an estimator with low bias towards prior mean
    - ↳ introduces bias (not much if prior is good)
    - ↳ reduces variance
    - ↳ overall, improved MSE if prior is "good", but often worse if prior is "bad".

Plan for next few days:

What are characteristics of these estimators that justify their use, or are useful for other purposes (building CI's, conducting hypothesis tests)

Justify use:

- Consistency: as $n \to \infty$, $\hat{\Theta}$ is close to Θ with high probability
  - MLE & Bayesian posterior mean are both consistent in many settings
- Efficiency: Compare Variance/MSE (want small)
- Asymptotic Efficiency: as $n \to \infty$, variance is as small as possible (among unbiased estimators)
  - MLE has this property in many settings

Useful for CI's & hypothesis tests:

- For n large, the sampling distribution of $\hat{\Theta}^{MLE}$ is approximately normal. in many settings
- For n large, the posterior distribution of Θ in a Bayesian analysis is approximately normal.