STAT 135 Lecture 10/Midterm Review

Henry Liev

15 July 2025

Remark 0.1 (Estimation Big Picture)

- Have data $X_1, X_2, \dots, X_n \overset{iid}{\sim} f(x|\theta)$, data ($X$'s) random, θ fixed (not random)
- Construct $T(X_1, X_2, \dots, X_n)$ to guess θ
- T is a random variable
- $\sqrt{Var(T)}$ is called standard error of T (Measure of uncertainty)
- T has a sampling distribution, which we can use to
  1. Quantify how good T is
  2. Expressing uncertainty
     - $SE(T)$
     - Confidence intervals involving T

Method 0.2 (Constructing Estimators)

- **Method of Moments**
  - Calculate population moments $\mu_k = E(X^k)$
  - Write θ in terms of the moments
  - Replace μ_k with $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i^k$
- **Maximum Likelihood Estimation**
  - Write likelihood function $L(\theta) = f(x_1, x_2, \dots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$
  - $\hat{\theta}_{MLE}$ is θ maximizing $L(\theta)$, $\hat{\theta}_{MLE} = \arg\max_{\theta} L(\theta)$
  - $l(\theta) = \log(L(\theta))$
  - $\arg\max_{\theta} l(\theta)$, set $l'(\theta) = 0$ solve for θ.

Remark 0.3 (Nice Properties)
$\hat{\theta}$ an estimator of θ

- **Unbiasedness**: $E(\hat{\theta}) = \theta, \forall\theta$
  - $\bar{X}$ is always unbiased
  - X_1 is always unbiased
  - MoM not always unbiased
    $\mathcal{N}(\mu, \sigma^2) \to \hat{\sigma}_{MoM} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$
  - MLE not always unbiased
- **Consistency**: $\hat{\theta}_n \xrightarrow{p} \theta$ as $n \to \infty$
  - A property of sequences of estimators
  - Low bar
  - $X_1, \ldots, X_n$ iid with mean μ, $\bar{X} \xrightarrow{p} \mu$ (Weak law of large numbers)
  - $f(\bar{X}) \xrightarrow{p} f(\mu)$ f continuous (Continuous mapping theorem)
  - MoM are consistent
  - MLE is consistent

Remark 0.4 (Confidence Intervals via Math)
"Reasonable range of values" for θ have seen some cases where $\hat{\theta} \pm 1.96SE(\hat{\theta})$ covers θ with probability 95%.

Example 0.5

Situation	How we get SE
$\hat{\theta} = \bar{X}$ or p	Use CLT/Calculate directly
$\hat{\theta} = f(\bar{X})$ or $f(\hat{p})$	Delta Method
$\hat{\theta}$ is MLE	$SE(\hat{\theta}_{MLE}) \approx \frac{1}{\sqrt{nI(\theta)}}$

Method 0.6 (Confidence Intervals via Simulation)
$X_1, X_2, \ldots, X_n \overset{iid}{\sim} f(x|\theta)$
Have some estimator $\hat{\theta}$ of θ
Get approximate CI with parametric bootstrap
Parametric Bootstrap

1. By simulation, generate large B iid samples of size n from $f(x|\hat{\theta})$, For each i, calculate θ_i^*

Remark 0.7 (Measuring Goodness of Estimators)

- Focused on MSE (Mean Squared Error) $E\left[(\hat{\theta}-\theta)^2\right]$
- $MSE(\hat{\theta}) = Var(\hat{\theta}) + Bias^2(\hat{\theta})$, Noise + Systematic Error
- Bias-Variance Tradeoff
- Optimality
  - Can't find a "best overall" estimator
  - We can find (sometimes) a best unbiased estimator
  - Cramer-Rao Lower Bound: Variance of unbiased estimator $\geq \frac{1}{nI(\theta)}$
    * If we have an unbiased estimator with variance $\frac{1}{nI(\theta)}$, it is the best possible amongst (unbiased estimators)
  - MLE is "asymptotically optimal", $\hat{\theta}_{MLE} \approx \mathcal{N}(\theta, \frac{1}{nI(\theta)})$

Remark 0.8 (**Sufficiency**)

- Main idea: A sufficient statistic T for θ has all the info in the sample about θ
- Formal definition: Conditional distribution of X's on T ($f(x_1, x_2, \ldots, x_n|\theta)$) does not depend on θ
- Factorization Criterion
  - T is sufficient $\Leftrightarrow f(x_1, \ldots, x_n|\theta) = g(T(x_1, \ldots, x_n), \theta)h(x_1, \ldots, x_n)$, where g involves T and θ, whereas h does not involve θ
  - Can help find a sufficient statistic T and can help check that T is sufficient
- Why do we care? (Two ways to say sufficient statistics are important or necessary building blocks)
  - MLE must be based on a sufficient statistic
  - For any estimator $\hat{\theta}$ of θ, if we have a sufficient statistic, T for θ, we can make a new estimator $\tilde{\theta} = E(\hat{\theta}|T)$, which has a lower MSE (Rao-Blackwell Theorem)

Remark 0.9 (MLE)

- Can usually carry out (at least numerically)
- Consistent
- Asymptotically efficient ($\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} \mathcal{N}(0, \frac{1}{I(\theta)})$)
- Equivariance: $\hat{\theta}$ is the MLE for θ then $f(\hat{\theta})$ is the MLE for $f(\theta)$