STAT 135 Lecture 6

Henry Liev

3 July 2025

Remark 0.1

Parametric Bootstrap Recap:

1. Get a CI for estimators with no theory, just simulation.
2. Don't even need closed form for estimate

Example 0.2

True parameter θ_0
Estimator: $\hat{\theta}$
$\hat{\theta} - \theta_0$ distribution can give us a CI for θ_0
95% CI $(\hat{\theta} - b, \hat{\theta} - a)$ is a 95% CI for θ_0, come up with estimates for $\hat{b}$ and $\hat{a}$
$X_1, \ldots, X_n \overset{i.i.d}{\sim} f(x|\theta_0)$ (Unknown distribution)
Replace θ_0 with $\hat{\theta}$
$X_1^*, \ldots, X_n^* \overset{i.i.d}{\sim} f(x|\hat{\theta})$
Take 10000 samples of size 100 from $f(x|\hat{\theta})$
Compute $\hat{\theta}^*$ for each sample

Method 0.3

Measuring Goodess of Estimators:

- unbiased
- Consistency
- Small Variance

"Concentrated around the true value" Suppose θ is the truth then $\forall\theta$, $E[\hat{\theta}] = \theta$

Definition 0.4

Loss Function θ True value
$\hat{\theta}$ Estimator
$L(\hat{\theta}, \theta)$ "Cost of estimating θ as $\hat{\theta}$"
$L(\hat{\theta}, \theta) = 0$ ideally
Non-decreasing in $|\hat{\theta} - \theta|$
Larger mistake are worse
$E[L(\hat{\theta}, \theta)]$ Try to minimize this (Risk)

Method 0.5

In statistics, we very often use squared loss: $L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$

Definition 0.6

Mean Squared Error: $MSE = E\left[(\hat{\theta} - \theta)^2\right]$

$E(\hat{\theta}) = \mu$

$E\left[(\hat{\theta} - \mu + \mu - \theta)^2\right]$

$= E\left[(\hat{\theta} - \mu)^2 + (\mu - \theta)^2 + 2(\hat{\theta} - \mu)(\mu - \theta)\right]$

$= E\left[(\hat{\theta} - \mu)^2\right] + (\mu - \theta)^2 + 2(\mu - \theta)E[\hat{\theta} - \mu]$

$= Var(\hat{\theta}) + (\mu - \theta)^2$

$= Var(\hat{\theta}) + Bias^2$

Definition 0.7

Bias-Variance Decomposition $MSE(\hat{\theta}) = Var(\hat{\theta}) + Bias^2$

$Var(\hat{\theta}) \to$ 'Noise'

$Bias^2 \to$ 'Systematic Error'

Often we encounter bias-variance tradeoff

Example 0.8

$X_1, \ldots, X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$

$\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \to$ unbiased, higher variance, higher MSE

$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \to$ biased, lower variance, lower MSE

Example 0.9

Suppose we have $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mu = \mu, \sigma^2 = \sigma^2$

- What's an estimate of μ with zero bias, but very high variance?
- What's an estimate of μ with very low variance, but very high bias?
- Suppose $\mu = 3$

Solution.

- $\hat{\mu} = X_1$ We can just use one observation
- Pick any random number, numbers do not have variance, but they have mean, so we can pick $3, \pi, etc.$
- Constant estimate, $\hat{\mu} = 3$, when the true mean $\mu = 3$, as 3 has no MSE
  'There is no uniformly most wonderful estimator'
  'Have to constrain the problem to find "optimal" estimator.

■

Definition 0.10

Cramer-Rao inequality

Restrict to unbiased estimators

$E[\hat{\theta}] = \theta, \forall\theta$

Then we can find the “best” lowest variance estimator

Theorem 0.11

Cramer-Rao inequality

$X_1, \ldots, X_n \overset{i.i.d}{\sim} f(x|\theta)$

Let T be an unbiased estimate of θ

Then $Var(T) \geq \frac{1}{nI(\theta)}$

Method 0.12

Cramer-Rao Playbook:

1. Have an unbiased estimator
2. Calculate the variance
3. Check if it achieves the lower bound
4. If it does, then it is the “best” unbiased estimator
5. Use to verify that candidate estimator is best possible unbiased estimator

Example 0.13

Let $X_1, \ldots, X_n \overset{i.i.d}{\sim} Poisson\ f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$

$\mu = \lambda \to \overline{x}$ reasonable estimator

$\sigma^2 = \lambda$

Consider $\overline{X}$. It is unbiased

$Var(\overline{X}) = \frac{\lambda}{n}$

Compute $I(\lambda)$

$I(\lambda) = -E\left[\frac{\partial^2}{\partial\lambda^2}\log f(X|\lambda)\right]$

$= -E\left[\frac{\partial^2}{\partial\lambda^2}\log\frac{\lambda^x e^{-\lambda}}{x!}\right]$

$= -E\left[\frac{\partial^2}{\partial\lambda^2}(x\log\lambda - \lambda - \log x!)\right]$

$= -E\left[\frac{\partial}{\partial\lambda}\left(\frac{x}{\lambda} - 1\right)\right]$

$= -E\left[-\frac{x}{\lambda^2}\right] = \frac{1}{\lambda}$

$\frac{1}{nI(\lambda)} = \frac{\lambda}{n}$

Example 0.14

Toss a biased coin with $P(heads) = p$ n times
Get x heads. $f(x|p)$ Binomial
Estimate p $\frac{x}{n}$
$Var(\frac{x}{n}) = \frac{1}{n^2}Var(x) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$
$f(x|p) = \binom{n}{x}p^x(1-p)^{n-x}$
$l(X|p) = \log f(X|p) = X\log p + (n-X)\log(1-p) + log\binom{n}{X}$
$\frac{\partial l(X|p)}{\partial p} = \frac{X}{p} - \frac{n-X}{1-p}$
$\frac{\partial^2 l(X|p)}{\partial p^2} = -\frac{X}{p^2} - \frac{n-X}{(1-p)^2}$
$I(p) = -E\left[\frac{\partial^2}{\partial p^2}\log f(X|p)\right] = -E\left[-\frac{X}{p^2} - \frac{n-X}{(1-p)^2}\right] = \frac{np}{p^2} + \frac{n(1-p)}{(1-p)^2} = \frac{n}{p} + \frac{n}{1-p} = \frac{n}{p(1-p)}$
$\frac{1}{nI(p)} = \frac{p(1-p)}{n}$

Remark 0.15

Connection to Asymptotic Efficiency:
We learned that $\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} \mathcal{N}(0, \frac{1}{I(\theta)})$
$\hat{\theta}_{MLE} \approx \mathcal{N}(\theta, \frac{1}{I(\theta)})$
For large n, approximately unbiased, variance is approximately bounded by Cramer-Rao lower bound

Takeaways:

1. MSE is a way to measure goodness
2. $MSE = Variance + Bias^2$
3. Often encounter bias-variance tradeoff
4. Cramer-Rao inequality