BSTA 552: Final Exam Review

Spring 2019

/

Final Review

The final exam is technically cumulative but with an emphasis on Chapter 10 material (C&B). It is closed note and closed book except for two sides of *two* 8.5x11 inch “cheat sheets" of notes.

Convergence

- A sequence of random variables is a random variable that depends on n, i.e. $\bar{X}_n$ or $\widehat{\theta}_n$.
- Convergence in probability of $W_n \rightarrow_p W$ or $W_n \rightarrow_p \theta$ describes the limit of a probability statment regarding the distance between W_n and what it is converging to, i.e.
$$\lim_{n\to\infty} P_\theta(|W_n - W| \geq \epsilon) = 0$$
or
$$\lim_{n\to\infty} P_\theta(|W_n - \theta| \geq \epsilon) = 0$$
- Convergence in distribution $W_n \rightarrow_d W$ describes the limit of the CDF of a sequence of random variables to the CDF of another random variable (or a constant θ)
$$\lim_{n\to\infty} F_{W_n}(w) = F_W(w)$$
- Chebychev’s Theorem can help us prove convergence in probability
- The Weak Law of Large Numbers (WLNN) proves convergence in probability of a sample mean to its expectation, $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$:
$$\bar{X}_n \rightarrow E(X_i)$$
- Central Limit Theorem shows convergence in distribution of a sample mean of *iid* random variables:
$$\sqrt{n}(\bar{X}_n - E(X_i)) \rightarrow_d \mathcal{N}(0, Var(X_i))$$
- Convergence in distribution to a constant implies convergence in probability to a constant
- Slutsky’s Theorem is to prove convergence in distribution of linear combinations of random variables
- Delta Method proves that *continuous functions* (not functions with n in them) of random variables that converge to a normal distribution also converge to a normal distribution and it tells us what the variance is
If
$$\sqrt{n}(W_n - \theta) \rightarrow_d \mathcal{N}(0, \sigma^2)$$
then
$$\sqrt{n}(g(W_n) - g(\theta)) \rightarrow_d \mathcal{N}(0, \sigma^2 g'(\theta)^2)$$

Point Estimation

Consistency

Consistency is when a sequence of random variables (depends on n) converges to a constant (i.e. θ) in probability. We want our estimators to be consistent.

- Review concepts (in terms of definitions and how to find/describe them): sequence of estimators, convergence of sequence of estimators, MSE, consistency, limits as $n \to \infty$, convergence in probability, unbiased, asymptotically unbiased
- Mean Square Error = Variance + Bias^2
- Consistent sequence of estimators formal definition is convergence in probability to a constant value, θ:
$$\lim_{n\to\infty} P_\theta(|W_n - \theta| < \epsilon) = 1$$
- If $MSE = E(W_n - \theta)^2 \to 0$ then the estimator W_n is consistent for θ.
- Be able to prove consistency using these theorems
    1. Theorem 10.1.3: If an estimator is asymptotically unbiased and the limiting variance is 0 then we have MSE $\to$ 0, therefore we have consistency.
    2. Theorem 10.1.5: Certain linear combinations of consistent estimators are consistent
    3. Theorem 10.1.6: Consistency of MLEs and continous functions of MLEs

Efficiency

Efficiency is when an estimator has the smallest variance. In *finite samples* that means its variance is equal to the Cramer Rao Lower Bound of the expectation of the estimator (see below). *Asymptotic efficiency* is when the estimator converges to a normal distribution with variance equal to the CRLB for one observation.

Finite sample efficiency:
$$Var(W_n) = \frac{[\frac{d}{d\theta}E(W_n)]^2}{nI_1(\theta)}$$

Asymptotic efficiency:
$$\sqrt{n}(W_n - \theta) \to_d \mathcal{N}\left(0, \frac{1}{I_1(\theta)}\right)$$

- Review concepts: efficiency, limiting variance, convergence in distribution, asymptotic normality, asymptotic variance, asymptotically efficient, Cramer Rao Lower Bound (CRLB), Fisher's Information, Central Limit Theorem, Delta Method, Slutksy's Theorem, Asymptotic relative efficiency
- Asymptotic variance of W_n is the variance of the limiting normal distribution (σ^2) for a series of constants k_n:
$$k_n(W_n - \tau(\theta)) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

- Asymptotically unbiased estimator of θ means

$$\lim_{n\to\infty} E(W_n) - \theta = 0$$

and if

$$\sqrt{n}(W_n - \theta) \to_d \mathcal{N}\left(0, \frac{1}{I_1(\theta)}\right)$$

then W_n is asymptotically unbaised for θ.
- Theorem 10.1.12 Asymptotic efficiency of MLEs (under regularity conditions)
  - MLE $\widehat{\theta}$ is asymptotically efficient, with asymptotic variance $v(\theta) = 1/I_1(\theta) =$ CRLB

$$\sqrt{n}(\widehat{\theta} - \theta) \to_d \mathcal{N}\left(0, \frac{1}{I_1(\theta)}\right)$$

  - Delta Method: functions of asymptotically normal variables are also asymptotically normal, so we also have asymptotic efficiency of a continuous function of the MLE $\tau(\theta)$:

$$\sqrt{n}\left[\tau(\widehat{\theta}) - \tau(\theta)\right] \xrightarrow{d} \mathcal{N}\left(0, \frac{\tau'(\theta)^2}{I_1(\theta)}\right)$$

  - Asymptotically unbiased, asymptotically normal, minimum variance
  - Proof relies on Taylor Series expansion of score function (derivative of log likelihood function)
  - This theorem also implies consistency of the MLEs (by Slutsky's Theorem)
- Asymptotic relative efficiency compares the asymptotic variances of two estimators. Since the MLE is asymptotic efficient estimator, any other estimator will have relative efficiency > 1 when compared to the MLE.
- When estimating the asymptotic variance in finite samples, this is the "large sample variance" of the estimator. We need to estimate Fisher's information. There are two ways to do this:
  - Expected Information is using Fisher's information as we usually define it $I(\theta) = E[-\ell''(\theta|\mathbf{x})]$ and plugging in $\widehat{\theta}$ after taking the expecatation to estimate $I(\theta)$, i.e. $I(\widehat{\theta}) = E[-\ell''(\theta|\mathbf{x})]|_{\theta=\widehat{\theta}}$.

$$Var\left[\sqrt{n}(\widehat{\theta} - \theta)\right] \approx \frac{1}{I_1(\widehat{\theta})} \Longrightarrow Var(\widehat{\theta}) \approx \frac{1}{nI_1(\widehat{\theta})}$$

  - Observed information is when we do not take the expectation and just plug in $\widehat{\theta}$ for θ, i.e. $\hat{I}(\widehat{\theta}) = -\ell''(\theta|\mathbf{x})|_{\theta=\widehat{\theta}}$.

$$Var(\widehat{\theta}) \approx \frac{1}{n\hat{I}_1(\widehat{\theta})}$$

Hypothesis Testing

1. Review concepts: asymptotic LRTs, Wald tests, Score tests, asymptotic pivot, Score function
2. Asymptotic distribution of LRTs
   - Proof uses distribution of second term of Taylor's series of log-likelihood around $\widehat{\theta}$

- Theorem 10.3.1 Asymptotic distribution of LRT simple hypothesis $-2\log\lambda(\mathbf{X}) \to \chi_1^2$ in distribution
- Theorem 10.3.3 null hypothesis concerns a vector of parameters $-2\log\lambda(\mathbf{X}) \to \chi_\nu^2$

3. Wald Tests
$$\frac{\widehat{\theta}-\theta}{\sqrt{v(\widehat{\theta})/n}} \xrightarrow{d} \mathcal{N}(0,1)$$
   - Proof uses asymptotic normality of MLEs, and slutsky's theorem for using an approximation of the variance (CRLB).
   - Any consistent estimator of the standard error of $\widehat{\theta}$ can be in the denominator.
4. Score Tests
$$\frac{S(\theta|\mathbf{X})}{\sqrt{I_n(\theta)}} \xrightarrow{d} \mathcal{N}(0,1)$$
   - Proof uses Central Limit Theorem since the score function is a sum of iid variables and has variance equal to the information
5. Convergence in distribution of the test statistics is *under the null* H_0 to preserve the approximate size of the test
6. Examples 10.3.2, 10.3.4, 10.3.5, 10.3.6

Interval Estimation

1. Review concepts: inverting asymptotic tests to confidence sets, asymptotic confidence sets/intervals, asymptotic LRT confidence intervals, Wald intervals, Score intervals, intervals with asymptotic pivots (large sample pivot), approximate
2. Asymptotic pivotal quantities are sequences of random variables that converge in distribution to a random variable that has CDF free of θ.
3. Wald interval can be constructed using a pivot (the wald statistic) or using inversion of hypothesis test
4. Use the Delta Method to obtain intervals for $g(\theta)$ based on the Wald interval for $\widehat{\theta}$.
5. Score intervals (invert Score tests)
6. Asymptotic Likelihood ratio intervals (invert LRT tests)
7. Examples 10.4.1, 10.4.2, 10.4.3, 10.4.4, 10.4.5, 10.4.6, 10.4.9