BSTA 552: Midterm Review

Spring 2019

Midterm Review

Hypothesis Testing

1. Hypothesis testing consists of comparing null and an alternative hypothesis using our data $\mathbf{X}$. The null hypothesis gives us a null parameter space Θ_0 while the alternative gives us Θ_1 and the full parameter space is $\Theta = \Theta_0 \bigcup \Theta_1$. Under assumptions about how $\mathbf{X}$ is distributed, we develop a test statistic, and a rejection region $\mathcal{R}$ that specifies for what values of that test statistic do we reject the null hypothesis.
   - For LRT and Neyman-Pearson tests, the test statistic is a ratio of likelihoods (or pdfs) with theta representing the null or alternative spaces. We *reject when the likelihood under the null (i.e. when $\theta \in \Theta_0$) is small.* For LRT this means rejecting when $\lambda(\mathbf{x})$ is small. For Neyman-Pearson, this means rejecting when $f(x|\theta_1)/f(x|\theta_0)$ is large.
   - For Karlin-Rubin one sided tests, if we have an MLR in a sufficient statistic, our rejection region is simplified to just be based on that sufficient statistic being large or small (depending on the side of the test).
   - Lastly, for all tests, we choose our cutoff value (i.e. c) that defines our rejection region to give us a size or level α test. Neyman-Pearson and Karlin Rubin tell us that the size can be evaluated when $\theta = \theta_0$, so $\alpha = P_{\theta_0}(\mathbf{X} \in \mathcal{R})$. For a general LRT, we need to take the supremum of $P_{\theta_0}(\mathbf{X} \in \mathcal{R}) = P_\theta(\lambda(\mathbf{x}) \leq c)$ over all values of θ in Θ_0. Solving for c for a given α often requires knowing the distribution of the test statistic and using a critical value of that distribution or the cdf of that distribution.
   - Review concepts (in terms of definitions and how to find/describe them) such as: type I error, type II error, power function, unbiased tests, rejection region, size and level of tests, uniformly most powerful tests, p-values
   - Note we are talking about *finite sample* tests and rejection regions. These tests have appropriate size for all sample sizes (i.e. "exact tests"). We have not yet covered asymptotic tests (i.e. Wald test, score test, asymptotic LRTs). This is Chapter 10 of C&B.
2. Evaluate tests:
   - $\beta(\theta)$ is the power function of a test, which is the probability that we reject when θ is the true parameter: $P_\theta(\mathbf{x} \in \mathcal{R})$ or $P(\mathbf{x} \in \mathcal{R}|\theta)$, a function of θ.
   - Note the power function is evaluated over all possible values of θ, that is, both the null and alternative spaces. UMP specifies what the power function looks like on the alternative space. Size and level of test specifies what the power function looks like on the null space.
   - UMP (Uniformly Most Powerful) tests of a certain class $\mathcal{C}$ are tests such that the power function is greater than or equal to every other test of class $\mathcal{C}$, for all θ in the alternative space $\theta \in \Theta_0^C$.

- level α tests have power function less than or equal to α for all $\theta \in \Theta_0$, while size α tests have maximum (supremum) power equal to α over all $\theta \in \Theta_0$.
- When θ is in the null space, $\beta(\theta)$ is the probability of a type I error. When θ is in the alternative space, $\beta(\theta)$ is *1* - the probability of type II error.
  - The probability of a type I error is $\beta(\theta_0)$ for a specified $\theta_0 \in \Theta_0$. It is just one point on the curve. If the null space consists of just one point i.e $\Theta_0 = \{\theta_0\}$, then the size and level of the test are the same, and the type I error is equal to the size. In this special case, there is only one point in the null space to evaluate the power function at, so we don't need to take the supremum to find the size of the test.
  - The probability of a type II error is $1 - \beta(\theta_1)$ for a specified $\theta_1 \in \Theta_1$. It is just one point on the curve.
  - "Power" as defined in your applied classes is just referring to one point on the power curve, that is $\beta(\theta_1)$ for a specified $\theta_1 \in \Theta_1$.
3. Likelihood ratio tests
   - Likelihood ratio tests are used for general types of hypothesis testing (complex, simple). They are very flexible and can be used to test one or multiple parameters. They are *not guaranteed to be uniformly most powerful (UMP) unless you use either a Neyman-Pearson or Karlin Rubin theorem* to show that the LRT reduces to the same exact rejection region as specified in those theorems.
   - The likelihood ratio test statistic and rejection region for testing $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^C$:
     $$\mathcal{R} = \{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$$
     where
     $$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} \mathcal{L}(\theta|\mathbf{x})}{\sup_{\Theta} \mathcal{L}(\theta|\mathbf{x})}$$
   - The test statistic can reduce to
     $$\lambda(\mathbf{x}) = \frac{\mathcal{L}(\widehat{\theta}_0|\mathbf{x})}{\mathcal{L}(\widehat{\theta}|\mathbf{x})}$$
     where $\widehat{\theta}$ is the MLE under the full likelihood space and $\widehat{\theta}_0$ is the MLE under the restricted likelihood under the null space, or θ_0 in the case where $\Theta_0 = \{\theta_0\}$.
   - If we have a sufficient statistic, $T(\mathbf{x})$, we can use the LRT statistic:
     $$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} \mathcal{L}^*(\theta|T(\mathbf{x}))}{\sup_{\Theta} \mathcal{L}^*(\theta|T(\mathbf{x}))}$$
     where $\mathcal{L}^*$ is based on the distribution function of $T(\mathbf{x})$.
   - To obtain a size α test we need to choose our cutoff c such that:
     $$\sup_{\theta \in \Theta_0} P_\theta(\mathbf{X} \in \mathcal{R}) = \sup_{\theta \in \Theta_0} P_\theta(\lambda(\mathbf{X}) < c) = \alpha$$
4. Neyman-Pearson Lemma and the sufficient statistic Corollary 8.3.13
   - *Simple* hypothesis $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$.
   - UMP level α tests where $\alpha = P_{\theta_0}(\mathbf{X} \in \mathcal{R})$ (note it's easy to calculate α since just one point in the null space)

- We can define rejection region $\mathcal{R} = \{\mathbf{x} : \frac{f_{\mathbf{X}}(\mathbf{x}|\theta_1)}{f_{\mathbf{X}}(\mathbf{x}|\theta_0)} > k\}$ or $\mathcal{R} = \{\mathbf{x} : \frac{g_T(t|\theta_1)}{g_T(t|\theta_0)} > k\}$ (assuming the ratio exists and you aren't dividing by 0)
- So to obtain a UMP α level test if $\mathcal{R} = \{\mathbf{x} : \frac{f_{\mathbf{X}}(\mathbf{x}|\theta_1)}{f_{\mathbf{X}}(\mathbf{x}|\theta_0)} > k\}$, we choose k such that $P_{\theta_0}\left(\frac{f_{\mathbf{X}}(\mathbf{x}|\theta_1)}{f_{\mathbf{X}}(\mathbf{x}|\theta_0)} > k\right) = \alpha$
- Notice how this ratio is different than the likelihood ratio test we learned first, but it still rejects when the likelihood under the null is sufficiently small.
- For some discrete distributions you can calculate the ratio for all possible rejection regions, order the rejection regions by k (increasing or decreasing) and choose your cut k so that you reject whenever the ratio is larger than that k.
- Corollary says we can use the distribution of a sufficient statistic T instead $\mathcal{R} = \{\mathbf{x} : \frac{g_T(t|\theta_1)}{g_T(t|\theta_0)} > k\}$.

5. Karlin-Rubin Theorem
   - *One-sided composite* hypothesis
   - Need monotone likelihood ratio (MLR) in sufficient statistic T – note this is not the same likelihood ratio as in a LRT (as C&B defines it). Show either by:
     - For $\theta_1 > \theta_2$, show
     $$\frac{f(t|\theta_1)}{f(t|\theta_2)} = V(t)$$
     where $V(t)$ is a monotone non-decreasing function of T.
     - Show the joint distribution is a regular exponential family with $f(x|\theta) = h(x)c(\theta)e^{w(\theta)T(x)}$ where $w(\theta)$ is a non-decreasing function in θ.
   - The rejection region is simply in terms of the sufficient statistic T. No likelihood ratios needed (MLR takes care of this)!!
   - UMP level α tests where $\alpha = P_{\theta_0}(\mathbf{X} \in \mathcal{R})$ (note we don't need the supremum over Θ_0 since we have an MLR guarantees the supremum is at θ_0)
   - $\mathcal{R} = \{T > c\}$ or $\mathcal{R} = \{T < c\}$ and UMP level $\alpha = P(\mathbf{x} \in \mathcal{R}|\theta_0)$. Often we know the distribution of T or some monotone function of T so we can define c in terms of the CDF of that distribution, or a critical value of that distribution (i.e. z_α).

6. p-values
   - A p-value is a test statistic (it is random because the data is random!)
   - Obtain a p-value of a test based on another test statistic $W(\mathbf{X})$ where we reject H_0 for large values of $W(\mathbf{X})$. If we have observed data $W(\mathbf{x})$ we can find a valid p-value by defining it as:
   $$p(\mathbf{x}) = \sup_{\theta \in \Theta_0} P_\theta(W(\mathbf{X}) \geq W(\mathbf{x}))$$
   - Valid p-values are such that $P_\theta(p(\mathbf{X}) \leq \alpha) \leq \alpha$ for all $\theta \in \Theta_0$ and for any $\alpha \in [0, 1]$. This means that we reject the null hypothesis with maximum type I error α by rejecting when the p-value is less than α.

7. Good practice problems in C&B: 8.1, 8.2, 8.6a&b, 8.12 8.15, 8.16, 8.17, 8.19, 8.23, 8.25, 8.27, 8.28, 8.31, 8.37, 8.38, 8.41, 8.45

Interval Estimation

1. We learned that a confidence interval (or more generally, a confidence region or set) is a random interval based on the data $\mathbf{X}$ that is used to estimate θ with a certain amount of precision. It is more than just a point estimate of θ. We are most interested in the confidence coefficient and coverage probability of our confidence set. For example, if we have a confidence interval $[L(\mathbf{X}), U(\mathbf{X})]$ for a parameter θ, the coverage probability is

$$P_\theta(L(\mathbf{X}) \leq \theta \leq U(\mathbf{X}))$$

and the confidence coefficient is the infimum over all coverage probabilities for all possible θ:

$$\inf_\theta P_\theta(L(\mathbf{X}) \leq \theta \leq U(\mathbf{X}))$$

Coverage probability can depend on θ, so the confidence coefficient gives us a minimal probability that our interval estimator is covering the true parameter θ, no matter what it is.
2. When we say a "$1 - \alpha$ confidence interval" we mean the confidence interval has confidence coefficient $\geq 1 - \alpha$.
3. A confidence set or region can be a union of multiple disjoint intervals. Hence, in general we refer to this procedure as "interval estimation."
4. We write a confidence interval/set as $C(\mathbf{X})$ as in $C(\mathbf{X}) = \{\theta : \theta \in [L(\mathbf{X}), U(\mathbf{X})]\}$ to denote it is a random interval based on data $\mathbf{X}$. When we sample from our underlying distribution of $\mathbf{X}$ and obtain observed data $\mathbf{x}$, we have $C(\mathbf{x})$ which is one observed interval based on observed data (it is no longer random just as $\mathbf{x}$ is not random but an observation; it is "fixed").
5. We learned how to construct confidence sets/intervals by inverting test statistics, by using the distribution of a pivot, or by pivoting the CDF of a statistic
6. Note, again, that these intervals have been based on *finite sample* distributions of our data. We will learn about asymptotic based intervals with approximate $1 - \alpha$ coverage when covering Chapter 10.
7. Review concepts: interval estimate, interval estimator, inverting test statistics, coverage probability, confidence coefficient, pivot
8. Theorem 9.2.2 Inverting a test statistic $\to$ need to invert a family of tests/null hypotheses (using the acceptance region for each $\theta_0 \in \Theta$) to obtain a confidence set; shows how the level of a test corresponds to the coverage of an confidence set.
   - As C&B states, "The hypothesis test fixes the parameter and asks what sample values (the acceptance region) are consistent with that fixed value [the null space]. The confidence set fixes the sample value [$\mathbf{x}$] and asks what parameter values (the confidence interval) make this sample value most plausible."
   - The confidence interval $C(\mathbf{X})$ is the set of θ's such that, for the given data $\mathbf{X}$ and for each $\theta_0 \in C(\mathbf{X})$ you would not be able to reject the null hypothesis H_0.
   - In hypothesis testing, acceptance region is the set of $\mathbf{X}$ which are very likely for a fixed Θ_0. In interval estimation, a confidence interval is a set of θ's which make $\mathbf{X}$ very likely for a fixed $\mathbf{X}$.
   - Thm 9.2.2 also tells us that if we start with a confidence interval, we can invert the interval back to obtain an acceptance region and hence formulate a rejection region based on a confidence interval. The relationship is 1 to 1. So, even when we are making a

confidence interval with a pivot, we would be able to construct a rejection region for a test of a specific H_0 based off of that confidence interval.

9. Pivotal quantities
   - A pivot is a random variable $Q(\mathbf{X}, \theta)$ that is a function of the data (or a sufficient statistic T) and the parameter θ, such that the distribution of a pivot does not depend on the parameter (i.e. $Q(\mathbf{X}, \theta) \sim$ a standard normal(0,1), or χ_n^2, or $\Gamma(n, 1)$ or $Uniform(0, 1)$ etc)
   - using a pivot we can create a confidence set and find the endpoints using the distribution of the pivot
   - can also pivot the CDF (since the CDF of a random variable has Uniform(0,1) distribution). We can use the theorem that tells us how to find the endpoints $U(\mathbf{X}) = \theta_U(\mathbf{X})$ and $L(\mathbf{X}) = \theta_L(\mathbf{X})$.
   - You can show something is a pivot by either:
     (a) Using the transformation theorem or logic/known facts to determine the distribution of $Q(\mathbf{X}, \theta)$ or $Q(T, \theta)$. For instance, if we know the distribution of T and $Q(T, \theta) = T/\theta$ we can use the transformation theorem to obtain the distribution of $Q(T, \theta)$. Remember the transformation theorem (Theorem 2.1.8 in C&B) tells us if we know the distribution of T then we can find the distribution of $Y = Q(T)$ (if it is monotone or piece-wise monotone in T):

     $$f_Y(y) = f_T(Q^{-1}(y)) \left| \frac{d}{dy} Q^{-1}(y) \right|$$

     (b) Show the distribution of a statistic T, $f(t|\theta)$ can be expressed in the form:

     $$f_T(t|\theta) = g(Q(t, \theta)) \left| \frac{\partial}{\partial t} Q(t, \theta) \right|$$

     for some function g (that turns out to be the distribution function of $Q(T, \theta)$) and some monotone function Q (monotone in t for each θ).
   - In general, differences (i.e. $\bar{X} - \mu$) are pivotal for location problems, while ratios or products (i.e. $\sum X_i/\theta$) are pivotal for scale problems.

10. Evaluate intervals $[L(\mathbf{X}), U(\mathbf{X})]$ of a certain coverage by the length $U(\mathbf{X}) - L(\mathbf{X})$ or expected length $E_\theta[U(\mathbf{X}) - L(\mathbf{X})]$.
   - Often, given a confidence coefficient $1 - \alpha$, we wish to find the CI with the shortest length.
   - Theorem 9.3.2 tells us that for a unimodal pdf $f(x)$ (not necessarily symmetric) the shortest interval $[a, b]$ is that which $f(a) = f(b)$ where $a \leq x^* \leq b$ with x^* the mode of $f(x)$.

11. Good practice problems in C&B: 9.1, 9.4a,c, 9.5, 9.6, 9.7,9.9, 9.11, 9.13, 9.16, 9.17, 9.23, 9.25, 9.34, 9.35, 9.36, 9.37

Other concepts to know

1. How to find an MLE of θ
2. How to find a sufficient statistic

3. Transformation theorem to find the distribution of a monotone function of $\mathbf{X}$
4. How to find the distribution of a order statistic: $X_{(1)}$ or $X_{(n)}$
5. How to calculate the CDF of a random variable from a pdf distribution

Other Midterm Review Practice Problems

1. Suppose $X_1, \ldots, X_n \sim \mathcal{N}(0, \sigma^2)$. We wish to test $H_0 : \sigma^2 \leq \sigma_0^2$ vs $H_1 : \sigma^2 > \sigma_0^2$. What is the UMP level α test? Invert this test to obtain a $1 - \alpha$ level confidence region.
2. Suppose $X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ where μ is unknown and σ is known. Consider two confidence intervals for μ :
   (a) $\mu \in [\bar{X} - a/\sqrt{n}, \bar{X} + a/\sqrt{n}], a > 0$
   (b) $\mu \in [\bar{X} - \sigma a/\sqrt{n}, \bar{X} + \sigma a/\sqrt{n}], a > 0$
   What is the coverage probability and confidence coefficient for each interval?
3. Find a $1 - \alpha$ confidence interval for θ, using pivots, given $X_1, \ldots, X_n$ iid with pdf
   (a) $f(x|\theta) = 1, \theta - \frac{1}{2} < x < \theta + \frac{1}{2}$.
   (b) $f(x|\theta) = \frac{2x}{\theta^2}, 0 < x < \theta, \theta > 0$.
4. Let $X_1, \ldots X_n$ be iid with pdf $f(x|\mu) = e^{-(x-\mu)} I_{[\mu,\infty)}(x)$. Then $Y = X_{(1)}$ is sufficient for μ with pdf
   $$f_Y(y|\mu) = n e^{-n(y-\mu)} I_{[\mu,\infty)}(y).$$

   In example 9.2.13 in class, we constructed a $1 - \alpha$ confidence interval for μ using the method of pivoting the CDF. Compare the length of that interval to lengths of $1 - \alpha$ intervals obtained by likelihood ratio test and pivotal methods.