Lecture 17 — November 19

Lecturer: Lester Mackey *Scribe: Guanyang Wang, Luke Lefebure*

Warning: *These notes may contain factual and/or typographic errors.*

17.1 Recap

In the last lecture, we saw how to use the strategy of conditioning on a sufficient statistic to find a UMPU test. We begin by extending an example seen in the last lecture.

Example 1. $X_1, ..., X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, where both μ and σ^2 are unknown. The joint density of $X_1, ..., X_n$ is

$$f(x_1, \ldots, x_n) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_i x_i^2 + \frac{\mu}{\sigma^2}\sum_i x_i\right)$$

Testing the mean In the last lecture, we considered testing $H_0 : \sigma \geq \sigma_0$ vs. $H_1 : \sigma < \sigma_0$. Next, let's consider testing $H_0 : \mu \leq 0$. Since $\mu \leq 0 \Leftrightarrow \frac{n\mu}{\sigma^2} \leq 0$, let

$$\theta = \frac{n\mu}{\sigma^2}, \lambda = -\frac{1}{2\sigma^2}$$
$$U = \bar{X}, T = \sum_{i=1}^n X_i^2.$$

An UMPU test has the form $\phi(X) = \mathbb{I}\left(\bar{X} > C_\alpha(\sum_{i=1}^n X_i^2)\right)$, set according to $\mathbb{P}_{\mu=0}(\bar{X} > C_\alpha(\sum_{i=1}^n X_i^2) | \sum_{i=1}^n X_i^2) = \alpha$. In order to simplify the selection of $C_\alpha(\sum_{i=1}^n X_i^2)$ and the form of this test, let us try to generalize the argument employed for variance testing.

Consider a general multiparameter exponential family of the form

$$\mathbb{P}_{\theta,\lambda}(x) = \exp\left(\theta U(x) + \sum_{i=1}^k \lambda_i T_i(x) - A(\theta, \lambda)\right) h(x).$$

Suppose there exists $V = h(U, T)$ which is independent of T when $\theta = \theta_0$ and increasing w.r.t U given fixed T. Then we can find an UMPU test of the form

$$\phi = \begin{cases} 1 & \text{if } V > C_\alpha \\ \gamma & \text{if } V = C_\alpha \\ 0 & \text{if } V < C_\alpha \end{cases}$$

for a single constant C_α. To see this, consider

$$\begin{aligned}\alpha &= \mathbb{P}_{\theta_0}\left[U > C_\alpha(T)|T\right] \\ &= \mathbb{P}_{\theta_0}\left[h(U,T) > h(C_\alpha(T),T)|T\right] = \mathbb{P}_{\theta_0}\left[V > C'_\alpha(T)|T\right] \\ &\qquad\qquad (V \text{ is increasing w.r.t } U \text{ for fixed } T) \\ &= \mathbb{P}_{\theta_0}\left[V > C_\alpha\right] \\ &\qquad\qquad (V,T \text{ are independent} \Rightarrow \text{can choose } C'_\alpha(T) \text{ constant w.r.t } T)\,.\end{aligned}$$

To apply this observation in our setting, we notice that

$$h(U,T) = \frac{U}{\sqrt{T - nU^2}} = \frac{\bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}}$$

is increasing in U for each T and independent of T (since it is ancillary for $\lambda = -\frac{1}{2\sigma^2}$, and T is complete sufficient when θ fixed). Hence, we can find an UMPU test that rejects iff $h(U,T) > C_\alpha$. Equivalently, this test rejects iff

$$t = \sqrt{n(n-1)}h(U,T) = \frac{\sqrt{n}\bar{X}}{\sqrt{\frac{\sum_i (X_i - \bar{X})^2}{n-1}}}$$

is large. The statistic t has a Student's t_{n-1} distribution when $\mu = 0$, so we can set the critical value for t to be a $(1-\alpha)$ quantile of the Student's t_{n-1} distribution.

17.2 UMP Invariant Tests

When a testing problem is unchanged by certain transformations of the input data, it is often desirable to employ a testing procedure that is similarly invariant to these transformations. Applying such invariance constraints often greatly reduces the class of valid tests and allows for the construction of a UMP *invariant* test even when no unconstrained UMP test exists. Let's begin with the example introduced in the previous lecture.

Example 2. Suppose that we observe $X = (X_1, \ldots, X_d)$, with independent coordinates $X_i \sim \mathcal{N}(\theta_i, 1)$ and that we are interested in the hypothesis testing problem

$$H_0 : \theta_1 = \ldots = \theta_d = 0 \quad \text{vs.} \quad H_1 : \exists i \in \{1, \ldots, d\} : \theta_i \neq 0.$$

If we transform our data so that $X' = OX$ for $O \in \mathbb{O}(d)$, the set of $d \times d$ orthogonal matrices, i.e., the set of square matrices such that $O^\top O = OO^\top = I$, then we have $X'_i \overset{\text{ind}}{\sim} \mathcal{N}(\theta'_i, 1)$ for $\theta' = O\theta$, and our hypothesis testing problem can be seen to be equivalent to testing

$$H_0 : \theta'_1 = \ldots = \theta'_d = 0 \quad \text{vs.} \quad H_1 : \exists i \in \{1, \ldots, d\} : \theta'_i \neq 0.$$

Hence, when searching for a test, the **principle of invariance** would suggest constraining $\phi(X) = \phi(OX)\ \forall O \in \mathbb{O}(d)$. In this case, it can be checked that ϕ is invariant in this way iff it is a function of the magnitude of the vector of samples, i.e. of $T = \sum_i X_i^2$. This tells us that if we only care about invariant tests, our optimality goal is to search for UMP tests among functions of T.

In this example, T is non-central chi-squared distributed with d degrees of freedom, i.e., $T \sim \chi_d^2(\psi^2)$ with a non-centrality parameter $\psi^2 = \sum_{i=1}^d \theta_i^2$. Thus, we can simplify the null and alternative hypotheses of our testing problem to

$$H_0 : \psi = 0 \quad \text{vs.} \quad H_1 : \psi > 0. \quad \text{(one-sided test)}$$

Note that we have reduced both the relevant data and the relevant parameter space for our testing problem; these are common advantages of imposing invariance constraints. To derive our UMP invariant test, we will check that the $\chi_d^2(\psi^2)$ has monotone likelihood ratios in T. Note that the density of non-central chi-squared distribution has the form:

$$f(t;\psi) = e^{\frac{-\psi^2}{2}} \sum_{k=0}^{\infty} \frac{(\frac{\psi^2}{2})^k}{k!} \frac{t^{\frac{d}{2}-1+k} e^{-\frac{t}{2}}}{2^{k+\frac{d}{2}}\Gamma(k+\frac{d}{2})}.$$

The likelihood ratio can therefore be computed as

$$\frac{p_{\psi^2}(t)}{p_{\psi^2=0}(t)} = \frac{f(t;\psi)}{\frac{t^{\frac{d}{2}-1}e^{-t/2}}{2^{\frac{d}{2}}\Gamma(\frac{d}{2})}} = e^{-\frac{\psi^2}{2}} \sum_k c_k \left(\frac{\psi^2}{2}\right)^k t^k.$$

where c_k are non-negative constants. We can see that each term in the sum above increases in t, and thus the ratio as a whole is increasing in t. (We only need to compare each parameter value with 0, because our null hypothesis is simple.) Thus this family has MLR, and the UMPI test rejects when T is large.

17.3 A General Framework for Invariant Tests

Let us now consider the general case of $X \in \mathcal{X}$, $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$, and hypotheses

$$H_0 : \theta \in \Omega_0, \quad H_1 : \theta \in \Omega_1.$$

Our general notion of invariance will be with respect to a **group of transformations**:

Definition 1. A **group** $\mathcal{G}$ is a set equipped with an operation (composition) satisfying certain axioms:

- Closure: $g_1 g_2 \in \mathcal{G}$ for all $g_1, g_2 \in \mathcal{G}$
- Associativity: $(g_1 g_2) g_3 = g_1 (g_2 g_3)$ for all $g_1, g_2, g_3 \in \mathcal{G}$
- Identity: $\exists e \in \mathcal{G} : e g_1 = g_1 = g_1 e$ for all $g_1 \in \mathcal{G}$
- Inverses: for any $g \in \mathcal{G}, \exists g^{-1}$ satisfying $g g^{-1} = g^{-1} g = e$.

We restrict our focus to groups $\mathcal{G}$ in which each $g \in \mathcal{G}$ is a bijection from $\mathcal{X} \to \mathcal{X}$, and if $X \sim P_\theta$, then $X' = gX \sim P_{\theta'}$ for some θ'. Let $\overline{g}$ be the mapping on Ω induced by g, so that $\theta' = \overline{g}(\theta)$ when $X \sim P_\theta$ and $X' = gX \sim P_{\theta'}$.

We will say that a testing problem is invariant under the group of transformations if $\mathcal{G}$ preserves the structure of the testing problem. More precisely,

Definition 2. A testing problem **remains invariant** under $\mathcal{G}$ if $\overline{g}\Omega_0 = \Omega_0$ and $\overline{g}\Omega_1 = \Omega_1$ $\forall g \in \mathcal{G}$.

Hence, $\mathcal{G}$ must map null hypotheses to null hypotheses and alternative hypotheses to alternative hypotheses. We've seen this behavior in the orthogonal matrix example. Similarly, we introduce a notion of invariance for a test.

Definition 3. A test ϕ is **invariant** under $\mathcal{G}$ if $\phi(gx) = \phi(x)$ $\forall x \in \mathcal{X},\ g \in \mathcal{G}$.

Example 3. Suppose $X_1, \ldots X_n \sim$ i.i.d. on $(0,1)$. Taking f to be a known density supported on $(0,1)$, we desire to test:

$$H_0 : X_i \sim \text{i.i.d. } U(0,1), \quad H_1 : X_i \sim \text{i.i.d. } g_1(x) = f(x) \text{ or } g_2(x) = f(1-x)$$

Using the transformation $X_i' = 1 - X_i$ results in invariance for the problem. This is because under H_0 we have that $1 - X_i \sim$ i.i.d. $U(0,1)$, and the alternative is invariant by construction. What does it mean for ϕ to be invariant here?

$$\phi \text{ invariant } \Leftrightarrow \phi(X_1, \ldots, X_n) = \phi(1 - X_1, \ldots, 1 - X_n)$$

For any invariant test ϕ, $\mathbb{E}_{g_1}[\phi(X)] = \mathbb{E}_{g_1}[\phi(1-X)] = \mathbb{E}_{g_2}[\phi(X)]$, where the first equality follows by invariance of ϕ and the second equality follows by definition of g_2. Hence, we have

$$\begin{aligned}
\mathbb{E}_{g_1}[\phi(X)] &= \mathbb{E}_{g_2}[\phi(X)] \\
&= \frac{1}{2}\mathbb{E}_{g_1}[\phi(X)] + \frac{1}{2}\mathbb{E}_{g_2}[\phi(X)] \text{ (from the above fact that both quantities are equal)} \\
&= \mathbb{E}_p[\phi(X)] \quad (*)
\end{aligned}$$

where $(*)$ comes because in this case, taking the average power is equivalent to taking the power of the average test. Here $\mathbb{E}_p[\cdot]$ is an expectation with respect to the measure with the following invariant density $p(x_1, \ldots, x_n)$:

$$p(x_1, \ldots, x_n) = \frac{\prod_{i=1}^n g_1(x_i) + \prod_{i=1}^n g_2(x_i)}{2} = \frac{\prod_{i=1}^n f(x_i) + \prod_{i=1}^n f(1-x_i)}{2}$$

As a result, ϕ is UMP invariant iff it is MP invariant against $H_1' = p$. For the simple testing problem of H_0 versus H_1', Neyman-Pearson implies that the most powerful ϕ' rejects for large $p(x_1, \ldots, x_n)$. Since ϕ' is invariant (because $p(\cdot)$ is invariant), it is UMPI.

17.4 Characterizing Invariant Tests

Now that we've defined invariant tests, we will develop a characterization that will ease our derivation of invariant tests for new testing problems. We begin by exploring the relationship between invariance and the **orbits** of a group.

Definition 4. Let G be a group of transformations $\mathcal{X} \to \mathcal{X}$. Then x and $y \in \mathcal{X}$ are in the same **orbit** of G if and only if there is some transformation $g \in G$ such that $y = gx$.

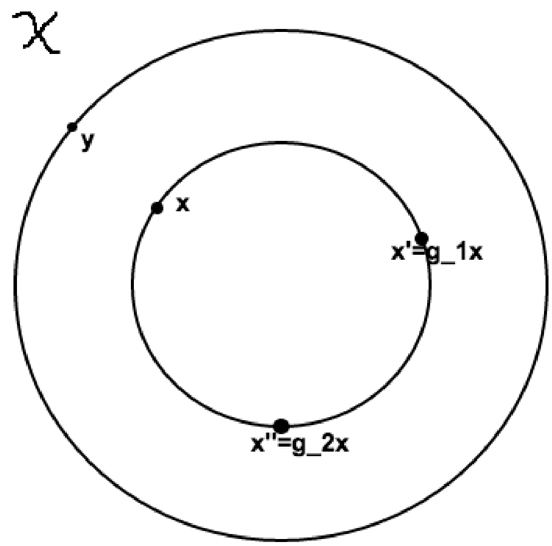

Figure 17.1. Orbit of transformations: x, x', and x'' are in the same orbit, since we can move between any of them with transformation $g \in G$. y is not in the same orbit, since there is no transformation in G such that $y = gx$.

In fact, invariant functions are constant on orbits. Indeed, this is what it means to be invariant. Let T be an invariant function, and consider $x \in \mathcal{X}$; we can reach any y that is in the same orbit as x by applying one of the transformations $g \in G$. By the definition of invariance, $T(x) = T(gx) = T(y)$.

Some invariant functions achieve greater compression of the data X than others. As an extreme example, $T(x) \equiv 0$ is invariant and achieves maximal compression. However, we are interested in the *least* compressed invariant functions. These **maximal invariant** functions maintain the most information about the original variable X, and we will see that any other invariant function is a function of these maximal invariants.

Definition 5. A function $T(x)$ is **maximal invariant** if

(a) $T(x)$ is invariant, i.e., $T(x) = T(gx)$ for all $x \in \mathcal{X}$ and $g \in G$, and

(b) $T(x)$ takes on distinct values on distinct orbits.

So in summary, $T(x)$ is maximal invariant if $T(x) = T(y) \iff \exists g \in G$ such that $y = gx$.

Example 4. Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, and let $G = \{g | g(\mathbf{x}) = (x_1 + a, x_2 + a, \ldots, x_n + a), a \in \mathbb{R}\}$ (shift every coordinate by a common constant). Then $T(\mathbf{x}) = (x_1 - x_n, x_2 - x_n, \ldots, x_{n-1} - x_n)$ is maximal invariant.

To see this, let $T(\mathbf{x}) = T(\mathbf{y})$. Then it must be that $y_1 - y_n = x_1 - x_n, \ldots, y_{n-1} - y_n = x_{n-1} - x_n$, so $y_1 = x_1 - x_n + y_n, \ldots, y_{n-1} = x_{n-1} - x_n + y_n$. Let $a = y_n - x_n$. Then there is a transformation $g \in G$ such that $g(\mathbf{x}) = (x_1 + (y_n - x_n), \ldots, x_{n-1} + (y_n - x_n), x_n + (y_n - x_n)) = (y_1, \ldots, y_{n-1}, y_n)$. So $T(\mathbf{x}) = T(\mathbf{y})$ implies that x and y are in the same orbit of G.

Now suppose $\mathbf{y} = g(\mathbf{x})$. Then there is some $a \in \mathbb{R}$ such that $y_1 = x_1 + a, \ldots, y_n = x_n + a$. So $T(\mathbf{y}) = (y_1 - y_n, \ldots, y_{n-1} - y_n) = (x_1 + a - (x_n + a), \ldots, x_{n-1} + a - (x_n + a)) = (x_1 - x_n, \ldots, x_{n-1} - x_n)$. So $\mathbf{y} = g(\mathbf{x})$ implies $T(\mathbf{x}) = T(\mathbf{y})$.

Now we can precisely characterize an invariant test.

Theorem 1. A test ϕ is invariant if and only if it is a function of a maximal invariant function T.

Proof. Let $\phi(x) = f \circ T(x)$. By the definition of invariance, $\phi(x) = f \circ T(gx) = \phi(gx)$. So if ϕ is a function of a maximal invariant T, then ϕ is invariant.

Now suppose ϕ is invariant, and let $T(x_1) = T(x_2)$ for a maximal invariant function T. Then we know that $x_2 = gx_1$ for some $g \in G$, and so $\phi(x_2) = \phi(gx_1) = \phi(x_1)$ by the invariance of ϕ. So for an invariant ϕ and a maximal invariant T, $T(x_1) = T(x_2)$ implies $\phi(x_1) = \phi(x_2)$, and therefore ϕ is a function of T. □

Example 5. Let $X = (X_1, ..., X_n)$, and consider the group of transformations $X_i{}' = aX_i$, $\forall i$ and $a \neq 0$. If X is the set of points in $\mathbb{R}^n$ for which none of the coordinates is zero, then the ratios $(\frac{X_1}{X_n}, ..., \frac{X_{n-1}}{X_n})$ form a maximal invariant.

Example 6. Let $X = (X_1, ..., X_n)$, and for any permutation π, consider $X_i{}' = X_{\pi(i)}$, $\forall i$. Then the set of order statistics $(X_{(1)}, ..., X_{(n)})$ is maximal invariant.

Example 7. Consider the group of transformations $(X_1, \ldots, X_n) \to (f(X_1), \ldots, f(X_n))$. Here, f is any strictly increasing and continuous function. If $\mathcal{X}$ contains the points in $\mathbb{R}^n$ with distinct coordinates, then the ranks $(r_1, r_2, ..., r_n)$ of $(x_1, x_2, .., x_n)$ are maximal invariants. Here the rank $r_i = 1$ means X_i is smallest, and $r_i = j$ means X_i is the j-th smallest.

17.5 Finding UMPI Tests

Now that we have successfully found several maximal invariants in some examples, let us move on and find the Uniformly Most Powerful Invariant (UMPI) test in some problems.

Example 8 (Find the UMPI). Let $X_1, X_2, ..., X_n$ be i.i.d and distributed as $f_i(x - \theta)$, $i \in \{0, 1\}$ with θ unknown. For testing H_0: $i = 0$ versus H_1: i = 1, if we make the transform that $X_i{}' = X_i + a$,($\forall a \in \mathbb{R}$), then we leave the problem invariant. If $Y_i = X_i - X_n$, we know

$(Y_1, Y_2, ..., Y_{n-1})$ is its maximal invariant from the previous example. Moreover, the joint density of Y is

$$\int f_i(y_1 + t, ..., y_{n-1} + t, t)dt$$

(we derived a similar expression in our discussion of Pitman estimators). Notice that Y has no dependence on the nuisance parameter θ. Indeed, by restricting our attention to Y, we have transformed our composite vs. composite testing problem into a simple vs. simple testing problem. Thus, to find a UMPI test we need only find a MP test based on Y via Neyman-Pearson. Neyman-Pearson implies that a MP test rejects when

$$\frac{\int f_1(y_1 + t, ..., y_{n-1} + t, t)dt}{\int f_0(y_1 + t, ..., y_{n-1} + t, t)dt} \tag{17.1}$$

is large. Since

$$(17.1) \iff \frac{\int f_1(x_1 - x_n + t, ..., x_{n-1} - x_n + t, t)dt}{\int f_0(x_1 - x_n + t, ..., x_{n-1} - x_n + t, t)dt},$$

letting $u = t - x_n$, we have

$$\iff \frac{\int f_1(x_1 + u, ..., x_n + u)du}{\int f_0(x_1 + u, ..., x_n + u)du} \tag{17.2}$$

Thus the UMPI test rejects when (17.2) is large.

In many cases, we can find a UMPI test by first compressing data into sufficient statistics and then applying invariance considerations to our compressed data. This strategy succeeds in the example below, but there are cases (that we will not discuss in this course) in which the strategy fails. See TSH Thm. 6.5.3 for sufficient conditions for the strategy to succeed.

Example 9. Suppose $X_1, ..., X_n$ are i.i.d distributed as $\mathcal{N}(\mu, \sigma^2)$, where μ and σ are both unknown. Our goal is to find the UMPI test of $H_0 : \sigma = \sigma_0$ versus $H_1 : \sigma > \sigma_0$. This problem remains invariant under transformations of the form $X_i{}' = X_i + a$.

Before searching for maximal invariants, let us first reduce the data to the sufficient statistics $(\bar{X}, \sum_i (X_i - \bar{X})^2) = (S_1, S_2)$. The original transformations on X induce a group of transformations $(S_1{}', S_2{}') = (S_1 + a, S_2)$ on the sufficient statistics. We will search for tests that are functions of (S_1, S_2) and invariant with respect to the induced group of transformations.

A maximal invariant with respect to the induced group in sufficient statistic space is S_2. Therefore the UMPI test can only depend on S_2. Note, that S_2 has the distribution of $\sigma^2\chi^2_{n-1} = \text{Gamma}(\frac{n-1}{2}, 2\sigma^2)$, where $\frac{n-1}{2}$ is the shape parameter and $2\sigma^2$ is the scale parameter. Since this one-parameter exponential family has monotone likelihood ratio, it follows from Theorem 3.4.1 that the UMP test for this family rejects when $\frac{S_2}{\sigma_0^2} > C_{n-1,1-\alpha}$, where $C_{n-1,1-\alpha}$ is the $1 - \alpha$ quantile of χ^2_{n-1}. This test is UMPI for the induced group on (S_1, S_2), and it so happens that this test is also UMPI for the original transformation group on $\mathcal{X}$. This is typical for the examples that we will deal with in this class but not guaranteed.

Note that this UMPI test is the same as the UMPU test that we previously derived for this example. This is not a coincidence. This happens whenever the UMPU is unique and a UMP *almost* invariant test exists. We will not discuss almost invariance in this class, but you can learn more about it in TSH chapter 6.5.