Admissible tests and maximization of power subject to size (WIP)

Gustavo Pereira

November 12, 2019

In lecture notes 9-10, Proposition 1 characterizes admissible tests in terms of the solution of a problem of maximizing power subject to a size constraint. I reproduce the statement of that proposition below.

Proposition 1. *Suppose that for any set* $A \subseteq \mathbf{X}$

$$\int_A f(x, \theta_0)dx > 0 \implies \int_A f(x, \theta_1)dx > 0.$$

A randomized test ϕ *is admissible if and only if there exists* $\alpha \in [0, 1]$ *such that* ϕ *maximizes power subject to having size at most* α*; that is*

$$\phi \in \arg\max_{\phi} (1 - R(\phi, \theta_1)) \tag{1}$$

s.t.

$$R(\phi, \theta_0) \leq \alpha \tag{2}$$

That proposition is actually really nice. In standard statistics courses, we sometimes take this maximization problem as the starting point, as if it's somehow self-evident that we should seek tests that *maximize power subject to size*. With the decision theoretic framework we built in the first few lectures, we can actually understand why tests that solve this maximization problem are of any interest to us. The reason is that this procedure yields tests that aren't dominated.

Another way of framing the proposition is the following. For a fixed $\alpha \in [0, 1]$, let $\Phi^*(\alpha)$ denote the set of all tests ϕ^* that maximize (1) subject to (2).

The correspondence $\Phi^*(\alpha)$ depends on a single parameter $\alpha \in [0, 1]$. What proposition 1 says is that, as we vary α, we cover all possible admissible tests. In other words,

$$\mathcal{A} = \bigcup_{\alpha \in [0,1]} \Phi^*(\alpha)$$

is *exactly* the set of all admissible tests.

1 Understanding Proposition 1

I modify the proposition's exposition to make it a bit more digestible.

First, let's define the following.

Definition 1. Let $\{f_\theta(x)\}_{\theta \in \Theta}$ be a statistical model. We say that

$$f_{\theta_0} \ll f_{\theta_1}$$

(in plain English: f_{θ_0} is *dominated* by f_{θ_1}) if, for every measurable set A,

$$\mathbf{P}_{\theta_1}(A) = 0 \implies \mathbf{P}_{\theta_0}(A) = 0$$

$\triangleleft$

Important remark. The relation $\ll$ has *nothing* to do with risk, loss, etc. It also has nothing to do with stochastic dominance.

Let's translate the definition above. What it means for f_{θ_0} to be dominated by f_{θ_1} is that, if the statistical model under θ_1 assigns zero probability to a set A – that is, there is a zero probability that we observe data in the set A under the alternative – then the probability that we observe data in the set A under the null must also be zero.

In other words, if that condition didn't hold, there would be a set of data realizations that are "impossible" under the alternative, but "possible" under the null.

Note that we can rewrite the definition in terms of integrals, since

$$\mathbf{P}_\theta(A) = \int_A f_\theta(x)dx$$

Hence, $f_{\theta_0} \ll f_{\theta_1}$ if and only if

$$\int_A f_{\theta_1}(x)dx = 0 \implies \int_A f_{\theta_0}(x)dx = 0$$

Or yet (by contraposition): $f_{\theta_0} \ll f_{\theta_1}$ iff

$$\int_A f_{\theta_0}(x)dx > 0 \implies \int_A f_{\theta_1}(x)dx > 0$$

All of these are restatements of the assumption that we can't observe under the null things that can't be observed under the alternative.

That assumption gives us an important result, that I state as a lemma.

Lemma 1. *Let $\{f_\theta\}_{\theta\in\Theta}$ be a statistical model with $\Theta = \{\theta_0, \theta_1\}$. Suppose $f_{\theta_0} \ll f_{\theta_1}$.*

Then any test ϕ achieving full power must have size equals one. Mathematically:

$$\mathbf{E}_{\theta_1}[\phi(X)] = 1 \implies \mathbf{E}_{\theta_0}[\phi(X)] = 1$$

Moreover, tests achieving zero size must have trivial power:

$$\mathbf{E}_{\theta_0}[\phi(X)] = 0 \implies \mathbf{E}_{\theta_1}[\phi(X)] = 0$$

Proof. Since $\phi(X) \leq 1$, full power – ie $\mathbf{E}_{\theta_1}\phi(X) = 1$ – implies that the set $A = \{x \in \mathcal{X} : \phi(x) < 1\}$ has zero probability under θ_1. Thus

$$\int_{\phi(x)<1} f_{\theta_1}(x)dx = 0$$

Since f_{θ_0} is dominated by f_{θ_1},

$$\mathbf{E}_{\theta_0}\phi(X) = \int_{\{\phi(x)=1\}} \phi(x) f_{\theta_0}(x)dx + \cancelto{0}{\int_{\{\phi(x)<1\}} \phi(x) f_{\theta_0}(x)dx} = 1$$

□

I'll now restate one directions of Proposition 1, for the particular case when $0 < \alpha < 1$.

Proposition 2. *Let $\{f_\theta\}_{\theta\in\Theta}$ be a statistical model with $\Theta = \{\theta_0, \theta_1\}$.*

Suppose $f_{\theta_0} \ll f_{\theta_1}$. Then any (randomized) test ϕ^* that solves the problem below is admissible in a decision problem with 0-1 loss, when $\alpha \in (0, 1)$.

$$\begin{aligned} \max_{\phi} \quad & E_{\theta_1}\phi(X) \\ s.t. \quad & E_{\theta_0}\phi(X) \leq \alpha \end{aligned} \tag{P}$$

Proof. Let's proceed by contradiction. Assume that ϕ^* solves the maximization problem but is not admissible. Then there exists some test ϕ that dominates ϕ^*, that is:

$$R(\phi, \theta_0) = \mathbf{E}_{\theta_0}[\phi(X)] \leq \mathbf{E}_{\theta_0}[\phi^*(X)] = R(\phi^*, \theta_0) \tag{3}$$

$$R(\phi, \theta_1) = 1 - \mathbf{E}_{\theta_1}[\phi(X)] \leq 1 - \mathbf{E}_{\theta_1}[\phi^*(X)] = R(\phi^*, \theta_1) \tag{4}$$

where one of the equalities holds strictly. We consider the two cases below.

1. Suppose 3 holds strictly, and 4 holds weakly. Since ϕ^* solves the maximization problem (P), the size constraint must be satisfied so

   $$\mathbf{E}_{\theta_0}[\phi(X)] < \mathbf{E}_{\theta_0}[\phi^*(X)] \leq \alpha < 1$$

   This first thing to note, which will only be used later on, is that since $\mathbf{E}_{\theta_0}[\phi(X)] < 1$, it must be that $\mathbf{E}_{\theta_1}[\phi(X)] < 1$ by the first part of Lemma 1.

   The idea of the proof is to construct yet another test that will use up the slack that ϕ has in the size constraint, $\mathbf{E}_{\theta_0}[\phi(X)] < \alpha$, to achieve higher power.

   We can do that by mixing ϕ with the test that rejects the null for any realization,

   $$\phi_R(X) \equiv 1$$

   and by picking the right mix, we will increase power relative to ϕ, while still controlling for size. By 4, we will also improve relative to ϕ^*, a contradiction.

   Now how do we find that combination? Consider, for arbitrary $\lambda \in [0, 1]$, the test

   $$\phi_\lambda(X) \equiv \lambda\phi^R(X) + (1 - \lambda)\phi(X)$$

   (Make sure you understand why we combine ϕ with ϕ^R, in particular why we don't combine ϕ^R with ϕ^*.) Its rate of type I error is given by

   $$\mathbf{E}_{\theta_0}[\phi_\lambda(X)] = \lambda + (1 - \lambda)E_{\theta_0}[\phi(X)]$$

   We pick $\bar{\lambda}$ that gives size exactly equal to α by setting

   $$\bar{\lambda} = \frac{\alpha - \mathbf{E}_{\theta_0}[\phi(X)]}{1 - \mathbf{E}_{\theta_0}[\phi(X)]}$$

   Since $0 \leq E_{\theta_0}[\phi(X)] < \alpha < 1$, we have $\bar{\lambda} \in (0, 1)$.

By construction, $\phi_{\bar{\lambda}}$ has a rate of type I error of exactly α. Its power on the other hand is given by

$$\mathbf{E}_{\theta_1}[\phi_{\bar{\lambda}}(X)] = \bar{\lambda} \cdot 1 + (1 - \bar{\lambda}) \cdot \mathbf{E}_{\theta_1}[\phi(X)]$$

Because $\mathbf{E}_{\theta_1}[\phi(X)] < 1$ and $\bar{\lambda} \in (0, 1)$, the above expression implies

$$\mathbf{E}_{\theta_1}[\phi_{\bar{\lambda}}(X)] > \mathbf{E}_{\theta_1}[\phi(X)] \geq \mathbf{E}_{\theta_1}[\phi^*(X)]$$

Where the last inequality comes from the assumption (4). That is a contradiction with the fact that ϕ^* is solves problem (P).

2. Suppose now that (4) holds strictly, while (3) holds weakly. Then (3) implies ϕ satisfies the size constraint, and

$$\mathbf{E}_{\theta_1}[\phi(X)] > \mathbf{E}_{\theta_1}[\phi^*(X)]$$

implies that ϕ achieves strictly higher power than ϕ^*, in direct contradiction with the fact that ϕ^* solves problem (P).

$\square$