A layman’s note on a class of frequentist hypothesis testing problems

Michele Pavon

Dipartimento di Matematica,
Università di Padova,
via Trieste 63, 35121 Padova, Italy

pavon@math.unipd.it

(Dated: February 5, 2015)

Abstract

It is observed that for testing between simple hypotheses where the cost of Type I and Type II errors can be quantified, it is better to let the optimization choose the test size.

Keywords: hypothesis testing; Neyman-Pearson; optimization; test size
I. HYPOTHESIS TESTING

Let \((X, \mathcal{F}, \mu)\) be a \(\sigma\)-finite measure space and let \(\mathcal{P}\) be the family of probability measures \(\mathbb{P}\) on \((X, \mathcal{F})\) which are absolutely continuous with respect to \(\mu\) so that, for \(A \in \mathcal{F}\),

\[\mathbb{P}(A) = \int_A p(x) d\mu.\]

Here \(p = d\mathbb{P}/d\mu\) is the density (Radon-Nikodym derivative) of \(\mathbb{P}\) with respect to \(\mu\). We are mostly interested in two cases: The first is when \(X\) is a Euclidean space \(\mathbb{R}^N\) equipped with the Borel \(\sigma\)-field and \(\mu\) is Lebesgue measure. The second is when \(X = \mathbb{Z}^N\) or \(X = \mathbb{N}^N\) and \(\mu\) is counting measure on all subsets of \(X\). This allows us treat probability densities and discrete probability distributions simultaneously.

Let \(\mathbb{P}_0, \mathbb{P}_1 \in \mathcal{P}\) and let \(p_0\) and \(p_1\) be the corresponding densities with respect to \(\mu\). Let \((X_1, \ldots, X_N)\) be the available sample taking values in \(X\). We seek a test \(\varphi : X \to \{0, 1\}\) such that, if \((x_1, \ldots, x_N)\) are the observed values, \(\varphi(x_1, \ldots, x_N) = 0\) if we accept \(H_0 = \{\mathbb{P}_0\}\) and \(\varphi(x_1, \ldots, x_N) = 1\) if we accept \(H_1 = \{\mathbb{P}_1\}\). Let \(C\) be the critical region, namely the subset of observations \(x = (x_1, \ldots, x_N)\) such that \(\varphi(x_1, \ldots, x_N) = 1\), namely where we reject the null hypothesis, cf. e.g. [1, Chapter 8].

II. A CLASS OF INFERENCE PROBLEMS

Consider a simple hypothesis testing problem where we can quantify the cost of each error. Namely, if we reject \(H_0\) when it is true we incur the cost \(c_0 > 0\) and if we reject \(H_1\) when it is true we incur the cost \(c_1 > 0\). This is the case in many applications such as when, on the basis of a sample, we need to decide whether to halt the production of an item which should meet certain required standards. Both producing a whole stock not meeting the requirements or halting the production process when the requirements are met causes certain quantifiable costs. A type I error occurs with probability \(\alpha = \mathbb{P}_0(C)\) while a type II error occurs with probability \(\beta = \mathbb{P}_1(C^c)\). It is then natural to try to minimise the cost

\[J(C) = c_0\mathbb{P}_0(C) + c_1\mathbb{P}_1(C^c).\]

This is a simple unconstrained optimisation problem which can be formalized as follows.
**Problem 1** Find a measurable set \( C \subset X \) such that the following cost function

\[
J(C) = c_0 P_0(C) + c_1 P_1(C^c) = \int_C [c_0 p_0(x) - c_1 p_1(x)] \, d\mu + c_1
\]

is minimised or, equivalently abusing notation, minimize

\[
J(1_C) = \int_X 1_C [c_0 p_0(x) - c_1 p_1(x)] \, d\mu
\]

where \( 1_C \) is the indicator function of the set \( C \).

Let us introduce the set

\[
Q = \{ f \in L^\infty(X, \mathcal{F}, \mu) | f : X \to [0, 1] \},
\]

and consider the following “relaxed” version of Problem 1:

**Problem 2**

Minimize \( f \in Q J(f) \),

where

\[
J(f) = \int_X f(x) [c_0 p_0(x) - c_1 p_1(x)] \, d\mu.
\]

Observe that the cost function is linear in \( f \) and \( Q \) is convex. Thus, this is a convex optimization problem. We recall a few basic facts from convex optimization. Let \( K \) be a convex subset of the vector space \( V \), let \( F : K \to \mathbb{R} \) be convex and let \( x_0 \in K \). Then, the one-sided directional derivative or hemidifferential of \( F \) at \( x_0 \) in direction \( x - x_0 \)

\[
F_+''(x_0; x - x_0) := \lim_{\epsilon \searrow 0} \frac{F(x_0 + \epsilon(x - x_0)) - F(x_0)}{\epsilon}
\]

exists for every \( x \in K \) (this is a consequence of the monotonicity of the difference quotients). We record next the characterisation of optimality for convex problems, see e.g. [2, p.66].

**Theorem 3** Let \( K \) be a convex subset of the vector space \( V \) and let \( F : K \to \mathbb{R} \) be convex. Then, \( x_0 \in K \) is a minimum point for \( F \) over \( K \) if and only if it holds

\[
F_+''(x_0; x - x_0) \geq 0, \quad \forall x \in K. \tag{1}
\]

We can then apply this result to Problem 2.
Proposition 4 The minimum in Problem 2 is attained for
\[ C^* = \{ x \in X | c_0 p_0(x) \leq c_1 p_1(x) \}. \] (2)

Proof. We apply Theorem 3 to Problem 2 and get that a necessary and sufficient condition for \( f^* \in Q \) to be a minimum point of \( J(f) \) over \( Q \) is
\[ J'(f^*; f - f^*) = \int_X [f(x) - f^*(x)] [c_0 p_0(x) - c_1 p_1(x)] d\mu \geq 0, \quad \forall f \in Q. \] (3)

Observe now that \( f^* = 1_{C^*} \) satisfies (3). Indeed
\[ \int_X [f(x) - 1_{C^*}(x)] [c_0 p_0(x) - c_1 p_1(x)] d\mu = \int_{C^*} [f(x) - 1_{C^*}(x)] [c_0 p_0(x) - c_1 p_1(x)] d\mu + \int_{(C^*)^c} f(x) [c_0 p_0(x) - c_1 p_1(x)] d\mu \geq 0, \]
since both integrals in the last line are nonnegative. Indeed, \( f(x) - 1 \leq 0 \) and, on \( C^* \), \( c_0 p_0(x) - c_1 p_1(x) \leq 0 \) imply that the integrand in the first integral is nonnegative. The integrand of the second integral is the product of two nonnegative functions and is therefore also nonnegative. Finally, since \( f^* = 1_{C^*} \) is an indicator function, it also solves Problem 1. \( \square \)

Remark 5 We can rewrite the optimal critical region in the familiar form
\[ C^* = \left\{ x \in X | \Lambda(x) \geq \frac{c_0}{c_1} \right\}, \quad \Lambda(x) = \frac{p_1(x)}{p_0(x)}. \] (4)

Thus, the ratio of the two costs \( c_0/c_1 \) plays the role of the multiplier associated to the size constraint in the usual Neyman-Pearson approach. The size of the test and its power, are simply
\[ \alpha^* = \mathbb{P}_0 \left( \Lambda(x) \geq \frac{c_0}{c_1} \right), \quad \beta^* = \mathbb{P}_1 \left( \Lambda(x) \geq \frac{c_0}{c_1} \right). \] (5)

III. EXAMPLE

We illustrate this approach in the simple case of testing the mean of a normal distribution with known variance. Let \( \mu \) be Lebesgue measure on \( \mathbb{R} \), \( p_0 = \mathcal{N}(0, 36) \) and \( p_1 = \mathcal{N}(1.2, 36) \). Suppose \((x_1, \ldots, x_N)\) are the observed values from a random sample and
let \( \bar{x}_N = \frac{1}{N} \sum_{i=1}^{N} x_i \) be the sample mean. Let us fix \( \alpha = 0.05 \) and let \( N = 100 \). Then the optimal Neyman-Pearson test has critical region \( C_{NP} = \{ \bar{x}_{100} \geq 0.987 \} \). The corresponding error of the second type is \( \beta = 0.36 \). Since only the ratio \( (c_0/c_1) \) matters in the minimisation of Problem 1, we take from here on \( c_1 = 1 \). Thus applying the Neyman-Pearson approach with tests of size 0.05, we incur the cost

\[
J(C_{NP}) = c_0(0.05) + 0.36.
\]

Next, we compare \( J(C_{NP}) \) with \( J(C^*) = c_0\alpha^* + \beta^* \), with \( C^* \) given by (2) and \( \alpha^* \) and \( \beta^* \) given by (5), for different values of \( c_0 \) and \( c_1 = 1 \). We get the results of Table I.

| \( c_0 \)   | \( J(C_{NP}) \)  | \( J(C^*) \)          |
|------------|-----------------|------------------------|
| 1          | 0.05 + 0.36 = 0.41 | 0.1587 + 0.1587 = 0.3174 |
| \( e \)    | \( 2.718 \times 0.05 + 0.36 = 0.495914 \) | \( 2.718 \times 0.06681 + 0.30854 = 0.490129 \) |
| \( e^2 \)  | \( 7.387 \times 0.05 + 0.36 = 0.7293762 \) | \( 7.387 \times 0.02275 + 0.5 = 0.668066171 \) |
| \( e^3 \)  | \( 20.07929 \times 0.05 + 0.36 = 1.3639645 \) | \( 20.07929 \times 0.00621 + 0.69 = 0.81469239 \) |

We see that in all cases, as expected since \( C^* \) gives the minimum cost, fixing \( \alpha \) a priori without considering the costs of type I and II errors, leads to a higher cost. The costs are closer when \( \alpha^* \) is close to 0.05. Indeed, if \( \alpha^* \) happens to be 0.05, given the form (11) of \( C^* \), we have \( C^* = C_{NP} \).

In conclusion, when the cost of the two errors is known, it appears wiser to let the optimization determine the size of the test through (5).

[1] J. C. Kiefer, Introduction to Statistical Inference, Springer-Verlag, 1987.
[2] P. Kosmol, Optimierung und Approximation, De Gruyter Lehrbuch, Berlin, 1991.
