Towards a general theory of optimal testing

Luis R. G. Pericchi and Carlos A. B. Pereira

Citation: AIP Conf. Proc. **1490**, 31 (2012); doi: 10.1063/1.4759586
View online: http://dx.doi.org/10.1063/1.4759586
View Table of Contents: http://proceedings.aip.org/dbt/dbt.jsp?KEY=APCPCS&Volume=1490&Issue=1
Published by the American Institute of Physics.

Towards a General Theory of Optimal Testing

Luis R. G. Pericchi* and Carlos A. B. Pereira†

*Department of Mathematics and Center for Biostatistics and Bioinformatics, University of Puerto Rico, Rio Piedras, San Juan, PR
†Instituto de Matemática e Estatística, Univ. Sao Paulo, Brasil

Abstract. In Pericchi and Pereira [1] it is argued against the traditional way on which testing is based on fixed significance level, either using p-values (with fixed levels of evidence, like the 5% rule) or α values. We instead, follow an approach put forward by [2], on which an optimal test is chosen by minimizing type I and type II errors.

Morris DeGroot in his authoritative book [2], Probability and Statistics 2nd Edition, stated that it is more reasonable to minimize a weighted sum of Type I and Type II error than to specify a value of type I error and then minimize Type II error. He showed it beyond reasonable doubt, but only in the very restrictive scenario of simple VS simple hypothesis, and it is not clear how to generalize it. We propose here a very natural generalization for composite hypothesis, by using general weight functions in the parameter space. This was also the position taken by [3, 4, 5]. We show, in a parallel manner to DeGroot's proof and Pereira's discussion, that the optimal test statistics are Bayes Factors, when the weighting functions are priors with mass on the whole parameter space. On the other hand when the weight functions are point masses in specific parameter values of practical significance, then a procedure is designed for which the sum of Type I error and Type II error in the specified points of practical significance is minimized. This can be seen as bridge between Bayesian Statistics and a new version of Hypothesis testing, more in line with statistical consistency and scientific insight.

Keywords: Hypothesis testing; optimal test; type I and II errors
PACS: 02.50.-r

MINIMIZING THE SUM OF WEIGHTED ERRORS

Morris DeGroot [2], argued that instead of fixing Type I error and minimize Type II, a better hypothesis testing paradigm is to choose the test as to minimize a weighted sum of the errors, namely

$$Min_{\delta} \quad [a \cdot \alpha_{\theta_0}(\delta) + b \cdot \beta_{\theta_1}(\delta)], \tag{1}$$

where δ denotes the test: $\delta(\mathbf{x}) = 1_{\mathtt{R}}(\mathbf{x})$, where $\mathtt{R}$ is the Rejection Region of H_0, $I_S(y)$ is the indicator function, equal to 1 if and only if $y \in S$. Notice the apparently slight difference of (1) with the traditional approach of Significance Testing:

$$\text{Restricting to those tests } \delta \text{ on which Type I error: } \alpha(\delta) \leq \alpha_0, \ \text{Min}_{\delta} \ \beta_{\theta_1}(\delta). \tag{2}$$

However, the difference is far reaching as we will see in the sequel. In the following the superiority of DeGroot's approach is distinctly apparent.
DeGroot in his book, only presented the approach for Simple VS Simple hypothesis. In the next section we generalize for a general setting.

XI Brazilian Meeting on Bayesian Statistics
AIP Conf. Proc. 1490, 31-35 (2012); doi: 10.1063/1.4759586

A General Formal Approach

Suppose we are testing the following two general hypotheses:

$$H_0 : \theta \in \Theta_0 \text{ VS } H_1 : \theta \in \Theta_1 \tag{3}$$

We define, in the Neyman-Pearson tradition, Type I and Type II error, of the test δ at the parameter point θ as,

$$\alpha_\theta(\delta) = Pr(\text{Rejecting } H_0|\theta \in \Theta_0), \tag{4}$$

$$\beta_\theta(\delta) = Pr(\text{Accepting } H_0|\theta \in \Theta_1). \tag{5}$$

Definition: The weighted Type I and Type II errors are defined respectively as:

$$\alpha(\delta) = \int_{\Theta_0} \alpha_\theta(\delta)\pi_0(\theta)d\theta, \tag{6}$$

and

$$\beta(\delta) = \int_{\Theta_1} \beta_\theta(\delta)\pi_1(\theta)d\theta, \tag{7}$$

where $\pi_j(\theta) \geq 0$ are such that $\int_{\Theta_j} \pi_j(\theta) = 1, j = 0, 1$.

How to understand the weight (prior) measures?

We pause here to discuss interpretations, since there are various possible interpretations of the weight measures $\pi_j(\theta), j = 0, 1$.

1. Prior Measures: The most obvious is the assumed prior density of the parameter values conditional on each hypothesis, which is the natural interpretation under a Bayesian framework. Notice that not necessarily this interpretation leads to a subjective approach. If a general method for generating conventional priors like the Intrinsic prior method, then this can be considered a non-subjective approach. But there is room for other interpretations.
2. Point Markers of Statistical Importance: These are point masses signalling specific points for which the errors ought to be controlled by design. For example if for a particular value of $\theta_1 \in \Theta_1$, where there is "practical significance", for example a novel medical treatment improvement of 20%, say, then the weight function may be set as a point mass on 20% of improvement. (This typically would work for monotone likelihood ratio families).

The Criterion

Define the weighted likelihoods, for the Data = $\mathbf{y}$ under each hypothesis as

$$\varpi_0(\mathbf{y}) = \int_{\Theta_0} f(\mathbf{y}|\theta)\pi_0(\theta)d\theta, \tag{8}$$

and

$$\varpi_1(\mathbf{y}) = \int_{\Theta_1} f(\mathbf{y}|\theta)\pi_1(\theta)d\theta. \tag{9}$$

Lemma 1: It is desired to find a test function δ that minimizes, for specified $a >$ 0, and ,$b > 0$:

$$\text{SERRORS}(\delta) = a\cdot\alpha(\delta) + b\cdot\beta(\delta). \tag{10}$$

The test δ^* is defined as: accept H_0, if

$$a\cdot\varpi_0(\mathbf{y}) > b\cdot\varpi_1(\mathbf{y}), \tag{11}$$

and H_1 is accepted if

$$a\cdot\varpi_0(\mathbf{y}) < b\cdot\varpi_1(\mathbf{y}), \tag{12}$$

and accept any if $a\cdot\varpi_0(\mathbf{y}) = b\cdot\varpi_1(\mathbf{y})$. Then for any other test function δ:

$$\text{SERRORS}(\delta^*) = a\cdot\alpha(\delta^*) + b\cdot\beta(\delta^*) \leq \text{ SERRORS}(\delta). \tag{13}$$

`Proof`: Denote by R, the rejection region of the test δ, that is those data points on which H_0 is rejected. Then, under the mild assumptions of Fubbini's Theorem that allows interchanging the order of the integrals, for any test function δ,

$$a\alpha(\delta) + b\beta(\delta) = a\int_{\Theta_0}[\int_{\mathrm{R}} f(\mathbf{y}|\theta)d\mathbf{y}]\pi_0(\theta)d\theta + b\int_{\Theta_1}[\int_{\mathrm{R}^C} f(\mathbf{y}|\theta)d\mathbf{y}]\pi_1(\theta)d\theta =$$

$$a\int_{\Theta_0}\int_{\mathrm{R}} f(\mathbf{y}|\theta)\pi_0(\theta)d\mathbf{y}d\theta + b\int_{\Theta_1}\int_{\mathrm{R}^C} f(\mathbf{y}|\theta)\pi_1(\theta)d\mathbf{y}d\theta =$$

$$a\int_{\Theta_0}\int_{\mathrm{R}} f(\mathbf{y}|\theta)\pi_0(\theta)d\mathbf{y}d\theta + b[1 - \int_{\Theta_1}\int_{\mathrm{R}} f(\mathbf{y}|\theta)\pi_1(\theta)d\mathbf{y}d\theta] =$$

$$b + \int_{\mathrm{R}}[a\int_{\Theta_0} f(\mathbf{y}|\theta)\pi_0(\theta)d\theta - b\int_{\Theta_1} f(\mathbf{y}|\theta)\pi_1(\theta)d\theta]d\mathbf{y} =$$

$$b + \int_{\mathrm{R}}[a\varpi_0(\mathbf{y}) - b\varpi_1(\mathbf{y})]d\mathbf{y}.$$

The results follows from application of the definition of δ^* in expressions 9 and 10, since every point on which $a\cdot\varpi_0(\mathbf{y}) - b\cdot\varpi_1(\mathbf{y}) < 0$ is in R, but no point on which $a\cdot\varpi_0(\mathbf{y}) - b\cdot\varpi_1(\mathbf{y}) > 0$, is included. Therefore δ^*, minimizes the last term in the sum and the first does not depend on the test. So the result has been established.

Regarding the assessment of the constants a and b, notice that it suffices to specify its ratio $a/b = \mathrm{r}$. This can be made as: i) By finding the *implicit a and b* of a carefully designed experiment, ii) Conventional table of ratio of evidences, like in Jeffreys Scale-Table of evidences, or iii) by a ratio of prior probabilities of H_0 times the loss incurred by false rejection of H_0 over the prior probability of H_1 times the loss incurred by false acceptance of H_0, in symbols:

$$\mathrm{r} = \frac{a}{b} = \frac{P(H_0)\cdot L_1}{P(H_1)\cdot L_0}.$$

To see this, notice that the risk function can be written as $R(\theta,\delta) = L_1\alpha_\theta(\delta)$ if $\theta \in \Theta_0$, and $R(\theta,\delta) = L_0\beta_\theta(\delta)$ if $\theta \in \Theta_1$. Assuming that a priori the probability of the Null Hypothesis is $P(H_0)$, then the average (Bayesian) risk, taking expectations with respect to $(P(H_0),\pi_0)$ and $((1-p(H_0)),\pi_1)$, we get the averaged risk

$$r(\delta) = P(H_0)\cdot L_1\cdot\alpha(\delta) + (1-P(H_0))\cdot L_0\cdot\beta(\delta), \tag{14}$$

and it is seen that the correspondence with expression (8) is: $a \mapsto P(H_0)\cdot L_1$ and $b \mapsto (1-P(H_0))\cdot L_0$.

The Rejection Region R in (10) takes two different shapes according to the interpretations 1 and 2.

- For interpretation 1, R is defined as

$$\frac{m_0(\mathbf{y})}{m_1(\mathbf{y})} < b/a, \tag{15}$$

  that is the Null Hypothesis is rejected if the Bayes Factor of H_0 over H_1 is small enough.
- For interpretation 2, R becomes,

$$\frac{f(\mathbf{y}|\theta_0)}{f(\mathbf{y}|\theta_1)} < b/a \tag{16}$$

  the relative likelihood ratios on the null over the alternative.

For examples and further discussions, see [1].

DEGROOT'S APPROACH AND THE LIKELIHOOD PRINCIPLE.

One of the usual criticisms against significance testing is that it does not obey the Likelihood Principle, a Principle not only shared by Bayesians, but that was actually enunciated or defended by eminent non-Bayesians. Loosely speaking the Likelihood Principle establishes that if two likelihoods are proportional to each other, the information about the parameter vector θ is the same.

The good news, surprising to many we would guess, is that the violation of the Likelihood is avoided by DeGroot's method, that is if the weighted sum of type I and type II errors is minimized.

Corollary 1. Testing by Minimizing a weighted sum of errors, automatically obeys the Likelihood Principle.

Proof: From Lemma 1, the optimal test is Reject H_0 if

$$\frac{\varpi_0(\mathbf{y})}{\varpi_1(\mathbf{y})} = \frac{\int_{\Theta_0} f(\mathbf{y}|\theta)\pi_0(\theta)d\theta}{\int_{\Theta_1} f(\mathbf{y}|\theta)\pi_1(\theta)d\theta} < b/a,$$

and in the ratio of the left hand side, the constant in the likelihood cancel out.

CONCLUSIONS

DeGroot's approach, and our implementation here in a general setting highlight the following:

1. A way out of the main *disagreement*(that of divergent measures of statistical significance of tests) among statisticians is to implement DeGroot's framework: This is no less than a powerful template on which the ideas of risk decision theory, types of error, tail probabilities, relative loss, and likelihood ratios are put together and make them coherent.
2. In order to generalize the framework, we have shown that by considering the averaged Type I and Type II error, the theory and the main result, (the optimality of the ratio of the evidences, i.e. of the Bayes Factor) is completely general.
3. The implementation of this framework, still depends on: i) the ratio of b/a, the relative weights of the two types of errors, and on the priors densities on Θ_0 and Θ_1. But now there are several methods on how to assess those (like Intrinsic Priors [6] and [5]).
4. This general theory, that yields at once: compliance with the Likelihood Principle, solution of the disagreement, reconciliation between statistical significance and practical significance, is in contrast with the traditional theory of fixed significance level, on which the optimal solutions are the exception rather than the norm, and the use of Likelihood Ratio Test which lacks a clear justification.

For further results and examples the reader is referred to [1].

REFERENCES

1. L. R. Pericchi, C. A. B. Pereira, *Testing Hypothesis by Minimizing Sum of Errors Type I and Type II: Changing the paradigm of fixed significance levels*. Tech. Report Center for Biostatistics and Bioinformatics, University of Puerto Rico, College of Natural Sciences, Rio Piedras Campus, 2012.
2. M. DeGroot, *Probability and Statistics*, 2nd Edition. Addison-Wesley, New York, 1975.
3. C. A. B. Pereira, *Teste de Hipóteses Definida em Espaços de Diferentes Dimensões: Visão Bayesiana e Interpretação Clássica*. Thesis presented to the Institute of Mathematics and Statistics. Universidade de Sâo Paulo, 1985.
4. C. A. B. Pereira, S. Wechsler, On the concept of P-Value, *Braz J Prob Stat* 7, 159–77, 1993.
5. L. Varuzza, C. A. B.Pereira, Significance test for comparing digital gene expression profiles: Partial likelihood application, *Chilean J Statistics* 1(1):91-102, *http://chjs.soche.cl/papers/volumen1_2010/ChJS-01-01-06.pdf*
6. J. O. Berger and L. R. Pericchi, The Intrinsic Bayes Factor for Model Selection and Prediction. *J*. Am. Stat. Assoc., 91, 433, 109–122 (1996).