

THE BAYES RISK CRITERION IN HYPOTHESIS TESTING*

Clayton Scott
Robert Nowak

This work is produced by The Connexions Project and licensed under the
Creative Commons Attribution License †

The design of a hypothesis test/detector often involves constructing the solution to an optimization problem. The optimality criteria used fall into two classes: Bayesian and frequent.

In the Bayesian setup, it is assumed that the *a priori* probability of each hypothesis occurring (π_i) is known. A cost C_{ij} is assigned to each possible outcome:

$$C_{ij} = Pr [\text{say } H_i \text{ when } H_j \text{ true}]$$

The optimal test/detector is the one that minimizes the Bayes risk, which is defined to be the expected cost of an experiment:

$$\bar{C} = \sum_{i,j} C_{ij} \pi_i Pr [\text{say } H_i \text{ when } H_j \text{ true}]$$

In the event that we have a binary problem, and both hypotheses are simple¹, the decision rule that minimizes the Bayes risk can be constructed explicitly. Let us assume that the data is continuous (*i.e.*, it has a density) under each hypothesis:

$$H_0 : x \sim f_0(x)$$

$$H_1 : x \sim f_1(x)$$

Let R_0 and R_1 denote the decision regions² corresponding to the optimal test. Clearly, the optimal test is specified once we specify R_0 and $R_1 = R_0'$.

The Bayes risk may be written

$$\begin{aligned} \bar{C} &= \sum_{ij=0}^1 C_{ij} \pi_i \int f_j(x) dx \\ &= \int C_{00} \pi_0 f_0(x) + C_{01} \pi_1 f_1(x) dx + \int C_{10} \pi_0 f_0(x) + C_{11} \pi_1 f_1(x) dx \end{aligned} \quad (1)$$

Recall that R_0 and R_1 **partition** the input space: they are disjoint and their union is the full input space. Thus, every possible input x belongs to precisely one of these regions. In order to minimize the Bayes

*Version 1.6: May 25, 2004 3:45 pm -0500

†<http://creativecommons.org/licenses/by/1.0>

¹"Hypothesis Testing": Section Simple Versus Composite Hypotheses <<http://cnx.org/content/m11531/latest/#svch>>

²"Hypothesis Testing": Section Tests and Decision Regions <<http://cnx.org/content/m11531/latest/#tdr>>

risk, a measurement x should belong to the decision region R_i for which the corresponding integrand in the preceding equation is smaller. Therefore, the Bayes risk is minimized by assigning x to R_0 whenever

$$\pi_0 C_{00} f_0(x) + \pi_1 C_{01} f_1(x) < \pi_0 C_{10} f_0(x) + \pi_1 C_{11} f_1(x)$$

and assigning x to R_1 whenever this inequality is reversed. The resulting rule may be expressed concisely as

$$\Lambda(x) \equiv \frac{f_1(x)}{f_0(x)} \underset{H_0}{\overset{H_1}{>}} \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})} \equiv \eta$$

Here, $\Lambda(x)$ is called the **likelihood ratio**, η is called the threshold, and the overall decision rule is called the Likelihood Ratio Test³ (LRT). The expression on the right is called a **threshold**.

Example 1

An instructor in a course in detection theory wants to determine if a particular student studied for his last test. The observed quantity is the student’s grade, which we denote by r . Failure may not indicate studiousness: conscientious students may fail the test. Define the models as

- \mathcal{M}_0 : did not study
- \mathcal{M}_1 : did study

The conditional densities of the grade are shown in Figure 1.

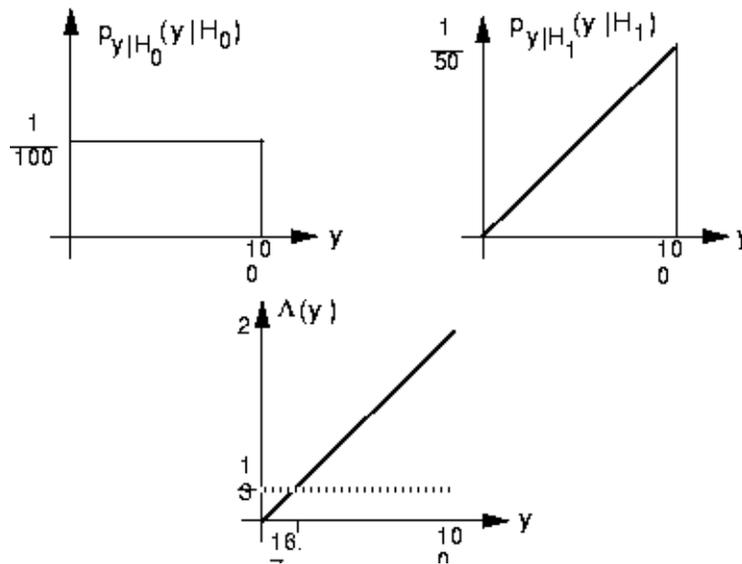


Figure 1: Conditional densities for the grade distributions assuming that a student did not study (\mathcal{M}_0) or did (\mathcal{M}_1) are shown in the top row. The lower portion depicts the likelihood ratio formed from these densities.

Based on knowledge of student behavior, the instructor assigns *a priori* probabilities of $\pi_0 = \frac{1}{4}$ and $\pi_1 = \frac{3}{4}$. The costs C_{ij} are chosen to reflect the instructor’s sensitivity to student feelings: $C_{01} = 1 = C_{10}$ (an erroneous decision either way is given the same cost) and $C_{00} = 0 = C_{11}$. The likelihood ratio is plotted in Figure 1 and the threshold value η , which is computed from the a

³<http://workshop.molecularevolution.org/resources/lrt.php>

priori probabilities and the costs to be $\frac{1}{3}$, is indicated. The calculations of this comparison can be simplified in an obvious way.

$$\frac{r}{50} \underset{\mathcal{M}_0}{\overset{\mathcal{M}_1}{\gtrless}} \frac{1}{3}$$

or

$$r \underset{\mathcal{M}_0}{\overset{\mathcal{M}_1}{\gtrless}} \frac{50}{3} = 16.7$$

The multiplication by the factor of 50 is a simple illustration of the reduction of the likelihood ratio to a sufficient statistic. Based on the assigned costs and *a priori* probabilities, the optimum decision rule says the instructor must assume that the student did not study if the student's grade is less than 16.7; if greater, the student is assumed to have studied despite receiving an abysmally low grade such as 20. Note that as the densities given by each model overlap entirely: the possibility of making the wrong interpretation **always** haunts the instructor. However, no other procedure will be better!

A special case of the minimum Bayes risk rule, the minimum probability of error rule⁴, is used extensively in practice, and is discussed at length in another module.

1 Problems

Exercise 1

Denote $\alpha = Pr$ [declare H_1 when H_0 true] and $\beta = Pr$ [declare H_1 when H_1 true]. Express the Bayes risk \bar{C} in terms of α and β , C_{ij} , and π_i . Argue that the optimal decision rule is not altered by setting $C_{00} = C_{11} = 0$.

Exercise 2

Suppose we observe x such that $\Lambda(x) = \eta$. Argue that it doesn't matter whether we assign x to R_0 or R_1 .

⁴"Minimum Probability of Error Decision Rule" <<http://cnx.org/content/m11534/latest/>>