

BEYOND TWO MODELS*

Don Johnson

This work is produced by OpenStax-CNX and licensed under the Creative Commons Attribution License 1.0[†]

Frequently, more than two viable models for data generation can be defined for a given situation. The **classification** problem is to determine which of several models best "fits" a set of measurements. For example, determining the type of airplane from its radar returns forms a classification problem. The model evaluation framework has the right structure if we can allow more than two models. We happily note that in deriving the likelihood ratio test we did not need to assume that only two possible descriptions exist. Go back and examine the expression for the maximum probability correct decision rule. If K models seem appropriate for a specific problem, the decision rule maximizing the probability of making a correct choice is

$$\forall i, i \in \{1, \dots, K\} : (\max \{ \pi_i p_{r | \mathcal{M}_i}(r) \})$$

To determine the largest of K quantities, exactly $K - 1$ numeric comparisons need be made. When we have two possible models ($K = 2$), this decision rule reduces to the computation of the likelihood ratio and its comparison to a threshold. In general, $K - 1$ likelihood ratios need to be computed and compared to a threshold. Thus the likelihood ratio test can be viewed as a specific method for determining the largest of the decision statistics $\pi_i p_{r | \mathcal{M}_i}(r)$.

Since we need only the relative ordering of the K decision statistics to make a decision, we can apply any transformation $T(\cdot)$ to them that does not affect ordering. In general, possible transformations must be positively monotonic to satisfy this condition. For example, the needless common additive components in the decision statistics can be eliminated, even if they depend on the observations. Mathematically, "common" means that the quantity does not depend on the model index i . The transformation in this case would be of the form $T(z_i) = z_i - a$, clearly a monotonic transformation. A **positive** multiplicative factor can also be "canceled"; if negative, the ordering would be reversed and that cannot be allowed. The simplest resulting expression becomes the sufficient statistic $\Upsilon_i(r)$ for the model. Expressed in terms of the sufficient statistic, the maximum probability correct or the Bayesian decision rule becomes

$$\forall i, i \in \{1, \dots, K\} : (\max \{ C_i + \Upsilon_i(r) \})$$

where C_i summarizes all additive terms that do not depend on the observation vector r . The quantity $\Upsilon_i(r)$ is termed the **sufficient statistic associated with model i** . In many cases, the functional form of the sufficient statistic varies little from one model to another and expresses the necessary operations that summarize the observations. The constants C_i are usually lumped together to yield the threshold against which we compare the sufficient statistic. For example, in the binary model situation, the decision rule becomes

$$\Upsilon_1(r) + C_1 \underset{\mathcal{M}_0}{\overset{\mathcal{M}_1}{\gtrless}} \Upsilon_0(r) + C_0$$

*Version 1.3: Aug 11, 2003 4:12 pm -0500

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

or

$$\Upsilon_1(r) - \Upsilon_0(r) \underset{\mathcal{M}_0}{\overset{\mathcal{M}_1}{\geq}} C_0 - C_1$$

Thus, the sufficient statistic for the decision rule is $\Upsilon_1(r) - \Upsilon_0(r)$ and the threshold γ is $C_0 - C_1$.

Example 1

In the Gaussian problem just discussed, the logarithm of the likelihood function is

$$\ln(p_{r|\mathcal{M}_i}(r)) = \left(- \left(\frac{L}{2} \ln(2\pi\sigma^2) \right) \right) - \frac{1}{2\sigma^2} \sum_{l=0}^{L-1} (r_l - m^{(i)})^2$$

where $m^{(i)}$ is the mean under model i . After appropriate simplification that retains the ordering, we have

$$\Upsilon_i(r) = \frac{m^{(i)}}{\sigma^2} \sum_{l=0}^{L-1} r_l$$

$$C_i = - \left(1/2 \frac{Lm^{(i)2}}{\sigma^2} \right) + c_i$$

The term c_i is a constant defined by the error criterion; for the maximum probability correct criterion, this constant is $\ln(\pi_i)$.

When employing the Neyman-Pearson test, we need to specify the various error probabilities $Pr[\text{say } \mathcal{M}_i | H_j \text{ true}]$. These specifications amount to determining the constants c_i when the sufficient statistic is used. Since $K - 1$ comparisons will be used to home in on the optimal decision, only $K - 1$ error probabilities need be specified. Typically, the quantities $Pr[\text{say } H_i | \mathcal{M}_0 \text{ true}]$, $i \in \{1, \dots, K - 1\}$ are used, particularly when the model \mathcal{M}_0 represents the situation when no signal is present (see this problem).