Non-Random Parameters*

Don Johnson

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

In those cases where a probability density for the parameters cannot be assigned, the model evaluation problem can be solved in several ways; the methods used depend on the form of the likelihood ratio and the way in which the parameter(s) enter the problem. In the Gaussian problem we have discussed so often, the threshold used in the likelihood ratio test η may be unity. In this case, examination of the resulting computations required reveals that implementing the test **does not require knowledge of the variance of the observations** (see this problem). Thus, if the common variance of the underlying Gaussian distributions is not known, this lack of knowledge has **no effect** on the optimum decision rule. This happy situation - knowledge of the value of a parameter is not required by the optimum decision rule - occurs rarely, but should be checked before using more complicated procedures.

A second fortuitous situation occurs when the sufficient statistic as well as its probability density under one of the models do **not** depend on the unknown parameter(s). Although the sufficient statistic's threshold γ expressed in terms of the likelihood ratio's threshold η depends on the unknown parameters, γ may be computed as a single value using the Neyman-Pearson criterion **if the computation of the false-alarm probability does not involve the unknown parameters**.

Example 1

Continuing the example of the previous section, let's consider the situation where the value of the mean of each observation under model $\mathcal{M}_1$ is not known. The sufficient statistic is the sum of the observations (that quantity doesn't depend on m) and the distribution of the observation vector under model $\mathcal{M}_0$ does not depend on m (allowing computation of the false-alarm probability). However, a subtlety emerges; in the derivation of the sufficient statistic, we had to divide by the value of the mean. The critical step occurs once the logarithm of the likelihood ratio is manipulated to obtain

$$m \sum_{l=0}^{L-1} r_l \underset{\mathcal{M}_0}{\overset{\mathcal{M}_1}{\gtrless}} \left(\sigma^2 \ln(\eta) + \frac{Lm^2}{2}\right)$$

Recall that only **positively** monotonic transformations can be applied; if a negatively monotonic operation is applied to this inequality (such as multiplying both sides by -1), the **inequality reverses**. If the sign of m is known, it can be taken into account explicitly and a sufficient statistic results. If, however, the sign is not known, the above expression cannot be manipulated further and the left side constitutes the sufficient statistic for this problem. The sufficient statistic then depends on the unknown parameter and we cannot develop a decision rule in this case. If the sign is known, we can proceed. Assuming the sign of m is positive, the sufficient statistic is the sum of the observations and the threshold γ is found by

$$\gamma = \sqrt{L}\sigma Q^{-1}(P_F)$$

*Version 1.4: Aug 27, 2003 1:50 pm -0500

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

Note that if the variance σ^2 instead of the mean were unknown, we could not compute the threshold. The difficulty lies not with the sufficient statistic (it doesn't depend on the variance), but with the false alarm probability as the expression indicates. Another approach is required to deal with the unknown-variance problem.

When this situation occurs - the sufficient statistic **and** the false-alarm probability can be computed without needing the parameter in question, we have established what is known as a **uniformly most powerful test** (or UMP test) (*Cramér; p.529-531*[1]), (*van Trees; p.89ff*[3]). If an UMP test does not exist, which can only be demonstrated by explicitly finding the sufficient statistic and evaluating its probability distribution, then the composite hypothesis testing problem cannot be solved without some value for the parameter being used.

This seemingly impossible situation - we need the value of the parameter that is assumed unknown - can be approached by noting that some data is available for "guessing" the value of the parameter. If a reasonable guess could be obtained, it could then be used in our model evaluation procedures developed in this chapter. **The data available for estimating unknown parameters are precisely the data used in the decision rule**. Procedures intended to yield "good" guesses of the value of a parameter are said to be **parameter estimates**. Estimation procedures are the topic of the next chapter; there we will explore a variety of estimation techniques and develop measure of estimate quality. For the moment, these issues are secondary; even if we knew the size of the estimation error, for example, the more pertinent issue is how the imprecise parameter value affects the performance probabilities. We can compute these probabilities **without** explicitly determining the estimate's error characteristics.

One parameter estimation procedure that fits nicely into the composite hypothesis testing problem is the **maximum likelihood estimate**. [1] Letting r denote the vector of observables and θ a vector of parameters, the maximum likelihood estimate of θ, θ_{ML}, is that value of θ that maximizes the conditional density $p_{r\,|\,\theta}\left(r\right)$ of the observations given the parameter values. To use θ_{ML} in our decision rule, we estimate the parameter vector **separately** for each model, use the estimated value in the conditional density of the observations, and compute the likelihood ratio. This procedure is termed the **generalized likelihood ratio test** for the unknown parameter problem in hypothesis testing (*Lehmann; p.16*[2]), (*van Trees; p.92ff*[3]).

$$\Lambda\left(r\right) = \frac{max_{\theta}\left\{\theta, p_{r\,|\,\mathcal{M}_1\theta}\left(r\right)\right\}}{max_{\theta}\left\{\theta, p_{r\,|\,\mathcal{M}_0\theta}\left(r\right)\right\}} \tag{1}$$

Note that we do **not** find that value of the parameter that (necessarily) maximizes the likelihood ratio. Rather, we estimate the parameter value most consistent with the observed data in the context of each assumed model (hypothesis) of data generation. In this way, the estimate conforms with each potential model rather than being determined by some amalgam of supposedly mutually exclusive models.

Example 2

Returning to our Gaussian example, assume that the variance σ^2 is known but that the mean under $\mathcal{M}_1$ is unknown.

$$\mathcal{M}_0 :\ r \sim \left(0, \sigma^2 I\right)$$

$$\mathcal{M}_1 :\ r \sim \left(m, \sigma^2 I\right)$$

$$m = \begin{pmatrix} m \\ \dots \\ m \end{pmatrix},\ m = ?$$

The unknown quantity occurs only in the exponent of the conditional density under $\mathcal{M}_1$; to maximize this density, we need only to maximize the exponent. Thus, we consider the derivative of the

[1] The maximum likelihood estimation procedure and its characteristics are fully described in this section.

exponent with respect to m.

$$\left(\frac{\partial\left(\left(-\frac{1}{2\sigma^2}\right)\sum_{l=0}^{L-1}\left(r_l-m\right)^2\right)}{\partial m}\Big|_{m=m_{\mathrm{ML}}=0}\right)\Rightarrow\left(\sum_{l=0}^{L-1}r_l-m_{\mathrm{ML}}=0\right)$$

The solution of this equation is the average value of the observations

$$m_{\mathrm{ML}}=\frac{1}{L}\sum_{l=0}^{L-1}r_l$$

To derive the decision rule, we substitute this estimate in the conditional density for $\mathcal{M}_1$. The critical term, the exponent of this density, is manipulated to obtain

$$\left(-\frac{1}{2\sigma^2}\right)\sum_{l=0}^{L-1}\left(r_l-\frac{1}{L}\sum_{k=0}^{L-1}r_k\right)^2=\left(-\frac{1}{2\sigma^2}\right)\left(\sum_{l=0}^{L-1}r_l^2-\frac{1}{L}\left(\sum_{l=0}^{L-1}r_l\right)^2\right)$$

Noting that the first term in this exponent is identical to the exponent of the denominator in the likelihood ratio, the generalized likelihood ratio becomes

$$\Lambda\left(r\right)=e^{\frac{1}{2L\sigma^2}\left(\sum_{l=0}^{L-1}r_l\right)^2}$$

The sufficient statistic thus becomes the square (or equivalently the magnitude) of the summed observations. Compare this result with that obtained in Example 1. There, an UMP test existed **if** we knew the sign of m and the sufficient statistic was the sum of the observations. Here, where we employed the generalized likelihood ratio test, we made no such assumptions about m; this generality accounts for the difference in sufficient statistic. Which test do you think would lead to a greater detection probability for a given false-alarm probability?

Once the generalized likelihood ratio is determined, we need to determine the threshold. If the *a priori* probabilities π_0 and π_1 are known, the evaluation of the threshold proceeds in the usual way. If they are not known, all of the conditional densities must not depend on the unknown parameters lest the performance probabilities also depend upon them. In most cases, the original model evaluation problem is posed in such a way that one of the models does not depend on the unknown parameter; a criterion on the performance probability related to that model can then be established via the Neyman-Pearson procedure. If not the case, the threshold cannot be computed and the threshold must be set experimentally: we force one of the models to be true and modify the threshold on the sufficient statistic until the desired level of performance is reached. Despite this non-mathematical approach, the overall performance of the model evaluation procedure will be optimum because of the results surrounding the Neyman-Pearson criterion.

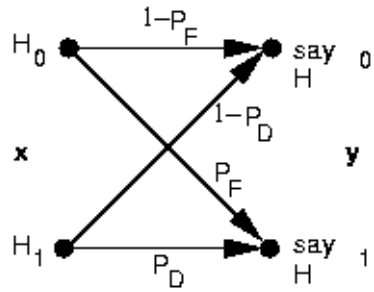

Figure 1: The two-model testing problem can be abstractly described as a communication channel where the inputs are the models and the outputs are the decisions. The transition probabilities are related to the false-alarm (P_F) and detection (P_D) probabilities.

References

[1] H. Cram[U+FFFD] *Mathematical Methods of Statistics.* Princeton University Press, Princeton, NJ, 1946.

[2] E.L. Lehmann. *Testing Statistical Hypotheses.* John Wiley and Sons, New York, second edition edition, 1986.

[3] H.L. van Trees. *Detection, Estimation, and Modulation Theory, Part I.* John Wiley and Sons, New York, 1968.