International Journal of Psychological Research, 2010. Vol. 3. No. 1.
ISSN impresa (printed) 2011-2084
ISSN electrónica (electronic) 2011-2079

Dunn, J.C., (2010). How to fit models of recognition memory data using maximum likelihood. *International Journal of Psychological Research, 3* (1), 140-149.

How to fit models of recognition memory data using maximum likelihood.

Cómo ajustar modelos de datos en experimentos sobre la memoria de reconcomiendo usando métodos de máxima verosimilitud.

John C. Dunn
University of Adelaide

ABSTRACT

The aim of this paper is to provide an introductory tutorial to how to fit different models of recognition memory using maximum likelihood estimation. It is in four main parts. The first part describes how recognition memory data is collected and analysed. The second part introduces four current models that will be fitted to the data. The third part describes in detail how a model is fit using maximum likelihood estimation. The fourth part examines how the fit of a model can be evaluated and the appropriate statistical test applied.

Key words: Recognition memory, maximum likelihood estimation, signal detection theory, mixture models, high threshold models.

RESUMEN

El propósito de este artículo es proveer un tutorial sobre cómo ajustar diferentes modelos de la memoria de reconocimiento usando estimación de máxima verosimilitud. El artículo presenta cuatro partes. Primero se describe cómo se analizan y obtienen datos en experimentos sobre la memoria de reconocimiento. En segundo lugar se presentan cuatro modelos recientes que serán ajustados a los datos. La tercera parte describe en detalle cómo se ajusta un modelo usando el procedimiento de estimación de máxima verosimilitud. Por último se examina cómo el modelo ajustado pueden ser evaluado y qué pruebas estadísticas pueden aplicarse para ello.

Palabras clave: memoria de reconocimiento, estimación de máxima verosimilitud, teoría de detección de señales, modelos mixtos, modelos de umbral alto.

Article received/Artículo recibido: December 15, 2009/Diciembre 15, 2009, Article accepted/ Artículo aceptado: March 15, 2009/Marzo 15/2009
Dirección correspondencia/Mail Address:
John C. Dunn, School of Psychology, University of Adelaide, Adelaide, SA, 5005, Australia. Email: john.c.dunn@adelaide.edu.au

INTERNATIONAL JOURNAL OF PSYCHOLOGICAL RESEARCH esta incluida en PSERINFO, CENTRO DE INFORMACION PSICOLOGICA DE COLOMBIA, OPEN JOURNAL SYSTEM, BIBLIOTECA VIRTUAL DE PSICOLOGIA (ULAPSY-BIREME), DIALNET y GOOGLE SCHOLARS. Algunos de sus articulos aparecen en SOCIAL SCIENCE RESEARCH NETWORK y está en proceso de inclusion en diversas fuentes y bases de datos internacionales.
INTERNATIONAL JOURNAL OF PSYCHOLOGICAL RESEARCH is included in PSERINFO, CENTRO DE INFORMACIÓN PSICOLÓGICA DE COLOMBIA, OPEN JOURNAL SYSTEM, BIBLIOTECA VIRTUAL DE PSICOLOGIA (ULAPSY-BIREME), DIALNET and GOOGLE SCHOLARS. Some of its articles are in SOCIAL SCIENCE RESEARCH NETWORK, and it is in the process of inclusion in a variety of sources and international databases.

International Journal of Psychological Research, 2010. Vol. 3. No. 1.
ISSN impresa (printed) 2011-2084
ISSN electrónica (electronic) 2011-2079

Dunn, J.C., (2010). How to fit models of recognition memory data using maximum likelihood. *International Journal of Psychological Research, 3* (1), 140-149.

In a typical recognition memory experiment, participants must discriminate previously studied items, called *targets*, from other items, called *lures*. In the simplest version of this experiment, participants are only required to make a yes/no decision. In a more complicated version, they may also provide confidence ratings. The advantage of the latter is that it enables different models of recognition memory to be compared. Over the last 10 to 15 years, several models of recognition have been proposed and much current research is focused on determining which may be correct. In fact, progress has been made and several models can been firmly rejected.

Recognition memory models are explicit mathematical descriptions that attempt to account for the distribution of responses to targets and lures across the available response categories. The aim of this paper is to outline how to fit such models to the data using maximum likelihood estimation (MLE). This procedure provides estimates of the model parameters which may be useful for descriptive purposes. In addition, it also provides an estimate of how well the model fits the data which can be used for model evaluation. The focus will be on a practical introduction to these techniques rather than its mathematical underpinning.

Table 1. *Observed number of responses for targets and lures across a 6-point rating scale.*

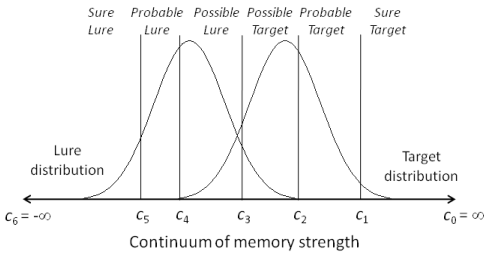
	Confidence Rating					
Item Type	*Sure Target*	*Probable Target*	*Possible Target*	*Possible Lure*	*Probable Lure*	*Sure Lure*
Number of responses:						
Lure	111	216	349	540	625	895
Target	1230	496	358	272	215	165
Cumulative number of responses:						
Lure	111	327	676	1216	1841	2736
Target	1230	1726	2084	2356	2571	2736
Cumulative proportion of responses:						
Lure	0.04	0.12	0.25	0.44	0.67	1.00
Target	0.45	0.63	0.76	0.86	0.94	1.00

The data

In a typical recognition memory study, a participant may be asked to classify a test item on a 6-point confidence scale. Each point on this scale may be given a label to indicate the appropriate level of confidence. A typical set of labels may be *sure target*, *probable target*, *possible target*, *possible lure*, *probable lure*, and *sure lure*. If participants are able to discriminate targets from lures then their responses to these items should be distributed differently across these categories. Table 1 shows a typical distribution of responses, aggregated over participants, observed in an experiment conducted recently in my laboratory. The first two rows of data show the number of responses in each category to lures and targets, respectively. As can be seen, the distribution of responses is different for the two types of item, indicating that participants are able to discriminate, albeit imperfectly, targets from lures.

A useful way of picturing recognition memory data is to construct a *receiver operating characteristic* or *ROC* curve. This is a plot of the *hit rate* against the *false alarm rate* across different decision criteria. ROC curves arise in the application of signal detection theory (for a general overview of signal detection theory and its application to psychology see Macmillan & Creelman, 2004). Memory researchers discovered early on that it is often useful to analyse recognition memory data using the methods of signal detection theory (Lockhart & Murdock, 1970). Figure 1 illustrates the picture that emerges. According to signal detection theory, targets and lures give rise to different distributions of *memory strength* with targets having, on average, greater memory strength than lures. These are illustrated in Figure 1 by two normal distributions, labelled lures and targets. It is further assumed that the different confidence judgments correspond to intervals on the memory strength continuum marked off by different decision criteria. These are illustrated in Figure 1 by the set of vertical lines labelled, c_1 to c_5. If the memory strength of an item falls between two adjacent criteria, it is allocated to the corresponding response category, indicated by the labels at the top of Figure 1.

Figure 1. *Signal detection interpretation of recognition memory rating task*

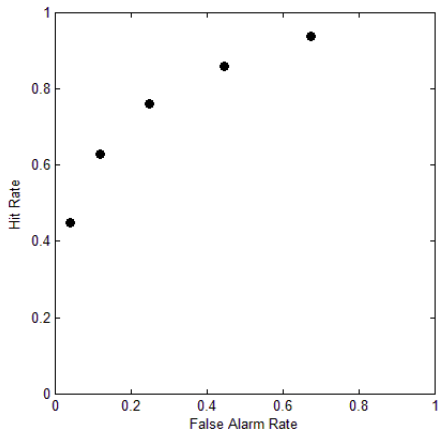


International Journal of Psychological Research, 2010. Vol. 3. No. 1.
ISSN impresa (printed) 2011-2084
ISSN electrónica (electronic) 2011-2079

Dunn, J.C., (2010). How to fit models of recognition memory data using maximum likelihood. *International Journal of Psychological Research, 3* (1), 140-149.

Each decision criterion defines a corresponding pair of hit and false alarm rates. Take, for example, the most stringent criterion corresponding to c_1 in Figure 1. The hit rate corresponding to this criterion is the proportion of targets whose memory strength is greater than c_1 ï those that have been allocated to the *sure target* category. The corresponding false alarm rate is then the proportion of lures whose memory strength is greater than c_1 ï those that have been (incorrectly) allocated to the *sure old* category. Similar reasoning applies to the other decision criteria. Thus the hit rate for c_2 is the proportion of targets that have been allocated to either the *sure target* or *probable target* categories and the corresponding false alarm rate is the proportion of lures so allocated. The result is that if there are *k* response categories there are *k* ï 1 pairs of corresponding hit and false alarm rates.

Figure 2. *ROC curve corresponding to the data in Table 1.*



As stated above, the ROC curve is a plot of hit rate against false alarm rate across the different decision criteria. Although it is called a curve, it is actually in this case no more than a set of points. The remaining rows of Table 1 show how these points are calculated. The middle two rows of Table 1 correspond to the *cumulative number of responses*. These are the number of hits and the number of false alarms, respectively, and are calculated by cumulating the number of responses in each category from *sure target* to *sure lure* (i.e. right to left in Figure 1). The final two rows in Table 1 express each cumulative number as a rate by dividing each by the total number of responses (given in the last column of the cumulative number of responses). These rates are then used to plot the ROC curve shown in Figure 2.

The ROC curve in Figure 2 reveals two distinguishing features that are typical of recognition memory data. First, the set of points trace out a curved rather than a straight line. Second, this curve tends to be asymmetrical. If you look carefully at Figure 2, you should be able to see that the points tend to rise sharply on the left and then decline more gently on the right. Any successful model of recognition memory must be able to account for these two features.

The models

Models or theories serve several important functions in science. A model is a description of the underlying processes that give rise to the observed data. They serve to *organize* and to *explain* these data and to *predict* and to *control* future data (for an excellent introduction to the role of models in psychology see Lewandowsky and Farrell, 2010). For example, Newtonôs Law of Gravity is a model of a range of phenomena such as weight, the trajectories of falling bodies, and the orbits of planets and satellites. It organizes these data by unifying them within a common explanatory scheme. It explains these data through the idea that gravity is a force attracting any two bodies having mass. It predicts future data, such as the motion of the planets, and can be used to control events subject to gravity.

A model may be usefully viewed as potentially consisting of two parts ï a verbal or pictorial description and a mathematical description. Newtonôs Law of Gravity verbally describes gravity as an attractive force between two bodies that is proportional to their masses and inversely proportional to their distance apart. In so doing, it provides a picture of what gravity is or what it is like and this description is often sufficient to organize and to provide an initial explanation of the phenomena. But Newtonôs Law does more than this ï it also provides a precise mathematical description of how gravity operates and, in so doing, allows it to predict and to control the phenomena. However, to do so, it must specify the verbal or pictorial description in a more precise way. Ultimately, this leads to the well-known formula for the force of gravity (*F*) between two bodies as being proportional to the product of their masses (m_1 and m_2) divided by the square of the distance (*d*) between them. That is,

$$F = G\frac{m_1 m_2}{d^2}$$

where *G* is the constant of proportionality, also called the universal gravitational constant.

THE EQUAL VARIANCE SIGNAL DETECTION MODEL

You have already been introduced to one model of recognition memory, that shown in Figure 1. This model pictures recognition memory decisions in terms of signal

International Journal of Psychological Research, 2010. Vol. 3. No. 1.
ISSN impresa (printed) 2011-2084
ISSN electrónica (electronic) 2011-2079

Dunn, J.C., (2010). How to fit models of recognition memory data using maximum likelihood. *International Journal of Psychological Research, 3* (1), 140-149.

detection theory according to which targets and lures have different distributions of memory strength and are evaluated against one or more decision criteria. In so doing, it organizes and explains the data. Beyond this, the model also provides a precise mathematical description of the kind of data shown in Table 1 and, in so doing, accounts for these data and, potentially, predicts future data. This mathematical description is generated by assuming that the target and lure distributions are both normal and have equal variances. This corresponds exactly to picture shown in Figure 1 and leads to what has been called the *equal variance signal detection model* of recognition memory. For this model, the hit and false alarm rates for any particular decision criterion are given by the following two equations:

$$f(c) = \Phi(-c)$$
$$h(c) = \Phi(d - c)$$

where *c* is a decision criterion and *f*(*c*) and *h*(*c*) are the corresponding false alarm and hit rates, respectively. The function, Φ(.), is the normal cumulative distribution function which returns the area under the normal curve to the left of its argument. For example, Φ(0) = 0.5, Φ(1) = 0.8413, and so on. Finally, *d* is the difference between the means of the target and lure distributions[1].

Because we cannot measure it directly, the continuum of memory strength that is posited by the equal variance model does not have any particular scale ï it does not have a defined zero point or unit of measurement. In many ways, this is similar to the way in which we measure temperature which also does not come with a standard zero point or unit of measurement[2]. The two scales in most common use, the Farenheit and Celsius scales, both choose different zero points (0°C is the freezing point of water but this occurs at 32°F) and different units of measurement (a change of 1°C is equivalent to a change of 1.8°F). In the same way that the lack of standard temperature scale does not mean that there is no such thing as temperature, so the lack of a standard memory strength scale does not mean that there is no such thing as memory strength. Rather it means that we are free to define a scale in any way we choose, as is done for temperature. A scale that is commonly used when applying the equal variance model (as well as the other models discussed below) is to set the zero point at the mean of the lure distribution and to set the unit of measurement to be the standard deviation of this distribution. The careful reader will note that this scale is built into the equations for the equal variance model presented above.

As we will discover later, the equal variance signal detection model fails to properly account for the asymmetrical shape of the ROC curve, as shown in Figure 2. For this and other reasons, researchers have proposed a number of modifications of this model to bring it into better alignment with the data. Three of these models are outlined below. Each is currently regarded as a viable model of recognition memory and is the subject of active research programs.

THE UNEQUAL VARIANCE SIGNAL DETECTION MODEL

The first of the three models is called the *unequal variance signal detection model* and is a straightforward extension of the equal variance model (Wixted, 2007). This model preserves the idea of normal distributions but allows the variance of the distributions to be different. The idea behind this is that when participants study a target item, the increase in memory strength that results is not a constant, as assumed by the equal variance model, but may vary from item to item. The unequal variance model assumes that this increase is itself normally distributed and (largely) independent of the item that is studied. As a consequence, the variance of the target distribution must be greater than that of the lure distribution. The equations for this model are,

$$f(c) = \Phi(-c)$$
$$h(c) = \Phi\left(\left(d - c\right)/_{s}\right)$$

where *s* is the standard deviation of the target distribution.

THE MIXTURE SIGNAL DETECTION MODEL

The second model is called the *mixture signal detection model* and is another way of extending the equal variance model to better account for the data (DeCarlo, 2002). According to this model, participants either pay attention to an item in the study phase or they do not. If they do pay attention, the item receives a constant increase in memory strength, just as is assumed by the equal variance model. However, if they donôt pay attention then there is no increase in memory strength ï the target at test has the same memory strength as a lure. As a result, the target distribution becomes a mixture of the target and lure distributions from the equal variance model. The equations for the mixture model are,

[1] In most treatments of signal detection theory, this quantity is called *d*ô, pronounced *d-prime*. Here, it is labelled as *d* in order to preserve its generalizability to other models to be discussed.

[2] It can be argued that absolute zero (-273°C) is the true or natural zero point of temperature. In the same way, it might also be possible to define a zero point of memory strength. While some theorists have attempted to do this, it requires replacing the assumption of normal distributions with the assumption of other distributions that are always positive (e.g. Gamma distribution).

International Journal of Psychological Research, 2010. Vol. 3. No. 1.
ISSN impresa (printed) 2011-2084
ISSN electrónica (electronic) 2011-2079

Dunn, J.C., (2010). How to fit models of recognition memory data using maximum likelihood. *International Journal of Psychological Research, 3* (1), 140-149.

$$f(c) = \Phi(-c)$$
$$h(c) = (1-\lambda)\Phi(-c) + \lambda\Phi(d-c)$$

where λ is the probability of paying attention to a target item at study.

THE HIGH THRESHOLD SIGNAL DETECTION MODEL

The third model to be considered is called the *high threshold signal detection model* and extends the equal variance model by proposing a second kind of memory component and is therefore often called a dual-process model of memory (Yonelinas, 1994). According to this model, the continuum of memory strength proposed by the equal variance model (and the other two models discussed above) reflects only one component of recognition memory, called *familiarity*. This is defined as a general sense that an item is a target (i.e. has been studied in the relevant list) but without being able to retrieve any further information about the study episode. In addition to familiarity, participants are also able to rely on recollection which is defined as the retrieval of information from the study episode. This information may include the appearance of the item or associations that were formed when the item was studied. In the original formulation of this model, recollection was viewed as *all-or-none*. That is, even if only some information from the study episode can be retrieved then this is still sufficient to identify the item as a target. It is further assumed that recollection and familiarity make independent contributions to the recognition memory judgment in which case the equations for the model are,

$$f(c) = \Phi(-c)$$
$$h(c) = r + (1-r)\Phi(d-c)$$

where r is the probability that some information from the study episode is recollected and d is the difference in familiarity between the target and lure distributions.

Fitting the models

Each of the four models listed above posit two or more unknown quantities. The equal variance model posits two quantities, d and c; the unequal variance model posits three, d, c, and s; as does the mixture model with the quantities d, c, and λ; and the high threshold model with d, c, and r. Depending on the values of these quantities, the respective models generate different predicted hit and false alarm rates. With one important difference, these are analogous to the quantities posited by Newtonôs Law of Gravity, G, m_1, m_2, and d, where, by Newton's law, the values of these quantities can be directly measured. In the case of, different values of the predicted force of gravity can be obtained. The important difference is that in the case of Newtonôs law, the values of these quantities of the models of recognition memory, we are not in a position to independently measure any of the posited quantities. Instead, we obtain values for these quantities by fitting the corresponding model to the data.

The idea of obtaining values of the quantities posited by the model by fitting the model seems a bit like lifting yourself by your shoelaces. However, in most cases it turns out to be a practical and legitimate exercise. This is because, in many cases, the data *overdetermine* the model. This means that there are more observed data points than are required to calculate the quantities of interest. This yields two related benefits. First, it is possible to determine how well the model fits or successfully predicts the data. Second, it is possible to attach an error estimate to each of the quantities whose values are obtained. The idea of error or mis-fit is important because then it becomes possible to decide whether or not a model should be accepted as a good account of the data or whether it should be rejected.

MAXIMUM LIKELIHOOD ESTIMATION

What does it mean to say that a model fits the data? In general terms, it means that it is possible to find values of the quantities posited by the model such that the resulting outcomes, in this case hit and false alarm rates, are sufficiently close to the corresponding observed outcomes. The notion of being ñsufficiently closeò can be given a precise meaning in terms of the *maximum likelihood* of the data given the model (for an excellent introduction to this concept and its application to psychological models, see Myung, 2003). For the data given in Table 1, there are 6 response categories for each type of item, target or lure. Each of these categories has an observed probability of being used in the experiment which is estimated by the observed number of responses divided by the total number of responses for that item type. For a particular choice of values of its posited quantities, a given model attaches an expected probability to each of these categories. Let o_i be observed number of responses to the ith category, let e_i be the expected number of responses and let $p_i = e_i/N$ be the expected probability of a response to this category[3]. Then, assuming that the responses are independent, the likelihood of exactly o_i responses is $p_i^{o_i}$, or p_i raised to the power of o_i. If there are a lot of responses then this is a very small number and so it is often more convenient to use the *log likelihood* of the data where $\log(p_i^{o_i}) = o_i \log(p_i)$. Then the total log likelihood of the data given the model is simply the sum of the log likelihoods of each category. That is,

[3] If the ith category is from the set of lures then N is the total number of lures, otherwise N is the total number of targets.

International Journal of Psychological Research, 2010. Vol. 3. No. 1.
ISSN impresa (printed) 2011-2084
ISSN electrónica (electronic) 2011-2079

Dunn, J.C., (2010). How to fit models of recognition memory data using maximum likelihood. *International Journal of Psychological Research, 3* (1), 140-149.

$$LL = \sum_i o_i \log(p_i)$$

Because each p_i is a probability, $\log(p_i)$ is a negative number and hence the total log likelihood is also a negative number. To the extent that a model predicts that the data are likely, the total log likelihood will be at a maximum. The aim of maximum likelihood estimation (MLE) is to find a set of model predictions that maximizes *LL*.

Table 2. *Predicted hit rates and false alarm rates for the equal variance signal detection model.*

		Decision Criteria				
Item Type	Displacement	c_1	c_2	c_3	c_4	c_5
Lure	0	$\Phi(\text{ï}\ c_1)$	$\Phi(\text{ï}\ c_2)$	$\Phi(\text{ï}\ c_3)$	$\Phi(\text{ï}\ c_4)$	$\Phi(\text{ï}\ c_5)$
Target	*d*	$\Phi(d\ \text{ï}\ c_1)$	$\Phi(d\ \text{ï}\ c_2)$	$\Phi(d\ \text{ï}\ c_3)$	$\Phi(d\ \text{ï}\ c_4)$	$\Phi(d\ \text{ï}\ c_5)$

FITTING THE EQUAL VARIANCE MODEL

To illustrate the foregoing, we can fit the equal variance model to the observed data from Table 1. The first thing to note is the structure of these data. The observed data consist of the number of responses in each category. It is these data that the model will attempt to fit and from which *LL* will be calculated. However, the model equations do not express these quantities directly. Instead the equations specify a hit rate and a false alarm rate for a given decision criterion as shown in Table 2. Altogether, there are six quantities to be estimated ï the five decision criteria and the displacement, *d*. These six quantities are called the *parameters* of the model.

For a particular choice of parameter values, the expected hit and false alarm rates for each decision criterion can be calculated (see Table 2). By multiplying each of these rates by the total number of observed responses for lures and targets, the expected cumulative number of responses for the first five responses categories can be calculated. The expected cumulative number of responses for the sixth category is given by the total number of responses. From the expected cumulative number of responses, the expected number of responses in each category can be obtained. It is these that enter the formula for the log likelihood (*LL*).

The foregoing is an iterative process that searches for the set of parameter values that maximizes *LL* and, generally speaking, a computer program is required to do this. One such program that is both powerful and easy to use is the Solver™ tool in Microsoft Excel (Fylstra, Lasdon, Watson & Waren, 1998) although it might be necessary to install it before use. Figure 3 shows an Excel worksheet for fitting the equal variance signal detection model[4]. The tables on the left hand side of the sheet recapitulate the entries in Table 1. The tables on the right hand side show the calculation and evaluation of the model predictions. Just as the tables on the left are to be read from top to bottom, the tables on the right, starting with the one labelled ñEXPECTED RATEò are to be read from bottom to top. The ñEXPECTED RATEò table calculates the expected hit and false alarm rates and is laid out in the same way as Table 2. The values in **bold** indicate the parameters that are being fitted[5]. The table immediately above, labelled ñEXPECTED CUM. NUMBERò gives the expected cumulative number of responses in each category and can be calculated directly from the ñRATEò table below it. Finally, the table labelled ñEXPECTED NUMBERò gives the expected number of responses and is, in turn, calculated from the cumulative numbers in the table below. It is these values, in conjunction with the corresponding observed numbers that are used to calculate the log likelihood. This is done in the table labelled ñ*LL*ò in which each cell contains the observed number of responses for the corresponding cell in the ñOBSERVED NUMBERò table multiplied by the (natural) logarithm of the expected *proportion* of responses derived from the corresponding cell in the ñEXPECTED NUMBERò table. The total log likelihood is the sum of these values and is indicated in *italics* in the cell immediately below that labelled ñ*LL*ò. After Solver has been invoked, this cell is selected as the target cell (the value to be maximized) and the cells in **bold** are selected as the cells to be changed. If a solution is found, these cells will contain the best fitting parameter values.

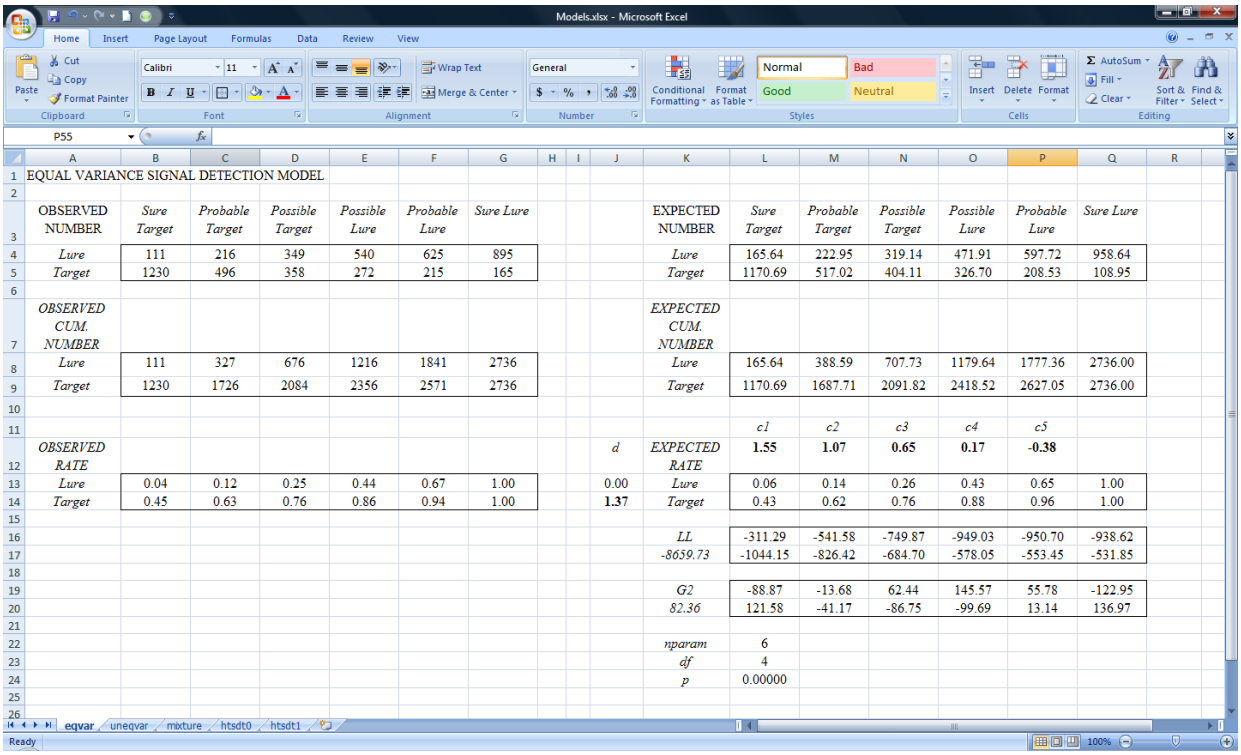
Because it is the log likelihood that is maximized, these values are called the *maximum likelihood parameter estimates*.

[4] The file containing this worksheet as well as additional sheets for each of the other three models can be obtained either from the author or from the IJPR website at http://mvint.usbmed.edu.co:8002/ojs/index.php/web

[5] This is a useful mnemonic to help identify the cells containing the parameter values. Similarly, the cell containing the value to be optimized is identified by italics.

Dunn, J.C., (2010). How to fit models of recognition memory data using maximum likelihood. *International Journal of Psychological Research, 3* (1), 140-149.

Figure 3. *Screen shot of an Excel spreadsheet for fitting the equal variance signal detection model.*



Models.xlsx - Microsoft Excel

EQUAL VARIANCE SIGNAL DETECTION MODEL

OBSERVED NUMBER	Sure Target	Probable Target	Possible Target	Possible Lure	Probable Lure	Sure Lure
Lure	111	216	349	540	625	895
Target	1230	496	358	272	215	165

OBSERVED CUM. NUMBER						
Lure	111	327	676	1216	1841	2736
Target	1230	1726	2084	2356	2571	2736

OBSERVED RATE							d
Lure	0.04	0.12	0.25	0.44	0.67	1.00	0.00
Target	0.45	0.63	0.76	0.86	0.94	1.00	**1.37**

EXPECTED NUMBER	Sure Target	Probable Target	Possible Target	Possible Lure	Probable Lure	Sure Lure
Lure	165.64	222.95	319.14	471.91	597.72	958.64
Target	1170.69	517.02	404.11	326.70	208.53	108.95

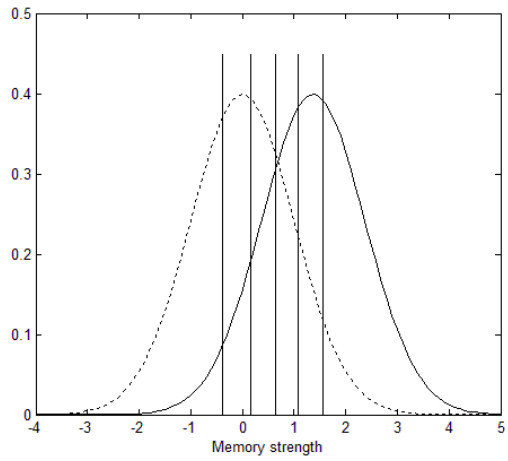
EXPECTED CUM. NUMBER						
Lure	165.64	388.59	707.73	1179.64	1777.36	2736.00
Target	1170.69	1687.71	2091.82	2418.52	2627.05	2736.00

	c1	c2	c3	c4	c5
EXPECTED RATE	**1.55**	**1.07**	**0.65**	**0.17**	**-0.38**
Lure	0.06	0.14	0.26	0.43	0.65	1.00
Target	0.43	0.62	0.76	0.88	0.96	1.00

LL	-311.29	-541.58	-749.87	-949.03	-950.70	-938.62
-8659.73	-1044.15	-826.42	-684.70	-578.05	-553.45	-531.85
G2	-88.87	-13.68	62.44	145.57	55.78	-122.95
82.36	121.58	-41.17	-86.75	-99.69	13.14	136.97

nparam	6
df	4
p	0.00000

Figure 4. *Best-fitting solution of the equal variance signal detection model.*



Evaluating the models

The cells marked in bold in Figure 3 show the maximum likelihood parameter estimates for the equal variance signal detection model of the data shown in Table 1. It is often a good idea to inspect these to see if they are plausible. In the present case they are. The estimate of *d* is 1.37. This means, according to the model, that the mean of the target distribution is 1.37 units (i.e. standard deviations of the lure distribution) greater than the mean of the lure distribution. This seems plausible. Similarly, the values for the five decision criteria also seem plausible. Figure 4 shows the estimated distributions and decision criteria based on these maximum likelihood estimates. The similarity to the idealized picture in Figure 1 is clear.

EVALUATING THE EQUAL VARIANCE MODEL

At this point, as well as obtaining the maximum likelihood parameter estimates, it is important to know whether the model in question provides a good fit to the data. If it does then it may be accepted as a satisfactory account of the data and the parameter values used in some further way, e.g. to examine changes across different experimental conditions. If the model does not fit the data well then we have grounds for rejecting it and evaluating alternative models to see if they do a better job.

How can model fit be evaluated? Fortunately, maximum likelihood estimation leads naturally to a statistic, called the *likelihood ratio test statistic* or the G^2

Dunn, J.C., (2010). How to fit models of recognition memory data using maximum likelihood. *International Journal of Psychological Research, 3* (1), 140-149.

statistic[6], that can be used to evaluate how well a model fits the data ï its *goodness of fit*. In the current context, this statistic can be defined as follows;

$$G^2 = 2\sum_i o_i \log\left({}^{o_i}/_{e_i}\right)$$

where o_i and e_i are the observed and expected number of responses, respectively, in category *i*. The table labelled ñG2ò in Figure 3 shows the calculation of this quantity. Each cell in this table contains $2o_i \log\left({}^{o_i}/_{e_i}\right)$ for the corresponding category *i*. The value beneath the cell labelled ñG2ò contains the final sum.

Another way of defining G^2 is in terms of the difference between two log likelihoods. That is,

$$G^2 = 2\left(\sum_i o_i \log\left({}^{o_i}/_{N}\right) - \sum_i o_i \log\left({}^{e_i}/_{N}\right)\right)$$

The second term in the parentheses is just the *LL* value that has been maximized. The first term can also be thought of as a maximized log likelihood, in this case of a model that predicts the data perfectly ï where the expected number of responses in each cell is exactly equal to the observed number. Such a model can be easily devised ï it would have one parameter for each hit or false alarm rate. Since there are 10 such rates in the present example, this model would have 10 parameters. And since it can predict the data perfectly with this number of parameters, it is said to be *saturated*.

Looked at as a difference in log likelihoods, G^2 can be thought of as a measure of how close a particular model is to the best possible model ï the saturated model. If the difference is small, the model is doing a good job, while if the difference is large, it is doing more poorly. This difference though depends upon two factors. The first is the intrinsic fit of the model. The second is the number of parameters in the model. All things being equal, the more parameters a model has the better it will fit the data. We have seen this already in the case of the saturated model ï it has the most parameters and fits the data perfectly.

It turns out that there is a test for the size of G^2 that takes into account both the intrinsic fit of the model and how many parameters it has. This is because, for large samples, G^2 is distributed as a chi-squared test statistic with degrees of freedom given by the difference between the number of parameters in the saturated model and the number of parameters in the model of interest. Since the saturated model has 10 parameters and the equal variance model has 6 parameters (consisting of *d* and the five decision criteria), it has 4 degrees of freedom. The critical χ^2 with 4 d.f. and $\alpha = 0.01$ is 13.28. The observed G^2 value is 82.36 and so we can conclude that the equal variance model does not provide a satisfactory fit to the data. In fact, it provides a very poor fit.

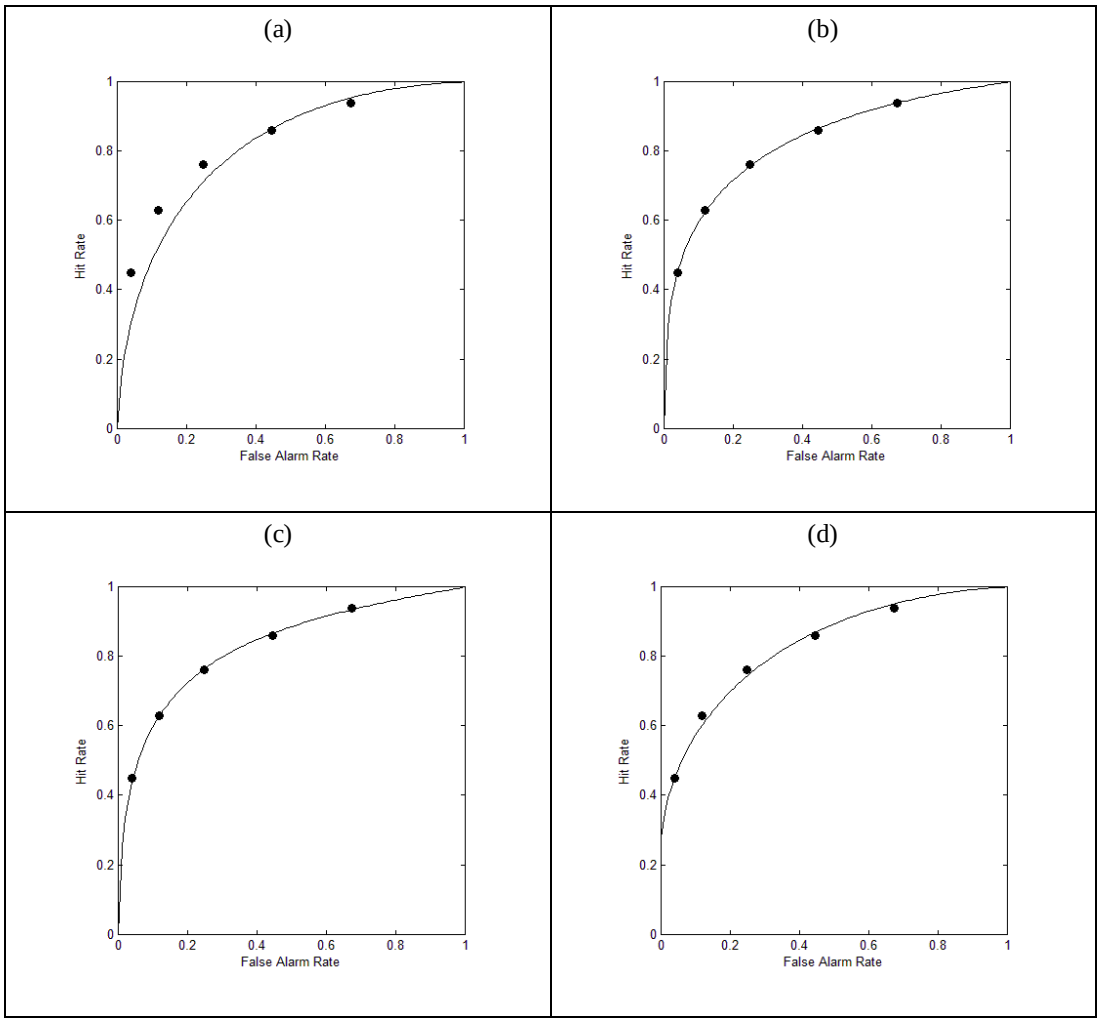
The failure of the equal variance model is shown in Figure 5a. This shows the same ROC plot of the data as in Figure 2 along with the ROC curve predicted by the best fitting equal variance model. Although the model was fit only to the data shown in the Figure, it is possible to use the maximum likelihood parameter estimates to construct an entire ROC curve. What this curve demonstrates is the major weakness of the equal variance model ï its failure to account for the asymmetrical nature of the recognition memory ROC curve. The equal variance model ROC curve is symmetrical and so tends to underestimate data on the left of the curve and to overestimate data on the right of the curve. This is a systematic effect and because each response category contains a large number of observations, even relatively small departures from the data are statistically significant. Therefore we can reject the equal variance model.

EVALUATING THE ALTERNATIVE MODELS

The remaining panels in Figure 5 show the best-fitting ROC curves for the other three models that we have considered ï the unequal variance model (Panel b), the mixture model (Panel c) and the high threshold model (Panel d). The G^2 values obtained for these models are 3.03, 4.37, and 23.89, respectively. As noted earlier, each of these models has one more parameter than the equal variance model and so the associated degrees of freedom of each are reduced by one to 3. The critical χ^2 with 3 d.f. and $\alpha = 0.01$ is 11.34 which suggests that both the unequal variance and mixture models fit the data well (the actual *p* values are 0.39 and 0.22, respectively) while the high threshold model can be rejected (actual *p* value is 0.00003). As with the equal variance model, it is instructive to see how the high threshold model fails to account for the data. In this case, inspection of Figure 5d suggests that its major failing is that it is not sufficiently curvilinear to capture the present data (even though the deviations seem quite small). It should be borne in mind that the present data are the results of only one experiment and have been summed over a number of individual participants and it is possible that a different outcome might be found if each participants is modelled individually (for an investigation of the relative benefits of group or individual level modelling, see Cohen, Sanborn & Shiffrin, 2008).

[6] This statistic is most frequently referred to as the G^2 statistic in the psychology literature and I will maintain that usage here.

Figure 5. *Observed and predicted ROC curves for each model. (a) Equal variance signal detection model. (b) Unequal variance signal detection model. (c) Mixture signal detection model. (d) High threshold signal detection model*



CONCLUSIONS

The aim of the present paper has been to provide an introductory tutorial on how to fit models of recognition memory data using maximum likelihood estimation. However, the relevance of this approach extends beyond the relatively narrow field of models of recognition memory. Many psychological models, if expressed with sufficient mathematical precision, can be fit to data using MLE. The value of this approach is that it includes a wide range of additional statistical machinery, such as the G^2 statistic, which can be used to answer many questions about models and their relationships to the data. It is my hope that as more researchers become acquainted with these techniques, they will become increasingly prepared to propose models that go beyond a verbal or pictorial description and include a precise mathematical description which would then allow it to be formally evaluated against data in order to determine if it satisfactorily accounts for the data and, if it does not, how and why it fails.

REFERENCES

Cohen, A. L., Sanborn, A. N. & Shiffrin, R. M. (2008). Model evaluation using grouped or individual data. *Psychonomic Bulletin & Review, 15(4)*, 692-712.

DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: Theoretical developments with applications to recognition memory. *Psychological Review, 109(4),* 710ï 721.

Fylstra, D., Lasdon, L., Watson, J. & Waren, A. (1998). Design and use of the Microsoft Excel Solver, *Informs Interfaces, 28(5)*, 29-55.

Lewandowsky, S. & Farrell, S. (2010). *Elements of cognitive modeling*.

Lockhart, R. S. & Murdock, B. B. (1970). Memory and the theory of signal detection. *Psychological Bulletin, 74(2),* 100-109.

Macmillan, N. A. & Creelman, C. D. (2004). *Detection theory: A userôs guide*. Mahwah, NJ: Erlbaum.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology, 47*, 90ï 100.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114(1),* 152ï 176.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20(6),* 1341-1354

Dunn, J.C., (2010). How to fit models of recognition memory data using maximum likelihood. *International Journal of Psychological Research, 3* (1), 140-149.