LIKELIHOOD RATIO TESTS

FOR

PROBABILISTIC CONCEPT LEARNING

Jan de Leeuw

June 1969

PSYCHOLOGICAL INSTITUTE - UNIVERSITY OF LEYDEN - NETHERLANDS

1: Introduction

In probabilistic concept learning experiments such as those discussed by De Klerk (1968), stimuli are sampled from two multinormal populations (with identical dispersion matrices and equal a priori probabilities). A stimulus that is sampled from the first population is called a positive instance (A) of the concept, a stimulus from the second population a negative instance (B). Subjects are instructed to give their subjective probabilities that a particular stimulus is a positive instance of the concept. Mathematical models for these experiments suppose that the subjects build up two subjective sampling distributions (SSD's), that these SSD's are multinormal with identical dispersion matrices, and that the subjective probabilities are computed by applying Bayes' Rule to the SSD's. Moreover it is assumed that the subjective posterior probabilities add up to one. By the Baysian assumption

$$\Psi(A/x) = \frac{\Psi(A)\,\Psi(x/A)}{\Psi(A)\Psi(x/A) + \Psi(B)\Psi(x/B)}\,. \tag{1}$$

By additivity

$$1 - \Psi(A/x) = \Psi(B/x) = \frac{\Psi(B)\,\Psi(x/B)}{\Psi(A)\Psi(x/A) + \Psi(B)\Psi(x/B)}\,. \tag{2}$$

Combining (1) and (2)

$$\frac{\Psi(A/x)}{1 - \Psi(A/x)} = \frac{\Psi(A)}{\Psi(B)} \cdot \frac{\Psi(x/A)}{\Psi(x/B)}\,, \tag{3}$$

or

$$\log \frac{\Psi(A/x)}{1 - \Psi(A/x)} = \log \frac{\Psi(x/A)}{\Psi(x/B)} + \log \frac{\Psi(A)}{\Psi(B)}\,. \tag{4}$$

By the multinormality assumption

$$\Psi(y/A) = (2\pi)^{-\frac{1}{2}m}|V|^{-\frac{1}{2}}\exp\left\{-\tfrac{1}{2}(y-m_A)'V^{-1}(y-m_A)\right\}, \tag{5a}$$

$$\Psi(y/B) = (2\pi)^{-\frac{1}{2}m}|V|^{-\frac{1}{2}}\exp\left\{-\tfrac{1}{2}(y-m_B)'V^{-1}(y-m_B)\right\}, \tag{5b}$$

where m is the dimensionality of the subjective sampling space, y is an element of that space, m_A and m_B are the vectors of means, and V is the common dispersion matrix. Substitution of (5a) and (5b) into (4) proves

$$S(y) =_{\text{def}} \log \frac{\psi(A/y)}{1-\psi(A/y)} =$$

$$= (m_A - m_B)'V^{-1}y - \tfrac{1}{2}(m_A - m_B)'V^{-1}(m_A + m_B) + \gamma, \tag{6}$$

with

$$\gamma =_{\text{def}} \log \frac{\psi(A)}{\psi(B)}. \tag{7}$$

Without loss of generality we may assume V diagonal. It follows that the density function of S(y) is a mixture of two normal density functions with means

$$m_1 = \tfrac{1}{2}(m_A - m_B)'V^{-1}(m_A - m_B) + \gamma, \tag{8a}$$

$$m_2 = -\tfrac{1}{2}(m_A - m_B)'V^{-1}(m_A - m_B) + \gamma, \tag{8b}$$

and variances

$$s_1^2 = s_2^2 = (m_A - m_B)'V^{-1}(m_A - m_B). \tag{9}$$

For a detailed proof of these results cf De Leeuw (1968). In the objective situation we have the similar results

$$B(x) =_{\text{def}} (\mu_A - \mu_B)'\Lambda^{-1}x - \tfrac{1}{2}(\mu_A - \mu_B)'\Lambda^{-1}(\mu_A + \mu_B) \tag{10}$$

and

$$\mu_1 = \tfrac{1}{2}(\mu_A - \mu_B)'\Lambda^{-1}(\mu_A - \mu_B) = -\mu_2, \tag{11}$$

$$\zeta_1^2 = \zeta_2^2 = (\mu_A - \mu_B)'\Lambda^{-1}(\mu_A - \mu_B). \tag{12}$$

We may add two additional assumptions: if S_1 is the subjective space and S_2 the objective one, then there is a matrix A (not necessarily square) such that $Ay = x$ for all $y \in S_1$ and $x \in S_2$ (the linearity assumption). Moreover there is a real constant δ such that $(\mu_A - \mu_B)'\Lambda^{-1}x = (\mu_A - \mu_B)'\Lambda^{-1}Ay = \delta(m_A - m_B)'V^{-1}y$ for all $y \in S_1$ (the proportionality assumption). (In De Klerk (1968) and De Leeuw (1968) somewhat stronger assumptions are used). If both S_1 and S_2 are one-dimensional these assumptions reduce to the assertion that the subjective dimension values are proportional to the objective ones. It follows from linearity and proportionality that S(y) is a linear function of B(x), or, in other words that there are real numbers α and β such that

$$\Psi(A/x) = \frac{1}{1 + \exp\{-\alpha B(x) - \beta\}} . \tag{13}$$

In the study of choice reaction times (CRT) there are two competitative models. One model, proposed by Falmagne, is essentially based on a dichotomous way of thinking. A particular stimulus may be expected by the subject, or it may come unexpected. Both types of stimuli generate a normal density of CRT's (with, of course, the mean of the 'expected' density less than of the 'unexpected'). The alternative is a model propose by Oppe. In this model expectancy is a continuous variable related in a simple way to the objective posterior probabilities of the stimuli. Stimuli with equal posterior probabilities have equal expectancy and generate one single normal distribution.

2: General case

The most general distribution we shall consider in this paper is

$$dF = \frac{P}{\sqrt{2\pi \rho_1}} \exp\left\{-\tfrac{1}{2} \frac{(x-\mu_1)^2}{\rho_1}\right\} + \frac{(1-P)}{\sqrt{2\pi \rho_2}} \exp\left\{-\tfrac{1}{2} \frac{(x-\mu_2)^2}{\rho_2}\right\} dx \tag{14}$$

This is a natural generalization of the theoretical distributions discussed in the introduction. We have a sample of n elements x_i. Let

$$y_{1i} = \frac{1}{\sqrt{2\pi \rho_1}} \exp\left\{-\tfrac{1}{2} \frac{(x_i - \mu_1)^2}{\rho_1}\right\}, \tag{15}$$

$$y_{2i} = \frac{1}{\sqrt{2\pi \rho_2}} \exp\left\{-\tfrac{1}{2} \frac{(x_i - \mu_2)^2}{\rho_2}\right\}, \tag{16}$$

$$z_{1i} = \frac{x_i - \mu_1}{\sqrt{\rho_1}} , \tag{17}$$

$$z_{2i} = \frac{x_i - \mu_2}{\sqrt{\rho_2}} , \tag{18}$$

$$h_i = Py_{1i} + (1-P)y_{2i} . \tag{19}$$

The likelihood function is

$$L = L(\mu_1, \mu_2, \rho_1, \rho_2, P) = \sum_{i=1}^{n} h_i \ , \tag{20}$$

and its logarithm

$$\mathcal{L} = \sum_{i=1}^{n} \log h_i . \tag{21}$$

If α, β are any two paramaters

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{i=1}^{n} \frac{1}{h_i} \frac{\partial h_i}{\alpha} , \tag{22}$$

and

$$\frac{\partial \mathcal{L}}{\partial \alpha \partial \beta} = \sum_{i=1}^{n} \left[\frac{1}{h_i} \frac{\partial h_i}{\partial \alpha \partial \beta} - \frac{1}{h_i^2} \frac{\partial h_i}{\partial \alpha} \frac{\partial h_i}{\partial \beta} \right] . \tag{23}$$

Now

$$\frac{\partial h_i}{\partial P} = y_{1i} - y_{2i} , \tag{24}$$

$$\frac{\partial h_i}{\partial \mu_1} = \frac{P y_{1i} z_{1i}}{\sqrt{\rho_1}} , \tag{25}$$

$$\frac{\partial h_i}{\partial \mu_2} = \frac{(1-P) y_{2i} z_{2i}}{\sqrt{\rho_2}} , \tag{26}$$

$$\frac{\partial h_i}{\partial \rho_1} = \frac{P y_{1i} (z_{1i}^2 - 1)}{2 \rho_1} , \tag{27}$$

and

$$\frac{\partial h_i}{\partial \rho_2} = \frac{(1-P) y_{2i} (z_{2i}^2 - 1)}{2 \rho_2} . \tag{28}$$

There are 15 second-order partial derivatives:

$$\frac{\partial^2 h_i}{\partial P^2} = 0 , \tag{29}$$

$$\frac{\partial^2 h_i}{\partial P \partial \mu_1} = \frac{y_{1i} z_{1i}}{\sqrt{\rho_1}} , \tag{30}$$

$$\frac{\partial^2 h_i}{\partial P \, \partial \mu_2} = \frac{y_{2i} z_{2i}}{\sqrt{\rho_2}} , \tag{31}$$

$$\frac{\partial^2 h_i}{\partial P \, \partial \rho_1} = \frac{y_{1i}(z_{1i}^2 - 1)}{2\rho_1} , \tag{32}$$

$$\frac{\partial^2 h_i}{\partial P \, \partial \rho_2} = \frac{y_{2i}(z_{2i}^2 - 1)}{2\rho_2} , \tag{33}$$

$$\frac{\partial^2 h_i}{\partial \mu_1^2} = \frac{P y_{1i}(z_{1i}^2 - 1)}{\rho_1} , \tag{34}$$

$$\frac{\partial^2 h_i}{\partial \mu_1 \, \partial \mu_2} = 0, \tag{35}$$

$$\frac{\partial^2 h_i}{\partial \mu_1 \, \partial \rho_1} = \frac{P y_{1i} z_{1i}(z_{1i}^2 - 3)}{2\rho_1 \sqrt{\rho_1}} , \tag{36}$$

$$\frac{\partial^2 h_i}{\partial \mu_1 \, \partial \rho_2} = 0, \tag{37}$$

$$\frac{\partial^2 h_i}{\partial \mu_2^2} = \frac{(1-P) y_{2i}(z_{2i}^2 - 1)}{\rho_2} , \tag{38}$$

$$\frac{\partial^2 h_i}{\partial \mu_2 \, \partial \rho_1} = 0, \tag{39}$$

$$\frac{\partial^2 h_i}{\partial \mu_2 \, \partial \rho_2} = \frac{(1-P) y_{2i} z_{2i}(z_{2i}^2 - 3)}{2\rho_2 \sqrt{\rho_2}} , \tag{40}$$

$$\frac{\partial^2 h_i}{\partial \rho_1^2} = \frac{P y_{1i}(z_{1i}^4 - 6 z_{1i}^2 + 3)}{4\rho_1^2} , \tag{41}$$

$$\frac{\partial^2 h_i}{\partial \rho_1 \partial \rho_2} = 0, \tag{42}$$

$$\frac{\partial^2 h_i}{\partial \rho_2^2} = \frac{(1-P)y_{2i}(z_{2i}^4 - 6z_{2i}^2 + 3)}{4\rho_2^2}. \tag{43}$$

These expressions for the first and second partial derivatives enable us to use the Newton-Raphson method to obtain ML-estimates. If $x^{(k)}$ is an estimate of the (five) parameters, g is the vector of first partial derivatives evaluated at $x^{(k)}$, and V is the matrix of second partial derivatives at $x^{(k)}$, then we use the iterative scheme

$$x^{(k+1)} = x^{(k)} + V^{-1}g. \tag{44}$$

If we start with a suitable $x^{(o)}$ then, for $k \to \infty$, $x^{(k)}$ converges to the vector of ML-estimates. If n is large then the ML-estimators are multinormally distributed with vector of means $x^{(\infty)}$ and dispersion matrix $-V^{-1}$ (the inverse of the matrix of second order partial derivatives at the maximum).

3: Testing of hypotheses

The hypothesis $P = \pi$: Maximize

$$\mathcal{L}_\pi(\mu_1, \mu_2, \rho_1, \rho_2) = \sum \log h_i, \tag{45}$$

with

$$h_i = \pi y_{1i} + (1 - \pi)y_{2i}, \tag{46}$$

where π is some constant value (in PCL we may want to test $\pi = \frac{1}{2}$). We use the derivatives and second derivatives from the previous section with π substituted for P and without considering the derivatives with respect to P. Asymptotic theory for LR-tests says that

$$\lambda = -2\left\{\mathcal{L}_\pi(\mu_1, \mu_2, \rho_1, \rho_2) - \mathcal{L}(\mu_1, \mu_2, \rho_1, \rho_2, P)\right\} \tag{47}$$

is distributed as χ^2 with one df if $n \to \infty$.

Special case $\pi = 1$: In this case

$$dF = \frac{1}{\sqrt{2\pi\rho_1}} \exp\left\{-\tfrac{1}{2}\frac{(x-\mu_1)^2}{\rho_1}\right\} dx, \tag{48}$$

and the ML-estimators of μ_1 and ρ_1 are the sample mean m and sample variance s^2. Moreover the asymptotic variances of these estimates are, respectively, s^2/n and $2s^4/n$, while their convariance vanishes asymptotically (Kendall & Stuart 1967, p 57). Of course this is the hypothesis we are interested in in CRT learning situations (the hypothesis $\pi = 0$ is trivially equivalent to $\pi = 1$).

<u>Hypothesis $P = \frac{1}{2}$, $\mu_1 = -\mu_2 = \frac{1}{2}\alpha$, $\rho_1 = \rho_2 = \alpha$</u> : Case I:

$$\alpha = (\mu_A - \mu_B)'\Lambda^{-1}(\mu_A - \mu_B) \tag{49}$$

We test the hypothesis that the subject copies the objective situation exactly. The log-likelihood function for this case has no unknowns, it is simply a single real number $\mathcal{L}$.

$$\mathcal{L} = \sum \log h_i, \tag{50}$$

with

$$h_i = \tfrac{1}{2}(y_{1i} + y_{2i}), \tag{51}$$

$$y_{1i} = \frac{1}{\sqrt{2\pi\alpha}} \exp\left\{-\tfrac{1}{2}\frac{(x_i - \frac{1}{2}\alpha)^2}{\alpha}\right\}, \tag{52}$$

$$y_{2i} = \frac{1}{\sqrt{2\pi\alpha}} \exp\left\{-\tfrac{1}{2}\frac{(x_i + \frac{1}{2}\alpha)^2}{\alpha}\right\}. \tag{53}$$

Asymptotic theory for LR-tests says that, if $n \to \infty$, then

$$\lambda = -2\left\{\mathcal{L} - \mathcal{L}(\mu_1, \mu_2, \rho_1, \rho_2, P)\right\} \tag{54}$$

is distributed as χ^2 with 5 df. Moreover, if $n \to \infty$,

$$\lambda = -2\left\{\mathcal{L} - \mathcal{L}_\pi(\mu_1, \mu_2, \rho_1, \rho_2)\right\} \tag{55}$$

with $\pi = \frac{1}{2}$ is distributed as χ^2 with 4 df.

Case II: We may also consider α in (52) and (53) as a variable. The hypothesis $P = \frac{1}{2}$, $\mu_1 = -\mu_2 = \frac{1}{2}\alpha$, $\rho_1 = \rho_2 = \alpha$ means in this case that the subject is a perfect Baysian strategist but he may be conservative (or whatever the opposite of conservative may be). $\mathcal{L}(\alpha)$ can be maximized by the one-dimensional Newton-Raphson method with α as the only parameter. The corresponding χ^2 has 4 df. By dropping the assumption that $\mu_1 = -\mu_2$ we introduce the possibility of bias. Of course many more combinations of H_0 and H_1 are possible, the only thing that must always be true is that H_0 is a restriction on the parameters of H_1. Our advise for PCL-experiments is to test the following sequence of hypotheses:

i) $\mu_1 = -\mu_2$.

ii) $\rho_1 = \rho_2$.

iii) $\mu_1 = -\mu_2 = \frac{1}{2}\rho_1 = \frac{1}{2}\rho_2$.

iv) $\mu_1 = -\mu_2 = \frac{1}{2}\rho_1 = \frac{1}{2}\rho_2 = \frac{1}{2}\alpha$.

4: References

De Klerk, L.F.W. — Probabilistic Concept Learning. Voorschoten, VAM, 1968.

De Leeuw, J. — Probabilistic Concept Learning. Department of Data Theory, University of Leyden, Report RN 002-68, 1968.

Kendall, M.G. & Stuart, A. — The advanced theory of statistics. Volume II, Second Edition, London, Griffin, 1967.