Jan de Leeuw

PROBABILISTIC CONCEPT LEARNING

RB - 001 - 72

0 Introduction

In this paper we discuss some applications of the general techniques discussed in De Leeuw (1971) to functional learning experiments (De Kler De Leeuw, & Oppe 1970).

1 Probabilistic concept learning (PCL): I

A thorough discussion of PCL-tasks can be found in De Klerk & Oppe (1966 De Klerk (1968), De Leeuw (1968 a,b,c, 1969), Lee (1963, 1966), Lee & Ja (1964), De Klerk, De Leeuw, & Oppe (1970).

1.1 Situation

1: A stimulus space X, can be identified with a p-dimensional Euclidea space.

2: Two multinormal distributions η_A and η_B on X with densities $N(x;\mu_A,\Sigma)$ and $N(x;\mu_B,\Sigma)$. The dispersion matrix Σ is nonsingula

3: An element $x \in X$.

4: The two hypotheses H_A: η_A and H_B: η_B, and the two corresponding decisions D_A and D_B. The hypotheses have prior probabilities π_A and π_B, $\pi_A + \pi_B = 1$, $0 < \pi_A, \pi_B < 1$.

5: A pay-off structure

	H_A	H_B
D_A	λ_{AA}	λ_{AB}
D_B	λ_{BA}	λ_{BB}

with $\lambda_{AA} - \lambda_{BA}, \lambda_{BB} - \lambda_{AB} > 0$.

1.2 Statistical analysis

1: Posterior probabilities

$$\pi(H_A \mid x) = \frac{\pi_A N(x;\mu_A,\Sigma)}{\pi_A N(x;\mu_A,\Sigma) + \pi_B N(x;\mu_B,\Sigma)},$$

$$\pi(H_B \mid x) = \frac{\pi_B N(x;\mu_B,\Sigma)}{\pi_A N(x;\mu_A,\Sigma) + \pi_B N(x;\mu_B,\Sigma)}.$$

Alternatively

$$\pi(H_A \mid x) = \frac{1}{1 + \exp\{-\delta'\Sigma^{-1}(x-\mu) - \gamma\}},$$

$$\pi(H_B \mid x) = \frac{\exp\{-\delta'\Sigma^{-1}(x-\mu) - \gamma\}}{1 + \exp\{-\delta'\Sigma^{-1}(x-\mu) - \gamma\}},$$

with

$\delta = \mu_A - \mu_B$,

$\mu = \frac{1}{2}(\mu_A + \mu_B)$,

$\gamma = \log \pi_A - \log \pi_B$.

2: Posterior expected pay-off

$P(D_A \mid x) = \lambda_{AA}\pi(H_A \mid x) + \lambda_{AB}\pi(H_B \mid x)$.

$P(D_B \mid x) = \lambda_{BA}\pi(H_A \mid x) + \lambda_{BB}\pi(H_B \mid x)$.

3: Optimal strategy

D_A/D_B if $P(D_A \mid x) \gtrless P(D_B \mid x)$ iff

$\text{logit}\ \pi(H_A \mid x) = \delta'\Sigma^{-1}(x-\mu) + \gamma \gtrless \eta$,

with

$$\eta = \log \frac{\lambda_{BB} - \lambda_{AB}}{\lambda_{AA} - \lambda_{BA}}.$$

1.3 Psychological analysis

1: Training

The subject does not know the basis parameters μ_A, μ_B, Σ, π_A. In fact he does not even know that the hypotheses specify two different multinormal distributions. In a training-run he is shown a random sample from the mixture $\pi_A\eta_A + \pi_B\eta_B$, i.e. he is shown a number of $x \in X$ and in each instance he is told: 'Here H_A/H_B is true'.

2: Basic assumption

It is assumed that during training the subject builds up two 'subjective' multinormal distributions on X with densities $N(x;\, m_A,\, S)$ and $N(x;\, m_B,\, S)$ and prior probabilities p_A and p_B, and that the subject computes posterior probabilities in the usual way, i.e.

$\text{logit}\ p(H_A \mid x) = d'S^{-1}(x - m) + c$,

where d, m, c are the subjective analogues of δ, μ, γ.

3: All-or-none strategy

$$p(D_A \mid x) = \begin{cases} 1 & \text{if } d'S^{-1}(x - m) > y - c, \\ 0 & \text{if } d'S^{-1}(x - m) < y - c, \end{cases}$$

with

$$y = \log \frac{g_{BB} - g_{AB}}{g_{AA} - g_{BA}},$$

and $g_{..}$ the utility of $\lambda_{..}$.

4: Event-matching strategy

$$p(D_A \mid x) = p(H_A \mid x).$$

5: Logistic strategy

$$\text{logit } p(D_A \mid x) = v'x + u.$$

The event-matching strategy is a special case, the all-or-none strategy is another (limiting) special case.

6: Criticism

Of course 1.3.2 is, essentially, a very primitive type of assumption. In 1.2 we described an optimal decision theoretical strategy, in the analysis we describe the subjects output by using a model of the same mathematical form. We estimate the free paparameters and compare them with those of the ideal observer.

1.4 Data

1: In a test-run we show the subject a number of elements of X and ask him in each instance to make either decision D_A or D_B on the basis of the information he has accumulated in the previous training-runs.

2: We suppose each dimension X_p is divided into n_p discrete classes. This means that the data consist of an $n_1 \times n_2 \times \ldots \times n_p \times 2$ table in which the first p variables are factors corresponding with the linear dimensions of X and the last variable is a variate corresponding with the responses D_A and D_B.

3: The use of grouping makes the role of the multinormal distributions somewhat doubtful. Actually, in later experiments, we started with discrete classes and constructed discrete multidimensional probability distributions Π_A and Π_B which merely looked like multinormal distributions, and which had the property that the optimal strategy defined a hyperplane cut-off strategy of the type 1.2.3.

1.5 Model, hypotheses, techniques

1: Notation

From now on we suppose $1 \leq p \leq 3$. This implies that, for $p = 3$ for example, we can write p_{ijk} for $p(D_A | x)$ with $i=1,\dots,n$; $j=1,\dots,m$; $k=1,\dots,l$. Moreover z_{ijk} = logit p_{ijk}. Suppose x_{ijk} occurred N_{ijk} times in the test-run and the subject has made decision A in n_{ijk} out of the N_{ijk} cases.

2: Model

In order to analyze the data we can use the modified analysis of variance techniques outlined by Gabriel (1963), and (assuming repeated independent trials with constant probabilities p_{ijk}) also the full table logarithmic models of Birch (1963), Goodman (1970, 1971), or the split table models of Bishop (1969). We have chosen for the logit model of 'rates' which seems particularly appropriate here because of 1.2.3. (In fact I cannot think of any application in which it is more appropriate).

3: Decomposition

It is well known from the analysis of variance that we can write

(p=1): $z_i = \mu_i + \alpha$,

(p=2): $z_{ij} = s_{ij} + \mu_i + \lambda_j + \alpha$,

(p=3): $z_{ijk} = \delta_{ijk} + s_{ij} + t_{ik} + u_{jk} + \mu_i + \lambda_j + \gamma_k + \alpha$,

with

$$\sum_i \mu_i = \sum_j \lambda_j = \sum_k \gamma_k = 0,$$

$$\sum_i s_{ij} = \sum_i t_{ik} = \sum_j s_{ij} = \sum_j u_{jk} = \sum_k t_{ik} = \sum_k u_{jk} = 0,$$

$$\sum_i \delta_{ijk} = \sum_j \delta_{ijk} = \sum_k \delta_{ijk} = 0.$$

In the following table we have collected the linear dimension of the sets of parameters, i.e. the number of degrees of freedom we loose if we set these parameters equal to a set of known constants.

Subset	Dimension
α	1
μ	n-1
λ	m-1
γ	l-1
ξ	(n-1)(m-1)
t	(n-1)(l-1)
u	(m-1)(l-1)
δ	(n-1)(m-1)(l-1)

If we require, for example, $\delta_{ijk} = 0$ for p=3 we loose (n-1)(m-1)(l-1) dfr and nml - (n-1)(m-1)(l-1) free parameters remain to be fitted. If we require $\mu_i = 0$ and $\gamma = 1$ for p=2 we loose (n-1)+1 = n dfr and nm - n = n(m-1) free parameters are left.

4: Hypotheses

In terms of the decomposition of the previous section the most general hypotheses we are interested in is that a particular subset of the parameters lies in a linear subspace of dimension q, which is not larger than the maximum dimension given in the table. Thus we can require, for example, $\mu_i = \theta a_i + \varepsilon b_i$, where a and b are known vectors of real numbers (linearly independent, $\sum a_i = \sum b_i = 0$). In this case we loose (n-1) - 2 = n-3 dfr. In the more general case that we require that λ_j is a polynomial function of degree $\leq$ q of a given set of constants (with q $\leq$ m-1) we loose (m-1) - q dfr. It is obvious how to generalize this computation of the degrees of freedom to more general cases.

5: Technique

In order to estimate parameters and test hypotheses we can use the maximum likelihood techniques of Dyke and Patterson (1952) and the minimum logit chi-squared techniques of Berkson (1944, 1953, 1955, 1956, 1968). We have chosen for the minimum logit chi-squared because they are more simple computationally and probably more accurate in small samples (Berkson 1955, 1956, Odoroff 1970, cf however also Silverstone 1957). The theoretical work of Gart and Zweifel (1967) and the computations of Odoroff (1970) suggest that a good estimate of the logit is $\hat{z}_{ijk} = \log \bar{n}_{ijk}/\underline{n}_{ijk}$, with $\bar{n}_{ijk} = n_{ijk} + \frac{1}{2}$ and $\underline{n}_{ijk} = n_{i.k} - n_{ijk} + \frac{1}{2}$. A good estimate of the variance of the logit is $\hat{v}_{ijk} = 1/\underline{n}_{ijk} + 1/\bar{n}_{ijk}$. By 'good' we mean in this context that the small-sample bias is usually less than that of the obvious maximum likelihood estimate, while the asymptotic properties are the same. Consequently we must minimize

$$S = \sum\sum\sum \hat{w}_{ijk}(\hat{z}_{ijk} - z_{ijk})^2$$

over all free parameters of the model. If v is the number of degrees of freedom we have lost from the original nml ones, then we know that the limiting distribution of S_{min} is χ^2 with v degrees of freedom. We also know that if model 1 implies model 2 then $S^1_{min} \geq S^2_{min}$ and $v_1 \geq v_2$. Moreover $S_{12} = S^1_{min} - S^2_{min}$ is asymptotically distributed as $\chi^2(v_1 - v_2)$, and S_{12} is asymptotically independent of S^1_{min}. This analysis easily extends to systems of hypotheses which are partially ordered by implication.

2 Probabilistic concept learning (PCL): II

Truyens (1969),

Some generalizations of the work outlined in section 1 can be found in De Klerk, De Leeuw, & Oppe (1970), De Leeuw (1971), Oppe (in preparation).

2.1 Generalization

1: The first obvious generalization is to multiple decision problems with more than two hypotheses. In fact suppose that we are dealing with $H_1, H_2, \ldots, H_n$. For the posterior probabilities we can write

$$\pi(H_k \mid x) \propto \pi_k N(x; \mu_k, \Sigma).$$

The posterior expected pay-off is

$$P(D_k \mid x) = \sum_{l=1}^{n} \lambda_{kl} \pi(H_k \mid x).$$

Making the symmetry assumption that $\bar{\lambda} = \lambda_{kk} > \lambda_{kl} = \underline{\lambda}$ for all $k \neq l$

we find

$$p(D_k \mid x) = \pi(H_k \mid x) + [illegible](1 - \pi(H_k \mid x)),$$

which means that with 'regular' pay-off we must select the hypothesis with the highest posterior probability. If the prior probabilities are also equal this is, of course, equivalent to choosing the hypothesis which maximizes the likelihood.

2: The all-or-none and event-matching strategies are easily translated, the logistic strategy is based on

$$\log \frac{\pi(H_k \mid x)}{\pi(H_l \mid x)} = \hat{\delta}'_{kl} \Sigma^{-1}(x - \mu_{kl}) + \gamma_{kl},$$

and becomes

$$\log \frac{p(D_k \mid x)}{p(D_l \mid x)} = (v'_k - v_l)'x + (u_k - u_l).$$

Observe that more general logistic strategies are possible here like

$$\log \frac{p(D_k \mid x)}{p(D_l \mid x)} = v'_{kl}x + u_{kl},$$

with $v_{kl} = -v_{lk}$, $u_{kl} = -u_{lk}$. The difference is that in this last model additive cancellation conditions like

$$u_{kl} + u_{lm} = u_{km},$$

$$v_{kl} + v_{lm} = v_{km},$$

and no longer true for $k \neq m$.

3: An obviously equivalent model is

$$z(x;k) = \log p(D_k \mid x)/p(D_n \mid x) = v'_k x + u_k$$

for all $k=1,\ldots,n-1$. This means that our model is simply the conjunction of n-1 submodels of the type 1.3.5 with no further complications because the parameters are neatly separated.

4: The logistic analysis from section 1.5 can consequently be very easily applied. We use the generalized logistic transform

$$\hat{z}(x;k) = \log (2n(D_k \mid x) + 1)/(2n(D_n \mid x) + 1)$$

for all $k=1,\ldots,n-1$. This gives us n-1 tables with $\hat{z}$-values, on which n-1 separate logit analyses must be performed.

2 Generalization

1: A second, equally obvious, generalization is to take $N(x; \mu_A, \Sigma_A)$ and $N(x; \mu_B, \Sigma_B)$ as hypotheses, with $\Sigma_A \neq \Sigma_B$ in general. This makes the optimal cut-off strategy quadratic, and the corresponding logistic model consequently must allow for quadratic (and interaction) parameters. The complications which result are not very essential.

2: The next logical step is to abandon the multinormal distribution altogether. This makes the logistic model somewhat less natural, and suggests using the more general decomposition models for nominal data.

.3 Generalization

1: As a third generalization we do not ask the subject to make a decision as to what hypothesis is true, but we ask him: 'If this $x \in X$ occurs M times in a sequence of random independent trials, how many times will H_A be true ?' Our previous problem was the special case with $M = 1$. This generalization is an attempt to get more information about the $p(H_A \mid x)$.

2: In previous PCL experiments people simply asked: 'What is your posterior probability that H_A is true ?' Responses were treated as if they were estimates of the posterior probabilities on which, for example, the logistic transformation could be applied.. Compared with the procedure in the previous section this has a number of serious disadvantages, and can not be recommended any more.

3: In a series of M Bernouilli trials as described the probability of m successes is

$$\pi(m \mid x) = \binom{M}{m} \pi(H_A \mid x)^m \, \pi(H_B \mid x)^{M-m}.$$

The expected pay-off of decision D_1 is

$$P(D_1 \mid x) = \sum_{m=0}^{M} \lambda_{1m} \pi(m \mid x).$$

If $\bar{\lambda} = \lambda_{11} > \lambda_{1m} = \underline{\lambda}$ for all $1 \neq m$ then

$$P(D_1 \mid x) = \bar{\lambda}\, \pi(1 \mid x) + \underline{\lambda}(1 - \pi(1 \mid x)),$$

which is monotone with $\Pi(1 \mid x)$, the binomial expression.

4: The corresponding generalized logistic strategy becomes

$$p(D_m \mid x) = \binom{M}{m} p(D_A \mid x)^m \left(1 - p(D_A \mid x)\right)^{M-m},$$

with

$$\operatorname{logit} p(D_A \mid x) = v'x + u.$$

5: The data can be collected in an $n_1 \times n_2 \times \dots \times n_p \times (M+1)$ table with frequencies $n(D_m \mid x)$.

6: The first hypothesis of interest is

H_B: $p(D_m \mid x)$ is a binomial distribution for each $x \in X$.

Further hypotheses can be formulated within H_B on the structure of the parameters of these binomial distributions, which means that we are back in the situation discussed in section 1, and the binomial parameters now replace the posterior probabilities.

7: In the case of binomial PCL maximum likelihood seems to have some advantages over minimum logit chi-squared. We introduce the notation $p_m(x)$ for $p(D_m \mid x)$, $n_m(x)$ for $n(D_m \mid x)$, $p(x)$ for $p(D_A \mid x)$, $z(x)$ for logit $p(D_A \mid x)$, $n(x)$ for $\sum n_m(x)$, and $e(x)$ for $\sum n_m(x)m$. The hypothesis is that $z(x)$ is a particular linear function of the parameters, i.e. we can write

$$H: z(x) = \sum a_v(x)\, \theta_v,$$

with $a_v(x)$ a known set of constants. For the logarithm of the LF we find

$$\mathcal{L} = \sum_{x \in X} \sum_{m=0}^{M} n_m(x) \log p_m(x) =$$

$$= \sum_{x \in X} z(x)e(x) + M \sum_{x \in X} n(x) \log\left(1 - p(x)\right).$$

If H is true we find for the derivatives

$$\frac{\partial \mathcal{L}}{\partial \theta_v} = \sum_{x \in X} e(x)a_v(x) - M \sum_{x \in X} n(x)p(x)a_v(x),$$

$$\frac{\partial^2 \mathcal{L}}{\partial \theta_v \partial \theta_w} = -M \sum_{x \in X} n(x)a_v(x)a_w(x)p(x)(1-p(x)).$$

It follows that $\mathcal{L}$ is a concave function of the parameters, and that the usual Newton-Raphson-type working logit methods will be very efficient.

2: Define $\hat{p}(x) = \Sigma(x)/M$. An obvious consequence of the binomial hypothesis H_b is

$H_b \hat{p} : \hat{p}(x) = p(x)$.

It follows that a consequence of the hypothesis H of the previous section is

$\mathcal{E}\,\hat{p} : \text{logit } \hat{p}(x) = \sum a_v(x)\,\vartheta_v$.

Starting points for the iterative ML procedure of the previous section can be found by applying the usual minimum logit chi-squared methods for estimating $\hat{\vartheta}$ on the values of logit $\hat{p}(x)$, where $\hat{p}(x)$ is estimated as $\Sigma(x)/M$.

4 Generalization

1: Further generalizations result if we take samples of size > 1. Mathematically this is trivial, because a sample of size r is equivalent to a sample of size 1 from the r-fold cartesian product with the product distribution.

2: Psychologically the true generalization is taking a sample of size r sequentially while telling the subject that one of the hypotheses is true all the time. The idea is that the subject has to revise his posterior probabilities with each new element. Asymptotically this procedure becomes identical to the familiar bookbag & pokerchip experiments.

3 Functional learning (FL)

A discussion of FL-tasks can be found in Carroll (1963), Björkman (1965a, 1965b), De Klerk, De Leeuw, and Oppe (1966, 1968, 1970).

3.1 Situation

1: A stimulus space X, can be identified with a p-dimensional Euclidean space.

2: A finite set Y of decisions.

3: For each $x \in X$ a probability distribution $p_x(y)$ over Y.

4: A pay-off function λ on Y x Y.

.2 Relations with PCL

1: FL is a generalizations of PCL in the sense that the $p_x(y)$ are not supposed to be generated by applying Bayes' rule to likelihoods and prior probabilities any more, the $p_x(y)$ are given directly.

2: Again $p_x(y)$ must be interpreted as the probability that y is true (reinforced) given x.

3: The statistical analysis of sections 1 and 2 again is identical if we substitute the $p_x(y)$ for the posterior probabilities.

4: We talk about functional learning only if the probability distributions $p_x(y)$ vary smoothly with x, for example; the expected values $e_x(y)$ are a low-degree polynomial function of x, the variances are constant (maybe even zero). Or: the expected values are constant, the variances increase slowly with (the real numbers) x. Previous attempts to define a clear boundary between functional learning experiments and other experiments satisfying 3.1 have failed rather miserably. A recent proposal is to call all experiments in class 3.1 <u>decision learning</u> (DL) experiments, and to speak of functional learning only if we have tried to make the relationship between x and $p_x(y)$ smooth enough. An experiment is a functional learning experiment only if we have designed it as such.

4 Decision learning (DL)

A discussion of DL-tasks is given in De Leeuw (1971). General approaches for constructing models and analytic techniques are also descüssed there. The paper restricts itself to the case where both X and Y are finite, but this causes no real loss of generality.

[illegible]

In stead of 2.3.7 and 2.3.8 the following sections seem more appropriate.

7: We introduce the notation $p_m(x)$ for $p(D_m \mid x)$, $n_m(x)$ for $n(D_m \mid x)$, $p(x)$ for $p(D_A \mid x)$, $z(x)$ for logit $p(D_A|x)$, $n(x)$ for $\sum n_m(x)$, and $\tilde{p}(x)$ for $\sum m\, p_m(x) / M$, and $\hat{p}_m(x)$ for $n_m(x)/n(x)$. The hypothesis H_B is equivalent to

$$H_B:\ q_m(x) = \log \frac{p_m(x)}{p_{m+1}(x)} = t(m) + z(x)$$

for all $m=0,\ldots,M-1$ and for all $x \in X$. Here

$$t(m) = \log \binom{M}{m} - \log \binom{M}{m+1}.$$

If we consider the $t(m)$ as a set of free parameters, the weaker hypothesis H_P specifies merely the additivity of the matrix Q. It is equivalent to

$$H_P:\ p_m(x) = A(x)B(m)\,[\theta(x)]^m,$$

i.e. to the hypothesis that $p_m(x)$ is a generalized power series distribution with finite range.

8: For each $x \in X$ the $\hat{q}_m(x)$ are asymptotically normal with dispersion matrix $S(x)$ defined by

$$S^{(x)}_{m\ell} = \begin{cases} (n(x))^{-1}\,(p_m(x))^{-1} + (n(x))^{-1}\,(p_\ell(x))^{-1}, & (m = \ell) \\ -(n(x))^{-1}\,(p_\ell(x))^{-1}, & (m = \ell - 1) \\ -(n(x))^{-1}\,(p_m(x))^{-1}, & (m = \ell + 1) \\ 0. & (\text{otherwise}) \end{cases}$$

This is a tridiagonal matrix, which is easy to invert. Its (generalized) inverse is written as $T(x)$. The modified chi-square measure we use is

$$S = \sum_{x \in X} (\hat{q}(x) - \underline{z}(x) - t)'\ \hat{T}(x)\ (\hat{q}(x) - \underline{z}(x) - t).$$

Here $\hat{T}(x)$ is $T(x)$ with $(n_m(x) + \frac{1}{2})^{-1}$ substituted for $(n(x))^{-1}(p_m(x))^{-1}$, $\hat{q}(x)$ is the M-vector with elements $\log(n_m(x) + \frac{1}{2}) - \log(n_{m+1}(x) + \frac{1}{2})$, $\underline{z}(x)$ is the M-vector with all elements equal to $z(x)$. If we want to test H_P we must minimize S over z and t, and compare S_{min} with $\chi^2(n-1)(M-1)$. If we set t equal to the known constants from the previous section and minimize over z we have a test of H_B with $n(M-1)$ dfr. Linear structural hypotheses of the form discussed in

4.5 can again be treated by minimizing over the free parameters of the model.

5. An obvious consequence of the hypothesis H_B is

H_E: $\tilde{p}(x) = p(x)$,

or

H_E: logit $\tilde{p}(x) =$ logit $p(x)$.

Structural hypotheses about the logits can be tested within H_E by the usual minimum logit chi squared methods, in which we estimate $p(x)$ by $\hat{p}(x) = \sum m \, \hat{p}_m(x) \,/\, M$ and the variance of the logit by

$$\hat{v}(x) = \frac{\sum \hat{p}_m(x)\, m^2 - (\sum \hat{p}_m(x)\, m)^2}{n(x)\left[M \, \hat{p}(x)(1-\hat{p}(x))\right]^2} .$$

References

Berkson, J. Application of the logistic function to bioassay
JASA 39, 1944, 357-365

— A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based on the logistic function
JASA 48, 1953, 565-599

— Maximum likelihood and minimum X^2 estimates of the logistic function
JASA 50, 1955, 130-162

— Estimation by least squares and maximum likelihood
Proc. Third Berkeley Symp. I, 1956, 1-11

— Application of minimum logit X^2 estimate to a problem of Grizzle with a notation on the problem of no interaction
Biometrics, 24, 1968, 75-97

Birch, M.W. Maximum likelihood in three-way contingency tables
JRSS B, 25, 220-233

Bishop, Y.M.M. Full contingency tables, logits, and split contingency tables
Biometrics, 25, 1969, 383-400

Björkman, M. Studies in prediction behaviour: explorations into predictive judgments based on functional learning and defined by estimation, categorization, and choice
Scand. J. Psychol., 6, 1965, 129-156

— Learning of linear functions: a comparison between a positive and a negative slope
Rep. Psychol. Lab., Univ. of Stockholm, No 183, 1965

De Klerk, L.F.W. de Probabilistic concept learning
Voorschoten, VAM, 1968

—, de Leeuw, J., and Oppe, S Funktioneel leren
Hypothese, 11, 1966, 10-19

— Functional learning
Report E 020-68, Psychol. Institute, Univ. Leiden, 1968

— Functional learning
Report E 024-70, Psychol. Institute, Univ. Leiden, 1970

De Klerk, L.F.W., and Oppe, S. Probabilistic concept learning
Report E 013-66, Psychol. Institute, Univ. Leiden, 1966

De Leeuw, J. Notes on the definition of a concept
Report RN 001-68, Psychol. Institute, Univ. Leiden, 1968

— Probabilistic concept learning
Report RN 002-68, Psychol. Institute, Univ. Leiden, 1968

— The influence of a cut-off strtegy on the alpha-prime measure
Report RN 003-68, Psychol. Institute, Univ. Leiden, 1968

— Likelihood ratio tests for probabilistic concept learning
Report RN 002-69, Psychol. Institute, Univ. Leiden, 1969

— Analyse van nominale gegevens
Hypothese, 16, 1971, in press

— Finite decision learning experiments
Mimeographed paper, Univ. Leiden, 1971

Carroll, J.D. Functional learning
Report RB 63-26, ETS, Princeton, 1963

Dyke, G.V. & Patterson, H.D. Analysis of factorial arrangements when the data are proportions
Biometrics, 8, 1952, 1-12

Gabriel, K.R. Analysis of variance of proportions with unequal frequencies
JASA 58, 1963, 1133-1157

Gart, J.J. & Zweifel, J.R. On the bias of various estimators of the logit and its variance with applications to quantal bioassay
Biometrika, 54, 1967, 181-187

Goodman, L.A. The multivariate analysis of qualitative data
JASA 65, 1970, 226-256

— The analysis of multidimensional contingency tables
Technometrics, 13, 33-61

Lee, W. Choosing among confusably distributed stimuli with specified likelihood ratios
Perc. Motor Skills, 6, 1963, 445-467

Lee, W. Conditioning parameter model for reinforcement generalization in probabilistic discrimination learning
J. Math. Psychol., 3, 1966, 184-196

Lee, W. & Janke, M. Categorizing externally distributed stimulus samples for three continua
J. Exp. Psychol., 68, 1964, 376-382

Odoroff, C.L. A comparison of minimum logit chi-square estimation and maximum likelihood estimation
JASA 65, 1970, 1617-1631

Silverstone, H. Estimating the logistic curve
JASA 52, 1957, 567-577

Truyens, C.L. Probabilistic concept learning with a curved concept boundary
Unpublished Master's thesis, Univ. of Leiden, 1969