Chapter 5

Optimal Designs for Latent Variable Models: A Review

Martijn P.F. Berger

University of Maastricht
The Netherlands

The application of statistical models is often hampered by problems associated with the estimation of their parameters, and in some cases the data simply do not contain enough information to obtain efficient estimates of the parameters. Especially in latent variable models this problem may arise. Extremly large samples are often needed to provide enough information for efficient parameter estimation. Then it may be worthwhile to search for an optimum design of the data that maximizes the information with a minimum sample size, i.e., with a minimum cost. The purpose of this paper is to review the problems of finding optimal designs for latent variable models and to place these problems within the general statistical theory of optimal design. The two distinct optimal design problems for latent variable models will be discussed in detail.

A great amount of research has been devoted to the optimization of experimental designs in statistical literature. The corresponding statistical theory is referred to as *optimal design theory*. Extensive reviews of the developments on optimal designs are given by Atkinson (1982), Silvey (1980), and Steinberg and Hunter (1984), among others.

In general, an optimal design will depend on the assumed model with its parameters and on the selected optimization criterion. An optimal design for a linear (regression) model may not be optimal for a nonlinear (exponential) model. Although, at first, the focus in optimal design research was on linear models, later developments were centered more around nonlinear models. The problem of finding an optimal design for a regression model is relatively easy. If, for example, we can assume that data are adequately described by a linear regression function, then the problem of finding an optimal design is actually a sampling problem: How should we sample the different levels of the independent variable to obtain the most efficient estimators for the regression parameters. The answer to this problem can be found by straightforward maximization of some function of Fisher's information matrix for the regression parameters.

For nonlinear models this problem remains essentially the same. However, nonlinear models have an additional problem. Fisher's information is a function of the derivatives of the likelihood. Since the likelihood for nonlinear models is nonlinear in the parameters, the derivatives are also functions of the parameters, and optimality will thus depend on the parameter values. This means, that the maximum of the optimality criterion based on Fisher's information may differ for each combination of parameter values, and that an optimal design for a nonlinear model will usually only be optimal locally, i.e., for a given set of parameter values. Because these parameters are usually unknown, no straightforward solution will be possible. Of course, there are a number of ways to circumvent this problem. Three strategies have been studied in the literature, the imputation of prior estimates for the parameters, Bayes methods, and sequential methods.

Latent variable models are nonlinear in the parameters, and therefore the above described problem will also occur when latent variable models are used to analyze the data, and the same strategies to circumvent this problem can be applied. However, in latent variable models the persons are usually characterized by one or more latent dimensions or variables, and latent variable models thus include additional parameters to describe them. In contrast with ordinary linear models the independent variables in latent variable models are represented by (incidental) parameters, and because these are unknown, the selection of the levels of the latent variables cannot be done exactly.

Two different optimal design problems can be distinguished in latent variable models. The first problem is centered around the optimal selection of the levels for the independent (latent) variable, and is concerned with the question of how to select a sample of persons with a latent variable distribution to obtain efficient estimates for the structural variables in the model. This is often referred to as the *optimal sampling design problem*. The second problem is the *optimal test design problem*, where an optimal set of parameters which describe the structural variables is searched for to obtain efficient estimates for the parameters describing the latent variable distribution. It should be emphasized that for both problems optimality will depend on the unknown parameter values.

In the next paragraphs, some practical situations in which optimal designs for latent variable models can be applied will be described briefly, and notation will be introduced. Then the optimal design problem will be reviewed.

1. Applications of Optimal Designs

In educational and psychological measurement, the interest in optimal test and sampling designs started with the work of Lord (1962) and Pandey and Carlson (1976), who used so-called multiple matrix sampling designs to increase the efficiency of the mean performance of a population for a certain item domain. This work was centered around classical test theory, and has been applied to large scale assessment programs in the United States, such as NAEP and CAP (Bock et al., 1982).

As latent trait models gained currency in educational testing, and large scale assessment studies were performed more frequently, the need for more efficient test designs became urgent. If it were possible to replace a whole test by a small but optimally selected set of items, and still obtain efficient estimates for the abilities of the a sample of persons, then a lot of classroom administration time could be saved. Studies on efficient test designs for latent trait models were done by Adema (1990), Berger (1994b), Boekkooi-Timminga (1989), Theunissen (1985), and van der Linden and Boekkooi-Timminga (1989), among others, by using mathematical programming. An alternative way of increasing the precision of ability estimates is to design an optimal test during the administration of that test. An example of such a process is adaptive testing, where items that optimally match the provisional ability estimate are selected during test administration. The selection of items in an adaptive test is an optimal test design problem. Wainer (1990) gives a review of the problems in designing adaptive tests.

With the introduction of item banks and computerized testing, large amounts of calibrated items became necessary. If it were possible to select optimal samples of persons to be used for item estimation, then testing time and the cost associated with the calibration of items could be reduced considerably. A lot of research has been conducted in this area (cf. Berger (1991, 1992, 1994a), Berger and van der Linden (1992), Jones and Jin (1994), Stocking (1990), Thissen and Wainer (1982), and Wingersky and Lord (1985), among others).

The research described above on optimal sampling designs and test designs has been mainly centered around achievement testing and latent trait models. However, other fields of research, such as survey research, attitude measurement, panel studies, consumer research, and epidemiological research, where latent class models are often used to analyze the polytomous data, can also benefit from optimal design research. Although not much research has yet been done on optimal designs for latent class models, the main results on latent trait models will also be expected to hold for these models due to the similarities between latent trait models and latent class models.

2. Optimal Design Notation

Suppose we have a set of responses $X = \{x_{vi}\}$ from N persons $(v = 1, \ldots, N)$ to k structural variables $(i = 1, \ldots, k)$, and let the probability of obtaining these responses be given by the function

$$p(X_{vi} = x) = p(\theta_v, \xi_i) \;, \tag{1}$$

where x is a dichotomous or a polytomous response, and θ_v is an element from a vector $\theta' = (\theta_1, \theta_2, \theta_3, \ldots, \theta_c)$,characterizing the latent variable. Corresponding with each value θ_v is a weight w_v which is an element of the vector $W' = (w_1, w_2, w_3, \ldots w_c)$ The weights in W can be defined in different ways. Veerkamp and Berger (1994), for example, proposed some special weight functions for adaptive testing.

In this paper it will be assumed that the weights in W characterize the shape of the latent variable distribution, i.e., it will be assumed that $w_j \geq 0$, $\sum_j w_j = N$, and $1 \leq c \leq N$. This means that any change of weights will change the shape of the latent variable distribution. When, for example, all weights w_v are equal, then the latent variable distribution will have a uniform shape. This notation also makes it possible to consider only a part of the latent variable distribution by setting some weights equal to zero.

The vector $\xi_i' = (\xi_{1,i}, \xi_{2,i}, \xi_{3,i}, \ldots, \xi_{p,i})$ contains the parameters characterizing the structural variable i. The number of parameters p will, of course depend on the chosen latent variable model. For the well-known Rasch model p will be $p = 1$, while $p = 2$ for the two-parameter logistic model. For the two-parameter logistic model described in the first chapter, paragraph 1.3, the vector ξ_i' will have elements $\xi_i' = (\beta_i, \sigma_i)$.

The mean and variance of the parametric family are $E\{X_{vi}\} = p(X_{vi} = x)$ and $Var\{X_{vi}\} = p(X_{vi} = x)\,[1 - p(X_{vi} = x)]$, respectively, and under the assumption of local independence the likelihood associated with the set of responses X and the vectors θ and ξ is given by:

$$L\left(X|\theta,\xi\right)=\prod_{v=1}^{c}\prod_{i=1}^{k}\ p\left(X_{vi}=x\right)^{w_v p_{iv}}\left[1-p\left(X_{vi}=x\right)\right]^{w_v\left(1-p_{iv}\right)}, \tag{2}$$

where p_{iv} is the proportion of correct responses in each of the c categories of vector θ.

A measure for the efficiency of parameter estimators is Fisher's information (Kendall & Stuart, 1973, p. 10), and for small samples and small numbers of structural variables Fisher's information is an upper bound for the actual information on the parameters. Fisher's information $J(\theta_v)$ on the parameters θ_v of the latent variable distribution {θ,W} can be grouped into a diagonal matrix:

$$J\left(\theta|\xi\right)=\text{Diag}\left\{J\left(\theta_1\right),J\left(\theta_2\right),\ldots,J\left(\theta_c\right)\right\}, \tag{3}$$

and the information on the (structural) parameters ξ_i can be grouped in a super-diagonal matrix M(ξ|θ,W) with main diagonal matrices:

$$M\left(\xi_i|\theta,W\right)=\sum_{v=1}^{c}\ w_v\ G\left(X_v\right)YY',$$

where (4)

$$G\left(X_v\right)=\frac{\left[p'\left(X_{vi}=x\right)\right]^2}{p\left(X_{vi}=x\right)\left[1-p\left(X_{vi}=x\right)\right]},$$

with $p'(X_{vi} = x)$ being the first derivative of the response function. Vector Y contains parameters of the latent variable model. For example, for the two-parameter logistic model vector Y' contains $Y' = [-(\theta_v - \sigma_i), \beta_i]$. The information on all the parameters of a latent variable model can be grouped into the total partitioned information matrix:

$$Inf=\begin{bmatrix} J\left(\theta\,|\,\xi\right) & \text{joint } J(.)\,M(.) \\ \text{joint } M(.)\,J(.) & M\left(\xi\,|\,\theta,W\right)\end{bmatrix}, \tag{5}$$

where joint J(.)M(.) stands for the joint information on the parameters θ and ξ.

Since the information matrix *Inf* contains all the information on the parameters of the latent variable model, a design that produces the most efficient estimators of the parameters should have the most information in *Inf*. Roughly speaking, the more information in *Inf*, the more optimal the design will become.

Finally, it should be noted that some parts of *Inf* may contain redundant information, and that *Inf* may not be of full rank. In that case, some of the parameters in the model will have to be fixed.

3. The Optimization Problem

Several functions of the information matrix *Inf* have been studied in the literature on optimal designs. These functions are all a member of a class of functions Φ(.) having certain properties within an approximate equivalence theory (Kiefer, 1974). Each of these functions may have advantages in different situations (cf., for example, Kiefer (1959) and Federov (1972) for more details).

The best known criterion is the D-optimality or determinant criterion. This criterion maximizes the determinant of Fisher's information matrix on the parameters in the model, and has some nice properties. Not only does this criterion use all relevant information, it is also invariant under linear transformation of the parameter scale, and has some useful upper bounds (Khan & Yazdi, 1988). These are perhaps the reasons why the D-optimality criterion has been applied so often.

Another well-known criterion is the A-optimality or trace criterion. The trace criterion maximizes the sum of the diagonal elements of Fisher's information matrix on the parameters. This means that not all available information is used, and that the criterion will depend on the scale of the independent variable.

A third criterion is the so-called E-optimality criterion. This criterion maximizes the smallest root of Fisher's information matrix. This also means that the E-optimality criterion does not implement all available information on the parameters. An alternative criterion, which is often referred to as a weighted E-optimality criterion, is the MAXIMIN criterion. This criterion actually uses the weights w_j in the maximization procedure. The MAXIMIN strategy, which can be seem as optimizing the worst possible value, is described by Silvey (1980, p. 59), and was utilized by van der Linden and Boekkooi-Timminga (1989) for optimal test design.

Depending on the design problem, the function $\Phi(.)$ can be defined for different parts of the information matrix *Inf*. If the main interest is on finding an optimal sample for the estimation of test items, i.e., when an optimal sampling design problem is encountered, then $\Phi(.)$ should be based on the super-diagonal matrix $M(\xi|\theta,W)$. On the other hand, when the main interest is on selecting a small but optimal set of items for the efficient estimation of abilities of a sample of persons with an ability distribution $\{\theta,W\}$ then the optimality criterion $\Phi(.)$ should be based on the diagonal matrix $J(\theta|\xi)$.

It is be clear that the problem of finding an optimal sampling design or an optimal test design for latent variable models is actually the problem of maximizing the function$\Phi(.)$, under certain constraints. Different (numerical) optimization procedures can be used. For the A-optimality and E-optimality criterion relatively easy linear (programming) procedures may be applied, while for the D-optimality criterion a nonlinear approach will be necessary. Algorithmic aspects were studied by Wu (1978) and Böhning (1986). See Cook and Nachtsheim (1980) and Yonchev (1988) for a comparison of the different algorithms for D-optimal designs.

The following definitions for an optimal sampling design and an optimal test design can now be given:

Locally Optimal Test Design: A test design with a set of vectors with parameters for k items or structural variables $\xi^{*'} = (\xi_1^*, \xi_2^*, \xi_3^*, \ldots, \xi_k^*)$ is locally Φ -optimal if:

$$\Phi\left\{J\left(\theta \middle| \xi^*\right), W\right\} \geq \Phi\left\{J\left(\theta \middle| \xi\right), W\right\}, \tag{6}$$

for a given pair of vectors $\{\theta,W\}$ where with weights W and any $\xi' = (\xi_1,\xi_2,\xi_3,\ldots\xi_k)$.

Locally Optimal Sampling Design: A sampling design with pair $\{\theta^*,W^*\}$ where $\theta^* = (\theta_1^*, \theta_2^*, \ldots, \theta_c^*)$ and $W^* = (w_1^*, w_2^*, \ldots, w_c^*)$, is locally Φ -optimal if:

$$\Phi\left\{M\left(\xi \mid \theta^*, W^*\right)\right\} \geq \Phi\left\{M\left(\xi \mid \theta, W\right)\right\} \tag{7}$$

for a given set of parameters ξ and any θ.

Constraints

Maximization of the function $\Phi(.)$ can be performed with or without certain practical constraints. Test constructors often have to take into account certain requirements for the items that are included in an achievement test. These requirements often have to do with the content of the items. In many situations, the items in an achievement test are required to cover the whole content area of the course. Sometimes it will be necessary to include at least one item on every chapter of a book, and in other cases the items in a test may use the same common subject text. These requirements will, of course, influence the optimization procedure. In linear programming procedures these requirements can easily be implemented. A complete list of such requirements on administration time, test composition, and on inclusion or exclusion of related items has been given by van der Linden and Boekkooi-Timminga (1989).

Various different kinds of achievement tests have been used in educational measurement. Apart from the standard fixed-form test, where all persons take the same set of items, adaptive tests are used to obtain more efficient estimates of the abilities, with a limited, but optimally selected set of items. Another form, which has become more popular in the last few years, is the so-called testlet. Instead of administering single items in each step, as is done in adaptive testing, multi-item bundles with a predetermined path that a person may follow, were proposed by Wainer and Kiely (1987). Testlets more or less combine the advantages of adaptive testing with the advantages of the standard fixed-form test. Berger (1994b) shows that the procedures from optimal design research can be used to find optimal solutions for each of these test forms by implementing special constraints.

4. Three Approaches

As mentioned in the introduction of this chapter, an optimal design for a latent variable model will depend on the unknown parameters. Several approaches to circumvent this problem have been proposed. In the next sections three procedures will briefly be described.

Sequential procedures

The problem that an optimum design for latent variable models will depend on the unknown parameters can be circumvented by applying sequential optimization and estimation methods. Sequential optimal design procedures have been proposed by Ford, Titterington, and Wu (1985) and Ford, Kitsos, and Titterington (1989). Iterative search procedures for optimal designs where studied by Wynn (1970) and Wu and Wynn (1978).

Sequential procedures can be applied fully sequential, i.e., single experimental units or levels of the independent variable are added in each step, in such a way that the optimality criterion $\Phi(.)$ is optimized in every step. Such a procedure, however, may be very time consuming. Especially when samples are selected to obtain efficient estimates for item parameters, it will often be much faster to include a whole batch of persons in each step. Such a procedure is often referred to as a multi-stage or batch-sequential design procedure.

Berger (1994a) utilizes both fully- and batch-sequential optimization of D-optimal sampling designs for item parameter estimation, and showed that batch-sequential designs appeared to produce consistent estimates of the item parameters. In the context of item parameter estimation a two-stage design, which is a batch design with the total sample is divided into two batches, was described by Berger (1992). An alternative sequential approach to D-optimal sampling design, which takes into account the estimation errors of the latent trait variable, was presented by Jones and Jin (1994).

Bayesian procedures

Another way to circumvent the problem that an optimal design for latent variable models will depend on the unknown parameter values is to use Bayesian methods. When suitable priors are assigned to the parameters, they can then be eliminated by taking expectations. This was done by Zacks (1977) and Tsutakawa (1980) by pre-posterior analysis. Prior information may be assigned subjectively or empirically. When, for example, the form of the ability distribution $\{\theta,W\}$ can be estimated from previous studies, then the distribution can be implemented as a so-called empirical prior. And, of course, empirical priors can be updated sequentially. In a so-called empirical Bayes approach, the values for the independent variable are selected by maximizing the expected information on the parameters with respect to the posterior distribution. Such an empirical Bayes approach to the calibration of items by means of the Rasch model was proposed by van der Linden and Eggen (1986).

Imputation of parameter estimates

In some cases good approximations of the parameter values may be obtained from previous studies. In other cases the range of the parameter values may be limited. In the two-parameter logistic model, for example, the range of the item difficulty parameters σ_i and the ability parameter θ_j is usually limited to -3 and +3. In this case it will often be possible to obtain a good overall picture of the efficiency by considering a limited number of combinations of the parameter values within that range. Examples of such an approach are the studies on the standard errors of sampling designs for latent trait models presented by Thissen and Wainer (1982) and Berger (1991).

5. Conclusion

The application of optimal design theory to educational testing has become of increasing importance in the educational measurement and psychometric literature. Because efficient estimation of the parameters in latent variable models usually requires large amounts of data, these models are ideal candidates for the application of optimal design theory. The optimal designing of samples and tests will increase information on the parameters at a minimum of cost. Although up till now most results have been obtained for latent trait models for the analysis of dichotomous response data, it may be predicted that in future research more results will be obtained for latent class models and polytomous responses.

References

Adema, J.J. (1990). *Models and algorithms for the construction of achievement tests*. Ph.D. thesis, Enschede: University of Twente.

Atkinson, A.C. (1982). Developments in the design of experiments. *International Statistical Review, 50*, 161-177.

Berger, M.P.F. (1991). On the efficiency of IRT models when applied to different sampling designs. *Applied Psychological Measurement, 15*, 293-306.

Berger, M.P.F. (1992). Sequential sampling designs for the two-parameter item response theory model, *Psychometrika, 57*, 521-538.

Berger, M.P.F. (1994a). D-optimal sequential sampling designs for item response theory models. *Journal of Educational Statistics, 19*, 43-56.

Berger, M.P.F. (1994b). A general approach to algorithmic design of fixed-form tests, adaptive tests and testlets. *Applied Psychological Measurement, 18*, 141-153.

Berger, M.P.F., & van der Linden, W.J. (1992). Optimality of sampling designs in item response theory models. In: M. Wilson (Ed.), *Objective measurement: Theory into practice* (Vol 1). Norwood, N.J.:Ablex.

Bock, R.D., Mislevy, R.J., & Woodson, C.E. (1982). The next stage in educational assessment. *Educational Researcher, 11*, 4-11.

Boekkooi-Timminga, E. (1989). Models for computerized test construction. Ph.D. thesis, Enschede: de Lier, Academisch Boeken Centrum, University of Twente.

Böhning, D. (1986). A vertex exchange method in D-optimal design theory. *Metrika, 33*, 337-347.

Cook, R.D., & Nachtsheim, C.J. (1980). A comparison of algorithms for constructing exact D-optimal designs. *Technometrics, 22*, 315-324.

Federov, V.V. (1972). *Theory of optimal experiments*. New York: Academic Press.

Ford, I., Kitsos, C.P., and Titterington, D.M. (1989). Recent developments in nonlinear experimental design. *Technometrics, 31*, 49-60.

Ford, I., Titterington, D.M., and Wu, C.F.J. (1985). Inference and sequential design. *Biometrika, 72*, 545-551.

Jones, D.H. & Jin, Z. (1994). Optimal sequential designs for on-line item estimation. *Psychometrika, 59*, 59-75.

Kendall, M.G., & Stuart, A. (1973). *The advanced theory of statistics* (Vol. 2). New York: Hafner.

Khan, M.K., & Yazdi, A.A. (1988). D-optimal designs. *Journal of Statistical Planning and Inference, 18*, 83-91.

Kiefer, J. (1959). Optimum experimental designs (with discussion). *Journal of the Royal Statistical Society, Series B, 21*, 271-319.

Kiefer, J. (1974). General equivalence theory for optimum designs (Approximate theory). *The Annals of Statistics, 2*, 849-879.

Lord, F.M. (1962). Estimating norms by item-sampling. *Educational and Psychological Measurement, 22*, 259-267.

Pandey, T.N. & Carlson, D. (1976). Assessing payoffs in the estimation of the mean using multiple matrix sampling designs. In D.N.M. de Gruijter & L.J. van der Kamp (Eds.) *Advances in psychological and educational measurement*. London: Wiley.

Silvey, S.D. (1980). *Optimal design*. London: Chapman and Hall.

Steinberg, D.M. & Hunter, W.G. (1984). Experimental design: Review and comment. *Technometrics, 26*, 71-130.

Stokking, M. (1990). Specifying optimum examinees for item parameter estimation in item response theory. *Psychometrika, 55*, 461-475.

Theunissen, T.J.J.M. (1985). Binary programming and test design. *Psychometrika, 50*, 411-420.

Thissen, D. & Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika*, *47,* 397-412.

Tsutakawa, R.K. (1980). Selection of dose levels for estimating a percentage point of a logistic quantal response curve. *Applied Statistics*, *29*, 25-33.

Veerkamp, W.J.J., & Berger, M.P.F. (1994) *Some new item selection criteria for adaptive testing.* Research Report, 94-06. Enschede: University of Twente, the Netherlands.

van der Linden, W.J., & Boekkooi-Timminga, E. (1989). A maximin model for test design with practical constraints. *Psychometrika*, *53*, 237-247.

van der Linden, W.J., & Eggen, T.J.J.M. (1986). An empirical Bayes approach to item banking. *Applied Psychological Measurement, 10*, 345-354.

Wainer, H. (1990). *Computerized adaptive testing: A primer*. Hilsdale, NJ:Lawrence Erlbaum.

Wainer, H. & Kiely, G.L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement*, *24*, 185-202.

Wingersky, M.S., & Lord, F.M. (1985). An investigation of methods for reducing sampling error in certain IRT procedures. *Applied Psychological Measurement*, *8*, 347-364.

Wu, C.F. & Wynn, H.P. (1978). The convergence of general step-length algorithms for regular optimum design criteria. *The Annals of Statistics*, *6*, 1273-1285.

Wu, C.F. (1978). Some algorithmic aspects of the theory of optimal designs. *The Annals of Statistics*, *6*, 1286-1301.

Wynn, H.P. (1970). The sequential generation of D-optimum experimental designs. *Annals of Mathematical Statistics*, *41*, 1655-1664.

Yonchev, H. (1988). New computer procedures for constructing D-optimal designs. In Y.Dodge, V.V. Federov, & H.P. Wynn (Eds.) Optimal design and analysis of experiments (pp. 71-81). New York, NY: Elsevier Science Publishing Company, Inc.

Zacks, S. (1977). Problems and approaches in design of experiments for estimation and testing in non-linear models. In P.R. Krishnaiah (Ed.), *Multivariate analysis IV* (pp.209-223) Amsterdam: North-Holland.