Research Report

Adaptive Numerical Integration for Item Response Theory

Tamás Antal

Andreas Oranje

Research & Development

March 2007
RR-07-06

Adaptive Numerical Integration for Item Response Theory

Tamás Antal and Andreas Oranje
ETS, Princeton, NJ

March 2007

As part of its educational and social mission and in fulfilling the organization's nonprofit charter and bylaws, ETS has and continues to learn from and also to lead research that furthers educational and measurement research to advance quality and equity in education and assessment for all users of the organization's products and services.

ETS Research Reports provide preliminary and limited dissemination of ETS research prior to publication. To obtain a PDF or a print copy of a report, please visit:

http://www.ets.org/research/contact.html

Abstract

Well-known numerical integration methods are applied to item response theory (IRT) with special emphasis on the estimation of the latent regression model of NAEP. An argument is made that the Gauss-Hermite rule enhanced with Cholesky decomposition and normal approximation of the response likelihood is a fast, precise, and reliable alternative for the numerical integration in NAEP and in IRT in general.

Key words: Gauss-Hermite quadrature, adaptive integration, latent regression, item reponse theory, EM algorithm

1 Introduction

Marginal item response theory (IRT) inherently involves integration. The goal in this study is to obtain fast and accurate computation of integrals appearing in any normal population IRT model. The log-likelihood of these models takes the form

$$L = \sum_{i=1}^{N} \log \int_{\mathbb{R}^K} P(y_i \mid \theta, \beta) \varphi(\theta; \mu, \Sigma) \mathrm{d}^K \theta, \tag{1}$$

where

$$L_i(\theta) := P(y_i \mid \theta, \beta) = \prod_{j=1}^{J} P(y_{ij} | \theta, \beta_j) \tag{2}$$

is the likelihood of the response y_i of student $i = 1, \ldots, N$ given the ability θ and item parameters β_j for item $j = i, \ldots, J$. Also, y assumes the value of 1 for a correct response and 0 otherwise. Moreover, the population distribution is multivariate normal by our main underlying assumption:

$$\varphi(\theta; \mu, \Sigma) = \frac{1}{(2\pi)^{k/2} \sqrt{\det(\Sigma)}} e^{-\frac{1}{2} \langle \theta - \mu \mid \Sigma^{-1} \mid \theta - \mu \rangle}. \tag{3}$$

Here, μ is the common population mean and Σ is the common population covariance of the subscales.

The estimation of this model will include, independently of the estimation method chosen, the computation of integrals in the form

$$\mathcal{E}(g) = \int_{\mathbb{R}^K} g(\theta) P(y_i \mid \theta, \beta) \varphi(\theta; \mu, \Sigma) \mathrm{d}^K \theta, \tag{4}$$

where g is any smooth function.

Since the integrand has a special Gaussian factor, Gauss-Hermite methods for numerical integration lend themselves naturally for the computation of the integral in (4). The applicability of these methods are actively investigated and well-documented (Genz & Kass, 1998); in statistical practice, however, they are seldom used (Genz & Kass, 1997).

The most frequently used method for numerical integration—even in widely used IRT programs such as Parscale (Muraki & Bock, 1999)—employs a trivial rectangle rule with a fixed number of quadrature points over a fixed ability range. Apart from being the least sophisticated numerical integration tool, there are several obvious drawbacks to the usual implementations of the method. Often, the number of quadrature points and the range of the integration is fixed. This, in general, prohibits evaluation of the convergence of the numerical integral. In addition, the

sampling of the function based on a fixed grid can become insufficient when the *essential mass* of the function is concentrated over a region that is commensurate with the size of the grid. A region of essential mass is a set E so that the difference between $\int_{\mathbb{R}} f(x)\mathrm{d}x$ and $\int_E f(x)\mathrm{d}x$ is negligible.[1] The choice of the integration domain can easily jeopardize computational precision as well. If a sizable nonzero mass of the function lies outside the area of integration, the rectangle rule will loose significant contribution. In IRT estimations, it is not always transparent how to set up the parameters of the numerical integration to avoid the above mentioned problems. By performing a sequence of runs on the same data sets with different integration parameters, one may gain a good understanding of the efficiency of the numerical integration. This time-consuming process is seldom applied in practice, however. Many testing programs run on a tight time line, and they rarely find the time to do dozens of runs just to confirm that the numerical integration is adequate.

Even a well-respected testing program such as the National Assessment of Educational Progress (NAEP) carries the burden of the rectangle rule (Allen, Donoghue, & Schoeps, 2001). To complicate matters further, in NAEP it is not unusual to have $K = 5$ subscales, resulting in five-dimensional numerical integrals. With the midpoint rule, the computations quickly become unfeasible. The current practice is to apply the rectangle rule only in Dimensions 1 and 2 and to resort to Laplace approximation for higher dimensional computations (Thomas, 1993). The main drawback of the Laplace approximation is that its precision cannot be easily adjusted. In NAEP, the second order Laplace approximation is hard-coded into to existing software and even that already contains Order 4 derivatives of the response likelihood. Due to the size of the data sets in NAEP, a sequence of runs to evaluate the precision of numerical integration does not appear feasible even in low dimensions.

There are, of course, drawbacks to the Gauss-Hermite rules as well. They are also prone to any imprecision caused by fixed number of quadrature points. This can be avoided, however, by performing an extra analysis of the integrand, especially by finding the location of any sharp peak, as can be done with any other method. The main property of the Gauss-Hermite rule of being exact on the *whole* of $\mathbb{R}^K$ for polynomials up to a certain degree makes it the ideal choice for numerical integration for many problems where polynomial approximation is suitable.

2 Numerical Integration

In normal population marginal IRT, the implemented numerical integration procedures aim to compute multidimensional integrals of the form

$$I := \int_{\mathbb{R}^K} f(x) e^{-\langle x \mid x \rangle} \mathrm{d}^K x, \tag{5}$$

where f is a smooth function. The rectangle rule approximates the integral by

$$I \cong \sum_{q \in \mathcal{QP}} f(q) e^{-\langle q \mid q \rangle} \Delta q, \tag{6}$$

where $\mathcal{QP} = \{q_1, \ldots, q_Q\}^K$ with $q_i = q_{min} + \frac{q_{max} - q_{min}}{Q-1}(i-1)$, $(i = 1, \ldots, Q)$. $\Delta q = \left(\frac{q_{max} - q_{min}}{Q-1}\right)^K$. An example would be $[-4, 4]$ or $[-5, 5]$ with $Q = 41$.

Gaussian quadrature (Stoer & Bulirsch, 2002, pp. 171–180) provides a tool that makes it possible to compute higher dimensional integrals without compromising computational precision. With the Rth Gauss-Hermite quadrature, the integral is approximated as

$$I \cong \sum_{q \in \mathcal{QP}_{GH_R}^K} f(q) w_q, \tag{7}$$

where $\mathcal{QP}_{GH_R}$ is the set of the zeros of the Rth Hermite polynomial H_R and $\mathcal{QP}_{GH_R}^K$ is the Kth Cartesian power of $\mathcal{QP}_{GH_R}$. The *weights* are given by $w_q = w_{q_1} w_{q_2} \ldots w_{q_K}$, where

$$w_{q_i} = \frac{2^{R-1} R! \sqrt{\pi}}{R^2 H_{R-1}(q_i)^2}. \tag{8}$$

Table 1 contains the minimal Gauss-Hermite quadrature point for some choice of R. It seems that the choice $R = 12$ would result in coverage approximately the same as the usual $[-4, 4]$ used in IRT. This comparison could be misguided, though, since the Gauss-Hermite approximation of I is *exact* on the *whole* of $\mathbb{R}^K$ if f is a polynomial with degree at most $2R - 1$. That is, with $R = 12$ as long as the approximation of f with a Degree 23 polynomial is reasonable the integral is well-approximated by the Gauss-Hermite quadrature on the whole of $\mathbb{R}^K$.

2.1 Gauss-Hermite Integration With Cholesky Decomposition

To be precise, it has to be noted that the functions to be integrated assume a form slightly different from (5), since the integration weight is not $e^{-\langle x \mid x \rangle}$ but the multivariate normal density

$$\frac{1}{(2\pi)^{k/2}\sqrt{\det(\Sigma)}} e^{-1/2 \langle x-\mu \mid \Sigma^{-1} \mid x-\mu \rangle}.$$

Table 1

Minimal Gauss-Hermite Quadrature Points and Weights

R	q_{min}	$w_{q_{min}}$
6	-2.35	$4 \cdot 10^{-3}$
8	-2.93	$2 \cdot 10^{-4}$
10	-3.44	$7 \cdot 10^{-6}$
12	-3.89	$3 \cdot 10^{-7}$
14	-4.30	$9 \cdot 10^{-9}$
16	-4.69	$3 \cdot 10^{-10}$

The simplest approach uses the factorization

$$e^{-1/2\langle x-\mu \mid \Sigma^{-1} \mid x-\mu\rangle} = e^{-1/2\langle x-\mu \mid \Sigma^{-1} \mid x-\mu\rangle + \langle x \mid x\rangle} e^{-\langle x \mid x\rangle}$$

and applies the Gauss-Hermite quadrature for $f(x)e^{-1/2\langle x-\mu \mid \Sigma^{-1} \mid x-\mu)+\langle x \mid x\rangle}$. Another way would be to perform a change of integration variable by first finding a decomposition $2\Sigma = TT^t$ and introducing $z = T^{-1}(x - \mu)$. Then, for the integral,

$$\int_{\mathbb{R}^K} f(x) \det(\Sigma)^{-1/2} e^{-1/2\langle x-\mu \mid \Sigma^{-1} \mid x-\mu\rangle} \mathrm{d}^K x = \int_{\mathbb{R}^K} f(Tz + \mu) e^{-\langle z \mid z\rangle} \mathrm{d}^K z. \tag{9}$$

For the positive definite symmetric matrix 2Σ, many decompositions of the form $2\Sigma = TT^t$ exist. Any such T can be used here. The special case, when T is upper triangular, the decomposition is called Cholesky decomposition (Stoer & Bulirsch, 2002, p. 204).

2.2 Using Normal Approximation to Response Likelihood

If the number of items is relatively large, it is possible that the response likelihood $P(y_i \mid \theta, \beta)$ has a sharp peak at a location depending on the item parameters β and the item responses y_i. An integration technique based on finite number of function evaluations can then fail to sufficiently capture the behavior of the response likelihood. While this is very uncommon in NAEP, for which the number of items per subscale rarely exceeds 10, this issue is addressed here for the sake of completeness.[2] A method more cognizant of the actual behavior of the response likelihood may be computationally more efficient even for tamer response likelihoods.

One way of taking the peak of the response likelihood into consideration first finds the modal multivariate normal approximation

$$P(y_i \mid \theta, \beta) \cong \varphi(\theta; \theta_i^m, \Sigma_i^m), \tag{10}$$

where θ_i^m is the mode of $P(y_i \mid \theta, \beta)$ and Σ_i^m is the modal covariance matrix of $P(y_i \mid \theta, \beta)$. More precisely, θ_i^m is obtained as the solution of

$$\frac{\partial P(y_i \mid \theta, \beta)}{\partial \theta} = 0, \quad (\theta =?), \tag{11}$$

and the modal covariance is defined by

$$\Sigma_i^m = \left(-\frac{\partial^2 \log P(y_i \mid \theta, \beta)}{\partial \theta^2} \right)^{-1} \Bigg|_{\theta=\theta_i^m}. \tag{12}$$

For an arbitrary smooth function $g(\theta)$, the integration proceeds as follows:

$$\begin{aligned}
\mathcal{E}(g)_i &= \int_{\mathbb{R}^K} g(\theta) \frac{P(y_i \mid \theta, \beta)}{\mathcal{E}(L_i)} \varphi(\theta; \Gamma x_i, \Sigma) \mathrm{d}^K\theta \\
&= \int_{\mathbb{R}^K} g(\theta) \frac{P(y_i \mid \theta, \beta)}{\mathcal{E}(L_i)\varphi(\theta, \theta_i^m, \Sigma_i^m)} \varphi(\theta; \theta_i^m, \Sigma_i^m) \varphi(\theta; \Gamma x_i, \Sigma) \mathrm{d}^K\theta \\
&= \int_{\mathbb{R}^K} g(\theta) \frac{P(y_i \mid \theta, \beta)}{\mathcal{E}(L_i)\varphi(\theta; \theta_i^m, \Sigma_i^m)} C_i \varphi(\theta; \theta_i^p, \Sigma_i^p) \mathrm{d}^K\theta,
\end{aligned}$$

where

$$\Sigma_i^p = \left(\Sigma^{-1} + (\Sigma_i^m)^{-1} \right)^{-1}, \tag{13}$$

$$\theta_i^p = \Sigma_i^p (\Sigma^{-1} \Gamma x_i + (\Sigma_i^m)^{-1} \theta_i^m), \tag{14}$$

and

$$C_i = \frac{\sqrt{|\Sigma_i^p|}}{(2\pi)^{K/2} \sqrt{|\Sigma_i^m||\Sigma|}}. \tag{15}$$

Then, one finds the Cholesky decomposition $T_i T_i^t = 2\Sigma_i^p$ and performs the change of variables

$$z_i = T_i^{-1}(\theta - \theta_i^p), \quad \theta = T_i z_i + \theta_i^p \tag{16}$$

to obtain the Gauss-Hermite rule

$$\mathcal{E}(g)_i \cong C_i \sum_{q \in \mathcal{QP}_{GH_R}^K} g(T_i q + \theta_i^p) \frac{P(y_i | T_i q + \theta_i^p, \beta)}{\mathcal{E}(L_i)\varphi(T_i q + \theta_i^p, \theta_i^m, \Sigma_i^m)} w_q. \tag{17}$$

When the approximation (10) is good, then the function $\frac{P(y_i \mid \theta, \beta)}{\varphi(\theta; \theta_i^m, \Sigma_i^m)}$ is approximately constant in the range for which the normal integration weight $\varphi(\theta; \theta_i^p, \Sigma_i^p)$ is not negligible.

Because this computation uses additional information about the integrand, that is, the method adapts itself to the integrand, the technique is sometimes referred to as adaptive numerical integration.

3 Results

3.1 Latent Regression

The above described integration methods are compared using the framework of NAEP's latent regression model (Allen et al., 2001; Mislevy, 1984). In short, the goal of the latent regression estimation is to find regression coefficient matrix $\hat{\Gamma}$ and scale covariance matrix $\hat{\Sigma}$ so that

$$L = \sum_{i=1}^{N} \log \int_{\mathbb{R}^K} P(y_i \mid \theta, \beta)\varphi(\theta; \Gamma x_i, \Sigma)\mathrm{d}^K\theta, \tag{18}$$

is maximized. The subtle difference between (18) and (1) is that the latent regression model uses group means Γx_i instead of the common population mean μ of (1). Here, $x_i \in \mathbb{R}^M$ is the vector of known background variables and $\Gamma \in M_{K,N}(\mathbb{R})$ is the matrix of regression coefficients. The implemented estimation method is the EM-algorithm. Again, interested readers are referred to Mislevy (1984) for the exact definition of the latent regression model and for the derivation of the estimation procedure.

The efficacy of numerical integration is very important here, because the integrations have to be carried out in higher dimensions for hundreds of thousands of response patterns. Finally, the list of the three functions for which numerical intergations should be computed is given as follows:

$$g_1(\theta) = 1, \quad g_2(\theta) = \theta, \quad g_3(\theta) = \theta_l\theta_m, \quad (1 \leq l, m \leq K). \tag{19}$$

3.2 QP41 Versus Gauss-Hermite

Table A3 provides a comparison of the two approaches to the Gauss-Hermite method (with and without Cholesky decomposition, but without normal approximation) and the current NAEP practice in terms of running time and precision. This comparison uses a 20-item, two-subscale test (10 items each) with 20 background variables and 500 students. The two-parameter logistic IRT model was utilized. More precisely,

$$P(y_{ij}|\theta, \beta_j = (a_j, b_j)) = \frac{1}{1 + e^{(1-2y_{ij})a_j(\theta - b_j)}}. \tag{20}$$

In accordance with current NAEP practice, item parameters were kept constant throughout the estimation; in this case, they were kept at their true value.

The listed running time is in seconds, but running times can only be used as relative to one another at best because they depend on the given computing environment. Precision is measured as the distance of the given covariance matrix estimate from the most precise estimate obtained here. This was assumed to be the one obtained from the Cholesky decomposition with 16 quadrature points and 200 EM cycles. Note that this comparison is not performed to decide which method is more precise because that is clear from their construction. The main reason for the comparison is to identify the magnitude of the running time and precision differences.

There are several observations concerning Table A1:

- The two Gauss-Hermite methods (factorization and Cholesky-decomposition-based) are very close to one another in terms of both running time and precision. The only sizable difference can be observed when only a few quadrature points are used because then the precision of the factorization method suffers significantly.

- The current NAEP practice (QP41) achieves the same precision as the most precise Gauss-Hermite, but this comes with a serious running time penalty— 24.7% more time is needed to obtain the results of comparable precision.

- There is a significant difference between the current NAEP estimates with quadrature range $[-4, 4]$ and $[-5, 5]$.

- To keep the presentation transparent, the threshold (the distance between estimates in consecutive iteration steps) is omitted from Table A3. In both the Gauss-Hermite and QP41 schemes, it is almost independent of the number of quadrature points and the method used. The dependence on the number of EM cycles is as follows: $t_{10} = 3.5 \cdot 10^{-3}$, $t_{30} = 2.5 \cdot 10^{-4}$, $t_{60} = 6.2 \cdot 10^{-6}$, $t_{200} = 5.0 \cdot 10^{-13}$. Using a not so unusual convergence criterion ($t = 0.005$), *all* of the parameter estimates could have been kept. This may appear to be a too optimistic approach, keeping in mind that the estimates after 10 EM cycles are still 0.024 away from the best estimate.

While it is always useful to have numerical evidence behind any statements, it is worthwhile to emphasize the obvious pitfalls of the QP41 approach with fixed quadrature range. The main

effects of the fixed quadrature range are that, first, it cuts out sizable contributions from the item likelihood when the range is not large enough. Second, the rectangle rule essentially redefines the response likelihoods to be step functions. The estimated parameters reflect convergence of the model with respect to these new response function alternatives. That is, without changing the number of quadrature points, there is no chance of improving the estimates.

The Gauss-Hermite rule replaces the response likelihoods with their polynomial approximations of appropriate degree (based on the given number of function evaluations), which, while able to produce much better approximations, may not always be desirable, either.

It is this understanding of the effect of the redefinition of the response likelihood that explains why the quality of convergence does not depend on the chosen numerical integration method. From the point of view of the convergence of the EM algorithm, it is almost irrelevant which approximation is chosen: The method will converge to a solution relative to the given response likelihood. This response likelihood is determined by the subtle interaction of the logistic IRT model and the numerical integration method. The numerical integration method is deemed better when this approximation is more appropriate.

3.3 Normal Approximation With Cholesky Method

Table A2 compares the two Cholesky-decomposition-based Gauss-Hermite methods (with or without normal approximation) in terms of their running time and precision. The distance again is defined as the distance of the given covariance matrix from the best estimate obtained, which in this case was the normal-approximation-based estimate obtained after 200 EM cycles with 16 quadrature points. The normal approximation method requires fewer quadrature points to reach convergence than its standard counterpart. However, due to the increased number of function evaluations, each step requires more time. This comparison shows that the two methods reaches the same level of precision in about the same time. The convergence thresholds are independent of both the method and number of quadrature points used; they depend only on the number of EM cycles similar to what was observed before. The running time comparison is relatively weak because there were no steps taken to optimize the underlying algorithms. There will be a significant cut in running time when the above described algorithms are implemented with optimization.

Next, to test the efficacy of the normal approximation model, the number of items were

increased. This change makes the peak of the response likelihoods sharper and the normal approximation more favorable.

Table A3 displays results from a run where the data is similar to the one used before, with the only exception being the number of items, which is now changed to 80. This setup speeds up the convergence, and the orders of the thresholds are $t_{10} = 10^{-3}$, $t_{30} = 10^{-7}$, $t_{60} = 10^{-13}$, and $t_{200} = 10^{-16}$.

The presence of sharp peaks in the likelihood makes the estimation intuitively more dependent on the number of quadrature points. To capture this dependence better, computations with a larger number of quadrature points (18-24) were performed as well. Since the convergence was considerably faster in this case, only 60 EM cycles were run when the number of quadrature points was large. Accordingly, to compute precision, the normal approximation and Cholesky-decomposition-based run with 24 quadrature points and 60 EM cycles were chosen as a reference.

Table A3 shows the following:

- The normal approximation performs well even with four quadrature points.
- Without the normal approximation, one has to increase the number of quadrature points to reach reasonable convergence.
- Monitoring the convergence threshold does not provide sufficient information about the precision because this latter is independent of the integration method, which determines the precision along with the number of EM cycles.

4 Conclusion

Numerical integration is an important but sometimes neglected part in IRT estimation. Numerical precision and running time carry equal importance when evaluating the merits of any numerical integration method. This paper showed that the least sophisticated rectangle method is favored undeservedly and its drawbacks can be easily identified. It also showed that the normal-approximation-based Gauss-Hermite rule coupled with Cholesky decomposition (NGHC) can be successfully applied in IRT computations. Some issues, however, call for discussion. First, the normal approximation method requires finding the mode of the response likelihood. The mode, however, does not always exist. The probability of a flat response likelihood increases as

the number of items decreases. One may replace the modal normal approximation with a generic normal approximation for a flat response likelihood, however, without losing precision. To justify this, it is important to note that the method does not replace the response likelihood with its modal normal approximation. The modal normal approximation is only used to tame the response likelihood without actually changing the integrand in question.

Second, while the NGHC is capable of capturing the behavior of the response likelihood, when the number of quadrature points is fixed, it is still necessary to assess the convergence of the numerical integrals. One operationally feasible way would be to draw a small sample of responses that is representative of the whole assessment prior to the full analysis and to closely analyze the convergence of the typical integrals of the estimation using this sample. Assuming that students in one assessment respond to fairly comparable set of items (otherwise, the assessment would have other, more serious problems), this sampling should provide, very quickly, sufficient information about the necessary number of quadrature points needed to reach accuracy and precision targets.

References

Allen, N. L., Donoghue, J. R., & Schoeps, T. L. (2001). *The NAEP 1998 technical report* (NCES 2001-509). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Genz, A., & Kass, R. E. (1997). Subregion adaptive integration of functions having a dominant peak. *Journal of Computational and Graphical Statistics*, *6*, 92–111.

Genz, A., & Kass, R. E. (1998). *BAYESPACK: A collection of numerical integration software for bayesian analysis.* Retrieved March 6, 2007, from http://www.math.wsu.edu/faculty/genz/homepage.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika*, *49*(3), 359–381.

Muraki, E., & Bock, R. D. (1999). PARSCALE: IRT item analysis and test scoring for rating-scale data [Computer software]. Chicago, IL: Scientific Software International.

Stoer, J., & Bulirsch, R. (2002). *Introduction to numerical analysis.* New York: Springer.

Thomas, N. (1993). Asymptotic corrections for multivariate posterior moments with factored likelihood functions. *Journal of Computational and Graphical Statistics*, *2*(3), 309–322.

Notes

[1] The precision goal of the specific computation can always serve as a guideline to make the definition of negligible more precise. Note that the notion of essential mass is closely related to that of the confidence region.

[2] Note that, in general, the more items, the sharper the peak.

Appendix

Table A1

Running Time and Precision of the EM-Algorithm With 10, 30, 60, and 200 Cycles

	Time		Distance	
QP	F	Chol	F	Chol
10 EM cycles				
4	18	17	$2.5 \cdot 10^{-1}$	$3.0 \cdot 10^{-2}$
6	30	28	$5.4 \cdot 10^{-2}$	$2.7 \cdot 10^{-2}$
8	46	43	$2.0 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$
10	68	63	$2.3 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$
12	94	88	$2.4 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$
14	124	116	$2.4 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$
16	160	150	$2.4 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$
41[a]	1673		$2.2 \cdot 10^{-2}$	
41[b]	1645		$2.4 \cdot 10^{-2}$	
30 *EM* cycles				
4	52	52	$2.5 \cdot 10^{-1}$	$1.1 \cdot 10^{-2}$
6	85	88	$6.2 \cdot 10^{-2}$	$7.2 \cdot 10^{-3}$
8	132	131	$1.1 \cdot 10^{-2}$	$2.5 \cdot 10^{-3}$
10	189	190	$1.8 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$
12	261	263	$1.8 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$
14	346	345	$1.9 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$
16	444	450	$2.0 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$
41[a]	4998		$3.1 \cdot 10^{-3}$	
41[b]	4913		$2.0 \cdot 10^{-3}$	
60 *EM* cycles				
4	103	104	$2.5 \cdot 10^{-1}$	$9.5 \cdot 10^{-3}$
6	167	170	$6.2 \cdot 10^{-2}$	$5.6 \cdot 10^{-3}$
8	256	262	$1.2 \cdot 10^{-2}$	$9.4 \cdot 10^{-4}$
10	373	382	$2.0 \cdot 10^{-3}$	$1.0 \cdot 10^{-4}$
12	514	527	$2.9 \cdot 10^{-4}$	$6.3 \cdot 10^{-5}$
14	680	699	$5.0 \cdot 10^{-5}$	$6.2 \cdot 10^{-5}$
16	873	898	$5.4 \cdot 10^{-5}$	$5.8 \cdot 10^{-5}$
41[a]	9999		$3.8 \cdot 10^{-3}$	
41[b]	9833		$5.2 \cdot 10^{-5}$	
200 *EM* cycles				
4	339	349	$2.5 \cdot 10^{-1}$	$9.4 \cdot 10^{-3}$
6	550	568	$6.2 \cdot 10^{-2}$	$5.6 \cdot 10^{-3}$
8	848	879	$1.2 \cdot 10^{-2}$	$9.1 \cdot 10^{-4}$
10	1227	1275	$2.0 \cdot 10^{-3}$	$1.1 \cdot 10^{-4}$
12	1693	1764	$3.2 \cdot 10^{-4}$	$1.8 \cdot 10^{-5}$
14	2238	2334	$4.8 \cdot 10^{-5}$	$5.7 \cdot 10^{-6}$
16	2873	2994	$7.9 \cdot 10^{-6}$	N/A
41[a]	33418		$3.8 \cdot 10^{-3}$	
41[b]	37768		$5.7 \cdot 10^{-5}$	

Note. Time is in seconds. F = factorization, Chol = Cholesky decomposition.

[a] On $[-4, 4]$. [b] On $[-5, 5]$.

Table A2

Running Time and Precision of the EM-Algorithm With 10, 30, 60, and 200 Cycles

	Precision								Time	
EM	10		30		60		200		200	
QP	Ch	NCh	Ch	NCh	Ch	NCh	Ch	NCh	Ch	NCh
4	$8.1 \cdot 10^{-2}$	$6.3 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	$5.4 \cdot 10^{-3}$	$1.1 \cdot 10^{-2}$	$3.7 \cdot 10^{-3}$	$1.1 \cdot 10^{-2}$	$3.7 \cdot 10^{-3}$	447	832
6	$6.6 \cdot 10^{-2}$	$6.3 \cdot 10^{-2}$	$9.3 \cdot 10^{-3}$	$4.2 \cdot 10^{-3}$	$7.7 \cdot 10^{-3}$	$2.0 \cdot 10^{-4}$	$7.7 \cdot 10^{-3}$	$2.3 \cdot 10^{-4}$	665	1551
8	$6.4 \cdot 10^{-2}$	$6.4 \cdot 10^{-2}$	$5.5 \cdot 10^{-3}$	$4.3 \cdot 10^{-3}$	$1.9 \cdot 10^{-3}$	$9.4 \cdot 10^{-5}$	$1.9 \cdot 10^{-3}$	$3.3 \cdot 10^{-5}$	994	2565
10	$6.4 \cdot 10^{-2}$	$6.4 \cdot 10^{-2}$	$4.6 \cdot 10^{-3}$	$4.3 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$	$9.7 \cdot 10^{-5}$	$9.7 \cdot 10^{-4}$	$6.7 \cdot 10^{-6}$	1414	3849
12	$6.4 \cdot 10^{-2}$	$6.4 \cdot 10^{-2}$	$4.3 \cdot 10^{-3}$	$4.3 \cdot 10^{-3}$	$1.9 \cdot 10^{-4}$	$9.8 \cdot 10^{-5}$	$1.6 \cdot 10^{-4}$	$1.2 \cdot 10^{-6}$	1929	5428
14	$6.4 \cdot 10^{-2}$	$6.4 \cdot 10^{-2}$	$4.3 \cdot 10^{-3}$	$4.3 \cdot 10^{-3}$	$9.5 \cdot 10^{-5}$	$9.8 \cdot 10^{-5}$	$5.4 \cdot 10^{-6}$	$1.8 \cdot 10^{-7}$	2535	7306
16	$6.4 \cdot 10^{-2}$	$6.4 \cdot 10^{-2}$	$4.3 \cdot 10^{-3}$	$4.3 \cdot 10^{-3}$	$1.0 \cdot 10^{-4}$	$9.8 \cdot 10^{-5}$	$7.1 \cdot 10^{-6}$	N/A	3298	9479

Note. Time is in seconds. Ch = Cholesky, NCh = Normal Approximation with Cholesky. Number of items = 20.

Table A3

Running Time and Precision of the EM-Algorithm With 10, 30, 60, and 200 Cycles and 80 Items

	Precision								Time	
EM	10		30		60		200		200	
QP	Ch	NCh	Ch	NCh	Ch	NCh	Ch	NCh	Ch	NCh
4	$1.4 \cdot 10^{-1}$	$3.9 \cdot 10^{-3}$	$1.4 \cdot 10^{-1}$	$2.4 \cdot 10^{-3}$	$1.4 \cdot 10^{-1}$	$2.4 \cdot 10^{-3}$	$1.4 \cdot 10^{-1}$	$2.4 \cdot 10^{-3}$	835	1222
6	$4.0 \cdot 10^{-2}$	$2.9 \cdot 10^{-3}$	$4.0 \cdot 10^{-2}$	$1.1 \cdot 10^{-3}$	$4.0 \cdot 10^{-2}$	$1.1 \cdot 10^{-3}$	$4.0 \cdot 10^{-2}$	$1.1 \cdot 10^{-3}$	1363	2227
8	$2.1 \cdot 10^{-2}$	$2.6 \cdot 10^{-3}$	$2.1 \cdot 10^{-2}$	$5.1 \cdot 10^{-4}$	$2.1 \cdot 10^{-2}$	$5.1 \cdot 10^{-4}$	$2.1 \cdot 10^{-2}$	$5.1 \cdot 10^{-4}$	2106	3631
10	$9.1 \cdot 10^{-3}$	$2.5 \cdot 10^{-3}$	$8.9 \cdot 10^{-3}$	$2.3 \cdot 10^{-4}$	$8.9 \cdot 10^{-3}$	$2.3 \cdot 10^{-4}$	$8.9 \cdot 10^{-3}$	$2.3 \cdot 10^{-4}$	3066	5443
12	$1.9 \cdot 10^{-3}$	$2.4 \cdot 10^{-3}$	$9.9 \cdot 10^{-4}$	$1.1 \cdot 10^{-4}$	$9.9 \cdot 10^{-4}$	$1.1 \cdot 10^{-4}$	$9.9 \cdot 10^{-4}$	$1.1 \cdot 10^{-4}$	4259	7673
14	$3.4 \cdot 10^{-3}$	$2.4 \cdot 10^{-3}$	$2.7 \cdot 10^{-3}$	$4.7 \cdot 10^{-5}$	$2.7 \cdot 10^{-3}$	$4.7 \cdot 10^{-5}$	$2.7 \cdot 10^{-3}$	$4.7 \cdot 10^{-5}$	5658	10302
16	$2.6 \cdot 10^{-3}$	$2.4 \cdot 10^{-3}$	$1.4 \cdot 10^{-3}$	$2.0 \cdot 10^{-5}$	$1.4 \cdot 10^{-3}$	$2.0 \cdot 10^{-5}$	$1.4 \cdot 10^{-3}$	$2.0 \cdot 10^{-5}$	7285	13345
18	$2.3 \cdot 10^{-3}$	$2.4 \cdot 10^{-3}$	$2.2 \cdot 10^{-4}$	$8.5 \cdot 10^{-6}$	$2.2 \cdot 10^{-4}$	$8.4 \cdot 10^{-6}$			18176	33513
20	$2.5 \cdot 10^{-3}$	$2.4 \cdot 10^{-3}$	$7.5 \cdot 10^{-4}$	$3.4 \cdot 10^{-6}$	$7.5 \cdot 10^{-4}$	$3.3 \cdot 10^{-6}$			22293	41146
22	$2.5 \cdot 10^{-3}$	$2.4 \cdot 10^{-3}$	$8.0 \cdot 10^{-4}$	$1.1 \cdot 10^{-6}$	$8.0 \cdot 10^{-4}$	$1.0 \cdot 10^{-6}$			26787	49626
24	$2.5 \cdot 10^{-3}$	$2.4 \cdot 10^{-3}$	$6.0 \cdot 10^{-4}$	$3.1 \cdot 10^{-7}$	$6.0 \cdot 10^{-4}$	N/A			31751	58863

Note. Time is in seconds. Ch = Cholesky, NCh = normal approximation with Cholesky. Number of items = 80.