A Discussion of Quadrature for Latent Variable Models with Categorical Responses

Xiangxiang Meng, Xinming An

SAS Institute, Inc.

Abstract

The estimation in latent variable models with categorical responses, such as generalized linear mixed models (GLMM) and item response models (IRT), often involves integration that cannot be solved analytically. Numerous approaches have been proposed to tackle this problem, among which numerical integration, for example Gauss-Hermite quadrature, has proven to generate more accurate estimations as compared with other methods. G-H quadrature with 20 points per dimension is widely considered to be able to produce quality estimations. However, several researches have reported unstable or biased estimations even with more than 20 quadrature points per dimension under some special conditions, for example when the variances of latent variables are large. The objective of this research is to investigate when and why standard G-H quadrature will be inadequate even with a large number of quadrature points, and demonstrate why adaptive quadrature can effectively solve this problem.

KEYWORDS: Latent Variable model, Quadrature, Generalized Linear Mixed Model, Item Response Model

1. Introduction

Latent variable model refers to statistical models with unobserved variables that explain the key features of observed data. Notable examples include factor analysis models, finite mixture models and linear mixed models. They have been widely used to tackle various practical modeling difficulties. For example, factor analysis models are often used for high dimensional data analysis, and linear mixed models enjoy several appealing properties for clustered data analysis. During the last twenty years, latent variable models have been extended in many different ways, which have greatly increased the practical usefulness of these models. Among these extensions, latent variable models for categorical responses, such as generalized linear mixed model (GLMM), are probably the most valuable. The estimation of latent variable models with categorical responses, however, often involves intractable integrations. A large number of approaches have been proposed to handle this problem. The most commonly used approaches include Laplace approximation, numerical integration and Monte Carlo integration (e.g., Breslow and Clayton 1993; Bock and Aitkin 1981; Meng and Schilling 1996). Estimates based on numerical integration, mostly Gauss-Hermite (G-H) quadrature, have shown to be the most accurate, but there are two scenarios that G-H quadrature becomes inadequate. The first one occurs when the number of latent variables is large, since the number of quadrature points grows exponentially as the number of latent variable increases. The other problem is related to the approximation accuracy of G-H quadrature under certain situations. It is commonly believed that 20 G-H quadrature points per dimension would produce accurate approximation of the likelihood and as a results reliable parameter estimates. However, several cases have been reported where a large number of quadrature points is required to obtain valid estimates (e.g., Lesaffre and Spiessens 2001; Rabe-Hesketh et al. 2002).

While the issue associated with high dimensional quadrature is widely appreciated, the approximation accuracy issue has only been reported for GLMM under certain cases. The underlying cause and most importantly its effect on other latent variable models, such as item response theory (IRT), has not been investigated. The purpose of this paper is to investigate and illustrate the approximation accuracy issues associated with G-H quadrature for latent variable models with categorical responses and why adaptive rule (Naylor and Smith 1982) can greatly reduce this problem. Cases where adaptive quadrature may potentially fail will also be discussed. The rest of this paper is organized at follows. Parameterization for latent variable models with categorical responses will be presented in the next section. In section 3, we will illustrate when the G-H quadrature will become inefficient. Using GLMM and IRT as examples, we will also review the cases when G-H quadrature may fail to produce reliable estimates. The paper is concluded with some discussions on the implications of this issue for applied researchers.

2. Latent Variable Model with Categorical Responses

In this section we will briefly review latent variable models with categorical responses and their estimations based on marginal likelihood. To simplify notations binary responses will be adopted for illustration purpose, while results presented in this paper also apply to ordinal and nominal responses. A general parameterization of one dimensional latent variable models with binary response, $u_i = u(u_{i1}, \ldots, u_{iJ}),\ i = 1, \ldots, N$, can be expressed as follows

$$y_i = X_i\beta + \Lambda_i\eta_i + \epsilon_i \tag{1}$$

$$P(u_{ij} = 1) = P(y_{ij} > 0) \tag{2}$$

where $y_i = (y_{i1}, \ldots, y_{iJ})$ are the continuous latent responses underlying u_i, X_i are the covariates, η_i and ϵ_i are the latent factor and the random error, and they are often assumed to be normally distributed and independent, $\eta_i \sim N(0, \sigma^2),\ \epsilon_i \sim N_J(0, I),\ \eta_i \perp \epsilon_i$. In IRT, σ is fixed at 1 for identification purpose. Structure matrix Λ_i can include either variables or parameters. For example, Λ_i is the covariates for random effects in mixed effects models, or the factor loading matrix in factor analysis model. Based on the above model specification, we have

$$P_{ij} = P(u_{ij} = 1|\eta_i) = \int_0^\infty p(y; x_{ij}\beta_j + \lambda_{ij}\eta_i, 1)dy = 1 - Q_{ij} \tag{3}$$

Parameter estimates for this model are often obtained by maximizing the marginal likelihood, which can be expressed as

$$L(\theta|U) = \prod_{i=1}^{N} \int \prod_{j=1}^{J} P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}} \phi(\eta)d\eta \tag{4}$$

where $\phi(\eta)$ is the density function of the prior distribution for latent factor η. The corresponding log marginal likelihood is

$$ll(\theta|U) = \sum_{i=1}^{N} log \int L_i(\eta)d\eta = \prod_{i=1}^{N} log \int \sum_{j=1}^{J} P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}} \phi(\eta)d\eta \tag{5}$$

Integrations involved in the above likelihood can not be solved analytically, and is approximated with numeric integration techniques, most often Gauss-Hermite quadrature.

3. Gauss-Hermite Quadrature

In general the G-H quadrature can be presented as follows

$$\int_{-\infty}^{\infty} g(x)dx = \int_{-\infty}^{\infty} f(x)\phi(x)dx \approx \sum_{g=1}^{G} f(x_g)w_g \tag{6}$$

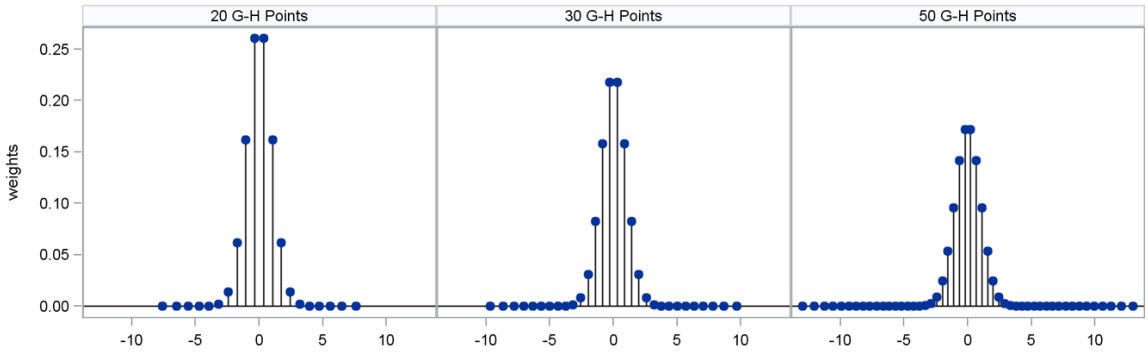

Figure 1: Illustrations of Gaussian-Hermite quadrature with 20, 30, 50 quadrature points.

where G is the number of quadrature, x_g and w_g are the integration points and weights, which are uniquely determined by the integration domain and the weighting kernel $\phi(x)$. Traditional G-H quadrature often uses e^{-x^2} as the weighting kernel. In statistics the density of standard normal distribution is widely used because for the estimation of various statistical models, the Gaussian density is often a factor of the integrand. In the case when a Gaussian density is not a factor of the integrand, the integral is transformed in to form in 6 by dividing and multiplying the original integrand by the standard normal density. Graphic illustrations of G-H quadrature with the number of quadrature points ranging from 20 to 50 are included in Figure 1. In Figure 1 we can observe that (1) the positions and weights of the quadrature points are symmetric around zero; (2) As the number of quadrature increases, the quadrature points extend gradually to the two ends.

The G points quadrature approximation is exact if $f(x)$ or $g(x)/\phi(x)$ is a polynomial of order $2G-1$. For example, the rth moment of standard normal distribution can be approximated exactly by $(r+1)/2$ points G-H quadrature. However, as pointed out by many researches, $f(x)$ for various statistical models often has a sharp peak and cannot be well approximated by a low degree polynomial. Furthermore, the peak may be far from zero so that substantial contribution to the integral is lost unless a large number of quadrature points have been used. Three situations when the G-H quadrature will become inadequate are illustrated by Figure 2. The red curve represents the case when $f(x)$ has a sharp peak. The green curve shows the case where the peak of the integrand is far from zero. The most troublesome case is illustrated by the black curve, in which $f(x)$ has a sharp peak and the peak is far from zero. Among these 30 quadrature points, only 4, 7 and 3 quadrature points make significant contributions to the approximation for the above three cases.

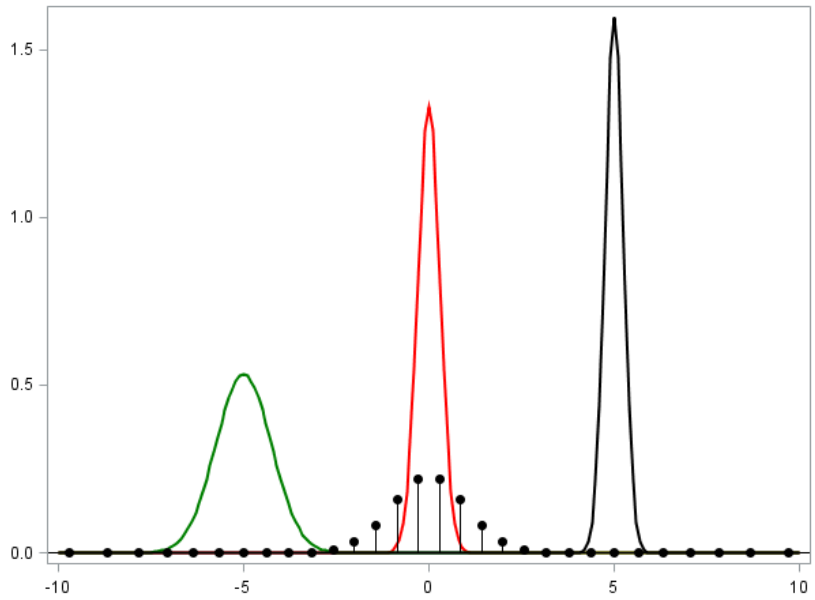

Figure 2: Three situations when the Gaussian-Hermite quadrature becomes inadequate.

4. G-H quadrature for latent variable models with categorical responses

G-H quadrature has been widely used along with optimization techniques, such as Expectation-Maximization and Newton methods, to estimate various latent variable models with categorical responses. Several cases have been reported that a larger number of quadrature points are needed to get reliable estimates for generalized linear mixed models with categorical responses. These problems are often caused by the fact that the peak for some integrands involved in the model are far from zero and(or) the integrand is very sharp around the peak. These problems also apply to other latent variable models with categorical responses, such as the item response theory model, but have not be recognized.

Using GLMM and IRT with binary responses as examples, we will investigate how the integrands, the location of peak and the sharpness, are affected by the setting of the model, for example cluster size and parameter values. Equation 5 suggests that the integrand, $L_i(\eta) = \prod_{j=1}^{J} P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}} \phi(\eta)$, can be considered as the unnormalized posterior distribution of latent variable η for cluster (or subject) i with response $u_i = (u_{i1}, \ldots, u_{iJ})$. Let $\hat{\eta}_i$ and H_i denote the location of the peak and the corresponding Hessian of $-logL_i(\eta)$. Then $\hat{\eta}_i$ and $\frac{1}{H}$ can be used as an estimate for the posterior mean and variance of latent variable η_i. The default setting for GLMM is a random intercept model with 30 observations for each cluster. The variance of random effect is set to 1 and the population intercept is set to 0. The default IRT model includes 30 items and one latent factor. Factor loadings are set to 1. Under this setting,

J	MaxH_i ($\hat{\eta}_i$)	G-H	AveH_i	Max abs($\hat{\eta}_i$)(H_i)	G-H	Ave abs($\hat{\eta}_i$)
30	20.10(0)	60	15.88	1.926(4.83)	15	0.72
100	64.66(0)	180	49.61	2.378(6.71)	20	0.77
200	128.32(0)	350	97.69	2.616(7.88)	25	0.78

Table 1: Number of Gaussian-Hermite quadrature points needed for the integrands with the sharpest peak (MaxH_i) or the peak furthest from zero (Max abs($\hat{\eta}_i$)) in GLMM models with different cluster sizes ($J = 30, 100, 200$) .

responses have the same marginal mean will produce roughly the same integrand. Thus instead of keeping track of all the possible response patterns, we will only consider the J+1 responses patterns with different marginal means. In the following studies, the average and maximum value of $\hat{\eta}_i$ and H_i across these J+1 response patterns will be reported under different settings. For integrand with the maximum $\hat{\eta}_i$ or H_i, the number of quadrature points that is needed to approximate the integration with an error smaller than 0.01 will also be reported.

It is well known that as the cluster size increases, the shape of the integrand becomes closer to a normal density function but meanwhile its peak becomes sharper (the posterior variance become smaller) which will cause problems for the quadrature. Results for cluster size of 30, 100, and 200 are summarized in Table 1, including the average and maximum $\hat{\eta}_i$ and H_i and the number of quadrature points needed to accurately approximate these integrand with the maximum $\hat{\eta}_i$ or H_i. These results suggest that the cluster size has an significant impact on the posterior variance ($\frac{1}{H_i}$) which in turn strongly affects the number of quadrature points required for numerical accuracy. As the cluster size moves from 30 to 200, the maximum H_i increases from 20 to 128; and to obtain equal accuracy, the number of quadrature points move from 60 to 350. Same effects also applies to IRT models, in which the cluster size refer to the number of items.

It has been reported that the high intraclass correlation can also cause problems for G-H quadrature for the estimation of GLMM model. Since the error variance is fixed as constant for identification purpose for GLMM model, higher intraclass correlation suggest larger variance of the random effect. Results correspond to different random effects variance, σ^2, are summarized in Table 2. To our surprise, the average and maximum value of $\hat{\eta}_i$ and H_i do not change a lot as σ^2 increases from 0.5 to 10. As a results, large σ^2 or equivalently hight intraclass correlation does not cause problems for the quadrature.

Last, we want to examine how the mean and scale of z_i affect the maximum and average value of H_i and $\hat{\eta}_i$ and whether they will cause problems for the G-H quadrature. The model used here is a random slope model that includes 30 observations for each cluster and the random effect variance is fixed to 1. The population intercept and slope are set to 0 and 1 respectively. By default,

	σ^2	MaxH_i ($\hat{\eta}_i$)	G-H	AveH_i	Max abs($\hat{\eta}_i$)(H_i)	G-H	Ave abs($\hat{\eta}_i$)
	0.5	65.66(0)	180	51.10	2.124(11.21)	25	0.742
J=100	1	64.66(0)	180	49.61	2.378(6.71)	20	0.766
	10	63.76(0)	180	48.20	3.116(1.07)	10	0.797
	0.5	21.10(0)	60	17.26	1.641(7.74)	20	0.661
J=30	1	20.10(0)	60	15.89	1.926(4.83)	20	0.717
	10	19.20(0)	60	14.54	2.748(0.86)	8	0.797

Table 2: Number of Gaussian-Hermite quadrature points needed for the integrands with the sharpest peak (MaxH_i) or the peak furthest from zero (Max abs($\hat{\eta}_i$)) in GLMM models with different intraclass correlations ($\sigma^2 = 0.5, 1, 10$) and cluster size ($J = 30, 100$).

μ_z	σ_z	MaxH_i ($\hat{\eta}_i$)	G-H	AveH_i	Max abs($\hat{\eta}_i$)(H_i)	G-H	Ave abs($\hat{\eta}_i$)
0	1	23.64(-1.0029)	60	18.76	2.620(4.32)	15	1.0551
0	4	91.59(-1.0124)	250	70.99	2.187(5.69)	15	1.0165
0	1/4	6.65(-1.0411)	20	5.65	3.007(3.10)	8	1.1504
5	1	675.45(-0.1644)	400+	506.64	0.4825(8.02)	25	0.1975
-5	1	334.96(0.1866)	400+	240.45	0.6897(94.64)	250	0.2623

Table 3: Number of Gaussian-Hermite quadrature points needed for the integrands with the sharpest peak (MaxH_i) or the peak furthest from zero (Max abs($\hat{\eta}_i$)) in random slope models with different mean and variance (μ_z σ_z)of the random slope.

z_i is drawn from a standard normal distribution and later shifted or rescaled to create other settings. Results summarized in Table 3 suggest that both the mean and variance of z_i have great impact on the maximum and average value of H_i and in turn the G-H quadrature. The mean of z_i is especially influential.

The element in IRT corresponding to the z_i in GLMM is the factor loading matrix Λ, which is unknown and need to be estimated. Table 4 summarizes the results for three cases where factor loadings are set to 0.5, 1 and 2 respectively. These results suggest that these factor loadings have a similar effects as z_i.

λ	MaxH_i ($\hat{\eta}_i$)	G-H	AveH_i	Max abs($\hat{\eta}_i$)(H_i)	G-H	Ave abs($\hat{\eta}_i$)
0.5	5.77(0)	15	4.95	2.680(3.03)	9	1.1611
1	20.10(0)	55	15.89	1.926(4.83)	15	0.7165
2	77.39(0)	210	59.11	1.221(7.02)	18	0.3885

Table 4: Number of Gaussian-Hermite quadrature points needed for the integrands with the sharpest peak (MaxH_i) or the peak furthest from zero (Max abs($\hat{\eta}_i$)) in IRT models with different factor loadings (λ).

5. Discussion

In this paper, we investigate several different factors that affect the G-H quadrature for the estimation of latent variable models with categorical responses. Results suggest that the cluster size, and the mean and variance of random covariates or factor loadings have great impact on G-H quadrature. The same problem also applies to other latent variable models, such as Structured Equation Modeling with categorical responses, where G-H quadrature is used for model estimation. One of our observation that contradicts with previous research results is that high intraclass correlation is not a cause of numerical accuracy problem for G-H quadrature. Adaptive G-H quadrature (Liu and Pierce 1994) has shown to be able to solve this problem effectively (Lesaffre and Spiessens 2001) and has also been used to improve computational efficiency for high dimensional latent variable models (e.g., Rabe-Hesketh et al. 2002; Schilling and Bock 2005). While several statistical packages, such as the GLMMIX procedure in SAS, have changed their default numeric integration technique from G-H to adaptive G-H quadrature, a large number of packages for latent variable models still use G-H as their default or may not even have adaptive G-H available. Implementation from this research to applied researchers is that make the default integration technique to adaptive G-H quadrature if it is available. Otherwise, increasing the number of quadrature until the change in parameter estimate becomes very small.

A. Item Response Theory Model with Binary Responses

An item response theory model (Bock and Aitkin 1981; Bock et al. 1988) with binary responses, $u_i = (u_{i1}, \ldots, u_{ip}), i = 1, \ldots, n$, can be expressed with the following equations:

$$y_i = \mu + \Lambda\eta_i + \epsilon_i \tag{7}$$

$$u_{ij} = \begin{cases} 0 & \text{if} \quad y_{ij} < 0 \\ 1 & \text{if} \quad y_{ij} > 0 \end{cases}, \tag{8}$$

where $y_i = (y_{i1}, \ldots, y_{ip})$ is a p elements vector, $\eta_i = (\eta_{i1}, \ldots, \eta_{id}) \sim N_d(0, \Phi)$ and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{ip}) \sim N_p(0, \Psi)$ are factor scores and residuals, μ is an intercept, Λ is a p by d factor loading matrix, and u_{ij} is the observed binary response of the ith respondent on the jth variable. Ψ is fixed as the identity matrix I for identification purpose, which suggests that $y_{ij}, j = 1, \ldots, p$ are independent conditional on latent factor η_i. This conditional independence assumption plays an important role in the development of the direct sampling based MCEM algorithm. To identify the model, we also need to fix some parameters in Λ and Φ. In exploratory analysis, Λ_{ij} is fixed at 0, for $j \geq i$, and Φ is fixed to be identity matrix I. In confirmatory analysis, the research design will specify enough zeros in Λ such that factor rotation will be impossible. To fix factor scales, we can either restrict one element in each column of Λ to be

1 or restrict the diagonal elements of Φ to be 1.

B. Generalized Linear Mixed Effect Model with Binary Responses

Assume that we have n subjects, and for each subject, there are n_i binary responses $u_i = (u_{i1}, \ldots, u_{in_i})$. Also assume that $y_i = (y_{i1}, \ldots, y_{in_i})$ is the latent response underlying u_i. Then, based on the latent response formulation (McCulloch 1994; Chan and Kuk 1997), the mixed effects model with binary responses can be expressed by the following two equations

$$y_i = X_i\beta + Z_i b_i + \epsilon_i \tag{9}$$

and

$$u_{ij} = \begin{cases} 0 & \text{if} \quad y_{ij} < 0 \\ 1 & \text{if} \quad y_{ij} > 0 \end{cases}, \tag{10}$$

where X_i and Z_i are matrices of known covariates with $(n_i \times p)$ and $(n_i \times q)$ dimensions respectively, β is a p dimensional vector of fixed effects, b_i is a q dimensional vector of random effects with $b_i \sim N(0, D)$, ε_i is a n_i dimensional vector of residuals, and b_i and ε_i, $i = 1, \ldots, n$ are independent. It is often assumed that $\varepsilon_i \sim N(0, I_{n_i})$, which will lead to the conditional independence assumption that $y_{i1}, \ldots, y_{in_i}$ are independent conditional on random effects b_i. As will be shown in the next section, this assumption is critical for the development of the DSMCEM algorithm. The unknown parameters in this model are $\theta = \{\beta, D\}$.

References

R. Bock, R. Gibbons, and E. Muraki. Full-information item factor analysis. *Applied Psychological Measurement*, 12(3):261, 1988.

R. D. Bock and M. Aitkin. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4):443–459, 1981.

N. Breslow and D. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88(421):9–25, 1993.

J. S. K. Chan and A. Y. C. Kuk. Maximum likelihood estimation for probit-linear mixed models with correlated random effects. *Biometrics*, 53(1): 86–97, 1997. ISSN 0006341X.

E. Lesaffre and B. Spiessens. On the effect of the number of quadrature points in a logistic random effects model: an example. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 50:325–335, 2001.

Q. Liu and D. Pierce. A note on Gauss–Hermite quadrature. *Biometrika*, 81 (3):624, 1994.

C. McCulloch. Maximum likelihood variance components estimation for binary data. *Journal of the American Statistical Association*, 89(425):330–335, 1994.

X. L. Meng and S. Schilling. Fitting full-information item factor models and an empirical investigation of bridge sampling. *Journal of the American Statistical Association*, 91(435):1254–1267, 1996.

J. Naylor and A. Smith. Applications of a method for the efficient computation of posterior distributions. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, 31(3):214–225, 1982.

S. Rabe-Hesketh, A. Skrondal, and A. Pickles. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2(1):1–21, 2002.

S. Schilling and R. D. Bock. High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika*, 70(3):533–555, 2005.