University of Wollongong
Research Online

Centre for Statistical & Survey Methodology
Working Paper Series

Faculty of Engineering and Information Sciences

2010

Asymptotic normality and valid inference for Gaussian variational approximation

P. Hall
University of Wollongong

T. Pham
University of Wollongong

M. Wand
University of Wollongong, mwand@uow.edu.au

S.S.J. Wang
University of Wollongong

Recommended Citation

Hall, P.; Pham, T.; Wand, M.; and Wang, S.S.J., Asymptotic normality and valid inference for Gaussian variational approximation, Centre for Statistical and Survey Methodology, University of Wollongong, Working Paper 04-10, 2010, 24p.
http://ro.uow.edu.au/cssmwp/55

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Centre for Statistical and Survey Methodology

The University of Wollongong

Working Paper

04-10

ASYMPTOTIC NORMALITY AND VALID INFERENCE FOR GAUSSIAN VARIATIONAL APPROXIMATION

Hall, P., Pham, T., Wand,M.P. and Wang, S.S.J.

Copyright © 2008 by the Centre for Statistical & Survey Methodology, UOW. Work in progress, no part of this paper may be reproduced without permission from the Centre.

Centre for Statistical & Survey Methodology, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Submitted to the Annals of Statistics
arXiv: math.PR/0000000

ASYMPTOTIC NORMALITY AND VALID INFERENCE FOR GAUSSIAN VARIATIONAL APPROXIMATION

By Peter Hall *, Tung Pham † M.P. Wand † and S.S.J. Wang †

University of Melbourne and University of Wollongong†*

9th November, 2010

We derive the precise asymptotic distributional behavior of Gaussian variational approximate estimators of the parameters in a single-predictor Poisson mixed model. These results are the deepest yet obtained concerning the statistical properties of a variational approximation method. Moreover, they give rise to asymptotically valid statistical inference. A simulation study demonstrates that Gaussian variational approximate confidence intervals possess good to excellent coverage properties, and with precision similar to their exact likelihood counterparts.

1. Introduction. Variational approximation methods are enjoying an increasing amount of development and use in statistical problems. This raises questions regarding their statistical properties, such as consistency of point estimators and validity of statistical inference. We make significant inroads into answering such questions via thorough theoretical treatment of one of the simplest non-trivial settings for which variational approximation is beneficial: the Poisson mixed model with a single predictor variable and random intercept. We call this the model *simple Poisson mixed model*.

The model treated here is also treated in [5], but there attention is confined to bounds and rates of convergence. We improve upon their results by obtaining the asymptotic distributions of the estimators. The results reveal that the estimators are asymptotically normal, have negligible bias, and that their variances decay at least as fast as m^{-1}, where m is the number of groups. For the slope parameter, the faster $(mn)^{-1}$ rate is obtained, where n is the number of repeated measures.

An important practical ramification of our theory is asymptotically valid statistical inference for the model parameters. In particular, a form of studentization leads to theoretically justifiable confidence intervals for all model parameters. Unlike those based on the exact likelihood, all Gaussian variational approximate point estimates and confidence intervals can be computed without the need for numerical integration. Simulation results reveal that the confidence intervals have good to excellent coverage and have about the same length as exact likelihood-based intervals.

Variational approximation methodology is now a major research area with Computer Science; see, for example, Chapter 10 of [2]. It is beginning to have a presence in Statistics as well (e.g. [8], [11]) A summary of the topic from a statistical perspective is given in [10]. Late 2010 saw the first non-beta release of a software library, Infer.NET [9], for facilitation of variational approximate inference. A high proportion of variational approximation methodology is framed within Bayesian hierarchical structures and offers itself as a faster alternative to Markov chain Monte Carlo methods. The chief driving force is applications where speed is at a premium and some accuracy can be sacrificed. Examples of such applications are cluster analysis of gene-expression data [13], fitting spatial models to neuroimage data [4], image segmentation [3] and

AMS 2000 subject classifications: Primary 62F12; secondary 62F25
Keywords and phrases: Generalized linear mixed models, Longitudinal data analysis, Maximum likelihood estimation, Poisson mixed models

genome-wide association analysis [6].

As explained in [2] and [10], there are many types of variational approximations. The most popular is *variational Bayes* (also known as *mean field* approximation), which relies on product restrictions applied to the joint posterior densities of a Bayesian model. The present article is concerned with *Gaussian* variational approximation in frequentist models containing random effects. There are numerous models of this general type. One of their hallmarks is the difficulty of exact likelihood-based inference for the model parameters due to presence of non-analytic integrals. Generalized linear mixed models (e.g. Chapter 7 of [7]) form a large class of models for handling within-group correlation when the response variable is non-Gaussian. The simple Poisson mixed model lies within this class. From a theoretical standpoint, the simple Poisson mixed model is attractive because it possesses the computational challenges that motivate Gaussian variational approximation — exact likelihood-based inference requires quadrature — but its simplicity makes it amenable to deep theoretical treatment. We take advantage of this simplicity to derive the asymptotic distribution of the Gaussian variational approximate estimators, although the derivations are still quite intricate and involved. These results represent the deepest statistical theory yet obtained for a variational approximation method.

Moreover, for the first time, asymptotically valid inference for a variational approximation method is manifest. Our theorem reveals that each estimator is asymptotically normal, centered on the true parameter value and with a studentizable variance. Replacement of the unknown quantities by consistent estimators results in asymptotically valid confidence intervals and Wald hypothesis tests. A simulation study shows that Gaussian variational approximate confidence intervals possess good to excellent coverage properties — especially in the case of the slope parameter.

Section 2 describes the simple Poisson mixed model and Gaussian variational approximation. An asymptotic normality theorem is presented in Section 3. In Section 4 we discuss the implications for valid inference and perform some numerical evaluations. Section 5 contains the proof of the theorem.

2. Gaussian variational approximation for the simple Poisson mixed model. The simple Poisson mixed model that we study here is identical to that treated in [5]. Section 2 of that paper provides a detailed description of the model and the genesis of Gaussian variational approximation for estimation of the model parameters. Here we give just a rudimentary account of the model and estimation strategy.

The simple Poisson mixed model is

$$Y_{ij}|X_{ij}, U_i \text{ independent Poisson with mean } \exp(\beta_0^0 + \beta_1^0 X_{ij} + U_i), \tag{2.1}$$

$$U_i \text{ independent } N(0, \sigma_0^2). \tag{2.2}$$

The X_{ij} and U_i, for $1 \le i \le m$ and $1 \le j \le n$, are totally independent random variables, with the X_{ij}s distributed as X. We observe values of (X_{ij}, Y_{ij}), $1 \le i \le m$, $1 \le j \le n$, whilst the U_i are unobserved latent variables. See, for example, Chapter 7 and Section 14.3 of [7] for further details on this model and its use in longitudinal data analysis. In applications it is typically the case that $m \gg n$.

Let $\boldsymbol{\beta} \equiv (\beta_0, \beta_1)$ be the vector of fixed effects parameters. The conditional log-likelihood of $(\boldsymbol{\beta}, \sigma^2)$ is the logarithm of the joint probability mass function of the Y_{ij}'s, given the X_{ij}'s, as a function of the parameters:

$$\begin{aligned}\ell(\boldsymbol{\beta}, \sigma^2) = &\sum_{i=1}^m \sum_{j=1}^n \{Y_{ij}(\beta_0 + \beta_1 X_{ij}) - \log(Y_{ij}!)\} - \tfrac{m}{2}\log(2\pi\sigma^2) \\ &+ \sum_{i=1}^m \log \int_{-\infty}^{\infty} \exp\left(\sum_{j=1}^n Y_{ij} u - e^{\beta_0 + \beta_1 X_{ij} + u} - \frac{u^2}{2\sigma^2}\right) du.\end{aligned} \tag{2.3}$$

Maximum likelihood estimation is hindered by the presence of m intractable integrals in (2.3). One of several ways around this is *Gaussian variational approximation*, in which (2.3) is replaced by:

$$\underline{\ell}(\boldsymbol{\beta},\sigma^2,\boldsymbol{\mu},\boldsymbol{\lambda}) = \sum_{i=1}^{m}\sum_{j=1}^{n}\{Y_{ij}(\beta_0+\beta_1 X_{ij}+\mu_i) - e^{\beta_0+\beta_1 X_{ij}+\mu_i+\frac{1}{2}\lambda_i} - \log(Y_{ij}!)\} - \frac{m}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(\mu_i^2+\lambda_i) + \tfrac{1}{2}\sum_{i=1}^{m}\log(\lambda_i) \tag{2.4}$$

and the vectors $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_m)$ and $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_m)$ contain introduced *variational* parameters. The Gaussian variational approximate maximum likelihood estimators are:

$$(\underline{\widehat{\boldsymbol{\beta}}},\underline{\widehat{\sigma}}^2) = (\boldsymbol{\beta},\sigma^2) \text{ component of } \operatorname*{argmax}_{\boldsymbol{\beta},\sigma^2,\boldsymbol{\mu},\boldsymbol{\lambda}} \underline{\ell}(\boldsymbol{\beta},\sigma^2,\boldsymbol{\mu},\boldsymbol{\lambda}).$$

3. Asymptotic normality results. Consider random variables (X_{ij}, Y_{ij}, U_i) satisfying (2.1) and (2.2). Put

$$Y_{i\bullet} = \sum_{i=1}^{n} Y_{ij} \quad \text{and} \quad B_i = \sum_{j=1}^{n}\exp(\beta_0+\beta_1 X_{ij})$$

and consider the following decompositions of the exact log-likelihood and its Gaussian variational approximation:

$$\begin{aligned}\ell(\boldsymbol{\beta},\sigma^2) &= \ell_0(\boldsymbol{\beta},\sigma^2) + \ell_1(\boldsymbol{\beta},\sigma^2) + \text{DATA}\,,\\ \underline{\ell}(\boldsymbol{\beta},\sigma^2,\boldsymbol{\mu},\boldsymbol{\lambda}) &= \ell_0(\boldsymbol{\beta},\sigma^2) + \ell_2(\boldsymbol{\beta},\sigma^2,\boldsymbol{\mu},\boldsymbol{\lambda}) + \text{DATA}\,,\end{aligned}$$

where

$$\ell_0(\boldsymbol{\beta},\sigma^2) = \sum_{i=1}^{m}\sum_{j=1}^{n} Y_{ij}\,(\beta_0+\beta_1 X_{ij}) - \tfrac{1}{2}\,m\,\log\sigma^2\,, \tag{3.1}$$

$$\ell_1(\boldsymbol{\beta},\sigma^2) = \sum_{i=1}^{m}\log\left\{\int_{-\infty}^{\infty}\exp\left(Y_{i\bullet}\,u - B_i\,e^u - \tfrac{1}{2}\,\sigma^{-2}\,u^2\right)du\right\},$$

$$\ell_2(\boldsymbol{\beta},\sigma^2,\boldsymbol{\mu},\boldsymbol{\lambda}) = \sum_{i=1}^{m}\left\{\mu_i\,Y_{i\bullet} - B_i\,\exp\left(\mu_i+\tfrac{1}{2}\,\lambda_i\right)\right\} - \tfrac{1}{2}\,\sigma^{-2}\sum_{i=1}^{m}\left(\mu_i^2+\lambda_i\right) + \tfrac{1}{2}\sum_{i=1}^{m}\log\lambda_i\,. \tag{3.2}$$

and DATA denotes a quantity depending on the Y_{ij} alone, and not on $\boldsymbol{\beta}$ or σ^2. Note that

$$\underline{\ell}(\boldsymbol{\beta},\sigma^2) = \max_{\boldsymbol{\mu},\boldsymbol{\lambda}}\underline{\ell}(\boldsymbol{\beta},\sigma^2,\boldsymbol{\mu},\boldsymbol{\lambda}) = \underline{\ell}_0(\boldsymbol{\beta},\sigma^2) + \max_{\boldsymbol{\mu},\boldsymbol{\lambda}}\underline{\ell}_2(\boldsymbol{\beta},\sigma^2,\boldsymbol{\mu},\boldsymbol{\lambda}).$$

Our upcoming theorem relies on the following assumptions:

(A1) the moment generating function of X, $\phi(t) = E\{\exp(tX)\}$, is well-defined on the whole real line;
(A2) the mapping that takes β to $\phi'(\beta)/\phi(\beta)$ is invertible;
(A3) in some neighbourhood of β_1^0 (the true value of β_1), $(d^2/d\beta^2)\log\phi(\beta)$ does not vanish;
(A4) $m = m(n)$ diverges to infinity with n, such that $n/m \to 0$ as $n \to \infty$;
(A5) and, for a constant $C > 0$, $m = O(n^C)$ as m and n diverge.

Define

$$\tau^2 = \frac{\exp\{-\frac{1}{2}\sigma_0^2 - \beta_0^0\}\,\phi(\beta_1^0)}{\phi''(\beta_1^0)\phi(\beta_1^0) - \phi'(\beta_1^0)^2}\,. \tag{3.3}$$

The precise asymptotic behavior of $\underline{\widehat{\beta}}_0$, $\underline{\widehat{\beta}}_1$ and $\underline{\widehat{\sigma}}^2$ is conveyed by:

THEOREM 3.1. *Assume that conditions (A1)–(A5) hold. Then:*

$$\underline{\widehat{\beta}}_0 - \beta_0^0 = m^{-1/2}\,N_0 + o_p(n^{-1} + m^{-1/2}), \tag{3.4}$$

where the random variable N_0 is normal $N(0, \sigma_0^2)$;

$$\underline{\widehat{\beta}}_1 - \beta_1^0 = (mn)^{-1/2}\,N_1 + o_p\{n^{-2} + (mn)^{-1/2}\}, \tag{3.5}$$

where the random variable N_1 is normal $N(0, \tau^2)$; and

$$\underline{\widehat{\sigma}}^2 - \sigma_0^2 = m^{-1/2}\,N_2 + o_p(n^{-1} + m^{-1/2}), \tag{3.6}$$

where the random variable N_2 is normal $N(0, 2\{\sigma_0^2\}^2)$.

REMARK. All three Gaussian variational approximate estimators have asymptotically normal distibutions with asymptotically negligible bias. The estimators $\underline{\widehat{\beta}}_0$ and $\underline{\widehat{\sigma}}^2$ have variances of size m^{-1}, as m and n diverge in such a manner that $n/m \to 0$. The estimator $\underline{\widehat{\beta}}_1$ has variance of size $(mn)^{-1}$. Hence, the estimator $\underline{\widehat{\beta}}_1$ is distinctly more accurate than either $\underline{\widehat{\beta}}_0$ or $\underline{\widehat{\sigma}}^2$, since it converges to the respective true parameter value at a strictly faster rate. For the estimator $\underline{\widehat{\beta}}_1$, increasing both m and n reduces variance. However, in the cases of the estimators $\underline{\widehat{\beta}}_0$ or $\underline{\widehat{\sigma}}^2$ only an increase in m reduces variance.

4. Asymptotically valid inference. Theorem 3.1 reveals that $\underline{\widehat{\beta}}_0$, $\underline{\widehat{\beta}}_1$ and $\underline{\widehat{\sigma}}^2$ are each asymptotically normal with means corresponding to the true parameter values. The variances depend on known functions of the parameters and $\phi(\beta_1^0)$, $\phi'(\beta_1^0)$ and $\phi''(\beta_1^0)$. Since the latter three quantities can be estimated unbiasedly via

$$\begin{aligned} \widehat{\phi(\beta_1^0)} &= \tfrac{1}{mn}\textstyle\sum_{i=1}^m \sum_{j=1}^n \exp(X_{ij}\underline{\widehat{\beta}}_1), \\ \widehat{\phi'(\beta_1^0)} &= \tfrac{1}{mn}\textstyle\sum_{i=1}^m \sum_{j=1}^n X_{ij}\,\exp(X_{ij}\underline{\widehat{\beta}}_1) \\ \text{and}\quad \widehat{\phi''(\beta_1^0)} &= \tfrac{1}{mn}\textstyle\sum_{i=1}^m \sum_{j=1}^n X_{ij}^2\,\exp(X_{ij}\underline{\widehat{\beta}}_1) \end{aligned}$$

we can consistently estimate the asymptotic variances for inferential procedures such as confidence intervals and Wald-type hypothesis tests. For example, the quantity τ^2 appearing in the expression for the asymptotic variance of $\underline{\widehat{\beta}}_1$ can be consistently estimated by

$$\widehat{\tau^2} = \frac{\exp\big(-\frac{1}{2}\underline{\widehat{\sigma}}^2 - \underline{\widehat{\beta}}_0\big)\,\widehat{\phi(\beta_1^0)}}{\widehat{\phi''(\beta_1^0)}\widehat{\phi(\beta_1^0)} - \widehat{\phi'(\beta_1^0)}^2}.$$

Approximate $100(1-\alpha)\%$ confidence intervals for β_0^0, β_1^0 and σ_0^2 are

$$\underline{\widehat{\beta}}_0 \pm \Phi(1-\tfrac{1}{2}\alpha)\sqrt{\frac{\underline{\widehat{\sigma}}^2}{m}},\quad \underline{\widehat{\beta}}_0 \pm \Phi(1-\tfrac{1}{2}\alpha)\sqrt{\frac{\widehat{\tau^2}}{mn}}\quad\text{and}\quad \underline{\widehat{\sigma}}^2 \pm \Phi(1-\tfrac{1}{2}\alpha)\,\underline{\widehat{\sigma}}^2\sqrt{\frac{2}{m}}. \tag{4.1}$$

where Φ denotes the $N(0,1)$ distribution function. These confidence intervals are *asymptotically valid* since they involve studentization based on consistent estimators of all unknown quantities.

We ran a simulation study to evaluate the coverage properties of the Gaussian variational approximate confidence intervals (4.1). The true parameter vector $(\beta_0^0, \beta_1^0, \sigma_0^2)$ was allowed to vary over $\{(-0.3, 0.2, 0.5), (2.2, -0.1, 0.16), (1.2, 0.4, 0.1), (0.02, 1.3, 1), (-0.3, 0.2, 0.1)\}$ and the distribution of the X_{ij} was taken to be either $N(0,1)$ or Uniform$(-1,1)$, the uniform distribution over the interval $(-1,1)$. The number groups m varied over $100, 200, \ldots, 1000$ with n fixed at $m/10$ throughout the study. For each of the ten possible combinations of true parameter vector and X_{ij} distribution, and sample size pairs, we generated 1000 samples and computed 95% confidence intervals based on (4.1).

Figure 4 shows the actual coverage percentages for the nominally 95% confidence intervals. In the case of β_1^0, the actual and nominal percentages are seen to have very good agreement — even for $(m,n) = (100, 10)$. This is also the case for β_0^0 for the first four true parameter vectors. For the fifth one, which has a relatively low amount of within-subject correlation, the asymptotics take a bit longer to become apparent and we see that $m \geq 400$ is required to get the actual coverage above 90%, i.e. within 5% of the nominal level. For σ_0^2, a similar comment applies, but with $m \geq 800$. The superior coverage of the β_1^0 confidence intervals is in keeping with the faster convergence rate apparent from Theorem 3.1.

Lastly, we ran a smaller simulation study to check whether or not the lengths of the Gaussian variational approximate confidence intervals are compromised in achieving the good coverage apparent in Figure 4. For each of the same settings used to produce that figure we generated 100 samples and computed the exact likelihood-based confidence intervals using adaptive Gauss-Hermite quadrature (via the R language [12] package `lme4` [1]). In almost every case, the Gaussian variational approximate confidence intervals were slightly shorter than their exact counterparts. This reassuring result indicates that the good coverage performance is not accompanied by a decrease in precision.

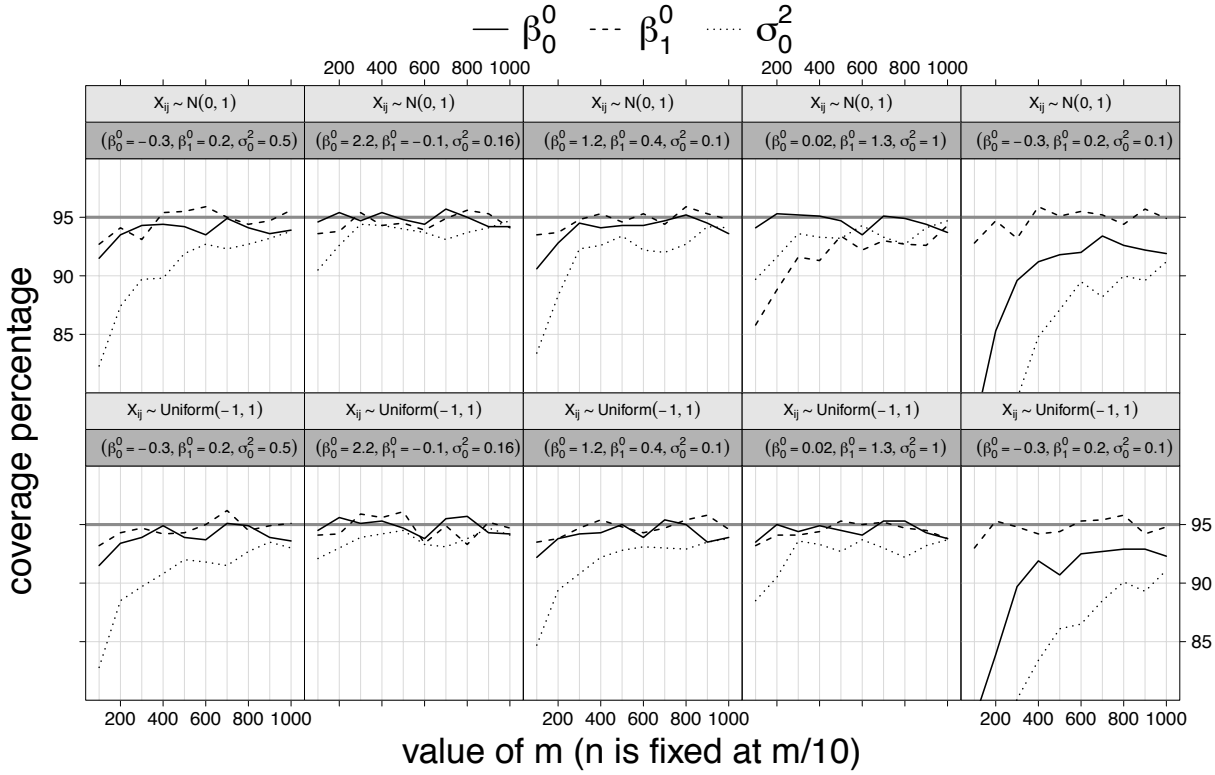

FIG 1. *Actual coverage percentage of nominally 95% Gaussian variational approximate confidence intervals for the parameters in the simple Poisson mixed model. The nominal percentage is shown as a thick grey horizontal line. The percentages are based on 1000 replications. The values of* m *are* $100, 200, \ldots, 1000$. *The value of* n *is fixed at* $n = m/10$.

TABLE 1
Definitions of the $O_{(k)}$ notation used in the proofs.

notation	meaning
$O_{(1)}$	$O_p(m^{-1/2} + n^{-1})$
$O_{(2)}$	$O_p(m^{-1} + n^{-2})$
$O_{(3)}$	$O(n^{\varepsilon-(1/2)})$, uniformly in $1 \le i \le m$, for each $\varepsilon > 0$
$O_{(4)}$	$O(n^{\varepsilon-1})$, uniformly in $1 \le i \le m$, for each $\varepsilon > 0$
$O_{(5)}$	$O(n^{\varepsilon-(3/2)})$, uniformly in $1 \le i \le m$, for each $\varepsilon > 0$
$O_{(6)}$	$O_p(m^{-1} + n^{\varepsilon-(3/2)})$, uniformly in $1 \le i \le m$, for each $\varepsilon > 0$
$O_{(7)}$	$O_p\{(m^{-1} + n^{-2})\, n^{\varepsilon-(1/2)}\}$, uniformly in $1 \le i \le m$, for each $\varepsilon > 0$
$O_{(8)}$	$O_p\{(m^{-1/2} + n^{-1})^3\, n^{\varepsilon}\}$, uniformly in $1 \le i \le m$, for each $\varepsilon > 0$
$O_{(9)}$	$O_p\{(mn)^{-1/2} + n^{\varepsilon-(3/2)}\}$, uniformly in $1 \le i \le m$, for each $\varepsilon > 0$
$O_{(10)}$	$O_p\{(m^{-1/2} + n^{-5/2})\, n^{\varepsilon}\}$, uniformly in $1 \le i \le m$, for each $\varepsilon > 0$
$O_{(11)}$	$O_p\{(m^{-1/2}\, n^{-1} + n^{-2})\, n^{\varepsilon}\}$, uniformly in $1 \le i \le m$, for each $\varepsilon > 0$

5. Proof of Theorem 3.1. The proof Theorem 3.1 requires some additional notation, as well as several stages of asymptotic approximation. This section provides full details, beginning with definitions of the necessary notation.

5.1. *Notation.* Recall that β_0^0, β_1^0 and σ_0^2 denote the true values of parameters and that $\underline{\widehat{\beta}}_0$, $\underline{\widehat{\beta}}_1$ and $\underline{\widehat{\sigma}}^2$ denote their respective Gaussian variational approximate estimators.

The proofs use "$O_{(k)}$" notation, for $k = 1, \ldots, 11$, as defined in Table 1.

5.2. *Formulae for estimators.* First we give, in (5.1)–(5.5) below, the results of equating to zero the derivatives of $\ell_0(\beta, \sigma^2) + \ell_2(\beta, \sigma^2, \lambda, \mu)$ with respect to β_0, β_1, σ^2, λ_i and μ_i, respectively:

$$\sum_{i=1}^{m} \{Y_{i\bullet} - B_i \exp(\underline{\widehat{\mu}}_i + \tfrac{1}{2}\underline{\widehat{\lambda}}_i)\} = 0\,, \tag{5.1}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}\left\{Y_{ij} - \exp\left(\underline{\widehat{\beta}}_0 + \underline{\widehat{\mu}}_i + \tfrac{1}{2}\underline{\widehat{\lambda}}_i + \underline{\widehat{\beta}}_1 X_{ij}\right)\right\} = 0\,, \tag{5.2}$$

$$\frac{1}{m}\sum_{i=1}^{m}\left(\underline{\widehat{\lambda}}_i + \underline{\widehat{\mu}}_i^2\right) = \underline{\widehat{\sigma}}^2\,, \tag{5.3}$$

$$\underline{\widehat{\lambda}}_i^{-1} - B_i \exp(\underline{\widehat{\mu}}_i + \tfrac{1}{2}\underline{\widehat{\lambda}}_i) - (\underline{\widehat{\sigma}}^2)^{-1} = 0\,, \quad 1 \le i \le m, \tag{5.4}$$

$$Y_{i\bullet} - B_i \exp(\underline{\widehat{\mu}}_i + \tfrac{1}{2}\underline{\widehat{\lambda}}_i) - (\underline{\widehat{\sigma}}^2)^{-1}\,\underline{\widehat{\mu}}_i = 0\,, \quad 1 \le i \le m. \tag{5.5}$$

These are the analogues of the likelihood equations in the conventional approach to inference.

The next step is to put (5.1), (5.2) and (5.5) into more accessible form, in (5.6), (5.11) and (5.12), respectively. Adding (5.5) over $1 \le i \le m$ and subtracting the result from (5.1) we deduce that

$$\sum_{i=1}^{m} \underline{\widehat{\mu}}_i = 0\,. \tag{5.6}$$

Defining

$$\Delta = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}\,\{Y_{ij} - \exp(\beta_0^0 + \beta_1^0 X_{ij} + U_i)\}$$

we deduce that (5.2) is equivalent to:

$$\begin{aligned}\Delta + \exp(\beta_0^0)\,\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}\,\exp(U_i + \beta_1^0 X_{ij}) \\ - \exp(\beta_0)\,\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}\,\exp(\underline{\widehat{\mu}}_i + \tfrac{1}{2}\underline{\widehat{\lambda}}_i + \beta_1 X_{ij}) = 0\,.\end{aligned} \tag{5.7}$$

Define ξ_i, η_i and ζ_i by, respectively,

$$\frac{1}{n}\sum_{j=1}^{n} X_{ij}\,\exp\left(\beta_1^0\,X_{ij}\right) \;=\; \phi'(\beta_1^0)\,\exp(\xi_i)\,, \tag{5.8}$$

$$\frac{1}{n}\sum_{j=1}^{n} X_{ij}\,\exp\left(\underline{\widehat{\beta}}_1\,X_{ij}\right) \;=\; \phi'(\underline{\widehat{\beta}}_1)\,\exp\left(\eta_i\right), \tag{5.9}$$

$$\begin{aligned}&\exp\left(\underline{\widehat{\beta}}_0+\underline{\widehat{\mu}}_i+\tfrac{1}{2}\,\underline{\widehat{\lambda}}_i\right)\frac{1}{n}\sum_{j=1}^{n}\{\exp(\underline{\widehat{\beta}}_1\,X_{ij})-\phi(\underline{\widehat{\beta}}_1)\}=\exp\left(\beta_0^0+U_i\right)\\ &\times\Bigg[\phi(\beta_1^0)\,\{1-\exp(\zeta_i)\}+\frac{1}{n}\sum_{j=1}^{n}\left\{Y_{ij}\,\exp\left(-\beta_0^0-U_i\right)-\phi(\beta_1^0)\right\}\Bigg]-(\underline{\widehat{\sigma}}^2 n)^{-1}\,\underline{\widehat{\mu}}_i.\end{aligned} \tag{5.10}$$

With probability converging to 1 as $n\to\infty$ the definitions at (5.8)–(5.10) are valid simultaneously for all $1\le i\le m$, because the variables ξ_i, η_i and ζ_i so defined converge to zero, uniformly in $1\le i\le m$, in probability. See (5.30), (5.31) and (5.25) below for approximations to ξ_i, η_i and ζ_i; indeed, those formulae quickly imply that each of ξ_i, η_i and ζ_i equals $O_{(3)}$.

Without loss of generality, $\phi'(t)$ is bounded away from zero in a neighbourhood of β_1^0. Indeed, if the latter property does not hold, simply add a constant to the random variable X to ensure that $\phi'(\beta_1^0)\neq 0$. We assume that β_1^0 is in the just-mentioned neighbourhood, and we consider only realizations for which β_1 is also in the neighbourhood. (The latter property holds true with probability converging to 1 as $n\to\infty$.) The definition of ζ_i at (5.10) can be justified using the fact that $\underline{\widehat{\mu}}_i < Y_{i\bullet}$, as shown in Theorem 2 of [5].

In this notation we can write (5.7) as:

$$\Delta+\phi'(\beta_1^0)\,\frac{1}{m}\sum_{i=1}^{m}\exp(\beta_0^0+U_i+\xi_i)=\phi'(\underline{\widehat{\beta}}_1)\,\frac{1}{m}\sum_{i=1}^{m}\exp\left(\underline{\widehat{\beta}}_0+\underline{\widehat{\mu}}_i+\tfrac{1}{2}\,\underline{\widehat{\lambda}}_i+\eta_i\right), \tag{5.11}$$

and write (5.5) as:

$$\exp\left(\underline{\widehat{\beta}}_0+\underline{\widehat{\mu}}_i+\tfrac{1}{2}\,\underline{\widehat{\lambda}}_i\right)\phi(\beta_1)=\exp\left(\beta_0^0+U_i+\zeta_i\right)\phi(\beta_1^0)\,. \tag{5.12}$$

Substituting (5.12) into (5.11) we obtain:

$$\begin{aligned}&\Delta\,\exp\left(-\beta_0^0\right)\phi(\beta_1^0)^{-1}+\phi'(\beta_1^0)\,\phi(\beta_1^0)^{-1}\,\frac{1}{m}\sum_{i=1}^{m}\exp(U_i+\xi_i)\\ &\qquad=\phi'(\underline{\widehat{\beta}}_1)\,\phi(\underline{\widehat{\beta}}_1)^{-1}\frac{1}{m}\sum_{i=1}^{m}\exp(U_i+\eta_i+\zeta_i)\,.\end{aligned} \tag{5.13}$$

5.3. *Approximate formulae for U_i and $\underline{\widehat{\lambda}}_i$.* The formulae are given at (5.16) and (5.18), respectively. To derive them, note that (5.5) implies that

$$(1+O_{(3)})\,\phi(\beta_1^0)\,\exp\left(\beta_0^0+U_i\right)-(1+O_{(3)})\,\phi(\beta_1^0)\,\exp\left(\beta_0^0+\underline{\widehat{\mu}}_i+\tfrac{1}{2}\,\underline{\widehat{\lambda}}_i\right)-(n\underline{\widehat{\sigma}}^2)^{-1}\,\underline{\widehat{\mu}}_i=0\,.$$

Here we have used the fact that, by [5],

$$\underline{\widehat{\beta}}_0-\beta_0^0=O_{(1)}\,,\quad \underline{\widehat{\beta}}_1-\beta_1^0=O_{(1)}\,, \tag{5.14}$$

and that by (1.3), $\max_{1\le i\le m}|X_i|=O_p(n^{\varepsilon})$ for all $\varepsilon>0$. Therefore,

$$(1+O_{(3)})\,\exp(U_i)=(1+O_{(3)})\,\exp\left(\underline{\widehat{\mu}}_i+\tfrac{1}{2}\,\underline{\widehat{\lambda}}_i\right)+(cn\underline{\widehat{\sigma}}^2)^{-1}\,\underline{\widehat{\mu}}_i\,, \tag{5.15}$$

where $c = \phi(\beta_1^0)\,\exp(\beta_0^0)$. The result $\max_{1\le i\le m}|U_i| = O_p\{(\log n)^{1/2}\}$ follows from properties of extrema of Gaussian variables and the fact that $m = O(n^C)$ for a constant $C > 0$. Moreover, by Theorem 2 of [5], $0 < \widehat{\underline{\lambda}}_i < \widehat{\underline{\sigma}}^2$. Therefore (5.15) implies that $\max_{1\le i\le n}|\widehat{\underline{\mu}}_i| = O_p\{(\log n)^{1/2}\}$. (Note that, for any constant $C > 0$, $\exp\{-C\,(\log n)^{1/2}\} = n^{-C\,(\log n)^{-1/2}}$, which is of larger order than $n^{-\varepsilon}$ for each $\varepsilon > 0$.) Hence, by (5.15),

$$(1 + O_{(3)})\,\exp(U_i) = (1 + O_{(3)})\,\exp\left(\widehat{\underline{\mu}}_i + \tfrac{1}{2}\,\widehat{\underline{\lambda}}_i\right),$$

and so, taking logarithms,

$$U_i = \widehat{\underline{\mu}}_i + \tfrac{1}{2}\,\widehat{\underline{\lambda}}_i + O_{(3)}\,. \tag{5.16}$$

Formula (5.4) and property (5.14) entail:

$$(n\widehat{\underline{\lambda}}_i)^{-1} - (1 + O_{(3)})\,\phi(\beta_1^0)\,\exp\left(\widehat{\underline{\mu}}_i + \tfrac{1}{2}\,\widehat{\underline{\lambda}}_i + \beta_0^0\right) - (n\widehat{\underline{\sigma}}^2)^{-1} = 0\,. \tag{5.17}$$

Using (5.16) to substitute $U_i + O_{(3)}$ for $\widehat{\underline{\mu}}_i + \frac{1}{2}\,\widehat{\underline{\lambda}}_i$ in (5.17) we deduce from that result that

$$\begin{aligned}(n\widehat{\underline{\lambda}}_i)^{-1} &= (1 + O_{(3)})\,\phi(\beta_1^0)\,\exp\left(U_i + \beta_0^0\right) + (n\widehat{\underline{\sigma}}^2)^{-1}\\ &= (1 + O_{(3)})\,\phi(\beta_1^0)\,\exp\left(U_i + \beta_0^0\right),\end{aligned}$$

where to obtain the second identity we again used the fact that

$$\max_{1\le i\le m}|U_i| = O_p\{(\log n)^{1/2}\}.$$

Therefore,

$$\begin{aligned}\widehat{\underline{\lambda}}_i &= (1 + O_{(3)})\,\{n\,\phi(\beta_1^0)\,\exp\left(U_i + \beta_0^0\right)\}^{-1}\\ &= \{n\,\phi(\beta_1^0)\,\exp\left(U_i + \beta_0^0\right)\}^{-1} + O_{(5)}\,,\end{aligned} \tag{5.18}$$

where $O_{(5)}$ is as defined in Table 1. To obtain the second identity in (5.18) we used the fact that $\max_{1\le i\le m}\exp(-U_i) = O(n^{\varepsilon})$ for all $\varepsilon > 0$.

5.4. *Initial approximations to* $\widehat{\underline{\beta}}_0 - \beta_0^0$ *and* $\widehat{\underline{\beta}}_1 - \beta_1^0$. These approximations are given at (5.19), (5.21) and (5.29), and lead to central limit theorems for $\widehat{\underline{\beta}}_1 - \beta_1^0$, $\widehat{\underline{\beta}}_0 - \beta_0^0$ and $\widehat{\underline{\sigma}}^2 - \sigma_0^2$, respectively. To derive the approximations, write $\gamma(\beta_1) = \phi'(\beta_1)\,\phi(\beta_1)^{-1}$ and note that, defining $O_{(2)}$ as in Table 1, we have:

$$\begin{aligned}\gamma(\widehat{\underline{\beta}}_1) &= \gamma(\beta_1^0) + (\widehat{\underline{\beta}}_1 - \beta_1^0)\,\gamma'(\beta_1^0) + O_p(|\widehat{\underline{\beta}}_1 - \beta_1^0|^2)\\ &= \gamma(\beta_1^0) + \{1 + O_p(m^{-1/2} + n^{-1})\}\,(\widehat{\underline{\beta}}_1 - \beta_1^0)\,\gamma'(\beta_1^0)\,.\end{aligned}$$

(Here we have used (5.14).) Therefore, by (5.13) and for each $\varepsilon > 0$,

$$\begin{aligned}&\Delta\,\exp\left(-\beta_0^0\right)\phi(\beta_1^0)^{-1} + \gamma(\beta_1^0)\,\frac{1}{m}\sum_{i=1}^m \exp(U_i + \xi_i)\\ &\quad = \left[\gamma(\beta_1^0) + \{1 + O_p(m^{-1/2} + n^{-1})\}\,(\widehat{\underline{\beta}}_1 - \beta_1^0)\,\gamma'(\beta_1^0)\right] \times \frac{1}{m}\sum_{i=1}^m \exp(U_i + \eta_i + \zeta_i)\,.\end{aligned}$$

That is,

$$\begin{aligned}&(\widehat{\underline{\beta}}_1 - \beta_1^0)\,\gamma'(\beta_1^0)\,\frac{1}{m}\sum_{i=1}^m \exp(U_i + \eta_i + \zeta_i)\\ &= \gamma(\beta_1^0)\,\frac{1}{m}\sum_{i=1}^m \exp(U_i)\,\{\exp(\xi_i) - \exp\left(\eta_i + \zeta_i\right)\} + \Delta\,\exp\left(-\beta_0^0\right)\phi(\beta_1^0)^{-1} + O_{(2)}.\end{aligned} \tag{5.19}$$

Taking logarithms of both sides of (5.12) we obtain:

$$\log\{\phi(\underline{\widehat{\beta}}_1)/\phi(\beta_1^0)\} = \beta_0^0 - \underline{\widehat{\beta}}_0 + U_i + \zeta_i - \underline{\widehat{\mu}}_i - \tfrac{1}{2}\,\underline{\widehat{\lambda}}_i\,, \tag{5.20}$$

which, on adding over i and dividing by m, implies that

$$\log\{\phi(\underline{\widehat{\beta}}_1)/\phi(\beta_1^0)\} = \beta_0^0 - \underline{\widehat{\beta}}_0 + \frac{1}{m}\sum_{i=1}^m \left(U_i + \zeta_i - \underline{\widehat{\mu}}_i - \tfrac{1}{2}\,\underline{\widehat{\lambda}}_i\right),$$

which in turn gives

$$\begin{aligned}
\underline{\widehat{\beta}}_0 - \beta_0^0 &= -(\underline{\widehat{\beta}}_1 - \beta_1^0)\,\gamma(\beta_1^0) + \frac{1}{m}\sum_{i=1}^m (U_i + \zeta_i - \underline{\widehat{\mu}}_i - \tfrac{1}{2}\,\underline{\widehat{\lambda}}_i) + O_{(2)} \\
&= -(\underline{\widehat{\beta}}_1 - \beta_1^0)\,\gamma(\beta_1^0) + \frac{1}{m}\sum_{i=1}^m (U_i + \zeta_i) \\
&\qquad -\{2n\,\phi(\beta_1^0)\,\exp(\beta_0^0 - \tfrac{1}{2}\,\sigma_0^2)\}^{-1} + O_{(6)}\,,
\end{aligned} \tag{5.21}$$

where we used (5.18) to substitute for $\underline{\widehat{\lambda}}_i$ and (5.6) to eliminate $\underline{\widehat{\mu}}_i$ from the right-hand side, and employed (5.14) to bound $(\underline{\widehat{\beta}}_1 - \beta_1^0)^2$. Note too that $E\{\exp(-U_i)\} = \exp(\frac{1}{2}\,\sigma_0^2)$; a term involving $E\{\exp(-U_i)\}$ arises from $\sum_i \underline{\widehat{\lambda}}_i$ via (5.18).

5.5. *Approximation to ζ_i.* The approximation is given at (5.25). First we derive an expansion, at (5.22) below, of $\underline{\widehat{\mu}}_i$. Reflecting (5.16), define the random variable δ_i by $\underline{\widehat{\mu}}_i = U_i - \frac{1}{2}\,\underline{\widehat{\lambda}}_i + \delta_i$. Then, by (5.16), $\delta_i = O_{(3)}$. Define too $B_{ik}^0 = \sum_j X_{ij}^k \exp(\beta_0^0 + \beta_1^0\,X_{ij})$ for $k = 0, 1, 2$, and $\Delta_i = Y_{i\bullet} - B_{i0}^0 \exp(U_i)$; and let $\mathcal{F}_i$ denote the sigma-field generated by U_i and $X_{i1}, \ldots, X_{in}$. Then $E(\Delta_i \,|\, \mathcal{F}_i) = 0$ and

$$\begin{aligned}
B_i &= \{1 + \underline{\widehat{\beta}}_0 - \beta_0^0 + \tfrac{1}{2}\,(\underline{\widehat{\beta}}_0 - \beta_0^0)^2\}\,B_{i0}^0 + \{\underline{\widehat{\beta}}_1 - \beta_1^0 + (\underline{\widehat{\beta}}_0 - \beta_0^0)\,(\underline{\widehat{\beta}}_1 - \beta_1^0)\}\,B_{i1}^0 \\
&\quad + \tfrac{1}{2}\,(\underline{\widehat{\beta}}_1 - \beta_1^0)^2\,B_{i2} + O_{(8)}\,,
\end{aligned}$$

uniformly in $1 \le i \le m$ for each $\varepsilon > 0$, where $O_{(8)}$ is as in Table 1. Therefore,

$$\begin{aligned}
Y_{i\bullet} - B_i \exp(U_i + \delta_i) &= Y_{i\bullet} - \Big[\{1 + \underline{\widehat{\beta}}_0 - \beta_0^0 + \tfrac{1}{2}\,(\underline{\widehat{\beta}}_0 - \beta_0^0)^2\}\,B_{i0}^0 \\
&\qquad + \{\underline{\widehat{\beta}}_1 - \beta_1^0 + (\underline{\widehat{\beta}}_0 - \beta_0^0)\,(\underline{\widehat{\beta}}_1 - \beta_1^0)\}\,B_{i1}^0 + \tfrac{1}{2}\,(\underline{\widehat{\beta}}_1 - \beta_1^0)^2\,B_{i2}^0\Big] \\
&\qquad \times \exp(U_i)\,(1 + \delta_i + \tfrac{1}{2}\,\delta_i^2 + O_{(5)}) + n\,O_{(8)}\,,
\end{aligned}$$

where $O_{(5)}$ is as in Table 1. Therefore, defining

$$\begin{aligned}
\chi_i &= \{\underline{\widehat{\beta}}_0 - \beta_0^0 + \tfrac{1}{2}\,(\underline{\widehat{\beta}}_0 - \beta_0^0)^2\}\,B_{i0}^0 + \{\underline{\widehat{\beta}}_1 - \beta_1^0 + (\underline{\widehat{\beta}}_0 - \beta_0^0)\,(\underline{\widehat{\beta}}_1 - \beta_1^0)\}\,B_{i1}^0 \\
&\quad + \tfrac{1}{2}\,(\underline{\widehat{\beta}}_1 - \beta_1^0)^2\,B_{i2}^0\,,
\end{aligned}$$

we see that the left-hand side of (5.5) equals:

$$\begin{aligned}
&Y_{i\bullet} - B_i \exp(U_i + \delta_i) - (\underline{\widehat{\sigma}}^2)^{-1}\,\underline{\widehat{\mu}}_i \\
&\quad = \Delta_i - B_{i0}^0 \exp(U_i)\,(\delta_i + \tfrac{1}{2}\,\delta_i^2 + O_{(5)}) - \chi_i \exp(U_i)\,(1 + \delta_i + \tfrac{1}{2}\,\delta_i^2 + O_{(5)}) \\
&\qquad -(\underline{\widehat{\sigma}}^2)^{-1}\,(U_i - \tfrac{1}{2}\,\underline{\widehat{\lambda}}_i + \delta_i) + n\,O_{(8)} \\
&\quad = \Delta_i - \{\chi_i \exp(U_i) + (\underline{\widehat{\sigma}}^2)^{-1}\,(U_i - \tfrac{1}{2}\,\underline{\widehat{\lambda}}_i)\} - \delta_i\,\{(B_{i0}^0 + \chi_i)\,\exp(U_i) + (\underline{\widehat{\sigma}}^2)^{-1}\} \\
&\qquad \tfrac{1}{2}\,\delta_i^2\,(B_{i0}^0 + \chi_i)\,\exp(U_i) + n\,O_{(5)} + n\,O_{(8)}\,.
\end{aligned}$$

Hence, (5.5) implies that:

$$\delta_i + \tfrac{1}{2}\,\delta_i^2\,\frac{(B_{i0}^0+\chi_i)\,\exp(U_i)}{(B_{i0}^0+\chi_i)\,\exp(U_i)+(\underline{\widehat{\sigma}}^2)^{-1}} = \frac{\Delta_i-\chi_i\,\exp(U_i)-(\underline{\widehat{\sigma}}^2)^{-1}\,(U_i-\frac{1}{2}\,\underline{\widehat{\lambda}}_i)}{(B_{i0}^0+\chi_i)\,\exp(U_i)+(\underline{\widehat{\sigma}}^2)^{-1}} + O_{(5)} + O_{(8)}\,,$$

which implies that:

$$\begin{aligned}\delta_i &= \frac{\Delta_i-\chi_i\,\exp(U_i)}{(B_{i0}^0+\chi_i)\,\exp(U_i)} + O_{(4)} = \{n\,\exp\,(\beta_0^0)\,\phi(\beta_1^0)\}^{-1}\,\{\Delta_i\,\exp(-U_i)-\chi_i\} + O_{(4)}\\ &= \{n\,\exp\,(\beta_0^0)\,\phi(\beta_1^0)\}^{-1}\,\Delta_i\,\exp(-U_i) - (\underline{\widehat{\beta}}_0-\beta_0^0) - (\underline{\widehat{\beta}}_1-\beta_1^0)\,\gamma(\beta_1^0) + O_{(4)}\,.\end{aligned}$$

Here we have defined $O_{(4)}$ is as in Table 1 and have used the fact that $n^{-1}\,B_{i0}^0 = \exp(\beta_0^0)\,\phi(\beta_1^0) + O_{(3)}$ and

$$n^{-1}\,B_{i1}^0 = \exp\,(\beta_0^0)\,\phi'(\beta_1^0) + O_{(3)} = \exp\,(\beta_0^0)\,\phi(\beta_1^0)\,\gamma(\beta_1^0) + O_{(3)}\,.$$

Therefore,

$$\begin{aligned}\underline{\widehat{\mu}}_i &= U_i - \tfrac{1}{2}\,\underline{\widehat{\lambda}}_i + \delta_i\\ &= U_i + \{n\,\exp\,(\beta_0^0)\,\phi(\beta_1^0)\}^{-1}\,\Delta_i\,\exp(-U_i) - (\underline{\widehat{\beta}}_0-\beta_0^0) - (\underline{\widehat{\beta}}_1-\beta_1^0)\,\gamma(\beta_1^0) + O_{(4)}\\ &= U_i - \bar{U} + \{n\,\exp\,(\beta_0^0)\,\phi(\beta_1^0)\}^{-1}\,\Delta_i\,\exp(-U_i) + O_{(4)}\,,\end{aligned} \tag{5.22}$$

where to obtain the second identity we used (5.18) to place $\underline{\widehat{\lambda}}_i$ into the remainder, and to obtain the third identity we used (5.21) to show that $\underline{\widehat{\beta}}_0 - \beta_0^0 + (\underline{\widehat{\beta}}_1-\beta_1^0)\,\gamma'(\beta_1^0) = \bar{U} + O_{(4)}$. Here we have used the property, deducible from (5.10), (5.16) and (5.18), that $\zeta_i = O_{(3)}$ and $\bar{\zeta} = O_{(4)}$.

The next step is to substitute the right-hand side of (5.22) for $\underline{\widehat{\mu}}_i$, and the right-hand side of (5.18) for $\underline{\widehat{\lambda}}_i$, in (5.10), and derive an expansion, at (5.25) below, of ζ_i. We obtain:

$$\begin{aligned}&\Big[1 + \{n\,\exp\,(\beta_0^0)\,\phi(\beta_1^0)\}^{-1}\,\Delta_i\,\exp(-U_i) - \bar{U}\Big]\,\frac{1}{n}\sum_{j=1}^n \{\exp(\underline{\widehat{\beta}}_1\,X_{ij}) - \phi(\underline{\widehat{\beta}}_1)\}\\ &\quad = -\phi(\beta_1^0)\,(\zeta_i + \tfrac{1}{2}\,\zeta_i^2) + \frac{1}{n}\sum_{j=1}^n \{Y_{ij}\,\exp\,(-\beta_0^0-U_i) - \phi(\beta_1^0)\}\\ &\qquad - \exp\,(-\beta_0^0-U_i)\,(\underline{\widehat{\sigma}}^2 n)^{-1}\,U_i + O_{(5)}\,,\end{aligned}$$

whence

$$\begin{aligned}&\phi(\beta_1^0)\,(\zeta_i + \tfrac{1}{2}\,\zeta_i^2)\\ &= \frac{1}{n}\sum_{j=1}^n \{Y_{ij}\,\exp\,(-\beta_0^0-U_i) - \phi(\beta_1^0)\} - \frac{1}{n}\sum_{j=1}^n \{\exp(\underline{\widehat{\beta}}_1\,X_{ij}) - \phi(\underline{\widehat{\beta}}_1)\}\\ &\quad - \Big[\{n\,\exp\,(\beta_0^0)\,\phi(\beta_1^0)\}^{-1}\,\Delta_i\,\exp(-U_i) - \bar{U}\Big]\,\frac{1}{n}\sum_{j=1}^n \{\exp\,(\underline{\widehat{\beta}}_1\,X_{ij}) - \phi(\underline{\widehat{\beta}}_1)\}\\ &\quad - \exp\,(-\beta_0^0-U_i)\,(\underline{\widehat{\sigma}}^2 n)^{-1}\,U_i + O_{(5)}\,.\end{aligned} \tag{5.23}$$

However, defining

$$D_{ik}(b) = \frac{1}{n}\sum_{j=1}^n \{X_{ij}^k\,\exp(b\,X_{ij}) - \phi^{(k)}(b)\} = O_{(3)} \tag{5.24}$$

for $k = 0, 1, 2$, and $\Delta_i = Y_{i\bullet} - B_{i0}^0 \exp(U_i)$, we see that

$$\begin{aligned}
& \sum_{j=1}^n \{Y_{ij} \exp(-\beta_0^0 - U_i) - \phi(\beta_1^0)\} - \sum_{j=1}^n \{\exp(\widehat{\underline{\beta}}_1 X_{ij}) - \phi(\widehat{\underline{\beta}}_1)\} \\
= & \sum_{j=1}^n \{Y_{ij} \exp(-\beta_0^0 - U_i) - \phi(\beta_1^0)\} \\
& \quad -n\{D_{i0}(\beta_1^0) + (\widehat{\underline{\beta}}_1 - \beta_1^0) D_{i1}(\beta_1^0)\} + O_{(3)} \\
= & \Delta_i \exp(-\beta_0^0 - U_i) - n(\widehat{\underline{\beta}}_1 - \beta_1^0) D_{i1}(\beta_1^0) + O_{(3)},
\end{aligned}$$

and so, by (5.23),

$$\begin{aligned}
\phi(\beta_1^0)(\zeta_i + \tfrac{1}{2}\zeta_i^2) &= n^{-1} \exp(-\beta_0^0 - U_i) \left[\Delta_i \{1 - \phi(\beta_1^0)^{-1} D_{i0}(\beta_1^0)\} - (\widehat{\underline{\sigma}}^2)^{-1} U_i\right] \\
& \quad -(\widehat{\underline{\beta}}_1 - \beta_1^0) D_{i1}(\beta_1^0) + \bar{U} D_{i0}(\beta_1^0) + O_{(5)}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\phi(\beta_1^0)\zeta_i &= n^{-1} \exp(-\beta_0^0 - U_i) \left[\Delta_i \{1 - \phi(\beta_1^0)^{-1} D_{i0}(\beta_1^0)\} - (\widehat{\underline{\sigma}}^2)^{-1} U_i\right] \\
& \quad -(\widehat{\underline{\beta}}_1 - \beta_1^0) D_{i1}(\beta_1^0) + \bar{U} D_{i0}(\beta_1^0) \\
& \quad -\tfrac{1}{2}\phi(\beta_1^0)^{-1} \{n^{-1} \exp(-\beta_0^0 - U_i)\Delta_i\}^2 + O_{(5)} \qquad (5.25)
\end{aligned}$$

Result (5.25), and the fact that $n/m \to 0$ as $n \to \infty$, imply that

$$\begin{aligned}
\phi(\beta_1^0)\frac{1}{m}\sum_{i=1}^m U_i\zeta_i &= -\frac{1}{mn}\frac{\exp(-\beta_0^0)}{(\sigma^2)^0}\sum_{i=1}^m U_i^2 \exp(-U_i) \\
& \quad -\frac{1}{2m}\phi(\beta_1^0)^{-1}\sum_{i=1}^m U_i\Big\{n^{-1}\exp(-\beta_0^0 - U_i)\Delta_i\Big\}^2 + o_p(n^{-1}) \\
&= -\frac{1}{n}\exp\{\tfrac{1}{2}(\sigma^2)^0 - \beta_0^0\}(1 + \tfrac{1}{2}(\sigma^2)^0) + o_p(n^{-1}) \qquad (5.26)
\end{aligned}$$

Here we have used the fact that $E\{U_i^2 \exp(-U_i)\} = \exp(\frac{1}{2}\sigma_0^2)\,\sigma_0^2\,(1 + \sigma_0^2)$.

5.6. *Initial approximation to $\sigma^2 - \sigma_0^2$.* Starting from (5.20); using (5.21) to substitute for $\widehat{\underline{\beta}}_0 - \beta_0^0$; using (5.18) to substitute for $\widehat{\underline{\lambda}}_i$; and defining $\bar{U} = m^{-1}\sum_i U_i$ and $\bar{\zeta} = m^{-1}\sum_i \zeta_i$; we obtain:

$$\begin{aligned}
\widehat{\underline{\mu}}_i &= U_i + \zeta_i - \tfrac{1}{2}\widehat{\underline{\lambda}}_i - \log\{\phi(\widehat{\underline{\beta}}_1)/\phi(\beta_1^0)\} - (\widehat{\underline{\beta}}_0 - \beta_0^0) \\
&= U_i + \zeta_i - \tfrac{1}{2}\widehat{\underline{\lambda}}_i - (\widehat{\underline{\beta}}_1 - \beta_1^0)\gamma(\beta_1^0) - (\widehat{\underline{\beta}}_0 - \beta_0^0) + O_{(2)} \\
&= U_i + \zeta_i - \{2n\,\phi(\beta_1^0)\exp(U_i + \beta_0^0)\}^{-1} - (\bar{U} + \bar{\zeta}) \\
& \quad +\{2n\,\phi(\beta_1^0)\exp(\beta_0^0 - \tfrac{1}{2}\sigma_0^2)\}^{-1} + O_{(6)}. \qquad (5.27)
\end{aligned}$$

Hence, squaring both sides of (5.27) and adding,

$$\begin{aligned}
\frac{1}{m}\sum_{i=1}^m \widehat{\underline{\mu}}_i^2 &= \frac{1}{m}\sum_{i=1}^m (U_i + \zeta_i - \bar{U} - \bar{\zeta})^2 \\
& \quad -\{mn\phi(\beta_1^0)\exp(\beta_0^0)\}^{-1}\sum_{i=1}^m \exp(-U_i)(U_i + \zeta_i - \bar{U} - \bar{\zeta}) + O_{(6)}. \qquad (5.28)
\end{aligned}$$

Combining (5.3), (5.18), (5.25) and (5.28) we deduce that:

$$\begin{aligned}
\widehat{\underline{\sigma}}^2 &= \frac{1}{m}\sum_{i=1}^m (\widehat{\underline{\lambda}}_i + \widehat{\underline{\mu}}_i^2) = \sigma_0^2 + \frac{1}{m}\sum_{i=1}^m \{(U_i + \zeta_i - \bar{U} - \bar{\zeta})^2 - \sigma_0^2\} \\
& \quad +\{n\,\phi(\beta_1^0)\exp(\beta_0^0 - \tfrac{1}{2}\sigma_0^2)\}^{-1}(1 + \sigma_0^2) + O_{(6)}. \qquad (5.29)
\end{aligned}$$

5.7. *Approximations to ξ_i and η_i.* The approximations are given at (5.30) and (5.31), respectively, and are derived as follows. Note the definition of $D_{ik}(b)$ at (5.24). In that notation, observing that $n/m \to 0$ and recalling (5.14), it can be deduced from (5.8) and (5.9) that, uniformly in $1 \le i \le m$,

$$\xi_i = \phi'(\beta_1^0)^{-1} D_{i1}(\beta_1^0) - \tfrac{1}{2}\{\phi'(\beta_1^0)^{-1} D_{i1}(\beta_1^0)\}^2 + O_{(5)}\,, \tag{5.30}$$

$$\begin{aligned} \eta_i &= \phi'(\beta_1^0)^{-1}\Big[D_{i1}(\beta_1^0) + (\underline{\widehat{\beta}}_1 - \beta_1^0)\Big\{D_{i2}(\beta_1^0) - \phi'(\beta_1^0)^{-1}\phi''(\beta_1^0)\, D_{i1}(\beta_1^0)\Big\}\Big] \\ &\quad -\tfrac{1}{2}\{\phi'(\beta_1^0)^{-1} D_{i1}(\beta_1^0)\}^2 + O_{(5)}\,. \end{aligned} \tag{5.31}$$

Result (5.30) is derived by writing (5.8) as:

$$\phi'(\beta_1^0)^{-1} D_{i1}(\beta_1^0) = \exp(\xi_i) - 1 = \xi_i + \tfrac{1}{2}\xi_i^2 + O_p(|\xi_i|^3)\,, \tag{5.32}$$

and then inverting the expansion. (The result $\max_{1\le i\le m}|\xi_i| = o_p(1)$, in fact $O_{(3)}$, used in this argument, is readily derived.) To obtain (5.31), note that the analogue of (5.32) in that case is:

$$\phi'(\underline{\widehat{\beta}}_1)^{-1} D_{i1}(\underline{\widehat{\beta}}_1) = \exp(\eta_i) - 1 = \eta_i + \tfrac{1}{2}\eta_i^2 + O_p(|\eta_i|^3)\,, \tag{5.33}$$

and that, uniformly in $1 \le i \le m$,

$$\begin{aligned} \phi'(\underline{\widehat{\beta}}_1)^{-1} D_{i1}(\underline{\widehat{\beta}}_1) &= \Big\{\phi'(\beta_1^0) + (\underline{\widehat{\beta}}_1 - \beta_1^0)\,\phi''(\beta_1^0) + O_{(2)}\Big\}^{-1} \\ &\qquad \times\Big\{D_{i1}(\beta_1^0) + (\underline{\widehat{\beta}}_1 - \beta_1^0)\, D_{i2}(\beta_1^0) + O_{(7)}\Big\} \\ &= \phi'(\beta_1^0)^{-1}\Big\{1 - (\underline{\widehat{\beta}}_1 - \beta_1^0)\,\phi'(\beta_1^0)^{-1}\phi''(\beta_1^0)\Big\} \\ &\qquad \times\Big\{D_{i1}(\beta_1^0) + (\underline{\widehat{\beta}}_1 - \beta_1^0)\, D_{i2}(\beta_1^0)\Big\} + O_{(7)} \\ &= \phi'(\beta_1^0)^{-1}\Big[D_{i1}(\beta_1^0) + (\underline{\widehat{\beta}}_1 - \beta_1^0)\Big\{D_{i2}(\beta_1^0) \\ &\qquad -\phi'(\beta_1^0)^{-1}\phi''(\beta_1^0)\, D_{i1}(\beta_1^0)\Big\}\Big] + O_{(7)}\,. \end{aligned} \tag{5.34}$$

Result (5.31) follows from (5.33) and (5.34) on inverting the expansion at (5.33).

5.8. *Another approximation to $\underline{\widehat{\beta}}_1 - \beta_1^0$, and final approximations to $\underline{\widehat{\beta}}_0 - \beta_0^0$ and $\sigma^2 - \sigma_0^2$.* Next we use the expansions (5.30), (5.31) and (5.25) of ξ_i, η_i and ζ_i to refine the approximations derived in section 2.3. The results are given in (5.40), (5.41) and (5.45) in the cases of $\underline{\widehat{\beta}}_0 - \beta_0^0$, $\underline{\widehat{\beta}}_1 - \beta_1^0$ and $\sigma^2 - \sigma_0^2$, respectively.

It can be deduced from (5.31) and (5.25) that

$$\frac{1}{m}\sum_{i=1}^m \exp(U_i + \eta_i + \zeta_i) = \exp\left(\tfrac{1}{2}\sigma_0^2\right) + O_{(3)}\,. \tag{5.35}$$

By (5.30), (5.31) and (5.25),

$$\begin{aligned} &\frac{1}{m}\sum_{i=1}^m \exp(U_i)\,\{\exp(\xi_i) - \exp(\eta_i + \zeta_i)\} \\ &\quad = \frac{1}{m}\sum_{i=1}^m \exp(U_i)\left[\xi_i - \eta_i - \zeta_i + \tfrac{1}{2}\{\xi_i^2 - (\eta_i + \zeta_i)^2\}\right] + O_{(5)} \\ &\quad = -\frac{1}{m}\sum_{i=1}^m \exp(U_i)\,\{\zeta_i + \tfrac{1}{2}(2\,\eta_i\,\zeta_i + \zeta_i^2)\} + O_{(5)} + O_p(|\underline{\widehat{\beta}}_1 - \beta_1^0|\, n^{\varepsilon-(1/2)}). \end{aligned} \tag{5.36}$$

Defining $O_{(9)}$ as at Table 1 we deduce from (5.25) that

$$\begin{aligned} \frac{1}{m}\sum_{i=1}^m \exp(U_i)\,\zeta_i &= -\tfrac{1}{2}\,\phi(\beta_1^0)^{-2}\,\frac{1}{mn^2}\sum_{i=1}^m \exp\big(-2\,\beta_0^0 - U_i\big)\,\Delta_i^2 \\ &\qquad\qquad + O_p\{(mn)^{-1/2}\} + O_{(5)} \\ &= -(2n)^{-1}\,\phi(\beta_1^0)^{-1}\,\exp\big(-\beta_0^0\big) + O_{(9)}\,, \end{aligned} \tag{5.37}$$

where we have used the fact that $n/m \to 0$ and, since $Y_{i\bullet}$, conditional on $\mathcal{F}_i$, has a Poisson distribution with mean $B_{i0}^0\,\exp(U_i)$, then

$$\begin{aligned} E\big\{\exp(-U_i)\,\Delta_i^2\big\} &= E\big[\exp(-U_i)\,\{Y_{i\bullet} - E(Y_{i\bullet}\,|\,\mathcal{F}_i)\}^2\big] = E\{\exp(-U_i)\,\mathrm{var}(Y_{i\bullet}\,|\,\mathcal{F}_i)\} \\ &= E\big\{\exp(-U_i)\,B_{i0}^0\,\exp(U_i)\big\} = E(B_{i0}^0) = n\,\exp\big(\beta_0^0\big)\,\phi(\beta_1^0)\,. \end{aligned}$$

$$\begin{aligned} \frac{1}{m}\sum_{i=1}^m \exp(U_i)\,\zeta_i^2 &= \phi(\beta_1^0)^{-2}\,\frac{1}{mn^2}\sum_{i=1}^m \exp\big(-2\,\beta_0^0 - U_i\big)\,\Delta_i^2 + O_{(9)} \\ &= n^{-1}\,\phi(\beta_1^0)^{-1}\,\exp\big(-\beta_0^0\big) + O_{(9)}\,. \end{aligned}$$

Moreover, since by (5.31) and (5.25),

$$\eta_i = \phi'(\beta_1^0)^{-1}\,D_{i1}(\beta_1^0) + O_{(4)}\,, \quad \zeta_i = \phi(\beta_1^0)^{-1}\,n^{-1}\,\exp\big(-\beta_0^0 - U_i\big)\,\Delta_i + O_{(4)}\,,$$

and for $k \ge 0$,

$$E\big\{\exp(U_i)\,D_{ik}(\beta_1^0)\,\exp\big(-U_i\big)\,\Delta_i\big\} = E\big\{D_{ik}(\beta_1^0)\,E(\Delta_i\,|\,\mathcal{F}_i)\big\} = 0\,,$$

then

$$\frac{1}{m}\sum_{i=1}^m \exp(U_i)\,\eta_i\,\zeta_i = O_{(5)}\,. \tag{5.38}$$

Together, (5.36), (5.37), (5.38) and (5.38) imply that

$$\begin{aligned} &\frac{1}{m}\sum_{i=1}^m \exp(U_i)\,\big\{\exp(\xi_i) - \exp\big(\eta_i + \zeta_i\big)\big\} \\ &\qquad = (2n)^{-1}\,\phi(\beta_1^0)^{-1}\,\exp\big(-\beta_0^0\big) - (2n)^{-1}\,\phi(\beta_1^0)^{-1}\,\exp\big(-\beta_0^0\big) \\ &\qquad\qquad + O_{(9)} + O_p\big(|\underline{\widehat{\beta}}_1 - \beta_1^0|\,n^{\varepsilon-(1/2)}\big) \\ &\qquad = O_{(9)} + O_p\big(|\underline{\widehat{\beta}}_1 - \beta_1^0|\,n^{\varepsilon-(1/2)}\big)\,. \end{aligned} \tag{5.39}$$

Combining (5.19), (5.35) and (5.39), and noting that $\Delta = O_p\{(mn)^{-1/2}\}$ and $n/m \to 0$, we deduce that:

$$\underline{\widehat{\beta}}_1 - \beta_1^0 = O_{(9)}\,. \tag{5.40}$$

Together, (5.21) and (5.40) imply that

$$\underline{\widehat{\beta}}_0 - \beta_0^0 = \bar{U} + \bar{\zeta} - c_0\,n^{-1} + o_p\big(m^{-1/2} + n^{-1}\big)\,, \tag{5.41}$$

where

$$c_0 = \big\{2\,\phi(\beta_1^0)\,\exp\big(\beta_0^0 - \tfrac{1}{2}\,\sigma_0^2\big)\big\}^{-1}\,.$$

Result (3.4) Theorem 3.1 is a direct consequence of (5.41) and the property:

$$\begin{aligned}\bar{\zeta} &= -\frac{1}{m}\sum_{i=1}^{m} U_i\{n(\sigma^2)^0\exp(U_i+\beta_0^0)\phi(\beta_1^0)\}^{-1} \\ &\qquad -\tfrac{1}{2}\phi(\beta_1^0)^{-2}E\{n^{-1}\exp(-\beta_0^0-U_i)\Delta_i\}^2+o_p(n^{-1}) \\ &= c_0 n^{-1}+o_p(n^{-1}). \qquad (5.42)\end{aligned}$$

Results (5.25) and (5.40), and the property

$$E\{\exp(-2\,U_i)\,\Delta_i^2\} = E\{B_{i0}^0\,\exp(-U_i)\} = n\,\exp\left(\beta_0^0+\tfrac{1}{2}\,\sigma_0^2\right)\phi(\beta_1^0)\,,$$

imply that

$$\begin{aligned}\frac{1}{m}\sum_{i=1}^{m}\zeta_i^2 &= \phi(\beta_1^0)^{-2}\,\frac{1}{mn^2}\sum_{i=1}^{m}\exp\left(-2\,\beta_0^0-2\,U_i\right)\Delta_i^2+o_p(1) \\ &= n^{-1}\,\phi(\beta_1^0)^{-1}\,\exp\left(\tfrac{1}{2}\,\sigma_0^2-\beta_0^0\right)+o_p(n^{-1}) = 2\,c_0\,n^{-1}+o_p(n^{-1})\,. \qquad (5.43)\end{aligned}$$

By (5.26),

$$\frac{1}{m}\sum_{i=1}^{m}U_i\,\zeta_i = -\frac{1}{n}\,\phi(\beta_1^0)^{-1}\,\exp\left(\tfrac{1}{2}\,\sigma_0^2-\beta_0^0\right)\left(1+\tfrac{1}{2}\sigma_0^2\right)+o_p(n^{-1})\,. \qquad (5.44)$$

Together, (5.42)–(5.44) give:

$$\begin{aligned}&\frac{1}{m}\sum_{i=1}^{m}\left\{(U_i+\zeta_i-\bar{U}-\bar{\zeta})^2-(\sigma^2)^0\right\} \\ &= \frac{1}{m}\sum_{i=1}^{m}(U_i^2-(\sigma^2)^0)+\frac{1}{m}\sum_{i=1}^{m}\zeta_i^2-\bar{\zeta}^2+\frac{2}{m}\sum_{i=1}^{m}U_i\zeta_i-2\bar{U}\bar{\zeta}+O_p(m^{-1}) \\ &= \frac{1}{m}\sum_{i=1}^{m}(U_i^2-(\sigma^2)^0)+2n^{-1}c_0-2n^{-1}c_0(2+(\sigma^2)^0)+o_p(m^{-1/2}+n^{-1}) \\ &= \frac{1}{m}\sum_{i=1}^{m}(U_i^2-(\sigma^2)^0)-2n^{-1}c_0(1+\sigma_0^2)+o_p(m^{-1/2}+n^{-1})\end{aligned}$$

Hence, by (5.29),

$$\underline{\widehat{\sigma}}^2-\sigma_0^2 = \frac{1}{m}\sum_{i=1}^{m}\left(U_i^2-\sigma_0^2\right)+o_p(m^{-1/2}+n^{-1})\,. \qquad (5.45)$$

Result (3.6) of Theorem 3.1 is a direct consequence of (5.45).

5.9. *Final approximation to* $\underline{\widehat{\beta}}_1-\beta_1^0$. Our first step is to sharpen the expansion of (5.5) at (5.15); see (5.49), which leads to (5.54), the principal analogue of (5.15).

Recall that

$$\Delta_i = Y_{i\bullet}-\exp\left(\beta_0^0+U_i\right)\sum_{j=1}^{n}\exp(\beta_1^0\,X_{ij}) = Y_{i\bullet}-\exp(U_i)\,B_{i0}^0\,. \qquad (5.46)$$

Also, in view of (5.40) and (5.41),

$$\begin{aligned}
B_i &= \exp\left(\underline{\widehat{\beta}}_0\right) \sum_{j=1}^{n} \exp(\underline{\widehat{\beta}}_1 X_{ij}) \\
&= \exp\left(\beta_0^0\right) \left\{1 + \left(\underline{\widehat{\beta}}_0 - \beta_0^0\right) + \tfrac{1}{2}\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)^2 + \tfrac{1}{6}\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)^3\right\} \\
&\qquad \times \sum_{j=1}^{n} \left\{1 + \left(\underline{\widehat{\beta}}_1 - \beta_1^0\right) X_{ij} + \tfrac{1}{2}\left(\underline{\widehat{\beta}}_1 - \beta_1^0\right)^2 X_{ij}^2\right\} \\
&\qquad \times \exp\left(\beta_1^0 X_{ij}\right) + O_p\big(m^{-2}\, n + m^{-3/2}\, n^{-1/2} + m^{-1} + n^{\varepsilon-3}\big) \\
&= \exp\left(\beta_0^0\right) \sum_{j=1}^{n} \left\{1 + \left(\underline{\widehat{\beta}}_0 - \beta_0^0\right) + \tfrac{1}{2}\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)^2 + \tfrac{1}{6}\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)^3\right. \\
&\qquad \left. + \left(\underline{\widehat{\beta}}_1 - \beta_1^0\right) X_{ij} + \tfrac{1}{2}\left(\underline{\widehat{\beta}}_1 - \beta_1^0\right)^2 X_{ij}^2 + \left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)\left(\underline{\widehat{\beta}}_1 - \beta_1^0\right) X_{ij}\right\} \exp\left(\beta_1^0 X_{ij}\right) \\
&\qquad + O_p\big(m^{-1/2}\, n^{\varepsilon} + n^{\varepsilon-(5/2)}\big) \\
&= \left\{1 + \left(\underline{\widehat{\beta}}_0 - \beta_0^0\right) + \tfrac{1}{2}\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)^2 + \tfrac{1}{6}\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)^3\right\} B_{i0}^0 \\
&\qquad + \left\{1 + \left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)\right\} \left(\underline{\widehat{\beta}}_1 - \beta_1^0\right) B_{i1}^0 + \tfrac{1}{2}\left(\underline{\widehat{\beta}}_1 - \beta_1^0\right)^2 B_{i2}^0 + O_{(10)}\,,
\end{aligned}$$

where $O_{(10)}$ is defined in Table 1. Hence, recalling that $\delta_i = \underline{\widehat{\mu}}_i + \frac{1}{2}\,\underline{\widehat{\lambda}}_i - U_i$, we see that, for each $\varepsilon > 0$, we have, uniformly in $1 \le i \le n$,

$$\begin{aligned}
Y_{i\bullet} - B_i \exp(\delta_i + U_i) &= Y_{i\bullet} - B_{i0}^0 \exp(\delta_i + U_i) \\
&\quad - \Big[\left\{\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right) + \tfrac{1}{2}\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)^2 + \tfrac{1}{6}\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)^3\right\} B_{i0}^0 \\
&\quad + \left\{1 + \left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)\right\} \left(\underline{\widehat{\beta}}_1 - \beta_1^0\right) B_{i1}^0 + \tfrac{1}{2}\left(\underline{\widehat{\beta}}_1 - \beta_1^0\right)^2 B_{i2}^0\Big] \exp(\delta_i + U_i) + O_{(10)}\,. \qquad (5.47)
\end{aligned}$$

Combining (5.46) and (5.47) we obtain:

$$\begin{aligned}
Y_{i\bullet} - B_i \exp(\delta_i + U_i) &= \Delta_i - \exp(U_i) \Bigg(\left\{\exp(\delta_i) - 1\right\} B_{i0}^0 \\
&\qquad + \Big[\left\{\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right) + \tfrac{1}{2}\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)^2 + \tfrac{1}{6}\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)^3\right\} B_{i0}^0 \\
&\qquad + \left\{1 + \left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)\right\} \left(\underline{\widehat{\beta}}_1 - \beta_1^0\right) B_{i1}^0 + \tfrac{1}{2}\left(\underline{\widehat{\beta}}_1 - \beta_1^0\right)^2 B_{i2}^0\Big] \exp(\delta_i)\Bigg) + O_{(10)}\,.
\end{aligned}$$

Therefore, (5.5) implies that:

$$\begin{aligned}
(\underline{\widehat{\sigma}}^2)^{-1}\, \underline{\widehat{\mu}}_i &= \Delta_i - \exp(U_i) \Bigg(\left\{\exp(\delta_i) - 1\right\} B_{i0}^0 \\
&\qquad + \Big[\left\{\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right) + \tfrac{1}{2}\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)^2 + \tfrac{1}{6}\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)^3\right\} B_{i0}^0 \\
&\qquad + \left\{1 + \left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)\right\} \left(\underline{\widehat{\beta}}_1 - \beta_1^0\right) B_{i1}^0 + \tfrac{1}{2}\left(\underline{\widehat{\beta}}_1 - \beta_1^0\right)^2 B_{i2}^0\Big] \exp(\delta_i)\Bigg) + O_{(10)}\,,
\end{aligned}$$

or equivalently,

$$\begin{aligned}
&\exp(U_i) \Bigg(\left\{\exp(\delta_i) - 1\right\} B_{i0}^0 \\
&\qquad + \exp(\delta_i) \Big[\left\{\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right) + \tfrac{1}{2}\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)^2 + \tfrac{1}{6}\left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)^3\right\} B_{i0}^0 \\
&\qquad + \left\{1 + \left(\underline{\widehat{\beta}}_0 - \beta_0^0\right)\right\} \left(\underline{\widehat{\beta}}_1 - \beta_1^0\right) B_{i1}^0 + \tfrac{1}{2}\left(\underline{\widehat{\beta}}_1 - \beta_1^0\right)^2 B_{i2}^0\Big]\Bigg) + (\underline{\widehat{\sigma}}^2)^{-1} \left(\delta_i + U_i - \tfrac{1}{2}\,\underline{\widehat{\lambda}}_i\right) \\
&\qquad = \Delta_i + O_{(10)}\,. \qquad (5.48)
\end{aligned}$$

Substituting the far right-hand side of (5.18) for $\widehat{\underline{\lambda}}_i$ in (5.48) we deduce that:

$$\begin{aligned}\exp(\delta_i) - 1 + \exp(\delta_i) \Big\{ (\widehat{\underline{\beta}}_0 - \beta_0^0) + \tfrac{1}{2} (\widehat{\underline{\beta}}_0 - \beta_0^0)^2 + (\widehat{\underline{\beta}}_1 - \beta_1^0) (B_{i1}^0 / B_{i0}^0) \Big\} \\ + \{ \widehat{\underline{\sigma}}^2 B_{i0}^0 \exp(U_i) \}^{-1} (\delta_i + U_i) \\ = \{ B_{i0}^0 \exp(U_i) \}^{-1} \Delta_i + O_{(11)} ,\end{aligned} \tag{5.49}$$

where $O_{(11)}$ is as defined in Table 1. Result (5.49) implies that

$$\delta_i + \tfrac{1}{2} \delta_i^2 G_{i2} + \tfrac{1}{6} \delta_i^3 G_{i3} = G_i + O_{(11)} , \tag{5.50}$$

where, putting

$$G_{i1} = 1 + (\widehat{\underline{\beta}}_0 - \beta_0^0) + \tfrac{1}{2} (\widehat{\underline{\beta}}_0 - \beta_0^0)^2 + (\widehat{\underline{\beta}}_1 - \beta_1^0) (B_{i1} / B_{i0}^0) + \{ \widehat{\underline{\sigma}}^2 B_{i0}^0 \exp(U_i) \}^{-1} , \tag{5.51}$$

we define G_i, G_{i2} and G_{i3} by $G_{i3} G_{i1} = 1$,

$$G_{i2} G_{i1} = 1 + (\widehat{\underline{\beta}}_0 - \beta_0^0) + (\widehat{\underline{\beta}}_1 - \beta_1^0) (B_{i1} / B_{i0}^0) , \tag{5.52}$$

$$\begin{aligned} G_i G_{i1} &= \{ B_{i0}^0 \exp(U_i) \}^{-1} \Delta_i - \{ \widehat{\underline{\sigma}}^2 B_{i0}^0 \exp(U_i) \}^{-1} U_i \\ &\quad - \Big\{ (\widehat{\underline{\beta}}_0 - \beta_0^0) + \tfrac{1}{2} (\widehat{\underline{\beta}}_0 - \beta_0^0)^2 + (\widehat{\underline{\beta}}_1 - \beta_1^0) (B_{i1}^0 / B_{i0}^0) \Big\} . \end{aligned} \tag{5.53}$$

Solving (5.50) for δ_i we deduce that, for each $\varepsilon > 0$,

$$\delta_i = G_i - \tfrac{1}{2} G_{i2} G_i^2 - (\tfrac{1}{6} G_{i3} - \tfrac{1}{2} G_{i2}^2) G_i^3 + O_{(11)} , \tag{5.54}$$

uniformly in $1 \le i \le n$. Now, G_{i1}, G_{i2} and G_{i3} each equal $1 + O_p(m^{-1/2} + n^{\varepsilon-1})$. Therefore, $\frac{1}{6} G_{i3} - \frac{1}{2} G_{i2}^2 = -\frac{1}{3} + O_p(m^{-1/2} + n^{\varepsilon-1})$. Using (5.51), (5.52) and (5.53) we deduce that

$$G_{i2} = 1 - \{ \widehat{\underline{\sigma}}^2 B_{i0}^0 \exp(U_i) \}^{-1} + O_p(m^{-1} + n^{\varepsilon-2}) , \quad G_i = H_i + O_{(11)} ,$$

where

$$\begin{aligned} H_i &= \Big[\{ B_{i0}^0 \exp(U_i) \}^{-1} \Delta_i - \{ \widehat{\underline{\sigma}}^2 B_{i0}^0 \exp(U_i) \}^{-1} U_i \\ &\quad - \Big\{ (\widehat{\underline{\beta}}_0 - \beta_0^0) + \tfrac{1}{2} (\widehat{\underline{\beta}}_0 - \beta_0^0)^2 + (\widehat{\underline{\beta}}_1 - \beta_1^0) (B_{i1}^0 / B_{i0}^0) \Big\} \Big] \\ &\quad \times \Big[1 - (\widehat{\underline{\beta}}_0 - \beta_0^0) - (\widehat{\underline{\beta}}_1 - \beta_1^0) (B_{i1} / B_{i0}^0) - \{ \widehat{\underline{\sigma}}^2 B_{i0}^0 \exp(U_i) \}^{-1} \Big] . \end{aligned}$$

Note too that $G_{i2} H_i^2 = H_i^2 + O_p(m^{-1/2} n^{\varepsilon-1} + n^{\varepsilon-2})$. Combining the results from (5.54) down we see that

$$\delta_i = H_i - \tfrac{1}{2} H_i^2 + \tfrac{1}{3} H_i^3 + O_{(11)} . \tag{5.55}$$

Note that, as $a \to 0$, $\exp(a - \frac{1}{2} a^2 + \frac{1}{3} a^3) - 1 = a + O(a^4)$ as $a \to 0$. This property, (5.55) and the fact that $H_i^4 = O_p(n^{\varepsilon-2})$ imply that

$$\exp(\delta_i) - 1 = H_i + O_{(11)} . \tag{5.56}$$

The formula immediately preceding (5.19) is equivalent to:

$$\begin{aligned} &\{ 1 + O_p(m^{-1/2} + n^{-1}) \} \gamma'(\beta_1^0) (\widehat{\underline{\beta}}_1 - \beta_1^0) \frac{1}{m} \sum_{i=1}^m \exp(U_i + \eta_i + \zeta_i) \\ &= \Delta \exp(-\beta_0^0) \phi(\beta_1^0)^{-1} + \gamma(\beta_1^0) \frac{1}{m} \sum_{i=1}^m \{ \exp(\xi_i) - \exp(\eta_i + \zeta_i) \} \exp(U_i) . \end{aligned} \tag{5.57}$$

Since η_i and ζ_i both equal $O_{(3)}$ (see (5.25) and (5.31)), and $m^{-1}\sum_{i=1}^m \exp(U_i) = E\{\exp(U_1)\} + o_p(1) = \exp\{\sigma_0^2/2\} + o_p(1)$, then (5.57) implies that

$$\begin{aligned}&\{1+o_p(1)\}\,\gamma'(\beta_1^0)\,(\underline{\widehat{\beta}}_1 - \beta_1^0)\exp\{\sigma_0^2/2\}\\ &\quad = \Delta\,\exp\big(-\beta_0^0\big)\,\phi(\beta_1^0)^{-1} + \gamma(\beta_1^0)\,\frac{1}{m}\sum_{i=1}^m \{\exp(\xi_i) - \exp(\eta_i+\zeta_i)\}\,\exp(U_i)\,. \end{aligned} \tag{5.58}$$

Formulae (5.8) and (5.9) are together equivalent to:

$$\phi'(\beta_1^0)\,\{\exp(\xi_i)-1\} = \frac{1}{n}\sum_{j=1}^n \{X_{ij}\,\exp\big(\beta_1^0\,X_{ij}\big) - \phi'(\beta_1^0)\}\,, \tag{5.59}$$

$$\phi'(\underline{\widehat{\beta}}_1)\,\{\exp(\eta_i)-1\} = \frac{1}{n}\sum_{j=1}^n \{X_{ij}\,\exp(\underline{\widehat{\beta}}_1\,X_{ij}) - \phi'(\underline{\widehat{\beta}}_1)\}\,. \tag{5.60}$$

Result (5.60) implies that, for each $\varepsilon > 0$,

$$\begin{aligned}&\{\phi'(\beta_1^0) + O_p(|\underline{\widehat{\beta}}_1 - \beta_1^0|)\}\,\{\exp(\eta_i)-1\}\\ &\quad = \frac{1}{n}\sum_{j=1}^n \{X_{ij}\,\exp\big(\beta_1^0\,X_{ij}\big) - \phi'(\beta_1^0)\} + O_p(|\underline{\widehat{\beta}}_1 - \beta_1^0|\,n^{\varepsilon-(1/2)})\,,\end{aligned}$$

uniformly in $1 \le i \le n$. Therefore, since $\eta_i = O_{(3)}$ (see (5.31)), then

$$\phi'(\beta_1^0)\,\{\exp(\eta_i)-1\} = \frac{1}{n}\sum_{j=1}^n \{X_{ij}\,\exp\big(\beta_1^0\,X_{ij}\big) - \phi'(\beta_1^0)\} + O_p(|\underline{\widehat{\beta}}_1 - \beta_1^0|\,n^{\varepsilon-(1/2)}),$$

which in company with (5.60) implies that

$$\phi'(\beta_1^0)\,\{\exp(\eta_i) - \exp(\xi_i)\} = O_p(|\underline{\widehat{\beta}}_1 - \beta_1^0|\,n^{\varepsilon-(1/2)})\,,$$

uniformly in $1 \le i \le n$. Hence, since $\eta_i = O_{(3)}$ and $\zeta_i = O_{(3)}$ (see (5.25) and (5.31)),

$$\begin{aligned}\exp(\xi_i) - \exp(\eta_i+\zeta_i) &= \{\exp(\xi_i) - \exp(\eta_i)\}\,\exp(\zeta_i) + \exp(\xi_i)\,\{1-\exp(\zeta_i)\}\\ &= \exp(\xi_i)\,\{1-\exp(\zeta_i)\} + O_p(|\underline{\widehat{\beta}}_1 - \beta_1^0|\,n^{\varepsilon-(1/2)})\,,\end{aligned} \tag{5.61}$$

uniformly in i. Combining (5.58) and (5.61) we deduce that

$$\begin{aligned}&\{1+o_p(1)\}\,\gamma'(\beta_1^0)\,(\underline{\widehat{\beta}}_1 - \beta_1^0)\,\exp\{\tfrac{1}{2}\sigma_0^2\}\\ &\quad = \Delta\,\exp\big(-\beta_0^0\big)\,\phi(\beta_1^0)^{-1} + \gamma(\beta_1^0)\,\frac{1}{m}\sum_{i=1}^m \exp(\xi_i+U_i)\,\{1-\exp(\zeta_i)\}\,.\end{aligned} \tag{5.62}$$

Next we return to (5.10), which we write equivalently as

$$\begin{aligned}\phi(\beta_1^0)\,\{1-\exp(\zeta_i)\} &= \exp\big(\underline{\widehat{\beta}}_0 - \beta_0^0 + \delta_i\big)\,\frac{1}{n}\sum_{j=1}^n \{\exp(\underline{\widehat{\beta}}_1\,X_{ij}) - \phi(\underline{\widehat{\beta}}_1)\}\\ &\quad -\frac{1}{n}\sum_{j=1}^n \{Y_{ij}\,\exp\big(-\beta_0^0 - U_i\big) - \phi(\beta_1^0)\} + (\underline{\widehat{\sigma}}^2 n)^{-1}\,\underline{\widehat{\mu}}_i\,\exp\big(-\beta_0^0 - U_i\big).\end{aligned} \tag{5.63}$$

So that we might replace $\underline{\widehat{\beta}}_1$ by β_1^0 on the right-hand side of (5.63) we observe that

$$\frac{1}{n}\sum_{j=1}^{n}\{\exp(\underline{\widehat{\beta}}_1 X_{ij}) - \phi(\underline{\widehat{\beta}}_1)\}$$

$$= \frac{1}{n}\sum_{j=1}^{n}\{\exp(\beta_1^0 X_{ij}) - \phi(\beta_1^0)\} + O_p(|\underline{\widehat{\beta}}_1 - \beta_1^0|\, n^{\varepsilon-(1/2)}). \tag{5.64}$$

Combining (5.62)–(5.64) we obtain:

$$\begin{aligned}
&\{1+o_p(1)\}\,\gamma'(\beta_1^0)\,(\underline{\widehat{\beta}}_1 - \beta_1^0)\,\exp\{\tfrac{1}{2}\sigma_0^2\}\\
&= \Delta\,\exp(-\beta_0^0)\,\phi(\beta_1^0)^{-1}\\
&\quad + \frac{\phi'(\beta_1^0)}{\phi(\beta_1^0)^2}\,\frac{1}{m}\sum_{i=1}^{m}\exp(\xi_i+U_i)\Big[\exp(\underline{\widehat{\beta}}_0 - \beta_0^0 + \delta_i)\,\frac{1}{n}\sum_{j=1}^{n}\{\exp(\beta_1^0 X_{ij}) - \phi(\beta_1^0)\}\\
&\quad - \frac{1}{n}\sum_{j=1}^{n}\{Y_{ij}\exp(-\beta_0^0 - U_i) - \phi(\beta_1^0)\} + (\underline{\widehat{\sigma}}^2 n)^{-1}\,\underline{\widehat{\mu}}_i\exp(-\beta_0^0 - U_i)\Big].
\end{aligned} \tag{5.65}$$

(Recall that $\gamma = \phi'\phi^{-1}$, and so $\gamma/\phi = \phi'\phi^{-2}$.)

Since $\exp(\xi_i) - 1 = D_{i1}(\beta_1^0)\,\phi'(\beta_1^0)^{-1}$ (see (5.8)) and $\underline{\widehat{\beta}}_0 - \beta_0^0 = O_p(m^{-1/2} + n^{-1})$ (see (5.41)) then

$$\begin{aligned}
&\frac{1}{m}\sum_{i=1}^{m}\exp(\xi_i+U_i)\,\exp(\underline{\widehat{\beta}}_0 - \beta_0^0 + \delta_i)\,\frac{1}{n}\sum_{j=1}^{n}\{\exp(\beta_1^0 X_{ij}) - \phi(\beta_1^0)\}\\
&\quad = \Big\{1 + (\underline{\widehat{\beta}}_0 - \beta_0^0) + \tfrac{1}{2}(\underline{\widehat{\beta}}_0 - \beta_0^0)^2\Big\}\,\frac{1}{m}\sum_{i=1}^{m}\exp(\xi_i+\delta_i+U_i)\,D_{i0}(\beta_1^0)\\
&\qquad + O_p(m^{-3/2} + n^{-3})\\
&\quad = \Big\{1 + (\underline{\widehat{\beta}}_0 - \beta_0^0) + \tfrac{1}{2}(\underline{\widehat{\beta}}_0 - \beta_0^0)^2\Big\}\\
&\qquad \times\frac{1}{m}\sum_{i=1}^{m}\exp(\delta_i+U_i)\Big\{1 + D_{i1}(\beta_1^0)\,\phi'(\beta_1^0)^{-1}\Big\}\,D_{i0}(\beta_1^0)\\
&\qquad + O_p(m^{-3/2} + n^{-3}).
\end{aligned} \tag{5.66}$$

Likewise,

$$\begin{aligned}
&\frac{1}{m}\sum_{i=1}^{m}\exp(\xi_i+U_i)\,\frac{1}{n}\sum_{j=1}^{n}\{Y_{ij}\exp(-\beta_0^0 - U_i) - \phi(\beta_1^0)\}\\
&= \frac{1}{m}\sum_{i=1}^{m}\exp(U_i)\Big\{1 + D_{i1}(\beta_1^0)\,\phi'(\beta_1^0)^{-1}\Big\}\Big\{n^{-1}\,\Delta_i\exp(-\beta_0^0 - U_i) + D_{i0}(\beta_1^0)\Big\},
\end{aligned} \tag{5.67}$$

and, since $\sum_i \underline{\widehat{\mu}}_i = 0$ (see (5.6)),

$$\begin{aligned}
&\frac{1}{m}\sum_{i=1}^{m}\exp(\xi_i+U_i)\,(\underline{\widehat{\sigma}}^2 n)^{-1}\,\underline{\widehat{\mu}}_i\exp(-\beta_0^0 - U_i) = \frac{1}{\underline{\widehat{\sigma}}^2\,mn}\sum_{i=1}^{m}\exp(\xi_i - \beta_0^0)\,\underline{\widehat{\mu}}_i\\
&\quad = \exp(-\beta_0^0)\,\frac{1}{\underline{\widehat{\sigma}}^2\,mn}\sum_{i=1}^{m}\Big\{1 + D_{i1}(\beta_1^0)\,\phi'(\beta_1^0)^{-1}\Big\}\,\underline{\widehat{\mu}}_i\\
&\quad = \exp(-\beta_0^0)\,\phi'(\beta_1^0)^{-1}\,\frac{1}{\underline{\widehat{\sigma}}^2\,mn}\sum_{i=1}^{m}D_{i1}(\beta_1^0)\,\underline{\widehat{\mu}}_i = O_p(m^{-1/2}\,n^{-3/2}).
\end{aligned} \tag{5.68}$$

Combining (5.65)–(5.68) we see that

$$\begin{aligned}
&\{1+o_p(1)\}\,\gamma'(\beta_1^0)\,(\underline{\widehat{\beta}}_1-\beta_1^0)\exp\{\tfrac{1}{2}\sigma_0^2\}\\
&\quad=\Delta\,\exp\big(-\beta_0^0\big)\,\phi(\beta_1^0)^{-1}+\frac{\phi'(\beta_1^0)}{\phi(\beta_1^0)^2}\Big[\Big\{1+(\underline{\widehat{\beta}}_0-\beta_0^0)+\tfrac{1}{2}\,(\underline{\widehat{\beta}}_0-\beta_0^0)^2\Big\}\\
&\qquad\times\frac{1}{m}\sum_{i=1}^m\exp(\delta_i+U_i)\Big\{1+D_{i1}(\beta_1^0)\,\phi'(\beta_1^0)^{-1}\Big\}\,D_{i0}(\beta_1^0)\\
&\qquad-\exp\big(-\beta_0^0\big)\,\frac{1}{m}\sum_{i=1}^m\Big\{1+D_{i1}(\beta_1^0)\,\phi'(\beta_1^0)^{-1}\Big\}\\
&\qquad\times\{n^{-1}\,\Delta_i+\exp\big(\beta_0^0+U_i\big)\,D_{i0}(\beta_1^0)\}\Big]\\
&\qquad+O_p(m^{-1/2}\,n^{-1}+n^{-3})\,.
\end{aligned}\tag{5.69}$$

Using the fact that $E(\Delta_i\,|\,\mathcal{F}_i)=0$ and $D_{i1}(\beta_1^0)=O_{(3)}$ it can be proved that, for all $\varepsilon>0$,

$$\begin{aligned}
&\frac{1}{mn}\sum_{i=1}^m\exp\big(-\beta_0^0\big)\Big\{1+D_{i1}(\beta_1^0)\,\phi'(\beta_1^0)^{-1}\Big\}\,\Delta_i\\
&\qquad=\exp\big(-\beta_0^0\big)\,\frac{1}{mn}\sum_{i=1}^m\Delta_i+O_p(m^{-1/2}\,n^{-1})\,.
\end{aligned}\tag{5.70}$$

Also,

$$\begin{aligned}
\Delta'\;&\equiv\;\Delta\,\exp\big(-\beta_0^0\big)\,\phi(\beta_1^0)^{-1}-\frac{\exp\big(-\beta_0^0\big)\phi'(\beta_1^0)}{\phi(\beta_1^0)^2}\,\frac{1}{mn}\sum_{i=1}^m\Delta_i\\
&=\;\phi(\beta_1^0)^{-1}\exp\big(-\beta_0^0\big)\,\frac{1}{mn}\sum_{i=1}^m\sum_{j=1}^n\Big\{X_{ij}-\frac{\phi'(\beta_1^0)}{\phi(\beta_1^0)}\Big\}\,\{Y_{ij}-\exp\big(\beta_0^0+\beta_1^0\,X_{ij}+U_i\big)\}.
\end{aligned}\tag{5.71}$$

Moreover, using (5.41) and the fact that $D_{i0}(\beta_1^0)=O_{(3)}$ and $E\{D_{i0}(\beta_1^0)\,|\,U_i\}=0$, it can be shown that

$$\begin{aligned}
&(\underline{\widehat{\beta}}_0-\beta_0^0)\,\frac{1}{m}\sum_{i=1}^m\exp(U_i)\Big\{1+D_{i1}(\beta_1^0)\,\phi'(\beta_1^0)^{-1}\Big\}\,D_{i0}(\beta_1^0)\\
&\qquad=O_p\Big\{(m^{-1/2}+n^{-1})\cdot(m^{-1/2}\,n^{\varepsilon-(1/2)})\Big\}\\
&\qquad=O_p(m^{-1/2}\,n^{\varepsilon-1})\,.
\end{aligned}\tag{5.72}$$

Combining (5.69)–(5.72) we deduce that

$$\begin{aligned}
&\{1+o_p(1)\}\,\gamma'(\beta_1^0)\,(\underline{\widehat{\beta}}_1-\beta_1^0)\,\exp\{\tfrac{1}{2}\sigma_0^2\}\\
&\quad=\Delta'+\frac{\phi'(\beta_1^0)}{\phi(\beta_1^0)^2}\,\frac{1}{m}\sum_{i=1}^m\exp(U_i)\,\{\exp(\delta_i)-1\}\Big\{1+D_{i1}(\beta_1^0)\,\phi'(\beta_1^0)^{-1}\Big\}\,D_{i0}(\beta_1^0)\\
&\qquad+O_p(m^{-1/2}\,n^{\varepsilon-1}+n^{-3})\,.
\end{aligned}\tag{5.73}$$

Using (5.56) to substitute for $\exp(\delta_i)-1$ in (5.73), and noting that $D_{ik}(\beta_1^0)=O_{(3)}$ for $k=0,1$, we deduce from (5.73) that

$$\begin{aligned}
&\{1+o_p(1)\}\,\gamma'(\beta_1^0)\,(\underline{\widehat{\beta}}_1-\beta_1^0)\,\exp\{\tfrac{1}{2}\sigma_0^2\}\\
&\qquad=\Delta'+\frac{\phi'(\beta_1^0)}{\phi(\beta_1^0)^2}\,\psi(H)+O_p(m^{-1/2}\,n^{\varepsilon-1}+n^{\varepsilon-(5/2)})\,,
\end{aligned}\tag{5.74}$$

where $H = (H_1, \ldots, H_m)$, H_i is as defined at (5.55), and, given a sequence of random variables $K = (K_1, \ldots, K_m)$, we put

$$\psi(K) = \frac{1}{m} \sum_{i=1}^{m} \exp(U_i)\, K_i \left\{1 + D_{i1}(\beta_1^0)\, \phi'(\beta_1^0)^{-1}\right\} D_{i0}(\beta_1^0)\,.$$

Note again that $|D_{i0}(\beta_1^0)| = O_{(3)}$, and the dominant term on the right-hand side of formula (5.55) for H_i is $\{B_{i0}^0 \exp(U_i)\}^{-1}\, \Delta_i$. Moreover, $|\underline{\widehat{\beta}}_0 - \beta_0^0| = O_p(m^{-1/2} + n^{-1})$ (see (5.41)), $|\underline{\widehat{\beta}}_1 - \beta_1^0| = O_p\{(mn)^{-1/2} + n^{\varepsilon-(3/2)}\}$ (see (5.41)),

$$\{\underline{\widehat{\sigma}}^2\, B_{i0}^0 \exp(U_i)\}^{-1} = \{n\, \sigma_0^2\, \phi(\beta_1^0)\, \exp\left(\beta_0^0 + U_i\right)\}^{-1} + O_p(n^{\varepsilon-(3/2)})\,,$$

$$B_{i1}^0\, B_{i0}^{0\,-1} = \phi'(\beta_1^0)\, \phi(\beta_1^0)^{-1} + O_{(3)}\,.$$

Combining these properties we deduce that

(5.75) (5.74) continues to hold if, on the right-hand side, $\psi(H)$ is replaced by $\psi(H')$ where $H' = (H_1', \ldots, H_m')$ and $H_i' = H_i^{(1)} - H_i^{(2)} - H_i^{(3)}$,

with

$$\begin{aligned} H_i^{(1)} &= \{B_{i0}^0 \exp(U_i)\}^{-1}\, \Delta_i \left[1 - (\underline{\widehat{\beta}}_0 - \beta_0^0) - \{n\, \sigma_0^2\, \phi(\beta_1^0)\, \exp(U_i)\}^{-1}\right], \\ H_i^{(2)} &= \{\underline{\widehat{\sigma}}^2\, B_{i0}^0 \exp(U_i)\}^{-1}\, U_i \\ \text{and} \quad H_i^{(3)} &= (\underline{\widehat{\beta}}_0 - \beta_0^0) + \tfrac{1}{2}\, (\underline{\widehat{\beta}}_0 - \beta_0^0)^2 + (\underline{\widehat{\beta}}_1 - \beta_1^0)\, \{\phi'(\beta_1^0)/\phi'(\beta_1^0)\}\,. \end{aligned}$$

(Note that $H_i^{(3)}$ does not depend on i.) It can be proved from the properties $E(\Delta_i \,|\, \mathcal{F}_i) = 0$ and $|D_{i0}(\beta_1^0)| = O_{(3)}$ that, with $H^{(j)}$ denoting $(H_1^{(j)}, \ldots, H_m^{(j)})$, we have:

$$\psi(H^{(1)}) = O_p(m^{-1/2}\, n^{-1})\,. \tag{5.76}$$

More simply, since $E(U_i \,|\, X_{i1}, \ldots, X_{in}) = 0$ then

$$\begin{aligned} \psi(H^{(2)}) &= \frac{1}{m} \sum_{i=1}^{m} (\underline{\widehat{\sigma}}^2\, B_{i0}^0)^{-1}\, U_i \left\{1 + D_{i1}(\beta_1^0)\, \phi'(\beta_1^0)^{-1}\right\} D_{i0}(\beta_1^0) \\ &= O_p(m^{-1/2}\, n^{-3/2})\,. \end{aligned}$$

Furthermore, writing $\mathbf{1} = (1, \ldots, 1)$, an n-vector; and noting that the properties $E\{D_{ik}(\beta_1^0) \,|\, U_i\} = 0$, $\mathrm{var}\{D_{ik}(\beta_1^0) \,|\, U_i\} = O(n^{-1})$ and $E\{\exp(U_i)\} = \exp\left(\frac{1}{2}\, \sigma_0^2\right)$ imply that

$$\begin{aligned} \psi(\mathbf{1}) &= \frac{1}{m} \sum_{i=1}^{m} \exp(U_i) \left\{1 + D_{i1}(\beta_1^0)\, \phi'(\beta_1^0)^{-1}\right\} D_{i0}(\beta_1^0) \\ &= \phi'(\beta_1^0)^{-1}\, \frac{1}{m} \sum_{i=1}^{m} \exp(U_i)\, D_{i1}(\beta_1^0)\, D_{i0}(\beta_1^0) + O_p(m^{-1/2}\, n^{-1/2}) \\ &= n^{-1}\, \{\phi'(2\beta_1^0)\phi'(\beta_1^0)^{-1} - \phi(\beta_1^0)\}\, \exp\left(\tfrac{1}{2}\, \sigma_0^2\right) + O_p(m^{-1/2}\, n^{-1/2} + n^{-3/2})\,; \end{aligned}$$

we obtain:

$$\begin{aligned} \psi(H^{(3)}) &= \left[(\underline{\widehat{\beta}}_0 - \beta_0^0) + \tfrac{1}{2}\, (\underline{\widehat{\beta}}_0 - \beta_0^0)^2 + (\underline{\widehat{\beta}}_1 - \beta_1^0)\, \{\phi'(\beta_1^0)/\phi'(\beta_1^0)\}\right] \psi(\mathbf{1}) \\ &= \left\{\frac{1}{n}\, [\phi'(2\beta_1^0)\phi'(\beta_1^0)^{-1} - \phi(\beta_1^0)]\, \exp(\tfrac{1}{2}(\sigma^2)^0) + O_p(m^{-1/2} n^{-1/2} + n^{-3/2})\right\} \\ &\qquad \times \left[(\underline{\widehat{\beta}}_0 - \beta_0^0) + \frac{1}{2}(\underline{\widehat{\beta}}_0 - \beta_0^0)^2 + (\underline{\widehat{\beta}}_1 - \beta_1^0)\{\phi'(\beta_1^0)/\phi(\beta_1^0)\}\right] \\ &= O_p(m^{-1/2} n^{-1}) \end{aligned}$$

To obtain the last line here we used (3.4) of Theorem 3.1, already proved in Section 5.8 above.

Combining (5.74)–(5.77), and noting that the function ψ is linear, so that

$$\psi(H) = \psi(H^{(1)}) - \psi(H^{(2)}) - \psi(H^{(3)}),$$

we deduce that

$$\{1 + o_p(1)\}\, \gamma'(\beta_1^0)\, (\widehat{\underline{\beta}}_1 - \beta_1^0) \exp(\tfrac{1}{2}\sigma_0^2) = \Delta' + o_p\{(mn)^{-1/2} + n^{-2}\}\,.$$

Furthermore, the random variable Δ', defined at (5.71), is asymptotically normally distributed with zero mean and variance

$$\begin{aligned}
&\frac{1}{mn}\, \exp(-2\,\beta_0^0)\, E\Big(\Big\{X_{11} - \frac{\phi'(\beta_1^0)}{\phi(\beta_1^0)}\Big\}^2\, E\Big[E\{Y_{11} - E(Y_{11}\,|\,X_{11}, U_1)\}^2 \;\Big|\; X_{11}, U_1\Big]\Big) \\
&\quad = \frac{1}{mn}\, \exp\,(-2\beta_0^0)\, E\Big[\Big\{X_{11} - \frac{\phi'(\beta_1^0)}{\phi(\beta_1^0)}\Big\}^2\, \exp\,(\beta_0^0 + \beta_1^0\, X_{11} + U_1)\Big] \\
&\quad = \frac{1}{mn}\, \exp\,(\tfrac{1}{2}\,\sigma_0^2 - \beta_0^0)\, E\Big[\Big\{X_{11} - \frac{\phi'(\beta_1^0)}{\phi(\beta_1^0)}\Big\}^2\, \exp\,(\beta_1^0\, X_{11})\Big] \\
&\quad = (mn)^{-1}\, \gamma'(\beta_1^0)^2\, \exp\{\sigma_0^2\}\tau^2\,,
\end{aligned}$$

where τ^2 is as at (3.3). Result (3.5) of the Theorem 3.1 is implied by this property and (5.9).

Acknowledgments. The authors are grateful to John Ormerod and Mike Titterington for their assistance in the preparation of this paper. This research was partially supported by Australian Research Council grants to the University of Melbourne and University of Wollongong.

References.

[1] Bates, D. and Maechler, M. (2010). `lme4`: linear mixed-effects models using S4 classes. R package. `http://www.R-project.org`

[2] Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer, New York.

[3] Boccignone, G, Napoletano, P and Ferraro, M. (2008). Embedding diffusion in variational Bayes: A technique for segmenting images . *International Journal of Pattern Recognition and Artificial Intelligence*, **22** 811–827.

[4] Flandin, G. and Penny, W.D. (2007). Bayesian fMRI data analysis with sparse spatial basis function priors. *NeuroImage*, **45** S173–S186.

[5] Hall, P., Ormerod, J.T. and Wand, M.P. (2011). Theory of Gaussian variational approximation for a Poisson linear mixed model. *Statistica Sinica* **21** 369–389.

[6] Logsdon, B.A., Hoffman, G.E. and Mezey, J.G. (2010). A variational Bayes algorithm for fast and accurate multiple locus genome-wide association analysis *BMC Bioinformatics*, **11:58** 1–13.

[7] McCulloch, C.E., Searle, S.R. and Neuhaus, J.M. (2008). *Generalized, Linear, and Mixed Models*, 2nd Edition. John Wiley & Sons, New York.

[8] McGrory, C.A., Titterington, D.M., Reeves, R. and Pettitt, A.N. (2009). Variational Bayes for estimating the parameters of a hidden Potts model. *Statistics and Computing* **19** 329–340.

[9] Minka, T., Winn, J., Guiver, J. and Kannan, A. (2010). Infer.Net 2.4, Microsoft Research Cambridge, Cambridge, UK.

[10] Ormerod, J.T. and Wand, M.P. (2010). Explaining variational approximations. *The American Statistician*, **64** 140–153.

[11] Ormerod, J.T. and Wand, M.P. (2011). Gaussian variational approximate inference for generalized linear mixed models. *Journal of Computational and Graphical Statistics*, **20** in press.

[12] R Development Core Team (2010). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, `http://www.R-project.org`

[13] Teschendorff, A.E., Wang, Y., Barbosa-Morais, N.L., Brenton, J.D. and Caldas C. (2005). A variational Bayesian mixture modelling framework for cluster analysis of gene-expression data. *Bioinformatics*, **21** 3025–3033.

Department of Mathematics and Statistics
University of Melbourne
Melbourne 3000
AUSTRALIA
E-mail: halpstat@ms.unimelb.edu.au

Centre for Statistical and Survey Methodology
School of Mathematics and Applied Statistics
University of Wollongong
Wollongong 2522
AUSTRALIA
E-mail: tung@uow.edu.au
E-mail: mwand@uow.edu.au
E-mail: sw918@uow.edu.au
URL: http://www.uow.edu.au/∼mwand.html