UPPSALA
UNIVERSITET

Working Paper 2013:9

Department of Statistics

Some Approximations of the Logistic Distribution with Application to the Covariance Matrix of Logistic Regression

Ronnie Pingel

Working Paper 2013:9
June 2013
Department of Statistics
Uppsala University
Box 513
SE-751 20 UPPSALA
SWEDEN

Working papers can be downloaded from www.statistics.uu.se

Title: Some Approximations of the Logistic Distribution with Application to the Covariance Matrix of Logistic Regression

Author: Ronnie Pingel
E-mail: ronnie.pingel@statistics.uu.se

(2013), *pp.* 1–8

Some Approximations of the Logistic Distribution with Application to the Covariance Matrix of Logistic Regression

BY R. PINGEL

Department of Statistics, Uppsala University,
SE-751 20 Uppsala, Sweden
ronnie.pingel@statistics.uu.se

SUMMARY

We show that a two component normal mixture model provides a very close approximation to the logistic distribution. This is an improvement over using the normal distribution and is on par with using the t-distribution as approximating distributions. The result from using the mixture model is exemplified by finding an approximative analytic expression for the covariance matrix of logistic regression with normally distributed random regressors.

Some key words: Correlation; Density; Distribution; Gaussian; Logit; Mixture; Normal; Shape; t-distribution

1. INTRODUCTION

If a logistic function, $F(\cdot)$, and its derivative, $f(\cdot)$ are functions of random variables then it is in general not possible to find analytic expressions for the moments of these two functions unless some heavy restrictions are used. In this paper we attempt to solve this problem by substituting the logistic function with some other function that closely resembles the logistic function. More specifically, this paper examines how well $F(\cdot)$ and $f(\cdot)$ are approximated by a normal distribution, a t-distribution and a normal mixture distribution, extending previous results such as Haley (1952) and the *Biometrika* remark by Mudholkar & George (1978). The resulting approximation using the mixture model is then applied to approximate the asymptotic covariance matrix in logistic regression having normally distributed regressors.

2. APPROXIMATION USING A NORMAL DISTRIBUTION OR A t-DISTRIBUTION

Consider a standard logistic random variable X having mean zero and variance $\pi^2/3$. Its cumulative distribution function is $F(x) = (1 + e^{-x})^{-1}$ and its density is $f(x) = F(x)\,[1 - F(x)]$. The logistic distribution is a member of a location-scale family and although not belonging to the exponential family of distributions, it is generally known that the logistic distribution is very similar to the normal distribution. Not only is the shape of the respective distributions determined by a location and a scale parameter, both distributions are bell shaped. However, the logistic distribution has heavier tails than the normal distribution. More specifically, the excess kurtosis of the logistic distribution is 1.2. Still, due to the similarities it is appealing to approximate a logistic distribution using a normal distribution. We denote the distribution function and density of a normal distribution having mean zero and standard deviation σ with $G(x)$ and $g(x)$ respectively.

A competing approximation is proposed by Mudholkar & George (1978). Due to the larger tails of the Student's t-distribution compared to the normal distribution, they suggest using a

t-distribution as approximation of the logistic distribution function. However, one drawback is that the expressions of the distribution function and the density of the t-distribution are more complicated than those of the normal distribution. In the following $T(x)$ and $t(x)$ denote the distribution function and the density of a central t-distributed random variable with ν degrees of freedom and scale parameter s.

To measure how well the approximations perform we consider two measures of accuracy: (i) the maximum absolute deviations, $||\cdot||_\infty$, or (ii) the square-root of the average of all squared deviations, $||\cdot||_2$. Only few analytic results exists in the literature, but Haley (1952) shows that $||F(x) - G(x)||_\infty$ is minimized at $\sigma \approx 1.702$. Other results, e.g. Mudholkar & George (1978), matches the moments, which yield some degree of similarity between distributions but without actually addressing (i) or (ii). In this paper we rely on numerical optimization to minimize (i) or (ii). See the comments to Table 1 regarding the optimization routine.

According to the results in Table 1, the normal distribution provides a fairly good approximation of the logistic distribution. The maximum absolute error between the normal distribution and the logistic distribution is minimized to be 0.0095 for the distribution function and 0.0115 for the density. Also, the minimum square-root of the average absolute error is 0.0231 for the distribution function and 0.0207 for the pdf. However, there are two important caveats regarding the conclusion of the approximation. First, what is considered good depends on the application at hand. Second, if another function improves the approximation and is just as easily implemented, then there is no reason not to use that function.

As can be seen in Table 1, using the t-distribution leads to a large decrease of the approximation errors. Depending on which error and function we are studying, the errors using the t-distribution is between 12 to 16 times smaller compared to the errors when using the normal distribution. In fact, the maximum absolute error between the t-distribution and the logistic

Table 1. *Minimum approximation error when approximating a logistic distribution with a normal distribution, a t-distribution or a normal mixture model*

	Minimum error ($\times 100$)	Value(s) that minimize error		
		σ		
$\|\|F(x) - G(x)\|\|_2$	2.31	1.699		
$\|\|F(x) - G(x)\|\|_\infty$	0.95	1.702		
$\|\|f(x) - g(x)\|\|_2$	2.07	1.630		
$\|\|f(x) - g(x)\|\|_\infty$	1.15	1.618		
		ν	s	
$\|\|F(x) - T(x)\|\|_2$	0.19	7.031	1.549	
$\|\|F(x) - T(x)\|\|_\infty$	0.06	6.974	1.548	
$\|\|f(x) - t(x)\|\|_2$	0.15	6.424	1.540	
$\|\|f(x) - t(x)\|\|_\infty$	0.07	6.400	1.539	
		v	ω_1	ω_2
$\|\|F(x) - H(x)\|\|_2$	0.15	0.567	1.304	2.300
$\|\|F(x) - H(x)\|\|_\infty$	0.07	0.505	1.247	2.227
$\|\|f(x) - h(x)\|\|_2$	0.17	0.478	1.243	2.168
$\|\|f(x) - h(x)\|\|_\infty$	0.08	0.460	1.231	2.143

Note: For the univariate optimization, a mixture of golden section search and parabolic interpolation is used. For the multivariate optimization the Nelder-Mead method is applied (Nelder & Mead, 1965). Since the Nelder-Mead method may not converge at a global optimum (McKinnon, 1998), we increase robustness by restarting the optimization $1,000$ times with uniformly distributed random vectors as start values. All numerical calculations are done using `R version 2.15.1`.

distribution is minimized to be 0.0006 for the distribution function and 0.0007 for the density, while the minimum square-root of the average absolute error are 0.0019 and 0.0017 for the distribution and the density respectively. Clearly, the t-distribution is far superior when it comes to minimizing the approximation errors. On the other hand, it still suffers from having complicated expressions for the distribution and density functions.

3. THE NORMAL MIXTURE APPROXIMATION

The normal mixture model is widely used when consideration must be taken to the occurrences of rare events, e.g. heavy-tailed probability models. Intuitively, the normal mixture distribution should therefore be able to take into account the heavier tails of the logistic distribution. For this purpose, we suggest as approximation the following two component normal mixture model

$$H(x) = \upsilon H_1(x) + (1-\upsilon)H_2(x), \quad -\infty < x < \infty,\, 0 < \omega_1, \leq \omega_2 < \infty, 0 < \upsilon < 1, \quad (1)$$
$$h(x) = \upsilon h_1(x) + (1-\upsilon)h_2(x), \quad -\infty < x < \infty,\, 0 < \omega_1 \leq \omega_2 < \infty, 0 < \upsilon < 1, \quad (2)$$

where $H_1(x)$, $H_2(x)$, $h_1(x)$ and $h_{_2}(x)$ are the distribution functions and density functions of two normal distributions with zero means and standard deviations ω_1 and ω_2. Again, we seek to minimize (i) and (ii), now with respect to the three parameters that govern the shape of $H(x)$ and $h(x)$. Note that a mixture model may approximate any function arbitrarily well depending on the number of components (Sorenson & Aspach, 1971). Still, a two component mixture models provides a balance between having a parsimonious model and providing a good approximation

The results in Table 1 show that the normal mixture model is a good approximation. Regarding the maximum absolute error the approximation is more or less on par with the t-distribution, with 0.0007 and 0.0008 being the maximum errors for the distribution and density function respectively. Also, this conclusion is more or less similar when instead looking at the square-root of the average absolute error. To be more specific, the error is 0.0015 for the distribution function and 0.0017 for density. Overall, the results suggests that the normal mixture model is a good alternative when trying to approximate the logistic distribution. It performs as well as the t-distribution and have the important benefit of having a simple form. The similarities of the magnitude of the errors are also apparent from Figure 1.

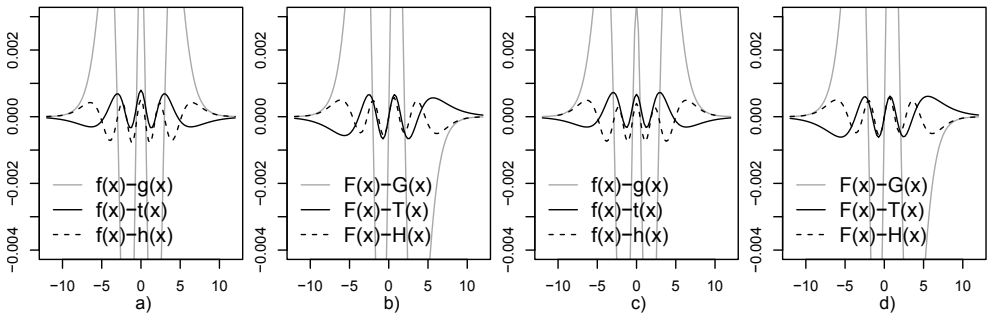

Fig. 1. Comparison of the logistic distribution $F(x)$, a normal distribution $G(x)$, a t-distribution $T(x)$, and a normal mixture distribution $H(x)$. Lower case letters denotes densities. Panels a) and b) contain distributions with values minimizing $||\cdot||_2$. Panels c) and d) contain distributions with values minimizing $||\cdot||_\infty$.

4. ON THE COVARIANCE OF LOGISTIC REGRESSION

In this section previous results are applied to facilitate modelling in logistic regression. More specifically, we are interested in the situation with stochastic regressors. This is not commonly dealt with in the literature on logistic regression, although some authors, e.g. Fahrmeir & Kaufmann (1986), Lee (1990), and most recently Ding & Chen (2006), have provided some results.

To clarify some notation define $X = \mu + Z'\gamma$, where is X a linear function of k random variables Z, a parameter μ and a parameter vector γ. Next, consider a binary outcome, W, generated by a Bernoulli trial with probability $F(X)$. From this we define a relationship between the variables Z and the outcome W. In practice we would like to be able to estimate this relationship using a sample of size n. The natural thing to do would be to regress W on Z and estimate μ and γ using maximum likelihood estimators, $\hat{\mu}$ and $\hat{\gamma}$. In other words, we would carry out a logistic regression. Under certain regularity conditions, we know that $\hat{\mu}$ and $\hat{\gamma}$ are asymptotically normally distributed with covariance matrix equal to the inverse of the information matrix. The Fisher information matrix for μ and γ can be expressed as

$$I = E\begin{pmatrix} F(X)\left(1-F(X)\right) & F(X)\left(1-F(X)\right)Z' \\ F(X)\left(1-F(X)\right)Z & F(X)\left(1-F(X)\right)ZZ' \end{pmatrix}. \tag{3}$$

See e.g. Lee (1990) for further details. Also, note that the expected information matrix is identical to the observed information matrix, as described by Nelder & Wedderburn (1972). The question that arises is if the asymptotic covariance matrix based on (3) could be calculated directly assuming that the data generating process is known. Thus the following theorem is introduced.

THEOREM 1. *Consider a logistic regression logit$[p(x)] = \mu + Z'\gamma = X$, where μ and γ are constants. Assume that Z is a k-dimensional random vector that follows a multivariate normal distribution with a zero mean vector and covariance matrix Σ, implying that $X \sim N(\mu, \sigma^2)$. The asymptotic covariance matrix for the ML estimators $\hat{\gamma}$ can then be expressed as*

$$\Lambda = \begin{pmatrix} \Lambda^{11} & \Lambda^{12} \\ \Lambda^{21} & \Lambda^{22} \end{pmatrix}, \quad \textit{where} \tag{4}$$

$$\begin{aligned}
\Lambda^{11} &= \frac{E[f(X)X^2] - 2E[f(X)X]\mu + E[f(X)]\mu^2}{E[f(X)]E[f(X)X^2] - (E[f(X)X])^2}, \\
\Lambda^{21} &= -\gamma\frac{E[f(X)X] - E[f(X)]\mu}{E[f(X)]E[f(X)X^2] - (E[f(X)X])^2}, \quad (\Lambda^{12} \textit{ being the transpose of } \Lambda^{21}) \\
\Lambda^{22} &= \frac{\Sigma^{-1}}{E\left[f(X)\right]} - \frac{\gamma\gamma'}{E\left[f(X)\right]\sigma^2} + \frac{\gamma\gamma' E\left[f(X)\right]}{E\left[f(X)\right]E\left[f(X)X^2\right] - \left(E\left[f(X)X\right]\right)^2}.
\end{aligned}$$

See the appendix for proof. Even though Σ^{-1} may be formulated in such a way that it has a closed form expression (e.g. defining it to be a diagonal or an equi-correlation matrix), Theorem 1 still involves expectations for which no analytic results are available. However, having already concluded that $f(X) \approx h(X)$ it would make sense to substitute $f(X)$ in Theorem 1 with $h(X)$. Lemma 1 provides necessary expectations.

LEMMA 1. *Let $X \sim N(\mu, \sigma^2)$ and let $h(X)$ be as defined in (2). Then $E[h(X)] = \upsilon\Omega_1 + (1-\upsilon)\Omega_2$, $E[h(X)X] = \upsilon\Omega_1\mu_1 + (1-\upsilon)\Omega_2\mu_2$ and $E[h(X)X^2] = \upsilon\Omega_1(\sigma_1^2 + \mu_1^2) + (1-\upsilon)\Omega_2(\sigma_2^2 + \mu_2^2)$, where $\mu_1 = \mu\omega_j^2/(\omega_1^2 + \sigma^2)$, $\sigma_1^2 = \omega_1^2\sigma^2/(\omega_1^2 + \sigma^2)$ and $\Omega_1 = \exp(-\frac{\mu^2}{2(\omega_1^2+\sigma^2)})(2\pi)^{-1/2}(\omega_1^2 + \sigma^2)^{-1/2}$. Substituting ω_1 with ω_2 gives μ_2, σ_2^2 and Ω_2.*

See the appendix for proof. Thus, the covariance matrix in Theorem 1 is approximated with the matrix $\tilde{\Lambda}$ where the corresponding expectations in (4) are simply replaced with $E[h(X)]$, $E[h(X)X]$ and $E[h(X)X^2]$. Table 1 provides the values of v, ω_1 and ω_2 that minimize $||f(x) - h(x)||_2$, which are used to get $E[h(X)]$. Using the same optimization routine as before $||xf(x) - xh(x)||_2$ is minimized at $v = 0.6095$, $\omega_1 = 1.3344$ and $\omega_2 = 2.3802$ and $||x^2 f(x) - x^2 h(x)||_2$ is minimized at $v = 0.6135$, $\omega_1 = 1.3456$ and $\omega_2 = 2.3709$. These values are inserted into the expressions in Lemma 1 to get $E[h(X)X]$ and $E[h(X)X^2]$.

To check how well the approximation works recall Theorem 1 and let λ_{ij} denote an element in Λ. Similarly, $\tilde{\lambda}_{ij}$ is an element in $\tilde{\Lambda}$. The approximation is evaluated using the maximum relative error, $\max\{|(\tilde{\lambda}_{ij} - \lambda_{ij})|/\lambda_{ij}|\}$, and we consider two different setups, A and B. In setup A, X has a bivariate normal distribution with a zero mean vector and a correlation matrix with correlation ρ. We let γ_1 and γ_2 each take the 199 values $\{0.1, 0.15 \ldots, 10\}$ and ρ take the 10 values $\{0, 0.1, \ldots, 0.9\}$. Also, $\mu = 0.1$. For each of the $199,000$ unique combinations the maximum relative error is calculated. In setup B, X has a trivariate normal distribution, again with a zero mean vector, but now with a correlation matrix having correlations ρ_{12} ρ_{13} and ρ_{23} set to be 0, 0.3 or 0.9. Furthermore, an element in γ may take the value 0.2, 1 or 5. Again, we put $\mu = 0.1$.

The results from both setups give similar conclusions. In general the approximation seem to work well. In Setup A, 99 percent of the approximated covariance matrices have a relative error less than 10 percent, which drops to a maximum relative error of one percent for the 90th percentile. Also, most of the approximations in Setup B seem to work well having maximum relative errors ranging from 0.001 to 0.0786.

However, what seems to be case is that small values of γ could make the approximations less accurate, and sometimes really bad. This can be seen by comparing the middle and right panels in Figure 2. The proportion of large errors is bigger among the errors based on combinations when values in γ are small, and for one percent of the cases the relative error is at least 100 percent. However, this does not mean that the absolute error is large. In fact, in many cases small numerical differences could inflate the relative error when variances and covariances in the covariance matrix are close to zero. The same conclusion regarding small values in γ is suggested by Table 2, where the largest maximum relative error occurs when γ is at its smallest. Also, it

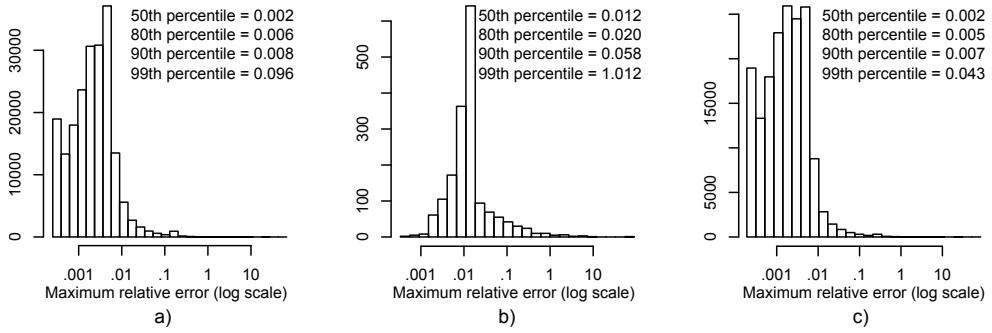

Fig. 2. Maximum relative error when approximating the covariance matrix from logistic regression using normal mixture approximation. Panel a) shows errors for 199,000 combinations generated by setup A. Panel b) shows errors for a subset of A containing 1,710 combinations where γ_1 and γ_2 are smaller than 1. Panel c) shows errors for a subset of A, containing 164,710 combinations where γ_1 and γ_2 are at least 1.

Table 2. *Maximum relative error when approximating the covariance matrix of a logistic distribution using a normal mixture model (Setup B)*

γ	ρ	Maximum relative error
$(0.2, 0.2, 0.2)$	$(0, 0, 0)$	0.0786
$(0.2, 0.2, 0.2)$	$(0.3, 0.3, 0.3)$	0.0115
$(0.2, 0.2, 0.2)$	$(0.9, 0.9, 0.9)$	0.0098
$(0.2, 0.2, 0.2)$	$(0, 0.3, 0.9)$	0.0051
$(1, 1, 1)$	$(0, 0, 0)$	0.0040
$(1, 1, 1)$	$(0.3, 0.3, 0.3)$	0.0107
$(1, 1, 1)$	$(0.9, 0.9, 0.9)$	0.0050
$(1, 1, 1)$	$(0, 0.3, 0.9)$	0.0030
$(5, 5, 5)$	$(0, 0, 0)$	0.0023
$(5, 5, 5)$	$(0.3, 0.3, 0.3)$	0.0010
$(5, 5, 5)$	$(0.9, 0.9, 0.9)$	0.0005
$(5, 5, 5)$	$(0, 0.3, 0.9)$	0.0013
$(0.2, 1, 5)$	$(0, 0, 0)$	0.0057
$(0.2, 1, 5)$	$(0.3, 0.3, 0.3)$	0.0262
$(0.2, 1, 5)$	$(0.9, 0.9, 0.9)$	0.0046
$(0.2, 1, 5)$	$(0, 0.3, 0.9)$	0.0048
$(0.2, 1, 5)$	$(0.9, 0.3, 0)$	0.0056

should be emphasized that small values in γ does not necessarily give large approximation error. Actually, for most cases the approximation still works well.

As a final check this section ends with a small simulation exercise. For the 17 designs in Setup B we compare the coverage rates of 95% Wald-type confidence intervals for μ and γ using standard errors based on either (4) or the approximative covariance matrix. Using 10,000 replicates, each with a sample size of 1,000, we find that we are not able to distinguish between the coverage rates, which in most cases are close to the nominal level of 0.95. Moreover, both methods also fail when $\gamma = (5, 5, 5)$ having coverage rates around 0.92 for γ.

ACKNOWLEDGEMENT

I would like to thank Rolf Larsson, Johan Lyhagen and Ingeborg Waernbaum for valuable comments and suggestions.

APPENDIX

Proof of Theorem 1. Assume that a k-dimensional random vector Z has a multivariate normal distribution with a zero mean vector and covariance matrix Σ and let μ and γ be a scalar and a k-dimensional parameter vector. Defining $X = \mu + Z'\gamma$, then X is normally distributed, with mean μ and variance $\gamma'\Sigma\gamma = \sigma^2$. It also follows that Z and X are distributed as a $k+1$ multivariate normal distribution with mean vector $(0', \mu)$ and covariance matrix

$$\Psi = \begin{pmatrix} \Sigma & \gamma'\Sigma \\ \Sigma\gamma & \gamma'\Sigma\gamma \end{pmatrix}.$$

Standard results (see e.g. Mardia et al. (1979)) provides the conditional means $E[Z|X] = (X-\mu)\Sigma\gamma/\sigma^2$ and $E[ZZ'|X] = \Sigma + \Sigma\gamma\gamma'\Sigma[(X-\mu)^2 - \sigma^2]/\sigma^4$. We now derive an expression for the Fisher information matrix for μ and γ in (3). The logistic model gives $F(X)(1-F(X)) = f(X) = e^X/(1+e^X)^2$

which makes

$$I = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix} = E \begin{pmatrix} f(X) & f(X)Z' \\ f(X)Z & f(X)ZZ' \end{pmatrix}$$

a natural partitioning. Except for the obvious $I_{11} = E\left[f(X)\right]$, we then apply iterated expectations and insert the previously shown conditional means $E[Z|X]$ and $E[ZZ'|X]$ giving

$$\begin{aligned} I_{21} &= E\left[f(X)Z\right] = E\left[f(X)(X-\mu)\right]\Sigma\gamma/\sigma^2, \\ I_{22} &= E\left[f(X)ZZ'\right] = E\left[f(X)\left(\Sigma + \Sigma\gamma\gamma'\Sigma\left[(X-\mu)^2 - \sigma^2\right]/\sigma^4\right)\right]. \end{aligned}$$

Note that I_{12} is just the transpose of I_{21}. In order to find I^{-1}, we make use of results found in e.g. Abadir & Magnus (2005). For a partitioned matrix with scalar blocks, such as I,

$$I^{-1} = \begin{pmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{pmatrix} = \begin{pmatrix} 0 & 0' \\ 0 & I_{22}^{-1} \end{pmatrix} + \frac{1}{I_{11}^{-1} - I_{12}I_{22}^{-1}I_{21}} \begin{pmatrix} -1 \\ I_{22}^{-1}I_{21} \end{pmatrix} \begin{pmatrix} -1 & I_{12}I_{22}^{-1} \end{pmatrix}.$$

To find I_{22}^{-1} define $A \equiv E\left[f(X)\right]\Sigma$, $C \equiv \sqrt{B}\Sigma\gamma$ and $B \equiv E\left[f(X)\left[(X-\mu)^2 - \sigma^2\right]/\sigma^4\right]$. Applying the Binomial Inverse Theorem on I_{22}^{-1} gives

$$I_{22}^{-1} = (A + CC')^{-1} = \frac{\Sigma^{-1}\sigma^2 - \gamma\gamma'}{E\left[f(X)\right]\sigma^2} + \frac{\gamma\gamma'}{E\left[f(X)(X-\mu)^2\right]}.$$

Now we can show that

$$\begin{aligned} I_{11} - I_{12}I_{22}^{-1}I_{21} =& I_{11} - \left(E\left[f(X)(X-\mu)\right]\right)^2 \times \frac{1}{\sigma^4}\gamma'\Sigma\left(\frac{\Sigma^{-1}\sigma^2 - \gamma\gamma'}{E\left[f(X)\right]\sigma^2} + \frac{\gamma\gamma'}{E\left[f(X)(X-\mu)^2\right]}\right)\Sigma\gamma \\ =& E\left[f(X)\right] - \frac{\left(E\left[f(X)(X-\mu)\right]\right)^2}{\left(E\left[f(X)(X-\mu)^2\right]\right)}, \\ I_{22}^{-1}I_{21} =& \gamma\frac{E[f(X)(X-\mu)]}{E[f(X)(X-\mu)^2]}, \\ I_{22}^{-1}I_{21}I_{12}I_{22}^{-1} =& \gamma\gamma'\frac{\left(E\left[f(X)(x-\mu)\right]\right)^2}{\left(E\left[f(X)(X-\mu)^2\right]\right)^2}. \end{aligned}$$

From these results it is straightforward, albeit somewhat tedious, to show the asymptotic covariance matrix in Theorem 1 where $\Lambda = I^{-1}$. □

Proof of Lemma 1. Consider $h_1(x)u(x)$ where $h_1(x)$ is as defined in (2) and $u(x)$ is the pdf of a random variable $X \sim N(\mu, \sigma^2)$. Completing the squares in the exponents and multiplying with constants directly gives that $h_1(x)u(x)$ is proportional to the density of a normal distribution with mean $\mu_1 = \mu\omega_1^2/(\omega_1^2 + \sigma^2)$ and variance $\sigma_1^2 = \omega_1^2\sigma^2/(\omega_1^2 + \sigma^2)$ and proportionality constant $\Omega_1 = \exp(-\frac{\mu^2}{2(\omega_1^2+\sigma^2)})(2\pi)^{-1/2}(\omega_1^2 + \sigma^2)^{-1/2}$. The results of $h_2(x)u(x)$ follows from substituting ω_1 with ω_2. From here it is straightforward to show $E[h(X)]$, $E[h(X)X]$ and $E[h(X)X^2]$

REFERENCES

ABADIR, K. & MAGNUS, J. (2005). *Matrix Algebra*. Cambridge: Cambridge University Press.

DING, J. L. & CHEN, X. R. (2006). Asymptotic properties of the maximum likelihood estimate in generalized linear models with stochastic regressors. *Acta Mathematica Sinica* **22**, 1679–1686.

FAHRMEIR, L. & KAUFMANN, H. (1986). Asymptotic inference in discrete response models. *Statistische Hefte* **27**, 179–205.

HALEY, D. (1952). Estimation of the dosage mortality relationship when the dose is subject to error. *Applied Mathematics and Statistics Laboratory, Stanford University, Technical report* .

LEE, K.-W. (1990). Consistency and asymptotic normality of the maximurn likelihood estimator in generalized linear models. *Communications in Statistics - Theory and Methods* **19**, 2527–2539.

MARDIA, K., KENT, J. & BIBBY, J. (1979). *Multivariate Analysis*. London: Academic Press.

MCKINNON, K. I. M. (1998). Convergence of the Nelder–Mead simplex method to a non-stationary point. *SIAM Journal on Optimization* **9**, 148–158.

MUDHOLKAR, G. & GEORGE, O. (1978). A remark on the shape of the logistic distribution. *Biometrika* **65**, 667–668.

NELDER, J. A. & MEAD, R. (1965). A simplex method for function minimization. *Computer Journal* **7**, 308–313.

NELDER, J. A. & WEDDERBURN, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A* **135**, 370–384.

SORENSON, H. W. & ASPACH, D. L. (1971). Recursive bayesian estimation using gaussian sums. *Automatica* **7**, 465–479.