

# Variances of Fitted Category Probabilities and Logits by the Delta Method for the Nested Logit Model

John Fox

May 25, 2023

## 1 Notation

Let  $\phi^{(j)}$  represent the probability that the dichotomous response in the  $j$ th nested dichotomous logit model is  $Y_j = 1$  (i.e., a “success”),  $j = 1, \dots, m-1$ , where  $m$  is the number of response categories for the polytomy. Then  $1 - \phi^{(j)}$  is the probability that  $Y_j = 0$  (i.e., a “failure”). I assume that the regression coefficients and their covariance matrix for each dichotomous logit model are estimated in the usual manner. Let  $\lambda^{(j)} = \log[\phi^{(j)}/(1 - \phi^{(j)})]$  represent the (estimated) logit (log-odds) for the  $j$ th dichotomous logit model, with variance  $V(\lambda^{(j)})$  (see Section 2.1).

Let  $\phi_k$ ,  $k = 1, \dots, m$  represent the probability that the polytomous response is  $Y = k$ . Let  $\hat{\phi}^{(j)}$  and  $\hat{\phi}_k$  represent the estimates of these probabilities. In the sequel, which involves only the estimates of these and other parameters, I’ll omit the hats so as to simplify the notation.

In the nested logit model, the polytomous probabilities  $\phi_k$  are each products of probabilities  $\phi^{(j)}$  or  $1 - \phi^{(j)}$  for  $j \in \mathcal{M}_k \subset \{1, \dots, m-1\}$ ; that is  $\mathcal{M}_k$  is the subset of the dichotomous logit models that enter into  $\phi_k$ . Let  $\phi^{(j,k_j)}$  represent either  $\phi^{(j)}$  or  $1 - \phi^{(j)}$ , as appropriate for category  $k$  of the polytomous response. Then

$$\phi_k = \prod_{j \in \mathcal{M}_k} \phi^{(j,k_j)}$$

for  $k = 1, \dots, m$ .

Finally, the individual-category probabilities  $\phi_k$  can be converted into logits,  $\lambda_k = \log[\phi_k/(1 - \phi_k)]$ . The estimates of these logits should approach asymptotic normality more rapidly than the estimates of the corresponding probabilities.

### 1.1 An Example

I’ll use the following example to illustrate the results in this document: Suppose that we have a three-category response variable  $Y$  with categories  $A$ ,  $B$ , and  $C$ , and define the two nested dichotomies  $Y_1$  coded 0 or 1 for categories  $\{A\}$  and

$\{B, C\}$ , respectively, and  $Y_2$  coded 0 and 1 for categories  $\{B\}$  and  $\{C\}$ . Then  $\phi^{(1)} = \Pr(Y_1 = 1) = \Pr(\{B, C\})$ ;  $1 - \phi^{(1)} = \Pr(Y_1 = 0) = \Pr(\{A\})$ . As well,  $\mathcal{M}_A = \{1\}$  and  $\mathcal{M}_B = \mathcal{M}_C = \{1, 2\}$ . Consequently,

$$\begin{aligned}\phi_A &= \phi^{(1, A_1)} = 1 - \phi^{(1)} \\ \phi_B &= \phi^{(1, B_1)} \phi^{(2, B_2)} = \phi^{(1)} (1 - \phi^{(2)}) \\ \phi_C &= \phi^{(1, C_1)} \phi^{(2, C_2)} = \phi^{(1)} \phi^{(2)}\end{aligned}$$

Here, I abuse the notation slightly in the interest of clarity, using letters rather than numbers for the response categories, so the index of response categories,  $k$ , takes on the values  $A$ ,  $B$ , and  $C$ , rather than 1, 2, and 3.

## 2 Variances of the Estimated Probabilities

### 2.1 Variances of Estimated Probabilities for the Dichotomous Logit Models

The estimated probability of success  $\phi^{(j)}$  for the  $j$ th dichotomous logit model is

$$\phi^{(j)} = \frac{1}{1 + e^{-\lambda^{(j)}}}$$

Then

$$\begin{aligned}\lambda^{(j)} &= \alpha^{(j)} + \beta_1^{(j)} x_1 + \cdots + \beta_p^{(j)} x_p \\ &= \mathbf{x}^T \boldsymbol{\beta}\end{aligned}$$

is a function of the regression coefficients, where  $\mathbf{x}^T = [1, x_1, \dots, x_p]$  (an arbitrary vector of values of the regressors) and  $\boldsymbol{\beta} = [\alpha, \beta_1, \dots, \beta_p]^T$ . The probability of failure is

$$1 - \phi^{(j)} = \frac{1}{1 + e^{\lambda^{(j)}}}$$

The variance of the logit is  $V(\lambda^{(j)}) = \mathbf{x}^T V(\boldsymbol{\beta}) \mathbf{x}$ ,

The derivatives of  $\phi^{(j)}$  and  $1 - \phi^{(j)}$  with respect to  $\lambda^{(j)}$  are

$$\begin{aligned}\frac{d\phi^{(j)}}{d\lambda^{(j)}} &= \frac{e^{-\lambda^{(j)}}}{(1 + e^{-\lambda^{(j)}})^2} \\ \frac{d(1 - \phi^{(j)})}{d\lambda^{(j)}} &= -\frac{e^{\lambda^{(j)}}}{(1 + e^{\lambda^{(j)}})^2}\end{aligned}$$

By the univariate delta method,

$$\begin{aligned}
V(\phi^{(j)}) &\approx \left( \frac{d\phi^{(j)}}{d\lambda^{(j)}} \right)^2 V(\lambda^{(j)}) \\
&= \left[ \frac{e^{-\lambda^{(j)}}}{(1 + e^{-\lambda^{(j)}})^2} \right]^2 V(\lambda^{(j)}) \\
V(1 - \phi^{(j)}) &\approx \left[ \frac{d(1 - \phi^{(j)})}{d\lambda^{(j)}} \right]^2 V(\lambda^{(j)}) \\
&= \left[ \frac{e^{\lambda^{(j)}}}{(1 + e^{\lambda^{(j)}})^2} \right]^2 V(\lambda^{(j)}) \\
&= V(\phi^{(j)})
\end{aligned}$$

## 2.2 Variances of Estimated Probabilities for the Nested Logit Model

The variances of the estimated response-category probabilities for the polytomous response can be obtained similarly by the multivariate delta method, recognizing that these probabilities are products of the dichotomous probabilities. The result is greatly simplified because the dichotomies are independent, and so the covariance matrix of the estimated dichotomous probabilities is diagonal.

The required derivatives are

$$\frac{\partial \phi_k}{\partial \phi^{(j, k_j)}} = \prod_{j' \in \mathcal{M}_k, j' - \{j\}} \phi^{(j', k_j)}$$

for  $j \in \mathcal{M}_k$  and  $k = 1, \dots, m$ . Here,  $-$  denotes set difference. Because  $V(\phi^{(j)}) = V(1 - \phi^{(j)})$ , it's always the case that  $V(\phi^{(j, k_j)}) = V(\phi^{(j)})$ , and so

$$\begin{aligned}
V(\phi_k) &\approx \sum_{j \in \mathcal{M}_k} \left( \frac{\partial \phi_k}{\partial \phi^{(j, k_j)}} \right)^2 V(\phi^{(j)}) \\
&= \sum_{j \in \mathcal{M}_k} \left( \prod_{j' \in \mathcal{M} - \{j\}} \phi^{(j', k_j)} \right)^2 V(\phi^{(j)})
\end{aligned}$$

for  $k = 1, \dots, m$ .

Applying these results to the example, recall, first, that  $\mathcal{M}_A = \{1\}$  and so the set for the product  $\prod_{j' \in \mathcal{M}_A - \{j\}} \phi^{(j', A_j)}$ , that is,  $j' \in \{1\} - \{1\}$ , is empty. In this case, the product is taken to be 1, and  $V(\phi_A) = V(\phi^{(1)})$ . That makes intuitive sense, because, as noted previously,  $\phi_A = 1 - \phi^{(1)}$ .

Proceeding with  $B$  and  $C$ ,  $\mathcal{M}_B = \mathcal{M}_C = \{1, 2\}$ . Consequently, each product  $\prod_{j' \in \mathcal{M}_B - \{j\}} \phi^{(j', B_j)}$  and  $\prod_{j' \in \mathcal{M}_C - \{j\}} \phi^{(j', C_j)}$  has only one term, for  $j' = 2$  in

the case of  $B$  or  $j' = 1$  in the case of  $C$ :

$$\begin{aligned}
V(\phi_B) &= \phi^{(2,B_2)} V(\phi^{(1)}) + \phi^{(1,B_1)} V(\phi^{(2)}) \\
&= (1 - \phi^{(2)}) V(\phi^{(1)}) + \phi^{(1)} V(\phi^{(2)}) \\
V(\phi_C) &= \phi^{(2,C_2)} V(\phi^{(1)}) + \phi^{(1,C_1)} V(\phi^{(2)}) \\
&= \phi^{(2)} V(\phi^{(1)}) + \phi^{(1)} V(\phi^{(2)})
\end{aligned}$$

Yet another application of the delta method produces approximate variances for the individual-category logits. The relevant derivative is

$$\frac{d\lambda_k}{d\phi_k} = \frac{1}{\phi_k(1 - \phi_k)}$$

for  $k = 1, \dots, m$ , and so

$$\begin{aligned}
V(\lambda_k) &\approx \left( \frac{d\lambda_k}{d\phi_k} \right)^2 V(\phi_k) \\
&= \left[ \frac{1}{\phi_k(1 - \phi_k)} \right]^2 V(\phi_k)
\end{aligned}$$