Solution to Exercise 11.21

⋆11.21 In terms of the notation of the DCAR, regression (11.42), the probability Π_{tj} that $y_t = j$, $j = 0, \ldots, J$, for the nested logit model is given by expression (11.40). Show that, if the index $i(j)$ is such that $j \in A_{i(j)}$, the partial derivative of Π_{tj} with respect to θ_i, evaluated at $\theta_k = 1$ for $k = 1, \ldots, m$, where m is the number of subsets A_k, is

$$\frac{\partial \Pi_{tj}}{\partial \theta_i} = \Pi_{tj}\big(\delta_{i(j)i} v_{tj} - \sum_{l \in A_i} \Pi_{tl} v_{tl}\big). \tag{S11.26}$$

Here $v_{tj} \equiv -\boldsymbol{W}_{tj}\boldsymbol{\beta}^j + h_{ti(j)}$, where h_{ti} denotes the inclusive value (11.39) of subset A_i, and δ_{ij} is the Kronecker delta.

When $\theta_k = 1$, $k = 1, \ldots, m$, the nested logit probabilities reduce to the multinomial logit probabilities (11.34). Show that, if the Π_{tj} are given by (11.34), then the vector of partial derivatives of Π_{tj} with respect to the components of $\boldsymbol{\beta}^l$ is $\Pi_{tj}\boldsymbol{W}_{tl}(\delta_{jl} - \Pi_{tl})$.

From equation (11.40), we have

$$\Pi_{tj} = \frac{\exp(\boldsymbol{W}_{tj}\boldsymbol{\beta}^j/\theta_{i(j)})}{\sum_{l \in A_{i(j)}} \exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l/\theta_{i(j)})} \frac{\exp(\theta_{i(j)} h_{ti(j)})}{\sum_{k=1}^m \exp(\theta_k h_{tk})}, \tag{S11.27}$$

where, from (11.39),

$$h_{ti} = \log\Big(\sum_{l \in A_i} \exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l/\theta_i)\Big). \tag{S11.28}$$

Recall from Exercise 11.17, that, when $\theta_k = 1$ for $k = 1, \ldots, m$, we have

$$\Pi_{tj} = \frac{\exp(\boldsymbol{W}_{tj}\boldsymbol{\beta}^j)}{\sum_{l=0}^J \exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)}. \tag{S11.29}$$

Suppose first that $i \neq i(j)$. Then the first big fraction in (S11.27) does not depend on θ_i. Since the only one of the θ_k on which h_{ti} depends is θ_i, the numerator of the second big fraction in (S11.27) does not depend on θ_i either.

The only term in the denominator of the second fraction that depends on θ_i is $\exp(\theta_i h_{ti})$, and the derivative of this term with respect to θ_i is

$$\exp(\theta_i h_{ti})\Big(h_{ti} + \frac{\partial h_{ti}}{\partial \theta_i}\Big). \tag{S11.30}$$

Thus we can see from (S11.28) that

$$\frac{\partial h_{ti}}{\partial \theta_i} = -\frac{1}{\theta_i^2} \frac{\sum_{l \in A_i} \boldsymbol{W}_{tl}\boldsymbol{\beta}^l \exp\big(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l/\theta_i\big)}{\sum_{l \in A_i} \exp\big(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l/\theta_i\big)}. \tag{S11.31}$$

To compute the derivative at $\theta_k = 1$, $k = 1, \ldots, m$, we note that the denominator of the second fraction, evaluated at $\theta_k = 1$, reduces to

$$\begin{aligned}\sum_{k=1}^{m} \exp(h_{tk}) &= \sum_{k=1}^{m} \exp\log\Big(\sum_{l\in A_k} \exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)\Big)\\ &= \sum_{k=1}^{m}\sum_{l\in A_k} \exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l) = \sum_{l=0}^{J} \exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l). \qquad \text{(S11.32)}\end{aligned}$$

Thus when $\theta_k = 1$, the derivative $\partial\Pi_{tj}/\partial\theta_i$, for $i \neq j(i)$, is the whole expression (S11.27), evaluated at $\theta_k = 1$, divided by the negative of the denominator given in (S11.32), and multiplied by the derivative (S11.30), which is also evaluated at $\theta_k = 1$.

Using (S11.31) for $\partial h_{ti}/\partial\theta_i$, we find that

$$\begin{aligned}&\frac{\partial\Pi_{tj}}{\partial\theta_i}\bigg|_{\theta_k=1} = \\ &\quad -\Pi_{tj}\frac{\sum_{l\in A_i}\exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)}{\sum_{l=0}^{J}\exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)}\bigg(h_{ti} - \frac{\sum_{l\in A_i}\boldsymbol{W}_{tl}\boldsymbol{\beta}^l\exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)}{\sum_{l\in A_i}\exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)}\bigg).\end{aligned} \qquad \text{(S11.33)}$$

From equation (S11.29), we see that the ratio of the two summations immediately following the factor of Π_{tj} on the right-hand side of (S11.33) is, since everything is now evaluated at $\theta_k = 1$, equal to $\sum_{l\in A_i}\Pi_{tl}$. If we next look at the ratio of the two summations in the large parentheses at the end of the expression, we see that the denominator cancels with the numerator of the ratio outside the parentheses. What is left of the product of the two ratios is therefore

$$\frac{\sum_{l\in A_i}\boldsymbol{W}_{tl}\boldsymbol{\beta}^l\exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)}{\sum_{l=0}^{J}\exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)} = \sum_{l\in A_i}\boldsymbol{W}_{tl}\boldsymbol{\beta}^l\Pi_{tl},$$

where we make use of (S11.29). Putting together these last two simplifications, we find that (S11.33) reduces to

$$\frac{\partial\Pi_{tj}}{\partial\theta_i}\bigg|_{\theta_k=1} = \Pi_{tj}\sum_{l\in A_i}\Pi_{tl}(-h_{ti} + \boldsymbol{W}_{tl}\boldsymbol{\beta}^l). \qquad \text{(S11.34)}$$

Recall that, in the question, we made the definition

$$v_{tj} = -\boldsymbol{W}_{tj}\boldsymbol{\beta}^j + h_{ti(j)}.$$

If $l \in A_i$, then $i(l) = i$, and $v_{tl} = -\boldsymbol{W}_{tl}\boldsymbol{\beta}^l + h_{ti}$. Thus the right-hand side of equation (S11.34) is just $-\Pi_{tj}\sum_{l\in A_i}\Pi_{tl}v_{tl}$, as given in (S11.26).

If $i = i(j)$, there are three other contributions to the derivative. The first, coming from the numerator of the first big fraction in (S11.27), is the whole expression (evaluated at $\theta_k = 1$) multiplied by $-\boldsymbol{W}_{tj}\boldsymbol{\beta}^j$, that is $-\Pi_{tj}\boldsymbol{W}_{tj}\boldsymbol{\beta}^j$. The second, coming from the numerator of the second big fraction, is the whole expression multiplied by the derivative of $\theta_i h_{ti}$. The third, coming from the denominator of the first fraction, is the whole expression divided by the negative of that denominator, multiplied by the derivative of the denominator at $\theta_k = 1$, which is $-\sum_{l\in A_i} \boldsymbol{W}_{tl}\boldsymbol{\beta}^l \exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)$. Thus this third contribution is Π_{tj} times

$$\frac{\sum_{l\in A_i} \boldsymbol{W}_{tl}\boldsymbol{\beta}^l \exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)}{\sum_{l\in A_i} \exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)}.$$

The derivative of $\theta_i h_{ti}$ at $\theta_i = 1$ is $h_{ti} + \partial h_{ti}\partial\theta_i$, and this has already been calculated; it is the expression in large parentheses at the end of (S11.33). Thus the sum of the second and third contributions is just $\Pi_{tj}h_{ti}$. Adding in the first contribution gives $\Pi_{tj}(h_{ti} - \boldsymbol{W}_{tj}\boldsymbol{\beta}^j) = \Pi_{tj}v_{tj}$, since $i(j) = i$. This is just the term in (S11.26) that is multiplied by the Kronecker delta $\delta_{i(j)i}$, and so (S11.26) is now seen to be fully correct.

For the second part of the question, the probabilities are given by expression (S11.29), which we rewrite here for convenience:

$$\Pi_{tj} = \frac{\exp(\boldsymbol{W}_{tj}\boldsymbol{\beta}^j)}{\sum_{l=0}^J \exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)}, \tag{S11.29}$$

When $j \neq l$, the derivative of the numerator is 0. Therefore

$$\frac{\partial\Pi_{tj}}{\partial\boldsymbol{\beta}^l} = \frac{-\exp(\boldsymbol{W}_{tj}\boldsymbol{\beta}^j)\exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)\boldsymbol{W}_{tl}}{\big(\sum_{l=0}^J \exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)\big)^2} = -\Pi_{tj}\Pi_{tl}\boldsymbol{W}_{tl} \tag{S11.35}$$

for $j \neq l$. When $j = l$, the derivative of the numerator is $\exp(\boldsymbol{W}_{tj}\boldsymbol{\beta}^j)\boldsymbol{W}_{tj}$. Therefore, there are two terms instead of one, and we see that

$$\begin{aligned}\frac{\partial\Pi_{tj}}{\partial\boldsymbol{\beta}^j} &= \frac{\exp(\boldsymbol{W}_{tj}\boldsymbol{\beta}^j)\big(\sum_{l=0}^J \exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)\big)\boldsymbol{W}_{tj}}{\big(\sum_{l=0}^J \exp(\boldsymbol{W}_{tl}\boldsymbol{\beta}^l)\big)^2} - \Pi_{tj}\Pi_{tj}\boldsymbol{W}_{tj} \\ &= \Pi_{tj}\boldsymbol{W}_{tj} - \Pi_{tj}\Pi_{tj}\boldsymbol{W}_{tj} = \Pi_{tj}(1-\Pi_{tj})\boldsymbol{W}_{tj}. \end{aligned} \tag{S11.36}$$

Using the Kronecker delta, the results (S11.35) for the case in which $j \neq l$ and (S11.36) for the case in which $j = l$ can be written more compactly as

$$\frac{\partial\Pi_{tj}}{\partial\boldsymbol{\beta}^l} = \Pi_{tj}\boldsymbol{W}_{tl}(\delta_{jl} - \Pi_{tl}), \tag{S11.37}$$

which is what we were asked to show.