Bayesian Analysis (2011) **6**, Number 2, pp. 353–356

Bayesian Estimation of the Multinomial Logit Model: A Comment on Holmes and Held, "Bayesian Auxiliary Variable Models for Binary and Multinomial Regression"

Ralf van der Lans*

Abstract. This note provides two corrections to the pseudo-code of the algorithm for the Bayesian estimation of the multinomial logit model using auxiliary variables as developed by Holmes and Held (2006). After incorporating the two corrections, the algorithm works correctly for the multinomial as well as the binary logit model.

Keywords: Auxiliary variables, Bayesian multinomial regression, Markov chain Monte Carlo

This note provides two corrections to the pseudo-code of the algorithm for the Bayesian estimation of the multinomial logit model using auxiliary variables as developed by Holmes and Held (2006). The first correction involves the computation of C_{ij} (line -3, p. 166) as introduced in line 2 of equation (15) of Holmes and Held. This value should be $C \leftarrow \text{sum}\left(\exp\left(X\left[j,\right]\beta\left[,-q,i\right]\right)\right)+1$, instead of $C \leftarrow \text{sum}\left(\exp\left(X\left[j,\right]\beta\left[,-q,i\right]\right)\right)$. The second correction involves sampling of the regression coefficients β_j. Sampling of these coefficients can be corrected in two ways. One solution is to add the value C_{ij} to the auxiliary variables z_{ij}, such that $Z\left[j,q\right] \sim \text{Lo}\left(m-\log C,1\right)\text{Ind}\left(Y\left[j,q\right],Z\left[j,q\right]\right)+\log C$, instead of $Z\left[j,q\right] \sim \text{Lo}\left(m-\log C,1\right)\text{Ind}\left(Y\left[j,q\right],Z\left[j,q\right]\right)$ (line 2, p. 167)[1]. An alternative solution is to subtract $\log C_{ij}$ from m (line -4, p. 166), such that $m \leftarrow X\left[j,\right]\beta\left[,q,i\right]-\log C$ instead of $m \leftarrow X\left[j,\right]\beta\left[,q,i\right]$, and to incorporate C_{ij} in the computation of B (line -9, p. 166), such that $B \leftarrow VX^T\Lambda\left[,,q\right]^{-1}\left(Z\left[,q\right]+\log C\right)$ instead of $B \leftarrow VX^T\Lambda\left[,,q\right]^{-1}Z\left[,q\right]$. In this alternative solution, because we now correct m in its definition, we need to remove this correction from the computation of the auxiliary variables z_{ij}, such that $Z\left[j,q\right] \sim \text{Lo}\left(m,1\right)\text{Ind}\left(Y\left[j,q\right],Z\left[j,q\right]\right)$, instead of $Z\left[j,q\right] \sim \text{Lo}\left(m-\log C,1\right)\text{Ind}\left(Y\left[j,q\right],Z\left[j,q\right]\right)$ (line 2, p. 167). Furthermore, because we need $\log C$ to compute m and B in this second solution, we need to compute $\log C$ before computing m and B. Note that after incorporating these corrections in Algorithm A5, we can use this algorithm for the multinomial as well as the binary logit model. Below, I explain these two corrections in more detail.

Correction 1: $C \leftarrow \text{sum}\left(\exp\left(X\left[j,\right]\beta\left[,-q,i\right]\right)\right)+1$

On page 155, Holmes and Held state under expression (13) that β_Q is set to zero to identify the parameter estimates of the multinomial logit. This restriction sets the

*Department of Marketing, Hong Kong University of Science and Technology, Hong Kong, mailto:rlans@ust.hk

[1]Note that the definition of C_{ij} (line 2, equation 15, p. 155), equals $\log C$ in the pseudo-code of Appendix A5.

 DOI:10.1214/11-BA613

term $\exp(x_i\beta_Q) = 1$. Subsequently, in Section 3.1, when doing Gibbs sampling, the polychotomous problem is brought back to a binary problem by considering $y_i = j$ versus $y_i \neq j$. To apply the auxiliary variable Gibbs sampler as introduced in Section 2.3, Holmes and Held correct the mean of the auxiliary variables z_i (line 2, equation 8) by the term C_{ij}. According to formula (15, line 2) in Holmes and Held, C_{ij} equals $\log \sum_{k\neq j} \exp(x_i\beta_k)$. Note that the sum in this expression runs over all classes $k = 1, .., j-1, j+1, .., Q$, except class j, but including class Q. Therefore, in the pseudo-code in Algorithm A5, for each iteration i, for each observation j, and for $q = q, .., Q-1$, we must take the term $\exp(x_i\beta_Q) = 1$ into account. Hence, in the expression for C in Algorithm A5 (line -3, p. 166), we need to add one so that $C \leftarrow \text{sum}(\exp(X[j,]\beta[,-q,i])) + 1$ instead of $C \leftarrow \text{sum}(\exp(X[j,]\beta[,-q,i]))$. Note that, after this correction, in the explanatory note in Algorithm A5 under expression C (lines -1 and -2, p. 166), C in this case records the sum of the Q-1 terms (including the reference category), and not Q-2 terms, as incorrectly mentioned by Holmes and Held.

Correction 2: sampling the regression coefficients β_j

As shown by Holmes and Held, $\eta_{ij} = \frac{\exp(x_i\beta_j - C_{ij})}{1+\exp(x_i\beta_j - C_{ij})}$, which has the form of a binary logistic regression on class indicator $I(y_i = j)$. Therefore, by incorporating the correction term C_{ij}, we have that, following line 2 of equation (8) of Holmes and Held, for each observation i and for $j = 1, .., Q-1$,

$$z_{ij} = x_{ij}\beta_j - C_{ij} + \varepsilon_{ij}. \tag{E1}$$

Note that (E1) is not equivalent to line 2 of equation (8) of Holmes and Held due to the term $-C_{ij}$, and therefore we cannot directly apply the algorithm of the binary logistic regression. The first solution is to add the value of C_{ij} to the draws of z_{ij}. In this case, z_{ij} again equals line 2 of equation (8) of Holmes and Held, and hence we may apply the algorithm of the binary logistic regression. Therefore, in this solution we need to add $\log C$ to the draws of $Z[j,q]$ (line 2 p. 167) such that $Z[j,q] \sim \text{Lo}(m - \log C, 1)\,\text{Ind}(Y[j,q], Z[j,q]) + \log C$ instead of
$Z[j,q] \sim \text{Lo}(m - \log C, 1)\,\text{Ind}(Y[j,q], Z[j,q])$ as presented in Holmes and Held.

An alternative solution is to keep the expression for Z as provided by Holmes and Held, but to change the computation of m (line -4, p. 166), and the computation of B (line -9, p. 166). First, in order to draw the individual specific variances λ_{ij} of z_{ij}, using the rejection sampling procedure as outlined in A4 of Holmes and Held, we need to set R to the difference between z_{ij} and its (corrected) mean m_{ij}. According to (E1), the corrected mean m_{ij} of z_{ij} equals $x_{ij}\beta_j - C_{ij}$, and hence $m \leftarrow X[j,]\beta[,q,i] - \log C$ instead of $m \leftarrow X[j,]\beta[,q,i]$ as indicated on page 166 (line -4) of Holmes and Held. In this case, in order to compute m, we first need to compute C. In addition, since we changed m, we need to change the computation of z_{ij} accordingly (line 2, p. 167) to $Z[j,q] \sim \text{Lo}(m, 1)\,\text{Ind}(Y[j,q], Z[j,q])$ instead of $Z[j,q] \sim \text{Lo}(m - \log C, 1)\,\text{Ind}(Y[j,q], Z[j,q])$. Second, when computing B_j in line 2 of equation (9), we need to take into account the additional term $-C_{ij}$ in equation (E1), and hence B_j now becomes

$$B_j = V(v_j^{-1} b_j + x_j^{'} W(z_j + C_j)). \tag{E2}$$

Accordingly, the computation of B on page 166 (line -9) should be
$B \leftarrow VX^T \Lambda\,[,,q]^{-1}\,(Z\,[,q] + \log C)$ instead of $B \leftarrow VX^T \Lambda\,[,,q]^{-1}\, Z\,[,,q]$ as indicated by Holmes and Held. Similar to the computation of m, we need to compute first C in order to compute B, and hence the computation of C should be executed at the beginning of the pseudo-code in A5, before line -9 on page 166.

References

Holmes, C. C. and Held, L. (2006). "Bayesian Auxiliary Variable Models for Binary and Multinomial Regression." *Bayesian Analysis*, 1(1): 145–168. 353

Acknowledgments

The author wishes to thank the editor and referees for valuable suggestions on a previous version of this note.