Albert and Chib (1993) introduced a data-augmentation idea for Bayesian inference in regression models for binary and polychotomous response data. This paper is a seminal work on introduction of latent data into Bayesian analysis of categorical response data.

By using continuous latent data response to connect the binary response in a probit model with the normal linear model, this approach has several advantages. First, it allows us to perform exact inference for binary regression models, which is likely preferable to maximum likelihood methods for small sample sizes. Second, sampling from the joint posterior of unknowns in the model then only require Gibbs sampling, which allows us to directly draw standard distributions such as the multivariate normal and thus is easy to implement. To sum up, this data-augmentation scheme leads to simple, effective methods for Bayesian posterior inference, which circumvent the need for analytic approximations, numerical integration or Metropolis-Hastings in probit models.

Moreover, we can easily extend this approach to model the latent data beyond the probit link. For example, our faculty members Dr. Ming-Hui Chen and Dr. Dipak K. Dey used this kind of data-augmentation idea in their paper (2008) titled "Flexible Generalized T-Link Models for Binary Response Data". I also employed this idea in my paper (2013) titled "Bayesian Analysis of Dynamic Item Response Models in Educational Testing" for logistic regression.

However, in comparison to the probit model, Bayesian inference for the logistic regression model has long been recognized as a hard problem for the inconvenient analytic form of the likelihood function of the model. A paper "Bayesian inference for logistic models using Polya-Gamma latent variables" written by Polson, Scott and Windle (2013) introduced a Pólya-Gamma latent variable as a new data-augmentation scheme for binomial likelihoods. Their approach avoided the Metropolis-Hasting steps for logistic regression as shown in my paper (2013), which made their method very useful, including for logistic regression, negative binomial regression, nonlinear mixed-effect models, and spatial models of count data.

The article of Albert and Chib (1993) is one of the papers I always encourage my graduate students to read when they begin to study Bayesian methods. It reflects a very important idea for designing an effective Markov chain Monte Carlo algorithm in the Bayesian analysis.

References:

1. Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. Journal of the American statistical Association, 88(422), 669-679.

2. Kim, S., Chen, M. H., & Dey, D. K. (2007). Flexible generalized t-link models for binary response data. Biometrika, 95(1), 93-106.

3. Wang, X., Berger, J. O., & Burdick, D. S. (2013). Bayesian analysis of dynamic item response models in educational testing. The Annals of Applied Statistics, 7(1), 126-153.

4. Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. Journal of the American statistical Association, 108(504), 1339-1349.