

Bayesian inference of the cumulative logistic principal component regression models

Minjung Kyung^{1,a}

^aDepartment of Statistics, Duksung Women's University, Korea

Abstract

We propose a Bayesian approach to cumulative logistic regression model for the ordinal response based on the orthogonal principal components via singular value decomposition considering the multicollinearity among predictors. The advantage of the suggested method is considering dimension reduction and parameter estimation simultaneously. To evaluate the performance of the proposed model we conduct a simulation study with considering a high-dimensional and highly correlated explanatory matrix. Also, we fit the suggested method to a real data concerning sprout- and scab-damaged kernels of wheat and compare it to EM based proportional-odds logistic regression model. Compared to EM based methods, we argue that the proposed model works better for the highly correlated high-dimensional data with providing parameter estimates and provides good predictions.

Keywords: Bayesian inference, principal components regression, shrinkage priors

1. Introduction

Logistic regression has been considered as the standard approach for the analysis of ordered categorical outcome data in application area such as biomedical and social science studies. For the ordinal outcomes, the most common regression models fall under the set of cumulative logit link models proposed by Walker and Duncan (1967) because of a direct result of its simplicity and ease of interpretation. Later it is called the proportional odds model by McCullagh (1980) since it assumes that the relationship between the cumulative log-odds and the explanatory variables does not depend on the response category. The cumulative logit model guarantees that for C ordered groups, the cumulative probability of belonging to group c is less than or equal to the cumulative probability of belonging to group $c + 1$, for $c = 1, \dots, C - 1$ (McKinley *et al.*, 2015). Various inference processes have been proposed for the cumulative logit link models and comprehensive details about models and the analysis of ordered category data can be found in Agresti (2010).

Another commonly used form of modeling the cumulative logit model is based on an unobserved continuous latent variable. McCullagh and Nelder (1989) showed that the cumulative log-odds model could be derived from an underlying unobserved continuous random variable falling into an interval on the real line. A fully Bayesian method for modeling polychotomous ordinal categories based on the latent variable with probit link was developed in Albert and Chib (1993) by using the data augmentation approach. It might be easier to use a normal distribution with probit link instead of logistic

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. 2021R1F1A1049834).

¹ Department of Statistics, Duksung Women's University, 33 Samyang-ro, 144-gil, Dobong-gu, Seoul 01369, Korea.
E-mail: mkyung@duksung.ac.kr

distribution. For example, Yi *et al.*(2004) proposed a Bayesian approach via MCMC algorithms to extend the composite model space approach to detect genomewide multiple interaction quantitative trait loci (QTL) for ordinal traits on the basis of a threshold model. A Bayesian method for the cumulative logit link model has been proposed by Lang (1999) such that because of the complexity of the likelihood function, a random-walk Metropolis-Hastings (MH) algorithm with a multivariate-Gaussian proposal distribution is used to estimate the posterior distribution of parameter. Holmes and Held (2006) suggested Gibbs samplers based on the full posterior distribution for logistic multinomial regression models based on auxiliary variable approach, and O'Brien and Dunson (2004) developed a multivariate logistic regression framework that provides a marginal logistic structure for each of the outcomes via a multivariate t -distribution approximation to the multivariate logistic distribution.

For the mean trend in regression, the linear model is commonly used in various field. However, with many explanatory variables, there might be chance of the inference problem of high-dimensional covariates. Principal component (PC) regression has been known as a tool for data analysis and dimension reduction in applications throughout science and engineering. The standard principal component regression can be represented using the reduced singular-value decomposition (SVD) of a matrix of explanatory variables. Then with orthogonal linear combination of predictors, PC, from SVD, the p -predictor linear regression transforms to the k factor linear regression considering subset of PC with high variance. With relevant subset of PC, PC regression model can be used for the prediction of responses. Because only high variance PC corresponding to the higher eigenvalues is considered in the model, the low variance components are discarded in the model, thus for the purpose of the prediction, other shrinkage method such as the ridge regression estimator is perhaps a better choice compared to the PC regression which has a discrete shrinkage effect (Frank and Friedman, 1993). Even though there are many controversies for the interpretation and usage of PC regression, it is still a useful tool when there exists a multicollinearity problem among explanatory variables in the regression models.

The standard PC regression performs principal component analysis (PCA) on the centered explanatory variables matrix, and selects subset of the PC based on some appropriate criteria. Then, the observed vector of responses is regressed linearly on the selected PC as covariates, and the regression coefficients are estimated via ordinary least square method. The estimated PC regression coefficients are transformed back to the parameters of the actual predictors in the original model using the selected PCA loadings. In these steps, the principal components are obtained from the SVD of the explanatory variables matrix, thus the PC regression estimator obtained from using these principal components may produce unsatisfactory results for the prediction of the responses. Bair *et al.*(2006) suggested a technique called supervised principal components which is similar to conventional PCA except that it uses a subset of the predictors selected based on their association with the outcome. To choose the reduced rank of data, a threshold is considered and cross-validation of the log-likelihood ratio test was considered to estimate the best value of threshold. In Bayesian framework, West (2003) discussed the standard principal component regression with shrinkage prior on the PC regression coefficients allowing for varying degrees of shrinkage in each of the orthogonal PC dimensions.

In this paper, we propose a Bayesian approach for the cumulative log-odds PC regression model to deal with ordinal response based on an approximation of a t -distribution to a logistic distribution. The purpose that we propose a Bayesian cumulative logit PC regression model is to make MCMC implementation easier and to consider multicollinearity problem with interpreting parameters as the cumulative log-odds model. The close correspondence between the logistic and t distribution was noted by Albert and Chib (1993) such that for probabilities between 0.001 and 0.999, logistic quantiles are approximately a linear function of $t(8)$ quantiles. O'Brien and Dunson (2004) instead considered a multivariate t -distribution with $\sigma^2 = \pi^2(\nu - 2)/3\nu$, a value chosen to make the variance of the

univariate t and logistic distribution equal, and $\nu = 7.3$, a value chosen to minimize the integrated squared distance between the univariate t and univariate logistic densities. The reason for considering a t -distribution approximation to the logistic is that if we take back transformation of logistic link, it will facilitate interpretability by explaining the parameter of interest in the form of odd ratios. Even with an approximation, it becomes possible to interpret the effects of predictors to the probabilities of each category directly. In contrast, with a probit link, coefficients for probit models can be interpreted as the difference in standard normal score associated with each one-unit difference in a predictor variable, which is not intuitive to interpret its effect. Thus we prefer to use a logistic model based on the t -distribution approximation. Also, for the mean, we consider PC regression with shrinkage priors of West (2003) on the regression coefficients, and we choose the number of subset of PC based on the Bayesian information criterion (BIC). Usually, BIC seems to be a popular criterion for model selection because of a simple form and an easy application. Raftery (1995) argued that Bayes factors allow the direct comparison of non-nested models, in a simple way and BIC provides a simple and accurate approximation to Bayes factors for the Bayesian approach to hypothesis testing, model selection and accounting for model uncertainty. In the frame of Bayesian inference, it is easy to obtain BIC using posterior mean from Gibbs samples of posterior distribution of parameters. Thus we consider that for our Bayesian inference of the cumulative log-odds model with PC regression, BIC might be a good model selection criterion to choose the number of subset of PC within regression model framework.

In the remainders of the paper, we present the details about the cumulative log-odds model with the PC based regression and subsequently describe about the Bayesian inference with conjugate priors. We then apply the proposed method to real data concerning sprout- and scab-damaged kernels of wheat. We finally discuss about conclusion.

2. Cumulative logistic principal component regression model

For the Bayesian analysis of ordered categorical outcomes, we express the ordinal logistic regression model via underlying continuous variables. Then with PC regression, we extend to the cumulative log-odds PC regression model.

2.1. Cumulative logistic model

For the polychotomous ordinal outcomes of C categories, we usually assume that

$$Y_i \sim \text{Multinomial}(1, (p_1, \dots, p_C)).$$

The correspondence between the observed outcome and the latent variable is defined by

$$P(Y_i = j | \mathbf{X}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2) = \int \sum_{j=1}^C \mathbf{I}(\gamma_{j-1} \leq z_i \leq \gamma_j) \mathcal{L}(z_i | \mathbf{X}_i \boldsymbol{\beta}, \sigma^2) dz_i,$$

where \mathbf{X}_i is the i^{th} observed predictors of length p , $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression parameter, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_C)'$ is a vector of cutpoints with $-\infty = \gamma_0 < \gamma_1 < \dots < \gamma_{C-1} < \gamma_C = \infty$, and $\mathcal{L}(\mu, \sigma^2)$ is a density function of logistic distribution with mean μ and scale parameter σ^2 . With given notation and $\sigma^2 = 1$, the cumulative logit model for the j^{th} ordinal outcome can be expressed as

$$\text{logit } P(Y_i \leq j | \mathbf{X}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \gamma_j - \mathbf{X}_i \boldsymbol{\beta}, \quad (2.1)$$

for $j = 1, 2, \dots, C - 1$. Here, the logit link is monotone increasing, and for the ease of notation, we let $P(Y_i \leq j | \mathbf{X}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}) \equiv \eta_{ij}$, then the probabilities of each category are computed by

$$P(Y_i = j | \mathbf{X}_i, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma^2) = \begin{cases} \eta_{ij} & \text{for } j = 1 \\ \eta_{ij} - \eta_{i(j-1)} & \text{for } j = 2, \dots, C - 1 \\ 1 - \eta_{i(C-1)} & \text{for } j = C \end{cases}$$

In (2.1), the effect of the predictors is independent of the category involved, and subject to the constraint $-\infty = \gamma_0 < \gamma_1 < \dots < \gamma_{C-1} < \gamma_C = \infty$, the cumulative probability will be such that $0 < \eta_{i1} < \eta_{i2} < \dots < \eta_{i(C-1)} < 1$, *strict stochastic ordering* (McCullagh, 1980).

As we discussed in Section 1, the cumulative logistic model is known as the proportional odds model (McCullagh, 1980). It is because the cumulative log-odds ratio for two sets of predictors, \mathbf{X}_1 and \mathbf{X}_2 is given by

$$\begin{aligned} & \text{logit } P(Y_i \leq j | \mathbf{X}_1, \boldsymbol{\beta}, \boldsymbol{\gamma}) - \text{logit } P(Y_i \leq j | \mathbf{X}_2, \boldsymbol{\beta}, \boldsymbol{\gamma}) \\ &= \gamma_j - \mathbf{X}_1 \boldsymbol{\beta} - (\gamma_j - \mathbf{X}_2 \boldsymbol{\beta}) = (\mathbf{X}_2 - \mathbf{X}_1) \boldsymbol{\beta}. \end{aligned}$$

Hence, the ratio of corresponding odds is independent of category j and depends only on the difference between the covariate values $\mathbf{X}_2 - \mathbf{X}_1$.

2.2. Principal component regression model

In statistics, principal component regression (PCR) is a regression analysis technique that is based on principal component analysis. Massy (1965) introduced PCR to simplify the sample space of explanatory variables by a transformation to PC, for the cases such that the original independent variables are highly collinear with one another or there are a great number of potential explanatory variables. PCR regresses on a small number of regressors accounting for most of the variability of the explanatory variables. Massy (1965) also proposed a method choosing PCs which are most highly correlated with the response to be include in the regression and thus to explain the most variation therein.

Reiss and Ogden (2007) extended PCR to a functional version of PCR with splines, and introduced the reduced singular-value decomposition (SVD) of a matrix of explanatory variables to form a new design matrix of PCR. We consider the SVD of $n \times p$ explanatory matrix \mathbf{X} , $\mathbf{X} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^T$. Here, \mathbf{U} is a $n \times n$ orthonormal matrix and the column vectors are the left singular vectors of \mathbf{X} which can be obtained by the spectral decomposition of $\mathbf{X} \mathbf{X}^T$. \mathbf{V} is a $p \times p$ orthonormal matrix and the column vectors are the right singular vector of \mathbf{X} from the spectral decomposition of $\mathbf{X}^T \mathbf{X}$. Also $\boldsymbol{\Sigma}$ is a $n \times p$ diagonal matrix of singular values (square root of eigenvalues) of $\mathbf{X} \mathbf{X}^T$ or $\mathbf{X}^T \mathbf{X}$ in descending order.

Let \mathbf{U}_K , $\boldsymbol{\Sigma}_K$, and \mathbf{V}_K be the selected K versions of the three SVD matrices, then we assume that

$$\mathbf{X} \approx \mathbf{U}_K \boldsymbol{\Sigma}_K \mathbf{V}_K^T \equiv \mathbf{F} \mathbf{A},$$

where $\mathbf{F} = \mathbf{U}_K \boldsymbol{\Sigma}_K$ is a $n \times K$ factor matrix and $\mathbf{A} = \mathbf{V}_K^T$ is a $K \times p$ SVD loading matrix with $K < \min(n, p)$. Here, $\mathbf{A} \mathbf{A}^T = \mathbf{I}$, $\mathbf{F}^T \mathbf{F} = \boldsymbol{\Sigma}_K^2$ and $\boldsymbol{\Sigma}_K$ is the $K \times K$ upper left submatrix of $\boldsymbol{\Sigma}$ which we assume to be nonsingular. For the linear model, we usually assume that $\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon}$ where \mathbf{y} is the n -vector of responses and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ is the n -vector error term. Then K component PCR can be expressed as such that

$$\mathbf{y} = \mathbf{F} \boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad (2.2)$$

where $\epsilon \sim \text{MVN}_n(\mathbf{0}, \sigma^2 \mathbf{I})$. Here, the regression transforms via $\mathbf{X}\beta = \mathbf{F}\theta$ where $\theta = \mathbf{A}\beta$ is the length K vector of regression parameters for the PCR. The original regression parameter β can be obtained by the least-norm inverse $\beta = \mathbf{A}^T \theta$.

2.3. Cumulative logistic principal component regression model

A useful feature of the logistic distribution is that the logistic density and the t -distribution density closely approximate one other when the degrees of freedom ν and the variance ξ^2 are chosen appropriately (O'Brien and Dunson, 2004). This fact allows us that an approximation to the posterior distribution can be derived by considering an appropriate t -distribution for the logistic distribution. Albert and Chib (1993) noted that logistic quantiles are approximately a linear function of $t(8)$ quantiles. O'Brien and Dunson (2004) considered a multivariate t -distribution with $\xi^2 = \pi^2(\nu - 2)/3\nu$, a value chosen to make the variance of the univariate t and logistic distribution equal, and $\nu = 7.3$, a value chosen to minimize the integrated squared distance between the univariate t and univariate logistic densities. Thus with properties of cumulative logistic regression that are discussed in Section 2.1, if we combine the t -approximation to the logistic regression and the PCR to consider the multicollinearity among \mathbf{X} which is discussed in section 2.2, the cumulative logistic PC regression model with latent variable Z_i can be expressed that for $i = 1, \dots, n$ and $j = 1, \dots, C$,

$$\begin{aligned} Y_i &= j \quad \text{if } \gamma_{j-1} \leq Z_i \leq \gamma_j \\ Z_i | \mathbf{F}_i, \boldsymbol{\theta}, \phi_i, a_0 &\sim N(a_0 + \mathbf{F}_i \boldsymbol{\theta}, \psi \phi_i^{-1} \sigma^2) \\ \phi_i | Z_i, \mathbf{F}_i, \boldsymbol{\theta}, a_0 &\sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \end{aligned} \tag{2.3}$$

where $\nu = 7.3$ and $\psi = \pi^2(\nu - 2)/3\nu$, a value chosen to make the variance of the univariate t and logistic distribution equal in our model.

For the t -distribution with location μ , scale ξ and degrees of freedom ν , the density function can be expressed in the following scale mixtures of normal form,

$$t_\nu(y | \mu, \xi) = \int_0^\infty N\left(y \mid \mu, \frac{\xi^2}{\eta}\right) \text{Gamma}\left(\eta \mid \frac{\nu}{2}, \frac{\nu}{2}\right) d\eta,$$

where $\text{Gamma}(\cdot | a, b)$ is the gamma density function. It means that the t -distribution can be expressed as the following hierarchical form such that

$$Y | \mu, \xi^2, \eta, \nu \sim N\left(\mu, \frac{\xi^2}{\eta}\right) \quad \text{and} \quad \eta | \nu \sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right).$$

This representation is considered for the model in (2.3), and based on the model in (2.3), we consider the Bayesian methods for the inference of parameters, and the details are in the following section.

3. Bayesian inference on the cumulative logistic principal component regression model

In Bayesian framework, usually an unknown parameter is treated as a random variable, and uncertainty about the parameters is inferred by its probability distribution, the so-called posterior distribution. The posterior distribution is a combination of evidence collected from data that is formally expressed by the likelihood function and a prior belief that is specified with a prior distribution. To

update a posterior distribution for the interest parameter, a Markov chain Monte Carlo (MCMC) algorithm is used, through which it enables to iteratively draw sample from even a complex distribution without analytically solving for the point estimate.

3.1. Specifications of prior distribution

In our work, we consider shrinkage priors on the underlying factor regression parameters θ (West, 2003). West (2003) discussed that a particularly flexible class of such priors assumed independent t -distributions for the elements of θ , so allowing for varying degrees of shrinkage in each of the orthogonal factor dimensions. The shrinkage priors on the PC regression parameters are the following: for $k = 1, \dots, K$,

$$\theta_k \sim N\left(0, \frac{\rho}{k^2 \delta_k}\right) \quad \text{and} \quad \delta_k \sim \text{Gamma}\left(\frac{r}{2}, \frac{r}{2}\right), \quad (3.1)$$

independently, for some $r > 0$. The tuning parameter r is the degree-of-freedom parameter for the implied t -distribution for θ_k that follows on marginalization over the random precision δ_k . Other prior distributions on the parameters of the model in (2.3) and prior distribution on the hyper-parameter ρ are the following,

$$\begin{aligned} \rho &\sim \pi(\rho) \propto \text{constant}, \quad \sigma^2 \sim \text{IG}(a, b), \quad r \sim U(0, 10), \\ a_0 &\sim N(\alpha_0, \eta^2), \quad \text{and} \quad \gamma_j \sim \pi(\gamma_j) \propto \text{I}(\gamma_{j-1} < \gamma_j < \gamma_{j+1}), \end{aligned} \quad (3.2)$$

where $j = 1, \dots, C-1$ and IG is the inverse gamma distribution. Here, the normal distribution of a_0 is indexed by two hyperparameters, which means we can specify a particular prior distribution by fixing two features of the distribution. However, without loss of generality, we usually set $\alpha_0 = 0$ and η^2 with large values such as 100.

3.2. Sampling scheme

With model in (2.3) and priors in (3.1) and (3.2), the Gibbs sampling is straightforward because the most of conditional posterior distributions are in closed form except for the tuning parameter r . The conditional posterior distribution of r is no closed form, but the sampling from the conditional posterior distribution can be processed by a Metropolis-Hastings algorithm. Thus, we iterate sampling from their full conditional posterior distributions until convergence.

With given the number of PC, K , the joint posterior distribution has the form of

$$\begin{aligned} &\pi(\boldsymbol{\theta}, \boldsymbol{\phi}, a_0, \sigma^2, \boldsymbol{\delta}, r, \rho, \boldsymbol{\gamma}, \mathbf{Z} | \mathbf{X}, \mathbf{Y}) \\ &\propto \left[\prod_{i=1}^n \left\{ \prod_{j=1}^C \text{I}(\gamma_{j-1} \leq Z_i \leq \gamma_j) \right\} \right] \left\{ (\phi^{-1} \sigma^2)^{\frac{1}{2}} \exp \left\{ -\frac{\phi_i}{2\psi\sigma^2} (Z_i - a_0 - \mathbf{F}_i \boldsymbol{\theta})^T (Z_i - a_0 - \mathbf{F}_i \boldsymbol{\theta}) \right\} \right\} \\ &\times \phi_i^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu}{2}\phi_i\right) \left[(\sigma^2)^{-(a+1)} \exp\left(-\frac{b}{\sigma^2}\right) \exp\left\{-\frac{1}{2\eta^2}(a_0 - \alpha_0)^2\right\} \prod_{j=1}^{C-1} \text{I}(\gamma_{j-1} < \gamma_j < \gamma_{j+1}) \right] \\ &\times \rho^{-\frac{K}{2}} \left(\prod_{k=1}^K \delta_k^{\frac{1}{2}} \right) \exp\left(-\frac{1}{2}\boldsymbol{\theta}^T \mathbf{G}^{-1} \boldsymbol{\theta}\right) \frac{\left(\frac{r}{2}\right)^{\frac{r}{2}K}}{\left\{\Gamma\left(\frac{r}{2}\right)\right\}^K} \left(\prod_{k=1}^K \delta_k^{\frac{r}{2}-1} \right) \exp\left(-\frac{r}{2} \sum_{k=1}^K \delta_k\right), \end{aligned}$$

where $\mathbf{G} = \text{diag}(\rho/k^2\delta_k)_{k=1,\dots,K}$ and $\boldsymbol{\theta}^T \mathbf{G}^{-1} \boldsymbol{\theta} = 1/\rho \sum_{k=1}^K k^2 \delta_k \theta_k^2$. Then, the posterior sampling step is carried out using the Gibbs sampling with priors as discussed earlier. For the tuning parameter r , we consider a Metropolis-Hastings algorithm because the conditional distribution is no closed form. But for other parameters, we use the Gibbs because of the conjugacy. Further details on the sampling scheme with the full conditional distributions are provided in the Appendix A.

There is an identifiability problem in intercept a_0 , cutpoints $\boldsymbol{\gamma}$, and variance σ^2 . Hirk *et al.*(2019) pointed out that in ordinal models absolute location and absolute scale of the underlying latent variable are not identifiable. It means that in our model, only the quantities $\boldsymbol{\theta}/\sigma^*$ and $(\gamma_j - a_0)/\sigma^*$, $j = 1, \dots, C - 1$ are identifiable. Here, σ^{*2} is the full variance of the error term. Thus, to secure the identifiability of scaled intercepts a_0/σ^* and scaled cutpoints γ_j/σ^* , we consider a typical constraints on the parameters such that $\gamma_1 = 0$ and $\sigma^2 = 1$ in our model. It is a common practice in Bayesian logistic regression models for both univariate and multivariate binary outcomes. The fixing of the cutpoints at zero in the process of MCMC is to prevent the identifiability problem of the parameter estimation. Thus, with these constraints, we update the posterior samples of the parameters iteratively.

As we discussed in Section 1, we choose the number of subset of PC, K , based on the Bayesian information criterion (BIC). In Bayesian analysis, Bayes factors is a tool of model selection. To compute the Bayes factor, calculating the integrated likelihood over the parameters space is involved and it can be a high-dimensional and intractable integral. Exact analytic integration might be possible for exponential family distributions with conjugate priors including normal linear models. However, Kass and Raftery (1995) argued that it is more often intractable and thus must be computed by numerical methods, and introduced various asymptotic approximation to the marginal density of the data. BIC (The Schwarz criterion) was one of the asymptotic approximation tools which Kass and Raftery (1995) introduced. Raftery (1995) argued that Bayes factors allow the direct comparison of non-nested models in a simple way and BIC provides a simple and accurate approximation to Bayes factors for the Bayesian approach to hypothesis testing, model selection and accounting for model uncertainty. BIC can be applied as a standard procedure even when the priors of the parameters are hard to set precisely. Another advantage of BIC is that in the frame of Bayesian inference, it is easy to obtain BIC using posterior mean from Gibbs samples of posterior distribution. Thus we consider that for our Bayesian inference of the cumulative log-odds model with PC regression, BIC might be a good model selection criterion to choose the number of subset of PC within regression model framework.

4. Illustrative examples

For the simulation study and data analysis, to analyze the data in the lack of the information of prior, diffuse prior distributions were considered. For this, we set the hyperparameters of prior distribution of intercept, the mean and variance, of a_0 as $\alpha_0 = 0$ and $\eta^2 = 100$. For the posterior inference, we obtained the posterior median and the 95% credible intervals of the parameter estimates from the proposed MCMC algorithm ran for 30,000 iterations with a burn-in period of 15,000 iterations and a thinning by taking every fifth posterior sample. Also, to fit the Bayesian cumulative logistic PC regression on data, we need to decide the optimal number of components. As discussed in Section 3.2, we compute the BIC of the Bayesian cumulative logistic PC regression for the number of components from 2 to a large value L (L can be vary because of the data structure). Based on the BIC plot, we choose the number of components with the smallest BIC value.

For the comparison of the suggested model, we consider the EM based inference with linear mean trend using *polr* (proportional-odds logistic regression) function in MASS package (Venables and Ripley, 2002) and *bayespolr* function in arm package (Gelman and Su, 2020) in R (R core team,

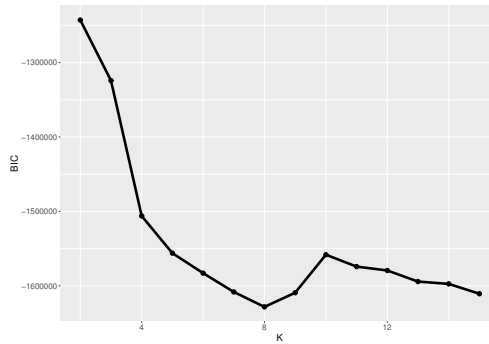


Figure 1: BIC plot of Bayesian cumulative logistic PC regression for the simulated samples.

2021). The *polr* function is developed based on the traditional maximum likelihood method using Fisher scoring or the Newton-Raphson algorithm. The *bayespolr* function is developed based on the weakly informative priors for logistic regression (Gelman *et al.*, 2008, 2015) assuming a Student-*t* prior distribution with mean 0, degrees-of-freedom (df) and scale with df and scale chosen to provide minimal prior information to constrain the coefficients to lie in a reasonable range. The full Bayesian computation using the usual MCMC methods can be applied, but for a quick iterative calculation to get a point estimate of the regression coefficients and standard errors, an approximate EM algorithm is applied.

4.1. Simulation study

We conduct a simulation study to evaluate how well the proposed method performs with considering a high-dimensional and highly correlated explanatory matrix. For a high-dimensional predictors, we set the number of observations as $n = 50$ and the number of explanatory variables as $p = 70$. To generate highly correlated explanatory matrix, we set the correlation coefficients of X_i and X_j as $\sigma_{ij}^X = (0.8)^{|i-j|}$ and the standard deviation of X_i is 0.1. With given covariance matrix and mean zero vector, we generate $n \times p$ explanatory matrix from a multivariate normal distribution. For the true value of regression parameter, we set 35 ones and 35 zeros such that $\beta = (1, \dots, 1, 0, \dots, 0)$. Incorporating a regression model into the mean $X\beta$, an ordered categorical outcome y were determined based on the logistic distribution and the predefined cutpoints $(-\infty, -1.9, 0, 1.8, \infty)$.

To decide the number of components that are considered in the Bayesian cumulative logistic PC regression, we computed BIC for each number of components and the BIC plot up to 15 number of components is in Figure 1. In Figure 1, we observe that the minimum value of BIC is obtained when the number of principal components is 8, $k = 8$. Thus, for the remaining analysis and model comparison, we employ 8 principal components of explanatory matrix as regressors and after fitting the Bayesian cumulative logistic PC regression, we re-transfer it to the regression parameter as we discussed in Section 2.2. Also, to check up the convergence of MCMC, we draw trace plots and autocorrelation plots for few selected θ s and the plot is in Figure B.1 in Appendix B. The plots how that the proposed Gibbs sampler converged fast and mixed well.

Subset of estimated values and 95% confidence intervals of selected regression parameters in EM based proportional-odds logistic regression (*polr*) and in EM based Bayesian proportional-odds logistic regression (*bayespolr*), and estimated values (posterior median) and 95% credible intervals of regression parameters in Bayesian cumulative logistic PC regression (*bayesPC(K=3)*) for simulated data are in Table B.1 in Appendix B. Also, estimated values and 95% confidence intervals of re-

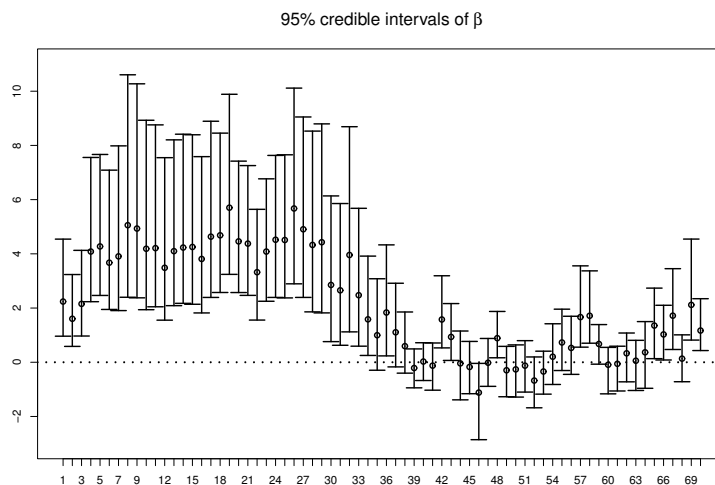


Figure 2: Estimated values and 95% credible intervals of re-transformed regression parameters in Bayesian cumulative logistic PC regression for the simulated samples.

transformed regression parameters of the suggested model can be found in Figure 2. Because the explanatory matrix is not full rank, EM based proportional-odds logistic regression model does not use full predictors, and it only estimated the regression parameter of 49 explanatory variables and cut-points. However, all the estimated regression parameters and cutpoints were not significant because of very largely computed variance of estimators. Thus, it is not enough to use this model for the prediction and analysis in this highly correlated high-dimensional data. Bayesian proportional-odds logistic regression shows almost same results of EM based proportional-odds logistic regression model. However, because of dimension reduction effects on PC regression and the considered prior is a generalized shrinkage prior, the propose model provides reasonable parameter estimates. Even though somewhat amplified estimates and largely valued variances were observed, the proposed model detects the true zero values of the last 35 regression parameters in the setting of highly correlated high-dimensional data. Also, the model fit based on the mean squared error could be computed and the computed MSE was 0.06 with given explanatory matrix X which is used to fit the model. Thus, we argue that the proposed model works better for the highly correlated high-dimensional data with providing parameter estimates and good prediction.

4.2. A real data example

4.2.1. Single-Kernel density of wheats

Background : Damage from sprout and scab kernels can be a critical problem in wheat production because preharvest sprouting causes harvest losses, reduced test weight, and loss of seed viability. Thus, kernel density measurement is important, and to facilitate this identification process, technological advances toward an objective grain inspection system have been developed and have resulted in the rapid characterization of single-kernel wheat properties. Martin *et al.*(1998) conducted a study to investigate the effects of sprout and scab damage on single-kernel density and explore the relationship between density and single-kernel weight, size, moisture, and hardness and measured on a sample of

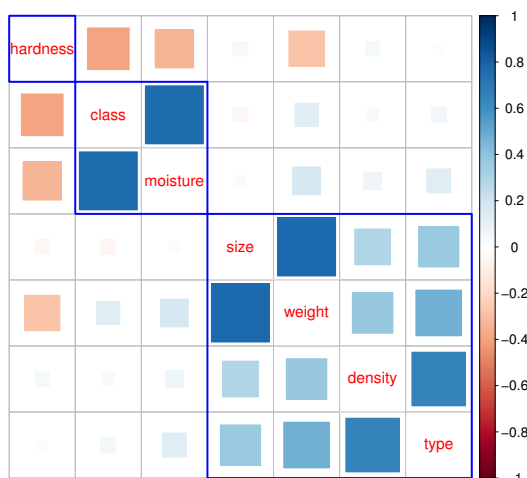


Figure 3: Spearman's rank correlation coefficient matrix of variables in wheat kernel data.

wheat kernels from two different classes of wheat, hard red winter (hrw) and soft red winter (srw). Sprout- and scab-damaged kernels were separated from non-damaged (healthy) kernel by the Federal Grain Inspection Service personnel. Fifteen kernels for each damage type and an equal number of healthy kernels were selected from four lots of hard red winter wheat and four lots of soft red winter wheat. Bilder and Loughin (2015) had been provided this wheat kernels data by Martin *et al.*(1998) and originally there were 275 wheat kernel measurements in the data set. Bilder and Loughin (2015) used a portion of the data to fit nominal and ordinal response regression models. We use here the same data set which Bilder and Loughin used to show how the suggested Bayesian cumulative logistic PC regression model works.

Results: In the wheat kernels data, we consider *type* (“scab” < “sprout” < “healthy”) as the response variable with orders. For the explanatory variables, we consider a binary *class* variable (hrw=0 & srw=1), *density* (single kernel density), *hardness* (endosperm texture), *size* (single kernel size), *weight* (single kernel weight), and *moisture* (single kernel moisture content). A graphical display of a correlation coefficients matrix of variables in wheat kernel data is in Figure 3. To compute the correlation coefficients among variables in wheat kernel data, the response variable *class* is expressed as numerical values such that “scab”=1, “sprout”=2 & “healthy”=3, and because of the order in response variable, we use a rank-based measure of association, Spearman's rho. Also, the results based on hierarchical clustering are shown as a box on the correlation plot in Figure 3. In Figure 3, we observe that *class* variable is highly correlated in positive direction with *moisture* and somewhat negatively correlated with *hardness*, but the relationship is not strong. Also, *moisture* is negatively correlated with *hardness*, but the relationship is not strong. As we expected *size* and *weight* is strongly correlated in positive direction. *density*, *weight* and *size* are correlated with numeritized response variable *type* in positive direction decreasingly.

Bilder and Loughin (2015) presented parallel coordinate plot of the wheat data. It is in Figure B.2 in Appendix. Three different line types, (solid = “healthy”, long dash=“sprout”, dot dash=“scab”) , have been used to understand some of the features that differentiate the healthy, sprout, and scab conditions. Y-axis of plot is the observed value of each variables of kernels. Also, to compare explanatory variables, , in each type of wheat, mean and 95% confidence intervals are computed in Table B.3 in Appendix. As Bilder and Loughin (2015) discussed, based on parallel coordinate plot and basic

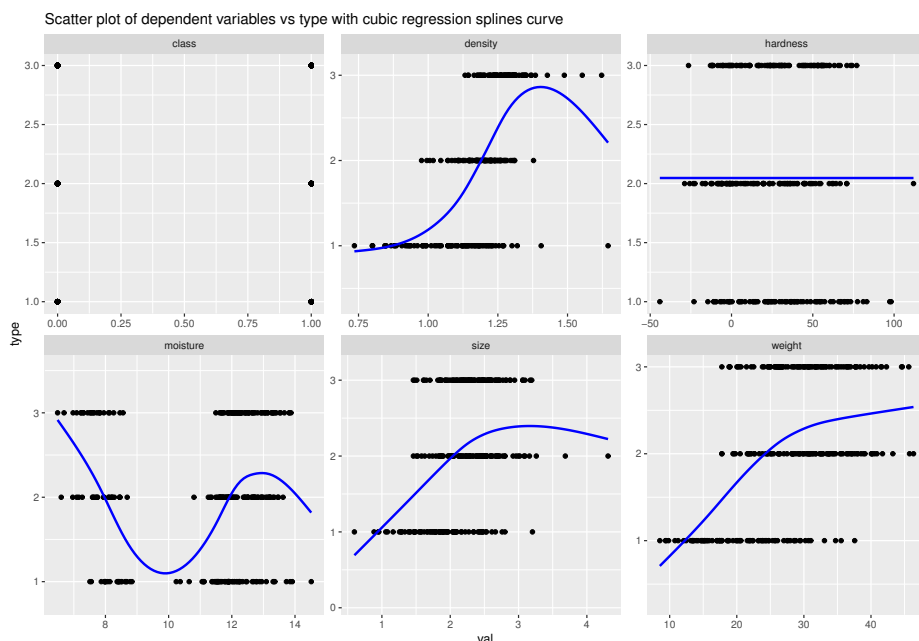


Figure 4: Scatter plot matrix of the wheat data.

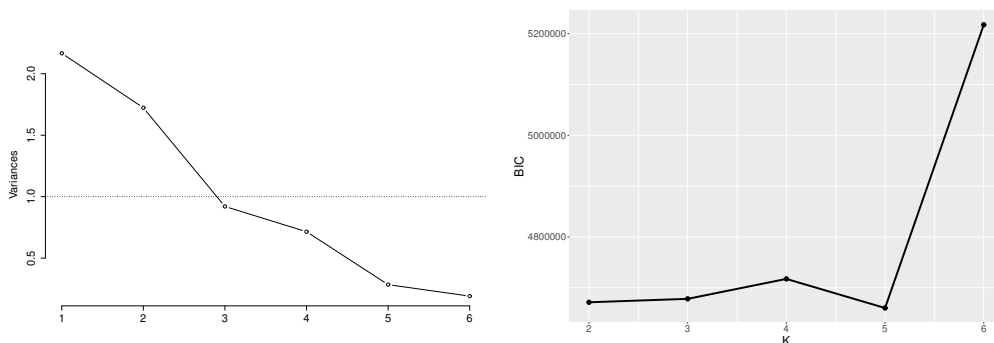


Figure 5: Scree plot of principal components on explanation variables matrix and BIC plot of Bayesian cumulative logistic PC regression.

statistics of each wheat type, we observe that (1) scab kernels generally have smaller density, size, and weight values, (2) healthy kernels may have higher densities, (3) there is much overlap for healthy and sprout kernels, and (4) the moisture content appears to be related to hard or soft red winter wheat class.

To explore the relationship between an explanatory variable and the response variable, we draw scatter plot of dependent variable vs. response variable (*type*) with cubic regression splines curve, and a scatter plot matrix is in Figure 4. We observe that as a wheat kernel is getting healthier, single kernel density is higher. A sprout and healthy kernel have similar *size* and *weight*, and a strong positive relationship exists between *size* and *weight* as we discussed above. The *hardness* and *moisture* do not seem to show any distinct properties for *type* of kernel.

Table 1: Estimated values and 95% confidence intervals of regression parameters in EM based proportional-odds logistic regression (polr) and in EM based Bayesian proportional-odds logistic regression (bayespqr), and estimated values (posterior median) and 95% credible intervals of regression parameters in Bayesian cumulative logistic PC regression (bayesPC(K=3))

	polr			bayespqr			bayesPC(K=3)		
	estimates	lb*	ub*	estimates	lb	ub	estimates	2.5%	97.5%
class	0.17	-0.13	0.51	0.18	-0.12	0.52	-0.00	-0.01	0.01
density	13.51	7.75	39.98	13.11	7.72	38.79	0.00	-0.00	0.01
hardness	0.01	0.01	0.03	0.01	0.01	0.03	0.01	0.00	0.02
size	-0.29	-0.63	-0.87	-0.29	-0.62	-0.86	0.01	0.00	0.01
weight	0.13	0.13	0.38	0.13	0.13	0.38	0.15	0.12	0.19
moisture	-0.04	-0.05	-0.12	-0.04	-0.06	-0.12	-0.02	-0.12	0.09
Scab Sprout	17.57	7.68	52.01	17.11	7.75	50.65	0.00	0.00	0.00
Sprout Healthy	20.04	9.32	59.33	19.55	9.42	57.88	1.84	1.52	2.20

* lb* and ub* are lower bound and upper bound of 95% confidence intervals.

To fit the Bayesian cumulative logistic PC regression on wheat kernel data, we need to decide the optimal number of components. What we had tried first is the principal components analysis on the explanatory variable matrix such that components with the decreasing order of variance are the corresponding eigenvector of the decreasing order of eigenvalues of the variance-covariance matrix of the centered and scaled explanatory variables. Secondly, as discussed in section 3.2, we computed BIC of the Bayesian cumulative logistic PC regression for the number of components from 2 to 6: 6 is the number of explanatory variables. Scree plot of PC analysis on the explanatory variable matrix and BIC plot of the Bayesian cumulative logistic PC regression are in Figure 5. We observe that only with explanatory matrix, 3 components ($K = 3$) is a good candidate for the optimal number of components. The computed BIC of 5 components is the smallest, but the BIC of 2, 3, and 5 components are quite similar. By the least-norm inverse $\beta = A^T \theta$, the original regression parameter β can be obtained and the estimated regression parameters $\hat{\beta} = A^T \hat{\theta}$ based on the Bayesian cumulative logistic PC regression with 5 component (bayesPC(K=5)) and with 3 component (bayesPC(K=3)) are in table B.2 in Appendix. As we discussed above, because of the identifiability, we fix the first cutpoint as $\gamma_1 = 0$ and the variance parameter in (2.3) as $\sigma^2 = 1$. With 5 components, *hardness* and *weight* seem to be significant, and with 3 components, *hardness*, *weight*, and *size* seem to be significant based on the 95% credible intervals. With fixed the first cutpoint, the posterior median of the second cutpoint is 1.84, and the 95% credible interval with 5 components model is slightly smaller. However, the number of explanatory variables is 6, and if we choose 5 components regression as a best model, it considers most part of variance of the model, thus we consider 3 components Bayesian cumulative logistic PC regression model for the comparison of other preexisted methods. To check the convergence of MCMC, we draw trace plots and autocorrelation plots for few selected θ s and the plot is in Figure B.3 in Appendix B. The plots show that the proposed Gibbs sampler converged fast and mixed well.

Estimated values and 95% confidence intervals of regression parameters from the EM based proportional-odds logistic regression (polr) and from the EM based Bayesian proportional-odds logistic regression (bayespqr) are in Table 1. Also the posterior median and the 95% credible interval of the least-norm inverse estimates of regression parameters based on the Bayesian cumulative logistic PC regression with $K = 3$ components are in Table 1. The values in bold in Table are statistically significant values. From the table, we observe that the estimated *density* parameters of polr and bayespqr have larger values than other estimated parameters, and the estimated first and second cutpoints are around 17~18 and around 19~20, respectively. However, the estimated *density* parameter of Bayesian

Table 2: Predictive MSE and MAPE of EM based proportional-odds logistic regression (polr), EM based Bayesian proportional-odds logistic regression (bayespolr), and the Bayesian cumulative logistic PC regression (bayesPC(K=3)) from 100 repeats

	polr		bayespolr		bayesPC(K=3)	
	mean	standard deviation	mean	standard deviation	mean	standard deviation
MSE	0.333	0.050	0.337	0.050	0.395	0.052
MAPE	0.213	0.036	0.214	0.035	0.243	0.039

cumulative logistic PC regression model is not significant based on the 95% credible interval and the estimated value is small. Also with fixed first cutpoint $\gamma_1 = 0$, the estimated second cutpoint is 1.84 and which is much smaller than the estimated values of polr and bayespolr. For the EM based methods, all parameter estimators are significant except a binary *class* variable, but for the suggested method, *hardness*, *weight*, and *size* are significant. These different properties might be because of the difference of inference method and the identifiability restriction of the Bayesian cumulative logistic PC regression model. Also, because of the shrinkage priors on the PC regression parameters θ , the estimates are small values as $(-0.09, 0.10, 0.07)$, then the estimates of regression parameter might have small values. In our proposed model, MCMC works with singular value decomposed explanatory matrix. The principal components with higher variances are selected regressors and these are only a subset of all the principal components for regression. Using a subset seem to make the principal components a kind of regularized procedure and also a type of shrinkage estimator. Therefore *density* and *size* parameter estimates in our proposed model shrink toward zero and *wieght* parameter estimates are similar to estimates of other methods because there exist high correlation among the *density*, *size* and *weight* as we checked in correlation plot in Figure 3.

4.2.2. Comparison of models

To compare the model fit, we compute the mean squared error (MSE) and the mean absolute percentage error (MAPE) of EM based models and the proposed model after fitting the models. MSE is the average squared difference between the estimated *type* based on the original explanatory matrix and the observed *type* and MSPE is the average of the absolute percentage errors. The estimated *type* is computed with the largest estimated category probability. The computed MSE of EM based proportional-odds logistic regression (polr) and of EM based Bayesian proportional-odds logistic regression (bayespolr) are 0.375 and 0.375, respectively. And the computed MSE of the Bayesian cumulative logistic PC regression with 3 components is 0.384. Also, the computed MAPE of EM based proportional-odds logistic regression (polr) and of EM based Bayesian proportional-odds logistic regression (bayespolr) are 0.240 and 0.242, respectively. And the computed MAPE of the Bayesian cumulative logistic PC regression with 3 components is 0.248 (Table of MSE and MAPE of data fit is omitted). As we discussed above, the proposed model uses principal components of explanatory matrix with higher variance as regressors and the principal components regression conducts a type of shrinkage estimator. Thus, computed MSE and MAPE of the Bayesian cumulative logistic PC regression are slightly largely valued compared to other EM based methods.

Considering prediction, we also compute the MSE and MAPE of predicted values such that for $n = 275$ observations, about 70% data ($n.train = 193$) is used to fit the models and about 30% ($n.test = 82$) observation is used to compute the predictive MSE and MAPE. We repeat the process 100 times and random training and testing data has been created in each iteration. The mean of predictive MSE and MAPE with standard deviations of EM based proportional-odds logistic regression (polr), EM based

Bayesian proportional-odds logistic regression (bayespolr), and the Bayesian cumulative logistic PC regression (bayesPC(K=3)) based on 100 repeats are in Table 2. The mean of predictive MSE for EM based methods are 0.333(0.050) and 0.337(0.050) respectively, and the computed values are quite similar, but the mean of predictive MSE of the suggested model is 0.395(0.052) and it is slightly larger than the values from EM based methods. Also, the computed MSPE of polr and bayespolr and the suggested model are 0.213(0.036), 0.214(0.035), and 0.243(0.039), respectively. We argue that the MSE and MAPE of the suggested model is slightly larger than the EM based methods, because the suggested model use part of X space for the prediction but the other methods use all of X . For the 100 repeats, the number of components of the Bayesian cumulative logistic PC regression are decided based on BIC, and the barplot of the number of components are in Figure B.4 in Appendix B. The most frequent number of component is $K = 3$ that is the number what we considered in whole data analysis.

5. Discussion

A Bayesian approach to cumulative logistic regression model with orthogonal principal components based on singular value decomposition was proposed to consider the multicollinearity among predictors. The suggested method has advantage such as considering dimension reduction and parameter estimation simultaneously. However, if the aim of the analysis is the prediction of the response variable, the suggested model has limitation, because the suggested method uses components with largely valued variance and ignores components with small valued variance. Also, the considered prior is a generalized shrinkage prior and West (2003) discussed the standard principal component regression with shrinkage prior on the PC regression coefficients allowing for varying degrees of shrinkage in each of the orthogonal PC dimensions. Thus, the suggested model might not be good to be considered for the prediction. However, if the multicollinearity problem is serious among explanatory variables, there might be the inference problem of high-dimensional covariates.

We conducted a simulation study to show how well the proposed method performs with considering a high-dimensional and highly correlated predictors. Because the predictor matrix is not full rank, compared to EM based proportional-odds logistic regression (polr) and to EM based Bayesian proportional-odds logistic regression (bayespolr), the proposed Bayesian cumulative logistic PC regression works better for the highly correlated high-dimensional data with providing parameter estimates and good prediction. Also, we applied the proposed method to real data concerning sprout- and scab-damaged kernels of wheat. Because of the shrinkage priors on the components, the estimated regression parameters were small valued estimates. Compared to EM based proportional-odds logistic regression (polr) and to EM based Bayesian proportional-odds logistic regression (bayespolr), the proposed Bayesian cumulative logistic PC regression showed limitation on the prediction. Also, the estimated parameters of *density*, and the estimates of the first and second cutpoints of the suggested method were very different from other methods numerically. We argue here that the scale of *density* variable much larger than other explanatory variables and it affects the position of cutpoints in somewhat ways.

Methodologically, the proposed method can be further extended to accommodate latent factor regression models for partitioning variation in the predictor variables into multiple component that reflect common patterns and separating these from noise variation. Although the suggested model considered the orthogonal components, there was a limitation on prediction. With latent factors, there might not be the orthogonal property, but we might be able to find out common structure in explanatory matrix isolating idiosyncratic variation for dimension reduction in predictor space.

Acknowledgement

The authors would like to thank the Editor-in-Chief, the Associate Editor, and anonymous reviewers for their thoughtful and constructive comments on prior versions of this manuscript.

Appendix A: Posterior distribution

The sampling scheme is as follows. With $\nu = 7.3$ and $\psi = \pi^2(\nu - 2)/3\nu$

- update \mathbf{Z} : for $i = 1, \dots, n$ and $j = 1, \dots, C$

$$Z_i | \theta, \phi_i, a_0, \sigma^2, \delta, r, \rho, \gamma, \mathbf{X}, \mathbf{Y} \sim N(a_0 + \mathbf{F}_i \boldsymbol{\theta}, \psi \phi_i^{-1} \sigma^2),$$

with Z_i is truncated at the left and right by γ_{j-1} and γ_j if $Y_i = j$.

- update γ : for $j = 1, \dots, C - 1$

$$\begin{aligned} & \gamma_j | \gamma_{-j}, \boldsymbol{\theta}, \boldsymbol{\phi}, a_0, \sigma^2, \delta, r, \rho, \mathbf{Z} | \mathbf{X}, \mathbf{Y} \\ & \sim U \left(\max \{ \max (Z_i : Y_i = j), \gamma_{j-1} \}, \min \{ \min (Z_i : Y_i = j + 1), \gamma_{j+1} \} \right) \mathbf{I}(\gamma_{j-1} < \gamma_j < \gamma_{j+1}), \end{aligned}$$

where $\boldsymbol{\gamma}_{-j} = (\gamma_1, \dots, \gamma_{j-1}, \gamma_{j+1}, \dots, \gamma_C)$.

- update $\boldsymbol{\theta}$

$$\begin{aligned} & \boldsymbol{\theta} | \boldsymbol{\phi}, a_0, \sigma^2, \delta, r, \rho, \gamma, \mathbf{Z}, \mathbf{X}, \mathbf{Y} \\ & \sim \text{MVN}_K \left(\left(\frac{1}{\psi \sigma^2} \sum_{i=1}^n \phi_i \mathbf{F}_i^T \mathbf{F}_i + \mathbf{G}^{-1} \right)^{-1} \left(\frac{1}{\psi \sigma^2} \sum_{i=1}^n \phi_i \mathbf{F}_i^T (Z_i - a_0) \right), \left(\frac{1}{\psi \sigma^2} \sum_{i=1}^n \phi_i \mathbf{F}_i^T \mathbf{F}_i + \mathbf{G}^{-1} \right)^{-1} \right), \end{aligned}$$

where MVN_K is the K -dimensional multivariate normal distribution.

- update $\boldsymbol{\phi}$: for $i = 1, \dots, n$

$$\phi_i | \boldsymbol{\theta}, a_0, \sigma^2, \delta, r, \rho, \gamma, \mathbf{Z}, \mathbf{X}, \mathbf{Y} \sim \text{Gamma} \left(\frac{\nu}{2} + \frac{1}{2}, \frac{\nu}{2} + \frac{1}{2\psi \sigma^2} (Z_i - a_0 - \mathbf{F}_i \boldsymbol{\theta})^T (Z_i - a_0 - \mathbf{F}_i \boldsymbol{\theta}) \right),$$

- update a_0

$$\begin{aligned} & a_0 | \boldsymbol{\theta}, \boldsymbol{\phi}, \sigma^2, \delta, r, \rho, \gamma, \mathbf{Z}, \mathbf{X}, \mathbf{Y} \\ & \sim N \left(\left(\frac{1}{\psi \sigma^2} \sum_{i=1}^n \phi_i + \frac{1}{\eta^2} \right)^{-1} \left(\frac{1}{\psi \sigma^2} \sum_{i=1}^n \phi_i (Z_i - \mathbf{F}_i \boldsymbol{\theta}) + \frac{\alpha_0}{\eta^2} \right), \left(\frac{1}{\psi \sigma^2} \sum_{i=1}^n \phi_i + \frac{1}{\eta^2} \right)^{-1} \right), \end{aligned}$$

- update σ^2

$$\sigma^2 | \boldsymbol{\theta}, \boldsymbol{\phi}, a_0, \delta, r, \rho, \gamma, \mathbf{Z}, \mathbf{X}, \mathbf{Y} \sim \text{IG} \left(a + \frac{n}{2}, b + \sum_{i=1}^n \phi_i (Z_i - a_0 - \mathbf{F}_i \boldsymbol{\theta})^T (Z_i - a_0 - \mathbf{F}_i \boldsymbol{\theta}) \right),$$

- update δ : for $k = 1, \dots, K$

$$\delta_k | \boldsymbol{\theta}, \boldsymbol{\phi}, a_0, \sigma^2, r, \rho, \boldsymbol{\gamma}, \mathbf{Z}, \mathbf{X}, \mathbf{Y} \sim \text{Gamma} \left(\frac{r}{2} + \frac{1}{2}, \frac{r}{2} + \frac{1}{2\rho} k^2 \theta_k^2 \right),$$

- update r

$$r | \boldsymbol{\theta}, \boldsymbol{\phi}, a_0, \sigma^2, \boldsymbol{\delta}, \rho, \boldsymbol{\gamma}, \mathbf{Z} | \mathbf{X}, \mathbf{Y} \\ \sim \pi \left(r | \boldsymbol{\theta}, \boldsymbol{\phi}, a_0, \sigma^2, \boldsymbol{\delta}, \rho, \boldsymbol{\gamma}, \mathbf{Z} | \mathbf{X}, \mathbf{Y} \right) \propto \frac{\left(\frac{r}{2}\right)^{\frac{r}{2}K}}{\left\{\Gamma\left(\frac{r}{2}\right)\right\}^K} \exp \left\{ -r \left(\frac{1}{2} \sum_{k=1}^K \delta_k - \frac{1}{2} \sum_{k=1}^K \ln \delta_k \right) \right\}.$$

For the ease of notation, let the target distribution as $P(r) = \{(r/2)^{(r/2)K} / \{\Gamma(r/2)\}^K\} \exp\{-r((1/2) \sum_{k=1}^K \delta_k - (1/2) \sum_{k=1}^K \ln \delta_k)\}$. Consider an exponential distribution with parameter $\lambda = (1/2) \sum_{k=1}^K \delta_k - (1/2) \sum_{k=1}^K \ln \delta_k$ as a candidate distribution

$$g^*(r | r_t) = \text{Exponential} \left(r \left| \lambda = \frac{1}{2} \sum_{k=1}^K \delta_k - \frac{1}{2} \sum_{k=1}^K \ln \delta_k \right. \right),$$

then for each iteration t ,

- generate a random candidate r' from the candidate distribution $r' \sim g^*(r | r_t)$
- calculate the acceptance probability

$$A(r', r_t) = \min \left(1, \frac{P(r') g^*(r_t | r')}{P(r_t) g^*(r' | r_t)} \right)$$

- $r_{t+1} = r'$ if $u \leq A(r', r_t)$ or $r_{t+1} = r_t$ if $u > A(r', r_t)$, where $u \sim U(0, 1)$

- update ρ

$$\rho | \boldsymbol{\theta}, \boldsymbol{\phi}, a_0, \sigma^2, \boldsymbol{\delta}, r, \boldsymbol{\gamma}, \mathbf{Z} | \mathbf{X}, \mathbf{Y} \sim \text{IG} \left(\frac{K}{2} - 1, \frac{1}{2} \sum_{k=1}^K k^2 \delta_k \theta_k^2 \right).$$

Appendix B: Supplemental Tables and Plots

Table B.1: Estimated value and standard error (s.e.) of selected regression parameters in EM based proportional-odds logistic regression (polr) and in EM based Bayesian proportional-odds logistic regression (bayespolr), and in Bayesian cumulative logistic PC regression (bayesPC(K=3)) for simulated data

	true	polr		bayespolr		bayesianPC	
		estimate	s.e	estimate	standard error	s.e	s.e
X1	1.00	-72.29	8055354.67	-0.19	440.37	2.31	0.13
X2	1.00	-989.25	23279508.27	17.21	509.22	1.64	0.09
X3	1.00	1160.6	3065818.64	3.83	749.09	2.21	0.11
X4	1.00	1125.13	21542791.4	4.23	578.85	4.18	0.19
X5	1.00	-2067.27	10574041.95	14.75	672.63	4.36	0.18
X6	1.00	1136.96	11201451.06	5.86	531.71	3.80	0.18
X7	1.00	-184.73	6851054.61	0.86	594.45	4.06	0.22
X8	1.00	413.57	14165806.42	8.13	524.29	5.27	0.30
X9	1.00	-540.99	27327338.17	5.02	638.53	5.12	0.29
X10	1.00	1501.02	9970514.25	13.33	710.27	4.36	0.26
X21	1.00	172.77	9156189.68	22.05	599.75	4.44	0.17
X22	1.00	345.29	10475712.67	0.09	584.88	3.36	0.14
X23	1.00	1195.68	7540582.87	10.23	681.82	4.12	0.16
X24	1.00	230.59	12162209.05	-7.42	633.15	4.55	0.18
X25	1.00	333.14	4259936.06	2.18	641.95	4.53	0.18
X26	1.00	-174.64	11016684.56	3.49	568.32	5.71	0.25
X27	1.00	-348.29	6888168.51	16.08	710.90	4.94	0.23
X28	1.00	823.1	6345472.98	0.70	477.04	4.38	0.23
X29	1.00	-406.82	12339415.15	6.75	613.20	4.48	0.24
X30	1.00	864.38	5988978.42	14.71	580.43	2.88	0.19
X41	0.00	-465.59	7455209.5	12.52	551.48	-0.13	0.06
X42	0.00	-140.18	9274035.47	3.39	669.30	1.60	0.09
X43	0.00	1201.24	10358985.13	-16.64	666.33	0.96	0.07
X44	0.00	311.66	6866423.31	-4.13	569.38	-0.04	0.09
X45	0.00	174.68	10334135.45	15.01	661.61	-0.17	0.07
X46	0.00	-540.74	10310530.19	14.61	796.02	-1.16	0.10
X47	0.00	455.75	3635108.05	3.49	586.07	-0.01	0.06
X48	0.00	325.7	8039198.69	-2.51	627.92	0.90	0.06
X49	0.00	-64.21	26844484.18	-10.95	706.34	-0.29	0.07
X50	0.00	NA	NA	-3.29	511.10	-0.27	0.07
X51	0.00	NA	NA	-0.18	479.21	-0.14	0.07
X52	0.00	NA	NA	-2.52	719.77	-0.70	0.07
X53	0.00	NA	NA	-5.73	755.87	-0.35	0.06
X54	0.00	NA	NA	2.86	586.26	0.20	0.08
X55	0.00	NA	NA	3.63	638.93	0.72	0.08
X56	0.00	NA	NA	-13.43	532.67	0.53	0.07
X57	0.00	NA	NA	-2.71	587.05	1.71	0.11
X58	0.00	NA	NA	-3.15	557.46	1.75	0.09
X59	0.00	NA	NA	2.07	551.61	0.67	0.05
X60	0.00	NA	NA	-2.50	569.40	-0.13	0.06
X1.2	-1.90	-164.42	24273592.35	-7.97	3.46	0.00	0.00
X2.3	0.00	-12.02	5147311.3	0.02	2.81	8.25	0.42
X3.4	1.80	93.3	35373841.15	8.41	3.43	14.29	0.59

Table B.2: Estimated values and 95% credible intervals of regression parameter of Bayesian cumulative logistic PC regression with 5 components (bayesPC(K=5)) and with 3 components (bayesPC(K=3)) of wheat data

	bayesPC(K=5)			bayesPC(K=3)		
	estimates	2.5%	97.5%	estimates	2.5%	97.5%
class	0.08	-0.31	0.48	-0.00	-0.01	0.01
density	-0.02	-0.11	0.08	0.00	-0.00	0.01
hardness	0.01	0.00	0.02	0.01	0.00	0.02
size	-0.02	-0.40	0.37	0.01	0.00	0.01
weight	0.15	0.11	0.19	0.15	0.12	0.19
moisture	-0.04	-0.16	0.09	-0.02	-0.12	0.09
Scab Sprout	0.00	0.00	0.00	0.00	0.00	0.00
Sprout Healthy	1.84	1.52	2.17	1.84	1.52	2.20

Table B.3: Mean and 95% confidence intervals of explanatory variables in wheat types)

	Healty			Sprout			Scab		
	estimates	2.5%	97.5%	estimates	2.5%	97.5%	estimates	2.5%	97.5%
density	1.28	1.26	1.29	1.19	1.18	1.21	1.08	1.04	1.13
hardness	28.22	23.25	33.18	18.06	12.72	23.40	31.17	24.88	37.47
size	2.33	2.26	2.41	2.37	2.28	2.46	1.86	1.76	1.97
weight	30.42	29.16	31.69	30.65	29.36	31.94	20.48	19.07	21.90
moisture	11.28	10.84	11.72	11.18	10.78	11.59	11.10	10.71	11.49

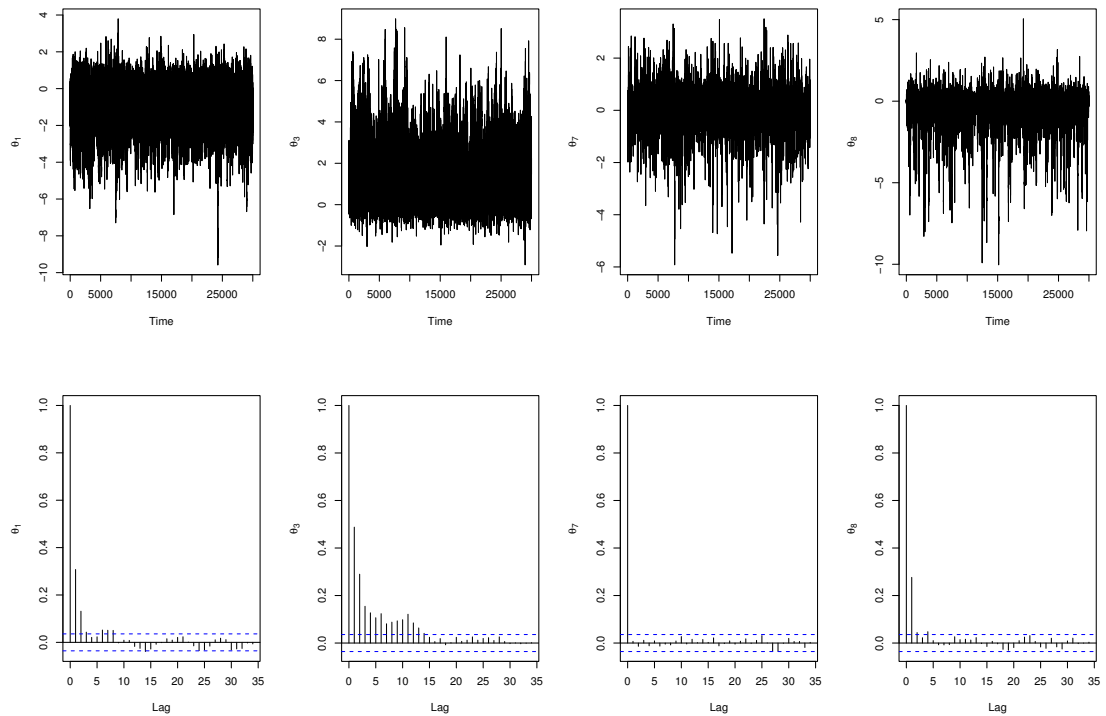


Figure B.1: The convergence plot of the selected posterior parameters for the simulated data.

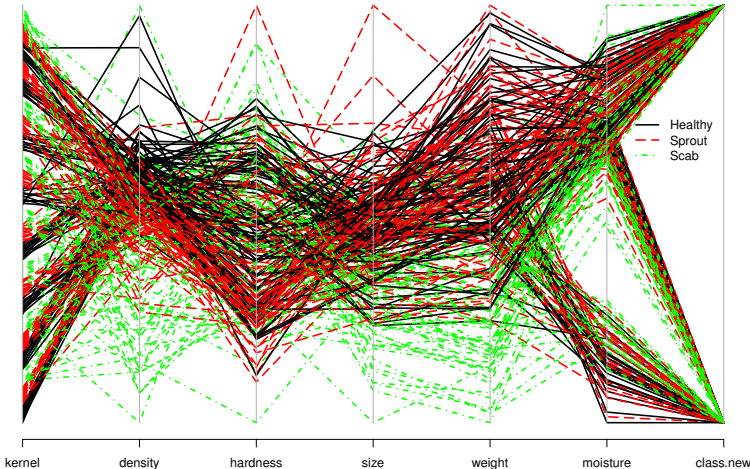


Figure B.2: Parallel coordinate plot of the wheat data.

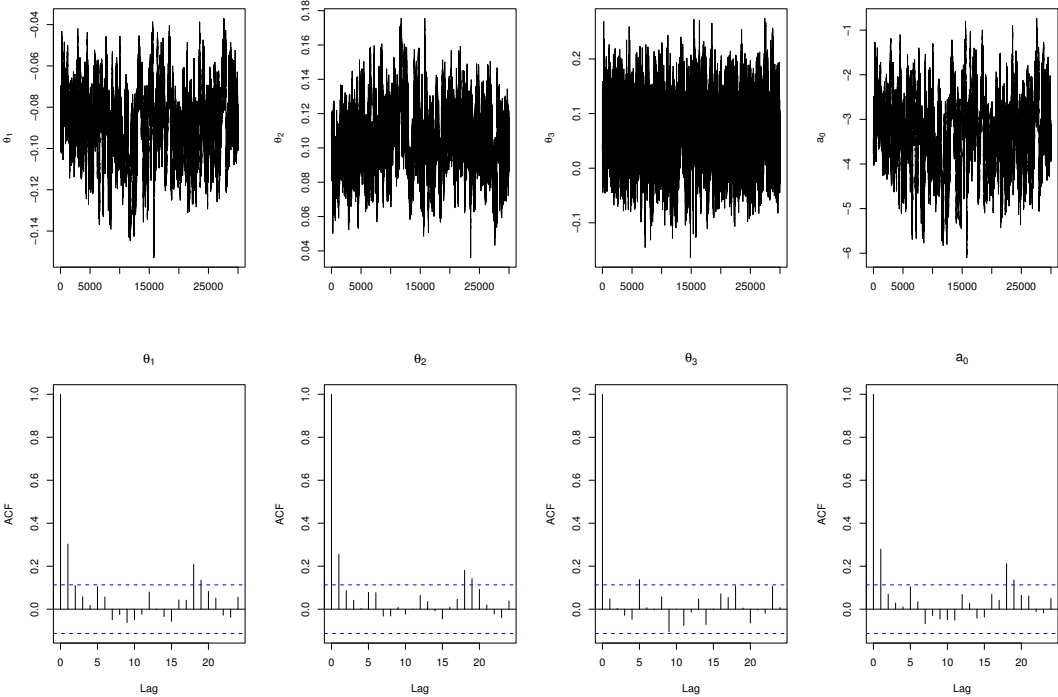


Figure B.3: The convergence plot of the posterior parameters for the wheat data.

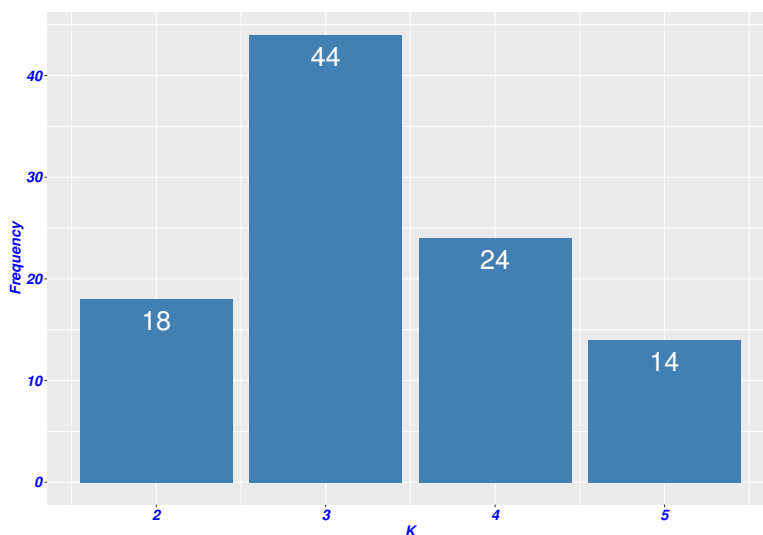


Figure B.4: Bar plot of number of components K of Bayesian cumulative logistic PC regression from 100 repeats.

References

- Agresti A (2010). *Analysis of Ordinal Categorical Data (2nd Edition)*, Wiley.
- Albert JH and Chib S (1993). Bayesian analysis of binary and polychotomous response data, *Journal of the American Statistical Association*, **88**, 669–679.
- Bair E, Hastie T, Paul D, and Tibshirani R (2006). Prediction by supervised principal components, *Journal of the American Statistical Association*, **101**, 119–137.
- Bilder CR and Loughin TM (2015). *Analysis of Categorical Data with R*, CRC Press.
- Frank LE and Friedman JH (1993). A statistical view of some chemometrics regression tools, *Technometrics*, **35**, 109–135.
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtary A, and Rubin DB (2015). *Bayesian Data Analysis (3rd ed)*, CRC press.
- Gelman A, Jakulin A, Pittau MG, and Su YS (2008). A weakly informative default prior distribution for logistic and other regression models, *The Annals of Applied Statistics*, **2**, 1360–1383.
- Gelman A and Su YS (2020). Arm: Data Analysis Using Regression and Multilevel/Hierarchical Models, R package version 1.11-2, <https://CRAN.R-project.org/package=arm>
- Hirk R, Hornik K, and Vana, L. (2019). Multivariate ordinal regression models: an analysis of corporate credit ratings, *Statistical Methods and Applications*, **28**, 507–539.
- Holmes CC and Held L (2006). Bayesian auxiliary variable models for binary and multinomial regression, *Bayesian Analysis*, **1**, 145–168.
- Kass RE and Raftery AE (1995). Bayes factors, *Journal of the American Statistical Association*, **90**, 773–795.
- Lang JB (1999). Bayesian ordinal and binary regression models with a parametric family of mixture links, *Computational Statistics & Data Analysis*, **32**, 59–87.
- Martin C, Herrman TJ, Loughin T, and Oentong S (1998). Micropycnometer measurement of single-kernel density of healthy, sprouted, and scab-damaged wheats, *Cereal Chemistry*, **75**, 177–180.
- Massy WF (1965). Principal components regression in exploratory statistical research, *Journal of the*

- American Statistical Association*, **60**, 234–256.
- McCullagh P (1980). Regression models for ordinal data, *Journal of the Royal Statistical Society, Series B.*, **42**, 109–142.
- McCullagh P and Nelder JA (1989). *Generalized Linear Models (2nd Edition)*, Chapman and Hall, London.
- McKinley TJ, Morters M, and Wood JLN (2015). Bayesian model choice in cumulative link ordinal regression models, *Bayesian Analysis*, **10**, 1–30.
- O’Brien SM and Dunson DB (2004). Bayesian multivariate logistic regression, *Biometrics*, **60**, 739–746.
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>
- Raftery AE (1995). Bayesian model selection in social research, *Sociological Methodology*, **25**, 111–163.
- Reiss PJ and Ogden TR (2007). Functional principal component regression and functional partial least squares, *Journal of the American Statistical Association*, **102**, 984–997.
- Sha N, Vannucci M, Tadesse MG, *et al.* (2004). Bayesian variable selection in multinomial probit models to identify molecular signatures of disease stage, *Biometrics*, **60**, 812–819.
- Venables WN and Ripley BD (2002). *Modern Applied Statistics with S (4th ed.)*, Springer, New York.
- Walker SH and Duncan DB (1967). Estimation of the probability of an event as a function of several independent variables, *Biometrika*, **54**, 167–179.
- West M (2003). Bayesian factor regression models in the “Large p, Small n” paradigm, *Bayesian Statistics 7*, 723–732.
- Yi N, Banerjee S, Pomp D, and Yandell BS (2007). Bayesian mapping of genomewide interacting quantitative trait loci for ordinal traits, *Genetics*, **176**, 1855–1864.

Received September 04, 2021; Revised January 24, 2022; Accepted February 21, 2022