Target Journal: Statistics & Probability Letters or Journal of Applied Statistics (Review Article)

Formatted: Highlight

Tentative Title: Bayesian Modeling of Complex Latent Structures in Regression Analysis

Authors: Chase J. Sakitis[1], Daniel B. Rowe[2]

[1]Division of Health Services & Outcomes Research, Children's Mercy, Kansas City, MO, USA
[2]Department of Mathematical and Statistical Sciences, Marquette University, Milwaukee, WI, USA

None of the authors have conflict-of-interest or financial disclosures to declare.

*Correspondence to:
Chase J. Sakitis, Ph.D.
2401 Gillham Rd
Children's Mercy
Kansas City, MO 64108
cjsakitis@cmh.edu

Daniel B. Rowe – daniel.rowe@marquette.edu

NOTES:

- I looked through the list of titles you had sent and liked the one that is currently the tentative title at the top
- This paper reads as if I took the PowerPoint presentation and condensed it into a paper
- I did not insert any references or discuss particular applications yet as I wanted to see your thoughts on it so far
- I think this would probably fit more in the Statistics & Probability Letters since we are focusing on the methodology and don't have any certain application (as of now)
- More needs to go in the Discussion section but I think that depends on the direction we want to go in
- Some of the equations can be moved to an appendix or supplemental page

Outline

1. Introduction

1.1. Linear Regression

Linear regression is a common statistical model utilized to predict the behavior of a variable based on other variables. This linear regression model can be expressed with no y-intercept ($\beta_0 = 0$) as shown in Eq. 1.1

$$y_k = \beta_1 x_{k1} + \cdots + \beta_p x_{kp} + \varepsilon_i, \quad k = 1, \ldots, n \qquad [1.1]$$

where p is the number of regression coefficients and n is the number of observations. The model expressed in Eq. 1.1 can be compactly written as

$$y = X\beta + \varepsilon \qquad [1.2]$$

Deleted: ,

Where $\varepsilon \sim N(0, \sigma^2 I_n)$, $y \in \mathbb{R}^{n\times 1}$ is the observed data, $X \in \mathbb{R}^{n\times p}$ is the design matrix, $\beta \in \mathbb{R}^{p\times 1}$ is the regression coefficients, $\varepsilon \in \mathbb{R}^{n\times 1}$ is the residuals (measurement error), and σ^2 is the variance of the residuals. One purpose of utilizing a linear regression model is to estimate the behavior of the regression coefficients β based on the design matrix X (predictor variables) given the observed data y. Estimating β can also provide hidden results yielded from the observed data and given design matrix. These regression coefficients can be estimated via

$$\hat{\beta} = (X'X)^{-1}(X'y). \qquad [1.3]$$

Utilizing Eq. 1.3 for estimating β is a simple and very common approach when the parameters and data are real-valued and the design matrix is observed or known. However, there are numerous applications that have an unknown design matrix, the parameters are complex-valued instead of real-valued or have both these occurrences.

1.2. Overview

In the real-world cases of not having a known design matrix and complex-valued data, along with β, we will also need to estimate X and account for complex-valued nature of the data in these applications. Here, we will explain how to handle this linear regression model in these scenarios building up from a real-valued linear regression model with an observed X to a complex-valued linear regression model with an unobserved X. In Section 2, we will discuss the standard linear regression model with an observed design matrix. We will then discuss the linear regression model when the design matrix is unobserved. In Section 4, we will introduce complex-valued parameters and how we handle the latent regression model. We will then finish with a discussion in Section 5.

2. Model 1: Real-Valued Parameters with an Observed Design Matrix

In this first model, we will describe the linear regression model with real-valued parameters and an observed design matrix. In this “textbook scenario”, the observed parameters are dependent data y and the design matrix X. The straightforward approach to estimating β given the observed data is to utilize Eq. 1.3.

Deleted: evaluate

2.1. Bayesian Approach to Simple Linear Regression

For a Bayesian approach to this model, we will first need to define the data likelihood, prior distributions, and the posterior distribution. From these, we can then derive posterior marginals and posterior conditionals to estimate the parameters of interest from the posterior distribution. Since the observed data y is assumed to follow a normal distribution, the data likelihood is expressed as

Deleted: To analyze

$$f(y|X,\beta,\sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)\right]. \qquad [2.1]$$

For this model, the regression coefficiets β and the variance of the residuals σ^2 are the unobserved parameters that need to be estimated. This requires prior distributions to be placed on both parameters. These prior distributions are

$$f(\beta|\beta_0,\sigma^2,n_\beta) \propto (\sigma^2)^{-\frac{p}{2}} \exp\left[-\frac{n_\beta}{2\sigma^2}(\beta-\beta_0)'(\beta-\beta_0)\right], \qquad [2.2]$$

$$f(\sigma^2|\alpha,\gamma) \propto (\sigma^2)^{-(\alpha+1)} \exp\left[-\frac{\gamma}{\sigma^2}\right]. \qquad [2.3]$$

Eqs. 2.2 and 2.3 exhibit a normal distribution for β and an inverse gamma distribution for σ^2, and are conjugate priors, simplifying the determination of the posterior distribution. Combining the data likelihood and prior distributions, we determine the posterior to be

Deleted: leaving

Deleted: easing

$$f(\beta,\sigma^2|X,y) \propto (\sigma^2)^{-\left(\frac{n+p+2\alpha}{2}+1\right)} \exp\left[-\frac{h}{2\sigma^2}\right], \qquad [2.4]$$

where $h = (y-X\beta)'(y-X\beta) + n_\beta(\beta-\beta_0)'(\beta-\beta_0) + 2\gamma$.

2.2. Posterior Marginals and Conditionals and Parameter Estimation

For the posterior marginals, we aim to integrate out the other unobserved parameters from the posterior, leaving the parameter we are currently estimating. So, for β, we would integrate out

σ^2 from the posterior, leaving only observed data to estimate β and then vice versa for σ^2. This will give us

$$f(\beta|X,y) \propto \left[1+\frac{1}{\nu}(\beta-\hat{\beta})'\left[\frac{\nu(X'X+n_\beta I_p)}{\phi}\right](\beta-\hat{\beta})\right]^{-\frac{\nu+p}{2}}, \qquad [2.5]$$

$$f(\sigma^2|X,y) \propto (\sigma^2)^{-(\alpha+1)}\exp\left[-\frac{\gamma_*}{\sigma^2}\right], \qquad [2.6]$$

where β follows a Student-t distribution such that $\beta|X,y \sim t(\nu,\hat{\beta},\mathrm{T})$, $E(\beta|X,y)=\hat{\beta}=(X'X+n_\beta I_p)^{-1}(X'y+n_\beta\beta_0)$, $Var(\beta|X,y)=\frac{\nu}{\nu-2}T$, and $\nu=2\alpha+n$ and σ^2 follows an inverse gamma distribution such that $\sigma^2\,|X,y \sim IG\left(\alpha_*=\frac{n+p+2\alpha}{2},\gamma_*=\frac{\phi}{2}\right)$, $E(\sigma^2|X,y)=\frac{\gamma_*}{\alpha_*-1}$, $Var(\sigma^2|X,y)=\frac{\gamma_*^2}{(\alpha_*-1)^2(\alpha_*-2)}$, and $\phi=y'y-\hat{\beta}'(X'X+n_\beta I_p)\hat{\beta}+n_\beta\beta_0'\beta_0+2\gamma$. Since X and y in the expected value equations ($E(\beta|X,y)$ and $E(\sigma^2|X,y)$) are known, we can simply calculate the expected value for β and σ^2 without requiring any computational techniques.

Deleted: s

For the posterior conditionals, instead of integrating out β and σ^2, we can leave the unobserved parameters in, mathematically determine the posterior conditionals, and utilize a computational technique to estimate β and σ^2. These yields

$$f(\beta|X,\sigma^2,y) \propto |\Sigma_\beta|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\beta-\hat{\beta})'(\Sigma_\beta)^{-1}(\beta-\hat{\beta})\right], \qquad [2.7]$$

$$f(\sigma^2|\beta,X,y) \propto (\sigma^2)^{-(\alpha_*+1)}\exp\left[-\frac{\gamma_{**}}{\sigma^2}\right], \qquad [2.8]$$

where β follows a normal distribution such that $\beta|X,\sigma^2,y \sim N\left(\hat{\beta}=(X'X+n_\beta I_p)^{-1}(X'y+n_\beta\beta_0),\Sigma_\beta=\sigma^2(X'X+n_\beta I_p)^{-1}\right)$ and σ^2 follows an inverse gamma distribution such that $\sigma^2|\beta,X,y \sim IG(\alpha_*,\gamma_{**})$ and $\gamma_{**}=[(\beta-\hat{\beta})'(X'X+n_\beta I_p)(\beta-\hat{\beta})+\phi]/2$. Since β and σ^2 are unobserved, we require a computational technique to estimate them from their posterior conditionals. In this case, we can use a Markov chain Monte Carlo (MCMC) technique, such as Gibbs sampling, to estimate β and σ^2.

Deleted: $\frac{1}{2}$

3. Model 2: Real-Valued Parameters with an Unobserved Design Matrix

In this second model, we will describe the linear regression model with real-valued parameters and an unobserved design matrix. In this scenario, the only observed parameter is the

Deleted: evaluate

dependent data y. The approach for estimating β, X, and σ^2 in this scenario requires a computational technique like Gibbs sampling.

3.1. Bayesian Approach to Linear Regression with Unobserved Design Matrix

Similar to the previous model, we first need to define the data likelihood, the prior distributions, and the posterior distribution before deriving the posterior marginals and conditionals. For the data likelihood, since the observed data y still follows a normal distribution, we can use Eq. 2.1 as the data likelihood for this model. For the prior distributions of β and σ^2, they will still follow normal and inverse gamma distributions as expressed in Eqs. 2.2 and 2.3, respectively. Since the design matrix X is also unobserved, it also requires a prior distribution to be placed on it. To keep consistency with conjugate priors, X will follow a normal distribution as shown in Eq. 3.1.

Deleted: Like before

$$f(X|X_0,\sigma^2,n_x) \propto (\sigma^2)^{-\frac{np}{2}} \exp\left[-\frac{n_x}{2\sigma^2} tr\{(X-X_0)(X-X_0)'\}\right], \quad [3.1]$$

where tr is the trace of the $(X-X_0)(X-X_0)'$ matrix. Combining the data likelihood and prior distributions, the posterior distribution we obtain is

Deleted: get

$$f(\beta,\sigma^2|X,y) \propto (\sigma^2)^{-\left(\frac{np+n+p+2\alpha}{2}+1\right)} \exp\left[-\frac{h_*}{2\sigma^2}\right], \quad [3.2]$$

where $h_* = (y-X\beta)'(y-X\beta) + n_\beta(\beta-\beta_0)'(\beta-\beta_0) + n_x tr\{(X-X_0)(X-X_0)'\} + 2\gamma$.

3.2. Posterior Marginals and Conditionals and Parameter Estimation

With the addition of X being an unobserved variable, we can no longer properly integrate out both X and σ^2 to estimate β. We can still integrate out σ^2 for both β and X, yielding

$$f(\beta|X,y) \propto \left[1+\frac{1}{v_\beta}(\beta-\hat{\beta})'\left[\frac{v_\beta(X'X+n_\beta I_p)}{\phi}\right](\beta-\hat{\beta})\right]^{-\frac{v_\beta+p}{2}}, \quad [3.3]$$

$$f(X|\beta,y) \propto \left[1+\frac{1}{v_x}(X-\hat{X})\left[\frac{v_x(\beta\beta'+n_x I_p)}{\theta}\right](X-\hat{X})'\right]^{-\frac{v_x+np}{2}}, \quad [3.4]$$

where β follows student-t distributions such that $\beta|X,y \sim t(v_\beta,\hat{\beta},T)$, $E(\beta|X,y)=\hat{\beta}=(X'X+n_\beta I_p)^{-1}(X'y+n_\beta\beta_0)$, $Var(\beta|X,y)=\frac{v_\beta}{v_\beta-2}T$, and X follows a multivariate Student-t distribution such that $X|\beta,y \sim Mt(v_x,\hat{X},\Delta)$, $E(X|\beta,y)=\hat{X}=(y\beta'+n_xX_0)(\beta\beta'+n_xI_p)^{-1}$, $Var(X|\beta,y)=\frac{v_x}{v_x-2}\Delta$. Since these posteriors marginals still have unobserved parameters, a computational technique, like an MCMC Gibbs sampler, is necessary for estimating β and X.

Deleted: s

Like Model 1, for the posterior conditionals, instead of integrating out β, X, or σ^2, we can retain the unobserved parameters, mathematically determine the posterior conditionals, and utilize a computational technique to estimate β, X, and σ^2. These yield

$$f(\beta|X,\sigma^2,y) \propto |\Sigma_\beta|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\beta-\hat{\beta})'(\Sigma_\beta)^{-1}(\beta-\hat{\beta})\right], \qquad [3.5]$$

$$f(X|\beta,\sigma^2,y) \propto |\Sigma_X|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}tr\{(X-\hat{X})(\Sigma_X)^{-1}(X-\hat{X})'\}\right], \qquad [3.6]$$

$$f(\sigma^2|\beta,X,y) \propto (\sigma^2)^{-(\alpha_*+1)} \exp\left[-\frac{\gamma_{**}}{\sigma^2}\right], \qquad [3.7]$$

where β follows a normal distribution such that $\beta|X,\sigma^2,y \sim N\left(\hat{\beta} = (X'X + n_\beta I_p)^{-1}(X'y + n_\beta\beta_0), \Sigma_\beta = \sigma^2(X'X + n_\beta I_p)^{-1}\right)$, X follows a multivariate normal distribution such that $X|\beta,\sigma^2,y \sim MN\left(\hat{X} = (y\beta' + n_x X_0)(\beta\beta' + n_x I_p)^{-1}, \Sigma_X = \sigma^2(\beta\beta' + n_x I_p)^{-1}\right)$, and σ^2 follows an inverse gamma distribution such that $\sigma^2|\beta,X,y \sim IG(\alpha_*,\gamma_{**})$, $\alpha_* = (np + n + p + 2\alpha)/2$, and $\gamma_{**} = [(y - X\beta)'(y - X\beta) + n_\beta(\beta - \beta_0)'(\beta - \beta_0) + n_x tr\{(X - X_0)(X - X_0)'\} + 2\gamma]/2$. Similar to the posterior marginals, the poster conditionals also require the use of an MCMC Gibbs sampler (or another computational method) for estimating β, X, and σ^2. Since posterior marginals and conditionals for β are the Student-t and normal distributions respectively, the expected values (mean) $\hat{\beta}$ are theoretically equivalent. This means that you can use the posterior marginals as a method of confirming the estimates from using the posterior conditionals (or vice versa). This can similarly be said for X.

4. Model 3: Complex-Valued Parameters with an Unobserved Design Matrix

In this third model, we will evaluate the linear regression model with complex-valued parameters and an unobserved design matrix. Like Model 2, the only observed parameter is the dependent data y and the approach for estimating β, X, and σ^2 also requires a computational method. For this model, we will introduce a real-valued isomorphic representation for the model.

4.1. Real-valued Isomorphic Representation

Transforming the complex-valued parameters into a real-valued representation allows for easier parameter estimation compared to the complex-valued counterpart. The complex-valued linear regression can be expressed as

$$y_c = X_c\beta_c + \varepsilon_c, \qquad [4.1]$$

Deleted: leave
Deleted: in
Deleted: 0.5
Deleted: 0.5
Deleted: Like
Deleted: ing
Deleted: s

with "*c*" subscripts representing that the parameter is complex-valued. For the real-valued isomorphic representation, each complex-valued parameter will be split into the real component (represented with an "*R*" subscript) and the imaginary component (represented with an "*I*" subscript). We can then express Eq 4.1 as

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

$$\begin{bmatrix} y_R \\ y_I \end{bmatrix} = \begin{bmatrix} x_R & -x_I \\ x_I & x_R \end{bmatrix} \begin{bmatrix} \beta_R \\ \beta_I \end{bmatrix} + \begin{bmatrix} \varepsilon_R \\ \varepsilon_I \end{bmatrix}, \quad [4.2]$$

where $(\varepsilon_R, \varepsilon_I)' \sim N(0, \sigma^2 I_{2n})$. This linear model can be rewritten as Eq 1.2 where $y \in \mathbb{R}^{2n\times 1}$, $X \in \mathbb{R}^{2n\times 2p}$, $\beta \in \mathbb{R}^{2p\times 1}$, and $\varepsilon \in \mathbb{R}^{2n\times 1}$. We note that estimating $\hat{\beta}$ via Eq. 1.3 with an observed design matrix X_c yields different results when the parameters are kept in complex-valued form compared with using its real-valued isomorphic representation.

Commented [DR1]: I think you need to show the two results here with derivation in an appendix.

4.2. Bayesian Approach using the Isomorphic Representation

For the real-valued isomorphic model, there are two representations for the design matrix X. The first one, expressed as X, is the skew-symmetric form expressed in Eq. 4.2. The second representation is expressed as $H = [x_R \quad x_I]$ yielding $H \in \mathbb{R}^{n\times 2p}$. This second representation is utilized to estimate the real and imaginary components of the design matrix only once since each component is expressed only one time in H, but twice in X. After estimating the real and imaginary components of the design matrix using the H representation, these components are then formulated into the X representation for estimating β and σ^2.

With this real-valued isomorphic representation, despite similarity to Model 2, we still need to define the data likelihood and prior distributions to obtain our posterior distribution. For the observed data, the real and imaginary components follow a normal distribution, yielding

$$f(y|X,\beta,\sigma^2) \propto (\sigma^2)^{-\frac{2n}{2}} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right]. \quad [4.3]$$

As for the prior distributions, β, X, and σ^2 will follow a multivariate normal distribution, matrix normal distribution, and an inverse gamma distribution, respectively, expressed as,

Deleted: ultivariate

$$f(\beta|\beta_0,\sigma^2,n_\beta) \propto (\sigma^2)^{-\frac{2p}{2}} \exp\left[-\frac{n_\beta}{2\sigma^2}(\beta - \beta_0)'(\beta - \beta_0)\right], \quad [4.4]$$

$$f(H|H_0,\sigma^2,n_x) \propto (\sigma^2)^{-\frac{2np}{2}} \exp\left[-\frac{n_x}{2\sigma^2} tr\{(H - H_0)(H - H_0)'\}\right], \quad [4.5]$$

$$f(\sigma^2|\alpha,\gamma) \propto (\sigma^2)^{-(\alpha+1)} \exp\left[-\frac{\gamma}{\sigma^2}\right]. \quad [4.6]$$

One notable difference between this model and Model 2 with the data likelihood and the priors is the exponent of the σ^2 component of each equation (except Eq. 4.6) due to the dimensions of the parameters from the isomorphic representation. The resulting combination of the likelihood and the prior distributions is

$$f(\beta,\sigma^2|X,y) \propto (\sigma^2)^{-\left(\frac{2np+2n+2p+2\alpha}{2}+1\right)} \exp\left[-\frac{h_{**}}{2\sigma^2}\right], \qquad [4.7]$$

where $h_{**} = (y - X\beta)'(y - X\beta) + n_\beta(\beta - \beta_0)'(\beta - \beta_0) + n_x tr\{(H - H_0)(H - H_0)'\} + 2\gamma$.

4.3. Posterior Marginals and Conditionals and Parameter Estimation

Similar to Model 2, we cannot properly integrate out both X and σ^2 to estimate β, but can still integrate out σ^2 for both β and X to obtain the posterior marginals. These posterior marginals are

Deleted: Like

$$f(\beta|X,y) \propto \left[1 + \frac{1}{\nu_\beta}(\beta - \hat{\beta})'\left[\frac{\nu_\beta(X\prime X + n_\beta I_{2p})}{\phi}\right](\beta - \hat{\beta})\right]^{-\frac{\nu_\beta+2p}{2}}, \quad [4.8]$$

$$f(H|\beta,y) \propto \left|I_n + \frac{1}{\nu_x}(H - \hat{x})\left[\frac{\nu_x(CC\prime + n_x I_{2p})}{\theta}\right](H - \widehat{H})'\right|^{-\frac{\nu_x+2np}{2}}, \quad [4.9]$$

where β follows student-t distributions such that $\beta|X,y \sim t(\nu_\beta, \hat{\beta}, T)$, $E(\beta|X,y) = \hat{\beta} = (X'X + n_\beta I_{2p})^{-1}(X'y + n_\beta\beta_0)$, $Var(\beta|X,y) = \frac{\nu_\beta}{\nu_\beta-2}T$, and X follows a multivariate student-t distribution such that $H|\beta,y \sim Mt(\nu_x, \widehat{H}, \Delta)$, $E(H|\beta,y) = \widehat{H} = (YC' + n_x x_0)(CC' + n_x I_{2p})^{-1}$, $Var(H|\beta,y) = \frac{\nu_x}{\nu_x-2}\Delta$, $C = \begin{bmatrix} \beta_R & \beta_I \\ -\beta_I & \beta_R \end{bmatrix}$, and $Y = [y_R \quad y_I]$.

For the posterior conditionals for β, X, or σ^2, we have

$$f(\beta|X,\sigma^2,y) \propto |\Sigma_\beta|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\beta - \hat{\beta})'(\Sigma_\beta)^{-1}(\beta - \hat{\beta})\right], \qquad [4.10]$$

$$f(H|\beta,\sigma^2,y) \propto |\Sigma_X|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(H - \widehat{H})(\Sigma_X)^{-1}(H - \widehat{H})'\right], \qquad [4.11]$$

$$f(\sigma^2|\beta,X,y) \propto (\sigma^2)^{-(\alpha_*+1)} \exp\left[-\frac{\gamma_{**}}{\sigma^2}\right], \qquad [4.12]$$

where β follows a multivariate normal distribution such that $\beta|X,\sigma^2,y \sim N\left(\hat{\beta} = (X'X + n_\beta I_p)^{-1}(X'y + n_\beta\beta_0), \Sigma_\beta = \sigma^2(X'X + n_\beta I_p)^{-1}\right)$, H follows a matrix normal distribution

Deleted: ultivariate

such that $H|\beta,\sigma^2,y \sim MN\left(\widehat{H} = (YC' + n_x x_0)(CC' + n_x I_{2p})^{-1}, \Sigma_X = \sigma^2(CC' + n_x I_{2p})^{-1}\right)$, and σ^2 follows an inverse gamma distribution such that $\sigma^2|\beta,X,y \sim IG(\alpha_*,\gamma_{**})$, $\alpha_* = np + n + p + \alpha$), and $\gamma_{**} = [(y - X\beta)'(y - X\beta) + n_\beta(\beta - \beta_0)'(\beta - \beta_0) + n_x tr\{(H - H_0)(H - H_0)'\} + 2\gamma]/2$. Like Model 2, the posterior marginals and conditionals require a computational technique, such as a Gibbs sampler, for estimating β, X, and σ^2. We can also use the posterior marginals of β and X to theoretically validate the results of the distributional mean of their respective posterior conditionals.

I think for each model you should add an Estimation subsection describing Gibbs sampler and ICM algorithm.

5. Discussion

In this paper, we reviewed the effects unobservable and complex-valued parameters have on a complex-valued latent variable linear regression model. We also discussed that a computational technique is required to estimate the parameters β, X, and σ^2 in Models 2 and 3. For β and X in these models, either the posterior marginals or posterior conditionals can be utilized for parameter estimation.

I think perhaps the addition of a table for the three models with conditional dist for each parameter, and mode for ICM.

What about some sort of illustrative example?

Deleted: 0.5

Deleted: like

Commented [DR2]: Describe this more with potential details in an appendix.