

# Gibbs Sampling with Data Augmentation for Bayesian Analysis of Binary and Polychotomous Response Data

STAT 230 Final Report

*Damon Bayer & Corey Katz*

*12/16/2019*

## Contents

<b>1</b>	<b>Abstract</b>	<b>1</b>
<b>2</b>	<b>Introduction</b>	<b>1</b>
<b>3</b>	<b>Methodology</b>	<b>1</b>
<b>4</b>	<b>Results</b>	<b>1</b>
4.1	Small-Cell Carcinoma . . . . .	1
4.2	Baseball . . . . .	4
4.3	Breast Cancer Data . . . . .	4
<b>5</b>	<b>Discussion</b>	<b>4</b>
<b>6</b>	<b>References</b>	<b>5</b>
<b>7</b>	<b>Appendix</b>	<b>6</b>

## 1 Abstract

Here's the abstract.

## 2 Introduction

We cite Albert and Chib (1993)

Here's the introduction.

## 3 Methodology

Here's the methodology.

## 4 Results

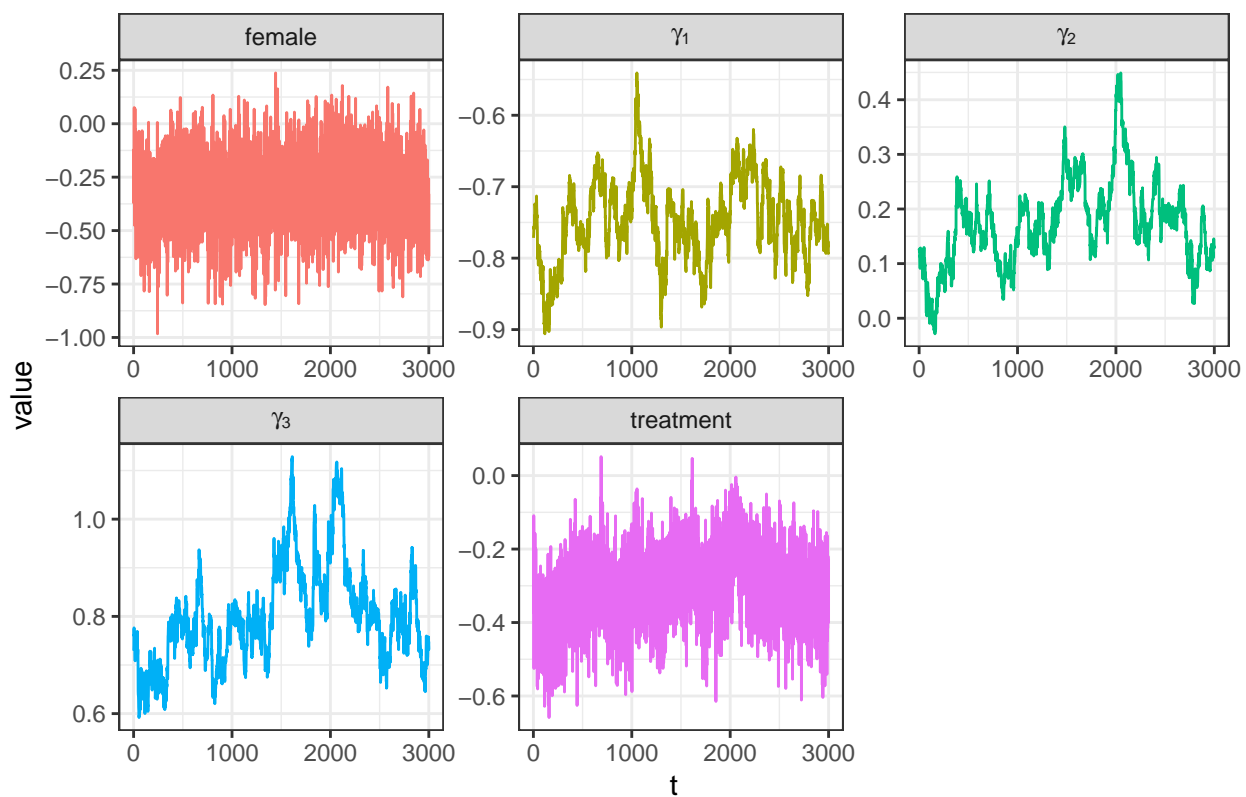
For three datasets, we apply the appropriate previously discussed model as well as standard maximum likelihood estimation and custom models written in Stan (Carpenter et al. 2017) and implemented in the RStan package (Stan Development Team 2019).

Here's the results.

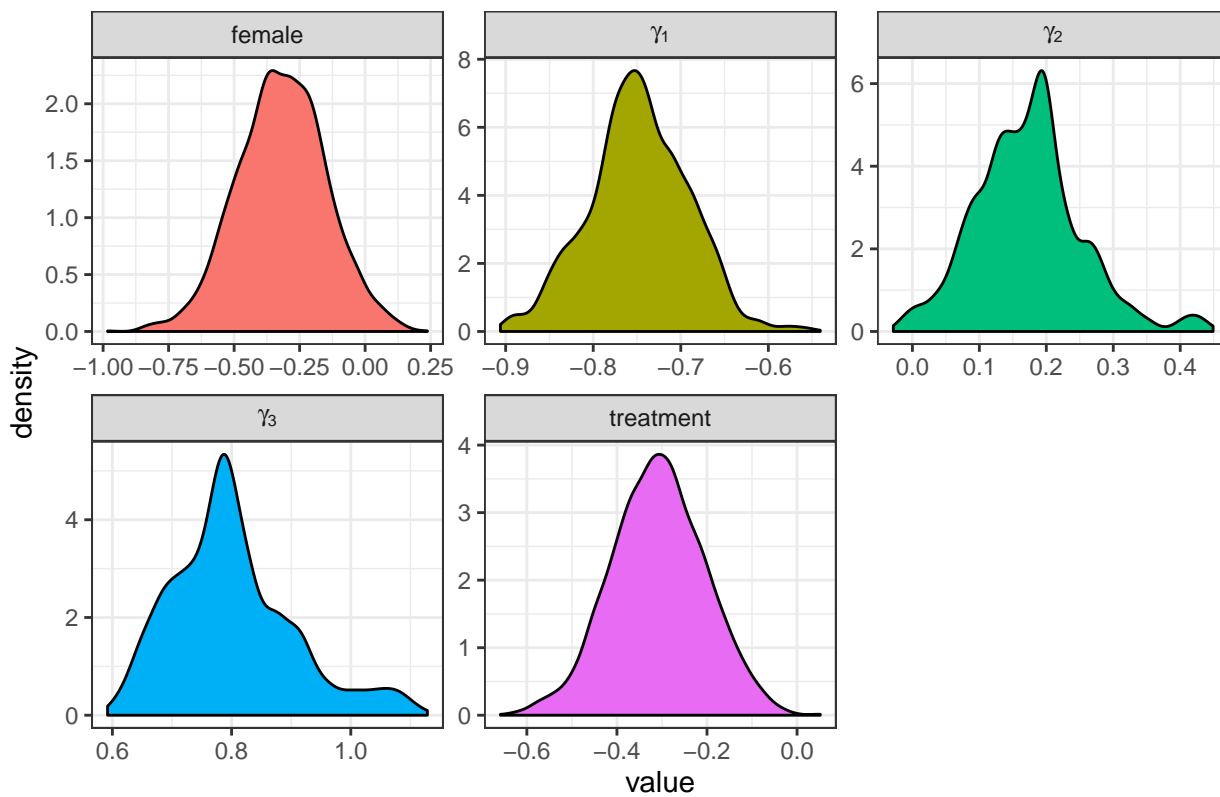
### 4.1 Small-Cell Carcinoma

Variable	mean	sd	2.5%	50%	97.5%
female	-0.3177006	0.1695306	-0.6483304	-0.3154434	0.0143534
gamma[1]	-0.7478856	0.0569057	-0.8576571	-0.7492453	-0.6420117
gamma[2]	0.1726966	0.0795173	0.0235177	0.1732758	0.3510630
gamma[3]	0.8011110	0.1027040	0.6390310	0.7890847	1.0615511
treatment	-0.3081612	0.1039444	-0.5136893	-0.3092713	-0.1059862

Trace Plot



Posterior Distributions



## 4.2 Baseball

## 4.3 Breast Cancer Data

We evaluate the probit regression methods on the Wisconsin Breast Cancer dataset, obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg [?]. The data reports 9 discrete measurements for 699 observations of clumps of breast cancer cells. The variables are Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, and Mitoses. We remove 16 missing observations from the dataset to arrive at a final dataset with  $n = 683$ ,  $p = 9$ .

Stan: 5000 iteration, 1000 burn in, 4000 posterior samples, 29.4429 sec R: 8.235 seconds

## 5 Discussion

Here's the discussion.

## 6 References

Albert, James H., and Siddhartha Chib. 1993. “Bayesian Analysis of Binary and Polychotomous Response Data.” *Journal of the American Statistical Association* 88 (422). Taylor & Francis: 669–79. <https://doi.org/10.1080/01621459.1993.10476321>.

Carpenter, Bob, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. “Stan: A Probabilistic Programming Language.” *Journal of Statistical Software, Articles* 76 (1): 1–32. <https://doi.org/10.18637/jss.v076.i01>.

Stan Development Team. 2019. “RStan: The R Interface to Stan.” <http://mc-stan.org/>.

## 7 Appendix