Published in final edited form as:

Stat Med. 2010 May 20; 29(11): 1250–1258. doi:10.1002/sim.3878.

Analysis of Extreme Drinking in Patients with Alcohol Dependence Using Pareto Regression†

Sourish Das¹, Ofer Harel^{2,*}, Dipak K. Dey², Jonathan Covault³, and Henry R. Kranzler³

¹SAMSI and Duke University, Durham, North Carolina, USA

²Department of Statistics, University of Connecticut, Storrs, Connecticut, USA

³Department of Psychiatry, University of Connecticut Health Center, Farmington, Connecticut, USA

SUMMARY

We developed a novel Pareto regression model with an unknown shape parameter to analyze extreme drinking in patients with Alcohol Dependence (AD). We used the generalized linear model (GLM) framework and the log-link to include the covariate information through the scale parameter of the generalized Pareto distribution. We proposed a Bayesian method based on Ridge prior and Zellner's *g*-prior for the regression coefficients. Simulation study indicated that the proposed Bayesian method performs better than the existing likelihood based inference for the Pareto regression.

We examined two issues of importance in the study of AD. First, we tested whether a single nucleotide polymorphism within *GABRA2* gene, which encodes a subunit of the *GABA_A* receptor, and that has been associated to AD, influences 'extreme' alcohol intake and second, the efficacy of three psychotherapies for alcoholism in treating extreme drinking behavior. We found an association between extreme drinking behavior and *GABRA2*. We also found that, at baseline, men with a high-risk *GABRA2* allele had a significantly higher probability of extreme drinking than men with no high-risk allele. However, men with a high-risk allele responded to the therapy better than those with two copies of the low-risk allele. Among men, cognitive behavioral therapy was the worst for the treatment of extreme drinking behavior. Women with high-risk alleles also responded to the therapy better than those with two copies of the low-risk allele, while women who received the cognitive behavioral therapy had better outcomes than those receiving either of the other two therapies.

1. INTRODUCTION

Alcohol dependence (AD), a disorder associated with significant adverse medical and psychosocial consequences, has been shown to be 50 – 60% heritable [1]. Bauer et al. [2] demonstrated an association between a single nucleotide polymorphism (SNP) within *GABRA2*, the gene encoding the α -2 subunit of the *GABA_A* receptor, and the probability of daily drinking and heavy drinking behavior. Heavy drinking is defined as 5 or more standard drinks per day for men and 4 or more standard drinks per day for women, where a standard

†This material was based upon work partially supported by the National Science Foundation under Grant DMS-0635449 to the Statistical and Applied Mathematical Science Institute. Any opinions, findings, conclusions and recommendations expressed in this material are those of authors and do not necessarily reflect the view of National Science Foundation. Additional support was provided by NIH grants P50 AA03510 and K24 AA013736.

Copyright © 2000 John Wiley & Sons, Ltd.

*Correspondence to: Department of Statistics, University of Connecticut, Storrs, CT-06269, oharel@stat.uconn.edu.

drink is 0.5 oz of ethanol [3]. Following this definition Bauer et al. [2] modeled heavy drinking behavior as a binary response and fit a logistic regression model. However, there are levels of heavy drinking, with values more extreme than the 2.5 or 2.0 oz thresholds having more of an adverse impact on health and behavior, leading to overdose and death in extreme cases.

To improve upon binary response logistic models, it would be natural to define an ordered categorical response with $z = 1$ if $5 \leq y < a$ (as heavy drinking), $z = 2$ if $a \leq y < b$ (very heavy drinking), $z = 3$ if $b \leq y < c$ (extreme drinking), and $z = 4$ if $c \leq y < \infty$ (fatal drinking). One can then fit a regression model with z as the response [4]. However, it is well known that categorization of a continuous variable results in it being sensitive to the choice of cut points a, b, c . If the intervals that are chosen are too wide, there may be considerable difference in sensitivity within an interval, while when the intervals are narrow, problems arise since many intervals may contain zero or a small number of subjects. This can lead to the lack of unique maximum likelihood estimates (MLE) or the posterior mode under the improper prior. We avoid categorization and instead focus on modeling the tail distribution. Tail data for the exceedances over high thresholds are often modeled by fitting a generalized Pareto distribution. Davison and Smith [5] presented the generalized Pareto regression and proposed the maximum likelihood based inference by incorporating the covariate information through the scale parameter. The application of such methods for modeling the extreme values are limited, since there may be few observations in the extreme tail, even for large samples. In addition, the existing methods cannot address the issues concerning the collinearity. This motivated us to use a Bayesian approach. We used the Ridge prior [6] and Zellner's g -prior [7], which improve the efficiency by reducing the mean square error compared to the likelihood method. To our knowledge we are the first to consider the Bayesian Pareto regression to model extreme drinking behavior.

The paper includes an additional five sections. In section 2, we present the description of the study. In section 3, we present the Pareto regression for extreme drinking and the Bayesian method to analyze the data. We discuss the rationale for choosing the Ridge and Zellner's g -prior over the other popular choices, like uniform flat prior. In section 4, we present a simulation study to check the performance of the proposed Bayesian method compared to the existing likelihood approach. In section 5, we analyze and present the data from patients in treatment for AD. Section 6 concludes the paper with a brief discussion of these findings.

2. STUDY DESCRIPTION

The present study examined a SNP marker in the AD-associated haplotype block in *GABRA2*, the gene encoding the α -2 subunit of the *GABA_A* receptor [8] using DNA samples from AD patients who participated in Project MATCH (Matching Alcoholism Treatment to Client Heterogeneity), a psychotherapy study conducted at 10 sites around the United States [9,10]. The project included two parallel, independent, clinical arms in which patients were recruited from outpatient settings ($n = 952$; 72% male) or aftercare settings following inpatient or day hospital treatment ($n = 774$; 80% male). In each of the two arms, patients were randomly assigned to one of the three manual-guided, individually delivered treatments: motivational enhancement therapy (MET), cognitive behavioral therapy (CBT), or twelve-step facilitation (TSF). In this study, we focus on the sub sample of 812 European-American patients who provided a DNA sample for analysis. This included 540 patients from outpatient settings and 272 from aftercare settings. The study included a 3-month treatment period and a 12-month post treatment period. The background characteristics of patients as a function of *GABRA2* genotype and treatment assignment are presented in Table I. The patient sample can generally be described as 73.4% male, with an average age of 40.8 years and an average 13.5 years of education.

3. PARETO REGRESSION MODEL FOR EXTREME DRINKING

Heavy drinking is defined as 5 or more standard drinks per day for men and 4 or more standard drinks per day for women. A standard drink is 0.5 oz of ethanol [3]. Following this definition of heavy drinking, we modeled extreme drinking behavior as the values exceeding these thresholds. Note that Bauer et al. [2] modeled heavy drinking as a binary response following this definition of heavy drinking. Our objective is to model fatal extreme drinking behavior as the tail probability model and not just regular heavy drinking behavior. In practice, tail data for the exceedances over high thresholds are often modeled by fitting a generalized Pareto distribution. There is a rich corresponding literature on the statistics of extreme value. The first publication on the topic was by Gnedenko [9], which was followed by Pickands [10]. Recent efforts include those by Smith [11]; Davison and Smith [5]; and Coles and Powell [12]. The Bayesian analysis of extreme rainfall data was presented by Coles and Tawn [13]. The dynamic mixture model approach for tail data was proposed by Frigessi et al. [14] and Behrens et al. [15] presented the Bayesian methods for modeling the extreme value for stock prices with threshold estimation. Recently, Castellanos and Cabras [16] developed the default Bayesian method for the generalized Pareto distribution and Tsionas [17] developed a hierarchical Pareto regression model using the generalized Pareto distribution for a macroeconomic data set using Markov Chain Monte Carlo (MCMC) techniques [18]. Here we present a Bayesian method to analyze the Pareto regression for extreme drinking behavior. A hierarchical Pareto regression can be viewed as a class of general linear model. In general, GLMs have three components: viz. (i) a random component, (ii) a systematic component and (iii) a link function. We define the Pareto regression as follows.

Random Component

Let Y be the number of standard drinks (exceeding the thresholds of heavy drinking) consumed by a patient on a particular day with a generalized Pareto distribution, say $G P D(\tau, \sigma_i, \kappa)$ with survival function as

$$G(y_i) = P(Y > y_i | Y \geq \tau) = \left\{ 1 - \kappa \frac{y_i - \tau}{\sigma_i} \right\}^{1/\kappa},$$

where $i = 1, \dots, n$, and τ is a known positive constant, $\sigma_i > 0$ is the unknown scale parameter and κ is an unknown shape parameter. Following [3] we chose $\tau = 5$ for men and $\tau = 4$ for women. The range of y_i is $\tau < y_i < \infty$ if $\kappa < 0$, $\tau < y_i < \tau + \sigma_i / \kappa$ if $\kappa > 0$. The case $\kappa = 0$ is interpreted as $\kappa \rightarrow 0$, that is, an exponential distribution with mean σ_i . Smith [19] showed that the MLE exists in a large sample, provided $\kappa < 1$ and is asymptotically normal and

efficient if $\kappa < \frac{1}{2}$.

Systematic Component

We incorporate the covariate information through the systematic component of GLM,

$$\eta_i = x_i' \beta, \quad i = 1, 2, \dots, n.$$

Link function

We use the log-link function as

$$\eta_i = \log(\sigma_i), \quad i = 1, 2, \dots, n.$$

Combining all three components of GLM, the Pareto regression is

$$G(y_i) = P(Y > y_i | Y \geq \tau) = \left\{ 1 - \frac{\kappa(y_i - \tau)}{\exp\{x_i^T \beta\}} \right\}^{1/\kappa}.$$

Note that we measure the effect of extreme drinking that exceeds the heavy drinking threshold as a continuous measure (as 0.5 oz of ethanol is defined as a standard drink and number of standard drinks can take a value up to two decimal places; see [3]). This contrasts with the binary measures used by Bauer et al. [2].

3.1. Pareto Regression for Extreme Drinking of AD patients—Despite a substantial sample size, the number of subjects and the frequency of the minor (i.e., less common) allele were not sufficient to support an analysis wherein genotype could be stratified by 3 levels, given the need to stratify by 3 levels of treatment assignment and 2 levels of sex. Instead, genotype was stratified by 2 levels: homozygotes for the low AD risk A-allele (i.e., A/A, which comprises the low-risk genotype group) versus carriers of the AD-associated G-allele (i.e., A/G or G/G, which comprise the high-risk genotype group). This strategy is consistent with the grouping scheme employed in prior studies [8,20] which showed an association of the G-allele with risk of AD, as well as the response to the acute effects of alcohol among moderate drinkers [21].

The Pareto regression for the extreme drinking of i^{th} AD patient is

$$\ln(\sigma_i) = \beta_1 + \beta_2 \ln(Day_i) + \beta_3 \ln(Age_i) + \beta_4 I(\text{High risk genotype of } GABRA2 \text{ gene})_i + \beta_5 \ln(Day)_i \times I(\text{High risk genotype of } GABRA2 \text{ gene})_i$$

where $I(A) = 1(0)$, if A is true (otherwise), is an indicator variable and $i = 1, 2, \dots, n$. Since the definition of heavy drinking differs by gender, we fit a separate model for each gender.

3.2. Bayesian Analysis and Different Choices of Prior—In analyzing complex models using Bayesian methods, we need to develop the posterior distribution of the parameters θ . In our case the parameters of interest are the regression coefficients β and the shape parameters κ . The posterior distribution is

$$\pi(\theta | y, X) \propto L(\theta, y, X) \pi(\theta),$$

where $\theta = (\beta, \kappa)$, and $L(\theta, y, X)$ is the likelihood function. After introducing the covariates, the log-likelihood function is

$$\ln(L(\theta, y, X)) = - \sum_{i=1}^n x_i^T \beta + (1/\kappa - 1) \sum_{i=1}^n \log \left\{ 1 - \frac{\kappa(y_i - \tau)}{\exp\{x_i^T \beta\}} \right\}, \quad (1)$$

where the parameters of interest are the regression coefficients $\beta = (\beta_1, \dots, \beta_p)^T$ and the shape parameter κ , i.e., $\theta = (\beta, \kappa)$, and n is the sample size.

We specify the recently proposed Jeffreys' prior by Castellanos and Cabras [16] over the shape parameter κ . Note that Castellanos and Cabras [16] did not consider the Pareto regression in their paper. Hastie et al. [6] discuss that for a linear regression model under the Gaussian prior, also known as the Ridge type prior, and impose L_2 -norm penalty over the likelihood, which improves the efficiency of the linear regression model. We consider the p variate Gaussian distribution, $\boldsymbol{\beta} \sim N_p(\mu_\beta, \Sigma_\beta)$, with mean vector μ_β and dispersion matrix Σ_β .

The log posterior (ignoring the normalizing constant) is

$$\ln \pi(\theta | y, \mathbf{X}) = - \sum_{i=1}^n x_i^T \beta + (1/\kappa - 1) \sum_{i=1}^n \ln \left\{ 1 - \frac{\kappa(y_i - \tau)}{\exp\{x_i^T \beta\}} \right\} \\ - \ln(1 - \kappa) - \frac{1}{2} \ln(1 - 2\kappa) - \ln \pi(\beta),$$

where $\pi(\boldsymbol{\beta})$ is the density of the prior on $\boldsymbol{\beta}$. According to the Bayesian paradigm, the posterior distribution of $\boldsymbol{\theta}$ contains all the information about $\boldsymbol{\theta}$ and the posterior probability is the maximum around the neighborhood of the mode of the posterior distribution under unimodality. The method of maximum a posteriori estimation (MAP) then estimates $\boldsymbol{\theta}$ as the mode of the posterior distribution and is defined as

$$\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}} \pi(\theta | y, X) = \underset{\theta}{\operatorname{argmax}} L(\theta | y, X) \pi(\theta).$$

The denominator of the posterior distribution does not depend on $\boldsymbol{\theta}$ and therefore plays no role in the optimization. In our case, we estimate the posterior modes of our parameters of interest by using the general purpose numerical optimization technique [22]. We used the R 2.9.0 software to implement our analysis, where the direct search Nelder-Mead [22] scheme is implemented in the 'optim' function of the base package of the R 2.9.0 software.

There are different possible choices of prior for the regression coefficients. One choice is a diffuse Gaussian prior (popularly known as 'flat prior') with mean at zero and a large specified variance or 'uniform flat prior'. In spite of the popularity of such priors, one should be cautious when using these priors [23]. It follows that a Bayesian method with 'flat prior' produces results similar to the likelihood method and yields no improvement over the likelihood method. In addition, in many cases (such as 'large p small n ' data or multicollinearity), where likelihood analysis performs poorly, flat prior tail performs in the same way.

One common choice is the Ridge prior for $\boldsymbol{\beta}$, which is a Gaussian distribution with $\mu_\beta = \mathbf{0}$ and $\Sigma_\beta = \sigma_\beta^2 \mathbf{I}_p$, where $\mathbf{I}_p$ is the identity matrix of order p . The posterior distribution for simple linear regression with the Ridge prior is still Gaussian and the posterior mean is the Ridge estimator for the regression coefficients [24]. The Ridge regression method can be viewed as a special case of the Tikhonov regularization [25] of "ill-posed" problems, which automatically addresses inherent non-orthogonality of the data. Here, by the term non-orthogonality, we mean 'multicollinearity' and/or 'large p small n ' data. Though in our data 'multicollinearity' and/or 'large p small n ' is not an issue, these additional advantages of the Ridge prior for the linear regression make it an attractive choice for the Pareto regression.

Another popular choice is Zellner's g -prior [7], i.e., $\boldsymbol{\beta} \sim N_p(0, \sigma^2 g^{-1} (X'X)^{-1})$, where a common choice of g is the sample size, i.e., $g = n$. A nice feature of the g -prior is that it automatically adjusts the units of covariates in the prior. However, when using this prior,

one should also be careful because the data-driven g -prior allows the posterior to process the data twice; once in likelihood and once in prior. This is also known as dual use of the data.

It is important to check the performance of these priors for Pareto regression. That is why, in section 4, we show a simulation study to compare the performance of the Bayesian method with different priors and the existing likelihood method.

4. SIMULATION STUDY

The purpose of this simulation study is to compare the performance of the maximum likelihood approach and the Bayesian methods based on the Ridge prior. We simulate the data from the true model

$$\mathbf{y} \sim GPD(\mathbf{y} | \tau, \sigma = \exp(\mathbf{x}^T \boldsymbol{\beta}), \kappa), \quad (2)$$

where $\kappa = -0.05$, $\tau = 5$ and GPD denotes the generalized Pareto distribution. Such choices of κ and τ are compatible with our main study. Since, for men ‘heavy drinking’ is defined as having 5 or more standard drinks on a particular day, $\tau = 5$ is a natural choice for our simulation study. In the main analysis of the study, the shape parameter κ for both male and female populations was in the range of about 0 to -0.1 (see Tables III and IV) and we chose $\kappa = -0.05$ for our simulation study. Four examples are presented here.

- a. In example 1, we simulated 500 data sets consisting of sample size n with eight predictors. We let $\boldsymbol{\beta} = (3, 1.5, 0, 0, 2, 0, 0, 0)^T$ and $\sigma = 3$. The pairwise correlation between $\mathbf{x}_i$ and $\mathbf{x}_j$ was set to be $\rho(i, j) = 0.5^{|i-j|}$ and $i, j = 1, 2, \dots, p (= 8)$. First we generate the design matrix with the sample of size n , i.e., $\mathbf{X}_{n \times p}$ are drawn from $N_p(\boldsymbol{\mu}, \Sigma)$, where $\Sigma = \sigma \rho^{|i-j|}$. Then given $\mathbf{x}_i$, $i = 1, 2, \dots, n$; we generate $\mathbf{y}_n = \{y_i | i = 1, 2, \dots, n\}$ from (2). We generate $M (= 500)$ such data sets and denote each data set as $D^s = \{(\mathbf{y}_n^s, \mathbf{X}_{n \times p}^s)\}$, where $s = 1, 2, \dots, M (= 500)$ and the true parameter is denoted as $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \kappa)$.
- b. Example 2 is the same as example 1, except that $\boldsymbol{\beta} = (0.85, 0.85, 0, 0, 0.85, 0, 0, 0)^T$.
- c. Example 3 is the same as example 1, except that $\beta_j = 0.85$ for all j .
- d. In example 4 we simulated $M = 50$ data sets, each of sample size n and consisting of 40 predictors. We set

$$\boldsymbol{\beta} = (\underbrace{0, \dots, 0}_{10}, \underbrace{2, \dots, 2}_{10}, \underbrace{0, \dots, 0}_{10}, \underbrace{2, \dots, 2}_{10})$$

and $\sigma = 3$; $\rho(i, j) = 0.75^{|i-j|}$ for all i and j .

We consider the sample size $n = 5, 10, 50, 100$ for all four examples. The $(p + 1)$ parameters, i.e., $\boldsymbol{\theta} = (\boldsymbol{\beta}, \kappa)$, are estimated for each data set D^s by the MAP and the MLE

methods. The estimates of $\boldsymbol{\theta}$, from the s^{th} data set, are denoted as $\hat{\boldsymbol{\theta}}_{MAP}^s$ and $\hat{\boldsymbol{\theta}}_{MLE}^s$ respectively. We estimate the mean square error (MSE) for the s^{th} data set as

$$MSE^s = \frac{(\hat{\boldsymbol{\theta}}^s - \boldsymbol{\theta}_0)^T (\hat{\boldsymbol{\theta}}^s - \boldsymbol{\theta}_0)}{(p+1)}$$

for both MAP and MLE methods and denote those as MSE_{MLE}^s and MSE_{MAP}^s respectively, for $s = 1, 2, \dots, M$.

Comparing two methods

To compare the performance of different methods, two different approaches are available. One is an empirical approach and the other is a decision theory approach. In the empirical approach, the results (e.g., average of estimates, coverage 95% CI over repetition) for each parameter should be compared to analyze the performance of ML and the Bayes methods. Though the empirical method is popular, it is inconvenient for a large simulation study with many parameters in the model. In a large simulation study (as example 4 has 41 parameters in the model), the decision theory approach is more convenient than the empirical approach. In the decision theory approach, we compute the MSE of $\hat{\theta}$ for each method to minimize the expected squared error loss and we define the mean MSE as

$$MSE^{(mean)} = \frac{1}{M} \sum_{s=1}^M MSE^s$$

and median MSE as

$$MSE^{(m)} = median\{MSE^s \mid s=1, 2, \dots, M\}.$$

We prefer to compute the median MSE, instead of the mean MSE for our purpose. Berger [23] discussed that the MSE has a convex tail utility function, which makes the distribution of MSE^s positively skewed. We compute the MSE for each method and if

$$MSE_{MLE}^{(m)} > MSE_{MAP}^{(m)},$$

this indicates that the Bayesian method yields higher efficiency than the likelihood method.

We conclude otherwise if $MSE_{MLE}^{(m)} < MSE_{MAP}^{(m)}$.

Discussion of the simulation study

The results are presented in Table II. The estimated median mean square error of MAP estimates for the Bayesian method are less than the existing MLE method. This indicates that the MAP estimates from the Bayesian method perform better than the standard likelihood approach. In particular, except for a large sample size of 50 and 100 in example 3 and example 4, in all other cases the estimated median MSE of the Bayesian method with the Ridge prior is smallest. Zellner's g -prior yields the smallest MSE for examples 3 and 4 with sample size 50 and 100.

5. ANALYSIS OF EXTREME BEHAVIOR OF PATIENTS DIAGNOSED WITH AD

Since the definition of heavy drinking differs by gender, we present separate analyses for female and male patients for extreme drinking. We fit the Bayesian Pareto regression under the standard likelihood and proposed Bayesian methods. Based on the Bayesian information criterion (BIC) we observed that the model with the Ridge prior fits the data best. Henceforth all of the analyses presented here are based on the Bayesian Pareto regression with the Ridge prior.

5.1. Analysis of Female Patients

We fit the Pareto regression model, using the likelihood method and the Bayesian method with Ridge prior and Zellner's g -prior, for women with AD and present the results in Table III. Based on the BIC, we observed that the model with the Ridge prior fits the data best. Hereafter, we present our analysis based on Pareto regression with the Ridge prior. We found that the day effect $\{-0.0712$ with 95% CI $(-0.0924, -0.0499)\}$ is statistically significant. The negative estimate implies that among women the probability of extreme drinking decreases during the treatment (or study) period. The effect of age, $\{0.6371$ with 95% CI $(0.5860, 0.6820)\}$ is also significant. There was also a significant interaction effect of *GABRA2* genotype with time $\{-0.0401$ with 95% CI $(-0.0641, -0.0162)$ over extreme drinking behavior among women with AD. The negative estimate shows that women with different risk alleles responded to the treatment differently. This phenomenon is shown in Figure 1.a, where women with one or two copies of the high-risk allele showed a better response to treatment and their probability of extreme drinking decreased faster than women with two copies of the low-risk allele. In Figure 1.b, we present the $P(Y > 30 | Y \geq 4)$ along the vertical axis and treatment period along the horizontal axis showing the efficacy of the three different treatments for AD. It is clear from Figure 1.b that among women CBT was the best psychotherapy for the treatment of extreme drinking behavior.

5.2. Analysis of Male Patients

We fit the Pareto regression for men separately and present the results in Table IV. We found that the day effect $\{-0.0420$ with 95% CI $(-0.0512, -0.0329)\}$ was statistically significant. As in women, a negative estimate implies that among men the probability of extreme drinking decreased during the treatment (or study) period. The effect of age $\{0.4788$ with 95% CI $(0.4495, 0.5081)\}$ was also statistically significant. The main effect of *GABRA2* genotype, $\{0.6107$ with 95% CI $(0.5542, 0.6673)\}$ on extreme drinking behavior among men with AD was significant. In Figure 2.a, it appears that the high-risk group responded better to therapy than the low-risk group with the significant interaction effect between *GABRA2* and *Day* with estimate $\{-0.1358$ with 95% CI $(-0.1475, -0.1242)\}$. In Figure 2.b we present the same $P(Y > 30 | Y \geq 5)$ along the vertical axis and study period along the horizontal axis. Here we present the efficacy of the three different psychotherapy treatments for AD. It is clear from Figure 2.b that among men CBT was the worst therapy for reducing extreme drinking.

6. DISCUSSION

In this paper, we present a detailed analysis of extreme drinking behavior among patients with AD. We present a Bayesian Pareto regression with unknown shape parameter to model extreme drinking behavior. We found that the probability of extreme drinking diminished over the treatment period for both genders. Men with one or two copies of the high-risk allele of *GABRA2* had a significantly higher probability of extreme drinking than those in the low-risk group. However, individuals of both genders with the high-risk allele responded to treatment better and their probability of extreme drinking decreased faster than individuals with the low-risk alleles. Because extreme drinking may elicit particular concern due to its potential to result in serious harm, treatment providers may have intervened more intensively with men who drank at extreme levels, resulting in a greater treatment effect among these individuals. Among women, cognitive behavioral therapy was superior to the other two treatments, while among men cognitive behavioral therapy had the poorest effect on decreasing extreme drinking.

We present a Bayesian method to model the data. The Bayesian method produces an improved result with reduced MSE compared to the existing likelihood method. In addition,

the Bayesian method with the Ridge prior automatically takes care of the ‘multicollinearity’ and/or ‘large p small n’ issue, if the data are characterized by these. The empirical results from the simulation study demonstrate the good performance of the proposed Bayesian method. Since $\tau = 5$ (4) was previously established as the definition of ‘heavy drinking’ for men (women), the Pareto model with known τ was the natural choice for our case. An extension to a Pareto regression model with unknown τ would be a much harder problem and will be investigated in the future. In addition, although our methodology is motivated by the regression problems in the analysis of extreme drinking behavior among patients with AD, in the future this method can be easily extended to classification and feature selection problems.

Acknowledgments

Contract/grant sponsor: National Science Foundation; contract/grant number: DMS-0635449

Contract/grant sponsor: National Institute of Health; contract/grant number: P50 AA03510 and K24 AA013736

references

1. Kendler K. Twin studies of psychiatric illness: an update. *Archives of General Psychiatry*. 2001; 58:1005–1014. [PubMed: 11695946]
2. Bauer L, Covault J, Harel O, Das S, Gelernter J, Anton R, Kranzler HR. Variation in *GABRA2* predicts drinking behavior in Project Match subjects. *Alcoholism: Clinical and Experimental Research*. 2007; 31:1780–1787.
3. Helping patients who drink too much: A clinician’s guide. National Institute on Alcohol Abuse and Alcoholism; NIH Publication; 2005.
4. Hastie TJ, Botha JL, Schnitzler CM. Regression with an ordered categorical response. *Statistics in Medicine*. 1989; 8:785–794. [PubMed: 2772438]
5. Davison AC, Smith RL. Models for exceedances over high thresholds. *Journal of Royal Statistical Society, Series B*. 1990; 52:393–442.
6. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer; 2009.
7. Zellner, A. On assessing prior distribution and bayesian regression analysis with *g*-prior distributions. In: Goel, P.; Zellner, A., editors. *In Bayesian Inference and Decision Technique*. Amsterdam: Elsevier Science Publication; 1986.
8. Covault J, Gelernter J, Hesselbrock V, Nellissery M, Kranzler HR. Allelic and haplotypic association of *GABRA2* with alcohol dependence. *Am J Med Genet B Neuropsychiatr Genet*. 2004; 129:104–109. [PubMed: 15274050]
9. Gnedenko B. Sur la distribution limite du terme maximum d’une série aléatoire. *Annals of Mathematics*. 1945; 44:423–453.
10. Pickands J. Statistical inference using extreme order statistics. *Annals of Statistics*. 1975; 3:119–131.
11. Smith RL. Estimating tails of probability distributions. *Annals of Statistics*. 1987; 15:1174–1207.
12. Coles SG, Powell EA. Bayesian methods in extreme value modeling: a review and new developments. *International Statistical Review*. 1996; 64:119–136.
13. Coles SG, Tawn JA. A bayesian analysis of extreme rainfall data. *Applied Statistics*. 1996; 45:463–478.
14. Frigessi A, Haug O, Rue H. A dynamic mixture model for unsupervised tail examination without threshold selection. *Extremes*. 2002; 5:219–235.
15. Behrens CN, Lopes HF, Gamerman D. Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling*. 2004; 4:227–244.
16. Castellanos EM, Cabras S. A default bayesian procedure for the generalized pareto distribution. *Journal of Statistical Planning and Inference*. 2007; 137:473–483.

17. Tsonas EG. Pareto regression: a bayesian analysis. *Communication in Statistics: Theor and Method*. 2003; 32:1213–1225.
18. Robert, C.; Casella, G. *Monte Carlo Statistical Methods*. Springer; 2005.
19. Smith RL. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*. 1985; 72:67–90.
20. Lappalainen J, Krupitsky E, Remizov M, Pchelina S, Taraskina A, Zvartau E, Somberg LK, Covault J, Kranzler HR, Krystal JH, Gelernter J. Association between alcoholism and gamma-amino butyric acid $\alpha 2$ receptor subtype in a Russian population. *Alcoholism: Clinical and Experimental Research*. 2005; 29:493–498.
21. Pierucci-Lagha A, Covault J, Feinn R, Nellissery M, Hernandez-Avila C, Oncken C, Morrow AL, Kranzler HR. *GABRA2* alleles moderate the subjective effects of alcohol, which are attenuated by finasteride. *Neuropsychopharmacology*. 2005; 30:1193–1203.
22. Burmen A, Puhan J, Tuma T. Grid restrained nelder-mead algorithm. *Computational Optimization and Applications*. 2006; 34:359–375.
23. Berger, JO. *Statistical Decision Theory and Bayesian Analysis*. Springer; 1985.
24. Hoerl EA, Kennard WR. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. 1970; 12:55–67.
25. Marroquin J, Mitter S, Poggio T. Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*. 1987; 82:76–89.

Figure 1.

Part a: $P(Y > 30 \mid Y \geq 4)$ for $GABRA2$ genotype group for female patients over the study period and **b:** $P(Y > 30 \mid Y \geq 4)$ for different treatments over the study period

Figure 2.

Part a: $P(Y > 30 \mid Y \geq 5)$ for *GABRA2* genotype for male patients over the study period and

Part b: $P(Y > 30 \mid Y \geq 5)$ for different treatments over the study period

Background characteristics of patients as a function of *GABRA2* genotype and treatment assignment

Table 1

Demographics	High-risk & CBT	High-risk & MET	High-risk & TSF	Low-risk & CBT	Low-risk & MET	Low-risk & TSF
	n=171	n=185	n=191	n=97	n=75	n=93
Age (sd)	41.2 (11.9)	40.6 (11.7)	41.2 (11.4)	41.2 (11.9)	40.0 (10.1)	39.84 (10.9)
Education (sd)	13.3 (1.9)	13.7 (1.9)	13.3 (2.1)	13.3 (2.1)	13.8 (2.2)	13.3 (2.1)
Gender, % female	33.3	24.9	27.2	21.6	25.3	22.6
ASI psychiatric severity score	0.19 (0.20)	0.20 (0.18)	0.23 (0.19)	0.20 (0.19)	0.19 (0.18)	0.22 (0.19)
BDI score	9.4 (7.1)	10.5 (8.2)	11.0 (8.6)	10.9 (9.7)	9.5 (6.8)	11.7 (8.6)

* High-risk: presence of an alcohol dependence associated allele.
Low-risk: absence of an alcohol dependence associated allele.

CBT: Cognitive behavioral therapy
MET: Motivational enhancement therapy
TSF: Twelve-step facilitation

BDI: Beck Depression Inventory (scores vary from 0–63 with higher scores showing greater depression)
ASI: Addiction severity index (scores vary from 0–1.0 with higher scores showing greater psychiatric improvement)

Table II

Median mean-squared errors for the four simulated examples, based on M replications. The numbers in parentheses are the corresponding standard error based on the M mean-squared errors. The underlined values indicate that the method yielded the smallest MSE compared to others.

<i>Results for the following examples</i>				
<i>Method</i>	<i>Example 1</i>	<i>Example 2</i>	<i>Example 3</i>	<i>Example 4</i>
	<i>M = 500</i>	<i>M = 500</i>	<i>M = 500</i>	<i>M = 50</i>
<i>Results for n=5</i>				
MLE	35.086 (3.025)	0.646 (0.061)	8.818 (0.827)	0.136 (0.005)
MAP (default)	25.734 (2.132)	0.499 (0.033)	6.901 (0.412)	0.126 (0.005)
MAP (Ridge)	<u>2.994</u> (0.127)	<u>0.254</u> (0.009)	<u>1.529</u> (0.061)	<u>0.087</u> (0.002)
MAP (Zellner's g)	4.571 (0.061)	0.297 (0.010)	2.637 (0.050)	0.124 (0.004)
<i>Results for n=10</i>				
MLE	29.948 (1.067)	0.443 (0.031)	6.743 (0.207)	0.125 (0.004)
MAP (default)	24.494 (1.021)	0.425 (0.023)	5.707 (0.161)	0.123 (0.003)
MAP (Ridge)	<u>4.272</u> (0.180)	<u>0.179</u> (0.006)	<u>1.909</u> (0.089)	<u>0.119</u> (0.003)
MAP (Zellner's g)	5.204 (0.072)	0.297 (0.010)	3.036 (0.050)	0.121 (0.003)
<i>Results for n=50</i>				
MLE	19.546 (0.489)	0.277 (0.012)	3.626 (0.051)	0.139 (0.004)
MAP (default)	18.441 (0.437)	0.268 (0.011)	3.459 (0.049)	0.138 (0.004)
MAP (Ridge)	<u>4.353</u> (0.168)	<u>0.035</u> (0.002)	3.250 (0.089)	0.131 (0.003)
MAP (Zellner's g)	7.261 (0.159)	0.119 (0.006)	<u>2.853</u> (0.043)	<u>0.127</u> (0.002)
<i>Results for n=100</i>				
MLE	16.998 (0.346)	0.244 (0.009)	3.259 (0.039)	0.135 (0.003)
MAP (default)	16.754 (0.356)	0.241 (0.009)	3.233 (0.041)	0.134 (0.003)
MAP (Ridge)	<u>3.667</u> (0.139)	<u>0.022</u> (0.002)	2.734 (0.059)	0.130 (0.003)
MAP (Zellner's g)	6.332 (0.188)	0.061 (0.002)	<u>2.531</u> (0.054)	<u>0.121</u> (0.002)

Table III

Result with estimates and 95% CI for corresponding regression coefficients of covariates from the model as described in section 3 for extreme drinking of female patients using both the Bayesian and the maximum likelihood methods.

Covariates	Methods	Estimates	95% CI
Intercept	MLE	0.1344	(-0.1097 , 0.3785)
	MAP (Ridge)	0.1485	(-0.0569 , 0.3541)
	MAP (Zellner's g)	-0.2221	(-0.4665 , 0.0221)
Day (t)	MLE	-0.0880	(-0.1109 , -0.0651)
	MAP (Ridge)	-0.0712	(-0.0924 , -0.0499)
	MAP (Zellner's g)	-0.0172	(-0.0401 , 0.0056)
Age	MLE	0.6693	(0.6103 , 0.7282)
	MAP(Ridge)	0.6371	(0.5860 , 0.6820)
	MAP(Zellner's g)	0.6721	(0.6132 , 0.7308)
GABRA2	MLE	0.0136	(-0.1128 , 0.1400)
	MAP (Ridge)	0.0321	(-0.0839 , 0.1481)
	MAP (Zellner's g)	0.2905	(0.1669 , 0.4142)
GABRA2 $\times$ Day	MLE	-0.0324	(-0.0583 , -0.0064)
	MAP (Ridge)	-0.0401	(-0.0641 , -0.0162)
	MAP (Zellner's g)	-0.1029	(-0.1286 , -0.0771)
MET	MLE	-0.1350	(-0.1687 , -0.1017)
	MAP (Ridge)	0.0088	(-0.0405 , 0.0223)
	MAP (Zellner's g)	0.0420	(0.0088 , 0.0752)
CBT	MLE	-0.3180	(-0.3517 , -0.2843)
	MAP (Ridge)	-0.1645	(-0.1966 , -0.1324)
	MAP (Zellner's g)	-0.1472	(-0.1807 , -0.1138)
Shape (κ)	MLE	-0.0661	(-0.0818 , -0.0504)
	MAP (Ridge)	-0.0241	(-0.0381 , -0.0102)
	MAP (Zellner's g)	-0.0595	(-0.0761 , -0.0429)
BIC	MLE	1497.195	
	MAP (Ridge)	1346.615	
	MAP (Zellner's g)	1437.955	

Table IV

Result with estimates and 95% CI for corresponding regression coefficients of covariates from the model as described in section 3 for extreme drinking of male patients using both the Bayesian and the maximum likelihood methods.

Covariates	Methods	Estimates	95% CI
Intercept	MLE	-0.0037	(-0.1302 , 0.1227)
	MAP (Ridge)	0.7036	(0.5851 , 0.8221)
	MAP (Zellner's g)	0.7036	(0.5852 , 0.8220)
Day (t)	MLE	-0.0786	(-0.0882 , -0.0690)
	MAP (Ridge)	-0.0420	(-0.0512 , -0.0329)
	MAP (Zellner's g)	-0.0421	(-0.0512 , -0.0329)
Age	MLE	0.7567	(0.7253 , 0.7881)
	MAP (Ridge)	0.4788	(0.4495 , 0.5081)
	MAP (Zellner's g)	0.4788	(0.4495 , 0.5081)
GABRA2	MLE	0.1150	(0.0563 , 0.1736)
	MAP (Ridge)	0.6107	(0.5542 , 0.6673)
	MAP (Zellner's g)	0.6107	(0.5542 , 0.6673)
GABRA2 $\times$ Day	MLE	-0.0441	(-0.0561 , -0.0320)
	MAP (Ridge)	-0.1358	(-0.1475 , -0.1242)
	MAP (Zellner's g)	-0.1358	(-0.1475 , -0.1242)
MET	MLE	-0.1840	(-0.2018 , -0.1664)
	MAP (Ridge)	-0.0227	(-0.0399 , -0.0055)
	MAP (Zellner's g)	-0.0225	(-0.0397 , -0.0053)
CBT	MLE	0.0359	(0.0181 , 0.0538)
	MAP (Ridge)	0.2812	(0.2644 , 0.2980)
	MAP (Zellner's g)	0.2812	(0.2644 , 0.2980)
Shape (κ)	MLE	-0.0262	(-0.0332 , -0.0192)
	MAP (Ridge)	-0.0051	(-0.0117 , 0.0015)
	MAP (Zellner's g)	-0.0052	(-0.0118 , 0.0016)
BIC	MLE	4465.055	
	MAP (Ridge)	2522.655	
	MAP (Zellner's g)	2524.045	