

Approximate Bayesian Inference for Quantiles

David B. Dunson^{1,*} and Jack A. Taylor²

¹ Biostatistics Branch and ² Epidemiology Branch,
MD A3-03, National Institute of Environmental Health Sciences
P.O. Box 12233, Research Triangle Park, NC 27709

**email:* dunson1@niehs.nih.gov

SUMMARY

Suppose data consist of a random sample from a distribution function F_Y , which is unknown, and that interest focuses on inferences on θ , a vector of quantiles of F_Y . When the likelihood function is not fully specified, a posterior density cannot be calculated and Bayesian inference is difficult. This article considers an approach which relies on a substitution likelihood characterized by a vector of quantiles. Properties of the substitution likelihood are investigated, strategies for prior elicitation are presented, and a general framework is proposed for quantile regression modeling. Posterior computation proceeds via a Metropolis algorithm that utilizes a normal approximation to the posterior. Results from a simulation study are presented, and the methods are illustrated through application to data from a genotoxicity experiment.

KEY WORDS: Comet assay; Nonparametric; Median regression; Order constraints; Prior elicitation; Quantile regression; Single cell electrophoresis; Substitution likelihood.

1. Introduction

A common problem in applications, which has motivated a vast literature in nonparametric methods, is the lack of a known parametric distribution for an outcome variable. For example, in genotoxicity experiments that rely on single-cell gel electrophoresis to assess DNA damage on the level of individual cells, measures of the frequency of DNA strand breaks seldom follow standard parametric distributions, even after transformation (Lovell et al., 1999). Numerous frequentist methods are available which do not require specification of the likelihood function. In contrast, Bayesian nonparametric methods typically require full specification of the likelihood, with parametric assumptions relaxed by choosing a prior distribution for the sampling density.

Although Bayesian nonparametric methods are extremely useful when there is uncertainty about distributional forms (cf., Walker et al., 1999), such approaches are often highly computationally intensive. For this reason, it is appealing to consider simple approximations, which can be used to obtain fairly accurate results quickly and easily. Suppose data consist of a random sample from a distribution function F_Y , which is unknown, and that interest focuses on inferences on θ , a vector of quantiles of F_Y . One possibility is to approximate the likelihood function by a substitution likelihood that depends only on θ , an idea conceptually related to Bayesian empirical likelihood (Lazar, 2003) and to the approach of Chernozhukov and Hong (2003). Lavine (1995) proposed a form for the substitution likelihood, based on a generalization of an approach developed by Jeffreys (1961, §4.4). An alternative approximation was proposed by Boos and Monahan (1986) for a special case.

As motivation, consider the application to studies using single-cell gel electrophoresis (a.k.a., the comet assay) to measure DNA repair kinetics following damage from different environmental carcinogens. Data consist of measures of DNA damage for each cell in a sample, with samples exposed to a variety of experimental conditions after being drawn from cell lines for individuals with different haplotypes. It is desirable to have approaches

for rapidly fitting different models in exploring relationships between the high-dimensional set of haplotypes and DNA repair. Since differences between groups may be more pronounced in the tails of the response distribution, a quantile regression approach is preferable to median or mean-based regression models.

Although quantile regression models are often useful in applications, most of the Bayesian literature in this area has focused on median regression. Walker and Mallick (1999) proposed a median regression accelerated failure time model based on a Polya tree prior for the error distribution. Hanson and Johnson (2002) generalized this approach to smooth out the partitioning effect by using a mixture of Polya trees. Kottas and Gelfand (2001) and Gelfand and Kottas (2003) proposed alternative approaches for modeling of median 0 residuals distributions using mixtures. Yu and Moyeed (2001) proposed the use of the asymmetric Laplace distribution for modeling of error residuals in a linear regression model in order to obtain estimates that minimize the appropriate quantile regression loss function.

Our proposed approach is based on generalizing the substitution likelihood idea of Lavine (1995) to the setting in which a vector of quantiles, including not just the median but also percentiles in the tails, can vary with covariates through a linear regression structure. Our hope is to obtain a computationally simple approach for rapidly implementing approximate Bayesian inferences on differences in quantiles. Section 2 considers properties of the substitution likelihood and proposes an approach for prior elicitation and regression modeling. Section 3 outlines a Metropolis algorithm for posterior computation. Section 4 contains results from a simulation study. Section 5 applies the method to genotoxicity data, and Section 6 discusses the results.

2. Substitution Likelihood

2.1 Notation and Basic Formulation

Suppose that n iid measurements y_1, \dots, y_n are available from a sampling density f_Y , which

is unknown, and let $\theta(p)$ denote the p th quantile of F_Y . Suppose that $\mathbf{p} = (p_1, \dots, p_m)'$ is a known vector satisfying $0 = p_0 < p_1 < \dots < p_m < p_{m+1} = 1$, and let $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)'$ be a vector of unknown quantiles such that $\theta_l = \theta(p_l)$ for $l = 1, \dots, m$. Following Lavine (1995), we replace the likelihood $l(\boldsymbol{\theta}) = \prod_{i=1}^n f_Y(y_i)$ with the substitution likelihood:

$$s(\boldsymbol{\theta}) = \binom{n}{u_1(\boldsymbol{\theta}) \dots u_{m+1}(\boldsymbol{\theta})} \prod_{l=1}^{m+1} \Delta p_l^{u_l(\boldsymbol{\theta})}, \quad (1)$$

where $\theta_0 = -\infty$, $\theta_{m+1} = \infty$, $u_l(\boldsymbol{\theta}) = \sum_{i=1}^n 1_{(\theta_{l-1}, \theta_l]}(y_i)$ for $l = 1, \dots, m+1$, and $\Delta \mathbf{p} = (p_1, p_2 - p_1, \dots, 1 - p_m)'$.

Jeffreys (1961, §4.4) suggests that using a substitution likelihood $s(\boldsymbol{\theta})$ in place of $l(\boldsymbol{\theta})$ yields a valid uncertainty. Although Monahan and Boos (1992) claim that such an approach is invalid, since $s(\boldsymbol{\theta})$ is not a true likelihood function, use of a substitution likelihood based on the quantiles has intuitive appeal. In addition, as shown by Lavine (1995), inferences on $\boldsymbol{\theta}$ based on (1) are asymptotically conservative at the truth in the sense that $s(\boldsymbol{\theta})$ distinguishes between two distributions less well than the log-likelihood in large samples.

2.2 Some Properties

It is of interest to consider some additional properties of substitution likelihood (1). Let $\mathbf{y} = (y_1, \dots, y_n)'$ denote the data vector, and let $y_{[1]}, \dots, y_{[n]}$ denote the corresponding order statistics. As noted in Theorem 1 (proof in Appendix A), the substitution likelihood obtains a maximum at simple empirical estimates of the quantiles. Due to the empirically-based structure of $s(\boldsymbol{\theta})$, which avoids assigning probability at locations in the parameter space where data are not informative, the substitution likelihood surface is in the form of a step function consisting of flat planes with data-dependent jumps.

THEOREM 1. *Suppose $f_Y(y)$ exists and is continuous for all $y \in \mathbb{R}^1$, and hence there are no ties in \mathbf{y} . Let $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_m)'$ with*

$$\hat{\theta}_l = \frac{\max\{y_{[k]} : np_l \geq k\} + \min\{y_{[k]} : np_l \leq k\}}{2}, \quad \text{for } l = 1, \dots, m. \quad (2)$$

Then (i) $s(\hat{\boldsymbol{\theta}})$ is at a maximum of the surface $s(\boldsymbol{\theta})$; (ii) identical values of $s(\boldsymbol{\theta})$ are obtained for $\boldsymbol{\theta}$ within some bounds of $\hat{\boldsymbol{\theta}}$ (shown in Appendix A); and (iii) the surface $s(\boldsymbol{\theta})$ is a step function, consisting of flat regions and discontinuity points.

Bayesian inferences on $\boldsymbol{\theta}$, with $s(\boldsymbol{\theta})$ replacing the likelihood function, are based on the quasi-posterior density $\pi(\boldsymbol{\theta}|\mathbf{y}) \propto s(\boldsymbol{\theta})\pi(\boldsymbol{\theta})$, where $\pi(\boldsymbol{\theta})$ denotes the prior density for $\boldsymbol{\theta}$. Although choosing an informative prior is appealing when outside information is available about the quantiles $\boldsymbol{\theta}$, one may sometimes wish to consider a non-informative or reference prior

$$\pi(\boldsymbol{\theta}) \propto 1(\boldsymbol{\theta} \in \Omega), \quad (3)$$

where Ω denotes the restricted space $\Omega = \{\boldsymbol{\theta} : \theta_1 < \dots < \theta_m\}$. However, as stated in Theorem 2 (proof in Appendix B), the quasi-posterior is improper when the prior density (3) is chosen, and hence inferences are invalid.

THEOREM 2. *The quasi-posterior density $\pi(\boldsymbol{\theta}|\mathbf{y}) \propto s(\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ is improper when an improper prior $\pi(\boldsymbol{\theta}) \propto 1(\boldsymbol{\theta} \in \Omega)$, which is uniform in the restricted space $\Omega = \{\boldsymbol{\theta} : \theta_1 < \dots < \theta_m\}$, is chosen for $\boldsymbol{\theta}$.*

The impropriety result under a uniform prior on Ω is due to the fact that $s(\boldsymbol{\theta})$ is constant as $\theta_1 \rightarrow -\infty$ and $\theta_m \rightarrow \infty$. Essentially, all values of θ_1 below $y_{[1]}$ and all values of θ_m above $y_{[n]}$ are equally likely according to $s(\boldsymbol{\theta})$. To obtain a proper quasi-posterior, one can potentially use the bounded uniform prior

$$\pi(\boldsymbol{\theta}) \propto 1(\boldsymbol{\theta} \in \Omega) 1(\theta_1 \geq c_L) 1(\theta_m \leq c_U), \quad (4)$$

where c_L and c_U are known finite constants, with $c_U > c_L$. However, in some cases, strict bounds on θ_1 and θ_m may not be known a priori. As a generalization of (4), one can choose

$$\pi(\boldsymbol{\theta}) \propto 1(\boldsymbol{\theta} \in \Omega) g(\theta_1, \theta_m), \quad (5)$$

where $g(\theta_1, \theta_m)$ is chosen to satisfy the propriety constraint:

$$\int_{-\infty}^{\infty} \int_{\theta_1}^{\infty} g(\theta_1, \theta_m) (\theta_m - \theta_1)^{m-2} d\theta_m d\theta_1 < \infty. \quad (6)$$

As shown in Appendix C, prior (5) is proper iff (6) holds. Conditional on θ_1 and θ_m , prior (5) allocates probability to $\theta_2, \dots, \theta_{m-1}$ uniformly within Ω .

A useful prior that satisfies (6) (as shown in Appendix C) takes

$$g(\theta_1, \theta_m) = \left\{ e^{-\phi_L(c_L - \theta_1)} \right\}^{1(\theta_1 < c_L)} \left\{ e^{-\phi_U(\theta_1 - c_U)} \right\}^{1(\theta_1 > c_U)} \left\{ e^{-\nu(\theta_m - \theta_1 - d)} \right\}^{1(\theta_m - \theta_1 > d)}, \quad (7)$$

where ϕ_L, ϕ_U, ν are known positive rate parameters, and c_L, c_U , and d are known finite constants with $c_U > c_L$. The resulting prior density is exponentially decreasing for θ_1 outside the interval $[c_L, c_U]$ and for $\theta_m - \theta_1 > d$, and is uniform elsewhere. Strict constraints: $\theta_1 \geq c_L$, $\theta_1 \leq c_U$, and $\theta_m - \theta_1 \leq d$ can be incorporated in the prior by letting $\phi_L = \infty$, $\phi_U = \infty$, and $\nu = \infty$, respectively. One should avoid choosing values for the rate parameters close to 0, a large negative value for c_L , or large positive values for c_U and d , since the resulting prior will be close to improper, resulting in slow-mixing problems in computation.

Alternatively, if prior information is available about the different quantiles $\boldsymbol{\theta}$, one can instead choose a truncated Normal prior,

$$\pi(\boldsymbol{\theta}) \stackrel{d}{=} N_m(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}_0) \quad s.t. \quad \boldsymbol{\theta} \in \Omega, \quad (8)$$

where $\boldsymbol{\theta}_0$ and $\boldsymbol{\Sigma}_0$ are investigator-specified hyperparameters. Truncated normal priors are widely used in order-constrained Bayesian models (cf., Gelfand et al., 1992).

2.3 Quantile Regression and Prior Specification

In many applications, there is interest in relating a vector of covariates $\mathbf{x}_i = (x_{i1}, \dots, x_{iq})' \in \mathcal{X}$ to the quantiles $\boldsymbol{\theta}_i$, where the subscript i denotes that $\boldsymbol{\theta}$ can vary for individuals with different values of \mathbf{x}_i in the sample space \mathcal{X} . To accommodate this case, we propose the

following quantile regression model:

$$\boldsymbol{\theta}_i = \begin{bmatrix} \theta_{i1} \\ \vdots \\ \theta_{im} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}'_1 \\ \vdots \\ \boldsymbol{\alpha}'_m \end{bmatrix} \mathbf{x}_i = \boldsymbol{\alpha} \mathbf{x}_i, \quad (9)$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m)'$ is a matrix of unknown regression coefficients, with the l th row vector $\boldsymbol{\alpha}'_l$ consisting of parameters relating \mathbf{x}_i to the p_l th quantile, θ_{il} . Although we focus on the simple case in which the same predictors apply to each element of $\boldsymbol{\theta}_i$, generalizations to allow different sets of covariates are straightforward.

In order for the elements of $\boldsymbol{\theta}_i$ to be interpretable as quantiles, $\boldsymbol{\theta}_i$ must belong to the order-restricted space Ω , for $i = 1, \dots, n$. This restriction implies that

$$\boldsymbol{\alpha}'_1 \mathbf{x}_i < \boldsymbol{\alpha}'_2 \mathbf{x}_i < \dots < \boldsymbol{\alpha}'_m \mathbf{x}_i, \quad \text{for } i = 1, \dots, n. \quad (10)$$

This constraint can be enforced for values of \mathbf{x} in the sample by choosing a prior for $\boldsymbol{\alpha}$ with support $\{\boldsymbol{\alpha} : \boldsymbol{\theta}_i \in \Omega, i = 1, \dots, n\}$. Alternatively, suppose \mathcal{X} can be expressed as $\mathcal{X}_1 \times \mathcal{X}_2$, where $\mathcal{X}_1 = \{(x_1, \dots, x_{q_1}) : x_l \in \{a_l, a_l + 1, \dots, b_l\} \forall l\}$ is a discrete space and \mathcal{X}_2 is a compact convex subset of \mathbb{R}^{q_2} . For example, the first q_1 predictors could be 0/1 indicators and the remaining q_2 predictors could be bounded continuous variables, with the bounds implied by the application or chosen to include the sample range and values at which predictions may be of interest. For $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$, we can develop a class of priors having the property that the order constraint is satisfied not only for the values of \mathbf{x} in the sample but for all $\mathbf{x} \in \mathcal{X}$.

THEOREM 3. *Let $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$, where $\mathcal{X}_1 = \{\mathbf{x}_1 : x_{1l} \in \{a_l, a_l + 1, \dots, b_l\} \forall l\}$ and \mathcal{X}_2 is a compact convex subset of \mathbb{R}^{q_2} , $\{\mathbf{a}, \tilde{\mathbf{x}}_{11}, \dots, \tilde{\mathbf{x}}_{1,q_1-1}\}$ is a linearly independent subset of \mathcal{X}_1 , and $\{\tilde{\mathbf{x}}_{21}, \dots, \tilde{\mathbf{x}}_{2,q_2+1}\}$ is a set of extreme points of \mathcal{X}_2 . Letting $\widetilde{\mathbf{X}}_1 = [\tilde{\mathbf{x}}_{11}, \dots, \tilde{\mathbf{x}}_{1,q_1-1}]'$, $\widetilde{\mathbf{X}}_2 = [\tilde{\mathbf{x}}_{21}, \dots, \tilde{\mathbf{x}}_{2,q_2+1}]'$,*

$$\widetilde{\mathbf{X}} = \begin{bmatrix} \widetilde{\mathbf{X}}_1 & \mathbf{0}_{(q_1-1) \times q_2} \\ \mathbf{a}' \otimes \mathbf{1}_{(q_2+1) \times 1} & \widetilde{\mathbf{X}}_2 \end{bmatrix}$$

and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{q+1}]' = \widetilde{\mathbf{X}} \boldsymbol{\alpha}'$, $\boldsymbol{\beta}_j \in \Omega$, for $j = 1, \dots, q$, implies $\boldsymbol{\theta} = \boldsymbol{\alpha} \mathbf{x} \in \Omega$ for all $\mathbf{x} \in \mathcal{X}$.

From Theorem 3 (proof in Appendix D), we can choose a prior with appropriately restricted support by specifying a prior for $\boldsymbol{\beta}$ such that $\boldsymbol{\beta}_j \in \Omega$, for $j = 1, \dots, q$. This will induce a prior on $\boldsymbol{\alpha} = \boldsymbol{\beta}'\widetilde{\mathbf{X}}^{-1}$. To specify a vague prior, we can use the form shown in (5), with $\boldsymbol{\theta}$ replaced by $\boldsymbol{\beta}_j$, for $j = 1, \dots, q$. As illustrated in Section 4, $\widetilde{\mathbf{X}}$ can often be chosen so that the row vectors define subpopulations of subjects for which bounds or other information about $\boldsymbol{\theta}$ is available. An informative prior, which may be useful in producing shrinkage estimators and in cases in which prior information is available, is the truncated normal density:

$$\pi(\boldsymbol{\alpha}) \stackrel{d}{=} \mathbf{N}(\boldsymbol{\alpha}_0, \boldsymbol{\Sigma}_0) \quad s.t. \quad \boldsymbol{\beta}_j \in \Omega \quad \forall j.$$

Potentially, one can replace $\boldsymbol{\alpha}$ by $\boldsymbol{\beta}$ in this expression.

3. Posterior Computation

A Markov chain Monte Carlo (MCMC) algorithm can be used for posterior computation (c.f., Casella and Roberts, 1999; Chen, Shao and Ibrahim, 2000 for recent books on MCMC). Due to the non-conjugate form of the quasi-posterior density, we use a Metropolis algorithm, which samples candidates based on a normal approximation to the posterior density of $\boldsymbol{\alpha}$. This normal approximation is also useful as a rough estimate of the posterior.

For simplicity, we focus on the case where the elements of \mathbf{x}_i are indicator variables, and hence the subjects can be grouped into $G < n$ strata. For $s = 1, \dots, G$, let $\widehat{\boldsymbol{\theta}}_s$ denote the estimate for $\boldsymbol{\theta}$ obtained by applying expression (2) to $\mathbf{y}_{(s)}$, the subvector of \mathbf{y} for strata s , and let $\widehat{\boldsymbol{\theta}}_{sb}$ denote the corresponding estimate for the b th bootstrap sample from $\mathbf{y}_{(s)}$, for $b = 1, \dots, B$. A bootstrap estimate of the empirical covariance of $\widehat{\boldsymbol{\theta}}_s$ can be calculated as

$$\widehat{\boldsymbol{\Sigma}}_s = \frac{1}{B} \sum_{b=1}^B (\widehat{\boldsymbol{\theta}}_s - \widehat{\boldsymbol{\theta}}_{sb})(\widehat{\boldsymbol{\theta}}_s - \widehat{\boldsymbol{\theta}}_{sb})'.$$

Letting $\mathbf{x}_{(s)}$ denote the value of \mathbf{x}_i for strata s , our normal approximation to the posterior

of $\boldsymbol{\alpha}$ is $\tilde{\pi}(\boldsymbol{\alpha} | \mathbf{y}) \stackrel{d}{=} \mathcal{N}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}})$, where

$$\begin{aligned}\hat{\boldsymbol{\alpha}} &= \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}} \left(\boldsymbol{\Sigma}_0^{-1} \boldsymbol{\alpha}_0 + \sum_{s=1}^G \mathbf{X}'_{(s)} \hat{\boldsymbol{\Sigma}}_s^{-1} \hat{\boldsymbol{\theta}}_s \right) \quad \text{and} \\ \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}} &= \left(\boldsymbol{\Sigma}_0^{-1} + \sum_{s=1}^G \mathbf{X}'_{(s)} \hat{\boldsymbol{\Sigma}}_s^{-1} \mathbf{X}_{(s)} \right)^{-1}\end{aligned}$$

with $\boldsymbol{\alpha} = (\boldsymbol{\alpha}'_1, \dots, \boldsymbol{\alpha}'_m)'$ and $\mathbf{X}_{(s)} = \mathbf{x}'_{(s)} \otimes \mathbf{I}_{m \times m}$. In the case where a vague prior is chosen for $\boldsymbol{\alpha}$, we let $\boldsymbol{\Sigma}_0^{-1} = \mathbf{0}$. For finite n , samples from this normal approximation need not satisfy $\boldsymbol{\theta}_i \in \Omega$, but as $n \rightarrow \infty$ (within each strata) violations of the constraint will become increasingly unlikely.

After setting $\boldsymbol{\alpha}^{(0)} = \hat{\boldsymbol{\alpha}}$, our Metropolis algorithm alternates between the steps:

1. Sample a candidate $\tilde{\boldsymbol{\alpha}} \sim \mathcal{N}_{mq}(\boldsymbol{\alpha}^{(t-1)}, \kappa \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}})$.
2. Calculate $\tilde{\boldsymbol{\beta}} = \tilde{\mathbf{X}} \tilde{\boldsymbol{\alpha}}'$. If $\tilde{\boldsymbol{\beta}}_j \notin \Omega$, for any j , then reject the candidate and let $\boldsymbol{\alpha}^{(t)} = \boldsymbol{\alpha}^{(t-1)}$.

Otherwise, if $\tilde{\boldsymbol{\beta}}_j \in \Omega$, for $j = 1, \dots, q$, let $\boldsymbol{\alpha}^{(t)} = \tilde{\boldsymbol{\alpha}}$ with probability

$$\min \left\{ 1, \frac{s(\tilde{\boldsymbol{\alpha}}) \pi(\tilde{\boldsymbol{\alpha}})}{s(\boldsymbol{\alpha}^{(t-1)}) \pi(\boldsymbol{\alpha}^{(t-1)})} \right\},$$

and otherwise let $\boldsymbol{\alpha}^{(t)} = \boldsymbol{\alpha}^{(t-1)}$. Note that $\pi(\boldsymbol{\alpha})$ in this expression is unnormalized.

Steps 1-2 are repeated for $t = 1, \dots, T$. The parameter κ is a tuning constant which can be chosen during an initial run to generate an acceptance probability of approximately 0.25 (c.f., Roberts, Gelman and Gilks, 1997). This algorithm is easy to implement, and has outperformed alternative random walk Metropolis algorithms we have considered, which did not incorporate the normal approximation.

4. Simulation Study

We evaluate the performance of the proposed substitution likelihood approach through a simulation study. In particular, we simulated data for a variety of sample sizes (25, 50, 100, 200) and distributional shapes, applying the proposed approach to each simulated data set. As

a flexible class of distributions with support on the real line, we chose mixtures of normals, weighting each of five components equally but allowing the component means and standard deviations to vary under three cases:

1. Standard Normal: $\boldsymbol{\mu} = (0, 0, 0, 0, 0)'$, $\boldsymbol{\sigma} = (1, 1, 1, 1, 1)'$
2. Unimodal, Moderate Positive Skew:

$$\boldsymbol{\mu} = (-0.6, -0.5, -0.4, 0.5, 1.0)', \boldsymbol{\sigma} = (0.4, 0.5, 0.6, 0.95, 1.15)'$$

3. Bi-modal, High Positive Skew:

$$\boldsymbol{\mu} = (-.5, -.5, -.5, .75, 1.5)', \boldsymbol{\sigma} = c(0.25, 0.25, 0.25, .5, 1)'$$

These densities (shown in Figure 1) are chosen to have mean 0 and variance 1 (approximately).

For each case and $n = 25, 50, 100, 200$, we simulated 100 data sets and analyzed each of these data sets using the proposed MCMC algorithm, collecting 4,000 iterations after a 1,000 iteration burn-in, with the empirical quantiles chosen as starting values. We focus on the simple case in which there are no predictors, and interest is in estimating $\boldsymbol{\theta} = (\theta_1, \dots, \theta_5)'$, the quantiles corresponding to $\mathbf{p} = (0.1, 0.25, 0.5, 0.75, 0.9)'$. In each case, we choose $\pi(\boldsymbol{\theta}) \propto 1(\boldsymbol{\theta} \in \Omega)N_5(\mathbf{0}, \mathbf{I}_{5 \times 5})$ as a prior for the quantiles. Based on the collected draws from the quasi-posterior, we estimated posterior means and 95% credible intervals for each of the simulated data sets. Tables 1-3 presents results, including the mean of the empirical quantiles across the simulations, the mean of the posterior mean estimates, 95% empirical intervals calculated from percentiles of the posterior means across the simulations, and coverage probabilities for the 95% credible intervals.

The mean of the quasi-posterior was close to unbiased in each case, except in the tails ($p_l = 0.1, 0.9$) in very small samples ($n = 25$) in which there was a slight tendency to estimate values that are too extreme. This is likely an artifact of the high degree of skewness

in the quasi-posterior densities in such cases, and the bias rapidly declines as sample size increases. The 95% credible intervals tended to be slightly conservative for very small samples, but were close to the nominal level overall. We conclude that the proposed substitution likelihood approach has excellent frequentist operating characteristics for a wide variety of true distributional shapes, even for small to moderate sample sizes.

5. Genotoxicity Example

5.1 Data and Priors

Data are drawn from a genotoxicity experiment assessing the effect of oxidative damage on the frequency of DNA strand breaks. Samples of cells exposed to different levels of hydrogen peroxide (0, 5, 20, 50, 100 μM H_2O_2) were prepared for use in the comet assay. After electrophoresis, the DNA from the nucleus of cells with a high frequency of DNA strand breaks exhibit a comet-type shape, with the nucleus forming the ball-like head and the cut DNA strands the tail. Cells with a low frequency tend to maintain an approximately spherical shape with less tail. The goal of the experiment was to evaluate the sensitivity of the comet assay in detecting genotoxic effects of hydrogen peroxide, a known genotoxic agent.

Data consist of the % DNA in the comet tail for 100 cells in each of the five dose groups. We have limited prior information about the distribution of % DNA in the comet tail, since our laboratory has just started using the comet assay under these conditions. Thus, we initially rely on a non-informative prior that allocates probability uniformly under the condition that the quantiles are all within [0%,100%].

Let y_i denote the % DNA in the comet tail for the i th cell divided by 100, so that $y_i \in [0, 1]$, for $i = 1, \dots, n$, and let $\mathbf{x}_i = (x_{1i}, x_{2i})'$ with $x_{1i} = 1$ and x_{2i} denoting the log-transformed dose of hydrogen peroxide (obtained by taking the natural logarithm of dose + 1). We let $\mathbf{p} = (0.1, 0.25, 0.5, 0.75, 0.9)'$, motivated by our prior belief that differences

between the treated and untreated groups may depend on the quantile of the distribution. Although we could potentially choose additional quantiles (e.g., 0.05 or 0.95), estimation of the additional parameters seems unlikely to improve sensitivity to detecting a dose response.

Following the strategy of Section 2.3, we have $\widetilde{\mathbf{X}} = [(1, 0)', (1, \log 101)']'$, where $\mathcal{X}_1 \equiv \{1\}$ and $\mathcal{X}_2 = \{x : x \in [0, \log 101]\}$, so that we restrict attention to the range of doses studied in the experiment. It follows that $\boldsymbol{\beta}_1 = (\alpha_{11}, \dots, \alpha_{51})'$, $\boldsymbol{\beta}_2 = (\alpha_{11} + \alpha_{12} \log 101, \dots, \alpha_{51} + \alpha_{52} \log 101)'$, and we can induce a prior on $\boldsymbol{\alpha}$ with appropriately restricted support by letting

$$\pi(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) \propto 1(\boldsymbol{\beta}_1 \in \Omega) 1(\boldsymbol{\beta}_2 \in \Omega) 1(\beta_{11} \geq 0) 1(\beta_{15} \leq 1) 1(\beta_{21} \geq 0) 1(\beta_{25} \leq 1),$$

which is equivalent to choosing uniform priors for the quantile-specific intercepts and slopes subject to the constraint that the quantiles fall in the $[0, 1]$ interval and are appropriately ordered.

Histograms and empirical quantiles of the distribution for % DNA in the comet tail are plotted in Figure 2 for cells in the different dose groups. The proportion of DNA in the comet tail (and hence the frequency of DNA strand breaks) appears to increase stochastically as dose of hydrogen peroxide increases. However, H_2O_2 exposure does not effect the quantiles uniformly; the largest dose effects occur in the upper quantiles of the distribution, corresponding to those cells with the highest frequency of DNA strand breaks.

5.2 Analysis and Results

We used the Metropolis algorithm described in Section 3 to obtain samples from the quasi-posterior distribution of $\boldsymbol{\alpha}$. We ran three chains for 100,000 iterations, discarding the first 5,000 iterations from each chain as a burn-in. There was no evidence of lack of convergence based on examination of trace plots and comparisons of the different chains. In addition, autocorrelation was low and sampling was reasonably efficient, with similar estimates obtained based on the full 285,000 iterates and on only the first 5,000 iterates from a single chain. We collected a large number of draws in order to better verify convergence and to obtain precise

estimates of marginal posterior densities, with minimal Monte Carlo error. In addition, given that the algorithm has only a few simple steps, long chains can be implemented quickly and easily.

Our primary interest is in inference on the slope parameters $\alpha_{12}, \dots, \alpha_{52}$ measuring changes with dose in the 0.1, 0.25, 0.5, 0.75, and 0.9 quantiles, respectively, of the distributions of % DNA in the comet tail. Estimates of the marginal posterior densities of these parameters are plotted in Figure 3. The solid lines represent the substitution likelihood-based estimates obtained using the Metropolis algorithm, while the dotted lines are the estimates based on the normal approximation. To assess robustness to choice of \mathbf{p} and m in estimating particular quantiles, we also present estimates for sensitivity analysis in which we set $\mathbf{p} = (0.1, 0.175, 0.25, 0.375, 0.5, 0.625, 0.75, 0.825, 0.9)$ ($m = 9$) and $\mathbf{p} = (0.05, 0.1, 0.15, 0.2, 0.25, 0.375, 0.5, 0.625, 0.75, 0.8, 0.85, 0.9, 0.95)$ ($m = 13$).

Interestingly, each of the estimated densities are quite similar, though the normal approximation has noticeably lower variance than the substitution-likelihood based estimates for α_{12} and α_{22} . The higher variance for the substitution likelihood-based estimates is consistent with the theoretical results of Lavine (1995) and with the slight conservatism of the credible intervals observed in the simulation study. Analyses with different values of m produced similar estimates of the posterior densities, so adding additional “nuisance” quantiles between quantiles of interest appears to have little effect. We also repeated the analysis using a truncated $N_{10}(\mathbf{0}, \mathbf{I})$ prior for $\boldsymbol{\alpha}$, and the resulting estimates are indistinguishable from those presented in Figure 3.

There is clear evidence of an increase in the frequency of DNA strand breaks with increasing dose of hydrogen peroxide, with $\widehat{\Pr}(\alpha_{l2} > 0 | \text{data}) > 0.99$ for $l = 1, \dots, 5$. In addition, there is clear evidence of an increasing trend in the direction of larger slopes for

the higher quantiles, with

$$\widehat{\Pr}\left\{\sum_{l=1}^5 p_l(a_{l2} - \bar{\alpha}_{+2}) > 0\right\} > 0.99.$$

This pattern is illustrated in Figure 4, which shows the empirical quantiles and fitted dose response curves, along with 95% pointwise credible intervals. From this plot, it is clear that the linear quantile regression model provides a good fit to the data, but a median regression model would be insufficient.

6. Discussion

This article has presented a simple and easy-to-apply alternative to fully Bayesian non- and semiparametric methods for quantile regression analysis. The method relies on the substitution likelihood idea of Lavine (1995), which we have extended to a general regression setting. A potentially useful feature of the approach is that it allows regression on more than one quantile simultaneously, while existing Bayesian methods for quantile regression focus on a single quantile (typically, the median). Hence, we can accommodate cases in which the distributional shape changes across values of the predictors. Such changes in shape are common in genotoxicity applications and other biomedical studies. An interesting area for future research is the development of fully Bayesian nonparametric methods that capture such changes.

We have presented some theoretical results on the substitution likelihood, prior choice, and resulting quasi-posterior density. Computation can proceed by a straightforward Metropolis algorithm that relies on a bootstrap approximation to the posterior density. Because analyses can be conducted rapidly, this approach should be particularly useful for exploratory analyses and in settings in which it is not feasible to implement a fully Bayesian nonparametric approach. The computational speed allowed us to conduct a simulation study of the frequentist operating characteristics of the approach for a variety of sample sizes and distributional shapes. The excellent performance in terms of low estimation bias and appro-

priate coverage rates of credible intervals suggests that the approach is a valid alternative to existing Bayesian and frequentist methods.

ACKNOWLEDGEMENTS

The authors thank Shyamal Peddada and Zhen Chen for their critical reading of the manuscript. Thanks also to the Associate Editor and two referees for their helpful comments.

REFERENCES

- Boos, D.D. and Monahan, J.F. (1986). Bootstrap methods using prior information. *Biometrika* **73**, 77-83.
- Chen, M.-H., Shao, Q.-M., Ibrahim, J.G. (2000). *Monte Carlo Methods in Bayesian Computation*. New York: Springer-Verlag.
- Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation. *Journal of Econometrics* **115**, 293-346.
- Gelfand, A.E. and Kottas, A. (2003). Bayesian semiparametric regression for median residual life. *Scandinavian Journal of Statistics* **30**, 651-665.
- Gelfand, A.E., Smith, A.F.M., and Lee, T.M. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association* **87**, 523-532.
- Hanson, T. and Johnson, W.O. (2002). Modeling regression error with a mixture of Polya trees. *Journal of the American Statistical Association* **97**, 1020-1033.
- Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97-109.

- Jeffreys, H. (1961). *Theory of Probability*, 3rd ed. Oxford: Claredon.
- Kottas, A. and Gelfand, A.E. (2001). Bayesian semiparametric median regression modeling. *Journal of the American Statistical Association* **96**, 1458-1468.
- Lavine, M. (1995). On an approximate likelihood for quantiles. *Biometrika* **82**, 220-222.
- Lazar, N.A. (2003). Bayesian empirical likelihood. *Biometrika* **90**, 319-326.
- Lovell, D.P., Thomas, G., and Dubow, R. (1999). Issues related to the experimental design and subsequent statistical analysis of in vivo and in vitro comet studies. *Teratogenesis, Carcinogenesis, and Mutagenesis* **19**, 109-119.
- Monahan, J.F. and Boos, D.D. (1992). Proper likelihoods for Bayesian analysis. *Biometrika* **79**, 271-278.
- Robert, C.P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. New York: Springer-Verlag.
- Roberts, A.W. and Varberg, D.E. (1973). *Convex Functions*. New York-London: Academic Press.
- Roberts, G.O., Gelman, A., and Gilks, W.R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability* **7**, 110-120.
- Walker, S.G., Damien, P., Laud, P.W. and Smith, A.F.M. (1999). Bayesian nonparametric inference for random distributions and related functions (with discussion). *J. R. Statist. B*, **61**, 485-527.
- Walker, S. and Mallick, B.K. (1999). A Bayesian semiparametric accelerated failure time model. *Biometrics* **55**, 477-483.

Yu, K. and Moyeed, R.A. (2001). Bayesian quantile regression. *Statistics & Probability Letters* **54**, 437-447.

APPENDIX A

Proof of Theorem 1

Let $\hat{\boldsymbol{\theta}}_\delta$ be shorthand for $\hat{\boldsymbol{\theta}} + \boldsymbol{\delta} = (\hat{\boldsymbol{\theta}}_1 + \delta_1, \dots, \hat{\boldsymbol{\theta}}_m + \delta_m)'$ and let $\boldsymbol{\delta}_l$ denote the $m \times 1$ vector containing δ in the l th position and 0s elsewhere. If

$$\frac{s(\hat{\boldsymbol{\theta}})}{s(\hat{\boldsymbol{\theta}}_{\delta_l})} \geq 1, \quad \text{for } l = 1, \dots, m \text{ and } \forall \delta : \hat{\boldsymbol{\theta}}_{\delta_l} \in \Omega, \quad (11)$$

then $s(\hat{\boldsymbol{\theta}})$ is at a maximum of the surface $s(\boldsymbol{\theta})$. It follows from expression (1) that

$$\frac{s(\hat{\boldsymbol{\theta}})}{s(\hat{\boldsymbol{\theta}}_{\delta_l})} = \left[n! \prod_{l=1}^{m+1} \frac{\Delta \mathbf{p}_l^{u_l(\hat{\boldsymbol{\theta}})}}{u_l(\hat{\boldsymbol{\theta}})!} \right] / \left[n! \prod_{l=1}^{m+1} \frac{\Delta \mathbf{p}_l^{u_l(\hat{\boldsymbol{\theta}}_{\delta_l})}}{u_l(\hat{\boldsymbol{\theta}}_{\delta_l})!} \right].$$

From the definition of $u_l(\boldsymbol{\theta})$, $u_h(\hat{\boldsymbol{\theta}})$ and $u_h(\hat{\boldsymbol{\theta}}_{\delta_l})$ are equivalent unless $h \in \{l, l+1\}$. Therefore, the ratio simplifies to

$$\frac{s(\hat{\boldsymbol{\theta}})}{s(\hat{\boldsymbol{\theta}}_{\delta_l})} = \frac{\Delta \mathbf{p}_l^{u_l(\hat{\boldsymbol{\theta}})} u_l(\hat{\boldsymbol{\theta}}_{\delta_l})! \Delta \mathbf{p}_{l+1}^{u_{l+1}(\hat{\boldsymbol{\theta}})} u_{l+1}(\hat{\boldsymbol{\theta}}_{\delta_l})!}{u_l(\hat{\boldsymbol{\theta}})! \Delta \mathbf{p}_l^{u_l(\hat{\boldsymbol{\theta}}_{\delta_l})} u_{l+1}(\hat{\boldsymbol{\theta}})! \Delta \mathbf{p}_{l+1}^{u_{l+1}(\hat{\boldsymbol{\theta}}_{\delta_l})}},$$

where $u_l(\hat{\boldsymbol{\theta}}) = \sum 1_{(\hat{\theta}_{l-1}, \hat{\theta}_l]}(y_i)$, $u_{l+1}(\hat{\boldsymbol{\theta}}) = \sum 1_{(\hat{\theta}_l, \hat{\theta}_{l+1}]}(y_i)$, $u_l(\hat{\boldsymbol{\theta}}_{\delta_l}) = \sum 1_{(\hat{\theta}_{l-1}, \hat{\theta}_l + \delta]}(y_i)$, and $u_{l+1}(\hat{\boldsymbol{\theta}}_{\delta_l}) = \sum 1_{(\hat{\theta}_l + \delta, \hat{\theta}_{l+1}]}(y_i)$. It is clear that for δ sufficiently small, the set membership of the elements of \mathbf{y} does not change (i.e., $u_l(\hat{\boldsymbol{\theta}}) = u_l(\hat{\boldsymbol{\theta}}_{\delta_l})$ and $u_{l+1}(\hat{\boldsymbol{\theta}}) = u_{l+1}(\hat{\boldsymbol{\theta}}_{\delta_l})$), implying that $s(\hat{\boldsymbol{\theta}}) = s(\hat{\boldsymbol{\theta}}_{\delta_l})$. Formally, we have $s(\hat{\boldsymbol{\theta}}) = s(\hat{\boldsymbol{\theta}}_{\delta_l})$ for

$$\frac{\max \{y_{[k]} : np_l \geq k\} - \min \{y_{[k]} : np_l \leq k\}}{2} < \delta < \frac{\min \{y_{[k]} : np_l \leq k\} - \max \{y_{[k]} : np_l \geq k\}}{2}.$$

For δ equal to the upper bound of this interval, a single observation is removed from u_{l+1} and added to u_l , resulting in the identities:

$$u_l(\hat{\boldsymbol{\theta}}) = u_l(\hat{\boldsymbol{\theta}}_{\delta_l}) - 1 \quad \text{and} \quad u_{l+1}(\hat{\boldsymbol{\theta}}) = u_{l+1}(\hat{\boldsymbol{\theta}}_{\delta_l}) + 1.$$

Under these identities, the ratio reduces to

$$\frac{s(\hat{\boldsymbol{\theta}})}{s(\hat{\boldsymbol{\theta}}_{\delta_l})} = \left(\frac{u_l(\hat{\boldsymbol{\theta}}) + 1}{p_l - p_{l-1}} \right) \left(\frac{p_{l+1} - p_l}{u_{l+1}(\hat{\boldsymbol{\theta}})} \right). \quad (12)$$

From the definition of $u_h(\hat{\boldsymbol{\theta}})$, it is clear that $u_h(\hat{\boldsymbol{\theta}})/(p_h - p_{h-1})$ is either equal to 1 or strictly less than 1, while $\{u_h(\hat{\boldsymbol{\theta}}) + 1\}/(p_h - p_{h-1})$ is strictly greater than 1. It follows that ratio (12) is greater than 1. A similar result is obtained by taking δ equal to any value outside the above interval (under the restriction that $\hat{\boldsymbol{\theta}}_{\delta_l} \in \Omega$). It follows that $\hat{\boldsymbol{\theta}}$ is at a maximum of the surface $s(\boldsymbol{\theta})$. Points (ii) and (iii) of Theorem 1 are also clear from the above argument.

APPENDIX B

Proof of Theorem 2

The posterior distribution of $\boldsymbol{\theta}$ is proper iff the marginal density of the data $\int_{\Omega} s(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}$ is integrable. To demonstrate impropriety under a uniform prior on Ω , it suffices to show that

$$\int_{\Omega} s(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\theta_2} s(\boldsymbol{\theta}) d\theta_1 \cdots d\theta_m \not\leq \infty.$$

We focus on the integral $\int_{-\infty}^{\theta_2} s(\boldsymbol{\theta}) d\theta_1$, holding $\theta_2, \dots, \theta_m$ constant at finite values satisfying $y_{[1]} < \theta_2 < \dots < \theta_m$. We can express this integral as

$$\int_{-\infty}^{\theta_2} s(\boldsymbol{\theta}) d\theta_1 = \int_{-\infty}^{y_{[1]}} s(\boldsymbol{\theta}) d\theta_1 + \int_{y_{[1]}}^{\theta_2} s(\boldsymbol{\theta}) d\theta_1. \quad (13)$$

Letting $A = \sum_{i=1}^n 1_{(-\infty, \theta_2]}(y_i)$ denote the number of observations less than or equal to θ_2 , we have $u_{i1}(\theta) = 0$ and $u_{i2}(\theta) = A$ for all $\theta_1 < y_{[1]}$. Thus, for $\theta_2, \dots, \theta_m$ held constant, $s(\theta)$ has a constant positive value of

$$n! \frac{\Delta \mathbf{p}_2^A}{A!} \prod_{l=3}^{m+1} \frac{\Delta \mathbf{p}_l^{u_l(\boldsymbol{\theta})}}{u_l(\boldsymbol{\theta})!} > 0 \quad \text{for all } \theta_1 \in (-\infty, y_{[1]}]. \quad (14)$$

Since $s(\boldsymbol{\theta})$ takes a constant positive value over an infinite width interval, the integral in (13) is infinite, and the posterior distribution of $\boldsymbol{\theta}$ is improper.

APPENDIX C

Proof of Propriety of the Prior

To demonstrate propriety of the prior it suffices to show that

$$\int_{\Omega} \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int_{-\infty}^{\infty} \int_{-\infty}^{\theta_m} \cdots \int_{-\infty}^{\theta_3} \int_{-\infty}^{\theta_2} \pi(\boldsymbol{\theta}) d\theta_1 \cdots d\theta_m < \infty.$$

Substituting expression (5) for $\pi(\boldsymbol{\theta})$ and rearranging the order of integration, we have

$$\int_{-\infty}^{\infty} \int_{\theta_1}^{\infty} \left\{ \int_{\theta_1}^{\theta_m} \int_{\theta_2}^{\theta_m} \cdots \int_{\theta_{m-2}}^{\theta_m} d\theta_{m-1} \cdots d\theta_2 \right\} g(\theta_1, \theta_m) d\theta_m d\theta_1.$$

The integral in parentheses equals $(\theta_m - \theta_1)^{m-2}/(m-2)!$, and hence the above integral is proportional to

$$\int_{-\infty}^{\infty} \int_{\theta_1}^{\infty} g(\theta_1, \theta_m) (\theta_m - \theta_1)^{m-2} d\theta_m d\theta_1,$$

which is $< \infty$ under (6).

Under (7), $g(\theta_1, \theta_m)$ can be expressed as $g_1(\theta_1) \times g_2(\theta_m^*)$, where $\theta_m^* = \theta_m - \theta_1$. Reparameterizing in terms of θ_1 and θ_m^* , expression (6) equals

$$\int_{-\infty}^{\infty} g_1(\theta_1) \left\{ \int_0^{\infty} g_2(\theta_m^*) (\theta_m^*)^{m-2} d\theta_m^* \right\} d\theta_1. \quad (15)$$

The integral in parentheses equals

$$\begin{aligned} & \int_0^{\infty} \left\{ e^{-\nu(\theta_m^*-d)} \right\}^{1(\theta_m^*>d)} (\theta_m^*)^{m-2} d\theta_m^* \\ &= \int_0^d (\theta_m^*)^{m-2} d\theta_m^* + \int_d^{\infty} e^{-\nu(\theta_m^*-d)} (\theta_m^*)^{m-2} d\theta_m^* \\ &= \text{constant} < \infty. \end{aligned}$$

Thus, integral (15) is proportional to a finite constant \times

$$\begin{aligned} \int_{-\infty}^{\infty} g_1(\theta_1) d\theta_1 &= \int_{-\infty}^{c_L} e^{-\phi_L(c_L-\theta_1)} d\theta_1 + \int_{c_L}^{c_U} d\theta_1 + \int_{c_U}^{\infty} e^{-\phi_U(\theta_1-c_U)} d\theta_1 \\ &= \phi_L^{-1} + (c_L - c_U) + \phi_U^{-1} < \infty. \end{aligned}$$

It follows directly that the prior is proper.

APPENDIX D

Proof of Theorem 3

Let $\mathbf{x} = (\mathbf{x}'_1, \mathbf{x}'_2)'$ represent an arbitrary point in \mathcal{X} , with $\mathbf{x}_1 \in \mathcal{X}_1$ and $\mathbf{x}_2 \in \mathcal{X}_2$. From the definition of $\tilde{\mathbf{X}}_1$, all $\mathbf{x}_1 \in \mathcal{X}_1$ can be expressed as $\mathbf{x}_1 = \mathbf{a}_1 + w_{11}\tilde{\mathbf{x}}_{11} + \dots + w_{1,q_1-1}\tilde{\mathbf{x}}_{1,q_1-1}$ for some vector of non-negative integers $\mathbf{w}_1 = (w_{11}, \dots, w_{1,q_1-1})'$. In addition, from Roberts and Varberg (1973), all $\mathbf{x}_2 \in \mathcal{X}_2$ can be expressed as a convex combination of the extreme points $\mathbf{x}_2 = w_{21}\tilde{\mathbf{x}}_{21} + \dots + w_{2,q_2+1}\tilde{\mathbf{x}}_{2,q_2+1}$, with $\mathbf{w}_2 = (w_{21}, \dots, w_{2,q_2+1})'$ non-negative constants summing to one. It follows that the elements of $\boldsymbol{\theta} = \boldsymbol{\alpha}\mathbf{x}$ can be expressed as:

$$\begin{aligned}
\theta_l &= \alpha_{l1}(a_{11} + w_{11}\tilde{x}_{111} + \dots + w_{1,q_1-1}\tilde{x}_{1,q_1-1,1}) + \dots + \alpha_{l,q_1}(a_{1,q_1} + w_{11}\tilde{x}_{11,q_1} + \dots \\
&\quad + w_{1,q_1-1}\tilde{x}_{1,q_1-1,q_1}) + \alpha_{l,q_1+1}(w_{21}\tilde{x}_{211} + \dots + w_{2,q_2+1}\tilde{x}_{2,q_2+1,1}) + \dots + \\
&\quad \alpha_{l,q}(w_{21}\tilde{x}_{21,q_2} + \dots + w_{2,q_2+1}\tilde{x}_{2,q_2+1,q_2}), \\
&= \boldsymbol{\alpha}'_{l1}\mathbf{a} + w_{11}\boldsymbol{\alpha}'_{l1}\tilde{\mathbf{x}}_{11} + \dots + w_{1,q}\boldsymbol{\alpha}'_{l1}\tilde{\mathbf{x}}_{1,q_1} + w_{21}\boldsymbol{\alpha}'_{l2}\tilde{\mathbf{x}}_{21} + \dots + w_{2,q_2+1}\boldsymbol{\alpha}'_{l2}\tilde{\mathbf{x}}_{2,q_2+1} \\
&= \boldsymbol{\alpha}'_{l1}\mathbf{a} + w_{11}\beta_{1l} + \dots + w_{1,q_1-1}\beta_{q_1-1,l} + w_{21}(\beta_{q_1,l} - \boldsymbol{\alpha}'_{l1}\mathbf{a}) + \dots + w_{2,q_2+1}(\beta_{ql} - \boldsymbol{\alpha}'_{l1}\mathbf{a}) \\
&= w_{11}\beta_{1l} + \dots + w_{1,q_1-1}\beta_{q_1-1,l} + w_{21}\beta_{q_1,l} + \dots + w_{2,q_2+1}\beta_{ql}, \tag{16}
\end{aligned}$$

for $l = 1, \dots, m$, with $\boldsymbol{\alpha}_l = (\boldsymbol{\alpha}'_{l1}, \boldsymbol{\alpha}'_{l2})'$, $\boldsymbol{\alpha}_{l1} \sim q_1 \times 1$, $\boldsymbol{\alpha}_{l2} \sim q_2 \times 1$, and $\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jm})'$. Since $\boldsymbol{\beta}_j \in \Omega$ implies $\beta_{j1} < \dots < \beta_{jm}$, for $j = 1, \dots, q$, each of the elements in the weighted sum (16) increase as l increases. Hence, since the weights are all non-negative, it follows directly that $\theta_1 < \theta_2 < \dots < \theta_m$ as required.

Table 1

Results from simulation study for data generated from the standard normal (Case 1)

Case	n	p	true ^a	Emp ^b	Mean ^c	95% CI	Cover ^d
1	25	0.10	-1.28	-1.20	-1.34	[-1.67,-0.89]	0.98
1	25	0.25	-0.67	-0.67	-0.7	[-1.09,-0.31]	0.99
1	25	0.50	0.00	-0.02	-0.02	[-0.38,0.27]	0.99
1	25	0.75	0.67	0.64	0.66	[0.33,0.95]	0.98
1	25	0.90	1.28	1.17	1.31	[0.95,1.66]	0.98
1	50	0.10	-1.28	-1.26	-1.33	[-1.58,-1.02]	0.98
1	50	0.25	-0.67	-0.67	-0.69	[-0.98,-0.47]	0.98
1	50	0.50	0.00	-0.01	-0.01	[-0.31,0.25]	0.96
1	50	0.75	0.67	0.64	0.65	[0.27,0.94]	0.97
1	50	0.90	1.28	1.22	1.29	[0.97,1.56]	0.99
1	100	0.10	-1.28	-1.27	-1.30	[-1.57,-1.05]	0.97
1	100	0.25	-0.67	-0.66	-0.67	[-0.92,-0.46]	0.97
1	100	0.50	0.00	0.00	0.00	[-0.26,0.21]	0.95
1	100	0.75	0.67	0.66	0.67	[0.47,0.89]	0.99
1	100	0.90	1.28	1.24	1.28	[0.97,1.61]	0.94
1	200	0.10	-1.28	-1.27	-1.28	[-1.51,-1.07]	0.98
1	200	0.25	-0.67	-0.65	-0.66	[-0.83,-0.49]	0.94
1	200	0.50	0.00	0.01	0.01	[-0.16,0.16]	0.94
1	200	0.75	0.67	0.68	0.69	[0.52,0.86]	0.96
1	200	0.90	1.28	1.28	1.30	[1.14,1.51]	0.98

^a true quantiles^b average of empirical quantiles^c average of estimated quasi-posterior means^d coverage of 95% credible intervals

Table 2

Results from simulation study for data generated from a moderately positively-skewed
unimodal mixture of normals (Case 2)

Case	n	p	true ^a	Emp ^b	Mean ^c	95% CI	Cover ^d
2	25	0.10	-1.05	-1.00	-1.24	[-1.48,-0.94]	0.97
2	25	0.25	-0.69	-0.66	-0.72	[-0.93,-0.47]	0.99
2	25	0.50	-0.23	-0.21	-0.18	[-0.44,0.22]	0.98
2	25	0.75	0.51	0.53	0.55	[0.14,1.07]	0.97
2	25	0.90	1.46	1.36	1.37	[0.91,1.86]	0.99
2	50	0.10	-1.05	-1.00	-1.10	[-1.34,-0.85]	0.95
2	50	0.25	-0.69	-0.68	-0.70	[-0.89,-0.46]	0.97
2	50	0.50	-0.23	-0.24	-0.22	[-0.46,0.08]	0.97
2	50	0.75	0.51	0.49	0.52	[0.20,0.93]	0.98
2	50	0.90	1.46	1.37	1.38	[0.95,1.90]	0.97
2	100	0.10	-1.05	-1.04	-1.07	[-1.24,-0.91]	0.97
2	100	0.25	-0.69	-0.69	-0.70	[-0.83,-0.54]	0.96
2	100	0.50	-0.23	-0.24	-0.23	[-0.4,-0.05]	0.98
2	100	0.75	0.51	0.51	0.52	[0.24,0.89]	0.95
2	100	0.90	1.46	1.43	1.43	[1.05,1.82]	0.99
2	200	0.10	-1.05	-1.03	-1.05	[-1.14,-0.95]	0.98
2	200	0.25	-0.69	-0.69	-0.69	[-0.79,-0.58]	0.95
2	200	0.50	-0.23	-0.24	-0.23	[-0.37,-0.07]	0.94
2	200	0.75	0.51	0.49	0.51	[0.23,0.82]	0.91
2	200	0.90	1.46	1.42	1.44	[1.14,1.71]	0.97

^a true quantiles

^b average of empirical quantiles

^c average of estimated quasi-posterior means

^d coverage of 95% credible intervals

Table 3

Results from simulation study for data generated from a highly positively-skewed bimodal mixture of normals (Case 3)

Case	n	p	true ^a	Emp ^b	Mean ^c	95% CI	Cover ^d
3	25	0.10	-0.75	-0.73	-1.03	[-1.54,-0.72]	0.90
3	25	0.25	-0.56	-0.55	-0.61	[-0.77,-0.43]	0.97
3	25	0.50	-0.28	-0.23	-0.18	[-0.45,0.21]	0.98
3	25	0.75	0.77	0.69	0.63	[0.10,1.17]	0.96
3	25	0.90	1.60	1.49	1.48	[0.96,2.03]	1.00
3	50	0.10	-0.75	-0.74	-0.84	[-1.36,-0.67]	0.95
3	50	0.25	-0.56	-0.55	-0.56	[-0.68,-0.44]	0.97
3	50	0.50	-0.28	-0.25	-0.19	[-0.43,0.14]	0.93
3	50	0.75	0.77	0.69	0.68	[0.14,1.14]	0.92
3	50	0.90	1.60	1.51	1.53	[1.05,1.99]	0.99
3	100	0.10	-0.75	-0.73	-0.76	[-0.87,-0.67]	0.92
3	100	0.25	-0.56	-0.56	-0.56	[-0.65,-0.48]	0.96
3	100	0.50	-0.28	-0.27	-0.23	[-0.39,0.04]	0.93
3	100	0.75	0.77	0.73	0.72	[0.29,1.10]	0.88
3	100	0.90	1.60	1.58	1.58	[1.21,1.95]	0.99
3	200	0.10	-0.75	-0.74	-0.76	[-0.81,-0.68]	0.99
3	200	0.25	-0.56	-0.56	-0.56	[-0.6,-0.51]	0.97
3	200	0.50	-0.28	-0.27	-0.26	[-0.35,-0.15]	0.96
3	200	0.75	0.77	0.74	0.72	[0.43,0.97]	0.96
3	200	0.90	1.60	1.59	1.60	[1.37,1.88]	0.98

^a true quantiles

^b average of empirical quantiles

^c average of estimated quasi-posterior means

^d coverage of 95% credible intervals

Figure 1. Probability density functions for the different mixtures of normals used in the simulation study.

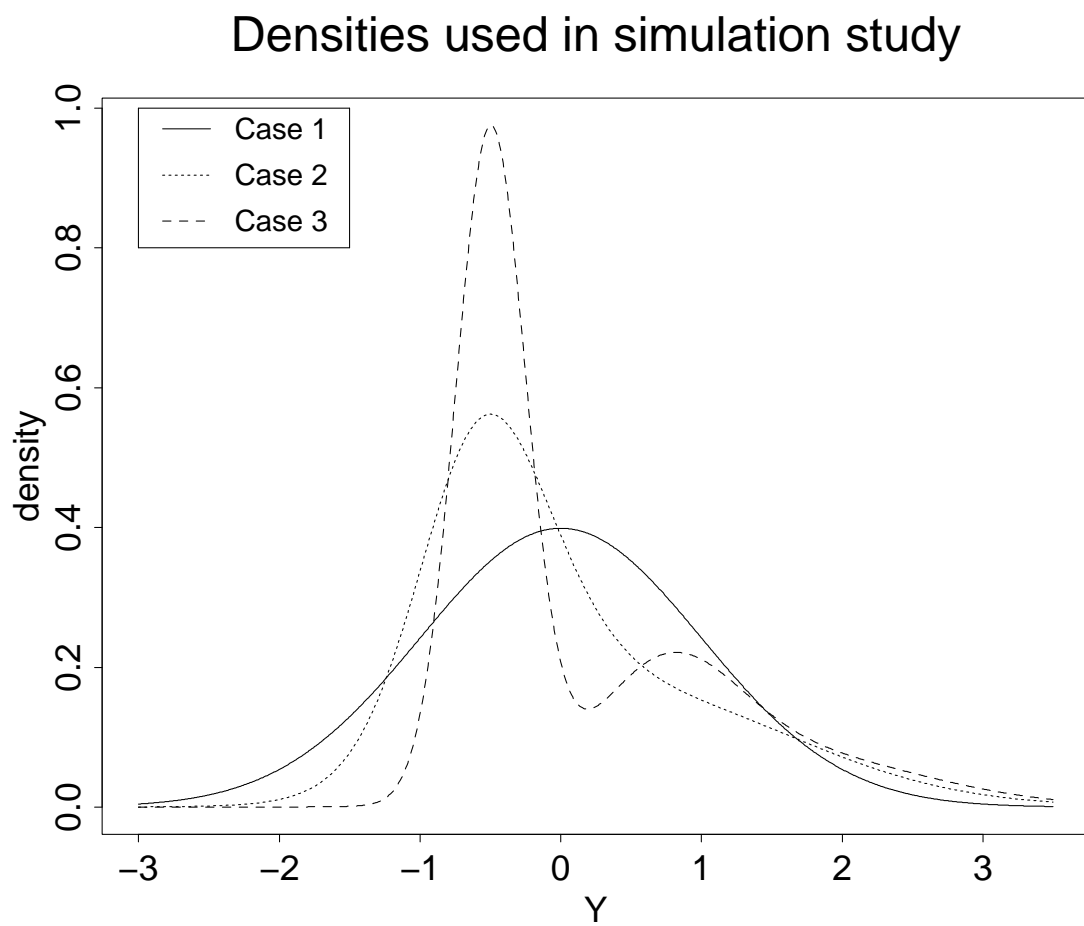


Figure 2. Histograms for the % DNA in the comet tail (divided by 100) for cells in each of the dose groups (0, 5, 20, 50, 100 μ M hydrogen peroxide). The vertical lines represent the 0.10, 0.25, 0.5, 0.75, and 0.9 quantiles of the empirical distribution.

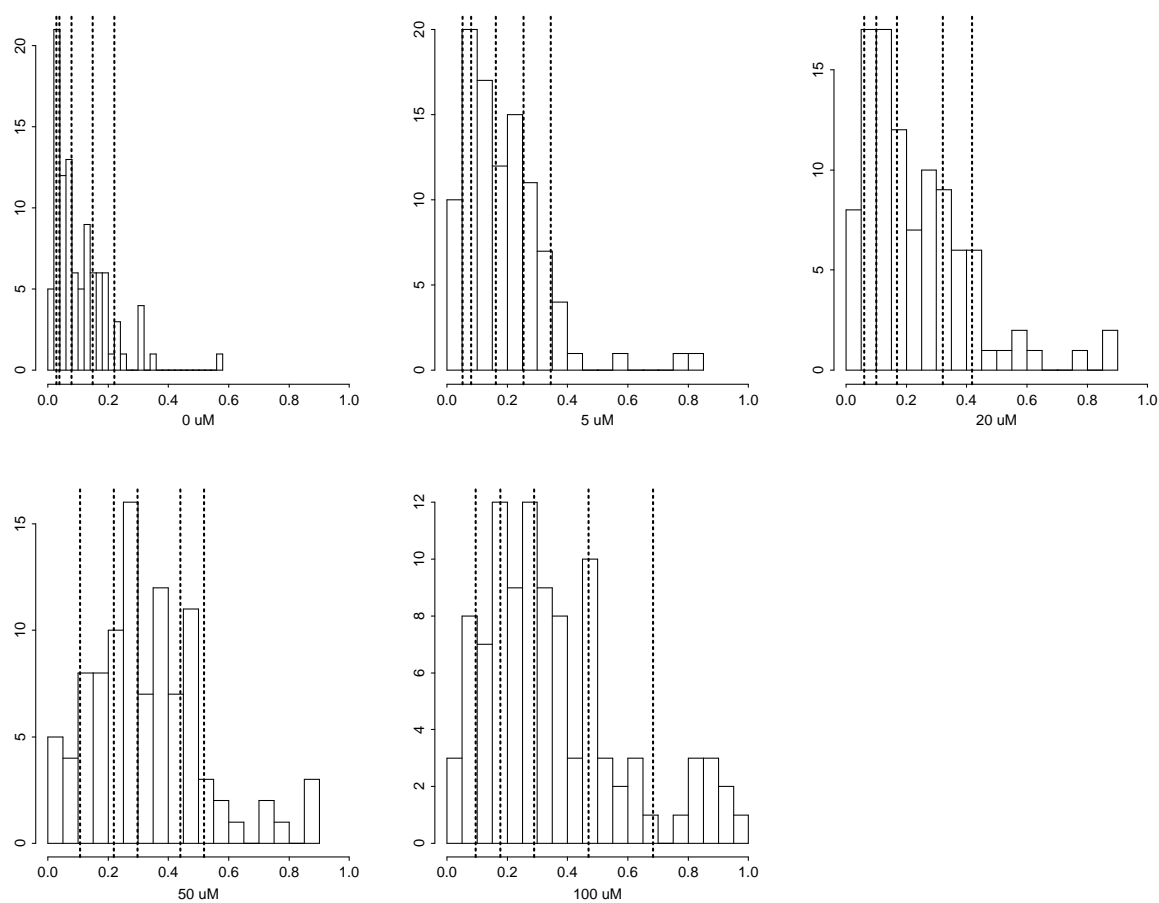


Figure 3. Estimated quasi-posterior densities for $\alpha_{12}, \alpha_{22}, \alpha_{32}, \alpha_{42}, \alpha_{52}$, the slope parameters measuring changes with dose in quantiles of the distribution of % DNA in the comet tail (divided by 100). The solid lines are estimates based on MCMC, while the dashed lines are estimates based on the normal approximation.

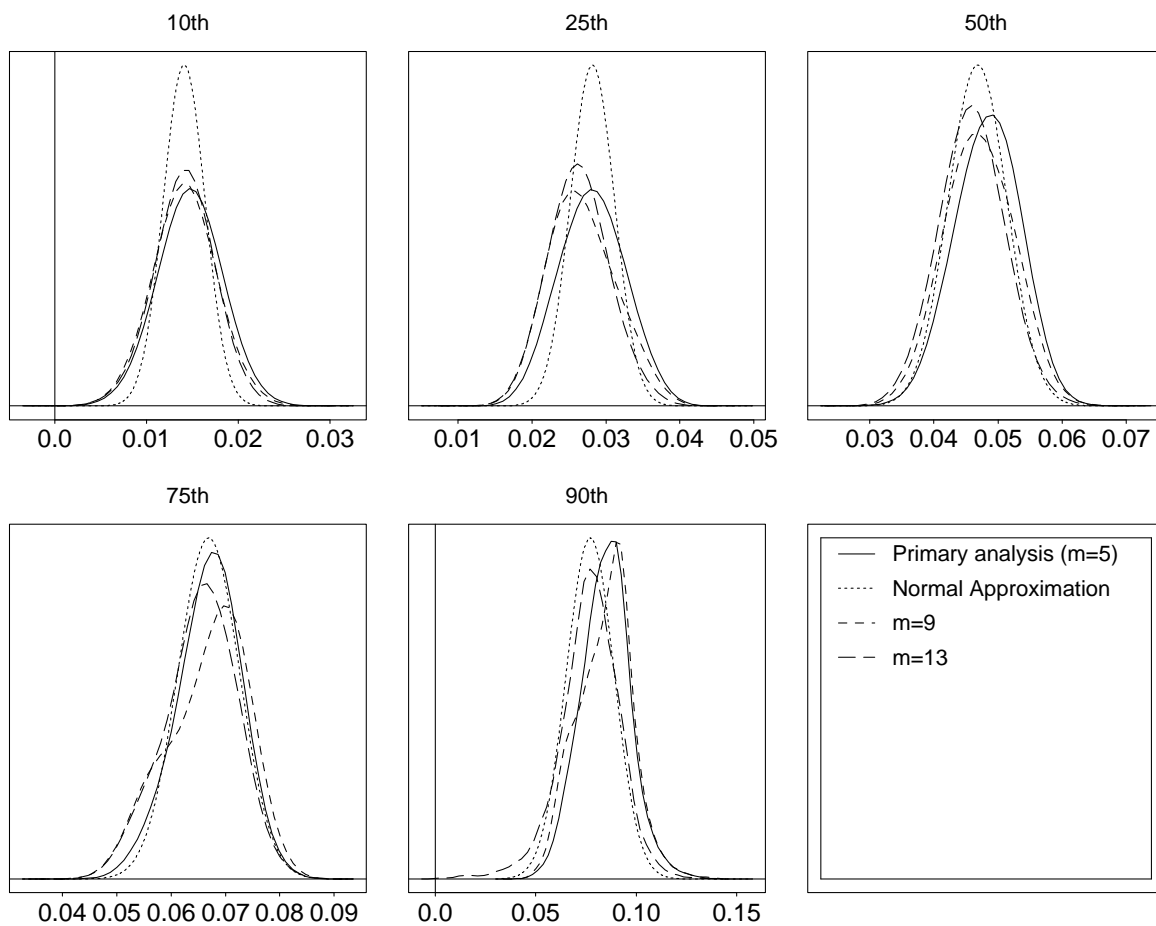


Figure 4. Empirical quantiles of the % DNA in the comet tail and fitted curves for each dose group. The solid line is the quasi-posterior mean, and the dashed lines are 95% pointwise credible intervals.

