BAYESIAN SEMI-PARAMETRIC ANALYSIS

OF DEVELOPMENTAL TOXICOLOGY DATA

Francesca Dominici

Department of Biostatistics, Johns Hopkins University

Giovanni Parmigiani

Departments of Oncology and Biostatistics, Johns Hopkins University

Summary

Modeling of developmental toxicity studies often requires simple parametric analyses of the dose-response relationship between exposure and probability of a birth defect, but poses challenges because of non-standard distributions of birth defects for a fixed level of exposure. This paper is motivated by two such experiments, in which the distribution of the outcome variable is challenging to both the standard logistic model with Binomial response, and its parametric multi-stage elaborations.

We approach our analysis using a Bayesian semi-parametric model that we tailored specifically to developmental toxicology studies. It combines parametric dose-response relationships with a flexible non-parametric specification of the distribution of the response, obtained via a Product of Dirichlet Process Mixtures approach (PDPM). Our formulation achieves three goals: 1) the distribution of the response is modeled in a general way; 2) the degree to which the distribution of the response adapts non-parametrically to the observations is driven by the data; 3) the marginal posterior distribution of the parameters of interest is available in closed form. The logistic regression model, as well as many of its extensions, such as the beta-binomial model and finite mixture models, are special cases.

In the context of the two motivating examples and a simulated example, we provide model comparisons; illustrate over-dispersion diagnostics that can assist model specification; show how to derive the posterior distributions of the effective dose parameters, and the predictive distribution of the response; and discuss sensitivity of the results to the choice of prior distribution.

1 Introduction

Developmental toxicity studies investigate the relationship between the exposure to a potentially toxic compound, and the frequency of birth defects. In the most common designs, pregnant laboratory animals (or dams) are exposed to varying doses of the compound. After they have given birth, the number of birth defects of interest among the offspring (or pups) is recorded. Inference focuses on the presence and magnitude of a dose effect, and on the so-called effective doses —defined either as the dose level at which the probability of malformation reaches a certain value, or that at which it is increased by a certain factor compared to the background rate. A simple and common approach to the analysis of data arising in developmental toxicology experiments is the logistic regression model with binomial response, which considers all dams to be identical, and all pups exposed to the same dose to be independent.

It is well documented that actual data frequently display evidence of departure from the logistic-binomial model (Catalano and Ryan, 1994). One source of difficulties is the heterogeneity of the reaction of different dams to the compound studied. For example, dams may vary in a) whether or not they are susceptible to the exposure; b) the extent to which they are susceptible if they are; or both. Other litter-specific events, such as those that may be induced by reabsorbings of implanted pups, can sum to these factors, inducing additional correlations among pups from the same litter. These correlations can be both positive and negative (e.g. when there is competition for limited nutrients). When these factors combine, the distribution of the birth defect counts at each dose will be far from a binomial, and challenging to any parametric approach. Inaccurate parametric models can seriously affect estimates of interest (Liang and McCullagh, 1993).

This paper is motivated by the analysis of two challenging developmental toxicity studies: one regarding the effects of diethylhexalphthalate, or DEHP, (Tyl et al., 1983) and the other regarding the effects of 2,4,5–Trichlorophenoxiacetic, or 2,4,5–T, (Holson et al., 1991). Datasets are displayed in Figure 1, and discussed in more detail in Section 3. Both studies present hard-

1

to-model response distributions, displaying combinations of zero inflation, n-inflation (that is an excessive number of dams with birth defects in n out of n pups,) overdispersion, and kurtosis, possibly as the result of multiple sources of heterogeneity. In addition, the extent of departures from the binomial model can vary significantly with the dose.

We approach our analysis using a Bayesian semi-parametric model. The small number of dose levels used in these experiments, and the interest in inverse problems such as finding effective doses, naturally lead to parsimonious parametric dose-response relationships between the exposure and the probability of a defect. On the other hand, the complexity of patterns in the distribution of birth defect counts strongly argues in favor of a flexible non-parametric specification of the distribution of the response. We achieve both goals by an approach based on Products of Dirichlet Process Mixtures (Cifarelli and Regazzini, 1978; Carota and Parmigiani, 1997).

Our approach consists of a "parametric backbone", comprising a parametric dose-response relationship and a mean distribution function for the birth defect counts; and a flexible specification for the distribution of the response, allowing departures from the backbone towards the empirical distribution. The mean distribution provides structure if the data are sparse, but may be progressively overruled as evidence accrues. An unknown dispersion parameter controls the extent to which the model is adapting non-parametrically to the data. In discrete models, the observations generally provide information about this dispersion parameter, so that the degree to which the model departs from the backbone to fit features of the empirical data is itself data driven. This flexibility can be important in capturing the results of the combined action of several biological mechanisms on the response distribution, as described earlier.

While nonparametric modeling can effectively accommodate for any pattern in the distribution of the response, including overdispersion as well as underdispersion, a natural comparison arises with models addressing heterogeneity and the resulting overdispersion. A classic approach is the beta-binomial model (Williams, 1975; Williams, 1982), assuming that each dam has a specific

probability of birth defects, and describing the variation in these probabilities by a beta distribution. This model is simple to fit and interpretable, but the resulting marginal distribution of birth defects may not be adequately flexible. The beta-binomial model exemplifies a general approach to correlated teratological data, in which each dam is assigned a dam-specific parameter. Classical implementations include random–effects and mixed–effects GLM's (Cox, 1983; Prentice, 1986; Lefkopoulou et al., 1989; Rosner, 1989), latent variable models, and finite mixture models for zero inflation (Lambert, 1992). Related Bayesian approaches are based on hierarchical models, in which the distribution of the dam-specific parameters may in turn be parametric (Wong and Mason, 1985; Zeger and Karim, 1991; Albert and Chib, 1993) or based on more flexible specifications (Müller and Rosner, 1997; Mukhopadhyay and Gelfand, 1997).

In these approaches, it can be challenging to model distributions of random effects, mixture components, or other latent variables, while accurately reflecing the biological variation. In practice, many of the parametric assumptions made in both Bayesian and standard analysis are dictated by mathematical convenience. Exceptions (Müller and Rosner, 1997; Mukhopadhyay and Gelfand, 1997) tend to require a substantial computational burden. Quasi-likelihood and generalized estimating equations (GEE) have provided simpler alternatives for accounting for correlation in binary data, without explicitly modeling higher stage distributions of dam-specific effects (Williams, 1982; Liang and Zeger, 1986; Lipsitz et al., 1991; Catalano and Ryan, 1992; Bowman and George, 1995; George and Wu, 1997). However, GEE procedures often rely on asymptotic approximations, while teratologic experiments involve relatively small samples.

The goal of the methodology described here is to achieve simplicity and flexibility comparable to those of GEE approaches, while maintaining all the advantages of a full probabilistic model specification, with associated exact Bayesian analysis. Our approach encompasses the marginal distributions of birth defect counts obtained under many common random effect models— and does not require specifying a model for the sources of heterogeneity. A result discussed in

Section 2.2 provides a closed form for the marginal likelihood of the parameters of interest, which simplifies inference and computing substantially. It is also straightforward to obtain full probabilistic inference on the unknown distribution of the response. This includes all the quantiles, which via the parametric backbone, vary with the dose. Therefore this modeling technique can also be used for exact small-sample inference in quantile regression with correlated binary outcomes.

2 Modeling

2.1 Semi–parametric dose–response model

We will consider an experiment with D groups: one control group and $D-1$ groups receiving an active dose of the compound being studies. M^d pregnant dams are exposed to dose d. Dam j at dose d has a litter of size N_j^d live pups. A common endpoint for the evaluation of developmental toxicity effects is the number y_j^d of pups with a certain birth defects of interest. We will denote by $f_j^d(y)$, with $0 \leq y \leq N_j^d$, the probability distribution functions (pdf) of y_j^d. If all dams were identical in their response to the compound, and all pups at dose d were independent, then we could represent all the variability in the data by choosing f_j^d to be a binomial pdf with parameter, say θ^d, independent of j, and sample size N_j^d. In general, this assumption is too restrictive, and greater flexibility in the distribution of the response is desirable.

Here, we wish to entertain all possible pdf's for y_j^d, by treating the whole f_j^d as an unknown parameter. The set of all possible values for f_j^d is the N_j^d-dimensional simplex. A natural and flexible distribution on the simplex is the Dirichlet distribution, taking the form $p(f|A, f_0) \propto \prod_{y=0}^{N} f(y)^{Af_0(y)-1}$. We omitted the dependence on d and j for notational simplicity. We will use the notation $\mathcal{D}(A, f_0)$ to denote the Dirichlet distribution. In this parameterization, f_0 is the mean of the random pdf f, while A can be interpreted as a precision parameter, controlling the amount of variation of f around the mean f_0.

The Dirichlet distribution is the natural conjugate distribution for multinomial sampling

4

(Bernardo and Smith, 1994). More specifically, after observing a sample with counts $k_0, \ldots, k_N$ in the $N+1$ categories $y = 0, \ldots, y = N$, the posterior distribution of f conditional on the sample and on hyperparameters f_0 and A is still a Dirichlet with posterior precision $A + \sum_{y=0}^{N} k_y$, and posterior mean $wf_0(y) + (1 - w)k_y$, where $w = A \left(A + \sum_{y=0}^{N} k_y \right)^{-1}$ denotes the prediction weight. Because the support of the distribution of y_j^d depends on the litter size N_j^d, we need to consider each combination of dose and litter size separately, and we specify separate, although related, Dirichlet distribution for each combination. To provide the necessary structure to model the dose-response relationship, we assume that the mean of the random vector f_j^d is element-wise equal to a Binomial pdf $f_0(\cdot \mid \theta^d, N_j^d)$. In our application, the θ^d's will follow a standard logistic link, namely $\mathrm{logit}(\theta^d) = \beta_0 + \beta_1 d$. Interest usually focuses on the β's, which retain their usual interpretations: β_0 is the log–odds of malformation at baseline, and β_1 is the change in the log–odds of malformation associated with a unit change in the dose.

The dispersion of f_j^d around its mean $f_0(\cdot \mid \theta^d, N_j^d)$ is controlled by a dose-specific precision parameters A^d. Large values of A^d favor the parametric backbone, while small values favor departures. In our application we either assume a common A_0 for all doses, or assume that each A^d is different. If desired, it is possible to model the A^d as a function of the dose, as illustrated in Carota and Parmigiani (1997) . In summary the model specification for Section 3 is:

$$
\begin{aligned}
y_j^d \mid f_j^d &\sim f_j^d & j = 1, \ldots, M^d \quad d = 1, \ldots, D \\
F_j^d \mid A^d, \theta^d &\sim \mathcal{D}\{A^d, f_0(\theta^d, N_j^d)\} & j = 1, \ldots, M^d \quad d = 1, \ldots, D \\
\mathrm{logit}(\theta^d) &= \beta_0 + \beta_1 d \quad d = 1, \ldots, D
\end{aligned}
$$

Because a change in the form of the mean pdf often alters the interpretation of the linear model parameters, and because enough flexibility is gained via the Dirichlet elaboration, for interpretability we favor the choice of the binomial as the parametric backbone.

Specification (2.1) is an instance of a Product of Dirichlet Process Mixtures (PDPM) (Cifarelli and Regazzini, 1978; Carota and Parmigiani, 1997). The random cdf F is said to have a Dirichlet

Process distribution (Ferguson, 1973) with parameter $\boldsymbol{\alpha}$, if for every $p \leq H + 1$, and every partition $(0, h_1], \dots, (h_{p-1}, H]$ of $\{0, \dots, H\}$ in p subsets: $F(h_1), \dots, F(H) - F(h_{p-1}) \sim$ $\mathcal{D} \left(\sum_{h=0}^{h_1} \alpha_h, \dots, \sum_{h=h_{p-1}}^{H} \alpha_h \right)$, where $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_H)$ is a fixed vector of non-negative weights and $\mathcal{D}$ is the Dirichlet distribution. A DPM is a Dirichlet process in which some features of the vector $\boldsymbol{\alpha}$ are unknown, and are assigned a further distribution, as we do here. Approaches modeling the mean of a Dirichlet process mixture as a function of predictors have precedents in the pioneering work of Cifarelli, Muliere and Scarsini (1981) for normal linear models.

2.2 Marginal Inference

Inference can be carried out by drawing a sample from the posterior distribution $p(\boldsymbol{\beta}, \boldsymbol{A}, \mathbf{f} \mid$ $\mathbf{y}^1, \dots, \mathbf{y}^D)$, where $\mathbf{y}^d = (y_1^d, \dots, y_{M^d}^d)$, $\boldsymbol{A} = (A^1, \dots, A^d)$, and $\mathbf{f}$ denotes the ensemble of the f_j^d, for $j = 1, \dots, M^d$, $d = 1, \dots, D$. A key simplifying feature of the formulation considered here is that the joint posterior distribution of $\boldsymbol{\beta}, \boldsymbol{A}$ and $\mathbf{f}$ can be factored into the marginal posterior distribution of $(\boldsymbol{\beta}, \boldsymbol{A})$ and the posterior distribution of $\mathbf{f}$ given $(\boldsymbol{\beta}, \boldsymbol{A})$, both of which are available analytically. Thus, we can simulate a draw from the joint posterior by 1) drawing $\boldsymbol{\beta}$ and $\boldsymbol{A}$ from the marginal posterior distribution; and 2) drawing the pdf functions f_j^d from their conditional posterior distribution given $(\boldsymbol{\beta}, \boldsymbol{A})$. Step 2 is simple: conditionally on $(\boldsymbol{\beta}, \boldsymbol{A})$ each f_j^d is, as seen earlier, a Dirichlet distribution:

$$ f_j^d \mid \boldsymbol{\beta}, \boldsymbol{A}, y_j^d, N_j^d \sim \mathcal{D} \left\{ A^d + \sum_{j=1}^{M^d} \hat{f}_j^d, \frac{A^d}{A^d + \sum_{j=1}^{M^d} \hat{f}_j^d} f_0(\theta^d, N_j^d) + \frac{\hat{f}_j^d}{A^d + \sum_{j=1}^{M^d} \hat{f}_j^d} \hat{f}_j^d \right\}. \quad (1) $$

In (1), $\hat{f}_j^d$ is the empirical frequency of y_j^d among all the dams having litter-size N_j^d at dose d.

Generating f_j^d from (1) is straightforward. Step 1 is based on adapting to this context results of Antoniak (1974) and Carota and Parmigiani (1997), who show how to marginalize the distribution of data generated by a discrete PDPM with respect to the unknown pdf, for a fixed $\boldsymbol{\alpha}$. Let N_l^d, $l = 1, \dots L^d$ be the unique litter sizes at the dose d, let $\tilde{\mathbf{y}}_l^d = (\tilde{y}_{l1}^d, \dots, \tilde{y}_{lr^d}^d)$ be the vector of the r^d unique observations corresponding to the litter size N_l^d, and let m_l^d be the

number of dams having litter size N_l^d. We assume that the generic element $\tilde{y}_{li}^d$ occurs n_{li}^d times. The marginal posterior distribution of $\boldsymbol{\beta}$ and $\boldsymbol{A}$ is then proportional to:

$$\pi(\boldsymbol{\beta}, \boldsymbol{A}) \prod_{d=1}^{D} \prod_{l=1}^{L^d} \{A_{(m_l^d)}^d\}^{-1} \prod_{i=1}^{r^d} A^d p_0(\tilde{y}_{li}^d \mid \theta^d, N_l^d) \{A^d p_0(\tilde{y}_{li}^d \mid \theta^d, N_l^d) + 1\}_{(n_{li}^d - 1)} \tag{2}$$

where $x_{(n)} = x(x+1) \dots (x+n-1)$, and $p_0(\tilde{y}_{li}^d \mid \theta^d, N_l^d)$ is the binomial probability mass function. The derivation of (2) is based on replicating the argument given in Lemma 1 of Antoniak (1974) for each dose and litter size, and using conditional independence (given $\boldsymbol{\alpha}$) of the pdf of the response in different doses. $\pi(\boldsymbol{\beta}, \boldsymbol{A})$ denotes the prior distribution on $\boldsymbol{\beta}$ and $\boldsymbol{A}$ and it is discussed in Section 2.3.

Drawing a sample from the joint posterior distribution (2) is typically not challenging for this model. A practical strategy is to use a Metropolis-Hastings algorithm (Tierney, 1994) with random walk proposal. In our application, we use a normal proposal with variance obtained from the output of standard generalized linear model analysis as a rough approximations to the location and dispersion of $\boldsymbol{\beta}$. This initial approximation is used to obtain a preliminary sample, based on which we further tune the proposal distribution.

2.3 Prior distributions

Bayesian analysis of (2.1) requires the specification of a prior distributions. One advantage of our semiparametric formulation is that we only need to elicit a prior distribution on a small number of parameters $\boldsymbol{\beta}$ and $\boldsymbol{A}$. Dispersed but proper priors on $\boldsymbol{\beta}$ will work in many applications, and are especially attractive in regulatory settings. We adopt them here in Section 3 by assuming $\boldsymbol{\beta} \sim N_2((0,0), 3I)$.

The choice of prior distribution on $\boldsymbol{A}$ is similar to the choice of a smoothness constant in other nonparametric settings. The more mass is assigned to large values of the A^d's, the more the model will be close to its parametric backbone. In general, even though the data can provide relevant information about the vector $\boldsymbol{A}$, it is important to specify a proper prior distribution. In

7

particular, it is possible for the likelihood function not to be integrable, so that a proper prior is required to obtain a proper posterior (Carota and Parmigiani, 1997).

In Section 3 we assign prior distributions on A by considering the implied distribution on the prediction weights $A(/A + N)$ (Leonard and Novick, 1986; Carota and Parmigiani, 1997) which control how close the posterior mean of the unknown pdf is to parametric backbone f_0. Here N indicates the number of dams in a litter–size/dose combination. Our prior states that all possible weights are equally likely a priori for a predictive distribution with a "typical" N. In our application we specify the typical value to be $N = 10$, and carry out a sensitivity analysis of results to the choice of N. When we consider dose–dependent A^d we assign them independent priors such that each $A^d/(N + A^d)$ is uniform.

3 Data Analysis

In this section, we discuss the analysis of a simulated data set and two developmental toxicity studies in rodents. The simulated data are from a Binomial($\theta^d, 15$) with logit(θ^d) = $-3 + .04d$, $M^d = 20$, and dose levels $0, 25, 50, 100$, and 150. The two studies consider respectively the effects of diethylhexalphthalate or DEHP (Tyl et al., 1983) and the effects of 2,4,5–Trichlorophenoxiacetic, or 2,4,5–T, (Holson et al., 1991; Bowman and George, 1995). In the DEHP data set the outcome is the number of malformations among live pups, in the 2,4,5–T data set the outcome is number of resorbed embryos or dead fetuses plus the number of fetuses with cleft palate malformation.

3.1 Exploratory Analyses

The DEHP and 2,4,5–T datasets are displayed in the top panels of Figure 1. Dose levels are indicated at the bottom, while the numbers of dams exposed to each dose level are displayed at the top. Each circle corresponds to a dam. The circles' areas are proportional to the litter size. The circles' coordinates are the dose level and the observed relative frequency of malformations

for the corresponding dam. Both experiments present evidence of overdispersion, at most doses. There is also evidence of zero inflation and n-inflation. Sometimes these are both occurring at the same dose, as in the intermediate doses of the 2,4,5-T data.

We performed over-dispersion diagnostics using the convexity-plot (C-plot) (Lambert and Roeder, 1995). The C-plot for logistic regression is a graph of $C(\theta)$ versus θ, where: $C(\theta) = \{\sum_{d=1}^{D} M^d\}^{-1} \times \sum_{d=1}^{D} \sum_{j=1}^{M^d} \left(\theta/\hat{\theta}^d\right)^{y_j^d} \left\{(1-\theta)/(1-\hat{\theta}^d)\right\}^{N_j^d - y_j^d}$ and $\hat{\theta}^d = \left\{1 - \exp(\hat{\beta}_0 + \hat{\beta}_1 d)\right\}^{-1}$. The shape of the C-plot is used to detect over-dispersion: the more convex $C(\theta)$ is in the range $(\hat{\theta}^0, \hat{\theta}^D)$, the greater the evidence for over-dispersion. The second row of Figure 1 shows the C-plots for the two data sets, for θ's in the range $(\hat{\theta}^0, \hat{\theta}^D)$. The C-plot for both data are convex, suggesting that the logistic regression model with Binomial error may not be appropriate.

3.2 Results and sensitivity analysis

We analyzed each data set using the Semiparametric Bayesian (SB), Beta-Binomial (BB) and Binomial (B) models. All regression coefficients are assigned the same vague prior $\boldsymbol{\beta} \sim N_2((0,0), 3I)$. The BB model is parameterized by setting the parameters of the second-stage beta distribution to $\alpha_1^d = \theta^d(\tau^{-2} - 1)$ and $\alpha_2^d = (1 - \theta^d)(\tau^{-2} - 1)$. The logit of τ^2 is assigned a flat prior. The SB model is parameterized with a common precision parameter A for the DEHP analysis, and dose-dependent precision parameters A^d for the 2,4,5-T analysis. For each of the analyses we generated a MCMC sequence of 10,000 iterations. Standard convergence diagnostics do not reveal convergence problems.

Table 1 compares posterior medians and quartiles of the posterior distribution of the probability of malformations at dose d, φ^d, in each of the nine model / data combinations. In the B model, $\varphi^d = \theta^d$; in the BB model $\varphi^d \sim Beta\{\theta^d(\tau^{-2} - 1), (1 - \theta^d)(\tau^{-2} - 1)\}$; in the SB model, $\varphi^d \sim Beta\{\theta^d A^d, (1 - \theta^d)A^d\}$. Samples of θ^d, τ^2 and A^d are available from the MCMC runs. Obtaining samples of φ^d in the BB and SB models requires an additional draw from the beta

9

distributions above, which is straightforward.

While means of the sampled θ^d estimate the posterior means of the malformation probability in all three models, their variances are not directly comparable. On the other hand, the entire distributions of probabilities of malformations are directly comparable across model. As expected, for the simulated data set, results are similar across the three models. For the other two data sets, differences are more pronounced at the low and high doses, possibly as the result of the presence of zero–inflation and or n–inflation in the data. The IQR's obtained with the SB approach are the widest and they often encompass the union of those obtained using the B and BB model, highlighting the model–robustness of the approach.

A parameter of interest in toxicology studies is the dose level at which the probability of malformation is k times larger than it is at background, or $ED(k)$. Formally, $\theta^{ED(k)} = k\theta^0$, so that $ED(k) = -\{\log(k^{-1}(1 + \exp(-\beta_0)) - 1) + \beta_0\}/\beta_1$ for $k > 1 + \exp(-\beta_0)$. Using the samples available from the MCMC runs, it is straightforward to evaluate the distribution of the $ED(k)$. The top panels of Figure 2 show boxplots of samples from these posterior distributions for plausible ranges of k.

Posterior inference on the precision parameters A^d quantifies the extent to which the data favor the parametric backbone model. The bottom left panel of Figure 2 shows the posterior (histogram) and prior (solid line) distributions of the parameter $\log(A)$ for the DEHP analysis. The prior is such that $A/(A + 10)$ is uniform, so $\log(A)$ is a logistic with location $\log(N)$. This is a rather dispersed prior, as it encompasses values such as $A = 1$, where the backbone has very little weight, and $A = \exp(5)$ where the model is close to the backbone. The posterior is more concentrated than the prior on both tails, indicating that the data provide valuable information about A. The solid vertical line is placed at the posterior mean.

The bottom right panel of Figure 2 shows boxplots of posterior samples of the precision parameters $\log(A^d)$ corresponding to the six doses for the 2,4,5–T data set. The six priors are

independent and identical to the prior in the bottom left panel. Here the posterior medians of the $\log(A^d)$ decrease with the dose, indicating that the data favor the non-parametric model as the dose increases. The distributions of $\log(A^0)$ and $\log(A^{90})$ are the most extreme. The reason is that the data (see Figure 1) are highly compatible with the binomial model at baseline, while they display an extreme deviation at dose 90. The probability of malformation at dose 90 under the B model from Table 1 is .87, with a narrow IQR, clearly a poor fit of the data. Using the SB model, the probability of malformation is .99, with IQR from .76 to 1, a much closer fit.

Using (1), we can draw samples of the cdf functions $\tilde{F}_j^d$ from the posterior distribution given the sampled values of β and A. These can be use for prediction (future data would be draws from such distributions) and for assessing goodness of fit. In Figure 3 we compare samples of $\tilde{F}_j^d$'s from SB, BB and B, for the two cells defined by doses 100 (DEHP) and 90 (2,4,5–T) and litter-sizes 11 (DEHP) and 12 (2,4,5–T). The empirical frequencies of the malformation counts are displayed at the top. For example in DEHP there were 3 dams with 6 out of 11 malformations, 1 with 5 out of 11, and so on.

Both empirical distributions display bimodality, resulting from dam heterogeneity that would be difficult to capture using parametric random–effects models. The 2,4,5–T data show both zero and n–inflation. The results highlight the flexibility of the SB model, which adapts to the observations and fits the data better than the BB and B counterparts. The SB prediction are a data–driven compromise between the parametric backbone and the empirical distribution, and are both adaptable and smooth. In the 2,4,5–T the parametric models are forced to effectively choose among the subpopulations of dams, and make a prediction that ignores the presence of the subpopulation that appears to be less sensitive to the compound. The opposite error could also happen in a different configuration. Figure 3 also stresses the strength of the SB approach in making inferences about any of the quantiles of the distribution of the response given the dose.

Finally, we have assessed the sensitivity of our results to the choice of the prior distributions.

The parameters that are of interest from a regulatory standpoint can be assigned vague, or even flat priors in this model. So our efforts concentrated on the specification of the prior on the precision parameters A's. We have assumed a uniform distribution on the weight $A/(A + N)$. This still requires the choice of a "target" sample size N. For our sensitivity analysis we selected three additional alternative scenarios, with the goal of covering a very broad spectrum of values: $N = 10$ (baseline); $N = 5$; $N = 20$; $N = 50$. The results are summarized in Table 2. Our model revealed little sensitivity of these parameters to the choice of N, and even less for the parameters β, not shown. The most sensitive parameter is A^0. At dose 0 the data are in agreement with the binomial model, and support large values of A^0.

4   Discussion

In this paper we introduced a Bayesian semiparametric approach to dose–response models for teratological data. Our model is designed to address situations in which the scientific focus is on a simple parametric dose response curve, and the distribution of the response needs to be modeled flexibly. Our formulation achieves several goals: 1) the distribution of the response is modeled in a general way; 2) the degree to which the distribution of the response adapts non-parametrically to the observations is driven by the data; 3) the parameters of teratological interest maintain their usual interpretation; 4) the marginal posterior distribution of the parameters of interest is available in closed form; 5) the specification of the prior distribution is very low–dimensional, interpretable, and amenable to "default" choices. The logistic regression model, as well as many extensions are special cases of the model proposed here. Our model is not designed to address estimation of the dam-to-dam variability in the response. Random–effects and hierarchical model are more appropriate in that case.

Our model specification is simple to handle and to interpret. Inference about any of the quantiles of the distribution of the response is easily available. While estimation requires an MCMC

approach, the simulation involved is typically not challenging and can be programmed easily. Extensions to more complex models for the birth defect probability for the precision parameter can be implemented without adding substantial complexity to the analysis. For example, it is possible to use the alternative parametric forms for f_0. Similarly, alternative to the linear logit specification are straghtforward.

Some practitioners may be dissatisfied with methodology that requires specification of a priori distributions. In this context however, the parameters that are of interest from a regulatory standpoint can be assigned vague, or even flat priors. The specification of the precision parameters A's effectively replaces the specification of a probability distribution in standard parametric analyses. It can be viewed as a probabilistic, non-binding, alternative to specifying a family of marginal probability distributions. In this sense it requires less stringent subjective input than parametric models, be they classical or Bayesian, fixed– or random–effect.

For our analysis we chose a Bayesian approach. However, the general modeling strategy of Section 2 is also amenable to maximum likelihood inference based on the marginal likelihood in expression (2). More generally, standard advantages of the Bayesian approach apply, including: 1) an accurate assessment of parameter uncertainty, that does not depend on asymptotic approximations, even in the presence of a semi-parametric specification; 2) practical computation of the predictive distribution for birth defects at a given dose, including accurate assessment of uncertainty about any feature of the predictive distribution; 3) ease of inference on nonlinear transformations of the parameters, such as the ED.

Acknowledgments

13

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669–679.

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to Bayesian non-parametric problems. *The Annals of Statistics* **2**, 1152–1174.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. New York: Wiley.

Bowman, D. and George, E. O. (1995). A satured model for analyzing exchangeable binary data: Applications to clinical and developmental toxicity studies. *Journal of the American Statistical Association* **90**, 871–879.

Carota, C. and Parmigiani, G. (1997). *Semiparametric Regression for Count Data*. Discussion paper 97-17, Institute of Statistics and Decision Sciences, Duke University, `http://ftp.isds.duke.edu/WorkingPapers/97-17.html`.

Catalano, P. J. and Ryan, L. M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association* **87**, 651–658.

Catalano, P. J. and Ryan, L. M. (1994). Statistical methods in developmental toxicology. In: *Handbook of Statistics Volume 12: Environmental Statistics*, 507– 533.

Cifarelli, D. M., Muliere, P., and Scarsini, M. (1981). The linear model in the Bayesian non parametric approach (italian). *Pub. Ist. Mat. "G. Castelnuovo"* .

Cifarelli, D. M. and Regazzini, E. (1978). Nonparametric statistical problems under partial exchangeability. The use of associative means (in italian). *Annali dell' Istituto di Matematica Finanaziaria dell'Università di Torino* **3(12)**.

Cox, D. R. (1983). Some remarks on overdispersion. *Biometrika* **70**, 269–274.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals*

of Statistics **1**, 209–230.

George, O. E. and Wu, H. (1997). Testing for treatment related trend with partially exchangeable clustered data. *Biometrics* **53**, 479–487.

Holson, J. F., Gaines, T. B., Nelson, C. J., LaBorde, J. B., Gaylor, D. W., Sheehan, D. M., and Young, J. F. (1991). Developmental toxicity of 2,4,5-Trichlorophenoxiacetic acid I: Dose response studies in four inbred strains and one outbred stock of mice. *Fundamentals of Applied Toxicology* **19**, 286–297.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**, 1–14.

Lambert, D. and Roeder, K. (1995). Overdispersion diagnostics for generalized linear models. *Journal of the American Statistical Association* **90**, 1225–1236.

Lefkopoulou, M., Moore, D., and Ryan, L. (1989). The analysis of multiple correlated binary outcomes: Application to rodent teratology experiments. *Journal of the American Statistical Association* **84**, 810–815.

Leonard, T. and Novick, M. R. (1986). Bayesian full rank marginalization for two-way contingency tables. *Journal of Educational Statistics* **11**, 33–56.

Liang, K.-Y. and McCullagh, P. (1993). Case studies in binary dispersion. *Biometrics* **49**, 623–630.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Lipsitz, S. R., Laird, N. M., and Harrington, D. P. (1991). Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. *Biometrika* **78**, 153–160.

Mukhopadhyay, S. and Gelfand, A. E. (1997). Dirichlet process mixed generalized linear models. *Journal of the American Statistical Association* **92**, 633–639.

Müller, P. and Rosner, G. (1997). A Bayesian population model with hierarchical mixture priors applied to blood count data. *Journal of the American Statistical Association* **92**.

Prentice, R. L. (1986). Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *Journal of the American Statistical Association* **81**, 321–327.

Rosner, B. (1989). Multivariate methods for clustered binary data with more than one level of nesting. *Journal of the American Statistical Association* **84**, 373–380.

Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics* **22**(4), 1701–1762.

Tyl, R. W., Jones-Price, C., Marr, M. C., and Kimmel, C. A. (1983). *Teratologic Evaluation od diethylhexyl phthalate (CAS No.111-81-7)*. Cd-1 mice, final study report for NCTR/NTP contract no.222-80-2031 (c), NITS no. PB85105674, National Technical Information Service, Springfield, Virginia.

Williams, D. A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics* **31**, 949–952.

Williams, D. A. (1982). Extra-binomial variation in logistic linear models. *Applied Statistics* **31**, 144–148.

Wong, G. Y. and Mason, W. M. (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association* **80**, 513–524.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association* **86**, 79–86.

SIMULATED BINOMIAL		$\beta_0 = -3,\ \beta_1 \times 100 = 4$			
	φ^0	φ^{25}	φ^{50}	φ^{100}	φ^{150}
SB	0.047 (0.020,0.087)	0.13 (0.079,0.19)	0.28 (0.21,0.35)	0.71 (0.64,0.78)	0.95 (0.91,0.98)
BB	0.046 (0.039,0.051)	0.12 (0.110,0.13)	0.27 (0.26,0.28)	0.74 (0.73,0.76)	0.96 (0.95,0.96)
B	0.047 (0.043,0.051)	0.12 (0.110,0.13)	0.27 (0.26,0.28)	0.74 (0.72,0.75)	0.95 (0.95,0.96)

DEHP					
	φ^0	φ^{25}	φ^{50}	φ^{100}	φ^{150}
SB	0.0017 (0.00,0.017)	0.022 (0.0039,0.07)	0.09 (0.037,0.17)	0.45 (0.34,0.57)	0.86 (0.77,0.93)
BB	0.00026 (0.00,0.011)	0.012 (0.00,0.065)	0.077 (0.022,0.18)	0.47 (0.32,0.63)	0.90 (0.78,0.97)
B	0.019 (0.016,0.023)	0.045 (0.04,0.05)	0.11 (0.09,0.12)	0.43 (0.41,0.45)	0.82 (0.79,0.85)

2,4,5–T					
	φ^0	φ^{30}	φ^{45}	φ^{60}	φ^{75}	φ^{90}
SB	0.042 (0.02,0.063)	0.16 (0.10, 0.24)	0.32 (0.24, 0.40)	0.51 (0.36, 0.66)	0.72 (0.54, 0.86)	0.99 (0.76,1.00)
BB	0.003 (0.00,0.065)	0.14 (0.026 0.40)	0.34 (0.12, 0.63)	0.58 (0.30, 0.83)	0.80 (0.52, 0.95)	0.94 (0.73,0.99)
B	0.031 (0.029,0.033)	0.16 (0.150, 0.16)	0.31 (0.31, 0.32)	0.53 (0.52, 0.53)	0.73 (0.72, 0.74)	0.87 (0.86,0.87)

Table 1: *Comparison of estimated malformation probabilities φ^d using the Semiparametric Bayesian (SB), the Beta-binomial (BB), and the Binomial (B) models, for each data set. Entries are the median and, in parenthesis, the quartiles of the distribution of the probability of malformation in a pup for a hypothetical new dam.*

2,4,5−T						
	$\log(A^0)$	$\log(A^{30})$	$\log(A^{45})$	$\log(A^{60})$	$\log(A^{75})$	$\log(A^{90})$
N=5	3.71 (1.05)	2.41 (0.60)	2.71 (0.39)	1.67 (0.31)	1.28 (0.38)	-0.30 (0.72)
N=10	3.73 (1.04)	2.47 (0.59)	2.75 (0.38)	1.67 (0.31)	1.35 (0.41)	-0.31 (0.74)
N=20	3.87 (1.20)	2.63 (0.65)	2.84 (0.38)	1.72 (0.30)	1.31 (0.39)	-0.14 (0.92)
N=50	4.55 (1.39)	2.63 (0.67)	2.83 (0.43)	1.70 (0.29)	1.34 (0.39)	-0.12 (0.81)

Table 2: *Sensitivity analysis for the 2,4,5−T data set. Entries are posterior means and standard deviations (in parenthesis) of* $\log(A^d)$ *under 4 alternative choices of the prior hyperparameter N.*
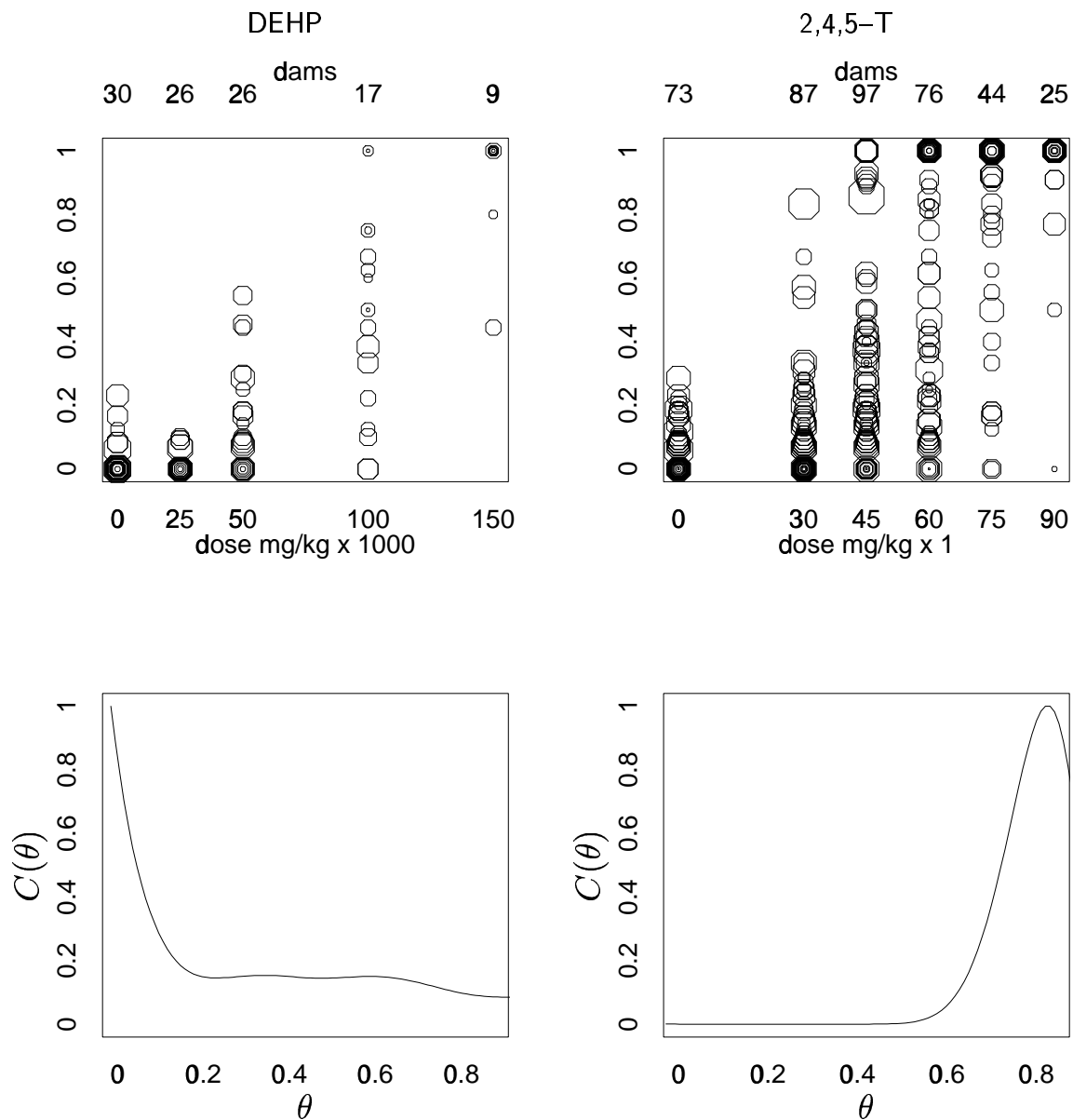
Figure 1: *The DEHP (left) and 2,4,5–T (right) data sets, used in Section 3. The top panels display the raw data. Each circle corresponds to a dam. The circles' areas are proportional to the litter sizes; the circles' coordinates are the dose level and the observed relative frequency of malformations. In addition, the numbers of dams exposed to each dose level is displayed at the top. The bottom panels are the corresponding C-plots. Convexity suggests over-dispersion for both data sets.*
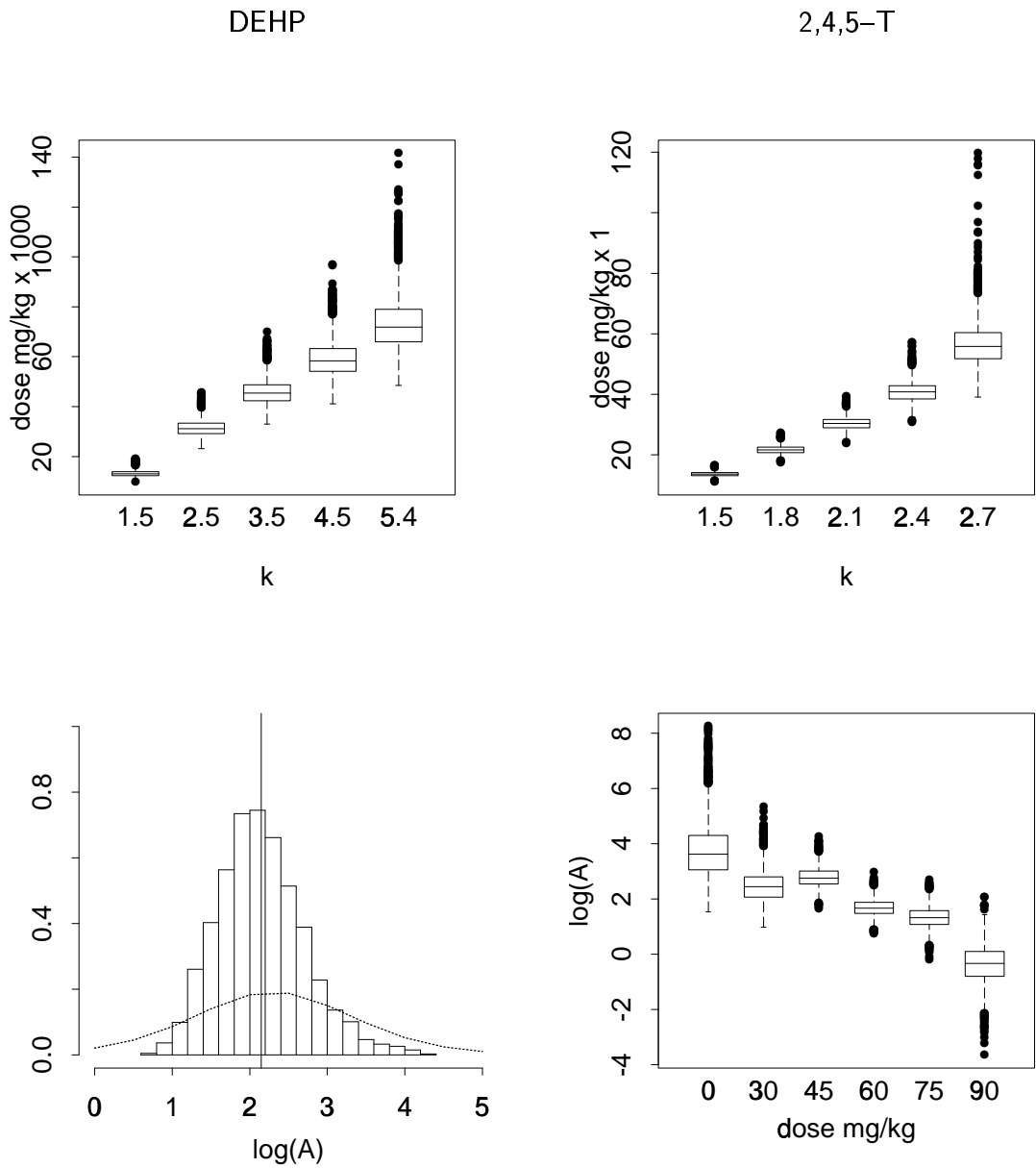
Figure 2: *Inference on the $ED(k)$ (top) and precision parameters (bottom) for the DEHP and 2,4,5,-T data. Top panels are boxplots of samples from the posterior distributions of effective doses $ED(k)$ corresponding to a k–fold increase of the probability of malformations above the background rate. The bottom left panel displays the posterior (histogram) and prior (solid line) distributions of the precision parameter $\log(A)$ for the DEHP data set. The bottom right panel shows boxplots of posterior samples of the precision parameters $\log(A^d)$ corresponding to the six doses for the 2,4,5-T data set.*
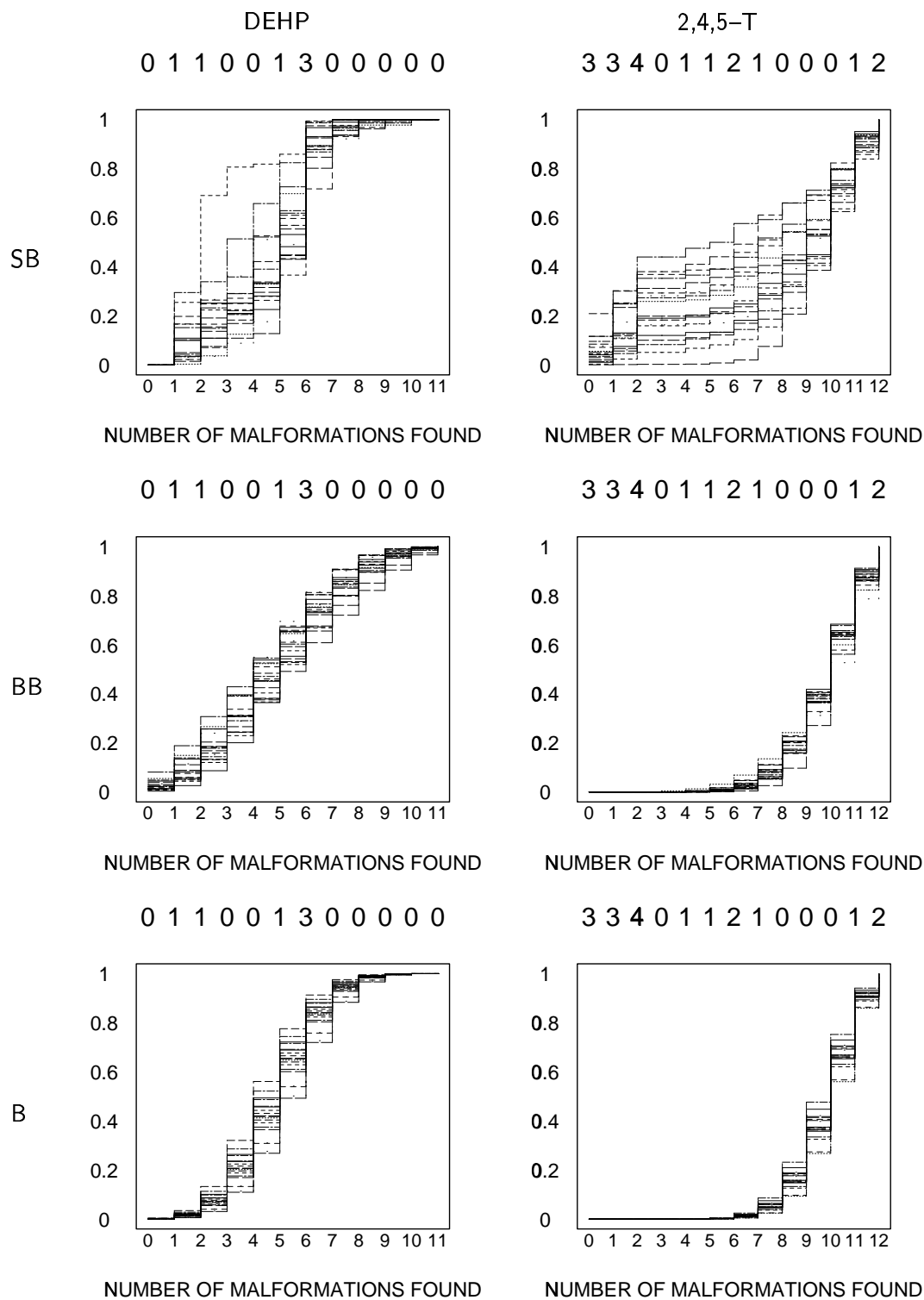
Figure 3: *Samples of 20 cumulative distributions functions $\tilde{F}_j^d$ obtained using the Semiparametric Bayesian (SB), the Beta-binomial (BB), and the Binomial (B) models. For the DEHP data set we choose a dose of $100\mu_g/m^3$ and a litter size of 11; for the 2,4,5-T we choose a dose of $90\mu_g/m^3$ and a litter size of 12. Empirical frequencies corresponding to the selected dose/litter size are displayed at the top of each picture.*