STATISTICS IN MEDICINE
Statist. Med. 2000; **00**:1–6 *Prepared using* simauth.cls *[Version: 2002/09/18 v1.11]*

A Unified Approach for Analyzing Exchangeable Binary Data with Applications to Developmental Toxicity Studies

Xin Dang[a, ‡], Stephine Lena Keeton[b, §] and Hanxiang Peng[c, *, †, ‡]

[a] *Department of Mathematics, University of Mississippi, University, MS 38677, USA*
[b] *Center for Veterinary Medicine, Food and Drug Administration, Rockville, MD 20855, USA*
[c] *Department of Mathematical Sciences, Indiana University – Purdue University At Indianapolis, Indianapolis, IN 46202-3216, USA*

SUMMARY

In this article, we present a general procedure to analyze exchangeable binary data that may also be viewed as realizations of binomial mixtures. Our approach unifies existing models and is practical and computationally easy. Resulting from completely monotonic functions, we give a rich family of parametric parsimonious binomial mixtures, including the incomplete Gamma-, Normal-, Poisson-, and Beta-binomial, generalizing the Beta-binomial. We show that the family is closed under convex linear combinations, products, and composites. We also give the moments and the Markov property for this family of mixtures. With such distributions, we can perform statistical inference on correlated binary data and, in particular, overdispersed data. We propose a regression procedure which generalizes logistic regression. We provide a forward model selection procedure about how a possible optimal model from the family can be achieved. We run a small simulation to validate the inclusion of the binomial distribution. Finally, we apply the proposed procedure to analyze the 2, 4, 5-T and E2 data and compare the results with existing procedures.

1. Introduction

The binomial distribution is widely used in modeling binary response data in many different areas of science. The binomial distribution, however, assumes that the binary responses are independent, an assumption which is often not valid for real data. Non-independence usually leads to a variance that is greater than the nominal variance of the binomial distribution. This is known as *over-dispersion* (or *extra-binomial variation*). The perception of *exchangeability*, intensely studied over the past century, is meant to capture the notion of symmetry in a collection of random variables and is often used as an alternative to independence. In this article, we relax independence to exchangeability and introduce a rich family of parsimonious distributions resulting from *completely monotonic functions*. We present a general framework that unifies existing procedures for modeling correlated binary responses and demonstrate the procedure using real correlated binary and overdispersed data.

As a specific example from the familial correlated data, let us look at a typical developmental toxicity experiment in animal studies, where fetuses from the same litter will respond *more similarly* to a stimulus than fetuses from different litters. Consider a litter of m fetuses with binary responses $B_1, ..., B_m$, where $B_i = 1$ or 0 denotes death or no death, malformation or no malformation. These binary responses are not

*Correspondence to: Department of Mathematical Sciences, Indiana University – Purdue University At Indianapolis, Indianapolis, IN 46202-3216, USA.
†Email: hpeng@math.iupui.edu.
‡This research is supported by the US National Science Foundation under Grant No. DMS-0707074.
§The second author is affiliated with the U.S. Food and Drug Administration and declares that the opinions and information in this review article are those of the authors, and do not represent the views and/or policies of the U.S. Food and Drug Administration.

Received 17 January 2000
Revised 18 September 2002

independent but correlated. One simple way to model them is to assume that the responses are exchangeable. As another specific example, let us look at the risk analysis in Finance such as CreditRisk$^+$, where the default probability of a company is assumed to depend on a set of common economic factors; given these common factors, defaults of the individual obligors are conditionally independent. Consider a portfolio of m obligors with concentration on the binary outcomes B_1, ..., B_m of default or non-default of the obligors. One way in practice to formalize the notion of a *homogeneous* group is to assume that these binary outcomes are exchangeable. For a full description, see e.g. Frey and McNeil [1]. Specifically, binary events $B_1, ..., B_m$ are exchangeable if for every $\{0,1\}$-valued variables $b_1, ..., b_m$, we have

$$\mathbb{P}(B_1 = b_1, ..., B_m = b_m) = \mathbb{P}(B_{\pi_1} = b_1, ..., B_{\pi_m} = b_m),$$

for every permutation $\pi_1, ..., \pi_m$ of $1, ..., m$. Let $Y = B_1 + ... + B_m$ be the total number of "successes". Under exchangeability, the distribution of the number Y of successes (in a litter) is given by

$$\mathbb{P}(Y = y) = \binom{m}{y} \sum_{k=0}^{m-y} (-1)^k \binom{m-y}{k} \lambda_{y+k}, \quad y = 0, 1, ..., m, \tag{1.1}$$

where $\lambda_0 = 1$ and $\lambda_k = \mathbb{P}(B_1 = 1, ..., B_k = 1)$, termed in the literature as the marginal probability, is the probability of k consecutive successes for $k = 1, ..., m$. See e.g. Kendall [2], George and Bowman [3], or Chow and Teicher [4]. We write $Y \sim \mathbf{EB}(\boldsymbol{\lambda})$ with $(\boldsymbol{\lambda} = \{\lambda_k\})$ and refer it to as the *Exchangeable Binomial (EB)* distribution. Clearly, if binary events $B_1, ..., B_m$ are independent, then $\lambda_k = \lambda_1^k, k = 1, ..., m$ and probability (1.1) reduces to the binomial probability $\mathbb{P}(Y = y) = \binom{m}{y}\lambda_1^y(1-\lambda_1)^{m-y}$ for $y = 0, 1, ..., m$, where λ_1 is the probability of a single success.

Applications are not limited to the aforementioned financial and familial data. A wide variety can be found in other fields of science. For example, in social science, see e.g. Conaway [5]; in clinical trails, see Fanaroff [6]; and in botany, see Example 1.3, Collett [7].

Data consisting of independent clusters or groups of dependent binary random variables commonly arise in many fields, including developmental toxicity studies, financial credit risk analysis, longitudinal studies with repeated measurement of subjects, studies of familial diseases, cluster sample surveys. The literature describing the analysis of correlated binary data is extensive (see, e.g., Joe [8]; Collett [7]; Neuhaus [9]. The different approaches account for cluster effects in different ways and may be broken into three main approaches. The *exchangeable model* is one approach. A second approach is the *random effects model*. The third approach is the *quasi-likelihood or generalized estimating equations (GEE) model*.

In a random effects model, the within-cluster (litter) correlation is assumed to be induced by random effects. Given the random effects, the within-cluster responses are conditionally independent and identically distributed, hence the sum of the responses follows the (conditional) binomial distribution. Williams [10] and Kupper and Haseman [11] proposed to use the Beta distribution to model the random effects on the response probability, leading to the commonly used Beta-Binomial distribution. Stiratelli, Laird and Ware [12] modeled the logit of the response probability to be normally distributed and introduced the logit-normal-binomial distribution. Modeling the probit of the response probability to be normally distributed, Ochi and Prentice [13] proposed the probit-normal-binomial model. Conaway [5] modeled the iterated logarithm of the response probability to have a log-Gamma distribution. Coull and Agresti [14] modeled the joint distribution of multiple binomial responses by the multivariate Logit-Normal-Binomial. Instead of assuming a parametric mixing distribution, Follmann and Lambert [15] left the mixing distribution completely unspecified and estimated it by a nonparametric maximum likelihood. Although these models are easily interpreted, most of them, like many Bayesian procedures, involve formidable numerical integration.

In the quasi-likelihood or GEE approach, the first two moments are specified, while the higher order moments and correlations are approximated by a "working matrix". (see, e.g., Williams [16], Liang, Qaqish and Zeger [17], and Lipsitz, Laird and Harrrington [18].) With this approach, estimates can be obtained by iterative weighted least squares. However, the estimates are generally inefficient when the correlation structure is of primary interest. Furthermore, as pointed out by George and Bowman [3] and Kuk [19], the GEE approach often cannot provide satisfactory estimates of quantities which depend on higher order moments. For example, it cannot satisfactorily estimate the probability of affected litters which is of interest in teratological risk assessment.

The proposed framework has many advantages over some commonly used procedures. The statistical inference based on the conditional analysis does not use any information about the distribution of the latent effects since it is lost in the conditioning. The EM algorithm method by Mislevy [20] is conceptually simple, but the computations may be formidable because each E-step in the computation may require a numerical integration. The empirical Bayes approach by Stiratelli, Laird, and Ware [12] may also be computationally difficult. The quasi-likelihood and GEE approach only use the first two moments. By exploiting the notions of exchangeability and complete monotonicity, our approach uses all of the distributional information including, of course, all order moments, so that it is a full likelihood approach. The probability mass functions of the proposed models are both mathematically and computationally simple. The likelihoods can be computed without numerical integration and parameter estimation can be obtained by maximum likelihood using either R or SAS.

From the celebrated de Finetti theorem, a sequence of exchangeable binary random variables is a mixture of binomials, whereas the mixture can be characterized by a completely monotonic sequence. Using this characterization between exchangeability and complete monotonicity, George and Bowman [3] initiated an investigation on correlated binary data and proposed the *folded logistic link*. In Section 2, we point out that it is not a valid link, where we also provide a modified version. Kuk [19] proposed the *power link* and established its complete monotonicity. George and Bowman [3] and Kuk [19] applied their approaches in clinical and developmental toxicity studies and compared their results with existing models such as the generalized estimating equations and the Beta-binomial. Stefanescu and Turnbull [21] used the EM algorithm to model exchangeable binary data with varying cluster sizes. Xu and Prorok [22] showed that, in general, there is no closed form for the maximum likelihood estimates of the marginal probabilities and that the MLE's can only be calculated by numerical methods. They applied their results to a double-blind randomized clinical trial for comparing two antibiotics, cefaclor and amoxicillin and performed several simulation studies. It is our thinking that further study on this approach is worthwhile, namely, to provide several completely monotonic links that are of practical use in conducting real statistical inference; and to give a procedure as a guideline for practitioners for finding possible optimal models.

In this article, we further develop these ideas and present a unified theory. We introduce a rich family of parsimonious exchangeable binomials (i.e., parametric binomial mixtures) via completely monotonic links, including the *incomplete Beta-*, *Gamma-*, *Normal-*, and *Poisson- binomial*. These mixtures are called incomplete because their probability mass functions contain incomplete special functions. For example, the incomplete Beta-binomial include the incomplete Beta function. Since the Beta function is a special case of the incomplete Beta function, the commonly used Beta-binomial is a special case of the incomplete Beta-binomial. Further, the incomplete Beta-binomial includes the binomial as a special case, unlike the Beta-binomial which does not include the binomial as a special case in its parameter domain. Indeed, all the introduced incomplete binomial mixtures include the binomial as a special case, while existing binomial mixtures usually do not; for example, the random effects model for binary data by Conaway [5], the logistic regression by nonparametric mixing by Follmann and Lambert [15], and the mixtures resulting from random variables (see Section 2). Further, we demonstrate that the family is closed under convex linear combinations, products and composites of links. These properties enormously enlarge the family. In addition, we present the Markov property of the stochastic binomial mixture process. We give a stepwise forward model selection procedure. A simulation is conducted to validate the inclusion of the binomial distribution. We apply the proposed procedure to analyze the 2, 4, 5-T and the E2 data and compare the results with existing models. Our results indicate that the proposed procedure improves these models and, in particular, the Gamma-binomials outperforms the other models in analyzing the 2, 4, 5-T data.

The proposed procedure unifies existing approaches. In addition to the aforementioned approaches, additional examples include the usual binomial, the correlated binomial (Kupper and Haseman [11]), the additive Model (Altham [23]), and the generalized logistic regression by nonparametric mixing (Follmann and Lambert [15]). The model by Conaway [5], interestingly, turns out to be a special case of our *incomplete Gamma Power binomial*. Brooks *et al.* [24] explored finite mixture models for proportions. These models can be recovered from the linear combinations of the proposed family. From this standpoint, our work generalizes the results of Brooks *et al.* to conclude that the products and composites of the CM links (see Section 2) are also included in the family (see Section 4).

The rest of the article is organized as follows. The parsimonious distributions are introduced in Section 2, followed by a description of the methods for obtaining such distributions and the Markov property. Section 3 provides examples of parsimonious distributions. Section 4 investigates the methods for obtaining new parsimonious distributions from existing ones. The moments and correlations are also given. Section 5 is devoted to regression with a discussion of computational issues and model selection. We give a small simulation and apply our proposed procedure to analyze two real datasets in Section 6. Technical details can be found in the Appendix.

2. Exchangeable Binomials and the Markov Property

In this section, we first introduce the parsimonious exchangeable binomials, followed by the methods for obtaining such distributions. The Markov property is given in the end of this section.

The marginal probabilities $\boldsymbol{\lambda} = \{\lambda_i : i = 0, 1, ..., m\}$ ($\lambda_0 = 1$) in (1.1) form an (finite) completely monotonic (CM) sequence as pointed out by George and Bowman [3], namely,

$$(-1)^k \Delta^k \lambda_i \geq 0, \quad i = 0, 1, ..., m, \quad k + i \leq m, \tag{2.2}$$

where Δ is the difference operator defined by $\Delta a_i = a_{i+1} - a_i$, $\Delta^2 = \Delta(\Delta)$ and Δ^0 the identity operator for a sequence $\{a_i\}$. A useful formula for checking complete monotonicity is

$$(-1)^k \Delta^k a_i = \sum_{r=0}^{k} \binom{k}{r} (-1)^r a_{i+r}. \tag{2.3}$$

See Feller [25] (page 221) for a proof. Applying this formula with $k = 1$ and $k = 2$, we see $1 \geq \lambda_1 \geq ... \geq \lambda_m \geq 0$ (decreasing) and $\lambda_{i+2} - 2\lambda_{i+1} + \lambda_i \geq 0, i = 0, ..., m-2$ respectively. Conversely, for $\boldsymbol{\lambda}$ satisfying (2.2), define

$$f_{\text{eb}}(y; \boldsymbol{\lambda}) = \binom{m}{y} \sum_{k=0}^{m-y} (-1)^k \binom{m-y}{k} \lambda_{y+k}, \quad y = 0, 1, 2, ..., m. \tag{2.4}$$

Then $f_{\text{eb}}(y; \boldsymbol{\lambda}) \geq 0$ for all y and $\sum_{y=0}^{m} f_{\text{eb}}(y; \boldsymbol{\lambda}) = 1$, hence f_{eb} well defines a probability distribution. This definition is based upon finitely many completely monotone numbers $\{\lambda_1, ..., \lambda_m\}$, while (1.1) is based upon finitely many exchangeable events $B_1, ..., B_m$. However, it follows from Kendall [2] that the two definitions are equivalent. Because the probability in (1.1) involves only finitely many exchangeable events, a slightly negative correlation is allowed in (1.1) (or (2.4)), while an infinite sequence of exchangeable binary random variables has a nonnegative correlation, see Kingman [27].

The mean, variance and second order correlation are

$$\mathbb{E}(Y) = m\lambda_1, \quad \mathbb{V}\text{ar}(Y) = m(\lambda_1 - \lambda_2) + m^2(\lambda_2 - \lambda_1^2), \quad \phi = (\lambda_2 - \lambda_1^2)/\lambda_1(1 - \lambda_1). \tag{2.5}$$

In terms of ϕ, we can rewrite $\mathbb{V}\text{ar}(Y) = \sigma_{\text{ind}}^2\{1 + (m-1)\phi\}$, where $\sigma_{\text{ind}}^2 = m\lambda_1(1-\lambda_1)$ is the variance of the binomial with parameters m and λ_1. Typically $\phi > 0$; it follows $\mathbb{V}\text{ar}(Y) > \sigma_{\text{ind}}^2$, manifesting that the EB can be used to model overdispersed data. Williams' model is, in fact, an approximation to the EB model because the model variance is $\mathbb{V}\text{ar}(Y)$. Thus, Williams' model can be viewed as a generalized estimating equation resulting from the EB model. From the first equality in (2.5), we observe that Y/m is an unbiased estimate of λ_1. However, from the second equality in (2.5), we see that Y/m does not converge in the second moment to λ_1 as the number m of trials tends to infinity, unless $\lambda_2 = \lambda_1^2$, which corresponds to independent binary events $\{B_i : i = 1, 2, ...\}$.

In teratological risk assessment, in addition to the marginal probability p of an affected fetus (which is given by λ_1) and the intra-litter correlation ϕ, the probability $q = P(Y \geq 1)$ of affected litters is also of interest. Kuk [19] used q to assess and determine an acceptable low risk or safe dose level. From (2.4), obviously

$$q = P(Y \geq 1) = 1 - P(Y = 0) = \sum_{k=1}^{m} (-1)^{k-1} \binom{m}{k} \lambda_k. \tag{2.6}$$

In the case of independence (i.e., $\lambda_k = \lambda_1^k$), equation (2.6) becomes the usual formula $q = 1 - (1-p)^m$. Unlike the correlation ϕ, the probabilities of affected litters are observable from the data. Therefore, they can be used as a criterion to compare different models. Models with expected probabilities closer to observed ones are preferred. The use of the observed probability q of affected litters as a criterion is a better indicator because it involves all marginal probabilities $\lambda_1, \lambda_2, ..., \lambda_m$, which contain the total distributional information.

Note that $\mathbf{EB}(\boldsymbol{\lambda})$ has m parameters and is the saturated model with a parameter space $\Lambda = \{\boldsymbol{\lambda} \in \mathbb{R}^m : \boldsymbol{\lambda} \text{ satisfies (2.2)}\}$. Direct estimation of the parameters is challenging due to the high dimensionality and complicated constraints of the parameter space. To overcome these difficulties, we consider parsimonious models, motivated from George and Bowman [3], by mapping a lower d-dimensional subset $\Theta \subset \mathbb{R}^d$ into the saturated m-dimensional parameter space Λ. Consider such a map from Θ into Λ defined by $\boldsymbol{\lambda} = \mathbf{h}(\theta)$, where $\theta \in \Theta$. Write $\mathbf{h} = (h_1, ..., h_m)^\top$ so that $\lambda_j = h_j(\theta), j = 1, ..., m$. Substituting the expressions in (1.1), we obtain a parsimonious model, which can be expressed as

$$f(y;\theta) = \binom{m}{y} \sum_{k=0}^{m-y} (-1)^k \binom{m-y}{k} h_{y+k}(\theta), \quad \theta \in \Theta, \quad y = 0, 1, ..., m, \tag{2.7}$$

where $h_0(\theta) = 1$ for every $\theta \in \Theta$. In order to ensure that the above expression is a valid probability mass function, these $h_1(\theta)$, ..., $h_m(\theta)$ must be completely monotone:

$$(-1)^k \Delta^k h_i(\theta) \geq 0, \quad \theta \in \Theta, \quad i = 0, 1, ..., m, \quad i + k \leq m. \tag{2.8}$$

We shall call such $\mathbf{h}$ *a completely monotonic link (CM link)* and write the resulting submodel as $\mathbf{EB}(\theta, m; \mathbf{h})$ (or simply $\mathbf{EB}(\mathbf{h})$). The above approach turns the high dimensional parameter $\boldsymbol{\lambda}$ in the complicated parameter space Λ to a low dimensional parameter θ in a simple parameter space Θ via a completely monotonic link $\mathbf{h}(\theta)$. The complete monotonicity of the link preserves the parameter structure and the choice of a suitable link to a particular data set can yield an optimal fitting. With such a low dimensional parameter θ, the usual procedures for parameter estimation such as maximum likelihood become feasible.

Each CM link $\mathbf{h}$ gives a parsimonious model. One trivial CM link is $h_t(\theta) = \theta^t$ for $\theta \in (0, 1)$, *the independence link*, corresponding to independent binary responses (the binomial model). We can use (2.3) for a direct verification of CM links. A sufficient condition is that $h_j(\theta) = \bar{h}(j;\theta), j = 0, 1, ...$ with $\bar{h}(0;\theta) = 1$ for some completely monotonic function $\bar{h}(t;\theta)$ defined on $t \in [0, \infty)$ in the sense that $\bar{h}$ has all order derivatives $\bar{h}^{(k)}(t;\theta)$ w.r.t. t satisfying

$$(-1)^k \bar{h}^{(k)}(t;\theta) \geq 0, \quad k = 0, 1, ..., \quad \theta \in \Theta, \quad t \in [0, \infty). \tag{2.9}$$

Henceforth, we shall assume the existence of such a CM function $\bar{h}$ unless otherwise explicitly stated. This assumption implies that the finite sequence of exchangeable binary random variables $B_1, ..., B_n$ is part of some infinite sequence of exchangeable binary random variables.

George and Bowman [3] mentioned this same sufficient condition. However, their proposed *folded logistic link*

$$\bar{h}_t(\theta) = 2/(1 + (1+t)^\theta), \quad \theta \in (0, \infty), \tag{2.10}$$

is not completely monotone on $\theta \geq 0$. Thus it is not a valid link. More specifically, it *is CM only* on $0 < \theta \leq 1$ but *not* on $\theta > 1$. As a counter example, let $\theta = 2$. It is easily checked that the folded logistic link does not satisfy (2.8) for $i = 6$ and $k = 4$th difference. A proof can be found in the Appendix. We modify the link and refer to it as the *piecewise folded logistic (Piecewise Flogit) link*,

$$\bar{h}_t(\theta) = \begin{cases} 2/(1 + (1+t)^\theta), & 0 < \theta \leq 1, \\ 1/(1 + t/2)^\theta, & \theta > 1. \end{cases} \tag{2.11}$$

This is a simple link and has, indeed, demonstrated good behavior when used fitting two real datasets, see Proposition 1 for the proof of complete monotonicity.

Remark 1. *For $\bar{\mathbf{h}}(\theta)$ satisfying (2.9), let $h_t = \bar{h}_{t_0+t}/\bar{h}_{t_0}$ for some $t_0 \geq 0$. Then $\{h_t : t \geq 0\}$ is CM with $h_0 = 1$.*

We shall say that a completely monotonic sequence is *normalized* at t_0. Therefore, every completely monotonic function can be used to obtain a CM link. Moreover, CM links resulting from the same CM function but normalized at different t_0 lead to different models. We now provide several constructive methods to obtain CM links. Examples of links are given in the next section.

CM Links from Laplace Transforms, MGFs and CHFs. It is well known that if a function is completely monotone on $[0,\infty)$, then it is a Laplace transform. Specifically, we have the following most useful theorem quoted from Feller [25].

Theorem 1. *A function $\bar{h}$ on $[0,\infty)$ is completely monotone with $\bar{h}(0)=1$ if and only if it is a Laplace transform of some probability distribution H on $[0,\infty)$, i.e.,*

$$\bar{h}(t)=\int_0^\infty \exp(-tx)\,dH(x),\quad t\in[0,\infty). \tag{2.12}$$

The distribution H is uniquely determined by $\bar{h}$.

By the above theorem, we can obtain CM links from the existing tables of Laplace transforms and, in particular, moment generating functions (mgf's). Specifically, suppose that the mgf $M_H(t)$ of a distribution H exists for all $t\le t_0$ for some $t_0\ge 0$, then

$$\bar{h}(t)=M_H(-t),\quad t\in[0,\infty) \tag{2.13}$$

gives a CM link. Using the relationship between Laplace transforms and characteristic functions (chf's), we can also get CM links. Suppose that the chf of H is φ_H. Then it is easily seen that

$$\bar{h}(t)=\varphi_H(\mathbf{i}t),\quad t\in[0,\infty),\quad \mathbf{i}=\sqrt{-1} \tag{2.14}$$

is a CM link.

CM Links from Moments. A substitution $x=-\log p$ in (2.12) yields a useful representation

$$\bar{h}(t)=\int_0^1 p^t\,dG(p),\quad t\in[0,\infty), \tag{2.15}$$

where G is the induced probability distribution from H by the logarithm transform. Therefore, we can also obtain CM links from the moments of a distribution G. Suppose that G has the jth moment $\mathbb{E}_G(p^j)$ for $j=1,2,...,m$. Then

$$h_j=\mathbb{E}_G(p^j),\quad j=0,1,...,m \tag{2.16}$$

gives a CM link.

Substituting (2.15) in (2.7), we obtain

$$f(y;\theta)=\binom{m}{y}\sum_{k=0}^{m-y}(-1)^k\binom{m-y}{k}\int_0^1 p^{y+k}\,dG(p)=\int_0^1\binom{m}{y}p^y(1-p)^{m-y}\,dG(p). \tag{2.17}$$

This demonstrates that the EB is a distribution of binomial mixtures. Alternatively, the EB can also be viewed as a *random effects model*. Indeed, *the random effects model for binary data* by Conaway [5] is a special parsimonious EB distribution. It is well established in the literature (e.g., de Finetti theorem) that if $B_1,B_2,...$ is an infinite sequence of exchangeable binary random variables, then the finite sum $B_1+...+B_n$ follows a binomial mixture distribution with the mixing distribution G uniquely determined by the infinite binary sequence. Conversely, summing up the above, we conclude that if $B_1,B_2,...,B_m$ are binary random variables for which the finite sum $B_1+...+B_m$ has a binomial mixture with a mixing distribution G, then $B_1,B_2,...,B_m$ are exchangeable with marginal probabilities $\lambda_j=\int_0^1 p^j\,dG(p)$ for $j=0,1,...,m$. Thus, a binomial mixture implies finite exchangeability. From our viewpoint, the difference among exchangeable modeling, mixing modeling, and the random effects approach is more for presentation and interpretation and less for mathematical substance.

CM Links from Incomplete Integrals. If the mixing distribution G is parametric $G = G_\vartheta$ with parameter $\vartheta \in \Theta$, then the resulting CM link $\bar{h}(t;\vartheta)$ is parametric. Additional parameters in a distribution lends itself to increased modeling flexibility. Analogous to the method of *tolerance distributions* for finding link functions in generalized linear models, a CM link containing additional parameters can be obtained by putting additional parameters in the limits of the above integrals (2.12) or (2.15). Specifically,

$$\bar{h}(t;\theta) = \int_{\theta_1}^{\theta_2} p^t \, dG_\vartheta(p), \quad \theta = (\theta_1, \theta_2, \vartheta) \in [0,1]^2 \times \Theta. \tag{2.18}$$

Or equivalently,

$$\bar{h}(t;\theta) = \int_a^b \exp(-tx) \, dH_\vartheta(x), \quad \theta = (a, b, \vartheta) \in [0,\infty)^2 \times \Theta. \tag{2.19}$$

From Remark 1 it follows that

$$h_t(\theta) = \bar{h}(t + t_0; \theta_1, \theta_2, \vartheta)/\bar{h}(t_0; \theta_1, \theta_2, \vartheta), \quad \theta = (\theta_1, \theta_2, \vartheta) \in [0,1]^2 \times \Theta \tag{2.20}$$

is a CM link with $h_0(\theta_1, \theta_2, \vartheta) = 1$ for $t_0 \geq 0$. Observe that the above CM link is not defined when $\theta_1 = \theta_2$. Nevertheless, it can be easily shown that the limit exists,

$$\lim_{\theta_2 \to \theta_1} h_t(\theta_1, \theta_2, \vartheta) = \theta_1^t. \tag{2.21}$$

Interestingly, this corresponds to the independence link (the binomial model). Thus, we can extend the definition of $h_t(\theta_1, \theta_2, \vartheta)$ to admit the equality $\theta_1 = \theta_2$ by defining the value of the function to be the limit, i.e.,

$$h_t(\theta_1, \theta_2, \vartheta) = \theta_1^t, \quad \theta_1 = \theta_2. \tag{2.22}$$

Then $h_t(\theta)$ is well defined for all $\theta \in [0,1]^2 \times \Theta$. Hereafter, we shall adopt this definition without explicitly referring to it. We shall call such links *incomplete links*.

It is interesting to observe that even though the positivity of (2.18) or (2.19) requires $\theta_1 < \theta_2$, the above definition renders the inequality constraint to unconstraint for the positivity of $h_t(\theta_1, \theta_2, \vartheta)$. The unconstraint is very useful both theoretically and computationally because, otherwise, the maximum likelihood estimation of the parameters is a cumbersome inequality-constrained maximization. We sum up our findings in the following theorem.

Theorem 2. *Suppose* $\{\bar{h}(t;\vartheta) : t \geq 0\}$ *satisfies (2.9) for* $\vartheta \in \Theta$. *If (2.21) holds with* h_t *given in (2.20), then* $\mathbf{h}(\theta) = \{h_j(\theta) : j = 0, 1, ..., m\}$ *with* $\theta = (\theta_1, \theta_2, \vartheta) \in [0,1]^2 \times \Theta$ *well defines a parsimonious* $\mathbf{EB}(\theta; \mathbf{h})$ *model. Moreover, the binomial distribution is recovered when* $\theta_1 = \theta_2$.

CM Links from Random Variables. Here we describe another method to get CM links commonly used in the literature. If a random variable ξ has a range of $[0,1]$, then its distribution is concentrated on $[0,1]$ and can be taken as the mixing distribution to obtain a CM link. One convenient way to obtain such a random variable ξ is from the composite $\xi = F(\eta)$, where F is a nonnegative measurable function with range $[0,1]$ and η is a random variable taking values in the domain of F. For example, equating F to the distribution function Φ of the standard normal distribution and η to the standard normal random variable Z gives the *probit-normal mixing* model $\xi = \Phi(\mu + \sigma Z)$, where μ and σ are location and scale parameters. Setting $\xi = 1/(1 + \exp(-\mu - \sigma Z))$ yields the *logit-normal mixing* model. These two mixing distributions are widely used in practice such as in the risk analysis in Finance. Such CM links result in the same models as the random effects approach, see Prentice [26] and Stiratelli, Laird and Ware [12]. It is noteworthy to mention that the resulting binomial mixtures usually do not include the binomial as a special case, whereas the proposed *incomplete binomial mixtures* do. Here we shall not provide further details of this method due to the computational difficulty. As pointed out in Frey and McNeil [1], the binomial mixture and the latent random variables approach are closely related. The difference lies in the presentation and interpretation rather mathematical content. The following Markov property describes the relationships among exchangeability, the latent variables approach, and binomial mixtures.

Parsimonious Exchangeable Binomials as Markov Chains. Let $Y_m \sim \mathbf{EB}(\theta, m; \mathbf{h})$ and consider the stochastic process $\{Y_m : m = 1, 2, ...\}$. Denote by G_θ the probability measure concentrated on $[0, 1]$ determined by the infinite completely monotonic sequence $\mathbf{h}(\theta) = \{h_j(\theta) : j = 0, 1, ...\}$ and write $p \sim G_\theta$ the random variable distributed like G_θ. We give a theorem below with the proof delayed to the Appendix.

Theorem 3. *Suppose that* $\{\mathbf{h}(\theta) : \theta = (\theta_1, \theta_2, \vartheta) \in [0, 1]^2 \times \Theta\}$ *is completely monotone. Then for every* θ *in the parameter space the following hold.*
(1) There exist binary random variables $\tilde{B}_1$, ..., $\tilde{B}_m$ *which, when conditioned on the latent random variable* $p \sim G_\theta$, *are independent and have a common Bernoulli distribution with probability* p *of success, such that* Y_m *has a stochastic representation* $Y_m = \tilde{B}_1 + ... + \tilde{B}_m$ *with*

$$\mathbb{P}(Y_m = y|p) = \binom{m}{y} p^y (1-p)^{m-y}, \quad y = 0, 1, ..., m.$$

(2) $\{Y_m\}$ *forms a non-homogeneous Markov chain with transition probabilities given by*

$$\begin{aligned}
&\mathbb{P}(Y_{m+1} = y + 1|Y_m = y, Y_{m-1} = y_{m-1}, ..., Y_1 = y_1) \\
&= \mathbb{P}(Y_{m+1} = y + 1|Y_m = y) = 1 - \mathbb{P}(Y_{m+1} = y|Y_m = y) \\
&= \sum_{k=0}^{m-y} (-1)^k \binom{m-y}{k} h_{y+k+1}(\theta) \Big/ \sum_{k=0}^{m-y} (-1)^k \binom{m-y}{k} h_{y+k}(\theta).
\end{aligned}$$

(3) $\mathbb{C}\mathrm{ov}(Y_m, Y_n) = mh_1(\theta)[1 - h_1(\theta)] + m(n-1)[h_2(\theta) - h_1^2(\theta)]$ *for* $n \geq m$.
(4) $\{Y_m - mh_1(\theta) : m = 1, 2, ...\}$ *is a martingale.*
(5) Y_m/m *converges in distribution to* Q_θ *as* $m \to \infty$.

The Random Generator. Based on Theorem 3, we have a recipe for generating exchangeable Bernoulli variables B_1, ..., B_m and hence Y_m having a parsimonious $\mathbf{EB}(\mathbf{h})$. First generate $\boldsymbol{\lambda}$ from G_θ determined by $\mathbf{h}(\theta)$. Then generate $B_1, ..., B_m$ i.i.d. from the Bernoulli distribution with probability $\boldsymbol{\lambda}$ of success and $Y_m = B_1 + ... + B_m$. Another approach to generate Y_m is the usual method of simulating discrete random variables using probability mass functions.

3. Examples of Parsimonious Exchangeable Binomials

In this section, we give various CM links using (2.12)-(2.16), (2.18), and (2.19). The CM links resulting from (2.18) and (2.19) usually contain incomplete special functions and are thus referred to as *incomplete CM links*. The CM links resulting from (2.12)-(2.16) are special cases of incomplete CM links and are referred to as *(complete) CM links*.

Complete CM Links. By Theorem 1, each distribution concentrated on $[0, \infty)$ gives a CM link. An explicit formula for a CM link may be found from an existing explicit formula of either a Laplace transform,

Table I. **Completely Monotonic Links.** $\theta = \theta$ **or** (θ_1, θ_2)**.**

Name	Link $(j = 0, 1, ...)$	Domain of θ
Ind-Bin	θ^j	$(0, 1)$
MM-Bin	$\theta/(\theta + j)$	$(0, \infty)$
Beta-Bin	$B(\theta_1 + j, \theta_2)/B(\theta_1, \theta_2)$	$(0, \infty)^2$
Gamma-Bin	$(1 + \theta_2 j)^{-\theta_1}$	$(0, \infty)^2$
Poisson-Bin	$\exp(\theta(e^{-j} - 1))$	$(0, \infty)$
Normal-Bin	$2\exp((\theta j)^2/2)(1 - \Phi(\theta j))$	$(0, \infty)$

a moment generating function, or a characteristic function. For example, the *Gamma binomial (Gamma-Bin)* link is obtained from (2.12) or (2.13) with H being the gamma distribution. It is interesting to note that the Gamma-Bin link yields the model considered by Conaway [5]. As a special case of the Gamma binomial link, taking H to be the exponential distribution in the Laplace transform, we get the simple *Michaelis-Menten (MM-Bin)* link, which is frequently used in biology. An explicit formula for a CM link can also be found from moment formula (2.16). The binomial *(Ind-Bin)* is the special case of the exchangeable binomial with the mixing G being a point mass. Taking G to be the Beta distribution yields the *Beta binomial (Beta-Bin)* link in Table I. The Beta-binomial link yields the the Beta-Binomial distribution. Even though the normal distribution with mean zero is not concentrated on $[0,\infty)$, it is symmetric about zero, so that it can be folded to be concentrated on $[0,\infty)$. The folded distribution resulting from a normal with mean zero and variance σ^2 yields the *normal binomial (Normal-Bin)* link given in Table I. In this way, any symmetric distribution yields a CM link. More generally, any distribution can be used to produce a CM link. We list some of the CM links in Table I. The links are referred to as XX-binomial link, where XX is the name of the mixing distribution such as Gamma and Poisson. Additional CM links can be analogously derived, such as the Inverse-Gamma-Bin, the Negative-Binomial-Bin, the Noncentral-χ^2-Bin, the Discrete-Uniform-Bin, the Positive-Stable-Bin, and the Weibull-Bin.

Incomplete CM Links. We now derive incomplete CM links from formulas (2.18) and (2.19). Our first incomplete CM link is from the independence link $h_t(p) = p^t, 0 < p < 1$. Consider for given t the area $A_t(\theta)$ under the polynomial $y = p^{t-1}$ over $[\theta_1, \theta_2]$, so that $A_t(\theta) = \int_{\theta_1}^{\theta_2} p^{t-1}\,dp = (\theta_2^t - \theta_1^t)/t$ for $t > 0$ with $\theta = (\theta_1, \theta_2)$. The kth derivative w.r.t. t is

$$A_t^{(k)}(\theta) = \int_{\theta_1}^{\theta_2} (\ln p)^k p^{t-1}\,dp, \quad t > 0, \quad k = 1, 2, ..., \quad \theta \in [0,1]^2, \tag{3.23}$$

so $(-1)^k A_t^{(k)}(\theta) \geq 0$. This shows that $A_t(\theta)$ is completely monotone. Since $\lim_{t\to 0} A_t(\theta) = \ln(\theta_2/\theta_2)$, normalizing at $t_0 = 0$ in Remark 1, we immediately get a CM link of two parameters, referred to as the *Inc-A binomial* link in Table II. From (3.23) it also follows that $A_t(\theta)$ is *absolutely monotone (AM)* for $1 < \theta_1 < \theta_2$, i.e., $d^k\phi(t)/dt^k \geq 0, k = 0, 1, ...$ for $1 < \theta_1 < \theta_2$. AM functions can be used to construct new CM links, see Theorem 4 in Section 4. Interestingly, the Inc-A binomial link is a special case of the incomplete Beta-binomial link below.

Alternatively, we can also normalize at $t_0 = 1$, we then get the CM link,

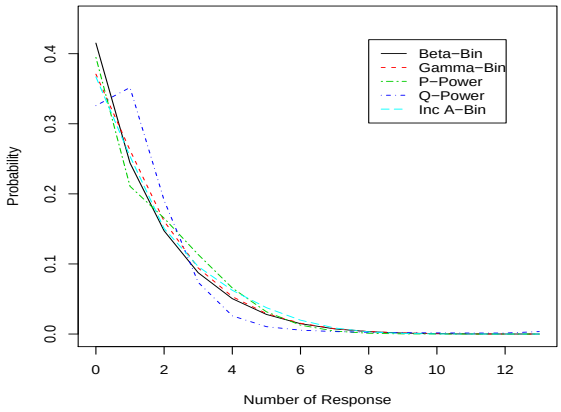
$$h_t(\theta) = \frac{1}{t+1}(\theta_2^t + \theta_2^{t-1}\theta_1 + ... + \theta_2\theta_1^{t-1} + \theta_1^t), \quad t \geq 0, \quad \theta \in [0,1]^2. \tag{3.24}$$

This is a polynomial in θ_1, θ_2 while the Inc-A link contains a nonlinear log function. However, the range of the second order correlation ϕ reduces from $[0, 1/2]$ of the former link to $[0, 1/3]$ of the latter. A detailed discussion about the range of correlation can be found in the Section 4. Hereafter, we shall normalize at $t_0 = 0$.

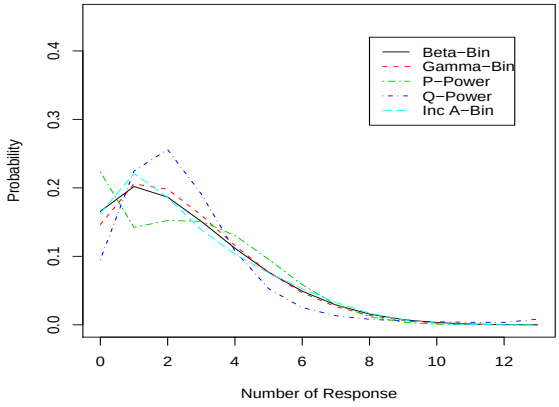
Mixing with the beta distribution, $dG_\vartheta(\lambda) = \lambda^{\theta_3-1}(1-\lambda)^{\theta_4-1}/B(\theta_3, \theta_4)d\lambda, \lambda \in (0,1)$ for $\theta_3 > 0, \theta_4 > 0$ in (2.18), we obtain a CM link, referred to as the *incomplete Beta (Inc-Beta)* binomial link in Table II. Here $B(x;\theta_1,\theta_2) = \int_0^x u^{\theta_1-1}(1-u)^{\theta_2-1}\,du$ is the incomplete Beta function and F_{beta} is the distribution function of the Beta distribution. The case $(\theta_1, \theta_2) = (0, 1)$ recovers the *(complete) Beta-Bin link*, which yields the Beta-binomial distribution. Mixing with the exponential distribution, $dH(x) = \theta_3 \exp(-\theta_3 x)dx$ for $\theta_3 > 0$ in (2.19), we obtain the *Inc-MM* binomial link. The case $(\theta_1, \theta_2) = (0, \infty)$ corresponds to the Michaelis-Menten equation used frequently in biology and recovers the MM-Bin link.

Mixing with the Gamma distribution, $dH(x) = (x/\theta_4)^{\theta_3-1}\exp(-(x/\theta_4))/\theta_4\Gamma(\theta_3)dx, x \geq 0, \theta_3 \in (0,\infty)$, we get the *incomplete Gamma (Inc-Gamma)* binomial link. Here $\Gamma(x,\theta) = \int_x^\infty z^{\theta-1}\exp(-z)\,dz$ is the incomplete Gamma function. The case $(\theta_1, \theta_2) = (0, \infty)$ recovers the *Gamma* binomial link.

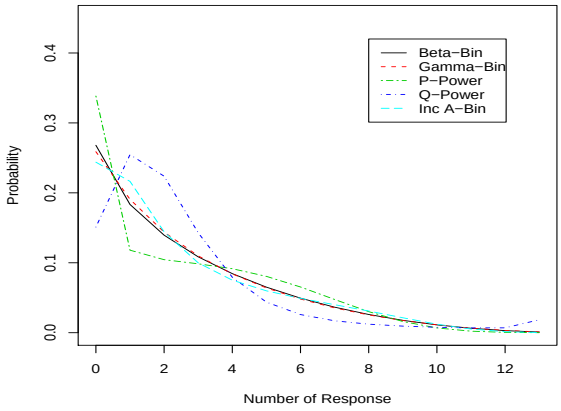
Displayed in Fig. 1 are the two-parameter probability curves with a typical litter size of 13. In each panel, all five distributions share a common response probability p and intra-litter correlation ϕ. The choices of $(p,\ \phi)$ are typical of those in toxicological experiments, corresponding to cases of low, medium and high dose levels. From these plots, we observe the following phenomena. (1) The curves of the Gamma- and



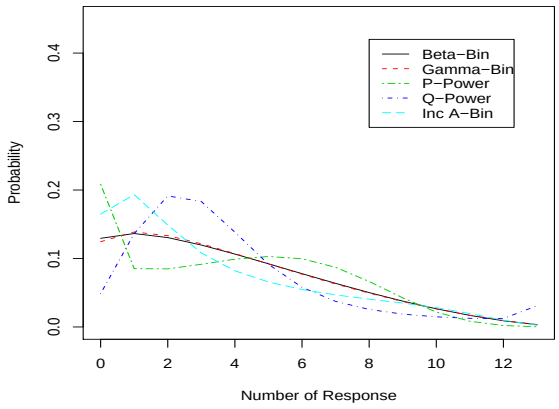
(a) $p = 0.1$, $\phi = 0.1$



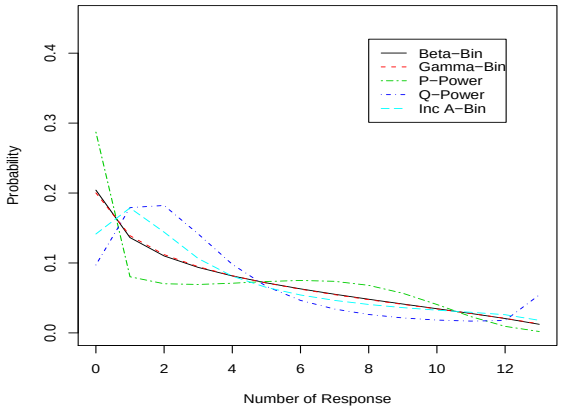
(b) $p = 0.2$, $\phi = 0.1$



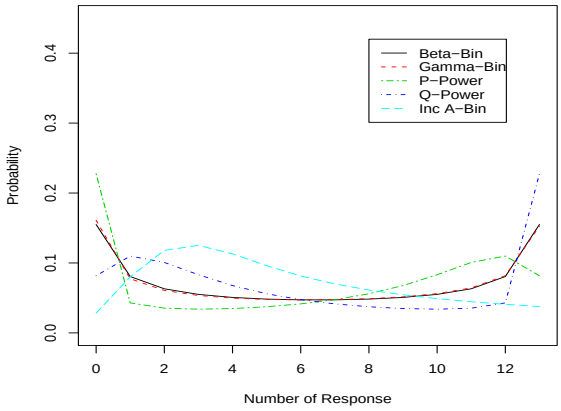
(a) $p = 0.2$, $\phi = 0.2$



(b) $p = 0.3$, $\phi = 0.2$



(c) $p = 0.3$, $\phi = 0.3$



(d) $p = 0.5$, $\phi = 0.5$

Figure 1. Probability curves under a common response probability p and intralitter correlation ϕ.

Table II. **Incomplete CM Links. $\theta = (\theta_1, \theta_2)$, or $(\theta_1, \theta_2, \theta_3)$, etc.**

Name	Link ($j = 0, 1, ...$)	Domain of θ
Inc-A-Bin	$\frac{1}{j}\frac{\theta_2^j - \theta_1^j}{\ln\theta_2 - \ln\theta_1}$	$[0,1]^2$
Inc-MM-Bin	$\frac{\theta_3}{\theta_3+j}\frac{\theta_2^{j+\theta_3} - \theta_1^{j+\theta_3}}{\theta_2^{\theta_3} - \theta_1^{\theta_3}}$	$(0,\infty)^3$
Inc-Beta-Bin	$\frac{B(\theta_3+j;\theta_4)}{B(\theta_3;\theta_4)}\frac{F_{\rm beta}(\theta_2;\theta_3+j,\theta_4) - F_{\rm beta}(\theta_1;\theta_3+j,\theta_4)}{F_{\rm beta}(\theta_2;\theta_3,\theta_4) - F_{\rm beta}(\theta_1;\theta_3,\theta_4)}$	$[0,1]^2 \times (0,\infty)^2$
Inc-Gamma-Bin	$\frac{1}{(1+\theta_4 j)^{\theta_3}}\frac{\Gamma((\theta_4 j+1)\theta_2;\theta_3) - \Gamma((\theta_4 j+1)\theta_1;\theta_3)}{(\Gamma(\theta_2;\theta_3) - \Gamma(\theta_1;\theta_3))}$	$(0,\infty)^4$
Inc-Normal-Bin	$\exp(j^2\theta_3^2/2)\frac{\Phi(\theta_2/\theta_3+\theta_3 j) - \Phi(\theta_1/\theta_3+\theta_3 j)}{\Phi(\theta_2/\theta_3) - \Phi(\theta_1/\theta_3)}$	$(0,\infty)^3$

Beta-binomials are similar with a slight difference in the situation where the response probability p and intra-litter correlation ϕ are small. The Beta-binomial places a little bit more probability mass at zero than the Gamma-binomial. This phenomenon explains why the Beta-binomial underestimates the probability of affected-litters in the E2 data, see Section 6 for more discussion. (2) The P-power and Q-power, proposed by Kuk [19], seem not very suitable for modeling the situations where there are high response probabilities and strong correlations, because they either put a bit too much probability mass at the extreme zero or at the extreme 13. (3) The Incomplete-A binomial link seems to have a very flexible shape. The Inc-A binomial probability curves are close to (half) bell-shaped in all cases, especially for $p = 0.5$ and $\phi = 0.5$, which is the typical situation in toxicological experiments at high dose levels.

4. Linear combinations, Products and Composites

In this section, we demonstrate that the family of parsimonious distributions is closed under convex linear combinations, products and composites of CM links. As an application, we show the complete monotonicity of the links in Tables I and II. We now give a useful theorem, of which the first two results are the quick consequences of Criteria 1, 2 of Feller [25] (page 441), and the third is from Widder [29].

Theorem 4. *(LP) If φ, ψ are CM, then the convex linear combination $\alpha\varphi + (1-\alpha)\psi$ for $\alpha \in [0,1]$ and the product $\varphi\psi$ are also CM.*
(C.I) If h is CM and ψ is a positive function with a CM derivative, then the composite $h(\psi)$ of h and ψ is CM.
(C.II) If h is CM and φ is absolutely monotone*, then the composite $\varphi(h)$ of φ and h is CM.*

The following remark gives several useful functions, which satisfy Theorem 4 (C.I) and (C.II).

Remark 2. *(1)The logarithm function $\psi_1(t;\theta) = \theta\log(1+t)$ with $\theta > 0$, the power function $\psi_2(t;\nu) = t^\nu$ with $0 \le \nu \le 1$, and $\psi_3(t) = 1 - \exp(-t)$ are positive with CM derivatives.*
(2) The positive polynomial $\varphi_1(t;\theta) = \theta_1 + \theta_2 t + ... + \theta_k t^k$ for $\theta = (\theta_1, ..., \theta_k)^\top \in [0,\infty)^k$ and the exponential function $\varphi_2(t;\theta) = \theta^t$ with $\theta > 1$ are absolutely monotone.

Applying Theorem 4 (C.I) and (C.II) to the existing CM links, we can obtain numerous new CM links. For example, the composite of $h_s(s) = \exp(-s)$ and $\psi_2(t;\theta) = t^\theta$ yields the CM link, which is called the *positive stable* link, see Joe [8]. The composite of the logarithm $\ln(1+t)$ and the Gamma-Bin link gives the Gamma-Log-Bin link. Moreover, additional parameters can be introduced via the two types of composite of CM links. The composite of the power t^ν and the independence binomial link $h(t) = p^t$ gives the P-power link of Kuk [19]. The following remark applies to each link in Tables I and II.

Remark 3. *If t in Tables I and II is replaced with the logarithm function $\ln(1+t)$ or power function t^ν, then the resulting links are still CM and referred to as XX-Log-Bin links and XX-Power-Bin links respectively,*

where XX is the name of the mixing distribution such as Gamma, Poisson, etc.

Moments and Correlations. There are convenient formulas for the means and variances of the parsimonious exchangeable binomials.

Theorem 5. *Suppose that* $\mathbf{g}, \mathbf{h}$ *are two CM links. Then we have the following.*
(1) The mgf of $\mathbf{EB}(\mathbf{h})$ *is* $M_{\mathbf{h}}(t) = \sum_{k=0}^{n}(-1)^k\binom{n}{k}h_k(\theta)(1-e^t)^k$. *Hence, the mean is* $\mu_{1,\mathbf{h}} = mh_1(\theta)$ *and the variance is* $\sigma^2_{\mathbf{h}} = m(h_1(\theta) - h_2(\theta)) + m^2(h_2(\theta) - h_1^2(\theta))$.
(2) The mgf of $\mathbf{EB}(\alpha_1\mathbf{g} + \alpha_2\mathbf{h})$ *is* $M(t) = \alpha_1 M_{\mathbf{g}}(t) + \alpha_2 M_{\mathbf{h}}(t)$, *where* $\alpha_1, \alpha_2 \in [0,1]$, $\alpha_1 + \alpha_2 = 1$. *Hence, the first and second moments are* $\mu_1 = \alpha_1\mu_{1,\mathbf{g}} + \alpha_2\mu_{1,\mathbf{h}}$ *and* $\mu_2 = \alpha_1\mu_{2,\mathbf{g}} + \alpha_2\mu_{2,\mathbf{h}}$.

The Range of the Correlation. The kth order correlation ϕ_k of the exchangeable binary random variables $B_1, ..., B_m$ is an important quantity which was given in Theorem 2.2 of George and Bowman [3]. The second order correlation of the $\mathbf{EB}(\theta; \mathbf{h})$ model is

$$\phi(\theta) = \phi_2(\theta) = (h_2(\theta) - h_1^2(\theta))/(h_1(\theta) - h_1^2(\theta)), \quad \theta \in \Theta. \tag{4.25}$$

This is useful, for example, to test exchangeability. Specifically, $\phi(\theta) = 0$ if and only if the sequence $B_1, B_2, ...$ is independent and identically distributed, see the simulation in Section 6. It is easily calculated that the ranges of the correlation $\phi(\theta)$ for the Kuk's power family, the MM-power-Bin, the Gamma-power-binomial, and the A-power binomial are $[0, 1]$, which is the full range of the correlation for an exchangeable binary sequence. A parsimonious distribution with a full range [0, 1] can fit any possible correlation considered. The range of correlation $\phi(\theta_1, \theta_2, 1)$ for the Gamma-power binomial is only $[0, 1/2]$. This corresponds to the Gamma link normalized at $t_0 = 0$ in Remark 1. If the Gamma link is normalized at $t_0 = 1$ in Remark 1, then the range of the correlation $\phi(\theta_1, \theta_2, 1)$ reduces to $[0, 1/3]$.

5. Regression and Parameter Estimation

Suppose now that for a given litter size M, we have observation Y from an $\mathbf{EB}(\theta, M; \mathbf{h})$ distribution, associated with a covariate X. Typically X is a vector of factors such as dose, weight, etc., whereas M is often considered to be a random number such as in toxicity data, see Zhu, *et al.* [30]. In the simulation study in the next section, we treat M as random. Let $(Y_i, M_i, X_i), i = 1, ..., n$ be independent observations of (Y, M, X). Abusing notation, we write θ as (θ, ϑ), where $\theta \in \mathbb{R}$ is a parameter of interest, while ϑ is treated as a nuisance parameter. Allowing θ to depend on the linear systematic part $\eta = \boldsymbol{\beta}^\top X$ gives $\theta = \eta$, where $\boldsymbol{\beta} \in \mathbf{B}$ is a regression parameter for some nonempty open $\mathbf{B}$. A common and equivalent expression is

$$\lambda_j = h_j(\boldsymbol{\beta}^\top X; \vartheta), \quad j = 0, 1, ..., M.$$

These equations suggest two extensions of the proposed modeling from generalized linear models, namely, the exchangeable binomials generalize the binomial and the completely monotonic links generalize the usual Logistic, Probit or Log-Log links.

Relation to the Logistic Regression. By Theorem 2, the binomial model is a special case of the proposed incomplete exchangeable binomial model. Thus, regression under the proposed model simplifies to logistic regression via equality $\theta_1 = \theta_2$ when data are from binomial model. This property is useful in modeling. We can always fit a dataset with the proposed exchangeable binomials. If the dataset is indeed resulted from independent binary responses (i.e., the binomial model), then the estimates of the parameters $\hat{\theta}_1, \hat{\theta}_2$ of θ_1, θ_2 should be very close $\hat{\theta}_1 \approx \hat{\theta}_2$, so that the fitted model is simplified to a logistic model. Our simulations and applications in section 6 validate this fact. Another approach is to make use of the convex linear combination. A sample of independent binary responses should yield an estimator of the α to be close to zero.

Estimation and the Exponential Technique. Parameter estimation is based upon maximum likelihood. The usual MLE's of parameters are the solutions of score equations and the numerical solutions

are found via Newton-Raphson iteration. Under regularity assumptions, we have asymptotic normality of the MLE's, in particular, the asymptotic covariance matrix, which is used to calculate the standard errors of the MLE's. Note that with CM links $\mathbf{h}(\theta)$, the complicated complete monotonicity constraints on the high dimensional parameter $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_m)$ is reduced to simple box constraints, for example, the parameter θ must be inside $(0, \infty]$ or $(0, 1)$. For regression $\eta = \beta^\top X$, the box constraint requires the parameter β to be linearly constrained. The usual procedure does not give an efficient method to control the linear constraint. One way to find a constrained MLE is to use the subroutines such as *nlpnra* in SAS or *constrOptim* in R. The standard errors of the parameters are obtained by inverting the Hessian matrix. Another method for handling the constraints is to apply the following Exponential Technique. Replacing η with $\exp(\eta)$, the constrained maximization is simplified to an unconstrained one. Then the usual MLE procedure may apply. This is a straightforward application of Remark 2 in Theorem 4. Interestingly, the MM link applied with this exponential technique and normalized at $t_0 = 1$ in Remark 1 gives the commonly used logistic link used in generalized linear models.

Criteria for Model Comparison. We use the well-known Akaike and Bayesian information criteria (AIC and BIC) for model comparison. Recall

$$\text{AIC} = -2\log \text{L} + 2 * \text{npr}, \quad \text{BIC} = -2\log \text{L} + \text{npr} * \log(\text{nobs}),$$

where npr and nobs are the number of parameters and number of observations respectively. Both criteria reward goodness of fit and penalize the increasing number of parameters. The penalty discourages over-fitting of the model. Although the BIC has a heavier penalty than AIC, the preferred model is the one with the lowest AIC or BIC value. We also look at the response probability p and the affected-litter probability q. Models where expected probabilities are closer to observed probabilities are preferred.

Model Selection. The Laplace transforms (including moment generating functions, characteristic functions) together with convex linear combinations, products, and composites provide a vast family of parsimonious parametric binomial mixtures, furnishing various choices and great flexibility for statistical inference such as model fitting and regression. From a practitioner's point of view, it is always helpful to have some guidelines on how to obtain the best possible model from a family. Here we suggest a procedure, referred to as the *Forward Model Selection* procedure. We borrow the idea from the *forward variable selection* procedure commonly used in multiple linear regression. However, in our forward model selection procedure, increasing the complexity of the model means increasing the number of parameters rather than increasing the number of covariates.

We start from simple one-parameter links such as MM, MM-log, one-parameter Gamma, one-parameter Gamma-log-bin, Poisson-log, and positive stable link. According to the criteria such as AIC or BIC, the best two or three models are then selected. Next, consider two-parameter models which must include the previously selected models as sub-models. At this step, we may look at models resulting from linear combinations, power composites, and nonnegative polynomial composites of the previous selected links. We test the significance between the large and reduced model by the asymptotic likelihood ratio test (LRT). If the hypothesis is rejected, then more parameters are added to the models and the significance of additional parameters is tested until the hypothesis is accepted or the model reaches four parameters. A model with more than four parameters is of no practical interest due to its complexity. Using this procedure, we may find an optimal model with considerably less computing time than the procedure that tries all possible models. The whole procedure is easy to implement. The R codes of a relatively complete list of link functions are available at the website `http://www.olemiss.edu/~xdang`.

6. Simulations and Applications

In this section, we present a small simulation and a comprehensive study of the applications of the proposed framework to the 2, 4, 5-T and E2 data. We fit the E2 data with various parsimonious EB distributions and compare the results with existing models from the literature. We also conduct a regression analysis on the

2, 4, 5-T data based upon a variety of the parsimonious EB models and compare the results with existing models.

A Small Simulation. We conduct a small simulation to validate the inclusion of the binomial as a special case of the parsimonious incomplete EB. Namely, given the cluster size m, if the observations are resulted from independent binary responses (i.e., the binomial), then the incomplete EB will degenerate to the binomial through the equality $\theta_1 = \theta_2$ in the EB model. To mimic the real situation of unequal cluster sizes as in the CD1 mice data (see below), we first generate a cluster size M from a typical range, $[5, 15]$, of litter size. In fact, it is useful both in theory and application to treat litter size M as random, see, e.g., Stefanescu and Turnbull [21], Xu and Prorok [22], and Zhu, *et al.* [30]. Given M, we generate a sample of size 400 from the binomial with parameters $p = 0.1$ and M. Here we choose a small $p = 0.1$ because the estimation for small p may become difficult. We calculate the MLE's of the parameters (θ_1, θ_1) in the incomplete A-, Beta- and Gamma-binomials. They are $(\hat{\theta}_1, \hat{\theta}_2) = (0.0940, 0.0940)$, $(0.0940, 0.0940)$, and $(0.0941, 0.0944)$, respectively, all exhibiting $\hat{\theta}_1 \approx \hat{\theta}_2$. Moreover, they all have the same maximized log likelihood -488.69, the same estimate 0.094 of the true probability (i.e. 0.1) from $h_1(\theta)$, and the same estimate $\hat{\phi} \approx 0$ of the second order correlation. This, again, indicates that the data are from the binomial model.

Table III. **Fitting the E2 Data**. The observed values of $(p, q) = (0.110, 0.640)$.

Models (npr)	$\hat{p}(s.d.)$	$\hat{\phi}(s.d.)$	$\hat{q}$	$-2\log L$	AIC	BIC
Bin (1)	0.113(.006)	0.000		765.6	767.6	767.9
Correlated Bin (2)	0.131(.010)	0.073(.012)		720.5	724.5	731.2
Beta-Bin (2)	0.112(.009)	0.101(.017)	0.612	689.8	693.8	700.5
Two Bin (3)	0.111	0.114		682.4	688.4	698.5
Three Bin (5)	0.121	0.101		679.7	689.7	706.5
Beta-Bin with Bin (4)	0.135	0.189		680.2	688.2	701.6
Kuk's Q-power (2)	0.119	0.209	0.648	687.1	691.1	697.8
Kuk's P-power (2)	0.109	0.080	0.595	698.8	702.8	709.5
Gamma-Bin with $\theta_2 = 1$(1)	0.118(.0015)	0.191(.0009)	0.543	697.8	699.8	700.1
Gamma-Bin (2)	0.110(.0087)	0.093(.0181)	0.619	679.9	683.9	**684.5**
Inc. Gamma-Bin (4)	0.109(.0118)	0.101(.0301)	**0.633**	**675.2**	**683.2**	696.6
Piecewise-Flogit (1)	0.111(.0154)	0.112(.0362)	0.601	680.8	**682.8**	**683.1**
Piecewise-Flogit Power (2)	0.111(.0175)	0.112(.0411)	0.601	680.8	684.8	691.5
Inc. Beta-Bin (4)	0.110(.0055)	0.102(.0384)	0.632	676.4	684.4	697.8
Inc. A-Bin (2)	0.115(.0085)	0.096(.0172)	0.624	681.6	685.6	692.3

The estimated response probability $\hat{p}$, intra-litter correction $\hat{\phi}$, probability $\hat{q}$ of affected litters, along with negative twice log-likelihood, AIC and BIC. The upper, middle, and lower table are from Brooks *et* al. [24] and Brooks [31], Kuk [19], and the proposed framework, respectively. The standard errors are included in parentheses as they are available. The optimal models are in bold.

Fitting the E2 Data. The E2 data (Brooks *et al.* [24]) records fetal control mortality in mouse litters. There are 211 litters in total with litter sizes varying from as small as 3 to as large as 19, with mean litter size of 12.9 and a standard deviation of 2.68. The proportion of dead fetuses is 0.110. Among the 211 litters, there are 135 litters which have at least one fatal fetus.

We now fit the data and select an optimal model by the proposed forward model selection procedure. We start with one-parameter models. The two-parameter Gamma-binomial link with $\theta_2 = 1$ and the Piecewise Flogit link are the best two chosen by the principle of maximum log-likelihood. These two models fit the data well in terms of the response probability, but the estimates of the intra-correlation are quite different. The intra-correlation for the Gamma-binomial is 0.191, whereas the Piecewise Flogit model is 0.112. It seems that it is very difficult to assess which model is more accurate since we do not know the true value of ϕ, as the observed value of ϕ is difficult to obtain from the data. This raises the concern that too simple models

may not have additional parameters to flexibly model the correlation structure. This phenomenon was also pointed out by Kuk [19].

Next, we consider larger models which include the previously selected two models as sub-models. The composite of the Piecewise Flogit with the power function yields the Piecewise Flogit Power link (specifically, replace t in (2.11) with t^ν). The additional parameter ν, however, turns out to be 1 with the p-value close to 1. This is a strong indication in favor of the reduced model. The results for the two-parameter Gamma-binomial model are significantly different from the one-parameter Gamma-binomial model as indicated by the p-value 0.00002 of the maximized log-likelihood ratio test. In Table III, we observe that there are not only reductions in the values of the negative twice maximized log-likelihood, the AIC and BIC from the two-parameter Gamma-Binomial model to the one-parameter Gamma-Binomial model, the two-parameter Gamma-Binomial model also improves the estimates of the probability of affected litters over the one-parameter Gamma-Binomial model, namely from $\hat{q} = 0.543$ of the latter to $\hat{q} = 0.619$ of the former in estimating the observed $q = 0.640$.

We continue to consider larger models by including more parameters in the previously selected models. Two additional parameters in the incomplete Gamma-binomial slightly reduces the values of the negative twice maximized log-likelihood and the AIC from 679.9 and 683.9 of the Gamma-binomial to 675.2 and 683.2, respectively, of the incomplete Gamma-binomial, whereas the value of the BIC is increased from 684.5 of the Gamma-Bin to 696.6 of the incomplete Gamma-Bin. The p-value of the LRT is 0.095. All these indicate no strong evidence to differentiate the incomplete Gamma-Binomial from the simpler complete Gamma-Binomial. Thus the forward model selection procedure ends with three models: the Piecewise Flogit, the Gamma-binomial, and the incomplete Gamma-binomial, selected by the criteria of AIC, BIC and -2 log L. It is interesting to observe that the selected three models are related to the Gamma distribution.

As Zhu, *et al.* [30] pointed out, one of several respects unique in toxicity data is that the number M of fetuses per litter is random. Another quantity of interest related the probability (2.6) of affected litters is the number L of affected litters, which is also random. Under the "interpretability" assumption (see Stefanescu and Turnbull [21]), the conditional expectation of L given $M = m$ is

$$\mathbb{E}(L|M = m) = L(m)q(m),$$

where $q(m)$ is the conditional probability q in (2.6) of affected litters given $M = m$, and $L(m)$ is the number of affected litters among the litters with size m. Thus, the expected value of affected litters is

$$\mathbb{E}(L) = \mathbb{E}(L(M)q(M)) = \sum_m L(m)q(m) = \sum_m L(m)\left[1 - \sum_{k=0}^{m}(-1)^k\binom{m}{k}\lambda_k\right], \tag{6.26}$$

where the summation runs over all possible values of m, and λ_k's are the marginal probabilities. In the E2 data, m ranges from 3 to 19. The observed overall probability of affected litters is the proportion of the observed expected number of affected litters to the total number, 211, of the litters. This proportion in the E2 data is $q = 0.640$. Unlike the response probability and second order correlation coefficient, which are determined by λ_1 and λ_2 only, the probability of affected litters involves all the marginal probabilities λ_1,...,λ_m. Among all the models, the incomplete Gamma-Binomial model yields the closest estimate, $\hat{q} = 0.633$, of the probability of affected litters. The estimate of the probability of affected litters by the Piecewise Flogit model is 0.601, which underestimates the observed value of 0.640. This is perhaps due to the fact that too simple models may not have enough flexibility to fit all the λ_i's. Therefore, our final preferred model is the Gamma-binomial type models, the Gamma-binomial and incomplete Gamma-binomial, see Table III.

For comparison with the Beta-binomial model used by Brooks *et al.* [24] and Brooks [31], we have included in Table III the incomplete Beta-binomial model, even though it is not selected by the forward model selection procedure. Also included in Table III is the incomplete A-binomial for comparison with other incomplete models. The additional two parameters in the incomplete Beta-binomial give more modeling flexibility than the Beta-binomial. The negative twice maximized log-likelihood of the incomplete Beta-binomial is 676.4, which is smaller than the negative twice maximized log-likelihood 689.8 of the Beta-binomial. The p-value of the likelihood ratio test comparing the two models is 0.001, indicating significance of the additional two

Table IV. **Summary of the CD1 Data.**

Dose	Litters(n)	Implants(n)	Littersize(avg)	Littersize(std)	Malfs(n)	Malf(prop)
0	73	777	10.64	2.69	59	0.076
30	87	952	10.94	2.88	124	0.130
45	98	1124	11.47	2.93	338	0.301
60	76	806	10.61	2.99	390	0.484
75	44	482	10.95	2.23	372	0.772

parameters. Even though the Beta-binomial and incomplete Beta-binomial yield similar estimates of the fetal probability and intra-litter correlation, the incomplete Beta-binomial provides a better estimate of the probability of affected litters. Indeed, the incomplete Beta-binomial gives the estimate, 133.1, of the expected number of affected litters, which agrees with the observed number, 135, of affected litters, whereas the Beta-binomial underestimates this quantity with 129.1, out of a total 211 litters. One reason, as pointed by Kuk [19], is due to the fact that the Beta-binomial overestimates the probability that none of litter mates is affected and hence underestimates the number of the litters affected. The additional parameters in the incomplete Beta-binomial provide more fitting flexibility and mitigate the problems of the the Beta-binomial. Another possible reason is that the fetal probability and intra-litter correlation are determined by up to two marginal moments (i.e. λ_1, λ_2), whereas the expected number (6.26) of affected litters involves all the marginal moments λ_i's. It should be noted that among all the incomplete links, the incomplete A-link is appealing because of its model simplicity, i.e., fewer number of parameters.

Reported in Table III are the MLE's of the marginal response probability p, the intra-litter correlation ϕ, the probability q of affected litters, -2*log-likelihood, AIC and BIC, along with the standard errors of $\hat{p}$ and $\hat{\phi}$. The standard errors are computed by the Delta method based on the asymptotic normality of the MLE. Comparing the results with Brook's and Kuk's models, all the proposed models adequately fit the data in terms of the values of the likelihoods, AIC and BIC as well as the matching of the estimated response probabilities with the observed ones. Note, however, that both Brook's and Kuk's models are special cases of our proposed parsimonious EB's.

Regression Analysis on the CD1 Data. We now apply the proposed regression procedure to analyze a real dataset from a developmental toxicology study conducted at the National Center for Toxicological Research. The study involves replicate experiments with 9 strains of female mice exposed to the herbicide 2,4,5-Trichlorophenoxyacetic acid. We use the data for the CD1 mice, which was analyzed by many authors, e.g., George and Bowman [3] and Kuk [19]. As Gaylor and Razzaghi [32] point out, the classical approach to bioassay presents a particular problem. The investigator in an animal study is usually interested in the results for low doses, where typical human exposure occurs. In a bioassay, though, there will be few responses in the treatment groups given low doses of the toxin. Thus, the parameter estimation for the dose-response curve is greatly affected by what happens at the high doses. However, two different distributions that fit the observed data adequately at the high doses may give very different estimates at the lower end of the curve. Furthermore, many animals are unnecessarily sacrificed at high doses, where response rates are high. This raises the question of animal allocation. Thus, we shall focus on the dose levels 0, 30, 45, 60, and 75 (mg/kg/day), excluding the high dose level 90 at which more than 95% of the fetuses were malformed. Table IV reports a summary of the CD1 data.

We shall follow the proposed forward model selection procedure to select possible optimal models. Starting from one-parameter links, we model parameter θ to linearly depend upon dose level D, i.e., $\theta = \alpha + \beta D$, where α, β are regression parameters. Among the one-parameter links, by the criterion of maximal log-likelihood, we select the Gamma-binomial link $(1 + \theta_2 t)^{-\theta_1}$ and Gamma-log-binomial link $(1 + \theta_2 \ln(1 + t))^{-\theta_1}$, both with $\theta_2 = 1$, denoted by Link1 and Link2, respectively in Tables V and VI.

The Gamma-type models outperform others in modeling this data. This is also confirmed by the stronger linearity demonstrated in the curves of the inverse link function of the observed probability of malformations versus dose levels in the left panel of Fig. 2. Each one-parameter link evaluated at $j = 1$ gives an estimate of the response probability p, namely, $p = \lambda_1 = h_1(\alpha + \beta D)$, so that $\alpha + \beta D = g(p)$, where $g = h_1^{-1}$ is the

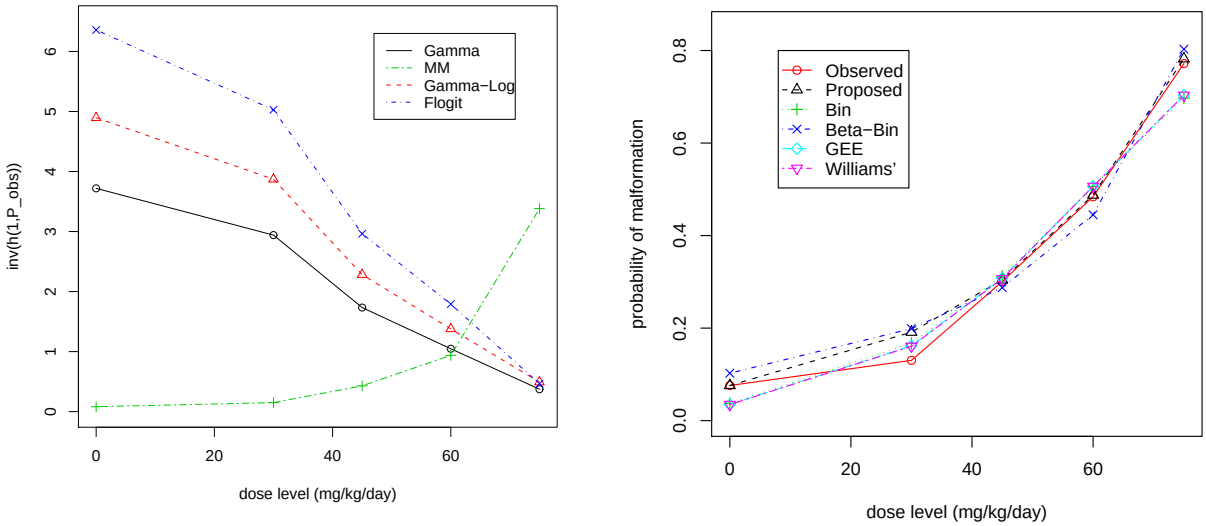


Figure 2. Left panel: The curves of the inverse links of the observed probability of malformations versus dose level for the CD1 data. Stronger linearity of the Gamma link than the MM link may serve as an indicator of its better suitability. Right panel: The probability curves of malformation versus dose levels under different models for the CD1 data. The curve of the proposed model closely matches the curve of the observed and outperforms the others except a slight overestimation at dose level 30 mg/kg/day.

inverse link. The left panel in Fig. 2 can be used to check the linearity assumption of the parameter θ on dose D. Here we look at four one-parameter models. From the curves, we observe that the linearity in the MM link is the weakest. The strongest linearity is demonstrated by the Gamma-binomial, followed by the Gamma-log-binomial. The latter two models are the ones selected by the AIC or BIC.

Next, we consider larger models, Link3=$(1+\theta_2 t)^{-\theta_1}$, Link4=$(1+\theta_2 \ln(1+t))^{-\theta_1}$, and the linear combination of Link1 and Link2, a*Link1+(1-a)*Link2 with $0 \leq a \leq 1$. They, of course, include the previous Link1 or Link2 as sub-models. We test the null hypothesis $\theta_2 = 1$ in Link3 and Link4 by the LRT and find that θ_2 is not significant in Link3 but *significant* in Link4 at an extremely high significant level p value ≈ 0. The MLE of a is $\hat{a} = 0.678$ with a standard error of 0.043. This value is not near 0 or 1, indicating a significant difference between the large and reduced models. Again, this significance is also confirmed by the LRT.

The linear combination of Link3 and Link4 includes Link3 and Link4 as sub-models with an extra parameter, a, and also reduces to the linear combination of Link1 and Link2 with additional parameter, $\theta_2 = 1$. The MLE's are $\hat{\theta_2} = 0.491(0.062)$ and $\hat{a} = 0.443(0.033)$. The maximized log-likelihood of the large model increases to -719.3. The LRT's have a significance level of 0.007 for θ_2 and reject $a = 0$ and $a = 1$ at the p-values ≈ 0, respectively. The procedure is terminated at this step due to model complexity. By the forward model selection procedure, we now conclude that the possible optimal link is the linear combination of the Gamma-binomial and Gamma-log-binomial links. From Table VI and the right panel of Fig. 2, we observe that the estimates from the proposed optimal model closely matches the observed values and outperforms the others except a slight overestimation at dose level 30 mg/kg/day.

In addition to the linear combinations, we also suggest examining the product, since it sometimes yields a better fit without introducing new parameters. In this case, we cannot directly use the LRT for model comparison because they are not nested. But other criteria are possible, for example, *the sum of the absolute differences between the estimated and observed number of malformations*, or *the square root of the sum of the squared differences*, which will be introduced below. The product model of Link3 and Link4 is not significantly different from the product of Link1 and Link2 by the LRT. These models together with the modified (piecewise) folded logistic link are reported in Table V. We include the modified folded logistic here not just for illustration but for its simplicity and relatively good behavior in fitting the data, even though it neither constitutes nested-relation with other models nor is chosen by the criteria.

Table V. **The Estimates Under Various Models for the CD1 Data**.

Model (npr)	α(s.d.)	β(s.d.)	-2logL	χ^2	AIC	BIC
Binomial (2)	-3.235(.113)	5.430(.217)	2295.0	1514.2	2298.9	2306.8
Beta-Bin (4)	0.433(.081)	3.788(.194)				
	0.503(.064)	-4.786(.118)	1464.8	336.43	1472.8	1488.5
GEE(Logit, Ex) (3)	-3.323(.205)	5.580(.438)		411.43		
Williams' (3)	-3.237(.231)	5.587(.442)		370.87		
Kuk's Q-power (3)	0.884(.071)	-0.525(.115)	1459.6	298.95	1465.6	1471.3
Link1 (2)	3.838(.174)	-4.712(.265)	1460.8	372.22	1464.8	1472.7
Link2 (2)	4.633(.259)	-5.447(.411)	1472.2	247.64	1476.2	1484.1
Link3 (3)	3.340(.161)	-4.082(.243)	1458.9	343.81	1464.9	1476.7
Link4 (3)	15.03(.759)	-18.16(1.16)	1450.1	312.63	1456.1	1467.9
aLink1+$(1-a)$Link2 (3)	4.116(.203)	-4.975(.315)	1445.8	310.53	1451.8	1463.6
aLink3+$(1-a)$Link4 (4)	7.884(.640)	-9.571(.878)	**1438.6**	338.84	**1446.6**	**1462.3**
Link1*Link2 (2)	2.119(.105)	-2.568(.161)	1459.4	309.66	1463.4	1471.3
Link3*Link4 (3)	2.404(.062)	-2.927(.292)	1458.4	326.75	1464.5	1476.2
Piecewise-Flogit (2)	6.792(.265)	-8.441(.393)	1501.8	510.81	1505.8	1513.7

Link1: $(1+j)^{-\theta_1}$, Link2: $(1+\ln(1+j))^{-\theta_1}$,
Link3: $(1+\theta_2 j)^{-\theta_1}$(Gamma-Bin), Link4: $(1+\theta_2\ln(1+j))^{-\theta_1}$(Gamma-log-Bin).

Table VI. **The Expected Number of Malformations of the CD1 Data**.
d_i: difference of the expected & observed number of malformations in the i^{th} dose group.

	Dose level (mg/kg)						
Models	0	30	45	60	75	$\sum\lvert d_i\rvert$	$\sqrt{\sum d_i^2}$
Observed	59	124	338	390	372		
Binomial (Logit)	29	159	351	408	336	132	62.56
Beta-binomial	80	189	323	359	387	147	77.96
GEE (Logit,ex)	27	154	346	408	339	121	58.32
Williams'	27	153	345	408	339	119	57.68
Kuk's Q-power	90	260	396	347	246	394	201.36
Link1	54	177	342	400	390	90	57.22
Link2	68	196	356	393	361	113	75.62
Link3	56	180	344	400	387	90	59.21
Link4	58	183	345	396	378	79	60.02
aLink1+$(1-a)$Link2	59	182	342	393	377	70	58.43
aLink3+$(1-a)$Link4	58	179	339	391	379	**65**	**55.47**
Link1*Link2	59	184	347	398	381	86	61.85
Link3*Link4	57	182	345	399	384	88	60.35
Piecewise-Flogit	50	169	333	400	405	102	57.61

Link1: $(1+j)^{-\theta_1}$, Link2: $(1+\ln(1+j))^{-\theta_1}$,
Link3: $(1+\theta_2 j)^{-\theta_1}$(Gamma-Bin), Link4: $(1+\theta_2\ln(1+j))^{-\theta_1}$(Gamma-log-Bin).

We now compare our proposed models with the existing ones: the binomial, the Beta-binomial, the GEE approach, Williams's procedure and Kuk's power family. Reported in Table V are the MLE's of the parameters under various models: the Pearson χ^2, the AIC and BIC for the goodness of fit and model comparison. The Pearson χ^2 statistic, defined as $\chi^2=\sum_{i=1}^{n}(Y_i-\hat{Y}_i)^2/\widehat{\mathbb{V}\mathrm{ar}}(\hat{Y}_i)$, is asymptotically distributed as a Chisquare $\chi^2(n-p)$, under mild regularity assumptions, where p is the number of fitted parameters.

The binomial model is a poor fit because it completely ignores the intra-litter correlation. The Pearson

χ^2 value of the binomial model is 1514.2 in Table V, which is much greater than its expected value 376. This indicates strong evidence of overdispersion in this data. The Beta-binomial model assumes, in each dose group, that the probability of malformation has a Beta binomial distribution with parameter θ_1, θ_2, which are modeled as

$$\theta_1 = \alpha_1 + \alpha_2 D, \quad \theta_2 = \beta_1 + \beta_2 D.$$

The MLE's of $\alpha_1, \alpha_2, \beta_1, \beta_2$ constrained on $\theta_1, \theta_2 > 0$ are reported in Table V. For the GEE approach, we use the R package *geepack*. We specify the logistic link on the mean response, and the correlation structure as "exchangeable". The estimate of the intra-litter correlation is 0.2647 with a standard deviation of 0.1166. Williams [16] presented an iterative algorithm to estimate the extra-binomial variation, ϕ, which was incorporated into the reweighed least squares procedure. Here we use the software Arc developed by Cook and Weisberg [33] to get the results of Williams' procedure. The overdispersion parameter in Williams' procedure is 0.2997 with a standard deviation of 0.1231. It was estimated by equating the Pearson χ^2 to its degrees of freedom, so that the goodness of fit statistic in Williams' procedure is no longer informative. As anticipated, the GEE and Williams' procedures produce similar results because both only model the first two moments. With Kuk's Q-power model, we directly model $q = \alpha + \beta D$, and treat ν as a nuisance parameter. The MLE of ν is $\hat{\nu} = 0.608$ with a standard deviation of 0.022 and the MLEs $\hat{\alpha}, \hat{\beta}$ of α, β are reported in Table V. Although Kuk's Q-power can achieve a relatively high likelihood, it gives a poor estimate to the number of malformations in each dose level; a severe over-estimation occurs at dose level 30 mg/kg and a under-prediction at dose level 75 mg/kg. In this example, we have found that the AIC and BIC are sometimes not reliable for model comparison. The AIC and BIC criteria should be used in conjunction with other criteria.

We use two other criteria to assess the models. Both criteria consider overall discrepancy between the estimated and observed numbers of malformations. The first is the *sum of the absolute differences between the estimated and observed numbers of malformations* across all dose groups. From Table VI, we can see that most results under the proposed models have superior estimates to the binomial, the Beta-binomial, the GEE and Williams' procedures. The second criterion is the *square root of the sum of squared differences*. The proposed models still exhibit competitive performance among the existing models. In particular, the model resulting from the linear combination of the Gamma-Bin and Gamma-log-Bin links, chosen by the forward model selection procedure, demonstrates the best performance by the two criteria, the AIC and BIC.

7. Concluding Remarks

In this article, we have proposed a general approach to analyze exchangeable binary response data. First, the proposed approach can be used to model *correlated and over-dispersed binary response data*. Our viewpoint is based upon the relationships among infinite exchangeability, finite exchangeability, complete monotonicity and binomial mixtures. Second, our approach unifies existing models including the binomial, the beta-binomial, Kuk's power family, Brook's correlated binomial, random effects models, and others. Our approach provides methods for finding new models either from existing Laplace transforms, moment generating functions and characteristic functions, or from existing completely monotonic links via linear combinations, products or composites. The incomplete parsimonious exchangeable binomials are new and have demonstrated great modeling capability. Third, we have presented and demonstrated, with real data, a forward model selection procedure, which may serve as a guideline to practitioners for building a possible optimal model. Last, we have applied the proposed framework to two real datasets from developmental toxicology and compared the results with commonly used procedures. Our proposed models exhibit great modeling flexibility and superior performance, the Gamma-type models have particularly demonstrated well.

Acknowledgments We would like to thank Dr. Xueqin Wang for insightful discussions. We are also grateful to Dr. Eric Adams for his help in improving the English language of this article.

Appendix: Technical Details

Here we collect some technical details. We first give the following proposition.

Proposition 1. *The folded logistic link (2.10) is completely monotone on* $0 < \theta \leq 1$ *and not completely monotone on* $\theta > 1$. *The piecewise folded logistic link (2.11) is completely monotone on* $\theta > 0$.

Proof. Clearly for $\theta = 1$, the folded logistic link is completely monotone. Suppose $0 < \theta < 1$. Then $\psi(t) = t^\theta$ is a positive function and its derivative $\psi'(t) = \theta t^{\theta-1}$ is obviously completely monotone. Let $h(t) = 1/(1+t)$. Then h is completely monotone. From Theorem 4 (C.I) it follows that the composite $h(\psi(t+1)) = 1/(1+(t+1)^\theta)$ is completely monotone. This shows that the folded logistic link is CM on $0 < \theta \leq 1$. This also shows the complete monotonicity of the piecewise folded logistic link (2.11) on $0 < \theta \leq 1$, whereas on $\theta > 1$, the link is the Gamma-binomial link. Suppose now $\theta > 1$. By the Taylor expansion, we have

$$h_\theta(t) = t^{-\theta}(1+t^{-\theta})^{-1} = \sum_{k=1}^{\infty}(-1)^{k-1}t^{-k\theta}, \quad t \geq 2, \quad \theta > 0.$$

Then $(-1)^n h_\theta^{(n)}(t) = \sum_{k=1}^{\infty}(-1)^{k-1}a_{n,k}(t,\theta)$, where $a_{n,k} = (k\theta)(k\theta+1)...(k\theta+n-1)t^{-k\theta-n}$. Since $(k\theta+j)/((k+1)\theta+j)$ is decreasing in θ, it follows that for $\theta \geq 1$ and arbitrarily fixed $t \geq 0$ and every $k \geq 1$,

$$\begin{aligned}\frac{a_{n,k}(t,\theta)}{a_{n,k+1}(t,\theta)} &= \frac{k\theta}{(k+1)\theta} \cdot \frac{k\theta+1}{(k+1)\theta+1} \cdots \frac{k\theta+n-1}{(k+1)\theta+n-1}t^\theta \\ &\leq \frac{k}{k+1} \cdot \frac{k+1}{k+2} \cdots \frac{k+n-1}{k+n}t^\theta = \frac{k}{k+n}t^\theta \to 0, \quad n \to \infty.\end{aligned}$$

Thus $(-1)^n h_\theta^{(n)}(t) < 0$ for large n. This shows that the folded logistic link is not completely monotone on $\theta > 1$ and the proof is complete. □

Proof of Theorem 3(1)–(4). Set $dQ_\theta(p) = \mathbf{1}[\theta_1, \theta_2]/\bar{h}_{t_0}(\theta)dQ_\vartheta(p)$ with $\theta = (\theta_1, \theta_2, \vartheta) \in [0,1]^2 \times \Theta$. Then this determines a probability measure on $[0,1]$ and rewrites (2.18) as $\bar{h}(t;\theta) = \int_0^1 p^t\, dQ_\theta(p)$ for $\theta \in \Theta$. Note that the distribution Q_θ is uniquely determined by the infinite completely monotonic sequence $\mathbf{h} = \{h_k(\theta), k = 1, 2, ...\}$. By Kendall [2], there exists a sequence of exchangeable events $A_1, A_2, ...$ such that $\bar{h}(k;\theta) = \mathbb{P}_\theta(A_{r_1} \cap A_{r_2} ... \cap A_{r_k})$ and $\mathbb{P}_\theta(A_{r_1} \cap A_{r_2} ... \cap A_{r_k}|p) = p^k$ almost surely, where $p \sim G_\theta$. Here $r_1, ..., r_k$ are different. Let $\tilde{B}_i = \mathbf{1}[A_i]$ be the indicator function of set A_i and $\tilde{Y}_m = \tilde{B}_1 + ... + \tilde{B}_m$. Then the distribution of $\tilde{Y}_m$ is given by (2.7), so that $\tilde{Y}_m$ and Y_m have the same distribution. Abusing notation to write $\tilde{B}_i = B_i$, we obtain a stochastic representation $Y_m = B_1 + ... + B_m$, where $B_1, B_2, ...$ is the infinite sequence of exchangeable binary random variables given in the early part of Section 2. Further, conditional on $p \sim G_\theta$, B_1, ..., B_m are independent and have a common Bernoulli distribution with probability p of success. Accordingly,

$$\mathbb{P}(Y_m = y|p) = \binom{m}{y}p^y(1-p^y)^{m-y}, \quad y = 0, 1, ..., m, \quad a.s.$$

The rest of the proof is straightforward verification analogous to Srivastava and Wu [34] except (5) which is proved below.

Proof of Theorem 3 (5)(Asymptotic Distribution of the Mean Y_m/m). By the de Finetti theorem, we find the moment generating function of Y_m/m,

$$M_{Y_m/m}(t) = \int_0^1 (pe^{t/m} + 1 - p)^m\, dQ_\theta(p) \to \int_0^1 e^{pt}\, dQ_\theta(p) \equiv M_{Q_\theta}(t), \quad \text{say},$$

where, to claim the limit, we used the Taylor approximation $e^{t/m} = 1 + t/m + o((t/m)^2)$ and the limit $[1 + pt/m + o((t/m)^2)]^m \to e^{pt}$ as $m \to \infty$. Note that M_{Q_θ} is the mgf of Q_θ. Hence, it follows from the continuity theorem that Y_m/m converges in distribution to the mixing distribution Q_θ. □

REFERENCES

1. Frey R and McNeil AJ. Dependent defaults in models of portfolio credit risk. *J. Risk* 2003; **6**.
2. Kendall DG. On finite and infinite sequences of exchangeable events. *Studia Scientiarum Mathematicarum Hungarica* 1967; **2**: 319-327.
3. George EO and Bowman D. A Full likelihood procedure for analyzing exchangeable binary data. *Biometrics* 1995; **51**: 512-523.
4. Chow YS and Teicher H. *Probability Theory: Independence, Interchangeability, Martingales*. Springer, 2003.
5. Conaway MR. A random effects models for binary data. *Biometrics* 1990; **46**: 317-328.
6. Fanaroff, AA. A Controlled trial of intravenous immune globulin to reduce nosocomial infections in very-low-birth-weight infants. *New Engl. J. Medicine* 1994; **330**: 1107-1113.
7. Collett D. *Modelling Binary Data*. Chapman & Hall/CRC, 2003.
8. Joe H. *Multivariate Models and Dependence Concepts*. Chapman & Hall, 1997.
9. Neuhaus JM. Statistical methods for longitudinal and clustered designs with binary responses. *Statist. Meth. in Medical Research* 1992; **1**: 249-273.
10. Williams D. The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics* 1975; **31**: 949-952.
11. Kupper L and Haseman J. The use of a correlated binomial model for the analysis of certain toxicological experiments. *Biometrics* 1978; **34**: 69-76.
12. Stiratelli R, Laird N and Ware JH. Random-effects models for serial observations with binary response. *Biometrics* 1984; **40**: 961-971.
13. Ochi Y and Prentice RL. Likelihood inference in a correlated probit regression model. *Biometrika* 1984; **71**: 531-543.
14. Coull BA and Agresti A. Random effects modeling of multiple binomial response using the multivariate binomial logit-normal distribution. *Biometrics* 2000; **56**: 73-80.
15. Follmann DA and Lambert D. Generalizing logistic regression by nonparametric mixing. *J. Amer. Statist. Assoc.* 1989; **84**: 295-300.
16. Williams D. Extra-binomial variation in logistic linear models. *Appl. Statist.* 1982; **31**: 144-148.
17. Liang KY, Qaqish B and Zeger S. Multivariate regression analyses for categorical data. *J. Royal Statist. Soc., Ser. B* 1992; **54**: 3-40.
18. Lipsitz SR, Laird NM and Harrington DP. Generalized estimation equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika* 1991; **78**: 153-160.
19. Kuk AYC. A litter-based approach to risk assessment in developmental toxicity studies via a power family of completely monotone functions. *Appl. Statist.* 2004; **53**: 369-386.
20. Mislevy RJ. Estimation of latent group effects. *J. Amer. Statist. Assoc.* 1985; **80**: 993-997.
21. Stefanescu C and Turnbull BW. Likelihood inference for exchangeable binary data with varying cluster sizes. *Biometrics* 2003; **59**: 18-24.
22. Xu J and Prorok CP. Modelling and analyzing exchangeable binary data with random cluster sizes. *Statist. in Medicine* 2003; **22**: 2401-2416.
23. Altham P. Two generalizations of binomial distribution. *Appl. Statist.* 1978; **27**: 162-167.
24. Brooks SP, Morgan B, Ridout M and Pack S. Finite mixture models for proportions. *Biometrics* 1997; **53**: 1097-1115.
25. Feller W. *An Introduction to Probability Theory and Its Applications*. Wiley: New York, 1971.
26. Prentice R. Correlated binary regression with covariates specific to each binary observation. *Biometrics* 1988; **44**: 1033-1048.
27. Kingman JFC. The 1977 Wald memorial lectures: uses of exchangeability. *Ann. Probabil.* 1978; **6**: 183-197.
28. Prentice R. Binary regression using extended beta-binomial distribution with discussion of correlation induced by covariate measure errors. *J. Amer. Statist. Assoc.* 1986; **81**: 321-327.
29. Widder D. *The Laplace Transforms*. Princeton University Press, 1941.
30. Zhu Y, Krewski D and Ross WH. Dose-Response Models for Correlated Multinomial Data from Developmental Toxicity Studies. *Appl. Statist.* 1994; **43**: 583-598.
31. Brooks SP. On Bayesian analysis and finite mixtures for proportions. *Statist. and Comput.* 2001; **11**: 179-190.
32. Gaylor DW and Razzaghi M. Process of building biologically based dose-response models for developmental defects. *Teratology* 1992; **46**: 573-581.
33. Cook D and Weisberg S. *Applied Regression Including Computing and Graphics*. Wiley: New York, 1999.
34. Srivastava MS and Wu Y. Local efficiency of moment estimators in Beta-binomial model. *Commun. Statist.–Theory Meth.* 1993; **22**: 257-261.