A Taxonomy of Mixing and Outcome Distributions Based on Conjugacy and Bridging

Peer-reviewed author version

Kenward, Michael G. & MOLENBERGHS, Geert (2016) A Taxonomy of Mixing and Outcome Distributions Based on Conjugacy and Bridging. In: COMMUNICATIONS IN STATISTICS-THEORY AND METHODS, 45 (7), pag. 1953-1968.

DOI: 10.1080/03610926.2013.870205
Handle: http://hdl.handle.net/1942/16505

A Taxonomy of Mixing and Outcome Distributions Based on Conjugacy and Bridging

Michael G. Kenward[1] **and Geert Molenberghs**[2]

[1] *I-BioStat, Universiteit Hasselt and Katholieke Universiteit Leuven, B-3590 Diepenbeek, Belgium*

[2] *Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London WC1E7HT, United Kingdom*

Abstract

The generalized linear mixed model is commonly used for the analysis of hierarchical non-Gaussian data. It combines an exponential family model formulation with normally distributed random effects. A drawback is the difficulty of deriving convenient marginal mean functions with straightforward parametric interpretations. Several solutions have been proposed, including the marginalized multilevel model (directly formulating the marginal mean, together with a hierarchical association structure) and the bridging approach (choosing the random-effects distribution such that marginal and hierarchical mean functions share functional forms). Another approach, useful in both a Bayesian and a maximum likelihood setting, is to choose a random-effects distribution that is conjugate to the outcome distribution. In this paper, we contrast the bridging and conjugate approaches. For binary outcomes, using characteristic functions and cumulant generating functions, it is shown that the bridge distribution is unique. Self-bridging is introduced as the situation in which the outcome and random-effects distributions are the same. It is shown that only the Gaussian and degenerate distributions have well-defined cumulant generating functions for which self-bridging holds.

Some Keywords: Cauchy distribution; Characteristic Function; Cumulant; Degenerate Distribution; Identity Link; Logit link; Log link; Marginalization; Mixed Models; Mixture Distribution; Probit Link; Random Effects; Random-effects Distribution.

1 Introduction

The class of generalized linear mixed models (GLMM; Breslow and Clayton 1993, Wolfinger and O'Connell 1993; Molenberghs and Verbeke 2005) is now a standard framework for handling hierarchical data of a non-Gaussian nature. In this a generalized linear model, based on the exponential family and conditional upon the random effects, is combined with a Gaussian distribution for the latter.

While convenient and flexible, a commonly-quoted drawback for such models is the lack of a closed-form marginal representation. Zeger, Liang, and Albert (1988) derived exact marginal mean functions for certain cases, in particular the probit-normal and Poisson-normal, as well as an approximate marginal mean for the logit-normal case. Molenberghs, Verbeke, and Demétrio (2007) and Molenberghs *et al* (2010) extended the GLMM to incorporate overdispersion random effects, typically chosen to be conjugate to the outcome distribution, that allow a more flexible accommodation of overdispersion (McCullagh and Nelder 1989, Hinde and Demétrio 1998ab). For example, the Poisson-normal model is extended with gamma random effects, and termed the Poisson-gamma-normal combined model. Other instances include the logit-beta-normal and probit-beta-normal model for binary data, and the Weibull-gamma-normal for time-to-event responses. In their work, these authors derived more general marginalization results, with closed-form marginal means, variances, higher-order moments, and even joint distributions. Predictably, the sole exception is the probit-beta-normal model, for which only approximate expressions exist. But, even when closed forms arise, they typically fail to provide functions with convenient parameterizations. For example, the Poisson-gamma-normal model in general, and the Poisson-normal GLMM in particular, do not produce a single group effect parameter in nearly all the cases for which the hierarchical model nevertheless does contain such a parameter. The same is true for all other instances, the sole exception being the linear mixed model (Laird and Ware 1982, Verbeke and Molenberghs 2000).

This is the reason why alternative approaches have been proposed that facilitate the derivation of marginal functions and corresponding marginal parameters.

First, Heagerty (1999) and Heagerty and Zeger (2000) proposed the *marginalized multilevel model* (MMM), by specifying simultaneously the marginal mean and the mean conditional upon the random effects. They achieve this by including a *connector function* that is derived by solving an integral equation. Second, Wang and Louis (2003, 2004) proposed, for hierarchical binary data and, in the first instance restricted to a random intercept, an approach where both the hierarchical and the marginal mean is of a logistic form. To achieve this goal, they chose a particular form for the random-intercept distribution, termed the $bridge$ distribution, which is derived through the solution of an integral equation. As well as the logit link, they also consider the probit, complementary log-log, and Cauchy links. Iddi and Molenberghs (2011) brought together overdispersion correction and MMM, into a so-called COMMM.

Molenberghs *et al* (2012) built upon this work in several ways. First, they showed that the bridge distribution in the binary-data case is unique. Second, also for the binary case they established that marginalizing a GLMM, using the MMM framework, and using the bridge concept results in three different approaches. Third, they derived bridge distributions for several further links and/or for vector rather than scalar random effects. Fourth, they showed that for the identity and logarithmic links, vast classes of bridge distributions exist. Fifth, they formally established a relationship between the three operations mentioned above: (1) marginalizing a GLMM or a combined model; (2) finding the connector function for a MMM or a COMMM; and (3) deriving the bridge distribution. They showed that, for the log and identity links, used when outcome variables have supports in the form of a half line and the real line, respectively, that the three specifications are identical in a number of situations and exhibit close connections in others. Sixth, they applied the various approaches to two sets of data, one with a binary outcome, the other with counts. They made generic SAS code available for all three approaches and for both data types.

Wang and Louis (2003) and Molenberghs *et al* (2012) used, for binary outcomes, the Fourier transform and its inverse to compute bridge distributions. While elegant as a general solution, it is not straightforward in practice even for fairly standard distributions when used as inverse link functions. This is already clear from the motivating case in Wang and Louis (2003), i.e., the logit link. In this paper, progress is made by exploiting the near identity between a Fourier transform and a characteristic function of a distribution. Some distributions admit a fairly simple characteristic function, in spite of a complicated density form, or vice versa. Moreover, when the characteristic function is expanded using a version of the Gram Charlier A series, i.e., admits an exponentiated cumulant generating function, further progress can be made. After this, two important problems are considered. First, the bridge problem can be reversed: the random-effects density is given, and a link function is then derived. Second, uniqueness results are derived.

The aim of this paper is to further develop the concepts of bridging, inverse bridging, and self bridging, exploiting properties of the characteristic function and its series expansion. It is not our intention to provide additional applications; for those interested in these we refer to Molenberghs *et al* (2012), who provide sufficient tools to apply the models discussed here. In the next section, we review existing work on mixing and mixing distributions, as well as the three operations of marginalizing a GLMM, deriving an MMM, and deriving a bridge distribution. In Section 3, we then introduce the characteristic function in this context. Inverse bridging and uniqueness results are studied in Section 4. These results are illustrated by means of several commonly encountered distributions in Section 5. The concept of self-bridging is introduced and existence and uniqueness results derived in Section 6.

2 Mixing Distributions

2.1 The Concept of Mixing

To fix notation, suppose a hierarchical outcome Y_{ij} is measured for each independent unit $i = 1, \ldots, N$ at occasion $j = 1, 2, \ldots, n_i$. The response for the ith unit, $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})^T$ is assumed to follow a distribution $f(\boldsymbol{y}_i|\boldsymbol{b}_i)$, where $\boldsymbol{b}_i$ is a random effect, also called mixing variable. We allow for the special case that $\boldsymbol{Y}_i$ is scalar and hence $n_i = 1$. The *conditional distribution* $f(\boldsymbol{y}_i|\boldsymbol{b}_i)$ is allowed to depend on, perhaps unknown, parameters. These are not essential here and will be dropped from notation to the extent possible. By $f(\boldsymbol{b}_i)$ we refer to the *mixing distribution*. The *joint distribution* is $f(\boldsymbol{y}_i, \boldsymbol{b}_i)$, which also admits the reverse factorization into the *marginal distribution* $f(\boldsymbol{y}_i)$ and the *predictive distribution* $f(\boldsymbol{b}_i|\boldsymbol{y}_i)$. In Bayesian terms, the mixing distribution is referred to as the prior, with the predictive distribution the posterior. Continuing in the Bayesian setting, the marginal distribution is termed the data distribution and the conditional one the likelihood.

$$\begin{array}{ccccccccc} \text{conditional/lik} & & \text{mixing/prior} & & \text{joint} & & \text{marginal/data} & & \text{predictive/posterior} \\ f(\boldsymbol{y}_i|\boldsymbol{b}_i) & \cdot & f(\boldsymbol{b}_i) & = & f(\boldsymbol{y}_i, \boldsymbol{b}_i) & = & f(\boldsymbol{y}_i) & \cdot & f(\boldsymbol{b}_i|\boldsymbol{y}_i). \end{array} \tag{1}$$

The above notation refers either to densities or probability mass functions, depending on whether the corresponding arguments are continuous or discrete. Generically, we will refer to them as distributions and the set of five will be termed a *quintuple*.

In likelihood terms, the conditional and mixing distributions are specified, but inferences made on the marginal distribution. Bayesians will formulate a prior and combine it with the likelihood to draw inferences in terms of the posterior. These conventional routes imply that one needs to pass from some components in the quintuple to others. Depending on the parametric forms chosen, such operations may be simple or cumbersome. We turn to this next.

2.2 Operations on the Quintuple

When both the conditional and the mixing distribution are chosen to be normal, the so-called linear mixed model results (Laird and Ware 1982, Verbeke and Molenberghs 2000) and it is well known that then the other three components of the quintuple are normal too. This facilitates both likelihood and Bayesian inferences. A generic expression, as well as expressions for several particular cases are given in Appendix A.1. For what follows, a few points need to be made. First, conjugacy is elegant and convenient for Bayesian inference, because the prior and the posterior share the same distributional form, with the parameters of the latter updates of the former. In particular, this implies that the dimensions of the prior and the posterior are identical. Second, these identical distribution forms do not hold for the likelihood case, where one would use the marginal distribution as given in (1). While in this case the marginal distribution is expressed as an exponentiated linear combination of the other three normalizing constants, this can lead to a form that is very different from that of the conditional distribution. For example, the marginal may be multivariate and the conditional a product of univariate distributions. Unsurprisingly, from Section A.1, the normal-normal is an exception, but in all other cases, the marginal distribution is fairly intractable.

Note however that, although the use of conjugacy can lead to marginal models with closed-form expressions for the moments and marginal density, it does not in general allow us to construct marginal expectations with pre-specified parametric form. This has the implication that, when interest lies in a marginal distribution with a pre-specified form in terms of expectation, conjugacy is not a viable route. We now review three possible alternatives and then focus on the third of these, bridging.

In a generalized linear mixed model (Wolfinger and O'Connell 1993, Breslow and Clayton 1993, Molenberghs and Verbeke 2005), the conditional distribution is chosen from the exponential family:

$$f_{ij}\left(y_{ij}|\theta_{ij},\boldsymbol{b}_i,\phi\right)=\exp\{\phi^{-1}\left[y_{ij}\theta_{ij}-\psi(\theta_{ij})\right]+c(y_{ij},\phi)\},\tag{2}$$

with the mixing distribution remaining normal. In (2), θ_{ij} is the natural parameter, $\psi(\cdot)$ is the generating function, $c(\cdot)$ the normalizing constant, and ϕ an overdispersion parameter. The dependence on $\boldsymbol{b}_i$ is usually through the natural parameter, e.g., $\theta_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i$. Here, $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$ are design vectors and $\boldsymbol{\beta}$ is a vector of fixed but unknown parameters.

An important tool for likelihood inference is the marginal mean $\boldsymbol{\mu}_i^m = E(\boldsymbol{Y}_i)$. It can be obtained from the marginal distribution, or from the conditional mean $\boldsymbol{\mu}_i^c = E(\boldsymbol{Y}_i|\boldsymbol{b}_i)$ by integrating over the random effects. Again here, for notational convenience, dependence of the mean on other information than the random effects is suppressed.

In general though, obtaining the marginal distribution, the marginal mean, or the posterior can be cumbersome and there are a number of important cases where no closed-form solution exists. This is because the conditional mean may not be a linear function of the random effects. For the exponential family, it is customary to introduce the concept of an inverse link function $g(\cdot)$ such that $\theta_{ij} = g^{-1}(\mu_{ij}^c)$.

When the conditional distribution is Bernoulli and the logit link is chosen together with a normal mixing distribution, then there is no closed form for μ_{ij}^m, let alone for the marginal distribution. While Molenberghs *et al* (2010) have shown that for a number of cases closed forms exist for the marginal distributions, means, variances, and higher moments, it remains of value to consider choices for the distributions such that marginals and posteriors remain tractable.

Heagerty (1999) formulated a so-called *marginalized multilevel model* (MMM) approach by requiring

$$g^{-1}(\mu_{ij}^m) = \boldsymbol{x}_{ij}'\boldsymbol{\beta}, \tag{3}$$

$$g^{-1}(\mu_{ij}^c) = \Delta_{ij} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i, \tag{4}$$

Here, the function Δ_{ij} connects the marginal and conditional means through the same link function.

Wang and Louis (2003) specified a model by requesting that the marginal mean and the conditional mean are specified by identical link functions, with predictors that are the same up to a multiplicative factor ϕ and an offset k:

$$g^{-1}(\mu_{ij}^m) = k + \phi\boldsymbol{x}_{ij}'\boldsymbol{\beta}, \tag{5}$$

$$g^{-1}(\mu_{ij}^c) = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i. \tag{6}$$

Specification (5)–(6) is similar to (3)–(4), but now the random-effects distribution of $\boldsymbol{b}_i$ is unknown, rather than the connector Δ_{ij} in (4). From (5) and (6) the random-effects distribution needs to be solved. We term this the *bridge* operation.

Molenberghs *et al* (2012) drew attention to the fact that marginalization, MMM, and bridging can be formulated in terms of integral equations. Assume that the fixed predictor $\boldsymbol{x}_{ij}'\boldsymbol{\beta}$ and random predictor $\boldsymbol{z}_{ij}'\boldsymbol{b}_i$ are pre-specified, as well as the link function $g^{-1}(\cdot)$. The integral equations are as follows.

Marginalization: The mixing distribution $f(\boldsymbol{b}_i)$ is given, while the marginal mean is unknown. This leads to the explicit integral:

$$\mu_{ij}^m = \int g(\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i) f(\boldsymbol{b}_i) d\boldsymbol{b}_i. \tag{7}$$

It is also possible to marginalize the entire distribution, as given in (1).

Marginalized Multilevel Model Integral Equation: The random-effects density $f(\boldsymbol{b}_i)$ is given, with the connector function Δ_{ij} unknown and identified through the relationship:

$$\mu_{ij}^m = g(\boldsymbol{x}_{ij}'\boldsymbol{\beta}) = \int g(\Delta_{ij} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i) f(\boldsymbol{b}_i) d\boldsymbol{b}_i. \tag{8}$$

Bridge Integral Equation: The random-effects density $f(\boldsymbol{b}_i)$ is unknown, as well as the constants k and ϕ, but are identified through:

$$\mu_{ij}^m = g(k + \phi\boldsymbol{x}_{ij}'\boldsymbol{\beta}) = \int g(\boldsymbol{x}_{ij}'\boldsymbol{\beta} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i) f(\boldsymbol{b}_i) d\boldsymbol{b}_i. \tag{9}$$

As before, the identifying relationship is the solution to an integral equation.

MMM and bridging are intended to produce convenient expressions that facilitate parametric inferences based on the marginal model. Alternatively, one can ensure that the components in (1) satisfy conjugacy (Cox and Hinkley 1974, p. 370; McCullagh and Nelder 1989). The conditional and mixing distributions are said to be *conjugate* if and only if they can be written in the generic forms:

$$f(y_{ij}|\theta_{ij}) = \exp\left\{\phi^{-1}[y_{ij}h(\theta_{ij}) - \psi(\theta_{ij})] + c(y_{ij}, \phi)\right\}, \tag{10}$$

$$f(\theta_{ij}) = \exp\left\{\gamma[\lambda h(\theta_{ij}) - \psi(\theta_{ij})] + c^*(\gamma, \lambda)\right\}, \tag{11}$$

where $\psi(\theta_{ij})$ and $h(\theta_{ij})$ are functions, ϕ, γ, and λ are parameters, and the additional functions $c(y_{ij}, \phi)$ and $c^*(\gamma, \psi)$ are so-called normalizing constants. Note that (10) differs slightly from (2) in that θ_{ij} is now replaced by a function. This is not essential because the function can be absorbed into the parameter itself. More fundamental is the fact that in (2) $\boldsymbol{b}_i$ is random and θ_{ij} is fixed, whereas in (10) $\boldsymbol{b}_i$ is not present but rather θ_{ij} is random. Both formulations give rise to different classes of models. The earlier one leads to GLMM, whereas the current formulation is behind overdispersion models. Molenberghs *et al* (2007, 2010) combined both into a single model; this will not be considered here.

It can then be shown, upon constructing the joint distribution from (10) and (11) and then constructing the reverse factorization, that the marginal and predictive distributions are:

$$\begin{aligned} f(y_{ij}) &= \exp\left[c(y_{ij}, \phi) + c^*(\gamma, \lambda) - c^*\left(\phi^{-1} + \gamma, \frac{\phi^{-1} y_{ij} + \gamma\lambda}{\phi^{-1} + \gamma}\right)\right], \qquad (12) \\ f(\theta_{ij}|y_{ij}) &= \exp\left\{(\phi^{-1} + \gamma)\left[\frac{\phi^{-1} y_{ij} + \gamma\lambda}{\phi^{-1} + \gamma} h(\theta_{ij}) - \psi(\theta_{ij})\right.\right. \\ &\qquad \left.\left. + c^*\left(\phi^{-1} + \gamma, \frac{\phi^{-1} y_{ij} + \gamma\lambda}{\phi^{-1} + \gamma}\right)\right]\right\}. \qquad (13) \end{aligned}$$

The marginal model (12) is used for likelihood inferences. It is expressed in terms of the three normalizing constants of (10), (11), and (13). The posterior (13) is used for Bayesian inference and can be seen as an update of (11) by replacing the prior by the posterior parameters:

$$\gamma \to \phi^{-1} + \gamma, \qquad \lambda \to \frac{\phi^{-1} y_{ij} + \gamma\lambda}{\phi^{-1} + \gamma}.$$

In other words, conjugacy leads in Bayesian terms to retention of the distributional form when going from prior to posterior, upon updating the parameters. In likelihood terms, where $f(\boldsymbol{y})$ is used, conjugacy leads to a marginal consisting of the three normalizing constants, as is clear from (12). This, while elegant, is a different parametric form than the conditional distribution (10). It is here that the bridge operation comes in, to preserve, not the entire distribution, but the parametric form of the mean function.

In the Appendix, several examples of conjugate quintuples are given.

2.3 Bridging

Aa stated at the end of the previous section and, in contrast to conjugacy, which operates at the level of the distributions, bridging is defined through (5) and (6) and operationalized in (9) at the level of the means. Molenberghs *et al* (2012) considered three generic cases.

The first one is where the mean function takes values on the entire real line. A natural link is then the identity and the bridging problem is trivial. In that case, the bridge equation is satisfied by any integrable function. Moreover, marginalization, bridging, and the MMM operations are identical. This is seen by assuming that $\boldsymbol{b}_i$ follows a distribution with finite mean and density $f(\boldsymbol{b}_i)$, producing $\phi = 1$ and $k_{ij} = \boldsymbol{z}'_{ij} E(\boldsymbol{b}_i)$.

The second case arises for mean functions taking non-negative values only over a half line; without loss of generality, assume that they take values over the nonnegative real numbers. A natural link is then the logarithmic function. Marginalizing the corresponding Poisson-normal GLM is then also straightforward (Zeger, Liang, and Albert 1988, Molenberghs, Verbeke, and Demétrio 2007, Molenberghs *et al* 2012):

$$E(Y_{ij}) = \int e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i} \varphi(\boldsymbol{b}_i) d\boldsymbol{b}_i = e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \frac{1}{2}\boldsymbol{z}'_{ij} D \boldsymbol{z}_{ij}}. \tag{14}$$

The connector function for the MMM, in the Poisson case, then is:

$$\Delta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} - \frac{1}{2}\boldsymbol{z}'_{ij} D \boldsymbol{z}_{ij}. \tag{15}$$

For the bridge distribution, the corresponding equation is:

$$e^{k_{ij} + \phi \boldsymbol{x}'_{ij}\boldsymbol{\beta}} = \int e^{\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i} f(\boldsymbol{b}_i) d\boldsymbol{b}_i, \tag{16}$$

satisfied by not only the normal but a wide class of mixing distributions. For the normal, $\phi = 1$, and $k_{ij} = \frac{1}{2}\boldsymbol{z}'_{ij} D \boldsymbol{z}_{ij}$. As in the case of the identity link, the three operations are strongly related, though not identical. Indeed, for the marginalized GLMM and for the bridge, the constant k_{ij} appears in the marginal mean, whereas the same constant appears in the connector when considering the MMM. Molenberghs *et al* (2012) considered this log-linear case for the Poisson distribution, the Weibull distribution, the gamma distribution, and the Inverse Gaussian distribution.

The third case, for which the mean function takes values through an interval, is fundamentally different from the other two and will be the focus in the remainder of the paper. The generic example is binary data where the conditional distribution is Bernoulli or, more generally, binomial, although a variety of other examples, such as percentages, could be considered as well. While intervals other than the unit interval could be considered, we focus here without loss of generality on the unit interval. Because the link function should be a one-to-one map between the unit interval and the real line, it is necessarily a distribution function. So, in contrast to the linear and log-linear cases, the bridge equation (9) can be seen as a relationship between two density functions. Although the function $g(\cdot)$ is a distribution rather than density function, Wang and Louis (2003) have shown how the equation can be transformed to give an integral equation between two densities. Writing $H(\cdot) \equiv g^{-1}(\cdot)$, Wang and Louis (2003) showed that the bridge equation becomes:

$$H(k + \phi\eta) = \int H(b + \eta) f(b) db. \tag{17}$$

Here, $\eta \equiv \boldsymbol{x}'_{ij}\boldsymbol{\beta}$ for ease of notation. Their derivation was focused on a random intercept only.

Because our development builds upon Wang and Louis's (2003) work, we provide a brief review of the relevant steps. Taking derivatives on both sides of (17) with respect to η, leads to

$$\phi h(k + \phi\eta) = \int h(b + \eta) f(b) db, \tag{18}$$

with $h = H'$, the first derivative. This is a convolution: $h * f_{-b}(\eta) = \phi h(k + \phi\eta)$. The subscript $-b$ refers to sign reversal. They then transformed this equation to the Fourier domain and applied properties of the Fourier transform to yield:

$$\mathcal{F} f_{-b}(\xi) = e^{ik\xi/\phi} \mathcal{F} h(\xi/\phi) \Big/ \mathcal{F} h(\xi), \tag{19}$$

where

$$(\mathcal{F} h)(\xi) = \int e^{-i\xi x} h(x) dx, \tag{20}$$

the Fourier transform. Clearly, (20) maps the density corresponding to the inverse link function to its characteristic function, as also noted by Wang and Louis (2003). This property will be taken up in the next section. Applying the inverse Fourier transform yields the generic solution to (17):

$$f(b) = \frac{1}{2\pi} \int e^{i(k/\phi - b)\xi} \frac{(\mathcal{F} h)(\xi/\phi)}{(\mathcal{F} h)(\xi)} d\xi. \tag{21}$$

Wang and Louis (2003) also showed that, for symmetric $h(\cdot)$, $k = 0$, and that $0 \leq \phi \leq 1$.

We note as well that the existence and uniqueness of the bridge density is guaranteed. Existence holds whenever $h(\cdot)$ is integrable and non-degenerate, which is satisfied for all conventional links, precisely stemming from the connection with density functions. Uniqueness follows by construction and from the uniqueness of the Fourier transform. Uniqueness will be revisited in terms of the characteristic function in the next section.

In what follows, we will be using three particular instances: the logit, probit, and Cauchy. We now set out existing results on these.

First, Molenberghs *et al* (2012) drew attention to the fact neither the marginal mean nor the connector function in the MMM, for a normal random effect, have a closed form. It then follows that the normal distribution is not bridge. In fact, Wang and Louis (2003) derived an entirely different solution, by solving:

$$\text{expit}(k + \phi \boldsymbol{x}_{ij}' \boldsymbol{\beta}) = \int_b \text{expit}(\boldsymbol{x}_{ij}' \boldsymbol{\beta} + b_i) f(b_i) db_i. \tag{22}$$

Their solution, derived based on (21) and Fourier transform operations, reads $k = 0$ and

$$f(b_i) = \frac{1}{2\pi} \frac{\sin(\phi\pi)}{\cosh(\phi b_i) + \cos(\phi\pi)}, \tag{23}$$

with

$$\phi = \frac{1}{\sqrt{1 + \frac{3}{\pi^2} d}}, \tag{24}$$

and d the random-intercept variance. Wang and Louis (2003) studied the properties of (22).

Second, matters are entirely different for a probit link. In that case, the marginal mean function is explicit (Zeger, Liang, and Albert 1988, Griswold and Zeger 2004, Molenberghs *et al* 2010):

$$E(Y_{ij}) = \Phi\left(\phi_{ij} \cdot \boldsymbol{x}_{ij}' \boldsymbol{\beta}\right) = \int_b \Phi(\boldsymbol{x}_{ij}' \boldsymbol{\beta} + \boldsymbol{z}_{ij}' \boldsymbol{b}_i) \varphi(\boldsymbol{b}_i) d\boldsymbol{b}_i, \tag{25}$$

with

$$\phi_{ij} = \frac{1}{\sqrt{1 + \boldsymbol{z}_{ij}' D \boldsymbol{z}_{ij}}}, \tag{26}$$

as is the connector function for the probit-normal GLMM:

$$\Delta_{ij} = \phi_{ij}^{-1} \boldsymbol{x}_{ij}' \boldsymbol{\beta} = \sqrt{1 + \boldsymbol{z}_{ij}' D \boldsymbol{z}_{ij}} \cdot \boldsymbol{x}_{ij}' \boldsymbol{\beta}. \tag{27}$$

Molenberghs *et al* (2012) derived from (25) that the normal density is the bridge with $k = 0$ and ϕ_{ij} as in (26). Thus, as opposed to the logit case, all three operations lead to closed forms and can be said to coincide, up to perhaps a multiplicative factor of the form (26).

It is important to realize that a probit link admits the normal distribution as a bridge. Stated differently, the normal is "self-bridging," a concept that we will formally define and return to in Section 6.

Third, when the Cauchy distribution

$$H(\eta) = \frac{1}{\pi}\left(\frac{\pi}{2} + \arctan\eta\right)$$

is used, the corresponding density is $h(\eta) = \left[\pi(1+\eta^2)\right]^{-1}$, with Fourier transform $(\mathcal{F}h)(\xi) = \exp(-|\xi|)$, a well-known special Fourier transform (Spiegel 1968, p. 176, Section 33.18). The bridge density then simply follows:

$$f(b) = \frac{1}{2\pi}\int_0^{+\infty} e^{-ib\xi - \frac{\xi}{\phi} + \xi} d\xi + \frac{1}{2\pi}\int_{-\infty}^{0} e^{-ib\xi + \frac{\xi}{\phi} - \xi} d\xi = \frac{1}{\pi}\frac{(\phi^{-1}-1)}{(\phi^{-1}-1)^2 + b^2}.$$

Hence, the standard Cauchy link produces a Cauchy bridge with parameter $c^2 = (\phi^{-1}-1)^2$, leading to $\phi = (c+1)^{-1}$. Interestingly, the Cauchy is also "self-bridging," and this will be revisited as well in Section 6.

3 Bridging Using Characteristic Function

As seen in the previous section, the Fourier operation employed by Wang and Louis (2003) is closely linked with the use of the characteristic function. Given the vast body of results on the characteristic function, it is of interest to examine this further.

Because the characteristic function of a distribution H is defined as

$$\varphi_h(\xi) = \int e^{i\xi x} h(x) dx, \tag{28}$$

the connection with Fourier transform (20) is immediate, a fact also noted by Wang and Louis (2003). This implies that (19) can be restated as

$$\varphi_h(\xi) \cdot \varphi_f(\xi) = e^{-ik\xi/\phi} \cdot \varphi_h(\xi/\phi). \tag{29}$$

The importance of this lies in the fact that there are many useful results known about the characteristic function of a distribution. We will make use of these in what follows, especially results about series expansions of such functions. Clearly the characteristic function representation of the bridge integral equation transforms in into multiplicative equation (29). This is hardly surprising, because it restates the approach of Wang and Louis. However, we may consider whether the other integral-equation operations, marginalization and MMM construction, can be restated in a parallel way.

For marginalization equation (7), a characteristic function representation is possible only when all functions involved are densities. However, we do have densities for binary datae real line, because a link functions is a 1-to-1 map between the unit interval and the real line. Thus, upon writing $\ell(\eta_{ij}) = \mu_{ij}^m$, with $\eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}$, and further restricting attention to a random intercept example $\boldsymbol{z}'_{ij}\boldsymbol{b}_i = b_i$, we obtain the following characteristic function representation of (7):

$$\varphi_\ell(\xi) = \varphi_h(\xi) \cdot \varphi_{\widetilde{f}}(\xi), \tag{30}$$

where $h = g^{-1}$ in this case, and $\widetilde{f}$ reflects the necessary sign change of the argument, because of the argument $\eta_{ij} + b_i$ of h leads to a convolution only if the sign of the argument of f is reversed. Evidently, for symmetric densities f this is immaterial.

Applying (30) to the logistic normal model, with

$$\varphi_h(\xi) = B(1 - i\xi, 1 + i\xi) = \Gamma(1 - i\xi) \cdot \Gamma(1 + i\xi) = \frac{2\pi\xi}{e^{\pi\xi} - e^{-\pi\xi}},$$

and $\varphi_f(\xi) = \exp\left[-\frac{1}{2}\xi^2\sigma^2\right]$, we immediately find that

$$\varphi_\ell(\xi) = \frac{2\pi\xi e^{-\frac{1}{2}\xi^2\sigma^2}}{e^{\pi\xi} - e^{-\pi\xi}}.$$

Thus, the characteristic function of the "marginal mean density," (i.e., inverse link function) has a closed form, even though the marginal mean itself does not.

For MMM, the situation is not so straightforward. We write (8) in simplified notation as:

$$h(\eta) = \int h(\Delta + b) f(b) db,$$

then the transform equation in terms of Δ is

$$\int e^{i\Delta\xi}h(\eta)d\Delta = \int\int e^{i\Delta\xi}h(\Delta+b)f(b)d\Delta, \tag{31}$$

where Δ and η are functions of each other. Writing $\eta = \eta(\Delta)$, (31) becomes

$$\int e^{i\Delta\xi}\mathcal{H}(\Delta)d\Delta = \varphi_h(\xi)\cdot\varphi_{\tilde{f}}(\xi), \tag{32}$$

where $\mathcal{H}(\Delta) = h(\eta(\Delta))$. This is not a very practicable result. As we will see next, the main use of the characteristic-form representation is with bridging.

4 Bridging, Inverse Bridging, and Uniqueness

Characteristic function representation (29) of the bridge equation is useful in its own right. This is clearly the case for distribution with a relatively simple characteristic function. For example if h is chosen to be the standard normal density, the $\varphi_h(\xi) = \exp(-0.5\xi^2)$ and hence

$$\varphi_f(\xi)\exp\left[\frac{(i\xi)}{1!}\frac{k}{\phi} + \frac{(i\xi)^2}{2!}\left(\frac{1}{\phi^2}-1\right)\right], \tag{33}$$

which implies that f can only be the normal density with mean $\mu = k/\phi$ and variance $\sigma^2 = 1/\phi^2 - 1$.

The exponentiated Taylor series expansion form of (33) is not a coincidence and is reminiscent of the Gram-Charlier A series (Kolassa 2006) expression of a characteristic function, used to write a characteristic function of an arbitrary density as a series expansion "around" the normal density, with the same possible for the density itself. Rather than considering the Gram-Charlier A series, we will adopt the general exponential series expansion:

$$\varphi_h(\xi) = \exp\left(\sum_{r=1}^{\infty}\frac{\kappa_r^{(h)}(i\xi)^r}{r!}\right), \tag{34}$$

where $\kappa_r^{(h)}$ is the rth cumulant of h. Using (34) for both h and f, characteristic-function condition (29) becomes:

$$\sum_{r=1}^{\infty}\frac{\kappa_r^{(f)}(i\xi)^r}{r!} + \sum_{r=1}^{\infty}\frac{\kappa_r^{(h)}(i\xi)^r}{r!} = -(i\xi)\frac{k}{\phi} + \sum_{r=1}^{\infty}\frac{\kappa_r^{(h)}(i\xi)^r}{\phi^r r!}. \tag{35}$$

This is equivalent to the set of conditions:

$$\kappa_1^{(f)} + \kappa_1^{(h)} = \frac{\kappa_1^{(h)}}{\phi} - \frac{k}{\phi}, \tag{36}$$

$$\kappa_r^{(f)} + \kappa_r^{(h)} = \frac{\kappa_r^{(h)}}{\phi^r}, \qquad (r \geq 2). \tag{37}$$

Conditions (36) and (37) allow us to find the cumulants and hence characteristic function of f in terms of these of h:

$$\kappa_1^{(f)} = -\frac{k}{\phi} + \left(\frac{1}{\phi} - 1\right)\kappa_1^{(h)}, \tag{38}$$

$$\kappa_r^{(f)} = \left(\frac{1}{\phi^r} - 1\right)\kappa_r^{(h)}, \qquad (r \geq 2), \tag{39}$$

and vice versa:

$$\kappa_1^{(h)} = \frac{\phi}{1-\phi}\left(\frac{k}{\phi} + \kappa_1^{(f)}\right), \tag{40}$$

$$\kappa_r^{(h)} = \frac{\phi^r}{1-\phi^r}\kappa_r^{(f)}, \qquad (r \geq 2). \tag{41}$$

These equations establish the unique correspondence between f and h, provided that both have existing, finite cumulants of all orders. Of course, this is up to some parametric freedom at the first order.

Equations (40)–(41) show that not only one can derive the bridge density corresponding to a given density (inverse link function), but also that the reverse operation can be conducted. We will refer to this as the *inverse bridge* operation.

The above establishes the theorem:

Proposition 1 *Provided that an inverse link function f and a random-effects distribution h both admit a cumulant generating function and that they are in a bridge relationship, then h is unique to f and vice versa.*

5 Illustration

Using bridging and inverse bridging we recover in a straightforward way the results from Wang and Louis and from our previous paper (normal, Cauchy, t,...) and several others that we present in the supplementary material.

Note that the Cauchy distribution is a special case in the sense that it does not admit a cumulant generating function, which is necessary for the results of the previous section. Nevertheless, from its characteristic function we have shown, along with Wang and Louis (2003), that it too is self-bridging. We will return to the Cauchy once more in the next section.

On the other hand, the degenerate distribution, often considered irregular, is regular when its characteristic function is considered. Indeed, if $f(x) = \delta_a$, then $\varphi_f(\xi) = \exp(i\xi a)$, and (29) is satisfied for both h and f degenerate. Indeed, the condition implies that

$$i\xi a_h + i\xi a_f = -\frac{k}{\phi} i\xi + i\xi \frac{a_h}{\phi},$$

immediately yielding the equations linking f to h and reverse relationship:

$$a_f = a_h \left(\frac{1}{\phi} - 1 \right) - \frac{k}{\phi}, \tag{42}$$

$$a_h = \frac{\phi}{1-\phi} a_f + \frac{1}{1-\phi} k. \tag{43}$$

Relations (42)–(43) are derived from the characteristic function and not from the cumulant generating representation. However, the same would follow immediately, given the form of the characteristic function. In fact, the characteristic function of the degenerate distribution is a simple special case of that of the normal, with the second cumulant now also equal to zero, apart from the ones of order 3 and higher. The degenerate distribution is also self bridging so, provided, $\phi < 1$, f degenerate implies that h is degenerate and vice versa.

As another example, take h to be the loggamma density:

$$h(x) = \frac{e^{-\alpha x} e^{-e^{-x}}}{\Gamma(\alpha)}, \tag{44}$$

with characteristic function

$$\varphi_h(\xi) = \frac{\Gamma(\alpha - i\xi)}{\Gamma(\alpha)}, \tag{45}$$

leading to the condition:

$$\frac{\Gamma(\alpha - i\xi)}{\Gamma(\alpha)} \cdot \varphi_f(\xi) = e^{-\frac{i\xi k}{\phi}} \cdot \frac{\Gamma\left(\alpha - i\frac{\xi}{\phi}\right)}{\Gamma(\alpha)}, \tag{46}$$

which, in turn, leads to the following expression for the characteristic function of f:

$$\varphi_f(\xi) = e^{-\frac{i\xi k}{\phi}} \cdot \frac{\Gamma\left(\alpha - i\frac{\xi}{\phi}\right)}{\Gamma(\alpha - i\xi)}. \tag{47}$$

Conversely, start from f loggamma, i.e., $\varphi_f(\xi)$ takes the form (45), leading to condition:

$$\varphi_h(\xi) \cdot \frac{\Gamma(\alpha - i\xi)}{\Gamma(\alpha)} = e^{-\frac{i\xi k}{\phi}} \cdot \varphi_h\left(\frac{\xi}{\phi}\right). \tag{48}$$

Clearly, condition (48) is more involved than its counterpart (46). We therefore use the characteristic-function form:

$$\varphi_h(\xi) = \exp\left[\sum_{r=1}^{+\infty} \frac{\kappa_r^{(h)}(i\xi)^r}{r!}\right],$$

which allows us to rewrite (48) as

$$\sum_{r=1}^{+\infty} \frac{\kappa_r^{(h)}(i\xi)^r}{r!} + \ln\Gamma(\alpha - i\xi) - \ln\Gamma(\alpha) = \frac{-i\xi k}{\phi} + \sum_{r=1}^{+\infty} \frac{\kappa_r^{(h)}(i\xi)^r}{\phi^r r!}. \tag{49}$$

Write

$$\ln\Gamma(\alpha - i\xi) = \ln\Gamma(\alpha) + \sum_{r=1}^{+\infty} \gamma^{(r)} \frac{(-i\xi)^r}{r!},$$

then (49) becomes

$$\sum_{r=1}^{+\infty} \frac{\kappa_r^{(h)}(i\xi)^r}{r!} + \sum_{r=1}^{+\infty} \gamma^{(r)}(-i)^r \frac{\xi^r}{r!} = \frac{-i\xi k}{\phi} + \sum_{r=1}^{+\infty} \frac{\kappa_r^{(h)}}{\phi^r} \frac{(i\xi)^r}{r!}.$$

Equating powers leads to

$$\begin{aligned}
\kappa_1^{(h)} &= \frac{\phi}{1-\phi}\left(\frac{k}{\phi} - \gamma^{(1)}\right), \\
\kappa_5^{(h)} &= \frac{\phi^r}{1-\phi^r}(-1)^r \gamma^{(r)}, \quad (r \geq 2),
\end{aligned}$$

with $\gamma^{(r)}$ the rth derivative of $\ln\Gamma(\alpha + x)$, evaluated at $x = 0$. In other words, $\gamma^(1)(\alpha) = \psi(\alpha)$, the digamma function and, for $r \geq 2$,

$$\gamma^{(r)}(\alpha) = \psi^{(r-1)}(\alpha) = \frac{d^r}{d\alpha^r} \ln\Gamma(\alpha) = \frac{d^{r-1}}{d\alpha^{r-1}} \psi(\alpha),$$

the polygamma function.

Returning to expression (47) and using the conventional characteristic function expansion for f, the equation can be written:

$$\sum_{r=1}^{+\infty} \frac{\kappa_r^{(f)}(i\xi)^r}{r!} = -\frac{i\xi k}{\phi} + \sum_{r=1}^{+\infty} \frac{\gamma^{(r)}}{\phi^r} \frac{(-i\xi)^r}{r!} - \sum_{r=1}^{+\infty} \gamma^{(r)} \frac{(-i\xi)^r}{r!}.$$

This leads to

$$\begin{aligned}
\kappa_1^{(f)} &= -\frac{k}{\phi} - \gamma^{(1)}\left(\frac{1}{\phi} - 1\right), \\
\kappa_r^{(f)} &= \gamma^{(r)}(-1)^r \left(\frac{1}{\phi^r} - 1\right), \quad (r \geq 2).
\end{aligned}$$

6 Self Bridging

We have found up to this point that three distributions are so-called self bridging, in the sense that their f and h have the same parametric form, up to the values taken by the parameters. These are the normal, Cauchy, and degenerate distributions.

We term a distribution *regular* if all of its cumulants exist and are finite, implying that it permits a cumulant generating function. This allows us to formulate and prove the following theorem:

Proposition 2 *If a regular distribution is self-bridging, then it is either the normal or the degenerate distribution.*

Proof. Conditions (29) for self-bridging distributions become

$$\varphi_h(\xi)e^{-\frac{\xi\widetilde{k}}{\widetilde{\phi}}} \cdot \varphi_h\left(\frac{\xi}{\widetilde{\phi}}\right) = e^{-\frac{\xi k}{\phi}} \cdot \varphi_h\left(\frac{\xi}{\phi}\right), \tag{50}$$

where $\widetilde{k}$ and $\widetilde{\phi}$ are introduced to ensure that the inverse link and mixing distributions are the same up to a change in location and scale. From regularity, (50) is equivalent to the conditions:

$$\kappa_1\left(1+\widetilde{\phi}^{-1}-\phi^{-1}\right) = \frac{\widetilde{k}}{\widetilde{\phi}} - \frac{k}{\phi}, \tag{51}$$

$$\kappa_r\left(1+\widetilde{\phi}^{-r}-\phi^{-r}\right) = 0. \tag{52}$$

The range for ϕ is $[0,1]$ while $\widetilde{\phi}$ can take any non-negative value, and from the presence of the location parameters, (51) can always be satisfied, whatever the scale parameters and first cumulant. Next, (52) is clearly satisfied if all cumulants from order 2 onwards up to at most one are equal to zero. If all are zero, then the degenerate distribution results. If $\kappa_2 \neq 0$ but all others equal zero, then the normal distribution results. Further, if for one given $r \geq 3$, together with κ_1, then a polynomial cumulant generating function results of order at least 3. However, Lukacs (1970) showed that there are no distributions with such a cumulant generating function.

The one remaining situation is where $\kappa_r \neq 0$ and $\kappa_s \neq 0$ for $r \neq s$. Note that, if this does occur, then there must be infinitely many of them unequal to zero, because otherwise a polynomial characteristic function would again result. But we can show that already for two values, this is not possible. Indeed, in this case, the second factors on the left hand side of (52) would have to be zero for r and s, implying that $(1+\widetilde{\lambda}^r)^{1/r} = \lambda = (1+\widetilde{\lambda}^s)^{1/s}$, where $\lambda = \phi^{-1}$ and $\widetilde{\lambda} = \widetilde{\phi}^{-1}$. This can be written as $||(1,\widetilde{\lambda})||_r = ||(1,\widetilde{\lambda})||_s$. However, the norms are ordered, with equality holding only when $\lambda = 1$ and $\widetilde{\lambda} = 0$, which does not correspond to a finite solution for $\widetilde{\phi}$. This completes the proof.

Now, we have seen in Section 2.3 that also the Cauchy distribution is self-bridging. This is not in conflict with Theorem 2 because the Cauchy distribution does not admit a regular cumulant generating function. Indeed, for the general Cauchy density:

$$h(x) = \frac{1}{\pi\gamma_h} \cdot \frac{1}{\left[1+\left(\frac{x-\mu_h}{\gamma_h}\right)\right]}, \tag{53}$$

the characteristic function is:

$$\gamma_h(\xi) = e^{\mu_h i\xi - \gamma_h|\xi|}. \tag{54}$$

Requiring that also f is of the form (53), with parameters μ_f and γ_f, then the correspondence equations become:

$$\begin{array}{rclcrcl} \mu_f &=& \mu_h\left(\phi^{-1}-1\right), && \mu_h &=& \frac{\phi}{1-\phi}\cdot\mu_f, \\ \gamma_f &=& \gamma_h\left(\phi^{-1}-1\right), && \gamma_h &=& \frac{\phi}{1-\phi}\cdot\gamma_f. \end{array} \tag{55}$$

Note that, when $\phi = 0.5$, then f and h are not only both Cauchy, but identical Cauchy distributions.

So, the Cauchy distribution is self bridging, even though the characteristic function is not of the usual cumulant generating function form, because none of the cumulants exist.

From a practical perspective, only the normal distribution is self-bridging and a candidate for data analysis. This implies that the three operations in all other cases differ from each other, to varying degrees. So, the choice between them has genuine substantive implications.

7 Concluding Remarks

This paper is concerned with the general problem of constructing marginal expectations and their parameters from generalized linear mixed models (GLMM) and their extensions. We focused on one particular framework, the so-called bridging operation. Because the expectation operator is linear and multiplicative, the identity link case is trivial, while the logarithmic case is straightforward, and these have been dealt with in Molenberghs *et al* (2012). The identity link corresponds to random variables with real-line support, the logarithmic with half lines. This leaves the important case of variables with finite-interval support. Without loss of generality, we restricted attention to the unit interval. This is the setting considered by both Wang and Louis (2003) and Molenberghs *et al* (2012).

We have developed these results for interval data, using the concept of conjugacy on the one hand and the bridging operation on the other. We saw that conjugacy is a natural device for the Bayesian paradigm, but is not when interest lies in the marginal distribution with a pre-specified form in terms of expectation.

This suggests considering quintuples with a conditional and marginal factor with the same parametric form, which can be achieved through the bridging operation. We have shown that using the characteristic function and its series expansions facilitates bridging and also allows for inverse bridging, which means that the parametric form for the random-effects distribution is pre-specified and the inverse link defining the mean function derives from the operation. Through this representation, uniqueness results can be readily derived.

Finally, we have introduced the concept of self-bridging, meaning that the random-effects distribution and the inverse link function are of the same form. We established that it is an extremely strong condition in the sense that there is only one practically useful regular instance: the normal function. The other regular instance, in the sense of admitting a characteristic function with series expansion, is the degenerate distribution. Other instances necessarily will be irregular. One example, studied here, is the Cauchy. This irregular case, as well as others, is not considered here, and is unlikely to be useful in practice.

Acknowledgment

The authors gratefully acknowledge support from IAP research Network P7/06 of the Belgian Government (Belgian Science Policy).

References

Breslow, N.E. and Clayton, D.G. (1993), "Approximate inference in generalized linear mixed models," *Journal of the American Statistical Association*, 88, 9–25.

Griswold, M.E. and Zeger, S.L. (2004), "On marginalized multilevel models and their computation (November 2004)," *Johns Hopkins University, Department of Biostatistics Working Paper #99*.

Heagerty, P.J. (1999), "Marginally specified logistic-normal models for longitudinal binary data," *Biometrics*, 55, 688–698

Heagerty, P.J. and Zeger, S.L. (2000), "Marginalized multilevel models and likelihood inference (with discussion)," *Statistical Science*, 15, 1–26

Hinde, J. and Demétrio, C.G.B. (1998a), "Overdispersion: Models and estimation," *Computational Statistics and Data Analysis*, 27, 151–170.

Hinde, J. and Demétrio, C.G.B. (1998b), *Overdispersion: Models and Estimation*, São Paulo: XIII Sinape.

Iddi, S., and Molenberghs, G. (2011), "A combined overdispersed and marginalized multilevel model," *Computational Statistics and Data Analysis*, 56, 1944–1951.

Kolassa, J.E. (2006). *Series Approximation Methods in Statistics.* (3rd ed.). Lecture Notes in Statistics 88. New York: Springer.

Krishnamoorthy, K. (2006), *Handbook of Statistical Distributions with Applications*, Boca Raton: Chapman & Hall/CRC.

Laird, N.M. and Ware, J.H. (1982), "Random effects models for longitudinal data," *Biometrics*, 38, 963–974.

Lukacs, E. (1970). *Characteristic Functions. (2nd ed.).* London: Griffin.

McCullagh, P. and Nelder, J.A. (1989), *Generalized Linear Models*, London: Chapman & Hall/CRC.

Molenberghs, G., Kenward, M.G., Verbeke, G., Efendi, A., and Iddi, S. (2012). On the connection between bridge distributions, marginalized multilevel models, and generalized linear mixed models. *Submitted for publication.*

Molenberghs, G. and Verbeke, G. (2005), *Models for Discrete Longitudinal Data*, New York: Springer.

Molenberghs, G., Verbeke, G., and Demétrio, C. (2007), "An extended random-effects approach to modeling repeated, overdispersed count data," *Lifetime Data Analysis*, 13, 513–531.

Molenberghs, G., Verbeke, G., Demétrio, C., and Vieira, A. (2010), "A family of generalized linear models for repeated measures with normal and conjugate random effects," *Statistical Science*, 25, 325–347.

Searle, S.R., Casella, G., and McCulloch, C.E. (1996). *Variance Components*. New York: John Wiley & Sons.

Spiegel, M.R. (1968). *Mathematical Handbook of Formulas and Tables*, Schaum's Outline Series in Mathematics, New York: McGraw-Hill.

Verbeke, G. and Molenberghs, G. (2000), *Linear Mixed Models for Longitudinal Data*, New York: Springer.

Wang, Z. and Louis, T.A. (2003), "Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function," *Biometrika*, 90, 765–775.

Wang, Z. and Louis, T.A. (2004), "Marginalizing binary mixed-effect models with covariate-dependent random effects and likelihood inference," *Biometrics*, 60, 884–891.

Wolfinger, R. and O'Connell, M. (1993), "Generalized linear mixed models: a pseudo-likelihood approach," *Journal of Statistical Computation and Simulation*, 48, 233–243.

Zeger, S.L., Liang, K.-Y., and Albert, P.S. (1988), "Models for longitudinal data: a generalized estimating equation approach," *Biometrics*, 44, 1049–1060.

A Taxonomy of Mixing and Outcome Distributions Based on Conjugacy and Bridging

Michael G. Kenward[1] **and Geert Molenberghs**[2]

[1] *I-BioStat, Universiteit Hasselt and Katholieke Universiteit Leuven, B-3590 Diepenbeek, Belgium*

[2] *Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London WC1E7HT, United Kingdom*

Supplementary Materials

A Conjugate Quintuples

To simplify notation, indices will be removed from outcome and parameters.

A.1 The Normal-normal Model

The quintuple is:

$$
\begin{aligned}
f(\boldsymbol{y}|\boldsymbol{b}) &\sim N(X\boldsymbol{\beta}+Z\boldsymbol{b},\Sigma),\\
f(\boldsymbol{b}) &\sim N(\mathbf{0},D),\\
f(\boldsymbol{y},\boldsymbol{b}) &\sim N\left[\left(\begin{array}{c}X\boldsymbol{\beta}\\ \hline \mathbf{0}\end{array}\right),\left(\begin{array}{c|c}ZDZ'+\Sigma & ZD\\ \hline DZ' & D\end{array}\right)\right],\\
f(\boldsymbol{y}) &\sim N(X\boldsymbol{\beta},ZDZ'+\Sigma),\\
f(\boldsymbol{b}|\boldsymbol{y}) &\sim N[DZ'(ZDZ'+\Sigma)^{-1}(\boldsymbol{Y}-X\boldsymbol{\beta}),(Z'\Sigma^{-1}Z+D^{-1})^{-1}].
\end{aligned}
$$

These expressions can be found, among others, in Searle, Casella, and McCulloch (1996).

A.2 The Beta-binomial Model

The quintuple is:

$$
\begin{aligned}
f(y|\theta) &= \binom{n}{y}\theta^{y}(1-\theta)^{n-y},\\
f(\theta) &= \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} = \mathsf{Beta}(\alpha,\beta),\\
f(y,\theta) &= \binom{n}{y}\frac{B(\alpha,\beta)}{\theta^{y+\alpha-1}\beta^{n-y+\beta-1}},\\
f(y) &= \binom{n}{y}\frac{B(y+\alpha,n-y+\beta)}{B(\alpha,\beta)},\\
f(\theta|y) &= \frac{\theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}}{B(y+\alpha,n-y+\beta)} = \mathsf{Beta}(y+\alpha,n-y+\beta).
\end{aligned}
$$

A.3 The Poisson-gamma (Negative Binomial) Model

The quintuple is:

$$
\begin{aligned}
f(y|\theta) &= \frac{\theta^y e^{-y}}{y!}, \\
f(\theta) &= \frac{\theta^{\alpha-1}e^{-\theta/\beta}}{\Gamma(\alpha)} \; = \; \mathsf{Gamma}(\alpha,\beta), \\
f(y,\theta) &= \frac{\theta^{y+\alpha-1}e^{-\theta(1+1/\beta)}}{y!\Gamma(\alpha)}, \\
f(y) &= \frac{1}{y!}\cdot\frac{\Gamma(y+\alpha)}{\Gamma(\alpha)}, \\
f(\theta|y) &= \frac{\theta^{y+\alpha-1}e^{-\theta/[\beta/(\beta+1)]}}{\Gamma(y+\alpha)} \; = \; \mathsf{Gamma}[y+\alpha,\beta/(\beta+1)].
\end{aligned}
$$

A.4 The Gamma-gamma Model

The quintuple is:

$$
\begin{aligned}
f(y|\beta) &= \frac{1}{\Gamma(\alpha)}y^{\alpha-1}e^{-\beta y}\beta^{\alpha} \; = \; \mathsf{Gamma}(\alpha,\beta), \\
f(\beta) &= \frac{1}{\Gamma(\alpha_0)}\beta^{\alpha_0-1}e^{-\beta\beta_0}\beta_0^{\alpha_0} \; = \; \mathsf{Gamma}(\alpha_0,\beta_0), \\
f(y,\beta) &= \frac{1}{\Gamma(\alpha)\Gamma(\alpha_0)}y^{\alpha-1}e^{\beta(y+\beta_0}\beta^{\alpha+\alpha_0-1}\beta_0^{\alpha_0}, \\
f(y) &= \frac{\Gamma(\alpha+\alpha_0)}{\Gamma(\alpha)\Gamma(\alpha_0)}\cdot\frac{\beta_0^{\alpha_0}}{(y+\beta_0)^{\alpha+\alpha_0}} \; = \; \frac{1}{y}\mathsf{Beta}\left(\frac{y}{y+\beta_0};\alpha,\alpha_0\right), \\
f(\beta|y) &= \frac{1}{\Gamma(\alpha+\alpha_0)}e^{-\beta(y+\beta_0))}(y+\beta_0)^{\alpha+\alpha_0}\beta^{\alpha+\alpha_0-1} \; = \; \mathsf{Gamma}(\alpha+\alpha_0,y+\beta_0).
\end{aligned}
$$

A.5 The Dirichlet-multinomial Model

The quintuple is:

$$
\begin{aligned}
f(y_1,\ldots,y_n|\theta_1,\ldots,\theta_n) &= \begin{pmatrix} N \\ y_1,\ldots,y_n \end{pmatrix}\theta_1^{y_1}\ldots\theta_n^{y_n} = \begin{pmatrix} N \\ \boldsymbol{y} \end{pmatrix}\boldsymbol{\theta}^{\boldsymbol{y}}, \\
f(\boldsymbol{\theta}) &= \frac{1}{B(\boldsymbol{\alpha})}\prod_{i=1}^{n}\theta_i^{\alpha_i-1} = \mathsf{Dirichlet}(\boldsymbol{\alpha}), \\
f(\boldsymbol{y},\boldsymbol{\theta}) &= \begin{pmatrix} N \\ \boldsymbol{y} \end{pmatrix}\frac{1}{B(\boldsymbol{\alpha})}\prod_{i=1}^{n}\theta_i^{y_i+\alpha_i-1}, \\
f(\boldsymbol{y}) &= \begin{pmatrix} N \\ \boldsymbol{y} \end{pmatrix}\frac{1}{B(\boldsymbol{\alpha})}B(\boldsymbol{y}+\boldsymbol{\alpha}), \\
f(\boldsymbol{\theta}|\boldsymbol{y}) &= \frac{1}{B(\boldsymbol{y}+\boldsymbol{\alpha})}\prod_{i=1}^{n}\theta_i^{y_i+\alpha_i-1} = \mathsf{Dirichlet}(\boldsymbol{y}+\boldsymbol{\alpha}).
\end{aligned}
$$

with $\sum_{i=1}^{n} y_i = N$, $0 \leq \alpha_i \leq 1$, and $\sum_{i=1}^{n} \alpha_i = 1$.

A.6 General Conjugate Exponential Families

The quintuple is:

$$\begin{aligned}
f(y|\theta) &= \exp\left\{\phi^{-1}[yh(\theta)-\psi(\theta)]+c(y,\phi)\right\},\\
f(\theta) &= \exp\left\{\gamma[\lambda h(\theta)-\psi(\theta)]+c^*(\gamma,\lambda)\right\},\\
f(y,\theta) &= \exp\left[(\phi^{-1}y+\gamma\lambda)h(\theta)-(\phi^{-1}+\gamma)\psi(\theta)+c(y,\phi)+c^*(\gamma,\lambda)\right],\\
f(y) &= \exp\left[c(y,\phi)+c^*(\gamma,\lambda)-c^*\left(\phi^{-1}+\gamma,\frac{\phi^{-1}y+\gamma\lambda}{\phi^{-1}+\gamma}\right)\right],\\
f(\theta|y) &= \exp\left\{(\phi^{-1}+\gamma)\left[\frac{\phi^{-1}y+\gamma\lambda}{\phi^{-1}+\gamma}h(\theta)-\psi(\theta)+c^*\left(\phi^{-1}+\gamma,\frac{\phi^{-1}y+\gamma\lambda}{\phi^{-1}+\gamma}\right)\right]\right\}.
\end{aligned}$$

B Additional Bridge Distribution Results

Apart from the cases listed in the main paper, Wang and Louis (2003) and Molenberghs *et al* (2012) derived a couple of others. These are reviewed here.

For the complementary log-log $H(\eta)=1-\exp[-\exp(\eta)]$, Wang and Louis (2003) derived the bridge:

$$f(b)=\frac{1}{2\pi}\int e^{i(k/\phi-b)\xi}\frac{\Gamma\left(1-\frac{i\xi}{\phi}\right)}{\Gamma(1-i\xi)}d\xi,$$

with $\phi=\left(1+6\pi^{-2}d\right)^{-1}$. Thus, the log-positive stable distribution is the bridge for the complementary log-log link. $\Gamma(\cdot)$ is the gamma function.

Molenberghs *et al* (2012) considered Student's t distribution as an inverse link function, with density

$$h_\nu(\eta)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{\eta^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

Hurst (1995) shows that the characteristic function is

$$(\mathcal{F}h_\nu)(\xi)=\frac{K_{\nu/2}\left(\sqrt{\nu}|\xi|\right)\cdot\left(\sqrt{\nu}|\xi|\right)^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2-1}}, \tag{56}$$

where $K_\alpha(\xi)$ is the modified Bessel function of the second kind, with index α (Abramowitz and Stegun 1964, p. 375). Using generic expression (21) and (56), together with the fact that symmetry implies $k=0$, the bridge distribution for the t link can be shown to be:

$$f_\nu(b)=\frac{1}{2\pi\phi^{\nu/2}}\int e^{-ib\xi}\frac{K_{\nu/2}\left(\sqrt{\nu}\left|\frac{\xi}{\phi}\right|\right)}{K_{\nu/2}\left(\sqrt{\nu}|\xi|\right)}d\xi. \tag{57}$$