

# Random coefficient based dropout models: a finite mixture approach

Alessandra Spagnoli and Marco Alfó

**Abstract** In longitudinal studies, subjects may be lost to follow up (a phenomenon which is often referred to as attrition) or miss some of the planned visits thus generating incomplete responses. When the probability for nonresponse, once conditioned on observed covariates and responses, still depends on the unobserved responses, the dropout mechanism is known to be informative. A common objective in these studies is to build a general, reliable, association structure to account for dependence between the longitudinal and the dropout processes. Starting from the existing literature, we introduce a random coefficient based dropout model where the association between outcomes is modeled through discrete latent effects; these latent effects are outcome-specific and account for heterogeneity in the univariate profiles. Dependence between profiles is introduced by using a bidimensional representation for the corresponding distribution. In this way, we define a flexible latent class structure, with possibly different numbers of locations in each margin, and a full association structure connecting each location in a margin to each location in the other one. By using this representation we show how, unlike standard (unidimensional) finite mixture models, an informative dropout model may properly nest a non informative dropout counterpart.

**Key words:** Finite mixtures, informative dropout, concomitant latent variables

---

Alessandra Spagnoli

Dipartimento di Scienze Statistiche, Sapienza Università di Roma, e-mail: [alessandra.spagnoli@uniroma1.it](mailto:alessandra.spagnoli@uniroma1.it)

Marco Alfó

Dipartimento di Scienze Statistiche, Sapienza Università di Roma, e-mail: [marco.alf@uniroma1.it](mailto:marco.alf@uniroma1.it)

## 1 Introduction

In longitudinal studies, measurements from the same individuals (units) are taken repeatedly over time. These kind of studies often suffer of attrition, since individuals may dropout of the study before the scheduled completion time and thus present incomplete data. When the reasons for dropout is related to unobserved responses, even after controlling for available covariates and responses, the missingness is known to be informative. In such studies, scientific interest may focus on the association structure between the longitudinal measurements and the missingness process. In a seminal paper, [10] discuss a class of statistical models for non ignorable dropout, referred to as Random Coefficient Based Dropout Models (RCBDM), where marginal association between the longitudinal and the survival process arises due only to dependent, outcome-specific, random coefficients. Separate models are hypothesized for the two partially observed processes, which share a common (correlated) set of random coefficients. In the context of binary responses, [2] propose an extension of these models by defining a semi-parametric selection model where the longitudinal and the dropout processes are linked through correlated random effects. The random effects are usually assumed to be Gaussian, but this assumption has been questioned by some authors, see eg [14], since the resulting inferences can be sensitive to assumptions that cannot be verified from the available data. In this perspective, [13], investigated the effect of misspecifying the random effect distribution on parameter estimates and standard errors when a shared parameter model is considered. They showed that, as the number of repeated longitudinal measurements per individual grows, the effect of misspecifying the random effect distribution vanishes for certain parameter estimates, thus referring, implicitly, to theoretical results in [4]. But in several contexts, for example in clinical research, the follow up times are usually short, and individual sequences include only a few information on the random effects; therefore, the choice of the random effect distribution may be important. As far as selection models are entailed, just to mention a few, [19] used a Monte Carlo EM algorithm for linear mixed model with Gaussian random effects, [8] propose a Laplace approximation to overcome the high-dimensional integration over the distribution of the random effects. Numerical integration techniques, such as standard or adaptive Gaussian quadrature, can be used as well. In this paper, we are interested to investigate the association structure between measurement and dropout processes when the random coefficient distribution is left completely unspecified, adopting a finite mixture perspective. We consider a bivariate distribution for the random coefficients that is equal to the product of the marginal distributions only when the dropout mechanism is ignorable. The structure of the paper follows. Section 2 discusses a random coefficient based dropout model where the association between outcomes is modeled through discrete latent effects. Section 3 describes the proposed ML algorithm. Last section contains concluding remarks.

## 2 Random coefficient-based models

Let  $Y_{it}$  represent a set of longitudinal measurements recorded on  $i = 1, \dots, n$  subjects at time  $t = (1, \dots, T)$ , associated to a row vector of  $p$  covariates  $\mathbf{x}_{it} = (x_{it1}, \dots, x_{itp})$ . Let us assume that the observed responses  $y_{it}$  are realizations of a random variable with density in the exponential family and canonical parameter,  $\theta_{it}$ . The canonical parameter is defined as follows:

$$\theta_{it} = \mathbf{x}_{it}^\top \boldsymbol{\beta} + \mathbf{x}_{it}^\top \mathbf{b}_i \quad (1)$$

The terms  $\mathbf{b}_i$ ,  $i = 1, \dots, n$ , are used to model unobserved individual-specific (time-invariant) heterogeneity common to each lower-level unit (time) within the same  $i$ th upper-level unit (individual), while  $\boldsymbol{\beta}$  is a  $p$ -dimensional vector of fixed regression parameters. Those effects that vary across individuals are collected in the design vector  $\mathbf{z}_{it} = (z_{it1}, \dots, z_{itm})$ . We denote with  $\mathbf{R}_i$  the missing data indicator vector, with generic element defined as  $R_{it} = 1$  if the  $i$ th unit drops out at any point in the windows  $(t-1, t)$ ,  $R_{it} = 0$  otherwise. Using this representation, we are implicitly assuming a discrete structure for the time to dropout; however the following arguments apply to continuous time survival process as well. We assume that, once a person drops out, he or she is out forever (attrition). If the designed completion time is denoted by  $T$ , we will have  $T_i \leq T$  measures for each unit. We may introduce an explicit model for the dropout mechanism, conditioning on a set of dropout specific covariates,  $\mathbf{v}_i$ , and the random coefficients in the longitudinal response model:

$$h(\mathbf{r}_i | \mathbf{v}_i, \mathbf{y}_i, \mathbf{b}_i) = h(\mathbf{r}_i | \mathbf{v}_i, \mathbf{b}_i) = \prod_{t=1}^{T_i} h(r_{it} | \mathbf{v}_i, \mathbf{b}_i) \quad i = 1, \dots, n \quad (2)$$

where the corresponding canonical parameter is:  $\phi_{it} = \mathbf{v}_i^\top \boldsymbol{\gamma} + \mathbf{d}_{it}^\top \mathbf{b}_i$ . These models are usually referred to as shared parameter models, see [21],[22], and are based on the assumption of conditional independence between the longitudinal response and the dropout indicator; as it can be easily noticed, they assume a perfect linear correlation between the latent variables in the two equations. In this framework, the joint density of the measurements process  $Y_{it}$  and the missigness process  $R_{it}$  may be written as:

$$\int \left[ \prod_{t=1}^T f(y_{it} | \mathbf{x}_{it}, \mathbf{b}_i) \prod_{t=1}^{T_i} h(r_{it} | \mathbf{v}_i, \mathbf{b}_i) \right] dG(\mathbf{b}_i) \quad (3)$$

where  $G(\cdot)$  represents a discrete or a continuous random coefficient distribution. Here, measurement and missigness processes are assumed to be independent given the random effects  $\mathbf{b}_i$ ; therefore, if any, association is completely accounted for by this latent structure. Correlated random effects represent a further alternative, see eg [1]: the unobservable latent characteristics control for potential overdispersion in the univariate profiles and for association between the measurements and missingness processes; this structure, however, avoids unit correlation estimates, and represents a more flexible approach when compared to shared random effects, where conditional

independence still hold. Let  $\mathbf{b}_i = (\mathbf{b}_{1i}, \mathbf{b}_{2i})$  denote a set of subject and outcome specific random coefficients; then, the joint density of the measurement process  $Y_{it}$  and the missigness process  $R_{it}$  can be factorized as:

$$\int \left[ \prod_{t=1}^T f(y_{it} | \mathbf{x}_{it}, \mathbf{b}_{1i}) \prod_{t=1}^{T_i} h(r_{it} | \mathbf{v}_{it}, \mathbf{b}_{2i}) \right] dG(\mathbf{b}_{1i}, \mathbf{b}_{2i}) \quad (4)$$

An extension of this association structure between random coefficients in the two equations may e defined following [5] where a general random effect model is introduced, where common, partially shared and independent (response-specific) random effects influences the measurement and the dropout processes. While it is common to assume that random effects follows a Gaussian distribution, this does not usually lead a tractable form of the integral in eqs (3) and (4). Among others, [20], [15], [17], show that the choice of the random effect distribution does not have great impact on parameter estimates, except for extreme cases, such as discrete distributions. On the same line, [13] show that when all subjects have a relatively large number of repeated measurements, the effects of a misspecifying the random effect distribution became minimal for model parameter estimates. However, [18] observe that the choice of an appropriate random effect distribution is generally difficult for, at least, three reasons. There is often little information about these unobservables, thus any distributional assumption is difficult to justify, by looking only at observed data. When high dimensional random coefficients are considered, the use of a parametric multivariate distribution imposing the same shape on every dimension can be restrictive. A potential dependence of the random effects on unobserved covariates induces heterogeneity that cannot be captured by common parametric assumptions. In studies where some subjects have few measurements, ie due to dropout, the choice of the random coefficient distribution may therefore be important. A finite mixture approach avoids any unverifiable assumptions upon this distribution, frequently referred to as the *mixing* distribution. In this perspective, [18] propose a semi-parametric shared parameter model to analyze continuous longitudinal responses while adjusting for non monotone missingness. On the same line, [2] jointly analyze longitudinal binary responses subject to dropout trough a selection model with correlated, outcome-specific, random coefficients. Using a finite mixture approach, the log-likelihood function in equation (4) can be written as follows:

$$\ell(\cdot) = \sum_{i=1}^n \left\{ \sum_{k=1}^K f(\mathbf{y}_i | \mathbf{x}_i, \mathbf{b}_{1k}) h(\mathbf{r}_i | \mathbf{v}_i, \mathbf{b}_{2k}) \pi_k \right\} = \sum_{i=1}^n \left\{ \sum_{k=1}^K f(\mathbf{y}_i, \mathbf{r}_i | \mathbf{x}_i, \mathbf{v}_i, \mathbf{b}_k) \pi_k \right\} \quad (5)$$

where  $\pi_k = \Pr(\mathbf{b}_k) = \Pr(\mathbf{b}_{1k}, \mathbf{b}_{2k})$  is the joint probability of locations  $\mathbf{b}_k = (\mathbf{b}_{1k}, \mathbf{b}_{2k})$ ,  $k = 1, \dots, K$ . The use of finite mixture has several significant advantages over parametric models; for instance, this approach is computationally efficient, and the discrete nature of the estimate may help classify subjects in components corresponding to clusters characterized by homogeneous values of random parameters. However, we may notice that the latent variables, as well as the corresponding number of locations, considered in the model to account for individual extra-model departures can

be different when the longitudinal and the missingness processes are considered. For this reason, according to [3], we propose to consider different number of components, locations and/or masses for the latent variables in the two equations. When compared to previously mentioned proposals, see eg equation (5), this is a more flexible representation for the random coefficient distribution and, in particular, this model properly nests a model which describes the dropout as being non informative. That is, the proposed MNAR model properly nests a MAR counterpart, while in case of equation (5) this is not true. Let us suppose the joint bivariate distribution of the random effects has the following marginal representation [9]:

$$P_1 = (u_{1g}, \pi_{1g}), g = 1, \dots, K_1 \quad P_2 = (u_{2l}, \pi_{2l}), l = 1, \dots, K_2$$

with  $\pi_{1g} = \Pr(b_{1i} = b_{1g}), g = 1, \dots, K_1, \pi_{2l} = \Pr(b_{2i} = b_{2l}), l = 1, \dots, K_2$ . That is, we associate to each couple of random coefficients, say  $(\mathbf{b}_{1g}, \mathbf{b}_{2l}), g = 1, \dots, K_1, l = 1, \dots, K_2$ , a mass  $\pi_{gl} = \Pr(\mathbf{b}_{1i} = \mathbf{b}_{1g}, \mathbf{b}_{2i} = \mathbf{b}_{2l})$ , where we do not restrict to consider the same number of components in each profile. While marginals control for heterogeneity in the univariate profiles, joint probabilities describe the association between latent effects in the two submodels. This approach can be related to a standard finite mixture approach where  $K = K_1 \times K_2$  components are used and each of the  $K_1$  locations in the first profile appears in a couple with each of the  $K_2$  locations corresponding to the second profile. Theorem 1 in [6] shows that the elements of any probability matrix  $\pi \in \Pi_{K_1 K_2}$ , where the latter represents the set of  $K_1 \times K_2$  probability matrices, can be decomposed as:

$$\pi_{gl} = \sum_{h=1}^M \tau_h \pi_{1g|h} \pi_{2l|h} \quad (6)$$

for an appropriate choice of  $M$ . Obviously, the following constraints hold:

$$\sum_h \tau_h = \sum_g \pi_{1g|h} = \sum_l \pi_{2l|h} = \sum_g \sum_l \pi_{gl} = 1$$

Therefore, the two set of random coefficients  $\mathbf{b}_{1i}$  and  $\mathbf{b}_{2i}, i = 1, \dots, n$  are independent conditional on belonging to the  $h$ -th (upper level) latent class  $h = 1, \dots, M$ . Random coefficients control for heterogeneity in the univariate profiles, while the hierarchy of the latent components control for potential dependence between outcome-specific random coefficients; this somewhat leads to separability of univariate heterogeneity and bivariate dependence. In some way, the hierarchical structure for  $\pi_{gl}$  resembles a copula-based model, where dependence between profiles is modeled through a copula function joining the marginal distributions for the outcome-specific random coefficients, see [13]. The independence case arises simply when  $M = 1$ ; in this case, the dropout mechanism is non ignorable. The dropout probability still depends on unobserved sources of variation, but these are independent on those influencing the longitudinal measurements. When  $M \geq 2$  we have some form of dependence and we can define different non ignorable dropout mechanisms according to the values assumed by the parameter  $M$ . In this sense, it may be interesting to investigate the

sensitivity of the results with respect to model assumptions when  $M$  moves away from 1, as for example in [16].

### 3 ML Parameter Estimation

The data vector is composed by an observable part  $\mathbf{y}_i$  and by unobservables  $\mathbf{z}_i = (z_{i1}, \dots, z_{iK})$  and  $\zeta_i = (\zeta_{i1}, \dots, \zeta_{iM})$  representing lower and upper level membership vectors. For fixed  $K_1, K_2$  and  $M$ , we assume  $\mathbf{z}_i$  and  $\zeta_i$  have multinomial distributions, with probabilities  $\pi_{gl}, g = 1, \dots, K_1, l = 1, \dots, K_2$  and  $\tau_h, h = 1, \dots, M$ . The complete data likelihood is given by

$$\begin{aligned} L_c(\cdot) &= \prod_{i=1}^n \prod_{g=1}^{K_1} \prod_{l=1}^{K_2} \left\{ f(\mathbf{y}_i, \mathbf{r}_i \mid z_{igl}) \left[ \prod_{h=1}^M \pi_{1g|h} \pi_{2l|h} \tau_h \right]^{\zeta_{ih}} \right\}^{z_{igl}} \\ &= \prod_{i=1}^n \prod_{g=1}^{K_1} \prod_{l=1}^{K_2} \left\{ \prod_{t=1}^{T_i} f(y_{it} \mid z_{igt}) h(r_{it} \mid z_{igt}) \left[ \prod_{h=1}^M \pi_{1g|h} \pi_{2l|h} \tau_h \right]^{\zeta_{ih}} \right\}^{z_{igl}} \end{aligned}$$

where  $\tau_h$  is the prior probability for the  $h$ -th upper level latent class,  $\pi_{1g|h}$  and  $\pi_{2l|h}$  are the conditional probabilities of belonging to the  $g$ -th and the  $l$ -th lower level components, conditional on being in the  $h$ -th class. We partition the parameter vector  $\Psi = (\Psi_g, \Psi_l, \Psi_{glh})$ , where  $\Psi_g$  and  $\Psi_l$  denote the parameter vectors for the longitudinal and the dropout process, respectively, while  $\Psi_{glh} = \{(\pi_{1g|h}, \pi_{2l|h}), \tau_h\}$ . By writing  $f_{igl} = f(\mathbf{y}_i, \mathbf{r}_i \mid z_{igl}) = f(\mathbf{y}_i \mid z_{igl}) h(\mathbf{r}_i \mid z_{igl})$ , the score function is:

$$\begin{aligned} S_c(\Psi_g) &= \sum_{i=1}^n \sum_{l=1}^{K_2} w_{igl} \frac{\partial}{\partial \Psi_g} [\log(f_{igl}) + \log(\pi_{gl})] = \sum_{i=1}^n w_{ig} \frac{\partial}{\partial \Psi_g} [\log(f_{ig})] \\ S_c(\Psi_l) &= \sum_{i=1}^n \sum_{g=1}^{K_1} w_{igl} \frac{\partial}{\partial \Psi_l} [\log(f_{igl}) + \log(\pi_{gl})] = \sum_{i=1}^n w_{il} \frac{\partial}{\partial \Psi_l} [\log(h_{il})] \\ S_c(\Psi_{glh}) &= \sum_{i=1}^n w_{igl} \omega_{ih|gl} \frac{\partial}{\partial \Psi_{glh}} [\log(\pi_{g|h}) + \log(\pi_{l|h}) + \log(\tau_h)] \end{aligned}$$

where  $f_{ig} = f(\mathbf{y}_i \mid z_{igl})$ ,  $h_{il} = f(\mathbf{r}_i \mid z_{igl})$  and  $\omega_{ih|gl}$  is the posterior probability that the  $i$ -th belongs to the  $h$ -th upper level component, given the observed data, the lower level components and the current parameter estimates  $\hat{\Psi}^{(r)}$ . Terms  $w_{igl}$  represent the posterior probability of the unit being in the  $g$ -th component and the  $l$ -th component in the measurement and dropout profiles, respectively. In this way, we may test for independence of the two processes, through standard Wald-type or  $\chi^2$ -based statistics; in particular, when the probability of dropout depends on unobserved sources of variation, eg unobserved heterogeneity, which influences also the longitudinal response, then the dropout process is non ignorable. Molemberghs et al.(2007) show

that for every MNAR model, there is an MAR counterpart that produces exactly the same fit to observed data. This can be more easily understood if we look at previous score equations that resemble the score equations for univariate mixture regression models, representing a potential MAR solution. The ML estimates can be achieved, conditional on  $w_{igl}^{(r)}$ , in subsequent maximization steps. To speed up the EM algorithm, and to ensure identifiability of a two-level latent structure with only one observation level, we may proceed by discretizing  $w_{igl}^{(r)}$  using a MAP rule, as in CEM algorithm (condition *choice*="C" in the algorithm below), or by drawing the component indicator  $\hat{z}_{igl}^{(r)}$  from a multinomial distribution using posterior probabilities as in SEM algorithms (condition *else* below), see [11]. In this case, the last the score equation resembles the one for a polytomous latent class model. The resulting EM algorithm is sketched below.

```

begin
  initialize  $w_{igl}^{(0)}, \Psi^{(0)}, \varepsilon > 0$  repeat
    update  $w_{igl}^{(t)}$  Expectation Step
    if (choice="C")
       $z_{igl}^{(t)} = 1 \iff w_{igl}^{(t)} = \max_{r,v} w_{irv}^{(t)}$ ; Classification Step
    else
      draw  $z_{igl}^{(t)}$  with probs given by  $w_{igl}^{(t)}$ ; Stochastic Step
      estimate  $\beta_1^{(t)}, \beta_2^{(t)}, \mathbf{u}_1, \mathbf{u}_2$  given  $z_{igl}^{(t)}$  Maximization Step
      estimate  $\pi_{g|h}^{(t)}, \pi_{l|h}^{(t)}, \tau_h^{(t)}$  Maximization Step
    until  $Q(\cdot)^{(t)} - Q(\cdot)^{(t-1)} < \varepsilon$ ;
end

```

**Algorithm 1:** Pseudo-code of the proposed SEM-CEM algorithm

## 4 Conclusions

We have defined a random coefficient based dropout model where the association between the longitudinal and the dropout processes is modeled through discrete outcome-specific latent effects. A bidimensional representation for the random effect distribution is used with possibly different numbers of locations in each margin, and a full association structure connecting each location in a margin to each location in the other one. The proposed approach may also be used, for example, in clinical context, where we have only few repeated measurements for subjects. The main advantage of a more flexible representation for the random effects distribution is that the general MNAR model properly nests a model where the dropout mechanism is non informative. This opens to a sensitivity analysis of changes in model parameter estimates as the number of upper level components,  $M$ , moves from 1.

## References

1. Aitkin, M., Alfó, M.: Variance component models for longitudinal count data with baseline information: epilepsy data revisited. *Stat. Comput.* **3**, 291–303 (2003)
2. Alfó, M., Maruotti, A.: A selection model for longitudinal binary responses subject to non-ignorable attrition. *Statist. Med.* **28**, 2435–2450 (2009)
3. Alfó, M., Rocchetti, I.: A flexible approach to finite mixture regression models for multivariate mixed responses. Submitted (2012)
4. Carlin, BP, Louis TA *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman and Hall: New York, 2000.
5. Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., Kenward, M.: Generalized shared-parameter models and missingness at random. *Statist. Mod.* **11**, 279–310 (2011)
6. Dunson, D., Chuanhua, X.: Nonparametric Bayes Modeling of Multivariate Categorical Data. *J. Am. Statist. Assoc.* **104**, 1042–1051 (2009)
7. Follmann, D., Wu, M.: An Approximate Generalized Linear Model with Random Effects for Informative Missing Data. *Biometrics* **51**, 151–168 (1995)
8. Gao, S.: A shared random effect parameter approach for longitudinal dementia data with non-ignorable missing data. *Statist. Med.* **23**, 211–219 (2004)
9. Lagona, F.: Model-based classification of clustered binary data with non ignorable missing values. In *Proceedings of Italian Statistical Society*. CLEUP, Padova (2010)
10. Little, R.J.A.: Modeling the drop-out mechanism in repeated-measures studies. *J. Amer. Statist. Assoc.* **90**, 1112–1121 (1995)
11. McCulloch, C. Maximum likelihood algorithms for generalized linear mixed models. *J. Am. Statist. Assoc.* **92**, 162–170 (1997)
12. Molenberghs, G., Beunckens, C., Sotito, C., Kenward, M. G.: Every missing not at random model has got a missing at random counterpart with equal fit. *J. Roy. Statist. Soc. Ser. B* **70**, 371–388 (2008)
13. Rizopoulos, D., Verbeke, G., Lesaffre, E., Vanrenterghem, Y.: A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics* **64**, 611–619 (2008)
14. Scharfstein, D., Rotnitzky, A., Robins, J.: Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models. *J. Am. Statist. Assoc.* **94**, 1096–1120 (1999)
15. Song, X., Davidian, M., Tsiatis, A.: semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* **58**, 742–753 (2002)
16. Troxel, AB, Ma, G, Heitjan, DF: An Index of Local Sensitivity to Non Ignorability, *Statist. Sinica* **14**, 1221–1237 (2004)
17. Tsiatis, A., Davidian, M.: An overview of joint modeling of longitudinal and time-to-event data. *Statist. Sinica* **14**, 793–818 (2004)
18. Tsonaka, R., Verbeke, G., Lesaffre, E.: A Semi-Parametric Shared Parameter Model to Handle Nonmonotone Nonignorable Missingness. *Biometrics* **65**, 81–87 (2009)
19. Verzilli, C.J., Carpenter, J.R.: A Monte Carlo EM algorithm for random-coefficient-based dropout models. *J. Appl. Statist.* **29**, 1011–1021 (2002)
20. Wang, Y., Taylor, J.: Jointly Modeling Longitudinal and Event Time Data with Application to Acquired Immunodeficiency Syndrome. *J. Am. Statist. Assoc.* **96**, 895–905 (2001)
21. Wu, M., Carroll, R.: Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175–188 (1988)
22. Wu, M., Bailey, K.: Estimation and comparison of changes in the presence of informative censoring by modeling the censoring process. *Biometrics* **44**, 175–188 (1989)