Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models

Anders Skrondal
&
Sophia Rabe-Hesketh

Contents

CHAPTER 1

Modelling different response processes

Alexande James Dumasian
University of Detroit

1.1 Introduction

In this chapter we do not yet introduce latent variables. However, the models discussed represent a building block for the general model framework to be presented in Chapter 4.

There are two general approaches to specifying response processes. In statistics and biostatistics, the most common approach is generalized linear models, whereas the latent response formulation is popular in econometrics and psychometrics. Although very different in appearance, the approaches can generate equivalent models for most response types. However, as we will see in later chapters, the choice of formulation can have implications for estimation and identification.

We start by describing generalized linear models and their extensions. We then introduce the latent response formulation and point out correspondences between approaches. Finally, durations or survival are discussed separately because they do not fit entirely into either of the frameworks. Both continuous and discrete time models are considered.

1.2 Generalized linear models

1.2.1 Conventional generalized linear models

In generalized linear models the explanatory variables affect the response only through the *linear predictor* ν_i for unit i and the response process is fully described by specifying the conditional probability (density) of y_i given the linear predictor. The linear predictor has the form

$$\nu_i = \mathbf{x}_i{}'\boldsymbol{\beta}$$

where $\mathbf{x}_i$ is a vector of covariates and $\boldsymbol{\beta}$ are the corresponding coefficients.

The simplest response process is the continuous. A *linear model*

$$y_i = \nu_i + \epsilon_{ij} \tag{1.1}$$

is usually specified in this case, where

$$\epsilon_{ij} \sim N(0, \sigma^2). \tag{1.2}$$

Depending on the type of covariates in the linear predictor, this could be an analysis of variance (ANOVA) model, an analysis of covariance (ANCOVA) model or a (multiple) linear regression model.

This linear model can also be defined by setting the expectation of the response equal to the linear predictor ν_i,

$$E(y_i|\nu_i) = \nu_i$$

and specifyig that, conditionally on ν_i, the y_i are independently normally distributed with variance σ^2.

For dichotomous responses, the probability that $y_i = 1$, or the expectation of y_i, can be modeled as a *logit*

$$E(y_i|\nu_i) = \frac{\exp(\nu_i)}{1 + \exp(\nu_i)}$$

or a *probit*

$$E(y_i|\nu_i) = \Phi(\nu_i),$$

where Φ is the standard normal cumulative distribution function. Conditional on ν_i, the y_i are independently Bernouilli distributed.

Counts are non-negative integers. A common model for counts is *Poisson regression* with expectation

$$E(y_i|\nu_i) = \exp(\nu_i)$$

and Poisson distribution

$$y_i \sim \frac{\exp(-\mu)\mu^{y_i}}{y_i!}.$$

Counts have a Poisson distribution if the events being counted occur at a constant rate and are mutually independent.

Generalised linear models, including the examples above, can be defined by specifying

1. the functional relationship between the expectation of the response and the linear predictor as

   $$\mu_i = E[y_i|\nu_i] = g^{-1}(\nu_i)$$

   where g is a *link function*, and
2. the conditional probability distribution of the responses as a member of the exponential family with expectation μ_i and, possibly, a common scale parameter ϕ.

Table 1.1 *Percentage heroin measurements in nine illicit heroin preparations.*

Sample	% Heroin			
1	2.2	2.3	2.2	2.3
2	8.4	8.7	2.2	2.3
3	7.6	7.5	2.2	2.3
4	11.9	12.6	2.2	2.3
5	4.3	4.2	2.2	2.3
6	1.1	1.0	2.2	2.3
7	14.4	14.8	2.2	2.3
8	21.9	21.1	2.2	2.3
9	8.8	8.4	2.2	2.3

The exponential family has the form

$$f(y_i|\theta_i,\phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi} + c(y_i,\phi)\right\}$$

where θ_i is the canonical or natural parameter, ϕ is the scale or dispersion parameter and b and c are functions depending on the type of exponential family.

The canonical parameter θ is a function of the mean, $\theta_i = \theta(\mu_i)$. The conditional expectation and variance of the response are given by

$$E(y_i|\nu_i) = \partial b(\theta)/\partial\theta\,|_{\theta=\theta_i}$$

and

$$\mathrm{var}(y_i|\nu_i) = \phi\partial^2 b(\theta)/\partial\theta^2\,|_{\theta=\theta_i}\ .$$

The *canonical link* function is

$$g(\mu) = \theta(\mu).$$

Important members of the exponential family are the normal, binomial, Poisson, gamma and inverse Gassian distributions, see Table **??**.

The canonical link is a natural choice of link function, but in some applications, a non-canonical link function may be more appropriate. The most common link functions are given in Table **??**. One consideration in choosing a link function is the range of values it generates for the mean $\mu = g^{-1}(\nu)$ when $-\infty \le \nu \le \infty$. For example, for binary responses, the logit and probit links are popular choices because they restrict μ to lie in the interval (0,1). Another important consideration relates to the interpretation of the regression parameters. Since $\nu_i = \mathbf{x}_i{}'\boldsymbol{\beta}$, using an identity link corresponds to additive effects of the covariates on the mean and a log link corresponds to multiplicative effects. Using a logit link for binary responses gives a multiplicative model for the odds, $\mu/(1-\mu)$. One reason for using this link is that odds-ratios are invariant with respect to retrospective sampling as in the popular case-control design in epidemiology (*ref*). The choice of distribution determines the

Table 1.2 *Percentage heroin measurements in nine illicit heroin preparations.*

Sample	% Heroin	
1	2.2	2.3
2	8.4	8.7
3	7.6	7.5
4	11.9	12.6
5	4.3	4.2
6	1.1	1.0
7	14.4	14.8
8	21.9	21.1
9	8.8	8.4

residual variance as a function of the mean,

$$\text{var}(y_i|\mu_i) = \phi V(\mu_i)$$

where $V(\mu_i)$ is known as the *variance function* and ϕ as a *dispersion parameter*. For example in the Poisson distribution the variance equals the mean whereas the variance is an independent but constant parameter $\phi = \sigma^2$ under the normal distribution (see Table **??**).

1.2.2 Extensions of generalized linear models

Modeling heteroscedasticity

A classical assumption in linear models is homoscedasticity, i.e. the residual standard deviation σ is assumed constant. However, it is often the case that the residual variance depends on covariates. For example, when comparing the heights of boys and girls aged 12, we would expect the girls' heights to be more variable because many of the girls would have entered puberty while most of the boys would be prepubertal. Since the standard deviation must be positive, it is convenient to model heteroscedasticity as

$$\ln\sigma \;=\; \mathbf{x}'\boldsymbol{\iota}.$$

A similar model can be used for the scale parameter in the scaled probit model, but special care must in this case be exercised to ensure identification of $\boldsymbol{\iota}$.

Models for ordinal responses

The generalized linear model framework is confined to continuous or dichotomous responses because the inverse link of the linear predictor represents the *expectation* of the responses. For ordinal responses, this approach would not be meaningful since the values assigned to the categories are arbitrary.

Models for ordinal responses can instead be defined by linking the cumulative probability $\Pr(y_i < s)$ with the linear predictor,

$$\Pr(y_i < s) \ = \ g^{-1}(-\nu_i + \kappa_s), \quad s = 0, \cdots, S-1 \tag{1.3}$$

where s labels the response category and κ_s is a *threshold parameter*, $\kappa_0 = -\infty < \kappa_1 < \kappa_2 < \cdots < \kappa_S = \infty$. Such models are often called *cumulative models*. By considering the argument of the inverse link function for the case of no covariates, $-\beta_0 + \kappa_s$ it is clear that we cannot separately identify all the thresholds and the constant in the linear predictor. We therefore set $\kappa_1 = 0$ so that a constant can be included among the covariates making the parameterization identical to that used for dichotomous response models in the previous section. Alternatively, κ_1 could be a model parameter if we omit the constant.

The probability of a particular response s becomes

$$\Pr(y_i = s) \ = \ \Pr(y_i < s+1) - \Pr(y_i < s).$$

A binomial distribution is specified for the responses with the probit, logit and complimentary log-log as common links.

The effects of the covariates are assumed to be constant across response categories. When a logit link is used, this property inplies that the odds ratio of y exceeding s, is independent of s, an assumption known as *proportional odds*. This can be seen by noting that the log odds that y exceeds s is

$$\log\left(\frac{1-\Pr(y_i < s)}{\Pr(y_i < s)}\right) = \mathbf{x}_i{}'\boldsymbol{\beta} - \kappa_s$$

so that the odds ratio for two individuals i and j is

$$\exp(\mathbf{x}_i - \mathbf{x}_j)$$

which is the same for all s.

The assumption of constant effects of the covariates across response categories can be relaxed by allowing the thresholds to depend on some covariates $\mathbf{x}_{2i}$ (Terza, 1985)

$$\kappa_{is} \ = \ \mathbf{x}_{2i}{}'\boldsymbol{\varsigma}.$$

The model then becomes

$$\mathrm{P}(y_i \leq y_s) = \Phi(\kappa_{si} - \nu_i) = \Phi(\mathbf{x}_{2i}{}'\boldsymbol{\varsigma} - \mathbf{x}_i{}'\boldsymbol{\beta}), \tag{1.4}$$

It is clear that the coefficients of any variables included in both $\mathbf{x}_{2i}$ and $\mathbf{x}_i$ are not identified.

An alternative device for effectively allowing the thresholds to depend on covariates is to use an ordinal probit link in which the scale parameter is modelled as

$$\ln\sigma_i \ = \ \mathbf{x}_{2i}{}'\boldsymbol{\iota}.$$

The model then becomes

$$\mathrm{P}(y_i \leq y_s) \ = \ \Phi((\kappa_s - \nu_i)/\sigma_i) = \Phi(\kappa^*_{si} - \nu^*_i), \tag{1.5}$$

where

$$\kappa_{sj}^* = \kappa_s/\sigma_i$$

and

$$\nu_i^* = \nu_i/\sigma_i.$$

Such a model was suggested and fitted by Skrondal (1996).

For identification, we will set the intercept in the model for σ, ι_0, to zero. The need for such a restriction is made clear by considering the case $\mathbf{x}_2 = \mathbf{0}$ so that the thresholds become

$$\kappa_{sj}^* = \kappa_s \exp(-\iota_0).$$

With this restriction, the scaled linear predictor has the form

$$\nu_i^* = (\mathbf{x}_i{}'\boldsymbol{\beta}) \exp(-\mathbf{x}_{2i}{}'\boldsymbol{\iota}).$$

It is clear that the model will be at most weakly identified if any covariates are inlcuded in both the linear predictor and the model for σ_i.

Since the thresholds are estimated freely, the intercept in the model for σ_i can be set to an arbitrary constant e.g. $\iota_0 = 0$. The model therefore effectively allows a different linear transformation of the thresholds for each item, i.e. the thresholds can be shifted and rescaled for each item.

Models for nominal responses

There are two kinds of nominal responses; unordered polytomous responses and permutations. The outcome for a unit in the unordered polytomous case is one among several objects, whereas the outcome in the permutation case is a particular ordering of objects. The objects are nominal in the sense that they do not possess an inherent ordering shared by all units as is assumed for ordinal variables. Nominal responses are often the results of decisions in which case the objects are denoted *alternatives*, unordered polytomous variables denoted *first choices* and permutations denoted *rankings*. For instance, in election studies a central outcome variable is the first choice of a voter (say Conservatives) among a set of alternatives (say Labour, Conservatives and Liberals). Sometimes additional information is obtained in the form of rankings of the alternatives (say Liberals preferred to Labour preferred to Conservatives). Note that in some consituencies there may not be a candidate for each of the parties so that the alternative sets vary across voters. We find it useful to use this choice terminology even when decisions are not involved.

For polytomous responses, a separate linear predictor is specified for each alternative. In this respect, the response can be viewed as multivariate and is sometimes represented as a vector having a one for the chosen alternative and zeros for the other alternatives (e.g. Fahrmeir and Tutz, 2001). The probability of the ath alternative category is typically modeled as a *multinomial logit*

$$\Pr(y_i = a) = \frac{\exp(\nu_i^a)}{\sum \exp(\nu_i^b)},$$

where ν_i^a is the linear predictor for alternative a and person i and the sum is over the alternatives available to person i.

We can include unit and alternative-specific covariates or attributes. For instance, consider the choice of supermarket. The linear predictors could include customer specific variables $\mathbf{x}_i$ such as income as well as customer and supermarket specific variables $\mathbf{x}_i^a$ such as travelling time to supermarket

$$\nu_i^a = \boldsymbol{\beta}^{a\prime}\mathbf{x}_i + \boldsymbol{\beta}'\mathbf{x}_i^a.$$

The coefficients of $\mathbf{x}_i^a$ could also differ between alternatives if for example the effect of travelling time is greater for 'corner-shops' than for large supermarkets.

Turning to rankings, let r_i^ℓ be the alternative given rank ℓ among A_i alternatives and $\boldsymbol{R}_i \equiv (r_i^1, r_i^2, \ldots, r_i^{A_i})$ be the ranking of unit i. The logistic model for rankings (e.g. Luce, 1959; Plackett, 1975) is then specified as

$$\Pr(\boldsymbol{R}_i) \;=\; \prod_{\ell=1}^{A_i-1} \frac{\exp(\nu_i^{r_i^\ell})}{\sum_{a=\ell}^{A_i} \exp(\nu_i^{r_i^a})}. \tag{1.6}$$

The model is often denoted the *exploded logit* (Chapman & Staelin, 1982) since the ranking probability is written as a product of first choice probabilities for successively remaining alternatives.

Partial rankings result when unit i only ranks a subset of the full set of alternatives, for example when experimental designs are used in presenting specific subsets of alternatives to different units (e.g. Durbin, 1951; Böckenholt, 1992). Such designs are easily handled by the present methodology by letting the alternative sets $\mathcal{A}_i$ vary over units i. Another kind of partial ranking is a top-ranking where not all alternatives are ranked but only the subset of the $P_i < A_i$ most preferred alternatives. The probability of a top-ranking is simply the product of the first P_i terms in equation (1.6). Note that the first choice probability is obtained as the special case of the ranking probability when $P_i = 1$ for all i.

Consider then the situation where two or more alternatives are *tied* in the sense that they are given the same rank. Although the probability of tied rankings is theoretically zero since the utilities are continuous, ties are often observed in practice. As we will se in Section 1.4, (1.6) has the same form as the partial likelihood as Cox's regression from survival analysis. Exploiting this duality, we can utilise methods for handling ties previously suggested in the survival literature. We hence assume that the units have preference orderings for the tied alternatives, but the ordering is hidden from us. A unit i produces $P_i^* < P_i$ ranks when there are ties. We let t_i^ℓ denote the number of alternatives tied at rank ℓ for the unit, give the tied alternatives arbitrary labels $m = 1, .., t_i^\ell$ and define $\nu_i^{r_i^{\ell[m]}}$ to be the linear predictors of the tied alternatives. The exact expression for the ranking probability is very complex when there are ties (e.g. Kalbfleisch & Prentice, 1980). Following Breslow (1974) we suggest using the

Table 1.3 *Percentage heroin measurements in nine illicit heroin preparations.*

Sample	% Heroin	
1	2.2	2.3
2	8.4	8.7
3	7.6	7.5
4	11.9	12.6
5	4.3	4.2
6	1.1	1.0
7	14.4	14.8
8	21.9	21.1
9	8.8	8.4

following approximation:

$$\Pr(\boldsymbol{R}_i \mid \mathcal{A}_i) \;=\; \prod_{\ell=1}^{P_i^*-1} \prod_{m=1}^{t_i^\ell} \frac{\exp(\nu_i^{r_i^{\ell[m]}})}{\sum_{\ell'=\ell}^{P_i^*} \sum_{m'=1}^{t_i^{\ell'}} \exp(\nu_i^{r_i^{\ell'[m']}})}. \tag{1.7}$$

This approximation amounts to assuming that all tied alternatives are still available when any of the tied alternatives are chosen. Breslow's method appears to work well as long as the number of ties is moderate (Farewell & Prentice, 1980). Note that no approximation is required if there is only one set of tied alternatives and these alternatives have the lowest rank P_i^*, since this represents a top-ranking.

An advantage of using first choice and ranking designs is that the responses are comparable across individuals (see e.g. Brady, 1989). Use of rating scales on the other hand invokes the assumption that individuals use the scale in the same way. However, different subjects tend to use scales in different ways. For instance, some subjects tend to use the high end of the scale whereas others use the low end. In addition there could be differences in the range of ratings used. *find a reference*

1.3 Latent response formulation

An observed discrete response can often be viewed as a partial observation or coarsening of a continuous latent (unobserved) response. *Introduce term 'limited dependent variable'?* Following Pearson (1900), the coarsening is formulated in terms of threshold functions.

The columns of Table **??** list different types of coarsening with the resulting types of variables given in the rows.

1.3.1 Grouped, interval censored, ordinal and dichotomous responses

Let the underlying or *latent response* be denoted y_i^*

$$y_i^* = \nu_i + \epsilon_i.$$

The observed response takes on one of S response categories s, $s = 0, \cdots, S-1$. The relationship between observed and latent response can be written as

$$y_i = \begin{cases} 0 & \text{if} \quad \kappa_{i0} < y_i^* < \kappa_{i1} \\ 1 & \text{if} \quad \kappa_{i1} < y_i^* \leq \kappa_{i2} \\ \vdots & \vdots \qquad \vdots \\ S-1 & \text{if} \quad \kappa_{iS-1} < y_i^* < \kappa_{iS} \end{cases} \tag{1.8}$$

where $\kappa_{i0} = -\infty$ and $\kappa_{iS} = \infty$.

For grouped responses $\kappa_{is} = \kappa_s$, the thresholds do not vary between units and are known apriori. An example of grouped data are salaries grouped into prespecified income brackets with boundaries κ_s. This situation was considered by Stewart (1983).

For interval censored responses the κ_{is} vary between units and are known apriori. For example, time of onset of an illness may not be known exactly but only to lie within a censoring interval between two clinic visits, with the timing of visits varying between individuals.

In the ordinal case, the thresholds κ_s do not vary between units and are unknown parameters. For example, severity of pain may be described as "none", "moderate" or "severe". These outcomes may literally be considered as resulting from pain severity, an unobserved continuous latent response, exceeding certain thresholds. Sometimes we can relax the assumption of constant thresholds to model individual differences in pain tolerance.

If a normal distribution is assumed for the error term, then the model is a *probit* model in the case of two categories (Bliss, 1934) and an *ordinal probit* model for more than 2 categories. The ordinal probit model was suggested by Aitchison and Silvey (1957) and rediscovered by McKelvey and Zavoina (1975). If the error term is assumed to have a logistic distribution,

$$F(\tau) = \Pr(\epsilon_i \leq \tau) = \frac{\exp(\tau)}{1 - \exp(\tau)},$$

the *logit* model is obtained in the case of two categories and an *ordinal logit* or *proportional odds* model in the case of more than two categories. To show the correspondence between the latent response and generalised linear model formulations, we consider the dichotomous case. The latent response model is

$$y_i^* = \nu_i + \epsilon_i, \quad y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

It follows from this that

$$\mathrm{E}(y_i|\nu_i) \quad = \quad \mathrm{P}(y_i = 1|\nu_i) = \mathrm{P}(y_i^* > 0|\nu_i) = \mathrm{P}(\nu_i + \epsilon_i > 0|\nu_i)$$

$$= \quad \mathrm{P}(\epsilon_i > -\nu_i|\nu_i) = \mathrm{P}(\epsilon_i < \nu_i|\nu_i) = F(\nu_i),$$

where the penultimate equality hinges on the symmetry of the density of ϵ_i. Here F corresponds to the inverse link g^{-1}, the standard normal or logistic cumulative distribution functions for the probit and logit, respectively.

As we will see later the latent response formulation is useful even for applications where interpretation in terms of a latent response appears contrived. *Pearson - Yule debate life-death*

1.3.2 Censored responses

The threshold model for doubly censored responses can be written as

$$y_i = \begin{cases} \kappa_{i1} & \text{if} \quad \quad y_i^* < \kappa_{i1} \\ y_i^* & \text{if} \quad \kappa_{i1} < y_i^* \leq \kappa_{i2} \\ \kappa_{i2} & \text{if} \quad \kappa_{i2} < y_i^* \end{cases} \tag{1.9}$$

For right censored responses, $\kappa_{i1} = -\infty$. For example, when measuring ability using a test, we do not know the ability of anyone achieving the maximum score. All we know is that their ability is greater than or equal to that required to achieve the maximum score (*ceiling effect*). For left censored responses $\kappa_{i2} = \infty$. An example is a (*floor effect*) occuring if the minimum score is attained on a test.

The different types of censored responses are prominent in survival analysis. Right-censoring is typically due to the event not having occured by the end of the period of observation. Left-censoring occurs if all we know is that the event had already happened before observation began. If both types of censoring can occur, the responses are doubly censored. One-sided censoring was introduced by Tobit (1958) and is hence denoted the *Tobit*. The model for double censoring is due to Rosett and Nelson (1975) and is often denoted the *Two-limit Probit*.

1.3.3 Nominal responses

As mentioned in Section 1.2.2, there are two types of nominal responses; polytomous and rankings. We can model polytomous responses by assuming that each person assigns a utility u_i^{a*} to each alternative a and that the alternative with the greatest utility is chosen, i.e.

$$y_i = a \text{ if } u_i^{a*} - u_i^{b*} > 0 \, \forall b \; b \neq a. \tag{1.10}$$

If we model the utilities as

$$u_i^{a*} = \nu_i^a + \epsilon_i^a \tag{1.11}$$

where the linear predictors ν_i^a can take on different values for different alternatives and the ϵ_i^a are independently extreme value (Gumbel) distributed,

$$\mathrm{Pr}(\epsilon_i^a \leq \tau) = \exp(-\exp(-\tau)),$$

then it can be shown (McFadden, 1973; Yellott, 1977) that the probability of a particular choice is

$$\Pr(y_i = a|\boldsymbol{\nu}_i) = \frac{\exp(\nu_i^a)}{\sum \exp(\nu_i^b)}$$

where $\boldsymbol{\nu}_i = (\nu_i^1, \nu_i^2, \cdots, \nu_i^{A_i})'$. This is the *multinomial logit* model. The *multinomial probit* model assumes that the ϵ_i^a have a multivariate normal distribution.

The probability of a ranking can be construed as *utility ordering* and be expressed in terms of $A_i - 1$ binary utility comparisons. Under the Gumbel specification of random utility this leads to the closed form exploded logit specification (1.6). That such an explosion results was proven by Luce & Suppes (1965) and Beggs, Cardell & Hausman (1981). The latent response perspective reveals that the exploded logit can be derived without making any behavioural assumption that the choice process is sequential.

Importantly, an analogous explosion is not obtained under normally distributed utilities. The Gumbel model is not reversible in the sense that successive choices starting with the worst alternative would lead to a different ranking probability. Another essential feature of the model is the so-called *independence from irrelevant alternatives*, which we will discuss in Section **??**. indexlatent response formulation—)

1.4 Modeling durations or survival

The outcome of interest is often time to some event. In medicine, the architypal example is survival time from the onset of a condition or treatment to death. In studies of the reliability of products or components, for instance light bulbs, lifetime to failure is often investigated. Instead of using such application specific terms, economists usually refer to durations to events. Generally, we will adhere to this terminology but occasionally we lapse by referring to survival or failure times.

There are two distinguishing features of duration data: durations are always nonnegative and some durations are typically not known because the event has not occurred before the end of the study period. All that is known is that the duration exceeds the period of observation so that it is right-censored. These features imply that one cannot simply apply standard models for continuous responses.

The duration models to be considered are usually not defined as generalized linear models. However, it turns out that generalized linear models can often be adapted to yield likelihoods that are proportional to those implied by duration models.

Durations are either considered in continuous or discrete time, and these cases will be discussed in the two subsequent sections.

In this section we confine the discussion to so called *absorbing events* or states where the unit can only experience the event once. Treatment of multiple events, such as recurring headaches, is deferred to Chapter **??** since the dependence among events for the same unit must be accommodated.

1.4.1 Continuous time

Let the density function for the duration T_i of unit i be denoted $f_i(t)$ and the cumulative density function $F_i(t)$. The survival function, the probability that duration exceeds t, is then defined as $S_i(t) = 1 - F_i(t)$. The *hazard*, sometimes also called the *incidence rate* or *instantaneous risk*, is defined as

$$h_i(t) = \lim_{\Delta \to 0} \left\{ \frac{\mathrm{P}(t \leq T_i < t + \Delta | T_i \geq t)}{\Delta} \right\}. \tag{1.12}$$

Somewhat loosely, this is the 'risk' of an event at time t for unit i given that it has not yet occured, or that unit i is still 'at risk'. It follows from these definitions that

$$h_i(t) = \frac{f_i(t)}{S_i(t)}. \tag{1.13}$$

Accelerated failure time models

Accelerated failure time models can be expressed as

$$\ln T_i = \nu_i + \epsilon_i. \tag{1.14}$$

It follows from this model that the ratio of the durations for two subjects i and i' becomes

$$\ln T_i = \nu_i + \epsilon. \tag{1.15}$$

Proportional hazards models

We assume that the hazards of any two units are proportional and can be modelled as

$$h_i(t) = h^0(t) \exp(\nu_i), \tag{1.16}$$

where $h^0(t)$ is the 'baseline' hazard, the hazard when all covariates are zero (the linear predictor does not include a constant).

If a unit is observed from time t_0 and fails or is censored at time t as indicated by δ_i which is 1 if the unit failed and 0 otherwise, the unit's contribution to the likelihood is

$$l_i = h_i(t)^{\delta_i} \exp(-\int_{t_0}^{t} h_i(T) dT) \tag{1.17}$$

An exponential model assumes that the baseline hazard is constant. This property is relaxed in the piecewise exponential model where the baseline hazard function is assumed to be piecewise constant, with $h^0(T) = h_s$ for $t_{s-1} \leq T < t_s$, $s = 1, 2, \ldots S$ and interval lengths $y_s = t_s - t_{s-1}$.

Interestingly, it turns out that the likelihood of the proportional hazards model is proportional to that of a Poisson model when the data are expanded appropriately.

Let $\theta_i = \exp(\nu_i)$. Clayton (1988) shows that for a unit that was censored or failed in the kth interval, the unit's contribution to the likelihood becomes

$$l_i = (h_k\theta_i)^{\delta_i} \exp(-\sum_{s=1}^{k} h_s\theta_i y_s) \tag{1.18}$$

and this can be rewritten as

$$l_i = \prod_{s=1}^{k} (h_s\theta_i)^{d_{is}} \exp(-h_s\theta_i y_s) \tag{1.19}$$

where $d_{is} = 0$ for $s < k$ and $d_{ik} = \delta_i$. This is proportional to the contribution to the likelihood of k independent Poisson processes with means $h_s\theta_i y_s$. Therefore, by representing each unit by a number of observations (or 'risk sets') equal to the number of time intervals preceding that unit's failure (or censoring) time, the model may be fitted by Poisson regression using d_{is} as the dependent variable, $\log(y_s)$ as an offset and dummies for the time intervals as explanatory variables. Explicitly,

$$d_{is} \sim \text{Poisson}(\mu_{is}),$$

where

$$\ln(\mu_{is}) = \ln(y_s) + \ln(h_s) + \nu_i.$$

Add a small table to show the expansion.

1. $\int |x| dF(x) < \infty$ for all $F \in \mathcal{F}$,
2. for each $m \in R$ there is an $F \in \mathcal{F}$ with $\int x dF(x) = m$, and
3. if $F, G \in \mathcal{F}$, then $\lambda F + (1-\lambda)G \in \mathcal{F}$ for $0 \le \lambda \le 1$.

1. $\int |x| dF(x) < \infty$ for all $F \in \mathcal{F}$,
2. for each $m \in R$ there is an $F \in \mathcal{F}$ with $\int x dF(x) = m$, and
3. if $F, G \in \mathcal{F}$, then $\lambda F + (1-\lambda)G \in \mathcal{F}$ for $0 \le \lambda \le 1$.

1. $\int |x| dF(x) < \infty$ for all $F \in \mathcal{F}$,
2. for each $m \in R$ there is an $F \in \mathcal{F}$ with $\int x dF(x) = m$, and
3. if $F, G \in \mathcal{F}$, then $\lambda F + (1-\lambda)G \in \mathcal{F}$ for $0 \le \lambda \le 1$.

1. $\int |x| dF(x) < \infty$ for all $F \in \mathcal{F}$,
2. for each $m \in R$ there is an $F \in \mathcal{F}$ with $\int x dF(x) = m$, and
3. if $F, G \in \mathcal{F}$, then $\lambda F + (1-\lambda)G \in \mathcal{F}$ for $0 \le \lambda \le 1$.

1. $\int |x| dF(x) < \infty$ for all $F \in \mathcal{F}$,
2. for each $m \in R$ there is an $F \in \mathcal{F}$ with $\int x dF(x) = m$, and
3. if $F, G \in \mathcal{F}$, then $\lambda F + (1-\lambda)G \in \mathcal{F}$ for $0 \le \lambda \le 1$.

1. $\int |x| dF(x) < \infty$ for all $F \in \mathcal{F}$,
2. for each $m \in R$ there is an $F \in \mathcal{F}$ with $\int x dF(x) = m$, and
3. if $F, G \in \mathcal{F}$, then $\lambda F + (1 - \lambda)G \in \mathcal{F}$ for $0 \leq \lambda \leq 1$.

1. $\int |x| dF(x) < \infty$ for all $F \in \mathcal{F}$,
2. for each $m \in R$ there is an $F \in \mathcal{F}$ with $\int x dF(x) = m$, and
3. if $F, G \in \mathcal{F}$, then $\lambda F + (1 - \lambda)G \in \mathcal{F}$ for $0 \leq \lambda \leq 1$.

1. $\int |x| dF(x) < \infty$ for all $F \in \mathcal{F}$,
2. for each $m \in R$ there is an $F \in \mathcal{F}$ with $\int x dF(x) = m$, and
3. if $F, G \in \mathcal{F}$, then $\lambda F + (1 - \lambda)G \in \mathcal{F}$ for $0 \leq \lambda \leq 1$.

1. $\int |x| dF(x) < \infty$ for all $F \in \mathcal{F}$,
2. for each $m \in R$ there is an $F \in \mathcal{F}$ with $\int x dF(x) = m$, and
3. if $F, G \in \mathcal{F}$, then $\lambda F + (1 - \lambda)G \in \mathcal{F}$ for $0 \leq \lambda \leq 1$.

1. $\int |x| dF(x) < \infty$ for all $F \in \mathcal{F}$,
2. for each $m \in R$ there is an $F \in \mathcal{F}$ with $\int x dF(x) = m$, and
3. if $F, G \in \mathcal{F}$, then $\lambda F + (1 - \lambda)G \in \mathcal{F}$ for $0 \leq \lambda \leq 1$.

Therefore, one approach to survival modelling is to divide the follow-up period into intervals over which the hazard can be assumed to be constant and use Poisson regression. Another approach is to define as many intervals as there are unique failure times with each interval starting at (just after) a unique failure time and ending at (just after) the next largest unique failure time. This corresponds to the famous *Cox proportional hazards model* since a 'saturated' or nonparametric model with a separate constant for each risk set *explain somewhere* is used for the baseline hazard. The famous *partial likelihood* of Cox's regression is obtained by eliminating the baseline hazard using the profile likelihood approach (johansen, 1980) giving

$$\prod_i \frac{\exp(\nu_i)}{\sum \exp(\nu_j).}$$

think about notation

1.4.2 Discrete time

In practice, durations are often observed discretely. This could be due to interval censoring or due to the opportunities for the events arising at discrete as in the number of elections to victory *?*.

Proportional odds model

The proportional odds model introduced in Section 1.2.2 can be used in modelling discrete time durations.

For uncensored durations, the probability that T is less than t_s becomes

$$P_{is} = \mathrm{P}(T_i < t_s) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i + \kappa_s)}{1 + \exp(\boldsymbol{\beta}'\mathbf{x}_i + \kappa_s)} \tag{1.20}$$

and the probability that the survival time lies in the kth interval $t_{s-1} \leq T_i < t_s$ is

$$\Pr(t_{s-1} \leq T_i < t_s) = \mathrm{P}(T_i < t_s) - \mathrm{P}(T_i < t_{s-1}) \tag{1.21}$$

with $\mathrm{P}(T_i < t_0) = 0$ and $\mathrm{P}(T_i < t_1) = 1$. In the absence of censoring, this is the likelihood contribution of all observations whose survival times lie in the sth interval. For observations that are censored after the sth interval, the likelihood contribution is $1 - P_s$. The proportional odds model has been used for discrete survival time data by Bennett, Ezzet and Whitehead and Hekeker.

The proportional odds model can also be interpreted as a linear model for an underlying or latent continuous response y^*,

$$y^* = \boldsymbol{\beta}'\mathbf{x} + \epsilon \tag{1.22}$$

where ϵ has a logistic distribution. (If a standard normal distribution is assumed for ϵ, the ordinal probit model is obtained.) The event occurs in the kth interval if $\kappa_{s-1} \leq y^* < \kappa_s$, i.e.,

$$\Pr(T < t_s) = \Pr(y^* < \kappa_s). \tag{1.23}$$

The latent response y^* can therefore be thought of as a monotonic transformation of T so that $y^* = \kappa_s$ corresponds to $T = t_s$. By constraining the threshold parameters κ_s to be equally spaced, the appropriateness of the linear regression model in (1.22) for the (untransformed) continuous time can be assessed. *accelerated failure time model for latent response?*

Models based on the discrete time hazard

The discrete time hazard h_s for the sth interval is defined as the probability that the event occurs in the sth interval given that it has not already occurred,

$$h_i(s) = \Pr(t_{s-1} \leq T_i < t_s | T_i \geq t_{s-1}) = \frac{\Pr(t_{s-1} \leq T_i < t_s)}{\Pr(T_i \geq t_{s-1})}. \tag{1.24}$$

The likelihood contribution of someone whose survival time lies in the sth interval is

$$h_i(s) \prod_{l=1}^{s-1} [1 - h_i(l)] = \prod_{l=1}^{s} h_i(l)^{y_{il}} [1 - h_i(l)]^{(1-y_{il})} \quad \text{with } y_{is} = 1 \tag{1.25}$$

where y_{il} is an indicator variable that is equal to 1 if the event occurred in the lth interval and equal to 0 otherwise, i.e. $y_i = 0$ when $l < s$ and $y_i = 1$ when $l = s$.

The likelihood contribution of someone who was censored after the kth interval has the same form with

$$\prod_{l=1}^{k}[1-h_i(l)] = \prod_{l=1}^{s} h_i(l)^{y_{il}}[1-h_i(l)]^{(1-y_{il})} \quad \text{with } y_{is}=0. \tag{1.26}$$

The likelihood contributions of both censored and non-censored observations are just the likelihood contributions of s independent binary responses y_{il}, $l = 1, \cdots, s$ with Bernouilli probabilities $h_i(l)$. Therefore, by expanding the data to s records per person and constructing the indicator variable y_{il}, discrete time survival models can be written as generalized linear models for binary responses. One possibility is to use logistic regression with a separate constant for each interval,

$$\log \frac{h_i(l)}{1-h_i(l)} = \nu_i + \kappa_l. \tag{1.27}$$

Note that this model is often referred to as a proportional odds model. However, whereas proportionality here applies to the conditional odds of the event happening in an interval given that it has not already happened, proportionality in the previous section applied to the odds of the event happening in a given interval or earlier. Another term for this model is the continuation ratio logit model or the logistic model for discrete time survival data. Continuatuion ratio models are useful for sequential processes in which stages (such as education attainment levels) cannot be skipped and interest focuses on the odds of (not) continuing beyond a stage given that the stage has been reached. See Jenkins and Singer and Willett for introductions to the model. *mention non-proportionality can be introduced as interactions.*

If a Cox proportional hazards model is assumed for the unobserved continuous survival times and the observed discrete survival times are treated as interval censored, it can be shown that the likelihood contributions are equal to those in (1.25) and (1.26) if a complementary log-log link is used for the discrete time hazard , i.e.

$$\log(-\log[1-h_i(l)]) = \nu_i + \kappa_l. \tag{1.28}$$

1.5 Summary

We have introduced a wide range of responce processes, including dichotomous, grouped, censored, ordinal, polytomous, ranking, counts, and continuous and discrete time durations. Most of the processes can more or less directly be expressed as generalised linear models, and many as latent response models. The models for the response processes will serve as building blocks for the more general models introduced later. Furthermore, the application chapters are structured according to type of response process.

1.6 References

A. R. Adams, “Band-structure engineering for low-threshold high-efficiency semiconductor lasers”, Electron. Lett., Vol. 22, No. 5, pp. 249–250, Feb. 1986.

E. Yablonovitch and E. O. Kane, "Reduction of lasing threshold current density by the lowering of valence band effective mass", J. Lightwave Technol., Vol. LT-4, No. 5, pp. 504–506, May. 1986.

T. J. Andersson et al.: Appl. Phys. Lett., 51 (1987) 752.

M. Fukuda, M. Okayasu, J. Temmyo, and J. Nakano, "Degradation behavior of 0.98mm strained quantum well InGaAs/AlGaAs lasers under high-power operation", IEEE J Quantum Electron., Vol. 30, pp. 471–476, 1994.

M. Asada, A. R. Adams, K. E. Stubkjaer, Y. Suematsu, Y. Itaya and S. Arai, "The temperature dependence of the threshold current of GaInAsP/InP DH lasers", IEEE J Quantum Electron., Vol. QE-17, No. 5, pp. 611–619, May 1981.

A. Kasukawa, Y. Imajo, I. J. Murgatroyd and H. Okamoto, "Effect of well number in 1.3μm GaInAsP graded-index separate-confinement-heterostructure multiple quantum well (GRIN-SCH-MQW) laser diodes", Trans. IEICE, E 73, pp. 59–62, 1990.

W. T. Tsang, "Extremely low threshold (AlGa)As modified multiquantum well heterostructure lasers grown by molecular-beam epitaxy", Appl. Phys. Lett., vol. 39, No. 10, pp. 786–788, Nov. 1981.

W. T. Tsang, "Extremely low threshold (AlGa)As graded-index waveguide separate- confinement heterostructure lasers grown by molecular-beam epitaxy", Appl. Phys. Lett., vol. 40, No. 3, pp. 217–219, Feb. 1982.

U. Koren, B. I. Miller, Y. K. Su, T. L. Koch and J. E. Bowers, "Low internal loss separate confinement heterostructure InGaAs/InGaAsP quantum well laser", Appl. Phys. Lett., Vol. 51, No. 21, pp. 1744–1746, 1987.

P. J. A. Thjis, L. F. Tiemeijer, J. J. M. Binsma, and T. van Dogen, "High-performance 1.5μm wavelength InGaAs-InGaAsP strained quantum well lasers and amplifiers", IEEE J Quantum Electron., Vo. 27, No. 6, pp. 1426–1439, Jun. 1991.

A. Kasukawa, I. J. Murgatroyd, Y. Imajo, T. Namegaya, H. Okamoto and S. Kashiwa, "1.5μm GaInAs/GaInAsP graded index separate confinement heterostructure multiple quantum well (GRIN-SCH-MQW) laser diodes grown by metalorganic chemical vapor deposition (MOCVD)", Electron. Lett., 25, pp. 659–660, 1989.

T. Tanbun-Ek, R. a. Logan, N. A. Olsson, H. Temkin, A. M. Sergent and K. W. Wecht, "High power output 1.48-1.51μm continuously graded index separate confinement strained quantum well lasers", Appl. Phys. Lett., Vol. 57, No. 16, pp. 224–226, Jul. 1990.

H. Kakibayashi, F. Nagata, "Composition analysis of GaxAl1-xAs superstructure by equal thickness fringes", Pro. XI Int. Cong. on Electron Microscopy, Kyoto, Japan, pp. 1495–1496, 1986.

H. Asano, S. Takano, M. Kawaradani, M. Kitamura, and I. Mito, "1.48mm high-power InGaAs/InGaAsP MQW LD's for Er-doped fiber amplifiers", IEEE, Phton. Technol. Lett., Vol. 3, No. 5, pp. 415–417, 1991.

H. Hirayama, Y. Miyake and M. Asada,"Analysis of current injection efficiency of separate-confinement-heterostructure quantum-film lasers", IEEE J. Quantum Electron., Vol. 28, No. 1, pp. 68–74, Jan. 1992.

P. Brosson, C. Rabourie, L. Le Gouezigou, J. L. Lievin, J. Jacqest, F. Leblond, A. Olivier and D. Leclerc, "Experimental determination of carrier-induced differential loss in 2-section GaInAsP/InP laser-waveguide", Electron. Lett., Vol. 25, No. 24, pp. 1623–1624, Nov. 1989.

A. Kasukawa, T. Namegaya, N. Iwai, N. Yamanaka, Y. Ikegami, and N. Tsukiji, "Extremely high power 1.48mm GaInAsP/InP GRIN-SCH strained MQW lasers", IEEE, Phton. Technol. Lett., Vol. 6, No. 1, pp. 4–6, 1994.

D. Fekete, K. Chan, J. M. Ballantyne, and L. F. Eastman, "Graded-index separate-confinement InGaAs/GaAs strained-layer quantum well laser grown by metalorganic chemical vapor deposition", Appl. Phys. Lett., Vol. 49, No. 24, pp. 1659–1660, 1986.

M. Okayasu, M. Fukuda, T. Takeshita, and S. Uehara, "Stable operation (over 5000 h) of high-power 0.98mm InGaAs-GaAs strained quantum well ridge waveguide lasers for pumping Er3+-doped fiber amplifiers", IEEE Photon. Technol. Lett., Vol. 2, No. 10, pp. 689–691, 1990.

S. Ishikawa, K. Fukagai, H. Chida, T. Miyazaki, H. Fujii, and K. Endo, "0.98-1.02mm strained InGaAs/AlGaAs double quantum-well high-power lasers with GaInP buried waveguide", IEEE J. Quantum Electron., Vol. 29, No. 6, pp. 1936–1942, 1993.

H. Meier: Technical Digest of 20th European Conference on Optical Communications, Firenze, Sep., 1994, p. 974.

J. S. Major, W. E. Plano, D. F. Welch, and D. Scifres, "Single-mode InGaAs-GaAs laser diodes operating at 980nm", Electron. Lett., Vol. 27, No. 6, pp. 539–541, 1991.

M. Ohkubo, T. Ijichi, A. Iketani, and T. Kikuta, "980nm aluminum-free InGaAs/InGaAsP/InGaP GRIN-SCH SL-SQW lasers", IEEE J Quantum Electron., Vol. 30, pp. 408–414, 1994.

H. Horie, N. Arai, Y. Mitsuishi, N. Komuro, H. Kaneda, H. GotohCM. Usami, and Y. Matsushima, "Greater than 500-mW CW kink-free single transverse-mode operation of weakly index guided buried-stripe type 980nm laser diodes", IEEE, Phton. Technol. Lett., Vol. 12, No. 10, pp. 1304–1306, 2000.

H. Horie, H. Ohta, and T. Fujimori, "Reliability improvement of 980nm laser diodes with a new facet passivation process", IEEE, J. Select. Topics Quantum Electron., Vol. 5, pp. 832–838, 1999.

CHAPTER 2

General model framework

Alexande James Dumasian
University of Detroit

2.1 Introduction

The general model framework unifies and generalises multilevel, structural equation and longitudinal models. All the response processes described in Chapter 2 are accommodated. The latent structure is generalised to combine random coefficients and factors in the same model and allow latent variables of both kinds to vary at different levels. We will also relax the assumption of multivariate normality of the latent variables by using other continuous distributions or discrete distributions. Latent class models are also accommodated.

We first unify conventional random coefficient and factor models and then introduce the general model framework which consists of three parts:

- the response model,
- the structural model for the latent variables and
- the distribution of the latent variables.

2.2 Unifying random coefficient and factor models

Random coefficient and structural equation models are more similar than generally acknowledged. Recall the random coefficient model from equation (**??**)

$$\boldsymbol{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \boldsymbol{Z}_j\mathbf{u}_j + \boldsymbol{\epsilon}_j \tag{2.1}$$

and the measurement part of the structural equation model in equation (**??**)

$$\boldsymbol{y}_j = (\boldsymbol{\nu} + \boldsymbol{K}\mathbf{x}_j) + \boldsymbol{\Lambda}\boldsymbol{\eta}_j + \boldsymbol{\epsilon}_j \tag{2.2}$$

where some subscripts and superscripts have been omitted. In the random coefficient model $\boldsymbol{y}_j$ represents the vector of responses for the level 1 units within the jth level 2 unit and in the measurement model the responses on the items. Although different in interpretation, these models have a similar structure. Both have an error term $\boldsymbol{\epsilon}_j$, the random effects $\mathbf{u}_j$ correspond to the factors $\boldsymbol{\eta}_j$ and the design matrix for the random effects $\boldsymbol{Z}_j$ corresponds to the factor loading matrix $\boldsymbol{\Lambda}$. The main difference is that $\boldsymbol{Z}_j$ is a known matrix of covariates whereas $\boldsymbol{\Lambda}$ is an unknown parameter matrix. While $\boldsymbol{Z}_j$ can differ between level 2 units, $\boldsymbol{\Lambda}$ is constant. The fixed parts $\mathbf{X}_j\boldsymbol{\beta}$ and $(\boldsymbol{\nu} + \boldsymbol{K}\mathbf{x}_j)$ serve the same purpose. In the case of a single covariate, the terms for the ith row are $\beta_0 + x_{ij}\beta$ and $\nu_i + k_i x_j$, respectively. Whereas the former assumes constant effects and varying covariates, the latter assumes (non-randomly) varying effects and constant covariates. However, this difference is superficial since dummy variables for the i can be used to allow coefficients to depend on i in the random coefficient model and different covariates used for different i to represent a varying covariate in the measurement model.

To facilitate the subsequent development, we will refer to the elementary observations i as *level 1 units* whether they are the lowest level units in a multilevel setting or items in a measurement model. Furthermore, we will refer to both factors and random effects as *latent variables*, denoted $\boldsymbol{\eta}_j$. $\boldsymbol{Z}_j$ and $\boldsymbol{\Lambda}$ will henceforth be called the *structure matrix* and denoted $\boldsymbol{\Lambda}_j$. The model relating responses to factors and/or random coefficients will be called a *response model* and can then be written as

$$\boldsymbol{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \boldsymbol{\Lambda}_j\boldsymbol{\eta}_j + \boldsymbol{\epsilon}_j \tag{2.3}$$

where $\boldsymbol{\Lambda}_j$ can contain both variables and parameters.

We can make the form of $\boldsymbol{\Lambda}_j\boldsymbol{\eta}_j$ explicit by writing it as

$$\boldsymbol{\Lambda}_j\boldsymbol{\eta}_j = \sum_{m=1}^{M} \eta_{mj}\boldsymbol{Z}_{mj}\boldsymbol{\lambda}_m,$$

where η_{jm} is the mth latent variable, $\boldsymbol{Z}_{mj}$ is an $n_j \times p_m$ (design) matrix of fixed constants and $\boldsymbol{\lambda}_m$ are p_m parameters associated with the mth latent variable. Here the product $\boldsymbol{Z}_{mj}\boldsymbol{\lambda}_m$ represents the mth column of $\boldsymbol{\Lambda}_j$. Note that $\boldsymbol{\lambda}_m$ is therefore not a vector in the matrix $\boldsymbol{\Lambda}_j$. The mth latent variable is a factor if $\boldsymbol{Z}_{mj}$ is a matrix of zeros and ones. This can be seen by considering a two factor model as an example:

$$\begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \end{bmatrix} \begin{bmatrix} \eta_{1j} \\ \eta_{2j} \end{bmatrix} = \eta_{1j} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \end{bmatrix} + \eta_{2j} \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda_{42} \\ \lambda_{52} \end{bmatrix}. \tag{2.4}$$

Here the $\boldsymbol{\lambda}_m$ contain the unique non-zero factor loadings for the mth factor and the role of $\boldsymbol{Z}_{mj}$ is to assign the correct factor loadings to the correct items.

The model in (2.3) can now be written as

$$\boldsymbol{y}_j = \mathbf{X}_j\boldsymbol{\beta} + \sum \eta_{mj}\boldsymbol{Z}_{mj}\boldsymbol{\lambda}_m + \boldsymbol{\epsilon}_j$$

and the ith row becomes

$$y_{ij} = \mathbf{x}_{ij}{}'\boldsymbol{\beta} + \sum_{m=1}^{M} \eta_{mj}\mathbf{z}_{mij}{}'\boldsymbol{\lambda}_m + \epsilon_{ij}, \tag{2.5}$$

where $\mathbf{x}_{ij}$ is the ith row of $\mathbf{X}_j$ and $\mathbf{z}_{mij}$ is the ith row of $\boldsymbol{Z}_{mj}$.

The mth latent variable is a random coefficient if $p_m = 1$ so that $\mathbf{z}_{mij}$ is a scalar with corresponding parameter $\lambda_m = 1$. Equation (2.5) then becomes

$$y_{ij} = \mathbf{x}_{ij}{}'\boldsymbol{\beta} + \sum_{m=1}^{M} \eta_{mj} z_{mij} + \epsilon_{ij}.$$

2.3 Response model

Conditional on the latent and explanatory variables, the response model is a generalised linear model specified via a linear predictor, a link and a distribution from the exponential family. Any of the conditional densities for a generalized linear model can be specified for the responses, including the extensions introduced in Chapter 1. Models for scale parameters and thresholds may also be specified. We will express the linear predictor in two different ways; using the generalized factor model (GF) formulation and the generalized random coefficient (GRC) formulation.

2.3.1 Linear predictor in GF formulation

The model in (2.3) can be defined by writing the linear predictors for the responses on unit j as

$$\boldsymbol{\nu}_j = \mathbf{X}_j\boldsymbol{\beta} + \boldsymbol{\Lambda}_j\boldsymbol{\eta}_j, \tag{2.6}$$

speficying an identity link and a normal density of the $\boldsymbol{y}_j$ given $\boldsymbol{\nu}_j$. We can now generalise the model to L levels as

$$\boldsymbol{\nu}_k = \mathbf{X}_k\boldsymbol{\beta} + \sum_{l=2}^{L} \boldsymbol{\Lambda}_k^{(l)}\boldsymbol{\eta}_k^{(l)}, \tag{2.7}$$

where $\boldsymbol{\nu}_k$ is the vector of linear predictors for all units in a particular level L unit k.

2.3.2 Linear predictor in GRC formulation

The linear predictor of the model in (2.5) is

$$\nu_{ij} = \mathbf{x}_{ij}{}'\boldsymbol{\beta} + \sum_{m=1}^{M} \eta_{mj}\mathbf{z}_{mij}{}'\boldsymbol{\lambda}_m. \tag{2.8}$$

This can be extended to L levels as

$$\nu = \mathbf{x}'\boldsymbol{\beta} + \sum_{l=2}^{L}\sum_{m=1}^{M_l} \eta_m^{(l)} \mathbf{z}_m^{(l)\prime} \boldsymbol{\lambda}_m^{(l)}, \tag{2.9}$$

where M_l is the number of latent variables at level l and we have omitted observation subscripts to simplify notation.

2.4 Structural model for the latent variables

The structural model for the latent variables has the form

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\mathbf{w} + \boldsymbol{\zeta}, \tag{2.10}$$

where $\mathbf{B}$ is an $M \times M$ matrix, $M = \sum_l M_l$, $\mathbf{w}$ is a vector of Q covariates, $\boldsymbol{\Gamma}$ is an $M \times Q$ design matrix and $\boldsymbol{\zeta}$ is a vector of M errors or disturbances. Note that (2.10) resembles one level structural equation models (Muthén, 1984). The important difference is that latent variables vary at different levels in our framework. Each element of $\boldsymbol{\zeta}$ varies at the same level as the corresponding element of $\mathbf{u}$.

We restrict the form of the regressions of latent variables on other latent variables in two ways:

1. The regressions among latent variables at a particular level are *recursive*
2. Higher level latent variables cannot be regressed on lower level latent variables

By the first restriction we mean that, assuming that the elements of $\boldsymbol{\eta}^{(l)}$ are suitably permuted, the expression for $\eta_{M_l}^{(l)}$ can be substituted into the expression for $\eta_1^{(l)}$ to $\eta_{M_l-1}^{(l)}$, the expression for $\eta_{M_l-1}^{(l)}$ can then be substituted into the regression for $\eta_1^{(l)}$ to $\eta_{M_l-2}^{(l)}$, etc., until all $\eta_m^{(l)}$ are eliminated from the right-hand side of the equation. Substituting the final expressions into equation (2.6) or (2.8) then yields the *reduced form* where the only latent variables remaining are the $\boldsymbol{\zeta}$. An implication of the first restriction is that we cannot have simultaneous effects with a particular latent variable regressed on another and vice versa. The two restrictions together imply that the matrix $\mathbf{B}$ is upper diagonal (if the elements of $\boldsymbol{\eta}^{(l)}$ are permuted appropriately), since the elements of $\boldsymbol{\eta} = (\boldsymbol{\eta}^{(2)\prime}, \cdots, \boldsymbol{\eta}^{(L)\prime})'$ are arranged in increasing order of l. Restrictions similar to the first also apply to $\boldsymbol{\Gamma}\mathbf{w}$. Latent variables cannot be regressed on observed variables varying at a lower level. *give a hypothetical example and path diagram for a multilevel SEM*

2.5 Distribution of the latent variables

The structure of the latent variables is specified by the number of levels L and the number of latent variables M_l at each level. Here a particular level may coincide with a level of clustering in the hierarchical dataset, for example, when a latent variable 'at the school level' varies between schools. However, there will often not be a direct correspondence between the levels of the model and the levels of the data hierarchy. For the models, we define 'level', 'unit at a level' and 'latent variable at a level' as follows:

1. A unit at level 1 is an elementary unit of observation,
2. a unit k at level l is a cluster of elementary observations,
3. if $l > 1$, these elementary observations form $n_k^{(l-1)}$ subsets of observations, representing units at level $l-1$,
4. a latent variable at level l varies between the units at level l but not within the units,
5. the units at level l are conditionally independent given the latent variables at level $l+1$ and above.

Here the basic assumption is that latent variables $\boldsymbol{\zeta}$ at the same level may be mutually correlated, whereas latent variables at different levels are independent. In the following subsections we describe different specifications of the distribution of $\boldsymbol{\zeta}$.

2.5.1 Continuous latent variables

When the latent variables are continuous, the models are often referred to as *latent trait models*. The predominant distributional assumption in this case is multivariate normality with mean zero and covariance matrix $\boldsymbol{\Psi}^{(l)}$ at level l. Importantly, the likelihood cannot be expressed in closed form in this case unless the responses are conditionally normally distributed. However, closed form expressions exist for some combinations of latent variable and response distributions in the case of simple random intercept models. Examples include conditionally Poisson distributed responses combined with a log-gamma distribution for the random intercept, and conditionally binomially distributed responses compared with a beta distribution for the random intercept. *If there are no within-cluster covariates?*, these combinations correspond to the negative binomial and beta-binomial models respectively. (Heckman and Willis?) ?Bayesian nonparametric prior

check out terms conjugate etc. and refer to Bayesian books Hougaard In order to avoid making strong assumptions about the distribution of the latent variables, flexible parametric distributions can be used such as finite mixtures of multivariate normal distributions (Magder and Zeger, 1996). Another approach is to use a truncated Hermite series expansion as suggested by Gallant and Nychka (1987) and Davidian and Gallant (1992).

2.5.2 Discrete latent variables

We consider three types of models with discrete latent variables:

- Semiparametetric mixture models
- Latent class models
  - Ordered latent class models
  - Unordered or conventional latent class models

None of these models has previously been considered in multilevel settings with more than two levels. While the semiparametric mixture model can easily be extended to an arbitrary number of levels, we will confine our discussion of the latent class models to the two level case.

Consider a unidimensional latent variable. If one is unwilling to make any distributional assumptions regarding the latent variables, a semiparametric approach can be acopted. Under certain conditions (Laird, 1978; Lindsay *et al.*, 1991 , ??), the non-parametric cumulative distribution function of the latent variable is a step function with the number of steps, their location and size determined to maximize the likelihood. The latent variables therefore have a discrete distribution having non-zero probabilities π_r at a finite number of points, $e_r, r = 1, \cdots, R$ often known as *masses* as illustrated in Figure 2.1. For a multivariate latent variables with M elements, the masses are located at points $\boldsymbol{e}_r$ in M dimensions. This approach is known is non-parametric maximum likelihood (NPML) and the models are also known as semi-parametric mixture models. Review of maximum likelihood theory can be found in ... In some cases, the locations are fixed apriori (Mislevy, 1984; Heinen, 1996).

put in a diagram of spikes and cdf Considering the two level case, the same type of model can be used if the population is believed to consist of latent classes c whose members have identical values $\boldsymbol{e}_c$ of the latent variables. The number of classes C would in this case be determined either apriori or using model selection criteria. The latent classes are in this case ordered along each latent dimension. For example, if the response model is a unidimensional factor model, the classes are ordered according to the magnitude of the factor. This is in contrast to unordered or conventional latent class models which allow factor loadings to differ between classes so that is not meaningful to compare the magnitude of the factor between classes.

In a conventional latent class model, the vector of linear predictors for the jth level 2 unit in the cth latent class is given by

$$\boldsymbol{\nu}_c = \mathbf{X}_j\boldsymbol{\beta}_c + \boldsymbol{\Lambda}_c^{(2)}\boldsymbol{\eta}_c^{(2)}, \tag{2.11}$$

Equivalently, the linear predictor for the ith level one unit belonging to the jth level 2 unit in class c can be written as

$$\nu_c = \mathbf{x}_{ij}{}'\boldsymbol{\beta}_c + \sum_{m=1}^{M_2} \eta_{mc}^{(2)}\boldsymbol{\lambda}_{mc}^{(2)\prime}\mathbf{z}_{mij}^{(2)}, \tag{2.12}$$

The probability of belonging to a particular class can itself depend on covariates through a multinomial logit

$$\pi_{cj} = \frac{\exp(\boldsymbol{\varrho}_c{}'\mathbf{v}_j)}{\sum_d \exp(\boldsymbol{\varrho}_d{}'\mathbf{v}_j)}.$$

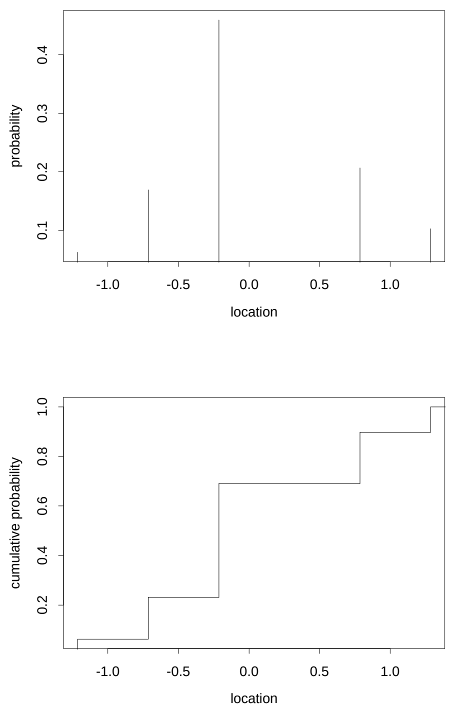

Figure 2.1 *Discrete distribution.*

The multinomial logit parameterization is useful even if the class membership does not depend on covariates since it enforces the restriction that the probabilities sum to one.

2.5.3 Mixed continuous and discrete latent variables

mention finite mixture again

2.6 Summary

We have introduced a general model framework unifying multilevel, structural equation and longitudinal models. The framework accommodates factors and random coefficients at different levels, regression structures among them and a wide range of response processes, and flexible specifications of latent variable distributions.

For further reading a must is multsem paper for advanced reading and GLLAMM book for introduction.

2.7 References

A. R. Adams, “Band-structure engineering for low-threshold high-efficiency semiconductor lasers”, Electron. Lett., Vol. 22, No. 5, pp. 249–250, Feb. 1986.

E. Yablonovitch and E. O. Kane, “Reduction of lasing threshold current density by the lowering of valence band effective mass”, J. Lightwave Technol., Vol. LT-4, No. 5, pp. 504–506, May. 1986.

T. J. Andersson et al.: Appl. Phys. Lett., 51 (1987) 752.

M. Fukuda, M. Okayasu, J. Temmyo, and J. Nakano, “Degradation behavior of 0.98mm strained quantum well InGaAs/AlGaAs lasers under high-power operation”, IEEE J Quantum Electron., Vol. 30, pp. 471–476, 1994.

M. Asada, A. R. Adams, K. E. Stubkjaer, Y. Suematsu, Y. Itaya and S. Arai, “The temperature dependence of the threshold current of GaInAsP/InP DH lasers”, IEEE J Quantum Electron., Vol. QE-17, No. 5, pp. 611–619, May 1981.

A. Kasukawa, Y. Imajo, I. J. Murgatroyd and H. Okamoto, “Effect of well number in 1.3μm GaInAsP graded-index separate-confinement-heterostructure multiple quantum well (GRIN-SCH-MQW) laser diodes”, Trans. IEICE, E 73, pp. 59–62, 1990.

W. T. Tsang, “Extremely low threshold (AlGa)As modified multiquantum well heterostructure lasers grown by molecular-beam epitaxy”, Appl. Phys. Lett., vol. 39, No. 10, pp. 786–788, Nov. 1981.

W. T. Tsang, “Extremely low threshold (AlGa)As graded-index waveguide separate- confinement heterostructure lasers grown by molecular-beam epitaxy”, Appl. Phys. Lett., vol. 40, No. 3, pp. 217–219, Feb. 1982.

U. Koren, B. I. Miller, Y. K. Su, T. L. Koch and J. E. Bowers, “Low internal loss separate confinement heterostructure InGaAs/InGaAsP quantum well laser”, Appl. Phys. Lett., Vol. 51, No. 21, pp. 1744–1746, 1987.

P. J. A. Thjis, L. F. Tiemeijer, J. J. M. Binsma, and T. van Dogen, “High-performance 1.5μm wavelength InGaAs-InGaAsP strained quantum well lasers and amplifiers”, IEEE J Quantum Electron., Vo. 27, No. 6, pp. 1426–1439, Jun. 1991.

A. Kasukawa, I. J. Murgatroyd, Y. Imajo, T. Namegaya, H. Okamoto and S. Kashiwa, “1.5μm GaInAs/GaInAsP graded index separate confinement heterostructure multiple quantum well (GRIN-SCH-MQW) laser diodes grown by metalorganic chemical vapor deposition (MOCVD)”, Electron. Lett., 25, pp. 659–660, 1989.

T. Tanbun-Ek, R. a. Logan, N. A. Olsson, H. Temkin, A. M. Sergent and K. W. Wecht, “High power output 1.48-1.51μm continuously graded index separate confinement strained quantum well lasers”, Appl. Phys. Lett., Vol. 57, No. 16, pp. 224–226, Jul. 1990.

H. Kakibayashi, F. Nagata, “Composition analysis of GaxAl1-xAs superstructure by equal thickness fringes”, Pro. XI Int. Cong. on Electron Microscopy, Kyoto, Japan, pp. 1495–1496, 1986.

H. Asano, S. Takano, M. Kawaradani, M. Kitamura, and I. Mito, “1.48mm high-power InGaAs/InGaAsP MQW LD’s for Er-doped fiber amplifiers”, IEEE, Phton. Technol. Lett., Vol. 3, No. 5, pp. 415–417, 1991.

H. Hirayama, Y. Miyake and M. Asada,“Analysis of current injection efficiency of separate-confinement-heterostructure quantum-film lasers”, IEEE J. Quantum Electron., Vol. 28, No. 1, pp. 68–74, Jan. 1992.

P. Brosson, C. Rabourie, L. Le Gouezigou, J. L. Lievin, J. Jacqest, F. Leblond, A. Olivier and D. Leclerc, “Experimental determination of carrier-induced differential loss in 2-section GaInAsP/InP laser-waveguide”, Electron. Lett., Vol. 25, No. 24, pp. 1623–1624, Nov. 1989.

A. Kasukawa, T. Namegaya, N. Iwai, N. Yamanaka, Y. Ikegami, and N. Tsukiji, “Extremely high power 1.48mm GaInAsP/InP GRIN-SCH strained MQW lasers", IEEE, Phton. Technol. Lett., Vol. 6, No. 1, pp. 4–6, 1994.

D. Fekete, K. Chan, J. M. Ballantyne, and L. F. Eastman, "Graded-index separate-confinement InGaAs/GaAs strained-layer quantum well laser grown by metalorganic chemical vapor deposition", Appl. Phys. Lett., Vol. 49, No. 24, pp. 1659–1660, 1986.

M. Okayasu, M. Fukuda, T. Takeshita, and S. Uehara, "Stable operation (over 5000 h) of high-power 0.98mm InGaAs-GaAs strained quantum well ridge waveguide lasers for pumping Er3+-doped fiber amplifiers", IEEE Photon. Technol. Lett., Vol. 2, No. 10, pp. 689–691, 1990.

S. Ishikawa, K. Fukagai, H. Chida, T. Miyazaki, H. Fujii, and K. Endo, "0.98-1.02mm strained InGaAs/AlGaAs double quantum-well high-power lasers with GaInP buried waveguide", IEEE J. Quantum Electron., Vol. 29, No. 6, pp. 1936–1942, 1993.

H. Meier: Technical Digest of 20th European Conference on Optical Communications, Firenze, Sep., 1994, p. 974.

J. S. Major, W. E. Plano, D. F. Welch, and D. Scifres, "Single-mode InGaAs-GaAs laser diodes operating at 980nm", Electron. Lett., Vol. 27, No. 6, pp. 539–541, 1991.

M. Ohkubo, T. Ijichi, A. Iketani, and T. Kikuta, "980nm aluminum-free InGaAs/InGaAsP/InGaP GRIN-SCH SL-SQW lasers", IEEE J Quantum Electron., Vol. 30, pp. 408–414, 1994.

H. Horie, N. Arai, Y. Mitsuishi, N. Komuro, H. Kaneda, H. GotohCM. Usami, and Y. Matsushima, "Greater than 500-mW CW kink-free single transverse-mode operation of weakly index guided buried-stripe type 980nm laser diodes", IEEE, Phton. Technol. Lett., Vol. 12, No. 10, pp. 1304–1306, 2000.