Modeling Strategies - Marginal Models, Random Effects Models, and Transition Models

Outline

- Summary of Modeling Strategies
  - Modeling Mean Responses
  - Modeling Covariances
  - Interpretation
- Continuous Response
- Binary Response
- Count Data

Summary of Modeling Strategies

Modeling Mean Responses

- Marginal models:
  - The marginal expectation of the response, $\mathrm{E}(Y_{ij}) = \mu_{ij}$, depends on the covariates, $\boldsymbol{x}_{ij}$, through a link function g,

$$\mathrm{E}(Y_{ij}) = \mu_{ij} = g^{-1}(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}).$$

- Random effects (conditional) models:
  - Conditional on subject specific, unobserved random variables $\boldsymbol{b}_i$,

$$\mathrm{E}(Y_{ij} \mid \boldsymbol{b}_i) = \mu_{ij}^* = g^{-1}(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}^* + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i).$$

- Transition models:

  - Let $\mathcal{H}_{ij} = (Y_{i1}, \ldots, Y_{ij-1})$ denote the history of Y_{ij},

$$\mathrm{E}(Y_{ij} \mid \mathcal{H}_{ij}) = \mu_{ij}^{**} = g^{-1}\left\{ \boldsymbol{x}_{ij}^T \boldsymbol{\beta}^{**} + \sum_{r=1}^{s} f_r(\mathcal{H}_{ij}, \boldsymbol{\alpha}) \right\}.$$

Modeling Covariances

- Marginal models:
  - The marginal variance of the response depends on the marginal mean,

$$\text{Var}(Y_{ij}) = \phi V(\mu_{ij}),$$

    where V is a known function and the scale parameter ϕ may also depend on some covariates.
  - The correlation between Y_{ij} and Y_{ik} is a function of the marginal mean:

$$\text{Cor}(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}, \boldsymbol{\alpha}),$$

    where ρ is a known function and the correlation parameters $\boldsymbol{\alpha}$ may depend on covariates.

- Random effects (conditional) models:
  - Typically, conditioning on the random effects, $\boldsymbol{b}_i$, the responses $Y_{i1}, \ldots, Y_{in_i}$ are *independent* with an *exponential family* distribution.
  - The random effects $\boldsymbol{b}_i$ have mean $\mathbf{0}$ and variance D.
  - Typically the random effects are assumed to be multivariate Gaussian.
  - Correlation among observations from the same person arises from their sharing unobservable variables, i.e., random effects.

- Transition models:
  - Typically, $Y_{ij} \mid \mathcal{H}_{ij}$ is assumed to have an *exponential family* distribution with variances

$$\text{Var}(Y_{ij} \mid \mathcal{H}_{ij}) = \phi V(\mu_{ij}^{**}).$$

  - Correlation among $Y_{i1}, \ldots, Y_{in_i}$ exists because the past response values explicitly influence the present observation.

Interpretation

- Marginal model: β is the mean (expected) difference in the response for the two populations (individuals) with the identical covariate values but differ in X by one unit.
- Conditional model: β^* is the expected difference in the response for two individuals with the identical covariate values AND identical random effects but differ in X by one unit.
  - It is somewhat artificial to have two individuals with the same random effects but interpretation of between-subject covariate effects (gender, treatment, etc.) is difficult otherwise.
  - For within-subject covariate effects (age, time), it is perhaps more natural to interpret it as the expected difference for the *same* individual when its X changes by one unit while other covariates (including random effects) being held as the same.
  - Alternatively, we can interpret β^* conditional on the random effects being 0, that is, the effect of the variable X on an average individual.

  - Gaussian case: $\beta = \beta^*$
  - Poisson case (log link): $\beta = \beta^*$ for non-intercept predictor and when only the random intercept is present.
  - Binary case (logistic link): $\beta \neq \beta^*$. β^* is usually larger than β because it is less noisy when using its own observation as control.

- Transitional models: β^{**} is the effect of X while other covariates and the history are held equal.
  - β^{**} is potentially misleading to study relationship between Y and X.

Choices of Models

- The conditional and transition models can, in principle, be "marginalized" by integrating out the random effects or the history.
- In general, the regression parameters $\boldsymbol{\beta}$, $\boldsymbol{\beta}^*$ and $\boldsymbol{\beta}^{**}$ have different interpretations and not directly comparable.
- The choice of model should depend on the scientific question of interest.

Continuous Response

In the linear case, it is possible to formulate the three regression approaches to have coefficients with the same interpretations. That is, coefficients from random effects and transition models can have marginal interpretations as well.

- Consider the linear regression model for the orthodontic measurement data,

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij},$$

  where y_{ij} is distance and x_{ij} is age of child i at time t_{ij}.

- The marginal modelling approach is to assume

$$\begin{aligned} \mathrm{E}(Y_{ij}) &= \beta_0 + \beta_1 x_{ij}, \\ \mathrm{Corr}(\epsilon_{ij}, \epsilon_{ik}) &= \rho(t_{ij}, t_{ik}, \boldsymbol{\alpha}) \end{aligned}$$

- The linear random effects model is

$$Y_{ij} = \beta_0^* + U_{i0} + (\beta_1^* + U_{i1}) x_{ij} + Z_{ij}$$

  – The regression coefficients also have a marginal interpretation since $\mathrm{E}(Y_{ij}) = \beta_0^* + \beta_1^* x_{ij}$.

- A transition model can have the form

$$Y_{ij} = \beta_0^{**} + \beta_1^{**} x_{ij} + \epsilon_{ij} \quad (1)$$

$$\epsilon_{ij} = \alpha \epsilon_{ij-1} + Z_{ij}, \quad (2)$$

where Z_{ij} are independent mean-zero with variance σ^2, and the process is initiated by $\epsilon_{i0} \sim N\{0, \sigma^2/(1-\alpha^2)\}$.

  - This model can be re-expressed as

$$Y_{ij} = \beta_0^{**} + \beta_1^{**} x_{ij} + \alpha(Y_{ij-1} - \beta_0^{**} - \beta_1^{**} x_{ij-1}) + Z_{ij} \quad (3)$$

  - Equations (1) and (2) imply that $\mathrm{E}(Y_{ij}) = \beta_0^{**} + \beta_1^{**} x_{ij}$, so that this form of linear transition model has coefficients which also have a marginal interpretation.
  - Note that this model is equivalent to a marginal model with an exponential autocorrelation function.

- Econometricians use a different formulation of transition model

$$Y_{ij} = \boldsymbol{x}_{ij}^T \boldsymbol{\beta}^\dagger + \alpha Y_{ij-1} + Z_{ij}. \tag{4}$$

  - The response is regressed on the previous outcome itself without adjusting for its expectation, like (3) does.
  - Equation (4) implies that $\mathrm{E}(Y_{ij}) = \sum_{r=0} \alpha^r \boldsymbol{x}_{ij-r}^T \boldsymbol{\beta}^\dagger$, so that $\boldsymbol{\beta}^\dagger$ does not have a marginal interpretation.
  - The interpretation of $\boldsymbol{\beta}^\dagger$ depends on the assumed form of the autocorrelation model. For example, if we add additional lagged values of Y's in Equation (4), both the correlation and the interpretation of $\boldsymbol{\beta}^\dagger$ will change.

Binary Response

Marginal Model

Consider a single covariate X and a binary response Y, the marginal model is given by:

$$\begin{aligned}
\Pr(Y_{ij} = 1) &= \mathrm{E}(Y_{ij} \mid x_{ij}) = \mu_{ij} \\
\mathrm{logit}(\mu_{ij}) &= \log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \beta_0 + \beta_1 x_{ij} \qquad \text{(log odds)} \\
\mathrm{Var}(Y_{ij}) &= \mu_{ij}(1-\mu_{ij}) \\
\mathrm{Cor}(Y_{ij}, Y_{ik}) &= \rho \qquad \text{for example}
\end{aligned}$$

ICHS example: For the jth visit of the ith individual

$$x_{ij} = \begin{cases} 1 & \text{Vitamin A deficient} \\ 0 & \text{otherwise} \end{cases}$$

$$y_{ij} = \begin{cases} 1 & \text{respiratory infection} \\ 0 & \text{otherwise} \end{cases}$$

- Odds of being infected ($y_{ij} = 1$) given $x_{ij} = 0$ is e^{β_0}
- Odds of being infected ($y_{ij} = 1$) given $x_{ij} = 1$ is $e^{\beta_0+\beta_1}$
- Odds ratio (Vitamin A deficient vs Vitamin A replete) is e^{β_1}.
- β_0 and β_1 are population-average parameters.
- Interpretations of β_0 and β_1 remain the same regardless of n_i and the magnitude of within-cluster dependence.

Random Effects Model

For binary data, a random intercept model can be written as

$$\text{logit}\,\Pr(Y_{ij} = 1 \,|\, u_i) = \beta_0^* + u_i + \beta_1^* x_{ij}$$

$$u_i \overset{\text{iid}}{\sim} \mathcal{N}\left(0, v^2\right)$$

ICHS example

- Probability of infection for child i when vitamin A replete ($x_{ij} = 0$) is

$$\Pr(Y_{ij} = 1 \,|\, u_i, x_{ij} = 0) = \frac{e^{\beta_0^* + u_i}}{1 + e^{\beta_0^* + u_i}},$$

  with corresponding odds of infection:

$$e^{\beta_0^* + u_i}$$

- $e^{\beta_0^*}$ is the odds of respiratory infection for a typical child with $u_i = 0$.

- If that child becomes vitamin A deficient ($x_{ij} = 1$), the odds of respiratory infection is

$$e^{\beta_0^* + u_i + \beta_1^*},$$

giving an odds ratio of infection (vitamin A deficient vs vitamin A replete)

$$e^{\beta_1^*}.$$

- $e^{\beta_1^*}$ is the odds ratio of respiratory infection when a child is vitamin A deficient relative to when that same child is not vitamin A deficient.

- How are (β_0^*, β_1^*) and (β_0, β_1) related?

The marginal probability of infection is

$$\begin{aligned}\Pr(Y_{ij}=1\,|\,x_{ij}) &= \int_{-\infty}^{\infty}\Pr(Y_{ij}=1\,|\,u_i,x_{ij})f(u_i;\upsilon^2)du_i \\ &= \int_{-\infty}^{\infty}\frac{e^{\beta_0^*+u_i+\beta_1^*x_{ij}}}{1+e^{\beta_0^*+u_i+\beta_1^*x_{ij}}}\frac{e^{-\frac{u_i^2}{2\upsilon^2}}}{\sqrt{2\pi\upsilon^2}}du_i \\ &= \frac{e^{\beta_0+\beta_1x_{ij}}}{1+e^{\beta_0+\beta_1x_{ij}}} \quad \text{under marginal model}\end{aligned}$$

The relationship between (β_0^*,β_1^*) and (β_0,β_1) is complicated. Some theoretical results (for binary data with logit link):

- (Neuhaus et al., 1991) If $\text{Var}(u_i)=\upsilon^2>0$,

$$|\beta_k|\le|\beta_k^*|, \qquad k=1,\ldots,p$$

  where the equality holds only when $\beta^*=0$. Moreover, the discrepancy between β and β^* increases with υ^2.

- (Zeger et al., 1988) If $u_i \sim \mathcal{N}\left(0, \upsilon^2\right)$,

$$\beta \approx \frac{1}{\sqrt{0.346\upsilon^2 + 1}}\beta^*.$$

As υ^2 increases, β is closer to zero.

A simulation study (using the ICHS data's notation)

Assume $\beta_0^* = -2.0$, $\beta_1^* = 0.4$ and $\text{Var}(u_i) = v^2 = 2.0$.

The random effects logistic model logit $\Pr(Y_{ij} = 1|U_i) = \beta_0^* + U_i + \beta_1^* x_{ij}$.

Value of u_i	$\Pr(1 \mid u_i, 0)$	$\Pr(1 \mid u_i, 1)$	odds ratio
u_i at $2.5\% = -2\sqrt{2}$	0.008	0.013	1.49
u_i at $50\% = 0$	0.12	0.17	1.49
u_i at $97.5\% = 2\sqrt{2}$	0.68	0.76	1.49

Marginal model:

$$\Pr(y_{ij} = 1 \,|\, x_{ij} = 0) = \int_{-\infty}^{\infty} \frac{e^{-2+u_i}}{1+e^{-2+u_i}} \frac{e^{-u_i^2/4}}{\sqrt{4\pi}} \mathrm{d}u_i = 0.18$$
$$\Pr(y_{ij} = 1 \,|\, x_{ij} = 1) = \int_{-\infty}^{\infty} \frac{e^{-2+u_i+0.4}}{1+e^{-2+u_i+0.4}} \frac{e^{-u_i^2/4}}{\sqrt{4\pi}} \mathrm{d}u_i = 0.23$$

The odds ratio for vitamin A deficiency is

$$e^{\beta_1} = \frac{0.23/(1-0.23)}{0.18/(1-0.18)} \approx 1.36.$$

So $\beta_1 = \log(1.36) = 0.31$.

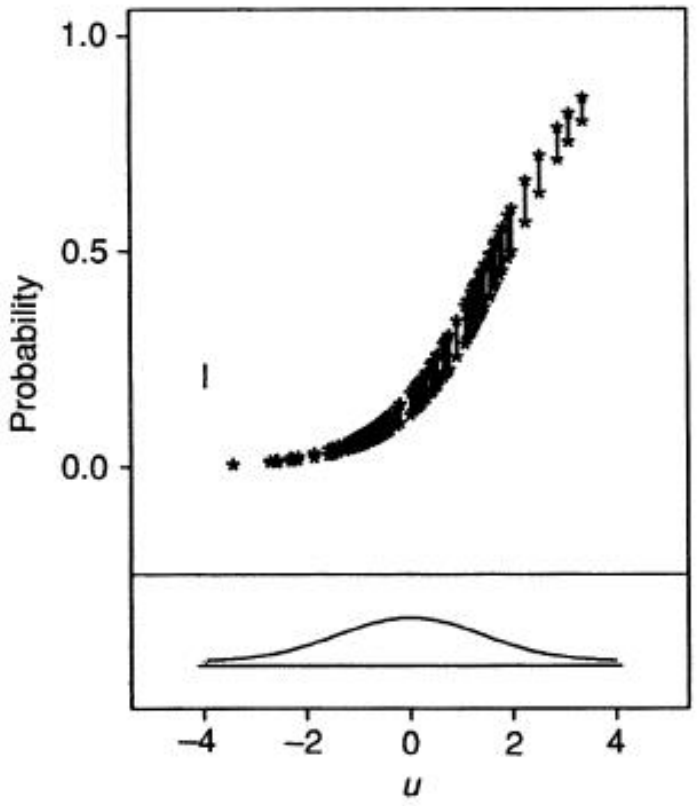

Fig. 7.2. Simulation of risk of infection with and without dichotomous risk factor, for 100 children in a logistic model with random intercept. Population-average risks with and without infection are indicated by the vertical line on the left. The Gaussian density for the random intercepts is shown.

- Note that the odds ratio for vitamin A deficiency is the same for every child, equal to 1.49.
- Children with a lower propensity for infection ($U_i < 0$) have a smaller change in absolute risk than those with propensities near the center ($U_i = 0$).
- The vertical line on the left of Fig 7.2 shows the change in the population rates with and without vitamin A.
- The marginal and random effects model parameters differ in the logistic model, i.e. $\beta_1 \neq \beta_1^*$.
- Actually, $\beta_1 < \beta_1^*$ in this case.
- This attenuation of parameter values is illustrated in Fig. 7.3, which displays $\Pr(Y_{ij} = 1|U_i)$ as a function of x_{ij} for a number of randomly generated U_i's.

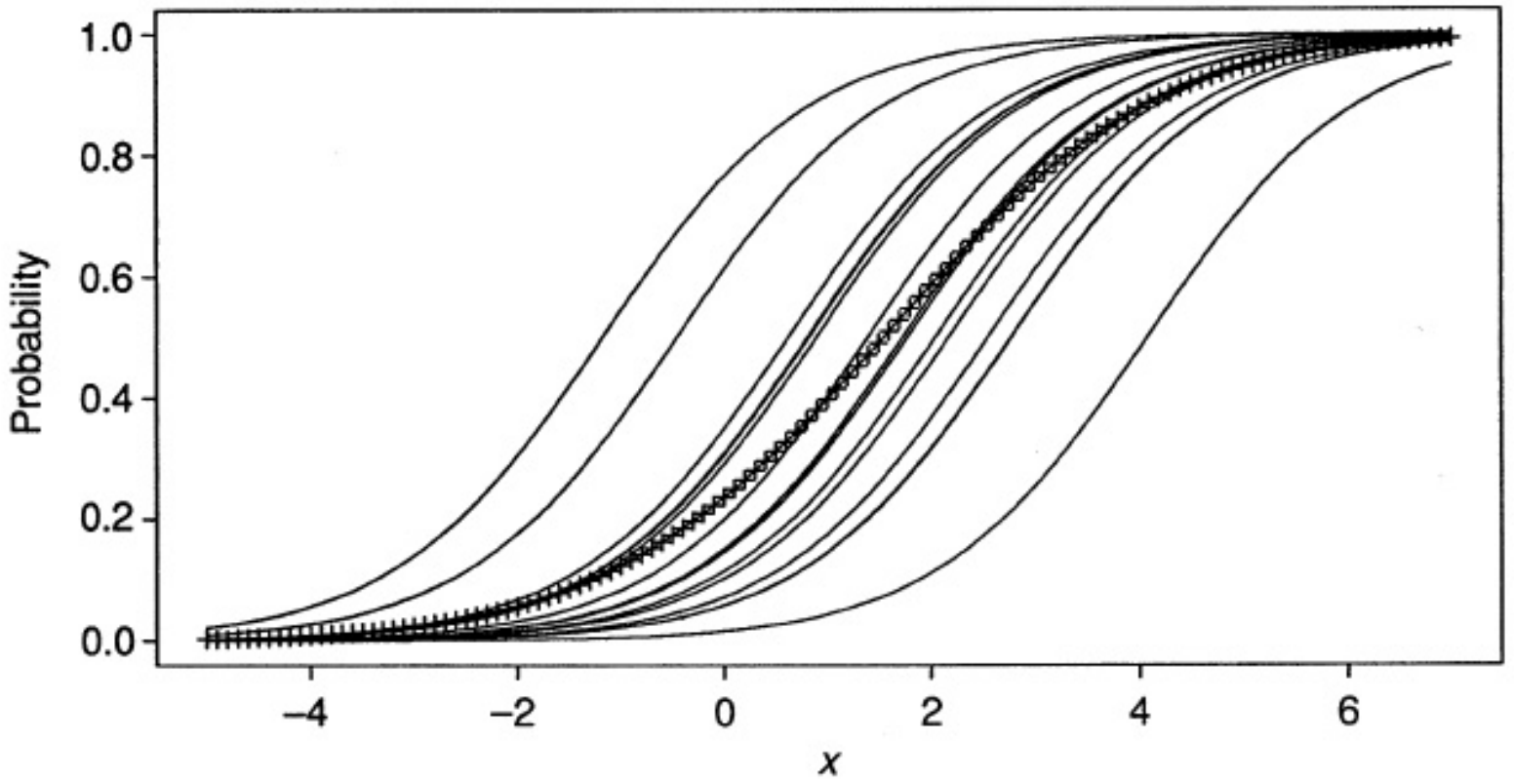

Fig. 7.3. Simulation of the probability of a positive response in a logistic model with random intercept and a continuous covariate. ——: sample curves; ++++++: average curve.

Transition Models

If Y_{ij} depends only on the last observation (Markov model, discrete autoregressive model with lapse 1)

$$\text{logit}\,\Pr(y_{ij} = 1 \,|\, x_{ij}, H_{ij}) = x_{ij}\beta^{**} + \alpha y_{i,j-1}$$

- $\exp(\beta^{**})$ is the odds ratio for vitamin A deficient vs. replete among children who were free of infection at last visit (or same status of infection at last visit)
- $\exp(\alpha)$ is the odds ratio for children who did and did not have infection at last visit.

The Markov chain has transition probability matrix:

	$y_{ij} =$	
	0	1
$y_{i,j-1} =$ 0	$\frac{1}{1+\exp(x_{ij}\beta^{**})}$	$1 - \frac{1}{1+\exp(x_{ij}\beta^{**})}$
$y_{i,j-1} =$ 1	$\frac{1}{1+\exp(x_{ij}\beta^{**}+\alpha)}$	$1 - \frac{1}{1+\exp(x_{ij}\beta^{**}+\alpha)}$

- Transition probability matrices vary with the covariate and therefore can vary across individuals.
- The initial distribution (for stationary chain, or otherwise) can be difficult to specify.
- The transition model parameters differ from either the random effects or marginal models. The relationship between marginal and transition parameters can be established only in limited cases. See Zeger and Liang (1992) for further discussion.

Count Data

In log-linear models for count data, random effects and marginal parameters can be equivalent in some special cases.

Consider the random intercept only model

$$\log E(Y_{ij}|U_i) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}^* + U_i.$$

$$E(Y_{ij}) =$$

$$=$$

$$=$$

$$=$$

- If we fit a marginal model assuming $E(Y_{ij}) = \exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta})$, all the coefficients except the intercept will have the same value and interpretation as in the random effects model.

Further Reading

- Chapter 7 of DHLZ.
- Neuhaus JM, Kalbfleisch JD, Hauck WW (1991). A comparison of cluster-specific and population averaged approaches for analyzing correlated binary data. *International Statistical Review*, **59**, 25-36.
- Zeger SL, Liang K-Y, Albert PS (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, **44**, 1049-60.
- Zeger SL and Liang K-Y (1992). An overview of methods for the analysis of longitudinal data. *Statistics in Medicine*, **11**, 1825-39.