Random effects GLM

- $Y_{ij} \mid (x_{ij}, u_i)'s \overset{indep.}{\sim} f_Y(y_{ij} \mid x_{ij}, u_i, \beta, \phi), j = 1, \ldots, m_i.$

- $U_i \overset{iid}{\sim} f_U(u \mid G)$

Problem 1. Maximum likelihood estimation:

Assume further that $u_i \sim N_q(0, G)$.

Let $\delta = (\beta, G)$. The likelihood function of δ is

$$L(\delta \mid y) = \prod_{i=1}^{n} \int \left(\prod_{j=1}^{m_i} f\left(y_{ij} \mid x_{ij}, u_i\right) \right) f_U\left(u_i \mid G\right) du_i$$

$$\propto \prod_{i=1}^{n} \int \left(\prod_{j=1}^{m_i} \exp\left(\beta^T x_{ij} y_{ij} + u_i^T d_{ij} y_{ij} - \psi\left(\theta_{ij}\right)\right) \right) |G|^{-\frac{1}{2}} \exp\left(\frac{-u_i^T G^{-1} u_i}{2} \right) du_i.$$

The score function for β, based on the complete data (y, U), is

$$S_\beta(\delta \mid \underset{\sim}{y}, U) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} x_{ij} \left(y_{ij} - \mu_{ij}\left(u_i\right)\right) = 0,$$

where

$$\mu_{ij}\left(u_i\right) = h^{-1}\left(x_{ij}^{T} \beta + d_{ij}^{T} u_i\right).$$

Show that the score function for G is

$$S_G(\delta \mid \underset{\sim}{y}, U) = \frac{1}{2} \left[G^{-1} \left(\sum_{i=1}^{n} u_i u_i^T \right) G^{-1} - nG^{-1} \right].$$

The observed score functions are then defined to be

$$\begin{cases} S_\beta(\delta \mid \underset{\sim}{y}) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} x_{ij} \left(y_{ij} - E\left[\mu_{ij}\left(u_i\right) \mid y_{ij}\right]\right) \\ S_G(\delta \mid \underset{\sim}{y}) = \frac{1}{2} \left[G^{-1} \left(\sum_{i=1}^{n} E\left[u_i u_i^\tau \mid y_{ij}\right]\right) G^{-1} - nG^{-1} \right] \end{cases}.$$

$$L(\delta \mid y, U) = \prod_{i=1}^{n} \left(\prod_{j=1}^{m_i} f\left(y_{ij} \mid x_{ij}, u_i\right) \right) f_U\left(u_i \mid G\right)$$

$$\propto \prod_{i=1}^{n} \left(\prod_{j=1}^{m_i} \exp\left(\beta^T x_{ij} y_{ij} + u_i^T d_{ij} y_{ij} - \psi\left(\theta_{ij}\right)\right) \right) |G|^{-\frac{1}{2}} \exp\left(\frac{-u_i^T G^{-1} u_i}{2} \right).$$

$$l(\delta \mid y, U) = \ln L(\delta \mid y, U)$$

$$\propto \sum_{i=1}^{n} \left\{ \left(\sum_{j=1}^{m_i} \exp\left(\beta^T x_{ij} y_{ij} + u_i^T d_{ij} y_{ij} - \psi\left(\theta_{ij}\right)\right) \right) - \frac{1}{2} \ln\left(|G|\right) - \left(\frac{u_i^T G^{-1} u_i}{2} \right) \right\}$$

$$S_G(\delta \mid y, U) = \frac{\partial l(\delta \mid y, U)}{\partial G} = \frac{\partial}{\partial G} \sum_{i=1}^{n} \left\{ -\frac{1}{2} \ln\left(|G|\right) - \left(\frac{u_i^T G^{-1} u_i}{2} \right) \right\}$$

$$\overset{*}{=} \sum_{i=1}^{n} \left\{ -\frac{1}{2} G^{-1} + \frac{1}{2} u_i^T G^{-1} G^{-1} u_i \right\}$$

$$= \frac{1}{2} \left[G^{-1} \left(\sum_{i=1}^{n} u_i u_i^T \right) G^{-1} - n G^{-1} \right].$$

It can be shown that $\frac{\partial}{\partial G} \ln(|G|) = (G^{-1})^T = G^{-1}$.

Problem 2. Logistic regression for binary responses

Consider $\log it P \left(Y_{ij} = 1 \mid U_i \right) = \beta_0 + U_i + X_{ij}^T \beta$: random intercept logistic model.

Let $r_i = \beta_0 + U_i$. The joint likelihood function for β and r_i is

$$L \left(\beta, r_1, \cdots, r_n \right) = \prod_{i=1}^{n} \exp \left[r_i \sum_{j=1}^{m_i} y_{ij} + \left(\sum_{j=1}^{m_i} y_{ij} x_{ij}^T \right) \beta - \sum_{j=1}^{m_i} \ln \left(1 + \exp \left(r_i + x_{ij}^T \beta \right) \right) \right].$$

The conditional likelihood, which is equivalent to that derived in stratified case-control studies, is

$$L_c(\beta) = \prod_{i=1}^{n} \left[\frac{\exp \left(\sum_{j=1}^{m_i} y_{ij} x_{ij}^T \beta \right)}{\sum_{\left\{ \underset{\sim}{y_i} : \sum_{i=1}^{m_i} y_{it} = t_{i2} \right\}} \exp \left(\sum_{l=1}^{m_i} y_{il} x_{il}^T \beta \right)} \right].$$

....

Example:

2×2 cross-over trial

Group \ Outcome	$(1,1)$	$(1,0)$	$(0,1)$	$(0,0)$
placebo-treatment $(0,1)$	a_1	b_1	c_1	d_1
treatment-placebo $(1,0)$	a_2	b_2	c_2	d_2

....

Show that

$$L_c(\beta) = \left(\frac{\exp \left(\beta_1 \right)}{1 + \exp \left(\beta_1 \right)} \right)^{b_2 + c_1} \left(\frac{1}{1 + \exp \left(\beta_1 \right)} \right)^{b_1 + c_2}.$$

$\log it P \left(Y_{ij} = 1 \mid U_i \right) = \beta_0 + U_i + X_{ij} \beta_1,$

$m_i = 2, \ x_{ij} = \begin{cases} 0 & \text{placebo} \\ 1 & \text{treatment} \end{cases}, \ y_{ij} = \begin{cases} 0 & \text{ineffective} \\ 1 & \text{effective} \end{cases}$

For $d_1 + d_2$	$\sum_{i=1}^{2} y_{it} = t_{i2} = 0$	$\sum_{j=1}^{2} y_{ij} x_{ij} = 0$
For $b_1 + b_2 + c_1 + c_2$	$\sum_{i=1}^{2} y_{it} = t_{i2} = 1$	$\sum_{j=1}^{2} y_{ij} x_{ij} = 0$ for $b_1 + c_2$ $\sum_{j=1}^{2} y_{ij} x_{ij} = 1$ for $b_2 + c_1$
For $a_1 + a_2$	$\sum_{i=1}^{2} y_{it} = t_{i2} = 2$	$\sum_{j=1}^{2} y_{ij} x_{ij} = 1$

$$L_c(\beta) = \prod_{i=1}^{n} \left[\frac{\exp\left(\sum_{j=1}^{m_i} y_{ij} x_{ij}^T \beta\right)}{\sum_{\left\{\underset{\sim}{y_i}: \sum_{i=1}^{m_i} y_{it} = t_{i2}\right\}} \exp\left(\sum_{l=1}^{m_i} y_{il} x_{il}^T \beta\right)} \right]$$

$$= \prod_{i=1}^{n} \left[\frac{\exp\left(\sum_{j=1}^{2} y_{ij} x_{ij} \beta_1\right)}{\sum_{\left\{\underset{\sim}{y_i}: \sum_{i=1}^{2} y_{it} = t_{i2}\right\}} \exp\left(\sum_{l=1}^{2} y_{il} x_{il} \beta_1\right)} \right]$$

$$= \frac{\exp(\beta_1)^{a_1+a_2+b_2+c_1}}{\{\exp(\beta_1)^{a_1+a_2}\}\{[1+\exp(\beta_1)]^{b_1+b_2+c_1+c_2}\}\{[1]^{d_1+d_2}\}}$$

$$= \frac{\exp(\beta_1)^{b_2+c_1}}{[1+\exp(\beta_1)]^{b_1+b_2+c_1+c_2}}$$

$$= \left(\frac{\exp(\beta_1)}{1+\exp(\beta_1)}\right)^{b_2+c_1} \left(\frac{1}{1+\exp(\beta_1)}\right)^{b_1+c_2} .$$

Problem 3. Poisson-Gamma mixture

Example:

- $Y_{ij}'s \mid u_i \sim \text{Poisson}(u_i)$.

- $u_i \sim \text{Gamma}(\alpha, \beta)$, where $\alpha\beta \triangleq \mu$ and $\alpha\beta^2 \triangleq \phi\mu^2$.

Show that Y_{ij} is Negative-binomial with $E\left[Y_{ij}\right] = \mu$ and $V\left[Y_{ij}\right] = \mu + \phi\mu^2$.

$$p(y) = \int_0^{\infty} p(y \mid u)p(u)\, du$$

$$= \int_0^{\infty} \frac{u^y e^{-u}}{y!} \left[\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\right] u^{\alpha-1} e^{-u/\beta}\, du$$

$$= \frac{1}{y!\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} u^{y+\alpha-1} e^{-\frac{1+\beta}{\beta}u}\, du$$

$$\overset{*}{=} \frac{1}{y!\Gamma(\alpha)\beta^{\alpha}} \frac{\Gamma(y+\alpha)}{\left[\frac{1+\beta}{\beta}\right]^{y+\alpha}}$$

$$= \frac{\Gamma(y+\alpha)}{y!\Gamma(\alpha)} \frac{\beta^y}{(1+\beta)^{y+\alpha}}$$

$$= \frac{(y+\alpha-1)!}{y!(\alpha-1)!} \left[\frac{\beta}{1+\beta}\right]^y \left[\frac{1}{1+\beta}\right]^{\alpha} .$$

Thus,

$$y \sim NB(\alpha, \frac{1}{1+\beta}),$$

and

$$\mathbb{E}[Y] = \frac{\alpha\frac{\beta}{1+\beta}}{\frac{1}{1+\beta}} = \alpha\beta = \mu,$$

$$\mathbf{Var}[Y] = \frac{\alpha\frac{\beta}{1+\beta}}{\left[\frac{1}{1+\beta}\right]^2} = \alpha\beta(1+\beta) = \alpha\beta + \alpha\beta^2 = \mu + \phi\mu^2.$$

Integration of expressions of the type $f(t)e^{g(t)}$

[a] The primary reason for the gamma function's usefulness in such contexts is the prevalence of expressions of the type $f(t)e^{g(t)}$ which describe processes that decay exponentially in time or space. Integrals of such expressions can occasionally be solved in terms of the gamma function when no elementary solution exists. For example, if f is a power function and g is a linear function, a simple change of variables $u : a \cdot t$ gives the evaluation

$$\int_0^\infty t^b e^{-at}\, dt = \frac{1}{a^b} \int_0^\infty u^b e^{-u}\, d\left(\frac{u}{a}\right) = \frac{\Gamma(b+1)}{a^{b+1}}.$$

....

$$\int_0^\infty u^{y+\alpha-1} e^{-\left[\frac{1+\beta}{\beta}\right]u}\, du = \frac{\Gamma(y+\alpha)}{\left[\frac{1+\beta}{\beta}\right]^{y+\alpha}}.$$

---

[a] *Gamma Function*. Wikipedia. Oct. 20, 2023. URL: `https://en.wikipedia.org/w/index.php?title=Gamma_function&oldid=1181093976#Integration_problems` (visited on 11/04/2023).

Paper[1] Summary

- **Research Problem**: Traditional cross-sectional analysis has its limitation. This paper addresses the challenge of analyzing longitudinal data on disease markers in AIDS, which involves problems of (left and right) censoring and (left) truncation [2].

- **Methodology**: The paper proposes a likelihood-based approach that models the joint distribution of the disease markers ($Z(t)$, such as T4 count and T4/T8 ratio [3]), the time of infection, and the time to AIDS using parametric models. The paper also handles the censoring and truncation problems using standard survival analysis techniques. The study considered different patterns of censoring, such as left censoring, right censoring, and interval censoring, and derived different likelihood formulas for each pattern. For truncation, the study divided the individual likelihood term by the probability of observing the disease occurring after the beginning of the ascertainment period B. The study provides a method for predicting the time to AIDS given a series of disease marker measurements.

- **Data and Application**: The paper applies the proposed method to the Toronto AIDS cohort study, which consists of 159 cases who were infected with HIV or seroconverted during the 5.5 years' follow-up. The paper analyzes the longitudinal data on T4 counts and T4/T8 ratio as disease markers and compares three models with different assumptions about the infection time.

- **Results and Conclusion**: The paper finds that there is a significant association between the rate of decline of T4 counts or T4/T8 ratio and the time to AIDS, and that the scaled incubation time may be a more natural scale for viewing the progression of HIV infection. The paper also shows that the disease marker information improves the prediction of the time to AIDS. The paper concludes that the proposed method is useful for understanding the natural history of AIDS and developing treatment strategies.

---

[1] Yudi Pawitan and Steve Self. "Modeling Disease Marker Processes in AIDS". in: *Journal of the American Statistical Association* 88.423 (Sept. 1993), pp. 719–726.

[2] Censoring and truncation are two types of incomplete data problems that occur in survival analysis. In this study, there are

- Left censoring: This happens when the infection time is unknown but known to be before a certain time. For example, if a person was seropositive at entry to the study, then his infection time is left censored by the entry time. Left censoring can also be interval censored, meaning that the infection time is known to be within a certain interval, such as between the first and last contact with the primary case.

- Right censoring: This happens when the time to AIDS is unknown but known to be after a certain time. For example, if a person did not develop AIDS by the end of follow-up, then his time to AIDS is right censored by the last follow-up time. Right censoring can also be informative, meaning that the censoring time is related to the event of interest. Right censoring due to death may be informative.

- Left truncation: This happens when the cohort members are selected based on a certain criterion that depends on the survival time. The dataset does not include individuals who developed the disease prior to the beginning of this study.

[3] The T4 count refers to the number of CD4 T cells, which are crucial for immune function. The T4/T8 ratio is the ratio of CD4 cells (T helper cells) to CD8 cells (regulatory T cells).