- Fit a log-linear model $Y_{\mathbf{x}r} \sim \mathbf{x}_c + r + r{:}X$, where

Poisson

$Y_{\mathbf{x}r}$: $(y_{11}, \ldots, y_{1J}, y_{21}, \ldots, y_{2J}, \ldots, y_{I1}, \ldots, y_{IJ})^T$

$\mathbf{x}_c$: a nominal factor which treats each covariate class as a level (i.e., $\mathbf{x}_c$ has I levels)

even though some covariate classes have same value, still treated as different levels

LNp.6-2

$h_1(x), h_2(x), \ldots, h_{p-1}(x)$

$r{:}X$: the interaction term with the same X (without the intercept term) in the multinomial logit model

$X\beta = \beta_0 + \beta_1 h_1(x) + \beta_2 h_2(r) + \beta_3 h_1(x)\cdot h_2(r) = (\beta_0 + \beta_2 h_2(r)) + (\beta_1 + \beta_3 h_2(r))\cdot h_1(x)$

- The parameters associated with $\mathbf{x}_c$ will represent the total for the cases ($\Leftrightarrow n_1, n_2, \ldots, n_I$) marginal row totals
- the parameters associated with r will represent the category totals ($\Leftrightarrow$ intercepts in multinomial logit model) marginal column totals β_{0j}
- The interaction terms $r{:}X$ tell us how the probability of falling in the different categories changes with X ($\Leftrightarrow$ coefficients of X in the multinomial logit model) $(\beta_{1j}, \ldots, \beta_{p-1,j})$
- This multinomial log-linear model would have same deviance as its corresponding multinomial logit model

- **Q**: what information in the log-linear model not valid? main effects of n_c

answer to the question on the bottom LNp6-2

➢ Multinomial logit can be viewed as a GLM-type model, which allows us to apply all the methodology developed for GLM

- **Q**: why can multinomial log-linear model be connected to multinomial logit model?

Suppose that

fixed but now treated as random

$$n_{\mathbf{x}} \overset{\text{indep}}{\sim} \text{Poisson}(\lambda_{\mathbf{x}})$$

average # of subjects observed at x. ⇒ false information when n_x's are fixed.

$$(y_{\mathbf{x}1}, \ldots, y_{\mathbf{x}J} | n_{\mathbf{x}}) \overset{\text{indep}}{\sim} \text{multinomial}(n_{\mathbf{x}}, p_{\mathbf{x}1}, \ldots, p_{\mathbf{x}J})$$

Then, $\quad y_{\mathbf{x}j} \overset{\text{indep}}{\sim} \text{Poisson}(\mu_{\mathbf{x}j} = \lambda_{\mathbf{x}} p_{\mathbf{x}j})$

r ⇒ r, x are predictors

If we fit a log-linear model for $y_{\mathbf{x}j}$, then:

LNp.6-2

$$\log\left(\frac{\mu_{\mathbf{x}j}}{\mu_{\mathbf{x}1}}\right) = \log\left(\frac{\lambda_{\mathbf{x}} p_{\mathbf{x}j}}{\lambda_{\mathbf{x}} p_{\mathbf{x}1}}\right) = \log\left(\frac{p_{\mathbf{x}j}}{p_{\mathbf{x}1}}\right) = \eta_{\mathbf{x}j} = X\boldsymbol{\beta}_j$$

$$\Rightarrow \log(\mu_{\mathbf{x}j}) = \log(\mu_{\mathbf{x}1}) + X\boldsymbol{\beta}_j$$

invalid information

$$= \log(\lambda_{\mathbf{x}} p_{\mathbf{x}1}) + X\boldsymbol{\beta}_j$$

indicator function $\sum_{j=2}^{J} I_{\{j\}}(r)\cdot\beta_{p-1,j}$

main effects of X main effects of Y

$$\beta_{0j} + h_1(\mathbf{x})\beta_{1j} + \cdots + h_{p-1}(\mathbf{x})\beta_{p-1,j}$$

interaction between r & X

❖ **Reading**: Faraway (2006, 1st ed.), 5.1

Ordinal Multinomial Responses

- Suppose that
  - We have a multinomial response with J ordered categories, and (order1 < order2 <…< orderJ)
  - For individual $\mathbf{x}$, observe ordinal response $(y_{\mathbf{x}1}, \ldots, y_{\mathbf{x}J})$ (same likelihood as nominal response)
- For an ordered multinomial response, it is often easier to work with the cumulative probabilities

$$\underset{\text{cdf}}{\gamma_{\mathbf{x}j}} = \underset{\text{pmf}}{p_{\mathbf{x}1}} + \cdots + p_{\mathbf{x}j},\ j = 1, \ldots, J$$

  - $p_{\mathbf{x}1}=\gamma_{\mathbf{x}1}$; $p_{\mathbf{x}j}=\gamma_{\mathbf{x},j}-\gamma_{\mathbf{x},j-1}$, $j=2,\ldots, J-1$; $\gamma_{\mathbf{x}J}=1$
  - **Q**: why cumulative probabilities are better?
    - the cumulative probabilities are increasing and invariant to combining adjacent categories
    - $\gamma_{\mathbf{x}J}=1$, so we need only model $J-1$ probability

$p_{\mathbf{x}j}$ 1 0 1 2 … J $\gamma_{\mathbf{x}j}$ 1 0 1 2 … J

Q: why not use $\gamma_{\mathbf{x}j}$ for nominal response? no unique $\gamma_{\mathbf{x}j}$ in the case of nominal response

- **Q**: how to link $\boldsymbol{\gamma}_{\mathbf{x}} = (\gamma_{\mathbf{x}1}, \gamma_{\mathbf{x}2}, \ldots, \gamma_{\mathbf{x},J-1})$ with $\mathbf{x}$?

$g: [0,1] \to (-\infty, \infty)$

$$\begin{array}{ccc} (\gamma_{\mathbf{x}1}, & \cdots, & \gamma_{\mathbf{x},J-1}) \\ g \downarrow & & g\downarrow \\ \eta_{\mathbf{x}1} = X\boldsymbol{\beta}_1 & \cdots & \eta_{\mathbf{x},J-1} = X\boldsymbol{\beta}_{J-1} \end{array}$$

same (X) different ($\boldsymbol{\beta}_j$)

cf. $X\boldsymbol{\beta}_j$ in LN p.6-2

# of parameters in $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{J-1}$: $(J-1)+(p-1)$

$$\begin{aligned} X\boldsymbol{\beta}_j \;=\;& \beta_{0j} - \\ & h_1(\mathbf{x})\beta_{1\cancel{j}} - \cdots - \\ & h_{p-1}(\mathbf{x})\beta_{p-1\cancel{j}} \end{aligned}$$

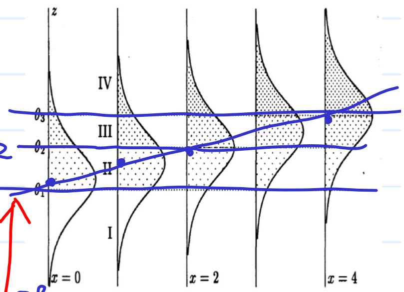
- some choices of g: (1) logit (2) probit (3) complementary log-log
  - Notice that the intercepts, β_{0j}, are *different* for the J categories
  - $\boldsymbol{\beta}=(\beta_1, \ldots, \beta_{p-1})$ *do not* depend on $j \Rightarrow$ we assume that the predictors have a uniform effect on the probabilities of response categories in a sense that we will shortly make clear
  - Negative sign before $\boldsymbol{\beta}$
- Latent variable approach for ordinal variables:
  - let $z_{\mathbf{x}}$ be some unobserved continuous variable that might be thought of as the real underlying latent response
  - we only observe a discretized version of $z_{\mathbf{x}}$ in the form of $y_{\mathbf{x}}$ where $y_{\mathbf{x}}=j$ is observed if $\beta_{0,j-1}<z_{\mathbf{x}} \le \beta_{0,j}$
  - suppose that $z_{\mathbf{x}}-\boldsymbol{\beta}^T\mathbf{x}^*$ has a distribution F:

Use multinomial model to study the relationship between $y_{\mathbf{x}}$ & $\mathbf{x}$ (ordinal)

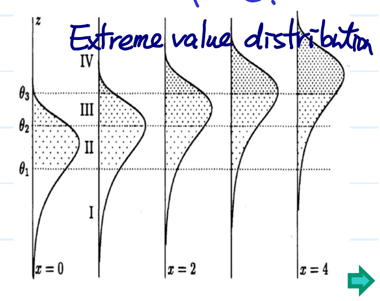
$y_{\mathbf{x}}$: a function of $z_{\mathbf{x}}$, contain less information than $z_{\mathbf{x}}$

Recall tolerance dist. in LN p.3-16~17

β_{03} β_{02} $\beta_{0,1}$



If $z_{\mathbf{x}}$ observed, can fit a linear model $z_{\mathbf{x}} = \mathbf{x}\boldsymbol{\beta} + \varepsilon$.



$$\begin{aligned} \gamma_{\mathbf{x}j} \leftarrow P(y_{\mathbf{x}} \le j) &= P(z_{\mathbf{x}} \le \beta_{0j}) \\ &= P(z_{\mathbf{x}} - \boldsymbol{\beta}^T\mathbf{x}^* \le \beta_{0j} - \boldsymbol{\beta}^T\mathbf{x}^*) \\ z_{\mathbf{x},j} \leftarrow &= F(\beta_{0j} - \boldsymbol{\beta}^T\mathbf{x}^*) \end{aligned}$$

$\mathbf{x}^* = (h_1(\mathbf{x}), h_2(\mathbf{x}), \ldots, h_{p-1}(\mathbf{x}))$

Why not depends on j

- If F follows the logistic distribution, i.e., $F(z)=e^z/(1+e^z)$, then

symmetric about 0, i.e., $F(-z)=1-F(z)$, pdf $f(-z)=f(z)$

$\gamma_{x,j}=e^{\eta_{x,j}}/1+e^{\eta_{x,j}}$

$\Leftrightarrow \eta_{x,j}=\log\left(\frac{\gamma_{x,j}}{1-\gamma_{x,j}}\right)$

$\Leftrightarrow g$: logit → on cdf, not pmf

$$\gamma_{\mathbf{x}j}=\frac{\exp(\beta_{0j}-\boldsymbol{\beta}^T\mathbf{x}^*)}{1+\exp(\underbrace{\beta_{0j}-\boldsymbol{\beta}^T\mathbf{x}^*}_{\eta_{x,j}})}$$

  - so, we would have a logit model for the cumulative probabilities $\gamma_{\mathbf{x}j}$

g: symmetric about 0 ← Φ $\gamma_{xj}=\Phi(\eta_{x,j})$

- Choosing Normal for latent variable (i.e, F) leads to probit model
- F=Extreme value distribution leads to complementary log-log
- Notice that if $\boldsymbol{\beta}>0$, as $\underbrace{\mathbf{x}\text{ increases}}_{\to \eta_{x,j}\downarrow}$, $P(y_{\mathrm{x}}>j)$ ($1-\gamma_{x,j}$) will also increase

  $\Rightarrow$ this explain the use of the minus sign in the definition of the model because it allows for the more intuitive interpretation of the sign of $\boldsymbol{\beta}$

- *Proportional odds* model (F: logistic distribution)
  - Let $\gamma_{\mathbf{x}j}=\gamma_j(\mathbf{x})=P(y_{\mathrm{x}}\le j|\mathbf{x})$, then the proportional odds model, which use the *logit* link, is:

odds for multinomial in LNp.6-2 (cf.) cumulative odds for ordinal multinomial

$$\log\left(\frac{\gamma_j(\mathbf{x})}{1-\gamma_j(\mathbf{x})}\right)=\boxed{\beta_{0j}-\boldsymbol{\beta}^T\mathbf{x}^*}\ //\ \eta_{x,j}$$

- It is so called because the relative odds ($\gamma_{x,j}/1-\gamma_{x,j}$) for the event $y_{\mathrm{x}}\le j$ comparing $\mathbf{x}=\mathbf{x}_1$ and $\mathbf{x}=\mathbf{x}_2$ are:

say, $x_1^*=x_2^*+1$

$$\left(\frac{\gamma_j(\mathbf{x}_1)}{1-\gamma_j(\mathbf{x}_1)}\right)\Big/\left(\frac{\gamma_j(\mathbf{x}_2)}{1-\gamma_j(\mathbf{x}_2)}\right)=\exp\left(-\boldsymbol{\beta}^T(\mathbf{x}_1^*-\mathbf{x}_2^*)\right)$$

irrelevant to j.

because we use the information of order

  - This does not depend on j
  - Of course, this assumption of proportional odds does need to be checked for a given dataset
  - It can be checked by computing the sample (observed) odds proportions with respect to $\mathbf{x}$ (check an example in lab)
- Some advantages
  - The model use fewer parameters than the multinomial logit model (**Q**: the former nested in the latter?)

    how many fewer? Compare $x\beta_j$ in LNp.6-2 & 6-6: $(p-1)\cdot(J-1-1)=(p-1)(J-2)$ — for nominal response

  - Typically, the output from the proportional odds model is easier to interpret

    interpretation of β_{0j}: the odds of γ_{0j} at $x^*=0$ is $\exp(\beta_{0j})$

- Inferences follows the usual GLM likelihood-based approach
- Interpretation of the fitted coefficients: odds of moving from one category to another increase by a factor of $\exp(\boldsymbol{\beta})$ as $\mathbf{x}$ increases by one unit (check lab)

- *ordered probit* model (F is normal)
  - If the latent variable $z_{\mathbf{x}}$ has a standard normal distribution, then

on cdf, not pmf

$$\Phi^{-1}(\gamma_j(\mathbf{x})) = \beta_{0j} - \boldsymbol{\beta}^T\mathbf{x}^*$$

  - Coefficients estimates could be very different from proportional odds model, but predicted values usually are very similar $p_{x,1}, p_{x,J} \geq 0.2$

$\Phi(z) \propto e^z/1+e^z$, $0.2 \leq \Phi(z) \leq 0.8$

- *proportional hazards* model (F = extreme value dist.)
  - Concept of hazard $\boxed{\text{pdf} \Leftrightarrow \text{survival function} \Leftrightarrow \text{cdf} \Leftrightarrow \text{hazard}}$

proportion hazard: $h_x(t) = h_0(t) \cdot \exp(-\beta^T x^*)$ $\Leftrightarrow$ $S_x(t) = S_0(t)^{\exp(-\beta^T x^*)}$ ← survival function at x

Consider the case z_x (LNp.6-7): lifetime & F: a lifetime distribution

    - Developed in insurance application: when issuing a life insurance policy, the insurer is interested in the prob. that the person will die during the term of the policy given that they are alive now ← $P(\text{die at } 55 \mid \text{age} \geq 55)$

$$\frac{S_x(t_2)}{S_x(t_1)} = \left[\frac{S_0(t_2)}{S_0(t_1)}\right]^{\exp(-\beta^T x^*)}$$

    - This is not the same as the unconditional probability of death ← $P(\text{die at } 55)$
  - Suppose we use the complementary log-log link, i.e.,

$$\log(1-\gamma_{x,j}) = -\exp(\beta_{0j} - \beta^T x^*) \Leftarrow \log(-\log(1-\gamma_j(\mathbf{x}))) = \beta_{0j} - \boldsymbol{\beta}^T\mathbf{x}^*$$

Then, the hazard of category j is the probability of falling in category j given that your category is greater than j: