Brief Summary

(a)

1)

Section 2 Summary

* Use of Proportional odds model *

x: indicator variable / predictor

k: Total k ordered categories of the responses

$\pi_1(x), \pi_2(x_2), \ldots, \pi_k(x)$ are the response probabilities respectively.

$\boxed{\gamma \leq j}$

$$\gamma_j(x) = \kappa_j \exp(-\beta^T x)$$

$\therefore$ Taking ratio of corresponding odds ratio

$$\kappa_j(x_1) / \gamma_j(x_2) = \exp\{\beta^T(x_2 - x_1)\}$$

$\therefore$ identical to the linear logistic model.

$$\therefore \boxed{\log[\gamma_j(x) / \{1 - \gamma_j(x)\}] = \theta_j - \beta^T x)}$$

NOTE: This is For Binary data [Two response categories].

$\therefore$ As the Example given in the paper demonstrate the Application of above by taking 3 ordered categories.

∴ Uses data related to the tonsil size and the presence of Streptococcus pyogenes. [with 2 sample problems]

Results : odds of having greatly enlarged tonsils are 1.8 times greater for carriers of Streptococcus pyogenes than non-carriers. and,

The odds of normal sized tonsils are 1.8 times greater for non carriers than for carriers.

∴ Paper also discussed a generalized empirical logit transform for parameter estimation and model verification.

∴ For equal probability to all categories here $1/3$.

$$1/3 - 1/3 \sum_{j=1}^{k} \pi_j^3$$

Using one of the equivalent forms.

* The quality Z_i with weights given by.

$$\therefore \boxed{w_j \propto R_j (n - R_j)(n_j + n_{j+1})}$$

Generalized Empirical Logit transform for the i^{th} group.

$w_j \Rightarrow w_1 = 0.638$ with standard error 0.225

$\hat{\Delta} = 0.603 \pm 0.225$

(Δ) parameter indicates the tonsil size is larger.

* Use of proportional Hazards model

* Section 3 Summary

Here $\lambda(t;x)$ is the Hazard function represents instantaneous risk for an event happening at time t, defined by the equation:

$$\lambda(t;x) = \lambda_0(t)\exp(-\beta^T x),$$

To able to view data more accurately for discrete data, model is transformed into complementary log-log form.

$\gamma_j(x)$ represent the complementary probability of being in category j given the covariate values x:

$$\boxed{\log[-\log\{1-\gamma_j(x)\}] = \theta_j - \beta^T x}$$

From the family distribution example. The logit difference, having same sign is the indication of strict stochastic ordering between the two distributions.

In the example, difference between complementary log-logs for 1960 and 1970 are relatively constant with median value 0.49.

Therefore the $p_1(x)$ is the proportion of the population in the North east earning more than \$$x$ in 1960.

→ This example also emphasizes the power of quantitative, parametric approach in explaining systematic components in the data, even when non-parametric tests might suggest model rejection.

* Properties of Related linear Models *

→ The proportional odds and the proportional hazards models have the same general form.

$$\boxed{\text{link}\{\gamma_j(x)\} = \theta_j - \beta^T x.}$$

logit / complementary log-log function.

→ $\boxed{1)}$ key property: $\boxed{\text{strict stochastic ordering}}$

which means if two groups of subpopulations having different covariates x_1 & x_2, the categories' probabilities for x_1 will either be greater or less than those for x_2.

→ $\boxed{2)}$ General property → To All log-linear models is that they don't use "scores".

* Reversibility and Invariance

→ where the categories of the response can be rearranged without affecting the model's fit.

→ But this property is generally inappropriate for the ordinal data. Instead the concept of palindromic invariance is introduced, where considers reversals of category but not arbitrary permutations.

* Similar Rank Tests *

→ The Efficient Score is introduced as a test statistic that provides maximum local power. For the linear logistic model, the efficient score involves weighted cross-products of covariates with the average rank for the response category.

* More Complex Covariate Structure *

→ To figure out we use different pieces of information ~~are~~ called "explanatory variables".

→ As the Age increases, the frequency of dreams also stated that it decreases. Although the transformation values $\{z_{ij}\}$ together with simple variance estimate $(0.297 n_i)^{-1}$ are not strictly monotone decreasing with Age, the relationship is nevertheless very marked. To check, maximum likelihood is also ~~p~~ performed

$$\therefore \log\{\gamma_{ij}/(1-\gamma_{ij})\} = \theta_j - \beta x_i$$

$$\therefore \log\{\gamma_{ij}/(1-\gamma_{ij})\} = \theta_j - \alpha_i,$$

* The only conclusion drawn is that, to a close approximation, the odds, for disturbed or severely disturbed dreams among boys decreases by a factor of 0.80 per year from 5 to 15 years. But, there is inconclusive evidence to suggest that this decrease may not be uniform over the 10-year period. Qualitatively similar conclusions would be obtained by an analysis on the probit or other suitable scale.

* Parameter Estimation in the General model

→ non-linear models.

In context of GLM, which describes strict stochastic ordering among responses.

- Therefore, if introduces a multiplicative model to relax the Assumption of Constant Variance or Scale parameters along an underlying continuum.

$\tau_1 = 1$ or $\sum \log \tau_i = 0$

$$\text{link}(\gamma_{ij}) = (\theta_j - \beta^T x_i)/\tau_i,$$

location scale

∴ $\tau \to$ unknown parameters to be estimated.

→ This model is applies to quality of vision data and finds location and scale differences, demonstrating that the ratio of logits is nearly constant.

* Maximum likelihood Estimation.

→ Though linear systematic structure in the model and presence of multinomial variation. Iteratively reweighted least square algorithm for parameter estimation. where,

The likelihood function is expressed in terms of probabilities and cumulative transformations.

→ In most of the cases, it's easy to solve. But in some cases, it can be tricky. we might mix them up, and may be put them at wrong places, "identifiability" of the model. This Identifiability problem which applies to linear models generally and is intimately related to the Rank of the design matrix.

→ The concavity of the likelihood function.

* Alternative models and the scope of Possible inferences *

→ Alternative structural equations to general linear model for describing strict stochastic order in data. One popular alternative is the log-linear model, which Assigns scores to partition interaction statistics into two components

1) strict stochastic ordering
2) representing higher order interactions

→ This scope of inferences based on this model is limited to the actual categories present in the data, and it lacks invariance under grouping of adjacent response categories.

→ In Contrast models like proportional odds or proportional hazards allow for broader inferences and often provide more interpretable parameters.

log-linear models	Logit linear models
Symmetric	Assymmetric

→ Their Conceptual approaches are distinct, but have mathematical similarities.

* An Analysis of Residuals *

→ Why Analysis of Residuals!!!

⇒ They are used to evaluate the goodness of fit and identify any patterns or discrepancies between the model's predictions and the actual observed ordinal responses. The analysis of residuals for multinomial response models can be challenging due to several factors, including the limited number of possible values.

→ These residuals are positive and provide a measure of how well the model explains the observed data.

→ Different patterns of cell residuals can indicate different inadequacies in the model, helping identify where model may need improvement.

⇒ Likelihood and odd ratio for ordinal regression.

→ It represents the probability of observing the given set of ordinal responses based on the predictor variable.

→ This model are based on the cumulative probabilities of observing values in or above a specific ordinal category.

→ The likelihood quantifies how well the ordinal regression model fits the observed data.

⇒ MLE finds the parameter values that maximize the likelihood of observing the given data under this regression model.

* $\boxed{\text{Odd Ratio in Ordinal Regression}}$ *

odds ratio's measures the effect of predictor variables on the odds of an observation falling into & above a particular ordinal categories relative to a reference category.

→ For three-category ordinal responses, the odds ratio indicates how the odds of being in a higher category change.

⇒ likelihood quantifies the goodness of fit of the model to the observed data, while odds ratios provide a measure of the effect of predictor values / variables on the odds of belonging to a specific ordinal category.

-> How's it different from multi-class classification.

=> In terms of their objectives, Assumption, outputs.

=> They differ in terms of nature of Dependent Variable.

=> In multiclass classification, the dependent Variable is categorical, But the categories are not inherently ordered. Each observation is assigned to one of several classes or categories, and there is no inherent Ranking or order among these classes.

=> In terms of modeling objective.

* Ordinal Regression : model the relationship between the ordinal response Variable and predictor Variables while considering nature of the categories. It aims to predict the ordinal category or rank of an observation.

* multiclass classification : The Primary objective of multiclass classification is to Assign observations to one of several classes or categories. It focuses on Accurately classifying each observation into a specific class, and there is no inherent consideration of ordinality or Rank.

→ $\boxed{\text{In terms of output,}}$ In multiclass.

: It provides class labels or probabilities for each class to indicate the likelihood of belonging to each class.

⇒ In ordinal regression, describes how these variables influence the odds of being in a higher ordinal category relative to a reference category. It may also provide odds ratios & predicted probabilities for each category.

* $\boxed{\text{Model Selection:}}$

For $\boxed{\text{Ordinal Regression}}$: ordinal regression models (proportional odds, probit) are specifically designed to handle ordinal response variables. Various model selection criteria and tests are available to assess model fit.

For $\boxed{\text{multiclass classification}}$ (logistic regression, decision trees, neural networks) are more general and can be applied to both categorical and continuous response variables with multiple classes.