The proportional-odds model

The proportional odds model is a class of generalized linear models used for modelling the dependence of an ordinal response on discrete or continuous covariates. Let Y denote the response category in the range $1, \ldots, k$, with $k \geq 2$, and let $\gamma_j = \mathrm{pr}(Y \leq j \mid x)$ be the cumulative response probability when the covariate is held at x. The most general form of linear logistic model for the jth cumulative response probability,

$$\mathrm{logit}(\gamma_j) = \alpha_j - \beta_j^T x, \tag{1}$$

is one in which both the intercept α and the regression coefficient β depend on the category j. The proportional-odds model is a linear logistic model in which the intercepts depend on j, but the slopes are all equal. Thus we arrive at the model

$$\mathrm{logit}(\gamma_j) = \alpha_j - \beta^T x, \tag{2}$$

asserting that the graph of the $k-1$ cumulative logits against x is a series of parallel lines or planes with intercepts $\alpha_1, \ldots, \alpha_{k-1}$.

Ordinal response variables are common in a number of areas, notably survey research, food testing, industrial quality assurance, radiology and clinical research. In a study of disease severity, for example, the degree of impairment might be described by one of a small collection of labels such as 'none,' 'slight,' 'moderate,' 'severe,' and 'incapactiating.' One of the most effective ways to construct a model for an ordinal response such as this is to invoke the concept of a latent, or unobserved, response Z. The actual recorded response Y is envisaged as a crude manifestation of the latent variable in such a way that the relationship is monotone:

$$\alpha_{j-1} < Z \leq \alpha_j \Longleftrightarrow Y = j. \tag{3}$$

The 'cut-points' α_j are envisaged as unknown points on the latent scale. In the example described, the z-interval $(-\infty, \alpha_1]$ is interpreted as no impairment; the interval $(\alpha_1, \alpha_2]$ as slight impairment, and so on. Unless the latent variables are close to one of the boundaries, similar values of the latent variable are not distinguished and give rise to identical responses.

This description of the model seems to require the observer to have a precisely measured latent variable Z available, if only to himself or herself, and to make the comparison (3) before reporting Y. Like all mathematical models of behaviour, this is an idealization of what actually occurs, and is not to be taken literally, particularly at the edges. In fact, however, the model does not make these extreme demands on the observer. Although the model is capable of this mechanastic interpretation, it is not a necessary interpretation. What is important is not so much the mechanism but the prediction. If the model predictions are sufficiently close to observations and known limiting behaviour, all is well.

The dependence of the latent variable on the covariates may be specified by means of a linear or non-linear model, as appropriate. In the case of a linear model, we have $Z = \beta^T x + \epsilon$, where ϵ is a random variable with cumulative distribution function F. Then the probability $\mathrm{pr}(Z \leq z)$ is $F(z - \beta^T x)$. Relationship (3) between the latent variable and the response gives the implied model for Y in the form

$$\gamma_j = \mathrm{pr}(Y \leq j) = \mathrm{pr}(Z \leq \alpha_j) = F(\alpha_j - \beta^T x),$$

or in linearized form

$$F^{-1}(\gamma_j) = \alpha_j - \beta^T x.$$

If $F(z) = e^z/(1+e^z)$, implying that ϵ has the logistic density, this scheme produces the proportional-odds model (2) illustrated in Fig. 1. Other choices for F produce generalized linear models of the same type. The cumulative probit model arises if ϵ is normal, and the grouped proportional hazards model, or complementary log-log model, arises if ϵ has the extreme-value distribution, in other words, if $\exp(\epsilon)$ has the exponential or Weibull distribution. This derivation explains the unorthodox choice of sign for the regression coefficients in (1) and (2).

Various extensions of this scheme are possible. Suppose, for example, that the covariates affect

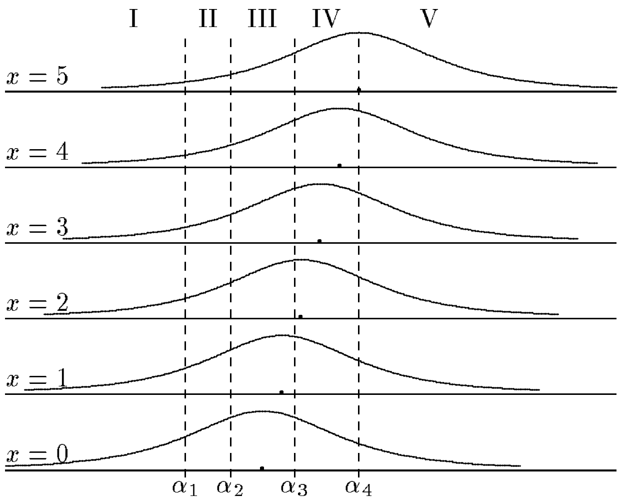

Figure 1. *Diagram illustrating how the distribution of the latent variable Z changes with x in the proportional-odds model. The horizontal axis represents the latent variable, and the recorded categories are denoted by roman numerals attached to the five contiguous Z-intervals. Over the range of x-values shown, the probability for category* IV *is almost constant. By contrast, the probabilities for categories* I *and* V *vary by factors of 3–4 over the same range.*

both the location and scale of the latent variable according to the model

$$Z = \beta^T x + \exp(\tau^T x)\epsilon.$$

Models incorporating dispersion effects of this type are used in industrial quality assurance in order to detect factors whose effect is primarily on the variability of the product. The implied model for Y is then

$$F^{-1}(\gamma_j) = (\alpha_j - \beta^T x)/\exp(\tau^T x), \tag{4}$$

which is no longer linearizable. Models (1), (2) and (4) have the limiting property for extreme covariate values, that as $\beta^T x \to \pm\infty$, all the probability accumulates in one of the extreme categories.

The class of models derivable in this way using a latent variable all have an important invariance, or closure, property connected with the amalgamation of adjacent response categories. Suppose that model (4) with $k > 2$ response categories is correct. If categories j and $j+1$ are combined into a single new response category, then model (4) still applies, but with k reduced to $k-1$ and with α_j deleted. In general, information is lost when categories are amalgamated, so the maximum-likelihood estimate is affected. In extreme cases, the parameters might not be estimable from the reduced data. The model, however, is invariant, and the same regression parameters apply to the reduced data. By contrast, most of the competing models described at the end of this article are not closed under category amalgamation.

The term 'proportional-odds' stems from the fact that in model (3) the odds of the event $Y \le j$ satisfies

$$\text{odds}(Y \le j \,|\, x) = \exp(\alpha_j - \beta^T x).$$

Consequently, the ratio of the odds of the event $Y \le j$ for x_1 and x_0 is

$$\frac{\text{odds}(Y \le j \,|\, x_1)}{\text{odds}(Y \le j \,|\, x_0)} = \exp\bigl(-\beta^T(x_1 - x_0)\bigr),$$

which is a constant independent of j. If we arrange matters such that $x_0 = 0$ is the baseline value of the covariates, it follows that $\exp(\alpha_j)$ is the baseline odds for the event $Y \le j$. From this point of view, the proportional-odds model simply takes the baseline odds, which can be set arbitrarily, and multiplies by the factor $\exp(-\beta^T x)$ to obtain the response odds at a non-baseline covariate value. Neither of the extended models (1) or (4) is a proportional-odds model in this sense.

By definition, the cumulative respone probabilities are ordered $\gamma_1 \leq \cdots \leq \gamma_{k-1} \leq 1$. The logit transformation is strictly monotone from $(0, 1)$ to the real line. The proportional-odds model (2) must therefore satisfy the constraints $\alpha_1 \leq \cdots \leq \alpha_{k-1}$. This condition is both necessary and sufficient to ensure that the fitted response-category probabilities are non-negative for all values of the covariate and for all values of the regression coefficient β. The same condition is necessary and sufficient for the non-linear model (4).

The analogous condition to ensure non-negative response probabilities in model (1) is much more complicated. Non-negativity requires that

$$\alpha_1 + \beta_1 x \leq \cdots \leq \alpha_{k-1} + \beta_{k-1} x$$

for all values of x in some set, $\mathcal{X}$. At a minimum, $\mathcal{X}$ must include the the observed covariates, but the set could be much larger, particularly if the model is to be used for extrapolation or prediction. Suppose for simplicity that there is a single covariate x taking values in the range $[0, \infty)$. Then, considering the logits at $x = 0$ and $x \to \infty$, we require both the intercepts and slopes to be non-decreasing in j. This condition is necessary and sufficient. Likewise, if the covariate space is bounded, say $\mathcal{X} = [-1, 1]$, a necessary and sufficient condition is that

$$|\Delta\beta| \leq \Delta\alpha,$$

componentwise, where $\Delta\alpha$ is the difference vector with components $\alpha_{j+1} - \alpha_j$ for $j = 1, \ldots, k-2$. If there are p covariates, all in the interval $[-1, 1]$, the necessary and sufficient condition $\sum |\Delta\beta_j| \leq \Delta\alpha$ suggests that most models in class (1) are close to (2) if p is large.

The proportional odds model and related family (4) is only one of several families that are designed to be used for the analysis of ordinal data. The three main competing classes are as follows:

(i) Log-linear model with pre-assigned category scores (Haberman 1974, Goodman 1979);

(ii) Canonical regression models (Anderson 1984, Goodman 1981, Greenland 1994);

(iii) Continuation-ratio models (Cox 1972, Fienberg 1980).

When the categories represent temporally ordered stages of development, such as educational attainment, it is natural to consider the conditional probability of failure at stage j conditional on survival up to stage j. The conditional probability of failure at stage j, or the hazard of stage j, or the attrition rate of stage j, is $\pi_j/(1-\gamma_{j-1})$. The natural linear logistic model, in this context called a continuation-ratio model or discrete-time proportional-hazards model, is

$$\text{logit}\big(\pi_j/(1-\gamma_{j-1})\big) = \log\big(\pi_j/(1-\gamma_j)\big) = \alpha_j - \beta_j x.$$

No order constraints are required on the parameters. However, depending on the context, it may be sensible to assume that $\beta_j = \beta$, or, less commonly, that $\alpha_j = \alpha$.

The adequacy of the proportional odds model can in principle be tested by a likelihood ratio test of model (2) against either (1) or (4). Readily available commercial software, such as SAS PROC LOGISTIC, is available for fitting the proportional odds model. Regrettably, such software is rarely sufficiently flexible to fit alternatives such as (1) or (4), so likelihood-ratio testing may require specially-written computer programs. In the absence of special purpose programs, a feasible alternative for model testing is to compute the residuals, and to examine them for patterns, either by plotting or by visual inspection. However, particularly for ordinal data, the visual appearance of a residual plot can be drastically affected by the definition of residual. Cumulative residuals seem to be more appropriate than cell residuals for many plots (McCullagh and Nelder 1989, section 5.6).

The proportional odds model goes back to early work of Snell (1964), Williams and Grizzle (1972), and Simon (1974). Similar ideas, particularly the notion of a latent variable and its use for modelling an ordinal response, can be found in Karl Pearson's early work. For illustrations and numerical examples of the proportional-odds model, see McCullagh (1980), Agresti (1984), Armstrong (1989), and McCullagh and Nelder (1989, chapter 5).

Cross-references:

generalized linear model

ordinal response

linear logistic model

log-linear model

latent variable

Bayesian inference

proportional hazards model

probit model

scores

continuation-ratio model

complementary-log-log model

References

Agresti, A. (1984) *Analysis of Ordinal Categorical Data.* J. Wiley & Sons, New York.

Armstrong, B.G. and Sloan, M. (1989) Ordinal regression models for epidemiologic data. *Amer. J. Epidemiol.* **129**, 191–204.

Anderson, J.A. (1984) Regression and ordered categorical variables (with discussion). *J. R. Statist. Soc.* B **46**, 1–30.

Cox, D.R. (1972) Regression models and life tables (with discussion). *J. R. Statist. Soc.* B **34**, 187–220.

Fienberg, S.E. (1980) *The analysis of Cross-classified Categorical Data.* MIT Press, Cambridge, MA.

Goodman, L.A. (1979) Simple models for the analysis of association in cross-classifications having ordered categories. *J. Am. Statist. Assoc.* **74**, 537–52.

Goodman, L.A. (1981) Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *J. Am. Statist. Assoc.* **76**, 320–34.

Greenland, S. (1994) Alternative models for ordinal logistic regression. *Statistics in Medicine* **13**, 1665–77.

Haberman, S.J. (1974) Log-linear models for frequency tables with ordered classifications. *Biometrics* **36**, 589–600.

McCullagh, P. (1980) Regression models for ordinal data (with discussion). *J. R. Statist. Soc.* B **42**, 109–42.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models.* Chapman and Hall, London.

Simon, G. (1974) Alternate analyses for the singly ordered contingency table. *J. Am. Statist. Assoc.* **69**, 971–6.

Snell, E.J. (1964) A scaling procedure for ordered categorical data. *Biometrics* **20**, 592–607.

Williams, O.D. and Grizzle, J.E. (1972) Analysis of contingency tables having ordered response categories. *J. Am. Statist. Assoc.* **67**, 55–63.

Peter McCullagh
University of Chicago
Department of Statistics
5734 University Ave.,
Chicago Il 60637
Email: `pmcc@galton.uchicago.edu`