Metrika, Volume 32, 1985, page 65–72.

Some Characterizations of Distributions by Regression Models for Ordinal Response Data

By *J. Engel*, Eindhoven[1])

Summary: Let a random variable X be classified into k classes. By doing so, a new random variable is obtained, measured on ordinal scale. If this variable is a response variable in certain regression models for ordinal response data, the distribution of X is characterized by the models. In this paper, characterizations of the distribution of X by the proportional odds model and the proportional hazards model are given.

1. Introduction

Regression models for discrete data receive increasing attention during the last years, especially when the response variable is measured on ordinal scale. In a survey publication, [*McCullagh*] introduces a regression model generalizing two special cases, which are the proportional odds model and the proportional hazards model.

Now suppose an ordinal response variable is obtained by classification of a random variable X with distribution function $F(x; \Theta)$, where Θ is location or scale parameter; the classification does not depend on Θ. If parameter Θ is considered as being an explanatory variable in one of the regression models mentioned above, it is clear that not every distribution can be used. This is also remarked by [*Agresti*, 1980, 1981], referring to a paper of Fleiss, who showed that a normal distribution with mean Θ does not satisfy the proportional odds model. In this paper it will be shown, what distributions can be used, for each of the two models.

In other words, it will be shown what distributions are being characterized by them: a survey is given in table 1.

[1]) *J. Engel*, N.V. Philips' Gloeilampenfabrieken, Centre for Quantitative Methods, Eindhoven, The Netherlands.
0026-1335/85/020065-72$2.50 © *1985 Physica-Verlag, Vienna.*

regression model	location parameter	scale parameter
proportional odds model	logistic distribution	log logistic distribution
proportional hazards model	extreme value distribution	Weibull distribution

Tab. 1: Characterizations of distributions by proportional models

In section 2, characterizations by the proportional odds model are derived; for the proportional hazards model, this is done in section 3.

2. Characterizations by the Proportional odds Model

Characterizations by the proportional odds model will be given for distributions with distribution function $F(x; \Theta)$, where Θ is location or scale parameter. A short description of the model will be given, and a first characterization is proved in theorem 2.1, where Θ is location parameter; the case of a scale parameter will be treated in theorem 2.3, using a monotone transformation of the random variable X.

Definition

Let the real valued outcome space of a random variable X with distribution function $F(x; \Theta)$, parametrized by Θ, be classified into k ordered classes, the boundaries of which do not depend on Θ. Then a new random variable is created, taking k ordered values; let it be the response in a regression model, where Θ is the explanatory variable. Let $\gamma_j(\Theta)$ be the sum of lower class probabilities up to an including class j, at parameter value Θ. Then the proportional odds model is additive on logit scale and defined by

$$\log \{\gamma_j(\Theta)/(1 - \gamma_j(\Theta))\} = c_j - \psi(\Theta)$$

where ψ is a function of Θ only, to be specified later.

Notice that, because of additivity, for any Θ_1 and Θ_2 the log odds ratio

$$\log \left\{ \frac{\gamma_j(\Theta_1)(1 - \gamma_j(\Theta_2))}{(1 - \gamma_j(\Theta_1))\gamma_j(\Theta_2)} \right\} = \psi(\Theta_2) - \psi(\Theta_1)$$

does not depend on j. So, if a $2 \times k$ contingency table is collapsed into a 2×2 table with probabilities $\gamma_j(\Theta_1)$ and $\gamma_j(\Theta_2)$ in the first column, the odds ratio of that 2×2 table does not depend on j.

Suppose that with respect to the distribution of X, the cutting point x is defined by $F(x; \Theta) = \gamma_j(\Theta)$ then an equivalent notation for the log odds is

$$\log\left\{\frac{F(x;\Theta_1)(1-F(x;\Theta_2))}{(1-F(x;\Theta_1))F(x;\Theta_2)}\right\} = \psi(\Theta_2) - \psi(\Theta_1)$$

which should be true for all x belonging to the support of X. In case Θ is location parameter, i.e. $F(x;\Theta) = F(x-\Theta)$ we have

$$\log\left\{\frac{F(x-\Theta_1)(1-F(x-\Theta_2))}{(1-F(x-\Theta_1))F(x-\Theta_2)}\right\} = \psi(\Theta_2) - \psi(\Theta_1)$$

and by putting $\Theta = \Theta_2 - \Theta_1$ this reduces to

$$\frac{F(x)(1-F(x-\Theta))}{(1-F(x))F(x-\Theta)} = e^{\psi(\Theta_2)-\psi(\Theta_1)} = \phi(\Theta)$$

which defines a function $\phi(\Theta)$ of Θ, if $\psi(\Theta_2) - \psi(\Theta_1)$ is required to depend on Θ_1 and Θ_2 via $\Theta_2 - \Theta_1$ only.

Now the following theorem gives a characterization of distribution function $F(x-\Theta)$.

Theorem 2.1

Let $F(x)$ be a distribution function of a random variable X, having the whole interval $(-\infty, \infty)$ as support, for which

$$\frac{F(x)(1-F(x-\Theta))}{F(x-\Theta)(1-F(x))} = \phi(\Theta) \tag{1}$$

holds for all real x and Θ, where $\phi(\Theta)$ is a non-constant function, not depending on x.

Then

$$F(x) = (1 + e^{-b(x-a)})^{-1},$$

for real a and $b > 0$, being the distribution function of the logistic distribution.

Proof

The proof will consist of two steps.

Step 1

Define $H(x) = F(x)/(1 - F(x))$ then it is easily verified that $H(x)$ is a non-decreasing, positive and finite function of x.

Relation (1) can be written as

$$H(x)/H(x - \Theta) = \phi(\Theta) \text{ for all real } x \text{ and } \Theta \tag{2}$$

$$\text{Putting } x = \Theta \text{ yields } H(\Theta) = H(0)\,\phi(\Theta) \text{ for all real } \Theta \tag{3}$$

and so $\phi(\Theta)$ is a non-decreasing function of Θ.

From the relation $\phi(\Theta_1 + \Theta_2) = \phi(\Theta_1)\,\phi(\Theta_2)$, which results from $\phi(\Theta_1 + \Theta_2) = H(x)/H(x - \Theta_1 - \Theta_2) = H(x)/H(x - \Theta_1)\,H(x - \Theta_1)/H(x - \Theta_1 - \Theta_2) = \phi(\Theta_1)\,\phi(\Theta_2)$ and which is true for all real Θ_1 and Θ_2, it follows that $\phi(\Theta) = e^{b\Theta}$ for $b > 0$ and all real Θ.

To see this, define

$$Q(\Theta) = 1/\phi(\Theta) = H(0)\left(\frac{1 - F(\Theta)}{F(\Theta)}\right).$$

It can be easily shown, that $1 - Q(\Theta)$, for $\Theta \geqslant 0$ is a distribution function on the non-negative real numbers, non degenerate at zero, for which $Q(\Theta_1 + \Theta_2) = Q(\Theta_1)\,Q(\Theta_2)$ holds for all non-negative Θ_1 and Θ_2. Application of the well known lemma characterizing the exponential distribution by the lack of memory property [see e.g. *Galambos/Kotz*, p. 8] results into

$$Q(\Theta) = 1/\phi(\Theta) = e^{-b\Theta} \quad \text{for } b > 0 \text{ and all } \Theta \geqslant 0.$$

Then, as $1 = \phi(0) = \phi(\Theta)\,\phi(-\Theta)$ for real Θ,

$$\phi(\Theta) = e^{b\Theta} \text{ for } b > 0 \text{ and all real } \Theta \tag{4}$$

which is the result of step 1 of the proof.

Step 2

Using the result of step 1, we may combine (3) and (4) to write $H(\Theta) = H(0)\,e^{b\Theta} = e^{b(\Theta - a)}$ say, for all real Θ, defining a as $a = 1/b \log H(0)$.

Finally,

$$F(x) = H(x)/(H(x) + 1) = \frac{e^{b(x-a)}}{1 + e^{b(x-a)}} = (1 + e^{-b(x-a)})^{-1}$$

which proves theorem 2.1.

Remark

In [*Galambos/Kotz*, p. 27], a characterization theorem is proved, which shows much ressemblance with theorem 2.1; it is stated in theorem 2.2.

Theorem 2.2

Let distribution function $F(x)$ of a random variable X be continuous and symmetric about the origin. Then

$$F(x) = (1 + e^{-bx})^{-1}, \; b > 0$$

if and only if

$$\frac{F(x)\,(1 - F(x + \Theta))}{F(x + \Theta)\,(1 - F(x))} = \frac{1 - F(\Theta)}{F(\Theta)} \quad \text{for all } x, \; \Theta \geqslant 0. \tag{6}$$

Proof

[See *Galambos/Kotz*].

Comparing this result with theorem 2.1 shows, that in the latter a characterization is given, using a more general function $\phi(\Theta)$, while *Galambos/Kotz* use $(1 - F(\Theta))/F(\Theta)$, and no requirements are made with respect to continuity and symmetry of F.

On the other hand, equation (6) needs to be satisfied only for non-negative x and Θ, and not for all real values as is the case in theorem 2.1. Apparently, the characterization has a practical value in building models for discrete data, which was questioned by the authors.

The scale-variant of theorem 2.1 is stated and proved in theorem 2.3; the location parameter Θ is substituted by scale parameter λ.

Theorem 2.3

Let $G(y)$ be a distribution function of a positive random variable Y, having the whole interval $(0, \infty)$ as support, for which

$$\frac{G(y)\,(1 - G(\lambda y))}{(1 - G(y))\,G(\lambda y)} = \chi(\lambda) \tag{7}$$

holds, for all $y, \lambda > 0$, where $\chi(\lambda)$ is a non-constant function not depending on y.

Then $G(y) = (1 + c\,y^{-b})^{-1}$ for $b, c > 0$ being the distribution function of the so-called log-logistic distribution [see *Johnson/Kotz*, p. 17].

Proof

Defining X as $X = \log Y$, with distribution function $F(x)$, then

$$F(x) = P(X \leqslant x) = P(\log Y \leqslant x) = P(Y \leqslant e^x) = G(e^x)$$

so

$$\frac{F(x)\,(1 - F(x + \log\lambda))}{(1 - F(x))\,F(x + \log\lambda)} = \frac{G(e^x)\,(1 - G(\lambda e^x))}{(1 - G(e^x))\,G(\lambda e^x)} = \chi(\lambda)$$

the last step resulting from (7), for all positive λ, $\chi(\lambda)$ not depending on x.

Using theorem 2.1, the above relation characterizes $F(x)$ as being

$$F(x) = (1 + e^{-b(x-a)})^{-1} \text{ for certain } a \text{ and } b > 0,$$

so

$$G(y) = F(\log y) = (1 + e^{ab}\, y^{-b})^{-1} = (1 + c\, y^{-b})^{-1}$$

with $c = e^{ab} > 0, b > 0$.

Remark

Characterization of random variables, which are monotonic transformations of random variables being characterized, is a familiar operation. It is applied by Galambos and Kotz to give some characterizations of the exponential distribution.

3. Characterizations by the Proportional Hazards Model

Using theorem 2.1, characterizations by the proportional hazards model of location or scale parametrized distributions will be given in this section.

Firstly, the model will be defined.

Definition

Let X be a random variable as in the definition of the proportional odds model of section 2. Then the proportional hazards model is additive on loglog scale and defined by

$$\log\{-\log(1 - \gamma_j(\Theta))\} = c_j - \psi(\Theta)$$

where ψ depends on Θ only; for any Θ_1, Θ_2 the difference

$$\log\left\{\frac{\log(1-\gamma_j(\Theta_1))}{\log(1-\gamma_j(\Theta_2))}\right\} = \psi(\Theta_2) - \psi(\Theta_1)$$

does not depend on j.

It is required again, that $\psi(\Theta_2) - \psi(\Theta_1)$ only depends on $\Theta = \Theta_2 - \Theta_1$. Some manipulation, in case of location parameter family $F(x - \Theta)$ leads to

$$(1 - F(x - \Theta_1)) = (1 - F(x - \Theta_2))^{\exp(\psi(\Theta_2) - \psi(\Theta_1))}$$

or

$$(1 - F(x)) = (1 - F(x - \Theta))^{k(\theta)}$$

for some positive function $k(\Theta)$.

We will characterize F in the next theorem.

Theorem 3.1

Let $F(x)$ be a distribution function of a random variable X, having the whole interval $(-\infty, \infty)$ as support, for which

$$(1 - F(x)) = (1 - F(x - \Theta))^{k(\Theta)} \tag{8}$$

holds, for all real x and Θ, where $k(\Theta)$ is a non-constant (positive) function, not depending on x.

Then $F(x) = 1 - e^{-e^{b(x-a)}}$ for real a and $b > 0$, being the distribution function of the extreme value distribution.

Proof

By taking minus logarithms, (8) can be written

$$-\log(1 - F(x)) = k(\Theta)(-\log(1 - F(x - \Theta))) \tag{9}$$

for all real x, Θ.

Comparing (9) with relation (2) of section 2, in the form

$$H(x) = \phi(\Theta) H(x - \Theta)$$

which characterized $H(x)$ as $H(x) = e^{b(x-a)}$ it is seen that, replacing $H(x)$ by $-\log(1 - F(x))$, $F(x)$ is characterized as $F(x) = 1 - e^{-e^{b(x-a)}}$

Theorem 3.2

Let $G(y)$ be a distribution function of a positive random variable Y, having the whole interval $(0, \infty)$ as support, for which

$$(1 - G(y)) = (1 - G(\lambda y))^{k(\lambda)} \tag{10}$$

holds for all y, $\lambda > 0$, where $k(\lambda)$ is a non-constant function, not depending on y.

Then $G(y) = 1 - e^{-c^{-1}y^b}$ for b, $c > 0$, being the distribution function of the two-parameter Weibull distribution.

Proof

As before define X as $X = \log Y$, having distribution function $F(x) = G(e^x)$, then (10) can be written

$$(1 - F(\log y)) = (1 - F(\log y + \log\lambda))^{k(\lambda)} \text{ for all } y, \lambda > 0,$$

which characterizes $F(x)$ because of theorem 3.1

So a characterization of $G(y)$ is obtained, and

$$G(y) = F(\log y) = 1 - e^{-e^{-ab} y^b} = 1 - e^{-c^{-1} y^b}$$

which is the distribution function of the Weibull distribution, with shape parameter $b > 0$ and scale parameter $c = e^{ab} > 0$.

Remark

Of course, distribution functions may depend on other parameters than location and scale, which is so in the examples given. Other characterizations may be found, for instance by considering shape parameters using other transformation than the logarithmic.

Finally, the proof of theorem 3.1 shows an interesting similarity between the two proportional models; study of the transformation used could be worthwhile.

Acknowledgement

The author wishes to thank his colleague Mynt Zijlstra for some of his critical remarks as well as the referee for his suggestions considering the proof of the main theorem of this paper.

References

Agresti, A.: Generalized odds ratios for ordinal data. Biometrics **36**, 1980, 59–67.

–: Measures of nominal-ordinal association. J. Am. Statist. Ass., **76**, 1981, 524–529.

Galambos, J., and *S. Kotz*: Characterizations of probability distributions. Springer-Verlag, Berlin 1978.

Johnson, N.L., and *S. Kotz*: Distributions in statistics: continuous univariate distributions – 2. Houghton Mifflin Company, Boston 1970.

McCullagh, P.: Regression models for ordinal data. J. Roy. Statist. Soc. B **42**, 1980, 109–142.

Received April 14, 1982
(Revised version August 4, 1982)