Faculty Working Papers

A GENERALIZATION OF THE LOGIT TRANSFORMATION

Takamitsu Sawa

#433

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

September 5, 1977

A GENERALIZATION OF THE LOGIT TRANSFORMATION

Takamitsu Sawa

#433

A Generalization of the Logit Transformation

Takamitsu Sawa

University of Illinois, Champaign-Urbana

August 22, 1977

Prepared under National Science Foundation Grant SOC 76-22232

Abstract

We start with postulating a Poisson regression model with a random error term

$$p_Y(y|x,\xi) = \frac{[\lambda(x)\xi]^y}{y!} e^{-\lambda(x)\xi}, \quad y = 0, 1, 2, \ldots$$

where x is assumed to be a nonstochastic variable; ξ is a random variable having an χ^2 distribution with 2r degrees of freedom. Then the marginal distribution of Y is the negative binomial distribution with probability function

$$p_Y(y) = \frac{\Gamma(r+y)}{\Gamma(y+1)\Gamma(r)} \left[\frac{1}{1+\lambda(x)}\right]^r \left[\frac{\lambda(x)}{1+\lambda(x)}\right]^y$$

for $y = 0, 1, 2, \ldots; r > 0$. We define a binary response variable Z such that $Z = 1$ iff $Y \geq 1$ and $Z = 0$ iff $Y = 0$. If $\lambda(x) = \exp(\alpha+\beta x)$, a generalized logit model

$$\log \left[\frac{1 - p_0(x)^{1/r}}{p_0(x)^{1/r}}\right] = \alpha + \beta x$$

follows, where $p_0(x) = P(Z=0|x)$. To see the practical usefulness of the generalization, we fitted the model to Ashford-Sowden's data. As compared with an ordinary logit model, significant improvement in goodness-of-fit has been observed in terms of the χ^2 goodness-of-fit statistic. Especially, it is remarkable for the tail areas of $p_0(x)$. Some more versions of qualitative response models will be also discussed in their connection to the Poisson process.

1. Introduction

The purpose of the present paper is to propose a possible generalization of the logit transformation. The logit transformation of a binomial probability has been widely used to analyze qualitative data in socio-economic investigations as well as in biometric research. The generalized logit model developed here involves only one more parameter than the conventional logit model. Therefore, the simplicity of the latter model is essentially preserved by our generalization. We base our derivation on the following assumption: a binary response may be observed as an indicator for an underlying (possibly unobservable) nonhomogenous compound Poisson process: i.e., it indicates whether or not the number of events occurred in the process exceeds a fixed unknown threshold. The Poisson rate parameter is assumed to depend on some exogenous factors as well as a multiplicative random component. Since the Poisson process is derived on the basis of a few weak assumptions, it would be fair to claim that our approach gives another natural interpretation to the logit model.

The genesis of the probability integral model, including the logit and probit models, is usually described by postulating a hypothetical random variable called <u>tolerance</u>, the variation of which causes the randomness in the binary response. If the tolerance has a logistic distribution, the logit model follows (See, for instance, Cox (1970).) In Section 2 we incidentally propose a probability integral model of a chi-square distribution. That is, if we have a Poisson process with nonhomogenous rate parameter as an underlying structure for a binary response, what we call the <u>chisquit model</u> immediately follows.

In order to see practical relevance of our generalization, we fit th model to some empirical data. The results given in Section 4 show that our generalization improves the statistical fit of the model quite substantially, albeit it causes no essential difficulties in computation.

2. Poisson Regression and Chisquit Model

To begin with let us consider a random phenomenon where events occur in a Poisson process with non-homogeneous rate parameter λ. For the time being, we assume that the variation of λ is fully explained by some independent variables $\underset{\sim}{x}$. Then the number, say Y, of events to occur in an interval of fixed length t and for fixed x is a Poisson random variable with probability function

$$(2.1) \qquad P_Y(y\ x) = \frac{[\lambda(\underset{\sim}{x})t]^y}{y!}\, e^{-\lambda(\underset{\sim}{x})t}$$

for y = 0, 1, 2, ..., where $\lambda(\underset{\sim}{x})$ is a function of $\underset{\sim}{x}$ and its range is limited to positive half of the real line. If $\lambda(\underset{\sim}{x})$ is specified up to its functional form and a random sample of Y is observed with corresponding value of $\underset{\sim}{x}$, then we can make inferences about unknown parameters in $\lambda(\underset{\sim}{x})$. This is called the Poisson regression analysis, special cases of which have been investigated by Gart [1964] and Jorgenson [1961].[1/]

If we let T be the continuous amount of time (or area, distance, etc.) required to observe the r-th event in the Poisson process with rate parameter $\lambda(\underset{\sim}{x})$, starting from an arbitrary point in the process, the nonnegative random variable T has a Gamma distribution with density function $f(t) = \lambda(\lambda t)^{r-1}e^{-\lambda t}/\Gamma(r)$ for $t > 0$ and $f(t) = 0$ elsewhere. Hence we obtain an obvious equality

$$(2.2) \qquad P(Y \geq k | x, t) = P(T \leq t) = F_{\chi^2_{2k}} [2t\lambda(x)] ,$$

where $F_{\chi^2_{2k}}$ is the cumulative distribution function (cdf) of χ^2-distribution with 2k degrees of freedom. (See, for instance, Johnson and Kotz [1969], p. 98.)

Now let us define the following binary response model on the Poisson process: a qualitative change (catastrophe) that concerns us occurs if and only if $Y \geq k$; namely, a binary random variable Z equals one if and only if $Y \geq k$ and zero elsewhere. This is a version of multi-hit model used in biological application. As a practical example this relates to the case where Y stands for a random accumulation of causes of a certain catastrophic change: if the number of an individual's accumulated causes exceeds a threshold k, then a catastrope occurs to him; otherwise, he remains in the same state. The degree of the change would be somehow related to the amount by which Y exceeds k. However, what we are concerned with and actually observe is a binary response: whether or not the catastrophe occurred to each individual. The rate parameter λ which is intrinsic to each individual is regarded as indicating his proneness to the catastrophe which is supposed to be determined by his characteristics as well as some exogenous factors.

Another example, presented below, is referring to limitation of observability. It very often happens that due to certain limitation of observability or some other reasons, we witness only an all-or-none response: whether or not at least one event has occurred to each individual in an interval of fixed length. To put it differently, the number of events which might have occurred is unobservable or outside our concern. A typical example may be a survey research on possession

of a certain durable goods, say a car. The survey is often concerned only with a binary response: whether or not each individual possesses a car. It conceals the number of cars he possesses as well as the quality of his car.

If we adjust the scale of measuring length so that t = 1/2, then we obtain a binary response model[2/]

$$P(Z=1|\underset{\sim}{x}) = F_{\chi^2_{2k}}[\lambda(\underset{\sim}{x})] \qquad (2.3)$$

The unknown parameter k may or may not have definite physical meaning. The function $\lambda(\underset{\sim}{x})$ is usually specified as either linear, exponential, or multiplicative function. If the value of k is not determined theoretically, then it should be regarded as a parameter that must be estimated from data simultaneously with the parameters in $\lambda(\underset{\sim}{x})$.

The underlying structure of the probit and logit models is often described by postulating the existence of an unobservable (hypothetical) random variable called the <u>tolerance</u>: namely, Z=1 if and only if the tolerance, say U, falls below the threshold, $c(\underset{\sim}{x})$ say, determined by an individual's characteristics $\underset{\sim}{x}$.[3/] If the cdf of the tolerance is F_U, then

$$P(Z=1|x) = F_U[c(x)] \ . \qquad (2.4)$$

If U has either normal or logistic distribution, the probit or logit model follows, respectively. Our model, straightforwardly derived from a nonhomogeneous Poisson process, may be regarded as an alternative specification of the tolerance model, and it could be appropriately called the <u>chisquit model</u>.

Since U is intrinsically a hypothetical variable, there is no reason at all to confine its distribution to a class of symmetric distributions. In some practical applications it might be more adequate to assume that the tolerance is distributed with some skewness and its range is bounded below. Since χ^2-distribution is asymptotically normal as its degrees of freedom become large, it may be fair to say that the probit model is obtained as a limiting form of the chisquit model.

As for estimation, no particular difficulty arises if the value of k is specified a priori. We can employ essentially similar methods to that used to estimate the probit and logit models. If k is unspecified a priori, it could be estimated simultaneously with the parameters in $\lambda(x)$.[4/]

The distribution of $\log(\chi^2)$ approaches normal distribution more rapidly than that of χ^2 itself. (See Johnson and Kotz [1970], p. 181.) Therefore, if the varying rate parameter is reasonably specified as an exponential function such as $\exp(\alpha+\beta x)$, then

$$P(Z=1|x) = P(\log \chi^2_{2k} \leq \alpha + \beta x) \quad , \tag{2.5}$$

the right-hand side of which could be very closely approximated by the cdf of the normal distribution with mean log(2k) and variance 1/k, unless k is extremely small. Moreover, if $\lambda(x)$ is a multiplicative function such as αx^{β}, we have a log-linear function of x instead of a linear one on the right-hand side of (2.5). The above consideration leads us to the following.

Suppose that the linear or log-linear probit model fits given data very well. Then this suggests that the varying rate parameter in

the assumed Poisson process might be just appropriately specified as an exponential or multiplicative function of x. Of course, it is fair to say that there is no way of discriminating a model with linear $\lambda(x)$ and k large enough to permit normal approximation from another alternative model with exponential $\lambda(x)$ and small k. In either of these two cases the linear probit model will fit data very well.

3. Generalized Logit Transformation

One of the apparent shortcomings of the chisquit model is the following: we assume that the rate parameter λ for each individual is completely determined by a finite number of explanatory variables $\underset{\sim}{x}$. This is obviously unrealistic and necessitates modification of the model. Also, in practice, we need to keep the number of the variables as small as possible. To cope with this we permit the rate parameter to be a random function of the characteristics set $\underset{\sim}{x}$, i.e.

(3.1) $$\lambda = \lambda(\underset{\sim}{x})\xi ,$$

where 2ξ is a random variable having χ^2 distribution with 2r degrees of freedom. It should be noted that 2r need not be an integer. Also, as the model is multiplicative, no loss of generality is caused by assuming the distribution of 2ξ is χ^2 instead of a Gamma distribution. Given ξ, Y has a conditional Poisson distribution. It is straightforward to show that Y is unconditionally distributed as negative binomial with probability function[5/]

(3.2) $$p_Y(y) = \frac{\Gamma(r+y)}{\Gamma(y+1)\Gamma(r)} \left[\frac{1}{1+\lambda(x)}\right]^r \left[\frac{\lambda(x)}{1+\lambda(x)}\right]^y$$

for y = 0, 1, 2, ..., and r > 0.

Now let k be a threshold: i.e., we have a binary response Z such that Z = 1 if $Y \geq k$ and Z = 0 otherwise. We note that

(3.3) $$P(Y \geq k) = \sum_{y=k}^{\infty} \frac{\Gamma(r+y)}{\Gamma(y+1)\Gamma(r)} \left[\frac{1}{1+\lambda(x)}\right]^r \left[\frac{\lambda(x)}{1+\lambda(x)}\right]^y$$

$$= I_{\psi(x)}(k,r) \; ,$$

where

(3.4) $$\psi(x) = \frac{\lambda(x)}{1+\lambda(x)} \; ;$$

I is the incomplete Beta function. A generalized logit transformation

(3.5) $$\log \left\{\frac{1-p_0(x)^{1/r}}{p_0(x)^{1/r}}\right\} = \log \lambda(x)$$

where $p_0(x) = P(Z=0|x)$ and r is a positive constant, may be derived from either of the following two models. First, let us suppose that k = 1, i.e., the binary response indicates whether or not at least one event occurs. Then we have

(3.6) $$p_0(x) = P(Z=0|x) = \frac{1}{[1+\lambda(x)]^r}$$

It is straightforward to verify that this implies (3.5). Second, let ξ be exponentially distributed, i.e., r = 1. Then model (3.6) again follows with r replaced by k. Which model is more appropriate is a problem that should be answered case by case on a priori ground; that is, the two models may not be discriminated objectively by data. The conventional logit transformation (3.5) with r = 1 corresponds to all-or-none binary response (i.e., k = 1) defined on a compound nonhomogeneous Poisson process with exponential error distribution (i.e., r = 1).

Moreover, if we specify $\lambda(x) = e^{\alpha+\beta x}$ and $r = 1$, then the familiar linear logit model follows. If we specify a multiplicative model $\lambda(x) = \alpha x^{\beta}$, the log-linear logit model follows. Since r is an unknown parameter appearing in the error distribution, it should be simultaneously estimated with the parameters in $\lambda(x)$ from sample observations.

When we have grouped data, the simplest way of estimating the model would be the so-called Berkson's minimum chi-squares method. The asymptotic variance of the empirical generalized logit transformation, $\log \{[1-\hat{p}_0^{1/r}]/\hat{p}_0^{1/r}\}$, is given by

$$(3.7) \qquad \frac{1-p_0}{n\, r^2 (1-p_0^{1/r})^2 p_0}$$

where $p_0 = P(Z=0|x)$; $\hat{p}_0$ is the estimate of p_0 based on a sample of size n. Replacing the unknown p_0 by its sample estimate $\hat{p}_0$, we can apply generalized least squares to obtain estimates for parameters in $\lambda(x)$ for a given value of r. The optimal value of r might be found by minimizing, for example, the χ^2 goodness-of-fit test statistic with respect to r.

4. Numerical Results

To examine practical relevance of our generalization we fitted the generalized logit model to Ashford and Sowden's [1970] data. Their data, presented in tables after aggregation, consisted of the number of coal miners in nine 5-year-wide age group reporting either, neither, or both of the respiratory symptoms, breathlessness and wheeze. They developed the bivariate probit model to analyze this data, which was later reanalyzed by Grizzle [1971] and Mantel and Brown [1973]. In

fact, the Ashford-Sowden data should be adequately analyzed by a certain bivariate model, but for simplicity we neglect the multivariate as well as multinomial aspect of the data and treat it as if it consisted of two separate sets of binomial data, one for each symptom.

The optimal value of r was determined so that the χ^2 goodness-of-fit test statistic

$$\sum_{i=1}^{9} \sum_{j=1}^{2} \frac{(y_{ij}-\hat{y}_{ij})^2}{\hat{y}_{ij}} \tag{4.1}$$

be minimized respectively for each symptom, where y_{ij} and $\hat{y}_{ij}$ are the observed and interpolated frequencies in the cell of the i-th age group and either having or not having each symptom (j=1 corresponds to "yes" and j=2 corresponds to "no").

Following Grizzle [1971] and Mantel and Brown [1973], the normalized median age of each group x = (median age-17)/5 is taken as the explanatory variable. For simplicity, we assume a linear function of x for the right-hand side of (3.5) and employ Berkson's minimum chi-squares method to estimate coefficients for each given value of r.

The estimated generalized logit models are

$$P(Z_1=0|x) = \frac{1}{[1 + \exp(-4.012+0.632x)]^{0.293}} \tag{4.2}$$

with χ^2 = 4.998 for breathlessness;

$$P(Z_2=0|x) = \frac{1}{[1 + \exp(-2.226+0.400x)]^{0.360}} \tag{4.3}$$

with χ^2 = 3.653 for wheeze, as compared to the ordinary logit models

(4.4) $$P(Z_1=0|x) = \frac{1}{1 + \exp(-4.804+0.510x)}$$

with χ^2 = 16.554 for breathlessness;

(4.5) $$P(Z_2=0|x) = \frac{1}{1 + \exp(-3.116+0.326x)}$$

with χ^2 = 8.027 for wheeze. The degrees of freedom are 14 and 15, respectively, for the generalized and ordinary logit models. The interpolated values are tabulated in Tables 1 and 2 with observed values. It may be fair to say that the improvement is significant on the whole. In particular, it is remarkable for the tail areas.

To see the sensitivity of the model to the change of r we present in Table 3 the estimates, $\hat{\alpha}$ and $\hat{\beta}$, of a constant term and coefficient to x with the associated value of χ^2 statistic for different values of r.

To give contrast to the Ashford-Sowden data the model was also fitted to Morimune's [1976] data relating private ownership of a house to income. It turned out that the log-linear model is far more appropriate than the linear model. The estimated generalized logit model is

(4.6) $$P(Z=0|x) = \frac{1}{[1 + \exp(-9.393+0.911 \log x)]^{3.794}}$$

with χ^2 = 8.616, as compared to the ordinary logit model

(4.7) $$P(Z=0|x) = \frac{1}{1 + \exp(-11.102+1.298 \log x)}$$

with χ^2 = 9.568. Also, the estimated chisquit model is

(4.8) $$P(Z=0|x) = F_{\chi^2_{7.158}}(-1.573+0.403 \log x)$$

with χ^2 = 8.530.

In this example the improvement of fit is not striking. If we take into account the decrease of the degrees of freedom, almost no significant gain is observed by generalizing the logit model by introducing a transformation parameter r. However, if you look at Table 4, you will realize that in tail areas of the distribution the goodness-of-fit was improved, albeit slightly, by generalizing the model. Also, it is interesting to note that the generalized logit model and the chisquit model gave almost the same interpolated numbers.

It is straightforward to extend the model to the case of multinomial ordered response. Also, possible further developments of the work in this note will include the expansion to multivariate cases by postulating a multivariate Poisson process.

Table 1. Ashford-Sowden Data on Breathlessness

Age	x	Yes			No		
		Obs.	r=.293	r=1	Obs.	r=.293	r=1
20-24	1	16	19.0	26.3	1936	1933.0	1925.7
25-29	2	32	32.3	39.8	1759	1758.7	1751.2
30-34	3	73	69.3	77.1	2040	2043.7	2035.9
35-39	4	169	161.8	165.2	2614	2621.2	2617.8
40-44	5	223	225.0	216.2	2051	2049.0	2057.8
45-49	6	357	379.8	356.4	2036	2013.2	2036.6
50-54	7	521	494.4	471.7	1569	1595.6	1618.3
55-59	8	558	570.6	571.9	1192	1179.4	1178.1
60-64	9	478	475.2	507.8	658	660.8	628.2

The columns headed by r=.293 and r=1 contain interpolated values by the models (11) and (13), respectively.

Table 2. Ashford-Sowden Data on Wheeze

Age	x	Yes			No		
		Obs.	r=.360	r=1	Obs.	r=.360	r=1
20-24	1	104	102.0	112.9	1848	1850.0	1839.1
25-29	2	128	133.4	140.4	1663	1657.6	1650.6
30-34	3	231	220.3	222.8	1882	1892.7	1890.2
35-39	4	378	397.1	390.7	2405	2385.9	2392.3
40-44	5	442	432.2	419.6	1832	1841.8	1854.4
45-49	6	593	587.5	571.1	1800	1805.5	1821.9
50-54	7	649	641.8	632.8	1441	1448.2	1457.2
55-59	8	631	651.0	657.4	1119	1099.0	1092.6
60-64	9	504	495.0	514.6	628	637.0	617.4

The columns headed by r=.360 and r=1 contain interpolated values by the models (12) and (14).

Table 3. Estimates and χ^2 Statistic for Different Values of r: Ashford-Sowden Wheeze Data

r	$\hat{\alpha}$	$\hat{\beta}$	χ^2
10	-5.345	.289	14.74
8	-5.124	.290	14.50
6	-4.840	.292	14.11
4	-4.442	.295	13.35
2	-3.769	.306	11.29
1	-3.116	.326	8.03
.8	-2.913	.336	6.79
.6	-2.657	.354	5.23
.4	-2.313	.388	3.75
.3	-2.083	.423	3.97
.2	-1.785	.493	7.86
.1	-1.363	.709	30.33

Table 4. Morimune's Data on House Ownership and Income

		House Owner				Non House Owner			
Income	x	Obs.	r=3.794	r=1	k=3.579	Obs.	r=3.794	r=1	k=3.579
0- 5000	3500	99	101	97	101	159	157	161	157
5000- 6500	6500	132	137	139	137	104	99	98	99
6500- 8000	7350	178	167	169	168	99	110	108	109
8000- 9500	8950	151	146	147	146	69	74	73	74
9500-11500	10600	165	167	168	167	70	68	67	68
11500-13500	12750	156	169	170	169	67	54	53	54
13500-16000	14750	212	212	212	212	54	54	55	54
16000-20000	17550	216	210	208	210	34	40	42	41
20000-30000	23000	223	224	220	224	29	28	32	28

The columns in each cell are the observed number and the interpolated numbers by the models (15), (16) and (17) in order.

Footnotes

1. Gart [1964] analyzed the case when $\lambda(x)$ is a linear function of a single explanatory variable without a constant term. Jorgenson [1961] developed the maximum likelihood estimation for the case when $\lambda(x)$ is a linear function of several variables. As far as I know, there have been quite a few applications of the Poisson regression analysis to real problems.

2. The rate parameter is not independent of the choice of the scale of measuring length. However, if we assume an exponential or multiplicative function for $\lambda(\underset{\sim}{x})$, it has no effect on relevant coefficients of the variables $\underset{\sim}{x}$; i.e., only a constant term is affected by the choice of the scale.

3. More details of the tolerance model are referred to Cox [1970]. In the context of econometric analysis, the underlying structure of the model is often described by postulating the existence of the random utility instead of the tolerance.

4. The chibit model defined by (3) may be regarded as a reduced form of the binary response model defined on a Poisson process with varying parameter $\lambda(x)$. In this case the parameter k must be an integer. It is possible, however, to view (3) as a version of the tolerance model. Then k need not be an integer.

5. The derivation of the negative binomial distribution as a compounding Poisson and Gamma distribution is found in most textbooks. See, for instance, Johnson and Kotz [1969]. This distribution is also derived by assuming different sorts of underlying chance mechanisms. A comprehensive review is given by Boswell and Patil [1970]. Under certain circumstances it may produce a more reasonable physical interpretation to postulate another underlying chance mechanism instead of the compound Poisson regression.

References

Ashford, J. R. and R. R. Sowden [1970], "Multivariate Probit Analysis," Biometrics, 26, 535-46.

Bowell, M. T. and G. P. Patil [1970], "Chance Mechanisms Generating the Negative Binomial Distributions," Random Counts in Scientific Work, 1, 3-22.

Cox, D. R. [1970], "The Analysis of Binary Data," London: Methuen.

Gart, J. J. [1964], "The Analysis of Poisson Regression with an Application in Virology," Biometrika, 51, 517-521.

Grizzle, D. J. [1971], "Multivariate Logit Analysis," Biometrics, 27, 1057-62.

Johnson, N. I. and S. Kotz [1969], Discrete Distributions, Houghton Mifflin Company, Boston.

Johnson, N. I. and S. Kotz [1970], Continuous Univariate Distributions-1, Houghton Mifflin Company, Boston.

Jorgenson, D. W. [1961], "Multiple Regression Analysis of a Poisson Process," Journal of the American Statistical Association, 56, 235-245.

Mantel, N. and C. Brown [1973], "A Logistic Reanalysis of Ashford and Sowden's Data on Respiratory Symptoms in British Coal Miners," Biometrics, 29, 649-65.

Morimune, K. [1976], Some Minimum Chi-Square Estimators and Comparisons of Normal and Logistic Models in Qualitative Response Analysis, Ph.D. dissertation, Department of Economics, Stanford University, Stanford, California.