Regression Model from Binomial / Multinomial distribution

Mingwei Fei

December 3, 2022

1 Regression model

One thing I could learn is to understand the theory from the real practice, it would be much easier to understand why we do that in the reverse way.

1.1 logistic regression model

1.1.1 distribution assumption

Pay attention to the $x_i, i = 1, ...N$, it says that N different x_i categories.
Binomial distribution assumption: $y_i|X_i \sim \mathbf{B}(n_i; \pi(x_i))$, for $i = 1, ..N$

1.1.2 structural assumption

$$g(\pi(x_i)) = logit(\pi_i) = log\frac{\pi(x_i)}{1-\pi(x_i)} = x_i^T\beta,$$

is related to x_i by Using different link functions in the structural assumption of the logistic regression model leads to other regression models for binomial and binary responses. Let's consider four link functions as follows

(i) $g_1(t) = log(t/(1-t))$ (logit link),

(ii) $g_2(t) = \Phi^{-1}(t)$ (probit link or inverse Normal function),

(iii) $g_3(t) = log{-}log(1-t)$(complementary log-log link),

(iv) $g_4(t) = -log{-}log(t)$ (log-log link).

Each link function leads to a regression model. For instance, the regression model with the probit link is called a probit regression. All four link functions are obtained as the inverses of well-known cumulative distribution functions having support on the entire real line. Moreover, they are continuous and increasing in (0, 1). Because the logistic and probit functions are almost linearly related over the interval [0.1, 0.9], it is difficult to discriminate between them based on goodness-of-fit statistics. Thus, the use of a probit regression usually produces results which are similar to those from logistic regression.

1.1.3 Likelihood function

To solve the logistic regression model, we will look at log-likelihood of the logistic regression model (ignore the constant)

$$ln(\beta) = \sum_{i=1}^{N} \left[y_i x_i^T \beta - n_i log(1 + exp(x_i^T \beta)) \right]$$

Taking the first and second derivatives, we have

$$\frac{\partial ln(\beta)}{\partial \beta} = \sum_{i=1}^{N} \left[y_i x_i - n_i \frac{exp(x_i^T \beta) x_i}{1 + exp(x_i^T \beta)} \right] = \sum_{i=1}^{N} \left[y_i x_i - n_i \pi_i x_i \right]$$
$$\frac{\partial^2 ln(\beta)}{\partial \beta \, \partial \beta} = \sum_{i=1}^{N} -n_i \frac{exp(x_i^T \beta) x_i}{(1 + exp(x_i^T \beta))^2} = \sum_{i=1}^{N} -n_i \pi_i (1 - \pi_i) x_i^{\otimes 2}$$

The Newton-Raphson algorithm is given by

$$\frac{\partial ln(\hat{\beta})}{\partial \beta} = 0 = \frac{\partial ln(\hat{\beta^*})}{\partial \beta} + \frac{\partial^2 ln(\beta^*)}{\partial \beta \, \partial \beta} (\hat{\beta} - \beta^*) + o_p(1)$$
$$\hat{\beta} = \beta^* - \{ \frac{\partial^2 ln(\beta^*)}{\partial \beta \, \partial \beta} \}^{-1} \frac{\partial ln(\beta^*)}{\partial \beta}$$
$$\beta^{k+1} = \beta^k + \{ \sum_{i=1}^{N} n_i \pi_i (1 - \pi_i) x_i^{\otimes 2} \}^{-1} \sum_{i=1}^{N} \left[y_i x_i - n_i \pi_i x_i \right]$$

The deviance function is the twice of the difference of log-likelihood between regression model and saturated model, (treat each y_i as μ_i)

$$ln(\pi_i) = \sum_{i=1}^{n} y_i log(\pi_i) + (n_i - y_i) log(1 - \pi_i)$$
$$D(y, \hat{\pi}) = 2[ln(y_i) - ln(\beta)] = 2 \sum_{i=1}^{n} [y_i ln(\frac{y_i}{\pi_i}) - (n_i - y_i) ln(\frac{1 - y_i}{1 - \pi_i})]$$

The asymptotic distribution of β

$$I(\beta) = E[-\frac{\partial^2 ln(\beta^*)}{\partial \beta \, \partial \beta}], \qquad I(\beta) = \frac{1}{n} \sum_{i=1}^{N} [-n_i \pi_i (1 - \pi_i) x_i^{\otimes 2}]$$
$$I_n(\hat{\beta}) = E[-\frac{\partial^2 ln(\hat{\beta})}{\partial \beta \, \partial \beta}], \qquad I(\beta) = \frac{1}{n} I_n(\hat{\beta})$$
$$\sqrt{n}(\hat{\beta} - \beta) \to N(0, I(\beta)^{-1} I_p) = N(0, n I_n(\hat{\beta})^{-1} I_p)$$

The odds ratio comparing the two categories of the covariate can be obtained

by exponentiating the coefficient of the covariate in the logistic regression model.

$$
\begin{aligned}
logit(y_i|x_i) &= \beta_1 + \beta_2 x_i \\
logit(y_1 = 1|x_i = 1) &= \beta_1 + \beta_2, logit(y_1 = 1|x_i = 0) = \beta_1 \\
\beta_2 &= logit(y_1 = 1|x_i = 1) - logit(y_1 = 1|x_i = 0) = log\frac{\frac{p(y_i=1|x_i=1)}{1-p(y_i=1|x_i=1)}}{\frac{p(y_i=1|x_i=0)}{1-p(y_i=1|x_i=0)}} \\
OR &= exp(\beta_2)
\end{aligned}
$$

Generalize the odds ratio in logistic regression, the covariance of odds ratio by delta method

$$
\begin{aligned}
exp(X_A - X_b)^T\beta_2 &= \frac{odds((y_1 = 1|x_A))}{odds((y_1 = 1|x_B))} \\
s.e.(OR) &= \hat{OR}\sqrt{(X_A - X_b)^T Cov(\beta)(X_A - X_b)}
\end{aligned}
$$

Don't mess the $Cov(\beta)$ with the variance of $logOR$ which expressed in π

1.2 ordinal regression model

(a) Distribution assumption.

$$
\begin{aligned}
Z_i|X_i &\sim Multi\,(1; \pi_1(x_i), \pi_2(x_i), \pi_3((x)_i, ...\pi_I((x)_i) \\
p(Z_i) &= \pi_1^{n_1}\pi_2^{n_2}\pi_3^{n_3}(1 - \pi_1 - \pi_2 - \pi_3)^{n_4} \\
p(Z_i <= j) &=
\end{aligned}
$$

(b) Structural assumption

$$
\begin{aligned}
Z_i &\in (1, 2, ..I), \quad i = 1, 2...n \\
Z_i = j \quad Z_{ij} &= 1, \quad Z_{ik} = 0(k \neq j), \\
P(Z_{ij}) &= E[Z_{ij}] = \pi_{ij} \\
P(Z_{ij} \leq j|X_i) &= \sum_{k=1}^{j} \pi_k(x_i)
\end{aligned}
$$

is related to x_i by

$$g(P(Z_{ij} \leq j|X_i)) = \alpha_j + x_i^T\beta, \qquad j = 1, 2...I - 1, i = 1, ...n, \alpha_1 \leq \alpha_2 \leq ...\alpha_{I-1}$$

$\dot{g()}$ is an increasing link function, x_i is a $p \times 1$ covariate vector, and $\beta = (\beta_1, \beta_2, ..\beta_p)^T$ is defined in a subset $\mathcal{B}$ of $R^p(p < n)$. The ordinal regression model transform the multinomial into binomial, and assume the odds ratio effect the same between covariates, which β is the same, only the α_i is different.That means that the effect of the independent variable is the

same for different logit functions. That's an assumption you have to check. That's also the reason the model is also called the proportional odds model.

$$\theta_1 = \frac{P(x=1)}{1-P(x=1)}$$
$$\theta_2 = \frac{P(x=1,2)}{1-P(x=1,2)}$$
$$\theta_3 = \frac{P(x=1,2,3)}{1-P(x=1,2,3)}$$
$$ln(\theta) = \alpha_j + \beta X$$

(c) Link function

The link function is the function of the probabilities that results in a linear model in the parameters. It defines what goes on the left side of the equation. It's the link between the random component on the left side of the equation and the systematic component link on the right. In the criminal rating example, the link function is the logit function, since the log of the odds results is equal to the linear combination of the parameters.

Using different link functions in the structural assumption of ordinal regression leads to various models for ordinal responses.Similar to logistic regression, we may choose four link functions as follows:

(i) $g_1(t) = log(t/(1-t))$ (logit link),

(ii) $g_2(t) = \Phi^{-1}(t)$ (probit link or inverse Normal function),

(iii) $g_3(t) = log{-}log(1-t)$(complementary log-log link),

(iv) $g_4(t) = -log{-}log(t)$ (log-log link).

These link functions are increasing functions of t in $(0,1)$. In particular, the logit link has been widely used in the literature. The ordinal regression model for the logit link is known as the proportional odds model.
The proportional odds model assumes that the cumulative logits are defined as

$$logit[P(Z_{ij} \leq j|X_i)] = \alpha_j + x_i^T\beta$$

Thus, each cumulative logit has its own intercept. Because the $\{\alpha_j\}$ are increasing, then $P(Z \leq j|x)$ is an increasing function of j, indicating the ordering feature of the ordinal response. Moreover, the model has the same covariate effect β for each logit. The proportional odds model makes the strong assumption that the odds ratio is invariant dichotomization of the ordinal response. Specifically, for any x_1 and x_2,

$$\frac{odds(P(Z \leq j|x_1))}{odds(P(Z \leq j|x_2))} = exp\left((x_1 - x_2)^T\beta\right)$$

holds for all j. Thus, the log cumulative odds ratio is proportional to the distance between x_1 and x_2. Because of this property, the ordinal response model based on the cumulative logits is called the proportional odds model.

1.2.1 Log-likelihood

(a) Write out the likelihood function for all the parameters.

$$logit(p(Z_j <= j)) = \alpha_j + x_i^T\beta, \qquad logit(p(Z_j <= j-1)) = \alpha_{j-1} + x_i^T\beta,$$

$$p(Z_j <= j) = \frac{exp(\alpha_j + x_i^T\beta)}{1 + exp(\alpha_j + x_i^T\beta)}, \qquad p(Z_j <= j-1) = \frac{exp(\alpha_{j-1} + x_i^T\beta)}{1 + exp(\alpha_{j-1} + x_i^T\beta)}$$

$$p(Z_j = j) = \pi_{ij} = p(Z_j <= j) - p(Z_j <= j-1) = \frac{exp(\alpha_j + x_i^T\beta)}{1 + exp(\alpha_j + x_i^T\beta)} - \frac{exp(\alpha_{j-1} + x_i^T\beta)}{1 + exp(\alpha_{j-1} + x_i^T\beta)}$$

$$p(Z) = \prod_{i=1}^{n}\prod_{j=1}^{I} \pi_{ij}^{I(Z_j=j)}$$

$$= \prod_{i=1}^{n}\prod_{j=1}^{I} \left[\frac{exp(\alpha_j + x_i^T\beta)}{1 + exp(\alpha_j + x_i^T\beta)} - \frac{exp(\alpha_{j-1} + x_i^T\beta)}{1 + exp(\alpha_{j-1} + x_i^T\beta)}\right]^{I(Z_j=j)}$$

(b) Compute $\dot{l}_n$ and $\ddot{l}_n$

$$ln(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{I} I(Z_j = j) log \left[\frac{exp(\alpha_j + x_i^T\beta)}{1 + exp(\alpha_j + x_i^T\beta)} - \frac{exp(\alpha_{j-1} + x_i^T\beta)}{1 + exp(\alpha_{j-1} + x_i^T\beta)}\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{I} I(Z_j = j) log \frac{exp(\alpha_j + x_i^T\beta) - exp(\alpha_{j-1} + x_i^T\beta)}{[1 + exp(\alpha_j + x_i^T\beta)][1 + exp(\alpha_{j-1} + x_i^T\beta)]}$$

$$\frac{\partial ln(\theta)}{\partial \alpha_j} =$$

(c) Write out the likelihood function for probit link (ordinal probit model).

$$g(p(Z_j) <= j) = \Phi^{-1}(p(Z_j)) = \alpha_j + x_i^T\beta$$

$$p(Z_j <= j) = \Phi(\alpha_j + x_i^T\beta), \qquad p(Z_j <= j-1) = \Phi(\alpha_{j-1} + x_i^T\beta)$$

$$p(Z_j = j) = \pi_{ij} = p(Z_j <= j) - p(Z_j <= j-1) = \Phi(\alpha_j + x_i^T\beta) - \Phi(\alpha_{j-1} + x_i^T\beta)$$

$$p(Z) = \prod_{i=1}^{n}\prod_{j=1}^{I} \pi_{ij}^{I(Z_j=j)}$$

$$= \prod_{i=1}^{n}\prod_{j=1}^{I} \left[\Phi(\alpha_j + x_i^T\beta) - \Phi(\alpha_{j-1} + x_i^T\beta)\right]^{I(Z_j=j)}$$

1.2.2 Odds Ratio

Show that Odds Ratio does not depend on category index $i = 1, 2, ...I$.

$$
\begin{aligned}
logit(p(Z_j <= j)) = \frac{p(Z_j <= j)}{1 - p(Z_j <= j)} &= \alpha_j + x_i^T \beta \\
logOR = logit(p(Z_j <= j)|x_A) - logit(p(Z_j <= j)|x_B) &= (x_A - x_B)^T \beta \\
OR = exp((x_A - x_B)^T \beta)
\end{aligned}
$$

OR does not depend on α_j

1.3 Nominal regression model

$$
\begin{aligned}
Z_i|X_i &\sim Multi\,(1; \pi_1(x_i), \pi_2(x_i), \pi_3((x)_i, ...\pi_I((x)_i) \\
p(Z_i) &= \pi_1^{n_1} \pi_2^{n_2} \pi_3^{n_3} (1 - \pi_1 - \pi_2 - \pi_3)^{n_4} \\
ln(Z_i) &= n_1 log(\pi_1) + n_2 log(\pi_2) + n_3 log(\pi_3) + (n - n_1 - n_2 - n_3) log(\pi_4) \\
&= n_1 log(\frac{\pi_1}{\pi_4}) + n_2 log(\frac{\pi_2}{\pi_4}) + n_3 log(\frac{\pi_3}{\pi_4}) + n log(\pi_4) \\
log(\frac{\pi_{i1}}{\pi_{i4}}) = x_i^T \beta_1, \qquad log(\frac{\pi_{i2}}{\pi_{i4}}) &= x_i^T \beta_2, \qquad log(\frac{\pi_{i3}}{\pi_{i4}}) = x_i^T \beta_3
\end{aligned}
$$

is related to x_i by $g(\dot{})$ is an increasing link function, x_i is a $p \times 1$ covariate vector, and $\beta = (\beta_1, \beta_2, ..\beta_p)^T$ is defined in a subset $\mathcal{B}$ of $R^p (p < n)$.
β_{Ip} dimension depends on how many y_j and how many columns in x_i. Here we choose π_4 as the referent category, and set $\beta_4 = 0$.

For the multicategorical logit model, we often choose a referent category and compare other categories with the referent category. Statistically, the use of the referent category is associated with the issue of parameter identification. Specifically, we have

$$log \frac{P(Z = j|x_i)}{P(Z = 1|x_i)} = x_i^T (\beta_j - \beta_1)$$

Thus only $\beta_j - \beta_1, j = 2, ..I$ is estimable. For identification purposes, we may set $\beta_1 = 0$. Therefore, for $j = 2, ...I$, we have

$$log \frac{P(Z = j|x_i)}{P(Z = 1|x_i)} = x_i^T (\beta_j)$$

in which β_j the log odds for category j with respect to the referent category 1. Thus, we can obtain $(I - 1)$ odds ratios for the polytomous response.

1.3.1 Likelihood function

$$ln(\beta) = \sum_{i=1}^{N} \left[\sum_{j=2}^{I} y_i x_i^T \beta_j - n_i log(1 + \sum_{j=2}^{I} exp(x_i^T \beta_j)) \right]$$

Taking the first and second derivatives, we have

$$\frac{\partial ln(\beta)}{\partial \beta_j} = \sum_{i=1}^{N} \left[\sum_{j=2}^{I} y_i x_i - n_i \frac{exp(x_i^T \beta_j) x_i}{1 + \sum_{j=2}^{I} exp(x_i^T \beta_j)} \right] = \sum_{i=1}^{N} [y_i x_i - n_i \pi_{ij} x_i]$$

$$\frac{\partial^2 ln(\beta)}{\partial \beta_j \, \partial \beta_j} = \sum_{i=1}^{N} -n_i \frac{exp(x_i^T \beta_j) x_i}{(1 + \sum_{j=2}^{I} exp(x_i^T \beta_j))^2} = \sum_{i=1}^{N} -n_i \pi_{ij} (1 - \pi_{ij}) x_i^{\otimes 2}$$

$$\frac{\partial^2 ln(\beta)}{\partial \beta_j \, \partial \beta_k} = \sum_{i=1}^{N} -n_i \frac{exp(x_i^T \beta_j) exp(x_i^T \beta_k) x_i x_k}{(1 + \sum_{j=2}^{I} exp(x_i^T \beta_j))^2} = \sum_{i=1}^{N} -n_i \pi_{ij} (\pi_{ik}) x_i^{\otimes 2}$$

If we compare two groups with covariates x_A and x_B for all odds ratios, then in the jth category, we obtain the odds ratio for comparing group A and B as

$$R_{Avs.B}(j) = \frac{P(Z = j|x_A) P(Z = 1|x_B)}{P(Z = 1|x_A) P(Z = j|x_B)} = exp[(x_A - x_B)^T \beta]$$

2 Practice

2.1 Proportional Odds Model