Paper Notes: Generalized R Squared (Also called Pseudo R Squared)

Yingbo Li

10/05/2020

Table of Contents

Generalized R^2 by Cox and Snell

Generalized R^2 by Nagelkerke

- Generalized R^2 for binary data

R^2 for normal linear regression

- R^2, also called coefficient of determination or multiple correlation coefficient, is defined for normal linear regression, as the proportion of variance "explained" by the regression model

$$R^2 = \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{1}$$

- Note that under the MLE, where $\hat{\sigma}^2 = \sum_i (y_i - \hat{y}_i)^2 / n$, the deviance (i.e., negative two times log likelihood) is

$$\begin{aligned} -2l\left(\hat{\beta}\right) &= -2 \log L(\hat{\beta}) \\ &= n \log(2\pi\hat{\sigma}^2) + \frac{\sum_i (y_i - \hat{y}_i)^2}{\hat{\sigma}^2} \\ &= n \left[\log\left(\frac{\sum_i (y_i - \hat{y}_i)^2}{n} \right) + \log(2\pi) + 1 \right] \end{aligned}$$

  – I list this derivation here to make clear that the following generalized R^2 contains (1) as a special case for normal linear regression

Generalized R^2, proposed by Cox and Snell [1989] (and also Magee [1990] and Maddala [1983])

- The genralized R^2 for more general models where
  1. the concept of residual variance cannot be easily define, and
  2. maximum likelihood is the criterion of fit, is

$$R^2 = 1 - \exp\left\{-\frac{2}{n}\left[l\left(\hat{\beta}\right) - l(\hat{0})\right]\right\} = 1 - \left[L(0)/L\left(\hat{\beta}\right)\right]^{2/n} \tag{2}$$

- Here, $L\left(\hat{\beta}\right)$ and $L(0)$ are the likelihood of the fitted and the null models, respectively.
- For normal linear regression, this generalized R^2 (2) becomes the classical R^2 (1)

Desirable properties of the generalized R^2, as in Eq (2)

1. Consistent with classical R^2
2. Consistent with maximum likelihood as an estimation method
3. Asymptotically independent of the sample size n
4. $1 - R^2$ has an interpretation as the propotion of unexplained "variation"
   - For example, if we have three nested models, from smallest to largest, M_1, M_2, and M_3, then we have

$$(1 - R^2_{3,1}) = (1 - R^2_{3,2})(1 - R^2_{2,1})$$

- For more desirable properties (7 in total), please check out the Nagelkerke[1991] paper

Generalized R^2, proposed by Nagelkerke [1991]

- An undesirable property: for discrete models, the maximum R^2 is always less than 1

$$\max(R^2) = 1 - L(0)^{2/n}$$

  - This is because the likelihood of discrete target variables are from pmf (rather than from pdf, as of continuous targets)

- A new definition of the generalized R^2

$$\bar{R}^2 = \frac{R^2}{\max(R^2)} = \frac{1 - \left[L(0)/L\left(\hat{\beta}\right)\right]^{2/n}}{1 - L(0)^{2/n}} \tag{3}$$

  - Majority of the desirable properties of (2), including the ones listed on the previous page, are still satisfied
  - Nagelkerke's general R^2 (3) seems to be a popular version. For example, the biostat textbook by Steyerberg uses this version

Generalized R^2 for binary data

- Denote the estimated binary probabilities as $\hat{p}_i$ for the fitted model, and $\bar{p}$ for the null model
- Cox and Snell R^2

$$R^2 = 1 - \left[L(0)/L\left(\hat{\beta}\right)\right]^{2/n} = 1 - \left[\prod_i \left(\frac{\bar{p}}{\hat{p}_i}\right)^{y_i} \left(\frac{1-\bar{p}}{1-\hat{p}_i}\right)^{1-y_i}\right]^{2/n}$$

- Nagelkerke R^2

$$\bar{R}^2 = \frac{1 - \left[L(0)/L\left(\hat{\beta}\right)\right]^{2/n}}{1 - L(0)^{2/n}} = \frac{1 - \left[\prod_i \left(\frac{\bar{p}}{\hat{p}_i}\right)^{y_i} \left(\frac{1-\bar{p}}{1-\hat{p}_i}\right)^{1-y_i}\right]^{2/n}}{1 - \left[\prod_i \bar{p}^{y_i} \left(1-\bar{p}\right)^{1-y_i}\right]^{2/n}}$$

References

- Nagelkerke, N. J. D. (1991). A Note on a General Definition of the Coefficient of Determination. Biometrika, 78(3), 691-692.
- A nice comparison of different versions of generalized R^2: https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-pseudo-r-squareds/
- Steyerberg, E. W. (2019). Clinical prediction models. Springer International Publishing.