Statistical Odds & Ends

What is deviance?

Posted on March 27, 2019

I encountered the term "deviance" in the first year of my PhD program and I've never fully understood what it means. This post tries to clarify my understanding of the concept of deviance and how it is used.

Let's say we have data $(\boldsymbol{x}_1, y_1), \dots, (\boldsymbol{x}_n, y_n)$, where $\boldsymbol{x}_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}$ for each i. In supervised learning, we are interested in constructing a model which uses the $\boldsymbol{x}_i$'s to predict the y_i's. If we use a generalized linear model (GLM) to model the relationship, ***deviance is a measure of goodness of fit: the smaller the deviance, the better the fit***.

The exact definition of deviance is as follows: for a particular GLM (denoted $\mathcal{M}$), let $L_\mathcal{M}$ denote the maximum achievable likelihood under this model. Let $L_\mathcal{S}$ denote the likelihood under the "saturated model". Then the deviance of the GLM is defined as

$$D_\mathcal{M} = -2\log\left(\frac{L_\mathcal{M}}{L_\mathcal{S}}\right) = -2(\log L_\mathcal{M} - \log L_\mathcal{S}).$$

What is the "saturated model"? According to Agresti in *Categorical Data Analysis*, this refers to "the most general model, having a separate parameter for each observation and the perfect fit". In other words, it is the model which perfectly fits the observed response. This model also has the maximum achievable log likelihood among all possible models within the GLM framework. (Why isn't the log likelihood of the saturated model always zero? See this answer, which shows that in Poisson GLMs, the saturated model doesn't always have a log likelihood of zero.)

If you are confused about what the saturated model is, do not fear! First, note that the term $2\log L_\mathcal{S}$ is the same across **all** models $\mathcal{M}$. Hence, ***you can think of the deviance of a model as twice the negative log likelihood plus a constant.*** Second, deviance is used to compare two different models, and we do so by subtracting one deviance statistic from another. Once we do this subtraction, the $2\log L_\mathcal{S}$ term gets canceled out, and hence is fit. (For more precise statements, see Reference 2.) The test statistic is the difference of deviances:

$$\begin{aligned} D &= D_{\mathcal{M}_0} - D_{\mathcal{M}_1} \\ &= -2(\log L_{\mathcal{M}_0} - \log L_{\mathcal{S}}) + 2(\log L_{\mathcal{M}_1} - \log L_{\mathcal{S}}) \\ &= -2(\log L_{\mathcal{M}_0} - \log L_{\mathcal{M}_1}). \end{aligned}$$

Assume that $\mathcal{M}_0$ and $\mathcal{M}_1$ have p_0 and p_1 parameters respectively. By Wilk's theorem, under the null hypothesis, as the sample size $n \to \infty$, the statistic has asymptotic distribution $D \sim \chi^2_{p_1 - p_0}$.

The last thing to note about deviance is that ***it can be viewed as a generalization of residual sum of squares in linear models.*** Assuming that the true model is $Y_i = \sum_{j=1}^p \beta_j X_{ij} + \epsilon_i$ with $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$, under the linear model we have

$$\begin{aligned} L_{\mathcal{M}} &= \max_\beta \prod_{i=1}^n \mathbb{P}(Y_i = y_i \mid x_{i1}, \dots, x_{ip}) \\ &= \max_\beta \prod_{i=1}^n \mathbb{P}\left(\epsilon_i = y_i - \sum_{j=1}^p \beta_j X_{ij}\right) \\ &= \max_\beta \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(y_i - \sum_{j=1}^p \beta_j X_{ij}\right)^2}{2\sigma^2}\right), \\ \log L_{\mathcal{M}} &= -\frac{n}{2}\log(2\pi\sigma^2) + \max_\beta \sum_{i=1}^n \left(-\frac{\left(y_i - \sum_{j=1}^p \beta_j X_{ij}\right)^2}{2\sigma^2}\right) \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \min_\beta \sum_{i=1}^n \frac{\left(y_i - \sum_{j=1}^p \beta_j X_{ij}\right)^2}{2\sigma^2}, \end{aligned}$$

and

$$\begin{aligned} L_{\mathcal{S}} &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - y_i)^2}{2\sigma^2}\right) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}}, \\ \log L_{\mathcal{S}} &= -\frac{n}{2}\log(2\pi\sigma^2). \end{aligned}$$

Putting them together,

$$\begin{aligned} D_{\mathcal{M}} &= -2(\log L_{\mathcal{M}} - \log L_{\mathcal{S}}) \\ &= -2\left[-\min_\beta \sum_{i=1}^n \frac{\left(y_i - \sum_{j=1}^p \beta_j X_{ij}\right)^2}{2\sigma^2}\right] \end{aligned}$$

References:

1. Agresti, A. Categorical Data Analysis 3rd ed. (p115-6).
2. http://www.maths.qmul.ac.uk/~bb/MS_Lectures_23and24.pdf (Section 2.6.3, Theorem 2.9).

SHARE THIS:

Twitter Facebook

Like

Be the first to like this.

RELATED

Quantile regression in R What is the Durbin- Change of basis formula

models, **hypothesis testing** by **kjytay**. Bookmark the **permalink**

[https://statisticaloddsandends.wordpress.com/2019/03/27/what-is-deviance/] .