UNIVERSITY OF EDINBURGH
Business School

Predictive Analytics and Modelling of Data

CMSE11428 (2020-2021)

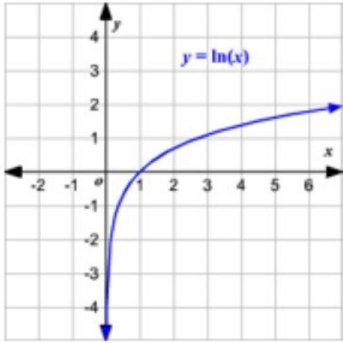
Dr Xuefei Lu
The University of Edinburgh Business School

Logistic Regression

To start

Now you have the opportunity to practise switching between probabilities, odds and log-odds.

Please pause the video and try to answer the following questions:



Q1. What are the odds for the probability of an event of 0.5?

Q2. What are the log-odds (natural logarithm) for the probability of an event of 0.5?

Q3. What are the log-odds (natural logarithm) for the probability of an event of 0?

Q4. What are the log-odds (natural logarithm) for the probability of an event of 1?

To start

Now you have the opportunity to practise switching between probabilities, odds and log-odds.

Please pause the video and try to answer the following questions:

Q1. What are the odds for the probability of an event of 0.5?

1

Q2. What are the log-odds (natural logarithm) for the probability of an event of 0.5?

0

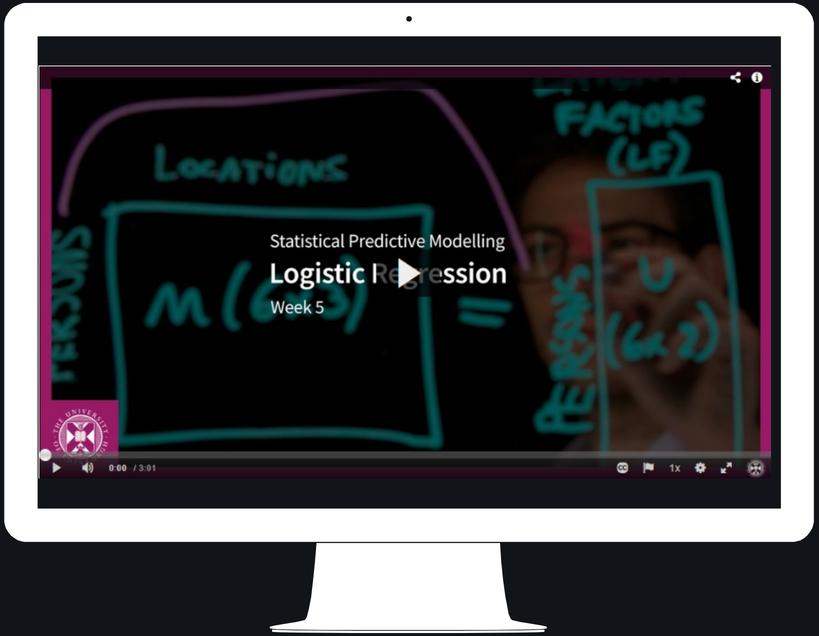
Q3. What are log-odds (natural logarithm) for the probability of an event of 0?

Cannot be defined since logarithm of 0

Q4. What are log-odds (natural logarithm) for the probability of an event of 1?

Cannot be defined since division of 0

Logistic Regression



Please watch the following video

https://media.ed.ac.uk/media/Logistic+Regression/0_y4d65hdj/122596071

Logistic regression

$$E(Y) = P(Y = 1 | x_1, x_2, \ldots x_p)$$

$$E(Y) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}}{1 + e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}}$$

logistic form

$$ln\left[\frac{P(Y=1)}{1-P(Y=1)}\right] = \boxed{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p}$$

logit transformation of P

- How do logistic and linear regression differ?
- Please study the following file:

12 - How do logistic and linear regression differ.ipynb + german.csv

Activity: Logistic vs linear regression

Maximum likelihood estimation (MLE)

To understand the principle of maximum likelihood, we use **Binomial model** for illustration. Binomial model looks at a binary variable (whether the event occurs or not).

In our Credit Miracle example, the event is whether a borrower defaults. For the sake of argument, if we assume that the probability of default is the same for all borrowers, that is, it does not depend on any of their characteristics (which is, of course, totally unrealistic), then we can model the probability of default with a binomial distribution.

Table: Customers with defaulting B (Bad) and non-defaulting with G (Good)

Customer	1	2	3	4	5	6	7	8	9	10
Status	G	G	G	G	B	G	G	G	B	G

Maximum likelihood estimation (MLE)

What is the true probability of being Bad in the customer population of portfolio of loans?

We want to estimate the parameter π, which denotes the probability of an event in a binomial model:

$$P(r) = \binom{n}{r} \pi^r (1-\pi)^{n-r}$$

where r is the number of events in n trials. Note that this formula is used to calculate the probability of having a certain number of events in a sample of size n, and the π is known or given.

For $n = 10$ and different values of π and r, we have:

r, π	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.349	0.107	0.028	0.006	0.001	0	0	0	0
1	0.387	0.268	0.121	0.04	0.01	0.002	0	0	0
2	0.194	0.302	0.233	0.121	0.044	0.011	0.001	0	0
3	0.057	0.201	0.267	0.215	0.117	0.042	0.009	0.001	0
4	0.011	0.088	0.2	0.251	0.205	0.111	0.037	0.006	0
5	0.001	0.026	0.103	0.201	0.246	0.201	0.103	0.026	0.001
6	0	0.006	0.037	0.111	0.205	0.251	0.2	0.088	0.011
7	0	0.001	0.009	0.042	0.117	0.215	0.267	0.201	0.057
8	0	0	0.001	0.011	0.044	0.121	0.233	0.302	0.194
9	0	0	0	0.002	0.01	0.04	0.121	0.268	0.387
10	0	0	0	0	0.001	0.006	0.028	0.107	0.349

Maximum likelihood estimation (MLE)

But we do not know the true probability of default π ➔ to estimate using our sample.

Maximum likelihood estimation (MLE) results in unknown model parameters that are most likely to generate the patterns we observe in the sample.

The likelihood function measures the probability of observing a particular set of values in the sample:

$$L(parameter|data) = P_1 \times P_2 \times \cdots \times P_n = \prod_{i=1}^{n} P_i$$

P_i: The prob of the i-th experiment

Taking our sample of 10 customers we can write (p_i: prob of being bad at i-th experiment):

$$\boxed{\begin{aligned} L(\pi|GGGGBGGGBG) &= (1-p_1)(1-p_2)(1-p_3)(1-p_4)p_5(1-p_6)(1-p_7)(1-p_8)p_9(1-p_{10}) \\ &= \pi^2(1-\pi)^8 \end{aligned}}$$

The probability distribution function looks the same, but the interpretation is quite different:

$$P(GGGGBGGGBG|\pi) = (1-p_1)(1-p_2)(1-p_3)(1-p_4)p_5(1-p_6)(1-p_7)(1-p_8)p_9(1-p_{10})$$
$$= \pi^2(1-\pi)^8$$

With the latter case, we are interested in obtaining the probability of observing our data given that parameter equals a certain value. With the likelihood it is the other way around.

Maximum likelihood estimation (MLE)

The likelihood function for the binomial model is the same as probability distribution function:

$$L(\pi|r,n) = \binom{n}{r}\pi^r(1-\pi)^{n-r}$$

r, π	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0	0.349	0.107	0.028	0.006	0.001	0	0	0	0
1	0.387	0.268	0.121	0.04	0.01	0.002	0	0	0
2	0.194	0.302	0.233	0.121	0.044	0.011	0.001	0	0
3	0.057	0.201	0.267	0.215	0.117	0.042	0.009	0.001	0
4	0.011	0.088	0.2	0.251	0.205	0.111	0.037	0.006	0
5	0.001	0.026	0.103	0.201	0.246	0.201	0.103	0.026	0.001
6	0	0.006	0.037	0.111	0.205	0.251	0.2	0.088	0.011
7	0	0.001	0.009	0.042	0.117	0.215	0.267	0.201	0.057
8	0	0	0.001	0.011	0.044	0.121	0.233	0.302	0.194
9	0	0	0	0.002	0.01	0.04	0.121	0.268	0.387
10	0	0	0	0	0.001	0.006	0.028	0.107	0.349

Maximum likelihood estimation (MLE)

$$L(\pi|r,n) = \binom{n}{r}\pi^r(1-\pi)^{n-r}$$

MLE consists of finding the maximum of the function above or of **log-likelihood function**. We take logarithms because it is easier to work with an additive model rather than a multiplicative one.

$$\ln\big(L(\pi)\big) = \ln\left(\binom{n}{r}\right) + r\ln(\pi) + (n-r)\ln(1-\pi)$$

Differentiating the log-likelihood function for the binomial model with respect to π gives:

$$\frac{\mathrm{d}\ln\big(L(\pi)\big)}{\mathrm{d}\,\pi} = 0 + \frac{r}{\pi} + (n-r)\frac{1}{1-\pi}(-1) = \frac{r}{\pi} - \frac{n-r}{1-\pi}$$

Setting the derivative to 0 gives $\pi = \frac{r}{n}$.

Therefore, sample proportion is unbiased MLE estimator of π.

- Read and study the following file:
- 13 - Activity 7 - Binomial maximum likelihood.ipynb + credit_regress.csv

Activity: Binomial maximum likelihood

Logistic regression and MLE

For logistic regression, the likelihood of the pattern of occurrences [Y = 1] and non-occurrences [Y = 0] of an event is given by the product of the corresponding probabilities for each observation i in a sample of the size n:

$$L = \prod_{i=1}^{n} P_i(X)^{Y_i}(1 - P_i(X))^{1-Y_i}$$

If the observation is an event $[Y_i = 1]$, it contributes to the probability of experiencing an event $P_i(X)$ in the expression above, since $[1 - Y_i] = 0$, if the observation is a non-event $[Y_i = 0]$, then it contributes to the probability of not experiencing the event $1 - P_i(X)$.

We would like to estimate the coefficients, β's:

$$P_i(X) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}}{1+e^{\beta_0\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}} = \frac{e^{\boldsymbol{\beta X}}}{1+e^{\boldsymbol{\beta X}}} \quad \Rightarrow \quad L(\boldsymbol{\beta}) = \prod_{i=1}^{n}\left(\frac{e^{\boldsymbol{\beta X}}}{1+e^{\boldsymbol{\beta X}}}\right)^{Y_i}\left(\frac{1}{1+e^{\boldsymbol{\beta X}}}\right)^{1-Y_i}$$

Logistic regression and MLE

$$P_i(X) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}}{1+e^{\beta_0 \beta_1 X_1+\beta_2 X_2+\ldots+\beta_p X_p}} = \frac{e^{\boldsymbol{\beta X}}}{1+e^{\boldsymbol{\beta X}}}$$

$$\Rightarrow$$

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{e^{\boldsymbol{\beta X}}}{1+e^{\boldsymbol{\beta X}}}\right)^{Y_i} \left(\frac{1}{1+e^{\boldsymbol{\beta X}}}\right)^{1-Y_i}$$

$$\ell = ln(L(\boldsymbol{\beta})) = \sum_{i=1}^{n} [Y_i ln(P_i|\boldsymbol{\beta}\mathbf{X}) + (1-Y_i) ln(1-P_i|\boldsymbol{\beta}\mathbf{X})$$

Note that the likelihood function ranges between 0 and 1 → the log-likelihood is a negative number

The objective is to find the values of β to maximize the log-likelihood, which the higher value is better. For example, a log-likelihood value of -3 is better than -7.

However, in contrast to the binomial model, there is no simple solution.

Logistic regression and MLE

Let's take an example using the same dataset as before. We will use just two variables:

1. Telephone (X_1; 1- telephone known; 0 - otherwise)
2. Age (X_2) to model

Y: The credit default

The model to be estimated is:

$$ln\frac{P(Y=1)}{1-P(Y=1)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Model 1: $\beta_0 = 0.2, \beta_1 = 0.1, \beta_2 = 0.05$

Model 2: $\beta_0 = 0.2, \beta_1 = -0.1, \beta_2 = -0.05$

Account no.#	Y, Status/Default	Telephone	Age
1	0	1	27
2	0	1	55
3	0	1	60
4	0	0	40
5	1	0	22
6	0	1	30
7	0	1	35
8	0	1	42
9	1	0	25
10	0	0	65

Logistic regression and MLE

$$ln\frac{P(Y=1)}{1-P(Y=1)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Model 1: $\beta_0 = 0.2, \beta_1 = 0.1, \beta_2 = 0.05$								
No.	$\beta_1 X_1$	$\beta_2 X_2$	Logit	Odds	P	P *Y	(1-P) * (1-Y)	Contribution to L
1	2	3	4	5	6	7	8	9
1	0.1	1.35	1.65	5.2070	0.8389	0.0000	0.1611	0.1611
2	0.1	2.75	3.05	21.1153	0.9548	0.0000	0.0452	0.0452
3	0.1	3	3.3	27.1126	0.9644	0.0000	0.0356	0.0356
4	0	2	2.2	9.0250	0.9002	0.0000	0.0998	0.0998
5	0	1.1	1.3	3.6693	0.7858	0.7858	0.0000	0.7858
6	0.1	1.5	1.8	6.0496	0.8581	0.0000	0.1419	0.1419
7	0.1	1.75	2.05	7.7679	0.8859	0.0000	0.1141	0.1141
8	0.1	2.1	2.4	11.0232	0.9168	0.0000	0.0832	0.0832
9	0	1.25	1.45	4.2631	0.8100	0.8100	0.0000	0.8100
10	0	3.25	3.45	31.5004	0.9692	0.0000	0.0308	0.0308
Likelihood, L								6.81215E-10
Log-Likelihood, LnL								-21.1071435

Apply the same procedure for Model 2:

Likelihood, L	0.025041
Log-Likelihood, LnL	-3.68724

What conclusions can we make?

Logistic regression and MLE

$$ln\frac{P(Y=1)}{1-P(Y=1)} = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

The parameters considered in this example are arbitrary. Computer packages rely on algorithms that search for the best parameters in an efficient way by successively improving the coefficient estimates that lead larger log-likelihoods. The search stops when any further increases in the log-likelihood function are very small and it is not beneficial to continue with the procedure.

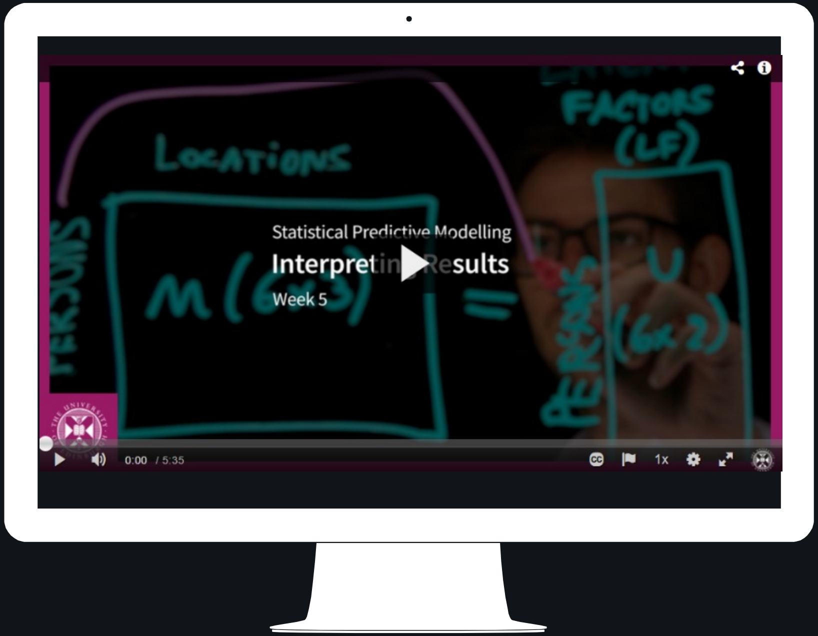
5	0	1.1	1.3	3.6693	0.7858	0.7858	0.0000	0.7858
6	0.1	1.5	1.8	6.0496	0.8581	0.0000	0.1419	0.1419
7	0.1	1.75	2.05	7.7679	0.8859	0.0000	0.1141	0.1141
8	0.1	2.1	2.4	11.0232	0.9168	0.0000	0.0832	0.0832
9	0	1.25	1.45	4.2631	0.8100	0.8100	0.0000	0.8100
10	0	3.25	3.45	31.5004	0.9692	0.0000	0.0308	0.0308
Likelihood, L								6.81215E-10
Log-Likelihood, LnL								-21.1071435

Apply the same procedure for Model 2:

Likelihood, L	0.025041
Log-Likelihood, LnL	-3.68724

Model 2 has higher values for both Likelihood and Log-Likelihood, so it is a better model.

Interpreting Results



Please watch the following video

https://media.ed.ac.uk/media/Interpreting+Results/1_4jzylf8z/122596071

Interpreting Results

Estimated model: Y ∈ {1 = Bad/Default, 0 = Good}

$$ln\left[\frac{P(Y=1)}{1-P(Y=1)}\right] = 0.2 \boxed{-0.49} X_1 + 0.35X_2 - 0.02X_3$$

$$\frac{P(Y=1)}{1-P(Y=1)} = e^{0.2} \boxed{e^{-0.49}} {}^{X_1} e^{0.35X_2} e^{-0.02X_3}$$

Homeowner 0.61<1

- You will now practise obtaining parameter estimates for logistic regression in Python and practise interpreting them. You will also look at how to assess predictive accuracy.
- 14 - Activity 8 - Interpreting logistic regression.ipynb + credit_regress.csv

Activity: Logistic Regression

UNIVERSITY OF EDINBURGH
Business School