Week 4: **Predicting the Academy Award Winners (Oscars)**

Tool: Multinomial Logit

The Analytics Edge: Each year the Oscars awards create huge interest in the movie industry, fans and box-office. There is tremendous consequences of taking an Oscar home, be it future earnings or fame. Using a simple multinomial logit model with data available before the awards are given such as other awards (Golden Globe), other nomination in the Oscars, it is possible to obtain simple models that can predict the winners in four categories - Best Picture, Best Director, Best Leading Actor and Best Leading Actress. This provides an alternate prediction model to expert opinions.

Binary Choice

Response: 0 or 1 (No or Yes)

$$y^* = \beta_0 + \sum_{j=1}^{p} \beta_j x_j + \epsilon$$

Here y^* is an unobserved latent variable,x_j are attributes of the choice and ϵ is the error term.

$$y = \begin{cases} 1 \text{ if } y^* \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

y is observed response.

$$\begin{aligned}
\therefore P(y=1|x) &= P(y^* \geq 0|x) \\
&= P(\beta_0 + \sum_{j=1}^{p} \beta_j x_j + \epsilon \geq 0|x) \\
&= P(\epsilon \geq -(\beta_0 + \sum_{j=1}^{p} \beta_j x_j)|x) \\
&= P(\epsilon \leq \beta_0 + \sum_{j=1}^{p} \beta_j x_j|x) \text{ (Assume } \epsilon \text{ is symmetric)} \\
&= F(\beta_0 + \beta' x) \text{ (where } F(.) \text{ is the cumulative distribution function of } \epsilon)
\end{aligned}$$

Note that with $F(z) = \frac{e^z}{1+e^z}$ we get,

$$P(y=1|x) = \frac{e^{\beta_0 + \beta' x}}{1 + e^{\beta_0 + \beta' x}} \text{ (Logit model for binary choice)}$$

Multinomial Choice

Given $k = 1, \ldots, K$ choices (alternatives), $i = 1, \ldots, n$ consumers (observations), the utility of consumer i for alternative k is,

$$U_{ik} = \beta' x_{ik} + \epsilon_{ik}$$

Here β is the weight given to the observable information x_{ik} that includes aspects specific to the individual i and choice (alternative) k. The term ϵ_{ij} captures the noise term that models aspect of choice making not captured by attributes included in x_{ik}.

For example x_{ik} might include demographic information, information such as price of a product.

$$\underbrace{P(y_i = k)}_{\text{Probability that the } i^{th} \text{ individual chooses k}} = P(U_{ik} > U_{il} \forall l \neq k)$$

In some cases, you might have alternate specific constants ASC_k for each alternative.

$$U_{ik} = ASC_k + \beta' x_{ik} + \epsilon_{ik}$$

Assuming that the ϵ_{ij} are independent and identically distributed with Gumbel (type 1 extreme value) distributions,

$$F(\epsilon_{ij}) = e^{-e^{-\epsilon_{ij}}}$$

It was shown by McFadden (1974) that,

$$P(y_i = k) = \frac{e^{\beta' x_{ik}}}{\sum_{l=1}^{k} e^{\beta' x_{il}}}$$

This model is known as the conditional logit model (also referred to as multinomial logit often).

Special Case of the model (Multinomial Logistic Regression):

Assuming only individual specific data and suppose we want to categorize the individual into one of several categories it gives rise to a multinomial logistic regression model.

$$P(y_i = k | x_i) = \frac{e^{\beta' x_i}}{\sum_{l=1}^{k} e^{\beta_l' x_i}}$$

Here β_k is a vector of weights corresponding to category k.

Properties of Multinomial Logit Model

Independence of irrelevant alternatives

$$\frac{P(y_i = k)}{P(y_i = l)} = \left(\frac{e^{\beta' x_{ik}}}{\sum_{t=1}^{K} e^{\beta' x_{it}}} \right) \frac{1}{\frac{e^{\beta' x_{il}}}{\sum_{t=1}^{K} e^{\beta' x_{it}}}} = e^{\beta'(x_{ik} - x_{il})}$$

$$\log \frac{P(y_i = k)}{P(y_i = l)} = \beta'(x_{ik} - x_{il})$$

This property of the MNL model is known as the Independence of Irrelevant Alternatives wherein a choice set consisting of two alternatives k and l, adding in a third alternative does not change the ratio of $P(y_i = k)/P(y_i = l)$. Namely the new alternate gains share proportionately from the choice share of existing alternatives in the set.

Example: Blue bus/ Red bus.

Alternatives: Car, Red Bus

$P(C) = P(R) = 0.5$ (Suppose a commuter chooses between a car and a red bus with equal probability)

Assume a new blue bus is added and reasonably commuters don't care about the color of the bus.

Alternatives: Car, Red Bus, Blue Bus

$P(C) = 0.5, P(R) = P(B) = 0.25$: Reasonable prediction

$P(C) = P(R) = P(B) = 0.33$: MNL prediction

This happens because blue bus and red bus are perfect substitutes here and not captured by MNL model.

Note this might be a smaller issue in the case of Oscars predictions since it is not clear if nominees are likely to be close substitutes except if an individual receives multiple nominations in a category in the same year or nominated movies might be from similar genres making them closer substitutes.

Maximum Likelihood Estimation

Given the observations:

x_{ik} (attributes of individual i and alternative k).

$$z_{ik} = \begin{cases} 1 & \text{if individual i chooses k} \\ 0 & \text{otherwise. (Note } y_i = k \iff z_{ik} = 1) \end{cases}$$

The problem of estimating the β (weights) that maximises the likelihood of the observations is given as:

$$\max_{\beta} \underbrace{L(\beta)}_{\text{Likelihood given } \beta} = \max_{\beta} \prod_{i=1}^{n} \prod_{k=1}^{K} P(y_i = k)^{z_{ik}}$$

Taking logarithms, this problem can be reformulated as maximizing the log-likelihood.

$$\max_{\beta} \underbrace{LL(\beta)} = \max_{\beta} \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} \log(P(y_i = k)) = \max_{\beta} \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} \log \left(\frac{e^{\beta' x_{ik}}}{\sum\limits_{l=1}^{k} e^{\beta' x_{il}}} \right)$$

This problem can be efficiently solved as the objective function is concave. This is one of the few formulas for choice probabilities where the objective function is known to be concave.

Testing quality of fit

1. AIC (Akaike Information Criterion)

   $$AIC = -2LL + 2(p+1)$$

   where LL: Log-likelihood, p : no. of parameters.

   Smaller the AIC, the better.

2. Likelihood ratio index (McFadden's index)

   $$\rho = 1 - \frac{LL(\hat{\beta})}{LL(0)}$$

   where $\hat{\beta}$ is the estimated value of the parameters. LL(0) refers to the log-likelihood when all the parameters are set to 0 (no model).

   This compares the quality and fit of the model in comparison to a model in which all the parameters are equal to 0. The likelihood ratio index ranges from 0 (estimated model is no better than zero parameters) to 1 (estimated model perfectly predicts the choice observed).

3. Percent correctly predicted

   This identifies the alternative with the highest probability for each individual observation and determining whether or not this was what the actual choice was.