Stats and Probability Information

April 11, 2014

Continuous and marginal distributions

The marginal distribution of x in a two-variable distribution is equal to the sum of the joint distribution over y.

$$Pr(X = x) = \sum_y Pr(X = x, Y = y) = \sum_y Pr(X = x|Y = y)Pr(Y = y) \tag{1}$$

From Wikipedia

For the continuous case

$$p_X(x) = \int_y p_{X,Y}(x, y)dy = \int_y p_{X|Y}(x|y)p_Y(y)dy \tag{2}$$

There are three related distributions: the marginal, the joint and the conditional.

1 Mixture Model

This is a probabilistic model that relates some random variables to some other variables. The model has sub-populations. The properties of the sub-population are different from those of the parent. The sub-populations may not be observable. For example, the distribution of returns may be different in different sub-population or regime.

A *mixture distribution* is the probability distribution of a random variable whose values are derived from an underlying set of random variables. The *mixture components* are individual distributions with *mixture weights*. Even in cases where the mixture comonents have a normal distribution, the mixture distribution is likely to be non-normal. Mixture models are used to understand the sub-population when there is only access to the information about the pooled population.

The mixture model will be comprised of N random varibles distributed according to K components, with each component belonging to the same distribution. The k mixture weights sum to one. Each component will have parameters (mean and variance in the case of normal distribution).

The method will try to estimate the all the parameters of the model from the data. The underlying data is known (x_i); the number of mixture components is set (K); the parameters of the distribution of each mixture component ($\theta_{i=1...K}$); mixture weight ($\Phi_{i=1...K}$); $\boldsymbol{\Phi}$ K-dimensional vector summing to 1; $F(x|\theta)$ probability distribution of observations parameterised on θ; α shared hyperparameter for component weights; β shared hyperparameter for mixture weights; $H(\theta|\alpha)$ prior probability distribution of component parameters;

2 Adjusted R squared

Adjusted R squared applied a penalty to the basic R squard to account for additional variables. The equartion is

$$R_A^2 = 1 - \left[\frac{(n-1)}{(n-k)}\right]\left[1 - R^2\right] \tag{3}$$

Adding a regressor to the equation will increase (reduce)) the $R_A^{"}$ when the absolute value of the t-statistic is greater (less) than one. Adding a group of regressors to the model will reduce (increase) the $R_A^{"}$ when the absolute value of the F-statistic is greater than one.

Proof http://davegiles.blogspot.com/2014/04/proof-of-result-about-adjusted.html