Multivariate Statistical Analysis, Exercise 10, Fall 2011, Prof.S.K.Jeng

December 23, 2011 TA: H.C. Cheng

(Likelihood Ratio Tests for Regression)

There are many techniques for hypothesis testing problem. Likelihood Ratio Test (LRT) is the popular one for its asymptotic properties. In this exercise, we will show the nice properties of likelihood ratio test and see how it be applied in the problems of linear model.

Suppose that the distribution of $(x_i)_{i=1}^n$, $x_i \in \mathbb{R}^p$, dependes on a parameter vector θ. We will consider two hypotheses:

$$H_0 : \quad \theta \in \Omega_0$$
$$H_1 : \quad \theta \in \Omega_1.$$

The null hypothesis H_0 corresponds to the "reduced model" and H_1 to the "full model". This notation was already used before. Define $L_j^* = \max_{\theta \in \Omega_j} L(X;\theta)$, the maxima of the likelihood for each of the hypotheses. Consider the likelihood ratio (LR)

$$\Lambda(X) = \frac{L_1^*}{L_0^*}.$$

One tends to favor H_0 if the LR is high and H_1 if the LR is low. A likelihood ratio test has the rejection region

$$R = \{X : \quad \Lambda(X) > c\},$$

where c is deterined so that $\max \theta \in \Omega_0 \Pr_\theta(X \in R) = \alpha$. The difficulty here is to express c as a function of α, because $\Lambda(X)$ might be a complicated function of X. Instead of Ω we may equivalently use the log-likelihood

$$2 \log \Lambda = 2(l_1^* - l_0^*).$$

In this case the rejection region will be $R = \{X : 2 \log \Lambda(X) > k\}$. What is the distribution of Λ or of $2 \log \Lambda$ from which we need to compute c or k? Let me see the theorem.

Theorem If $\Omega_1 \subset \mathbb{R}^q$ is a q-dimensional space and if $\Omega_0 \subset \Omega_1$ is an r-dimensional subspace, then under some regularity conditions

$$\forall \theta \in \Omega_0 : \quad 2 \log \Lambda \to \chi^2_{q-r} \text{ in distribution as } n \to \infty.$$

An asymptotic rejection region can now be given by simply computing the $1-\alpha$ quantile $k = \chi^2_{1-\alpha;q-r}$. The LRT rejection region is therefore

$$R = \{X : \quad 2\log\Lambda(X) > \chi^2_{1-\alpha;q-r}\}.$$

The above theorem is very helpful that it gives a general way of building rejection regions in many problems. Unfortunately, it is only an asymptotic result, meaning that the approximation becomes better when the sample size n increases. The question is "how large should n be?". There is no definite rule. It is data dependent. For more detailed results, please see Asymptotic Statistics and the part of Central Limit Theorem in the Probability Theory.

Now we consider the linear regression model $y_i = \beta^T x_i + \epsilon_i$ for $i = 1, \dots, n$, where ϵ_i is i.i.d. and $N(0, \sigma^2)$ and where $x_i \in \mathbb{R}^p$. Here $\theta = (\beta^T, \sigma)$ is a $(p+1)$=dimensional parameter vector. Denote $y = [y_1, \dots, y_n]^T$, $X = [x_1, \dots, x_n]^T$. then

$$L(y, X; \theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{1}{2\sigma^2}(y_i - \beta^T x_i)^2$$

and

$$\begin{aligned} l(y, X; \theta) &= \log\Big(\frac{1}{(2\pi)^{n/2}\sigma^n}\Big) - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \beta^T x_i)^2 \\ &= -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta). \end{aligned}$$

Differentiating with respect to the parameters yields

$$\begin{aligned} \frac{\partial}{\partial\beta} l &= -\frac{1}{2\sigma^2}(2X^TX\beta - 2X^Ty) \\ \frac{\partial}{\partial\sigma} l &= \frac{n}{\sigma} + \frac{1}{\sigma^3}\{(y - X\beta)^T(y - X\beta)\}. \end{aligned}$$

For the first equation we get

$$X^TX\hat\beta = X^Ty \Rightarrow \hat\beta = (X^TX)^{-1}X^Ty.$$

Plugging $\hat\beta$ into the second equation gives

$$\frac{n}{\hat\sigma} = \frac{1}{\hat\sigma^3}(y - X\hat\beta)^T(y - X\hat\beta) \Rightarrow \hat\sigma^2 = \frac{1}{n}\|y - X\hat\beta\|^2.$$

Now we consider a test problem that $y_1, \dots, y_n$ are independent random vectors with $y_i \sim N_1(\beta^T x_i, \sigma^2)$, $x_i \in \mathbb{R}^(p+1)$.

$$H_0 : \beta = \beta_0, \quad \sigma^2 \text{ unknown versus } H_1 : \text{no constraints}.$$

Under H_0 we have $\beta = \beta_0$, $\hat{\sigma}_0^2 = \frac{1}{n}\|y - X\beta_0\|^2$ and under H_1 we have $\hat{\beta} = (X^T X)^{-1} X^T y$, $\hat{\sigma}^2 = \frac{1}{n}\|y - X\beta\|^2$. Hence

$$\begin{aligned} 2\log\Lambda &= 2(l_1^* - l_0^*) \\ &= n\log\Big(\frac{\|y - X\beta_0\|^2}{\|y - X\hat{\beta}\|^2}\Big) \\ &\to \chi^2_{p+1} \text{ in distribution.} \end{aligned}$$

Then from the definition of F-distribution:

$$F = \frac{\{SS(\text{reduced}) - SS(\text{full})/\{df(r) - df(f)\}}{SS(\text{full})/df(f)}.$$

we get

$$F = \frac{(n-p-1)}{p+1}\Big(\frac{\|y - X\beta_0\|^2}{\|y - X\hat{\beta}\|^2} - 1\Big) \sim F_{p+1,n-p-1}.$$

Now with the data from Exercise 8:

	Sales	Price	Advert.	Ass. Hours
1	230	125	200	109
2	181	99	55	107
3	165	97	105	98
4	150	115	85	71
5	97	120	0	82
6	192	100	150	103
7	181	80	85	111
8	189	90	120	93
9	172	95	110	86
10	170	125	130	78

we know that $\hat{\beta} = [65.6696, -0.2158, 0.4852, 0.8437]^T$. We can see that the slope of the regression curve is rather small. Hence we might ask if $\beta_0 = [65.67, 0, 0, 0]$. With significance level $\alpha = 0.05$, please use the LR-test and F-test to see if we will reject this hypothesis.