Confidence set for parameter vector in linear regression

This questions is in reference to equation 3.15 in the book *Elements of Statistical Learning* by Tibshirani and coll.

I do understand the individual beta confidence interval estimation as provided in equation 3.14, but equation 3.15 just bowls me:

$$\beta|(\hat{\beta}-\beta)^T X^T X(\hat{\beta}-\beta) \leq \sigma^2\chi^2$$

What is the idea being expressed here? What is a confidence set? Can we not estimate the intervals of all the betas as per equation 3.14?

regression

edited Nov 14 '11 at 12:32
mpiktas
24.5k 4 48 102

asked Nov 13 '11 at 20:11
bgbgh
11 1

3 Please make the question self-contained. I do not have access to the book. – varty Nov 13 '11 at 21:13

1 @varty: The book is freely available (legally) at the link provided by the OP. That said, I still agree the question should be made self-contained if for no other reason than that future versions, or even printings, of the text may have different equation numbering.) – cardinal ♦ Nov 13 '11 at 21:31

Sorry this is the eqn I am referring to – bgbgh Nov 13 '11 at 22:47

$$\beta|(\hat{\beta}-\beta)^T X^T X(\hat{\beta}-\beta) \leq \sigma^2\chi^2$$

– bgbgh Nov 13 '11 at 22:48

1 @varty: I agree. My intent was not to be argumentative, but simply to (kindly) point out that you *did* have access to the book, in case you were interested. – cardinal ♦ Nov 14 '11 at 0:24

2 Answers

To make things clearer recall that

$$\hat{\beta} \sim N(\beta, \hat{\sigma}^2(X^T X)^{-1}),$$

When you isolate β_j you get that

$$\hat{\beta}_j - \beta_j \sim N(0, \sigma^2 v_j)$$

where v_j are the diagonal elements of $X^T X$. We can write this alternatively as

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{v_j}} \sim N(0, \sigma),$$

which is the same as

$$\left(\frac{\hat{\beta}_j - \beta_j}{\sqrt{v_j}}\right)^2 = (\hat{\beta}_j - \beta_j)(v_j)^{-1}(\hat{\beta}_j - \beta_j) \sim \sigma\chi^2_1.$$

Note that those β_j that satisfy the condition

$$\left(\frac{\hat{\beta}_j - \beta_j}{\sqrt{v_j}}\right)^2 \leq \sigma^2\chi^2_{1,1-\alpha}$$

fall in the confidence interval described in the equation 3.14. Hence the confidence interval is the set in real line.

Now similarly we get

$$(X^T X)^{1/2}(\hat{\beta} - \beta) \sim N(0, \sigma^2 I),$$

so

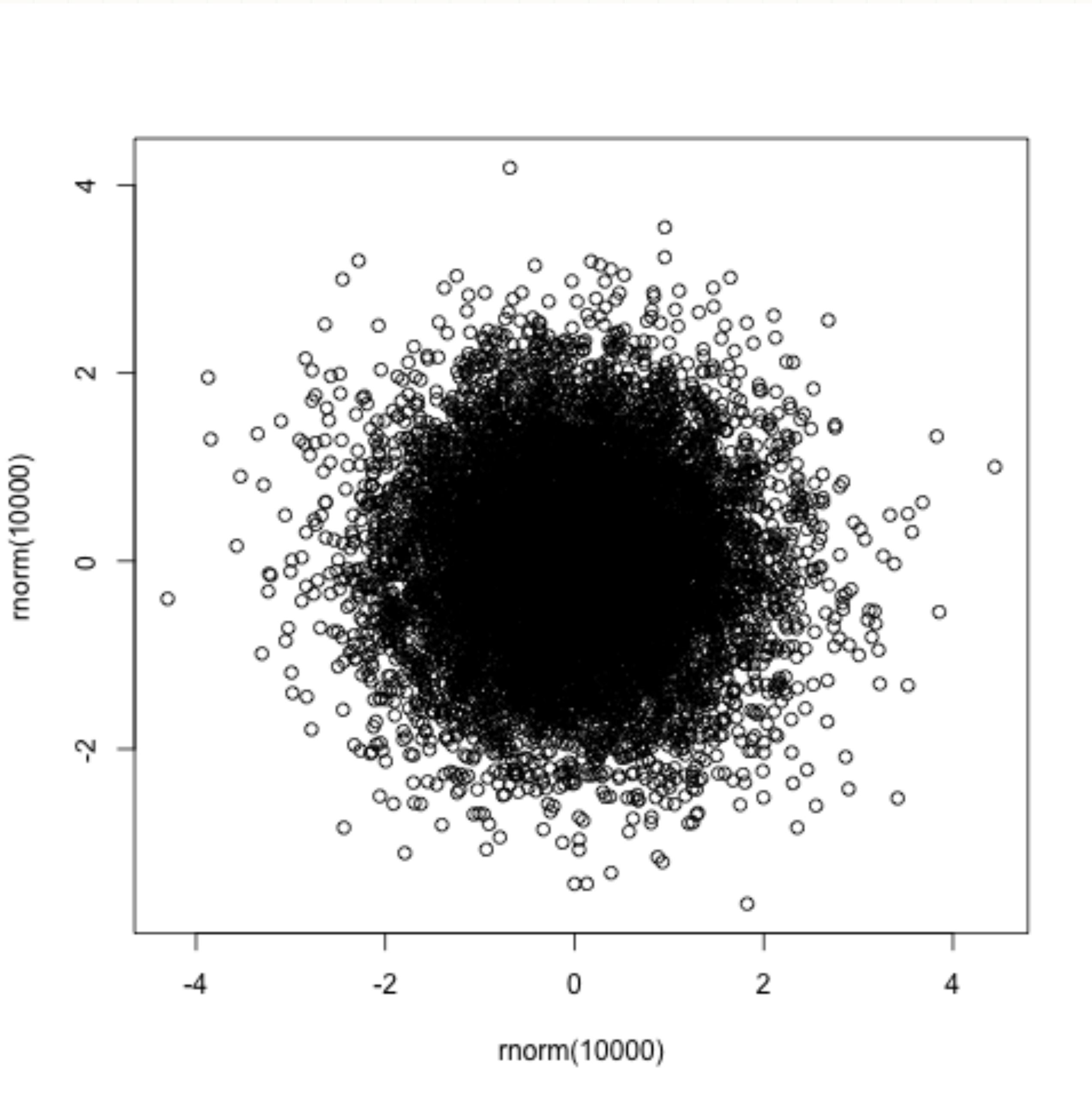
$$(\hat{\beta}-\beta)X^TX(\hat{\beta}-\beta) \sim \sigma^2\chi^2_{p+1}$$

where p is the number of the regressors. Using the same analogy we can look for vector points $\beta \in \mathbb{R}^{p+1}$ which satisfy the condition

$$(\hat{\beta}-\beta)X^TX(\hat{\beta}-\beta) \leq \sigma^2\chi^2_{p+1,1-\alpha}.$$

For $p = 1$ this set will be the interior of the ellipsis.

The confidence set is used since it accounts for interactions between β_i and β_j. Look at the scatter plot of two independent normal variables (which would be the case for orthogonal regressors with the same variance):



The circular shape is evident. Using the univariate confidence intervals the confidence set would be square, and this graph illustrates that it will actualy estimate the confidence incorrectly.

answered Nov 14 '11 at 13:44

mpiktas
24.5k 4 48 102

Thanks a lot for the response mpiktas. Here are some follow up questions Can you please elaborate on the step in which you were able to show that (betahat-beta)XtX(betahat-beta) has a chi-squared distribution? Also how does one show that for p = 1, we get an ellipse? What ellipse are you talking about? Then is that wrong to build individual tests on the isolated coefficients? – user7413 Nov 17 '11 at 3:21

Like also the equation in the book says that its an "approximate" confidence interval, what makes it approximate. Forgive my ignorance but I wasnt able to see how you got the scatter plot, are you just plotting you independent normally distributed random variables? – user7413 Nov 17 '11 at 3:22

To supplement, if $X \in \mathbb{R}^{N\times(p+1)}$ and $\hat{\beta}$ is the LS estimation for β in the linear regression model $Y = X\beta + \epsilon$ with $\epsilon \sim \mathcal{N}(0,\sigma^2)$,

$$\frac{(\hat{\beta}-\beta)^TX^TX(\hat{\beta}-\beta)}{\hat{\sigma}^2} \sim \chi^2_{p+1}$$

holds asymptotically when $N \to +\infty$. To see this, we first have

$$\begin{aligned}(\hat{\beta}-\beta)^TX^TX(\hat{\beta}-\beta) &\sim \sigma^2\chi^2_{p+1} \quad (\text{from } \hat{\beta} \sim \mathcal{N}(\beta,\sigma^2(X^TX)^{-1}))\\ (N-p-1)\hat{\sigma}^2 &\sim \sigma^2\chi^2_{N-p-1}\end{aligned}$$

which gives

$$\frac{(\hat{\beta}-\beta)^T X^T X(\hat{\beta}-\beta)}{(p+1)\hat{\sigma}^2} \sim F_{p+1,N-p-1}$$

On the other hand, one can prove if $S \sim F_{m,n}$, $T = \lim_{n\to+\infty} mS \sim \chi_m^2$ by directly computing the limit of mS's PDF, with the help of the relation between gamma function and beta function and Stirling's formula. With this claim, we have

$$\frac{(\hat{\beta}-\beta)^T X^T X(\hat{\beta}-\beta)}{\hat{\sigma}^2} \sim \chi_{p+1}^2 \quad (N \to +\infty)$$

.

edited Feb 19 '14 at 13:10

answered Feb 19 '14 at 13:04

ziyuang

624 3 20