

Problem Set 5 (Lectures 9-10-11-12)

Problem 1 (Parametric Bootstrap, 40 points) In this problem I would like you to code the parametric bootstrap algorithm described in the lecture notes. The context is the Homoskedastic Gaussian Linear Regression

$$Y \sim \mathcal{N}_n(X\beta, \sigma^2 \mathbb{I}_n),$$

with parameters β and σ^2 . Set $k = 1$ and let $X = (1, 1, \dots, 1)'$, so that the model becomes

$$Y \sim \mathcal{N}_n(\beta, \sigma^2 \mathbb{I}_n).$$

1. (20 points) Below is some matlab code to generate a bootstrap confidence set. Fill in the blanks. Run it in your computer to make sure it gives the desired output.

```
function [CIbeta, CIsig] = Boots_PSET(y,coverage,I)

%---
%INPUTS
%---
%- y : n times 1 vector of data
%- coverage : nominal coverage (scalar)
%- I : number of bootstrap draws
%---
%OUTPUT
%---
%- CIbeta : 1 x 2 vector (confidence interval for beta)
%- CIsig : 1 x 2 vector (confidence interval for sigma)

%% Compute the Maximum Likelihood estimator of beta and sigma2

betaML = ;
sigmaML = ;

%% Generate I artificial datasets of size n
n = size(y,1);

bootsdata = betaML +  randn(n,I)
```

```

%% Generate I estimators of betaML and sigmaML
betaboos = mean(bootsdata,1);
sigmaboos = var(bootsdata,1,1);
%% Bootstrap CSETs
alpha = ;
CIbeta = quantile(,[alpha/2,1-(alpha/2)],2);
CIsig = quantile(,[alpha/2,1-(alpha/2)],2);
end

```

2. (20 points) Now we are going to check if the $1 - \alpha$ parametric bootstrap covers the true parameter with probability $1 - \alpha$. Write a matlab script that does the following. First, let's take the role of nature and select values of β and σ^2 (let's take $\beta = 0$ and $\sigma^2 = 1$). Let's generate $D = 1000$ data sets y_1, y_2, \dots, y_D from the model $\mathcal{N}_n(0, \mathbb{I}_n)$. (imagine that nature and the econometrician are playing the game D times). To be more explicit, start by setting $n = 50$ so that each data set that we see has 50 observations. For each data set y_d do

```
[CIbeta, CIsig] = Boots_PSET(y_d, .95, 5000),
```

and for $d = 1, \dots, D$ create the variable `covbeta` as a dummy that equals 1 if the true β (0 in our example) belongs to `CIbeta`. Create an analogous variable `covsigma`. The mean of these variables is a Monte-Carlo approximation of the coverage probability of the parametric bootstrap confidence interval at $(\beta, \sigma^2) = (0, 1)$.

How often are the true values of β and σ covered by the parametric bootstrap when $n = 50$ according to the Monte Carlo approximation above? Create a graph of the coverage for different values of n , say 100, 150, 200, 250, 300, 500, 1000.

Problem 2 (Credible Set Based on the quantiles of the posterior, 40 points) Keep the same assumptions as in the previous problem:

$$Y \sim \mathcal{N}_n(\beta, \sigma^2 \mathbb{I}_n).$$

- (a) (20 points) Instead of writing a function that computes the parametric bootstrap confidence interval, I would like you to now write a function that—for a given data set—computes the credible set described in the lecture notes based on the prior. I am looking for a function

`[Credbeta, Credsig] = Bayes_PSET(y,credibility,I,lambda,a0,b0)`

that takes as inputs a data set y , a desired credibility level, a fixed number of posterior draws (I), and the “hyperparameters” λ , a_0 , b_0 .

- (b) Give a Monte Carlo approximation of the coverage of the Bayesian credible set for different values of n , say 100, 150, 200, 250, 300, 500, 1000, and for hyperparameters $\lambda = 1$, $a_0 = 0$, $b_0 = 0$? Can we say that in large samples, the Bayes 95% credible set has coverage of approximately 95%?

Problem 3 (Parametric Bootstrap as a tool to approximate the critical values of a test, 10 points) In class we discussed how to use an heuristic argument to approximate the critical values of the Likelihood Ratio Test in equation (1.9). Keep the same assumptions as in the previous problem:

$$Y \sim \mathcal{N}_n(\beta, \sigma^2 \mathbb{I}_n),$$

and suppose that we test the null hypothesis $\beta = 0$. Under the null, the distribution of the LR test:

$$n(\ln(\hat{\sigma}_0^2) - \ln(\hat{\sigma}_{\text{ML}}^2))$$

depends on $f(Y|\beta_0, \sigma^2)$, where σ^2 is unknown. The parametric bootstrap suggests that if $\hat{\sigma}_0^2$ is a good approximation to σ^2 (under the null), then

$$f(Y|\beta_0, \hat{\sigma}_0^2)$$

should be a decent approximation of $f(Y|\beta_0, \sigma^2)$. Thus, we could approximate the distribution of the LR statistic by taking I draws (say 10,000) from $f(Y|\beta_0, \hat{\sigma}_0^2)$; we could then recompute the LR statistic for each data set, and then obtain the $1 - \alpha$ quantile of the I statistics.

Conduct this exercise for $n = 50, 100, 200, 300, 500, 1000, 2000$ and $1 - \alpha = .95$? Does the critical value approach the 95% quantile of a χ_1^2 ?

References