ESCOLA BRASILEIRA DE ECONOMIA E FINANÇAS - EPGE

Luciano Fabio Busatto Venturim

Econometrics 1 - Problem Set 3

Rio de Janeiro

4th Quarter - 2021

Question 1: The benefits of the BFGS optimization algorithm compared to the Newton-Raphson optimization is that the BFGS does not need a Hessian matrix, and does not need to invert matrices, which is computationally expensive. The following equations contain the updating rule for the parameters and the updating rule for the approximation to the Hessian matrix in the BFGS algorithm. Suppose we are minimizing an objective function J with respect to 3 parameters β. Denote q_t as the gradient of the objective function at β_t. Denote as $g_t := q_{t+1} - q_t$ the difference between the gradients at β_{t+1} and β_t. Denote as $p_t := \beta_{t+1} - \beta_t$ the difference between the current parameter guess and the next guess.

$$\beta_{t+1} = \beta_t - A_t q_t$$
$$A_{t+1} = A_t + \frac{p_t p_t'}{p_t' g_t} + \frac{A_t g_t g_t' A_t}{g_t' A_t g_t} + g_t' A_t g_t h_t h_t'$$
$$h_t = \frac{p_t}{p_t' g_t} - \frac{A_t g_t}{g_t' A_t g_t}$$

1. Carefully write down the dimensions of all components on the right hand side, and show that this is a valid summation, i.e. show that we are summing across matrices of identical dimensions.

2. Prove that if A_t is symmetric, A_{t+1} is also symmetric.

Solution:

1. β is 3×1, so q_t is 3×1. Thus, A_t must be 3×3, g_t must be 3×1, and p_t must be 3×1. For each term:

   (a) $\frac{p_t p_t'}{p_t' g_t}$: $p_t p_t'$ is 3×3, $p_t' q_t$ is 1×1, so $\frac{p_t p_t'}{p_t q_t'}$ is 3×3.

   (b) $\frac{A_t g_t g_t' A_t}{g_t' A_t g_t}$: $g_t g_t'$ is 3×3 and $g_t' A_t g_t$ is 1×1, so $A_t g_t g_t' A_t$ is 3×3 and $\frac{A_t g_t g_t' A_t}{g_t' A_t g_t}$ is 3×3.

   (c) $h_t = \frac{p_t}{p_t' g_t} - \frac{A_t g_t}{g_t' A_t g_t}$: $\frac{p_t}{p_t' g_t}$ is 3×1, $A_t g_t$ is 3×1, thus h_t is 3×1;

   (d) $g_t' A_t g_t h_t h_t'$: $h_t h_t'$ is 3×3, $g_t' A_t g_T$ is 1×1, so $g_t' A_t g_t h_t h_t'$ is 3×3.

   Therefore, each term of the right hand side have dimension 3×3.

2. Assuming that A_t is symmetric, $(A_t + \frac{p_t p_t'}{p_t' g_t} + \frac{A_t g_t g_t' A_t}{g_t' A_t g_t} + g_t' A_t g_t h_t h_t')' = A_t' + \frac{p_t p_t'}{p_t' g_t} + \frac{A_t' g_t g_t' A_t'}{g_t' A_t g_t} + h_t h_t' g_t' A_t' g_t = A_t + \frac{p_t p_t'}{p_t' g_t} + \frac{A_t g_t g_t' A_t}{g_t' A_t g_t} + g_t' A_t g_t h_t h_t'$, since $g_t' A_t g_t$ is a number. Thus, $A_{t+1} = A_{t+1}'$.

Question 2: Suppose we have an i.i.d sample $Y = (y_1, \ldots, y_n)$ drawn from a population with $E(y_i) = \mu$. Suppose we draw a "bootsample"from Y by drawing n times with replacement. Denote the bootsample with $(y_1^*, \ldots, y_n^*)$ and the bootstrap sample mean with $\overline{y}^* = \frac{1}{n}\sum_{i=1}^n y_i^*$. Remember that P^* is the probability that an event happens conditional on the empirical distribution function F_n

$$P^*(y^* \leq z) := P(y^* \leq z | F_n)$$

and that "bootstrap convergence"$z_n^* \xrightarrow{P^*} z$ is defined as

$$P^*(||z_n^* - z|| > \epsilon) \xrightarrow{P^*} 0$$

Show that the bootstrap sample mean $\overline{y}^*$ "bootstrap converges"to the true population mean μ, i.e, $\overline{y}^* \xrightarrow{P^*} \mu$ as the sample size goes to infinity.

Solution: First, we claim that if $z_n \xrightarrow{P} z$, then $z_n \xrightarrow{P^*} z$. In fact, for any $\epsilon > 0$, there is n sufficiently large such that $P(||z_n - z|| > \epsilon) < \epsilon$. For this n, conditional on F_n, the event $||z_n - z|| > \epsilon$ is non-random, thus $P^*(||z_n - z|| > \epsilon) = 0$ with probability more than $1 - \epsilon$ or $P^*(||z_n - z|| > \epsilon) = 1$ with probability less than ϵ. Since ϵ is arbitrary, we have our result.

Now, $P^*(||\overline{y}^* - \overline{y}|| > \epsilon) \leq \frac{\mathbb{E}^*[||\overline{y}^* - \overline{y}||^2]}{\epsilon^2} = \frac{tr(var^*(\overline{y}^*))}{\epsilon^2} = \frac{tr(\frac{1}{n}\hat{\Sigma})}{\epsilon^2} \leq \frac{\sum_{i=1}^n y_i' y_i}{\epsilon^2 n^2} \xrightarrow{P} 0$, i.e., $\overline{y}^* \xrightarrow{P^*} \overline{y}$. And since $\overline{y} \xrightarrow{P} \mu$, $\overline{y} \xrightarrow{P^*} \mu$. Therefore, $\overline{y}^* - \mu = \overline{y}^* - \overline{y} + \overline{y} - \mu \xrightarrow{P^*} 0$, i.e., $\overline{y}^* \xrightarrow{P^*} \mu$.

Question 3: Think of an example for a differentiated good (not cars and not TV packages) and assume that consumers can choose only one good from J available options. Think of at least one observable characteristic that differentiates consumers and affects their enjoyment of a characteristic of the good. Write down the utility maximization problem for a consumer i and allow your utility function to capture differences in utility that arise from (1) differences in product characteristics and (2) observable differences in consumer characteristics.

Solution: An example of differentiated goods are recreational clubs in a city. Some of the characteristics are: (i) number of pools; (ii) presence of sauna; (iii) presence of playground for children; (iv) presence of soccer field; (v) price; and (vi) presence of food court. An observable characteristic that differentiated consumers and affects their enjoyment of the characteristic ”presence of playground for children”is the number of kids that consumer has.

Let $x_{j,k}$ be the characteristic (k), $k = 1, \ldots, 6$ from above, of the j-th good. Then, the utility for a consumer i is given by

$$u_i(x_j) = \sum_{k=1}^{6} \beta_{i,k} x_{j,k} + \epsilon_{i,j}$$

where $\beta_{i,k} = \overline{\beta}_k \ \forall \ k \neq 3$, $\beta_{i,3} = \overline{\beta}_3 + \beta_3^0 z_i$, z_i is the number of children consumer i has, $\overline{\beta}_k$ is a common enjoyment parameter of the population towards characteristic k, β_3^0 is a parameter for the individual characteristic ”number of children”that affects the enjoyment of characteristic 3, and $\epsilon_{i,j}$ is an unobserved individual valuation of good j for consumer i.

The utility maximization problem for consumer i is, therefore,

$$\max_{j=1,\ldots,J} u_i(x_j)$$