ECON 899b, Problem Set 1

Jackson Crawford, Dalya Elmalt, Jon Kroah

Due 4 Nov 2022

Setup:

$$Y_i = \begin{cases} 1 & \text{if } \beta_0 + X_i\beta_1 + \epsilon_{i1} > \epsilon_{i0} \\ 0 & \text{otherwise} \end{cases}$$

$$\epsilon_{ij} \overset{\text{iid}}{\sim} \text{T1EV}(0,1)$$

1 Question 1

1.1 Derive log-likelihood: $l(y_1, \ldots, y_n | x_1, \ldots, x_n; \beta_0, \beta_1)$

First, note that

$$\begin{aligned} P(Y_i = 1 | X_i = x_i; \beta) &= P(\beta_0 + x_i\beta_1 + \epsilon_{i1} > \epsilon_{i0}) \\ &= P(\epsilon_{i1} - \epsilon_{i0} > -(\beta_0 + x_i\beta_1)) \\ &= \frac{\exp(\beta_0 + x_i\beta_1)}{1 + \exp(\beta_0 + x_i\beta_1)} \end{aligned}$$

Assume $\{(Y_i, X_i)\}_{i=1}^n$ is an i.i.d. sample. Then we can derive the log-likelihood:

$$\begin{aligned} &P(Y_1 = y_1, \ldots, Y_n = y_n | X_1 = x_1, \ldots, X_n = x_n; \beta_0, \beta_1) \\ &= \prod_{i=1}^n P(Y_i = y_i | X_i = x_i; \beta_0, \beta_1) \\ &= \prod_{i=1}^n P(Y_i = 1 | X_i = x_i; \beta_0, \beta_1)^{1\{y_i = 1\}} P(Y_i = 0 | X_i; \beta_0, \beta_1)^{1\{y_i = 0\}} \\ \Rightarrow &l(y_1, \ldots, y_n | x_1, \ldots, x_n; \beta_0, \beta_1) \\ &= \sum_{i=1}^n \Big[1\{y_i = 1\} \log P(Y_i = 1 | X_i; \beta_0, \beta_1) + 1\{y_i = 0\} \log P(Y_i = 0 | X_i; \beta_0, \beta_1) \Big] \\ &= \sum_{i=1}^n \left[1\{y_i = 1\} \log \left(\frac{\exp(\beta_0 + X_i\beta_1)}{1 + \exp(\beta_0 + X_i\beta_1)} \right) + 1\{y_i = 0\} \log \left(\frac{1}{1 + \exp(\beta_0 + X_i\beta_1)} \right) \right] \\ &= \boxed{\sum_{i=1}^n \Big[1\{y_i = 1\}(\beta_0 + x_i\beta_1) - \log \Big(1 + \exp(\beta_0 + x_i\beta_1)\Big) \Big]} \end{aligned}$$

1.2 Score of the log-likelihood:

$$\frac{\partial l}{\partial \beta_0} = \sum_{i=1}^{n} \left[1\{y_i = 1\} - \frac{\exp(\beta_0 + x_i\beta_1)}{1 + \exp(\beta_0 + x_i\beta_1)} \right] \qquad (1 \times 1)$$

$$\frac{\partial l}{\partial \beta_1} = \sum_{i=1}^{n} \left[x_i 1\{y_i = 1\} - \frac{x_i \exp(\beta_0 + x_i\beta_1)}{1 + \exp(\beta_0 + x_i\beta_1)} \right] = \sum_{i=1}^{n} x_i \left[1\{y_i = 1\} - \frac{\exp(\beta_0 + x_i\beta_1)}{1 + \exp(\beta_0 + x_i\beta_1)} \right] \qquad (K \times 1)$$

Score is:

$$\nabla l(\beta; y, x) = (\partial l/\partial \beta_0, \partial l/\partial \beta_1)' \qquad \text{(this is } (K+1) \times 1)$$

1.3 Hessian of the log-likelihood

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta_0^2} &= -\sum_i \left[\frac{\exp(\beta_0 + x_i\beta_1)}{1 + \exp(\beta_0 + x_i\beta_1)} - \frac{\exp(\beta_0 + x_i\beta_1)^2}{[1 + \exp(\beta_0 + x_i\beta_1)]^2} \right] \\ &= -\sum_i \frac{\exp(\beta_0 + x_i\beta_1)}{1 + \exp(\beta_0 + x_i\beta_1)} \left(1 - \frac{\exp(\beta_0 + x_i\beta_1)}{1 + \exp(\beta_0 + x_i\beta_1)} \right) \qquad \text{(this is } 1 \times 1) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta_1 \partial \beta_1'} &= -\sum_i x_i \left[x_i' \frac{\exp(\beta_0 + x_i\beta_1)}{1 + \exp(\beta_0 + x_i\beta_1)} - x_i' \frac{\exp(\beta_0 + x_i\beta_1)^2}{[1 + \exp(\beta_0 + x_i\beta_1)]^2} \right] \\ &= -\sum_i x_i x_i' \frac{\exp(\beta_0 + x_i\beta_1)}{1 + \exp(\beta_0 + x_i\beta_1)} \left(1 - \frac{\exp(\beta_0 + x_i\beta_1)}{1 + \exp(\beta_0 + x_i\beta_1)} \right) \qquad \text{(this is } K \times K) \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta_0 \partial \beta_1'} &= -\sum_i \left[x_i' \frac{\exp(\beta_0 + x_i\beta_1)}{1 + \exp(\beta_0 + x_i\beta_1)} - x_i' \frac{\exp(\beta_0 + x_i\beta_1)^2}{[1 + \exp(\beta_0 + x_i\beta_1)]^2} \right] \\ &= -\sum_i x_i' \frac{\exp(\beta_0 + x_i\beta_1)}{1 + \exp(\beta_0 + x_i\beta_1)} \left(1 - \frac{\exp(\beta_0 + x_i\beta_1)}{1 + \exp(\beta_0 + x_i\beta_1)} \right) \qquad \text{(this is } K \times 1) \end{aligned}$$

Hessian is:

$$H = \begin{pmatrix} \partial^2 l/\partial \beta_0^2 & \partial^2 l/\partial \beta_0 \partial \beta_1'{}' \\ \partial^2 l/\partial \beta_0 \partial \beta_1' & \partial^2 l/\partial \beta_1 \partial \beta_1' \end{pmatrix} \qquad \text{(this is } (K+1) \times (K+1))$$

1.4 Results:

Let $\tilde{\beta} = (-1, 0, \ldots, 0)$.

Log-likelihood: $l(\tilde{\beta}, y, x) = -6924.395$

Analytic score $\nabla l(\tilde{\beta}, y, x)$:

k	β_k = Coefficient on:	$\partial l / \partial \beta_k$
(0)	Intercept	-2597.5
(1)	Large loan	-554.4
(2)	Medium loan	-1153.0
(3)	Rate spread	-220.4
(4)	Refi	-929.1
(5)	Age	-1210.2
(6)	Combined LTV	-2102.7
(7)	Mtg. Debt-to-Income	-944.3
(8)	Credit union	-5033.7
(9)	First mtg.	-4519.7
(10)	FICO year 0	-19342.5
(11)	FICO year 1	-19103.9
(12)	FHA	-913.1
(13)	Open 2014	-350.1
(14)	Open 2015	-464.4
(15)	Open 2016	-580.8
(16)	Open 2017	-544.3

Analytic Hessian $\nabla^2 l(\tilde{\beta}, y, x)$

	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(0)	-3215.6	-878.3	-1424.8	-385.8	-1301.4	-1541.1	-2611.9	-1206.8	-6286.9	-5745.2	-23719.5	-23534.5	-1399.3	-662.4	-678.7	-672.4	-581.4
(1)	-878.3	-878.3	0.0	-10.0	-403.4	-420.0	-684.4	-331.0	-1716.4	-1651.0	-6592.6	-6547.2	-388.5	-163.4	-188.7	-211.4	-169.1
(2)	-1424.8	0.0	-1424.8	-164.5	-558.2	-674.0	-1189.2	-543.1	-2788.9	-2544.9	-10487.9	-10403.7	-584.9	-283.1	-307.1	-298.7	-266.4
(3)	-385.8	-10.0	-164.5	-713.1	-184.9	-186.8	-323.4	-151.3	-780.5	-690.9	-2727.1	-2708.5	44.9	-59.4	-103.8	-91.9	-77.2
(4)	-1301.4	-403.4	-558.2	-184.9	-1301.4	-690.6	-969.7	-499.3	-2547.7	-2584.3	-9523.6	-9556.4	-499.6	-213.7	-289.4	-298.5	-192.1
(5)	-1541.1	-420.0	-674.0	-186.8	-690.6	-802.7	-1226.8	-582.8	-3013.6	-2831.0	-11394.8	-11318.6	-656.9	-310.9	-324.4	-324.3	-281.9
(6)	-2611.9	-684.4	-1189.2	-323.4	-969.7	-1226.8	-2218.0	-989.3	-5111.0	-4607.4	-19165.8	-18997.3	-1223.2	-543.6	-548.8	-538.4	-476.1
(7)	-1206.8	-331.0	-543.1	-151.3	-499.3	-582.8	-989.3	-526.7	-2362.2	-2162.1	-8842.7	-8771.0	-555.3	-247.3	-255.9	-252.6	-219.6
(8)	-6286.9	-1716.4	-2788.9	-780.5	-2547.7	-3013.6	-5111.0	-2362.2	-12429.4	-11233.2	-46358.1	-45992.1	-2707.5	-1293.9	-1326.9	-1315.3	-1131.5
(9)	-5745.2	-1651.0	-2544.9	-690.9	-2584.3	-2831.0	-4607.4	-2162.1	-11233.2	-10804.4	-42497.9	-42187.5	-2409.7	-1166.9	-1217.8	-1210.5	-1030.4
(10)	-23719.5	-6592.6	-10487.9	-2727.1	-9523.6	-11394.8	-19165.8	-8842.7	-46358.1	-42497.9	-176268.1	-174611.3	-10022.4	-4894.4	-4995.0	-4956.5	-4260.5
(11)	-23534.5	-6547.2	-10403.7	-2708.5	-9556.4	-11318.6	-18997.3	-8771.0	-45992.1	-42187.5	-174611.3	-173634.8	-9938.3	-4836.8	-4956.1	-4931.3	-4235.4
(12)	-1399.3	-388.5	-584.9	44.9	-499.6	-656.9	-1223.2	-555.3	-2707.5	-2409.7	-10022.4	-9938.3	-1399.3	-292.0	-304.9	-304.4	-267.2
(13)	-662.4	-163.4	-283.1	-59.4	-213.7	-310.9	-543.6	-247.3	-1293.9	-1166.9	-4894.4	-4836.8	-292.0	-662.4	0.0	0.0	0.0
(14)	-678.7	-188.7	-307.1	-103.8	-289.4	-324.4	-548.8	-255.9	-1326.9	-1217.8	-4995.0	-4956.1	-304.9	0.0	-678.7	0.0	0.0
(15)	-672.4	-211.4	-298.7	-91.9	-298.5	-324.3	-538.4	-252.6	-1315.3	-1210.5	-4956.5	-4931.3	-304.4	0.0	0.0	-672.4	0.0
(16)	-581.4	-169.1	-266.4	-77.2	-192.1	-281.9	-476.1	-219.6	-1131.5	-1030.4	-4260.5	-4235.4	-267.2	0.0	0.0	0.0	-581.4

2 Question 2

2.1 Formulas for numerical derivatives

Let

$$\beta^j := (\beta_1, \ldots, \beta_j + \epsilon, \ldots, \beta_K) \text{ (perturb element j)}$$
$$\beta^{jl} := (\beta_1, \ldots, \beta_j + \epsilon, \ldots, \beta_l + \epsilon \ldots, \beta_K) \text{ (perturb elements j and l)}$$

Numerical first and second derivatives:

$$\frac{\partial l}{\partial \beta_j} \approx \frac{f(\beta^j) - f(\beta)}{\epsilon}, \qquad \frac{\partial^2 l}{\partial \beta_j \beta_l} \approx \frac{1}{\epsilon}\left[\frac{f(\beta^{jl}) - f(\beta^l)}{\epsilon} - \frac{f(\beta^j) - f(\beta)}{\epsilon}\right]$$

2.2 Results:

Numerical score (evaluated at $\tilde{\beta} = (-1, 0, \ldots, 0)$, $\epsilon = 10^{-5}$):

k	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Approx. $\partial l/\partial \beta_k$	-2597.5	-554.4	-1153.0	-220.4	-929.1	-1210.2	-2102.7	-944.3	-5033.7	-4519.7	-19342.5	-19103.9	-913.1	-350.1	-464.4	-580.8	-544.3

Numerical Hessian (evaluated at $\tilde{\beta} = (-1, 0, \ldots, 0)$, $\epsilon = 10^{-5}$):

	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(0)	-3215.3	-878.2	-1424.7	-385.5	-1301.3	-1540.8	-2611.5	-1206.6	-6286.0	-5744.5	-23719.6	-23534.6	-1399.2	-662.3	-678.6	-672.4	-581.3
(1)	-878.2	-878.2	-0.0	-9.9	-403.4	-419.9	-684.4	-331.0	-1716.2	-1650.8	-6592.7	-6547.3	-388.5	-163.4	-188.7	-211.3	-169.1
(2)	-1424.7	-0.0	-1424.7	-164.4	-558.1	-673.8	-1189.0	-543.0	-2788.6	-2544.6	-10488.0	-10403.8	-584.9	-283.1	-307.1	-298.6	-266.4
(3)	-385.5	-9.9	-164.4	-712.9	-184.7	-186.6	-323.2	-151.2	-780.1	-690.5	-2727.0	-2708.3	45.0	-59.3	-103.7	-91.8	-77.2
(4)	-1301.3	-403.4	-558.1	-184.7	-1301.3	-690.5	-969.6	-499.2	-2547.4	-2583.9	-9523.7	-9556.5	-499.5	-213.7	-289.4	-298.4	-192.1
(5)	-1540.8	-419.9	-673.8	-186.6	-690.5	-802.6	-1226.7	-582.7	-3013.3	-2830.7	-11394.9	-11318.6	-656.7	-310.9	-324.3	-324.2	-281.8
(6)	-2611.5	-684.4	-1189.0	-323.2	-969.6	-1226.7	-2217.9	-989.1	-5110.6	-4607.0	-19165.9	-18997.4	-1223.1	-543.5	-548.8	-538.3	-476.0
(7)	-1206.6	-331.0	-543.0	-151.2	-499.2	-582.7	-989.1	-526.4	-2361.9	-2161.8	-8842.7	-8771.0	-555.2	-247.3	-255.8	-252.5	-219.6
(8)	-6286.0	-1716.2	-2788.6	-780.1	-2547.4	-3013.3	-5110.6	-2361.9	-12429.2	-11232.8	-46358.7	-45992.6	-2707.2	-1293.7	-1326.7	-1315.2	-1131.3
(9)	-5744.5	-1650.8	-2544.6	-690.5	-2583.9	-2830.7	-4607.0	-2161.8	-11232.8	-10804.2	-42498.4	-42188.0	-2409.4	-1166.7	-1217.6	-1210.4	-1030.3
(10)	-23719.6	-6592.7	-10488.0	-2727.0	-9523.7	-11394.9	-19165.9	-8842.7	-46358.7	-42498.4	-176274.0	-174617.1	-10022.4	-4894.4	-4995.0	-4956.5	-4260.5
(11)	-23534.6	-6547.3	-10403.8	-2708.3	-9556.5	-11318.6	-18997.4	-8771.0	-45992.6	-42188.0	-174617.1	-173640.5	-9938.3	-4836.8	-4956.1	-4931.3	-4235.5
(12)	-1399.2	-388.5	-584.9	45.0	-499.5	-656.7	-1223.1	-555.2	-2707.2	-2409.4	-10022.4	-9938.3	-1399.2	-291.9	-304.9	-304.3	-267.2
(13)	-662.3	-163.4	-283.1	-59.3	-213.7	-310.9	-543.5	-247.3	-1293.7	-1166.7	-4894.4	-4836.8	-291.9	-662.3	-0.0	0.0	0.0
(14)	-678.6	-188.7	-307.1	-103.7	-289.4	-324.3	-548.8	-255.8	-1326.7	-1217.6	-4995.0	-4956.1	-304.9	-0.0	-678.6	0.0	0.0
(15)	-672.4	-211.3	-298.6	-91.8	-298.4	-324.2	-538.3	-252.5	-1315.2	-1210.4	-4956.5	-4931.3	-304.3	0.0	0.0	-672.4	0.0
(16)	-581.3	-169.1	-266.4	-77.2	-192.1	-281.8	-476.0	-219.6	-1131.3	-1030.3	-4260.5	-4235.5	-267.2	0.0	0.0	0.0	-581.3

The numerical score is very close to the analytic score. The numerical Hessian is also close to the analytic Hessian, but is often off at the 1st decimal place.

3 Questions 3 & 4

Newton algorithm:

1. Take initial guess β^0
2. h^{th} guess $\beta^h = \beta^{h-1} - H(\beta^{h-1})^{-1}\nabla l(\beta^{h-1})$
3. Keep updating the guess until $||\beta^h - \beta^{h-1}|| < \varepsilon$.

For the Newton algorithm, we use the analytic score and Hessian (the algorithm fails to converge if we use the numerical approximations, depending on the tolerance).

Results:

All 3 methods (Newton, BFGS, and simplex) produce identical estimates. The Newton method is the fastest and least computationally intensive, but relies on analytic expressions for the score and Hessian. Nelder-Mead is the slowest and most computationally intensive.

For all 3 algorithms, we use an initial guess of $\beta^0 = (-1, 0, \ldots, 0)$.

	Estimates: $\hat{\beta}_k$		
k	Newton	BFGS	Simplex
(0)	-6.083	-6.083	-6.083
(1)	0.872	0.872	0.872
(2)	0.532	0.532	0.532
(3)	0.604	0.604	0.604
(4)	0.163	0.163	0.163
(5)	0.877	0.877	0.877
(6)	-0.061	-0.061	-0.061
(7)	0.227	0.227	0.227
(8)	1.011	1.011	1.011
(9)	0.335	0.335	0.335
(10)	-0.287	-0.287	-0.287
(11)	0.194	0.194	0.194
(12)	0.765	0.765	0.765
(13)	1.154	1.154	1.154
(14)	0.771	0.771	0.771
(15)	0.378	0.378	0.378
(16)	0.244	0.244	0.244
Iterations	7	30	4729
Time (Seconds)	0.05	1.32	2.57