Logit/Probit Exercises

Prof. Conlon

2025-10-05

1. Let's load the Boston HMDA data.

The function should take the following arguments:

- dir: debt to income ratio
- hir : housing to income ratio
- single : dummy for single borrower
- self : dummy for self-employed

```
data("Hmda")

data <- Hmda %>%
  mutate(
    deny = 1 * (deny == "yes"),
    black = 1 * (black == "yes"),
    pbcr = 1 * (pbcr == "yes"),
    dmi = 1 * (dmi == "yes"),
    lvr_high = (lvr > 0.95),
    lvr_med  = (0.8 <= lvr) & (lvr <= 0.95)
  )

formula <- deny ~ dir + hir + single + self
logit    <- feols(formula, data = data,family = binomial(link = "logit"))
```

```
## Warning: In fixest_env(fml = fml, data = data, weights = weig...:
##  feols(fml = for...: family is not a valid argument for function feols()
```

```
## NOTE: 1 observation removed because of NA values (RHS: 1).
```

```
summary(logit)
```

OLS estimation, Dep. Var.: deny Observations: 2,380 Standard-errors: IID Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.095945 0.021766 -4.40804 1.0895e-05 ***dir 0.734388 0.097445 7.53646 6.8251e-14*** hir -0.202476 0.108210 -1.87114 6.1448e-02 .
singleyes 0.049009 0.013336 3.67487 2.4324e-04 ** *selfyes 0.044464 0.020332 2.18690 2.8847e-02*
— Signif. codes: 0 '***' ***0.001*** '' *0.01* '' 0.05 '.' 0.1 ' ' 1 RMSE: 0.316736 Adj. R2: 0.046653

2. Consider the regression model of the logit regression:

$$deny_i = F(\beta_1 \cdot dir_i + \beta_2 \cdot hir_i + \beta_3 \cdot single_i + \beta_4 \cdot self_i)$$

For a single observation compute the contribution to the log-likelihood (analytically)

3. For a single observation compute the Score (analytically).
4. Compute the Hessian Matrix and Fisher information (analytically).
5. Code up the Fisher Information for the logit model above $I(\hat{\beta})$ using the Hessian Matrix.
6. Code up the Fisher Information for the logit model above $I(\hat{\beta})$ using the score method.
7. Compute the standard errors from the Fisher information and compare them to the standard errors reported from the regression. How do they compare?
8. Generate $n = 100$ observations where $\lambda = 15$ from a poisson model:

$$Y_i \sim Pois(\lambda)$$

9. The poisson distribution is a discrete distribution for count data where the p.m.f. is given by:

$$Pr(Y_i = k) = \frac{\lambda^k e^{-\lambda}}{k!}a$$

10. Write the log-likelihood $\ell(y_1, \dots, y_n; \lambda)$ (analytically).
11. Write the Score contribution $\mathcal{S}_i(y_i; \lambda)$ (analytically).
12. Write the Hessian Contribution $\mathcal{H}_i(y_i; \lambda)$ (analytically).
13. Code up the log-likelihood function

```
pois_log_lik <- function(lambda,y){

  return(ll)
}
```

14. Find the value of λ that maximizes your log likelihood using `optim` in R.
15. Write a function that returns the standard error of $\hat{\lambda}$:

```
pois_se <- function(lambda_hat,y){

  return(se)
}
```