Exercise 1

$y_i = x_i'\beta + u_i$. Assume that $u_i|x_i \sim \mathcal{N}(0,1)$ are i.i.d.

Part (a)

What is the distribution of $y_i|x_i$? Condition both sites with respect to x_i:

$$y_i|x_i = x_i'\beta|x_i + u_i|x_i.$$

Since $u_i|x_i$ is standard normal, and we are adding $x_i'\beta|x_i = x_i\beta$ to it, it follows that $y_i|x_i \sim \mathcal{N}(x_i'\beta, 1)$. Recall that the density function for a normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$ is

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}.$$

In our case, we have $\sigma^2 = 1$ and $\mu = x_i'\beta$. Therefore

$$p(y_i|x_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_i - x_i'\beta)^2}{2}} = \phi(y_i - x_i'\beta).$$

Part (b)

The likelihood function can be written as

$$L(\beta|x,y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y_i - x_i'\beta)^2}{2}},$$

and therefore the log-likelihood function is

$$\ell(\beta|x,y) = \sum_{i=1}^{n} \left[-\frac{1}{2}\log(2\pi) - \frac{(y_i - x_i'\beta)^2}{2} \right].$$

Part (c)

The first order condition is

$$\sum_{i=1}^{n} x_i(y_i - x_i'\hat{\beta}) := 0.$$

Part (d)

Uh, the closed form solution is

$$\left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i y_i = \hat{\beta}.$$

So we need invertibility of $\sum_{i=1}^{n} x_i x_i'$.

Part (e)

Let $s(y_i, x_i; \beta) = x_i(y_i - x_i'\beta)$. It follows that the Hessian is $H(y_i, x_i; \beta) = -x_i x_i'$. Then this process is exactly the same as in the OLS case, as seen in the previous homework.

Exercise 2

Part (a)

Suppose $y_i^* = x_i'\beta + u_i$, but we only observe a censored version of it, y_t, where

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > c, \\ c & \text{otherwise.} \end{cases}$$

Suppose that $u_i | x_i \sim \mathcal{N}(0, \sigma^2)$. Let $\theta = (\beta', \sigma^2)'$. **Find the maximum likelihood estimator.**

First of all, $y_i^* | x_i \sim \mathcal{N}(x_i'\beta, \sigma^2)$. Because of the censoring, we have to define the density of y_i piece-wise. If $y_i > c$, then we have

$$f(y_i^* | x_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i - x_i'\beta)^2}{2\sigma^2}}.$$

The other part needs some brain-thinking. Notice that

$$\begin{aligned} P(y_i = c) &= P(y_i^* \leq c) \\ &= P(x_i'\beta + u_i \leq c) \\ &= P(u_i \leq c - x_i'\beta). \end{aligned}$$

We know that $u_i | x_i \sim \mathcal{N}(0, \sigma^2)$. Therefore

$$P(u_i \leq c - x_i'\beta) = P\left(\frac{u_i}{\sigma} \leq \frac{c - x_i'\beta}{\sigma}\right),$$

which is now a standard normal distribution—and therefore

$$P(y_i = c) = \Phi\left(\frac{c - x_i'\beta}{\sigma}\right).$$

So then we can write the likelihood function as

$$L(\beta, \sigma | y_i, x_i) = \prod_{i=1}^{n} \left[\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i - x_i'\beta)^2}{2\sigma^2}}\right]^{1(y_i > c)} \left[\Phi\left(\frac{c - x_i'\beta}{\sigma}\right)\right]^{1(y_i = c)},$$

where $1(\cdot)$ is an indicator random variable. For example, $1(y_i > c) = 1$ if $y_i > c$ and is zero otherwise.

We can log-likelihood it up by introducing another indicator random variable D_i such that

$$D_i = \begin{cases} 1 & \text{if } y_i > c, \\ 0 & \text{if } y_i = c. \end{cases}$$

Then the log-likelihood is

$$\ell(\beta, \sigma | y_i, x_i) = \sum_{i=1}^{n} D_i \left[-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(y_i - x_i'\beta)^2}{2\sigma^2}\right] + (1 - D_i) \ln \left[\Phi\left(\frac{c - x_i'\beta}{\sigma}\right)\right].$$

Yikes.

Part (b)

We have two regressions,

$$y_{1i} = x_i'\gamma_1 + \epsilon_i,$$
$$y_{2i} = \mu y_{1i} + x_i'\gamma_2 + \eta_i.$$

Assume that ϵ_i and η_i are independent of each other. Have $\epsilon_i | x_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ i.i.d. and $\eta_i | y_{1i}, x_i \sim \mathcal{N}\left(0, \sigma_\eta^2\right)$ i.i.d. Let $\theta = (\gamma_1', \mu, \gamma_2', \sigma_\epsilon^2, \sigma_\eta^2)$. **Write down the maximum likelihood estimator.**

The conditional joint density of two functions can be written as the product of a conditional and a marginal one, that is,

$$f(y_1, y_2 | x) = f(y_2 | y_1, x) f(y_1 | x).$$

First of all, notice that $p(y_{1i}|x_i) = \mathcal{N}\left(x_i'\gamma_1, \sigma_\epsilon^2\right)$, which has a density of

$$\frac{1}{\sigma_\epsilon\sqrt{2\pi}}e^{-\frac{(y_{1i}-x_i'\gamma_1)^2}{2\sigma_\epsilon^2}}.$$

Second, $y_{2i|x_i} = \mu y_{1i} + x_i'\gamma_2$, so $p(y_{2i}|x_i, y_{1i}) = \mathcal{N}\left(\mu y_{1i} + x_i'\gamma_2, \sigma_\eta^2\right)$, which has density

$$\frac{1}{\sigma_\eta\sqrt{2\pi}}e^{-\frac{(y_{2i}-\mu y_{1i}-x_i'\gamma_2)^2}{2\sigma_\eta^2}}.$$

The joint density then is

$$\frac{1}{\sigma_\epsilon\sqrt{2\pi}}e^{-\frac{(y_{1i}-x_i'\gamma_1)^2}{2\sigma_\epsilon^2}}\frac{1}{\sigma_\eta\sqrt{2\pi}}e^{-\frac{(y_{2i}-\mu y_{1i}-x_i'\gamma_2)^2}{2\sigma_\eta^2}}.$$

Jesus.

Problem 3

Part (a)

$y \sim Bernoulli(p)$ takes on two values—either $y = 0$ with probability $1 - p$ or $y = 1$ with probability p. For the density function, when $y = 0$ the function will be $1 - p$. And when $y = 1$, the function will be p. So, in general, the density will be

$$f(y_i; p) = p^y(1-p)^{1-y}.$$

For this problem, we are given $p = F(x_i'\beta)$. Therefore

$$p(y_i|x_i) = \begin{cases} F(x_i'\beta) & \text{if } y_i = 1, \\ 1 - F(x_i'\beta) & \text{if } y_i = 0, \end{cases}$$

which we could also write as

$$f(y_i|x_i;\beta) = F(x_i'\beta)^{y_i}\left(1 - F(x_i'\beta)\right)^{1-y_i}.$$

- $P(y_i = 1|x_i;\beta) = F(x_i'\beta)$
- $P(y_i = 0|x_i;\beta) = 1 - F(x_i'\beta)$
- $E[y_i|x_i;\beta] = F(x_i'\beta)$
- $\dfrac{\partial E[y_i|x_i;\beta]}{\partial x_i} = \beta f(x_i\beta)$

Part (b)

Much like in one of the previous questions, we'll use an indicator to write the log-likelihood. Incidentally, we already *have* an indicator, y_i. So we can log-likelihood it up as

$$\ell(\beta; y, x) = \sum_{i=1}^{n} y_i \log\big(F(x_i'\beta)\big) + (1-y_i)\log\big(1-F(x_i'\beta)\big).$$

Part (c)

The first order condition be

$$\sum_{i=1}^{n} x_i \frac{y_i}{F(x_i'\hat{\beta})} f(x_i'\hat{\beta}) - x_i \frac{1-y_i}{1-F(x_i'\hat{\beta})} f(x_i'\hat{\beta}) := 0,$$

which can be rewritten as

$$\sum_{i=1}^{n} x_i f(x_i'\hat{\beta}) \left[\frac{y_i - F(x_i'\hat{\beta})}{F(x_i'\hat{\beta})\big(1-F(x_i'\hat{\beta})\big)}\right] = 0.$$

The probit case is

$$\sum_{i=1}^{n} x_i \phi(x_i'\hat{\beta}) \left[\frac{y_i - \Phi(x_i'\hat{\beta})}{\Phi(x_i'\hat{\beta})\big(1-\Phi(x_i'\hat{\beta})\big)}\right] = 0,$$

and the logit case is

$$\sum_{i=1}^{n} x_i \lambda(x_i'\hat{\beta}) \left[\frac{y_i - \Lambda(x_i'\hat{\beta})}{\Lambda(x_i'\hat{\beta})\big(1-\Lambda(x_i'\hat{\beta})\big)}\right] = 0.$$

Part (d)

Give conditions for the consistency of the general case.

- **Existence.** We just assume that $\hat{\theta} = \arg\max_\theta n^{-1} \sum_{i=1}^{n} \log\big(f(y_i|x_i;\theta)\big)$ exists.
- **Identification.** We need $f(y_i|x_i;\beta)$ to be a **proper density**. That is, it needs to be nonnegative and must integrate to 1. It is clearly nonnegative since $F(x_i'\beta)$ is a distribution function. For integration, note that we only have two values: 0 and 1. So

$$\begin{aligned} f(y_i = 0|x_i;\beta) + f(y_i = 1|x_i;\beta) &= F(x_i'\beta)^0\big(1-F(x_i'\beta)\big)^1 + F(x_i'\beta)^1\big(1-F(x_i'\beta)\big)^0 \\ &= \big(1-F(x_i'\beta)\big) + F(x_i'\beta) \\ &= 1. \end{aligned}$$

  We also need uniqueness of β_0, which means we need $f(y_i|x_i;\beta)$ to be the **true conditional density**. One way of stating this is to assume that $y_i \sim Bernoulli\big(F(x_i'\beta_0)\big)$

i.i.d. for unique β_0. Logit and probit already satisfy these criteria.

- **ULLN.** Same stuff as before.

  (a) **i.i.d**. Just state it.

  (b) **Compactness.** Just state it.

  (c) **Continuity.** $f(y_i|x_i;\theta) > 0$ for all (y_i, x_i') and θ. Furthermore, it is continuous in θ for all (y_i, x_i'). We **assume** continuity. We **assume** that F is strictly positive, from which it follows that f is also strictly positive.

  (d) **Measurability.** f is measurable in (y_i, x_i') for all θ. Just state it.

  (e) **Dominance.** Exists $d(y_i, x_i)$ such that $\left|\log\left(f(y_i|x_i;\theta)\right)\right| \leq d(y_i, x_i)$ for all θ, and $E[d] < \infty$.

Continuity is trivially satisfied by both probit and logit. Dominance is really hard so we just won't do it.

Part (e)

To mean value expand this bad boy, we want to find the score and the Hessian. In our case, summand of the log-likelihood function is the m function, so the derivative of that is the score. We already found it while calculating the first order condition, so

$$s(y_i, x_i; \beta) = x_i f(x_i'\beta)\left[\frac{y_i - F(x_i'\beta)}{F(x_i'\beta)\big(1 - F(x_i'\beta)\big)}\right].$$

The Hessian is going to be... unpleasant. But differentiate the score. It's tedious and stupid to do by hand and I'd rather lose the points than waste my time doing it. Then, as usual, asymptotic normality is expressed as

$$\sqrt{n}(\hat{\beta} - \beta_0) = -\left(\frac{1}{n}\sum_{i=1}^{n} H(y_i, x_i; \beta^*)\right)^{-1} \frac{\sqrt{n}}{n}\sum_{i=1}^{n} s(y_i, x_i; \beta_0).$$

- **Consistency** is required. Been there, done that.
- **Mean Value Validity.** We need $f(y_i|x_i;\beta)$ to be twice continuously differentiable on the interior for all y_i, x_i. And β_0 needs to be interior.
- **Consistency of sample average of Hessian.** Each element of the Hessian is bounded in absolute value of a function $b(y_i, x_i)$ with finite expectation. Just say it. Also need positive definiteness of the (negative) expected value of the Hessian. Just say it.

- **Asymptotic Normality of sample average of score.** Integration and differentiation need to be interchangeable so we can go from

$$\frac{\partial E[m(y_i, x_i; \theta)]}{\partial \theta} = E\left[\frac{\partial m(y_i, x_i; \beta)}{\partial \theta}\right] = E[s(y_i, x_i; \theta)],$$
$$\frac{\partial E[s(y_i, x_i; \theta)]}{\partial \theta'} = E\left[\frac{\partial s(y_i, x_i; \beta)}{\partial \theta'}\right] = E[H(y_i, x_i; \theta)].$$

  This can just be stated.

Then the normality is

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, E[H]^{-1}E[ss']E[H]^{-1}\right).$$

The unconditional information equality says that

$$B = E[ss'] = -E[H] = A.$$

So hell, we can simplify it to just

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, E[ss']^{-1}\right).$$

Exercise 4

Suppose i.i.d. $y_i = Bernoulli\big(\Lambda(x_i\beta_1 + z_i\beta_2)\big)$.

Part (a)

The density is

$$p(y_i|x_i) = \begin{cases} \Lambda(x_i\beta_1 + z_i\beta_2) & \text{if } y_i = 1, \\ 1 - \Lambda(x_i\beta_1 + z_i\beta_2) & \text{if } y_i = 0, \end{cases}$$

which can also be written as

$$\Lambda(x_i\beta_1 + z_i\beta_2)^{y_i}\big(1 - \Lambda(x_i\beta_1 + z_i\beta_2)\big)^{1-y_i},$$

which therefore gives a log-likelihood function of

$$\ell = \sum_{i=1}^{n} y_i \log\big(\Lambda(x_i\beta_1 + z_i\beta_2)\big) + (1 - y_i)\log\big(1 - \Lambda(x_i\beta_1 + z_i\beta_2)\big).$$

Part (b)

We are told that $\lambda(z) = \Lambda(z)\big(1 - \Lambda(z)\big)$. Now recall that the information matrix is the *expectation of the outer product of the score.* So we need the score function. In this case, it is the derivative of the summand of the log-likelihood. We have two parameters, so let's first take the derivative with respect to β_1. We get

$$\begin{aligned} s_i(\beta) &= \frac{y_i}{\Lambda(x_i\beta_1 + z_i\beta_2)}\lambda(x_i\beta_1 + z_i\beta_2)x_i + \frac{1-y_i}{1-\Lambda(x_i\beta_1 + z_i\beta_2)}\big(-\lambda(x_i\beta_1 + z_i\beta_2)\big)x_i \\ &= \frac{y_i\Lambda(x_i\beta_1 + z_i\beta_2)\big(1-\Lambda(x_i\beta_1 + z_i\beta_2)\big)}{\Lambda(x_i\beta_1 + z_i\beta_2)}x_i - \frac{(1-y_i)\Lambda(x_i\beta_1 + z_i\beta_2)\big(1-\Lambda(x_i\beta_1 + z_i\beta_2)\big)}{1-\Lambda(x_i\beta_1 + z_i\beta_2)}x_i \\ &= y_i\big(1-\Lambda(x_i\beta_1 + z_i\beta_2)\big)x_i - (1-y_i)\Lambda(x_i\beta_1 + z_i\beta_2)x_i \\ &= [y_i - \Lambda(x_i\beta_1 + z_i\beta_2)]x_i. \end{aligned}$$

This is all very symmetric, so we can write

$$I(\beta_1,\beta_2) = E\left[\begin{pmatrix} [y_i - \Lambda(x_i\beta_1 + z_i\beta_2)]x_i \\ [y_i - \Lambda(x_i\beta_1 + z_i\beta_2)]z_i \end{pmatrix}\begin{pmatrix} [y_i - \Lambda(x_i\beta_1 + z_i\beta_2)]x_i & [y_i - \Lambda(x_i\beta_1 + z_i\beta_2)]z_i \end{pmatrix}\right].$$

With the constraint that $\beta_1 = \beta_2$, the log-likelihood function is instead

$$\ell_R = \sum_{i=1}^n y_i \log\big(\Lambda(x_i\beta_1 + z_i\beta_1)\big) + (1-y_i)\log\big(1-\Lambda(x_i\beta_1 + z_i\beta_1)\big).$$

So then the score function is

$$s = \big(y_i - \Lambda(x_i\beta_1 + z_i\beta_1)\big)(x_i + z_i),$$

and the information matrix is

$$I(\beta_1) = E\left[\big(y_i - \Lambda(x_i\beta_1 + z_i\beta_1)\big)^2(x_i + z_i)^2\right].$$

Part (c)

According to the information matrix equality, the asymptotic variance of $\sqrt{n}(\hat\beta - \beta_0)$ is equal to the inverse of the information matrix. So an estimate can be found by taking the inverse sample analogue of the information matrix evaluated at hats. The unconstrained version is

$$\widehat{\text{Avar}}\big(\sqrt{n}(\hat\beta - \beta_0)\big) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n y_i - \Lambda(x_i\hat\beta_1 + z_i\hat\beta_2)x_i \\ \frac{1}{n}\sum_{i=1}^n y_i - \Lambda(x_i\hat\beta_1 + z_i\hat\beta_2)z_i. \end{pmatrix}^{-1}.$$

The estimator of $\widehat{\text{Avar}}(\hat{\beta})$ is to just divide $\widehat{\text{Avar}}\big(\sqrt{n}(\hat{\beta}-\beta_0)\big)$ by n, so

$$\widehat{\text{Avar}}\big(\hat{\beta}\big) = \begin{pmatrix} \sum_{i=1}^n y_i - \Lambda(x_i\hat{\beta}_1 + z_i\hat{\beta}_2)x_i \\ \sum_{i=1}^n y_i - \Lambda(x_i\hat{\beta}_1 + z_i\hat{\beta}_2)z_i. \end{pmatrix}^{-1}.$$

The constrained case looks similar,

$$\widehat{\text{Avar}}(\hat{\beta}_1) = \left[\sum_{i=1}^n \big(y_i - \Lambda(x_i\hat{\beta}_{1R} + z_i\hat{\beta}_{1R})\big)^2 (x_i+z_i)^2\right]^{-1},$$

where β_{1R} is the maximizer of ℓ_R, the restricted log-likelihood function.

Part (d)

We're testing the null hypothesis of $H_0 = \beta_1 + \beta_2 = 1$.

Likelihood Ratio Test. First, rewrite the hypothesis as $H_0 : \beta_2 = 1 - \beta_1$. We'll need both the constrained and unconstrained estimators for this one. The unconstrained log-likelihood is still

$$\ell = \sum_{i=1}^n y_i \log\big(\Lambda(x_i\beta_1 + z_i\beta_2)\big) + (1-y_i)\log\big(1 - \Lambda(x_i\beta_1 + z_i\beta_2)\big).$$

This gives $\hat{\beta}_1$ and $\hat{\beta}_2$. For the constrained problem, we have

$$\tilde{\ell}_R = \sum_{i=1}^n y_i \log\big(\Lambda(x_i\beta_1 + z_i[1-\beta_1]\big) + (1-y_i)\log\big(1 - \Lambda(x_i\beta_1 + z_i[1-\beta_1]\big),$$

which gives rise to $\tilde{\beta}_{1R}$. The likelihood ratio statistic is then

$$2\ln\big(\ell(\hat{\beta}_1, \hat{\beta}_2) - \tilde{\ell}(\tilde{\beta}_{1R})\big) \xrightarrow{d} \mathcal{X}_1^2.$$

Lagrange Multiplier. First, evaluate the sample average of the unrestricted score, evaluated at the restricted estimates, and multiply it by $\sqrt{n}$.

$$S = \sqrt{n}\frac{1}{n}\sum_{i=1}^n s_i(\hat{\beta}_{1R}, 1-\hat{\beta}_{1R}) = \sqrt{n}\frac{1}{n}\sum_{i=1}^n \begin{pmatrix} y_i - \Lambda(x_i\hat{\beta}_{1R} + z_i[1-\hat{\beta}_{1R}])x_i \\ y_i - \Lambda(x_i\hat{\beta}_{1R} + z_i[1-\hat{\beta}_{1R}])z_i \end{pmatrix}.$$

Second, let $c(\beta) = \beta_1 + \beta_2 - 1$, which is essentially the hypothesis. Take the derivative with respect to β' and you get $\partial c / \partial \beta' = [1 \quad 1]$; call this C.

Third, get an estimate of $A_0 = -E[H]$ evaluated at the constrained estimates:

$$A_{nR} = -\frac{1}{n}\sum_{i=1}^{n} H(y_i, x_i, z_i; \hat{\beta}_{1R}, 1-\hat{\beta}_{1R}).$$

Then the score statistic is the following monstrosity:

$$LM = S' A_{nR}^{-1}\left[\widehat{\text{Avar}}\left(CA_{nR}^{-1}\sqrt{n}\frac{1}{n}S\right)\right]^{-1} CA_{nR}^{-1}S \xrightarrow{d} \mathcal{X}_1^2.$$

Yeah, good luck memorizing that thing.

Wald Test. The general formula is

$$nc(\hat{\beta})'\left[C\hat{A}_n^{-1}\hat{B}_n\hat{A}_n^{-1}C'\right]^{-1} c(\hat{\beta}).$$

We already know some of these things, so

$$n(\hat{\beta}_1+\hat{\beta}_2-1)\left[[1 \quad 1]\hat{A}_n^{-1}\hat{B}_n\hat{A}_n^{-1}[1 \quad 1]'\right]^{-1}(\hat{\beta}_1+\hat{\beta}_2-1).$$

And then let's use the information equality where we can replace $-\hat{A}_n^{\;-1} = \hat{B}_n$, so

$$n(\hat{\beta}_1+\hat{\beta}_2-1)\left[[1 \quad 1]\hat{B}_n^{-1}[1 \quad 1]'\right]^{-1}(\hat{\beta}_1+\hat{\beta}_2-1) \xrightarrow{d} \mathcal{X}_1^2.$$

Part (e)

Now we have the function $c(\beta) = (\beta_1+\beta_2)^3 - 1$ instead, which gives a different Jacobian, namely,

$$C = \begin{pmatrix} 3(\beta_1+\beta_2)^2 & 3(\beta_1+\beta_2)^2 \end{pmatrix}.$$

And therefore the Wald estimate becomes

$$n\big((\hat{\beta}_1-\hat{\beta}_2)^3-1\big)\left[\begin{pmatrix}3(\beta_1+\beta_2)^2 & 3(\beta_1+\beta_2)^2\end{pmatrix}\hat{B}_n^{-1}\begin{pmatrix}3(\beta_1+\beta_2)^2 & 3(\beta_1+\beta_2)^2\end{pmatrix}'\right]^{-1}\big((\hat{\beta}_1-\hat{\beta}_2)^3-1\big) \xrightarrow{d} \mathcal{X}_1^2,$$

which is clearly a different thing, even though the hypothesis itself is essentially the same expression.

Part (e)

The LM statistic and the LR statistic are invariant, so nothing would change.