PS5 Solutions

Jingle Fu

Problem 1

The dataset `dat_SalesCustomers.csv` contains data on sales of shopping malls in Istanbul. It includes the following variables: *invoice no* (identifier of transaction or invoice), *customer id* (identifier of customer), *category* (type of goods sold), *price* (in TRY, Turkish Lira), *invoice date*, *shopping mall*, *gender*, *age*, and *payment method* (cash vs. credit card vs. debit card payment).

You are interested in shedding light on the determinants of cash vs card payment. For this purpose, you set up a probit model:

$$y_i^* = x_i'\beta + u_i \quad u_i \mid x_i \sim N(0,1) \tag{1}$$

whereby we observe $y_i = \mathbf{1}\{y_i^* > 0\}$, a dummy variable for cash payment. Recall that the Maxi- mum Likelihood (ML) estimator for β solves

$$\hat{\beta} = \arg\min_\beta Q_n(\beta; Z_n) \quad Q_n(\beta; Z_n) = -\frac{1}{n}\ell(\beta; Z_n) \tag{2}$$

where

$$\ell(\beta; Z_n) = \sum_{i=1}^n \left[y_i \log(\Phi(x_i'\beta)) + (1 - y_i) \log(\Phi(-x_i'\beta)) \right]$$

is the log-likelihood and $Z_n = \{y_i, x_i\}_{i=1}^n$ comprises all of the data you have available (outcome-variables and covariates for the n observations in your sample).

(a)

Are there missing values in your data? Delete all observations with a missing value in the variables *category*, *price*, *gender*, *age* or *payment method*. How many observations do you have left?

(b)

Based on the variable *payment method*, generate a dummy variable for cash payment and call it *paid in cash*. Also, based on *gender*, create a dummy for males, *male*. What fraction

of transactions were carried out in cash? What fraction of the overall sales (in TRY) were carried out in cash?

(c)

To decrease computational costs, consider only the first $n = 1000$ observations for the following questions. Based on the variable category, create a dummy for each of the following four categories: i) clothes and shoes, ii) cosmetics, iii) food, iv) technology. In this way, we divide the categories into five groups, whereby the fifth is made up by the rest, i.e. goods that do not belong to either of the four categories. How are the transactions split across these five categories? How are the sales split across these five categories?

(d)

Taking *paid in cash* as your outcome variable y_i and *price, male, age* and all category-dummies but one as your covariates x_i, use a numerical optimization-command from the software of your choice to solve the optimization problem in Eq. (2) and obtain $\hat{\beta}$ for your sample. If manual optimization does not work, you can use a pre-programmed command to estimate the probit model.

(e)

Based on your estimate, compute the effect of age increasing by 5 years on the expected probability of using cash for a 30 year-old male who bought clothes for 500 TRY, i.e. for an observation with $x_i = x_i^* \equiv [500, 1, 30, 0, \ldots, 0, 1, 0, \ldots, 0]$. Put differently, this is the difference in expected probabilities of cash payment between a 60 year-old and a 30 year-old male who bought clothes/shoes for 500 TRY. We will call this quantity $\gamma_1(\hat{\beta})$. Also, compute the same effect without conditioning on the category of goods sold in two steps: (i) compute the effect for each of the five categories and (ii) take a weighted average of them, with weights given by the proportions of these goods-categories in overall sales (see your answer to (c)). We will call this quantity $\gamma_2(\hat{\beta})$.

(f)

Suppose that your probit model in Eq.(1) is correctly specified. Is your estimator $\hat{\beta}$ consistent? Use the simplified version of the extremum estimation theory we discussed in class to answer this question.

(g)

Use bootstrapping to find a numerical approximation of the finite sample distribution of $\hat{\beta}$ as well as the two marginal effects $\gamma_1(\hat{\beta})$ and $\gamma_2(\hat{\beta})$: draw $M = 100$ different samples of

n observations with replacement from your dataset and compute (numerically) $\hat{\beta}$, $\gamma_1(\hat{\beta})$ and $\gamma_2(\hat{\beta})$ for each of them. Plot the finite sample distributions you obtained (regarding $\hat{\beta}$, you can limit yourself to the coefficient on age).

(h)

Another approach to approximate the finite sample distribution of $\hat{\beta}$ and functions of it like the marginal effects is to use their asymptotic distribution. Use the simplified version of the extremum estimation theory we discussed in class to show that the asymptotic distribution of $\hat{\beta}$ is given by

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N\left(0, H^{-1}\right),$$

with

$$H = E\left[\frac{\phi(x_i'\beta_0)^2}{\Phi(x_i'\beta_0)\Phi(-x_i'\beta_0)} x_i x_i'\right]. \tag{3}$$

Then, use the asymptotic distribution in Eq. (3) to approximate the finite sample distribution of $\hat{\beta}$ in your sample. How does this approximate finite sample distribution of the estimated coefficient on age compare to the one obtained via bootstrapping?

Hint: The numerator and the denominator in the fraction that appears in H are often both very close to zero. Rather than computing it as-is, first compute the log of it and then take the exponential, i.e. compute

$$\frac{\phi(x_i'\beta_0)^2}{\Phi(x_i'\beta_0)\Phi(-x_i'\beta_0)} \text{ as } \exp\left\{2\log\phi(x_i'\beta_0) - \log\Phi(x_i'\beta_0) - \log\Phi(-x_i'\beta_0)\right\}$$

(i)

Use the asymptotic distribution of $\hat{\beta}$ from Eq. (3) and the Delta method to find the asymptotic distribution of $\gamma_1(\hat{\beta})$. Then, use it to approximate the finite sample distribution of $\gamma_1(\hat{\beta})$ in your sample. How does this approximate finite sample distribution compare to the one obtained via bootstrapping?

(j)

Now let's test whether the true partial effect $\gamma_1(\hat{\beta})$ (i.e. the true change in the expected probability of cash payment for a 30 year-old male buying clothes for 500 TRY when this individual becomes 5 years older) is significantly different from 0 at the $\alpha = 0.05$ level:

$$H_0 : \gamma_1(\beta) = 0 \quad \text{vs.} \quad H_1 : \gamma_1(\beta) \neq 0.$$

(In other words, we are testing whether the expected probabilities of cash payment for a 30 year-old and a 35 year-old male buying clothes for 500 TRY are different.) One approach

to do so uses the finite sample distribution of $\gamma_1(\hat{\beta})$ approximated via its asymptotic distribution,

$$\gamma_1(\hat{\beta}) \overset{approx}{\sim} N\left(\gamma_1(\beta), \frac{1}{n}\hat{V}\right),$$

for some $\hat{V}$ you had to find. Use this expression to construct a t-test. What do you conclude?

Also, use the above expression to construct a 95% confidence interval for $\gamma_1(\beta)$. (If you couldn't find $\hat{V}$, just state the test statistic and critical value for a general $\hat{V}$.)

Solution (a).

Yes, here missing values in the data set. The initial number of observations is 99457, The dataset's reported missing values indicate that only *age* has missing observations (119 missing values). After dropping these, we end up with 99338 observations.

```r
rm(list = ls())
library(tidyverse)
library(ggplot2)
library(dplyr)
library(broom)
library(stats)
library(stargazer)
library(car)


set.seed(2024)


dat <- read.csv("dat_SalesCustomers.csv")
variables_to_check <- c("category", "price", "gender", "age", "payment_
    method")
missing_counts <- sapply(dat[variables_to_check], function(x) sum(is.na(
    x)))
print("Number of missing values in each variable:")
print(missing_counts)


dat_clean <- dat[complete.cases(dat[variables_to_check]), ]


num_observations <- nrow(dat_clean)
print(paste("Number of observations after removing missing values:", num
    _observations))

```

Solution (b).

Define

$$\text{paid\_in\_cash}_i = \mathbf{1}\{\text{payment\_method}_i = \text{Cash}\}$$

$$\text{male}_i = \mathbf{1}\{\text{gender}_i = \text{Male}\}$$

The fraction of transactions carried out in cash is

$$\frac{1}{n} \sum_{i=1}^{n} \text{paid\_in\_cash}_i.$$

Empirically, this is about 44.69%.

The fraction of overall sales carried out in cash is

$$\frac{\sum_{i=1}^{n} \text{paid\_in\_cash}_i \cdot price_i}{\sum_{i=1}^{n} price_i}.$$

Empirically, this fraction is about 44.79%.

These results indicate that cash payments represent nearly half of all transactions and sales value.

```r
dat_clean$paid_in_cash <- ifelse(dat_clean$payment_method == "Cash", 1,
    0)
dat_clean$male <- ifelse(dat_clean$gender == "Male", 1, 0)
fraction_cash_transactions <- mean(dat_clean$paid_in_cash)
print(paste("Fraction of transactions carried out in cash:",
            round(fraction_cash_transactions * 100, 2), "%"))

total_sales <- sum(dat_clean$price)
cash_sales <- sum(dat_clean$price[dat_clean$paid_in_cash == 1])
fraction_cash_sales <- cash_sales / total_sales
print(paste("Fraction of overall sales carried out in cash:",
            round(fraction_cash_sales * 100, 2), "%"))

```

Solution (c).

We now consider only the first $n = 1000$ observations. Let the categories be divided into five mutually exclusive groups: Clothes and Shoes (C), Cosmetics (Cos), Food (F), Technology (T), and Other (O). Define indicator variables:

$$d_{C,i} = \mathbf{1}\{\text{category}_i = \text{Clothes and Shoes}\}$$
$$d_{Cos,i} = \mathbf{1}\{\text{category}_i = \text{Cosmetics}\}$$
$$d_{F,i} = \mathbf{1}\{\text{category}_i = \text{Food}\}$$
$$d_{T,i} = \mathbf{1}\{\text{category}_i = \text{Technology}\}$$
$$d_{O,i} = 1 - (d_{C,i} + d_{Cos,i} + d_{F,i} + d_{T,i}).$$

The fraction of transactions in category j is

$$\frac{1}{1000} \sum_{i=1}^{1000} d_{j,i}.$$

The fraction of sales in category j is

$$\frac{\sum_{i=1}^{1000} d_{j,i} \cdot price_i}{\sum_{i=1}^{1000} price_i}.$$

Empirically:

- Transactions fraction: Clothes/Shoes: 43.8%, Cosmetics: 14.8%, Food: 14.0%, Technology: 5.0%, Other: 22.4%.

- Sales fraction: Clothes/Shoes: 70.58%, Cosmetics: 2.72%, Food: 0.32%, Technology: 23.9%, Other: 2.49%.

The result shows that most transactions and sales are in the Clothes/Shoes category. Technology, though having the lowest transaction fraction, has the second-highest sales fraction, meaning that it has the highest average price.

```r
dat_1000 <- dat_clean[1:1000, ]

dat_1000$clothes_shoes <- ifelse(dat_1000$category %in% c("Clothing", "
    Shoes"), 1, 0)
dat_1000$cosmetics <- ifelse(dat_1000$category == "Cosmetics", 1, 0)
dat_1000$food <- ifelse(dat_1000$category %in% c("Food", "Food &
    Beverage"), 1, 0)
dat_1000$technology <- ifelse(dat_1000$category == "Technology", 1, 0)

dat_1000$other_category <- ifelse(dat_1000$clothes_shoes + dat_1000$
    cosmetics + dat_1000$food + dat_1000$technology == 0, 1, 0)

all(dat_1000$clothes_shoes + dat_1000$cosmetics + dat_1000$food + dat_
    1000$technology + dat_1000$other_category == 1)

fraction_transactions <- c(
  "Clothes and Shoes" = mean(dat_1000$clothes_shoes),
  "Cosmetics" = mean(dat_1000$cosmetics),
  "Food" = mean(dat_1000$food),
  "Technology" = mean(dat_1000$technology),
  "Other" = mean(dat_1000$other_category)
)
print("Fraction of transactions in each category:")
print(round(fraction_transactions * 100, 2))

total_sales_1000 <- sum(dat_1000$price)
sales_clothes_shoes <- sum(dat_1000$price[dat_1000$clothes_shoes == 1])
sales_cosmetics <- sum(dat_1000$price[dat_1000$cosmetics == 1])
sales_food <- sum(dat_1000$price[dat_1000$food == 1])
sales_technology <- sum(dat_1000$price[dat_1000$technology == 1])
sales_other <- sum(dat_1000$price[dat_1000$other_category == 1])
```

```
29 fraction_sales <- c(
30   "Clothes and Shoes" = sales_clothes_shoes / total_sales_1000,
31   "Cosmetics" = sales_cosmetics / total_sales_1000,
32   "Food" = sales_food / total_sales_1000,
33   "Technology" = sales_technology / total_sales_1000,
34   "Other" = sales_other / total_sales_1000
35 )
36
37 print("Fraction of sales in each category:")
38 print(round(fraction_sales * 100, 2))
39
```

Solution (d).

To find the Maximum Likelihood Estimator (MLE) $\hat{\beta}$, we differentiate the log-likelihood with respect to β. Let $\phi(\cdot)$ denote the standard normal PDF. We use:

$$\frac{d}{dt}\log(\Phi(t)) = \frac{\phi(t)}{\Phi(t)}, \quad \text{and} \quad \frac{d}{dt}\log(1 - \Phi(t)) = -\frac{\phi(t)}{1 - \Phi(t)}.$$

For each element β_j of β, the derivative of the log-likelihood is:

$$\frac{\partial \ell(\beta; Z_n)}{\partial \beta_j} = \sum_{i=1}^{n} \left[y_i \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)} - (1 - y_i)\frac{\phi(x_i'\beta)}{1 - \Phi(x_i'\beta)} \right] x_{ij}.$$

Stacking all partial derivatives together, the gradient (score vector) is:

$$\nabla_\beta \ell(\beta; Z_n) = \sum_{i=1}^{n} \left[\frac{y_i - \Phi(x_i'\beta)}{\Phi(x_i'\beta)(1 - \Phi(x_i'\beta))} \phi(x_i'\beta) \right] x_i.$$

Often written more simply as:

$$\nabla_\beta \ell(\beta; Z_n) = \sum_{i=1}^{n} \left[y_i \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)} - (1 - y_i)\frac{\phi(x_i'\beta)}{1 - \Phi(x_i'\beta)} \right] x_i.$$

Step 1: Characterizing the MLE $\hat{\beta}$

The MLE $\hat{\beta}$ sets the gradient to zero:

$$\nabla_\beta \ell(\hat{\beta}; Z_n) = 0.$$

Substituting back:

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

This is a system of k nonlinear equations in the k unknowns $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_k)'$.

Step 2: No Closed-Form Solution

Unlike in linear regression or the logit model (even the logit doesn't have a closed form), the Probit model does not admit a closed-form solution for $\hat{\beta}$. The equation above must be solved using numerical optimization techniques such as the Newton-Raphson algorithm or other iterative methods.

Step 3: Numerical Optimization

A common iterative procedure is:

Algorithm 1: Newton-Raphson Method

Input: Initialize β_0, tolerence level $\varepsilon > 0$

1 for $m = 1$ **to** M **do**

2   Given β^m, compute $\nabla_\beta \ell(\beta^{(m)}; Z_n)$ and $[H(\beta^{(m)}; Z_n)]$;

3   Set $\beta^{(m+1)} = \beta^{(m)} - [H(\beta^{(m)}; Z_n)]^{-1} \nabla_\beta \ell(\beta^{(m)}; Z_n)$;;

4   **if** $\|\beta^{m+1} - \beta^m\| < \varepsilon$ **then**

5 $\hat{\beta} = \beta^{m+1}$;

6   **else**

7   Proceed to the next iteration;

8   **end**

9 end

---

where $H(\beta; Z_n)$ is the Hessian matrix of second derivatives evaluated at β. Convergence is achieved when changes in β or the norm of the gradient are below a given tolerance. The regression result is as follows:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_{price} \\ \hat{\beta}_{male} \\ \hat{\beta}_{age} \\ \hat{\beta}_{cosmetics} \\ \hat{\beta}_{food} \\ \hat{\beta}_{technology} \end{bmatrix} = \begin{bmatrix} 0.0682 \\ 0.000112 \\ -0.0502 \\ -0.00183 \\ -0.2879 \\ -0.1195 \\ 0.0640 \\ -0.4195 \end{bmatrix}.$$

Interpretation:

- The coefficient on price is positive but very small, suggesting a tiny positive association of price with the probability of cash payment (not statistically significant).

- male is negative, but not significant, suggesting no strong gender effect on the probability of cash usage.

- age coefficient is negative and small, not statistically significant either.

- Some category dummies (like Clothes/Shoes) are significantly different from zero, indicating that the reference category (likely "Other") differs in payment method probability.

Table 1: Optimization model

	Dependent variable:
	paid_in_cash
price	0.0001
	(0.0001)
male	−0.050
	(0.081)
age	−0.002
	(0.003)
clothes_shoes	−0.288**
	(0.130)
cosmetics	−0.120
	(0.133)
food	0.064
	(0.135)
technology	−0.420
	(0.314)
Constant	0.068
	(0.148)
Observations	1,000
Log Likelihood	−685.217
Akaike Inf. Crit.	1,386.434
Note:	*p<0.1; **p<0.05; ***p<0.01

```r
X <- as.matrix(cbind(1, dat_1000[, c("price", "male", "age", "clothes_
    shoes", "cosmetics", "food", "technology")]))
y <- dat_1000$paid_in_cash

neg_log_likelihood <- function(beta, X, y) {
  X_beta <- X %*% beta
  log_phi_Xb <- pnorm(X_beta, log.p = TRUE)
  log_phi_minus_Xb <- pnorm(-X_beta, log.p = TRUE)
  ll <- sum(y * log_phi_Xb + (1 - y) * log_phi_minus_Xb)
  return(-ll)
}

neg_log_likelihood_grad <- function(beta, X, y) {
  X_beta <- X %*% beta
  phi_Xb <- dnorm(X_beta)
  Phi_Xb <- pnorm(X_beta)
  Phi_minus_Xb <- pnorm(-X_beta)
  epsilon <- 1e-16
  Phi_Xb <- pmax(Phi_Xb, epsilon)
  Phi_minus_Xb <- pmax(Phi_minus_Xb, epsilon)
  gradient <- -t(X) %*% ((y * phi_Xb / Phi_Xb) - ((1 - y) * phi_Xb / Phi
    _minus_Xb))
  return(as.vector(gradient))
}

initial_beta <- rep(0, ncol(X))

result <- optim(par = initial_beta, fn = neg_log_likelihood, gr = neg_
    log_likelihood_grad, X = X, y = y, method = "BFGS")

if (result$convergence == 0) {
  cat("Optimization converged.\n")
} else {
  cat("Optimization did not converge.\n")
}

beta_hat <- result$par
print("Estimated coefficients (beta_hat):")
print(beta_hat)

### Programming Method
model <- glm(paid_in_cash ~ price + male + age + clothes_shoes +
    cosmetics + food + technology,
      data = dat_1000, family = binomial(link = "probit"))
stargazer(model, type = "latex", title = "Optimization model", out = "d.
    tex")

beta_hat2 <- coef(model)
```

```
44 print("Estimated coefficients (beta_hat):")
45 print(beta_hat2)
46
```

Solution (e).

We define

$$\gamma_1(\beta) = \Phi(x_2'\beta) - \Phi(x_1'\beta),$$

where x_1' is a vector for a 30-year-old male buying Clothes/Shoes for 500 TRY, and x_2' is the same vector but with age increased to 60 years old. Only the age element of x_i changes from 30 to 60.

For our estimated $\hat{\beta}$,

$$\gamma_1(\hat{\beta}) \approx -0.02096.$$

This suggests that increasing age from 30 to 60 reduces the probability of cash payment by about 2.1 percentage points for this specific profile.

For $\gamma_2(\beta)$, we do not condition on category. We take a weighted average of the partial effects across the five categories, with weights given by their share in total sales:

$$\gamma_2(\hat{\beta}) = \sum_j w_j \left[\Phi(x_{2,j}'\beta) - \Phi(x_{1,j}'\beta) \right],$$

where w_j is the sales fraction for category j.

Empirically,

$$\gamma_2(\hat{\beta}) \approx -0.02077,$$

very close to $\gamma_1(\hat{\beta})$, indicating a similar overall effect once categories are averaged by their sales importance.

```
1  x_age_30 <- c(1, 500, 1, 30, 1, 0, 0, 0)
2
3  x_age_60 <- x_age_30
4  x_age_60[4] <- 60  # Update age to 35
5
6  prob_age_30 <- pnorm(sum(x_age_30 * beta_hat))
7  prob_age_60 <- pnorm(sum(x_age_60 * beta_hat))
8
9  gamma_1 <- prob_age_60 - prob_age_30
10 print(paste("Gamma_1 (effect of age increasing by 5 years):", gamma_1))
11
12 gamma_c <- numeric(length(fraction_sales))
13 names(gamma_c) <- names(fraction_sales)
14
15 for (cat in names(fraction_sales)) {
16   clothes_shoes <- ifelse(cat == "Clothes and Shoes", 1, 0)
17   cosmetics <- ifelse(cat == "Cosmetics", 1, 0)
18   food <- ifelse(cat == "Food", 1, 0)
```

```
19   technology <- ifelse(cat == "Technology", 1, 0)
20
21   x_age_30_2 <- c(1, 500, 1, 30, clothes_shoes, cosmetics, food,
       technology)
22   x_age_60_2 <- x_age_30_2
23   x_age_60_2[4] <- 60  # Update age to 35
24
25   prob_age_30_2 <- pnorm(sum(x_age_30_2 * beta_hat))
26   prob_age_60_2 <- pnorm(sum(x_age_60_2 * beta_hat))
27
28   gamma_c[cat] <- prob_age_60_2 - prob_age_30_2
29 }
30
31 gamma_2 <- sum(fraction_sales * gamma_c)
32 print(paste("Gamma_2 (weighted effect over categories):", gamma_2))
33
```

Solution (f).

Consider the linear model

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

where $u_i \mid x_i \sim N(0, 1)$.

Step 1: Define the Objective Function

Define $\mathcal{B} = \{\beta \in \mathbb{R} : \|\beta\| \leq c\}$ for some very large c.

The objective function is given by:

$$\hat{\beta} = \arg\min_{\beta \in \mathcal{B}} Q_n(\beta) = \arg\min_{\beta \in \mathcal{B}} \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i'\beta)^2$$

Step 2: Express the Limiting Behavior of the Objective Function

The expected form of the objective function $Q_n(\beta)$ is:

$$\begin{aligned}
Q(\beta) &= E[(y_i - x_i'\beta)^2] \\
&= E[(x_i'\beta_0 + u_i - x_i'\beta)^2] \\
&= E[(x_i'(\beta_0 - \beta) + u_i)^2] \\
&= E[(x_i'(\beta_0 - \beta))^2] + 2E[x_i'(\beta_0 - \beta)u_i] + E[u_i^2] \\
&= E[(x_i'(\beta_0 - \beta))^2] + \sigma^2 \\
&= (\beta_0 - \beta)'E[x_i x_i'](\beta_0 - \beta) + 1
\end{aligned}$$

since $E[u_i \mid x_i] = 0$ and $\sigma^2 = 1$.

Step 3: Show Minimization of $Q(\beta)$ at β_0

As $E[x_i x_i']$ is positive definite, the function $Q(\beta)$ is minimized at β_0 because:

$$(\beta_0 - \beta)'E[x_i x_i'](\beta_0 - \beta) \geq 0$$

which is zero if and only if $\beta = \beta_0$.

Step 4: Prove Uniform Convergence of $Q_n(\beta)$ to $Q(\beta)$

It's obvious that $m(x_i, y_i, \beta) = (y_i - x_i'\beta)^2$ satisfies the first three conditions of the Uniform Law of Large Numbers(ULLN).

We then prove the fourth one:

$$E\left[\sup_{\beta \in \mathcal{B}} \|m(x_i, y_i, \beta)\|\right] \leq E\left[|y_i|^2\right] + \sup_{\beta \in \mathcal{B}} 2E\left[|y_i|\,|x_i|\,\|\beta\|\right] + \sup_{\beta \in \mathcal{B}} E\left[|x_i|^2\,\|\beta^2\|\right]$$
$$< \infty$$

By ULLN:

$$Q_n(\beta) \xrightarrow{p} Q(\beta), n \to \infty$$

Step 5: Demonstrate the Consistency of $\hat{\beta}_n$

According to extremum estimator theory, if $Q_n(\beta)$ converges uniformly to $Q(\beta)$ and $Q(\beta)$ has a unique global minimum at β_0, then:

$$\hat{\beta}_n \xrightarrow{p} \beta_0$$

First, we define $m(w_i; \theta) = y_i \log(\Phi(x_i'\beta)) + (1 - y_i)\log(\Phi(-x_i'\beta))$.

Step 1: Uniform convergence

By the Uniform Law of Large Numbers (ULLN), we have that

$$\frac{1}{n}\sum_{i=1}^{n} m(w_i; \theta) \xrightarrow{p} E[m(w_i; \theta)].$$

Solution (g).

```
1  M <- 100
2  n <- nrow(dat_1000)
3  beta_age_bootstrap <- numeric(M)
4  gamma_1_bootstrap <- numeric(M)
5  gamma_2_bootstrap <- numeric(M)
6  set.seed(2024)
7
8  for (m in 1:M) {
9     indices <- sample(1:n, size = n, replace = TRUE)
10    dat_bootstrap <- dat_1000[indices, ]
11
12    model_boot <- glm(paid_in_cash ~ price + male + age + clothes_shoes +
        cosmetics + food + technology, data = dat_bootstrap, family =
        binomial(link = "probit"))
13    beta_hat_boot <- coef(model_boot)
14
15    beta_age_bootstrap[m] <- beta_hat_boot["age"]
16
```

```
17    x_age_30 <- c(1, 500, 1, 30, 1, 0, 0, 0)
18    x_age_60 <- x_age_30
19    x_age_60[4] <- 60   # Update age to 35
20    prob_age_30 <- pnorm(sum(x_age_30 * beta_hat_boot))
21    prob_age_60 <- pnorm(sum(x_age_60 * beta_hat_boot))
22    gamma_1_bootstrap[m] <- prob_age_60 - prob_age_30
23
24    gamma_c <- numeric(length(fraction_sales))
25    names(gamma_c) <- names(fraction_sales)
26
27    for (cat in names(fraction_sales)) {
28      clothes_shoes <- ifelse(cat == "Clothes and Shoes", 1, 0)
29      cosmetics <- ifelse(cat == "Cosmetics", 1, 0)
30      food <- ifelse(cat == "Food", 1, 0)
31      technology <- ifelse(cat == "Technology", 1, 0)
32
33      x_age_30_2 <- c(1, 500, 1, 30, clothes_shoes, cosmetics, food,
      technology)
34      x_age_60_2 <- x_age_30
35      x_age_60_2[4] <- 60   # Update age to 35
36
37      prob_age_30_2 <- pnorm(sum(x_age_30_2 * beta_hat_boot))
38      prob_age_60_2 <- pnorm(sum(x_age_60_2 * beta_hat_boot))
39      gamma_c[cat] <- prob_age_60_2 - prob_age_30_2
40    }
41
42    gamma_2_bootstrap[m] <- sum(fraction_sales * gamma_c)
43  }
44
45  hist(beta_age_bootstrap, main = "Bootstrap Distribution of Coefficient
      on Age", xlab = "Coefficient on Age", breaks = 20)
46  hist(gamma_1_bootstrap, main = "Bootstrap Distribution of Gamma_1", xlab
      = "Gamma_1", breaks = 20)
47  hist(gamma_2_bootstrap, main = "Bootstrap Distribution of Gamma_2", xlab
      = "Gamma_2", breaks = 20)
48
```

Solution (h).

For a sample $\{(y_i, x_i)\}_{i=1}^{n}$, the log-likelihood function is:

$$\ell(\beta; Z_n) = \sum_{i=1}^{n} \left[y_i \log(\Phi(x_i'\beta)) + (1 - y_i) \log(1 - \Phi(x_i'\beta)) \right].$$

To derive the Hessian, focus on a single observation i and then we will take expectations:

$$\ell_i(\beta) = y_i \log(\Phi(t_i)) + (1 - y_i) \log(1 - \Phi(t_i)), \quad t_i = x_i'\beta.$$

First Derivative (Score):

First, take the derivative with respect to t_i:

$$\frac{\partial \ell_i(\beta)}{\partial t_i} = y_i \frac{\phi(t_i)}{\Phi(t_i)} - (1 - y_i)\frac{\phi(t_i)}{1 - \Phi(t_i)},$$

where $\phi(\cdot)$ is the standard normal PDF.

Combine the fractions:

$$\frac{\partial \ell_i(\beta)}{\partial t_i} = \phi(t_i)\left[\frac{y_i}{\Phi(t_i)} - \frac{1 - y_i}{1 - \Phi(t_i)}\right].$$

Find a common denominator $\Phi(t_i)(1 - \Phi(t_i))$:

$$\frac{\partial \ell_i(\beta)}{\partial t_i} = \frac{\phi(t_i)}{\Phi(t_i)(1 - \Phi(t_i))}(y_i - \Phi(t_i)).$$

Since $y_i - \Phi(t_i) = y_i - p_i$, we have:

$$\frac{\partial \ell_i(\beta)}{\partial t_i} = \frac{\phi(t_i)}{p_i(1 - p_i)}(y_i - p_i).$$

To get the gradient w.r.t. β, use chain rule:

$$\frac{\partial \ell_i(\beta)}{\partial \beta} = \frac{\partial \ell_i(\beta)}{\partial t_i} x_i = \frac{\phi(t_i)}{p_i(1 - p_i)}(y_i - p_i)x_i.$$

This is the score vector for a single observation.

Second Derivative (Hessian):

Now differentiate again with respect to β:

$$\frac{\partial^2 \ell_i(\beta)}{\partial \beta \partial \beta'} = \frac{\partial}{\partial \beta}\left(\frac{\phi(t_i)}{p_i(1 - p_i)}(y_i - p_i)x_i\right).$$

Since $t_i = x_i'\beta$, $\frac{\partial t_i}{\partial \beta} = x_i$. Thus, second derivatives w.r.t. β come through differentiating w.r.t. t_i, then applying chain rule again:

$$\frac{\partial^2 \ell_i(\beta)}{\partial \beta \partial \beta'} = \left(\frac{\partial^2 \ell_i(\beta)}{\partial t_i^2}\right) x_i x_i'.$$

So the main task is to find:
$$\frac{\partial^2 \ell_i(\beta)}{\partial t_i^2}.$$

We have:

$$\frac{\partial \ell_i(\beta)}{\partial t_i} = \frac{\phi(t_i)}{p_i(1 - p_i)}(y_i - p_i).$$

Take the derivative w.r.t. t_i:

$$\frac{\partial^2 \ell_i(\beta)}{\partial t_i^2} = \frac{\partial}{\partial t_i} \left[\frac{\phi(t_i)}{p_i(1 - p_i)}(y_i - p_i) \right].$$

This involves the product rule and quotient rule. However, *the key simplification occurs when we take expectations at the true parameter* β_0. Under the true model, $E[y_i] = p_i$, so $E[y_i - p_i] = 0$. Terms involving $(y_i - p_i)$ vanish when taking expectation.

At the true parameter, the Fisher information (which is $-E[\partial^2 \ell_i(\beta_0)/\partial\beta\partial\beta']$) simplifies dramatically. Instead of going through the full complex algebra of the second derivative in the y_i form, we use the known result from standard Probit derivations:

Under correct specification, the expected Hessian w.r.t. t_i at β_0 is known to be:

$$E\left[\frac{\partial^2 \ell_i(\beta_0)}{\partial t_i^2} \right] = -\frac{\phi(t_i)^2}{p_i(1 - p_i)}.$$

Thus:

$$E\left[\frac{\partial^2 \ell_i(\beta_0)}{\partial\beta\partial\beta'} \right] = E\left[\frac{\partial^2 \ell_i(\beta_0)}{\partial t_i^2} x_i x_i' \right] = E\left[-\frac{\phi(t_i)^2}{p_i(1 - p_i)} x_i x_i' \right].$$

Multiplying by -1, the Fisher Information matrix (which is H in the problem) is:

$$H = E\left[\frac{\phi(x_i'\beta_0)^2}{\Phi(x_i'\beta_0)[1 - \Phi(x_i'\beta_0)]} x_i x_i' \right].$$

Since $1 - \Phi(t_i) = \Phi(-t_i)$:

$$H = E\left[\frac{\phi(x_i'\beta_0)^2}{\Phi(x_i'\beta_0)\Phi(-x_i'\beta_0)} x_i x_i' \right].$$

```r
X <- as.matrix(cbind(1, dat_1000[, c("price", "male", "age", "clothes_
    shoes", "cosmetics", "food", "technology")]))
X_beta_hat <- X %*% beta_hat

log_phi_Xb <- dnorm(X_beta_hat, log = TRUE)
log_Phi_Xb <- pnorm(X_beta_hat, log.p = TRUE)

log_Phi_minus_Xb <- pnorm(-X_beta_hat, log.p = TRUE)
log_factor <- 2 * log_phi_Xb - log_Phi_Xb - log_Phi_minus_Xb
factor <- exp(log_factor)
factor[!is.finite(factor)] <- 0
factor <- as.vector(factor)
X_weighted <- sweep(X, 1, factor, FUN = "*")

H_hat <- t(X) %*% X_weighted / nrow(X)
V_hat <- solve(H_hat)

variance_beta_age <- V_hat[4, 4]
```

```r
18  beta_age_sd <- sqrt(variance_beta_age / nrow(X))
19
20  if (!is.finite(beta_age_sd)) {
21    stop("Standard error for the coefficient on age is not finite.")
22  }
23
24  mean_age <- beta_hat[4]
25  simulated_draws <- rnorm(1000, mean = mean_age, sd = beta_age_sd)
26
27  library(ggplot2)
28
29  bootstrap_data <- data.frame(Distribution = "Bootstrap", Values = beta_
        age_bootstrap)
30  asymptotic_data <- data.frame(Distribution = "Asymptotic", Values =
        simulated_draws)
31  combined_data <- rbind(bootstrap_data, asymptotic_data)
32
33  ggplot(combined_data, aes(x = Values, fill = Distribution)) +
34    geom_histogram(aes(y = ..density..),
35                   bins = 20, alpha = 1, position = "identity", color = "
        black") +
36    scale_fill_manual(values = c("Bootstrap" = "grey", "Asymptotic" = "red
        ")) +
37    labs(title = "Comparison of Bootstrap and Asymptotic Distributions",
38         x = "Coefficient on Age", y = "Density") +
39    theme_minimal() +
40    theme(plot.title = element_text(hjust = 0.5, size = 14),
41          legend.title = element_blank(),
42          legend.position = "topright") +
43    guides(fill = guide_legend(reverse = TRUE))
44
```

Solution (i).

The Delta Method states that if

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, H^{-1}),$$

and $g(\cdot)$ is a continuously differentiable function at β_0, then

$$\sqrt{n}(g(\hat{\beta}) - g(\beta_0)) \xrightarrow{d} N(0, \nabla_\beta g(\beta_0)' H^{-1} \nabla_\beta g(\beta_0)).$$

In our case, $g(\beta) = \gamma_1(\beta)$.

Computing the Gradient $\nabla_\beta \gamma_1(\beta)$:

We have:

$$\gamma_1(\beta) = \Phi(x_2'\beta) - \Phi(x_1'\beta).$$

The gradient with respect to β is:

$$\nabla_\beta \gamma_1(\beta) = \frac{\partial}{\partial \beta}\left[\Phi(x_2'\beta)\right] - \frac{\partial}{\partial \beta}\left[\Phi(x_1'\beta)\right].$$

Since $\frac{d}{dt}\Phi(t) = \phi(t)$, we get:

$$\nabla_\beta \gamma_1(\beta) = \phi(x_2'\beta)x_2 - \phi(x_1'\beta)x_1.$$

Asymptotic Distribution of $\gamma_1(\hat{\beta})$:

Applying the Delta Method at β_0:

$$\sqrt{n}(\gamma_1(\hat{\beta}) - \gamma_1(\beta_0)) \xrightarrow{d} N(0, \nabla_\beta \gamma_1(\beta_0)' H^{-1} \nabla_\beta \gamma_1(\beta_0)).$$

In finite samples, we replace β_0 with $\hat{\beta}$, and H with its estimator $\hat{H}$, thus:

$$\gamma_1(\hat{\beta}) \overset{approx}{\sim} N\left(\gamma_1(\hat{\beta}), \frac{1}{n}\nabla_\beta \gamma_1(\hat{\beta})' \hat{H}^{-1} \nabla_\beta \gamma_1(\hat{\beta})\right),$$

where $\hat{H}$ and $\nabla_\beta \gamma_1(\hat{\beta})$ are computed from the sample and the estimated parameters. This gives us an asymptotic approximation to the finite sample distribution of $\gamma_1(\hat{\beta})$.

To summarize, the asymptotic variance of $\gamma_1(\hat{\beta})$ is:

$$\widehat{\text{Var}}(\gamma_1(\hat{\beta})) = \frac{1}{n}\nabla_\beta \gamma_1(\hat{\beta})' \hat{H}^{-1} \nabla_\beta \gamma_1(\hat{\beta}).$$

Empirical Implementation:

1. Estimate $\hat{\beta}$ using the Probit model.
2. Compute $\nabla_\beta \gamma_1(\hat{\beta}) = \phi(x_2'\hat{\beta})x_2 - \phi(x_1'\hat{\beta})x_1$.
3. Compute $\hat{H}^{-1}$ (the inverse of the estimated H matrix).
4. Approximate:

$$\gamma_1(\hat{\beta}) \sim N\left(\gamma_1(\hat{\beta}), \frac{1}{n}\nabla_\beta \gamma_1(\hat{\beta})' \hat{H}^{-1} \nabla_\beta \gamma_1(\hat{\beta})\right).$$

```r
phi_age_60 <- dnorm(sum(x_age_60 * beta_hat))
phi_age_30 <- dnorm(sum(x_age_30 * beta_hat))
grad_g <- phi_age_60 * x_age_60 - phi_age_30 * x_age_30

print(grad_g)

# Compute asymptotic variance
var_gamma_1 <- t(grad_g) %*% V_hat %*% grad_g / nrow(dat_1000)
gamma_1_sd <- sqrt(var_gamma_1)

simulated_gamma <- rnorm(1000, mean = gamma_1, sd = gamma_1_sd)

bootstrap_data2 <- data.frame(Value = gamma_1_bootstrap, Distribution =
```

```
      "Bootstrap")
14 simulated_data2 <- data.frame(Value = simulated_gamma, Distribution = "
      Asymptotic")
15
16 # Combine data
17 combined_data2 <- rbind(bootstrap_data2, simulated_data2)
18
19 # Create the plot
20 ggplot(combined_data2, aes(x = Value, fill = Distribution)) +
21   geom_histogram(aes(y = ..density..), bins = 20, position = "identity",
       alpha = 0.7, color = "black") +
22   scale_fill_manual(values = c("Bootstrap" = "grey", "Asymptotic" = "
      orange")) +
23   labs(title = "Comparison of Bootstrap and Asymptotic Distributions of
       Gamma_1",
24        x = "Gamma_1 Estimation", y = "Density") +
25   theme_minimal() +
26   theme(plot.title = element_text(hjust = 0.5, size = 16),
27         legend.title = element_blank(),
28         legend.position = "topright",
29         axis.title = element_text(size = 14),
30         axis.text = element_text(size = 12)) +
31   guides(fill = guide_legend(reverse = TRUE))
32
```

Solution (j).

We test

$$H_0 : \gamma_1(\beta) = 0 \quad \text{vs.} \quad H_1 : \gamma_1(\beta) \neq 0.$$

Under the asymptotic approximation,

$$\gamma_1(\hat{\beta}) \approx N\left(\gamma_1(\beta_0), \frac{1}{n}\hat{V}\right),$$

where

$$\hat{V} = \nabla_\beta \gamma_1(\hat{\beta})' \hat{H}^{-1} \nabla_\beta \gamma_1(\hat{\beta}).$$

The t-statistic is:

$$t = \frac{\gamma_1(\hat{\beta})}{\sqrt{\hat{V}/n}}.$$

Empirically, $t \approx -0.68$.

A 95% confidence interval for $\gamma_1(\beta)$ is:

$$\left[\gamma_1(\hat{\beta}) - 1.96\sqrt{\frac{\hat{V}}{n}}, \gamma_1(\hat{\beta}) + 1.96\sqrt{\frac{\hat{V}}{n}}\right].$$

Empirically, the 95% Confidence Interval for $\gamma_1(\beta)$ is: -0.0815 to 0.0396, covering 0.

So, we conclude that the expected probabilities of cash payment for a 30 year-old and a 60 year-old male buying clothes for 500 TRY are not significantly different.

Graphical Analysis and Econometric Interpretation:

- The bootstrap histograms of the coefficient on age, $\gamma_1(\hat{\beta})$, and $\gamma_2(\hat{\beta})$ all resemble roughly normal shapes, centered near their estimated values. This supports the use of normal approximations.

- The asymptotic approximations (shown by smooth curves on histograms) align reasonably with the bootstrap distributions, suggesting that for $n = 1000$, the asymptotic theory is not grossly inaccurate.

- The negative partial effects of age, though small in magnitude, show that older individuals might be slightly less likely to pay in cash. However, statistical tests indicate this difference is not significant at conventional levels.

- Category-specific differences matter, as seen from the significance of some category dummies. Clothes/Shoes, for example, differ in their propensity of cash usage compared to the base category.

```r
t_statistic <- gamma_1 / gamma_1_sd
critical_value <- qnorm(0.975)  # 1.96 for 95% confidence

if (abs(t_statistic) > critical_value) {
  conclusion <- "Reject the null hypothesis."
} else {
  conclusion <- "Fail to reject the null hypothesis."
}

print(paste("t-statistic:", round(t_statistic, 2)))
print(paste("Conclusion:", conclusion))

lower_bound <- gamma_1 - critical_value * gamma_1_sd
upper_bound <- gamma_1 + critical_value * gamma_1_sd

print(paste("95% Confidence Interval for gamma_1:", round(lower_bound,
    4), "to", round(upper_bound, 4)))
```