Mikroekonometria
6

Mikołaj Czajkowski
Wiktor Budziński

'Nietypowe' zmienne objaśniane

- Problemy mikroekonometryczne – często zmienna objaśniana nie jest ciągła lub jej wartość nie ma bezpośredniej interpretacji ilościowej
  - Zmienną objaśnianą jest jakiś wskaźnik o ocenzurowanych wartościach – np. decyzja o zakupie, głosowanie, wybór spośród dostępnych alternatyw, deklarowany poziom jakiejś wielkości (w podanej skali) itp.
  - Wynikiem optymalizacji może być rozwiązanie brzegowe
  - Estymatory MNK nie są zgodne -> estymacja MNW

'Nietypowe' zmienne objaśniane

- Przykłady modeli dla nieciągłych zmiennych objaśnianych
  - Wybór binarny (ang. *binary choice*)
    - Wybór jednej z dwóch możliwości
      - Np. podjęcie działania czy nie – dojazd transportem publicznym, kupno ubezpieczenia
    - 0/1 oznacza nie/tak
  - Wybór wielomianowy (ang. *multinomial choice*)
    - Wybór jednej z więcej niż dwóch możliwości
      - Tej, która dostarcza największej użyteczności
      - Pozwala modelować preferencje konsumentów
      - Np. wybór marki, produktu, sposobu dojazdu do pracy
    - 0/1 oznacza nie wybraną/wybraną alternatywę

'Nietypowe' zmienne objaśniane

- Przykłady modeli dla nieciągłych zmiennych objaśnianych
  - Wybór uporządkowany (ang. *ordered choice*)
    - Wybór jednej z wielkości na podanej skali
    - Ujawnia siłę preferencji
      - Np. ocena filmu, własnego stanu zdrowia, zadowolenia z produktu, miejsce w rankingu
    - Wartości liczbowe są na skali porządkowej, nie absolutnej
      - 4 oznacza lepiej niż 2, ale niekoniecznie 2 razy lepiej
  - Liczność zdarzeń (ang. *event counts*)
    - Obserwowane są wielkości całkowite
      - Np. liczba wizyt u lekarza, w parku narodowym, na basenie, liczba zachorowań, aresztowań, dzieci, liczba wadliwych sztuk w procesie produkcyjnym w ciągu roku

Modele binarne

- Obserwujemy wynik ($Y = 0$ lub $Y = 1$)
- Interesuje nas jak można go przewidywać
  - Prawdopodobieństwo $Y = 0$ / prawdopodobieństwo $Y = 1$
  - Funkcja zmiennych objaśniających $\mathbf{X}$, $\Pr(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$
  - Jakie powinny być parametry ($\boldsymbol{\beta}$)?

- Jaka jest interpretacja takiego modelu?
- Jaki jest wzór na to prawdopodobieństwo, jako funkcja $\mathbf{X}\boldsymbol{\beta}$?
- W jaki sposób wykorzystać dane (obserwacje) do oszacowania $\boldsymbol{\beta}$?

Modele binarne – modele dla zmiennej ukrytej

- Modele binarne można zinterpretować jako modele funkcji wskaźnikowej
  - Wyjaśniana jest losowa, ciągła zmienna *y** ale obserwowana jest tylko zmienna binarna *y*, która przyjmuje wartość 0 lub 1 w zależności od tego, czy *y** przekracza wartość graniczną (np., po normalizacji, 0)
    - $y^* = \boldsymbol{\beta}'\mathbf{X} + \varepsilon$
    - y^* nie jest obserwowalne, obserwujemy tylko: $y = \begin{cases} 1 & \textit{jeśli} \quad y^* > 0 \\ 0 & \textit{jeśli} \quad y^* \le 0 \end{cases}$
    - $$\begin{aligned} \Pr[y = 1 \mid \mathbf{X}] &= \Pr\left[y^* > 0\right] \\ &= \Pr[\boldsymbol{\beta}'\mathbf{X} + \varepsilon > 0] \\ &= \Pr[-\varepsilon < \boldsymbol{\beta}'\mathbf{X}] \\ &= F(\boldsymbol{\beta}'\mathbf{X}) \end{aligned}$$
    - *F* jest dystrybuantą ε
      - Zwykle o gęstości symetrycznej względem 0
    - Pozostaje tylko wybrać odpowiedni rozkład ε

Modele binarne – normalizacja składnika losowego

- Wariancja przyjętego rozkładu składnika losowego nie ma znaczenia

$$y = 1 \Leftrightarrow \boldsymbol{\beta}'\mathbf{X} + \varepsilon > 0$$

  - ε ma średnią = 0 (jeśli model ma stałą to to też nie ma znaczenia) i jakąś wariancję σ (nie wiadomo jaką)
  - Model będzie ten sam jeśli policzymy

$$y = 1 \Leftrightarrow \frac{\boldsymbol{\beta}'\mathbf{X}}{\sigma} + \frac{\varepsilon}{\sigma} > 0$$

  - Co jest tożsame z

$$y = 1 \Leftrightarrow \boldsymbol{\gamma}'\mathbf{X} + w > 0$$

  - Gdzie w to jakiś rozkład o średniej 0 i wariancji 1
  - Można przyjąć wariancję = 1, przeskalują się tylko parametry, model ten sam
  - Normalizacja wariancji (lub jednego z parametrów) jest konieczna dla identyfikowalności modelu

Logit / probit

- Typowo zakładane – składnik losowy (ε) ma rozkład:
  - Logistyczny

$$f(x)=\frac{\exp\left(-\frac{x-\mu}{\sigma}\right)}{\sigma\left(1+\exp\left(-\frac{x-\mu}{\sigma}\right)\right)^2}$$

  - Normalny

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

  - Zakładamy rozkład standardowy (μ = 0, σ = 1)
    - Nie ma to znaczenia, bo najwyżej parametry β będą większe, żeby poziom 'losowości' był odpowiedni
  - Dystrybuanta

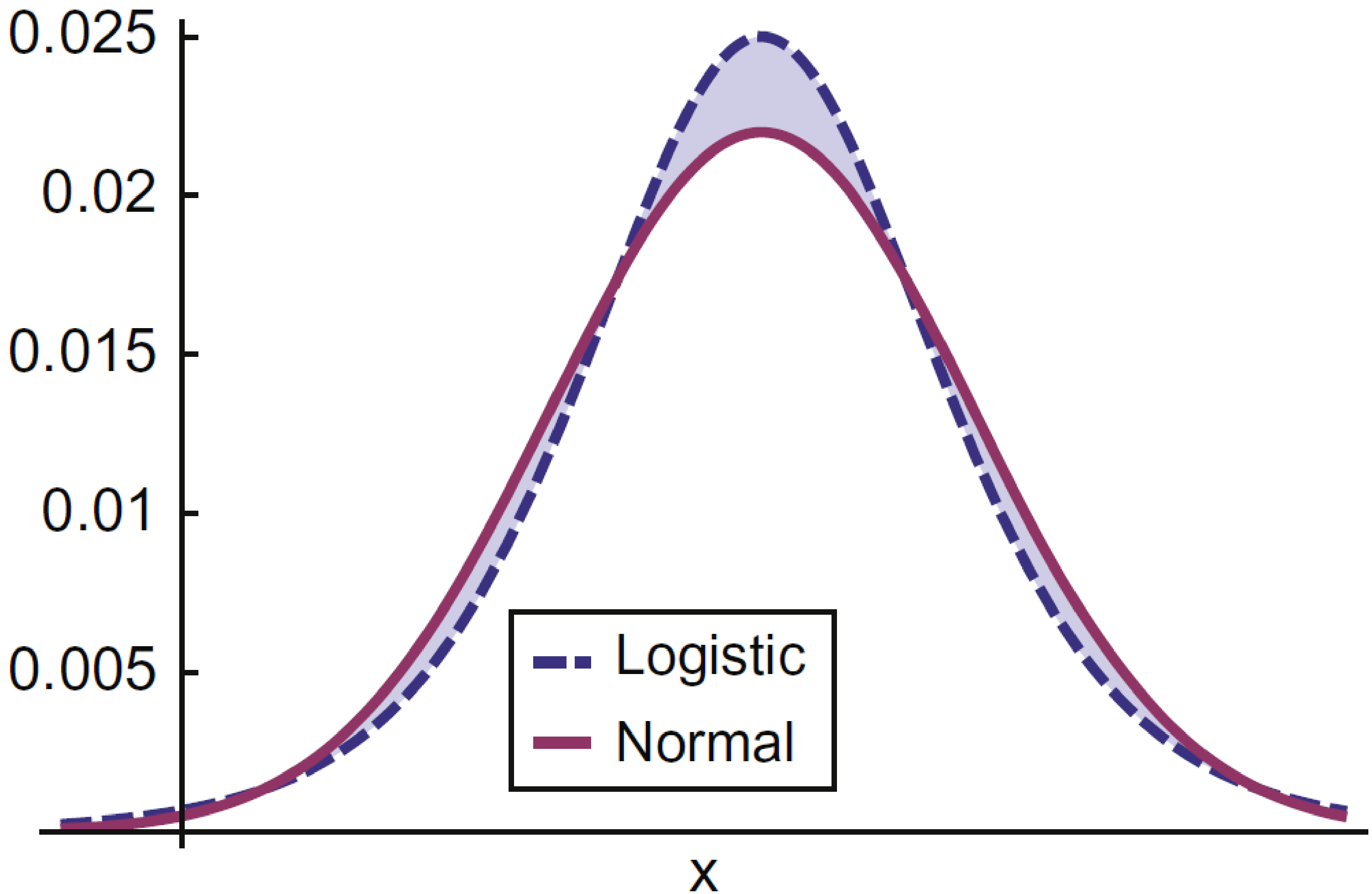
$$\Lambda(x)=\frac{\exp(x)}{1+\exp(x)}$$

  - Dystrybuanta

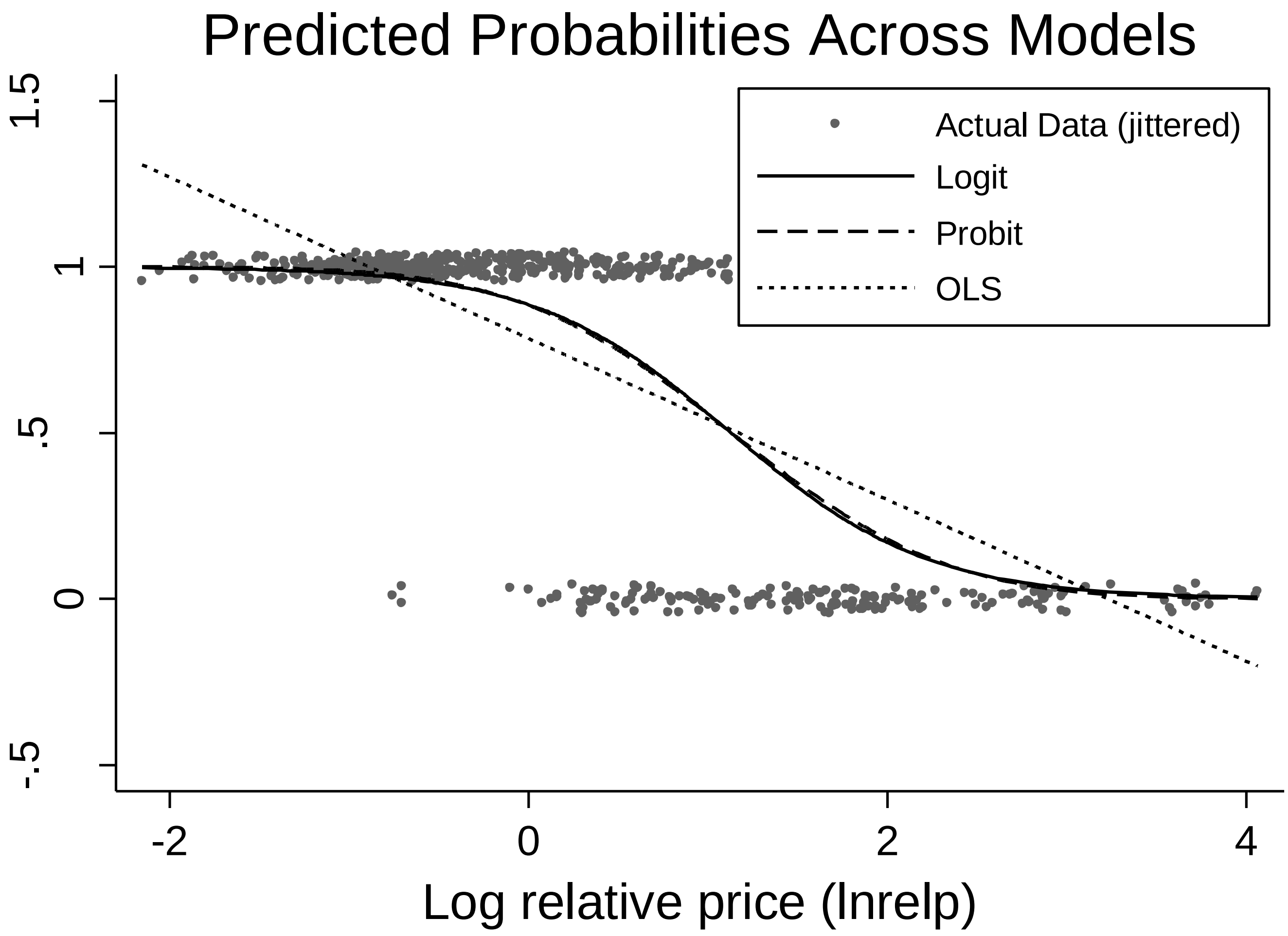
$$\Phi(x)=\int_{-\infty}^{x}\phi(z)dz$$

  - Mamy wzór na Pr, który możemy wykorzystać w estymacji!
  - -> Model logit
  - -> Model probit

Rozkład logistyczny vs. normalny



Logit / probit / liniowy model prawdopodobieństwa



Logit / probit – który wybrać?

- Rozkład logistyczny – przypomina normalny (symetryczny dzwon, trochę grubsze ogony), ale ma wzór na dystrybuantę w postaci zamkniętej
  - Upraszcza obliczenia / umożliwia rozwiązanie analityczne
  - Pozwala interpretować parametry jako stosunek log-oddsów
- Probit
  - Może być bardziej uzasadniony w przypadku normalnego rozkładu zmiennej ukrytej
- Praktycznie nie ma różnicy
  - Istnieją testy dla porównywania niezagnieżdżonych modeli, ale w tym przypadku prawie zawsze nie dają jednoznacznego wyniku
  - Ponieważ logit ma grubsze ogony, w próbach w których obserwacje są bardzo niesymetryczne (znacznie więcej 0 lub 1) lub jest bardzo duża wariancja ważnej zmiennej objaśniającej – różnice mogą być większe

Logit / probit – parametry

- Ponieważ logit i probit wykorzystują różne funkcje prawdopodobieństwa, oszacowania parametrów będą różne
  - Jako 'reguła kciuka':

$$\begin{aligned} \hat{\beta}_{\text{logit}} &\approx 4\hat{\beta}_{\text{MNK}} \\ \hat{\beta}_{\text{probit}} &\approx 2{,}5\hat{\beta}_{\text{MNK}} \\ \hat{\beta}_{\text{logit}} &\approx 1{,}6\hat{\beta}_{\text{probit}} \end{aligned}$$

  - Ale to oczywiście bardziej skomplikowane
  - Absolutne wartości parametrów i tak nie mają znaczenia
    - Nie tak jak w MNK!
    - Dlaczego? Bo parametry wykorzystywane w skomplikowanych funkcjach obliczających prawdopodobieństwo
    - Jeśli już, to lepiej porównywać *efekty krańcowe* (pochodne)
      - Jak zmienne objaśniające wpływają na zmianę prawdopodobieństwa danego wyboru

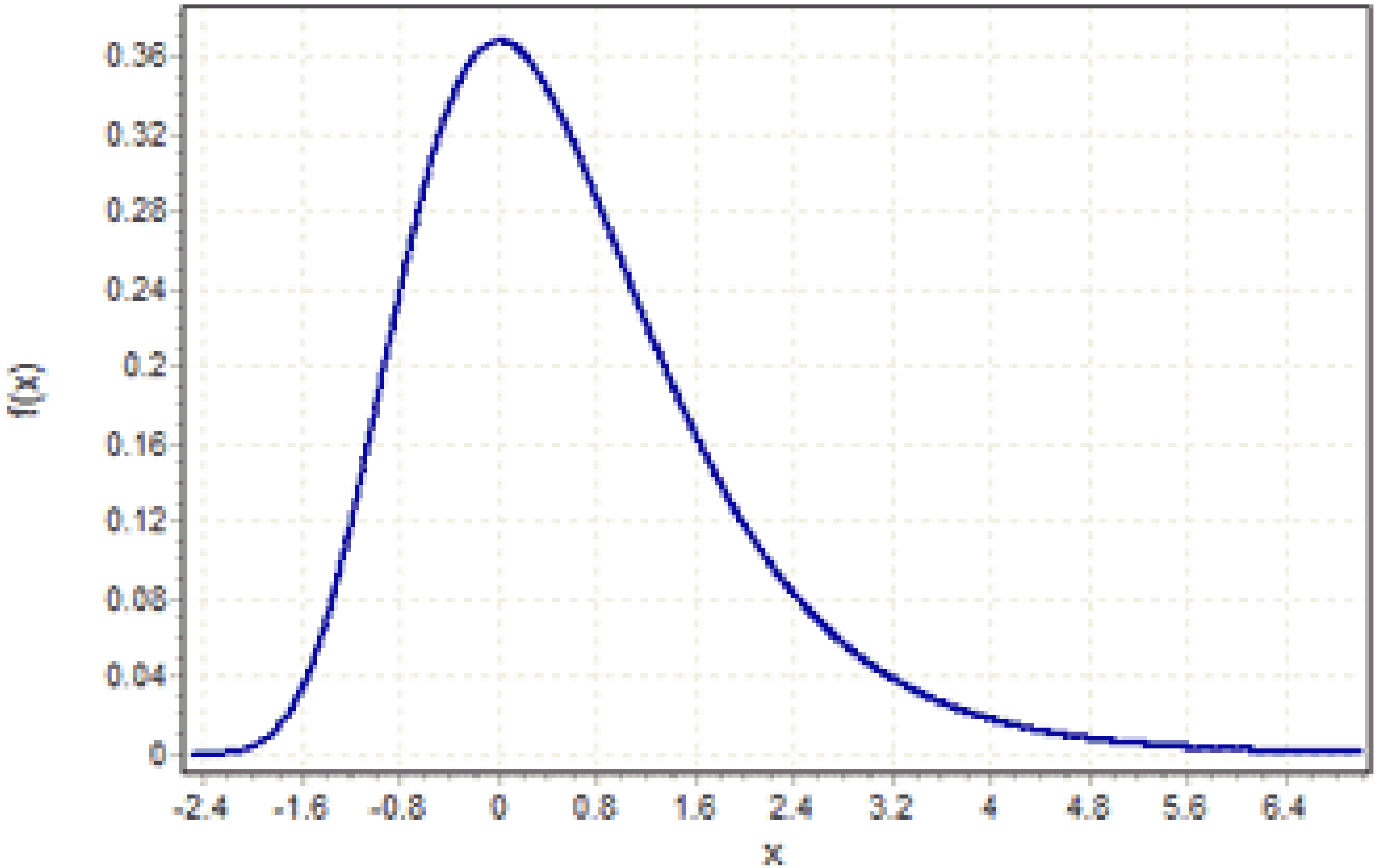
Logit / probit i inne modele dla wyboru binarnego

- Inne rozkłady składnika losowego, które nie zakładają symetryczności:
  - Gumbel, Gompit, Gompertz (ε ma rozkład wartości ekstremalnych)

    $$CDF(x) = \exp(-\exp(-x))$$
  - *Complementary log log*

    $$CDF(x) = 1 - \exp(-\exp(-x))$$
  - I inne (np. arctangens, burr (scobit), ...)
  - Wyniki mogą się czasem znacznie różnić od wyników logitu i probitu
  - Przydatne, gdy jeden z wyników jest rzadki
    - Np. mały odsetek 'tak' w próbie

Standardowy rozkład wartości ekstremalnych



Estymacja – metoda największej wiarygodności

- Zaobserwowaliśmy $(Y_i, \mathbf{X}_i)$, dla $i = 1, \ldots, N$
- Zmienna zależna (Y_i) ma rozkład Bernoulliego (binarny z 1 powtórzeniem)
  - Funkcja prawdopodobieństwa masy

$$f(Y_i | \mathbf{X}_i) = p_i^{Y_i} (1 - p_i)^{1-Y_i}$$

  - Co się dzieje z tą funkcją gdy $Y_i = 1$? A gdy $Y_i = 0$?
    - Zawsze dostajemy p_i lub $(1 - p_i)$
  - U nas $p_i = F(\boldsymbol{\beta}'\mathbf{X})$ więc

$$f(Y_i | \mathbf{X}_i) = F(\boldsymbol{\beta}'\mathbf{X}_i)^{Y_i} (1 - F(\boldsymbol{\beta}'\mathbf{X}_i))^{1-Y_i}$$

Estymacja – metoda największej wiarygodności

- Funkcja prawdopodobieństwa masy

$$f\left(Y_i \mid \mathbf{X}_i\right) = F\left(\boldsymbol{\beta}'\mathbf{X}_i\right)^{Y_i}\left(1 - F\left(\boldsymbol{\beta}'\mathbf{X}_i\right)\right)^{1-Y_i}$$

- Obserwacje są niezależne. Jaka jest szansa, że dostaniemy taki zbiór obserwacji, jaki mamy?

$$L\left(\boldsymbol{\beta} \mid \text{dane}\right) = \prod_{i=1}^{N} F\left(\boldsymbol{\beta}'\mathbf{X}_i\right)^{Y_i}\left(1 - F\left(\boldsymbol{\beta}'\mathbf{X}_i\right)\right)^{1-Y_i}$$

  - Funkcja wiarygodności (ang. *likelihood function*)
- Chcemy znaleźć takie $\boldsymbol{\beta}$, żeby zmaksymalizować tę funkcję
  - Czyli, żeby nasz model miał takie parametry, żeby przewidywał takie prawdopodobieństwa, które dają największą szansę takiego wyniku, jaki dostaliśmy

Estymacja – metoda największej wiarygodności

- Szukamy maksimum funkcji wiarygodności:

$$L(\boldsymbol{\beta}|\text{dane}) = \prod_{i=1}^{N} F(\boldsymbol{\beta}'\mathbf{X}_i)^{Y_i}\left(1 - F(\boldsymbol{\beta}'\mathbf{X}_i)\right)^{1-Y_i}$$

  - Jest ono w tym samym punkcie, co maksimum logarytmu z funkcji wiarygodności (logarytm to funkcja monotonicznie rosnąca), a wygodniej maksymalizować coś takiego:

$$\ln L(\boldsymbol{\beta}|\text{dane}) = \sum_{i=1}^{N}\left(Y_i \ln F(\boldsymbol{\beta}'\mathbf{X}_i) + (1 - Y_i)\ln\left(1 - F(\boldsymbol{\beta}'\mathbf{X}_i)\right)\right)$$

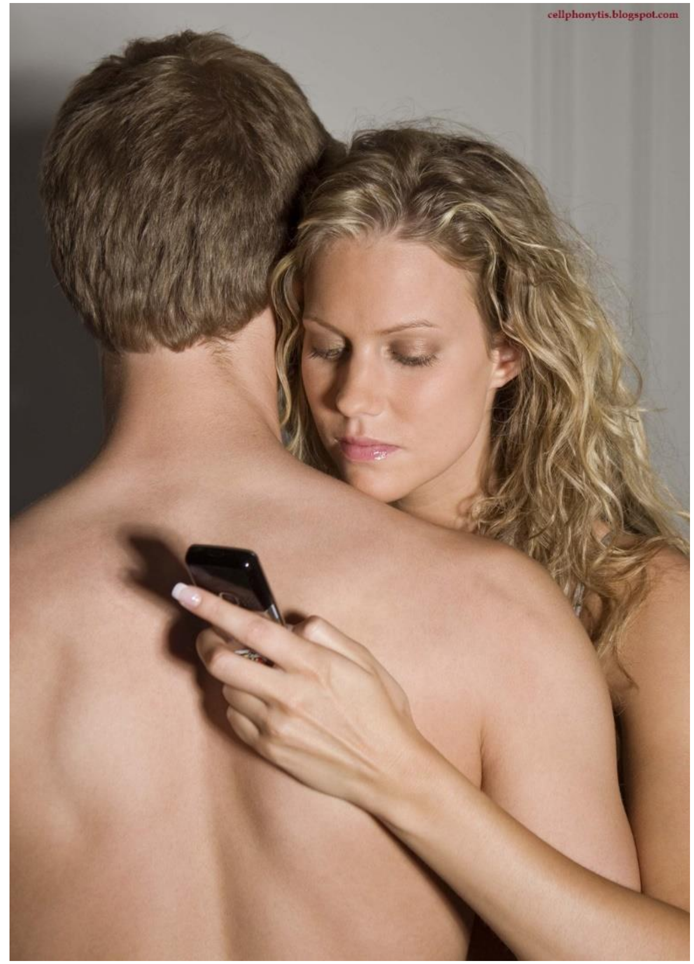
    - W praktyce – dość prosto się liczy
      - Dla każdej obserwacji wziąć prawdopodobieństwo uzyskanego wyniku (tak lub nie)
      - Zsumować
      - Zmaksymalizować funkcję po β

Estymacja – metoda największej wiarygodności

- Estymację robią za nas zwykle pakiety statystyczne ☺
  - Ale jak widać jest to dość proste, da się policzyć na piechotę
- Hesjan modelu logit/probit jest ujemnie określony
  - Funkcja LL jest globalnie wklęsła
  - Istnieje globalne maksimum
- Estymator MNW jest zgodny
  - Estymator MNK nie jest
- Parametry mają asymptotyczny rozkład normalny
- Asymptotyczna macierz wariancji-kowariancji (ang. *asymptotic variance covariance matrix*, AVC) może być oszacowana jako odwrotność hesjanu dla optymalnych oszacowań MNW
  - Stąd mamy błędy standardowe parametrów

Zadanie 1. – Posiadanie kochanka

1. Wczytaj projekt me.affairs.dta
   - Dane z ankiety 6366 mężatek (magazyn *Redbook*)
2. Ile % kobiet w próbie miało kiedykolwiek romans?
3. Sprawdź, jakie zmienne mogą pomóc to przewidywać
4. Porównaj wyniki z Liniowym Modelem Prawdopodobieństwa



Wartość funkcji LL zawsze ujemna

- Funkcja LL:

$$\ln L(\boldsymbol{\beta} | \text{dane}) = \sum_{i=1}^{N} \Big(Y_i \ln F(\boldsymbol{\beta}' \mathbf{X}_i) + (1 - Y_i) \ln \big(1 - F(\boldsymbol{\beta}' \mathbf{X}_i)\big) \Big)$$

- Logarytm z prawdopodobieństwa (<1) zawsze ujemny
  - Funkcja LL zawsze ujemna
  - Osiąga teoretyczne maksimum w 0

Zadanie 2. – Zakup eko-jabłek

1. Wczytaj projekt me.apples.dta
2. Oszacuj model, w którym gotowość do zakupu dodatnich ilości eko-jabłek wyjaśniana jest przez:
   - Poziom edukacji respondenta
   - Cenę zwykłych jabłek
   - Cenę eko-jabłek
   - To, czy wybór dokonywany jest w sezonie
   - To czy respondent był mężczyzną
   - Dochód rodziny
   - Liczbę członków rodziny w każdej kategorii wiekowej
3. Co dzieje się z modelem, jeśli dodatkowo uwzględnić wielkość rodziny?

Pomiar 'dopasowania' modelu

- Po estymacji modelu zwykle raportowane są:
  - Wartość funkcji LL w optimum (w punkcie konwergencji)
  - Wartość funkcji LL 'w 0' (model wykorzystujący tylko stałą)
    - Odpowiada założeniu, że wszystkie parametry w modelu nieistotne
- Nie ma miary R^2 z MNK
- Pewnym rozwiązaniem jest Pseudo-R^2 (kilka wersji)
  - Miary oparte na wartości funkcji LL (wiele wersji):
    - McFadden's pseudo-$R^2 = \frac{1 - \ln L(\boldsymbol{\beta})}{\ln L(\mathbf{Y})} = 1 - \ln L / \ln L_0 \qquad \in [0,1)$
    - Im model lepiej dopasowany tym R2 bliższe 1
      - Ale przedział między 0 a 1 nie ma interpretacji naturalnej
        - Nie wiemy, gdzie jest zero
        - Z powodu istnienia składnika losowego model nie osiąga 1

Pomiar 'dopasowania' modelu

- Kryterium informacyjne Akaike (AIC)

$$AIC = 2k - 2\ln L(\boldsymbol{\beta})$$

- AIC skorygowane (dla skończonej próby)

$$AICc = AIC + \frac{2k(k+1)}{N-k-1}$$

- Bayesowskie kryterium informacyjne (BIC, Schwarza)

$$BIC = k\ln N - 2\ln L(\boldsymbol{\beta})$$

  - Znormalizowane – podzielone przez *N*
  - Jak w modelu liniowym – nie jest to kryterium 'statystycznego' porównania, ale często używane
    - Mówi, który model lepiej pasuje do danych
    - Nie mówi czy dobrze, o ile lepiej i czy istotnie lepiej
    - W porównaniu z pseudo-R^2 przynajmniej bierze pod uwagę liczbę zmiennych objaśniających

Pomiar 'dopasowania' modelu

- Miary oparte na przewidywaniu wyników
  - Tabela poprawnych i niepoprawnych predykcji
    - Błędy I i II typu – trade-off
    - Ile % poprawnych predykcji ma naiwna reguła $Y_i = 1$?
- Miary oparte na przewidywaniu prawdopodobieństw
  - BenAkiva i Lerman's pseudo-$R^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i\hat{F}(\mathbf{X}_i\boldsymbol{\beta}) + (1-Y_i)\left(1-\hat{F}(\mathbf{X}_i\boldsymbol{\beta})\right)\right)$
    - Średnie prawdopodobieństwo poprawnych predykcji
    - Problem gdy obserwacje w próbie są niesymetryczne

Zadanie 2. – Zakup eko-jabłek c.d.

4. Wypróbuj inne modele dla zmiennych binarnych
   - Probit
   - Scobit
   - Complementary log-log
5. Porównaj dopasowanie wykorzystując różne miary

Praca domowa ME.6 – wydatki na lekarza w USA

1. Wczytaj projekt me.usahealth.dta
2. Sprawdź czy następujące zmienne mogą pomóc przewidzieć, czy ktoś poniósł jakieś wydatki medyczne (tzn. czy zmienna *med* jest większa niż 0)
   - Udział własny w kosztach opieki medycznej (zmienna *coins*)
   - Stan zdrowia
   - Czynniki socjodemograficzne

```
set seed 10+"Nr indeksu"
sample 90, by(coins)
```

3. Wypróbuj różne modele dla zmiennych binarnych i wybierz najlepszy
4. Wykorzystaj wybrane przekształcenia nieliniowe, w celu poprawy dopasowania (LL) wybranego modelu

- Do przygotowania w grupach trzyosobowych