Το Υπόδειγμα Probit

18 Μαΐου 2026

Οικονομική θεμελίωση: απόφαση ως σύγκριση ωφελειών I

Θεωρούμε έναν οικονομικό δρώντα που επιλέγει μεταξύ δύο εναλλακτικών:

$$d \in \{0, 1\}$$

Ορίζουμε τις αντίστοιχες ωφέλειες:

$$U_i(1), \quad U_i(0)$$

Κανόνας απόφασης:

$$Y_{(i)} = 1 \iff U_i(1) > U_i(0)$$

Ορίζουμε τη διαφορά ωφελειών:

$$Y^*_{(i)} = U_i(1) - U_i(0)$$

Οικονομική θεμελίωση: απόφαση ως σύγκριση ωφελειών II

Γραμμική εξειδίκευση:

$$Y^*_{(i)} = X_{(i)}\beta_0 + \varepsilon_{(i)}$$

όπου:

- $X_{(i)}$: παρατηρήσιμα χαρακτηριστικά (εισόδημα, εκπαίδευση κ.λπ.)
- β: παράμετροι που μετρούν επίδραση
- $\varepsilon_{(i)}$: μη παρατηρήσιμοι παράγοντες (προτιμήσεις, πληροφορία-πηγές ετερογένειας)

Παρατηρήσιμη μεταβλητή:

$$Y_{(i)} = \mathbf{1}\{Y^*_{(i)} > 0\}$$

Κεντρική ιδέα: Παρατηρούμε μόνο την (διακριτή) επιλογή, όχι τη ωφέλεια.

Από το μη παρατηρήσιμο υπόδειγμα στις πιθανότητες

Από:

$$Y_{(i)} = \mathbf{1}\{X_{(i)}\beta_0 + \varepsilon_{(i)} > 0\}$$

έχουμε:

$$\mathbb{P}(Y_{(i)} = 1 | X_n) = \mathbb{P}(\varepsilon_{(i)} > -X_{(i)}\beta)$$

Προφανώς η μορφή της παραπάνω πιθανότητας (και του αντίστοιχου στατιστικού υποδείγματος) καθορίζεται από την κατανομή του $\varepsilon_{(i)}$:

- $\varepsilon_{(i)} | X_n \sim N(0, 1) \Rightarrow$ Probit
- $\varepsilon_{(i)} | X_n \sim \mathit{Logistic} \Rightarrow$ Logit

Κεντρική ιδέα:

$$\text{Κατανομή σφάλματος} \;\Rightarrow\; \text{μορφή πιθανοτήτων}$$

Ταυτοποίηση στο Probit: το πρόβλημα

Θεωρούμε το μη παρατηρήσιμο υπόδειγμα:

$$Y^*_{(i)} = X_{(i)}\beta + \varepsilon_{(i)}, \quad Y_{(i)} = \mathbf{1}\{Y^*_{(i)} > 0\}, \quad \beta \in \Theta$$

Παρατηρούμε μόνο:

$$Y_{(i)} = \mathbf{1}\{X_{(i)}\beta + \varepsilon_{(i)} > 0\}$$

Κρίσιμο σημείο: η απόφαση εξαρτάται μόνο από το πρόσημο του $Y^*_{(i)}$.

Για κάθε $c > 0$:

$$\mathbf{1}\{X_{(i)}\beta + \varepsilon_{(i)} > 0\} = \mathbf{1}\{cX_{(i)}\beta + c\varepsilon_{(i)} > 0\}$$

Κεντρική ιδέα: Το υπόδειγμα είναι **αναλλοίωτο στην κλίμακα (scale invariance)**.

Ταυτοποίηση: συνέπειες και κανονικοποίηση

Από την προηγούμενη ιδιότητα η διακύμανση της δεσμευμένης κατανομής του $\varepsilon_{(i)}$ δεν είναι ταυτοποιήσιμη.

Διαισθητικά:

- μόνο το *σχετικό μέγεθος* $X_{(i)}\beta$ ως προς το $\varepsilon_{(i)}$ έχει σημασία
- η απόλυτη κλίμακα δεν παρατηρείται

Κανονικοποίηση:

$$\mathrm{Var}(\varepsilon_{(i)}|_n) = 1$$

και περαιτέρω (υποθέτουμε και δεσμευμένη ανεξαρτησία και ομοιογένεια-θα μπορούσαμε να το γενικεύσουμε):

$$\varepsilon_n | X_n \sim N(\mathbf{0}_n, I_{n\times n})$$

Από τα παραπάνω έχουμε

$$\mathbb{P}(Y_{(i)} = 1 \mid X_n) = \Phi(X_{(i)}\beta), \qquad \mathbb{P}(Y_{(i)} = 0 \mid X_n) = 1 - \Phi(X_{(i)}\beta),$$

με $\Phi(X_{(i)}\beta) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{X_{(i)}\beta} \exp(-\frac{1}{2}z^2)dz$.
Επειδή $Y_{(i)} \in \{0, 1\}$, η υπό συνθήκη πιθανότητα της παρατήρησης $Y_{(i)}$ γράφεται ενιαία ως

$$\mathbb{P}(Y_{(i)} \mid X_n; \beta) = \Phi(X_{(i)}\beta)^{Y_{(i)}} \left[1 - \Phi(X_{(i)}\beta)\right]^{1-Y_{(i)}}.$$

Άρα, υπό την υπό συνθήκη ανεξαρτησία η από κοινού κατανομή του δείγματος Y_n δεδομένων των X_n αν η πραγματική τιμή της παραμέτρου ήταν το β είναι η,

$$L_n(\beta) = \prod_{i=1}^{n} \Phi(X_{(i)}\beta)^{Y_{(i)}} \left[1 - \Phi(X_{(i)}\beta)\right]^{1-Y_{(i)}}.$$

Εξαγωγή της πιθανοφάνειας II

Λαμβάνοντας λογαρίθμους και ολοκληρώνοντας ως προς την εμπειρική κατανομή:

$$\ell_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left[Y_{(i)}\log\Phi(X_{(i)}\beta) + (1 - Y_{(i)})\log(1 - \Phi(X_{(i)}\beta))\right].$$

Ο εκτιμητής μέγιστης πιθανοφάνειας ορίζεται ως

$$\hat{\beta} \in \operatorname*{arg\,max}_{\beta \in \Theta} \ell_n(\beta).$$

Άρα:

- το κριτήριο προκύπτει από την πιθανοθεωρητική δομή του μοντέλου,
- το πρόβλημα είναι μη γραμμικό και η λύση δεν έχει κλειστή μορφή.

Αριθμητική βελτιστοποίηση: δομή προβλήματος I

Θέλουμε:

$$\hat{\beta} \in \arg\max_{\beta \in \Theta} \ell_n(\beta)$$

$$\ell_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \left[Y_{(i)} \log \Phi(X_{(i)}\beta) + (1 - Y_{(i)}) \log(1 - \Phi(X_{(i)}\beta)) \right]$$

Παράγωγος: (χρήσιμο π.χ. όταν $\Theta = \mathbb{R}^p$)

$$\nabla \ell_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \frac{(Y_{(i)} - \Phi(X_{(i)}\beta))\phi(X_{(i)}\beta)}{\Phi(X_{(i)}\beta)(1 - \Phi(X_{(i)}\beta))} X'_{(i)}$$

Αριθμητική βελτιστοποίηση: δομή προβλήματος II

Εσσιανή μήτρα:

- περιλαμβάνει μεταξύ άλλων παρονομαστές της μορφής

$$\Phi(X_{(i)}\beta)^2,\ (1-\Phi(X_{(i)}\beta))^2$$

οι οποίοι μπορεί να γίνουν αριθμητικά πολύ μικροί για μεγάλες θετικές ή αρνητικές τιμές του $X'_{(i)}\beta$ οπότε είναι δυνατόν να εμφανίζονται προβλήματα αριθμητικής ευστάθειας

Ενδεικτικοί Αλγόριθμοι βελτιστοποίησης

Newton-Raphson:

$$\beta_{k+1} = \beta_k - \left[\nabla^2 \ell_n(\beta_k)\right]^{-1} \nabla \ell_n(\beta_k)$$

με προβολή στο Θ αν δεν είναι πλήρης.

- χρησιμοποιεί την πλήρη Εσσιανή
- τοπική τετραγωνική προσέγγιση του κριτηρίου
- πολύ γρήγορη σύγκλιση κοντά στο βέλτιστο
- ευαίσθητος στην επιλογή αρχικής τιμής

Fisher Scoring:

- αντικαθιστά την Εσσιανή με την εμπειρική μήτρα πληροφοριών

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\phi^2(X_{(i)}\beta)}{\Phi(X_{(i)}\beta)(1-\Phi(X_{(i)}\beta))} \cdot X_{(i)}' X_{(i)}$$

BFGS:

- γενικότερα προσεγγίζει την Εσσιανή χωρίς τον υπολογισμό δεύτερων παραγώγων

Πλεονεκτήματα:

- υψηλή σταθερότητα
- καλή απόδοση σε μεσαία/μεγάλα δείγματα

Σύγκριση:

Μέθοδος	Ταχύτητα	Σταθερότητα	Κόστος
Newton	πολύ υψηλή	μέτρια	υψηλό
Fisher	υψηλή	υψηλή	μέτριο
BFGS	καλή	πολύ υψηλή	χαμηλό

Trade-off: ταχύτητα σύγκλισης vs υπολογιστική σταθερότητα

Υπολογιστική δομή του Probit MLE I

Σε κάθε επανάληψη ενός αλγορίθμου (π.χ. BFGS / Newton) υπολογίζουμε:

1. Τιμές γραμμικού δείκτη:

$$z_i = X_{(i)}\beta \quad \Rightarrow \quad \mathcal{O}(np)$$

2. Πιθανότητες και πυκνότητες:

$$\Phi(z_i), \quad \phi(z_i) \quad \Rightarrow \quad \mathcal{O}(n)$$

3. Gradient:

$$\nabla \ell_n(\beta) = \sum_{i=1}^{n} X_{(i)} \cdot w_i \quad \Rightarrow \quad \mathcal{O}(np)$$

(Newton μόνο) Υπολογισμός Εσσιανής:

$$\nabla^2 \ell_n(\beta) \quad \Rightarrow \quad \mathcal{O}(np^2)$$

Υπολογιστική δομή του Probit MLE II

Συνολικό κόστος ανά iteration:

$$\mathcal{O}(np) \quad \text{(BFGS)} \qquad \mathcal{O}(np^2) \quad \text{(Newton)}$$

Αριθμός επαναλήψεων:

$$20 - 50 \ (\text{τυπικά})$$

Συνολικό αριθμητικό κόστος:

$$\mathcal{O}(np^\star \times \#\,\text{επαναλήψεις}), \quad p^\star = p \ (\text{BFGS}), \qquad p^\star = p^2 \ (\text{Newton})$$

Οι υπολογισμοί μπορεί να επιβαρύνονται από την ύπαρξη περιορισμών όταν $\Theta \subset \mathbb{R}^p$.

Python snippet: Probit MLE (BFGS)

```
import numpy as np
from scipy.stats import norm
from scipy.optimize import minimize

def probit_loglik(beta, X, y):
    z = X @ beta
    p = norm.cdf(z)

    # numerical stability
    eps = 1e-8
    p = np.clip(p, eps, 1 - eps)

    ll = y * np.log(p) + (1 - y) * np.log(1 - p)
    return -np.sum(ll)   # negative log-likelihood

def estimate_probit(X, y):
    p = X.shape[1]
    beta0 = np.zeros(p)

    res = minimize(probit_loglik, beta0,
                   args=(X, y),
                   method="BFGS")

    return res.x

# usage:
# beta_hat = estimate_probit(X, y)
```

Monte Carlo I: καλή εξειδίκευση

$$Y_n^* = X_n\beta_0 + \varepsilon_n, \quad \varepsilon \sim N(0,1)$$

$$\beta_0 = (1,1)$$

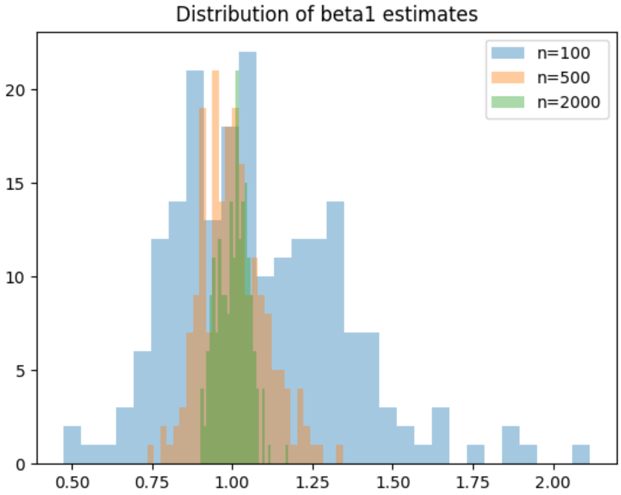
- $\Theta = \mathbb{R}^2$
- $X = (\mathbf{1}, N(\mathbf{0}_n, I_{n\times n}))$ ανεξάρτητα του ε
- $n = 100, 500, 2000$
- 200 MC επαναλήψεις
- βελτιστοποίηση με BFGS

Monte Carlo I: αποτελέσματα

n	$\bar{\beta}_0$	$\bar{\beta}_1$	σ_0	σ_1
100	1.071	1.081	0.230	0.268
500	1.007	1.002	0.087	0.101
2000	1.004	1.004	0.041	0.048

Διακρίνεται: σύγκλιση προς β_0

Monte Carlo I: εμπειρική κατανομή εκτιμητή



Monte Carlo II: κακή εξειδίκευση

$$\varepsilon \sim t(3)$$

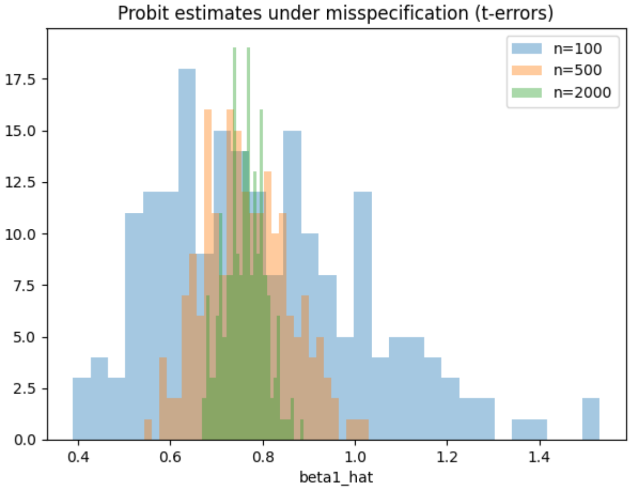
εκτίμηση: Probit

Κεντρική ιδέα: το υπόδειγμα κακώς εξειδικευμένο

Monte Carlo II: αποτελέσματα

n	$\bar{\beta}_0$	$\bar{\beta}_1$	σ_0	σ_1
100	0.800	0.801	0.189	0.224
500	0.784	0.766	0.074	0.092
2000	0.787	0.761	0.040	0.043

Monte Carlo II: εμπειρική κατανομή εκτιμητή)



Ψευδο-αληθής (pseudo-true) τιμή της παραμέτρου

$$\hat{\beta} \to \beta^* \neq \beta_0$$

$$\beta^* = \arg\max \mathbb{E}[\log f(Y|X, \beta)]$$

Κεντρική ιδέα: KL-τύπου προβολή στην εγγύτερη κατανομή στο στατιστικό υπόδειγμα.

Υπολογιστικοί χρόνοι πειραμάτων

- 200 βελτιστοποιήσεις
- κάθε βελτιστοποίηση $\approx 0.01 - 0.03$ sec

Συνολικός χρόνος:

$$2 - 6 \text{ δευτερόλεπτα}$$

Hardware:

- σε τυπικό laptop (i5 / M1)