

1 Binarinis pasirinkimas

2005/2006 m.m. rudens semestras

Tegu

$$I = Z\gamma + V \quad (1)$$

$$D = 1(I > 0), \quad (2)$$

čia Z ir γ gali būti vektoriai. Stebimi dydžiai yra D ir Z . Tarkime, kad

$$V \sim N(0, \sigma_V^2)$$

ir

(Z, V) tarpusavyje nepriklausomi.

1. Parašykite didžiausio tikėtinumo funkciją šios problemos sprendimui.
2. Paaiškinkite, kodėl reikalinga prielaida $\sigma_V^2 = 1$ (ar kokiai kitokiai konstantai, nebūtinai vienetui).
3. Išspręskite šią didžiausio tikėtinumo problemą naudodamiesi savo pasirinktu statistiniu paketu. T.y. parašykite funkciją (programą), kuri gražintų didžiausio tikėtinumo funkcijos reikšmes duotam γ , kai D ir Z žinomi, bei ją maksimizuokite. (Galite naudoti kokį tik norite statistinį paketą, funkcija turi veikti nepriklausomai nuo to kokios yra duomenų dimensijos). Nuo šiol kitoms užduotims atlikti laikykite, kad $\sigma_V^2 = 1$. Toliau naudosime duomenis, kurie yra generuojami taip:

$$V \sim N(0, 1),$$

Z_0 yra konstanta (tarkime vienetą). Tada

$$Z_1 \sim N(0, 1),$$

taip kad Z_1 ir V būtų nepriklausomi. Dabar generuokime

$$I = 0.5Z_0 + Z_1 + V$$

ir tegu

$$D = 1(I > 0).$$

4. Aprašykite, kaip sugeneruoti tokius duomenis. (Pageidautina, kad tai atliktų funkcija, kurios parametras būtų imties dydis).
5. Išspręskite šį didžiausio tikėtinumo probit, naudodamiesi savo parašyta funkcija bei sugeneruotais duomenimis. Pakartokite tai bent su trim skirtingais imties dydžiais. Kaip keičiasi koeficientų įverčiai priklausomai nuo imties dydžio? Paaiškinkite. (Jei viską teisingai atlikote γ_0 įvertis turėtų būti apie 0.5, o γ_1 apie 1).
6. Dabar generuokime duomenis šiek tiek kitaip. Tegu vėl

$$Z_1^* \sim N(0, 1),$$

bet

$$Z_1^* = Q + \frac{1}{\sqrt{3}}V + \varepsilon^*,$$

čia Q nepriklauso nuo V ir ε^* , ε^* nepriklauso nuo V ir

$$Q \sim N\left(0, \frac{1}{3}\right)$$

$$\varepsilon^* \sim N\left(0, \frac{1}{3}\right)$$

Gauname

$$\text{cov}(Z_1^*, V) = \frac{1}{\sqrt{3}}$$

taip, kad Z_1^* yra pasiskirstęs pagal normalųjį dėsnį $(0, 1)$ kaip ir Z_1 , bet dabar jis jau yra endogeninis kintamasis.

Apibrėžkime

$$I^* = 0.5Z_0 + Z_1^* + V$$

Parašykite funkciją generuojančią naujus duomenis D^* , Z_0 ir Z_1^* .

7. Pakartokite su jais probit didžiausio tikėtino sprendimą.
8. Jeigu viską atlikote teisingai, koeficientų įverčiai turėtų būti apie 0.6 ir 2, nors modelis yra generuotas su koeficientais 0.5 ir 1. Paaiškinkite iš kur atsirado skirtumas.
9. Tegu

$$\widehat{V} \sim 0.5N(\sqrt{0.5}, 0.5) + 0.5N(-\sqrt{0.5}, 0.5),$$

čia sudedamos dvi pasiskirstymo funkcijos. Taigi V yra dviejų normaliųjų dydžių mišinys, su dispersija 1 ir vidurkiu 0. Tegu Z 'ai yra egzogeniški ir

$$\widehat{I} = 0.5Z_0 + Z_1 + \widehat{V}$$

bei

$$\widehat{D} = 1(\widehat{I} > 0).$$

Parašykite funkciją generuojančią šiuos duomenis.

10. Kas nutiks kai šitam modeliui pritaikysime probit? Kiek svarbus yra tikrasis paklaidų pasiskirstymas?
11. Įvertinkite sukurtus modelius (iš viso buvo 3 modeliai) naudodami logit ir tiesinės tikimybės modelius. (Naudokite savo statistinio paketo galimybes). Kiek skiriasi rezultatai? Ar vis dėl to svarbu, kad duomenų pasiskirstymas sutaptų su teorinio modelio?