

# 1 Binarinis pasirinkimas

2006/2007 m.m. rudens semestras

Tegu

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

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

čia  $Z$  ir  $\gamma$  gali būti vektoriai. Stebimi dydžiai yra  $D$  ir  $Z$ . Tarkime, kad

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

ir

$(Z, V)$  tarpusavyje nepriklausomi.

1. Išveskite didžiausio tikėtinumo funkciją šios problemos sprendimui. Kas šioje funkcijoje bus nežinomas parametras, o kas bus turimi duomenys?
2. Paašškinkite ir pademonstruokite, kodėl reikalinga prielaida  $\sigma_V^2 = 1$  (ar kokiai kitokiai konstantai, nebūtinai vienetui).
3. Naudodamiesi savo pasirinktu statistiniu paketu parašykite funkciją, kuri gražina šitos problemos didžiausio tikėtinumo funkcijos reikšmes, duotiems duomenimis ir konkrečiai nežinomo parametro reikšmei.
4. 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

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

Parašykite funkciją generuojančią tokius duomenis.

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? Paašškinkite. (Jei viską teisingai atlikote  $\gamma_0$  įvertis turėtų būti apie 0.5, o  $\gamma_1$  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  $\varepsilon^*$ ,  $\varepsilon^*$  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

$$\begin{aligned} I^* &= 0.5Z_0 + Z_1^* + V \\ D^* &= I(I^* > 0) \end{aligned}$$

Parašykite funkciją generuojančią naujus duomenis  $D^*$ ,  $Z_0$  ir  $Z_1^*$ . Sugeneruokite duomenis ir pademonstruokite, kad  $Z_1^*$  yra tikrai endogeninis kintamasis.

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

8. 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 normalių dydžių mišinys, su dispersija 1 ir vidurkiu 0. Parodykite tai. Parašykite funkciją generuojančią  $\widehat{V}$  ir nubrėžkite jo histogramą.

9. Tegu  $Z$ 'ai yra egzogeniški (kaip ir 4 užduotyje) 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?